

Polygenic risk for prostate cancer: Decreasing relative risk with age but little impact on absolute risk

Authors

Daniel J. Schaid, Jason P. Sinnwell,
Anthony Batzler, Shannon K. McDonnell

Correspondence

schaid@mayo.edu



Polygenic risk for prostate cancer: Decreasing relative risk with age but little impact on absolute risk

Daniel J. Schaid,^{1,*} Jason P. Sinnwell,¹ Anthony Batzler,¹ and Shannon K. McDonnell¹

Summary

Polygenic risk scores (PRSs) for a variety of diseases have recently been shown to have relative risks that depend on age, and genetic relative risks decrease with increasing age. A refined understanding of the age dependency of PRSs for a disease is important for personalized risk predictions and risk stratification. To further evaluate how the PRS relative risk for prostate cancer depends on age, we refined analyses for a validated PRS for prostate cancer by using 64,274 prostate cancer cases and 46,432 controls of diverse ancestry (82.8% European, 9.8% African American, 3.8% Latino, 2.8% Asian, and 0.8% Ghanaian). Our strategy applied a novel weighted proportional hazards model to case-control data to fully utilize age to refine how the relative risk decreased with age. We found significantly greater relative risks for younger men (age 30–55 years) compared with older men (70–88 years) for both relative risk per standard deviation of the PRS and dichotomized according to the upper 90th percentile of the PRS distribution. For the largest European ancestral group that could provide reliable resolution, the log-relative risk decreased approximately linearly from age 50 to age 75. Despite strong evidence of age-dependent genetic relative risk, our results suggest that absolute risk predictions differed little from predictions that assumed a constant relative risk over ages, from short-term to long-term predictions, simplifying implementation of risk discussions into clinical practice.

Introduction

Polygenic risk scores (PRSs), also called genomic risk scores, provide a single measure of a large number of genetic variants associated with common diseases and have potential to improve personalized medical care and public health by informing subjects of their future risk of developing disease. Because common diseases increase with age with increasing impact of lifetime exposures, it is critical to evaluate whether the association of a PRS with disease changes with age and the practical implications of ignoring age-dependent risks. As stressed by others,¹ understanding the age dependency of PRSs for a disease is important not only for personalized medical care and population health but also to improve understanding of disease etiology.

A PRS for an individual is a weighted sum over the doses of selected risk variants, on the order of hundreds to millions of genetic variants,² and so creation of a PRS depends on which variants are chosen and how weights are assigned. A variety of methods to create a PRS have been developed,^{3–13} many of which result in a large number of selected variants. The purpose of this report is to propose a strategy to evaluate a PRS for clinical risk predictions by determining whether the relative risk for a PRS depends on age and whether age-dependent relative risks have practical implications. Our empirical evaluations are based on a PRS for prostate cancer that was developed on a large number of men of diverse ancestry and has been replicated. Hence, our starting point is based on a chosen PRS and not development of a new PRS.

Conti et al.¹⁴ developed a PRS for prostate cancer based on a multi-ancestry meta-analysis of genome-wide associa-

tion summary statistics from a total of 107,247 cases and 127,006 controls with European, African, East Asian, and Hispanic ancestries. They discovered a total of 269 variants associated with prostate cancer risk and constructed a PRS by using multi-ancestry weights. Although 269 variants might seem like a small number for a PRS, these variants were selected on the basis of stepwise selection to determine independent associations within each genomic region as well as fine-mapping with joint analysis of marginal summary statistics (JAM)¹⁵ to determine population-specific variants that were independently associated with prostate cancer. This PRS was validated in an independent study of 13,628 US men.¹⁶

Conti et al.¹⁴ found that men with prostate cancer in the top 10% of the PRS distribution were diagnosed 2.84 years younger than men in the bottom 10% of the distribution. Others have also reported that larger values of PRS based on 110 genetic variants were associated with younger age of prostate cancer diagnosis for white men.¹⁷ These observations raise concerns regarding whether statistical models to predict prostate cancer should allow for PRS relative risks to depend on age. A complication is that even if the relative risk for a PRS is constant over all ages, the men with highest risk will succumb to disease at earlier ages, resulting in observations that men with larger values of PRS tend to be diagnosed at younger ages. See [supplemental information](#) for theoretical derivations and [Figure S1](#) for numerical illustration. Hence, because odds ratios and relative risks are used to predict future risk of disease, it is important to evaluate whether these risk parameters change with age. Nonetheless, Conti et al.¹⁴ found that among men of European ancestry, those with PRSs in the top decile of the PRS distribution had an

¹Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA

*Correspondence: schaid@mayo.edu

<https://doi.org/10.1016/j.ajhg.2022.03.008>

© 2022 American Society of Human Genetics.

odds ratio of 6.71 (95% CI, 5.99–7.52), compared with PRSs in the 40th–60th percentile, for men aged ≤ 55 years in contrast to a smaller odds ratio of 4.39 (95% CI, 4.19–4.60) for men older than 55 years.

The observation of weakening association of PRS with disease risk as age increases has been observed for breast cancer¹⁸ and for cardiovascular disease, particularly when many non-genetic risk factors are known to have effects at older ages.¹⁹ Furthermore, by a study of genetic relative risk for 24 common diseases within the British ancestry subset of the UK Biobank, Jiang²⁰ found evidence for age-varying relative risk for hypertension, skin cancer, atherosclerotic heart disease, hypothyroidism, and calculus of gallbladder. The predominant pattern was genetic risk largest at younger ages, and relative risk decreased as age increased. Because risk due to a PRS can change with age, it is important to understand the full impact of age on risk predictions, including future absolute risk of disease, conditional on current age and PRS.

In addition to the influence of age, it is critical to consider the influence of ancestry. Large genome-wide association studies (GWASs) are needed to determine the most relevant genetic variants and their weights, and many GWASs have focused on European ancestry.²¹ The work by Conti et al.¹⁴ attempted to overcome this limitation by gathering as much GWAS summary statistics as possible across different ancestries. They found that the distribution of PRSs varied across different ancestral populations, even for controls. This is expected when allele frequencies of the variants in the PRSs differ across different populations. It can be shown that the distribution of the PRSs in controls is a normal distribution with mean μ and variance σ^2 that both depend on the allele frequencies and the PRS weights.¹⁸ Furthermore, if the risk of disease is due to a large number of alleles of small effect, combining multiplicatively, the distribution among cases is also a normal distribution with the same variance as for controls, but with a mean among cases that is approximately $\mu + \sigma^2$, illustrating that the distribution among cases is shifted to larger values. Because of these theoretical expectations and empirical data that support these expectations, it is critical to account for not only the association of PRSs with prostate cancer in different populations but also the difference in distribution of PRSs across different populations.

In summary, multiple factors complicate the modeling of the effect of PRSs on prostate cancer risk: the population distribution of PRSs, which depends on ancestry; the influence of PRSs on prostate cancer risk, which depends on age; and family history of prostate cancer, which can be confounded with age of diagnosis. Men with a family history of prostate cancer tend to have a younger age of diagnosis,¹⁷ and a younger age of diagnosis has been reported when a close relative had prostate cancer.^{22,23} In this report, we refined the analyses reported by Conti et al. by a more extensive evaluation of age, beyond the dichotomy of ≤ 55 years versus >55 years, and adjusting for the differences in distribution of PRSs across different ancestries.

Furthermore, we evaluated the role of family history in addition to effects of age. Finally, following the recommendation to convey absolute risks to lay people in order to simplify interpretation of personal risks,^{24,25} we evaluated the impact of a PRS relative risk changing over ages on predicting the future absolute risk of prostate cancer.

Material and methods

Studies

Prostate cancer case-control GWASs were obtained from dbGaP after approval of project request # 25202 “Development and Testing of Polygenic Risk Scores for Prostate Cancer.” All data were de-identified, and by dbGaP policy, no review by an institutional review board was necessary. The seven case-controls studies are illustrated in Table S1 and described detail in the supplemental information. Advantages of these studies are their large size and diverse ancestry. Note that some of the studies were used to develop the PRS by Conti et al.,¹⁴ so our results should not be viewed as an independent validation of the original PRS.

Genotype quality control and imputation

Genotype quality control (QC) prior to imputation was conducted separately for each study and each genotyping platform. We removed SNPs with a call rate $< 98\%$, indels, duplicate SNPs, or monomorphic SNPs and men with a call rate $< 95\%$. SNPs were also excluded if they failed Hardy-Weinberg equilibration (HWE) test p value $< 10^{-6}$. Because admixture of different ancestries can influence tests of HWE, we applied the software ADMIXTURE²⁶ to the genetic data to classify men into major ancestral groups (European, African, Amerindian, East Asian, South Asian) and then tested HWE within major ancestral groups. Genetic sex was verified by PLINK with markers on the X and Y chromosomes, and subjects that were not consistent with male were excluded. Samples were removed if they displayed a call rate $< 80\%$ on any given chromosome or if they had unusually low heterozygosity ratio < 0.4 (observed/expected heterozygosity) on any chromosome, presuming poor quality genotype data that would unduly influence imputation. The relatedness between each pair of men was evaluated by estimation of the kinship coefficient via King robust²⁷ that is implemented in the R package *SNPRelate*. We randomly removed one subject from each strongly related pair (i.e., duplicates, parent-offspring, full siblings, and third-degree relatives, with an estimated kinship coefficient at least 0.0442). This approach allowed us to identify men whose samples were included in more than one study and remove duplicates.

After the above QC processing, samples were uploaded to the TOPMed imputation server,^{28,29} where additional QC steps were completed including removal of multi-allelic SNPs, removal of indels, removal of monomorphic SNPs, and removal of SNPs with large allele frequency differences compared with the TOPMed Imputation Reference panel. Imputed variants with an imputation $R^2 \geq 0.3$ were retained.

Data harmonization

The Breast and Prostate Cancer Cohort Consortium (BPC3) data included both men and women; women were excluded from our analyses. The set of variables that were common across all dbGaP studies were case-control status, ancestry, family history of prostate cancer, and age (age of disease diagnosis for cases and age of study

Table 1. Characteristics of men by case-control status

	Case (No. = 64,274)	Control (No. = 46,432)	Total (No. = 110,706)
Ancestry group			
African American	5,505 (8.6%)	5,370 (11.6%)	10,875 (9.8%)
Asian	1,574 (2.4%)	1,483 (3.2%)	3,057 (2.8%)
European	54,564 (84.9%)	37,059 (79.8%)	91,623 (82.8%)
Ghanaian	461 (0.7%)	452 (1.0%)	913 (0.8%)
Latino	2,170 (3.4%)	2,068 (4.5%)	4,238 (3.8%)
Family history of prostate cancer			
Unknown	25,094	16,016	41,110
No	29,206 (74.5%)	27,189 (89.4%)	56,395 (81.0%)
Yes	9,974 (25.5%)	3,227 (10.6%)	13,201 (19.0%)
Age group, years			
[30,45)	330 (0.5%)	991 (2.1%)	1,321 (1.2%)
[45,50)	1,290 (2.0%)	1,707 (3.7%)	2,997 (2.7%)
[50,55)	4,288 (6.7%)	4,945 (10.6%)	9,233 (8.3%)
[55,60)	10,122 (15.7%)	8,740 (18.8%)	18,862 (17.0%)
[60,65)	12,586 (19.6%)	10,840 (23.3%)	23,426 (21.2%)
[65,70)	15,769 (24.5%)	9,635 (20.8%)	25,404 (22.9%)
[70,75)	11,335 (17.6%)	5,631 (12.1%)	16,966 (15.3%)
[75,88)	8,554 (13.3%)	3,943 (8.5%)	12,497 (11.3%)
Age, year: median (range)	66 (30–87)	62 (30–87)	64 (30–87)

Excludes men with missing age, age < 30 years, or missing ancestry.

enrollment for controls). Age was recorded differently across studies. Some studies recorded exact age, others recorded in 5-year intervals, and others recorded in 10-year intervals. To determine a common set of intervals, 10-year intervals were recoded as the last 5 years of an interval. For example, 50–59 was recoded to 55–59. For analyses that used yearly ages, we used the mid-point of an age interval when exact age was missing.

Polygenic risk score

Conti et al.¹⁴ developed a trans-ethnic PRS based on 269 SNPs and their associations with prostate cancer across four ancestries: European, African, East Asian, and Hispanic. The variants and weights used to create the PRS are available from their Table S4 and also at <https://www.pgscatalog.org/publication/PGP000122/>. After QC and imputation in our data, there were 220 variants that overlapped with our imputed data. Because the distribution of PRSs differs across different ancestries, we evaluated different approaches to correct for ancestry (see [supplemental information](#)) and chose to center and scale the PRS within each ancestry group by using the mean and standard deviation for controls within each ancestry group.

Age-specific incidence rates for prostate cancer and for death

We used age-specific prostate cancer incidence (hazard) rates to create weights for Cox proportional hazards models as well as

combine with death hazard rates to compute absolute risks. Prostate cancer incidence rates were downloaded from the CDC US Cancer Statistics (https://www.cdc.gov/cancer/uscs/dataviz/download_data.htm). The cancer rates were reported for 5-year intervals, and we used linear interpolation to determine the rate at each year of age from age 30 to 87 years. CDC rates were available for ancestries of US White, Latino, African American, and Asian. We used the incidence rates of US African American men for men from Ghana. [Figure S3](#) in the [supplemental information](#) illustrates how the incidence rates and cumulative risk for prostate cancer vary over ages for different ancestries.

Death incidence rates were obtained from CDC Health Statistics (https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/NVSR/61_03/) with documentation from https://www.cdc.gov/nchs/products/life_tables.htm. The life tables for 2008 were for each sex and race (White, non-Hispanic White, Black, non-Hispanic Black, and Hispanic). Based on life analytic methods, the death hazard rate at age t , h_t was calculated according to $h_t = q_t / (1 - q_t / 2)$, where q_t is the life table probability of dying between ages t and $t + 1$. The division by 2 assumes deaths on average occur mid-way during each year.³⁰ We used non-Hispanic White death rates for Asian ancestry.

