The Microbiomes of Urine and the Prostate are Linked to Human Prostate Cancer Risk Groups

Rachel Hurst[a,†], Emma Meader[b,†], Abraham Gihawi[a,†], Ghanasyam Rallapalli[a,†], Jeremy Clark[a], Gemma L. Kay[a,c], Martyn Webb[a], Kate Manley[a,d], Helen Curley[a], Helen Walker[d], Ravi Kumar[a], Katarzyna Schmidt[b], Lisa Crossman[a,c], Rosalind A. Eeles[e,f], David C. Wedge[g,h], Andy G. Lynch[i], Charlie E. Massie[j], CRUK-ICGC Prostate Group[k], Marcelino Yazbek-Hanna[d], Mark Rochester[d], Robert D. Mills[d], Richard F. Mithen[c,l], Maria H. Traka[c], Richard Y. Ball[m], Justin O' Grady[a,c‡], Daniel S. Brewer[a,n‡], John Wain[a,c‡], Colin S. Cooper[a,‡,*]

[a]Norwich Medical School, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK

[b]Microbiology Department, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Lane, Norwich, NR4 7UY, UK

[c]Quadram Institute Biosciences, Colney Lane, Norwich, NR4 7UQ, UK

[d]Department of Urology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Lane, Norwich, NR4 7UY, UK

[e]The Institute of Cancer Research, London, SW7 3RP, UK

[f]Royal Marsden NHS Foundation Trust, London and Sutton, SM2 5PT, UK

[g]Oxford Big Data Institute, University of Oxford, Old Road Campus, Oxford, OX3 7LF, UK

[h]University of Manchester, Oxford Rd, Manchester, M13 9PL, UK

[i]School of Medicine and School of Mathematics and Statistics, University of St Andrews, St Andrews, UK

[j]Hutchison/MRC Research Centre, Cambridge University, Cambridge, CB2 0XZ, UK

[k]The CRUK ICGC Prostate Group

[l]Liggins Institute, University of Auckland, 85 Park Road, Grafton, Auckland 1023, New Zealand

[m]Norfolk and Waveney Cellular Pathology Service, Norfolk and Norwich University Hospitals NHS Foundation Trust, Colney Lane, Norwich, NR4 7UB, UK

[n]Earlham Institute, Norwich Research Park Innovation Centre, Colney Lane, Norwich NR4 7UZ

† These authors contributed equally to this work.

ǂ These authors jointly supervised this work.

*Corresponding author. Norwich Medical School, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK. E-mail address: colin.cooper@uea.ac.uk

Keywords: Anaerobic bacteria; Cancer progression; Metagenomic sequencing; Metastases; Microbiota; Microbiome; Prognosis; Prostate cancer; Urine; 16S Ribosomal Amplicon Sequencing.

Word count of the text: 2747

Word count of the abstract: 293

Abstract

Background: Bacteria play a suspected role in the development of several cancer types and associations between the presence of particular bacteria and prostate cancer have been reported.

Objective: To provide an improved characterisation of the prostate and urine microbiome and to investigate the prognostic potential of the bacteria present.

Design, Setting, and Participants: Microbiome profiles were interrogated in sample collections of patient urine (sediment microscopy: *n*=318, 16S ribosomal amplicon sequencing: *n*=46; extracellular vesicle RNA-Seq: *n*=40) and cancer tissue (*n*=204).

Outcome Measurements and Statistical Analysis: Microbiomes were assessed using anaerobic culture, population level 16S analysis, RNA-Seq and whole genome DNA sequencing.

Results and Limitations: We demonstrate an association between the presence of bacteria in urine sediments and higher D'Amico risk prostate cancer (discovery, *n*=215 patients, $p<0.001$; validation, *n*=103, $p<0.001$, χ^2 test for trend). Characterisation of the bacterial community led to (i) the identification of four novel bacteria (*Porphyromonas sp. nov., Varibaculum sp. nov., Peptoniphilus sp. nov.*, and *Fenollaria sp. nov.)* that were frequently found in patient urine; and (ii) to the definition of a patient subgroup associated with metastasis development ($p=0.015$, log-rank test). The presence of five specific anaerobic genera, which includes three of the novel isolates, was associated with cancer risk group: in urine sediment ($p=0.045$, log-rank test), urine extracellular vesicles ($p=0.039$), and cancer tissue ($p=0.035$); with a meta-analysis hazard ratio for disease progression of 2.60 (95% CI:1.39-4.85; $p=0.003$; Cox regression). A limitation is that functional links to cancer development are not yet established.

Conclusions: This study characterises prostate and urine microbiomes and indicates that specific anaerobic bacteria genera have prognostic potential.

Patient Summary: In this study we investigated bacteria present in patient urine and prostates,

69 we identify four novel bacteria and suggest a potential prognostic utility for the microbiome in
70 prostate cancer.
71

1. Introduction

Prostate cancer is the most common non-skin malignancy in men in developed countries, with over 250,000 deaths annually worldwide [1]. The clinical course of prostate cancer is highly heterogeneous and critical decisions are made about the likelihood of aggressive disease based on information obtained at presentation, including histopathological Gleason score determined following biopsy [2]. Determining urinary biomarkers to identify aggressive prostate cancer is an area of growing interest. Material secreted by the prostate gland appears in the urine, and reflux of urine into the prostate is well established supporting the existence of a prostate-urine loop [3-5]. Urine biomarkers identified include assessment of gene methylation [6], or gene expression profiles including the PCA3, and various gene combinations [7-9]. However, none of these tests are in widespread clinical use and the challenge remains to find a combination of biomarkers and clinical data that at initial patient assessment can reliably predict prostate cancer risk groups and disease progression.

Genetic inheritance and ethnicity have established roles [10, 11] in prostate cancer development while chronic inflammation has also been proposed as an aetiological factor [12-14]. *Helicobacter pylori* has an established role in the development of gastric cancer [15] stimulating the search for microbial involvement in the development of other cancers. Bacteria are known to be present in the urogenital tract and in prostate tissue [12, 16] and bacteria isolated from the prostate can cause inflammation in animal models [12-14]. Encouragingly microbes present in prostate tissue differ between patients with different Gleason grades [17] and there are links between the presence of prostate cancer and distinct microbial profiles of the urine [13, 18] and the gastrointestinal tract [18-20].

