Identifying Rare Germline Variants Associated with Metastatic Prostate Cancer Through an Extreme Phenotype Study

Yen-Yi Lin[1,27], Hamideh Sharifi Noghabi[1,27], Stanislav Volik[1], Robert Bell[1], Funda Sar[1], Anne Haegert[1], Hee Chul Chung[1], Ladan Fazli[1], Htoo Zarni Oo[1,2], Mads Daugaard[1,2], Ming-Han Kuo[3,4], Sheng-Chieh Hsu[3,4,5,§], Eddie L. Imeda[6,*], Claudio Zanettini[6,*], Lucio Queiroz[6,*], Balthasar Schlotmann[7,8,*], Kazzem Gheybi[9,*], Colin Cooper[10,11,*], Zsofia Kote-Jarai[10,12,*], Rosalind Eeles[10,12,*], Pan Prostate Cancer Group (PPCG)[10], Hsing-Jien Kung[3,4,5,13], Luigi Marchionni[6,*], Joachim Weischenfeldt[7,8,14,*], Keith D. Miller[15], Alan Rabinowitz[16,17,18], Yuzhuo Wang[1,2,19], Hai-Feng Zhang[20,21], Poul H. Sorensen[20,21], Mark S. Carey[22], Martin Gleave[1,2], Vanessa M. Hayes[9,23,24,*], William T. Gibson[25,26,28] and Colin C. Collins[1,2,28,*]

1. *Vancouver Prostate Centre, Vancouver, British Columbia, Canada*
2. *Department of Urologic Sciences, University of British Columbia, Vancouver, Canada*
3. *Research Center of Cancer Translational Medicine, Taipei Medical University, Taipei, Taiwan*
4. *Ph.D. Program for Cancer Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan*
5. *Institute of Molecular and Genomic Medicine, National Health Research Institutes, Zhunan, Miaoli County, Taiwan*
6. *Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, USA*
7. *Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark*
8. *The Finsen Laboratory, Rigshospitalet, Copenhagen, Denmark*
9. *Ancestry and Health Genomics Laboratory, Charles Perkins Centre, School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Camperdown, New South Wales, Australia*
10. *The Institute of Cancer Research, London, UK*
11. *Norwich Medical School, University of East Anglia, Norwich, UK*
12. *The Royal Marsden NHS Foundation Trust, London, UK*
13. *Department of Biochemistry and Molecular Medicine, University of California Davis, Sacramento, CA, USA*
14. *Department of Urology, Charité-Universitätsmedizin Berlin, Berlin, Germany*
15. *Hexamer Therapeutics, Pullman, WA, USA*
16. *Rural Coordination Center of British Columbia, Vancouver, British Columbia, Canada*
17. *Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada*
18. *St. Paul's Hospital, Vancouver, British Columbia, Canada*
19. *Department of Experimental Therapeutics, BC Cancer, Vancouver, British Columbia, Canada*
20. *Department of Molecular Oncology, BC Cancer, Vancouver, British Columbia, Canada*
21. *Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada*

22. Department of Obstetrics and Gynecology, University of British Columbia, Vancouver, Canada
23. School of Health Systems and Public Health, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa
24. Manchester Cancer Research Centre, University of Manchester, Manchester, UK
25. Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada
26. British Columbia Children's Hospital Research Institute, Vancouver, British Columbia, Canada
27. These authors are joint first authors
28. These authors are joint last authors

($) Current Address: The University of Texas Southwestern Medical Center , Dallas, US
(*) PPCG members

Abstract

Background: Studies of germline variants in prostate cancer (PCa) have largely focused on their connections to cancer predisposition. However, an understanding of how heritable factors contribute to cancer progression and metastasis remain limited.
Objective: To identify low frequency to rare germline nonsynonymous variants associated with increased risk for metastatic PCa (mPCa), while providing functional validation.
Design: We assembled an extreme phenotype cohort (EPC) of 52 patients diagnosed with predominantly high-grade (Gleason Score (GS) ≥ 8) PCa and > 7 years of follow-up for which localized treatment naïve tumor tissues were available. In half of the cases, the tumor had metastasized to bone, providing an even distribution of bone mPCa and non mPCa cases. Tumor and matched distant benign DNA samples were exome sequenced and analyzed for germline variants with population-wide minor allelic frequencies ≤ 2%. Findings were validated using two independent PCa germline cohorts, including a closely matched Australian study biased to aggressive disease (n = 53) and Pan Prostate Cancer Group (PPCG, n = 976). Two mPCa-promoting candidate variants in *KDM6B* and *BRCA2* were engineered into cell lines and functionalized.
Results: Germline nonsynonymous rare variants (gnsRVs) identified in 25 DNA Damage Repair (DDR) genes were significantly enriched in the mPCa patients (p=4.57e-06). Conversely, the prevalence of synonymous variants at minor allele frequencies of ≤ 2% were similar between the mPCa and non mPCa patients. The predictive power of variants in 53 non-DDR genes was validated in the Australian cohort (p=0.028) and correlated with high-risk PCa in PPCG (p=0.03). *KDM6B* K973Q showed functional significance despite being annotated as benign in ClinVar, while *BRCA2* I1962T showed sensitivity to Olaparib. In total, six EPC variants related to DNA repair or epigenetics were found to alter enzymatic activity.
Conclusions: EPCs coupled with low frequency/rare variant analyses may advance understanding of interactions between the germline and tumor in PCa. We identified a series of germline variants that were enriched among mPCa patients. Moreover, we showed that one of these variants confers a metastatic phenotype. Our findings suggest that germline testing at diagnosis may improve treatment stratification in PCa.

86 **Patient summary:** The presence of specific genetic variants among men with PCa may elevate
87 the risk of mPCa once PCa develops. Knowledge of the variant burden at time of diagnosis may
88 enable accurate stratification of some patients for aggressive therapeutic interventions.
89
90

1. Introduction

Prostate cancer (PCa) affects roughly 1.4 million men worldwide, with its lethal form, metastatic (mPCa), causing significant morbidity and resulting in ~ 397,000 annual deaths worldwide [1], this is expected to worsen significantly with an aging population [2]. Most deaths are a consequence of high-risk, aggressive disease. Patients with high-risk localized PCa have a 50-70% probability of progressing to metastasis after local therapy [3], but the basis of this aggressive phenotype remains incompletely understood [4]. Identifying the genetic factors underlying the conversion from localized to metastatic disease is crucial for improving patient stratification, prognosis, and treatment.

Recent advances in PCa genomics and molecular biology have shed light on the molecular mechanisms underpinning therapy resistance and metastasis, including alterations in the androgen receptor (AR) signaling pathway, DNA Damage repair (DDR) genes, and the activation of alternative oncogenic pathways [5, 6] and the tumor microenvironment [7-9]. Despite these advances, there remains much to learn about the interplay of germline factors contributing to the aggressive PCa phenotype. A deeper understanding of these genetic determinants is poised to unlock novel targeted therapies and more personalized treatment stratification for PCa patients.

Rare variant analyses aim to uncover the contribution of genetic variants to disease causality [10]. Typically, rare (Minor Allele Frequency(MAF) <1%) and a lesser extent low-frequency variants (1%-5%), have larger effect sizes than common variants (>5%), accounting for a substantial proportion of the disease heritability, depending on context [11]. Also, they may play unique roles in complex genetic disorders, allowing for hypothesis-free identification of disease-associated gene function [12]. Generally, a variant is rare either because it has been selected against during evolution, or it is under neutral selection [13, 14]. In a recent study, Burns et al., whole genome sequenced 850 germlines from PCa patients to identify rare germline variants predictive of time to biochemical recurrence (BCR), a surrogate for aggressive disease [15]. They found rare deleterious coding germline variants in "Hallmark" cancer gene sets consistently associated with time to BCR, including independent cohort validation.

This use of rare variant analysis in extreme phenotypic cohorts (EPCs) is a powerful approach for identifying variants that contribute to aggressive disease, with particular relevance to cancer studies such as PCa which displays considerable clinical, biological and genomic heterogeneity [16]. EPC preselection aims to enrich the study population for rare variant discovery, providing a surrogate for classical study power [17]. Furthermore, by reducing phenotypic heterogeneity, EPC studies can minimize the effect of confounding factors, such as polygenic background or environmental influences [18]. This technique reduces the “noise” and amplifies causative genetic signals, thereby enabling the use of smaller cohorts [19].

In this study, we leveraged the power of combining an EPC comprised of mPCa and non-mPCa with rare variant germline analyses. This allows a smaller cohort size for variants with a MAF ≤ 2%, which are referred to as rare varaints. The EPC includes 52 ethnically and age-matched (mean age 65 years) treatment-naïve patients who underwent surgery for high-grade disease with > 7 years follow-up with (n=26) or without (n=26) bone metastasis. Ultimately, the identification

of functionally-relevant rare variants provides valuable insights into the biological mechanisms underlying metastasis, with benefits to improve screening and rationale for targeted interventions.

2. Patients and Methods.

2.1. **EPC selection**. An extreme phenotype cohort of 52 localized and predominantly (73%) pathologic high-grade (GS ≥ 8) formalin-fixed and paraffin-embedded (FFPE) tumor specimens was assembled from treatment-naive patients with a minimum of 7 years of follow-up. Outcomes data allowed for two study arms including patients presenting either with (n=26) or without (n=26) post-surgical bone metatstisis. We note that due to the challenge of assembling such a cohort, nine tumors (five metastatic, four non-metastatic) pathologically scored as GS 4+3 were included. All were GS ≥8 based on diagnostic biopsies. Five additional non-metastatic GS 3+4 tumors were also included by the pathologist based on diagnostic GS ≥8, advanced pathological stage (T3a), > 50% positive biopsy cores or positive surgical margins. Careful review identified a single GS 3+4 lacking high risk feature in both arms. The cohort arms were further matched for positive margins 10 *versus* 13, posivitive lymph nodes 1 *versus* 2, average PSA 11.8 (4.1-31.26) *versus* 11.3 (3.6-61), average biopsy GS 8 versus 8.1, average pathological GS 8.1 versus 8.4, average positive biopsy cores 3.6 versus 3.8 (Supplementary Table 1). The patient ancestry information, determined by Somalier [20], corresponded with the predominantly Caucasian and East Asian demographics in Vancouver (32 Caucasians and 13 East Asians (further details in Table 1.))

