medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Evaluating a polygenic hazard score to predict risk of developing metastatic or fatal prostate cancer in the multi-ancestry Million Veteran Program cohort

Meghana S. Pagadala, BS[1,2,3]
Julie Lynch, PhD, RN, MBA[4,5]
Roshan Karunamuni, PhD[6]
Patrick R. Alba, MS[4,5]
Kyung Min Lee, PhD[4,5]
Fatai Y. Agiri, BS[4]
Tori Anglin, MHA[4,5]
Hannah Carter, PhD[7]
Guneet Kaur Jasuja, MPH, PhD[9,10]
Rishi Deka, PhD[1,6,8]
Brent S. Rose, MD[1,6,11]
Matthew S. Panizzon, PhD[1,12]
Richard L. Hauger, MD[1,12,13]
Tyler M. Seibert, MD, PhD[1,6,14,15]

[1] Research Service, VA San Diego Healthcare System, San Diego, CA
[2] Medical Scientist Training Program, University of California San Diego, La Jolla, CA
[3] Biomedical Science Program, University of California San Diego, La Jolla, CA
[4] VA Informatics and Computing Infrastructure, VA Salt Lake City Healthcare System (VINCI), Salt Lake City, UT
[5] Department of Internal Medicine, Division of Epidemiology, University of Utah School of Medicine, Salt Lake City, UT
[6] Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA
[7] Department of Medicine, University of California San Diego, La Jolla, CA
[8] Department of Psychiatry, University of California San Diego, La Jolla, CA
[9] Center for Healthcare Organization and Implementation Research (CHOIR), VA Bedford Healthcare System, Bedford, MA
[10] Section of General Internal Medicine, Boston University School of Medicine, Boston, MA
[11] Department of Urology, University of California San Diego, La Jolla, CA
[12] Center for Behavioral Genetics of Aging, University of California San Diego, La Jolla, CA
[13] Center of Excellence for Stress and Mental Health (CESAMH), VA San Diego Healthcare System, San Diego, CA
[14] Department of Radiology, University of California San Diego, La Jolla, CA
[15] Department of Bioengineering, University of California San Diego, La Jolla, CA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract:

Importance: Early detection of prostate cancer to reduce mortality remains controversial because there is often also overdiagnosis of low-risk disease and unnecessary treatment. Genetic scores may provide an objective measure of a man's risk of dying from prostate cancer and thus inform screening decisions, especially in men of African ancestry, who have a higher average risk of prostate cancer death but are often treated as a homogeneous group.

Objective: Determine whether a polygenic hazard score based on 290 genetic variants (PHS290) is associated with risk of metastatic or fatal prostate cancer in a racially and ethnically diverse population.

Design: Million Veteran Program (MVP) cohort study, 2011-2021.

Setting: Nation-wide study of United States military veterans.

Participants: Population-based volunteer sample of male participants.

Exposure(s): Genotype data were used to calculate the genetic score, PHS290. Family history of prostate cancer and ancestry group (harmonized genetic ancestry and self-reported race/ethnicity: European, African, Hispanic, or Asian) were also studied.

Main Outcome(s) and Measure(s): Study designed after MVP data collected. Primary outcome: age at death from prostate cancer. Key secondary outcome: age at diagnosis of prostate cancer metastases. Hypothesis: A germline genetic score (PHS290) is associated with risk of fatal (or metastatic) prostate cancer.

Results: 513,997 MVP participants were included. Median age at last follow-up: 69 years. PHS290 was associated with age at death from prostate cancer in the full cohort and for each ancestry group ($p<10^{-16}$). Comparing men in the highest 20% of PHS290 to those in the lowest 20%, the hazard ratio for death from prostate cancer was 4.41 [95% CI: 3.9-5.02]. Corresponding hazard ratios for European, African, Hispanic, and Asian subsets were 4.26 [3.66-4.9], 2.4 [1.77-3.23], 4.72 [2.68-8.87], and 10.46 [2.01-101.0]. When accounting for family history and ancestry group, PHS290 remained a strong independent predictor of fatal prostate cancer. PHS290 was also associated with metastasis. PHS290 was higher, on average, among men with African ancestry.

Conclusions and Relevance: PHS290 stratified US veterans of diverse ancestry for lifetime risk of metastatic or fatal prostate cancer. Predicting genetic risk of lethal prostate cancer with PHS290 might inform individualized decisions about prostate cancer screening.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction:

Prostate cancer is the most diagnosed and second deadliest cancer in men[1]. Despite the enormous mortality from this disease, early detection of prostate cancer remains controversial. Screening all men via prostate-specific antigen (PSA) testing, regardless of underlying risk, has been shown to reduce prostate cancer deaths by 27% but also results in frequent overdiagnosis of indolent prostate cancer that may never have become symptomatic[2–4]. These overdiagnoses often lead to unnecessary treatment, with attendant side effects and societal costs. A better strategy is to target PSA screening to those men at higher risk of developing metastatic or fatal prostate cancer.

As one of the most heritable cancers[5], genetic risk stratification is a promising approach for identifying individuals at higher risk of developing metastatic or fatal prostate cancer[1,3,6].Measures of genetic risk have proven highly effective for predicting lifetime risk of being diagnosed with prostate cancer, outperforming family history or other clinical risk factors[7–10]. Rather than only predicting lifetime risk, however, an ideal genetic test would focus on clinically significant prostate cancer and estimate age-specific risk. Prostate cancer is highly age dependent, with very low incidence before 50 years of age and increasing exponentially as men get older[11,12]. Absolute incidence of *aggressive* prostate cancer also increases with age[11,12]. Meanwhile, some men with high genetic risk develop aggressive prostate cancer at a younger age and are at particular risk of dying from this disease. Age-specific genetic risk could inform individualized decisions about PSA testing, in the context of a given man's overall health and competing causes of mortality.

