Characterizing the combined role of rare and common variants for prostate cancer risk

Alexander Chuckuma Dunkwu

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science in Genetic Epidemiology

University of Washington
2026

Committee:
Burcu Darst
Sara Lindström

Program Authorized to Offer Degree:
Institute for Public Health Genetics

University of Washington

Abstract

Characterizing the combined role of rare and common variants for prostate cancer risk

Alexander Chuckuma Dunkwu

Chair of the Supervisory Committee:

Burcu Darst

Department of Epidemiology

Prostate cancer (PCa) remains a public health problem as it is one of the most frequently diagnosed and most heritable non-skin cancers in men globally. Unfortunately, the disease etiology of PCa is not yet fully understood, but age, family history, race and ethnicity, and genetic factors are widely reported risk factors. Previous studies have shown that risk for PCa depends on both rare and common variant risk separately, but the combined effect of them has not been explored in depth. Here, we used logistic regression models to validate a common variant polygenic risk score (PRS) of 451 PCa risk variants and REGENIE to conduct targeted gene-based analyses of rare, pathogenic coding variants, defined as having a 'HIGH' VEP Impact or a 'Pathogenic/Likely Pathogenic' ClinVar annotation. Subsequently, we conducted a combined PRS and rare variant carrier status interaction analysis in 82,802 participants from the All of Us (AoU) Research Program through logistic regression models. Our results validated the PRS' ability to distinguish between PCa cases and controls and confirmed associations for 6 previously reported PCa risk genes (*BIK, FAM111A, HOXB13, BRCA2, CHEK2* and *ATM)*. We found evidence of an additive interaction between the PRS and rare variant carrier status, with carriers defined as those with 1 pathogenic coding variant in any of the 6 genes mentioned

above. We observe significantly higher risk estimates in carriers in the 80-100% PRS (OR = 4.61; 95% CI: 3.86,5.50) than non-carriers in the 80-100% PRS (OR = 3.11; 95% CI: 2.89, 3.36). Taken together, our results demonstrate the importance of jointly considering the effect of common and rare variants in PCa risk estimation to improve risk stratification.

Introduction

Prostate cancer (PCa) remains a public health problem as it is one of the most frequently diagnosed and most heritable non-skin cancers in men globally.[1,2] This is further exacerbated by disparities in PCa incidence and mortality that is seen across different racial and ethnic groups.[3] For example, Black men have the highest incidence rate of any racial or ethnic group in the United States, are more likely to be diagnosed with more aggressive forms of PCa, and are more likely to die from the disease.[3–9] Although race and ethnicity are social constructs, they have been considered proxies for genetic ancestry when studying genetic or environmental contributions to disease. Genetic ancestry has also been recommended by the National Academies of Sciences, Engineering, and Medicine (NASEM) as an appropriate population descriptor to be used in studies involving gene discovery and prediction for complex, polygenic traits.[10] Previous studies employing ancestry, as opposed to race and ethnicity, have identified risk factors that were more common in African genetic ancestry groups.[11,12] As such, considering genetic ancestry is an important, early step for improving PCa risk stratification, so we will be using genetic ancestry in these analyses.

Unfortunately, the disease etiology of PCa is not yet fully understood, but age, family history, race and ethnicity, and genetic factors are widely reported risk factors.[13,14] Importantly, genetic factors are estimated to be responsible for up to 57% of PCa risk variability.[2] Although the role of genetic factors has not yet been fully characterized, there have been numerous studies that have identified genetic factors that contribute to PCa risk. Previous genome-wide association studies (GWAS) have identified 451 independent risk variants, which have been aggregated into a multi-ancestry polygenic risk score (PRS) that is highly predictive of PCa risk across ancestry groups.[11,15,16] This PRS of 451 variants has also been shown to have better

predictive ability than previous PRS iterations with less risk variants and other PRS generated through genome-wide approaches.[15] However, this PRS primarily only aggregates common variant (minor allele frequency (MAF) >1%) risk and does not incorporate risk from rare variants. This remains problematic as rare variants (MAF <1%) are widely believed to have distinct effects on PCa risk.[16–18]

Rare variant analyses have identified several genes with pathogenic, likely pathogenic, and deleterious variants (P/LP/D) that are associated with PCa risk.[16,19] Most notably, *HOXB13* is a widely reported risk gene for PCa, with rare variants such as the Northern European ancestry-specific G84E (rs138213197) and Western African ancestry-specific X285K (rs77179853) variants being associated with an increased risk of PCa.[20,21] *BRCA2* is another widely reported risk gene that is associated with increased risk for PCa and other cancers.[22,23] Other reported genes associated with PCa risk include *CHEK2, ATM, PALB2, NBN, MSH2, BIK, FAM111A, SAMHD1, EPCAM, PMS2, RAD51D,* and *TP53*, along with DNA mismatch repair genes that are implicated in Lynch syndrome (*MLH1, MSH6)*.[4,24–32]

When considered separately, common and rare variant analyses have improved risk prediction for PCa and our understanding of its disease etiology. PRS-based screenings have started to be tested to guide traditional, annual cancer screenings as seen in the WISDOM study for breast cancer.[33] In a recent clinical trial, a PCa PRS of about 130 variants was shown to identify patients with clinically significant disease who were missed by prostate specific antigen (PSA) or MRI based screening.[34] Several of the reported PCa associated genes, including *BRCA2* and *HOXB13*, are also included in multi-gene panel tests for hereditary PCa.[35] Given the utility of understanding common and rare variant risk for PCa on their own, combining our understanding in regards to both may also improve genetic risk prediction.

Given that common and rare variants are often considered separately, their combined impact is still not fully explored.[36–39] Although previous studies have shown that the risk profile for individuals with a similar common variant PRS can vary by rare pathogenic variant carrier status, these studies have not investigated the most recent multi-ancestry PRS of 451 variants or had a large and diverse sample.[40–42] By considering the joint effect of common and rare risk variants with a more up-to-date PRS and a more ancestrally diverse sample, we may be able to further improve risk stratification and genetic risk prediction for PCa.

Here, we conducted PRS and targeted gene-based rare variant analyses, followed by a combined interaction analysis in the multi-ancestry All of Us (AoU) Research Program to: 1) validate the separate effect of rare variants in known PCa risk genes and a multi-ancestry PRS with risk of PCa and 2) assess how PCa risk depends on both common and rare variant risk.

Methods

Participants / Cohort

The All of Us (AoU) Research Program is a longitudinal, cohort study of over 633,000 people aged 18 and older living in the United States that has aimed to enroll a diverse set of people to accelerate biomedical research and improve human health.[43] The program began accepting participants in May 2018 and has since continued to enroll participants from around the United States. After the provision of informed consent, participants are first given baseline health surveys. Participants are then asked to provide authorization to share electronic health records (EHRs) and undergo physical measurement and biospecimen collection. EHRs included the capture of structured data such as billing codes and laboratory results from various health care provider organizations. Biospecimens from blood and saliva were collected and used to generate

short-read whole genome sequencing data (srWGS).[44] Currently, over 350,000 participants have agreed to shared EHRs and over 410,000 participants have had srWGS data generated from collected biospecimens.

Our inclusion criteria included having male genetic sex assigned at birth, srWGS data collected as part of AoU, and authorization of the release of EHR data to AoU.

Our primary outcome of interest was diagnosis of PCa, as defined via the eMERGE algorithm for PCa.[45] Cases had at least one PCa diagnosis code (International Classification of Disease (ICD) 9 codes 185 or V10.46, ICD10 codes C61, Z85.45 or R97.21, or corresponding SNOMED codes in AoU for malignant neoplasm of the prostate) with at least two separate occurrences on different dates. Controls had none of these ICD9, ICD10, or SNOMED codes, as well as no prostatectomy or self-reported history of PCa. We further restricted our controls to those who were over 50 years of age at their last clinical visit. This resulted in a final sample size of 7,577 cases and 75,225 controls across 6 different genetic ancestry groups: African (AFR), Americas (AMR), East Asian (EAS), European (EUR), Middle Eastern (MID), and South Asian (SAS). In brief, the 6 genetic ancestry groups were generated by AoU by measuring the genetic similarity of each participant to global reference populations from the Human Genome Diversity Project and 1000 Genomes Project.[46–48]

Genetic Data

The srWGS SNP and Indel dataset from the AoU research program version 8 were used for these analyses, which were generated through the Illumina NovaSeq 6000 instrument and DRAGEN platform with the GRCh38 reference genome.[44] As previously described, the AoU Research Program applied standard, predefined quality control (QC) measures.[44] In brief, single

sample QC were applied, variants were called, and QC was applied to the joint call set. The entire joint call set was rendered as a Hail Variant Dataset (VDS), before subsections were extracted and made available to researchers in different formats. Any variants that did not meet QC criteria were excluded. For PRS analysis, we extracted the 451 variants from the srWGS data in PLINK format from the AoU Controlled Tier Dataset v8. For rare variant coding analyses, we used the same srWGS dataset but limited to only coding regions. This resulted in approximately 41 million variants for downstream analysis.

