

# UCSF

## UC San Francisco Previously Published Works

### Title

A genome-wide association study of prostate cancer in Latinos

### Permalink

<https://escholarship.org/uc/item/6845042z>

### Journal

International Journal of Cancer, 146(7)

### ISSN

0020-7136

### Authors

Du, Zhaohui  
Hopp, Hannah  
Ingles, Sue A  
[et al.](#)

### Publication Date

2020-04-01

### DOI

10.1002/ijc.32525

Peer reviewed

# A genome-wide association study of prostate cancer in Latinos

Zhaohui Du<sup>1</sup>, Hannah Hopp<sup>1</sup>, Sue A. Ingles<sup>1</sup>, Chad Huff<sup>2</sup>, Xin Sheng<sup>1</sup>, Brandi Weaver<sup>3</sup>, Mariana Stern<sup>1</sup>, Thomas J. Hoffmann<sup>4,5</sup>, Esther M. John<sup>6</sup>, Stephen K. Van Den Eeden<sup>7,8</sup>, Sara Strom<sup>2</sup>, Robin J. Leach<sup>3</sup>, Ian M. Thompson Jr.<sup>3</sup>, John S. Witte<sup>4,5,8</sup>, David V. Conti<sup>1,9</sup> and Christopher A. Haiman<sup>1,9</sup>

<sup>1</sup>Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA

<sup>2</sup>The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>3</sup>Department of Urology, University of Texas Health Science Center, San Antonio, TX

<sup>4</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA

<sup>5</sup>Institute for Human Genetics, University of California, San Francisco, San Francisco, CA

<sup>6</sup>Department of Medicine and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA

<sup>7</sup>Division of Research, Kaiser Permanente, Northern California, Oakland, CA

<sup>8</sup>Department of Urology, University of California San Francisco, San Francisco, CA

<sup>9</sup>Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA

Latinos represent <1% of samples analyzed to date in genome-wide association studies of cancer. The clinical value of genetic information in guiding personalized medicine in populations of non-European ancestry will require additional discovery and risk locus characterization efforts across populations. In the present study, we performed a GWAS of prostate cancer (PrCa) in 2,820 Latino PrCa cases and 5,293 controls to search for novel PrCa risk loci and to examine the generalizability of known PrCa risk loci in Latino men. We also conducted a genetic admixture-mapping scan to identify PrCa risk alleles associated with local ancestry. Genome-wide significant associations were observed with 84 variants all located at the known PrCa risk regions at 8q24 (128.484–128.548) and 10q11.22 (*MSMB* gene). In admixture mapping, we observed genome-wide significant associations with local African ancestry at 8q24. Of the 162 established PrCa risk variants that are common in Latino men, 135 (83.3%) had effects that were directionally consistent as previously reported, among which 55 (34.0%) were statistically significant with  $p < 0.05$ . A polygenic risk model of the known PrCa risk variants showed that, compared to men with average risk (25th–75th percentile of the polygenic risk score distribution), men in the top 10% had a 3.19-fold (95% CI: 2.65, 3.84) increased PrCa risk. In conclusion, we found that the known PrCa risk variants can effectively stratify PrCa risk in Latino men. Larger studies in Latino populations will be required to discover and characterize genetic risk variants for PrCa and improve risk stratification for this population.

## Introduction

Prostate cancer (PrCa) is the most common nonskin cancer and the second leading cause of cancer death among men in the U.S.,

with large differences in incidence rates observed across ethnic groups.<sup>1</sup> Age-adjusted incidence rates (per 100,000) are highest in African Americans (AA; 178.3), lower in non-Hispanic whites

**Key words:** prostate cancer, Latino men, GWAS, admixture mapping

**Abbreviations:** 1KGP: 1000 Genomes Project; AA: African Americans; AFR: African; AMR: Amerindian; BMI: body mass index; CI: confidence interval; EUR: European; GERA: Genetic Epidemiology Research on Aging study; GWAS: genome-wide association study; HWE: Hardy–Weinberg equilibrium; KP: Kaiser Permanente; KPNCCR: KP Northern California Cancer Registry; KPSCCR: KP Southern California Cancer Registry; LAAPC: Los Angeles Aggressive Prostate Cancer Study; MAF: minor allele frequency; MDA: MD Anderson; MEC: multiethnic cohort; NHGRI: National Human Genome Research Institute; NHW: non-Hispanic whites; OR: odds ratio; PAGE: Population Architecture using Genomics and Epidemiology (PAGE) Consortium; PC: principle component; PrCa: prostate cancer; PRS: polygenic risk score; PSA: prostate-specific antigen; QC: quality control; RAF: risk allele frequency; RPGEH: Research Program on Genes, Environment and Health cohort; SABOR: San Antonio Biomarkers of Risk; SNP: signal nucleotide polymorphism

Additional Supporting Information may be found in the online version of this article.

**Grant sponsor:** National Cancer Institute; **Grant numbers:** R01 CA84979, U01 CA086402, U01 CA164973

**DOI:** 10.1002/ijc.32525

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

**History:** Received 19 Feb 2019; Accepted 15 May 2019; Online 21 Jun 2019

**Correspondence to:** Christopher A. Haiman, Harlyne Norris Research Tower, 1450 Biggy Street, Room 1504, Los Angeles, CA 90033, USA, Tel.: +1-323-442-7755, Fax: +1-323-442-7749, E-mail: haiman@usc.edu

**What's new?**

There is strong evidence for a genetic predisposition to prostate cancer (PrCa). Most of this information has come from European ancestry populations, with Latinos representing less than 1% of samples in cancer genome-wide association studies (GWAS). In this study, the majority of established PrCa risk variants (83.3%) were consistently associated with PrCa risk in Latinos. A polygenic risk score comprised of GWAS-identified risk variants could identify 10% of Latino men with a ~three-fold increase in PrCa risk. These findings suggest that common germline variants for PrCa can stratify risk in Latino men, which has implications for targeted screening and prevention.