Statistical methods

Statistical analyses were conducted with R version 4.0.3. The associations between PRS and age of disease diagnosis were estimated by weighted Cox regression models. Although logistic regression is typically used to analyze case-control studies, the resulting odds ratios are an approximation of the relative risk estimated in studies when disease is rare, yet odds ratios over-estimate the relative risk,³¹ which impacts models to estimate absolute risk. Furthermore, additional information about the risk of disease at different ages is potentially available from the age of diagnosis, and the age at which controls are free of disease. Studies have shown that applying Cox regression models to case-control data, using age information, can lead to greater power than logistic regression,³² although a naive analysis that fails to account for over-sampling of cases in case-control studies can lead to biased estimates of the relative risks.³³ For these reasons, we estimated relative risks by use of the Cox model with sampling weights based on population incidence rates to account for how cases and controls were sampled.^{34–36} We assigned weights of 1 to cases and weights of $1/inc_t$ to controls, where inc_t is the age-specific incidence rate of prostate cancer in a defined ancestry group. In addition, we used the *survSplit* function in the *survival* package to fit piece-wise proportional hazards models to allow the relative risks to differ across different age categories, as well as the time-transform functions of *coxph* to model continuous time-dependent coefficients. Methods to compute the future absolute risk of disease, conditional on a man alive and free of disease at a specified age and with a standardized PRS, are described in [Appendix A](#).

Results

The cases and controls included in analyses are characterized in [Table 1](#). Details about how samples were evaluated for quality of genetic results, how genetically related men were removed, and selection criteria to exclude men with missing age (N = 2,800), missing ancestry (N = 27), or young age < 30 years (N = 39) and characteristics of men

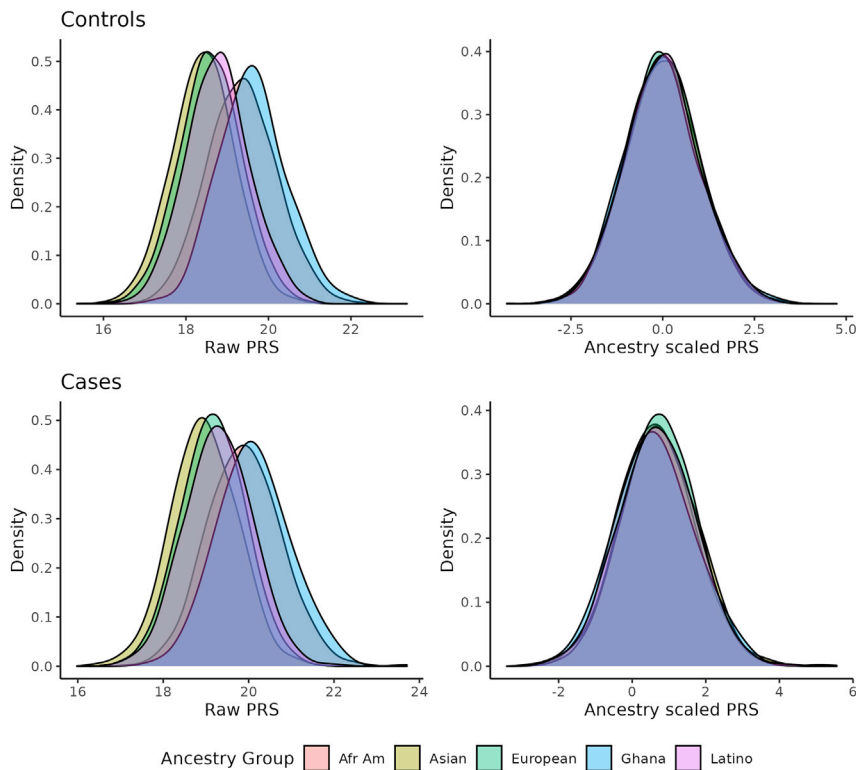


Figure 1. Distribution of PRSs by ancestry and disease status before and after adjustment for ancestry by centering and scaling according to mean and standard deviation within controls of each ancestry

that the distributions were centered at zero for controls, as expected, while the distributions for the cases were shifted to greater values.

The association of the standardized PRS with age of onset of prostate cancer in the pool of all data, assessed by a weighted Cox proportional hazards model assuming a constant hazard ratio, estimated a relative risk of 2.14 per standard deviation (SD) of the PRS (95% confidence interval, CI, 2.09–2.19), allowing for adjusting covariates ancestry group and dbGaP study. Sensitivity analyses of the weights showed that a 10-fold decrease or increase in the weights had little impact on results (relative risks of 1.93 and 2.17, respectively). More refined weights that at-

tempted to account for potential preferential sampling of cases at different ages (see supplemental information) gave results identical to the initial proposed weights. Relative risks for each ancestry are presented in Figure 2 for relative risks per PRS SD as well as for PRS dichotomized according to the upper 90th percentile. These results show that relative risks per PRS SD were similar for European, African American, and Asian ancestries, slightly less for Latino, and much less for Ghanaian men (heterogeneity p value = 0.008). The relative risks for the upper 90th percentile were much larger than the relative risk per SD, as expected, and less heterogeneity of relative risks (p value = 0.308), although the larger standard errors of the estimates for the upper 90th percentile relative risks decreased power to detect heterogeneity. See Table S7 for the log-relative risk estimates and their standard errors.

according to study are provided in the [supplemental information](#) (see [Tables S2, S3, and S4](#)). There were 64,274 prostate cancer cases and 46,432 controls included in analyses. The ancestry groups in [Table 1](#) are based on self-report, with 82.8% European, 9.8% African American, 3.8% Latino, 2.8% Asian, and 0.8% Ghanaian. Overall, 19% reported a family history of prostate cancer. However, family history varied across the studies with the International Consortium for Prostate Cancer Genetics (ICPCG) reporting the largest fraction with family history (69.6%) because the ICPCG group focused on ascertaining cases with a family history; some studies failed to collect family history information (see [Table S4](#)). For this reason, analyses that focused on family history were considered secondary. The median age of diagnosis of prostate cancer for cases was 66 years and the median age of blood collection for controls was 62 years.

The distribution of the PRSs is illustrated in [Figure 1](#) for cases and controls of different ancestries, both the raw PRS and the PRS standardized by the mean and standard deviation within controls of each ancestry group. This figure illustrates that the distribution of the PRS is shifted to larger values for men of African American and Ghanaian ancestry, shifted to smaller values for men of Asian ancestry, and similar distributions for men of European and Latino ancestries. These shifted distributions occurred for both controls and cases, emphasizing the impact of the allele frequency differences across different ancestries. In contrast, when the PRS was centered and scaled according to the controls of each ancestry group, the distributions of the PRS overlap for the different ancestral groups. Note

Because the pooled analysis of all men showed a strong departure from a constant relative risk ($p < 2e-16$), we performed piece-wise proportional hazards analyses by partitioning age into five age groups and found significant differences in the relative risks across the age groups ($p < 1e-30$). Piece-wise proportional hazards were fit separately for each ancestry group and results in [Figure 3](#) illustrate that relative risks per SD were greatest for the youngest age group (30–55 years) and least for the oldest age group (70–88 years). For European ancestry, there was a clear trend of decreasing risk with age, from a relative risk of 2.56 for 30–55 years old (95% CI 2.47–2.65) to a relative risk of 1.86 for 70–88 years old (95% CI 1.76–1.98), with no overlapping confidence intervals throughout the different age groups. The relative risks for European

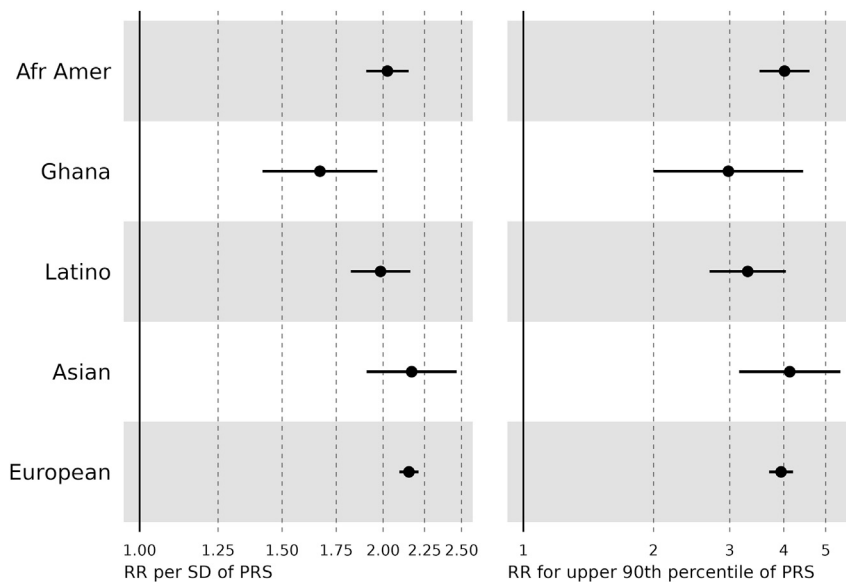


Figure 2. Relative risk per PRS SD (left panel) and for upper 90th percentile cutoff of PRS (right panel) according to ancestry. The bars represent 95% confidence intervals.

ancestry differed statistically across the age groups ($p < 0.001$). The other ancestries in [Figure 3](#) showed that the youngest and oldest age groups had different relative risks, but the patterns of risk in the intermediate age groups were not as distinct as for European ancestry, presumably because of the much smaller sample sizes of the other ancestry groups. For these non-European ancestries, the relative risks were not statistically significantly different across the age groups (see [Tables S8](#) and [S9](#) for details).

We performed a similar analysis with the PRS dichotomized according to the upper 90th percentile and similar patterns of relative risk decreasing with age were found, as illustrated in [Figure 3](#). The relative risks implied by the upper 90th percentile ranged 3–5 for most ancestries and age groups, except the lesser relative risks for Ghanaian men. The wide confidence intervals for this group reflect the small sample size. These relative risks differed significantly across the age groups ($p < 0.001$) for European ancestry but not for the non-European ancestries (see [Tables S8](#) and [S9](#) for details). Because of the large number of men with European ancestry, we were able to refine analyses of the age dependence of relative risk. We created the age group 30–45-year olds, then groups in 5-year intervals (45–50, to 70–75, then 75–88) and fit-piece-wise hazard ratios for each age group. We also fit a model with the log-relative risk depending linear on age: $\beta_0 PRS + \beta_1(a - 30)PRS$, where age a ranged 30–87 and we offset by the minimum age of 30 years in our dataset. The estimated effect of PRS was $\beta_0 = 1.2504$ (SE = 0.049) and the estimated gradient of age was $\beta_1 = -0.0138$ (SE = 0.0015). The estimated relative risks and their 95% confidence intervals for the piecewise and linear models are presented in [Figure 4](#), showing that the log-relative risk decreased approximately linearly from age 50 to age 75. We developed methods to account for preferential inclusion of cases at different ages, to evaluate the sensitivity of the Cox model weights (see supplemental information), and found

results identical to the proposed weights. These results were also consistent with fitting logistic regression models to the case-control data (see [supplemental information](#)). To compare this linear decrease across different diseases and populations, one can estimate for the relative risk per adjusted standard error (OPERA)³⁷ per year of age, which is $\exp(-0.0138/.00155) = 0.00014$.

Because men with a family history of prostate cancer have been reported to have a younger age of diagnosis,¹⁷ we performed secondary analyses to attempt to sort out the role of family history. Secondary analyses were necessary because of the relatively large number of men without family history information. For this focus, we subset to men with European ancestry to obtain sufficient sample size. The results in [Figure 5](#) illustrate that men with a family history of prostate cancer had greater relative risks associated with PRS at all ages compared with men with a negative family history (p value = 0.007) and that relative risks decreased with increasing age for both men with a family history of prostate cancer (heterogeneity of relative risks over ages, p value = 0.015) and men without a family history of prostate cancer (p value < 0.001). See [Table S10](#) for more details. The gradient of the log-relative risk with increasing age was -0.0108 (SE = 0.0045) for men with a family history and -0.0122 (SE = 0.0019) for men with a negative family history, and these gradients were not statistically significantly different (p value = 0.78). The greater relative risk among men with a family history of prostate cancer across all ages emphasizes the importance of obtaining family history information when attempting to predict future risk of prostate cancer with PRSs.

Given the decreasing relative risk of PRS with increasing age, it is important to evaluate how much the decreasing relative risk impacts prediction of future risk when attempting to use PRS for personalized medical recommendations. To view this, we computed the future prostate cancer absolute risk, conditional on men's current age and value of a standardized PRS. These future risks are based on the relative risks estimated from our data, population disease incidence rates, as well as death rates to account for competing risks. We present in [Figure 6](#) the future absolute risk for men of European and of African American ancestry, assuming current ages of 50, 60, and 70 years, for future remaining years at 1-year increments until age 80 years. These results illustrate that even though

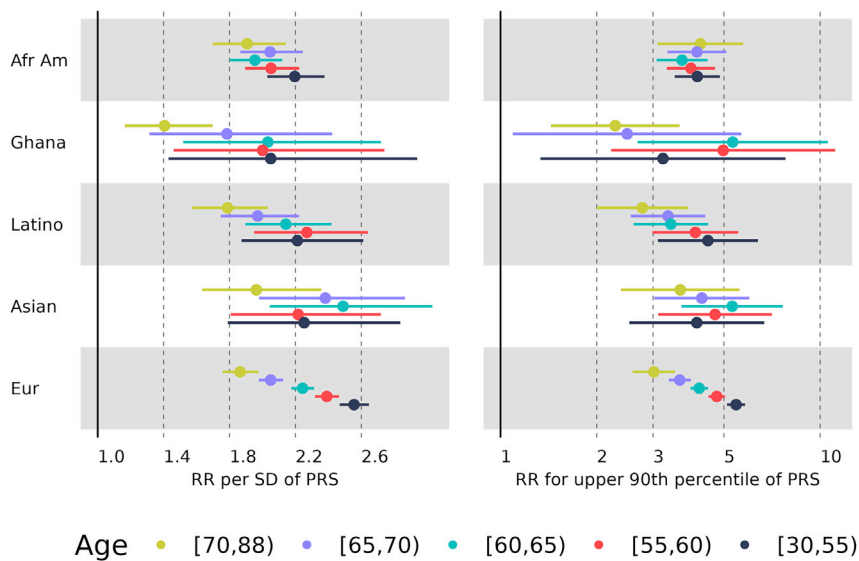


Figure 3. Relative risk per PRS SD (left panel) and for upper 90th percentile cutoff of PRS (right panel) for age groups according to ancestry
The bars represent 95% confidence intervals.

our results show strong evidence of relative risks due to PRS decreasing with increasing age, predicting future absolute risk while allowing for decreasing relative risk differed little from predictions that assumed a constant relative risk over ages. The largest difference was 2.7% for future predictions for a 70-year-old man of European ancestry.

