available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Prostate Cancer

Rare Germline Variants Are Associated with Rapid Biochemical Recurrence After Radical Prostate Cancer Treatment: A Pan Prostate Cancer Group Study

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,l,m], *Niall M. Corcoran*[n,o,p], *Chol-Hee Jung*[q], *Bernard J. Pope*[n,r], *Chris M. Hovens*[p,q,s], *Géraldine Cancel-Tassin*[t,u], *Olivier Cussenot*[t,u], *Massimo Loda*[v], *Chris Sander*[w], *Vanessa M. Hayes*[x,y], *Karina Dalsgaard Sorensen*[z,aa], *Yong-Jie Lu*[ab], *Freddie C. Hamdy*[ac], *Christopher S. Foster*[a], *Vincent Gnanapragasam*[ad], *Adam Butler*[ae], *Andy G. Lynch*[af,ag], *Charlie E. Massie*[ah], *CR-UK/Prostate Cancer UK, ICGC, The PPCG*[ai] *Dan J. Woodcock*[aj], *Colin S. Cooper*[ak], *David C. Wedge*[al], *Daniel S. Brewer*[ak,am,†], *Zsofia Kote-Jarai*[a,*,†], *Rosalind A. Eeles*[a,an,†]

[a] *The Institute of Cancer Research, London, UK;* [b] *Manchester Cancer Research Centre and CRUK Manchester Institute, The University of Manchester, Manchester, UK;* [c] *Princess Margaret Cancer Centre/University Health Network, Toronto, Ontario, Canada;* [d] *Computational Biology Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada;* [e] *Department of Medical Biophysics & Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada;* [f] *Charité - Universitätsmedizin Berlin, Berlin, Germany;* [g] *University Medical Centre Hamburg - Eppendorf, Hamburg, Germany;* [h] *German Cancer Research Center (DKFZ), Deutsches Krebsforschungszentrum, Heidelberg, Germany;* [i] *European Molecular Biology Laboratory (EMBL), Heidelberg, Germany;* [j] *Biotech Research & Innovation Centre (BRIC) & Finsen Laboratory, University of Copenhagen, Rigshospitalet, Copenhagen, Denmark;* [k] *Centre for Molecular Medicine Norway (NCMM), Nordic EMBL Partnership, University of Oslo and Oslo University Hospital, Oslo, Norway;* [l] *Department of Neurology, University of California, San Francisco, San Francisco, CA, USA;* [m] *Department of Pediatric Research, Division of Pediatric and Adolescent Medicine, Rikshospitalet, Oslo University Hospital, Oslo, Norway;* [n] *Department of Surgery, The University of Melbourne, Grattan Street, Parkville, Victoria, Australia;* [o] *Department of Urology, Royal Melbourne Hospital, Parkville, Victoria, Australia;* [p] *Melbourne Bioinformatics, The University of Melbourne, Grattan Street, Victoria, Australia;* [q] *The University of Melbourne, Grattan Street, Parkville, Victoria, Australia;* [r] *Royal Melbourne Hospital, Melbourne, Parwille, Victoria, Australia;* [s] *University of Melbourne Centre for Cancer Research, The Victorian Comprehensive Cancer Centre, Parkville, Victoria, Australia;* [t] *CeRePP, Hopital Tenon, Paris, France;* [u] *Sorbonne Universite, GRC n°5 Predictive Onco-Urology, APHP, Tenon Hospital, Paris, France;* [v] *Department of Pathology & Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA;* [w] *cBio Center, Dana-Farber Cancer Institute, Boston, MA, USA;* [x] *Garvan Institute of Medical Research, The Kinghorn Cancer Centre, Darlinghurst, NSW, Australia;* [y] *School of Medical Sciences, University of Sydney, Charles Perkins Centre, Camperdown, NSW, Australia;* [z] *Department of Molecular Medicine, Aarhus University Hospital, Aarhus N, Denmark;* [aa] *Department of Clinical Medicine, Aarhus University Hospital, Aarhus N, Denmark;* [ab] *Centre for Biomarker and Therapeutics, Barts Cancer Institute, Queen Mary University of London, London, UK;* [ac] *Nuffield Department of Surgical Sciences University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK;* [ad] *Department of Surgery, Division of Urology, University of Cambridge, Cambridge, UK;* [ae] *Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge, UK;* [af] *School of Medicine, University of St Andrews, St Andrews, Fife, UK;* [ag] *School of Mathematics & Statistics, St Andrews, Fife, UK;* [ah] *CRUK Cambridge Institute, Hutchison MRC Research Centre, University of Cambridge, Li Ka Shing Centre, Cambridge, UK;* [aj] *Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK;* [ak] *Norwich Medical School, University of East Anglia, Norwich, UK;* [al] *Manchester Cancer Research Centre, The University of Manchester, Manchester, UK;* [am] *The Earlham Institute, Norwich Research Park, Norwich, UK;* [an] *The Royal Marsden NHS Foundation Trust, London, UK;* [ai] *CR-UK/Prostate Cancer UK, ICGC, The Pan Prostate Cancer Group, UK*