A major limitation of early studies of polygenic risk was an exclusive focus on men of European ancestry[13,14]. Such systematic bias may exacerbate existing health disparities in prostate cancer incidence and health outcomes[15,16]. This is particularly worrisome for men of African ancestry, who have a higher overall incidence of metastatic and fatal prostate cancer than men of European or Asian ancestry[17,18].

Our group has developed a novel risk prediction tool called a polygenic hazard score (PHS) that identifies men who are likely to develop clinically significant prostate cancers at younger ages. This score, which can be calculated from a single saliva sample at any point in a man's life, was strongly associated with age at diagnosis of clinically significant prostate cancer in large datasets[10,19,20]. The score also improved the accuracy of conventional screening with PSA[7,10,12]. We subsequently expanded the model to optimize performance in men of all ancestries, particularly men with African ancestry[19–21]. Here, we seek to validate the ability of the PHS to identify men at risk of metastatic or fatal prostate cancer within the Million Veteran Program (MVP), one of the largest and most racially and ethnically diverse populations studied to date[22].

Methods:

Participants

We retrospectively obtained data from the MVP, composed of individuals between ages 19 to over 100 years who were recruited from 63 Veterans Affairs Medical Centers across the United States (US). Recruitment for the MVP started in 2011, and all veterans were eligible for participation. Consent to participate and permission to re-contact was provided after counseling by research staff and mailing of informational materials. Study participation included consenting to access the participant's electronic health records for research purposes. The MVP received ethical and study protocol approval from the VA Central Institutional Review Board in accordance with the principles outlined in the Declaration of Helsinki.

Only men were included in this prostate cancer study, comprising 513,997 individuals of European (73.3%), African (17.2%), Hispanic (8.2%), and Asian ancestry (1.3%) **(Table 1).** There were no inclusion or exclusion criteria for age. Age at last follow-up was 69 years (interquartile range 59-74 years). Men not meeting the endpoint for each analysis were censored at age at last follow-up. Clinical information used for analyses was retrieved as described below in the Clinical Data Extraction section.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Genotype Data

All study participants provided blood samples for DNA extraction and genotyping. Researchers are provided data that is de-identified except for dates. Blood samples were collected by phlebotomists and banked at the VA Central Biorepository in Boston, MA, where DNA was extracted and shipped to two external centers for genotyping. DNA extracted from buffy coat was genotyped using a custom Affymetrix Axiom biobank array. The MVP 1.0 genotyping array contains a total of 723,305 single nucleotide polymorphisms (SNPs), enriched for low frequency variants in African and Hispanic populations, and variants associated with diseases common to the VA population[22].

Harmonized Ancestry and Race/Ethnicity (HARE)

The MVP has previously categorized individuals into ancestry groups called HARE (Harmonized Ancestry and Race/Ethnicity) groups using a machine learning algorithm[23]. HARE utilizes a support vector machine to estimate probabilities of an individual belonging to one of four ancestry groups using both self-identified race/ethnicity and genetic ancestry[23]. Information on race and ethnicity was obtained based on self-report through centralized VA data collection methods using standardized survey forms or using information from the VA Corporate Data Warehouse or Observational Medical Outcomes Partnership data. All but 9,989 (1.52%) MVP participants were assigned to a HARE ancestry group. The support vector machine was trained with the top 30 principal components of population stratification analysis and self-identified ancestry. Regularization constant *C* and inverse variance of kernel were optimized through 2-dimensional grid searching and 5-fold cross-fold validation. Individuals were categorized as predominantly European, Hispanic, African, or Asian based on output probabilities.

Clinical Data Extraction

Each participant's electronic health record is integrated into the MVP biorepository. These records include International Classification of Diseases (ICD) diagnosis codes (ICD-9-CM and ICD-10-CM), procedure codes (ICD, Current Procedural Terminology, and Healthcare Common Procedure Coding (HCPCS)), laboratory values, medications, and clinical notes documenting VA care (inpatient and outpatient) and non-VA care paid for by the VA.

Prostate cancer diagnosis, age at diagnosis, and date of last follow-up were retrieved from the VA Corporate Data Warehouse based on ICD codes and VA Central Cancer Registry data. Age at diagnosis of metastasis was determined via a validated natural language processing tool and a search of individual participant's medical records in the Veterans Affairs system, as described previously[24]. This tool was developed using data from over 1 million VA patients with prostate cancer; compared to manual chart review, the natural language processing tool had 92% sensitivity and 98% specificity for diagnosis of metastatic prostate cancer. Cause and date of death was collected from National Death Index. Participants with ICD10 code "C61" as underlying cause of death were considered to have died from prostate cancer. Age of death was determined from difference between year of death and year of birth.

