nature cancer

Article

https://doi.org/10.1038/s43018-025-01103-0

Genomic risk model to implement precision prostate cancer screening in clinical care: the ProGRESS study

Received: 4 December 2024

Accepted: 24 November 2025

Published online: 26 January 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

Precision healthcare aims to tailor disease prevention and early detection to individual risk. Prostate cancer screening may benefit from genomics-informed approaches. We developed and validated the P-CARE model, a prostate cancer risk prediction tool combining a polygenic score, family history and genetic ancestry, using data from over 585,000 male participants in the Million Veteran Program. The model was externally validated in diverse cohorts and implemented via a blended genome–exome assay for clinical use. Here we show that the P-CARE model identifies clinically meaningful gradients of prostate cancer risk among men, with higher scores associated with increased risk of any, metastatic and fatal prostate cancer. The model is now being used in a clinical trial of precision prostate cancer screening. This work demonstrates the potential for genomics-enabled health systems to improve prostate cancer screening and prevention in men. ClinicalTrials.gov registration: NCT05926102.

Preventive healthcare is moving from a one-size-fits-all approach to more personalized, risk-adapted strategies. Individual risk prediction is an important step for developing tailored strategies for disease prevention and early detection. Risk prediction models can now incorporate larger and more complex arrays of clinical, genetic, environmental and other risk factors from more diverse populations[1]. Genomics specifically is increasingly demonstrating its potential to inform risk stratification and preventive strategies for several diseases[2–5]; however, this potential clinical utility remains only theoretical in the absence of prospective intervention studies demonstrating improved patient outcomes.

Healthcare systems linked to genomic biobanks thus have the opportunity both to generate knowledge about the clinical validity of genomic risk prediction and to demonstrate the clinical utility of implementing that knowledge in care[6]. These genomics-enabled learning healthcare systems can leverage knowledge-generating infrastructure to determine whether genomics and other novel risk predictors improve the effectiveness of disease screening and prevention within the healthcare system. The result is not only improved care for patients within that system but also potentially generalizable knowledge for patients in other settings.

Prostate cancer screening is one clinical context where a genomics-enabled learning healthcare system approach might be particularly beneficial. Prostate cancer is one of the most heritable cancers, and recent genomic discoveries have characterized the rare and common genetic variation underlying much of this heritability[7,8]. At the same time, clinical guidelines differ on which patient populations are most likely to experience net benefit from prostate cancer screening, including Black men or those with a family history of the disease[9–11]. Universal screening with prostate-specific antigen (PSA) reduces prostate cancer mortality but can also overdiagnose indolent disease and lead to unnecessary procedures and treatments[12–15]. The result is substantial variation in prostate cancer screening practices[16,17]. Clinical consensus is even less developed on whether genotype should play a role in prostate cancer risk stratification, despite the discovery of robust associations between risk and both single rare variants and polygenic scores[8].

Given this context, genomics-enabled learning health systems can lead the development of genomics-tailored approaches to prostate cancer screening and then evaluate the effectiveness of those approaches in clinical care. This evidence generation is an important step toward the development of clinical guidelines. Here, we

e-mail: jvassy@bwh.harvard.edu; tseibert@health.ucsd.edu

https://doi.org/10.1038/s43018-025-01103-0

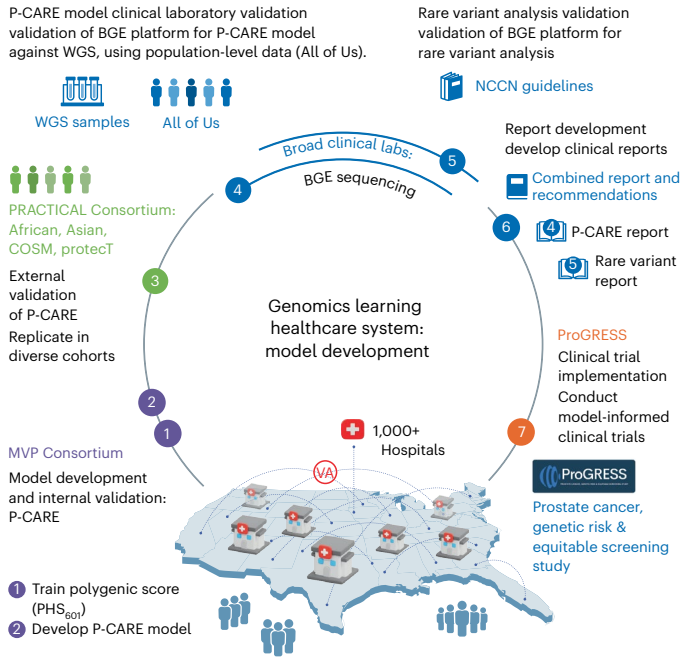


Fig. 1 | Translating prostate cancer genomic risk discovery to clinical trial implementation. A prostate cancer polygenic score (PHS_{601}) is trained in the MVP biobank of the VA using known prostate cancer and other prostate trait-associated loci (1). The P-CARE model is developed in MVP from PHS_{601}, genetic principal components and prostate cancer family history (2). Both PHS_{601} and P-CARE are replicated in external multi-ancestry datasets from the PRACTICAL Consortium (3). A BGE platform is validated for the P-CARE model, including imputation, analytic validation of PHS_{601} against WGS and clinical laboratory validation of P-CARE in AoU Research Program data (4). BGE platform is validated for gene panel annotation, filtering and analysis of rare variants in guideline-informed prostate cancer-associated genes (5). Clinical P-CARE and rare variant reports with summary recommendations are developed (6). Clinical laboratory analysis and reporting pipeline is implemented in a pragmatic clinical trial of precision prostate cancer screening across the VA (7).

describe the development, validation and clinical implementation of a genomics-informed prostate cancer risk model, developed to enable a randomized clinical trial of precision prostate cancer screening in a large national healthcare system (Clinicaltrials.gov NCT05926102).

Results

Overview of risk prediction model development

Figure 1 illustrates our discovery-to-implementation approach. The Prostate CAncer integrated Risk Evaluation (P-CARE) model was developed and validated to enhance genomic risk assessment for prostate cancer. Using data from the Million Veteran Program (MVP), a large biobank linked to the US Veterans Health Administration, we refined a prostate cancer polygenic score and integrated it with family history and genetic ancestry to create P-CARE. This model was externally validated in four diverse prostate cancer cohorts from the PRACTICAL Consortium. To facilitate clinical implementation, we developed a blended genome–exome (BGE) assay to assess both P-CARE and rare monogenic variants associated with prostate cancer risk. The assay is now being deployed in a randomized clinical trial (ProGRESS, NCT05926102) to evaluate genomics-informed prostate cancer screening in a real-world healthcare setting.

Association of polygenic score with prostate cancer outcomes

We assessed whether a polygenic hazard score (PHS_{601}) was significantly associated with prostate cancer risk, metastasis and mortality in both MVP and PRACTICAL as well as across ancestry groups. We hypothesized that PHS_{601} would show a strong, consistent association with these outcomes across diverse populations.

The final model included 601 of the 707 unique candidate variants evaluated (Supplementary Data). The resulting polygenic hazard score (PHS_{601}) was associated with age at diagnosis of prostate cancer, metastatic prostate cancer and prostate cancer death in MVP (Table 1). Among the overall MVP cohort, the hazard ratio (HR) per s.d. increase in PHS_{601} for prostate cancer, metastatic prostate cancer and prostate cancer death were 2.02 (95% CI 1.97–2.07), 2.07 (95% CI 1.95–2.17) and 1.96 (95% CI 1.75–2.18), respectively. The associations between PHS_{601} and prostate cancer outcomes were similar in each ancestry-stratified analysis with >100 events in MVP and within each ancestry-specific PRACTICAL dataset (Table 1). Among the East Asian subgroup in MVP, which had small case numbers, associations with metastatic and fatal prostate cancer were not statistically significant but had consistent directions of effect; the association with clinically significant disease was statistically significant in the Asian cohort in PRACTICAL (HR 2.11 95% CI 1.90–2.39). Among the American subgroup in MVP, the association with fatal prostate cancer was not statistically significant but had a consistent direction of effect (HR 2.22, 95% CI 0.98–4.25).

Association of P-CARE with prostate cancer outcomes

We evaluated whether P-CARE, which integrates a polygenic risk score with genetic ancestry and family history, improves prostate cancer risk stratification and correlates with disease severity. Family history was independently statistically significant for prostate cancer risk stratification (Supplementary Table 1) and inclusion of genetic ancestry improved performance of our previous PHS[18], so both were included in the model a priori. As hypothesized, the P-CARE model that integrated PHS_{601}, genetic ancestry and family history described a strong gradient of risk for any, clinically significant, metastatic and fatal prostate cancer across MVP and PRACTICAL datasets (Table 2 and Supplementary Tables 2–5). Among the overall MVP cohort, the HR per s.d. increase in P-CARE for prostate cancer, metastatic prostate cancer and prostate cancer death were 2.04 (95% CI 1.99–2.08), 2.05 (95% CI 1.93–2.16), and 1.95 (95% CI 1.76–2.15), respectively. Across the MVP and PRACTICAL datasets, compared to men with median P-CARE values, men in the lowest P-CARE quintile had HR 0.35–0.46 for the four prostate cancer outcomes ($HR_{20/50}$), whereas men in the highest P-CARE quintile had HR 2.48-4.03 ($HR_{80/50}$; Table 2). The direction and magnitude of association between P-CARE and the prostate cancer outcomes were similar in analyses of subgroups defined by genetic ancestry (Supplementary Table 6) and, alternatively, by self-reported race and ethnicity (Supplementary Table 7), in each subgroup with adequate case counts. As additional validation, time-dependent area under the curve analysis, sensitivity and specificity of defined P-CARE risk-category thresholds, and random forest survival modeling confirmed consistent model discrimination and robustness (Supplementary Tables 8-10)

Within the ProtecT (Prostate Testing for Cancer and Treatment) dataset, the positive predictive value (PPV) of a PSA value ≥ 3 ng ml^{-1} for clinically significant prostate cancer was 0.13 (95% CI 0.12–0.14) in the overall dataset and 0.19 (95% CI 0.16–0.21) and 0.23 (0.17-0.28) in the subsets in the top 20% and top 5% of P-CARE values, respectively (Fig. 2, stratified by PSA level in Extended Data Fig. 1). The percentage of true positive cases within the ProtecT dataset that fall into high P-CARE categories is shown in Extended Data Fig. 2.

We defined P-CARE risk categories by HR thresholds (HR 0.75 and HR 1.5 for metastatic prostate cancer) and evaluated both cumulative incidence and risk-equivalent age for any, metastatic and fatal prostate cancer. Overall, the model categorized 25.1%, 37.3% and 37.6% of MVP participants as low, average and high risk, respectively (Table 3). The model categorized 68.7% of participants with positive family history as high risk and only 5.6% as low risk. Among participants self-reporting Black or African-American race, only 2.8% were categorized as low risk. Figure 3 shows cumulative prostate cancer incidence curves in MVP

https://doi.org/10.1038/s43018-025-01103-0

Table 1 | Association of polygenic score with prostate cancer outcomes in MVP and PRACTICAL cohorts

