Investigating the diagnostic potential of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy: from research to clinic.

Davina Gillian Gale BSc. (Hons)

PhD by Publication

University of East Anglia

Faculty of Medicine and Health Sciences

March 2019

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Recent advances in oncology have led to the development of targeted therapies, enabling patients to be treated based on their tumour molecular profile. Whilst tumour biopsies are routinely used for profiling, they can be highly invasive, may not fully reflect the heterogeneity present within the tumour mass, or accurately represent the genomic profile as the tumour evolves over time. Recent interest has focussed on the use of circulating tumour DNA (ctDNA) as a non-invasive 'liquid biopsy'. Cell-free ctDNA, released from cancer cells, is highly fragmented, and carries the same genetic modifications present in the originating tumour, so has potential to be an exquisitely specific biomarker.

This thesis will focus on research I have performed over the last decade to investigate the diagnostic potential of ctDNA. I assessed the hypothesis that ctDNA is a clinically useful biomarker, able to correlate with disease burden, monitor tumour dynamics, and be used to guide treatment. I developed novel digital PCR and next generation sequencing (NGS) assays for the highly sensitive detection of ctDNA, and led the development and analytical validation of a clinical diagnostic ctDNA test to ISO15189:2012 regulatory standards, which is now being used in the clinic to stratify advanced non-small cell lung cancer patients to treatment. This thesis involves critical analysis of 14 publications that I have co-authored investigating the use of ctDNA in high-grade serous ovarian, breast and lung cancer, and the development of novel methods to improve sensitivity of detection.

When I started this work in 2009, very little was known about the clinical relevance of ctDNA. Since this time, work by myself and others has led to an explosion of interest in this area, leading to significant advances in the use of ctDNA for cancer diagnosis, treatment selection, patient monitoring and detection of minimal residual disease.

Table of Contents

Page

Abstract………3

List of Figures………………………………………………………………………………………………11

List of Tables………………………………………………………………………………………………15

List of accompanying material……………………………………………………………………17

Acknowledgements………………………………………………………………………………………23

Chapter 1: Introduction………………………………………………………………………………27

1.1: Cancer and tumour heterogeneity…………………………………………………27

1.2: Targeted therapies, 'personalised medicine' & companion diagnostics……………………………………………………………………………………28

1.3: Clinical diagnostic assay development…………………………………………30

1.4: Overview of ctDNA……………………………………………………………………32

1.5: Development of digital PCR for ctDNA analysis……………………………34

1.6: Development of next generation sequencing for ctDNA analysis…………36

1.7: Thesis - Aims and objectives………………………………………………………39

Chapter 2: Development of digital PCR for ctDNA analysis in ovarian cancer………41

2.1: High-grade serous ovarian cancer…………………………………………………41

2.2: Parkinson*, Gale* et al., PLoS Medicine, 2016……………………………………41

2.2.1: Aims……………………………………………………………………………41

2.2.2: Results…………………………………………………………………………42

2.2.3: Monitoring tumour dynamics using *TP53* ctDNA ……………………42

2.2.4: Correlation of *TP53* ctDNA with tumour volume……………………44

2.2.5: Assessment of *TP53* ctDNA as an early response marker…………44

Page

2.2.6: **Discussion**……47

2.3: **Schwarz et al., PLoS Medicine, 2015**……52

2.3.1: **Aims**……52

2.3.2: **Results**……53

2.3.3: **Discussion**……53

Chapter 3: Development of NGS for ctDNA analysis - TAm-Seq & exome sequencing……55

3.1: **Forshew*, …, Gale* et al., Science Translational Medicine, 2012**……55

3.1.1: **Aims**……55

3.1.2: **Results**……55

3.1.3: **Discussion**……58

3.2: **Murtaza et al., Nature, 2013**……64

3.2.1: **Aims**……64

3.2.2: **Results**……64

3.3: **Murtaza et al., Nature Communications, 2015**……68

3.3.1: **Aims**……68

3.3.2: **Results**……68

3.4: **Discussion**……72

Chapter 4: Clinical diagnostic ctDNA assay development……75

4.1: **Risberg, ..., Gale† et al., Journal of Molecular Diagnostics, 2018**……75

4.1.1: **Aims**……75

4.1.2: **Results**……76

4.1.3: **Module 1: Assessment of the effects of delayed processing**……76

Page

4.1.4: Module 2: Assessment of the effects of different storage temperatures........76

4.1.5: Module 3: Assessment of different blood collection tubes..........76

4.1.6: Module 4: Assessment of different centrifugation protocols.......79

4.1.7: Module 5: Assessment of the effects of sample shipment...........79

4.1.8: Discussion........79

4.2: Gale[‡] et al., PLoS ONE, 2018........81

4.2.1: Aims........81

4.2.2: Results........82

4.2.3: Analytical Performance of eTAm-Seq™: SNVs........82

4.2.4: Analytical Performance of eTAm-Seq™: Specificity........84

4.3: Plagnol et al., PLoS ONE, 2018........84

4.3.1: Aims........84

4.3.2: Results........88

4.3.3: Analytical Performance of InVisionFirst™-Lung: SNVs & indels....88

4.3.4: Analytical Performance of InVisionFirst™-Lung: SVs........88

4.3.5: Analytical Performance of InVisionFirst™-Lung: SCNAs........88

4.3.6: Analytical Performance of InVisionFirst™-Lung: Specificity..........90

4.4: Discussion........90

4.4.1: Challenges of analytical validation of an NGS diagnostic assay......90

4.4.2: Clinical impact of eTAm-Seq™ and the InVisionFirst™-Lung assay........93

Chapter 5: Clinical applications of ctDNA........95

5.1: Molecular stratification........95

Page

5.1.1: **Remon et al., Annals of Oncology, 2017**……………………………………95

5.1.1.1: **Aims**………………………………………………………………………………95

5.1.1.2: **Results**……………………………………………………………………………95

5.1.1.3: **Discussion**………………………………………………………………………96

5.2: **Monitoring treatment response**…………………………………………………97

5.2.1: **Dawson et al., New England Journal of Medicine, 2013**……………98

5.2.1.1: **Aims**………………………………………………………………………………98

5.2.1.2: **Results**……………………………………………………………………………98

5.2.2: **Tsui et al., EMBO Molecular Medicine, 2018**……………………………99

5.2.2.1: **Aims**………………………………………………………………………………99

5.2.2.2: **Results**……………………………………………………………………………99

5.2.3: **Discussion**………………………………………………………………………105

5.3: **Early detection of cancer and minimal residual disease**………………106

5.3.1: **Mair et al., Cancer Research, 2019**……………………………………………106

5.3.1.1: **Aims**………………………………………………………………………………106

5.3.1.2: **Results**……………………………………………………………………………107

5.3.2: **Mouliere et al., bioRxiv, 2017; Mouliere et al., Science Translational Medicine, 2018**……………………………………………………107

5.3.2.1: **Aims**………………………………………………………………………………107

5.3.2.2: **Results**……………………………………………………………………………107

5.3.3: **Discussion**………………………………………………………………………112

Chapter 6: **Summary**…………………………………………………………………………115

Definitions……………………………………………………………………………………………119

References……………………………………………………………………………………………127

Page

Appendix 1: **Summary of each submitted publication and details of my contribution**……………………………………………………………………………………..143

Appendix 2: **Letters from senior authors detailing my contribution to each publication**………………………………………………………………………………………….157

Appendix 3: **Full list of all author contributions to each submitted publication**……….213

Appendix 4: **Reproduction of each submitted publication**………………………………………..221

Appendix 5: **List of FDA-approved targeted therapies for solid malignancies and their molecular targets**……………………………………………………………………….415

Appendix 6: **List of FDA-cleared or approved companion diagnostic tests or devices**…………………………………………………………………………………………………..421

List of Figures

Page

1.1: **Total frequency of somatic mutations in exome of different cancer types per megabase** ……29

1.2: **2x2 table for reporting results comparing a test outcome with a reference standard**……31

1.3: **Origins and types of alterations in circulating tumour DNA (ctDNA)**……33

1.4: **Difference in recurrence-free survival in colorectal cancer patients with detected versus undetected post-operative ctDNA and CEA**……35

1.5: **ctDNA levels in patients with different advanced cancers**……37

1.6: **Levels of ctDNA detected at different stages of disease**……38

2.1: **Schematic workflow for mutant *TP53* ctDNA analysis in HGSOC**……43

2.2: **Monitoring ctDNA and CA-125 kinetics during chemotherapy**……45

2.3: **Comparison of TP53MAF and tumour volume**……46

2.4: **ROC curve and Kaplan-Meier plot for decrease in TP53MAF after one cycle of chemotherapy, excluding courses with ascitic drains**……48

2.5: **Concordance between tumor volume and pretreatment pg/mL ctDNA**…50

2.6: **Tumour volume correlation with mean clonal plasma VAF**……51

3.1: **Overview of TAm-Seq**……56

3.2: **Identification of mutations in ovarian cancer FFPE samples by TAm-Seq**……58

3.3: **Retrospective analysis by TAm-Seq of plasma collected during patient follow-up and biopsy specimens collected at initial surgery**……59

3.4: **Monitoring mutant DNA in plasma of HGSOC and breast cancer patients over time**……60

Page

3.5: Retrospective analysis of samples from synchronous primary tumours (ovarian and bowel) and plasma collected at relapse……………….61

3.6: Analysis of cfDNA using dPCR, targeted sequencing and WGS……………….63

3.7: Overview of workflow for the identification of treatment-associated mutational changes by exome sequencing of plasma…………………………….65

3.8: Genome-wide analysis of exome data from plasma and synchronous metastatic tumour DNA…………………………………………………………………….66

3.9: Mutations showing evidence of genomic tumour evolution…………………..67

3.10: Inference of clonal structure from multiregional tumour biopsies…………69

3.11: Distribution of validated mutations in tumour and plasma samples………70

3.12: Mutant AFs of stem, metastatic-clade and private mutations during treatment……………………………………………………………………………………….71

4.1: Effect of storage temperature on (A) cfDNA (B) AF (C) ctDNA levels following delayed storage at RT/4°C prior to plasma processing….77

4.2: Effects of using different blood tubes on (A) cfDNA levels (B) mutant AF in K_3EDTA tubes; (C) Background error rates …………………..78

4.3: InVision™ liquid biopsy tumour profiling panel……………………………………..83

4.4: Sensitivity and inter-/intra-operator variability of eTAm-Seq™ using low, medium and high input DNA………………………………………………………………..86

4.5: Allele Fractions determined by analysis with eTAm-Seq™ and dPCR……..87

4.6: Overview of the InVisionFirst™ workflow……………………………………………….89

4.7: Sensitivity of InVisionFirst™ for detection of SNVs and indels………………..91

5.1: Waterfall plot of p.T790M-positive NSCLC patients treated with osimertinib, showing best percentage change in lesion size…………………..96

Page

5.2: Monitoring multiple SNVs and SVs in cfDNA from breast cancer patients……..100

5.3: Comparison of ctDNA, CTCs, CA15-3 and imaging data to monitor tumour dynamics and predict survival…….101

5.4: ctDNA dynamics in 45 NSCLC patients reveals distinct patterns of resistance mechanisms……..103

5.5: Global SCNA profiles in primary tumour and plasma from patient who underwent histological transformation to SCLC……104

5.6: Fragmentation patterns of plasma DNA in rat xenograft model of glioblastoma……108

5.7: Plasma fragmentation patterns on a pan-cancer scale……..110

5.8: Enhancing tumour fraction from plasma sequencing with size selection…….111

List of Tables

Page

4.1: eTAm-Seq™ analytical validation experiments……………………………………….83

4.2: Sensitivity of eTAm-Seq™ using median input DNA at different AFs tested in two independent laboratories…………………………………………..85

List of accompanying material

My supervisors were Dr. Barbara Jennings and Dr. Daniel Brewer at the University of East Anglia. I can confirm that I have completed all required courses.

This thesis is based on critical analysis of the following 14 publications that I have co-authored, presented in the order cited in Chapters 1 - 6. **My publications are highlighted in bold in the thesis text,** and reproduced in Appendix 4, including full results and discussion. The length of the critical analysis is 19,989 words.

Key

*** Joint first co-author**

‡ First author and senior co-corresponding author

† Senior co-corresponding author

Note

Appendix 1:	Summary of each submitted publication and details of my contribution.
Appendix 2:	Letters from senior authors detailing my contribution to each publication.
Appendix 3:	Full list of all author contributions to each submitted publication.
Appendix 4:	Reproduction of each submitted publication, with kind permission from the journals.

List of submitted publications

1. **Parkinson C*, Gale D*, Piskorz A, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, Earl H, Qian W, Rosenfeld N†, Brenton JD†; *PLoS Medicine*; 2016; 13 (12): e1002198 doi:10.1371/ journal. pmed.1002198; 'Exploratory Analysis of *TP53* Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study.' Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (1).**

2. Schwarz R, Ng C, Cooke S, Newman S, Temple J, Piskorz A, Gale D, Sayal K, Murtaza M, Baldwin P, Rosenfeld N, Earl H, Sala E, Jimenez-Linan M, Parkinson C, Markowetz F[†], Brenton JD[†]; *PLoS Medicine*; 2015; Feb 24;12(2): e1001789; ‘Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis.’ Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (2).

3. Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DWY*, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD[†], Rosenfeld N[†]; *Science Translational Medicine*; 2012; 4(136):136ra68; ‘Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.’ Publication and Figures reprinted with kind permission of the journal (3).

4. Murtaza M*, Dawson SJ*, Tsui DWY*, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin S, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD[†], Caldas C[†], Rosenfeld N[†]*; Nature*; 2013 May 2; 497(7447):108-12. ‘Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.’ Publication and Figures reprinted with kind permission of the journal (4).

5. Murtaza M*, Dawson SJ*, Pogrebniak K, Rueda O, Provenzano E, Grant J, Chin SF, Tsui D, Marass F, Gale D, Ali HR, Shah P, Contente-Cuomo T, Farahani H, Shumansky K, Kingsbury Z, Humphray S, Bentley D, Shah S, Wallis M, Rosenfeld N[†], Caldas C[†]; *Nature Communications*; 2015; 6:8760 doi: 10.1038/ncomms9760 ‘Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer’. Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (5).

6. Risberg B*, Tsui DWY*[†], Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, Brenton JD, Gale D[†]; *The Journal of Molecular Diagnostics*, 2018; *Volume 20, (6), 883 - 892*. *Senior co-corresponding author; ‘Effects of collection and processing procedures on plasma circulating

cell-free DNA from cancer patient'. Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (6).

7. **Gale D‡, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G, Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale A, Huggett J, Foy CA, Jones G, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N†; *PLoS ONE*; 2018; 13(3): e0194630; 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.' Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (7).**

8. **Plagnol V*, Woodhouse S*, Howarth K, Lensing S, Smith M, Epstein M, Madi M, Smalley S, Leroy C, Hinton J, De Kievit F, Musgrave-Brown E, Herd C, Neblett K, Brennan W, Dimitrov P, Campbell N, Morris C, Rosenfeld N, Clark J, Gale D, Platt J, Calaway J, Jones G, Forshew T†; *PLoS ONE*; 2018; 13(3): e01938022018; 'Analytical validation of a Next Generation Sequencing liquid biopsy assay for high sensitivity broad molecular profiling.' Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (8).**

9. **Remon J, Caramella C, Jovelet C, Lacroix L, Lawson ARJ, Smalley S, Howarth K, Gale D, Green E, Plagnol V, Rosenfeld N, Planchard D, Bluthgen MV, Gazzah A, Pannet C, Nicotra C, Auclin E, Soria JC, Besse B†; *Annals of Oncology*; 2017; 28:784-790; 'Osimertinib benefit in EGFR-mutant NSCLC patients with T790M mutation detected by circulating tumour DNA.' Publication and Figures reprinted with kind permission of the journal (9).**

10. **Dawson SJ*, Tsui DWY*, Murtaza M, Biggs H, Rueda O, Chin SF, Dunning M, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C†, Rosenfeld N†; *New England Journal of Medicine*; 'Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer.' Publication and**

Figures reproduced with permission from N Engl J Med 2013; 368:1199-1209; DOI:10.1056/NEJMoa1213261, Copyright Massachusetts Medical Society (10).

11. Tsui DWY*, Murtaza M*, Wong ASC, Rueda OM, Smith CG, Chandrananda D, Soo RA, Lim H, Goh B, Caldas C, Forshew T, Gale D, Liu W, Morris J, Marass F, Eisen T, Chin T†, Rosenfeld N†; *EMBO Molecular Medicine*; 2018; e794; DOI 10.15252/ emmm. 201707945; ‘Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in NSCLC.’ Publication and Figures reprinted with kind permission from the journal (Creative Commons Attribution License) (11).

12. Mair R*, Mouliere F*, Smith CG, Chandrananda D, Gale D, Marass F, Tsui DWY, Massie CE, Wright AJ, Watts C, Rosenfeld N†, Brindle KM†; *Cancer Research;* 2019; 79 (1): 220 LP-230; https://doi.org/10.1158/0008-5472.CAN-18-0074; “Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models.’ Publication and Figures reprinted with kind permission from AACR (12).

13. Mouliere F*, Piskorz A*, Chandrananda D*, Moore E*, Morris J, Smith C, Goranova T, Heider K, Mair R, Supernat A, Gounaris I, Ros S, Wan J, Jimenez-Linan M, Gale D, Brindle K, Massie C, Parkinson C, Brenton JD†, Rosenfeld N†; *bioRxiv*; 2017; *(Pre-print, not peer-reviewed)*; ‘Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA.’ Publication and Figures reprinted with kind permission of the senior authors (13).

14. Mouliere F*, Chandrananda D*, Piskorz A*, Moore E*, Morris J, Barlebo Ahlborn L, Mair R, Goranova T, Marass F, Heider K, Wan J, Supernat A, Hudecova I, Gounaris I, Ros S, Jimenez-Linan M, Garcia-Corbacho J, Patel K, Østrup O, Murphy S, Eldridge M, Gale D, Stewart G, Burge J, Cooper W, van der Heijden M, Massie C, Watts C, Corrie P, Pacey S, Brindle K, Baird R, Mau-Sørensen M, Parkinson C, Smith C , Brenton JD†, Rosenfeld N†; *Science Translational Medicine*; 2018; Vol. 10, Issue 466, eaat4921; DOI: 10.1126/scitranslmed.aat4921; ‘Enhanced detection of circulating tumor DNA by fragment size analysis’. Publication and Figures reprinted with kind permission from AAAS (14).

Acknowledgements

First and foremost, I would like to thank Steve, my husband, for all his love and support, for encouraging me to do a PhD by Publication, and for being the best soul mate and travel companion I could have in life's many adventures! Special thanks also to my Mum and sister Nicky, for their love and support throughout the years, and for always being there.

In my career, I have had the honour and privilege to work with some amazingly talented people, working together as a true team on the most exciting and rewarding projects, which I hope has made an impact, however small, on patient healthcare. My most exciting times were in the start-up years at the Cancer Research UK Cambridge Institute and Inivata. Special thanks to Nitzan Rosenfeld, Tim Forshew, Muhammed Murtaza and Dana Tsui – there were only 5 of us, but together we had different yet complementary skills, and showed that you really can achieve amazing things if you are all working together on a common, rewarding and challenging goal if motivated for the right reasons. Thanks also to colleagues at the CRUK-CI, including past and present members of the Rosenfeld Lab, Jill Temple, Heather Biggs, James Brenton, Anna Piskorz, Sarah-Jane Dawson, the Genomics Core facility, all co-authors on my publications and all patients who have kindly donated samples.

Particular thanks to Nitzan for giving me the opportunity to realise my dream and take a genomic research test to a clinical diagnostic assay. I was also involved in the naming and branding of the InVision™ assay, which really was the best bit! It has been an honour to have had the opportunity to develop a ctDNA assay which has been of benefit to cancer patients. Many thanks to John Wells, Director of Operations at the CRUK-CI, who supported my move to Inivata and to do a PhD by Publication. At the start-up of Inivata, I was fortunate to work with a great group of talented and motivated people. Special thanks to Andrew R.J. Lawson - at the beginning, we were the only 2 full-time staff, and it was a privilege to work together. To Janice, Amanda, Andrew, Karen, Sam and Michael - thank you for your friendship.

Special mention to James Hadfield, who encouraged me to do a PhD by Publication at the University of East Anglia. He kindly put me in touch with Dr. Barbara Jennings, the perfect supervisor for me, with specialist knowledge in the area of personalised medicine and clinical diagnostics. Many thanks to Barbara for agreeing to being my supervisor, to my second supervisor, Dr. Daniel Brewer, and to the UEA Postgraduate Research Office for supporting my application.

Finally, and most importantly, I would like to dedicate this thesis and achievement to my Dad, Grandad, Grannie, Auntie Madge, Auntie Bebe, Auntie Olive, and the many members of my family who are sadly no longer with us, but will always remain in loving memory. Thank you for all the wonderful memories when I was growing up, and for teaching me what is important in life. I know you would be very proud of this achievement.

Chapter 1: Introduction

1.1: Cancer and tumour heterogeneity

Cancer is the second leading cause of mortality worldwide, responsible for an estimated 9.6 million deaths in 2018 (15). It is caused by abnormal cell growth and loss of normal control mechanisms, which result in uncontrolled cell proliferation and formation of a tissue mass, or tumour. A malignant tumour can invade normal adjacent tissue, and spread or metastasize to other parts of the body, eventually leading to death. The characteristic hallmarks of cancer include the acquisition of multiple biological processes during tumour development, including angiogenesis, the growth of new blood vessels that supply nutrients to tumour cells, and genome instability (16,17). Cancer is a heterogenous group of over 200 different diseases, which may be divided into five broad subtypes - carcinomas, lymphomas, leukaemias, brain tumours and sarcomas, depending on cell of origin. To assess the advancement of disease, cancer is often classified into different stages, with Stage I involving a small, localised tumour, Stage II involving larger tumour growth and possible spread to the lymph nodes, Stage III involving spread to additional tissues and lymph nodes, and Stage IV where the cancer has metastasized to additional organs or other parts of the body. The earlier the cancer is detected and treated, the greater the chance of patient survival. In ovarian cancer, the 5-year survival rate is 90% at Stage I, compared to just 3.5% if diagnosed at Stage IV (18).

Cancer is a disease of the genome, caused by germline and somatic mutations in the genetic sequence (19). Somatic mutations may occur by substitution of single DNA bases, resulting in single nucleotide variants (SNVs) or point mutations, by insertion or deletion of bases (indels), or by duplication, deletion, inversion or translocation of large chromosomal regions (structural variants, SVs). Alternatively chromosomal regions may be amplified or deleted, resulting in somatic copy number alterations (SCNAs). Furthermore, epigenetic modifications, such as DNA methylation or histone modifications, can affect chromatin structure and gene expression patterns, which are then replicated during cell division. Mutations may arise following exposure to carcinogens such as tobacco smoke, radiation or ultraviolet sunlight, or may be acquired as a failure of the DNA repair process during DNA replication. The probability of errors occurring are low, but accumulate during the ageing process. Many of these mutations are silent, or “passenger” mutations, with no physiological effect, but other “driver” mutations may incur a selective advantage, driving cell division and growth, thereby accumulating additional somatic mutations at a higher rate. Certain driver mutations are prevalent in specific cancers. For example, *BRAF* p.V600E mutations are common in melanoma (corresponding to a single point mutation in the *BRAF* gene resulting in a valine (V) to glutamic acid (E) amino acid substitution at codon 600), whereas mutations in *TP53*, a tumour-suppressor gene, are more ubiquitous, found in the majority of cancers, with mutations occurring throughout the gene, predominantly in exons 5-8.

In 1990, Vogelstein proposed that tumourigenesis occurs by an evolutionary process resulting from sequential accumulation of mutations in ‘oncogenes’, where mutations lead to a gain-of-function which drives the oncogenic process, and in ‘tumour suppressor genes’, where mutations lead to loss-of-function (20). Tumour mutation burden (TMB) in the exome [corresponding to the number of somatic mutations in expressed exonic regions of the genome per megabase (Mb)] has been found to differ in different cancer types (Figure 1.1). The carcinogen-induced cancers melanoma and lung cancer have the highest mutational burden, induced by UV irradiation and tobacco respectively. Recent evidence suggests that TMB may predict clinical response to immune checkpoint inhibitors across different cancers, with patients with higher somatic TMB associated with better overall survival (21).

Different selective pressures can be exerted by the tumour microenvironment, depending on the location of tumour cells. As a result, tumours evolve and can be highly heterogeneous, composed of different subclones with different genomic profiles within a single tumour mass, or in different metastatic deposits. Treatment with different drugs can also drive tumour evolution, and may result in the acquisition of additional mutations leading to the development of treatment resistance.

1.2: Targeted therapies, ‘personalised medicine’ & companion diagnostics

In recent years, knowledge of the tumour molecular profile has led to the development of targeted therapies which are effective in patients with specific somatic mutations. Appendix 5 shows a list of targeted therapies approved by the US Food and Drug Administration (FDA) for solid malignancies, together with their molecular targets (22). Vemurafinib, for example, is beneficial in melanoma patients with a *BRAF* p.V600E mutation, and tyrosine kinase inhibitors (TKIs), such as gefitinib and erlotinib, are effective

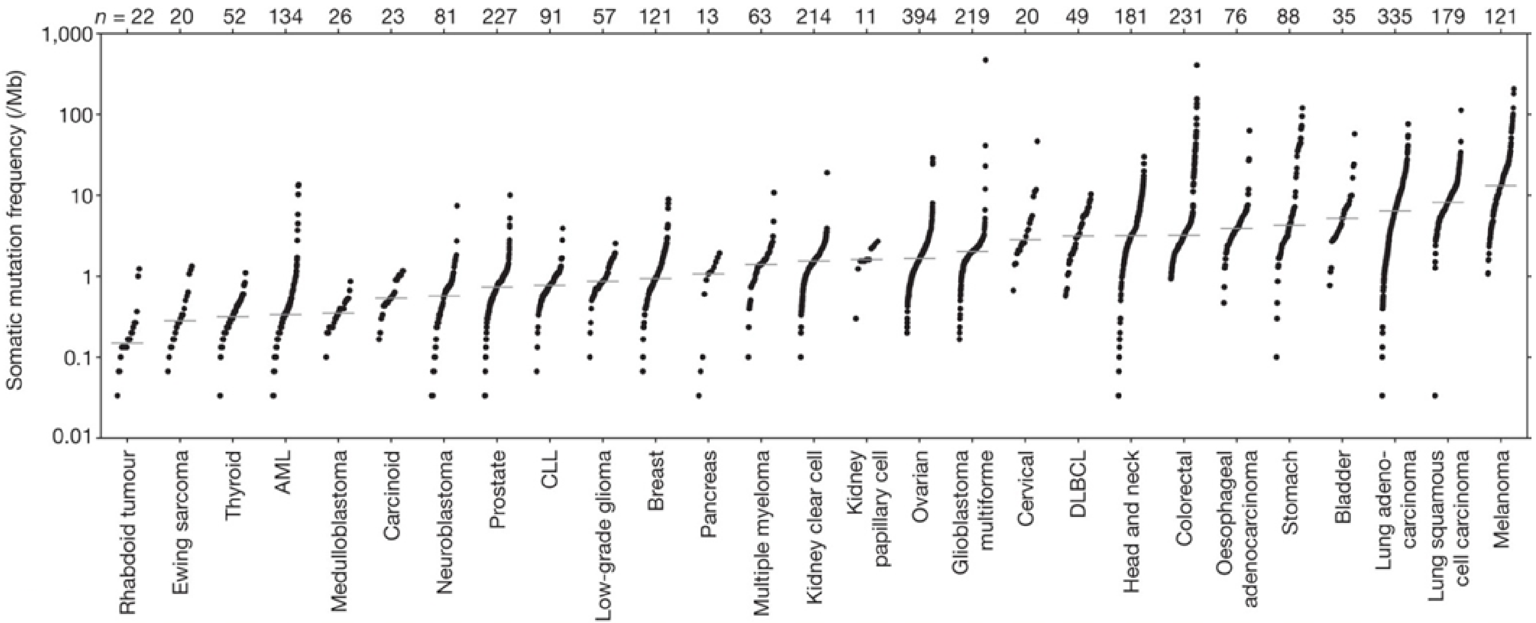


Figure 1.1: Total frequency of somatic mutations in exome of different cancer types per megabase (Mb).

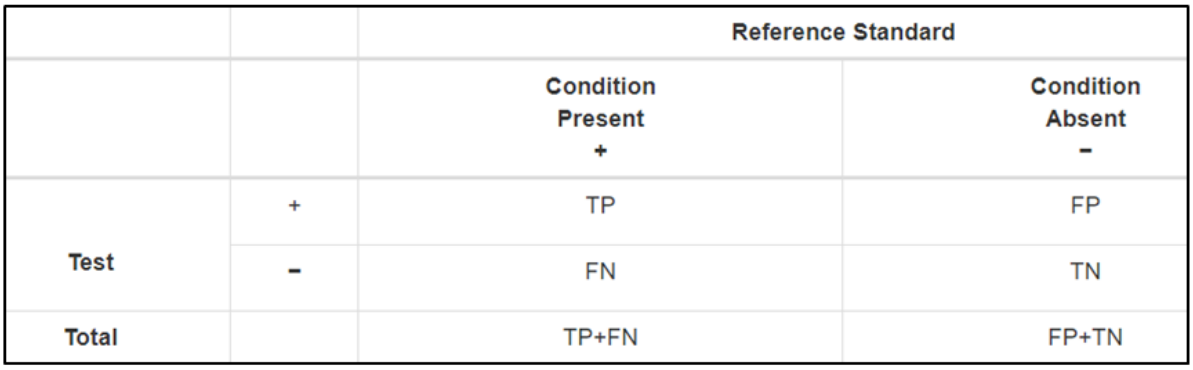
Dot: tumour-normal pair. Tumour types ordered by median somatic mutation frequency. Figure and Figure Legend adapted from Lawrence et al.(23)

in non-small cell lung cancer (NSCLC) patients with known activating *EGFR* exon 19 deletions or p.L858R mutations. In 50% of NSCLC patients, TKI resistance is caused by an *EGFR* p.T790M mutation, which affects the ATP binding pocket of the EGFR kinase domain (24). Third generation TKIs, such as osimertinib, have now been developed which are effective in patients with p.T790M resistance mutations. Tailoring treatment has led to the concept of 'personalised medicine' or 'precision medicine', which takes into account individual variability to stratify patients to the most appropriate treatment based on their tumour molecular profile. Where a drug treatment is available for a patient with a specific tumour mutation, the mutation is considered 'actionable', with potential downstream clinical benefit for that patient.

In recent years, different 'companion diagnostics' (CDx) tests, essential for the safe and effective use and prescription of a particular drug, and 'complementary diagnostics', which aid the therapeutic decision process, have been developed to determine which patients would benefit from a particular drug treatment. A list of CDx tests and devices which have been approved by the FDA is given in Appendix 6 (25). Diagnostic test methods include PCR (polymerase chain reaction) to amplify and sequence specific targeted regions of interest, immuno-histochemistry (IHC), using antibody staining to visualise the presence of proteins of interest, and fluorescent *in situ* hybridisation (FISH), using fluorescent probes that bind to complementary chromosomal regions of interest to assess chromosomal breakpoints.

1.3: Clinical diagnostic assay development

The ideal clinical diagnostic assay has high sensitivity and specificity. To calculate these, for example when using a test to identify patients with and without a health condition, there are four possible outcomes to be taken into consideration (Figure 1.2), involving the identification of:
a) *true positives (TP),* where the test correctly identifies a patient has the condition;
b) *false positives (FP)*, where the patient does not have the condition, but the test is positive;
c) *true negatives (TN)*, where the patient does not have the condition, and the test is negative;
d) *false negatives (FN)*, where the patient has the condition, but the test is negative.
The sensitivity of a clinical test is calculated as TP/(TP + FN). Specificity refers to the ability of the test to correctly identify individuals without the condition, calculated as TN/(TN + FP). The Positive Predictive Value (PPV) determines how likely it is that a patient has the condition given that the test result is positive [TP/(TP + FP)], with the Negative Predictive Value (NPV) calculated as TN/(TN + FN). The LoD90 of a test is the limit of detection where the test analyte can be detected 90% of the time.



		Reference Standard	
		Condition Present +	Condition Absent −
Test	+	TP	FP
	−	FN	TN
Total		TP+FN	FP+TN

Figure 1.2: 2x2 table for reporting results comparing a test outcome with a reference standard. *TP: true positive; FP: false positive; TN: true negative; FN: false negative. Table and Table Legend adapted from US FDA statistical guidelines* (26)

To ensure patient safety, a diagnostic test must be performed to rigorous quality standards to ensure results are reliable and can be used for clinical decision-making. Prior to implementation in an NHS (UK National Health Service) or clinical diagnostic lab, a research assay must be performed to high quality management standards, such as ISO15189:2012, the International Organisation for Standardization framework for medical laboratories. ISO requirements include using well-maintained equipment, ensuring staff are adequately trained, standard operating procedures are version-controlled, and assay, computational and data analysis processes are fully validated. In the US, the Centers for Medicare & Medicaid Services (CMS) regulates laboratory testing to CLIA (Clinical Laboratory Improvement Amendments) standards, requiring clinical laboratories to be certified before human samples can be accepted for diagnostic testing (27).

Clinical diagnostic test validation involves three stages - analytical validation, clinical validation, and the assessment of clinical utility. Analytical validation verifies a test is suitable for its intended use, and involves in-depth analysis of the analytical performance of the assay (eg. sensitivity, specificity, LoD90 etc.) to ensure the reliability of analytical test results. Clinical validation assesses how consistently a test is able to detect or predict diagnostic information in the target patient population. Finally, clinical utility assesses whether the test can be used to improve clinical outcomes for a patient.

Currently the majority of clinical cancer diagnostic tests require tumour biopsy material for analysis. Whilst knowledge of a cancer patient's tumour molecular profile can help select appropriate treatment, sampling of a tissue biopsy is highly invasive and expensive. In addition, a single biopsy may not fully reflect the heterogeneity present within the tumour mass or metastatic deposits, or accurately represent the genomic profile as the tumour evolves over time or in response to treatment, representing only a single 'snapshot' in time. Recently, much interest has focussed on circulating tumour DNA (ctDNA) as a non-invasive 'liquid biopsy'. With its ease of collection, ctDNA has potential to enable multiple repeat sampling and more effectively guide treatment decisions as the tumour evolves (28). However, there are considerable technical challenges associated with this approach, as detailed in the next sections.

1.4: Overview of ctDNA

Since 1948, it has been known that fragmented DNA is present in blood (29). Levels were shown to be higher in serum from cancer patients compared to healthy controls (30), and to partly originate from cancer cells (31). In 1994, circulating tumour DNA carrying tumour-specific mutations in the *KRAS* gene, present as a sub-fraction of the total cell-free DNA (cfDNA), was first identified in plasma from pancreatic cancer patients (32). This opened up the possibility to use plasma as a non-invasive source of diagnostic information to guide cancer treatment. The precise mechanism of release of cfDNA into the bloodstream is relatively unknown, although believed to be largely due to apoptosis (programmed cell death), in addition to necrosis (premature cell death) and active secretion from cells in the tumour microenvironment, haemopoietic and immune system and other organs (33). cfDNA may be free, or associated with nucleosomes or extracellular vesicles, such as exosomes (34). Cell-free and tumour-specific DNA has been detected in other bodily fluids, including urine, saliva, ascites and cerebrospinal fluid (CSF), and foetal DNA is known to be present in maternal plasma (28,35). ctDNA, originating from a tumour, carries the same genetic and epigenetic modifications present in that tumour, as depicted in Figure 1.3 (28). For this reason, ctDNA has potential to be an exquisitely-specific biomarker for use in cancer diagnosis, treatment selection, monitoring, and detection of minimal residual disease (MRD).

The half-life of cfDNA in blood is believed to be short, between 16 – 150 minutes, and may be cleared from the bloodstream and excreted by the kidneys into urine, or degraded by macrophages in the liver and spleen (28,36). Plasma cfDNA is highly fragmented, with a mode around 167bp in length, which corresponds to the length of DNA wrapped around a 'chromatosome' - a nucleosome

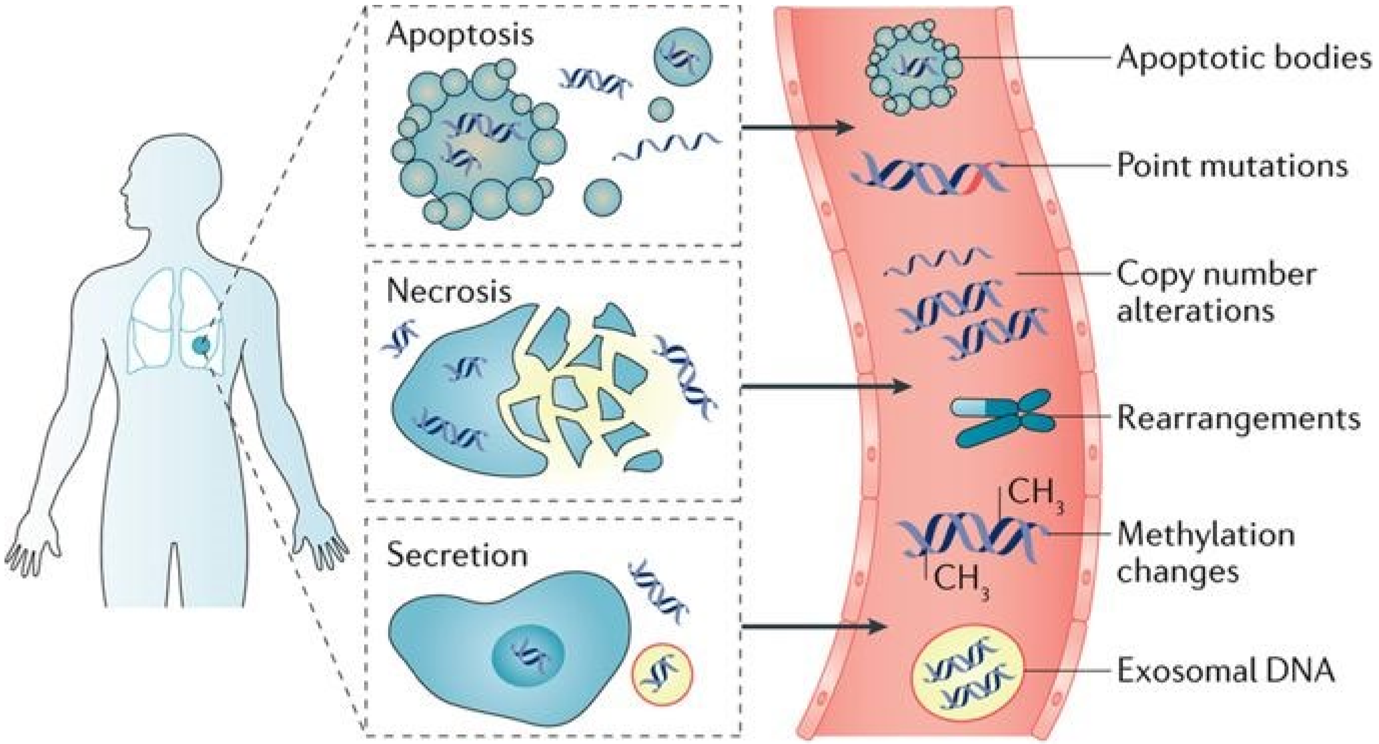

Apoptosis
Necrosis
Secretion
Apoptotic bodies
Point mutations
Copy number alterations
Rearrangements
CH_3
CH_3
Methylation changes
Exosomal DNA

Figure 1.3: Origins and types of alterations in circulating tumour DNA (ctDNA). *Figure and Figure Legend adapted from Wan et al.* (28)

histone core (~147bp), plus a 20bp linker region associated with histone H1 (37). Fragment lengths of 167bp are believed to largely result from caspase-dependent endonuclease cleavage during apoptosis (38). Fragmentation patterns of cfDNA in blood show a distinct 10bp periodic ladder, possibly caused by nuclease-cleavage at exposed sites at each turn of the DNA helix. In urine, cfDNA has a wide range of fragment sizes including very short fragments, possibly due to its harsh nuclease environment (28,39).

Recent evidence suggests that cfDNA and tumour-specific DNA differ in size, with ctDNA being shorter than non-mutant DNA centred around 147bp, corresponding to the length of DNA around the protected nucleosome, minus the linker region (12–14,37,38,40). The reasons for this shortening are unclear, although may result from modifications in chromatin organisation and compaction (41,42). Evidence suggests that cfDNA fragmentation may not be a random process, with specific genomic signatures present at fragment ends (43,44). Mice with a DNAse1l3 deletion (a DNase1-like nuclease which digests DNA in chromatin) have been shown to have aberrations in the fragmentation pattern, with an increase in longer nucleosomal fragments (44). In addition, these mice develop antibodies to chromatin and DNA, indicating that the enzyme may be involved in autoimmunity. Deletion of both DNAse1 and DNAse1l3 genes result in an increase in shorter fragment sizes <120bp.

1.5: Development of digital PCR for ctDNA analysis

Early research into ctDNA largely focussed on analysis of key single point mutations in driver genes that promote tumourigenesis. The challenge has been that ctDNA may only constitute a small fraction of the total cfDNA [down to <0.1% allele fraction (AF)], such that we can be searching for a ‘needle in a haystack’. As a result, highly sensitive methods are required to detect and quantify mutant alleles. Methods historically used include quantitative PCR (qPCR), allele-specific PCR and ARMS-PCR (amplification-refractory mutation system), using fluorescent probes for ‘relative’ quantification using standard reference DNA to estimate levels (45,46). Digital PCR (dPCR) was subsequently developed, which enables ‘absolute’ quantification of tumour-specific DNA by partitioning DNA molecules into multiple wells, microfluidic chambers or water-in-oil droplets, and performing qPCR using fluorescent probes to mutant or wild-type sequences. Unlike a single-tube qPCR reaction, where rare mutant molecules cannot be detected amongst high levels of background wild-type DNA, partitioning reduces background levels, enabling sensitive detection (47). BEAMing is a modified dPCR method that combines bead-based emulsion PCR and flow cytometry, using allele-specific fluorescent probes. In 2008, Diehl et al. used BEAMing to monitor tumour dynamics in 18 metastatic colorectal cancer (mCRC) patients (36), demonstrating for the first time that ctDNA had increased sensitivity compared to the protein biomarker CEA (carcinoembryonic antigen), and was detected in 78% patients compared to 56% for CEA. ctDNA was not detected post-surgery in four patients who did not progress over time, but was detected in patients who later progressed (Figure 1.4).

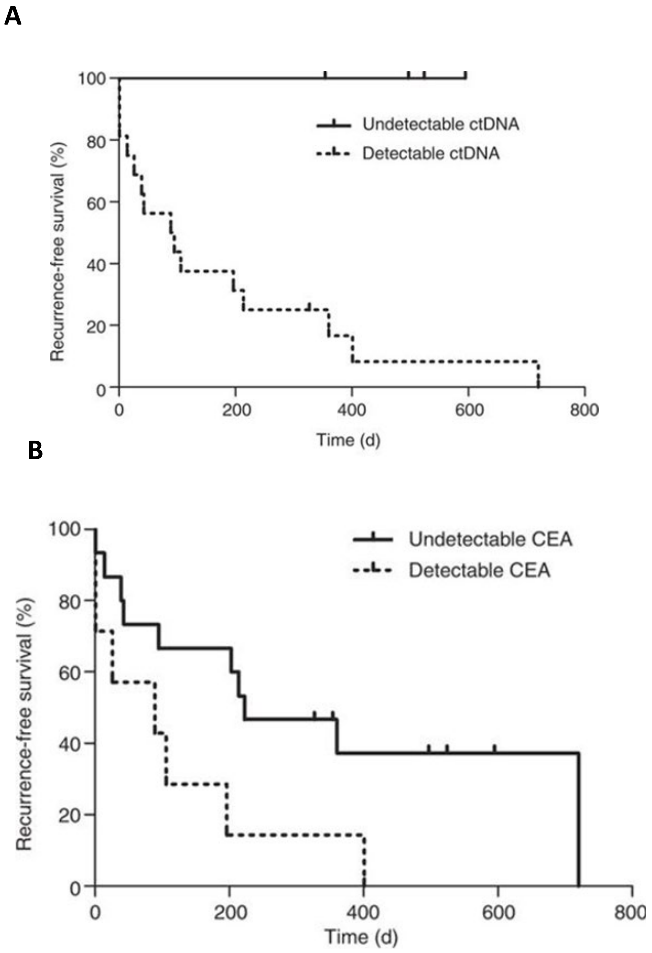


Figure 1.4: Difference in recurrence-free survival in colorectal cancer patients with detected versus undetected post-operative (A) ctDNA *($p=0.006$)*; **(B) CEA** (carcinoembryonic antigen, *$p=0.03$). Mantel-Cox log-rank test; d: days. Figures and Figure Legend adapted from Diehl et al.* (36)

Bettagowda et al. assessed ctDNA levels in 640 patients with different types of cancer, using BEAMing or a PCR-ligation reaction involving hybidisation of fluorescent mutant probes (48,49). Whilst the numbers of patients with each cancer type was low, ctDNA was detected in 75% of patients with advanced bladder, colorectal, gastroesophageal, ovarian, pancreatic, breast, melanoma, hepatocellular and head and neck cancer, yet in <10% of glioma patients and <50% of patients with thyroid, renal, prostate and medulloblastoma cancer (Figure 1.5). The lower levels detected in brain cancer may possibly reflect the influence of the blood brain barrier, although in some cases and subtypes this is broken down during disease progression. ctDNA levels correlated with disease stage, as expected if ctDNA reflects tumour burden. In patients with localised Stage I-III colorectal, gastroesophageal, pancreatic and breast cancer, ctDNA was detected in 49% - 78% of patients, compared to 86%-100% patients with Stage IV disease (Figure 1.6A). In analysis of all cancer types, ctDNA was detected in 47% of Stage I patients, compared to 55%, 69% and 82% of patients with Stage II-IV cancer respectively (Figure 1.6B,C).

1.6: Development of next generation sequencing for ctDNA analysis

Over the years, different sequencing methods have been developed to assess the sequence of DNA bases and identify tumour-specific mutations. Original methods include dideoxy, or ‘Sanger’ sequencing, which incorporates fluorescently-tagged chain-terminating dideoxynucleotides during *in vitro* DNA replication to determine the sequence of bases (50,51). More recently, next generation sequencing (NGS) methods have been developed which enable high-throughput ‘massively parallel sequencing’ (52,53). The most commonly used method is a ‘sequencing-by-synthesis’ method by Illumina, which uses reversible dye-terminators and clonal amplification to determine the sequence of bases in a library of fragmented adaptor-ligated DNA molecules which are clustered on a flow cell (53). Different DNA library preparation methods can be used to assess sequences on different scales. Whole genome sequencing (WGS, ~30x coverage) can be used to sequence the entire genome **(10)**, although would be prohibitively expensive as a clinical assay. Shallow whole genome sequencing (sWGS) is a cost-effective alternative, using low depth sequencing (~0.1 – 0.4x coverage) to assess SCNAs on a genome-wide scale (54,55). Whole exome sequencing (WES) can be performed to capture and sequence exonic regions which make up ~1% of the genome **(4)**, and targeted sequencing can be used for high-depth sequencing of specific regions of interest **(3)**.

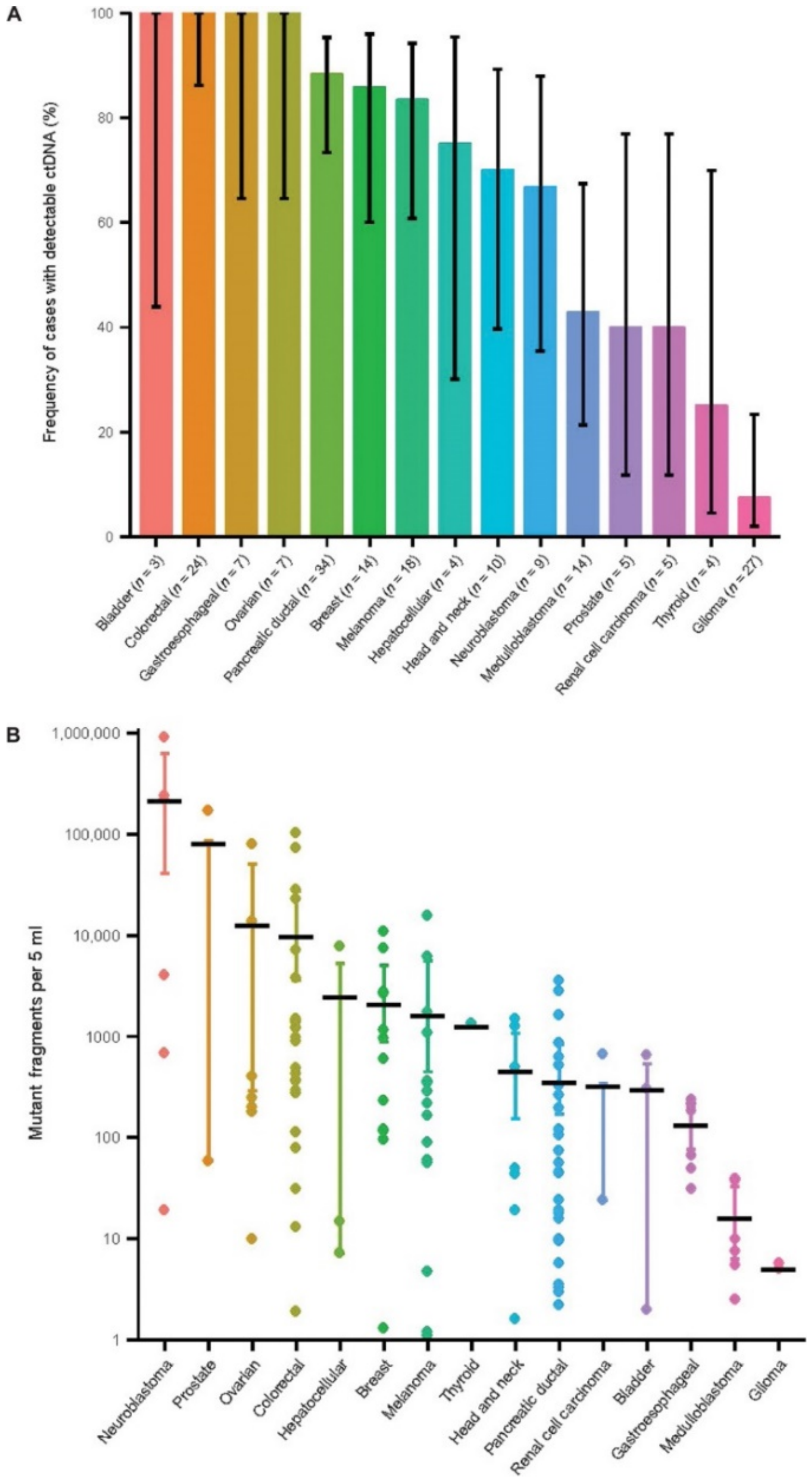


Figure 1.5: ctDNA levels in patients with different advanced cancers (*A). Frequency of cases where ctDNA was detected (%). (B). Quantification of mutant fragments per 5 mL plasma. Error bars: 95% bootstrapped confidence interval of the mean. Figure and Figure Legend adapted from Bettegowda et al.* (48)

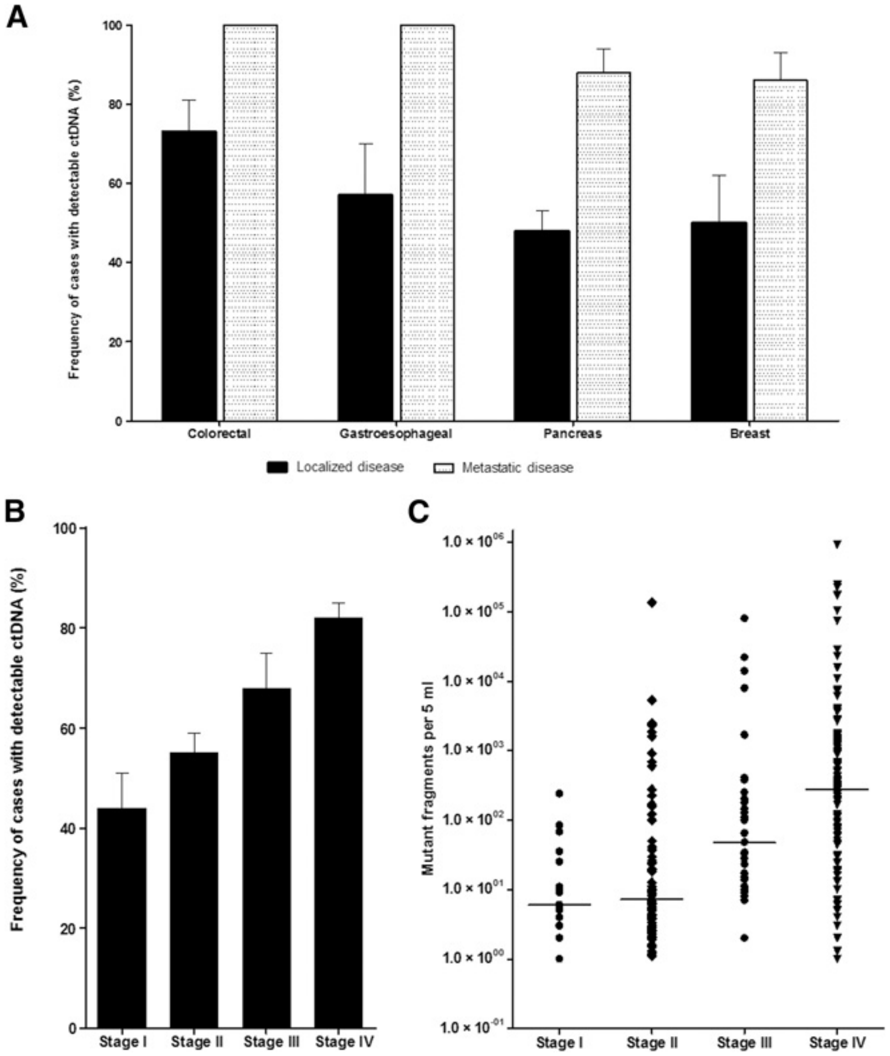


Figure 1.6: Levels of ctDNA detected at different stages of disease. *Frequency of cases with ctDNA detected in (A) localised (Stage I-III) and metastatic (Stage IV) colorectal, gastroesophageal, pancreatic and breast cancer. (B) in Stage I-IV (C) Quantification of mutant fragments/5 mL plasma in Stage I-IV patients with 16 cancer types. Error bars: standard error of mean. Figure and Figure Legend adapted from Bettegowda et al.* (48)

The identification of the presence of foetal DNA in maternal plasma led to seminal work by Professor Dennis Lo to analyse cell-free DNA for non-invasive prenatal testing (NIPT) for the detection of trisomy 21 (Down's syndrome) and other aneuploidies (56). The use of a non-invasive blood test has distinct advantages, given the fact that amniocentesis and chorionic villus sampling, diagnostic methods previously routinely used in the clinic, are invasive and may increase the risk of miscarriage by ~1%. Advances have led to the recent widespread implementation of NGS in the clinic to analyse cfDNA for NIPT to assess the foetal fraction present in maternal plasma and determine the presence of chromosomal abnormalities (57).

In NIPT, the foetal fraction can be relatively high, with limits for accurate detection of aneuploidies at approximately 4% (58). In oncology, the situation is more challenging, given the low fraction of ctDNA, particularly in early stage disease. Digital PCR is a sensitive method, but is limited to analysis of single or a few loci. If the tumour evolves, resulting in loss of the mutation in a minor subclone, this may in theory provide false information to a clinician. NGS has the potential to provide more comprehensive profiling of the tumour genomic landscape in ctDNA by monitoring multiple mutations in parallel. This is challenging, however, given the limited amounts of fragmented DNA in plasma, and high levels of background noise. To circumvent this, we developed TAm-Seq (tagged amplicon deep sequencing), a novel amplicon-based targeted sequencing method for the identification of low frequency mutations in plasma by sequencing large genomic regions encompassing key cancer genes. This work, discussed in Chapter 3, demonstrated for the first time that NGS can be used to detect and monitor multiple low frequency mutations in parallel in ctDNA **(3)**.

1.7: Thesis - Aims and objectives

This thesis will focus on research I performed between 2009 and 2019 to investigate the diagnostic potential of ctDNA as a non-invasive liquid biopsy, and assess the hypothesis that ctDNA is a clinically useful biomarker able to monitor tumour dynamics, correlate with disease burden, and be used to guide treatment. It will include research I have performed to develop dPCR and NGS methods for the highly sensitive detection of ctDNA (Chapter 2, 3), the development and analytical validation of clinical diagnostic tests to ISO15189:2012 regulatory standards for patient benefit (Chapter 4), and research into the ability to use ctDNA for molecular stratification, monitoring and the earlier detection of cancer (Chapters

2, 5). In the next Chapter, I will discuss work I have published using dPCR to assess levels of tumour-specific DNA in the plasma of patients with high-grade serous ovarian cancer.

Chapter 2: Development of digital PCR for ctDNA analysis in ovarian cancer

2.1: High-grade serous ovarian cancer

High-grade serous ovarian carcinoma (HGSOC) is the most common type of ovarian cancer, an aggressive disease with poor prognosis, characterised by ubiquitous *TP53* mutations and a large number of copy number aberrations (59–62). Approximately 13% of HGSOC cases are caused by germline mutations in *BRCA1* and *BRCA2* genes, resulting in deficiencies in the homologous recombination DNA repair pathway (60). Patients are often treated with platinum-based chemotherapy, and may initially have a good response, but invariably relapse. Women relapsing within 6 months are considered to have ‘platinum-resistant’ disease, and those who relapse >12 months after initial treatment are considered ‘platinum-sensitive’. Poly ADP ribose polymerase (PARP) inhibitors, such as olaparib, have recently been developed which are effective in mutant *BRCA1/2* patients by preventing single-stranded DNA break repair, resulting in the accumulation of double-stranded DNA (dsDNA) breaks and tumour cell death.

Unfortunately, most HGSOC patients present with late-stage disease. Cancer antigen 125 (CA-125) is a serum glycoprotein routinely used in clinical practice to monitor treatment response in HGSOC and assess pelvic masses, but unfortunately has relatively low specificity, as it is expressed in normal tissues and other benign conditions. There is a need for biomarkers with higher sensitivity and specificity. In the next section I will discuss work I published using dPCR to assess: [1] mutant *TP53* ctDNA as a treatment response biomarker in HGSOC **(1)**; and [2] an *NF1* deletion in an HGSOC patient to determine if subclonal populations were present prior to treatment **(2).**

2.2: Parkinson*, Gale* et al., PLoS Medicine, 2016 (1)

2.2.1: Aims

I co-led this study to test the hypothesis that mutant *TP53* ctDNA is a clinically useful biomarker in HGSOC by comparison with CA-125 and disease volume, to establish whether dPCR can be used to monitor ctDNA dynamics, and assess whether ctDNA can be used as an early response marker and predict time to progression (TTP). An overview of the workflow is given in Figure 2.1. As >97% patients have a *TP53* mutation, HGSOC is an ideal cancer in which to study ctDNA dynamics (59,60).

2.2.2: Results

Retrospective analysis was performed on 318 serial plasma samples from 40 HGSOC patients undergoing heterogeneous chemotherapy treatment. *TP53* mutations were identified in formalin-fixed paraffin-embedded (FFPE) tumour DNA. To assess both the *TP53* mutant allele fraction (TP53MAF) in ctDNA and total cfDNA levels, I designed and validated 31 unique patient-specific dPCR assays, using dual-labelled fluorogenic probes to wild-type and mutant sequences. The number of amplifiable copies/mL (AC/mL) of plasma was calculated, with a limit of detection (LoD) at ≥20 AC/mL. CA-125 levels were assessed using a two-site sandwich immunoassay, with ≤30 IU/mL defined as the institutional upper limit of normal (ULN). 3D volumetric analysis was performed using CT (computed tomography) and positron emission tomography (PET) scans. Tumour responses were evaluated using RECIST v1.1 (Response Evaluation Criteria in Solid Tumours) criteria (63), to assess whether there was progressive disease (PD), complete response (CR), partial response (PR) or stable disease (SD) since previous scans.

The two most common mutations were *TP53* p.R175H and p.R273H, each identified in four patients (20% cases). ctDNA was detected in 82% (42/51) of treatment courses (defined as the period of a specific treatment) in relapsed patients, and in 86% (6/7) untreated newly diagnosed Stage IIIC/IV cases. CA-125 was detected in 100% of relapsed and untreated cases. Analysis showed that the median pre-treatment TP53MAF was 8% in 51 relapsed treatment courses, 0.7% in newly diagnosed cases, and 0.2% in 4 patients after primary surgery.

2.2.3: Monitoring tumour dynamics using *TP53* ctDNA

To address whether dPCR can be used to monitor ctDNA kinetics, TP53MAF was plotted over time during the course of treatment, together with CA-125 and imaging response data. In the majority of cases ctDNA tracked the course of disease, but with a more rapid decrease in TP53MAF compared to CA-125 (Figure 2.2). The median time to nadir (the

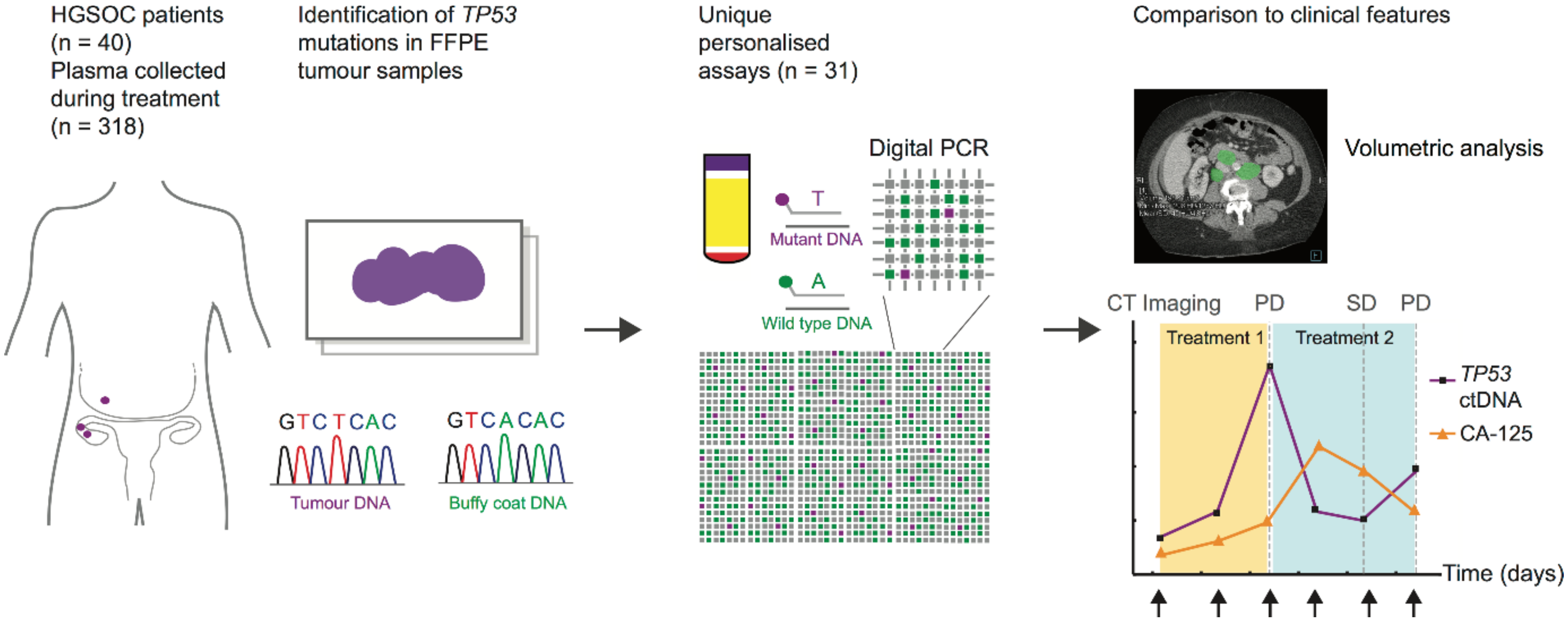

HGSOC patients
(n = 40)
Plasma collected
during treatment
(n = 318)
Identification of TP53
mutations in FFPE
tumour samples
GTCTCAC
Tumour DNA
GTCACAC
Buffy coat DNA
Unique
personalised
assays (n = 31)
Digital PCR
T
Mutant DNA
A
Wild type DNA
Comparison to clinical features
Volumetric analysis
CT Imaging
PD
SD
PD
Treatment 1
Treatment 2
TP53
ctDNA
CA-125
Time (days)

Figure 2.1: Schematic workflow for mutant *TP53* ctDNA analysis in high-grade serous ovarian cancer (HGSOC) patients

FFPE: formalin-fixed paraffin-embedded; CT: Computed tomography: PD: Progressive disease; SD: Stable disease. Figure and Figure Legend adapted from ***Parkinson*, Gale* et al.*** **(1)**

lowest point during a treatment course) was 37 days for TP53MAF (median decrease: 98%), compared to 84 days for CA-125 (median decrease: 55%). Two cases showed discrepant TP53MAF and CA-125 kinetics, with ctDNA more accurately reflecting clinical data (Figure 2.2D/E). Ascites fluid can accumulate in HGSOC patients, so drains may be inserted to remove excess fluid and improve patient comfort. Ascitic drainage appeared to reduce TP53MAF levels (Figure 2.2F).

2.2.4: Correlation of *TP53* ctDNA with tumour volume

To assess if ctDNA levels correlated with tumour volume, volumetric analysis was performed on imaging data from 51 relapsed treatment courses from 32 patients, and from 7 newly diagnosed cases. ctDNA was detected (at >20 AC/mL) in all relapsed courses with tumour volume >32cm^3, with the exception of one patient with 50cm^3 disease detected at 15 AC/mL (Figure 2.3A). All patients with <20 cm^3 tumour had <20 AC/mL ctDNA, with the exception of one case with 1cm^3 tumour volume, but with large volume ascites. Analysis of 35 treatment courses from relapsed cases showed that tumour volume positively correlated with TP53MAF (Pearson $r = 0.59$, $p < 0.001$), and this increased when those with ascites were excluded (Pearson $r = 0.82$, $p < 0.001$, Figure 2.3B). CA-125 only moderately correlated with tumour volume, with and without ascites (Pearson $r = 0.52$, $p = 0.001$, and $r = 0.51$, $p = 0.016$ respectively, Figure 2.3C).

2.2.5: Assessment of *TP53* ctDNA as an early response marker

We next wanted to assess whether a decrease in ctDNA within one cycle of chemotherapy could predict time to progression. Analysis of relapsed patients in 32 courses of chemotherapy showed a median 74% decrease in TP53MAF from cycle 1 (C1) to pre-cycle 2 (C2), compared to 18% for CA-125. The median TTP was 189 days. Pre-treatment TP53MAF, total cfDNA amplifiable copies, age, platinum sensitivity, number of lines of chemotherapy and disease volume were all shown to be significant predictors of TTP in univariate analysis ($p < 0.05$; Cox regression model). In multivariate analysis, adjusted using Cox regression analysis, only pre-treatment TP53MAF [Hazard Ratio (HR) 1.03, 95% CI: 1.01 – 1.06, $p = 0.019$] and platinum sensitivity (HR 0.43, 95% CI: 0.19 – 0.99, $p = 0.048$) remained significant.

A Receiver Operating Characteristic (ROC) curve was plotted, providing a graphical display of the proportions of true and false positives at all possible pre-determined values to

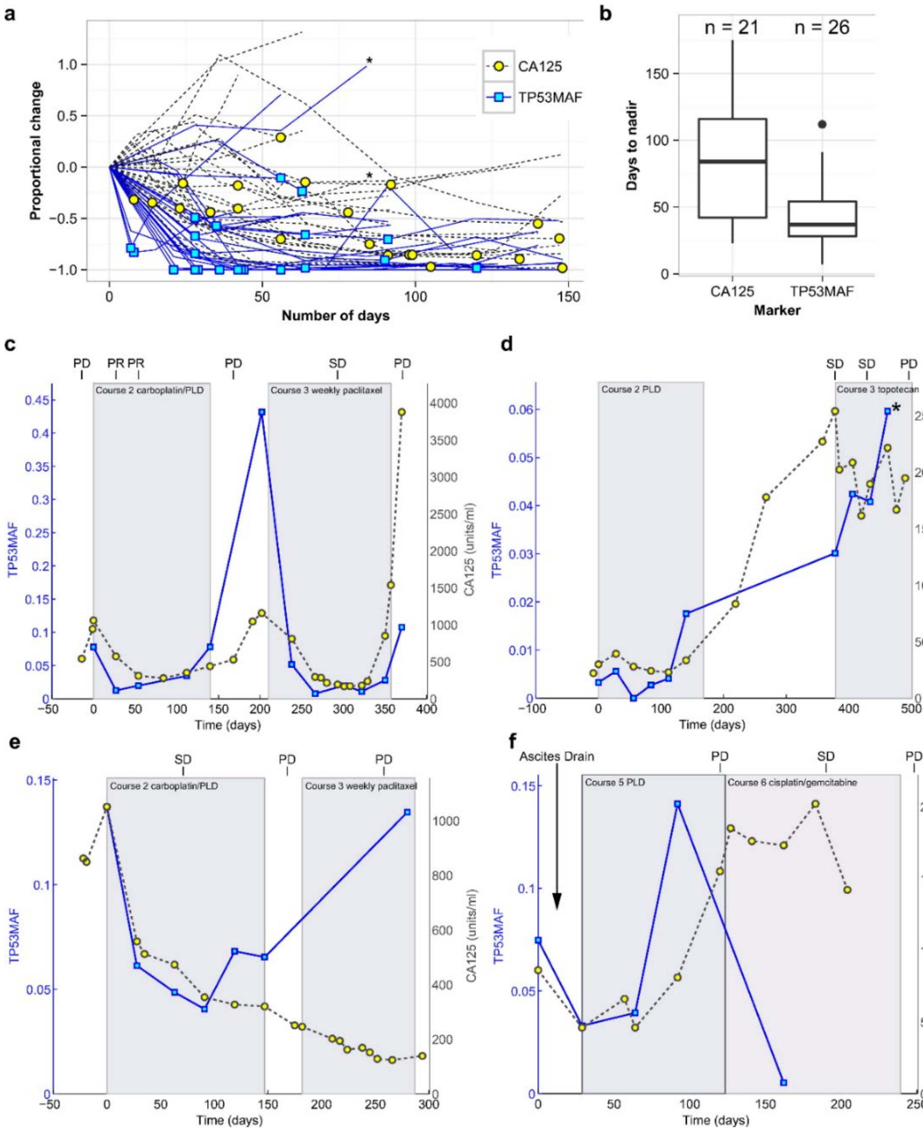


Figure 2.2: Monitoring ctDNA and CA-125 kinetics during chemotherapy.

*(a) CA-125 (yellow), TP53MAF (blue) ctDNA kinetics, normalised to pre-treatment levels. (b) Time to nadir (lowest point during a treatment course) following start of chemotherapy for CA-125 and TP53MAF (p<0.01). (c-f) Illustrative cases of TP53MAF and CA-125 kinetics. (c) Faster time to nadir of TP53MAF compared to CA-125. (d-e) Discrepant TP53MAF and CA-125 kinetics on third-line chemotherapy: (d) CA-125 decreased, TP53MAF increased, PD on imaging. (e) CA-125 decreased, TP53MAF increased, PD on imaging. (f) Effect of ascitic drainage, before start of treatment: TP53MAF fell 7.5% to 3.3% AF; CA-125 decreased 86 IU/mL to 46 IU/mL. * New brain metastasis; PLD: Pegylated liposomal doxorubicin; PD: Progressive disease; SD: Stable disease; PR: Partial response; Figure and Figure Legend adapted from* ***Parkinson*, Gale* et al.*** **(1)*****.***

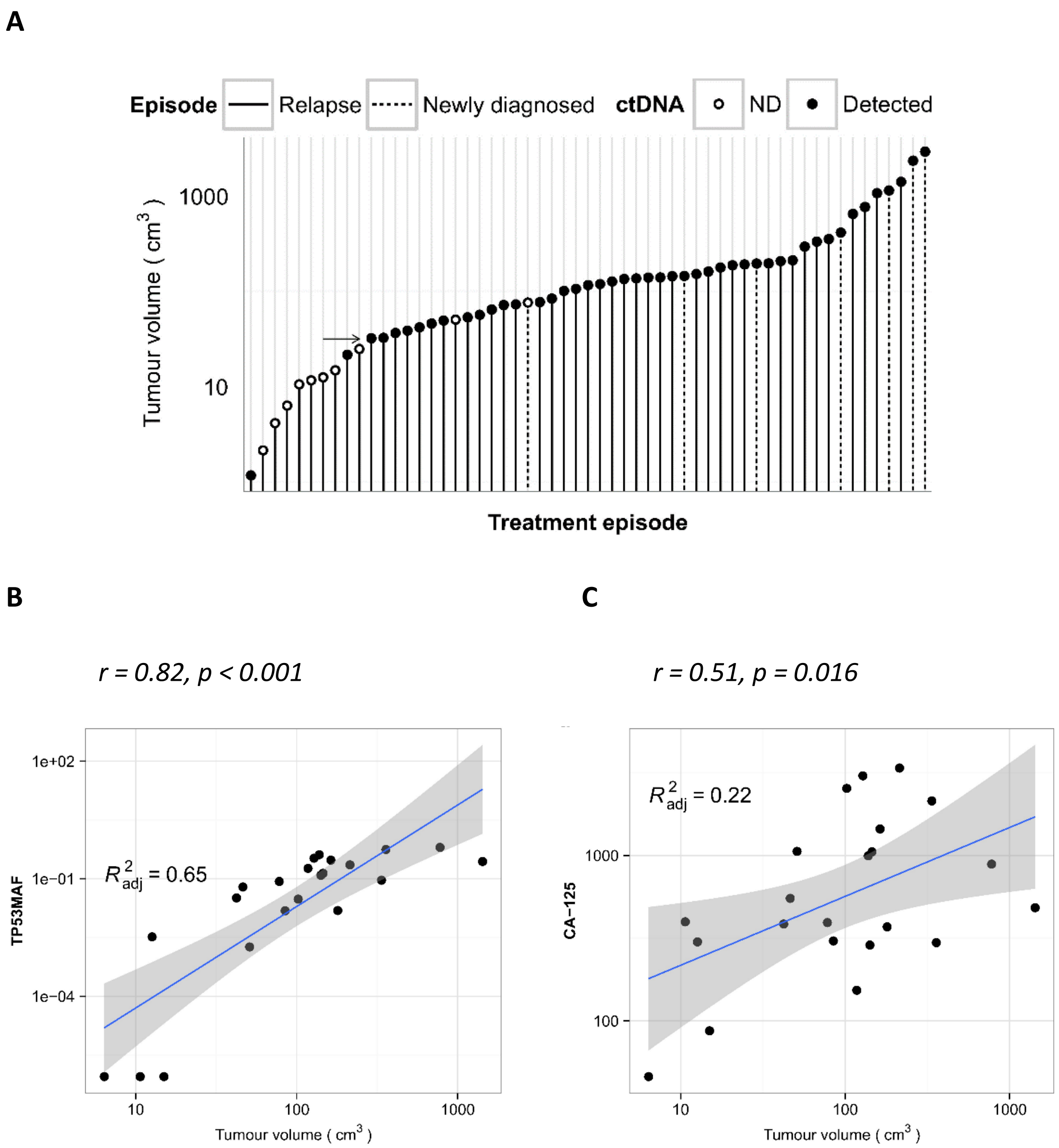


Figure 2.3: Comparison of TP53 mutant allele fraction (TP53MAF) and tumour volume. *(A) Ranked total tumour volume at start of treatment course (episode). Filled circles: cases with ≥20 AC/mL TP53MAF; Arrow: 32cm^3 tumour volume; ND: not detected (B, C) Linear regression analysis of TP53MAF and CA-125 with tumour volume in 22 relapsed patients without ascites. R^{2}adj: For log correlation calculations, zero values for TP53MAF and total cfDNA amplifiable copies were adjusted by addition of 0.001 times the lowest value in the series. Blue line: Best line of fit; Grey shading: 95% confidence intervals; Figure and Figure Legend adapted from* ***Parkinson*, Gale* et al.*** **(1)**

determine the optimal cut-point for predicting which patients will relapse within 6 months. Six months was selected as a clinically significant endpoint (Figure 2.4). Multivariate analysis was performed on 31 relapsed courses evaluable for response using the following covariates: TP53MAF decrease >60% from C1 to C2; CA-125 decrease >50%; age; performance status; platinum sensitivity; number of lines of chemotherapy; volume of disease and presence of ascites. A decrease in TP53MAF of >60% from C1 to C2 was the only independent predictor of TTP (HR 0.22, 95% CI: 0.07 – 0.67, *p* = 0.008). A decrease of ≤60% was associated with poor prognosis and progression within 6 months with 71% sensitivity (95% CI 42% - 92%) and 88% specificity (95% CI: 64% - 99%). CA-125 did not show the same effect. (HR 0.86, 95% CI: 0.28 – 2.71, *p* = 0.802; 93% sensitivity, 29% specificity). In analysis of TP53MAF, specificity was improved further when patients with a recent ascites drain were excluded [75% sensitivity (95% CI: 43% - 95%); 100% specificity (95% CI 74% - 100%)].

2.2.6: Discussion

In HGSOC, there is a need for more specific biomarkers that correlate with tumour burden and can monitor treatment response. Whilst CA-125 may indicate a trend in response, it does not provide an accurate reflection of absolute tumour volume. ctDNA represents an attractive alternative, as it is tumour-derived, highly specific, and has a much shorter half-life. Swisher et al. analysed blood and peritoneal fluid from ovarian cancer patients with known *TP53* tumour mutations, and detected mutations in plasma from 30% (21/69) patients, and in peritoneal fluid from 28/30 (93%) cases using a ligase-detection reaction (64). Otsuka identified *TP53* mutations in just 16.7% (2/12) cases in pre-operative plasma (65). In **Parkinson et al**., I used dPCR to detect ctDNA in 82% of treatment courses and in 86% newly diagnosed Stage IIIC/IV cases. This publication provides an important contribution to the literature, being the first to correlate ctDNA with tumour volume in HGSOC. ctDNA tracked the course of disease, with a more rapid decrease and dynamic range compared to CA-125. Response to chemotherapy was seen earlier with ctDNA, with a faster median time to nadir (37 v 84 days). A ≤60% decrease in TP53MAF between cycle 1 and cycle 2 was associated with shorter TTP, indicating that mutant *TP53* has potential as an early response marker in HGSOC.

Tie et al. provided comparable data in analysis of 53 metastatic colorectal cancer patients, demonstrating early changes in ctDNA during chemotherapy predicted radiologic response, and was superior to CEA (66). A decrease of ≥10 fold in ctDNA pre-cycle 2 was associated

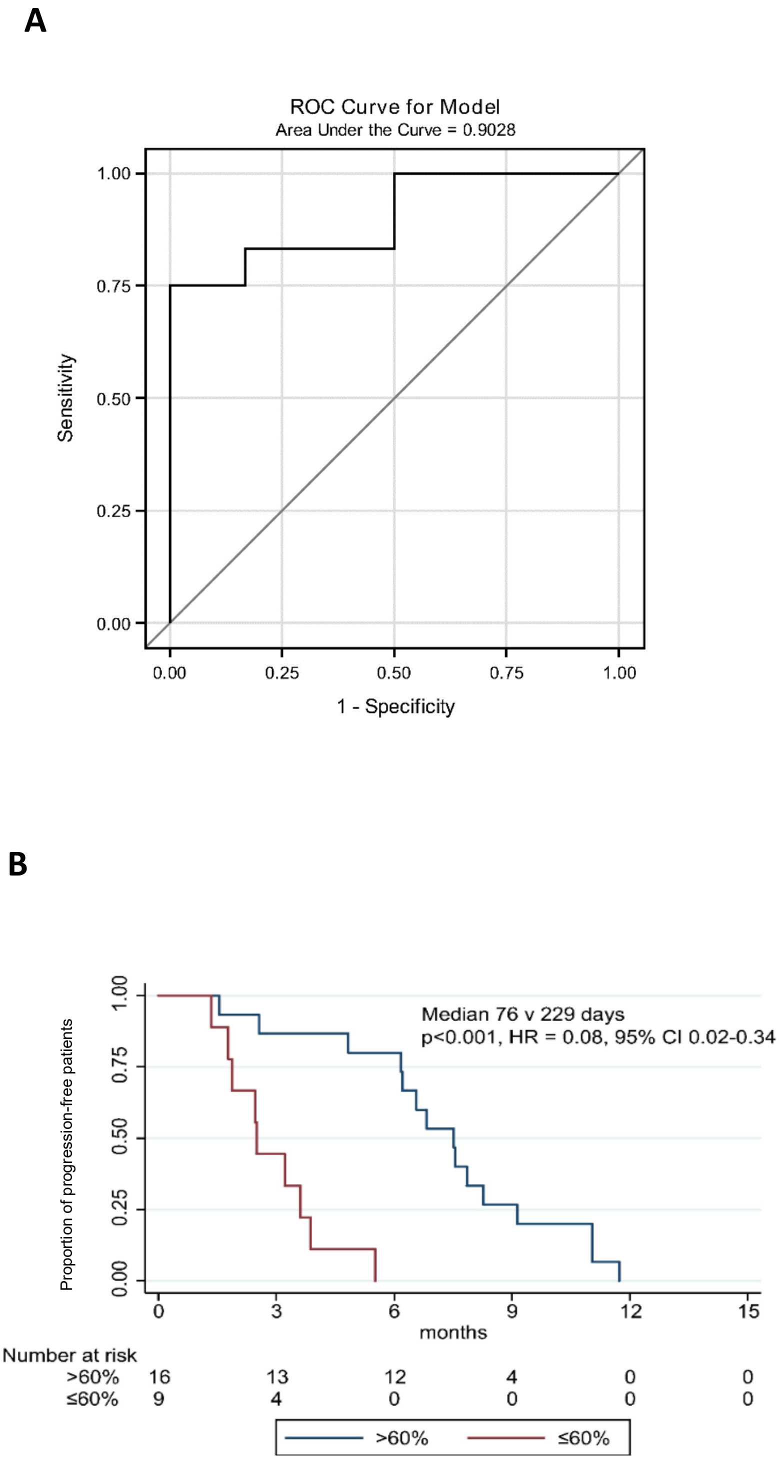


Figure 2.4: (A) Receiver Operating Characteristic (ROC) curve and (B) Kaplan-Meier plot for TP53MAF decrease after one cycle of chemotherapy, excluding courses with recent ascitic drains. *HR: hazard ratio; CI: confidence interval. Figure and Figure legend adapted from* ***Parkinson*, Gale* et al.*** **(1)**

with increased progression-free survival (PFS; median 14.7 v 8.1 months; HR = 1.87; *p* = 0.266). Newman et al. were able to demonstrate that pre-treatment ctDNA levels from 9 NSCLC patients significantly correlated with tumour volume, as assessed by CAPP-Seq (a sensitive NGS method discussed in Chapter 3), CT and PET imaging [Linear regression; R^2 = 0.89, *p* = 0.0002, Figure 2.5 (67)]. ctDNA was detected in patients with 5 – 20 cm^3 of disease, which is comparable with our data where ctDNA was detected in relapsed HGSOC patients with >32 cm^3 disease, and in one patient with 1cm^3 disease and high volume ascites. McEvoy et al. demonstrated that ctDNA correlated with tumour volume in patients with metastatic melanoma (68).

Seminal work by Professor Charlie Swanton has focussed on the study of tumour heterogeneity, defining 'truncal' mutations carried in all tumour cells, and 'branch' mutations, not present in all, to explain the phylogenetic evolutionary process. Abbosh, Swanton et al. demonstrated that it is possible to track tumour branch mutations in ctDNA in patients with early stage lung cancer on the TRACERx trial, using patient-specific multiplex panels targeting clonal and subclonal SNVs (69). Tumour volume measured by CT correlated with mean clonal variant allele frequency (VAF) [Figure 2.6]. Linear modelling predicted a primary NSCLC tumour volume of 10 cm^3 would give a 0.1% VAF. In **Parkinson et al.**, we were able to demonstrate that the median TP53MAF/volume in patients without ascites was 0.08% per cm^3 (IQR 0.02%–0.13% per cm^3). Taken together, these publications provide good agreement with our results, and demonstrate that ctDNA correlates with tumour burden in these cancer types.

The main limitation of the **Parkinson et al.** study is that it was a retrospective study, performed in a non-regulated laboratory, analysing a small cohort of heterogeneously-treated patients who entered the clinic at different stages of disease. To validate these findings, further analysis is required in large prospective clinical trials receiving standardised treatment, using clinically validated assays performed to appropriate diagnostic regulatory standards.

One of the key challenges was that analysis of mutant *TP53* by dPCR required initial tumour sequencing, and bespoke design and validation of 31 unique patient-specific assays to assess ctDNA levels in plasma. This is both time consuming and expensive, compared to CA-125 testing. For clinical implementation, pre-validated assays to commonly mutated regions such as *TP53* p.R175H and p.R273H are available, but custom assays would need to be designed and validated to assay rarer mutations. For this reason, in the clinical setting,

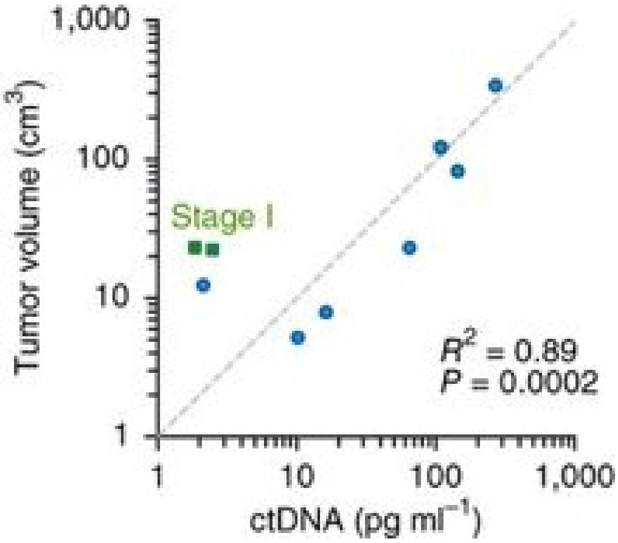


Figure 2.5: Concordance between tumor volume and pre-treatment pg/mL ctDNA in 9 NSCLC patients. *Linear regression analysis performed in non-log space; log-log axes and dashed diagonal line for display purposes only. Figures and Figure Legend adapted from Newman et al.* (67)

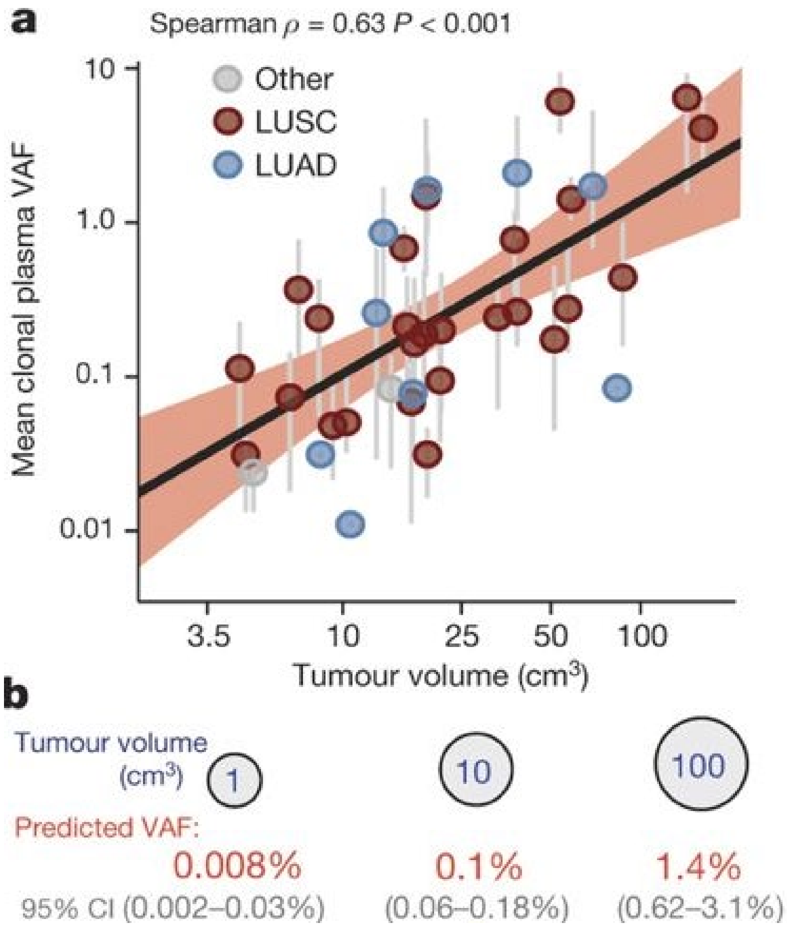


Figure 2.6: (a) Tumour volume (cm^3) correlation with mean clonal plasma VAF *(n=37). Grey vertical lines: range of clonal VAF, red shading: 95% CI; LUAD = lung adenocarcinoma; LUSC = lung squamous cell carcinoma.* **(b) Model-based prediction of mean clonal VAF at hypothetical volumes ranging 1-100 cm^3.** *Figure and Figure Legend adapted from Abbosh et al.* (69)

dPCR is best suited to analysis of single driver mutations. dPCR analysis of multiple mutations requires the sample to be sub-divided into different assays, introducing sampling bias and limiting sensitivity of detection of low frequency mutant alleles. For clinical implementation, the ideal assay would be able to analyse multiple genomic regions in parallel without using additional sample material. For this reason, we developed TAm-Seq, a novel amplicon-based sequencing method, which can screen large genomic regions in parallel in the same assay, as discussed in Chapter 3.1 **(3)**.

2.3: Schwarz et al., PLoS Medicine, 2015 (2)

2.3.1: Aims

The aim of this study was to assess whether intra-tumour heterogeneity (ITH) in HGSOC is associated with development of disease resistance through clonal evolution and emergence of sub-clonal populations, and determine whether quantitative assessment of ITH could predict outcomes. Schwarz et al. developed the MEDICC (Minimal Event Distance for Intra-tumour Copy Number Comparisons) algorithm to quantitatively assess heterogeneity by analysis of copy number profiles generated by array CGH (comparative genomic hybridisation) and WGS (70). MEDICC was used to determine estimates of evolutionary distances between tumour samples and assess the degree of clonal expansion [CE] (2). I designed and validated dPCR assays to assess levels of an *NF1* deletion and *TP53* mutation to determine if subclonal populations were present prior to treatment.

2.3.2: Results

Copy number data was generated on temporally and spatially separated tumour samples (n=138) from 17 HGSOC patients undergoing platinum-based chemotherapy. MEDICC was used to reconstruct tumour evolution, demonstrating profiles grouped primarily by patient, rather than sensitive or resistant subtypes. Evolutionary distances indicated that in 8/9 patients with ≥3 samples, tumours originated from different metastatic lesions, rather than from the same primary tumour, indicating that heterogeneity in these patients generally resulted from ongoing clonal evolution.

MEDICC was used to assess the degree of clonal expansion, using median CE to divide patients into CE-high and CE-low groups. CE-high patients had shorter PFS and overall survival (OS) compared to CE-low patients (PFS: median 12.7 v 10.1 months, $p = 0.009$; OS:

42.6 v 23.5 months, *p* = 0.003). In a CE-high patient, a focal *NF1* deletion was identified in a relapsed sample, but not detected in pre-chemotherapy and interval debulking surgery (IDS) samples by array CGH. *NF1* is an inhibitor of RAS signalling, and a potential driver gene, commonly mutated or deleted in HGSOC.

For enhanced sensitivity, I designed patient-specific dPCR assays to the *NF1* deletion and *TP53* p.R175H mutations, to assess whether resistant sub-clonal populations were already present prior to treatment. The *NF1* assay used forward and reverse primers spanning either side of the breakpoint, and EvaGreen® DNA binding dye, which becomes fluorescent upon binding dsDNA during amplification. dPCR demonstrated that the *NF1* deletion was present prior to treatment, detected at 5% and 26% AF in the pre-treatment samples, in 25% - 100% of IDS samples, at 1.2% AF in a primary invasive carcinoma in the fallopian tube, and at 7.9% AF in an adjacent left ovarian metastasis.

2.3.3: Discussion

In **Schwarz et al**., the MEDICC algorithm was shown to be able to quantify heterogeneity in HGSOC. CE-high patients were shown to have shorter PFS and OS compared to CE-low patients. Unlike array profiling, dPCR analysis showed that an *NF1* deletion was present in a subclonal population pre-treatment, and in IDS samples and relapsed ascites fluid, with the ratio of *NF1*/*TP53* increasing over time.

Diaz et al. provided further evidence that resistant subclones may be present prior to treatment by demonstrating the emergence of *KRAS* mutations in serum from mCRC patients approximately 5-6 months after treatment with panitumumab (an anti-EGFR antibody therapy). Mathematical modeling indicated that these mutations were already present in subclones prior to the initiation of treatment (71), consistent with our data.

The main limitation of our study is that it was performed on a small patient cohort, which may over-estimate the effect of clonal expansion on survival. Watkins and Schwarz have done further work to show that multi-region sampling can help detect clonal expansion and provide a quantitative measure of ITH (72). Bashashati et al. have also demonstrated using targeted sequencing, copy number analysis, exome sequencing, and gene expression profiling that HGSOC patients have a high degree of tumour heterogeneity, which occurs in early stages of tumourigenesis (73).

In Chapter 3, I will discuss work I have published on the use of next generation sequencing to identify low frequency mutations in cfDNA **(3–5)**.

Chapter 3: Development of NGS for ctDNA analysis - TAm-Seq & exome sequencing

As discussed in Chapter 2, digital PCR is limited to analysis of single or few loci of interest. To address this, we investigated the hypothesis that next generation sequencing could be used to detect and monitor multiple mutations in parallel in ctDNA. TAm-Seq was developed to analyse rare mutations in plasma **(3)**, and exome sequencing was used to study potential mechanisms of treatment resistance **(4,5)**.

3.1: Forshew*..Gale* et al., Science Translational Medicine, 2012 (3)

3.1.1: Aims

To test whether NGS could be used to detect low frequency mutations in plasma, TAm-Seq, a novel amplicon-based sequencing method, was developed using primers designed to amplify entire coding regions of *TP53* and *PTEN* genes, and selected regions of *EGFR*, *BRAF*, *KRAS* and *PIK3C*A. I was involved in establishing TAm-Seq for analysis of tumour DNA, and generated key dPCR validation data, demonstrating the quantitative accuracy of TAm-Seq in analysis of plasma DNA.

3.1.2: Results

The TAm-Seq workflow is outlined in Figure 3.1. Primer pairs, incorporating universal 5' adaptors, were designed to tile and amplify regions of interest, spanning 5995 bases in total. A limited-cycle multiplex pre-amplification step was initially performed to amplify template molecules, and a second amplification performed on pre-amplified material using the Fluidigm Access Array 48.48 microfluidic chip. This enabled 48 single-plex amplifications to be performed in parallel, each using 48 different primer sets. Finally, sequencing adaptors and sample-specific barcodes were incorporated using primers containing sequences complementary to the universal adaptors, with forward and reverse sequencing adaptors added separately then pooled, to enable sequencing in both directions. Reactions were performed in duplicate, with each replicate tagged with a different sample-specific barcode.

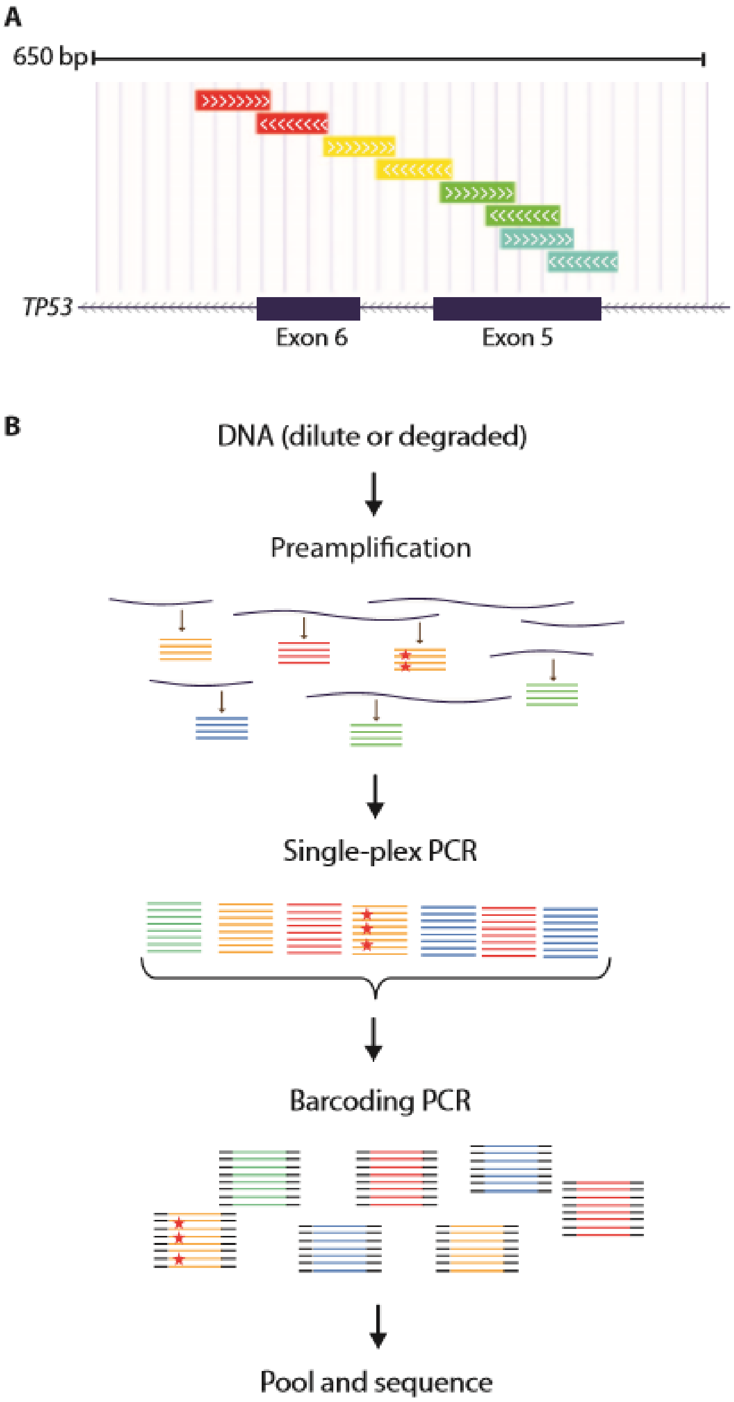

A
650 bp
TP53
Exon 6
Exon 5
B
DNA (dilute or degraded)
Preamplification
Single-plex PCR
Barcoding PCR
Pool and sequence

Figure 3.1: Overview of TAm-Seq. ***(A)*** *Amplicon design using tiled short overlapping amplicons.* ***(B)*** *Multiple regions amplified in parallel, using initial pre-amplification, followed by single-plex PCR. DNA molecules carrying mutations (red stars) are amplified alongside wild-type molecules. Sequencing adapters and sample-specific barcodes are attached by PCR. Figure and Figure Legend adapted from* ***Forshew et al*** **(3)***.*

TAm-Seq was tested by sequencing FFPE tumour DNA samples from HGSOC patients, and compared with Sanger sequencing of the *TP53* gene to identify mutations. TAm-Seq data was obtained for 44 samples, with SNVs detected in 43/44, and 100% concordance between *TP53* mutations identified by both techniques. Three additional mutations were identified by TAm-Seq at low frequency, below the detection threshold of Sanger sequencing. In a dilution experiment, using diluted mixtures of 6 FFPE samples with known *TP53* mutations, all expected mutations were detected by TAm-Seq at >1% AF, with the exception of one false positive at 1.9% AF (Figure 3.2). There was a significant correlation between the mutant AF and cellularity estimates from histological analysis of the FFPE biopsy (t-test; $r = 0.422$; $p = 0.0049$).

The background level of non-reference (mutant) AFs was ~0.1%, although this varied based on the particular loci and mutation class of each base substitution. Mutations were called if non-reference sequences were above the substitution-specific background distribution at high confidence (≥0.9995), and ranked high in the list of non-reference AFs in both replicates.

TAm-Seq was used to analyse plasma DNA from 7 HGSOC patients. In all cases, *TP53* mutations identified were concordant with Sanger sequencing. In one case, a *de novo EGFR* mutation, not previously identified, was also observed at 6% AF in plasma taken at relapse, and verified using both dPCR and replicate Sanger sequencing of highly diluted template. Analysis of additional specimens from this patient showed this mutation was present in a second plasma sample taken 25 months post-diagnosis at 5% AF, at 0.7% AF in two omental tissues taken at IDS, but not in left and right ovary samples (Figure 3.3). This indicated that the mutation originated in omental tissue. Further TAm-Seq analysis of 62 plasma samples from 37 HGSOC patients demonstrated 92.86% concordance with matched dPCR data.

To assess the hypothesis that TAm-Seq can be used to non-invasively monitor tumour dynamics in blood, serial plasma samples collected during the course of treatment were analysed from patients with HGSOC and breast cancer. Results demonstrated the concordance of TAm-Seq with dPCR data, and the ability to track 10 concomitant mutations in plasma to monitor disease burden (Figure 3.4).

Figure 3.5 shows results of analysis of a patient with synchronous primary ovarian and bowel cancer. Retrospective TAm-Seq analysis was performed on primary tumour and plasma DNA collected following relapse, showing that a *TP53* p.R273H mutation initially

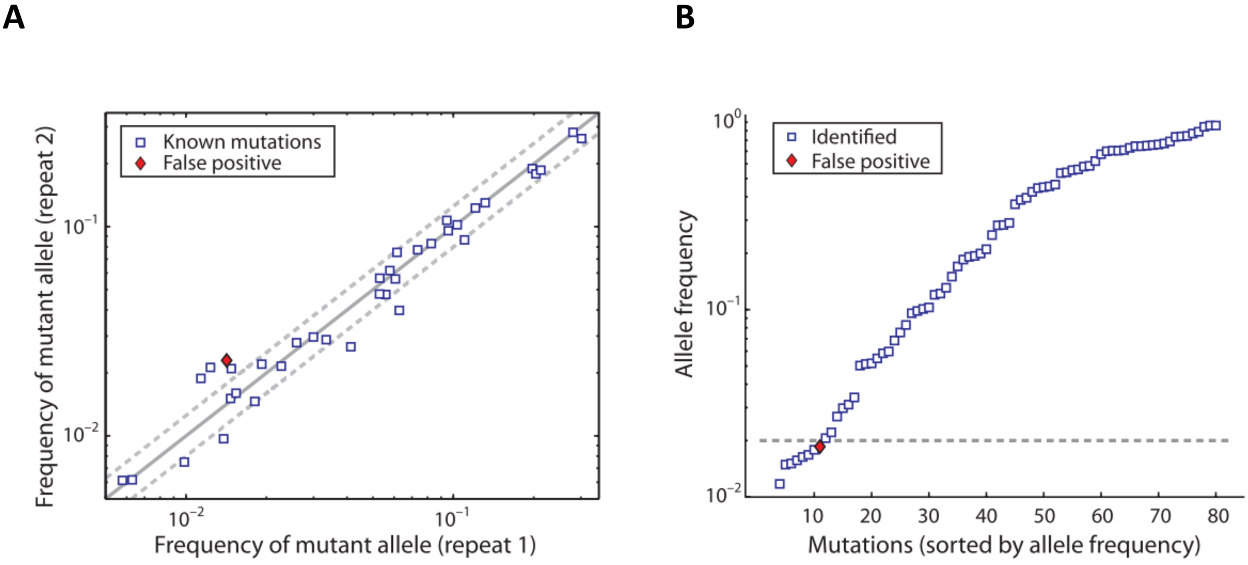


Figure 3.2: Identification of mutations in ovarian cancer FFPE samples by TAm-Seq. ***(A)*** *Concordance between duplicate measurements of allele frequencies (AFs) in mutations identified in a FFPE DNA mixture. Solid line: Equality. Dotted lines: 0.05 difference in AF* ***(D)*** *Summary of FFPE mutations identified by TAm-Seq. Dotted line: 2% AF. Figure and Figure Legend adapted from* ***Forshew et al.*** (3)

detected in the ovarian primary was present in relapsed plasma samples, whilst *TP53* p.R248W, *KRAS* p.G12V and *PIK3CA* p.E545K mutations, originally present in the bowel primary, could not be detected post-relapse.

3.1.3: Discussion

In **Forshew et al**., we were able to demonstrate for the first time that NGS could be used to non-invasively identify *de novo* mutations from solid tumours directly in plasma cfDNA, and monitor concurrent mutations in parallel, providing an important contribution to the field. Using TAm-Seq, large genomic regions spanning six genes, including all exonic regions of *TP53*, were screened directly in plasma cfDNA. Using dPCR, this would first require the identification of mutations in tumour DNA, the design and validation of bespoke assays followed by quantification of each mutation in plasma DNA [as used in **Parkinson*,**

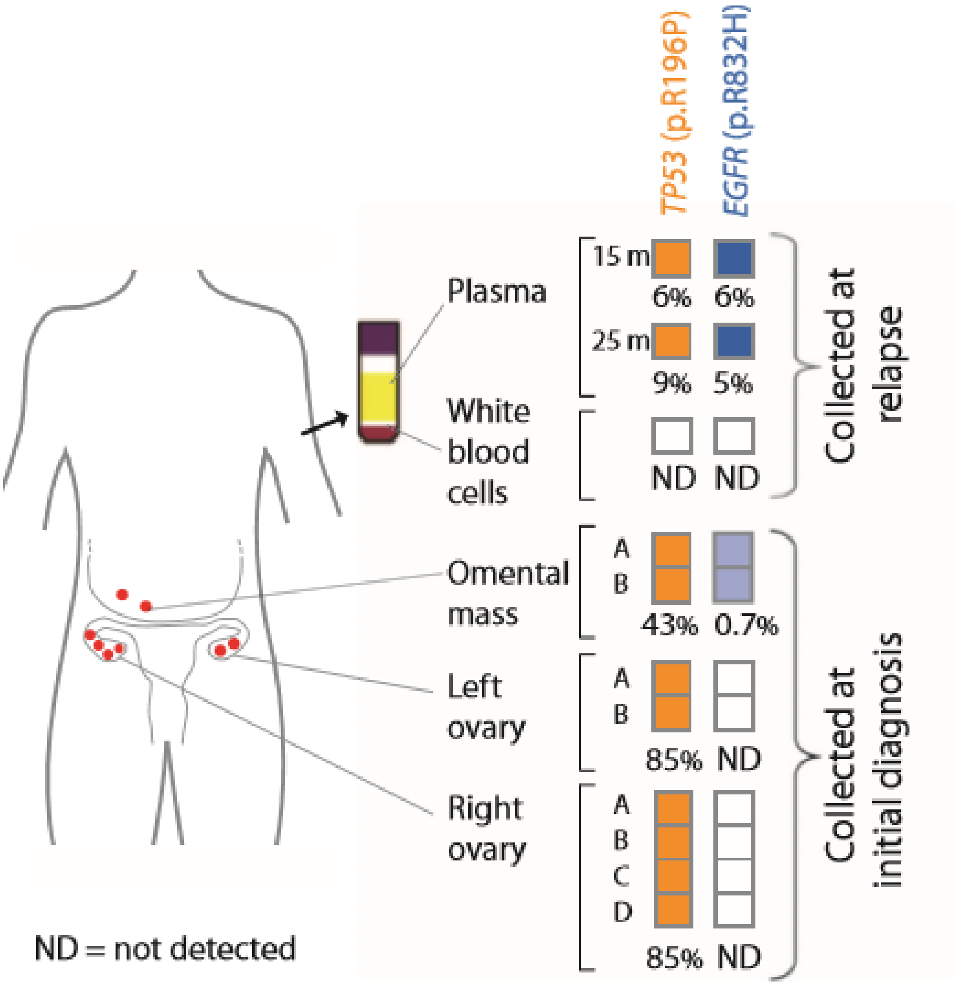

TP53 (p.R196P)
EGFR (p.R832H)
Plasma
White blood cells
Omental mass
Left ovary
Right ovary
15 m
6%
6%
25 m
9%
5%
ND
ND
A
B
43%
0.7%
A
B
85%
ND
A
B
C
D
85%
ND
Collected at relapse
Collected at initial diagnosis
ND = not detected

Figure 3.3: Retrospective analysis by TAm-Seq of plasma collected during patient follow-up and biopsy specimens collected at initial surgery. *EGFR p.R832H mutation identified in 2 plasma samples (dark blue boxes), collected 15 and 25 months after initial surgery, and in 2 omental specimens from initial surgery, but not identified in 6 ovarian specimens. TP53 p.R196P mutation identified in all tumour and plasma samples, but not in buffy coat. Percentages: mutant AFs. Empty boxes/"ND": no mutation identified/detected (below 0.8 confidence margin). Figure and Figure Legend adapted from* ***Forshew et al.*****(3)**

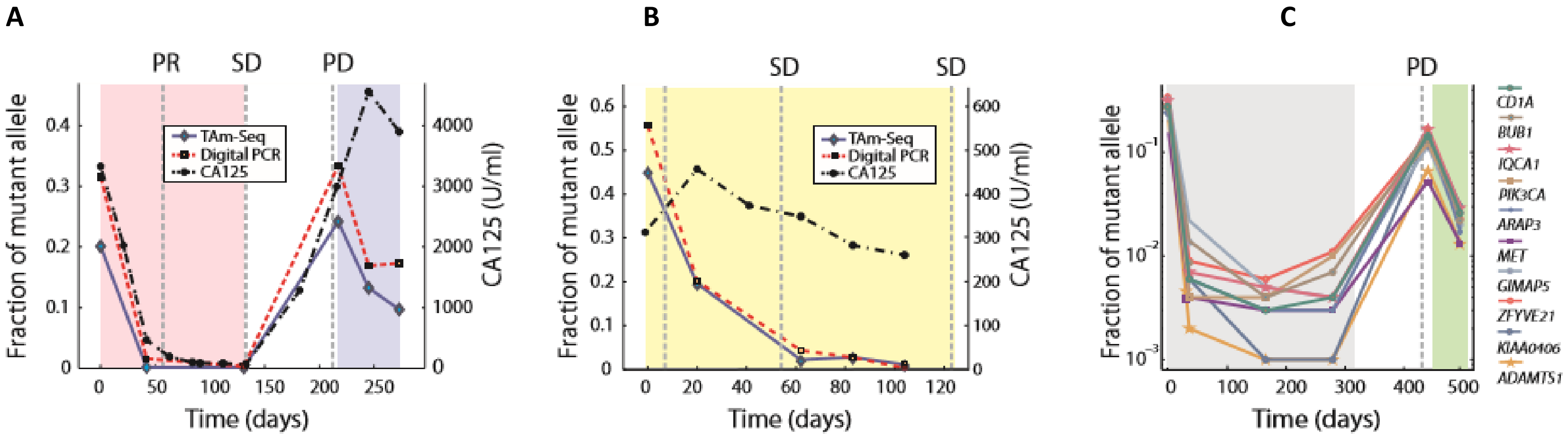


Figure 3.4: (A, B) Monitoring mutant DNA in plasma of HGSOC patients over time using TAm-Seq, mutant *TP53* by dPCR, compared with serum CA-125. (C) Dynamics of 10 concomitant mutations in plasma of a breast cancer patient. *Figures and Figure Legend adapted from* ***Forshew et al.*** **(3)**

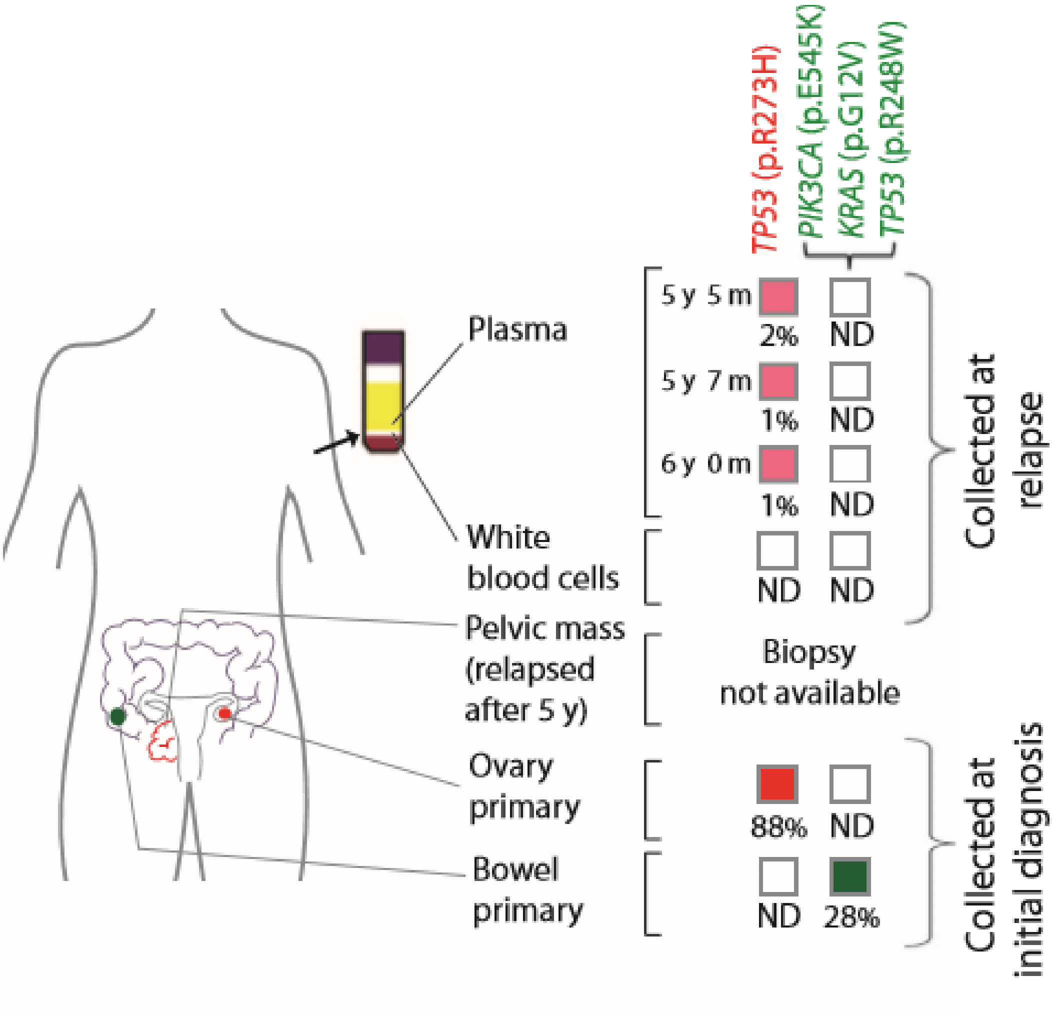

TP53 (p.R273H)
PIK3CA (p.E545K)
KRAS (p.G12V)
TP53 (p.R248W)
Plasma
5 y 5 m
2%
ND
5 y 7 m
1%
ND
6 y 0 m
1%
ND
White blood cells
ND
ND
Collected at relapse
Pelvic mass (relapsed after 5 y)
Biopsy not available
Ovary primary
88%
ND
Bowel primary
ND
28%
Collected at initial diagnosis

Figure 3.5: Retrospective analysis of samples from synchronous ovarian (red) and bowel (green) primary tumours collected at initial surgery and three plasma samples collected at relapse. *Percentages: mutant AFs indicated. Empty boxes/ "ND": no mutation identified/detected (below 0.8 confidence margin).Figure and Figure Legend adapted from* ***Forshew et al.*** **(3)**

Gale* et al.(**1**)]. TAm-Seq can be performed at relatively low-cost, compared to the alternatives at the time involving the bespoke design of multiple dPCR assays, or expensive WGS (Figure 3.6). It has potential as a clinical assay as it provides an ideal mid-point between sensitivity, ease of use, genomic coverage and cost (see Chapter 4).

Monitoring multiple mutations allows for tracking newly arising resistance mutations within the same assay, and opens up the possibility of patients being guided to treatment with other targeted therapies, such as osimertinib following the detection of an *EGFR* p.T790M resistance mutation. The ability to concurrently monitor multiple mutations using the same assay circumvents the issue of having to subdivide plasma DNA into different reactions, avoiding sampling bias. The Qiagen Therascreen *EGFR* RGQ PCR Kit version 2 is an *in vitro* diagnostic (IVD) assay for NSCLC patients, which assesses 29 different *EGFR* mutations to guide treatment with gefitinib, afatinib and dacomitinib. However, to assay all these mutations requires subdividing DNA into 7 different assays. Where levels are low, sampling noise may result in a higher proportion of false negatives.

Using TAm-Seq, ctDNA was detected at >2% AF at a sensitivity and specificity of >97%, with detection down to 0.14% AF for SNVs. This was a significant advance on standard practice at the time for analysis of mutations in FFPE tumour DNA, which generally involved performing Sanger sequencing to detect mutations >5% AF. Since TAm-Seq was published, several alternative amplicon-based sequencing methods have been developed including Signatera™, a patient-specific multiplex PCR technology developed by Natera™ and used in the TRACERx study; FireFly™, a rolling circle amplification method developed by Accuragen™; and SiMSen-Seq which uses hairpin-protected barcode primers to increase sensitivity (69,74,75).

At the time of publication, the main limitation of TAm-Seq was its sensitivity, compared to single-locus dPCR assays which are able detect down to 1 mutant molecule in 10,000 wild-type alles (0.01% AF). Since this time, however, TAm-Seq technology has been transferred to Inivata Ltd., leading to the development of enhanced TAm-Seq™ (eTAm-Seq™) technology. The InVisionFirst™-Lung assay is able to perform comprehensive genomic profiling of SNVs, indels, SVs, fusions and SCNAs in 36 genes, with an LoD90 of 0.25% AF for SNV detection, and detection down to 0.02% AF. The development and validation of these clinical diagnostic assays is described in more detail in Chapter 4 **(7,8).**

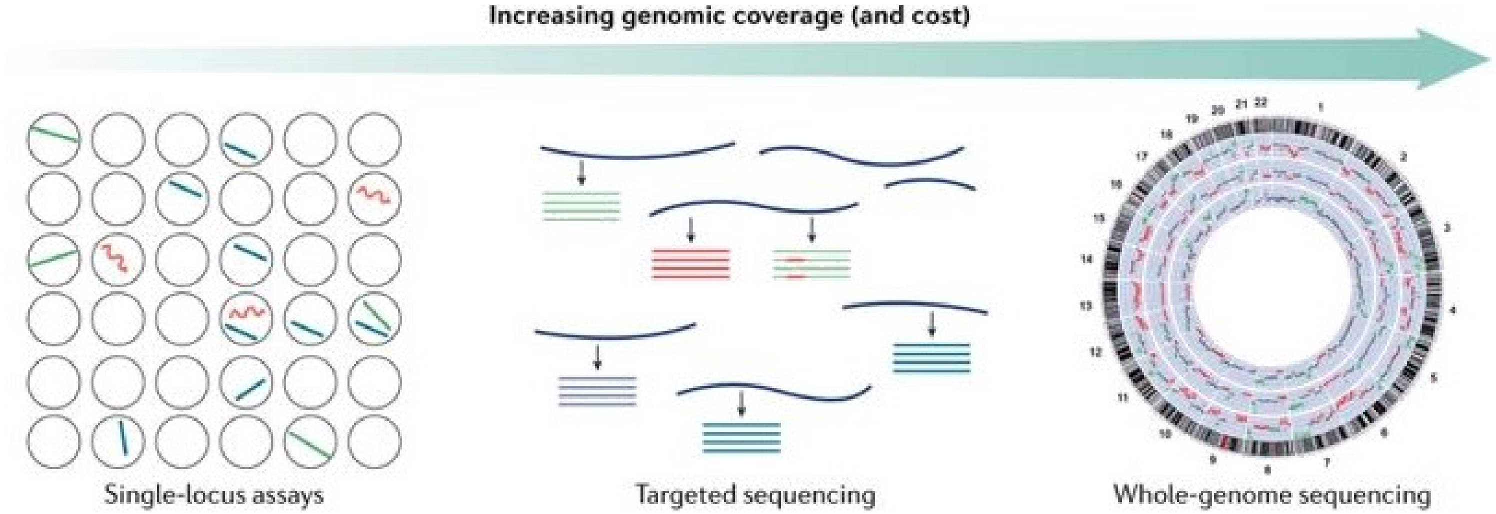

Increasing genomic coverage (and cost)
Single-locus assays
Targeted sequencing
Whole-genome sequencing

Figure 3.6: Analysis of cfDNA ranging from the interrogation of individual loci (eg. dPCR), targeted sequencing of multiple loci (eg. TAm-Seq) and whole genome analysis (eg. WGS), with increasing coverage and cost. *Figure and Figure Legend adapted from Wan et al.* (28)

3.2: Murtaza et al., Nature, 2013 (4)

3.2.1: Aims

The aim of this study was to use whole exome sequencing to determine if plasma taken before and after the development of treatment resistance can be used to assess genomic evolution, and identify potential mechanisms of acquired resistance to therapy (Figure 3.7). To achieve optimal results, timepoints were selected where the ctDNA allele fraction was >10%, based on dPCR and TAm-Seq data.

3.2.2: Results

WES was performed on cfDNA extracted from 1 NSCLC, 2 breast cancer and 3 ovarian cancer patients undergoing treatment. Libraries were prepared using the ThruPLEX-FD™ kit (Rubicon Genomics), optimised for use with fragmented DNA. AFs were compared between exome, dPCR and TAm-Seq data and showed good concordance (Pearson's, exome v dPCR: 0.79; $p < 0.0001$; exome v TAm-Seq: 0.83; $p < 0.0001$).

For two cases, tumour biopsies were available, collected both at the time of initial presentation (9 and 4.5 years earlier) and at the same time as plasma. Genome-wide copy number analysis showed similar SCNA profiles in the plasma and metastatic tumour data (Figure 3.8A/B), demonstrating that the tumour genome is represented in plasma. In breast cancer case 1, of 151 mutations identified in either the plasma or metastatic biopsy, 93 were identified in both, with AFs generally higher in the plasma. The correlation of AFs was high for mutations present in the primary tumour (Pearson's, 0.71), but did not correlate when taking into account all mutations also in the synchronous biopsy (Pearson's. -0.22, Figure 3.8C).

A modified version of the genome-wide aggregated allelic loss (GAAL) algorithm developed by Chan et al. was used to determine the fractional concentrations of tumor-derived DNA in plasma (76). A range of 15 – 121 (median:49) non-synonymous mutations were identified in each case. Mutations with increased AF following treatment included those in known cancer genes, and in genes associated with drug resistance, including an *EGFR* p.T790M mutation in the NSCLC patient following treatment with gefitinib, and a splicing mutation in *GAS6*, the ligand for AXL, which is a tyrosine kinase receptor (Figure 3.9).

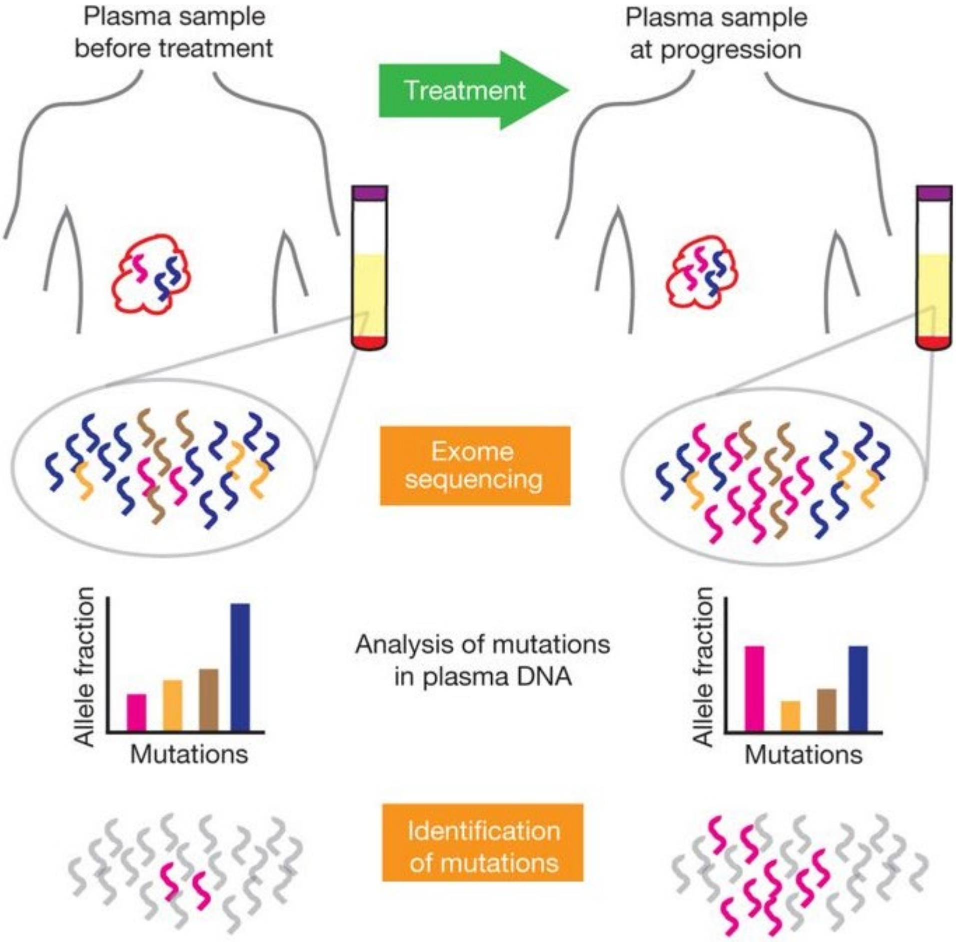

Plasma sample before treatment
Treatment
Plasma sample at progression
Exome sequencing
Allele fraction
Mutations
Analysis of mutations in plasma DNA
Allele fraction
Mutations
Identification of mutations

Figure 3.7: Overview of workflow for the identification of treatment-associated mutational changes by exome sequencing of plasma before and after treatment. *Figure and Figure Legend adapted from* ***Murtaza et al.*** **(4)**

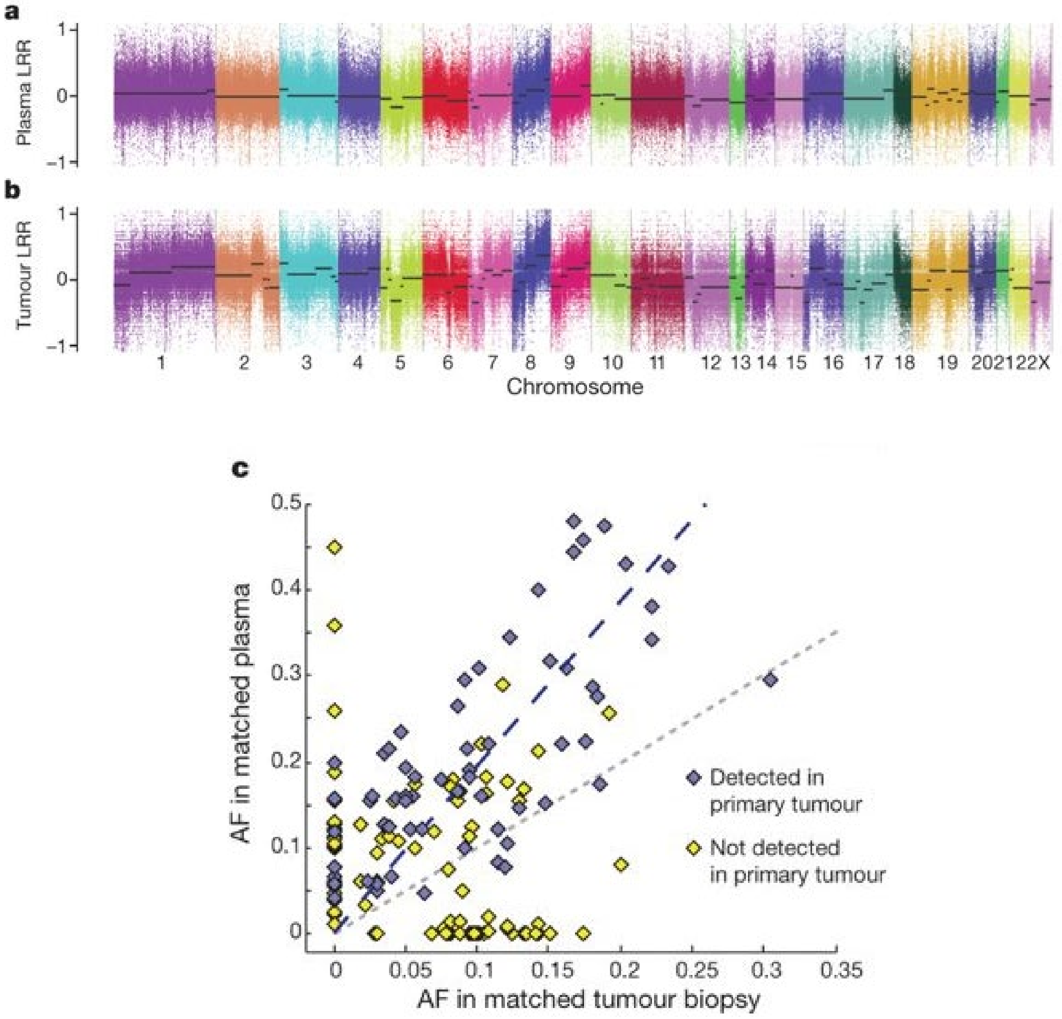


Figure 3.8: Genome-wide analysis of exome data from plasma DNA and synchronous metastatic tumour DNA. *Copy number profiles in (a) plasma and (b) tumour biopsy in ovarian patient. LogR ratios (LRR) calculated from exome data from (a) plasma and germline DNA (b) tumour and germline DNA; (c) AF of mutations identified in plasma or tumour in breast cancer patient. Grey dotted line: Equality; Blue dashed line: Median of the AF ratio for mutations in both samples (slope: 1.93). Figures and Figure Legend adapted from* ***Murtaza et al.*** **(4)**

Activation of the AXL pathway is known to cause resistance to TKIs in NSCLC, and lapatinib resistance in ER+HER2+ breast cancer cell lines (77,78).

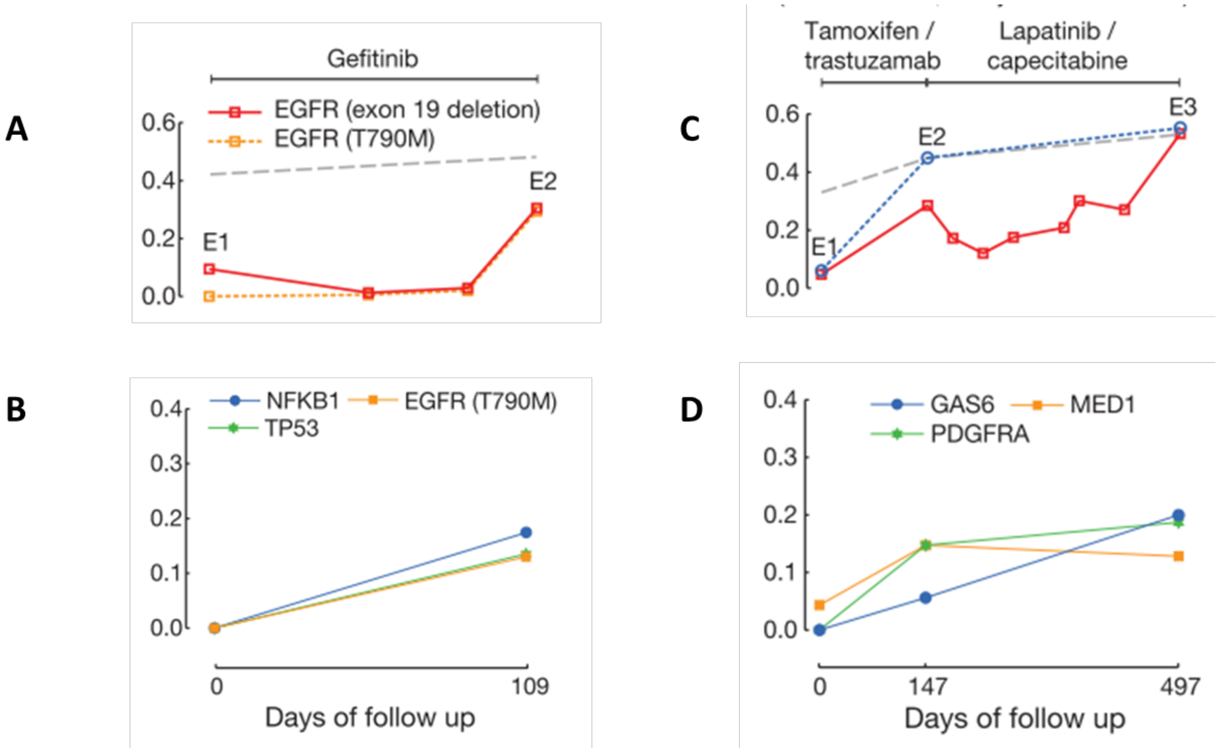


Figure 3.9: Mutations showing evidence of genomic tumour evolution. *(A, B) Lung cancer case 6 (C, D) Breast cancer case 2. Upper subpanels (A, C): Timecourses for allele fractions of 'anchor mutations' used to quantify ctDNA, and fractional concentrations using (A) dPCR of EGFR exon 19 deletion & p.T790M. (C) TAm-Seq of ATM p.I2948F; Grey dotted line: Tumour burden estimate using a modification of Genome-wide aggregated allelic loss. Exome sequencing timepoints indicated by E1, E2 or E3. Lower subpanels (B, D) AF in exome sequencing data of different mutations indicated by different colour lines. Figure and Figure Legend adapted from* ***Murtaza et al.*** **(4)**

3.3: Murtaza et al., Nature Communications, 2015 (5)

3.3.1: Aims

A further study was performed to provide in-depth analysis of plasma and multi-region tumour specimens taken from a patient with metastatic ER+HER2+ (oestrogen receptor-positive, human epidermal growth factor receptor 2-positive) breast cancer treated with tamoxifen and trastuzumab, followed by lapatinib **(5)**. The aim was to determine whether plasma could be used to identify changes in the tumour genome and non-invasively assess clonal evolution. WES and targeted deep sequencing was used to analyse 9 plasma samples and 8 tumour biopsies, including specimens taken from different metastases and at different timepoints, including at autopsy (Figure 3.10A).

3.3.2: Results

Analysis identified stem mutations (present in all tumour biopsies), metastatic-clade mutations (present in all metastases) and private mutations (unique for a specific biopsy). 362 candidate non-synonymous SNVs were identified by WES, and 207 mutations were validated by targeted deep sequencing. PyClone was used to cluster the data based on clonal architecture (Figure 3.10B, 3.11). Ubiquitous stem mutations (n=23 at >2% AF) showed the highest levels in plasma with mean AFs ranging 3.8% - 34.9% AF, followed by metastatic-clade mutations (n=26, 2.5% - 19.1% AF) then private muations (n=126, Figure 3.12A). 13/26 metastatic-clade mutations were also detected in lymph node biopsy P1.2 at low frequency, indicating a common ancestor. Plasma abundance was calculated for each high-confidence private mutation, relative to timepoint 1 (T1). This demonstrated an increase in abundance of mutations present in the chest wall mass following lapatinib resistance, which agreed with imaging data (Figure 3.12B). The most abundant private mutation at this time was *ERBB2* p.H809G, with elevated levels observed in plasma following lapatinib resistance (Figure 3.12C). A *PIK3CA* p.E542K mutation was elevated at the time of progression on tamoxifen and trastuzumab, but decreased following treatment with lapatinib (Figure 3.12D). This is consistent with data indicating that the PI3K/AKT pathway is associated with resistance to trastuzumab and endocrine therapy.

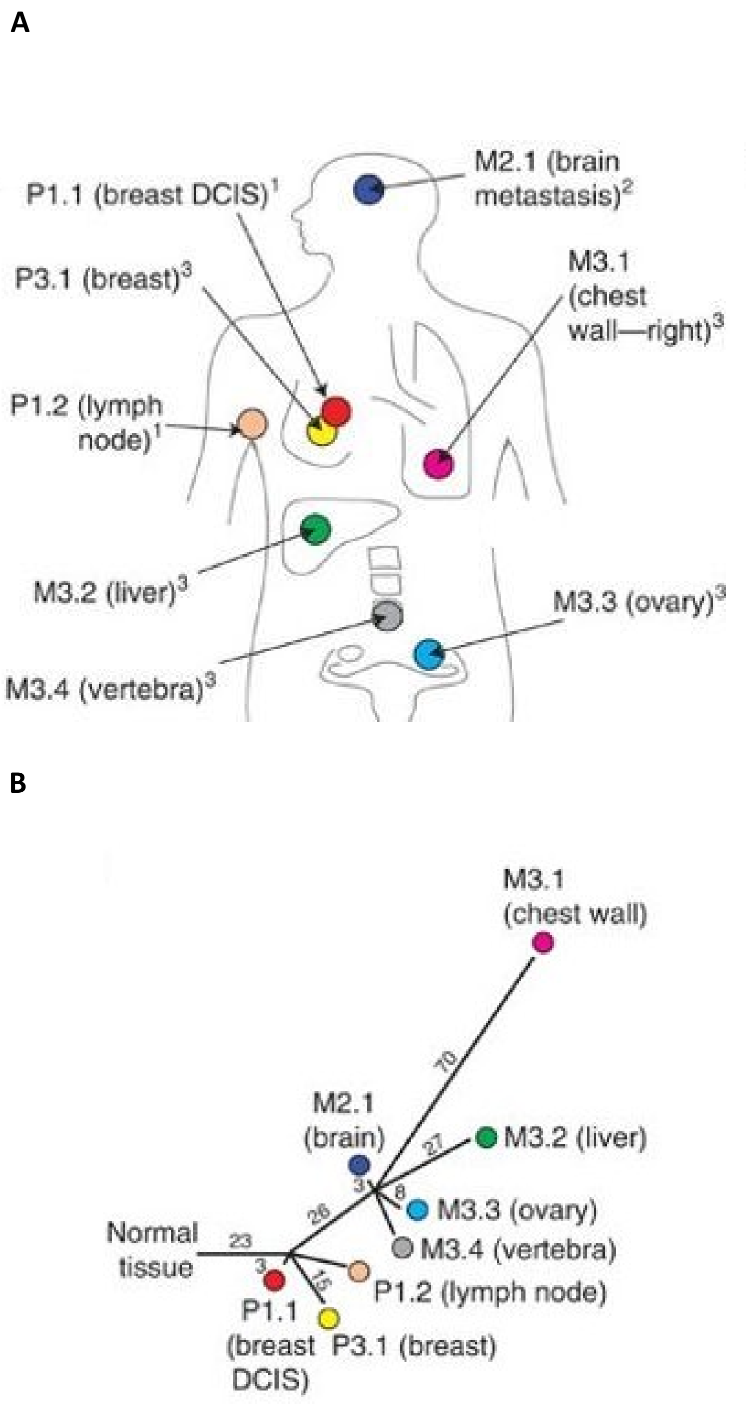

A
P1.1 (breast DCIS)¹
M2.1 (brain metastasis)²
P3.1 (breast)³
M3.1 (chest wall—right)³
P1.2 (lymph node)¹
M3.2 (liver)³
M3.3 (ovary)³
M3.4 (vertebra)³
B
M3.1 (chest wall)
70
M2.1 (brain)
27
M3.2 (liver)
3
8
M3.3 (ovary)
26
Normal tissue
23
M3.4 (vertebra)
3
15
P1.2 (lymph node)
P1.1 (breast DCIS)
P3.1 (breast)

Figure 3.10: Inference of clonal structure from multiregional tumour biopsies (A) Tumour samples collected. *P = Primary; M = Metastasis; 1 = Collected at diagnosis; 2 = Collected at time of resection of brain metastasis; 3 = Collected at autopsy.* ***(B) Tumour phylogenetic tree.*** *Length of each branch correlates to number of mutations. Figure and Figure Legend adapted from* ***Murtaza et al.*** **(5)*.***

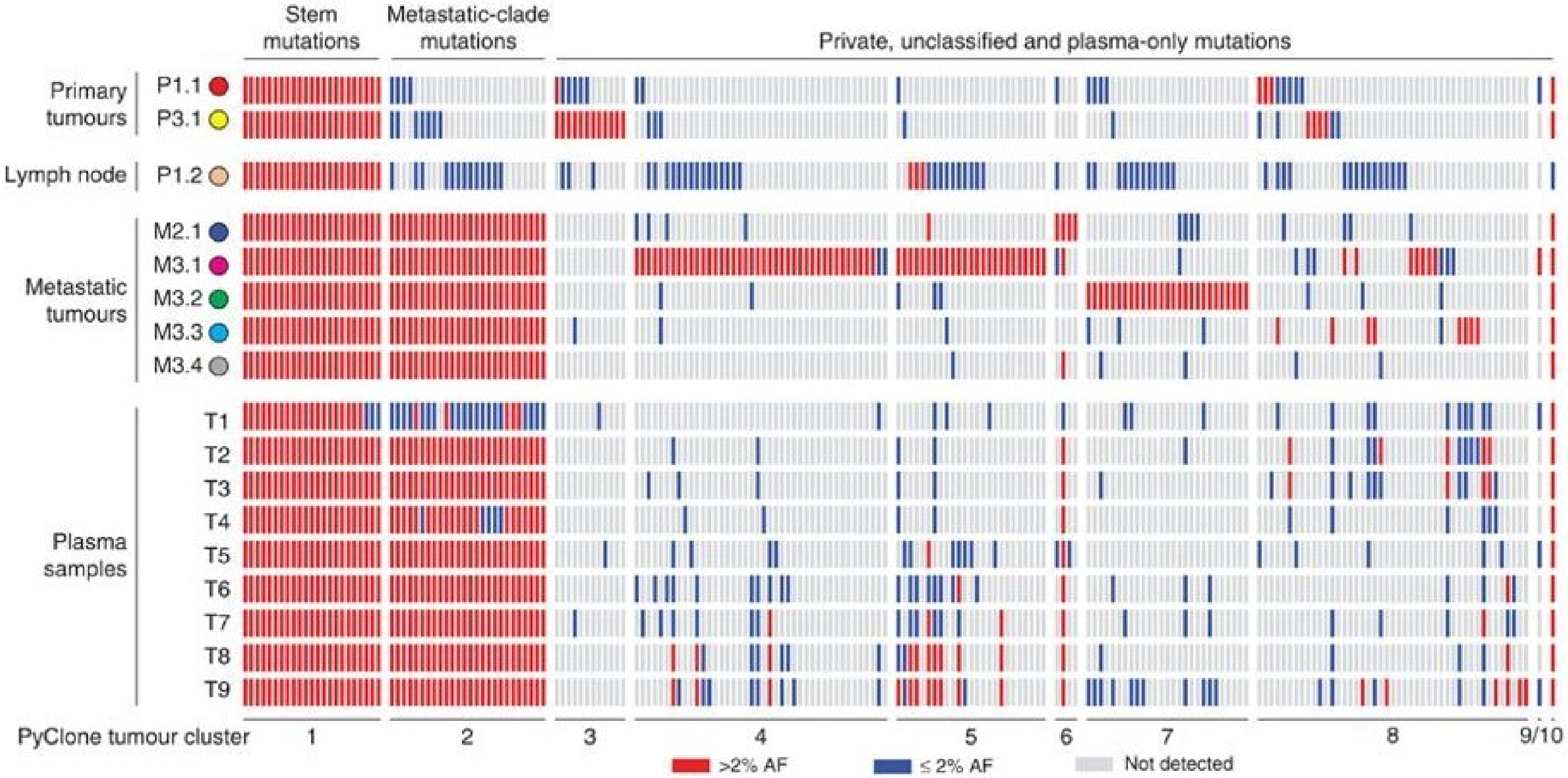


Figure 3.11: Distribution of validated mutations in tumour and plasma samples. *Ordered by clusters inferred using PyClone. Stem: observed in all tumour samples; metastatic-clade: in all metastatic tumours; Red rectangles: mutations with AF >2%; Blue rectangles: Mutations with AF ≤2% significantly above background; Grey: Not detected. P = Primary; M = Metastasis; 1 = Collected at diagnosis; 2 = Collected at time of resection of brain metastasis; 3 = Collected at autopsy. Figure and Figure Legend adapted from* ***Murtaza et al*****.** **(5)**

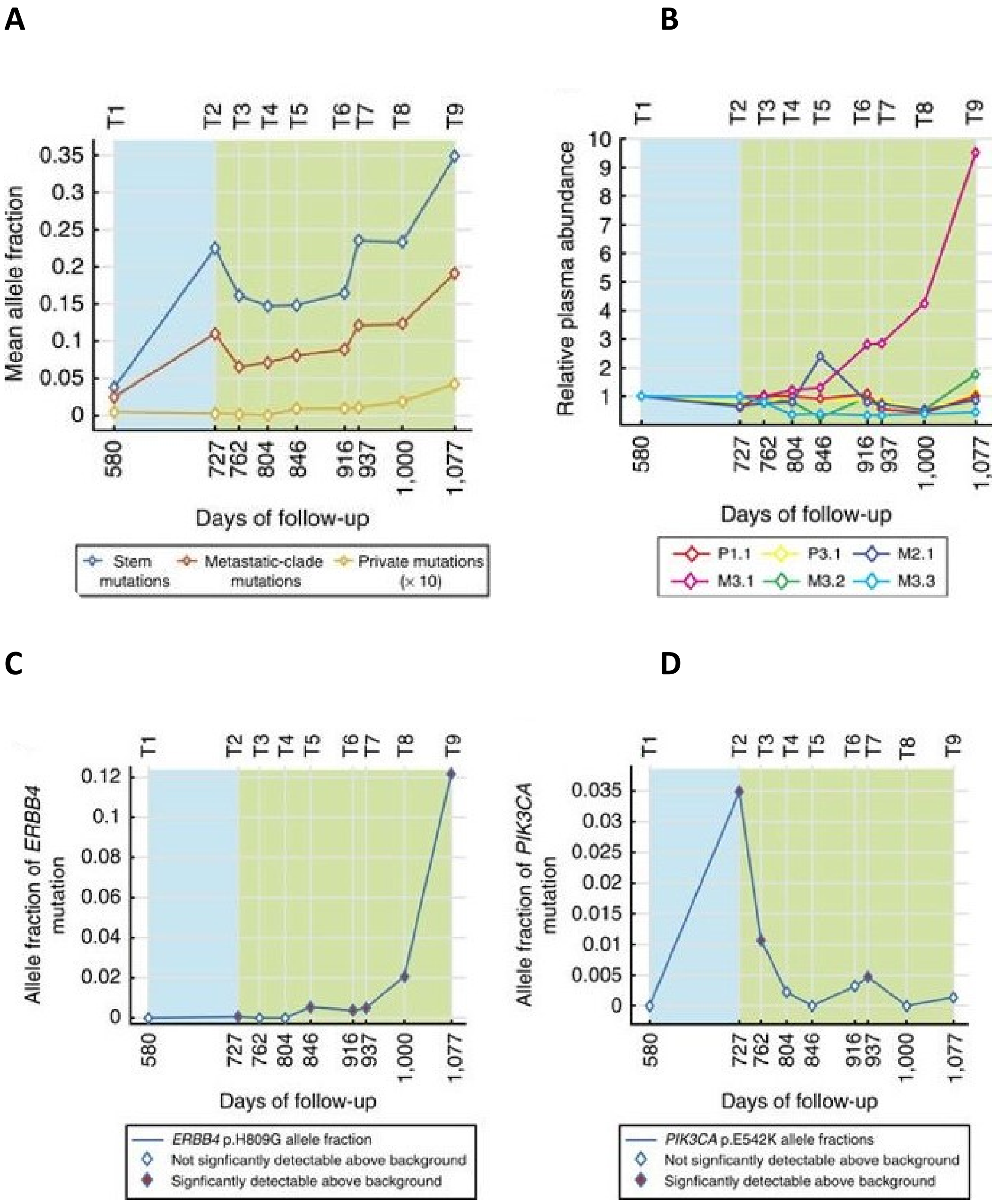


Figure 3.12: (A) Mean allele fractions (AFs) of stem, metastatic-clade and private mutations) during treatment *(private x 10 to see trend)*. *Shaded areas: Treatment.* **(B) Plasma abundance relative to plasma timepoint 1 (T1)** *[calculated as product of AF in tumour (normalised to mean of stem mutations) and AF in plasma, summed across all private mutations for each tumour], normalised to T1 to account for different number of private mutations in each tumour* **(C) Dynamics of *ERBB4* p.H809G mutation, and (D) Dynamics of *PIK3CA* p.E542K mutation over time by deep amplicon sequencing.** *P = Primary; M = Metastasis; 1 = Collected at diagnosis; 2 = Collected at time of resection of brain metastasis; 3 = Collected at autopsy. Figure and Figure Legend adapted from* ***Murtaza et al.*** **(5)**

3.4. Discussion

The **Murtaza et al.** (**Nature)** paper demonstrated for the first time that exome sequencing using hybrid capture could be used to analyse ctDNA, and investigate tumour heterogeneity and mechanisms of acquired resistance to therapy **(4)**. This provided an important contribution to the literature. Whilst hybrid capture had previously been used to analyse tumour DNA, this was the first time that it had been used to non-invasively analyse the tumour genome in plasma. Exome sequencing of plasma samples, taken before and after the development of treatment resistance, identified known resistance mutations, providing proof-of-concept data, and demonstrating the value of this approach.

Hybrid capture involves hybridisation of biotinylated DNA or RNA baits to pull down and enrich for sequences of interest (79), which can be performed on a custom or exome-wide scale. It has the advantage over amplicon-based sequencing methods, such as TAm-Seq, in that it can target larger genomic regions, and is more amenable to analysis of SVs, SCNAs, tumour mutation burden and detection of microsatellite instability (MSI). Amplicon sequencing is limited to targeting regions incorporated within the PCR primers. This is particularly challenging when analysing fragmented DNA, as a few fragments may be lost to analysis that do not encompass both priming sites. It may also potentially result in more uneven sequencing coverage, although the two-step amplification process employed in TAm-Seq, involving multiplex pre-amplification followed by single-plex PCR in a microfluidic chip, resulted in much improved evenness of coverage. Hybrid capture has the disadvantage that a significant proportion of molecules are lost during library preparation due to the inefficiency of adaptor ligation. Attachment of adaptors by PCR is considerably more efficient, resulting in less loss of material, which is vital when analysing low frequency mutant alleles in patient plasma.

The main limitation of this study is that it is small cohort, and required analysis of samples with allele fraction >10% AF, which had to be pre-determined. However, the study demonstrated that plasma can non-invasively represent the tumour genome, and provide a summary of different mutations present in different tumour lesions that may be missed in a single tumour biopsy.

In the follow-up **Murtaza et al**. **(Nature Communications)** paper, a single patient with metastatic ER+HER2+ breast cancer was analysed using exome and targeted deep sequencing of plasma and tumour biopsies, including some taken at autopsy **(5)**. This provided valuable information demonstrating that stem mutations had the highest levels in plasma, followed by metastatic-clade, then private mutations. This study provided valuable insight into the use of ctDNA as an alternative to tissue profiling. However, methods used are too expensive for routine clinical testing, requiring high depth exome sequencing, and targeted deep sequencing of candidate mutations to get the required sensitivity.

Since these publications, hybrid capture has been developed with much improved sensitivity, now comparable with amplicon-based sequencing. Similar methods are being commercially developed by several companies as clinical diagnostic ctDNA assays. This includes the Guardant360® test by Guardant Health, covering 73 genes with detection of >99.9% mutations at >0.25% AF, and down to 0.04% AF (80,81). Foundation Medicine™ have developed a FoundationOne Liquid™ test, covering 70 genes, with 95.8% mutations detected at 0.25% - 0.5% AF (82,83). PlasmaSELECT™ 64+MSI has been developed by Personal Genome Diagnostics, which assays 64 genes, with 99.4% sensitivity for detection of SNVs ≥0.5% AF, and can detect MSI to assess potential response to checkpoint inhibitors (84).

The recent increase in sensitivity of hybrid capture has been afforded by the introduction of molecular barcodes into the NGS library preparation process. In 2011, Kinde et al. developed Safe-SeqS, which incorporates unique identifiers (UIDs) to individually tag different molecules to suppress error rates. This was applied to the analysis of Pap smear tests for endometrial and ovarian cancer (85,86). In 2012, Schmitt et al. developed duplex sequencing which independently tags *each strand* of a DNA duplex with unique identifiers, to distinguish true mutations, present at the same place on both strands, from PCR or sequencing errors, which occur at random on a single strand (87). Newman et al. developed a modified molecular barcoding approach and iDES (integrated digital error suppression) to analyse ctDNA using CAPP-Seq (Cancer Personalized Profiling by deep Sequencing) (88). CAPP-Seq, first published in 2014, uses a hybrid capture panel, termed a CAPP-Seq selector library, specific for patients with NSCLC, based on analysis of WES data from 407 patients profiled by The Cancer Genome Atlas (TCGA) (67). Using CAPP-Seq, ctDNA was detected in 50% patients with Stage I disease, and 100% of Stage II-IV patients, with 96% specificity and detection down to 0.02% AF. In their follow-up paper, further enhancements were included to suppress error, including the introduction of both single and double-stranded barcoding, incorporating single-stranded ‘index’ and double-stranded ‘insert’ UIDs as exogenous barcodes, and endogenous barcodes corresponding to the mapped genomic co-ordinates of molecules (88). Further error-suppression bioinformatics methods were included, performing ‘background polishing’ to remove position-specific errors. Using iDES-enhanced CAPP-Seq, ctDNA could be detected down to 4 in 10^5 cfDNA molecules. Analytical performance showed 92% sensitivity and >99.99% specificity, with 90% clinical sensitivity and 96% specificity.

A further NGS method that has proved valuable in the clinic for analysing ctDNA is the use of shallow whole genome sequencing (sWGS, 0.1-0.4x coverage) to identify tumour-associated copy number changes to screen for tumour-specific aneuploidy in plasma samples with high ctDNA levels (54,55). This approach is discussed in more detail in Chapter 5.

In Chapter 4, I will discuss work I performed to develop enhanced TAm-Seq™ to ISO15189:2012 regulatory standards for patient benefit, and to assess appropriate blood tubes and processing conditions for liquid biopsy studies **(6–8).**

Chapter 4: Clinical diagnostic ctDNA assay development

Prior to clinical implementation, all diagnostic assays must be rigorously tested, and developed to appropriate regulatory quality standards to ensure they are fit-for-purpose and do not compromise patient safety. In this Chapter, I will discuss work I co-led, published by **Risberg et al.**, to assess the effects of different collection and plasma processing protocols on the levels of ctDNA **(6)**. In addition, I will discuss work by **Gale et al.** and **Plagnol et al**. on the development and analytical validation of enhanced TAm-Seq™ (eTAm-Seq™) and the InVisionFirst™-Lung assay to ISO15189:2012 standards for the detection of low frequency alleles in plasma of advanced NSCLC patients **(7,8)**.

4.1: Risberg, …, Gale† et al., Journal of Molecular Diagnostics, 2018 (6)

4.1.1: Aims

Detection of ctDNA is challenging given the potentially low levels present in plasma. Pre-analytic factors, such as a delay in plasma processing of blood collected into EDTA (Ethylenediaminetetraacetic acid) tubes, can result in increased levels of genomic DNA (gDNA) due to leukocyte lysis, which may affect the detection of low frequency mutant alleles, so it is important to collect samples using optimised protocols. Processing recommendations include performing an initial centrifugation step to separate plasma within hours of venepuncture to avoid leukocyte lysis, followed by a second higher speed centrifugation step to clear cellular material (89). More recently, cell-stabilising blood collection tubes, such as Streck Cell-free DNA BCT® tubes (BCT), have been developed which stabilise nucleated blood cells, enabling plasma processing to be delayed, thereby facilitating processing under more controlled conditions and within centralised laboratories. In **Risberg et al**., five different Modules were performed to assess the effects of the following on the levels of ctDNA and cfDNA using dPCR, TAm-Seq or sWGS: [1] delayed plasma processing, [2] different storage temperatures, [3] different blood collection tubes, [4] different centrifugation protocols and [5] sample shipment.

4.1.2: Results

4.1.3: Module 1: Assessment of the effects of delayed processing

In this Module, we assessed the effects of delaying plasma processing of K_3EDTA-collected blood for 0, 6, 24, 48, 96 hours and 1 week at room temperature (RT, 19°C-25°C) by quantifying cfDNA, mutant AF and ctDNA levels using dPCR. Data was expressed as a ratio of the immediately processed EDTA 0 hour sample (denoted E.RT.0h) to enable direct comparisons to be made. cfDNA levels significantly increased following a 48, 96 hour and 1-week delay ($p < 0.05$, Mann-Whitney Rank Sum test), with a corresponding decrease in mutant AF, whilst ctDNA levels remained constant.

4.1.4: Module 2: Assessment of the effects of different storage temperatures

In Module 2, K_3EDTA-collected blood tubes were stored at RT or 4°C for 24, 48 or 96 hours prior to plasma processing to assess the effects of different storage temperatures. cfDNA levels significantly increased after 48 hours at RT ($p < 0.05$, Mann-Whitney Rank Sum test, Figure 4.1). Whilst still elevated, levels were lower following 4°C storage. cfDNA levels significantly increased after 96 hours at both storage temperatures.