2.2 **DNA Sequencing**. Both tumor and respective distant matched benign prostate H&E sections were selected from different FFPE tissue blocks and reviewed by a research pathologist (LF). The cancerous regions were selected based on the highest pathological presenting tumor content. For DNA extraction, two to four 5 µm sections from archived tissue blocks were collected in 1.5 mL microcentrifuge tubes. The genomic DNA from tumour and matched distant benign tissues was extracted using the Maxwell RSC DNA FFPE kit and with the Maxwell RSC instrument (Promega) according to the manufacturer's instructions. The quality and quantity of the extracted DNA were assessed with TapeStation 42100 (Agilent Technologies) and measured using the Qubit ssDNA Assay Kit (ThermoFisher), respectively. NEBNext FFPE DNA Repair Mix was used to minimize FFPE-related artifacts. Whole-exome sequencing (WES) was performed using the KAPA HyperCap workflow and SeqCap EZ MedExome panel (Roche NimbleGen) and the libraries were sequenced on the Illumina NextSeq500 (PE150). The mean sequencing depths achieved were 89X and 57X for tumor and benign samples, respectively. The identical WES protocol was applied to freshly collected cells for all clones in 2.7.

2.3 **Variant Calling and Annotation.** Sequencing data were aligned to the GRCh38 Human Genome (Ensembl Release 92) using BWA 0.7.17. PCR duplicates marked by Picard 2.18.11 and all pairs with more than five mapped mismatches (3% error rates) removed. Strelka2 2.9.10 [21] was employed for variant calling, and a variant was considered germline if it was covered by at least 15 reads with the alternative allele frequency (AAF, defined as the ratio of reads with alternative allele over the total number of reads covering the base) ≥0.2 in both tumor and matching benign tissues. Rare variants (defined here as MAF ≤ 2%) in both WGS and WES datasets were derived from population-wide gnomAD v.2.1.1 data [22]. Our analysis focused on nonsynonymous protein-altering variants (missense, nonsense, splice-site) in GENCODE v27. The

deleteriousness of all alternative allele variants was assessed using integral methods, namely fathmm-MKL [23] and CADD PHRED scores [22].

2.4 **Copy Number Variation (CNV).** We used Bionano Nexus Copy Number to determine the copy number profile for each tumour using the following settings: minimum read depth for CNV calculation – 20, minimum mapping quality – 30, minimum base quality – 20, 10^{-6} significance threshold, at least 3 calls for calling copy number status, maximum 1Mb distance between calls. The copy number status for genes was called only when a CNV region completely overlapped a gene's genomic location using Nexus Copy Number Query functionality. We also applied scarHRD [24] to calculate Homologous Recombination Deficiency (HRD) scores that are based on loss-of-heterozygosity (LOH), telomeric allelic imbalance, and large-scale state transitions. Higher HRD scores imply defective DNA Damage Response (DDR) and potential sensitivity to PARP inhibitors.

2.5 **Statistical analysis**

2.5.1 **Burden and Variance Test.** We use PPCG cohort to test the associations of detected gnsRVs with various clinical phenotypes (see Results). For each clinical phenotype, we will first test the association of every single gnsRV using the Wald test [25]. Next, collapsed all gnsRVs in a gene and test the gene-level associations using the Burden test. Finally, we aggregated the scores of all gnsRVs in a combined test for each phenotype using the sequence kernel association test (SKAT). SKAT, or variance component test in general, is known to be robust in the presence of opposite association directions and in the presence of many non-causal variants [26].

2.5.2 **Bootstrapping.** To evaluate the significance of gnsRV enrichments in metastatic *versus* non-metastatic tumors within our EPC, we employed a bootstrapping approach. The observed enrichments in the 25 genes were compared to the enrichments of gnsRVs in an equal number of randomly selected protein-coding genes. This process was iterated 10,000 times to generate robust empirical distributions for both metastatic and non-metastatic samples. Subsequently, we compared the resulting distributions of gnsRV enrichments in target genes to those in randomly selected protein-coding genes using the Wilcoxon rank-sum test with continuity correction. This statistical analysis allowed us to assess the significance of the observed enrichments in the target genes relative to the background distribution of gnsRVs in the genome, providing confidence in our findings. Analysis of this nature should be robust to any evolutionary selection pressure that is not related specifically to the phenotype under consideration (i.e. mPCa *versus* PCa).

2.6 Validation Cohorts. The Australian cohort is a closely matched alternative EPC high-risk resource (86.8% of patients with aggressive diseases), with available WGS germline data (buffy coats) from 53 genetically confirmed European ancestral PCa patients. It also includes 21 metastases (17 bone or/and distant visceral (1 treated), 4 nodal), 12 BCR, 3 PSA and local recurrence, 3 non-curative radical prostatectomies (RP), and 14 cases reported as no BCR. [27]. The PPCG cohort was comprised of WGS germline data from 976 PCa patients: including 200 with metastases out of 531 high-risk (54.6%), 305 intermediate, and 140 low-risk of metastasis.

2.7 CRISPR primer editing and isolation of KDM6B knock-in clones. KDM6B K973Q was engineered into the wildtype (WT) *KDM6B* gene in the PCa cell line LNCaP cells using

CRISPR/Cas9 prime editing[28, 29]. LNCaP cells were cultured in RPMI-1640 medium supplemented with 10% FBS and 1% P/S at 37°C in a 5% CO_2 incubator. The cells were seeded into 6-cm dishes 18 hours before transfection to ensure 70-80% confluency by the next day. The KDM6B_2917C pegRNA was cloned into the piggyPrime vector pPBT-PE2-PuroTK-pegRNA_GG [29]. For transfection, 2.5 µg of plasmid DNA was mixed in a 5:1 ratio (pegRNA:hyPBase) and delivered using 7.5 µl of PolyJet (SignaGen), following the manufacturer's protocol. After transfection, cells were selected with 2 µg/ml puromycin. The oligonucleotides used for pegRNA cloning and KDM6B sequencing are listed below. Single-cell clones were isolated, and the editing efficiency verified using Sanger sequencing. Successful editing was confirmed in two independent LNCaP clones (homozygote C3 and heterozygote C6 knock-ins) selected for further study. WES analysis of these clones revealed no additional edits in the LNCaP cells. The edited LNCaP cells were grown in media with charcoal-stripped serum.

2.8 Cell migration and invasion assays.
WT LNCaP and CRISPR edited clones C3 and C6 were grown to 90% confluency on a 96-well ImageLock plate (Essen BioScience Cat# 4379) pre-coated with 50 µL of 50 µg/mL Poly-D-Lysine. A uniform wound was created in the cell monolayer using the 96-pin IncuCyte Wound Maker Tool. The cells were then gently washed twice with PBS to remove debris, and the medium was replaced with serum-starved medium to minimize cell proliferation. The plates were incubated for 48 hours, during which the progression of wound closure was monitored and captured using the IncuCyte Live-Cell Analysis System (Sartorius). To better measure the effects of cell migration and invasion, respectively, in the context of metastasis, the Transwell/ Boyden chamber migration and invasion assays were utilized. WT LNCaP and C3 and C6 cells were seeded onto the inserts and grown using the starvation media. The insert placed in a well containing a medium enriched with growth factors or chemoattractants (e.g., serum), and the cells incubated in a humidified incubator at 37 °C to allow migration for 24 hours. After incubation, migrated cells on the lower side of inserts were fixed using methanol for analysis. Once fixed, the migrated cells (located on the outer side of insert) are stained using Crystal violet, a nuclei-specific histological dye, and quantified. The percentage of cells in each insert were measured or quantified using ImageJ.

2.9 PARPi sensitivity. To test if BRCA2 variant rs1060502377 (I1962T) attenuates BRCA2 activity in response to PARPi the variant was engineered into HEK293T cells to compare its sensitivity to Olaparib treatment with DMSO controls on cell viability using the CCK-8 assay Selleckchem #B34304 [30]. WT and BRCA2-edited HEK293T cells were seeded in 96-well plates at 5,000 cells per well and treated with 100 µM Olaparib or an equivalent volume of 0.1% DMSO as a control. After 48 hours of treatment, 10 µL of CCK-8 reagent was added to each well, and the cells were incubated for an additional 2 hours. Cell viability was assessed by measuring absorbance at 450 nm using a microplate reader, with the treated cells' absorbance compared to that of untreated control cells.

2.10 RNA sequencing and biological characterizations. RNA-Seq data were aligned to the GRCh38 release 90 reference transcriptome using STAR v2.6.0a [31]. Gene expression was quantified with HTSeq v0.11.2 [32], and differential expression was analyzed using DESeq2 v1.16.1 [33]. Genes with a fold change of ≥ 2 or ≤ 0.5 and an adjusted p-value less than 0.05 were considered significantly differentially expressed. These genes were then tested for enrichment with Hallmark gene sets, canonical pathways, and ontology gene sets from MSigDB [34].

3. Results

Focused on gnsRVs (gnomAD population frequency ≤ 2%) that alter protein sequences (missense, stop-gain, stop-loss, splice-site variants), we identified 10,854 gnsRVs present in the mPCa, yet absent in the non mPCa study arm. mPCa-specific gnsRVs were further assessed for potential pathogenicity using fathmm-MKL[23], identifying 6,080 variants with potential to impact protein function. Of these, 56 variants from 53 genes appeared among at least three mPCa patients at 11.5% penetrence.

Our germline variant calling relies on DNA extracted from FFPE histologically benign prostate tissue distant from the tumor site (usually in another prostate quadrant). This method poses two potential problems, the first being sequence artifact caused by the FFPE protocol , and the second being tumor admixture in the normal surrogate DNA. The fixing process in FFPE is known to cause deamination and other chemical modifications of DNA that require repair prior to sequencing and variant calling [35]. To determine the efficiency of repair matched flash frozen (FF) tissues from 7 patients (3 and 4 are from the metastatic and nonmetastatic groups respectively) were exome sequenced and variants called. 97.72% of the FFPE calls were observed in FF tissues indicating very efficient repair of artifacts. Moreover, most sequence artifacts occurred at sites AAF < 10% whereas our strict cut-off for variants is AAF > 20% thus eliminating most remaining artifacts. Next, we physically ruled out tumor admixture in the DNA isolated from distant benign prostate tissue using CNV analysis. First, we used random, non-matched benign DNA as a baseline to derive tumor CNV profiles. In every case, the resulting CNV profiles were identical to those obtained using matched benign DNA as a baseline. Next, we generated CNV profiles from benign samples, using other benign samples as the baseline. In all cases, these profiles were flat, while the profiles from random tumor samples showed significant variation. These results combined with our variant calling threshold cut-off of AAF > 20% rule out significant tumor admixture in our normal germline surrogate DNA.