(NHW; 105.7), and slightly lower still in Hispanics/Latinos (91.8).<sup>1,2</sup> In the only prospective study of PrCa in Latinos, risk was observed to be higher among Latinos compared to NHW after adjustment for potential confounders, including lifestyle factors and PSA screening history.<sup>3</sup> Though classified as a single ethnic group, the Latino population consists of genetically admixed individuals from populations that display considerable diversity in PrCa incidence and mortality rates. For example, analyses of cancer registry data in Florida revealed that Latinos of Mexican origin had a remarkably lower age-adjusted incidence rate of PrCa compared to those of Cuban or Puerto Rican origin or to NHW,<sup>4</sup> whereas Latinos with Dominican and Cuban origins had a significantly higher PrCa mortality rate compared to NHW.<sup>5</sup> Possible explanations for these differences include variation across subgroups in place of birth, acculturation, socioeconomic status (SES), access to care, lifestyle factors, and genetic ancestry and susceptibility. Latinos are extensively admixed from multiple ancestries including Amerindian (AMR), European (EUR) and African (AFR),<sup>6</sup> with large variation in ancestry proportions observed across subgroups and individuals. For example, the proportion of AFR ancestry is small among Mexicans (<10%) but quite large in Dominicans and Puerto Ricans (20–40%). Throughout the Americas, and even within a single country, AMR ancestry proportions vary widely.<sup>6,7</sup>

GWAS in non-European populations have provided insight into ancestry-specific variation and have revealed regions of susceptibility that are of particular importance in certain populations. For example, GWAS in Latinos of phenotypes such as central corneal thickness,<sup>8</sup> asthma<sup>9</sup> and diabetes,<sup>10</sup> have discovered novel susceptibility not reported in other populations. For breast cancer, an admixture-mapping study discovered higher AMR ancestry at chromosome 6q25 and protective variants within this region<sup>11</sup> that are only found in AMR populations. Genetic studies of PrCa in Latino men have been limited but are needed to evaluate the potential for novel germline variants for PrCa risk in men of AMR ancestry and to test the generalizability of established PrCa genetic markers in this admixed population. The extensive diversity of ancestry proportions within Latinos also provides the opportunity to investigate the interaction between genetic background and genetic risk loci on disease risk.<sup>12,13</sup>

In the present study, we carried out a GWAS of PrCa in Latinos to search for novel risk alleles and to examine how well the known PrCa risk alleles may stratify PrCa risk in Latino men. We also leveraged genetic admixture to conduct a

genome-wide admixture mapping analysis to scan for PrCa risk alleles associated with local ancestry. In addition to generating a polygenic risk score (PRS) to test the cumulative effect of all known PrCa risk variants in Latinos, we also explored whether genetic background/ancestry modified associations with single variants and a PRS for PrCa.

**Materials and Methods****Study participants and GWAS genotyping**

Our study includes Latino PrCa cases and controls from five studies that were genotyped with different GWAS array platforms and denoted as Sets 1–3.

Set 1 consisted of 1,079 incident Latino PrCa cases and 1,083 controls from the multiethnic cohort (MEC).<sup>14</sup> In brief, the MEC is a large population-based cohort study including 215,251 men and women recruited from Hawaii and California between 1993 and 1996. Incident Latino PrCa cases were identified by linking with the cancer registries in Hawaii and California. Controls were men with no prostate cancer diagnosis that were selected from a control pool who provided specimens for genetic analysis and were frequency-matched to cases ( $\pm 5$  years). Genotyping of Set 1 was performed with the Illumina Human660W array [database of Genotypes and Phenotypes (dbGaP) phs000306.v3.p1].<sup>14</sup>

Set 2 included 1,253 cases and 1,069 controls from four studies: the MEC, the Los Angeles Aggressive Prostate Cancer (LAAPC), MD Anderson (MDA) and San Antonio Biomarkers of Risk (SABOR). These studies were genotyped with the Illumina OncoArray (260K GWAS backbone),<sup>15</sup> as part of the ELLIPSE GAME-ON Consortium (dbGaP phs001391.v1.p1).<sup>16</sup> The MEC included 152 incident and prevalent Latino PrCa cases and 162 controls (not included in Set 1). The LAAPC is a population-based case-control study of aggressive prostate cancer in Los Angeles County.<sup>17</sup> Eligible cases ( $n = 320$ ) were Latinos of any age diagnosed with primary prostate cancer. Controls ( $n = 331$ ) were Latino men without PrCa diagnosis and were frequency matched with cases on age ( $\pm 5$  years), who were identified *via* a neighborhood walk algorithm.<sup>18</sup> MDA cases ( $n = 521$ ) were Latino men enrolled in epidemiological PrCa studies conducted at the University of Texas M.D. Anderson Cancer Center.<sup>19,20</sup> Controls ( $n = 316$ ) were men of self-reported Mexican origin recruited by random digit dialing in Texas<sup>20</sup> or enrolled in the Mexican American Cohort Study, an ongoing population-based cohort in Houston, TX.<sup>21</sup> MDA controls had no diagnosis of invasive cancer and were frequency matched with cases on age

( $\pm 5$  years). SABOR is a cohort study which has been enrolling healthy male volunteers in San Antonio and South Texas area since 2001.<sup>22,23</sup> Participants were examined annually/biannually by digital rectal exam and serum prostate-specific antigen level, and prostate biopsy was recommended for men with positive results. In total, 260 incident Latino PrCa cases, who had been biopsy-confirmed, were enrolled. Controls ( $n = 260$ ) were Latino men  $\geq 45$  years old who had normal digital rectal exams and prostate-specific antigen levels  $\leq 2.5$  ng/ml on all annual visits.