Discussion

Based on a large number of men of diverse ancestry with publicly available genome-wide genetic variants, we demonstrated decreasing PRS relative risks for prostate cancer as age increased. Our results were most accurate for men of European ancestry because of the large number of men in

this group. By applying a novel weighted proportional hazards model to case-control data, we were able to fully utilize age information (diagnosis among cases; enrollment age among controls) to refine how the genetic relative risk decreased with age. For men of European ancestry, we observed a linear decrease of the log-relative risk from age 50 to 75, the ages at which most men are diagnosed with prostate cancer. Thomas et al. also

observed a linear decrease of the log-relative risk for colorectal cancer with age, for age > 50 years with 72,791 subjects of European ancestry and with 1,311 colorectal cancer cases.^{38,39} This reduced risk could result from non-genetic risk factors accumulating over a lifetime such that genetic effects that influence developmental pathways at younger ages have relatively less influence as non-genetic risk factors accumulate. Although a reduction in genetic relative risk with increasing age is expected when the highest risk individuals succumb to disease at younger ages, and hence are preferentially removed from the at-risk population at older ages,⁴⁰ it is important to evaluate the implications of age-varying risk. In contrast to prostate and colon cancers, breast cancer has not shown a declining risk with age for ER-negative disease, and only a weak decline has been observed for ER-positive disease.⁴¹

Despite the accumulating environmental risk with age, we observed that men of European ancestry had increased genetic relative risks at all ages if they had a family history of prostate cancer compared with a negative family history, suggesting that additional unmeasured genetic risk factors could be causing this difference, or perhaps clustering of environmental risk factors within families. This also emphasizes the importance of obtaining accurate pedigree disease information to combine with a PRS to improve age-dependent risk predictions.⁴²

As with most GWASs conducted to date, our data had a limited number of subjects with non-European ancestry, making it difficult to accurately refine how genetic relative risk decreased with age in other ancestries. Nonetheless, ancestries of Ghanaian, African American, Asian, and Latino all showed the PRS relative risk to be greatest for the youngest age group (30–55 years) and least for the oldest age group (70–88 years). These patterns were observed for both the relative risk per SD of the PRSs and binary classification based on the upper 90th percentile of the PRS distribution.

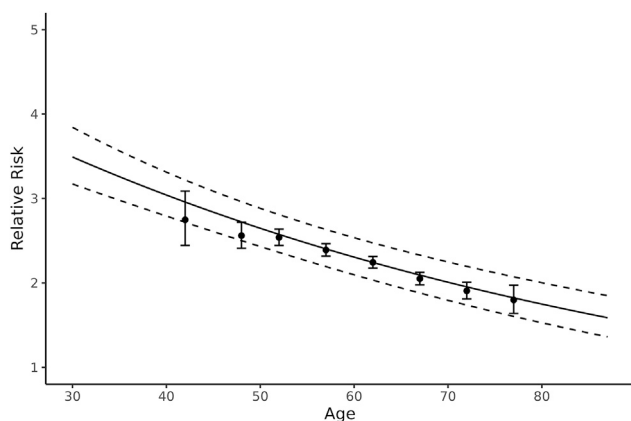


Figure 4. Piece-wise relative risk models and log-relative risk modeled as linear with age for European ancestry
The solid line is the model assuming log-relative risk depends linearly on age and dashed lines are the 95% confidence interval over ages. The estimated dependence on age was $1.2504 - 0.0138 * (\text{age} - 30)$. The piece-wise relative risks are represented as “*” and their 95% confidence intervals represented as whiskers. The piece-wise results are positioned on the x axis at the median value of age within each age group.

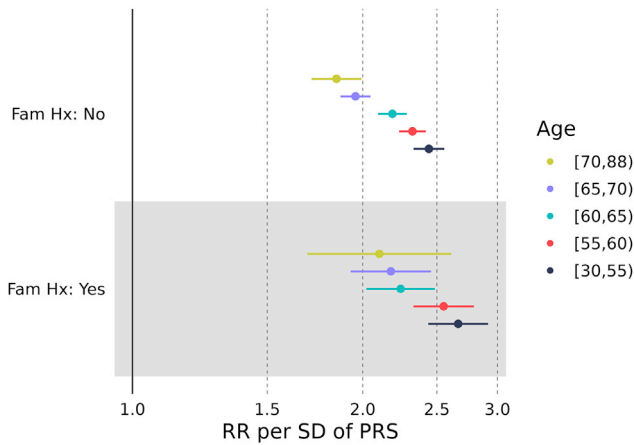


Figure 5. Relative risk per PRS SD according to age groups and family history for European ancestry
The bars represent 95% confidence intervals.

The influence of a decreasing genetic relative risk with age on personalized medical decisions should consider how the risk will be used. Categorizing into highest risk, such as above the 90th percentile of the PRS distribution, is a common approach, yet the amount of risk also depends on a man's current age, as our results show. Because absolute risks are important for interpretation of personal risks,^{24,25} we evaluated the impact of decreasing genetic relative risk with age on predicting future absolute risk. Despite strong evidence of age-dependent genetic relative risk, our results suggest that absolute risk predictions differed little between predictions that assumed a constant relative risk and those that allowed relative risks to decrease with age. These findings covered a broad range, from short-term (e.g., 1 year) to long-term (e.g., to age 80). This may be due to the calculation of absolute risk depending on both the relative risk and the baseline incidence rates; large rela-

tive risks at younger ages have less impact on absolute risk because the incidence of prostate cancer is much smaller at young ages. Assuming a constant relative risk over age simplifies the approach to calculate and present risk predictions to lay persons as well as simplifies implementation of risk discussions into clinical practice. Our strategy to evaluate how genetic relative risks vary with age and the impact of changing relative risks with age on absolute risk predictions is worth considering for other common diseases.

Appendix A

The future absolute risk of disease up to age a_h , conditional on a man alive and free of disease at age a_l and with a standardized PRS of z , depends on the baseline population age-specific incidence of disease, $\lambda_o(t)$, the population age-specific death rate, $\delta_o(t)$, and the log hazard ratio estimated by Cox regression, β_t . For our purposes, we use the subscript t to account for piece-wise proportional hazards (e.g., β_t constant over an interval), or β_t could be constant over all time. From this information, we calculate the future risk that a man will have disease by age a_h , given he is free of disease at a_l , as

$$Risk(a_h|a_l, z) = \frac{F(a_h|z) - F(a_l|z)}{S_{disease}(a_l)S_{death}(a_l)},$$

where $F(a|z) = \sum_{t=0}^a e^{\beta_t z} \lambda_o(t) S_{disease}(t|z) S_{death}(t)$ is the cumulative probability of disease up to age a accounting for competing risk of death, $S_{disease}(a|z) = \exp[-\sum_{t=0}^a e^{\beta_t z} \lambda_o(t)]$ is the probability of being free of disease at age a , and $S_{death}(a) = \exp[-\sum_{t=0}^a \delta_o(t)]$ is the probability of being alive at age a . Absolute risk calculations were achieved by estimates of β_t determined in our data.

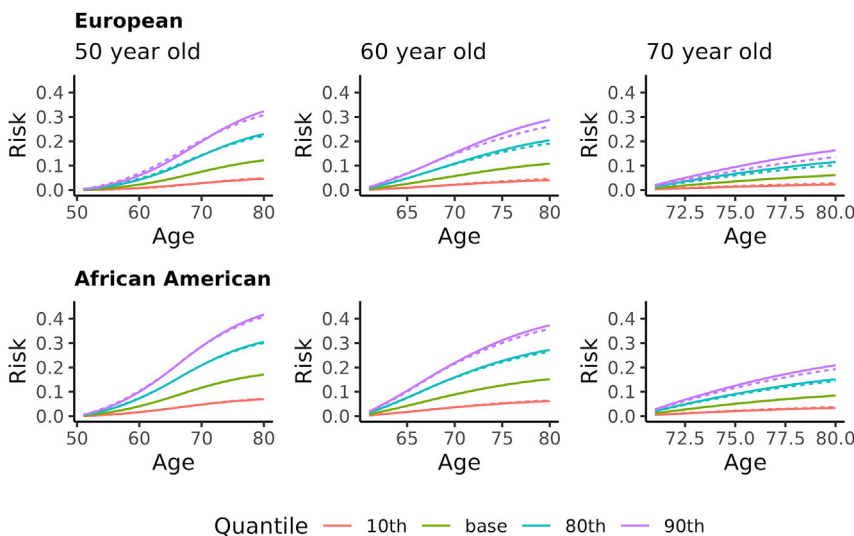


Figure 6. Future absolute risk of prostate cancer for European ancestry (left three panels) and African American ancestry (right three panels) conditional on current age and quantile of PRS

The solid lines assume a relative risk constant over ages, and the broken lines assume piece-wise relative risks to account for relative risks changing over ages. The quantiles (10, 80, 90) are for a standard normal distribution as expected for a standardized PRS in a population. The baseline represents the absolute risk assuming the PRS is unknown.

Data and code availability

All data are available through dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>) based on the dbGaP Study accession numbers provided in the supplemental information in Table S1. Code was written in the R statistical language and linux shell commands in multiple scripts that are available upon request from the corresponding author.

Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2022.03.008>.

Acknowledgments

This work was supported by the U.S. Public Health Service and National Institutes of Health (R35 GM140487).

Declaration of interests

The authors declare no competing interests.

Received: October 25, 2021

Accepted: March 9, 2022

Published: March 29, 2022

References

- Li, S., and Hopper, J.L. (2021). Age dependency of the polygenic risk score for colorectal cancer. *Am. J. Hum. Genet.* *108*, 525–526.
- Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., et al. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* *53*, 420–425.
- Choi, S.W., and O'Reilly, P.F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* *8*, giz082.
- Vilhjálmsón, B.J., Yang, J., Finucane, H.K., Gusev, A., Lindström, S., Ripke, S., Genovese, G., Loh, P.R., Bhatia, G., Do, R., et al. (2015). Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* *97*, 576–592.
- Privé, F., Arbel, J., and Vilhjálmsón, B.J. (2020). LDpred2: better, faster, stronger. *Bioinformatics* *36*, 5424–5431.
- Fritsche, L.G., Patil, S., Beesley, L.J., VandeHaar, P., Salvatore, M., Ma, Y., Peng, R.B., Taliun, D., Zhou, X., and Mukherjee, B. (2020). Cancer PRSweb: An Online Repository with Polygenic Risk Scores for Major Cancer Traits and Their Evaluation in Two Independent Biobanks. *Am. J. Hum. Genet.* *107*, 815–836.
- Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A., and Smoller, J.W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* *10*, 1776.
- Ruan, Y., Feng, Y.-C.A., Chen, C.-Y., Lam, M., Initiatives, S.G.A., Sawa, A., Martin, A.R., Qin, S., Huang, H., and Ge, T. (2020). Improving Polygenic Prediction in Ancestrally Diverse Populations. Preprint at medRxiv. <https://doi.org/10.1101/2020.12.27.20248738>.
- Lloyd-Jones, L.R., Zeng, J., Sidorenko, J., Yengo, L., Moser, G., Kemper, K.E., Wang, H., Zheng, Z., Magi, R., Esko, T., et al. (2019). Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* *10*, 5086.
- Chu, B.B., Keys, K.L., German, C.A., Zhou, H., Zhou, J.J., Sobel, E.M., Sinsheimer, J.S., and Lange, K. (2020). Iterative hard thresholding in genome-wide association studies: Generalized linear models, prior weights, and double sparsity. *Giga-science* *9*, giaa044.
- Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W., Khera, A., Okada, Y., Martin, A., Finucane, H., Price, A.L.; and The Biobank Japan Project (2021). Leveraging fine-mapping and non-European training data to improve trans-ethnic polygenic risk scores. Preprint at medRxiv. <https://doi.org/10.1101/2021.01.19.21249483>.
- Atkinson, E.G., Maihofer, A.X., Kanai, M., Martin, A.R., Karczewski, K.J., Santoro, M.L., Ulirsch, J.C., Kamatani, Y., Okada, Y., Finucane, H.K., et al. (2021). Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* *53*, 195–204.
- Amariuta, T., Ishigaki, K., Sugishita, H., Ohta, T., Koido, M., Dey, K.K., Matsuda, K., Murakami, Y., Price, A.L., Kawakami, E., et al. (2020). Improving the trans-ancestry portability of polygenic risk scores by prioritizing variants in predicted cell-type-specific regulatory elements. *Nat. Genet.* *52*, 1346–1354.
- Conti, D.V., Darst, B.F., Moss, L.C., Saunders, E.J., Sheng, X., Chou, A., Schumacher, F.R., Olama, A.A.A., Benlloch, S., Dadaev, T., et al. (2021). Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet.* *53*, 65–75.
- Newcombe, P.J., Conti, D.V., and Richardson, S. (2016). JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. *Genet. Epidemiol.* *40*, 188–201.
- Plym, A., Penney, K.L., Kalia, S., Kraft, P., Conti, D.V., Haiman, C., Mucci, L.A., and Kibel, A.S. (2021). Evaluation of a Multi-ethnic Polygenic Risk Score Model for Prostate Cancer. *J. Natl. Cancer Inst.*, djab058.
- Na, R., Labbate, C., Yu, H., Shi, Z., Fantus, R.J., Wang, C.H., Andriole, G.L., Isaacs, W.B., Zheng, S.L., Helfand, B.T., and Xu, J. (2019). Single-Nucleotide Polymorphism-Based Genetic Risk Score and Patient Age at Prostate Cancer Diagnosis. *JAMA Netw. Open* *2*, e1918145.
- Mavaddat, N., Pharoah, P.D., Michailidou, K., Tyrer, J., Brook, M.N., Bolla, M.K., Wang, Q., Dennis, J., Dunning, A.M., Shah, M., et al. (2015). Prediction of breast cancer risk based on profiling with common genetic variants. *J. Natl. Cancer Inst.* *107*, djv036.
- Isgut, M., Sun, J., Quyyumi, A.A., and Gibson, G. (2021). Highly elevated polygenic risk scores are better predictors of myocardial infarction risk early in life than later. *Genome Med.* *13*, 13.
- Jiang, X., Holmes, C., and McVean, G. (2021). The impact of age on genetic risk for common diseases. *PLoS Genet.* *17*, e1009723.
- Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* *51*, 584–591.
- Cotter, M.P., Gern, R.W., Ho, G.Y., Chang, R.Y., and Burk, R.D. (2002). Role of family history and ethnicity on the mode and age of prostate cancer presentation. *Prostate* *50*, 216–221.