[†] These authors are joint last authors.
[*] Corresponding author. The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP, UK.
E-mail address: Zsofia.Kote-Jarai@icr.ac.uk (Z. Kote-Jarai).

https://doi.org/10.1016/j.eururo.2022.05.007

Article info

Article history:
Accepted May 10, 2022

Associate Editor:
James Catto

Statistical Editor:
Melissa Assel

Keywords:
Germline variants
Prostate cancer
Biochemical recurrence
Pan Prostate Cancer Group

EU ACME
www.eu-acme.org/europeanurology

Please visit www.eu-acme.org/europeanurology to answer questions on-line. The EU-ACME credits will then be attributed automatically.

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 the UK, Canada, Germany, Australia, and France. Findings were validated using 383 patients from The Cancer Genome Atlas (TCGA) dataset.
Outcome measurements and statistical analysis: A total of 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 and limitations: 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: Prostate cancer patients with particular genetic mutations have a higher chance of relapsing after initial radical treatment, potentially providing opportunities to identify patients who might need additional treatments earlier.

1. 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, an understanding of potential heritable genetic factors contributing to tumour 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 a 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.

2. Patients and methods

2.1. 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, and 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 the criteria outlined in the Supplementary material. Collection was subject to the International Cancer Genome Consortium (ICGC) standards of ethical consent. Collection and

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)	2917 (1884)	3728 (2473)
Canadian Prostate Cancer Genome Network	EGAD00001004170	288 (92)	63 (22)	4579 (2637)	5900 (3154)
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)	3787 (2160)	4761 (2404)
CRUK-ICGC Prostate Group, UK	EGAC00001000852	184 (36)	63 (22)	3365 (2073)	4071 (2401)
Total		850 (285)	336 (169)	8455 (5792)	15, 822 (9006)

QC = quality control.

analysis of the Australian samples received institutional review board approval (Epworth Health 34506; Melbourne Health 2019.058). WGS was performed using Illumina technology to ≥30× depth. The Burrows-Wheeler Aligner [6] was used to align sequencing data to the GRCh37 human genome (human_g1k_v37) with polymerase chain reaction duplicates removed [7]. Sequencing data have been deposited at the European Genome-phenome Archive (https://ega-archive.org; dataset IDs in Table 1) and are available upon request.

2.2. Variant calling

Variant calling was performed with the Genome Analysis Toolkit pipeline (GATK v4.0) [8] following GATK best practice recommendations for germline single nucleotide variant and indel calling (Supplementary material) [9,10], apart from for the German samples that were called using FreeBayes v1.1.0 [11] and processed as described by Gerhauser et al [12], normalised with vt v0.5 (Supplementary material) [13]. This analysis was restricted to variants within protein-coding transcript sequences according to GENCODE v29 [14].

2.3. Quality control, variant annotation, and prioritisation

Low-quality variants and samples were removed based on established quality control (QC) protocols [15–17]. We excluded samples from related individuals (using R package SNPRelate method identity by descent [18]) or those with non-European ancestry (using Principal Component Analysis relative to 2504 samples from the 1000 Genomes Project [19]). We used Picard tools v2.23.8 [20] to remove samples with a mean insert size of <250 bp, AT or GC dropout >5%, <95% aligned reads, >5% mismatch rate, <80% with ≥20× 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, those with monomorphic loci, those in repetitive regions (simple repeats, segmental duplications, and centromeric regions), and those for which the ExAC minor allele frequency in any population was >1%. Of the submitted samples, 3% were excluded based on ancestry, whilst 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 (PD) missense variants with a CADD PHRED score of >30 (Table 1) [23].

2.4. Pathways and gene sets

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

2.5. Statistical analysis

2.5.1. Software and libraries

All statistical analyses were applied using Python v3.8 [25]. Data in VCF format were 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 Matplotlib v3.3.4 [33], to_precision [34], and Maftools v2.6.5 [35].

2.5.2. 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 PH model were generated by recording the presence of any gene with PD mutations in the selected gene sets across all samples. Pathological 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 RT. Preoperative PSA and age at the time of surgery were continuous variables. Gleason score had a baseline of ≤3 + 4 (Gleason grade group 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 RP was defined as two consecutive post-RP PSA measurements of >0.2 ng/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 >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 penalisation 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 penalisation.