Set 3 included 488 Latino PrCa cases and 3,141 controls from three cohorts within Kaiser Permanente (KP), an integrated health care delivery system: the Research Program on Genes, Environment and Health (RPGEH) cohort, the ProHealth Study, and the California Men's Health Study. Incident PrCa cases were identified from the KP Northern California Cancer Registry (KPNCCR), the KP Southern California Cancer Registry (KPSCCR) or through review of clinical electronic health records by the end of 2012. Controls were all Latino men in RPGEH Genetic Epidemiology Research on Aging (GERA) study without PrCa diagnosis. These studies were genotyped using the Affymetrix Axiom v2 reagent as previously described (dbGaP phs000674.p1).<sup>24</sup>

#### Genotyping quality control, imputation and GWAS analysis

In Set 1, samples were excluded based on call rate  $< 95\%$  and first-degree relatedness, with the final analysis sample size of 2,080 (1,034 cases and 1,046 controls). SNPs with call rate  $< 0.95$  or with MAF  $< 1\%$  were excluded and 528,023 SNPs were retained for imputation. Imputation was performed using the cosmopolitan reference panel in the 1,000 Genomes Project (1KGP) using Minimac3 Version 1.0.12. A total of 10,441,344 SNPs with MAF  $\geq 0.01$  and imputation quality score  $\geq 0.3$  were included in the analysis. Principal components (PC) were estimated using EIGENSTRAT<sup>25</sup> and per-allele odds ratios (OR) and  $p$  values were estimated using unconditional logistic regression for each SNP, adjusting for age and the first 10 PCs.

In Set 2, genotyping quality control (QC) was conducted together with a larger number of samples from the ELLIPSE consortium as described previously.<sup>15,16</sup> Briefly, samples were removed if they were gender/sex mismatches ( $n = 6$ ), first-degree relative pairs ( $n = 19$ ) or had a call rate  $< 0.95$  ( $n = 9$ ). We calculated the shared IBD for Set 1 and Set 2 using PLINK to remove related samples across sets. We further excluded 43 cases and 1 control from Set 2, leaving a final sample size of 2,244 (1,192 cases and 1,052 controls). We excluded SNPs with call rate  $< 0.95$  or replicate concordance  $< 99.8\%$  based on QC replicate samples, or due to poor clustering after visual inspection. We further removed SNPs with estimated MAF that deviated or had mismatched alleles in comparison to the AMR individuals in phase III 1KGP data; 456,809 SNPs were available for imputation. Imputation was performed with Minimac3 Version 1.0.12 using the phase III 1KGP cosmopolitan reference panel. A total of 10,595,258 SNPs with MAF  $\geq 0.01$  and imputation quality score

$\geq 0.3$  were included in the analysis. PCs were estimated using EIGENSTRAT and per-allele ORs and  $p$  values were estimated using unconditional logistic regression, adjusting for age, study and the first 10 PCs.

In Set 3, QC exclusions were based on call rate  $< 97\%$ , ancestry outliers and relatedness as previously described.<sup>24</sup> The final analysis sample size was 3,629 (488 cases and 3,141 controls). Problematic SNPs were removed if they had MAF  $< 1\%$ , call rate  $< 95\%$  or Hardy-Weinberg equilibrium (HWE)  $p < 1 \times 10^{-5}$ , leaving 568,496 SNPs for imputation. Imputation was performed to the phase III 1KGP using IMPUTE2 v2.3.1. A total of 10,748,756 SNPs with MAF  $\geq 0.01$  and imputation quality score  $\geq 0.3$  were included in the analysis. PCs were estimated using EIGENSTRAT v4.2. ORs and  $p$  values were estimated using unconditional logistic regression for each SNP, adjusting for age, body mass index (BMI) and the first 10 PCs.

#### Statistical analyses

A fixed-effect meta-analysis with inverse variance weights was used to obtain the combined results of the three sets for the overlapping SNPs ( $n = 10,330,976$ ). The combined sample size for meta-analysis was 2,714 cases and 5,239 controls. Risk allele frequencies (RAF) were derived by averaging the case/control RAFs of the three sets, weighted by the corresponding case/control numbers in each study. Regional association plots were generated using LocusZoom<sup>26</sup> for regions with genome-wide significant variants. All tests were two-sided with the genome-wide significance level being  $\alpha = 5.0 \times 10^{-8}$ . Unlike in Sets 1 and 2, Set 3 was additionally adjusted for BMI because it was found to be associated with PrCa risk in ProHealth. However, BMI was not found to confound the SNP associations or alter the PRS meta-analysis results (data not shown).

To assess the number of independent signals in the genome-wide significant risk regions, we performed forward-selection logistic regression in a pooled dataset of Set 1 and 2 (primary data were not available for Set 3), adjusting for global ancestry, age, and study. The correlation ( $r^2$ ) between the independent SNPs identified by the stepwise regression procedure, and previously reported PrCa risk alleles in these regions, was calculated within the phase III 1KGP AFR/AMR/EUR populations.

**Admixture analysis.** We performed an admixture-based genome-wide scan using the primary genotype data for Sets 1 and 2. We first computed PCs with the reference panels including 1KGP ( $n = 2,504$ ;  $n = 347$  AMR) and the National Human Genome Research Institute (NHGRI) Population Architecture using Genomics and Epidemiology (PAGE) Consortium reference panel ( $n = 1,553$ ;  $n = 630$  AMR) to visualize the ancestry distribution of our samples. We used individuals (i.e., European, African and Amerindian) of PAGE as the reference samples for local ancestry estimation. We conducted random sampling of the AMR population to get balanced sample sizes across ethnicities, leaving a total of 393 individuals in the final reference panel (AMR = 147, EUR = 150, AFR = 96). We estimated genome-wide local ancestry

using RFMix.<sup>27</sup> We calculated individual EUR/AMR/AFR global ancestry ( $Q_{EUR/AMR/AFR}$ ) by taking an average of an individual's local ancestry estimates across 1–22 autosome chromosomes. The association between global ancestry and PrCa risk was examined by a logistic regression model adjusting for age and study. We also tested the association between global ancestry and PrCa aggressiveness using case-only analysis adjusting for age and study. Aggressive PrCa was defined as cases with Gleason score  $\geq 8$ .

To search for regions of the genome where local ancestry (EUR vs. AMR vs. AFR) may be associated with PrCa risk, we regressed the difference between an individual's local ancestry from their global ancestry using linear regression and compared this difference between cases and controls adjusting for age, study and global ancestry. We also performed case-only analyses using linear regression adjusting for age and study, comparing a case's local ancestry with his global ancestry. A fixed-effect meta-analysis with inverse variance weights was conducted to combine results of Sets 1 and 2, using  $p < 1 \times 10^{-5}$  as criteria for genome-wide significance. Continuous regions (adjacent regions with  $p < 1 \times 10^{-4}$ ) that were significant in both case-only and case-control comparisons were considered suggestive PrCa risk associations. In order to assess whether the local ancestry signal could be explained by risk alleles within that region, we did two conditional analyses for local ancestry: one adjusting for the top independent risk alleles identified in our Latino GWAS, the other one adjusting for both independent Latino risk alleles and the known risk variants in the detected region.