		HARE groups*			
	All	European	African	Hispanic	Asian
Partipants					
All participants	582,515	425,251	104,582	45,668	7014
Prostate cancer	68,538	48,642	16,129	3375	392
Metastases from prostate cancer	6,365	4357	1650	319	39
Death from prostate cancer	1,847	1,384	377	79	7

Table 1 Participant characteristics, n=582,515

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Polygenic Hazard Score (PHS290)

The most recent version of the PHS, called PHS290, was calculated as the vector product of participants' genotype dosage (X_i) for 290 SNPs and the corresponding parameter estimates (β_i) from Cox proportional hazards regression:

$$PHS = \sum_{i}^{n} X_i \beta_i$$

The development of this score has been described elsewhere[21]. Briefly, 299 previously identified SNPs associated with prostate cancer risk (in single-ancestry or all-ancestry analyses) were simultaneously evaluated using a machine-learning least absolute shrinkage and selection operator (LASSO) approach to generate an optimal combined model for association with age at prostate cancer diagnosis.

We calculated PHS290 for each MVP participant. Distributions were visualized using histograms for each ancestry group. Differences in mean PHS290 between ancestry groups were assessed via ANOVA. In all statistical analyses, significance was set at a two-tailed alpha of 0.01. As in prior studies, *p*-values less than 10^{-16} were truncated at this value, as comparison of miniscule values is not likely to be meaningful[7,10,12,20].

Cox Proportional Hazards Analysis

We evaluated association of PHS290 with two important clinical endpoints extracted from clinical data: age at death from prostate cancer (i.e., lifetime prostate-cancer-specific mortality) and age at diagnosis of metastases from prostate cancer (i.e., lifetime distant-metastasis-free survival). Secondarily, we also tested for association with age at diagnosis of any prostate cancer. Cox proportional hazards models were used to assess these associations in the full dataset and in each ancestral group (European, African, Hispanic, Asian). Individuals not meeting the endpoint of interest were censored at age at last follow-up. Sample-weight corrections were applied to all Cox models to correct for potential bias due to a higher number of prostate cancer cases compared to a general population and permit direct comparisons to other studies[7,9,19,25]. Although underlying population incidence may vary (e.g., across ancestry groups), a constant correction factor was used for all analyses (in this case, based on previously reported population data from Sweden, for consistency with prior work), as effect size estimates have been shown to be robust to wide variation in population incidence[19].

Effect sizes were estimated using hazard ratios (HRs) between risk strata, as described previously[7,9,10,12,19,20,26]. Percentiles of genetic risk were calculated using percentile thresholds defined in a prior study of men of European ancestry less than 70 years old and with no diagnosis of prostate cancer[21]. HRs for each ancestry group were calculated to make the following comparisons: $HR_{80/20}$, men in the highest 20% vs. lowest 20%; $HR_{95/50}$, men in the highest 5% of genetic risk vs. those with average risk (30–70th percentile); and $HR_{20/50}$, men in the lowest 20% vs. those with average risk. These risk groups were chosen to mirror prior work and permit direct comparison of HRs[7,19,20].

Ancestry, Family History, and PHS290

To assess the added value of PHS290 beyond commonly used clinical risk factors, we tested a multivariable Cox proportional hazards model with ancestry group, family history, and PHS290[7,9,19]. This combined model was limited to the 374,455 participants who provided family history information in baseline survey data. Family history was recorded as either the presence or absence of (one or more) first-degree relatives with prostate cancer. Cox proportional hazards models tested associations with fatal, metastatic, or any prostate cancer. For PHS290, the effect size was illustrated via the hazard ratio for the highest 20% vs. lowest 20% of genetic risk. Hazard ratios for ancestry groups were estimated using European as the reference.

A univariable Cox proportional hazards model was applied to test for association of ancestry group with prostate cancer risk. Similarly, a univariable model tested for association of family history alone. The @anova function

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

from the *R* 'survival' package (version 3.2-13; Therneau 2021) was used to compare the nested Cox models (multivariable vs. univariable), based on the log partial likelihood of the model fits. Significance was set at a two-tailed alpha of 0.01 for the test of whether the multivariable model performed better than either univariable model alone.

Results:

PHS290 Score

The distribution of PHS290 in the European ancestry group was similar to that reported previously for men of European ancestry (mean=9.37, SD=0.37)[21]. Mean PHS290 did vary by ancestry group with statistically significant differences between all groups (ANOVA $p < 10^{-16}$); all pair-wise *t*-tests also $p < 10^{-16}$). The distribution for the Hispanic ancestry group overlapped closely with that of the European group (mean=9.35, SD=0.37), while PHS290 tended to be lower among men of Asian ancestry (mean=9.17, SD=0.36) and higher among men of African ancestry (mean=9.56, SD=0.34, t-test) (**Figure 1**).

