Polygenic risk of any, metastatic, and fatal prostate cancer in the Million Veteran Program

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]
J. Michael Gaziano, MD, MPH[9,10]
Guneet Kaur Jasuja, MPH, PhD[11,12]
Rishi Deka, PhD[1,6,8]
Brent S. Rose, MD[1,6,12]
Matthew S. Panizzon, PhD[1,13]
Richard L. Hauger, MD[1,13,14]
Tyler M. Seibert, MD, PhD[1,6,15,16,#]

[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] Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, MA
[10] Department of Medicine, Brigham and Women's Hospital, Boston, MA
[11] Center for Healthcare Organization and Implementation Research (CHOIR), VA Bedford Healthcare System, Bedford, MA
[12] Section of General Internal Medicine, Boston University School of Medicine, Boston, MA
[12] Department of Urology, University of California San Diego, La Jolla, CA
[13] Center for Behavioral Genetics of Aging, University of California San Diego, La Jolla, CA
[14] Center of Excellence for Stress and Mental Health (CESAMH), VA San Diego Healthcare System, San Diego, CA
[15] Department of Radiology, University of California San Diego, La Jolla, CA
[16] Department of Bioengineering, University of California San Diego, La Jolla, CA

#Corresponding author:
Tyler M. Seibert
9500 Gilman Dr.
La Jolla, CA 92093-0861
T: (858) 246-5988
F: (858) 246-0501
tseibert@ucsd.edu

Keywords: prostate cancer; polygenic hazard score; polygenic risk; screening; early cancer detection; family history; race/ethnicity; health disparities; Million Veteran Program

Abstract:

Background: Genetic scores may provide an objective measure of a man's risk of prostate cancer and thus inform screening decisions. We evaluated whether a polygenic hazard score based on 290 genetic variants (PHS290) is associated with risk of prostate cancer in a diverse population, including Black men, who have higher average risk of prostate cancer death but are often treated as a homogeneous, high-risk group

Methods: This was a retrospective analysis of Million Veteran Program (MVP), a national, population-based cohort study of United States military veterans conducted 2011-2021. Cox proportional hazards analyses tested for association of genetic and other risk factors (including self-reported race/ethnicity and family history) with age at death from prostate cancer, age at diagnosis of metastatic (nodal or distant) prostate cancer, and age at diagnosis of any prostate cancer.

Results: 590,750 male participants were included. Median age at last follow-up was 69 years. PHS290 was associated with fatal prostate cancer in the full cohort and for each racial/ethnic group ($p<10^{-10}$). Comparing men in the highest 20% of PHS290 to those in the lowest 20%, the hazard ratio for fatal prostate cancer was 4.42 [95% CI: 3.91-5.02]. When accounting for guideline-recommended risk factors (family history, race/ethnicity), PHS290 remained the strongest independent predictor of any, metastatic, and fatal prostate cancer.

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

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 numerous false positive results and 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]. Recent efforts have incorporated data from more diverse populations, yielding improved performance in these groups[19–23].

Our group has developed a 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[10,19,22]. 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[19,22,24]. 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) longitudinal cohort, one of the largest and most racially and ethnically diverse populations studied to date[25].

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 590,750 individuals of diverse self-reported race/ethnicity, over 100,000 of whom were Black or African American **(Table 1).** There were no inclusion or exclusion criteria for age. Median 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.

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 variants, enriched for low frequency variants in African

and Hispanic populations and variants associated with diseases common to the VA population[25]. The details on the quality control and imputation have been described previously[26].

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 (nodal and/or distant) 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[27]. 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.

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 variants 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[24]. Briefly, previously identified common variants associated with prostate cancer risk 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 for association with clinical endpoints 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[7,10,12,19]. Subgroup analyses with less than 100 events are reported in the Supplemental Material[28].

Cox Proportional Hazards Analysis

We evaluated association of PHS290 with age at diagnosis of prostate cancer and with two important clinical endpoints extracted from clinical data: age at diagnosis of nodal and/or distant metastases from prostate cancer and age at death from prostate cancer (i.e., lifetime prostate-cancer-specific mortality). To visualize the association in the full dataset, we generated cause-specific cumulative incidence curves for each endpoint and each of several PHS290 risk groups. Cox proportional hazards models were used to assess these associations in the full dataset and in each racial/ethnic group. Individuals not meeting the endpoint of interest were censored at age at last follow-up.

Effect sizes were estimated using hazard ratios (HRs) between risk strata, as described previously[7,9,10,12,19,20,22,23] and with previously defined thresholds for PHS290: 9.004659 (20th quantile), 9.123500 (30th quantile), 9.519703 (70th quantile), 9.639068 (80th quantile), 9.946332 (95th quantile)[24]. 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.

Race/Ethnicity, Family History, and PHS290

To assess the independent predictive value of PHS290 beyond commonly used clinical risk factors, we tested a multivariable Cox proportional hazards model with self-reported race/ethnicity, family history, and PHS290[7,9,22]. 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 any, metastatic, or fatal 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 racial/ethnic groups were estimated using Non-Hispanic White as the reference.