**Association testing of known risk regions.** We examined the associations of the 181 established risk variants from previous PrCa GWAS and fine-mapping studies (Supporting Information Table S2). Consistent directionality of effect were alleles with ORs in the same direction as those previously described (i.e., OR > 1). A nominal  $p$ -value of 0.05 was used to determine statistical significance. For each risk loci, we tested the interaction between continuous local ancestry (AMR/EUR) estimates and risk allele dosage on PrCa risk. For alleles with a nominal significant interaction term ( $p_{\text{interaction}} < 0.05$ ), we conducted stratified analysis by local ancestry (i.e., number of AMR chromosomes:  $\leq 0.5$ , 0.5–1.5, >1.5).

**Polygenic risk score analyses.** The aggregate effect of the known risk alleles was examined using a weighted polygenic risk score (PRS),  $PRS_i = \sum_{m=1}^C \beta_m g_{im}$ , for each individual.  $g_{im}$  is the risk allele dosage for individual  $i$  at SNP  $m$ ;  $C$  defines a set of 176 reported risk loci with  $MAF \geq 0.001$  and imputation  $r^2 \geq 0.3$  in Latino men (five risk variants were excluded based on this criteria).  $\beta_m$  is the weight for SNP  $m$ . For an EUR-weighted PRS, weights were the conditional log ORs derived from men of European ancestry<sup>28</sup>; for a Latino-weighted PRS, weights were the conditional log ORs obtained from meta-analyses in Latino men (Sets 1 and 2). The PRS in each set (Sets 1 and 2) was categorized by percentile (<10, 10–25,

25–75, 75–90,  $\geq 90\%$ ), and the risk for each category was estimated relative to the interquartile range (25–75%) using logistic regression adjusting for the first 10 PCs, age and study. The estimates were then meta-analyzed using the *metafor* package in R. We also examined the association between PRS and PrCa risk by strata of EUR and AMR global ancestry.

## Results

Demographic and clinical characteristics of individuals in the study are presented in Supporting Information Table S1. The mean age of cases was 61.8–73.7 across studies with mean ages being comparable in controls. The frequency of cases with Gleason score  $\geq 8$  ranged from 13.4% to 33.6% across studies, with LAAPC (Set 2) containing a higher proportion of aggressive cases (by design). Family history was more common among cases than controls in all studies and was significantly associated with PrCa risk (OR = 2.8; 95% CI: 2.2, 3.5,  $p = 3.1 \times 10^{-16}$ ) after adjusting for age and study.

The degree of European/AmerIndian admixture in the Set 1 and Set 2 samples is shown in Supporting Information Figure S1, with the majority of the current study samples spread along the European and AmerIndian axis. The PAGE reference panel revealed two AMR clusters: with the Set 1 and 2 samples congregated more closely with samples from Venezuela/Colombia/Brazil/Mexico (vs. Peru; Supporting Information Fig. S2). European ancestry was the major ancestral component with average values ranging from 48.5% to 58.4% in controls across studies, followed by AmerIndian ancestry (36.9–46.7% in controls), with African ancestry being a minor component (4.7–5.7% in controls). AMR global ancestry was negatively associated with PrCa risk after adjusting for age and study, with a 0.1 increase in AMR ancestry percentage associated with a 16% decrease in PrCa risk (OR = 0.84; 95% CI: 0.81, 0.88,  $p = 1.01 \times 10^{-15}$ ). This difference between cases and controls was variable across studies (Supporting Information Fig. S3) and when excluding MDA, the inverse association between AMR and PrCa risk was attenuated (OR = 0.94 per 0.1 increase in AMR; 95% CI: 0.90, 0.99), but still statistically significant ( $p = 0.01$ ). AFR global ancestry was not significantly associated with PrCa risk. In the case-only analysis, neither AMR or AFR global ancestry was significantly associated with PrCa aggressiveness ( $p_{AMR} = 0.62$ ,  $p_{AFR} = 0.28$ ).

The GWAS meta-analysis indicated no evidence of inflation in association test statistics (e.g., due to confounding by population stratification;  $\lambda = 1.03$ ). Genome-wide statistically significant associations were detected with 84 variants in known risk regions at 10q11.22 (SNP  $n = 74$ ) and 8q24.21 (SNP  $n = 10$ ; Supporting Information Table S2 and Figs. S4–S6). The most statistically significant variant was the known risk allele rs10993994 (OR = 1.29; 95% CI: 1.19, 1.39,  $p = 1.08 \times 10^{-10}$ ) located upstream of *MSMB* at 10q11.22. At 8q24.21, the strongest association was with rs7843031 (OR = 1.53; 95% CI: 1.34, 1.74,  $p = 5.12 \times 10^{-10}$ ), which is highly correlated with the known risk variant rs7812894

( $r_{AFR}^2 = 0.57$ ,  $r_{EUR}^2 = 0.89$ ,  $r_{AMR}^2 = 0.83$ , 1KGP phase III) at 128.52 Mb. All other associated SNPs ( $p < 5 \times 10^{-8}$ ) at 10q11.22 and 8q24.21 were correlated with either rs7843031 ( $r^2 \geq 0.3$ ) or rs10993994 ( $r^2 \geq 0.6$ ). At 8q24, a second variant, rs56005245, was found to be independently associated with risk ( $p < 1 \times 10^{-5}$ ) from the forward selection procedure. Variant rs56005245 is highly correlated with the previously reported risk allele rs72725879 ( $r_{AFR}^2 = 0.11$ ,  $r_{EUR}^2 = 0.71$ ,  $r_{AMR}^2 = 0.29$ ).