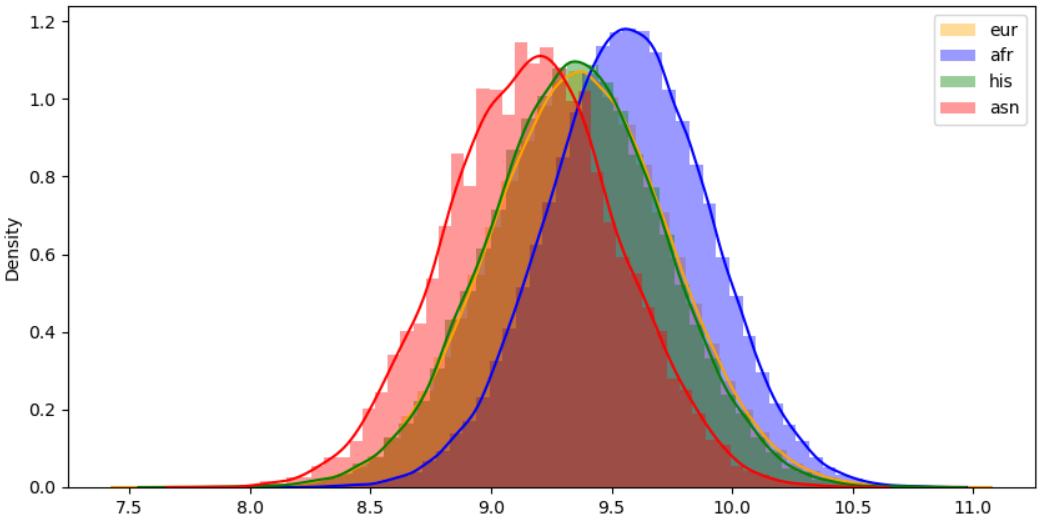

Figure 1. PHS290 score density plot in European, African, Hispanic, and Asian ancestry groups.

Association of PHS290 with Fatal Prostate Cancer

PHS290 was associated with age at death from prostate cancer in the full dataset and in each of the four ancestry groups **(Table 2)**. Comparing 80th and 20th percentiles of genetic risk in the full dataset, men with higher PHS290 had an $HR_{80/20}$ of 4.41 [95% CI: 3.9-5.02]. Cumulative incidence curves for various PHS290 demonstrated risk stratification **(Figure 2)**. Hazard ratios quantified significant risk stratification using PHS290 in the full MVP dataset and in each ancestry group (though confidence intervals were large in the Asian ancestry group). For European, African, Hispanic, and Asian men, $HRs_{80/20}$ were 4.26 [95% CI: 3.66-4.9], 2.4 [1.77-3.23], 4.72 [2.68-8.87] and 10.46 [2.01-101.0], respectively.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

