

Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants

A full list of authors and affiliations appears at the end of the article.

Introduction

The transferability and clinical value of genetic risk scores (GRS) across populations remains limited due to an imbalance in genetic studies across ancestrally diverse populations. We conducted a multi-ancestry genome-wide association study (GWAS) of 156,319 prostate cancer cases and 788,443 controls of European, African, Asian, and Hispanic men, reflecting a 57% increase in the number of non-European cases over previous prostate cancer GWAS. We identified 187 novel risk variants for prostate cancer, increasing the total number of risk variants to 451. An externally replicated multi-ancestry GRS was associated with risk that ranged from 1.8 (per standard deviation (SD)) in African ancestry men to 2.2 in European ancestry men. The GRS was associated with a greater risk of aggressive versus non-aggressive disease in men of African ancestry ($P=0.03$). Our study

*Corresponding Author: Christopher A. Haiman, Harlyne J. Norris Cancer Research Tower, USC Norris Comprehensive Cancer Center, 1450 Biggy Street, Rm 1504, Los Angeles, CA 90033 or haiman@usc.edu.

†Drs. Conti and Haiman contributed equally to this article.

Author Contributions

CAH, DVC, RAE and ZK-J contributed to study conception. AWang, CAH, DVC, EJS and NM wrote the manuscript. EJS, YX, XS, PW, MB, AAR, RKM and TD provided data management and bioinformatics support. AWang, JShen, AAR, DVC and CAH contributed to data analysis and interpretation. All authors contributed data to the study, revised, critically reviewed and approved the final version of the manuscript: AWang, JShen, AAR, EJS, FChen, RJanivara, BFDarst, XS, YX, AJC, SB, TD, MNB, AP, AS, TJH, ATakahashi, KMatsuda, YM, MF, TL, JF, KMuir, SI, XL, YU, MKubo, YK, ALophatananon, PW, CA, ALori, PPC, JSchleutker, TLT, CSipeky, AAuvinen, GGG, MCSouthey, RJM, CC, DW, JLubinski, CTR, KC, BHM, DEN, JLD, FCH, RMMartin, BGN, SFN, MW, SEB, MAR, HVS, JB, SC, LH, JAC, WTilly, GPR, HG, MA, RS, ME, TN, NP, AMD, MGhousaini, RCT, TJK, ER, JYP, TAS, HYL, DA, SWeinstein, MBC, LAM, EG, SLindstrom, PK, DJH, KLP, CTurman, CMT, PJG, IMT, RJH, NEF, AF, MEP, JLS, EAO, SK, LEBF, MS, AWolk, NH, GLA, RNH, MJM, KDS, MB, WJB, WZ, EDY, JEM, YJL, HWZ, NF, XM, YW, SCZ, ZS, SNT, SKM, DJS, CMW, GB, CM, TS, ML, ASK, BFDrake, OC, GCT, FM, TT, YAK, EMJ, EMG, LMK, KTK, SAI, MCStern, AV, AGC, LFachal, BSR, SLK, HO, MRT, PPaulo, AB, SWatya, ALubwama, JTB, ENB, JLM, JAT, MKogevinas, TDS, GCV, LCA, CCT, CDH, PPilie, YY, RJB, JG, SSS, LM, PB, LB, RK, CSlavov, VM, RJL, HB, XC, BH, BS, EAK, AWH, RAK, ABM, CJL, JK, SLN, LS, YCD, WBI, BN, AJH, JCarpten, HP, AM, KDR, GDM, PO, JX, AR, JLim, SHT, LFN, DWL, JHF, CMND, BAR, MGamulin, DL, TK, NU, AAbraham, SSinghal, MP, FClæssens, SJ, TVDB, MGD, JEC, MEM, SLarkin, PAT, CA-H, WSB, MCA, DCC, SSRivastava, JCullen, GP, GCasey, YW, YT, JLachance, WTang, RBB, AAA, ETay, AT, SN, KY, KG, APC, JMK, JNH, PEC, MJ, SMGueye, LN, OO, OS, OA, AOA, OIAS, HOA, MAJ, OPO, MN, BA, SM, ADA, HD, SMGundell, MJR, GJ, RHVS, JJH, MS, LK, RV, RMC, MT, MHP, RJL, MZ, SZ, ZL, SKVDE, DFE, SA, TLE, RM, TRR, LFritsche, SJC, SIB, FW, HN, JSW, JMG, ACJ, NM, CTerao, RAE, ZKJ, RKM, DVC, and CAH. CAH and RKM had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Code Availability