### Admixture analysis

We found no genome-wide significant associations between local EUR or AMR ancestry and PrCa risk in the case-control or case-case analyses. We did detect genome-wide significant ( $p < 1 \times 10^{-5}$ ) PrCa risk associations with AFR local ancestry at the 8q24 PrCa susceptibility region (127.0–127.8 MB), and each AFR-derived chromosome at this region was associated with an average of 1.60-fold increased PrCa risk (95% CI: 1.31, 1.95); the continuous suggestive risk associations ( $p < 1 \times 10^{-4}$ ) extended from 126.9 to 128.1 Mb (Supporting Information Fig. S7). We performed a conditional analysis for AFR local ancestry in the genome-wide significant risk region, with additional adjustment for the two independent risk alleles rs7843031 and rs56005245 identified above. This resulted in a general increase of less than two orders of magnitude for the AFR local ancestry  $p$  values ( $p = 9.5 \times 10^{-5}$ – $5.9 \times 10^{-4}$ , Supporting Information Figure S8). Conditioning on all the 14 risk 8q24 variants (two independent and 12 known risk alleles<sup>16,29</sup>), each AFR-derived chromosome at 8q24 was associated with 1.30-fold increased PrCa risk (95% CI: 1.03, 1.61) and the increase of the  $p$  values were much greater ( $p = 0.02$ – $0.06$ ).

### Association testing of known risk regions

Of the 181 previously reported PrCa risk loci, one (rs138213197) was not imputed in Set 1 and Set 2; 162 were polymorphic with MAF  $\geq 0.01$  and imputation quality score  $\geq 0.3$  in all three sets of Latino men (Supporting Information Table S3). Of the remaining 162 variants, directional consistency was noted for 135 (83.3%) in

the meta-analysis, among which 55 (34.0%) were nominally significant ( $p < 0.05$ ). In comparing the frequency of the known risk alleles between populations, the average risk allele frequency in Latino controls was only 0.005 larger than that observed in the European population ( $p = 0.48$ ,  $t$ -test), with 18 (11.3%) having opposite minor alleles. Local ancestry was estimated for 157 autosomal risk alleles and 11 variants demonstrated nominally statistically significant ( $p < 0.05$ ) interactions between local ancestry (EUR or AMR) and risk allele on PrCa risk, although no variant was statistically significant after accounting for the number of interaction tests. Of note, there was suggestive evidence that variant rs10993994 at 10q11.22 is more strongly associated with risk among Latino men with AMR local ancestry (OR<sub>AMR>1.5</sub> = 1.40, 95% CI: 1.10, 1.77,  $p = 5.97 \times 10^{-3}$ ; OR<sub>AMR0.5–1.5</sub> = 1.36, 95% CI: 1.17, 1.58,  $p = 6.28 \times 10^{-5}$ ) compared to men with little AMR local ancestry in this region (OR<sub>AMR  $\leq$  0.5</sub> = 1.19; 95% CI: 1.04, 1.36,  $p = 1.23 \times 10^{-2}$ ). The same suggestive trend, with an association being stronger or limited to men with AMR ancestry in the region, was observed for another four known PrCa risk variants (rs9443189 [6q14.1], rs10875943 [12q13.12], rs12956892 [18q21.32] and rs1978060 [22q11.21]), while six variants had greater effect sizes among men with lower AMR local ancestry proportions (rs2028900 [2p11.2], rs4976790 [5q35.3], rs5875234 [6p22.1], rs630045 [6q22.1], rs17790938 [20q13.13] and rs909666 [22q13.2]; Supporting Information Table S4).

### Polygenic risk score

In estimating a EUR-weighted PRS, Latino men in the top 10% PRS stratum had a 3.19-fold (95% CI: 2.65, 3.84) elevated risk and those in the top 1% had a 4.02-fold (95% CI: 2.46, 6.55) increased risk compared to men with average risk (PRS in 25th–75th percentiles; Table 1). Among Latinos with a higher proportion of European global ancestry (in the 4th quantile of EUR global ancestry in controls), we observed a more pronounced increase in PrCa risk (OR = 3.68; 95% CI: 2.56, 5.29) for men in the top 10% EUR-weighted PRS risk

**Table 1.** Associations between categorized polygenic risk scores (PRS) and prostate cancer risk in Latino men

Polygenic risk score category	European-weighted PRS <sup>1</sup>				Latino-weighted PRS <sup>1</sup>			
	No. of cases	No. of controls	OR (95% CI) <sup>2</sup>	$p$ -value <sup>3</sup>	No. of cases	No. of controls	OR (95% CI) <sup>2</sup>	$p$ -value <sup>3</sup>
0–1%	5	22	0.25 (0.09,0.67)	$6.11 \times 10^{-3}$	2	22	0.14 (0.03,0.68)	$1.44 \times 10^{-2}$
1–10%	68	189	0.38 (0.28,0.51)	$1.93 \times 10^{-10}$	53	189	0.32 (0.23,0.45)	$1.17 \times 10^{-11}$
10–25%	169	314	0.60 (0.48,0.74)	$2.02 \times 10^{-6}$	150	314	0.57 (0.45,0.71)	$5.30 \times 10^{-7}$
25–75% (baseline)	952	1,048	–	–	835	1,048	–	–
75–90%	445	314	1.58 (1.33,1.88)	$2.34 \times 10^{-7}$	540	314	2.25 (1.90,2.67)	$1.10 \times 10^{-20}$
90–99%	507	189	3.10 (2.55,3.76)	$2.37 \times 10^{-30}$	533	189	3.87 (3.18,4.71)	$9.57 \times 10^{-42}$
99–100%	80	22	4.02 (2.46,6.55)	$2.43 \times 10^{-8}$	113	22	6.87 (4.27,11.06)	$2.22 \times 10^{-15}$

<sup>1</sup>PRS was calculated using 176 known SNPs (MAF  $> 0.001$  and imputation score  $\geq 0.3$  in Set 1 and Set 2); for EUR-weighted PRS, the weights were conditional log ORs derived in men of European ancestry; for Latino-weighted PRS, the weights were conditional log ORs derived in men of Latino ancestry (Set 1 and Set 2).

<sup>2</sup>Odds ratios (ORs) were adjusted for age, study and the first 10 principal components.