In this study we used fluorescent microscopy, anaerobic culture, 16S ribosomal amplicon sequencing, mRNA sequencing, and whole genome DNA sequencing to detect bacteria in post-

DRE urine and prostate cancer tissue. We (i) use non-biased "tree of life" [21] methodology to isolate and classify novel bacteria and (ii) search for associations between the presence of bacteria and prostate cancer risk groups.

2. Patients and methods

Detailed methods can be found in the Supplementary materials.

2.1 Patient recruitment and specimen collection

Ethical approval was obtained from the local research ethics committee, 12/EE/0058. Patients were categorised into clinical groups (Supplementary Material), prostate cancers were stratified according to D'Amico risk group [22]. Urine samples were collected [7] (Apr-2012 to Jan-2015) post-Digital Rectal Examination (DRE, prostate massage 3 strokes per lobe from base to apex), prior to biopsy, from patients undergoing assessment for prostate cancer or haematuria at the Norfolk and Norwich University Hospital and processed immediately using sterile techniques. Urine sediments and extracellular vesicle fractions were prepared as previously described [7], with an additional step for detection of bacteria by microscopy (Supplementary methods). Prostate secretions (100-400µl) were collected (May-2017 to Feb-2020) via manual compression of the excised prostate <20 minutes post-prostatectomy. Samples were processed immediately, snap frozen on dry-ice and stored at -80^{o}C.

2.2 Metagenomics, 16S ribosomal amplicon DNA sequencing and RNA-seq metatranscriptomics

DNA extraction from urine sediment was similar to Yu and Morrison, 2004 [23] with repeated bead-beating extraction to maximise bacterial DNA yield. Bacterial 16S DNA was amplified and sequenced (V1-V3/V3-V5 hypervariable regions). Controls included no template controls, elution buffer controls and blank bead-beating extraction samples. Quantitative PCR (qPCR)

assays detected several bacterial genera and species. Urine extracellular vesicle total RNA was extracted as previously described [7] from 40 urine samples and were sequenced and processed with the SEPATH [24] pipeline.

2.3 Detection of bacteria in ICGC prostate tissue whole genome sequences

Unmapped reads from human-aligned whole genome sequencing data (International Cancer Genome Consortium, ICGC prostate cancer tissue *n*=204, collected from Mar-2004 to Jun-2014) were classified using a curated BWA database containing GRCh38, 75 study isolates and strains frequently identified by Kraken (Supplementary material). Reads were filtered to have a minimum mapping quality of 20, 50bp minimum alignment and were subject to complexity filtering. Assemblies with 200bp or more of their genome covered were considered present in the sample.

2.4 Isolate anaerobic culture, whole genome sequencing and assembly, phylogenetic and metabolic pathway analyses

Urine or prostate secretion samples were inoculated into pre-reduced PY broth or Brucella blood agar plates with 5% sheep blood and vitamin K1/hemin supplementation (Beckton Dickinson GmbH Heidelberg, Germany) and grown in an anaerobic cabinet supplied with 5% Hydrogen, 10% CO_2, and 85% Nitrogen at 37^{o}C. Pure colonies were picked and prepared for DNA extraction and sequenced with Nextera XT library preparation on Illumina MiSeq using V3 reagents (2x300bp). MinION nanopore sequencing was used on three novel species for hybrid assembly (Unicycler). Phylogenetic analysis was carried out as previously described [21] using multiple sequence alignments of 16S ribosomal proteins from isolates and known strains. Metabolic pathways were predicted using InterProScan REST api v5.29-68.053 (Additional Data File 1).

2.5 Statistical analysis

For the urine sediment microscopy dataset (*n*=318), patients were categorised into clinical groups including low-, intermediate- and high-risk prostate cancer (for further details see Supplementary Material) [22]. Further data for each cohort including clinical characteristics are provided in Supplementary Tables 1-4. Follow-up for the clinical cohorts was over 3-4 years (median 2.7 years) or up to 6 years post sample collection (median 5.2 years) for the 16S and RNAseq datasets respectively. For the cancer tissue dataset (n=204) follow up data was up to 9.8 years (median 3.5 years) (Supplementary Material). Progression events were detection of prostate cancer metastasis or PSA biochemical failure following initial treatment (2 PSA tests ≥0.2ng/ml). Survival analyses include Kaplan-Meier curves, Cox proportional hazards models, and the log-rank test. Random-effect meta-analysis based on log hazard ratios was carried out with metagen function (meta R package).

3. Results

3.1 Bacteria in urine are associated with increased risk groups of prostate cancer

Examination of post-DRE urine sediments revealed background DNA staining of bacteria in ~50% samples, supported by scanning electron microscopy and by 16S ribosomal RNA gene detection (Supplementary Fig. 1A,B, Supplementary Fig. 2, Supplementary Table 5). To further investigate this observation we analysed urine samples from men undergoing assessment for prostate cancer (*n*=300) or from a haematuria clinic (*n*=18). Background DNA staining of bacteria was more common from men with intermediate and high D,Amico risk group and advanced prostate cancer (Table 1) with a statistically significant association between the presence of bacteria and increased risk group of prostate cancer (χ^2 test for trend in proportions: discovery set $p<0.001$, $n=215$; validation $p<0.001$,$n=103$). A similar significant

association with the presence of bacteria was observed in each of the component parts of risk groups, including PSA, Gleason Score and clinical stage (Supplementary Fig. 3). For the combined data set large aggregates of bacteria in urine were also significantly associated with increased prostate cancer risk group (p=0.006; Supplementary Fig. 2). The bacteria also appeared, in some cases, to be intracellular within human cells (Supplementary Fig. 4).