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

Genetic Ancestry and Genetic Principal Components

MVP participants have been assigned genetic ancestry groups based on previous analyses[29]. Briefly, a reference panel of 1000 Genomes Project[30] and Human Genome Diversity Project[31] individuals was constructed for preliminary PCA analysis. PC loadings of reference panel were projected onto PC loadings of MVP participant PC loadings, and assignments were made based on a random forest classifier. We repeated the prostate cancer analyses described above using genetic ancestry groups instead of self-reported race/ethnicity. We also repeated the above analyses when including the top 10 genetic principal components.

Results:

PHS290 Score

The distribution of PHS290 in Non-Hispanic Whites was similar to that reported previously for men of European ancestry (mean=9.37, SD=0.37)[24]. Mean PHS290 did vary by self-reported race/ethnicity with statistically significant differences between all groups (ANOVA $p < 10^{-16}$; all pair-wise *t*-tests $p < 10^{-14}$). 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 Asian men (mean=9.18, SD=0.35) and higher among Black men (mean=9.56, SD=0.34, t-test) (**Figure 1, Supplementary Figure 1**).

Association of PHS290 with Prostate Cancer

PHS290 was associated with age at diagnosis of prostate cancer, with age at development of prostate cancer nodal/distant metastases, and with age at death from prostate cancer **(Table 2)**. These associations also held in all racial/ethnic subgroup analyses with >100 events. Comparing 80th and 20th percentiles of genetic risk in the full dataset, men with higher PHS290 had an $HR_{80/20}$ of 5.20 [95% CI: 5.09-5.31] for any prostate cancer, $HR_{80/20}$ of 4.89 [95% CI: 4.57-5.21] for metastatic prostate cancer, and $HR_{80/20}$ of 4.42 [95% CI: 3.91-5.02] for fatal prostate cancer. Cause-specific cumulative incidence curves for various PHS290 percentile groups demonstrated risk stratification **(Figure 2)**. Consistent with prior reports, Black men had a higher average incidence of prostate cancer than Non-Hispanic White men. However, the incidence among Black men with low PHS290 was comparable to that of the average among Non-Hispanic White men **(Figure 2)**.

Race/Ethnicity, Family History, and PHS290

Race/ethnicity and family history were each associated with every clinical endpoint in univariable models **(Supplementary Tables 4, 5)**. Men with a family history of prostate cancer had a HR of 1.83 [1.53-2.17] for dying of prostate cancer. The race/ethnicity associations were largely driven by an increased risk for Black men. Compared to Non-Hispanic White men, Black men had a HR of 2.53 [2.14-2.92] for dying of prostate cancer.

PHS290 remained an independent predictor of prostate cancer risk—including prostate cancer death—when accounting for race/ethnicity and family history (**Table 3**). The multivariable model improved prediction for each clinical endpoint over the common risk factors alone (ANOVA $p < 2.2 \times 10^{-16}$). 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 3)**.

Genetic Ancestry and Genetic Principal Components

Results of the above analyses using genetic ancestry and the top 10 genetic principal components are reported in the Supplementary Material **(Supplementary Tables 2, 3, 6)**. PHS290 was significantly and independently associated with all three prostate cancer endpoints in each case.

Discussion:

PHS290 was associated with lifetime prostate-cancer-specific mortality in this large and diverse dataset. Even when accounting for current guideline-recommended risk factors (family history and 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 (nodal or distant) and with age at diagnosis of any prostate cancer. Metastatic prostate cancer has poor prognostic outcomes and is a major driver of pain, disability and aggressive medical therapy[32]. 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 racially/ethnically diverse independent validation of the association of a polygenic score with lifetime risk of fatal prostate cancer.

Black men 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[33–37]. National guidelines recommend consideration of prostate screening in men of African ancestry at a younger age and that screening occur at more frequent intervals[38]. The results of the present study confirm a generally great risk of prostate cancer among Black men but also demonstrate that Black men have variable levels of lifetime risk and should not be treated as a homogeneous group. Black men with low PHS290 had a prostate cancer risk comparable to average-risk Non-Hispanic White men, while Black men with high PHS290 had the highest risk of all subgroups. PHS290 can identify those more likely to develop lethal prostate cancer and may facilitate personalized screening recommendations.

Intriguingly, typical PHS290 scores differed between racial/ethnic groups, with the mean PHS290 slightly higher among Black men and slightly lower among East Asian men, compared to Hispanic and Non-Hispanic White men. These shifts in PHS290 distribution are consistent with reported differences in prostate cancer incidence across racial groups[39–43] and with a previous polygenic risk meta-analysis[21]. 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[44].

Family history is another important clinical consideration in prostate cancer screening decisions[38,45–48]. Prior studies have found polygenic scores to be the most important risk factor for prostate cancer (in men without known rare pathogenic mutations), with family history typically also independently predictive in multivariable models, possibly by capturing yet unknown genetic factors and/or shared familial environmental factors[7,9,22,49,50]. Among MVP participants, family history of prostate cancer was independently associated with prostate cancer risk in a multivariable model that included race/ethnicity and PHS290. The relationship of environmental exposures, family history, and prostate cancer risk merit further investigation[43], particularly in groups like veterans who may have been exposed to rare carcinogens[51].