<sup>3</sup> $p$  values were Wald  $p$ -value from fixed-effect meta-analysis.

stratum (Table 2). This association was slightly reduced (OR = 2.94) among men in the 4th quartile of Amerindian ancestry (Supporting Information Table S5). The *p* values for interaction between PRS and EUR and AMR global ancestry were 0.26 and 0.04, respectively. The PRS odds ratios were larger using weights among Latino men from our study; the top 10% PRS stratum had a 4.18-fold (95% CI: 3.47, 5.04) elevated risk and those in the top 1% had a 6.87-fold (95% CI: 4.27, 11.06; Table 1). Effect modification of the Latino-weighted PRS by EUR and AMR global ancestry was also observed (Table 2 and Supporting Information Table S5), with *p* values for interaction of 0.08 and 0.01, respectively.

## Discussion

In our study, among Latinos, two known risk regions, at 8q24.21 and 10q11.22, achieved genome-wide significance, and admixture mapping highlighted the 8q24 region as harboring PrCa risk variants related to local African ancestry. The majority of established risk alleles were also replicated in Latinos in terms of directional consistency, and among them, ~30% achieved nominal significance. In the PRS analysis, the established risk alleles were found to be strongly associated with PrCa risk, with a larger PRS effect observed for men with more European ancestry.

Previous GWAS studies of PrCa have identified more than 170 common risk variants, with the majority of discovery populations being of European or Asian ancestry.<sup>28,30</sup> As found in previous studies in men of African ancestry,<sup>16</sup> directional consistency was also observed for the majority (>80%) of risk variants in Latinos, among which ~30% were nominally statistically significant, suggesting that most of the known genetic susceptibility *loci* for PrCa generalize to the Latino population, which may not be surprising given their high degree of European ancestry. Two regions, 8q24 and 10q11.22, achieved genome-wide significance. The risk region at 8q24 harbors multiple independent risk variants and is consistently recognized as the most significant PrCa risk region across ethnic populations.<sup>16,31</sup> However, in the Latino population, the 10q11.22 surpassed 8q24 as the most significant risk region. At 10q11.22, the risk variant rs10993994 has been consistently associated with PrCa risk across populations,<sup>32–34</sup> and is likely to be the putative causal variant within the region.<sup>35</sup> The risk allele rs10993994-T is more common among populations of African ancestry (RAF<sub>AFR</sub> = 0.65, RAF<sub>AMR</sub> = 0.40, RAF<sub>EUR</sub> = 0.39) in Phase III 1KGP. In our Latino men, it was associated with a 1.29-fold (95%CI: 1.19, 1.39) increased risk, which is similar to that reported in the largest European PrCa GWAS (OR = 1.23, 1.21, 1.25),<sup>28</sup> while larger than that reported in the AA PrCa GWAS (OR = 1.12, 95%CI: 1.07, 1.16).<sup>16</sup> This allele is located close to the transcription start site of the microseminoprotein-beta (*MSMB*) gene and was reported to be significantly related with gene expression abundance.<sup>36</sup> The encoded microseminoprotein (MSP) is one of the three major proteins secreted by the

Table 2. Associations between categorized polygenic risk scores (PRSs) and prostate cancer risk in Latino men by European global ancestry strata

European global ancestry strata <sup>1</sup>	Polygenic risk score category	European-weighted PRS <sup>2</sup>				Latino-weighted PRS <sup>2</sup>			
		No. of cases	No. of controls	OR (95% CI) <sup>3</sup>	<i>p</i> -value <sup>4</sup>	No. of cases	No. of controls	OR (95% CI) <sup>3</sup>	<i>p</i> -value <sup>4</sup>
≤25%	0–10%	12	53	0.27 (0.14,0.54)	1.64 × 10 <sup>-4</sup>	9	53	0.21 (0.10,0.45)	5.71 × 10 <sup>-5</sup>
	10–25%	30	79	0.51 (0.32,0.83)	6.41 × 10 <sup>-3</sup>	15	79	0.27 (0.15,0.49)	2.46 × 10 <sup>-5</sup>
	25–75%	183	261	–	–	174	261	–	–
	75–90%	77	79	1.47 (1.00,2.15)	4.70 × 10 <sup>-2</sup>	88	79	1.64 (1.12,2.40)	1.04 × 10 <sup>-2</sup>
	90–100%	104	53	3.08 (2.05,4.63)	5.50 × 10 <sup>-8</sup>	120	53	3.81 (2.53,5.73)	1.46 × 10 <sup>-10</sup>
>75%	0–10%	25	53	0.43 (0.25,0.74)	2.13 × 10 <sup>-3</sup>	17	53	0.32 (0.18,0.59)	2.25 × 10 <sup>-4</sup>
	10–25%	61	79	0.68 (0.46,1.01)	5.50 × 10 <sup>-2</sup>	45	79	0.59 (0.38,0.91)	1.75 × 10 <sup>-2</sup>
	25–75%	299	262	–	–	247	262	–	–
	75–90%	132	78	1.43 (1.01,2.03)	4.60 × 10 <sup>-2</sup>	192	78	2.70 (1.93,3.78)	7.53 × 10 <sup>-9</sup>
	90–100%	216	53	3.68 (2.56,5.29)	1.84 × 10 <sup>-12</sup>	232	53	4.95 (3.43,7.15)	1.36 × 10 <sup>-17</sup>

<sup>1</sup>Strata were created by categorizing European global ancestry score according to its percentiles (≤25%, >75%) in controls.

<sup>2</sup>PRS was calculated using 176 known SNPs (MAF ≥ 0.001 and imputation score ≥ 0.3 in Set 1 and Set 2); for European-weighted PRS, the weights were the conditional log ORs derived from men of European ancestry; for Latino-weighted PRS, weights were the conditional log ORs obtained from our Latino men (Set 1 and Set 2).

<sup>3</sup>Odds ratios (ORs) were adjusted for age, the first 10 principal components, and studies.

<sup>4</sup>*p*-values were Wald *p*-values from fixed-effect meta-analyses.

prostate, and we have shown that reduced serum levels are strongly associated with PrCa risk.<sup>37</sup> In comparison to whites and blacks, the geometric mean plasma MSP level was observed to be lower in PrCa-free Latinos after adjusting for rs10993994 genotype, age, BMI and PSA level.<sup>38</sup> However, in contrast to the association observed with the risk SNP, the magnitude of the association between blood MSP concentration and PrCa risk is smaller in Latinos than whites, yet the difference was not statistically significant.<sup>37</sup> Additional studies will be needed to better understand the strong association between the 10q11.22 risk SNP and PrCa risk in Latinos.