4.1.5: Module 3: Assessment of different blood collection tubes

Prior to implementing cell-preservation tubes in a clinical study, it is important to assess if the tubes can reproducibly stabilise cfDNA levels for prolonged periods, and if the preservative has any adverse effects on DNA integrity. In Module 3, BCT tubes stabilised cfDNA levels up to 1 week at RT prior to plasma processing, unlike K_3EDTA tubes where levels were significantly elevated (Figure 4.2). TAm-Seq analysis was performed analysing DNA extracted from BCT-collected plasma delayed for 0, 96 hours and 1 week, demonstrating no significant effect on background NGS error using either blood tube or different processing conditions.

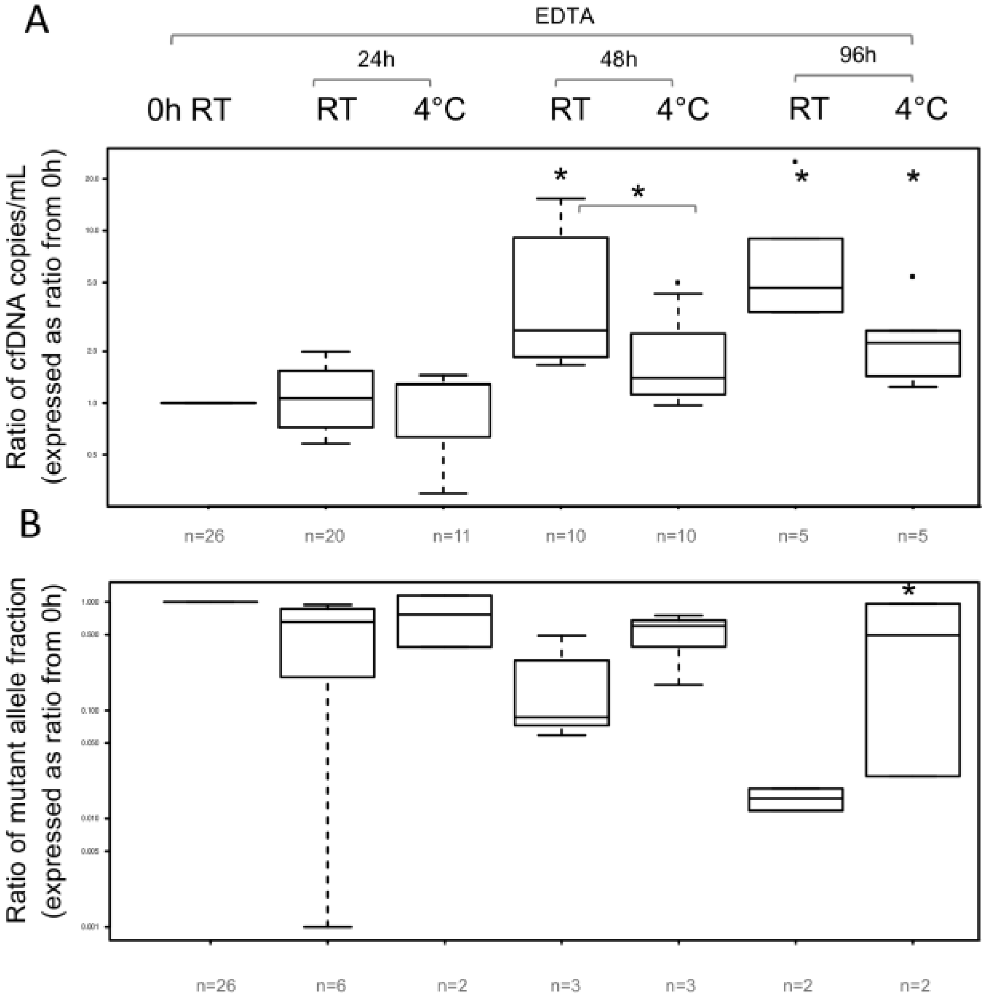


Figure 4.1: Effect of storage temperature on K$_3$EDTA-collected blood on (A) cfDNA copies/mL; (B) AF following storage at RT and 4°C at 24, 48 and 96 hours prior to plasma processing. *Bottom and top of boxplots represent first and third quartiles; Line: median. Data expressed as ratio to each patients' EDTA Room Temperature immediately processed sample (E.RT.0h); *p < 0.05 (Mann-Whitney Rank Sum test, versus E.RT.0h). Figures and Figure Legend adapted from* ***Risberg et al.*** **(6)**

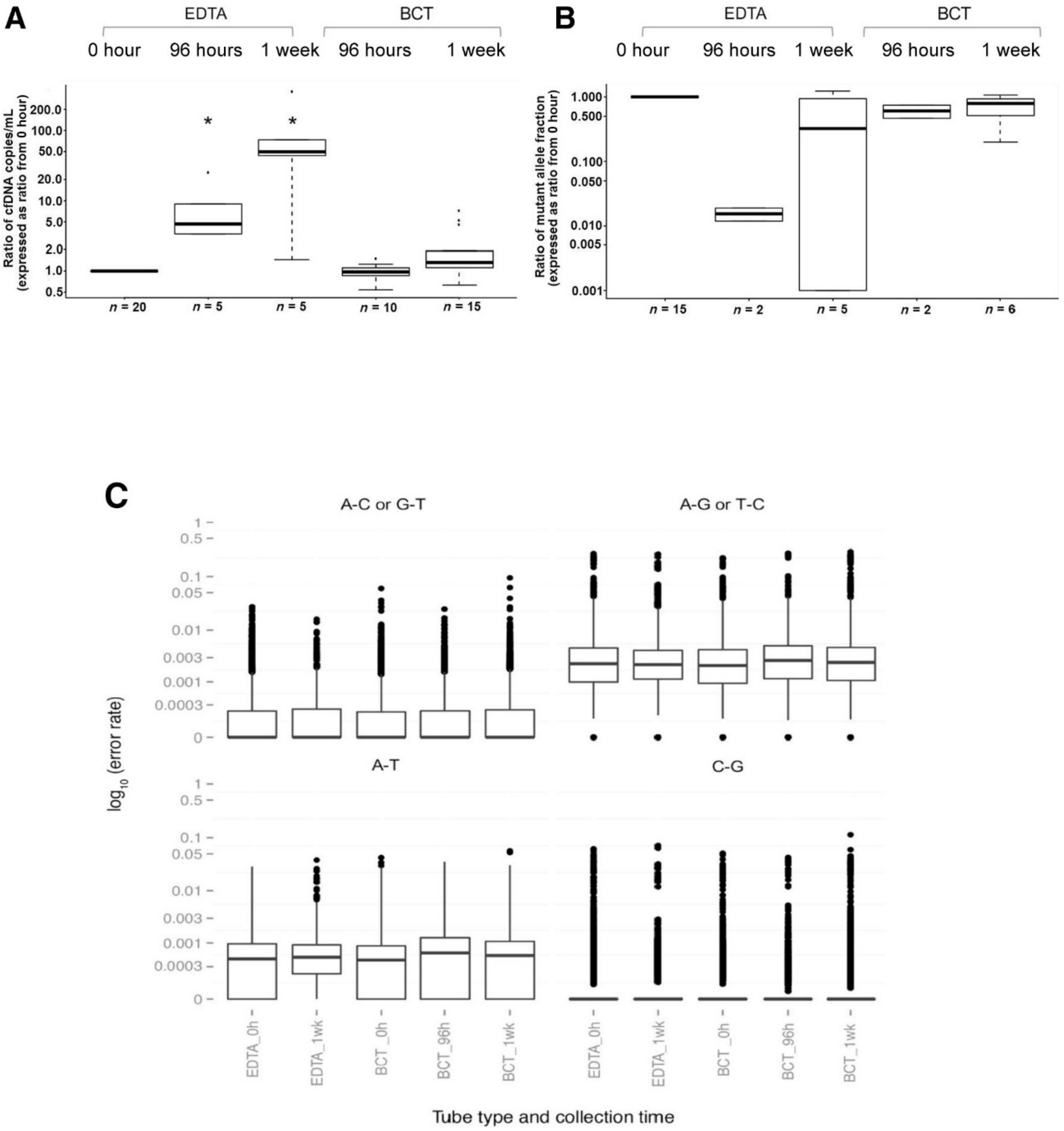


Figure 4.2: Effects of using different blood tubes on (A) cfDNA copies/mL and (B) mutant allele fraction in K$_3$EDTA tubes delayed for 0 hours, 96 hours and 1 week v BCT tubes delayed for 96 hours and 1 week. (C) Background error rates: Distribution of ratio of non-reference (ie. mutant)/reference alleles according to mutation types using TAm-Seq. *Bottom and top boxplots represents first and third quartiles; Line: median; (A, B) Data expressed as ratio to each patients' EDTA Room Temperature immediately processed sample (E.RT.0h) (C) log10 scale; *p < 0.05 (Mann-Whitney Rank Sum test, versus E.RT.0h). Figures and Figure Legend adapted from* ***Risberg et al.*** **(6)**

4.1.6: Module 4: Assessment of different centrifugation protocols

Our standard plasma processing protocol involves double centrifugation of whole blood, with the initial centrifugation performed in a mega-centrifuge, followed by fast centrifugation of plasma in a bench-top microcentrifuge. Unfortunately, many hospitals do not have access to a bench-top microcentrifuge.

In Module 4, three different centrifugation protocols were tested:
Protocol A: Centrifugation at 820 x g for 10 minutes (mega-centrifuge), then 14,000 x g for 10 minutes (bench-top microcentrifuge);
Protocol B: Centrifugation at 1,600 x g for 10 minutes (mega-centrifuge), then 14,000 x g for 10 minutes (bench-top microcentrifuge);
Protocol C: Centrifugation at 1600 x g for 10 minutes, then 3000 x g for 10 minutes (both steps performed in a mega-centrifuge).
Results showed no statistically significant difference in total cfDNA levels by dPCR, or mutant AF by TAm-Seq, between the three centrifugation protocols.

4.1.7: Module 5: Assessment of the effects of sample shipment

In Module 5, blood was collected from 13 HGSOC patients into either K_3EDTA tubes and processed immediately, or BCT tubes and stored at RT or shipped using Royal Mail postage, and processed at different times (from 48 hours to 5 days). Analysis of cfDNA levels showed no statistical difference in cfDNA levels and *TP53* mutant AF between the different collection methods. sWGS data also showed no difference in SCNAs or background noise profiles using these different collection protocols.

4.1.8: Discussion

Risberg et al. assessed whether different pre-analytic factors affect plasma quality. Results demonstrate a 48 hour delay in processing of K_3EDTA-collected blood resulted in significantly elevated levels of total cfDNA, and a corresponding decrease in mutant ctDNA levels, believed to be due to contaminating gDNA released during leukocyte lysis (90). A small but non-significant increase was also observed at 6 and 24 hours. Furthermore, results demonstrated storage of K_3EDTA blood at 4°C for 48 hours reduced cfDNA levels compared to room temperature storage, although levels were significantly elevated by 96 hours at both temperatures. Although this would need to be verified in a larger cohort, this provides practical guidance as to an appropriate storage temperature of K_3EDTA tubes in the event of an unavoidable delay in plasma processing. BCT tubes stabilised cfDNA and mutant AF levels for up to 1 week prior to plasma processing, although a slight but non-significant increase was also observed between 96 hours and 1 week.

These results show a similar trend to previous studies using K_2EDTA tubes (containing spray dried di-potassium salt as anti-coagulant, rather than K_3EDTA tubes containing tri-potassium salt in liquid form). Parpart-Li et al. demonstrated significantly elevated gDNA levels when processing was delayed 24 hours ($p < 0.05$), and 48 hours ($p < 0.01$, Student *t* test) (91). Total genomic equivalents/mL plasma (GE/mL) were not significantly elevated following 4°C storage of K_2EDTA tubes for 72 hours, but total mutant AF/mL decreased. BCT tubes stabilised cfDNA levels up to 1 week, with a 6% increase in total GE/mL compared to a 13-fold increase using K_2EDTA tubes. Medina Diaz et al. showed cfDNA levels were stable in BCT tubes for up to 5 days (92), and **Plagnol et al**. showed mutant DNA spiked into BCT blood was stable for 10 days **(8)**.

Interestingly, Parpart-Li and Medina Diaz both observed a small decrease in plasma volume in both K_2EDTA and Streck BCT tubes stored over time. This was most notable in BCT tubes stored for 5 days at 4°C and 40°C (91,92), outside the manufacturer's storage recommendations of 6°C - 37°C. Whilst **Risberg et al.** did not find any differences in SCNAs following shipment within the UK, Medina Diaz et al. did find that extremes of temperature appeared to visibly affect the interface layer above the buffy coat, affecting 20%-50% of the plasma fraction at 4°C following 3 and 5 day storage, and the appearance of haemolysed plasma when stored at >40°C. qPCR was used to assess ratios of both long (402bp) and short (96bp) fragments to determine the release of long gDNA fragments, demonstrating an increased ratio of 402:96bp fragments at extreme temperatures. As a result, the authors recommend BCT tubes are shipped using temperature-controlled insulated packaging to avoid compromising cfDNA quality.

Risberg et al. next assessed whether the BCT cell preservative had any adverse effect on DNA used in downstream TAm-Seq and sWGS assays, which could potentially result in artefacts that may be mis-interpreted as genuine tumour-specific mutations. Fortunately, background levels were similar using DNA from both BCT and K_3EDTA tubes, in agreement with other published data using different NGS assays (91,92). This indicates that BCT tubes may provide a practical alternative to K_3EDTA tubes for analysis using TAm-Seq and sWGS, although larger studies should be performed to verify this result prior to clinical implementation.

One of the main challenges I have encountered in establishing multi-centre clinical collections of plasma for ctDNA studies is that a number of hospitals do not have access to a bench-top microcentrifuge required for the second centrifugation step. **Risberg et al.** tested alternative protocols. Similar results were obtained if the second centrifugation was performed in a megacentrifuge at 3000 x g, compared to centrifugation at 14,000 x g in a bench-top microcentrifuge. This provides valuble data, opening up the possibility to expand clinical collections to additional sites that may have previously been limited based on lack of centrifuge equipment, particularly where expensive cell-preservation tubes are not an option. These studies should be repeated on a larger scale to verify these findings prior to clinical implementation.

The main limitations of the **Risberg et al.** study are the small number of samples analysed, that available blood volumes limited the number of comparisons that could be performed, and that the study was conducted in a non-regulated lab. In addition, only two different blood collection tubes were assessed. Since our study was initiated, new cfDNA preservation tubes have been commercially developed, including PAXgene Blood ccfDNA tubes (Qiagen/BD), Cell-free DNA collection tubes (Roche), cf-DNA Preservation tubes (Norgen), Blood STASIS 21-ccfDNA tubes (MagBio Genomics) and LBgard Blood tubes (Biomatrica). Some have an advantage over the BCT tubes that we used in that they are not glass, so less prone to breakage. Further larger studies should be performed to fully assess these tubes to ensure they are fit-for-purpose. Selection of appropriate tube type for ctDNA clinical studies comes down to a number of factors, including cost. Cell-preservation tubes offer significant advantages over EDTA tubes, that need to be processed promptly after venepuncture by an on-site research nurse, but are currently relatively expensive, limiting their use in some studies.

4.2: Gale‡ et al, PLoS ONE, 2018 (7)

4.2.1: Aims

In order to enable patients to access to TAm-Seq for clinical diagnostic testing, Dr. Nitzan Rosenfeld, Dr James Brenton, Dr. Tim Forshew and I co-founded Inivata Ltd, a clinical cancer genomics company focussed on analysis of ctDNA to improve patient healthcare in oncology, funded by £4M Seed and £31.5M Series A venture capital funding. As Head of Molecular Diagnostics, I led the technology transfer of TAm-Seq™, and the development and analytical validation of enhanced TAm-Seq™ (eTAm-Seq™) for detection of clinically-relevant mutations in patients from advanced NSCLC patients.

Gale et al. describes the development and analytical validation of eTAm-Seq™ technology **(7)**. The InVision™ liquid biopsy tumour profiling panel was developed to profile key driver point mutations and entire coding regions of 35 genes spanning 10.61kb, including SNVs, indels and SCNAs (Figure 4.3). The assay and bioinformatic workflows were fully re-designed and optimised using a two-step multiplex PCR amplification protocol to enable high-throughput library preparation. To assess the sensitivity, specificity and LoD90 of eTAm-Seq™, validation studies were performed in two independent laboratories – Laboratory 1 to ISO15189:2012 standards, and Laboratory 2 to CLIA (Clinical Laboratory Improvement Amendments) regulatory standards.

eTAm-Seq™ was validated by 6 operators on different days, with one operator performing validations in both laboratories. Libraries were prepared from sheared Horizon Tru-Q6 and Tru-Q7 DNA reference standards, using different DNA input amounts (Table 4.1), or from plasma DNA from presumed healthy volunteers. Tru-Q6 is a mixture of 20 isogenic genetically-engineered cell lines, carrying mutations at known allele fractions (predominantly ~2%-2.5% AF), with 21 mutations covered by the InVision™ liquid biopsy tumour profiling panel. Tru-Q7 is a mix of 40 cell-lines predominantly at ~1%-1.3% AF, with 38 mutations targeted by the panel. To test the LoD90, a dilution series was prepared using Tru-Q0 wild-type DNA as diluent..

4.2.2: Results

4.2.3: Analytical Performance of eTAm-Seq™: SNVs

Analytical validation studies demonstrated that eTAm-Seq™ has high sensitivity. In Laboratory 1, overall sensitivity was 100% (90% CI: 99.01%-100%) using low input samples (2000 AC/sample) at 2%-2.5% AF; 99.17% (90% CI: 97.40%-99.85%) using median input samples (8000 AC) at 1%-1.3% AF; and 95.45% (90% CI: 93.09%-97.18%) using high input samples (16,000 AC) at 0.25%-0.33% AF. To further assess the LoD90, a dilution series was

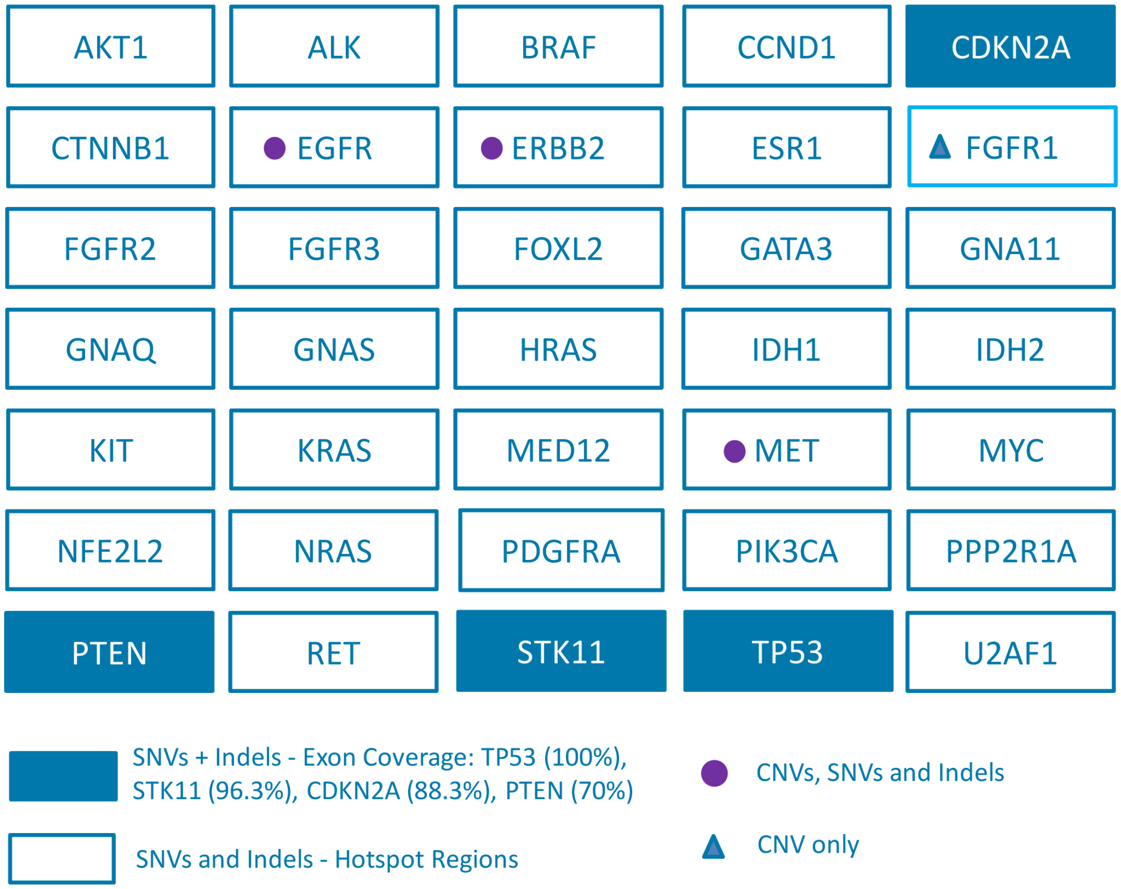

AKT1
ALK
BRAF
CCND1
CDKN2A
CTNNB1
● EGFR
● ERBB2
ESR1
▲ FGFR1
FGFR2
FGFR3
FOXL2
GATA3
GNA11
GNAQ
GNAS
HRAS
IDH1
IDH2
KIT
KRAS
MED12
● MET
MYC
NFE2L2
NRAS
PDGFRA
PIK3CA
PPP2R1A
PTEN
RET
STK11
TP53
U2AF1
SNVs + Indels - Exon Coverage: TP53 (100%), STK11 (96.3%), CDKN2A (88.3%), PTEN (70%)
● CNVs, SNVs and Indels
SNVs and Indels - Hotspot Regions
▲ CNV only

Figure 4.3: InVision™ liquid biopsy tumour profiling panel, *including SNVs and comprehensive or full coverage of coding regions. %: Tiling coverage per gene. SCNAs: labelled as CNVs (copy number variants). Figure and Figure Legend adapted from* ***Gale et al.*** **(7)**

Input DNA (AC)	AF (%)*	Number of repeats/operator	Number of operators/laboratory
16,000	0.25%-0.33%	4	3
8,000	1%-1.3%	3	3
2,000	2%-2.5%	7	3
8,000	0.5%-0.65%	3	3
8,000	0.25%-0.33%	3	3
8,000	0.13%-0.16%	3	3
8,000	0.06%-0.08%	3	3

Table 4.1: eTAm-Seq™ analytical validation experiments, *detailing range of DNA input amounts (amplifiable copies, AC), range of AF's tested, and number of operators/laboratory and replicates performed. *AF shows indicative ranges for Tru-Q reference matrial. Table and Table Legend adapted from* ***Gale et al. (7)***

performed using median input DNA. Results were comparable across both laboratories (Table 4.2). Overall, across the two laboratories, 98.65% mutations were detected at 0.5%-0.65% AF, and 89.73% were detected at 0.25%-0.33% AF. Using optimal input DNA (16,000 copies), sensitivity was 94% at 0.25%-0.33% AF. Mutant alleles were detected down to 0.02% AF, with >30% sensitivity for detection at 0.06% AF.

Inter- and intra-operator variablilty was assessed by performing replicate eTAm-Seq™ experiments on different days in two laboratories, using different operators and NGS sequencing runs. Results show consistent results across replicates using different DNA input amounts at different AFs (Figure 4.4). To assess the quantitative performance of eTAm-Seq™, data was compared with dPCR data analysing Horizon Tru-Q reference DNA, supplied by the manufacturer. Results shown in Figure 4.5 demonstrate the quantitative accuracy of eTAm-Seq™ and its ability to detect low frequency alleles.

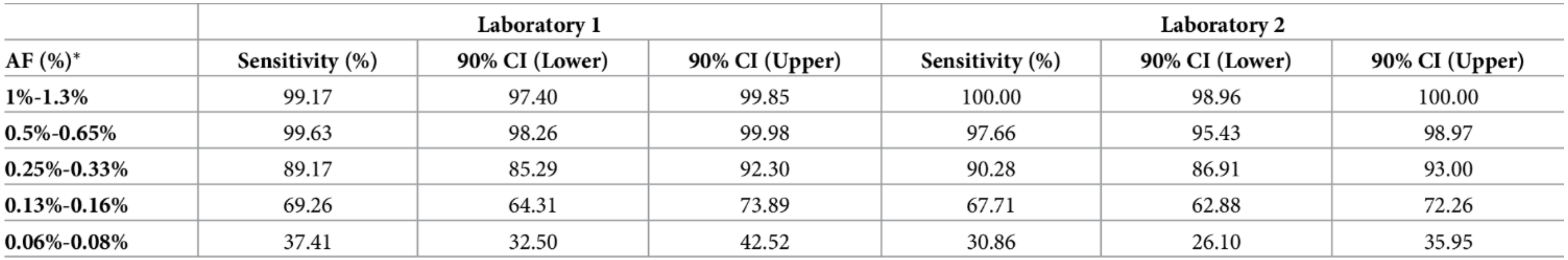
4.2.4: Analytical Performance of eTAm-Seq™: Specificity

To assess specificity, plasma from 79 presumed healthy donors were analysed by eTAm-Seq™. Five mutations were identified at ≤0.5% AF. 3/4 were validated as true positives in repeat experiments by re-analysis of samples with sufficient material. This resulted in two potential false positives - a *GATA* T419T mutation with insufficient material to repeat, and a *TP53* L308L mutation, detected at 0.19% AF, that was not verified in repeat experiments. The per-base specificity was 99.9997% (95% CI: 99.9989%-99.99996%).

4.3: Plagnol et al, PLoS ONE, 2018 (8)

4.3.1: Aims

Approximately 5% of NSCLC patients are known to have gene fusions with *ALK* and *EML4* genes, and ~1% of patients have *ROS1* gene rearrangements. ALK and ROS1 inhibitors, such as crizotinib, alectinib and ceretinib, have been shown to be effective in *ALK* and *ROS1*-mutated NSCLC (93). To provide comprehensive genomic profiling of cfDNA, eTAm-Seq™ was further developed to enable detection of *EML4-ALK*, *SLC34A2-ROS1, CD74-ROS1, EZR-ROS1 and SDC4-ROS1* gene fusions. **Plagnol et al.** describes the analytical validation of the InVisionFirst™-Lung assay, encompassing 36 genes commonly mutated in NSCLC and other cancers, including SNVs, indels, *ALK* and *ROS1* fusions, and SCNAs in *ERBB2*, *FGFR1*, *MET* and *EGFR* (8). An overview of the InVisionFirst™ workflow is detailed in Figure 4.6.



	Laboratory 1			Laboratory 2		
AF (%)*	**Sensitivity (%)**	**90% CI (Lower)**	**90% CI (Upper)**	**Sensitivity (%)**	**90% CI (Lower)**	**90% CI (Upper)**
1%-1.3%	99.17	97.40	99.85	100.00	98.96	100.00
0.5%-0.65%	99.63	98.26	99.98	97.66	95.43	98.97
0.25%-0.33%	89.17	85.29	92.30	90.28	86.91	93.00
0.13%-0.16%	69.26	64.31	73.89	67.71	62.88	72.26
0.06%-0.08%	37.41	32.50	42.52	30.86	26.10	35.95

Table 4.2: Sensitivity of eTAm-Seq™ using median input DNA at different AFs tested in two independent laboratories.

AF shows indicative ranges for Tru-Q reference matrial. CI: confidence interval. Table and Table Legend adapted from* ***Gale et al.***(7)**

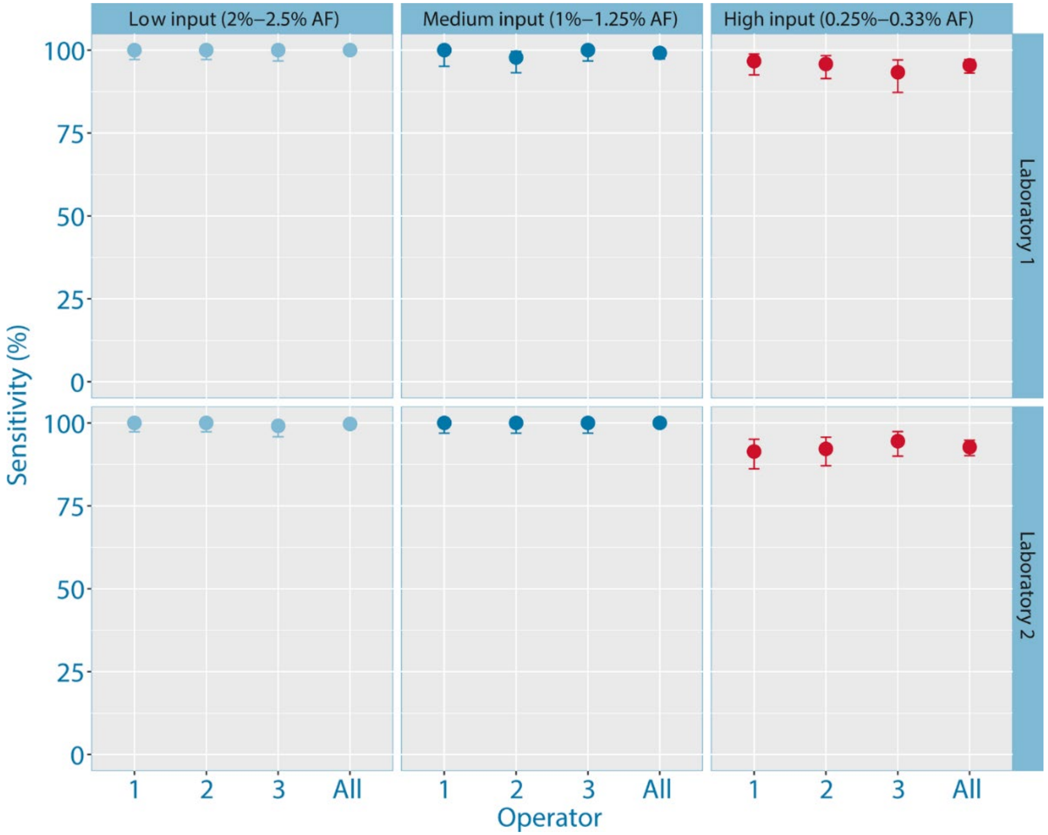


Figure 4.4: Sensitivity and inter-/intra-operator variability of eTAm-Seq™ using low, medium and high input DNA.

Laboratory 1: upper; Laboratory 2: lower. Error bars: 90% Confidence Interval; Figure and Figure Legend adapted from ***Gale et al.*** **(7)**

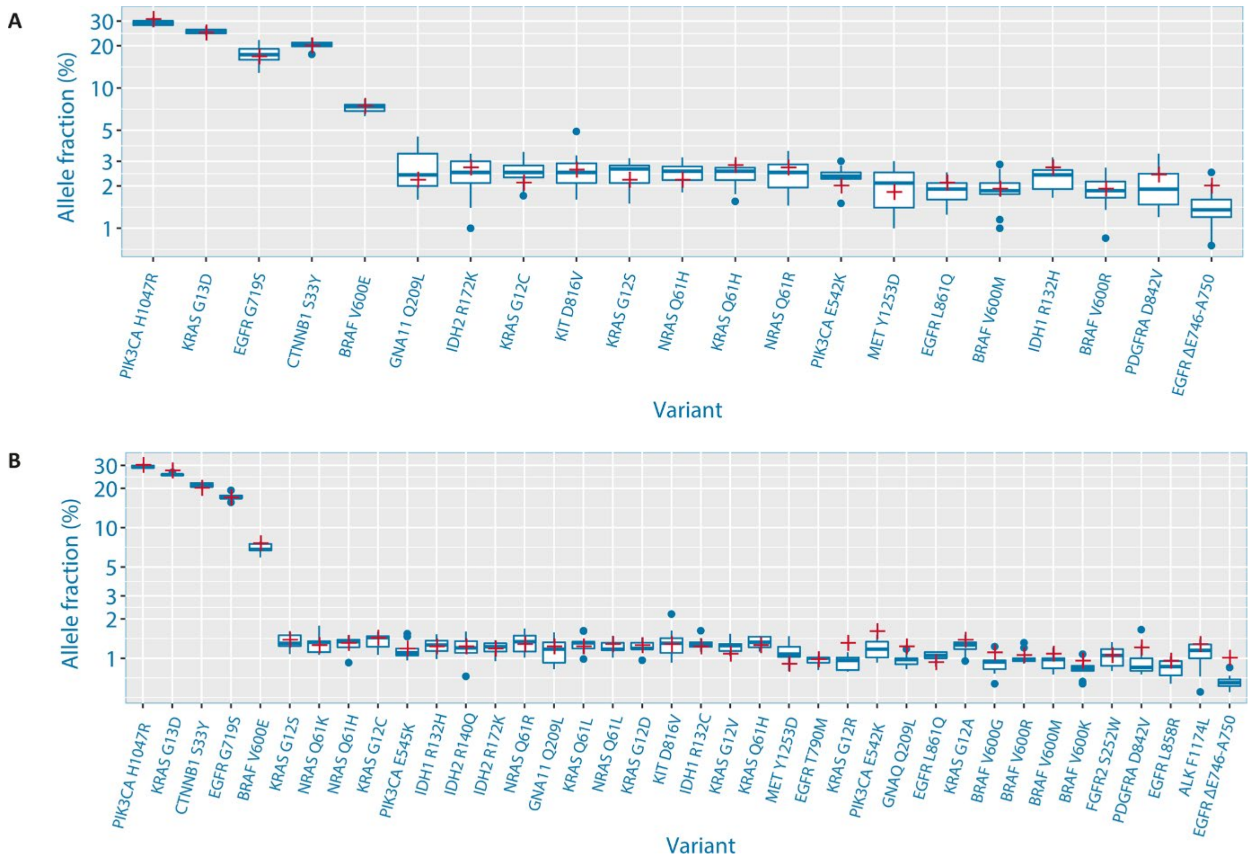


Figure 4.5: Allele Fractions determined by analysis with eTAm-Seq™ (blue boxplot) and dPCR (red cross) *for analysis of mutations present in both the InVision™ liquid biopsy tumour profiling panel and (A) Tru-Q6 (B) Tru-Q7. Figure and Figure Legend adapted from* ***Gale et al.*** **(7)**

4.3.2: Results

4.3.3: Analytical Performance of InVisionFirst™-Lung: SNVs & indels

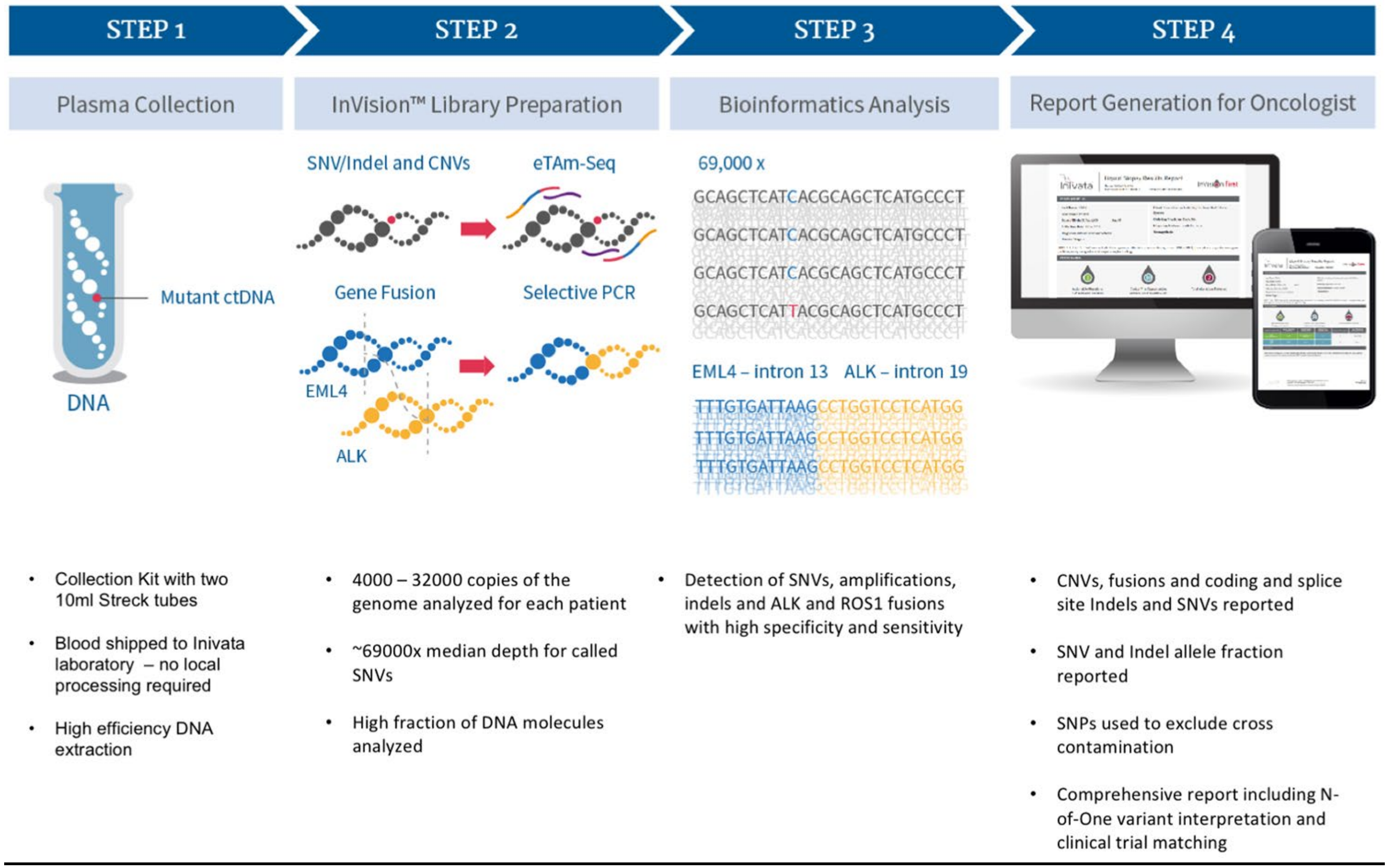
A dilution series was performed to assess the LoD90. Tru-Q7 contains 36 mutations covered by the InVisionFirst™ gene panel, including 32 mutations at ~1%-1.3% AF (Range 1%-30% AF). Replicate samples were analysed in Laboratory 2 by three operators, each using two different reagent lots (Figure 4.7). The assay showed high repeatability and reproducibility, with 100% SNVs detected at 0.5% AF in all replicates. The LoD90 was 0.25% AF, with 99.48% SNVs detected at 0.25%-0.33% AF. There was no noticeable difference in inter-and intra-operator variability or lots at 0.25%-0.33% AF. A SeraCare custom reference standard was manufactured to assess 18 indels (ranging -24bp-+12bp), using a dilution series down to 0.1% AF. 99.7% indels were detected at 0.5% AF, with 92.46% detected at 0.25% AF.

4.3.4: Analytical Performance of InVisionFirst™-Lung: SVs

Due to the scarcity of obtaining specimens from patients with specific *ALK* and *ROS1* gene fusions, analytical validation of rearrangements was performed using either fragmented Horizon custom cell-line mix DNA, carrying known mutations in *EML4-ALK* or *SLC34A2-ROS1*; synthetic dsDNA fragments carrying published and random fusion breakpoints; and analysis of Horizon cell-line fusion reference DNA spiked into donor blood. In analysis of a total of 87 samples in a dilution series, 36 fusions present in the cell-line mix were detected down to 0.13% AF, with 90% detected at 0.06% AF. Analysis of synthetic fusions detected 90% rearrangements at 0.5% AF, and 2.5% AF cell-line fusion DNA was detected in all 19 donors, collected into both BCT or EDTA tubes. Overall, 104 variants were analysed in 54 unique samples, and fusions were detected in all but 3, resulting in a positive percentage agreement (PPA) of 97.1% at ≥0.5% AF.

4.3.5: Analytical Performance of InVisionFirst™-Lung: SCNAs

To validate *ERBB2, FGFR1, MET* and *EGFR* amplifications, synthetic dsDNA blocks carrying complementary sequences were spiked into sheared wild-type DNA, resulting in copy number amplification ratios (CNARs) of 1.25x, 1.5x and 2x. Each amplification was tested at each dilution level 22-24 times by the 3 operators. All 4 amplifications were detected at 2x CNAR, with a mean of 97.7% detected at 1.5x, and 75% [range 59% (*FGFR1*) to 90.91%


STEP 1
STEP 2
STEP 3
STEP 4
Plasma Collection
InVision™ Library Preparation
Bioinformatics Analysis
Report Generation for Oncologist
Mutant ctDNA
DNA
SNV/Indel and CNVs
eTAm-Seq
Gene Fusion
Selective PCR
EML4
ALK
69,000 x
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATTACGCAGCTCATGCCCT
EML4 – intron 13
ALK – intron 19
TTTGTGATTAAGCCTGGTCCTCATGG
TTTGTGATTAAGCCTGGTCCTCATGG
TTTGTGATTAAGCCTGGTCCTCATGG
Invivata
InVisionFirst

- Collection Kit with two 10ml Streck tubes
- Blood shipped to Inivata laboratory – no local processing required
- High efficiency DNA extraction

- 4000 – 32000 copies of the genome analyzed for each patient
- ~69000x median depth for called SNVs
- High fraction of DNA molecules analyzed

- Detection of SNVs, amplifications, indels and ALK and ROS1 fusions with high specificity and sensitivity

- CNVs, fusions and coding and splice site Indels and SNVs reported
- SNV and Indel allele fraction reported
- SNPs used to exclude cross contamination
- Comprehensive report including N-of-One variant interpretation and clinical trial matching

Figure 4.6: Overview of the InVisionFirst™ workflow *for detection of SNVs, indels, ALK and ROS1 gene fusions and SCNAs (labelled as CNVs, copy number variants) from patient blood to oncologists report. Figure and Figure Legend adapted from* ***Plagnol et al.*****(8)**

(*EGFR*) at 1.25x CNAR. Further samples were prepared between 2x and 50x CNAR. 51/52 variants were detected in 49 unique samples at ≥1.5x CNAR, resulting in a PPA of 98.1%.

4.3.6: Analytical Performance of InVisionFirst™-Lung: Specificity

To assess specificity, blood from presumed healthy volunteers were collected into K_3EDTA or BCT tubes. In analysis of up to 109 samples, the per-base specificity was 99.9997%. Three coding or splice variants, but no SCNAs or gene fusions were identified. A *TP53* p.L369X mutation was verified as real by dPCR, but two (at 0.13% and 0.29% AF) were not re-confirmed and may be false positives, or true positives that could not be detected at this low frequency.

4.4: Discussion

In **Gale et al.**, analytical validation studies demonstrated 94% sensitivity for detection of SNVs at 0.25%-0.33% AF using optimal input DNA, with mutant alleles detected down to 0.02% AF. This was a 10-fold enhancement in sensitivity compared to original TAm-Seq **(3),** with the assay expanded from analysis of SNVs in 6 genes spanning <6kb, to an assay targeting 35 genes spanning 10.61kb, covering SNVs, indels and SCNAs. In **Plagnol et al.**, the assay was further developed to target SNVs and indels in 36 genes, including *NTRK1* and *NTRK3* for which newly developed targeted inhibitors have recently been developed (94). A major limitation of original TAm-Seq and eTAm-Seq™ was the inability to target gene rearrangements. The InVisionFirst™-Lung assay was further developed to enable detection of *ALK and ROS1* gene fusions. As a result, InVisionFirst™-Lung provides comprehensive genomic profiling (CGP) of all four mutation classes in ctDNA for patients with advanced NSCLC, with 99.48% sensitivity for detection of SNVs at 0.25%-0.33% AF, 92.46% sensitivity for indels at 0.25% VAF, a PPA of 97.1% for detection of *ALK* and *ROS1* fusions at ≥0.5% AF, a PPA of 98.1% for detection of SCNAs at ≥1.5x CNAR, and a per-base specificity of 99.9997%.

4.4.1: Challenges of analytical validation of an NGS diagnostic assay

The main challenge for both **Gale et al.** and **Plagnol et al.** was to design robust analytical validation studies appropriate for analysis of NGS data. As only a few NGS assays had entered clinical diagnostic use at the time of performing this study, only limited analytical validation studies had been published, meaning that there was no fixed blueprint for how

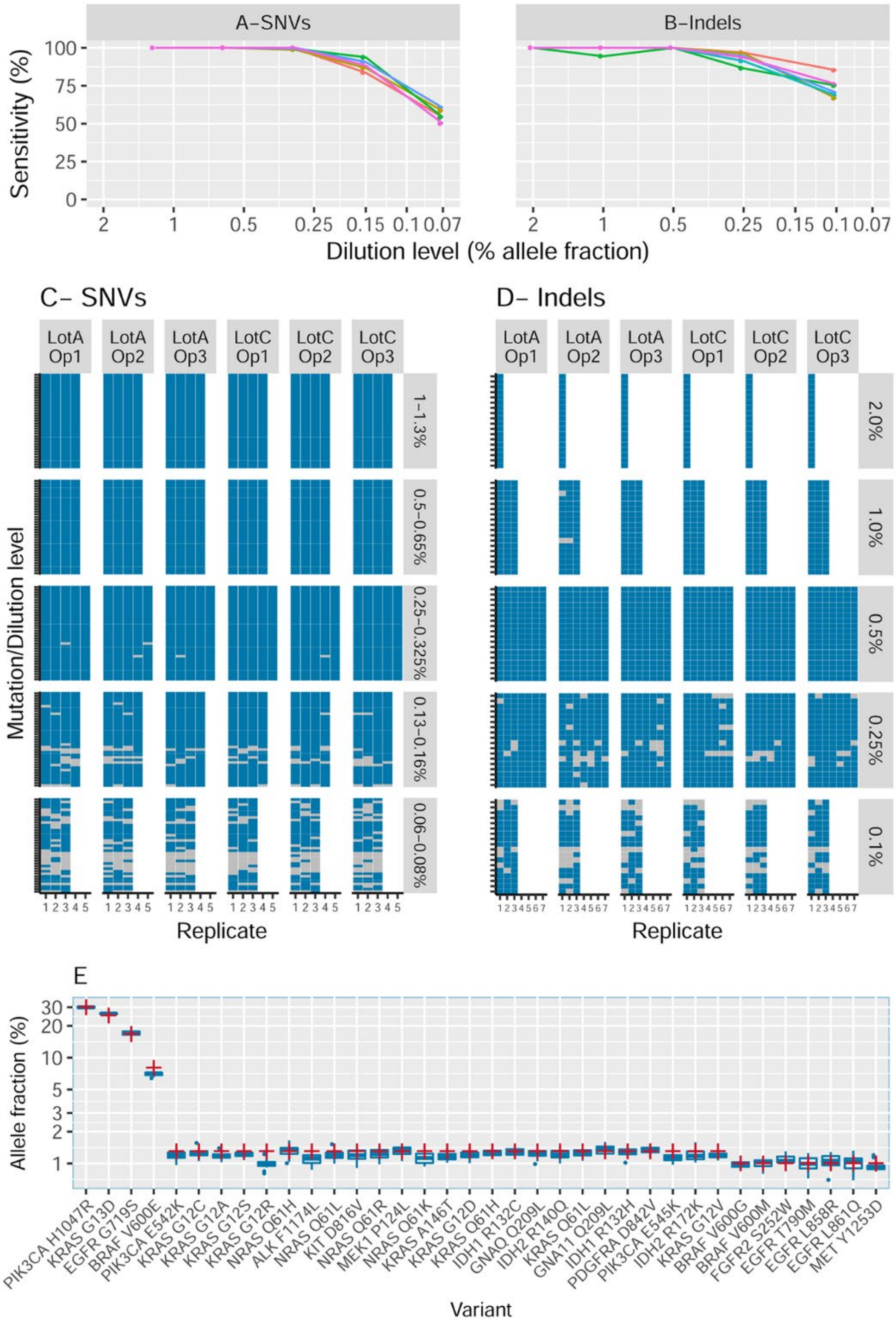


Figure 4.7: Sensitivity of InVisionFirst™ for detection of SNVs (A, C) and indels (B, D). **(A,*B*)**: *Each line: Different operator/reagent lot.* **(*C,D*):** *All calls for each dilution/variant (vertical) and repeater/operator/lot (horizontal). Blue rectangles: Mutations detected; Grey: Not detected. (E)***:** *Comparison of AFs generated using InVisionFirst™ (blue) and dPCR (red crosses, supplied by manufacturer). Op: Operator*; *Figures and Figure Legend adapted from* ***Plagnol et al*.** **(8)**

to conduct such studies. Mattocks et al. published a standardized framework for validation and verification of clinical molecular genetic tests (95). More recently, further guidelines have been published by the Association for Molecular Pathology (AMP) and College of American Pathologists (CAP), covering NGS assay validation of oncology panels and bioinformatics pipelines respectively (96,97). Since initiating this work, Guardant Health and Foundation Medicine have also published analytical validation of their ctDNA assays (80,82).

One of the considerable challenges in analytical validation of an NGS ctDNA assay is obtaining appropriate controls for validation of rare mutated sequences that appear in only a small proportion of the population. In such circumstances, it is difficult to obtain rare tissue or blood specimens to appropriately validate assays targeting these regions. For this reason, we used contrived samples, using a custom fragmented cell-line mix containing *ALK* and *ROS1* fusions, using synthetic dsDNA fragments designed to published and random fusion breakpoints, and Horizon cell-line fusion reference standard spiked into donor blood. One issue with using contrived material is that there may be unavoidable differences with genuine control material. For example, acoustic shearing of tumour or reference standard DNA may introduce artefacts due to 8-oxoguanine lesions, resulting in an increase in G>T/C>A transversion errors (98,99). Using an highly sensitive NGS assay, these may be identified as mutations, and would need to be verified as true or false positives. Where possible, repeat analysis should be performed using an alternative method to exclude technical errors. Enzymatic fragmentation could also be used rather than acoustic shearing to avoid DNA damage.

Due to issues with shearing, the most appropriate control is plasma cfDNA as it is naturally fragmented. However, when using control plasma from presumed healthy individuals, such as for specificity studies, it is important to be aware that some ‘healthy’ individuals who have not been diagnosed with cancer may also carry low frequency sequence changes. These alterations may be benign, pre-malignant mutations that accumulate prior to cancer or during ageing, originate from an undetected tumour, or result from changes occurring during clonal hematopoiesis (100,101). As assays become more sensitive, it is becoming increasingly important to rule out clonal hematopoiesis of indeterminate potential (CHIP), a common age-related process involving clonal expansion of a sub-population of blood cells from hematopoietic stem cells or blood cell progenitors, as a confounder by analysis of buffy coat germline DNA .

A further challenge of reporting NGS ctDNA results was highlighted by Torga et al., who compared results from analysing the same plasma specimens from 40 prostate cancer patients using two different commercial assays, Guardant360® and PlasmaSELECT™. (102). Of concern, data showed poor agreement in mutations reported by both assays, with only 7.5% (3/40) patients with one or more alterations showing complete congruence, and 15% (6/40) showing partial congruence. Unsurprisingly, this caused considerable controversy, with both companies disputing findings, raising concerns on study design and analysis, sample collection, and the differential effects of germline variants reported by each company (103,104). More reassuringly, Thress et al. compared the cobas® EGFR mutation test and BEAMing, and showed >90% concordance. Further comparative studies must be expected to ensure patient safety.

4.4.2: Clinical impact of eTAm-Seq™ and InVisionFirst™-Lung assay

Analytical validation of the InVisionFirst™-Lung assay demonstrated that the test has high sensitivity. A prospective multi-centre clinical validation study has now been performed analysing plasma from 264 NSCLC patients (178 with tissue available) (105). ctDNA was detected in 77.3% of patients, and 53.8% of patients had an actionable alteration, or an alteration that is mutually exclusive for actionable changes (eg. *KRAS* or *STK11*). Overall concordance with matched tissue profiling was 97.8%, with 83.0% PPV, 98.5% NPV, 70.6% sensitivity and 99.2% specificity. Analysis of the eight most clinically relevant genes (*EGFR*, *ALK*, *ROS1*, *ERBB2*, *MET*, *BRAF*, *KRAS*, *STK11*) in patients where ctDNA was detected resulted in a sensitivity of 87.3%, specificity of 99.3%, 93.7% PPV and 98.4% NPV. 26% of patients had more actionable alterations identified by InVisionFirst™-Lung than by tissue testing, and ctDNA genomic profiles were consistent across patients with or without tissue. The assay is now commercially available in the United States. Unlike in the UK, where healthcare is free to its citizens for validated interventions [following recommendations by the National Institute for Health and Care Excellence (NICE)], the US largely operates a healthcare system that needs to be paid for by medical insurance. As such, diagnostic tests need to be approved by health insurance programs, such as Medicare, in order to get reimbursement. Palmetto GBA, a Medicare Administrative Contractor, recently provided a final Local Coverage Decision (LCD) for the InVisionFirst™-Lung assay to be used as a plasma-based somatic CGP test for patients with Stage IIIB/IV NSCLC (106,107). The recommended intended use for the test is:

At diagnosis: when test results for *EGFR* SNVs/indels, *ALK*/*ROS1* rearrangments, and *BRAF* SNVs are not available and tissue-based testing is not possible (medically contraindicated or quantity not sufficient [QNS])

OR

At progression: for patients progressing on or after chemotherapy or immunotherapy who have never been tested for *EGFR* SNVs/indels, *ALK*/*ROS1* rearrangments, and *BRAF* SNVs, and for whom tissue-based testing in not possible (QNS from original biopsy).

OR

For patients progressing on EGFR TKIs. If no genetic alteration is detected or if ctDNA is insufficient/non-detected, tissue-based genotyping should be used. (108)

This is a major milestone in getting the assay reimbursed for all fee-for-service Medicare patients with advanced Stage III/IV NSCLC in the US who meet the appropriate clinical criteria. It is only the second NGS ctDNA assay, together with Guardant360®, to have achieved this status.

In the next Chapter, I will discuss potential applications of ctDNA in the clinic, including work I have performed to assess the clinical utility of eTAm-Seq™, and use of TAm-Seq™ and dPCR for monitoring treatment response **(9–11)**. I will also discuss the development of size selection to improve the sensitivity of detection of ctDNA **(12–14)**. Improved sensitivity is required for earlier diagnosis of cancer and detection of minimal residual disease.

Chapter 5: Clinical applications of ctDNA

5.1: Molecular stratification

Following the demonstration by us and others that ctDNA has potential as a non-invasive liquid biopsy, work has progressed at a rapid pace to further evaluate its utility and implement assays in the clinic. The most immediate impact has been in the molecular stratification of NSCLC patients to treatment, since activating *EGFR* mutations predict sensitivity to tyrosine kinase inhibitors, and osimertinib is effective in patients with the *EGFR* p.T790M resistance-conferring mutation. In **Remon et al.**, we used eTAm-Seq™ for plasma-based testing to guide treatment with osimertinib in advanced NSCLC patients **(9)**.

5.1.1: Remon et al, Annals of Oncology, 2017 (9)

5.1.1.1: Aims

The aim of this study was to assess the clinical utility of eTAm-Seq™. Analysis was performed in a prospective cohort of 48 advanced NSCLC patients with known activating *EGFR* mutations, without an available tissue biopsy, who were progressing on current treatments. p.T790M status was assessed in plasma cfDNA using eTAm-Seq™ to guide treatment with osimertinib. The aim was to determine treatment response from imaging data using RECIST v1.1 criteria (63), assess progression-free survival, and evaluate plasma-based testing using eTAm-Seq™. In this study, I led the clinical lab team generating data for this study, using the assay developed by **Gale et al. (7)**

5.1.1.2: Results

EGFR p.T790M was detected using eTAm-Seq™ in 24/48 patients (50%). 18 patients received osimertinib on progression, and 16 were evaluated for response. 10/16 (62.5%) patients had a partial response and 6/16 (37.5%) had stable disease. Responses were re-confirmed in 90% patients, with one patient having progressed (Figure 5.1). Median PFS was not achieved (NA) after a median follow-up of 8-months (95% CI: 4-NA), with 6-months and 12-months PFS of 66.7% and 52% respectively. Overall survival was not achieved, with 4 patients having died at the time of study cut-off, with one year OS at 78% (95% CI: 59-97).

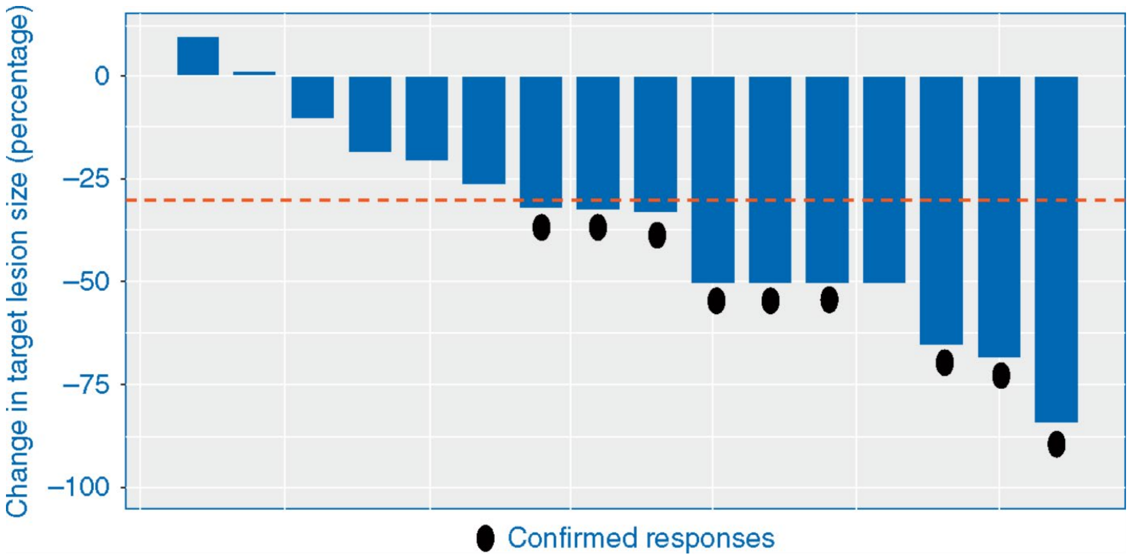


Figure 5.1: Waterfall plot of p.T790M-positive NSCLC patients treated with osimertinib, showing best percentage change in lesion size. *Red dotted line: 30% decrease. Figure and Figure Legend adapted from* ***Remon et al.*** **(9)**

5.1.1.3: Discussion

The most immediate clinical impact of ctDNA has been in the stratification of NSCLC patients to treatment. In patients treated with erlotinib or gefitinib, the 5-year survival of *EGFR*-mutant metastatic lung adenocarcinoma patients was shown to be 14.6%, compared to 4.5% for unselected patients with distant-stage NSCLC (109,110). The cobas *EGFR* Mutation Test v2 is an FDA-approved real-time PCR test for the qualitative detection of *EGFR* mutations including exon 19 deletions, p.L858R and p.T790M. Importantly, it was the first companion diagnostic to be approved for use with either tissue *or plasma* to guide treatment with erlotinib and osimertinib (111). However, the test has an input of just 2mL plasma and an LoD of 25-100 copies/mL plasma (dependent on mutation). (112). Recent advances in the development of dPCR and NGS ctDNA assays means that it is now possible to more sensitively detect low frequency *EGFR* mutations.

Remon et al. was the first published prospective study in a real-world setting to assess p.T790M status and efficacy of osimertinib based on plasma rather than tissue profiling. p.T790M was detected in 50% of *EGFR*-mutant patients using the eTAm-Seq™ assay, which is consistent with tumour profiling in other studies (113,114). Results demonstrated good response rates, showing 62.5% patients with a partial response, also comparable with tissue-based testing (114,115). This study provides data supporting the clinical utility of eTAm-Seq™ for detection of clinically-relevant mutations in plasma of advanced NSCLC patients.

Retrospective analysis of plasma samples from the Phase I AURA trial using BEAMing demonstrated detection of the *EGFR* resistance mutation in 31% of p.T790M-negative tumours (assessed using the cobas assay), although these may potentially be false positives or result from tissue sampling biases (116). Results from the AURA Phase II extension trial demonstrate that treatment of p.T790M-positive NSCLC patients with osimertinib showed high response rates, encouraging PFS and durable responses (117). Interestingly, our study demonstrates that response to osimertinib could be seen in patients with mutant allele fractions <0.5% AF. Currently little is known about the optimal time to treat based on the detection of low frequency mutant alleles. To fully evaluate the utility of sensitive plasma testing, clinical trials need to be performed to determine whether it is best to treat a patient immediately following detection of low levels of the p.T790M mutation, or whether patients would have better overall survival if treated later. Hypothetically, earlier treatment may lead to the early emergence of new resistance mutations, which will require new therapies targeting these mutations. Thress et al. have already shown that some patients treated with osimertinib acquire an *EGFR* p.C797S resistance mutation (118).

The **Remon et al.** study was limited in that it analysed a relatively low number of patients, many of whom had undergone multiple lines of heterogenous treatment. To fully assess performance, studies should be performed on a larger cohort of patients. This can now happen given that plasma-based testing for patients eligible for osimertinib treatment has been approved.

5.2: Monitoring treatment response

The ability to monitor treatment response is important to avoid over-treatment with an ineffective therapy, prevent unnecessary side effects, indicate when new treatments should be initiated and provide an opportunity to assess novel drug compounds. We tested whether ctDNA can accurately monitor tumour dynamics in breast cancer by comparison with other circulating biomarkers and imaging data [**Dawson et al., (10)**], and whether ctDNA can be used to study different resistance mechanisms in NSCLC [**Tsui et al.,(11)**]. In these studies, I developed optimised protocols for clinical collections and analysis of ctDNA, trained the lead co-authors in these techniques and provided scientific guidance.

5.2.1: Dawson et al., New England Journal of Medicine, 2013 (10)

5.2.1.1: Aims

The aim of this study was to assess whether ctDNA could be used to monitor tumour burden and response to treatment in patients with metastatic breast cancer, by comparison with circulating tumour cells (CTCs), cancer antigen 15-3 (CA15-3) and radiographic imaging. Serum CA15-3 is a clinically used biomarker, with sensitivity of only 60%-70% (119,120). CTCs were evaluated using the CellSearch System, with sensitivity of ~65% for detection of ≥1 cell/7.5mL blood, with levels deemed to be elevated and associated with poor prognosis if ≥5 cells/7.5mL blood (119,121). Mutations were first identified in FFPE tumour DNA using either TAm-Seq, to identify *PIK3CA* and *TP53* mutations, or WGS to identify SNVs or SVs.

5.2.1.2: Results

PIK3CA or *TP53* mutations were identified in tumour DNA from 25/52 patients. WGS identified SVs in a further 8 cases, including 5 where no mutations had previously been identified. ctDNA levels were subsequently quantified in 141 serial plasma samples from 30 patients using either dPCR or TAm-Seq. ctDNA was detected in 97% (29/30) cases, and in 115/141 (82%) plasma specimens. Figure 5.2 shows monitoring of ctDNA levels using either SNVs or SVs. Three deletions in an amplified locus show a similar dynamic pattern to a *PIK3CA* mutation but at higher levels, as expected using a quantitative assay (Figure 5.2A). Similar dynamics were generally observed in cases with multiple mutations (Figure 5.2B), although in some cases there was clonal heterogeneity, with different dominant clones observed post-treatment (Figure 5.2C/D).

In cases where samples were available for comparative analysis, CA15-3 was elevated in 21/27 women (78%), and in 71/114 samples (62%), compared to detection of ctDNA in 26/27 patients (96%) and 94/114 samples (82%). Sensitivity of CA15-3 was 59%, and 85% for ctDNA. CTCs were detected at ≥1 cell/7.5mL blood in at least one timepoint in 26/30 patients (87%), and elevated at ≥5 cells/7.5mL blood in 18/30 women (60%). CTCs were not detected in 50/126 samples (40%). By comparison, ctDNA was detected in 29/30 patients

(97%), and in 106/126 samples (84%). Sensitivity of CTCs was 67%, and 90% for ctDNA. The median number of amplifiable copies of ctDNA was 133 times greater than the number of CTCs, with a greater dynamic range.

In 20 patients with available data, the performance of different circulating biomarkers was compared to imaging data by CT. Examples are shown in Figure 5.3. ctDNA was detected in 19/20 patients (95%), and ctDNA dynamics generally correlated with treatment responses seen on imaging. CTC data agreed with imaging data in 10/20 patients with elevated CTCs (≥5 CTCs/7.5mL blood), but not in the remaining 50% cases with lower levels. CA15-3 showed similar profiles in patients with >50 U/mL, but with a smaller dynamic range, and these fluctuations were not observed in patients with lower levels. In 10/19 patients (53%), ctDNA gave the earliest measure of treatment response, with levels elevated on average five months before detection of progressive disease by imaging. In addition, higher levels of ctDNA were associated with poorer overall survival ($p < 0.001$, Cox regression model). CTCs were also shown to be prognostic ($p = 0.03$), but CA15-3 was not.

5.2.2: Tsui et al., EMBO Molecular Medicine, 2018 (11)

5.2.2.1: Aims

The aim of this study was to monitor tumour dynamics and investigate co-existent resistance mechanisms in 50 NSCLC patients with *EGFR*-positive tumour DNA undergoing treatment with gefitinib and hydroxchloroquine **(11)**. 392 plasma specimens were analysed using dPCR and TAm-Seq. sWGS was used to analyse three patients undergoing histological transformation to small-cell lung cancer (SCLC).

5.2.2.2: Results

Analysis showed that 41/43 (95%) patients had identical *EGFR* mutations in plasma and the tumour biopsy. Additional mutations were identified in plasma, including in *TP53*, *PIK3CA* and *PTEN*. Analysis of 19 TKI treatment-naïve cases showed that patients with low pre-treatment levels of *EGFR*-activating mutations tended to have better PFS and OS, although this did not reach significance (Cox; $p = 0.06$). Levels correlated with tumour burden, with patients with low mutant AF having lower tumour burden (median: 17 mm), compared to those with intermediate (median: 42 mm) and high levels (median: 80 mm). In

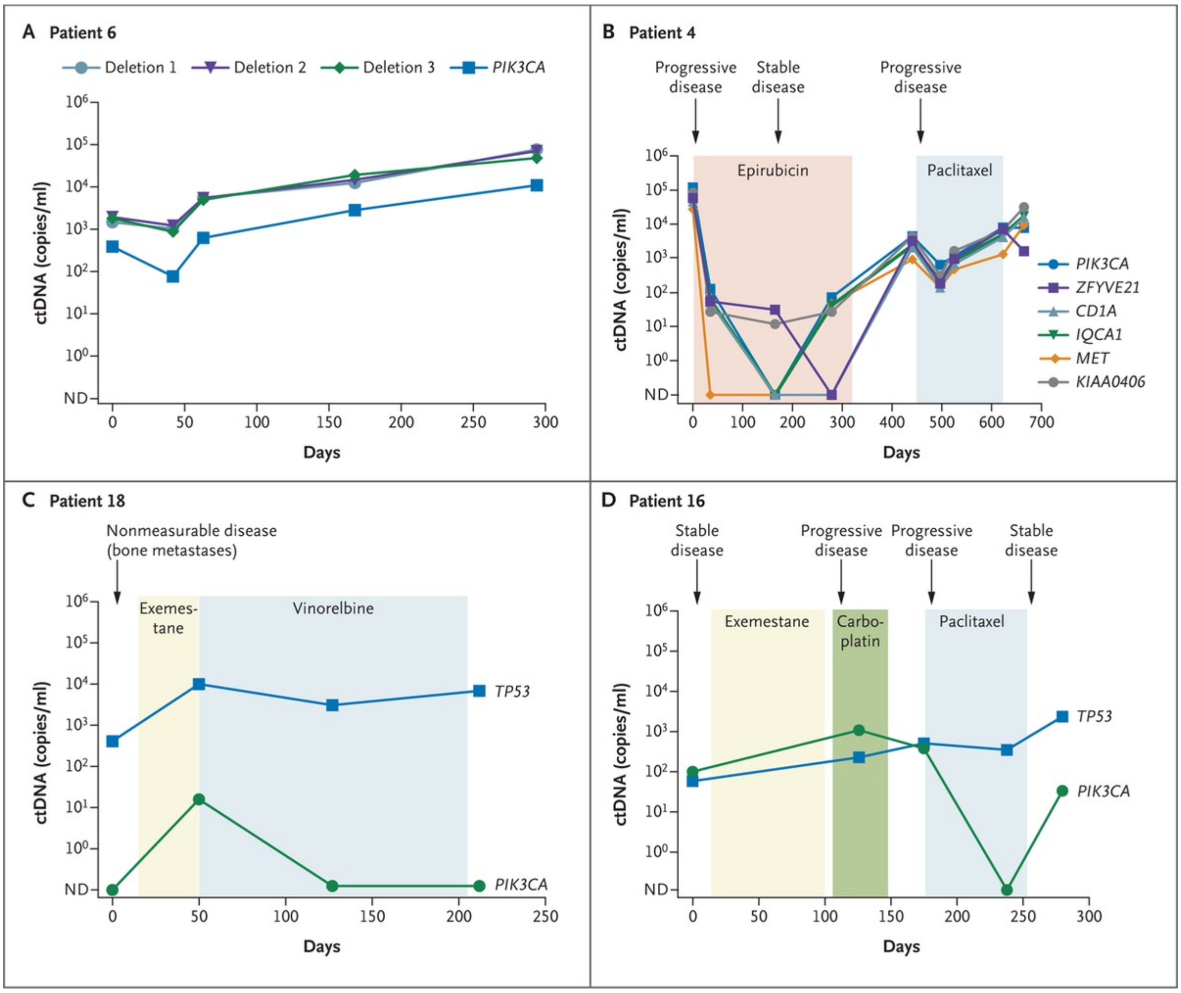


Figure 5.2: Monitoring multiple SNVs and SVs in cfDNA from breast cancer patients. *(A) Monitoring 3 SVs and 1 PIK3CA mutation in one patient by dPCR; (B) Monitoring 6 SNVs in one patient by dPCR; C) Monitoring PIK3CA and TP53 SNVs by dPCR, indicating TP53 mutation was dominant in the circulation; (D) Differences in PIK3CA and TP53 levels during paclitaxel treatment using TAm-Seq; Shaded areas: Periods of treatment; RECIST v1.1 imaging responses indicated; ND: Not detected; Figure and Figure Legend adapted from* ***Dawson et al.*** **(10)**

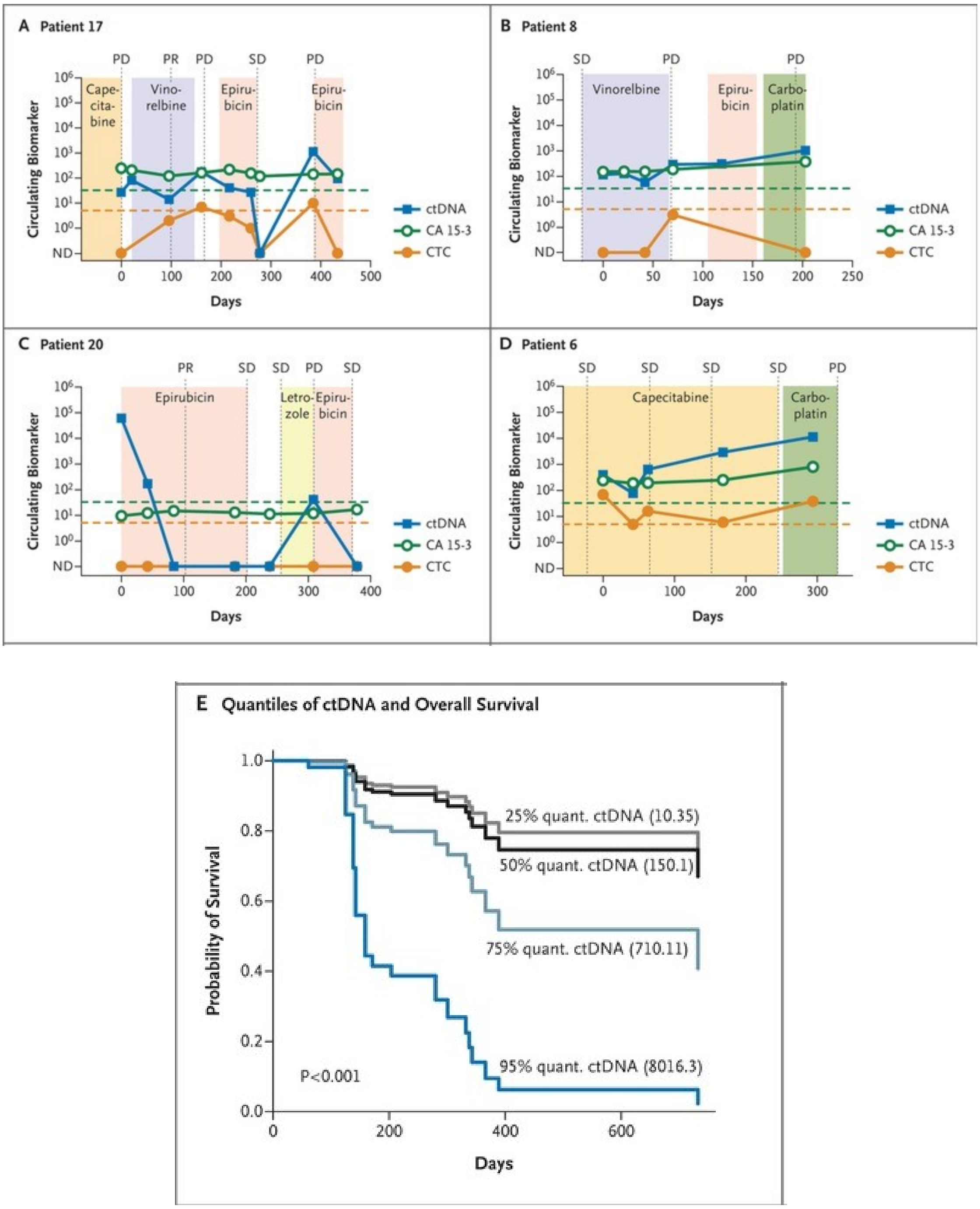


Figure 5.3 (A-D): Comparison of ctDNA, CTCs, CA15-3 and imaging data to monitor tumour dynamics in 4 patients. *ctDNA: copies/mL plasma; CTCs/7.5mL whole blood (orange line: 5 CTCs/7.5mL); CA15-3: U/mL (green line: 32.4 U/mL threshold); Shaded areas: Periods of treatment; RECIST v1.1 imaging responses: PD: Progressive Disease; SD: Stable disease; ND: Not detected;* **(E) Quantiles of ctDNA copies/mL compared to overall survival.** *Increasing levels associated with poor overall survival (p < 0.001, Cox regression model, ctDNA - continuous time-dependent variable). Figure and Figure Legend adapted from* ***Dawson et al.*** **(10)**

addition, patients with both *EGFR* and *TP53* mutations in pre-treatment plasma had worse prognosis.

Three distinct mutational profiles were observed (Figure 5.4). 62% of patients had a p.T790M mutation in plasma following treatment, with the resistance mutation detected a median of 6.8 months prior to progression. 22% of patients had an activating *EGFR* mutation before and after progression, but no p.T790M mutation. TKI-naïve patients who did not have a p.T790M mutation in plasma had worse PFS (log-rank test, *p* = 0.008). Finally, in a third sub-cohort (16%), no *EGFR* activating or resistance mutations were detected in plasma on progression, suggesting resistance in these patients may develop through alternative pathways.

Approximately 2%-3% NSCLC cases develop TKI-resistance through transformation to a histological classification of SCLC. In this study, this was observed in three patients by analysis of tissue re-biopsies at progression (122). sWGS analysis of plasma from these patients showed the presence of SCNAs post-progression, including in genes known to be associated with SCLC, and these correlated with tumour burden and radiological response (Figure 5.5).

Figure 5.4 ***(on next page)*** **: ctDNA dynamics in 45 NSCLC patients reveals distinct patterns of resistance mechanisms***.* *(A) EGFR-activating mutations before and after disease progression, and newly arising p.T790M mutation (B) EGFR-activating mutations before and after disease progression, but no p.T790M mutation; (C) No EGFR-activating or p.T790M mutation. Coloured lines: specific mutations; PD: Progressive Disease; PD*: PD with new lesions; PD**: PD defined by presentation of symptoms on brain or bone scan. Figure and Figure Legend adapted from* ***Tsui et al.*** **(11)**

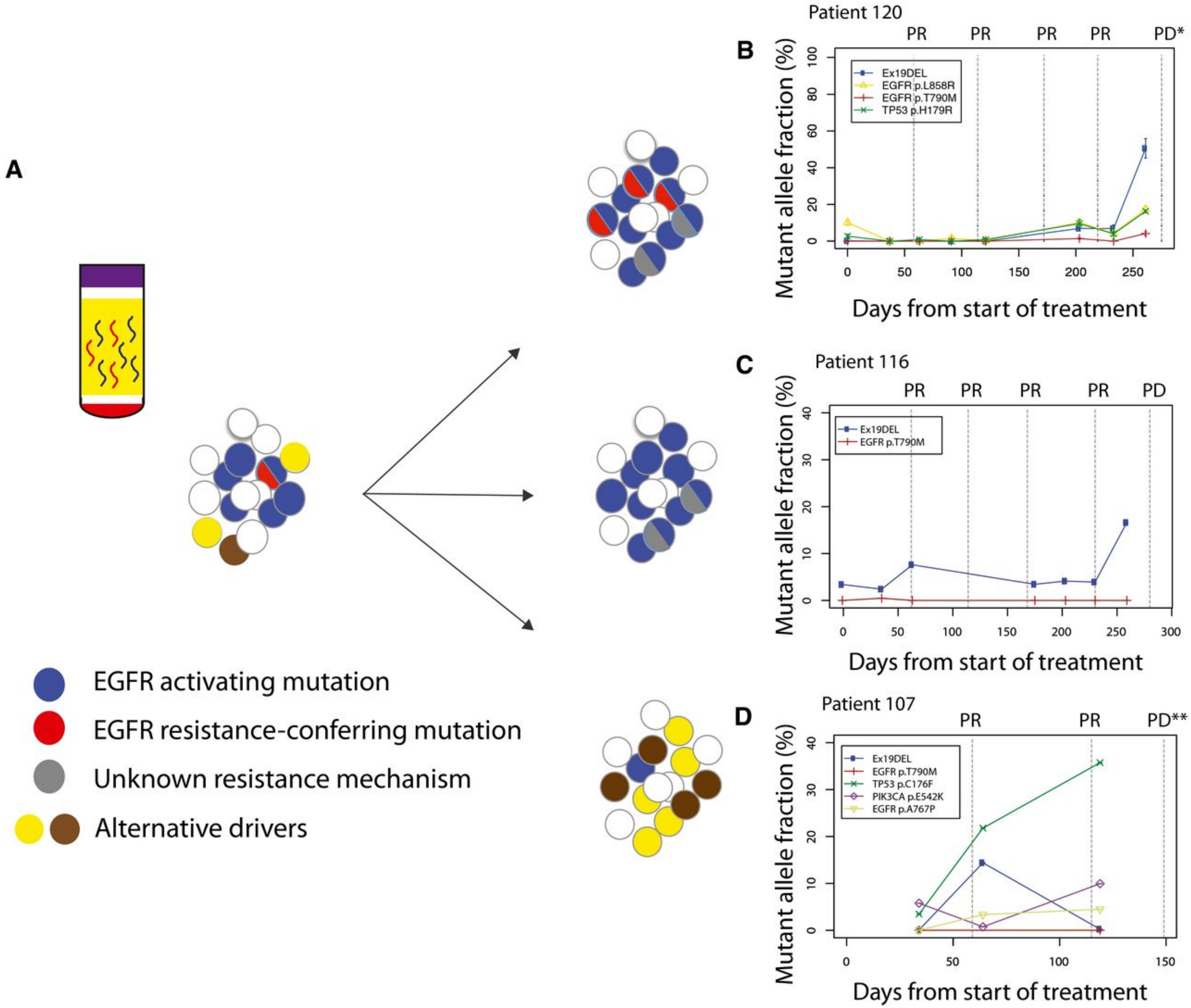

A
EGFR activating mutation
EGFR resistance-conferring mutation
Unknown resistance mechanism
Alternative drivers
B
Patient 120
PR
PR
PR
PR
PD*
Ex19DEL
EGFR p.L858R
EGFR p.T790M
TP53 p.H179R
Mutant allele fraction (%)
0
20
40
60
80
100
0
50
100
150
200
250
Days from start of treatment
C
Patient 116
PR
PR
PR
PR
PD
Ex19DEL
EGFR p.T790M
Mutant allele fraction (%)
0
10
20
30
40
0
50
100
150
200
250
300
Days from start of treatment
D
Patient 107
PR
PR
PD**
Ex19DEL
EGFR p.T790M
TP53 p.C176F
PIK3CA p.E542K
EGFR p.A767P
Mutant allele fraction (%)
0
10
20
30
40
0
50
100
150
Days from start of treatment

Figure 5.4

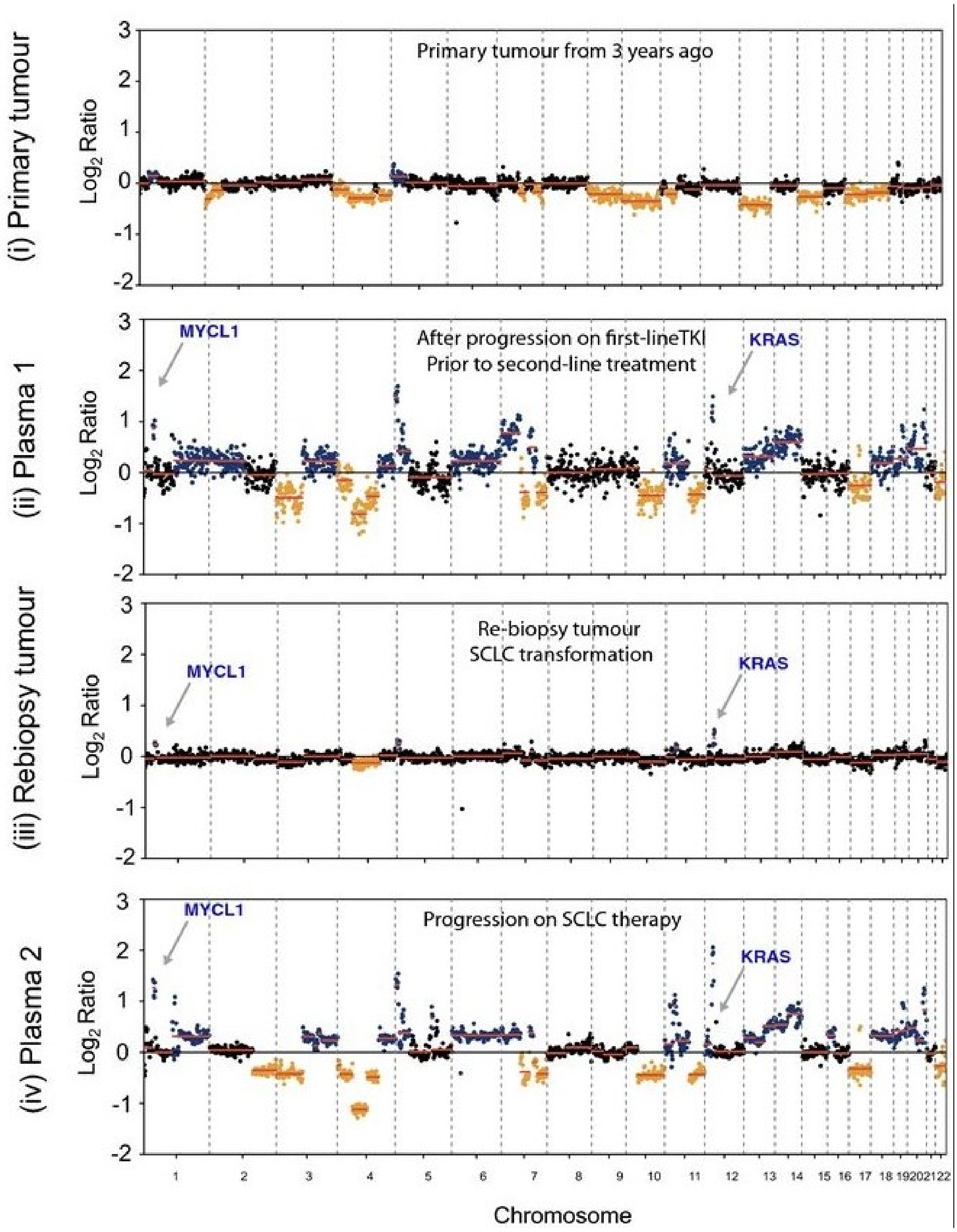


Figure 5.5: (i) Global SCNA profiles in primary tumour at diagnosis of NSCLC; (ii) Plasma SCNA profiles prior to small cell lung cancer (SCLC) transformation; (iii) Global SCNA profiles in tumour at SCLC transformation (iv) Plasma SCNA profiles after SCLC transformation and progression on cisplatin and irinotecan. *Blue: gain; Orange: loss. Figure and Figure Legend adapted from* ***Tsui et al.*** **(11)**

5.2.3: Discussion

Dawson et al. demonstrated for the first time that ctDNA could be used to monitor tumour burden, had greater sensitivity compared to CTCs and CA15-3, and often provided the earliest measure of response in patients with metastatic breast cancer. Subsequent publications by **Parkinson et al**. and Tie et al. supported these results, demonstrating that ctDNA correlated with tumour burden in ovarian and colorectal cancer, that elevated levels of ctDNA could be observed prior to progressive disease on imaging, and that ctDNA was superior to CA-125 and CEA (1,66). In **Tsui et al.**, p.T790M was detected in NSCLC patients a median of 6.8 months before clinical progression.

In **Dawson et al.**, SVs in amplified regions were shown to have elevated levels of ctDNA compared to a point mutation in the same patient. SVs can be used to monitor tumour burden. As chromosomal rearrangements are not present in germline DNA, this results in a highly specific assay with no background noise, unlike SNVs which differ by a single nucleotide. Leary et al. demonstrated that patient-specific rearrangements can be detected down to 1 molecule in 100,000 wild-type alleles (0.001%) (123). However, some SVs are passenger mutations. In **Dawson et al**., ctDNA analysis showed clonal heterogeneity, with different dominant clones observed post-treatment. This highlights that a monitoring assay should ideally track driver rather than passenger mutations, which may be in minor subclones that are lost during clonal evolution.

In **Tsui et al.**, ctDNA was used to monitor NSCLC patients treated with gefitinib and hydroxchloroquine. Three different sub-cohorts were observed in plasma – patients with an *EGFR*-activating mutation, with or without a p.T790M resistance mutation, and patients with no *EGFR*-activating or p.T790M resistance mutation. In NSCLC, in addition to p.T790M-mediated resistance, there are other mutations, including *MET* and *HER2* amplifications and *PIK3CA* mutations that may lead to drug resistance. Additional known resistance pathway include transformation to a histological classification of SCLC. In **Tsui et al.**, sWGS was used to observe changes in SCNAs in 3 patients undergoing transformation. Patients with both *EGFR* and *TP53* mutations in pre-treatment plasma had poor prognosis. **Parkinson et al.** also showed that pre-treatment ctDNA levels were prognostic, and associated with time to progression. **Tsui et al.** demonstrated that TKI-naïve patients who were p.T790M-negative had a significantly worse PFS, in agreement with Oxnard et al. who observed that p.T790M-negative patients are less likely to benefit from TKI continuation (124).

Longitudinal monitoring enables clinicians to determine whether a particular treatment is working to avoid ineffective treatment and unnecessary side-effects. Should newly arising actionable mutations be detected, then patients may become eligible for additional treatment options. Detection of SCLC histology, for example, may open up the possibility of additional SCLC treatment options (125).

The ideal monitoring assay would perform comprehensive genomic profiling, and be relatively inexpensive to enable repeat sampling. sWGS only detects SCNAs, but has potential as a monitoring assay, as is relatively inexpensive yet informative. Heitzer et al. first demonstated that sWGS to a depth of ~0.1x coverage is sufficient to detect SCNAs on a genome-wide scale (54). Chromosomal aneuploidy has been shown to predict response to immunotherapy in melanoma and other cancers (126–128). A limitation, however, is that sWGS needs to be performed on samples with sufficiently high AF (>5-10%). Methods to enrich for ctDNA could be potentially valuable to increase detection of SCNAs in samples with lower tumour fraction. As discussed in the next section, **Mouliere et al.** investigated whether size selection could enrich for tumour-specific DNA and enhance detection of clinically-relevant SCNAs.

5.3: Early detection of cancer and minimal residual disease

Initial ctDNA research largely focussed on patients with advanced cancer, where pre-treatment ctDNA levels are at their highest. More recently, focus has shifted to analysis of ctDNA in the detection of minimal residual disease, and in patients with early stage cancer. This is more challenging given the low number of mutant fragments available for analysis. **Mair et al.** and **Mouliere et al.** used paired-end sWGS to study fragmentation patterns of cfDNA **(12–14)**. **Mouliere et al.** subsequently tested whether size selection could be used to enrich for tumour-specific ctDNA fragments to improve assay sensitivity **(13,14)**.

5.3.1: Mair et al., Cancer Research, 2019 (12)

5.3.1.1: Aims

The aim of this study was to analyse human ctDNA in a patient-derived rat xenograft model of glioblastoma (GBM) **(12)**. sWGS was performed to analyse fragmentation lengths of human (tumour) and rat (host) DNA. As GBM is known to have low concentrations of ctDNA, tumour mitochondrial DNA (tmtDNA) was also analysed in rat plasma, urine and cerebrospinal fluid (CSF) by dPCR using human and rat-specific assays. tmtDNA may potentially be a more sensitive marker than ctDNA as there are 10^2-10^5 copies of the 16.5kb mitochondrial genome in each cancer cell (129).

5.3.1.2: Results

sWGS from the xenograft model showed that human tumour-specific and CSF DNA fragments centred around 145bp in length, compared to 167bp for rat DNA (Figure 5.6). Human and rat mitochondrial DNA showed a peak <100bp.

Tumour mitochondrial DNA was detected in plasma in 82% of rats, compared to 24% for ctDNA, at ~190-fold higher levels, indicating the potential of using tmtDNA to enhance detection in GBM. tmtDNA was also detected in CSF and urine. ctDNA levels in CSF were 5-8-fold higher than in plasma, possibly due to higher background levels of host DNA in plasma.

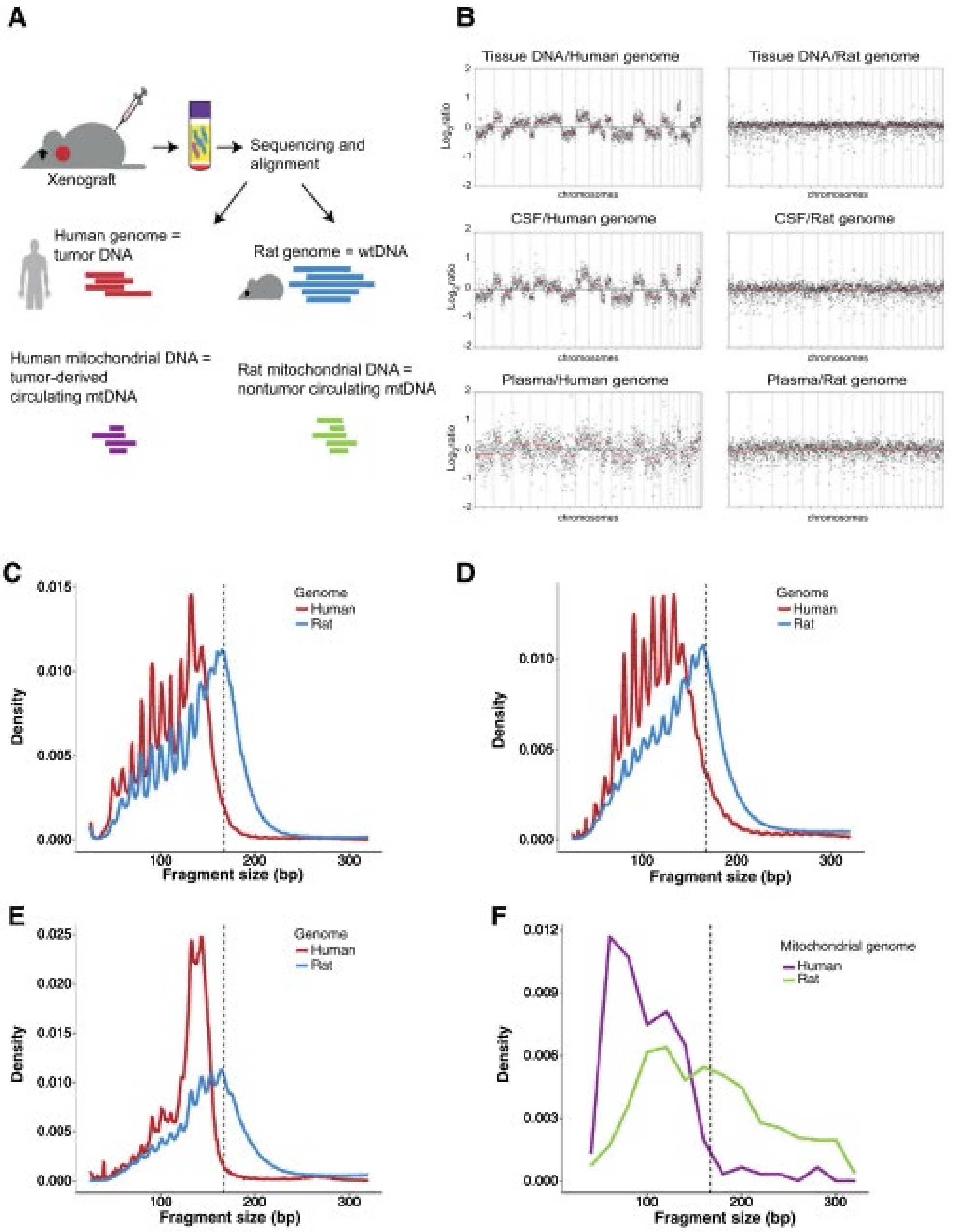
5.3.2: Mouliere et al., bioRxiv, 2017; Mouliere et al., Science Translational Medicine, 2018 (13,14)

5.3.2.1: Aims

Mouliere et al. tested whether size selection of 90bp-150bp fragments could enrich for ctDNA. This data, published initially as a bioRxiv pre-print **(13)**, was followed by further analysis of fragmentation patterns in 344 plasma samples from 200 patients with 18 different cancer types, and 65 presumed healthy controls **(14)**.

5.3.2.2: Results

In the **bioRxiv** pre-print, *in vitro* size selection was performed on plasma from 13 relapsed HGSOC patients using a PippinHT 3% agarose cassette, and analysed by TAm-Seq and sWGS. Results demonstrated that tumour-specific DNA could be enriched up to 11-fold. t-MAD (trimmed Mean Absolute Deviation from copy-number neutrality) was developed to analyse sWGS data to measure the median deviation from the copy number neutral state, and quantitatively assess SCNAs on a genome-wide scale. Size selection resulted in a median 1.5-fold increase in t-MAD score across all samples, and a median 2.9-fold increase



,

Figure 5.6: Fragmentation patterns of plasma DNA in rat xenograft model of glioblastoma. *(A) sWGS reads aligned to human (tumour) or rat (host) genomic reference sequences (B) SCNAs identified in tumour, CSF and plasma by alignment to human and rat genome. (C,D) Size distribution of human ctDNA (red) and rat DNA (blue) in two rat GBM models. Vertical line: 167bp. (E) Size distribution of CSF DNA. (F) Size distribution of mitochondrial DNA (tmtDNA) in plasma from xenografted rat. Human tmtDNA (purple); Rat tmtDNA (green). wtDNA: wild-type DNA. Figure and Figure Legend adapted from* ***Mair et al.*** **(12)**

in the majority of samples collected post-treatment where ctDNA levels were low. Two samples were highly degraded with no observed 167bp peak, and size selection resulted in a decrease in t-MAD score.

In **Mouliere et al.** (**Science Translational Medicine**), sWGS showed different fragmention patterns in patients with different cancer types and healthy controls (Figure 5.7). Longer fragments were observed in healthy controls and in patients with 'low ctDNA' cancers (renal, glioblastoma, bladder, pancreatic). 'High ctDNA' cancers [melanoma, breast, ovarian, lung, colorectal and cholangiocarcinoma (ChC)], showed an increased proportion of fragments <150bp, and an enrichment of 250bp-320bp fragments.

In silico and *in vitro* size selection was next used to enhance detection of ctDNA. t-MAD analysis of 45 pre- and post-treatment plasma samples from 35 HGSOC patients showed a mean 2.5-fold increase in 98% samples following *in vitro* size selection. In addition, SCNAs were detected which were not previously observed in pre-treatment samples (Figure 5.8), including in clinically-relevant genes, including *NF1*, *TERT* and *MYC*. Comparison of t-MAD with dPCR and WES data showed high correlation (Pearson, $r = 0.80$) between t-MAD and mutant AF in samples above the detection threshold (0.015, based on highest t-MAD score in control samples) and AF>0.025. A spike-in dilution series showed linearity of t-MAD and mutant fraction down to ~0.01 AF. t-MAD also correlated with tumour volume by RECIST v1.1 in analysis of 35 patients (Pearson, $r = 0.6$).

Machine learning algorithms were used to classify 'healthy' and 'cancer' samples by analysing cfDNA fragmentation features in sWGS data. Features used included the proportion (P) of fragments of specific size ranges, the ratio of fragments in different size ranges, and the amplitude of the 10bp periodicity oscillations below 150bp. Random Forest (RF) and Logistic Regression (LR) models were trained on 153 samples, and cross-validated on 2 independent datasets of [1] 94 samples, and [2] 83 samples from 'low ctDNA' cancers. The RF model was most predictive, using t-MAD, 10bp amplitude, P(160 to 180), P(180 to 220) and P(250 to 320) features, with an AUC of 0.994 in analysis of the 94 validation cohort, and 0.914 in 'low ctDNA' samples. It correctly classified cancer in 94% of samples from 'high ctDNA' cancers, and 65% of 'low ctDNA' cancers. Using just four fragmentation features without t-MAD resulted in an AUC of 0.989 ('high ctDNA') and 0.891 ('low ctDNA'), indicating that fragmentation patterns of cfDNA are the most predictive indicator.

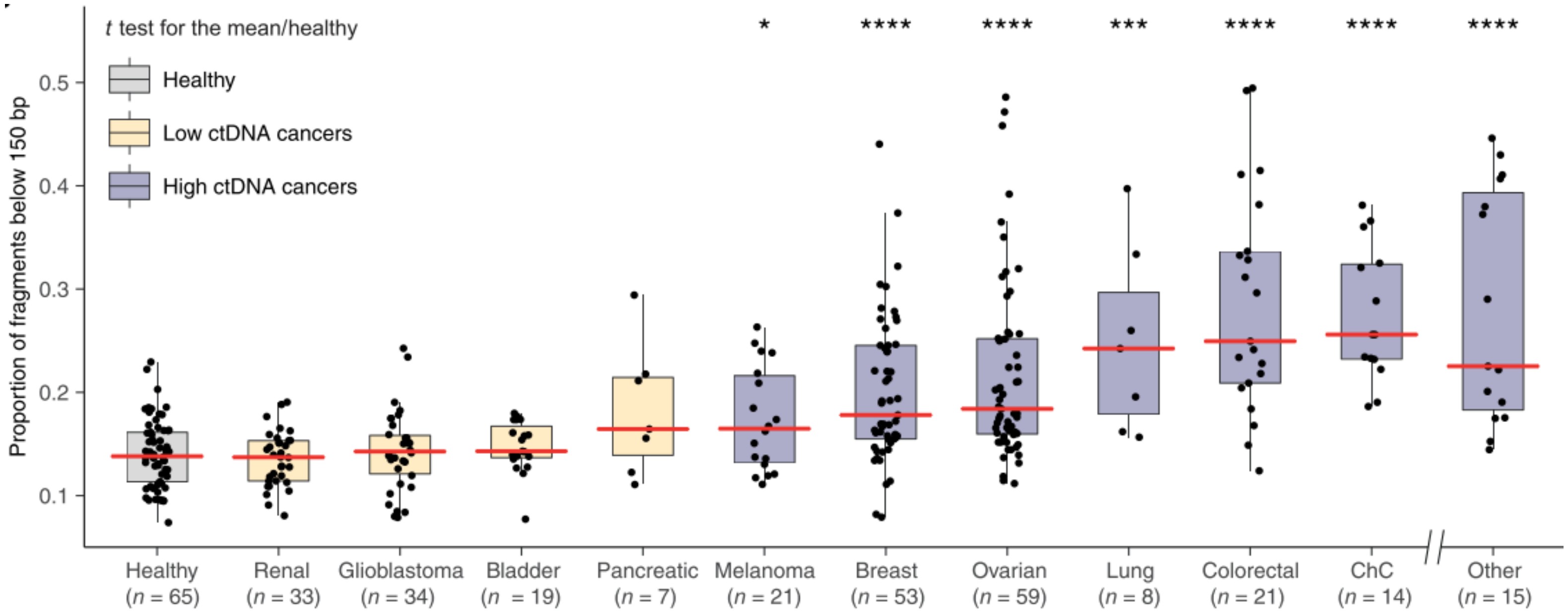


Figure 5.7: Plasma fragmentation patterns on a pan-cancer scale. *Proportion of fragments <150bp by cancer type. ChC: Cholangiocarcinoma. 'Other': Cancer types represented by <4 individuals. Red lines: median proportion for each cancer type. (Kruskal-Wallis, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 versus healthy and 'low ctDNA' cancers.. Figure and Figure Legend from* ***Mouliere et al.*** **(14)**

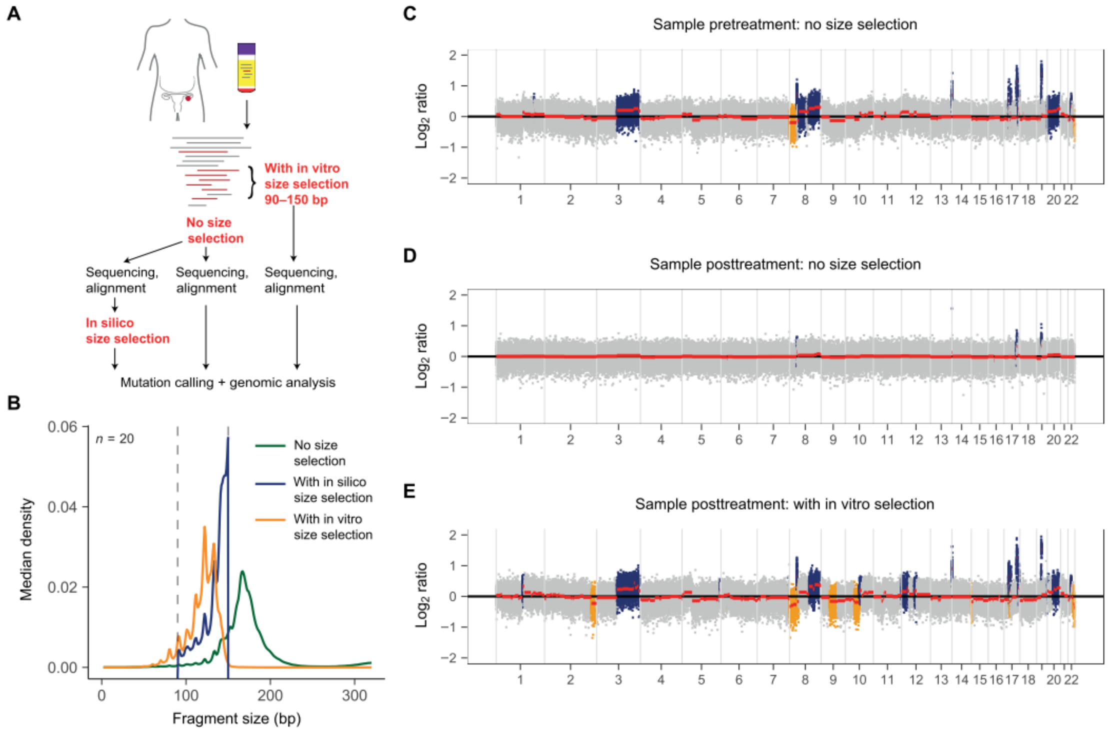


Figure 5.8: Enhancing tumour fraction from plasma sequencing with size selection. (A) *Plasma samples from HGSOC patients analysed without size selection, or using either in silico or in vitro size selection. (B) Accuracy of 90-150bp in vitro or in silico size selection demonstrated on 20 healthy controls. Green: Before size selection; Blue: After in silico size selection; Orange: After in vitro size selection. Vertical lines: 90bp, 150bp. (C) SCNA analysis of pre-treatment plasma from HGSOC patient (D) SCNA analysis of non-size selected plasma 3 weeks post-treatment (E) SCNA analysis of same plasma in (D) with 90bp-150bp in vitro size selection; Blue: amplifications; Orange: deletions: Grey: copy number neutral regions. Figures and Figure Legend from* **Mouliere et *al.* (14)**

5.3.3: Discussion

In **Mair et al.**, a rat xenograft model of glioblastoma was used to study fragmentation patterns of human tumour-specific ctDNA, and rat cfDNA, demonstrating that the ctDNA had shorter fragments, around 147bp in length. This is in agreement with previous observations that tumour-specific DNA has a different size profile (37,38,40). Following these findings**, Mouliere et al.** used size selection to enrich for tumour-specific DNA, enabling the detection of clinically-relevant SCNAs that had not previously been observed. Analysis of fragment sizes on a pan-cancer scale showed that plasma from 'high ctDNA' cancers showed an increased proportion of fragments <150bp, compared to healthy controls and 'low ctDNA' cancers. The ranking order was very similar to that observed by Bettegowda et al. who analysed the number of mutant fragment across different cancer types (42). Importantly, unlike the Bettagowda study, no prior knowledge of specific mutations was required to generate this data.

Two novel approaches were developed to analyse sWGS data. t-MAD was used to analyse sWGS data to quantitatively assess copy number data and levels of enrichment on a genome-wide scale. Results showed that t-MAD correlated with tumour volume and mutant allele fraction down to 0.01. Secondly, machine learning was used to build a model incorporating t-MAD scores and size fragmentation features to predict the presence of ctDNA, with the RF model shown to be the most predictive. These studies should be repeated on a larger scale and require further analysis to determine if other biological features of cfDNA can be incorporated to enhance detection in early stage disease or MRD.

The ultimate goal for non-invasive cancer diagnostics is to enable cancer to be detected earlier, when patients can be treated with curative intent and improve survival. Size selection and use of machine learning may potentially help in the earlier detection of cancer or MRD. Several studies have used sensitive methods to study whether ctDNA can be a prognostic biomarker in these cohorts. Garcia-Murillas et al. used dPCR to analyse ctDNA to detect MRD in a prospective cohort of early-stage breast cancer patients receiving neoadjuvant chemotherapy (130). Analysis of post-surgical plasma specimens showed that detection of ctDNA was prognostic, predicting relapse with high accuracy in a single post-surgical plasma or in serial follow-up samples (HR: 25.1; CI: 4.08-130.5; HR: 12.0; CI: 3.36-43.07; log-rank $p < 0.0001$, respectively). Chaudhuri et al. demonstrated that CAPP-Seq could be used to detect MRD in the first post-treatment plasma sample in 94% of

Stage I-III NSCLC patients treated with curative intent who went on to relapse (131). Tie et al. used Safe-SeqS to detect ctDNA in 7.9% of Stage II colon cancer patients not treated with chemotherapy, 79% of whom relapsed within a median of 27 months. In contrast, only 9.8% of patients with no detectable ctDNA subsequently relapsed (HR 18; 95% CI: 7.9 – 40; $p < 0.001$) (132). Phallen et al. used TEC-Seq (targeted error correction sequencing), involving deep sequencing of a 58-gene panel, to detect ctDNA in 71%, 59%, 59% and 68% of patients with Stage I-II colorectal, breast, lung or ovarian cancer, respectively (133).

More recently, Cohen et al. developed CancerSEEK, a multi-analyte blood test designed to detect both ctDNA, using a 61-gene panel, and 8 clinically-used protein biomarkers to analyse patients with eight common cancer types (ovarian, liver, stomach, pancreatic, oesophageal, colorectal, lung, breast cancer) (134). Analysis of 1005 Stage I-III cancer patients showed a median sensitivity of 70% ($p < 10^{-96}$ one-sided binomial test; range: 33%-98% in different cancers types), with a specificity >99%. The unique aspect of this test is the use of multiple analytes to enhance detection of cancer. This multi-modal approach shows promise, and should be explored further to enable early detection of cancer.

Chapter 6: Summary

This thesis represents nearly 10 years of research investigating the diagnostic potential of circulating tumour DNA, and assessing the hypothesis that ctDNA is a clinically useful biomarker able to monitor tumour dynamics, correlate with disease burden, and be used to guide treatment. When I initiated this work in 2009, ctDNA was known to be present in the plasma of cancer patients, but its clinical relevance had not been fully ascertained.

In **Parkinson et al.**, I was able to demonstrate for the first time that ctDNA levels correlated with tumour burden in high-grade serous ovarian cancer **(1)**. Using patient-specific digital PCR assays, mutant *TP53* ctDNA levels were monitored over time. Response to chemotherapy was seen earlier with ctDNA than CA-125, and pre-treatment TP53MAF was shown to be associated with time to progression. These studies demonstrate the potential of ctDNA in HGSOC as an early response marker. Additional dPCR analysis of specimens from a HGSOC patient with high clonal expansion demonstrated an *NF1* deletion was already present in subclonal populations prior to treatment **(2).**

The development of TAm-Seq, an amplicon-based sequencing assay, demonstrated for the first time that next generation sequencing could be used to non-invasively identify low-frequency mutations in cfDNA, and be used to monitor multiple mutations in parallel **(3)**. This opened up the possibility to use NGS for the detection and monitoring of ctDNA. TAm-Seq was subsequently used to analyse plasma specimens from ovarian, breast and lung cancer patients and demonstrate its ability to monitor tumour dynamics **(3,10,11)**. **Murtaza et al.** provided the first demonstration that exome sequencing could be used to identify potential mechanisms of resistance in plasma **(4)**. Analysis of multiple specimens from an ER+ve, HER2+ve breast cancer patient demonstrated that plasma DNA can non-invasively reflect the tumour genome, and be used to study clonal evolution **(5)**. Furthermore, **Dawson et al.** were able to demonstrate that ctDNA had greater correlation with tumour burden than CTCs and CA15-3, and often provided the earliest measure of treatment response in patients with metastatic breast cancer **(10)**.

To enable patients to have access to TAm-Seq technology, I , together with my colleagues, co-founded Inivata, which analyses ctDNA to improve patient healthcare in oncology. I led the development and analytical validation of enhanced TAm-Seq™ technology to ISO15189:2012 regulatory standards, demonstrating 94% mutations were detected at

0.25%-0.33% AF, with mutant alleles detected down to 0.02% AF **(7)**. **Remon et al.** demonstrated the clinical utility of this assay in the first prospective study of a cohort of *EGFR*-mutant NSCLC patients treated with osimertinib based on plasma profiling alone **(9)**. Results showed good response rates, with 62.5% with partial response, comparable with tissue-based testing. Further enhancements led to the development of the InVisionFirst-Lung™ assay, able to perform comprehensive genomic profiling of ctDNA, with detection of *ALK* and *ROS1* gene fusions in addition to SNVs, indels and SCNAs **(8)**. This assay has received a final Local Coverage Determination by Palmetto GBA to be used as a plasma-based test for patients with Stage IIIB/IV NSCLC. This is a major milestone in getting the assay reimbursed for use in the US, and only the second NGS ctDNA assay to have achieved this status.

Further work evaluated different pre-analytic factors that may affect ctDNA levels in plasma collected in different blood tubes and using different processing conditions **(6)**. In addition, analysis of cfDNA fragmentation patterns in a pan-cancer study and in a rat xenograft model demonstrated that tumour-specific DNA is shorter than cfDNA **(12–14)**. Furthermore, size selection and use of machine learning algorithms, incorporating size fragmentation features, enhanced detection of ctDNA and identified clinically-relevant SCNAs.

This has been an exciting decade to be involved in ctDNA research. In this time, I have seen it advance from academic research to clinical implementation for patient benefit. Encouragingly, the first companion diagnostic based on plasma profiling has now been approved for use to guide treatment with erlotinib and osimertinib for patients with advanced NSCLC. The challenge is now to improve the sensitivity of detection to enable early diagnosis of disease, at a stage when patients can be treated with curative intent, whilst ensuring assays have high specificity to limit detection of false positives. Studies by Cohen et al. demonstrate improved sensitivity of detection by incorporating both ctDNA and protein biomarker analysis (134). The use of multi-analyte assays, incorporating a combination of ctDNA and methylation, mitochondrial DNA, RNA, CTCs, exosomes, tumour-educated platelets (135) and/or protein biomarkers for example, may hopefully go some way towards achieving the ultimate goal of detecting cancer early and improving survival rates.

Definitions

AC/mL	Amplifiable copies per mL
AF	Allele fraction
AKT pathway	Signal transduction pathway involving Protein kinase B
AMP	Association for Molecular Pathology
APC	HGNC nomenclature for gene that encodes adenomatous polyposis coli protein
ARMS-PCR	Amplification-refractory mutation system PCR
Array CGH	Array comparative genomic hybridisation
AXL	HGNC nomenclature for AXL receptor tyrosine kinase
AUC	Area under the curve
BEAMing	Beads, emulsion, amplification and magnetics (36)
bp	Base pairs of DNA
BRAF	HGNC nomenclature for gene that encodes Proto-oncogene B-Raf
BRAF p.**V600E**	Specific mutation in *BRAF* gene at codon 600 resulting in an amino acid substitution from valine (V) to glutamic acid (E)
BRCA1	HGNC nomenclature for gene that encodes BReast CAncer 1 protein
BRCA2	HGNC nomenclature for gene that encodes BReast CAncer 2 protein
CA-125	Cancer antigen 125
CA15-3	Cancer antigen 15-3
CAP	College of American Pathologists
CAPP-Seq	Cancer personalized profiling by deep sequencing

CDK12	HGNC nomenclature for gene that encodes cyclin-dependent kinase 12
CDx	Companion diagnostic
CE	Clonal expansion
CEA	Carcinoembryonic antigen
cfDNA	Cell-free DNA
CFTR	HGNC nomenclature gene that encodes cystic fibrosis transmembrane conductance regulator
ChC	Cholangiocarcinoma
CHIP	Clonal hematopoiesis of indeterminate potential
CI	Confidence interval
CLIA	Clinical Laboratory Improvement Amendments (United States regulatory standards for clinical laboratory testing)
CNAR	Copy number amplification ratio
CGP	Comprehensive genomic profiling
CMS	Centers for Medicare & Medicaid Services
CR	Complete response
CRUK-CI	Cancer Research UK Cambridge Institute
CSF	Cerebrospinal fluid
CT	Computed tomography
CTCs	Circulating tumour cells
ctDNA	Circulating tumour DNA
DNA	Deoxyribonucleic acid
dPCR	Digital PCR
dsDNA	Double-stranded DNA

EGFR	HGNC nomenclature for gene that encodes epidermal growth factor receptor
***EGFR* p.L858R**	Specific activating mutation in *EGFR* gene at codon 858 resulting in an amino acid substitution from leucine (L) to arginine (R)
***EGFR* p.T790M**	Specific resistance mutation in *EGFR* gene at codon 790 resulting in an amino acid substitution from threonine (T) to methionine (M)
EDTA	Ethylenediaminetetraacetic acid
ER	HGNC nomenclature for gene encoding oestrogen receptor
ERBB2	HGNC nomenclature for gene encoding erb-b2 receptor tyrosine kinase 2 (also known as HER2)
ERBB4	HGNC nomenclature for gene encoding erb-b2 receptor tyrosine kinase 4
E.RT.0h	EDTA blood tube stored at room temperature and processed immediately (at 0 hours)
eTAm-Seq™	Enhanced tagged-amplicon deep sequencing
FDA	US Food and Drug Administration
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescent in situ hybridisation
FN	False negative
FP	False positive
GAAL	Genome-wide aggregated allelic loss(76)
GAS6	HGNC nomenclature for growth arrest specific 6 gene
GBM	Glioblastoma
GCIG	Gynaecological Cancer InterGroup
gDNA	Genomic DNA
GE/mL	Haploid genomic equivalents per mL plasma

HER2	HGNC nomenclature for the human epidermal growth factor receptor 2 gene, also known as ERBB2 (erb-b2 receptor tyrosine kinase 2)
HGNC	HUGO gene nomenclature committee (136)
HGSOC	High-grade serous ovarian carcinoma
HR	Hazard ratio
HUGO	Human Genome Organisation (137)
iDES	Integrated digital error suppression
IDS	Interval debulking surgery
IHC	Immunohistochemistry
Indels	Insertions/deletions
IQR	Interquartile range
ISO	International Organisation for Standardization
ISO15189:2012	International standard specifying quality management system requirements for medical laboratories
ITH	Intra-tumour heterogeneity
IVD	In vitro diagnostic
IU	International units
K_2EDTA	EDTA tubes containing dipotassium salt
K_3EDTA	EDTA tubes containing tripotassium salt
KRAS	HGNC nomenclature for the gene that encodes kirsten rat sarcoma virus oncogene
KW	Kruskal-Wallis test for differences in size distributions
LCD	Local Coverage Decision
LoD	Limit of detection
LoD90	Limit of detection where can detect a mutation 90% of the time

LR	Logistic regression model
LRR	LogR ratio
LUAD	Lung adenocarcinoma
LUSC	Lung squamous cell carcinoma
mCRC	Metastatic colorectal cancer
Mb	Megabase (1 million base pairs of DNA)
MEDICC	Minimal Event Distance for Intra-tumour Copy Number Comparisons
m*EGFR*	mutant *EGFR* gene
MRD	Minimal residual disease
MSI	Microsatellite instability
mtDNA	Mitochondrial DNA
MYC	HGNC nomenclature for gene that encodes MYC proto-oncogene, bHLH transcription factor
NA	Not achieved
NEJM	New England Journal of Medicine
NF1	HGNC nomenclature for gene that encodes Neurofibromin 1
NGS	Next-generation sequencing
NHS	National Health Service (UK)
NICE	National Institute for Health and Care Excellence
NIPT	Non-invasive prenatal testing
NPV	Negative predictive value
NSCLC	Non-small cell lung cancer
OS	Overall survival
PARP	HGNC nomenclature for gene that encodes Poly (ADP ribose) polymerase

PCR	Polymerase chain reaction
PD	Progressive disease
PET	Positron emission tomography
PFS	Progression-free survival
PhD	Doctor of Philosophy
PI3K pathway	Signal transduction pathway involving phosphatidylinositol-4,5-bisphosphate 3-kinase
PIK3CA	HGNC nomenclature for gene that encodes phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PLD	Pegylated liposomal doxorubicin
PPA	Positive percentage agreement
PPV	Positive predictive value
PTEN	HGNC nomenclature for gene that encodes phosphatase and tensin homolog protein
PR	Partial Response
QNS	Quantity not sufficient
qPCR	Quantitative PCR
RAS	‘Rat Sarcoma’ signalling pathway
RB1	HGNC nomenclature for gene that encodes RB transcriptional corepressor 1 (also known as retinoblastoma-associated protein)
RECIST 1.1	Response Evaluation Criteria in Solid Tumours
RF	Random forest model
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
RT	Room temperature
SCLC	Small-cell lung cancer

SCNA	Somatic copy number aberration/alteration
SD	Stable disease
SNVs	Single nucleotide variants
ssDNA	Single-stranded DNA
SVs	Structural variants
sWGS	Shallow whole genome sequencing
TAm-Seq	Tagged-amplicon deep sequencing
TCGA	The Cancer Genome Atlas
TEC-Seq	Targeted error correction sequencing
TERT	HGNC nomenclature for gene encoding telomerase reverse transcriptase
TKI	Tyrosine kinase inhibitor
t-MAD	Trimmed Mean Absolute Deviation from copy-number neutrality
TMB	Tumour mutation burden
TP53	HGNC nomenclature for gene encoding tumor protein p53
TP53MAF	*TP53* mutant allele fraction
TN	True negative
TP	True positive
TTP	Time to progression
UEA	University of East Anglia
UK	United Kingdom
ULN	Upper limit of normal
VAF	Variant allele frequency
WGS	Whole genome sequencing
WES	Whole exome sequencing

References

1. Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, et al. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. PLoS Med. 2016 Dec;13(12):e1002198.

2. Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015 Feb;12(2):e1001789.

3. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012 May;4(136):136ra68.

4. Murtaza M, Dawson S-J, Tsui DWY, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013 May;497(7447):108–12.

5. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun. 2015 Nov;6:8760.

6. Risberg B, Tsui DWY, Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson S-J, Hodgkin C, et al. Effects of Collection and Processing Procedures on Plasma Circulating Cell-Free DNA from Cancer Patients. J Mol Diagn. 2018 Nov;20(6):883–92.

7. Gale D, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, et al. Development of a highly sensitive liquid biopsy platform to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA. PLoS One. 2018;13(3):e0194630.

8. Plagnol V, Woodhouse S, Howarth K, Lensing S, Smith M, Epstein M, et al. Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling. PLoS One. 2018;13(3):e0193802.

9. Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, et al. Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by

circulating tumour DNA. Ann Oncol Off J Eur Soc Med Oncol. 2017 Apr;28(4):784–90.

10. Dawson S-J, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin S-F, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013 Mar;368(13):1199–209.

11. Tsui DWY, Murtaza M, Wong ASC, Rueda OM, Smith CG, Chandrananda D, et al. Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in non-small cell lung cancer. EMBO Mol Med. 2018 Jun;10(6).

12. Mair R, Mouliere F, Smith CG, Chandrananda D, Gale D, Marass F, et al. Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models. Cancer Res [Internet]. 2019 Jan 1;79(1):220 LP-230. Available from: http://cancerres.aacrjournals.org/content/79/1/220.abstract

13. Mouliere F, Piskorz AM, Chandrananda D, Moore E, Morris J, Smith CG, et al. Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA. bioRxiv [Internet]. 2017 Jan 1; Available from: http://biorxiv.org/content/early/2017/05/05/134437.abstract

14. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med. 2018 Nov;10(466).

15. World Health Organisation. Cancer - Key facts [Internet]. 2018. Available from: http://www.who.int/news-room/fact-sheets/detail/cancer

16. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan;100(1):57–70.

17. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar;144(5):646–74.

18. Cancer Research UK. Ovarian cancer survival statistics [Internet]. 2018. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/survival#heading-Three

19. Macconaill LE, Garraway LA. Clinical implications of the cancer genome. J Clin Oncol [Internet]. 2010 Dec 10;28(35):5219–28. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20975063

20. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990 Jun;61(5):759–67.

21. Samstein RM, Lee C-H, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet. 2019 Jan;

22. Abramson R. Overview of Targeted Therapies for Cancer. My Cancer Genome [Internet]. 2018;(Updated May 25). Available from: https://www.mycancergenome.org/content/molecular-medicine/overview-of-targeted-therapies-for-cancer/

23. Lawrence MS, Stojanov P, Polak P, Kryukov G V, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013 Jul;499(7457):214–8.

24. Sequist L V, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med. 2011 Mar;3(75):75ra26.

25. Food and Drug Administration. List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools) [Internet]. US FDA website. 2018. Available from: https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm301431.htm

26. US Food and Drug Administration. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests - Guidance for Industry and FDA Staff [Internet]. 2007. Available from: https://www.fda.gov/RegulatoryInformation/Guidances/ucm071148.htm

27. Centers for Medicare and Medicaid Services. Clinical Laboratory Improvement Amendments (CLIA) [Internet]. 2019. Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html

28. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: Towards implementation of circulating tumour DNA. Nat Rev Cancer [Internet]. 2017;17(4):223–38. Available from: http://dx.doi.org/10.1038/nrc.2017.7

29. Mandel P, Metais P. [Not Available]. C R Seances Soc Biol Fil. 1948 Feb;142(3–4):241–3.

30. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977 Mar;37(3):646–50.

31. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology. 1989;46(5):318–22.

32. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev. 1994;3(1):67–71.

33. Thierry AR, El Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. Cancer Metastasis Rev. 2016 Sep;35(3):347–76.

34. Palmirotta R, Lovero D, Cafforio P, Felici C, Mannavola F, Pelle E, et al. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. Ther Adv Med Oncol. 2018;10:1758835918794630.

35. Lo YM, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. Presence of fetal DNA in maternal plasma and serum. Lancet (London, England). 1997 Aug;350(9076):485–7.

36. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008 Sep;14(9):985–90.

37. Jiang P, Lo YMD. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. Trends Genet. 2016 Jun;32(6):360–71.

38. Mouliere F, Rosenfeld N. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. Proc Natl Acad Sci U S A. 2015 Mar;112(11):3178–9.

39. Dudley JC, Schroers-Martin J, Lazzareschi D V, Shi WY, Chen SB, Esfahani MS, et al. Detection and surveillance of bladder cancer using urine tumor DNA. Cancer Discov. 2018 Dec;

40. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment Length of Circulating Tumor DNA. PLoS Genet. 2016;12(7):1–24.

41. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell. 2016 Jan;164(1–2):57–68.

42. Sun K, Jiang P, Chan KCA, Wong J, Cheng YKY, Liang RHS, et al. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci U S A. 2015 Oct;112(40):E5503-12.

43. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc Natl Acad Sci U S A. 2018 Nov;115(46):E10925–33.

44. Serpas L, Chan RWY, Jiang P, Ni M, Sun K, Rashidfarrokhi A, et al. Dnase1l3 deletion causes aberrations in length and end-motif frequencies in plasma DNA. Proc Natl Acad Sci. 2019 Jan 8;116(2):641-649. doi: 10.1073/pnas.1815031116.

45. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome Res. 1996 Oct;6(10):986–94.

46. Little S. Amplification-refractory mutation system (ARMS) analysis of point mutations. Curr Protoc Hum Genet. 2001 May;Chapter 9:Unit 9.8.

47. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999 Aug;96(16):9236–41.

48. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014 Feb;6(224):224ra24.

49. Wu J, Matthaei H, Maitra A, Dal Molin M, Wood LD, Eshleman JR, et al. Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development. Sci Transl Med [Internet]. 2011 Jul 20;3(92):92ra66-92ra66. Available from: https://www.ncbi.nlm.nih.gov/pubmed/21775669

50. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463–7.

51. Rosenblum BB, Lee LG, Spurgeon SL, Khan SH, Menchen SM, Heiner CR, et al. New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res.

1997 Nov;25(22):4500–4.

52. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000 Jun;18(6):630–4.

53. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov;456(7218):53–9.

54. Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Med. 2013;5(4):30.

55. Belic J, Koch M, Ulz P, Auer M, Gerhalter T, Mohan S, et al. Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. Clin Chem. 2015 Jun;61(6):838–49.

56. Lo YMD, Lun FMF, Chan KCA, Tsui NBY, Chong KC, Lau TK, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci U S A. 2007 Aug;104(32):13116–21.

57. Renga B. Non invasive prenatal diagnosis of fetal aneuploidy using cell free fetal DNA. Eur J Obstet Gynecol Reprod Biol. 2018 Jun;225:5–8.

58. Ashoor G, Syngelaki A, Poon LCY, Rezende JC, Nicolaides KH. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks’ gestation: relation to maternal and fetal characteristics. Ultrasound Obstet Gynecol. 2013 Jan;41(1):26–32.

59. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol. 2010 May;221(1):49–56.

60. The Cancer Genome Atlas Research. Integrated genomic analyses of ovarian carcinoma. Nature. 2011 Jun;474(7353):609–15.

61. Macintyre G, Goranova TE, De Silva D, Ennis D, Piskorz AM, Eldridge M, et al. Copy number signatures and mutational processes in ovarian carcinoma. Nat Genet [Internet]. 2018;50(9):1262–70. Available from: https://doi.org/10.1038/s41588-018-0179-8

62. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer Biol Med. 2017 Feb;14(1):9–32.

63. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). Eur J Cancer [Internet]. 2009 Jan 1;45(2):228–47. Available from: https://doi.org/10.1016/j.ejca.2008.10.026

64. Swisher EM, Wollan M, Mahtani SM, Willner JB, Garcia R, Goff BA, et al. Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. Am J Obstet Gynecol. 2005 Sep;193(3 Pt 1):662–7.

65. Otsuka J, Okuda T, Sekizawa A, Amemiya S, Saito H, Okai T, et al. Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. Int J Gynecol Cancer. 2004;14(3):459–64.

66. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. Ann Oncol Off J Eur Soc Med Oncol. 2015 Aug;26(8):1715–22.

67. Newman AM, Bratman S V., To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20(5); 548-54. doi: 10.1038/nm.3519.

68. McEvoy AC, Warburton L, Al-Ogaili Z, Celliers L, Calapre L, Pereira MR, et al. Correlation between circulating tumour DNA and metabolic tumour burden in metastatic melanoma patients. BMC Cancer [Internet]. 2018 Jul 9;18:726. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC6038195/

69. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017 Apr;545(7655):446–51.

70. Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, Markowetz F. Phylogenetic quantification of intra-tumour heterogeneity. PLoS Comput Biol. 2014 Apr;10(4):e1003535.

71. Diaz LA, Williams R, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature [Internet]. 2012 Jun 28;486(7404):537–40. Available from:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3436069/

72. Watkins TBK, Schwarz RF. Phylogenetic Quantification of Intratumor Heterogeneity. Cold Spring Harb Perspect Med. 2018 Apr;8(4).

73. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, Roth A, et al. Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol [Internet]. 2013 Jun 18;231(1):21–34. Available from: https://doi.org/10.1002/path.4230

74. Xu T, Kang X, You X, Dai L, Tian D, Yan W, et al. Cross-Platform Comparison of Four Leading Technologies for Detecting EGFR Mutations in Circulating Tumor DNA from Non-Small Cell Lung Carcinoma Patient Plasma. Theranostics [Internet]. 2017 Apr 2;7(6):1437–46. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28529628

75. Stahlberg A, Krzyzanowski PM, Jackson JB, Egyud M, Stein L, Godfrey TE. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. Nucleic Acids Res. 2016;44(11):1–7.

76. Chan KCA, Jiang P, Zheng YWL, Liao GJW, Sun H, Wong J, et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. Clin Chem. 2013 Jan;59(1):211–24.

77. Zhang Z, Lee JC, Lin L, Olivas V, Au V, LaFramboise T, et al. Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat Genet. 2012 Jul;44(8):852–60.

78. Liu L, Greger J, Shi H, Liu Y, Greshock J, Annan R, et al. Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. Cancer Res. 2009 Sep;69(17):6871–8.

79. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009 Feb;27(2):182–9.

80. Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA. PLoS One. 2015;10(10):e0140712.

81. Guardant Health. Guardant360 Specification Sheet [Internet]. 2018. Available from: https://www.therapyselect.de/sites/default/files/downloads/guardant360/guardant360_specification-sheet_en.pdf

82. Clark TA, Chung JH, Kennedy M, Hughes JD, Chennagiri N, Lieber DS, et al. Analytical Validation of a Hybrid Capture–Based Next-Generation Sequencing Clinical Assay for Genomic Profiling of Cell-Free Circulating Tumor DNA. J Mol Diagnostics [Internet]. 2018;20(5):686–702. Available from: https://doi.org/10.1016/j.jmoldx.2018.05.004

83. Foundation Medicine. FoundationOne Liquid Technical Specification Sheet [Internet]. 2018. Available from: https://assets.ctfassets.net/vhribv12lmne/3SPYAcbGdqAeMsOqMyKUog/d0eb51659e08d733bf39971e85ed940d/F1L_TechnicalInformation_MKT-0061-04.pdf

84. Personal Genome Diagnostics. PlasmaSELECT 64+MSI [Internet]. 2017. Available from: http://www.personalgenome.com/wp-content/uploads/2017/04/PS-64_MSI-_TESTBUL_CS-002-002.pdf

85. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A. 2011 Jun;108(23):9530–5.

86. Kinde I, Bettegowda C, Wang Y, Wu J, Agrawal N, Shih I-M, et al. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. Sci Transl Med. 2013 Jan;5(167):167ra4.

87. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A. 2012 Sep;109(36):14508–13.

88. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol. 2016 May;34(5):547–55.

89. El Messaoudi S, Rolet F, Mouliere F, Thierry AR. Circulating cell free DNA: Preanalytical considerations. Clin Chim Acta. 2013 Sep;424:222–30.

90. Zhao Y, Li Y, Chen P, Li S, Luo J, Xia H. Performance comparison of blood collection tubes as liquid biopsy storage system for minimizing cfDNA contamination from genomic DNA. J Clin Lab Anal. 2018 Sep;e22670.

91. Parpart-Li S, Bartlett B, Popoli M, Adleff V, Tucker L, Steinberg R, et al. The Effect of Preservative and Temperature on the Analysis of Circulating Tumor DNA. Clin Cancer Res. 2017 May;23(10):2471–7.

92. Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F, Holtrup F. Performance of Streck cfDNA Blood Collection Tubes for Liquid Biopsy Testing. PLoS One [Internet]. 2016 Nov 10;11(11):e0166354–e0166354. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27832189

93. Arbour KC, Riely GJ. Diagnosis and Treatment of Anaplastic Lymphoma Kinase-Positive Non-Small Cell Lung Cancer. Hematol Oncol Clin North Am [Internet]. 2017 Feb;31(1):101–11. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27912826

94. Fuse MJ, Okada K, Oh-hara T, Ogura H, Fujita N, Katayama R. Mechanisms of Resistance to NTRK Inhibitors and Therapeutic Strategies in NTRK1-Rearranged Cancers. Mol Cancer Ther [Internet]. 2017 Oct 1;16(10):2130 LP-2143. Available from: http://mct.aacrjournals.org/content/16/10/2130.abstract

95. Mattocks CJ, Morris MA, Matthijs G, Swinnen E, Corveleyn A, Dequeker E, et al. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur J Hum Genet [Internet]. 2010/07/28. 2010 Dec;18(12):1276–88. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20664632

96. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn. 2017 May;19(3):341–65.

97. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. J Mol Diagnostics [Internet]. 2018 Jan 1;20(1):4–27. Available from: https://doi.org/10.1016/j.jmoldx.2017.11.003

98. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013 Apr;41(6):e67.

99. Park G, Park JK, Shin S-H, Jeon H-J, Kim NKD, Kim YJ, et al. Characterization of background noise in capture-based targeted sequencing data. Genome Biol. 2017 Jul;18(1):136.

100. Krimmel JD, Schmitt MW, Harrell MI, Agnew KJ, Kennedy SR, Emond MJ, et al. Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues. Proc Natl Acad Sci U S A. 2016 May;113(21):6005–10.

101. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014 Dec;371(26):2477–87.

102. Torga G, Pienta KJ. Patient-Paired Sample Congruence Between 2 Commercial Liquid Biopsy Tests. JAMA Oncol. 2018 Jun;4(6):868–70.

103. Forbes. Liquid Biopsy Controversy Rages On [Internet]. 2018. Available from: https://www.forbes.com/sites/victoriaforster/2018/07/19/liquid-biopsy-controversy-rages-on/#63434933dac1

104. SV A, JR W, Jones S. Regarding the congruence between 2 circulating tumor dna sequencing assays. JAMA Oncol [Internet]. 2018 Oct 1;4(10):1430–1. Available from: http://dx.doi.org/10.1001/jamaoncol.2018.2323

105. Inivata. Inivata to Present Results of Prospective Clinical Validation Study of its InVisionFirst-Lung Liquid Biopsy Test in Advanced Non-Small Cell Lung Cancer (NSCLC) at the World Conference on Lung Cancer [Internet]. 2018. Available from: https://www.inivata.com/inivata-to-present-results-of-prospective-clinical-validation-study-of-its-invisionfirst-lung-liquid-biopsy-test-in-advanced-non-small-cell-lung-cancer-nsclc-at-the-world-conference-on-lung-cancer/

106. Inivata. Inivata Receives Medicare Final Coverage Decision for InVisionFirst-Lung Liquid Biopsy Test in Advanced Non-Small Cell Lung Cancer [Internet]. 2019. Available from: https://www.inivata.com/inivata-receives-medicare-final-coverage-decision-for-invisionfirst-lung-liquid-biopsy-test-in-advanced-non-small-cell-lung-cancer/

107. GenomeWeb. https://www.genomeweb.com/liquid-biopsy/inivata-medicare-local-coverage-decision-finalized-invisionfirst-lung-liquid-biopsy [Internet]. 2019.

Available from: https://www.genomeweb.com/liquid-biopsy/inivata-medicare-local-coverage-decision-finalized-invisionfirst-lung-liquid-biopsy#.XH6qKIj7Q2w

108. Palmetto GBA. Proposed Local Coverage Determination (LCD): MolDX: Inivata, InVisionFirst, Liquid Biopsy for Patients with Lung Cancer (DL37870) [Internet]. 2018. Available from: https://www.cms.gov/medicare-coverage-database/license/cpt-license.aspx?from=~/overview-and-quick-search.aspx&npage=/medicare-coverage-database/details/lcd-details.aspx&LCDId=37869&ver=12&name=391*1%257C394*1%257C392*1%257C395*1%257C393*1%257C396*1%257C373*1%257C374*1%25

109. Lin JJ, Cardarella S, Lydon CA, Dahlberg SE, Jackman DM, Janne PA, et al. Five-Year Survival in EGFR-Mutant Metastatic Lung Adenocarcinoma Treated with EGFR-TKIs. J Thorac Oncol. 2016 Apr;11(4):556–65.

110. Howlader N, Noone AM, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ CK(. SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Bethesda [Internet]. 2004. Available from: https://seer.cancer.gov/archive/csr/1975_2011/

111. FDA. Cobas EGFR MUTATION TEST V2 Pre-Market Approval [Internet]. 2016. Available from: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P150044

112. FDA. Device Approvals and Clearances. Cobas®EGFR Mutation Test v2— P150047. Patient Labeling. Table 22. [Internet]. 2016. Available from: https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150044C.pdf

113. Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res. 2013 Apr;19(8):2240–7.

114. Janne PA, Yang JC-H, Kim D-W, Planchard D, Ohe Y, Ramalingam SS, et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med. 2015 Apr;372(18):1689–99.

115. Mok TS, Wu Y-L, Ahn M-J, Garassino MC, Kim HR, Ramalingam SS, et al. Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. N Engl J Med [Internet]. 2017;376(7):629–40. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/27959700

116. Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, et al. Association Between Plasma Genotyping and Outcomes of Treatment With Osimertinib (AZD9291) in Advanced Non-Small-Cell Lung Cancer. J Clin Oncol [Internet]. 2016/06/27. 2016 Oct 1;34(28):3375–82. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27354477

117. Yang JC-H, Ahn M-J, Kim D-W, Ramalingam SS, Sequist L V, Su W-C, et al. Osimertinib in Pretreated T790M-Positive Advanced Non-Small-Cell Lung Cancer: AURA Study Phase II Extension Component. J Clin Oncol. 2017 Apr;35(12):1288–96.

118. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med. 2015 Jun;21(6):560–2.

119. Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med. 2004 Aug;351(8):781–91.

120. Mego M, De Giorgi U, Dawood S, Wang X, Valero V, Andreopoulou E, et al. Characterization of metastatic breast cancer patients with nondetectable circulating tumor cells. Int J cancer. 2011 Jul;129(2):417–23.

121. Pierga J-Y, Hajage D, Bachelot T, Delaloge S, Brain E, Campone M, et al. High independent prognostic and predictive value of circulating tumor cells compared with serum tumor markers in a large prospective trial in first-line chemotherapy for metastatic breast cancer patients. Ann Oncol Off J Eur Soc Med Oncol. 2012 Mar;23(3):618–24.

122. Sequist L V, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, et al. Genotypic and Histological Evolution of Lung Cancers Acquiring Resistance to EGFR Inhibitors. Sci Transl Med [Internet]. 2011 Mar 23;3(75):75ra26 LP-75ra26. Available from: http://stm.sciencemag.org/content/3/75/75ra26.abstract

123. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med. 2010 Feb;2(20):20ra14.

124. Oxnard GR, Arcila ME, Sima CS, Riely GJ, Chmielecki J, Kris MG, et al. Acquired

resistance to EGFR tyrosine kinase inhibitors in EGFR-mutant lung cancer: distinct natural history of patients with tumors harboring the T790M mutation. Clin Cancer Res. 2011 Mar;17(6):1616–22.

125. Niederst MJ, Sequist L V, Poirier JT, Mermel CH, Lockerman EL, Garcia AR, et al. RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. Nat Commun. 2015 Mar;6:6377; doi: 10.1038/ncomms7377.

126. Silva S, Danson S, Teare D, Taylor F, Bradford J, McDonagh AJG, et al. Genome-Wide Analysis of Circulating Cell-Free DNA Copy Number Detects Active Melanoma and Predicts Survival. Clin Chem. 2018 Jun; 64(9):1338-1346; doi:10.1373/ clinchem. 2018.290023

127. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science. 2017 Jan;355(6322).

128. Vanderstichele A, Busschaert P, Smeets D, Landolfo C, Van Nieuwenhuysen E, Leunen K, et al. Chromosomal Instability in Cell-Free DNA as a Highly Specific Biomarker for Detection of Ovarian Cancer in Women with Adnexal Masses. Clin Cancer Res. 2017 May;23(9):2223–31.

129. Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. Elife. 2014; Oct 1;3. doi: 10.7554/eLife.02935.

130. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015; Aug 26;7(302):302ra133. doi: 10.1126/scitranslmed.aab0021.

131. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, Stehr H, Azad TD, et al. Early Detection of Molecular Residual Disease in Localized Lung Cancer by Circulating Tumor DNA Profiling. Cancer Discov [Internet]. 2017/09/24. 2017 Dec;7(12):1394–403. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28899864

132. Tie J, Wang Y, Tomasetti C, Li L, Springer S, Kinde I, et al. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. Sci Transl Med [Internet]. 2016 Jul 6;8(346):346ra92-346ra92. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27384348

133. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med [Internet]. 2017 Aug 16;9(403):eaan2415. Available from: http://stm.sciencemag.org/content/9/403/eaan2415.abstract

134. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science. 2018 Feb;359(6378):926–30.

135. Best MG, Wesseling P, Wurdinger T. Tumor-Educated Platelets as a Noninvasive Biomarker Source for Cancer Detection and Progression Monitoring. Cancer Res. 2018 Jul;78(13):3407–12.

136. HUGO Gene Nomenclature Committee. Curated online repository of HGNC-approved gene nomenclature, gene families and associated resources including links to genomic, proteomic and phenotypic information [Internet]. 2018. Available from: https://www.genenames.org/

137. Human Genome Organisation. Website of the Human Genome Organisation (HUGO) [Internet]. 2018. Available from: http://www.hugo-international.org/

Appendix 1 : Summary of each submitted publication and details of my contribution

1. **Parkinson C*, Gale D*, Piskorz A, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, Earl H, Qian W, Rosenfeld N†, Brenton JD†; *PLoS Medicine*; 2016; 13 (12): e1002198 doi:10.1371/ journal. pmed.1002198; 'Exploratory Analysis of *TP53* Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study.'**

Summary

The aim of this study was to investigate circulating tumour DNA (ctDNA) as an early response marker in relapsed high-grade serous ovarian cancer (HGSOC). Plasma specimens (n=318) from 40 patients were analysed using patient-specific digital PCR assays (n=31) to assess the *TP53* mutant allele fraction (TP53MAF), compared to serum CA-125 and tumour burden by volumetric analysis. ctDNA levels correlated with disease volume, and pre-treatment TP53MAF, not CA-125, was associated with poor response and time to progression (TTP). A decrease of ≤60% TP53MAF after one cycle of chemotherapy was associated with TTP in <6 months, demonstrating the potential of ctDNA as an early response marker in ovarian carcinoma.

Author contributions

I co-led this study and was joint first co-author on this PLoS Medicine publication. I established protocols and lab infrastructure for the analysis of ctDNA from start-up of the lab, including protocols for the collection of high-quality specimens from the gynaecological clinic, extraction protocols and digital PCR (dPCR) analysis of *TP53* mutations in cell-free DNA (cfDNA). I designed and validated 31 patient-specific digital PCR assays, and used these to detect and monitor low frequency mutations in 318 plasma specimens, the largest study at the time. I, together with other team members, analysed data and correlated with clinical findings. I was actively involved in writing and reviewing the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

2. **Schwarz R, Ng C, Cooke S, Newman S, Temple J, Piskorz A, Gale D, Sayal K, Murtaza M, Baldwin P, Rosenfeld N, Earl H, Sala E, Jimenez-Linan M, Parkinson C, Markowetz F†, Brenton JD†; *PLoS Medicine*; 2015; Feb 24;12(2): e1001789;**
'Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis.'

Summary

This paper involved study of intra-tumour heterogeneity (ITH) in patients with high-grade serous ovarian cancer to determine whether quantitative measures of ITH could predict outcome. Spatially and temporally-separated tumour specimens (n=135) were analysed from 14 patients receiving platinum-based chemotherapy. Evolutionary inference and phylogenetic quantification of heterogeneity was performed by analysis of copy number profiles and WGS data using the MEDICC algorithm. Patients with high clonal expansion (CE-high) had shorter progression-free and overall survival than CE-low patients. Analysis of an *NF1* deletion in a CE-high patient pre- and post-relapse indicated that resistant subclonal populations were already present prior to treatment.

Author contributions

I designed and validated digital PCR assays for the detection of patient-specific *NF1* and *TP53* mutations in tumour and ascites fluid, and was able to confirm the presence of these mutations in subclonal populations prior to treatment, which less sensitive methods had been unable to do. I analysed dPCR data and prepared figures and text for inclusion in the manuscript. All author contributions are detailed in Appendix 3.

3. Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DWY*, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD†, Rosenfeld N†; *Science Translational Medicine*; 2012; 4(136):136ra68;
'Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.'

Summary

Our team developed TAm-Seq™ (tagged-amplicon deep sequencing), a novel amplicon-based next-generation sequencing (NGS) assay to non-invasively identify low-frequency mutations in cell-free DNA. Previous studies demonstrated the potential of ctDNA by analysis of single hotspot mutations. This study provided the first demonstration of the use of NGS to screen large genomic regions to monitor multiple mutations in parallel to detect mutations at an allele fraction (AF) of 2%, with >97% sensitivity and specificity. TAm-Seq was used to identify a *de-novo EGFR* mutation in plasma, not previously identified in the initial tumour biopsy, and monitor tumour dynamics by tracking 10 concomitant mutations in parallel.

Author contributions

I designed and validated digital PCR assays to demonstrate the quantitative accuracy of TAm-Seq, including developing a dPCR assay to validate a low frequency *de novo* mutation identified by TAm-Seq. I performed the first TAm-Seq experiments in the group to analyse FFPE (formalin-fixed paraffin-embedded) tumour DNA. I worked with the team to optimise the TAm-Seq protocol for detection of low frequency mutations in ctDNA, and provided technical advice and training. I revised and modified the TAm-Seq paper prior to publication. I devised and optimised collection and extraction protocols for the collection of high-quality ovarian and breast cancer specimens. All author contributions are detailed in Appendix 3.

4. **Murtaza M*, Dawson SJ*, Tsui DWY*, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin S, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD†, Caldas C†, Rosenfeld N†*; Nature*; 2013 May 2; 497(7447):108-12**
'Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.'

Summary

Exome sequence analysis was performed on serial plasma samples from patients with advanced breast, ovarian and lung cancer to study potential mechanisms of acquired drug resistance. This paper provided the first proof-of-principle demonstration of the use of exome sequencing to non-invasively assess tumour heterogeneity on an exome-wide scale, and study genomic evolution pre- and post-treatment following resistance. Known resistance-causing mutations were identified including an *EGFR* T790M mutation in a patient treated with gefitinib, and a *GAS6* mutation (ligand for AXL). Activation of the AXL pathway is known to cause TKI and lapatinib resistance.

Author contributions

I established lab protocols and infrastructure for the analysis of ctDNA, and trained a PhD student, post-doc and clinical fellow that led this study on these methods. I established protocols for collection and DNA extraction of clinical specimens from the local ovarian and breast cancer oncology clinics used in this study. I advised the lead author on the optimal method for exome sequencing from cfDNA which had proved challenging using standard protocols. I was involved in generating libraries, the interpretation of data and critically reviewing the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

5. **Murtaza M*, Dawson SJ*, Pogrebniak K, Rueda O, Provenzano E, Grant J, Chin SF, Tsui DWY, Marass F, Gale D, Ali HR, Shah P, Contente-Cuomo T, Farahani H, Shumansky K, Kingsbury Z, Humphray S, Bentley D, Shah S, Wallis M, Rosenfeld N†, Caldas C†; *Nature Communications*; 2015; 6:8760 doi: 10.1038/ncomms9760**
'Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer'.

Summary

Following on from the Murtaza et al. *Nature* paper, this study performed exome sequence analysis and targeted amplicon deep sequencing of specimens from a patient with metastatic ER+ve and HER2+ve breast cancer. Nine plasma samples and 8 tumour biopsies, including autopsy specimens, were analysed across 1,193 days of clinical follow-up. This study provided a proof-of-principle demonstration of the ability of ctDNA to be able to non-invasively reflect the genomic architecture and clonal evolution inferred by sequencing of the tumours, and study serial changes in stem and private mutations to assess tumour dynamics following treatment.

Author contributions

I established lab protocols and infrastructure for the analysis of ctDNA, and trained the PhD student and clinical fellow that led this study on these methods. I established protocols for collection and DNA extraction of clinical specimens from the local breast cancer oncology clinic used in this study. I advised the lead author on the optimal method for exome sequencing from cfDNA which had proved challenging using standard protocols. I was involved in critically reviewing the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

6. **Risberg B*, Tsui DWY*†, Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, Brenton JD, Gale D†; *The Journal of Molecular Diagnostics*, 2018; *Volume 20 , Issue 6 , 883 - 892*. *Senior co-corresponding author; 'Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patient'.**

Summary

This study analysed different pre-analytical factors that can affect cfDNA levels, including delays in plasma processing, storage temperatures, use of different blood tubes (EDTA v Streck cell-free DNA BCT cell preservation tubes), centrifugation protocols and sample shipment. Analysis was performed using dPCR, TAm-Seq or shallow whole genome sequencing (sWGS). BCT tubes stabilised DNA release, whilst EDTA storage at 4°C showed less cfDNA variation compared to room temperature storage. Similar cfDNA levels were observed using second centrifugation at 3000g v 14,000g, providing protocols for hospitals with limited equipment. NGS analysis of DNA from ETDA or BCT tubes showed similar mutational profiles, background error and copy number alterations.

Author contributions

I was senior co-corresponding author on this publication. Having initiated studies evaluating delayed processing of Streck and EDTA tubes in 2010, I devised and co-led this study with Dr. Dana Tsui, a postdoctoral research associate in our group. I worked with Heather Biggs, a research assistant in the ovarian and breast cancer clinics at Addenbrooke's Hospital, Cambridge, to train and arrange collection and processing of blood samples for evaluation. I oversaw the project with Dr. Tsui, and trained team members in different techniques. In addition, I designed and validated digital PCR assays which were used by Bente Risberg, first co-author, to assess cfDNA and ctDNA levels. I reviewed data, revised and improved the manuscript and figures, and led interactions with the journal editors prior to publication. All author contributions are detailed in Appendix 3.

7. Gale D‡, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G, Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N†; *PLoS ONE*; 2018; 13(3): e0194630; 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.'

Summary

In this paper, I led the development and analytical validation of the Inivata InVision™ next-generation sequencing (NGS) assay to detect low frequency mutations in cfDNA. This clinical diagnostic assay is based on eTAm-Seq™ technology, designed to identify clinically-relevant somatic mutations in a panel of 35 cancer-related genes. Analytical validation was performed in two independent clinical laboratories to CLIA and ISO15189:2012 quality standards. Analysis of optimal input DNA detected mutant alleles down to 0.02%, with 94% mutations at 0.25%-0.33% AF, and per-base specificity of 99.9997%. These studies demonstrate that InVision is a highly sensitive assay for the detection of clinically-relevant mutations in plasma.

Author contributions

As first and senior co-corresponding author on this publication, I led technology transfer of TAm-Seq™ from CRUK-CI to Inivata, led the development of eTAm-Seq™ technology, including development to ISO15189:2012 quality standards, and led the team involved in processing clinical specimens for several studies and for analytical validation of the assay. I was involved in development of the analytical validation plan and data analysis, wrote the first draft and subsequent modifications of the manuscript, and was the senior author in discussions with journal editors prior to publication. All author contributions are detailed in Appendix 3.

8. Plagnol V*, Woodhouse S*, Howarth K, Lensing S, Smith M, Epstein M, Madi M, Smalley S, Leroy C, Hinton J, De Kievit F, Musgrave-Brown E, Herd C, Neblett K, Brennan W, Dimitrov P, Campbell N, Morris C, Rosenfeld N, Clark J, Gale D, Platt J, Calaway J, Jones G, Forshew T†; *PLoS ONE*; 2018; 13(3): e01938022018; 'Analytical validation of a Next Generation Sequencing liquid biopsy assay for high sensitivity broad molecular profiling.'

Summary

This study performed analytical validation of the InVisionFirst™ assay, based on eTAm-Seq™ technology. The assay was developed to profile 36 genes commonly mutated in NSCLC, and able to detect SNVs, indels, copy number amplifications, and *ALK* and *ROS1* gene fusions, thereby providing comprehensive genomic profiling of low frequency mutations in ctDNA. The InVisionFirst™ assay demonstrated 99.48% sensitivity for detection of SNVs at 0.25%-0.33% AF, 92.46% sensitivity for indels at 0.25% AF, and 99.9997% per-base specificity. Comparison with digital PCR showed high concordance. These results demonstrate InVisionFirst™ is a highly sensitive and specific assay, suitable for use in clinical applications.

Author contributions

Following the development of eTAm-Seq, I was responsible for implementation of assay automation, which led to a significant improvements in assay sensitivity and robustness. I provided technical advice to the technology development team on development of the assay, helped with design of the analytical validation plan to standards required for reimbursement and contributed to the dossier for reimbursement, reviewed data and helped revise the manuscript prior to publication. All author contributions are detailed in Appendix 3.

9. Remon J, Caramella C, Jovelet C, Lacroix L, Lawson ARJ, Smalley S, Howarth K, Gale D, Green E, Plagnol V, Rosenfeld N, Planchard D, Bluthgen MV, Gazzah A, Pannet C, Nicotra C, Auclin E, Soria JC, Besse B†; *Annals of Oncology*; 2017; 28:784-790; 'Osimertinib benefit in EGFR-mutant NSCLC patients with T790M mutation detected by circulating tumour DNA.'

Summary

This was the first prospective study of ctDNA in a cohort of *EGFR*-mutant non-small cell lung cancer (NSCLC) patients treated with osimertinib, a third-generation TKI active against *EGFR* T790M resistance mutations. Analysis was performed on 48 activating m*EGFR*+ patients with acquired resistance to TKIs, who had no available tissue biopsy. T790M status was detected in plasma in 50% patients using InVision eTAm-Seq™ technology. 62.5% patients gave a partial response following treatment, and 37.5% had stable disease. These studies indicate the clinical utility of InVision™ to detect T790M mutations in plasma where no tissue biopsy is available, demonstrating good response rates.

Author contributions

As Head of Molecular Diagnostics at Inivata, I led the development of eTAm-Seq technology (see Gale et al., Chapter 4.2), and its implementation to ISO15189:2012 quality standards for clinical use. I also led the team responsible for processing clinical specimens in this study. I was involved in data review and quality control, and reviewed and edited the manuscript ahead of publication. In my academic role at the Cancer Research UK Cambridge Institute, I participated in discussions with clinicians at the Institut Gustave Roussy, the clinical site where patients were recruited for this study, which ultimately led to this collaboration. All author contributions are detailed in Appendix 3.

10. Dawson SJ*, Tsui D.W.Y*, Murtaza M, Biggs H, Rueda O, Chin SF, Dunning M, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C†, Rosenfeld N†; *New England Journal of Medicine*; 'Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer.'

Summary

This study compared ctDNA with CA 15-3, circulating tumour cells (CTCs) and radiographic tumour imaging in 30 patients with metastatic breast cancer. ctDNA was detected in 97% cases, compared to CA 15-3 (78%) and CTCs (87%) respectively, showing a greater dynamic range and correlation with changes in tumour burden than the other circulating biomarkers. Importantly, ctDNA provided the earliest measure of treatment response in 53% patients, with detection of elevated ctDNA levels on average 5 months earlier than detection of progressive disease by imaging. This study provided an important proof-of-concept demonstration of the potential of ctDNA in metastatic breast cancer.

Author contributions

For this study, I established novel protocols for the collection, extraction and analysis of cfDNA, and provided expertise, training and technical advice to co-authors on this study to help establish collections and process samples. I established digital PCR and helped develop TAm-Seq which were critical in this study for the analysis of low frequency mutations. I provided critical review of the manuscript prior to publication. All author contributions are detailed in Appendix 3.

11. **Tsui DWY*, Murtaza M*, Wong ASC, Rueda OM, Smith CG, Chandrananda D, Soo RA, Lim H, Goh B, Caldas C, Forshew T, Gale D, Liu W, Morris J, Marass F, Eisen T, Chin T†, Rosenfeld N†; *EMBO Molecular Medicine*; 2018; e794; DOI 10.15252/ emmm. 201707945; 'Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in NSCLC.'**

Summary

In this study, tumour dynamics and resistance mechanisms were assessed in plasma from 50 *EGFR*-mutant NSCLC patients treated with gefitinib and hydroxychloroquine, using TAm-Seq and dPCR. Three cases who underwent histological transformation to SCLC were analysed by sWGS. Plasma *EGFR* mutations were detected in 95% cases with known tumour *EGFR* mutations. TAm-Seq identified additional mutations including T790M, *TP53*, *PIK3CA* and *PTEN* mutations. Patients with both *TP53* and *EGFR* mutations had worse overall survival compared to *EGFR*-mutant patients. Patients who progressed without T790M-mediated resistance had worse PFS during TKI continuation, and developed alternative mutations including *TP53* and SCLC-associated copy number changes.