To evaluate the validity of the pipeline and test our hypothesis, we first assessed for DDR genes [36, 37]. DDR gene variants can disrupt DNA homologous recombination (HR), leading to a BRCA1/2 genome structure [38] or BRCAness [39]. Alterations in DDR genes are also among the most common genetic events observed in metastatic PCa [40]. We selected 25 DDR genes most relevant for PCa and compared DDR gnsRVs distributions between the two arms of the cohort. We identified 71 gnsRVs (81 total 6 recurrent) exclusive to the metastatic arm and 21 gnsRVs exclusive to the nonmetastatic arm. The gnsRV burden of these 25 DDR genes is significantly higher in the metastatic arm compared to the nonmetastatic arm (p-value=4.57e-06), which is evidence of strong association with metastasis **(Fig. 1A&B).** For scientific rigor the 5 non metastatic patients having only diagnostic GS and five random metastatic patients were deleted and the burden analysis repeated (p-value=5.9e-05) remaining highly significant. As a control variant set, we investigated germline synonymous rare variants (gsRVs) between the mPCa and non-mPCa arms. Not only was there no difference between the two arms (10 *versus* 13) but there were far fewer gsRVs than gnsRVs in each arm **(Fig. 1C&D)**. This was further confirmed through bootstrapping based on gnsRVs from randomly selected protein-coding genes where the Wilcoxon rank sum test p-value for the metastatic cohort being enriched with DDR gnsRVs is <2.2e-16 whereas the non-metastatic cohort showed no significance.

A higher HRD score (≥21) implies defective DDR. To investigate if *BRCA2* variants associate with HRD, HRD scores were generated for all 52 tumors. The 26 metastatic tumors had a significantly higher combined HRD score (21.58 versus 14.62, t-test p-value=0.04), which is in line with PCa and monoallelic loss of *BRCA2* [41]. In addition, 13 of the 26 metastatic tumors had high HRD scores, compared to 7 in the nonmetastatic cohort. *BRCA2* gnsRV are 100% concordant with HRD scores ≥21 (including 1 non-metastatic tumor) suggesting impaired BRCA2 mediated DNA repair (**Fig. 1A&B**).

Using genome-wide interrogation we identified 56 gnsRVs in 53 genes **(Fig. 2)** that occurred exclusively among 3 or more patients in the metastatic arm suggesting that they are candidates for potentiating aggressive high-risk PCa. The variant burden within these 53 genes also showed a significant difference between the two patient groups (p-value = 2.98e-08), an observation that implies pre-existing germline variation within the 53 genes modified the risk of metastasis. If this is true, then at least some of these genes should be under a selection, positive or negative, in prostate tumors that develop during the lifetime of men who carry these gnsRVs. To test this, gene-level analysis was performed for a subset of the 53 genes in public PCa cohorts. Most tested genes were either mostly gained or lost suggesting biological selection in tumors (**Fig. 3A**). This provided evidence that these genes do undergo selection in PCa and collectively impact overall survival once PCa develops (p-value=1.88e-09.) **(Fig. 3A,B&C).** Notable gnsRV include *POLB* rs3136797 (P242R, gnomAD MAF= 0.011) a DDR gene and a *KDM6B* rs61764072 (K973Q, gnomAD MAF= 0.009), a histone demythlase gene.

Having developed the above Discovery Set of gnsRVs, we then sought a validation set within a similarly-enriched patient cohort. An independent and closely matched cohort of Australian PCa patients was interrogated for the presence of these 56 gnsRVs. The number of patients carrying at least one of these variants is higher than the expected carrier frequency as estimated using binomial distributions. The statistical significance for variants using our rare variant criteria (p=0.05 mets, 0.04 mets +BCR, and 0.25 non-mets) is consistently stronger in the metastatic phenotype group compared to the non-metastatic-phenotype group (**Table 3**). These results are concordant with the findings in our EPC, supporting that the recurrent variants identified in the EPC do associate with, and may contribute to the metastatic phenotype for high-risk patients.

Finally, as a second Replication Set we analyzed germline WGS data from 976 PCa patients from the PPCG consortium for the presence of gnsRVs that include 81 gnsRVs in 25 DDR genes (Fig. 1A), 56 gnsRVs exclusive to and recurrent to our metastatic arm, and 343 gnsRVs from 206 genes with gnsRVs exclusive to (1-2 per germline) the metastatic patients, resulting in 476 unique gnsRVs. 356 of those were detected as germline variants in the PPCG cohort. We found at least 3X enrichment in PPCG relative to gnomAD Caucasian population frequency. We further tested gene-level associations with clinical-pathologic features in the PPCG cohort using burden tests. Significant associations with younger age at radical prostatectomy (*FOCAD*, *CHEK2*), PSA at RP (*ACSF3, SIVA1, HPGDS*), metastasis (*SPATA9*), and high risk (*ZSWIM4, DNAJC10*) were identified for the genes that contain selected gnsRVs. The combined score of all gnsRVs using SKAT test shows a significant association with high-risk disease in PPCG (p-value < 0.03) **(Table 4)**. This suggests that gnsRVs specific to the metastatic arm in EPC also correlate to aggressive PCa in the PPCG cohort.

We also highlighted the functional and clinical significance of gnsRVs identified in our study to demonstrate the effectiveness of our strategy. The *KDM6B* gnsRV rs61764072 (K973Q, gnomAD MAF=0.009) was selected to test whether variants annotated as "benign" in ClinVar [42] could still alter gene activity. KDM6B is a histone demethylase involved in epigenetic regulation. It targets and demethylates lysine 27 on histone H3 (H3K27me3), activating gene expression [43-45]. KDM6B is also known to play a role in PCa [46]. Given its involvement in eoigentic regulation, we hypothesized that the selected variant might impact global gene expression, making its effect easier to detect. The *KDM6BgnsRV* rs61764072 (K973Q) occurs in 11.5% of germlines of the EPC metastatic arm alone. While this variant lies outside of the catalytic JMJD domain and thus may not directly affectenzyme activity, it could alter interactions with other proteins to influence chromatin binding sites and gene expression patterns(**Fig. 4A**). AlphaFold modeling [47] suggests that rs61764072 is located on the protein surface, potentially functioning as a rotamer [48] (**Fig. 4B**). Analyses utilizing several predictive algorithms suggest that rs61764072 may modify KDM6B ability to reprogram global gene expression. To test that, we engineered the *KDM6B* K973Q variant into WT LNCaP cells using CRISPR/Cas9 prime editing [28, 29]. Successful editing was verified by WES, and two independent LNCaP clones (homozygote C3 and heterozygote C6) knock-ins were selected for further analysis. WES confirmed that rs61764072/ K973Q is the only variant present in the edited *KDM6B* gene and that no additional recurrent edits were introduced outside of *KDM6B*. The KDM6B protein abundance in LNCaP-C3 and LNCaP-C6 is comparable to that of the parental clone (**Fig 4C**).

Proliferation assays revealed that clone C6 (heterozygous) grew faster that clone C3 (homozygous) (**Fig. 4D**). Scratch assays and Boyden chamber further demonstrated that the *KDM6B* K973Q confers both migratory and invasive phenotypes to LNCaP cells (**Fig. 4E-I**), with clone C3 exhibiting greater migration and invasion than C6. Remarkably, all assays showed a gene dosage effect, highlighting *KDM6B* K973Q-specific functional relevance. These findings provide evidence that *KDM6B* K973Q is functionally relevant, supporting our hypothesis that gnsRVs classified as benign or a VUS may, in a context-dependent manner, potentiate a metastatic phenotype in extant tumours.

A second rare variant (gnomAD MAF=0.00001), *BRCA2* rs1060502377 (I1962T), from the metastatic arm of the cohort, was assessed for attenuated activity in response to PARP inhibition. We showed that *BRCA2*-I1962T-edited HEK293T cells to be more sensitive to Olaparib compared to WT HEK293T and the positive control *BRCA1*-edited HEK293T cells (**Fig. 4J and K**). This suggests that *BRCA2*-I1962T may rely more heavily on PARP-mediated DNA repair mechanisms, making the patient carrier a promising candidate for PARP inhibitor treatment.

4. Discussion

While germline studies of common varaints have been instrumental in identifying genetic variants associated with increased susceptibility to PCa [49, 50], with the development of polygenic risk scores [51, 52], rare pathogenic variant identification is leading to an uptake in PCa germline testing [53, 54]. In contrast, the impact of germline variants on tumor phenotype have been largely overlooked - although this is changing [55-57]. Here, using our unique EPC high-risk metastatic

versus non metastatic PCa study design, we interrogate for rare potentially pathogenic variants. Our study overcomes notable challenges associated with such studies, including the requirement for significant and lengthy clinical follow-up (over 7 years), the ability to generate accurate sequence data from FFPE specimens, and the use of matched distant benign tissue as a surrogate source of germline DNA. While our results clearly demonstrate that the use of pathologist-reviewed distant benign prostate tissue from FFPE blocks imposes no significant limitation to our study, our approach for technical validation shows the reliability of our FFPE sequenced data.

To test the hypothesis that combining an EPC study with a rare variant analysis can identify variants that impact the metastatic phenotype of a tumor, we assessed a panel of 25 DDR genes in each arm of the study for nonsynonymous and synonymous variants. This revealed an asymmetric distribution of nonsynonymous variants between the two arms (p-value=4.57e-06) **(Fig.1A&B)**, whereas no difference was detectable for synonymous variants. Bootstrapping indicates this frequency difference was unlikely to have occurred by chance (p-value <2.2e-16).

The analysis of variants identified in our EPC across independent cohorts further strengthens the link between our findings and PCa metastasis. A strict unbiased whole genome analysis was performed identifying 56 variants in 53 genes that are exclusive to and enriched in the metastatic arms. A subset of these genes is preferentially gained or lost in public PCa cohorts and predict for poor outcome, suggesting biological selection of these genes in tumors **(Fig.3A)**. These variants were first validated in a closely-matched Australian PCa cohort, where the number of patients carrying at least one of these coding variants exceeded the expected carrier frequency, as estimated using binomial distributions. The statistical significance for gnsRV carrier number was notably stronger in the mPCa phenotype group compared to the non mPCa group [p=0.05 mets, 0.04 mets +BCR and 0.25 non-mets] **(Table 3).** We also analyzed the PPCG cohort of 967 PCa germlines including 200 mPCa and found significant enrichment for DDR variants, particularly those predicted to be pathogenic, as well as for some non-DDR genes. An association with metastasis was also identified (p=0.03). Rare variants in individual genes associated with age at RP (*FOCAD* and *CHEK2*), PSA at RP (*SIVA1* and *ACSF3*), metastasis (*SPATA9*), High Risk (*ZSWIM4* and *DNAJC10*). Three of these *FOCAD*, *SIVA1*, and *DNACJ10* had substantially higher population frequencies in the African American (AA) population. These investigations support our assertion that the recurrent germline variants identified in the mPCa EPC are linked to the tumor phenotype associated with the progression of PCa to mPCa.