Latinos are a highly heterogeneous population; the ancestry structure varies widely across subgroups. Previous literature reported that compared to other Latino subgroups, Mexicans had the highest proportion of Native American ancestries.<sup>39</sup> Coincidentally, studies have also shown that self-reported Mexican Americans have lower PrCa incidence and mortality rates than whites and other Latino-subgroups,<sup>4,40,41</sup> suggesting that AmerIndian genetic ancestry might be a protective factor for PrCa risk. A previous study showed that the estimated global AMR ancestry was inversely associated with breast cancer risk.<sup>42</sup> Similarly, our results support the hypothesis that global AMR ancestry was inversely associated with PrCa risk, even after excluding the outlier study MDA with Mexican American controls but a more diverse representation of Latino cases. To note, global genetic ancestry estimates not only reflect potential genetic differences in disease susceptibility but may also capture cultural, behavioral and lifestyle factors, including SES as well as access and adherence to medical care and cancer screening. For some chronic diseases, associations between genetic ancestry and disease risk have been shown to be greatly attenuated or extinguished after accounting for such factors.<sup>43</sup> However, this is not the case in some other studies of lung cancer,<sup>44</sup> myocardial infarction and impaired fasting glucose.<sup>45</sup> Thus, further investigation is required to disentangle genetic ancestry representing genetic vs. nongenetic/social/behavioral influences on PrCa risk.

While none of the interactions between known risk alleles and local ancestry were significant after correcting for multiple tests, there was a suggestion that variant rs10993994 was more strongly associated with risk among men with greater local AMR ancestry. In men with a high or moderate proportion of local AMR ancestry, the OR was 1.4 vs. 1.2 in men with lower local AMR ancestry (<25%), which may explain

the observed strong association in 10q11.22 risk region among Latinos. Testing interaction effects by local ancestry in Latinos will require a larger sample size.

Previous PRS analyses in populations of European ancestry have reported a ~threefold difference in risk comparing people in the top 10% risk stratum to the population average,<sup>46–49</sup> with the magnitude of effect being similar in African Americans.<sup>16</sup> Similar to the previously reported effect size in studies among men of European descent, we observed a 3.2-fold increased PrCa risk in Latino men. A multiethnic study, which contained a part of our samples, demonstrated that when comparing the highest to the lowest risk score decile, the effect size was larger among non-Hispanic whites than in Latinos (OR = 6.2 vs. 5.8).<sup>24</sup> Consistent with their results, we observed a stronger effect of the PRS on PrCa risk among Latino men with higher proportion of European global ancestry: among them, the effect size comparing the top 10% to the population average risk stratum increased to 3.7-fold. These observations may be due to ethnic differences in the frequencies of risk alleles and to the LD patterns surrounding causal SNPs and suggest that global ancestry background might modify the effect of PrCa risk variants, further supporting the need to construct ethnicity-specific PRS. We also found the PRS associations to be larger when using weights from Latino men in our study; however, since weights came from the same population, the effect sizes are likely to be overestimated. An independent Latino replication sample is needed to validate this observation.

Although our analysis represented the largest study of PrCa genetic susceptibility among Latinos, it remained underpowered for less common risk alleles with small effect size; for the genome-wide analysis ( $\alpha = 5 \times 10^{-8}$ ), our study only had 80% *a priori* power to detect common risk alleles (MAF = 10%) with moderate effect size (OR  $\geq$  1.40); for known risk alleles with MAF of 5%, the power to detect ORs of 1.20 at a nominal significant level ( $p < 0.05$ ) was only 70%. However, for common variants with MAF > 10%, we had more than adequate power (90%) to detect a moderate effect of 1.20.

In summary, we found that the known PrCa risk variants can stratify PCa risk in Latino men. Larger studies in Latino populations, both in the US and in other countries, which will expand AMR ancestral diversity, will be required to characterize genetic risk variants and improve risk stratification for this population.

## References

1. Jemal A, Ward EM, Johnson CJ, et al. Annual report to the nation on the status of cancer, 1975–2014, featuring survival. *J Natl Cancer Inst* 2017;109:djx030.
2. Latini DM, Elkin EP, Cooperberg MR, et al. Differences in clinical characteristics and disease-free survival for Latino, African American, and non-Latino white men with localized prostate cancer. *Cancer* 2006;106:789–95.
3. Park S-Y, Haiman CA, Cheng I, et al. Racial/ethnic differences in lifestyle-related factors and prostate cancer risk: the multiethnic cohort study. *Cancer Causes Control* 2015;26:1507–15.
4. Pinheiro PS, Sherman RL, Trapido EJ, et al. Cancer incidence in first generation U.S. Hispanics: Cubans, Mexicans, Puerto Ricans, and new Latinos. *Cancer Epidemiol Biomarkers Prev* 2009;18:2162–9.
5. Pinheiro PS, Callahan KE, Siegel RL, et al. Cancer mortality in Hispanic ethnic groups. *Cancer Epidemiol Biomarkers Prev* 2017;26:376–82.
6. Bryc K, Durand EY, Macpherson JM, et al. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet* 2015;96:37–53.
7. Beuten J, Halder I, Fowler SP, et al. Wide disparity in genetic admixture among Mexican