3.2 Culture confirms new species of bacteria from the urine of prostate cancer patients.

To identify the bacteria involved, we applied 16S sequencing on urine sediments from 46 men (24 with a diagnosis of prostate cancer) using accepted protocols and controls to avoid contamination [25, 26]. The bacterial community structure identified revealed 1614 bacterial operational taxonomic units (OTUs). No significant association was found between the number of OTUs detected and prostate cancer risk group: an average of 168 OTUs (range: 64-265) in samples from non-cancer patients; 130 OTUs (range: 67-237) in samples from low/intermediate risk prostate cancer and 171 OTUs (range: 81-290) in samples from high risk and advanced prostate cancer. Many OTUs lacked assignment at levels lower than genus or family level. We found no exact matches of these unassigned OTU sequences in the NCBI dataset suggesting that post-DRE urine contains novel bacterial species. Therefore, we attempted to culture them.

We applied a fastidious anaerobic culture protocol for culturing bacteria previously considered to be "unculturable" [27] to post-DRE urine sediments, and to prostate fluid secretions obtained by squeezing the prostate after prostatectomy. Previous studies identified anaerobic bacteria in prostate tissue [28, 29] but have not fully characterised the species present. In this study, strict anaerobic culture protocols yielded 39 bacterial isolates from post-DRE urine (Supplementary Table 6) and 8 isolates from prostate cancer secretions. Assembly of whole genome sequencing

data (Illumina sequencing for all isolates and Oxford Nanopore sequencing for candidate novel species) resulted in 1-515 contigs per isolate. Most anaerobic bacterial isolates from post-DRE urine sediments were from the phyla *Firmicutes*, class *Clostridia*, including genera *Peptoniphilus, Fenollaria* and *Anaerococcus* (Supplementary Table 6). Sixteen isolates of *Propionimicrobium lymphophilium* from three different urine samples demonstrated considerable genetic variation (Supplementary Fig. 5). Prostate secretions yielded bacteria from the genera *Porphyromonas*, *Staphylococcus*, *Streptococcus* and *Cutibacterium* (Supplementary Table 6).

Higher-resolution phylogenetic analysis was obtained by aligning selected full-length ribosomal gene protein sequences from unclassified isolates to the same genes from known bacterial species [21], (Fig. 1A). This allowed us to identify four novel species (Table 2). The novel species, defined as sequence similarity less than 97% to the closest published assemblies [30], were from the phyla *Firmicutes*, (*Fenollaria sporofastidiosus* sp. nov. and *Peptoniphilus rachelemmaiella* sp.nov.), *Actinobacteria* (*Varibaculum prostatecancerukia* sp. nov.) and *Bacteroidetes (Porphyromonas bobii* sp. nov.). Further details on novel species and isolates are in Supplementary Fig. 6A-D and Supplementary Table 7.

We confirmed the presence of all four novel species in urine cell sediment samples with in-house qPCR assays (6 to 65% of samples, Supplementary Table 8). Two novel species (*Peptoniphilus* sp. nov. and *Varibaculum* sp. nov.) were detected by qPCR in prostate tissue (2.8-8.6%) and all four novel species in prostate secretions (2.8-17%). This is consistent with evidence for a prostate-urine reflux loop where there is an exchange of bacteria between the urine and prostate [3-5].

3.3 16S ribosomal amplicon sequencing identifies bacterial genera potentially associated with prostate cancer risk groups

Having identified putative novel species, we investigated 16S OTUs. Clustering on the relative abundance of the 16S OTUs at the family level from 46 men, using *k*-means on Principal Coordinates Analysis (PCoA), revealed three clusters (Fig 2A, B and Supplementary Fig. 7). The first three principal coordinates explained 49% of variance (Fig. 2A and Supplementary Fig. 7 A-C). Patients demonstrating metastases at investigation or during follow-up were overrepresented in Cluster 1 (Fig. 2A, diamonds, Supplementary Fig. 7) compared to the other two clusters ($p = 0.015$; log-rank test. Fig. 2C).

We identified eight genera with a significantly higher abundance in cluster 1 (metastatic group) relative to the rest (Supplementary Fig. 7D, Supplementary Table 9, supplementary methods). Four genera were selected for further study based on significance and value of median relative abundance of cluster 1 (supplementary methods). These were the strict anaerobes *Fenollaria/Ezakiella, Peptoniphilus*, *Porphyromonas* and *Anaerococcus*. *Fusobacterium,* another anaerobe detected in the 16S amplicon data, was also included due to growing evidence of association with the development of a range of cancers [31, 32]. Co-occurrence plots of the genera in the urine sediment 16S data demonstrated that several of these strict anaerobes are commonly found together in high-risk and advanced/metastatic disease (Supplementary Fig. 8). The five selected bacteria genera (Table 3, Supplementary Table 10) are referred to as the ABBS (Anaerobic Bacteria Biomarker Set) and includes three of the novel isolates.

3.4 Use of the Anaerobic Bacteria Biomarker Set from the prostate–urine reflux loop as a prognostic biomarker.

Two fractions were produced from processing urine: the sedimentary faction that was used to detect bacterial DNA fluorescence staining and generate 16S OTU data and a supernatant fraction that contains prostate derived extracellular vesicles. Both 16S OTU data (*n*=24) from urine sediment and RNA sequencing data of the urine extracellular vesicle supernatant fraction (*n*=25) demonstrated more clinically aggressive cancer when at least one ABBS genus was detected (*p*=0.045 & *p*=0.039 respectively; log-rank test; Fig. 3 A,B).

The 16S and RNA sequencing investigations are limited by small sample sizes and so for validation we examined a much larger and entirely independent prostate cancer tissue whole genome sequencing dataset for association of bacterial genera (Supplementary Tables 4, 11) with clinical outcome after prostatectomy (*n*=204). Although this is human cancer genome data, bacterial DNA is concomitantly sequenced if present. There is a significantly high rate of biochemical recurrence in donors with at least one of the ABBS genera (log-rank *p*=0.035, Fig. 3C).