The results of the gnsRV analysis in both our EPC and two independent cohorts suggests that the clinical and functional significance of variants of unknown significance (VUS) or benign gnsRVs are underestimated, and that the aggregate contributions of these gnsRVs interact with somatic mutations to influence outcome once an individual develops a PCa tumor. This combined low frequency/rare variant and extreme phenotype study identified variants in four DDR genes. These are *CHEK2* I157T [rs17879961], *POLB* P242R [rs3136797], *POLB* R137Q, and *WRN* R834C [rs3087425], all shown to alter enzymatic activity and/or impair DNA repair [58, 59]. The *POLB* variants are reported to induce genome instability, while the activity of the *CHEK2* variant is altered and segregates with familial nonmedullary thyroid cancer [60, 61]. In the PPCG cohort *CHEK2* I157T was associated with younger age at diagnosis (p-value=1.0e-06). We functionalized two variants annotated as benign and VUS in ClinVar in *KDM6B* and *BRCA2*, respectively [61]. Engineering the *KDM6B* K973Q into LNCaP dramatically altered global gene expression leading

to a migratory and invasive phenotype in a dosage-dependent manner. Edited *BRCA2* I1962T [rs1060502377] in HEK293T conferred PARPi sensitivity on HEK293T. Significantly, PARPi sensitivity was intermediate between the negative and positive controls suggesting attenuated BRCA2 activity. Further studies are required to appreciate the detailed mechanism whereby rs61764072 and rs1060502377 modify tumor phenotypes.

5. Conclusion

The results of sequence analysis and functionalization demonstrate that our strategy, which combines EPC and low frequency/rare variant analyses, has the potential to uncover clinically and functionally significant variants related to PCa metastasis, even those annotated as benign. In aggregate low frequency and rare variants may be the equivalent of "death by a thousand cuts". A potential shortcoming of this study is the small EPC size, however the results make a compelling case for much larger EPC studies aimed at delineating the contribution of the germline to shaping prostate tumor phenotypes.

Figure and Table Legends

Table 1. Ancestry of the Vancouver Prostate Centre Extreme Phenotype Cohort (EPC), Australian, and Pan Prostate Cancer Group (PPCG) cohorts.

Table 2. The allele frequency enrichment of germline nonsynonymous rare variants (gnsRVs) discovered from the VPC cohort in the independent PPCG cohort compared to healthy Caucasians (CAs) in gnomAD. We identified gnsRVs exclusive to metastatic patients in the EPC cohort and determined their population frequency within the independent PPCG cohort of 976 PCa germlines. We then compared these frequencies to those of healthy Caucasians (CA) in gnomAD. The average fold changes were 3.82, 4.42, and 5.86 for DDR gnsRVs, DDR gnsRVs predicted to be deleterious, and gnsRVs occurring at least three times, respectively. The average CA allele frequency closely matches the overall population allele frequencies reported in gnomAD. When considering only variants with ≥ 1.2-fold positive enrichment, we observed values of 8.25, 10.61, and 17.9. (t) Notes that the values of 5.86 and 17.9 in 3 recurrent gnsRVs are driven by one very rare gnsRV rs76761697 in the NIPSNAP3A gene that occurs in three PPCG germlines and in three metastatic patients from the EPC. If this gnsRV is removed, the values drop to 2 and 5-fold enrichment, respectively.

Table 3. Association of 56gnsRVs discovered from the EPC cohort in the independent Australian cohort of 53 Caucasian PCa germilnes. This independent CA cohort consisted of 21 metastatic cases (17 with bone and/or distant visceral metastases, and 4 with lymph node metastases), 12 with biochemical recurrence (BCR), and 14 without metastases (with clinical follow-up of more than 10 years). We determined the number of patients carrying at least one of these gnsRVs in each group and compared these frequencies to the expected carrier frequencies, estimated using binomial distributions. The enrichment is statistically significant for aggressive cancer groups, further supporting our hypothesis that the recurrent variants identified in the EPC cohort are associated with the progression of PCa to metastatic castration-resistant prostate cancer (mPCa).

Table 4. Representative genes containing gnsRVs that associated (nominal p-values) with phenotypes in the PPCG cohort. The FOCAD and CHEK2 gnsRV are linked to multiple cancers and may be associated with early-onset cancers. SIVA1 is an E3 ubiquitin ligase that regulates cell cycle progression, proliferation, and apoptosis in an ARF/p53-dependent manner. ACSF3 (PSA at RP, high-risk, and metastasis) is involved in fatty acids metabolism, and the gnsRV is likely pathogenic. SPATA9 associates with multiple clinical-pathologic features in the PPCG cohort, including metastasis. DNAJC10 is involved in misfolded protein degradation and is associated with poor survival in breast cancer and germline variants associated with aggressive PCa. *FOCAD, SIVA1, and DNAJC10 have higher minor allele frequency in African American populations than those of CAs [62].*

Figure 1. DDR gene gnsRVs and germline rare sysnonymous variants (gsRVs) in both groups of the EPC. Patterns of 102 DDR gnsRVs exclusive to either the metastatic **(A)** or non-metastatic **(B)** arm of the extreme phenotype cohort: 81 exclusive to mets **(A)** and 21 exclusive to non-mets **(B).** Green cells: HRD scores ≥ 21. ✓: gnsRV. ✗: gnsRV predicted to be deleterious. ✓ or ✗ indicate compound gnsRVs. Yellow cells: allele frequency of the gnsRV in the matched tumor is at least 1.5-fold the allele frequency in benign suggesting positive selection. The column "M" and "N" are the number of metastatic and nonmetastatic tumors with copy number aberrations detected at the gene locus. BRCA2 gnsRVs are 100% concordant with high HRD scores. The average HRD score is 21.58 vs. 14.62 (t-test p-value=0.04) in metastatic vs. nonmetastatic patients, respectively. There are fewer patients with HRD scores ≥21 (7 versus 13) in the non-metastatic cohort. The WRN gnsRV in patients 66 and 36 is known to impair helicase activity (PMID: 15489508). It is enriched 153-fold (2 out of 26 versus 0.0005 in gnomAD) over the general population. The POLB and CHEK2 variants are known to be deleterious(PMID: 23144635, 35127508). ***DDR gnsRV are enriched in metastatic vs non-metastatic tumors (p-value=4.57e-06)*** *using a Wilcoxon rank sum test) supporting positive selection for these gnsRVs in the metastatic arm.* No significant difference was detected for 11 synonymous RVs exclusive to mets in (C) and 13 exclusive to non-mets in (D).

Figure 2. 56 gnsRVs specific to and enriched in 26 localized metastatic tumors from 53 genes. Each gnsRV is recurrent in at least 3 (out of 26) tumors and is predicted or known to be deleterious. ✓: rare variant. ✗: predicted deleterious rare variant by fathmm-MKL. ✓ or ✗ indicate rare compound variants. Yellow cells: allele frequency of the gnsRV in the matched tumor is at least 1.5-fold the allele frequency in benign suggesting positive selection. The column "M" and "N" are the number of metastatic and nonmetastatic tumors with copy number aberrations detected at the gene locus. Genes in yellow have their gnsRVs reported in COSMIC. The enrichment of the gnsRVs of in our VPC cohort ranges from > 10-fold to >200-fold over gnomAD population frequencies. The number of gnsRVs per patient is shown at the bottom. The average is ~ eight. ***In the non-metastatic cohort, only LAMA5 and PKD1L2 had recurrent gnsRVs meeting our filtration requirement. These are shaded in gray.***

Figure 3. Genomic Analysis of Public PCa Cohorts for Genes with Recurrent gnsRVs in the EPC. Gene-level analysis was performed on 12 genes, selected from Figure 2, across 4 public prostate cancer (PCa) cohorts to determine if these genes are under selective pressure in PCa and

collectively impact overall survival after disease onset. (A) A subset of genes with recurrent gnsRVs from Figure 2 was analyzed using cBioPortal for copy number variant analysis. Among 4 selected PCa cohorts, most genes are either predominately gained (e.g., CRISPLD1) or lost (e.g., PKD1L2) and thus appear to be under selection. (B) Copy number aberrations of 12 genes increase as tumors progress from adenocarcinoma to metastasis. This suggests that as tumors progress from relatively low risk of metastasis to a high risk of metastasis, these genes undergo positive selection. (C) Survival analysis of 12 genes among 4 PCa cohorts. Taken together, copy number aberrations at these genes predict a poor outcome with a Log-rank Test P-value of 1.88e-9. This suggests that these genes may play a role in the aggressive phenotype and impact patient outcomes.

Figure 4. Functional assays and sequencing analysis of 2 gnsRVs discovered in the metastatic arm of EPC: KDM6B gnsRV rs61764072 (K973Q) in LNCaP and BRCA2 gnsRV rs1060502377(I1962T) in HEK293T. We edited two gnsRVs discovered in the EPC metastatic arm for functionalization. (A) KDM6B protein structure and location of the target gnsRV. KDM6B is a histone demethylase that demethylates lysine 27 on histone H3 (H3K27me3) thereby activating gene expression. The KDM6B gnsRV rs61764072 (K973Q) occurs in 11.5% of germlines of the extreme phenotype cohort metastatic arm (Fig. 2) and has a minor allele frequency approximately two-fold higher in the PPCG validation cohort compared to gnomAD. We hypothesized that KDM6B (K973Q) might impact global gene expression, making an effect straightforward to detect. rs61764072 (K973Q) lies outside the catalytic jmjd domain and may not directly impact enzyme activity, but it could influence interactions with other proteins, thereby affecting chromatin binding sites and gene expression patterns.(B) Despite its benign status in ClinVar, modeling using AlphaFold suggests an extrinsic positioning of rs61764072 (K973Q) on the protein surface, and several prediction algorithms predict it to be deleterious. To test this, we edited the K973Q variant into the KDM6B locus in wildtype LNCaP cells using CRISPR/Cas9 prime editing. (C) Two independent LNCaP clones, homozygote C3 and heterozygote C6 knock-ins, were selected for further study (successful editing in both was confirmed by whole exome sequencing.) The expression levels of KDM6B of LNCaP-C3 and LNCaP-C6 are comparable to that of the parental clone. (D) The growth rate of homozygous C3 is lower than that of heterozygous C6, with C6 growing slower than the WT. (E) Scratch assays reveal a mild migratory phenotype in edited LNCaP cells compared to WT. (G) Quantification shows C3 is more migratory than C6. (G,H,I) Boyden chamber assays reveal approximately a 4-fold increase in migration and invasion compared to WT LNCaP cells and again clone C3 (homozygous) displays stronger phenotypes than clone C6 (heterozygous) suggesting a dosage effect. (J,K) The BRCA2 gnsRV rs1060502377 (I1962T) (gnomAD MAF: 10e-7) confers sensitivity to PARP inhibitors, despite being classified as of unknown significance in ClinVar.