23. Brandt, A., Bermejo, J.L., Sundquist, J., and Hemminki, K. (2009). Age at diagnosis and age at death in familial prostate cancer. *Oncologist* 14, 1209–1217.
24. Fagerlin, A., Zikmund-Fisher, B.J., and Ubel, P.A. (2011). Helping patients decide: ten steps to better risk communication. *J. Natl. Cancer Inst.* 103, 1436–1443.
25. Zipkin, D.A., Umscheid, C.A., Keating, N.L., Allen, E., Aung, K., Beyth, R., Kaatz, S., Mann, D.M., Sussman, J.B., Korenstein, D., et al. (2014). Evidence-based risk communication: a systematic review. *Ann. Intern. Med.* 161, 270–280.
26. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.
27. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867–2873.
28. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299.
29. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287.
30. Arias, E. (2012). United States life tables, 2008. *Natl. Vital Stat. Rep.* 61, 1–63.
31. Davies, H.T., Crombie, I.K., and Tavakoli, M. (1998). When can odds ratios mislead? *BMJ* 316, 989–991.
32. van der Net, J.B., Janssens, A.C., Eijkemans, M.J., Kastelein, J.J., Sijbrands, E.J., and Steyerberg, E.W. (2008). Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *Eur. J. Hum. Genet.* 16, 1111–1116.
33. Leffondré, K., Abrahamowicz, M., and Siemiatycki, J. (2003). Evaluation of Cox's model and logistic regression for matched case-control data with time-dependent covariates: a simulation study. *Stat. Med.* 22, 3781–3794.
34. Chen, K., and Lo, S.-H. (1999). Case-cohort and case-control analysis with Cox's model. *Biometrika* 86, 755–764.
35. Therneau, T.M., and Li, H. (1999). Computing the Cox model for case cohort designs. *Lifetime Data Anal.* 5, 99–112.
36. Nan, B., and Lin, X. (2008). Analysis of case-control age-at-onset data using a modified case-cohort method. *Biom. J.* 50, 311–320.
37. Hopper, J.L. (2015). Odds per adjusted standard deviation: comparing strengths of associations for risk factors measured on different scales and across diseases and populations. *Am. J. Epidemiol.* 182, 863–867.
38. Thomas, M., Sakoda, L.C., Hoffmeister, M., Rosenthal, E.A., Lee, J.K., van Duijnhoven, F.J.B., Platz, E.A., Wu, A.H., Dampier, C.H., de la Chapelle, A., et al. (2021). Response to Li and Hopper. *Am. J. Hum. Genet.* 108, 527–529.
39. Thomas, M., Sakoda, L.C., Hoffmeister, M., Rosenthal, E.A., Lee, J.K., van Duijnhoven, F.J.B., Platz, E.A., Wu, A.H., Dampier, C.H., de la Chapelle, A., et al. (2020). Genome-wide Modeling of Polygenic Risk Score in Colorectal Cancer Risk. *Am. J. Hum. Genet.* 107, 432–444.
40. Aalen, O.O., Valberg, M., Grotmol, T., and Tretli, S. (2015). Understanding variation in disease risk: the elusive concept of frailty. *Int. J. Epidemiol.* 44, 1408–1421.
41. Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.H., Wang, Q., Bolla, M.K., et al. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* 104, 21–34.
42. So, H.C., Kwan, J.S., Cherny, S.S., and Sham, P.C. (2011). Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.* 88, 548–565.

The American Journal of Human Genetics, Volume 109

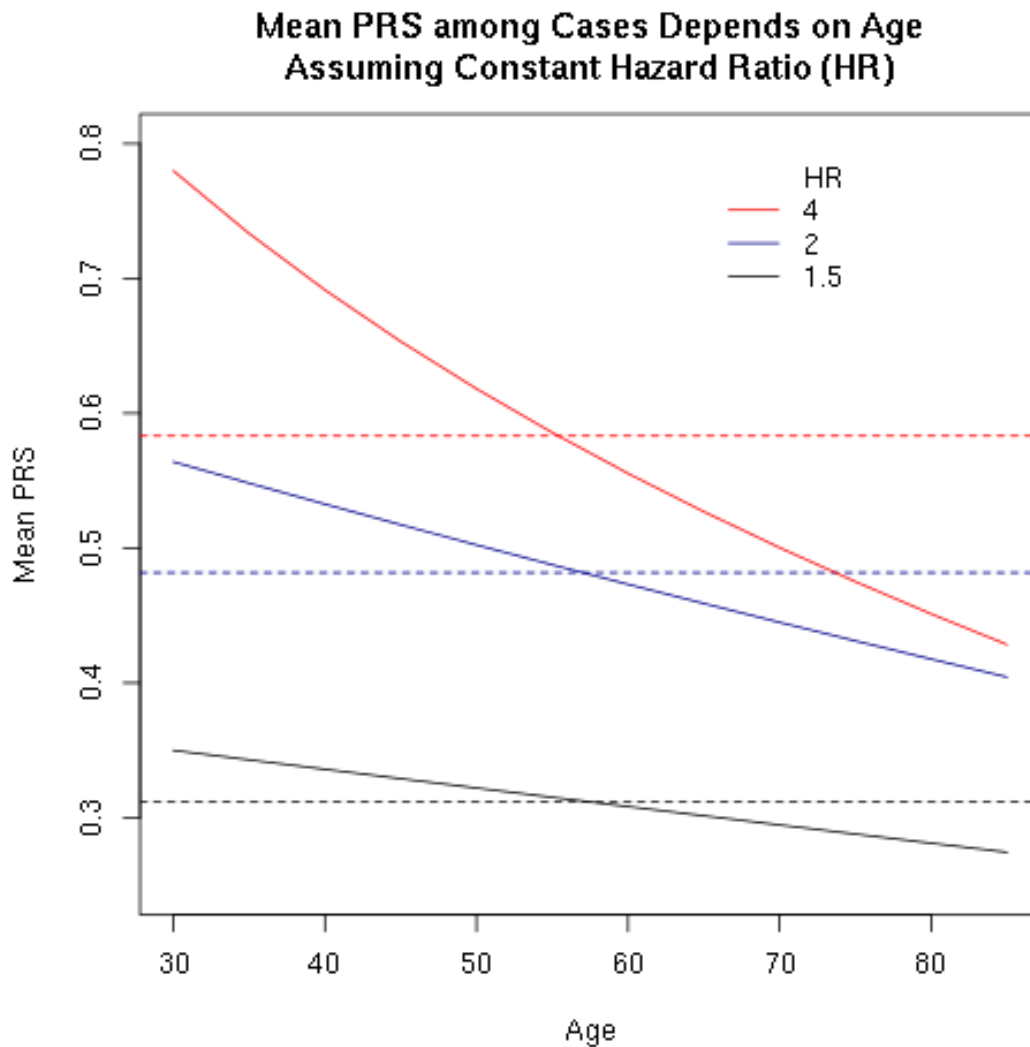
Supplemental information

**Polygenic risk for prostate cancer: Decreasing
relative risk with age but little impact on absolute risk**

Daniel J. Schaid, Jason P. Sinnwell, Anthony Batzler, and Shannon K. McDonnell

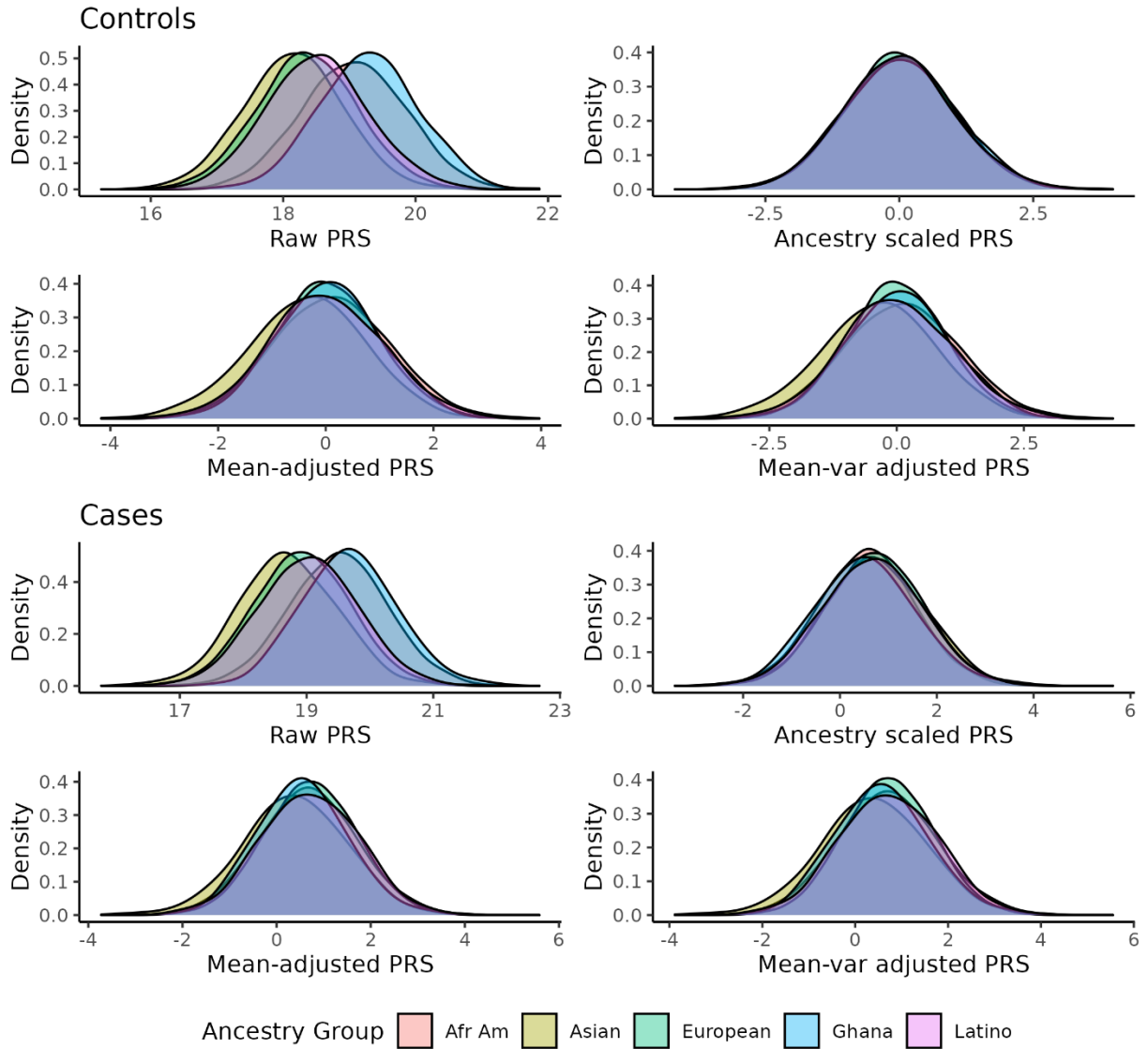
Supplemental Information

Figure S1.



