Rare germline variants are associated with rapid biochemical recurrence after radical prostate cancer treatment: a PPCG study

Keywords: germline variants; prostate cancer; biochemical recurrence; PPCG

Word Counts:

Abstract: 286

Text (including abstract): 3533

Corresponding Author:

Zsofia Kote-Jarai
The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP, UK
Zsofia.Kote-Jarai@icr.ac.uk

Authors:

Daniel Burns[a], Ezequiel Anokian[a], Edward J Saunders[a], Robert G Bristow[b], Michael Fraser[c,d], Jüri Reimand[d,e], Thorsten Schlomm[f], Guido Sauter[g], Benedikt Brors[h], Jan Korbel[i], Joachim Weischenfeldt[f,j], Sebastian M Waszak[k,al,am], Niall M Corcoran[l,m,n], Chol-Hee Jung[o], Bernard J Pope[l,p], Chris M Hovens[o,q,n], Géraldine Cancel-Tassin[r,s], Olivier Cussenot[r,s], Massimo Loda[t], Chris Sander[u], Vanessa M Hayes[v,w], Karina Dalsgaard Sorensen[x], Yong-Jie Lu[y], Freddie C Hamdy[z], Christopher S Foster[a],

Vincent Gnanapragasam[aa], Adam Butler[ab], Andy G Lynch[ac,ad], Charlie E Massie[ae], CR-UK/Prostate Cancer UK, ICGC, The PPCG[af], Dan J Woodcock[ag], Colin S Cooper[ah], David C Wedge[ai], Daniel S Brewer[ah,aj,*], Zsofia Kote-Jarai[a,*,&], Rosalind A Eeles[a,ak,*]

[*] Joint last authors

[&] corresponding author

[a]The Institute of Cancer Research, 123 Old Brompton Road, London, SW7 3RP, UK

[b]Manchester Cancer Research Centre and CRUK Manchester Institute, The University of Manchester, 555 Wilmslow Rd, Manchester, M20 4GJ, UK

[c]Princess Margaret Cancer Centre/University Health Network, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada

[d]Computational Biology Program, Ontario Institute for Cancer Research, 661 University Avenue, Toronto, Ontario, M5G 0A3, Canada

[e]Department of Medical Biophysics & Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

[f]Charité - Universitätsmedizin Berlin, Charitéplatz 1, Berlin, 10117, Germany

[g]University Medical Centre Hamburg - Eppendorf, Martinistrasse 52, Hamburg, 20246, Germany

[h]German Cancer Research Center (DKFZ), Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, Heidelberg, D-9120, Germany

[i]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg, 69117, Germany

[j]Biotech Research & Innovation Centre (BRIC) & Finsen Laboratory, University of

Copenhagen, Rigshospitalet, Blegdamsvej 9, Copenhagen, 2100, Denmark

[k]Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo and Oslo University Hospital, Oslo, 0349, Norway

[l]Department of Surgery, The University of Melbourne, Grattan Street, Parkville, Victoria, 3010, Australia

[m]Department of Urology, Royal Melbourne Hospital, Parkville, Victoria, 3050, Australia

[n]University of Melbourne Centre for Cancer Research, The Victorian Comprehensive Cancer Centre, Parkville, Victoria, 3000, Australia

[o]The University of Melbourne, Grattan Street, Parkville, Victoria, 3010, Australia

[p]Melbourne Bioinformatics, The University of Melbourne, Grattan Street, Victoria, 3010, Australia

[q]Royal Melbourne Hospital, Parkville, Victoria, 3050, Australia

[r]CeRePP, Hopital Tenon, 4 rue de la Chine, Paris, 75020, France

[s]Sorbonne Université, GRC n°5 Predictive Onco-Urology, APHP, Tenon Hospital, 4 rue de la Chine, Paris, 75020, France

[t]Weill Cornell Medicine, Department of Pathology & Laboratory Medicine, 1300 York Avenue, New York, NY10065, USA

[u]cBio Center, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts, MA02215, USA

[v]Garvan Institute of Medical Research, The Kinghorn Cancer Centre, 370 Victoria Street, Darlinghurst, NSW, 2010, Australia

[w]School of Medical Sciences, University of Sydney, Charles Perkins Centre, Camperdown, NSW, 2050, Australia

[x]Department of Molecular Medicine, Aarhus University Hospital & Department of

Clinical Medicine, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark

[y]Centre for Biomarker and Therapeutics, Barts Cancer Institute, Queen Mary University of London, John Vane Science Building, Charterhouse Square, London, EC1M 6BQ, UK

[z]Nuffield Department of Surgical Sciences University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU

[aa]Department of Surgery, Division of Urology, University of Cambridge, Cambridge Biomedical Campus, Addenbrooke's Hospital Site, S Wards Building, Keith Day Road, Cambridge, CB2 0SL, UK

[ab]Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

[ac]School of Medicine, University of St Andrews, St Andrews, Fife, KY16 9TF, UK

[ad]School of Mathematics & Statistics, St Andrews, Fife, KY16 9SS, UK

[ae]CRUK Cambridge Institute, Hutchison MRC Research Centre, University of Cambridge, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK

[af]CR-UK/Prostate Cancer UK, ICGC, The Pan Prostate Cancer Group

[ag]Nuffield Department of Surgical Sciences University of Oxford, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK

[ah]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, NR4 7JT, UK

[ai]Manchester Cancer Research Centre, The University of Manchester, 555 Wilmslow Rd, Manchester, M20 4GJ, UK

[aj]The Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK

[ak]The Royal Marsden NHS Foundation Trust, London, UK

[al]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

[am]Department of Pediatric Research, Division of Pediatric and Adolescent Medicine, Rikshospitalet, Oslo University Hospital, Oslo, Norway

ABSTRACT

Background: Germline variants explain more than a third of prostate cancer (PrCa) risk, but very few associations have been identified between heritable factors and clinical progression.

Objective: To find rare germline variants that predict time to biochemical recurrence (BCR) after radical treatment in men with PrCa, and understand the genetic factors associated with such progression**.**

Design, Setting and Participants: Whole-genome sequencing data from blood DNA were analysed for 850 PrCa patients with radical treatment from the Pan Prostate Cancer Group (PPCG consortium) from UK, Canada, Germany, Australia and France. Findings were validated using 383 patients from The Cancer Genome Atlas (TCGA).