We then adopted the PRS of 451 known PCa risk variants reported by the largest multi-ancestry PCa GWAS meta-analysis performed to date.[15] In brief, the 451 PRS variants were extracted from the AoU-provided Hail MatrixTable and then converted to PLINK format. PLINK2 was then used to generate the PRS as a sum of the amount of risk alleles carried, with each variant weighted by the previously reported multi-ancestry conditional weights for each variant.[15,49]

As reported by others, we used the Ensembl Variant Effect Predictor (VEP) release 112 with the Combined Annotation Dependent Depletion (CADD) v1.7, ClinVar, and the Rare Exome Variant Ensemble Learner (REVEL) plugins to generate functional annotations for rare variants in the coding regions of 16 previously reported PCa associated genes.[50–53] These functional annotations were used to generate 3 rare variant sets for each gene: predicted loss of function (pLOF), predicted deleterious missense, and synonymous variants. We chose to include a synonymous set to serve as a class of genetic variation with 'neutral' effects toward PCa risk. Across all rare variant sets, variants had to have a minor allele frequency (MAF) <1% in our controls to be included. We defined pLOF variant sets as having a 'High' VEP Impact annotation or a ClinVar 'pathogenic' or 'likely pathogenic' annotation. Variants with the

following consequences are given a 'High' VEP Impact: transcript ablation, splice acceptor/donor variant, stop gained, frameshift variant, stop lost, start lost, transcript amplification, feature elongation, and feature truncation. Predicted deleterious missense variants were defined as having a 'Missense' VEP consequence and a Phred CADD score greater than 20, a REVEL Score greater than 0.5, or a ClinVar annotation of 'pathogenic' or 'likely pathogenic'.[51,52,54] Synonymous variants were required to have a 'synonymous variant' VEP consequence or a 'benign' or 'likely benign' ClinVar annotation. Variants were only included if they met the standard AoU QC criteria.

Statistical Analysis

We standardized the PRS score for all individuals within their respective genetic ancestry groups and tested its association with PCa by modeling the PRS as a continuous variable in logistic regression analyses. Analyses were conducted overall and stratified by ancestry, and models were adjusted for age at sample draw and the first 10 principal components (PCs) of genetic ancestry. Estimates were reported as odds ratios per standard deviation unit increase in the PRS with corresponding 95% confidence intervals. We also modeled the PRS as a categorical variable by grouping individuals into separate risk groups: 0-10%, 10-20%, 20-30%, 30-40%, 40-60% (reference group), 60-70%, 70-80%, 80-90%, 90-100%, defined by PRS distributions in cases and controls combined. For analyses where sample size allowed, the highest risk group was further split into a 90-99% and a 99-100% risk group. We calculated category-specific odds ratios using the 40-60% category as our reference group. To assess discriminative ability of the PRS, we generated receiver operating curves (ROC) and calculated the area under the curve (AUC) for a logistic regression model with the PRS, age at sample draw, and first 10 PCs, as well as a logistic regression model with only age at sample draw and the first 10 PCs. DeLong's

test for two correlated ROC curves was then used to assess statistically significant differences between the two models.

REGENIE v4.0 was used to conduct gene-based burden tests of rare variants in the previously mentioned variant sets.[55] In REGENIE step 1, we used a set of 123,755 autosomal variants from our exon-filtered srWGS dataset to fit a whole genome regression model with block sizes of 1,000 consecutive markers.[55] Variants in this set had to meet the following criteria: genotype missingness rates less than 10%, Hardy-Weinberg equilibrium p-value above 10^{-15}, and a minor allele frequency above 1%. For REGENIE step 2, we created 4 gene-based burden masks to collapse all variants below MAF 1% in our previously defined variant sets into a single variable for each of our 16 previously reported PCa associated genes: 1) pLOF, 2) predicted deleterious missense, 3) synonymous, and 4) a combined pLOF and deleterious missense mask. We then tested the association of these burden masks with PCa risk through REGENIE, adjusting for age and the first 10 PCs. As previously proposed by Wang (2014), Firth correction was used for all p-values less than 0.999 to avoid inflation in our rare variant gene-based tests.[56] By default, variants with MAC <5 were not tested.

For the subsequent PRS-rare variant analyses, we selected the genes that were significantly associated ($p<0.05$) with PCa when considering pLOF and deleterious missense variants. A binary carrier status variable was created whereby carriers needed to have at least one copy of the rare allele for at least 1 pLOF variant in any of those genes. We first tested for effect heterogeneity of the PRS across carrier status through carrier status stratified logistic regression models, adjusting for age and the first 10 PCs. Then, we added an interaction term between the PRS and carrier status into the model to formally test for effect heterogeneity on the multiplicative scale.

We then defined new risk categories that were based on PRS risk group and carrier status (0-20% in carriers, 0-20% in non-carriers, 20-40% in carriers, 20-40% in non-carriers, 40-60% in carriers, 40-60% in non-carriers (reference group), 60-80% in carriers, 60-80% in non-carriers, 80-100% in carriers, 80-100% in non-carriers). For this step, we used risk quintiles instead of deciles to allow for each risk group to have an adequate number of cases. We also recalculated our PRS excluding two known rare risk variants: *HOXB13* G84E and *CHEK2* 1100delC.

To assess any improvements in our model when including a carrier status variable, we then used the likelihood ratio test to compare a model with the PRS, carrier status, age, and first 10 PCs to a model with only the PRS, age, and the first 10 PCs. We also generated a ROC curve and calculated the AUC for the model with PRS, carrier status, age, and the first 10 PCs. DeLong's test for two correlated ROC curves was then used to assess statistically significant differences between the two models.

We used a p-value threshold of 0.05 to assess significance for all statistical analyses.

Results

All analysis were performed in the AoU Research Program with 7,577/75,225 (cases/controls) overall. Of our overall sample, 5,757/47,351 were predicted to be EUR, 1,212/16,532 AFR, 492/9,301 AMR, 63/1,123 EAS, 27/687 SAS, and 26/231 MID genetic ancestry groups (Table 1). The overall mean age at sample draw was 64.11 years (SD: 9.92), while the mean age at sample draw for cases and controls was 71.00 years (SD: 8.56) and 63.43 years (SD: 9.78), respectively (Table 1). Given that we standardized our PRS, the overall mean PRS was 0.00 (1.00). Cases had a mean (SD) PRS of 0.59 (1.01), while controls had a mean (SD) PRS of -0.06 (0.98) (Table 1; Figure 1). About 24% of cases were assigned to the highest

risk group of 90-100% based on their PRS score, while only 8.5% of controls were assigned to this risk group (Table 1). Conversely, about 11% of controls and only 3.2% of cases were in the lowest risk group (Table 1).

The PRS as a continuous predictor was significantly associated with PCa overall (OR = 2.08 per SD unit increase in PRS; 95% CI: 2.03,2.13; p-value < 2×10^{-16}), and in AFR, AMR, EAS, and SAS genetic ancestry groups, with nominal evidence in MID (Figure 2 and Table 2). The PRS had the largest effect estimate in the AMR genetic ancestry group (OR = 2.17; 95% CI: 1.96, 2.41; p-value < 2×10^{-16}), then the EUR genetic ancestry group (OR = 2.13; 95% CI: 2.07, 2.19; p-value < 2×10^{-16}), the EAS genetic ancestry group (OR = 2.06; 95% CI: 1.55, 2.78; p-value = 9.85×10^{-7}), the AFR genetic ancestry group (OR = 1.99; 95% CI: 1.86, 2.12; p-value < 2×10^{-16}), the SAS genetic ancestry group (OR = 1.68; 95% CI: 1.14, 2.51; p-value = 0.009), and the MID genetic ancestry group (OR = 1.57; 95% CI: 1.00, 2.56; p-value = 0.058) (Figure 2; Table 2). The AUC for a model with covariates and the PRS was 0.783 (95% CI: 0.778, 0.788), in comparison to the AUC of 0.725 (95% CI: 0.720, 0.730) for the model with covariates only (DeLong's Test p-value < 2×10^{-16}) (Figure S1).

Overall, those in the top 1% risk group (99-100%) had an OR of 7.98 (95% CI: 6.76; 9.41; p-value <2×10^{-16}) when compared to our reference group (40-60%) (Figure 3; Table 3). When making these same comparisons stratified by ancestry groups with a large enough case group, we observe an OR of 8.27 (95% CI: 6.81, 10.03; p-value < 2×10^{-16}) in EUR genetic ancestry, 7.72 (95% CI: 5.02, 11.64; p-value < 2×10^{-16}) in AFR genetic ancestry, and 8.15 (95% CI: 4.37, 14.68; p-value = 9.84×10^{-12}) in AMR genetic ancestry (Figure 4; Table 3). We could not assess the risk associated with the top 1% in EAS, MID and SAS genetic ancestries because of limited sample sizes.

We identified 2,325 pLOF variants, 7,234 predicted deleterious missense variants, 9,437 pLOF or predicted deleterious missense variants (some overlapped), and 10,049 synonymous variants across 16 previously reported PCa-associated genes (Table 4). When considering only pLOF variants, *HOXB13* (OR = 4.14; 95% CI: 3.06, 5.60; p-value < 2×10^{-16}), *BRCA2* (OR = 1.94; 95% CI: 1.40, 2.70; p-value = 6.61×10^{-5}) , *ATM* (OR = 1.53; 95% CI: 1.07, 2.19; p-value = 0.02), *BIK* (OR = 1.40; 95% CI: 1.19, 1.64; p-value = 6.09×10^{-5}), *FAM111A* (OR = 1.31; 95% CI: 1.06, 1.62; p-value = 0.01), and *EPCAM* (OR = 0.13, 95% CI: 0.02, 0.86; p-value = 0.04) were associated with PCa (Table 5A; Figure 5). With only predicted deleterious missense variants, *HOXB13* (OR = 2.31; 95% CI: 1.74, 3.05; p-value = 5.0×10^{-9}), *CHEK2* (OR = 1.32; 95% CI: 1.14, 1.53; p-value = 2.4×10^{-4}), and *ATM* (OR = 1.09; 95% CI: 1.01, 1.19; p-value = 0.04) were associated with PCa (Table 5B; Figure 5). When considering both pLOF and predicted deleterious missense variants, *HOXB13* (OR = 2.37; 95% CI: 1.84, 3.07; p-value = 4.5×10^{-11}), *BIK* (OR = 1.37; 95% CI: 1.17, 1.62; p-value = 1.2×10^{-4}), *CHEK2* (OR = 1.31; 95% CI: 1.15, 1.50; p-value = 7.9×10^{-5}), *BRCA2* (OR = 1.17; 95% CI: 1.05, 1.31; p-value = 0.004), *FAM111A* (OR = 1.22; 95% CI: 1.01, 1.47; p-value = 0.035), and *ATM* (OR = 1.11; 95% CI: 1.02, 1.20; p-value = 0.015) were associated with PCa (Table 5C; Figure 5). For synonymous variants, only *ATM* (OR = 1.11; 95% CI: 1.05, 1.17; p-value = 3.77×10^{-4}) was associated with PCa (Table 5D; Figure 5). We did not observe any significant associations in any of the variant sets for *MLH1, NBN, PALB2, PMS2, RAD51D, MSH6, MSH2, TP53* and *SAMHD1*. (Table 5D; Figure 5)