Table 2 – Multifactor Cox model results for predicted-deleterious mutations in 850 germline samples, grouped into 52 gene sets

	HR (95% CI)	*p* value	Bootstrap HR (95% CI)	Bootstrap *p* value
Gleason (≥4 + 3 : <4 + 3)	1.98 (1.47–2.67)	**<0.001**	2.01 (1.99–2.04)	<0.001
Stage (T3–T4 : T1–T2)	1.69 (1.29–2.21)	**<0.001**	1.75 (1.74–1.77)	<0.001
PI3K/AKT/mTOR signalling	1.55 (1.06–2.25)	**0.023**	1.58 (1.56–1.60)	0.012
Age	1.53 (1.20–1.96)	**<0.001**	1.03 (1.03–1.03)	0.001
Inflammatory response	1.35 (1.00–1.82)	**0.048**	1.37 (1.35–1.38)	0.028
KRAS signalling (up)	1.35 (1.01–1.79)	**0.041**	1.37 (1.36–1.38)	0.020
Fatty acid metabolism	1.29 (0.96–1.71)	0.087	1.32 (1.30–1.33)	0.040
G2-M checkpoint	1.25 (0.94–1.66)	0.13	1.27 (1.26–1.28)	0.074
Myc targets v2	1.23 (0.84–1.81)	0.3	1.26 (1.24–1.27)	0.16
Mitotic spindle	1.21 (0.94–1.56)	0.14	1.22 (1.21–1.23)	0.10
DRG	1.16 (0.90–1.51)	0.3	1.18 (1.17–1.19)	0.15
p53 pathway	1.16 (0.85–1.60)	0.4	1.18 (1.16–1.19)	0.2
IL-2/STAT5 signalling	1.06 (0.77–1.46)	0.7	1.07 (1.06–1.09)	0.4
Preop PSA	1.04 (1.01–1.06)	**0.006**	1.01 (1.01–1.01)	0.004
Coagulation	1.01 (0.76–1.36)	0.9	1.01 (1.00–1.02)	0.5
Glycolysis	0.81 (0.61–1.08)	0.16	0.82 (0.81–0.82)	0.080
UV response (dn)	0.71 (0.51–0.99)	**0.042**	0.72 (0.71–0.73)	0.038
Cholesterol homoeostasis	0.58 (0.34–1.00)	**0.048**	0.59 (0.58–0.60)	0.013

BCR = biochemical recurrence; CI = confidence interval; dn = down; DRG = DNA repair gene; HR = hazard ratio; PSA = prostate-specific antigen.
Shown are *p* values and hazard ratios of LASSO-selected gene sets as well as clinical variables reported at the time of BCR or last check-up, impacting the predicted time until BCR.

2.5.3. *Univariate Cox regression*

Each gene set was modelled individually along with clinical covariates (preoperative PSA, pathological T stage, Gleason score, and age) and *p* values were adjusted for multiple testing using false discovery rate (FDR).

2.5.4. *Validation*

We performed harmonised variant filtering for PD mutations on germline PrCa samples from The Cancer Genome Atlas (TCGA) PRAD project. From the original 500 TCGA PRAD samples, those from contributing institutions with <15 samples were excluded, and models were stratified by institution, resulting in 383 samples used in the analysis. Of these, 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 (HRs) in both sets.

2.5.5. *Kaplan-Meier analysis*

A Kaplan-Meier plot measuring time to BCR in the event of a relapse was used to visualise the impact of mutations within significant gene sets on the risk of BCR. This was applied separately to the whole dataset and the 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.

2.5.6. *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. The *p* values were computed for each predictor as a percentage of the iterations where the coefficient was in a different direction than expected.

3. Results

We analysed germline WGS data from 850 patients across five studies in the PPCG consortium (Table 1 and 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 PD. No individual variants or genes demonstrated a significant association with time to BCR (Cox regression analysis; $p > 0.05$), although the available sample size of 850 cases is underpowered for such an 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 PD 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 to 200, the DRG panel containing 175 DNA repair genes [16], and the extended BROCA gene panel containing 65 genes (Supplementary Table 4) [17].

After variable selection by LASSO, the optimal model for predicting time to BCR contained 14 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 Fig. 1A). Clinical variables at the time of radical treatment (preoperative 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; 95% confidence interval [CI] 1.06–2.25; $p = 0.023$), Inflammatory response (HR = 1.35; 95% CI 1.00–1.82; $p = 0.048$), and KRAS signalling (up) (HR = 1.35; 95% CI 1.01–1.79; $p = 0.041$). These gene sets are associated with shortened average time to BCR. The UV response (down) (HR = 0.71; 95% CI 0.51–0.99; $p = 0.042$) and cholesterol homoeostasis (HR = 0.58; 95% CI 0.34–1.00; $p = 0.048$) gene sets were borderline significantly protective. Application of this model to multiple bootstrap resamplings showed that these results are robust, with all risk-elevating gene sets having HR >1 in >97% of resamples and *p* values indicating the same coefficient direction.