Outcome Measurements and Statistical analysis: 15,822 rare (MAF<1%) predicted-deleterious coding germline mutations were identified. Optimal multifactor and univariate *Cox* regression models were built to predict time to BCR after radical treatment, using germline variants grouped by functionally annotated gene-sets. Models were tested for robustness using bootstrap resampling.

Results: Optimal *Cox* regression multifactor models showed that rare predicted-deleterious germline variants in “Hallmark” gene-sets were consistently associated with altered time to BCR. Three gene-sets had a statistically significant association with risk-elevated outcome when modelling all samples: PI3K/AKT/mTOR, Inflammatory response and KRAS signalling (up). PI3K/AKT/mTOR and KRAS signalling (up) were also associated among patients with higher grade cancer, as were

Pancreas-beta cells, TNFA signalling via NKFB and Hypoxia, the latter of which was validated in the independent TCGA dataset.

Conclusions: We demonstrate for the first time that rare deleterious coding germline variants robustly associate with time to BCR after radical treatment, including cohort-independent validation. Our findings suggest that germline testing at diagnosis could aid clinical decisions by stratifying patients for differential clinical management.

Patient summary: PrCa patients with particular genetic mutations have a higher chance of relapsing after initial radical treatment, potentially providing opportunities to identify which patients might need additional treatments earlier.

Introduction

Prostate cancer (PrCa) is the most common cancer in men in the developed world. Although the majority of PrCa cases are diagnosed with low or intermediate risk disease, approximately 10% of patients develop metastatic disease with poor survival rates [1, 2]. Genetic predisposition to the overall disease risk of PrCa of any severity is well researched; however, understanding of potential heritable genetic factors contributing to tumor progression is limited [3].

Biochemical recurrence (BCR) is often used as a prostate-specific antigen (PSA)-based predictor of progression to poor prognosis phenotype, and is observed in approximately 25% of patients after radical prostatectomy (RP) [4]. Identification of men at high-risk for progression to lethal disease and who are likely to relapse after primary treatment would present the possibility to triage treatment intensification using current or novel systemic therapies. Most research into BCR to date has focused on gene expression or mutational signatures in prostate tumour tissue, or specific candidate genes only [5]. In this study, we investigate for the first time whether rare germline variants across the full exome are predictive of poor prognosis after radical treatment. This information could aid clinical management of the disease, particularly at diagnosis, pre- or post-treatment staging and prognostication.

Materials and Methods

Sequencing of DNA from PrCa Patients

Whole-genome sequencing (WGS) data derived from whole blood samples were collated for PrCa patients from member countries of the Pan Prostate Cancer Group (PPCG, http://panprostate.org; Australia n=133, Canada n=288, France n=15, Germany n=230, UK n=184; *Table 1*, further characteristics in *Supplementary Table 1*). The study presented here combines data from patients following RP, and a small subset of samples with radical radiotherapy (RT; 8%) from the Canadian study group. We refer to the samples collectively as having radical treatment.

Samples were collected according to criteria outlined in *Supplementary Method 1.* Collection was subject to the International Cancer Genome Consortium (ICGC) standards of ethical consent. Collection and analysis of the Australian samples received institutional review board approval (Epworth Health 34506; Melbourne Health 2019.058). WGS was performed using Illumina technology to ≥30x depth. Burrows-Wheeler Aligner (BWA, [6]) was used to align sequencing data to the GRCh37 human genome (human_g1k_v37) with PCR duplicates removed [7]. Sequencing data have been deposited at the European Genome-phenome Archive (https://ega-archive.org, study IDs in *Table 1*) and is available upon request.

Variant Calling

Variant calling was performed with *The Genome Analysis Toolkit* pipeline (GATK v4.0) [8] following GATK best practice recommendations for germline SNV and indel calling [9, 10] (*Supplementary Method 2)*, apart from for the German samples which were

called using *FreeBayes* v1.1.0 [11] and processed as described by Gerhauser *et al.* [12], normalised with vt v0.5 *[13]* (*Supplementary Method 3*). This analysis was restricted to variants within protein-coding transcript sequences according to GENCODE v29 [14].

Quality Control, Variant Annotation and Prioritization

Low-quality variants and samples were removed based on established QC protocols [15-17]. We excluded samples from related individuals (using R package *SNPRelate* method identity-by-descent [18]), or with non-European ancestry (using Principal Component Analysis relative to 2,504 samples from the 1000 Genomes Project [19]). We used *Picard tools* v2.23.8 [20] to remove samples with a mean insert size <250bp, AT or GC dropout >5%, <95% aligned reads, >5% mismatch rate, <80% with ≥20x coverage or >5% missing call rate. Using *verifyBamID* v1.1.2 [21] we removed samples with >3% sample contamination. We excluded variants with a missing call rate in >5% of the samples, monomorphic loci, those in repetitive regions (simple repeats, segmental duplications and centromeric regions) and where the ExAC major allele frequency in any population was >1%. 3% of the submitted samples were excluded based on ancestry, while 2% were removed because of sequencing quality. One sample was removed due to relatedness.

Post-QC variants were annotated using the *Variant Effect Predictor* (VEP v101) and loss-of-function transcript effect estimator (LOFTEE) package [22]. For downstream analyses we retained only variants categorised as deleterious/loss-of-function, comprising those with protein-truncating mutations (nonsense, frameshift and splice site variants) occurring in the first 95% of the protein, as well as predicted-deleterious missense variants with a CADD PHRED score >30 [23] (*Table 1*).

Pathways and Gene-sets

For pathway level analysis, all 50 “Hallmark” gene-sets from GSEA MsigDB were considered [24 (Downloaded April 2017)], along with the BROCA extended panel of 66 genes and 175 curated DNA repair genes (DRG) [16, 17] (*Supplementary Table 2*).

Statistical analysis

Software and libraries

All statistical analyses were applied using *Python* v3.8 [25]. Data in VCF format was converted using *PyVCF* v0.6.7 [26] and processed using *pandas* v1.3.0 [27], *SciPy* v1.4.1 [28], *NumPy* v1.18.3 [29], *IPython* v7.14 [30] and *Scikits.bootstrap* v1.1 [31]. Survival analysis for *Cox*’s proportional hazard (PH) model and *Kaplan-Meier* estimates were performed using the *Lifelines* v0.25 package [32]. Tables and graphs were output using *Matplolib* v3.3.4 [33], *to_precision* [34] and *Maftools* v2.6.5 [35].