Given these findings, subsequent analyses focused on carrier status for pLOF variants in either *BIK, FAM111A, HOXB13, BRCA2, CHEK2,* or *ATM*, which in aggregate were carried by 6.74% of cases and 4.6% of controls (Table 6). When stratifying by carrier status

across the six genes, we observed that the PRS association with PCa was slightly larger in non-carriers (OR = 2.08; 95% CI: 2.02, 2.14; p-value < 2×10^{-16}), as opposed to carriers (OR = 1.98; 95% CI: 1.80, 2.12; p-value < 2×10^{-16}), but a PRS and carrier status interaction term was not significant (p-value = 0.11). In a model adjusting for PRS and including a PRS-carrier status interaction term, the main effect of carrier status was significant (OR = 1.38; 95% CI: 1.22, 1.56; p-value = 2.03×10^{-7}).

When generating new risk categories that were determined by PRS and carrier status combined, we observed that carriers in the top 20% PRS risk group (80-100%) had an OR of 4.61 (95% CI: 3.86, 5.50; p-value <2×10^{-16}) when compared to the reference group (40-60% PRS in non-carriers) (Figure 6; Table 7). Among non-carriers, those in the top 20% PRS risk group had an OR of 3.11 (95% CI: 2.89, 3.36; p-value < 2×10^{-16}) when compared to the same reference group. We did not stratify these analyses by ancestry due to the limited number of pLOF variant carriers. From the likelihood ratio test, we see that the inclusion of carrier status in the model significantly improved its performance (χ^2(2 d.f.) = 27.32; p-value = 1.17×10^{-6}). The AUC for the model with carrier status (AUC = 0.783, 95% CI: 0.778, 0.788) did not have a clinically meaningful difference than the AUC for the model without carrier status (AUC = 0.782, 95% CI: 0.777, 0.787).

Discussion

In this investigation, we first validated the performance of the most recent multi-ancestry PRS of 451 variants for PCa[15] across most of the evaluated genetic ancestry groups. Although AoU is a diverse cohort, it is still heavily represented by European genetic ancestry participants, and for some genetic ancestry groups with more limited representation, we could not robustly assess the PRS. As expected, we found that cases had on average higher PRS when compared to

controls. As reported in previous studies, the PRS was significantly associated with PCa risk, in both the overall sample and in separate, predicted genetic ancestry groups, whether we modeled the PRS continuously or categorically. This was true even for genetic ancestry groups where the number of cases was relatively small, such as the MID and SAS groups. When creating categorical risk groups based on the PRS, we found that risk was increased for higher PRS risk groups, particularly the 90-100% and 99-100% PRS categories, as expected. This was found to be true for all ancestries where we had large enough sample sizes to conduct the analyses. Lastly, the AUC for the model with the PRS also supports the strong performance of the PRS as a tool for genetic risk prediction. Specifically, we see a significant +0.058 improvement in AUC when adding the PRS to the model with age and the first 10 PCs.

Our results for the rare variant analyses validated the associations of 6 previously reported PCa associated genes: *BIK, FAM111A, HOXB13, BRCA2, CHEK2,* and *ATM*. As expected, *HOXB13* had the strongest effect across all targeted genes when considering pLOF variants, deleterious missense variants, or both. *HOXB13*, or Homeobox 13, encodes a homeobox transcription factor and is a widely known high-risk gene for PCa with some literature also pointing towards it increasing risk for other cancers as well.[20,21,57] *ATM* pLOF variants, deleterious missense variants, or both were all found to be significantly associated with PCa. *ATM* is a tumor-suppressor kinase with known roles in DNA-damage repair, apoptosis, and cell-cycle arrest.[58–62] Given its role as a tumor suppressor, it is not surprising that variants causing a loss of function may increase risk for PCa. Unexpectedly though, even synonymous variants in *ATM* were found to increase risk for PCa. Interestingly, previous cancer-type agnostic studies have shown that some synonymous variants may affect cancer risk through mechanisms involving aberrant splicing, which may be the case here.[63,64] However, the true underlying

biological pathway behind this phenomenon is unclear, so this association should be explored further in future studies.

While *BRCA2* was significantly associated with PCa when considering pLOF variants or pLOF and deleterious missense variants, we did not observe the same effect for deleterious missense variants on their own. This is not unexpected as previous *BRCA2* rare variant studies only considered variants predicted to cause a loss of function. [65,66] When considering pLOF variants, we observed effects similar to what has been reported elsewhere.[65,66]

Similarly, *BIK* and *FAM111A* were also found to only be significant when considering pLOF variants or pLOF and deleterious missense variants. *BIK* has roles in mediating apoptosis, and the gene has also been linked to aggressive breast cancer.[67,68] However, the functional role of *BIK* is unclear, so it may be difficult to correctly classify rare missense variants in *BIK* as deleterious when using variant annotation tools that incorporate functional data. This is further supported by *BIK* having the lowest number of deleterious missense variants across all tested genes (Table 4). *FAM111A* is a protease with implications in DNA replication and genome integrity maintenance, as well as the genetic basis for Kenny-Caffey Syndrome type 2 (KCS2).[69,70] KCS2 patients are known to show symptoms of hypoparathyroidism and hypocalcemia.[69–71] Interestingly, there have been previous studies that have implicated high serum calcium and high serum parathyroid hormone (PTH) with increased risk of PCa.[72,73] Given this, the association of pLOF variants in *FAM111A* may be mediated by a biological pathway involving serum calcium and serum PTH.

Unexpectedly, we did not see a significant association of pLOF variants in *CHEK2* with PCa, but we did see an association when considering deleterious missense or deleterious missense and pLOF variants. *CHEK2* encodes for a kinase with a critical role in DNA repair, and

several variants in *CHEK2* have been previously reported as associated with PCa.[74,75] Given that association of *CHEK2* pathogenic variants with PCa is widely reported elsewhere, we believe our study is underpowered to quantify the effect of *CHEK2* on PCa risk.

We observed a protective effect for pLOF variants in *EPCAM*. Previous studies have reported that *EPCAM* is highly expressed in many cancer types and focused on a potential prognostic biomarker role of the protein.[76,77] Given its versatility across cancers and roles in cell signaling and proliferation, *EPCAM* may participate in many steps of cancer progression.[78] This supports *EPCAM*'s role as a potentially protective gene as variants that cause a loss of function may reduce *EPCAM*-mediated cancer progression.

In our total sample, 3,974 (4.80%) individuals were carriers for at least 1 pLOF variant in either of the following genes: *BIK, FAM111A, CHEK2, HOXB13, ATM,* and *BRCA2.* The frequency of carriers across all six genes differed across case or control status, but *HOXB13* carrier frequency was especially higher among cases. We did not observe a significant multiplicative interaction between the PRS and carrier status, which could reflect that our study was underpowered to detect an effect. Since there are no well-suited methods for estimating excess risk due to additive interaction with a continuous predictor, we defined new risk categories based on PRS and carrier status. With this analysis, we found evidence of additive interaction, such that carriers in the 80-100% PRS (OR = 4.61; 95% CI: 3.86,5.50) had significantly higher ORs than non-carriers in the 80-100% PRS (OR = 3.11; 95% CI: 2.89, 3.36). Similarly, carriers in the 20-40% PRS category had higher risk 1.41-fold higher odds (95% CI: 1.09, 1.83; p-value = 0.001) than non-carriers in the 40-60% PRS category. Lastly, we found discordant evidence that including carrier status and a joint PRS-carrier status interaction term improved the performance of our model. Taken all together, this leads us to conclude that future

analysis needs to be conducted to more definitively assess a potential role of rare variant carrier status in known PCa-risk genes, in improving genetic risk prediction.

This study adds to existing literature on the joint effect of common and rare variant risk for PCa. We used and validated a PRS from the largest multi-ancestry meta-analysis GWAS for PCa conducted to date.[15] This was done in a cohort that includes individuals from ancestry groups that are typically underrepresented in genetic research. We also used a combination of variant annotation tools and definitions to ensure the inclusion of a broad range of variants.

However, there were several notable limitations in this study that must be considered. Although AoU is a diverse and large cohort, the majority are of European genetic ancestry. Additionally, there was also selection bias, as we required that individuals had both srWGS and EHR data available. Although the PRS was strongly associated with PCa overall and in AFR, EUR, and AMR genetic ancestry groups, we could not adequately assess its performance in SAS, MID, and EAS genetic ancestry groups due to limited case numbers. This similarly limited our ability to see how the effect of carrier status may differ across ancestry groups. Additionally, there are population-specific rare variants, like X285K in *HOXB13* only found in populations with Western African genetic ancestry, that we were unable to adequately capture and evaluate in our study populations. Our rare variant analyses were also limited to coding regions, so any rare variants in noncoding regions were not considered.