Author contributions

I established protocols and lab infrastructure for the analysis of ctDNA, and trained the post-doc who led this study on these protocols, which were subsequently used for the analysis of plasma from non-small cell lung cancer patients in this publication. I provided scientific guidance throughout the study, and critically reviewed the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

12. **Mair R*, Mouliere F*, Smith CG, Chandrananda D, Gale D, Marass F, Tsui DWY, Massie CE, Wright AJ, Watts C, Rosenfeld N†, Brindle KM†; *Cancer Research;* 2019; 79 (1): 220 LP-230; https://doi.org/10.1158/0008-5472.CAN-18-0074; "Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models.'**

Summary

Human-derived ctDNA, tumour mitochondrial DNA (tmtDNA) and rat cfDNA was analysed in a patient-derived rat xenograft model of glioblastoma (n=64). Paired-end sWGS (<0.2x coverage) demonstrated that tumour-specific plasma and CSF DNA centred around 145bp in length, compared to host DNA at 167bp, and tmtDNA <100bp. tmtDNA was detected in 82% of plasma samples, and ctDNA in 24%, demonstrating that tmtDNA may be a more sensitive marker than ctDNA in GBM. Different mechanisms of ctDNA release were indicated in treatment-naïve rats and rats treated with temozolomide and radiotherapy, and levels were not affected by disruption of the blood brain barrier.

Author contributions

I was actively involved in the original concept and design of this study to analyse ctDNA in rat xenograft model of glioblastoma. I advised and trained the clinical fellow that led this study on ctDNA analysis, including advising on an optimised protocol for the collection of plasma from rat tail veins. I designed digital PCR assays for the analysis of the human ctDNA from GBM tumours, advised on appropriate rat assays to use to analyse host cfDNA, and performed digital PCR experiments and analysed dPCR data to assess levels of both rat and human DNA. I also advised on TAm-Seq experiments to further characterise the GBM tissue specimens that were implanted. I was involved in writing and critically reviewing the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

13. **Mouliere F*, Piskorz A*, Chandrananda D*, Moore E*, Morris J, Smith C, Goranova T, Heider K, Mair R, Supernat A, Gounaris I, Ros S, Wan J, Jimenez-Linan M, Gale D, Brindle K, Massie C, Parkinson C, Brenton JD†, Rosenfeld N†; *bioRxiv*; 2017; *(Pre-print, not peer-reviewed)*; 'Selecting Short DNA Fragments In Plasma Improves Detection Of Circulating Tumour DNA.'**

Summary

As clinical analysis of ctDNA can be limited by its low concentration, size selection was performed to determine whether this would enrich for tumour-specific fragments between 90-150bp. Plasma samples from 13 patients with recurrent high-grade serous ovarian cancer were collected before and during chemotherapy, and analysed by targeted and shallow whole genome sequencing following size selection. Size selection enabled detection of tumour alterations not previously detected, including a MYC amplification, and an up to 11-fold enrichment of mutated DNA. These studies provide a potential approach to overcome sensitivity limitations of ctDNA for early diagnosis and detection of minimal residual disease.

Author contributions

I established protocols for the collection, DNA extraction and analysis of ctDNA, including working with clinical teams to establish unique collections of plasma specimens from ovarian, melanoma and phase I trials, and provided optimised SOPs and advice to other groups collecting samples. The plasma cohorts provided a unique resource to analyse fragmentation patterns of cfDNA in this pan-cancer study. I trained the lead authors on ctDNA analysis techniques, and critically reviewed the manuscript ahead of submission to bioRxiv. All author contributions are detailed in Appendix 3.

14. **Mouliere F*, Chandrananda D*, Piskorz A*, Moore E*, Morris J, Barlebo Ahlborn L, Mair R, Goranova T, Marass F, Heider K, Wan J, Supernat A, Hudecova I, Gounaris I, Ros S, Jimenez-Linan M, Garcia-Corbacho J, Patel K, Østrup O, Murphy S, Eldridge M, Gale D, Stewart G, Burge J, Cooper W, van der Heijden M, Massie C, Watts C, Corrie P, Pacey S, Brindle K, Baird R, Mau-Sørensen M, Parkinson C, Smith C , Brenton JD†, Rosenfeld N†; *Science Translational Medicine*; 2018; Vol. 10, Issue 466, eaat4921; DOI: 10.1126/scitranslmed.aat4921; 'Enhanced detection of circulating tumor DNA by fragment size analysis'.**

Summary

ctDNA fragment sizes were analysed across 344 plasma samples from 200 patients with 18 different cancer types, and 65 healthy controls. Paired-end sWGS (<0.4x) demonstrated a lower proportion of fragments <150bp in healthy controls and in cancer-types known to have 'low ctDNA' levels (renal, GBM, bladder, pancreatic cancer), compared to those with 'high ctDNA' (melanoma, breast, ovarian, lung, colorectal, cholangiocarcinoma). Size selection of 90bp-150bp fragments resulted in a median >2-fold enrichment of tumour-specific DNA in >95% cases, and >4-fold enrichment in >10% cases. Supervised machine learning demonstrated that integration of SCNA data and different fragmentation features significantly improved ctDNA detection.

Author contributions

I established protocols for the collection, DNA extraction and analysis of ctDNA, including working with clinical teams to establish unique collections of plasma specimens from ovarian, melanoma and phase I trials, and provided optimised SOPs and advice to other groups collecting samples. The plasma cohorts provided a unique resource to analyse fragmentation patterns of cfDNA in this pan-cancer study. I trained the lead authors on ctDNA analysis techniques, and critically reviewed the manuscript ahead of publication. All author contributions are detailed in Appendix 3.

Appendix 2: Letters from senior authors detailing my contribution to each publication.

The following letters of support by senior authors detail the contributions I made to each co-authored publication, and the scientific impact of each of these papers.

Cancer Research UK Cambridge Institute
University of Cambridge
Li Ka Shing Centre
Robinson Way
Cambridge
CB2 0RE
Tel: 01223 769761

26 February 2019

To Whom It May Concern

Dear Sir/Madam

I am delighted to write in support of Davina Gale's application for award of PhD by Publication at the University of East Anglia, and to confirm Davina's significant involvement in the following publications:

1. 'Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study.'
   Parkinson C*, **Gale D*,** Piskorz A, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, Earl H, Qian W, Rosenfeld N‡, **Brenton JD‡**; **PLoS Medicine**; 2016; 13 (12): e1002198 doi:10.1371/journal. pmed.1002198.
2. 'Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis.'
   Schwarz R, Ng C, Cooke S, Newman S, Temple J, Piskorz A, **Gale D**, Sayal K, Murtaza M, Baldwin P, Rosenfeld N, Earl H, Sala E, Jimenez-Linan M, Parkinson C, Markowetz F‡, **Brenton JD‡**; **PLoS Medicine**; 2015; Feb 24;12(2): e1001789.
3. 'Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.'
   Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DW*, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, **Brenton JD‡**, Rosenfeld N‡; **Science Translational Medicine**; 2012; 4(136):136ra68.
4. 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.'
   Gale D*, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G, Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N; **PLoS ONE**; 2018; 13(3): e0194630.
5. 'Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.'

Murtaza M*, Dawson S-J*, Tsui DWY*, **Gale D**, Forshew T, Piskorz AM, Parkinson C, Chin S, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, **Brenton JD‡**, Caldas C‡, Rosenfeld N‡; **Nature**; 2013; 497(7447):108-12.

6. ‘Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patient’
Risberg B*, Tsui DWY*‡, Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, **Brenton JD**, **Gale D‡**; **The Journal of Molecular Diagnostics**, 2018; Volume 20 , Issue 6 , 883 – 892.
7. ‘Enhanced detection of circulating tumor DNA by fragment size analysis.’
Mouliere F*, Chandrananda D*, Piskorz AM*, Moore E*, Morris J, Barlebo Ahlborn L, Mair R, Goranova T, Marass F, Heider K, Wan J, Supernat A, Hudecova I, Gounaris I, Ros S, Jimenez-Linan M, Garcia-Corbacho J, Patel K, Østrup O, Murphy S, Eldridge M, **Gale D**, Stewart G, Burge J, Cooper W, van der Heijden M, Massie C, Watts C, Corrie P, Pacey S, Brindle K, Baird R, Mau-Sørensen M, Parkinson C, Smith C , **Brenton JD‡**, Rosenfeld N‡; **Science Translational Medicine**; 2018; Vol. 10, Issue 466, eaat4921; DOI: 10.1126/scitranslmed.aat4921.

I am a Senior Group Leader at the Cancer Research UK Cambridge Institute University of Cambridge and lead the Functional Genomics of Ovarian Cancer group; I am also an Honorary Consultant in Medical Oncology at Addenbrooke’s Hospital. I am a co-founder of Inivata Ltd., a clinical cancer genomics company using circulating tumour DNA (ctDNA) to improve personalised healthcare in oncology (with Nitzan Rosenfeld, Davina Gale, Tim Forshew).

I have worked closely with Davina for over 9 years dating from 2009, when she joined Dr. Nitzan Rosenfeld Molecular and Computational Diagnostics group at the Cancer Research UK Cambridge Institute.

Davina’s contributions are as follows:

1. Parkinson, Gale et al., PLoS Medicine, 2016

Davina established lab infrastructure, including novel protocols for the analysis of ctDNA, at a time when very little was known about ctDNA or its clinical relevance as a cancer diagnostic. Davina co-led this study with Christine Parkinson from my group to investigate mutant *TP53* ctDNA as a non-invasive biomarker for high-grade serous ovarian carcinoma (HGSOC). Davina initially determined optimal protocols for collection and extraction of plasma cell-free DNA from whole blood samples, and worked closely with my team in Addenbrooke’s Hospital to establish a unique plasma collection from patients attending the gynaecological oncology clinic. Davina next established digital PCR protocols using the Fluidigm microfluidic system to assess levels of mutant and wild-type *TP53* ctDNA in patient plasma. She designed and validated 31 unique patient-specific digital PCR assays to assess ctDNA levels in 318 plasma specimens from 40 HGSOC patients, in what I believe was the largest ctDNA study that had been performed at that time. In 2010, Davina’s experiments provided the first evidence that we were able to detect ctDNA in patient plasma. Following retrospective analysis, ctDNA levels were compared to serum biomarker CA-125 levels, and with volumetric analysis following radiographic imaging.

Davina was joint first co-author on this study and played a key role in the writing, editing and reviewing of the manuscript at every stage of it’s development. As senior co-author, I worked very closely with Davina during this process. We were able to demonstrate that response to chemotherapy was seen earlier with ctDNA than clinical marker CA-125. We also showed that the *TP53* mutant allele fraction (TP53MAF) correlated with tumour volume, and a decrease of pre-treatment TP53MAF of ≤60% after one cycle of chemotherapy was associated with better response. This was the first publication correlating ctDNA with volumetric imaging data and established the first cut point for early detection of response in HGSOC.

2. Schwarz et al., PLoS Medicine, 2015

Davina performed critical studies for this publication to confirm the allelic fraction of a subclonal population in a patient with HGSOC and progressive disease. She designed and validated digital PCR assays to assess the allelic fraction of a patient-specific *NF1* deletion by using *TP53* mutation to estimate tumour-specific contributions. This work provided strong evidence from spatially and temporally-separated tumour and ascites specimens that the clone had emerged over time but was present at diagnosis at very low levels. This work demonstrated for the first time that in HGSOC resistant subclones may be present prior to treatment and that increases in allele fraction during chemotherapy treatment may indicate emergence of adverse clonal populations. Davina used her extensive molecular biology expertise to design and validate these assays and to analyse the data. Davina was responsible for writing and reviewing the appropriate sections describing this work in our publication.

3. Forshew et al., Science Translational Medicine, 2012

At the time of this work, ctDNA studies had previously been performed using assays designed to analyse single hotspot mutations. Nitzan Rosenfeld's team developed TAm-Seq (tagged amplicon deep sequencing), a novel amplicon-based next-generation sequencing (NGS) method capable of analysing multiple mutations in parallel in ctDNA. Validation studies were performed on clinical plasma specimens from patients with HGSOC and metastatic breast cancer. Davina was involved in developing TAm-Seq, and also provided critical digital PCR validation data to assess its accuracy. This study was published in Science Translational Medicine in 2012, with Davina as joint first co-author, and myself as joint senior co-corresponding author, and demonstrated for the first time that NGS could be used to detect low frequency mutations in cell-free DNA and monitor multiple mutations in parallel. The study has had notable impact on subsequent widespread use of NGS for analysis of ctDNA.

4. Gale et al., PLoS ONE, 2018

Following the success of TAm-Seq, Dr. Nitzan Rosenfeld, Davina Gale, Tim Forshew and I co-founded Inivata Ltd, a spin-out company to make ctDNA assays accessible for clinical diagnostic use. Davina, as Head of Molecular Diagnostics, was instrumental in leading the development and validation of enhanced TAm-Seq (eTAm-Seq), performed to clinical regulatory standards, to analyse low-frequency mutations in cfDNA, with >20 increase in sensitivity compared to original TAm-Seq. These analytical validation studies, published in PLoS ONE with Davina as both first and senior co-corresponding author, demonstrated the sensitivity of the assay to detect low frequency mutations.

5. Murtaza et al., Nature, 2013

Davina was involved in the first exome sequencing analysis of ctDNA, published in Nature in 2013. This impactful publication demonstrated for the first time that NGS could be used to perform exome-wide profiling of plasma taken before and after development of treatment resistance to identify potential mechanisms of acquired resistance to therapy. Davina trained PhD and post-docs in Nitzan's and my group on relevant protocols and the collection of high quality specimens for ctDNA studies. She advised on the most suitable method for preparing exome libraries from low input amounts of cell-free DNA, which had been a key challenge for the study. In addition, she was involved in critically reviewing the manuscript, published in Nature in 2013.

6. Risberg et al., The Journal of Molecular Diagnostics, 2018
This study investigated pre-analytic factors that may affect the quality of plasma collected for cell-free DNA analysis. Different factors were assessed including delays in blood processing, storage temperatures, different blood collection tubes, centrifugation protocols and the effect of sample shipment. Davina was actively involved in the concept and design of this study, together with Dr Dana Tsui, a post-doc in Dr Nitzan Rosenfeld's lab, and she collaborated with a research assistant working in my clinic to collect appropriate samples for analysis. In addition, Davina designed and validated digital PCR and TAm-Seq assays used to assess levels of ctDNA and cell-free DNA in this study. Davina was senior co-corresponding author on this paper, and she led the writing and review, along with discussions with the editors at The Journal of Molecular Diagnostics. This paper will help others working in the field to assess appropriate methods for collection of high quality specimens for clinical ctDNA research.

7. Mouliere et al., Science Translational Medicine, 2018
In this paper, ctDNA fragment sizes were assessed in 344 plasma samples from 200 patients with cancer using low-pass whole-genome sequencing. We developed in vitro and in silico size selection methods and machine-learning algorithms to enhance sensitivity of detection of ctDNA, which may have important implications in early detection of cancer. Davina played an important role in establishing protocols for collection of high-quality plasma specimens used in this study, including helping establish collections of ovarian, melanoma and breast samples, which enabled this study to be performed. In addition, Davina critically reviewed the manuscript ahead of publication.

Our work on ctDNA along with others has catalysed a paradigm shift in the conduct of clinical trials and the possibilities of personalised treatment for patients. The vast majority of clinical trials in lung, colorectal and ovarian cancer are now incorporating ctDNA endpoints. It is not an exaggeration to say that liquid biopsies have transformed choices for patients and scientific insights into how patients respond to cancer treatments. Davina has been hugely instrumental in initiating and leading studies to quantitatively assess levels of ctDNA in patients.

I strongly support Davina's application for award of PhD by Publication at the University of East Anglia, and confirm here her very significant contributions in these publications which have been highly impactful in the field of liquid biopsy research. Please do not hesitate to contact me if you require further information.

Yours faithfully

James Brenton

Dr James D. Brenton PhD FRCP
Senior Group Leader and Honorary Consultant in Medical Oncology
Functional Genomics of Ovarian Cancer Laboratory
01223 769761 Assistant
james.brenton@cruk.cam.ac.uk
http://www.cruk.cam.ac.uk/research-groups/brenton-group

www.cuh.nhs.uk

NHS
Cambridge
University Hospitals
NHS Foundation Trust

Postgraduate Research Office University
of East Anglia Norwich Research Park
Norwich
NR4 7TJ

Oncology Department
Box 193
Addenbrooke's Hospital
Cambridge Biomedical Campus
Hills Road
Cambridge
CB2 0QQ

01223 217074 (Gynae Secretary)
paola.barbieri@addenbrookes.nhs.uk

15th March 2019

Dear Sir/Madam,

I have worked closely with Davina Gale over a number of years on a number of projects, most closely for the following paper published in 2016:

Parkinson CA, Gale D, Piskorz AM, et al. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. PLoS Med. 2016 Dec 20;13(12): PubMed PMID: 27997533.

Davina is joint first author with myself and I would like to confirm her contribution to this paper published in PLoS Med (Impact Factor 11.7).

This work investigates clinical significance of baseline circulating tumour DNA (ctDNA) levels in patients with ovarian cancer on chemotherapy treatment and also of changes in ctDNA whilst on treatment. We identify that baseline ctDNA are correlated significantly with tumour volume (assessment by the radiologists on volumetric CT). And following chemotherapy, we identified a threshold for a fall in ctDNA after 1 and 2 cycles that significantly predicts the time to progression. This paper when published was the largest series correlating ctDNA with tumour volume. The paper was also one of the very few pioneer papers to describe very early changes in ctDNA after just 1 or 2 cycles of treatment. The field has rapidly expanded since then with many groups looking at early changes in ctDNA as a biomarker of treatment response in cancer.

The project required very close working and discussions between the clinic and the laboratory. Davina played a key role at all stages of the study, starting initially with study design all the way through to the writing of the manuscript and submission. She coordinated the study and was the main scientist responsible for development of protocols for sample collection and blood processing, and for the development of the digital PCR (dPCR) assays. She carried out the majority of plasma dPCR assays herself (over 300 samples) and played a key role in the analysis and interpretation of the data.

I fully support her application for a PhD by Publication at the University of East Anglia.

Please do not hesitate to contact me for any further information if needed.

Yours sincerely

Dr Christine Parkinson
Consultant Medical Oncologist

UNIVERSITY OF
CAMBRIDGE

CANCER
RESEARCH
UK
CAMBRIDGE
INSTITUTE

CRUK Cambridge Institute
University of Cambridge
Li Ka Shing Centre
Robinson Way
Cambridge CB2 0RE

September 21st, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

I am delighted to write in support of Davina Gale's application for a PhD by Publication at the University of East Anglia. As part of that application I would like to confirm Davina's contribution to the following paper published in PLoS Medicine in 2018:

> Parkinson C*, **Gale D***, Piskorz A, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, Earl H, Qian W, Rosenfeld N, Brenton J.; 'Exploratory Analysis of *TP53* Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study.' *PLoS Medicine*; 2016; 13 (12): e1002198 doi:10.1371/journal. pmed.1002198

Davina is joint first author, and I am a senior co-corresponding author on this paper. This paper describes an extensive project which we carried out over several years, and constitutes an important advance in the field. In this study, we assessed the levels of circulating tumour DNA carrying *TP53* mutations in the plasma samples of patients with relapsed high-grade serous ovarian cancer (HGSOC), and the potential value of this measurement as a biomarkers of treatment response in those patients. Davina played a key leading role in designing and coordinating the study, and performed the laboratory work to measure circulating tumour DNA. In this multi-year project which Davina coordinated and conducted, Davina performed multiple stages of laboratory work including development of protocols for collection of plasma samples and extraction of DNA; development of multiple digital PCR (dPCR) assays to detect the mutant DNA; performing a large number of control experiments to assess the performance of those assays; and finally the digital PCR analysis of 318 plasma samples from 40 HGSOC patients. Additional data was obtained by our collaborators on levels of CA-125 and tumour burden as assessed by volumetric analysis. Davina played a key role in the analysis and interpretation of the data, and in the writing of the manuscript and preparation for publication.

When we compared these different metrics of clinical response and outcomes, we found that a rapid decrease in the levels of circulating tumour DNA in the plasma of patients as they started treatment by chemotherapy was a strong indicator of better prognosis. This paper has had important impact in providing a detailed comparison of circulating tumour DNA and tumour volumetric analysis in patients with high-grade serous ovarian cancer. To my knowledge this is the largest study to date describing the dynamics of circulating tumour DNA in the plasma of ovarian cancer patients. The paper, published in December 2016 has been cited 32 times according to Google Scholar, as was cited as one of the Top 50 most downloaded articles by PLoS Medicine in 2016.

If you require any further information, please do not hesitate to contact me.

Yours faithfully,

Dr. Nitzan Rosenfeld
Senior Group Leader
Cancer Research UK Cambridge Institute
University of Cambridge
Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE
Tel: 01223-769769, Fax: 01224-769510
Email: nitzan.rosenfeld@cruk.cam.ac.uk
Webpage: www.rosenfeldlab.org

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

7th August 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

It is my pleasure to write this reference for Davina Gale's PhD submission titled 'Investigating the diagnostic potential of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy: from research to clinic'.

Davina and I have worked together for approximately 8 years. In this time we have been authors on 6 papers together, 3 of which I have written separate letters of support for describing the research and Davina's role (Forshew* et al, 2012, Gale et al, 2018 and Plagnol et al, 2018).

I started working with Davina in 2010 as a Post Doc based at the Cancer Research UK Cambridge Institute. Our time working together has been without any question, the most exciting and productive of my career as we investigated the diagnostic potential of ctDNA. Due to the scale of the work it was very much a team effort of which Davina was a key part.

I have outlined in detail in the 3 associated letters Davina's contributions to, and the impacts of the 3 published studies for which Davina asked me to provide review.

Davina has made many important scientific contributions described throughout these letters but highlights include: leading the setup of both an academic then commercial lab to first explore then transition to clinic, methods for analysing cancer DNA in a patients blood stream. Planning then executing a large digital PCR study for assessing ctDNA in ovarian cancer patients which was also used to assess our NGS method. Planning then leading the development of our method from an academic assay to a much improved commercial test. She has also been critical in planning the validation of their performance as described in the 3 associated letters. Our first paper in this field has been cited over 600 times and since we first proved you can detect ctDNA through next generation sequencing, many different organisations have started developing methods for stratification, monitoring, residual disease detection and many other applications besides in many different types of cancer.

Davina is logical, highly organised and motivated. She has been critical to the 3 studies I describe and the many others which she has been part of. I strongly believe her contributions are deserving of a PhD and I would happy provide more information if helpful.

Yours Sincerely,

Dr Tim Forshew

Head of Science and Innovation at, and co-founder of Inivata Ltd.
Honorary Lecturer - UCL Cancer Institute

Forshew T*, Murtaza M*, Parkinson C*, **Gale D***, Tsui DW*, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N. 'Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.' **Science Translational Medicine**; 2012; 4(136):136ra68 **(* Joint first author)**

Gale D*, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N. 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.' **PLOS ONE**; 2018; 13(3): e0194630;
***First and co-corresponding senior author**

Plagnol V, Woodhouse S, Howarth K, Lensing S, Smith M, Epstein M, Madi M, Smalley S, Leroy C, Hinton J, De Kievit F, Musgrave-Brown E, Herd C, Neblett K, Brennan W, Dimitrov P, Campbell N, Morris C, Rosenfeld N, Clark J, **Gale D**, Platt J, Calaway J, Jones G, Forshew T. 'Analytical validation of a Next Generation Sequencing liquid biopsy assay for high sensitivity broad molecular profiling.' **PLoS ONE**; 2018; 13(3): e01938022018

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

7th August 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

It is my pleasure to write this letter of support for Davina Gale's PhD submission titled 'Investigating the diagnostic potential of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy: from research to clinic'.

In this letter I will outline the key findings of the following study as well as describing Davina's main contributions:

Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DW* et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Science Translational Medicine; 2012; 4(136):136ra68 (* Joint first authors)

It had been shown more than 15 years prior to the above study that some cancer mutations can be detected non-invasively through analysis of samples including blood plasma, urine, stool and sputum. All early studies however were limited to methods that assessed just individual mutations and typically with fairly low sensitivity. Whilst the detection of this DNA therefore raised a range of potentially powerful clinical applications, from early cancer detection through to monitoring cancer evolution, they were initially not practical due to the methods available.

To address this, we initiated the above study looking to apply next generation sequencing methods to analysing this cell free DNA (cfDNA). Prior to our work, next generation sequencing (NGS) methods typically required large volumes of DNA (>100ng DNA) and also were limited to detecting mutations at high allele fractions (>5% mutant).

We developed a PCR based method that amplified 6 genes including TP53, EGFR, BRAF, and KRAS in duplicate then enabled deep Illumina sequencing. We also developed a novel mutation calling algorithm that took advantage of this method. We first optimised and validated this method using dilution experiments and tumour DNA before progressing to analysing cfDNA. In order to analyse cell free DNA, we first developed patient specific digital PCR assays for a large series of ovarian cancer patients (in whom we had sequenced and found TP53 mutations in their tumours) and used these to assess whether we could detect known cancer mutations in the blood of each patient. We then used our sequencing method to test our ability to detect the same mutations in an unbiased way in 38 of these ovarian cancer patients. We showed we could detected the mutations we expected down to 2% allele fraction with 97% sensitivity. In plasma samples from one patient, we also identified a mutation in EGFR that had not previously been detected through tissue sequencing.

Finally we used this tagged-amplicon deep sequencing method (TAm-Seq) to analyse cfDNA in plasma samples collected throughout treatment from three women, showing the ability to monitor cancer non-invasively.

Due to the scale and complexity of this project, Davina and I were equal joint first authors along with 3 other people. Davina played a critical role throughout the project and contributed a number of important things. She first established the laboratory where this work was performed (before I joined). She then developed the methods and SOPs with which we collected, processed and extracted cfDNA along with methods and SOPs for analysis for example by digital PCR. A critical part of the success of this study was our ability to compare our NGS method with, what at the time was the largest digital PCR study of ctDNA in any tumour type to my knowledge. This was a very large piece of work and Davina guided the collection of the ovarian cancer patient samples, designed each of their personalised digital PCR assays then planned and performed the experiments validating these assays and assessing the patient samples. She then interpreted their results. Before this project started, Davina planned and performed early NGS experiments which this method was built upon then stayed involved with the NGS throughout. Finally, Davina contributed towards the drafting and editing of the manuscript.

The impact of this research has I believe been very significant. It has been cited now well over 600 times. To my knowledge this study was the first to show we can detect solid tumour mutations de novo in blood through NGS. It was also the first to show you could monitor patients cancer by sequencing cancer DNA in their blood and the first to show that if you develop patient specific assays this can be even more powerful.

This research was disseminated widely at the time including in The Telegraph, Daily Mail and the LA Times. Since then we have spun a company out based on this method (Inivata Ltd.) currently building tests to improve lung cancer patient care and many other people have entered this space, most notably a company called Grail's who have raised more than $1.5 billion since 2016 looking to used NGS analysis of ctDNA for early cancer detection. I would be very pleased to see Davina earn a PhD for her important contributions to this work.

Yours Sincerely,

Dr Tim Forshew

Head of Science and Innovation at, and co-founder of Inivata Ltd.
Honorary Lecturer - UCL Cancer Institute

tgen®
AN AFFILIATE OF City of Hope.

September 28, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Re: Contributions by Davina Gale as co-author of Forshew et al. Science Translational Medicine 2012

Forshew et al. described an approach for massively parallel sequencing of fragmented low-molecular weight circulating tumor-specific DNA in plasma to identify low-abundance somatic mutations in blood of cancer patients. I was co-first author of the manuscript and worked with Davina Gale at the Cancer Research UK Cambridge Institute. Davina is co-first author on the paper.

This was the first study to demonstrate next-generation sequencing could be used to directly identify, detect and quantify cancer-specific mutations in plasma DNA from cancer patients. Since it's publication, the study has been cited 684 times in ~6 years. As a result of this observation and additional results in the field that followed this work, noninvasive genotyping using plasma DNA sequencing is a reality in clinical oncology today. The particular approach we described in Forshew et al. has been developed commercially and Davina led that development as member of the founding team at Inivata (a liquid biopsy startup copmany).

Davina played a key role in the design and conduct of this study, in particular by helping identify the best time points for clinical sample collection, establishing best practices for sample processing and DNA extraction (drawing on her experience in these areas), developing necessary sample quality assessment assays, designing and optimizing dozens of digital PCR assays for patient-specific mutations, generating mutation quantification data using digital PCR, helping optimize the amplicon sequencing approach and implementing the amplicon sequencing approach for tumor and plasma samples. She directly contributed to the analysis of digital PCR data (results that served as gold standard in this manuscript) and it's comparison with amplicon sequencing results. She directly contributed to the writing and framing of this manuscript during numerous joint writing and revision sessions.

The contributions that Davina made to this paper drew on her unique background and experience and we would not have been able to execute this study without her enabling insights and help.

Please feel free to contact me if there is any further information I may provide.

Sincerely,

Murtaza

Muhammed Murtaza, MBBS, PhD
Assistant Professor, Co-Director, Center for Noninvasive Diagnostics, TGen
Assistant Professor of Medicine, Mayo Clinic
email: mmurtaza@tgen.org; phone: +1-602-343-8497

tgen®
AN AFFILIATE OF City of Hope

September 28, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Re: Contributions by Davina Gale as co-author of Murtaza et al. Nature 2013

Murtaza et al. described the use of low-input whole-exome sequencing of circulating tumor DNA (ctDNA) for non-invasive monitoring of clonal evolution in solid cancers and for identification of novel drivers of acquired therapeutic resistance. I led the study as co-first author of the manuscript and worked with Davina Gale at the Cancer Research UK Cambridge Institute.

This was a proof-of-principle study, demonstrating for the first time, that tracking clonal evolution of cancer was feasible using plasma DNA sequencing. Since it's publication, the study has been cited 967 times in ~5 years, highlighting the key role this paper has played in establishing the role of liquid biopsies in tracking cancer evolution and drug resistance.

Davina played an enabling role in the design of this study, in particular by identifying the most suitable patients and samples, based on data she generated using orthogonal technologies such as amplicon sequencing and digital PCR. She contributed to the conduct of this study, by guiding the evaluation of methods for sequencing library preparation using low amounts of DNA available in plasma samples, a key challenge we had to overcome and generating orthogonal data required for interpretation of our results. In addition, she developed and implemented pre-analytical processing steps that enabled generation of this data and it's reasonable interpretation. She contributed to the interpretation and framing of our results, based on her data and experience with ctDNA analysis in breast and ovarian cancer.

This study relied on longitudinal analysis of plasma samples from multiple cancer patients and the interpretation of our results was dependent on appropriate annotation and management of the samples involved. Longitudinal studies are logistically more challenging than single time point analyses. Davina's informed management of the study and her role in ensuring that these samples were collected, processed and stored in a timely, accessible and organized fashion was critical to the study's success. Without her oversight, it would have not been possible to conduct this study.

Please feel free to contact me if there is any further information I may provide.

Sincerely,

Murtaza

Muhammed Murtaza, MBBS, PhD
Assistant Professor, Co-Director, Center for Noninvasive Diagnostics, TGen
Assistant Professor of Medicine, Mayo Clinic
email: mmurtaza@tgen.org; phone: +1-602-343-8497

Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne Victoria
3000 Australia

Postal Address
Locked Bag 1 A'Beckett Street

Victoria 8006 Australia

Phone +61 3 8559 5000
Fax +61 3 8559 7379
ABN 42 100 504 883

petermac.org

Locations
Melbourne
Bendigo
Box Hill
Moorabbin
Sunshine

Peter Mac
Peter MacCallum Cancer Centre
Victoria Australia

Assoc. Prof. Sarah-Jane Dawson
Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne VIC 3000
Australia

20th May 2018

To whom it may concern,

Dear Sir/Madam,

I am writing to confirm Davina Gale's contribution to the following paper published in Nature in 2013, in support of her PhD by Publication at the University of East Anglia:

> Murtaza M*, **Dawson S-J.***, Tsui D.W.Y*, **Gale D**, Forshew T, Piskorz AM, Parkinson C, Chin SF, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD, Caldas C, Rosenfeld N. 'Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. 'Nature; 2013 May 2; 497(7447):108-12

I am joint first co-author of this publication, which was based on pioneering research performed at the Cancer Research UK Cambridge Institute between 2010 and 2013 to perform exome sequencing analysis of plasma DNA before and after treatment to identify potential mechanisms of acquired resistance to therapy in patients with advanced cancer.

Davina Gale's input was critical to the successful completion of this research. She played a fundamental role in developing the methods used in the analysis including digital PCR and TAm-Seq, which were not routinely established techniques in the field at this time. Importantly, she was specifically involved in the selection of the Rubicon protocol used for the exome hybrid capture which was the major component of the ctDNA analysis employed in this manuscript. She also established standardised operating procedures for clinical collection and processing of plasma samples, plasma DNA extractions and the downstream analysis of these samples which were all integral to the successful completion of this project. She provided critical input into the preparation of the final manuscript.

This publication has had a major impact in the field. It was the first proof-of-concept study to demonstrate the use of exome sequencing to study mechanisms of acquired resistance to therapy using circulating tumour DNA and helped establish a paradigm for the use of serial ctDNA analysis to track genomic mechanism of resistance in real-time. The influential nature of the manuscript is highlighted by its citation rate of 893 over the past 5 years.

Yours faithfully,

Dawson.

Associate Professor Sarah-Jane Dawson (MBBS, FRACCP, PhD)

Consultant Medical Oncologist
Group Leader & Head of Molecular Biomarkers and Translational Genomics Laboratory
Peter MacCallum Cancer Centre
Australia
Email: sarah-jane.dawson@petermac.org
Telephone: +61 3 8559 7132

tgen
AN AFFILIATE OF City of Hope

September 28, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Re: Contributions by Davina Gale as co-author of
Murtaza et al. Nature Communications 2015

Murtaza et al. tested the assumption that metastatic tumors at multiple sites in a patient shed DNA into the blood and that site-specific mutations can be detected directly in plasma (thereby overcoming tumor heterogeneity). We found that mutations shared by all tumor sites (ancestral mutations) were more likely to be detected in circulating tumor DNA (ctDNA) and observed at higher circulating levels, compared to mutations that were limited to a single tumor site. I was co-first author of the manuscript and worked with Davina Gale at the Cancer Research UK Cambridge Institute.

This study showed extensive comparison of ctDNA samples with a multi-regional analysis of tumors and informed the field's assessment of plasma-tumor concordance. The conclusions of this extensive but single patient case report were subsequently validated in larger studies. Since it's publication, the study has been cited 166 times in ~3 years.

Davina played a significant role in the design and conduct of this study, in particular by identifying the most suitable patients and samples, based on data she generated using orthogonal technologies such as amplicon sequencing and digital PCR, by guiding the evaluation of methods for sequencing library preparation using low amounts of DNA available in plasma samples, a key challenge we had to overcome and by generating orthogonal data required for interpretation of our results. In addition, she developed and implemented pre-analytical processing steps that enabled generation of this data and it's reasonable interpretation.

This study relied on longitudinal analysis of plasma samples from multiple cancer patients and the interpretation of our results was dependent on appropriate annotation and management of the samples involved. Longitudinal studies are logistically more challenging than single time point analyses. Davina's informed role in ensuring that these samples were collected, processed and stored in a timely, accessible and organized fashion was critical. Without her oversight, it would have not been possible to conduct this study.

Please feel free to contact me if there is any further information I may provide.

Sincerely,

Murtaza

Muhammed Murtaza, MBBS, PhD
Assistant Professor, Co-Director, Center for Noninvasive Diagnostics, TGen
Assistant Professor of Medicine, Mayo Clinic
email: mmurtaza@tgen.org; phone: +1-602-343-8497

Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne Victoria
3000 Australia

Postal Address
Locked Bag 1 A'Beckett Street

Victoria 8006 Australia

Phone +61 3 8559 5000
Fax +61 3 8559 7379
ABN 42 100 504 883

petermac.org

Locations
Melbourne
Bendigo
Box Hill
Moorabbin
Sunshine

Peter Mac
Peter MacCallum Cancer Centre
Victoria Australia

Assoc. Prof. Sarah-Jane Dawson
Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne VIC 3000
Australia

20th May 2018

To whom it may concern,

Dear Sir/Madam,

I am writing to confirm Davina Gale's contribution to the following paper published in Nature Communications in 2015, in support of her PhD by Publication at the University of East Anglia:

> Murtaza M*, **Dawson SJ***, Pogrebniak K, Rueda O, Provenzano E, Grant J, Chin SF, Tsui D, Marass F, Gale D, Ali HR, Shah P, Contente-Cuomo T, Farahani H, Shumansky K, Kingsbury Z, Humphray S, Bentley D, Shah S, Wallis M, Rosenfeld N, Caldas C. 'Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer'; *Nature Communications*; 2015; 6:8760 doi: 10.1038/ncomms9760

I am joint first co-author of this publication, which was based on comprehensive genomic profiling of 8 tissue and 9 plasma samples from a single patient with metastatic breast cancer, including analysis of autopsy specimens. This study was a proof-of-principal analysis highlighting that ctDNA can be used to reflect multi-focal clonal evolution, and study serial changes in stem and sub-clonal private mutations in plasma.

Davina Gale provided an important original contribution to this work. She played a key role in developing the methods used in the analysis including digital PCR, TAm-Seq and exome sequencing, and once established provided training and supervision for multiple laboratory members to execute this analysis effectively. She established standardised operating procedures for clinical collection and processing of plasma samples, plasma DNA extractions and the downstream analysis of these samples which were all integral to the successful completion of this project. She also provided critical input into the preparation of the final manuscript.

This publication has enabled substantial progress in our understanding of the relationship between ctDNA and tumour biopsy testing in clinical practice. It was the first study to demonstrate the ability of ctDNA to infer genomic heterogeneity and clonal heirarchy through comprehensive exome and targeted sequencing of longitudinal plasma and tumour samples. The influential nature of the manuscript is highlighted by its citation rate of 145 over the past 3 years.

Yours faithfully,

Dawson.

Associate Professor Sarah-Jane Dawson (MBBS, FRACCP, PhD)

Consultant Medical Oncologist
Group Leader & Head of Molecular Biomarkers and Translational Genomics Laboratory
Peter MacCallum Cancer Centre
Australia
Email: sarah-jane.dawson@petermac.org
Telephone: +61 3 8559 7132

Oslo
universitetssykehus

Oslo universitetssykehus HF
Postboks 4956 Nydalen
0424 Oslo

11th January 2019

To whom it may concern,

Dear Sir/Madam,
I am pleased to support Davina Gale's application for a PhD by Publication at the University of East Anglia, and confirm Davina's contribution to the following paper:

> 'Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patient'; ***The Journal of Molecular Diagnostics***, 2018; Volume 20, Issue 6, 883 – 892;
> **Risberg B***, Tsui DWY*, Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, Brenton JD, **Gale D**

Davina was senior co-corresponding author on this paper, and I was joint first author, together with Dr. Dana Tsui. Davina played an instrumental role in this paper. She established and designed this study together with Dana Tsui to collect plasma samples from a local oncology clinic at the Addenbrooke's Hospital from patients with high-grade serous ovarian cancer and metastatic breast cancer. She oversaw the study together with Dr Dana Tsui, providing important technical and scientific advice. Davina also provided digital PCR assays which she had designed and validated to quantitatively assess levels of tumour-specific DNA, to determine the effects of delayed processing and alternative processing methods. In addition, Davina played a key role in critically reviewing and revising the manuscript, and was the senior corresponding author with editors, leading to the publication of the paper.

Please do not hesitate to contact me if you require any further information.

Yours faithfully

Bente Risberg

Bente Risberg
MSc, BLS
Molecular Pathology
Department of Pathology
Oslo University Hospital, the Norwegian Radium Hospital
Oslo, NORWAY
Tel: +4793035961
Email: brp@ous-hf.no

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

September 21st, 2018

Dear Sir/Madam,

I am delighted to write in support of Davina Gale's application for a PhD by Publication at the University of East Anglia. As part of that application I would like to confirm Davina's contribution to the following paper, published in PLoS ONE in 2018:

> **Gale D**, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G, Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N.; 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.' PLoS ONE; 2018; 13(3): e0194630;

Davina is first author and senior/co-corresponding author, and I am the other co-corresponding author on this paper. This paper describes the development of the enhanced Tagged-Amplicon deep-Sequencing technology, a novel ultrasensitive amplicon-based technology for next-generation sequencing. This technology was used to develop Inivata's InVisionFirst-Lung™ assay to detect low frequency mutations in plasma samples from patients with non-small cell lung cancer.

The enhanced Tagged-Amplicon deep-Sequencing (eTAm-Seq) technology is a development of the TAm-Seq technology, originally developed by the Rosenfeld Lab at the Cancer Research UK Cambridge Institute. As co-founder and Head of Molecular Diagnostics from start-up at Inivata Ltd, Davina played a critical and leading role in the technology transfer of the TAm-Seq technology from the academic lab to Inivata, in the design and development of the new enhanced TAm-Seq technology and its evaluation as described in this paper. Davina led the writing of this manuscript and the correspondence with the journal editorial office.

The assay described in this paper was designed to identify clinically-relevant somatic mutations in a panel of 35 cancer-related genes. The study performed by Davina's team demonstrated that the assay could detect mutant alleles down to 0.02% allele frequency (AF), with approximately 90% of mutations detected at 0.25%-0.33% AF, and per-base specificity of 99.9997%. These results confirmed that this enhanced TAm-Seq technology had much greater sensitivity than the original TAm-Seq assay developed in the Rosenfeld lab. These results provided important data to support implementation of assay for clinical diagnostic use.

If you require any further information, please do not hesitate to contact me.

Yours faithfully,

Dr. Nitzan Rosenfeld
Chief Scientific Officer
Inivata Ltd
The Portway Building
Granta Park
Cambridge
CB21 6GS
United Kingdom
Tel: +44 (0)1223 790 880
Email: nitzan.rosenfeld@inivata.com

Inivata™

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

Dear Sir/Madam,

It is a pleasure to provide a letter in support of Davina Gale's unique contribution to the article Gale et al, 2018 "Development of a highly sensitive liquid biopsy platform to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA" PLOS ONE.

Davina made significant contributions to all aspects of this paper and was the driving force behind the development and implementation of eTAm-Seq™ for clinical use, which allows the identification of somatic mutations at low frequency across multiple genes. Davina provided all day to day line management and project oversight for the laboratory team responsible for carrying out the technology development. This included hands on wet lab training as well as experimental design, which Davina was able to provide following many years developing her expertise in molecular biology, cancer and ctDNA. In order to be suitable for clinical use, Davina instigated the automation of this technology so that it could be streamlined across multiple laboratories and effective on a larger scale. Davina led the design of a 2-site validation and managed the operators responsible for carrying out this process. Her attention to detail was fundamental to ensure all appropriate controls and established standards were considered and the process conducted within the framework of CLIA (U.S. site) and ISO15189:2012 quality standards (U.K. site).

Davina was part of a small team of scientists and bioinformaticians responsible for the data analysis. The assay demonstrated excellent sensitivity down to 0.02% allele fraction and a high base pair specificity of 99.997% across both sites.

The preparation of work for publication was led by Davina, including drafting and subsequent revisions of the manuscript.

This paper is the foundation of the InVision™ platform, and the assay has been used in many clinical research studies throughout the U.K., Europe and the U.S.A. Many non-small cell lung cancer (NSCLC) patients, for whom tissue is not available, have had the benefit of targeted therapy as a result of this assay.

Do not hesitate to contact me if you require further information.

Yours faithfully,

KHowarth

Karen Howarth
Head of Cancer Genomics
Email: karen.howarth@inivata.com
Tel: 01223 790881

Invivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

7th August 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

It is my pleasure to write this letter of support for Davina Gale's PhD submission titled 'Investigating the diagnostic potential of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy: from research to clinic'.

In this letter I will outline the key findings of the following study as well as describing Davina's main contributions:

Gale D, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N. 'Development of a highly sensitive liquid biopsy platform utilizing enhanced Tagged-Amplicon deep-Sequencing technology to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA.' PLOS ONE; 2018; 13(3)

As outlined in my letter of support describing the Forshew et al study of 2012, to our knowledge we were the first to show we could detect solid tumour mutations de novo through sequencing cell free DNA and the first to show you could use this to monitor a patient's cancer. The paper received very broad interest (over 600 citations so far) and I am confident it played some part in initiating many different efforts to bring cfDNA analysis to clinic. As one such effort, Davina, Nitzan Rosenfeld, James Brenton and I, co-founded a company called Inivata. The aim was to take the methods we had developed, refine them for clinical use then offer them to patients. Whilst our method published in 2012 was cutting edge at the time, an optimal assay for clinical use needed many very significant improvements. The assay needed to become more robust, higher throughput, significantly more sensitive and specific and it needed to be able to detect many more variants. The challenges of running an assay in a clinical as opposed to a research setting are also very significant. This paper describes the development of our first assay and validation of the performance of the assay. Through this we demonstrated that we have built a method with the highest levels of sensitivity for the key clinically actionable point mutations and indels.

For the first few years of Inivata's existence I worked for Inivata part time providing specific expertise in our NGS method and molecular approaches for optimising the assay. Davina was the only one of the co-founders to work full time in establishing Inivata. As a rapidly growing startup (we went from just 4 of us to 70 people in just 4 years) Davina and I both had many different hats working in various aspects of Inivata from intellectual property through to business development. Perhaps Davina's biggest roles however, and the roles that contributed directly to her being the first author on this paper include the following though:

She first led the team that built a clinical lab from scratch in Cambridge UK, then initiated a second clinical lab in North Carolina (USA). She also led the same group with my NGS support in planning then building our first assay. As described above, the assay changed significantly from the initial academic version so this development work was extremely complex. This covered all aspects from optimising, designing new primers, changing chemistry and bringing in initial automation through to developing extremely detailed SOPs describing how the assay must work. Key to this was building a strong team which Davina drove with my support in interviewing. Finally she was an important part of this analytical validation and we designed a number of aspects of the validations together. Davina wrote this manuscript with some input on specific details from myself and others.

In this study we were able to demonstrate that we had taken our assay from being able to analyse 6 genes to now being able to analyse 35. We also showed we had increased our sensitivity about 10 fold from 2% down to ~0.25% allele fraction.

This and the newer version of this assay described in Plagnol et al have now been run on a large number of patients. The findings of these blood tests have been both presented at meetings and published (e.g. Remon J, et al. Ann Oncol. 2017) further enhancing our collective understanding of what ctDNA can be used for. Clearly the most significant thing though is that I am now aware of many patients whose lung cancer treatment has been changed and improved based on the results of this test (see Remon J, et al as example). I would be very pleased to see Davina being awarded a PhD for her role in this and her many other pieces of work in the ctDNA space.

Yours Sincerely,

Dr Tim Forshew

Head of Science and Innovation at, and co-founder of Inivata Ltd.
Honorary Lecturer - UCL Cancer Institute

Invata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

7th August 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

It is my pleasure to write this letter of support for Davina Gale's PhD submission titled 'Investigating the diagnostic potential of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy: from research to clinic'.

In this letter I will outline the key findings of the following study as well as describing Davina's main contributions:

Plagnol V, Woodhouse S, Howarth K, Lensing S, Smith M, Epstein M, Madi M, Smalley S, Leroy C, Hinton J, De Kievit F, Musgrave-Brown E, Herd C, Neblett K, Brennan W, Dimitrov P, Campbell N, Morris C, Rosenfeld N, Clark J, **Gale D**, Platt J, Calaway J, Jones G, Forshew T. 'Analytical validation of a Next Generation Sequencing liquid biopsy assay for high sensitivity broad molecular profiling.' PLoS ONE; 2018; 13(3)

As outlined in my letters of support describing the Forshew et al study of 2012 and the Gale et al study of 2018, we were the first to our knowledge to demonstrate you can detect solid tumour mutations through sequencing cell free DNA and we then developed and validated a clinical assay to bring this approach to patients.

In this study we performed the complete analytical validation of what is our current assay. Following the completion of the assay described in the Gale et al paper, we updated the assay adding the ability to detect fusion genes and we enhanced many aspects of the assay such as further improvements in how we detect point mutations. We also fully automated the process.

As described in the letter for the Gale et al paper, Davina played central part in the development of this assay. She first led the team that built a clinical lab from scratch in Cambridge UK, then initiated a second clinical lab in North Carolina (USA). She also led the same group with my NGS support in planning then building our first assay. This covered all aspects from optimising and bringing in initial automation through to developing extremely detailed SOPs describing how the assay must work. Key to this was building a strong team which Davina drove with my support in interviewing. This led to this analytical validation.

One key aspect of the assay described in this paper was its automation. This enabled us to scale the assay and allowed significant improvements in robustness. The automation work was led by a very talented scientist whom Davina was line managing. Davina was also important in the planning phases of the validation guiding us on critical components of this validation. Finally, Davina helped with critically reviewing and editing the manuscript.

The current assay is designed to enable stratification of patients with non small cell lung cancer on to the most appropriate therapy. This assay is already improving the care of many patients and I believe it would be fitting for Davina's scientific contribution to be recognised with a PhD. Please do let me know if I can provide any further information.

Yours Sincerely,

Dr Tim Forshew

Head of Science and Innovation at, and co-founder of Inivata Ltd.
Honorary Lecturer - UCL Cancer Institute

Inivata™

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

Dear Sir/Madam,

I am delighted to provide a letter in support of Davina Gale's contribution to the Inivata publication Plagnol et al 2018 "Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling" PLOS ONE.

In this study, the eTAm-Seq™ technology was significantly developed and enhanced to produce the InVisionFirst™ assay that is now being used in many clinical research studies and is available as a commercial product in the U.S.A. This assay was developed to detect point mutations, indels, amplifications and gene fusions that commonly occur in NSCLC.

Davina contributed to the design and continued development of this assay and her knowledge in this competitive field was essential to its successful implementation. Davina provided input during the set-up and analysis, reviewed the analytical validation and provided feedback on the figures and text included in the final publication.

Data presented in this paper was submitted as part of the technical assessment presented to Palmetto GBA, a Medicare Administrative Contractor that evaluates diagnostic technology through its laboratory technology assessment group, MolDx. In September 2018, a draft local coverage determination (LCD) for the InVisionFirst-Lung™ assay was published. This was a significant step in making the power of liquid biopsy a realisation for many patients.

Do not hesitate to contact me if you require further information.

Yours faithfully,

KHowarth

Karen Howarth
Head of Cancer Genomics
Email: karen.howarth@inivata.com
Tel: 01223 790881

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

September 21st, 2018

Dear Sir/Madam,

I am delighted to write in support of Davina Gale's application for a PhD by Publication at the University of East Anglia. As part of that application I would like to confirm Davina's contribution to the following paper, published in Annals of Oncology in 2017:

> Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, Howarth K, **Gale D**, Green E, Plagnol V, Rosenfeld N, Planchard D, Bluthgen MV, Gazzah A, Pannet C, Nicotra C, Auclin E, Soria JC, Besse B; Annals of Oncology; 2017; 28:784-790 'Osimertinib benefit in EGFR-mutant NSCLC patients with T790M mutation detected by circulating tumour DNA.'

This pioneering paper demonstrated for the first time the outcomes for patients with non-small cell lung cancer, for whom a biopsy sample was not available, who were treated by the 3rd-generation EGFR inhibitor osimertinib, based on minimally-invasive detection of the *EGFR* T790M mutation in the plasma of those patients. Similar to prior reports of studies of osimertininb response based on tissue analysis, the resistance mutation was detected in 50% patients, and treatment with osimertinib in those patients led to 'partial response' for 62.5% of patients and to 'stable disease' for 37.5% of patients. This data helped support the implementation of liquid biopsy testing for patients with non-small cell lung cancer.

The detection of the *EGFR* T790M mutation in the plasma samples was achieved by an enhanced TAm-Seq assay, developed by Inivata Ltd. As co-founder and Head of Molecular Diagnostics from start-up at Inivata Ltd, Davina played a critical and leading role in the technology transfer of the TAm-Seq technology from the academic lab to Inivata, in the development of the enhanced TAm-Seq (eTAm-Seq) technology, and in its implementation for testing of clinical patient samples for research. Davina participated in discussions with the team at Gustave Roussy (as part of her role in the academic research lab), led the team processing clinical samples, and was involved in reviewing and editing the manuscript for publication. The paper, published in January 2017 has been cited more than 50 times according to Google Scholar.

If you require any further information, please do not hesitate to contact me.

Yours faithfully,

Dr. Nitzan Rosenfeld
Chief Scientific Officer
Inivata Ltd
The Portway Building
Granta Park
Cambridge
CB21 6GS
United Kingdom
Tel: +44 (0)1223 790 880
Email: nitzan.rosenfeld@inivata.com

IOMD
Instituto de Oncologia
Medica Delfos

Postgraduate Research Office

University of East Anglia

Norwich Research Park

Norwich, NR4 7TJ, UK

Barcelona 21 September 2018

Dear Sir/Madam,

I am writing to confirm Davina Gale's contribution to the following paper published in Annals of Oncology in 2017:

> Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, Howarth K, **Gale D**, Green E, Plagnol V, Rosenfeld N, Planchard D, Bluthgen MV, Gazzah A, Pannet C, Nicotra C, Auclin E, Soria JC, Besse B; Annals of Oncology; 2017; 28:784-790 'Osimertinib benefit in EGFR-mutant NSCLC patients with T790M mutation detected by circulating tumour DNA.'

This important paper was the first published prospective study to assess the presence of *EGFR* T790M resistance mutations using a liquid biopsy to guide treatment with osimertinib in non-small cell lung cancer patients with no available tissue biopsy. The resistance mutation was detected in 50% of patients (equal than data reported in tissue rebiopsy), and treatment with osimertinib, a third generation tyrosine kinase inhibitor, led to good response rates, with 62.5% with a Partial Response and 6 months progression free survival of 67%. This outcome mirrors the efficacy of osimertinib in this population according patients' selection by *T790M* positivity in a new tissue biopsy. This paper provided prospective evidence that the liquid biopsy test reported similar rate of detection to tissue biopsy testing, and provided critical clinical validation data to help support clinical implementation of the assay.

Davina Gale, as co-founder and Head of Molecular Diagnostics from start-up at Inivata Ltd. led the development of enhanced TAm-Seq (eTAm-Seq) technology and led the team processing the clinical samples for this paper. She was also involved in reviewing and editing the manuscript for publication.

If you require any further information, please do not hesitate to contact me .

Yours faithfully,

Dr. Jordi Remon

Medical Oncologist

Centro Integral Oncología Clara Campal (CIOCC) Barcelona – HM Delfos

Barcelona

Spain

Email: jordi.remon@delfos.cat. Tel: 0034 93 254 50 00

Inivata™

Inivata Limited
The Portway
Granta Park
Great Abington
Cambridge
CB21 6GS

Dear Sir/Madam,

As a co-author, I am happy to provide a letter of support detailing Davina Gale's contribution to the publication Remon et al 2017 "Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA" Ann Oncol.

This study was a research collaboration between Institute Gustave Roussy and Inivata to illustrate the utility of liquid biopsy for the personalised treatment of lung cancer patients in a real-world setting.

Davina is an expert in ctDNA analysis and provided knowledge of how the InVision™ assay could be utilised during the initial design of this investigation. Subsequently, she was instrumental in the implementation of this study by ensuring processes were in place for the accessioning, processing and analysis of plasma samples. Davina managed the team responsible for running the first clinical samples at Inivata, including those described in this manuscript. She provided oversight and trouble-shooting of results as necessary. Davina was actively involved in the analysis of these results, preparing data as required for presentation in the manuscript.

This work supports the use of liquid biopsies for personalised treatment in advanced NSCLC and is the first study to our knowledge to show the possible clinical significance of detecting T790M at very low allele fractions in plasma.

Do not hesitate to contact me if you require further information.

Yours faithfully,

KHowarth

Karen Howarth
Head of Cancer Genomics
Email: karen.howarth@inivata.com
Tel: 01223 790881

Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne Victoria
3000 Australia

Postal Address
Locked Bag 1 A'Beckett Street

Victoria 8006 Australia

Phone +61 3 8559 5000
Fax +61 3 8559 7379
ABN 42 100 504 883

petermac.org

Locations
Melbourne
Bendigo
Box Hill
Moorabbin
Sunshine

Peter Mac
Peter MacCallum Cancer Centre
Victoria Australia

Assoc. Prof. Sarah-Jane Dawson
Peter MacCallum Cancer Centre
305 Grattan Street
Melbourne VIC 3000
Australia

20th May 2018

To whom it may concern,

Dear Sir/Madam,

I am writing to confirm Davina Gale's contribution to the following paper published in the New England Journal of Medicine in 2013, in support of her PhD by Publication at the University of East Anglia:

> **Dawson S-J*,** Tsui D.W.Y*, Murtaza M, Biggs H, Rueda O, Chin SF, Dunning M, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N; 'Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer'; *New England Journal of Medicine*; 2013; DOI: 10.1056/NEJMoa121326

I am first co-author of this publication, which was based on innovative research performed at the Cancer Research UK Cambridge Institute between 2010 and 2013 to assess the clinical utility of circulating tumour DNA (ctDNA) and compare the performance of ctDNA testing to other biomarkers including circulating tumour cells, CA 15-3 and radiographic imaging in metastatic breast cancer patients. The paper used tagged amplicon deep sequencing (TAm-Seq), whole genome sequencing and personalised digital PCR assays to identify somatic mutations and quantify ctDNA in serially collected plasma specimens. This publication provided the first demonstration that ctDNA revealed the earliest measure of treatment response and showed a greater dynamic range and correlation with tumour burden than the other circulating biomarkers.

Davina Gale provided an instrumental contribution to this work. She played a fundamental role in developing the methods used in the analysis including digital PCR and TAm-Seq, which were not routinely established techniques in the field at this time. She established standardised operating procedures for clinical collection and processing of plasma samples, plasma DNA extractions and the downstream analysis of these samples which were all integral to the successful completion of this project. She provided critical input into the preparation of the final manuscript.

This publication has been highly influential in the field. It was the first study to show the superiority of circulating tumour DNA to circulating tumour cells for monitoring tumour dynamics in breast cancer and provided the first demonstration of the role of ctDNA for molecular disease monitoring in this disease. The influential nature of the manuscript is highlighted by its citation rate of 1053 over the past 5 years.

Yours faithfully,

Dawson.

Associate Professor Sarah-Jane Dawson (MBBS, FRACCP, PhD)

Consultant Medical Oncologist
Group Leader & Head of Molecular Biomarkers and Translational Genomics Laboratory
Peter MacCallum Cancer Centre
Australia
Email: sarah-jane.dawson@petermac.org
Telephone: + 61 3 8559 7132

tgen
AN AFFILIATE OF City of Hope

September 28, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Re: Contributions by Davina Gale as co-author of
Dawson et al. New England Journal of Medicine 2013

Dawson et al. showed that changes in circulating tumor DNA (ctDNA) levels during treatment reflected tumor response and disease progression in patients with metastatic breast cancer. I was co-author of the manuscript and worked with Davina Gale at the Cancer Research UK Cambridge Institute.

This was the first study to demonstrate that ctDNA levels reliably decreased in patients who responded to systemic therapy and that ctDNA levels increased before or at the time that patients progressed on these treatments. In addition, the paper compared ctDNA levels with other known biomarkers such as the number of circulating tumor cells or CA15-3 levels. Since it's publication, the study has been cited 1143 times in ~5 years. This paper cemented the confidence of cancer researchers and oncologists in ctDNA analysis as a reliable and quantitative biomarker.

Davina played a key role in the design and conduct of this study, in particular by helping identify the best time points for clinical sample collection, establishing best practices for sample processing and DNA extraction (drawing on her experience in these areas), developing necessary sample quality assessment assays, designing and optimizing dozens of digital PCR assays for patient-specific mutations and generating mutation quantification data using digital PCR. She directly contributed to the analysis of digital PCR data. She helped interpret the digital PCR results and present the results in this manuscript.

This study was one of the first papers to show quantitative changes in serial samples during treatment. The interpretation of our results was dependent on appropriate annotation and management of the samples involved. Longitudinal studies are logistically more challenging than single time point analyses, any sample mix-ups can lead to mis-interpretation of the results. Davina played an informed and critical role in ensuring that these samples were collected, processed, annotated, stored and retrieved in a timely, accessible and organized fashion. I have no doubt that without her oversight, it would have not been possible to conduct this study.

Please feel free to contact me if there is any further information I may provide.

Sincerely,

Murtaza

Muhammed Murtaza, MBBS, PhD
Assistant Professor, Co-Director, Center for Noninvasive Diagnostics, TGen
Assistant Professor of Medicine, Mayo Clinic
email: mmurtaza@tgen.org; phone: +1-602-343-8497

TGen | 445 North Fifth Street, Suite 600 | Phoenix, Arizona 85004
Main 602.343.8400 | Toll Free 1.866.370.8436 | Fax 602.343.8440 | tgen.org

Memorial Sloan Kettering
Cancer Center

January 21, 2019

Re: Letter of Support

Dear Sir/Madam,

I am delighted to have the opportunity to support and recommend Davina Gale in her PhD by Publication. I have known Davina since 2010, when I joined the Cancer Research UK Cambridge Institute as a postdoctoral research associate in Dr. Nitzan Rosenfeld's lab. In her role as Lab Manager and first employee of Nitzan's group, Davina was instrumental in establishing the groups protocols and lab infrastructure for the analysis of circulating tumour DNA (ctDNA) as a non-invasive liquid biopsy. She trained all members of the group, and provided key scientific guidance. Together we have published several key publications, including in Science Translational Medicine, Nature, Nature Communications, New England Journal of Medicine, The Journal of Molecular Diagnostics, and EMBO Molecular Medicine.

In this letter, I confirm Davina's contributions to the following publications:

Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DWY* et al.; Science Translational Medicine, 2012; 'Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA.'

In this paper, we published for the first time the ability to used next generation sequencing (NGS) analysis to analyse cell-free DNA for the detection of tumour-specific mutations in plasma from cancer patients. We developed TAm-Seq (tagged amplicon deep sequencing) and demonstrated that it could be used to monitor ctDNA during patient treatment to help understand tumour dynamics. We are both joint first co-authors on this publication. Davina played an instrumental role, both in establishing critical digital PCR data to quantify mutant TP53 ctDNA levels in plasma from ovarian cancer patients to help determine the quantitative accuracy of TAm-Seq, and in working to establish and develop TAm-Seq protocols and review data. Davina also helped write and critically review the manuscript ahead of publication. This publication had critical impact in the field, as it provided the first demonstration of the use of NGS to assay multiple mutations. In her role as Head of Molecular Diagnostics and co-founder of Inivata, Davina led further development of eTAm-Seq technology to develop ultrasensitive assays for clinical use, which is now being used for patient benefit.

Murtaza M*, Dawson S.-J*., Tsui D.W.Y*, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin S, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD, Caldas C, Rosenfeld N.; Nature; 2013; 'Non-invasive analysis of acquired

resistance to cancer therapy by sequencing of plasma DNA.'

In this publication, we were able to demonstrate for the first time that exome sequencing could be used to non-invasively analyse ctDNA before and after development of acquired resistance to cancer therapy. I am a joint first author on this paper. Davina played a critical role in this paper, in establishing protocols, sample collections and training lab team members, advising on the most appropriate method for exome sequencing of low input amounts of DNA, and critically reviewing the manuscript. This publication provided important proof-of-concept of the ability to use NGS and hybrid capture to non-invasively assess ctDNA on an exome-wide scale. A follow-up publication was published in Nature Communications **- Murtaza et al., Nature Communications, 2015, 'Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer'.** Davina and I are co-authors on this paper, which demonstrated the ability to use these methods to analyse ctDNA in depth in multiple specimens from the same patient, including at autopsy, to further understand tumour heterogeneity and evolution and the representation of ctDNA in plasma.

Dawson SJ*, Tsui D.W.Y*, Murtaza M, Biggs H, Rueda O, Chin SF, Dunning M, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N.; *New England Journal of Medicine*; 2013; 'Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer.'

I am joint first co-author on this publication, which provided the first demonstration of comparison of ctDNA with CTCs, CA 15-3 and imaging data, to show that ctDNA has clinical value in monitoring patients with metastatic breast cancer. Davina played an important role in this paper, developing important protocols, helping to establish high-quality clinical collections for analysis, and providing training. She also helped critically review the manuscript ahead of publication.

Risberg B*, Tsui DWY*, Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, Brenton JD, Gale D$^{\&}$; The Journal of Molecular Diagnostics; 2018; The Journal of Molecular Diagnostics, 2018; 'Effects of collection and processing procedures on plasma circulating cell-free DNA from cancer patient'

Davina and I are senior co-corresponding authors on this publication. Davina played an instrumental role in this paper. In 2010, she initiated studies to assess the performance of Streck Cell-Free DNA BCT tubes to enable delayed and centralised processing of plasma for liquid biopsy studies, and limit pre-analytic factors which may affect downstream analysis of circulating tumour DNA (ctDNA). Streck BCT tubes contain a cell preservative which stablilises nucleated blood cells, limiting release of contaminating genomic DNA. Following these studies, we worked together to design and establish this study to collect plasma samples from patients with ovarian and metastatic breast cancer, and process specimens in different ways. We worked together to oversee the project, and Davina provided important scientific advice on this study, including digital PCR assays, which she had designed and validated to assess mutant ctDNA levels. Davina played a vital role in getting the manuscript published, providing critical review of the manuscript and leading discussions with the journal editors to improve the manuscript prior to publication.

Tsui DWT*, Murtaza M*, Wong ASC, Rueda OM, Smith CG, Chandrananda D, Soo RA, Lim H, Goh B, Caldas C, Forshew T, Gale D, Liu W, Morris J, Marass F, Eisen T, Chin T, Rosenfeld N; EMBO Molecular Medicine; 2018; 'Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in NSCLC.'

I am the first co-author on this publication, which demonstrates the ability to assess multiple resistance mechanisms in *EGFR*-mutant non-small cell lung cancer (NSCLC) patients treated with gefitinib and hydroxychloroquine. Davina played a critical role in establishing digital PCR and TAm-Seq protocols used in this study, and provided important training and scientific guidance throughout my time at the Cancer Research UK Cambridge Institute.

Please feel free to contact me if you require any further information.

Sincerely,

Dana W. Y. Tsui, Ph.D.
Assistant Attending
Faculty, Department of Pathology
Member, Center for Molecular Oncology
tsuiw@mskcc.org
+1-646-888-2079

Memorial Sloan Kettering Cancer Center
1275 York Avenue, Box 20, New York, NY 10065, USA

UNIVERSITY OF CAMBRIDGE
CANCER RESEARCH UK
CAMBRIDGE INSTITUTE

CRUK Cambridge Institute
University of Cambridge
Li Ka Shing Centre
Robinson Way
Cambridge CB2 0RE

September 21st, 2018

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

I am delighted to write in support of Davina Gale's application for a PhD by Publication at the University of East Anglia. As part of that application I would like to confirm Davina's contribution to the following paper published in EMBO Molecular Medicine in 2018:

> Tsui DWT, Murtaza M, Wong ASC, Rueda OM, Smith CG, Chandrananda D, Soo RA, Lim H, Goh B, Caldas C, Forshew T, **Gale D**, Liu W, Morris J, Marass F, Eisen T, Chin T, Rosenfeld N; EMBO Molecular Medicine; 2018; 'Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in NSCLC.'; EMBO Mol Med; 2018; e794; DOI 10.15252/emmm.201707945;

I am senior co-corresponding author on this paper, which describes the analysis of 50 non-small lung cancer patients treated with gefitinib and hydroxychloroquine, using TAm-Seq and digital PCR. Three cases who underwent histological transformation to small cell lung cancer were analysed by shallow whole genome sequencing. The results demonstrated that *EGFR* mutations were detected in plasma in 95% of cases with known tumour *EGFR* mutations, indicating that plasma has potential to provide clinically relevant information to guide treatment decisions. Pre-treatment levels of *EGFR* mutations were prognostic and correlated with tumour burden. Analysis by TAm-Seq identified additional mutations including *EGFR* T790M, *TP53*, *PIK3CA* and *PTEN* mutations, demonstrating the importance of using a next-generation sequencing approach to analyse multiple mutations to study tumour dynamics and evolution. Patients with concomitant *TP53* and *EGFR* mutations had worse overall survival compared to *EGFR*-only mutant patients, and patients who progressed with no detectable levels of the EGFR T790M mutation had poorer prognosis.

Davina joined my group as Lab Manager in 2009, and played a critical role in establishing from the outset the lab capabilities and infrastructure which was important to enable us to perform such a study. This included the development of protocols for analysis of circulating tumour DNA, extraction of DNA from plasma, and establishing Pre- and Post-PCR facilities within the Institute. She played a critical role in the development of digital PCR and TAm-Seq for analysis of low frequency mutations in plasma. In addition, she trained new lab members and provided technical advice to PhD and post-doctoral students within my group, including Dr. Dana Tsui who is lead author of this paper. She also reviewed the data and manuscript in preparation for publication.

If you require any further information, please do not hesitate to contact me.

Yours faithfully,

Dr. Nitzan Rosenfeld
Senior Group Leader
Cancer Research UK Cambridge Institute
University of Cambridge
Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE
Tel: 01223-769769, Fax: 01224-769510
Email: nitzan.rosenfeld@cruk.cam.ac.uk
Webpage: www.rosenfeldlab.org

Cancer Research UK Cambridge Institute
University of Cambridge
Li Ka Shing Centre, Robinson Way
Cambridge CB2 0RE

Tel: +44 (0)1223 769 500

UNIVERSITY OF CAMBRIDGE
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

CRUK Cambridge Institute
Robinson Way,
Cambridge,
CB2 0RE

02nd January 2019

To whom it may concern,

I am writing in support of Ms Davina Gale's application for PhD by Publication. I have worked with Ms Gale for the past 5 years and she has contributed significantly towards a recent publication that I co-first authored entitled 'Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models'.

Specifically she was instrumental in the original design of the project whereby we used dPCR to identify specific sequences known to denote the human origin of 'circulating' DNA within patient-derived orthotopic xenografts of glioblastoma. She performed several of the early dPCR experiments herself as well as analysing the data from those investigations. When preparing the work for publication she contributed at every stage of the manuscript's preparation and provided invaluable insight that significantly improved the final document.

Overall I thoroughly support this application.

Yours faithfully

Mair

Mr Richard Mair PhD FRCS (Neuro. Surg)
University Lecturer in Neurosurgical Oncology
Honorary Consultant Neurosurgeon
University of Cambridge, UK

UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

CRUK Cambridge Institute
University of Cambridge
Li Ka Shing Centre
Robinson Way
Cambridge CB2 0RE

January 9th, 2019

Postgraduate Research Office
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

Dear Sir/Madam,

I am delighted to write in support of Davina Gale's application for a PhD by Publication at the University of East Anglia. As part of that I would like to confirm Davina's contribution to the following manuscripts published as a preprint in bioRxiv, and as a research article in Science Translational Medicine:

- Mouliere F, Piskorz A, Chandrananda D, Moore E, Morris J, Smith C, Goranova T, Heider K, Mair R, Supernat A, Gounaris I, Ros S, Wan J, Jimenez-Linan M, **Gale D**, Brindle K, Massie C, Parkinson C, Brenton J, **Rosenfeld N**; ***bioRxiv; 2017***, http://dx.doi.org/10.1101/134437; 'Selecting Short DNA Fragments in Plasma Improves Detection of Circulating Tumour DNA.'

- Mouliere F, Chandrananda D, Piskorz A, Moore E, Morris J, Barlebo Ahlborn L, Mair R, Goranova T, Marass F, Heider K, Wan J, Supernat A, Hudecova I, Gounaris I, Ros S, Jimenez-Linan M, Garcia-Corbacho J, Patel K, Østrup O, Murphy S, Eldridge M, **Gale D**, Stewart G, Burge J, Cooper W, van der Heijden M, Massie C, Watts C, Corrie P, Pacey S, Brindle K, Baird R, Mau-Sørensen M, Parkinson C, Smith C , Brenton JD, **Rosenfeld N**; ***Science Translational Medicine; 2018***; Vol. 10, Issue 466, eaat4921 DOI: 10.1126/scitranslmed.aat4921; 'Enhanced detection of circulating tumor DNA by fragment size analysis'

I am a senior co-corresponding author on both these manuscripts, which investigate the size distribution of cell-free DNA fragments. In the Science Translational Medicine paper, ctDNA fragment sizes were analysed in 344 plasma samples from 200 patients with 18 different types of cancer, and 65 healthy controls. We demonstrated that tumour-specific DNA fragments were shorter than DNA of non-tumour origin, and high ctDNA cancers (melanoma, breast, ovarian, lung, colorectal and cholangiocarcinoma) had an increased proportion of fragments less than 150bp, compared to healthy controls and low ctDNA cancers (renal, glioblastoma, bladder, pancreatic). We were able to demonstrate the ability to use *in vitro* and *in silico* size selection to enrich for shorter circulating tumour DNA (ctDNA) fragments, and machine learning algorithms to enhance detection of ctDNA. In the bioRxiv paper, *in vitro* size selection resulted in up to 11-fold enrichment of ctDNA in patients with high-grade serous ovarian carcinoma, and detection of clinically-relevant somatic copy number aberrations not previously detected without enrichment. These studies may have important implications for early detection of cancer and minimal residual disease.

Davina Gale made important contributions to these papers in establishing protocols for collection of high quality clinical samples, and working with several clinical teams to help establish important collections in different cancers including melanoma, ovarian and breast cancer. In particular, Davina had previously co-led a study analysing ctDNA in high-grade serous ovarian cancer, and samples from this study were included in both papers, where we were able to demonstrate the effects of size selection to enrich ctDNA and enhance detection. Davina also helped train some of the staff that carried out this project. Davina played a key role setting up collaborations and plans and designing lab protocols for translational studies (MelResist, CALIBRATE) from which samples were collected for this project. In preparation of the manuscripts, Davina also provided critical and insightful scientific feedback to help improve the papers.

If you require any further information, please do not hesitate to contact me.

Yours faithfully,

Dr. Nitzan Rosenfeld
Senior Group Leader
Cancer Research UK Cambridge Institute
University of Cambridge
Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE
Tel: 01223-769769, Fax: 01224-769510
Email: nitzan.rosenfeld@cruk.cam.ac.uk
Webpage: www.rosenfeldlab.org

Appendix 3: Full list of all author contributions to each submitted publication.

I would like to acknowledge the important contributions of my colleagues in all of the submitted publications, reprinted in Appendix 6. Author contributions are reproduced from the relevant publications, where available. My contributions are highlighted in bold. The number of times the papers have been cited by Google Scholar since publication are listed at the end of each section.

Parkinson*, Gale * et al.; PLoS Medicine, 2016

Parkinson C*, Gale D*, Piskorz A, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K, Hosking K, Gounaris I, Jimenez-Linan M, Earl H, Qian W, Rosenfeld N†, Brenton JD†.

- CAP and **DG** contributed equally to this work.
- Senior authors: NR, JDB.
- Conceptualization: CAP, **DG,** NR, JDB.
- Data curation: CAP, **DG,** AMP, KS, KH, NR, JDB.
- Formal analysis: CAP, **DG,** AMP, HA, SF, PM, ES, WQ, NR, JDB.
- Funding acquisition: NR, JDB.
- Investigation: CAP, **DG**, AMP, HB, CH, HA, SF, PM, ES, KS, KH, MJ-L, HME, IG.
- Methodology: CAP, **DG**, NR, JDB.
- Project administration: **DG**, AMP, HB, CH, KS.
- Resources: CAP HB CH IG HME JDB.
- Software: NR.
- Supervision: NR, JDB.
- Validation: CAP, **DG**, WQ, NR, JDB.
- Visualization: CAP, **DG**, NR, JDB.
- Writing – original draft: CAP, **DG**, NR, JDB.
- Writing – review & editing: CAP, **DG**, AMP, HB, CH, HA, SF, PM, ES, KS, KH, IG, MJ-, LH, ME, WQ, NR, JDB.
- According to Google Scholar, this paper has been cited 53 times since 2016.
- PLoS Medicine listed it as one of their Top 50 most downloaded articles in 2016. An editorial article by Elaine Mardis and Marc Landanyi at the time of publication cite that this 'retrospective study is important in laying the foundations to change the standard of care' (Mardis and Ladanyi 2016).

Schwarz et al., PLoS Medicine, 2015

Schwarz R, Ng C, Cooke S, Newman S, Temple J, Piskorz A, Gale D, Sayal K, Murtaza M, Baldwin P, Rosenfeld N, Earl H, Sala E, Jimenez-Linan M, Parkinson C, Markowetz F†, Brenton JD†.

- Conceived and designed the experiments: JDB, ES, FM, RFS, CKYN, SLC, AMP, **DG.**
- Performed the experiments: SLC, SN, JT, AMP, **DG**.
- Analysed the data: RFS, CKYN, SLC, SN, AMP, **DG,** KS, MM, NR, MJL, CAP, FM, JDB.
- Contributed reagents/materials/analysis tools: RFS, CKYN, SN, PJB, MM, NR, ES, CAP FM, JDB.
- Enrolled patients: HME, CAP, JDB.
- Wrote the first draft of the manuscript: RFS, FM, JDB.
- Wrote the paper: RFS, CKYN, AMP, **DG,** FM, JDB.
- Agree with manuscript results and conclusions: RFS, CKYN, SLC, SN, JT, AMP, **DG,** KS, MM, PJB, NR, HME, ES, MJL, CAP, FM, JDB.
- According to Google Scholar, this paper has been cited 170 times since 2015.

Forshew*...Gale* et al.; Science Translational Medicine, 2012

Forshew T*, Murtaza M*, Parkinson C*, Gale D*, Tsui DWY*, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD†, Rosenfeld N†.

- TF, MM, CP, **DG,** DT contributed equally to this work.
- Senior authors: JDB, NR.
- Designed the study:TF MM, CP, **DG**, DWYT, CC, JDB, and NR.
- Developed methods: TF, MM, DWYT, FK, JH, APM, and NR.
- Collected the data: TF, **DG**, DWYT, AMP, SJD.
- Analysed TAm-Seq data: TF, MM, NR.
- Designed clinical studies and collected samples and clinical data: CP, SJD, CC, JDB.
- Performed pathological analysis: MJL.
- Contributed sequencing data: DB.
- Interpreted data: TF, MM, CP, **DG**, DWYT, JH, APM, JDB, NR.
- Wrote the manuscript: TF, MM, NR with assistance from CP, **DG**, DWYT, APM, JDB, and other authors.
- All authors approved the final manuscript.
- According to Google Scholar, the paper has been cited 761 times since 2012.

Murtaza et al., Nature, 2013

Murtaza M*, Dawson SJ*, Tsui DWY*, Gale D, Forshew T, Piskorz AM, Parkinson C, Chin S, Kingsbury Z, Wong AS, Marass F, Humphray S, Hadfield J, Bentley D, Chin TM, Brenton JD†, Caldas C†, Rosenfeld N†.

- Designed the study: MM, SJD, TF, DWYT, **DG**, JDB, CC, NR.
- Developed methods: MM, DWYT, TF.
- Designed and conducted the prospective clinical studies: SJD, CP, ASCW, TMC, JDB, CC.
- Generated data: MM, SJD, DWYT, **DG**, TF, AMP.
- Contributed sequencing data: ZK, SH, DB.
- Analysed sequencing data: MM, FM, NR.
- Contributed to experiments and data analysis: SFC, JH.
- Interpreted data: MM, SJD, DWYT, TMC, JDB, CC, NR.
- Wrote the paper: MM, NR with assistance from SJD, DWYT, CC, JDB and other authors.
- All authors approved the final manuscript.
- Project co-leaders and joint senior authors: JDB, CC, NR.
- According to Google Scholar, this paper has been cited 1072 times since 2013.

Murtaza et al., Nature Communications, 2015

Murtaza M*, Dawson SJ&, Pogrebniak K, Rueda O, Provenzano E, Grant J, Chin SF, Tsui DWY, Marass F, Gale D, Ali HR, Shah P, Contente-Cuomo T, Farahani H, Shumansky K, Kingsbury Z, Humphray S, Bentley D, Shah S, Wallis M, Rosenfeld N†, Caldas C†.

- Joint first co-authors: MM, SJD.
- Designed the study: MM, SJD, NR, CC.
- Developed methods: MM, DWYT.
- Conducted prospective clinical, histopathological and imaging studies: SJD, EP, JG, MW, CC.
- Generated data: MM, SJD, SFC, DWYT, **DG**, PS, TC-C.
- Contributed sequencing data: ZK, SH, DB.
- Analysed sequencing data: MM, KP, OMR, FM, HF, KS, SPS.
- Contributed to experiments and data analysis: S-FC HRA.
- Interpreted data: MM, SJD, NR, CC.
- Wrote the paper: MM, CC; With assistance from SJD, NR and other authors.
- All authors approved the final manuscript.

- Project co-leaders and joint senior authors: NR, CC.
- According to Google Scholar, this paper has been cited 193 times since 2015.

Risberg et al., Journal of Molecular Diagnostics, 2018

Risberg B[*], Tsui DWY*[†], Biggs H, Ruiz-Valdepenas Martin de Almagro A, Dawson SJ, Hodgkin C, Jones L, Parkinson C, Piskorz A, Marass F, Chandrananda D, Moore E, Morris J, Plagnol V, Rosenfeld N, Caldas C, Brenton JD, Gale D[†].

- Conceived and designed the study: DWYT, HB., SJD, CP, AP, NR, CC, JDB, **DG**.
- Processed samples, collected clinical data and managed samples: DWYT, HB, SJD, CH, LJ, CP, AP.
- Performed experiments: BR, DWYT, AR-VMA, SJD, EM, **DG**.
- Analyzed NGS data: FM, DC, JM, VP.
- Wrote the manuscript: BR, DWYT, DC, **DG**.
- All authors approved the final manuscript
- According to Google Scholar, this paper has been cited once since 2018.

Gale et al., PLoS ONE, 2018

Gale D[‡], Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, Calaway J, Blais S, Jones G, Clark J, Dimitrov P, Pugh M, Woodhouse S, Epstein M, Fernandez-Gonzalez A, Whale AS, Huggett JF, Foy CA, Jones GM, Raveh-Amit H, Schmitt K, Devonshire A, Green E, Forshew T, Plagnol V, Rosenfeld N[†].

- Conceptualization: **DG**, ARJL, TF, VP, NR.
- Data curation: **DG**, ARJL, ME, TF, VP.
- Formal analysis: PD, ME, VP.
- Funding acquisition: CAF, KS, NR.
- Investigation: **DG**, ARJL, KH, MM, BD, SS, JCa, SB, SW, AFG, ASW, GMJ, AD.
- Methodology: **DG**, ARJL, SS, AFG, ASW, AD, TF, VP, NR.
- Project administration: **DG**, KS, AD.
- Resources: HRA, KS.
- Software: VP.
- Supervision: **DG**, GJ, JCl, JFH, CAF, TF, VP, NR.
- Validation: **DG**, ARJL, KH, SS, JC, SB, MP, EG, TF, VP.
- Visualization: AD.

- Writing – original draft: **DG**.
- Writing – review & editing: **DG**, ARJL, KH, MP, SW, CAF, AD, EG, TF, VP, NR.
- According to Google Scholar, this paper has been cited 10 times since 2018.

Plagnol et al., PLoS ONE, 2018

Plagnol V*, Woodhouse S*, Howarth K, Lensing S, Smith M, Epstein M, Madi M, Smalley S, Leroy C, Hinton J, De Kievit F, Musgrave-Brown E, Herd C, Neblett K, Brennan W, Dimitrov P, Campbell N, Morris C, Rosenfeld N, Clark J, Gale D, Platt J, Calaway J, Jones G, Forshew T†.

- Conceptualization: VP, SW, KH, SL, MS, ME, PD, CM, NR, JC, GJ, TF.
- Data curation: VP, KH, FdK, GJ, TF.
- Formal analysis: VP, SW, FdK, GJ, TF.
- Investigation: VP, SW, KH, SL, MS, ME, MM, FdK; EM-B, CH, KBN, WB, NC, CM, JCa, TF.
- Methodology: VP, SW, KH, SL, MS, ME, MM, SS, EMB, CH, WB, PD, **DG**, JCa, TF
- Project administration: JCl, JP.
- Resources: SS, JCl, JP.
- Software: VP, ME, CL, JH, FdK, PD.
- Supervision: VP, SW, KH, CL, JH, NC, CM, NR, JCl, **DG**, JP, JCa, GJ, TF.
- Validation: VP, SW, KH, MS, **DG**, JCa, GJ, TF.
- Visualization: VP, ME.
- Writing – original draft: VP, TF.
- Writing – review & editing: VP, SW, KH, SL, MS, ME, MM, SS, CL, JH, FdK, EMB, CH, KB-N, WB, PD, NC, CM, NR, JCl, **DG**, JP, JCa, GJ, TF.
- According to Google Scholar, this paper has been cited 9 times since 2018.

Remon et al., Annals of Oncology, 2017

Remon J, Caramella C, Jovelet C, Lacroix L, Lawson ARJ, Smalley S, Howarth K, Gale D, Green E, Plagnol V, Rosenfeld N, Planchard D, Bluthgen MV, Gazzah A, Pannet C, Nicotra C, Auclin E, Soria JC, Besse B†.

- **Note: Official details are not listed in the publication, so authorship details are reported to the best of my knowledge.**
- Conceptualization: JR, EG, NR, BB.
- Data curation: JR, ARJL, SS, KH, **DG**, EG, VP.
- Collection of clinical specimens: JR, CC, CJ, LL, DP, MVB, AG, CP, CN, EA, JCS, BB.

- Formal analysis: JR, VP.
- Funding acquisition: NR.
- Investigation: JR, CC, CJ, LL, ARJL, SS, KH, **DG**, VP.
- Methodology: **DG**, ARJL, SS, KH, VP.
- Project administration: JR, EG.
- Supervision: JR, **DG**, EG, NR.
- Writing – original draft: JR, EG.
- Writing – review & editing: JR, DG, EG, VP, NR, **DG**, BB.
- According to Google Scholar, this paper has been cited 69 times since 2017.

Dawson et al., New England Journal of Medicine, 2013

Dawson SJ*, Tsui DWY*, Murtaza M, Biggs H, Rueda O, Chin SF, Dunning M, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C†, Rosenfeld N†.

- **Note: Official details are not listed in the publication, so authorship details are reported to the best of my knowledge.**
- Conceptualization: SJD, DWYT, MM, **DG**, CC, NR.
- Data curation: SJD, DWYT, MM, OR.
- Collection of clinical specimens: SJD, HB, CC.
- Formal analysis: SJD, MM, OR, MD.
- Funding acquisition: CC, NR.
- Investigation: SJD, DWYT, MM,OR, SFC, MD, DG, TF, BM-A, SR, SH, JB, DH, MW, DB.
- Methodology: SJD, DWYT, MM, **DG**, TF, NR.
- Project administration: SJD.
- Writing – original draft: SJD, DWYT, CC, NR.
- Writing – review & editing: SJD, DWYT, MM, **DG**, CC, NR.
- According to Google Scholar, this paper has been cited 1258 times since 2013.

Tsui et al., EMBO Molecular Medicine, 2018

Tsui DWY*, Murtaza M*, Wong ASC, Rueda OM, Smith CG, Chandrananda D, Soo RA, Lim H, Goh B, Caldas C, Forshew T, Gale D, Liu W, Morris J, Marass F, Eisen T, Chin T†, Rosenfeld N†.

- Initiated and designed the study: DWYT, TE, TMC, NR.
- Developed methods: DWYT, CGS, DC, TF, MM, FM, **DG**, NR.

- Analysed the data: DWYT, MM, TMC and NR; With assistance from OMR, CC, CGS, DC, FM, JM and WL.
- Treating physicians of the patients included in this study, and collected samples and clinical data: ASCW, RAS, HLL, BCG, TMC.
- Interpreted the data and wrote the manuscript: DWYT, MM, TMC, NR with assistance from other authors.
- All authors approved the final manuscript.
- According to Google Scholar, this paper has been cited 4 times since 2018.

Mair et al., Cancer Research, 2019

Mair R*, Mouliere F*, Smith CG, Chandrananda D, Gale D, Marass F, Tsui DWY, Massie CE, Wright AJ, Watts C, Rosenfeld N†, Brindle KM†.

- Conception and design: RM, FM, CGS, CW
- Development of methodology: RM, FM, CGS, **DG**, CW
- Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): RM, FM, **DG**, DWYT, AJW, CW
- Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): RM, FM, CGS, DC, **DG**, **FM**, CEM, AJW, NR
- Writing, review, and/or revision of the manuscript: RM, FM, CGS, DC, **DG**, FM, DWYT, AJW, CW, NR, KMB.
- According to Google Scholar, this paper has been cited once since 2018.

Mouliere et al., bioRxiv, 2017

Mouliere F*, Piskorz A*, Chandrananda D*, Moore E*, Morris J, Smith C, Goranova T, Heider K, Mair R, Supernat A, Gounaris I, Ros S, Wan J, Jimenez-Linan M, Gale D, Brindle K, Massie C, Parkinson C, Brenton JD†, Rosenfeld N†.

- Conceptualised and designed the study: FM, AMP, DC, EM, JDB, NR.
- Performed experiments and collected data: FM, AMP, EM, KH, CGS, JCMW, DG, RM, TG, AS, IG, CAP.
- Conceptualised and designed the t-MAD index and performed sWGS bioinformatics analysis: DC.
- Performed TAm-Seq bioinformatics analysis: JM.
- Designed animal model: RM, SR.

- Performed histopathology revision: MJL.
- Wrote the manuscript: FM, AMP, DC, EM, JDB, NR.
- All co-authors critically reviewed the manuscript.
- Supervised the study: FM, AMP, DC, JDB, NR.
- According to Google Scholar, this paper has been cited 5 times since 2017.

Mouliere et al., Science Translational Medicine, 2018

Mouliere F*, Chandrananda D*, Piskorz A*, Moore E*, Morris J, Barlebo Ahlborn L, Mair R, Goranova T, Marass F, Heider K, Wan J, Supernat A, Hudecova I, Gounaris I, Ros S, Jimenez-Linan M, Garcia-Corbacho J, Patel K, Østrup O, Murphy S, Eldridge M, Gale D, Stewart G, Burge J, Cooper W, van der Heijden M, Massie C, Watts C, Corrie P, Pacey S, Brindle K, Baird R, Mau-Sørensen M, Parkinson C, Smith C , Brenton JD†, Rosenfeld N†.

- Conceptualized and designed the study: FM, AMP, DC, EKM, JDB, NR.
- Performed experiments and collected data: FM, AMP, EKM, LBA, KH, CGS, JCMW, **DG**, RM, TG, AS, IG, OØ, CAP, MM-S, IH, KP, WNC.
- Conceptualized the size selection approach: FM, AMP, DC, EKM, CGS.
- Designed and performed *in vitro* size selection: FM, AMP, EKM.
- Conceptualized and designed the fragmentation feature analysis: FM, DC, with input from FM, NR.
- Conceptualized and designed the t-MAD index: D.C. with input from FM.
- Bioinformatics analysis of somatic CNAs from sWGS: FM, DC.
- Performed bioinformatics analysis of TAm-Seq: JM.
- Designed the tailored captured sequencing and performed WES: FM, LBA.
- Performed bioinformatics analysis of the capture sequencing and WES: FM, JM.
- Developed and optimized mutation calling algorithms: MDE.
- Designed the animal model: RM, KB, SR.
- Collected human samples: JGC, SP, RDB, MMS, GDS, JB, SM, PC, CW, RM, MSvdH.
- Performed histopathology revision: MJL, JB.
- Wrote the manuscript: FM, DC, AMP, EKM, JDB, NR.
- All co-authors have critically reviewed the manuscript.
- Supervised the study: FM, AMP, DC, JDB, NR.
- Co-ordinated the study: FM
- According to Google Scholar, this paper has been cited 10 times since 2018.

Appendix 4: Reproduction of each submitted publication, with kind permission from the journals.

PLOS | MEDICINE

RESEARCH ARTICLE

Exploratory Analysis of *TP53* Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study

Christine A. Parkinson[1,2,3,4,5☯], Davina Gale[1,5☯], Anna M. Piskorz[1,5], Heather Biggs[2,3,5], Charlotte Hodgkin[2,5], Helen Addley[4,5], Sue Freeman[4,5], Penelope Moyle[4,5], Evis Sala[4¤], Karen Sayal[1], Karen Hosking[2,3,5], Ioannis Gounaris[1,5], Mercedes Jimenez-Linan[4,5], Helena M. Earl[2,3,4,5], Wendi Qian[4,5], Nitzan Rosenfeld[1,5]*, James D. Brenton[1,2,3,4,5]*

1 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom, **2** Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, Cambridge, United Kingdom, **3** NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom, **4** Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, **5** Cancer Research UK Major Centre–Cambridge, Cancer Research UK Cambridge Institute, Cambridge, United Kingdom

☯ These authors contributed equally to this work.
¤ Current address: Memorial Sloan Kettering Cancer Center, New York, New York, United States of America.
* Nitzan.Rosenfeld@cruk.cam.ac.uk (NR); James.Brenton@cruk.cam.ac.uk (JDB)

OPEN ACCESS

Citation: Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, et al. (2016) Exploratory Analysis of *TP53* Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. PLoS Med 13 (12): e1002198. doi:10.1371/journal.pmed.1002198

Academic Editor: Elaine Rene Mardis, Washington University School of Medicine, UNITED STATES

Received: July 16, 2016

Accepted: November 9, 2016

Published: December 20, 2016

Copyright: © 2016 Parkinson et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Cancer Research UK (http://www.cancerresearchuk.org/) Grant numbers: A15601 (JDB), A11906 (NR), A20240 (NR), A18072 (JDB). JDB was supported by the National Institute for Health Research Cambridge Biomedical Research Centre. CAP was supported in part by the Academy of Medical

Abstract

Background

Circulating tumour DNA (ctDNA) carrying tumour-specific sequence alterations may provide a minimally invasive means to dynamically assess tumour burden and response to treatment in cancer patients. Somatic *TP53* mutations are a defining feature of high-grade serous ovarian carcinoma (HGSOC). We tested whether these mutations could be used as personalised markers to monitor tumour burden and early changes as a predictor of response and time to progression (TTP).

Methods and Findings

We performed a retrospective analysis of serial plasma samples collected during routine clinical visits from 40 patients with HGSOC undergoing heterogeneous standard of care treatment. Patient-specific *TP53* assays were developed for 31 unique mutations identified in formalin-fixed paraffin-embedded tumour DNA from these patients. These assays were used to quantify ctDNA in 318 plasma samples using microfluidic digital PCR. The *TP53* mutant allele fraction (TP53MAF) was compared to serum CA-125, the current gold-standard response marker for HGSOC in blood, as well as to disease volume on computed tomography scans by volumetric analysis. Changes after one cycle of treatment were compared with TTP.

The median TP53MAF prior to treatment in 51 relapsed treatment courses was 8% (interquartile range [IQR] 1.2%–22%) compared to 0.7% (IQR 0.3%–2.0%) for seven untreated

Sciences, the Wellcome Trust, British Heart Foundation and Arthritis Research UK. CAP, DG, AMP, HB, CH, HA, SF, PM, KH, IG, MJL, HME, WQ, NR and JDB are affiliated with Cancer Research UK but the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: We have read the journal's policy and the authors of this manuscript have the following competing interests: DG, NR and JDB are co-founders, shareholders and officers/consultants of Inivata Ltd, a cancer genomics company that commercialises ctDNA analysis.

Abbreviations: AC/ml, amplifiable copies per millilitre of plasma; CRC, colorectal cancer; CT, computed tomography; ctDNA, circulating tumour DNA; FFPE, formalin-fixed paraffin-embedded; HGSOC, high-grade serous ovarian carcinoma; HR, hazard ratio; IQR, interquartile range; PS, performance status; RECIST, Response Evaluation Criteria in Solid Tumours; ROC, receiver operating characteristic; TP53MAC, *TP53* mutant allele count; TP53MAF, *TP53* mutant allele fraction; TP53TAC, *TP53* total allele count; TTP, time to progression.

newly diagnosed stage IIIC/IV patients. TP53MAF correlated with volumetric measurements (Pearson $r = 0.59$, $p < 0.001$), and this correlation improved when patients with ascites were excluded ($r = 0.82$). The ratio of TP53MAF to volume of disease was higher in relapsed patients (0.04% per cm^3) than in untreated patients (0.0008% per cm^3, $p = 0.004$). In nearly all relapsed patients with disease volume > 32 cm^3, ctDNA was detected at $\geq$20 amplifiable copies per millilitre of plasma. In 49 treatment courses for relapsed disease, pre-treatment TP53MAF concentration, but not CA-125, was associated with TTP. Response to chemotherapy was seen earlier with ctDNA, with a median time to nadir of 37 d (IQR 28–54) compared with a median time to nadir of 84 d (IQR 42–116) for CA-125. In 32 relapsed treatment courses evaluable for response after one cycle of chemotherapy, a decrease in TP53MAF of >60% was an independent predictor of TTP in multivariable analysis (hazard ratio 0.22, 95% CI 0.07–0.67, $p = 0.008$). Conversely, a decrease in TP53MAF of $\leq$60% was associated with poor response and identified cases with TTP < 6 mo with 71% sensitivity (95% CI 42%–92%) and 88% specificity (95% CI 64%–99%). Specificity was improved when patients with recent drainage of ascites were excluded. Ascites drainage led to a reduction of TP53MAF concentration. The limitations of this study include retrospective design, small sample size, and heterogeneity of treatment within the cohort.

Conclusions

In this retrospective study, we demonstrated that ctDNA is correlated with volume of disease at the start of treatment in women with HGSOC and that a decrease of $\leq$60% in TP53MAF after one cycle of chemotherapy was associated with shorter TTP. These results provide evidence that ctDNA has the potential to be a highly specific early molecular response marker in HGSOC and warrants further investigation in larger cohorts receiving uniform treatment.

Author Summary

Why Was This Study Done?

- The standard clinical blood test for measuring response in women receiving chemotherapy for high-grade serous ovarian cancer (HGSOC) is the serum protein cancer antigen 125 (CA-125).
- CA-125 is sensitive but it lacks specificity for detection of ovarian cancer, and in response to chemotherapy, CA-125 level does not change rapidly enough to suggest change in treatment after one or two cycles if chemotherapy treatment is ineffective.
- Better tumour markers are required, and circulating tumour DNA (ctDNA) is a promising candidate.
- ctDNA is cell-free DNA derived from tumour cells that can be detected in the bloodstream; ctDNA can be used as a highly specific marker because it carries mutations unique to the cancer.

What Did the Researchers Do and Find?

- HGSOC is an ideal cancer to test ctDNA as a biomarker because 99% of patients have a mutation in the *TP53* gene.
- We designed patient-specific *TP53* assays for a retrospective study of 40 patients with HGSOC, and these were used to quantify the amount of ctDNA in 318 plasma samples collected before, during, and after chemotherapy.
- We asked if ctDNA level was correlated with the amount of disease present before chemotherapy treatment measured using 3-D volume reconstruction from CT images taken as part of routine care.
- We also asked if the decrease in *TP53* ctDNA after one cycle of chemotherapy treatment could predict which patients would have progression of their cancer within six months.
- ctDNA level, but not CA-125 level, was strongly correlated with the total volume of disease.
- Patients whose ctDNA level exhibited a decrease of >60% after one cycle of chemotherapy had a significantly longer time to progression than those whose ctDNA level decreased by 60% or less.

What Do These Findings Mean?

- *TP53* ctDNA has the potential to be a clinically useful blood test to assess prognosis and response to treatment in women with HGSOC.
- The response findings from this retrospective study should be confirmed in larger, prospective studies with uniform treatment. If these findings are confirmed, *TP53* ctDNA could be used in HGSOC clinical trials and routine practice to identify earlier whether treatment is effective.

Introduction

The development of blood biomarkers that can be used for early detection of cancer or to measure tumour burden and response to treatment is a major goal of translational cancer research across all cancer types. Both tumour-derived proteins and DNA can be detected in circulating plasma and serum from cancer patients [1,2]. In epithelial ovarian cancer, particularly high-grade serous ovarian cancer (HGSOC), cancer antigen 125 (CA-125) is a serum glycoprotein biomarker used in standard clinical practice for the first assessment of pelvic masses [3] and for monitoring response to treatment [4,5]. However, CA-125 is limited by specificity, since it can also be expressed by normal tissues. Two large screening studies using CA-125 and ultrasound have failed to show an improvement in mortality on primary analysis [6,7]. CA-125 has heterogeneous intra- and inter-patient cellular expression and a long biological half-life in serum, resulting in utility for sequential clinical measurements to indicate the trend of treatment response, but not as a direct reflection of absolute tumour volume [8]. Whilst several studies have shown that pre-treatment CA-125 is prognostic and that early changes following chemotherapy are predictive [9,10], the positive and negative predictive values are not sufficient for use as a surrogate for radiological response or time to progression (TTP), or as the sole endpoint in registration trials. More recently, attention has turned to DNA-based biomarkers in blood as potentially superior measurements of response. In contrast to protein biomarkers, which typically are not specific to cancer cells, circulating tumour DNA (ctDNA) measures levels of mutations in plasma cell-free DNA and provides highly cancer-specific biomarkers. ctDNA fragments have a short half-life, and their levels have been shown in other cancer types to be related to tumour volume and response to treatment [11,12].

Although mutations in cell-free DNA have been studied for more than 20 years [13,14], recent improvements in assay sensitivity and in the ability to routinely identify patient-specific mutations have made it practical to accurately measure ctDNA levels in blood samples to monitor tumour response to treatment [12,15–18]. ctDNA may represent ~0.01%–90% of total circulating DNA and potentially offers greater specificity than protein biomarkers [19–21]. In initial studies performed on small numbers of patients, ctDNA has compared favourably with different serum tumour markers [16,20]. In a study of 18 metastatic colorectal cancer (CRC) patients following hepatic metastasectomy, Diehl and colleagues showed, using sensitive BEAMing assays, that ctDNA outperformed the serum marker carcinoembryonic antigen for the detection of microscopic disease [12]. Further studies in CRC have now been published correlating changes in ctDNA with response to chemotherapy [22]. Similarly, ctDNA outperformed the serum marker cancer antigen 15-3 for assessing tumour response to treatment in metastatic breast cancer [16].

HGSOC is an ideal cancer type in which to explore the clinical utility of ctDNA for response monitoring during treatment in comparison to a clinically accepted biomarker of response, as >99% of cases show mutations in *TP53* [23–25] and >90% of advanced HGSOC cases express the serum protein tumour marker CA-125 [26,27]. We and others have previously shown that *TP53* mutations can be detected in ctDNA from patients with advanced HGSOC and that, in a small number of patients studied, changes in ctDNA levels correlated with other clinical response measurements including CA-125 [21,28–30]. However, in HGSOC the relationship of ctDNA to tumour volume, dynamic ctDNA changes during chemotherapy, and the relationship of early changes to outcomes during chemotherapy have not been characterised.

The primary aim of this study was to define the distribution and dynamics of ctDNA in patients with recurrent HGSOC treated with standard of care chemotherapy, and to correlate ctDNA with the volume of disease. A secondary aim was to evaluate whether early change in ctDNA could predict TTP.

Methods

Ethics and Consent

Patients included in this report were enrolled in the prospective CTCR-OV04 clinical study, which collected blood and tissue samples for exploratory biomarker studies from patients treated at Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust. All patients provided written informed consent for participation in the study and for use of their donated tissue and blood specimens. The CTCR-OV04 study was approved by the Cambridgeshire Research Ethics Committee (reference 08/H0306/61).

Study Design and Patients

In order to quantify ctDNA levels, patient-specific *TP53* TaqMan assays were designed to target mutations identified in formalin-fixed paraffin-embedded (FFPE) tissue. Digital PCR was

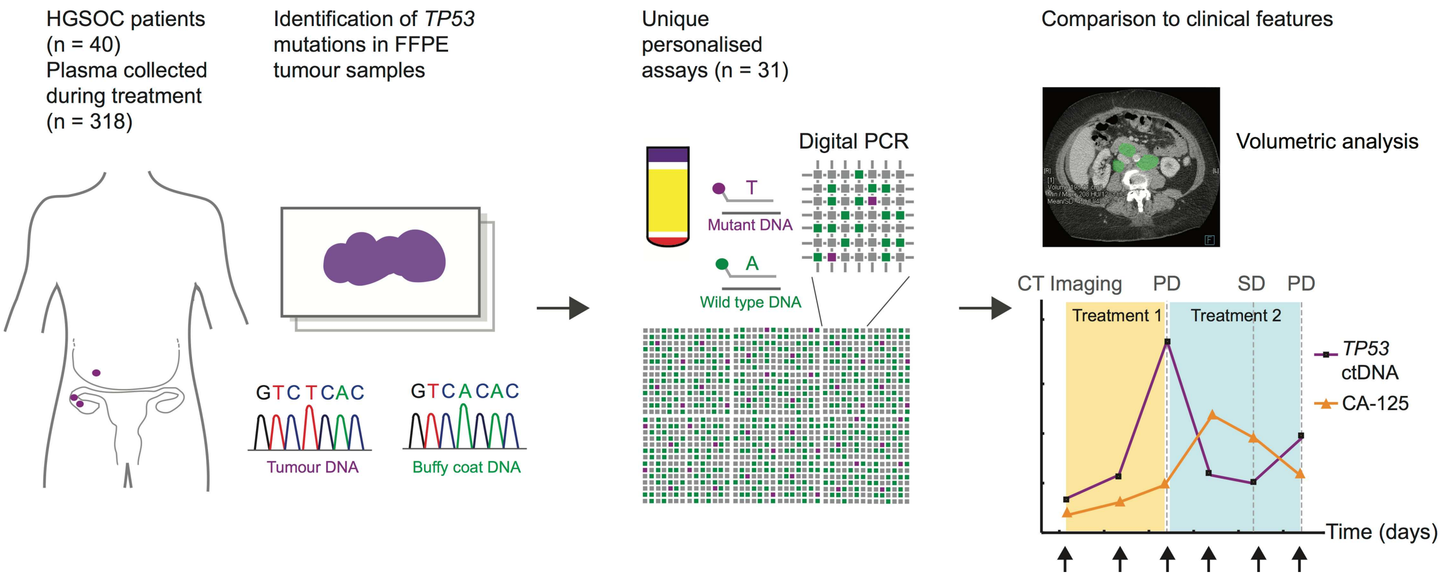

HGSOC patients
(n = 40)
Plasma collected
during treatment
(n = 318)
Identification of TP53
mutations in FFPE
tumour samples
GTCTCAC
Tumour DNA
GTCACAC
Buffy coat DNA
Unique
personalised
assays (n = 31)
Digital PCR
T
Mutant DNA
A
Wild type DNA
Comparison to clinical features
Volumetric analysis
CT Imaging
PD
SD
PD
Treatment 1
Treatment 2
TP53 ctDNA
CA-125
Time (days)

Fig 1. Schema of workflow for circulating tumour DNA analysis. ctDNA, circulating tumour DNA; FFPE, formalin-fixed paraffin-embedded; HGSOC, high-grade serous ovarian carcinoma; PD, progressive disease; SD, stable disease.

doi:10.1371/journal.pmed.1002198.g001

used to measure ctDNA levels in cell-free DNA from plasma samples collected during courses of chemotherapy, as shown in Fig 1.

This report has been written in accordance with REMARK and STROBE guidelines [31,32]. REMARK diagrams, the STROBE checklist, and a summary of the statistical analysis are found in S1 Checklist, S1 Text and S1 and S4 Figs. Patient selection criteria for this report were as follows: histological diagnosis of HGSOC of the ovary, primary peritoneum, or fallopian tube (hereafter HGSOC); ≥2 plasma samples from at least one course of chemotherapy including a pre-treatment sample collected before the start of the course; and *TP53* point mutation or short indel identified by sequencing of FFPE tumour DNA (S2 Fig).

The study was initiated as an exploratory retrospective analysis of samples collected as part of the CTCR-OV04 study protocol (see above). Further to initial findings, the cohort was expanded to 40 patients for whom missense mutations or short indels in *TP53* were identified by Sanger sequencing of tumour DNA. All patients were enrolled between 19 August 2009 and 13 June 2011. Patients were followed up during routine clinical practice. Follow-up was censored on 16 September 2015, with a median duration of 59 mo (range 43–70 mo). TTP for relapsed patients was defined as the interval from cycle 1 day 1 of chemotherapy to the date of progression measured by Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 [33]. Patients with non-cancer-related/unknown cause of death were censored at the date of death and included in the TTP analysis. Clinical details, including CA-125, stage, residual disease after surgery, chemotherapy, and procedure dates, were abstracted from clinical records by research staff.

TP53 Mutation Identification

FFPE tissue blocks were cut as 8-μm sections and tumour-enriched regions were recovered by macrodissection based on regions marked on an adjacent haematoxylin-and-eosin-stained section by the study pathologist. DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen) and quantified using a Qubit 2.0 Fluorometer (Invitrogen). Coding sequences of the *TP53* gene (exons 2–11) were PCR-amplified from FFPE DNA using primers and conditions as described previously [34] and sequenced using an ABI 3730 DNA Analyzer (Applied Biosystems), except that an additional forward instead of reverse primer (5′-CAGGTCTCCCCAA GGCGCAC-3′) was used for the sequencing of exon 7. Mutational analysis was performed using Mutation Surveyor software version 3.97 (SoftGenetics), and sequence data were aligned to *TP53* reference sequence NC_000017.10. In patients 127 and 200, mutations were identified in FFPE DNA by TAm-Seq (tagged-amplicon deep sequencing) using the 48.48 Access Array System (Fluidigm) and GAIIx Genome Analyzer (Illumina), as previously described [21].

Plasma, Buffy Coat, and Serum Collection

Serial plasma samples were collected from patients, including at the appointment closest to the start of their chemotherapy course, which in most cases was on the first day of treatment. Peripheral blood samples were collected into EDTA tubes (Sarstedt) and centrifuged within 1 h of collection at 820*g* for 10 min to limit leukocyte lysis and degradation of cell-free DNA. Plasma aliquots of 1 ml were centrifuged in a benchtop microfuge at 14,000 rpm for 10 min, and the supernatant was transferred to sterile 1.5-ml tubes and stored at −80°C prior to extraction. Buffy coat samples were collected at the time of plasma collection, and stored at −80°C prior to DNA extraction using the QIAamp DNA Mini Kit (Qiagen). CA-125 assessments were carried out as part of routine clinical care in a clinically accredited laboratory. Where routine clinical CA-125 results were missing, CA-125 results were assessed using research serum samples taken at the same time point. For serum collections, 7.5-ml peripheral blood samples were collected into serum collection tubes (Sarstedt), gently inverted 5–10 times, and left upright at room temperature for 45–60 min to enable clot formation. Tubes were centrifuged at 3,000 rpm for 10 min, and 1-ml aliquots of serum were transferred to sterile 1.5-ml tubes and stored at −80°C.

Extraction of Circulating DNA from Plasma

Circulating DNA was extracted from 0.85–2.8 ml (median 2.1 ml) of patient plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen), and a fraction of the extracted DNA was used for digital PCR analysis (5% of the extracted DNA for each panel). Carrier RNA was added to each sample prior to lysis, and eluant was passed twice through the QIAamp column to maximise yield. Control circulating DNA was extracted from a pool of plasma from five healthy female individuals (Sera Laboratories International).

Assay Design and Validation

Dual-labelled patient-specific TaqMan assays were designed for mutated and wild-type *TP53* sequences, labelled with 6FAM, VIC, or HEX fluorophores (Applied Biosystems; Sigma-Aldrich). Sequences of all primers and probes are shown in S1 Table. Each assay was validated by digital PCR using matched FFPE tumour and buffy coat template DNA extracted from the same patient, and tumour DNA from individuals carrying non-matching *TP53* mutations. To test the performance of the assays, circulating DNA from a control mix of plasma samples from healthy volunteers was extracted 86 independent times, and *TP53* alleles (both wild-type and mutant alleles) measured by digital PCR a total of 141 times, using the 31 assays designed plus an additional similar assay designed for a patient who was excluded from this cohort (S1 and S2 Figs). Values of wild-type *TP53* concentrations fit a normal distribution, and 95% of values were within 1.5-fold of the mean value (S2 Fig). Each measurement included 765 independent real-time PCR amplifications. Of the 141 measurements (107,865 reactions in total), 19 false-positive amplifications of mutant alleles in the control samples were observed (0.018%), 13 of which were obtained in two of the 32 assays (for mutations g.12458G>A [p.R175H] and g.12460T>A [p.C176S]).

Digital PCR

Digital PCR was performed using the Biomark microfluidic system (Fluidigm), as previously described [35]. Standard operating procedures were followed, and all assay setup and liquid handling were performed in a HEPA/UV sterilising PCR workstation (Ultra-Violet Products) in a PCR-free environment to prevent PCR contamination. Data analysis was performed using Digital PCR Analysis software version 3.02 (Fluidigm) and Matlab. For each assay, a threshold for positive amplification was determined by manual inspection of the PCR amplification curves in the patient and control circulating DNA samples, and this threshold was used to determine the number of observed amplifications. A Poisson correction was used to convert the number of observed amplifications to estimated targets assuming independent segregation of the DNA molecules into the multiple digital PCR reaction chambers. The measurement by microfluidic digital PCR was corrected for the relative fraction of the extracted DNA loaded on the microfluidic array (including correction for "dead volume" lost on the array) and was normalised to units of amplifiable copies per millilitre of plasma (AC/ml). In the procedure, 3.5 μl of the total volume of 70 μl of eluted DNA was loaded for each digital PCR panel, and with correction for dead volume (~54% lost), the total DNA assayed was equivalent to a median of 0.05 ml of plasma per sample. Correspondingly, we set a cutoff for ctDNA detection at 20 AC/ml.

Total circulating cell-free DNA was measured as the *TP53* total allele count (TP53TAC), the sum of estimated targets of mutated and wild-type copies of *TP53* amplified by the assay primers. The level of mutated *TP53* ctDNA was quantified in two ways: by the number of mutated *TP53* amplifiable copies (*TP53* mutant allele count [TP53MAC]), defined as the number of single-stranded fragments of DNA amplified by the assay primers and containing the mutation of interest, and by the *TP53* mutant allele fraction (TP53MAF), defined as TP53MAC divided by TP53TAC.

CA-125

Serum CA-125 level was routinely monitored using a two-site sandwich immunoassay on a Siemens Centaur XP auto-analyser (upper limit of normal $\leq$ 30 IU/ml). CA-125 response was assessed in accordance with Gynecologic Cancer InterGroup criteria [36].

Computed Tomography Imaging

Patients underwent computed tomography (CT) imaging as part of standard care. A subset of patients had PET/CT imaging data available. All scans were retrospectively evaluated according to RECIST 1.1 by consultant radiologists subspecialised in gynaecological oncology imaging. The measurement of tumour volume was performed by consultant radiologists who were blinded to the ctDNA variables. CT images were uploaded onto a dedicated workstation and retrospectively reviewed with syngo.via multi-modality software (Siemens). A region of interest was manually placed around each visible lesion (e.g., peritoneal deposits, subcapsular disease, omental disease, ovarian masses, nodal disease), and the total volume of disease was calculated. Cases with ill-defined stranding were assessed as non-measurable disease. The presence or absence of ascites on CT was recorded.

Statistics

Pre-specified analyses were determined after sample collection but before statistical analysis was performed. Additional exploratory analyses were carried out based on the results obtained (S1 Text). Baseline characteristics were summarised using the standard descriptive statistics:

mean ± standard deviation or median with interquartile range (IQR) for continuous variables and percentage for categorical variables. Comparisons of the ratio of TP53MAF to volume of disease and CA-125 at baseline between untreated and relapsed patients were assessed with the Wilcoxon rank-sum test). Where two pre-treatment ctDNA samples were available, the sample closest to the date of volumetric CT was chosen for volumetric correlation. For all other analyses, the sample closest to treatment start date was selected.

Correlations between TP53MAF and CA-125 values and tumour volumes were analysed by Pearson rank correlation using log10 values. For log correlation calculations, TP53MAF and TP53MAC values of zero were adjusted by the addition of 0.001 times the value of the lowest value in the series.

The Cox regression model was applied to investigate the value of TP53MAF pre-treatment, change after cycle 1, and change after cycle 2 in predicting TTP. Multivariable analysis was adjusted for the following factors: age (continuous), ECOG (Eastern Cooperative Oncology Group) performance status (PS) (continuous), platinum sensitivity (yes or no), number of previous lines of chemotherapy (2 versus $\geq$3), tumour volume (continuous), ascites (absent or present), and TP53TAC. Three courses from three separate patients had missing PS, and PS was imputed as the mean PS value for all courses in the model. Total cell-free DNA level (TP53TAC) was included in the multivariable model because it has been reported as a possible independent prognostic marker in ovarian and other cancers [37,38].

In analyses where treatment courses with recent ascitic or pleural drainage were excluded, we defined "recent" as within 28 d of the baseline ctDNA sample, or between baseline and the subsequent cycle of interest (cycle 2 or 3).

We defined ctDNA as being evaluable for analysis of response if baseline TP53MAC was $\geq$40 AC/ml (double the lower limit of detection of $\geq$20 AC/ml).

The optimum cut-points for determining 6-mo TTP were identified using receiver operating characteristic (ROC) curves. The standard log-rank test was applied for the comparisons on TTP. Statistical analyses were carried out using SAS version 9.4 and R [39]; confidence intervals for sensitivity and specificity were calculated using MedCalc (http://www.medcalc.org).

Results

Patients, Samples, and Treatments

A total of 318 plasma samples were collected from 40 patients with HGSOC. S1 Fig shows the REMARK diagram for selection of patients; Table 1 shows the summary clinical features for the 40 patients (see also S2 Table for demographic information by patient and S3 Table for further details of plasma samples). In all, 261 samples were collected during treatment of relapsed disease, and included 54 courses of chemotherapy for 32 individual patients (including multiple lines of treatment per patient). A further 57 samples were collected during first-line treatment with chemotherapy from 12 patients, including four patients who had samples taken during first-line treatment and again at relapse. The ECOG PS of patients was 0–2 for all courses, where known. The median number of courses of treatment per patient with ctDNA analysis was 2 (IQR 1–2), and the median number of cycles with ctDNA analysis per course of treatment was 5 (IQR 3–6) (S1 and S2 Data).

TP53 Mutation Identification and Assay Validation

Patient-specific TaqMan assays were designed for the 40 patients, who had in total 31 unique somatic *TP53* mutations (Figs 1 and S1; S1 Table). The most common mutations were g.12458G>A (p.R175H) and g.13744G>A (p.R273H), each identified in four different patients. Dual-labelled assays were designed to measure copies of wild-type and mutated *TP53* alleles

Table 1. Summary statistics.

Feature	Value
Number of patients	40
Age at diagnosis, median (range) (years)	63 (38–85)
Cancer type (number of patients)	
Ovarian	32
Primary peritoneal	5
Fallopian tube	3
Stage at diagnosis (number of patients)	
I	3
II	0
III	27
IV	10
Newly diagnosed untreated ($n_{courses}$ = 7)	
TP53MAF, median (IQR)	0.7% (0.3%–2.0%)
CA-125, median (IQR) (IU/ml)	964 (488–2,909)
Volume, median (IQR) (cm^3)	418 (172–1,770)
Relapsed disease ($n_{courses}$ = 51)	
TP53MAF, median (IQR)	8% (1.2%–22%)
CA-125, median (IQR) (IU/ml)	422 (205–1,108)
Volume, median (IQR) (cm^3)	93.5 (37.0–176.0)

IQR, interquartile range; TP53MAF, *TP53* mutant allele fraction.

doi:10.1371/journal.pmed.1002198.t001

using digital PCR. The length of the amplified regions was limited to effectively amplify fragmented cell-free circulating DNA (median 84 base pairs, range 58–177; S1 Table). The false-positive rates for the ctDNA assays were estimated at one mutated allele per 3,010 wild-type alleles ($<$0.033%, at 95% confidence) for 29 assays, and one mutated allele per 621 wild-type alleles ($<$0.16%, at 95% confidence) for assays for g.12458G>A (p.R175H) and g.12460T>A (p.C176S).

Mutated *TP53* Circulating Tumour DNA Is Frequently Detected Pre-treatment in Patients with High-Grade Serous Ovarian Carcinoma

For the 54 courses of chemotherapy at relapse, pre-treatment samples were available for 51/54 courses from 32 relapsed patients (S4 Fig). Digital PCR was used to measure both total and fractional concentration of ctDNA (TP53MAC and TP53MAF; see Methods). Mutated *TP53* alleles in plasma were detected at TP53MAC $\geq$ 20 AC/ml pre-treatment in 42/51 (82%) of courses from relapsed patients.

Mutated *TP53* alleles in plasma were detected at TP53MAC $\geq$ 20 AC/ml pre-treatment in 6/7 (86%) of newly diagnosed stage IIIC/IV patients. CA-125 was above the institutional upper limit of normal in 100% of relapsed and untreated patients.

Median TP53MAF pre-treatment was 8.0% in patients with recurrent disease, 0.7% in patients with newly diagnosed stage IIIC/IV disease, and 0.2% in four patients after primary surgery (see S4 Table for full description of statistics). TP53MAF correlated with CA-125 ($r = 0.49$, $p < 0.001$). (S5 Table shows all correlations.)

Circulating Tumour DNA Detection and Volume of Disease

We performed volumetric analysis of CT imaging performed prior to treatment for 51 relapsed courses (32 individual patients) and for seven newly diagnosed patients with stage IIIC/IV HGSOC (Fig 2A). One out of 51 relapsed treatment courses was excluded because CT imaging had non-measurable disease. Fig 2B shows the distribution of tumour volume across 57 treatment courses. All relapsed courses with tumour volume > 32 cm^3 had ctDNA detected at TP53MAC ≥ 20 AC/ml except for one patient who had a TP53MAC of 15 AC/ml and disease volume of 50 cm^3. All patients with tumour volume $<$ 20 cm^3 had TP53MAC < 20 AC/ml except for one patient with detected ctDNA with 1 cm^3 of disease in the presence of large volume ascites (see below).

We next considered the possibility that the presence of ascites or pleural effusions could impact the accuracy of correlations between ctDNA and measured tumour volume because volumetric CT measurements were made only for solid disease, and the presence of ascites frequently alters CA-125 levels in clinical practice. We correlated levels of ctDNA with volume of disease in relapsed courses with and without ascites (S6 Table) in a subset of patients with closely matched dates of CT scans and ctDNA, and without drainage of ascites between these two time points. Of 50 relapsed pre-treatment ctDNA samples with matched volumetric CT data, ten were excluded because of a >14-d interval between CT imaging and ctDNA sample collection, and five because of a pleural or ascitic drain performed between the CT scan and plasma collection (S3 Data).

For the 35 remaining courses, TP53MAF showed a positive correlation with volume (Pearson $r = 0.59$, $p < 0.001$; S6 Table). Of these, 13/35 CT images showed the presence of ascites. When cases with ascites were excluded from analysis, the correlation of TP53MAF with volume increased (Pearson $r = 0.82$, $p < 0.001$; S6 Table), indicating that ctDNA in ascites may contribute to blood ctDNA levels. CA-125 was moderately correlated with tumour volume in all 35 cases and in those without ascites (Pearson $r = 0.52$, $p = 0.001$, and $r = 0.51$, $p = 0.016$, respectively; S6 Table). The linear regressions for TP53MAF and CA-125 in cases without ascites are shown in Fig 2C and 2D.

Median TP53MAF/volume in the patients without ascites was 0.08% per cm^3 (IQR 0.02%–0.13% per cm^3). TP53MAC showed lower correlation with tumour volume (S6 Table), and the median value of TP53MAC/volume was 6.0 AC/ml per cm^3 (S3 Data).

Relationship of *TP53* Mutant Allele Fraction to Volume in Recurrent Compared with Untreated Disease

The factors determining ctDNA blood levels are not well understood. Comparison of pre-treatment TP53MAF/volume values between relapsed and untreated patients showed a significant difference ($p = 0.004$, Wilcoxon rank-sum test; Fig 2E). CA-125/volume was not significantly different between these groups ($p = 0.063$; Wilcoxon rank-sum test; Fig 2F). The median TP53MAF/volume was 0.04% per cm^3 in 50 relapsed courses compared with 0.0008% per cm^3 in seven newly diagnosed untreated patients.

Pre-treatment *TP53* Mutant Allele Fraction Is Associated with Time to Progression in Relapsed Patients

We next asked whether ctDNA measured prior to chemotherapy was associated with progression in patients with recurrent disease. We compared TP53MAF, CA-125, and total cell-free DNA (TP53TAC) to TTP estimates. Of the 50 relapse events with measurable disease, one treatment course was excluded because ctDNA was measured >14 d before start of treatment (S4 Fig). The median follow-up was 58 mo (range 43–70 mo), with all patients progressing during the follow-up period.

In univariable analysis, TP53MAF, CA-125, total cell-free DNA (TP53TAC), age, platinum sensitivity, the number of lines of chemotherapy, and volume of disease were all significant

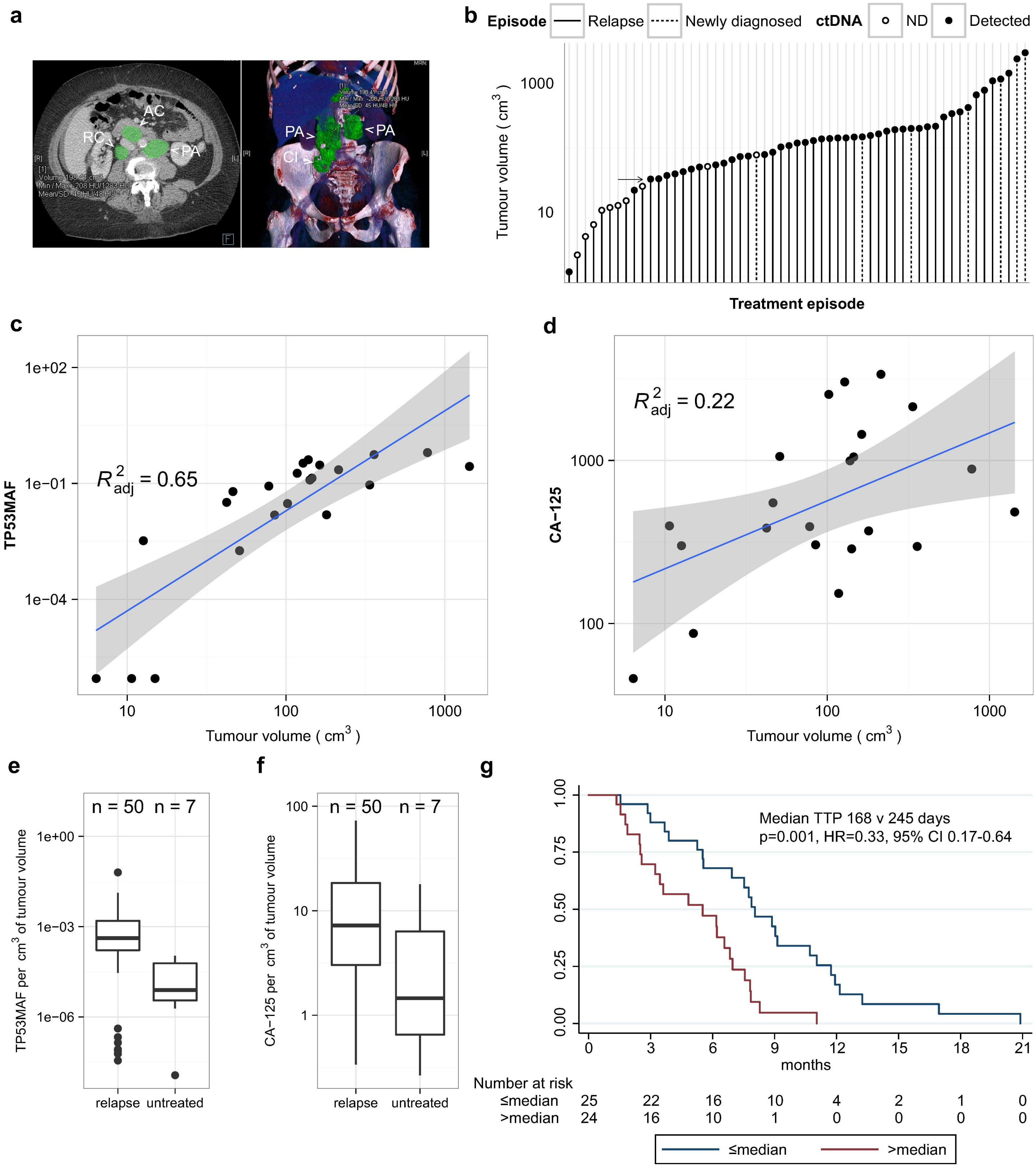


Fig 2. Comparison of *TP53* mutant allele fraction to tumour volume. (A) Example of volumetric analysis of high-grade serous ovarian cancer with relapsed disease in abdominal lymph nodes. Left panel shows cross-sectional view. Right panel shows 3-D reconstruction to show disease volume. Green

shading indicates regions of interest for volume measurements. Lymph node masses are indicated by arrowheads and labelled as follows: AC, aorto-caval; CI, common iliac; PA, para-aortic; RC, retro-caval. (B) Ranked total volume of tumour at start of treatment course. Filled circles indicate cases with TP53MAC $\geq$ 20 AC/ml. Arrow indicates tumour volume of 32 cm^3. (C and D) Linear regression analysis of TP53MAF and CA-125 with tumour volume in 22 relapsed events without ascites. Grey shading shows 95% confidence intervals. (E and F) Comparison of TP53MAF and CA-125 values adjusted for tumour volume between relapsed and newly diagnosed patients before treatment. (G) Time to progression analysis for relapsed patients with greater or less than the median pre-treatment TP53MAF. ctDNA, circulating tumour DNA; HR, hazard ratio; ND, not detected; TP53MAF, *TP53* mutant allele fraction.

doi:10.1371/journal.pmed.1002198.g002

predictors of TTP (Table 2). When adjusted using the Cox proportional hazards model in multivariable analysis, only TP53MAF (hazard ratio [HR] 1.03, 95% CI 1.01–1.06, $p = 0.019$) and platinum sensitivity (HR 0.43, 95% CI 0.19–0.99, $p = 0.048$) remained significant. TTP was significantly longer for treatment courses with pre-treatment levels of TP53MAF below the median level than for treatment courses with TP53MAF above the median ($p = 0.001$ by log-rank test; Fig 2G). Pre-treatment TP53MAC was also tested as a continuous variable and was a significant predictor of TTP in univariable but not multivariable analysis (S10 Table).

Response Kinetics and Nadir of *TP53* Mutant Allele Fraction during Chemotherapy

We next analysed the kinetics of TP53MAF during treatment by measuring the time to achieve a nadir value following chemotherapy cycles. Only a subset of chemotherapy courses had consecutive plasma samples at each cycle and were assessable for nadir (see S7 Table for nadir assessment criteria). Of these, 26/27 courses reached a nadir for TP53MAF, compared with 21 for CA-125 (Fig 3A). The only course without a decrease in TP53MAF occurred in a patient who developed new brain metastases (patient 57). The median time to nadir was 37 d (IQR 28–54) for TP53MAF and 84 d (IQR 42–116) for CA-125. The median decrease at nadir was 98% for TP53MAF and 55% for CA-125 (Fig 3B).

We observed a more rapid decrease and greater dynamic range of TP53MAF measurements compared with CA-125 (Fig 3C). In most cases, TP53MAF and CA-125 trends with sequential treatment cycles were similar. In two cases, however, we observed discrepant kinetics, with rising TP53MAF and decreasing CA-125. In both cases, CT scans confirmed progressive disease, including in patient 57, who had progressive brain metastases (Fig 3D and 3E). The effect of ascitic drainage on plasma TP53MAF level is shown in Fig 3F. Drainage of ascites resulted in a decrease in TP53MAF, demonstrating that drainage may introduce rapid changes in plasma ctDNA level, potentially confounding comparison to response measures. The TP53MAF and CA-125 plots for all patients in the study are available in S3 Fig.

Change in *TP53* Mutant Allele Fraction after One Cycle of Treatment Predicts Time to Progression

Early prediction of response or resistance to chemotherapy could have important implications for the clinical management of relapsed HGSOC patients. We therefore examined whether a decrease in TP53MAF after one cycle of chemotherapy could predict TTP, and compared this with CA-125. There were 32 courses of chemotherapy for relapsed patients that had matched samples at cycle 1 and cycle 2 and were evaluable for response (see S4 Fig for REMARK diagram). Of these, 22 were treated with platinum-based chemotherapy and ten with non-platinum chemotherapy. The median number of days from start of chemotherapy (cycle 1) to collection of plasma sample pre-cycle 2 was 28 d (IQR 21–28). The median TP53MAF decrease from pre-treatment to the pre-cycle-2 sample was 74% (IQR 55%–89%); median CA-125 decrease was 18% (IQR -46%-12%), and the median TTP was 189 d.

Table 2. Univariable and multivariable analysis of pre-treatment values as a predictor of time to progression.

Variable (Units), $n_{courses} = 49$	Univariable			Multivariable		
	HR	CI	*p*-Value	HR	CI	*p*-Value
TP53MAF (percent)	1.04	1.02–1.06	<0.001***	1.03	1.01–1.06	0.019*
CA-125 (10^2 IU/ml)	1.02	1.00–1.05	0.078	1.01	0.98–1.04	0.567
TP53TAC (10^3 copies/ml)	1.03	1.01–1.04	0.009**	1.02	1.00–1.05	0.079
Age (years)	0.96	0.93–1.00	0.030*	0.96	0.92–1.01	0.089
Performance status (0–2)	0.72	0.44–1.18	0.192	0.64	0.36–1.17	0.146
Platinum sensitive (yes/no)	0.35	0.18–0.68	0.002**	0.43	0.19–0.99	0.048*
Number of lines chemotherapy (2 lines/≥3 lines)	0.43	0.22–0.83	0.013	0.75	0.35–1.64	0.478
Volume of disease (10 cm^3)	1.02	1.01–1.03	0.002**	1.00	0.98–1.02	0.786
Ascites (no/yes)	0.95	0.53–1.71	0.858	0.82	0.42–1.58	0.548

For variables with HR > 1, an increase in the value is associated with a higher risk or number of events, and a decreased TTP. For binary variables, the HR listed is for the first option, with the second option being HR = 1 (if the HR listed is <1 then the first option is associated with lower risk and longer TTP).

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

CI, confidence interval; HR, hazard ratio; TP53MAF, *TP53* mutant allele fraction; TP53TAC, *TP53* total allele count; TTP, time to progression.

doi:10.1371/journal.pmed.1002198.t002

To determine the optimal cut-point for predicting TTP from a decrease in TP53MAF, a ROC plot was generated, and 6-mo TTP was selected as a clinically significant endpoint. The ROC curve identified a 60% decrease in TP53MAF as the optimal cut-point for sensitivity and specificity (Fig 4A; S8 Table). This threshold was used in subsequent analyses. Median TTP was 94 d versus 230 d for a decrease of ≤60% and >60% in TP53MAF, respectively, with an HR of 0.22 (95% CI 0.09–0.52, $p < 0.001$; Fig 4B).

In univariable analysis, volume of disease and a decrease in TP53MAF of >60% from pre-treatment to cycle 2 were significant predictors of TTP. In multivariable analysis, a TP53MAF decrease of >60% remained a significant predictive factor for 6-mo TTP (HR 0.22, 95% CI 0.07–0.67, $p = 0.008$; Table 3). CA-125 decrease was not significant.

TP53MAF decrease was also significant as a continuous variable in multivariable analysis (S9 Table). TP53MAC decrease after one cycle was also tested as a continuous variable and was not a significant predictor of TTP (S10 Table). Together with our observation that pre-treatment TP53MAC was not a significant predictor of TTP, these results indicate that TP53MAF was the most informative ctDNA parameter.

As recent ascitic drainage could interfere with response assessment, we carried out the same analysis excluding patients who had had a recent ascitic or pleural drain. When we excluded patients with recent ascitic drain, the ROC cut-point remained 60% (Fig 4C; S8 Table). Median TTP was 76 d versus 229 d for a decrease of ≤60% and >60% in TP53MAF, respectively, and the HR decreased from 0.22 to 0.08 (95% CI 0.02–0.34, $p < 0.001$; Fig 4D). The sensitivity and specificity of a TP53MAF decrease of ≤60% after one cycle of chemotherapy for predicting TTP < 6 mo was 71% and 88%, respectively, in the whole population, and 75% and 100% in patients without ascitic drains (Table 4; see also S8 and S12 Tables).

The predictive value of TP53MAF remained significant when we looked at changes from pre-treatment to pre-cycle 3 of chemotherapy (as compared with changes to pre-cycle 2, above). The optimal cut-point was selected at a decrease of 80% for this time point (S13–S17 Tables). Using an 80% decrease threshold, the HR for TTP after two cycles in multivariable analysis was 0.26 (95% CI 0.07–0.92, $p = 0.037$). Response classification after one cycle of chemotherapy and after two cycles of chemotherapy was consistent (S6 Fig).

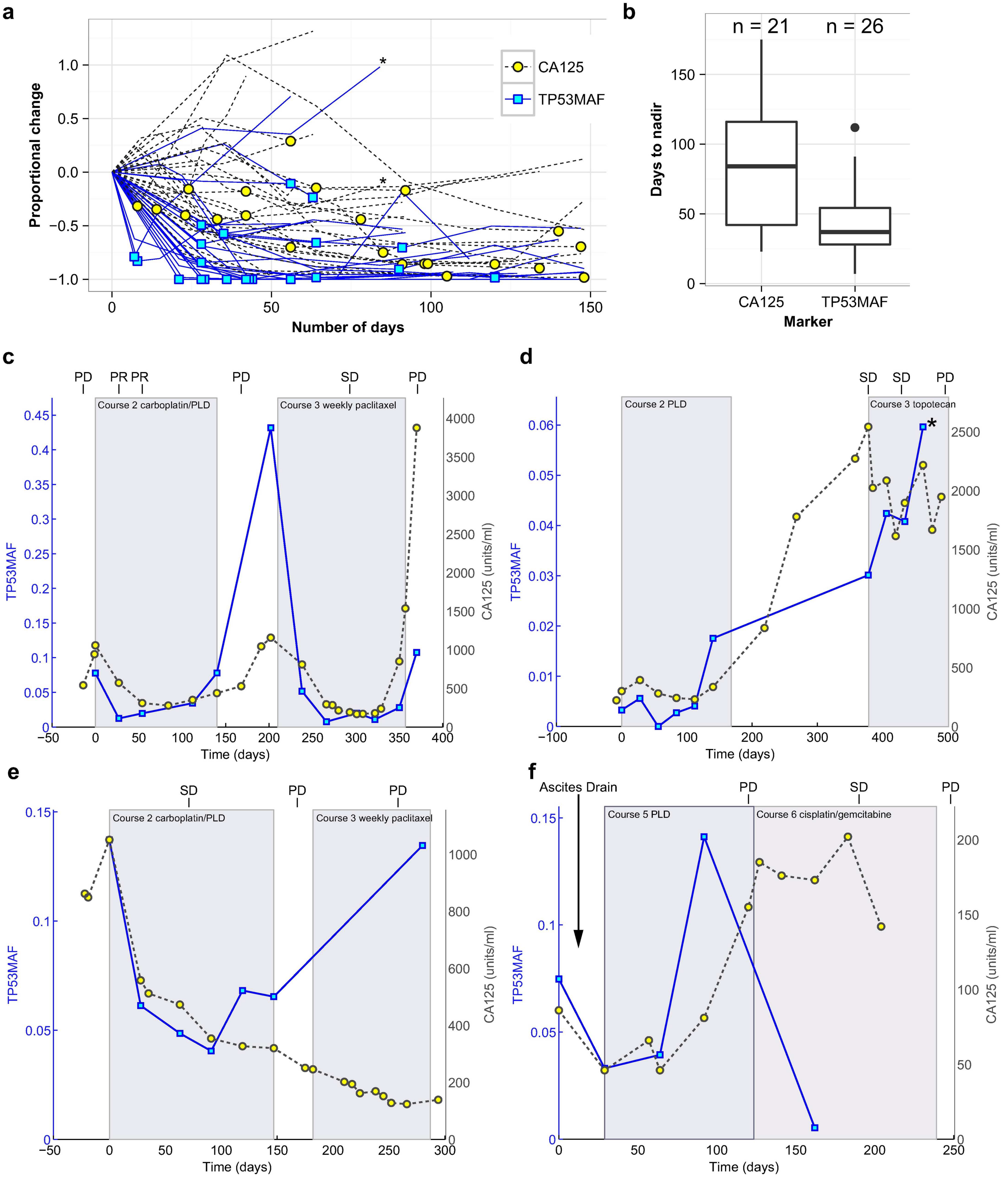

a
Proportional change
1.0
0.5
0.0
−0.5
−1.0
0
50
100
150
Number of days
CA125
TP53MAF
*
*
b
n = 21
n = 26
Days to nadir
150
100
50
0
CA125
TP53MAF
Marker
c
PD
PR
PR
PD
SD
PD
Course 2 carboplatin/PLD
Course 3 weekly paclitaxel
TP53MAF
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
CA125 (units/ml)
4000
3500
3000
2500
2000
1500
1000
500
0
−50
0
50
100
150
200
250
300
350
400
Time (days)
d
SD
SD
PD
Course 2 PLD
Course 3 topotecan
*
TP53MAF
0.06
0.05
0.04
0.03
0.02
0.01
0
CA125 (units/ml)
2500
2000
1500
1000
500
0
−100
0
100
200
300
400
500
Time (days)
e
SD
PD
PD
Course 2 carboplatin/PLD
Course 3 weekly paclitaxel
TP53MAF
0.15
0.1
0.05
0
CA125 (units/ml)
1000
800
600
400
200
0
−50
0
50
100
150
200
250
300
Time (days)
f
Ascites Drain
PD
SD
PD
Course 5 PLD
Course 6 cisplatin/gemcitabine
TP53MAF
0.15
0.1
0.05
0
CA125 (units/ml)
200
150
100
50
0
0
50
100
150
200
250
Time (days)

Fig 3. Circulating tumour DNA and CA-125 kinetics during chemotherapy. (A) TP53MAF and CA-125 kinetics from start of treatment, normalised to the pre-treatment levels. Asterisk denotes one treatment course where the patient developed new brain metastases. Yellow circles and blue boxes indicate nadir points. (B) Box plots show time to nadir following start of chemotherapy for CA-125 and TP53MAF. (C–F) Illustrative cases of TP53MAF and CA-125 kinetics. (C) Faster time to nadir and greater dynamic range of TP53MAF compared with CA-125. (D) Discrepant TP53MAF and CA-125 kinetics. This patient commenced on third-line chemotherapy and had a minor response on CT (stable disease by RECIST). CA-125 fell slightly whilst TP53MAF increased. After cycle 4, the patient developed new headaches, and a CT scan showed new brain metastases (marked by asterisk). (E) Discrepant TP53MAF and CA-125 kinetics. This patient commenced third-line chemotherapy, and the TP53MAF and CA-125 values diverged. CT scan showed progressive disease, in keeping with rise of TP53MAF. (F) The effect of ascitic drainage on plasma TP53MAF levels. This patient had an ascitic drain (at time = 4 d) before starting chemotherapy, with a ctDNA sample taken before (time = 0 d) and after (time = 29 d) the ascitic drain. Following drainage of 8 l of ascites, and before start of any further treatment, TP53MAF fell from 7.5% to 3.3%. CA-125 decreased from 86 IU/ml to 46 IU/ml. This patient had small-volume (1 cm^3) solid disease and large-volume ascites. CA-125, cancer antigen 125; ctDNA, circulating tumour DNA; PD, progressive disease; PLD, pegylated liposomal doxorubicin; PR, partial response; SD, stable disease; TP53MAF, *TP53* mutant allele fraction.

doi:10.1371/journal.pmed.1002198.g003

Discussion

We describe here our analysis of patient-specific *TP53* mutations in ctDNA in women with HGSOC. We used sequence-specific assays to detect predefined *TP53* mutations and to quantify the TP53MAF by digital PCR with sensitivity down to 0.15%. In plasma samples collected prior to treatment for relapsed disease, we were able to detect mutated alleles at ≥20 amplifiable copies/ml in 82% of treatment courses, and in 86% of newly diagnosed patients.

We compared ctDNA to tumour volume, using volumetric analysis of CT images. Although tumour volume appears likely to be an important prognostic factor, this has not been extensively studied [40–44]. In our limited sample set, 3-D tumour volume was significantly associated with TTP in univariable analysis, but did not emerge as significantly associated with TTP in multivariable analysis including TP53MAF, TP53MAC, and other data (Tables 2 and S10). Our data suggest that TP53MAF contains more information on prognosis than CT imaging. This finding agrees with findings from other studies using CT imaging to track metastatic cancer, for example in breast cancer, where a rising level of ctDNA was found to be an earlier indicator of disease relapse than CT imaging [16].

In our study, ctDNA (TP53MAF and TP53MAC) showed significant correlation with disease volume, particularly in patients without ascites. This is consistent with previous studies that showed that ctDNA level increases as stage increases across a range of different tumour types [15]. For example, in recurrent CRC, significant correlations have been demonstrated between pre-treatment ctDNA levels and both RECIST and carcinoembryonic antigen measurements [22]. In addition, in a study of untreated lung cancer, there was significant concordance between ctDNA level and tumour volume in nine patients [17]. The weaker correlation between tumour volume and TP53MAF and TP53MAC levels observed when including patients with ascites, and the rapid change in plasma TP53MAF observed after ascitic drainage (Fig 3F), suggests that ascites fluid may be a reservoir for cell-free tumour DNA. CA-125 was abnormal in all patients at baseline. However, CA-125 has low specificity and may be elevated by any malignant peritoneal process, which may explain its poorer correlation with tumour volume [45].

We found that the ratio of TP53MAF to disease volume was higher in relapsed patients than in newly diagnosed patients. Potential explanations for higher TP53MAF in patients with recurrent disease include disruption of the peritoneum after surgery, differences in tumour biology, different rates of ctDNA release from different metastatic sites, and differences in DNA lifetime in circulation as a result of other physiological changes. With a detection cut-off of 20 AC/ml, ctDNA was detected in all but one relapsed patient with more than 32 cm^3 of disease, and in some patients with lower volume of disease. Previous analysis of tumour volume in lung cancer showed detection of ctDNA with 5–20 cm^3 of disease [17]. These data support the notion that ctDNA has the potential for use in screening and earlier diagnosis of cancer.

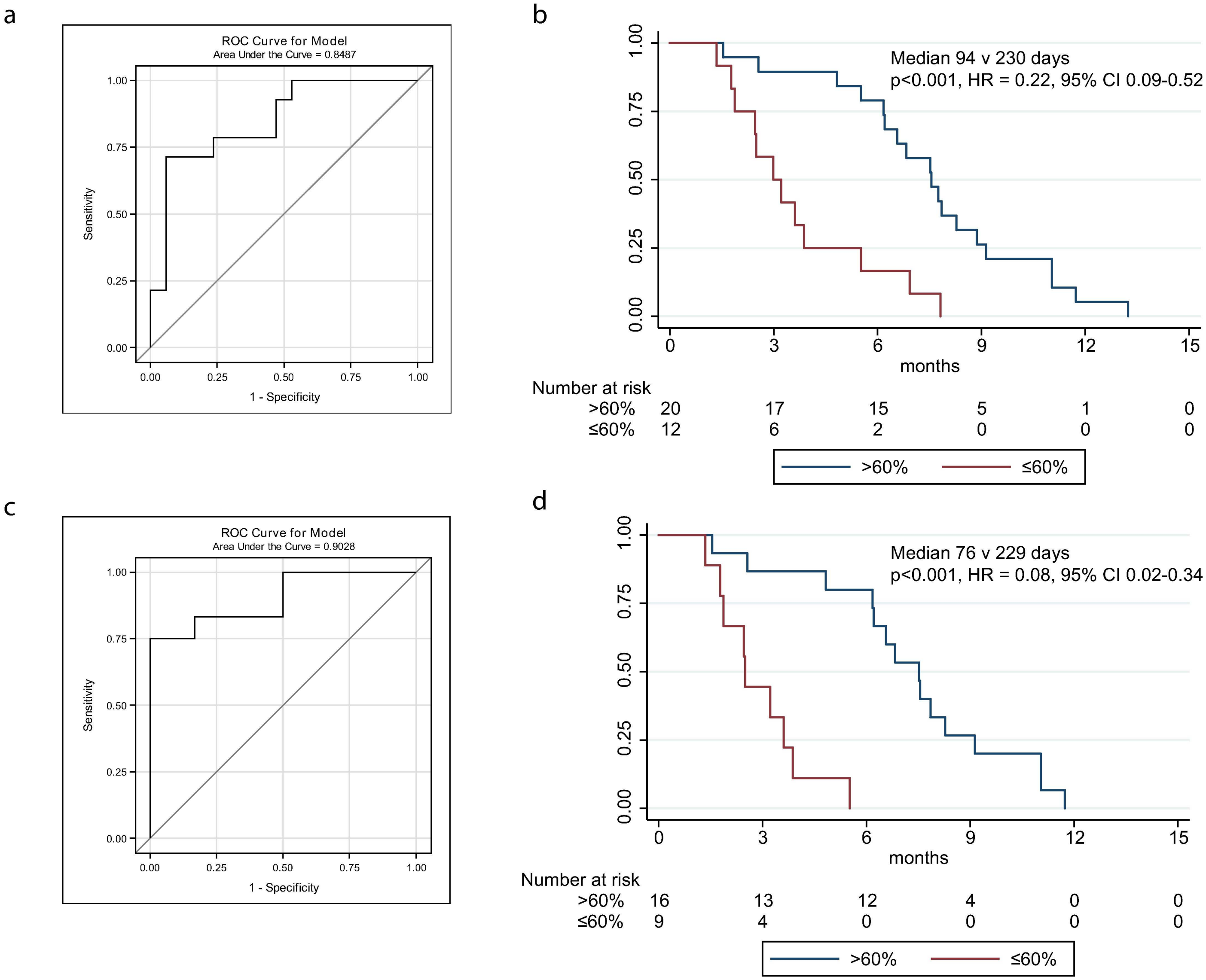


Fig 4. ROC curves and Kaplan-Meier plots for change in circulating tumour DNA after one cycle of chemotherapy, including and excluding courses with recent ascitic drains. (A) ROC plot identifies 60% decrease in TP53MAF as the most accurate threshold for predicting 6-mo TTP in all patients. (B) Kaplan-Meier curve showing TTP for patients with decrease of $\leq$60% or >60% after one cycle of chemotherapy. (C) ROC plot identifies a 60% decrease in TP53MAF as the most accurate threshold for predicting 6-mo TTP in patients without ascitic drains. (D) Kaplan-Meier curve for TP53MAF decrease after one cycle of chemotherapy to predict 6-mo progression-free survival in patients without ascitic drains. ctDNA, circulating tumour DNA; HR, hazard ratio; ROC, receiver operating characteristic; TP53MAF, *TP53* mutant allele fraction; TTP, time to progression.

doi:10.1371/journal.pmed.1002198.g004

For earlier diagnosis in symptomatic women, it may be advantageous to combine ctDNA and CA-125 assays to increase specificity as well as sensitivity.

We examined ctDNA (TP53MAF and TP53MAC) levels prior to the start of treatment as a possible marker for prognosis. Platinum sensitivity, as defined by disease-free or treatment-free interval, is currently the most clinically useful prognostic factor for TTP in relapsed patients [46–50]. We compared TP53MAF and TP53MAC to platinum sensitivity and to established prognostic markers for relapsed ovarian cancer including CA-125 and disease volume. In multivariable analysis, TP53MAF and platinum sensitivity remained significant predictive factors. At present, interventional trials in relapsed ovarian cancer stratify by platinum

Table 3. Univariable and multivariable analysis of 60% decrease in *TP53* mutant allele fraction after one cycle of chemotherapy as a predictor of time to progression.

Variable (Units), $n_{\text{courses}} = 32$	Univariable			Multivariable		
	HR	CI	*p*-Value	HR	CI	*p*-Value
TP53MAF decrease > 60% from C1 to C2 (yes/no)	0.22	0.09–0.52	<0.001***	0.22	0.07–0.67	0.008**
CA-125 decrease > 50% from C1 to C2 (yes/no)	0.58	0.23–1.43	0.234	0.86	0.28–2.71	0.802
Age (years)	1.00	0.95–1.04	0.841	0.97	0.91–1.03	0.276
Performance status (0–2)	0.78	0.35–1.76	0.549	0.74	0.29–1.87	0.522
Platinum sensitive (yes/no)	0.49	0.23–1.02	0.057	0.65	0.25–1.66	0.365
Number of lines chemotherapy (2 lines/$\geq$3 lines)	0.53	0.24–1.16	0.114	0.92	0.30–2.81	0.888
Volume of disease (10 cm^3)	1.02	1.00–1.03	0.031*	1.00	0.98–1.02	0.999
Ascites (no/yes)	1.36	0.66–2.82	0.409	1.76	0.75–4.16	0.197

For variables with HR > 1, an increase in the value is associated with a higher risk or number of events, and a decreased TTP. For binary variables, the HR listed is for the first option, with the second option being HR = 1 (if the HR listed is <1 then the first option is associated with lower risk and longer TTP).

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$.

C1, cycle 1; C2, cycle 2; CI, confidence interval; HR, hazard ratio; TP53MAF, *TP53* mutant allele fraction; TTP, time to progression.

doi:10.1371/journal.pmed.1002198.t003

sensitivity. If our results are confirmed in larger studies, then additional stratification by ctDNA level could increase the accuracy of outcome prediction.

There is also a clinical need for rapid detection of early response to therapy in HGSOC. CA-125, which has been previously evaluated for this purpose, is not sufficiently predictive to be used as a primary endpoint in clinical trials. A large retrospective analysis of the CALYPSO phase III trial in relapsed HGSOC unexpectedly showed that early decline in CA-125 was more likely in the inferior arm, leading the authors to conclude that early change in CA-125 is a poor surrogate for progression-free survival [51]. More recently, alternative criteria for CA-125 response in clinical trials have been explored [52]; however, no reliable model has yet been identified for biomarker-driven trials.

We therefore analysed the dynamics of *TP53* ctDNA during treatment with standard of care chemotherapy, to explore its potential as an early response marker. We found that the

Table 4. Sensitivity and specificity of a decrease in *TP53* mutant allele fraction and CA-125 for predicting 6-mo time to progression following one cycle of chemotherapy.

Predictor	Sensitivity (95% CI)	Specificity (95% CI)	Negative Predictive Value (95% CI)	Positive Predictive Value (95% CI)
TP53MAF: $\leq$60% decrease from cycle 1 to 2				
All ($n_{\text{courses}} = 31$)*	71% (42%–92%)	88% (64%–99%)	79% (54%–94%)	83% (52%–98%)
Excluding drains ($n_{\text{courses}} = 24$)*	75% (43%–95%)	100% (74%–100%)	80% (52%–96%)	100% (66%–100%)
CA-125: $\leq$50% decrease from cycle 1 to 2				
All ($n_{\text{courses}} = 31$)*	93% (66%–100%)	29% (10%–56%)	83% (36%–100%)	52% (31%–72%)
Excluding drains ($n_{\text{courses}} = 24$)*	92% (62%–100%)	33% (9%–65%)	80% (28%–99%)	58% (34%–80%)

*One course of chemotherapy was excluded from the sensitivity/specificity analysis for 6-mo time to progression since it was censored before 6 mo.

CI, confidence interval; TP53MAF, *TP53* mutant allele fraction.

doi:10.1371/journal.pmed.1002198.t004

mean time to nadir was shorter for TP53MAF than for CA-125 (37 d versus 84 d). This may be explained by the longer half-life of CA-125 of 10 d [53] compared with 1–2 h for ctDNA [11,12]. In relapsed treatment courses, a decrease of ≤60% in TP53MAF after one cycle identified a group with progression within 6 mo with high specificity (88% overall; 100% of courses when patients with recent ascitic drains were excluded). In multivariable analysis, TP53MAF was an independent predictor of 6-mo TTP after one and two cycles of chemotherapy. Importantly, patients in the group with poorer prognosis after one cycle did not change group after the second cycle. However, this is a small retrospective study, and variability in the study individuals and their treatments may account for these effects. Future studies are needed to validate a 60% decrease in TP53MAF after one cycle as the optimum clinically useful threshold and to validate other findings from this study.

A similar study in metastatic colon cancer with a prospective design analysed the predictive value of changes in ctDNA after one cycle of chemotherapy in 53 patients [22]. The single ctDNA response marker for each patient was chosen from a panel of genes that are commonly mutated in CRC, and an association with increased progression-free survival was observed in patients with a ≥90% decrease in ctDNA, although this outcome did not reach significance (HR = 1.87, $p = 0.266$). A potential advantage of quantifying *TP53* mutations for response in HGSOC is that *TP53* mutation is the earliest known driver event in HGSOC and is detectable in all metastatic disease [54,55]. By contrast, in other cancers, intratumoural heterogeneity and clonal diversity may reduce the accuracy of using any single mutation in ctDNA as a quantitative measure of tumour burden and as a predictor of response [56,57]. We observed rapid increase in TP53MAF in a patient who developed new cerebral metastasis, suggesting that, in HGSOC, ctDNA changes are not limited to abdominal disease.

These findings need replication but may have a significant impact for patients with HGSOC, particularly if ctDNA can be developed as an early predictor of treatment efficacy. The phase III SWOG S0500 trial tested whether circulating tumour cells could be used as an early predictor of response in metastatic breast cancer, but early switching of therapy did not show improved overall survival [58]. However, the negative results may reflect a lack of active drugs for these patients, or methods of insufficient sensitivity, rather than the usefulness of the predictor. In metastatic breast cancer, ctDNA was shown to have >100-fold higher copy numbers in plasma compared to the number of circulating tumour cells detected by the CELL-SEARCH system, which was the method used in the SWOG S0500 study [16].