Multifactor *Cox* Regression

Analyses were performed on the combined post-QC dataset (*Table 1*) and a subset of patients with high Gleason score tumours, with models stratified by study to compensate for differing baseline hazards. Gene-set predictors of the *Cox* Proportional Hazard (PH) model were generated by recording the presence of any gene with predicted-deleterious mutations in the selected gene-sets across all samples. Pathologic T-stage had a baseline of stage 1-2, and a second group for stage

3-4. Clinical T-stage was used for patients receiving radiotherapy (RT). Pre-operative PSA and age at time of surgery were continuous variables. Gleason score had a baseline of ≤3+4 (Gleason grade groups 1-2), and a group for ≥4+3 (Gleason grade group 3-5). Time was measured from radical treatment until BCR, which for samples with radical prostatectomy (RP) was defined as two consecutive post-RP PSA measurements of >0.2ng/ml on the last known follow-up date [36]. For the 72 Canadian samples with RT, BCR was defined as a rise in PSA concentration of more than 2.0 ng/ml above the nadir, backdated to first PSA>0.2 ng/ml if PSA continues to rise [37]. We performed a sensitivity analysis on a subset that excluded RT samples, which did not affect the significant risk-elevating gene-sets observed (*Supplementary Table 3*).

Variables included in the final models were selected by performing *Cox* regression with penalization based on the least absolute shrinkage and selection operator (LASSO) [38]. The optimal penalty factor (lambda) was determined as within 1 standard error of the optimum from the mean of 100 ten-fold cross-validation models. Only features with a non-zero coefficient were retained. The final prediction models were then built using *Cox* regression without penalization.

Univariate *Cox* regression

Each gene-set was modelled individually along with clinical covariates (pre-op PSA, pathologic T-stage, Gleason score, age), and *p*-values were adjusted for multiple testing using False Discovery Rate (FDR).

Validation

We performed harmonised variant filtering for predicted-deleterious mutations on germline PrCa samples from The Cancer Genome Atlas (TCGA) PRAD project. From the original 500 TCGA PRAD samples, any samples from contributing institutions with <15 samples were excluded, and models were stratified by institution, resulting in 383 samples used in the analysis. Of those, 233 were included in the high-Gleason subset analysis. We applied the variants to the predictors selected from the *Cox* model built using the combined PPCG samples, to compare the hazard ratios (HR) in both sets.

Kaplan-Meier analysis

A KM-plot measuring time to BCR in the event of relapse was used to visualise the impact of mutations within significant gene-sets on risk of BCR. This was applied separately to the whole dataset and high-Gleason subset, and reported alongside log-rank test *p*-values.

We performed a combined analysis, considering mutations in any of the gene-sets significant for the corresponding analyses, and subdivided to ascertain potential additive effects upon a patient's time to relapse.

Bootstrapping validation

To test model robustness, we produced new datasets of the same sample size by randomly choosing samples with replacement, without stratification, and building a *Cox* regression model from the resulting dataset. This was repeated 1000 times to derive a distribution of coefficients. *p*-values were computed for each predictor as a

percentage of the iterations where the coefficient was in a different direction than expected.

Results

We analysed germline WGS data from 850 patients across five studies in the PPCG consortium (*Table 1; Supplementary Tables 1 and 2*) for germline predictors of PrCa progression measured by BCR after radical treatment. This analysis was restricted to variants within protein-coding transcript sequences, resulting in 15,822 rare variants identified as deleterious or likely deleterious, jointly categorised as predicted-deleterious (PD). No individual variants or genes demonstrated significant association with time to BCR (*Cox* regression analysis; *p*-values >0.05), although the available sample size of 850 cases is underpowered for such analysis. Therefore, we focused on finding gene-sets or pathways with significant associations, to identify potential biological mechanisms linked with progression. To this end, we determined whether there was at least one predicted-deleterious germline alteration in 52 gene-sets, including 50 “Hallmark” gene-sets from the MsigDB database [24], containing over 4000 genes with sets varying in size from 30-200, the DRG panel containing 175 DNA repair genes [16], and the extended BROCA gene panel containing 65 genes [17] (*Supplementary Table 4*).

After variable selection by LASSO, the optimal model for predicting time to BCR contained fourteen gene-sets, three of which were significantly associated with time to BCR (*Cox* PH model for all samples; *p*-value threshold <0.05; *Table 2* and *Figure 1a*). Clinical variables at the time of radical treatment (pre-op PSA, pathological T-stage, age and Gleason score) were added to the model as covariates. The significant risk-elevating Hallmarks were PI3K/AKT/mTOR (HR=1.55; 1.06-2.25 95% CI; *p*=0.023~~26~~), Inflammatory response (HR=1.35; 1.00-1.82 95% CI; *p*=0.048~~3~~) and KRAS signalling (up) (HR=1.35; 1.01-1.79 95% CI; *p*=0.041~~3~~). These gene-sets are

associated with shortened average time to BCR. The UV response (dn) (HR=0.71; 0.51-0.99 95% CI; *p*=0.04218) and Cholesterol homestasis (HR=0.58; 0.34-1.00 95% CI; *p*=0.0483) gene-sets were borderline significantly protective. Applying this model to multiple bootstrap re-samplings showed that these results are robust, with all risk-elevating gene-sets HR>1 in >97% of resamples and *p*-values indicating the same coefficient direction.

The clinical covariates-only model built using all the samples determined that Gleason score, preop-PSA, age and pathological T-stage significantly associate with time to BCR (*Cox* PH; *p*-value threshold <0.05; *Supplementary Table 5*). This model is significantly improved by the addition of the selected gene-sets (likelihood ratio test *p*=0.04877; c-index 0.68 vs 0.66).