Imputation was performed using IMPUTE2, MACH 1.0, Beagle 4.1, Beagle 5.1, EAGLE v2.4, Minimac3, and Minimac4. Association testing was performed using PLINK 1.07 and 2.0, SNPtest v2.5.2, SAIGE v0.20, and R v3.6.3. Meta-analyses were conducted using METAL v2011-03-25 and fine-mapping with mJAM (<https://github.com/USCbiostats/hJAM/>). Genome-wide PRS was derived from PRS-CSx v1.0.0 (<https://github.com/getian107/PRScsx>). Variant annotation was performed with wANNOVAR (<https://wannovar.wglab.org/>, accessed 20 May, 2022) and R package rtracklayer v1.42.2. TWAS was performed with FUSION (https://github.com/gusevlab/fusion_twas, accessed 20 May, 2022); TWAS weights: GTExv8 and TCGA: <http://gusevlab.org/projects/fusion/>, MAYO RefZ: <https://www.mancusolab.com/prostate-twas/>, INTERVAL: <https://www.mancusolab.com/pwas/>) and GCTA v1.94.0beta. Data visualization was performed using ggplot2 v3.4.2 and gwasforest v1.0.0 packages in R software (v3.6.3).

Competing interests

The authors declare no competing interests.

presents novel prostate cancer susceptibility loci and a GRS with effective risk stratification across ancestry groups.

In men, prostate cancer is the most frequently diagnosed non-skin cancer globally¹. Variation in prostate cancer incidence is observed across populations globally, with the highest rates observed in men of African ancestry¹. Prostate cancer risk is heavily influenced by genetic factors, with 278 genetic risk variants identified through GWAS²⁻¹³. While the majority of samples in prostate cancer GWAS have been of European ancestry, multi-ancestry analysis has been demonstrated to improve discovery of novel risk variants¹⁴ and enhance genetic risk prediction for prostate cancer across populations².

We conducted a multi-ancestry GWAS meta-analysis with 122,188/604,640 (cases/controls) of European ancestry, 19,391/61,608 of African ancestry, 10,809/95,790 of East Asian ancestry and 3,931/26,405 of Hispanic ethnicity. Studies, genotyping, quality control and association testing methods are described in Supplementary Table 1 and 2 (Methods). Case sample size was increased by 43% in European, 87% in African, 26% in Asian and 45% in Hispanic groups (with a corresponding effective sample size 128% in each population accounting for controls), compared to previous multi-ancestry GWAS analyses². We performed a fixed-effect meta-analysis within each ancestry group and meta-analyzed the ancestry-specific GWAS results. The genomic inflation statistic (λ) was 1.158 in the multi-ancestry GWAS and ranged from 1.053 in Asian to 1.169 in European ancestry studies (Supplementary Table 3); the corresponding meta-analysis λ_{1000} (scaled to a sample size of 1,000 cases and 1,000 controls) was 1.001.

Overall, 42,428,922 variants with a minor allele frequency (MAF)>0.1% were examined for association with prostate cancer risk, with 55,241 variants reaching genome-wide significance ($P<5.0\times 10^{-8}$). To identify independent risk variants, we implemented a forward-selection conditional analysis using multi-population Joint Analysis of Marginal summary statistics (mJAM; Methods)^{2,15}. We identified 451 independent risk variants for prostate cancer that were genome-wide significant in multi-ancestry or ancestry-specific analyses (Supplementary Table 4), including 187 that were previously unreported (Fig. 1, Supplementary Tables 4 and 5). Of these, 61 were within 800 Kb of known variants but remained genome-wide significant after conditioning on nearby known variants. Of the 451 variants, 150 were known risk variants that were replaced by a more significant lead variant, while 114 remained the lead risk variant in the region. Eighteen variants previously reported as prostate cancer risk variants were dropped because they did not reach genome-wide significance (Supplementary Table 4).