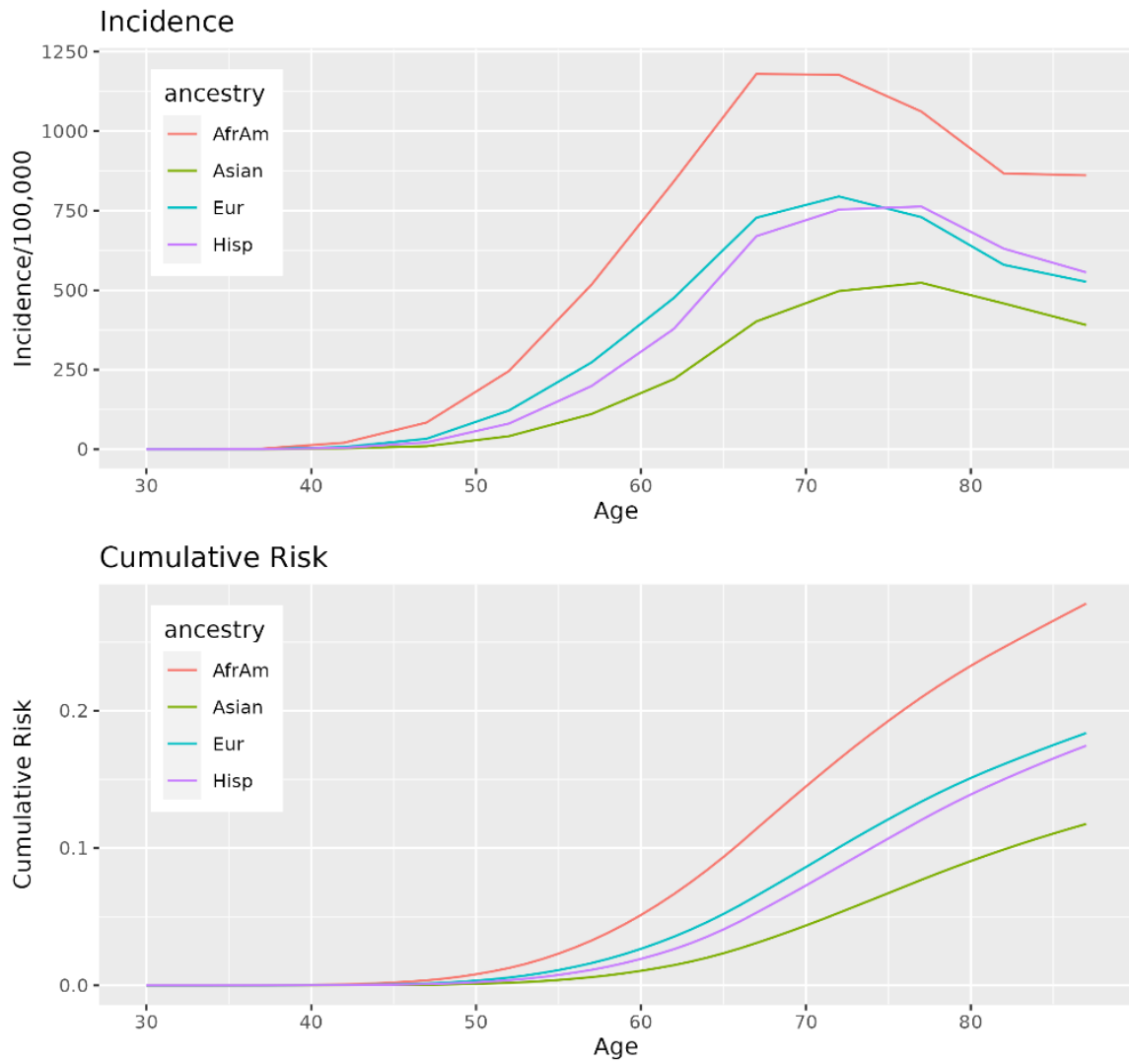
Mean PRS among cases over different ages, assuming a constant hazard ratio. The solid lines illustrate how the mean PRS is expected to decrease with age, and the horizontal dashed lines provide perspective on how the solid lines pivot from a constant value. See Methods section for Theoretical Mean PRS Among Cases & Age for derivations used to create Figure S1.

Figure S2



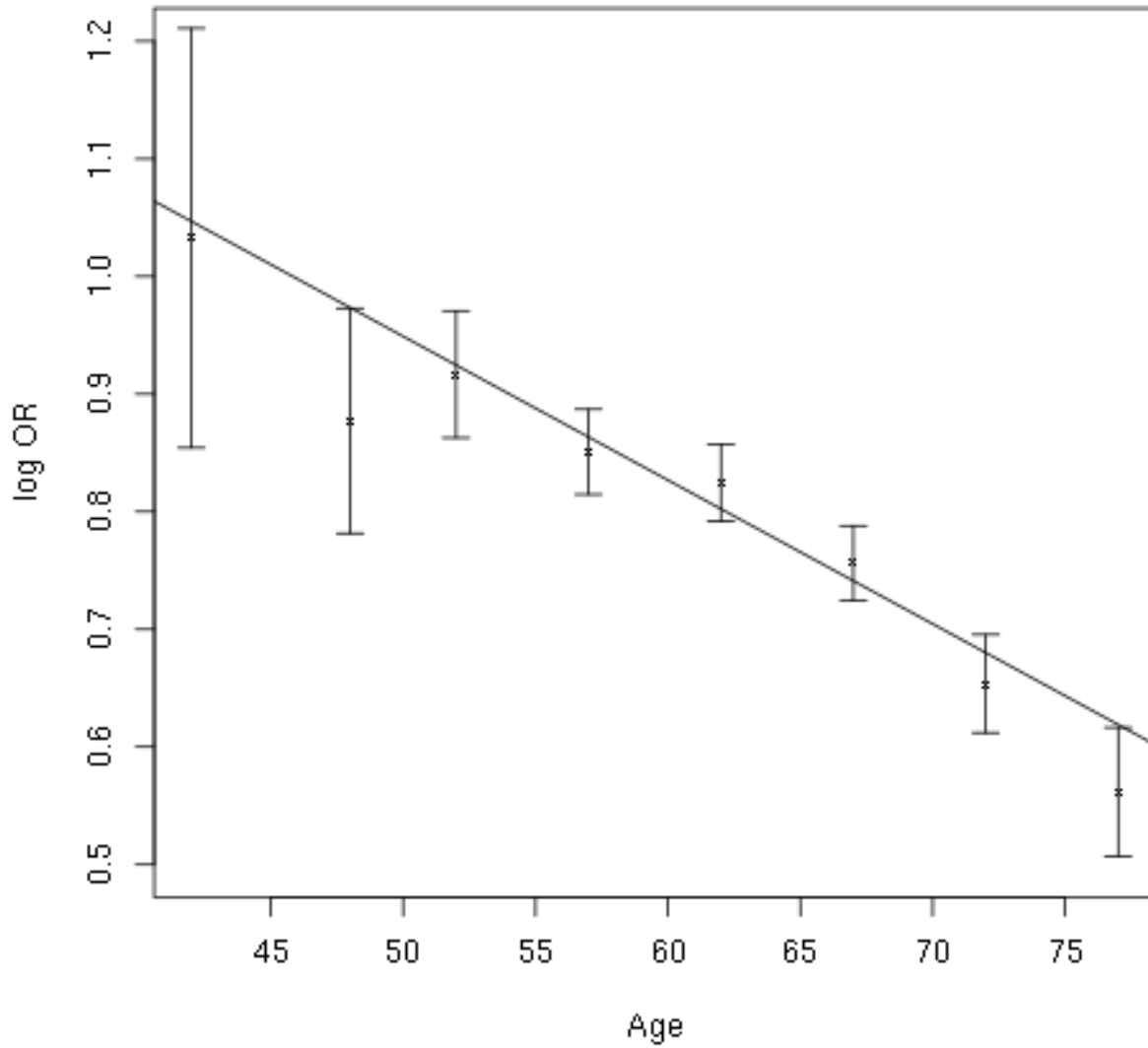
Distribution of PRS uncorrected (raw PRS) and corrected by various methods for controls and cases. See Methods section PRS Ancestry Correction by Projection onto 1,0000 Genome Reference Sample for details.

Figure S3.



Age-specific incidence of prostate cancer, per 100,000, for different ancestry groups (upper panel) and corresponding cumulative risk of prostate cancer (lower panel). See Methods section Age-Specific Incidence and Cumulative Risk by Ancestry for details.

Figure S4.



Piece-wise log-odds-ratios (“*”) and their 95% confidence intervals represented as whiskers) and model assuming log-odds-ratio depends linearly on age (solid line). The piece-wise results are positioned on the x-axis at the median value of age within each age group. See Methods section on Linear Decrease in PRS Log-Odds-Ratio with Age for details.

Table S1. Studies obtained from dbGaP for Prostate Cancer Genome Wide Association Study Data. See Methods section Description of dbGaP Studies for more details

dbGaP Study Accession		Study Label	Study Full Name
phs000207.v1.p1		CGEMS	CGEMS Prostate Cancer GWAS - Stage 1 – PLCO (Embargo Release Date: December 22, 2009)
phs000306.v4.p1		GENEVA	GENEVA Prostate Cancer (Embargo Release Date: February 01, 2013)
phs000812.v1.p1		BPC3	Characterizing Genetic Susceptibility to Breast and Prostate Cancer - BPC3 (Embargo Release Date: June 11, 2015)
phs000838.v1.p1		GHANA	Ghana Prostate Study (Embargo Release Date: July 10, 2015)
phs000882.v1.p1		PEGASUS	National Cancer Institute (NCI) Prostate Cancer Genome-wide Association Study for Uncommon Susceptibility Loci (PEGASUS) (Embargo Release Date: January 24, 2017)
phs001391.v1.p1		ONCO	OncoArray: Prostate Cancer (Embargo Release Date: March 21, 2018)
phs000733.v1.p1		ICPCG	The International Consortium for Prostate Cancer Genetics Genome Wide Association Study of Familial Prostate Cancer (Embargo Release Date: March 10, 2015)

Table S2. Sample Size before and after quality control, merging, and removal of related subjects.

Study	Downloaded Data		After Initial QC (1)		After Merging and Removing Related Subjects (2)	
	Cases	Controls	Cases	Controls	Cases	Controls
BPC3	2782	4458	2775	4451	2179	3573
CGEMS	1151	1101	1140	1097	520	414
GENEVA	4304	4529	4264	4496	4115	4371
ICPCG	2568	1422	2520	1392	2106	1022
ONCO	52700	37751	52658	37723	51489	36910
PEGASUS	4599	2841	4595	2840	3670	2290
GHANA	474	458	474	458	461	452
Total	68578	52560	68426	52457	64540	49032

(1) Reasons for exclusion include low call rate (<80%), low heterozygosity (<0.4) on any chromosome, self-reported sex inconsistent with chromosome X and Y data.

(2) One subject was randomly removed from each related pair (kinship coefficient ≥ 0.0442)

Table S3: Exclusions by study.

	BPC3 (N=5752)	CGEMS (N=934)	GENEVA (N=8486)	ICPCG (N=3128)	ONCO (N=88399)	PEGASUS (N=5960)	WAFR (N=913)	Total (N=113572)
Missing Age	0	0	0	123	2677	0	0	2800
Age < 30	0	0	0	0	39	0	0	39
Missing Ancestry	0	0	0	27	0	0	0	27

Table S4. Description of studies included in analyses ^a

	BPC3 (No.=5752)	CGEMS (No.=934)	GENEVA (No.=8486)	ICPCG (No.=2978)	ONCO (No.=85683)	PEGASUS (No.=5960)	WAFR (No.=913)	Total (No.=110706)
Status								
Case	2179 (37.9%)	520 (55.7%)	4115 (48.5%)	2072 (69.6%)	51257 (59.8%)	3670 (61.6%)	461 (50.5%)	64274 (58.1%)
Control	3573 (62.1%)	414 (44.3%)	4371 (51.5%)	906 (30.4%)	34426 (40.2%)	2290 (38.4%)	452 (49.5%)	46432 (41.9%)
Ancestry								
Afr.Amer	0	0	4521 (53.3%)	0	6354 (7.4%)	0	0	10875 (9.8%)
Asian	0	0	1935 (22.8%)	0	1122 (1.3%)	0	0	3057 (2.8%)
European	5752 (100.0%)	934 (100.0%)	0	2978 (100.0%)	75999 (88.7%)	5960 (100.0%)	0	91623 (82.8%)
Ghana	0	0	0	0	0	0	913 (100.0%)	913 (0.8%)
Latino	0	0	2030 (23.9%)	0	2208 (2.6%)	0	0	4238 (3.8%)
Family History PrCa								
Unknown	1483	0	3024	0	29730	5960	913	41110
No	3826 (89.6%)	844 (90.4%)	4923 (90.1%)	906 (30.4%)	45896 (82.0%)	0	0	56395 (81.0%)
Yes	443 (10.4%)	90 (9.6%)	539 (9.9%)	2072 (69.6%)	10057 (18.0%)	0	0	13201 (19.0%)
Age Group								
[30,45)	0	0	77 (0.9%)	79 (2.7%)	1163 (1.4%)	0	2 (0.2%)	1321 (1.2%)
[45,50)	22 (0.4%)	0	206 (2.4%)	210 (7.1%)	2557 (3.0%)	0	2 (0.2%)	2997 (2.7%)
[50,55)	86 (1.5%)	0	444 (5.2%)	448 (15.0%)	8121 (9.5%)	0	134 (14.7%)	9233 (8.3%)
[55,60)	339 (5.9%)	121 (13.0%)	905 (10.7%)	628 (21.1%)	16423 (19.2%)	300 (5.0%)	146 (16.0%)	18862 (17.0%)
[60,65)	929 (16.2%)	0	1296 (15.3%)	620 (20.8%)	19219 (22.4%)	1212 (20.3%)	150 (16.4%)	23426 (21.2%)
[65,70)	1669 (29.0%)	497 (53.2%)	1816 (21.4%)	562 (18.9%)	18923 (22.1%)	1768 (29.7%)	169 (18.5%)	25404 (22.9%)
[70,75)	1413 (24.6%)	0	1796 (21.2%)	257 (8.6%)	11651 (13.6%)	1674 (28.1%)	175 (19.2%)	16966 (15.3%)
[75,88)	1294 (22.5%)	316 (33.8%)	1946 (22.9%)	174 (5.8%)	7626 (8.9%)	1006 (16.9%)	135 (14.8%)	12497 (11.3%)
Age,median (range)	69 (45,87)	67 (57, 77)	67 (44, 77)	61 (33, 87)	63 (30, 87)	67 (57, 77)	65 (42, 87)	64 (30, 87)

^aExcluding men with missing age, age < 30 years, or missing ancestry

Table S5. Self-reported ancestry versus ancestry based on the maximum estimate of admixture probability. See Methods section Ancestry: Self—Reported and Genetically Informed for more details.