The major limitations of this study are its retrospective design, analysis of multiple courses from the same patient, limited sample size and sampling times, and the heterogeneity of treatment within the cohort. In addition, as this was a proof of concept study, we analysed DNA from only a small volume of plasma (median of approximately 0.1 ml per sample) to conserve research material. Even in such a limited volume of plasma, we detected ctDNA at ≥20 AC/ml in >80% of pre-treatment plasma samples and were able to show the association of TP53MAF with TTP. Future studies aiming to validate this approach could further enhance the sensitivity and accuracy of ctDNA measurements by using larger volumes of plasma. The technology for assessment of ctDNA that was used in this study is based on fluorescently labelled patient-specific probes. Such assays can be expensive and time-consuming to design and validate. Since this study was initiated, we and others have demonstrated that suitably designed next generation sequencing assays can be used with high sensitivity both for monitoring ctDNA levels and for direct identification of mutations for genotyping of tumour via plasma sampling [16,21]. The use of such panel assays can obviate the need to design patient-specific assays targeting individual mutations, simplifying the deployment of such approaches for clinical use, and allowing multiple mutations to be monitored simultaneously [17,21].

Recent studies have demonstrated the potential of ctDNA as a tool for minimally invasive real-time molecular profiling, to identify risk of progression based on residual disease, and to identify disease recurrence earlier. In this study, we showed the potential of ctDNA to identify, after 1–2 cycles of treatment, ovarian cancer patients with an expected poor response to chemotherapy. These findings have strong potential for clinical utility owing to the ease of assaying DNA in plasma and the low cost and speed of ctDNA testing. There is therefore a strong rationale for including ctDNA collection in current clinical trials as exploratory endpoints to support clinical validation of ctDNA as a potential early marker of response and prognosis. Having very early information on response would empower patients and physicians to test alternative treatment options and would have high utility in trials that link biomarkers to targeted therapy [59].

Supporting Information

S1 Checklist. STROBE statement—checklist of items that should be included in reports of observational studies.
(DOC)

S1 Data. Circulating tumour DNA CA-125 chemotherapy RECIST and procedures.
(XLSX)

S2 Data. Survival data.
(XLSX)

S3 Data. Volumetric data.
(XLSX)

S1 Fig. REMARK diagram of patient selection for circulating tumour DNA quantitation.
(PPTX)

S2 Fig. Assessment of assay performance in control experiments.
(DOCX)

S3 Fig. Graphs of *TP53* mutant fraction and CA-125 changes by patient.
(PDF)

S4 Fig. REMARK diagram of course selection for analysis of circulating tumour DNA pretreatment and for circulating tumour DNA change after the first and second cycle of chemotherapy.
(PPTX)

S5 Fig. Receiver operating characteristic curve for circulating tumour DNA decrease after two cycles of chemotherapy to predict 6-mo time to progression and Kaplan-Meier curves including and excluding patients with ascitic drains.
(DOCX)

S6 Fig. Consistency of predictive category by *TP53* mutant allele fraction decrease after one and two cycles in all patients and excluding patients with recent ascitic drains.
(DOCX)

S1 Table. Sequence *TP53* mutations, primer sequences, and amplicon size.
(XLSX)

S2 Table. Demographic information by patient.
(XLSX)

S3 Table. Context of circulating tumour DNA samples collected in study relative to chemotherapy cycle. (A) Samples from chemotherapy at relapse. (B) Samples collected during first-line treatment.
(PDF)

S4 Table. Pre-treatment descriptive statistics for *TP53* mutant allele fraction, *TP53* mutant allele count, *TP53* total allele count, and CA-125.
(DOCX)

S5 Table. Pearson correlation of pre-treatment *TP53* mutant allele fraction, *TP53* mutant allele count, *TP53* total allele count, and CA-125 in relapsed patients.
(DOCX)

S6 Table. Pearson correlation of blood parameters with volume of disease in relapsed patients.
(DOCX)

S7 Table. Inclusion criteria for courses of treatment eligible for nadir analysis.
(DOCX)

S8 Table. Sensitivity and specificity by *TP53* mutant allele fraction decrease after one cycle including and excluding courses with recent ascites drains.
(DOCX)

S9 Table. Univariable and multivariable analysis of decrease in *TP53* mutant allele fraction as a continuous variable to predict time to progression after one cycle of chemotherapy.
(DOCX)

S10 Table. Univariable and multivariable analysis of pre-treatment *TP53* mutant allele count as a continuous variable to predict time to progression.
(DOCX)

S11 Table. Univariable and multivariable analysis of decrease in *TP53* mutant allele count as continuous variable to predict time to progression after one cycle of chemotherapy.
(DOCX)

S12 Table. Sensitivity and specificity of CA-125 decrease after one cycle including and excluding courses with recent ascites drains.
(DOCX)

S13 Table. Univariable and multivariable analysis of decrease in *TP53* mutant allele fraction as a continuous variable to predict time to progression after two cycles of chemotherapy.
(DOCX)

S14 Table. Sensitivity and specificity of *TP53* mutant allele fraction decrease after two cycles including and excluding courses with recent ascites drains.
(DOCX)

S15 Table. Univariable and multivariable analysis of *TP53* mutant allele fraction decrease of >80% after two cycles of chemotherapy as a predictor of time to progression.
(DOCX)

S16 Table. Sensitivity and specificity of CA-125 decrease after two cycles including and excluding courses with recent ascites drains.
(DOCX)

S17 Table. Sensitivity and specificity of decrease in *TP53* mutant allele fraction and CA-125 for predicting 6-mo time to progression following one and two cycles of chemotherapy.
(DOCX)

S1 Text. Statistical analysis plan, September 2015.
(DOCX)

Acknowledgments

We would like to acknowledge our patients and their caregivers, and the help and support of the research nurses, trial staff, and the staff at Addenbrooke's Hospital. We thank the University of Cambridge, Cancer Research UK, and Hutchison Whampoa. The Human Research Tissue Bank is supported by the NIHR Cambridge Biomedical Research Centre. In addition, we acknowledge the help of Fiona Collen and Kay Lyons (Addenbrooke's Hospital) in data collection. We thank Keith Burling and the NIHR Cambridge Biomedical Research Centre Core Biochemistry Assay Laboratory for performing CA-125 assays; Sarah Dawson and Lorna Morris for initial statistical discussions; the Biorepository, Bioinformatics, and Genomics Core Facilities of the Cancer Research UK Cambridge Institute for technical support; and Dr. Dana Tsui for artwork and advice.

Author Contributions

Conceptualization: CAP DG NR JDB.

Data curation: CAP DG AMP KS KH NR JDB.

Formal analysis: CAP DG AMP HA SF PM ES WQ NR JDB.

Funding acquisition: NR JDB.

Investigation: CAP DG AMP HB CH HA SF PM ES KS KH MJ-L HME IG.

Methodology: CAP DG NR JDB.

Project administration: DG AMP HB CH KS.

Resources: CAP HB CH IG HME JDB.

Software: NR.

Supervision: NR JDB.

Validation: CAP DG WQ NR JDB.

Visualization: CAP DG NR JDB.

Writing – original draft: CAP DG NR JDB.

Writing – review & editing: CAP DG AMP HB CH HA SF PM ES KS KH IG MJ-L HME WQ NR JDB.

References

1. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013; 10(8):472–84. doi: 10.1038/nrclinonc.2013.110 PMID: 23836314

2. Aebersold R, Anderson L, Caprioli R, Druker B, Hartwell L, Smith R. Perspective: a program to improve protein biomarker discovery for cancer. J Proteome Res. 2005; 4(4):1104–9. doi: 10.1021/pr050027n PMID: 16083259
3. Jacobs I, Oram D, Fairbanks J, Turner J, Frost C, Grudzinskas JG. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol. 1990; 97(10):922–9. PMID: 2223684
4. Sturgeon CM, Duffy MJ, Stenman UH, Lilja H, Brunner N, Chan DW, et al. National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast, and ovarian cancers. Clin Chem. 2008; 54(12):e11–79. doi: 10.1373/clinchem.2008.105601 PMID: 19042984
5. Soletormos G, Duffy MJ, Othman Abu Hassan S, Verheijen RH, Tholander B, Bast RC Jr, et al. Clinical use of cancer biomarkers in epithelial ovarian cancer: updated guidelines from the European Group on Tumor Markers. Int J Gynecol Cancer. 2016; 26(1):43–51. doi: 10.1097/IGC.0000000000000586 PMID: 26588231
6. Buys SS, Partridge E, Black A, Johnson CC, Lamerato L, Isaacs C, et al. Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Randomized Controlled Trial. JAMA. 2011; 305(22):2295–303. doi: 10.1001/jama.2011.766 PMID: 21642681
7. Jacobs IJ, Menon U, Ryan A, Gentry-Maharaj A, Burnell M, Kalsi JK, et al. Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. Lancet. 2016; 387(10022):945–56. doi: 10.1016/S0140-6736(15)01224-6 PMID: 26707054
8. Saygili U, Guclu S, Uslu T, Erten O, Dogan E. The effect of ascites, mass volume, and peritoneal carcinomatosis on serum CA125 levels in patients with ovarian carcinoma. Int J Gynecol Cancer. 2002; 12 (5):438–42. PMID: 12366659
9. Gupta D, Lis CG. Role of CA125 in predicting ovarian cancer survival—a review of the epidemiological literature. J Ovarian Res. 2009; 2:13. doi: 10.1186/1757-2215-2-13 PMID: 19818123
10. Gadducci A, Cosio S, Tana R, Genazzani AR. Serum and tissue biomarkers as predictive and prognostic variables in epithelial ovarian cancer. Crit Rev Oncol Hematol. 2009; 69(1):12–27. doi: 10.1016/j.critrevonc.2008.05.001 PMID: 18595727
11. To EW, Chan KC, Leung SF, Chan LY, To KF, Chan AT, et al. Rapid clearance of plasma Epstein-Barr virus DNA after surgical treatment of nasopharyngeal carcinoma. Clin Cancer Res. 2003; 9(9):3254–9. PMID: 12960110
12. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008; 14(9):985–90. doi: 10.1038/nm.1789 PMID: 18670422
13. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev. 1994; 3(1):67–71. PMID: 8118388
14. Vasioukhin V, Anker P, Maurice P, Lyautey J, Lederrey C, Stroun M. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. Br J Haematol. 1994; 86(4):774–9. PMID: 7918071
15. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014; 6(224):224ra24. doi: 10.1126/scitranslmed.3007094 PMID: 24553385
16. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013; 368(13):1199–209. doi: 10.1056/NEJMoa1213261 PMID: 23484797
17. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014; 20(5):548–54. doi: 10.1038/nm.3519 PMID: 24705333
18. Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. Clin Cancer Res. 2009; 15(6):2076–84. doi: 10.1158/1078-0432.CCR-08-2622 PMID: 19276259
19. Diaz LA Jr, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014; 32 (6):579–86. doi: 10.1200/JCO.2012.45.2011 PMID: 24449238
20. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A. 2005; 102(45):16368–73. doi: 10.1073/pnas.0507904102 PMID: 16258065

21. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012; 4 (136):136ra68. doi: 10.1126/scitranslmed.3003726 PMID: 22649089

22. Tie J, Kinde I, Wang Y, Wong HL, Roebert J, Christie M, et al. Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. Ann Oncol. 2015; 26(8):1715–22. doi: 10.1093/annonc/mdv177 PMID: 25851626

23. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, et al. Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol. 2010; 221(1):49–56. doi: 10.1002/path.2696 PMID: 20229506

24. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474(7353):609–15. doi: 10.1038/nature10166 PMID: 21720365

25. Köbel M, Piskorz AM, Lee S, Lui S, LePage C, Marass F, et al. Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. J Pathol Clin Res. 2016; 2(4):247–58. doi: 10.1002/cjp2.53 PMID: 27840695

26. Bast RC Jr, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest. 1981; 68(5):1331–7. doi: 10.1172/JCI110380 PMID: 7028788

27. Moore RG, MacLaughlan S, Bast RC Jr. Current state of biomarker development for clinical application in epithelial ovarian cancer. Gynecol Oncol. 2010; 116(2):240–5. doi: 10.1016/j.ygyno.2009.09.041 PMID: 19879639

28. Swisher EM, Wollan M, Mahtani SM, Willner JB, Garcia R, Goff BA, et al. Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. Am J Obstet Gynecol. 2005; 193(3 Pt 1):662–7.

29. Otsuka J, Okuda T, Sekizawa A, Amemiya S, Saito H, Okai T, et al. Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. Int J Gynecol Cancer. 2004; 14(3):459–64. doi: 10.1111/j.1048-891x.2004.014305.x PMID: 15228418

30. Pereira E, Camacho-Vanegas O, Anand S, Sebra R, Catalina Camacho S, Garnar-Wortzel L, et al. Personalized circulating tumor DNA biomarkers dynamically predict treatment response and survival in gynecologic cancers. PLoS ONE. 2015; 10(12):e0145754. doi: 10.1371/journal.pone.0145754 PMID: 26717006

31. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med. 2007; 4(10):e296. doi: 10.1371/journal.pmed.0040296 PMID: 17941714

32. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. Reporting recommendations for tumor marker prognostic studies (REMARK). J Natl Cancer Inst. 2005; 97(16):1180–4. doi: 10.1093/jnci/dji237 PMID: 16106022

33. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009; 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026 PMID: 19097774

34. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006; 314(5797):268–74. doi: 10.1126/science.1133427 PMID: 16959974

35. Archibald KM, Kulbe H, Kwong J, Chakravarty P, Temple J, Chaplin T, et al. Sequential genetic change at the TP53 and chemokine receptor CXCR4 locus during transformation of human ovarian surface epithelium. Oncogene. 2012; 31(48):4987–95. doi: 10.1038/onc.2011.653 PMID: 22266861

36. Rustin GJ, Quinn M, Thigpen T, du Bois A, Pujade-Lauraine E, Jakobsen A, et al. Re: new guidelines to evaluate the response to treatment in solid tumors (ovarian cancer). J Natl Cancer Inst. 2004; 96 (6):487–8.

37. No JH, Kim K, Park KH, Kim YB. Cell-free DNA level as a prognostic biomarker for epithelial ovarian cancer. Anticancer Res. 2012; 32(8):3467–71. PMID: 22843932

38. Tissot C, Toffart AC, Villar S, Souquet PJ, Merle P, Moro-Sibilot D, et al. Circulating free DNA concentration is an independent prognostic biomarker in lung cancer. Eur Respir J. 2015; 46(6):1773–80. doi: 10.1183/13993003.00676-2015 PMID: 26493785

39. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016 [cited 2016 Nov 18]. Available from: https://www.R-project.org/.

40. Rockall AG, Avril N, Lam R, Iannone R, Mozley PD, Parkinson C, et al. Repeatability of quantitative FDG-PET/CT and contrast-enhanced CT in recurrent ovarian carcinoma: test-retest measurements for tumor FDG uptake, diameter, and volume. Clin Cancer Res. 2014; 20(10):2751–60. doi: 10.1158/1078-0432.CCR-13-2634 PMID: 24573555

41. Andreopoulou E, Andreopoulos D, Adamidis K, Fountzila-Kalogera A, Fountzilas G, Dimopoulos MA, et al. Tumor volumetry as predictive and prognostic factor in the management of ovarian cancer. Anticancer Res. 2002; 22(3):1903–8. PMID: 12168891

42. Eisenhauer EA, Vermorken JB, van Glabbeke M. Predictors of response to subsequent chemotherapy in platinum pretreated ovarian cancer: a multivariate analysis of 704 patients. Ann Oncol. 1997; 8 (10):963–8. PMID: 9402168

43. Mayoral M, Fernandez-Martinez A, Vidal L, Fuster D, Aya F, Pavia J, et al. Prognostic value of (18)F-FDG PET/CT volumetric parameters in recurrent epithelial ovarian cancer. Rev Esp Med Nucl Imagen Mol. 2016; 35(2):88–95. doi: 10.1016/j.remn.2015.08.005 PMID: 26541072

44. Vargas HA, Burger IA, Goldman DA, Micco M, Sosa RE, Weber W, et al. Volume-based quantitative FDG PET/CT metrics and their association with optimal debulking and progression-free survival in patients with recurrent ovarian cancer undergoing secondary cytoreductive surgery. Eur Radiol. 2015; 25(11):3348–53. doi: 10.1007/s00330-015-3729-9 PMID: 25916387

45. Van Gorp T, Cadron I, Despierre E, Daemen A, Leunen K, Amant F, et al. HE4 and CA125 as a diagnostic test in ovarian cancer: prospective validation of the Risk of Ovarian Malignancy Algorithm. Br J Cancer. 2011; 104(5):863–70. doi: 10.1038/sj.bjc.6606092 PMID: 21304524

46. Blackledge G, Lawton F, Redman C, Kelly K. Response of patients in phase II studies of chemotherapy in ovarian cancer: implications for patient treatment and the design of phase II trials. Br J Cancer. 1989; 59(4):650–3. PMID: 2713253

47. Markman M, Rothman R, Hakes T, Reichman B, Hoskins W, Rubin S, et al. Second-line platinum therapy in patients with ovarian cancer previously treated with cisplatin. J Clin Oncol. 1991; 9(3):389–93. doi: 10.1200/jco.1991.9.3.389 PMID: 1999708

48. Markman M, Markman J, Webster K, Zanotti K, Kulp B, Peterson G, et al. Duration of response to second-line, platinum-based chemotherapy for ovarian cancer: implications for patient management and clinical trial design. J Clin Oncol. 2004; 22(15):3120–5. doi: 10.1200/JCO.2004.05.195 PMID: 15284263

49. Lee CK, Simes RJ, Brown C, Gebski V, Pfisterer J, Swart AM, et al. A prognostic nomogram to predict overall survival in patients with platinum-sensitive recurrent ovarian cancer. Ann Oncol. 2013; 24 (4):937–43. doi: 10.1093/annonc/mds538 PMID: 23104722

50. Eng KH, Hanlon BM, Bradley WH, Szender JB. Prognostic factors modifying the treatment-free interval in recurrent ovarian cancer. Gynecol Oncol. 2015; 139(2):228–35. doi: 10.1016/j.ygyno.2015.09.011 PMID: 26383827

51. Lee CK, Friedlander M, Brown C, Gebski VJ, Georgoulopoulos A, Vergote I, et al. Early decline in cancer antigen 125 as a surrogate for progression-free survival in recurrent ovarian cancer. J Natl Cancer Inst. 2011; 103(17):1338–42. doi: 10.1093/jnci/djr282 PMID: 21840849

52. Wilbaux M, Henin E, Oza A, Colomban O, Pujade-Lauraine E, Freyer G, et al. Dynamic modeling in ovarian cancer: an original approach linking early changes in modeled longitudinal CA-125 kinetics and survival to help decisions in early drug development. Gynecol Oncol. 2014; 133(3):460–6. doi: 10.1016/j.ygyno.2014.04.003 PMID: 24726614

53. Yoshikawa T, Takano M, Kita T, Kudoh K, Sasaki N, Kato M, et al. Normal serum CA125 half-life and normal serum nadir CA125 level in patients with ovarian cancers. Eur J Gynaecol Oncol. 2012; 33 (3):269–73. PMID: 22873097

54. Chien J, Sicotte H, Fan JB, Humphray S, Cunningham JM, Kalli KR, et al. TP53 mutations, tetraploidy and homologous recombination repair defects in early stage high-grade serous ovarian cancer. Nucleic Acids Res. 2015; 43(14):6945–58. doi: 10.1093/nar/gkv111 PMID: 25916844

55. Schwarz RF, Ng CK, Cooke SL, Newman S, Temple J, Piskorz AM, et al. Spatial and temporal heterogeneity in high-grade serous ovarian cancer: a phylogenetic analysis. PLoS Med. 2015; 12(2): e1001789. doi: 10.1371/journal.pmed.1001789 PMID: 25710373

56. Jamal-Hanjani M, Wilson GA, Horswell S, Mitter R, Sakarya O, Constantin T, et al. Detection of ubiquitous and heterogeneous mutations in cell-free DNA from patients with early-stage non-small-cell lung cancer. Ann Oncol. 2016; 27(5):862–7. doi: 10.1093/annonc/mdw037 PMID: 26823523

57. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science. 2014; 346(6206):251–6. doi: 10.1126/science.1253462 PMID: 25301630

58. Smerage JB, Barlow WE, Hortobagyi GN, Winer EP, Leyland-Jones B, Srkalovic G, et al. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. J Clin Oncol. 2014; 32(31):3483–9. doi: 10.1200/JCO.2014.56.2561 PMID: 24888818

59. Wason JM, Abraham JE, Baird RD, Gournaris I, Vallier AL, Brenton JD, et al. A Bayesian adaptive design for biomarker trials with linked treatments. Br J Cancer. 2015; 113(5):699–705. doi: 10.1038/bjc.2015.278 PMID: 26263479

PLOS | MEDICINE

RESEARCH ARTICLE

Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis

Roland F. Schwarz[1¤], Charlotte K. Y. Ng[1,2], Susanna L. Cooke[1], Scott Newman[1], Jillian Temple[1], Anna M. Piskorz[1], Davina Gale[1], Karen Sayal[1], Muhammed Murtaza[1], Peter J. Baldwin[3], Nitzan Rosenfeld[1,2], Helena M. Earl[2,4], Evis Sala[5], Mercedes Jimenez-Linan[3], Christine A. Parkinson[1,2], Florian Markowetz[1,2]*, James D. Brenton[1,2]*

1 Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom, **2** Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, Cambridge, United Kingdom, **3** Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, **4** NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom, **5** University Department of Radiology, Addenbrooke's Hospital, Cambridge, United Kingdom

¤ Current address: European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom
* florian.markowetz@cruk.cam.ac.uk (FM); james.brenton@cruk.cam.ac.uk (JDB)

OPEN ACCESS

Citation: Schwarz RF, Ng CKY, Cooke SL, Newman S, Temple J, Piskorz AM, et al. (2015) Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis. PLoS Med 12(2): e1001789. doi:10.1371/journal.pmed.1001789

Editor: Christopher Kemp, Fred Hutchinson Cancer Research Center, UNITED STATES

Received: July 5, 2013

Accepted: January 8, 2015

Published: February 24, 2015

Copyright: © 2015 Schwarz et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant clinical data and analyses are within the paper and its Supporting Information files. Array data is available at the NCBI Gene Expression Omnibus under accession number GSE40546. http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40546

Funding: This work was supported by Cancer Research UK [grant numbers A15601, A17197, A18072]; the University of Cambridge; National Institute for Health Research Cambridge Biomedical Research Centre; Cambridge Experimental Cancer Medicine Centre and Hutchison Whampoa Limited. CAP was supported in part by the Academy of

Abstract

Background

The major clinical challenge in the treatment of high-grade serous ovarian cancer (HGSOC) is the development of progressive resistance to platinum-based chemotherapy. The objective of this study was to determine whether intra-tumour genetic heterogeneity resulting from clonal evolution and the emergence of subclonal tumour populations in HGSOC was associated with the development of resistant disease.

Methods and Findings

Evolutionary inference and phylogenetic quantification of heterogeneity was performed using the MEDICC algorithm on high-resolution whole genome copy number profiles and selected genome-wide sequencing of 135 spatially and temporally separated samples from 14 patients with HGSOC who received platinum-based chemotherapy. Samples were obtained from the clinical CTCR-OV03/04 studies, and patients were enrolled between 20 July 2007 and 22 October 2009. Median follow-up of the cohort was 31 mo (interquartile range 22–46 mo), censored after 26 October 2013. Outcome measures were overall survival (OS) and progression-free survival (PFS). There were marked differences in the degree of clonal expansion (CE) between patients (median 0.74, interquartile range 0.66–1.15), and dichotimization by median CE showed worse survival in CE-high cases (PFS 12.7 versus 10.1 mo, $p = 0.009$; OS 42.6 versus 23.5 mo, $p = 0.003$). Bootstrap analysis with resampling showed that the 95% confidence intervals for the hazard ratios for PFS and OS in the CE-high group were greater than 1.0. These data support a relationship between heterogeneity and survival but do not precisely determine its effect size. Relapsed tissue was available for

Medical Sciences, the Wellcome Trust, British Heart Foundation and Arthritis Research UK. The Human Research Tissue Bank is supported by the NIHR Cambridge Biomedical Research Centre. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: CE, clonal expansion; CNA, copy number alteration; HGSOC, high-grade serous ovarian cancer; HR, hazard ratio; IQR, interquartile range; MEDICC, Minimum Event Distance for Intra-tumour Copy Number Comparisons; OS, overall survival; PFS, progression-free survival; TH, temporal heterogeneity; WGS, whole genome sequencing.

two patients in the CE-high group, and phylogenetic analysis showed that the prevalent clonal population at clinical recurrence arose from early divergence events. A subclonal population marked by a *NF1* deletion showed a progressive increase in tumour allele fraction during chemotherapy.

Conclusions

This study demonstrates that quantitative measures of intra-tumour heterogeneity may have predictive value for survival after chemotherapy treatment in HGSOC. Subclonal tumour populations are present in pre-treatment biopsies in HGSOC and can undergo expansion during chemotherapy, causing clinical relapse.

Introduction

Intra-tumour genetic heterogeneity in cancer has been investigated for almost half a century [1,2], and recent advances in genomic technology have demonstrated diverse genetic changes within a single epithelial cancer [3–14]. Multiple sampling of primary and metastatic sites in breast, pancreas, and renal carcinoma has catalogued genetic divergence and shown that metastases from the same site can show organ-specific phylogenetic branches [5–8,12]. Deep sequencing of epithelial tumours has revealed the clonal compositions of individual clinical samples and has shown how major and minor subclones may co-exist [5,7,8,12,14,15]. These data extend earlier observations showing that there is significant intra-tumour heterogeneity in solid tumours, and suggest that tumours with sufficient heterogeneity may be able to explore the fitness landscape widely enough during selection pressure from chemotherapy to repopulate with a resistant subclone [16,17]. Although this phenomenon has been extensively demonstrated in haematological cancers [18,19], the sequence of clonal expansions (CEs) has not been comprehensively described in epithelial tumours or correlated with clinical outcome.

High-grade serous ovarian cancer (HGSOC) is genomically characterised by a ubiquitous *TP53* mutation, high-frequency somatic copy number alterations (CNAs), and whole genome duplications [20–22]. Oncogenic mutations are rare, and most nonsynonymous changes are seen in tumour suppressor genes, including somatic mutations in *TP53*, *BRCA1*, *BRCA2*, *RB1*, and *NF1* [22]. Loss of *NF1*, an inhibitor of RAS signalling, may occur by point mutation or structural rearrangement and may be present in subclonal populations [23–27]. The typical clinical presentation of HGSOC is with extensive abdominal disease, involving multiple implantation sites throughout the abdomen. Intra-tumour heterogeneity may contribute to acquired resistance in HGSOC [3,4,28–30], but quantitation of the degree of heterogeneity and its relationship to changes in the course of treatment or the development of resistance is unknown.

Accurately reconstructing the evolutionary history of cancer cells in an unbiased manner improves the quantification of tumour heterogeneity. However, inferring phylogenetic trees in cancers that have highly frequent somatic CNAs is particularly difficult because of the unknown phasing of parental alleles and the horizontal dependencies between adjacent genomic loci. Previous work has used ad hoc thresholds or visual analysis [15,31]. We recently developed the Minimum Event Distance for Intra-tumour Copy Number Comparisons (MEDICC) algorithm, which provides accurate estimates of evolutionary distances between tumour samples by determining the optimum phasing of major and minor alleles from copy number or whole genome sequencing (WGS) data [32]. A numerical measure of the degree of heterogeneity can also be derived by transforming the pairwise minimum event distances [32].

To address the hypothesis that quantitative measures of intra-tumour heterogeneity could predict outcome in HGSOC, we collected multiple spatially and temporally separated tumour samples from 14 women undergoing chemotherapy for HGSOC, and used formal methods to reconstruct the evolutionary history of the disease within each patient from whole genome copy number profiles.

Materials and Methods

Ethical Consent

Tissue samples were obtained from the prospective CTCR-OV03 and CTCR-OV04 clinical studies, which were designed to collect imaging, blood, and tissue samples for exploratory biomarker studies. All patients provided written, informed consent for participation in these studies and for the use of their donated tissue, blood specimens, and anonymized data for the laboratory studies carried out. The CTCR-OV03 and CTCR-OV04 studies were approved by the Suffolk Local Research Ethics Committee (reference 05/Q0102/160) and Cambridgeshire Research Ethics Committee (reference 08/H0306/61).

SNP Arrays

DNA extraction was performed using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. In total, 177 samples from 18 patients were profiled for copy number aberrations using Affymetrix Genome-Wide SNP 6.0 arrays (S1 Table). Hybridisation of DNA to SNP 6.0 arrays was performed by AROS Applied Biotechnology following the manufacturer's protocol. Array data are available online at the NCBI Gene Expression Omnibus under accession number GSE40546. The datasets were segmented using PICNIC [33] (using the "primary" option), which further corrects for cellularity and estimates integer major and minor copy numbers.

Evolutionary Inference and Tree Robustness

The MEDICC algorithm and methods for copy number reconstruction and quantification of heterogeneity have been described previously [32]. We determined the CE and temporal heterogeneity (TH) indices as described for patients with more than three samples and where paired biopsy and surgery samples were available.

Paired-End Sequencing and Breakpoint Validation

DNA extracted from tumour samples and from matched normal blood was processed using the Illumina Paired-End Sample Prep Kit. Paired-end WGS (41 bp; in some cases 50 bp trimmed to 41 bp) was performed on the Illumina Genome Analyzer IIx, where the median number of read pairs for each library was 153 million and the median sequencing depth was ×16.7. Sequencing data were processed using analysis pipelines as previously described [34]. Briefly, reads were aligned using BWA [35] and Novoalign (Novocraft Technologies), and discordantly mapped read pairs were used to identify putative structural variants using a custom pipeline. PCR primers for validating structural variants were designed using Primer3 [36].

Deletion and insertion breakpoints from WGS were considered confirmed if there was >50% reciprocal overlap of copy number decrease or increase in the SNP array data. Additionally, deletion, insertion, inter-chromosomal, and inversion breakpoints were considered confirmed if both ends of a breakpoint were within 10 kb of copy number breakpoints in any of the sequenced samples of the tumour.

Mutation Detection by Resequencing

The coding sequences of *TP53*, *BRCA1*, *BRCA2*, *PTEN*, *PIK3CA*, *EGFR*, and *APC* were sequenced using the TAm-Seq method using the Fluidigm Access Array 48.48 platform as described previously [37] with minor modifications: pre-amplification steps were omitted, as high-molecular-weight DNA was extracted from fresh-frozen tumour specimens, and 50 ng of sample DNA and multiplexed primers was used in the target-specific amplification step. Primer sequences are available upon request. Sequencing data analysis and variant verification was performed using an in-house-developed pipeline and IGV software as previously described [37,38].

Digital PCR

Digital PCR was performed using the Fluidigm Biomark microfluidic system according to the manufacturer's instructions. Primers were designed spanning the *NF1* deletion (forward: 5′-TTTTGTTTACGAGCACAGATAACC-3′; reverse: 5′-GAAACAGAAGATGACAGCAAA-GAA-3′). Reaction mixes were prepared containing 1× TaqMan Gene Expression Master Mix (Applied Biosystems), 1× EvaGreen DNA binding dye (Biotium), 1× DNA Binding Dye Sample Loading Reagent (Fluidigm), and 10 nM primers and template DNA. Prior to loading into a 12.765 Fluidigm digital chip, reactions were heated to 95°C for 1 min and placed on ice. Reactions were thermocycled at 50°C for 2 min, 95°C for 10 min, followed by 55 cycles of 95°C for 15 s and 56°C or 60°C for 1 min. Digital PCR was also performed on the same samples using an assay for the p.R175H mutation in *TP53* as previously described [39] (forward: 5′-CCATC-TACAAGCAGTCAC-3′; reverse: 5′-GTCACCATCGCTATCTGAG-3′; mutant-specific probe: [6FAM]-TTGTGAGGCACTGCCCCC-[BHQ1]; wild-type-specific probe: [HEX]-TTGTGAG-GCGCTGCCCCC-[BHQ1]). The proportion of tumour cells with the *NF1* deletion was calculated from the estimated counts of the assays for both the *NF1* deletion and mutant *TP53* p.R175H.

Study Design

Cases were retrospectively selected from available tumour samples from the CTCR-OV03 and CTCR-OV04 studies. The CTCR-OV03 study has been previously described [40] and was a prospective, single-institution, protocol-driven study with eligibility criteria of (a) clinical diagnosis of advanced ovarian cancer (International Federation of Gynecology and Obstetrics stage 3 or higher), (b) gynaecology-oncology multidisciplinary team recommendation for neoadjuvant chemotherapy treatment before interval debulking surgery, (c) measurable disease at staging based on computed tomography of the abdomen and pelvis, and (d) no contraindications to MRI [40]. Samples were stored in RNAlater (Life Technologies) immediately after acquisition and later histologically examined and scored for cellularity by a specialist gynaecological pathologist (M J.-L.). When selecting cases for analysis, 14/28 CTCR-OV03 patients were excluded from the planned analysis (six had tumours that were not HGSOC histology, five had no research tissue available, and three had samples from only one time point or one metastatic site), leaving 14 for genomic profiling. Four additional "convenience" cases were obtained from the CTCR-OV04 study and were selected by availability of ≥3 spatially discrete, fresh-frozen tumour biopsies. All patients were enrolled between 20 July 2007 and 22 October 2009. The follow-up was censored after 26 October 2013, with median duration of 31 mo (interquartile range [IQR] 22–46 mo; range 7–53 mo). Progression-free survival (PFS) was defined as the interval between the date of the original pathology report confirming ovarian cancer and the date of progression measured by RECIST 1.1, Gynecologic Cancer Intergroup CA 125 criteria, or symptomatic progression. Overall survival (OS) was defined as the interval between the date of the original pathology report confirming ovarian cancer and the date of death from any cause. Clinical details, including CA 125 measurements, stage, and residual disease after debulking surgery, were abstracted from clinical records by research staff.

Statistical Analysis Methods

A detailed transcript of all statistical analyses using R and Sweave is provided in S1 Protocol.

Results

We collected 177 temporally and spatially distinct HGSOC samples from 18 patients undergoing platinum-based chemotherapy (Fig. 1A). Copy number profiles were obtained with Affymetrix Genome-Wide SNP6.0 arrays (Table 1) and segmented using PICNIC [33]. Of the 18 patients, 17 had neoadjuvant chemotherapy (Table 1). The median number of chemotherapy cycles prior to interval debulking surgery was three (range 3–7). One patient had primary surgery followed by adjuvant chemotherapy. Data from 39/177 arrays were excluded after profiling because of tumour cellularity $< 50\%$ or high noise, resulting in removal of one patient from the analysis (6/6 samples excluded), leading to a final total of 17 patients included in the following analyses. In all, 31/39 excluded arrays were from samples obtained from interval debulking surgery following pre-operative chemotherapy (S1 Table). Analyses of clonal evolution were performed using whole genome bi-allelic integer copy number profiles of 135 tumour samples from 14 patients who had $\geq$3 samples. Ten patients had samples both from biopsy prior to chemotherapy treatment and from interval debulking surgery, allowing for comparison of temporal effects. Two patients had relapsed ascites samples (S1 Table).

To exclude potential confounding effects on heterogeneity, we performed a detailed pathology review of all paraffin blocks from each patient. No significant differences in morphology or growth pattern were observed between metastatic sites in any patient (S1 Fig). In addition, we performed tagged-amplicon resequencing of tumour tissue for genes commonly somatically mutated in HGSOC. *BRCA1* and *BRCA2* were also included in the sequencing panel as OS is significantly improved in women with germ line mutations. In all, 15/17 patients had a mutation in *TP53* consistent with HGSOC (Table 2) [20]. Of the two wild-type *TP53* cases, patient 12 had strong nuclear p53 protein accumulation consistent with p53 dysfunction, and patient 3 was reclassified as a synchronous HGSOC and high-grade uterine serous papillary carcinoma (based on simultaneous invasive uterine and ovarian lesions together with positive WT1 immunohistochemistry). No germ line mutations in *BRCA1* and *BRCA2* were identified. Patient 14 had a nonsense mutation in *BRCA2*, and patient 2 showed variants of uncertain significance in *BRCA1* somatic nonsense mutation in *BRCA2*, and patient 2 showed variants of uncertain significance in *BRCA1* and *BRCA2* (Table 2). We examined the copy number profiles for evidence of specific mutator phenotypes that would alter the propensity for evolutionary change, but did not find any patient with features of the tandem duplicator phenotype [34,41].

We recently developed the MEDICC package [32] to reconstruct patient-specific evolutionary trees and quantify heterogeneity in tumour samples using methods that employ a minimum evolution criterion to measure the genetic divergence between copy number profiles. This algorithm estimates evolutionary distances between samples based on the minimum number of segmental amplifications and deletions needed to transform one genomic profile into another using optimised allele-specific assignment of major and minor copy numbers. Using

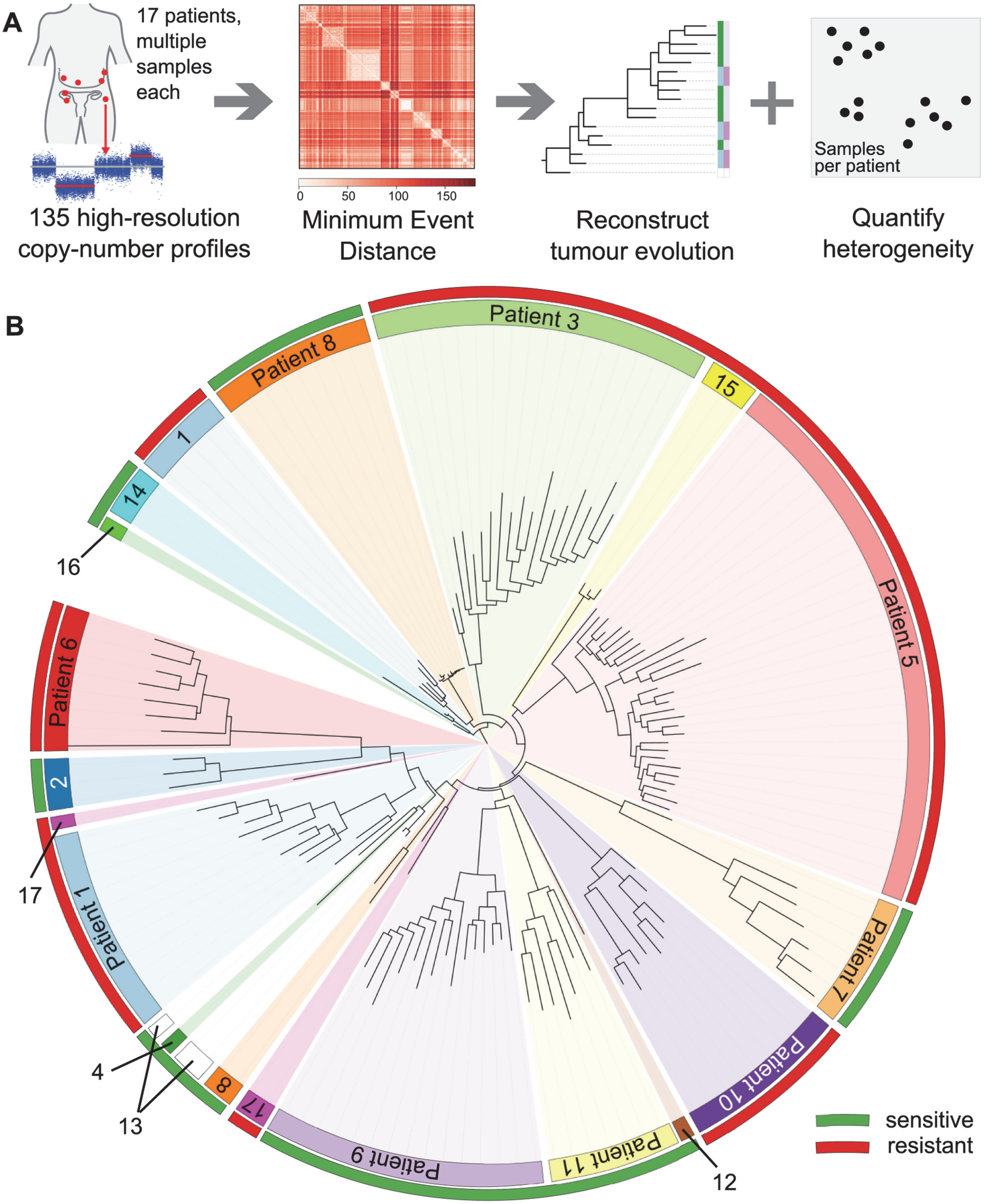

A
17 patients, multiple samples each
135 high-resolution copy-number profiles
0
50
100
150
Minimum Event Distance
Reconstruct tumour evolution
Samples per patient
Quantify heterogeneity
B
Patient 3
15
Patient 5
Patient 7
Patient 10
12
Patient 11
Patient 9
17
8
13
4
Patient 1
17
2
Patient 6
16
14
1
Patient 8
sensitive
resistant

Fig 1. Overview of the analysis and the clinical dataset. (A) Numerical quantification of intra-tumour genetic heterogeneity by evolutionary comparisons. Copy number profiles from 135 metastatic sites were obtained for 17 patients with HGSOC. The MEDICC algorithm was used to compute minimum event distances between profiles and to reconstruct the evolutionary history for each sample, enabling numerical quantification of both spatial and TH and CE for each patient. (B) HGSOC exhibits significant patient-specific intra-tumour genetic heterogeneity. Neighbour-joining tree of all samples based on total copy

number. Samples from each patient (coloured inner circular bar) cluster into clades. The outer circular bar indicates HGSOC classified as resistant versus sensitive to treatment based on survival: red, resistant, PFS < 12 mo; green, sensitive, PFS > 12 mo. No immediate clustering of patients into sensitive and resistant subgroups is visible.

doi:10.1371/journal.pmed.1001789.g001

MEDICC we reconstructed evolutionary trees for 14 patients with three or more samples (Fig. 1A). Circos plots and evolutionary trees for all patients can be found in S3–S16 Figs.

Patient-Specific Copy Number Changes Cluster by Anatomical Site

As expected, CNAs were highly patient-specific. A tree reconstructed from applying MEDICC to all 135 samples grouped samples by patient (Fig. 1B). However, no clustering into subgroups with HGSOC resistant or sensitive to a second line of treatment was evident by this analysis (outer colour bar, Fig. 1B). We next applied MEDICC to each patient individually. Copy number changes within each patient showed strong divergence overall, with a median of 54.5 (IQR 32.5–65.6) genomic events. Divergence from the hypothetical normal genome was, as expected, larger, with a median of 104 events (IQR 57.8–112). Three patients showed less marked changes; reconstruction of events for patients 1 and 8 had limited phylogenetic signal owing to

Table 1. Summary of samples from the CTCR-OV03/04 clinical studies.

Patient Number	Patient Age[a]	Tumour Stage	Response[b]	Treatment	CA 125 Reduction[c]	PFS	OS	Number of Samples[d]	TH Index	CE Index	*p*-Value for Star Topology[e]
1	I	IV	PR	Neo	−93	271	511	16/20	4.73	1.26	<0.001
2	IV	IV	PR	Neo	−92	363	977	3/5	NA	0.71	0.67
3	V	IV	SD	Neo	−92	153	209	18/20	3.74	1.24	<0.001
4	I	IIIC	PR	Neo	−91	616	625	1/3	NA	NA	NA
5	IV	IV	PR	Neo	−76	303	547	29/29	3.8	1.47	<0.001
6	III	IV	SD	Neo	−80	298	744	8/8	6.59	0.73	0.001
7	IV	IV	PR	Neo	−43	358	1,587	7/8	3	0.68	<0.001
8	II	IIIC	PR	Neo	−24	373	889	11/14	3.42	2.24	<0.001
9	VI	IV	SD	Neo	−100	563	1,278	15/16	4.49	0.65	<0.001
10	III	IIIC	PR	Neo	−87	303	1,139	9/11	4.72	0.87	<0.001
11	III	IIIC	PR	Neo	−98	382	1,556	7/17	5.7	0.48	0.28
12	III	IIIC	SD	Neo	−88	534	1,565	1/4	NA	NA	NA
13	III	IIIC	PR	Neo	−28	776	1,166	3/3	NA	0.62	0.48
14	IV	IIIC	PR	Neo	NA	601	1,513	3/4	4.62	0.61	0.74
15	II	IV	SD	Neo	NA	332	706	3/5	NA	0.74	0.74
16	III	IIIC	PR	Neo	NA	1,380	1,405	1/3	NA	NA	NA
17	III	IIIC	SD[f]	PS	NA	293	849	3/4	NA	0.86	0.64

The table shows patients identified for study and the number of samples available before and after quality control. Patients with <3 samples could not be evaluated for TH and CE indices. Patients with CE index but no TH index did not have paired pre-chemotherapy biopsy and interval debulking surgery samples available.

[a] Patient age was segmented into brackets as follows: I, 45–50 y; II, 46–55 y; III, 56–60 y; IV, 61–65 y; V, 66–70 y; VI, 71–75 y.

[b] Response according to RECIST evaluation: PD, progressive disease; PR, partial response; SD, stable disease.

[c] CA 125 tumour marker reduction (percentage),

[d] Number of samples used for analysis (out of all samples taken in study).

[e] Test for star topology (BH corrected).

[f] Not comparable to other SD cases as treatment modalities were different.

NA, not available; Neo, neoadjuvant; PS, primary surgery.

doi:10.1371/journal.pmed.1001789.t001

Table 2. Mutations detected in samples from CTCR-OV03 and CTCR-OV04 patients using TAm-Seq.

Patient Number	Effect	Gene	Protein Change	cDNA Change	RefSeq ID
1	MS	*TP53*	p.Y234C	c.A701G	NM_000546
2	NS	*TP53*	p.Y234X	c.C702A	NM_000546
2	MS	*BRCA1*	p.Y179C	c.A536G	NM_007294
2	MS	*BRCA1*	p.N550H	c.A1648C	NM_007294
2	MS	*BRCA1*	p.F486L	c.T1456C	NM_007294
2	MS	*BRCA2*	p.E1110V	c.A3329T	NM_000059
3	ND	*TP53*			
4	MS	*TP53*	p.H214R	c.A641G	NM_000546
5	MS	*TP53*	p.C141R	c.T421C	NM_000546
6	FS	*TP53*	p.P153fs	c.459_469del11	NM_000546
7	MS	*TP53*	p.R273C	c.C817T	NM_000546
7	MS	*APC*	p.S2596A	c.T7786G	NM_000038
8	MS	*TP53*	p.R175H	c.G524A	NM_000546
8	Silent	*BRCA2*	p.G1552G	c.T4656C	NM_000059
9	FS	*TP53*	p.I195fs	c.583_584dupA	NM_000546
10	MS	*TP53*	p.S215G	c.A643G	NM_000546
10	MS	*APC*	p.D1714N	c.G5140A	NM_000038
11	NS	*TP53*	p.R306X	c.C916T	NM_000546
12	ND	*TP53*			
13	MS	*TP53*	p.Y236S	c.A707C	NM_000546
13	Silent	*BRCA2*	p.V465V	c.A1395C	NM_000059
14	MS	*TP53*	p.V216L	c.G646T	NM_000546
14	NS	*BRCA2*	p.L2732X*	c.T8195A	NM_000059
15	MS	*TP53*	p.C135R	c.T403C	NM_000546
16	MS	*TP53*	p.C275Y	c.G824A	NM_000546
17	MS	*TP53*	p.R273H	c.G818A	NM_000546

Patient 14 had a deleterious somatic nonsense mutation (p.L2732X*) in *BRCA2*. This mutation was not detected in two independent germ line DNA samples from patient 14. All other *BRCA1/2* mutations were not pathogenic or were of no/unknown clinical importance according to the Breast Cancer Information Core Database and the LOVD-IARC database.

FS, frameshift; MS, missense; ND, no mutation detected; NS, nonsense; silent, silent mutation.

doi:10.1371/journal.pmed.1001789.t002

low divergence or limited numbers of samples, and in patient 1 we found very high heterogeneity and strong CE (discussed below).

Metastasis of HGSOC is thought to occur by physical shedding from the invasive lesions in the fallopian tube onto pelvic structures and into the abdomen, where physiological recirculation of ascites fluid facilitates widespread seeding of tumour cells. Evolutionary clades in the patient-specific trees often agreed with the anatomical sites where the sample was taken (see Fig. 2 and individual trees in S3–S16 Figs.). For example, patient 9 showed clear separation of omentum and small bowel mesentery samples (Fig. 2D), with right paracolic gutter and peritoneum as the respective out-groups. Genetic markers of this divergent evolution included Chromosomes 2q and 3p as well as amplifications on Chromosome 10 (Fig. 2C). Higher resolution analyses with paired-end WGS on samples from six patients confirmed additional divergent genetic change at higher resolution. For example, in patient 9, there were three deletion breakpoints and an insertion breakpoint that were present only in the omentum sample, and three deletion breakpoints that were only in the peritoneal sample (S18 Fig).

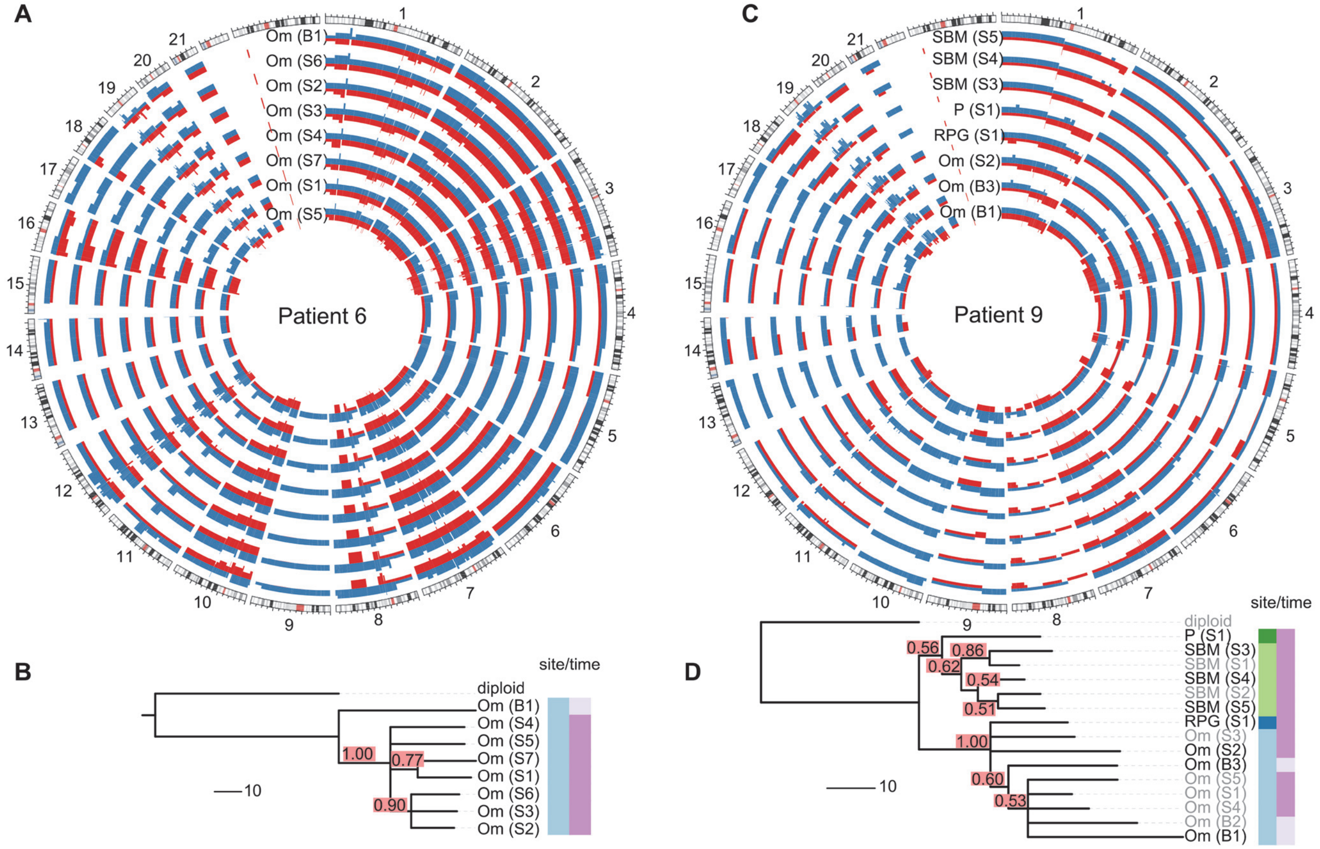


Fig 2. Examples of spatial and temporal heterogeneity in HGSOC. (A and C) Total copy number profiles show strong overall conservation. As examples, a representative subset of the allele-specific genomic copy number profiles of patients 6 and 9 are shown. Separate alleles are indicated in red and blue. (B and D) Genomic changes between biopsy and surgery reveal tumour evolution. The black sample names in the trees indicate the samples shown in the Circos plots. Confidence values for each split are printed in red boxes. The colour-coded bars on the right of the phylogenies indicate different sites (left column) and different sampling times (right column). Branch lengths indicate number of genetic events as determined by MEDICC (scale bar shows ten events). Om, omentum; P, peritoneum; RPG, right paracolic gutter; SBM, small bowel mesentery.

doi:10.1371/journal.pmed.1001789.g002

Analysis of Tree Topologies Suggests Metastasis-to-Metastasis Spread with Changing Evolutionary Rates

We next considered whether the observed spatial heterogeneity arose from metastasis-to-metastasis spread or by successive metastases from the primary cancer site by examining the pattern of the evolutionary relationships between metastatic samples within each patient. Given a fixed number of metastases, two scenarios are possible: if only the primary tumour gives rise to metastatic clones, the resulting evolutionary tree will have a star topology (Fig. 3A). By contrast, if cells retain their metastatic potential after metastasis, ongoing spread and associated genetic change will lead to a fully branched evolutionary tree (Fig. 3B).

We used MEDICC to test the null hypothesis for each patient that the evolutionary distances were derived from a star topology [32]. To verify and visualise the findings we applied the neighbour-net method [42], which captures non-tree-like evolutionary signals in distance data. From nine patients with $\geq$3 samples, eight showed significant branching ($p < 0.05$, chi-squared test for goodness of fit with Benjamini and Hochberg correction for false discovery rate), supporting the model of metastasis-to-metastasis spread. Patient 11 showed only a weak tree structure, and the null hypothesis could not be rejected ($p = 0.22$; Fig. 3A).

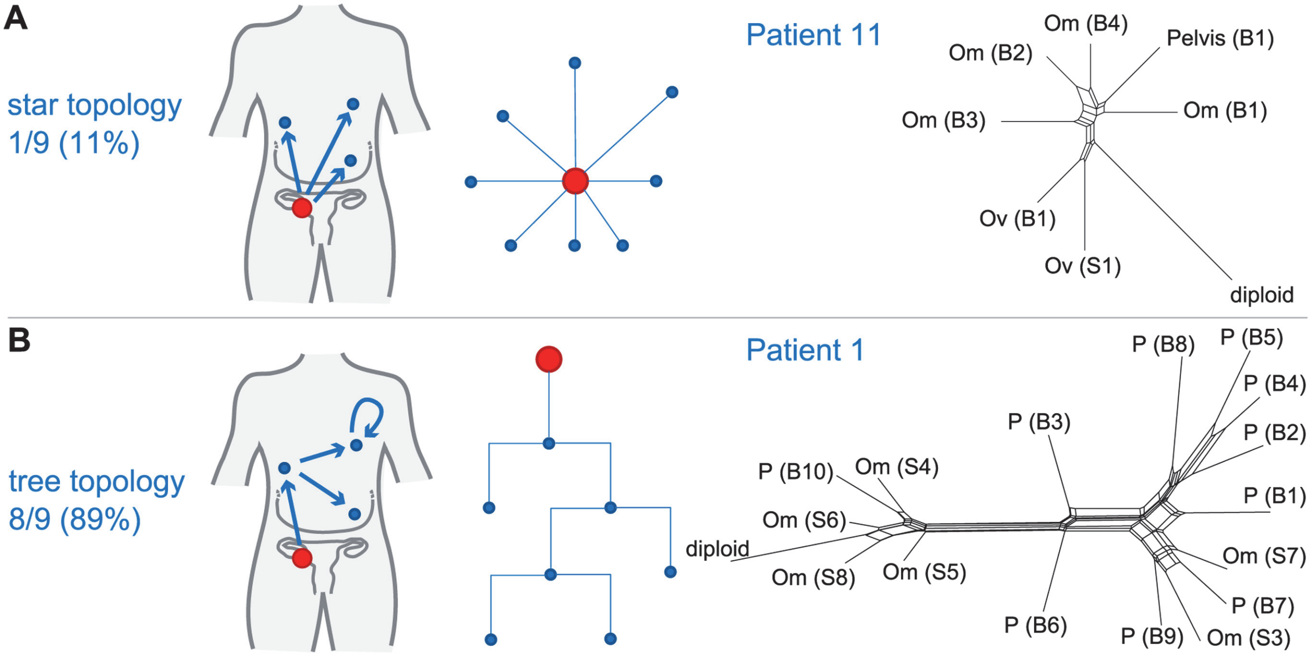


Fig 3. Branching patterns in HGSOC. (A) Radial pattern of metastatic spread leads to a star topology. The schematic shows how the evolutionary relationships are predicted to have a star-like topology if all metastases (blue) are derived from the primary lesion (red). A neighbour-net representation of the evolutionary distances from patient 11 shows deviation from a tree structure (right). (B) Branched metastatic spread leads to a tree topology. The schematic shows that evolutionary history is predicted to be tree-like if metastases create new metastases (including metastasis-to-metastasis spread). A neighbour-net representation of the distance matrix for patient 1 shows a tree-like structure (right). The number and proportion of patients classified to star or tree topology are shown. Labels on trees indicate site of metastasis (Om, omentum; Ov, ovary; P, peritoneum). Sample identifiers indicate whether the sample was collected from pre-chemotherapy biopsy (B) or interval debulking surgery (S).

doi:10.1371/journal.pmed.1001789.g003

Next we compared evolutionary distances within each patient to investigate whether evolutionary change occurs at a constant rate (clock-like evolution). After correction for multiple testing, two out of 14 patients (14.3%) showed significant non-clock-like evolutionary trajectories ($p < 0.05$). We conclude that HGSOC shows metastasis-to-metastasis spread and that heterogeneity is generated through ongoing clonal evolution with potentially unknown mutator phenotypes present.

Small Changes in Heterogeneity Occur during Neoadjuvant Therapy

As most metastases are established before onset of treatment, we next investigated the rate of ongoing clonal evolution by examining samples before and after neoadjuvant chemotherapy. The average genomic change during treatment (TH index) was quantified using MEDICC. To ensure that differences were not due to cellularity, we compared histopathology estimates between the pre-chemotherapy biopsies and surgical specimens and found no significant differences (*t*-test, $p = 0.7$). MEDICC measures TH by mapping genomes into a high-dimensional space, termed the "mutational landscape" [16,17], in which distances correspond to evolutionary distances between genomes. The TH index is then calculated as the distance between the robust centres of mass of the biopsy and surgery samples [32], leading to a robust estimate of change during treatment. Visual analysis of Circos plots showed strong overall conservation, indicating that the main karyotypes for each cancer were established before onset of treatment. Quantitative analysis with MEDICC detected genomic differences between biopsy and surgery samples, showing on average 46 (standard deviation 13) new genomic events (Figs. 2A, 2C, and S3–S16). For example, for patient 6 there was a profound difference between the two sample subgroups (TH index 0.66), with early and long branching of the pre-treatment omental biopsy sample away from the remaining omental samples, indicating divergent evolution (Fig. 2A and 2B). The copy number events responsible for this divergence included deletions on Chromosomes 1p, 1q, 3p, 7q, 9q, and 11p. In patient 9, one of the three omentum samples differed in 18.1% of its genome from the omentum samples at surgery (Fig. 2C and 2D). We concluded that HGSOC shows detectable changes during neoadjuvant chemotherapy (median 75 d), but these are minor compared to the overall changes from the onset of the disease.

HGSOC is Frequently Polygenomic and Shows Variable Clonal Expansion

It has previously been shown in breast cancer that CEs of minor subpopulations of cells lead to polygenomic tumours, while other tumours appear monogenomic [6]. These CEs are potentially modulated by selection pressure from chemotherapy (or other factors) and might have prognostic value. Using MEDICC allowed statistical quantification of the degree of CE on a continuous scale (CE index) by testing for local spatial clustering of genomes in the mutational landscape [32].

We found the CE index to be variable across the cohort (median 0.73, IQR 0.65–1.24). As there was no clinically defined cutoff point for CE, the median value was used to divide patents into two groups (CE-low versus CE-high). Patients with in the CE-low group, for example patient 11, showed linear emission of samples throughout the tree and had homogenous branch lengths (S12 Fig). By contrast, patients 5 and 8, in the CE-high group, showed marked CE (CE index 1.47 and 2.24, respectively), with multiple samples in strongly diverging subclades (Fig. 4).

Patients with Tumours with High Clonal Expansion Show Short Survival and Resistant Relapse

It has been proposed that for a tumour to overcome the selection pressure applied by chemotherapy, it needs to be able to efficiently explore the mutational landscape [17]. Therefore, we hypothesized that polygenomic tumours that have already undergone CEs are likely to be at an advantage for acquiring other mutations for survival during treatment.

We used the log-rank test to test for differences in PFS and OS between the CE-low and CE-high groups (Fig. 5). Survival was shorter in patients in the CE-high group (PFS 12.7 versus 10.1 mo, $p = 0.009$; OS 42.6 versus 23.5 mo, $p = 0.003$; Fig. 5). Being in the CE-high group was an independent predictor of survival in a multivariable Cox hazard regression analysis that included patient age, tumour stage, and residual disease after debulking surgery (PFS, $p = 0.001$; OS, $p = 0.004$). Survival differences were not significant between patients with low or high TH index (S1 Protocol).

We tested CE as a continuous variable in a Cox proportional hazard model, which assumes a linear relationship between CE and survival. In univariable analysis, the quantitative CE index had a borderline significant association with OS with hazard ratio (HR) = 2.7 (95% CI 0.96, 7.8; $p = 0.06$), but no significant association with PFS. In multivariable models that considered CE, patient age, tumour stage, and residual disease, CE as a quantitative variable was not significantly associated with OS or PFS (coefficient $p = 0.64$ and $p = 0.76$, respectively). We examined potential nonlinear effects of CE on survival using cubic spline methods. For PFS and OS, the relationship between CE and relative hazard was nonlinear, showing a step function effect with marked increase in hazard seen at CE values greater than 0.7–0.8 (Fig. 5B; S1 Protocol), similar to the median cut point. Given the small size of the patient cohort, we performed tests of robustness using bootstrap analysis with 10,000-fold resampling to test whether the HR for the CE-high group was likely to be greater than one (deleterious for outcome). The

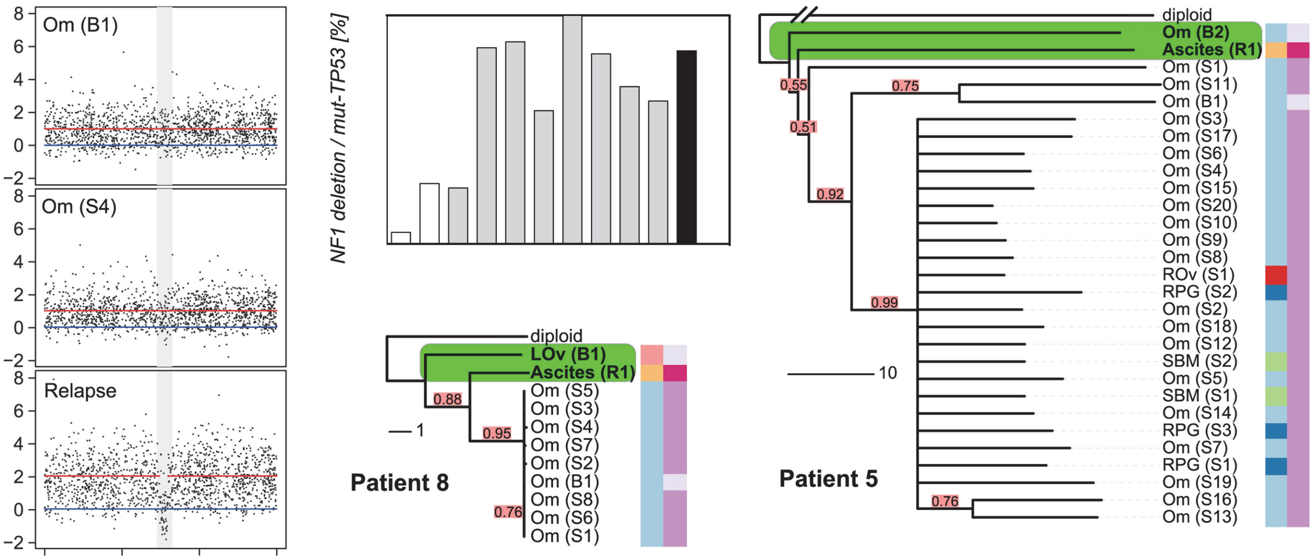


Fig 4. Relapse is an early diverged clonal expansion of a low-prevalence subclone of pre-treatment disease. Array copy number profiles (left) from patient 8 detected a focal *NF1* deletion in the relapsed ascites sample that was not observed in the pre-chemotherapy or interval debulking samples. The bar plot shows the results of digital PCR for the *NF1* breakpoint from pre-chemotherapy (white bars), interval debulking (grey bars), and relapsed ascites (black) samples. Phylogenetic trees for patients 8 and 5 are shown. The relapsed clonal population for each case is placed next to the pre-chemotherapy biopsy sample, indicating early branching events from the diploid. The length of each branch indicates the degree of divergence. Colour coding and sample identifiers are as for Fig. 3. LOv, left ovary; Om, omentum; SBM, small bowel mesentery; RPG, right paracolic gutter.

doi:10.1371/journal.pmed.1001789.g004

differences in PFS and OS remained significant ($p < 0.05$) in 82% and 92%, respectively, of the perturbed datasets. The bootstrapped derived median HR values for PFS and OS were HR = 7.1 (95% CI > 1) and HR = 11.4 (95% CI > 1), respectively.

Resistant Subclonal Populations Are Present in Pre-Treatment Disease

The survival analyses suggested that the degree of CE could have effects on PFS and OS consistent with the hypothesis that cancers with high CE may have increased genetic diversity favouring the emergence of drug-resistant clones. We were able to explore this hypothesis in patients 5 and 8, who had additional samples collected at progression. Patient 8 had symptomatic progressive disease, with the development of ascites at 12 mo after completing chemotherapy. The phylogenetic reconstruction of her cancer showed early divergence of the ascites sample from the root (Fig. 4). Examination of the relapsed copy number profile revealed a new focal deletion at *NF1* that was not present in the pre-chemotherapy and interval debulking surgery samples (Fig. 4). *NF1* is recurrently mutated in HGSOC [22,23], which suggests this was unlikely to be a passenger event. As copy number profiles detect the dominant clone in a sample, we investigated the population structure of earlier samples using WGS to map the new *NF1* deletion, and digital PCR to accurately estimate the number of cells containing the *NF1* deletion in each sample. To prevent confounding effects from differences in tumour cellularity, the counts for the *NF1* deletion were expressed as a proportion of all mutant *TP53* counts. The *NF1* deletion was detected at 5% and 26% in the pre-treatment samples (absolute cellularity 80% and 41%) and in 25%–100% of the interval debulking samples (median cellularity 49%). Histological analysis of the left fallopian tube specimen removed at interval debulking confirmed a tubal primary site (S1 Fig). We therefore extended the digital PCR analysis to DNA from microdissected tissues from formalin-fixed tissue blocks including the left fallopian tube fimbria. The *NF1*

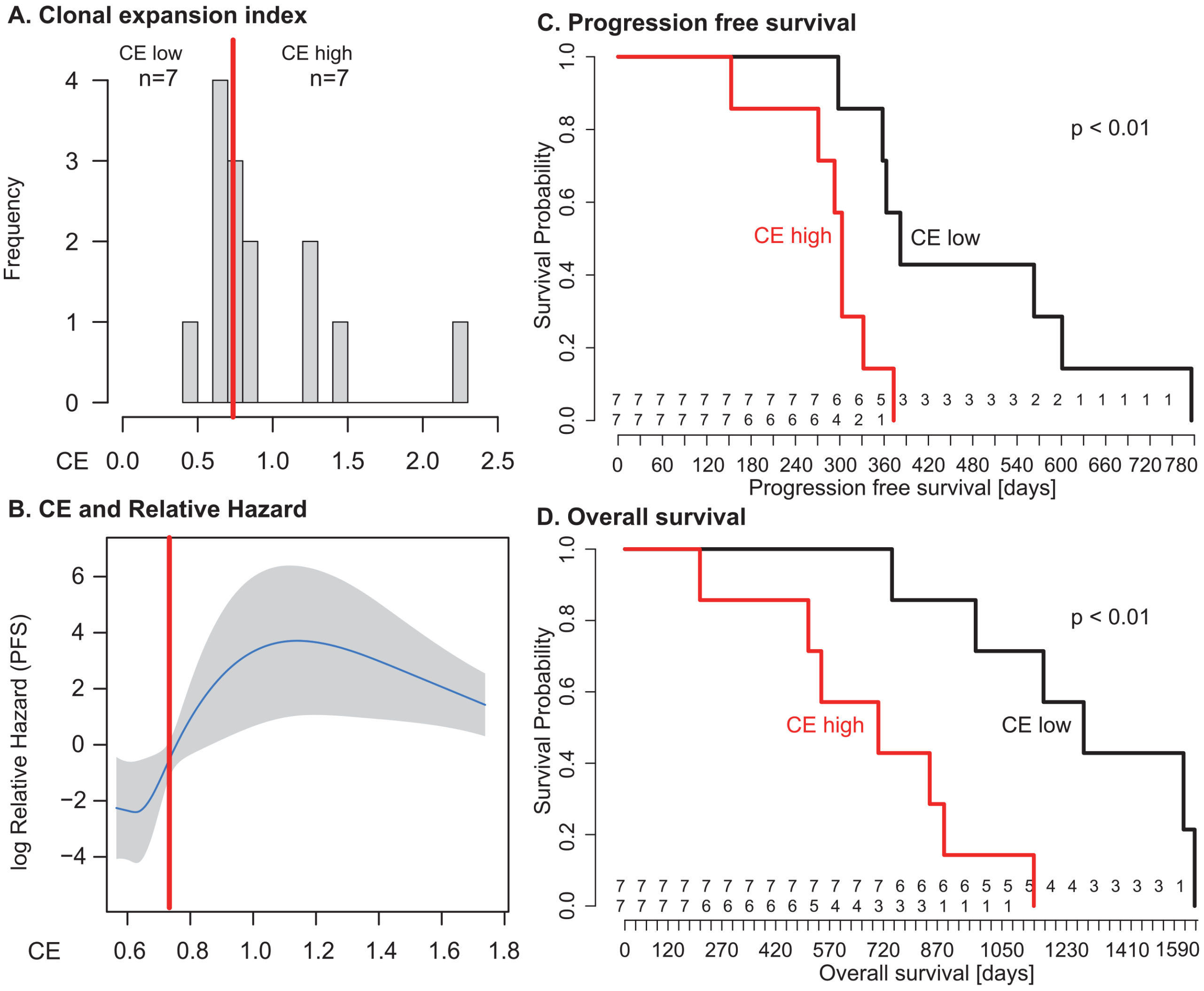


Fig 5. Clonal expansion index stratifies patients into prognostic subgroups. (A) Distribution of CE index over all patients and the respective group sample sizes (*n*). The red line indicates median CE = 0.73, dichotomizing the cases into equal-sized CE-low and CE-high groups. (B) The relationship between CE and relative hazard is nonlinear. The fit line is generated from the multivariable model incorporating penalised spline smoothing. Grey shading indicates the 95% confidence interval for log hazard. Extreme CE values are not shown as the spline smoothing algorithm disregards values outside the 95% range. The median (red line) separates a region of low hazard from a region of high hazard indicated by non-overlapping confidence intervals. (C and D) The CE-low and CE-high groups show a statistically significant difference in PFS (log-rank $p < 0.01$) and OS (log-rank $p < 0.01$). Numbers at risk are given above the *x*-axis for the CE-low (top) and CE-high (bottom) groups.

doi:10.1371/journal.pmed.1001789.g005

deletion was present in 1.2% of the primary invasive carcinoma in the fallopian tube and 7.9% of the biopsy from the adjacent left ovarian metastasis (S2 Table).

In patient 5, inspection of the tree showed that the relapsed ascites sample also diverged early, with a long branch (Fig. 4) indicating marked divergent evolution. This divergence was associated with deletions on Chromosomes 1q, 15q, and 18q (S6A Fig). In summary, the *NF1* deletion, while part of the dominant subpopulation at relapse, was already present pre-treatment. As it is highly unlikely that this specific deletion arose twice independently in the course of tumour evolution, we conclude that the relapse was a CE of a minor subclone of pre-chemotherapy disease.

Discussion

In this work, we tested the hypothesis that intra-tumour heterogeneity in HGSOC is correlated with survival. We also assessed whether certain subclonal populations contribute to treatment failure. Our experimental design combined two approaches: first, we analysed spatially and temporally separate tumour samples from 14 women undergoing chemotherapy. This approach estimates the genetic complexity of the cancer burden in an individual more accurately than sampling from a single location and time point [5,12,13,25]. Second, we applied formal methods (MEDICC) to infer the most parsimonious representation of genetic evolution in each patient's cancer [32]. Importantly, our methods are fully unsupervised and are derived independently of the clinical data.

Our analyses showed marked differences in CE between patients and negative effects of high CE on survival. In two patients with very high CE, we demonstrated that clonal populations detected at relapse arose from early branching events, followed by divergent evolution and CE. Indeed, digital PCR of a *NF1* deletion that marked the predominant clonal population at relapse conclusively showed that this deletion was present at very low fraction in pre-treatment samples including the tubal primary site. We further showed that HGSOC generally evolves and spreads in a branching process with frequently changing rates of evolution. Taken together, these findings are consistent with previous data from cell-based studies and circulating tumour DNA assays that suggested that CE occurs between diagnosis and relapse in HGSOC [28,37].

Although the number of HGSOC samples studied here is relatively large compared to those of other publications, the size of the patient cohort prevents strong conclusions about effect sizes and clinical impact. We used a median value for CE as an unbiased cut point to avoid strong assumptions about the relationship between CE and survival, but given the limited sample size, it is likely that our analyses overestimate the prognostic effect of CE. It is notable that the majority of the samples that failed quality assurance (and were therefore excluded from estimations of heterogeneity) were taken after chemotherapy treatment, suggesting that these samples may have had better response to treatment [28]. This implies that the samples from which our heterogeneity measures were calculated may be enriched for more chemoresistant disease. We have not defined the minimum number of samples per case that are required for accurate estimates of CE, and this will require larger patient studies. Collecting these samples remains a major challenge for heterogeneity research, owing to the difficulties of collecting multiple fresh tissue samples at different treatment time points and the costs of detailed genomic profiling. Further technological development to use shallow WGS data from formalin-fixed, paraffin-embedded samples may be a useful approach to increasing statistical power in future studies.

Comparison of the effects of CE and TH on survival showed that TH was not predictive of PFS or OS. This was surprising, as we expected strong TH effects to be correlated with response, and therefore survival. There are several factors that may explain our finding. First, we were unable to take samples from the same tumour deposit before and after chemotherapy treatment. Therefore, apparent differences in TH could be confounded by spatial differences in tumour heterogeneity, rather than representing intrinsic changes in subclonal populations caused by chemotherapy treatment. Second, the time window for evolutionary changes to occur during chemotherapy was short compared to the genetic lifespan of each cancer. Third, both CE and TH showed moderate correlation with sample size (S1 Protocol). Sample size was not significant in a multivariable Cox model, but could potentially contribute to the predictive power of CE (S1 Protocol).

Our results are in disagreement with recent findings where the analysis of nonsynonymous mutations did not show effects of ongoing evolutionary change in HGSOC [43]. These findings were based upon exome sequencing of three patients, and the power of this assay for evolutionary inference is dependent upon the depth of sequencing achieved. It is also likely that the majority of nonsynonymous changes detected by exome sequencing are passenger or private mutations, which may explain why other studies have not found evidence of the strong evolutionary patterns that we see using CNAs.

Our phylogenetic reconstructions further allowed us to assess the robustness of the evolutionary trees, and thereby the certainty of placement of a sample in the tree. With this we addressed the question of when in the course of disease the relapse clone evolved. In both patients 5 and 8, we were able to determine an early branching point as the origin of relapse that shared an immediate ancestor with a pre-treatment sample. In larger datasets these methods could be applied to the identification of early driver events and may mitigate the difficulties of identifying therapeutically relevant CNAs in heterogeneous patients.

In summary, our approach has been to define the evolutionary trajectories of HGSOC using robust and accurate methods to reconstruct the phylogenetic trees for individual patients. This approach has the potential to act as a patient-specific prognostic indicator and may be a powerful tool to identify and calibrate surrogate genomic markers of CE.

Supporting Information

S1 Data. Data file for survival analysis.
(ZIP)

S1 Fig. Histopathology for case 8.
(PDF)

S2 Fig. Cross-patient loss of heterozygosity frequencies.
(PDF)

S3 Fig. Copy number profile and evolutionary tree for patient 1. Only selected copy number profiles are shown (A); these are marked in bold in the evolutionary tree (B). Individual alleles are coloured in red and blue. Confidence values for each split in the tree are given in red boxes. The colour bars to the right of the tree indicate different sampling sites (left) and sampling times (right). Branch lengths are given in number of rearrangement events. Particularly long branches are marked in green.
(PDF)

S4 Fig. Copy number profile and evolutionary tree for patient 2. Caption as for S3 Fig.
(PDF)

S5 Fig. Copy number profile and evolutionary tree for patient 3. Caption as for S3 Fig.
(PDF)

S6 Fig. Copy number profile and evolutionary tree for patient 5. Caption as for S3 Fig.
(PDF)

S7 Fig. Copy number profile and evolutionary tree for patient 6. Caption as for S3 Fig.
(PDF)

S8 Fig. Copy number profile and evolutionary tree for patient 7. Caption as for S3 Fig.
(PDF)

S9 Fig. Copy number profile and evolutionary tree for patient 8. Caption as for S3 Fig.
(PDF)

S10 Fig. Copy number profile and evolutionary tree for patient 9. Caption as for S3 Fig.
(PDF)

S11 Fig. Copy number profile and evolutionary tree for patient 10. Caption as for S3 Fig.
(PDF)

S12 Fig. Copy number profile and evolutionary tree for patient 11. Caption as for S3 Fig.
(PDF)

S13 Fig. Copy number profile and evolutionary tree for patient 13. Caption as for S3 Fig.
(PDF)

S14 Fig. Copy number profile and evolutionary tree for patient 14. Caption as for S3 Fig.
(PDF)

S15 Fig. Copy number profile and evolutionary tree for patient 15. Caption as for S3 Fig.
(PDF)

S16 Fig. Copy number profile and evolutionary tree for patient 17. Caption as for S3 Fig.
(PDF)

S17 Fig. Tree shapes and evolutionary patterns in resistant and sensitive cases.
(PDF)

S18 Fig. Paired-end sequencing of selected samples.
(PDF)

S1 Protocol. Sweave file for survival analysis.
(RNW)

S1 Table. Distribution of samples.
(PDF)

S2 Table. Digital PCR results.
(PDF)

S3 Table. Primers for digital PCR.
(PDF)

S4 Table. Copy number alteration feature selection results.
(PDF)

S1 Text. Case-by-case study of the CTCR-OV03/04 cohort.
(PDF)

Acknowledgments

We thank H. Biggs, C. Hodgkin, L. Jones, the Addenbrooke's Hospital Human Research Tissue Bank, and the Addenbrooke's Hospital gynaecology-oncology multidisciplinary team for expert assistance in sample collection and study recruitment. We thank the Core Facilities at the Cancer Research UK Cambridge Institute for expert support.

Author Contributions

Conceived and designed the experiments: JDB ES FM RFS CKYN SLC AMP DG. Performed the experiments: SLC SN JT AMP DG. Analyzed the data: RFS CKYN SLC SN AMP DG KS MM NR MJL CAP FM JDB. Contributed reagents/materials/analysis tools: RFS CKYN SN PJB MM NR ES CAP FM JDB. Enrolled patients: HME CAP JDB. Wrote the first draft of the manuscript: RFS FM JDB. Wrote the paper: RFS CKYN AMP DG FM JDB. Agree with manuscript results and conclusions: RFS CKYN SLC SN JT AMP DG KS MM PJB NR HME ES MJL CAP FM JDB. All authors have read, and confirm that they meet, ICMJE criteria for authorship.

References

1. Nowell PC (1976) The clonal evolution of tumor cell populations. Science 194: 23–28. PMID: 959840
2. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, et al. (1978) Heterogeneity of tumor cells from a single mouse mammary tumor. Cancer Res 38: 3174–3181. PMID: 210930
3. Khalique L, Ayhan A, Weale ME, Jacobs IJ, Ramus SJ, et al. (2007) Genetic intra-tumour heterogeneity in epithelial ovarian cancer and its implications for molecular diagnosis of tumours. J Pathol 211: 286–295. PMID: 17154249
4. Khalique L, Ayhan A, Whittaker JC, Singh N, Jacobs IJ, et al. (2009) The clonal evolution of metastases from primary serous epithelial ovarian cancers. Int J Cancer 124: 1579–1586. doi: 10.1002/ijc.24148 PMID: 19123469
5. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, et al. (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 461: 809–813. doi: 10.1038/nature08489 PMID: 19812674
6. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, et al. (2010) Inferring tumor progression from genomic heterogeneity. Genome Res 20: 68–80. doi: 10.1101/gr.099622.109 PMID: 19903760
7. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, et al. (2010) The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 467: 1109–1113. doi: 10.1038/nature09460 PMID: 20981101
8. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. (2011) Tumour evolution inferred by single-cell sequencing. Nature 472: 90–94. doi: 10.1038/nature09807 PMID: 21399628
9. Marusyk A, Almendro V, Polyak K (2012) Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer 12: 323–334. doi: 10.1038/nrc3261 PMID: 22513401
10. Vermaat JS, Nijman IJ, Koudijs MJ, Gerritse FL, Scherer SJ, et al. (2012) Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clin Cancer Res 18: 688–699. doi: 10.1158/1078-0432.CCR-11-1965 PMID: 22173549
11. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJH, et al. (2012) Clonal selection drives genetic divergence of metastatic medulloblastoma. Nature 482: 529–533. doi: 10.1038/nature10825 PMID: 22343890
12. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366: 883–892. doi: 10.1056/NEJMoa1113205 PMID: 22397650
13. Shah SP, Roth A, Goya R, Oloumi A, Ha G, et al. (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 486: 395–399. doi: 10.1038/nature10933 PMID: 22495314
14. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, et al. (2014) Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 346: 251–256. doi: 10.1126/science.1253462 PMID: 25301630
15. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, et al. (2012) The life history of 21 breast cancers. Cell 149: 994–1007. doi: 10.1016/j.cell.2012.04.023 PMID: 22608083
16. Merlo LMF, Pepper JW, Reid BJ, Maley CC (2006) Cancer as an evolutionary and ecological process. Nat Rev Cancer 6: 924–935. PMID: 17109012
17. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481: 306–313. doi: 10.1038/nature10762 PMID: 22258609
18. Anderson K, Lutz C, van Delft FW, Bateman CM, Guo Y, et al. (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. Nature 469: 356–361. doi: 10.1038/nature09650 PMID: 21160474
19. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, et al. (2013) Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell 152: 714–726. doi: 10.1016/j.cell.2013.01.019 PMID: 23415222
20. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, et al. (2010) Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. J Pathol 221: 49–56. doi: 10.1002/path.2696 PMID: 20229506
21. Gorringe KL, Jacobs S, Thompson ER, Sridhar A, Qiu W, et al. (2007) High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications. Clin Cancer Res 13: 4731–4739. PMID: 17699850

22. Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474: 609–615. doi: 10.1038/nature10166 PMID: 21720365
23. Sangha N, Wu R, Kuick R, Powers S, Mu D, et al. (2008) Neurofibromin 1 (NF1) defects are common in human ovarian serous carcinomas and co-occur with TP53 mutations. Neoplasia 10: 1362–1372. PMID: 19048115
24. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, et al. (2012) Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 30: 413–421. doi: 10.1038/nbt.2203 PMID: 22544022
25. Bashashati A, Ha G, Tone A, Ding J, Prentice LM, et al. (2013) Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. J Pathol 231: 21–34. doi: 10.1002/path.4230 PMID: 23780408
26. Zhang J, Shi Y, Lalonde E, Li L, Cavallone L, et al. (2013) Exome profiling of primary, metastatic and recurrent ovarian carcinomas in a BRCA1-positive patient. BMC Cancer 13: 146. doi: 10.1186/1471-2407-13-146 PMID: 23522120
27. Hoogstraat M, de Pagter MS, Cirkel GA, van Roosmalen MJ, Harkins TT, et al. (2014) Genomic and transcriptomic plasticity in treatment-naive ovarian cancer. Genome Res 24: 200–211. doi: 10.1101/gr.161026.113 PMID: 24221193
28. Cooke SL, Ng CKY, Melnyk N, Garcia MJ, Hardcastle T, et al. (2010) Genomic analysis of genetic heterogeneity and evolution in high-grade serous ovarian carcinoma. Oncogene 29: 4905–4913. doi: 10.1038/onc.2010.245 PMID: 20581869
29. Cooke SL, Brenton JD (2011) Evolution of platinum resistance in high-grade serous ovarian cancer. Lancet Oncol 12: 1169–1174. doi: 10.1016/S1470-2045(11)70123-1 PMID: 21742554
30. Wang ZC, Birkbak NJ, Culhane AC, Drapkin R, Fatima A, et al. (2012) Profiles of genomic in-stability in high-grade serous ovarian cancer predict treatment outcome. Clin Cancer Res 18: 5806–5815. doi: 10.1158/1078-0432.CCR-12-0857 PMID: 22912389
31. Cowin PA, George J, Fereday S, Loehrer E, Van Loo P, et al. (2012) LRP1B deletion in high-grade serous ovarian cancers is associated with acquired chemotherapy resistance to liposomal doxorubicin. Cancer Res 72: 4060–4073. PMID: 22896685
32. Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, et al. (2014) Phylogenetic quantification of intra-tumour heterogeneity. PLoS Comput Biol 10: e1003535. doi: 10.1371/journal.pcbi.1003535 PMID: 24743184
33. Greenman CD, Bignell G, Butler A, Edkins S, Hinton J, et al. (2010) PICNIC: an algorithm to predict absolute allelic copy number variation with microarray cancer data. Biostatistics 11: 164–175. doi: 10.1093/biostatistics/kxp045 PMID: 19837654
34. Ng CKY, Cooke SL, Howe K, Newman S, Xian J, et al. (2012) The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer. J Pathol 226: 703–712. doi: 10.1002/path.3980 PMID: 22183581
35. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760. doi: 10.1093/bioinformatics/btp324 PMID: 19451168
36. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, et al. (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40: e115. PMID: 22730293
37. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, et al. (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 4: 136ra68. doi: 10.1126/scitranslmed.3003726 PMID: 22649089
38. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. (2011) Integrative genomics viewer. Nat Biotechnol 29: 24–26. doi: 10.1038/nbt.1754 PMID: 21221095
39. Archibald KM, Kulbe H, Kwong J, Chakravarty P, Temple J, et al. (2012) Sequential genetic change at the TP53 and chemokine receptor CXCR4 locus during transformation of human ovarian surface epithelium. Oncogene 31: 4987–4995. doi: 10.1038/onc.2011.653 PMID: 22266861
40. Sala E, Kataoka MY, Priest AN, Gill AB, McLean MA, et al. (2012) Advanced ovarian cancer: multiparametric MR imaging demonstrates response-and metastasis-specific effects. Radiology 263: 149–159. doi: 10.1148/radiol.11110175 PMID: 22332064
41. McBride DJ, Etemadmoghadam D, Cooke SL, Alsop K, George J, et al. (2012) Tandem duplication of chromosomal segments is common in ovarian and breast cancer genomes. J Pathol 227: 446–455. doi: 10.1002/path.4042 PMID: 22514011
42. Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21: 255–265. PMID: 14660700
43. Castellarin M, Milne K, Zeng T, Tse K, Mayo M, et al. (2012) Clonal evolution of high-grade serous ovarian carcinoma from primary to recurrent disease. J Pathol 229: 515–524. doi: 10.1002/path.4105 PMID: 22996961

Editors' Summary

Background

Every year, nearly 250,000 women develop ovarian cancer, and about 150,000 die from the disease. Ovarian cancer occurs when a cell on the surface of the ovaries (two small organs in the pelvis that produce eggs) or in the Fallopian tubes (which connect the ovaries to the womb) acquires genetic changes (mutations) that allow it to grow uncontrollably and to spread around the body (metastasize). For women whose ovarian cancer is diagnosed when it is confined to its site of origin, the outlook is good. About 90% of these women survive for at least five years. However, ovarian cancer is rarely diagnosed this early. Usually, by the time the cancer causes symptoms (often only vague abdominal pains and mild digestive disturbances), it has spread into the peritoneal cavity (the space around the gut, stomach, and liver) or has metastasized to distant organs. Patients with advanced ovarian cancer are treated with a combination of surgery and platinum-based chemotherapy, but only a quarter of such women are still alive five years after diagnosis, and the overall five-year survival rate for ovarian cancer is less than 50%.

Why Was This Study Done?

The major clinical challenge in the treatment of high-grade serous ovarian cancer (HGSOC; the most common type of ovarian cancer) is the development of resistance to platinum-based chemotherapy. If we knew how this resistance develops, it might be possible to improve the treatment of HGSOC. Tumors are thought to arise from a single mutated cell that accumulates additional mutations as it grows and divides. This process results in the formation of subpopulations of tumor cells, each with a different set of mutations. Experts think that this "intra-tumor heterogeneity" gives rise to tumor subclones that possess an evolutionary advantage over other subclones (they might, for example, grow faster or be resistant to chemotherapy) and that eventually dominate the tumor ("clonal expansion"). Here, the researchers investigate whether clonal evolution and the emergence of subclonal tumor populations explains the development of chemotherapy-resistant HGSOC by undertaking evolutionary inference and phylogenetic quantification of the heterogeneity of samples taken from women with HGSOC at different times and from different places in their body. Evolutionary inference and phylogenetic quantification are analytical approaches that can be used to reconstruct the evolutionary history ("family tree") of a tumor.

What Did the Researchers Do and Find?

The researchers used an algorithm (a step-by-step procedure for data processing) called MEDICC to analyze detailed genetic data obtained from 135 spatially and temporally separated samples taken from 14 patients with HGSOC who had received platinum-based chemotherapy. The researchers report that there were marked differences in the degree of clonal expansion among the patients. When they split the patients into two groups based on the degree of clonal expansion in their tumors (CE-high and CE-low), patients with tumors classified as CE-high had a shorter progression-free survival time than patients with tumors classified as CE-high (10.1 months compared to 12.7 months) and a shorter overall survival time (23.5 months compared to 42.6 months). Moreover, a type of statistical analysis called bootstrap analysis, which tests for the robustness of the result, indicated that having CE-high tumors was likely to increase a patient's risk of a poor outcome. Finally, phylogenetic analysis of samples taken from two patients before and after relapse and analysis of a *NF1* deletion (*NF1* encodes neurofibromin 1, a tumor suppressor protein that prevents uncontrolled cell growth; *NF1* is frequently mutated in HGSOC) indicated that a resistant subclonal population was already present in the patients' tumors before treatment began.

What Do These Findings Mean?

These findings show that clonal expansion occurs between diagnosis and relapse in HGSOC, that there are marked differences in the degree of clonal expansion among patients, and that a high degree of clonal expansion may have a negative effect on survival. The accuracy of these findings is limited by the small number of patients included in the study, and it is likely that the analyses reported here overestimate the effect of clonal expansion on patient outcomes. Nevertheless, the researchers suggest that, provided larger patient studies yield similar results, quantitative measures of intra-tumor heterogeneity might be useful as patient-specific prognostic markers in HGSOC. That is, measures of intra-tumor heterogeneity might eventually help clinicians to predict which of their patients with ovarian cancer are likely to have the best outcomes after platinum-based chemotherapy.

Additional Information

Please access these websites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.1001789.

- The US National Cancer Institute provides information about cancer and how it develops (in English and Spanish), including detailed information about ovarian cancer
- Cancer Research UK, a not-for-profit organization, provides general information about cancer and how it develops, and detailed information about ovarian cancer
- The UK National Health Service Choices website has information and personal stories about ovarian cancer
- The not-for-profit organization Healthtalk.org provides personal stories about dealing with ovarian cancer; Eyes on the Prize, an online support group for women who have had cancers of the female reproductive system, also includes personal stories; the not-for-profit organization Ovarian Cancer Action also provides information, support, and personal stories about ovarian cancer
- Wikipedia provides information about clonal evolution in cancer, tumor heterogeneity, and phylogenetics (note that Wikipedia is a free online encyclopedia that anyone can edit; available in several languages)
- More information about the MEDICC algorithm is available

CANCER GENOMICS

Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA

Tim Forshew,[1]* Muhammed Murtaza,[1,2]* Christine Parkinson,[1,2,3]* Davina Gale,[1]* Dana W. Y. Tsui,[1]* Fiona Kaper,[4†] Sarah-Jane Dawson,[1,2,3] Anna M. Piskorz,[1,2] Mercedes Jimenez-Linan,[3,5] David Bentley,[6] James Hadfield,[1] Andrew P. May,[4] Carlos Caldas,[1,2,3,7] James D. Brenton,[1,2,3,7‡] Nitzan Rosenfeld[1,2‡]

Plasma of cancer patients contains cell-free tumor DNA that carries information on tumor mutations and tumor burden. Individual mutations have been probed using allele-specific assays, but sequencing of entire genes to detect cancer mutations in circulating DNA has not been demonstrated. We developed a method for tagged-amplicon deep sequencing (TAm-Seq) and screened 5995 genomic bases for low-frequency mutations. Using this method, we identified cancer mutations present in circulating DNA at allele frequencies as low as 2%, with sensitivity and specificity of >97%. We identified mutations throughout the tumor suppressor gene *TP53* in circulating DNA from 46 plasma samples of advanced ovarian cancer patients. We demonstrated use of TAm-Seq to noninvasively identify the origin of metastatic relapse in a patient with multiple primary tumors. In another case, we identified in plasma an *EGFR* mutation not found in an initial ovarian biopsy. We further used TAm-Seq to monitor tumor dynamics, and tracked 10 concomitant mutations in plasma of a metastatic breast cancer patient over 16 months. This low-cost, high-throughput method could facilitate analysis of circulating DNA as a noninvasive "liquid biopsy" for personalized cancer genomics.

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

INTRODUCTION

Circulating cell-free DNA extracted from plasma or other body fluids has potentially transformative applications in cancer management (*1*–*7*). Characterization of tumor mutation profiles is required for informed choice of therapy, given that biological agents target specific pathways and effectiveness may be modulated by specific mutations (*8*–*11*). Yet, mutation profiles in different metastatic clones can differ significantly from each other or from the parent primary tumor (*12*, *13*). Evolutionary changes within the cancer can alter the mutational spectrum of the disease and its responsiveness to therapies, which may necessitate repeat biopsies (*14*–*17*). Biopsies are invasive and costly and only provide a snapshot of mutations present at a given time and location. For some applications, mutation detection in plasma DNA as a "liquid biopsy" could potentially replace invasive biopsies as a means to assess tumor genetic characteristics (*2*–*7*). Sensitive methods for detecting cancer mutations in plasma may find use in early detection screening (*1*), prognosis, monitoring tumor dynamics over time, or detection of minimal residual disease (*3*, *18*, *19*). In high-grade serous ovarian carcinomas (HGSOC), mutations in the tumor suppressor gene *TP53* have been observed in 97% of cases (*20*, *21*), but these are located throughout the gene and are difficult to assay. A cost-effective method that could detect and measure allele frequency (AF) of *TP53* mutations in plasma may be highly applicable as a biomarker for HGSOC (*22*).

Circulating DNA is fragmented to an average length of 140 to 170 base pairs (bp) and is present in only a few thousand amplifiable copies per milliliter of blood, of which only a fraction may be diagnostically relevant (*2*, *3*, *23*–*25*). Recent advances in noninvasive prenatal diagnostics highlight the clinical potential of circulating DNA (*25*–*28*), but also the challenges involved in analysis of circulating tumor DNA (ctDNA), where mutated loci and AFs may be more variable. Various methods have been optimized to detect extremely rare alleles (*1*, *2*, *6*, *7*, *29*–*31*), and can assay for predefined or hotspot mutations. These methods, however, interrogate individual or few loci and have limited ability to identify mutations in genes that lack mutation hotspots, such as the *TP53* and *PTEN* tumor suppressor genes (*32*). In patients with more advanced cancers, ctDNA can comprise as much as 1% to 10% or more of circulating DNA (*2*), presenting an opportunity for more extensive genomic analysis. Targeted resequencing has been recently used to identify mutations in selected genes at AFs as low as 5% (*33*–*35*). However, identifying mutations across sizeable genomic regions spanning entire genes at an AF as low as 2%, or in few nanograms of fragmented template from circulating DNA, has been more challenging.

In response, we describe a tool for noninvasive mutation analysis on the basis of tagged-amplicon deep sequencing (TAm-Seq), which allows amplification and deep sequencing of genomic regions spanning thousands of bases from as little as individual copies of fragmented DNA. We applied this technique for detection of both abundant and

---

[1] Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [2] Department of Oncology, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK. [3] Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [4] Fluidigm Corporation, 7000 Shoreline Court, Suite 100, South San Francisco, CA 94080, USA. [5] Department of Histopathology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. [6] Illumina Cambridge, Chesterford Research Park, Little Chesterford, Cambridge CB10 1XL, UK. [7] Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.

*These authors contributed equally to this work.

†Present address: Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA.

‡To whom correspondence should be addressed. E-mail: nitzan.rosenfeld@cancer.org.uk (N.R.); james.brenton@cancer.org.uk (J.D.B.)

rare mutations in circulating DNA from blood plasma of ovarian and breast cancer patients. This sequencing approach allowed us to monitor changes in tumor burden by sampling only patient plasma over time. Combined with faster, more accurate sequencing technologies or rare allele amplification strategies, this approach could potentially be used for personalized medicine at point of care.

RESULTS

Targeted deep sequencing of fragmented DNA by TAm-Seq

To amplify and sequence fragmented DNA, we designed primers to generate amplicons that tile regions of interest in short segments of about 150 to 200 bases (Fig. 1A and table S1), incorporating universal

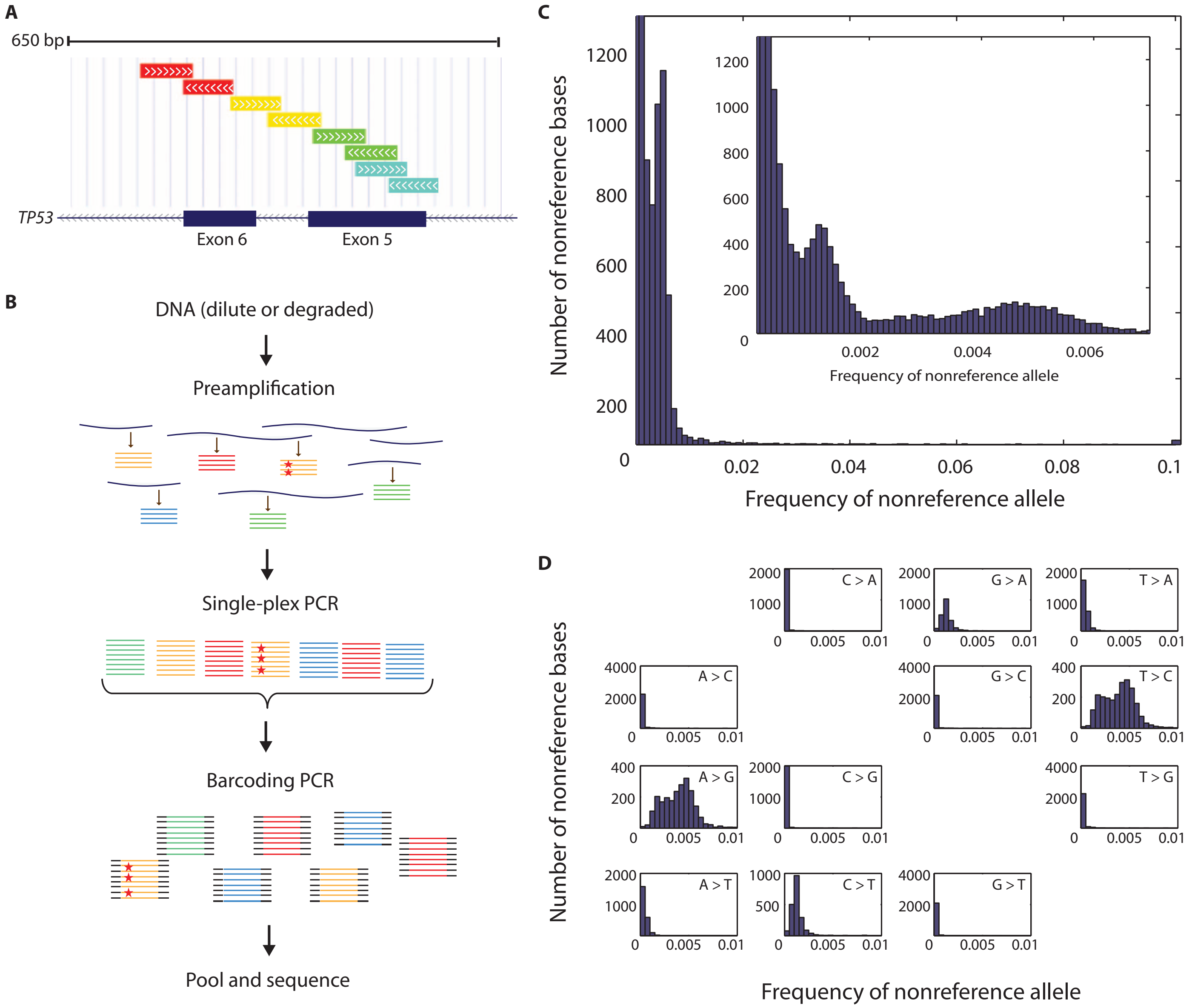


Fig. 1. Overview of tagged amplicon sequencing (TAm-Seq). (**A**) Illustration of amplicon design. Primers were designed to amplify regions of interest in overlapping short amplicons (table S1). Amplicon design is illustrated for a region covering exons 5 to 6 of *TP53*. Colored bars, segmented into forward and reverse reads, show regions covered by different amplicons (excluding primer regions). Sequencing adaptors are attached at either end, such that a single-end read generates separate sets of forward and reverse reads (fig. S1). Because amplicons are mostly shorter than 200 bp, the forward and reverse reads also partially overlap. Figure adapted from University of California, Santa Cruz, Genome Browser (http://genome.ucsc.edu/). (**B**) Workflow overview. Multiple regions were amplified in parallel. An initial preamplification step was performed for 15 cycles using a pool of the target-specific primer pairs to preserve representation of all alleles in the template material. The schematic diagram shows DNA molecules that carry mutations (red stars) being amplified alongside wild-type molecules. Regions of interest in the preamplified material were then selectively amplified in individual (single-plex) PCR, thus excluding nonspecific products. Finally, sequencing adaptors and sample-specific barcodes were attached to the harvested amplicons in a further PCR. (**C**) Distribution of observed nonreference read frequencies, averaged over 47 FFPE samples, across all loci and all nonreference bases. Inset expands the low-frequency range. (**D**) Distribution of the observed background nonreference read frequencies averaged over 47 FFPE samples for the 12 different A/C/G/T base substitutions.

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

adaptors at 5′ ends (fig. S1). Performing single-plex amplification with each of these primer pairs would require dispersing the initial sample into many separate reactions, considerably increasing the probability of sampling errors and allelic loss. Multiplex amplification using a large set of primers could result in nonspecific amplification products and biased coverage. We therefore applied a two-step amplification process: a limited-cycle preamplification step where all primer sets were used together to capture the starting molecules present in the template, followed by individual amplification to purify and select for intended targets (Fig. 1B) (Supplementary Methods). The final concentration of each primer in the preamplification reaction was 50 nM, reducing the potential for interprimer interactions, and 15 cycles of long-extension (4 min) polymerase chain reaction (PCR) were used to remain in the exponential phase of amplification. We used a microfluidic system (Access Array, Fluidigm) to perform parallel single-plex amplification from multiple preamplified samples using multiple primer sets. An additional PCR step attached sequencing adaptors (fig. S1) and tagged each sample by a unique molecular identifier or “barcode” (table S2). Sequencing adaptors were separately attached at either end and the products mixed together, such that single-end sequencing generated separate sets of forward and reverse reads. We performed 100-base single-end sequencing (GAIIx sequencer, Illumina), with an additional 10 cycles using the barcode sequencing primer, generating ~30 million reads per lane. This produced an average read depth of 3250 for each of 96 barcoded samples for 48 amplicons read in two possible orientations.

Validation and sensitivity for mutation identification in ovarian tumor samples

We designed a set of 48 primer pairs to amplify 5995 bases of genomic sequence covering coding regions (exons and exon junctions) of *TP53* and *PTEN*, and selected regions in *EGFR*, *BRAF*, *KRAS*, and *PIK3CA* (table S1) by overlapping short amplicons (Fig. 1A). The sequenced regions cover mutations that account for 38% of all point mutations in the COSMIC database (v55) (*32*). We used TAm-Seq to sequence DNA extracted from 47 formalin-fixed, paraffin-embedded (FFPE) tumor specimens of ovarian cancers (table S3), which were also sequenced for *TP53* by Sanger sequencing (*36*) (Supplementary Methods). DNA extracted from FFPE samples is generally degraded and fragmented as a result of fixation and long-term ambient storage. We amplified DNA from each sample in duplicate, tagging each replicate with a different barcode. Using a single lane of sequencing, we generated 3.5 gigabases of data passing signal purity filters, producing mean read depth of 3200 above Q30 for each of the 9024 expected read groups (48 amplicons × 2 directions × 94 barcoded samples). Background frequencies of nonreference reads were ~0.1% (median, 0.03%; mean, 0.2%; in keeping with Q30 quality threshold applied), yet varied substantially between loci and base substitutions (Fig. 1C) and showed a clear bias toward purine/pyrimidine conservation (Fig. 1D). Sixty-six percent of loci had mean background rate of <0.1%, and 96% of loci had background rate of <0.6%.

The data set interrogated nearly 18,000 possible single-base substitutions for each sample, which introduces a risk of false detection. To control for sporadic PCR errors and reduce false positives, we called point mutations in a sample only if nonreference AFs were above the respective substitution-specific background distribution at a high confidence margin (0.9995 or greater), and ranked high in the list of nonreference AFs, in both replicates (Supplementary Methods). Duplicate sequencing data were obtained for 44 samples, and 43 single-base substitutions were called (table S3). These matched 100% of mutations identified by Sanger sequencing and included three additional mutations at low AFs that were below detection thresholds of Sanger sequencing (fig. S2). The upper bound of AFs that may have been missed was estimated (Supplementary Methods) at <5% for 36 of 44 FFPE samples (82%) and <10% for 42 of 44 samples (95%), with median value of 1.3% and mean value of 2.7%. Mutant AFs were highly reproducible in duplicate samples. For 42 of 43 mutations called, the difference in measured frequency between duplicates was less than 0.08, and the relative difference was 25% or less (Fig. 2A). Mutant AFs correlated significantly with tumor cellularity in the FFPE block (correlation coefficient = 0.422; $P = 0.0049$, t test) (Fig. 2B).

In a separate run, we sequenced libraries prepared from six different diluted mixtures of six FFPE samples, with a different known point mutation in *TP53* in each, to mean read depth of 5600. Of more than 100,000 possible non-SNP (single-nucleotide polymorphism) substitutions, we identified all 33 expected point mutations present at AF >1%, including 6 mutations present at AF <2%, with one false-positive called with AF = 1.9%. Using less stringent parameters (Supplementary Methods), we identified three additional mutations present at AF = 0.6% (Fig. 2C), with no additional false positives. Thus, we obtained 100% sensitivity, identifying mutations at AFs as low as 0.6%. A positive predictive value (PPV) of 100% was calculated for mutations at AF >2%, and a PPV of 90% for mutations identified at AF <2% (Fig. 2D).

Quantitative limitations of mutation detection

When applying TAm-Seq to measure a predefined mutation (as opposed to screening thousands of possible substitutions), the frequency of the mutant allele can be read out directly from the data at the desired locus. False detection is less likely, and criteria for confident mutation detection for a predefined substitution can be less stringent than those described above for de novo mutation identification (Supplementary Methods). The minimal nonreference AFs that could be detected depend on the read depth and background rates of nonreference reads, which vary per locus and substitution type. Minimal detectable frequencies increase when higher confidence margins are used (Supplementary Methods) and had a median value of 0.14% at confidence margin of 0.95 and 0.18% at confidence margin of 0.99 (fig. S3). The minimal detectable frequency would also be limited if a minimal number of reads is applied for confident mutation detection; for example, a minimum of 10 reads implies that sequencing depth of 5000 would be required to detect mutations at AF as low as 0.2%. For alleles present at ~10 or fewer copies in the starting template, reproducibility would also be limited by sampling noise, because these alleles may be over- or underrepresented in any particular reaction.

To characterize the quantitative accuracy of TAm-Seq as applied to circulating DNA, we simulated rare circulating tumor mutations by mixing plasma DNA from two healthy individuals. Using the same set of primers as used for the FFPE experiment, we identified that these two individuals differed at five known SNP loci (table S4). Total amplifiable copies in both plasma DNA samples were determined by digital PCR and mixed to obtain minor AFs ranging from 0.16% to 40% (Supplementary Methods). We sequenced diluted templates containing between 250 and <1 expected copy of the minor allele (table S5). The coefficient of variation (CV) of the observed AFs was equal on average to the inverse square root ($1/\sqrt{n}$) of the expected number of copies of the rare allele (Fig. 3A), which is the theoretical

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

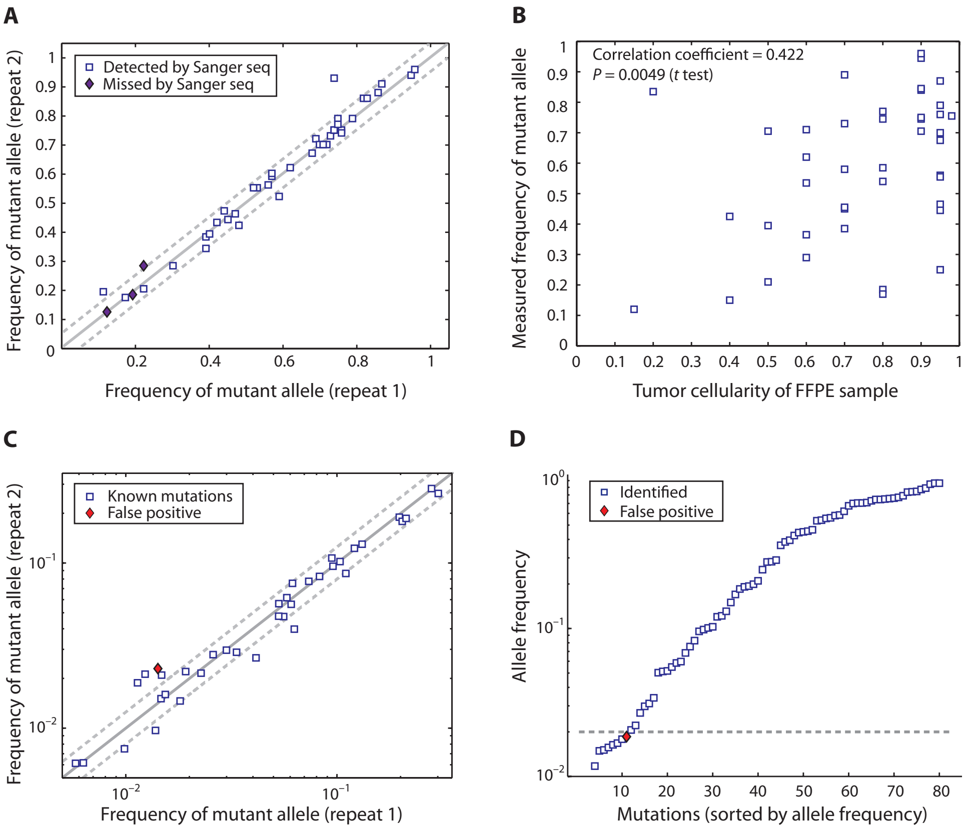


Fig. 2. Identification of mutations in ovarian cancer FFPE samples by TAm-Seq. (**A**) Concordance between duplicate measurements of AFs of mutations identified in fragmented DNA extracted from FFPE samples. The mutation frequency in each library was calculated as the fraction of reads with the mutant (nonreference) base. Solid line indicates equality. Dotted lines indicate a difference in AF of 0.05. (**B**) Correlation of AF with FFPE tumor cellularity. The measured mutant AF (average of both repeats) correlated significantly with the cellularity, estimated from histology (table S3). (**C**) Concordance between duplicate measurements of AFs of mutations identified in a mixture of DNA extracted from different FFPE samples. (**D**) Summary of mutations called in FFPE using TAm-Seq, sorted by increasing AF. Dotted line indicates AF of 2%.

limit of accuracy set by the Poisson distribution for independently segregating molecules. We compared the observed AF to the expected AF for cases where more than six copies of the minor allele were expected. Of 24 such cases, the root mean square (RMS) relative error between the expected and the observed frequency was 14%, with only 2 of 24 cases exhibiting more than 20% discrepancy. For samples with expected minor AF of 0.025, the RMS error was 23% (Fig. 3B).

Noninvasive identification of cancer mutations in plasma circulating DNA

We applied TAm-Seq to directly identify mutations in plasma of cancer patients. We studied a cohort of samples from individuals with HGSOC. These samples were first analyzed for tumor-specific mutations using digital PCR (Supplementary Methods), a method that is highly accurate (*2*, *3*, *7*, *37*) but requires design and validation of a different assay for every mutation screened and relies on previous identification of mutations in tumor samples from the same patients (*2*, *3*). We initially selected for analysis seven cases that had relatively high levels of circulating mutant *TP53* DNA in the plasma (as assessed by digital PCR). Using the equivalent amount of DNA present in 30 to 120 μl of plasma, we performed duplicate preamplification reactions for each sample. For all seven patients, *TP53* tumor mutations were identified in the circulating DNA at frequencies of 4% to 44% (Table 1). In one plasma sample collected from an ovarian cancer patient at relapse, we also identified a de novo mutation in the tyrosine kinase domain of *EGFR* (exon 21), at AF of 6% (patient 27, Table 1). We subsequently validated the presence of this mutation in plasma by performing replicate Sanger sequencing reactions of highly diluted template (Supplementary Methods), and 4 of 91 wells that were successfully Sanger-sequenced contained the *EGFR* mutation (fig. S4). We further validated the presence of this mutation by designing a sequence-specific TaqMan probe targeting this mutation and performing digital PCR (Table 1). The mutation was also identified by TAm-Seq in additional plasma collected from the same individual (sample 16, Table 2). This mutation in *EGFR* was not found in the ovarian mass removed by interval debulking surgery 15 months before the blood sample was collected, although the same sample did contain the concomitant *TP53* mutation found in the same patient's plasma, at AF of 85% (patient 27, table S3). We subsequently used TAm-Seq to sequence seven additional samples collected at the time of initial surgery including deposits in right and left ovaries and omentum. The *EGFR* mutation was detected in the two omental samples above the 0.99 confidence margin (fig. S3) at AF of 0.7%, but was not detected in the six ovarian samples (below the 0.8 confidence margin). Without previous identification in plasma, this mutation would not have been directly identified on screening those samples using high-specificity mutation identification criteria owing to its low AF. In contrast, the *TP53* mutation was identifiable in all biopsy and plasma samples (Fig. 4A). The frequency of mutant alleles in the relapsed tumor could not be directly assessed because a biopsy at relapse was not available.

We validated the TAm-Seq method on a larger panel of plasma samples in which levels of tumor-specific mutations were measured in parallel using patient-specific digital PCR assays. DNA extracted from 62 additional plasma samples collected at different time points from 37 patients with advanced HGSOC was amplified in duplicate (table S6), using DNA present in ~0.15 ml of plasma per reaction (range, 0.06 to 0.2 ml). Amplicon libraries were tagged and pooled together for sequencing with libraries prepared from 24 control samples. This generated an average sequencing depth of 650 for 62 plasma samples, sufficient to detect mutations present at AFs of 1% to 2%. Of >1.5 million possible substitutions, 42 mutations were called using the parameters previously optimized for FFPE analysis (table S6).

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

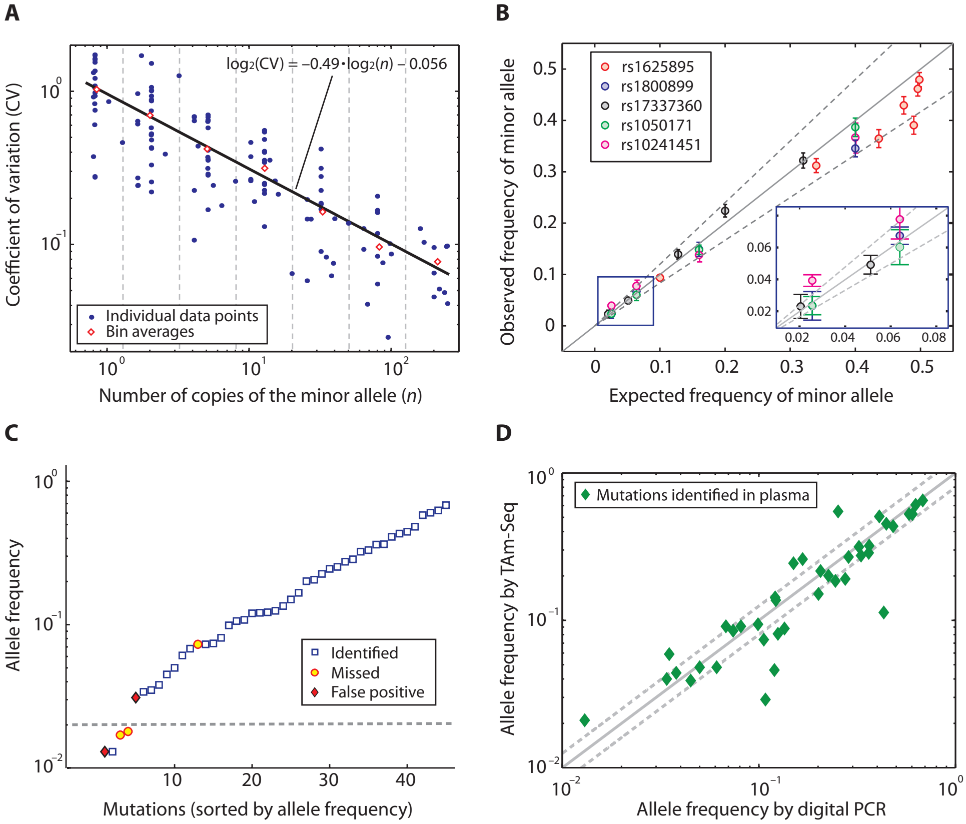


Fig. 3. Noninvasive identification and quantification of cancer mutations in plasma DNA by TAm-Seq. (**A**) Sampling noise in sequencing of sparse DNA using dilutions of plasma DNA from healthy individuals. CV of triplicate AF readings was calculated for each of the five SNPs in each of the mixes, which had varying numbers of copies of the minor allele (*n*) (blue dots). Bin averages (red diamonds) are the mean CVs calculated for each bin (bin edges denoted by the dotted vertical lines). A linear fit to the $\log_2$ of the mean CV as a function of the $\log_2$ expected copy number was calculated (black line). Two data points, with ($n = 100$, CV = 0.0064) and ($n = 32$, CV = 0.0185), were omitted from the figure for enhanced scaling. Three data points with minor allele copies of <0.8 were omitted from the analysis ($n = 0.51$, CV = 0.62; $n = 0.41$, CV = 0.86; $n = 0.20$, CV = 0.99). (**B**) Expected versus observed frequency of rare alleles in a dilution series of circulating DNA. Mean observed frequency was calculated for each of five SNPs for samples, where expected initial number of minor allele copies was greater than 6. Expected frequencies were calculated on the basis of quantification by digital PCR. Dotted lines represent 20% deviation from the expected frequencies. Inset highlights cases with expected minor AF <0.025. (**C**) Mutations identified in 62 plasma samples from patients with advanced HGSOC using TAm-Seq. AFs are based on digital PCR measurement for confirmed mutations (identified or missed by TAm-Seq), and on TAm-Seq for the false positives called using parameters optimized for analysis of FFPE samples. The dashed horizontal line indicates AF of 2%. Mutations detected by digital PCR at AF <1% are not shown. (**D**) AFs measured by TAm-Seq versus digital PCR for mutations identified in plasma DNA.

Table 1. Mutations identified by TAm-Seq in plasma samples from seven ovarian cancer patients. TAm-Seq was used to sequence DNA extracted from plasma of subjects with HGSOC (stage III/IV at diagnosis). Plasma was collected when patients presented with relapse disease, before initiation of chemotherapy. For patient 46, DNA from a formalin-fixed, paraffin-embedded (FFPE) sample was not included in the TAm-Seq set and the mutation was validated in FFPE by Sanger sequencing. CA125 was measured at time of plasma collection. Mean depth of coverage at the mutation locus in the TAm-Seq data was averaged over the repeats (RMS deviation = 850). AF, allele frequency; N, no; Y, yes.

Patient ID	Age at diagnosis	Time elapsed since surgery (months); number of previous lines of chemotherapy	CA125 (U/ml)	Plasma per amplification reaction (µl)	Gene	Mutation and base change (genome build hg19)		Protein change	Detected in FFPE	Mean depth (sequencing reads)	Mean AF using TAm-Seq	Mean AF using digital PCR
8	60	13; 1	2122	50	*TP53*	17:7577120	C>T	p.R273H	Y	5000	0.09	0.10
12	62	27; 3	365	50	*TP53*	17:7577579	G>T	p.Y234*	Y	5000	0.10	0.08
14	58	50; 3	260	120	*TP53*	17:7578212	G>A	p.R213*	Y	5800	0.15	0.12
25	61	9; 1	944	110	*TP53*	17:7578404	A>T	p.C176S	Y	4800	0.04	0.08
27†	68	15; 1	1051	90	*TP53*	17:7578262	C>G	p.R196P	Y	7700	0.06	0.14
					EGFR	7:55259437	G>A	p.R832H	N	5700	0.06	0.05
31	64	12; 1	313	30	*TP53*	17:7578406	C>T	p.R175H	Y	4500	0.44	0.56
46	56	30; 2	1509	30	*TP53*	17:7578406	C>T	p.R175H	Y	4200	0.23	0.30

*Indicates stop codon. †Both a *TP53* and an *EGFR* mutation were identified in this sample (Fig. 4A).

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Table 2. Mutations identified by TAm-Seq in a set of 62 plasma samples from ovarian cancer patients. Forty mutations were identified by TAm-Seq using stringent parameters for mutation calling. Plasma samples described in this table are distinct from those in Table 1, but patients included overlap. Additional data on patients and mutations are provided in table S6.

Sample number	Plasma volume per amplification reaction (μl)	DNA amount per amplification reaction (ng)	Gene	Protein change	Mean depth (sequencing reads)	Mean AF using TAm-Seq	Mean AF using digital PCR
1	70	0.9	*TP53*	p.R273C	640	0.260	0.167
2	160	4.2	*TP53*	p.R248Q	340	0.244	0.150
3	160	5.7	*TP53*	p.R248Q	640	0.507	0.410
4	120	9.9	*TP53*	p.R213X	810	0.059	0.035
5	120	1.4	*TP53*	p.C141Y	680	0.021	0.013
6	120	2.1	*TP53*	p.C141Y	720	0.044	0.038
7	190	17.9	*TP53*	p.I195N	800	0.091	0.081
8	160	14.8	*TP53*	p.R175H	510	0.608	0.627
9	160	10.7	*TP53*	p.R175H	550	0.526	0.604
10	160	6.1	*TP53*	p.R175H	530	0.651	0.682
11	160	4.9	*TP53*	p.R175H	490	0.526	0.581
13	160	2.8	*TP53*	p.C135R	480	0.039	0.045
14	160	2.5	*TP53*	p.C135R	610	0.046	0.120
15	160	3.0	*TP53*	p.C135R	470	0.091	0.068
16†	130	3.7	*TP53*	p.R196P	1070	0.088	0.135
			EGFR	p.R832H	614	0.048	0.050
17	160	4.2	*TP53*	p.C176S	580	0.113	0.432
18	160	4.4	*TP53*	p.C176S	620	0.029	0.108
20	140	5.2	*TP53*	p.R175H	650	0.201	0.226
21	140	3.6	*TP53*	p.R175H	650	0.085	0.074
22	140	4.1	*TP53*	p.R175H	630	0.081	0.125
23	140	3.7	*TP53*	p.R175H	710	0.074	0.106
24	140	7.1	*TP53*	p.R175H	760	0.269	0.286
25	130	3.9	*TP53*	p.R273H	750	0.094	0.099
26	160	5.7	*TP53*	p.R282W	640	0.048	0.061
27	150	3.6	*TP53*	p.C141Y	480	0.321	0.364
29	150	9.5	*TP53*	p.E258K	190	0.548	0.253
31	160	3.6	*TP53*	p.C135Y	620	0.040	0.034
32	140	2.4	*TP53*	p.E56X	1480	0.137	0.122
33	160	13.2	*TP53*	p.K132N	740	0.216	0.206
34	60	5.3	*TP53*	p.K132N	570	0.151	0.201
36	160	5.8	*TP53*	p.K132N	620	0.191	0.275
37	160	9.4	*TP53*	p.K132N	530	0.287	0.362
38	160	10.1	*TP53*	p.K132N	590	0.275	0.331
39	160	16.4	*TP53*	p.K132N	700	0.315	0.323
40	160	19.7	*TP53*	p.K132N	830	0.435	0.482
41	160	15.0	*TP53*	p.K132N	730	0.452	0.445
42	160	8.5	*TP53*	p.K132N	560	0.185	0.245
43	150	3.6	*TP53*	Splicing	680	0.143	0.121
44‡	170	5.2	*TP53*	p.C238R	1543	0.071	0.073

†Both a *TP53* and an *EGFR* mutation were identified in this sample, collected from patient 27 (Table 1), 25 months after initial surgery (Fig. 4A). ‡The amplicon containing the mutation failed amplification in this sample in the initial experiment and was identified successfully in repeat analysis.

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Thirty-nine of these matched mutations detected by digital PCR in those samples (Fig. 3C). Three potential false positives were called, at AF of 3.1%, 1.3%, and 0.7% (the latter in a control sample). Using higher-stringency parameters for mutation identification (Supplementary Methods), we retained only the 39 validated mutations called, with no false positives (Table 2).

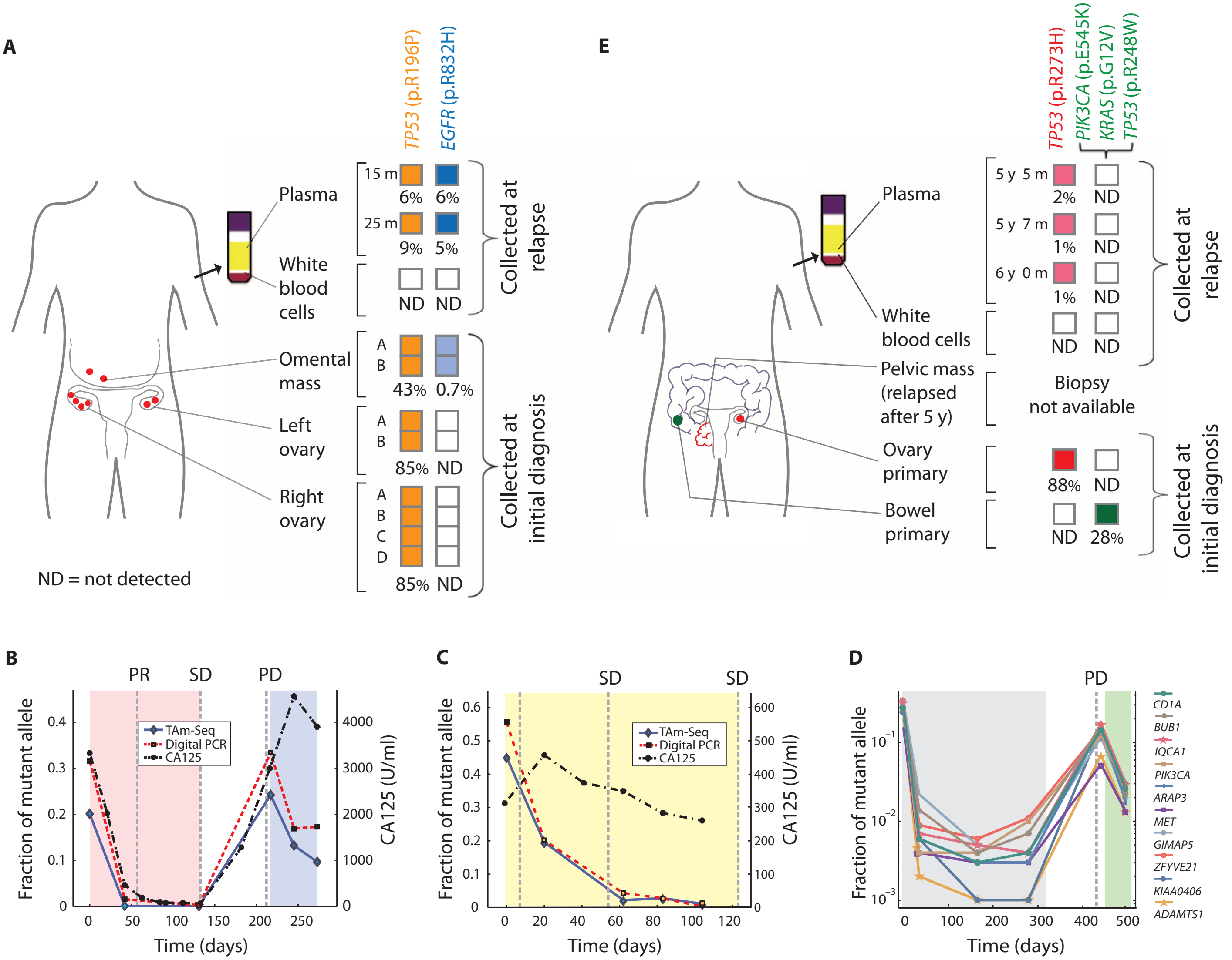


Fig. 4. Clinically relevant applications of plasma DNA sequencing using TAm-Seq. (**A**) Retrospective analysis by TAm-Seq of plasma samples collected during patient follow-up and biopsy specimens collected at initial surgery. We identified a mutation in exon 21 of *EGFR* (dark blue boxes) in two separate plasma samples, collected 15 and 25 months after initial surgery from patient 27 (Tables 1 and 2). This mutation was not directly identified in eight tumor biopsy specimens collected at the time of initial surgery (two from omental mass, two from left ovary, and four from right ovary). Having identified the mutation in the plasma samples, we examined this mutation using the lower-specificity criteria defined for mutation detection (Supplementary Methods) and detected the mutation in the two specimens that had been collected from the omentum at the time of surgery (light blue boxes) but not in the six ovarian specimens. A mutation in *TP53* was identified in all tumor and plasma samples collected from this patient (Tables 1 and 2 and table S3), but not in white blood cells (buffy coat). Percentages indicate mutant AFs. Empty boxes and "ND" indicate samples where a mutation was not identified or detected (below 0.8 confidence margin). (**B**) Monitoring frequency of mutant DNA in plasma of an ovarian cancer patient (patient 46) over time using TAm-Seq and digital PCR. TAm-Seq results are reported as the mean frequency of duplicate analyses. Parallel data are shown for digital PCR and serum CA125. Shaded regions indicate periods of chemotherapy, and vertical dashed lines indicate radiological assessment of patient responses: PR, partial response; SD, stable disease; PD, progressive disease. (**C**) Monitoring frequency of mutant DNA in plasma of an ovarian cancer patient (patient 31) over time. (**D**) Dynamics of 10 tumor-specific mutations in plasma of a breast cancer patient (not included in the other sets of samples analyzed). (**E**) Retrospective analysis of samples from synchronous primary tumors (bowel and ovarian) collected at the time of initial surgery and three plasma samples collected at relapse. In primary tumors from this patient (not included in the other sets of samples analyzed), a *TP53* mutation was identified in the ovarian cancer (red box), and mutations in *PIK3CA*, *KRAS*, and *TP53* were identified in the bowel cancer (green box). At relapse, a biopsy was not performed on the pelvic mass. The *TP53* mutation that was identified in the ovarian primary tumor (p.R273H) was detected in plasma, whereas the bowel-associated mutations were not detected.

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Of 40 point mutations detected at AF >2% by digital PCR, 38 (95%) were identified by TAm-Seq in a single experiment (Fig. 3C). One additional mutation was located in an amplicon that failed in that sample and was identified in repeated analysis; the other was likely missed by TAm-Seq owing to sampling noise, because it was found in one of the duplicate preamplified libraries but not the other (table S6). One of three mutations detected by digital PCR at 1% < AF < 2% was identified by TAm-Seq (Fig. 3C). Eleven additional point mutations detected by digital PCR at AF <1% were not detected by TAm-Seq at these settings. TAm-Seq and digital PCR measurements of AF had excellent agreement, with correlation coefficient of 0.90, increasing to 0.97 when discarding the two strongest outliers (Fig. 3D). Thus, we screened 62 samples across sizeable genomic stretches, using minute amounts of plasma DNA (median, 4 ng), and obtained 97.5% sensitivity with PPV of 100% for identifying mutations at AF >2% in plasma by TAm-Seq. Using parameters optimized for FFPE samples, one potential false positive was called at AF >2%, reducing the PPV to 97.5% (Table 3).

Monitoring levels of ctDNA

Various methods have been suggested to monitor changes in mutation load in plasma. These can have enhanced sensitivity compared to TAm-Seq for tracking individual mutations, but require design of personalized assays (*3*, *18*, *19*). None of these methods have been widely adopted. We therefore applied TAm-Seq as a generic tool to measure changes in the frequency of ctDNA over time. We studied serial plasma samples collected during follow-up and treatment of two patients with relapsed HGSOC, collected during 104 and 273 days of follow-up and treatment, respectively. Frequencies of mutant *TP53* alleles were measured by TAm-Seq and in parallel by digital PCR using a mutation-specific probe. The two methods of quantification had excellent agreement. Mutant AFs in plasma of ovarian cancer patients reflected well the clinical course of the disease compared to the serum marker CA125, showed marked decrease when systemic treatment was initiated, and increased in parallel to disease progression. In the first case (Fig. 4B), a 56-year-old woman with relapsed ovarian cancer (patient 46) was treated with fourth-line carboplatin + paclitaxel chemotherapy for six cycles (pink-shaded region). Radiology showed partial response on mid-treatment computed tomography (CT) scan. End-of-treatment CT showed stable disease. Twelve weeks from the end of her fourth-line treatment, the patient developed progressive disease. The patient then initiated fifth-line chemotherapy with liposomal doxorubicin (purple-shaded region). In the second case (Fig. 4C), a 64-year-old woman with relapsed ovarian cancer (patient 31) was treated with second-line ECX (epirubicin, cisplatin, and capecitabine) chemotherapy for six cycles. Radiology showed stable disease on mid- and end-of-treatment CT scans. The patient then remained off treatment, until she progressed 3 months later.

TAm-Seq can be flexibly adapted to sequence different genomic regions by designing primers to amplify regions of interest. We used this capability to study dynamics of multiple mutations in parallel.

Table 3. Summary of mutations identified in 69 plasma samples of ovarian cancer patients. Samples were analyzed by TAm-Seq and in parallel by digital PCR. Using parameters optimized for plasma DNA, false-positive calls were lost, whereas all confirmed calls were retained, resulting in specificity and PPV of 100%.

First set of plasma samples	
Plasma samples analyzed	7
Point mutations originally detected by digital PCR, using patient-specific assays targeting mutations identified in tumor samples	7
Point mutations identified directly in plasma by TAm-Seq	8
De novo mutations identified by TAm-Seq only, subsequently confirmed by digital PCR	1
Second set of plasma samples	
Plasma samples analyzed	62
Point mutations detected by digital PCR at AF >2%	40
Point mutations with AF >2% (by digital PCR) identified by TAm-Seq	39
Point mutations missed by TAm-Seq due to sampling error	1
Sensitivity of TAm-Seq for identifying mutations at AF >2%	97.5%
PPV of mutations called by TAm-Seq with AF >2%	97.5%*
ctDNA in ovarian cancer	
Advanced ovarian cancer patients in both sets†	38
Patients where TAm-Seq identified cancer mutations	20

*One unconfirmed substitution was called at AF >2% using parameters optimized for FFPE material. (table S6), 6 of whom overlap.

†The first set included 7 patients (Table 1), and the second set included 37 patients

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Whole-genome sequencing of tumor material was used to identify tumor mutations in a patient with metastatic breast cancer undergoing two phases of chemotherapy. Ten mutations were selected, and short amplicons (<120 bp) were designed to cover the mutation loci (table S7). Serial plasma samples were collected over the course of 497 days, both before and after treatment. We performed TAm-Seq in duplicate, using DNA from 0.08 ml of plasma per amplification, and tracked dynamics of all mutations in parallel (Fig. 4D). The patient was treated with single-agent epirubicin (gray-shaded region). After 4 months off treatment, a CT scan showed progressive disease and the patient commenced further treatment with paclitaxel chemotherapy. The 10 mutations followed a common pattern of sharp decline in AF upon onset of therapy and an increase in AF upon disease progression after termination of therapy (Fig. 4D).

Finally, we used TAm-Seq to study plasma from a patient who had a history of two synchronous primary cancers, bowel and ovarian, which were resected simultaneously. After a 5-year remission, a pelvic mass of uncertain origin was detected. A biopsy was considered to guide selection of therapy but was not performed owing to risk of complications and comorbidities. The patient commenced empirically on an ovarian cancer chemotherapy regimen, to which she responded. Retrospective analysis by TAm-Seq of FFPE from the primary tumors collected at initial surgery, and three plasma samples collected serially at the time of relapse (5 years and 5 months, 5 years and 7 months, and 6 years after initial surgery), showed that the patient's plasma at relapse contained the *TP53* (p.R273H) mutation identified in the ovarian primary tumor (exceeding the 0.98, 0.93, and 0.97 confidence margins, respectively), but not the *PIK3CA* (p.E545K), *KRAS* (p.G12V), or *TP53* (p.R248W) mutations identified in the primary bowel cancer (below the 0.8 confidence margin) (Fig. 4E). Had these results been available, uncertainty and treatment delays may have been avoided, as well as the risk of prescribing chemotherapy for an inappropriate tumor site. An alternative possible outcome may have involved a finding of the *PIK3CA* or *KRAS* mutations (present in the primary bowel cancer) in the patient's plasma at the time of relapse. Such a finding, if available to clinicians at the time, may not only have led to alternate chemotherapy being offered but may have also opened the possibility of enrolment into a trial for targeted therapy with mammalian target of rapamycin (mTOR), phosphatidylinositol 3-kinase (PI3K), or mitogen-activated protein kinase kinase (MEK) inhibitors (*11*).

DISCUSSION

Detection of rare mutations in circulating DNA has long been pursued owing to its potentially transformative impact on cancer diagnosis and management. Important progress has been made using sequence-specific assays that target predefined mutations and that detect extremely rare alleles. Assays such as PCR (*6*, *7*), ligation (*5*), and primer extension/mass spectrometry (*27*) can identify specific, predefined mutations in plasma samples. Enhanced detection down to 1 mutant allele in 10,000 or more wild-type alleles can be obtained using a variety of methods, such as peptide nucleic acid and primer extension ("PPEM") (*38*), ligation followed by quantitative PCR ("LigAmp") (*39*), bead-based digital PCR in emulsions ("BEAMing") (*2*, *3*), microfluidic-based (*7*) or droplet-based digital PCR (*40*), or microinsertion/deletion/indel-activated pyrophosphorolysis ("MAP") (*29*). Nonetheless, identification of rare mutations in tumor suppressor genes such as *TP53*, which are widely mutated in cancers but lack a well-defined hotspot region, remains an elusive goal.

In patients with advanced cancers, mutant alleles can reach a sizeable fraction of DNA. For example, Dukes' D colorectal cancers have median 8% mutant AF (*2*). Screening of entire genes for mutations would therefore be useful for some applications, even if analytical selectivity is limited to a few percent. Advances in massively parallel sequencing make new approaches possible. These have largely focused on large-scale analyses, including whole-genome or whole-exome sequencing (*41*). This generates a large amount of data on genomic regions that do not, at present, inform clinical decisions. Moreover, the depth of coverage for clinically significant loci is not sufficient to detect changes that occur at low frequency (<5%). Such approaches have recently been complemented by methods for examination of individual amplicons at great depth (*30*).

The intermediate scale of sequencing is most likely to have immediate impact on clinical genomics. Targeted sequencing has been applied for tumor DNA (*34*, *35*) and cyst fluid (*33*) to detect mutations down to 5% AF, but has not been applied for analysis of circulating tumor nucleic acids. Here, we demonstrate noninvasive identification of mutant alleles in plasma, at AFs as low as 2%, by targeted deep sequencing of circulating DNA. Our TAm-Seq method uses a combination of short amplicons, two-step amplification, sample barcodes, and high-throughput PCR. Because the amplicons are short, this method effectively amplifies even small amounts of fragmented DNA such as are present in circulating DNA. The two-step amplification permits extensive primer multiplexing that enables the amplification and sequencing of sizeable genomic regions by tiling short amplicons without loss of fidelity or efficiency. Duplicate sequencing of each sample is used to avoid false positives stemming from PCR errors. Sample barcodes and high-throughput PCR reduce the per-sample costs to a range where this may be widely applicable. Preparing TAm-Seq libraries for sequencing from 48 samples takes less than 24 hours and involves only few hours of hands-on time. New platforms for massively parallel sequencing allow for fast turnaround times, which make this approach practical in a clinical setting.

The sensitivity presently achieved can provide useful diagnostic information in certain advanced cancers. We studied a cohort of subjects with advanced HGSOC in which the tumor suppressor gene *TP53* is a driver mutation (*20*). Of the 69 plasma samples collected from 38 different individuals with advanced HGSOC, we identified mutations in *TP53* in 46 samples (67%) from 20 of the cases (53%). In contrast, a previous study using a ligase detection reaction with bespoke primers found mutated *TP53* sequences in plasma for only 30% of advanced ovarian cancer patients (*5*), and a study using single-strand conformation polymorphism found no ctDNA in preoperative plasma samples from high-grade serous cancer patients (*42*).

Targeted agents, such as inhibitors of poly(adenosine diphosphate–ribose) polymerase (PARP), or tyrosine kinase inhibitors targeting epidermal growth factor receptor (EGFR), may be applicable for systemic treatment of advanced HGSOC (*8*, *10*, *22*). In a recent study of 203 HGSOC tumors, *EGFR* was found to be the most frequently mutated oncogene and was mutated in nearly 10% of cases (*10*). In one case, we identified in plasma a de novo mutation in the tyrosine kinase domain (exon 21) of *EGFR*, located 26 amino acids upstream of the L858R activating mutation widely documented for lung cancer. In a subset of tumor samples collected from the same patient 15 months earlier, this mutation was detected at AF of 0.7%, but could not have

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

been identified by analysis of those samples alone without previous knowledge of the mutation identified in plasma (Fig. 4A). In a clinical setting, identification of such a mutation could potentially guide treatment with alternative molecularly targeted therapy (*10*). Current clinical recommendations in lung adenocarcinoma suggest mutation assessment in exons 18 to 21 of *EGFR* (a region of ~560 bp) in the tumor tissue to identify patients eligible for treatment with gefitinib or erlotinib (*9*). Using a commercial PCR-based in vitro diagnostic kit (Qiagen), 28 different *EGFR* variants can be assayed (not including the mutation we identified), but the sample needs to be subdivided into seven different reactions. When sample is limited or mutant alleles are rare, this could introduce sampling errors.

Using standard amplification primers tailored to the mutation loci, we also used TAm-Seq to monitor the dynamics of 10 mutations in plasma DNA of a single patient with metastatic breast cancer, using minute amounts of input DNA. Previous studies have followed up to two mutations in any individual patient (*3*, *19*). Tracking multiple mutations can provide insight into clonal evolution and, at the same time, increases the robustness for tumor monitoring by compensating for effects of sampling noise or mutational drift. For example, if a patient has only five copies of a mutant allele per milliliter of plasma (on average), there is a 37% probability that this mutation will not be present in a 0.2-ml sample, and even a perfect assay will fail to detect residual tumor, whereas a method that measures multiple mutations in parallel can have a low likelihood of a false-negative result even if the detection rate for each mutation is less than 50%.

A current limitation of TAm-Seq is the detection limit compared to assays that target individual loci (*2*, *3*, *7*, *40*), which have been shown to detect two to three orders of magnitude lower frequencies. Our approach may be sufficient for analyzing plasma from patients with certain advanced cancers, but further improvement may be necessary before this method can be more widely used in the clinic. Higher read depth or fidelity, additional replicates, or improved algorithms could allow for enhanced mutation detection without change to protocols. An alternative strategy is through rare allele enrichment, for example, by combining TAm-Seq with protocols such as COLD-PCR (co-amplification at lower denaturation temperature PCR) (*31*).

Previously proposed methods for personalized monitoring of tumor dynamics relied on expensive custom-designed probes (*3*) or identification of rearrangements using whole-genome sequencing (*18*, *19*). These have better analytical sensitivity than currently achieved by TAm-Seq, but are difficult to implement on a routine basis. TAm-Seq strikes a balance between sensitivity and ease of use and could facilitate study and application of circulating DNA. Using TAm-Seq, we identified cancer mutations in the plasma of most advanced ovarian cancer patients and tracked dynamics of *TP53* mutations without requiring any specially designed probes. In summary, TAm-Seq is a flexible and cost-effective platform for applications in noninvasive cancer genomics and diagnostics. We have shown that this method can be used for high-throughput sequencing of plasma samples to identify and monitor levels of multiple cancer mutations in circulating DNA. This could also be applied to screen for rare mutations in a variety of heterogeneous sample types such as low-cellularity tumor specimens, cytological samples, or circulating tumor cells (*16*). With further developments, this and derivative methods may be applied in molecular screening for earlier detection or for differential diagnosis of cancer from benign masses. For genetic analysis of FFPE or small biopsy samples, TAm-Seq can be applied as is, as a cost-effective clinical aid.

MATERIALS AND METHODS

Sample collection

FFPE blocks were obtained from the pathology archives at Addenbrooke's Hospital (Cambridge, UK). Plasma samples were collected upon disease relapse, before and during chemotherapy treatment. Sample collection for this study was approved by Cambridgeshire Research Ethics Committee (REC 08/H0306/61 and 07/Q0106/63). Peripheral blood samples were collected into EDTA tubes and centrifuged at 820*g* for 10 min within 1 hour of collection to limit degradation of cell-free DNA and leukocyte lysis. Aliquots (1 ml) of plasma were centrifuged in a benchtop microfuge at 14,000 rpm for 10 min. The supernatant was transferred to sterile 1.5-ml tubes and stored at −80°C before extraction.

Extraction of DNA from FFPE and blood plasma

Paraffin blocks were cut as 8-μm sections on plain glass slides. Targeted regions for sampling were marked on adjacent hematoxylin and eosin sections by the study pathologist and recovered by scrape macrodissection. Between 3 and 20 sections were macrodissected depending on the tissue sample's size. DNA from FFPE sections was extracted with QIAamp DNA FFPE Tissue Kit (Qiagen) according to the manufacturer's instructions.

Circulating DNA was extracted from between 0.85 and 2.2 ml of plasma with the QIAamp Circulating Nucleic Acid kit (Qiagen), following the manufacturer's instructions, and with the QIAvac 24 Plus vacuum manifold. Carrier RNA was added to ACL lysis buffer to enhance binding of nucleic acids to the QIAamp membrane with the aim to enhance yields.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/4/136/136ra68/DC1

Methods

Fig. S1. PCR strategy and primer design.

Fig. S2. Sanger traces for mutations identified by tagged-amplicon sequencing.

Fig. S3. Background frequencies and detection limits for base substitutions.

Fig. S4. Replicate dilute Sanger sequencing of a mutation identified in plasma.

Table S1. Target-specific primers.

Table S2. Unique sequencing barcodes.

Table S3. Mutations identified in FFPE samples.

Table S4. SNPs identified in circulating DNA from two plasma control samples.

Table S5. Frequency of SNP alleles in dilution series of DNA from control plasma.

Table S6. Additional data for Table 2 for mutations identified in plasma samples.

Table S7. Mutations and amplicons studied in one breast cancer patient.

REFERENCES AND NOTES

1. E. Gormally, E. Caboux, P. Vineis, P. Hainaut, Circulating free DNA in plasma or serum as biomarker of carcinogenesis: Practical aspects and biological significance. *Mutat. Res.* **635**, 105–117 (2007).
2. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz Jr., S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16368–16373 (2005).
3. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
4. H. Schwarzenbach, D. S. Hoon, K. Pantel, Cell-free nucleic acids as biomarkers in cancer patients. *Nat. Rev. Cancer* **11**, 426–437 (2011).
5. E. M. Swisher, M. Wollan, S. M. Mahtani, J. B. Willner, R. Garcia, B. A. Goff, M. C. King, Tumor-specific p53 sequences in blood and peritoneal fluid of women with epithelial ovarian cancer. *Am. J. Obstet. Gynecol.* **193**, 662–667 (2005).

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

6. R. E. Board, A. M. Wardley, J. M. Dixon, A. C. Armstrong, S. Howell, L. Renshaw, E. Donald, A. Greystoke, M. Ranson, A. Hughes, C. Dive, Detection of *PIK3CA* mutations in circulating free DNA in patients with breast cancer. *Breast Cancer Res. Treat.* **120**, 461–467 (2010).
7. T. K. Yung, K. C. Chan, T. S. Mok, J. Tong, K. F. To, Y. M. Lo, Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non–small cell lung cancer patients. *Clin. Cancer Res.* **15**, 2076–2084 (2009).
8. S. Banerjee, S. Kaye, The role of targeted therapy in ovarian cancer. *Eur. J. Cancer* **47** (Suppl. 3), S116–S130 (2011).
9. V. L. Keedy, S. Temin, M. R. Somerfield, M. B. Beasley, D. H. Johnson, L. M. McShane, D. T. Milton, J. R. Strawn, H. A. Wakelee, G. Giaccone, American Society of Clinical Oncology provisional clinical opinion: Epidermal growth factor receptor (*EGFR*) mutation testing for patients with advanced non–small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy. *J. Clin. Oncol.* **29**, 2121–2127 (2011).
10. U. A. Matulonis, M. Hirsch, E. Palescandolo, E. Kim, J. Liu, P. van Hummelen, L. MacConaill, R. Drapkin, W. C. Hahn, High throughput interrogation of somatic mutations in high grade serous cancer of the ovary. *PLoS One* **6**, e24433 (2011).
11. J. A. Engelman, L. Chen, X. Tan, K. Crosby, A. R. Guimaraes, R. Upadhyay, M. Maira, K. McNamara, S. A. Perera, Y. Song, L. R. Chirieac, R. Kaur, A. Lightbown, J. Simendinger, T. Li, R. F. Padera, C. García-Echeverría, R. Weissleder, U. Mahmood, L. C. Cantley, K. K. Wong, Effective use of PI3K and MEK inhibitors to treat mutant *Kras* G12D and *PIK3CA* H1047R murine lung cancers. *Nat. Med.* **14**, 1351–1356 (2008).
12. P. J. Campbell, S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. A. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, C. A. Griffin, J. Burton, H. Swerdlow, M. A. Quail, M. R. Stratton, C. Iacobuzio-Donahue, P. A. Futreal, The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
13. S. P. Shah, R. D. Morin, J. Khattra, L. Prentice, T. Pugh, A. Burleigh, A. Delaney, K. Gelmon, R. Guliany, J. Senz, C. Steidl, R. A. Holt, S. Jones, M. Sun, G. Leung, R. Moore, T. Severson, G. A. Taylor, A. E. Teschendorff, K. Tse, G. Turashvili, R. Varhol, R. L. Warren, P. Watson, Y. Zhao, C. Caldas, D. Huntsman, M. Hirst, M. A. Marra, S. Aparicio, Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
14. M. Inukai, S. Toyooka, S. Ito, H. Asano, S. Ichihara, J. Soh, H. Suehisa, M. Ouchida, K. Aoe, M. Aoe, K. Kiura, N. Shimizu, H. Date, Presence of *epidermal growth factor receptor gene* T790M mutation as a minor clone in non–small cell lung cancer. *Cancer Res.* **66**, 7854–7858 (2006).
15. S. L. Edwards, R. Brough, C. J. Lord, R. Natrajan, R. Vatcheva, D. A. Levine, J. Boyd, J. S. Reis-Filho, A. Ashworth, Resistance to therapy caused by intragenic deletion in *BRCA2*. *Nature* **451**, 1111–1115 (2008).
16. S. Maheswaran, L. V. Sequist, S. Nagrath, L. Ulkus, B. Brannigan, C. V. Collura, E. Inserra, S. Diederichs, A. J. Iafrate, D. W. Bell, S. Digumarthy, A. Muzikansky, D. Irimia, J. Settleman, R. G. Tompkins, T. J. Lynch, M. Toner, D. A. Haber, Detection of mutations in *EGFR* in circulating lung-cancer cells. *N. Engl. J. Med.* **359**, 366–377 (2008).
17. B. Norquist, K. A. Wurz, C. C. Pennil, R. Garcia, J. Gross, W. Sakai, B. Y. Karlan, T. Taniguchi, E. M. Swisher, Secondary somatic mutations restoring *BRCA1/2* predict chemotherapy resistance in hereditary ovarian carcinomas. *J. Clin. Oncol.* **29**, 3008–3015 (2011).
18. R. J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F. M. De La Vega, K. W. Kinzler, B. Vogelstein, L. A. Diaz Jr., V. E. Velculescu, Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl. Med.* **2**, 20ra14 (2010).
19. D. J. McBride, A. K. Orpana, C. Sotiriou, H. Joensuu, P. J. Stephens, L. J. Mudie, E. Hämäläinen, L. A. Stebbings, L. C. Andersson, A. M. Flanagan, V. Durbecq, M. Ignatiadis, O. Kallioniemi, C. A. Heckman, K. Alitalo, H. Edgren, P. A. Futreal, M. R. Stratton, P. J. Campbell, Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer* **49**, 1062–1069 (2010).
20. A. A. Ahmed, D. Etemadmoghadam, J. Temple, A. G. Lynch, M. Riad, R. Sharma, C. Stewart, S. Fereday, C. Caldas, A. Defazio, D. Bowtell, J. D. Brenton, Driver mutations in *TP53* are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221**, 49–56 (2010).
21. Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
22. R. C. Bast Jr., Molecular approaches to personalizing management of ovarian cancer. *Ann. Oncol.* **22** (Suppl. 8), viii5–viii15 (2011).
23. K. C. Chan, S. F. Leung, S. W. Yeung, A. T. Chan, Y. M. Lo, Persistent aberrations in circulating DNA integrity after radiotherapy are associated with poor prognosis in nasopharyngeal carcinoma patients. *Clin. Cancer Res.* **14**, 4141–4145 (2008).
24. H. C. Fan, Y. J. Blumenfeld, U. Chitkara, L. Hudgins, S. R. Quake, Analysis of the size distributions of fetal and maternal cell-free DNA by paired-end sequencing. *Clin. Chem.* **56**, 1279–1286 (2010).
25. Y. M. Lo, K. C. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
26. R. W. Chiu, R. Akolekar, Y. W. Zheng, T. Y. Leung, H. Sun, K. C. Chan, F. M. Lun, A. T. Go, E. T. Lau, W. W. To, W. C. Leung, R. Y. Tang, S. K. Au-Yeung, H. Lam, Y. Y. Kung, X. Zhang, J. M. van Vugt, R. Minekawa, M. H. Tang, J. Wang, C. B. Oudejans, T. K. Lau, K. H. Nicolaides, Y. M. Lo, Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: Large scale validity study. *BMJ* **342**, c7401 (2011).
27. Y. M. Lo, N. B. Tsui, R. W. Chiu, T. K. Lau, T. N. Leung, M. M. Heung, A. Gerovassili, Y. Jin, K. H. Nicolaides, C. R. Cantor, C. Ding, Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. *Nat. Med.* **13**, 218–223 (2007).
28. Y. M. Lo, F. M. Lun, K. C. Chan, N. B. Tsui, K. C. Chong, T. K. Lau, T. Y. Leung, B. C. Zee, C. R. Cantor, R. W. Chiu, Digital PCR for the molecular detection of fetal chromosomal aneuploidy. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 13116–13121 (2007).
29. Z. Chen, J. Feng, C. H. Buzin, Q. Liu, L. Weiss, K. Kernstine, G. Somlo, S. S. Sommer, Analysis of cancer mutation signatures in blood by a novel ultra-sensitive assay: Monitoring of therapy or recurrence in non-metastatic breast cancer. *PLoS One* **4**, e7220 (2009).
30. I. Kinde, J. Wu, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9530–9535 (2011).
31. J. Li, L. Wang, H. Mamon, M. H. Kulke, R. Berbeco, G. M. Makrigiorgos, Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. *Nat. Med.* **14**, 579–584 (2008).
32. S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton, P. A. Futreal, COSMIC: Mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
33. J. Wu, H. Matthaei, A. Maitra, M. Dal Molin, L. D. Wood, J. R. Eshleman, M. Goggins, M. I. Canto, R. D. Schulick, B. H. Edil, C. L. Wolfgang, A. P. Klein, L. A. Diaz Jr., P. J. Allen, C. M. Schmidt, K. W. Kinzler, N. Papadopoulos, R. H. Hruban, B. Vogelstein, Recurrent *GNAS* mutations define an unexpected pathway for pancreatic cyst development. *Sci. Transl. Med.* **3**, 92ra66 (2011).
34. O. Harismendy, R. B. Schwab, L. Bao, J. Olson, S. Rozenzhak, S. K. Kotsopoulos, S. Pond, B. Crain, M. S. Chee, K. Messer, D. R. Link, K. A. Frazer, Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. *Genome Biol.* **12**, R124 (2011).
35. N. Wagle, M. F. Berger, M. J. Davis, B. Blumenstiel, M. DeFelice, P. Pochanard, M. Ducar, P. Van Hummelen, L. E. MacConaill, W. C. Hahn, M. Meyerson, S. B. Gabriel, L. A. Garraway, High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.* **2**, 82 (2012).
36. T. Sjöblom, S. Jones, L. D. Wood, D. W. Parsons, J. Lin, T. D. Barber, D. Mandelker, R. J. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. D. Markowitz, J. Willis, D. Dawson, J. K. Willson, A. F. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. H. Park, K. E. Bachman, N. Papadopoulos, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
37. B. Vogelstein, K. W. Kinzler, Digital PCR. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9236–9241 (1999).
38. X. Sun, K. Hung, L. Wu, D. Sidransky, B. Guo, Detection of tumor mutations in the presence of excess amounts of normal DNA. *Nat. Biotechnol.* **20**, 186–189 (2002).
39. C. Shi, S. H. Eshleman, D. Jones, N. Fukushima, L. Hua, A. R. Parker, C. J. Yeo, R. H. Hruban, M. G. Goggins, J. R. Eshleman, LigAmp for sensitive detection of single-nucleotide differences. *Nat. Methods* **1**, 141–147 (2004).
40. L. B. Pinheiro, V. A. Coleman, C. M. Hindson, J. Herrmann, B. J. Hindson, S. Bhat, K. R. Emslie, Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal. Chem.* **84**, 1003–1011 (2012).
41. M. Meyerson, S. Gabriel, G. Getz, Advances in understanding cancer genomes through second-generation sequencing. *Nat. Rev. Genet.* **11**, 685–696 (2010).
42. J. Otsuka, T. Okuda, A. Sekizawa, S. Amemiya, H. Saito, T. Okai, M. Kushima, Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. *Int. J. Gynecol. Cancer* **14**, 459–464 (2004).