In summary, we have validated the performance of the most recent multi-ancestry PCa PRS of 451 variants in predicting PCa risk. We have also validated the association of 6 previously reported PCa risk genes and identified 1 potentially protective gene. Our results suggest that risk for PCa depends on both common variant risk (here in the form of a PRS) and rare variant risk. However, we had limited power to comprehensively characterize the predictive

ability of rare variants. As such, the inclusion of rare variants in future risk stratification and genetic risk prediction models should continue to be explored more in-depth. Future absolute risk calculations should be done to quantify the clinical and population impact of considering rare variant risk in conjunction with a common variant PRS. Future studies should also seek to meta-analyze across studies to increase sample sizes and power, particularly across studies that have been underrepresented in genetic investigations. It may also be helpful to extend the work presented here in future cohorts of cases with younger ages of onset or more aggressive forms of PCa.

Figures/Tables

Table 1: Cohort Demographics Table

	Disease Status		
	Overall	**Control**	**Case**
Variable	N = 82,802[1]	N = 75,225[1]	N = 7,577[1]
Age at Sample Draw (Years)	64.11 (9.92)	63.42 (9.78)	71.00 (8.56)
Mean PRS Score (Std)	0.00 (1.00)	-0.06 (0.98)	0.59 (1.01)
Risk Group			
0-10%	8,281 (10%)	8,038 (11%)	243 (3.2%)
10-20%	8,281 (10%)	7,942 (11%)	339 (4.5%)
20-30%	8,282 (10%)	7,860 (10%)	422 (5.6%)
30-40%	8,277 (10.0%)	7,749 (10%)	528 (7.0%)
40-60% (reference)	16,560 (20%)	15,294 (20%)	1,266 (17%)
60-70%	8,280 (10.0%)	7,522 (10.0%)	758 (10%)
70-80%	8,280 (10.0%)	7,384 (9.8%)	896 (12%)

80-90%	8,281 (10%)	7,012 (9.3%)	1,269 (17%)
90-100%	8,280 (10.0%)	6,424 (8.5%)	1,856 (24%)
Predicted Genetic Ancestry			
African	17,744 (21%)	16,532 (22%)	1,212 (16%)
Americas	9,793 (12%)	9,301 (12%)	492 (6.5%)
East Asian	1,186 (1.4%)	1,123 (1.5%)	63 (0.8%)
European	53,108 (64%)	47,351 (63%)	5,757 (76%)
Middle Eastern	257 (0.3%)	231 (0.3%)	26 (0.3%)
South Asian	714 (0.9%)	687 (0.9%)	27 (0.4%)

[1] Mean (SD); n (%)

Figure 1: Distribution of standardized PRS, stratified by prostate cancer case/control status

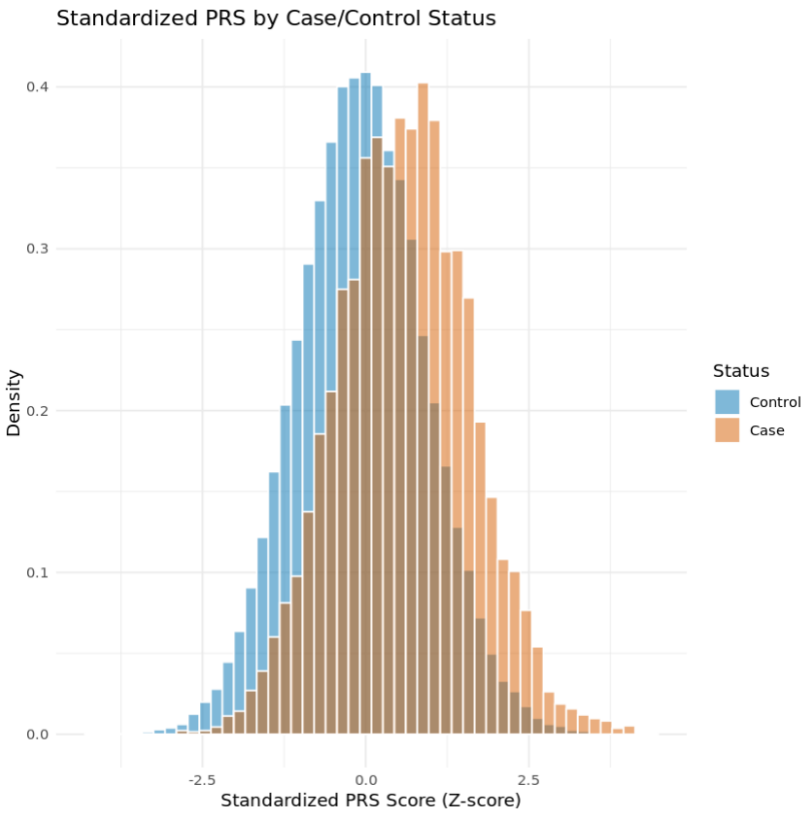


Figure 2: Association of PRS with prostate cancer, overall and stratified by ancestry

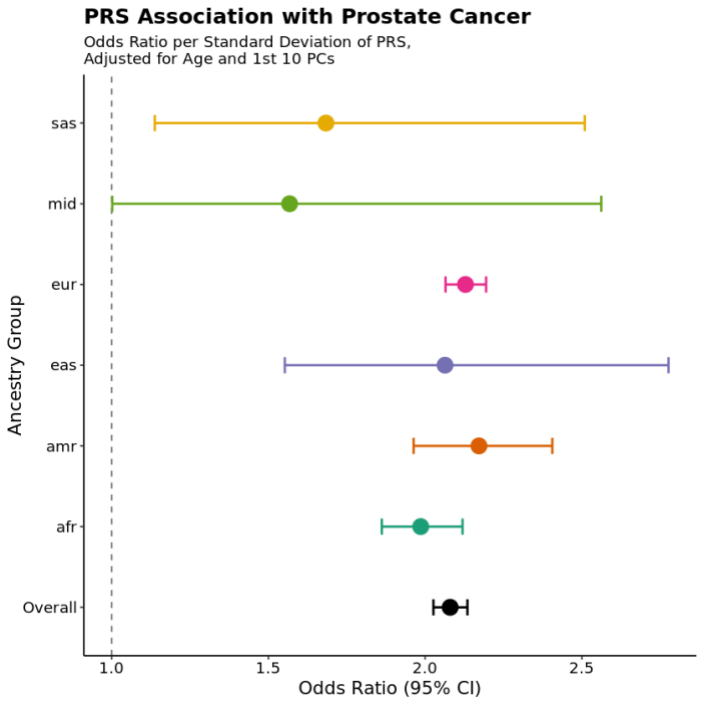


Table 2: Association Results for PRS as continuous analysis

Odds Ratios (OR) per 1 Standard Deviation unit increase in PRS

Ancestry	OR	95% CI	p-value
Overall	**2.08**	**(2.03, 2.13)**	**<2e-16**
European	2.13	(2.07, 2.19)	<2e-16
African	1.99	(1.86, 2.12)	<2e-16
Americas	2.17	(1.96, 2.41)	<2e-16
East Asian	2.06	(1.55, 2.78)	9.85e-07
South Asian	1.68	(1.14, 2.51)	9.46e-03
Middle Eastern	1.57	(1.00, 2.56)	5.81e-02

Figure 3: Association between PRS risk groups and prostate cancer in the overall sample