Ackonwledgement

C.S.C would like to thank those men with prostate cancer and the subjects who have donated their time and samples for this study. Part of the research presented in this paper was carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia. We acknowledge support from Cancer Research UK C5047/A29626, C5047/A22530, C309/A11566, C368/A6743, A368/A7990, and C14303/A17197; We acknowledge support from Prostate Cancer UK (MA-TIA23-002, MA-ETNA19-003, TLD-CAF22-011), the Dallaglio Foundation; and Prostate Cancer Research. We also acknowledge support from the Bob Champion Cancer Research Trust, Movember, Big C

Cancer Charity, the Masonic Charitable Foundation, the King Family, the Stephen Hargrave Foundation, Alan Boswell Group, The Bob Willis Fund and the Bedford Memorial Trust. R.A.E thank those men with prostate cancer and the subjects who have donated their time and their samples to the consortium. We also acknowledge support of the research staff in the groups who so carefully curated the samples and the follow-up data. We acknowledge support from Cancer Research UK C5047/A14835/A22530/ A17528, C309/A11566, C368/A6743, A368/A7990, C14303/A17197, Prostate Cancer UK (MA-TIA23-002, the Dallaglio Foundation (CR-UK Prostate Cancer ICGC Project and Pan Prostate Cancer Group), PC-UK/Movember, the NIHR support to The Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.Genomic sequencing and interrogation of Southern African Prostate Cancer Study (SAPCS) data was supported by the Ancestry and Health Genomics Laboratory at the University of Sydney through National Health and Medical Research Council (NHMRC) of Australia funding, including a Project Grant (APP1165762) and Ideas Grants (APP2001098, APP2010551), and U.S.A. Congressionally Directed Medical Research Programs (CDMRP) Prostate Cancer Research Program (PCRP) funding, including an Idea Development Award (PC200390, TARGET Africa) and HEROIC Consortium Award (PC210168 and PC230673, HEROIC PCaPH Africa1K). Further support for SAPCS analytical costs was provided by the U.S.A. National Institute of Health (NIH) National Cancer Institute (NCI) Award (1R01CA285772-01) and a U.S.A. Prostate Cancer Foundation (PCF) Challenge Award (2023CHAL4150). V. M. H was further supported by the Petre Foundation via the University of Sydney Foundation (Australia). J. W was supported by the Danish Cancer Society #R147-A9843, Danish Cancer Society #R374-A22518, Danish Council for Independent Research #8020-00282, Danish Council for Independent Research #3101-00177A, Novo Nordisk Foundation #NNF200C0060141, Sygeforsikringen Danmark # 2022-0198. C.S.C, D.C.W, R.A.E, D.B and Z.K-J made a substantial contribution to the organisation and conduct of the study and critiqued the output for important intellectual content. H.J.K acknowledges the support of MOHW 114-TDU-B222-134002, NSTC-114-2634-039-001 and 114-2326-B-038-001 W.T.G gratefully acknowledges funding from the BC Children's Hospital Research Institute IGAP award. C.C.C. acknowledges the support of Canadian Institue of Health Research PJT-175238 and PJT-153073, and the support of MITACS through the MITACS Accelerate program IT19627, in partnership with PHEMI Systems Corporation and the University of British Columbia.

Competing Interests

R. A. E holds Honoraria from GU-ASCO, Janssen, University of Chicago, Dana Farber Cancer Institute USA as a speaker. Educational honorarium from Bayer and Ipsen, member of external expert committee to Astra Zeneca UK and Member of Active Surveillance Movember Committee. Member of the SAB of Our Future Health. Undertakes private practice as a sole trader at The Royal Marsden NHS Foundation Trust and 90 Sloane Street SW1X 9PQ and 280 Kings Road SW3 4NX, London, UK. M.S.C received consultancy fees and honoraria from Hexamer Therapeutics, Verastem Oncology, and AbbVie, and holds shares in Hexamer Therapeutics, Dental Today, aiGENE, Illumina, Bausch Health, BioAtla, Exact Sciences, and Zymeworks.

Abbreviations

AR: androgen receptor
BCR: biochemical recurrence

CNV: copy number variation
DDR: DNA damage repair
EPC: extreme phenotype cohort
FF: flash frozen
FFPE: formalin-fixed and paraffin-embedded
GS: Gleason Score
HR: homologous recombination
HRD: homologous recombination deficiency
LOH: loss-of-heterozygosity
MAF: minor Allele Frequency
PCa: prostate cancer
PSA: prostate-Specific Antigen
WES: whole-exome sequencing
WT: wildtype
gnsRV: germline nonsynonymous rare variants
mPCa: metastatic PCa

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians. 2021;71:209-49.
[2] James ND, Tannock I, N'Dow J, Feng F, Gillessen S, Ali SA, et al. The Lancet Commission on prostate cancer: planning for the surge in cases. The Lancet. 2024;403:1683-722.
[3] D'Amico AV, Chen M-H, Catalona WJ, Sun L, Roehl KA, Moul JW. Prostate cancer-specific mortality after radical prostatectomy or external beam radiation therapy in men with 1 or more high-risk factors. Cancer. 2007;110:56-61.
[4] Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, et al. Genomic and phenotypic heterogeneity in prostate cancer. Nature Reviews Urology. 2021;18:79-92.
[5] Howard N, Clementino M, Kim D, Wang L, Verma A, Shi X, et al. New developments in mechanisms of prostate cancer progression. Seminars in Cancer Biology. 2019;57:111-6.
[6] Bancroft EK, Page EC, Brook MN, Thomas S, Taylor N, Pope J, et al. A prospective prostate cancer screening programme for men with pathogenic variants in mismatch repair genes (IMPACT): initial results from an international prospective study. The Lancet Oncology. 2021;22:1618-31.
[7] Mo F, Lin D, Takhar M, Ramnarine VR, Dong X, Bell RH, et al. Stromal Gene Expression is Predictive for Metastatic Primary Prostate Cancer. European Urology. 2018;73:524-32.
[8] Tyekucheva S, Bowden M, Bango C, Giunchi F, Huang Y, Zhou C, et al. Stromal and epithelial transcriptional map of initiation progression and metastatic potential of human prostate cancer. Nature Communications. 2017;8:420.
[9] Pakula H, Omar M, Carelli R, Pederzoli F, Fanelli GN, Pannellini T, et al. Distinct mesenchymal cell states mediate prostate cancer progression. Nature Communications. 2024;15:363.
[10] Gibson G. Rare and common variants: twenty arguments. Nature reviews Genetics. 2012;13:135-45.

[11] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747-53.
[12] Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. Journal of Human Genetics. 2021;66:11-23.
[13] Lu Y-F, Goldstein DB, Angrist M, Cavalleri G. Personalized Medicine and Human Genetic Diversity. Cold Spring Harbor Perspectives in Medicine. 2014;4.
[14] Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proceedings of the National Academy of Sciences. 2012;109:17758-64.
[15] Burns D, Anokian E, Saunders EJ, Bristow RG, Fraser M, Reimand J, et al. Rare Germline Variants Are Associated with Rapid Biochemical Recurrence After Radical Prostate Cancer Treatment: A Pan Prostate Cancer Group Study. European urology. 2022;82:201-11.
[16] Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40:695-701.
[17] Li D, Lewinger JP, Gauderman WJ, Murcray CE, Conti D. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genetic epidemiology. 2011;35:790-9.
[18] Amanat S, Requena T, Lopez-Escamez JA. A Systematic Review of Extreme Phenotype Strategies to Search for Rare Variants in Genetic Studies of Complex Disorders. Genes (Basel). 2020;11.
[19] Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. American journal of human genetics. 2014;95:5-23.
[20] Pedersen BS, Bhetariya PJ, Brown J, Kravitz SN, Marth G, Jensen RL, et al. Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. Genome Med. 2020;12:62.
[21] Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods. 2018;15:591-4.
[22] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434-43.
[23] Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics. 2015;31:1536-43.
[24] Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. npj Breast Cancer. 2018;4:16.
[25] Pirie A, Wood A, Lush M, Tyrer J, Pharoah PD. The effect of rare variants on inflation of the test statistics in case-control analyses. BMC Bioinformatics. 2015;16:53.
[26] Wu Michael C, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. The American Journal of Human Genetics. 2011;89:82-93.
[27] Jaratlerdsiri W, Jiang J, Gong T, Patrick SM, Willet C, Chew T, et al. African-specific molecular taxonomy of prostate cancer. Nature. 2022;609:552-9.
[28] Petri K, Zhang W, Ma J, Schmidts A, Lee H, Horng JE, et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. Nature Biotechnology. 2022;40:189-93.

[29] Wolff JH, Haldrup J, Thomsen EA, Andersen S, Mikkelsen JG. piggyPrime: High-Efficacy Prime Editing in Human Cells Using piggyBac-Based DNA Transposition. Frontiers in genome editing. 2021;3.
[30] Zhang S, Li E, Liu Z, Shang H, Chen Y, Jing H. Nanoparticle-based Olaparib delivery enhances its effect, and improves drug sensitivity to cisplatin in triple negative breast cancer. Journal of Drug Delivery Science and Technology. 2022;76:103731.
[31] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15-21.
[32] Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166-9.
[33] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
[34] 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.
[35] Lim J, Heo D-h, Kim S-G, Kim M, Park J, Park M, et al. Optimizing clinical genomic analysis of FFPE specimens for enhanced data quality, reduced artefacts, and ensured reliability2024.
[36] Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Rep. 2018;23:239-54 e6.
[37] Lange SS, Takata K, Wood RD. DNA polymerases and cancer. Nat Rev Cancer. 2011;11:96-110.
[38] Marquard AM, Eklund AC, Joshi T, Krzystanek M, Favero F, Wang ZC, et al. Pan-cancer analysis of genomic scar signatures associated with homologous recombination deficiency suggests novel indications for existing cancer drugs. Biomark Res. 2015;3:9.
[39] Lord CJ, Ashworth A. BRCAness revisited. Nature Reviews Cancer. 2016;16:110-20.
[40] Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018;50:645-51.
[41] Lotan TL, Kaur HB, Salles DC, Murali S, Schaeffer EM, Lanchbury JS, et al. Homologous recombination deficiency (HRD) score in germline BRCA2- versus ATM-altered prostate cancer. Mod Pathol. 2021;34:1185-93.
[42] Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 2020;48:D835-D44.
[43] Yıldırım-Buharalıoğlu G. Lysine Demethylase 6B Regulates Prostate Cancer Cell Proliferation by Controlling c-MYC Expression. Molecular pharmacology. 2022;101:106-19.
[44] Macedo-Silva C, Miranda-Goncalves V, Tavares NT, Barros-Silva D, Lencart J, Lobo J, et al. Epigenetic regulation of TP53 is involved in prostate cancer radioresistance and DNA damage response signaling. Signal Transduct Target Ther. 2023;8:395.
[45] Pecci V, Troisi F, Aiello A, De Martino S, Carlino A, Fiorentino V, et al. Targeting of H19/cell adhesion molecules circuitry by GSK-J4 epidrug inhibits metastatic progression in prostate cancer. Cancer Cell Int. 2024;24:56.
[46] Cao Z, Shi X, Tian F, Fang Y, Wu JB, Mrdenovic S, et al. KDM6B is an androgen regulated gene and plays oncogenic roles by demethylating H3K27me3 at cyclin D1 promoter in prostate cancer. Cell Death & Disease. 2021;12:2.