Combining the three data sets in a meta-analysis gives a hazard ratio for disease progression of 2.60 (95% CI: 1.39-4.85; *p*=0.003; Cox proportional hazards regression; Supplementary Fig. 9). The ICGC dataset was additionally subject to a multivariable analysis including covariates: PSA at radical prostatectomy, age at diagnosis, tumour size at diagnosis and Gleason score (supplementary methods). The predicted hazard ratio for the multivariable analysis was 2.02 (95% CI: 0.97-4.2, *p*=0.061). Overall, these results indicate that detecting anaerobic bacteria that comprise the ABBS in the urinary tract may constitute a prognostic test for prostate cancer biochemical failure.

To explore common biological features of ABBS bacteria we used assemblies to predict genes

and their function. We found the following genes enriched in ABBS compared to non-ABBS isolates that are potentially relevant to cancer development (Supplementary Fig. 10): (i) components of metabolic pathways that can convert cholesterol to androstenedione, an immediate precursor for testosterone that is required for prostate cancer growth [33]; (ii) flavin-dependent (FAD) bacterial specific thymidylate synthase; (iii) a predicted citrate lyase complex (reduced citrate is a known predictor of cancer aggression in prostate cancer [34]); and (iv) the glycine cleavage complex and components of the pathway for biotin synthesis that can impact host metabolic pathways [12, 34-36]. We currently have no evidence of causality.

3. Discussion

A review published in 2019 [12], describes the association of the microbiome with prostate pathologies but concluded that major difficulties remain: sampling contamination, obtaining effective control tissue and, classifying the often-novel bacteria involved. Addressing these concerns, we implemented several improvements. First, we used protocols to minimise contamination during OTU data generation [25, 37]. Secondly, we used strictly anaerobic culture conditions (Supplementary Fig. 11) leading to the isolation of novel bacteria and then their qPCR detection in the urine and prostate. We also used short and long read DNA sequencing to obtain complete genomes for three of the novel bacteria. Overall, our results provide a more complete characterization of the urine and prostate microbiomes and provided a solid foundation for examining the relationships between the presence of specific bacteria and clinical outcome.

Several separate lines of evidence support the role of bacteria as a prognostic marker of disease progression. We demonstrated a significant correlation between increased risk groups of prostate cancer and the presence of bacteria as determined by the fluorescence microscopic

296 detection of bacterial cell DNA, both in discovery (n=215 patients, p<0.001) and validation
297 (n=103, p<0.001) datasets. Secondly, Principal Coordinate Analysis of 16S OTU data
298 identified a cluster of patients with higher incidence of metastatic disease. This observation led
299 to the development of the Anaerobic Bacteria Biomarker Set (ABBS) consisting of 5 genera of
300 strictly anaerobic bacteria (Table 3). Thirdly, analysis of RNAseq libraries prepared from the
301 extracellular vesicle fraction of urine and of OTU data from the urine sedimentary fraction
302 provided indicative results supporting the importance of the ABBS. Remarkably, ABBS
303 importance was validated by a distinct detection technology in tissue: namely by interrogating
304 a large (n=204) dataset whereby whole genome DNA sequencing captured information from
305 bacteria present in prostatectomy tumour samples. Taken together these studies provide a
306 strong case for a role of specific anaerobic bacteria (ABBS) present in the prostate-urine reflux
307 loop in predicting aggressive prostate cancer.
308
309 Recent studies undertaken by others [38, 39] and by our own laboratory [40] have provided
310 comprehensive analyses of microbiomes and viromes associated with human cancers. A
311 consistent observation is that microbiomes present in cancer tissue or blood can act as
312 diagnostic markers across multiple cancer types [38, 39]. Analyses of published datasets [38,
313 39] indicated that ABBS bacteria were also present in other cancer types (results not shown),
314 hence their relevance in determining aggression may extend beyond prostate cancer. We also
315 provide predicted functions of the ABBS, adding to previous studies on prostate cancer that
316 have investigated the association between the presence of microorganisms and inflammation
317 [14, 29], and identifying a variety of molecular mechanisms that are of potential interest for
318 tumour progression and therapeutic exploitation. Further research is needed to determine if
319 ABBS specific bacteria may cause cancer subsequently to identify potential treatment options that
320 would eradicate the anaerobic pathogens. In conclusion our results establish the importance of

321 bacteria present in urine and prostate as potential prognostic markers and, when considered
322 together with data from other studies [12-14, 16, 18], provide a starting point for future
323 investigations into the role of bacteria in prostate cancer.
324

Author contributions

Study concept and design: Hurst, Rallapalli, Clark, O' Grady, Brewer, Wain, Cooper

Acquisition of data: Hurst, Meader, Clark, Kay, Webb, Manley, Curley, Walker, Kumar, Schmidt, Hanna, Rochester, Mills, Ball

Analysis and interpretation of data: Hurst, Meader, Gihawi, Rallapalli, Clark, Kay, Crossman, Wedge, O' Grady, Brewer, Wain, Cooper. Eeles, Wedge, Lynch, Massie & The ICGC Prostate Group provided genome data from prostate cancers advised on data analysis and results.

Drafting of the manuscript: Hurst, Rallapalli, Gihawi, Brewer, Cooper

Critical revision of the manuscript for important intellectual content: Hurst, Gihawi, Rallapalli, Clark, Eeles, Wedge, Lynch, Massie, Mithen, Traka, O' Grady, Brewer, Wain, Cooper

Statistical analysis: Hurst, Gihawi, Rallapalli, Crossman, Brewer, Cooper

Obtaining funding: Hurst, Clark, O' Grady, Brewer, Wain, Cooper

Administrative, technical or material support: Curley, Webb, Walker

Supervision: Hurst, Clark, O' Grady, Brewer, Wain, Cooper

Other: None

Funding/Support and role of the sponsor: This work was funded by Prostate Cancer UK (research grant ref RIA15-ST2-029, MA-ETNA19-003), The Bob Champion Cancer Trust and Big C Cancer Charity (ref 16-09R). We are grateful for and acknowledge support from The Masonic Charitable Foundation Successor to The Grand Charity, Movember, The Prostate Cancer Research Centre, The King Family and the Stephen Hargrave Trust. We also acknowledge funding from Cancer Research UK and Dallaglio Foundation that supported the ICGC project (research grant ref. C5047/A22530). J.W. and J.O'G. are funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1. The authors confirm that the funding organisations did not inform or influence the design and conduct of

the study; collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Acknowledgements:

The authors acknowledge the staff in the Urology Department at the Norfolk and Norwich University Hospital for facilitating the research, in particular, Elizabeth Rooney, Andrea Clayton and Robert Collins. Also, the authors acknowledge Kim Findlay at the John Innes Centre Bioimaging facility for expertise with the scanning electron microscopy and to Dr Paul Thomas at the Henry Wellcome laboratory for cell imaging for advice with microscopy at UEA. We are grateful for the advice regarding the culture for NCBI entry: *Sporobacterium WAL_1855D* taxonomy ID 507843 from Dr Sydney Finegold, Nisha Patel and Dr Paul Lawson. Much of the computational reported in this manuscript was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia. The authors would especially like to thank the participants who have donated their time and samples for this study. Cooper, Brewer, Hurst, Rallapalli, Gihawi, Wain, O'Grady and Meader are coinventors on a patent application (UK Patent Application No. 2200682.9) from the University of East Anglia/UEA Enterprises Limited regarding the application of ABBS genera in prostate cancer.

Supplementary Material

Supplementary Material including Supplementary Methods, Figures, Tables and Additional Data File 1 related to this article are attached.

Take Home Message: A considerably improved characterisation of the urine and prostate microbiomes are provided including the identification of four novel bacteria. These discoveries

375 provided a platform for the identified group of five anaerobic bacterial genera called ABBS
376 associated with aggressive prostate cancer.
377
378

References:

[1] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380:2095-128.

[2] NICE. National Institute for Health and Care Excellence, NICE clinical guideline 175. Prostate cancer: diagnosis and treatment. Issued January 2014 pg.1-48. ISBN: 978-1-4731-0404-4. Guidance.nice.org.uk/cg175. 2014.

[3] Blacklock NJ. Anatomical factors in prostatitis. Br J Urol. 1974;46:47-54.

[4] Kirby RS, Lowe D, Bultitude MI, Shuttleworth KE. Intra-prostatic urinary reflux: an aetiological factor in abacterial prostatitis. Br J Urol. 1982;54:729-31.

[5] Butel R, Ball R. The distribution of BCG prostatitis: A clue for pathogenetic processes? Prostate. 2018;78:1134-9.

[6] O'Reilly E, Tuzova AV, Walsh AL, Russell NM, O'Brien O, Kelly S, et al. epiCaPture: A Urine DNA Methylation Test for Early Detection of Aggressive Prostate Cancer. JCO Precis Oncol. 2019;2019.

[7] Connell SP, Hanna M, McCarthy F, Hurst R, Webb M, Curley H, et al. A Four-Group Urine Risk Classifier for Predicting Outcome in Prostate Cancer Patients. BJU Int. 2019.

[8] Connell SP, O'Reilly E, Tuzova A, Webb M, Hurst R, Mills R, et al. Development of a multivariable risk model integrating urinary cell DNA methylation and cell-free RNA data for the detection of significant prostate cancer. Prostate. 2020;80:547-58.

[9] Tomlins SA, Day JR, Lonigro RJ, Hovelson DH, Siddiqui J, Kunju LP, et al. Urine TMPRSS2:ERG Plus PCA3 for Individualized Prostate Cancer Risk Assessment. Eur Urol. 2016;70:45-53.