The present study builds on prior work that reported the performance of polygenic scores in non-Europeans[20–22,24,52] and is consistent with those prior studies in showing a strong association of polygenic scores with prostate cancer risk, including death from prostate cancer[9,22,49]. 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[3,6]. Population-level analyses of benefit, harm, and cost-effectiveness support incorporation of genomic risk into screening[24]. The present study adds to the literature an independent validation in a dataset of over half a million men with diverse race/ethnicity and ancestry. Current clinical guidelines try to achieve targeted, or risk-stratified, screening by recommending each man discuss his individual risk factors, emphasizing race/ethnicity[38,46–48]. It is particularly important, therefore, that this study was able to combine race/ethnicity and genetic risk to estimate the relative impact of each and to demonstrate that a polygenic score adds considerable information beyond race/ethnicity 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. Smaller effect sizes were also seen when comparing the strength of association with age at diagnosis of prostate cancer within ancestry

groups[49,50]. 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 healthcare in a single US-based 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 variants 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.

The present work shows that adding PHS290 to guideline-recommended risk factors improves risk stratification for meaningful clinical endpoints of death or metastases from prostate cancer. Men at highest risk of metastatic or fatal prostate cancer are potentially those most likely to benefit from screening. Prior studies have further suggested genetic scores could also add value even after results from screening or diagnostic tests are already available, but this needs further investigation[53,54]. For example, one early detection strategy with strong evidence is early baseline PSA (e.g., at age 45-49)[55,56]. This strategy has not yet been widely adopted in the U.S.[57], but future studies should evaluate whether PHS290 adds value in men where early baseline PSA is known and whether PHS290, if known prior to PSA testing, should inform the decision of whether to obtain an early baseline PSA test. Another compelling avenue for future studies is whether high genetic risk can be mitigated by lifestyle or other preventive intervention[58,59].

Notwithstanding the possible advantages to risk-based screening, men should still be cautioned on the risks of screening: false positive PSA testing results and overdiagnosis of non-threatening cancer. While men with high PHS290 have a higher absolute incidence of prostate cancer metastasis or death, they also remain at risk of developing low-grade cancers likely to be detected with screening. Strategies to mitigate screening harms should be appropriately applied, including multiparametric MRI prior to biopsy and active surveillance for cancer diagnoses with favorable prognosis[38].

Limitations of this study include heterogeneity of prostate cancer screening and diagnostic pathways by clinicians across VA and other hospitals in the US that could potentially introduce noise, although this

heterogeneity likely leads to underestimation of associations with prostate cancer. The natural language processing tool used to identify men with metastatic disease does not reliably distinguish regional nodal from distant metastases, so these were all considered as one endpoint. Finally, we acknowledge that while we have used race/ethnicity, genetic ancestry, and principal genetic components, none of these groups can 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[22].

We show that PHS290 stratified US men for lifetime risk of any, metastatic, and fatal prostate cancer. Critically, this genetic risk stratification was successful within racial/ethnic subgroups in this diverse dataset. PHS290 was higher, on average, among Black men, who were also at higher risk from prostate cancer. The combination of race/ethnicity, family history, and PHS290 performed better than any single risk factor 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.

Notes:

Role of Funder:

Dr. Hauger was funded by the VISN-22 VA Center of Excellence for Stress and Mental Health (CESAMH) and National Institute of Aging RO1 grant AG050595 (*The VETSA Longitudinal Twin Study of Cognition and Aging VETSA 4).* This research was supported by VA MVP022. Meghana S. Pagadala was supported by the National Institutes of Health (#1F30CA247168, #T32CA067754). Tyler Seibert and Roshan Karunamuni were supported by the National Institutes of Health (NIH/NIBIB #K08EB026503), the Prostate Cancer Foundation, and the University of California (#C21CR2060).

Disclosures:

TMS reports honoraria from Varian Medical Systems and WebMD; he has an equity interest in CorTechs Labs, Inc. and serves on its Scientific Advisory Board; he has received in-kind research support from GE Healthcare via a research agreement with 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.

Author Contributions:

MSP and TMS wrote paper with assistance from JL, RK, PA, KL, FA, TA, HC, JG, GJ, RD, BR, MP, RH. JL, RK, KL, FA, TA, and RD contributed to phenotyping and MSP contributed to genetic analyses, RK and TMS contributed to original PHS290 score development

Acknowledgements:

This research used data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration. This research was supported by the Million Veteran Program MVP022 award # I01 CX001727 (PI: Richard L. Hauger MD). This publication does not represent the views of the Department of Veterans Affairs or the United States Government.

Data Availability Statement:

Requests regarding data access may be directed to MVPLOI@va.gov.

	All	Non-Hispanic White	Black or African American	Hispanic White	Asian	Native American	Pacific Islander	Other	Unknown
Participants									
All participants	590,750 (378,366)	420,473 (299,266)	102,203 (47,838)	27,651 (13,407)	6,644 (3,876)	5,835 (3,927)	3,246 (740)	8,226 (8,224)	16,472 (1,088)
Prostate cancer	69,137 (49,400)	48,339 (37,922)	15,748 (8,613)	2,120 (1,325)	376 (287)	504 (378)	236 (60)	720 (720)	1,094 (95)
Metastases from prostate cancer	6,413 (4,274)	4,299 (3,199)	1,604 (804)	213 (126)	33 (24)	48 (35)	18 (5)	68 (68)	130 (13)
Death from prostate cancer	1,858 (1,314)	1,354 (1,059)	363 (192)	59 (30)	5 (4)	10 (6)	4 (2)	19 (19)	44 (2)
Age demographics									
Age at diagnosis, median and IQR	67 [62-72]	68 [63-73]	63 [58-68]	66 [60-71]	67 [61-74]	65 [61-69]	65 [60-70]	64 [60-69]	67 [61-73]
Age at last follow-up, median and IQR	66 [59-74]	68 [62-75]	62 [55-70]	59 [47-71]	56 [41-70]	63 [56-72]	59 [49-70]	63 [57-71]	56 [42-70]

Table 1 Participant characteristics for self-reported race/ethnicity groups, n=590,750. Numbers in parentheses indicate participants with family history information also.