Within the PPCG set, patients presenting with higher-grade localised PrCa (a subset of 336 patients where Gleason score was 4+3 or higher; Gleason grade group 3-5) had a higher proportion of BCR events (50.2% compared to 33.5% for all samples; *Table 1*). We developed an optimal multifactor *Cox* regression model (*Cox* PH; *p*-value threshold <0.05; *Table 3* and *Figure 1b)* for this subset of high-Gleason samples with poorer prognosis disease. After feature selection by LASSO, we identified five significant risk-elevating gene-sets: Pancreas-beta cells (HR=2.52; 1.01-6.29 95% CI; *p*=0.0470), PI3K/AKT/mTOR signalling (HR=1.95; 1.21-3.15 95% CI; *p*=0.0065.91x10^{-3}), TNFA signalling via NFKB (HR=1.79; 1.19-2.68 95% CI; *p*=0.0054.85x10^{-3}), Hypoxia (HR=1.73; 1.14-2.63 95% CI; *p*=0.0101) and KRAS signalling (up) (HR=1.58; 1.08-2.32 95% CI; *p*=0.01989). PI3K/AKT/mTOR has a higher HR and lower *p*-value than in the all samples model. The Glycolysis gene-set shows here as significantly protective (HR=0.60; 0.40-0.91 95% CI; *p*=0.017766). The

bootstrap re-samplings for the significant gene-sets have the same coefficient direction in >96% of resamples.

Examining each gene-set in individual univariate models with all samples, none had a significant association with outcome after multiple testing correction (FDR; *p*-value threshold <0.1; *Supplementary Table 6*). PI3K/AKT/mTOR signalling (*q*=0.143), KRAS signalling (up) (*q*=0.203) and TNFA signalling via NKFB (*q*=0.1657) had *p*-values close to the significance threshold, and achieve the threshold of significance in the high-Gleason subset (*Table 3*). In the high-Gleason subset, performing a log-rank test on each gene-set revealed four gene-sets that had a significant association with time to BCR: TNFA signalling via NFKB (*p*=0.0272), PI3K/AKT/mTOR signalling (*p*=0.02548), KRAS signalling (up) (*p*=0.0132) and Pancreas-beta cells (*p*=0.0233). In the multifactor high-Gleason *Cox* model these four gene-sets are also statistically significant (*Table 3*), alongside Hypoxia.

Applying the all sample *Cox* multifactor model to the TCGA validation set results in two significant gene-set predictors that are not reflected in the PPCG data: Myc targets v2 (HR=4.46; 1.73-11.5 95% CI; *p*=0.0021.99x10^{-3}) and Coagulation (HR=3.49; 1.47-8.30 95% CI; *p*=0.0054.64x10^{-3}) (*Cox* PH; *p*-value threshold <0.05; *Supplementary Table 7*). Performing the same high-Gleason filtering on TCGA samples and applying that set to the high-Gleason PPCG model identifies three significant risk-elevating predictors: Myc targets v2 (HR=2.90; 1.00-8.40 95% CI; *p*=0.0492) and Coagulation (HR=3.53; 1.30-9.59 95% CI; *p*=0.01435), and additionally Hypoxia (HR=3.18; 1.04-9.74 95% CI; *p*=0.04325) (*Cox* PH; *p*-value threshold <0.05; *Table 4* and *Figure 1c*). The consistent significance, and same direction of coefficient of Hypoxia in patients

with more advanced disease, is compelling evidence that germline variations in genes within this pathway contributes to clinical progression.

We used *Kaplan-Meier* plots to visualise the additive effect of mutations in the corresponding risk-elevating gene-sets for the all samples and high-Gleason sets (*Figure 2*). In both plots we show a significant difference in survival when multiple gene-sets carry predicted-deleterious mutations. In the all samples analysis 285 of 850 patients had a mutation in one significant gene-set and 58 patients had mutations in two or more gene-sets, whilst in the high-Gleason subset analysis, 129 of 336 patients had a mutation in one significant gene-set, 36 patients had mutations in two gene-sets and 12 had mutations in three or more gene-sets, which was the clearest detrimental impact (*Figure 2B).*

To search for individual genes mutated more frequently in patients with BCR, we calculated the odds ratio (OR) between the BCR positive and negative groups (*Supplementary Table 8*). 12 genes within the significant gene-sets for all samples (*PIKFYVE, MYD88, CAB39, RPS6KA1, IRAK2, IL2RB, MSR1, ITGB8, PIK3R5, MMP10, HKDC1, RBM4*) and 17 genes within the significant gene-sets in the high-Gleason subset (*GAPDHS, GRHPR, PGM1, SELENBP1, NAGK, SLC6A6, PIKFYVE, MYD88, CAB39, RPS6KA1, DDX58, KYNU, NR4A1, DENND5A, MMP10, HKDC1, RBM4)* had an OR at least 2-fold higher and a mutation count difference of 2 or more between samples with a mutation and BCR and those with a mutation and no BCR (*Supplementary Table 9*). The overwhelming majority (92.7%) of the PD mutations identified in these combined 22 risk-elevating genes are missense variants (*Supplementary Figure 2)*, although patients with BCR exhibited more non-missense

variants (*Supplementary Figure 3)* compared with those without BCR (*Supplementary Figure 4*).

Discussion

The primary aim of genetic profiling of germline or tumour DNA is to aid clinical decisions, such as targeted screening of asymptomatic individuals and treatment options for cancer patients. Germline signatures in particular would have the advantage of helping to stratify patients in both pre- and post-operative settings. Follow-up strategies and decisions on further treatments could be aided by predicting which individuals are likely to develop prostate tumours, progress to clinically significant disease or relapse. This study is the first to our knowledge to evaluate association of rare germline mutations across the full exome as opposed to specific plausible candidate genes, and provides evidence that germline mutation status is predictive for BCR after radical treatment for PrCa. Our multifactor *Cox* model identified that rare predicted-deleterious variants in three Hallmark gene-sets are associated with time to BCR after radical treatment (PI3K/AKT/mTOR, KRAS signalling (up) and Inflammatory response), and five gene-sets associated with BCR in a subset of cases with more aggressive phenotype at diagnosis (PI3K/AKT/mTOR, KRAS signalling (up), Hypoxia, TNFA signalling via NFKB and Pancreas-beta cells). Importantly, we also show that these gene-sets remained an independent predictive biomarker of time to BCR, over and above the inclusion of clinical variables. We further demonstrate that the Hypoxia gene-set replicated in an independent cohort of high-Gleason tumour cases from TCGA. With additional confirmation and refinement, these signatures could inform prognosis and clinical decision making.