- Americans from San Antonio, TX. *Ann Hum Genet* 2011;75:529–38.
8. Gao X, Nannini DR, Corrao K, et al. Genome-wide association study identifies WNT7B as a novel locus for central corneal thickness in Latinos. *Hum Mol Genet* 2016;25:5035–45.
  9. Galanter JM, Gignoux CR, Torgerson DG, et al. GWAS and admixture mapping identify different asthma-associated loci in Latinos: the GALA II study. *J Allergy Clin Immunol* 2014;134:295–305.
  10. Consortium T2D. Sequence variants in *SLC16A11* are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014;506:97–101.
  11. Fejerman L, Ahmadiyeh N, Hu D, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. *Nat Commun* 2014;5:5260.
  12. González Burchard E, Borrell LN, Choudhry S, et al. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. *Am J Public Health* 2005;95:2161–8.
  13. Choudhry S, Ung N, Avila PC, et al. Pharmacogenetic differences in response to albuterol between Puerto Ricans and Mexicans with asthma. *Am J Respir Crit Care Med* 2005;171:563–70.
  14. Cheng I, Chen GK, Nakagawa H, et al. Evaluating genetic risk for prostate cancer among Japanese and Latinos. *Cancer Epidemiol Biomarkers Prev* 2012;21:2048–58.
  15. Amos CI, Dennis J, Wang Z, et al. The OncoArray consortium: a network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev* 2017;26:126–35.
  16. Conti DV, Wang K, Sheng X, et al. Two novel susceptibility loci for prostate cancer in men of African ancestry. *J Natl Cancer Inst* 2017;109:djx084.
  17. Schwartz GG, John EM, Rowland G, et al. Prostate cancer in African-American men and polymorphism in the calcium-sensing receptor. *Cancer Biol Ther* 2010;9:994–9.
  18. Pike MC, Peters RK, Cozen W, et al. Estrogen-progestin replacement therapy and endometrial cancer. *J Natl Cancer Inst* 1997;89:1110–6.
  19. Strom SS, Gu Y, Zhang H, et al. Androgen receptor polymorphisms and risk of biochemical failure among prostatectomy patients. *Prostate* 2004;60:343–51.
  20. Strom SS, Yamamura Y, Flores-Sandoval FN, et al. Prostate cancer in Mexican-Americans: identification of risk factors. *Prostate* 2008;68:563–70.
  21. Wilkinson AV, Spitz MR, Strom SS, et al. Effects of nativity, age at migration, and acculturation on smoking among adult Houston residents of Mexican descent. *Am J Public Health* 2005;95:1043–9.
  22. Beuten J, Gelfond JAL, Martinez-Fierro ML, et al. Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men. *Carcinogenesis* 2009;30:1372–9.
  23. Beuten J, Gelfond JAL, Franke JL, et al. Single and multigenic analysis of the association between variants in 12 steroid hormone metabolism genes and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2009;18:1869–80.
  24. Hoffmann TJ, Eeden SKVD, Sakoda LC, et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov* 2015;5:878–91.
  25. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–9.
  26. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–7.
  27. Maples BK, Gravel S, Kenny EE, et al. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 2013;93:278–88.
  28. Schumacher FR, Olama AAA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018;50:928–36.
  29. Matejčić M, Saunders EJ, Dadaev T, et al. Germline variation at 8q24 and prostate cancer risk in men of European ancestry. *Nat Commun* 2018;9:4616.
  30. Park SL, Cheng I, Haiman CA. Genome-wide association studies of cancer in diverse populations. *Cancer Epidemiol Biomarkers Prev* 2018;27:405–17.
  31. Takata R, Akamatsu S, Kubo M, et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 2010;42:751–4.
  32. Guy M, Kote-Jarai Z, Giles GG, et al. Identification of new genetic risk factors for prostate cancer. *Asian J Androl* 2009;11:49–55.
  33. Mhatre DR, Mahale SD, Khatkhatay MI, et al. The rs10993994 in the proximal MSMB promoter region is a functional polymorphism in Asian Indian subjects. *Springerplus* 2015;4:380.
  34. Chang B-L, Cramer SD, Wiklund F, et al. Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. *Hum Mol Genet* 2009;18:1368–75.
  35. Dadaev T, Saunders EJ, Newcombe PJ, et al. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat Commun* 2018;9:2256.
  36. Pomerantz MM, Shrestha Y, Flavin RJ, et al. Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS Genet* 2010;6:e1001204.
  37. Haiman CA, Stram DO, Vickers AJ, et al. Levels of Beta-Microseminoprotein in blood and risk of prostate cancer in multiple populations. *J Natl Cancer Inst* 2013;105:237–43.
  38. Waters KM, Stram DO, Marchand L-L, et al. A common prostate cancer risk variant 5' of MSMB (microseminoprotein-beta) is a strong predictor of circulating MSP (microseminoprotein) in multiple populations. *Cancer Epidemiol Biomarkers Prev* 2010;19:2639–46.
  39. Bryc K, Velez C, Karafet T, et al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci USA* 2010;107:8954–61.
  40. Sheinfeld Gorin S, Heck JE. Cancer screening among Latino subgroups in the United States. *Prev Med* 2005;40:515–26.
  41. Martinez-Tyson D, Pathak EB, Soler-Vila H, et al. Looking under the Hispanic umbrella: cancer mortality among Cubans, Mexicans, Puerto Ricans and other Hispanics in Florida. *J Immigr Minor Health* 2009;11:249–57.
  42. Fejerman L, John EM, Huntsman S, et al. Genetic ancestry and risk of breast cancer among U.S. Latinas. *Cancer Res* 2008;68:9723–8.
  43. Pereira L, Zamudio R, Soares-Souza G, et al. Socioeconomic and nutritional factors account for the Association of Gastric Cancer with Amerindian ancestry in a Latin American admixed population. *PLoS One* 2012;7:e41200.
  44. Aldrich MC, Selvin S, Wrensch MR, et al. Socioeconomic status and lung cancer: unraveling the contribution of genetic admixture. *Am J Public Health* 2013;103:e73–80.
  45. Ruiz-Narváez EA, Bare L, Arellano A, et al. West African and Amerindian ancestry and risk of myocardial infarction and metabolic syndrome in the Central Valley population of Costa Rica. *Hum Genet* 2010;127:629–38.
  46. Al Olama AA, Kote-Jarai Z, Berndt SI, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 2014;46:1103–9.
  47. Fritsche LG, Gruber SB, Wu Z, et al. Association of polygenic risk scores for multiple cancers in a Phenome-wide study: results from the Michigan genomics initiative. *Am J Hum Genet* 2018;102:1048–61.
  48. Al Olama AA, Benlloch S, Antoniou AC, et al. Risk analysis of prostate cancer in PRACTICAL, a multinational consortium, using 25 known prostate cancer susceptibility loci. *Cancer Epidemiol Biomarkers Prev* 2015;24:1121–9.
  49. Machiela MJ, Chen C-Y, Chen C, et al. Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. *Genet Epidemiol* 2011;35:506–14.