			Hazard ratios			
Group	**Clinical Endpoint**	**p**	**$HR_{80/20}$**	**$HR_{20/50}$**	**$HR_{80/50}$**	**$HR_{95/50}$**
All (n=590,750)	**Fatal Prostate Cancer**	$<10^{-16}$	4.42 [3.91-5.02]	0.48 [0.45-0.51]	2.12 [2.0-2.27]	3.0 [2.75-3.3]
Non-Hispanic White (n=420,473)		$<10^{-16}$	4.37 [3.77-5.05]	0.48 [0.45-0.52]	2.11 [1.96-2.27]	3.0 [2.69-3.34]
Black or African American (n=102,203)		$<10^{-16}$	2.37 [1.73-3.29]	0.66 [0.56-0.77]	1.57 [1.33-1.85]	1.9 [1.5-2.42]
Hispanic White (n=27,651)		-	-	-	-	-
Asian (n=6,644)		-	-	-	-	-
Native American (n=5,835)		-	-	-	-	-
Pacific Islander (n=3,246)		-	-	-	-	-
Unknown (n=16,472)		-	-	-	-	-
Other (n=8,226)		-	-	-	-	-
All (n=590,750)	**Metastatic Prostate Cancer**	$<10^{-16}$	4.89 [4.57-5.21]	0.46 [0.44-0.47]	2.23 [2.16-2.31]	3.23 [3.07-3.39]
Non-Hispanic White (n=420,473)		$<10^{-16}$	4.62 [4.27-5.01]	0.47 [0.45-0.49]	2.17 [2.08-2.26]	3.13 [2.95-3.32]
Black or African American (n=102,203)		$<10^{-16}$	3.03 [2.62-3.51]	0.59 [0.55-0.63]	1.78 [1.65-1.92]	2.28 [2.05-2.54]
Hispanic White (n=27,651)		$<10^{-7}$	2.92 [1.96-4.43]	0.58 [0.47-0.71]	1.71 [1.4-2.11]	2.22 [1.65-3.01]
Asian (n=6,644)		-	-	-	-	-
Native American (n=5,835)		-	-	-	-	-
Pacific Islander (n=3,246)		-	-	-	-	-
Unknown (n=16,472)		-	-	-	-	-
Other (n=8,226)		-	-	-	-	-
All (n=590,750)	**Prostate Cancer**	$<10^{-16}$	5.2 [5.09-5.31]	0.44 [0.44-0.45]	2.31 [2.28-2.33]	3.39 [3.33-3.44]
Non-Hispanic White (n=420,473)		$<10^{-16}$	5.12 [4.99-5.26]	0.45 [0.44-0.45]	2.29 [2.26-2.32]	3.38 [3.31-3.44]
Black or African American (n=102,203)		$<10^{-16}$	3.22 [3.06-3.39]	0.57 [0.56-0.58]	1.83 [1.79-1.88]	2.39 [2.3-2.48]
Hispanic White (n=27,651)		$<10^{-16}$	4.24 [3.74-4.85]	0.49 [0.45-0.52]	2.06 [1.93-2.2]	2.92 [2.66-3.22]
Asian (n=6,644)		$<10^{-16}$	5.32 [4.04-7.09]	0.43 [0.37-0.49]	2.27 [1.98-2.61]	3.49 [2.83-4.34]

Native American (n=5,835)		$<10^{-16}$	5.1 [4.04-6.55]	0.45 [0.4-0.51]	2.31 [2.05-2.62]	3.4 [2.85-4.12]
Pacific Islander (n=3,246)		$<10^{-10}$	3.58 [2.5-5.37]	0.53 [0.43-0.63]	1.9 [1.57-2.33]	2.6 [1.98-3.57]
Unknown (n=16,472)		$<10^{-16}$	5.27 [4.39-6.3]	0.44 [0.4-0.48]	2.32 [2.12-2.55]	3.45 [3.0-3.94]
Other (n=8,226)		$<10^{-16}$	5.52 [4.45-6.89]	0.43 [0.39-0.48]	2.36 [2.12-2.64]	3.59 [3.06-4.25]

Table 2: Association of PHS290 with any, metastatic and fatal prostate cancer. Cox Proportional Hazards model results from association with age at prostate cancer, metastatic prostate cancer (nodal or distant) and death from prostate cancer. *P*-values reported are from univariable models using PHS290 as the sole predictor variable. Hazard ratios (HRs) compare 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. For subgroup analyses with less than 100 events of the endpoint, the box is marked '-'; these statistically less reliable results are reported in Table S2.