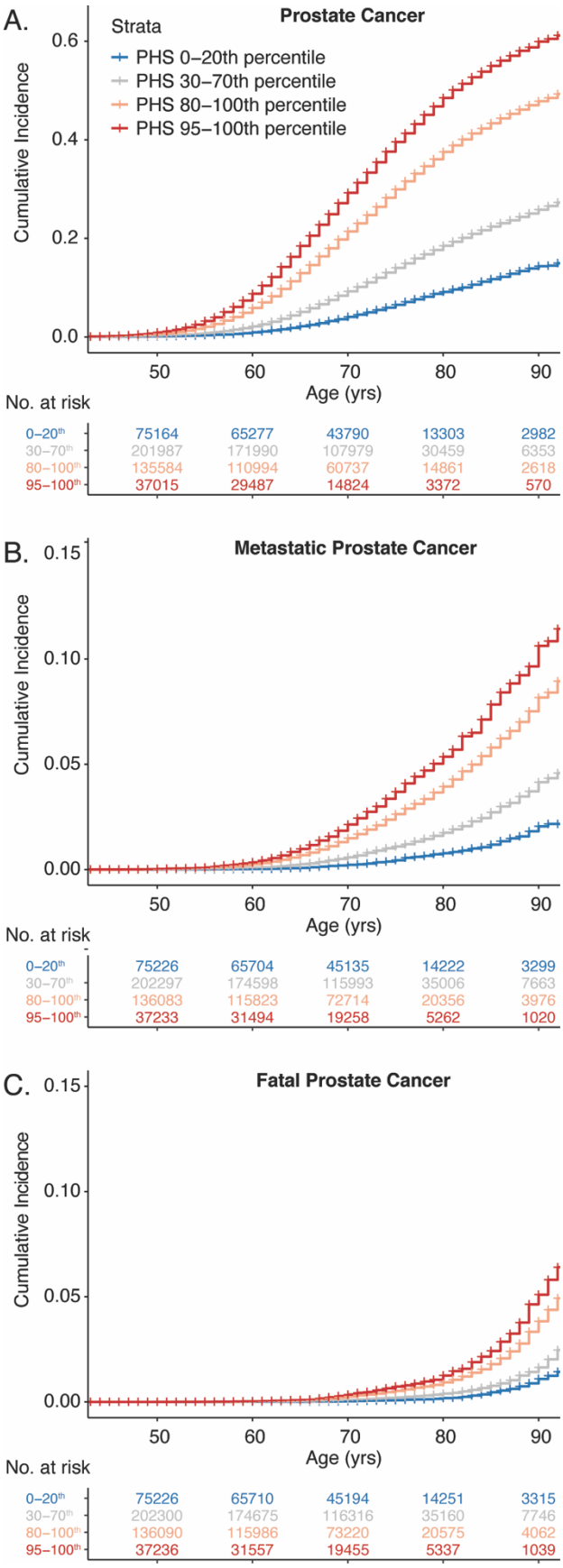

Figure 2. Cumulative Incidence Curves in by MVP PHS290 Strata. **(A)** Cumulative incidence curves for prostate cancer. **(B)** Cumulative incidence curves for metastatic prostate cancer. (C) Cumulative incidence curves for fatal prostate cancer. For all curves, 0-20th, 30-60th, 80-100th, and 95-100th PHS290 percentile groups are plotted.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ancestry group	p Value	Hazard ratios $HR_{80/20}$	$HR_{20/50}$	$HR_{80/50}$	$HR_{95/50}$
All (n=582,516)	$<10^{-16}$	4.41 [3.9-5.02]	0.48 [0.45-0.51]	2.12 [1.99-2.27]	2.99 [2.74-3.3]
European (n=425,252)	$<10^{-16}$	4.26 [3.66-4.9]	0.49 [0.46-0.53]	2.08 [1.93-2.24]	2.94 [2.64-3.27]
African (n=104,583)	2.14E-08	2.4 [1.77-3.23]	0.66 [0.57-0.76]	1.58 [1.34-1.84]	1.92 [1.53-2.39]
Hispanic (n=45,669)	1.75E-06	4.72 [2.68-8.87]	0.46 [0.34-0.61]	2.17 [1.64-2.97]	3.17 [2.09-5.04]
Asian (n=7015)	0.01	10.46 [2.01-101.0]	0.3 [0.09-0.7]	3.13 [1.41-9.47]	5.71 [1.69-29.76]

Table 2: Association of PHS290 with fatal prostate cancer. *P*-values reported are for the overall performance of Cox models using PHS290 as the sole predictor variable. Hazard ratios (HRs) are shown comparing men in various percentiles of genetic risk. $HR_{80/20}$: highest 20% (≥80th percentile of PHS290, using previously published thresholds for men <70 years old and no diagnosis of cancer) vs. average risk (30-70th percentile). $HR_{20/50}$: lowest 20% (≤20th percentile) vs. average risk. $HR_{80/50}$: highest 20% vs. average risk. $HR_{95/50}$: highest 5% (≥95th percentile) vs. average risk. Numbers in brackets are 95% confidence intervals.

Association of PHS290 with Metastatic Prostate Cancer

PHS290 was associated with age at diagnosis of metastases from prostate cancer in the full dataset and in each of the four ancestry groups **(Table 3).** Comparing 80th and 20th percentiles of genetic risk, men with higher PHS290 had an $HR_{80/20}$ of 4.94 [95% CI: 4.58-5.28]. For European, African, Hispanic, and Asian men, $HRs_{80/20}$ were 4.64 [4.28-5.03], 3.02 [2.6-3.57], 3.95 [2.87-5.37] and 6.85 [2.84-17.12], respectively.

Ancestry group	p Value	Hazard ratios $HR_{80/20}$	$HR_{20/50}$	$HR_{80/50}$	$HR_{95/50}$
All (n=582,516)	$<10^{-16}$	4.94 [4.58-5.28]	0.45 [0.44-0.47]	2.25 [2.16-2.32]	3.26 [3.09-3.43]
European (n=425,252)	$<10^{-16}$	4.64 [4.28-5.03]	0.47 [0.45-0.49]	2.17 [2.09-2.27]	3.14 [2.95-3.33]
African (n=104,583)	$<10^{-16}$	3.02 [2.6-3.57]	0.59 [0.54-0.63]	1.78 [1.64-1.94]	2.28 [2.04-2.58]
Hispanic (n=45,669)	$<10^{-16}$	3.95 [2.87-5.37]	0.5 [0.43-0.59]	1.99 [1.7-2.32]	2.77 [2.2-3.48]
Asian (n=7015)	$<10^{-4}$	6.85 [2.84-17.12]	0.37 [0.23-0.58]	2.54 [1.66-3.94]	4.17 [2.14-8.43]

Table 3: Association of PHS290 with metastatic prostate cancer. *P*-values reported are for the overall performance of Cox models using PHS290 as the sole predictor variable. Hazard ratios (HRs) are shown comparing men in various percentiles of genetic risk. $HR_{80/20}$: highest 20% (≥80th percentile of PHS290, using previously published thresholds for men <70 years old and no diagnosis of cancer) vs. average risk (30-70th percentile). $HR_{20/50}$: lowest 20% (≤20th percentile) vs. average risk. $HR_{80/50}$: highest 20% vs. average risk. $HR_{95/50}$: highest 5% (≥95th percentile) vs. average risk. Numbers in brackets are 95% confidence intervals.

Association of PHS290 with Prostate Cancer

PHS290 was associated with age at prostate cancer diagnosis in all 4 ancestry groups **(Table 4)**. Comparing 80th and 20th percentiles of genetic risk, men with higher PHS290 had an $HR_{80/20}$ of 6.29 [95% CI: 6.14-6.46]. For European, African, Hispanic, and Asian men, $HRs_{80/20}$ were 6.19 [6.01-6.38], 3.83 [3.61-4.08], 4.75 [4.22-5.32] and 5.52 [3.98-7.56], respectively.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ancestry group	p Value	Hazard ratios $HR_{20/50}$	$HR_{80/20}$	$HR_{80/50}$	$HR_{95/50}$
All (n=582,516)	$<10^{-16}$	0.4 [0.4-0.41]	6.29 [6.14-6.46]	2.54 [2.51-2.57]	3.89 [3.83-3.97]
European (n=425,252)	$<10^{-16}$	0.41 [0.4-0.41]	6.19 [6.01-6.38]	2.52 [2.48-2.55]	3.89 [3.81-3.98]
African (n=104,583)	$<10^{-16}$	0.52 [0.51-0.54]	3.83 [3.61-4.08]	2.01 [1.95-2.08]	2.72 [2.6-2.85]
Hispanic (n=45,669)	$<10^{-16}$	0.46 [0.43-0.49]	4.75 [4.22-5.32]	2.18 [2.06-2.31]	3.18 [2.91-3.47]
Asian (n=7015)	$<10^{-16}$	0.41 [0.35-0.49]	5.52 [3.98-7.56]	2.29 [1.96-2.67]	3.56 [2.79-4.57]