Clinical end point	*n*	Event, *n*	HR (95% CI)				
			HR_{SD}	$HR_{80/20}$	$HR_{20/50}$	$HR_{80/50}$	$HR_{95/50}$
MVP development and validation							
Any prostate cancer							
All	585,418	68,618	2.02 (1.97–2.07)	6.27 (5.85–6.72)	0.43 (0.41–0.45)	2.72 (2.62–2.82)	4.06 (3.86–4.29)
African	105,014	16,178	1.94 (1.84–2.06)	5.81 (5.06–6.79)	0.48 (0.45–0.52)	2.81 (2.59–3.08)	3.74 (3.37–4.23)
European	420,722	48,178	2.04 (1.97–2.11)	5.74 (5.26–6.20)	0.43 (0.41–0.45)	2.46 (2.35–2.55)	3.78 (3.54–4.03)
American	50,590	3,775	2.03 (1.83–2.26)	5.55 (4.24–7.07)	0.44 (0.39–0.50)	2.44 (2.13–2.79)	3.72 (3.06–4.52)
East Asian	9,092	487	2.10 (1.59–2.81)	5.81 (2.90–10.65)	0.44 (0.32–0.60)	2.41 (1.72–3.36)	3.83 (2.31–6.04)
Metastatic prostate cancer							
All	585,418	6,606	2.07 (1.95–2.17)	6.69 (5.70–7.62)	0.42 (0.40–0.45)	2.81 (2.58–3.01)	4.27 (3.78–4.71)
African	105,014	1,726	1.90 (1.65–2.19)	5.60 (3.81–7.84)	0.50 (0.43–0.58)	2.73 (2.18–3.39)	3.62 (2.70–4.77)
European	420,722	4,467	1.96 (1.80–2.19)	5.20 (4.19–6.83)	0.45 (0.39–0.50)	2.33 (2.08–2.68)	3.50 (2.97–4.29)
American	50,590	369	2.07 (1.42–2.64)	6.02 (2.33–10.47)	0.44 (0.32–0.67)	2.51 (1.55–3.40)	3.92 (1.91–6.08)
East Asian	9,092	44	–	–	–	–	–
Fatal prostate cancer							
All	585,418	1,709	1.96 (1.75–2.18)	5.81 (4.31–7.65)	0.45 (0.40–0.51)	2.60 (2.22–3.03)	3.82 (3.06–4.72)
African	105,014	365	1.62 (1.16–2.12)	3.81 (1.49–7.26)	0.61 (0.44–0.85)	2.15 (1.27–3.22)	2.68 (1.35–4.45)
European	420,722	1,250	1.92 (1.60–2.20)	4.99 (3.18–6.86)	0.47 (0.39–0.57)	2.27 (1.81–2.69)	3.39 (2.41–4.32)
American	50,590	87	2.22 (0.98–4.25)	8.45 (0.95–32.23)	0.48 (0.19–1.03)	2.78 (0.97–6.14)	4.81 (0.96–14.32)
East Asian	9,092	8	–	–	–	–	–
PRACTICAL replication							
Any prostate cancer							
COSM	3,415	2,298	2.27 (2.11–2.46)	9.18 (6.66–12.93)	0.36 (0.31–0.42)	3.34 (2.97–3.98)	5.35 (4.26–6.92)
ProtecT	6,411	1,583	1.87 (1.78–2.01)	5.67 (4.75–6.73)	0.44 (0.40–0.48)	2.78 (2.47- 3.05)	3.78 (3.31–4.29)
African	6,253	3,240	1.84 (1.72–1.98)	8.55 (6.64–11.08)	0.41 (0.36–0.46)	3.47 (2.93–3.99)	4.49 (3.71–5.37)
Asian	2,320	1,164	2.15 (1.92–2.38)	8.80 (6.73–11.09)	0.35 (0.30–0.39)	3.02 (2.63–3.39)	5.26 (4.24–6.48)
Clinically significant prostate cancer							
COSM	3,415	1,487	2.30 (2.09–2.53)	9.35 (7.32–12.20)	0.36 (0.31–0.41)	2.81 (2.58–3.01)	4.27 (3.78–4.71)
ProtecT	6,411	628	2.02 (1.88–2.21)	7.02 (5.65–8.42)	0.40 (0.36–0.44)	2.73 (2.18–3.39)	4.45 (3.76–5.11)
African	6,253	1,424	1.85 (1.69–2.02)	8.61 (6.50–11.61)	0.41 (0.35–0.47)	2.51 (1.55–3.40)	3.92 (1.91–6.08)
Asian	2,320	716	2.11 (1.90–2.39)	7.88 (5.94–10.68)	0.36 (0.32–0.42)	2.85 (2.44–3.32)	4.83 (3.84–6.09)
Fatal prostate cancer							
COSM	3,415	278	1.91 (1.65–2.28)	5.88 (3.51–9.19)	0.45 (0.36–0.56)	2.58 (1.97–3.32)	3.82 (2.59–5.35)

Association of PHS_{601} with any, metastatic and fatal prostate cancer in MVP (total and genetic ancestry-stratified groups) and with any, clinically significant, and fatal prostate cancer in four PRACTICAL Consortium datasets. Results with fewer than 50 events per subset were excluded given the unstable nature of the HR estimates. COSM, Cohort of Swedish Men; PHS, polygenic hazard score.

both by P-CARE percentile groups and by P-CARE risk category. As shown in Table 4, by age 80 years, men in the high-risk P-CARE group had a cumulative risk of any, metastatic and fatal prostate cancer of 37.4%, 4.4% and 0.8%, respectively. The expected age of any and metastatic prostate cancer occurred 5 years earlier in the high-risk group compared to the men in the standard risk; specifically, a man in the high-risk group reached a prostate cancer detection risk equivalent to the 55-year standard at an age of 50 years and a metastatic prostate cancer risk equivalent to the 70-year standard at an age of 63.5 years (Supplementary Table 11).

Development and validation of clinical laboratory assay for genetic prostate cancer risk

We then used the BGE platform to develop a clinical laboratory assay for inherited prostate cancer risk by combining P-CARE, which evaluates polygenic risk, with targeted testing for 12 genes known to be associated with hereditary prostate cancer. First, both PHS_{601} and the integrated P-CARE model were again externally validated in the All of Us (AoU) cohort, demonstrating strong associations with prostate cancer across diverse ancestry groups. Within the AoU dataset, the PHS_{601} was associated with prostate cancer with an odds ratio per s.d. of 1.91 (95% CI 1.85–1.98). In the same dataset, for the full P-CARE model (PHS_{601} plus genetic principal components and family history) we found an odds ratio of 2.41 (95% CI 2.25–2.60) for individuals in the high-risk category to be diagnosed with prostate cancer, compared to individuals classified as average risk. Similarly, individuals in the low-risk category show an odds ratio of 0.48 (95% CI 0.44–0.54), compared to individuals classified as average risk. Notably, this strong association holds across different ancestries. (Extended Data Fig. 3).

Next, the accuracy and reliability of the BGE assay in detecting both polygenic and rare monogenic variants associated with prostate cancer risk were evaluated against known reference samples; the BGE

https://doi.org/10.1038/s43018-025-01103-0

Table 2 | Association of P-CARE model with prostate cancer outcomes in MVP and PRACTICAL cohorts

Clinical end point	*n*	HR (95% CI)				
		HR_{SD}	$HR_{80/20}$	$HR_{20/50}$	$HR_{80/50}$	$HR_{95/50}$
Any prostate cancer						
MVP	585,418	2.04 (1.99–2.08)	6.33 (5.95–6.71)	0.43 (0.42–0.45)	2.75 (2.66–2.84)	4.09 (3.89–4.29)
COSM	3,415	2.33 (2.14–2.58)	9.40 (6.88–13.20)	0.37 (0.31–0.42)	3.43 (2.91–4.08)	5.45 (4.31–7.00)
ProtecT	6,411	1.87 (1.77–2.01)	5.53 (4.60–6.66)	0.45 (0.41–0.49)	2.48 (2.27–2.73)	3.71 (3.23–4.23)
African	6,253	2.00 (1.86–2.16)	9.68 (7.55–12.49)	0.40 (0.36–0.46)	3.88 (3.35–4.42)	5.16 (4.35–6.03)
Asian	2,320	2.17 (1.95–2.42)	8.74 (6.60–11.28)	0.35 (0.30–0.40)	3.06 (2.65–3.46)	5.23 (4.09–6.47)
Clinically significant prostate cancer						
COSM	3,415	2.34 (2.12–2.55)	9.39 (7.25–12.27)	0.37 (0.32–0.42)	3.43 (2.98–3.99)	5.47 (4.47–6.76)
ProtecT	6,411	2.01 (1.87–2.18)	6.81 (5.42–8.20)	0.40 (0.36–0.44)	2.77 (2.48–3.04)	4.35 (3.71–4.12)
African	6,253	2.04 (1.86–2.23)	10.30 (7.66–13.47)	0.39 (0.36–0.44)	4.03 (3.39–4.86)	5.41 (4.38–6.77)
Asian	2,320	2.11 (1.91–2.35)	7.60 (5.70–10.15)	0.38 (0.33–0.43)	2.84 (2.42–3.31)	4.68 (3.71–5.73)
Metastatic prostate cancer						
MVP	585,418	2.05 (1.93–2.16)	6.50 (5.51–7.38)	0.43 (0.40–0.46)	2.78 (2.54–2.99)	4.17 (3.68–4.59)
Fatal prostate cancer						
MVP	585,418	1.95 (1.76–2.15)	5.71 (4.33–7.30)	0.45 (0.41–0.52)	2.59 (2.22–2.97)	3.77 (3.05–4.57)
COSM	3,415	1.95 (1.66–2.37)	5.95 (3.61–9.29)	0.46 (0.37–0.56)	2.65 (2.03–3.41)	3.87 (2.69–5.49)

Association of P-CARE model with any prostate cancer, clinically significant prostate cancer, metastatic prostate cancer, and fatal prostate cancer in MVP and four PRACTICAL Consortium datasets. As described, P-CARE model consists of PHS_{601}, first-degree family history of prostate cancer and genetic principal components.

platform produced nearly identical results for polygenic risk scores and ancestry estimates, with Pearson correlations exceeding $r > 0.998$ for PHS_{601} and $r > 0.999$ for both principal components. For the 12 genes related to hereditary prostate cancer risk, the assay met quality thresholds for coverage and variant detection in 11 of 12 genes. The one exception was *PMS2*, a technically challenging gene due to its similarity to a nearby pseudogene, which can interfere with accurate sequencing. Within the *PMS2* gene, exons 13, 14 and 15 were undercovered in a subset of samples, with 80% and 20% of samples missing full coverage in those regions, respectively. Of the 18 samples assessed for monogenic rare variants, all 18 variants of interest were successfully detected, including seven single-nucleotide variants (SNVs), five insertions/deletions (indels) and six copy-number variants (CNVs); however, three of the CNVs were classified as low quality based on prespecified thresholds for clinical reporting (QUAL ≥ 50 for duplications, QUAL ≥ 100 for heterozygous deletions and QUAL ≥ 400 for homozygous deletions) and would have not been clinically reported in a real-world setting. This is consistent with known limitations in detecting small CNVs involving fewer than three exons. Despite this, the platform showed excellent technical performance with 100% precision across repeated tests, both within and between sequencing runs.

Clinical P-CARE and monogenic reports

Here we describe the implementation of the P-CARE and monogenic risk reports in clinical use, linking them to personalized screening recommendations in the ProGRESS trial. An example of the resulting laboratory report package is shown in the Supplementary Information. The cover page summarizes the results of both the monogenic and P-CARE analyses and provides an overall risk category for the individual based on these results. An individual with a pathogenic or likely pathogenic variant in one of the 12 prostate cancer-associated genes is categorized as high risk, regardless of P-CARE results. Individuals without such a variant are categorized as low, average or high risk according to their P-CARE result, with thresholds at HR = 0.75 and HR = 1.5, as described in the Methods. The cover page also links these risk categories to tailored prostate cancer screening recommendations for the individual. After this cover page summary, separate P-CARE and rare variant reports provide further detail about these individual result types, including information about P-CARE model development and validation, technical descriptions of the analyses performed, relevant gene and disease information and literature references. These reports are now being used in the national ProGRESS randomized clinical trial, in which 5,000 VA patients who are prostate cancer screen-eligible are randomly assigned to usual care versus precision screening recommendations informed by P-CARE and rare variants.

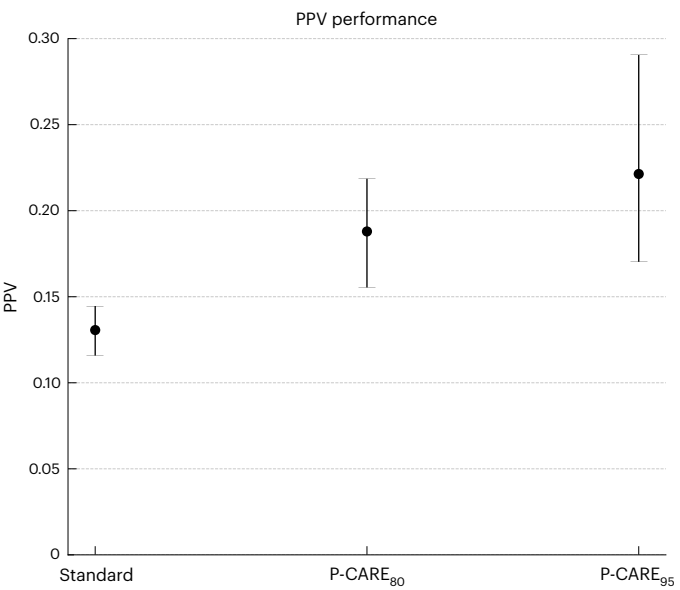


Fig. 2 | Positive predictive value of PSA in ProtecT by P-CARE values. Illustrated are mean PPV (95% CI) for a PSA value ≥ 3 ng ml^{-1} for clinically significant prostate cancer among three groups of men in the ProtecT study (n = 6,411): all men (regardless of P-CARE value), men in the top 20% of P-CARE values ($P\text{-}CARE_{80}$) and men in the top 5% of P-CARE values ($P\text{-}CARE_{95}$).

https://doi.org/10.1038/s43018-025-01103-0

Table 3 | Characteristics of P-CARE risk categories for metastatic prostate cancer among 585,418 MVP participants

		P-CARE risk category, *n* (%)		
	n	Low risk (HR<0.75)	Average risk (HR 0.75-1.5)	High risk (HR>1.5)
Total	585,418	146,826 (25.08)	218,530 (37.33)	220,062 (37.59)
Positive family history	28,358	1,595 (5.62)	7,270 (25.63)	19,493 (68.73)
Genetic ancestry groups				
African	105,014	2,607 (2.48)	19,314 (18.39)	83,093 (79.12)
European	420,722	128,096 (30.44)	173,774 (41.30)	118,852 (28.24)
American	50,590	12,842 (25.38)	21,518 (42.53)	16,230 (32.08)
East Asian	9,092	3,281 (36.08)	3,924 (43.15)	1,887 (20.75)
Self-reported race/ethnicity groups				
American Indian or Alaska Native	5,507	1,346 (24.44)	2,236 (40.60)	1,925 (34.95)
Asian	6,210	2,403 (38.69)	2,684 (43.22)	1,123 (18.08)
Black or African American	101,920	2,812 (2.75)	18,986 (18.62)	80,122 (78.61)
Hispanic white	26,037	6,871 (26.38)	11,067 (42.50)	8,099 (31.10)
Native Hawaiian or Pacific Islander	3,042	755 (24.81)	1,259 (41.38)	1,028 (33.79)
Non-Hispanic white	418,387	126,633 (30.26)	172,661 (41.26)	119,093 (28.46)
Other	8,077	1,938 (23.99)	3,303 (40.89)	2,836 (35.11)
Unknown	16,238	4,068 (25.05)	6,334 (39.00)	5,836 (35.94)

P-CARE risk categories are defined by thresholds of HR 0.75 and HR 1.5 for metastatic prostate cancer. Participants with a HR for metastatic prostate cancer <0.75 and >1.5 are defined as low and high risk, respectively.