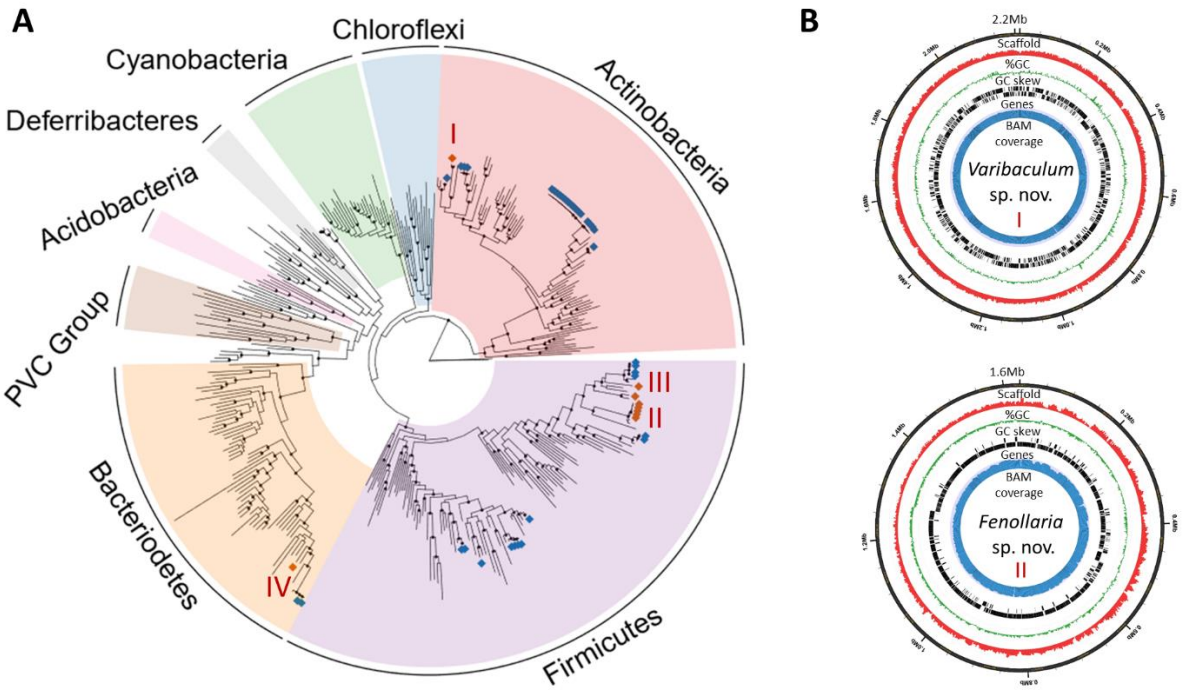
403 [10] Jones AL, Chinegwundoh F. Update on prostate cancer in black men within the UK.
404 Ecancermedicalscience. 2014;8:455.
405 [11] Hjelmborg JB, Scheike T, Holst K, Skytthe A, Penney KL, Graff RE, et al. The
406 heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol
407 Biomarkers Prev. 2014;23:2303-10.
408 [12] Wheeler KM, Liss MA. The Microbiome and Prostate Cancer Risk. Curr Urol Rep.
409 2019;20:66.
410 [13] Shrestha E, White JR, Yu SH, Kulac I, Ertunc O, De Marzo AM, et al. Profiling the
411 Urinary Microbiome in Men with Positive versus Negative Biopsies for Prostate Cancer. J
412 Urol. 2018;199:161-71.
413 [14] Puhr M, De Marzo A, Isaacs W, Lucia MS, Sfanos K, Yegnasubramanian S, et al.
414 Inflammation, Microbiota, and Prostate Cancer. Eur Urol Focus. 2016;2:374-82.
415 [15] Chang AH, Parsonnet J. Role of bacteria in oncogenesis. Clin Microbiol Rev.
416 2010;23:837-57.
417 [16] Cavarretta I, Ferrarese R, Cazzaniga W, Saita D, Luciano R, Ceresola ER, et al. The
418 Microbiome of the Prostate Tumor Microenvironment. Eur Urol. 2017;72:625-31.
419 [17] Banerjee S, Alwine JC, Wei Z, Tian T, Shih N, Sperling C, et al. Microbiome signatures
420 in prostate cancer. Carcinogenesis. 2019;40:749-64.
421 [18] Liss MA, White JR, Goros M, Gelfond J, Leach R, Johnson-Pais T, et al. Metabolic
422 Biosynthesis Pathways Identified from Fecal Microbiome Associated with Prostate Cancer.
423 Eur Urol. 2018;74:575-82.
424 [19] Sha S, Ni L, Stefil M, Dixon M, Mouraviev V. The human gastrointestinal microbiota
425 and prostate cancer development and treatment. Investig Clin Urol. 2020;61:S43-S50.
426 [20] Alanee S, El-Zawahry A, Dynda D, Dabaja A, McVary K, Karr M, et al. A prospective
427 study to examine the association of the urinary and fecal microbiota with prostate cancer

428 diagnosis after transrectal biopsy of the prostate using 16sRNA gene analysis. Prostate.
429 2019;79:81-7.
430 [21] Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new
431 view of the tree of life. Nat Microbiol. 2016;1:16048.
432 [22] D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, et al.
433 Biochemical outcome after radical prostatectomy, external beam radiation therapy, or
434 interstitial radiation therapy for clinically localized prostate cancer. JAMA. 1998;280:969-74.
435 [23] Yu Z, Morrison M. Improved extraction of PCR-quality community DNA from digesta
436 and fecal samples. Biotechniques. 2004;36:808-12.
437 [24] Gihawi A, Rallapalli G, Hurst R, Cooper CS, Leggett RM, Brewer DS. SEPATH:
438 benchmarking the search for pathogens in human tissue whole genome sequence data leads to
439 template pipelines. Genome Biol. 2019;20:208.
440 [25] Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and
441 laboratory contamination can critically impact sequence-based microbiome analyses. BMC
442 Biol. 2014;12:87.
443 [26] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al.
444 QIIME allows analysis of high-throughput community sequencing data. Nature Methods.
445 2010;7:335.
446 [27] Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culturing
447 of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. Nature.
448 2016;533:543-6.
449 [28] Cooper PN, Millar MR, Godwin PG. Anaerobes and carcinoma of the prostate. Br Med J
450 (Clin Res Ed). 1988;296:466-7.
451 [29] Sfanos KS, Sauvageot J, Fedor HL, Dick JD, De Marzo AM, Isaacs WB. A molecular
452 analysis of prokaryotic and viral DNA sequences in prostate tissue from patients with

453 prostate cancer indicates the presence of multiple and diverse microorganisms. Prostate.
454 2008;68:306-20.
455 [30] Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining
456 operational taxonomic units and estimating species richness. Appl Environ Microbiol.
457 2005;71:1501-6.
458 [31] Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-
459 Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various
460 Stages of Cervical Cancer: A Pilot Study. PLoS One. 2016;11:e0153274.
461 [32] Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, et al. Fusobacterium
462 nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and
463 disease outcome. Eur J Clin Microbiol Infect Dis. 2014;33:1381-90.
464 [33] Attard G, Reid AH, Yap TA, Raynaud F, Dowsett M, Settatree S, et al. Phase I clinical
465 trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant
466 prostate cancer commonly remains hormone driven. J Clin Oncol. 2008;26:4563-71.
467 [34] Kdadra M, Hockner S, Leung H, Kremer W, Schiffer E. Metabolomics Biomarkers of
468 Prostate Cancer: A Systematic Review. Diagnostics (Basel). 2019;9.
469 [35] Locasale JW. Serine, glycine and one-carbon units: cancer metabolism in full circle.
470 Nature Reviews Cancer. 2013;13:572-83.
471 [36] Lin S, Hanson RE, Cronan JE. Biotin synthesis begins by hijacking the fatty acid
472 synthetic pathway. Nat Chem Biol. 2010;6:682-8.
473 [37] Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in
474 Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends
475 Microbiol. 2019;27:105-17.