Acknowledgments: We thank H. Biggs, C. Hodgkin, S. Richardson, and L. Jones for assistance in sample collection, S. Aldridge for assistance in genomic analysis, and S. Tavaré, B. Davis, and M. Dunning for assistance in data analysis. **Funding:** We acknowledge the support of Cancer Research UK, the University of Cambridge, National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge Experimental Cancer Medicine Centre, and Hutchison Whampoa Limited. C.P. was supported in part by the Academy of Medical Sciences, Wellcome Trust, British Heart Foundation, and Arthritis Research UK. **Author contributions:** T.F., M.M., C.P., D.G., D.W.Y.T., C.C., J.D.B., and N.R. designed the study. T.F., M.M., D.W.Y.T., F.K., J.H., A.P.M.,

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

and N.R. developed methods. T.F., D.G., D.W.Y.T., A.M.P., and S.-J.D. collected the data. T.F., M.M., and N.R. analyzed TAm-Seq data. C.P., S.-J.D., C.C., and J.D.B. designed clinical studies and collected samples and clinical data. M.J.-L. performed pathological analysis. D.B. contributed sequencing data. T.F., M.M., C.P., D.G., D.W.Y.T., J.H., A.P.M., J.D.B., and N.R. interpreted the data. T.F., M.M., and N.R. wrote the manuscript with assistance from C.P., D.G., D.W.Y.T., A.P.M., J.D.B., and other authors. All authors approved the final manuscript.

Competing interests: A.P.M. and F.K. hold equity in Fluidigm and may stand to gain by publication of these findings. D.B. and F.K. hold equity in Illumina and may stand to gain by publication of these findings.

Submitted 24 October 2011
Accepted 18 April 2012
Published 30 May 2012
10.1126/scitranslmed.3003726

Citation: T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68 (2012).

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Science Translational Medicine

Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA

Tim Forshew, Muhammed Murtaza, Christine Parkinson, Davina Gale, Dana W. Y. Tsui, Fiona Kaper, Sarah-Jane Dawson, Anna M. Piskorz, Mercedes Jimenez-Linan, David Bentley, James Hadfield, Andrew P. May, Carlos Caldas, James D. Brenton and Nitzan Rosenfeld

Sci Transl Med **4**, 136ra68136ra68.
DOI: 10.1126/scitranslmed.3003726

Deep Sequencing Tumor DNA in Plasma

Five liters of circulating blood contain millions of copies of the genome, broken into short fragments; in cancer patients, a small fraction is circulating tumor DNA (ctDNA). An even smaller number harbor mutations that affect cancer outcome. Looking for diagnostic answers in circulating DNA is a challenge, but Forshew, Murtaza, and colleagues have risen to the occasion by developing a tagged-amplicon deep sequencing (TAm-Seq) method that can amplify and sequence large genomic regions from even single copies of ctDNA. By sequencing such large regions, the authors were able to identify low-level mutations in the plasma of patients with high-grade serous ovarian carcinomas.

Forshew *et al*. designed primers to amplify 5995 bases that covered select regions of cancer-related genes, including *TP53*, *EGFR*, *BRAF*, and *KRAS*. In plasma obtained from 38 patients with high levels of ctDNA, the authors were able to identify mutations in *TP53* at allelic frequencies of 2% to 65%. In plasma samples from one patient, they also identified a de novo mutation in *EGFR* that had not been detected 15 months prior in the tumor mass itself. Finally, the TAm-Seq approach was used to sequence ctDNA in plasma samples collected from two women with ovarian cancer and one woman with breast cancer at different time points, tracking as many as 10 mutations in parallel. Forshew and coauthors showed that levels of mutant alleles reflected the clinical course of the disease and its treatment––for example, stabilized disease was associated with low allelic frequency, whereas patients at relapse exhibited a rise in frequency.

Through several experiments, the authors were able to show that TAm-Seq is a viable method for sequencing large regions of ctDNA. Although this provides a new way to noninvasively identify gene mutations in our blood, TAm-Seq will need to achieve a more sensitive detection limit (<2% allele frequency) to identify mutations in the plasma of patients with less advanced cancers. Nevertheless, once optimized, this ''liquid biopsy'' approach will be amenable to personalized genomics, where the level and type of mutations in ctDNA would inform clinical decision-making on an individual basis.

ARTICLE TOOLS	http://stm.sciencemag.org/content/4/136/136ra68
SUPPLEMENTARY MATERIALS	http://stm.sciencemag.org/content/suppl/2012/05/25/4.136.136ra68.DC1

Use of this article is subject to the Terms of Service

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Translational Medicine* is a registered trademark of AAAS.

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

RELATED CONTENT
http://science.sciencemag.org/content/sci/339/6119/580.full
http://stm.sciencemag.org/content/scitransmed/5/207/207ps14.full
http://stm.sciencemag.org/content/scitransmed/6/224/224ra24.full
http://stm.sciencemag.org/content/scitransmed/6/224/224ra26.full
http://stm.sciencemag.org/content/scitransmed/6/241/241ra77.full
http://stm.sciencemag.org/content/scitransmed/6/243/243fs26.full
http://stm.sciencemag.org/content/scitransmed/7/283/283ra53.full
http://stm.sciencemag.org/content/scitransmed/7/293/293ra104.full
http://stm.sciencemag.org/content/scitransmed/7/302/302ra133.full
http://science.sciencemag.org/content/sci/339/6127/1546.full
http://science.sciencemag.org/content/sci/339/6127/1559.full
http://science.sciencemag.org/content/sci/339/6127/1563.full
http://science.sciencemag.org/content/sci/339/6127/1539.full
http://science.sciencemag.org/content/sci/339/6127/1543.full
http://science.sciencemag.org/content/sci/339/6127/1567.full
http://science.sciencemag.org/content/sci/339/6127/1540.full
http://science.sciencemag.org/content/sci/339/6127/1493.full

REFERENCES
This article cites 42 articles, 16 of which you can access for free
http://stm.sciencemag.org/content/4/136/136ra68#BIBL

PERMISSIONS
http://www.sciencemag.org/help/reprints-and-permissions

Downloaded from http://stm.sciencemag.org/ by guest on August 15, 2018

Use of this article is subject to the Terms of Service

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Translational Medicine* is a registered trademark of AAAS.

LETTER

doi:10.1038/nature12065

Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA

Muhammed Murtaza[1]*, Sarah-Jane Dawson[1,2]*, Dana W. Y. Tsui[1]*, Davina Gale[1], Tim Forshew[1], Anna M. Piskorz[1], Christine Parkinson[1,2], Suet-Feung Chin[1], Zoya Kingsbury[3], Alvin S. C. Wong[4], Francesco Marass[1], Sean Humphray[3], James Hadfield[1], David Bentley[3], Tan Min Chin[4,5], James D. Brenton[1,2,6], Carlos Caldas[1,2,6] & Nitzan Rosenfeld[1]

Cancers acquire resistance to systemic treatment as a result of clonal evolution and selection[1,2]. Repeat biopsies to study genomic evolution as a result of therapy are difficult, invasive and may be confounded by intra-tumour heterogeneity[3,4]. Recent studies have shown that genomic alterations in solid cancers can be characterized by massively parallel sequencing of circulating cell-free tumour DNA released from cancer cells into plasma, representing a non-invasive liquid biopsy[5–7]. Here we report sequencing of cancer exomes in serial plasma samples to track genomic evolution of metastatic cancers in response to therapy. Six patients with advanced breast, ovarian and lung cancers were followed over 1–2 years. For each case, exome sequencing was performed on 2–5 plasma samples (19 in total) spanning multiple courses of treatment, at selected time points when the allele fraction of tumour mutations in plasma was high, allowing improved sensitivity. For two cases, synchronous biopsies were also analysed, confirming genome-wide representation of the tumour genome in plasma. Quantification of allele fractions in plasma identified increased representation of mutant alleles in association with emergence of therapy resistance. These included an activating mutation in *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha) following treatment with paclitaxel[8]; a truncating mutation in *RB1* (retinoblastoma 1) following treatment with cisplatin[9]; a truncating mutation in *MED1* (mediator complex subunit 1) following treatment with tamoxifen and trastuzumab[10,11], and following subsequent treatment with lapatinib[12,13], a splicing mutation in *GAS6* (growth arrest-specific 6) in the same patient; and a resistance-conferring mutation in *EGFR* (epidermal growth factor receptor; T790M) following treatment with gefitinib[14]. These results establish proof of principle that exome-wide analysis of circulating tumour DNA could complement current invasive biopsy approaches to identify mutations associated with acquired drug resistance in advanced cancers. Serial analysis of cancer genomes in plasma constitutes a new paradigm for the study of clonal evolution in human cancers.

Serial sampling of the tumour genome is required to identify the mutational mechanisms underlying drug resistance[2]. Serial tumour biopsies are invasive and often unattainable. Tumours are heterogeneous and continuously evolve, and even if several biopsies are obtained, these are limited both spatially and temporally. Analysis of isolated circulating tumour cells (CTCs) has been proposed, but circulating tumour DNA (ctDNA) is more accessible and easier to process[15]. Previous studies of tumour mutations in plasma have analysed individual loci, genes or structural variants to quantify tumour burden and to detect previously-characterized resistance-conferring mutations[1,6,16–18]. Genome-wide sequencing of plasma samples is used in prenatal diagnostics, demonstrating comprehensive coverage of the genome[19]. More recently, genome-wide sequencing of plasma DNA has been demonstrated as a potential tool for detection of disease or analysis of tumour burden in patients with advanced cancers[5,7]. These studies established that plasma DNA contains representation of the entire tumour genome[7], mixing together variants originating from multiple independent tumours[5]. This suggests that deeper sequencing of plasma DNA, applied to selected samples with high tumour burden in blood, may allow assessment of clonal heterogeneity and selection. In this study, we applied exome sequencing of ctDNA as a platform for non-invasive analysis of tumour evolution during systemic cancer treatment (Fig. 1).

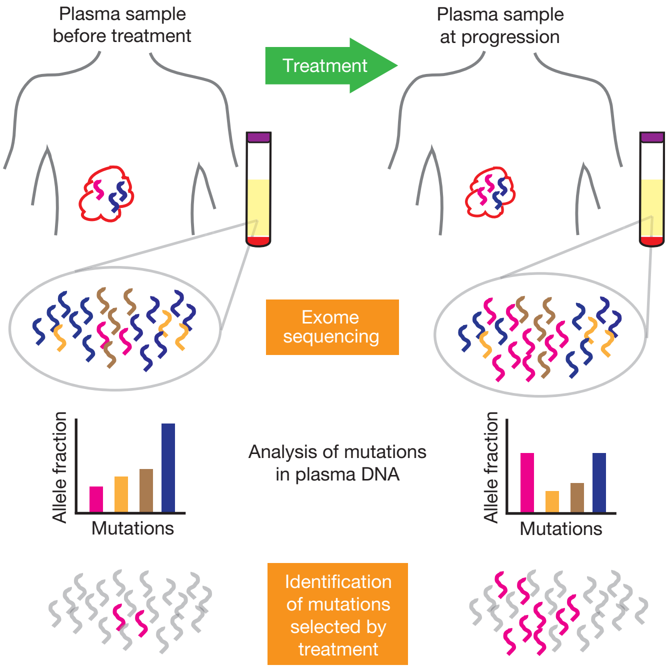

Plasma sample before treatment
Treatment
Plasma sample at progression
Exome sequencing
Allele fraction
Mutations
Analysis of mutations in plasma DNA
Allele fraction
Mutations
Identification of mutations selected by treatment

Figure 1 | Identification of treatment-associated mutational changes from exome sequencing of serial plasma samples. Overview of the study design: plasma was collected before treatment and at multiple time-points during treatment and follow-up of advanced cancer patients. Exome sequencing was performed on circulating DNA from plasma at selected time-points, separated by periods of treatment, and germline DNA. Mutations were identified across the plasma samples, and their abundance (allele fraction) at different time-points compared, generating lists of mutations that showed a significant increase in abundance, which may indicate underlying selection pressures associated with specific treatments. These lists contained mutations known to promote tumour growth and drug resistance, but also mutations of unknown significance. Accumulating such data across large cohorts could identify genes or pathways with recurrent mutations.

---

[1] Cancer Research UK Cambridge Institute and University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. [2] Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. [3] Illumina, Inc., Chesterford Research Park, Little Chesterford CB10 1XL, UK. [4] Department of Haematology-Oncology, National University Cancer Institute, National University Health System, 5 Lower Kent Ridge Road, Tower block level 7, 119074 Singapore. [5] Cancer Science Institute, National University of Singapore, Centre for Translational Medicine, 14 Medical Drive, #12-01, 117599 Singapore. [6] Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.
*These authors contributed equally to this work.

©2013 Macmillan Publishers Limited. All rights reserved

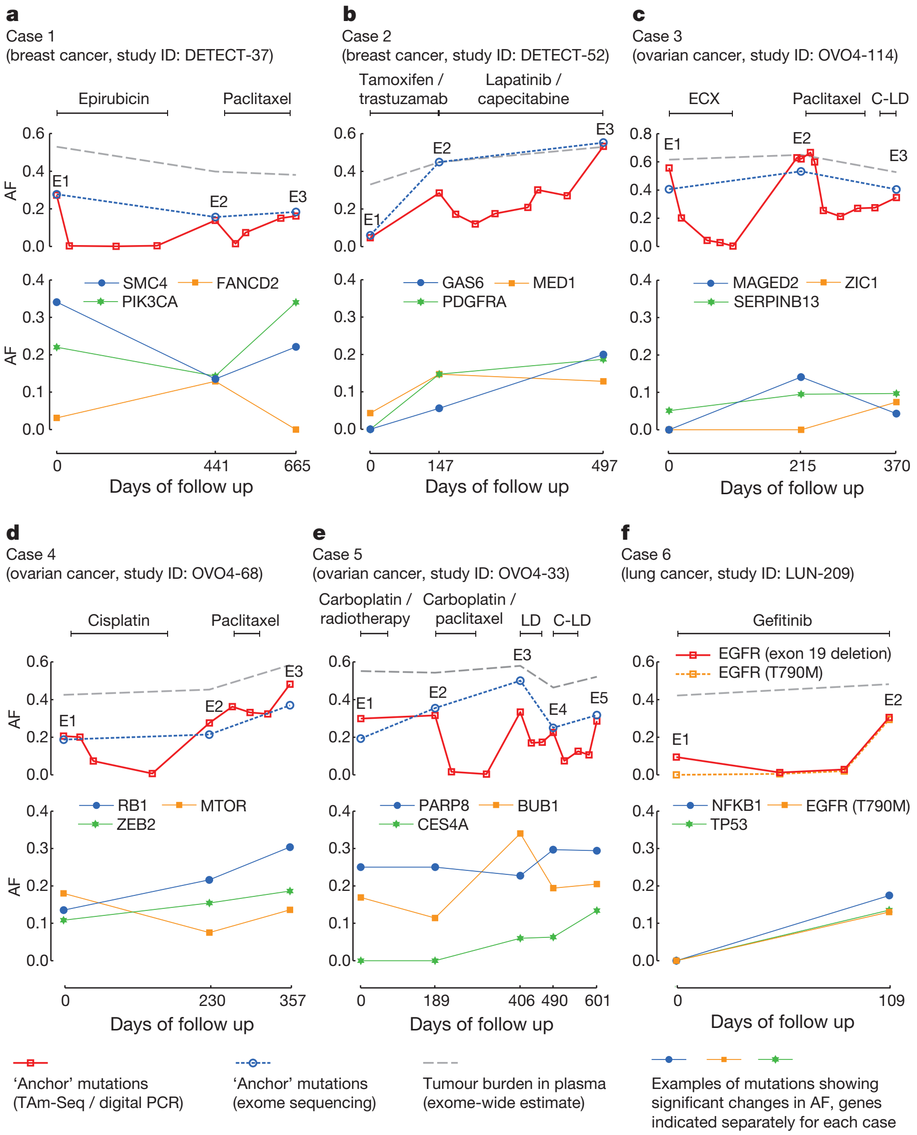


Figure 2 | Mutations showing evidence of genomic tumour evolution. All panels (**a**–**f**) are made up of an upper and a lower subpanel. Upper subpanels, time courses for allele fractions (AF; data points) of 'anchor' mutations used for initial quantification of ctDNA levels, and the fractional concentration of tumour DNA (tumour burden; grey dashed lines). 'Anchor' mutations were measured using digital PCR or TAm-Seq[6] for all available plasma samples, and using exome sequencing at selected time points indicated by E1, E2, E3 (and E4 and E5 for case 5). Tumour burden was estimated from exome data (an adaptation of genome-wide aggregated allelic loss[7]). In **a**, AF was averaged over six mutations measured in parallel using digital PCR. In **b**, a single mutation in *ATM* (predicted amino acid change I2948F) was measured by TAm-Seq. In **c**, **d** and **e**, a single mutation in *TP53* was measured by digital PCR for each case (R175H, K132N and R175H, respectively). In **f**, digital PCR was used to measure abundance of a deletion in exon 19 of *EGFR* (not quantified in exome sequencing data) and the EGFR T790M mutation. Lower subpanels, AF in exome data for selected mutations (blue, green and orange datapoints, see key) for each of the cases. Additional details are listed in Table 1, and a full list of mutations that showed a significant increase in abundance is included in Supplementary Tables 2–7. ECX, epirubicin, cisplatin and capecitabine; C-LD, carboplatin and liposomal doxorubicin; LD, liposomal doxorubicin.

We performed whole exome sequencing of plasma DNA in six patients with advanced cancers (Supplementary Table 1): two with breast cancer (cases 1 and 2), three with ovarian cancer (cases 3–5), and one with non-small-cell lung cancer (NSCLC, case 6). Exome sequencing was performed on multiple plasma samples from each patient separated by consecutive lines of therapy, spanning up to 665 days of clinical follow up (range 109–665 days, median 433 days). The ability to detect genomic events using redundant sequencing is dependent on the allele fraction (AF) of the mutant alleles in the samples analysed (ratio of mutant reads to depth of coverage at that locus), the sequencing depth, and the background noise rates of sequencing. Levels of ctDNA were previously quantified in these patients using digital PCR and tagged-amplicon deep sequencing[6] (TAm-Seq; Fig. 2, upper subpanels), allowing us to focus on samples with a high mutant AF in plasma, in which genomic changes related to the tumour could be identified even at relatively modest depth of sequencing. Comparison of AF measured using exome sequencing, digital PCR and TAm-Seq showed a high degree of concordance (correlation coefficient 0.8, $P < 0.0001$; Supplementary Fig. 1). Using as little as 2.3 ng of DNA (4%–20% of the DNA extracted from 2.0–2.2 ml of plasma), and an average of 169 million reads of sequencing per sample, we analysed the coding exons of all protein-coding genes at an average unique coverage depth ranging from 31-fold to 160-fold across 19 plasma samples (Supplementary Table 2). Consistent with previous reports[5,7], we observed copy number aberrations (CNAs, both gains and losses) in plasma samples in all patients across the whole genome (Supplementary Figs 2–7). These were strongly modulated by the fraction of tumour DNA in plasma and were particularly prominent in plasma samples in which mutant AF exceeded 50%.

©2013 Macmillan Publishers Limited. All rights reserved

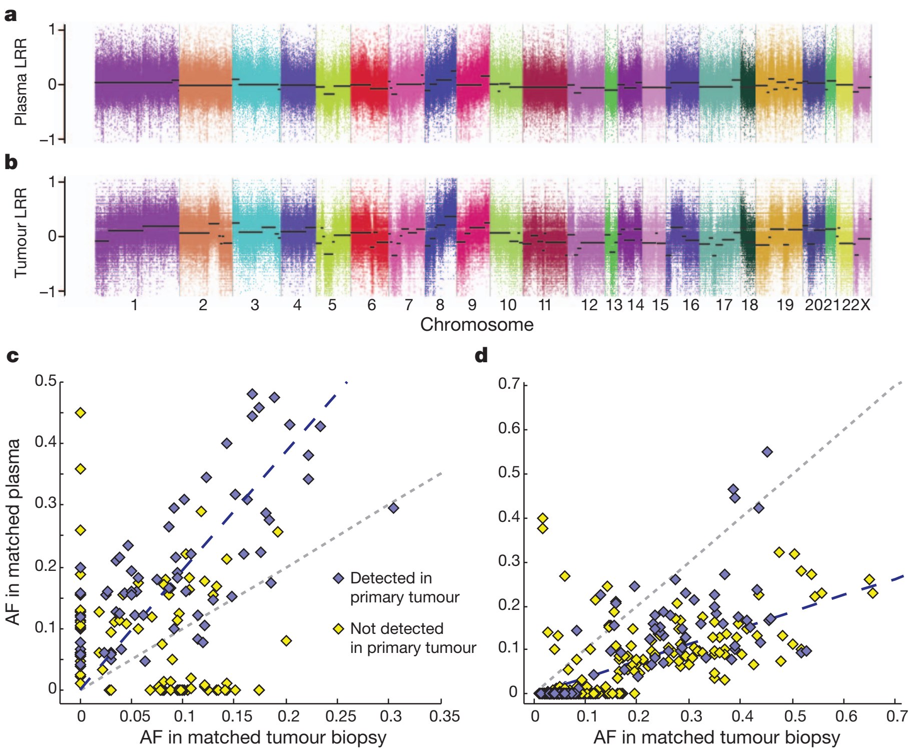


Figure 3 | Genome-wide concordance between plasma DNA and tumour DNA. a, **b**, Sequencing data were used to assess CNAs in the plasma sample (**a**) and in the synchronous metastatic tumour biopsy (**b**) from case 4. Panels show log *R* ratio (LRR), calculated on the basis of exome data, between plasma DNA and normal DNA (**a**) and between tumour and normal DNA (**b**). **c**, AF of mutations identified in exome data from plasma or metastatic biopsy for case 1. Grey dotted line shows equality. Blue dashed line has a slope of 1.93, indicating the median of the AF ratio for mutations found in both samples. Key applies to **c** and **d**. **d**, As **c** but for case 4, blue dashed line has a slope of 0.37.

For two cases, sequencing data were also available from metastatic tumour biopsies, collected at the same time as plasma samples (case 1 sample E1, and case 4 sample E2), and from tumour samples collected at the patients' initial presentation, 9 and 4.5 years earlier. CNAs were concordant between plasma and metastasis DNA in both patients (Fig. 3a, b, and Supplementary Fig. 7). Mutations identified in sequencing data[20–23] from the plasma or metastatic biopsy were compared (Supplementary Information). In case 1 with breast cancer, 151 mutations were identified in either the plasma or the synchronous biopsy. Of these, 93 mutations were found in both, and mutant AFs for these were higher in the plasma sample compared to the metastatic biopsy. The correlation coefficient of mutant AFs was positive (0.71) for mutations that were also found in the primary tumour, but negative (−0.22) for other mutations (Fig. 3c). In case 4 with ovarian cancer, 895 mutations were identified in either plasma or the tumour biopsy. For 172 mutations found in both, AFs were positively correlated (0.72) and were higher in the metastatic biopsy, which also contained 686 'private' mutations with AF < 0.2 that were not found in either the plasma or the earlier tumour sample (Fig. 3d).

To identify changes in the mutation profiles of the tumours, we compared the abundance of somatic mutations found in plasma before and after each course of systemic treatment. For each patient, we examined a conservative list of mutations, including all mutations that were called in any of the plasma samples with a Bonferroni-corrected binomial probability of <0.05 assuming a background sequencing error rate of 0.1%. For each mutation and course of treatment (spanned by a pair of plasma samples), a *P*-value for a possible change in mutant AF was calculated as the binomial probability of obtaining the observed number of mutant reads, given the sequencing depth and the observed abundance in the paired time-point, normalized by the fractional concentration of tumour-derived DNA in the plasma (based on genome-wide aggregated allelic loss[5], Supplementary Table 3). Overall, 364 non-synonymous mutations passed with false discovery rate of $<10\%$ for significant changes in normalized abundance, ranging from 15 to 121 for each case (median 49). These include mutations in well-known cancer genes, genes linked to drug resistance and drug metabolism, and genes not previously associated with carcinogenesis or therapy resistance (Supplementary Tables 4–9). Selected examples are shown in Table 1 and Fig. 2.

We highlight here five examples. In case 1 with breast cancer, a strong increase was observed in the abundance of an activating mutation in *PIK3CA* following treatment with paclitaxel (Fig. 2a and Table 1). This mutation has been shown to promote resistance to paclitaxel in mammary epithelial cells[8]. In case 2, a patient with an oestrogen-receptor (ER)-positive, HER2-positive breast cancer, treatment with tamoxifen in combination with trastuzumab led to an increase in abundance of a nonsense mutation near the carboxy terminus of *MED1*, an ER co-activator that has been shown to be involved in tamoxifen resistance[10,11]. After further treatment of this patient with lapatinib in combination with capecitabine, we observed an increase in abundance of a splicing mutation in *GAS6*, the ligand for the tyrosine kinase receptor AXL (Fig. 2b, Table 1). Activation of the AXL kinase pathway has been shown to cause resistance to tyrosine kinase inhibitors in NSCLC[13] and resistance to lapatinib in ER-positive, HER2-positive breast cancer cell lines[12]. In case 4 with ovarian cancer, following treatment with cisplatin, we observed increase in abundance of a truncating mutation in the tumour-suppressor *RB1* (Fig. 2d, Table 1), predicted to inactivate the RB1 protein (Supplementary Fig. 8). In the matched metastasis biopsy obtained after treatment, the mutation was found in 95% of sequencing reads (59 of 62), with apparent loss of heterozygosity at 13q containing the *RB1* gene (Fig. 3a, b). Loss of *RB1* has been linked with chemotherapy response[9]. Case 6 was a NSCLC patient with an activating mutation in *EGFR* who was treated with gefitinib but progressed on treatment. Analysis by digital PCR detected the EGFR T790M mutation in plasma at progression, but not at the start of treatment. This mutation inhibits binding of

©2013 Macmillan Publishers Limited. All rights reserved

Table 1 | Selected mutations whose mutant AF significantly increased following treatment

Patient	Cancer type	Gene	Effect	Potential biological interest	Associated treatment	Mutant AF in plasma	
						Before	After
Case 1	Breast	*PIK3CA*	E545K	PI-3-kinase. p.E545K mutation associated with chemoresistance in mammary epithelial cells[8].	Paclitaxel	14%	34%
Case 1	Breast	*BMI1*	S324Y	BMI1 polycomb ring finger oncogene. Associated with chemoresistance[25].	Paclitaxel	3%	12%
Case 1	Breast	*SMC4*	I1000S	Structural maintenance of chromosomes 4. Downregulated in taxane resistant cell lines[26].	Paclitaxel	14%	22%
Case 1	Breast	*FANCD2*	G56V	Fanconi anaemia complementation group D2. Chromatin dynamics and DNA crosslink repair[27].	Epirubicin	3%	13%
Case 2	Breast	*MED1*	S1179X	Mediator complex subunit 1. Co-activator of ER with functional role in tamoxifen resistance[10,11].	Tamoxifen/trastuzumab	4%	15%
Case 2	Breast	*ATM*	I2948F	Ataxia telangiectasia mutated.	Tamoxifen/trastuzumab	6%	45%
Case 2	Breast	*PDGFRA*	D714E	Platelet-derived growth factor alpha. Cell surface tyrosine kinase receptor.	Tamoxifen/trastuzumab	0%	15%
Case 2	Breast	*GAS6*	Splicing	Growth arrest-specific 6. Ligand for AXL, overexpression associated with TKI resistance[12,13].	Lapatinib/capecitabine	6%	30%
Case 2	Breast	*TP63*	Splicing / S551G	Tumour protein p63.	Lapatinib/capecitabine	4%	20%
Case 4	Ovarian	*RB1*	E580X	Retinoblastoma 1. Loss of RB1 associated with EMT and drug resistance[9].	Cisplatin	14%	22%
Case 4	Ovarian	*ZEB2*	Y663C	Zinc finger E-box binding homeobox 2. Overexpression associated with cisplatin resistance in ovarian cancer[28].	Cisplatin	11%	15%
Case 4	Ovarian	*MTOR*	K1655N	Mechanistic target of rapamycin. Activating mutations in mTOR confers resistance to antimicrotubule agents[29].	Paclitaxel	8%	14%
Case 5	Ovarian	*CES4A*	P55S	Carboxylesterase 4A. Hydrolysis or transesterification of various xenobiotics.	Carboplatin/paclitaxel Carboplatin/liposomal doxorubicin	0% 6%	6% 13%
Case 5	Ovarian	*BUB1*	M889K	Mitotic checkpoint serine/threonine-protein kinase.	Carboplatin/paclitaxel	11%	34%
Case 5	Ovarian	*PARP8*	P81T	Poly [ADP-ribose] polymerase family, member 8.	Liposomal doxorubicin	23%	30%
Case 6	Lung	*EGFR*	T790M	Epidermal growth factor receptor. Established to cause gefitinib resistance by inhibiting drug binding[14].	Gefitinib	0%	13%
Case 6	Lung	*TP53*	Y163C	Tumour protein p53[30].	Gefitinib	0%	14%
Case 6	Lung	*NFKB1*	G489V	Nuclear factor κB[30].	Gefitinib	0%	17%

Potential biological role and associations with drug resistance described in literature are highlighted. The "Effect" column lists predicted change in amino acid sequence.

gefitinib to EGFR and has been established as the main driver of acquired resistance to gefitinib[14]. Unbiased analysis of plasma DNA by exome sequencing identified selection for this mutation amongst genomic changes that occurred following therapy (Fig. 2f, Table 1).

In this proof of principle study, we demonstrate that exome analysis of plasma ctDNA represents a novel paradigm for non-invasive characterization of tumour evolution. Our data, together with recent reports[5,7], show that CNAs and somatic mutations identified in ctDNA are widely representative of the tumour genome and provide an alternative method of tumour sampling that can overcome limitations of repeated biopsies. Cell-free DNA fragments from multiple lesions in the same individual all mix together in the peripheral blood[5], therefore ctDNA is likely to contain a wider representation of the genomes from multiple metastatic sites, whereas mutations present in a single biopsy or minor sub-clone may be missed. This strengthens the case for the use of ctDNA as a biomarker for monitoring tumour burden or for the analysis of hotspot mutation regions[1,6,16,17], but also indicates that tracking different mutations for assessment of tumour heterogeneity and clonal evolution is now possible. Our data identified a subset of genes that were positively selected following treatment, many of which have been previously associated with drug resistance. Other changes may represent 'passenger' mutations or false-positives, but some are likely to contribute to resistance to therapy. Accumulating data across a large number of cases could identify new genes or pathways that are frequently mutated following specific treatment types, and help refine analysis algorithms.

The approach we describe here may be broadly applicable to a large fraction of advanced cancers, where the median mutation burden in plasma (before start of treatment) is 5%–10% (refs 6, 16, 24). Analysis of acquired drug resistance is of particular utility in advanced or metastatic cancers, which is the target population for nearly all early phase clinical trials. Improvements in sequencing and associated technologies may enable similar analysis in cases with a lower tumour burden in plasma. At present, this non-invasive approach for characterizing cancer exomes in plasma is readily applicable to patients with high systemic tumour burden, enabling detailed and comprehensive evaluation of clonal genomic evolution associated with treatment response and resistance.

METHODS SUMMARY

Patients and samples. Cases 1–5 were recruited as part of prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference nos 07/Q0106/63, 08/H0306/61 and 07/Q0106/63). Case 6 was recruited as part of the 'Hydroxychloroquine and gefitinib to treat lung cancer' study (NCT00809237) at the National University Health System, Singapore, approved by the National Healthcare Group NHG IRB—DSRB 2008/00196. Written informed consent was obtained from patients, and serial blood samples were collected at intervals of ≥3 weeks.

Extraction and sequencing of plasma DNA. DNA was extracted from plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions. Barcoded sequencing libraries were prepared using a commercially available kit (ThruPLEX-FD, Rubicon Genomics). Pooled libraries were enriched for the exome using hybridization (TruSeq Exome Enrichment Kit, Illumina), quantified using quantitative PCR and pooled in 1:1 ratio for paired-end sequencing on a HiSeq2500 (Illumina).

Variant calling and analysis. Sequencing data were demultiplexed and aligned to the hg19 genome using BWA[20]. Pileup files for properly paired reads with mapping quality ≥60 were generated using samtools[22]. AFs were calculated for all Q30 bases. A mutation was called if ≥4 mutant reads were found in plasma with ≥1 read on each strand, and no mutant reads were observed in germline DNA or in a prior plasma sample with ≥10-fold coverage. For comparison between consecutive plasma samples in a patient, we calculated the binomial probability of obtaining the observed AF (or greater) if the abundance of the mutant allele, normalized by tumour load in plasma (based on a modified genome-wide aggregated allelic loss method[5]), had remained constant between the two samples.

Full Methods and any associated references are available in the online version of the paper.

Received 5 October 2012; accepted 11 March 2013.

Published online 7 April 2013.

1. Diaz, L. A. Jr *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486,** 537–540 (2012).

©2013 Macmillan Publishers Limited. All rights reserved

2. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368,** 842–851 (2013).
3. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
4. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–399 (2012).
5. Chan, K. C. *et al.* Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin. Chem.* **59,** 211–224 (2013).
6. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4,** 136ra168 (2012).
7. Leary, R. J. *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci. Transl. Med.* **4,** 162ra154 (2012).
8. Isakoff, S. J. *et al.* Breast cancer-associated PIK3CA mutations are oncogenic in mammary epithelial cells. *Cancer Res.* **65,** 10992–11000 (2005).
9. Knudsen, E. S. & Knudsen, K. E. Tailoring to RB: tumour suppressor status and therapeutic response. *Nature Rev. Cancer* **8,** 714–724 (2008).
10. Cui, J. *et al.* Cross-talk between HER2 and MED1 regulates tamoxifen resistance of human breast cancer cells. *Cancer Res.* **72,** 5625–5634 (2012).
11. Nagalingam, A. *et al.* Med1 plays a critical role in the development of tamoxifen resistance. *Carcinogenesis* **33,** 918–930 (2012).
12. Liu, L. *et al.* Novel mechanism of lapatinib resistance in HER2-positive breast tumor cells: activation of AXL. *Cancer Res.* **69,** 6871–6878 (2009).
13. Zhang, Z. *et al.* Activation of the AXL kinase causes resistance to EGFR-targeted therapy in lung cancer. *Nature Genet.* **44,** 852–860 (2012).
14. Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2,** e73 (2005).
15. Punnoose, E. A. *et al.* Evaluation of circulating tumor cells and circulating tumor DNA in non-small cell lung cancer: association with clinical endpoints in a phase II clinical trial of pertuzumab and erlotinib. *Clin. Cancer Res.* **18,** 2391–2401 (2012).
16. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nature Med.* **14,** 985–990 (2008).
17. McBride, D. J. *et al.* Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosom. Cancer* **49,** 1062–1069 (2010).
18. Yung, T. K. *et al.* Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin. Cancer Res.* **15,** 2076–2084 (2009).
19. Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2,** 61ra91 (2010).
20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
21. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43,** 491–498 (2011).
22. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
23. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).
24. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl Acad. Sci. USA* **102,** 16368–16373 (2005).
25. Siddique, H. R. & Saleem, M. Role of BMI1, a stem cell factor, in cancer recurrence and chemoresistance: preclinical and clinical evidences. *Stem Cells* **30,** 372–378 (2012).
26. Chang, H. *et al.* Identification of genes associated with chemosensitivity to SAHA/taxane combination treatment in taxane-resistant breast cancer cells. *Breast Cancer Res. Treat.* **125,** 55–63 (2011).
27. Sato, K. *et al.* Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J.* **31,** 3524–3536 (2012).
28. Haslehurst, A. M. *et al.* EMT transcription factors snail and slug directly contribute to cisplatin resistance in ovarian cancer. *BMC Cancer* **12,** 91 (2012).
29. VanderWeele, D. J., Zhou, R. & Rudin, C. M. Akt up-regulation increases resistance to microtubule-directed chemotherapeutic agents through mammalian target of rapamycin. *Mol. Cancer Ther.* **3,** 1605–1613 (2004).
30. Wu, C. C., Yu, C. T., Chang, G. C., Lai, J. M. & Hsu, S. L. Aurora-A promotes gefitinib resistance via a NF-κB signaling pathway in p53 knockdown lung cancer cells. *Biochem. Biophys. Res. Commun.* **405,** 168–172 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Langmore and K. Solomon (Rubicon Genomics) for early access to library preparation products. We thank L. Jones, S. Richardson, C. Hodgkin and H. Biggs for recruiting patients into the DETECT and CTCR-OV04 studies, all medical and ancillary staff in the breast and gynaecological cancer clinic and patients for consenting to participate. We thank the Human Research Tissue Bank at Addenbrooke's Hospital which is supported by the NIHR Cambridge Biomedical Research Centre. We thank the Cancer Science Institute, National University of Singapore, and the Hematology-Oncology Research Group, National University Health System, Singapore for their support. We acknowledge the support of Cancer Research UK, the University of Cambridge, National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge Experimental Cancer Medicine Centre, Hutchison Whampoa Limited, and the National Medical Research Council, Singapore. S.-J.D. is supported by an Australian NHMRC/RG Menzies Early Career Fellowship that is administered through the Peter MacCallum Cancer Centre, Victoria, Australia.

Author Contributions M.M., S.-J.D., T.F., D.W.Y.T., D.G., J.D.B., C.C. and N.R. designed the study. M.M., D.W.Y.T. and T.F. developed methods. S.-J.D., C.P., A.S.C.W., T.M.C., J.D.B. and C.C. designed and conducted the prospective clinical studies. M.M., S.-J.D., D.W.Y.T., D.G., T.F. and A.M.P. generated data. Z.K., S.H. and D.B. contributed sequencing data. M.M., F.M. and N.R. analysed sequencing data. S.-F.C. and J.H. contributed to experiments and data analysis. M.M., S.-J.D., D.W.Y.T., T.M.C., J.D.B., C.C. and N.R. interpreted data. M.M. and N.R. wrote the paper with assistance from S.-J.D., D.W.Y.T., C.C., J.D.B. and other authors. All authors approved the final manuscript. J.D.B., C.C. and N.R. are the project co-leaders and joint senior authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.B. (james.brenton@cruk.cam.ac.uk), C.C. (carlos.caldas@cruk.cam.ac.uk) or N.R. (nitzan.rosenfeld@cruk.cam.ac.uk).

©2013 **Macmillan Publishers Limited. All rights reserved**

METHODS

Sample collection. Cases 1–5: patients were recruited as part of prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by local research ethics committee (REC reference nos 07/Q0106/63, 08/H0306/61 and 07/Q0106/63). Written informed consent was obtained from the patients. Serial blood samples were collected in EDTA tubes at intervals of ≥3 weeks, and centrifuged within 1 h at 820*g* for 10 min to separate the plasma from the peripheral blood cells. The plasma was then further centrifuged at 20,000*g* for 10 min to pellet any remaining cells. The plasma was then stored at −80°C until DNA extraction.

Case 6: this patient was recruited as part of the 'Hydroxychloroquine and gefitinib to treat lung cancer' study (NCT00809237) at the National University Health System, Singapore, approved by the National Healthcare Group NHG IRB-DSRB 2008/00196. Blood was collected in CPT tubes (BD Vacutainer) before gefitinib was started, and at monthly intervals while the patient was on treatment, until disease progression. Blood collected was spun within 1 h at 1,500*g* for 20 min, and the plasma fraction was frozen at −80 °C. Thawed samples were recentrifuged at 20,000*g* for 10 min to further separate any cellular portions.

Extraction of plasma DNA. DNA was extracted from aliquots of plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions (see Supplementary Table 1 for volumes used). DNA was eluted into buffer AVE, eluted twice through each column to maximize yield, and stored at −20 °C.

Extraction of normal and tumour DNA. DNA from tumour sections was extracted using DNeasy tissue or DNA Allprep kits (Qiagen) according to manufacturer's instructions. Matched germline DNA was derived from normal peripheral blood leucocytes. After the collection of plasma from each blood sample, the remaining layer of normal peripheral blood lymphocytes ('buffy coat') was removed. This layer was either subjected to red cell lysis using a red cell lysis buffer (155 mM NH_4Cl, 10 mM $KHCO_3$ and 0.1 mM EDTA pH 7.4) and DNA extracted using a standard phenol-chloroform extraction protocol; or frozen at −80 °C before extraction using QIAamp DNA mini kit (Qiagen).

Sequencing of plasma DNA. Concentration of DNA for each plasma sample was determined using digital PCR, with an assay targeting *RPP30* for case 2, *TP53* for cases 3–5 and *EGFR* for case 6. For case 1, DNA concentration and 'anchor' mutation AF were calculated by averaging results from six assays targeting *PIK3CA*, *MET*, *IQCA1*, *CD1A*, *KIAA0406* and *ZFYVE21*. Libraries were generated using a commercially available kit for fragmented DNA (ThruPLEX-FD, Rubicon Genomics). 2.3–40 ng of DNA (Supplementary Table 2) was used to generate a sequencing library using manufacturer's protocols. Separate unique molecular identifiers were used for each sample. 30 μl of the library volume was obtained for each sample. 2–5 plasma DNA libraries from each patient were made and pooled together for exome capture using hybridization (TruSeq Exome Enrichment Kit, Illumina). Pools were concentrated using vacuum (Eppendorf Vacuum Concentrator) and prepared to 40 μl volume. Exome enrichment was performed following manufacturer's protocols. Enriched libraries were quantified using quantitative PCR and pooled in 1:1 ratio for paired-end next generation sequencing on HiSeq2500 (Illumina).

Sequencing of normal and tumour DNA. Sequence data for tumour and germline samples for case 1 have been reported previously. In brief, genomic libraries from tumour and matched normal tissue were prepared using the standard Illumina paired-end sample preparation kit according to the manufacturer's instructions. DNA fragments of 300 bp in size were sequenced using paired-end 100 bp reads on a HiSeq2000 (Illumina) achieving a depth of >30×. Germline samples for cases 2–6 and tumour sample for case 4 were sheared using Covaris and exome sequenced as described above.

Digital PCR. The principle of microfluidic digital PCR and its use for quantification of tumour DNA has been described previously[6,18]. Assays were designed based on TaqMan chemistry. All digital PCR analysis was carried out on the BioMark system using 12.765 Digital Arrays (Fluidigm) following manufacturer's instructions and protocol. Briefly, 3.5 μl from the eluted DNA was heated to 95 °C for 1 min and placed on ice, then mixed with TaqMan Universal PCR Master Mix (Applied Biosystems) and sample loading buffer (Fluidigm) into a final reaction volume of 10 μl and loaded into each panel of the chip. The reaction mix was then automatically partitioned into 765 reaction chambers. The numbers of starting template DNA molecules were calculated using Poisson statistics based on the number of positive amplifications[6,18].

Analysis of sequencing data. Sequencing reads were demultiplexed allowing zero mismatches in barcodes. Paired-end alignment to the hg19 genome was performed using BWA version 0.5.9 for all exome sequencing data including germline samples, plasma samples and tumour metastasis where generated[20]. PCR duplicates were marked using Picard. Local realignment was performed using Genome Analysis Tool Kit (GATK)[21]. Pileup files were generated for the genomic regions targeted by exome enrichment using samtools v0.1.17[22]. For plasma samples, properly paired reads with mapping quality ≥60 were used to generate the pileup. AFs for each single-base locus were calculated for all bases with phred quality ≥30. For germline DNA, an additional pileup file was generated (using a mapping quality cut-off of ≥1 and without any base quality cut-offs) and was used as reference for calling somatic variants. A mutation was called if no mutant reads for an allele were observed in germline DNA at a locus that was covered at least 10 fold, and if at least 4 reads supporting the mutant were found in the plasma data with at least 1 read on each strand (forward and reverse). At loci with <10-fold coverage in normal DNA and no mutant reads, mutations were called in plasma if a prior plasma sample showed no evidence of a mutation and was covered adequately (10 fold or more). All mutations were annotated for genes and function as well as repeated genomic regions using ANNOVAR[23].

AF was defined as the number of high quality reads supporting a mutation as a fraction of the total number of high quality reads covering the locus. For each patient, AF and number of reads for any mutations called with the above parameters were identified in all plasma samples. A binomial probability of obtaining the observed number of reads given depth in each plasma sample was calculated. The minimum of these probability values was corrected using Bonferroni correction for 62 million × n hypotheses tested, where n was the number of plasma samples sequenced (3 samples for cases 1–4, 5 samples for case 5 and 2 samples for case 6). Mutations with corrected *P*-values under 0.05 were retained for further analysis in plasma samples.

Estimation of CNAs. To assess CNAs, plasma DNA and tumour sequencing data were compared to germline DNA data at single nucleotide polymorphisms (SNPs) covered within the targeted exome region. The SNPs were identified from the publicly available 1000 Genomes Project data.

Depth information was normalized by dividing the depth of each SNP by the median depth across all SNPs. The log *R* ratio (LRR) was computed as the base-10 logarithm of the sample depth (metastasis or plasma) divided by the depth of the normal. Each chromosome was segmented by an iterative process that considered non-overlapping blocks of 1,000 data points. Points lying at least 1.5 standard deviations away from the median LRR for the block were removed from the mean LRR computation. If the difference in mean LRR between two consecutive blocks was less than 0.12, the blocks were merged into a single segment whose mean LRR was re-computed using points from both blocks.

Segmentation of B allele frequency (BAF) plots was similarly performed, considering windows of 1,000 data points and starting new segments if the difference in median frequency was greater than 4%. Blocks whose median frequency was within 8% of the median chromosome frequency in the normal sample were considered consistent with the BAF of the normal sample.

Comparison of mutations between plasma and tumour. For tumour/plasma comparison presented for cases 1 and 4, we identified all mutations called in data from synchronous plasma and metastatic tumour samples, as described above. We retained all mutations adequately covered in both samples (minimum 50 reads in plasma, minimum 10 reads in synchronous tumour whole genome data for case 1, minimum 50 reads in synchronous tumour exome data for case 4). We further discarded all mutations with no coverage in archived tumour samples obtained earlier (9 years earlier for case 1, and 4.5 years earlier for case 4).

Identification of mutations that changed in representation over treatment. To estimate systemic tumour burden, we calculated fractional concentration of ctDNA in blood using an adaptation of genome-wide aggregated allelic loss[5]. AFs of SNPs from the 1000 Genomes Project were obtained for germline and plasma data. SNPs with $0 < AF < 1$ in germline DNA were identified. SNPs where the minor AF in the germline data deviated from heterozygosity were identified using a binomial probability of obtaining the observed number of minor allele reads given depth in germline DNA and expected AF of 0.5. SNPs with probability <0.25 were discarded from further analysis.

Of the remaining SNPs, significant deviation from heterozygosity in any of the sequenced plasma samples, determined by a binomial distribution using sequencing depth and expected AF of 0.5, was used to identify loss of heterozygosity (LOH). SNPs with a probability <0.01 in any of the sequenced plasma samples were retained for estimation of tumour burden as described previously[5]. Fractional ctDNA burden was calculated as follows:

1 – [(sum of reads in the lost alleles)/(sum of reads in the retained alleles)]

AFs for all mutations were normalized by the estimated tumour burden. For any comparison between two consecutive plasma samples in a patient, we calculated the binomial probability for the observed difference in AF assuming no difference in normalized abundance. For a comparison between (for example) E1 and E2, we calculated the probability of obtaining the observed number of mutant reads or greater in E2 if normalized abundance in E2 had remained the same as in E1; this probability was multiplied by the probability of the observed number of mutant reads or less in E1 if the normalized abundance in E1 was the same as observed in E2. Where no mutant reads were obtained in the E1, only the reverse direction was used for this analysis. Changes in representation with a false discovery rate of 10% or lower, which were exonic non-synonymous or splicing mutations, were retained and are presented in Supplementary Tables 2–7.

©2013 Macmillan Publishers Limited. All rights reserved

nature COMMUNICATIONS

ARTICLE

Received 11 Aug 2015 | Accepted 28 Sep 2015 | Published 4 Nov 2015

DOI: 10.1038/ncomms9760

OPEN

Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer

Muhammed Murtaza[1,2,3,4,*], Sarah-Jane Dawson[1,5,6,*], Katherine Pogrebniak[1,2], Oscar M. Rueda[1,2], Elena Provenzano[5,7], John Grant[7], Suet-Feung Chin[1,2], Dana W.Y. Tsui[1], Francesco Marass[1,2], Davina Gale[1], H. Raza Ali[1,2,5,7], Pankti Shah[3], Tania Contente-Cuomo[3], Hossein Farahani[8], Karey Shumansky[8], Zoya Kingsbury[9], Sean Humphray[9], David Bentley[9], Sohrab P. Shah[8], Matthew Wallis[5,10], Nitzan Rosenfeld[1,2,**] & Carlos Caldas[1,2,5,**]

Circulating tumour DNA analysis can be used to track tumour burden and analyse cancer genomes non-invasively but the extent to which it represents metastatic heterogeneity is unknown. Here we follow a patient with metastatic ER-positive and HER2-positive breast cancer receiving two lines of targeted therapy over 3 years. We characterize genomic architecture and infer clonal evolution in eight tumour biopsies and nine plasma samples collected over 1,193 days of clinical follow-up using exome and targeted amplicon sequencing. Mutation levels in the plasma samples reflect the clonal hierarchy inferred from sequencing of tumour biopsies. Serial changes in circulating levels of sub-clonal private mutations correlate with different treatment responses between metastatic sites. This comparison of biopsy and plasma samples in a single patient with metastatic breast cancer shows that circulating tumour DNA can allow real-time sampling of multifocal clonal evolution.

---

[1] Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB2 0RE, UK. [2] Department of Oncology, University of Cambridge, Cambridge CB2 0RE, UK. [3] Center for Noninvasive Diagnostics, Translational Genomics Research Institute, Phoenix, Arizona 85004, USA. [4] Mayo Clinic Center for Individualized Medicine, Scottsdale, Arizona 85259, USA. [5] Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital National Health Service Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, and the Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0QQ, UK. [6] Peter MacCallum Cancer Centre, East Melbourne, Victoria 3002, Australia. [7] Department of Histopathology, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge CB2 0QQ, UK. [8] BC Cancer Research Centre, Vancouver, British Columbia, Canada V5Z 1L3. [9] Illumina, Inc., Chesterford Research Park, Little Chesterford CB10 1XL, UK. [10] Department of Radiology, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust, Cambridge CB2 0QQ, UK. * These authors contributed equally to this work. ** These authors jointly supervised this work. Correspondence and requests for materials should be addressed to N.R. (email: nitzan.rosenfeld@cruk.cam.ac.uk) or to C.C. (email: carlos.caldas@cruk.cam.ac.uk).

© 2015 Macmillan Publishers Limited. All rights reserved.

Intra-tumour clonal heterogeneity limits efficacy and duration of response to targeted treatments in metastatic cancer[1–3]. Evaluating heterogeneity to guide choice and sequence of therapy could be achieved by multiregional and repeated metastatic tumour biopsies but this is impractical due to associated risk of complications and costs. In contrast, analysis of circulating tumour DNA in plasma (ctDNA) is a less-invasive approach that could provide a summary of somatic alterations contributed by distinct metastases[4,5], potentially circumventing the problem of spatial heterogeneity[1]. Serial analysis of ctDNA has been shown to track tumour burden[6–8] and to correlate with treatment-driven clonal evolution[3,4,9]. Most studies of concordance between tumour and plasma samples have compared individual mutations or relied on single tumour biopsies[9]. However, direct evidence comparing plasma with multiregional tumour samples to establish the extent of clonal heterogeneity captured in ctDNA is extremely limited[5,10–13].

Here we present extensive analysis of eight tumour biopsies and nine plasma samples collected from a patient with oestrogen receptor-positive (ER +) human epidermal growth factor receptor 2-positive (HER2 +) metastatic breast cancer treated with sequential targeted therapies (tamoxifen and trastuzumab, followed by lapatinib) over a 3-year clinical course. We performed whole-exome followed by deep amplicon sequencing to validate and quantify several hundred somatic mutations. We find that ubiquitous stem mutations (common to all tumour biopsies) have the highest circulating levels in plasma followed by metastatic-clade and private mutations. In addition, serial changes during treatment in circulating levels of private somatic mutations correlate with disease progression in their respective tumour lesions on imaging. These results, from a single patient with metastatic breast cancer, suggest that ctDNA reflects clonal tumour hierarchy and captures sub-clonal dynamics in real time.

Results

Clinical case. A 42-year-old woman presented with a right breast lump, lower back pain, loss of height, marked kyphosis and hepatomegaly. Core biopsies from the breast lump showed ductal carcinoma *in situ* (sample labelled P1.1; Supplementary Fig. 1 and Supplementary Table 1). An additional biopsy from an ipsilateral axillary lymph node (P1.2) revealed metastatic ductal adenocarcinoma (ER + (8/8) and HER2 + (3 +)). Computed tomography scan revealed widespread metastatic disease in bones, pleura and liver (Supplementary Fig. 2 and Supplementary Table 2). The patient was started on treatment with trastuzumab and taxane-based chemotherapy, with a significant partial response (Supplementary Fig. 3). After induction chemotherapy, she was maintained on tamoxifen and trastuzumab. After 19 months on treatment, she presented with seizures and head computed tomography revealed a large metastasis in the left frontal lobe (Supplementary Fig. 4), which was resected (M2.1). Therapy with tamoxifen and trastuzumab was continued and collection of plasma samples was initiated (samples T1–T9). Four months after surgery, she had enlarging liver lesions and a new metastatic deposit in the left ovary (Supplementary Fig. 5). Treatment was switched to a combination of lapatinib and capecitabine, resulting in stable disease for 12 months (Supplementary Fig. 6). General deterioration then occurred, with disease progression in the chest (new pulmonary nodules, bilateral pleural effusions and posterior chest wall mass, Supplementary Fig. 7; Eastern Cooperative Oncology Group performance status 2–3). Treatment was stopped and the patient died ~4 months later.

Tumour samples were obtained at diagnosis from the primary breast site (P1.1) and an axillary lymph node (P1.2); after 19 months from the brain metastasis area (M2.1); and at autopsy after 3 years on treatment (from the primary breast site, and from metastatic deposits in the chest, liver, ovary and vertebrae, labelled P3.1 and M3.1–M3.4, respectively). Serial plasma samples were obtained over the last 500 days of clinical follow-up (T1–T9). Tumour and plasma samples collected and the clinical course are summarized in Fig. 1a,b.

Inferring clonal structure from multiregional tumour biopsies. Exome sequencing of peripheral blood leukocytes (N1), 6/8 tumour samples and 3/9 plasma DNA samples (3 plasma exomes reported previously[4]) was performed. Single-nucleotide variants (SNVs) were further analysed by targeted amplicon deep sequencing in all samples for orthogonal validation and accurate measurement of allele fractions (AFs, Supplementary Table 3). Of the 362 candidate non-synonymous SNVs identified by exome sequencing in at least one sample, 310 were successfully tested by deep sequencing (median coverage: 288 × –8,248 × for plasma samples; 965 × –2,777 × for tumour samples). For each candidate SNV, a mutation was called if AF was at least three s.d.'s above the mean background error rate obtained by analysing 12 control samples[11].

Deep sequencing validated 207 functional mutations. We identified 8 major mutation clusters based on variation in their allele fractions across all tumour samples using Bayesian clustering with PyClone (Fig. 1c–e), a data-driven method we have developed and extensively validated for analysing clonal hierarchies and inferring cellular prevalence in tumour biopsies and to follow clonal dynamics in serially transplanted tumour xenografts[14–16]. We also inferred tumour phylogeny using clonal ordering of high-confidence mutations (with $>2\%$ allele fraction in a tumour sample). A total of 23 stem mutations were detected in all tumour samples (tumour cluster 1), 26 metastatic-clade mutations were detected only in metastatic tumour samples (tumour cluster 2) and 126 private mutations were detected at AF $>2\%$ only in one of the tumour samples (tumour clusters 3–8). The most parsimonious pathway of evolution in this cancer together with mutation clustering results is presented in Fig. 1d. Stem and metastatic-clade mutation clusters inferred using PyClone were identical to the results from clonal ordering. Similarly, mutations in clusters 3, 4/5, 6 and 7 correspond to private mutations in P3.1, M3.1, M2.1 and M3.2, respectively. A total of 13/26 metastatic-clade mutations were detectable at low levels in the lymph node biopsy samples (P1.2), consistent with a common ancestor for metastasis as a minor clone at the axillary lymph node site. The inferred phylogenetic structure was stable using 5 and 10% allele fraction cutoffs for high-confidence mutations (Supplementary Figs 8 and 9) and allele fractions for stem mutations were highly correlated between all tumour samples (Supplementary Fig. 10).

Serial plasma analysis and comparison with tumours. In plasma, stem mutations were highest in abundance, with mean plasma AFs ranging from 3.8 to 34.9% across the time series. Metastatic-clade mutations were lower in abundance with mean AFs ranging from 2.5 to 19.1% (Wilcoxon rank sum test $P<0.001$, except T5 $P=0.001$). The dynamic longitudinal changes in plasma AFs for both mutation groups reflected the observed overall tumour response, both clinically and on imaging (Fig. 2a). Mutation clusters statistically inferred using PyClone from variation in circulating mutant allele fractions (without relying on tumour data, referred to as 'plasma clusters') overlapped significantly with clusters identified from multiregional tumour sampling. A total of 21/23 stem mutations were assigned to plasma cluster 1 (with highest cellular prevalence), and 19/26

© 2015 Macmillan Publishers Limited. All rights reserved.

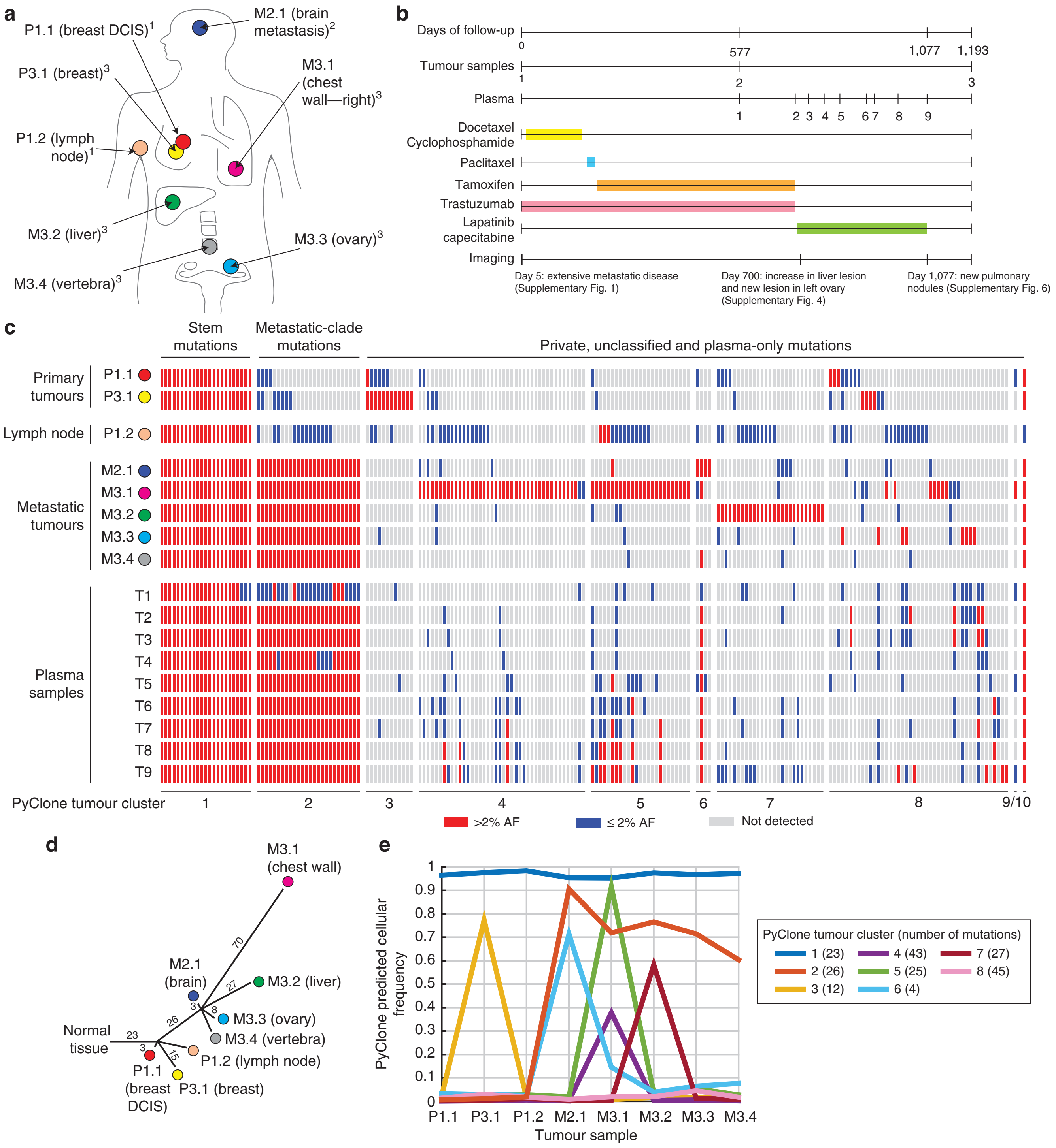


Figure 1 | Inference of clonal structure from multiregional tumour biopsies. (**a**) Tumour samples collected from the patient, labelled P (primary) and M (metastasis). Numbers preceding dot (1,2 and 3) correspond to time of collection: 1, collected at diagnosis; 2, collected at the time of resection of brain metastasis; 3, collected at autopsy. (**b**) Timeline describing clinical course, samples collected, treatments administered and selected imaging assessments. Plasma DNA samples are labelled 1 through 9. Imaging assessments were performed using computed tomography scans. Histopathological and imaging findings are summarized in Supplementary Tables 1 and 2 and Supplementary Figs 1–6. (**c**) Distribution of 207 validated functional mutations in tumour and plasma samples, ordered by mutation clusters inferred using PyClone from mutant allele fractions in all tumours. Red rectangles indicate high-confidence mutations with AF >2%. Blue rectangles indicated mutations detected significantly above background but with AF of 2% or lower. Stem mutations (observed ubiquitously in all tumour samples and comprising tumour cluster 1) and metastatic-clade mutations (high confidence in metastatic tumours and comprising tumour cluster 2) are readily identifiable/detectable in plasma samples. Detailed values of allele fractions are documented in Supplementary Data 1–3. (**d**) Tumour phylogenetic tree, inferred by clonal ordering given distribution of high-confidence mutations in tumour samples shown in **a**. Length of each branch of the tree correlates with the number of mutations on the branch as indicated. Exome-sequencing results for samples P1.2 and M3.4 were not available and therefore private mutations for these branches cannot be identified. Assignment of mutations to each branch is documented in Supplementary Data 1. (**e**) Mean predicted cellular frequency of each cluster identified by PyClone across the tumour samples.

© 2015 Macmillan Publishers Limited. All rights reserved.

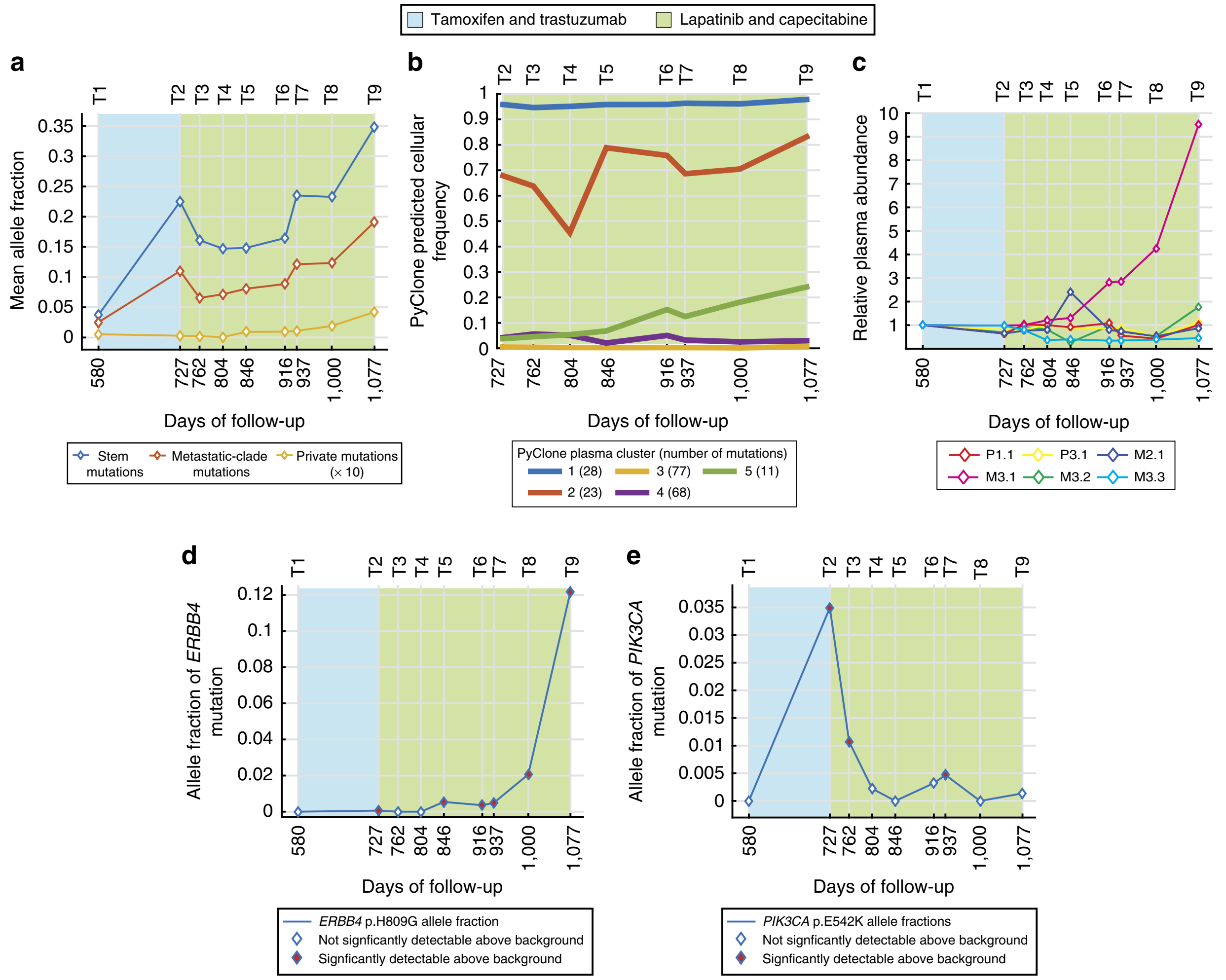


Figure 2 | Serial plasma analysis during systemic treatment. (**a**) Circulating levels of stem, metastatic-clade and private mutations during treatment. Mean allele fractions at each time point are presented. Mean AF for private mutations is multiplied by 10 to highlight trend. Shaded areas represent treatment lines. (**b**) Mean predicted cellular frequency of each cluster identified by PyClone across the plasma samples T2–T9. PyClone identified five mutation clusters from variation of circulating allele fractions (without reliance on tumour data). Clusters 1 and 2 are largely comprised of stem and metastatic-clade mutations. Cluster 5 is comprised of 11 mutations, 10 of which are private M3.1 mutations. (**c**) Plasma abundance calculated as the product of AF in a tumour sample (normalized for mean of stem mutations) and the corresponding AF in a plasma sample, summed across all private mutations for each tumour. To normalize for different number of private mutations in each tumour (3–70), we calculated plasma abundance relative to T1. (**d**) Dynamics of *ERBB4* mutation (p.H809G) in deep sequencing data. (**e**) Allele fractions measured by deep amplicon sequencing for the *PIK3CA* mutation (p.E542K) identified in exome sequencing of plasma sample T2. Mutation was significantly detectable (>3 s.d.'s above the mean allele fraction in control samples) on days 727, 762 and 937 (yellow diamonds).

metastatic-clade mutations were assigned to plasma cluster 2 (Fig. 2b and Supplementary Data 1).

To assess whether plasma DNA captured differential response across distinct metastatic sites during targeted treatment, the relative plasma abundance of high-confidence private mutations originating from each tumour site was calculated. During lapatinib treatment, a rapid increase in the circulating abundance of several mutations private to the chest mass was observed in plasma samples T4–T9 (Fig. 2c), coinciding with significant disease progression seen on imaging at this site. This was also reflected in plasma-based PyClone mutation clusters; plasma cluster 5 increased in circulating prevalence with disease progression on lapatinib treatment and 10/11 mutations in this cluster are private to M3.1 (and correspond to tumour clusters 4 and 5; Fig. 2b and Supplementary Data 1). At the time of lapatinib resistance, the most abundant private mutation in plasma was in the tyrosine kinase domain of *ERBB4* (p.H809G; plasma cluster 5; Fig. 2b,d and Supplementary Fig. 11). This mutation was private to the chest wall mass (28.2% AF) with its levels in plasma DNA increasing during lapatinib treatment up to an AF of 12.2% at the time of disease progression on imaging (compared with average stem and metastatic-clade AFs of 34.9 and 19.1% in the same plasma sample). The predicted functional effect of this mutation[17,18] and its exclusive molecular detection in the chest wall mass (the main site of disease progression on treatment) suggest it was a key determinant of resistance to lapatinib.

Interestingly, 11 non-synonymous high-confidence SNVs were identified and validated in plasma but not detectable at >2% AF in any of the analysed tumour biopsies. Amongst these was an actionable hotspot mutation in *PIK3CA* (p.E542K), identified in plasma with an AF of 3.5% at the time of progression on trastuzumab and tamoxifen (tumour cluster 8 and plasma cluster 4; Fig. 2e). After lapatinib treatment was started, the plasma levels dropped to AF of 1.1% and then became undetectable. This mutation was only marginally detectable (AF <1%) in two

© 2015 Macmillan Publishers Limited. All rights reserved.

tumour biopsies (axillary lymph node and vertebral metastasis). These results suggest this *PIK3CA* mutation originated from a minor tumour sub-clone that increased in size during treatment with tamoxifen and trastuzumab, and then regressed on treatment with lapatinib. Activation of the PI3K/AKT pathway has been associated with resistance to both endocrine therapy and trastuzumab[19,20].

Discussion

In this paper, we have presented an extensive comparison of biopsy and plasma samples collected from a metastatic breast cancer patient over a 3-year clinical course. Our results show that circulating tumour DNA provides a dynamic sampling of somatic alterations reflecting the size and activity of distinct tumour sub-clones. Analysis of ctDNA reflects the clonal hierarchy determined from multiregional tumour sequencing and tracks different treatment responses across metastases. Unlike previous reports, our results qualify tumour-plasma concordance of each somatic mutation in context of the tumour phylogeny. Truncal mutations that represent the majority of tumour lesions in a patient have higher circulating levels and therefore, are more likely to be detected in plasma, than clade or private mutations.

These results were obtained from deep analysis of a single patient and need to be confirmed in a larger cohort of patients with multiregional biopsies and serial plasma samples. If confirmed, our observations have important implications for future ctDNA studies. For monitoring tumour burden using ctDNA, our results suggest that truncal mutations are the best candidates, as they are highest in circulating levels and least likely to drop out during follow-up. For molecular treatment stratification, our results suggest that if multiple actionable somatic mutations, or alterations that are known to confer resistance to specific therapies, are identified in tumour analysis, their relative circulating levels in pretreatment plasma samples may inform the choice of targeted treatments for individual patients. The potential of using plasma DNA for molecular stratification and tracking of resistant clones in patients treated with targeted therapies heralds a new era for precision cancer medicine.

Methods

Sample collection and exome sequencing. Informed consent was obtained and research autopsy was performed under a study protocol approved by the Cambridgeshire Research Ethics Committee (Cambridgeshire 3 REC 07/Q0106/63MN.A). Collection, processing, DNA extraction and preparation of exome-sequencing libraries for plasma samples T1, T2 and T9 have been described previously[4]. Exome sequencing of tumour samples and additional sequencing of germline DNA (N1) was performed using commercially available kits. Tumour and germline DNA were sheared using sonication to a target fragment size of 200 bp. Whole-genome libraries were prepared from 32 to 50 ng of fragmented DNA using ThruPLEX-FD (Rubicon Genomics) as per the manufacturer's protocols, with unique sample-specific molecular barcodes. Genomic libraries were quantified using quantitative PCR and pooled for exome enrichment by hybridization using the TruSeq Exome Enrichment Kit (Illumina). Enriched libraries were quantified using quantitative PCR and pooled for sequencing on the HiSeq 2500 (Illumina).

Targeted amplicon sequencing. Targeted sequencing libraries were prepared using droplet-based PCR amplification following the manufacturer's protocols for ThunderBolts Cancer Panel with specific modifications (RainDance Technologies). Custom target-specific primers were designed using in-house primer design pipelines (see Supplementary Data 5 for the list of primer sequences). Universal adapters were added on the 5′-end to allow sample-specific barcoding. Target-specific amplification was performed using primers flanking 350 loci in multiplex in a 40-μl volume PCR mix. Primer concentration was limited to 3.5 nM per primer (an estimated 10,000 copies per 5 pl droplet). Droplets were generated on the RainDrop Source instrument (8,000,000 droplets for a 40-μl volume). An input of 2–18 ng (mean: 12.1 ng) of plasma DNA (1–10-μl volume of eluted DNA), corresponding to the DNA extracted from a volume of 40–400 μl (mean: 280 μl) of plasma, and 6–31 ng (mean: 21.5 ng) of genomic DNA from tumour and germline samples were used for library preparation. PCR was performed for 55 cycles using 1 °C s^{-1} ramp and following conditions: 94 °C for 30 s, 62 °C for 30 s and 68 °C for 1 min followed by a final extension at 68 °C for 10 min. Droplets were destabilized using manufacturer-supplied reagents. PCR product was purified using magnetic beads (SPRIworks) in 2:1 volume ratio. PCR product was eluted in 20 μl 1 × Tris–EDTA buffer (pH 8.0). A second 25 μl barcoding PCR was performed using 13 μl of the eluted product and primers specific to the universal adapter with sample-specific barcodes. PCR was performed for 10 cycles using 1 °C s^{-1} ramp and following conditions: 94 °C for 30 s, 56 °C for 30 s and 68 °C for 1 min followed by a final extension at 68 °C for 10 min. An additional purification was performed using magnetic beads (SPRIworks) in a 1.2:1 volume ratio. Libraries were quantified using KAPA SYBR FAST LightCycler 480 qPCR kit (KAPA Biosystems) and using DNA High Sensitivity Kit on BioAnalyzer (Agilent Technologies) and pooled in 1:1 ratio. Paired-end sequencing was performed using MiSeq 150-cycle v3 kit (Illumina).

Exome-sequencing analysis and mutation calling. Sequencing reads were demultiplexed allowing zero mismatches in barcodes. Paired-end alignment to the hg19 genome was performed using BWA version 0.5.9 for all exome-sequencing data including germline samples, plasma samples and tumour samples[21]. PCR duplicates were marked using Picard. Local realignment was performed using Genome Analysis Tool Kit[22]. Pileup files were generated for the genomic regions targeted by exome enrichment using samtools v0.1.1722 (ref. 23). For plasma samples, properly paired reads with mapping quality ≥ 60 were used to generate the pileup. AFs for each single-base locus were calculated for all bases with phred quality ≥ 30. For germline DNA, an additional pileup file was generated (using a mapping quality cutoff of ≥ 1 and without any base quality cutoffs) and was used as reference for calling somatic variants. All mutations were annotated for genes and function as well as repeated genomic regions using ANNOVAR[24].

A mutation was identified if (1) no mutant reads for an allele were observed in germline DNA (N1) at a locus that was covered at least 10-fold, (2) at least five reads supporting the mutant were observed in any tumour or plasma sample with at least one read on each strand (forward and reverse) and (3) the binomial probability of observing the number of mutant reads given total depth at that locus was < 0.001 assuming an error rate of 0.01.

Analysis of targeted sequencing data. Sequencing reads were extracted and demultiplexed using Picard allowing zero mismatches in barcodes and a base quality of ≥ 30. Sequencing reads were clipped to remove universal adapter sequences using ea-utils. Minimum amplicon length in our set was 80 bp. Therefore, we removed any sequencing reads < 70 bp in length following adapter clipping to discard nonspecific amplification and primer dimers. Clipped sequencing reads were aligned to the human genome hg19 using BWA version 0.7.10. Unmapped reads, unpaired reads and supplementary alignments were removed. As described previously, reads were demultiplexed to specific amplicons using known amplicon start and end positions and expected amplicon length (accounting for potential indels)[11]. Pileup files were generated using samtools including any reads with mapping quality ≥ 30 and base quality ≥ 30. Pileup data were imported into MATLAB.

For each locus and non-reference allele of interest, we assessed the allele fraction in eight control plasma samples and four control genomic DNA samples. We considered a mutation significantly detectable if the AF in a sample was > 3 s.d.'s higher than the mean AF in control samples.

Control samples. A volume of 250 ml pooled control plasma sample was purchased from BioreclamationIVT (Baltimore, MD, USA). The sample was prepared from equal number of male and female volunteers and collected with K2 EDTA additive. We performed independent cell-free DNA extractions from 1-ml aliquots of plasma and eight aliquots were used as control plasma samples. Four genomic DNA control samples were used from the Human Random Control DNA Panel 3 (Sigma-Aldrich).

Calculation of plasma abundance for private mutations. Plasma abundance was calculated as the product of AF of a private mutation in a tumour sample and the corresponding AF in a plasma sample, summed across all private mutations for each tumour. To account for cellularity of each tumour sample, we normalized the tumour AF of each mutation by mean tumour AF of stem mutations. To normalize for different number of private mutations in each tumour, we calculated plasma abundance relative to T1.

Bayesian clustering using PyClone. PyClone (a Bayesian clustering method) was used to infer the clonal population structures present in the tumour and plasma samples from the amplicon sequencing data. Given the mutation allele frequencies for each sample, PyClone clusters mutations that shift together across the samples and estimates cellular prevalence for each cluster in each sample (adjusting for copy number changes and normal cell contamination). To infer the clonal population structure of each sample (either tumour sample or plasma sample), copy number and depth of coverage information must be determined for each mutation under analysis.

Copy number information at each mutation location was generated from the whole-exome-sequencing data using the CopyWriteR Bioconductor package.

© 2015 Macmillan Publishers Limited. All rights reserved.

CopyWriteR uses off-target read information from targeted sequencing data files to determine copy number. CopyWriteR outputs segmented logarithmic depth of coverage ratios (logR), which are converted to absolute copy number predictions. Segments with logR below -0.25 were assigned a copy number of 1 and those with logR above 0.25 received a copy number of 3. A bin size of 100 kb was utilized with hg19 as the reference genome. Whole-exome-sequencing data were available from four metastatic samples, two primary tumour samples, three plasma samples and patient's germline DNA sample (used as the control in copy number determination). Copy number predictions for other samples were assigned as the median copy number calculated for all available samples of the same type (either plasma, primary or metastasis). If the algorithm was unable to deduce copy number at a given mutation locus, the sequentially nearest valid copy number assignment was used. Inferred total copy number information for tumour and plasma samples is presented in Supplementary Data 4.

Depth of coverage (for the normal and variant alleles at each mutation) was computed using the bam2R function in the deepSNV Bioconductor package. The amplicon sequencing files for each sample were used as input. Reads with a phred quality of 30 or greater were included in the recorded read counts.

Depth of coverage and copy number information for each mutation was then inputted into PyClone (a Bayesian clustering method) to infer the presence of clonal mutations in both the tumour and plasma samples. Two PyClone analyses were performed: one for the tumour samples and another for the plasma samples. For each simulation, the PyClone algorithm was run for 40,000 iterations with a burn-in of 20,000 iterations using the PyClone beta binomial model with the 'total_copy_number' option. A beta binomial value of 500 was utilized. Default values were used for all other parameters. Cellularity for each sample (including ctDNA samples) was estimated by computing the mean allele fraction for mutations classified as 'stem mutations'—these are reported in Supplementary Table 1. Mean predicted cellular frequencies (in the case of ctDNA these should be interpreted as clonal frequencies) for each cluster identified by PyClone are plotted in the Figs 1e and 2b. Because PyClone corrects for normal cell contamination, the predicted cellular frequencies shown in the figures represent the proportion of cancer cells containing each set of clonal mutations (hence stem mutation cluster in plasma being near 100% frequency). The T1 plasma sample was not included in the PyClone analyses; data from the T1 sample were uncharacteristically noisy due to the sample's low cellularity (3%)—a reflection of low systemic tumour burden mid-treatment. The PyClone inference results for two mutations (in the tumour sample simulation) were ambiguous (the 5th–95th percentile credible range from the PyClone post-burn-in trace data spanned more than 70% of the cellular frequency space), leading to singleton clusters for each. Two mutations are not shown in Fig. 1e.

References

1. Aparicio, S. & Caldas, C. The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.* **368,** 842–851 (2013).
2. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
3. Diaz, Jr L. A. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486,** 537–540 (2012).
4. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497,** 108–112 (2013).
5. Chan, K. C. *et al.* Cancer genome scanning in plasma: detection of tumour-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin. Chem.* **59,** 211–224 (2013).
6. Dawson, S. J. *et al.* Analysis of circulating tumour DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368,** 1199–1209 (2013).
7. Diehl, F. *et al.* Circulating mutant DNA to assess tumour dynamics. *Nat. Med.* **14,** 985–990 (2008).
8. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumour DNA with broad patient coverage. *Nat. Med.* **20,** 548–554 (2014).
9. Bettegowda, C. *et al.* Detection of circulating tumour DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6,** 224ra224 (2014).
10. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *J. Pathol.* **231,** 21–34 (2013).
11. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4,** 136ra168 (2012).
12. Heitzer, E. *et al.* Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **5,** 30 (2013).
13. De Mattos-Arruda, L. *et al.* Capturing intra-tumour genetic heterogeneity by de novo mutation profiling of circulating cell-free tumour DNA: a proof-of-principle. *Ann. Oncol.* **25,** 1729–1735 (2014).
14. Eirew, P. *et al.* Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* **518,** 422–426 (2015).
15. Roth, A. *et al.* PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11,** 396–398 (2014).
16. Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486,** 395–399 (2012).
17. Qiu, C. *et al.* Mechanism of activation and inhibition of the HER4/ErbB4 kinase. *Structure* **16,** 460–467 (2008).
18. Canfield, K. *et al.* Receptor tyrosine kinase ERBB4 mediates acquired resistance to ERBB2 inhibitors in breast cancer cells. *Cell Cycle* **14,** 648–655 (2015).
19. Berns, K. *et al.* A functional genetic approach identifies the PI3K pathway as a major determinant of trastuzumab resistance in breast cancer. *Cancer Cell* **12,** 395–402 (2007).
20. Chandarlapaty, S. *et al.* Frequent mutational activation of the PI3K-AKT pathway in trastuzumab-resistant breast cancer. *Clin. Cancer Res.* **18,** 6784–6791 (2012).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
22. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).
23. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
24. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

Acknowledgements

We thank L. Jones, S. Richardson and H. Biggs for recruiting patients into the DETECT study, all medical and ancillary staff in the breast cancer clinic and the patient for consenting to participate. We thank the Human Research Tissue Bank at Addenbrooke's Hospital, which is supported by the NIHR Cambridge Biomedical Research Centre. We acknowledge the support of Cancer Research UK, the University of Cambridge, National Institute for Health Research Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no 337905. Dr Dawson was supported by an Australian National Breast Cancer Foundation and Victorian Cancer Agency Early Career Fellowship. Dr Murtaza was supported by Science Foundation Arizona's Bisgrove Scholars Early Tenure Track award.

Author contributions

M.M., S.-J.D., N.R. and C.C. designed the study; M.M. and D.W.Y.T. developed methods; S.-J.D., E.P., J.G., M.W. and C.C. conducted the prospective clinical, histopathological and imaging studies; M.M., S.-J.D., S.-F.C., D.W.Y.T., D.G., P.S. and T.C.-C. generated data; Z.K., S.H. and D.B. contributed sequencing data; M.M., K.P., O.M.R., F.M., H.F., K.S. and S.P.S. analysed sequencing data; S.-F.C. and H.R.A. contributed to experiments and data analysis; M.M., S.-J.D., N.R. and C.C. interpreted data; M.M. and C.C. wrote the paper with assistance from S.-J.D., N.R. and other authors; all authors approved the final manuscript; N.R. and C.C. are the project co-leaders and joint senior authors.

Additional information

Accession codes: The sequencing data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI, under accession code EGAS00001001466.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: N.R., M.M., F.M., D.G., D.T. and C.C. are co-inventors or contributors on a patent 'A method for detecting a genetic variant', GB patent application no. GB1512626.1 and International patent application no. PCT/GB2015/052086, and may benefit from commercialization of technologies discussed in the manuscript. N.R. and D.G. are co-founders of Invitata, a cancer genomics company focused on ctDNA analysis. D.T. is a consultant for Invitata. Z.K., S.H. and D.B. are full-time employees of Illumina, Inc., providers of the sequencing technology used in this study. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat. Commun.* 6:8760 doi: 10.1038/ncomms9760 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© 2015 Macmillan Publishers Limited. All rights reserved.

The Journal of Molecular Diagnostics, Vol. 20, No. 6, November 2018

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Effects of Collection and Processing Procedures on Plasma Circulating Cell-Free DNA from Cancer Patients

Bente Risberg,[*†‡] Dana W.Y. Tsui,[*§] Heather Biggs,[¶] Andrea Ruiz-Valdepenas Martin de Almagro,[*§] Sarah-Jane Dawson,[*¶‖] Charlotte Hodgkin,[‖] Linda Jones,[¶] Christine Parkinson,[‖] Anna Piskorz,[*§] Francesco Marass,[*§] Dineika Chandrananda,[*§] Elizabeth Moore,[*§] James Morris,[*§] Vincent Plagnol,[**] Nitzan Rosenfeld,[*§] Carlos Caldas,[*§¶‖] James D. Brenton,[*§‖] and Davina Gale[*§]

From the Cancer Research UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, University of Cambridge, Cambridge, United Kingdom; the Department of Cancer Genetics,† Institute for Cancer Research, and the Department of Pathology,‡ Oslo University Hospital, Oslo, Norway; the Cancer Research UK Major Centre,§ Cambridge, United Kingdom; the Cambridge Breast Unit¶ and the Department of Oncology,‖ Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Cambridge, United Kingdom; and the UCL Genetics Institute,** University College London, London, United Kingdom*

Accepted for publication July 26, 2018.

Address correspondence to Dana W.Y. Tsui, Ph.D., or Davina Gale, B.Sc., Cancer Research UK Cambridge Institute, Robinson Way, Cambridge CB2 0RE, United Kingdom. E-mail: tsuiw@mskcc.org or davina.gale@cruk.cam.ac.uk.

Circulating tumor DNA (ctDNA) offers new opportunities for noninvasive cancer management. Detecting ctDNA in plasma is challenging because it constitutes only a minor fraction of the total cell-free DNA (cfDNA). Pre-analytical factors affect cfDNA levels contributed from leukocyte lysis, hence the ability to detect low-frequency mutant alleles. This study investigates the effects of the delay in processing, storage temperatures, different blood collection tubes, centrifugation protocols, and sample shipment on cfDNA levels. Peripheral blood ($n = 231$) from cancer patients ($n = 62$) were collected into K_3EDTA or Cell-free DNA BCT tubes and analyzed by digital PCR, targeted amplicon, or shallow whole-genome sequencing. To assess pre-analytic effects, plasma was processed under different conditions after 0, 6, 24, 48, 96 hours, and 1 week at room temperature or 4°C, or using different centrifugation protocols. Digital PCR showed that cfDNA levels increased gradually with time in K_3EDTA tubes, but were stable in BCT tubes. K_3EDTA samples stored at 4°C showed less variation than room temperature storage, but levels were elevated compared with BCT. A second centrifugation at $3000 \times g$ gave similar cfDNA yields compared with higher-speed centrifugation. Next-generation sequencing showed negligible differences in background error or copy number changes between K_3EDTA and BCT, or following shipment in BCT. This study provides insights into the effects of sample processing on ctDNA analysis. *(J Mol Diagn 2018, 20: 883*–*892; https://doi.org/10.1016/j.jmoldx.2018.07.005)*

Supported by Cancer Research UK, University of Cambridge grants A15601 (J.D.B.), A11906 (N.R.), A20240 (N.R.), and A18072 (J.D.B.), National Institute for Health Research, Cambridge Biomedical Research Centre and Cambridge Experimental Cancer Medicine Centre (J.D.B.), European Research Council under the European Union's Seventh Framework Programme FP/2007-2013/ERC grant 337905 (N.R.), NIHR (C.P.) and Academy of Medical Sciences (C.P.), the Wellcome Trust (C.P.), British Heart Foundation (C.P.), Arthritis Research UK (C.P.), and an Australian National Breast Cancer Foundation and Victorian Cancer Agency Early Career Fellowship (S.-J.D.).

B.R., D.W.Y.T., and H.B. contributed equally to this work.

Disclosures: V.P., N.R., J.D.B., and D.G. are cofounders, shareholders, consultants, and/or employees of Inivata Ltd., a cancer genomics company focused on ctDNA analysis; D.W.Y.T., F.M., N.R., C.C., J.D.B., and D.G. are co-inventors or contributors on a patent describing methods for analysis of rare DNA fragments; N.R. has received research support from AstraZeneca.

Current address of D.W.Y.T., Department of Pathology, Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY; of S.-J.D., Division of Research and Cancer Medicine, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia; of F.M., Department of Biosystems Science and Engineering, ETH Zurich, and SIB Swiss Institute of Bioinformatics, Basel, Switzerland; and of V.P., Inivata Ltd., The Portway Building, Granta Park, Cambridge, UK.

Copyright © 2018 American Society for Investigative Pathology and the Association for Molecular Pathology. Published by Elsevier Inc.
This is an open access article under the CC BY license (*http://creativecommons.org/licenses/by/4.0*).

https://doi.org/10.1016/j.jmoldx.2018.07.005

Circulating tumor DNA (ctDNA) in plasma offers new opportunities for noninvasive cancer management. Recent studies have demonstrated its potential for molecular stratification, monitoring tumor response, identifying resistance mutations, and patients at risk of relapse.[1,2] Detecting ctDNA in plasma is challenging because it constitutes only a minor fraction of the total cell-free DNA (cfDNA), particularly in early-stage cancers and in the minimal residual disease setting.[3,4] A proportion of background wild-type DNA is believed to originate from lysis of white blood cells.[5] Previous studies have highlighted the pre-analytic effects of different processing and collection protocols on plasma ctDNA levels from cancer patients and pregnant women.[6–9] On the basis of these results, it is recommended to process whole-blood samples for retrieval of plasma as soon as possible after collection, before *in vitro* cell lysis. At the same time, a double-centrifugation protocol has been recommended to obtain cell-free plasma, using an initial slow centrifugation speed to separate plasma, then fast centrifugation to clear cellular material.[7] However, some of these procedures may be difficult to perform in a clinical setting due to lack of appropriate personnel or equipment. To circumvent this, cell-stabilizing blood collection tubes have become available to stabilize cfDNA, enabling a delay in processing, which may be done under more controlled conditions and within centralized laboratories. This study performed a systematic comparison of the effects of different processing protocols and collection tubes on the levels of cfDNA and ctDNA from cancer patients using digital PCR (dPCR). With the growing use of next-generation sequencing (NGS) for the analysis of ctDNA, the effect of different protocols and collection tubes on the performance of targeted amplicon and shallow whole-genome sequencing (sWGS) for quantification of plasma DNA was also investigated.

Materials and Methods

Analysis Modules

The study was designed to include five different modules: Module 1 investigated the effects of delayed processing on the levels of circulating DNA (cfDNA and ctDNA) in plasma collected in K_3EDTA tubes (9 mL S-Monovette; Sarstedt, Nümbrecht, Germany). The separation of plasma was delayed for different durations: 0, 6, 24, 48, and 96 hours, and 1 week at room temperature (19°C to 25°C). Module 2 investigated the effects of storage temperature on the levels of circulating DNA in plasma collected in K_3EDTA tubes. Samples were stored at room temperature or at 4°C before processing at the following hours post-collection: 0, 24, 48, and 96 hours. Module 3 investigated the effects of collection devices on the levels of circulating DNA. Blood samples from each patient were collected at the same time point into K_3EDTA tubes and cell-stabilization blood collection tubes (10 mL Cell-Free DNA BCT; Streck, La Vista, NE), respectively. BCTs contain a proprietary formaldehyde-free preservative that stabilizes nucleated blood cells preventing the release of genomic DNA.[10,11] The samples were processed at the following times post-collection: 0, 96 hours, and 1 week. Module 4 investigated the effects of different centrifugation protocols on the levels of circulating DNA. Module 5 investigated the effects of shipment on samples collected in BCT tubes at ambient temperature. For modules 1, 2, 3, and 5, plasma was separated from blood using a double-centrifugation protocol (protocol A): a first centrifugation at 820 × *g* for 10 minutes in a mega-centrifuge (Thermo Sorvall Legend RT; Thermo Fisher Scientific, Waltham, MA), then subjected to a second centrifugation step of the plasma supernatant at 14,000 × *g* for 10 minutes in a benchtop micro-centrifuge (Heraeus Fresco 21; Thermo Fisher Scientific). For module 4, blood aliquots from the same patients were processed with three different protocols: protocol A as above, protocol B with the first centrifugation performed at 1600 × *g* and the second centrifugation at 14,000 × *g* for 10 minutes in a bench top micro-centrifuge, and protocol C with both first and second centrifugations performed in the same mega-centrifuge, initially at 1600 × *g* for 10 minutes, then at 3000 × *g* for 10 minutes.

Patient Samples and DNA Extractions

Peripheral whole blood was collected from 62 patients in total: 47 patients with high-grade serous ovarian cancer and 15 patients with metastatic breast cancer. Informed consent was obtained from each patient with protocols approved by an institutional ethics committee. Fifteen to 30 mL blood from each patient was processed according to each analysis module. DNA from all samples, except module 5, was extracted from an average 1.4 mL (range, 0.3 to 2.76 mL) plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol, except that 6.2 μg of carrier RNA was added per sample. DNA was eluted twice through the column to maximize yield. A nonhuman spike-in PCR product was added to each sample as an internal quality control to assess extraction efficiency.[12] In module 5, DNA was extracted from plasma on a QIAsymphony robot (Qiagen) using a 2-mL extraction protocol. Eluted DNA was stored at −20°C until analysis.

A total of 231 blood samples aliquots were analyzed in this study. Table 1 summarizes the number of plasma samples collected for each module. Note that the collection was designed in such a way that each sample from every processing condition (temperature, collection tube, delayed processing duration) had a matched sample that was collected in K_3EDTA and processed immediately (denoted E.RT.0h) using centrifugation protocol A, and was assigned as the reference sample for each condition. The levels of circulating DNA (either cfDNA or ctDNA), were expressed as a ratio of the respective data with the reference sample

Table 1 Summary of Data Available for Different Modules

Module	Collection devices	Temperature	Delay before sample processing					
			0 hours	6 hours	24 hours	48 hours	96 hours	1 week
Module 1	EDTA	Room temperature	26	21	20	10	5	5
Module 2	EDTA	Room temperature/4°C	26	21	20/11	10/10	5/5	5
Module 3	EDTA/BCT	Room temperature	10/5	-	-	-	5/10	5/15
Module 4	EDTA	Room temperature	13	-	-	-	-	-
Module 5	EDTA/BCT	Room temperature	13/-	-	-	-/10	-/2	-/1*

Module 1: The effects of delayed processing.

Module 2: The effects of storage temperature. Samples were stored both in room temperature and at 4°C. 20/11 indicates that 20 tubes were stored at room temperature and 11 at 4°C, and so on.

Module 3: The effects of collection devices (EDTA versus BCT).

Module 4: The effects of different centrifugation speeds.

Module 5: The effects of shipment in BCT.

Dashes indicate no data available.

(E.RT.0h). Therefore, data collected under the same processing conditions could be grouped together to evaluate the effect of the processing even though they were collected from different patients. A more detailed summary of the distribution of samples involved in each module is given in Supplemental Table S1.

Quantification of Circulating Plasma DNA by dPCR and Targeted Amplicon Sequencing

Plasma samples from ovarian and breast cancer patients were first quantified by dPCR (using the Biomark microfluidic system (Fluidigm, South San Francisco, CA) as previously described,[13] using an assay that targets a 65-bp amplicon in *RPP30*, a nonamplified region in the genome, to estimate cfDNA levels.[12,14] ctDNA levels were then determined by dPCR using dual-labelled patient-specific TaqMan assays designed to mutant and wild-type sequences in *TP53* or *PIK3CA*, or deletions in chromosome 8, 11, or 17. A summary of the samples analyzed is provided in Supplemental Table S1, and sequences of primers and fluorescent probes, amplicon sizes, and amplification conditions used in dPCR are detailed in Supplemental Table S2.

The levels of cfDNA and ctDNA were calculated from the number of observed amplifications above a set threshold, and Poisson statistics were used to convert the number of observed amplifications to estimated targets, assuming independent segregation of DNA molecules into the microfluidic reaction chambers. The total number of amplifiable copies of DNA molecules per mL of plasma (copies/mL) were calculated, taking into account the relative fraction of the extracted DNA loaded and the proportion of sample lost during the loading process through the microfluidic channels. The levels of ctDNA were calculated as mutant allele fraction (ie, the fraction of mutant DNA copies divided by the total cfDNA copies) expressed as a percentage or as mutant copies/mL plasma. For the purpose of comparing different protocols in the modules, the data are expressed at each processing condition as a ratio from the E.RT.0h reference sample that was collected in K_3EDTA and immediately processed according to protocol A, unless otherwise specified.

To investigate the effects of different collection devices and processing protocols on the performance of NGS, plasma samples from all modules were analyzed by Tagged Amplicon deep sequencing (TAm-Seq), as previously described.[13] TAm-Seq is a targeted amplicon sequencing method that allows identification and quantification of low-frequency mutant alleles in plasma across sizable genomic regions. Sequencing was performed using an Illumina HiSeq 2500 sequencer (Illumina, San Diego, CA) to an average of greater than 1000× sequencing depth. Mutations were identified and quantified as previously described.[13] To assess the effect of collection and processing procedures on the background error rates during NGS, the allelic read ratio (reference/alternative) was generated at each position within R software version 3.1.2[15] from the BAM files, using the Bioconductor 3.2 software packages Rsamtools version 1.22.0 and Biostrings version 2.38.4.[16] All positions flagged as polymorphic by the 1000 Genomes Project (*http://www.internationalgenome.org*, last accessed) or the COSMIC database (*https://cancer.sanger.ac.uk/cosmic*, last accessed), were filtered out.

To investigate the effects of shipping on global somatic copy number alterations, samples in module 5 were also subjected to sWGS.[17] Briefly, a DNA library was prepared from 2 to 10 ng of cfDNA from each sample using the ThruPLEX DNA-seq Kit (Takara Bio, Inc., Shiga, Japan) and sequenced on an Illumina HiSeq 4000 to 0.1× average depth using single-end sequencing. Sequence data were analyzed using a pipeline that involved the following: single-end sequence reads were aligned to the human reference genome (GRCh37) using BWA-mem software version 0.7.17[18] after removing any contaminant adapter sequences. SAMtools software version 1.7 (*http://samtools.sourceforge.net*)[19,20] was used to convert files to BAM format. PCR and optical duplicates were marked using Picard-Tools' MarkDuplicates software feature version

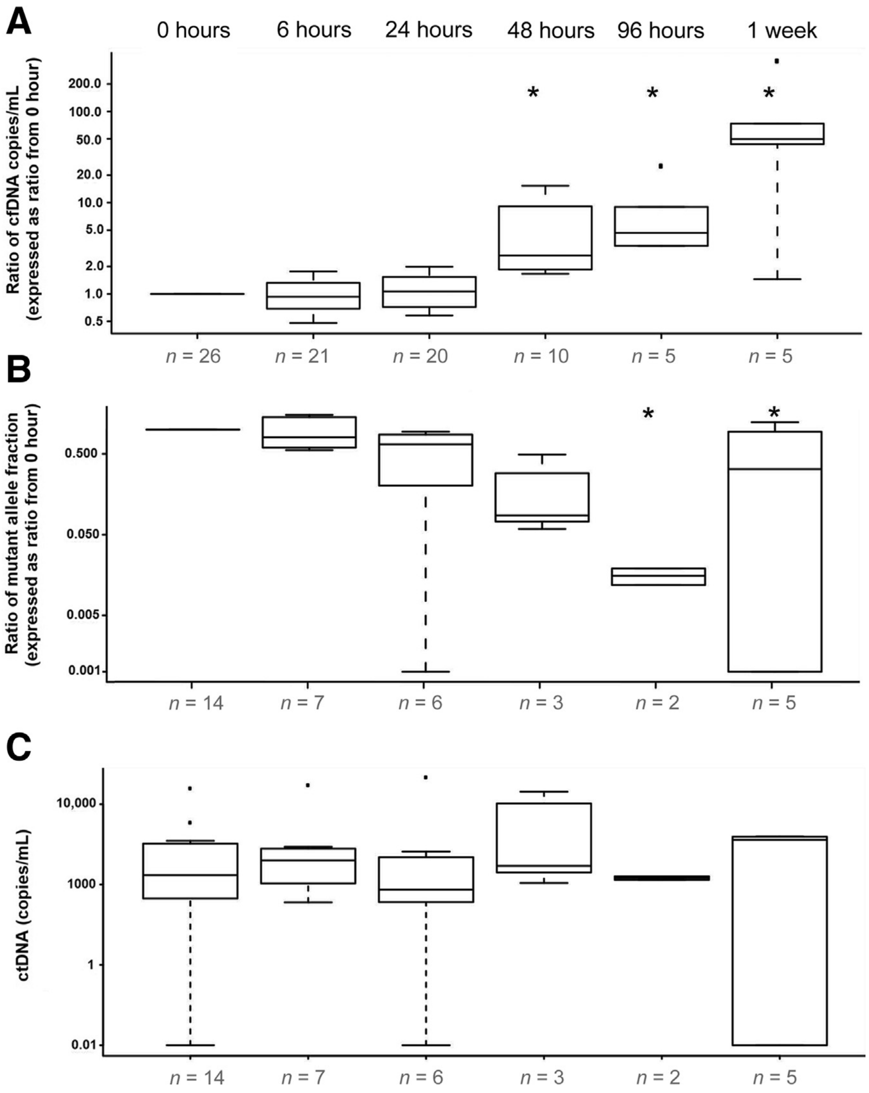


Figure 1 The effects of delayed processing on the levels of circulating DNA in plasma collected in K_3EDTA tubes. Blood samples were collected into K_3EDTA tubes and stored at room temperature for 0, 6, 24, 48, and 96 hours, and 1 week before plasma separation. Cell-free DNA (cfDNA) copies/mL plasma (**A**), mutant allele fraction (**B**). **C:** Circulating tumor DNA (ctDNA) copies/mL plasma in samples processed at different time of delay. The bottom and top of the box represent the first and third quartiles, respectively, and the band inside the box represents the median. Data are expressed as the ratio from E.RT.0h of each patient's immediately processed K_3EDTA sample. $^*P < 0.05$ versus E.RT.0h (Mann–Whitney rank sum test).

2.17.6 (*https://broadinstitute.github.io/picard*), and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments. Reads in each sample were down-sampled to approximately 3 million reads to have similar coverage between patients and conditions. Subsequently, copy number analysis was performed in R[15] using the R package CNAclinic version 1.0 (*https://github.com/sdchandra/CNAclinic*, last accessed December 21, 2017), a software suite that allows for robust copy number analysis of sWGS data. Briefly, sequence reads were allocated into equally sized (100 Mb) nonoverlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and regions corresponding to artifacts and probable germline changes were excluded from downstream analysis utilizing a cohort of 45 healthy controls. After median normalization, binned counts were segmented using both the Circular Binary Segmentation– and Hidden-Markov Model–based algorithms, and an averaged $\log_2$ R value per bin was calculated.

Statistical Analysis

The difference in circulating DNA levels between different subgroups in each module was analyzed using nonparametric Mann–Whitney rank sum test unless specified, and $P < 0.05$ was considered statistically significant. To assess the noise of sWGS data, values corresponding to the median of the absolute values of all pairwise differences were calculated between $\log_2$ R copy numbers. This metric provides a measure of the noise of the sample that is less dependent on true biological copy number variation and more on technical variation.[21] To compare the three collection methods in all patients, pairwise Spearman correlations were calculated between the binned copy number segments of the three collection methods. Furthermore, a nonparametric Wilcoxon signed rank test was

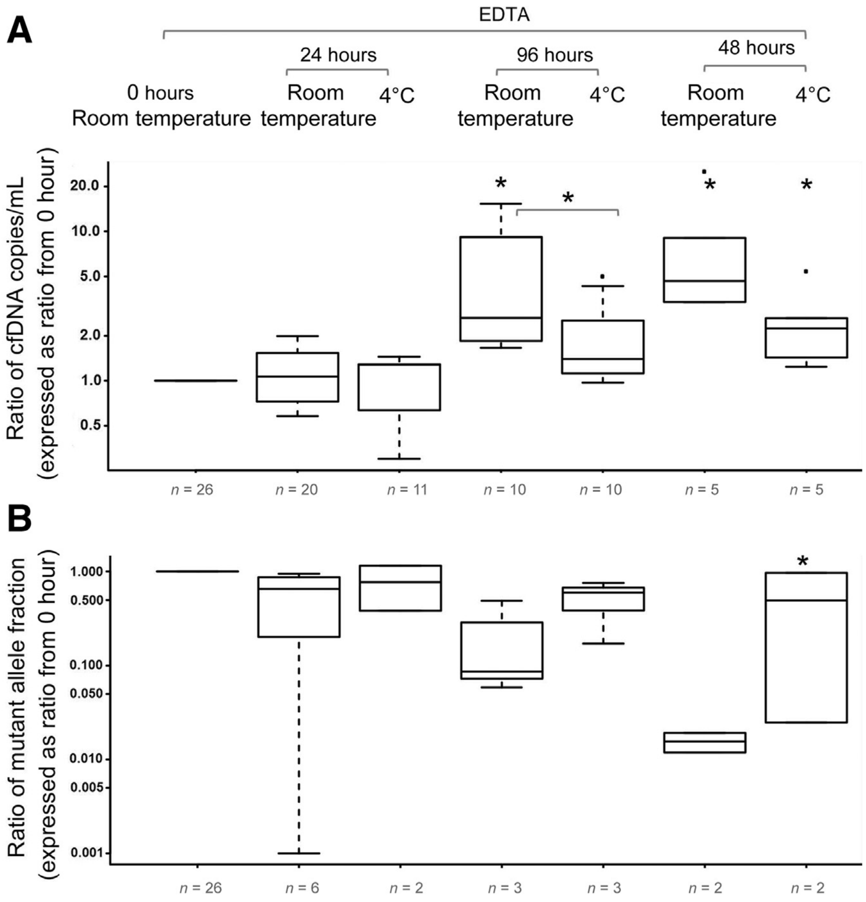


Figure 2 The effects of storage temperature on the levels of circulating DNA in plasma collected in K_3EDTA tubes. Blood samples collected into K_3EDTA tubes were stored at room temperature and at 4°C for 24, 48, and 96 hours, and 1 week before plasma was separated. Cell-free DNA (cfDNA) copies/mL plasma (**A**) and mutant allele (**B**) fraction. The bottom and top of the box represent the first and third quartiles, respectively, and the band inside the box represents the median. Data are expressed as the ratio from E.RT.0h of each patient's immediately processed K_3EDTA sample. $^*P < 0.05$ versus E.RT.0h (Mann–Whitney rank sum test).

applied on these values to test the similarity of the copy number profiles between all pairwise samples.

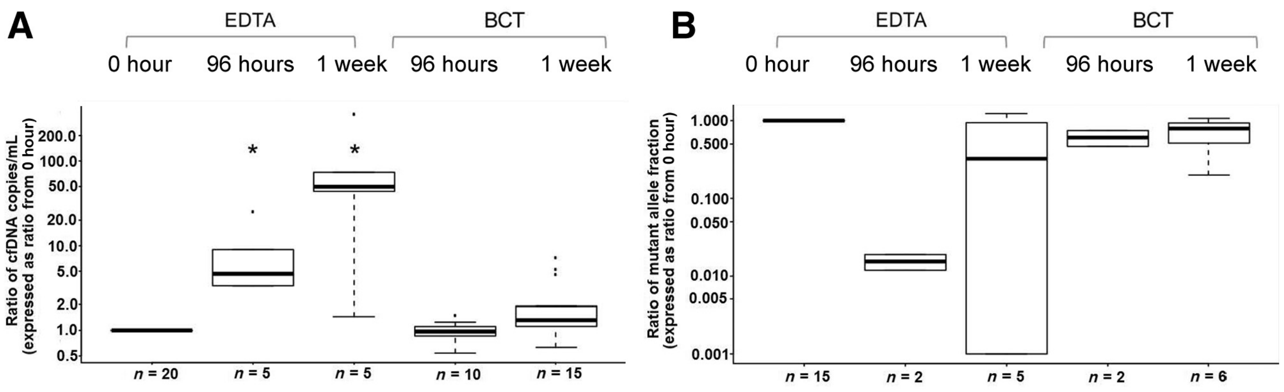
Results

Module 1: The Effects of Delayed Processing on the Levels of Circulating DNA in Plasma Collected in EDTA Tubes

In this module, all samples ($n = 26$) were collected in K_3EDTA tubes. One tube from each collection was processed immediately. The other tubes were stored at room temperature and processed at different prolonged time points: 6, 24, 48, 96 hours, and 1 week. Analysis by dPCR showed that the levels of cfDNA in the plasma samples increased gradually with increasing delay in the processing (Figure 1A), whereas the fraction of ctDNA decreased (Figure 1B). In particular, the levels of cfDNA increased significantly after 48, 96 hours, and 1 week of delay, whereas the mutant allele fraction of ctDNA decreased significantly after 96 hours and 1 week of delay (Mann–Whitney rank sum test, $P < 0.05$). Previous reports have indicated that in analysis of circulating cell-free DNA from maternal plasma, despite changes in total cfDNA, the levels of fetal DNA are relatively stable in different storage and processing conditions.[8,22] Indeed, our results confirm that the numbers of mutant molecules, expressed as copies/mL of plasma, were relatively stable across the different processing time points with no statistically significant difference observed compared to samples that were processed immediately (Figure 1C and Supplemental Figure S1).

Module 2: The Effects of Storage Temperature on the Levels of Circulating DNA in Plasma Collected in K_3EDTA Tubes

In this module, all samples ($n = 26$) were collected in K_3EDTA tubes and either processed to plasma immediately or after 24, 48, and 96 hours. The individual tubes were stored in two conditions: at room temperature (19°C to 25°C) or at 4°C. If kept at room temperature, dPCR showed that the levels of cfDNA significantly increased after 48 hours. If kept at 4°C, the levels increased after 48 hours but were significantly lower than those observed at room temperature (Figure 2A). If delayed for 96 hours, samples kept at room temperature and 4°C all increased significantly. The changes in mutant allele fraction showed an inverted



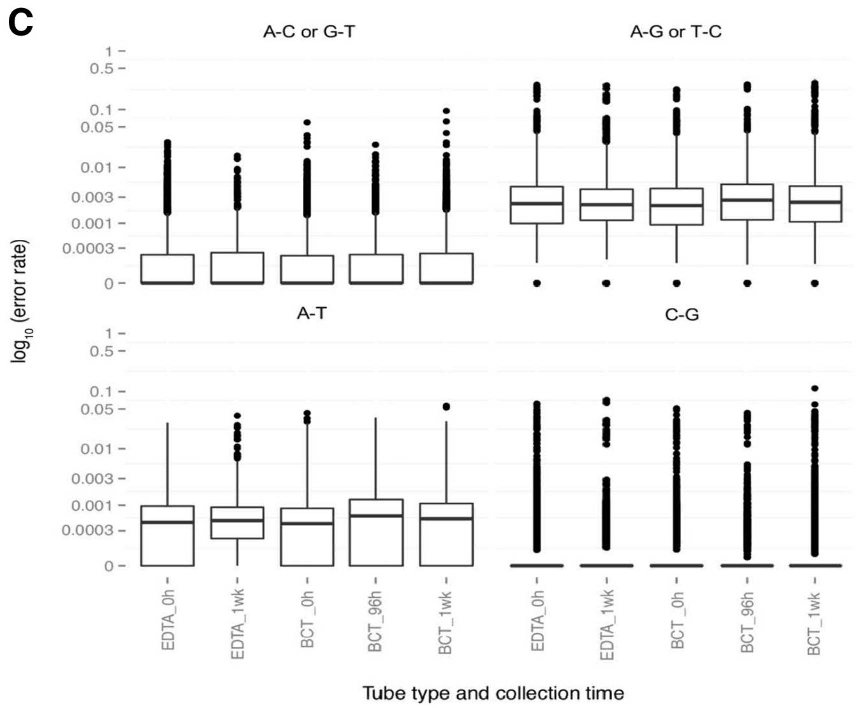


Figure 3 The effects of collection device (K_3EDTA versus BCT) on the levels of circulating DNA. Blood samples collected into K_3EDTA tubes were processed immediately, after 96 hours or 1 week at room temperature. Blood samples in BCT were stored at room temperature for 96 hours and 1 week before plasma separation. Cell-free DNA (cfDNA) copies/mL plasma (**A**) and mutant allele (**B**) fraction. **C:** The distributions of the ratio of nonreference/reference alleles as generated by targeted amplicon sequencing shown in boxplots. The bottom and top of the box represent the first and third quartiles, respectively, and the band inside the box represents the median. Data are expressed as the ratio from E.RT.0h of each patient's immediately processed K_3EDTA sample (**A** and **B**) or $\log_{10}$ scale (**C**). $^*P < 0.05$ versus E.RT.0h (Mann—Whitney rank sum test).

similar trend, although the amount of available data were too low for statistical analysis (Figure 2B).

Module 3: The Effects of Collection Devices (K_3EDTA versus Cell-Free DNA BCT) on the Levels of Circulating DNA

In this module, one K_3EDTA tube for each collection was processed immediately (E.RT.0h) and served as a reference sample ($n = 20$). The other K_3EDTA tubes were stored for 96 hours ($n = 5$) and 1 week ($n = 5$) at room temperature. Cell-free DNA BCT's were stored at room temperature and processed immediately ($n = 5$) or delayed for 96 hours ($n = 10$) and 1 week ($n = 15$). The cfDNA levels increased significantly after 1 week if kept in K_3EDTA tubes, but remained at similar levels if kept in BCT (Figure 3A). The changes in the mutant allele fraction showed an inverted similar trend, but the amount of data available were too low for statistical analysis (Figure 3B). The mutant allele fraction from six patients that were collected in K_3EDTA and processed immediately, versus the matched samples that were collected in BCT was compared and processed after 1 week's delay. The levels of ctDNA were similar for four patients but decreased twofold for two patient samples (Supplemental Figure S2). There was no statistically significant difference in the numbers of mutant copies/mL plasma between storage in the two tube types (Supplemental Figure S1).

The effects of collection and processing procedures on the background error rates during NGS analysis were next assessed using targeted amplicon sequencing. As previously described, different A/C/G/T base substitutions are associated with different error rates.[13] The distribution of the ratio of nonreference/reference alleles was plotted as box plots, shown according to mutation types. No difference was observed using different collection devices and processing conditions (Figure 3C).

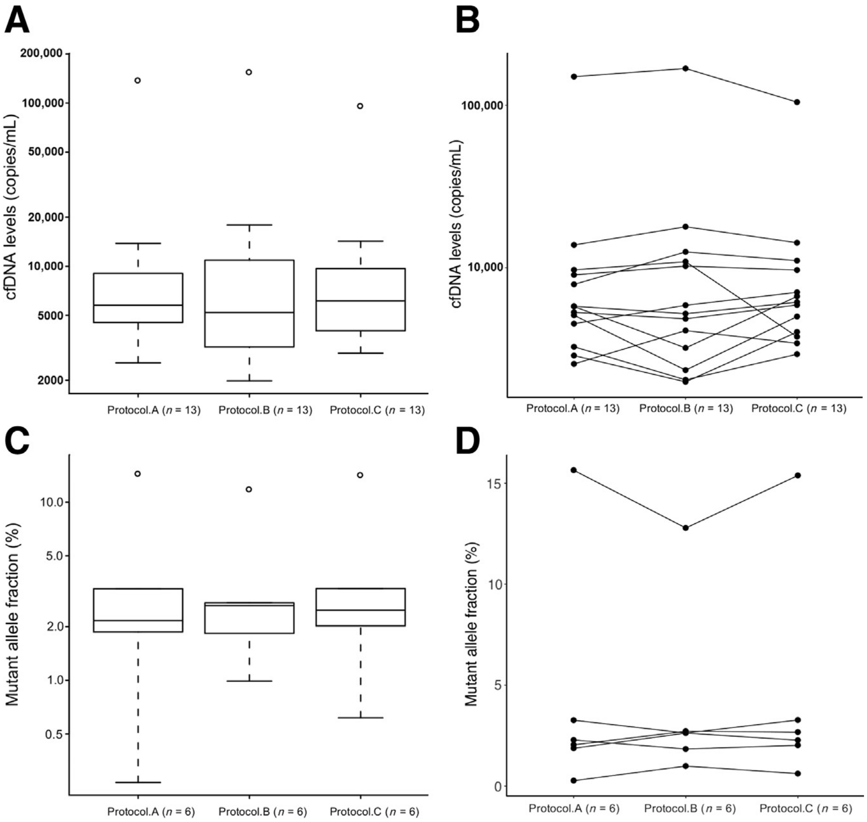


Figure 4 The effects of different centrifugation speeds on the levels of circulating DNA. Blood samples were collected into K_3EDTA tubes and processed to plasma with three different protocols. All protocols included two 10-minute centrifugation steps, the first on whole blood, and the second on plasma aliquots. Protocol A (820 and 14,000 × *g*), protocol B (1600 and 14,000 × *g*), Protocol C (1600 and 3000 × *g*). Cell-free DNA (cfDNA) copies/mL plasma (**A** and **B**) and mutant allele (**C** and **D**) fractions (%) in samples processed by different protocols. The bottom and top of the box represent the first and third quartiles, respectively, and the band inside the box represents the median.

Module 4: The Effects of Different Centrifugation Speeds on the Levels of Circulating DNA

In this module, all samples ($n = 13$) were collected in K_3EDTA tubes and processed immediately. Aliquots from the same patients were processed using three different centrifugation protocols (A to C) as defined in *Materials and Methods*. There were no statistically significant differences across the three protocols on the total circulating DNA levels as measured by dPCR (Figure 4, A and B), or in mutant allele fraction as measured by targeted amplicon sequencing (Figure 4, C and D).

Module 5: The Effects of Shipment of cfDNA BCT on Mutant Allele Fraction and Global Copy Number Changes

In this module, three tubes of blood were drawn from each patient ($n = 13$). K_3EDTA tubes were processed immediately (E.RT.0h), one cell-free DNA BCT was collected and stored at room temperature within the same centralized processing laboratory, whereas the other BCT was packaged and shipped back to the same laboratory. All shipped samples, apart from three, were received and processed within 48 hours from the time of collection. Of these, two BCTs were processed after 96 hours and one was processed after 5 days. The stored BCTs were processed at the same time as the matched shipped sample. There was no statistically significant difference in cfDNA levels between the three collection methods (Figure 5, A and B). *TP53* mutations were identified by TAm-Seq in four patients, and there were no statistically significant differences in mutant allele fraction using the different collection methods (Figure 5, C and D).

To further investigate the effects of collection methods on global copy number changes, sWGS analysis was performed on four patients with detectable *TP53* mutations (P161, P227, P479, P488) and four without (P615, P489, P464, P450). Data from one patient (P464) were excluded from further analysis because the total read count generated for one of the collection methods was below 1 million. This is below the threshold recommended for inference when analyzing shallow coverage.[23] The segmental copy number profiles among the three collection methods were highly similar, showing an average Spearman correlation of 0.76, range 0.44 to 0.98 (Supplemental Figure S3 and Supplemental Table S3). The paired Wilcoxon test P values indicated no significant differences in all 21 copy number distributions comparisons ($P > 0.001$). Supplemental Figure S4 shows an example of the copy number alterations in plasma samples processed with and without shipping. The same gains and losses in chromosomal arms were identified in all three protocols. Supplemental Figure S5 depicts the estimation of noise in the sWGS data using values that were the median of the absolute values of all pairwise differences. All patients showed very similar noise levels between the different tubes and protocols.

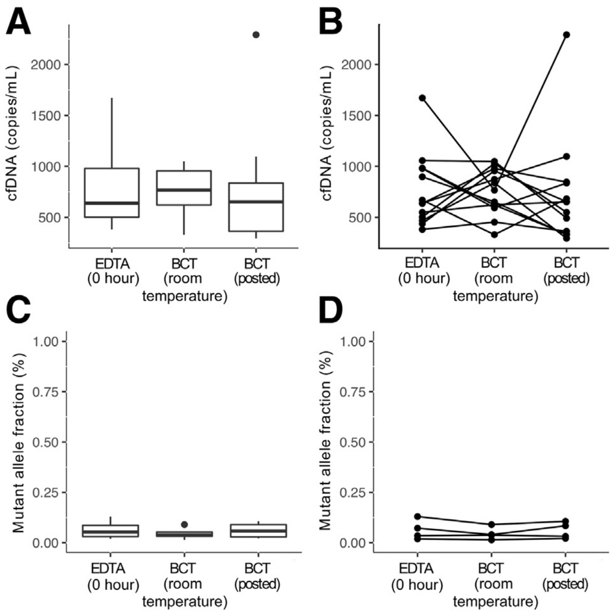


Figure 5 The effects of shipping using cell-free DNA BCT on the levels of circulating DNA. Blood samples were collected in K_3EDTA and cell-free DNA BCT tubes, and processed immediately except for one cell-free DNA BCT from each collection that was shipped by mail back to the same collection center [BCT (posted)]. Cell-free DNA (cfDNA) levels (copies/mL) (**A** and **B**) and mutant allele (**C** and **D**) fractions. The bottom and top of the box represent the first and third quartiles, respectively, and the band inside represents the median.

Discussion

Multiple research studies have demonstrated the potential of using plasma as a tool for noninvasive cancer management. There is increasing interest in incorporating ctDNA as a liquid biopsy in both clinical and research settings. Because the frequency of mutant alleles in plasma may be low, particularly in early-stage disease, it is crucial to optimize and standardize pre-analytic sample processing procedures to maintain the quality of samples for accurate quantification of rare mutant molecules. In this study, the pre-analytic effects of blood sample processing procedures, including the use of different blood collection tubes, storage conditions, and centrifugation speeds, were examined on downstream analysis of cfDNA using different molecular technologies including dPCR, targeted amplicon, and genome-wide sequencing. Our results show that levels of cfDNA are stable in K_3EDTA tubes at room temperature for up to 24 hours. If delayed beyond 24 hours, storage of K_3EDTA blood at 4°C appeared to delay the increase in background cfDNA. It is worth noting that a recent study demonstrated that storing the samples in K_2EDTA tubes at 4°C kept the cfDNA levels stable for a course of 3 days.[24] This agrees with the observations that storing K_3EDTA tubes at 4°C improved the stability of cfDNA compared with room temperature storage. Alternatively, collection into cell-free DNA BCT tubes at room temperature maintained stable cfDNA levels for at least a week. These tubes can facilitate delayed and centralized blood processing, circumventing issues arising with delayed plasma processing. Other researchers have evaluated alternative cell-stabilization tubes such as CellSave (CellSearch system; Menarini Silicon Biosystems, Huntington Valley, PA) and PAXgene Blood ccfDNA tubes (Qiagen) and demonstrated similar stability when sample processing was delayed.[9,25] New cell-free stabilization tubes have recently become available [eg, Cell-free DNA Collection tube (Roche, Basel, Switzerland), cf-DNA Preservation tube (Norgen Biotek, Thorold, ON, Canada), Blood STASIS 21-ccfDNA, (MagBio Genomics, Gaithersburg, MD), and LBgard Blood tubes, Biomatrica, San Diego, CA)], and it will be important to test these thoroughly to assess their performance for optimal sample processing procedures before next-generation sequencing and dPCR analysis of ctDNA.

These findings have addressed a few of the practical challenges in the blood-to-plasma sample processing workflow in a hospital setting. For example, in the clinic, processing may be delayed due to shortage of staff to enable immediate processing, or collection outside office hours. In some scenarios, when conducting multicenter clinical trials, many individual centers do not have access to the full spectrum of centrifuges with the higher second centrifugation speeds required to perform the recommended double-centrifugation procedures. The ability to delay processing by collecting into cell-stabilization tubes, or the flexibility to perform the centrifugation in a range of different types of centrifuges, or storing at 4°C after collection for a short period, will greatly improve the feasibility of collecting high-quality specimens. For samples collected across a wide geographical area, shipment may be necessary before central processing to standardize pre-analytic factors and maximize cost-effectiveness. This study showed no statistically significant difference in NGS background noise with or without shipment. However, other studies have shown that the shipping temperature of cell-free DNA BCT was deemed to be a critical factor to ensure delivery of high-quality specimens for downstream ctDNA analysis.[26] In these studies, variable results were observed at extreme temperatures, at $\leq$10°C and 40°C, which affected the cellular interface, resulted in an elevated ratio of long/short genomic DNA fragments, and a decrease in plasma volume. These studies indicate that shipment temperature should be carefully controlled by the use of insulated packages, gel blocks, or temperature logging devices to maintain stability.

Previous studies have mainly focused on locus-specific analysis using quantitative PCR or dPCR that examined one locus at a time. With technology advances, an increasing number of molecular profiling strategies have been developed using NGS,[27] which provides a higher resolution and larger genomic coverage than a locus-specific approach. It is therefore important to also understand the effects of cfDNA sample processing on the analytical performance of NGS-based analysis. It is particularly important to test whether using a collection tube containing a preservative has the potential to introduce DNA sequence modifications, which may be misinterpreted as true patient-specific genomic alterations. A recent study examined the influence of sample collection in CellSave tubes on the analysis of global copy number variations using NGS technology, and did not find differences in allele frequencies compared with EDTA blood.[9] In this study with BCT and K_3EDTA tubes, the effects of processing on the background error rates during targeted amplicon sequencing and sWGS were evaluated. As expected, different error rates were observed in different base substitutions, but there was no difference in background error rate regardless of the type of collection device and sample processing schedule. The sWGS analysis results agreed with previous findings in that copy number data were consistent across conditions.[28]

All of these findings provide important insights for the potential incorporation of routine NGS technology in plasma-based molecular diagnostics. Beyond the analysis of ctDNA, it is crucial to also understand the impact of pre-analytical factors on other nucleic acids or genomic variants, such as tumor-specific RNA (ctRNA), microRNA, or DNA methylation, some of which have been studied,[29] but more evidence is required. Their quantification would likely be affected by the levels of total RNA or methylated DNA that is derived from the blood cells. It is important to understand whether the effects of sample processing procedures could be addressed in a similar manner to the effects on circulating DNA.

With the increasing understanding of genomic alterations and matched targeted treatment options, the demand for a non-invasive molecular profiling tool is growing. Analyzing cell-free nucleic acids presents a unique opportunity for longitudinal follow-up during treatment of cancer patients. Initiatives have begun to pursue the standardization of methods for cell-free DNA analysis. Understanding the impact of different pre-analytic factors will help accelerate the process and drive large-scale cross-center validation studies to provide robust evidence for clinical utility of circulating tumor DNA and its integration into routine clinical practice.

Acknowledgments

We thank the patients for consenting to participate, the Human Research Tissue Bank at Addenbrooke's Hospital, which is supported by the NIHR Cambridge Biomedical Research Center, Frank Diehl for discussions about method standardization, Irena Hudecova for help preparing figures, and the Genomics Core of the Cancer Research UK Cambridge Institute for providing sequencing support.

D.W.Y.T., H.B., S.-J.D., C.P., A.P., N.R., C.C., J.D.B., and D.G. conceived and designed the study; D.W.Y.T., H.B., S.-J.D., C.H., L.J., C.P., and A.P. processed samples, collected clinical data, and managed samples; B.R., D.W.Y.T., A.R.-V.M.A., S.-J.D., E.M., and D.G. performed experiments; F.M., D.C., J.M., and V.P. analyzed NGS data; B.R., D.W.Y.T., D.C., and D.G. wrote the manuscript; all authors approved the final version.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2018.07.005*.

References

1. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R, Rosenfeld N: Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017, 17:223−238
2. Siravegna G, Marsoni S, Siena S, Bardelli A: Integrating liquid biopsies into the management of cancer. Nat Rev Clin Oncol 2017, 14: 531−548
3. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, Cheang M, Osin P, Nerukar A, Kozarewa I, Garrido JA, Dowsett M, Reis-Filho JS, Smith IE, Turner NC: Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med 2015, 7:302ra133
4. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al: Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014, 6:224ra24
5. Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM: Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem 2002, 48: 421−427
6. El Messaoudi S, Rolet F, Mouliere F, Thierry AR: Circulating cell free DNA: preanalytical considerations. Clin Chim Acta 2013, 424: 222−230
7. Chiu RWK, Poon LLM, Lau TK, Leung TN, Wong EM, Lo YMD: Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem 2001, 47:1607−1613
8. Barrett AN, Zimmermann BG, Wang D, Holloway A, Chitty LS: Implementing prenatal diagnosis based on cell-free fetal DNA: accurate identification of factors affecting fetal DNA yield. PLoS One 2011, 6:e25202
9. Rothwell DG, Smith N, Morris D, Leong HS, Li Y, Hollebecque A, Ayub M, Carter L, Antonello J, Franklin L, Miller C, Blackhall F, Dive C, Brady G: Genetic profiling of tumours using both circulating free DNA and circulating tumour cells isolated from the same preserved whole blood sample. Mol Oncol 2016, 10:566−574
10. Fernando MR, Chen K, Norton S, Krzyzanowski G, Bourne D, Hunsley B, Ryan WL, Bassett C: A new methodology to preserve the original proportion and integrity of cell-free fetal DNA in maternal plasma during sample processing and storage. Prenat Diagn 2010, 30:418−424
11. Norton SE, Lechner JM, Williams T, Fernando MR: A stabilizing reagent prevents cell-free DNA contamination by cellular DNA in plasma during blood sample storage and shipping as determined by digital PCR. Clin Biochem 2013, 46:1561−1565
12. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N: Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013, 368:1199−1209
13. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N: Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012, 4:136ra68

14. Wang J, Ramakrishnan R, Tang Z, Fan W, Kluge A, Dowlati A, Jones RC, Ma PC: Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. Clin Chem 2010, 56:623−632
15. R-Core-TEAM: R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing, 2015
16. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M: Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods 2015, 12:115−121
17. Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, Benezeder T, Auer M, Pischler C, Mannweiler S, Pichler M, Eisner F, Hauesler M, Riethdorf S, Pantel K, Samonigg H, Hoefler G, Augustin H, Geigl JB, Speicher MR: Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Med 2013, 5:30
18. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25:1754−1760
19. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25:2078−2079
20. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011, 27:2987−2993
21. Affymetrix: Median of the Absolute Values of all Pairwise Differences and Quality Control on Affymetrix Genome-Wide Human SNP Array 6.0. Santa Clara, CA: Affymetrix; 2008
22. Wong D, Moturi S, Angkachatchai V, Mueller R, DeSantis G, van den Boom D, Ehrich M: Optimizing blood collection, transport and storage conditions for cell free DNA increases access to prenatal testing. Clin Biochem 2013, 46:1099−1104
23. Gusnanto A, Wood HM, Pawitan Y, Rabbitts P, Berri S: Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. Bioinformatics 2012, 28:40−47
24. Parpart-Li S, Bartlett B, Popoli M, Adleff V, Tucker L, Steinberg R, Georgiadis A, Phallen J, Brahmer J, Azad N, Browner I, Laheru D, Velculescu VE, Sausen M, Diaz LA Jr: The effect of preservative and temperature on the analysis of circulating tumor DNA. Clin Cancer Res 2017, 23:2471−2477
25. Warton K, Yuwono NL, Cowley MJ, McCabe MJ, So A, Ford CE: Evaluation of Streck BCT and PAXgene stabilised blood collection tubes for cell-free circulating DNA studies in plasma. Mol Diagn Ther 2017, 21:563−570
26. Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F, Holtrup F: Performance of streck cfDNA blood collection tubes for liquid biopsy testing. PLoS One 2016, 11:e0166354
27. Katsanis SH, Katsanis N: Molecular genetic testing and the future of clinical genomics. Nat Rev Genet 2013, 14:415−426
28. Henao Diaz E, Yachnin J, Grönberg H, Lindberg J: The in vitro stability of circulating tumour DNA. PLoS ONE 2016, 11:e0168153
29. Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, Blighe K, Marchese SD, Hills A, Woodley L, Stebbing J, Coombes RC, Shaw JA: Influence of plasma processing on recovery and analysis of circulating nucleic acids. PLoS One 2013, 8:e77963

PLOS | ONE

RESEARCH ARTICLE

Development of a highly sensitive liquid biopsy platform to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA

Davina Gale[1]*, Andrew R. J. Lawson[1], Karen Howarth[1], Mikidache Madi[1], Bradley Durham[1], Sarah Smalley[2], John Calaway[2], Shannon Blais[2], Greg Jones[2], James Clark[1], Peter Dimitrov[1], Michelle Pugh[1], Samuel Woodhouse[1], Michael Epstein[1], Ana Fernandez-Gonzalez[3], Alexandra S. Whale[3], Jim F. Huggett[3,4], Carole A. Foy[3], Gerwyn M. Jones[3], Hadas Raveh-Amit[5], Karin Schmitt[5], Alison Devonshire[3], Emma Green[1], Tim Forshew[1], Vincent Plagnol[1], Nitzan Rosenfeld[1]*

1 Inivata Ltd, Granta Park, Cambridge, United Kingdom, **2** Inivata Inc, Research Triangle Park, NC, United States of America, **3** LGC, Teddington, Middlesex, United Kingdom, **4** School of Biosciences & Medicine, Faculty of Health & Medical Science, University of Surrey, Guildford, United Kingdom, **5** Horizon Discovery, Waterbeach, Cambridge, United Kingdom

* davina.gale@cruk.cam.ac.uk (DG); nitzan.rosenfeld@inivata.com (NR)

Check for updates

OPEN ACCESS

Citation: Gale D, Lawson ARJ, Howarth K, Madi M, Durham B, Smalley S, et al. (2018) Development of a highly sensitive liquid biopsy platform to detect clinically-relevant cancer mutations at low allele fractions in cell-free DNA. PLoS ONE 13(3): e0194630. https://doi.org/10.1371/journal.pone.0194630

Editor: Alvaro Galli, CNR, ITALY

Received: August 25, 2017

Accepted: February 19, 2018

Published: March 16, 2018

Copyright: © 2018 Gale et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. S4 Table describes all allele fractions (shown as percentages) for the common low frequency variants validated/quantified in Horizon Tru-Q 7 and in the InVision liquid biopsy panel for both laboratories (one tab per laboratory). S5 Table details the list of variants analyzed in the Horizon Tru-Q 6 and Horizon Tru-Q 7 dilution study. S6 Table shows the sensitivity analysis for all samples. All reportable calls (excluding SNPs and non-coding changes) made using eTAm-Seq

Abstract

Introduction

Detection and monitoring of circulating tumor DNA (ctDNA) is rapidly becoming a diagnostic, prognostic and predictive tool in cancer patient care. A growing number of gene targets have been identified as diagnostic or actionable, requiring the development of reliable technology that provides analysis of multiple genes in parallel. We have developed the InVision™ liquid biopsy platform which utilizes enhanced TAm-Seq™ (eTAm-Seq™) technology, an amplicon-based next generation sequencing method for the identification of clinically-relevant somatic alterations at low frequency in ctDNA across a panel of 35 cancer-related genes.

Materials and methods

We present analytical validation of the eTAm-Seq technology across two laboratories to determine the reproducibility of mutation identification. We assess the quantitative performance of eTAm-Seq technology for analysis of single nucleotide variants in clinically-relevant genes as compared to digital PCR (dPCR), using both established DNA standards and novel full-process control material.

Results

The assay detected mutant alleles down to 0.02% AF, with high per-base specificity of 99.9997%. Across two laboratories, analysis of samples with optimal amount of DNA

technology to analyse the novel full-process control are available in S7 Table. In addition, a completed dPCR dMIQE checklist and assay information files have been provided for the full-process control study.

Funding: The analytical validation study was funded by Inivata. Inivata provided support in the form of salaries for authors DG, AL, KH, MM, BD, SS, JCa, SB, GJ, JCl, PD, MP, SW, ME, EG, TF, VP and NR, and participated in study design, data collection and analysis, decision to publish, and preparation of the manuscript. The work on the novel full-process control material performed by LGC was funded by the UK government Department for Business, Energy & Industrial Strategy (BEIS) and Innovate UK under the project 'Enabling stratified medicine through cell and tissue reference standards for minimally invasive cancer testing' (Project Number: 101862) in partnership with Horizon Discovery. LGC provided support in the form of salaries for authors AF-G, AW, JH, CF, GMJ and AD, and Horizon Discovery provided support in the form of salaries for authors HR-A and KS. LGC and Horizon Discovery did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of all authors are articulated in the 'author contributions' section.

Competing interests: DG, AL, KH, MM, BD, SS, JCa, SB, GJ, JCl, PD, MP, SW, ME, EG, TF, VP and NR are current or former employees, officers, consultants and/or share-holders of Inivata Ltd or Inivata Inc. Inivata, InVision, TAm-Seq and eTAm-Seq are UK registered trademarks of Inivata Ltd. DG, TF and NR are co-founders of Inivata Ltd. NR is a Board member of Inivata. DG, SW, NR and TF are co-inventors or contributors on patent applications that describe methods for the analysis of DNA fragments and applications of circulating tumor DNA. AF-G, AW, JH, CF, GMJ and AD are current or former employees of LGC. HR-A and KS are former employees of Horizon Discovery. This does not alter our adherence to PLOS ONE policies on sharing data and materials, with restrictions described in the Data Availability Statement.

detected 94% mutations at 0.25%-0.33% allele fraction (AF), with 90% of mutations detected for samples with lower amounts of input DNA.

Conclusions

These studies demonstrate that eTAm-Seq technology is a robust and reproducible technology for the identification and quantification of somatic mutations in circulating tumor DNA, and support its use in clinical applications for precision medicine.

Introduction

Circulating cell-free DNA (cfDNA) from cancer cells, commonly referred to as circulating tumor DNA (ctDNA), is known to be present in the plasma of cancer patients. Since the first report of identical DNA mutations in plasma compared to a patient's tumor, ctDNA has been investigated as a tool for cancer diagnosis, detection, prognostication, treatment selection and monitoring [1–3]. Over the past decade, increasing evidence demonstrates the utility of ctDNA as a 'liquid biopsy' to supplement conventional biopsies for molecular characterization and monitoring of solid cancers [4–7]. Circulating tumor DNA can be readily accessed from a non-invasive blood draw, allowing easier access to genomic information from a patient's tumor or metastases as the cancer evolves, without the associated expense, complications or risk to patients during surgery or biopsy. Moreover, tissue testing may not be a viable option in many patients. In the Iressa Pan-Asia Study (IPASS), a phase III randomized study of gefitinib vs. carboplatin/paclitaxel in patients with pulmonary adenocarcinoma, *EGFR* mutation status could only be evaluated in 437/1038 (42%) patients that gave their consent for biomarker analyses [8]. The high failure rate may be due to a number of reasons, including insufficient biopsy material available, because the biopsy was of too poor quality for adequate analysis, or because surgery was not possible for medical reasons. In such cases, ctDNA can provide a valuable alternative for molecular stratification to select appropriate therapy. With the development of targeted therapies, the molecular profile of the cancer has been established to be informative to select therapies that are more likely to be effective in given patient groups. For example, tyrosine kinase inhibitors (TKIs), such as gefitinib and erlotinib, have been shown to be effective in non-small cell lung cancer (NSCLC) patients carrying activating *EGFR* exon 19 deletions or L858R mutations, and vemurafenib is known to be beneficial to patients with *BRAF* V600E mutations [9–11]. It has also been shown that it is possible to detect tumor evolution in plasma ctDNA [6, 12, 13]. *EGFR*-mutant NSCLC patients can now be tested and monitored to identify the emergence of newly arising *EGFR* T790M resistance mutations, and be effectively treated with osimertinib, a third generation TKI [14–15].

Studies have shown that ctDNA levels often correlate with tumor burden, and provide an earlier and potentially more reliable measure of treatment response than other clinical biomarkers, such as CA-15-3 in metastatic breast cancer, and CA-125 in advanced high-grade serous ovarian cancer [5, 7]. Recent exciting developments have shown that it is possible to use ctDNA as a tool to assess minimal residual disease [16], and be used to identify mutant DNA in early stage cancer, although this is much more challenging given the lower number of mutant molecules present in the bloodstream [17,18]. With such diversity in potential clinical applications, it is important to use a ctDNA assay that has high sensitivity and specificity, and can interrogate multiple mutations in parallel to detect, track and monitor clinically-relevant genomic changes as the cancer evolves. Several techniques are available for the analysis of ctDNA. Many of the earlier studies focused on analyzing single mutated regions. Digital PCR (dPCR) and BEAMing have both been established as sensitive techniques for the detection and quantification of specific 'hotspot' mutant alleles [19, 4]. The cobas *EGFR* Mutation Test v2 is a real-time PCR test for the qualitative detection of *EGFR* exon 19 deletions, L858R and T790M mutations, and is used to determine which NSCLC patients are eligible for treatment with erlotinib or osimertinib. The test has gained FDA-approval for testing on both plasma and tissue, making it the first companion diagnostic that allows the use of ctDNA analysis to guide treatment [20]. The FDA-approved assays, however, are less sensitive than digital PCR, with a limit of detection (LOD) at >25 copies/mL of plasma [21]. Analysis of single genomic loci is restricted to a limited number of pre-defined hotspots. To analyse multiple mutations, cfDNA must first be sub-divided for each assay, reducing the sensitivity of the test and introducing potential sampling bias for detection of low frequency alleles.

The development of next generation sequencing (NGS) has allowed for a broader application of ctDNA analysis. In 2012, Forshew et al. developed TAm-Seq technology, or tagged-amplicon deep sequencing which, for the first time, enabled interrogation of 6 genes across a large genomic region spanning 5995 bases to detect low frequency mutations in cell-free DNA [22]. The assay was evaluated in plasma from patients with high-grade serous ovarian cancer, and shown to have 97% sensitivity and specificity for detection of mutations at 2% allele fraction (AF), and was able to identify mutations down to 0.14% AF. Analysis of clinical samples showed that it is possible to use TAm-Seq technology to assay multiple mutations in parallel to monitor tumor dynamics, identify *de novo* mutations direct from patient cfDNA, and identify the origin of metastatic relapse. Since this time, other NGS technologies have been implemented for analysis of ctDNA, including the use of hybrid capture and the introduction of panel assays that use molecular barcodes to enable error suppression [6, 23, 24]. The ideal ctDNA assay needs to have high sensitivity and specificity, have good turnaround times and target clinically-relevant and clinically actionable genes. This will enable oncologists to make clear treatment decisions based on molecular profiling information, according to cancer care guidelines and used in conjunction with other clinical observations.

Here we describe the development of the InVision liquid biopsy platform which utilizes enhanced TAm-Seq (eTAm-Seq^{TM}) technology for the identification of low frequency mutations in ctDNA. The assay has been expanded to target hotspots and entire coding regions from 35 cancer-related genes, utilizing a primer design strategy that allows for amplification of highly fragmented DNA, typical of ctDNA. The calling algorithm has been revised, and in addition to improved detection of single nucleotide variants (SNVs) and short insertions/deletions (indels), it also identifies copy number variants (CNVs). The library preparation process has been adapted to remove the use of microfluidics and to reduce the background error rate. We present analytical validation of the InVision liquid biopsy platform across two laboratories to demonstrate its reproducibility and to support the use of this platform in clinical applications. We compare the performance of eTAm-Seq technology and digital PCR by analysis of sheared cell-line reference standard DNA and novel full-process control material developed by LGC and Horizon Discovery.

Materials and methods

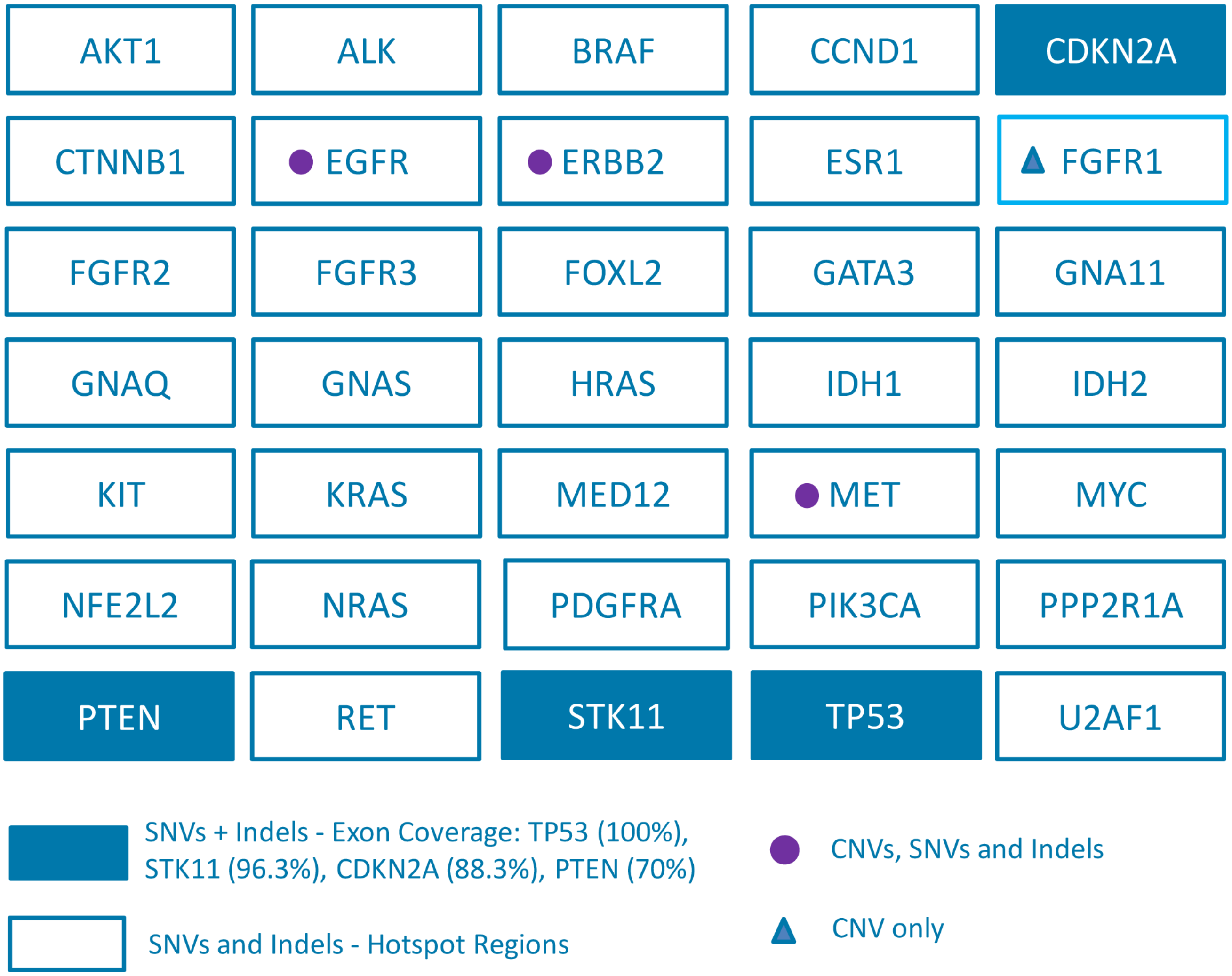
Analytical validation of eTAm-Seq technology

To assess the performance of the eTAm-Seq technology, analytical validation studies were performed in two laboratories within the scope of CLIA (Laboratory 2) and ISO 15189:2012 quality standards (Laboratory 1). Next-generation sequencing libraries were prepared using eTAm-Seq technology, analysing sheared reference standard DNA and cfDNA extracted from control plasma from presumed healthy controls. Healthy control samples used in this study were obtained on a commercial basis, from BioreclamationIVT (US) and Seralab (UK). Cell-free DNA (cfDNA) was extracted from 5 mL plasma using a QIAamp Circulating Nucleic Acid kit (Qiagen) as previously described [22], incorporating an internal control to monitor extraction efficiency. cfDNA and the internal control were both quantified by dPCR, using either the Fluidigm Biomark or Biorad QX200, with a 108 bp assay targeting a region of ribonuclease P/MRP subunit p30 (*RPP30*) gene (Forward = `5′-GGAGGTGGAGGAGGAGGATA-3′`; Reverse = `5′-ACGGAATACAGAACCCATGACT-3′`; Probe = `5′-FAM/AGCCTTGAG/ZEN/ AGACGAGAACCTGT/IABkF Q-3′`) and an assay targeting the internal control, as previously described [22]. Yields were expressed as amplifiable copies (AC) per 10 mL blood.

Preparation of sheared cell-line reference standard DNA for analytical validation studies. Horizon Tru-Q 6 (2.5% Tier) and Tru-Q 7 (1.3% Tier) reference cell-line DNA samples (Horizon Discovery), carrying cancer-related mutations at known allele fractions, were sheared to ~200 bp by acoustic shearing (Covaris) to mimic fragmented cfDNA. Tru-Q 6 contains a mix of 20 isogenic genetically-engineered cell lines with known engineered and endogenous mutations predominantly at ~2%-2.5% AF (range: <2%-30% AF), and Tru-Q 7 contains a mix of 40 cell lines with known mutations predominantly at ~1%-1.3% AF (range: <1%-30% AF). Dilutions were prepared using sheared Horizon Tru-Q 0 wild-type DNA as diluent.

InVision liquid biopsy tumor profiling panel. The InVision liquid biopsy platform utilizes an enhanced version of TAm-Seq technology to identify and quantify low frequency tumor-derived SNVs and indels in cfDNA. The technology is also able to identify CNVs in *EGFR*, *ERBB2* (also known as *HER2*), *FGFR1* and *MET* [25]. Full analytical validation of CNVs is not included in this study. The assay targets 35 cancer-related genes spanning 10.61kb, using primers designed to hotspots and entire coding regions of interest. Covered regions were chosen to maximise the mutation yield for common cancer types primarily NSCLC, focusing on clinically actionable mutations. We therefore included oncogenes *EGFR*, *BRAF*, *KRAS*, *ERBB2*, *MET* (exon 14), *U2AF1*, *CTNNB1*, *EGFR*/*MET* amplifications as well as tumour suppressor genes *TP53*, *STK11*, *PTEN*. We further included key regions of *ESR1*/*GATA3*, as well as *ERBB2*/*FGFR1* amplifications, and the most common mutation hotspots in common carcinomas as defined by COSMIC frequencies. The panel was designed optimizing primers for amplification of fragmented DNA with amplicon sizes ranging from 72bp-154bp. The primers were selected based on factors including GC content, similar Tm (target 60˚C), avoidance of primer dimer, avoidance of off-target products and avoidance of SNPs. Fig 1 shows an overview of the InVision liquid biopsy tumor profiling panel, and S1 Table provides detail of the exonic regions covered.

Library preparation using eTAm-Seq technology. eTAm-Seq technology is based on methods previously described [22, 25,26], with an optimized assay workflow utilizing multiplex PCR to enable high-throughput library preparation without the use of microfluidics. Next generation sequencing libraries were prepared using a two-step multiplex PCR amplification process incorporating replicate and patient-specific barcodes and Illumina sequencing adaptors. Different input amounts of DNA were used to assess the performance of the assay, using either 2,000 AC (low), 8,000 AC (medium) or 16,000 AC (high) input (~6.6ng to 53ng of amplifiable DNA). All regions were analysed multiple times using a fixed DNA input range for all samples to enable error correction [26]. As each sample is analysed multiple times, false positive and true positive calls can be readily identified, providing a robust analytical pipeline [22, 26]. After target enrichment, amplified regions were purified using SPRISelect beads (Beckman Coulter) following the manufacturer's protocol. Samples were quantified using the LabChip GX touch and DNA high sensitivity assay. Quantified samples were then pooled to generate a normalized library of 12 nM. This library was quantified using the Kapa Library



AKT1	ALK	BRAF	CCND1	CDKN2A
CTNNB1	● EGFR	● ERBB2	ESR1	▲ FGFR1
FGFR2	FGFR3	FOXL2	GATA3	GNA11
GNAQ	GNAS	HRAS	IDH1	IDH2
KIT	KRAS	MED12	● MET	MYC
NFE2L2	NRAS	PDGFRA	PIK3CA	PPP2R1A
PTEN	RET	STK11	TP53	U2AF1

■ SNVs + Indels - Exon Coverage: TP53 (100%), STK11 (96.3%), CDKN2A (88.3%), PTEN (70%)

● CNVs, SNVs and Indels

□ SNVs and Indels - Hotspot Regions

▲ CNV only

Fig 1. InVision liquid biopsy tumor profiling panel. The coverage per gene is indicated, including hotspots, comprehensive or full coverage of coding regions (70%–100% tiling coverage) and CNVs. SNVs = Single Nucleotide Variants; Indels = short insertions or deletions; CNVs = Copy Number Variants.

https://doi.org/10.1371/journal.pone.0194630.g001

Quantification Kit, and 1.8 pM libraries analysed on an Illumina NextSeq 500 (300 cycle PE) with 5% PhiX to monitor sequencing performance.

Data analysis. Sequencing files were analysed using the Inivata Somatic Mutation Analysis (ISoMA) pipeline to identify SNVs, CNVs and indels. A minimum Phred quality score of 30 for each base was required for inclusion in the analytics. The pipeline clipped primers and merged paired-end reads into synthetic reads (using Flash v1.2.11). A minimum Phred quality score of 2 was assigned to discordant positions at the merging step. Default settings were used for Flash and a Phred quality score of 2 was assigned to mismatched base pairs. These synthetic reads were subsequently aligned to the reference genome using BWA (v0.7.12). Samples passing sequencing QC were kept for further analysis.

To enable variant calling, the background noise for each potential SNV was compared to the variability observed from a set of control samples [22]. The same statistical principle was used for indels using samples from the same batch of samples in order to enable appropriate background calibration. In addition, each run was assessed using positive and negative controls. Common single nucleotide polymorphisms (SNPs) were used to identify potential cross sample contamination, as well as rule out potential swaps for longitudinal studies involving multiple samples from the same patient. The final determination of a call integrated the data across replicates for the sample within a maximum likelihood framework. Variants were annotated using the variant effect predictor [27] based on the canonical transcript for each gene. SNVs and indels that resulted in coding and splice-site mutations were reported. For CNVs, a normalized measure of read depth that corrects for sample and amplicon effects was used to infer the number of DNA copies. A mutation calling report was generated providing a comprehensive summary of somatic alterations identified.