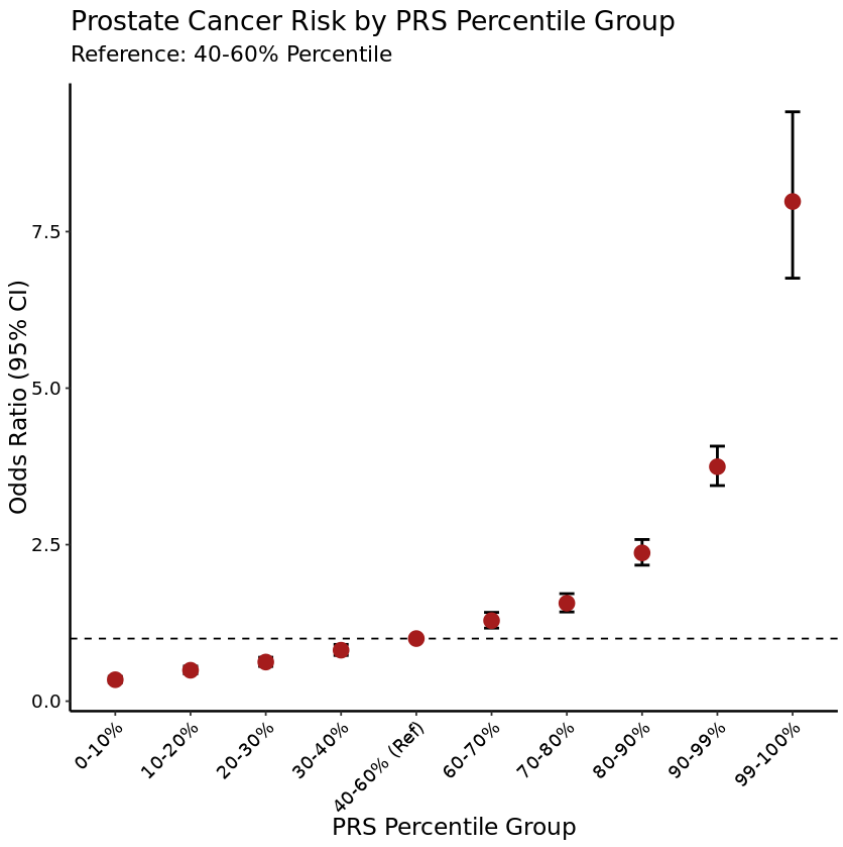


Figure 4: Plot for risk group analysis, ancestry stratified

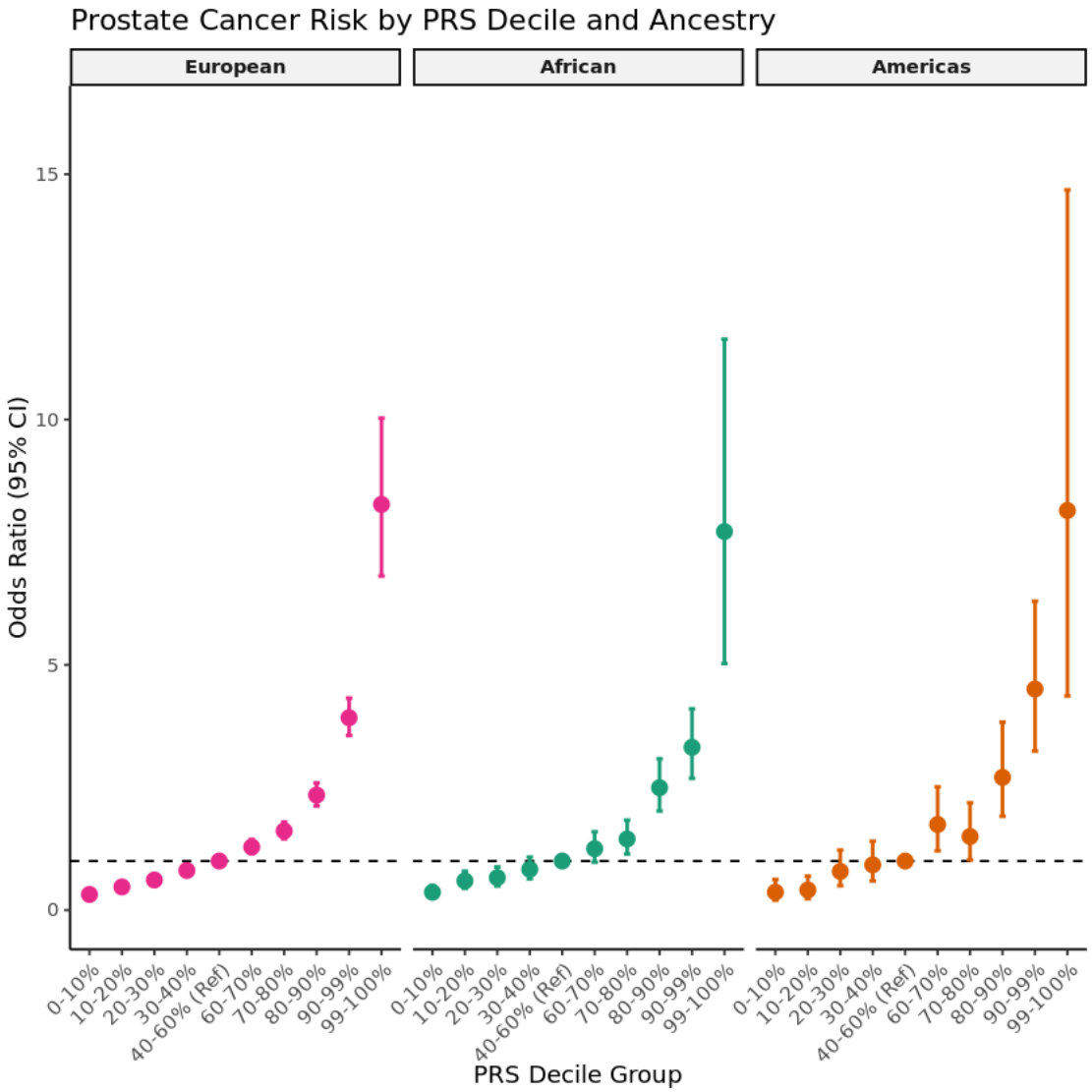


Table 3: Association Results for PRS risk group analysis (overall and ancestry-stratified)

Odds Ratios (OR) adjusted for age and principal components 1-10

PRS Risk Group	OR	95% CI	p-value
Overall (All Ancestries)			
0-10%	0.34	(0.30, 0.39)	<2e-16
10-20%	0.49	(0.44, 0.56)	<2e-16
20-30%	0.63	(0.56, 0.70)	2.40e-15
30-40%	0.81	(0.73, 0.91)	1.93e-04
40-60% (Ref)	*1.00*	*Reference*	*NA*
60-70%	1.29	(1.17, 1.42)	3.76e-07
70-80%	1.57	(1.43, 1.72)	<2e-16
80-90%	2.37	(2.17, 2.58)	<2e-16
90-99%	3.74	(3.44, 4.07)	<2e-16
99-100%	7.98	(6.76, 9.41)	<2e-16

Odds Ratios (OR) adjusted for age and principal components 1-10

PRS Risk Group	OR	95% CI	p-value
European			
0-10%	0.32	(0.27, 0.38)	<2e-16
10-20%	0.47	(0.41, 0.55)	<2e-16
20-30%	0.61	(0.54, 0.70)	7.92e-13
30-40%	0.81	(0.71, 0.91)	7.58e-04
40-60% (Ref)	*1.00*	*Reference*	*NA*
60-70%	1.29	(1.15, 1.44)	1.02e-05
70-80%	1.61	(1.45, 1.80)	<2e-16
80-90%	2.35	(2.12, 2.59)	<2e-16
90-99%	3.92	(3.56, 4.32)	<2e-16
99-100%	8.27	(6.81, 10.03)	<2e-16
African			

Odds Ratios (OR) adjusted for age and principal components 1-10

PRS Risk Group	OR	95% CI	p-value
0-10%	0.36	(0.26, 0.50)	2.08e-09
10-20%	0.60	(0.44, 0.79)	5.07e-04
20-30%	0.66	(0.49, 0.88)	5.33e-03
30-40%	0.83	(0.64, 1.08)	1.74e-01
40-60% (Ref)	*1.00*	*Reference*	*NA*
60-70%	1.25	(0.98, 1.60)	7.31e-02
70-80%	1.45	(1.14, 1.83)	2.08e-03
80-90%	2.50	(2.02, 3.09)	<2e-16
90-99%	3.32	(2.69, 4.10)	<2e-16
99-100%	7.72	(5.02, 11.64)	<2e-16
Americas			
0-10%	0.37	(0.20, 0.63)	4.42e-04

Odds Ratios (OR) adjusted for age and principal components 1-10

PRS Risk Group	**OR**	**95% CI**	**p-value**
10-20%	0.41	(0.23, 0.69)	1.45e-03
20-30%	0.79	(0.50, 1.22)	3.00e-01
30-40%	0.92	(0.60, 1.40)	7.11e-01
40-60% (Ref)	*1.00*	*Reference*	*NA*
60-70%	1.74	(1.21, 2.51)	2.79e-03
70-80%	1.50	(1.02, 2.19)	3.61e-02
80-90%	2.71	(1.91, 3.83)	1.82e-08
90-99%	4.51	(3.24, 6.30)	<2e-16
99-100%	8.15	(4.37, 14.68)	9.84e-12

Table 4: Counts of variants in each annotation category by gene

Note: Variants can be categorized into both pLOF and del. Missense, but were only counted once for the joint set

Chromosome	**Gene**	**pLOF**	**Deleterious missense (del. missense)**	**pLOF and del missense**	**Synonymous**
22	CHEK2	136	387	522	294
22	BIK	70	23	93	120
20	SAMHD1	73	226	294	254
17	TP53	76	209	244	428
17	RAD51D	67	206	273	255
17	HOXB13	45	254	298	162
16	PALB2	172	319	491	589
13	BRCA2	456	697	1136	1834
11	ATM	446	127	216	282
11	FAM111A	89	1614	2036	1727
8	NBN	138	326	464	451
7	PMS2	97	354	450	445
3	MLH1	81	527	688	550

2	EPCAM	63	203	266	242
2	MSH2	95	801	887	987
2	MSH6	221	961	1174	1429
Total	**16 genes**	**2325**	**7234**	**9437**	**10049**

Table 5: Rare Variant Analyses Results by Mask

Table 5A. ORs of Pathogenic Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
HOXB13 *	4.14	(3.064, 5.600)	2.46×10^{-20}
SAMHD1	2.04	(0.898, 4.626)	0.088
BRCA2 *	1.94	(1.402, 2.695)	6.61×10^{-5}
ATM *	1.53	(1.065, 2.194)	0.021
TP53	1.48	(0.743, 2.939)	0.265
BIK *	1.40	(1.186, 1.643)	6.09×10^{-5}
FAM111A *	1.31	(1.056, 1.620)	0.014
CHEK2	1.24	(0.906, 1.711)	0.177
MSH2	1.10	(0.864, 1.389)	0.451
MSH6	0.98	(0.703, 1.358)	0.891
RAD51D	0.91	(0.296, 2.788)	0.868
PMS2	0.88	(0.433, 1.797)	0.730

Table 5A. ORs of Pathogenic Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
PALB2	0.76	(0.346, 1.652)	0.483
NBN	0.74	(0.337, 1.629)	0.455
MLH1	0.53	(0.116, 2.411)	0.410
EPCAM *	0.13	(0.018, 0.861)	0.035

Table 5B. ORs of Deleterious Missense Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
HOXB13 *	2.31	(1.743, 3.052)	4.99×10^{-9}
SAMHD1	1.35	(0.830, 2.210)	0.225
CHEK2 *	1.32	(1.137, 1.526)	2.39×10^{-4}
BRCA2	1.12	(0.998, 1.254)	0.053
MSH2	1.11	(0.915, 1.346)	0.290
NBN	1.10	(0.966, 1.247)	0.153