Discussion

We used genomic, clinical and survey data from a large national biobank to develop a genomics-informed prostate cancer prediction model consisting of family history, genetic principal components and an updated polygenic score of 601 prostate trait-associated loci. Patients in the lowest and highest 20% of values under this model have 0.4-fold and 2.7-fold risk of prostate cancer, respectively, compared to those with median values; replication in external multi-ancestry cohorts confirmed these associations. Men at highest risk of developing advanced prostate cancer are most likely to benefit from screening; the P-CARE model is associated with risk of all, clinically significant, metastatic and fatal prostate cancer. When low and high risk were defined as HR < 0.75 and HR > 1.5, respectively, for metastatic prostate cancer, the cumulative incidence of metastatic prostate cancer by age 80 years in the biobank was 0.8% in the low-risk group and 4.4% in the high-risk group.

Unlike our previous PHS (PHS_{290}), both P-CARE and PHS_{601} have relatively similar performance at a population level to discriminate prostate cancer risk; family history and agnostic genetic ancestry have less prognostic value in the current multivariable model than in our previous model. However, the effect of family history is substantial for individuals so inclusion of family history could make a difference at an individual level in clinical decision-making. While ancestry seems to be mostly accounted for by PHS_{601}, we opted not to exclude this post hoc. We then developed and validated a clinical assay on a cost-efficient BGE platform for both the prediction model and rare pathogenic variants in known prostate cancer genes. This assay and associated clinical reports are now enabling a clinical trial of precision prostate cancer screening among patients receiving care from the national healthcare system from which the biobank data were derived. This approach illustrates the power of genomics-enabled learning health systems to generate translatable discoveries for implementation in preventive healthcare.

We designed the P-CARE model and ongoing prostate cancer screening trial to examine how the routine collection and interpretation of genomic data in preventive care might improve upon existing screening practices in a large integrated health system. Prostate cancer is highly prevalent, but despite randomized controlled trial evidence that screening with PSA testing can reduce prostate cancer mortality[14,19], guidelines vary by organization and country[11] on how to balance the benefits of screening (early detection and treatment, resulting in lower incidence of advanced and lethal disease) and its potential risks (overdiagnosis of apparently indolent disease and morbidity from unnecessary procedures and treatments). As a result, screening practices are highly variable[16,17,20–23]. Better models are needed to distinguish men most likely to benefit from screening from those for whom its risks might outweigh its benefits. A learning health system approach is ideal to improve prostate cancer screening for a few reasons. First, risk prediction models that inform the net benefit of cancer screening depend in large part on model calibration within a population; relative and absolute risk estimates derived from a healthcare system-linked biobank are thus particularly informative for patients receiving care in that system. In particular, age is a critical factor not only in the risk of advanced prostate cancer but also for the competing health risks that might make prostate cancer early detection less important[24]; our time-to-event analysis and age-specific cumulative incidence curves account for age and allow physicians to balance these with age-related competing risks for a given individual to guide age-based screening decisions. Second, the effect sizes of polygenic scores themselves, including for prostate cancer, can vary between biobanks[25,26]. Third, the net benefit of prostate cancer screening in a population is highly dependent on the downstream diagnostic and therapeutic management of elevated PSA values and abnormal prostate biopsy results[27,28]; nesting the evaluation of a new screening paradigm within its target healthcare delivery system helps ensure that system-specific clinical practice patterns are included.

Our approach also seeks to address controversies in prostate cancer screening that are intimately intertwined with health disparities. In the United States, Black men are more likely to be both diagnosed with and die from prostate cancer[29]. Possible causal factors include genetic, environmental and social determinants of health, including structural factors including access to screening and other healthcare[30–32]. Black men are highlighted in prostate cancer guidelines as a group whose high risk merits earlier screening[9,10]. This recommendation is appropriate to address racial disparities in prostate cancer outcomes; however, at the same time, the use of race in medical decision-making

https://doi.org/10.1038/s43018-025-01103-0

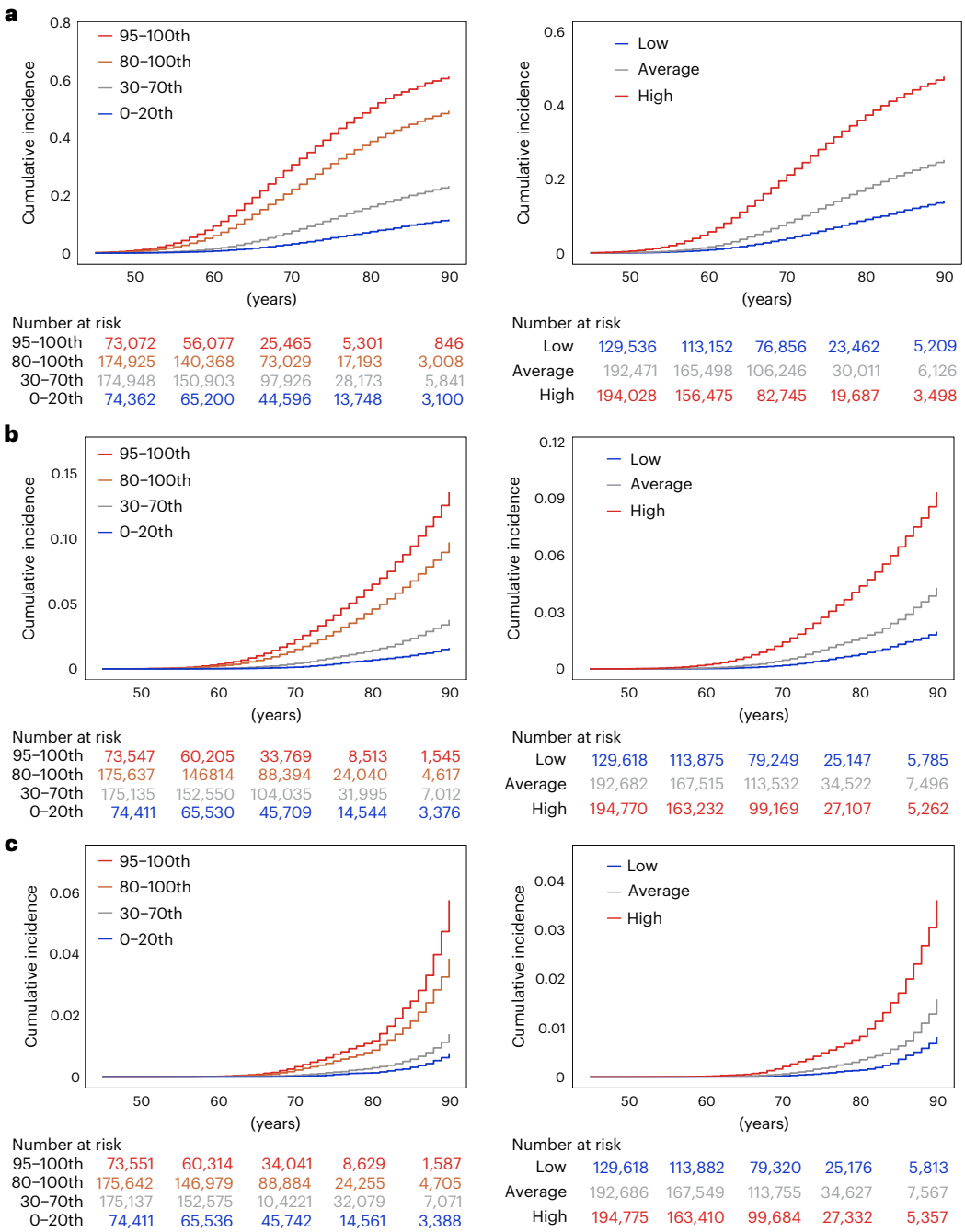


Fig. 3 | Prostate cancer cause-specific cumulative incidence in MVP by P-CARE strata. a–**c**, Cause-specific cumulative incidence within MVP for prostate cancer (**a**), metastatic prostate cancer (**b**) and fatal prostate cancer (**c**). The left column shows incidence for each end point by P-CARE percentile group: 0–20th, 30–70th, 80–100th and 95–100th. The right column shows incidence for each end point by P-CARE risk category: high, average and low risk.

can inappropriately ascribe to biology effects that arise from a complex social construct confounded by myriad social determinants of health; it also ignores the complex multiracial and multi-ancestry backgrounds of individuals in modern healthcare system populations. We thus set out to develop a prostate cancer risk prediction model that did not include discrete race or genetic ancestry categories, favoring instead, principal components as a continuous measure of genetic variation. At the same time, we confirmed that the resulting model performed well across categories of socially defined populations (race and ethnicity groups in MVP and external cohorts). Initial genome studies predominantly included individuals with European ancestry, but more recent work has improved genetic discovery and risk stratification in more diverse populations, including African ancestry[18,33–37]. The P-CARE model extends this work, confirming that most Black men, but not all, have high risk. While the model does not fully disentangle the confounded associations between genetic ancestry and social determinants of prostate cancer risk, it represents an advance toward a more equitable, tailored approach to risk stratification and screening that does not treat race as a biological construct.

Family history of prostate cancer and certain rare genetic variants are also known prostate cancer risk factors, independent of ancestry and polygenic score[18,33,34,37]. We designed the P-CARE model to build upon, not replace, these clinical risk factors, similar to breast cancer-screening models[38,39]. Rare variants in several genes, including *BRCA2* and *MSH2*, are known to increase prostate cancer risk and thus have separate screening guidelines for carriers[20]. Carrier status of

https://doi.org/10.1038/s43018-025-01103-0

Table 4 | Prostate cancer cause-specific cumulative incidence in MVP by P-CARE category

Clinical end point	Cumulative incidence (%)		
	Low risk (HR<0.75)	Average risk (HR 0.75-1.5)	High risk (HR>1.5)
Prostate cancer			
By age 70	4.02	8.38	21.22
By age 80	9.17	17.68	37.43
By age 90	14.11	25.29	47.87
Metastatic prostate cancer			
By age 70	0.18	0.46	1.41
By age 80	0.77	1.64	4.38
By age 90	1.96	4.29	9.34
Fatal prostate cancer			
By age 70	0.02	0.06	0.21
By age 80	0.13	0.34	0.82
By age 90	0.81	1.58	3.61

these variants is presently unknown for the vast majority of prostate cancer screen-eligible patients and yet might play a more prominent role in preventive care in a future when genomic testing is more commonplace. Despite aggregate analyses suggesting that polygenic scores can modify the effects of these rare variants[40–42], we determined that these modified associations are not yet robust enough for individual variant-level clinical reporting and should not supersede National Comprehensive Cancer Network (NCCN) guidelines for the clinical management of rare variants. We therefore chose a genomic analysis platform that could detect and interpret these important rare variants and will report them according to established clinical guidelines to participants. By combining high-coverage exome and low-coverage whole-genome sequencing (WGS) data, the new BGE technology provides a cost-efficient, scalable and accurate platform for implementing the P-CARE model in clinical care. In the ProGRESS trial (ClinicalTrials.gov ID NCT05926102) participants and their healthcare providers are now receiving clinical reports with P-CARE results and the results of rare variant analysis, enabling an evaluation of a precision screening approach on prostate cancer in the US Veterans Health Administration.

Previous modeling studies suggest that the use of polygenic scores can improve the cost-effectiveness of prostate cancer screening pathways with and without MRI[43,44]; the ProGRESS trial will provide additional empiric data to determine the costs and cost-effectiveness of such an approach in real-world implementation. Even if the cost-effectiveness of a polygenic approach to screening is marginal for single cancers[45,46], multiplex platforms such as the BGE enable both monogenic variant screening and polygenic risk stratification with one test. Modeling studies already suggest that population genomic screening for a select number of monogenic diseases is cost-effective[47,48]. More complex models incorporating polygenic risk for multiple diseases are needed, but it is plausible that as the costs of genomic testing decrease, the incremental cost-effectiveness of adding polygenic approaches to genomic medicine programs will be favorable, depending on patient population, healthcare setting and country.