Comparison of performance of eTAm-Seq technology and digital PCR by analysis of novel full-process control material

Preparation of pooled plasma. Sixteen human plasma samples of ~20 mL each from 6 male and 10 female donors were obtained from Seralab (UK). All plasma samples had undergone a second centrifugation step of 1,000 x *g* for 10 minutes at 4˚C, following initial centrifugation from whole blood. Samples were stored at -80˚C upon receipt. Samples were pooled and homogenized using a roller mixer for 30 minutes at 4˚C followed by preparation of 5.0 mL aliquots which were frozen at -80˚C.

cfDNA reference standards. Multiplex I cfDNA Reference Standards (Horizon Discovery) were generated from genomic DNA isolated from isogenic cell-lines, and fragmented to ~160 bp by acoustic shearing (Covaris). The standards, containing 8 known mutations in *EGFR* (L858R, Δ746–750, T790M, V769-D770insASV), *KRAS* (G12D), *NRAS* (A59T, Q61K) and *PIK3CA* (E545K), were diluted to 8 ng/μL for spiking into plasma.

cfDNA extraction. Following thawing of plasma aliquots, 50 μL (400 ng) of Multiplex I cfDNA Reference Standard containing target mutations at ~5%, ~1%, ~0.1% AF or 100% wild-type DNA was added to 5 mL pooled plasma and mixed by vortexing for 10 seconds. DNA was extracted from plasma samples using the QIAamp Circulating Nucleic Acid Kit (Qiagen) and eluted in 50 μL AVE buffer. Replicate extractions (n = 6) were performed for all four levels of Reference Standard (5%, 1%, 0.1% and 100% wild-type) and plasma only controls over 3 days (2 extractions per day). Extracts were divided into two aliquots (25 μL) and frozen, with one aliquot analysed by the eTAm-Seq technology and one aliquot analysed by digital PCR.

Mutational analysis. Samples were analysed using the eTAm-Seq technology in Laboratory 1 using an average of 12,450 AC per reaction. Digital PCR analysis was performed using a QX200 droplet dPCR system (Bio-Rad) with a C1000 Touch Thermal Cycler (Bio-Rad) at LGC. *KRAS* G12/WT and *EGFR* L858R/WT mutations were assessed using PrimePCR assays (Bio-Rad) and custom designed assays were used targeting *NRAS* A59T/WT and *PI3KCA* E545K/WT (S2A–S2C Table). Primers and BHQplus probes for custom assays were supplied by BioSearch and diluted in 1 x TE pH 8.0 (Sigma). Reactions (20 μL) were prepared (with 10% excess) and contained ddPCR Supermix for Probes with no dUTP (Bio-Rad), 20x primer/probe mix, 4 μL cfDNA extract (n = 1 per target mutation) with the remaining volume nuclease-free water (Ambion). Non-spiked Multiplex I cfDNA Reference Standards (32 ng/reaction) were analysed alongside the spiked extracts as controls (n = 3). Data was analysed using QuantaLife (Bio-Rad, version 1.6.6.0320) with classification of single positive, double positive and negative droplets as shown in S1A–S1E Fig. Copy number concentration was calculated based on a partition volume of 0.85 nL.

Calculation of LOD for dPCR assays. The LOD of dPCR assays were calculated using the approach described in Whale et al. based on modelling two binomial distributions, combining a 5% probability of a false positive ($\alpha = 0.05$) with a 5% probability of a false negative ($\beta = 0.05$) [28]. The false positive rate (FPR) for each assay was calculated from analysis of the 100% WT Multiplex I cfDNA Reference Standard (n = 6 reactions, 80 ng per reaction). The 'critical level' is the 95th percentile for a binomial distribution with n trials and probability given by the false positive rate per droplet (λ), where n is the mean number of droplets from the data. The LOD expressed as mutant copies per reaction is given by $n \times -\ln(1-p)$ where p is the probability of success for the binomial distribution with n trials and where the 5th percentile equals the critical level. The LOD expressed as AF% is the previous value relative to the total number of target copies (i.e. n times the mean concentration per droplet (λ) of the wild-type target, plus the mutant value) (S3 Table).

Results

Analytical validation of the eTAm-Seq technology

Analytical validation studies were performed to assess sensitivity of the eTAm-Seq technology for detection of SNVs and indels. Horizon Tru-Q 6 Tier 2.5% and Tru-Q 7 Tier 1.3% cell-line reference standard DNA, carrying mutations at known AF were sheared to ~200bp to approximate cfDNA. There are 21 mutations present in Tru-Q6, and 38 mutations in Tru-Q7 targeted by the InVision liquid biopsy tumor profiling panel. Dilutions were prepared using Horizon Tru-Q 0 wild-type DNA as diluent. Data previously published showed that concentrations of cfDNA in plasma of cancer patients was >1.65 ng/mL in 99% of patients, >6.6 ng/mL in 80% of patients, and >13.2 ng/mL in nearly 50% of patients [29]. This is consistent with cfDNA amounts in 10 mL blood samples (approximately 4~4.5 mL of plasma) from NSCLC patients which were previously shown to contain >2,000 AC (~6.6 ng) in >95% of samples, >8,000 AC (~26.4 ng) in >69% of samples and >16,000 AC (~52.8 ng) in >40% of samples [25]. Dilutions were therefore performed to prepare low (2,000 AC), medium (8,000 AC) and high (16,000 AC) input amounts. Limit of Detection (LOD), inter-operator and inter-laboratory variability were assessed by performing the assay across two laboratories by 6 operators on different days and sequenced on different NGS runs. Each operator independently performed the entire process, and one of the operators performed the assay at each of the two laboratories. Multiple assays were performed in each laboratory at different dilution levels, as shown in Table 1.

Analysis using eTAm-Seq technology showed that the assay had high sensitivity (Fig 2). In Laboratory 1, sensitivity was 100% (90% confidence interval (CI): 99.01%-100%) in low input samples at 2%-2.5% AF, 99.17% (90% CI: 97.40%-99.85%) in medium input samples at 1%-

Table 1. Details of analytical validation experiments performed to assess sensitivity of the eTAm-Seq technology, including range of input DNA (AC), AF (%), number of sample repeats per operator, and number of operators per laboratory.

Input DNA (AC)	AF (%)*	Number of repeats/operator	Number of operators/laboratory
16,000	0.25%-0.33%	4	3
8,000	1%-1.3%	3	3
2,000	2%-2.5%	7	3
8,000	0.5%-0.65%	3	3
8,000	0.25%-0.33%	3	3
8,000	0.13%-0.16%	3	3
8,000	0.06%-0.08%	3	3

* AF shows indicative ranges for Tru-Q reference material, full list of values presented in S4 Table.

https://doi.org/10.1371/journal.pone.0194630.t001

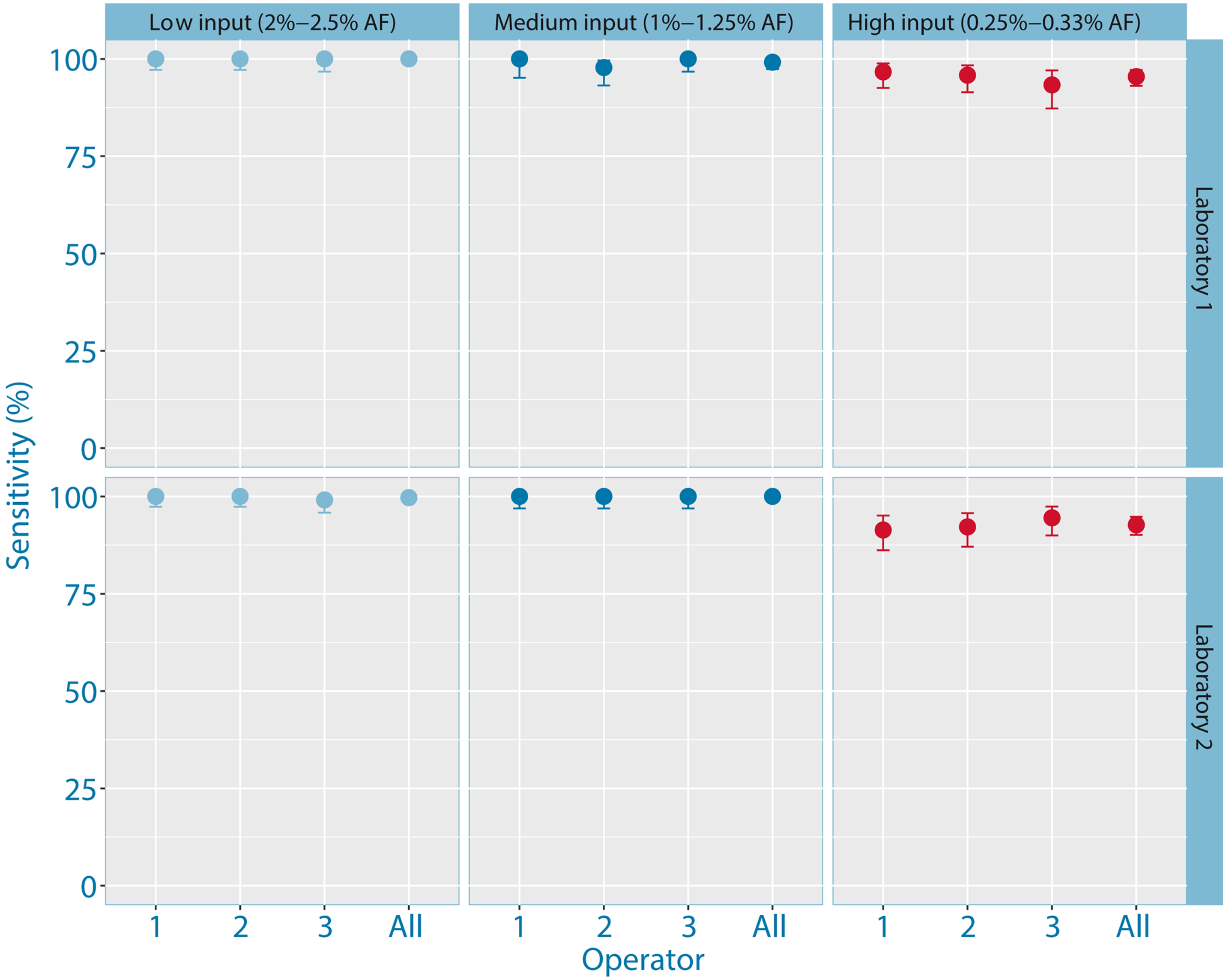


Fig 2. Plot showing sensitivity and inter-operator variability of eTAm-Seq technology using low, medium and high input DNA. Experiments were performed in two laboratories (Laboratory 1 –upper; Laboratory 2 –lower) by different operators, performed on separate days and different NGS runs.

https://doi.org/10.1371/journal.pone.0194630.g002

Table 2. Sensitivity of the eTAm-Seq technology with 8000 amplifiable copies of DNA input per sample.

	Laboratory 1			Laboratory 2		
AF (%)*	**Sensitivity (%)**	**90% CI (Lower)**	**90% CI (Upper)**	**Sensitivity (%)**	**90% CI (Lower)**	**90% CI (Upper)**
1%-1.3%	99.17	97.40	99.85	100.00	98.96	100.00
0.5%-0.65%	99.63	98.26	99.98	97.66	95.43	98.97
0.25%-0.33%	89.17	85.29	92.30	90.28	86.91	93.00
0.13%-0.16%	69.26	64.31	73.89	67.71	62.88	72.26
0.06%-0.08%	37.41	32.50	42.52	30.86	26.10	35.95

* AF shows indicative ranges for Tru-Q reference material, full list of values presented in S4, S5 and S6 Tables.

https://doi.org/10.1371/journal.pone.0194630.t002

1.3% AF, and 95.45% (90% CI: 93.09%-97.18%) in high input samples at 0.25%-0.33% AF. Comparable results were seen in the second laboratory, with a sensitivity of 99.7% (90% CI: 98.6%-99.98%) in low input samples, 100% (90% CI: 98.97%-100%) in medium input samples, and 92.71% (90% CI: 90.14%-94.77%) in high input samples.

To further assess the limit of detection (LOD) using medium input of 8000 amplifiable copies of DNA, a dilution series was created by spiking sheared Horizon Tru-Q 7 into Tru-Q 0 reference standard to approximate an AF range of 0.06%-1.25% AF. In Laboratory 1, 99.17% mutations were detected at 1%-1.3% AF, 99.63% at 0.5%-0.65% AF, 89.17% at 0.25%-0.33% AF, 69.26% at 0.13%-0.16% AF and 37.41% at 0.06%-0.08% AF. Comparable results were seen in Laboratory 2 (Table 2, S2 Fig, S4 Table, S5 Table) and for all samples (S6 Table).

Across the two laboratories using 8000 amplifiable copies of input DNA, 98.65% of mutations were detected at 0.5%-0.65% AF. For the 0.25%-0.33% dilution range, the sensitivity was 89.73%. The lowest frequency mutation identified was an *EGFR* indel (ΔE746-A750) detected at 0.02% AF.

Assessment of quantitative performance of eTAm-Seq technology by comparison with digital PCR analysis of reference cell-line DNA carrying mutations at known allele fraction

In order to determine the quantitative performance of the eTAm-Seq technology, data was compared with allele fractions generated by digital PCR analysis of Horizon Tru-Q 6 and Tru-Q 7, supplied by the manufacturer. As can be seen in Fig 3A and Fig 3B, there is good concordance between AFs determined by the eTAm-Seq technology and digital PCR analysis of 21 mutations present in both the InVision liquid biopsy tumor profiling panel and Tru-Q 6, and analysis of 38 common mutations in Tru-Q 7. This demonstrates the quantitative accuracy of eTAm-Seq technology for reliable detection of mutations at low allele frequency.

Assessment of specificity of the eTAm-Seq technology by analysis of plasma from presumed healthy donors

Tru-Q6 or Tru-Q7 reference DNA contains additional mutations outside of the validated mutations listed in these cell-line mixes, and is therefore not suitable for assessing specificity. Plasma samples from 79 presumed healthy donors were therefore analysed using eTAm-Seq technology to assess specificity. This analysis identified five low frequency coding mutations, all at ≤0.5% AF: three located in *TP53* [L308L at 0.19% AF (Laboratory 1); Y220C at 0.5% AF and P27L at 0.5% AF (Laboratory 2)] and two in *GATA3* [T323T at 0.1% AF and T419T at 0.317% AF (Laboratory 2)]. Sufficient material was available to enable re-extraction of plasma cfDNA from the same blood draw in four out of five cases (all but *GATA3* T419T). Analysis by eTAm-Seq technology was repeated for these 4 samples. Re-analysis confirmed the initial call for three of the four samples, failing only to detect the *TP53* L308L change originally identified at 0.19% AF. This resulted in two potential false positives, one is unconfirmed, and the other may be a false-negative of the replicate assay at ≤0.19% AF. The identification of 2 potential false positives in 79 healthy samples amounts to a per-base specificity of at least 99.9997% (95% confidence interval, 99.9989% to 99.99996% per-base specificity).

Analysis of novel full-process control material using eTAm-Seq technology and digital PCR

To explore the performance of a novel full-process control with spiked DNA reference standards and assess the ability of the eTAm-Seq technology to identify low frequency mutations,

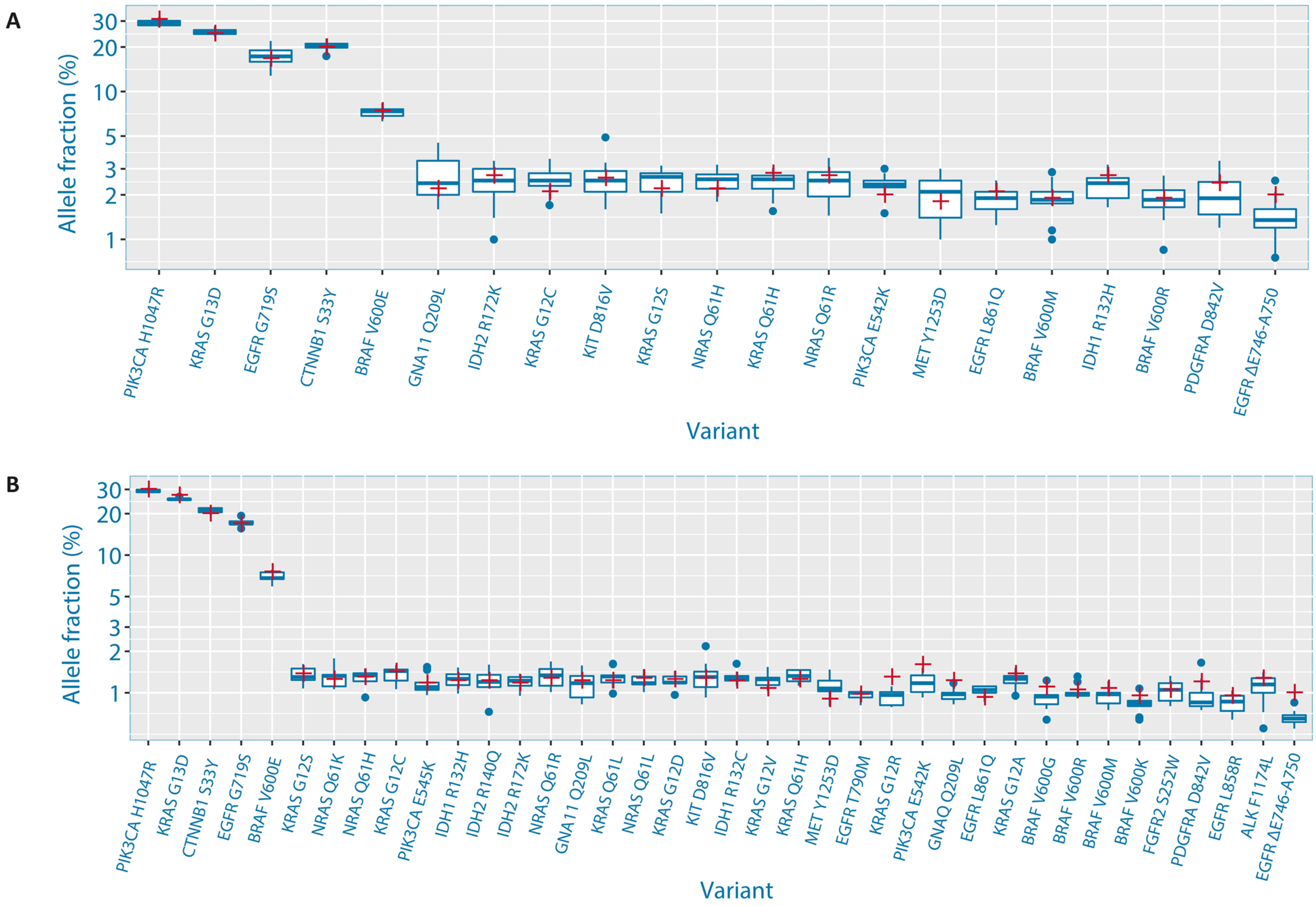


Fig 3. Plot showing allele fractions determined by analysis with eTAm-Seq technology (blue boxplot) and digital PCR (red cross) for analysis of mutations present in both the InVision liquid biopsy tumor profiling panel and (A) Tru-Q 6 and (B) Tru-Q 7.

https://doi.org/10.1371/journal.pone.0194630.g003

5 mL aliquots of pooled plasma from 6 male and 10 female presumed healthy donors were spiked with 400ng Multiplex I cfDNA Reference Standard. This reference standard, acoustically sheared to 160bp to mimic cfDNA, is derived from well-characterized isogenic cell-lines and contains 8 target mutations at ~5%, ~1% or ~0.1% AF. 100% wild-type DNA from non-modified cell-lines containing 100% wild-type DNA was used as a control. By spiking into plasma containing background DNA, the resulting mix would be expected to contain lower AFs than the original standards. For each of the four levels, replicate cfDNA extractions (n = 6) were performed over 3 days, together with replicate plasma-only controls. The cfDNA was sub-divided into two for analysis by both eTAm-Seq technology (Laboratory 1) and dPCR (LGC). dPCR analysis was performed targeting hotspot mutations in *EGFR* L858R, *KRAS* G12D, *NRAS* A59T and *PIK3CA* E545K. The observed extraction efficiency was highly reproducible between replicates, with ~50% recovery of the spike-in (S3 Fig). More variability was observed in the 0.1% AF-spiked sample, likely due to sampling noise when quantifying small numbers of mutant molecules. The extraction efficiency was slightly lower than has previously been reported (60%-80% recovery) for measurement of a spike-in control [30]. Quantification of the <5% AF and <1% AF samples using the eTAm-Seq technology and dPCR spiked

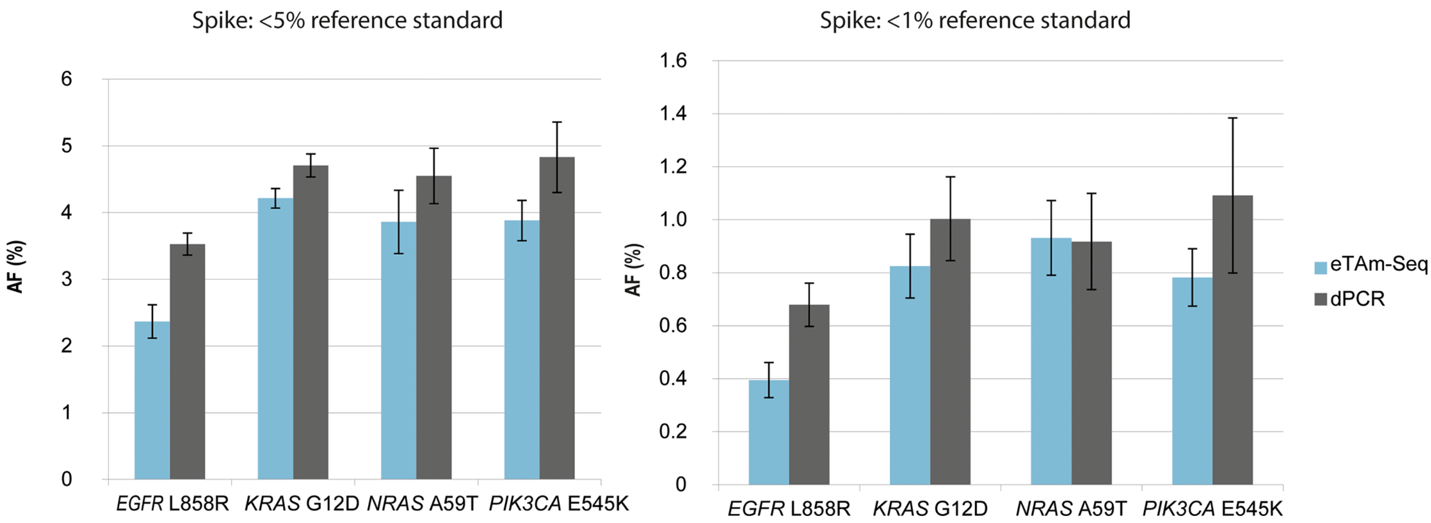


Fig 4. Quantitative agreement of 5% AF and 1% AF reference standard spiked into plasma, and measured by eTAm-Seq technology and dPCR. Mean mutant AF (%) ± SD are displayed for each technology (n = 5* (5% AF standard); n = 6 (1% AF standard)). By spiking into plasma containing background wild-type DNA, the resulting mix was confirmed to contain lower AFs than the original reference standards (original mutant AF values 5% standard: 5% (*EGFR*); 6.3% (*KRAS*, *NRAS*, *PIK3CA*); 1% standard: 1% (*EGFR*), 1.3% (*KRAS*, *NRAS*, *PIK3CA*). (*1 data point omitted due to anomalous extraction efficiency).

https://doi.org/10.1371/journal.pone.0194630.g004

plasma showed good concordance (Fig 4). The small deviation in AF observed in the contrived control samples may be related to differences in DNA fragment sizes between the sheared mutant DNA and the wild-type donor plasma, which may differentially affect the results of the two methods. Both methods showed good precision with low %CV for all 4 mutations in analysis of plasma spiked with 5% and 1% AF reference standard (S4 Fig).

100% of mutations known to be present in the <5% and <1% AF pool were detected by both eTAm-Seq technology and dPCR (S5 Fig, S7 Table). An additional 5 coding mutations in *BRAF* V600E (20.07% AF), *CTNNB1* S33Y (14.72% AF), *PIK3CA* H1047R (13.98% AF), *STK11* Q123Q (13.37% AF) and *EGFR* G719S (13.31% AF) were detected using the eTAm-Seq technology. These mutations were all confirmed to be present by exome sequence analysis of the original isogenic cell-lines that the reference standards were derived from. In the <0.1% spiked sample, across the 4 overlapping mutations analysed by both methods in the 6 replicate extractions, amplicon sequencing detected 11/24 mutations whilst dPCR detected 16/24. Overall, 25/48 (53%) mutations in the <0.1% AF sample were detected using eTAm-Seq technology, as expected given the limit of detection for the assay and stochastic sampling effects. Two potential false positives were identified: *TP53* F113V (GRCh38 chr17:7676032 A>C) at 0.15% AF and *GNA11* R214M (chr19:3118959 G>T) at 0.1% AF. The *GNA11* mutation was possibly caused by 8-oxoguanine (8-oxoG) lesions created during the shearing process used to create the original reference standard DNA. Many of the samples with spiked fragmented DNA had high background at this position and at other G bases, whilst non-sheared plasma did not show an aberrant base both in this run and in previous experiments. The *TP53* mutation was observed significantly above normal background, and may be a false positive or a true low frequency variant.

Discussion

It has long been known that genomic alterations in cancer can be detected in the plasma of cancer patients in the form of circulating tumor DNA. Increasing evidence indicates clinical utility of ctDNA as a diagnostic, prognostic and predictive tool with potential application throughout the continuum of cancer care. FDA approval of the first companion diagnostic permitting ctDNA-based mutation detection, and the emergence of several ctDNA-guided clinical trials [31–33] signals growing acceptance of its utility. ctDNA analysis offers important advantages over profiling single biopsies taken during invasive surgery, often many months or years before clinical progression. ctDNA enables repeat sampling and molecular assessment of tumor evolution during patient treatment, which may help guide subsequent therapy [5, 15]. Advances in NGS have shown it is possible to monitor tumor dynamics and assess evolution in plasma by analysis of multiple mutations in parallel across serially-collected samples, rather than focusing on single hotspot mutations. Digital PCR analysis of multiple mutations is possible to a limited degree but requires sub-dividing DNA into different assays. When large amounts of DNA are available, this can be achieved but where DNA is limited, such as in the analysis of cfDNA, this results in sampling noise and loss of sensitivity as rare mutant molecules are missed. NGS analysis, with a sensitive and appropriately validated platform, circumvents these issues, providing substantially more information on somatic alterations present in the bloodstream, which can be used to guide subsequent cancer therapy.

Currently, there is a limited but growing number of clinically actionable gene targets. The hope is that future advances will result in the development of new immunotherapies and targeted treatments effective against additional somatic alterations known to be present. One important factor in the development of a clinically useful ctDNA assay is to strike the right balance in the size of the genomic region analysed to enable optimal test sensitivity and specificity. By increasing the size of the genomic region covered, the correction for false positives needs to be more stringent. Hybrid capture-based enrichment methods have enabled analysis of focused genomic regions up to whole exomes [6, 23, 24]. However, analysis of larger regions either requires expensive high depth sequencing to identify low frequency mutations, or a compromise on depth and associated reduction in sensitivity. Hybrid capture can be used to target more focused regions but this leads to a high proportion of off-target sequencing reads. Like the sampling noise challenge for dPCR described above, a key limit for all NGS methods developed for cfDNA analysis is the fraction of DNA molecules successfully analysed. Through PCR enrichment, with suitably short amplicons, amplicon-based sequencing can achieve sensitivity comparable to dPCR by amplifying and thereby sampling the majority of cfDNA molecules accessible to PCR amplification. Since samples need not be split into multiple assays, the effective sensitivity of amplicon-based sequencing may even exceed that of dPCR [20, 34, 35]. Hybrid capture library preparation methods are not restricted to amplifying regions containing both priming sites. However, they require considerable pre-processing prior to enrichment or PCR-based amplification and therefore may lose a significant proportion of molecules during library preparation stages, particularly during adaptor ligation [36]. This is important for analysis of ctDNA given the low frequency of tumor-derived DNA molecules present in patient plasma, particularly in earlier stage cancer.

Here, we have described the InVision liquid biopsy platform which utilizes enhanced TAm-Seq technology for the identification of low frequency mutations in cell-free DNA. This amplicon-based method has been carefully optimized for efficient amplification from limited amounts of fragmented plasma DNA. The focused gene panel targets 35 clinically actionable and clinically-relevant genes, providing coverage of critical regions in 31 genes and near complete coverage of 4 genes of clinical significance. Analytical validation of the assay demonstrates high sensitivity and specificity for detection of low frequency mutations with 94.08% of mutations detected at 0.25% - 0.33% allele fraction (AF) with optimal DNA input, with a per-base specificity of 99.9997%. Validation across two laboratories demonstrates its reproducibility and supports its use in clinical applications. In addition, the assay is highly quantitative, demonstrating excellent concordance with digital PCR analysis of commercial cell-line reference standard DNA, and novel full-process control material developed by LGC and Horizon Discovery, that carry cancer-related mutations at known allele fractions.

Using this assay, mutant alleles were detected down to 0.02% AF, with >30% sensitivity for detection at 0.06% AF (1 mutant DNA copy in 1600 molecules). This study identifies two challenges when assessing assay specificity using either individual donors or acoustically-sheared commercial reference standards for analysis of low frequency mutations in ctDNA using an ultra-sensitive test. During analysis, five mutations at ≤0.5% AF were identified in presumed healthy donors, yet when 4 samples with sufficient material were re-analysed, the same mutations were repeatedly identified in 3, indicating these were true positives; the 4th change was originally detected at a low allele fraction of 0.19% so possibly missed on repeat due to its low allele fraction. Somatic mutations have previously been detected in presumed healthy individuals, and may represent pre-malignant mutations that accumulate prior to cancer or during the aging process [37], or changes that have arisen during clonal hematopoiesis [37, 38] or could originate from undetected tumors. More studies are needed using orthogonal assays with similarly high sensitivity to determine if changes are truly present. Using the commercial reference standard, a *GNA11* R214M G>T mutation was identified along with a signature of high background G>T/C>A errors at other bases. This is consistent with 8-oxoguanine (8-oxoG) lesions created during the acoustic shearing process or potentially evolution and heterogeneity of the cell lines. The phenomenon of G>T/C>A transversion artifacts was first identified by Costello et al. [39, 40], and highlights the potential risk of using acoustically-sheared DNA to validate specificity of sensitive ctDNA NGS-based assays capable of detecting low frequency sequence aberrations. One solution may be to limit the use of sheared material to the assessment of assay sensitivity, since this material performed well at the loci that were defined and tested for this purpose, and use donor (healthy volunteer) DNA for broader specificity assessment (given the caveats previously mentioned and repeat or orthogonal analysis for confirmation). Alternatively, different mechanisms could be investigated to fragment commercial reference standard DNA, such as enzymatic fragmentation, which may potentially introduce less DNA damage.

Given the restrictive requirement to immediately process EDTA-collected blood to plasma to prevent leukocyte lysis, it is important to validate the eTAm-Seq technology using blood collected into Streck Cell-free DNA BCT tubes. These tubes contain a proprietary cell preservative which stabilizes nucleated blood cells preventing contamination with background wild-type DNA. Analysis of the eTAm-Seq technology in EDTA and Streck tubes collected at the same time from patients has previously been presented [40], and showed high technical reproducibility between two independently processed blood tube types, indicating use of either tube type is suitable for clinical blood collection using this technology. The Streck Cell-free DNA BCT tubes provide a robust alternative to enable delayed and centralized processing which will help standardize pre-analytic factors during blood collection, and provides improved feasibility for introduction into routine ctDNA testing in the clinic.

In support of the use of the InVision liquid biopsy platform in clinical applications, data has previously been reported demonstrating a high level of concordance between this platform and dPCR in mutations detected in 35 patients with advanced breast cancer [41]. With 100% and 96% agreement for mutation detection in *ESR1* and *PIK3CA* respectively, amplicon sequencing identified additional mutations not covered by dPCR analysis and therefore substantially more mutations per patient which have possible clinical relevance. There was 100% concordance in the detection of *HER2* amplifications when compared to IHC and/or FISH of metastatic tumors [42]. Furthermore, Fribbens et al. demonstrated good concordance of eTAm-Seq technology with dPCR, with high levels of genetic heterogeneity and frequent subclonal mutations in advanced breast cancer patients progressing on first-line aromatase inhibitor therapy [42]. In this study, *ESR1* mutations were detectable in plasma median of 6.7 months before clinical progression. Another study compared the detection of mutations in *EGFR* between plasma and tissue and across platforms, and found amplicon-based plasma NGS to have exquisite sensitivity and specificity, with excellent quantitative concordance with an optimized dPCR assay [43]. Remon et al. have previously demonstrated the use of eTAm-Seq technology to aid in selection of targeted treatment in a prospective cohort of 48 *EGFR*-mutant advanced NSCLC patients with acquired resistance to *EGFR* TKIs, and without an available tissue biopsy [15]. cfDNA analysis identified resistance mutations in *EGFR* T790M at frequencies as low as 0.1% AF, and the study was able to demonstrate the benefit of osimertinib treatment in these patients. Strikingly, of the seven cases in that study with best response (decrease of 50% or more in size), three cases had T790M detected at <0.25% AF. Use of a less sensitive assay would miss such low frequency alleles.

Taken together, these studies demonstrate that the InVision liquid biopsy platform is a highly sensitive, quantitative and reproducible platform for detection of low frequency clinically-relevant cancer mutations in cell-free DNA. Additional larger cohorts are currently being analyzed to support clinical validation and clinical utility of the test and provide evidence to support introduction into routine testing for patient management.

Supporting information

S1 Fig. A-E dPCR data plots for analysis of full-process control samples. (A-D) dPCR 2D data plots for A. *EGFR* L858R/WT, B. *KRAS* G12D/WT, C. *NRAS* A59T/WT and D. *PIK3CA* E545K/WT assays showing negative controls (plasma only, NTCs), positive controls (non-spiked Multiplex I cfDNA Reference Standards) and analysis of full-process controls (plasma spiked with Multiplex I cfDNA Reference Standards). Data was analysed using QuantaLife (Bio-Rad, version 1.6.6.0320) with classification of single positive (mutant (blue), wild-type (green)), double positive (orange) and negative (black) droplets by manual crosshair setting (A, B, D) or lasso (C; lassos not shown by software post-analysis). E. Accepted droplet number for all four dPCR assays.
(EPS)

S2 Fig. Sensitivity of the eTAm-Seq technology.
(AI)

S3 Fig. Evaluation of extraction efficiency using full-process control samples. Copy number concentrations of each target in spiked extracts were quantified by dPCR and extraction efficiency calculated by comparison with the values assigned by dPCR using the same assay.
(AI)

S4 Fig. Precision of eTAm-Seq technology and dPCR measurements of full-process control materials. Precision is expressed as % coefficient of variation (%CV) for measurements of full-process control materials spiked with 5% and 1% AF Multiplex I cfDNA Reference Standards (n = 5 (5% Reference Standard); n = 6 (1% Reference Standard) and reflects both extraction and analytical variability.
(AI)

S5 Fig. Analytical sensitivity of eTAm-Seq technology and dPCR measurements of full-process control materials. Detection rate of mutant targets by eTAm-Seq technology and dPCR in full-process control materials spiked with 5%, 1% and 0.1% AF Multiplex I cfDNA Reference Standards expressed as number of positive measurements (n = 6).
(AI)

S1 File. dMIQE checklist.
(PDF)

S2 File. dPCR assay validation for dMIQE review.
(PDF)

S1 Table. Exonic regions covered by the InVision liquid biopsy tumor profiling panel.
(XLSX)

S2 Table. A-C. A. dPCR assay information B. dPCR assay information (custom design) C. PCR cycling conditions.
(DOCX)

S3 Table. LOD of dPCR assays.
(DOCX)

S4 Table. Table of allele fractions (shown as percentages) for the 32 common low frequency variants validated/quantified in Horizon Tru-Q 7 and in the InVision liquid biopsy panel for both laboratories (one tab per laboratory). Rows list all combination of variants (32 variants) and dilution levels (5 levels). Columns show the operator/repeats. The last column indicates the expected allele fraction based on dPCR data supplied by Horizon Discovery, and the dilution level. Missing values indicate that no call was made for that combination of variant/dilution/operator/repeat.
(XLSX)

S5 Table. List of variants analyzed in the Horizon Tru-Q 6 and Horizon Tru-Q 7 dilution study.
(XLSX)

S6 Table. Sensitivity analysis for all samples.
(XLSX)

S7 Table. Table of all reportable calls by eTAm-Seq technology using the novel full-process control material.
(XLSX)

Acknowledgments

We would like to thank management and staff at the Cancer Research UK Cambridge Institute (CRUK-CI), Cancer Research Technologies, CRUK, Cambridge Enterprise and the University of Cambridge. We would like to thank Femke de Snoo (Medex15) for medical writing assistance, funded by Inivata. We would like to thank Philip Wilson and Simon Cowen (LGC) for statistical advice.

Author Contributions

Conceptualization: Davina Gale, Andrew R. J. Lawson, Tim Forshew, Vincent Plagnol, Nitzan Rosenfeld.

Data curation: Davina Gale, Andrew R. J. Lawson, Michael Epstein, Tim Forshew, Vincent Plagnol.

Formal analysis: Peter Dimitrov, Michael Epstein, Vincent Plagnol.

Funding acquisition: Carole A. Foy, Karin Schmitt, Nitzan Rosenfeld.

Investigation: Davina Gale, Andrew R. J. Lawson, Karen Howarth, Mikidache Madi, Bradley Durham, Sarah Smalley, John Calaway, Shannon Blais, Samuel Woodhouse, Ana Fernandez-Gonzalez, Alexandra S. Whale, Gerwyn M. Jones, Alison Devonshire.

Methodology: Davina Gale, Andrew R. J. Lawson, Sarah Smalley, Ana Fernandez-Gonzalez, Alexandra S. Whale, Alison Devonshire, Tim Forshew, Vincent Plagnol, Nitzan Rosenfeld.

Project administration: Davina Gale, Karin Schmitt, Alison Devonshire.

Resources: Hadas Raveh-Amit, Karin Schmitt.

Software: Vincent Plagnol.

Supervision: Davina Gale, Greg Jones, James Clark, Jim F. Huggett, Carole A. Foy, Tim Forshew, Vincent Plagnol, Nitzan Rosenfeld.

Validation: Davina Gale, Andrew R. J. Lawson, Karen Howarth, Sarah Smalley, John Calaway, Shannon Blais, Michelle Pugh, Emma Green, Tim Forshew, Vincent Plagnol.

Visualization: Alison Devonshire.

Writing – original draft: Davina Gale.

Writing – review & editing: Davina Gale, Andrew R. J. Lawson, Karen Howarth, Michelle Pugh, Samuel Woodhouse, Carole A. Foy, Alison Devonshire, Emma Green, Tim Forshew, Vincent Plagnol, Nitzan Rosenfeld.

References

1. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. Clin Chim Acta. 2001 Nov; 313(1–2):139–42. PMID: 11694251
2. Sidransky D, Von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, et al. Identification of p53 gene mutations in bladder cancers and urine samples. Science. 1991 May 3; 252(5006):706–9. PMID: 2024123
3. Wan JC, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer. 2017 Apr; 17(4):223–238. https://doi.org/10.1038/nrc.2017.7 PMID: 28233803
4. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008; 14(9):985–90. https://doi.org/10.1038/nm.1789 PMID: 18670422
5. Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med. 2013 Mar 28; 368(13):1199–209. https://doi.org/10.1056/NEJMoa1213261 PMID: 23484797
6. Murtaza M, Dawson SJ, Tsui DW, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013 May 2; 497 (7447):108–12. https://doi.org/10.1038/nature12065 PMID: 23563269
7. Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, et al. Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. PLoS Med. 2016 Dec 20; 13(12): e1002198. https://doi.org/10.1371/journal.pmed.1002198 PMID: 27997533
8. Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N Engl J Med. 2009 Sep 3; 361(10):947–57. https://doi.org/10.1056/NEJMoa0810699 PMID: 19692680
9. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012 Mar; 13(3):239–46. Epub 2012 Jan 26. https://doi.org/10.1016/S1470-2045(11)70393-X PMID: 22285168
10. Cappuzzo F, Ciuleanu T, Stelmakh L, Cicenas S, Szczésna A, Juhász E, et al. Erlotinib as maintenance treatment in advanced non-small-cell lung cancer: a multicentre, randomised, placebo-controlled phase

3 study. Lancet Oncol. 2010 Jun; 11(6):521–9. Epub 2010 May 20. https://doi.org/10.1016/S1470-2045(10)70112-1 PMID: 20493771

11. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, et al. Safety and efficacy of vemurafenib in BRAF(V600E) and BRAF(V600K) mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study. Lancet Oncol. 2014 Mar; 15(3):323–32. Epub 2014 Feb 7. https://doi.org/10.1016/S1470-2045(14)70012-9 PMID: 24508103
12. Diaz LA Jr, Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature. 2012 Jun 28; 486(7404):537–40. https://doi.org/10.1038/nature11219 PMID: 22722843
13. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature. 2012 Jun 28; 486(7404):532–6. https://doi.org/10.1038/nature11156 PMID: 22722830
14. Mok TS, Wu Y-L, Ahn M-J, Garassino MC, Kim HR, Ramalingam SS, et al. Osimertinib or Platinum-Pemetrexed in EGFR T790M-Positive Lung Cancer. N Engl J Med. 2017 Feb 16; 376(7):629–640. https://doi.org/10.1056/NEJMoa1612674 PMID: 27959700
15. Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, et al. Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. Ann Oncol. 2017 Apr 1; 28(4):784–790. https://doi.org/10.1093/annonc/mdx017 PMID: 28104619
16. Garcia-Murillas I, Schiavon G, Weigelt B, Ng C, Hrebien S, Cutts RJ, Cheang M, et al. Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. Sci Transl Med. 2015 Aug 26; 7(302):302ra133. https://doi.org/10.1126/scitranslmed.aab0021 PMID: 26311728
17. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature. 2017 Apr 26; 545(7655):446–451. https://doi.org/10.1038/nature22364 PMID: 28445469
18. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies.; Sci Transl Med. 2014 Feb 19; 6(224):224ra24. https://doi.org/10.1126/scitranslmed.3007094 PMID: 24553385
19. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999 Aug 3; 96(16):9236–41. PMID: 10430926
20. US Food & Drug Administration. Premarket approval P150044, Cobas EGFR mutation test V2. 2017 FDA Available from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P150044
21. US Food & Drug Administration. Device Approvals and Clearances. Cobas® EGFR Mutation Test v2—P150047. Patient Labeling. Table 22. Available from https://www.accessdata.fda.gov/cdrh_docs/pdf15/P150044C.pdf
22. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med. 2012 May 30; 4(136):136ra68. https://doi.org/10.1126/scitranslmed.3003726 PMID: 22649089
23. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014 May; 20(5):548–54 https://doi.org/10.1038/nm.3519 PMID: 24705333
24. Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and Clinical Validation of a Digital Sequencing Panel for Quantitative, Highly Accurate Evaluation of Cell-Free Circulating Tumor DNA. PLoS One. 2015 Oct 16; 10(10):e0140712. https://doi.org/10.1371/journal.pone.0140712 PMID: 26474073
25. Gale D, Plagnol V, Lawson A, Pugh M, Smalley S, Howarth K, et al. Cancer Research, AACR 2016, Abstract 3639: Analytical performance and validation of an enhanced TAm-Seq circulating tumor DNA sequencing assay; https://doi.org/10.1158/1538-7445.AM2016-3639 Published July 2016
26. Rosenfeld N., Forshew, T., Marass, F. & Murtaza, M. A method for detecting a genetic variant. World Intellectual Property Organization patent WO2016009224A1 (2016).
27. McLaren W, Gil L, Hunt SE, Singh Riat H, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016 Jun 6; 17(1):122. https://doi.org/10.1186/s13059-016-0974-4 PMID: 27268795
28. Whale AS, Bushell CA, Grant PR, Cowen S, Gutierrez-Aguirre I, O'Sullivan DM, et al. Detection of Rare Drug Resistance Mutations by Digital PCR in a Human Influenza A Virus Model System and Clinical Samples. J Clin Microbiol. 2016 Feb; 54(2):392–400. Epub 2015 Dec 9. https://doi.org/10.1128/JCM.02611-15 PMID: 26659206
29. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med. 2017 Aug 16; 9(403). pii: eaan2415. https://doi.org/10.1126/scitranslmed.aan2415

30. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal Bioanal Chem. 2014 Oct; 406(26):6499–512. https://doi.org/10.1007/s00216-014-7835-3 PMID: 24853859

31. 209.US National Library of Medicine. ClinicalTrials.gov Available from: https://clinicaltrials.gov/ct2/show/NCT02284633 (2015).

32. 210, US National Library of Medicine. ClinicalTrials.gov Available from: https://clinicaltrials.gov/ct2/show/NCT02743910 (2016).

33. 211. International Standard Randomised Controlled Trials Number Registry. isrctn.com Available from: http://www.isrctn.com/ISRCTN16945804 (2016).

34. Oxnard GR, Paweletz CP, Kuang Y, Mach SL, O'Connell A, Messineo MM, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. Clin Cancer Res. 2014 Mar 15; 20(6):1698–1705. Epub 2014 Jan 15. https://doi.org/10.1158/1078-0432.CCR-13-2482 PMID: 24429876

35. Hrebien S, O'Leary B, Beaney M, Schiavon G, Fribbens C, Bhambra A, et al. Reproducibility of Digital PCR Assays for Circulating Tumor DNA Analysis in Advanced Breast Cancer. PLoS One. 2016 Oct 19; 11(10): e0165023. https://doi.org/10.1371/journal.pone.0165023 PMID: 27760227

36. Aigrain L, Gu Y, Quail MA. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays—a systematic comparison of DNA library preparation kits for Illumina sequencing. BMC Genomics. 2016 Jun 13; 17:458. https://doi.org/10.1186/s12864-016-2757-4 PMID: 27297323

37. Krimmel JD, Schmitt MW, Harrell MI, Agnew KJ, Kennedy SR, Emond MJ, et al. Ultra-deep sequencing detects ovarian cancer cells in peritoneal fluid and reveals somatic TP53 mutations in noncancerous tissues. Proc Natl Acad Sci U S A. 2016 May 24; 113(21):6005–10. Epub 2016 May 5. https://doi.org/10.1073/pnas.1601311113 PMID: 27152024

38. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014 Dec 25; 371 (26):2477–87. Epub 2014 Nov 26. https://doi.org/10.1056/NEJMoa1409405 PMID: 25426838

39. Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013 Apr 1; 41(6): e67. https://doi.org/10.1093/nar/gks1443 PMID: 23303777

40. Park G, Park JK, Shin SH, Jeon HJ, Kim NKD, Kim YJ, et al. Characterization of background noise in capture-based targeted sequencing data. Genome Biol. 2017 Jul 21; 18(1):136. https://doi.org/10.1186/s13059-017-1275-2 PMID: 28732520

41. Garcia-Murillas I, Beaney M, Epstein M, Howarth K, Lawson A, Hrebien S, et al. Comparison of enhanced Tagged-Amplicon Sequencing and digital PCR for circulating tumor DNA analysis in advanced breast cancer. Cancer Res 2017; 77(13 Suppl): Abstract nr 2743. https://doi.org/10.1158/1538-7445.AM2017-2743 http://www.inivata.com/wp-content/uploads/2017/04/AACR-2017-Inivata-FINAL.pdf (AACR 2017).

42. Fribbens C, Garcia Murillas I, Beaney M, Hrebien S, O'Leary B, Kilburn L, et al. Tracking evolution of aromatase inhibitor resistance with circulating tumour DNA analysis in metastatic breast cancer. Ann Oncol. 2017 Oct 4. https://doi.org/10.1093/annonc/mdx483.

43. Guibert NM, Paweletz C, Hu Y, Feeney NB, Plagnol V, Poole V, et al. J Clin Oncol 35, 2017 (suppl; abstr 11529). https://doi.org/10.1200/JCO.2017.35.15_suppl.11529 Journal of Clinical Oncology 35, no. 15_suppl (May 2017) 11529–11529. https://www.inivata.com/wp-content/uploads/2017/06/Guibert_ASCO.pdf (ASCO 2017).

RESEARCH ARTICLE

Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling

Vincent Plagnol[1☯], **Samuel Woodhouse**[1☯], **Karen Howarth**[1], **Stefanie Lensing**[1], **Matt Smith**[1], **Michael Epstein**[1], **Mikidache Madi**[1], **Sarah Smalley**[2], **Catherine Leroy**[1], **Jonathan Hinton**[1], **Frank de Kievit**[1], **Esther Musgrave-Brown**[1], **Colin Herd**[1], **Katherine Baker-Neblett**[3], **Will Brennan**[2], **Peter Dimitrov**[1], **Nathan Campbell**[4], **Clive Morris**[3], **Nitzan Rosenfeld**[1], **James Clark**[1], **Davina Gale**[1], **Jamie Platt**[2], **John Calaway**[2], **Greg Jones**[2], **Tim Forshew**[1]*

1 Research and Development, Inivata Ltd, Granta Park, Cambridge, United Kingdom, **2** Product Development, Inivata Inc, Research Triangle Park, North Carolina, United States of America, **3** Clinical Development, Inivata Inc, Research Triangle Park, North Carolina, United States of America, **4** Clinical Laboratory Operations, Inivata Inc, Research Triangle Park, North Carolina, United States of America

☯ These authors contributed equally to this work.
* tim.forshew@inivata.com

Check for updates

OPEN ACCESS

Citation: Plagnol V, Woodhouse S, Howarth K, Lensing S, Smith M, Epstein M, et al. (2018) Analytical validation of a next generation sequencing liquid biopsy assay for high sensitivity broad molecular profiling. PLoS ONE 13(3): e0193802. https://doi.org/10.1371/journal.pone.0193802

Editor: Alvaro Galli, CNR, ITALY

Received: November 9, 2017

Accepted: February 20, 2018

Published: March 15, 2018

Copyright: © 2018 Plagnol et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files including complete lists of all detected mutation in this study.

Funding: These studies were funded by Inivata Ltd, and Inivata Inc. The funder provided support in the form of salaries for authors VP, SW, KH, SL, MS, ME, MM, SS, CL, JH, FDK, EMB, CH, KBN, WB, PD, NC, CM, NR, J. Clark, DG, JP, J. Calaway, GJ & TF, but did not have any additional role in the study design, data collection and analysis, decision to

Abstract

Circulating tumor DNA (ctDNA) analysis is being incorporated into cancer care; notably in profiling patients to guide treatment decisions. Responses to targeted therapies have been observed in patients with actionable mutations detected in plasma DNA at variant allele fractions (VAFs) below 0.5%. Highly sensitive methods are therefore required for optimal clinical use. To enable objective assessment of assay performance, detailed analytical validation is required. We developed the InVisionFirst™ assay, an assay based on enhanced tagged amplicon sequencing (eTAm-Seq™) technology to profile 36 genes commonly mutated in non-small cell lung cancer (NSCLC) and other cancer types for actionable genomic alterations in cell-free DNA. The assay has been developed to detect point mutations, indels, amplifications and gene fusions that commonly occur in NSCLC. For analytical validation, two 10mL blood tubes were collected from NSCLC patients and healthy volunteer donors. In addition, contrived samples were used to represent a wide spectrum of genetic aberrations and VAFs. Samples were analyzed by multiple operators, at different times and using different reagent Lots. Results were compared with digital PCR (dPCR). The InVisionFirst assay demonstrated an excellent limit of detection, with 99.48% sensitivity for SNVs present at VAF range 0.25%-0.33%, 92.46% sensitivity for indels at 0.25% VAF and a high rate of detection at lower frequencies while retaining high specificity (99.9997% per base). The assay also detected *ALK* and *ROS1* gene fusions, and DNA amplifications in *ERBB2*, *FGFR1*, *MET* and *EGFR* with high sensitivity and specificity. Comparison between the InVisionFirst assay and dPCR in a series of cancer patients showed high concordance. This analytical validation demonstrated that the InVisionFirst assay is highly sensitive, specific and robust, and meets analytical requirements for clinical applications.

publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: These studies were funded by Inivata Ltd, and Inivata Inc. All authors are current or former employees and/or officers and/or share-holders of Inivata Ltd or Inivata Inc, the funder of this study. Patents have been filed relating to the technology described in this manuscript (WO2016009224 A1 - A method for detecting a genetic variant) and the study describes the validation of a product developed by Inivata Ltd. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

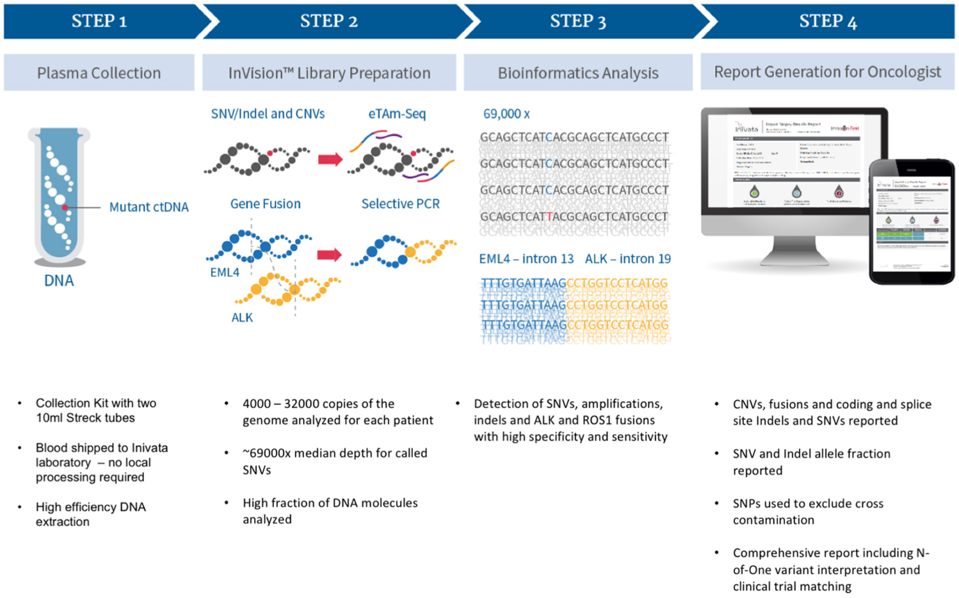
It has been shown more than 20 years ago that some cancer mutations can be detected non-invasively through analysis of samples including blood plasma, urine, stool and sputum [1–6]. Circulating tumor DNA (ctDNA) is believed to enter a patient's blood plasma largely through turnover of cancer cells and subsequent release of the resultant fragmented DNA into circulation. Early attempts to analyze this ctDNA were restricted to methods that focused on a small number of genomic changes with relatively limited sensitivity. It is now known that a significant fraction of mutations, especially in earlier stage cancer are present at extremely low variant allele fractions (VAF) in the blood.

The development of methods including digital PCR (dPCR) and its derivatives such as droplet-based digital PCR (ddPCR) subsequently enabled the sensitive and quantitative analysis of 'hotspot' mutations or individual mutant alleles [7–9]. These more sensitive methods demonstrated the potential of using ctDNA for a range of applications including cancer prognostication, treatment selection, monitoring and even early detection [9–11]. They were still however limited to assessing just a small number of changes.

We demonstrated for the first time in 2012 the ability to use next generation sequencing (NGS) of gene panels to detect solid tumor mutations through sequencing a patient's cell free DNA (cfDNA) and to monitor the VAF of multiple mutations in serially collected plasma samples over time [12]. The initial version of our assay using TAm-Seq® technology covered 6 genes and had 97% sensitivity and specificity for detecting single nucleotide variants (SNVs) and indels at 2% VAF and above and reported mutations down to 0.14% VAF. This demonstration was rapidly followed by examples of a range of different NGS approaches including hybrid capture and molecular barcoding that could be applied to broadly analyze ctDNA with varying performance characteristics [13–15].

The area where ctDNA analysis is most rapidly entering clinical use is in the molecular stratification of patients for treatment where tissue is limited, unavailable or of insufficient quality; most notably for non-small cell lung cancer (NSCLC) patients. This is due to the complexity of a lung biopsy, the risk and associated costs and the availability of appropriate effective targeted agents for treatment of NSCLC patients. The first assays to gain regulatory approval for testing in this setting were the therascreen® EGFR RGQ PCR kit and cobas® *EGFR* Mutation Test v2 assays which use real-time PCR for the qualitative detection of *EGFR* exon 19 deletions, L858R, T790M and other mutations in *EGFR*. Positive ctDNA results can be used to determine which NSCLC patients are eligible for treatment with 1st- or 3rd-generation *EGFR* inhibitors [16]. Due to the technology used however, these assays are less sensitive than dPCR and can only assess a limited number of mutations, reducing the number of patients these assays will successfully stratify to treatment.

To enable broad and highly sensitive ctDNA analysis we have developed eTAm-Seq technology, a significantly enhanced version of our original TAm-Seq technology (Fig 1). We previously described the development of an earlier version of this enhanced assay, which covered 35 genes and could detect SNVs, indels and copy number variations (CNVs) [17]. Here we describe the analytical validation of the InVisionFirst assay which utilizes this technology and has been updated to cover 36 genes for a range of SNVs, indels, CNVs and gene fusion events including the key mutations in *EGFR* and *ALK* and *ROS1* fusions (S1 Table). InVisionFirst is an NGS assay designed to detect the key actionable somatic NSCLC mutations in ctDNA, released into the blood stream of NSCLC patients which, when combined with standard clinical observations, can be used by the clinician to guide a patient to therapy. Based on the previously published NSCLC mutation spectrum, 94% of patients contain at least one mutation within the 36 genes targeted [18].


STEP 1
STEP 2
STEP 3
STEP 4
Plasma Collection
InVision™ Library Preparation
Bioinformatics Analysis
Report Generation for Oncologist
Mutant ctDNA
DNA
SNV/Indel and CNVs
eTAm-Seq
Gene Fusion
Selective PCR
EML4
ALK
69,000 x
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATCACGCAGCTCATGCCCT
GCAGCTCATTACGCAGCTCATGCCCT
EML4 – intron 13
ALK – intron 19
TTTGTGATTAAGCCTGGTCCTCATGG
TTTGTGATTAAGCCTGGTCCTCATGG
TTTGTGATTAAGCCTGGTCCTCATGG

- Collection Kit with two 10ml Streck tubes
- Blood shipped to Invivata laboratory – no local processing required
- High efficiency DNA extraction

- 4000 – 32000 copies of the genome analyzed for each patient
- ~69000x median depth for called SNVs
- High fraction of DNA molecules analyzed

- Detection of SNVs, amplifications, indels and ALK and ROS1 fusions with high specificity and sensitivity

- CNVs, fusions and coding and splice site Indels and SNVs reported
- SNV and Indel allele fraction reported
- SNPs used to exclude cross contamination
- Comprehensive report including N-of-One variant interpretation and clinical trial matching

Fig 1. Overview of the InVisionFirst workflow.

https://doi.org/10.1371/journal.pone.0193802.g001

There are a growing number of ctDNA assays being developed based on different technologies, with different performance characteristics and differing levels of concordance with tissue [15,19]. It is therefore critical that analytical validation studies are executed and performance testing schemes developed to determine the functional characteristics of each assay to enable clinicians to select the most suitable assay for their patient. In the current study, we first describe the assessment of the InVisionFirst assay's ability to call SNVs, indels, amplifications and gene fusions using contrived material and blood from donors not known to have cancer. We then compare our ability to call changes in NSCLC patients' blood with that of digital PCR. Finally, we demonstrate that the InVisionFirst assay gives concordant results whether blood is drawn in Streck Cell-Free DNA Blood collection tubes (Streck BCT) or EDTA tubes and we show, concordant with previously published results, that when DNA containing mutations was spiked into blood drawn into Streck BCT then the mutant allele fraction stayed stable for at least 10 days.

Results

SNV detection sensitivity, repeatability and reproducibility

To determine the ability of the InVisionFirst assay to call mutations at different allele fractions and thus its limit of detection (LoD) which we have defined as the point where we would detect a mutation ≥90% of the time (LoD90), a dilution series was created of sheared Tru-Q7 reference DNA in Tru-Q0 (both obtained from Horizon Discovery). Details of this and subsequent contrived materials are described in detail in the Materials and Methods section. Tru-Q7 contains 39 validated mutations that are covered in our targeted sequencing region, and 32 of these are SNVs present at low VAF, predominantly between 1%-1.3% (S2 Table). The dilution series created 5 different samples, containing respectively the majority of mutations at the following VAF levels: 1%-1.3%, 0.5%-0.65%, 0.25%-0.33%, 0.13%-0.16% and 0.06%-0.08%. Samples from this dilution series was analyzed multiple times by three operators, each using two different Lots of reagents. Full details of this and subsequent designs are available in S3 Table.

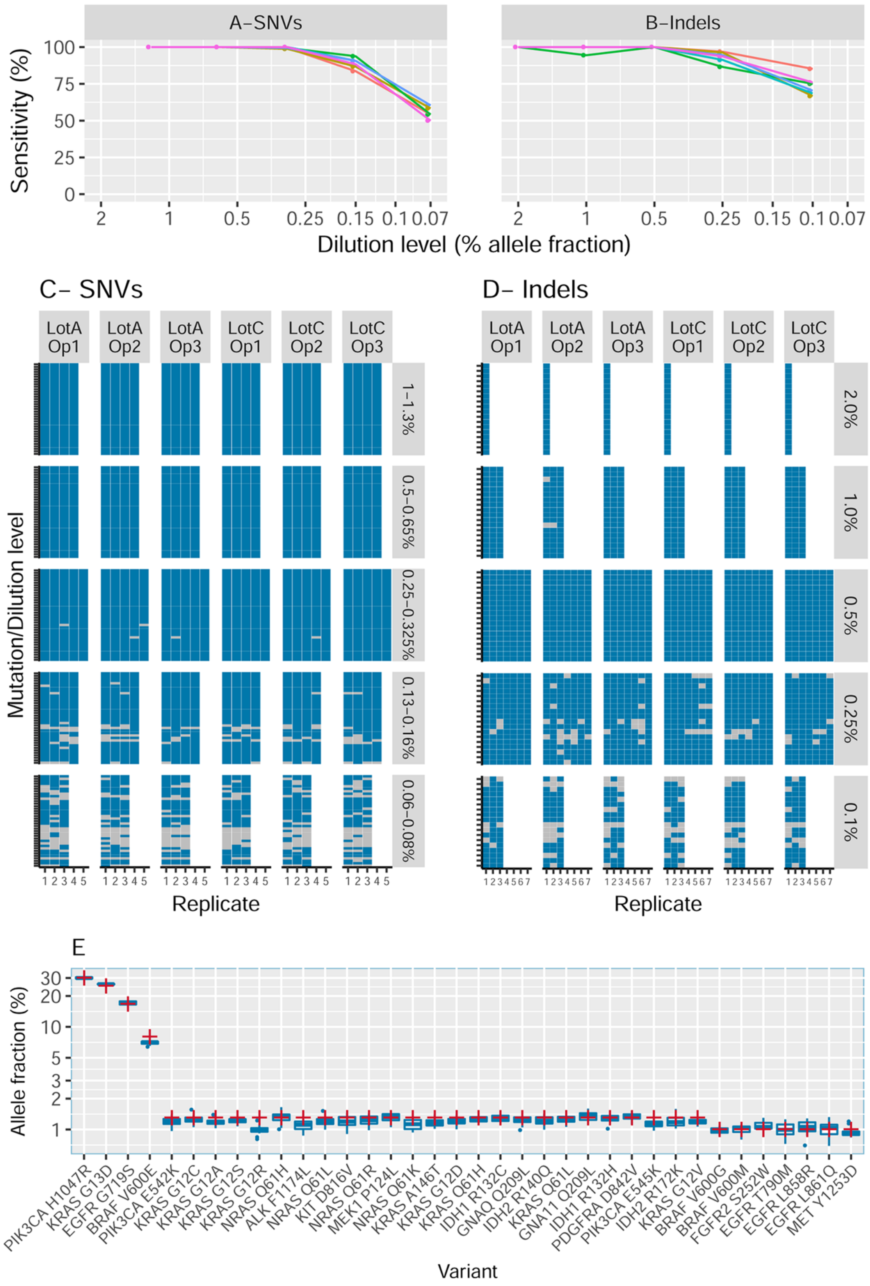


Fig 2. Sensitivity for SNVs (A, C) and indels (B, D). **A** and **B** show the sensitivity as a function of the allele fraction of the reference mutations. Each line represents a different operator/Lot combination. **C** and **D** show the full set of calls for all combinations of dilution/variant (vertical) and repeat/operator/lot (horizontal). Blue rectangles represent mutations that were detected and grey represents those missed. Panel E shows for SNVs the estimated allele fraction compared between InVisionFirst (blue box-plots) and the reference as estimated by Horizon using ddPCR (red crosses).

https://doi.org/10.1371/journal.pone.0193802.g002

100% of SNVs with an expected VAF of 0.5% and above were detected across all runs, all operators and all reagent Lots. For SNVs at VAF in the range 0.25%-0.33%, 99.48% were detected. 88.93% of SNVs were detected at the VAF range of 0.13%-0.16%, and 56.25% were detected at the VAF range 0.06%-0.08% (Fig 2A, S4 Table). This confirmed the LoD of our assay to be 0.25% VAF. A complete table describing all expected calls and whether they were made, along with depth of coverage is available (S5 Table). In total, there were 3,498 mutation calls made at low VAF ($\leq$1.3%) in this SNV section of the study. The median depth of sequencing for all the detected mutations was 69,061x and the lowest depth at which any of the 3,498 mutation calls were made was 10,223x. Coverage was extremely even, with 95% of calls having a depth no less than half the median and 99.9% of calls having a depth no less than 0.2x of the median.

Mutation calls between replicates, operators and reagent Lots showed high repeatability and reproducibility with all 32 mutations detected in all replicates at 0.5% VAF and above and no noticeable difference within or between operators or Lots at 0.25%-0.33% VAF (Fig 2C). To extend our analysis across a broader set of mutations we assessed a total of 43 unique samples containing a total of 605 unique variants at or above the LoD (S6 Table). We detected all SNVs giving a Positive percentage agreement (PPA) of 100.0% at $\geq$0.25% VAF.

Combining the replicates of the InVisionFirst assay, the average estimated VAF for the 36 validated SNVs closely correlated with the expected frequencies as stated by Horizon for the undiluted Tru-Q7 DNA (Pearson squared correlation coefficient $R^2 = 0.9987$) (Fig 2E).

Indel detection sensitivity, repeatability and reproducibility

To assess the InVisionFirst assay's ability to call indels, a custom reference material was created by SeraCare containing eighteen indels targeted by our panel ranging from -24bp to +12bp (S7 Table). Five separate samples were produced by SeraCare with all eighteen indels present at one of five different levels; 2%, 1%, 0.5%, 0.25% or 0.1% VAF. All five of these samples were analyzed multiple times by three operators each using two different Lots of reagents (S3 and S7 Tables).

For the 2%, 1% and 0.5% VAF all but 3 of the 1188 expected indels were detected (99.7%). At a VAF of 0.25%, 92.46% of indels were detected, whilst at 0.1% VAF, 234 out of an expected 324 indels (72.22%) were detected (Fig 2B and S8 Table). A complete table of all expected indels and whether they were detected are available in S9 Table.

As with the SNV calling, the sensitivity of the assay did not vary within runs, between operators or between reagent Lots demonstrating high assay sensitivity, repeatability and reproducibility (Fig 2D).

To extend indel analysis across a broader set of unique samples and mutations we assessed a total of 31 unique samples containing a total of 115 variants at or above our LoD ($\geq$0.25% VAF) demonstrating a PPA of 97.4% (S6 Table).

Fusion gene detection sensitivity, repeatability and reproducibility

The InVisionFirst assay detects the DNA breakpoints that create the common *EML4-ALK* and *ROS1* gene fusions. Due to the scarcity of DNA samples with such breaks, three separate approaches were used to assess our sensitivity to detect these fusions. Fragmented cell line DNA was created (Horizon Discovery) with one *EML4-ALK* and one *SLC34A2-ROS1* fusion. Dilutions with these fusions at five different levels were created (VAF of 1%, 0.5%, 0.25%, 0.13% and 0.06%) and a similar replication strategy as used for SNVs and indels was undertaken (S3 Table).

For each dilution level, other than the lowest, 36 fusions were tested. At all levels, down to 0.13% VAF, all 36 fusions were detected (Fig 3A). At the lowest level of 0.06% VAF, 90% (27/30) were detected.

To test the assay performance across a broader spectrum of fusions an additional 44 breakpoints were synthesized then spiked into fragmented genomic DNA. 18 of these were designed according to breakpoints previously reported in NSCLC cases [20]. A further 26 were

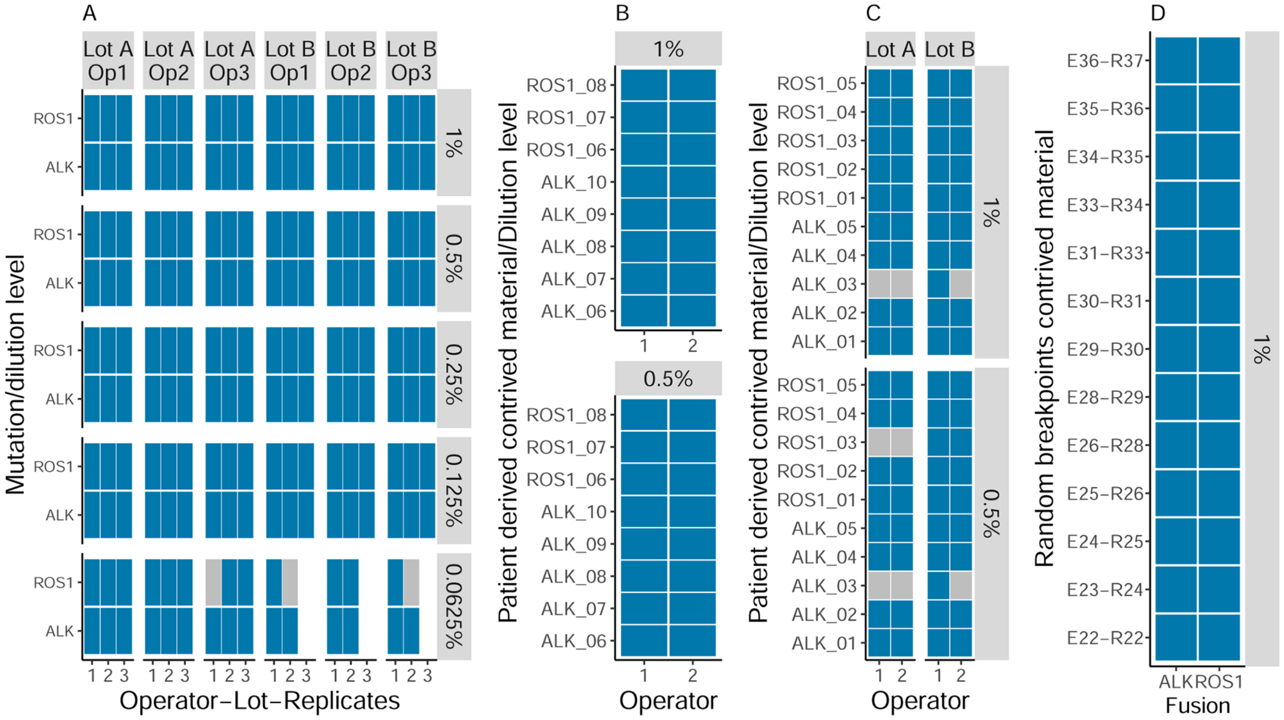


Fig 3. Fusion sensitivity analysis. Blue rectangles represent fusions that were detected and grey represents those missed. (A) Dilution of Horizon reference material containing 2 fusions (ALK and ROS1) across dilution levels (vertical) and operator/lot (horizontal), (B) Set 1 of contrived material based on published DNA breakpoints (AF 1% and 0.5%, 2 operators), (C) Set 2 of contrived material based on published DNA breakpoints (AF 1% and 0.5%, 2 operators, 2 reagent lots) and (D) Contrived material based on randomly generated fusion breakpoints. Different operators performed different parts of this fusion study.

https://doi.org/10.1371/journal.pone.0193802.g003

generated through random joining of fusion partners (S10 Table). The 18 published breakpoints where diluted to 1% and 0.5% in fragmented DNA and analyzed across multiple operators whilst the random breakpoints were diluted to 1% VAF and analyzed a single time.

Of the 44 unique synthetic fusions analyzed, 43 were detected in all repeats at 1% VAF (the remaining fusion was detected in 1 of 4 repeats) (Fig 3B, 3C & 3D). With the inclusion of Horizon cell line DNA, 97.8% (45/46) of the gene fusions were detected in all repeats at 1% VAF. At 0.5% VAF just one additional fusion was not detected in 2 of 4 repeats. All other fusions were called in all repeats at this level (Fig 3A, 3B & 3C) resulting in a total of 90% (18/20) of gene fusions that were detected at 0.5% VAF.

To extend the assessment of fusion detection sensitivity, the 2.5% VAF Horizon cell line fusion DNA was spiked into blood from 19 different donors (collected into both Streck BCT and EDTA tubes) at levels close to our LoD as described below. All fusions were detected in these samples (S11 Table). Collectively 54 unique samples with fusions and 104 variants at or above our LoD were analyzed. All but 3 were detected giving a PPA of 97.1% ($\geq$0.5% VAF).

Amplification detection sensitivity, repeatability and reproducibility

To determine the sensitivity of the InVisionFirst assay to *EGFR*, *FGFR1*, *ERBB2* and *MET* amplifications, double stranded DNA matching the parts of these genes targeted by the assay was manufactured, quantified by dPCR, then spiked into a background of sheared wild type DNA creating samples with copy number amplification ratios (CNAR) of 1.25x, 1.5x and 2x.

Each synthetic amplification was analyzed multiple times by three operators using two different Lots of reagents (S3 Table).

In total, each gene amplification was assessed at each dilution level between 22 and 24 times across the 3 operators. All amplifications for all four genes were detected at 2x CNAR, while 86 out of 88 amplifications were detected at 1.5x (97.7%). Detection at 1.25x CNAR ranged from 59% for *FGFR1* to 90.91% for *EGFR* (Table 1).

An additional series of samples were created with amplifications of all four genes between 2x and 50x to assess our reportable range and extend our assessment of PPA. A total of 49 unique samples were analyzed with 52 variants at or above our LoD of 1.5x CNAR and 51 variants were detected giving a PPA of 98.1% (S6 Table).

Specificity of the InVisionFirst assay

To determine assay specificity, blood was drawn into Streck BCT or EDTA tubes from donors not known to have cancer. 95 samples were analyzed for gene fusions and no calls were made. 109 samples (70 in Streck BCT and 39 in EDTA tubes) were analyzed for SNVs, indels and amplifications. No CNVs were detected in these 109 individuals. A total of 3 coding or splice altering variants were called at a VAF of between 0.13% and 1.57% (TP53 L369X, a TP53 splice alteration at chr17:7673838 and EGFR T790M, S12 Table). Digital PCR analysis was performed targeting all changes. The *TP53* mutation (g.chr17:7673838 C>A) at 1.57% was confirmed by dPCR. The mutations at 0.13% and 0.29% were not detected by dPCR. To determine the frequency with which we call these changes, we analyzed the presence of these changes in a further 92 samples from donors not known to have cancer and 242 samples from untreated NSCLC patients. None of these alterations were detected in this extended cohort.

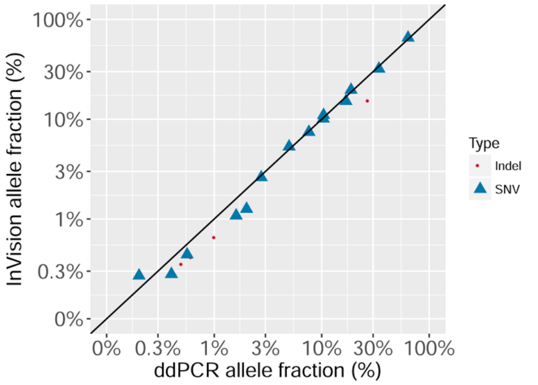
Comparison of the InVisionFirst assay with dPCR

To compare InVisionFirst with an orthogonal method, blood from 20 NSCLC patients was assessed with both the InVisionFirst assay and dPCR. Twenty patients were first identified with either a *KRAS* (p.G12C or p.G12D) or *EGFR* (p.L858R or p.E745_A750del/K) mutation above 0.25% VAF by the InVisionFirst assay (0.27%-65.55% VAF). In this cohort 40% of patients had a VAF <0.75% (S13 Table). cfDNA from a second tube of blood was then extracted from all 20 donors and shipped to an independent site (LGC, Teddington, UK) for blinded analysis. LGC analyzed all samples for all 4 mutations. Using dPCR they detected 19 of the 20 expected changes while not identifying any unexpected changes giving a PPA of 100% and a Positive Predictive Value (PPV) of 95% (Fig 4). The one change not detected by dPCR was a change identified at 0.3% by the InVisionFirst assay and was in a sample with comparatively low DNA input levels (an estimated 646 molecules were read by dPCR, implying 1–2 mutant molecules were to be expected in the sample). Comparison of VAF between dPCR and InVisionFirst showed excellent concordance (R-squared = 0.965) (Fig 4).

Table 1. Amplification sensitivity analysis for *FGFR1*, *EGFR*, *ERBB2* and *MET*.

Gene	1.25X (%)	1.5X (%)	2X (%)
MET	81.82	95.45	100
FGFR1	59.09	100	100
ERBB2	68.18	100	100
EGFR	90.91	95.45	100
Combined	75	97.73	100

https://doi.org/10.1371/journal.pone.0193802.t001



	dPCR.positive	dPCR.negative
InVision Positive	19	1
InVision Negative	0	60

Fig 4. Comparison between InVisionFirst assay and orthogonal dPCR generated by an independent laboratory. Four common cancer mutations were tested by dPCR in 20 samples selected to have one of these four mutations based on the InVisionFirst assay. The allele fraction of the mutation not detected by the orthogonal method is estimated by the InVisionFirst assay at 0.3%.

https://doi.org/10.1371/journal.pone.0193802.g004

Streck BCT and EDTA tube comparison

To demonstrate that the InVisionFirst assay can be used to analyze blood collected in either Streck BCT or EDTA tubes, reference DNA with SNVs, indels and fusions were spiked into the blood of donors not known to have cancer. 16,000 amplifiable copies of either sheared Tru-Q2, Tru-Q3 or the custom 2.5% VAF fusion cell line DNA (Horizon Discovery) were spiked into either blood tube type. Following extraction and successful sequencing of 36 spiked samples, all expected mutations were detected in both tube types (S11 Table).

Effect of delayed processing on mutant allele fraction

To assess the impact of delayed processing on mutant VAF for blood drawn into Streck BCT, Horizon 5% Multiplex I cfDNA Reference Standard DNA was spiked into whole blood from 4 donors then processed at 2, 3, 5, 7 or 10 days post blood draw. Variant allele fractions were assessed by the InVisionFirst assay and were shown to be stable following room temperature storage for up to 10 days (Fig 5).

Discussion

A number of assays with varying performance characteristics are available for molecular stratification of patients with NSCLC. In a recent study of patients treated with osimertinib following detection of an *EGFR* T790M mutation through ctDNA sequencing, 3 of the 7 best responders had the T790M mutation detected at VAF<0.25%, highlighting the potential benefit to patients of more sensitive assays [21].

Most assays have either high sensitivity for one or a limited number of mutations, such as assays based on dPCR, or a low sensitivity for a broader spectrum of changes. Recently, a limited number of assays that aim for both broad coverage and high sensitivity were introduced such as the InVisionFirst assay. In order for clinicians to differentiate assays and determine the one most suitable for their patient, detailed analytical validation, clinical validation and clinical utility studies will be needed in combination with factors such as turnaround time, cost and reproducibility.

Here we have demonstrated that the InVisionFirst assay has exceptionally high sensitivity for detecting SNVs, indels, amplifications and gene fusions. We have also shown high SNV and indel detection concordance between the InVisionFirst assay and dPCR in blood samples from NSCLC patients. Out of 80 possible changes assessed by both methods, 79 were

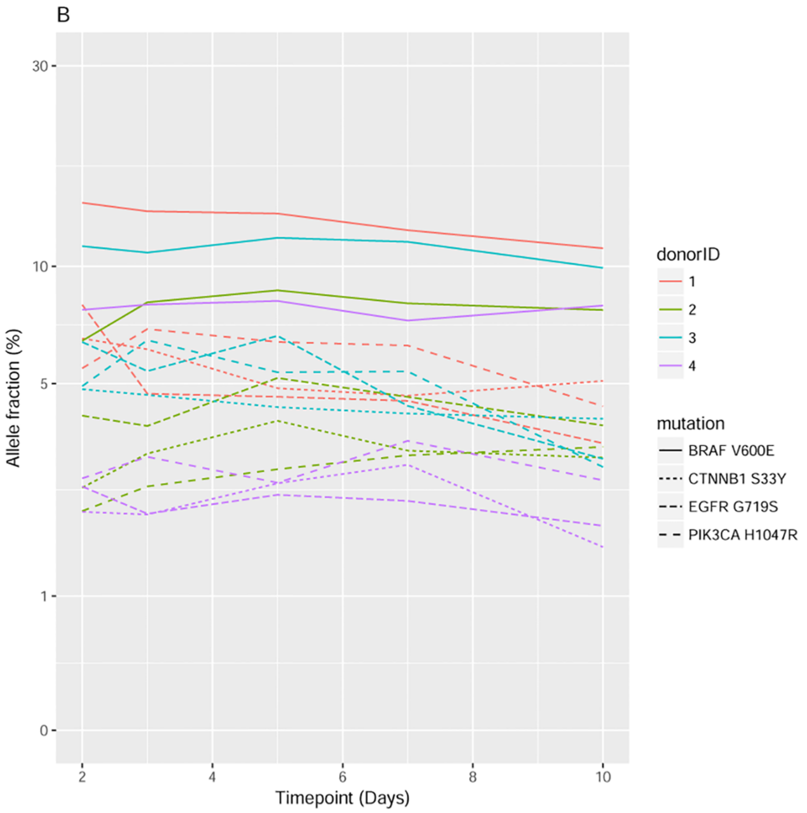


Fig 5. Stability of AF over time using Streck blood tubes spiked with Horizon reference material.

https://doi.org/10.1371/journal.pone.0193802.g005

concordant. A single *KRAS* p.G12C change that was detected by the InVisionFirst assay at VAF of 0.3% was not observed by dPCR. Multiplexing by dPCR is typically not practical and therefore all samples in this study had to be split 4 ways to analyse the 4 mutations by dPCR. The discordant *KRAS* change was found in a sample with low concentration of DNA and due to both this splitting of material and to sample loss (dead volume) common in dPCR (~39%); we therefore only expected to see ~ 1 to 2 mutant molecules in dPCR analysis, and the probability of allelic loss (zero representative molecules in the assay) was substantial (estimated probability of at least 14.4% to have no mutant copies present according to Poisson statistics). Separately a blinded study comparing InVsionFirst with both dPCR analysis of cfDNA and sequencing of tissue has shown high concordance [22].

In the analysis of donors not known to have cancer, no fusion or CNV calls were made. Three unexpected protein coding alterations were called in analysis of 109 samples (per base specificity 99.9997%). Through dPCR analysis, one of these alterations was confirmed whilst the other two were not detected. These calls were made at low VAF (0.13% and 0.29%) and could represent either false positives or true changes that could not be replicated at such low levels. We reviewed available data from a further 344 individuals that were either not known to have cancer or were newly diagnosed, untreated NSCLC patients, and neither change was detected in these samples, so each was detected only once in >450 samples. These data confirm the specificity of the assay. While false-positive calls in cfDNA at low VAF can occur, results from clinical studies have demonstrated that patients treated with osimertinib following detection of EGFR T790M mutation in plasma down to 0.06% VAF have achieved high rates of clinical response, demonstrating the importance of detecting mutations at low VAF [21].

Additional data is required to confirm that preliminary observation in larger populations. Based on the data presented here, the clinical specificity would appear to be greater than 99%. Measures such as setting a hard threshold based on a minimal allele fraction may thus result in the loss of clinically relevant true positive calls.

In order to attain high sensitivity and specificity in NGS analysis of ctDNA a number of factors are important. Key amongst these are conversion of sufficient DNA into a sequencing library and sequencing to sufficient depth to read mutant molecules multiple times and to have a low chance of missing mutant molecules in the sequencing step. Through the use of an amplicon based strategy, we previously showed that any molecule spanned by the amplification primers should be read [12]. The use of short amplicons results in a high fraction of DNA being analyzed. Other methods for targeted NGS analysis typically incorporate ligation steps and cleanup steps prior to amplification which may result in lost mutant molecules before analysis starts.

A second challenge is sequencing depth. By focusing on a panel of 36 key genes we can sequence more deeply than is routinely achieved for many larger panels. In this study using our routine process, the median depth at which an SNV was called was ~69,000x and 95% of calls had a depth greater than half this (S5 Table). By contrast, the current target depth for Foundation ACT is >5000x unique median coverage [23] and the target depth for Guardant 360 v2.10 is 15,000x [24]. We have a reportable range down to 0.0125% VAF for indels and SNVs, and with our high sequencing depth even a mutation at this level would typically result in 9 mutant reads.

The InVisionFirst assay has been developed to analyze blood drawn into either Streck BCT or EDTA tubes. Here we have demonstrated that when mutant DNA was spiked into both tube types at close to our LoD, neither inhibited mutation detection. As we assessed our specificity using blood drawn into both tube types, we have shown there is not a significant impact on specificity. The performance of Streck BCT for preventing white blood cell lysis and subsequent reduction in mutant allele fraction has already been demonstrated by others both in pregnant donors and cancer patients [25,26]. Our results support this showing that the mutant allele fractions of spiked DNA detected by the InVisionFirst assay stayed stable when blood processing was delayed for up to 10 days.

This study demonstrates that the InVisionFirst assay has high analytical sensitivity, specificity and reproducibility which are appropriate for clinical applications. Separate studies are ongoing to test clinical validity and utility in a range of settings.

Materials and methods

Healthy donor and cancer patient blood collection

For analysis of assay specificity, blood was collected from donors not known to have cancer. Blood was collected by a trained phlebotomist into both Streck BCTs and EDTA tubes by BioreclamationIVT (NY, USA). A minimum of two 10mL tubes were collected from each donor. For orthogonal assessment comparing the InVisionFirst assay to dPCR, blood was collected into Streck BCT from a series of NSCLC patients. All were analyzed using the InVisionFirst assay and the first 20 of those identified as having *KRAS* (p.G12C or p.G12D) or *EGFR* (p.L858R or p.E745_A750del/K) mutations at or above our LoD (0.25% VAF) with a second tube of blood available were selected for dPCR orthogonal testing. For extended assessment of 2 locations with potential false positives (chr7:55181378 C>T and chr17:7669684 C > -), an additional group of 242 untreated NSCLC patients from the same series as above and a further 92 individuals not known to have cancer were analysed.

Institutional Review Board (IRB) approval was obtained from the six centers collecting samples (Levine Cancer Institute, University of Colorado Lung Cancer Research Center, Holy Cross Hospital, Mid-Florida Hematology and Oncology, Christiana Care and North Shore Hematology Oncology). All patients provided written informed consent and data was de-identified so no patients could be identified by study personnel outside of the clinical trial site including the study authors.

Upon collection, the Streck BCTs were gently inverted 8–10 times before being shipped immediately to Inivata Inc (North Carolina) where they were processed within 7 days of collection. Here they were centrifuged at 1600 x g for 10 minutes at room temperature, plasma was removed, transferred to a new tube and a 2nd centrifugation step was performed at 20,000 x g for 10 minutes to pellet any remaining cellular debris before transferring the plasma to a new tube.

EDTA samples collected by BioreclamationIVT were processed immediately following collection before shipping to Inivata. The one significant modification from the Streck SOP was that the second spin was performed at 6500 x g as a faster centrifuge was not available. Upon completion of processing all cfDNA samples were frozen at -80 ˚C until ready for analysis.

Contrived ctDNA samples

SNVs. To assess SNV detection performance the Horizon Tru-Q reference material was used. Tru-Q7 and Tru-Q0 DNA were both sheared to ~200bp (Covaris) and quantified by Horizon Discovery. Tru-Q7 contains 39 validated mutations targeted by the InVisonFirst assay. 32 of these are at low allele fractions with the majority between ~1%-1.3% VAF (range: <1%-30% VAF). A full list of all mutations is available in S2 Table. All genomic changes described in this and subsequent tables use the hg38 human genome build. Two-fold dilutions were performed four times using Horizon Tru-Q 0 wild-type DNA as diluent to create the following mixes: 1%-1.3%, 0.5%-0.65%, 0.25%-0.33%, 0.13%-0.16% and 0.06%-0.08% VAF.

Indels. The Horizon Tru-Q7 reference DNA contains just a single indel (*EGFR* del746-A750). To assess the InVisionFirst assay's indel calling performance, SeraCare manufactured a custom indel reference material. 9 common insertions (+1 to +12bp) and 9 common deletions (-1 to -24bp) targeted by the InVisionFirst assay were synthesized by SeraCare. An additional 2 indels not currently covered by the InVisionFirst assay were also created in the mix. These mutations were mixed against "Genome in a Bottle" (GM24385) wild-type genomic DNA to produce mixes where the 18 targeted indels were present at approximately 2%, 1%, 0.5%, 0.25% or 0.1% VAF. The DNA was then sheared to ~150bp (Covaris) and the top three dilution levels were assessed by dPCR by SeraCare in order to confirm each indels VAF as compared to wild type background DNA (S7 Table). The lowest two dilutions were not tested by dPCR due to the expected low VAF of the indels.

Fusions. The InVisionFirst assay identifies *ALK* and *ROS1* gene fusions by detecting the genomic breakpoint junctions that bring the relevant genes together. As relatively few fusion-associated DNA breakpoints have been published to date and as there are only a small number of reference materials and cell lines with published *ALK* and *ROS1* DNA breakpoints, three different contrived materials were used.

A custom cell line mix was generated by Horizon Discovery. The resultant mix contained 1 *ROS1* and 1 *ALK* fusion (S10 Table). Following shearing by Horizon Discovery, they demonstrated with dPCR that the two fusions were present at ~2.5% VAF. These were subsequently diluted to 1%, 0.5%, 0.25%, 0.12% and 0.06% VAF.

To further assess the performance of the assay over a broad spectrum of breakpoints, the DNA junctions from 8 published *ROS1* gene fusions and 10 published *EML4-ALK* gene fusions were identified [20]. A further 26 synthetic fusions were then designed by computationally joining a 5' and 3' partner anywhere randomly within the common introns and exons targeted by the InVisionFirst assay (S10 Table).

A 500 bp sequence was designed for all fusions with the DNA breakpoint in the center. Synthetic fusions were manufactured as double stranded DNA fragments (IDT) then diluted and sheared (Covaris) to ~160 bp before dPCR quantification. Genomic DNA (Bioline) was also sheared to 160 bp and quantified using dPCR targeting a 108 bp region of the *RPP30* gene. The patient specific fusion DNA fragments were then spiked into genomic DNA at two different levels (1% VAF and at 0.5% VAF). Eight samples were created with both an *ALK* and a *ROS1* fusion and a further two were created with just a single *ALK* fusion. Thirteen samples were created containing one of the randomly generated *ALK* fusion sequences and one of the *ROS1* fusion sequences. All were present at 1% VAF (S10 Table).

CNVs. In order to assess the InVisionFirst assay's performance for CNV detection, a similar approach was taken to our analysis of fusions. Firstly, the amplicons we use for targeting the 4 genes currently assessed for amplification by the InVisionFirst assay were identified (*EGFR*, *FGFR1*, *ERBB2* and *MET*). 160 bp fragments of DNA were then selected encompassing these regions. These were synthesized as double stranded DNA fragments (IDT) then quantified by dPCR. Genomic DNA (Bioline) was then sheared and quantified as above to use for dilutions.

Each quantified double stranded DNA fragment was pooled by gene such that each targeted region was equally represented (for example *EGFR* had 11 targeted regions synthesized). Sheared background DNA and CNV pools were then combined to give the relevant amplified amounts.

Extended assessment of PPA and PPV. To further assess the InVisionFirst assay's ability to call a broad spectrum of SNVs, indels, fusions and CNVs across a large range of allele fractions above our LoD, an additional series of undiluted and diluted samples were created using DNA from the Horizon Tru-Q reference material series and SeraCare reference DNA (SeraCareTriLevel and SeraSeq). An extended set of CNV samples were created with CNAR of between 2-50x using the same method outlined above and mutant DNA was also spiked into a range of healthy donor samples as described below. Each of these samples was run just once. A full table of the different samples and their detected mutations is available in S6 Table.

Library preparation and analysis with the InVisionFirst assay

An earlier version of this assay based on eTAm-Seq has previously been described [17]. The InVisionFirst assay is based on the same approach but with the addition of the ability to call gene fusions. The updated assay also has an updated primer panel adding coverage to key *ALK* and *ROS1* inhibitor resistance mutations, and an amplicon size distribution of 73bp-155bp (median = 112bp). The targeted exons and introns targeted for fusion detection are described in S1 Table. cfDNA was first extracted from plasma using the QIAamp Circulating Nucleic Acid kit (Qiagen) followed by quantification by dPCR using the BioRad QX200 and an assay targeting a 108 bp region of the ribonuclease P/MRP subunit p30 (*RPP30*) gene. Contrived samples were quantified using the same assay. Yields were expressed as amplifiable copies (AC) of DNA. Two separate libraries were then setup in parallel from two blood tubes or from the contrived DNA. Where libraries were prepared using contrived samples, 16,000 amplifiable copies of the genome were used except in the amplification study where a mix of 16,000 and 2000 amplifiable copies (minimum input) were used. Both libraries were setup using a two-step amplification process that first targeted the desired regions then incorporated replicate and patient-specific barcodes and Illumina sequencing adaptors (See Fig 1). The first library targets SNVs, indels and CNVs whilst the second library has been designed to target all introns and exons brought together to create the three major *EML4-ALK* variants which collectively account for 93% of ALK fusions found in the COSMIC database (COSMIC Version 83). It also targets 92% (COSMIC V83) of the intronic and exonic bases brought together to create *CD74-ROS1*, *SLC34A2-ROS1*, *SDC4-ROS1* and *EZR-ROS1* fusions in lung carcinomas (S1 Table).

For both library types up to 48 samples were pooled together including positive and negative controls before sequencing on the Illumina NextSeq 500 (300 cycle PE) with 5% PhiX to monitor sequencing performance. Sequencing files were analyzed using the Inivata Somatic Mutation Analysis (ISoMA) pipeline to identify SNVs, CNVs and indels and the FUSP pipeline to call fusions. For the ISoMA pipeline a minimum Phred quality score of 30 for each base was required for inclusion in the analytics. In each run, in addition to the controls, we used the non reference allele fraction at common single nucleotide polymorphisms (SNPs) to detect potential contamination events. In addition, the overall sequencing depth at these common polymorphisms was used as part of quality control to confirm that sufficient sequencing depth had been generated.

For SNV and indel analysis, a background model was first established using samples from presumed healthy donors for each position/base pair change covered by our panel. The final determination of an SNV call integrated the data across multiple replicates for each sample in comparison with this background within a maximum likelihood framework. The same statistical principle was used for indels using samples from the same analytical batch in order to enable appropriate background calibration. The minimum depth at which any SNV or indel would be called was 1000x. In order to identify CNVs, a normalized measure of read depth was generated correcting for sample and amplicon effects in order to infer the copy number ratio between the 4 assessed genes (*ERBB2*, *FGFR1*, *MET* and *EGFR*) and the remainder of the genome.

Fusions were called by identifying the breakpoint sequences created when fusion partners joined. Patients with sequence reads matching to a 3' and 5' fusion partner were identified as fusion positive (e.g. *EML4* intron 13 and *ALK* intron 19). When the same breakpoint is detected twice, a fusion call is made. All variants were annotated using the canonical transcript for each gene. All SNVs and indels that resulted in coding and splice-site mutations were reported. Finally, a mutation calling report was generated providing a comprehensive summary of somatic alterations identified.

Orthogonal dPCR analysis

20 patients with NSCLC in whom a *KRAS* (p.G12C or p.G12D) or *EGFR* (p.L858R or p. E745_A750del/K) mutation was detected by the InVisionFirst assay above our LoD (0.25% VAF), were selected for dPCR orthogonal testing as described above. DNA was extracted from a second tube of blood from all 20 patients, assessed by dPCR using the *RPP30* assay then shipped on dry ice, anonymized, to LGC.

LGC had previously determined the suitability of the dPCR assays targeting the four mutations on the BioRad QX200 using both commercially available cfDNA standards (Horizon Diagnostics) and a set of in-house materials (*KRAS* G12C). The LoD for each assay was calculated at the start of the study using the false positive rate determined from ≥4 dPCR reaction per assay using ~116ng wild type gDNA per reaction. (S14 Table). Importantly although this LOD is achievable in samples with 116ng of DNA or greater, in samples with lower DNA inputs, sensitivity will be reduced in a predictable fashion based on the stochasticity of small numbers of mutant molecules.

LGC then performed a single dPCR with the *KRAS* G12C assay to determine the DNA concentration and whether the DNA could be run undiluted or needed to be diluted to be within the dynamic range of the BioRad QX200 dPCR platform. Finally, DNA from all twenty patients was assessed in triplicate (7μl per reaction) using the four assays and samples were called mutant or wild type depending on whether they were above or below the assays LoD (S13 and S14 Tables).

dPCR assays designed for assessment of unexpected calls (potential false positives)

In order to validate possible false positive calls made during the analysis of healthy donors, two digital PCR assays were ordered from BioRad and one was kindly donated by Dr Dana Tsui (Cancer Research UK Cambridge Institute, University of Cambridge, UK). The assays that were sourced from BioRad were designed using their online tool (https://www.bio-rad.com/digital-assays/#/). Synthetic mutant sequence for each assay was also designed using the online tool, and were ordered as double stranded DNA (IDT) and delivered pre-diluted (to 2000 copies per ul in 10mM Tris pH 8, 0.1mM EDTA and 0.1mg/mL Poly A). A mix of wild type DNA (BioLine) and ~5% synthetic mutant DNA was first tested with all 3 assays using a temperature gradient to determine optimal annealing temperatures. A dilution series of mutant to wild type DNA was then created then run in duplicate at two different concentrations along with wild type DNA to determine the background and limit of detection of each dPCR assay. Finally, each sample was run at least 4 times using 5μl of DNA.

Robustness of the InVisionFirst assay to Streck BCT and EDTA blood tube collection

To test that the InVisionFirst assay gives comparable results whether blood is collected in Streck BCT or EDTA blood tubes, blood from multiple donors was drawn into each tube type then this was processed to plasma as described above.

Both Streck and EDTA plasma was spiked with 16,000 amplifiable copies of either the Horizon fusion cell line reference material at 2.5% VAF or sheared Horizon reference material Tru-Q2 or Tru-Q3 which contain up to 14 variants (SNVs and indels) at 4 to 30% VAF. All samples were then mixed, extracted, then analyzed using the InVisionFirst assay.

Effect on mutant allele fraction of delayed Streck BCT processing

Whole blood was collected into five 10mL Streck BCTs from four individual donors as described above by BioreclamationIVT. This was shipped to Inivata and upon receipt of the tubes, 4000 copies of sheared (200bp) Horizon 5% Multiplex I cfDNA Reference Standard (Horizon Discovery) DNA was spiked into each sample. The five tubes from each donor were then randomized and kept for 2, 3, 5, 7 or 10 days at room temperature (~26˚C) before routine processing and analysis by the InVisionFirst assay. The allele fraction of each of the detected mutations from the Horizon reference standard was then compared to the matched sample from day 2 as a baseline to determine reduction in mutant VAF induced by delayed processing.

Supporting information

S1 Table. Details of the 36 genes targeted by InVisionFirst and which variant types are assessed for each gene. All exons covered for SNV, indel and CNV analysis are described. All

introns and exons covered for fusion gene detection are described.
(XLSX)

S2 Table. Tru-Q7 (Horizon) reference DNA mutations. A full list of validated mutations covered by the InVisionFirst panel are listed along with Horizon's validated variant allele fraction. All 39 variants are grouped by type.
(XLSX)

S3 Table. Overview of each study used to assess the four different variant types limit of detection.
(XLSX)

S4 Table. SNV sensitivity at different dilution levels of Horizon Tru-Q7 DNA.
(XLSX)

S5 Table. Full list of all SNVs in the Tru-Q7 dilution study and whether they were detected. DNA input (as assessed by dPCR), detected variant allele fraction (VAF) and total sequencing depth are all described.
(XLSX)

S6 Table. Details of samples used in CNV sensitivity analysis and the extended analysis of PPA for the four variant types. Section A describes the calculations used to determine PPA for all 4 variant types. Section B details the cell line mixes created and calls made in the extended assessment of PPA. Section C describes all contrived CNV samples and whether a call was made. This includes both the samples used to assess the assays LoD and those used for the extended assessment of PPA.
(XLSX)

S7 Table. List of all targeted indels in the custom SeraCare indel control material and comparison of SeraCare dPCR VAFs with average VAFs determined by InVisionFirst.
(XLSX)

S8 Table. Indel sensitivity at different dilution levels in the custom SeraCare indel control material.
(XLSX)

S9 Table. Full list of all indels in the custom SeraCare indel control material dilution study and whether they were detected. DNA input (as assessed by dPCR) and detected variant allele fraction (VAF) are both described.
(XLSX)

S10 Table. Details of all fusion reference materials. Part A describes the Horizon cell line mix. Section B describes the full sequence of the published patient specific fusions and section C describes all the randomly generated fusions.
(XLSX)

S11 Table. List of samples spiked with mutation positive DNA (Tru-Q2, Tru-Q3 or the custom 2.5% VAF fusion cell line DNA) and whether each mutation was detected. For SNVs and the one indel a detected VAF is reported.
(XLSX)

S12 Table. List of variants detected in donors not know to have cancer.
(XLSX)

S13 Table. List of variants detected by both InVisionFirst and blinded dPCR in an orthogonal analysis study.
(XLSX)

S14 Table. LoD for the 4 digital PCR assays analyzed by LGC in samples with optimal DNA input amounts.
(XLSX)

Acknowledgments

We would like to thank management and staff at the Cancer Research UK Cambridge Institute (CRUK-CI), Cancer Research Technologies, CRUK, Cambridge Enterprise and the University of Cambridge. Inivata, InVision, InVisionFirst, TAm-Seq and eTAm-Seq are UK registered trademarks of Inivata Ltd.

Author Contributions

Conceptualization: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Stefanie Lensing, Matt Smith, Michael Epstein, Peter Dimitrov, Clive Morris, Nitzan Rosenfeld, John Calaway, Greg Jones, Tim Forshew.

Data curation: Vincent Plagnol, Karen Howarth, Frank de Kievit, Greg Jones, Tim Forshew.

Formal analysis: Vincent Plagnol, Samuel Woodhouse, Frank de Kievit, Greg Jones, Tim Forshew.

Investigation: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Stefanie Lensing, Matt Smith, Michael Epstein, Mikidache Madi, Frank de Kievit, Esther Musgrave-Brown, Colin Herd, Katherine Baker-Neblett, Will Brennan, Nathan Campbell, Clive Morris, John Calaway, Tim Forshew.

Methodology: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Stefanie Lensing, Matt Smith, Michael Epstein, Mikidache Madi, Sarah Smalley, Esther Musgrave-Brown, Colin Herd, Will Brennan, Peter Dimitrov, Davina Gale, John Calaway, Tim Forshew.

Project administration: James Clark, Jamie Platt.

Resources: Sarah Smalley, James Clark, Jamie Platt.

Software: Vincent Plagnol, Michael Epstein, Catherine Leroy, Jonathan Hinton, Frank de Kievit, Peter Dimitrov.

Supervision: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Catherine Leroy, Jonathan Hinton, Nathan Campbell, Clive Morris, Nitzan Rosenfeld, James Clark, Davina Gale, Jamie Platt, John Calaway, Greg Jones, Tim Forshew.

Validation: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Matt Smith, Davina Gale, John Calaway, Greg Jones, Tim Forshew.

Visualization: Vincent Plagnol, Michael Epstein.

Writing – original draft: Vincent Plagnol, Tim Forshew.

Writing – review & editing: Vincent Plagnol, Samuel Woodhouse, Karen Howarth, Stefanie Lensing, Matt Smith, Michael Epstein, Mikidache Madi, Sarah Smalley, Catherine Leroy, Jonathan Hinton, Frank de Kievit, Esther Musgrave-Brown, Colin Herd, Katherine Baker-

Neblett, Will Brennan, Peter Dimitrov, Nathan Campbell, Clive Morris, Nitzan Rosenfeld, James Clark, Davina Gale, Jamie Platt, John Calaway, Greg Jones, Tim Forshew.

References

1. Sidransky D, Von Eschenbach A, Tsai YC, Jones P, Summerhayes I, Marshall F, et al. Identification of p53 gene mutations in bladder cancers and urine samples. Science. 1991; 252: 706–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/2024123 PMID: 2024123
2. Sidransky D, Tokino T, Hamilton SR, Kinzler KW, Levin B, Frost P, et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. Science. 1992; 256: 102–5. Available: http://www.ncbi.nlm.nih.gov/pubmed/1566048 PMID: 1566048
3. Mao L, Hruban RH, Boyle JO, Tockman M, Sidransky D. Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. Cancer Res. 1994; 54: 1634–7. Available: http://www.ncbi.nlm.nih.gov/pubmed/8137272 PMID: 8137272
4. Caldas C, Hahn SA, Hruban RH, Redston MS, Yeo CJ, Kern SE. Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. Cancer Res. 1994; 54: 3568–73. Available: http://www.ncbi.nlm.nih.gov/pubmed/8012983 PMID: 8012983
5. Takeda S, Ichii S, Nakamura Y. Detection of K-ras mutation in sputum by mutant-allele-specific amplification (MASA). Hum Mutat. 1993; 2: 112–117. https://doi.org/10.1002/humu.1380020209 PMID: 8318987
6. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev. 3: 67–71. Available: http://www.ncbi.nlm.nih.gov/pubmed/8118388 PMID: 8118388
7. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. Biotechniques. 1992; 13: 444–9. Available: http://www.ncbi.nlm.nih.gov/pubmed/1389177 PMID: 1389177
8. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A. 1999; 96: 9236–41. Available: http://www.ncbi.nlm.nih.gov/pubmed/10430926 PMID: 10430926
9. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med. 2008/08/02. 2008; 14: 985–990. https://doi.org/10.1038/nm.1789 PMID: 18670422
10. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer. 2017; 17: 223–238. https://doi.org/10.1038/nrc.2017.7 PMID: 28233803
11. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med. 2014/02/21. 2014; 6: 224ra24. https://doi.org/10.1126/scitranslmed.3007094 PMID: 24553385
12. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DWY, Kaper F, et al. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. Sci Transl Med. 2012/06/01. 2012; 4: 136ra68–136ra68. https://doi.org/10.1126/scitranslmed.3003726 PMID: 22649089
13. Murtaza M, Dawson S-J, Tsui DWY, Gale D, Forshew T, Piskorz AM, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. Nature. 2013; 497: 108–112. https://doi.org/10.1038/nature12065 PMID: 23563269
14. Newman AM, Bratman S V, To J, Wynne JF, Eclov NCW, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. Nature Publishing Group; 2014; 20: 548–554. https://doi.org/10.1038/nm.3519 PMID: 24705333
15. Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. PLoS One. 2015; 10: 1–27. https://doi.org/10.1371/journal.pone.0140712 PMID: 26474073
16. P150044. Premarket Approval (PMA). US Food Drug Adm Premarket Approv Cobas EGFR Mutat test V2 2017 FDA https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P150044.
17. Gale D, Plagnol V, Lawson A, Pugh M, Smalley S, Howarth K, et al. Abstract 3639: Analytical performance and validation of an enhanced TAm-Seq circulating tumor DNA sequencing assay. Cancer Res. American Association for Cancer Research; 2016; 76: 3639–3639. https://doi.org/10.1158/1538-7445.AM2016-3639
18. Jordan EJ, Kim HR, Arcila ME, Barron D, Chakravarty D, Gao J, et al. Prospective Comprehensive Molecular Characterization of Lung Adenocarcinomas for Efficient Patient Matching to Approved and

Emerging Therapies. Cancer Discov. NIH Public Access; 2017; 7: 596–609. https://doi.org/10.1158/2159-8290.CD-16-1337 PMID: 28336552

19. Kuderer NM, Burton KA, Blau S, Rose AL, Parker S, Lyman GH, et al. Comparison of 2 Commercially Available Next-Generation Sequencing Platforms in Oncology. JAMA Oncol. American Medical Association; 2017; 3: 996. https://doi.org/10.1001/jamaoncol.2016.4983 PMID: 27978570
20. Seki Y, Mizukami T, Kohno T. Molecular Process Producing Oncogene Fusion in Lung Cancer Cells by Illegitimate Repair of DNA Double-Strand Breaks. Biomolecules. 2015; 5: 2464–76. https://doi.org/10.3390/biom5042464 PMID: 26437441
21. Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, et al. Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. Ann Oncol. 2017; 28: 784–790. https://doi.org/10.1093/annonc/mdx017 PMID: 28104619
22. Guibert N, Hu Y, Feeney N, Kuang Y, Plagnol V, Jones G, et al. Amplicon-based next-generation sequencing of plasma cell-free DNA for detection of driver and resistance mutations in advanced non-small cell lung cancer. Ann Oncol Off J Eur Soc Med Oncol. 2018; 1–30. https://doi.org/10.1093/annonc/mdy005 PMID: 29325035
23. Allen JM, Schrock AB, Erlich RL, Miller VA, Stephens PJ, Ross JS, et al. Genomic Profiling of Circulating Tumor DNA in Relapsed EGFR-mutated Lung Adenocarcinoma Reveals an Acquired FGFR3-TACC3 Fusion. Clin Lung Cancer. 2017; 18: 219–22. https://doi.org/10.1016/j.cllc.2016.12.006 PMID: 28089157
24. Vowles J, Odegaard J, Mortimer S, Fairclough S, Sikora M, Abdueva D, et al. Abstract 5705: Analytical validation of Guardant360 v2.10. Cancer Res. American Association for Cancer Research; 2017; 77: 5705–5705. https://doi.org/10.1158/1538-7445.AM2017-5705
25. Wong D, Moturi S, Angkachatchai V, Mueller R, DeSantis G, van den Boom D, et al. Optimizing blood collection, transport and storage conditions for cell free DNA increases access to prenatal testing. Clin Biochem. 2013; 46: 1099–1104. https://doi.org/10.1016/j.clinbiochem.2013.04.023 PMID: 23643886
26. Medina Diaz I, Nocon A, Mehnert DH, Fredebohm J, Diehl F, Holtrup F. Performance of Streck cfDNA blood collection tubes for liquid biopsy testing. PLoS One. 2016; 11. https://doi.org/10.1371/journal.pone.0166354 PMID: 27832189

Annals of Oncology 28: 784–790, 2017
doi:10.1093/annonc/mdx017
Published online 19 January 2017

ORIGINAL ARTICLE

Osimertinib benefit in *EGFR*-mutant NSCLC patients with *T790M*-mutation detected by circulating tumour DNA

J. Remon[1], C. Caramella[2], C. Jovelet[3], L. Lacroix[3], A. Lawson[4], S. Smalley[4], K. Howarth[4], D. Gale[4,5], E. Green[4], V. Plagnol[4], N. Rosenfeld[4,5,6], D. Planchard[1], M. V. Bluthgen[1], A. Gazzah[1], C. Pannet[1], C. Nicotra[1], E. Auclin[1], J. C. Soria[1,7] & B. Besse[1,7]*

Departments of [1]Oncology Medicine; [2]Radiology, Gustave Roussy, Université Paris-Saclay, Villejuif; [3]Translational Research Laboratory, AMMICA, INSERM US23/CNRS UNS3655, Gustave Roussy, Villejuif, France; [4]Inivata Ltd, Cambridge; [5]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge; [6]Cambridge Cancer Centre, Cambridge, UK; [7]University Paris-Sud and Gustave Roussy Cancer Campus, Villejuif, France

*Correspondence to: Prof. Benjamin Besse, Department of Oncology Medicine, University Paris-Sud and Gustave Roussy Cancer Campus, 114 Rue Edouard Vaillant, 94805 Villejuif, France. Tel: +33-1-42-11-43-22; Fax: +33-1-42-11-52-19; E-mail: benjamin.besse@gustaveroussy.fr

Background: Approximately 50% of epidermal growth factor receptor (*EGFR*) mutant non-small cell lung cancer (NSCLC) patients treated with EGFR tyrosine kinase inhibitors (TKIs) will acquire resistance by the *T790M* mutation. Osimertinib is the standard of care in this situation. The present study assesses the efficacy of osimertinib when *T790M* status is determined in circulating cell-free tumour DNA (*ct*DNA) from blood samples in progressing advanced *EGFR*-mutant NSCLC patients.

Material and methods: *ct*DNA *T790M* mutational status was assessed by Inivata InVision^{TM} (eTAm-Seq^{TM}) assay in 48 *EGFR*-mutant advanced NSCLC patients with acquired resistance to EGFR TKIs without a tissue biopsy between April 2015 and April 2016. Progressing *T790M*-positive NSCLC patients received osimertinib (80 mg daily). The objectives were to assess the response rate to osimertinib according to Response Evaluation Criteria in Solid Tumours (RECIST) 1.1, the progression-free survival (PFS) on osimertinib, and the percentage of *T790M* positive in *ct*DNA.

Results: The *ct*DNA *T790M* mutation was detected in 50% of NSCLC patients. Among assessable patients, osimertinib gave a partial response rate of 62.5% and a stable disease rate of 37.5%. All responses were confirmed responses. After median follow up of 8 months, median PFS by RECIST criteria was not achieved (95% CI: 4–NA), with 6- and 12-months PFS of 66.7% and 52%, respectively.

Conclusion(s): *ct*DNA from liquid biopsy can be used as a surrogate marker for *T790M* in tumour tissue.

Key words: EGFR mutation, *T790M*, osimertinib, lung cancer, ctDNA liquid biopsies

Introduction

The activated epidermal growth factor receptor (*EGFR*) mutation is present in almost 50% of patients with advanced non-small cell lung cancer (NSCLC) who are of Asian ethnicity compared with only 12% in the Caucasian population [1].

These mutations predict sensitivity to first- and second-generation EGFR tyrosine kinase inhibitors (TKIs) such as erlotinib, gefitinib or afatinib. Response rate and progression-free survival (PFS) with EGFR TKIs are superior to standard first-line platinum doublet chemotherapy, making them the standard of care [2]. However, tumours invariably develop acquired resistance 9–13 months after treatment initiation. The substitution of threonine to methionine at amino acid position 790 (*T790M*) in exon 20 of the *EGFR* gene reduces first-generation EGFR TKIs binding, and accounts for over half of acquired resistance mechanisms [3, 4].

Knowledge of acquired resistance mechanisms to EGFR TKIs was one of the triggers behind the development of personalised therapies, with the introduction of the third-generation EGFR-TKIs, which are active against sensitive, as well as resistant *T790M EGFR* mutations, such as osimertinib [5]. Both the FDA

© The Author 2017. Published by Oxford University Press on behalf of the European Society for Medical Oncology.
All rights reserved. For Permissions, please email: journals.permissions@oup.com.

and the EMA recently approved osimertinib in patients with acquired EGFR *T790M* mutations tested in a tumour-tissue biopsy or in plasma [6, 7], but noted that osimertinib efficacy has not been prospectively established in patients where *T790M* mutation was determined in plasma with unknown status in the tissue. Lack of available tissue for performing molecular profile (such as when bone metastases are present, as reported in almost 50% of cases [8], requiring decalcification of the samples impairing DNA quality), the location or size of the tumour at progression, and the risk of complications, are serious limitations to re-biopsy NSCLC tumours. Moreover, single site biopsies may not provide a representative profile of the overall predominant resistance mechanisms for a given patient [9].

Liquid biopsies based on circulating cell-free tumour DNA (*ct*DNA) analysis have been described as surrogate samples for molecular analysis replacing solid tumours [10], and may allow real-time sampling of multifocal clonal evolution [11]. Here we assessed the feasibility of identifying *T790M* mutations in *ct*DNA isolated from blood samples in a cohort of *EGFR*-mutant NSCLC patients with progression under first- or second-generation EGFR TKIs without a tissue biopsy at progression, in order to detect acquired resistance. The efficacy of osimertinib in the *ct*DNA *T790M*-positive NSCLC patients was also assessed.

Patients and methods

Patients

Eligible patients treated at the Gustave Roussy (Villejuif, France) between April 2015 and April 2016 were included in this study. Patients had to have advanced NSCLC, the presence of a common activating *EGFR*-mutation in the initial biopsy (*Del19*, *L858R*), clinical or radiological progression to at least one first- or second-generation EGFR TKI [12], and ineligibility for a new tissue biopsy (due to lack of available tissue, localisation and/or patient's refusal) for testing *T790M* status at the time of progression. There was no upper limit for the number of prior EGFR-inhibitor or systemic therapies. All patients provided written informed consent for biomedical research (CEC-CTC IDRcb2008-AOO585-50) and the institutional ethics committee approved the protocol. Osimertinib at 80 mg daily was prescribed as a part of the French Expanded Access Program in France, which allow its prescription when *T790M* was present in tumour-tissue biopsy or in a liquid biopsy.

Outcomes

The primary endpoint was to determine the overall response rate with osimertinib in patients treated on the basis of a positive *T790M* mutational status from a liquid biopsy results. Secondary endpoints included: the percentage of *T790M* mutation-positive patients identified by *ct*DNA analysis from pretreated *EGFR*-mutant patients with progression to systemic treatment, PFS by radiological criteria and investigator's criteria and overall survival on osimertinib.

As an exploratory objective, correlation between RECIST radiological responses with osimertinib and three *ct*DNA predictors was evaluated: (i) *T790M* allele fraction (AF), (ii) *EGFR* activating mutation AF and (iii) ratio of *T790M* and *EGFR* activating mutation AF.

PFS was calculated from the initiation of osimertinib treatment until the date of progression by RECIST 1.1 or death (whichever came first), with censoring at the date of last follow-up if the patient had not progressed. PFS by investigator (time to off-osimertinib progression if osimertinib therapy was extended beyond progression at investigator discretion) was also assessed. Overall survival (OS) was calculated from the initiation of osimertinib treatment until the date of death.

InVision^{TM} (eTAm-seq^{TM}) analysis

Ten millilitres of blood were collected in K2-EDTA tubes and processed at the time of disease progression (clinical or radiological). DNA was extracted from <5 ml of plasma and analysed by the InVision assay, using enhanced Tagged Amplicon-Sequencing; eTAmSeq^{TM} [13], which was developed from TAm-Seq$^{®}$ assay [14] (Supplementary Appendix S1).

Radiologic assessments

Before prescribing osimertinib, all patients underwent tumour imaging, including computed tomography of the chest and abdomen and/or PET-scan. Brain imaging was performed in cases of symptoms. Restaging scans were obtained at least 4-weeks after treatment initiation and then every 6–8 weeks. Senior radiologist (CC) centrally reviewed the response rate and determined best response to osimertinib according to RECIST v1.1 [15]. The objective response rate was defined as the percentage of patients with response (complete or partial) at first restaging after osimertinib initiation. Confirmed responses were defined as persistent responses (partial or complete) at second radiological assessment. Only assessable patients who received osimertinib based on positivity for the *T790M* mutation from *ct*DNA liquid biopsies were evaluated for the response rate.

Results

Patient characteristics

Forty-eight advanced *EGFR*-mutant NSCLC patients with radiological or clinical progression on systemic treatment were evaluated for *T790M* status in a liquid biopsy. Median age was 65 years (range 37–83); 36 (75%) patients were women and 58% were never-smoker. *EGFR* mutation status was *Del19* in 33 (69%) and *L858R* in 15 (31%) NSCLC patients.

T790M status in a liquid biopsy

The *T790M* positivity in *ct*DNA was reported in 24 out of 48 (50%) NSCLC patients (supplementary Figure S1, available at *Annals of Oncology* online).

Activating *EGFR* mutational status in *ct*DNA analysis confirmed that the original mutation was maintained in 23 out of 24 *T790M*-positive samples. The *T790M* mutation positivity was more frequent among patients with the *EGFR Del19* mutation (20 out of 33 patients, 61%) compared with the EGFR *L858R* mutation (4 out of 15, 27%). Concomitant mutations to *T790M* mutation were reported in three patients (Table 1).

For 9 of the 24 patients with *ct*DNA *T790M*-positivity, the *T790M* AF was lower than 0.5% in the liquid biopsy (supplementary Table S1, available at *Annals of Oncology* online).

Osimertinib response rate

Of the 24 NSCLC patients with a *T790M* mutation in the *ct*DNA, 18 received osimertinib at progression and were evaluated for response (supplementary Figure S1, available at *Annals of Oncology* online).

Table 1 summarizes baseline demographic characteristics of NSCLC patients who were *T790M* positive by *ct*DNA and treated with osimertinib. Median age was 63 years, and a total of 78% of

Table 1. Patients' characteristics with *T790M* mutation positive in a liquid biopsy who received osimertinib

Patient	Gender	Age (years)	Pack-years	EGFR mutation	*T790M* AF (%)	Previous systemic treatments	Previous EGFR TKI	Other mutations[a]	Last treatment before osimertinib	RECIST osimertinib
1	M	51	0	Del19	0.41	3	1	TP53 (P151X, R273H)	Erlotinib	NE
2	F	56	0	Del19	15.96	3	2	TP53 (Q331*, V225A)	Erlotinib	SD (−10%)
3	M	54	6	Del19	0.86	3	1	TP53 (R337C) STK11 (P179L)	Erlotinib -BVZ	PR (−50%)
4	F	37	0	Del19	1.06	3	2	TP53 (Q165*)	Erlotinib	PR (−84%)
5	M	67	6	Del19	1.60	4	2	CTNBB1 (S37S)	Pem/Cis	PR (−50%)
6	F	83	10	Del19	6.96	2	2	CTNNB1 (S33C)	Erlotinib	SD (0%)
7	F	67	0	Del19	19.60	1	1	CDKN2A (frameshift) TP53 (frameshift)	Erlotinib	PR (−50%)
8	F	70	0	L858R	0.25	2	1	NRAS (A59G)	Pem	SD (−26%)
9	F	66	5	L858R	0.07	10	3[b]	–	Erlotinib	PR (−65%)
10	F	81	0	L858R	5.38	4	3	TP53 (P60X, splice) PIK3CA (E545K)	Pem	PR (−33%)
11	F	70	0	Del19	0.31	3	1	TP53 (R282W)	Pem	NE
12	F	58	0	Del19	0.24	6	2	–	Erlotinib	PR (−68%)
13	F	54	0	Del19	2.24	3	2	–	Pem/Cb	SD (9%)
14	F	59	10	Del19	0.14	2	1	TP53 (I232S)	Gefitinib	PR (−50%)
15	F	67	2	L858R	0.30	3	1	EGFR (K860I)	Erlotinib	SD (−20%)
16	M	61	20	Del19	0.70	5	3[c]	TP53 (E343*, C238Y, C135X)	Afatinib	SD (−18%)
17	F	54	3	Del19	3.95	2	1	TP53 (R249S)	Gefitinib	PR (−32%)
18	F	65	0	Del19	0.68	1	1	CTNNB1 (S37C)	Gefitinib	PR (−32%)

M, male; F, female; AF, allelic fraction; BVZ, bevacizumab; Pem/Cis, pemetrexed/cisplatin; Pem/Cb, pemetrexed/carboplatin; NE, not evaluable; SD, stable disease; PR, partial response.

[a] Other mutation at the moment of *T790M* positive in the liquid biopsy (all patients had the common EGFR mutation at the time of *T790M* mutation positive, except patient number 15 whom original EGFR Del19 mutation was not found at acquired resistance).

[b] This patient had already received rociletinib.

[c] This patient has already been treated with osimertinib.

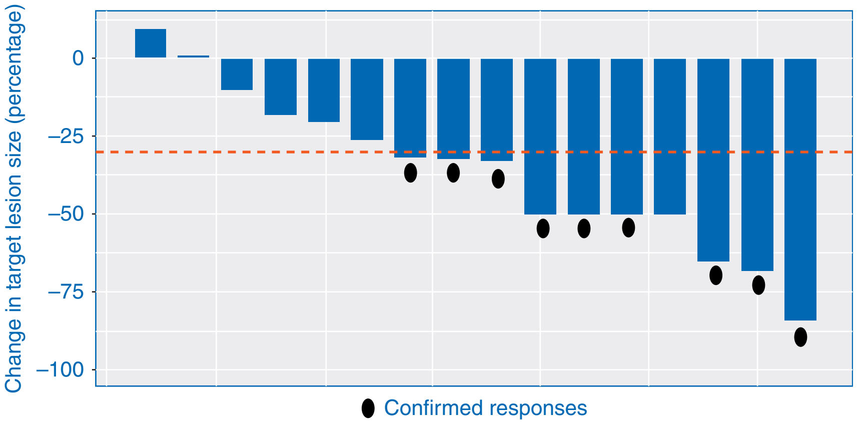


Figure 1. Best percentage change in target-lesion size (waterfall plot of *T790M* positive NSCLC patients in a liquid biopsy treated with osimertinib).

patients (14 of 18) were female. All the patients had received at least one prior EGFR TKI. Three or more previous systemic treatment lines were reported in up to 65% of patients and in 70% of cases an EGFR TKI was the last treatment before starting osimertinib.

Two patients were not evaluated for response: one having only bone metastases and the other died due to a treatment-unrelated cerebral haemorrhage. Of the 16 assessable patients, 10 had a partial response (62.5%), and 6 had stable disease (37.5%). No patients had complete response or disease progression as best response (Table 1 and Figure 1).

Among those patients with partial response ($n = 10$), all had second radiological assessment to confirm response, and the

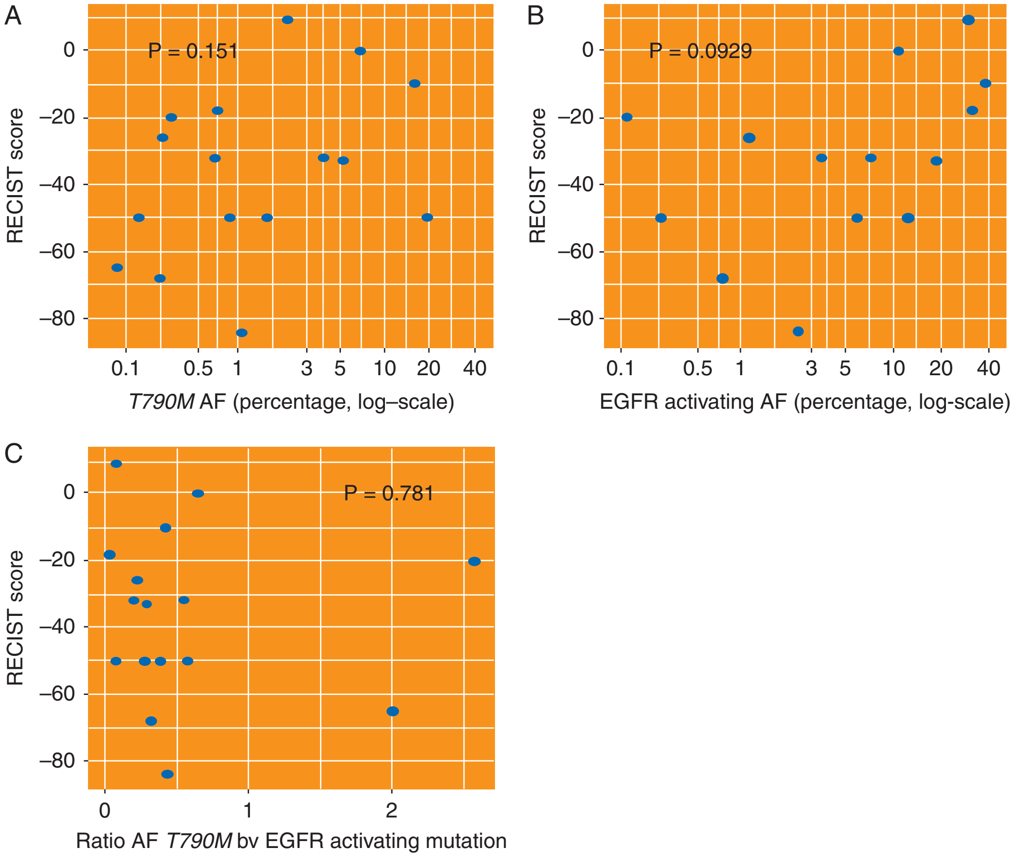


Figure 2. Correlation between RECIST radiological responses with osimertinib and three ctDNA predictors: (A) *T790M* AF, (B) EGFR activating mutation AF and (C) *T790M* by EGFR activating mutation AF ratio.

response was confirmed in 90% of patients (1 patient progressed at the second radiological assessment). Of note, one patient previously treated with rociletinib, received osimertinib as tenth line treatment achieving a partial response.

The median time between the blood draw in which *ct*DNA *T790M* positivity was detected and start of osimertinib treatment was 6 weeks.

Correlation between RECIST and ctDNA predictors

Correlations between RECIST radiological responses with osimertinib and three *ct*DNA predictors: (i) *T790M* AF, (ii) *EGFR* activating mutation AF and (iii) ratio of *T790M* and *EGFR* activating mutation AF were evaluated, however, none showed significance (Figure 2), but a trend (*P*-value 0.09–0.15) was observed for larger decrease in tumour size for smaller mutant AFs of *T790M* or *EGFR* activating mutations. Of the seven cases with best response (decrease of 50% or more in size), three cases had *T790M* detected at <0.25%.

Progression free survival and overall survival

After a median follow up of 8.5 months, median PFS on osimertinib by RECIST 1.1 criteria was not achieved (95% CI: 4–NA), with a 6- and 12-months PFS of 66.7% and 52%, respectively (Figure 3). By investigator, median PFS was 13 months (95% CI: 8–NA), with 6- and 12-months PFS of 79% and 70%, respectively (supplementary Figure S3, available at *Annals of Oncology* online). At the time of cut-off 4 patients had died; hence overall survival (OS) was not achieved. One-year OS was 78% (95% CI: 59–97) (supplementary Figure S2, available at *Annals of Oncology* online).

Discussion

Osimertinib is a third-generation oral EGFR TKI developed to treat tumours bearing sensitizing *EGFR* and acquired resistant *T790M*-mutations, that spares the wild type form of the receptor [16]. To the best of our knowledge, our analysis is the first to prospectively test in a real-world setting the efficacy of osimertinib according to *ct*DNA results. In this study, osimertinib achieved a 62.5% response rate and 12-months PFS of 52% among NSCLC patients who were *T790M*-mutation positive, based on *ct*DNA analysis by a multiplexed deep sequencing [13] assay. These results are comparable to the efficacy reported with osimertinib in patients with *T790M* mutation detected in a tumour tissue biopsy [5, 16]. In the phase 3 AURA3 study, osimertinib provided a 71% of response rate and 12-months PFS of 44% in pre-treated and tissue *T790M*-mutation positive NSCLC patients [16]. However, in the phase I AURA trial, some patients with *T790M*-mutation negative also responded to osimertinib [5] reflecting the inadequacy of tissue-biopsy for catching tumour heterogeneity. In the *post hoc* exploratory analysis of the samples from the phase I AURA trial, which included 216 patients (73% were *T790M*-positive in the tumour) osimertinib gave a response rate of 63%

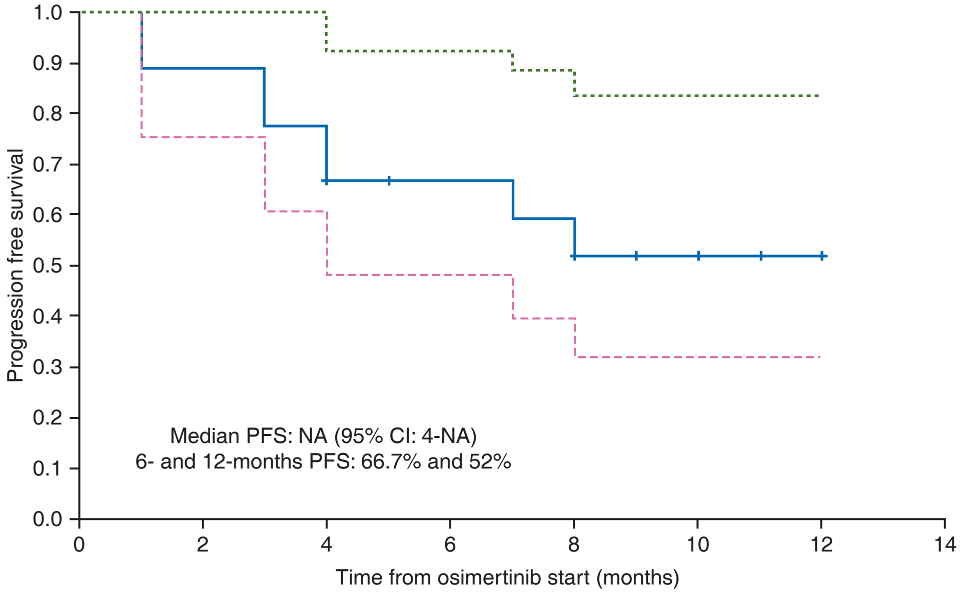


Figure 3. Progression-free survival (PFS) by RECIST 1.1 criteria. NA, not achieved.

among patients who were *T790M*-mutation positive according to central blood-test genotyping by the BEAMing method (allelic fraction for positive results for *T790M* mutation ≥0.06%) [17].

Liquid biopsies based on *ct*DNA analysis are described as surrogate samples for tumour molecular analysis [10], and also as potential dynamic markers for monitoring the efficacy of EGFR TKI [18, 19] and early detection of resistance mutations [20]. Liquid biopsy assays have been developed for analysis of hot-spot mutations and gene panels. Hot-spot assays can offer lower complexity and some PCR-based assays for detection of mutations in *EGFR* (including activating mutations and the *T790M* mutation) have received CE-mark [21] and approval by the FDA for in-vitro use [22]. Several commercial laboratories now offer sensitive assays for ctDNA using targeted deep sequencing of gene panels that include *EGFR*. In our study, the rate of *T790M* mutation positivity in a liquid biopsy among *EGFR*-mutant patients progressing on systemic treatment was 50%, which is consistent with previous biopsy series [3, 4] and clinical trials [5, 23]. In a recent prospective exploratory analysis, the resistance-associated mutation in *ct*DNA (tested by cobas *EGFR* Mutation Testv2) among *EGFR*-mutant NSCLC patients was detected in 50% of patients, and concordance with tumour biopsy-derived genotyping was 61% [24]. Among patients with sufficient material for concurrent *ct*DNA and tumour-derived genotyping, *ct*DNA identified the *T790M* mutation in 5 of 25 (20%) in whom the concurrent study biopsy was negative. Similarly, in the phase I AURA trial, *T790M* was detected in plasma of 30% of patients with *T790M*-negative tumours [17]. Discrepancies between tumour biopsy and *ct*DNA genotyping may result from technological differences, or sampling of different tumour cell populations in a heterogeneous setting [24]. Studies focusing on the discrepancy of *T790M* mutation between tissue and plasma samples are underway using amplification-refractory mutation system (ARMS) and droplet digital PCR methods (NCT02418234). Moreover, recent data suggest that *ctDNA T790M* mutation derived from NSCLC tumours can be detected with high sensitivity in urine as well as in plasma, enabling complementary modes of tissue and liquid biopsies in EGFR TKI resistant NSCLC [25]. Although sensitivity and specificity of *ct*DNA varies across different technology platforms [26], the establishment of robust and standardised protocols for blood sampling, processing, storage, DNA extraction and analysis will support liquid biopsies as new standard tests in the near future for tumour genotyping as well as predictive biomarkers [26].

In this setting, the relatively low number of patients, the heavy degree of pre-treatment population included in our analysis (median of four previous treatment lines, 33% with at least two EGFR TKIs before osimertinib initiation and two patients previously pre-treated with *T790M*-inhibitors), the lack of corresponding tumour sample for all patients, and the heterogeneity in terms of lines of treatment are all considered as potential limitations. Moreover, *ct*DNA cut-off points to define the clinical relevance of the findings specifically based on functional consequences, namely their ability to predict therapeutic responsiveness, are required before *ct*DNA can be routinely implemented in clinical practice. Interestingly, the observation in our cohort that the *T790M* AF was not significantly correlated with clinical response suggests that any level of *T790M* positivity may be clinically relevant, independent of the AF threshold. However, the relatively long time delay between establishment of *ct*DNA *T790M* positivity and osimertinib initiation may mean that the AF at the moment of treatment initiation may be higher than the reported results. Our data suggest a possible importance for detection of *T790M* at low AFs, but additional studies are needed to confirm the minimum biological threshold with clinical relevance.

Testing tumour tissue is so far the recommended method for detecting the presence of the resistant *T790M* mutation among *EGFR*-mutant NSCLC patients and tailoring treatment [6, 7], Prior biopsy-based studies have reported multiple acquired resistance mechanisms in ~5%–15% of NSCLC patients with EGFR TKIs [3, 4]. However, up to 23% of tumour tissue specimens available at the time of acquired resistance have been reported as providing limited, low quality material for tumour genotyping [4, 24], and may not be representative of the entire genomic landscape of the tumour [9, 27]. In addition, not all patients are suitable for new tissue biopsy at progression, which can thereby delay treatment initiation [28]. Recently, mechanisms of acquired resistance after first-line EGFR TKI were analysed in *ct*DNA by CAPP-Seq in 41 *EGFR*-mutant NSCLC patients. At least 46% of these tumours had developed another mechanism of acquired resistance in addition to *T790M* mutation, and these multiple resistance mechanisms were associated with poorer outcome to third generation EGFR TKIs [29]. In our analysis, blood samples from three patients reported concomitant mutations with no clear correlation with outcome: one *PIK3CA* mutation, previously reported as mechanism of acquired resistance [3]; and two other mutations, *STK11* and *NRAS* mutation, not previously described as acquired resistance mechanisms to first- or second-generation EGFR TKI. However, *NRAS* mutation has been recently reported as an acquired mechanism of resistance to osimertinib in preclinical models [30]. *ct*DNA analysis may allow the development of rational trials for personalised selection of combined therapies to address intra-tumoural heterogeneity; however, a risk-benefit assessment should be performed to avoid substantial increases in toxicity.

Conclusion

In this analysis of liquid biopsies in a small cohort of *EGFR*-mutant NSCLC patients with acquired resistance to systemic treatment, our results provide relevant clinical data about the efficacy of osimertinib in a real-world setting among patients where *T790M*-positivity was detected in *ct*DNA, supporting the use of such liquid biopsies for personalising treatment in lung cancer patients. Our results suggest a possible clinical importance for detection of *T790M* at low levels in plasma samples.

Funding

None declared.

Disclosure

Authors affiliated with Inivata Ltd (AL, SS, KH, DG, NR, EG and VP) are employees, officers and/or share-holders of Inivata Ltd. Inivata Ltd commercialises assays based on the technology described in this paper. All remaining authors have declared no conflicts of interest.

References

1. Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res 2015; 5(9): 2892–2911.
2. Reguart N, Remon J. Common EGFR-mutated subgroups (Del19/L858R) in advanced non-small-cell lung cancer: chasing better outcomes with tyrosine-kinase inhibitors. Future Oncol 2015; 1–13.
3. Sequist LV, Waltman BA, Dias-Santagata D et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. Sci Transl Med 2011; 3(75): 75ra26.
4. Yu HA, Arcila ME, Rekhtman N et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res 2013; 19(8): 2240–2247.
5. Jänne PA, Yang JC-H, Kim D-W et al. AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. N Engl J Med 2015; 372(18): 1689–1699.
6. http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm301431.htm (30 December 2016, date last accessed).
7. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/004124/WC500202022.pdf (30 December 2016, date last accessed).
8. Remon J, Faivre L, Facchinetti F et al. Radiogenomics in 332 metastatic non-small cell lung cancer (NSCLC) patients. ASCO Meeting Abstracts 2016; 34(15 Suppl): 11563.
9. Piotrowska Z, Nierdest MJ, Mino-Kenudson M et al. Variation in mechanisms of acquired reistantce among EGFR-mutant NSCLC patients with more than 1 postresistance biopsy. Int J Radiat Oncol 2014; 90: S6–S7.
10. Jovelet C, Ileana E, Le Deley M-C et al. Circulating cell-free tumor DNA analysis of 50 genes by next-generation sequencing in the Prospective MOSCATO Trial. Clin Cancer Res 2016; 22(12): 2960–2968.
11. Murtaza M, Dawson S-J, Pogrebniak K et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. Nat Commun 2015; 6: 8760.
12. Jackman D, Pao W, Riely GJ et al. Clinical definition of acquired resistance to epidermal growth factor receptor tyrosine kinase inhibitors in non-small-cell lung cancer. J Clin Oncol 2010; 28(2): 357–360.
13. Gale D, Plagnol V, Lawson A et al. Analytical performance and validation of an enhanced TAm-Seq circulating tumor DNA sequencing assay. AACR 2016, New Orleans. Cancer Res 2016; 76 (Suppl 14): 3639.
14. Forshew T, Murtaza M, Parkinson C et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012; 4(136): 136ra68.
15. Therasse P, Arbuck SG, Eisenhauer EA et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment Of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000; 92(3): 205–216.
16. Mok TS, Wu Y-L, Ahn M-J et al. Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. N Engl J Med 2016. doi:10.1056/NEJMoa1612674.
17. Oxnard GR, Thress KS, Alden RS et al. Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. J Clin Oncol 2016; 34(28): 3375–3382.
18. Marchetti A, Palma JF, Felicioni L et al. Early prediction of response to tyrosine kinase inhibitors by quantification of EGFR mutations in plasma of NSCLC patients. J Thorac Oncol 2015; 10(10): 1437–1443.
19. Mok T, Wu Y-L, Lee JS et al. Detection and dynamic changes of EGFR mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC patients treated with first-line Intercalated erlotinib and chemotherapy. Clin Cancer Res 2015; 21(14): 3196–3203.
20. Sorensen BS, Wu L, Wei W et al. Monitoring of epidermal growth factor receptor tyrosine kinase inhibitor-sensitizing and resistance mutations in the plasma DNA of patients with advanced non-small cell lung cancer during treatment with erlotinib. Cancer 2014; 120(24): 3896–3901.
21. https://www.qiagen.com/gb/shop/detection-solutions/personalized-healthcare/therascreen-egfr-plasma-rgq-pcr-kit-emea/ (30 December 2016, date last accessed).

22. http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm504540.htm (30 December 2016, date last accessed).
23. Sequist LV, Soria J-C, Goldman JW et al. Rociletinib in EGFR-mutated non-small-cell lung cancer. N Engl J Med 2015; 372(18): 1700–1709.
24. Sundaresan TK, Sequist LV, Heymach JV et al. Detection of T790M, the acquired resistance EGFR mutation, by tumor biopsy versus noninvasive blood-based analyses. Clin Cancer Res 2016; 22(5): 1103–1110.
25. Reckamp KL, Melnikova VO, Karlovich C et al. A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. J Thorac Oncol 2016; 11(10): 1690–1700.
26. Thress KS, Brant R, Carr TH et al. EGFR mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. Lung Cancer 2015; 90(3): 509–515.
27. Hata A, Katakami N, Yoshioka H et al. Spatiotemporal T790M heterogeneity in individual patients with EGFR-mutant non-small-cell lung cancer after acquired resistance to EGFR-TKI. J Thorac Oncol 2015; 10(11): 1553–1559.
28. Lim C, Sung M, Shepherd FA et al. Patients with advanced non-small cell lung cancer: are research biopsies a barrier to participation in clinical trials? J Thorac Oncol 2016; 11(1): 79–84.
29. Chabon JJ, Simmons A, Newman AM et al. Inter- and intra-patient heterogeneity of resistance mechanisms to the mutant EGFR selective inhibitor rociletinib. ASCO Meeting Abstracts 2016; 34(15 suppl): 9000.

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer

Sarah-Jane Dawson, F.R.A.C.P., Ph.D., Dana W.Y. Tsui, Ph.D.,
Muhammed Murtaza, M.B., B.S., Heather Biggs, M.A.,
Oscar M. Rueda, Ph.D., Suet-Feung Chin, Ph.D., Mark J. Dunning, Ph.D.,
Davina Gale, B.Sc., Tim Forshew, Ph.D., Betania Mahler-Araujo, M.D.,
Sabrina Rajan, M.D., Sean Humphray, B.Sc., Jennifer Becq, Ph.D.,
David Halsall, M.R.C.Path., Ph.D., Matthew Wallis, M.B., Ch.B.,
David Bentley, D.Phil., Carlos Caldas, M.D., F.Med.Sci.,
and Nitzan Rosenfeld, Ph.D.

ABSTRACT

BACKGROUND

The management of metastatic breast cancer requires monitoring of the tumor burden to determine the response to treatment, and improved biomarkers are needed. Biomarkers such as cancer antigen 15-3 (CA 15-3) and circulating tumor cells have been widely studied. However, circulating cell-free DNA carrying tumor-specific alterations (circulating tumor DNA) has not been extensively investigated or compared with other circulating biomarkers in breast cancer.

METHODS

We compared the radiographic imaging of tumors with the assay of circulating tumor DNA, CA 15-3, and circulating tumor cells in 30 women with metastatic breast cancer who were receiving systemic therapy. We used targeted or whole-genome sequencing to identify somatic genomic alterations and designed personalized assays to quantify circulating tumor DNA in serially collected plasma specimens. CA 15-3 levels and numbers of circulating tumor cells were measured at identical time points.

RESULTS

Circulating tumor DNA was successfully detected in 29 of the 30 women (97%) in whom somatic genomic alterations were identified; CA 15-3 and circulating tumor cells were detected in 21 of 27 women (78%) and 26 of 30 women (87%), respectively. Circulating tumor DNA levels showed a greater dynamic range, and greater correlation with changes in tumor burden, than did CA 15-3 or circulating tumor cells. Among the measures tested, circulating tumor DNA provided the earliest measure of treatment response in 10 of 19 women (53%).

CONCLUSIONS

This proof-of-concept analysis showed that circulating tumor DNA is an informative, inherently specific, and highly sensitive biomarker of metastatic breast cancer. (Funded by Cancer Research UK and others.)

From the Department of Oncology, University of Cambridge and Cancer Research UK Cambridge Institute, Li Ka Shing Centre (S.-J.D., D.W.Y.T., M.M., O.M.R., S.-F.C., M.J.D., D.G., T.F., C.C., N.R.), the Departments of Histopathology (B.M.-A.), Radiology (S.R., M.W.), and Clinical Biochemistry and Immunology (D.H.) and the Cambridge Breast Unit (S.-J.D., H.B., B.M.-A., S.R., M.W., C.C.), Addenbrooke's Hospital, Cambridge University Hospital National Health Service Foundation Trust and National Institute for Health Research Cambridge Biomedical Research Centre, and the Cambridge Experimental Cancer Medicine Centre (C.C.), Cambridge; and Illumina, Little Chesterford (S.H., J.B., D.B.) — all in the United Kingdom; and the Peter MacCallum Cancer Centre, East Melbourne, VIC, Australia (S.-J.D.). Address reprint requests to Dr. Rosenfeld or Dr. Caldas at Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, United Kingdom, or at carlos.caldas@cruk.cam.ac.uk.

Drs. Dawson and Tsui and Drs. Caldas and Rosenfeld contributed equally to this article.

This article was published on March 13, 2013, at NEJM.org.

N Engl J Med 2013;368:1199-209.
DOI: 10.1056/NEJMoa1213261
Copyright © 2013 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

Breast cancer is the most common cancer and the leading cause of cancer-related death in women worldwide.[1] Metastatic breast cancer remains an incurable disease but is treatable by means of serial administration of endocrine, cytotoxic, or biologic therapies. The monitoring of treatment response is essential to avoid continuing ineffective therapies, to prevent unnecessary side effects, and to determine the benefit of new therapeutics. Treatment response is generally assessed with the use of serial imaging, but radiographic measurements often fail to detect changes in tumor burden. Therefore, there is an urgent need for biomarkers that measure tumor burden with high sensitivity and specificity.

Cancer antigen 15-3 (CA 15-3) is a serum biomarker that is clinically useful in some patients with metastatic breast cancer but has a sensitivity of only 60 to 70%.[2-4] The enumeration of circulating tumor cells has emerged as a promising biomarker. Although there are numerous methods to detect circulating tumor cells in the research setting,[5-7] the CellSearch System is the only test approved by the Food and Drug Administration. The system has a sensitivity of approximately 65% for detecting circulating tumor cells (≥1 cell per 7.5 ml of blood) in patients with metastatic breast cancer.[8,9] Elevated levels of circulating tumor cells (defined as ≥5 cells per 7.5 ml of blood) have been associated with a worse prognosis.[8,10]

Circulating DNA fragments carrying tumor-specific sequence alterations (circulating tumor DNA) are found in the cell-free fraction of blood, representing a variable and generally small fraction of the total circulating DNA.[11,12] Advances in sequencing technologies have enabled the rapid identification of somatic genomic alterations in individual tumors, and these can be used to design personalized assays for the monitoring of circulating tumor DNA. Studies have shown the feasibility of using circulating tumor DNA to monitor tumor dynamics in a limited number of patients with various solid cancers, but few cases of breast cancer have been analyzed.[13-20] Here, we provide a direct comparison between circulating tumor DNA and other circulating biomarkers (CA 15-3 and circulating tumor cells) and medical imaging, the current standard of care, for the noninvasive monitoring of metastatic breast cancer.

METHODS

PATIENTS AND SAMPLE COLLECTION

We carried out a prospective, single-center study to compare the sensitivity of measuring circulating tumor DNA, CA 15-3, and circulating tumor cells for monitoring tumor burden in patients with metastatic breast cancer (see the Supplementary Appendix, available with the full text of this article at NEJM.org). The study was approved by the local institutional research ethics committee.

Eligible patients were women with metastatic breast cancer currently undergoing active treatment. A total of 52 women were recruited, and 30 had genomic alterations suitable for monitoring. All women provided written informed consent. Serial blood samples (30 ml each) were collected between April 2010 and April 2012 at intervals of 3 or more weeks. Computed tomography (CT) was performed and reviewed in a blinded fashion to document response to treatment according to the Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1.[21] All reagents and equipment used in the study were purchased.

IDENTIFICATION OF SOMATIC GENOMIC ALTERATIONS

Sequencing was performed on DNA from breast-cancer specimens and matched normal tissue specimens, with the use of one or both of two methods: tagged-amplicon deep sequencing[22] for *PIK3CA* (encoding the phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha protein) and *TP53* (encoding tumor protein p53) or paired-end whole-genome sequencing (see the Supplementary Appendix). Tagged-amplicon deep sequencing was done by means of the Fluidigm Access Array and sequencing on the Illumina GAIIx or HiSeq instruments. Paired-end sequencing was done with the use of the Illumina HiSeq2000 instrument. Candidate mutations and structural variants were validated and confirmed to be somatic with the use of Sanger sequencing.

ISOLATION AND QUANTIFICATION OF CIRCULATING TUMOR DNA

Blood samples that were collected in EDTA tubes were processed within 1 hour after collection and were centrifuged to separate the plasma from the peripheral-blood cells. DNA was extracted from

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

aliquots (2 ml) of plasma with the use of the QIAamp circulating nucleic acid kit (Qiagen). To measure the DNA carrying specific somatic genomic alterations in plasma, we carried out a microfluidic digital polymerase-chain-reaction (PCR) assay[17,23-25] (using the Fluidigm BioMark system) or direct plasma sequencing by means of tagged-amplicon deep sequencing[22] (using the Fluidigm Access Array and sequencing on the Illumina HiSeq2500 instrument) (see the Supplementary Appendix).

ASSAY OF CA 15-3 AND CIRCULATING TUMOR CELLS

We measured levels of CA 15-3 in aliquots (50 μl) of plasma by means of the ADVIA Centaur immunoassay system (Siemens Healthcare). Blood samples were collected in CellSave Preservative Tubes (Veridex) and were processed within 96 hours for the enumeration of circulating tumor cells with the use of the CellSearch System (Veridex). The counting of circulating tumor cells was performed in a manner blinded to the results of CT and assessments of CA 15-3 or circulating tumor DNA.

STATISTICAL ANALYSIS

To estimate the sensitivity of each of the circulating biomarkers, we used a modified bootstrapping method.[26] We randomly sampled the complete data set to obtain a new data set containing only one time point for each patient. This random sampling was repeated 1000 times to obtain 1000 data sets, each containing independent observations. For each data set, we calculated the sensitivity of each biomarker. The median sensitivity for each biomarker and the median difference in sensitivity between two biomarkers — circulating tumor DNA versus either CA 15-3 or circulating tumor cells — was then calculated across the 1000 data sets. The percentile method was used to obtain 95% confidence intervals.

Survival analysis was performed by fitting a different Cox regression model for each of the three variables of interest: circulating tumor DNA, circulating tumor cells, and CA 15-3. Each model was constructed with the use of the counting process notation (start, end, event),[27] such that for each time period, the date of the visit was taken as the start, and the date before the next visit (or the date of last follow-up) was considered the end. The predictors were modeled as time-dependent covariates that use splines to account for nonlinear relationships. Estimated survival curves were produced for different values of the covariates at the first visit. Wald statistic P values were reported for each model, and relative hazard plots were computed for each covariate, showing the linear predictor relative to the mean value of the covariate (for details, see the Supplementary Appendix).

RESULTS

IDENTIFICATION OF SOMATIC GENOMIC ALTERATIONS

Clinical details, results of CT imaging, and serial whole-blood samples were collected prospectively from 52 women undergoing therapy for metastatic breast cancer (Fig. 1, and Table S1 in the Supplementary Appendix). DNA extracted from archival-tumor tissue samples was analyzed to identify somatic genomic alterations, with the use of two approaches. First, we used targeted deep sequencing to screen for point mutations in *PIK3CA* and *TP53*,[28] which we identified in 25 of the 52 patients (Table S2 in the Supplementary Appendix). Second, we used whole-genome paired-end sequencing of tumor-tissue specimens and matched normal-tissue specimens in 9 of the 52 patients. We identified somatic structural variants[29] in 8 patients (Table S3 in the Supplementary Appendix), including 5 in whom no mutations were previously identified in *PIK3CA* or *TP53*, bringing the total number of patients with identified genomic alterations to 30 of 52 women (Fig. 1, and Fig. S1 in the Supplementary Appendix). In 3 patients, both mutations and structural variants were identified, enabling us to compare and contrast the use of point mutations[13] and structural variants[14,15] for serial monitoring of circulating tumor DNA. For 1 patient, we used whole-genome paired-end sequencing to identify multiple somatic mutations, enabling us to monitor multiple mutations in parallel in circulating tumor DNA (Table S2 in the Supplementary Appendix).

QUANTIFICATION OF CIRCULATING TUMOR DNA IN PLASMA

In the 30 women with somatic mutations or structural variants, circulating tumor DNA was quantified in a total of 141 serial plasma samples

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

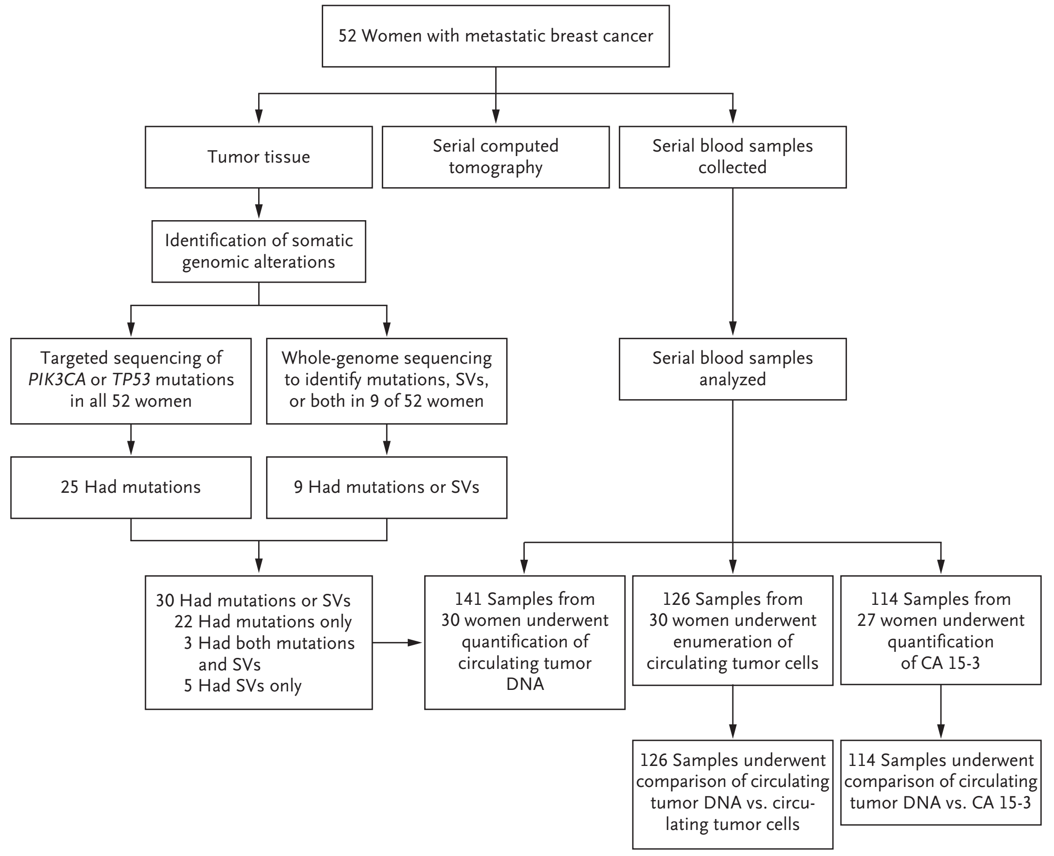

52 Women with metastatic breast cancer
Tumor tissue
Serial computed tomography
Serial blood samples collected
Identification of somatic genomic alterations
Targeted sequencing of *PIK3CA* or *TP53* mutations in all 52 women
Whole-genome sequencing to identify mutations, SVs, or both in 9 of 52 women
Serial blood samples analyzed
25 Had mutations
9 Had mutations or SVs
30 Had mutations or SVs
22 Had mutations only
3 Had both mutations and SVs
5 Had SVs only
141 Samples from 30 women underwent quantification of circulating tumor DNA
126 Samples from 30 women underwent enumeration of circulating tumor cells
114 Samples from 27 women underwent quantification of CA 15-3
126 Samples underwent comparison of circulating tumor DNA vs. circulating tumor cells
114 Samples underwent comparison of circulating tumor DNA vs. CA 15-3

Figure 1. Enrollment of Patients and Collection of Clinical Samples.