Table 5B. ORs of Deleterious Missense Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
ATM *	1.09	(1.006, 1.188)	0.035
TP53	1.05	(0.676, 1.637)	0.821
PALB2	1.04	(0.886, 1.228)	0.614
MSH6	1.04	(0.895, 1.214)	0.593
MLH1	1.03	(0.898, 1.174)	0.701
FAM111A	0.99	(0.679, 1.436)	0.948
RAD51D	0.92	(0.674, 1.264)	0.619
EPCAM	0.91	(0.747, 1.100)	0.319
PMS2	0.89	(0.747, 1.065)	0.205
BIK	0.65	(0.191, 2.241)	0.500

Table 5C. ORs of Pathogenic and Deleterious Missense Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
HOXB13 *	2.37	(1.835, 3.070)	4.52×10^{-11}
SAMHD1	1.49	(0.962, 2.310)	0.074
BIK *	1.37	(1.169, 1.615)	1.19×10^{-4}
CHEK2 *	1.31	(1.147, 1.502)	7.92×10^{-5}
TP53	1.25	(0.827, 1.875)	0.293
FAM111A *	1.22	(1.014, 1.472)	0.035
BRCA2 *	1.17	(1.054, 1.309)	0.004
MSH2	1.12	(0.967, 1.308)	0.128
ATM *	1.11	(1.020, 1.201)	0.015
NBN	1.08	(0.956, 1.230)	0.208
PALB2	1.03	(0.874, 1.204)	0.754
MLH1	1.02	(0.893, 1.166)	0.767

Table 5C. ORs of Pathogenic and Deleterious Missense Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
MSH6	1.01	(0.882, 1.167)	0.841
RAD51D	0.92	(0.676, 1.241)	0.572
PMS2	0.89	(0.748, 1.056)	0.180
EPCAM	0.88	(0.726, 1.065)	0.188

Table 5D. ORs of Synonymous Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
ATM *	1.11	(1.047, 1.171)	3.77×10^{-4}
CHEK2	1.11	(0.961, 1.274)	0.161
FAM111A	1.10	(0.987, 1.221)	0.087
NBN	1.08	(0.997, 1.169)	0.058
MSH6	1.05	(0.977, 1.127)	0.190

Table 5D. ORs of Synonymous Variant Sets in 16 PCa-associated Genes

Gene	Odds Ratio (OR)	95% CI	P-value
TP53	1.03	(0.934, 1.139)	0.539
BRCA2	1.02	(0.963, 1.091)	0.437
MLH1	1.01	(0.935, 1.098)	0.740
BIK	1.00	(0.819, 1.215)	0.980
RAD51D	0.99	(0.876, 1.120)	0.874
PALB2	0.97	(0.886, 1.062)	0.514
PMS2	0.97	(0.892, 1.055)	0.475
EPCAM	0.97	(0.879, 1.062)	0.472
MSH2	0.95	(0.867, 1.032)	0.213
SAMHD1	0.89	(0.753, 1.060)	0.196
HOXB13	0.88	(0.731, 1.060)	0.178

Figure 5: Rare Variant Analyses Results by gene

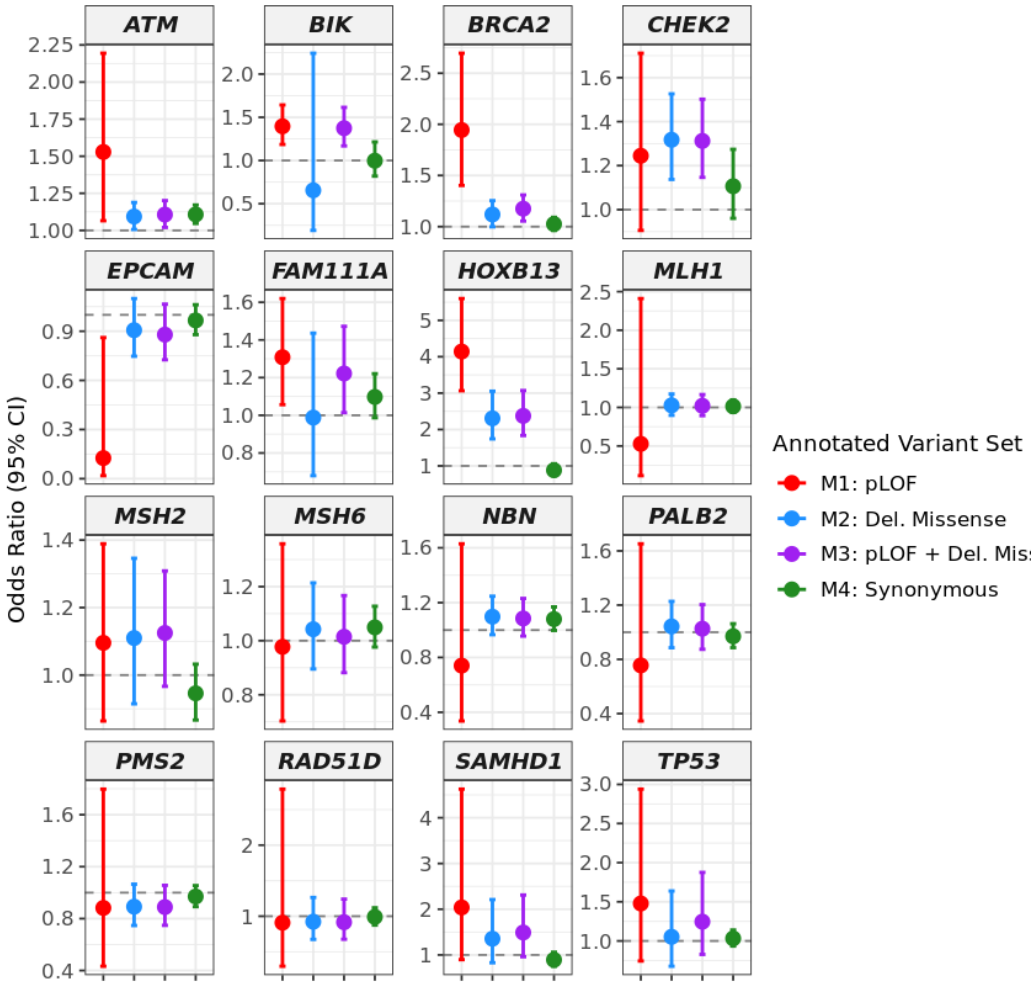


Table 6: Rare Variant Carrier Frequencies by Gene

Table 6. Carrier Frequencies by Gene & Disease Status

Carriers have at least 1 pLOF variant in any of the tested genes

	Overall Cohort			Controls			Cases		
Gene	N	Total	Freq (%)	N	Total	Freq (%)	N	Total	Freq (%)
Any Gene	**3974**	**82802**	**4.80%**	**3463**	**75225**	**4.60%**	**511**	**7577**	**6.74%**
BIK	1722	82802	2.08%	1530	75225	2.03%	192	7577	2.53%
FAM111A	950	82802	1.15%	839	75225	1.12%	111	7577	1.46%
HOXB13	292	82802	0.35%	211	75225	0.28%	81	7577	1.07%
BRCA2	375	82802	0.45%	323	75225	0.43%	52	7577	0.69%
CHEK2	404	82802	0.49%	355	75225	0.47%	49	7577	0.65%
ATM	306	82802	0.37%	266	75225	0.35%	40	7577	0.53%

Figure 6: Odds ratios for joint PRS and carrier status risk categories

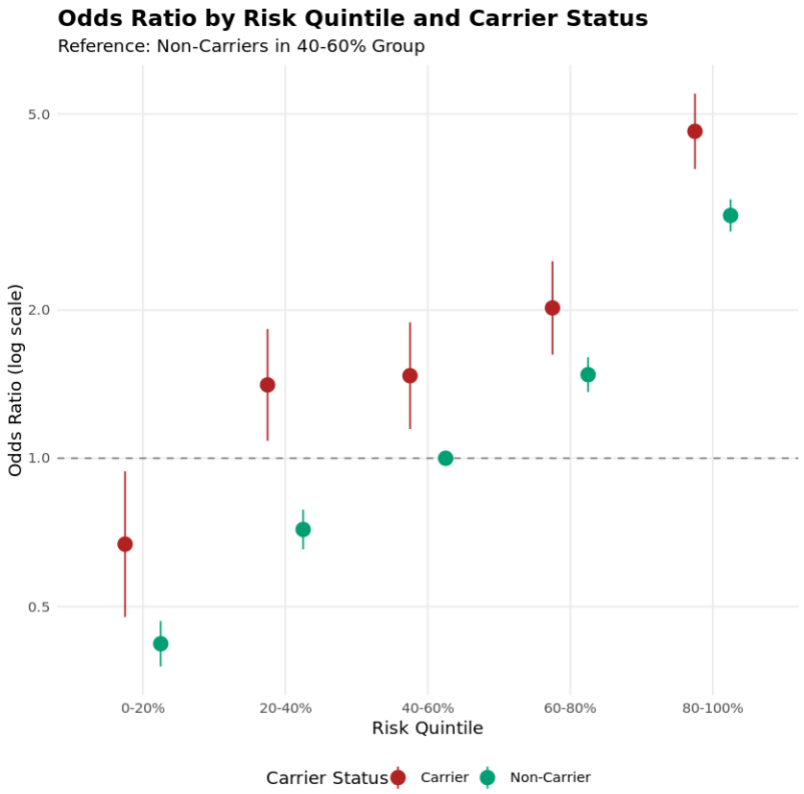


Table 7: Odds ratios for categories from joint PRS and carrier status analysis

PRS Quintile	Status	Beta	OR	95% CI	p-value
0-20%	Non-Carrier	-0.87	0.42	(0.38, 0.47)	2.35×10^{-57}
0-20%	Carrier	-0.40	0.67	(0.48, 0.94)	0.021
20-40%	Non-Carrier	-0.33	0.72	(0.65, 0.79)	1.78×10^{-12}
20-40%	Carrier	0.34	1.41	(1.09, 1.83)	0.010
40-60%	*Non-Carrier*	*0.00*	*1.00*	*Reference*	—
40-60%	Carrier	0.39	1.47	(1.15, 1.89)	0.002
60-80%	Non-Carrier	0.39	1.48	(1.36, 1.60)	5.11×10^{-21}
60-80%	Carrier	0.70	2.02	(1.62, 2.51)	2.96×10^{-10}