Our work has some limitations. Despite the strengths of our learning healthcare system approach described above, this system-specific model may not generalize to other settings with different population risks and screening practices. Model replication in the diverse PRACTICAL and AoU datasets mitigates this concern, but other healthcare systems should examine model calibration in their own data before implementation. In addition, while the inclusion of family history, polygenic score, genetic principal components and rare variants improves upon existing clinical prostate cancer screening approaches, the P-CARE model cannot disentangle the effects of genetic predisposition from environmental exposures and other social determinants that shape prostate cancer risk. Ongoing and future work should examine how to model and include other important risk factors in a clinically implementable risk stratification tool, including the consideration of other machine-learning-based prediction approaches[29,49,50]. Finally, BGE has many benefits, including genome-level variant information for polygenic scores as well as an exome backbone for monogenic reporting, but there are limitations that come with an exome-based approach that a purpose built capture panel may overcome, including lower sensitivity around complex regions of genes like *PMS2* and reduced sensitivity of small CNVs below 3 exons in size.

In summary, a healthcare system-linked biobank has enabled the development, replication and clinical laboratory validation of an updated prostate cancer risk model, now implemented in a clinical trial of precision prostate cancer screening. This approach exemplifies the power of genomics-enabled-learning health systems to accelerate the discovery and translation of precision technologies to improve population health outcomes.

Methods

Study overview

The VA Central Institutional Review Board approved this study (IRBNet 1735869 and 1735136). As described in detail below, we used data from a large biobank linked to a national healthcare system to update a previous prostate cancer polygenic score[18]. We then developed and cross-validated a prostate cancer prediction model based on the combination of that score and family prostate cancer history, now termed the P-CARE model. We further validated the P-CARE model in four external prostate cancer cohort datasets before the development and validation of a clinical BGE assay both for the P-CARE model and also for rare prostate cancer-associated monogenic variants. This assay is now being implemented in a randomized clinical trial of genomics-informed prostate cancer screening in a new cohort of patients from the national healthcare system in which the P-CARE model was first developed (ProGRESS; ClinicalTrials.gov ID NCT05926102).

Participants and phenotype definitions

Genotype and phenotype data were analyzed from the following cohorts[33,51,52] (also summarized in Supplementary Tables 12 and 13).

Million Veteran Program. Data from the MVP were used to update a previous prediction model[18] to develop the new P-CARE model. MVP is a mega-biobank linked to the national Veterans Health Administration healthcare system of the US Department of VA[51]. Participants provide biospecimens, consent to research access to their VA health records and complete surveys about family health history, health behaviors, military and environmental exposures and other health-related factors. For the present analyses, we used data from 585,418 male MVP participants to develop and cross-validate the P-CARE model. All study participants provided blood samples for DNA extraction and genotyping using a custom Affymetrix Axiom biobank array containing 723,305 variants, enriched for low-frequency variants in African and Hispanic populations[53]. Family history was defined as the presence or absence of paternal history of prostate cancer, as reported on the MVP survey. Prostate cancer diagnosis, age at diagnosis and date of last follow-up were retrieved from the VA Corporate Data Warehouse based on International Classification of Diseases diagnosis codes and VA Central Cancer Registry data[18,54]. Age at diagnosis of metastasis (nodal and/or distant, regardless of whether metastases were detected at diagnosis or at recurrence) was determined via a validated natural language processing tool developed in the VA system[18,55]. Cause and date of death were obtained from the National Death Index. Fatal prostate cancer

was defined by ICD9 code 185 or ICD10 code C61 as the underlying cause of death.

PRACTICAL Consortium. Data from four external cohorts from the PRACTICAL Consortium were used to externally validate the P-CARE model. Data from 18,457 men previously genotyped via OncoArray or iCOGs arrays[56,57] were divided into four datasets, as described in previous studies evaluating polygenic scores: (1) men of African ancestry (n = 6,253); (2) men of Asian ancestry (n = 2,320); (3) the COSM population-based cohort with long-term outcomes (n = 3,415); and (4) the population-based ProtecT screening trial (n = 6,411)[33]. Family history was defined as the presence or absence of a first-degree relative with a prostate cancer diagnosis. Clinically significant prostate cancer was defined as any case with Gleason score ≥7, PSA ≥ 10 ng ml^{-1}, T3-T4 stage, nodal metastases or distant metastases[33]. The COSM dataset additionally had age at prostate cancer death[58] and the ProtecT dataset had prostate biopsy results for both cases and controls with screening PSA ≥ 3 ng ml^{-1}[59,60].

All of Us Research Program

Data from the v.7 release of the AoU Research Program were used as an additional external validation cohort. Excluding samples flagged for failing quality control criteria, for being related or for lack of available electronic health record data, 74,331 samples with short-read WGS data and male sex assigned at birth were analyzed. Samples were classified as cases (n = 4,473) and controls (n = 69,858) based on the presence or absence of 'Malignant neoplasm of prostate' or 'Personal history of malignant neoplasm of prostate' in the AoU electronic health record data. Family history was determined based on AoU survey data as positive (responses 'Father,' 'Sibling' or 'Son' to the question 'Including yourself, who in your family has had prostate cancer?'; n = 3,034) or negative otherwise (no response or different response to survey question; n = 71,297). The model validity was evaluated within and across AoU-provided predicted genetic ancestries (African/African American, n = 16,733; American Admixed/Latino, n = 10,769; East Asian, n = 1,436; European, n = 43,917; Middle Eastern, n = 346; and South Asian, n = 1,130) (Supplementary Table 14).

Candidate variants and training for polygenic score

We considered variants previously identified from the following sources for potential inclusion in an updated polygenic score for the P-CARE model: 290 variants from a previous score, 613 variants identified as prostate cancer susceptibility loci in a multi-ancestry genome-wide association studies, 23 variants identified as susceptibility loci for benign elevation of PSA or benign prostatic hypertrophy, nine variants identified as prostate cancer susceptibility loci in men of African ancestry in a genome-wide meta-analysis and 128 variants identified as susceptibility loci for prostate cancer in a genome-wide multi-ancestry meta-analysis[33,35,36,61,62]. A machine-learning, least absolute shrinkage and selection operator (LASSO)-regularized Cox proportional hazards model approach was used in the MVP dataset to select the final variants for the polygenic score and estimate weights, using the R (v.4.4) glmnet package (v.4.1.8)[63–65]. To develop the polygenic score, age at any prostate cancer diagnosis in MVP was the time to event, as this gives the most statistical power; controls were censored at age of last follow-up. First, we identified pairs of variants with highly correlated genotype (defined as $r^2 > 0.95$) and used univariable Cox models to exclude the variant from each pair with weakest univariable association. Next, all remaining candidate variants were evaluated for inclusion in the new polygenic score using a Cox model with genotype allele counts of candidate variants and the first five FastPop principal components as predictor variables. Genetic principal components were estimated using 2,309 ancestry informative markers from FastPop[66]. Loadings for the first five principal components were estimated in the 1000 Genomes Project phase 3 dataset[67]. The final form of the LASSO model was estimated using the lambda value that minimized the mean cross-validated error[68].

We then used Cox proportional hazards models to evaluate the association of the new polygenic score with age at diagnosis of prostate cancer, age at diagnosis of nodal and/or distant metastatic prostate cancer, and age at prostate cancer death within the MVP dataset overall and in analyses stratified by continental population ancestry group, as in previous work[18,37,58,63,69–72]. Similarly, Cox models were used to evaluate the association between the new score and age at diagnosis of any prostate cancer, clinically significant prostate cancer and fatal prostate cancer (in the COSM dataset) in the PRACTICAL cohort[18].

P-CARE model development and validation

The resulting polygenic score was then carried forward for use in the development of an integrated clinical prediction model within MVP. We developed a Cox model for age at prostate cancer diagnosis as a function of the polygenic score, modeled as a continuous variable; family history of prostate cancer (Supplementary Table 1), modeled as a binary variable indicating presence or absence of at least one first-degree relative with prostate cancer; and population structure, modeled using the first two genetic principal components (PCs). Previous analyses showed that the first two PCs are sufficient to capture genetic variation for prostate cancer risk stratification compared to 5–10 PCs[37]. Individuals not meeting the end point of interest were censored at last follow-up. Training on metastatic disease did not give improved results, due to lower event rates.

The resulting P-CARE model was then validated internally within the MVP dataset and externally within the four PRACTICAL datasets. Where available, we evaluated the association of the P-CARE model with age of diagnosis of any prostate cancer, clinically significant prostate cancer, metastatic prostate cancer and fatal prostate cancer. As in our previous work[18,33,34,37,58,63,70–72], we estimated illustrative effect sizes using HRs and ten iterations of tenfold cross-validation, calculated to make the following comparisons: $HR_{80/20}$, men in the highest 20% versus lowest 20%; $HR_{95/50}$, men in the highest 5% versus those with median values; and $HR_{20/50}$, men in the lowest 20% versus those with median values. Within the MVP dataset, we generated cumulative incidence curves for each prostate cancer end point by P-CARE percentile groups, as in previous work[63,70]. We additionally generated cumulative incidence curves by P-CARE risk categories defined by risk of metastatic disease, given its morbidity and mortality and to counter the criticism that current prostate cancer screening approaches over-detect indolent disease[12–15]. The high-risk category was defined as an overall P-CARE HR > 1.5 for metastatic prostate cancer and the low-risk category was defined as HR < 0.75 (consistent with routine clinical prediction tools for other diseases, such as breast cancer, diabetes and cardiovascular disease[73–75]); all other risk values were considered average risk. The ages at which different P-CARE percentiles and P-CARE risk groups reached an equivalent cumulative risk of any and metastatic/fatal prostate cancer as that of the average risk man at 55 and 70 years old, respectively, were also determined. Because ProtecT systematically collected prostate biopsies, this dataset offered the opportunity to correlate PSA values with likelihood of clinically significant prostate cancer. Within the ProtecT dataset, we calculated the PPV of a PSA value ≥ 3 ng ml^{-1} for clinically significant prostate cancer on biopsy among participants in the top fifth (PPV_{95}) and top 20th P-CARE percentile (PPV_{80})[33,60].

Clinical laboratory assay development and validation

The P-CARE model was then carried forward to develop a clinical laboratory assay (Broad Clinical Labs) to enable precision prostate cancer screening informed by both the model and relevant rare variants, given their importance in prostate cancer risk.

Blended genome–exome assay. We constructed the assay on a BGE platform[76] modified to achieve deeper exome coverage. The assay achieves cost-efficiency for detecting rare and common variants by combining 2–3× WGS with 60–90× exome sequencing in a single

sequenced sample. The BGE platform has achieved >99% concordance with 30× genome sequencing data for both exome and genome short variants[76]. Short-variant calling was performed over the high coverage exome target regions using the Illumina DRAGEN Bio-IT platform v.4.2.7. Genotypes and dosage information over the whole genome were obtained from sequencing data through GLIMPSE2 imputation[77] using the gnomAD HGDP and 1000 Genomes Project callset[78]. CNVs were detected over the exome target regions using GATK-gCNV[79].

Analytic and clinical laboratory validation of polygenic score and P-CARE model. The analytic validity of the BGE platform for the polygenic score was assessed by comparing 60 clinical samples with previously identified variants; reference samples from Coriell Institute for Medical Research with curated reference variant datasets maintained by the National Institute of Standards and Technology; and samples with known SNVs, indels and CNVs from a combination of previous in-production clinical samples, previous eMERGE studies, previous CAP proficiency testing samples, Coriell samples and the Coriell Ancestral Panel. For each of these samples, representing six genetic ancestry groups (Admixed American, African, Non-Finnish European, East Asian, South-East Asian and Ashkenazi), we generated both BGE and WGS data and calculated the polygenic score and genetic PCs. Additional evidence of clinical validity for both the polygenic score and the P-CARE model was obtained using 74,331 samples from the AoU Research Program. Polygenic score and genetic PCs were calculated from the WGS genotypes provided by AoU. Individuals were classified as cases and controls based on the AoU electronic health record data. P-CARE values were calculated for each AoU participant using polygenic score, the first two genetic PCs, first-degree family history of prostate cancer and the MVP-derived coefficients. To determine the association between P-CARE and prostate cancer case status in AoU, we calculated odds ratios for an individual to be diagnosed with prostate cancer in the low and high P-CARE categories, relative to the average category, using logistic regression models controlling for age.