Clinical Endpoints	PHS290	Family History	Race/ Ethnicity						
	$HR_{80/20}$	Family History	Black or African American	Hispanic White	Asian	Native American	Other	Pacific Islander	Unknown
Fatal Prostate Cancer	4.17 [3.59-4.88]***	1.67 [1.4-1.96]*	1.97 [1.69-2.31]***	1.04 [0.69-1.42]	0.68 [0.15-1.43]	0.98 [0.32-1.85]	1.9 [1.12-2.81]*	1.74 [0.0-4.43]	0.73 [0.0-1.95]
Metastatic Prostate Cancer	4.15 [3.81-4.53]***	1.53 [1.38-1.68]***	2.24 [2.07-2.42]***	1.32 [1.1-1.57]*	1.31 [0.81-1.85]	1.4 [0.97-1.86]	1.67 [1.25-2.08]*	1.14 [0.23-2.29]	1.52 [0.79-2.45]
Prostate Cancer	4.69 [4.57-4.81]***	1.78 [1.73-1.83]***	1.83 [1.78-1.87]***	1.1 [1.04-1.16]*	1.28 [1.13-1.43]*	1.04 [0.95-1.15]	1.2 [1.1-1.29]*	0.98 [0.73-1.25]	0.9 [0.72-1.08]

Table 3: Multivariable models combining self-reported Race/Ethnicity, Family History (FH), and PHS290 for three prostate cancer clinical endpoints. Cox proportional hazards results for association with age at death from prostate cancer, at diagnosis of metastatic prostate cancer, and age at diagnosis with prostate cancer. *P*-values reported are from multivariable models using self-reported race/ethnicity, family history, and PHS290. 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 race/ethnicity were estimated using Non-Hispanic White as the reference. Hazard ratios for family history were for one or more first-degree relatives diagnosed with prostate cancer. This multivariable analysis was limited to the 375,763 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 *** ($p<10^{-16}$).

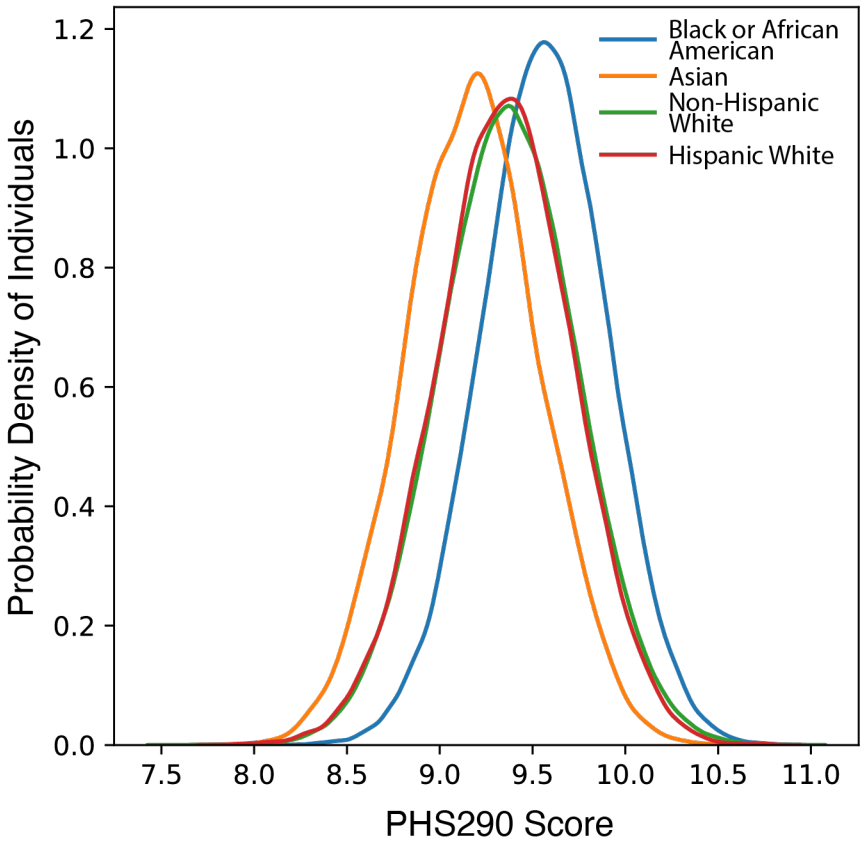

Figure 1. PHS290 Score Density Plots in Million Veteran Program. PHS290 score density plot in select self-reported race/ethnicity groups.