476 [38] Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, et al. The human tumor
477 microbiome is composed of tumor type-specific intracellular bacteria. Science.
478 2020;368:973-80.
479 [39] Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome
480 analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020;579:567-74.
481 [40] Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape
482 of viral associations in human cancers. Nat Genet. 2020;52:320-30.
483

484



485

486 **Fig. 1 - Phylogenetic tree and novel bacteria**. (A) Cultured fastidious anaerobes were isolated from urine and
487 prostate and their genomes were decoded using Illumina and Nanopore DNA sequencing. Each bacterial strain
488 was positioned on the phylogenetic tree as described in Methods. Bacteria with a known ID are highlighted with
489 blue diamonds, while novel species are highlighted with orange diamonds (I: *Varibaculum prostatecancerukia*
490 sp. nov., II: *Fenollaria sporofastidiosus* sp. nov., III: *Peptoniphilus rachelemmaiella* sp. nov. and IV:
491 *Porphyromonas bobii* sp. nov.) (B) Genome representation for two of the novel species. Data for : I, *Varibaculum*
492 sp. nov. isolate 39, 2.2Mb, GC content, 53%; II, *Fenollaria* sp. nov. isolate 24, 1.6Mb, GC content, 36%; III,
493 *Peptoniphilus* sp. nov. isolate 23, 1.9Mb, GC content 49%, and IV, *Porphyromonas* sp. nov. (isolate 6C, 2.2Mb,
494 GC content 56%) amongst other isolates are in Supplementary Tables 6, 7, Fig. 6A-D. PVC: *Planctomycetes*,
495 *Verrucomicrobiae*, *Chlamydiae* group.

496

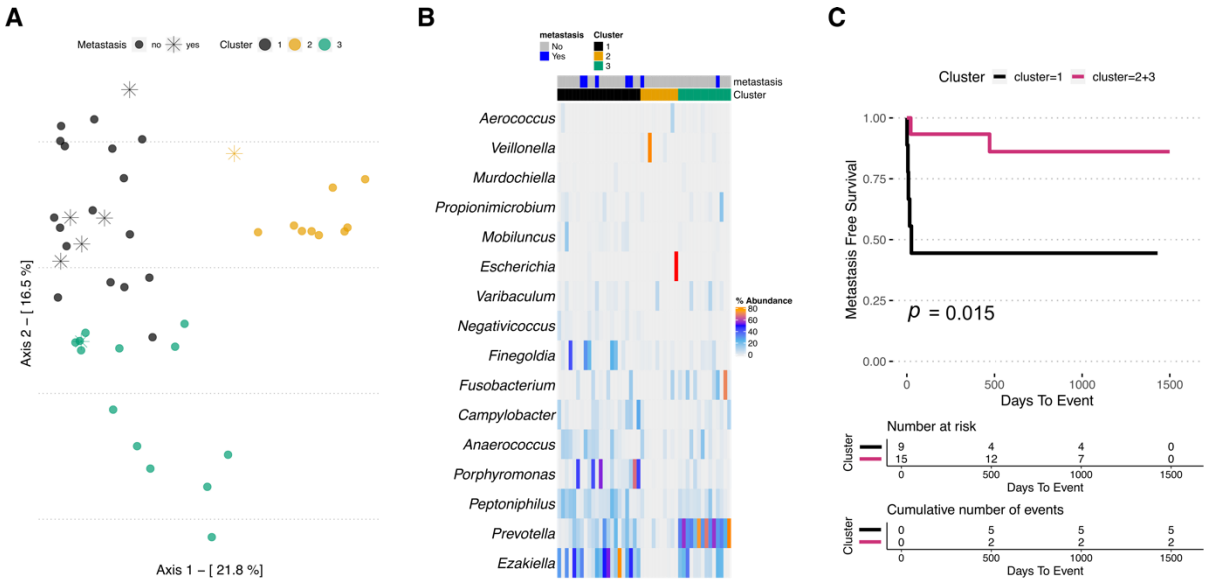


Fig. 2 - Presence and composition of urine microbiota identify participants with a poorer prognosis. Analysis of 16S OTU sequence from urine sediments (A-C). (A) Principal Coordinate Analysis (PCoA, Manhattan distance) of family level OTU data from urine sediments from 46 patients undergoing assessment for prostate cancer. Clustering with *k*-means suggested three clusters: Cluster 1 (black), Cluster 2 (yellow) and Cluster 3 (green). Samples from patients that developed skeletal metastases are indicated with diamonds. (B) Heatmap demonstrating a variety of bacterial genera selected to demonstrate differences across the 3 family level clusters. (C) Kaplan-Meier analysis investigating metastasis free survival: cluster one (black); clusters two plus three (pink).

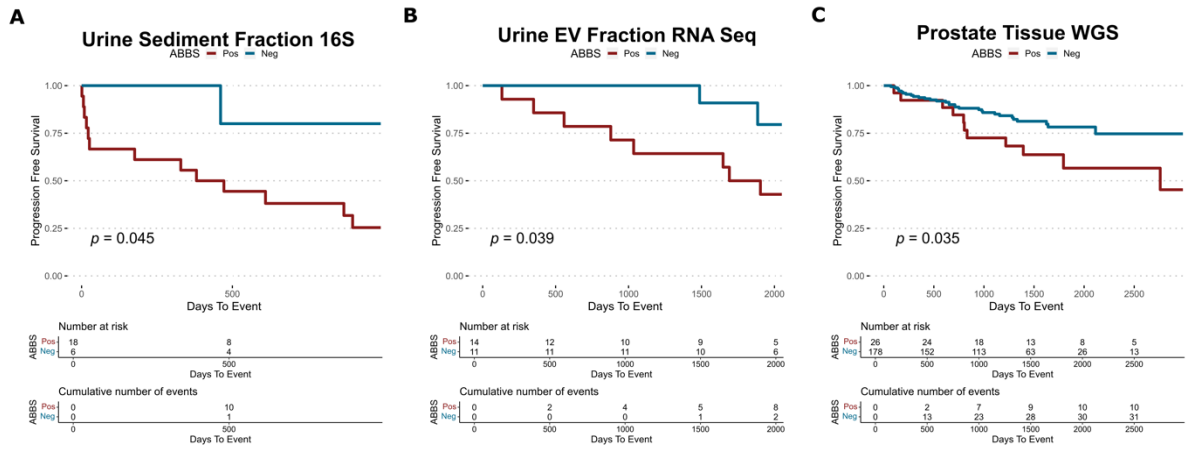


Fig. 3 - Anaerobic Bacteria Biomarker Set from the prostate–urine reflux loop are significantly associated with more rapid progression. Kaplan-Meier analysis investigating progression free survival: (A-C) The presence of the following genera was used to partition sample sets: *Fenollaria* (including hits to *Ezakiella* due to closely