Rare variant selection, validation, and interpretation. We identified known prostate cancer-associated genes for which the NCCN has issued clinical management recommendations[20,80,81]. This gene list informed the filtering for an in silico gene panel for rare variant analysis. The ability of the BGE to identify pathogenic or likely pathogenic variants in these genes was evaluated by assessing the overall technical performance of 12 genes related to hereditary prostate cancer risk (*BRCA1*, *BRCA2*, *ATM*, *PALB2*, *CHEK2*, *HOXB13*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *TP53* and *EPCAM*) and identification of known variants from previous clinical testing (SNVs, small indels and CNVs) within these genes in 18 clinical samples. Technical performance of these genes was assessed by determining the percentage of undercovered bases within a panel gene. A base is considered covered if it satisfies the following: coverage >20×, base quality >20 and mapping quality >20. This coverage analysis was performed with two sample fraction thresholds: ≥80% and ≥20%. We determined the sensitivity for the detection of rare monogenic variants if the variant of interest was identified in the variant call file and would meet quality and prioritization metrics to be flagged for manual review by our tertiary analysis platform. Additionally, inter and intra run precision was assessed by running samples in triplicate across different runs and within the same run, respectively. We developed a workflow to classify, review and prioritize variants in a tertiary analysis platform (Fabric Genomics) before in-house clinical interpretation and reporting of pathogenic and likely pathogenic variants by a team of board-certified geneticists.

Clinical report development

After clinical laboratory validation of the P-CARE and rare variant pipelines, we developed a laboratory report package suitable for the clinical implementation of these results, consistent in format and content with other clinical genetic test reports and with our previous work[4,5,82]. As described in the Results, the report package consisted of separate laboratory reports for the P-CARE and rare variant results and a summary report synthesizing the result types and providing prostate cancer screening recommendations for the patient and provider.

Statistics and reproducibility

This study was designed to develop, validate and clinically implement a genomic risk model for prostate cancer screening using large, diverse biobank-linked cohorts. Statistical analyses were conducted using Cox proportional hazards models to evaluate associations between polygenic scores, family history, genetic PCs and prostate cancer outcomes. Model development included internal cross-validation and external replication in multiple cohorts, with effect sizes estimated using HRs and cumulative incidence curves.

Sample sizes were determined by the availability of eligible participants in the MVP, PRACTICAL Consortium and AoU Research Program datasets. No statistical method was used to predetermine sample size. No data were excluded from the analyses unless flagged for failing quality control criteria, being from related individuals or lacking available electronic health record data, as described below. Reproducibility was assessed through internal cross-validation (ten iterations of tenfold cross-validation) and external validation in independent cohorts. Analytic validity of the BGE platform was confirmed by comparison with reference samples and repeated testing across sequencing runs. All statistical analyses were performed using R (v.4.4) and relevant packages. The statistical analyses in this study primarily utilized Cox proportional hazards models and related approaches. Data distribution, including normality and equal variances, were formally tested and met model assumptions.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated from our analyses are included in the text, tables, figures and supplementary information. The genetic loci included in the polygenic score and their effect sizes are included in the Supplementary information. Source data for Figs. 2 and 3 and Extended Data Figs. 1 and 2 have been provided as Source Data files. All other data supporting the findings of this study are available from the corresponding author on reasonable request. It is not possible for the authors to share individual-level data from the MVP due to constraints stipulated in the informed consent. Anyone wishing to gain access to this data should inquire directly to MVP (MVPLOI@va.gov). Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome (PRACTICAL) Consortium data are available upon request to the Data Access Committee (http://practical.icr.ac.uk/blog). Data from the AoU Research Program are accessible through the Researcher Workbench to researchers with an approved Data Use and Registration Agreement. Source data are provided with this paper.

Code availability

The code used for analyses is available at https://github.com/precimed/MVP-PCa-PHS.

References

1. Khan, S. S. et al. Development and validation of the American Heart Association's PREVENT equations. *Circulation* **149**, 430–449 (2024).
2. Huntley, C. et al. Utility of polygenic risk scores in UK cancer screening: a modelling analysis. *Lancet Oncol.* **24**, 658–668 (2023).
3. Barkas, F. et al. Advancements in risk stratification and management strategies in primary cardiovascular prevention. *Atherosclerosis* **395**, 117579 (2024).

4. Hao, L. et al. Development of a clinical polygenic risk score assay and reporting workflow. *Nat. Med.* **28**, 1006–1013 (2022).
5. Lennon, N. J. et al. Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat. Med.* **30**, 480–487 (2024).
6. Roundtable on Translating Genomic-Based Research for Health; Board on Health Sciences Policy; Institute of Medicine. *Genomics-Enabled Learning Health Care Systems: Gathering and Using Genomic Information to Improve Patient Care and Research: Workshop Summary* (National Academies Press, 2015).
7. Mucci, L. A. et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA* **315**, 68–76 (2016).
8. Hall, R., Bancroft, E., Pashayan, N., Kote-Jarai, Z. & Eeles, R. A. Genetics of prostate cancer: a review of latest evidence. *J. Med. Genet.* **61**, 915–926 (2024).
9. Grossman, D. C. et al. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. *JAMA* **319**, 1901–1913 (2018).
10. Garraway, I. P. et al. Prostate cancer foundation screening guidelines for black men in the United States. *NEJM Evid.* **3**, EVIDoa2300289 (2024).
11. Jackson, S. D. et al. Screening asymptomatic men for prostate cancer: a comparison of international guidelines on prostate-specific antigen testing. *J. Med. Screen.* **29**, 268–271 (2022).
12. Loeb, S. et al. Overdiagnosis and overtreatment of prostate cancer. *Eur. Urol.* **65**, 1046–1055 (2014).
13. Ilic, D. et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *Brit. Med. J.* **362**, k3519 (2018).
14. Hugosson, J. et al. A 16-yr follow-up of the European randomized study of screening for prostate cancer. *Eur. Urol.* **76**, 43–51 (2019).
15. Paschen, U. et al. Assessment of prostate-specific antigen screening: an evidence-based report by the German Institute for Quality and Efficiency in Health Care. *BJU Int* **129**, 280–289 (2022).
16. Stone, B. V. et al. The association of county-level prostate-specific antigen screening with metastatic prostate cancer and prostate cancer mortality. *Eur. Urol. Oncol.* **7**, 563–569 (2024).
17. Iyer, H. S. et al. Access to prostate-specific antigen testing and mortality among men with prostate cancer. *JAMA Netw. Open* **7**, e2414582 (2024).
18. Pagadala, M. S. et al. Polygenic risk of any, metastatic, and fatal prostate cancer in the Million Veteran Program. *J. Natl Cancer Inst.* **115**, 190–199 (2023).
19. Martin, R. M. et al. Prostate-specific antigen screening and 15-year prostate cancer mortality: a secondary analysis of the CAP randomized clinical trial. *JAMA* **331**, 1460–1470 (2024).
20. National Comprehensive Cancer Network. NCCN Guidelines: Prostate Cancer Early Detection, version 2 (NCCN, 2024).
21. Wei, J. T. et al. Early detection of prostate cancer: AUA/SUO guideline part I: prostate cancer screening. *J. Urol.* **210**, 46–53 (2023).
22. Parker, C. et al. Prostate cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **31**, 1119–1134 (2020).
23. Mottet, N. et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer-2020 update. part 1: screening, diagnosis, and local treatment with curative intent. *Eur. Urol.* **79**, 243–262 (2021).
24. Huynh-Le, M.-P. et al. Age dependence of modern clinical risk groups for localized prostate cancer: a population-based study. *Cancer* **126**, 1691–1699 (2020).
25. Hou, K. et al. Calibrated prediction intervals for polygenic scores across diverse contexts. *Nat. Genet.* **56**, 1386–1396 (2024).
26. Moreno-Grau, S. et al. Polygenic risk score portability for common diseases across genetically diverse populations. *Hum. Genomics* **18**, 93 (2024).
27. Roth, J. A., Gulati, R., Gore, J. L., Cooperberg, M. R. & Etzioni, R. Economic analysis of prostate-specific antigen screening and selective treatment strategies. *JAMA Oncol.* **2**, 890–898 (2016).
28. Shoag, J. E., Nyame, Y. A., Gulati, R., Etzioni, R. & Hu, J. C. Reconsidering the trade-offs of prostate cancer screening. *N. Engl. J. Med.* **382**, 2465–2468 (2020).
29. Bergengren, O. et al. 2022 Update on prostate cancer epidemiology and risk factors-a systematic review. *Eur. Urol.* **84**, 191–206 (2023).
30. Das, H. & Rodriguez, R. Health care disparities in urologic oncology: a systematic review. *Urology* **136**, 9–18 (2020).
31. Riviere, P. et al. Survival of African American and non-Hispanic white men with prostate cancer in an equal-access health care system. *Cancer* **126**, 1683–1690 (2020).
32. Dess, R. T. et al. Association of black race with prostate cancer-specific and other-cause mortality. *JAMA Oncol.* **5**, 975–983 (2019).
33. Huynh-Le, M.-P. et al. Prostate cancer risk stratification improvement across multiple ancestries with new polygenic hazard score. *Prostate Cancer Prostatic Dis.* **25**, 755–761 (2022).
34. Karunamuni, R. A. et al. Performance of African-ancestry-specific polygenic hazard score varies according to local ancestry in 8q24. *Prostate Cancer Prostatic Dis.* **25**, 229–237 (2022).
35. 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).
36. Kachuri, L. et al. Genetically adjusted PSA levels for prostate cancer screening. *Nat. Med.* **29**, 1412–1423 (2023).
37. Huynh-Le, M.-P. et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. *Nat. Commun.* **12**, 1236 (2021).
38. Mavaddat, N. et al. Incorporating alternative polygenic risk scores into the BOADICEA breast cancer risk prediction model. *Cancer Epidemiol. Biomark.* **32**, 422–427 (2023).
39. Yang, X. et al. Prospective validation of the BOADICEA multifactorial breast cancer risk prediction model in a large prospective cohort study. *J. Med. Genet.* **59**, 1196–1205 (2022).
40. 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).
41. 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).
42. 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).
43. Callender, T. et al. Polygenic risk-tailored screening for prostate cancer: a benefit–harm and cost-effectiveness modelling study. *PLoS Med.* **16**, e1002998 (2019).
44. Callender, T., Emberton, M., Morris, S., Pharoah, P. D. P. & Pashayan, N. Benefit, harm, and cost-effectiveness associated with magnetic resonance imaging before biopsy in age-based and risk-stratified screening for prostate cancer. *JAMA Netw. Open* **4**, e2037657 (2021).
45. Dixon, P., Keeney, E., Taylor, J., Wordsworth, S. & Martin, R. Can polygenic risk scores contribute to cost-effective cancer screening? A systematic review. *Genet. Med.* **24**, 1604–1617 (2022).
46. Jiang, S. et al. Cost-effectiveness of population-wide genomic screening for Lynch syndrome and polygenic risk scores to inform colorectal cancer screening. *Genet. Med.* **27**, 101285 (2025).
47. Guzauskas, G. F. et al. Population genomic screening for three common hereditary conditions: a cost-effectiveness analysis. *Ann. Intern. Med.* **176**, 585–595 (2023).