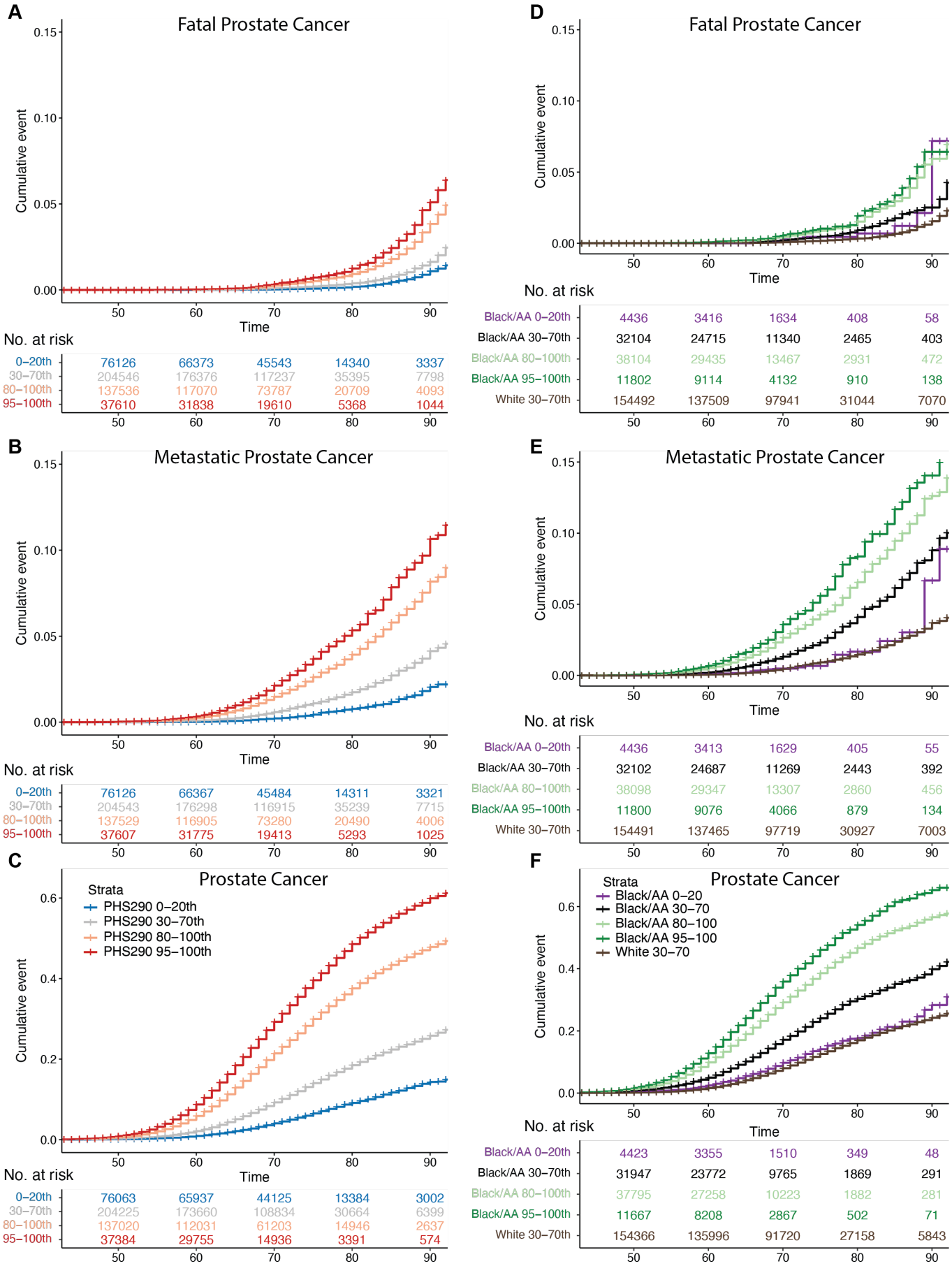
A
Fatal Prostate Cancer
Cumulative event
0.15
0.10
0.05
0.00
50
60
70
80
90
Time
No. at risk
0–20th 76126 66373 45543 14340 3337
30–70th 204546 176376 117237 35395 7798
80–100th 137536 117070 73787 20709 4093
95–100th 37610 31838 19610 5368 1044
B
Metastatic Prostate Cancer
Cumulative event
0.15
0.10
0.05
0.00
50
60
70
80
90
Time
No. at risk
0–20th 76126 66367 45484 14311 3321
30–70th 204543 176298 116915 35239 7715
80–100th 137529 116905 73280 20490 4006
95–100th 37607 31775 19413 5293 1025
C
Prostate Cancer
Strata
PHS290 0–20th
PHS290 30–70th
PHS290 80–100th
PHS290 95–100th
Cumulative event
0.6
0.4
0.2
0.0
50
60
70
80
90
Time
No. at risk
0–20th 76063 65937 44125 13384 3002
30–70th 204225 173660 108834 30664 6399
80–100th 137020 112031 61203 14946 2637
95–100th 37384 29755 14936 3391 574
D
Fatal Prostate Cancer
Cumulative event
0.15
0.10
0.05
0.00
50
60
70
80
90
Time
No. at risk
Black/AA 0–20th 4436 3416 1634 408 58
Black/AA 30–70th 32104 24715 11340 2465 403
Black/AA 80–100th 38104 29435 13467 2931 472
Black/AA 95–100th 11802 9114 4132 910 138
White 30–70th 154492 137509 97941 31044 7070
E
Metastatic Prostate Cancer
Cumulative event
0.15
0.10
0.05
0.00
50
60
70
80
90
Time
No. at risk
Black/AA 0–20th 4436 3413 1629 405 55
Black/AA 30–70th 32102 24687 11269 2443 392
Black/AA 80–100th 38098 29347 13307 2860 456
Black/AA 95–100th 11800 9076 4066 879 134
White 30–70th 154491 137465 97719 30927 7003
F
Prostate Cancer
Strata
Black/AA 0–20
Black/AA 30–70
Black/AA 80–100
Black/AA 95–100
White 30–70
Cumulative event
0.6
0.4
0.2
0.0
50
60
70
80
90
Time
No. at risk
Black/AA 0–20th 4423 3355 1510 349 48
Black/AA 30–70th 31947 23772 9765 1869 291
Black/AA 80–100th 37795 27258 10223 1882 281
Black/AA 95–100th 11667 8208 2867 502 71
White 30–70th 154366 135996 91720 27158 5843