The underlying rationale for conducting a cross-ancestry meta-analysis is based on the hypothesis that true causal variants are predominantly shared across populations. Of the 451 risk variants, 429 (95%) in European, 411 (91%) in African, 377 (84%) in Asian and 424 (94%) in Hispanic populations had MAF>1% (Extended Data Fig. 1), and 339 (75%), 47 (10%), 42 (9%) and 9 (2%) were genome-wide significant, respectively (Fig. 2a). Of these, nineteen (European), five (African) and three (Asian) were population-specific risk variants with MAF >1% in all other populations (Extended Data Fig. 1). For variants with a MAF>1% in all populations (n=370), 369, 247, 208 and 125 were nominally

significant in European, African, Asian and Hispanic populations, respectively (Fig. 2b). The effect sizes for variants with a MAF>1% were correlated between populations, with an $R=0.73$ for European versus African ancestry (398 variants), $R=0.58$ for European versus Asian ancestry (371 variants) and $R=0.72$ for European ancestry versus Hispanic men (414 variants; Fig. 2c, Supplementary Fig. 1). Heterogeneity in effect size was statistically significant ($P_{heterogeneity}<0.05$) for 78 variants (21%), with the largest average effect size in Asian men (odds ratio (OR)_{avg}=1.11) followed by European ancestry (OR_{avg} =1.09), African ancestry (OR_{avg} =1.08) and Hispanic men (OR_{avg} =1.08; Supplementary Table 6).

Of the 451 variants, 28 (6.2%) directly alter protein structure (Supplementary Table 7). We detected a novel association with a population-specific frameshift deletion in the *C9orf152* gene (European) and previously reported frameshift deletions in *ANO7* (African¹⁶) and *CHEK2* (European²) and a frameshift insertion in *FAM111A* (European⁴). The lead variants include 24 missense substitutions representing previously reported variants within *ANO7* (three lead variants⁴), *CDKN1B*, *CHEK2*, *COL23A1*, *HOXB13*, *INCENP*, *KLK3*, *POGLUT3*, *RASSF6*, *RFX7* and *SUN2*, replacement lead variants in *FAM118A*, *INHBB* and *SPDL1*, novel associations in *MMAB*, *PIMI*, *RPA1*, *SERPINA1*, *SIM2*, *SYTL1* and *ZBTB42*, and a second missense risk variant in *RASSF6* (Supplementary Table 7). Among the new genes implicated in prostate cancer risk, expression of *SIM2*, a transcription factor, has been shown to discriminate prostate cancer and non-cancerous tumor tissue¹⁷ and to be associated with poorer survival¹⁸, while *PIMI* is a serine/threonine kinase overexpressed in prostate cancer¹⁹, shown to modulate androgen receptor transcriptional activity through phosphorylation²⁰ and be a co-activator of c-MYC²¹.

Many lead variants were also implicated in regulation of gene expression in prostate tissues and cell-lines (Methods). Seventy-four variants (16.4%), including 19 novel associations, were located within regions of open chromatin, chromatin modifications consistent with regulatory elements, situated within transcription factor binding sites overlapping an association for differential gene expression or splicing (Supplementary Table 7), providing strong support for biological functionality. Candidate functional variants include rs1858800, correlated with expression of *ZFXH3*, a gene frequently somatically mutated in prostate cancer²²; rs10499188, correlated with expression of *SLC2A12*, a gene encoding a glucose transporter expressed in prostate cancer cell-lines but not benign prostatic hyperplasia²³ and regulated by androgen receptor signaling²⁴, and rs79186742, correlated with expression of *BARX2*, a homeobox transcription factor associated with poor prognosis for a range of solid tumors²⁵.

Overall, 219 of the 451 lead variants (48.6%) overlap with significant associations for differential expression in prostate tissues (Methods, Supplementary Table 7) of 439 distinct genes (eQTLs), while 69 (15.3%) correlate with significant associations for alternative splicing of 95 unique genes (sQTLs). Of the 439 differentially expressed genes, 204 (46.5%) had not been implicated as candidate mediators of prostate cancer risk by the previous panel of 269 prostate cancer risk variants² and were established through the identification of additional novel risk variants and replacement of lead variants. To assess the extent to which prostate cancer risk variants exhibit prostate-specific regulatory function compared with the genome-wide background, we performed a permutation test while controlling for

MAF and linkage disequilibrium (LD) patterns (Methods). Overall, we found evidence for enrichment of prostate cancer risk variants in regions of prostate-specific regulatory activity across eQTLs, sQTLs and candidate *cis*-regulatory elements (2.9-fold enrichment, $P < 0.0017$; Supplementary Table 8).