In the 30 women who were found to have somatic mutations, structural variants (SVs), or both, the genomic alterations were determined through targeted deep sequencing or whole-genome paired-end sequencing of tumor-tissue specimens and matched normal-tissue specimens. CA 15-3 denotes cancer antigen 15-3.

by means of either digital PCR assay or tagged-amplicon deep sequencing.

Digital PCR assay was performed in 97 plasma samples from 19 of the 30 patients to track both somatic mutations and structural variants. The sensitivity of digital PCR assay allowed for the detection of a mutant allele fraction of 0.1% or more (one mutant molecule in a background of 1000 wild-type molecules) (Fig. S2 in the Supplementary Appendix).[17] Circulating tumor DNA was detected in 18 of the 19 women and in 80 of the 97 plasma samples (82%) analyzed.

As a high-throughput alternative to digital PCR assay, the remaining 44 plasma samples from the remaining 11 patients were analyzed with the use of tagged-amplicon deep sequencing.[22] The sensitivity of tagged-amplicon deep sequencing allowed for the detection of a mutant allele fraction of 0.14% or more with a confidence margin of 0.95.[22] Using this approach, circulating tumor DNA was identified in all 11 patients and in 35 of the 44 plasma samples (80%) analyzed.

In a subset of plasma samples in which circulating tumor DNA was analyzed by both techniques, quantification of mutant allele fraction by means of either tagged-amplicon deep sequencing or digital PCR assay showed excellent agreement (Fig. S3 in the Supplementary Appendix).[22] Taken together, circulating tumor DNA was detected in 29 of the 30 women (97%) and in 115 of the 141 plasma samples (82%). The median quantity of circulating tumor DNA across all samples was 150 amplifiable copies per milliliter of plasma (interquartile range, 9 to 720) (Table S4 in the Supplementary Appendix). The median mutant allele fraction was 4% (interquar-

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

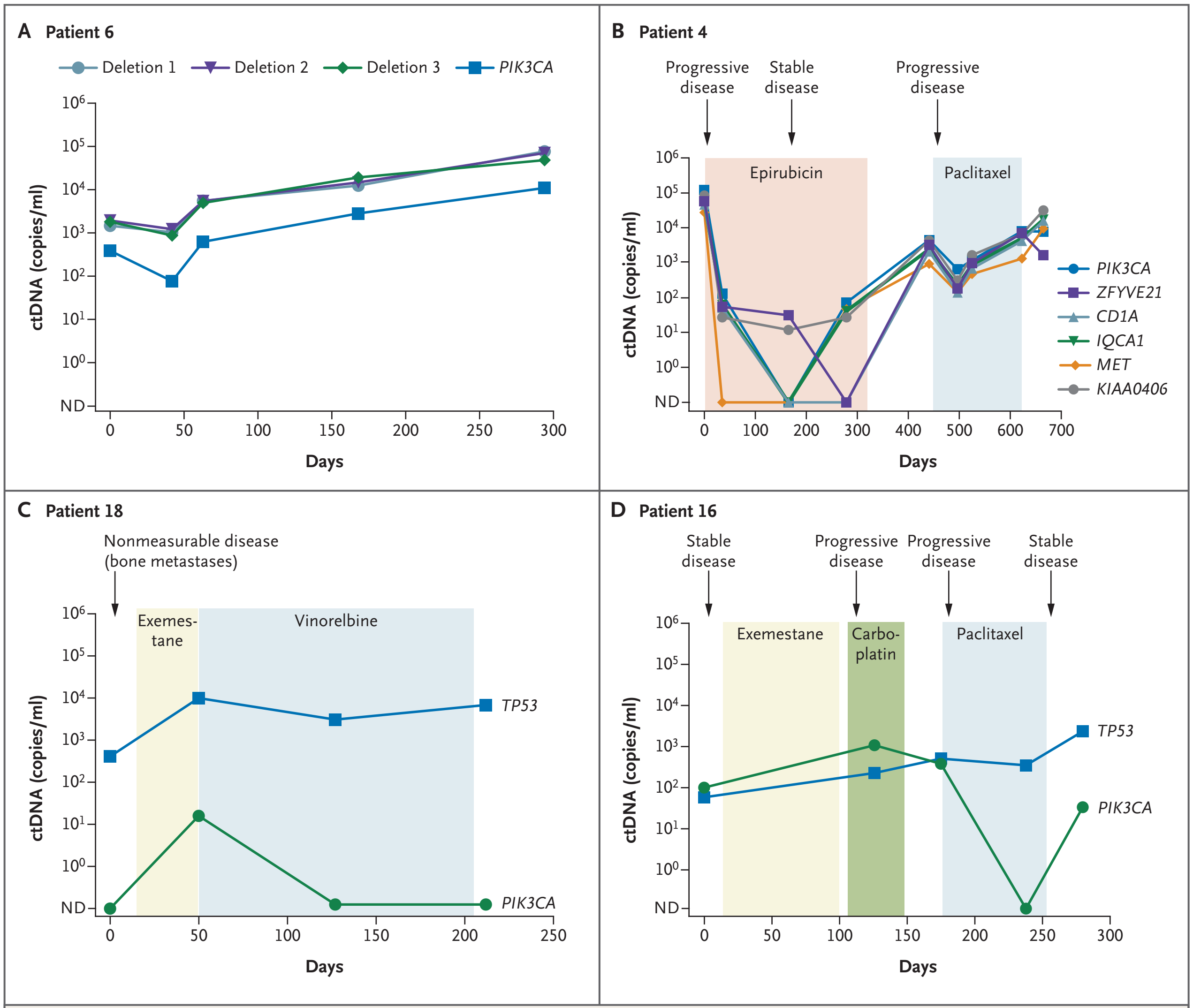


Figure 2. Monitoring Multiple Point Mutations and Structural Variants in Circulating DNA.

Panels A, B, and C show plasma levels of circulating tumor DNA (ctDNA) for three patients (one per panel), quantified in parallel by means of a digital polymerase-chain-reaction (PCR) assay across multiple time points. In Panels B, C, and D, the use of endocrine or cytotoxic therapy is indicated by colored shading, and disease status at various times (as ascertained on computed tomography) is shown. Panel A shows three structural variants (deletions) and a point mutation in *PIK3CA*. The three deletions occurred in the setting of a complex rearrangement associated with amplification. Panel B shows six point mutations, all of which showed similar dynamic patterns. Panel C shows point mutations in *PIK3CA* and *TP53*; the *TP53* mutation was dominant in the circulation as compared with the *PIK3CA* mutation. Panel D shows plasma levels of ctDNA for a fourth patient, with point mutations in *PIK3CA* and *TP53* quantified by means of tagged-amplicon deep sequencing. The *TP53* mutation was identified in plasma only, and levels remained elevated after paclitaxel chemotherapy despite a fall in the *PIK3CA* mutation level in the presence of stable disease. ND denotes not detected.

tile range, 1 to 14). The 1 patient in whom circulating tumor DNA was not detected (Patient 12) had a low burden of metastatic disease (small-volume mediastinal lymphadenopathy) and no evidence of disease progression during the study. Overall, levels of total plasma DNA were measured in parallel and had limited informative content (Fig. S4 in the Supplementary Appendix).

CONCURRENT MONITORING OF MULTIPLE SOMATIC GENOMIC ALTERATIONS IN PLASMA

Plasma levels of either mutations or structural variants identified in the tumor tissue of the same patient (Fig. S1C in the Supplementary Appendix) showed a similar dynamic pattern (Fig. 2A, and Table S4 in the Supplementary Appendix). This confirmed the utility and comparability of both

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

approaches. In women with tumors in which the genomic location of the structural variants overlapped with an amplified locus, such alterations were detected in the plasma at higher concentrations, confirming that the assay of circulating tumor DNA is quantitative (Fig. 2A, and Fig. S1B and Table S5 in the Supplementary Appendix).

When multiple mutations were identified in tumor-tissue samples (Fig. S1C in the Supplementary Appendix), they generally showed similar dynamic patterns in plasma (Fig. 2B, and Table S4 in the Supplementary Appendix). However, in some cases, we also observed evidence of clonal heterogeneity, whereby certain mutations dominated in the plasma (Fig. 2C, and Table S4 in the Supplementary Appendix). Tagged-amplicon deep sequencing also identified mutations in plasma that were not detected in archival-tumor DNA (Fig. S1C in the Supplementary Appendix).[22] In these cases, the archival primary tissue had been collected more than 10 years previously, and the discordance may have reflected tumor evolution.[30,31] These mutations showed diverging patterns over the course of disease progression and treatment (Fig. 2D, and Table S4 in the Supplementary Appendix), as compared with the mutations identified in the tumor, suggesting that they originated from different subclones.

SENSITIVITY OF CIRCULATING TUMOR DNA, CA 15-3, AND CIRCULATING TUMOR CELLS

Data comparing CA 15-3 values and circulating tumor DNA levels were available across 114 serial time points for 27 patients (Fig. 3A, and Table S4 in the Supplementary Appendix). CA 15-3 levels were elevated (>32.4 U per milliliter) at one or more time points in 21 of the 27 women (78%) and in 71 of the 114 samples (62%). In contrast, circulating tumor DNA was detected in 26 of 27 women (96%) and in 94 of 114 samples (82%). Of the 43 samples without elevated CA 15-3 levels, 27 (63%) had measurable levels of circulating tumor DNA. Using a modified bootstrapping method, we showed improved sensitivity of circulating tumor DNA as compared with CA 15-3 (85% vs. 59%), with a median difference in sensitivity of 26% (95% confidence interval [CI], 11 to 37; P<0.002).

Circulating tumor cells were quantified by means of the CellSearch System at 126 time points for all 30 women (Fig. 3B, and Table S4 in the Supplementary Appendix). Circulating tumor cells (≥1 cell per 7.5 ml of blood) were detected at one or more time points in 26 of the 30 women (87%), and elevated circulating tumor cells (≥5 cells per 7.5 ml of blood) were identified in 18 of the 30 women (60%). Of the 126 samples, 50 (40%) had no detected circulating tumor cells, and 76 (60%) had 1 or more cells per 7.5 ml, of which 46 (37% of all 126 samples) had 5 or more cells per 7.5 ml. In contrast, circulating tumor DNA was detected in 29 of the 30 women (97%) and at 106 of 126 time points (84%). In the 50 samples in which no circulating tumor cells were detected, 33 (66%) had measurable levels of circulating tumor DNA. According to the modified bootstrapping method, circulating tumor DNA had sensitivity superior to that of circulating tumor cells (90% vs. 67%), with a median difference in sensitivity of 27% (95% CI, 13 to 37; P<0.002). At the median, the number of amplifiable copies of circulating tumor DNA was 133 times the number of circulating tumor cells and had a greater dynamic range (Fig. 3B).

CT AND CIRCULATING BIOMARKERS FOR TUMOR MONITORING

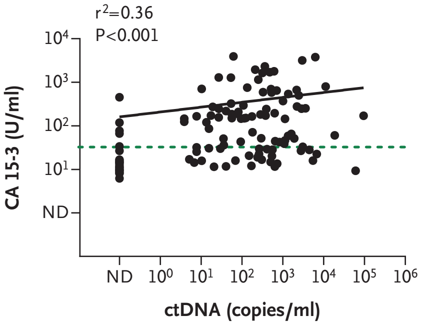
We compared the performance of circulating biomarkers with the performance of CT in 20 patients with measurable disease (as defined by RECIST[21]) and for whom circulating biomarker data were available at 3 or more time points over a period of more than 100 days of follow-up (Fig. S5 in the Supplementary Appendix). Circulating tumor DNA was detected and showed serial changes in 19 of 20 women (95%) with fluctuations in circulating tumor DNA generally correlating with treatment responses seen on imaging (Fig. 4A, and Fig. S5 in the Supplementary Appendix). Similar findings were noted for women with 5 or more circulating tumor cells per 7.5 ml of blood (10 of 20 patients [50%]) in which serial changes in circulating tumor cell counts were evident and corresponded with responses ascertained on CT (Fig. 4A). However, in the remaining 10 women with a maximal count of circulating tumor cells of fewer than 5 cells per 7.5 ml of blood, the number of circulating tumor cells was uninformative (Fig. 4B and 4C, and Fig. S5 in the Supplementary Appendix).

Similar to the findings regarding circulating tumor cells was the finding that women with high levels of CA 15-3 had fluctuations corresponding to responses on imaging but with a smaller dynamic range (Fig. 4A and 4B, and Fig.

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

A CA 15-3 vs. ctDNA

CA 15-3	ctDNA		
	Detected	Not detected	Total
Patients			
Elevated (>32.4 U/ml)	21	0	21
Not elevated (≤32.4 U/ml)	5	1	6
Total	26	1	27
ctDNA sensitivity, 26/27 (96%) CA 15-3 sensitivity, 21/27 (78%)			
Samples			
Elevated (>32.4 U/ml)	67	4	71
Not elevated (≤32.4 U/ml)	27	16	43
Total	94	20	114
ctDNA sensitivity, 94/114 (82%) CA 15-3 sensitivity, 71/114 (62%)			



B CTC vs. ctDNA

CTC	ctDNA		
	Detected	Not detected	Total
Patients			
Elevated (≥5)	18	0	18
Detected (1–4)	7	1	8
Not detected (0)	4	0	4
Total	29	1	30
ctDNA sensitivity, 29/30 (97%) CTC sensitivity (detected, >0), 26/30 (87%) CTC sensitivity (elevated, ≥5), 18/30 (60%)			
Samples			
Elevated (≥5)	45	1	46
Detected (1–4)	28	2	30
Not detected (0)	33	17	50
Total	106	20	126
ctDNA sensitivity, 106/126 (84%) CTC sensitivity (detected, >0), 76/126 (60%) CTC sensitivity (elevated, ≥5), 46/126 (37%)			
Median ratio of ctDNA copy numbers (per 3.75 ml of plasma) to number of CTCs (per 7.5 ml of whole blood)=133 (interquartile range, 51–528)			

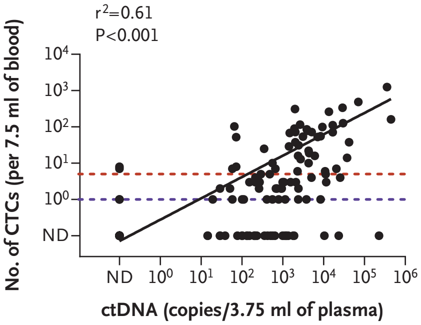


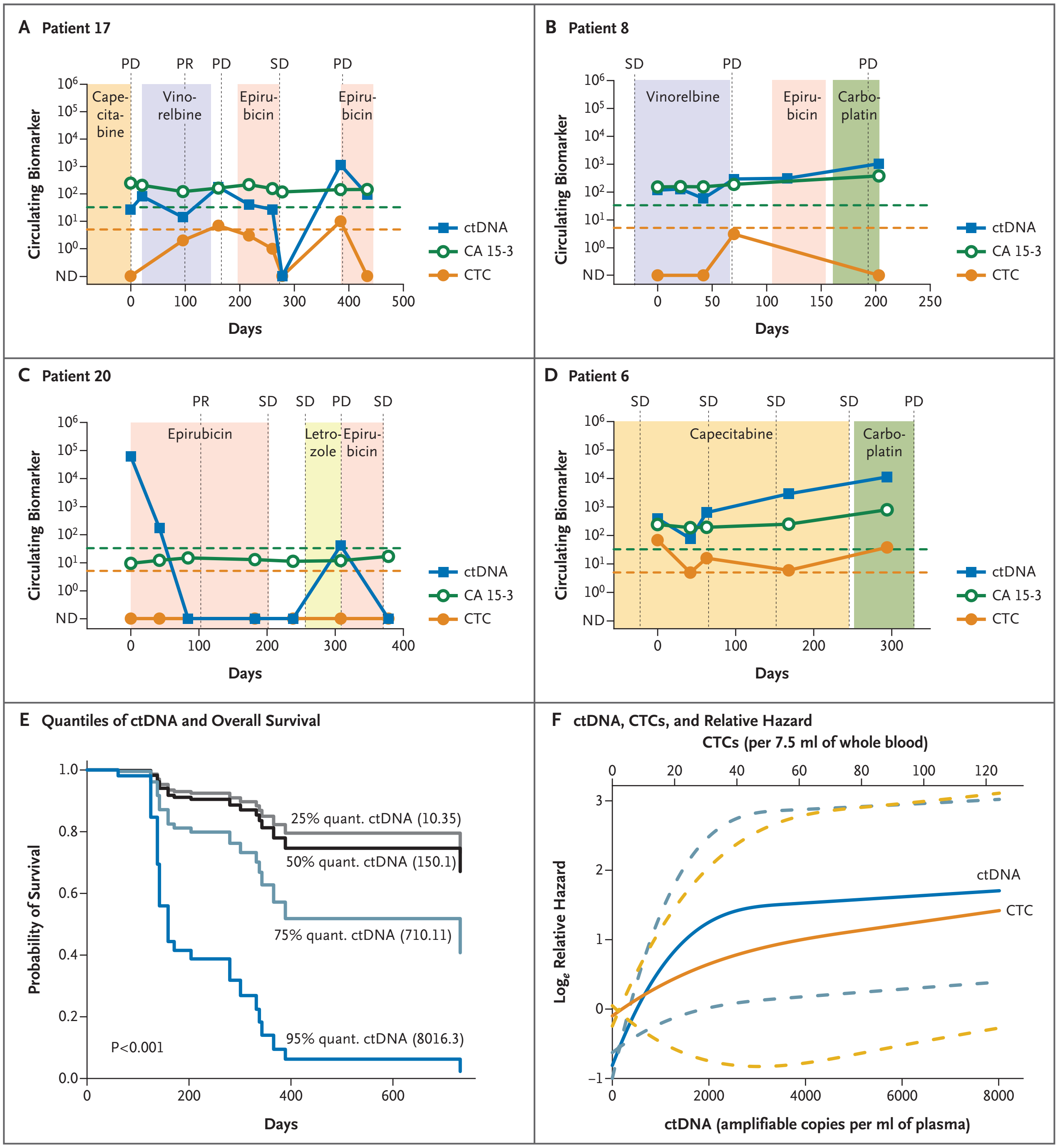
Figure 3. Comparison of Circulating Tumor DNA, CA 15-3, and Circulating Tumor Cells as Blood-Based Biomarkers.

Panel A shows comparisons of CA 15-3 levels (U per milliliter of plasma) and circulating tumor DNA (ctDNA) levels (amplifiable copies per milliliter of plasma) across the maximal value analyzed for individual patients and across all samples analyzed for all patients. The green horizontal dashed line indicates the CA 15-3 threshold of 32.4 U per milliliter. The Spearman correlation coefficient (r) between CA 15-3 levels and ctDNA levels across all time points was 0.36 (P<0.001). Panel B shows comparisons of circulating tumor cell (CTC) numbers (per 7.5 ml of whole blood) and ctDNA numbers (amplifiable copies per 3.75 ml of plasma) across the maximal value analyzed for individual patients and across all samples analyzed for all patients. Copy numbers of ctDNA were adjusted for direct comparison to the numbers of circulating tumor cells from an equivalent volume of whole blood (7.5 ml). The purple dashed line indicates the CTC threshold of 1 cell per 7.5 ml of blood, and the orange dashed line indicates the CTC threshold of 5 cells per 7.5 ml of blood. The Spearman correlation coefficient (r) between quantified ctDNA levels and numbers of CTCs across all time points was 0.61 (P<0.001). ND denotes not detected.

S5 in the Supplementary Appendix). In patients with levels of CA 15-3 of 50 U or less per milliliter (8 of 19 patients [42%]), no consistent serial changes in CA 15-3 levels were seen (Fig. 4C, and Fig. S5 in the Supplementary Appendix).

Progressive disease was documented on CT (as defined by RECIST) in 19 of 20 women during the follow-up period; CA 15-3 data were available for 18 of these women (95%) (Fig. S5 in the Supplementary Appendix). Increases in circulating tumor DNA levels reflected progressive disease in 17 of the 19 women (89%). In these women, on average, circulating tumor DNA levels increased by a factor of 505 (range, 2 to 4457) from the nadir before the establishment of progressive disease. The numbers of circulating tu-

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.



mor cells increased in 7 of the 19 women (37%), and CA 15-3 levels increased in 9 of 18 women (50%) (Fig. S5 in the Supplementary Appendix). In 10 of the 19 patients (53%), levels of circulating tumor DNA increased at one or more consecutive time points, on average 5 months (range, 2 to 9) before the establishment of progressive disease by means of imaging (Fig. 4D, and Fig. S5 in the Supplementary Appendix). In 2 women (Patients 9 and 22), increasing levels of circulating tumor DNA did not reflect the presence of progressive disease as assessed on CT (a detailed description

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

Figure 4 (facing page). Comparison of Circulating Biomarkers to Monitor Tumor Dynamics and Predict Survival.

Panels A, B, C, and D show serial circulating tumor DNA (ctDNA) levels (number of copies per milliliter of plasma), circulating tumor cell (CTC) numbers (per 7.5 ml of whole blood), CA 15-3 levels (U per milliliter), and disease status as ascertained on computed tomography (vertical dashed lines) for four patients (one in each panel). Details of endocrine or cytotoxic therapy are indicated by colored shading. The orange dashed line indicates the threshold of 5 CTCs per 7.5 ml of whole blood. The green dashed line indicates the CA 15-3 threshold of 32.4 U per milliliter. ND denotes not detected, PD progressive disease, PR partial response, and SD stable disease. Panel E shows the results of a Cox regression model, which identified an inverse relationship between quantiles (quant.) of ctDNA (indicated in copies per milliliter of plasma) and overall survival, with increasing levels significantly associated with poor overall survival ($P<0.001$). At 200, 400, and 600 days, a total of 23, 8, and 3 patients were at risk, respectively. Panel F shows that increasing ctDNA levels (copies per milliliter), as indicated on the bottom x axis, and increasing numbers of CTCs (per 7.5 ml of whole blood), as indicated on the top x axis, were associated with an increased $\log_e$ relative hazard. The prognostic discrimination power of circulating tumor DNA level was greatest with levels up to 2000 copies per milliliter. Patients with levels of more than 2000 copies per milliliter were uniformly found to have the worst prognosis. The prognostic power of CTCs increased according to the number of cells. Dashed lines represent 95% confidence intervals.

of these patients is provided in Fig. S5 in the Supplementary Appendix).

PROGNOSTIC USE OF CIRCULATING BIOMARKERS

Finally, we compared the circulating biomarkers with respect to prognostic use. Using a Cox proportional-hazards model in which circulating tumor DNA was treated as a continuous time-dependent variable, we found that increasing levels of circulating tumor DNA were associated with inferior overall survival ($P<0.001$) (Fig. 4E). Circulating tumor cells were also found to have prognostic significance ($P=0.03$) (Fig. S6A in the Supplementary Appendix). In contrast, CA 15-3 was not found to be prognostic in this series of patients (Fig. S6B in the Supplementary Appendix). Increasing numbers of circulating tumor cells and increasing levels of circulating tumor DNA were associated with an increased hazard (Fig. 4F), indicating that absolute levels of each is informative in guiding prognosis.

DISCUSSION

In the detection of metastatic breast cancer, circulating tumor DNA shows superior sensitivity to that of other circulating biomarkers and has a greater dynamic range that correlates with changes in tumor burden. Circulating tumor DNA often provides the earliest measure of treatment response, as has been supported by recent analyses of circulating tumor DNA in other solid cancers.[20,32]

The monitoring of circulating tumor DNA levels requires the identification of somatic alterations in individual patients. Future developments will reduce the cost of whole-genome paired-end sequencing, and targeted sequencing can be readily expanded to include other genes, in addition to *PIK3CA* and *TP53*, known to be recurrently mutated in breast cancer.[33-35] Here we have demonstrated the use of two strategies to quantify circulating tumor DNA: digital PCR assay and targeted deep sequencing. Digital PCR assay provides high accuracy and sensitivity but requires the design of personalized assays, an expensive and rate-limiting step. Targeted deep sequencing of plasma DNA provides a cost-effective alternative for high-throughput analysis and may overcome limitations of initial tumor-tissue assessment by virtue of allowing for the direct identification of mutations in plasma.[22] However, our findings on circulating tumor DNA are not limited to these molecular platforms. Other methods for the identification of somatic mutations (such as exome sequencing[33]) or for the quantification of circulating tumor DNA (e.g., BEAMing [beads, emulsions, amplification, and magnetics] technology[13] or Safe-SeqS [Safe-Sequencing System][36]) may be applied with even greater sensitivity. Recent studies have also shown the feasibility of performing genomewide analysis of tumor-associated copy-number changes and mutations in plasma.[37-39]

Our expanding knowledge of the genetic mechanisms underpinning breast cancer now provides a framework to better stratify patients.[30,33-35,40,41] The analysis of circulating tumor DNA represents a unique opportunity to integrate this knowledge into the clinical arena. Although the acquisition of tumor-tissue specimens will continue to be important, the use of biopsy specimens is limited, since such material may

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

not capture tumor heterogeneity; in addition, repeated biopsy is impractical. Circulating tumor DNA represents a "liquid biopsy" alternative, allowing for sensitive and specific serial sampling to be performed during the course of treatment.

Supported by grants from Cancer Research UK, and the Experimental Cancer Medicine Centre and National Institute for Health Research Cambridge Biomedical Research Centre, and by an Australian National Health and Medical Research Council–R.G. Menzies Early Career Fellowship (to Dr. Dawson).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Linda Jones and Susan Richardson for recruiting patients into the study; all the medical and ancillary staff in the breast-cancer clinic and the patients for consenting to participate; Sonia Bradbury from the Department of Clinical Biochemistry and Immunology, Addenbrooke's Hospital, for assistance with the CA 15-3 analysis; the Genomics, Histopathology, Bioinformatic, and Biorepository Core Facilities at the Cancer Research UK Cambridge Institute; the Addenbrooke's Human Research Tissue Bank (supported by the National Institute for Health Research Cambridge Biomedical Research Centre); and Sarah Dawson and Sarah Vowler for their contribution to the statistical analysis.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer 2010; 127:2893-917.

2. Lauro S, Trasatti L, Bordin F, et al. Comparison of CEA, MCA, CA 15-3 and CA 27-29 in follow-up and monitoring therapeutic response in breast cancer patients. Anticancer Res 1999;19:3511-5.

3. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287-312.

4. Duffy MJ, Evoy D, McDermott EW. CA 15-3: uses and limitation as a biomarker for breast cancer. Clin Chim Acta 2010; 411:1869-74.

5. Lianidou ES, Markou A. Circulating tumor cells in breast cancer: detection systems, molecular characterization, and future challenges. Clin Chem 2011;57: 1242-55.

6. Nagrath S, Sequist LV, Maheswaran S, et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature 2007;450:1235-9.

7. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. Nat Rev Cancer 2008;8:329-40.

8. Cristofanilli M, Budd GT, Ellis MJ, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 2004;351:781-91.

9. Mego M, De Giorgi U, Dawood S, et al. Characterization of metastatic breast cancer patients with nondetectable circulating tumor cells. Int J Cancer 2011;129: 417-23.

10. Pierga JY, Hajage D, Bachelot T, et al. High independent prognostic and predictive value of circulating tumor cells compared with serum tumor markers in a large prospective trial in first-line chemotherapy for metastatic breast cancer patients. Ann Oncol 2012;23:618-24.

11. Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer 2011;11: 426-37.

12. Gormally E, Caboux E, Vineis P, Hainaut P. Circulating free DNA in plasma or serum as biomarker of carcinogenesis: practical aspects and biological significance. Mutat Res 2007;635:105-17.

13. Diehl F, Schmidt K, Choti MA, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2008;14:985-90.

14. Leary RJ, Kinde I, Diehl F, et al. Development of personalized tumor biomarkers using massively parallel sequencing. Sci Transl Med 2010;2:20ra14.

15. McBride DJ, Orpana AK, Sotiriou C, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. Genes Chromosomes Cancer 2010;49: 1062-9.

16. Chen X, Bonnefoi H, Diebold-Berger S, et al. Detecting tumor-related alterations in plasma or serum DNA of patients diagnosed with breast cancer. Clin Cancer Res 1999;5:2297-303.

17. Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. Clin Cancer Res 2009;15:2076-84.

18. Nakamura T, Sueoka-Aragane N, Iwanaga K, et al. A noninvasive system for monitoring resistance to epidermal growth factor receptor tyrosine kinase inhibitors with plasma DNA. J Thorac Oncol 2011;6:1639-48.

19. Otsuka J, Okuda T, Sekizawa A, et al. Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. Int J Gynecol Cancer 2004;14:459-64.

20. Diaz LA Jr, Williams RT, Wu J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 2012;486:537-40.

21. Eisenhauer EA, Therasse P, Bogaerts J, et al. New Response Evaluation Criteria in Solid Tumours: revised RECIST guideline (version 1.1). Eur J Cancer 2009;45:228-47.

22. Forshew T, Murtaza M, Parkinson C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012;4:136ra68.

23. Vogelstein B, Kinzler KW. Digital PCR. Proc Natl Acad Sci U S A 1999;96: 9236-41.

24. Wang J, Ramakrishnan R, Tang Z, et al. Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. Clin Chem 2010;56: 623-32.

25. Lo YM, Lun FM, Chan KC, et al. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc Natl Acad Sci U S A 2007;104:13116-21.

26. Efron B, Tibshirani R. An introduction to the bootstrap. London: Chapman & Hall/CRC, 1993.

27. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat 1982;10:1100-20.

28. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 2011; 39:D945-D950.

29. Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature 2009;462:1005-10.

30. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395-9.

31. Higgins MJ, Jelovac D, Barnathan E, et al. Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. Clin Cancer Res 2012;18:3462-9.

32. Misale S, Yaeger R, Hobor S, et al. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature 2012;486:532-6.

33. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400-4.

34. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405-9.

35. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490: 61-70.

36. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

massively parallel sequencing. Proc Natl Acad Sci U S A 2011;108:9530-5.

37. Chan KC, Jiang P, Zheng YW, et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. Clin Chem 2013;59:211-24.

38. Shaw JA, Page K, Blighe K, et al. Genomic analysis of circulating cell-free DNA infers breast cancer dormancy. Genome Res 2012;22:220-31.

39. Leary RJ, Sausen M, Kinde I, et al. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. Sci Transl Med 2012;4:162ra154.

40. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 2012;486:353-60.

41. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346-52.

Copyright © 2013 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org at CAMBRIDGE UNIVERSITY LIBRARY on March 16, 2018. For personal use only. No other uses without permission.
Copyright © 2013 Massachusetts Medical Society. All rights reserved.

Research Article

TRANSPARENT PROCESS OPEN ACCESS

EMBO Molecular Medicine

Dynamics of multiple resistance mechanisms in plasma DNA during EGFR-targeted therapies in non-small cell lung cancer

Dana Wai Yi Tsui[1,2,§,†], Muhammed Murtaza[1,2,3,†,¶,††], Alvin Seng Cheong Wong[4], Oscar M Rueda[1,2], Christopher G Smith[1,2], Dineika Chandrananda[1,2], Ross A Soo[4,5], Hong Liang Lim[6], Boon Cher Goh[4,5], Carlos Caldas[1,2,3,7], Tim Forshew[1,2,‡‡], Davina Gale[1,2], Wei Liu[1,2,§§], James Morris[1,2], Francesco Marass[1,2,¶¶,†††], Tim Eisen[3,7,8], Tan Min Chin[4,5,9,‡,*] & Nitzan Rosenfeld[1,2,‡,**]

Abstract

Tumour heterogeneity leads to the development of multiple resistance mechanisms during targeted therapies. Identifying the dominant driver(s) is critical for treatment decision. We studied the relative dynamics of multiple oncogenic drivers in longitudinal plasma of 50 *EGFR*-mutant non-small-cell lung cancer patients receiving gefitinib and hydroxychloroquine. We performed digital PCR and targeted sequencing on samples from all patients and shallow whole-genome sequencing on samples from three patients who underwent histological transformation to small-cell lung cancer. In 43 patients with known EGFR mutations from tumour, we identified them accurately in plasma of 41 patients (95%, 41/43). We also found additional mutations, including *EGFR* T790M (31/50, 62%), *TP53* (23/50, 46%), *PIK3CA* (7/50, 14%) and *PTEN* (4/50, 8%). Patients with both *TP53* and *EGFR* mutations before treatment had worse overall survival than those with only *EGFR*. Patients who progressed without T790M had worse PFS during TKI continuation and developed alternative alterations, including small-cell lung cancer-associated copy number changes and *TP53* mutations, that tracked subsequent treatment responses. Longitudinal plasma analysis can help identify dominant resistance mechanisms, including non-druggable genetic information that may guide clinical management.

Keywords circulating tumour DNA; liquid biopsy; lung cancer; resistance mechanisms; targeted therapy

Subject Categories Cancer; Pharmacology & Drug Discovery; Respiratory System

DOI 10.15252/emmm.201707945 | Received 25 April 2017 | Revised 4 April 2018 | Accepted 9 April 2018

EMBO Mol Med (2018) 10: e7945

Introduction

Molecularly targeted therapies offer substantial clinical benefit in a subset of patients whose tumours harbour specific oncogenic drivers. Unfortunately, treatment resistance inevitably develops, partly driven by the evolving genetic landscape of cancer cells. For example, though non-small-cell lung cancer (NSCLC) patients carrying activating mutations in *EGFR* (epidermal growth factor receptor) initially respond to EGFR-targeted tyrosine kinase inhibitors (EGFR-TKIs; Lynch *et al*, 2004; Paez *et al*, 2004), the emergence of

---

1 Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Cambridge, UK
2 Cancer Research UK Major Center - Cambridge, Cambridge, UK
3 Department of Oncology, University of Cambridge, Cambridge, UK
4 Department of Haematology-Oncology, National University Cancer Institute, National University Health System, Singapore, Singapore
5 Cancer Science Institute, Centre for Translational Medicine, National University of Singapore, Singapore, Singapore
6 Parkway Cancer Center, Singapore, Singapore
7 Department of Oncology, Addenbrooke's Hospital, Cambridge University Health Partners, Cambridge, UK
8 Oncology Early Clinical Development, AstraZeneca, Cambridge, UK
9 Raffles Cancer Centre, Raffles Hospital, Singapore, Singapore
*Corresponding author. Tel: +65 63112306; E-mail: csictm@nus.edu.sg
**Corresponding author. Tel: +44 1223 769769; E-mail: nitzan.rosenfeld@cruk.cam.ac.uk

[†] These authors contributed equally to this work

[‡] These authors should be considered as (co-)senior authors and project co-leaders

[§] Present address: Department of Pathology, Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[¶] Present address: Center for Noninvasive Diagnostics, Translational Genomics Research Institute, Phoenix, AZ, USA

[††] Present address: Mayo Clinic, Center for Individualized Medicine, Scottsdale, AZ, USA

[‡‡] Present address: Inivata Ltd., Granta Park, Cambridge, UK

[§§] Present address: University of Glasgow, Glasgow, UK

[¶¶] Present address: Department of Biosystems Science and Engineering, ETH Zurich, Basel, Switzerland

[†††] Present address: SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

© 2018 The Authors. Published under the terms of the CC BY 4.0 license

mutations that confer resistance to these TKIs or activate alternative drivers (such as EGFR T790M, MET/HER2 amplifications, PIK3CA mutation) leads to eventual drug resistance (Yu *et al*, 2013; Camidge *et al*, 2014). Some of these resistance mechanisms are targetable, such as T790M (Janne *et al*, 2015; Sequist *et al*, 2015) and MET amplification (Sierra & Tsao, 2011). Apart from such individual genetic changes, a small subset of EGFR-mutant NSCLC patients develop resistance to EGFR-TKI therapy by undergoing histological transformation to small-cell lung cancer (SCLC) and become sensitive to standard SCLC treatment (Sequist *et al*, 2011; Niederst *et al*, 2015). Therefore, longitudinal monitoring of the dynamic genetic changes during the course of a patient's treatment has become increasingly important to guide treatment at progression or when resistance occurs. Plasma circulating tumour DNA (ctDNA) is a non-invasive method that has been used to identify EGFR mutations and other genetic drivers in NSCLC and in response to treatment of NSCLC patients with EGFR-TKIs (Yung *et al*, 2009; Couraud *et al*, 2014; Douillard *et al*, 2014; Newman *et al*, 2014, 2016; Weber *et al*, 2014; Paweletz *et al*, 2015; Wan *et al*, 2017). During treatment of NSCLC patients with first-generation EGFR-TKIs, serial assessment of EGFR mutations in plasma ctDNA has proved successful in allowing early detection of T790M-driven resistance prior to radiographic progression (Oxnard *et al*, 2014; Mok *et al*, 2015). However, a subset of the patients develop resistance that is independent of the EGFR pathway, and multiple resistance mechanisms may co-exist because of tumour heterogeneity (Sequist *et al*, 2015; Abbosh *et al*, 2017). Here, we performed longitudinal analysis of plasma ctDNA to study the dynamics of co-existing multiple resistance mechanisms during sequential therapy in NSCLC patients.

In this study, we analysed a cohort of 392 plasma samples collected longitudinally from 50 Stage IV NSCLC patients. All were treated with the first-generation TKI gefitinib in combination with hydroxychloroquine as part of the "Hydroxychloroquine and Gefitinib to Treat Lung Cancer" trial (NCT00809237). Thirty-four patients were TKI-naïve (i.e. not previously treated with EGFR-TKI), and 16 were TKI-treated (i.e. previously treated with TKI with a 2-week washout period). Eligibility for the trial and patient characteristics are summarized in the Appendix Supplementary Methods. This is a phase II study with a phase I lead in that studies the tolerability, safety profile and efficacy of hydroxychloroquine and gefitinib in advanced non-small-cell lung cancer. Appendix Fig S1 summarizes the number of patients in each arm (Appendix Fig S1). We performed tagged-amplicon deep sequencing (TAm-Seq; Forshew *et al*, 2012) for *de novo* identification and quantification of mutations in *EGFR* exons 18–21, coding regions of *TP53* and *PTEN*, and selected hotspot regions of *PIK3CA*, *KRAS* and *BRAF*; and digital PCR for detection and quantification of hotspot mutations in *EGFR*. For a subset of patients, we also performed shallow whole-genome sequencing to analyse global copy number changes during treatment (Heitzer *et al*, 2013).

Results

Mutational profiling by plasma DNA

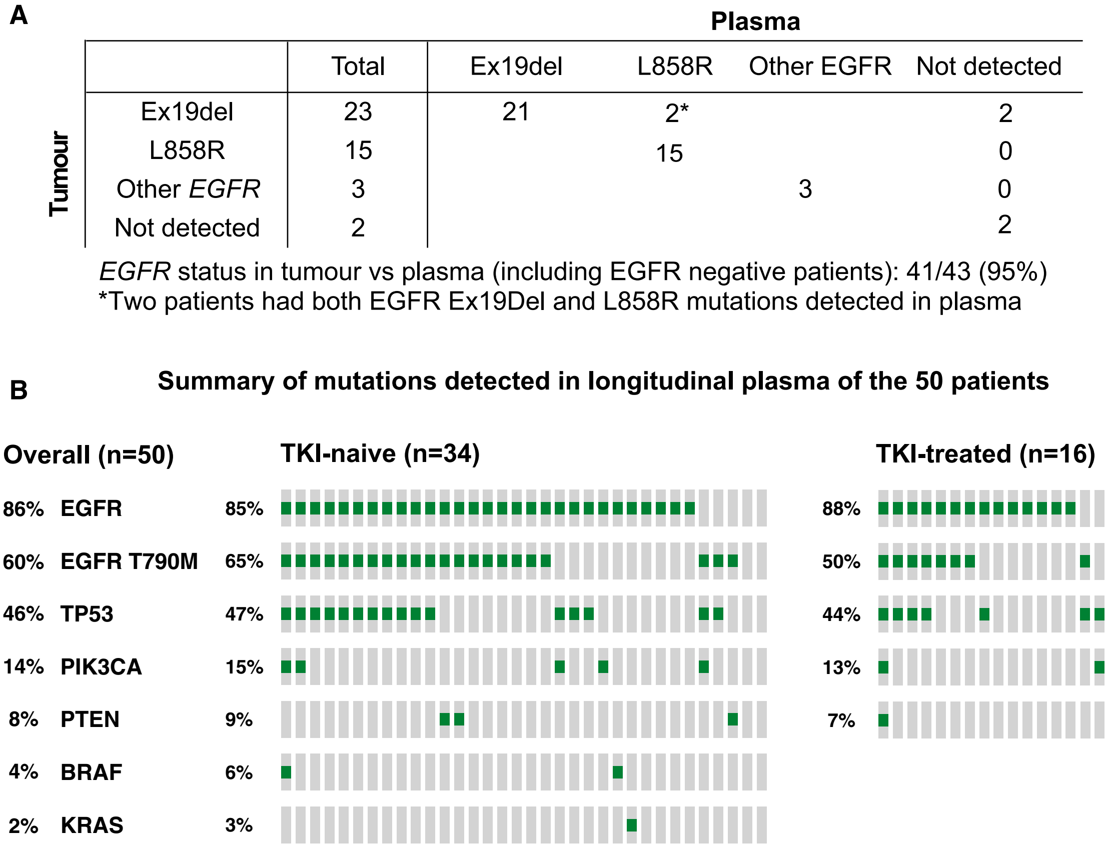
To determine whether plasma was a good surrogate of *EGFR* mutation status in the tumour, we compared the *EGFR* mutation status in plasma samples (as determined by our assays) with the tumour status reported in hospital records. The *EGFR* status was known in the tumour of 43 of the 50 patients, and we detected the same *EGFR* mutation in any follow-up plasma samples of 41 of 43 (95%) patients (Fig 1A and Appendix Table S1). In the remaining seven patients, two were found to be *EGFR* wild-type in both tumour and plasma, and the remaining five have *EGFR* mutations detected in plasma. In 24 patients who responded to the treatment within the initial 70 days, 19 of them showed a drop in *EGFR* cfDNA levels within that period (Appendix Fig S2 and Appendix Table S2). In addition to *EGFR*, somatic mutations in other cancer genes, such as in *TP53* or the PI3K/AKT/mTOR pathway (*PIK3CA* and *PTEN*), were also identified in the plasma of 29 patients (Fig 1B). Of the identified mutations, 25–43% are likely oncogenic (*TP53*, 10/23, 43%; *PIK3CA*, 3/7, 43%; *PTEN*, 1/4, 25%) according to OncoKB annotation (Chakravarty *et al*, 2017). To further compare molecular profiles between tumour and plasma, we studied paired tumour and plasma samples in four patients, where tumour samples were available before and after disease progression. The types of EGFR mutations identified in plasma and tumour (*EGFR* activating, resistance-conferring mutations in *EGFR* and other mutations) were identical for 11 of 12 (92%) mutations before treatment, and for 9 of 12 (75%) mutations after treatment (Appendix Fig S3 and Appendix Table S3). Plasma captured the same or more mutations than tumour in 23 of 24 cases (96%). These results confirmed that plasma analysis is informative for mutation profiling in NSCLC patients using our assays. Initial changes in EGFR ctDNA levels after start of treatment mirrored in most cases the radiographic assessment of clinical response.

Prognostic value of baseline plasma DNA

We studied the relationship between pre-treatment EGFR ctDNA levels and prognosis in 19 TKI-naïve patients (Appendix Table S4), for which at least one plasma sample was collected before initiation of treatment. Patients with low levels of *EGFR*-activating mutations pre-treatment tended to have better progression-free survival (PFS) and overall survival (OS; Fig 2A and B), though this did not reach statistical significance level of 0.05 (their corresponding Cox *P*-values were 0.06 for both PFS and OS). Of note, patients with low levels of *EGFR*-activating mutation allele fractions had reduced tumour burden (median 17 mm) by RECIST measurements, as compared to those with intermediate (median 42 mm) and high (median 80 mm) levels of *EGFR*-activating mutation (Appendix Table S4). These findings suggest that baseline mutation concentrations in the plasma correlate with tumour burden. In addition, patients with both *EGFR* and *TP53* detected in pre-treatment plasma tended to have worse prognosis (Fig 2C and D, Cox *P*-value 0.109 for PFS and 0.035 for OS). We repeated the analysis with copies/ml instead of mutant allele fractions, and the conclusions were the same. These data suggest that both the molecular profile of genomic alterations, and the quantification of ctDNA levels in baseline plasma, can have prognostic implications.

Mutation dynamics in plasma DNA reveals heterogeneous resistance mechanisms

For 45 of 50 patients, EGFR mutations were detected before treatment in tumour and/or plasma and more than one plasma sample



A

Tumour	Total	Plasma: Ex19del	L858R	Other EGFR	Not detected
Ex19del	23	21	2*		2
L858R	15		15		0
Other *EGFR*	3			3	0
Not detected	2				2

EGFR status in tumour vs plasma (including EGFR negative patients): 41/43 (95%)
*Two patients had both EGFR Ex19Del and L858R mutations detected in plasma

Figure 1. Summary of somatic mutations identified in the 50 NSCLC patients.

A Detection of tumour EGFR mutations in plasma. EGFR mutation status in tumour samples was documented in the clinical record for 43 patients (Appendix Table S1), of which 38 had verified hotspot activating mutations (deletion in exon 19 for 23 patients and the L858R mutation for 15 patients), three patients had other mutations in *EGFR* (one of these patients had two different mutations detected in the tumour sample), and two patients were wild-type for *EGFR* according to tumour analysis and confirmed by plasma analysis.

B Summary of the mutations identified in any of the plasma samples during longitudinal follow-up in the 50 patients. TKI-naïve (n = 34) and TKI-treated (n = 16) patients are presented separately.

was available from clinical follow-up. Longitudinal analysis of ctDNA in plasma revealed heterogeneity of resistance mechanisms (Fig 3A). During longitudinal follow-up, a large subset of patients retained the sensitizing mutation and developed resistance-conferring *EGFR* T790M mutation (n = 28/45, 62%, Fig 3B). To estimate the detection lead time (i.e. the interval between detection of the resistance-conferring mutation in plasma and radiographic evidence of disease progression), we focus on 28 patients where T790M was detected in plasma at any time during EGFR-TKI, including detection before disease progression became evident. In patients treated with first-line EGFR-TKI, we found that the median time-to-appearance of T790M in plasma was 4 months from the start of TKI treatment, with a lead time between T790M detection and clinical progression of 6.8 months. Patients with *EGFR* T790M can now be treated with third-generation, irreversible EGFR-TKIs (Janne *et al*, 2015; Piotrowska *et al*, 2015). One patient (220) had a biopsy of the lung tumour after progression, in which both activating *EGFR* exon 19 and T790M mutations were detected. The same mutations were detected in plasma at the time of progression. This patient was then treated with a third-generation EGFR-TKI (EGF-816, Novartis (NCT02108964)) and demonstrated partial radiological response. Subsequent plasma samples showed no further EGFR mutations (data shown in Dataset EV1).

In a second group of patients (n = 10/45, 22%), the activating *EGFR* mutation was detected in plasma before and after progression, with an average mutant allele fraction (AF, i.e. the fractional concentration of mutant allele over total DNA) of 7.9%, but not T790M (Fig 3C). The continued presence of activating mutations in plasma suggests possible positive selection of the mutations in the *EGFR* pathway in the corresponding cancers. In these patients, mutations in other pathways also emerged in plasma, such as *TP53* and *PIK3CA* (Dataset EV1). One possible hypothesis is that tumours of patients in this group may retain partial sensitivity to EGFR-TKI treatment, and may respond clinically if EGFR-TKI is used in combination with treatments targeting additional resistance pathway.

The third group of patients (n = 7/45, 15%) did not have EGFR-activating nor known resistance-conferring mutations in EGFR detected in plasma when they progressed. These patients initially had exon 19 deletion detected in the tumour (7/7) and their first plasma sample (6/7). Interestingly, comparing to the other two groups, this group of patients had EGFR-activating mutations

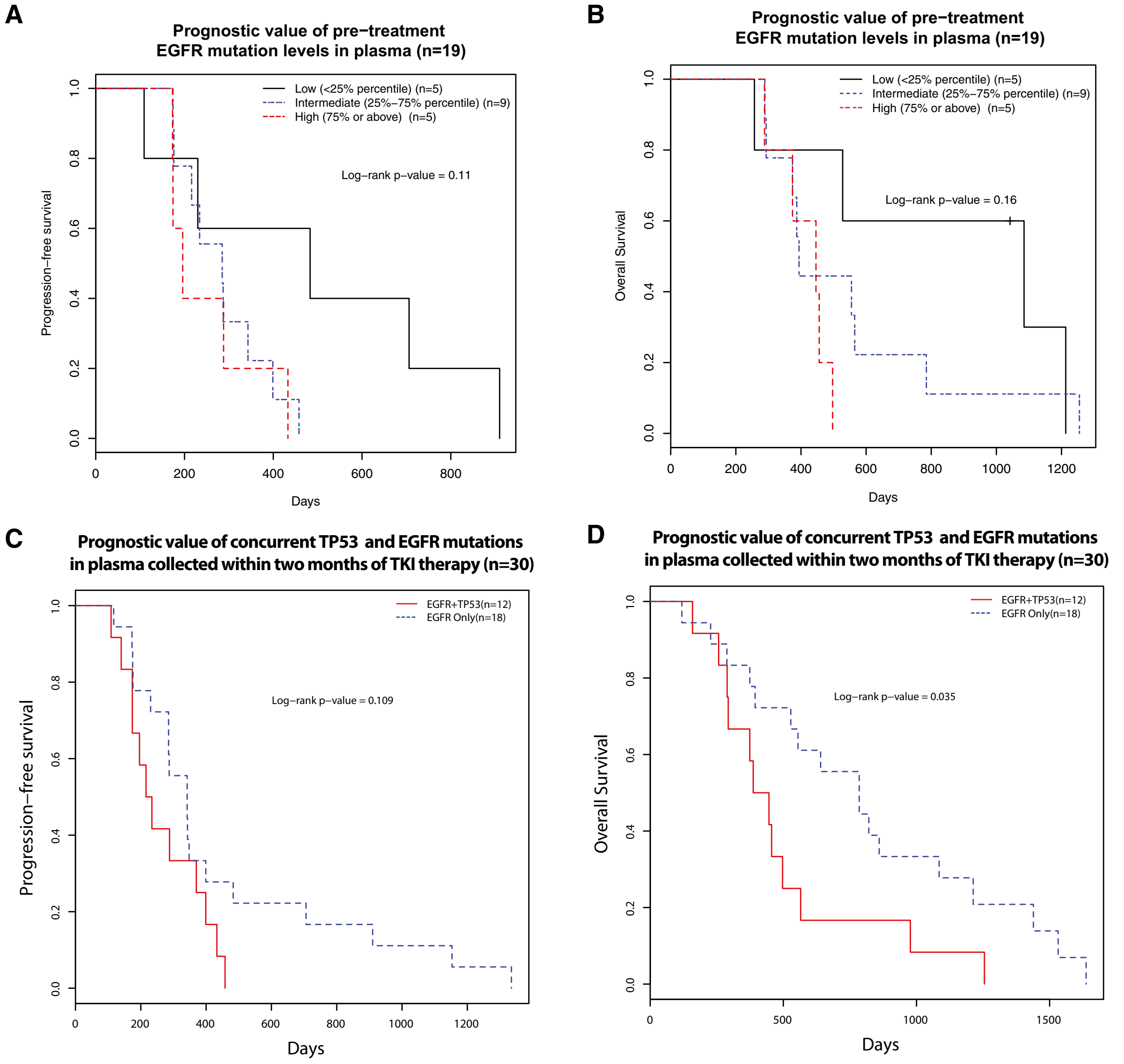


Figure 2. Prognostic value of qualitative and quantitative assessments of pre-treatment ctDNA.

A, B The relationship of pre-treatment *EGFR*-activating mutation levels (allele fractions) with progression-free survival (PFS) and overall survival (OS) of 19 first-line TKI-treated patients where baseline plasma samples (collected before the start of treatment) were available. Patients were grouped into three groups according to their pre-treatment ctDNA levels, as measured by *EGFR*-activating mutation allele fractions: low (< 25% quartile), intermediate (25–75% quartile) and high (> 75% quartile) ctDNA levels. Kaplan–Meier survival curves indicated that patients with high baseline pre-treatment *EGFR*-activating mutant allele fractions were non-significantly associated with unfavourable (A) PFS (log-rank *P*-value = 0.11) and (B) OS (log-rank *P*-value = 0.16), Cox *P*-value of 0.06 for either PFS or OS.

C, D The prognostic value of concurrent TP53 and EGFR mutations in pre-treatment plasma samples before EGFR-TKI therapy. This analysis was performed in 30 first-line EGFR-TKI patients where plasma samples were available within 2 months of start of treatment. The presence of both TP53 and EGFR mutations in plasma was associated with a trend of worse PFS (log-rank *P*-value = 0.109, hazard ratio and 95% confidence interval: 0.53 [0.24–1.17]) and significantly worse OS (log-rank *P*-value = 0.035, hazard ratio and 95% confidence interval: 0.43 [0.20–0.97]).

present at relatively lower allele fractions in their first plasma samples [groups 1 and 2: median EGFR mutations mutant allele fractions was 3% (range: 0.07–65.7%) versus group 3: median 0.23% (range: 0.06–2.11%)]. We do not rule out the possibility that the tumours of these patients might release less tumour-derived DNA into the circulation. In some of these patients, we detected alternative cancer mutations such as *TP53* and *PIK3CA* in plasma before treatment was initiated, and the levels of these mutations then increased to present the highest allele fractions in ctDNA when disease progressed (Fig 3D). We speculate that one possible explanation for the absence of EGFR mutations in cfDNA at disease progression could be that, EGFR mutations were subclonal in those patients initially, and under the selective pressure of the EGFR-targeting therapy, the EGFR-driven clones shrank below detection limit of the assay, while clones that were driven by alterative drivers (such as TP53 and PIK3CA) and did not carry the EGFR-sensitizing mutations, expanded. Based on our data from cfDNA, these alternative drivers pre-existed even before treatment initiation, but were present in parts of the tumour that were not analysed, or alternatively were present at very low cellularity such that

© 2018 The Authors

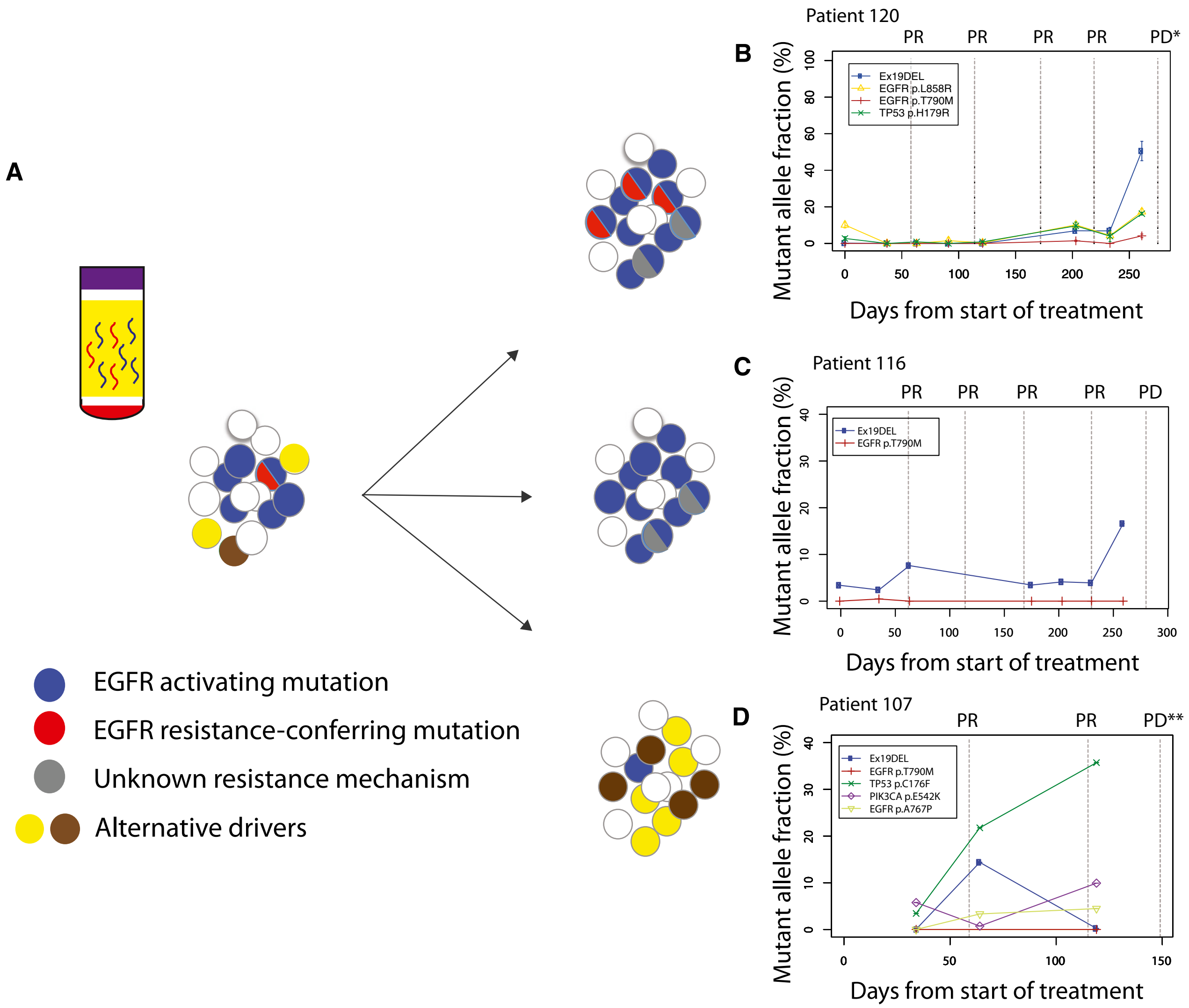


Figure 3. Longitudinal analysis of ctDNA dynamics reveals distinct patterns of resistance mechanisms.

A Longitudinal analysis of ctDNA dynamics in 45 NSCLC patients revealed three main groups of concurrent heterogeneous resistance mechanisms.

B In the first group (n = 28/45, 62%), patients retained EGFR-sensitizing mutations before and after disease progression, with the development of T790M in their plasma samples, indicating that at least some of the progressing clones developed resistance to TKI by acquiring T790M.

C In the second group (n = 10/45, 22%), patients retained EGFR-sensitizing mutations but progressed without developing T790M in their plasma samples, suggesting that resistance arose due to other mechanisms which were not analysed in this dataset.

D In the third group (n = 7/45, 15%), patients progressed without EGFR-sensitizing nor resistance-conferring T790M mutations detected in their plasma samples. Resistance possibly develops through dependence on alternative cancer driver pathways.

Data information: For patients where multiple *EGFR*-activating mutations were identified in plasma, only the most abundant one is shown here (complete data for all patients are shown in Dataset EV1). Clinical progression and CT imaging times are indicated with a dotted line, with RECIST classification: SD, stable disease; PR, partial response; PD, progressive disease. Progressive disease defined by presentation of symptoms on brain or bone scan is indicated by PD**.

their allele fractions in those samples were below the detection limit by standard clinical tumour sequencing assay. Recent data from tumour sequencing suggested EGFR may be subclonal in a small subset of EGFR-mutant NSCLC tumour (McGranahan *et al*, 2015), which agrees with our hypothesis. A recent plasma-based study also reported 4 out of 24 NSCLC patients had EGFR-sensitizing mutations detected in plasma at T0 but absent when the patients progressed, which agreed with our findings (Pecuchet *et al*, 2016). The fact that EGFR T790M was not detected in the third group suggested that the cancers have developed resistance mechanisms that are alterative to the EGFR pathway, which agreed with the observations of alternative drivers in plasma in some patients. For the patients whom we did not detect T790M nor other drivers were detected, they may have progressed due to other resistance mechanisms that were not covered by our targeted sequencing assay.

Of note, overall, TKI-naïve patients who progressed without T790M detection (13 patients selected from the second and the third groups) had a significantly worse PFS (Appendix Fig S4A and Appendix Table S5, P-value = 0.008), and a trend towards worse OS

© 2018 The Authors

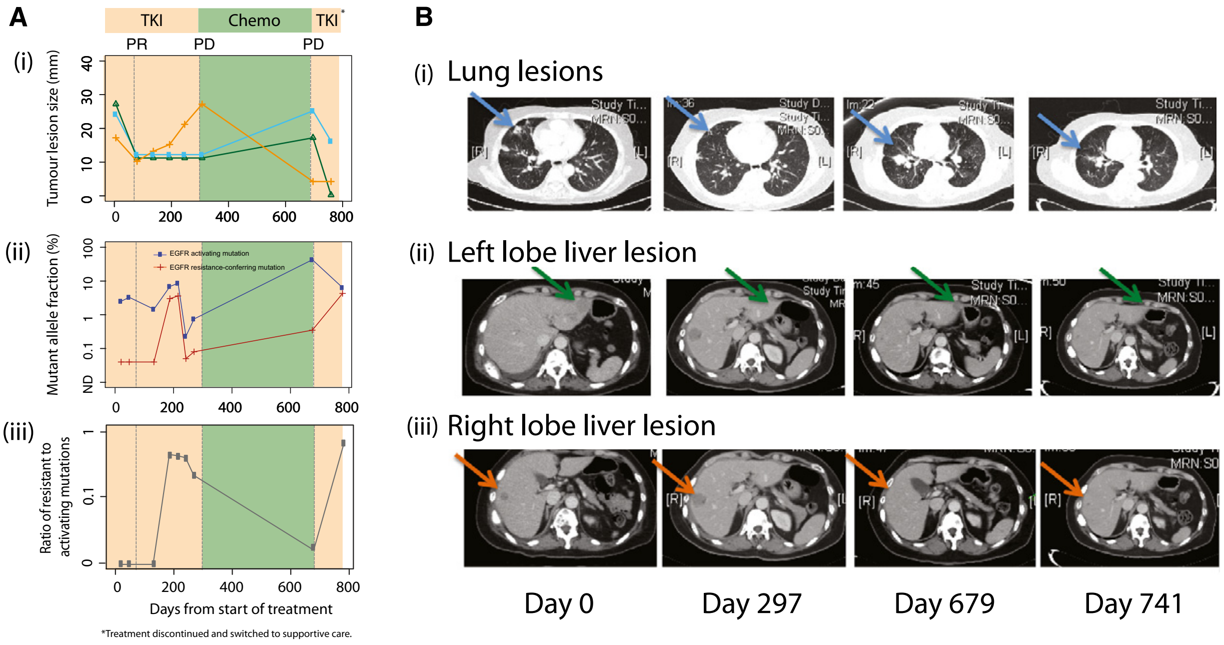
(Appendix Fig S4B, *P*-value = 0.22), as compared to T790M-positive patients. These observations confirm results from a re-biopsy study (Oxnard *et al*, 2011), suggesting that patients that progress without T790M are less likely to benefit from TKI continuation, possibly due to the more aggressive nature of their disease or reduced dependency on the EGFR pathway.

Dynamics of EGFR mutations in plasma across multiple lines of treatment

To explore the utility of monitoring both activating and resistance-conferring mutations, we investigated their relative representation in plasma in relation to radiographic assessment in one patient for whom samples were available spanning sequential lines of first-generation EGFR-TKI and chemotherapy (Fig 4). Three lesions were tracked by imaging (Fig 4B): in the lung (L1), left liver lobe (L2) and right liver lobe (L3). During treatment with TKI, all three lesions initially shrank, before a small incremental increase in the size of L3 from day 49, although this did not amount to RECIST progression. The mutant AF of EGFR-activating mutation (exon 19 deletion) decreased initially but increased from day 134 (Fig 4A). The T790M mutation was detected from day 189.

On day 297, L3 showed a substantial growth. The T790M mutation at that timepoint reached AF of 3.6%. Treatment was changed to platinum-based doublet chemotherapy at which point L3 showed a reduction on serial CT imaging, which coincided with a drop of the T790M mutation AF to 0.7% (Fig 4A). At day 679, despite the shrinkage of L3, L1 and L2 grew, coinciding with an increase in the AF of the activating mutation from 8.6% (when responding to initial TKI) to 43%. The patient was given a first-generation EGFR-TKI re-challenge (newer TKIs were not yet approved at the time this patient was treated), and activating mutation sharply dropped back to 6.4%, corresponding to a reduction in L1 and L2.

From day 217 to 244, both EGFR mutations exhibited a sharp drop in AF in plasma, for reasons we do not understand, and may be related to metabolic effects or technical artefacts. The increased AFs from days 244 to 272 may be a related transient effect. Similarly, the sample collected on day 300 from patient 103 presented an unexpectedly low total cfDNA level (> 10-fold different from the timepoint immediately before and after), which could potentially influence the interpretation of mutant allele fractions at that timepoint. Such variations could be contributed by effects of processing, collection or other technical reasons. We have therefore excluded that timepoint from the analysis. To normalize for these kinds of pre-analytic effects, and to explore the relative representation of the alleles, we calculated the ratio of AFs of the resistance-conferring/activating EGFR mutations (Fig 4A). The ratio was zero before the start of TKI. It reached a maximal value of 0.43 during first-line TKI (when L3 grew substantially), dropped to 0.02 after chemotherapy (corresponding to a reduction in L3), then rapidly increased upon TKI re-challenge to 0.67.



*Treatment discontinued and switched to supportive care.

Figure 4. Analysis of relative representation of activating and resistance-conferring mutations in EGFR during sequential therapy.

A (i) Sizes of three different lesions in patient 103 over time, measured from the start of first-line TKI treatment. Shading indicates duration of treatment with TKI (days 0–297), chemotherapy (days 297–679), and TKI re-challenge (days 679–783). From day 783, the patient was treated with supportive care. Dotted lines indicate the CT imaging assessment at select timepoints. (ii) Levels of activating EGFR mutations (exon 19 deletion) and resistance-conferring EGFR T790M mutations for patient 103 (Dataset EV1). (iii) Ratio of resistance-conferring/activating mutations, calculated from data shown in Appendix Table S1 (excluding the data at $T = 300$ days).

B CT imaging scans performed at the start of TKI treatment (day 0), at the change of treatment to chemotherapy (day 297), at the end of chemotherapy and start of TKI re-challenge (day 679), and after initiating TKI re-challenge (day 741). Sizes are assessed from CT imaging scans, and indicated by blue (i, lung), green (ii, left lobe liver) and orange (iii, right lobe liver) lines. Lesions identified in the lung (blue arrow), left lobe liver (green arrow) and right lobe liver (orange arrow) are indicated. PR, partial responses; PD, progressive disease.

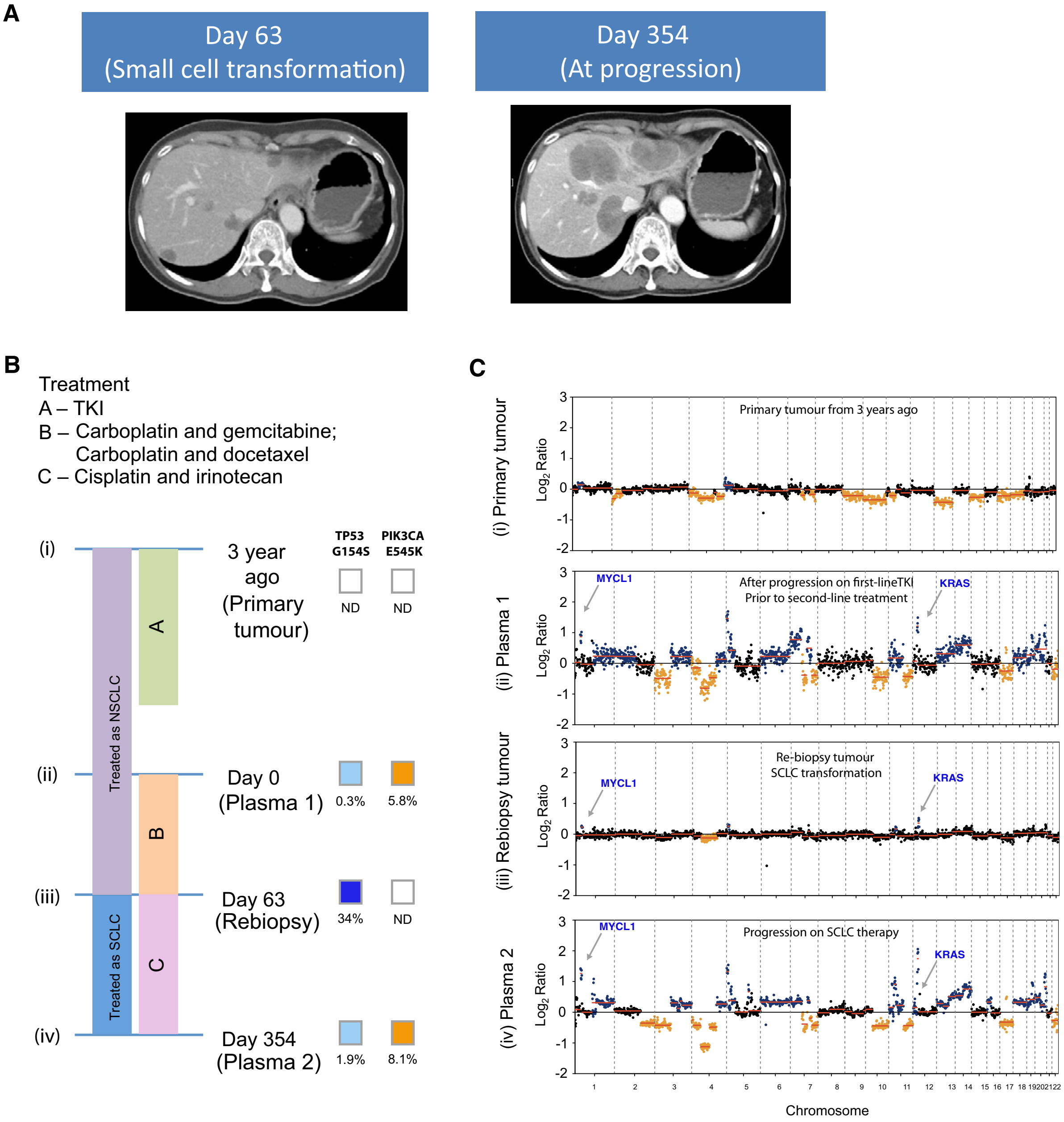


Figure 5. Plasma analysis reveals global copy number changes and ctDNA dynamics in patients who have undergone histological transformation to SCLC (patient 223).

A CT liver scans are shown: At day 63, the patient had progressed on platinum-based chemotherapy, and CT of the liver showed appearance of new liver lesions. Liver biopsy at that point confirmed small-cell lung cancer in new lesions in the liver. CT of the liver at day 354 shows marked growth in the liver lesions, after a period of transient response to cisplatin and irinotecan.

B The timeline of the patient's treatment is charted, alongside the timepoints where plasma and tumour samples are available for molecular analyses. Per diagram timeline, tumour samples were available from diagnosis (3 years prior to recruitment for study) (i) and at day 63 (iii). Plasma samples were available at day 0 (ii) and day 354 (iv). The mutations allele fractions for TP53 and PIK3CA are shown.

C Global copy number profiles in plasma samples collected prior to small-cell transformation (ii) and after SCLC transformation and progression on cisplatin and irinotecan (iv). Global copy number profiles in tumour samples collected at diagnosis of NSCLC (i) and at small-cell transformation (iii). CNA events that were significantly identified and coincide with literature-reported SCLC events are denoted in colours: blue for gain and orange for loss.

Although a liver biopsy was not performed, which may have been able to confirm the presence of T790M in L3, the clinical and radiological evidence in conjunction with the dynamics of EGFR mutations in plasma strongly suggests that the T790M was present in L3 but not in L1/L2. This suggests that the ratio of resistance-conferring/activating mutations in plasma can help identify dominant drivers of disease and progression in real time (Oxnard *et al*, 2016).

Copy number changes and mutations detected in plasma after SCLC transformation

Recent findings have shown that around 2–3% of NSCLC patients develop resistance to EGFR-targeted therapies by undergoing histological transformation to small-cell lung cancer (SCLC; Sequist *et al*, 2011). In our cohort, three patients (patients 122, 223, 218) were confirmed to present SCLC histology based on re-biopsy examination at progression. We used targeted deep sequencing and shallow whole-genome sequencing of plasma DNA to track and study the dynamics of somatic point mutations and global copy number alterations (CNAs) in samples collected before and after SCLC transformation in those patients.

T790M mutations were not detected in plasma at disease progression for any of the three SCLC-transformed patients, and two of them (patients 122 and 218) retained the EGFR-activating mutations in plasma after SCLC transformation. *TP53* mutations were detected before EGFR-TKI initiation in all three patients' baseline plasma samples, at low levels ($< 1\%$) compared to the EGFR-activating mutations, and their levels in plasma increased with disease progression (patient 122, 223) or decreased when patient demonstrated clinical response (patient 218) during small-cell lung cancer-directed chemotherapy (Figs 5 and 6, and Appendix Fig S5). Analysis of the plasma samples collected after transformation in all three patients revealed the emergence of CNAs that have been previously reported to be associated with SCLC, including *MYCL1, SOX2* and *SOX4* (George *et al*, 2015; Figs 5 and 6, and Appendix Fig S5). In each patient, we also identified gain or loss of cancer genes as part of larger chromosomal events (> 5 Mb; Figs 5 and 6, and Appendix Fig S5). These may have contributed to the biological change or may represent passenger events as a result of greater genomic instability of the TP53-mutant clones. We observed focal copy number changes at key oncogenic drivers that may play a role in driving disease progression, for example, amplification of *KRAS* in patient 223 (Fig 5) and amplification of *EGFR* in patients 122 and 218 (Appendix Fig S5 and Fig 6).

Patient 223 (Fig 5A) initially harboured an exon 19 activating EGFR mutation, had indolent disease and remained clinically and radiologically responsive to first-line EGFR-TKI for 2 years. Subsequently, there was development of PET FDG-avid, but subcentimetre, liver lesions that were not clearly appreciated on CT imaging. At this point, even though progression was only obvious in two new small spots on PET imaging (stable disease by CT and RECIST criteria), ctDNA showed multiple copy number changes (Fig 5B and C, plasma 1). There was subsequent rapid growth of the liver lesions, despite two lines of chemotherapy. A liver biopsy was performed at that point, and showed only small-cell carcinoma, with no further activating nor T790M EGFR mutations found in the small-cell cancer. Figure 5C demonstrates that a few months prior to the confirmation of small-cell transformation, plasma analysis already showed marked copy number changes, this time of focal amplification of *MYCL1* and *KRAS*, known oncogenic drivers in the KRAS pathway. These focal genomic changes were also observed in the subsequent liver biopsy, showing parallel changes in both tumour and plasma, suggesting that plasma is a good surrogate for study of genomic copy number alterations and evolution. In this particular case, the copy number changes were more markedly observed in the plasma, compared to the tumour, likely due to scarcity of tumour cells in the repeat biopsy.

Another patient with SCLC transformation (patient 218, Fig 6A) harboured an *EGFR* L858R-activating mutation at diagnosis, and again demonstrated a good clinical and radiological response, with progression-free survival of 14 months on first-line EGFR-TKI (Fig 6B). Upon clinical progression while on treatment with EGFR-TKI (day 78), the plasma analysis (Fig 6B plasma 1) showed increased levels of *EGFR* L858R as well as a *TP53* R175H mutation, one of the most frequently observed mutations in *TP53*. On treatment with platinum-based doublet chemotherapy with pemetrexed, the patient achieved stability of disease. There was however early progression on maintenance pemetrexed and despite a switch to carboplatin and docetaxel, the patient's lung mass progressed, and a biopsy confirmed SCLC, harbouring the original activating *EGFR* L858R mutation. At this point, the patient was switched to treatment for SCLC with carboplatin and etoposide, with a radiological response and corresponding reduction in *TP53* R175H mutation (Fig 6B, plasma 3). Unfortunately, the treatment response was transient, and there was development of widespread symptomatic metastases in both the brain and spine, necessitating radiotherapy to those areas. At the completion of radiotherapy, the patient had developed liver metastases with rapid progression. At this point, we observed marked copy number change in plasma 4 and 5 analyses (Fig 6C and Appendix Fig S6). This example illustrates that dynamics of genomic copy number changes in plasma reflect the mutational burden and radiologic responses.

The losses of TP53 and RB1 are common in SCLC (George *et al*, 2015), and we therefore attempted to look for RB1 somatic copy number alterations but did not find any significant signal by sWGS. In patient 218, RB1 did appear to have a reduced copy number, but this event could be part of a much larger chromosomal aberration, and thus, we cannot rule this out as being a passenger event. RB1

Figure 6. Plasma analysis reveals global copy number changes and ctDNA dynamics in patients who have undergone histological transformation to SCLC (patient 218).

- A CT images of the left lung tumour at baseline (day −730), and upon progression on EGFR-TKI (day 78). The patient was treated for non-small-cell lung cancer from days 78 to 329 with two lines of chemotherapy: initially carboplatin and pemetrexed, followed by docetaxel. The CT image corresponding to plasma 2 and re-biopsy of the tumour at day 329 was upon progression on the above two lines of chemotherapy. CT image corresponding to plasma 3 at day 379 was upon response to small-cell lung cancer chemotherapy. CT corresponding to days 500 and 524 are upon progression on cisplatin and etoposide. Marked growth of the lung and liver lesions are demonstrated, and this corresponds to marked CNA changes on plasma drawn on those respective days.
- B The timeline of the patient's treatment is shown, alongside timepoints where tumour and plasma samples were available for analyses. The bottom chart shows the respective mutations that were found, and the changes to the mutation allele fractions in a longitudinal timeline.
- C Global copy number profiles in tumour and plasma samples are shown: Tumour samples at baseline diagnosis of non-small-cell lung cancer (day −730) and at transformation to small-cell lung cancer (day 329). For plasma samples, the following were available: days 78 (upon progression on EGFR-TKI); 329 (at transformation to small-cell lung cancer); 379 (at response to small-cell lung cancer); 500 and 534 (progression on small-cell lung cancer treatment).

▶

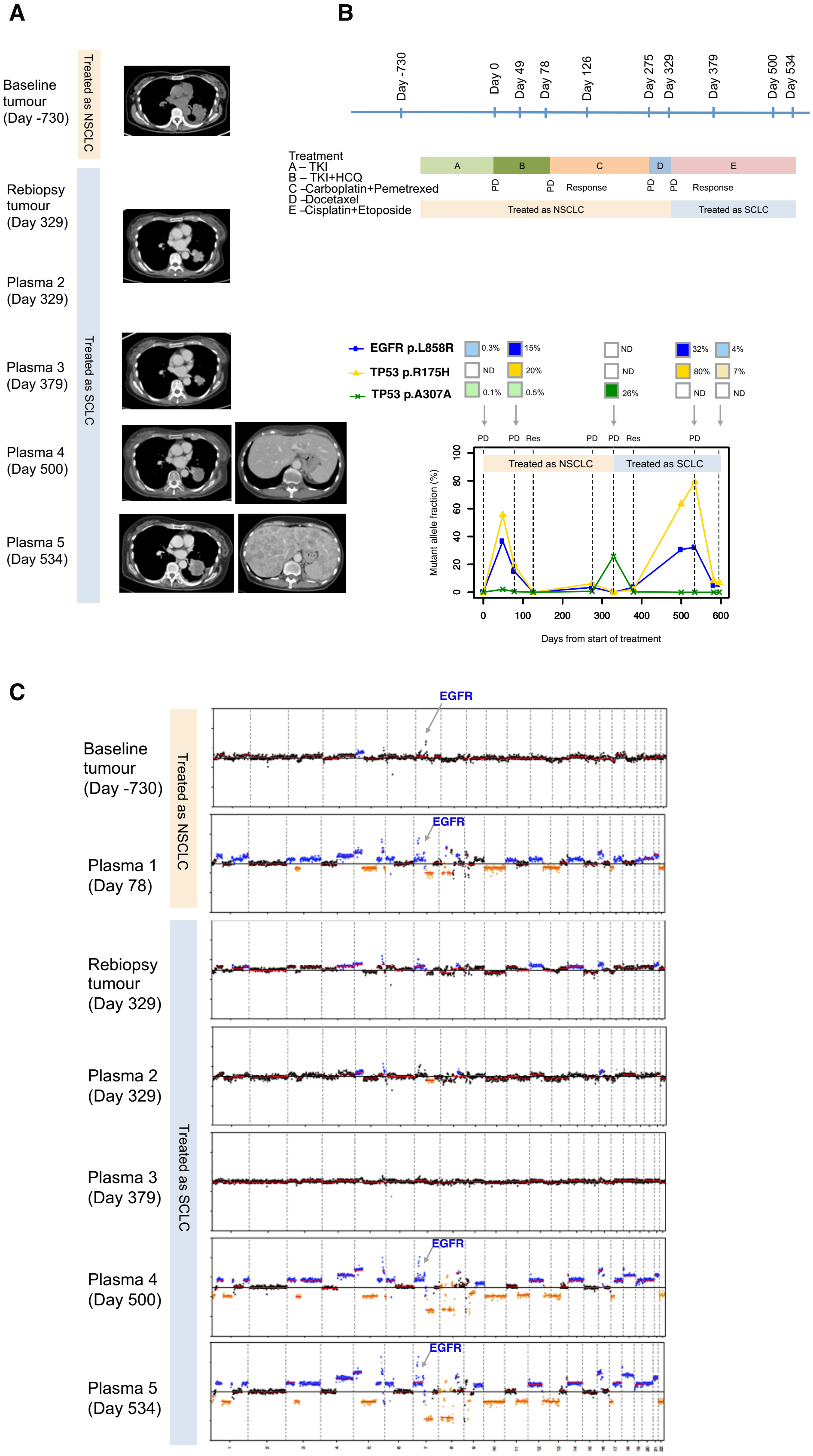

A
Baseline tumour (Day -730)
Treated as NSCLC
Rebiopsy tumour (Day 329)
Plasma 2 (Day 329)
Plasma 3 (Day 379)
Plasma 4 (Day 500)
Plasma 5 (Day 534)
Treated as SCLC
B
Day -730
Day 0
Day 49
Day 78
Day 126
Day 275
Day 329
Day 379
Day 500
Day 534
Treatment
A – TKI
B – TKI+HCQ
C –Carboplatin+Pemetrexed
D –Docetaxel
E –Cisplatin+Etoposide
A
B
C
D
E
PD
PD
Response
PD
PD
Response
Treated as NSCLC
Treated as SCLC
EGFR p.L858R
TP53 p.R175H
TP53 p.A307A
0.3%
15%
ND
32%
4%
ND
20%
ND
80%
7%
0.1%
0.5%
26%
ND
ND
PD
PD
Res
PD
PD
Res
PD
Treated as NSCLC
Treated as SCLC
Mutant allele fraction (%)
0
20
40
60
80
100
0
100
200
300
400
500
600
Days from start of treatment
C
EGFR
Baseline tumour (Day -730)
Treated as NSCLC
EGFR
Plasma 1 (Day 78)
Rebiopsy tumour (Day 329)
Plasma 2 (Day 329)
Plasma 3 (Day 379)
Treated as SCLC
EGFR
Plasma 4 (Day 500)
EGFR
Plasma 5 (Day 534)

Figure 6.

alterations are important driver in SCLC, but it is not necessarily the only driver (Karachaliou *et al*, 2016). All three SCLC-transformed patients have evidence of TP53 mutations in their SCLC re-biopsies, and pre- and post-transformation plasma suggest that TP53 is an important driver in these particular patients.

In all three patients, CNAs were more evident in plasma as compared to tumour DNA analysis, likely due to scarcity of tumour cells at repeat biopsy. The data also illustrate that the global CNA profile in plasma can act as an indicator of disease burden and can be used to track clinical progression, as previously suggested in other cancer types (Heitzer *et al*, 2013). Subtype-specific mutational and CNA signatures can be identified in plasma in association with histological transformation that warrants different treatment strategy, and their increasing levels in plasma can pre-date radiological progression (by CT imaging). These observations suggest that plasma genomic changes could be indicative of early progression, and may complement current imaging modalities in monitoring response.

Discussion

We studied the dynamics of concurrent somatic point mutations and copy number alterations in plasma DNA during treatment of NSCLC patients with EGFR inhibitor. Several observations may provide information for the design of future plasma DNA studies of patients treated with targeted therapies:

First, we found a strong concordance between *EGFR* status in tumour and plasma samples, and showed that mutations in multiple cancer-related genes can be identified directly in plasma by targeted sequencing. These results confirm findings from a previous validation study (Douillard *et al*, 2014; Weber *et al*, 2014; Huang *et al*, 2017), and lend further credibility to the application of circulating DNA in plasma for non-invasive molecular profiling and treatment stratification (Jamal-Hanjani *et al*, 2017; Remon *et al*, 2017). Second, high pre-treatment levels of ctDNA, and specifically of *EGFR* mutations in plasma prior to treatment with EGFR-TKI, correlated with increased tumour burden and were associated with poor prognosis, echoing previous findings (Mok *et al*, 2015). We also showed that early changes in levels of ctDNA (in our case, of *EGFR* mutations) may predict initial response (Parkinson *et al*, 2016). Both of these findings lend support to the analysis of baseline and subsequent plasma samples for *EGFR* mutations to track treatment responses. Third, we detected the emergence of the T790M mutation in approximately 50% of patients who progressed on TKI, at a median of 6.8 months before clinical progression. Early identification of emerging resistance highlights the potential to use ctDNA to guide clinical interventions such as therapies that target T790M-mutant cells (Janne *et al*, 2015; Sequist *et al*, 2015; Chabon *et al*, 2016; Remon *et al*, 2017).

In addition, we showed that profiling TP53 in plasma before EGFR-targeted therapy can provide prognostic value. Cancers harbouring both *TP53* and *EGFR* mutations in baseline plasma were associated with inferior overall survival in patients treated with EGFR-targeted TKI. This confirmed observations from a tumour sequencing study (Labbe *et al*, 2017), and because plasma DNA captures the mutations coming from different parts of the tumour, we found that in some patients, these TP53 mutations pre-existed at AF < 1% in plasma prior to treatment and later became the dominant mutations in plasma when patients progressed or exhibited histological transformation to SCLC. This echoes one of the recent findings that EGFR-TKI-resistant SCLCs can branch out from early events that pre-existed in NSCLC prior to transformation based on tumour biopsy analysis (Lee *et al*, 2017).

These data highlight the potential value for clinical management of analysing mutations which may not be perceived as "actionable", such as mutations in the tumour suppressor *TP53*, and suggests that the genetic context of an EGFR-mutant tumour may determine its dependence on the EGFR and/or other pathways and predict sensitivity or resistance to EGFR-directed treatment. Tracking the dynamics of multiple mutations showed that different resistance mechanisms co-existed, and are likely to be the result of tumour heterogeneity (Piotrowska *et al*, 2015). As illustrated by the above clinical cases, the response and progression of different lesions coincided with the changing levels of distinct mutations in plasma. As treatment selection pressure is dynamic, analysing the relative proportions of different oncogenic drivers in plasma at any one point may provide insight on a particular dominant oncogenic pathway dependence and guide decisions on subsequent treatment by targeting the current dominant clone.

Finally, we showed that tracking the dynamics of plasma EGFR mutations alone may not provide the most accurate estimate of tumour responses, as seen in the 14% of patients who progressed with decreasing levels of EGFR mutations in plasma. We speculate that this observation may be explained by the recent finding of subclonal EGFR driver mutations in 3 of 21 (14%) NSCLC cases (McGranahan *et al*, 2015), and suggest that monitoring EGFR-targeted therapies by ctDNA would require tracking mutations beyond EGFR. These findings will need to be confirmed by further studies in larger cohorts. Nonetheless, using a multi-gene assay, our data revealed the presence of concurrent oncogenic drivers before treatment. Parallel analysis of somatic point mutations and global CNA events (by shallow WGS) could more accurately track disease burden and detect subtype-specific events (e.g. marked copy number changes associated with histological transformation). In particular, we identified multiple genomic changes in the plasma of three patients who underwent transformation to SCLC (patients 122, 223 and 218), which correlated closely with burden of disease. All three patients presented increasing fractions of *TP53* mutations that pre-existed at < 1% levels in plasma before treatment, together with recurrent SCLC CNA events in ctDNA at the time of transformation. These changes correlated closely with burden of disease, and CNA signals in plasma reduced accordingly when computerized tomography imaging showed radiological response to chemotherapy. In these cases, it was interesting to note that all cases exhibited different drivers, *TP53* in the first case, *MYCL-1* and *KRAS* in the second, and two *TP53* mutations in the third. It may be that each patient has a unique signature of genomic copy number change that may reflect disease burden, and this can be used to determine responding or progressive disease at various timepoints, with relation to each line of treatment received. However, the complex and varied genomic landscape in these transformed small-cell lung cancers underscores the difficulty in targeting any one of these genomic signatures (even if actionable), and providing support for the use of chemotherapy, which is currently the most appropriate treatment for targeting these multiple genomic instabilities. Our

© 2018 The Authors

findings suggest that an observation of multiple genomic copy number changes in the plasma of a patient with rapid progression of disease on EGFR-TKI should prompt the need to re-biopsy to exclude the possibility of small-cell transformation. To date, EGFR T790M remains the main actionable resistance mechanisms in the context of EGFR-TKI. However, the ability to reveal TP53 mutations or other possible SCLC-associated genomic signatures in plasma would provide additional insight into possible resistance mechanisms that are particularly important in individuals that show no evidence of T790M or other known resistance mechanisms, and may justify the need for a re-biopsy to confirm the histological transformation.

Our study, spanning 392 clinical samples analysed by a combination of genomic techniques, describes multiple genomic changes in *EGFR*-mutant patients with acquired resistance to first-generation EGFR-TKIs, and may explain the heterogeneity of treatment response to EGFR-TKIs in *EGFR*-mutant patients with similar activating mutations. We studied key genomic driver events of lung cancer, and evaluated their significance in a temporal and dynamic way with respect to disease response and progression in patients with analyses of longitudinal plasma studies. As the majority of patients were treated when second- and third-generation EGFR-TKIs were not readily available, most patients with acquired *EGFR* T790M mutation were not routinely re-biopsied and were treated with chemotherapy. Analyses of changes in ctDNA in response to treatment with second- and third-generation EGFR-TKIs are not within the scope of this study. Nonetheless, some of the acquired resistance mechanisms described here may also apply to acquired resistance to a wide range of EGFR-TKIs. Importantly, our study represents the first report on ctDNA changes in EGFR-mutant cancers before and after histological transformation to SCLC and provides important insight into the management of this alterative form of resistance mechanism to EGFR-TKI.

In summary, our data show that in the NSCLC EGFR-targeted therapy setting, analysing the presence and dynamics of both actionable oncogenic drivers (such as *EGFR* mutations) and other, potentially “non-actionable” alterations (such as *TP53* mutations and global copy number changes), before and during treatment, can offer clinically relevant information to potentially guide subsequent clinical management.

Materials and Methods

Sample collection and processing

Patients with metastatic NSCLC treated by gefitinib in combination with hydroxychloroquine therapy attending the University Hospital of Singapore, Singapore, during January 2009 to May 2014 were recruited as part of the “Hydroxychloroquine and Gefitinib to Treat Lung Cancer” study (NCT00809237). This was a single-arm phase II study that recruited two groups of patients. In the first group, EGFR-TKI-naive patients who were known to have activating *EGFR* mutations were recruited to determine whether hydroxychloroquine improved the efficacy of gefitinib. The second group included patients who had previously responded to EGFR-TKIs for at least 12 weeks (per Jackman criteria for acquired resistance to EGFR-TKI) and aimed to determine whether the addition of hydroxychloroquine to gefitinib would reverse acquired resistance in these patients. Please see consort diagram (Appendix Fig S1) for further details including number of patients in each arm included for the purpose of this cDNA study (note that this represents a subset of the original clinical study—as this was subject to availability of plasma samples). Blood was collected from patients in a CPT sodium citrate tube (BD) every 4 weeks and stored at −80°C. CT imaging of relevant measurable sites of disease was performed every 8 weeks. This study was approved by the Singapore National Healthcare Group (Singapore National Healthcare Group Domain Specific Review Board NHG DSRB Reference: 2008/00196). Blood and tumor collection were also collected and approved by NHG DSRB 2014/00131. Informed consent has been obtained from all patients involved in this study. DNA was extracted from 0.8 to 2 ml of plasma using the Qiagen QIAamp Circulating Nucleic Acid Kit (Qiagen) and eluted into 50 μl buffer AVE. More details of sample processing are given in Appendix Supplementary Methods. A spike-in control, non-human DNA PCR product was added to the lysis buffer during DNA extraction to control for extraction efficiency.

Mutation identification and quantification by TAm-Seq and digital PCR

Analysis by tagged-amplicon deep sequencing (TAm-Seq) was performed using the panel described previously (Forshew *et al*, 2012), with the addition of an amplicon that covers exon 18 of *EGFR* (additional details in Appendix Supplementary Methods). All samples were analysed by at least two replicates to control for errors arising during PCR. The purified libraries were sequenced using paired-end 100 bp read length of a HiSeq 2000 or HiSeq 2500 System (Illumina, USA). Somatic mutations were identified based on filtering against the matched normal control (white blood cells) of the same patient. For quantification of known hotspot mutations in EGFR, namely exon 19 deletion, T790M and L858R, digital PCR analysis was developed and optimized: sensitivity and specificity and limit of detection was determined using samples with known mutations and samples from healthy volunteer controls (Appendix Fig S7 and Appendix Table S6). Assays were performed using the BioMark system using 12.765 Digital Arrays (Fluidigm, USA) following manufacturer’s instructions and protocol. Total DNA levels (amplifiable copies per ml) were also quantified in every plasma sample by digital PCR using a 65-bp assay targeting a region on the *RPP30* gene (Appendix Table S7), a region in the genome that is not amplified in lung cancer (Wang *et al*, 2010). Two samples were excluded from analysis due to an unexplained sharp drop in total circulating DNA levels extracted from plasma, with >10-fold drop in those levels compared to samples collected few weeks prior and no evident association with any clinical or treatment parameters. This is suspected to be related to technical fault either at collection or processing of samples. Evaluation of the specificity and sensitivity of the assays was described in Appendix Supplementary Methods. The primer and probe sequences of all the digital PCR assays are summarized in Appendix Table S7. The mutant allele fractions measured by TAM-Seq and ddPCR strongly correlate with each other (Appendix Fig S8).

Shallow whole-genome sequencing (sWGS)

Libraries were prepared from either plasma DNA (5–10 ng), sheared tumour DNA, or sheared buffy coat DNA using the Plasma-Seq

© 2018 The Authors

protocol (Rubicon, USA). Briefly, end repair and "A-tailing" of fragment ends preceded the ligation of truncated Illumina sequencer-compatible adapters to fragment ends. Thermocycling of libraries completed the adapters through the addition of sample-specific index sequences, and was performed as described in the Plasma-Seq protocol (Heitzer *et al*, 2013), using 10 (plasma) or 8 (tumour and buffy coat) amplification cycles. Upon amplification libraries were cleaned using Agencourt AMPure XP beads (Beckman Coulter, USA) at a 1:1 (v/v) ratio and eluted in 30 μl nuclease-free water. Successful library preparation was confirmed by running 1 μl of library on a High-Sensitivity Bioanalyser gel, and libraries were quantified using SYBR-green-based qPCR (Kapa Biosystems, USA). Libraries were pooled in an equimolar fashion, and 125-bp paired-end sequencing was performed on Illumina sequencers (Illumina, USA).

Paired-end sequence reads were aligned to the human reference genome (GRCh37) using BWA, SAMtools was used to convert files to BAM format, to which mate pair information was added. PCR duplicates were marked using Picard-Tools' "MarkDuplicates" feature and were excluded from downstream analysis. Fragment lengths were analysed using Picard-Tools' "CollectInsertSizeMetrics". CNA calling was performed in R using the program QDNAseq (Scheinin *et al*, 2014). Briefly, sequence reads were allocated into equally sized (here 1 Mb and 50-kb bins) non-overlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and regions corresponding to previously "blacklisted" regions (ENCODE) were excluded from downstream analysis. Within the QDNAseq package, bins were segmented using the "Circular Binary Segmentation" algorithm (Venkatraman & Olshen, 2007) and significantly "amplified" or "lost" regions were called using CGHcall (van de Wiel *et al*, 2007).

Data deposition

Sequence data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under Accession no. EGAS00001002908.

Survival analysis

Kaplan–Meier curves were computed for prognostic groups defined by their mutation fractions, and log-rank tests were computed for testing differences in survival. We measured pre-treatment ctDNA levels using allele fractions of the *EGFR*-activating mutations and computed Kaplan–Meier survival curves to evaluate the effects of different levels of pre-treatment ctDNA: We divided patients into three groups: low pre-treatment ctDNA levels (less than the lower quartile), intermediate (second and third quartiles) or high (upper quartile). All survival analyses were performed using the R package survival (Therneau, 2012). It should be noted that only the EGFR-TKI-naïve group of patients with available pre-treatment plasma samples ($n = 19$) were used for progression-free and overall survival analyses, and correlative prognostic study. This is to ensure analyses of a homogeneous population. The survival analyses of both groups of patients will be reported in a separate clinical paper that would include response rates and other clinical parameters.

The experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.

The paper explained

Problem

The cancer genome evolves under the selective pressure of targeted therapies. One of the key challenges is to identify resistance mechanisms and the most dominant drivers as early as possible. Analysis of plasma cell-free DNA allows one to track molecular dynamics non-invasively.

Results

This study found that cell-free DNA analysis reveals clinically important information during EGFR-targeted therapy in non-small-cell lung cancer (NSCLC): At baseline, quantitative tumour-derived cell-free DNA levels in plasma provided prognostic information and correlated with tumour burden. During treatment, multiple potential indications of resistance such as *EGFR* T790M or *TP53* can be detected in plasma months before disease progression became clinically evident. Longitudinal analysis of tumour-derived cell-free DNA levels tracks tumour responses and reveals heterogeneous resistance mechanisms: The majority depend on the EGFR pathway while a small subset developed alterative drivers that could be identified by tracking multiple mutations in plasma DNA. In patients who developed resistance by transforming to small-cell lung cancer (SCLC), we identified *TP53* mutations, one of the key drivers of SCLC, and SCLC-specific copy number events in plasma before the transformation.

Impact

Parallel analysis of multiple mutations and copy number alterations in plasma allows identification of dominant drivers at any given time during treatment. Tracking EGFR mutations alone in plasma during EGFR-targeted therapy may not accurately reflect tumour burden due to underlying tumour heterogeneity. The results of this study provide important insight about the implication of cell-free DNA analysis for management of targeted therapies.

Expanded View for this article is available online.

Acknowledgements

We are grateful to the following physicians who recruited patients on the study—Dr. Pei Jye VOON, Dr. Winnie LING, Dr. Thomas SOH, Dr. Chee Seng TAN, Dr. Angela PANG and Dr. Yew Oo TAN. We also thank the following for technical support—Maricel Tiemsin CODERO, Hui Hui SHEE, Shiau Hui Diong, and Dr. Azhar ALI from The Centre for Translational Research and Diagnostics, National University of Singapore and Cancer Science Institute, Singapore. We thank the sequencing support by the Genomics Core of the Cancer Research UK Cambridge Institute, and by David Bentley and Sean Humphray from Illumina. We would like to acknowledge the support of The University of Cambridge, Cancer Research UK (grant numbers A11906, A20240; to N.R.), the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement no. 337905 (to N.R.), and Hutchison Whampoa Limited (to N.R.), and National Medical Research Council, Singapore (to T.M.C.).

Author contributions

DWYT, TE, TMC and NR initiated and designed the study. DWYT, CGS, DC, TF, MM, FM, DG and NR developed methods. DWYT, MM, TMC and NR analysed the data, with assistance from OMR, CC, CGS, DC, FM, JM and WL. ASCW, RAS, HLL, BCG and TMC are the treating physicians of the patients included in this study, and collected samples and clinical data. DWYT, MM, TMC and NR interpreted the data and wrote the manuscript with assistance from other authors. All authors approved the final manuscript.

© 2018 The Authors

Conflict of interest

T.F., D.G. and N.R. are co-founders, shareholders and employees/consultants of Inivata Ltd., a company that seeks to commercialize ctDNA technologies and has licensed patents and technologies from Cancer Research Technology and the University of Cambridge. D.T. and F.M are former consultants of Inivata Ltd. D.T., M.M., T.F., F.M., J.M., D.G. and N.R. may receive royalties related to the licences of IP to Inivata Ltd, and the terms of these royalties are managed by Cancer Research Technology and Cambridge Enterprise. T.E. and N.R. have received research support from AstraZeneca. T.E. is an employee of AstraZeneca on leave of absence from the University of Cambridge.

References

Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, Le Quesne J, Moore DA, Veeriah S, Rosenthal R *et al* (2017) Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* 545: 446 – 451

Camidge DR, Pao W, Sequist LV (2014) Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat Rev Clin Oncol* 11: 473 – 481

Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, Kurtz DM, Stehr H, Scherer F, Karlovich CA *et al* (2016) Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. *Nat Commun* 7: 11815

Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH *et al* (2017) OncoKB: A precision oncology knowledge base. *JCO Precis Oncol* doi: 10.1200/PO.17.00011

Couraud S, Vaca-Paniagua F, Villar S, Oliver J, Schuster T, Blanche H, Girard N, Tredaniel J, Guilleminault L, Gervais R *et al* (2014) Noninvasive diagnosis of actionable mutations by deep sequencing of circulating free DNA in lung cancer from never-smokers: a proof-of-concept study from BioCAST/IFCT-1002. *Clin Cancer Res* 20: 4613 – 4624

Douillard JY, Ostoros G, Cobo M, Ciuleanu T, Cole R, McWalter G, Walker J, Dearden S, Webster A, Milenkova T *et al* (2014) Gefitinib treatment in EGFR mutated caucasian NSCLC: circulating-free tumor DNA as a surrogate for determination of EGFR status. *J Thorac Oncol* 9: 1345 – 1353

Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D *et al* (2012) Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 4: 136ra168

George J, Lim JS, Jang SJ, Cun Y, Ozretic L, Kong G, Leenders F, Lu X, Fernandez-Cuesta L, Bosco G *et al* (2015) Comprehensive genomic profiles of small cell lung cancer. *Nature* 524: 47 – 53

Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, Benezeder T, Auer M, Pischler C, Mannweiler S *et al* (2013) Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med* 5: 30

Huang WL, Chen YL, Yang SC, Ho CL, Wei F, Wong DT, Su WC, Lin CC (2017) Liquid biopsy genotyping in lung cancer: ready for clinical utility? *Oncotarget* 8: 18590 – 18608

Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, Shafi S, Johnson DH, Mitter R, Rosenthal R *et al* (2017) Tracking the evolution of non-small-cell lung cancer. *N Engl J Med* 376: 2109 – 2121

Janne PA, Yang JC, Kim DW, Planchard D, Ohe Y, Ramalingam SS, Ahn MJ, Kim SW, Su WC, Horn L *et al* (2015) AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* 372: 1689 – 1699

Karachaliou N, Pilotto S, Lazzari C, Bria E, de Marinis F, Rosell R (2016) Cellular and molecular biology of small cell lung cancer: an overview. *Transl Lung Cancer Res* 5: 2 – 15

Labbe C, Cabanero M, Korpanty GJ, Tomasini P, Doherty M, Mascaux C, Jao K, Pitcher B, Wang R, Pintilie M *et al* (2017) Prognostic and predictive effects of TP53 mutation in patients with EGFR-mutated non-small cell lung cancer. *Lung Cancer* 111: 23 – 29

Lee JK, Lee J, Kim S, Kim S, Youk J, Park S, An Y, Keam B, Kim DW, Heo DS *et al* (2017) Clonal history and genetic predictors of transformation into small-cell carcinomas from lung adenocarcinomas. *J Clin Oncol* 35: 3065 – 3074

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG *et al* (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350: 2129 – 2139

McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C (2015) Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. *Sci Transl Med* 7: 283ra254

Mok T, Wu YL, Lee JS, Yu CJ, Sriuranpong V, Sandoval-Tan J, Ladrera G, Thongprasert S, Srimuninnimit V, Liao M *et al* (2015) Detection and dynamic changes of EGFR mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC patients treated with first-line intercalated erlotinib and chemotherapy. *Clin Cancer Res* 21: 3196 – 3203

Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE *et al* (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* 20: 548 – 554

Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, Stehr H, Liu CL, Bratman SV, Say C *et al* (2016) Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* 34: 547 – 555

Niederst MJ, Sequist LV, Poirier JT, Mermel CH, Lockerman EL, Garcia AR, Katayama R, Costa C, Ross KN, Moran T *et al* (2015) RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer. *Nat Commun* 6: 6377

Oxnard GR, Arcila ME, Sima CS, Riely GJ, Chmielecki J, Kris MG, Pao W, Ladanyi M, Miller VA (2011) Acquired resistance to EGFR tyrosine kinase inhibitors in EGFR-mutant lung cancer: distinct natural history of patients with tumors harboring the T790M mutation. *Clin Cancer Res* 17: 1616 – 1622

Oxnard GR, Paweletz CP, Kuang Y, Mach SL, O'Connell A, Messineo MM, Luke JJ, Butaney M, Kirschmeier P, Jackman DM *et al* (2014) Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin Cancer Res* 20: 1698 – 1705

Oxnard GR, Thress KS, Alden RS, Lawrance R, Paweletz CP, Cantarini M, Yang JC, Barrett JC, Janne PA (2016) Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J Clin Oncol* 34: 3375 – 3382

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ *et al* (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304: 1497 – 1500

Parkinson CA, Gale D, Piskorz AM, Biggs H, Hodgkin C, Addley H, Freeman S, Moyle P, Sala E, Sayal K *et al* (2016) Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: a retrospective study. *PLoS Med* 13: e1002198

Paweletz CP, Sacher A, Raymond CK, Alden RS, O'Connell A, Mach SL, Kuang Y, Gandhi L, Kirschmeier P, English JM *et al* (2015) Bias-corrected targeted next-generation sequencing for rapid, multiplexed detection of actionable alterations in cell-free DNA from advanced lung cancer patients. *Clin Cancer Res* 22: 915 – 922

Pecuchet N, Zonta E, Didelot A, Combe P, Thibault C, Gibault L, Lours C, Rozenholc Y, Taly V, Laurent-Puig P *et al* (2016) Base-position error rate analysis of next-generation sequencing applied to circulating tumor DNA in non-small cell lung cancer: a prospective study. *PLoS Med* 13: e1002199

Piotrowska Z, Niederst MJ, Karlovich CA, Wakelee HA, Neal JW, Mino-Kenudson M, Fulton L, Hata AN, Lockerman EL, Kalsy A *et al* (2015) Heterogeneity underlies the emergence of EGFRT790 wild-type clones following treatment of T790M-positive cancers with a third-generation EGFR inhibitor. *Cancer Discov* 5: 713 – 722

Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, Howarth K, Gale D, Green E, Plagnol V *et al* (2017) Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. *Ann Oncol* 28: 784 – 790

Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, van Essen HF, Eijk PP, Rustenburg F, Meijer GA *et al* (2014) DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res* 24: 2022 – 2032

Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, Bergethon K, Shaw AT, Gettinger S, Cosper AK *et al* (2011) Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med* 3: 75ra26

Sequist LV, Soria JC, Goldman JW, Wakelee HA, Gadgeel SM, Varga A, Papadimitrakopoulou V, Solomon BJ, Oxnard GR, Dziadziuszko R *et al* (2015) Rociletinib in EGFR-mutated non-small-cell lung cancer. *N Engl J Med* 372: 1700 – 1709

Sierra JR, Tsao MS (2011) c-MET as a potential therapeutic target and biomarker in cancer. *Ther Adv Med Oncol* 3: S21 – S35

Therneau T (2012) A package for survival analysis in S. In *R package version 2.36-14*.

Venkatraman ES, Olshen AB (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23: 657 – 663

Wan JC, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R, Rosenfeld N (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 17: 223 – 238

Wang J, Ramakrishnan R, Tang Z, Fan W, Kluge A, Dowlati A, Jones RC, Ma PC (2010) Quantifying EGFR alterations in the lung cancer genome with nanofluidic digital PCR arrays. *Clin Chem* 56: 623 – 632

Weber B, Meldgaard P, Hager H, Wu L, Wei W, Tsai J, Khalil A, Nexo E, Sorensen BS (2014) Detection of EGFR mutations in plasma and biopsies from non-small cell lung cancer patients by allele-specific PCR assays. *BMC Cancer* 14: 294

van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B (2007) CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics* 23: 892 – 894

Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, Kris MG, Miller VA, Ladanyi M, Riely GJ (2013) Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin Cancer Res* 19: 2240 – 2247

Yung TK, Chan KC, Mok TS, Tong J, To KF, Lo YM (2009) Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res* 15: 2076 – 2084

License: This is an open access article under the terms of the Creative Commons Attribution 4.0 License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

Translational Science

Cancer Research

Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models

Richard Mair[1,2,3], Florent Mouliere[1,3,4], Christopher G. Smith[1,3], Dineika Chandrananda[1,3], Davina Gale[1,3], Francesco Marass[1,3], Dana W.Y. Tsui[1,5], Charles E. Massie[1,3], Alan J. Wright[1,3], Colin Watts[6], Nitzan Rosenfeld[1,3], and Kevin M. Brindle[1,3,7]

Abstract

The factors responsible for the low detection rate of cell-free tumor DNA (ctDNA) in the plasma of patients with glioblastoma (GBM) are currently unknown. In this study, we measured circulating nucleic acids in patient-derived orthotopically implanted xenograft (PDOX) models of GBM ($n = 64$) and show that tumor size and cell proliferation, but not the integrity of the blood–brain barrier or cell death, affect the release of ctDNA in treatment-naïve GBM PDOX. Analysis of fragment length profiles by shallow genome-wide sequencing (<0.2× coverage) of host (rat) and tumor (human) circulating DNA identified a peak at 145 bp in the human DNA fragments, indicating a difference in the origin or processing of the ctDNA. The concentration of ctDNA correlated with cell death only after treatment with temozolomide and radiotherapy. Digital PCR detection of plasma tumor mitochondrial DNA (tmtDNA), an alternative to detection of nuclear ctDNA, improved plasma DNA detection rate (82% vs. 24%) and allowed detection in cerebrospinal fluid and urine. Mitochondrial mutations are prevalent across all cancers and can be detected with high sensitivity, at low cost, and without prior knowledge of tumor mutations via capture-panel sequencing. Coupled with the observation that mitochondrial copy number increases in glioma, these data suggest analyzing tmtDNA as a more sensitive method to detect and monitor tumor burden in cancer, specifically in GBM, where current methods have largely failed.

Significance: These findings show that detection of tumor mitochondrial DNA is more sensitive than circulating tumor DNA analysis to detect and monitor tumor burden in patient-derived orthotopic xenografts of glioblastoma.

Introduction

Release of DNA fragments from solid tumors, which can be collected in body fluids and used to identify and quantify tumor mutations, has created new possibilities for minimally invasive diagnosis and therapy monitoring (1, 2). The concentration of cell-free tumor DNA (ctDNA) varies with cancer type, with some, such as glioblastoma (GBM), showing extremely low plasma concentrations (3), which has hindered clinical translation.

Although ctDNA levels have been correlated with tumor burden (2, 4), an understanding of the relationship between tumor biology and the release of ctDNA into the circulation is lacking, most notably for GBM. Detection and measurement of ctDNA may be affected by both technical and biological factors (1, 5). Recent work has related necrosis, tumor volume, and proliferation to detection of ctDNA in patients with non–small cell lung cancer (2). However, no investigation of the effect of tumor biology on ctDNA release in GBM has been performed.

Using a large cohort of patient-derived orthotopically implanted xenografts (PDOX; $n = 64$), we investigated combined detection of circulating tumor mitochondrial DNA (tmtDNA) and ctDNA. Custom digital PCR (dPCR) was used to differentiate human mitochondrial DNA, originating from grafted tumor cells, from the host rat mitochondrial DNA. We demonstrated a higher frequency of detection and higher copy number for tmtDNA when compared with ctDNA in the plasma, cerebrospinal fluid (CSF), and urine of the xenografted rats. We used this improved yield to analyze the factors affecting tumor DNA release.

Release of ctDNA and tmtDNA in treatment-naïve GBM was associated with tumor volume and cell proliferation but not cell death. However, following treatment with temozolomide and radiotherapy (6), plasma tmtDNA was correlated with the levels of tumor cell death. Finally, bypassing blood–brain

---

[1]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom. [2]Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom. [3]Cancer Research UK Major Centre - Cambridge, Cancer Research UK Cambridge Institute, Cambridge, United Kingdom. [4]Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Pathology, Cancer Center Amsterdam, Amsterdam, the Netherlands. [5]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York. [6]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom. [7]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

R. Mair and F. Mouliere are the co-first authors of this article.

N. Rosenfeld and K.M. Brindle are the co-senior authors of this article.

Corresponding Authors: Kevin M. Brindle, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, United Kingdom. Phone: 44-1223-769647; Fax: 44-1223-769510; E-mail: kmb1001@cam.ac.uk; and Nitzan Rosenfeld, nitzan.rosenfeld@cruk.cam.ac.uk

doi: 10.1158/0008-5472.CAN-18-0074

©2018 American Association for Cancer Research.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

barrier (BBB) integrity did not significantly affect the yield of ctDNA or tmtDNA.

Materials and Methods

Cell culture

Cells were obtained either locally or from the American Type Culture Collection (ATCC) and mycoplasma tested using RNA-capture ELISA. Cell line authentication was performed using short tandem repeat genotyping contemporaneously with the experiments. U87 cells (ATCC) were cultured in DMEM, 2 mmol/L L-glutamine (Gibco), and 10% FBS (Gibco). Patient-derived cell lines were derived using protocols compliant with the UK Human Tissue Act 2004 (HTA licence ref. 12315), approved by the Local Regional Ethics Committee (LREC ref. 04/Q0108/60), and in accordance with the Declaration of Helsinki. GBM tissue was minced, and cells were filtered (40 μm; Falcon) and washed with red blood cell lysis buffer. Live cells were seeded at 1.5×10^4 cm^2 and grown as monolayer cultures on extracellular matrix–coated flasks (Engelbreth-Holm-Swarm murine sarcoma—1:10 dilution, Sigma) in Neurobasal A (Gibco), 2 mmol/L L-glutamine (Sigma), 1% streptomycin/penicillin/amphotericin B (Invitrogen), 20 ng/mL hEGF (Sigma), 20 ng/mL hFGF (R&D Systems), 2% B27 (Invitrogen), and 1% N2 (Invitrogen) at 37.5°C in 5% CO_2.

Orthotopic tumor model

Procedures were performed in compliance with project and personal licenses issued under the United Kingdom Animals (Scientific Procedures) Act, 1986, and approved by the local Animal Welfare and Ethical Review body. Patient-derived cells, below passage 20, were resuspended at 2×10^5 cells μL^{-1}, and 5 μL were implanted stereotactically (2 mm anterior and 3 mm lateral to the bregma, right side) in 6 week-old female rnu/rnu athymic nude rats (Charles River; Harlan; $n = 64$).

Subcutaneous tumor model

Patient-derived cells (GBM4) were resuspended at 2.5×10^4 cells μL^{-1}, and 200 μL were injected subcutaneously into the right flank of 6 athymic nude rats.

Sample collection

Whole blood was taken via tail vein cannulation or peri-mortem via cardiac puncture and exsanguination. Coagulation was inhibited by adding 4.5 mmol/L EDTA to a maximum of 6 mL of blood. CSF was collected peri-mortem via cisterna magna puncture (7) and urine by direct bladder cannulation. Samples were centrifuged (4°C, 1,500 × *g* for 10 minutes then 20,000 × *g* for 10 minutes) before freezing (–80°C).

DNA extraction

DNA from plasma (~1 mL), CSF (~100 μL), and urine (~100 μL) was extracted with the QIAamp Circulating Nucleic Acids Kit (QIAGEN) and elution volume of 50 μL. Fragments of the *Xenopus Tropicalis* genome were spiked into the samples to estimate DNA extraction efficiency (Forward PCR primer: 5′-GTGATCATGGGATTTGTAGCTGTT 3′; Reverse PCR primer: 5′ AAACCAACCTGAAAACCATGGA-3′).

Western blot

Cell or tissue samples were lysed in RIPA buffer with 1% protease inhibitor (Thermo Fisher), run on BIS-TRIS gels (Thermo Fisher) transferred onto nitrocellulose membranes, and incubated with nestin (Atlas, 1:100) and β-actin (Abcam; 1:5,000) antibodies in LI-COR-Odyssey blocking buffer (LI-COR Biotechnology) overnight at 4°C. Primary antibodies were visualized using fluorescently-labeled anti-mouse or anti-rabbit LI-COR secondary antibodies and a LI-COR Odyssey CLx imaging system (LI-COR biotechnology).

Chemoradiation

Rats were anesthetized with 1% to 2% isoflurane (Isoflo, Abbotts Laboratories Ltd.), and tumors were irradiated via a lead collimator [15 Gy; Cs-137 irradiator (IBL 637; CIS Bio International)]. Temozolomide (100 mg kg^{-1}) was given by oral gavage 1 hour prior to radiotherapy.

Histopathology and immunohistochemistry

Brains were placed in 10% formalin (Sigma-Aldrich) for 24 hours, and then sectioned. Hematoxylin and eosin staining (ST020 Multistainer; Leica Microsystems) was performed on 5 μm sections. TUNEL staining and IHC were performed on 10 μm sections. TUNEL staining used Leica's Polymer Kit (Leica Microsystems) and Promega's DeadEnd Colorimetric TUNEL System (Promega). IHC was performed using Leica's Polymer Refine Kit and human-specific antibodies: Ki67 (1:200 dilution; M7240; Dako), cleaved caspase 3 (CC3; 1:200 dilution; 9664; Cell Signaling Technology), Glial Fibrillary Acid Protein (GFAP; 1:10,000 dilution; Z0334; Dako), and Carbonic Anhydrase 9 (CAIX; 1:1,000 dilution; AB1001; BioScience, Slovakia).

In situ hybridization

Pecam1 (CD31) mRNA was detected on 5 μm formalin-fixed paraffin-embedded tissue sections with a probe for rat *Pecam1* (NM_031591.1, region 861–1766; RNAscope 2.5 LS red detection kit, 322150, Advanced Cell Diagnostics) on a Leica Bond Rx (Leica Biosystems). Hybridization was detected using the Bond Polymer Refine Red detection Kit (Leica Biosystems, DS9390) followed by counterstaining with hematoxylin. Probes targeting *peptidylprolyl isomerase B (PPIB)* (NM_022536.2, region 95–830) and *Dabp* (EF191515, region 414–86) were used as positive and negative controls, respectively.

Image analysis

Images were annotated manually and analyzed using in-house algorithms (Aperio, Leica)

Digital PCR

dPCR was performed using Fluidigm 12.765 and 37k dPCR chips (Fluidigm). For targeting human nuclear DNA, 5 μL of TaqMan Gene Expression Master Mix, 0.5 μL of buffer, 0.5 μL of EVAGREEN (Biotium), and 1 μL of 10 μmol/L forward primer (5′-TCACTCAAAGCCGCTCAACTAC-3′; Invitrogen) and 10 μmol/L reverse primer (5′-TCTGCCTTCATTTCGTTATGTACC-3′; Invitrogen) were mixed with 3.5 μL of DNA. Primers for identifying human mitochondrial DNA were forward 5′-ATACCCATGGCCAACCTCCT-3′, reverse 5′-GGGCCTTTGCGTAGTTGTAT-3′. Primers for identifying rat DNA were forward 5′-CCACCCCCTGGGCTCTGTT-3′, reverse 5′-CCCGGATCCCCTGCGTGAGA-3′. Assays for human DNA (ctDNA) and rat DNA [nontumor cell-free DNA (nt cfDNA)] targeted the human (*RPP30* gene) and rat (*RPP30* gene) sequences, respectively, in copy-

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

number–neutral regions where there was no homology with the reciprocal rat and human genomes.

Shallow whole-genome sequencing

Libraries were prepared using an NEB ultra v2 kit (New England Biolabs). Ten ng of tumor issue DNA was sheared to 150 to 200 bp with an ultra-sonicator (Covaris). For plasma and CSF samples, we selected rats with concentrations of ctDNA greater than 1,000 copies/mL, as determined by dPCR. Libraries were pooled in equimolar amounts and sequenced on a HiSeq 2500 (Illumina) generating 125 bp paired-end reads. Reads were aligned, and localization of somatic copy-number aberrations was estimated by QDNAseq (8).

Magnetic resonance imaging

We used a 7T spectrometer (Agilent) and a 72 mm inner-diameter ^{1}H quadrature birdcage coil (Rapid Biomedical GMBH). Animals were anesthetized with 1% to 2% isoflurane in O_2. Axial T_2-weighted images were acquired using a fast spin-echo sequence [TR, 1.5 seconds; TE, 40 ms; 256 × 256 data points over a 40 × 40 mm field-of-view (FOV), 4–8 averages] from 15 2-mm-thick slices. A T_1-weighted spoiled gradient echo sequence (27° flip angle, TR 43 ms, TE 4.6 ms, FOV 40 mm x 40 mm, 256 × 128 data points) was used to acquire images before and 30, 60, and 90 seconds after injection of contrast agent (100 μmol/kg Dotarem; Guebert). Five axial slices, 1.5 mm thick and with a 0.3 mm gap between them, were acquired. Images were transferred to MATLAB (Mathworks), and difference maps calculated, on a voxel-by-voxel basis, as the postcontrast image minus the precontrast image divided by the precontrast image.

Disruption of the BBB

Mannitol (2.5 mL of a 25% solution in 0.9% saline) was administered via a tail vein cannula. Rats immediately underwent diuresis, evident from urinary incontinence under anesthesia.

Demonstration of BBB opening using dynamic contrast–enhanced magnetic resonance imaging

Images were acquired using the 72 mm diameter ^{1}H transmit coil and a 2-channel rat-head ^{1}H receiver coil placed over the brain. A fast spin-echo sequence (TR 2 seconds, TE 48 ms, FOV 4 cm × 4 cm, 2 mm thick slice, 256 × 256 data points) was used to acquire 4 axial brain slices from the same region where tumors were implanted in the other animals. Baseline T_1 measurements used an inversion recovery-spoiled gradient echo sequence (adiabatic inversion pulse, 8 inversion times between 0.05 and 10 seconds, scan repeat time 12 seconds, TR 2.08 ms, TE 0.92 ms, flip angle 10°, 4 × 1.8 mm thick slices with a 0.2 mm gap between slices). Dynamic contrast–enhanced (DCE) images were acquired using a gradient echo sequence (TR 25 ms, TE 2.85 ms, flip angle 30°). A series of 100 images (2 averages, 6.4 seconds per set of 4 images) were acquired. Dotarem (0.2 mmol/kg; Gadoteric acid, Guerbet) was injected via a tail vein after the 10th image. Mannitol was administered immediately prior to the start of DCE image acquisition. Signals from the DCE time course were converted, on a pixel-by-pixel basis, to a contrast-agent concentration by assuming an R_1 relaxivity for Dotarem of 3.1 s^{-1} mmol/L^{-1} (9). An elliptical region of interest was drawn in each of the four slices, covering the thalamus to the prefrontal cortex, and an average DCE profile was calculated (10) using the same population-derived double-exponential arterial-input-function for each dataset (11). The calculated extravascular and extracellular spaces per unit volume of tissue (Ve) accessible to the contrast agent were used as an indicator of BBB permeability.

Statistical analysis

Statistics were performed using GraphPad Prism (GraphPad Software Inc.) and R (www.r-project.org). Principal Component Analysis (PCA) was performed with R using the *factoextra* package.

Results

tmtDNA is a more sensitive marker of systemic tumor nucleic acids than ctDNA and is detected in multiple body fluids

There are 10^2 to 10^5 copies of the 16.5 kb mitochondrial genome per human cancer cell (12), and therefore, tmtDNA released into the circulation may be a more sensitive marker of tumor burden than ctDNA (13). We used dPCR to investigate the levels of tmtDNA and ctDNA in different rat PDOX models of GBM, which were derived from tumor material taken from different patients with GBM. The selected dPCR assays were chosen from among 9 dPCR assays. Specificity for human (in the PDOX models, this represents tumor DNA) and rat (host) DNA was determined using plasma DNA from 4 healthy human individuals and 4 nongrafted rat controls (Fig. 1A). Human nuclear DNA levels averaged 7,469 copies/mL, and human mitochondrial DNA averaged 38,091 copies/mL in the human plasma samples, where copies/mL represents the number of amplifiable copies in the dPCR reaction. Rat nuclear DNA was not detected in the human plasma, and human nuclear DNA was not detected in rat plasma, despite a high concentration of rat nuclear DNA (15,610 copies/mL). Only very low amounts of human mitochondrial DNA (mean 3 copies/mL, <0.02%) were detected in rat plasma.

The sensitivity of our selected ctDNA and tmtDNA assays was determined with a duplicate dilution series of human DNA in rat plasma DNA. The tmtDNA assay could detect the presence of human DNA at dilution levels 100x greater than the ctDNA assay, and could detect the presence of human mitochondrial DNA even when human nuclear DNA could no longer be detected (Fig. 1B).

Six representative PDOX models of GBM (GBM1, 8 rats; GBM2, 8 rats; GBM3, 3 rats; GBM4, 36 rats; GBM5, 6 rats; and GBM6, 3 rats) were studied. In total, 64 animals were analyzed using the dPCR assay. As shown previously (14), these models showed much slower growth rates than tumors arising from implantation of a GBM cell line (U87; Supplementary Fig. S1A and Supplementary Table S1) and much higher levels of expression of glial fibrillary acidic protein *in vivo* (Supplementary Fig. S1B; ref. 15) and nestin, a neural stem cell marker, *in vitro* (16), which were largely absent from U87 tumors and cells, respectively (Supplementary Fig. S1C). All showed histologic features of GBM (Supplementary Fig. S1D).

Plasma ctDNA was detected in all but one cohort (GBM1), with a detection rate of 24% across all animals (15/64) and at an average concentration of 27 tumor haploid genome equivalents per mL (copies/mL of the targeted human sequence; Fig. 2A and B). Plasma tmtDNA was identified in all the PDOX cohorts with a detection rate of 82% (52/64) and an average concentration of 5,081 copies/mL (~190-fold higher than the mean value for ctDNA; Fig. 2A and B). Nontumor (rat host) cell-free nuclear DNA was detected in all the animals at considerably higher

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

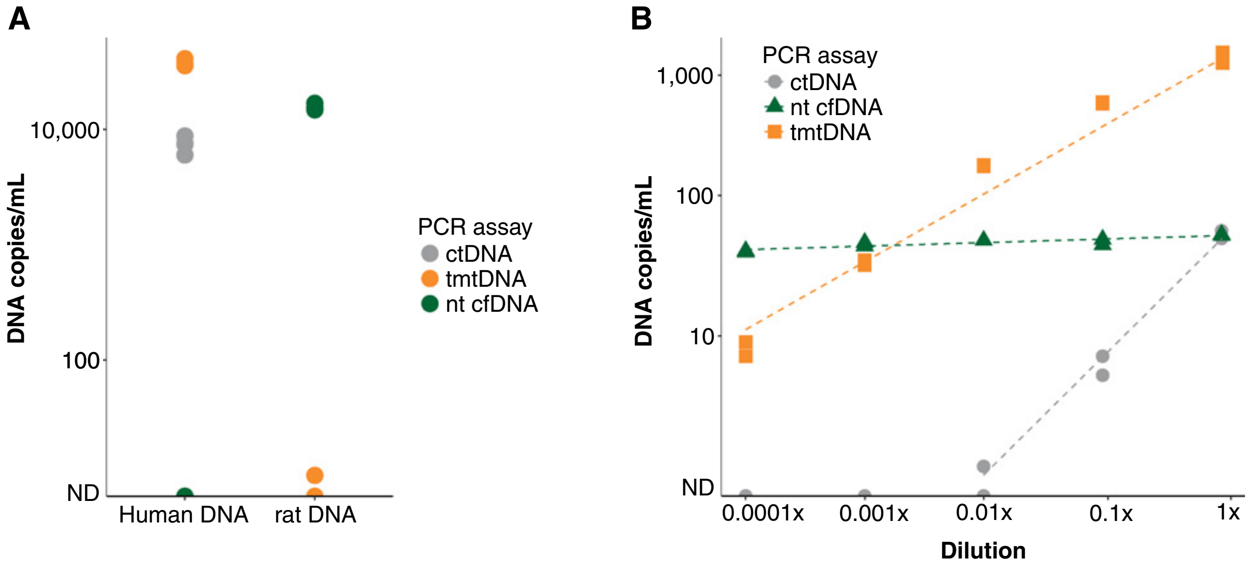


Figure 1.

Validation of the specificity and sensitivity of the ctDNA and tmtDNA dPCR assays. **A,** dPCR assays designed to detect ctDNA, tmtDNA, and nt cfDNA were tested with human and rat plasma DNA. Samples were tested in quadruplicate for each assay. ND, nondetectable. **B,** Dilution series of human (tumor) DNA in rat (nontumor) DNA, which was used to evaluate the sensitivity of tmtDNA detection in comparison with detection of ctDNA. tmtDNA was detected at 100× greater dilution than ctDNA. Each sample was measured in duplicate.

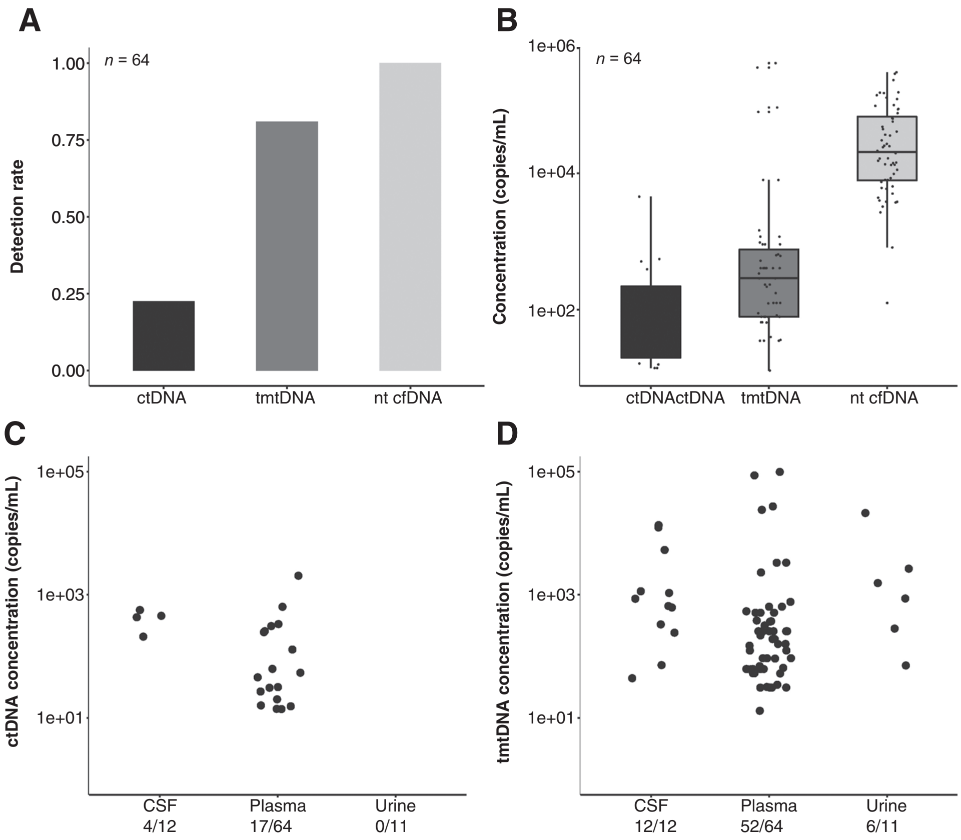


Figure 2.

tmtDNA was detected more frequently in plasma than ctDNA. **A,** Detection rates for plasma circulating ctDNA, tmtDNA, and nt cfDNA in 64 animals implanted orthotopically with cells derived from tumors from 6 different patients. **B,** Concentration (copies/mL) of circulating nucleic acids in the samples where these were detected. ctDNA (**C**) and tmtDNA concentrations (**D**) in CSF, plasma, and urine from the tumor-bearing animals. In the ratios shown below the plots in **C** and **D**, the numerator represents the number of samples containing the indicated DNA and the denominator the number of samples.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

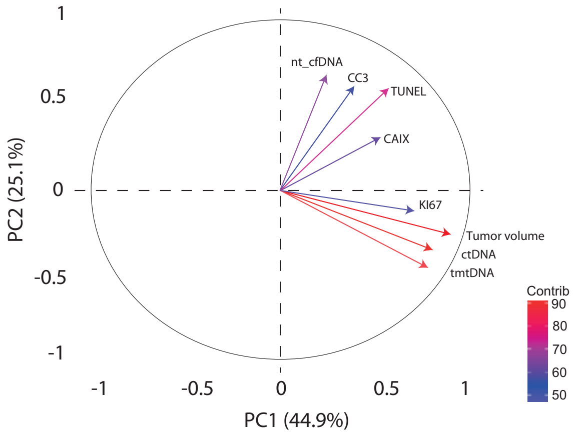


Figure 3.

Factors affecting the levels of ctDNA and tmtDNA in the plasma of treatment-naïve tumor-bearing rats. PCA of variables associated with tumor histology and circulating nucleic acids in the plasma of rats with GBM4 tumors ($n = 36$). The vectors represent ctDNA, tmtDNA, and nontumor cfDNA concentrations, tumor volume, tumor proliferation (Ki67), hypoxia (CAIX), necrosis (TUNEL), and apoptosis (CC3). PC1 (44.9%) and PC2 (25.1%) indicate the % variance accounted for by the two principal components.

concentrations than ctDNA (t test, $P < 0.001$) with a mean concentration of 6,989 copies/mL (Fig. 2A and B). Variable detection rates were observed between the different PDOX models, with tmtDNA detected in 66% of some models (GBM1; $n = 8$) and 100% in others (GBM5; $n = 6$; Supplementary Fig. S2A). ctDNA and tmtDNA were not detected in plasma from nongrafted animals ($n = 4$; Supplementary Fig. S2B).

ctDNA has been detected at low concentrations in urine from patients with nonbrain tumors (17). Urine samples from 11 tumor-bearing animals (10 GBM4 and 1 GBM5) had undetectable levels of ctDNA. However, tmtDNA was identified in 60% of samples with a median concentration of 606 copies/mL (Fig. 2C and D). The CSF presents another possible source of cell-free tumor DNA. We collected an average of 97 μL of CSF (7) from 12 PDOXs (10 GBM4, 1 GBM1, and 1GBM2). ctDNA was detected in 4 of 12 samples (median concentration of 222 copies/mL), and tmtDNA was detected in all samples (median concentration of 760 copies/mL; Fig. 2C and D). Rat host cell-free nuclear DNA was detected in all samples with a median concentration of 215 copies/mL.

ctDNA and tmtDNA levels correlate with tumor size and cell proliferation in treatment-naïve PDOXs

We performed PCA on 8 tumor-related variables in treatment-naïve GBM4 models ($n = 36$). The first component included plasma ctDNA and tmtDNA concentrations, tumor volume, and Ki67 staining, a marker of cell proliferation, and the second component the plasma concentration of host nontumor cell-free nuclear DNA, staining for TUNEL and CC3, which are cell death markers, and carbonic anhydrase 9 (CAIX), a marker of hypoxia (Fig. 3). Correlations (Pearson analysis) were observed between tmtDNA and ctDNA ($R^2 = 0.83$, $P < 0.001$), tumor-derived DNA and tumor volume (tmtDNA $R^2 = 0.86$, $P < 0.001$; ctDNA $R^2 = 0.83$, $P < 0.001$), and tmtDNA and ctDNA and the number of proliferating cells (Ki67-positive cells; tmtDNA $R^2 = 0.54$, $P < 0.001$; ctDNA $R^2 = 0.54$, $P < 0.001$). We also observed a correlation between nt cfDNA and cell death (TUNEL $R^2 = 0.62$, $P < 0.001$ and CC3 $R^2 = 0.47$, $P < 0.01$). Tumor microvessel density was not significantly different between the different PDOX models ($P = 0.27$; Supplementary Fig. S3).

ctDNA and tmtDNA levels correlate with cell death following treatment with temozolomide and radiotherapy

In GBM4 ($n = 36$), tmtDNA and ctDNA were highly correlated with tumor volume ($R^2 = 0.8$; $P \leq 0.0001$; Fig. 4A), suggesting that tmtDNA, like ctDNA, could be used to track tumor burden and monitor treatment response.

We analyzed plasma from GBM4 PDOX models 72 hours after treatment with temozolomide plus radiotherapy (15 Gy, $n = 7$). ctDNA detection frequency and concentration increased (from 40%, 7 copies/mL to 75%, 54 copies/mL, $P = 0.051$; Fig. 4B). tmtDNA concentration also increased (from a median 121 copies/mL to 256 copies/mL, $P = 0.094$; Fig. 4B), but detection frequency remained unchanged (6/7 cases). These increases in ctDNA and tmtDNA concentrations were associated with an increase in tumor cell death, as assessed by TUNEL ($P = 0.039$; Fig. 4C) and CC3 staining ($P = 0.037$) of tumor sections ($n = 7$; Fig. 4D), with a correlation being observed between ctDNA and CC3 staining ($R^2 = 0.58$, $P = 0.074$, Pearson analysis), which is a marker of early apoptosis (18).

These data indicate that in treatment-naïve models, tumor DNA release was related to tumor burden and cell proliferation, whereas following treatment, tumor DNA was released primarily through tumor cell death.

Genome-wide sequencing showed a different fragmentation pattern for ctDNA and host DNA in treatment-naïve PDOXs

We used genome-wide sequencing at low coverage (<0.2×) to determine copy-number profiles of host rat and human (tumor) nuclear genomes in plasma, CSF, and tumor tissue. Paired-end sequencing reads were aligned to rat (RGSC 6.0/rn6) and human (hg19) genomes, and assigned to the appropriate species (Fig. 5A). Similar copy-number profiles were found in tumor DNA from the different fluid compartments and from tumor tissue (Fig. 5B), even though the plasma compartment exhibited a lower tumor DNA fraction, relative to host DNA, when compared with tumor tissue and CSF. We also determined the size distribution of human (tumor) and rat (host) circulating nuclear DNA fragments (Fig. 5C–E) in the plasma of animals grafted with GBM6 (Fig. 5C) and with GBM4 (Fig. 5D). We also determined, for one animal implanted with GBM4, the size distribution of the DNA fragments from CSF (Fig. 5E). The fragment size distribution in plasma and CSF showed a peak at 133–145 bp for human (tumor) DNA, and a different fragmentation pattern for host rat DNA, with a peak at 167 bp (Fig. 5C and D). Mitochondrial DNA showed a peak below 100 bp for both human (tumor) and rat circulating mitochondrial DNA (Fig. 5F), in agreement with previous work (19).

The BBB has a limited effect on plasma ctDNA and tmtDNA concentrations

Despite extensive disruption of the BBB during gliomagenesis (20), the low levels of ctDNA observed in the plasma of patients with GBM and the apparent enrichment of tumor DNA in the CSF have been attributed to the impermeability of the BBB (3). This was supported by sequencing, where tumor mutations in DNA

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

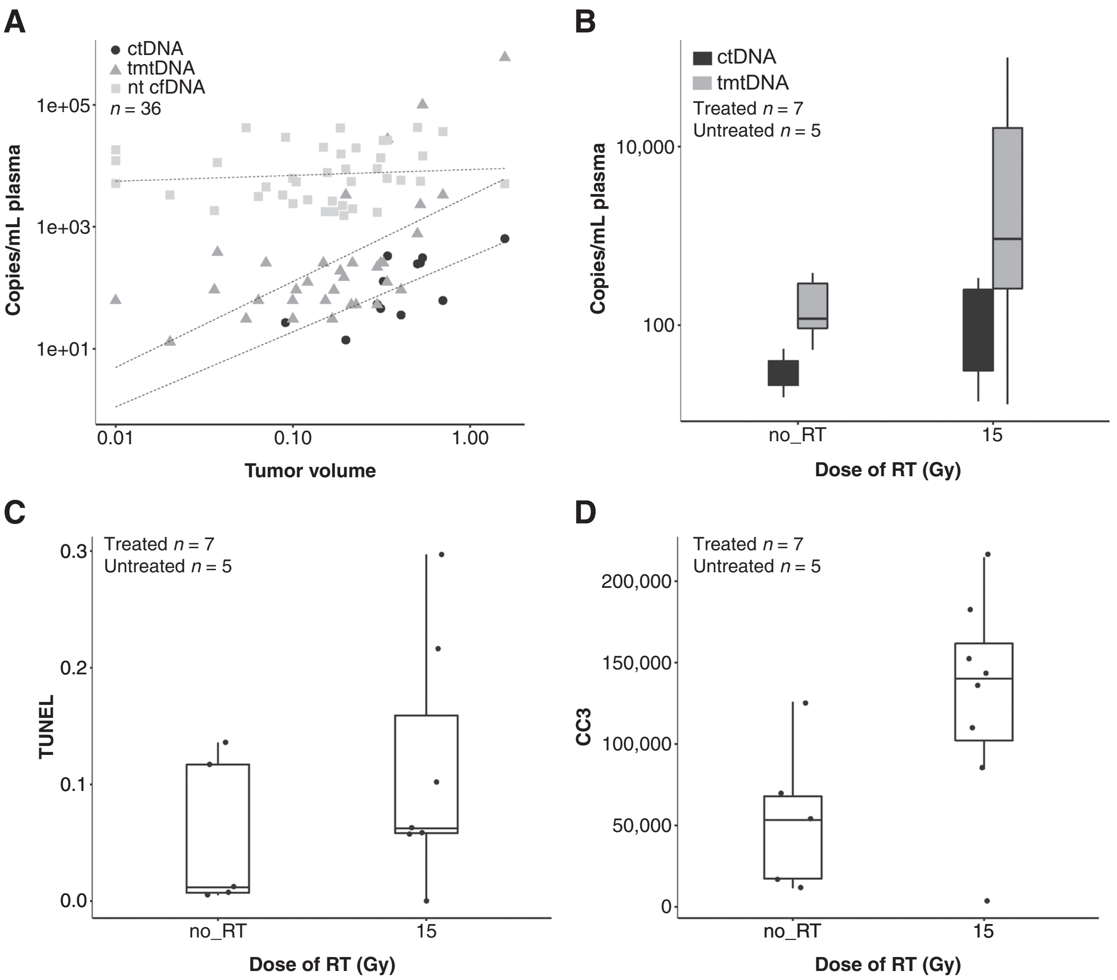


Figure 4.

Factors affecting the levels of ctDNA and tmtDNA in the plasma of tumor-bearing rats following concomitant temozolomide and radiotherapy treatment. **A,** Correlation between tumor volume and the concentrations of ctDNA, tmtDNA, and nt cfDNA in the plasma of animals with GBM4 tumors (*n* = 36) determined by dPCR. **B,** ctDNA and tmtDNA levels in a subset of 7 rats with GBM4 tumors that received 15 Gy with concurrent temozolomide and 5 rats with untreated GBM4 tumors that were analyzed as controls (no_RT). **C** and **D** show levels of cell death in the tumors of these GBM4 models determined by TUNEL and CC3 staining.

from the CSF of patients with GBM were detected more frequently than in plasma and at higher mutant allele fractions (21, 22). However, the absolute concentrations of tumor and nontumor DNA in CSF and in plasma of patients with GBM have not been reported previously. The data shown in Fig. 2 show that the higher detection rate of tumor DNA in CSF is due to a higher concentration of ctDNA relative to host nt cfDNA in CSF (222 copies/mL ctDNA vs. 215 copies/mL nt cfDNA) when compared with plasma (27 copies/mL ctDNA vs. 6,989 copies/mL nt cfDNA). We investigated this further by using dPCR to quantify the concentrations of tmtDNA in plasma and CSF samples collected from 12 of the tumor models [GBM1 (*n* = 1), GBM2 (*n* = 1), GBM4 (*n* = 10)]. tmtDNA concentration was higher in CSF as compared with plasma in each of the tumor models (Fig. 6A), with a median of 476 copies/mL in CSF and 93 copies/mL in plasma. However, CSF volume in the rat is approximately 90 μL and the plasma volume is approximately 6 mL (23), and therefore, the total amount of tmtDNA in the plasma (558 copies) is approximately 13 times higher than in the CSF (43 copies), showing therefore that the BBB does not prevent significant amounts of tumor DNA, at least tmtDNA, from reaching the circulation. Whereas the concentration of tumor-derived cfDNA was 5 to 8 times higher in CSF compared with plasma, the concentration of nt cfDNA was nearly 25 times higher in plasma compared with CSF. Therefore, lower detection rates of tumor-derived DNA in plasma are due, at least in part, to the presence of higher levels of background host DNA in plasma.

To investigate more directly the effect of the BBB on plasma tmtDNA and ctDNA concentrations, we used subcutaneous implantation of GBM4 cells to generate a GBM model that was outside the BBB. We also disrupted the BBB by intravenous administration of mannitol (24). Following mannitol injection, 60 minutes were allowed for ctDNA to escape into the circulation before plasma collection. If the BBB blocks release of tumor DNA into the circulation, then 60 minutes after mannitol injection, there should be an increase in tumor DNA levels in the circulation,

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

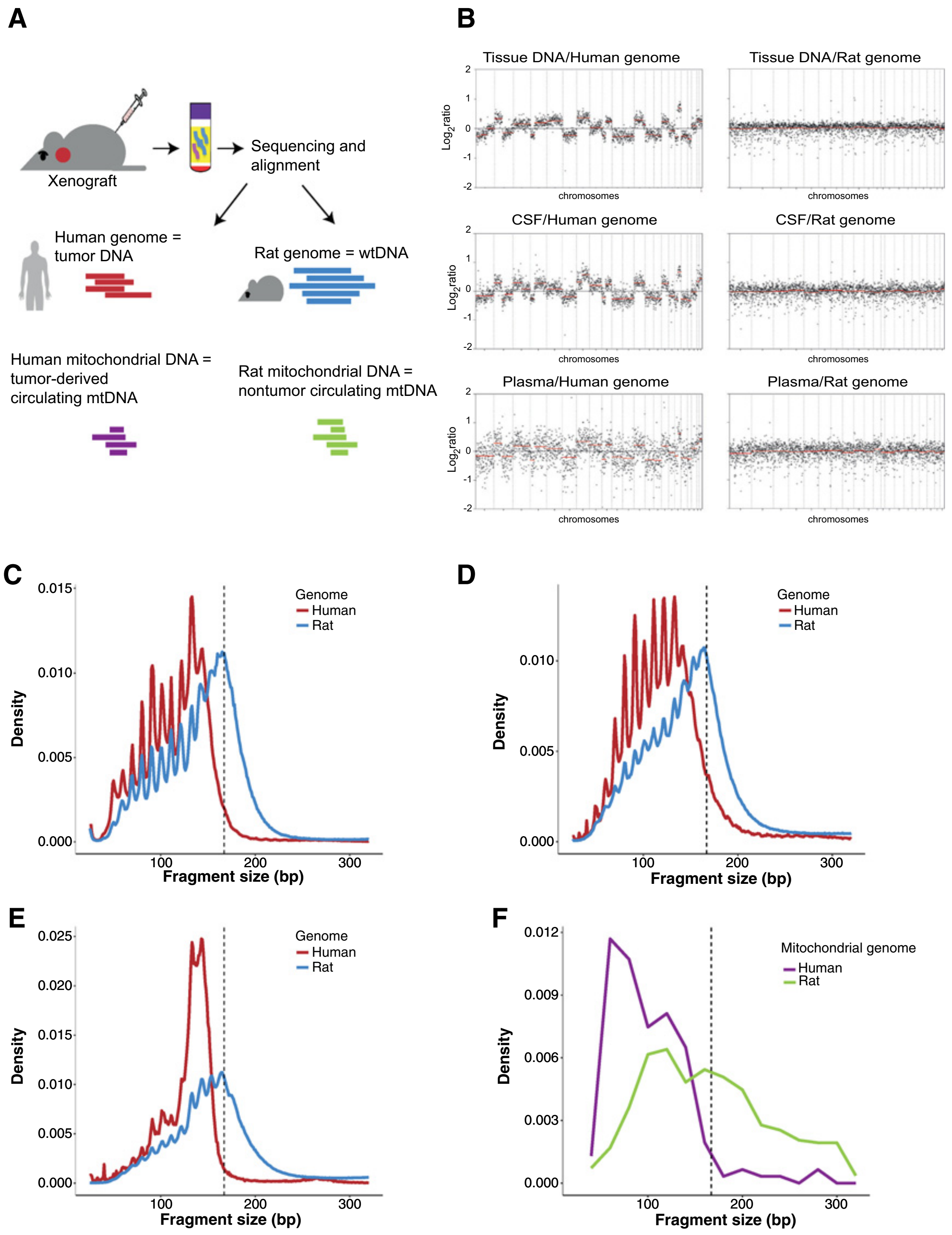


Figure 5.

The fragmentation patterns of plasma DNA in tumor-bearing animals. **A,** Sequencing reads were obtained by paired-end shallow WGS of plasma DNA and aligned to the human (tumor) and rat (host) genomes. **B,** Copy-number profiles obtained from sWGS of DNA from tumor tissue, CSF, and plasma from the GBM4 model ($n = 3$), separated into reads that aligned with the human and rat genome. **C,** Size distribution of DNA fragments of nuclear origin from a plasma sample from the GBM6 model. ctDNA fragments originating from tumor cells, aligned to the human genome, are shown in red, whereas nt cfDNA fragments from host cells, aligned to the rat genome, are shown in blue. A vertical line (at 167 bp) indicates fragment sizes associated with nt cfDNA of apoptotic origin. **D,** Size distribution of the DNA fragments of nuclear origin from a plasma sample from a GBM4 tumor–bearing animal. **E,** Size distribution of the DNA fragments of nuclear origin from a CSF sample from a GBM4 tumor–bearing animal. **F,** Size distribution of mitochondrial DNA fragments from a plasma sample. tmtDNA originating from tumor cells, aligned to the human mitochondrial genome, are shown in purple, and nontumor mitochondrial DNA, aligned to the rat genome, are shown in green.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

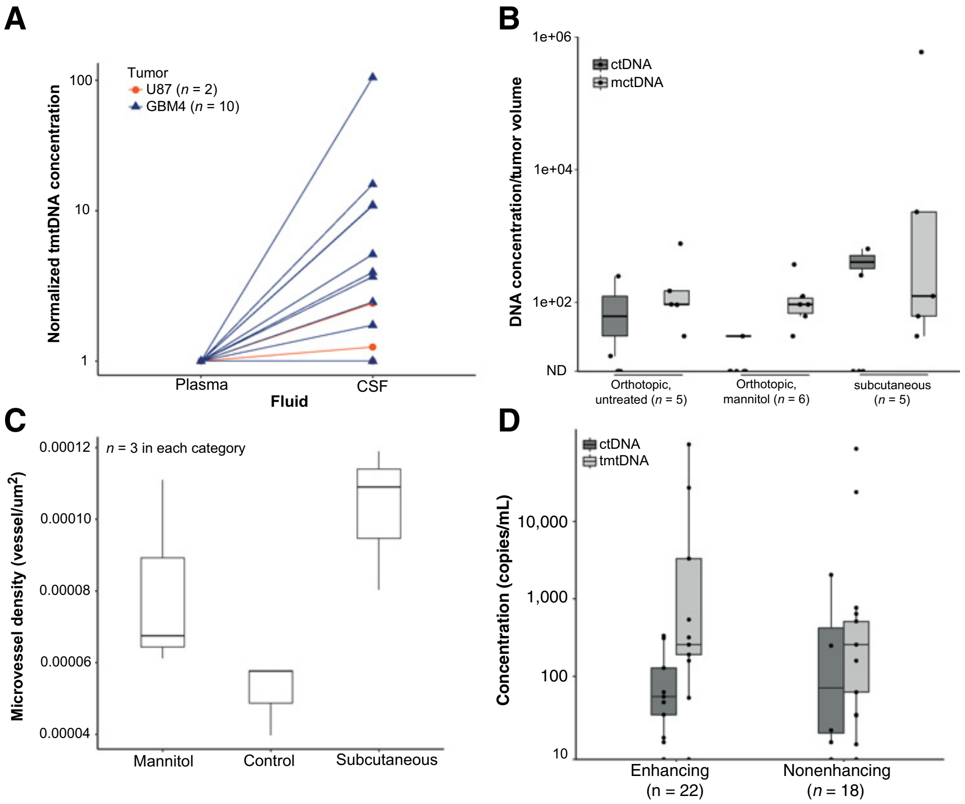


Figure 6.

The integrity of the BBB has little effect on ctDNA and tmtDNA levels in plasma. **A,** Pairwise comparison of tmtDNA in plasma and CSF collected at the same time from 10 GBM4 tumor–bearing animals and two U87 tumor–bearing animals. Concentrations determined in the CSF are plotted relative to the concentration detected in plasma samples. **B,** Concentrations of ctDNA and tmtDNA, normalized to tumor volume, in the plasma of GBM4 tumor–bearing rats, where the tumors were implanted orthotopically ($n = 5$), with or without disruption of the BBB by mannitol injection ($n = 6$), or where the tumors were implanted subcutaneously ($n = 5$). There were no significant differences in the concentrations of ctDNA or tmtDNA between these groups [ANOVA ($P = 0.57$) and individual paired *t* tests ($P > 0.2$); $n = 16$]. **C,** Microvessel density in each tumor model ($n = 3$ per cohort), as analyzed by *in situ* hybridization with a CD31 mRNA probe. **D,** Plasma ctDNA and tmtDNA concentrations in animal models with tumors that showed signal enhancement ($n = 22$; GBM1, 3, 4), and those that did not enhance ($n = 18$; GBM1, 2, 3, 4), in T_1-weighted MR images following administration of a gadolinium-based contrast agent. There was no significant difference between the groups ($P = 0.26$). Additional points represent outliers.

given that maximal BBB opening occurs 5 minutes following mannitol infusion (24), and the circulating DNA half-life is 16 minutes (25). Gadolinium-based contrast agents do not cross the intact BBB and are used commonly for MR imaging of BBB breakdown in GBM (26). We confirmed in 3 control rats that mannitol infusion caused BBB disruption using DCE MRI measurements. Within 10 minutes of mannitol administration, there was an increase in the fraction of tissue accessible to the contrast agent ($P < 0.02$) and in the contrast agent concentration [untreated, 6.3 ± 4.0 μmol/L (SD); post mannitol, 14.8 ± 1.8 μmol/L (SD), $P < 0.025$ (one sided Welch *t* test; Supplementary Fig. S4A and S4B)]. There were no significant differences in the concentrations of ctDNA or tmtDNA between the three groups, after normalization to tumor volume, which was determined using T_2-weighted MRI (orthotopic model) and caliper measurements (subcutaneous model; one-way ANOVA $P = 0.57$ and individual *t* tests; $n = 16$; Fig. 6B). Moreover, the tumor volume–corrected concentrations of ctDNA detected in animals with subcutaneous GBM tumors were much lower than those reported for animals implanted subcutaneously with other tumor types (27, 28). Analysis of CD31 expression (an endothelial cell marker) showed increased microvessel density in subcutaneous versus orthotopic tumors ($P = 0.0173$; Fig. 6C); however, there was no significant difference in the ctDNA or tmtDNA levels (Fig. 6B), suggesting that release from the subcutaneous tumors was not affected by vascular density. Comparison of ctDNA and tmtDNA concentrations in contrast agent enhancing (GBM4, GBM3) and nonenhancing (GBM1, GBM5) tumors showed no differences in ctDNA ($P = 0.65$) or tmtDNA concentrations ($P = 0.49$) between these groups (Fig. 6D).

Discussion

Detection of ctDNA in patients with GBM is challenging because of low plasma concentrations (3). Sampling of CSF has been proposed as a method for detecting ctDNA in GBM (21, 22, 29); however, lumbar puncture is contraindicated in patients with intracerebral space occupying lesions, and thus routine use of this technique is not clinically feasible (30, 31). Nevertheless, the requirement for minimally invasive techniques that avoid

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

repeated biopsies in patients with GBM remains due to current inadequacies in identifying treatment response/escape (32) and the evolving nature of the disease during treatment (33–35). We therefore pursued methods to improve detection of circulating tumor-derived nucleic acids through the use of PDOX models of GBM and used these methods to identify factors affecting DNA release.

dPCR was used to estimate plasma tmtDNA and ctDNA concentrations in a large number of PDOX models of GBM. The detection rate for tmtDNA was 82% in plasma samples ($n = 64$), at an average concentration of 5,081 copies/mL, versus a detection rate for ctDNA of 24%, at an average concentration of 27 copies/mL. Host cell-free nuclear DNA concentrations have a broad range, and the values we report are within the range reported previously for animal models (27, 28). tmtDNA was also detected in 60% of urine samples in which ctDNA was undetectable. Because tmtDNA is highly fragmented in plasma (Fig. 5F), *in vitro* or *in silico* size selection of fragments below 100 bp could be used to sieve tmtDNA from nuclear ctDNA, enriching the sample for tmtDNA and further enhancing the sensitivity of detection (36). The potential for tmtDNA to be used to detect smaller tumors, either at diagnosis or at recurrence, would be important clinically.