Among the gene-sets associated with greater risk of BCR in PrCa patients, genes involved in PI3K/AKT/mTOR and KRAS signalling (up) remained significant across all PPCG samples as well as when restricted to patients with high-Gleason tumours. In

somatic analyses, AKT expression and phosphorylation have previously been linked to risk of BCR after radical prostatectomy [39, 40] and poorer survival in patients with metastatic castrate-resistant PrCa [41]. Somatic loss of *PTEN*, a tumour suppressor that downregulates the AKT signalling pathway, is also associated with poorer prognosis PrCa [5] and disease recurrence [42, 43]. The fact that the other gene-sets were not significant in the TCGA replication set could result from power limitations owing to the lower sample size (383 vs 850 samples), but these signatures will require validation in independent cohorts.

In the analysis of patients with high-Gleason tumours, the Hypoxia gene-set was established at statistical significance in the PPCG cohort and also replicated in the independent TCGA validation cohort. This provides strong evidence that germline mutations within this gene-set contribute to recurrence in patients with more aggressive disease. Hypoxia has previously been reported to contribute to progression when analysing tumour samples [44, 45], with a 28 gene mRNA signature for hypoxia demonstrated to predict BCR and metastases after radical prostatectomy or radiotherapy and provide independent prognostic value after adjustment for clinical features [46]. Our results indicate for the first time that heritable mutations in genes upregulated in response to a low oxygen environment predispose PrCa patients towards greater likelihood of, and shorter time to, BCR.

A small number of additional gene-sets also achieved significance in a single analysis only (Inflammatory response in PPCG all samples, TNFA signalling via NFKB and Pancreas-beta cells in the PPCG high-Gleason subset, and Myc targets v2 and Coagulation in the TCGA validation cohort). Due to the less consistent selection of these gene-sets, the importance of these gene-sets in germline susceptibility towards

BCR is less compelling; however they would nonetheless represent potential gene-sets of interest for examination in future larger replication studies.

In this study, we observed significantly shorter time to BCR among the individuals carrying mutations in >1 of the risk-increasing gene-sets. 58 out of the 850 total patients having mutations in multiple of the three all samples gene-sets, and 48 out of the 336 patients having mutations in multiple of the five high-Gleason gene-sets identified through our multifactor analysis, compared to both non-carriers and individuals carrying mutations in a single gene-set only. This provides further support that mutations affecting multiple regulatory networks may co-operatively serve to negatively influence PrCa prognosis; and that for some men, intraprostatic features that determine an aggressive tumour environment may be predetermined in the germline. This has been suggested before, based on hypoxia associating with genetic instability and aggressive sub-pathologies as field defects in PrCa, and warrants further investigation [47].

The limitations of this study include multi-cohort biases, relatively small, European ancestry-only sample size and in turn limited statistical power to detect associations at the individual gene or variant levels, and the imperfect status of BCR as a definitive surrogate for clinical recurrence and survival. In addition, this analysis included only coding variants with strong evidence for deleterious effect, excluding variants of uncertain significance, copy number alterations and structural variants. It may be necessary to integrate different data types, including expression and methylation data, to fully understand the mechanisms behind our findings. Although it is very encouraging that genes curated within PI3K/AKT/mTOR signalling and KRAS signalling (up) remained significant across both the PPCG all samples and high

Gleason subset analyses, and the independent validation cohort confirmed evidence for genes curated as involved in Hypoxia, additional larger studies remain necessary to confirm these findings and disentangle which specific genes contribute towards increased risk of PrCa progression and invasiveness.

Conclusion

Our findings have potentially important clinical implications. Germline DNA can be sequenced at an early stage of disease or even for healthy individuals which could enable prediction of PrCa progression close to, or even in advance of, the point of diagnosis. This would allow clinicians to stratify and differentiate patients that are more likely to relapse, putting them on a different clinical treatment pathway comprising more radical intervention or more frequent follow-up.

Take home message

Prostate cancer patients with inherited mutations in specific genes demonstrate a greater likelihood of relapsing after initial radical treatment. In the future, we may be able to use genetic information to identify sooner which patients may require additional treatments, and in turn improve prognoses for these individuals.

Acknowledgements

The authors thank those men with prostate cancer and the subjects who have donated their time and their samples to the Cambridge, Oxford, The Institute of Cancer Research, John Hopkins and University of Tampere BioMediTech Biorepositories for this study. We also acknowledge support of the research staff in S4 who so carefully curated the samples and the follow-up data (J. Burge, M. Corcoran, A. George and S. Stearn). We thank M. Stratton for discussions when setting up the CR-UK Prostate Cancer ICGC Project. We acknowledge support from Cancer Research UK C5047/A14835/A22530/ A17528, C309/A11566, C368/A6743, A368/A7990, C14303/A17197 (Z.K.-J., S. Merson, N.C., S.E., D.L., T. Dadaev, M.A., E.B., J.B., G.A., P.W., B.A.-L., D.S.B., C.S.C., R.A.E.), the Dallaglio Foundation (CR-UK Prostate Cancer ICGC Project and Pan Prostate Cancer Group), PC-UK/Movember (Z.K.-J.), the NIHR support to The Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust (Z.K.-J., N.D., S. Merson, N.C., S.E., D.L., T. Dadaev, S. Thomas, M.A., E.B., C.F., N.L., D.N., V.K., N.A., P.K., C.O., D.C., A.T., E.M., E.R., T. Dudderidge, S. Hazell, J.B., G.A., P.W., B.A.-L., D.S.B., C.S.C., R.A.E.), Cancer Research UK funding to The Institute of Cancer Research and the Royal Marsden NHS Foundation Trust CRUK Centre, the National Cancer Research Institute (National Institute of Health Research (NIHR) Collaborative Study: “Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT)” (grant G0500966/75466) (D.E.N., V.G.), the Li Ka Shing Foundation (D.C.W., D.J.W.), Canadian Institutes of Health Research (CIHR) Project Grant (J.R.), and the Academy of Finland and Cancer Society of Finland (G.S.B.). We thank the National Institute for Health Research, Hutchison Whampoa Limited, University of Cambridge and the Human Research Tissue Bank