To further explore the molecular mechanisms underlying prostate cancer risk, we performed transcriptome- (TWAS) and proteome-wide association studies (PWAS)^{26–28} using predicted gene expression and protein levels from multiple prostate tissue^{29–31} and plasma³² studies (Methods). Across 19,352 tests performed, we identified 746 associations across 528 genes and 230 genomic regions (Supplementary Tables 9 and 10). Of the 746 associations, the greatest contribution was from predicted expression in histologically normal prostate tissue (351/746)³⁰. However, this is likely due to the larger reference panel sample size and, thus, number of association tests performed (Supplementary Table 9; ANOVA $P > 0.05$). Of the 451 genomic risk regions identified through GWAS, 237 colocalized within 250Kb of transcriptome- or proteome-wide significant associations, which is consistent with previous large-scale TWAS investigations of prostate cancer risk^{33,34}. Of the 230 TWAS/PWAS genomic risk regions identified, 45 did not colocalize within 250Kb of the 451 genome-wide significant variants, suggesting that increasing GWAS sample sizes will continue to identify novel risk regions (Supplementary Table 11).

The predictive ability of the GRS for prostate cancer has improved with the identification of additional risk variants^{2–6,8}. We compared the performance of GRSs based on past marker sets ($n=100^8$, $181^{5,6,35}$, 269^2) to the current set of 451 risk variants, with GRSs constructed by summing the risk allele dosage, weighted by the multi-ancestry per-allele log-ORs estimated from the current meta-analysis (Methods). With the discovery of more risk variants, there is greater stability in the assignment of unaffected men to GRS categories; 58% of men in the lowest or highest quintile remained in the same quintile between GRS₁₀₀ and GRS₁₈₁, whereas 69% to 70% remained between GRS₂₆₉ and GRS₄₅₁ (Supplementary Fig. 2a-6a). Likewise, the percentage of cases has increased for each population within higher GRS categories (e.g., from 40.5% in the highest quintile of GRS₁₀₀ to 51.2% in GRS₄₅₁) and decreased within lower GRS categories (e.g., from 7.5% in the lowest quintile of GRS₁₀₀ to 4.4% in GRS₄₅₁; Fig. 3, Supplementary Fig. 2b-6b). Risk classification with the GRS in addition to age was evaluated using the net reclassification index (NRI)³⁶ and showed substantial improvement from GRS₁₀₀ (range across populations: 30.2% in African to 49.5% in European) to GRS₄₅₁ (range across populations: 58.5% in African to 69.9% in European; Supplementary Table 12). Compared to a model with GRS₂₆₉, the population specific improvement for a model with GRS₄₅₁ resulted in a NRI ranging from 3.3% in Asian ancestry to 21.7% in Hispanics. The improvement in risk prediction of GRS₄₅₁ over previous GRS panels was confirmed in replication studies among men of European and African ancestry that were not included in the GWAS (Fig. 4a-b, Supplementary Table 13 and 14). Based on the high degree of variation in the association of GRS₄₅₁ with prostate cancer risk across sub-studies in the discovery and replication phases (Extended Data Fig. 2), a single summary OR per SD was estimated from the overall meta-analyzed sample: 2.32 [95% CI: 2.30–2.35], 2.04 [95% CI: 2.00–2.08], 2.15 [95% CI: 1.99–2.32] and 2.12 [95% CI: 2.03–2.23] for European, African, Asian and Hispanic men, respectively ($P_{heterogeneity}$ by population: 4.51×10^{-50} , 7.52×10^{-4} , 0.29 and 0.31, respectively). The ORs in the replication

studies were 2.19 [95%CI: 2.12–2.25] in European and 1.79 [95%CI:1.69–1.90] in African ancestry men (Fig. 4b). In replication studies, comparing GRS₄₅₁ to a genome-wide polygenic risk score (PRS) derived by PRS-CSx (Methods), the effect estimates of the genome-wide PRS were smaller than those of GRS₄₅₁ in both men of European (OR per SD = 2.00, 95%CI: 1.92–2.10) and African ancestry (OR per SD = 1.54, 95%CI: 1.44–1.64; Supplementary Table 15).