Table 4: Association of PHS290 with prostate cancer. *P*-values reported are for the overall performance of Cox models using PHS290 as the sole predictor variable. Hazard ratios (HRs) are shown comparing men in various percentiles of genetic risk. $HR_{80/20}$: highest 20% (≥80th percentile of PHS290, using previously published thresholds for men <70 years old and no diagnosis of cancer) vs. average risk (30-70th percentile). $HR_{20/50}$: lowest 20% (≤20th percentile) vs. average risk. $HR_{80/50}$: highest 20% vs. average risk. $HR_{95/50}$: highest 5% (≥95th percentile) vs. average risk. Numbers in brackets are 95% confidence intervals.

Ancestry, Family History, and PHS290

Ancestry group, alone, was associated with differential risk of fatal prostate cancer **(Supplementary Table 1)**. These associations were largely driven by an increased risk from African ancestry. Similar patterns were seen for age at diagnosis of prostate cancer and for age at diagnosis of prostate cancer metastasis **(Supplementary Table 1)**. Compared to the European group, men in the African ancestry group had a HR of 2.65 [2.37-2.96] for dying of prostate cancer.

Family history, alone, was also associated with fatal prostate cancer, as well as with diagnosis of prostate cancer and with age at diagnosis of prostate cancer metastasis (**Supplementary Table 2**). Compared to men with no family history of prostate cancer, men with one or more first-degree relatives who had prostate cancer had a HR of 1.84 [1.54-2.17] for dying of prostate cancer.

PHS290 added significant value beyond ancestry group or family history in a multivariable model that included all three variables and tested for association with age at prostate cancer death (**Table 5**). The multivariable model improved performance over the common risk factors alone ($p < 10^{-16}$). Similarly, the combination proved optimal when evaluating age at diagnosis of prostate cancer diagnosis or age at diagnosis of prostate cancer metastasis ($p<10^{-16}$). Independent of ancestry and family history, a high PHS290 (top 20%) approximately quadrupled a man's risk of death from prostate cancer, compared to a low PHS290 (bottom 20%) **(Table 5)**.

	PHS290	Ancestry			Family History
Endpoint	HR80_20	African	Hispanic	Asian	FH
Prostate Cancer	5.66 [5.48-5.84]***	1.84 [1.79-1.9]***	1.06 [1.01-1.12]*	1.25 [1.1-1.42]*	1.9 [1.84-1.97]***
Metastatic	4.18 [3.84-4.53]***	2.27 [2.09-2.46]***	1.3 [1.1-1.5]	1.42 [0.88-2.01]	1.52 [1.39-1.69]*
Fatal	4.15 [3.59-4.83]***	1.97 [1.67-2.29]***	0.97 [0.69-1.32]*	0.69 [0.17-1.45]	1.67 [1.4-1.99]*

Table 5: Ancestry, Family History, and PHS290 in multivariable models for three clinical endpoints. For PHS290, effect size was illustrated via the hazard ratio ($HR_{80/20}$) for the highest 20% vs. lowest 20% of genetic risk. Hazard ratios for ancestry groups were estimated using European as the reference. Hazard ratios for family history were for one or more first-degree relatives diagnosed with prostate cancer. This analysis was limited to the 374,455 participants who provided family history information in baseline survey data. Numbers in brackets are 95% confidence intervals. Significant predictors in the multivariable model are indicated by * ($p<0.01$) and *** ($p<10^{-16}$).

Discussion:

PHS290 was associated with lifetime prostate-cancer-specific mortality in this large and diverse dataset. Even when accounting for family history and ancestry group (harmonized genetic ancestry and self-reported

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

race/ethnicity), PHS290 remained a strong independent predictor of dying from prostate cancer. The genomic score was also associated with age at diagnosis of metastasis from prostate cancer and at age at diagnosis of any prostate cancer, consistent with previous reports that common genetic markers for overall prostate cancer risk overlap with those for aggressive prostate cancer risk[27,28]. Metastatic prostate cancer has poor prognostic outcomes and is major driver of pain, disability and aggressive medical therapy[29]. To our knowledge, this study is the first to show the association of a genomic score with lifetime risk of metastatic prostate cancer. This study also represents the largest and most ancestry-diverse independent validation of polygenic association with lifetime risk of fatal prostate cancer.

Men of African ancestry in the US are substantially more likely to develop metastatic disease and to die from prostate cancer[18]. The causes of this disparity are likely a combination of genetic, environmental, and social factors, including systemic racism[30–34]. National guidelines recommend consideration of prostate screening in men of African ancestry at a younger age and that screening occur at more frequent intervals[35]. The results of the present study demonstrate that men of African ancestry have highly variable levels of lifetime risk and should not be treated as a homogeneous group. This finding is consistent with prior results and with the known admixture and genetic diversity of the African American population[9,26]. Moreover, PHS290 can identify those more likely to develop lethal prostate cancer and may facilitate personalized screening recommendations.

Intriguingly, typical PHS290 scores differed between ancestry groups, with the mean PHS290 slightly higher in the African ancestry group and slightly lower in Asian ancestry group than in the European or Hispanic groups. These shifts in PHS290 distribution are consistent with reported differences in prostate cancer incidence across racial groups[36–40]. Higher overall PHS290 scores in African ancestry group may point to true differences in prostate cancer risk but could also be inflated by minor allele frequency (MAF) differences between ancestry groups. Incorporating approaches for local ancestry and admixture can also boost genetic model performance and should be explored further to improve the predictive accuracy of polygenic scores[41].