Figure 2. Million Veteran Program (MVP) Cause-specific Cumulative Incidence. Cause-specific cumulative incidence within MVP, stratified by PHS290, for **(A)** fatal prostate cancer, **(B)** metastatic prostate cancer, and **(C)** prostate cancer. PHS290 percentile groups shown for each endpoint: 0-20th, 30-70th, 80-100th, and 95-100th. Cumulative incidence for Black or African American (Black/AA) men in several PHS290 strata compared with average-risk Non-Hispanic White (PHS290 30-70th percentiles) for **(D)** fatal prostate cancer, **(E)** metastatic prostate cancer, and **(F)** prostate cancer. The *y*-axis scale was adjusted for (C) and (E) to show the higher incidence values for any prostate cancer.

References:

1. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19(9):581-590.

2. Schröder FH, Hugosson J, Roobol MJ, 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*. 2014;384(9959):2027-2035.

3. Callender T, Emberton M, Morris S, et al. Polygenic risk-tailored screening for prostate cancer: A benefit-harm and cost-effectiveness modelling study. *PLoS Med*. 2019;16(12):e1002998.

4. Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol*. 2014;65(6):1046-1055.

5. Mucci LA, Hjelmborg JB, Harris JR, et al. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA*. 2016;315(1):68-76.

6. Callender T, Emberton M, Morris S, Pharoah PDP, 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*. 2021;4(3):e2037657.

7. Seibert TM, Fan CC, Wang Y, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ*. 2018;360:j5757.

8. Schumacher FR, Olama AAA, Berndt SI, et al. Author Correction: Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2019;51(2):363.

9. Huynh-Le MP, Karunamuni R, Fan CC, et al. Common genetic and clinical risk factors: association with fatal prostate cancer in the Cohort of Swedish Men. *Prostate Cancer Prostatic Dis*. Published online March 15, 2021. doi:10.1038/s41391-021-00341-4

10. Karunamuni RA, Huynh-Le MP, Fan CC, et al. Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. *Prostate Cancer Prostatic Dis*. 2021;24(2):532-541.

11. Huynh-Le MP, Myklebust TÅ, Feng CH, et al. Age dependence of modern clinical risk groups for localized prostate cancer-A population-based study. *Cancer*. 2020;126(8):1691-1699.

12. Huynh-Le MP, Fan CC, Karunamuni R, et al. A genetic risk score to personalize Prostate Cancer screening, applied to population data. *Cancer Epidemiol Biomarkers Prev*. 2020;29(9):1731-1738.

13. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019;10(1):3328.

14. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol*. 2016;17(1). doi:10.1186/s13059-016-1016-y

15. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584-591.

16. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161-164.

17. DeSantis CE, Miller KD, Goding Sauer A, Jemal A, Siegel RL. Cancer statistics for African Americans, 2019. *CA Cancer J Clin*. 2019;69(3):211-233.

18. Tsodikov A, Gulati R, de Carvalho TM, et al. Is prostate cancer different in black men? Answers from 3 natural history models. *Cancer*. 2017;123(12):2312-2319.

19. Karunamuni RA, Huynh-Le MP, Fan CC, et al. African-specific improvement of a polygenic hazard score for age at diagnosis of prostate cancer. *Int J Cancer*. 2021;148(1):99-105.

20. Karunamuni RA, Huynh-Le MP, Fan CC, et al. Performance of African-ancestry-specific polygenic hazard score varies according to local ancestry in 8q24. *Prostate Cancer Prostatic Dis*. Published online June 14, 2021. doi:10.1038/s41391-021-00403-7

21. Conti DV, Darst BF, Moss LC, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet*. 2021;53(1):65-75.

22. Huynh-Le MP, Fan CC, Karunamuni R, et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. *Nat Commun*. 2021;12(1):1236.

23. Huynh-Le MP, Karunamuni R, Fan CC, et al. Prostate cancer risk stratification improvement across multiple ancestries with new polygenic hazard score. *Prostate Cancer Prostatic Dis*. Published online February 12, 2022. doi:10.1038/s41391-022-00497-7

24. Huynh-Le MP, Karunamuni R, Fan CC, et al. Prostate cancer risk stratification improved across multiple ancestries with new polygenic hazard score. Published online August 18, 2021. doi:10.1101/2021.08.14.21261931

25. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016;70:214-223.

26. Hunter-Zinck H, Shi Y, Li M, et al. Genotyping array design and data quality control in the Million Veteran Program. *Am J Hum Genet*. 2020;106(4):535-548.

27. Alba PR, Gao A, Lee KM, et al. Ascertainment of Veterans with Metastatic Prostate Cancer in Electronic Health Records: Demonstrating the Case for Natural Language Processing. *JCO Clin Cancer Inform*. 2021;5:1005-1014.

28. Karunamuni RA, Huynh-Le MP, Fan CC, et al. The effect of sample size on polygenic hazard models for prostate cancer. *Eur J Hum Genet*. 2020;28(10):1467-1475.