(Addenbrooke's Hospital), which is supported by the NIHR Cambridge Biomedical Research Centre; The Core Facilities at the Cancer Research UK Cambridge Institute, Orchid and Cancer Research UK, D. Holland from the Infrastructure Management Team, and P. Clapham from the Informatics Systems Group at the Wellcome Trust Sanger Institute. D.M.B. is supported by Orchid. C.V.'s academic time was supported by the NIHR Oxford Biomedical Research Centre (Molecular Diagnostics Theme/Multimodal Pathology sub-theme). The validation results shown are based upon data generated by the TCGA Research Network. We also acknowledge support from the Bob Champion Cancer Trust, The Masonic Charitable Foundation successor to The Grand Charity, The King Family and the Stephen Hargrave Trust (C.S.C., D.S.B.). P.W. is a Cancer Research Life Fellow. We acknowledge core facilities provided by CRUK funding to the CRUK ICR Centre, the CRUK Cancer Therapeutics Unit and support for canSAR C35696/A23187 (P.W., G.A.). We would like to acknowledge The D. J Fielding Medical Research Trust for its support.

Support for analysis of the Australian samples was provided through the PRECEPT program grant, co-funded by Movember and the Australian Federal Government (PI N.M.C.) as well as NHMRC projects grants #10413 (CIs C.M.H., N.M.C., R.E.A.) and #1104010 (CIs C.M.H., N.M.C.). B.P. was supported by a Victorian Health and Medical Research Fellowship from the Department of Health and Human Services in the State of Victoria. N.M.C. was supported by a David Bickart Clinician Researcher Fellowship from the Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, and more recently by a Movember – Distinguished Gentleman's Ride Clinician Scientist Award through the Prostate Cancer Foundation of Australia's Research Program.

The French Prostate ICGC project was founded by Institut National de la Santé et de la Recherche Médicale (INSERM) and Institut National du Cancer (INCa); grant INSERM CV_2011/023 (C18), with additional support from LYric (grant INCa-4662). This project would not have been possible without the contribution of the bioinformatics platform from Synergy Lyon Cancer (Centre Léon Bérard, Lyon, France).

The PPCG project at Weill Cornell Medicine is supported by NCI P50CA211024, DoD PC160357, DoD PC180582, and the Prostate Cancer Foundation.

References

[1] Dell'Oglio P, Stabile A, Gandaglia G, et al. New surgical approaches for clinically high-risk or metastatic prostate cancer. Expert Rev Anticancer Ther. 2017;17:1013-31.
[2] Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. BMJ. 2013;346:f2023.
[3] Saunders EJ, Kote-Jarai Z, Eeles RA. Identification of Germline Genetic Variants that Increase Prostate Cancer Risk and Influence Development of Aggressive Disease. Cancers (Basel). 2021;13.
[4] Hull GW, Rabbani F, Abbas F, Wheeler TM, Kattan MW, Scardino PT. Cancer control with radical prostatectomy alone in 1,000 consecutive patients. J Urol. 2002;167:528-34.
[5] Bostrom PJ, Bjartell AS, Catto JW, et al. Genomic Predictors of Outcome in Prostate Cancer. Eur Urol. 2015;68:1033-44.
[6] Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589-95.
[7] Cooper CS, Eeles R, Wedge DC, et al. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat Genet. 2015;47:367-72.
[8] McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-303.
[9] DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491-8.
[10] Van der Auwera GAOC, B.D. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media. 2020.
[11] Garrison E.; Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907 [q-bioGN]. 2012.
[12] Gerhauser C, Favero F, Risch T, et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell. 2018;34:996-1011 e8.
[13] Tan A, Abecasis GR, Kang HM. Unified Representation of Genetic Variants. Bioinformatics. 2015;31:2202-4.
[14] Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019;47:D766-D73.
[15] Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010;5:1564-73.
[16] Leongamornlert DA, Saunders EJ, Wakerell S, et al. Germline DNA Repair Gene Mutations in Young-onset Prostate Cancer Cases in the UK: Evidence for a More Extensive Genetic Panel. Eur Urol. 2019;76:329-37.
[17] Mijuskovic M, Saunders EJ, Leongamornlert DA, et al. Rare germline variants in DNA repair genes and the angiogenesis pathway predispose prostate cancer patients to develop metastatic disease. Br J Cancer. 2018;119:96-104.
[18] Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326-8.

[19] Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature. 2015;526:68-74.
[20] Picard Tools. Broad Institute, GitHub repository (2021; v2238; http://broadinstitutegithubio/picard/).
[21] Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012;91:839-48.
[22] Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434-43.
[23] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-5.
[24] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417-25.
[25] Van Rossum G, Drake Jr FL. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam. 1995.
[26] Casbon J, Dougherty J. PyVCF. (https://pyvcfreadthedocsio/). 2011.
[27] Reback J, jbrockmendel, McKinney W, et al. pandas-dev/pandas: v1.3.0. Zenodo. https://doi.org/10.5281/zenodo.5060318. 2021.
[28] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods. 2020;17:261-72.
[29] Harris CR, Jarrod Millman K, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585:357-62.
[30] Pérez F, Granger BE. IPython: A System for Interactive Scientific Computing. Computing in Science & Engineering. 2007;9:21-9.
[31] Evans C, Paolo F, Natsuo Kishimoto P, et al. cgevans/scikits-bootstrap: v1.1.0-pre.1. Zenodo. https://doi.org/10.5281/zenodo.5095250. 2021.
[32] Davidson-Pilon C, Kalderstam J, Jacobson N, et al. CamDavidsonPilon/Lifelines: v0.25.0. Zenodo. https://doi.org/10.5281/zenodo.3962065. 2020.
[33] Hunter JD. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering. 2007;9:90-5.
[34] Rusnack W, Moyer E, Taylor R. to_precision. [accessed on 16 November 2021]; Available online: https://bitbucketorg/william_rusnack/to-precision/src/master/. 2013.
[35] Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. 2018;28:1747-56.
[36] Budaus L, Schiffmann J, Graefen M, et al. Defining biochemical recurrence after radical prostatectomy and timing of early salvage radiotherapy : Informing the debate. Strahlenther Onkol. 2017;193:692-9.
[37] Fraser M, Sabelnykova VY, Yamaguchi TN, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541:359-64.
[38] Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Statist Soc B. 1996;58:267-88.
[39] Bedolla R, Prihoda TJ, Kreisberg JI, et al. Determining risk of biochemical recurrence in prostate cancer by immunohistochemical detection of PTEN expression and Akt activation. Clin Cancer Res. 2007;13:3860-7.
[40] Torrealba N, Rodriguez-Berriguete G, Fraile B, et al. PI3K pathway and Bcl-2 family. Clinicopathological features in prostate cancer. Aging Male. 2018;21:211-22.