48. Lacaze, P. et al. Combined population genomic screening for three high-risk conditions in Australia: a modelling study. *eClinicalMedicine* **66**, 102297 (2023).
49. Wongvibulsin, S., Wu, K. C. & Zeger, S. L. Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. *BMC Med. Res. Methodol.* **20**, 1 (2019).
50. Loef, B. et al. Using random forest to identify longitudinal predictors of health in a 30-year cohort study. *Sci. Rep.* **12**, 10372 (2022).
51. Gaziano, J. M. et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J. Clin. Epidemiol.* **70**, 214–223 (2016).
52. Bick, A. G. et al. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
53. Hunter-Zinck, H. et al. Genotyping array design and data quality control in the Million Veteran Program. *Am. J. Hum. Genet.* **106**, 535–548 (2020).
54. Pagadala, M. S. et al. Agent orange exposure and prostate cancer risk in the Million Veteran Program. *Acta Oncol. Stockh. Swed.* **63**, 373–378 (2024).
55. Alba, P. R. et al. Ascertainment of veterans with metastatic prostate cancer in electronic health records: demonstrating the case for natural language processing. *JCO Clin. Cancer Inform.* **5**, 1005–1014 (2021).
56. Amos, C. I. et al. The OncoArray Consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol. Biomark.* **26**, 126–135 (2017).
57. Eeles, R. A. et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.* **45**, 385–391. e1–2 (2013).
58. Huynh-Le, M.-P. et al. Common genetic and clinical risk factors: association with fatal prostate cancer in the Cohort of Swedish Men. *Prostate Cancer Prostatic Dis.* **24**, 845–851 (2021).
59. Discacciati, A. et al. Coffee consumption and risk of localized, advanced and fatal prostate cancer: a population-based prospective study. *Ann. Oncol.* **24**, 1912–1918 (2013).
60. Hamdy, F. C. et al. Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N. Engl. J. Med.* **388**, 1547–1558 (2023).
61. Gudmundsson, J. et al. Genome-wide associations for benign prostatic hyperplasia reveal a genetic correlation with serum levels of PSA. *Nat. Commun.* **9**, 4568 (2018).
62. Chen, F. et al. Evidence of novel susceptibility variants for prostate cancer and a multiancestry polygenic risk score associated with aggressive disease in men of African ancestry. *Eur. Urol.* **84**, 13–21 (2023).
63. Karunamuni, R. A. et al. Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. *Prostate Cancer Prostatic Dis.* **24**, 532–541 (2021).
64. Tibshirani, R. Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).
65. Tibshirani, R. The LASSO method for variable selection in the Cox model. *Stat. Med.* **16**, 385–395 (1997).
66. Li, Y. et al. FastPop: a rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC Bioinform.* **17**, 122 (2016).
67. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
68. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
69. Wendt, F. R. et al. Modeling the longitudinal changes of ancestry diversity in the Million Veteran Program. *Hum. Genomics* **17**, 46 (2023).
70. Huynh-Le, M.-P. et al. A genetic risk score to personalize prostate cancer screening, applied to population data. *Cancer Epidemiol. Biomarkers Prev.* **29**, 1731–1738 (2020).
71. Seibert, T. M. et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *Brit. Med. J.* **360**, j5757 (2018).
72. Karunamuni, R. A. et al. African-specific improvement of a polygenic hazard score for age at diagnosis of prostate cancer. *Int. J. Cancer* **148**, 99–105 (2021).
73. Yeh, H. C., Duncan, B. B., Schmidt, M. I., Wang, N. Y. & Brancati, F. L. Smoking, smoking cessation, and risk for type 2 diabetes mellitus: a cohort study. *Ann. Intern. Med.* **152**, 10–17 (2010).
74. Brentnall, A. R., Cuzick, J., Buist, D. S. M. & Bowles, E. J. A. Long-term accuracy of breast cancer risk assessment combining classic risk factors and breast density. *JAMA Oncol.* **4**, e180174 (2018).
75. Wang, T. J. et al. Plasma natriuretic peptide levels and the risk of cardiovascular events and death. *N. Engl. J. Med.* **350**, 655–663 (2004).
76. DeFelice, M. et al. Blended genome exome (BGE) as a cost efficient alternative to deep whole genomes or arrays. Preprint at *bioRxiv* https://doi.org/10.1101/2024.04.03.587209 (2024).
77. Rubinacci, S., Hofmeister, R. J., Sousa da Mota, B. & Delaneau, O. Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat. Genet.* **55**, 1088–1090 (2023).
78. Tiao, G. & Goodrich, J. gnomAD v3.1 new content, methods, annotations, and data availability. https://gnomad.broadinstitute.org/news/2020-10-gnomad-v3-1-new-content-methods-annotations-and-data-availability/#the-gnomad-hgdp-and-1000-genomes-callset (2023).
79. Babadi, M. et al. GATK-gCNV enables the discovery of rare copy number variants from exome sequencing data. *Nat. Genet.* **55**, 1589–1597 (2023).
80. National Comprehensive Cancer Network. NCCN Guidelines: Genetic/familial High-risk Assessment: Colorectal, version 2 (NCCN, 2023).
81. National Comprehensive Cancer Network. NCCN Guidelines: Genetic/familial High-risk Assessment: Breast, Ovarian, And Pancreatic, version 3 (NCCN, 2024).
82. Farmer, G. D., Gray, H., Chandratillake, G., Raymond, F. L. & Freeman, A. L. J. Recommendations for designing genetic test reports to be understood by patients and non-specialists. *Eur. J. Hum. Genet.* **28**, 885–895 (2020).

Acknowledgements

This research was funded by the US Department of Veterans Affairs Office of Research and Development (I01 CX002635 to J.L.V. and I01 CX001727 to R.L.H.), which played no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript. It was supported using resources and facilities of the Department of VA Informatics and Computing Infrastructure (VINCI) ORD 24-VINCI-01, including writing support from K. Pridgen, under the research priority to Put VA Data to Work for Veterans (VA ORD 24-D4V). Funding for salaries includes Department of VA (VISN22 Veterans Center of Excellence for Stress and Mental Health to R.L.H.), VA Office of Research and Development (1I01CX002709 and 1I01CX002622 to K.N.M.), National Institutes of Health (NIH) (R01AG050595 to R.L.H. and K08CA215312 to K.N.M.), the Department of Defense (DOD/CDMRP PC220521 to T.M.S.), the Prostate Cancer Foundation (23CHAL12 to T.M.S., 20YOUN02 to K.N.M., 22CHAL02 to I.P.G., B.S.R. and K.N.M.), the Burroughs Wellcome Foundation (no. 1017184 to K.N.M.), Basser Center for BRCA (K.N.M.). The authors thank the MVP staff, researchers, and volunteers, who have contributed to MVP, and especially who previously served their country in the military and now generously agreed to enroll in the study (see mvp.va.gov for more

information). The underlying work was based on data from the MVP, Office of Research and Development, Veterans Health Administration, and was supported by the Veterans Administration MVP award no. 000. This publication does not represent the views of the Department of VA or the US Government.

CAP: The CAP trial was funded by grants C11043/A4286, C18281/A8145, C18281/A11326, C18281/A15064 and C18281/A24432 from Cancer Research UK. The UK Department of Health, National Institute of Health Research (NIHR) provided partial funding.

ProtecT: The ProtecT trial was funded by project grants 96/20/06 and 96/20/99 from the UK National Institute for Health Research, Health Technology Assessment Programme. R.M.M. is an NIHR Senior Investigator (NIHR202411). R.M.M. is supported by a Cancer Research UK 25 (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme). R.M.M. is also supported by the NIHR Bristol Biomedical Research Centre which is funded by the NIHR (BRC-1215-20011) and is a partnership between University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. Department of Health and Social Care disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. A.V. is supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (PI22/00589, INT24/00023, DTS24/00083 and PI25/00744); and by the AECC (PRYES211091VEGA). A.S.K. is supported by NIH, grant/award numbers U01 - U01CA268810.

CRUK and PRACTICAL Consortium: This work was supported by the Canadian Institutes of Health Research (CIHR), European Commission's Seventh Framework Programme grant agreement no. 223175 (HEALTH-F2-2009-223175), Cancer Research UK grants (C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692 and C16913/A6135) and The NIH Cancer Post-Cancer GWAS initiative grant: no. 1 U19 CA 148537-01 (the GAME-ON initiative). We also thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The National Cancer Research Network UK and The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research, The Royal Marsden NHS Foundation Trust, and Manchester NIHR Biomedical Research Centre. The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, 396414, 450104, 504700, 504702, 504715, 623204, 940394 and 614296), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers and Tattersall's. E.A.O., D.M.K. and E.M.K. acknowledge the Intramural Program of the National Human Genome Research Institute for their support. Genotyping of the OncoArray was funded by the US NIH (U19 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research under contract number HHSN2682012000008I). Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). Research reported in this publication also received support from the National Cancer Institute of the NIH under award numbers U10 CA37429 (C.D. Blanke) and UM1 CA182883 (C.M. Tangen/I.M. Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement no. 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692 and C8197/A16565), the NIH (CA128978) and Post-Cancer GWAS initiative (1U19 CA 148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the CIHR for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund. PROtEuS received funding from the Canadian Cancer Society, the Cancer Research Society, the Fonds de Recherche du Québec–Santé, the Ministère du Développement Économique, de l'Innovation et de l'Exportation du Québec and the Canada Research Chairs Program

BPC3: The BPC3 was supported by the US NIH, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 to E.R. and U01-CA98758 to B.E.H., and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics).
CAPS: The CAPS GWAS study was supported by the Cancer Risk Prediction Center (CRisP; www.crispcenter.org), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, (grant no K2010-70X-20430-04-3), the Swedish Cancer Foundation (grant no 09-0677), the Hedlund Foundation, the Soederberg Foundation, the Enqvist Foundation, ALF funds from the Stockholm County Council. Stiftelsen Johanna Hagstrand och Sigfrid Linner's Minne, Karlsson's Fund for urological and surgical research.

PEGASUS: PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH.

Author contributions

J.L.V., A.M. Dornisch and T.M.S. conceived and designed the study, oversaw data analysis, interpreted results, and drafted the manuscript. R.K. contributed to study design, supervised analytic workflows and provided domain expertise in polygenic risk modeling. M.G., C.J.K., N.J.L., S.E.H., K.A.L., K.L., E.M., C.J.P. and D.M.T. coordinated sequencing, bioinformatics and data curation. C.A.B., M.E.D. and D.R. assisted with data interpretation, visualization and manuscript preparation. R.L.H., I.P.G., K.M.L., J.A.L., K.N.M., B.S. Rose, C.C.T., A.S.K. and G.J.X. provided expertise in prostate cancer phenotyping and critical review of study design and interpretation. J.L.D., F.H., R.M.M., D.E.N., E.L.T., O.A.A., A.M. Dale, I.G.M., A.A., J.B., J.C., O.C., C.C., R.A.E., J.H.F., E.M.G., H.G., R.J.H., J.L., Y.J.L., R.J.M., C.M., L.A.M., L.M., S.L.N., S.F.N., M.E.P., J.Y.P., G.P., A.P., A.R., B.S. Rosenstein, J.S., K.D.S., P.A.T., R.C.T., A.V., C.M.L.W., F.W. and W.Z. contributed data from consortium studies, performed genotyping and quality control and reviewed the manuscript for intellectual content. J.L.V. served as principal investigator and corresponding author. All authors reviewed the results and approved the final manuscript.

Competing interests

N.L. has received speaking honoraria from Illumina and is an advisory board member for FYR Diagnostics and Everygene; N.L. has also received research collaborative funding (for work unrelated to this publication) from Illumina and PacBio. J.A.L., K.M.L. and C.T.C. report grants from Alnylam Pharmaceuticals, Astellas Pharma, AstraZeneca Pharmaceuticals, Biodesix, Celgene Corporation, Cerner Enviza, GSK, IQVIA, Janssen Pharmaceuticals, Novartis International and the Parexel International Corporation through the University of Utah or Western Institute for Veteran Research outside the submitted work. A.S.K. reports funding (for work unrelated to this publication) from Janssen, Pfizer, Profound, Bristol Myers Squibb and Merck. S.L.D. reports grants from AstraZeneca Pharmaceuticals, Biodesix, Myriad Genetic Laboratories, Parexel, Moderna, GlaxoSmithKline, Cerner Enviza, Janssen Research & Development, Celgene, Novartis Pharmaceuticals, IQVIA, Astellas Pharma and Alnylam Pharmaceuticals. R.A.E. reports speaking honoraria from GU-ASCO, Janssen, University of Chicago and the Dana Farber Cancer Institute, educational honorarium from Bayer and Ipsen, being a member of external expert committee

to AstraZeneca UK and Member of Active Surveillance Movember Committee and is a member of the Scientific Advisory Board of Our Future Health; she additionally 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. L.A.M. reports research funding from AstraZeneca to Harvard University; she holds equity in Convergent Therapeutics. T.M.S. reports honoraria from Varian Medical Systems, WebMD, GE Healthcare and Janssen; he has an equity interest in CorTechs Labs and serves on its Scientific Advisory Board; he receives research funding from GE Healthcare through the University of California, San Diego. These companies might potentially benefit from the research results. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict-of-interest policies. All other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43018-025-01103-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43018-025-01103-0.

Correspondence and requests for materials should be addressed to Jason L. Vassy or Tyler M. Seibert.