The clinical covariate-only model built using all the samples determined that Gleason score, preoperative PSA, age,

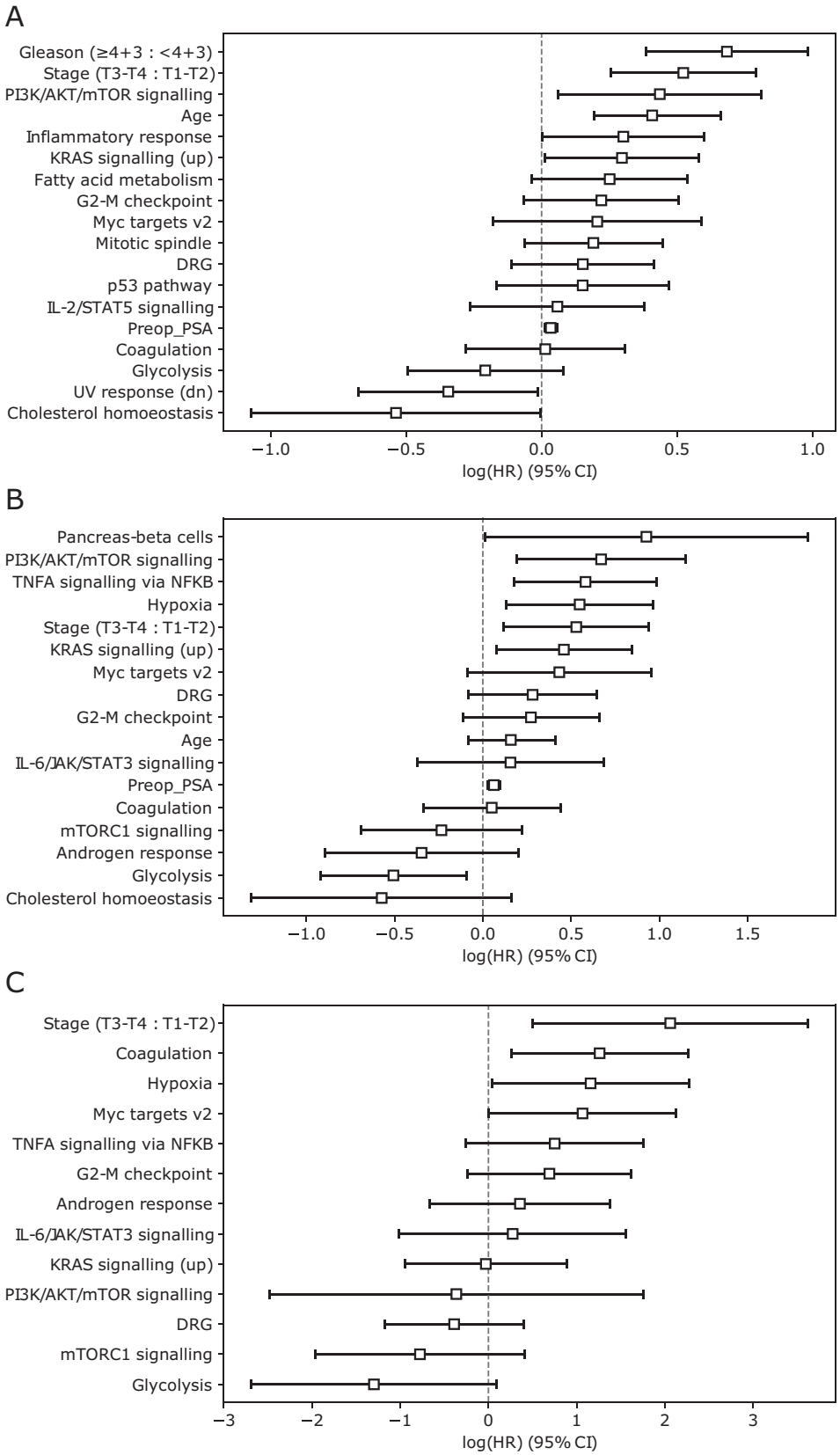

Fig. 1 – Horizontal box plot of the coefficient/log hazard ratios with lower and upper 95% confidence intervals for (A) Table 2, (B) Table 3, and (C) Table 4. CI = confidence interval; dn = down; DRG = DNA repair gene; HR = hazard ratio; PSA = prostate-specific antigen.