Genetic Admixture Max Probability	Self-reported Ancestry				
	African American	Asian	European	Latino	Ghana
African	10354	1	0	11	913
Amerindian	7	113	0	1637	0
Asian	33	2746	7	24	0
European	481	197	88638	2566	0

Table S6. Intercept and slope for models of log-risk of PRS as a function of age in years.

Model	Intercept (SE)	Slope (SE)
Weighted Cox	1.2504 (0.049)	-0.0138 (0.0015)
Logistic Regression	1.1935 (0.035)	-0.0122 (0.0010)

Table S7. Log-relative risk estimates (beta) and their standard errors (se) for Figure 2.

Ancestry	beta persd	se persd	beta up90	se up90
Afr				
Amer	0.7054	0.0309	1.3904	0.0681
Ghana	0.5132	0.0834	1.0916	0.2031
Latino	0.6858	0.0433	1.1945	0.1038
Asian	0.7743	0.0654	1.4182	0.1377
European	0.7666	0.0140	1.3720	0.0326

Table S8. Log-relative risk estimates (beta) and their standard errors (se) for Figure 3.

Ancestry	Age	beta persd	se persd	beta up90	se up90
Afr Am	[30, 55)	0.7867	0.0403	1.4177	0.0831
Afr Am	[55, 60)	0.7188	0.0408	1.3718	0.0888
Afr Am	[60, 65)	0.6699	0.0417	1.3081	0.0937
Afr Am	[65, 70)	0.7160	0.0473	1.4152	0.1084
Afr Am	[70, 88)	0.6454	0.0593	1.4397	0.1570
Ghana	[30, 55)	0.7179	0.1838	1.1702	0.4502
Ghana	[55, 60)	0.6935	0.1604	1.6039	0.4118
Ghana	[60, 65)	0.7092	0.1485	1.6726	0.3498
Ghana	[65, 70)	0.5790	0.1562	0.9123	0.4200
Ghana	[70, 88)	0.3404	0.0963	0.8264	0.2365
Latino	[30, 55)	0.7937	0.0848	1.4939	0.1839
Latino	[55, 60)	0.8194	0.0774	1.4036	0.1577
Latino	[60, 65)	0.7615	0.0622	1.2271	0.1374
Latino	[65, 70)	0.6782	0.0614	1.2064	0.1367
Latino	[70, 88)	0.5812	0.0651	1.0218	0.1673
Asian	[30, 55)	0.8122	0.1175	1.4138	0.2478
Asian	[55, 60)	0.7960	0.1041	1.5453	0.2087
Asian	[60, 65)	0.9118	0.1006	1.6689	0.1861
Asian	[65, 70)	0.8676	0.0942	1.4500	0.1744
Asian	[70, 88)	0.6744	0.0933	1.2949	0.2177
Eur	[30, 55)	0.9384	0.0177	1.6969	0.0332
Eur	[55, 60)	0.8715	0.0156	1.5575	0.0304
Eur	[60, 65)	0.8081	0.0157	1.4314	0.0326
Eur	[65, 70)	0.7178	0.0183	1.2914	0.0403
Eur	[70, 88)	0.6227	0.0296	1.1045	0.0784

Table S9. Tests of heterogeneity of relative risks across ages for parameters in Table S8.

Ancestry	Test of Heterogeneity of Relative Risk across Ages	
	Per SD PRS	Upper 90 th Percentile of PRS
Afr Am	0.221	0.902
Ghana	0.112	0.228
Latino	0.099	0.321
Asian	0.473	0.746
Eur	0.000	0.000

Table S10. Log-relative risk estimates (beta) and their standard errors (se) for Figure 5.

FamHx	Age	beta perSD	se perSD
No	[30, 55)	0.8922	0.0236
No	[55, 60)	0.8427	0.0204
No	[60, 65)	0.7824	0.0221
No	[65, 70)	0.6713	0.0230
No	[70, 88)	0.6141	0.0382
Yes	[30, 55)	0.9803	0.0459
Yes	[55, 60)	0.9366	0.0464
Yes	[60, 65)	0.8072	0.0526
Yes	[65, 70)	0.7777	0.0617
Yes	[70, 88)	0.7432	0.1107

Methods

Theoretical Mean PRS Among Cases & Age

The mean PRS among cases depends on the strength of association of the PRS with disease and it is possible for the mean PRS to be greater among younger cases than older cases, even if the hazard ratio associated with a PRS is constant over all ages. This is because men who have greater values of PRS are at the greatest susceptibility for disease and are more likely to succumb at a younger age.

Below we derive the expected PRS among cases, and how this expectation depends on age, when assuming a constant hazard ratio (e.g., proportional hazards model). The derivation follows standard methods for survival analyses. Assume that the standardized PRS, z , has a standard normal density, $\phi(z)$. The probability of disease at age a , conditional on z , is

$$P(a | z) = \lambda_o(a) e^{\beta z} S_o(a)^{\exp(\beta z)}, \text{ where } \lambda_o(a) \text{ is the baseline hazard rate, } S_o(a) = \exp\left[-\sum_{t=0}^a \lambda_o(t)\right],$$

and β is the log hazard ratio constant over age. From these, we determine the density of z conditional on disease at age a :

$$P(z | a) = \frac{\lambda_o(a) e^{\beta z} S_o(a)^{\exp(\beta z)} \phi(z)}{\int_{-\infty}^{\infty} \lambda_o(a) e^{\beta z} S_o(a)^{\exp(\beta z)} \phi(z) dz}. \quad (1)$$

The expected value of z among diseased cases at age a is then

$$E[z | a] = \int_{-\infty}^{\infty} zP(z | a)dP \quad (2)$$

To illustrate this numerically, we assume that age of disease diagnosis has an exponential distribution (i.e., constant hazard rate of $\lambda = .003$, the mean baseline incidence for European ancestry), and a constant log hazard ratio of β , making it easy to numerically integrate equations (1) and (2). Figure S1 illustrates how the mean PRS among cases decreases with age, while the mean PRS among controls is expected to be approximately zero.

PRS Ancestry Correction by Projection onto 1,000 Genome Reference Sample

Because the distribution of PRS differs across different ancestries due to SNP allele frequency differences, we evaluated three approaches to correct for population differences: 1) centering and scaling the PRS within each ancestry group, using the mean and standard deviation for controls within each ancestry group; 2) projection of data onto 1,000 Genome reference panel and correction of mean PRS;¹ 3) projection of data onto 1,000 Genome reference panel and correction of both mean PRS and variance of PRS. These latter two methods intend to provide a continuum of correction for men of different ancestries, some admixed.

The projection methods #2 and #3 (Christopher Kachulis, Broad Institute, personal communication), are based on projecting the study sample PRS onto a reference sample. This is accomplished by computing the PRS on the reference sample and using linear regression to regress the reference PRS on the top (maybe 10) principal components of the reference sample. The regression coefficients from this reference regression are used to predict the PRS in the study sample, by using the sample principal components. This predicted value is then subtracted from the sample PRS to adjust for ancestry. Because this approach only corrects for the mean of the distribution, it might not fully correct for ancestry if the variance of the PRS differs across ancestry. To adjust for the variance, one can create residuals from the linear regression in the reference sample, and then perform a second linear regression of the squared residuals on the principal components in the reference sample. This can then be used to predict the variance in the study sample, by using the regression coefficients with the sample principal components. To illustrate, the PRS adjusted score would be computed as

$$PRS_{adjusted} = \frac{PRS_{sample} - (\alpha_o + \sum \alpha_i PC_i)}{\sqrt{\beta_o + \sum \beta_i PC_i}},$$

where α coefficients are estimated by regression of the PRS on the principal components in the reference samples, β coefficients are estimated by regression of the squared residuals on the principal components in the reference samples, and PRS_{sample} and PC_i are from the study samples.

When computing the above PRS corrections, the genetic variants need to be available in both the study and reference samples. For the prostate cancer PRS, there were 220 variants available in our study sample, but only 212 of these were available in the 1000 Genome reference (8 variants were in our prostate cancer studies but not available in the 1000 Genome reference data). To compute the principal components, we removed the 212 risk variants and removed variants in linkage disequilibrium, resulting in approximately 100,000 variants.

The panels in Figure S2 below illustrate the distribution of the raw PRS, the PRS corrected by self-reported ancestry mean and standard deviation among controls (“Ancestry scaled PRS”), the PRS corrected by projection on the reference samples, correcting for the mean (“Mean-adjusted PRS”), and the PRS corrected by projection correcting for both mean and variance (“Mean-Var adjusted PRS”). For our prostate cancer study, it can be seen that the projection methods do not fully correct for ancestry, by viewing the non-overlapping distributions among controls. In contrast, the self-reported ancestry mean and standard deviation correction performed better.

Age-Specific Incidence and Cumulative Risk by Ancestry

The age-specific incidence rates of prostate cancer, as described in the main manuscript methods, are illustrated in the upper panel of Figure S3. This figure illustrates the greater incidence rate among African American ancestry across all ages, the lesser incidence among Asian ancestry and similar incidence among European and Latino (Hispanic) ancestries. The incidence rates, λ_t , can be used to compute the cumulative incidence of prostate cancer by age a ,

$F(a) = \exp[-\sum_{t=0}^a \lambda_t]$. The cumulative incidences in the ancestral populations are illustrated in the

lower panel of Figure S3. By age 87, the life-time risk of prostate cancer is expected to be 28% among African American men, 18% among European ancestry, 17% among Latino ancestry, and 12% among Asian ancestry.

Linear Decrease in PRS Log-Odds-Ratio with Age

The results in the main manuscript, illustrated in Figure 4, were based on fitting Cox proportional hazards models with weights the inverse of population incidence rates for controls. Since men were enrolled based on case-control studies, we evaluated the sensitivity of our conclusions of log-relative-risk decreasing linearly with age among European ancestry by fitting piece-wise logistic regression models for age partitioned into 5 year intervals, from age 30 to age 88. In addition, a logistic regression model assuming linear change in PRS risk according to age was fit by the form of $\text{status} \sim \text{cohort} + (\text{age}-30) + \text{PRS} + \text{I}((\text{age}-30)*\text{PRS})$. The results from piece-wise fits and the linear model are illustrated in Figure S4. The linear decrease in log-odds-ratio fits the data well for ages 50-70. Table S6 below shows that parameter estimates and their standard errors for the weighted Cox model and the logistic regression model are consistent, and quite close for the linear decrease in log risk.

Description of dbGaP Studies

A brief description of each the studies obtained from dbGaP and listed Table S1 is provided below. Complete descriptions are available from the dbGaP web site (<https://www.ncbi.nlm.nih.gov/gap/>)

CGEMS

The Cancer Genetic Markers of Susceptibility (CGEMS) prostate cancer genome-wide association study (GWAS) included genotyping approximately 550,000 SNPs (Phase 1A with HumanHap300 and Phase 1B HumanHap240, both from Illumina, San Diego, CA) in 1,172 prostate cancer patients and 1,157 controls of European ancestry from the Prostate, Lung, Colon and Ovarian (PLCO) Cancer Screening Trial. Selected publications include:²

Acknowledgement: Data submitted to dbGaP by Lead Principal Investigator Stephen J. Chanock. Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute and Core Genotyping Facility, Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes of Health, Department (NIH), Department of Health and Human Services (DHHS), Bethesda, MD, USA. dbGaP accession [phs000207.v1.p1](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs000207.v1.p1).

GENEVA

This study is part of the Gene Environment Association Studies initiative (GENEVA, <http://www.genevastudy.org>) funded by the trans-NIH Genes, Environment, and Health Initiative (GEI). The version 1 release of this dataset included genotype data for the Japanese and Latino populations in the study. Genotyping was performed at the Broad Institute of MIT and Harvard, a GENEVA genotyping center and at the University of Southern California. Selected publications include:^{3; 4}

Acknowledgement: Funding support for the GENEVA Prostate Cancer study was provided through the National Cancer Institute (R37CA54281, R01CA6364, P01CA33619, U01CA136792, and U01CA98758) and the National Human Genome Research Institute (U01HG004726). Assistance with phenotype harmonization, SNP selection, data cleaning, meta-analyses, data management and dissemination, and general study coordination, was provided by the GENEVA Coordinating Center (U01HG004789-01). dbGaP accession phs000306.v4.p1.

BPC3

The Breast and Prostate Cancer Cohort Consortium (BPC3) was established in 2003 to pool data and biospecimens from nine large prospective cohorts to conduct research on gene-environment interactions in cancer etiology. The BPC3 GWAS includes the following cohorts: the American Cancer Society Cancer Prevention Study-II (CPS-II); the European Prospective Investigation of Cancer (EPIC); the Physician's Health Study (PHS); the Nurses' Health Studies I and II (NHS and NHSII); the Health Professionals Follow-up Study (HPFS); the Multiethnic Cohort (MEC); the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial; and the Alpha-Tocopherol, Beta-Carotene (ATBC) Study. Selected publications include:⁵⁻⁷

Acknowledgement: The Breast and Prostate Cancer Cohort Consortium (BPC3) genome-wide association studies of advanced prostate cancer and estrogen-receptor negative breast cancer was supported by the National Cancer Institute under cooperative agreements U01-CA98233, U01-CA98710, U01-CA98216, and U01-CA98758 and the Intramural Research Program of the National Cancer Institute, Division of Cancer Epidemiology and Genetics. dbGaP accession [phs000812.v1.p1](https://www.ncbi.nlm.nih.gov/bioproject/1000000000)

GHANA

Participants were recruited through the Ghana Prostate Study (a population-based component and a clinical component) between 2004 and 2006. Additional prostate cancer cases were recruited between 2008 and 2012. Selected publications include:⁸⁻¹⁴

Acknowledgement: The genome-wide association study of prostate cancer in West African men project was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health, Department of Health and Human Services including Contract No.