Peer review information *Nature Cancer* thanks Irajizad Ehsan, Fumitaka Koga, Shintaro Narita and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Jason L. Vassy[1,2,3,178] ✉, Anna M. Dornisch[4,178], Roshan Karunamuni[4,5], Michael Gatzen[6], Christopher J. Kachulis[6,7], Niall J. Lennon[6,7], Charles A. Brunette[1,2], Morgan E. Danowski[1], Richard L. Hauger[8,9], Isla P. Garraway[10,11], Adam S. Kibel[12], Kyung M. Lee[13], Julie A. Lynch[13], Kara N. Maxwell[14,15], Dmitry Ratner[1], Brent S. Rose[4,5], Craig C. Teerlink[16,17], George J. Xu[4,5], Sean E. Hofherr[6], Katherine A. Lafferty[6,7], Katie Larkin[6,7], Edyta Malolepsza[6,7], Candace J. Patterson[6,7], Diana M. Toledo[6,7], Jenny L. Donovan[18], Freddie C. Hamdy[19], Richard M. Martin[18,20], David E. Neal[19,21,22], Emma L. Turner[18], Ole A. Andreassen[23], Anders M. Dale[24,25,26,27], Ian G. Mills[19], Aswin Abraham[28], Jyotsna Batra[29,30,31,32], Judith Clements[30,31,32], Olivier Cussenot[33,34], Cezary Cybulski[35], Rosalind A. Eeles[36,37], Jay H. Fowke[38], Eli Marie Grindedal[39], Henrik Grönberg[40], Robert J. Hamilton[41,42], Jasmine Lim[43], Yong-Jie Lu[44], Robert J. MacInnis[45,46], Christiane Maier[47,48], Lorelei A. Mucci[49], Luc Multigner[50], Susan L. Neuhausen[51], Sune F. Nielsen[52,53], Marie-Élise Parent[54,55], Jong Y. Park[56], Gyorgy Petrovics[57,58], Anna Plym[40,49], Azad Razack[43], Barry S. Rosenstein[59], Johanna Schleutker[60,61], Karina Dalsgaard Sørensen[62,63], Paul A. Townsend[64,65], Ruth C. Travis[66], Ana Vega[67,68,69], Catharine M. L. West[70], Fredrik Wiklund[40], Wei Zheng[71], Profile Steering Committee*, IMPACT Study Steering Committee and Collaborators*, PRACTICAL Consortium*, VA Million Veteran Program* & Tyler M. Seibert[4,5,25,72,73] ✉

[1]VA Boston Healthcare System, Boston, MA, USA. [2]Department of Medicine, Harvard Medical School, Boston, MA, USA. [3]Division of General Internal Medicine, Mass General Brigham, Boston, MA, USA. [4]Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA, USA. [5]Research Service, VA San Diego Healthcare System, San Diego, CA, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Broad Clinical Labs, Burlington, MA, USA. [8]Center of Excellence for Stress and Mental Health (CESAMH), VA San Diego Healthcare System, San Diego, CA, USA. [9]Center for Behavior Genetics of Aging, University of California San Diego, La Jolla, CA, USA. [10]Division of Urology, VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA. [11]Department of Urology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA. [12]Urology Division, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [13]VA Informatics and Computing Infrastructure, VA Salt Lake City Healthcare System, Salt Lake City, UT, USA. [14]Corporal Michael Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA. [15]Department of Medicine - Hematology/Oncology, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA. [16]VA Salt Lake City Healthcare System, Salt Lake City, UT, USA. [17]Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA. [18]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [19]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. [20]NIHR Biomedical Research Centre at the University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, UK. [21]University of Cambridge, Department of Oncology, Cambridge, UK. [22]Cancer Research UK, Cambridge Research Institute, Cambridge, UK. [23]Centre for Precision Psychiatry and KG Jebsen Centre for Neurodevelopment, Oslo University Hospital and University of Oslo, Oslo, Norway. [24]NORMENT, KG Jebsen Centre, Oslo University Hospital and University of Oslo, Oslo, Norway. [25]Department of Radiology, University of California San Diego, La Jolla, CA, USA. [26]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. [27]Halicioglu Data Science Institute, University of California San Diego, La Jolla, CA, USA. [28]Department of Oncology, Cross Cancer Institute, University of Alberta, Alberta, Canada. [29]School of Biomedicine, Faculty of Health and Medical Sciences, Bond University, Gold Coast, Queensland, Australia. [30]Centre for Genomics and Personalised Health, School of Biomedical Sciences, Faculty of Health, Queensland University of Technology (QUT), Brisbane, Queensland, Australia. [31]Australian Prostate Cancer Bioresource (APCB), QUT, Brisbane, Queensland, Australia. [32]Translational Research Institute, QUT, Brisbane, Queensland, Australia. [33]Sorbonne Université, Paris, France. [34]Tenon Hospital, Paris, France. [35]International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland. [36]The Institute of Cancer Research, London, UK. [37]Royal Marsden NHS Foundation Trust, London, UK. [38]Division of Epidemiology, Department of Preventive

Medicine, University of Tennessee Health Science Center, Memphis, TN, USA. [39]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [40]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden. [41]Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [42]Dept. of Surgery (Urology), University of Toronto, Toronto, Ontario, Canada. [43]Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [44]Centre for Cancer Biomarker and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, London, UK. [45]Cancer Epidemiology Division, Cancer Council Victoria, East Melbourne, Victoria, Australia. [46]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia. [47]Zentrum für Humangenetik Tuebingen, Tuebingen, Germany. [48]Institut für Humangenetik, University of Ulm, Ulm, Germany. [49]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [50]Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), Rennes, France. [51]Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA. [52]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark. [53]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [54]Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Institut national de la recherche scientifique, Laval, Quebec, Canada. [55]Department of Social and Preventive Medicine, School of Public Health, University of Montreal, Montreal, Quebec, Canada. [56]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA. [57]Uniformed Services University, Bethesda, MD, USA. [58]Center for Prostate Disease Research, Bethesda, MD, USA. [59]Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [60]Institute of Biomedicine, University of Turku, Turku, Finland. [61]Department of Medical Genetics, Genomics, Laboratory Division, Turku University Hospital, Turku, Finland. [62]Department of Molecular Medicine, Aarhus University Hospital, Aarhus N, Denmark. [63]Department of Clinical Medicine, Aarhus University, Aarhus N, Denmark. [64]Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, Manchester, UK. [65]The University of Surrey, Guildford, Surrey, UK. [66]Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [67]Fundación Pública Galega Medicina Xenómica, Santiago de Compostela, Spain. [68]Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela, Spain. [69]Centro de Investigación en Red de Enfermedades Raras (CIBERER), Madrid, Spain. [70]Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, UK. [71]Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. [72]Department of Bioengineering, University of California San Diego, La Jolla, CA, USA. [73]Department of Urology, University of California San Diego, La Jolla, CA, USA. [178]These authors contributed equally: Jason L. Vassy, Anna M. Dornisch. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: jvassy@bwh.harvard.edu; tseibert@health.ucsd.edu

Profile Steering Committee

Rosalind A. Eeles[74], Elizabeth K. Bancroft[74,75], Eva McGrowder[74], Zsofia Kote-Jarai[74], Mark N. Brook[74], Elizabeth C. Page[74], Jana McHugh[74], Holly Ni Raghallaigh[74], Denzil James[75], Pardeep Kumar[75], Steve Hazell[75], Aslam Sohaib[75], Alexander Dias[74], Christos Mikropoulos[74], Netty Kinsella[75] & Declan Cahill[75]

[74]Oncogenetics Team, Genetics and Epidemiology, The Institute of Cancer Research, London, UK. [75]The Royal Marsden NHS Foundation Trust, London, UK.

IMPACT Study Steering Committee and Collaborators

Rosalind A. Eeles[74], Elizabeth K. Bancroft[74,76], Elizabeth C. Page[74], Mark N. Brook[74], Zsofia Kote-Jarai[74], D. Gareth Evans[76], Geoffrey J. Lindeman[77,78,79], Paul James[78], Lucy Side[80], Karina Rønlund[81], Brian T. Helfand[82], Cezary Cybulski[83], Kai-Ren Ong[84], Monica Salinas[85], Jackie Cook[86], Kara Maxwell[87], Manuel R. Teixeira[88], Rosemarie Davidson[89], Annelie Liljegren[90], Ashraf Azzabi[91], Marc Tischkowitz[92], Ruth Cleaver[93], Julian Barwell[94], Neil K. Aaronson[95], Audrey Ardern-Jones[96], Chris H. Bangma[97], Elena Castro[98], David Dearnaley[99], Diana M. Eccles[80,100], Alison Falconer[101], Henrik Gronberg[102], Freddie C. Hamdy[19,103], Vincent Khoo[79,96,99,104], Hans Lilja[105,106], Jan Lubinski[83], Christos Mikropoulos[107], Anita Mitra[108], Judith Offman[109], Gad Rennert[110] & Mohnish Suri[111]

[76]Genomic medicine, Division of Evolution infection and genomic sciences, University of Manchester, Manchester, UK. [77]Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre and The Royal Melbourne Hospital, Melbourne, Victoria, Australia. [78]Cancer Biology and Stem Cells Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [79]Department of Medicine, The University of Melbourne, Parkville, Victoria, Australia. [80]Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK. [81]Department of Clinical Genetics, University Hospital of Southern Denmark, Vejle Hospital, Vejle, Denmark. [82]John and Carol Walter Center for Urological Health, Division of Urology, Endeavor Health NorthShore University HealthSystem, Evanston, IL, USA. [83]Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University in Szczecin, Szczecin, Poland. [84]Department of Clinical Genetics, Birmingham Women's Hospital, Birmingham, UK. [85]Hereditary Cancer Program, Catalan Institute of Oncology, ICO-IDIBELL (Bellvitge Biomedical Research Institute), 08908 Barcelona, CIBERONC, Madrid, Spain. [86]Sheffield Clinical Genetics Service, Sheffield Children's NHS Foundation Trust, Sheffield, UK. [87]Basser Research Center, University of Pennsylvania, Philadelphia, PA, USA. [88]Department of Laboratory Genetics and IPO Porto Research Center, Portuguese Oncology Institute (IPO Porto), Porto, Portugal. [89]West of Scotland Genetic Service, Queen Elizabeth University Hospital, Glasgow, UK. [90]Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden. [91]Northern Genetics Service, Newcastle upon Tyne Hospitals, Newcastle upon Tyne, UK. [92]East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Trust, Cambridge, UK. [93]Department of Clinical Genetics, Royal Devon and Exeter Hospital, Exeter, UK. [94]University of Leicester, Leicester, UK. [95]Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands. [96]Cancer Genetics Unit & Academic Urology Unit, The Royal Marsden NHS Foundation Trust, London, UK. [97]Department of urology, Erasmus University Medical Center, Rotterdam, The Netherlands. [98]Hospital Universitario 12 de Octubre, Madrid, Spain. [99]Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK. [100]Cancer Sciences, The University of Southampton, Southampton, UK. [101]Imperial College Healthcare NHS Trust, London, UK. [102]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [103]Churchill Hospital, Oxford, UK. [104]St George's Hospital, Tooting, London, UK. [105]Department of Translational Medicine, Lund University, Malmö, Sweden. [106]Departments of Pathology and Laboratory Medicine, Surgery and Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [107]Royal Surrey Hospital, Guildford, UK. [108]University

College London Hospitals NHS Foundation Trust, London, UK. [109]Centre for Cancer Screening, Prevention and Early Diagnosis, Wolfson Institute of Population Health, Queen Mary University of London, Charterhouse Square, London, UK. [110]Technion-Israel Institute of Technology, Haifa, Israel. [111]Nottingham City Hospital, Nottingham, UK.

PRACTICAL Consortium

Rosalind A. Eeles[36,37], Christopher A. Haiman[112], Zsofia Kote-Jarai[36], Fredrick R. Schumacher[113,114], Sara Benlloch[36,115], Ali Amin Al Olama[115,116], Kenneth R. Muir[117], Sonja I. Berndt[118], David V. Conti[112], Fredrik Wiklund[102], Stephen Chanock[118], Ying Wang[119], Catherine M. Tangen[120], Jyotsna Batra[29,30,31,32], Judith Clements[30,31,32], Henrik Grönberg[102], Nora Pashayan[121,122], Johanna Schleutker[60,61], Demetrius Albanes[118], Stephanie J. Weinstein[118], Alicja Wolk[123], Catharine M. L. West[70], Lorelei A. Mucci[49], Géraldine Cancel-Tassin[124,125], Stella Koutros[118], Karina Dalsgaard Sørensen[62,63], Eli Marie Grindedal[39], David E. Neal[126,127,128], Freddie C. Hamdy[19,129], Jenny L. Donovan[18], Ruth C. Travis[66], Robert J. Hamilton[41,42], Sue Ann Ingles[130], Barry S. Rosenstein[131], Yong-Jie Lu[132], Graham G. Giles[46,133,134], Robert J. MacInnis[46,133], Adam S. Kibel[135], Ana Vega[67,68,69], Manolis Kogevinas[136,137,138,139], Kathryn L. Penney[140], Jong Y. Park[56], Janet L. Stanford[141,142], Cezary Cybulski[35], Børge G. Nordestgaard[53,143], Sune F. Nielsen[53,143], Hermann Brenner[144,145], Christiane Maier[146], Jeri Kim[147], Esther M. John[148], Manuel R. Teixeira[149,150,151], Susan L. Neuhausen[51], Kim De Ruyck[152], Azad Razack[43], Lisa F. Newcomb[141,153], Davor Lessel[154], Radka Kaneva[155], Nawaid Usmani[156,157], Frank Claessens[158], Paul A. Townsend[65,159], Jose Esteban Castelao[160], Monique J. Roobol[161], Florence Menegaux[162], Kay-Tee Khaw[163], Lisa Cannon-Albright[17,164], Hardev Pandha[65], Stephen N. Thibodeau[165], David J. Hunter[166], Peter Kraft[167], William J. Blot[71,168] & Elio Riboli[169]