Table 3 – Multifactor Cox model results for predicted-deleterious mutations in 336 high-Gleason germline samples, grouped into 52 gene sets

	HR (95% CI)	*p* value	Bootstrap HR (95% CI)	Bootstrap *p* value
Pancreas-beta cells	2.52 (1.01–6.29)	**0.047**	3.58 (3.43–3.73)	0.034
PI3K/AKT/mTOR signalling	1.95 (1.21–3.15)	**0.006**	2.13 (2.09–2.17)	0.007
TNFA signalling via NFKB	1.79 (1.19–2.68)	**0.005**	1.86 (1.83–1.89)	0.005
Hypoxia	1.73 (1.14–2.63)	**0.010**	1.82 (1.79–1.85)	0.011
Stage (T3–T4 : T1–T2)	1.70 (1.13–2.56)	**0.012**	1.86 (1.84–1.89)	0.003
KRAS signalling (up)	1.58 (1.08–2.32)	**0.019**	1.65 (1.63–1.67)	0.016
Myc targets v2	1.54 (0.92–2.60)	0.10	1.60 (1.57–1.63)	0.081
DRG	1.33 (0.92–1.91)	0.13	1.38 (1.36–1.39)	0.071
G2-M checkpoint	1.31 (0.89–1.93)	0.17	1.41 (1.39–1.43)	0.092
Age	1.17 (0.90–1.52)	0.2	1.01 (1.01–1.01)	0.17
IL-6/JAK/STAT3 signalling	1.17 (0.69–1.98)	0.6	1.22 (1.19–1.24)	0.3
Preop PSA	1.06 (1.00–1.11)	0.039	1.00 (1.00–1.00)	0.028
Coagulation	1.05 (0.71–1.55)	0.8	1.08 (1.06–1.09)	0.4
mTORC1 signalling	0.79 (0.50–1.25)	0.3	0.80 (0.79–0.82)	0.17
Androgen response	0.71 (0.41–1.22)	0.2	0.73 (0.72–0.74)	0.12
Glycolysis	0.60 (0.40–0.91)	**0.017**	0.61 (0.60–0.62)	0.012
Cholesterol homoeostasis	0.564 (0.270–1.18)	0.13	0.586 (0.571–0.602)	0.063

CI = confidence interval; DRG = DNA repair gene; HR = hazard ratio; PSA = prostate-specific antigen.
Shown are *p* values and hazard ratios of LASSO-selected gene sets impacting the predicted time until biochemical recurrence.

Table 4 – Multifactor Cox model results for predicted-deleterious mutations in 233 high-Gleason TCGA germline samples, stratified by location and grouped into 52 gene sets

	HR (95% CI)	*p* value	Bootstrap HR (95% CI)	Bootstrap *p* value
Stage (T3–T4 : T1–T2)	7.85 (1.65–37.3)	**0.01**	6.24×10^{12} (1.73×10^{8}–3.73×10^{13})	0.001
Coagulation	3.53 (1.30–9.59)	**0.014**	11.3 (7.47–28.5)	0.022
Hypoxia	3.18 (1.04–9.74)	**0.043**	7.88×10^{6} (1.14×10^{6}–3.40×10^{7})	0.097
Myc targets v2	2.90 (1.00–8.40)	**0.049**	5.63 (5.29–6.07)	0.044
TNFA signalling via NFKB	2.12 (0.78–5.79)	0.14	3.95 (3.51–4.97)	0.11
G2-M checkpoint	2.00 (0.79–5.05)	0.14	2.89 (2.75–3.11)	0.10
Androgen response	1.43 (0.52–3.97)	0.5	1.81 (1.70–2.00)	0.3
IL-6/JAK/STAT3 signalling	1.32 (0.36–4.77)	0.7	2.86×10^{8} (5.67–1.71×10^{9})	0.3
KRAS signalling (up)	0.97 (0.39–2.43)	>0.9	1.36 (1.29–1.46)	0.5
PI3K/AKT/mTOR signalling	0.70 (0.08–5.77)	0.7	1.52×10^{6} (0.972–7.60×10^{6})	0.3
DRG	0.68 (0.31–1.49)	0.3	0.72 (0.70–0.75)	0.18
mTORC1 signalling	0.46 (0.14–1.50)	0.2	0.46 (0.43–0.49)	0.075
Glycolysis	0.27 (0.07–1.09)	0.067	0.34 (0.31–0.36)	0.047

CI = confidence interval; DRG = DNA repair gene; HR = hazard ratio; TCGA = The Cancer Genome Atlas.
Shown are *p* values and hazard ratios of the same predictors identified by the Pan Prostate Cancer Group Cox model (pancreas-beta cells and cholesterol homoeostasis were removed, as most samples had a mutation or no mutation in the gene set, respectively, which caused convergence errors).