29. Wendt FR, Pathak GA, Vahey J, et al. Modeling the longitudinal changes of ancestry diversity in the Million Veteran Program. *bioRxiv*. Published online January 25, 2022. doi:10.1101/2022.01.24.477583

30. The 1000 Genomes Project Consortium, Auton A, Abecasis GR, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

31. Li JZ, Absher DM, Tang H, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008;319(5866):1100-1104.

32. Sartor O, de Bono JS. Metastatic prostate cancer. *N Engl J Med*. 2018;378(7):645-657.

33. Barocas DA, Grubb R 3rd, Black 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*. 2013;119(12):2223-2229.

34. Han Y, Rand KA, Hazelett DJ, et al. Prostate cancer susceptibility in men of African ancestry at 8q24. *J Natl Cancer Inst*. 2016;108(7):djv431.

35. Mahal BA, Aizer AA, Ziehr DR, et al. Trends in disparate treatment of African American men with localized prostate cancer across National Comprehensive Cancer Network risk groups. *Urology*. 2014;84(2):386-392.

36. Yamoah K, Johnson MH, Choeurng V, et al. Novel biomarker signature that may predict aggressive disease in African American men with prostate cancer. *J Clin Oncol*. 2015;33(25):2789-2796.

37. Zhang H, Messing EM, Travis LB, 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*. 2013;3:312.

38. Moses KA, Sprenkle PC, Box G, et al. NCCN Guidelines Prostate Cancer Early Detection Version 1.2022. Published online 2021.

39. Siegel DA, O'Neil ME, Richards TB, Dowling NF, Weir HK. Prostate cancer incidence and survival, by stage and race/ethnicity - United States, 2001-2017. *MMWR Morb Mortal Wkly Rep*. 2020;69(41):1473-1480.

40. Cook MB, Rosenberg PS, McCarty FA, et al. Racial disparities in prostate cancer incidence rates by census division in the United States, 1999-2008. *Prostate*. 2015;75(7):758-763.

41. Atlas of cancer mortality in the United States, 1950-94. *Int J Epidemiol*. 2000;29(3):602-a-602.

42. Jemal A, Ward E, Wu X, Martin HJ, McLaughlin CC, Thun MJ. Geographic patterns of prostate cancer mortality and variations in access to medical care in the United States. *Cancer Epidemiol Biomarkers Prev*. 2005;14(3):590-595.

43. Curry SJ, Krist AH, Owens DK. Annual report to the nation on the status of cancer, part II: Recent changes in prostate cancer trends and disease characteristics. *Cancer*. 2019;125(2):317-318.

44. Marnetto D, Pärna K, Läll K, et al. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat Commun*. 2020;11(1):1628.

45. Powell IJ. The precise role of ethnicity and family history on aggressive prostate cancer: a review analysis. *Arch Esp Urol*. 2011;64(8):711-719.

46. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin*. 2010;60(2):70-98.

47. Horwich A, Hugosson J, de Reijke T, et al. Prostate cancer: ESMO Consensus Conference Guidelines 2012. *Ann Oncol*. 2013;24(5):1141-1162.

48. Qaseem A, Barry MJ, Denberg TD, Owens DK, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med*. 2013;158(10):761-769.

49. Shi Z, Platz EA, Wei J, et al. Performance of three inherited risk measures for predicting prostate cancer incidence and mortality: A population-based prospective analysis. *Eur Urol*. 2021;79(3):419-426.

50. Kachuri L, Graff RE, Smith-Byrne K, et al. Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nat Commun*. 2020;11(1):6084.

51. Schecter A, Needham L, Pavuk M, et al. Agent Orange exposure, Vietnam war veterans, and the risk of prostate cancer. *Cancer*. 2009;115(14):3369-3371.

52. Plym A, Penney KL, Kalia S, et al. Evaluation of a multiethnic polygenic risk score model for prostate cancer. *J Natl Cancer Inst*. Published online April 1, 2021. doi:10.1093/jnci/djab058

53. Eklund M, Jäderling F, Discacciati A, et al. MRI-targeted or standard biopsy in prostate cancer screening. *N Engl J Med*. 2021;385(10):908-920.

54. Jiang Y, Meyers TJ, Emeka AA, et al. Genetic factors associated with prostate cancer conversion from active surveillance to treatment. *HGG Adv*. 2022;3(1):100070.

55. Vickers AJ, Ulmert D, Sjoberg DD, et al. Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. *BMJ*. 2013;346(apr15 5):f2023.

56. Lilja H, Ulmert D, Björk T, et al. Long-term prediction of prostate cancer up to 25 years before diagnosis of prostate cancer using prostate kallikreins measured at age 44 to 50 years. *J Clin Oncol*. 2007;25(4):431-436.

57. US Preventive Services Task Force, Grossman DC, Curry SJ, et al. Screening for prostate cancer. *JAMA*. 2018;319(18):1901.

58. Plym A, Zhang Y, Stopsack K, et al. Abstract 822: Can the genetic risk of prostate cancer be attenuated by a healthy lifestyle. In: *Epidemiology*. American Association for Cancer Research; 2021. doi:10.1158/1538-7445.am2021-822

59. Ahmed M, Goh C, Saunders E, et al. Germline genetic variation in prostate susceptibility does not predict outcomes in the chemoprevention trials PCPT and SELECT. *Prostate Cancer Prostatic Dis*. 2020;23(2):333-342.