[112]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA. [113]Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA. [114]Seidman Cancer Center, University Hospitals, Cleveland, OH, USA. [115]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK. [116]University of Cambridge, Department of Clinical Neurosciences, Stroke Research Group, R3, Box 83, Cambridge Biomedical Campus, Cambridge, UK. [117]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Oxford Road, Manchester, UK. [118]Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA. [119]Department of Population Science, American Cancer Society, 250 Williams Street, Atlanta, GA, USA. [120]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [121]Department of Public Health & Primary Care, Strangeways Research Laboratory, Cambridge, UK. [122]Department of Applied Health Research, University College London, London, UK. [123]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [124]CeRePP, Tenon Hospital, Paris, France. [125]Sorbonne Universite, GRC n°5, AP-HP, Tenon Hospital, Paris, France. [126]Nuffield Department of Surgical Sciences, University of Oxford, Room 6603, Level 6, John Radcliffe Hospital, Headington, UK. [127]University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Cambridge, UK. [128]Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK. [129]Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK. [130]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA. [131]Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Box 1236, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [132]Centre for Cancer Biomarker and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London, UK. [133]Cancer Epidemiology Division, Cancer Council Victoria, 200 Victoria Parade, East Melbourne, Victoria, Australia. [134]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia. [135]Division of Urologic Surgery, Brigham and Womens Hospital, Boston, MA, USA. [136]ISGlobal, Barcelona, Spain. [137]IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain. [138]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [139]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [140]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA, USA. [141]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [142]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA. [143]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [144]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [145]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [146]Humangenetik Tuebingen, Tuebingen, Germany. [147]The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, TX, USA. [148]Departments of Epidemiology & Population Health and of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. [149]Department of Laboratory Genetics, Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal. [150]Cancer Genetics Group, IPO Porto Research Center (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal. [151]School of Medicine and Biomedical Sciences (ICBAS), University of Porto, Porto, Portugal. [152]Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Gent, Belgium. [153]Department of Urology, University of Washington, Seattle, WA, USA. [154]Institute of Human Genetics, University of Regensburg, and Institute of Clinical Human Genetics, University Hospital Regensburg, Regensburg, Germany. [155]Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, Bulgaria. [156]Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada. [157]Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada. [158]Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium. [159]Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, Univeristy of Manchester, Manchester, UK. [160]Genetic Oncology Unit, CHUVI Hospital, Complexo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), Vigo (Pontevedra), Spain. [161]Department of Urology, Erasmus University Medical Center, Cancer Institute, Rotterdam, The Netherlands. [162]"Exposome and Heredity", CESP (UMR 1018), Faculté de Médecine, Université Paris-Saclay, Inserm, Gustave Roussy, Villejuif, France. [163]Clinical Gerontology Unit, University of Cambridge, Cambridge, UK. [164]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA. [165]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [166]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [167]Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology,

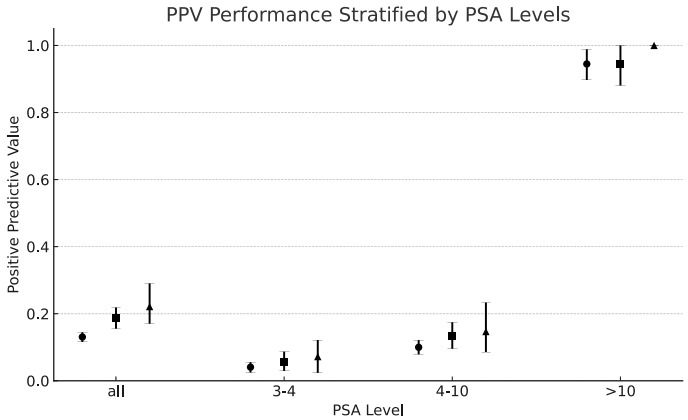
Harvard School of Public Health, Boston, MA, USA. [168]International Epidemiology Institute, Rockville, MD, USA. [169]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK.

VA Million Veteran Program

Sumitra Muralidhar[170], Jennifer Moser[170], Jennifer E. Deen[170], Philip S. Tsao[171], J. Michael Gaziano[1], Adriana Hung[172], Dave Oslin[173], Deepak Voora[174], Jessica V. Brewer[1], Mary T. Brophy[1], Kelly Cho[1], Lori Churby[171], Jacob T. Kean[175], Saiju Pyarajan[1], Robert Ringer[176], Luis E. Selva[1], Shahpoor Alex Shayan[1], Brady Stephens[177] & Stacey B. Whitbourne[1]

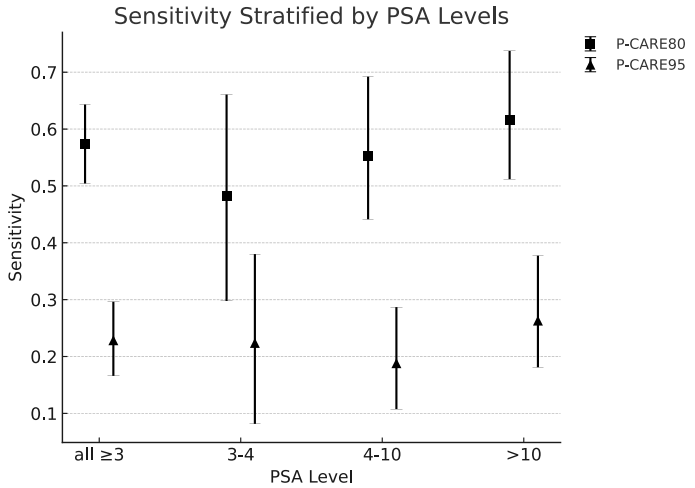
[170]US Department of Veterans Affairs, Washington, DC, USA. [171]VA Palo Alto Health Care System, Palo Alto, CA, USA. [172]VA Tennessee Valley Healthcare System, Nashville, TN, USA. [173]Philadelphia VA Medical Center, Philadelphia, PA, USA. [174]Durham VA Medical Center, Durham, NC, USA. [175]VA Salt Lake City Health Care System, Salt Lake City, UT, UK. [176]New Mexico VA Health Care System, Albuquerque, NM, USA. [177]Canandaigua VA Medical Center, Canandaigua, NY, USA.

https://doi.org/10.1038/s43018-025-01103-0



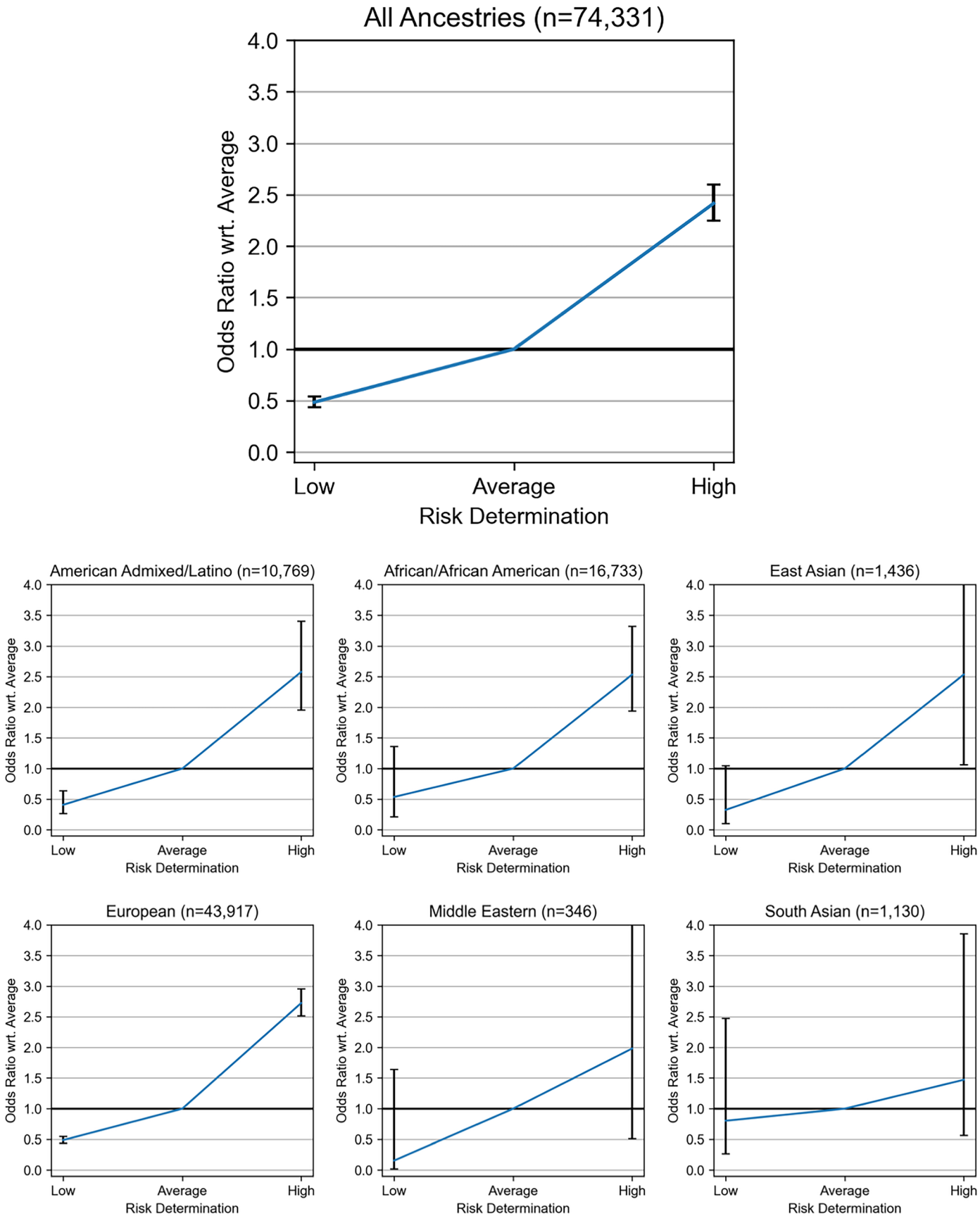
Extended Data Fig. 1 | Positive predictive value of PSA in ProtecT by P-CARE values stratified by PSA values. Illustrated are mean PPV (95% CI) for various PSA levels (all (≥ 3 ng/mL), 3–4 ng/mL, 4-10 ng/mL, and >10 ng/mL) for clinically significant prostate cancer among three groups of men in the ProtecT study (n = 6,411): all men (regardless of P-CARE value), men in the top 20% of P-CARE values ($P\text{-}CARE_{80}$), and men in the top 5% of P-CARE values ($P\text{-}CARE_{95}$). Abbreviations: CI, confidence interval; P-CARE, Prostate CA Risk and Evaluation; PPV, positive predictive value; ProtecT, Prostate Testing for Cancer and Treatment; PSA, prostate-specific antigen.

https://doi.org/10.1038/s43018-025-01103-0



Extended Data Fig. 2 | Percentage of true positive cases in ProtecT across P-CARE categories. Illustrated are the percentage of true positive cases for various PSA levels among 6,411 men in the ProtecT Study (all values ≥ 3 ng/mL, 3-4 ng/mL, 4-10 ng/mL, and >10 ng/mL) for clinically significant prostate cancer stratified by P-CARE risk percentiles, with comparisons between the top 5% ($P\text{-}CARE_{95}$) and top 20% ($P\text{-}CARE_{80}$). Error bars represent 95% confidence intervals.

https://doi.org/10.1038/s43018-025-01103-0



Extended Data Fig. 3 | Odds of prostate cancer in All of Us Research Program by P-CARE category. Shown are the odds ratios for an individual to be diagnosed with prostate cancer in the low and high P-CARE categories, relative to the average P-CARE category, derived from logistic regression models controlling for age. Error bars correspond to the 95% confidence intervals, and the shown ancestries are the predictions provided by All of Us. Abbreviations: P-CARE, Prostate CAncer Risk and Evaluation.

nature portfolio

Corresponding author(s): Jason L. Vassy and Tyler M. Seibert

Last updated by author(s): 10/20/2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used.
Data analysis	The code used for analyses is available at https://github.com/precimed/MVP-PCa-PHS.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data generated from our analyses are included in the manuscript text, tables, figures, and supplement. The genetic loci included in the polygenic score and their effect sizes are included in the Supplement. It is not possible for the authors to directly share individual-level data from the Million Veteran Program (MVP) due to constraints stipulated in the informed consent. Anyone wishing to gain access to this data should inquire directly to MVP at MVPLOI@va.gov. Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) Consortium data are available upon request to the Data Access

Committee (http://practical.icr.ac.uk/blog). Data from the All of Us Research Program are accessible through the Researcher Workbench to researchers with an approved Data Use and Registration Agreement.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	For this study about prostate cancer, we used samples from participants identified as "male" based on sex chromosome data. We do not have information about the gender identity of most participants.
Reporting on race, ethnicity, or other socially relevant groupings	We have been careful to distinguish the socially defined constructs of race and ethnicity groups (identified through self-report in these samples) and genetically similar populations (identified through genetic analysis).
Population characteristics	Participants were all adults of male sex, aged 18 and older, without prostate cancer at baseline
Recruitment	N/A - These cohorts are already established cohorts and there was no new participant recruitment.
Ethics oversight	The VA Central IRB approved this study (IRBNet 1735869 and 1735136).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available data meeting inclusion criteria were used.
Data exclusions	Genotype data were excluded if they did not meet standard QC criteria or were from related samples.
Replication	Analyses from the Million Veteran Program were replicated in PRACTICAL datasets and in the All of Us Research Program.
Randomization	Randomization is not relevant, as there was no group allocation.
Blinding	Blinding is not relevant, as there was no group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*