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 $p = 0.048$; 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; Gleason grade group 3–5) had a higher proportion of BCR events (50.2% compared with 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 Fig. 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; 95% CI 1.01–6.29; $p = 0.047$), PI3K/AKT/mTOR signalling (HR = 1.95; 95% CI 1.21–3.15; $p = 0.006$), TNFA signalling via NFKB (HR = 1.79; 95% CI 1.19–2.68; $p = 0.005$), Hypoxia (HR = 1.73; 95% CI 1.14–2.63; $p = 0.010$), and KRAS signalling (up) (HR = 1.58; 95% CI 1.08–2.32; $p = 0.019$). PI3K/AKT/mTOR has a higher HR and a lower *p* value than in the all-sample model. The Glycolysis gene set is shown here as significantly protective (HR = 0.60; 95% CI 0.40–0.91; $p = 0.017$). The bootstrap resamplings for the significant gene sets have the same coefficient direction in >96% of resamples.

After examining each gene set in individual univariate models with all samples, none was found to have a significant association with outcome after multiple testing correction (FDR; *p* value threshold <0.1; Supplementary Table 6). PI3K/AKT/mTOR signalling ($q = 0.14$), KRAS signalling (up) ($q = 0.20$), and TNFA signalling via NKFB ($q = 0.16$) 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 ($p = 0.027$), PI3K/AKT/mTOR signalling ($p = 0.025$), KRAS signalling (up) ($p = 0.013$), and Pancreas-beta cells ($p = 0.023$). In the multifactor high-Gleason Cox model, these four gene sets are also statistically significant (Table 3), alongside Hypoxia.

Application of 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; 95% CI 1.73–11.5; $p = 0.002$) and Coag-