As observed for GRS₂₆₉, age modifies the association of GRS₄₅₁ and prostate cancer risk (Fig. 4c, Supplementary Table 16, Methods)³⁷. In men of European ancestry, GRS₄₅₁ was associated with an OR per SD of 2.86 [95%CI: 2.76–2.97] for men ≤ 55 and 2.27 [95%CI: 2.25–2.30] for men > 55 years ($P_{heterogeneity} = 8.7 \times 10^{-33}$). Effect modification of GRS₄₅₁ by age was similarly observed in men of African ancestry: OR per SD = 2.41 [95%CI: 2.29–2.54] for men ≤ 55 years and 1.98 [95%CI: 1.94–2.03] for men > 55 years ($P_{heterogeneity} = 8.0 \times 10^{-12}$) and was reproducible in the replication studies (Supplementary Table 16).

In men of European and Asian ancestry and in Hispanic men, the GRS₄₅₁ was equally associated with risk of aggressive prostate cancer (stage T3/T4, regional lymph node involvement, metastatic disease, Gleason score ≥ 8 , prostate-specific antigen (PSA) level ≥ 20 ng/mL or prostate cancer as the underlying cause of death) and non-aggressive prostate cancer (no aggressive features; Fig. 4d, Supplementary Table 17, Methods). For men of African ancestry with prostate cancer, GRS₄₅₁ was associated with a greater risk of aggressive versus non-aggressive disease (OR per SD = 1.08, 95%CI: 1.04–1.12, $P=1.1 \times 10^{-4}$; Fig. 4d, Supplementary Fig. 7). A weak nominally significant association of GRS₄₅₁ with aggressive disease in African ancestry men was also observed in the African prostate cancer MADCaP replication sample (OR per SD= 1.12, 95%CI: 1.01–1.23, $P=0.03$).

Fifty-one of the 451 prostate cancer risk variants have been directly or indirectly (LD $R^2 > 0.8$) associated in GWAS of PSA at $P < 5 \times 10^{-8}$ (Supplementary Table 7, Methods). To assess whether the prostate cancer risk signals for PSA-associated variants reflect an increased likelihood of prostate cancer detection due to screening, particularly for low-stage disease, we examined their aggregate association with disease aggressiveness (Supplementary Table 18). When removing the prostate cancer-PSA variants from the GRS analysis we found the GRS (with 400 markers) to be more strongly associated with aggressive disease (versus GRS₄₅₁) in European ancestry men (OR per SD = 1.04, 95%CI: 1.03–1.06, $P=3.2 \times 10^{-7}$), African ancestry men (OR per SD = 1.10, 95%CI: 1.06–1.14, $P=7.0 \times 10^{-7}$) and Hispanic men (OR per SD = 1.05, 95%CI: 0.97–1.14, $P=0.12$), which suggests that some prostate cancer risk variants may be over-represented in men with less aggressive disease as the result of their association with PSA levels.

A man's cumulative risk of developing prostate cancer, including aggressive disease, is profoundly influenced by the GRS. For men of European ancestry, 20% of men have a 2-fold or greater risk compared to men at the 50% of GRS₄₅₁, and these men achieve an absolute risk comparable to the median risk in the population 16 years earlier. Specifically, these men reach a level of absolute risk of at least 7.8% (the risk at age 85 for men with a 50% GRS₄₅₁) by age 69 or earlier (Fig. 5). For African ancestry men, 16% of men

achieve a 2-fold or greater risk by age 66, with an absolute risk comparable to the risk reached by the average man by age 85 (11.6%), a full 19 years earlier. A GRS-informed approach to screening may improve early detection, as over 50% of cases, including those with aggressive and lethal disease, develop among men in the top GRS quintile, while fewer than 5% of cases develop among men in the bottom 20% (Fig. 3).

Increasing the size of genetic studies across ancestrally diverse populations is paramount for broad and equitable discovery of risk loci and clinical translation. The current multi-ancestry study reflects a 57% increase in the number of non-European cases over previous prostate cancer GWAS and resulted in the identification of 187 novel risk variants, which represents ~40% of all prostate cancer risk variants identified to date. We detected a 3% (Asian), 14% (European), 15% (Hispanic) and 23% (African) increase in the OR (per SD) for GRS_{451} versus GRS_{269} (Fig. 4a), which supports previous work demonstrating the ability of multi-ancestry studies to identify prostate cancer risk variants that improve risk prediction across populations². As shown previously in comparisons of GRS_{269} with genome-wide approaches³⁸, the greater predictive performance observed for GRS_{451} over a genome-wide PRS emphasizes our approach to select a limited set of multi-ancestry risk variants that capture risk across populations. The random selection of markers used for genome-wide PRS may not adequately capture risk across all risk regions resulting in poorer performance, particularly in some populations.