Using both ctDNA and tmtDNA, we investigated the factors influencing release of tumor-derived nucleic acids into the circulation. The levels of both were correlated with tumor size, in agreement with previous preclinical (27, 28) and clinical (3, 25) studies. Previous analyses of cfDNA fragment sizes in plasma showed these to be mostly distributed around 167 bp, and multiples thereof, characteristic of caspase-dependent cleavage and suggesting that the majority of cfDNA originates from apoptosis (37, 38). In patients with cancer, a shortening of cfDNA was observed (39, 40), which could reflect modifications in chromatin organization (41, 42). Recent work on fetal cfDNA suggested that methylation-related chromatin reorganization can result in shortening of fragment length (38, 43). The first comprehensive analysis of the relationship between tumor physiology and ctDNA in patients indicated that cell proliferation and tumor volume are more strongly correlated with ctDNA concentration than cell death (2). Here, we have shown, in treatment-naïve PDOX models, that there is a correlation between nontumor (host) cfDNA levels and cell death. Fragmentation analysis showed a distribution centered around 167 bp, consistent with release from apoptotic host cells. We observed a correlation between ctDNA levels and tumor volume, and to a lesser extent with cell proliferation, but not with cell death, as was observed previously (2). Analysis of ctDNA fragment sizes revealed a shift toward shorter fragment sizes, with a distribution centered around 145 bp, corresponding to the core nucleosome. These findings suggest that size selection could potentially be used to improve the yield of ctDNA fragments (36).

The concentrations of plasma ctDNA and tmtDNA were increased following temozolomide and radiotherapy treatment and, in this instance, were related to an increase in tumor cell death. However, CC3 and TUNEL staining only inform upon a proportion of dying cells and not those affected by mitotic catastrophe or senescence for example. In these treated animals, there was no longer any correlation between plasma levels of ctDNA and tmtDNA and cell proliferation. Therefore, it appears that release of tumor DNA before and after treatment occurs via different processes. DNA release via cell death after treatment may be explained by the requirement for tumor cells to be in close proximity to viable blood vessels, which provide the oxygen necessary for radiotherapy-induced tumor cell kill (44). Thus, when these cells die, they do so in a vessel-rich microenvironment, and are distinct from dying tumor cells in treatment-naïve GBM, where cell death may occur predominantly in cells with a poor blood supply.

The BBB has been proposed as the main reason for reduced ctDNA detection in GBM (21). Our experiments, in which we circumvented the BBB via heterotopic tumor engraftment or opened the BBB using mannitol, suggest that the effect of the BBB on release of tumor-derived DNA into the plasma may be less significant than previously thought. Recent studies have shown higher relative levels of mutant DNA in CSF compared with plasma of patients with GBM (21, 22), which has been interpreted as being due to enrichment of tumor DNA in the CSF. Using dPCR to measure absolute concentrations of tumor and host DNA, we found that higher relative levels of tumor DNA in CSF resulted primarily from lower concentrations of nontumor host DNA together with more modest increases in the quantity of tumor-derived DNA.

Although we used single-copy human mitochondrial sequences to identify tmtDNA in the PDOX models, this strategy is not directly applicable to a human patient. However, mitochondrial mutations are present in the majority of cancers, with frequencies depending upon the tumor of origin, and mutational "hotspot" regions have been identified (45), suggesting that mutated mitochondrial sequences could be used to detect tmtDNA in the clinic (12). Whole-genome sequencing (WGS) has enabled detection of mitochondrial DNA variant-allele fractions down to 1% (46); moreover, studies have shown that certain tumors positively select for nonsynonymous mitochondrial DNA mutations (47). Although WGS is expensive, the small size of the mitochondrial genome means that targeted and/or capture-sequencing-based methods could provide a more affordable alternative and may enable improved sequencing depth (48). Genome-wide or targeted sequencing of the tumor tissue DNA obtained at surgery may also permit strategies whereby dPCR probes or focused sequencing assays may be employed to track tmtDNA mutations in plasma.

Several cancers have higher mitochondrial copy numbers, thus further increasing the probability of detecting tmtDNA (49). Detection of ctDNA in IDH1-mutant glioma, for example, has demonstrated limited clinical efficacy (3); however, the high tmtDNA copy number in these tumors may make circulating tmtDNA analysis achievable (49). Recent studies have identified certain cancers with functional tmtDNA mutations that affect metabolism (46). This could be used to target metabolic therapies to tumors with known metabolic weaknesses (46). tmtDNA mutations have also been described that confer specific chemoresistant properties (50). Thus, their monitoring via serial liquid biopsy may enable therapy modulation, as has been demonstrated with the use of ctDNA (4).

In conclusion, release of tmtDNA and ctDNA is correlated with tumor volume and tumor cell proliferation in treatment-naïve tumors and with tumor cell death following treatment. The BBB appears to play only a minor role in preventing release of glioma-derived ctDNA into plasma. Analysis of circulating tmtDNA can improve the sensitivity of detection of tumor DNA in multiple body fluids and may make plasma liquid biopsy possible for

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

patients with gliomas, where detection rates for ctDNA have so far been very low.

Disclosure of Potential Conflicts of Interest

D. Gale is co-founder of, and has an ownership interest (including stock, patents, etc.) in, Inivata Ltd. F. Marass has an ownership interest (including stock, patents, etc.) in patent and shares of Inivata Ltd. D.W.Y. Tsui received honoraria from the speakers' bureau of Astrazenaca and National Taiwan University, and is a consultant/advisory board member for Inivata Ltd. N. Rosenfeld is CSO at Inivata Ltd., reports receiving commercial research grant from Astrazenaca, and has an ownership interest (including stock, patents, etc.) in patents/applications. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: R. Mair, F. Mouliere, C.G. Smith, C. Watts
Development of methodology: R. Mair, F. Mouliere, C.G. Smith, D. Gale, C. Watts
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Mair, F. Mouliere, D. Gale, D.W.Y. Tsui, A.J. Wright, C. Watts
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Mair, F. Mouliere, C.G. Smith, D. Chandrananda, D. Gale, F. Marass, C.E. Massie, A.J. Wright, N. Rosenfeld
Writing, review, and/or revision of the manuscript: R. Mair, F. Mouliere, C.G. Smith, D. Chandrananda, D. Gale, F. Marass, D.W.Y. Tsui, A.J. Wright, C. Watts, N. Rosenfeld, K.M. Brindle
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R. Mair, F. Mouliere
Study supervision: F. Mouliere, C. Watts, N. Rosenfeld, K.M. Brindle

Acknowledgments

The authors would like to acknowledge Professor Richard J. Gilbertson and Dr. Irena Hudecova for fruitful discussions. We wish to thank for their help and support the Cancer Research UK Cambridge Institute core facilities, in particular the biological resource unit, genomics, histopathology, and preclinical imaging sections. We wish also to thank the Cambridge Molecular Diagnostic Laboratory, and in particular Dr. Mikel Velganon. We would like to thank also Stephen Price, Dr. De-en Hu, Leigh-Anne McDuffus, Jodi Miller, Bev Wilson, Julia Jones, Mike Mitchell, Lisa Young, and Gemma Bullen.

N. Rosenfeld and K.M. Brindle are supported by the University of Cambridge, Cancer Research UK (grant numbers A11906, A20240, 17242, and 16465) and Hutchison Whampoa Limited. N. Rosenfeld has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 337905. C. Watts is supported by The Brain Tumour Charity grant 10/136.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 5, 2018; revised June 29, 2018; accepted October 26, 2018; published first November 2, 2018.

References

1. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 2017;17:223–38.
2. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 2017;545:446–51.
3. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014;6:224ra24.
4. Dawson S-J, Tsui DWY, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013;368:1199–209.
5. Wang J, Bettegowda C. Applications of DNA-based liquid biopsy for central nervous system neoplasms. J Mol Diagn 2017;19:24–34.
6. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJ, Janzer RC, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol 2009;10:459–66.
7. Mahat MYA, Ahamed NFA, Chandrasekaran S, Rajagopal S, Narayanan S, Surendran N. An improved method of transcutaneous cisterna magna puncture for cerebrospinal fluid sampling in rats. J Neurosci Methods 2012;211:272–9.
8. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res 2014;24: 2022–32.
9. Noebauer-Huhmann IM, Szomolanyi P, Juras V, Kraff O, Ladd ME, Trattnig S. Gadolinium-based magnetic resonance contrast agents at 7 Tesla: in vitro T1 relaxivities in human blood plasma. Invest Radiol 2010;45:554–8.
10. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging 1999;10:223–32.
11. McGrath DM, Bradley DP, Tessier JL, Lacey T, Taylor CJ, Parker GJ. Comparison of model-based arterial input functions for dynamic contrast-enhanced MRI in tumor bearing rats. Magn Reson Med 2009;61: 1173–84.
12. Ju YS, Alexandrov LB, Gerstung M, Martincorena I, Nik-Zainal S, Ramakrishna M, et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. eLife 2014; eLife.02935
13. Fliss MS, Usadel H, Caballero OL, Wu L, Buta MR, Eleff SM, et al. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. Science 2000;287:2017–9.
14. Mair RJ, Wright A, Ros S, Hu DE, Booth T, Kreis F, et al. Metabolic imaging of c-Myc expression in orthotopically implanted patient-derived xenograft models of glioblastoma. Cancer Res 2018; 78:5408–18.
15. Yung WK, Luna M, Borit A. Vimentin and glial fibrillary acidic protein in human brain tumors. J Neurooncol 1985;3:35–8.
16. Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, et al. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell 2006;9:391–403.
17. Su Y-H, Wang M, Brenner DE, Norton PA, Block TM. Detection of mutated K- rasDNA in urine, plasma, and serum of patients with colorectal carcinoma or adenomatous polyps. Ann N Y Acad Sci 2008;1137:197–206.
18. Armstrong RC, Aja TJ, Hoang KD, Gaur S, Bai X, Alnemri ES, et al. Activation of the CED3/ICE-related protease CPP32 in cerebellar granule neurons undergoing apoptosis but not necrosis. J Neurosci 1997;17:553–62.
19. Jiang P, Chan CW, Chan KC, Cheng SH, Wong J, Wong VW, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. Proc Natl Acad Sci U S A 2015;112:E1317–25.
20. Dubois LG, Campanati L, Righy C, D'Andrea-Meira I, Spohr TC, Porto-Carreiro I, et al. Gliomas and the vascular fragility of the blood brain barrier. Front Cell Neurosci 2014;8:418.
21. De Mattos-Arruda L, Mayor R, Ng CKY, Weigelt B, Martínez-Ricarte F, Torrejon D, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun 2015;6:1–6.
22. Wang Y, Springer S, Zhang M, McMahon KW, Kinde I, Dobbyn L, et al. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. Proc Natl Acad Sci USA 2015;112:9704–9.
23. Lee HB, Blaufox MD. Blood volume in the rat. J Nucl Med 1985;26:72–6.
24. Cosolo WC, Martinello P, Louis WJ, Christophidis N. Blood-brain barrier disruption using mannitol: time course and electron microscopy studies. Am J Physiol 1989;256:R443–7.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

25. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. Nat Med 2007;14:985–90.
26. Sage MR, Wilson AJ. The blood-brain barrier: an important concept in neuroimaging. AJNR Am J Neuroradiol 1994;15:601–22.
27. Thierry AR, Mouliere F, Gongora C, Ollier J, Robert B, Ychou M, et al. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. Nucleic Acids Res 2010;38:6159–75.
28. Rago C, Huso DL, Diehl F, Karim B, Liu G, Papadopoulos N, et al. Serial assessment of human tumor burdens in Mice by the Analysis of Circulating DNA. Cancer Res 2007;67:9364–70.
29. Pentsova EI, Shah RH, Tang J, Boire A, You D, Briggs S, et al. Evaluating cancer of the central nervous system through next-generation sequencing of cerebrospinal fluid. J Clin Oncol 2016;34:2404–15.
30. Hasbun R, Abrahams J, Jekel J, Quagliarello VJ. Computed tomography of the head before lumbar puncture in adults with suspected meningitis. N Engl J Med 2001;345:1727–33.
31. Engelborghs S, Niemantsverdriet E, Struyfs H, Blennow K, Brouns R, Comabella M, et al. Consensus guidelines for lumbar puncture in patients with neurological diseases. Alzheimers Dement (Amst) 2017;8:111–26.
32. Abdulla S, Saada J, Johnson G, Jefferies S, Ajithkumar T. Tumour progression or pseudoprogression? A review of post-treatment radiological appearances of glioblastoma. Clin Radiol 2015;70:1299–312.
33. Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science 2014;343:189–93.
34. Kim J, Lee I-H, Cho HJ, Park CK, Jung YS, Kim Y, et al. Spatiotemporal evolution of the primary glioblastoma genome. Cancer Cell 2015;28: 318–28.
35. Wang J, Cazzato E, Ladewig E, Frattini V, Rosenbloom DI, Zairis S, et al. Clonal evolution of glioblastoma under therapy. Nat Genet 2016;48: 768–76.
36. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. Sci Transl Med 2018;10. doi: 10.1126/scitranslmed.aat4921.
37. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659–65.
38. Lo YM, Chan KC, Sun H, Chen EZ, Jiang P, Lun FM, et al. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med 2010;2:61ra91.
39. Mouliere F, Robert B, Arnau Peyrotte E, Del Rio M, Ychou M, Molina F, et al. High fragmentation characterizes tumour-derived circulating DNA. PLoS One 2011;6:e23418.
40. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment length of circulating tumor DNA. PLoS Genet 2016;12: e1006162.
41. Snyder MW, Kircher M, Hill AJCell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 2016;164: 57–68.
42. Sun K, Jiang P, Chan KC, Wong J, Cheng YK, Liang RH, et al. Plasma DNA tissue mapping by genome wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. Proc Natl Acad Sci USA 2015;112:E5503–12.
43. Sun K, Jiang P, Wong AIC, Cheng YKY, Cheng SH, Zhang H, et al. Sizetagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. Proc Natl Acad Sci U S A 2018;115:E5106–E5114.
44. Gray LH, Conger AD, Ebert M, Hornsey S, Scott OC. The concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy. Br J Radiol 1953;26:638–48.
45. Yeung K, Dickinson A, Donoghue JF, Polekhina G, White SJ, Grammatopoulos DK, et al. The identification of mitochondrial DNA variants in glioblastoma multiforme. Acta Neuropathol Commun 2014;2:1.
46. Yuan Y, Ju YS, Kim Y, Li J, Wang Y, Yang Y, et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. bioRxiv 2017; https://doi.org/10.1101/161356.
47. Grandhi S, Bosworth C, Maddox W, Sensiba C, Akhavanfard S, Ni Y, et al. Heteroplasmic shifts in tumor mitochondrial genomes reveal tissue-specific signals of relaxed and positive selection. Hum Mol Genet 2017;26:2912–22.
48. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 2016;34:547–55.
49. Reznik E, Miller ML, Senbabaoglu Y, Riaz N, Sarungbam J, Tickoo SK, et al. Mitochondrial DNA copy number variation across human cancers. eLife 2016; eLife.10769.
50. Guerra F, Perrone AM, Kurelac I, Santini D, Ceccarelli C, Cricca M, et al. Mitochondrial DNA mutation in serous ovarian cancer: implications for mitochondria-coded genes in chemoresistance. J Clin Oncol 2012;30: e373–8.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

Published OnlineFirst November 2, 2018; DOI: 10.1158/0008-5472.CAN-18-0074

American Association for Cancer Research

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Measurement of Plasma Cell-Free Mitochondrial Tumor DNA Improves Detection of Glioblastoma in Patient-Derived Orthotopic Xenograft Models

Richard Mair, Florent Mouliere, Christopher G. Smith, et al.

Cancer Res 2019;79:220-230. Published OnlineFirst November 2, 2018.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-18-0074
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2018/11/02/0008-5472.CAN-18-0074.DC2 http://cancerres.aacrjournals.org/content/suppl/2018/10/31/0008-5472.CAN-18-0074.DC1

Cited articles	This article cites 46 articles, 17 of which you can access for free at: http://cancerres.aacrjournals.org/content/79/1/220.full#ref-list-1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/79/1/220. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.

Downloaded from cancerres.aacrjournals.org on January 24, 2019. © 2019 American Association for Cancer Research.

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

Selecting short DNA fragments in plasma improves detection of circulating tumour DNA

Florent Mouliere[1,2,*], Anna M. Piskorz[1,2,*], Dineika Chandrananda[1,2,*], Elizabeth Moore[1,2,3,*], James Morris[1,2], Christopher G. Smith[1,2], Teodora Goranova[1,2], Katrin Heider[1,2], Richard Mair[1,2], Anna Supernat[1,2,4], Ioannis Gounaris[1,2,3], Susana Ros[1,2], Jonathan C. M. Wan[1,2], Mercedes Jimenez-Linan[2,3], Davina Gale[1,2], Kevin Brindle[1,2,5], Charles E. Massie[1,2], Christine A. Parkinson[1,2,3,6,7], James D. Brenton[1,2,3,6,7,#], Nitzan Rosenfeld[1,2,#]

[*] co-first authors;

[#] corresponding authors: james.brenton@cruk.cam.ac.uk; nitzan.rosenfeld@cruk.cam.ac.uk

1. Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK
2. Cancer Research UK Major Centre – Cambridge, Cancer Research UK Cambridge Institute, Cambridge, UK
3. Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
4. Medical University of Gdansk, Gdansk, Poland
5. Department of Biochemistry, University of Cambridge, Cambridge, UK
6. Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, Cambridge, UK
7. NIHR Cambridge Biomedical Research Centre, Cambridge, UK

Introductory paragraph:

Non-invasive analysis of cancer genomes using cell-free circulating tumour DNA (ctDNA) is being widely implemented for clinical indications. The sensitivity for detecting the presence of ctDNA and genomic changes in ctDNA is limited by its low concentration compared to cell-free DNA of non-tumour origin. We studied the feasibility for enrichment of ctDNA by size selection, in plasma samples collected before and during chemotherapy treatment in 13 patients with recurrent high-grade serous ovarian cancer. We evaluated the effects using targeted and whole genome sequencing. Selecting DNA fragments between 90-150 bp before analysis yielded enrichment of mutated DNA fraction of up to 11-fold. This allowed identification of adverse copy number alterations, including *MYC* amplification, otherwise not observed. Size selection allows detection of tumour alterations masked by non-tumour DNA in plasma and could help overcome sensitivity limitations of liquid biopsy for applications in early diagnosis, detection of minimal residual disease, and genomic profiling.

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Text:

Analysis of circulating tumour DNA (ctDNA) by non-invasive sampling of cell-free DNA from plasma is now becoming an important tool in oncology for molecular stratification, monitoring tumour burden and analysis of genomic evolution during treatment[1–3]. Analysis of ctDNA is technically challenging, as ctDNA is often present in low concentrations in plasma and is mixed with cell-free DNA (cfDNA) of non-cancerous origin, which is generally present at much higher concentrations. In patients with advanced cancers, the median concentration of ctDNA can reach 10% or more of the total cfDNA, but this fraction is much lower in earlier stage cancer, and ctDNA may rapidly decrease following initiation of systemic treatment or surgery[1,2,4,5]. Various strategies have been proposed to improve the sensitivity of ctDNA analysis[2]. Such methods generally focus on a small subset of the genome, such as hot-spot PCR-based assays or ultra-deep sequencing across gene panels[2,6–9]. Recent observations that ctDNA fragments may be shorter than non-tumour cfDNA in plasma has led to suggestions that these differences may be exploited to enrich for the tumour-specific signal in plasma DNA[10–14]. *In-silico* analysis of ctDNA size differences has been used to enhance the signal for chromosomal changes[13]. However, physical size selection to filter out non-tumour DNA prior to large-scale genomic sequencing has not been demonstrated. Therefore, we tested the hypothesis that selecting DNA fragments of specific sizes could improve the sensitivity of detecting genomic alterations in cfDNA from plasma of cancer patients, enabling the detection of point mutations and copy number alterations that are previously undetectable[11].

Previous reports using paired-end sequencing reads revealed that cell-free DNA is mostly distributed around a mode at 167 bp, a length that could correspond to the chromatosome (core histones + linker)[15,16]. This size distribution pattern is characteristic of a caspase-dependant cleavage, therefore it was hypothesized that apoptosis releases a large fraction of cfDNA into the bloodstream[14,15,17]. Previous studies in non-invasive prenatal testing have explored the potential of size selection for enriching the fetal DNA fraction in maternal plasma with both physical and *in-silico* methods[14,18–20]. However, this analysis of the size distribution cannot be easily generalised to tumour-derived fragments as the characterisation of ctDNA-specific patterns requires analysing fragment sizes of DNA with tumour-derived alterations[13]. Plasma samples from xenografted animal models have shown ctDNA to be highly fragmented below 167 bp[10] and this distribution with a mode at 145 bp has been then confirmed with PCR, atomic force microscopy and recently with whole-genome sequencing[12,21]. If the distribution differs between tumour-derived and non-tumour derived fragments, sieving fragments by their size could reduce the (often overwhelming) fraction of non-tumour cfDNA and improve the signal to noise ratio in downstream analysis.

We first analysed paired-end reads from shallow whole genome sequencing (sWGS) of plasma DNA from animal models xenografted with a human ovarian cancer cells, and confirmed that the distribution of ctDNA differed in this model from non-tumour cfDNA (Fig. 1a, b). ctDNA in this model was enriched in the size range between 90 and 150 bp, while non-tumour cfDNA was dominant at sizes greater than 150 bp, and peaked at ~166 bp, similar to previous observations in animal models and in human samples[12,15,21,22]. Based on these observations we used an automated electrophoresis agarose gel selection method to isolate DNA fragments between 90 and 150 bp for downstream analysis. We sequenced size-selected DNA using both sWGS and tagged-amplicon deep sequencing (TAm-Seq[23]), in 26 samples collected from 13 patients with high-grade serous ovarian cancer (HGSOC), and compared the results to those of the same samples without size selection. For each

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

patient, two plasma samples were collected: one pre-treatment, when levels of ctDNA were generally high, and another several weeks after the start of treatment, when levels of ctDNA were often much lower due to treatment[24]. Analysis of the distribution of fragment lengths after *in-vitro* size selection and sWGS indicated that 96% of resulting reads were in the selected range (Fig. 1c). For two of the patients, the distribution of fragment sizes obtained by sWGS without size-selection exhibited a degraded pattern of cfDNA, without a prominent peak at 166 bp or the 10-bp increment peaks, which has been observed previously[12,13,15], and is present in all the other cases in this study (Suppl. Fig. 1).

Analysis of somatic copy number aberrations (SCNAs) was carried out with sWGS on all plasma samples before and after size selection of DNA fragments between 90 and 150 bp (Fig. 2a). One case is exemplified in Fig. 2b and 2c (see further data in Supplementary Fig. 2). Without size selection, a small number of SCNAs were detected with sWGS in DNA isolated from plasma collected 3 weeks after initiation of treatment from patient OV04-83 (Fig. 2b). These included focal amplifications in chromosomes 8p, 14p, 17, and 19q that were observed in this sample at very low levels (Suppl. Table 2). Analysis of the same DNA sample following size-selection for short DNA fragments revealed an increase in the level of these detected SCNAs, in addition to multiples other SCNAs that were not observed without size selection (Fig. 2c). The same pattern of SCNAs and focal amplifications was observed in DNA from plasma collected from the same patient before initiation of the treatment, when the fraction of tumour DNA in the plasma was higher (Suppl. Fig. 2).

The relative copy number signals in plasma DNA, across a list of 29 genes frequently mutated in HGSOC, were compared with and without size selection of DNA from plasma samples collected during treatment across the cohort of 13 patients. This showed that a large number of SCNAs that were not observed without size selection, could be detected after size selection for shorter DNA fragments, notably as amplifications in key genes such as *NF1*, *PARP2* and *MYC* (Fig. 2d and Suppl. Fig. 3). More SCNAs could be detected after size selection in 11/13 patients, and the absolute level of the log2ratio was significantly increased after size selection (t-test, $p=7.72.10^{-9}$). The 2 patients for whom the SCNA signal did not increase exhibited a degraded pattern of cfDNA, which could explain why the size selection had not enriched for ctDNA (Fig. 2d and Suppl. Fig. 2).

We next assessed the detection of SCNAs and point mutations, in plasma samples of the 13 patients collected at baseline (before treatment initiation) and 3 weeks after initiation of chemotherapy treatments, with and without size selection of the plasma DNA (Fig. 3a and Suppl. Table 1). The data from the baseline samples, where ctDNA levels are generally higher[24], was used to identify and confirm genomic changes, which were then studied in the samples after initiation of treatment, with generally, lower levels of ctDNA. The amplitudes of the absolute log2ratio for the SCNAs were higher, and the concordance of the alterations detected between the baseline and post-treatment samples were improved, with size selection of the plasma DNA (Fig. 3b and Suppl. Fig. 3).

Using sWGS data, we converted the amplitude of the CNAs into a quantitative metric called t-MAD (trimmed Mean Absolute Deviation from copy-number neutrality, see *Methods*). Size selection of plasma DNA resulted in a median of 1.5 fold (n=26) increase in the t-MAD score (figure 3c and 3d). However, in the samples collected after the initiation of treatment, when ctDNA content was low, the genome wide enrichment was higher (Fig. 3d), with a median increase of the t-MAD score of 2.9 fold (range: 0.6 - 4.5 fold), except for two samples. For those two samples we observed a decrease in

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the t-MAD score, and an analysis of the size profile before selection for these samples revealed that the DNA had been heavily fragmented, which could explain why the size selection did not result in enrichment for these cases (Suppl. Fig. 1). In this dataset, we did not identify a differential effect of size selection on the recovery of specific alterations, suggesting that there is a global genome enrichment post size selection. Additionally, size selection notably led to an increased detection of deletions (Suppl. Fig. 2). Analysis with greater sequencing depth, integration of samples from other cancer types and different stages of the disease would help to extend our observations and expand further our understanding of ctDNA biology and fragmentation.

In order to confirm that enrichment for tumour DNA could be observed irrespective of the sequencing approach, we further analysed the mutant allele fractions in the samples using Tagged-Amplicon Sequencing (TAm-Seq). We detected a relative enrichment in the ctDNA fraction in all samples exhibiting a typical pattern of cfDNA fragmentation, with a mode of fragment distribution at 166bp, before size selection (Suppl. Fig. 2). The enrichment effect was below 2-fold in samples collected pre-treatment, when the ctDNA fractions were high in plasma (20%-50% allele fractions for *TP53* mutations as detected by TAm-Seq) (Fig. 3e and 3f). Enrichment of the tumour fraction by size selection was much greater, between 5-fold and 11-fold for most samples, in samples collected approximately 3 weeks after initiation of treatment, when levels of ctDNA (without size selection) were low (ranging from <1% to 5% allele fraction for *TP53* mutations as previously detected by TAm-Seq) (Fig. 3e). For 8 of the 26 plasma samples (31%) we noted a decrease in the allele fractions of mutant *TP53* following size selection; this may be related to loss of rare mutant fragments during the size selection process, or by limitations of this assay for analysis of very short fragments. Increasing the starting amount of material used for size selection, and optimisation of assays for recovery of short fragment, could overcome such limitations.

Methods such as exome-wide sequencing of plasma DNA at multiple time-points of cancer treatment could be effective for the study of cancer evolution and for identification of possible resistance mechanisms to treatment[3,25]. However, analysis with broads-spectrum approaches such as whole-exome sequencing or sWGS are most effective when the ctDNA content is greater than approximately 5%[3,26,27]. In this study, we found that 4 out of 13 cases (31%) where ctDNA levels <5% may have made such analysis uninformative, have been “rescued” by size selection that resulted in >5% mutant allele fractions (Fig. 3e).

These results demonstrate a proof-of-principle that by a simple step of filtering of cfDNA and selection of shorter fragments, it is possible to increase the tumour DNA fraction in plasma cell-free DNA samples. A better understanding of the biology of ctDNA, and their mechanism of release in the circulation, could help select specific fragment ranges depending on their expected origin[28,29]. For example, necrotic cells may release long mutant fragments[14,17], and thus size selection for fragments of hundreds to thousands of bp may be appropriate for certain samples. Alternatively, a size selection of ultra-short fragments might help for enriching a sample for mitochondrial circulating DNA or bacterial DNA, as their DNA is further fragmented below 100bp[13,30]. Other selection or filtration methods can selectively enrich DNA fragments that are bound to specific molecules for example modified nucleosomes[31] Determining and accounting for inherent size-biases induced by DNA isolation method that are currently employed for recovery of cfDNA in practice will be important to standardise and optimise methods for liquid biopsy, as large fractions of short or long DNA could be lost at this step. In addition, recent reports have highlighted that different methods of

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint
(which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.
All rights reserved. No reuse allowed without permission.

library preparation for sequencing may enable more effective recovery of short DNA fragments from plasma samples, which have led to new observations on cfDNA size distributions[28,30,32]. Such methods should be further investigated to determine if these could help recover more ctDNA.

The size selection process we demonstrated here is based on inherent characteristics of ctDNA in comparison to cfDNA and does not require alteration of these fragments. The enrichment we observed is therefore compatible with any downstream genomic analysis, from locus-specific to wider genomic sequencing. This work shows that sWGS (and by extension, whole exome sequencing) can be performed on plasma DNA samples with low ctDNA content, and that this can facilitate the characterisation of mutations present in plasma at lower allele fractions and with lower sequencing depth. The compatibility of the cfDNA fragment size selection with wide-scale and sensitive genomic analysis could unlock the potential of liquid biopsies for the diagnosis of cancer at an earlier stage, and for the detection of minimal residual disease.

Methods:

Patients and sample preparation. 13 patients were recruited as part of prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference numbers 07/Q0106/63, 08/H0306/61 and 07/Q0106/63). Written informed consent was obtained from all patients and blood samples were collected before and after initiation of treatment with chemotherapeutic agents. DNA was extracted from 2 mL of plasma using the QIAamp circulating nucleic acid kit (Qiagen) according to the manufacturer's instructions.

Size selection. Between 8-10 ng of DNA were loaded into a 3% agarose cassette (HTC3010, Sage Bioscience) and size selection was performed on a PippinHT (Sage Bioscience) according to the manufacturer's protocol.

TAm-Seq. Tagged-Amplicon Sequencing libraries were prepared as previously described[23], using primers designed to assess single nucleotide variants (SNV) and small indels across selected hotspots and the entire coding regions of *TP53*. Libraries were sequenced using an HiSeq 4000 (Illumina).

sWGS. Indexed sequencing libraries were prepared using a commercially available kit (ThruPLEX-Plasma Seq, Rubicon Genomics). Libraries were pooled in equimolar amounts and sequenced on a HiSeq 4000 (Illumina) generating 150-bp paired-end reads. Sequence data was analysed using an in-house pipeline that consists of the following; Paired end sequence reads were aligned to the human reference genome (GRCh37) using BWA-mem following the removal of contaminating adapter sequences[33]. PCR and optical duplicates were marked using MarkDuplicates (Picard Tools) feature and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments.

SCNA analysis: Copy number analysis was performed in R using a modification of the QDNAseq pipeline[34], as follow: sequence reads were allocated into equally sized (50 kbp) non-overlapping bins throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and bins overlapping 'blacklisted' regions (ENCODE project) were excluded from downstream analysis. After median normalisation of the counts, bins were

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

segmented using both Circular Binary Segmentation and Locus-aware Circular Binary Segmentation algorithms, and an averaged log2R value per bin was calculated. The t-MAD score is calculated as the averaged absolute deviation from log2R = 0 after first trimming bin counts greater than 4 standard deviations from the mean count across all genomic regions.

References:

1. Siravegna, G., Marsoni, S., Siena, S. & Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* (2017). doi:10.1038/nrclinonc.2017.14
2. Wan, J. C. M. *et al.* Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* (2017). doi:10.1038/nrc.2017.7
3. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497,** 108–112 (2013).
4. Bettegowda, C. *et al.* Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci. Transl. Med.* **6,** 224ra24-224ra24 (2014).
5. Dawson, S.-J. *et al.* Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. *N. Engl. J. Med.* **368,** 1199–1209 (2013).
6. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci.* **102,** 16368–16373 (2005).
7. Taly, V. *et al.* Multiplex Picodroplet Digital PCR to Detect KRAS Mutations in Circulating DNA from the Plasma of Colorectal Cancer Patients. *Clin. Chem.* **59,** 1722–1731 (2013).
8. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34,** 547–555 (2016).
9. Khodakov, D., Wang, C. & Zhang, D. Y. Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Adv. Drug Deliv. Rev.* **105,** 3–19 (2016).
10. Mouliere, F. *et al.* High Fragmentation Characterizes Tumour-Derived Circulating DNA. *PLoS One* **6,** e23418 (2011).
11. Mouliere, F. & Rosenfeld, N. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proc. Natl. Acad. Sci.* **112,** 3178–3179 (2015).
12. Underhill, H. R. *et al.* Fragment Length of Circulating Tumor DNA. *PLOS Genet.* **12,** e1006162 (2016).
13. Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci.* **112,** E1317–E1325 (2015).
14. Jiang, P. & Lo, Y. M. D. The Long and Short of Circulating Cell-Free DNA and the Ins and Outs of Molecular Diagnostics. *Trends Genet.* **32,** 360–371 (2016).
15. Lo, Y. M. D. *et al.* Maternal Plasma DNA Sequencing Reveals the Genome-Wide Genetic and Mutational Profile of the Fetus. *Sci. Transl. Med.* **2,** 61ra91-61ra91 (2010).
16. Chandrananda, D. *et al.* High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genomics* **8,** 29 (2015).
17. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61,** 1659–65 (2001).

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

18. Yu, S. C. Y. *et al.* Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 8583–8 (2014).

19. Lun, F. M. F. *et al.* Noninvasive prenatal diagnosis of monogenic diseases by digital size selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 19920–5 (2008).

20. Minarik, G. *et al.* Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance. *PLoS One* **10,** e0144811 (2015).

21. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8,** 927–941 (2014).

22. Thierry, A. R. *et al.* Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res.* **38,** 6159–6175 (2010).

23. Forshew, T. *et al.* Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Sci. Transl. Med.* **4,** 136ra68-136ra68 (2012).

24. Parkinson, C. A. *et al.* Exploratory Analysis of TP53 Mutations in Circulating Tumour DNA as Biomarkers of Treatment Response for Patients with Relapsed High-Grade Serous Ovarian Carcinoma: A Retrospective Study. *PLOS Med.* **13,** e1002198 (2016).

25. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat. Commun.* **6,** 8760 (2015).

26. Heitzer, E. *et al.* Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **5,** 30 (2013).

27. Belic, J. *et al.* Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. *Clin. Chem.* **61,** 838–849 (2015).

28. Snyder, M. W. *et al.* Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164,** 57–68 (2016).

29. Ulz, P. *et al.* Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48,** 1273–1278 (2016).

30. Burnham, P. *et al.* Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6,** 27859 (2016).

31. Shema, E. *et al.* Single-molecule decoding of combinatorially modified nucleosomes. *Science (80-.).* **352,** (2016).

32. Vong, J. S. L. *et al.* Single-Stranded DNA Library Preparation Preferentially Enriches Short Maternal DNA in Maternal Plasma. *Clin. Chem.* (2017). doi:10.1373/clinchem.2016.268656

33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

34. Scheinin, I. *et al.* DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24,** 2022–2032 (2014).

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Notes:

Acknowledgements: The authors would like to thank all members of the Rosenfeld Lab and Brenton Lab for help and constructive discussion, in particular Francesco Marass, Wendy N. Cooper, Keval Patel, Jenny P.Y. Chan, Mareike Thompson, Lise Barlebo Ahlborn and Irena Hudecovà. The authors would like to also thank the Cancer Research UK Cambridge Institute core facilities for their support, in particular the genomics and biorepository facilities. We would like to acknowledge our patients and our caregivers, and the help and support of the research nurses, trial staff and the staff at Addenbrooke's Hospital. In particular, we would like to acknowledge Charlotte Hodgkin, Heather Biggs and Karen Hosking. We would like also to acknowledge the support of The University of Cambridge, Cancer Research UK (grant number A11906, A20240) and Hutchison Whampoa Limited. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 337905. This research is also supported by Target Ovarian Cancer and the Medical Research Council through their Joint Clinical Research Training Fellowship for Dr Moore. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: FM, AMP, DC, EM, JDB and NR conceptualised and designed the study; FM, AMP, EM, KH, CGS, JCMW, DG, RM, TG, AS, IG, CAP have performed experiments and collected data; DC has conceptualised and designed the t-MAD index and performed sWGS bioinformatics analysis; JM performed TAm-Seq bioinformatics analysis; RM and SR have designed the animal model; MJL performed histopathology revision; FM, AMP, DC, EM, JDB and NR have written the manuscript; all co-authors have critically reviewed the manuscript; FM, AMP, DC, JDB and NR supervised the study.

Author information: NR, JDB and DG are cofounders, shareholders and officers/consultants of Inivata Ltd, a cancer genomics company that commercialises ctDNA analysis. Inivata Ltd had no role in the conceptualisation, study design, data collection and analysis, decision to publish or preparation of the manuscript. NR and FM are co-inventors of patent applications that describe methods for the analysis of DNA fragments and applications of circulating tumour DNA. Other co-authors have declared no conflict of interests.

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 1:

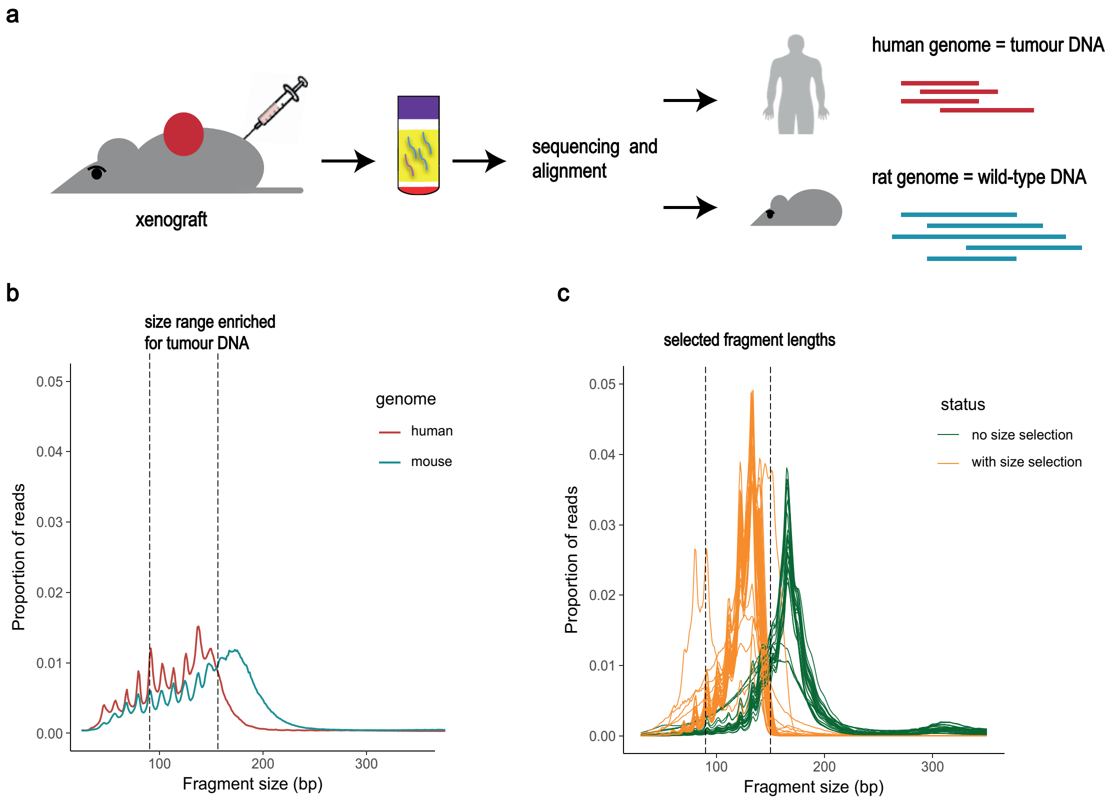


Figure 1: Plasma DNA originating from tumour and non-tumour cells have different sizes, enabling specific enrichment for ctDNA. a. Using an animal model with xenografted cells enabled the discrimination of DNA fragments released by the cancer cells (corresponding to the human DNA) from the DNA fragments released by the wild-type cells (corresponding to the rat DNA). b. Size distribution, assessed by sWGS, of DNA fragments from a plasma sample of a rat xenografted with a human glioblastoma tumour. c. Size distribution of DNA fragments from 26 plasma samples included in this study, assessed by sWGS. In green are the DNA fragments of the samples without size-selection, and in orange after size-selection. The two dotted vertical lines indicate the size selection range between 90 bp and 150 bp.

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2:

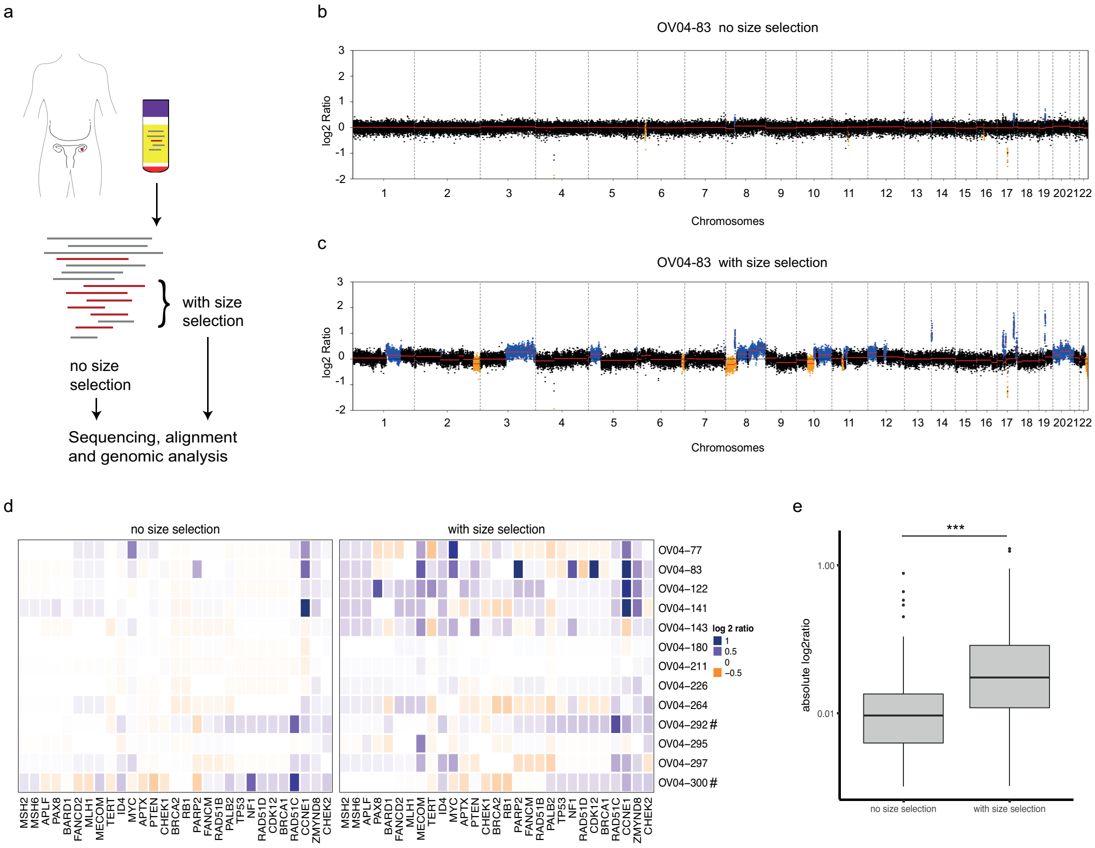


Figure 2: Recovery of short cfDNA fragments enriches for the representation of the cancer genome. a. Samples collected from 13 patients with HGSOC were analysed either with or without filtering by size selection. b. SCNA analysis based on the log2 ratios of regions along the genome of DNA extracted from a plasma sample collected during treatment for patient OV04-83. c. The same analysis of the same sample with size selection of fragments between 90 bp and 150 bp. Inferred amplifications are shown in blue and deletions in orange. d. SCNA analysis of the segmental log2ratio across a list of 29 genes frequently mutated in recurrent ovarian cancer, measured in plasma samples collected during treatment for all 13 patients, without size selection (left) and with size selection (right). The two samples which exhibited a degraded pattern of cfDNA fragmentation were OV04-292 and OV04-300 (both labelled by "#"). e. A comparison of the absolute level of log2ratio across the 29 genes of interest indicated a significant difference between the same samples without and with size selection ($p = 7.72.10^{-9}$). The 2 samples with the degraded pattern of cfDNA fragmentation have been excluded from this analysis.

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 3:

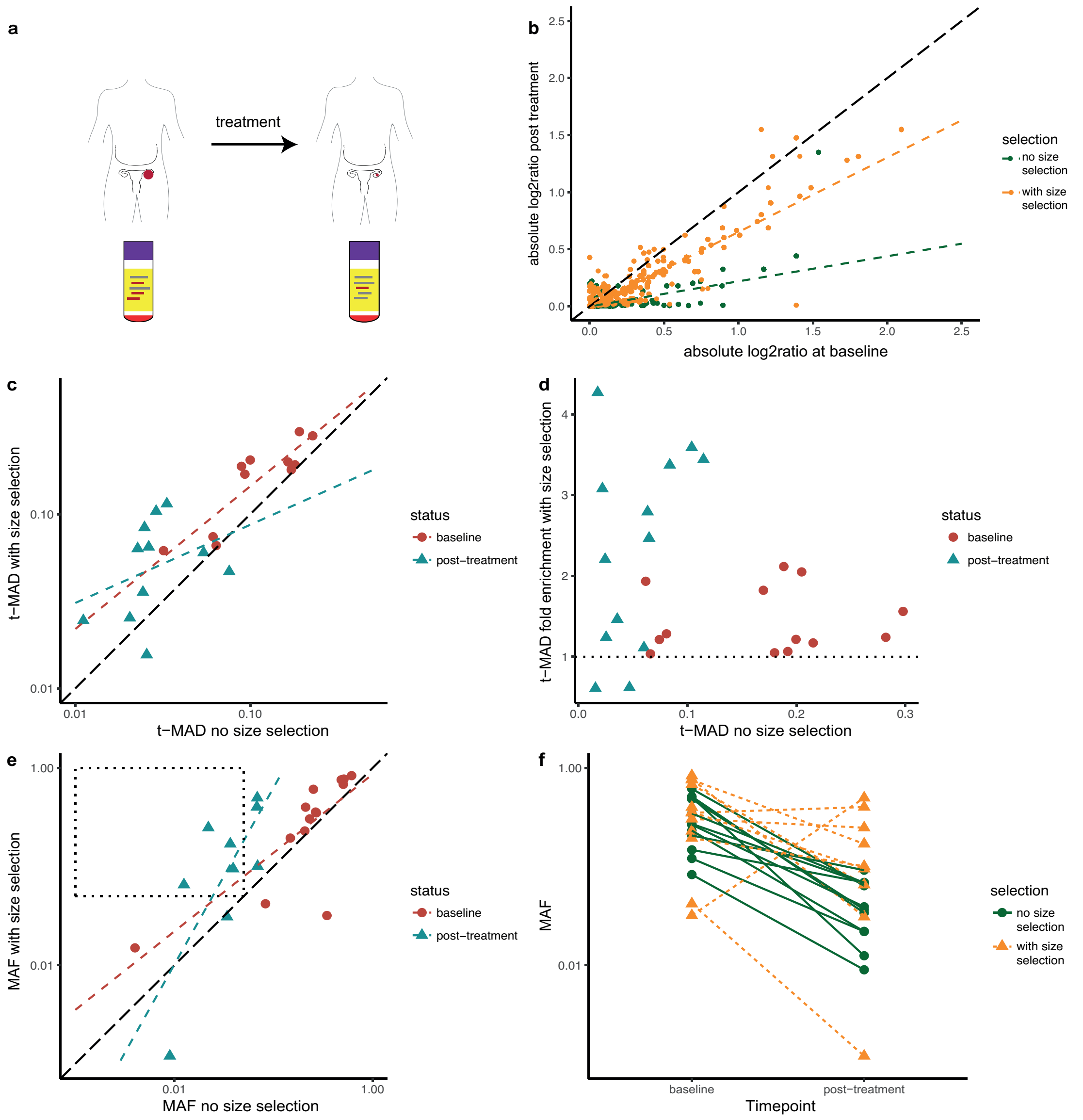


Figure 3: Analysis of the enrichment after size selection in 26 samples sequenced by sWGS and Tagged Amplicon Sequencing (TAm-Seq) revealed relative enrichment in tumour content. a. For each of 13 patients, we compared cfDNA from plasma samples collected before initiation of chemotherapy and 3 weeks or more after initiation of chemotherapy. Each of the 26 plasma samples was analysed with and without size selection. b. Comparison of the absolute value of the segmented log2ratio of the SCNAs called for the plasma samples of patient OV04-83 collected before and after initiation of the treatment. Data from the samples without size-selection is shown in green, and with size selection in orange. c. The t-MAD score determined from the sWGS with size selection (vertical) was higher than without size selection (horizontal) for most samples, including the samples collected at baseline (red circles) and after initiation of treatment (blue triangles). The 2 samples with no observed enrichment are OV04-292 and OV04-300. d. The enrichment factor with size selection,

bioRxiv preprint first posted online May. 5, 2017; doi: http://dx.doi.org/10.1101/134437. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

determined by t-MAD, varied per sample but was lower for was samples collected at baseline (red circles), which had high initial t-MAD score, compared to samples collected after treatment (blue triangle). e. The mutant allele fraction (MAF) determined by targeted sequencing with size selection (vertical) was higher than without size selection (horizontal) for most samples, including samples collected at baseline (red circles) and after initiation of treatment (blue triangles). The dotted area highlights samples with low MAF (<5%), where methods such as whole-exome sequencing (at sequencing depth of ~100x) would not be effective, where size selection enriched the mutant fraction to >5% and therefore accessible for wide-scale analysis. f. Comparison of the MAF detected by TAm-Seq before treatment and after initiation of treatment, as assessed by targeted sequencing, with size selection (yellow triangles) and without size selection (green circles).

CANCER

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim
to original U.S.
Government Works

Enhanced detection of circulating tumor DNA by fragment size analysis

Florent Mouliere[1,2]*†, Dineika Chandrananda[1,2]*, Anna M. Piskorz[1,2]*, Elizabeth K. Moore[1,2,3]*, James Morris[1,2], Lise Barlebo Ahlborn[4,5], Richard Mair[1,2,6], Teodora Goranova[1,2], Francesco Marass[1,2,7,8], Katrin Heider[1,2], Jonathan C. M. Wan[1,2], Anna Supernat[1,2,9], Irena Hudecova[1,2], Ioannis Gounaris[1,2,3], Susana Ros[1,2], Mercedes Jimenez-Linan[2,3], Javier Garcia-Corbacho[10], Keval Patel[1,2], Olga Østrup[5], Suzanne Murphy[1,2], Matthew D. Eldridge[1,2], Davina Gale[1,2], Grant D. Stewart[2,3,11], Johanna Burge[2,11], Wendy N. Cooper[1,2], Michiel S. van der Heijden[12,13], Charles E. Massie[1,2,14], Colin Watts[15], Pippa Corrie[3], Simon Pacey[3,14], Kevin M. Brindle[1,2,16], Richard D. Baird[17], Morten Mau-Sørensen[4], Christine A. Parkinson[1,2,3,18,19], Christopher G. Smith[1,2], James D. Brenton[1,2,3,18,19]‡§, Nitzan Rosenfeld[1,2]‡§

Existing methods to improve detection of circulating tumor DNA (ctDNA) have focused on genomic alterations but have rarely considered the biological properties of plasma cell-free DNA (cfDNA). We hypothesized that differences in fragment lengths of circulating DNA could be exploited to enhance sensitivity for detecting the presence of ctDNA and for noninvasive genomic analysis of cancer. We surveyed ctDNA fragment sizes in 344 plasma samples from 200 patients with cancer using low-pass whole-genome sequencing (0.4×). To establish the size distribution of mutant ctDNA, tumor-guided personalized deep sequencing was performed in 19 patients. We detected enrichment of ctDNA in fragment sizes between 90 and 150 bp and developed methods for in vitro and in silico size selection of these fragments. Selecting fragments between 90 and 150 bp improved detection of tumor DNA, with more than twofold median enrichment in >95% of cases and more than fourfold enrichment in >10% of cases. Analysis of size-selected cfDNA identified clinically actionable mutations and copy number alterations that were otherwise not detected. Identification of plasma samples from patients with advanced cancer was improved by predictive models integrating fragment length and copy number analysis of cfDNA, with area under the curve (AUC) >0.99 compared to AUC <0.80 without fragmentation features. Increased identification of cfDNA from patients with glioma, renal, and pancreatic cancer was achieved with AUC > 0.91 compared to AUC < 0.5 without fragmentation features. Fragment size analysis and selective sequencing of specific fragment sizes can boost ctDNA detection and could complement or provide an alternative to deeper sequencing of cfDNA.

INTRODUCTION

Blood plasma of patients with cancer contains circulating tumor DNA (ctDNA), but this valuable source of information is diluted by much larger quantities of DNA of noncancerous origins, such that ctDNA usually represents only a small fraction of the total cell-free DNA (cfDNA) (*1*, *2*). High-depth targeted sequencing of selected genomic regions can be used to detect low amounts of ctDNA, but broader analysis with methods such as whole-exome sequencing (WES) and shallow whole-genome sequencing (sWGS) is only generally informative when ctDNA content is ~10% or greater (*3*–*5*). The concentration of ctDNA can exceed 10% of the total cfDNA in patients with advanced-stage cancers (*6*–*8*), but is much lower in patients with low tumor burden (*9*–*12*) and in patients with some cancer types such as gliomas and renal cancers (*6*). Current strategies to improve ctDNA detection rely on increasing depth of sequencing coupled with various error correction methods (*2*, *13*, *14*). However, approaches that focus only on genomic alterations do not take advantage of the potential differences in chromatin organization or fragment sizes of ctDNA (*15*–*17*). Results of ever-deeper sequencing are also confounded by the likelihood of false-positive results from detection of mutations from noncancerous cells, clonal expansions in normal epithelia, or clonal hematopoiesis of indeterminate potential (CHIP) (*13*, *18*, *19*).

The cell of origin and the mechanism of cfDNA release into blood can mark cfDNA with specific fragmentation signatures, potentially

---

[1] Cancer Research UK Cambridge Institute, University of Cambridge, CB2 0RE Cambridge, UK. [2] Cancer Research UK Major Centre–Cambridge, Cancer Research UK Cambridge Institute, CB2 0RE Cambridge, UK. [3] Cambridge University Hospitals NHS Foundation Trust, CB2 0QQ Cambridge, UK. [4] Department of Oncology, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. [5] Centre for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark. [6] Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, CB2 0QQ Cambridge, UK. [7] Department of Biosystems Science and Engineering, ETH Zurich, 4058 Basel, Switzerland. [8] Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. [9] Department of Medical Biotechnology, Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, 80-211 Gdańsk, Poland. [10] Clinical Trials Unit, Clinic Institute of Haematological and Oncological Diseases, Hospital Clinic de Barcelona, 170 08036 Barcelona, Spain. [11] Academic Urology Group, Department of Surgery, University of Cambridge, CB2 0QQ Cambridge, UK. [12] Division of Molecular Carcinogenesis, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [13] Department of Medical Oncology, Netherlands Cancer Institute, 1066 CX Amsterdam, Netherlands. [14] Department of Oncology, University of Cambridge, CB2 0XZ Cambridge, UK. [15] Institute of Cancer Genomics Science, University of Birmingham, B15 2TT Birmingham, UK. [16] Department of Biochemistry, University of Cambridge, CB2 1QW Cambridge, UK. [17] Early Phase Clinical Trials and Breast Cancer Research Teams, Cancer Research UK Cambridge Centre, CB2 0QQ Cambridge, UK. [18] Department of Oncology, Hutchison/MRC Research Centre, University of Cambridge, CB2 0XZ Cambridge, UK. [19] NIHR Cambridge Biomedical Research Centre, CB2 0QQ Cambridge, UK.
*These authors contributed equally to this work as co-first authors.
†Present address: Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Pathology, Cancer Center Amsterdam, de Boelelaan 1117, 1081 HV Amsterdam, Netherlands.
‡These authors contributed equally to this work as co-senior authors.
§Corresponding author. Email: james.brenton@cruk.cam.ac.uk (J.D.B.); nitzan.rosenfeld@cruk.cam.ac.uk (N.R.)

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

providing precise information about cell type, gene expression, cell physiology or pathology, or action of treatment (*15*, *16*, *20*). cfDNA fragments commonly show a prominent mode at 167 bp, suggesting release from apoptotic caspase-dependent cleavage (Fig. 1A) (*21*–*24*). Circulating fetal DNA has been shown to be shorter than maternal DNA in plasma, and these size differences have been used to improve sensitivity of noninvasive prenatal diagnosis (*22*, *25*–*27*). The size distribution of tumor-derived cfDNA has only been investigated in a few studies, encompassing a small number of cancer types and patients, and showed conflicting results (*28*–*33*). A limitation of previous studies is that determining the specific sizes of tumor-derived DNA fragments requires detailed characterization of matched tumor-derived alterations (*30*, *33*), and the broader understanding and implications of potential biological differences have not previously been explored.

We hypothesized that we could improve the sensitivity for noninvasive cancer genomics by selective sequencing of ctDNA fragments and by leveraging differences in the biology that determine DNA fragmentation. To test this, we established a pan-cancer catalog of cfDNA fragmentation features in plasma samples from patients with different cancer types and healthy individuals to identify biological features enriched in tumor-derived DNA. We developed methods for selecting specific sizes of cfDNA fragments before sequencing and investigated the impact of combining cfDNA size selection with genome-wide sequencing to improve the detection of ctDNA and the identification of clinically actionable genomic alterations.

RESULTS

Surveying the fragmentation features of tumor cfDNA

We generated a catalog of cfDNA fragmentation features (Fig. 1A) in 344 plasma samples from 200 patients with 18 different cancer types and additional 65 plasma samples from healthy controls (Fig. 1B, fig. S1, and tables S1 and S2). The size distribution of cfDNA fragments in patients with cancer differed in the size ranges of 90 to 150 bp, 180 to 220 bp, and 250 to 320 bp compared to healthy individuals (Fig. 1B and fig. S2). cfDNA fragment sizes in plasma of healthy individuals and in plasma of patients with late-stage glioma, renal, pancreatic, and bladder cancers were significantly longer than in other late-stage cancer types including breast, ovarian, lung, melanoma, colorectal, and cholangiocarcinoma (Kruskal-Wallis, $P < 0.001$; Fig. 1C). Sorting the 18 cancer types according to the proportion of cfDNA fragments in the size range of 20 to 150 bp resulted in an order very similar to that obtained by Bettegowda *et al.* (*6*) based on the concentrations of ctDNA measured by individual mutation assays (Fig. 1D). In contrast to previous reports (*6*, *34*), this sorting was performed without any analysis or prior knowledge of the presence of mutations or somatic copy number alterations (SCNAs) yet allowed the investigation of ctDNA content in different cancers.

Sizing up mutant ctDNA

We determined the size profile of mutant ctDNA in plasma using two high-specificity approaches. First, we inferred the specific size profile of ctDNA and nontumor cfDNA with sWGS from the plasma of mice bearing human ovarian cancer xenografts (Fig. 2A). We observed a shift in ctDNA fragment sizes to less than 167 bp (Fig. 2B). Second, the size profile of mutant ctDNA was determined in plasma from 19 patients with cancer, using deep sequencing with patient-specific hybrid-capture panels developed from whole-exome profiling of matched tumor samples (Fig. 2C). By sequencing hundreds of mutations at a depth of >300× in cfDNA, we obtained allele-specific reads from mutant and normal DNA. Enrichment of DNA fragments carrying tumor-mutated alleles was observed in fragments between ~20 and 40 bp shorter than nucleosomal DNA sizes (multiples of 167 bp; Fig. 2D). We determined that mutant ctDNA is generally more fragmented than nonmutant cfDNA, with a maximum enrichment of ctDNA in fragments between 90 and 150 bp (fig. S3), as well as enrichment in the size range of 250 to 320 bp. These data also indicated that mutant DNA in plasma of patients with advanced cancer (before treatment) is consistently shorter than predicted mononuclesomal and dinucleosomal DNA fragment lengths (Fig. 2D).

Selecting tumor-derived DNA fragments

We evaluated whether the shorter cfDNA fragments in plasma can be harnessed to improve ctDNA detection. We determined the feasibility of selective sequencing of shorter fragments using in vitro size selection with a bench-top microfluidic device followed by sWGS in 48 plasma samples from 35 patients with high-grade serous ovarian cancer (HGSOC; Fig. 3A and figs. S4 and S5). We assessed the accuracy and quality of the size selection with the plasma from 20 healthy individuals (Fig. 3B and fig. S6). We also explored the utility of in silico size selection of fragmented DNA using read-pair positioning from unprocessed sWGS data (Fig. 3A). In silico size selection was performed once reads were aligned to the genome reference, by selecting the paired-end reads that corresponded to the fragment lengths in a 90- to 150-bp size range. Figure 3 (C to E) shows the effect of in vitro size selection for one HGSOC case (see all five samples in figs. S7 and S8). First, we identified SCNAs in plasma cfDNA before treatment, when the concentration of ctDNA was high (Fig. 3C). Only a small number of focal SCNAs were observed in the subsequent plasma sample collected 3 weeks after initiation of chemotherapy (without size selection; Fig. 3D). In vitro size selection of the same posttreatment plasma sample showed a median increase of 6.4× in the amplitude of detectable SCNAs without size selection. Selective sequencing of shorter fragments in this sample resulted in the detection of multiple other SCNAs that were not observed without size selection (Fig. 3E) and a genome-wide copy number profile that was similar to that obtained before treatment when ctDNA concentrations were four times higher, with additional copy number alterations identified in this sample despite the lower initial concentration of ctDNA (Fig. 3C). In silico size selection also enriched ctDNA but to a lower extent than using in vitro size selection (fig. S7). We concluded that selecting short DNA fragments in plasma can enrich tumor content on a genome-wide scale.

Quantifying the impact of size selection

To quantitatively assess the enrichment after size selection on a genome-wide scale, we developed a metric from sWGS data (<0.4× coverage) called t-MAD (trimmed median absolute deviation from copy number neutrality; see Fig. 4A). All sWGS data were down sampled to 10 million sequencing reads for comparison. To define the detection threshold, we measured the t-MAD score for sWGS data from 65 plasma samples from 46 healthy individuals and took the maximal value (median, 0.01; range, 0.004 to 0.015). We compared t-MAD to the mutant allele fraction (MAF) in high ctDNA cancer types as assessed by digital polymerase chain reaction (dPCR)

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

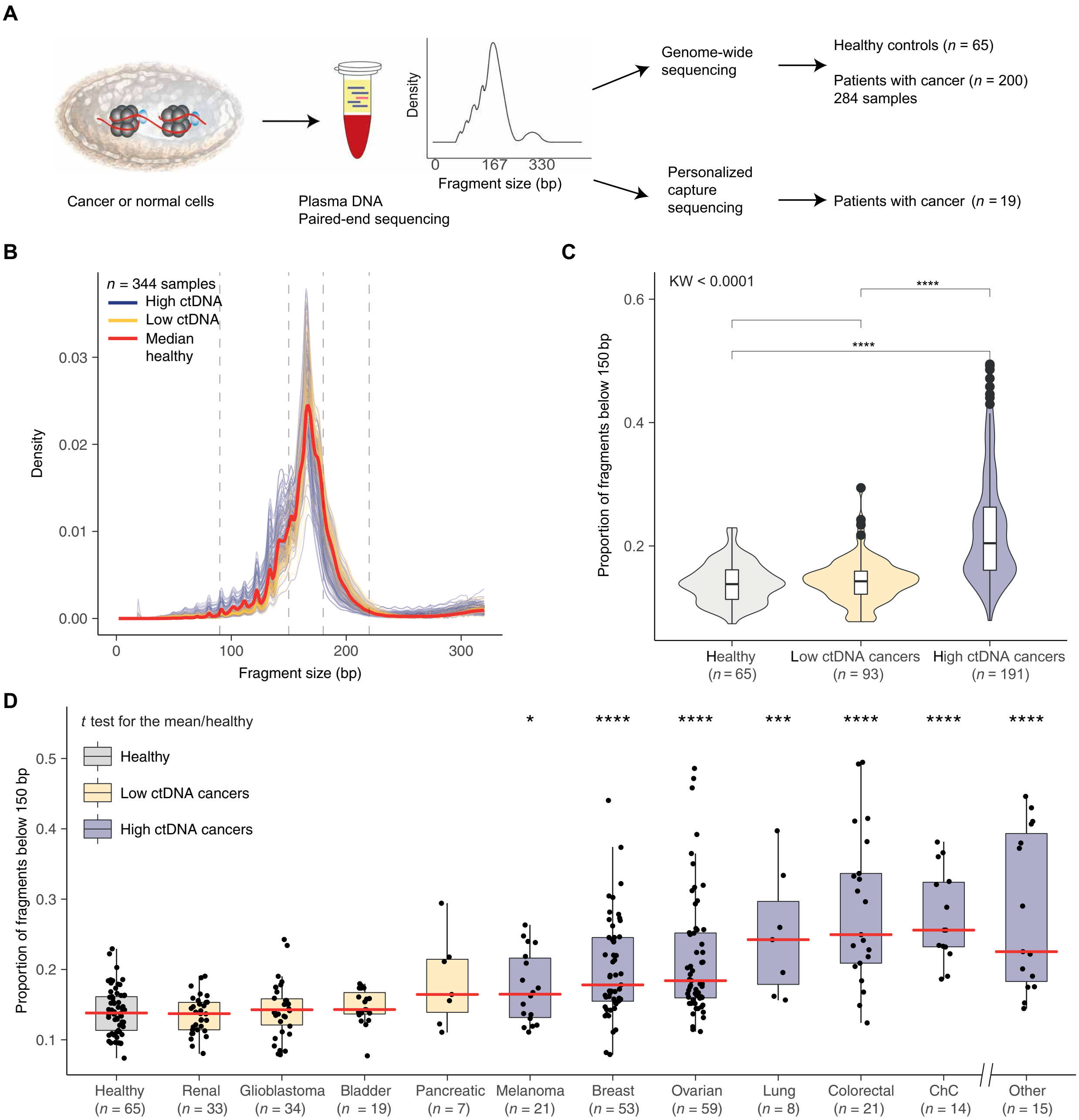


Fig. 1. Survey of plasma DNA fragmentation with genome-wide sequencing on a pan-cancer scale. (**A**) The size profile of cfDNA can be determined by paired-end sequencing of plasma samples and reflects its organization around the nucleosome. cfDNA is released into the blood circulation by various means, each of which leaves a signature on the DNA fragment sizes. We inferred the size profile of cfDNA by analyzing with sWGS ($n = 344$ plasma samples from 65 healthy controls and 200 patients with cancer) and the size profile of mutant ctDNA by personalized capture sequencing ($n = 19$ plasma samples). (**B**) Fragment size distributions of 344 plasma samples from 200 patients with cancer. Samples are split into two groups based on the previous literature (*6*), with orange representing samples from patients with cancer types previously observed to have low amounts of ctDNA (renal, bladder, pancreatic, and glioma) and blue representing samples from patients with cancer types previously observed to have higher amounts of ctDNA (breast, melanoma, ovarian, lung, colorectal, cholangiocarcinoma, and others; see table S1). (**C**) Proportion of cfDNA fragments below 150 bp in those samples, grouped into cancer types as defined in (B). The Kruskal-Wallis (KW) test for difference in size distributions indicated a significant difference between the group of samples from cancer types releasing high amounts of ctDNA and the group of samples from cancer types releasing low amounts, as well as the group of samples from healthy individuals). (**D**) Proportion of cfDNA fragments below 150 bp by cancer type (all samples). Cancer types represented by fewer than four individuals are grouped in the "other" category. Red lines indicate the median proportion for each cancer type. ChC, cholangiocarcinoma. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

or WES in 97 samples. We observed a high correlation (Pearson correlation, $r = 0.80$) between t-MAD and MAF (Fig. 4B) for samples with t-MAD greater than the detection threshold (0.015) or with MAF > 0.025. Figure S9 shows that the slope of t-MAD versus MAF fit lines differed between cancer types (range, 0.17 to 1.12), likely reflecting differences in the extent of SCNAs. We estimated the

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

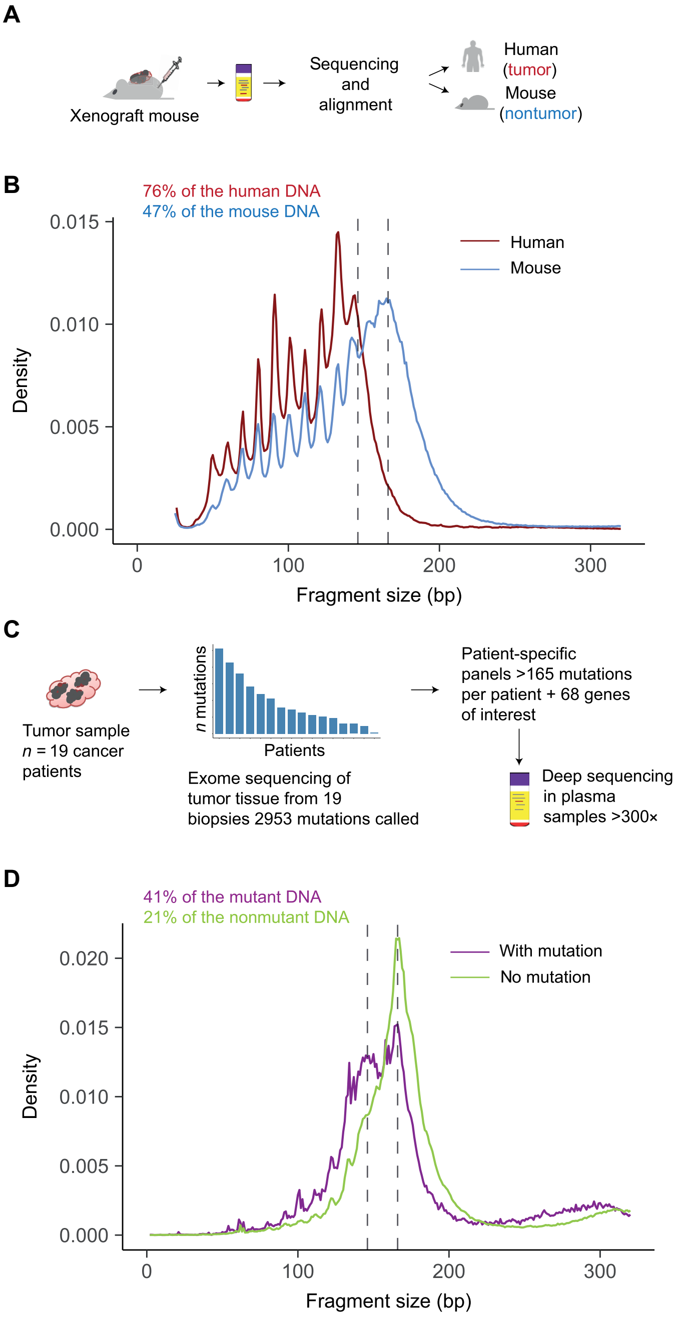


Fig. 2. Determining the size profile of mutant ctDNA with animal models and personalized capture sequencing. (**A**) A mouse model with xenografted human tumor cells enabled the discrimination of DNA fragments released by cancer cells (reads aligning to the human genome) from the DNA released by healthy cells (reads aligning to the mouse genome), with the use of sWGS. (**B**) Fragment size distribution from the plasma extracted from a mouse xenografted with a human ovarian tumor, showing ctDNA originating from tumor cells (red) and cfDNA from noncancerous cells (blue). Two vertical dashed lines indicate 145 and 167 bp. The fraction of reads shorter than 150 bp is indicated. (**C**) Design of personalized hybrid-capture sequencing panels developed to specifically determine the size profiles of mutant DNA and nonmutant DNA in plasma from 19 patients with late-stage cancers. Capture panels included somatic mutations identified in tumor tissue by WES. A mean of 165 mutations per patient was then analyzed from matched plasma samples. Reads were aligned and separated into fragments carrying either the reference or the mutant sequence. Fragment sizes for paired-end reads were calculated. (**D**) Size profiles of mutant DNA and nonmutant DNA in plasma from 19 patients with late-stage cancers were determined by tumor-guided capture sequencing. The fraction of reads shorter than 150 bp is indicated.

sensitivity of t-MAD for detecting low amounts of ctDNA using a spike-in dilution of DNA from a patient with a *TP53* mutation into DNA from a pool of seven healthy individuals (fig. S10), which confirmed that the t-MAD score was linear with ctDNA fraction down to MAF of ~0.01. In addition, t-MAD scores greater than the detection threshold (0.015) for samples were present even in samples with MAF as low as 0.004. t-MAD was also strongly correlated with tumor volume determined by RECIST1.1 (Pearson correlation, $r = 0.6$; $P < 0.0001$; $n = 35$; fig. S11).

Using t-MAD, we detected ctDNA from 69% (130 of 189) of the samples from cancer types where ctDNA concentrations were shown to be high (Fig. 4C). From cancer types for which ctDNA concentrations are suspected to be low (glioma, renal, bladder, and pancreatic), we detected ctDNA in 17% (10 of 57) of the cases (Fig. 4C). We used in silico size selection of the DNA fragments between 90 and 150 bp from the high ctDNA cancers ($n = 189$) and healthy controls ($n = 65$) to improve the sensitivity for detecting t-MAD (Fig. 4D). Receiver operating characteristic (ROC) analysis comparing the t-MAD score for the samples revealed an area under the curve (AUC) of 0.90 after in silico size selection, against an AUC of 0.69 without size selection (Fig. 4D).

We explored whether size-selected sequencing could improve the detection of response or disease progression. We used sWGS of longitudinal plasma samples from six patients with cancer (Fig. 4, E and F) and in silico size selection of the cfDNA fragments between 90 and 150 bp. In two patients, size-selected samples indicated tumor progression 60 and 87 days before detection by imaging or unselected t-MAD analysis (Fig. 4, E and F). Other longitudinal samples exhibited improvements in the detection of ctDNA with t-MAD and size selection (Fig. 4F).

Identifying more clinically relevant genomic alterations with size selection

We next tested whether size selection could increase the sensitivity for detecting cancer genomic alterations in cfDNA. To test effects on copy number aberrations, we studied 35 patients with HGSOC as the archetypal copy number–driven cancer (*35*). t-MAD was used to quantify the enrichment of ctDNA with in vitro size selection in 48 plasma samples, including samples collected before and after initiation of chemotherapy treatment. In vitro size selection resulted in an increase in the calculated t-MAD score from the sWGS data for 47 of 48 of the plasma samples (98%; *t* test, $P = 0.06$) with a mean of 2.5 and median of 2.1-fold increase (Fig. 5A and table S3). We compared the t-MAD scores against those obtained by sWGS for the plasma samples from healthy individuals. Thirty-nine of the 48 size-selected HGSOC plasma samples (82%) had a t-MAD score greater than the highest t-MAD value determined in the in vitro size-selected healthy plasma samples (Fig. 5A and figs. S6 and S12), compared to 24 of 48 without size selection (50%). ROC analysis comparing the t-MAD score for the samples from patients with cancer (pre- and posttreatment initiation, $n = 48$) and healthy controls ($n = 46$) revealed an AUC of 0.97 after in vitro size selection, with maximal sensitivity and specificity of 90 and 98%, respectively.

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

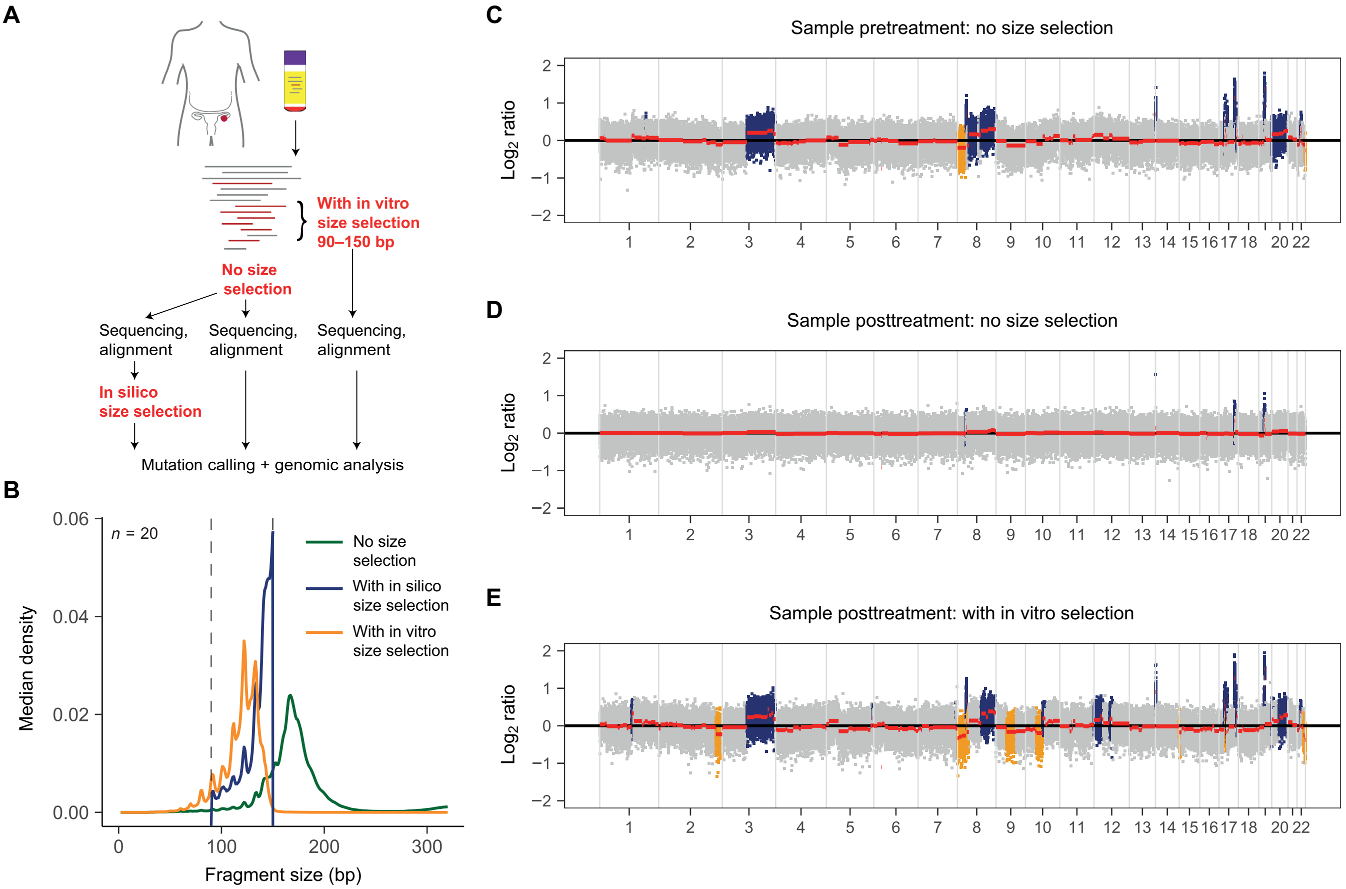


Fig. 3. Enhancing the tumor fraction from plasma sequencing with size selection. (**A**) Plasma samples collected from patients with ovarian cancer were analyzed in parallel without size selection or using either in silico or in vitro size selection. (**B**) Accuracy of the in vitro and in silico size selection determined on a cohort of 20 healthy controls. The size distribution before size selection is shown in green, after in silico size selection (with sharp cutoff at 90 and 150 bp) in blue and after in vitro size selection in orange. Vertical lines indicate 90 and 150 bp. (**C**) SCNA analysis with sWGS from plasma DNA of a patient with ovarian cancer collected before initiation of treatment, when ctDNA MAF was 0.271 for a *TP53* mutation as determined by tagged-amplicon deep sequencing (TAm-Seq). Inferred amplifications are shown in blue and deletions in orange. Copy number neutral regions are shown in gray. (**D**) SCNA analysis of a plasma sample from the same patient as in (C), collected 3 weeks after treatment start. The MAF for the *TP53* mutation at this time point was 0.068, and sWGS revealed only limited evidence of copy number alterations (before size selection). (**E**) Analysis of the same plasma sample as in (D) after in vitro size selection of fragments between 90 and 150 bp in length. The MAF for the *TP53* mutation increased to 0.402 after in vitro size selection, and SCNAs were apparent by sWGS. More SCNAs were detected in comparison to (C) and (D) (for example, in chr2, chr9, and chr10). SCNAs were also detected in this sample after in silico size selection (fig. S7).

This was superior to detection by sWGS without size selection (AUC, 0.64; Fig. 5B).

We then determined whether this improved sensitivity resulted in the detection of SCNAs with potential clinical value. Across the genome, t-MAD scores evaluating SCNAs were higher after size selection in 33 of 35 (94%) patients with HGSOC, and the magnitude of copy number (log$_2$ ratio) values significantly increased after in vitro size selection (t test for the means, $P = 0.003$; Fig. 5C). We compared the relative copy number values for 15 genes frequently altered in HGSOC (table S4). Analysis of plasma cfDNA after size selection revealed a large number of SCNAs that were not observed in the same samples without size selection (Fig. 5D), including amplifications in key genes such as *NF1*, *TERT*, and *MYC* (fig. S13).

We also tested whether similar enrichment was seen for substitutions to exclude the possibility that size selection might only increase the sensitivity for sWGS analysis. We performed WES of plasma cfDNA from 23 patients with seven cancer types (fig. S1). We used the WES data to compare the size distributions of fragments carrying mutant or nonmutant alleles (Fig. 6A) and to test whether size selection could identify additional mutations. We first selected six patients with HGSOC and performed WES of plasma DNA with and without in vitro size selection in the range of 90 to 150 bp, analyzing time points before and after initiation of treatment (*36*). In addition, in silico size selection for the same range of fragment sizes was performed (Fig. 6A). Analysis of the MAF of SNVs revealed statistically significant enrichment of the tumor fraction with both in vitro size selection (mean, 4.19-fold; median, 4.27-fold increase; t test, $P < 0.001$) and in silico size selection (mean, 2.20-fold; median, 2.25-fold increase; t test, $P < 0.001$; Fig. 6A and fig. S14). Three weeks after initiation of treatment, ctDNA fractions are often lower (*36*), and therefore, we further analyzed posttreatment plasma samples using TAm-Seq (*37*). We observed enrichment of MAFs by in vitro size selection between 0.9 and 11 times (mean, 2.1 times; median, 1.5 times), with one outlier sample exhibiting a relative enrichment of 118 times compared to the same samples without size selection (fig. S15).

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

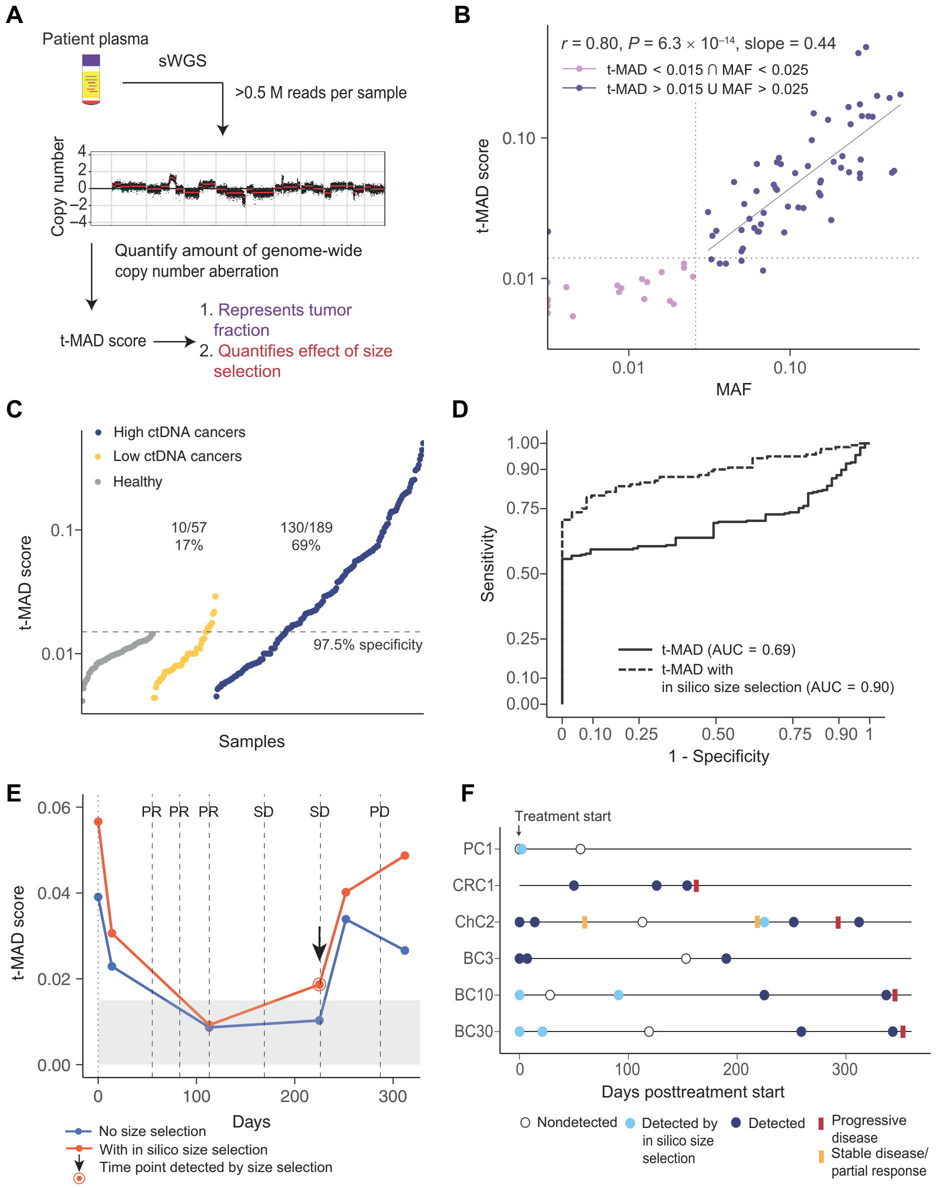


Fig. 4. Quantifying the ctDNA enrichment by sWGS with in silico size selection and t-MAD. (**A**) Workflow to quantify tumor fraction from SCNA as a genome-wide score named t-MAD. (**B**) Correlation between the MAF of single-nucleotide variants (SNVs) determined by dPCR or hybrid-capture sequencing and t-MAD score determined by sWGS. Data included 97 samples from patients with multiple cancer types with matched MAF measurements and t-MAD scores. Pearson correlation (coefficient *r*) between MAF and t-MAD scores was calculated for all cases with MAF > 0.025 and t-MAD > 0.015. Linear regression indicated a fit with a slope of 0.44 (purple solid line). (**C**) Comparison of t-MAD scores determined from sWGS between healthy samples and samples collected from patients with cancer types that exhibit low amounts of ctDNA and from patients with cancer types that exhibit high amounts of ctDNA (as in Fig. 1). All samples for which t-MAD could be calculated have been included. (**D**) ROC analysis comparing the classification of these plasma samples from high ctDNA cancer samples ($n = 189$) and plasma samples from healthy controls ($n = 65$) using t-MAD had an AUC of 0.69 without size selection (black solid curve). After applying in silico size selection to the samples from patients with cancer, we observed an AUC of 0.90 (black dashed curve). (**E**) Determination of t-MAD from longitudinal plasma samples of a patient with colorectal cancer. t-MAD was analyzed before and after in silico size selection of the DNA fragments between 90 and 150 bp and then compared to the RECIST status for this patient. PR, partial response; SD, stable disease; PD, progressive disease. (**F**) Application of in silico size selection to six patients with long-term follow-up. t-MAD score was determined before and after in silico size selection of the short DNA fragments. Dark blue circles indicate samples in which ctDNA was detected both with and without in silico size selection. Light blue circles indicate samples where ctDNA was detected only after in silico size selection. Open circles indicate samples where ctDNA was not detected by either analysis. Times when RECIST status was assessed are indicated by a red bar for progression or an orange bar for regression or stable disease. PC, prostate cancer; CRC, colorectal cancer; ChC, cholangiocarcinoma; BC, breast cancer. The numbers correspond to the patients.

Size selection with both in vitro and in silico methods increased the number of mutations detected by WES by an average of 53% compared to no size selection (Fig. 6B). We identified a total of 1023 mutations in the samples without size selection. An additional 260 mutations were detected by in vitro size selection, and an additional 310 mutations were called after in silico size selection (Fig. 6B and table S5). To exclude the possibility that the improved sensitivity for mutation detection was a result of sequencing artifacts, we validated whether new mutations were also detectable in tumor specimens. We used in silico size selection in an independent cohort of 16 patients for whom matched tumor tissue DNA was available (table S6). In silico size selection enriched the MAF for nearly all mutations (2061 of 2133, 97%), with an average increase of MAF of 1.7× (Fig. 6C). For 13 of 16 patients (81%), we identified additional mutations in plasma after in silico size selection. Of these 82 additional mutations, 23 (28%) were confirmed to be present in the matched tumor tissue DNA (Fig. 6D). This included mutations in key cancer genes including *BRAF*, *ARID1A*, and *NF1* (fig. S16).

Detecting cancer by supervised machine learning combining cfDNA fragmentation and somatic alteration analysis

Although in vitro and in silico size selection increase the sensitivity of detection, they also result in a loss of cfDNA for analysis. In analysis of ctDNA based on genomic signals, potentially informative data are lost because regions of the cancer genome that are not mutated or altered do not contribute to detection (fig. S17). We hypothesized that leveraging other biological properties of the cfDNA fragmentation profile could enhance the detection of ctDNA.

We defined other cfDNA fragmentation features from sWGS data including (i) the proportion of fragments in multiple size ranges, (ii) the ratios of proportions of fragments in different sizes, and (iii) the amplitude of oscillations in fragment size density with 10-bp periodicity (see Materials and Methods and Fig. 7A). These fragmentation features were compared between patients with cancer and healthy individuals (fig. S18),

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

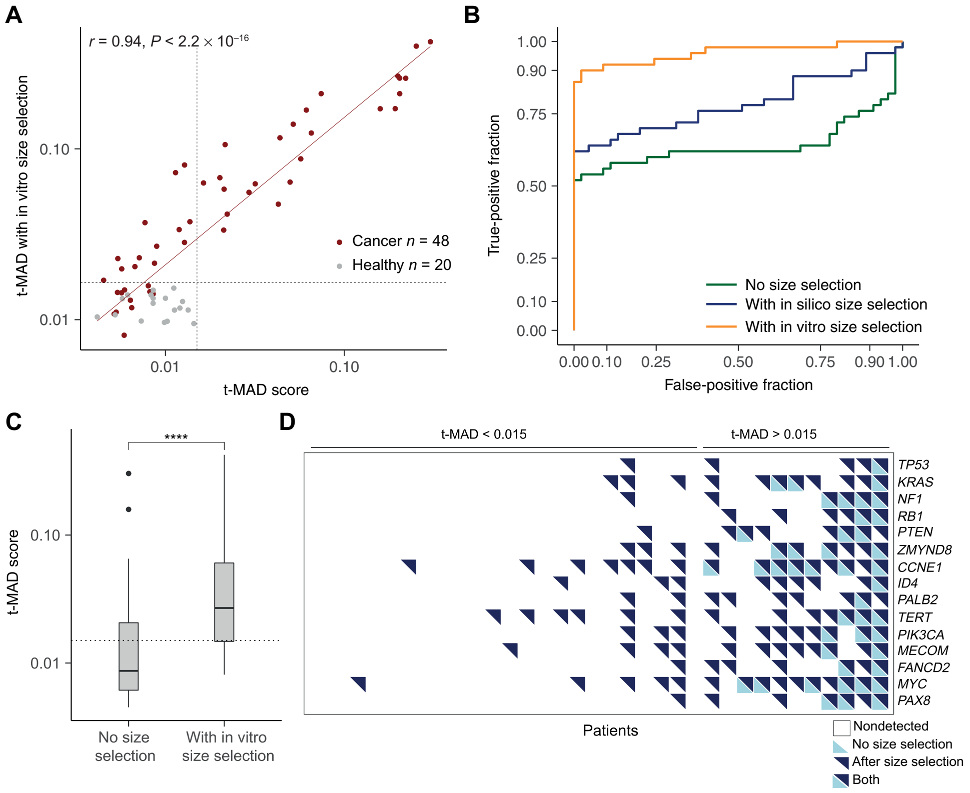


Fig. 5. Quantifying the ctDNA enrichment by sWGS with in vitro size selection. (**A**) The effect of in vitro size selection on the t-MAD score. For each of 48 plasma samples collected from 35 patients, the t-MAD score was determined from the sWGS after in vitro size selection (*y* axis) and without size selection (*x* axis). In vitro size selection increased the t-MAD score for nearly all samples, with a median increase of 2.1-fold (range from 1.1- to 6.4-fold). t-MAD scores determined from sWGS for 46 samples from healthy individuals were all <0.015 both before and after in vitro size selection. (**B**) ROC analysis comparing the classification of plasma samples from patients with cancer ($n = 48$) and plasma samples from healthy controls ($n = 46$) using t-MAD had an AUC of 0.64 without size selection (green curve). After applying in silico size selection to the samples from the patients and controls, we observed an AUC of 0.78 (blue curve), and after in vitro size selection, an AUC of 0.97 (orange curve). (**C**) Comparison of t-MAD scores determined from sWGS between matched ovarian cancer samples with and without in vitro size selection. The *t* test for the difference in means indicates a significant increase in tumor fraction (measured by t-MAD) with in vitro size selection (****$P < 0.0001$). (**D**) Detection of SCNAs across 15 genes frequently mutated in recurrent ovarian cancer, measured in plasma samples collected during treatment for 35 patients. Patients were ranked from left to right by increasing tumor fraction as quantified by t-MAD (before in vitro size selection). SCNAs were labeled as detected for a gene if the mean $\log_2$ ratio in that region was greater than 0.05. Empty squares represent copy number neutral regions, bottom left triangles in light blue indicate that SCNAs were detected without size selection, and top right triangles in dark blue represent SCNAs detected after in vitro size selection.

and the feature representing the proportion (P) of fragments between 20 and 150 bp exhibited the highest AUC (0.819). Principal components analysis (PCA) of the samples represented by t-MAD and fragmentation features showed a separation between healthy samples and samples from patients with cancer and identified fragment features that were aligned (in PCA) with t-MAD scores (Fig. 7B).

We next explored the potential of fragmentation features to enhance the detection of tumor DNA in plasma samples. A predictive analysis was performed using the t-MAD score and nine fragmentation features across 304 samples (239 from patients with cancer and 65 from healthy controls; Fig. 7C, fig. S19, and table S2). The nine fragmentation features determined from sWGS included five features based on the proportion (P) of fragments in defined size ranges: P(20 to 150), P(100 to 150), P(160 to 180), P(180 to 220), and P(250 to 320); three features based on ratios of those proportions: P(20 to 150)/P(160 to 180), P(100 to 150)/P(163 to 169), and P(20 to 150)/P(180 to 220); and a further feature based on the amplitude of the oscillations having 10-bp periodicity observed below 150 bp.

Variable selection and the classification of samples as "healthy" or "cancer" were performed using logistic regression (LR) and random forest (RF) models trained on 153 samples and validated on two datasets of 94 and 83 independent samples (Fig. 7C). The best feature set for the LR model included t-MAD, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320). The same five variables were independently identified using the RF model (with some differences in their ranking). Figure S20 shows performance metrics for the different algorithms on training set data using cross-validation. Using t-MAD alone in the validation pan-cancer dataset (Fig. 7D and fig. S19), we could distinguish cancer samples from healthy individuals with an AUC of 0.764. Using the LR model improved the classification of the samples to an AUC of 0.908. The RF model (trained on the 153-sample training set) could distinguish cancer from healthy individuals even more accurately in the validation dataset ($n = 94$) with an AUC of 0.994. On the second validation dataset containing low-ctDNA cancer samples ($n = 83$; Fig. 7E), t-MAD alone or the LR performed less well, with AUC values of 0.421 and 0.532, respectively. However, the RF model was still able to distinguish low-ctDNA cancer samples from healthy controls with an AUC of 0.914. At a specificity of 95%, the RF model correctly classified as cancer in 64 of 68 (94%) of the samples from high-ctDNA cancers (colorectal, cholangiocarcinoma, ovarian, breast, and melanoma) and 37 of 57 (65%) of the samples from low-ctDNA cancers (pancreatic, renal, and glioma; Fig. 7F). In a second iteration of model training, we omitted t-MAD using only the four fragmentation features (fig. S21). The RF model could still distinguish cancer from healthy controls, albeit with slightly reduced AUCs (0.989 for cancer types with high amounts of ctDNA and 0.891 for cancer types with low amounts of ctDNA), suggesting that the cfDNA fragmentation pattern is the most important predictive component.

DISCUSSION

Our results indicate that exploiting fundamental properties of cfDNA with fragment-specific analyses can allow more sensitive evaluation

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

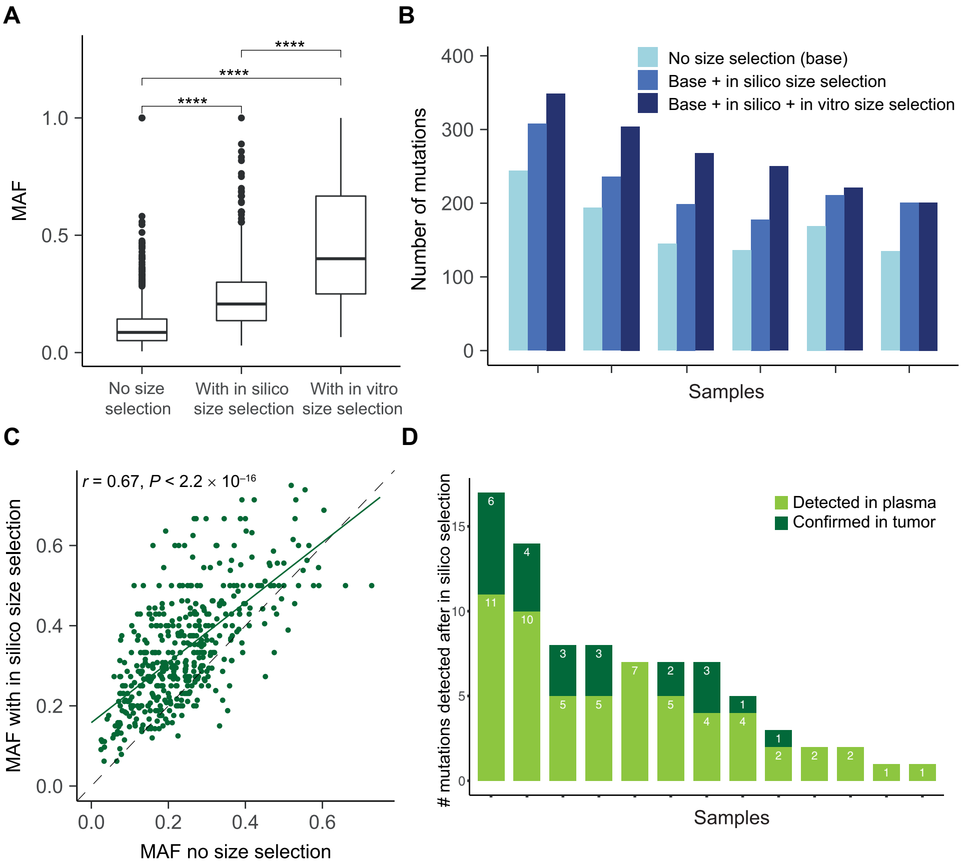


Fig. 6. Improving the detection of somatic alterations by WES in multiple cancer types with size selection. (**A**) Analysis of the MAF of mutations detected by WES in six patients with HGSOC without size selection and with either in vitro or in silico size selection. ****$P < 0.0001$. (**B**) Comparison of size-selected WES data with nonselected WES data to assess the number of mutations detected in plasma samples from six patients with HGSOC. For each patient, the first bar in light blue shows the number of mutations called without size selection, the second bar quantifies the number of mutations called after the addition of those identified with in silico size selection, and the third bar in dark blue shows the number of mutations called after addition of mutations called after in vitro size selection. (**C**) Patients ($n = 16$) were retrospectively selected from a cohort with different cancer types (colorectal, cholangiocarcinoma, pancreatic, and prostate) enrolled in early-phase clinical trials. Matched tumor tissue DNA was available for each plasma sample, and two patients also had a biopsy collected at relapse. WES was performed on tumor tissue DNA and plasma DNA samples, and in silico size selection was applied to the data. A total of 97% (2061 of 2133) of the shared mutations detected by WES showed higher MAF after in silico size selection. (**D**) Mutations detected only after in silico selection of WES data from 16 patients [as in (C)] compared to mutations called by WES of the matched tumor tissue. Three of 16 patients had no additional mutations identified after in silico size selection. Of the 82 mutations detected in plasma after in silico size selection, 23 (28%) had low signal in tumor WES data and were not identified in those samples without size selection.

of ctDNA. We based the fragment size selection criteria on a biological observation that ctDNA fragment size distribution is shifted from noncancerous cfDNA. Our work builds on a comprehensive survey of plasma cfDNA fragmentation patterns across 200 patients with multiple cancer types and 65 healthy individuals. We identified features that could determine the presence and amount of ctDNA in plasma samples, without a priori knowledge of somatic aberrations. We caution that this catalog is limited to double-stranded DNA from plasma samples and is subject to potential biases incurred by the DNA extraction and sequencing methods we used. Additional biological effects could contribute to further selective analysis of cfDNA. Other bodily fluids (urine, cerebrospinal fluid, and saliva), different nucleic acids and structures, altered mechanisms of release into circulation, or sample processing methods could exhibit varying fragment size signatures and could offer additional exploitable biological patterns for selective sequencing.

Previous work has reported the size distributions of mutant ctDNA but only considered limited genomic loci, cancer types, or cases (*30*, *32*, *33*). We identified the size differences between mutant and nonmutant DNA on a genome-wide and pan-cancer scale. We developed a method to size mutant ctDNA without using high-depth WGS. By sequencing >150 mutations per patient at high depth, we obtained large numbers of reads that could be unequivocally identified as tumor derived and thus determined the size distribution of mutant ctDNA and nonmutant cfDNA in patients with cancer. A potential limitation of our approach is that capture-based sequencing is biased by probe capture efficiency and, therefore, our data may not accurately reflect ctDNA fragments of <100 or >300 bp.

Our work provides strong evidence that the modal size of ctDNA for many cancer types is less than 167 bp, which is the length of DNA wrapped around the chromatosome. In addition, our work also shows that there is enrichment of mutant DNA fragments at sizes greater than 167 bp, notably in the range of 250 to 320 bp. These longer fragments may explain previous observations that longer ctDNA can be detected in the plasma of patients with cancer (*29*, *32*). The origin of these long fragments is still unknown, and their observation could be linked to technical factors. However, it is likely that mechanisms of compaction and release of cfDNA into circulation, which may differ depending on its origin, will be reflected by different fragment sizes (*38*). Improving the characterization of these fragments will be important, especially for future work combining analysis of ctDNA with that of other entities in blood such as microvesicles and tumor-educated platelets (*39*, *40*). Fragment-specific analyses not only increase the sensitivity for detection of rare mutations but could also be used to track modifications in the size distribution of ctDNA. Future work should address whether this approach could be used to elucidate mechanistic effects of treatment on tumor cells, for example, by distinguishing between necrosis and apoptosis based on fragment size (*41*).

Genome-wide and exome sequencing of plasma DNA at multiple time points during cancer treatment have been proposed as noninvasive means to study cancer evolution and for the identification of possible mechanisms of resistance to treatment (*3*). However, WGS and WES

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

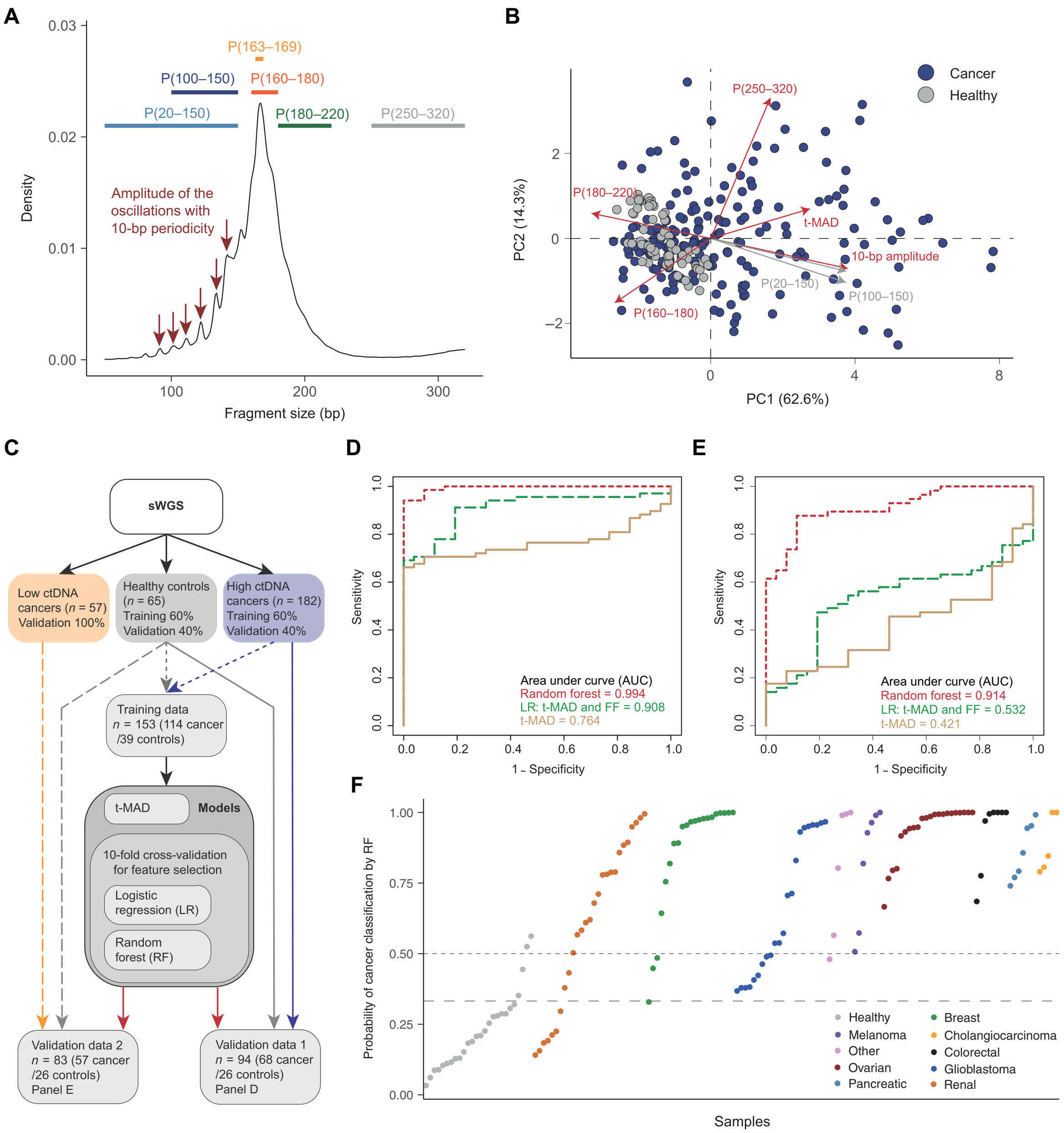


Fig. 7. Enhancing the potential for ctDNA detection by combining SCNAs and fragment size features. (**A**) Schematic illustrating the selection of different size ranges and features in the distribution of fragment sizes. For each sample, fragmentation features included the proportion (P) of fragments in specific size ranges, the ratio between certain ranges, and a quantification of the amplitude of the 10-bp oscillations in the 90- to 145-bp size range calculated from the periodic "peaks" and "valleys." (**B**) PCA comparing cancer and healthy samples using data from t-MAD scores and the fragmentation features. Red arrows indicate features that were selected as informative by the predictive analysis. (**C**) Workflow for the predictive analysis combining SCNAs and fragment size features. sWGS data from 182 plasma samples from patients with cancer types with high amounts of ctDNA (colorectal, cholangiocarcinoma, lung, ovarian, and breast) were split into a training set (60% of samples) and a validation set (validation data 1, together with the healthy individual validation set). A further dataset of sWGS from 57 samples of cancer types exhibiting low amounts of ctDNA (glioma, renal, and pancreatic) was used as validation data 2, together with the healthy individual validation set. Plasma DNA sWGS data from healthy controls were split into a training set (60% of samples) and a validation set (used in both validation data 1 and validation data 2). (**D**) ROC curves for validation data 1 (samples from patients with cancer with high ctDNA amounts, 68; healthy, 26) for three predictive models built on the pan-cancer training cohort (cancer, 114; healthy, 39). The beige curve represents the ROC curve for classification with t-MAD only, the long-dashed green line represents the LR model combining the top five features based on recursive feature elimination [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed line shows the result for a RF classifier trained on the combination of the same five features, independently chosen for the best RF predictive model. FF, fragment size features. (**E**) ROC curves for validation data 2 (samples from patients with cancer with low ctDNA amounts, 57; healthy, 26) for the same three classifiers as in (D). The beige curve represents the model using t-MAD only, the long-dashed green curve represents the LR model combining the top five features [t-MAD score, 10-bp amplitude, P(160 to 180), P(180 to 220), and P(250 to 320)], and the red dashed curve shows the result for a RF classifier trained on the combination of same five predictive features. (**F**) Plot representing the probability of classification as cancer with the RF model for all samples in both validation datasets. Samples are separated by cancer type and sorted within each by the RF probability of classification as cancer. The horizontal dashed line indicates 50% probability (achieving specificity of 24 of 26, 92.3%), and the long-dashed line indicates 33% probability (achieving specificity of 22 of 26, 84.6%).

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

approaches are costly and have thus far been applicable only in samples for which the tumor DNA fraction was >5 to 10% (*3*–*5*, *42*). We demonstrated that we could exploit the differences in fragment lengths using in vitro and in silico size selection to enrich for tumor content in plasma samples, which improved mutation and SCNA detection in sWGS and WES data. We demonstrated that size selection improved the detection of mutations that are present in plasma at low allelic fractions while maintaining low sequencing depth by sWGS and WES. Size selection can be achieved with simple means and at low cost and is compatible with a wide range of downstream genome-wide and targeted genomic analyses, greatly increasing the potential value and utility of liquid biopsies as well as the cost-effectiveness of cfDNA sequencing.

Size selection can be applied in silico, which incurs no added costs, or in vitro, which adds a simple and low-cost intermediate step that can be applied to either the extracted DNA or the libraries created from it. This approach, applied prospectively to new studies, could boost the clinical utility of ctDNA detection and analysis and creates an opportunity for reanalysis of large volumes of existing data (*4*, *34*, *43*). The limitation of this technique is a potential loss of material and information, because some of the informative fragments may be found in size ranges that are filtered out or deprioritized in the analysis. This may be particularly problematic if only a few copies of the fragments of interest are present in the plasma. Despite potential loss of material, we demonstrated that classification algorithms can learn from cfDNA fragmentation features and SCNA analysis and improve the detection of ctDNA with a cheap sequencing approach. Moreover, the cfDNA fragmentation features alone can be leveraged to classify cancer and healthy samples with a high accuracy [AUC, 0.989 (high ctDNA cancers) and 0.891 (low ctDNA cancers)].

Analysis of fragment sizes could provide improvements in other applications. Introducing fragment size information on each read could enhance mutation-calling algorithms from high-depth sequencing to distinguish tumor-derived mutations from other sources such as somatic variants or background sequencing noise. In addition, cfDNA from patients analyzed with CHIP is likely to be structurally different from ctDNA released during tumor cell proliferation (*18*, *19*). Thus, fragmentation analysis or selective sequencing strategies could be applied to distinguish clinically relevant tumor mutations from those present in clonal expansions of normal cells. This will be critical for the development of cfDNA-based methods for identification of patients with early-stage cancer.

Size selection could also have an impact on the detection of other types of DNA in body fluids or enrichment of signals from circulating bacterial or pathogen DNA and mitochondrial DNA. These DNA fragments are not associated with nucleosomes and are often highly fragmented below 100 bp. Filtering or selection of such fragments may prove to be important in light of the recently established link between the microbiome and treatment efficiency (*17*, *44*). Moreover, recent work highlights a stronger correlation of ctDNA detection with cellular proliferation than with cell death (*45*). We hypothesize that the mode of the distribution of ctDNA fragment sizes at 145 bp could reflect cfDNA released during cell proliferation, and the fragments at 167 bp may reflect cfDNA released by apoptosis or maturation/turnover of blood cells. The effect of other cancer hallmarks (*46*) on ctDNA biology, structure, concentration, and release is yet unknown.

In summary, ctDNA fragment size analysis, via size selection and machine learning approaches, boosts noninvasive genomic analysis of tumor DNA. Size selection of shorter plasma DNA fragments enriches ctDNA and assists in the identification of a greater number of genomic alterations with both targeted and untargeted sequencing at minimal additional cost. Combining cfDNA fragment size analysis and the detection of SCNAs with a nonlinear classification algorithm improved the discrimination between samples from patients with cancer and those from healthy individuals. Because the analysis of fragment sizes is based on the structural properties of ctDNA, size selection could be used with any downstream sequencing applications. Our work could help overcome current limitations of sensitivity for liquid biopsy, supporting expanded clinical and research applications. Our results indicate that exploiting the endogenous biological properties of cfDNA provides an alternative paradigm to deeper sequencing of ctDNA.

MATERIALS AND METHODS

Study design

Three hundred forty-four plasma samples from 200 patients with multiple cancer types were collected along with plasma from 65 healthy controls. Among the patients, 172 individuals, and notably the OV04 samples, were recruited through prospective clinical studies at Addenbrooke's Hospital, Cambridge, UK, approved by the local research ethics committee (REC reference number: 07/Q0106/63; and National Research Ethics Service Committee East of England–Cambridge Central 03/018). Written informed consent was obtained from all patients, and blood samples were collected before and after initiation of treatment with surgery or chemotherapeutic agents. DNA was extracted from 2 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (QIAGEN) or QIAsymphony (QIAGEN) according to the manufacturer's instructions. In addition, 28 patients were recruited as part of the Copenhagen Prospective Personalized Oncology (CoPPO) program (PMID reference number: 25046202) at Rigshospitalet, Copenhagen, Denmark, approved by the local research ethics committee. Baseline tumor tissue biopsies were available from all 28 patients, together with rebiopsies collected at relapse from two patients, and matched plasma samples. Brain tumor patients were recruited at Addenbrooke's Hospital, Cambridge, UK as part of the BLING (bopsies of liquids in new gliomas) study (REC reference number: 15/EE/0094). Patients with bladder cancer were recruited at the Netherlands Cancer Institute, Amsterdam, The Netherlands, and approval according to national guidelines was obtained (N13KCM/CFMPB250) (*47*). Sixty-five plasma samples were obtained from healthy control individuals using a similar collection protocol (Seralab). Plasma samples have not been freeze thawed more than two times to reduce artifactual fragmentation of cfDNA. A flowchart of the study is presented in fig. S1.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/10/466/eaat4921/DC1

Materials and Methods

Fig. S1. Flowchart summarizing the experiments performed in this study and the sample numbers used at each step.

Fig. S2. Size distribution of cfDNA determined by sWGS for different cancer types.

Fig. S3. Insert size distribution of mutant cfDNA determined with hybrid-capture sequencing for 19 patients.

Fig. S4. DNA fragment size distribution for plasma samples from patients with ovarian cancer.

Fig. S5. Quality control assessed for in vitro size selection.

Fig. S6. Quality control assessed for in vitro and in silico size selection on healthy control samples.

Fig. S7. SCNA analysis of the segmental $\log_2$ ratio determined after sWGS (<0.4× coverage) for the patient OV04-83.

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

Fig. S8. SCNA analysis of the segmental $\log_2$ ratio determined after sWGS (<0.4× coverage) for plasma samples from patients with ovarian cancer (from the OV04 study).
Fig. S9. MAF and t-MAD score compared for different cancer types.
Fig. S10. t-MAD score measured on a plasma DNA dilution series.
Fig. S11. t-MAD scores and fragmentation features compared to tumor volume.
Fig. S12. Changes to t-MAD after in vitro size selection.
Fig. S13. SCNA analysis in cfDNA from plasma samples collected at baseline and after treatment for 13 patients with HGSOC.
Fig. S14. MAF for SNVs called by WES with and without size selection.
Fig. S15. TAm-Seq before and after in vitro size selection.
Fig. S16. Mutations in clinically relevant genes detected by WES with and without in silico size selection.
Fig. S17. Size distribution of nonmutant DNA and ctDNA concentration.
Fig. S18. ROC curve for individual fragmentation features in high ctDNA cancers versus controls.
Fig. S19. t-MAD score compared with seven fragmentation features.
Fig. S20. Performance metrics for the two algorithms, LR and RF.
Fig. S21. LR and RF models using the fragmentation features without t-MAD.
Table S1. Summary table of the patients and samples included in this study.
Table S2. Values for nine fragmentation features determined from sWGS data for the samples included in the study.
Table S3. t-MAD score for the 48 plasma samples of the OV04 cohort before and after in vitro size selection.
Table S4. Log_2 of the signal ratio observed by sWGS of the plasma samples from the OV04 cohort.
Table S5. Mutations called by WES of six patients selected from the OV04 cohort.
Table S6. Mutations called by WES data of the plasma samples from 16 patients from the CoPPO cohort.
References (*48*, *49*)

REFERENCES AND NOTES

1. G. Siravegna, S. Marsoni, S. Siena, A. Bardelli, Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **14**, 531–548 (2017).
2. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
3. M. Murtaza, S.-J. Dawson, D. W. Y. Tsui, D. Gale, T. Forshew, A. M. Piskorz, C. Parkinson, S.-F. Chin, Z. Kingsbury, A. S. C. Wong, F. Marass, S. Humphray, J. Hadfield, D. Bentley, T. M. Chin, J. D. Brenton, C. Caldas, N. Rosenfeld, Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108–112 (2013).
4. V. A. Adalsteinsson, G. Ha, S. S. Freeman, A. D. Choudhury, D. G. Stover, H. A. Parsons, G. Gydush, S. C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. M. Francis, C.-Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M. S. Merrill, R. A. Santiago, E. P. O'Connor, S. H. Jeong, R. Leeson, R. M. Barry, J. F. Kramkowski, Z. Zhang, L. Polacek, J. G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N. M. Oliver, L. Marini, A. G. Waks, L. C. Harshman, S. M. Tolaney, E. M. Van Allen, E. P. Winer, N. U. Lin, M. Nakabayashi, M.-E. Taplin, C. M. Johannessen, L. A. Garraway, T. R. Golub, J. S. Boehm, N. Wagle, G. Getz, J. C. Love, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
5. E. Heitzer, P. Ulz, J. Belic, S. Gutschi, F. Quehenberger, K. Fischereder, T. Benezeder, M. Auer, C. Pischler, S. Mannweiler, M. Pichler, F. Eisner, M. Haeusler, S. Riethdorf, K. Pantel, H. Samonigg, G. Hoefler, H. Augustin, J. B. Geigl, M. R. Speicher, Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med.* **5**, 30 (2013).
6. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, I.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
7. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz Jr., S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16368–16373 (2005).
8. S.-J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
9. F. Diehl, K. Schmidt, M. A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S. A. Szabo, K. W. Kinzler, B. Vogelstein, L. A. Diaz, Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
10. J. Tie, Y. Wang, C. Tomasetti, L. Li, S. Springer, I. Kinde, N. Silliman, M. Tacey, H.-L. Wong, M. Christie, S. Kosmider, I. Skinner, R. Wong, M. Steel, B. Tran, J. Desai, I. Jones, A. Haydon, T. Hayes, T. J. Price, R. L. Strausberg, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, P. Gibbs, Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage II colon cancer. *Sci. Transl. Med.* **8**, 346ra92 (2016).
11. A. A. Chaudhuri, J. J. Chabon, A. F. Lovejoy, A. M. Newman, H. Stehr, T. D. Azad, M. S. Khodadoust, M. S. Esfahani, C. L. Liu, L. Zhou, F. Scherer, D. M. Kurtz, C. Say, J. N. Carter, D. J. Merriott, J. C. Dudley, M. S. Binkley, L. Modlin, S. K. Padda, M. F. Gensheimer, R. B. West, J. B. Shrager, J. W. Neal, H. A. Wakelee, B. W. Loo Jr., A. A. Alizadeh, M. Diehn, Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov.* **7**, 1394–1403 (2017).
12. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz Jr., C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
13. I. S. Haque, O. Elemento, Challenges in using ctDNA to achieve early detection of cancer. bioRxiv 237578 [preprint]. 21 December 2017. https://doi.org/10.1101/237578
14. A. M. Newman, A. F. Lovejoy, D. M. Klass, D. M. Kurtz, J. J. Chabon, F. Scherer, H. Stehr, C. L. Liu, S. V Bratman, C. Say, L. Zhou, J. N. Carter, R. B. West, G. W. Sledge Jr., J. B. Shrager, B. W. Loo Jr., J. W. Neal, H. A. Wakelee, M. Diehn, A. A. Alizadeh, Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
15. P. Ulz, G. G. Thallinger, M. Auer, R. Graf, K. Kashofer, S. W. Jahn, L. Abete, G. Pristauz, E. Petru, J. B. Geigl, E. Heitzer, M. R. Speicher, Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat. Genet.* **48**, 1273–1278 (2016).
16. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
17. P. Burnham, M. S. Kim, S. Agbor-Enoh, H. Luikart, H. A. Valantine, K. K. Khush, I. De Vlaminck, Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. *Sci. Rep.* **6**, 27859 (2016).
18. G. Genovese, A. K. Kähler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landén, M. Höglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Grönberg, C. M. Hultman, S. A. McCarroll, Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
19. Y. Hu, B. Ulrich, J. Supplee, Y. Kuang, P. H. Lizotte, N. B. Feeney, N. M. Guibert, M. M. Awad, K.-K. Wong, P. A. Jänne, C. P. Paweletz, G. R. Oxnard, False-positive plasma genotyping due to clonal hematopoiesis. *Clin. Cancer Res.* **24**, 4437–4443 (2018).
20. A. J. Bronkhorst, J. F. Wentzel, J. Aucamp, E. van Dyk, L. du Plessis, P. J. Pretorius, Characterization of the cell-free DNA released by cultured cancer cells, *Biochim. Biophys. Acta* **1863**, 157–165 (2016).
21. S. Jahr, H. Hentze, S. Englisch, D. Hardt, F. O. Fackelmayer, R.-D. Hesch, R. Knippers, DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).
22. Y. M. D. Lo, K. C. A. Chan, H. Sun, E. Z. Chen, P. Jiang, F. M. F. Lun, Y. W. Zheng, T. Y. Leung, T. K. Lau, C. R. Cantor, R. W. K. Chiu, Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl. Med.* **2**, 61ra91 (2010).
23. D. Chandrananda, N. P. Thorne, M. Bahlo, High-resolution characterization of sequence signatures due to non-random cleavage of cell-free DNA. *BMC Med. Genomics* **8**, 29 (2015).
24. P. Jiang, Y. M. D. Lo, The long and short of circulating cell-free DNA and the ins and outs of molecular diagnostics. *Trends Genet.* **32**, 360–371 (2016).
25. S. C. Y. Yu, K. C. A. Chan, Y. W. L. Zheng, P. Jiang, G. J. W. Liao, H. Sun, R. Akolekar, T. Y. Leung, A. T. J. I. Go, J. M. G. van Vugt, R. Minekawa, C. B. M. Oudejans, K. H. Nicolaides, R. W. K. Chiu, Y. M. D. Lo, Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8583–8588 (2014).
26. F. M. F. Lun, N. B. Y. Tsui, K. C. A. Chan, T. Y. Leung, T. K. Lau, P. Charoenkwan, K. C. K. Chow, W. Y. W. Lo, C. Wanapirak, T. Sanguansermsri, C. R. Cantor, R. W. K. Chiu, Y. M. D. Lo, Noninvasive prenatal diagnosis of monogenic diseases by digital size

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

selection and relative mutation dosage on DNA in maternal plasma. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19920–19925 (2008).

27. G. Minarik, G. Repiska, M. Hyblova, E. Nagyova, K. Soltys, J. Budis, F. Duris, R. Sysak, M. Gerykova Bujalkova, B. Vlkova-Izrael, O. Biro, B. Nagy, T. Szemes, Utilization of benchtop next generation sequencing platforms ion torrent PGM and MiSeq in noninvasive prenatal testing for chromosome 21 trisomy and testing of impact of in silico and physical size selection on its analytical performance. *PLOS ONE* **10**, e0144811 (2015).
28. M. B. Giacona, G. C. Ruben, K. A. Iczkowski, T. B. Roos, D. M. Porter, G. D. Sorenson, Cell-free DNA in human blood plasma: Length measurements in patients with pancreatic cancer and healthy controls. *Pancreas* **17**, 89–97 (1998).
29. N. Umetani, A. E. Giuliano, S. H. Hiramatsu, F. Amersi, T. Nakagawa, S. Martino, D. S. B. Hoon, Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J. Clin. Oncol.* **24**, 4270–4276 (2006).
30. F. Mouliere, B. Robert, E. Arnau Peyrotte, M. Del Rio, M. Ychou, F. Molina, C. Gongora, A. R. Thierry, High fragmentation characterizes tumour-derived circulating DNA. *PLOS ONE* **6**, e23418 (2011).
31. F. Mouliere, S. El Messaoudi, D. Pang, A. Dritschilo, A. R. Thierry, Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
32. P. Jiang, C. W. M. Chan, K. C. A. Chan, S. H. Cheng, J. Wong, V. W.-S. Wong, G. L. H. Wong, S. L. Chan, T. S. K. Mok, H. L. Y. Chan, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1317–E1325 (2015).
33. H. R. Underhill, J. O. Kitzman, S. Hellwig, N. C. Welker, R. Daza, D. N. Baker, K. M. Gligorich, R. C. Rostomily, M. P. Bronner, J. Shendure, Fragment length of circulating tumor DNA. *PLOS Genet.* **12**, e1006162 (2016).
34. O. A. Zill, K. C. Banks, S. R. Fairclough, S. A. Mortimer, J. V. Vowles, R. Mokhtari, D. R. Gandara, P. C. Mack, J. I. Odegaard, R. J. Nagy, A. M. Baca, H. Eltoukhy, D. I. Chudova, R. B. Lanman, A. Talasaz, The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin. Cancer Res.* **24**, 3528–3538 (2018).
35. G. Macintyre, T. E. Goranova, D. De Silva, D. Ennis, A. M. Piskorz, M. Eldridge, D. Sie, L.-A. Lewsley, A. Hanif, C. Wilson, S. Dowson, R. M. Glasspool, M. Lockley, E. Brockbank, A. Montes, A. Walther, S. Sundar, R. Edmondson, G. D. Hall, A. Clamp, C. Gourley, M. Hall, C. Fotopoulou, H. Gabra, J. Paul, A. Supernat, D. Millan, A. Hoyle, G. Bryson, C. Nourse, L. Mincarelli, L. N. Sanchez, B. Ylstra, M. Jimenez-Linan, L. Moore, O. Hofmann, F. Markowetz, I. A. McNeish, J. D. Brenton, Copy number signatures and mutational processes in ovarian carcinoma. *Nat. Genet.* **50**, 1262–1270 (2018).
36. C. A. Parkinson, D. Gale, A. M. Piskorz, H. Biggs, C. Hodgkin, H. Addley, S. Freeman, P. Moyle, E. Sala, K. Sayal, K. Hosking, I. Gounaris, M. Jimenez-Linan, H. M. Earl, W. Qian, N. Rosenfeld, J. D. Brenton, Exploratory analysis of TP53 mutations in circulating tumour DNA as biomarkers of treatment response for patients with relapsed high-grade serous ovarian carcinoma: A retrospective study. *PLOS Med.* **13**, e1002198 (2016).
37. T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68 (2012).
38. A. R. Thierry, S. El Messaoudi, P. B. Gahan, P. Anker, M. Stroun, Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev.* **35**, 347–376 (2016).
39. M. G. Best, N. Sol, B. A. Tannous, P. Wesseling, T. Wurdinger, RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. *Cancer Cell* **28**, 666–676 (2015).
40. M. G. Best, N. Sol, S. G. J. G. In 't Veld, A. Vancura, M. Muller, A.-L. N. Niemeijer, A. V Fejes, L.-A. Tjon Kon Fat, A. E. Huis In 't Veld, C. Leurs, T. Y. Le Large, L. L. Meijer, I. E. Kooi, F. Rustenburg, P. Schellen, H. Verschueren, E. Post, L. E. Wedekind, J. Bracht, M. Esenkbrink, L. Wils, F. Favaro, J. D. Schoonhoven, J. Tannous, H. Meijers-Heijboer, G. Kazemier, E. Giovannetti, J. C. Reijneveld, S. Idema, J. Killestein, M. Heger, S. C. de Jager, R. T. Urbanus, I. E. Hoefer, G. Pasterkamp, C. Mannhalter, J. Gomez-Arroyo, H.-J. Bogaard, D. P. Noske, W. P. Vandertop, D. van den Broek, B. Ylstra, R. J. A. Nilsson, P. Wesseling, N. Karachaliou, R. Rosell, E. Lee-Lewandrowski, K. B. Lewandrowski, B. A. Tannous, A. J. de Langen, E. F. Smit, M. M. van den Heuvel, T. Wurdinger, Swarm intelligence-enhanced detection of non-small-cell lung cancer using tumor-educated platelets. *Cancer Cell* **32**, 238–252.e9 (2017).
41. A. L. Riediger, S. Dietz, U. Schirmer, M. Meister, I. Heinzmann-Groth, M. Schneider, T. Muley, M. Thomas, H. Sültmann, Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients. *Sci. Rep.* **6**, 33505 (2016).
42. J. Belic, M. Koch, P. Ulz, M. Auer, T. Gerhalter, S. Mohan, K. Fischereder, E. Petru, T. Bauernhofer, J. B. Geigl, M. R. Speicher, E. Heitzer, Rapid identification of plasma DNA samples with increased ctDNA levels by a modified FAST-SeqS approach. *Clin. Chem.* **61**, 838–849 (2015).
43. D. G. Stover, H. A. Parsons, G. Ha, S. S. Freeman, W. T. Barry, H. Guo, A. D. Choudhury, G. Gydush, S. C. Reed, J. Rhoades, D. Rotem, M. E. Hughes, D. A. Dillon, A. H. Partridge, N. Wagle, I. E. Krop, G. Getz, T. R. Golub, J. C. Love, E. P. Winer, S. M. Tolaney, N. U. Lin, V. A. Adalsteinsson, Association of cell-free DNA tumor fraction and somatic copy number alterations with survival in metastatic triple-negative breast cancer. *J. Clin. Oncol.* **36**, 543–553 (2018).
44. B. Routy, E. Le Chatelier, L. Derosa, C. P. M. Duong, M. T. Alou, R. Daillère, A. Fluckiger, M. Messaoudene, C. Rauber, M. P. Roberti, M. Fidelle, C. Flament, V. Poirier-Colame, P. Opolon, C. Klein, K. Iribarren, L. Mondragón, N. Jacquelot, B. Qu, G. Ferrere, C. Clémenson, L. Mezquita, J. R. Masip, C. Naltet, S. Brosseau, C. Kaderbhai, C. Richard, H. Rizvi, F. Levenez, N. Galleron, B. Quinquis, N. Pons, B. Ryffel, V. Minard-Colin, P. Gonin, J.-C. Soria, E. Deutsch, Y. Loriot, F. Ghiringhelli, G. Zalcman, F. Goldwasser, B. Escudier, M. D. Hellmann, A. Eggermont, D. Raoult, L. Albiges, G. Kroemer, L. Zitvogel, Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science* **359**, 91–97 (2018).
45. C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann; TRACERx consortium; PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
46. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
47. K. M. Patel, K. E. van der Vos, C. G. Smith, F. Mouliere, D. Tsui, J. Morris, D. Chandrananda, F. Marass, D. van den Broek, D. E. Neal, V. J. Gnanapragasam, T. Forshew, B. W. van Rhijn, C. E. Massie, N. Rosenfeld, M. S. van der Heijden, Association of plasma and urinary mutant DNA with clinical outcomes in muscle invasive bladder cancer. *Sci. Rep.* **7**, 5554 (2017).
48. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
49. I. Scheinin, D. Sie, H. Bengtsson, M. A. van de Wiel, A. B. Olshen, H. F. van Thuijl, H. F. van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson, B. Ylstra, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).

Acknowledgments: We would like to thank all members of the Rosenfeld Lab and Brenton Lab for their help and constructive discussion, in particular, M. Thompson, A. Ruiz-Valdepenas, J. P. Y. Chan, and A. L. Riediger. We would also like to thank the Cancer Research UK Cambridge Institute core facilities for their support, in particular, the genomics, bioinformatics, and biorepository facilities. Support is also acknowledged from the Cancer Research UK Cambridge Cancer Centre, the Cambridge Experimental Cancer Medicine Centre (ECMC), Cancer Molecular Diagnostics Laboratory (CMDL), and NIHR Biomedical Research Centre (BRC). We would like to acknowledge our patients and caregivers and the help and support of the research nurses, trial staff, and the staff at Addenbrooke's Hospital and Rigshospitalet. In particular, we would like to acknowledge C. Hodgkin, H. Biggs, and K. Hosking. We would like to thank H. Carr and AstraZeneca for support for the CALIBRATE study. **Funding:** We would like to acknowledge the support of the University of Cambridge, Cancer Research UK, and the EPSRC [CRUK grant numbers A11906 (to N.R.), A20240 (to N.R.), A22905 (to J.D.B.), A15601 (to J.D.B.), A25177 (CRUK Cancer Centre Cambridge), A17242 (to K.M.B.), and A16465 (CRUK-EPSRC Imaging Centre in Cambridge and Manchester)]. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC grant agreement no. 337905. The research was supported by the National Institute for Health Research Cambridge, National Cancer Research Network, Cambridge Experimental Cancer Medicine Centre, and Hutchison Whampoa Limited. This research is also supported by Target Ovarian Cancer and the Medical Research Council through their Joint Clinical Research Training Fellowship for E.K.M. The CALIBRATE study was supported by funding from AstraZeneca. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **Author contributions:** F. Mouliere, A.M.P., D.C.,

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

E.K.M., J.D.B., and N.R. conceptualized and designed the study. F. Mouliere, A.M.P., E.K.M., L.B.A., K.H., C.G.S., J.C.M.W., D.G., R.M., T.G., A.S., I.G., O.Ø., C.A.P., M.M.-S., I.H., K.P., C.E.M., and W.N.C. performed experiments and collected data. F. Mouliere, A.M.P., D.C., E.K.M., and C.G.S. conceptualized the size selection approach. F. Mouliere, A.M.P., and E.K.M. designed and performed in vitro size selection. F. Mouliere and D.C. conceptualized and designed the fragmentation feature analysis, with input from F. Marass and N.R. D.C. conceptualized and designed the t-MAD index with input from F. Mouliere. F. Mouliere and D.C. carried out bioinformatics analysis of SCNAs from sWGS. J.M. performed bioinformatics analysis of TAm-Seq. F. Mouliere and L.B.A. designed the tailored captured sequencing and performed WES. F. Mouliere and J.M. performed bioinformatics analysis of the capture sequencing and WES. M.D.E. developed and optimized mutation calling algorithms. R.M., K.M.B., and S.R. designed the animal model. J.G.-C., S.P., R.D.B., M.M.-S., G.D.S., J.B., S.M., P.C., C.W., R.M., and M.S.v.d.H. have collected human samples. M.J.-L. and J.B. performed histopathology revision. F. Mouliere, D.C., A.M.P., E.K.M., J.D.B., and N.R. wrote the manuscript. All authors have critically reviewed the manuscript. F. Mouliere, A.M.P., D.C., J.D.B., and N.R. supervised the study. F. Mouliere coordinated the study. **Competing interests:** N.R., J.D.B., and D.G. are cofounders, shareholders, and officers/consultants of Inivata Ltd., a cancer genomics company that commercializes ctDNA analysis. Inivata Ltd. had no role in the conceptualization, study design, data collection and analysis, and decision to publish or preparation of the manuscript. J.D.B. received research funding from Aprea and NCI and has received advisory board fees from AstraZeneca. F. Marass and N.R. are co-inventors of patent WO/2016/009224 on "A method for detecting a genetic variant." F. Mouliere, J.C.M.W., K.H., C.E.M., C.G.S., N.R., and other authors may be listed as co-inventors on patent application number 1803596.4 on "Improvements in variant detection" and other potential patents describing methods for the analysis of DNA fragments and applications of ctDNA. I.G. is currently an employee of Novartis AG, a relationship that started after all his work contributing to this manuscript had been completed. Novartis had no role in the work presented in this manuscript. Other authors declare that they have no competing interests. **Data and materials availability:** Sequencing data for this study are deposited in the EGA database (accession number EGAS00001003258). Other data associated with this study are present in the paper or the Supplementary Materials.

Submitted 29 March 2018
Accepted 17 October 2018
Published 7 November 2018
10.1126/scitranslmed.aat4921

Citation: F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. van der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

Science Translational Medicine

Enhanced detection of circulating tumor DNA by fragment size analysis

Florent Mouliere, Dineika Chandrananda, Anna M. Piskorz, Elizabeth K. Moore, James Morris, Lise Barlebo Ahlborn, Richard Mair, Teodora Goranova, Francesco Marass, Katrin Heider, Jonathan C. M. Wan, Anna Supernat, Irena Hudecova, Ioannis Gounaris, Susana Ros, Mercedes Jimenez-Linan, Javier Garcia-Corbacho, Keval Patel, Olga Østrup, Suzanne Murphy, Matthew D. Eldridge, Davina Gale, Grant D. Stewart, Johanna Burge, Wendy N. Cooper, Michiel S. van der Heijden, Charles E. Massie, Colin Watts, Pippa Corrie, Simon Pacey, Kevin M. Brindle, Richard D. Baird, Morten Mau-Sørensen, Christine A. Parkinson, Christopher G. Smith, James D. Brenton and Nitzan Rosenfeld

Sci Transl Med **10**, eaat4921.
DOI: 10.1126/scitranslmed.aat4921

Sizing up circulating tumor DNA

Unlike solid tumors, which are often hidden deep within a patient's body, a patient's blood is easy to access safely. As a result, liquid biopsy, the analysis of tumor DNA in the blood, is an attractive alternative to conventional biopsy. Unfortunately, tumor DNA molecules are usually vastly outnumbered by the fragments of noncancer DNA in the blood, and detecting them can be a challenge, especially in early stages of cancer. Mouliere *et al*. identified characteristic differences in the size distribution of tumor-derived and noncancer DNA fragments and then used these observations to design a method of tumor DNA detection with greater sensitivity.

ARTICLE TOOLS	http://stm.sciencemag.org/content/10/466/eaat4921
SUPPLEMENTARY MATERIALS	http://stm.sciencemag.org/content/suppl/2018/11/05/10.466.eaat4921.DC1
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/10/466/eaav3873.full http://stm.sciencemag.org/content/scitransmed/9/403/eaan2415.full http://stm.sciencemag.org/content/scitransmed/8/364/364ra155.full http://stm.sciencemag.org/content/scitransmed/8/346/346ra92.full
REFERENCES	This article cites 48 articles, 17 of which you can access for free http://stm.sciencemag.org/content/10/466/eaat4921#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science Translational Medicine* is a registered trademark of AAAS.

Downloaded from http://stm.sciencemag.org/ by guest on January 24, 2019

Appendix 5: List of FDA-approved targeted therapies for solid malignancies and their molecular targets (22)

Agent	**Target(s)**	**FDA-approved indication(s)**
Ado-trastuzumab emtansine (Kadcyla)	HER2 (ERBB2/neu)	Breast cancer (HER2+)
Afatinib (Gilotrif)	EGFR (HER1/ERBB1), HER2 (ERBB2/neu)	Non-small cell lung cancer (with EGFR exon 19 deletions or exon 21 substitution (L858R) mutations)
Aldesleukin (Proleukin)		Renal cell carcinoma
		Melanoma
Alectinib (Alecensa)	ALK	Non-small cell lung cancer (with ALK fusion)
Alemtuzumab (Campath)	CD52	B-cell chronic lymphocytic leukemia
Atezolizumab (Tecentriq)	PD-L1	Urothelial carcinoma
		Non-small cell lung cancer
Avelumab (Bavencio)	PD-L1	Merkel cell carcinoma
		Urothelial cancer
Axitinib (Inlyta)	KIT, PDGFRβ, VEGFR1/2/3	Renal cell carcinoma
Belimumab (Benlysta)	BAFF	Lupus erythematosus
Belinostat (Beleodaq)	HDAC	Peripheral T-cell lymphoma
Bevacizumab (Avastin)	VEGF ligand	Cervical cancer
		Colorectal cancer
		Fallopian tube cancer
		Glioblastoma
		Non-small cell lung cancer
		Ovarian cancer
		Peritoneal cancer
		Renal cell carcinoma
Blinatumomab (Blincyto)	CD19/CD3	Acute lymphoblastic leukemia (precursor B-cell)
Bortezomib (Velcade)	Proteasome	Multiple myeloma
		Mantle cell lymphoma

Appendix 5

Agent	Target(s)	FDA-approved indication(s)
Bosutinib (Bosulif)	ABL	Chronic myelogenous leukemia (Philadelphia chromosomepositive)
Brentuximab vedotin (Adcetris)	CD30	Hodgkin lymphoma
		Anaplastic large cell lymphoma
Brigatinib (Alunbrig)	ALK	Non-small cell lung cancer (ALK+)
Cabozantinib (Cabometyx [tablet], Cometriq [capsule])	FLT3, KIT, MET, RET, VEGFR2	Medullary thyroid cancer
		Renal cell carcinoma
Canakinumab (Ilaris)	IL-1β	Juvenile idiopathic arthritis
		Cryopyrin-associated periodic syndromes
Carfilzomib (Kyprolis)	Proteasome	Multiple myeloma
Ceritinib (Zykadia)	ALK	Non-small cell lung cancer (with ALK fusion)
Cetuximab (Erbitux)	EGFR (HER1/ERBB1)	Colorectal cancer (KRAS wild type)
		Squamous cell cancer of the head and neck
Cobimetinib (Cotellic)	MEK	Melanoma (with BRAF V600E or V600K mutation)
Crizotinib (Xalkori)	ALK, MET, ROS1	Non-small cell lung cancer (with ALK fusion or ROS1 genealteration)
Dabrafenib (Tafinlar)	BRAF	Melanoma (with BRAF V600 mutation)
		Non-small cell lung cancer (with BRAF V600E mutation)
Daratumumab (Darzalex)	CD38	Multiple myeloma
Dasatinib (Sprycel)	ABL	Chronic myelogenous leukemia (Philadelphia chromosomepositive)
		Acute lymphoblastic leukemia (Philadelphia chromosomepositive)
Denosumab (Xgeva)	RANKL	Giant cell tumor of the bone
Dinutuximab (Unituxin)	B4GALNT1 (GD2)	Pediatric neuroblastoma
Durvalumab (Imfinzi)	PD-L1	Urothelial carcinoma
		Non-small cell lung cancer
Elotuzumab (Empliciti)	SLAMF7 (CS1/CD319/CRACC)	Multiple myeloma
Enasidenib (Idhifa)	IDH2	Acute myeloid leukemia (with IDH2 mutation)

Appendix 5

Agent	**Target(s)**	**FDA-approved indication(s)**
Erlotinib (Tarceva)	EGFR (HER1/ERBB1)	Non-small cell lung cancer (with EGFR exon 19 deletions or exon 21 substitution (L858R) mutations)
		Pancreatic cancer
Everolimus (Afinitor)	mTOR	Pancreatic, gastrointestinal, or lung origin neuroendocrine tumor
		Renal cell carcinoma
		Nonresectable subependymal giant cell astrocytoma associated with tuberous sclerosis
		Breast cancer (HR+, HER2-)
Gefitinib (Iressa)	EGFR (HER1/ERBB1)	Non-small cell lung cancer (with EGFR exon 19 deletions or exon 21 substitution (L858R) mutations)
Ibritumomab tiuxetan (Zevalin)	CD20	Non-Hodgkin's lymphoma
Ibrutinib (Imbruvica)	BTK	Mantle cell lymphoma
		Chronic lymphocytic leukemia
		Waldenstrom's macroglobulinemia
Idelalisib (Zydelig)	PI3Kδ	Chronic lymphocytic leukemia
		Follicular B-cell non-Hodgkin lymphoma
		Small lymphocytic lymphoma
Imatinib (Gleevec)	KIT, PDGFR, ABL	GI stromal tumor (KIT+)
		Dermatofibrosarcoma protuberans
		Multiple hematologic malignancies including Philadelphia chromosome-positive ALL and CML
Ipilimumab (Yervoy)	CTLA-4	Melanoma
		Renal cell carcinoma
Ixazomib (Ninlaro)	Proteasome	Multiple Myeloma
Lapatinib (Tykerb)	HER2 (ERBB2/neu), EGFR (HER1/ERBB1)	Breast cancer (HER2+)
Lenvatinib (Lenvima)	VEGFR2	Renal cell carcinoma
		Thyroid cancer
Midostaurin (Rydapt)	FLT3	acute myeloid leukemia (FLT3+)

Appendix 5

Agent	**Target(s)**	**FDA-approved indication(s)**
Necitumumab (Portrazza)	EGFR (HER1/ERBB1)	Squamous non-small cell lung cancer
Neratinib (Nerlynx)	HER2 (ERBB2/neu)	Breast cancer (HER2 overexpressed/amplified)
Nilotinib (Tasigna)	ABL	Chronic myelogenous leukemia (Philadelphia chromosomepositive)
Niraparib (Zejula)	PARP	Ovarian cancer
		Fallopian tube cancer
		Peritoneal cancer
Nivolumab (Opdivo)	PD-1	Colorectal cancer (dMMR and MSI-H)
		Head and neck squamous cell carcinoma
		Hepatocellular carcinoma
		Hodgkin lymphoma
		Melanoma
		Non-small cell lung cancer
		Renal cell carcinoma
		Urothelial carcinoma
Obinutuzumab (Gazyva)	CD20	Chronic lymphocytic leukemia
		Follicular lymphoma
Ofatumumab (Arzerra, HuMax-CD20)	CD20	Chronic lymphocytic leukemia
Olaparib (Lynparza)	PARP	Ovarian cancer (with BRCA mutation)
Olaratumab (Lartruvo)	PDGFRα	Soft tissue sarcoma
Osimertinib (Tagrisso)	EGFR	Non-small cell lung cancer (with EGFR T790M mutation)
Palbociclib (Ibrance)	CDK4, CDK6	Breast cancer (HR+, HER2-)
Panitumumab (Vectibix)	EGFR (HER1/ERBB1)	Colorectal cancer (KRAS wild type)
Panobinostat (Farydak)	HDAC	Multiple myeloma
Pazopanib (Votrient)	VEGFR, PDGFR, KIT	Renal cell carcinoma

Appendix 5

Agent	**Target(s)**	**FDA-approved indication(s)**
Pembrolizumab (Keytruda)	PD-1	Classical Hodgkin lymphoma
		Colorectal cancer (MSI-H/dMMR)
		Gastric cancer
		Melanoma
		Non-small cell lung cancer (PD-L1+)
		Head and neck squamous cell carcinoma
		Urothelial cancer
		Solid tumors (MSI-H/dMMR)
Pertuzumab (Perjeta)	HER2 (ERBB2/neu)	Breast cancer (HER2+)
Ponatinib (Iclusig)	ABL, FGFR1-3, FLT3, VEGFR2	Chronic myelogenous leukemia
		Acute lymphoblastic leukemia (Philadelphia chromosomepositive)
Ramucirumab (Cyramza)	VEGFR2	Colorectal cancer
		Gastric cancer or Gastroesophageal junction (GEJ) adenocarcinoma
		Non-small cell lung cancer
Regorafenib (Stivarga)	KIT, PDGFRβ, RAF, RET, VEGFR1/2/3	Colorectal cancer
		Gastrointestinal stromal tumors
		Hepatocellular carcinoma
Ribociclib (Kisqali)	CDK4, CDK6	Breast cancer (HR+, HER2-)
Rituximab (Rituxan, Mabthera)	CD20	Non-Hodgkin’s lymphoma
		Chronic lymphocytic leukemia
		Rheumatoid arthritis
		Granulomatosis with polyangiitis
Rituximab/hyaluronidase human (Rituxan Hycela)	CD20	Chronic lymphocytic leukemia
		Diffuse large B-cell lymphoma
		Follicular lymphoma

Appendix 5

Agent	**Target(s)**	**FDA-approved indication(s)**
Romidepsin (Istodax)	HDAC	Cutaneous T-cell lymphoma
		Peripheral T-cell lymphoma
Rucaparib (Rubraca)	PARP	Ovarian cancer (with BRCA mutation)
Ruxolitinib (Jakafi)	JAK1/2	Myelofibrosis
Siltuximab (Sylvant)	IL-6	Multicentric Castleman's disease
Sipuleucel-T (Provenge)		Prostate cancer
Sonidegib (Odomzo)	Smoothened	Basal cell carcinoma
Sorafenib (Nexavar)	VEGFR, PDGFR, KIT, RAF	Hepatocellular carcinoma
		Renal cell carcinoma
		Thyroid carcinoma
Temsirolimus (Torisel)	mTOR	Renal cell carcinoma
Tocilizumab (Actemra)	IL-6R	Rheumatoid arthritis
		Juvenile idiopathic arthritis
Tofacitinib (Xeljanz)	JAK3	Rheumatoid arthritis
Tositumomab (Bexxar)	CD20	Non-Hodgkin's lymphoma
Trametinib (Mekinist)	MEK	Melanoma (with BRAF V600 mutation)
		Non-small cell lung cancer (with BRAF V600E mutation)
Trastuzumab (Herceptin)	HER2 (ERBB2/neu)	Breast cancer (HER2+)
		Gastric cancer (HER2+)
Vandetanib (Caprelsa)	EGFR (HER1/ERBB1), RET, VEGFR2	Medullary thyroid cancer
Vemurafenib (Zelboraf)	BRAF	Melanoma (with BRAF V600 mutation)
Venetoclax (Venclexta)	BCL2	Chronic lymphocytic leukemia (with 17p deletion)
Vismodegib (Erivedge)	PTCH, Smoothened	Basal cell carcinoma
Vorinostat (Zolinza)	HDAC	Cutaneous T-cell lymphoma
Ziv-aflibercept (Zaltrap)	PIGF, VEGFA/B	Colorectal cancer

Appendix 6: List of FDA-cleared or approved companion diagnostic tests or devices (In Vitro and Imaging Tools) (25)

PMA – Pre-market approval; 510(k) – Pre-market notification; HDE – Humanitarian Device Extension
NDA – New drug application; BLA – Biologics Licence application

Diagnostic Name	PMA/ 510(k)/ HDE	Diagnostic Manufacturer	Trade Name (Generic) - NDA/BLA
BRACAnalysis CDx	P140020/S016	Myriad Genetic Laboratories, Inc.	Breast Cancer Lynparza (olaparib) - NDA 208558 Talzenna (talazoparib) – NDA 211651 Ovarian Cancer Lynparza (olaparib) - NDA 208558 Rubraca (rucaparib) – NDA 209115
therascreen EGFR RGQ PCR Kit	P120022/S018	Qiagen Manchester, Ltd.	Non-small cell lung cancer Iressa (gefitinib) - NDA 206995 Gilotrif (afatinib)- NDA 201292 Vizimpro (dacomitinib)- NDA 211288
cobas EGFR Mutation Test v2	P120019/S019	Roche Molecular Systems, Inc.	Non-small cell lung cancer (tissue and plasma) Tarceva (erlotinib) - NDA 021743 Tagrisso (osimertinib) - NDA 208065 Iressa (gefitinib) - NDA 206995
PD-L1 IHC 22C3 pharmDx	P150013/S011	Dako North America, Inc.	Non-small cell lung cancer, gastric or gastroesophageal junction adenocarcinoma, cervical cancer, and urothelial carcinoma Keytruda (pembrolizimab) - BLA 125514
PD-L1 (SP142)	P160002/S006	Ventana Medical Systems, Inc.	Non-small cell lung cancer and urothelial carcinoma Tecentriq (atezolizumab) – NDA 761034/S012

Appendix 6

Diagnostic Name	PMA/ 510(k)/ HDE	Diagnostic Manufacturer	Trade Name (Generic) - NDA/BLA
Abbott RealTime IDH1	P170041	Abbott Molecular, Inc.	Acute myeloid leukemia Tibsovo (ivosidenib) - NDA 211192
MRDx BCR-ABL Test	K173492	MolecularMD Corporation	Chronic myeloid leukemia Tasigna (nilotinib) - NDA 022068/S026
FoundationOne CDx	P170019	Foundation Medicine, Inc.	Non-small cell lung cancer Gilotrif (afatinib) - NDA 201292 Iressa (gefitinib) - NDA 206995 Tarceva (erlotinib) - NDA 021743 Tagrisso (osimertinib) NDA 208065 Alecensa (alectinib) - NDA 208434 Xalkori (crizotinib) - NDA 202570 Zykadia (ceritinib) - NDA 205755 Tafinlar (dabrafenib) - NDA 202806 in combination with Mekinist (trametinib) - NDA 204114 Melanoma Tafinlar (dabrafenib) - NDA 202806 Zelboraf (vemurafenib) - NDA 202429 Mekinist (trametinib) - NDA 204114 or Cotellic (cobimetinib) - NDA 206192 in combination with Zelboraf (vemurafenib) - NDA 202429 Breast cancer Herceptin (trastuzumab) - BLA 103792 Perjeta (pertuzumab) - BLA 125409 Kadcyla (ado-trastuzumab emtansine) - BLA 125427 Colorectal cancer Erbitux (cetuximab) - BLA 125084 Vectibix (panitumumab) - BLA 125147 Ovarian cancer Rubraca (rucaparib) - NDA 209115

Appendix 6

Diagnostic Name	PMA/ 510(k)/ HDE	Diagnostic Manufacturer	Trade Name (Generic) - NDA/BLA
cobas KRAS Mutation Test	P140023	Roche Molecular Systems, Inc.	Colorectal cancer Erbitux (cetuximab) - BLA 125084 Vectibix (panitumumab) - BLA 125147
therascreen KRAS RGQ PCR Kit	P110030 P110027	Qiagen Manchester, Ltd.	Colorectal cancer Erbitux (cetuximab) - BLA 125084 Vectibix (panitumumab) - BLA 125147
Dako EGFR pharmDx Kit	P030044/S002	Dako North America, Inc.	Colorectal cancer Erbitux (cetuximab) - BLA 125084 Vectibix (panitumumab) - BLA 125147
FerriScan	DEN130012/K124065	Resonance Health Analysis Services Pty Ltd	Non-transfusion-dependent thalassemia Exjade (deferasirox) – NDA 021882
Dako c-KIT pharmDx	P040011	Dako North America, Inc.	Gastrointestinal stromal tumors Gleevec (imatinib mesylate) – NDA 021335 Glivec (imatinib mesylate) – NDA 021588
INFORM HER-2/neu	P940004	Ventana Medical Systems, Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792
PathVysion HER-2 DNA Probe Kit	P980024	Abbott Molecular Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792
PATHWAY anti-Her2/neu (4B5) Rabbit Monoclonal Primary Antibody	P990081/S001-S028	Ventana Medical Systems, Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792
InSite Her-2/neu KIT	P040030	Biogenex Laboratories, Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792
SPOT-LIGHT HER2 CISH Kit	P050040/S001-S003	Life Technologies Corporation	Breast cancer Herceptin (trastuzumab) - BLA 103792
Bond Oracle HER2 IHC System	P090015	Leica Biosystems	Breast cancer Herceptin (trastuzumab) - BLA 103792
HER2 CISH pharmDx Kit	P100024	Dako Denmark A/S	Breast cancer Herceptin (trastuzumab) - BLA 103792

Appendix 6

Diagnostic Name	PMA/ 510(k)/ HDE	Diagnostic Manufacturer	Trade Name (Generic) - NDA/BLA
INFORM HER2 Dual ISH DNA Probe Cocktail	P100027	Ventana Medical Systems, Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792
HercepTest	P980018/S018	Dako Denmark A/S	Breast cancer Herceptin (trastuzumab) - BLA 103792 Perjeta (pertuzumab) - BLA 125409 Kadcyla (ado-trastuzumab emtansine) - BLA 125427 Gastic and gastroesophogeal cancer Herceptin (trastuzumab) - BLA 103792
HER2 FISH pharmDx Kit	P040005/S009	Dako Denmark A/S	Breast cancer Herceptin (trastuzumab) - BLA 103792 Perjeta (pertuzumab) - BLA 125409 Kadcyla (ado-trastuzumab emtansine) - BLA 125427 Gastic and gastroesophogeal cancer Herceptin (trastuzumab) - BLA 103792
THXID BRAF Kit	P120014	bioMérieux Inc.	Breast cancer Herceptin (trastuzumab) - BLA 103792 Mekinist (tramatenib) - NDA 204114 Tafinlar (dabrafenib) – NDA 202806
Vysis ALK Break Apart FISH Probe Kit	P110012	Abbott Molecular Inc.	Non-small cell lung cancer Xalkori (crizotinib) – NDA 202570
cobas 4800 BRAF V600 Mutation Test	P110020/S016	Roche Molecular Systems, Inc.	Melanoma Zelboraf (vemurafenib) - NDA 202429 Cotellic (cobimetinib) - NDA 206192 in combination with Zelboraf (vemurafenib) - NDA 202429