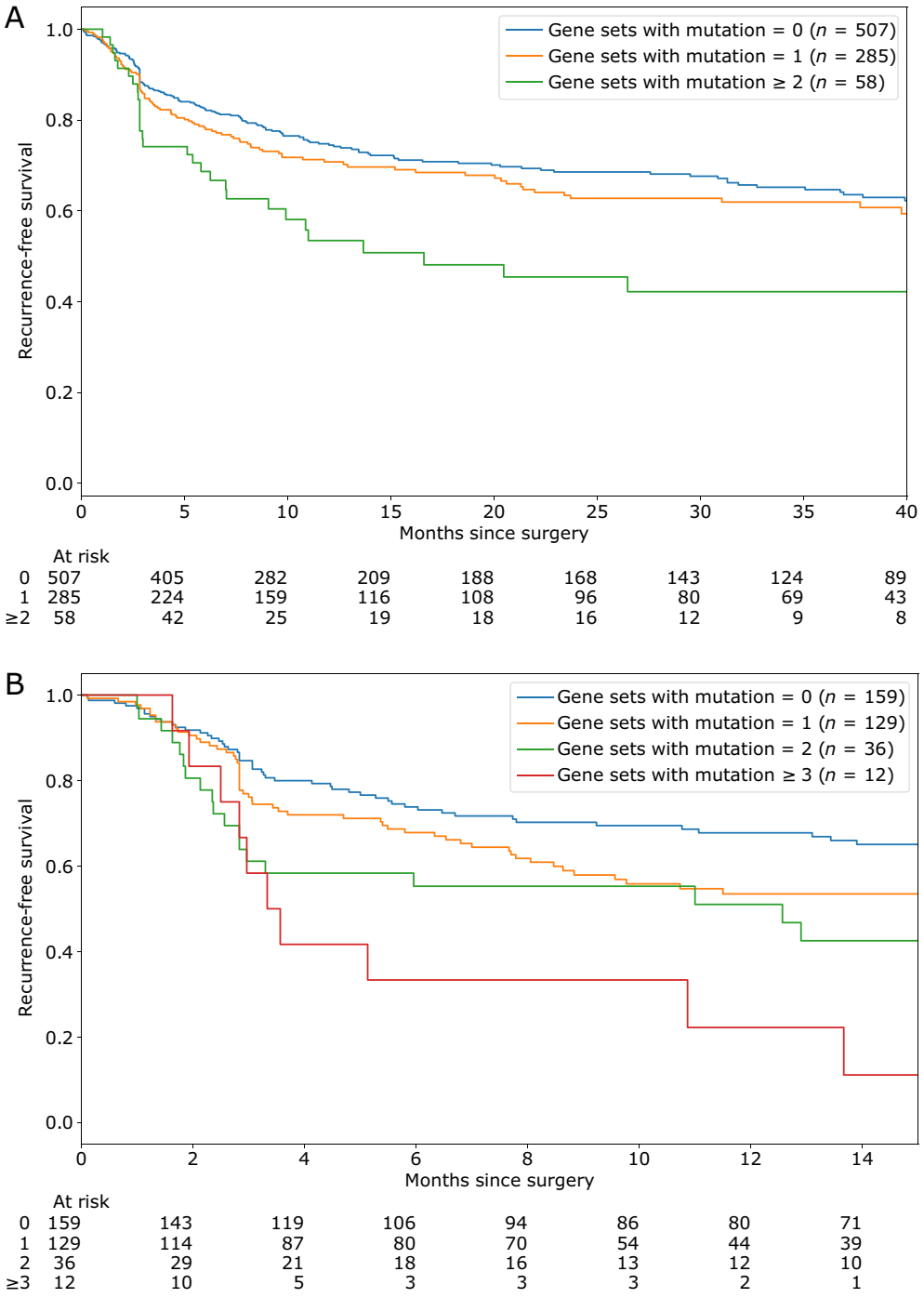

Fig. 2 – Kaplan-Meier plot showing survival probability against time in months until biochemical recurrence for (A) all samples and (B) the 336 samples in the high-Gleason subset (Gleason score >3 + 4; Gleason grade group 3–5). The impact of mutations in significant sets are subdivided by samples with mutations in multiple gene sets. Log-rank tests for each category: in Figure A—1 (p = 0.63) and ≥2 (p = 2.88 × 10^{-3}); and Figure B—1 (p = 0.27), 2 (p = 8.55 × 10^{-3}), and ≥3 (p = 3.29 × 10^{-3}).

ulation (HR = 3.49; 95% CI 1.47–8.30; p = 0.005; 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, three significant risk-elevating predictors are identified: Myc targets v2 (HR = 2.90; 95% CI 1.00–8.40; p = 0.049), Coagulation (HR = 3.53; 95% CI 1.30–9.59; p = 0.014), and additionally Hypoxia (HR = 3.18; 95% CI 1.04–9.74; p = 0.043; Cox PH; p value threshold <0.05; Table 4 and Fig. 1C). The consistent significance and same direction of coefficient of Hypoxia in patients with more advanced disease are compelling evidence that germline variations in genes within this pathway contribute 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-sample and high-Gleason sets (Fig. 2). In

both plots, we show a significant difference in survival when multiple gene sets carry PD mutations. In the all-sample analysis, 285 of 850 patients had a mutation in one significant gene set and 58 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 had mutations in two gene sets, and 12 had mutations in three or more gene sets, which was the clearest detrimental impact (Fig. 2B).

To search for individual genes mutated more frequently in patients with BCR, we calculated the odds ratio (OR) between the BCR-positive and BCR-negative groups (Supplementary Table 8). Twelve genes within the significant gene sets for all samples (*PIKFYVE*, *MYD88*, *CAB39*, *RPS6KA1*, *IRAK2*, *IL2RB*, *MSR1*, *ITGB8*, *PIK3R5*, *MMP10*, *HKDC1*, and *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*, and *RBM4*) had an OR at least two-fold higher and a mutation count difference of $\geq$2 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 Fig. 2), although patients with BCR exhibited more protein-truncating variants (Supplementary Fig. 3) than those without BCR (Supplementary Fig. 4).

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 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. To our knowledge, this study is the first to evaluate the 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 PD 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 independent predictive biomarkers of time to BCR, over and above the inclusion of clinical variables. We further demonstrate that the Hypoxia gene set was 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 a 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 the risk of BCR after RP [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 these 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 RP or RT 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 a 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). Owing to the less consistent selection of these gene sets, their importance 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 more than one of the risk-increasing gene sets, compared with both noncarriers 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 subpathologies as field defects in PrCa, and warrants further investigation [47]. Fifty-eight out of the 850 total patients were carriers of mutations in two or more of the three all-sample gene sets and 48 out of the 336 patients carried mutations within two or more of the five high-Gleason gene sets identified through our multifactor analysis.

The limitations of this study include multicohort 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-sample and the 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 an increased risk of PrCa progression and invasiveness.

5. Conclusions

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 the prediction of PrCa progression close to, or even in advance of, the point of diagnosis. This would allow clinicians to stratify and differentiate patients who are more likely to relapse, putting them on a different clinical treatment pathway comprising more radical intervention or more frequent follow-up.

***Author contributions*:** Zsofia Kote-Jarai had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Bristow, Fraser, Reimand, Schlomm, Brors, Weischenfeldt, Corcoran, Hovens, Cancel-Tassin, Cussenot, Hayes, Sorensen, Lynch, Massie, Cooper, Wedge, Brewer, Kote-Jarai, Eeles.

Acquisition of data: Bristow, Fraser, Reimand, Schlomm, Sauter, Brors, Korbel, Weischenfeldt, Waszak, Corcoran, Jung, Pope, Hovens, Cancel-Tassin, Cussenot, Loda, Sander, Lu, Hamdy, Foster, Gnanapragasam, Butler, Lynch, Massie, Woodcock, Wedge, Brewer.