[47] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583-9.
[48] Haddad Y, Adam V, Heger Z. Rotamer Dynamics: Analysis of Rotamers in Molecular Dynamics Simulations of Proteins. Biophys J. 2019;116:2062-72.
[49] Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. Nature Reviews Cancer. 2017;17:692-704.
[50] Benafif S, Kote-Jarai Z, Eeles RA, Consortium obotP. A Review of Prostate Cancer Genome-Wide Association Studies (GWAS). Cancer Epidemiology, Biomarkers & Prevention. 2018;27:845-57.
[51] Xiang R, Kelemen M, Xu Y, Harris LW, Parkinson H, Inouye M, et al. Recent advances in polygenic scores: translation, equitability, methods and FAIR tools. Genome Medicine. 2024;16:33.
[52] Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. Genome Medicine. 2020;12:44.
[53] Giri VN, Knudsen KE, Kelly WK, Cheng HH, Cooney KA, Cookson MS, et al. Implementation of Germline Testing for Prostate Cancer: Philadelphia Prostate Cancer Consensus Conference 2019. Journal of Clinical Oncology. 2020;38:2798-811.
[54] Giri VN, Morgan TM, Morris DS, Berchuck JE, Hyatt C, Taplin M-E. Genetic testing in prostate cancer management: Considerations informing primary care. CA: a cancer journal for clinicians. 2022;72:360-71.
[55] Saunders EJ, Kote-Jarai Z, Eeles RA. Identification of Germline Genetic Variants that Increase Prostate Cancer Risk and Influence Development of Aggressive Disease. Cancers. 2021;13:760.
[56] Burns D, Anokian E, Saunders EJ, Bristow RG, Fraser M, Reimand J, et al. Rare Germline Variants Are Associated with Rapid Biochemical Recurrence After Radical Prostate Cancer Treatment: A Pan Prostate Cancer Group Study. European Urology. 2022;82:201-11.
[57] Saunders EJ, Dadaev T, Brook MN, Wakerell S, Govindasami K, Rageevakumar R, et al. Identification of Genes with Rare Loss of Function Variants Associated with Aggressive Prostate Cancer and Survival. European Urology Oncology. 2024;7:248-57.
[58] Yamtich J, Nemec AA, Keh A, Sweasy JB. A Germline Polymorphism of DNA Polymerase Beta Induces Genomic Instability and Cellular Transformation. PLOS Genetics. 2012;8:e1003052.
[59] Sobol RW. Genome Instability Caused by a Germline Mutation in the Human DNA Repair Gene POLB. PLOS Genetics. 2012;8:e1003086.
[60] Kaczmarek-Ryś M, Ziemnicka K, Hryhorowicz ST, Górczak K, Hoppe-Gołębiewska J, Skrzypczak-Zielińska M, et al. The c.470 T > C CHEK2 missense variant increases the risk of differentiated thyroid carcinoma in the Great Poland population. Hereditary cancer in clinical practice. 2015;13:8.
[61] Pires C, Marques IJ, Valério M, Saramago A, Santo PE, Santos S, et al. CHEK2 germline variants identified in familial nonmedullary thyroid cancer lead to impaired protein structure and function. Journal of Biological Chemistry. 2024;300.
[62] Zins K, Frech B, Taubenschuss E, Schneeberger C, Abraham D, Schreiber M. Association of the rs1346044 Polymorphism of the Werner Syndrome Gene RECQL2 with Increased Risk and Premature Onset of Breast Cancer. International journal of molecular sciences. 2015;16:29643-53.

777

TABLES and FIGURES

Table 1. Ancestry of the Vancouver Prostate Centre Extreme Phenotype Cohort (EPC), Australian, and Pan Prostate Cancer Group (PPCG) cohorts.

Ancestry	Cohort Size	Caucasian	African American	African	East Asian and Others	South Asian
VPC EPC	52	61.5%	0.0%	0.0%	34.7%	3.8%
Australian Cohort	53	100.0%	0.0%	0.0%	0.0%	0.0%
PPCG Cohort	976	94.2%	0.0%	2.8%	1.8%	1.2%

Table 2. The allele frequency enrichment of germline nonsynonymous rare variants (gnsRVs) discovered from the VPC cohort in the independent PPCG cohort compared to healthy Caucasians (CAs) in gnomAD.

	Number gnsRVs	Total # gnsRVs (PPCG)	Average Allele Frequency (gnomAD Caucasian)	Average Fold Change In Allele Frequency (PPCG/ gnomAD Caucasians)	Average Fold Change In Allele Frequency (PPCG/gnomAD Caucasians) **>=1.2 enrichment**
Variant-Centric: >= 3 Recurrent gnsRVs*	56	51	0.012	5.86[t] (0.16 ~ 199) (21 variants enriched)	17.9[t] (15 variants)
DDR gnsRVs	81	38	0.004	3.82 (0.3 ~ 56.8) (16 variants enriched)	8.25 (13 variants)
DDR Predicted Deleterious	45	21	0.004	4.42 (0.45 ~ 56.8) (10 variants enriched)	10.61 (7 variants)

Table 3. Association of 56gnsRVs discovered from the EPC cohort in the independent Australian cohort of 53 Caucasian PCa germilnes.

	Cases with gnsRVs	Number of Patients	p-value
Met	9	17	0.05063704
Nodal	2	4	0.3686052
BCR	6	12	0.3072268
NonMet	5	14	0.2493105
Others	3	6	0.2772393
Met+BCR	14	29	0.03995316
Met+Nodal+BCR	16	33	0.02779074

Table 4. Representative genes containing gnsRVs that associated (nominal p-values) with phenotypes in the PPCG cohort.

Phenotype	Gene	P-value	Caucasian MAF in gnomAD	African American MAF in gnomAD
Age at RP	*FOCAD*	.0000001	.0001	.0007
	CHEK2	.00001	.005	.00008
PSA at RP	SIVA1	10e-80	.00009	.0328
	ACSF3	10e-80	.001	.0005
Metastasis	SPATA9[1]	.005	.0157	.0022
High-Risk	*ZSWIMJ*[2]	.001	.0257	.004
	DNAJC10	.001	.005	.027
High-Risk	Aggregated gnsRVs	.03	N/A	N/A

Figure 1. DDR gene gnsRVs and germline rare sysnonymous variants (gsRVs) in both groups of the EPC.

(A)

26 Metastasized Tumors

Patient	69	66	64	15	65	10	67	13	9	17	41	16	33	18	68	11	26	1	37	63	27	6	34	25	36	35	M	N
HRD	50	47	44	35	32	29	28	26	25	25	25	22	22	19	19	17	17	15	15	14	13	9	5	3	3	2		
ATM		✗							✓				✗							✓		✓		✗			5	2
ATR										✗																	3	4
BRCA1																		✓									9	4
BRCA2		✗	✓	✓	✓				✗																		11	3
BRIP1				✗			✗															✓				✓	3	2
CHEK2												✗	✗							✗							1	1
DDX11			✓						✗													✗					2	0
FANCD2									✓												✗						1	2
FMN2						✗															✗						10	4
EXO1														✗		✓					✓		✓				8	3
MET									✗		✗				✗								✗				4	3
MSH2							✗																				3	0
MSH3																		✓	✗	✗							7	2
MYC				✗					✗																		8	3
NBN																	✗										6	5
PALB2											✓																0	0
POLB	✗				✗															✗	✗						10	0
POLG															✓	✓	✓		✗								2	0
RAD51C																				✗							3	0
RIF1	✓				✓		✗										✗						✗	✗			5	4
TEX15								✓										✓		✓			✓				13	4
TP53							✗											✓									6	4
TP53BP1													✓		✓								✓				3	2
TRAIP			✗	✗																							3	2
WRN		✗												✗											✗		13	3
Total	2	4	3	5	3	1	5	1	7	1	2	1	3	1	5	1	4	4	3	7	5	3	6	2	1	1		

(C)

26 Metastasized Tumors

Patient	69	66	64	15	65	10	67	13	9	17	41	16	33	18	68	11	26	1	37	63	27	6	34	25	36	35	M	N
HRD	50	47	44	35	32	29	28	26	25	25	25	22	22	19	19	17	17	15	15	14	13	9	5	3	3	2		
ATM									✓														✓				5	2
ATR										✓																	3	4
BRCA1																											9	4
BRCA2																											11	3
BRIP1																											3	2
CHEK2																											1	1
DDX11																											2	0
FANCD2																											1	2
FMN2																											10	4
EXO1																											8	3
MET																											4	3
MSH2																					✓						3	0
MSH3																											7	2
MYC																											8	3
NBN																											6	5
PALB2																											0	0
POLB																											10	0
POLG																											2	0
RAD51C																											3	0
RIF1																											5	4
TEX15																											13	4
TP53													✓														6	4
TP53BP1													✓					✓									3	2
TRAIP												✓															3	2
WRN		✓																			✓						13	3
Total	0	1	0	0	0	0	0	0	2	1	0	1	2	0	0	0	0	1	0	0	2	0	1	0	0	0		

(B)

26 Non-Metastasized Tumors

Patient	19	32	24	14	56	60	4	2	7	28	54	5	61	58	57	51	52	12	21	29	20	22	53	3	59	55	M	N
HRD	43	28	27	25	25	22	21	19	19	19	19	18	17	13	12	11	10	7	6	5	4	4	3	2	1	0		
ATM		✗							✗											✗			✓				5	2
ATR								✗																			3	4
BRCA1																									✗		9	4
BRCA2	✓																										11	3
BRIP1																											3	2
CHEK2											✓																1	1
DDX11																		✗									2	0
FANCD2																											1	2
FMN2																											10	4
EXO1															✓					✓							8	3
MET																											4	3
MSH2													✗	✗													3	0
MSH3																											7	2
MYC																											8	3
NBN															✓												6	5
PALB2																		✓					✗				0	0
POLB																											10	0
POLG																											2	0
RAD51C																											3	0
RIF1																			✗						✓		5	4
TEX15																											13	4
TP53																											6	4
TP53BP1												✓															3	2
TRAIP																											3	2
WRN									✓																		13	3
Total	1	1	0	0	0	0	0	1	2	0	1	0	3	1	2	0	0	2	1	2	0	0	1	1	2	0		

(D)

26 Non-Metastasized Tumors

Patient	19	32	24	14	56	60	4	2	7	28	54	5	61	58	57	51	52	12	21	29	20	22	53	3	59	55	M	N
HRD	43	28	27	25	25	22	21	19	19	19	19	18	17	13	12	11	10	7	6	5	4	4	3	2	1	0		
ATM		✓																									5	2
ATR																								✓			3	4
BRCA1																											9	4
BRCA2																			✓				✓				11	3
BRIP1																											3	2
CHEK2																											1	1
DDX11																											2	0
FANCD2																											1	2
FMN2																											10	4
EXO1																			✓								8	3
MET																											4	3
MSH2			✓																								3	0
MSH3																											7	2
MYC																											8	3
NBN													✓														6	5
PALB2				✓																							0	0
POLB																											10	0
POLG																										✓	2	0
RAD51C																											3	0
RIF1																											5	4
TEX15			✓																								13	4
TP53																											6	4
TP53BP1																			✓								3	2
TRAIP																											3	2
WRN												✓															13	3
Total	0	1	0	2	1	0	0	0	0	0	0	1	1	0	0	0	0	0	3	0	1	0	0	2	0	1		

Figure 2. 56 gnsRVs specific to and enriched in 26 localized metastatic tumors from 53 genes.