[41] McCall P, Gemmell LK, Mukherjee R, Bartlett JM, Edwards J. Phosphorylation of the androgen receptor is associated with reduced survival in hormone-refractory prostate cancer patients. Br J Cancer. 2008;98:1094-101.
[42] Geybels MS, Fang M, Wright JL, et al. PTEN loss is associated with prostate cancer recurrence and alterations in tumor DNA methylation profiles. Oncotarget. 2017;8:84338-48.
[43] Chaux A, Peskoe SB, Gonzalez-Roibon N, et al. Loss of PTEN expression is associated with increased risk of recurrence after prostatectomy for clinically localized prostate cancer. Modern Pathol. 2012;25:1543-9.
[44] Bhandari V, Hoey C, Liu LY, et al. Molecular landmarks of tumor hypoxia across cancer types. Nat Genet. 2019;51:308-18.
[45] Lalonde E, Ishkanian AS, Sykes J, et al. Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol. 2014;15:1521-32.
[46] Yang L, Roberts D, Takhar M, et al. Development and Validation of a 28-gene Hypoxia-related Prognostic Signature for Localized Prostate Cancer. EBioMedicine. 2018;31:182-9.
[47] Chua MLK, Lo W, Pintilie M, et al. A Prostate Cancer "Nimbosus": Genomic Instability and SChLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. Eur Urol. 2017;72:665-74.

Tables

Table 1: Number of samples, genes and variants contributed, by study, also showing the number of samples with high-Gleason score (>3+4; Gleason grade group 3-5), the numbers of samples in each set with biochemical recurrence (BCR), numbers associated with mutations that are predicted-deleterious, and how many of those are known deleterious/loss-of-function (LoF) mutations.

Study	**European Genome-phenome Archive ID**	**Samples used in study after QC (with BCR)**	**Samples with high-Gleason score (with BCR)**	**Number of genes with predicted-deleterious mutations (LoF)**	**Number of predicted-deleterious mutations included in analysis (LoF)**
Melbourne, Australian research group	EGAD00001004182	133 (79)	110 (70)	2,917 (1,884)	3,728 (2,473)
Canadian Prostate Cancer Genome Network	EGAD00001004170	288 (92)	63 (22)	4,579 (2,637)	5,900 (3,154)
French ICGC Prostate Cancer Group	EGAD00001003835	15 (10)	15 (10)	409 (255)	393 (243)
Germany ICGC Prostate Cancer Group – Early Onset	EGAD00001005997	230 (68)	85 (45)	3,787 (2,160)	4,761 (2,404)
CRUK-ICGC Prostate Group, UK	EGAC00001000852	184 (36)	63 (22)	3,365 (2,073)	4,071 (2,401)
Total		**850 (285)**	**336 (169)**	**8,455 (5,792)**	**15,822 (9,006)**

Table 2: Multifactor *Cox* model results for predicted-deleterious mutations in 850 germline samples, grouped into 52 gene-sets. Shown are *p*-values and hazard ratios of LASSO-selected gene-sets as well as clinical variables reported at time of biochemical recurrence (BCR) or last check-up, impacting the predicted time until BCR.

	HR (95% CI)	*p*-value	Bootstrap HR (95% CI)	Bootstrap *p*-value
Gleason (≥4+3 : <4+3)	1.98 (1.47 – 2.67)	<0.0017.03x10^{-6}	2.01 (1.99 – 2.04)	0.00<0.001
Stage (T3-T4 : T1-T2)	1.69 (1.29 – 2.21)	<0.0011.32x10^{-4}	1.75 (1.74 – 1.77)	0.00<0.001
PI3K/AKT/mTOR signalling	1.55 (1.06 – 2.25)	**0.023**26	1.58 (1.56 – 1.60)	0.0120
Age	1.53 (1.20 – 1.96)	<0.0017.16x10^{-4}	1.03 (1.03 – 1.03)	0.0011.00x10^{-3}
Inflammatory response	1.35 (1.00 – 1.82)	**0.0483**	1.37 (1.35 – 1.38)	0.0280
KRAS signalling (up)	1.35 (1.01 – 1.79)	**0.0413**	1.37 (1.36 – 1.38)	0.0200
Fatty acid metabolism	1.29 (0.96 – 1.71)	0.08768	1.32 (1.30 – 1.33)	0.0400
G2-M checkpoint	1.25 (0.94 – 1.66)	0.130	1.27 (1.26 – 1.28)	0.0740
Myc targets v2	1.23 (0.84 – 1.81)	0.30	1.26 (1.24 – 1.27)	0.1655
Mitotic spindle	1.21 (0.94 – 1.56)	0.142	1.22 (1.21 – 1.23)	0.100
DRG	1.16 (0.90 – 1.51)	0.326	1.18 (1.17 – 1.19)	0.151
p53 pathway	1.16 (0.85 – 1.60)	0.435	1.18 (1.16 – 1.19)	0.21
IL-2/STAT5 signalling	1.06 (0.77 – 1.46)	0.72	1.07 (1.06 – 1.09)	0.438
Preop_PSA	1.04 (1.01 – 1.06)	**0.006**5.91x10^{-3}	1.01 (1.01 – 1.01)	0.0044.00x10^{-3}
Coagulation	1.01 (0.76 – 1.36)	0.928	1.01 (1.00 – 1.02)	0.51
Glycolysis	0.81 (0.61 – 1.08)	0.1655	0.82 (0.81 – 0.82)	0.0800

UV response (dn)	0.71 (0.51 – 0.99)	**0.042~~18~~**	0.72 (0.71 – 0.73)	0.038~~0~~
Cholesterol homeostasis	0.58 (0.34 – 1.00)	**0.048~~3~~**	0.59 (0.58 – 0.60)	0.013~~0~~

Table 3: Multifactor *Cox* model results for predicted-deleterious mutations in 336 high-Gleason germline samples, grouped into 52 gene-sets. Shown are *p*-values and hazard ratios of LASSO-selected gene-sets impacting the predicted time until biochemical recurrence.