Analysis and interpretation of data: Burns, Anokian, Saunders, Weischenfeldt, Hovens, Sander, Lu, Butler, Lynch, Massie, Woodcock, Cooper, Wedge, Brewer, Kote-Jarai, Eeles.

Drafting of the manuscript: Burns, Saunders, Brewer, Kote-Jarai.

Critical revision of the manuscript for important intellectual content: Burns, Anokian, Saunders, Bristow, Fraser, Reimand, Schlomm, Sauter, Brors, Korbel, Weischenfeldt, Waszak, Corcoran, Jung, Pope, Hovens, Cancel-Tassin, Cussenot, Loda, Sander, Hayes, Sorensen, Lu, Hamdy, Foster, Gnanapragasam, Butler, Lynch, Massie, Woodcock, Cooper, Wedge, Brewer, Kote-Jarai, Eeles.

Statistical analysis: Burns, Anokian, Brewer, Lynch.

Obtaining funding: Bristow, Hovens, Cussenot, Hamdy, Cooper, Wedge, Brewer, Kote-Jarai, Eeles.

Administrative, technical, or material support: None.

Supervision: None.

Other: None.

Financial disclosures: Zsofia Kote-Jarai certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: A patent application is pending/will be submitted by ICR and CRUK based on these results. P.W. is a Cancer Research Life Fellow.

***Funding/Support and role of the sponsor*:** We acknowledge support from Cancer Research UK C5047/A14835/A22530/ A17528, C309/A11566, C368/A6743, A368/A7990, and C14303/A17197 (Zsofia Kote-Jarai, S. Merson, Niall M. Corcoran, S.E., D.L., T. Dadaev, M.A., E.B., J.B., G.A., P.W., B.A.-L., Daniel S. Brewer, Colin S. Cooper, and Rosalind A. Eeles); the Dallaglio Foundation (CR-UK Prostate Cancer ICGC Project and Pan Prostate Cancer Group), PC-UK/Movember (Zsofia Kote-Jarai), the NIHR support to the Biomedical Research Centre at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust (Zsofia Kote-Jarai, N.D., S. Merson, Niall M. Corcoran, S.E., D.L., T. Dadaev, S. Thomas, M.A., E.B., Christopher S. Foster, 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., Daniel S. Brewer, Colin S. Cooper, and Rosalind A. Eeles), 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 (grant G0500966/75466; D.E.N. and Vincent Gnanapragasam): "Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT)", the Li Ka Shing Foundation (David C. Wedge and Dan J. Woodcock), Canadian Institutes of Health Research (CIHR) Project Grant (Jüri Reimand), and the Academy of Finland and Cancer Society of Finland (G.S.B.). 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 subtheme). 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 (Colin S. Cooper and Daniel S. Brewer). 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. and 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, cofunded by Movember and the Australian Federal Government (PI Niall M. Corcoran) as well as NHMRC projects grants #10413 (CIs Chris M. Hovens, Niall M. Corcoran, and R.E.A.) and #1104010 (CIs Chris M. Hovens and Niall M. Corcoran). Bernard J. Pope was supported by a Victorian Health and Medical Research Fellowship from the Department of Health and Human Services in the State of Victoria. Niall M. Corcoran 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). The PPCG project at Weill Cornell Medicine is supported by NCI P50CA211024, DoD PC160357, DoD PC180582, and the - Prostate Cancer Foundation.

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, Johns 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 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. The validation results shown are based upon data generated by the TCGA Research Network.
This project would not have been possible without the contribution of the bioinformatics platform from Synergy Lyon Cancer (Centre Léon Bérard, Lyon, France).

Peer Review Summary

Peer Review Summary and Supplementary data to this article can be found online at https://doi.org/10.1016/j.eururo.2022.05.007.

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;13:760.
[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 GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. ed. 1. 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:D766–73.
[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] Broad Institute, GitHub Repository. Picard tools; v2238. 2021. 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. Amsterdam, the Netherlands: Centrum voor Wiskunde en. Informatica 1995.
[26] Casbon J, Dougherty J. PyVCF. 2011. https://pyvcfreadthedocsio/.
[27] Reback J, McKinney W, Jbrockmendel, et al. pandas-dev/pandas: v1.3.0. Zenodo 2021.
[28] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat 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. Comput Sci Eng 2007;9:21–9.
[31] Evans C, Paolo F, Natsuo Kishimoto P, et al. cgevans/scikits-bootstrap: v1.1.0-pre.1. Zenodo 2021.
[32] Davidson-Pilon C, Kalderstam J, Jacobson N, et al. CamDavidsonPilon/Lifelines: v0.25.0. Zenodo 2020.
[33] Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9:90–5.
[34] Rusnack W, Moyer E, Taylor R. to_precision. 2013. https://bitbucketorg/william_rusnack/to-precision/src/master/.
[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.