80-100%	Non-Carrier	1.14	3.11	(2.89, 3.36)	1.57×10^{-192}
80-100%	Carrier	1.53	4.61	(3.87, 5.50)	6.07×10^{-65}

Figure S1: ROC Curves for model with covariates and PRS vs PC only

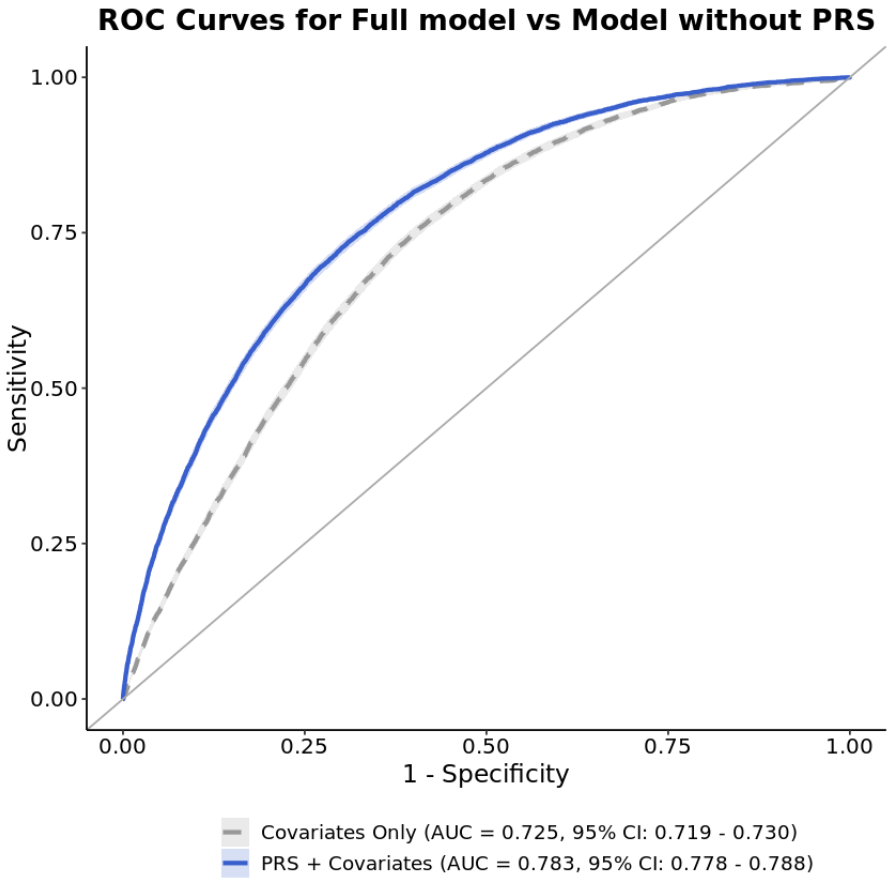
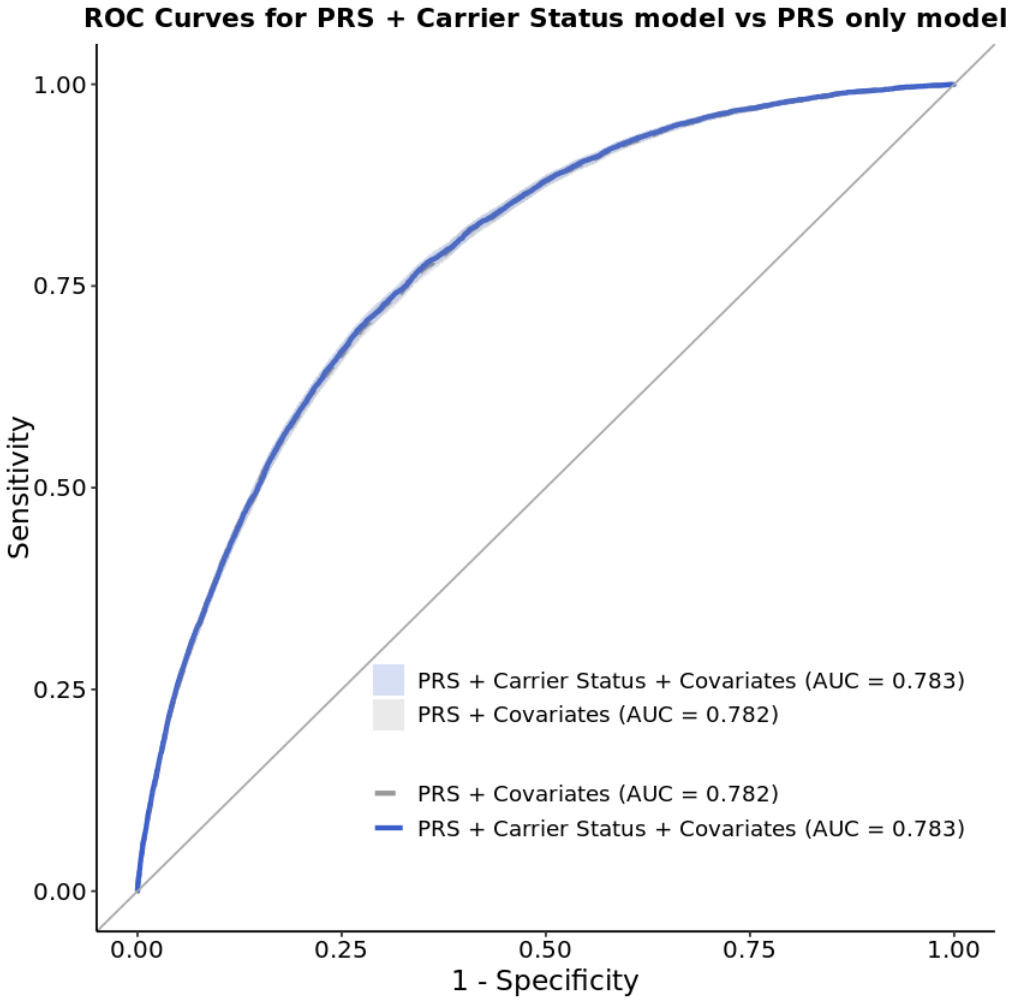


Figure S2: ROC curves for model with PRS, carrier status, and covariates vs PRS and covariates only



References

1. Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **74**, 229–263 (2024).
2. Mucci, L. A. *et al.* Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA* **315**, 68–76 (2016).
3. Kratzer, T. B. *et al.* Prostate cancer statistics, 2025. *CA: A Cancer Journal for Clinicians* **75**, 485–497 (2025).
4. Matejcic, M. *et al.* Pathogenic Variants in Cancer Predisposition Genes and Prostate Cancer Risk in Men of African Ancestry. *JCO Precis Oncol* **4**, PO.19.00179 (2020).
5. Lowder, D. *et al.* Racial disparities in prostate cancer: A complex interplay between socioeconomic inequities and genomics. *Cancer Lett* **531**, 71–82 (2022).
6. Bigler, S. A., Pound, C. R. & Zhou, X. A Retrospective Study on Pathologic Features and Racial Disparities in Prostate Cancer. *Prostate Cancer* **2011**, 239460 (2011).
7. Gaines, A. R. *et al.* The association between race and prostate cancer risk on initial biopsy in an equal access, multiethnic cohort. *Cancer Causes Control* **25**, 1029–1035 (2014).
8. Catalona, W. J., Antenor, J. A. V. & Roehl, K. A. SCREENING FOR PROSTATE CANCER IN HIGH RISK POPULATIONS.
9. Race-specific prostate cancer outcomes in a cohort of military health care beneficiaries undergoing surgery: 1990–2017.

https://onlinelibrary.wiley.com/doi/epdf/10.1002/cam4.4787?src=getftr&utm_source=sciencedirect_contenthosting&getft_integrator=sciencedirect_contenthosting.

10. National Academies of Sciences, Engineering, and Medicine; Division of Behavioral and Social Sciences and Education; Health and Medicine Division; Committee on Population; Board on Health Sciences Policy; Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. (National Academies Press (US), Washington (DC), 2023).
11. Conti, D. V. *et al.* Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet* **53**, 65–75 (2021).
12. Li, S. *et al.* Non-coding genetic variants underlying higher prostate cancer risk in men of African ancestry. *Nat Commun* **16**, 10202 (2025).
13. Prostate Cancer Risk Factors | Risk Factors for Prostate Cancer. https://www.cancer.org/cancer/types/prostate-cancer/causes-risks-prevention/risk-factors.html.
14. CDC. Prostate Cancer Risk Factors. *Prostate Cancer* https://www.cdc.gov/prostate-cancer/risk-factors/index.html (2025).
15. Wang, A. *et al.* Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. *Nat Genet* **55**, 2065–2074 (2023).
16. Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nat Genet* **48**, 30–35 (2016).