	HR (95% CI)	*p*-value	Bootstrap HR (95% CI)	Bootstrap *p*-value
Pancreas-beta cells	2.52 (1.01 – 6.29)	**0.047~~0~~**	3.58 (3.43 – 3.73)	0.034~~0~~
PI3K/AKT/mTOR signalling	1.95 (1.21 – 3.15)	**0.006~~5.91x10^{-3}~~**	2.13 (2.09 – 2.17)	0.007~~7.00x10^{-3}~~
TNFA signalling via NFKB	1.79 (1.19 – 2.68)	**0.005~~4.85x10^{-3}~~**	1.86 (1.83 – 1.89)	0.005~~5.00x10^{-3}~~
Hypoxia	1.73 (1.14 – 2.63)	**0.010~~1~~**	1.82 (1.79 – 1.85)	0.011~~0~~
Stage (T3-T4 : T1-T2)	1.70 (1.13 – 2.56)	**0.012~~15~~**	1.86 (1.84 – 1.89)	0.003~~3.00x10^{-3}~~
KRAS signalling (up)	1.58 (1.08 – 2.32)	**0.018~~8~~9**	1.65 (1.63 – 1.67)	0.016~~0~~
Myc targets v2	1.54 (0.92 – 2.60)	0.10~~4~~	1.60 (1.57 – 1.63)	0.081~~0~~
DRG	1.33 (0.92 – 1.91)	0.13~~28~~	1.38 (1.36 – 1.39)	0.071~~0~~
G2-M checkpoint	1.31 (0.89 – 1.93)	0.1~~6~~7	1.41 (1.39 – 1.43)	0.092~~0~~
Age	1.17 (0.90 – 1.52)	0.2~~4~~	1.01 (1.01 – 1.01)	0.17~~0~~
IL-6/JAK/STAT3 signalling	1.17 (0.69 – 1.98)	0.~~5~~6	1.22 (1.19 – 1.24)	0.3~~1~~
Preop_PSA	1.06 (1.00 – 1.11)	0.039~~1~~	1.00 (1.00 – 1.00)	0.028~~0~~
Coagulation	1.05 (0.71 – 1.55)	0.8~~0~~	1.08 (1.06 – 1.09)	0.4~~1~~

mTORC1 signalling	0.79 (0.50 – 1.25)	0.32	0.80 (0.79 – 0.82)	0.172
Androgen response	0.71 (0.41 – 1.22)	0.22	0.73 (0.72 – 0.74)	0.121
Glycolysis	0.60 (0.40 – 0.91)	**0.017**66	0.61 (0.60 – 0.62)	0.0120
Cholesterol homeostasis	0.564 (0.270 – 1.18)	0.1328	0.586 (0.571 – 0.602)	0.0630

Table 4: Multifactor *Cox* model results for predicted-deleterious mutations in 233 high-Gleason The Cancer Genome Atlas (TCGA) germline samples, stratified by location and grouped into 52 gene-sets. Shown are *p*-values and hazard ratios of the same predictors identified by the Pan Prostate Cancer Group (PPCG) *Cox* model (pancreas-beta cells and cholesterol homeostasis were removed as most samples had a mutation or had no mutation in the gene-set respectively, which caused convergence errors).

	HR (95% CI)	***p*-value**	**Bootstrap HR (95% CI)**	**Bootstrap *p*-value**
Stage (T3-T4 : T1-T2)	7.85 (1.65 – 37.3)	0.019.55x10^{-3}	6.24x10^{12} (1.73x10^{8} – 3.73x10^{13})	0.0011.00x10^{-3}
Coagulation	3.53 (1.30 – 9.59)	**0.01**435	11.3 (7.47 – 28.5)	0.0220
Hypoxia	3.18 (1.04 – 9.74)	**0.043**25	7.88x10^{6} (1.14x10^{6} – 3.40x10^{7})	0.0970
Myc targets v2	2.90 (1.00 – 8.40)	**0.049**2	5.63 (5.29 – 6.07)	0.0440
TNFA signalling via NFKB	2.12 (0.78 – 5.79)	0.143	3.95 (3.51 – 4.97)	0.110
G2-M checkpoint	2.00 (0.79 – 5.05)	0.144	2.89 (2.75 – 3.11)	0.102

Androgen response	1.43 (0.52 – 3.97)	0.549	1.81 (1.70 – 2.00)	0.32
IL-6/JAK/STAT3 signalling	1.32 (0.36 – 4.77)	0.768	$2.86x10^{8}$ ($5.67 – 1.71x10^{9}$)	0.33
KRAS signalling (up)	0.97 (0.39 – 2.43)	≥0.95	1.36 (1.29 – 1.46)	0.51
PI3K/AKT/mTOR signalling	0.70 (0.08 – 5.77)	0.74	$1.52x10^{6}$ ($0.972 – 7.60x10^{6}$)	0.31
DRG	0.68 (0.31 – 1.49)	0.34	0.72 (0.70 – 0.75)	0.181
mTORC1 signalling	0.46 (0.14 – 1.50)	0.2199	0.46 (0.43 – 0.49)	0.0750
Glycolysis	0.27 (0.07 – 1.09)	0.067 67	0.34 (0.31 – 0.36)	0.0470

Figure Legends

Figure 1: Horizontal box plot of the coefficient / log hazard ratios with lower and upper 95% confidence intervals for A) Table 2, B) Table 3 and C) Table 4.

Figure 2: *Kaplan-Meier* plot showing survival probability against time in months until biochemical recurrence (BCR) for A) all samples, and B) the 336 samples in the high-Gleason subset (Gleason score >3+4; Gleason grade group 3-5). The impact of mutations in significant sets are subdivided by samples with mutations in multiple gene-sets. Log-rank tests for each category: A) =1 (p=0.63); ≥2 (p=2.88x10^{-3}). B) =1 (p=0.27); =2 (p=8.55x10^{-3}); ≥3 (p=3.29x10^{-3}).