26 Metastasized Tumors																										
Patient	69	66	64	15	65	10	67	13	9	17	41	16	33	18	68	11	26	1	37	63	27	6	34	25	36	35	M	N
HRD	50	47	44	35	32	29	28	26	25	25	25	22	22	19	19	17	17	15	15	14	13	9	5	3	3	2		
AHNAK				✗						✗																✗	3	1
AKAP11						✗															✗		✗				14	5
CALCA														✗					✗			✗					0	0
CAMSAP2					✗										✗				✗								2	0
CAPN9													✓							✓		✓					5	3
CCDC112						✓											✓				✓		✓				6	4
CENPJ		✗	✗												✗												3	0
CLTCL1													✗							✗		✗					1	0
COL4A3								✗												✗				✗			5	0
CRISPLD1									✗	✗		✗															8	3
DNAJC10			✗					✗							✗			✗			✗						0	0
ERBIN													✗							✗						✗	6	4
FOCAD			✗			✗																	✗				1	2
FREM1			✓			✓	✓										✓										2	0
GJD4	✗																		✗					✗			4	1
HCK				✗										✗				✗						✗			0	0
IFIT2		✗								✗																✗	6	4
IL17RD								✗							✗	✗											5	2
ING1;ING1							✗														✗		✗				0	0
KDM6B									✗													✗			✗		8	5
KIAA1671				✓																✓						✓	1	1
KMT2D			✗																				✗		✗		5	6
KMT2E							✗			✗								✗					✗				0	0
LAMA5																				✗		✗				✗	1	3
LAMB1									✗									✗								✗	0	0
LSM14A	✗										✗														✗		0	0
MASP1					✗														✗						✗		2	6
MCM2	✗			✗												✗											4	5
MUC5AC	✓				✓									✓										✓			13	8
NECAB3		✗			✗					✗																	2	4
NID2										✗				✗					✗								0	0
NIPSNAP3A			✗				✗														✗						5	3
NPC1										✗	✗									✗							2	3
PEPD						✗	✗										✗				✗						2	1
PKD1L2				✗					✗			✗															0	0
PLEKHM3				✗						✗									✗								2	0
POLB	✗				✗															✗							10	0
PPM1E				✗										✗										✗			3	0
PUS1						✗									✗		✗										4	4
RFXANK							✗										✗				✗						1	1
RP1L1											✗	✗													✗		11	3
S1PR2	✗			✗								✗							✗								1	0
SCARF1				✗										✗	✗				✗								0	0
SPRY3										✗									✗			✗					4	5
SRP72						✗	✗									✗	✗						✗				5	2
SUSD1	✗			✗															✗								4	4
TENM4																				✗		✗			✗		0	1
TLL2						✗												✗	✗								3	2
TRAPPC12								✗						✗								✗					1	0
TTC38									✗				✗	✗													3	0
WNT10A							✗										✗						✗				3	2
ZC3H7A				✗					✗			✗															5	3
ZSWIM4						✗												✗	✗								0	1
Total	7	3	6	11	7	10	8	4	6	9	3	5	4	10	6	3	7	7	13	9	7	8	8	7	6	6		

Figure 3. Genomic Analysis of Public PCa Cohorts for Genes with Recurrent gnsRVs in the EPC.

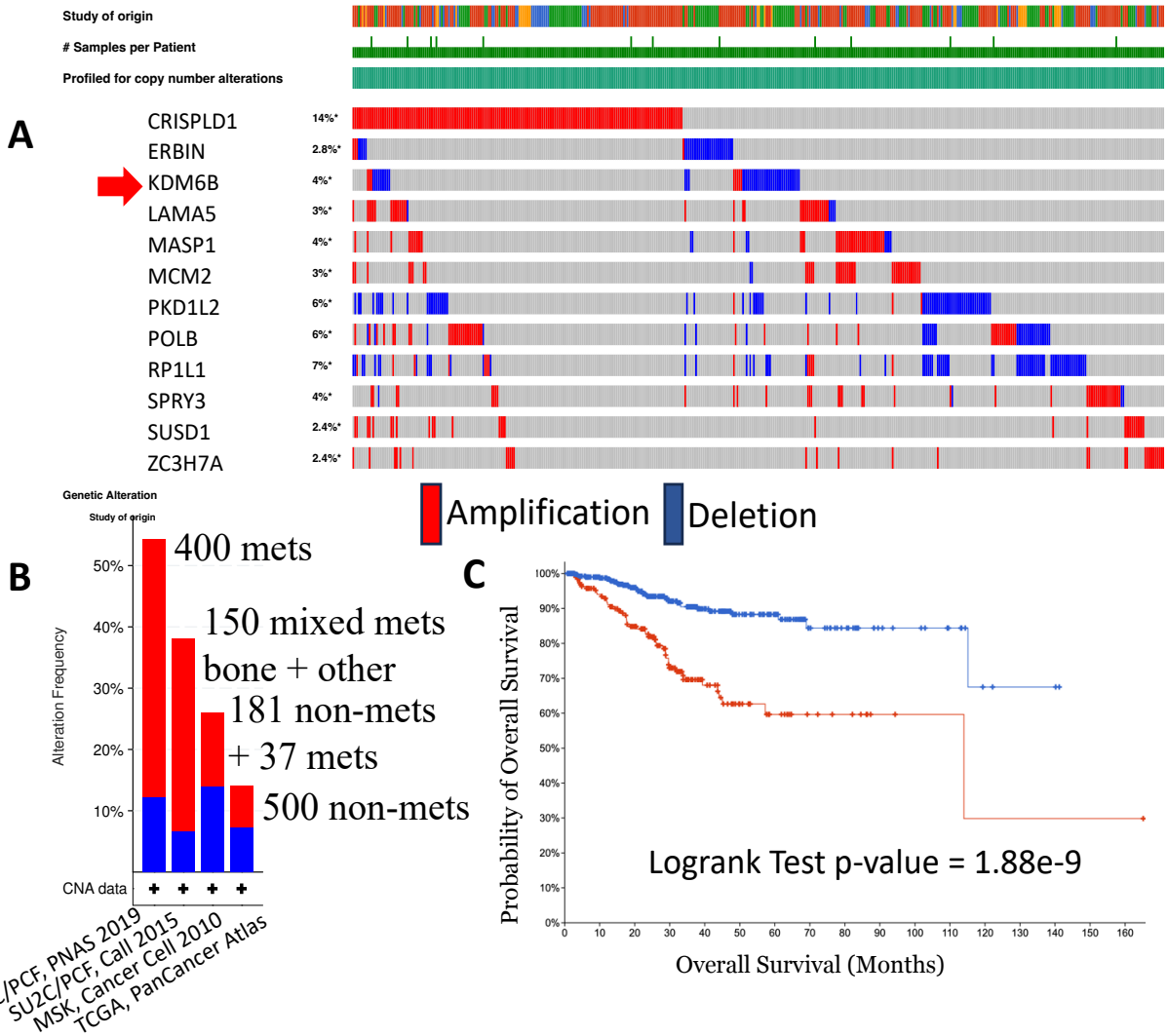
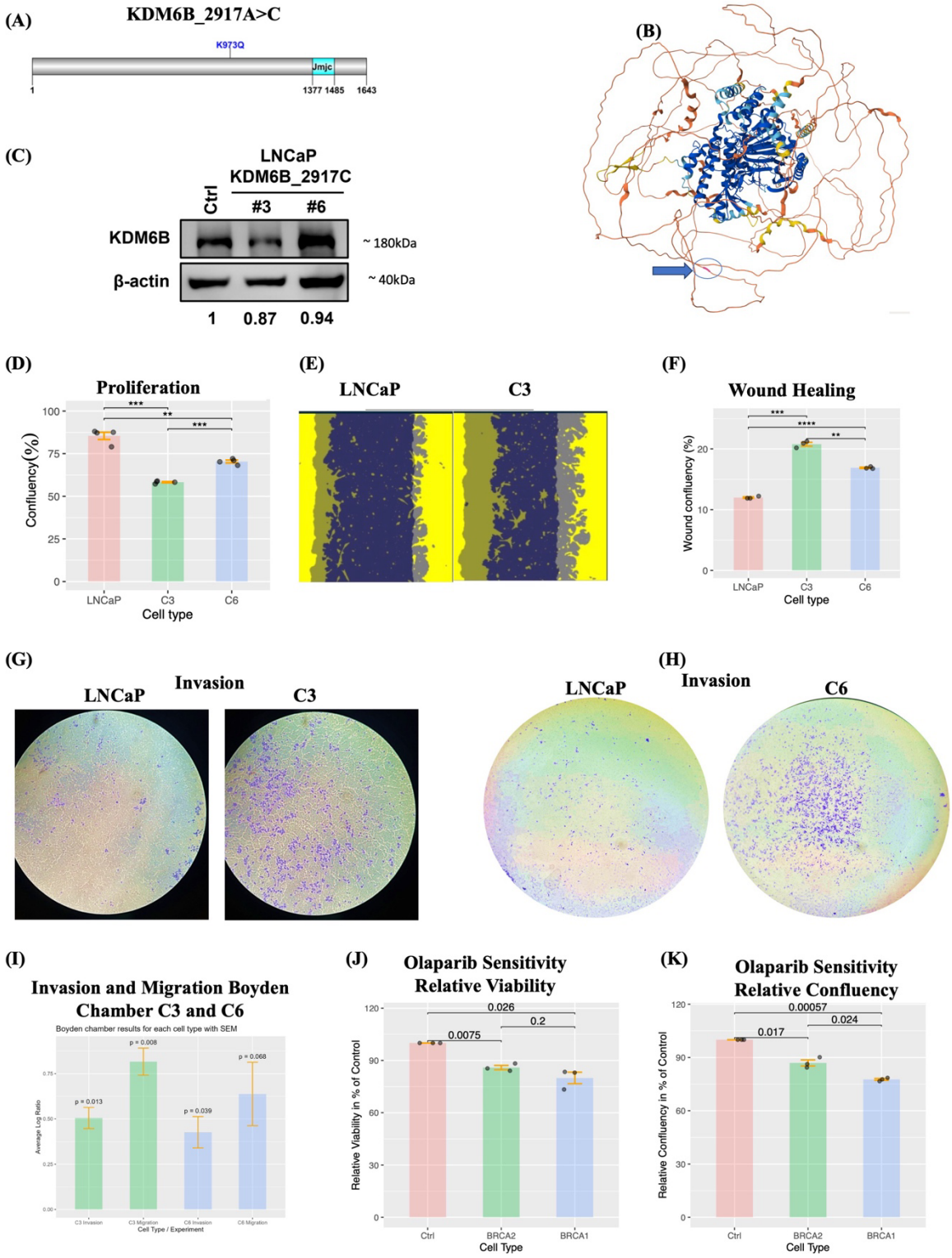