17. Claussnitzer, M. *et al.* A brief history of human disease genetics. *Nature* **577**, 179–189 (2020).
18. Wu, X. & Gu, J. Heritability of prostate cancer: a tale of rare variants and common single nucleotide polymorphisms. *Annals of Translational Medicine* **4**, 206–206 (2016).
19. Mitchell, J. *et al.* Assessing the contribution of rare protein-coding germline variants to prostate cancer risk and severity in 37,184 cases. *Nat Commun* **16**, 1779 (2025).
20. Ewing, C. M. *et al.* Germline Mutations in HOXB13 and Prostate-Cancer Risk. *N Engl J Med* **366**, 141–149 (2012).
21. Darst, B. F. *et al.* A Rare Germline *HOXB13* Variant Contributes to Risk of Prostate Cancer in Men of African Ancestry. *European Urology* **81**, 458–462 (2022).
22. BRCA Gene Changes: Cancer Risk and Genetic Testing Fact Sheet - NCI. https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet (2024).
23. Leongamornlert, D. *et al.* Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease. *Br J Cancer* **110**, 1663–1672 (2014).
24. Ivarsdottir, E. V. *et al.* Gene-based burden tests of rare germline variants identify six cancer susceptibility genes. *Nat Genet* **56**, 2422–2433 (2024).
25. Pavlovich, C. P. *et al.* Confirmation of BIK and SAMHD1 as Prostate Cancer Susceptibility Genes. *Prostate* **85**, 1556–1561 (2025).
26. Pritchard, C. C. *et al.* Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N Engl J Med* **375**, 443–453 (2016).

27. Bancroft, E. K. *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. *Lancet Oncol* **22**, 1618–1631 (2021).
28. Wokołorczyk, D. *et al.* PALB2 mutations and prostate cancer risk and survival. *Br J Cancer* **125**, 569–575 (2021).
29. Rusak, B. *et al.* Inherited NBN Mutations and Prostate Cancer Risk and Survival. *Cancer Res Treat* **51**, 1180–1187 (2019).
30. Schaid, D. J. *et al.* Two-stage Study of Familial Prostate Cancer by Whole-exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with the Risk of Prostate Cancer. *Eur Urol* **79**, 353–361 (2021).
31. Ni, J. *et al.* Epithelial cell adhesion molecule (EpCAM) is associated with prostate cancer metastasis and chemo/radioresistance via the PI3K/Akt/mTOR signaling pathway. *Int J Biochem Cell Biol* **45**, 2736–2748 (2013).
32. Shevach, J. W. *et al.* A population-based study of established cancer predisposition genes and single and multiple cancers. *JAMA Oncol* **11**, 1222–1230 (2025).
33. Esserman, L. J. *et al.* Risk-Based vs Annual Breast Cancer Screening: The WISDOM Randomized Clinical Trial. *JAMA* **335**, 763–774 (2026).
34. McHugh, J. K. *et al.* Assessment of a Polygenic Risk Score in Screening for Prostate Cancer. *New England Journal of Medicine* **392**, 1406–1417 (2025).
35. About the Multi-Gene Panel Test for Hereditary Prostate Cancer | Memorial Sloan Kettering Cancer Center. https://www.mskcc.org/cancer-care/patient-education/about-multi-gene-panel-test-hereditary-prostate.

36. Darst, B. F. *et al.* Combined Effect of a Polygenic Risk Score and Rare Genetic Variants on Prostate Cancer Risk. *Eur Urol* **80**, 134–138 (2021).
37. Kang, J. H. *et al.* Polygenic risk and rare variant gene clustering enhance cancer risk stratification for breast and prostate cancers. *Commun Biol* **7**, 1289 (2024).
38. Hughley, R. W. *et al.* Polygenic Risk Score Modifies Prostate Cancer Risk of Pathogenic Variants in Men of African Ancestry. *Cancer Res Commun* **3**, 2544–2550 (2023).
39. Darst, B. F. *et al.* Germline Sequencing DNA Repair Genes in 5545 Men With Aggressive and Nonaggressive Prostate Cancer. *J Natl Cancer Inst* **113**, 616–625 (2020).
40. Gallagher, S. *et al.* Association of a Polygenic Risk Score With Breast Cancer Among Women Carriers of High- and Moderate-Risk Breast Cancer Genes. *JAMA Netw Open* **3**, e208501 (2020).
41. Barnes, D. R. *et al.* Breast and Prostate Cancer Risks for Male BRCA1 and BRCA2 Pathogenic Variant Carriers Using Polygenic Risk Scores. *J Natl Cancer Inst* **114**, 109–122 (2022).
42. Lecarpentier, J. *et al.* Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores. *J Clin Oncol* **35**, 2240–2250 (2017).
43. null, null. The “All of Us” Research Program. *New England Journal of Medicine* **381**, 668–676 (2019).
44. Bick, A. G. *et al.* Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
45. Prostate Cancer | PheKB. https://phekb.org/phenotype/prostate-cancer-0.

46. All of Us Genomic Quality Report. *User Support* https://support.researchallofus.org/hc/en-us/articles/29390274413716-All-of-Us-Genomic-Quality-Report (2025).
47. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
48. Cavalli-Sforza, L. L. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet* **6**, 333–340 (2005).
49. Cheng, S. *et al.* Validation and context-dependent effects of a prostate cancer polygenic risk score in the All of Us Research Program. *Am J Hum Genet* **113**, 392–398 (2026).
50. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
51. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886–D894 (2019).
52. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877–885 (2016).
53. Hammermeister Suger, A., Lindström, S., Darst, B. F. & Wu, M. Rare genetic variant contributions to multiple cancers. (2026).
54. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* **46**, D1062–D1067 (2018).
55. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097–1103 (2021).

56. Wang, X. Firth logistic regression for rare variant association tests. *Front Genet* **5**, 187 (2014).
57. HOXB13, a high-risk prostate cancer gene, also confers risk for breast cancer: novel variants of clinical significance, especially in hormone-positive patients | Journal of Human Genetics. https://www.nature.com/articles/s10038-026-01481-y.
58. Phan, L. M. & Rezaeian, A.-H. ATM: Main Features, Signaling Pathways, and Its Diverse Roles in DNA Damage Response, Tumor Suppression, and Cancer Development. *Genes (Basel)* **12**, 845 (2021).
59. Bensimon, A. *et al.* ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci Signal* **3**, rs3 (2010).
60. Shiloh, Y. & Ziv, Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol* **14**, 197–210 (2013).
61. Berkovich, E., Monnat, R. J. & Kastan, M. B. Roles of ATM and NBS1 in chromatin structure modulation and DNA double-strand break repair. *Nat Cell Biol* **9**, 683–690 (2007).
62. Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
63. Kaissarian, N. M., Meyer, D. & Kimchi-Sarfaty, C. Synonymous Variants: Necessary Nuance in Our Understanding of Cancer Drivers and Treatment Outcomes. *J Natl Cancer Inst* **114**, 1072–1094 (2022).

64. Zhang, L. *et al.* Identification of Synonymous Pathogenic Variants in Monogenic Disorders by Integrating Exome with Transcriptome Sequencing. *The Journal of Molecular Diagnostics* **26**, 267–277 (2024).
65. Nyberg, T., Tischkowitz, M. & Antoniou, A. C. BRCA1 and BRCA2 pathogenic variants and prostate cancer risk: systematic review and meta-analysis. *Br J Cancer* **126**, 1067–1081 (2022).
66. Saunders, E. J. *et al.* Identification of Genes with Rare Loss of Function Variants Associated with Aggressive Prostate Cancer and Survival. *European Urology Oncology* **7**, 248–257 (2024).
67. Pandya, V. *et al.* BIK drives an aggressive breast cancer phenotype through sublethal apoptosis and predicts poor prognosis of ER-positive breast cancer. *Cell Death Dis* **11**, 448 (2020).
68. Chinnadurai, G., Vijayalingam, S. & Rashmi, R. BIK-the founding member of the BH3-only family proteins: mechanisms of cell death and role in cancer and pathogenic processes. *Oncogene* **27**, S20–S29 (2008).
69. Naicker, D., Rhoda, C., Sunda, F. & Arowolo, A. Unravelling the Intricate Roles of FAM111A and FAM111B: From Protease-Mediated Cellular Processes to Disease Implications. *Int J Mol Sci* **25**, 2845 (2024).
70. Unger, S. *et al.* FAM111A Mutations Result in Hypoparathyroidism and Impaired Skeletal Development. *Am J Hum Genet* **92**, 990–995 (2013).
71. Hatziagapiou, K. *et al.* Kenny-Caffey Syndrome Type 2 (KCS2): A New Case Report and Patient Follow-Up Optimization. *J Clin Med* **14**, 118 (2024).

72. Skinner, H. G. & Schwartz, G. G. Serum calcium and incident and fatal prostate cancer in the National Health and Nutrition Examination Survey. *Cancer Epidemiol Biomarkers Prev* **17**, 2302–2305 (2008).

73. Kim, W. T. *et al.* Parathyroid hormone is associated with prostate cancer. *Prostate Int* **8**, 116–120 (2020).

74. Wang, Y., Dai, B. & Ye, D. CHEK2 mutation and risk of prostate cancer: a systematic review and meta-analysis. *Int J Clin Exp Med* **8**, 15708–15715 (2015).

75. Kim, S. Y. *et al.* Genomic Ascertainment of CHEK2-Related Cancer Predisposition. *JAMA Netw Open* **8**, e2549730 (2025).

76. Liao, Y., Wu, M., Jia, Y., Mou, R. & Li, X. EpCAM as a Novel Biomarker for Survivals in Prostate Cancer Patients. *Front. Cell Dev. Biol.* **10**, (2022).

77. Ni, J. *et al.* Epithelial cell adhesion molecule (EpCAM) is associated with prostate cancer metastasis and chemo/radioresistance via the PI3K/Akt/mTOR signaling pathway. *Int J Biochem Cell Biol* **45**, 2736–2748 (2013).

78. Liu, Y. *et al.* Understanding the versatile roles and applications of EpCAM in cancers: from bench to bedside. *Exp Hematol Oncol* **11**, 97 (2022).