Along with race/ancestry, family history is another important clinical consideration in prostate cancer screening decisions[35,42–45]. Prior studies have found polygenic scores to be the most important risk factor for prostate cancer, with family history sometimes offering modest improvement in multivariable models, possibly by capturing yet unknown genetic factors and/or shared familial environmental factors[7,9,19,46]. Among the subset of MVP participants who provided family history information, family history of prostate cancer was independently associated with prostate cancer risk in a multivariable model that included PHS290 and ancestry group. The relationship of environmental exposures, family history, and prostate cancer risk are worth further investigation[9], particularly in groups like veterans who may have been exposed to rare carcinogens[47].

The present study builds on prior work that reported the performance of polygenic scores in non-Europeans[19,21,26,27,48] and is consistent with those prior studies in showing a strong association of polygenic scores with prostate cancer risk, including death from prostate cancer[9,19,46]. Polygenic hazard scores designed to incorporate the strong age-dependence of prostate cancer have also been shown to increase the accuracy of conventional prostate cancer screening[7,10,12]. Population-level analyses of benefit, harm, and cost-effectiveness support incorporation of genomic risk into screening[3,6]. The present study adds to the literature an independent validation in a dataset of over 500,000 men with diverse ancestry. Current clinical guidelines try to achieve targeted, or risk-stratified, screening by recommending each man discuss his individual risk factors, emphasizing racial/ethnic background[35,43–45]. It is particularly important, therefore, that this study was able to combine ancestry and genetic risk to estimate the relative impact of each and to demonstrate that a polygenic score adds considerable information beyond ancestry alone, for a man's individual risk of metastasis or death from prostate cancer.

While PHS290 performed well in the present study to stratify men by genetic prostate cancer risk, the effect sizes estimated here are lower than those reported in previous studies. For example, $HR_{80/20}$ for fatal prostate cancer using PHS290 was 7.73 [95% CI 6.45, 9.27] in a population-based Swedish cohort[21], compared to 4.41 [3.9-5.02] in all participants and 4.26 [3.66-4.9] in the European ancestry group in the present study. A similar pattern of smaller effect size was seen when comparing the strength of association with age at diagnosis of prostate cancer within MVP ancestry groups to the hazard ratios previously reported for European, Asian, and African genetic ancestry groups[21]. Most likely, the discrepancy arises in differences in the populations studied; for example, the MVP dataset comes exclusively from a population of US veterans, with many receiving care in a

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

single US-based healthcare system, whereas the prior study used data from multiple countries and widely varying recruitment strategies. Patterns of screening, detection, and treatment of prostate cancer in the present dataset could be different from clinical trial and case-control datasets used in previous work. Some of the difference in performance could also be explained by the fact that the testing datasets in the prior report for PHS290 had been included in the discovery of a majority of the 290 SNPs in the model; on the other hand, the testing datasets represented a very small proportion of the discovery datasets, and the model weights were estimated in an independent training dataset.

Limitations of this study include heterogeneity of phenotyping and smaller sample sizes for Asian ancestry group. Heterogeneity of prostate cancer screening and diagnostic pathways by clinicians across VA and other hospitals in the US could potentially introduce noise, although this heterogeneity likely leads to underestimation of associations with prostate cancer. Large confidence intervals in the Asian ancestry group may be due to relatively smaller sample sizes, but they also suggest increased heterogeneity compared to prior datasets (including Asian ancestry) of similar sample size[19,21,49]. Finally, we acknowledge that while HARE ancestry groups may be a reasonable attempt to harmonize genetic ancestry, race, and ethnicity, these groups cannot account for—much less, disentangle—the complex web of biological and social factors associated with these categories. Further work will attempt to incorporate agnostic genetic ancestry groups and address impacts of admixture and local/regional genetic ancestry on risk stratification with PHS[19].

We show that PHS290 stratified US men for lifetime risk of metastatic or fatal prostate cancer. Critically, this genetic risk stratification was successful within each of four ancestry groups in this diverse dataset. PHS290 was higher, on average, among men with African ancestry, who were also at higher risk from prostate cancer. The combination of ancestry, family history, and PHS290 performed better than any variable, alone, in identifying men at highest risk of prostate cancer metastasis and death. Predicting genetic risk of lethal prostate cancer with PHS290 might inform individualized decisions about screening and early cancer detection.

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References:

1. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
2. Schröder, F. H. *et al.* Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* **384**, 2027–2035 (2014).
3. Callender, T. *et al.* Polygenic risk-tailored screening for prostate cancer: A benefit-harm and cost-effectiveness modelling study. *PLoS Med.* **16**, e1002998 (2019).
4. Loeb, S. *et al.* Overdiagnosis and overtreatment of prostate cancer. *Eur. Urol.* **65**, 1046–1055 (2014).
5. Mucci, L. A. *et al.* Familial risk and heritability of cancer among twins in Nordic countries. *JAMA* **315**, 68–76 (2016).
6. 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).
7. Seibert, T. M. *et al.* Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ* **360**, j5757 (2018).
8. Schumacher, F. R. *et al.* Author Correction: Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **51**, 363 (2019).
9. 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.* (2021) doi:10.1038/s41391-021-00341-4.
10. 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).
11. 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).
12. 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).
13. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

14. Petrovski, S. & Goldstein, D. B. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* **17**, (2016).

15. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

16. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).