Figure 4. Functional assays and sequencing analysis of 2 gnsRVs discovered in the metastatic arm of EPC: KDM6B gnsRV rs61764072 (K973Q) in LNCaP and BRCA2 gnsRV rs1060502377(I1962T) in HEK293T.



Supplementary Texts

Table 2. The allele frequency enrichment of germline nonsynonymous rare variants (gnsRVs) discovered from the VPC cohort in the independent PPCG cohort compared to healthy Caucasians (CAs) in gnomAD. We identified gnsRVs exclusive to metastatic patients in the EPC cohort and determined their population frequency within the independent PPCG cohort of 976 PCa germlines. We then compared these frequencies to those of healthy Caucasians (CA) in gnomAD. The average fold changes were 3.82, 4.42, and 5.86 for DDR gnsRVs, DDR gnsRVs predicted to be deleterious, and gnsRVs occurring at least three times, respectively. The average CA allele frequency closely matches the overall population allele frequencies reported in gnomAD. When considering only variants with ≥ 1.2-fold positive enrichment, we observed values of 8.25, 10.61, and 17.9. (t) Notes that the values of 5.86 and 17.9 in 3 recurrent gnsRVs are driven by one very rare gnsRV rs76761697 in the NIPSNAP3A gene that occurs in three PPCG germlines and in three metastatic patients from the EPC. If this gnsRV is removed, the values drop to 2 and 5-fold enrichment, respectively.

Table 3. Association of 56gnsRVs discovered from the EPC cohort in the independent Australian cohort of 53 Caucasian PCa germilnes. This independent CA cohort consisted of 21 metastatic cases (17 with bone and/or distant visceral metastases, and 4 with lymph node metastases), 12 with biochemical recurrence (BCR), and 14 without metastases (with clinical follow-up of more than 10 years). We determined the number of patients carrying at least one of these gnsRVs in each group and compared these frequencies to the expected carrier frequencies, estimated using binomial distributions. Statistical significance was more pronounced in the metastatic phenotype group compared to the non-metastatic group, further supporting our hypothesis that the recurrent variants identified in the EPC cohort are associated with the progression of PCa to metastatic castration-resistant prostate cancer (mPCa).

Table 4. Representative genes containing gnsRVs that associated (nominal p-values) with phenotypes in the PPCG cohort. The FOCAD and CHEK2 gnsRV are linked to multiple cancers and may be associated with early-onset cancers. SIVA1 is an E3 ubiquitin ligase that regulates cell cycle progression, proliferation, and apoptosis in an ARF/p53-dependent manner. ACSF3 (PSA at RP, high-risk, and metastasis) is involved in fatty acids metabolism, and the gnsRV is likely pathogenic. SPATA9 associates with multiple clinical-pathologic features in the PPCG cohort, including metastasis. DNAJC10 is involved in misfolded protein degradation and is associated with poor survival in breast cancer and germline variants associated with aggressive PCa. *FOCAD, SIVA1, and DNAJC10 have higher minor allele frequency in African American populations than those of CAs.*

Figure 1. DDR gene gnsRVs and germline rare synonymous variants (gsRVs) in both groups of the EPC. Patterns of 102 DDR gnsRVs exclusive to either the metastatic **(A)** or non-metastatic **(B)** arm of the extreme phenotype cohort: 81 exclusive to mets **(A)** and 21 exclusive to non-mets **(B).** Green cells: HRD scores ≥ 21. ✓: gnsRV. ✗: gnsRV predicted to be deleterious. ✓ or ✗ indicate compound gnsRVs. Yellow cells: allele frequency of the gnsRV in the matched tumor is at least 1.5-fold the allele frequency in benign suggesting positive selection. The column “M” and ”N” are the number of metastatic and nonmetastatic tumors with copy number aberrations detected at the gene locus. BRCA2 gnsRVs are 100% concordant with high HRD scores. The average HRD score is 21.58 vs. 14.62 (t-test p-value=0.04) in metastatic vs. nonmetastatic patients, respectively. There are fewer patients with HRD scores ≥21 (7 versus 13) in the non-metastatic cohort. The WRN gnsRV in patients 66 and 36 is known to impair helicase activity (PMID: 15489508). It is enriched 153-fold (2 out of 26 versus 0.0005 in gnomAD) over the general population. The POLB and CHEK2 variants are known to be deleterious(PMID: 23144635, 35127508). ***DDR gnsRV are enriched in metastatic vs non-metastatic tumors (p-value=4.57e-06)*** *using a Wilcoxon rank sum test) supporting positive selection for these gnsRVs in the metastatic arm.* No significant difference was detected for 11 synonymous RVs exclusive to mets in **(C)** and 13 exclusive to non-mets in **(D)**.

Figure 2. 56 gnsRVs specific to and enriched in 26 localized metastatic tumors from 53 genes. Each gnsRV is recurrent in at least 3 (out of 26) tumors and is predicted or known to be deleterious. ✓: rare variant. ✗: predicted deleterious rare variant by fathmm-MKL. ✓ or ✗ indicate rare compound variants. Yellow cells: allele frequency of the gnsRV in the matched tumor is at least 1.5-fold the allele frequency in benign suggesting positive selection. The column "M" and "N" are the number of metastatic and nonmetastatic tumors with copy number aberrations detected at the gene locus. Genes in yellow have their gnsRVs reported in COSMIC. The enrichment of the gnsRVs of in our VPC cohort ranges from > 10-fold to >200-fold over gnomAD population frequencies. The number of gnsRVs per patient is shown at the bottom. The average is ~ eight. ***In the non-metastatic cohort, only LAMA5 and PKD1L2 had recurrent gnsRVs meeting our filtration requirement. These are shaded in gray.***

Figure 3. Genomic Analysis of Public PCa Cohorts for Genes with Recurrent gnsRVs in the EPC. Gene-level analysis was performed on 12 genes, selected from Figure 2, across 4 public prostate cancer (PCa) cohorts to determine if these genes are under selective pressure in PCa and collectively impact overall survival after disease onset. A. A subset of genes with recurrent gnsRVs from Figure 2 was analyzed using cBioPortal for copy number variant analysis. Among 4 selected PCa cohorts, most genes are either predominately gained (e.g., CRISPLD1) or lost (e.g., PKD1L2) and thus appear to be under selection. B. Copy number aberrations of 12 genes increase as tumors progress from adenocarcinoma to metastasis. This suggests that as tumors progress from relatively low risk of metastasis to a high risk of metastasis, these genes undergo selection. C. Survival analysis of 12 genes among 4 PCa cohorts. Taken together, copy number aberrations at these genes predict a poor outcome with a Log-rank Test P-value of 1.88e-9. This suggests that these genes may play a role in the aggressive phenotype and impact patient outcomes.

Figure 4. Functional assays and sequencing analysis of 2 gnsRVs discovered in the metastatic arm of EPC: KDM6B gnsRV rs61764072 (K973Q) in LNCaP and BRCA2 gnsRV rs1060502377(I1962T) in HEK293T. We edited two gnsRVs discovered in the EPC metastatic arm for functionalization. (A) KDM6B protein structure and location of the target gnsRV. KDM6B is a histone demethylase that demethylates lysine 27 on histone H3 (H3K27me3) thereby activating gene expression. The KDM6B gnsRV rs61764072 (K973Q) occurs in 11.5% of germlines of the extreme phenotype cohort metastatic arm (Fig. 2) and has a minor allele frequency approximately two-fold higher in the PPCG validation cohort compared to gnomAD. We hypothesized that KDM6B (K973Q) might impact global gene expression, making an effect straightforward to detect. rs61764072 (K973Q) lies outside the catalytic jmjd domain and may not directly impact enzyme activity, but it could influence interactions with other proteins, thereby affecting chromatin binding sites and gene expression patterns.(B) Despite its benign status in ClinVar, modeling using AlphaFold suggests an extrinsic positioning of rs61764072 (K973Q) on the protein surface, and several prediction algorithms predict it to be deleterious. To test this, we edited the K973Q variant into the KDM6B locus in wildtype LNCaP cells using CRISPR/Cas9 prime editing. (C) Two independent LNCaP clones, homozygote C3 and heterozygote C6 knock-ins, were selected for further study (successful editing in both was confirmed by whole exome sequencing.) The expression levels of KDM6B of LNCaP-C3 and LNCaP-C6 are comparable to that of the parental clone. (D) The growth rate of homozygous C3 is lower than that of heterozygous C6, with C6 growing slower than the WT. (E) Scratch assays reveal a mild migratory phenotype in edited LNCaP cells compared to WT. (G) Quantification shows C3 is more migratory than C6. (G,H,I) Boyden chamber assays reveal approximately a 4-fold increase in migration and invasion compared to WT LNCaP cells and again clone C3 (homozygous) displays stronger phenotypes than clone C6 (heterozygous) suggesting a dosage effect. (J,K) The BRCA2 gnsRV rs1060502377 (I1962T) (gnomAD MAF: 10e-7) confers sensitivity to PARP inhibitors, despite being classified as of unknown significance in ClinVar.