HHSN261200800001E. The datasets have been accessed through the NIH database for Genotypes and Phenotypes (dbGaP). A full list of acknowledgements can be found in the supplementary note ⁸. dbGaP accession [phs000838.v1.p1](#)

PEGASUS

This genome-wide association study was funded by the National Cancer Institute (NCI) to identify uncommon susceptibility loci for prostate cancer. A total of 7440 subjects of European ancestry from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial were genotyped using the Illumina HumanOmni2.5.

Acknowledgement: The National Cancer Institute (NCI) Prostate Cancer Genome-wide Association Study for Uncommon Susceptibility Loci (PEGASUS) was supported by the Intramural Research Program of the NCI. Please see publication number 14 dbGaP accession [phs000882.v1.p1](#).¹⁵

ONCO

Original description of the study: From ELLIPSE (linked to the PRACTICAL consortium), ~78,000 SNPs were contributed to the OncoArray. A large fraction of the content was derived from the GWAS meta-analyses in European ancestry populations (overall and aggressive disease; ~27K SNPs). An additional just over 10,000 SNPs were selected from the meta-analyses in the non-European populations, with a majority of these SNPs coming from the analysis of overall prostate cancer in African ancestry populations as well as from the multiethnic meta-analysis. A substantial fraction of SNPs (~28,000) were also selected for fine-mapping of 53 loci not included in the common fine-mapping regions (tagging at $r^2 > 0.9$ across ± 500 kb regions). A few thousand SNPs related with PSA levels and/or disease survival as well as SNPs from candidate lists provided by study collaborators, as well as from meta-analyses of exome SNP chip data from the Multiethnic Cohort and UK studies, were also selected. A large number of studies contributed to the total sample (99,622): see description at the web link https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001391.v1.p1.

Acknowledgement: dbGaP accession [phs001391.v1.p1](#). See below for **OncoArray: Prostate Cancer Acknowledgements**.

ICPCG

The aim of this study was to perform a GWAS for prostate cancer cases that came from pedigrees with multiple men affected with prostate cancer. Pedigrees were identified that had 3 or more related prostate cancer cases and have an average age at diagnosis ≤ 75 years. Only one case was chosen from each pedigree; if more than one case was available in a pedigree, the most

aggressive case was chosen, or if no aggressive cases, the case with the earliest age of diagnosis was chosen. Male controls were selected such that they were unrelated to cases and to each other and had a distribution of race and birth year similar to the cases. A GWAS was performed based on genotyping with the Illumina 5M plus exome SNP set. Selected references include:¹⁶⁻¹⁹

Acknowledgement: Data was provided by principal investigator Lisa Cannon Albright, PhD. The University of Utah, UT, USA, and funding provided by R01 CA089600. National Institutes of Health, Bethesda, MD, USA. The genotyping data was generated and provided by the International Consortium for Prostate Cancer Genetics (ICPCG). The ICPCG was funded by a grant from the National Institutes of Health, U01 CA89600. dbGaP accession phs000733.v1.p1.

Ancestry: Self-Reported and Genetically Informed

The program ADMIXTURE²⁰ was used to estimate genetic admixture probabilities using a reference sample of 1000 Genome supplemented with data from the Human Genome Diversity Project.²¹ We found the ADMIXTURE software to provide odd results for very large sample sizes (~100,000), presumably due to numerical accuracy when computing the log-likelihood. For this reason, we partitioned samples into batches of size no greater than 10,000. The results in Table S5 below illustrate that potential misclassification was minimal among subjects self-reported as African American, European and Ghana. Asian subjects had a small number genetically classified as Amerindian, yet there is close ancestry among Amerindian and Asians, so these might represent historical ancestries. Latino ancestry is known to be admixed among Amerindian, European, and African. Note that the classification by maximum genetic admixture probability includes subjects that are bordering 50:50 admixture between two ancestries, such as 51% European and 49% African who self-report as African.

Model to Adjust Weights to Fit Age Distribution Among Cases

To evaluate the sensitivity of the weights in the Cox model, we modified the population incidence (hazard) rates to better fit the age distribution of the cases to allow for the possibility that the cases were sampled with preference to certain ages. Based on theory of survival analysis, the probability density of an event occurring at age t is

$$f(t) = \lambda_t S(t)$$

where λ_t is the population hazard rate and $S(t) = \exp(-\sum_i^t \lambda_i)$, assuming age is partitioned into one year increments. Assuming multiplicative changes to λ_t to adapt our weights to the age of diagnosis among the cases, we modeled the hazards as $\lambda_t e^{\beta_t}$, where β_t are parameters to estimate. The likelihood for the cases is

$$L = \prod_{t=30}^{87} \left(\lambda_t e^{\beta_t} \exp(-\sum_i^t \lambda_i e^{\beta_i}) \right)^{N_t}$$

Where $30 \leq t \leq 87$ for our cases and N_t is the number of cases at age of diagnosis t . We estimated the β_t parameters by the Newton-Raphson method, and use the revised hazard rates $\lambda_t e^{\beta_t}$ to weight the controls (weight of $1 / (\lambda_t e^{\beta_t})$), and used ancestry-specific λ_t and β_t in these computations.

OncoArray: Prostate Cancer Acknowledgments

Aarhus: This study was supported by the Danish Strategic Research Council (now Innovation Fund Denmark) and the Danish Cancer Society. The Danish Cancer Biobank (DCB) is acknowledged for biological material.

AHS: This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics (Z01CP010119).

ATBC: This research was supported in part by the Intramural Research Program of the NIH and the National Cancer Institute. Additionally, this research was supported by U.S. Public Health Service contracts N01-CN-45165, N01-RC-45035, N01-RC-37004, HHSN26120100006C, and HHSN261201500005C from the National Cancer Institute, Department of Health and Human Services.

BioVu: The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the National Center for Research Resources, Grant UL1 RR024975-01 (which is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06).

Canary PASS: PASS was supported by Canary Foundation and the National Cancer Institute's Early Detection Research Network (U01 CA086402)

CCI: This work was awarded by Prostate Cancer Canada and is proudly funded by the Movember Foundation - Grant # D2013-36. The CCI group would like to thank David Murray, Razmik Mirzayans, and April Scott for their contribution to this work.

CeRePP: None reported

COH: SLN is partially supported by the Morris and Horowitz Families Endowed Professorship

COSM: The Swedish Research Council, the Swedish Cancer Foundation

CPCS1 & CPCS2: Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark CPCS1 would like to thank the participants and staff of the Copenhagen General Population Study for their important contributions.

CPDR: Uniformed Services University for the Health Sciences HU0001-10-2-0002 (PI: David G. McLeod, MD)

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study II cohort. CPS-II thanks the participants and Study Management Group for their invaluable contributions to this research. We would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

EPIC: The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by the Danish Cancer Society (Denmark); the Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation, Greek Ministry of

Health; Greek Ministry of Education (Greece); the Italian Association for Research on Cancer (AIRC) and National Research Council (Italy); the Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF); the Statistics Netherlands (The Netherlands); the Health Research Fund (FIS), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, Spanish Ministry of Health ISCIII RETIC (RD06/0020), Red de Centros RCESP, C03/09 (Spain); the Swedish Cancer Society, Swedish Scientific Council and Regional Government of Skåne and Västerbotten, Fundacion Federico SA (Sweden); the Cancer Research UK, Medical Research Council (United Kingdom).

EPICAP: The EPICAP study was supported by grants from Ligue Nationale Contre le Cancer, Ligue départementale du Val de Marne; Fondation de France; Agence Nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES)The EPICAP study group would like to thank all urologists, Antoinette Anger and Hasina Randrianasolo (study monitors), Anne-Laure Astolfi, Coline Bernard, Oriane Noyer, Marie-Hélène De Campo, Sandrine Margaroline, Louise N'Diaye, Sabine Perrier-Bonnet (Clinical Research nurses)

ERSPC: This study was supported by the DutchCancerSociety(KWF94-869,98-1657,2002-277,2006-3518, 2010-4800); The Netherlands Organisation for HealthResearch and Development (ZonMW-002822820,22000106,50-50110-98-311, 62300035), The Dutch Cancer Research Foundation(SWOP), and an unconditional grant from Beckman-Coulter-HybritechInc.

ESTHER: The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts.The ESTHER group would like to thank Hartwig Ziegler, Sonja Wolf, Volker Hermann, Heiko Müller, Karina Dieffenbach, Katja Butterbach for valuable contributions to the study.

FHCRC: The FHCRC studies were supported by grants R01-CA056678, R01-CA082664, and R01-CA092579 from the US National Cancer Institute, National Institutes of Health, with additional support from the Fred Hutchinson Cancer Research Center. FHCRC would like to thank all the men who participated in these studies.

Gene-PARE: The Gene-PARE study was supported by grants 1R01CA134444 from the U.S. National Institutes of Health , PC074201 and W81XWH-15-1-0680 from the Prostate Cancer Research Program of the Department of Defense and RSGT-05-200-01-CCE from the American Cancer Society

Hamburg-Zagreb: None reported

HPFS: The Health Professionals Follow-up Study was supported by grants UM1CA167552, CA133891, CA141298, and P01CA055075.HPFS are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

IMPACT: The IMPACT study was funded by The Ronald and Rita McAulay Foundation, CR-UK Project grant (C5047/A1232), Cancer Australia, AICR Netherlands A10-0227, Cancer Australia and Cancer Council Tasmania, NIHR, EU Framework 6, Cancer Councils of Victoria and South Australia, Philanthropic donation to Northshore University Health System. We acknowledge support from the National Institute for Health Research (NIHR) to the Biomedical Research Centre at The Institute of Cancer Research and Royal Marsden Foundation NHS Trust. IMPACT acknowledge the IMPACT study steering committee, collaborating centres and participants.

IPO-Porto: The IPO-Porto study was funded by Fundação para a Ciência e a Tecnologia (FCT; UID/DTP/00776/2013 and PTDC/DTP-PIC/1308/2014) and by IPO-Porto Research Center (CI-IPOP-16-2012 and CI-IPOP-24-2015). MC and MPS are research fellows from Liga Portuguesa Contra o Cancro, Núcleo Regional do Norte. SM is a research fellow from FCT (SFRH/BD/71397/2010). IPO-Porto would like to express our gratitude to all patients and families who have participated in this study.

Karuprostate: The Karuprostate study was supported by the the Frech National Health Directorate and by the Association pour la Recherche sur les Tumeurs de la Prostate Karuprostate thanks Séverine Ferdinand

KULEUVEN: F.C. and S.J. are holders of grants from FWO Vlaanderen (G.0684.12N and G.0830.13N), the Belgian federal government (National Cancer Plan KPC_29_023), and a Concerted Research Action of the KU Leuven (GOA/15/017). TVDB is holder of of a doctoral fellowship of the FWO.

LAAPC: This study was funded by grant R01CA84979 (to S.A. Ingles) from the National Cancer Institute, NIH.

Malaysia: The study was funded by the University Malaya High Impact Research Grant (HIR/MOHE/MED/35). Malaysia thanks all associates in the Urology Unit, University of Malaya, Cancer Research Initiatives Foundation (CARIF) and the Malaysian Men's Health Initiative (MMHI).

MCCS: MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.

MCC-Spain: The study was partially funded by the Accion Transversal del Cancer, approved on the Spanish Ministry Council on the 11th October 2007, by the Instituto de Salud Carlos III-FEDER (PI08/1770, PI09/00773-Cantabria, PI11/01889-FEDER, PI12/00265, PI12/01270, PI12/00715), by the Fundación Marqués de Valdecilla (API 10/09), by the Spanish Association Against Cancer (AECC) Scientific Foundation and by the Catalan Government DURSI grant 2009SGR1489. Samples: Biological samples were stored at the Parc de Salut MAR Biobank (MARBiobanc; Barcelona) which is supported by Instituto de Salud Carlos III FEDER (RD09/0076/00036). Also sample collection was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncologia de Catalunya (XBTC). MCC-Spain acknowledges the contribution from Esther Gracia-Lavedan in preparing the data. We thank all the subjects who participated in the study and all MCC-Spain collaborators.

MD Anderson: Prostate Cancer Case-Control Studies at MD Anderson (MDA) supported by grants CA68578, ES007784, DAMD W81XWH-07-1-0645 and CA140388.

MDACC_AS: None reported

MEC: Funding provided by NIH grant U19CA148537 and grant U01 CA 164973.

MIAMI (WFPCS): ACS

MOFFITT: The Moffitt group was supported by the US National Cancer Institute (R01CA128813, PI: J.Y. Park).

NMHS: Funding for the Nashville Men's Health Study (NMHS) was provided by the National Institutes of Health Grant numbers: RO1CA121060

PCaP only data: The North Carolina - Louisiana Prostate Cancer Project (PCaP) is carried out as a collaborative study supported by the Department of Defense contract DAMD 17-03-2-0052. For HCaP-NC follow-up data: The Health Care Access and Prostate Cancer Treatment in North Carolina (HCaP-NC) study is carried out as a collaborative study supported by the American Cancer Society award RSGT-08-008-01-CPHPS. For studies using both PCaP and HCaP-NC follow-up data please use: The North Carolina - Louisiana Prostate Cancer Project (PCaP) and the Health Care Access and Prostate Cancer Treatment in North Carolina (HCaP-NC) study are carried out as collaborative studies supported by the Department of Defense contract DAMD 17-03-2-0052 and the American Cancer Society award RSGT-08-008-01-CPHPS, respectively. For any PCaP data please include: The authors thank the staff, advisory committees and research subjects participating in the PCaP study for their important contributions. For studies using PCaP DNA/genotyping data please include: We would like to acknowledge the UNC BioSpecimen Facility and the LSUHSC Pathology Lab for our DNA extractions, blood processing, storage and sample disbursement (<https://genome.unc.edu/bsp>). For studies using PCaP tissue please include: We would like to acknowledge the RPCI Department of Urology Tissue Microarray and Immunoanalysis Core for our tissue processing, storage and sample disbursement. For studies using HCaP-NC follow-up data please use: The Health Care Access and Prostate Cancer Treatment in North Carolina (HCaP-NC) study is carried out as a collaborative study supported by the American Cancer Society award RSGT-08-008-01-CPHPS. The authors thank the staff, advisory committees and research subjects participating in the HCaP-NC study for their important contributions. For studies that use both PCaP and HCaP-NC please use: The authors thank the staff, advisory committees and research subjects participating in the PCaP and HCaP-NC studies for their important contributions.