17. DeSantis, C. E., Miller, K. D., Goding Sauer, A., Jemal, A. & Siegel, R. L. Cancer statistics for African Americans, 2019. *CA Cancer J. Clin.* **69**, 211–233 (2019).

18. Tsodikov, A. *et al.* Is prostate cancer different in black men? Answers from 3 natural history models. *Cancer* **123**, 2312–2319 (2017).

19. Huynh-Le, M.-P. *et al.* Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. *Nat. Commun.* **12**, 1236 (2021).

20. 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).

21. Huynh-Le, M.-P. *et al.* Prostate cancer risk stratification improved across multiple ancestries with new polygenic hazard score. (2021) doi:10.1101/2021.08.14.21261931.

22. 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).

23. Fang, H. *et al.* Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *Am. J. Hum. Genet.* **105**, 763–772 (2019).

24. Alba, P. R. *et al.* Ascertainment of Veterans with Metastatic Prostate Cancer in Electronic Health Records: Demonstrating the Case for Natural Language Processing. *In Press*.

25. Therneau, T. M. & Li, H. Computing the Cox model for case cohort designs. *Lifetime Data Anal.* **5**, 99–112 (1999).

26. Karunamuni, R. A. *et al.* Performance of African-ancestry-specific polygenic hazard score varies according to local ancestry in 8q24. *Prostate Cancer Prostatic Dis.* (2021) doi:10.1038/s41391-021-00403-7.

27. Conti, D. V. *et al.* Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet.* **53**, 65–75 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

28. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).

29. Sartor, O. & de Bono, J. S. Metastatic prostate cancer. *N. Engl. J. Med.* **378**, 645–657 (2018).

30. Barocas, D. A. *et al.* Association between race and follow-up diagnostic care after a positive prostate cancer screening test in the prostate, lung, colorectal, and ovarian cancer screening trial. *Cancer* **119**, 2223–2229 (2013).

31. Han, Y. *et al.* Prostate cancer susceptibility in men of African ancestry at 8q24. *J. Natl. Cancer Inst.* **108**, djv431 (2016).

32. Mahal, B. A. *et al.* Trends in disparate treatment of African American men with localized prostate cancer across National Comprehensive Cancer Network risk groups. *Urology* **84**, 386–392 (2014).

33. Yamoah, K. *et al.* Novel biomarker signature that may predict aggressive disease in African American men with prostate cancer. *J. Clin. Oncol.* **33**, 2789–2796 (2015).

34. Zhang, H. *et al.* Age and racial differences among PSA-detected (AJCC stage T1cN0M0) prostate cancer in the U.s.: A population-based study of 70,345 men. *Front. Oncol.* **3**, 312 (2013).

35. Carroll, P. R. *et al.* NCCN Guidelines Prostate Cancer Early Detection Version 1.2021. (2021).

36. Siegel, D. A., O'Neil, M. E., Richards, T. B., Dowling, N. F. & Weir, H. K. Prostate cancer incidence and survival, by stage and race/ethnicity - United States, 2001-2017. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 1473–1480 (2020).

37. Cook, M. B. *et al.* Racial disparities in prostate cancer incidence rates by census division in the United States, 1999-2008. *Prostate* **75**, 758–763 (2015).

38. Atlas of cancer mortality in the United States, 1950-94. *Int. J. Epidemiol.* **29**, 602-a-602 (2000).

39. Jemal, A. *et al.* Geographic patterns of prostate cancer mortality and variations in access to medical care in the United States. *Cancer Epidemiol. Biomarkers Prev.* **14**, 590–595 (2005).

40. Curry, S. J., Krist, A. H. & Owens, D. K. Annual report to the nation on the status of cancer, part II: Recent changes in prostate cancer trends and disease characteristics. *Cancer* vol. 125 317–318 (2019).

41. Marnetto, D. *et al.* Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat. Commun.* **11**, 1628 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2021.09.24.21264093; this version posted September 27, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

42. Powell, I. J. The precise role of ethnicity and family history on aggressive prostate cancer: a review analysis. *Arch. Esp. Urol.* **64**, 711–719 (2011).

43. Wolf, A. M. D. *et al.* American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J. Clin.* **60**, 70–98 (2010).

44. Horwich, A. *et al.* Prostate cancer: ESMO Consensus Conference Guidelines 2012. *Ann. Oncol.* **24**, 1141–1162 (2013).

45. Qaseem, A. *et al.* Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann. Intern. Med.* **158**, 761–769 (2013).

46. Shi, Z. *et al.* Performance of three inherited risk measures for predicting prostate cancer incidence and mortality: A population-based prospective analysis. *Eur. Urol.* **79**, 419–426 (2021).

47. Schecter, A. *et al.* Agent Orange exposure, Vietnam war veterans, and the risk of prostate cancer. *Cancer* vol. 115 3369–3371 (2009).

48. Plym, A. *et al.* Evaluation of a multiethnic polygenic risk score model for prostate cancer. *J. Natl. Cancer Inst.* (2021) doi:10.1093/jnci/djab058.

49. Karunamuni, R. A. *et al.* The effect of sample size on polygenic hazard models for prostate cancer. *Eur. J. Hum. Genet.* **28**, 1467–1475 (2020).