related 16S sequences to *Fenollaria)*, *Peptoniphilus*, *Porphyromonas*, *Anaerococcus* and *Fusobacterium*. These genera collectively constitute the anaerobic bacteria biomarker set (ABBS) in this manuscript. Data were from: (A) urine cell sediment fraction 16S sequencing (14 total events, 13 from $ABBS^+$ samples, 1 from $ABBS^-$ samples); (B) RNA sequencing of urine extracellular vesicle (EV) fraction (13 total events, 11 from $ABBS^+$ samples, 2 from $ABBS^-$ samples); and (C) whole human genome ICGC DNA sequence data from prostate cancer tissue (WGS) (42 total events, 12 from $ABBS^+$ samples and 31 from $ABBS^-$ samples.. All "*p*" values are calculated using the log-rank test. Curves were truncated at the point where there were ten remaining samples overall and each curve to ensure a minimum of four samples remaining. Univariate cox proportional hazards models are summarised in supplementary Fig. 9. The hazard ratios (95% confidence intervals; *p* values) are respectively as follows: 6.18 (95% CI: 0.81-47.33; p=0.023), 4.41 (95% CI: 0.95-20.53; p=0.059) and 2.07 (95% CI: 1.04-4.15; p=0.040). A meta-analysis of these three models gave HR 2.60 (95% CI: 1.39-4.85; p=0.003). A forest plot of this meta-analysis is available in supplementary Fig. 9.

Table. 1 | Presence of bacteria in urine cell sediment and association with increased risk groups of prostate cancer.

(A) Discovery set.

Category	Percentage Positive for microorganisms	Negative count	Positive Count
Normal PSA Range	31%	18	8
Low Risk PCa	46%	6	5
Intermediate Risk PCa	64%	20	36
High Risk PCa	88%	4	29
Advanced PCa	83%	2	10
Atypia/HG-PIN	26%	14	5
Raised PSA negative biopsy	40%	35	23

(B) Validation set.

Category	Percentage Positive for microorganisms	Negative count	Positive Count
Normal PSA Range	23%	10	3
Low Risk PCa	17%	5	1
Intermediate Risk PCa	77%	6	20
High Risk PCa	75%	4	12
Advanced PCa	100%	0	6
Atypia/HG-PIN	30%	7	3
Raised PSA negative biopsy	27%	19	7

Significant correlation of clinical D'Amico risk group and advanced disease (PSA>100ng/mL) with the presence of background DAPI stained bacteria fluorescence in post-DRE urine. Data are presented as percentage positive for bacteria/microorganisms. Samples examined by microscopy were divided into (A) a discovery (two-thirds of samples; n=215) and (B) validation dataset (one-third of samples; n=103) by random assignment, stratified by clinical group. Discovery set (χ^2 test for trend in proportions, p<0.001), Validation set (χ^2 test for trend in proportions, positive trend, p<0.001). Statistical analyses were performed on the first 5 groups. Data on Atypia/HG-PIN and Raised PSA negative biopsy were included for comparison.

541

Table 2 | Novel bacteria species isolated

Novel Species Isolated Bacteria ID	Novel species belonging to Phyla	Novel species belonging to Class	Novel species belonging to Genus	Reference Novel Strain ID
Fenollaria sp. nov.	*Firmicutes*	*Clostridia*	*Fenollaria*	*Fenollaria sporofastidiosus* sp. nov. (EMRHCC_24)
Peptoniphilus sp. nov.	*Firmicutes*	*Clostridia*	*Peptoniphilus*	*Peptoniphilus rachelemmaiella* sp. nov. (EMRHCC_23)
Varibaculum sp. nov.	*Actinobacteria*	*Actinobacteria*	*Varibaculum*	*Varibaculum prostatecancerukia* sp. nov. (EMRHCC_39)
Porphyromonas sp. nov.	*Bacteroidetes*	*Bacteroidia*	*Porphyromonas*	*Porphyromonas bobii* sp. nov. (EMRHCC_6C)

Novel species isolated from clinical samples, including taxonomy and proposed new species name, novel strain ID. *Fenollaria* sp.nov., *Peptoniphilus* sp. nov. and *Varibaculum* sp. nov. were isolated from urine and *Porphyromonas* sp. nov. from prostate secretion fluid (further details regarding the novel species are provided in Supplementary Table 7 and Supplementary Figure 6).

548 **Table 3 | Anaerobic Bacteria Biomarker Set (ABBS). Bacteria associated with poor**

549 **prognosis.**

Anaerobic Bacteria Biomarker Set: ABBS bacteria genera	ABBS belonging to Phylum	ABBS belonging to Class	ABBS belonging to Order	Novel Species and known species isolated by anaerobic culture belonging to ABBS genera
Fenollaria	*Firmicutes*	*Clostridia*	*Clostridiales*	*Fenollaria* sp. nov.◆
Peptoniphilus	*Firmicutes*	*Clostridia*	*Clostridiales*	*Peptoniphilus* sp. nov.◆; *Peptoniphilus harei*
Anaerococcus	*Firmicutes*	*Clostridia*	*Clostridiales*	*Anaerococcus prevotii*
Porphyromonas	*Bacteroidetes*	*Bacteroidia*	*Bacteroidales*	*Porphyromonas* sp. nov.◆; *Porphyromonas asaccharolytica*
Fusobacterium	*Fusobacteria*	*Fusobacteriia*	*Fusobacteriales*	*Fusobacterium nucleatum*

550 ABBS bacteria taxonomy, novel species isolated in this study (◆)