PCMUS: The PCMUS study was supported by the Bulgarian National Science Fund, Ministry of Education and Science (contract DOO-119/2009; DUNK01/2-2009; DFNI-B01/28/2012) with additional support from the Science Fund of Medical University - Sofia (contract 51/2009; 8I/2009; 28/2010;).

PHS: The Physicians' Health Study was supported by grants CA34944, CA40360, CA097193, HL26490 and HL34595. PHS members are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

PLCO: This PLCO study was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH. PLCO thanks Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention at the National Cancer Institute, the screening center investigators and staff of the PLCO Cancer Screening Trial for their contributions to the PLCO Cancer Screening Trial. We thank Mr. Thomas Riley, Mr. Craig Williams, Mr. Matthew Moore, and Ms. Shannon Merkle at Information Management Services, Inc., for their management of the data and Ms. Barbara O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial. We also thank the PLCO study participants for their contributions to making this study possible.

Poland: None reported

PROCAP: PROCAP was supported by the Swedish Cancer Foundation (08-708, 09-0677). PROCAP thanks and acknowledges all of the participants in the PROCAP study. We thank Carin Cavalli-Björkman and Ami Rönnberg Karlsson for their dedicated work in the collection of data. Michael Broms is acknowledged for his skilful work with the databases. KI Biobank is acknowledged for handling the samples and for DNA extraction. We acknowledge The NPCR steering group: Pär Stattin (chair), Anders Widmark, Stefan Karlsson, Magnus Törnblom, Jan Adolfsson, Anna Bill-Axelsson, Ove Andrén, David

Robinson, Bill Pettersson, Jonas Hugosson, Jan-Erik Damber, Ola Bratt, Göran Ahlgren, Lars Egevad, and Roy Ehrnström.

PROGReSS: The PROGReSS study was supported by grants from the Instituto de Salud Carlos III (FIS PI10/00164 and FIS PI13/02030) and Fondo Europeo de Desarrollo Regional (FEDER 2007-2013).

ProMPT : Founded by CRUK, NIHR, MRC, Cambridge Biomedical Research Centre

ProtecT: Founded by NIHR ProtecT and ProMPT would like to acknowledge the support of The University of Cambridge, Cancer Research UK. Cancer Research UK grants [C8197/A10123] and [C8197/A10865] supported the genotyping team. We would also like to acknowledge the support of the National Institute for Health Research which funds the Cambridge Bio-medical Research Centre, Cambridge, UK. We would also like to acknowledge the support of the National Cancer Research Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT) collaborative (grant code G0500966/75466) which has funded tissue and urine collections in Cambridge. We are grateful to staff at the Wellcome Trust Clinical Research Facility, Addenbrooke's Clinical Research Centre, Cambridge, UK for their help in conducting the ProtecT study. We also acknowledge the support of the NIHR Cambridge Biomedical Research Centre, the DOH HTA (ProtecT grant) and the NCRI / MRC (ProMPT grant) for help with the bio-repository. The UK Department of Health funded the ProtecT study through the NIHR Health Technology Assessment Programme (projects 96/20/06, 96/20/99). The ProtecT trial and its linked ProMPT and CAP (Comparison Arm for ProtecT) studies are supported by Department of Health, England; Cancer Research UK grant number C522/A8649, Medical Research Council of England grant number G0500966, ID 75466 and The NCRI, UK. The epidemiological data for ProtecT were generated through funding from the Southwest National Health Service Research and Development. DNA extraction in ProtecT was supported by USA Dept of Defense award W81XWH-04-1-0280, Yorkshire Cancer Research and Cancer Research UK. The authors would like to acknowledge the contribution of all members of the ProtecT study research group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Department of Health of England. The bio-repository from ProtecT is supported by the NCRI (ProMPT) Prostate Cancer Collaborative and the Cambridge BMRC grant from NIHR.

PROtEuS: PROtEuS was supported financially through grants from the Canadian Cancer Society [13149, 19500, 19864, 19865] and the Cancer Research Society, in partnership with the Ministère de l'enseignement supérieur, de la recherche, de la science et de la technologie du Québec, and the Fonds de la recherche du Québec - Santé. PROtEuS would like to thank its collaborators and research personnel, and the urologists involved in subjects recruitment. We also wish to acknowledge the special contribution made by Ann Hsing and Anand Chokkalingam to the conception of the genetic component of PROtEuS.

QLD: The QLD research is supported by The National Health and Medical Research Council (NHMRC) Australia Project Grants [390130, 1009458] and NHMRC Career Development Fellowship and Cancer Australia PdCCRS funding to J Batra. The QLD research is supported by The National Health and Medical Research Council (NHMRC) Australia Project Grants [390130, 1009458] and NHMRC Career Development Fellowship and Cancer Australia PdCCRS funding to J Batra.

RAPPER: RAPPER is funded by Cancer Research UK [C1094/A11728; C1094/A18504] and Experimental Cancer Medicine Centre funding [C1467/A7286]. The RAPPER group thank Rebecca Elliott for project management.

SABOR: The SABOR research is supported by NIH/NCI Early Detection Research Network, grant U01 CA086640-12. Also supported by the Cancer Center Support Grant to the Cancer Therapy and Research Center from the National Cancer Institute (US) P30 CA054174

SCCS: SCCS is funded by NIH grant R01 CA092447, and SCCS sample preparation was conducted at the Epidemiology Biospecimen Core Lab that is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry.

SCPCS: SCPCS is funded by CDC grant S1135-19/19, and SCPCS sample preparation was conducted at the Epidemiology Biospecimen Core Lab that is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485).

SEARCH: SEARCH is funded by a programme grant from Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge.

SNP_Prostate_Ghent: The study was supported by the National Cancer Plan, financed by the Federal Office of Health and Social Affairs, Belgium.

SPAG: Wessex Medical Research Hope for Guernsey, MUG, HSSD, MSG, Roger Allsopp

STHM2: STHM2 was supported by grants from The Strategic Research Programme on Cancer (StratCan), Karolinska Institutet; the Linné Centre for Breast and Prostate Cancer (CRISP, number 70867901), Karolinska Institutet; The Swedish Research Council (number K2010-70X-20430-04-3) and The Swedish Cancer Society (numbers 11-0287 and 11-0624); Stiftelsen Johanna Hagstrand och Sigfrid Linnérs minne; Swedish Council for Working Life and Social Research (FAS), number 2012-0073. STHM2 acknowledges the Karolinska University Laboratory, Aleris Medilab, Unilabs and the Regional Prostate Cancer Registry for performing analyses and help to retrieve data. Carin Cavalli-Björkman and Britt-Marie Hune for their enthusiastic work as research nurses. Astrid Björklund for skilful data management. We wish to thank the BBMRI.se biobank facility at Karolinska Institutet for biobank services.

SWOG-PCPT & SELECT: PCPT & SELECT are funded by Public Health Service grants U10CA37429 and 5UM1CA182883 from the National Cancer Institute. SWOG and SELECT thank the site investigators and staff and, most importantly, the participants who donated their time to this trial.

TAMPERE: The Tampere (Finland) study was supported by the Academy of Finland (251074), The Finnish Cancer Organisations, Sigrid Juselius Foundation, and the Competitive Research Funding of the Tampere University Hospital (X51003). The PSA screening samples were collected by the Finnish part of ERSPC (European Study of Screening for Prostate Cancer). TAMPERE would like to thank Riina Liikanen, Liisa Maaettaenen and Kirsi Talala for their work on samples and databases.

UGANDA: None reported

UKGPCS: UKGPCS would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. UKGPCS should also like to acknowledge the NCRN nurses, data managers and Consultants for their work in the UKGPCS study. UKGPCS would like to thank all urologists and other persons involved in the planning, coordination, and data collection of the CAPS study.

ULM: The Ulm group received funds from the German Cancer Aid (Deutsche Krebshilfe).

WUGS/WUPCS: WUGS would like to thank the following for funding support: The Anthony DeNovi Fund, the Donald C. McGraw Foundation, and the St. Louis Men's Group Against Cancer.

References

1. Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., et al. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet* 50, 1219-1224.
2. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., et al. (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 39, 645-649.
3. Kolonel, L.N., Henderson, B.E., Hankin, J.H., Nomura, A.M., Wilkens, L.R., Pike, M.C., Stram, D.O., Monroe, K.R., Earle, M.E., and Nagamine, F.S. (2000). A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 151, 346-357.
4. Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirmer, C., McDonald, G.J., et al. (2007). Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39, 638-644.
5. Siddiq, A., Couch, F.J., Chen, G.K., Lindstrom, S., Eccles, D., Millikan, R.C., Michailidou, K., Stram, D.O., Beckmann, L., Rhie, S.K., et al. (2012). A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* 21, 5373-5384.
6. Schumacher, F.R., Berndt, S.I., Siddiq, A., Jacobs, K.B., Wang, Z., Lindstrom, S., Stevens, V.L., Chen, C., Mondul, A.M., Travis, R.C., et al. (2011). Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* 20, 3867-3875.
7. Garcia-Closas, M., Brinton, L.A., Lissowska, J., Chatterjee, N., Peplonska, B., Anderson, W.F., Szeszenia-Dabrowska, N., Bardin-Mikolajczak, A., Zatonski, W., Blair, A., et al. (2006). Established breast cancer risk factors by clinically important tumour characteristics. *Br J Cancer* 95, 123-129.
8. Cook, M.B., Wang, Z., Yeboah, E.D., Tettey, Y., Biritwum, R.B., Adjei, A.A., Tay, E., Truelove, A., Niwa, S., Chung, C.C., et al. (2014). A genome-wide association study of prostate cancer in West African men. *Hum Genet* 133, 509-521.
9. Al Olama, A.A., Kote-Jarai, Z., Berndt, S.I., Conti, D.V., Schumacher, F., Han, Y., Benlloch, S., Hazelett, D.J., Wang, Z., Saunders, E., et al. (2014). A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 46, 1103-1109.
10. Wang, Z., Zhu, B., Zhang, M., Parikh, H., Jia, J., Chung, C.C., Sampson, J.N., Hoskins, J.W., Hutchinson, A., Burdette, L., et al. (2014). Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1L region on chromosome 5p15.33. *Hum Mol Genet* 23, 6616-6633.
11. Hsing, A.W., Yeboah, E., Biritwum, R., Tettey, Y., De Marzo, A.M., Adjei, A., Netto, G.J., Yu, K., Li, Y., Chokkalingam, A.P., et al. (2014). High prevalence of screen detected prostate cancer in West Africans: implications for racial disparity of prostate cancer. *J Urol* 192, 730-735.
12. Chokkalingam, A.P., Yeboah, E.D., Demarzo, A., Netto, G., Yu, K., Biritwum, R.B., Tettey, Y., Adjei, A., Jadallah, S., Li, Y., et al. (2012). Prevalence of BPH and lower urinary tract symptoms in West Africans. *Prostate Cancer Prostatic Dis* 15, 170-176.

13. Chu, L.W., Ritchey, J., Devesa, S.S., Quraishi, S.M., Zhang, H., and Hsing, A.W. (2011). Prostate cancer incidence rates in Africa. *Prostate Cancer* 2011, 947870.
14. Chung, C.C., Hsing, A.W., Edward, Y., Biritwum, R., Tettey, Y., Adjei, A., Cook, M.B., De Marzo, A., Netto, G., Tay, E., et al. (2014). A comprehensive resequence-analysis of 250 kb region of 8q24.21 in men of African ancestry. *Prostate* 74, 579-589.
15. Berndt, S.I., Wang, Z., Yeager, M., Alavanja, M.C., Albanes, D., Amundadottir, L., Andriole, G., Beane Freeman, L., Campa, D., Cancel-Tassin, G., et al. (2015). Two susceptibility loci identified for prostate cancer aggressiveness. *Nat Commun* 6, 6889.
16. Bailey-Wilson, J.E., Childs, E.J., Cropp, C.D., Schaid, D.J., Xu, J., Camp, N.J., Cannon-Albright, L.A., Farnham, J.M., George, A., Powell, I., et al. (2012). Analysis of Xq27-28 linkage in the international consortium for prostate cancer genetics (ICPCG) families. *BMC Med Genet* 13, 46.
17. Lu, L., Cancel-Tassin, G., Valeri, A., Cussenot, O., Lange, E.M., Cooney, K.A., Farnham, J.M., Camp, N.J., Cannon-Albright, L.A., Tammela, T.L., et al. (2012). Chromosomes 4 and 8 implicated in a genome wide SNP linkage scan of 762 prostate cancer families collected by the ICPCG. *Prostate* 72, 410-426.
18. Xu, J., Lange, E.M., Lu, L., Zheng, S.L., Wang, Z., Thibodeau, S.N., Cannon-Albright, L.A., Teerlink, C.C., Camp, N.J., Johnson, A.M., et al. (2013). HOXB13 is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 132, 5-14.
19. Teerlink, C.C., Leongamornlert, D., Dadaev, T., Thomas, A., Farnham, J., Stephenson, R.A., Riska, S., McDonnell, S.K., Schaid, D.J., Catalona, W.J., et al. (2016). Genome-wide association of familial prostate cancer cases identifies evidence for a rare segregating haplotype at 8q24.21. *Hum Genet* 135, 923-938.
20. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655-1664.
21. Bergstrom, A., McCarthy, S.A., Hui, R.Y., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, 1339-+.