Of critical importance for clinical utility of GRS in prostate cancer is the ability to differentiate risk of aggressive/lethal versus non-aggressive disease. We demonstrated that an understanding of the relationship between germline variants that influence both PSA levels and prostate cancer risk variants is needed to accurately estimate the GRS association with prostate cancer aggressiveness and prostate cancer outcomes. Evidence that GRS can differentiate risk of aggressive versus non-aggressive disease, albeit modestly, for men of African ancestry, an association that strengthened when accounting for PSA variants, suggests potential clinical utility of GRS in this high-risk population¹⁶. While GRS for prostate cancer is a highly effective tool for risk stratification and personalized risk assessment, how and when this information should be included in the decision-making process for prostate cancer screening and early detection needs to be determined.

Online Methods

Study subjects in the multi-ancestry GWAS.

The institutional review board at the University of Southern California approved the study protocol. The meta-analysis included 107,247 prostate cancer cases and 127,006 controls that were part of a previous multi-ancestry meta-analysis (Supplementary Table 1)². The present study included an additional 49,072 cases and 661,437 controls from the UK Biobank, the FinnGen study, the Electronic Medical Records and Genomics (eMERGE) Network, the BioVU Biobank, the BioMe Biobank, the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO), the MD Anderson prostate cancer study (MD Anderson), the California and Uganda Prostate Cancer Study (CA UG), the VA Million Veteran Program (MVP), and the Maryland Prostate Cancer Case-Control Study (NCI-MD) (Supplementary Table 1). Each study includes adult males over the age of 21 years. All

participants provided written informed consents, and study protocols were approved by the Institutional Review Board at each study site. In total, there were 122,188 cases and 604,640 controls of European ancestry, 19,391 cases and 61,608 controls of African ancestry, 10,809 cases and 95,790 controls of Asian ancestry, and 3,931 cases and 26,405 controls of Hispanic ancestry. The effective sample size for each population was calculated using the formula $N_{\text{eff}} = 4/(1/N_{\text{cases}} + 1/N_{\text{controls}})$.

Genotyping and imputation in the multi-ancestry GWAS.

The details of study design, inclusion and exclusion criteria, genotyping, imputation and quality control procedures are provided in Supplementary Tables 1 and 2. Imputation in each study was performed using Minimac3/Minimac4³⁹, Impute2⁴⁰, Eagle2⁴¹, or Beagle 4.1⁴² under the 1000 Genome phase 3⁴³, the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium freeze 5⁴⁴, Haplotype Reference Consortium (HRC), UK10K⁴⁵, or SISu v3 imputation⁴² panels. For most studies, single nucleotide polymorphisms (SNPs) and small insertion/deletions (indels) with MAF $\geq 0.1\%$ and imputation quality scores ≥ 0.3 were included in the association analysis. A higher cutoff of imputation quality score was applied in FinnGen (>0.6) and BioMe (≥ 0.8).

Statistical analysis for GWAS.

Genetic similarity was estimated with uncorrelated SNPs using principal component analysis in each study based. In total, 42,428,922 variants (SNPs and indels) were examined for association using logistic regression adjusting for age, sub-study (if applicable, see Supplementary Table 1) and up to 10 principal components. Per-allele ORs and standard errors from individual studies were combined by a fixed-effects inverse-variance weighted meta-analysis using METAL in ancestry-specific analyses as well as across all four ancestry groups to obtain multi-ancestry estimates of effects. Heterogeneity of effect sizes across ancestries were examined by the statistic I^2 with corresponding tests of significance (Supplementary Table 6). The genomic inflation factors (λ) were calculated in each study/consortium and within each population (Supplementary Table 3). Each inflation factor was then rescaled to λ_{1000} , which represents the inflation factor for an equivalent study of 1,000 cases and 1,000 controls⁴⁶.

Risk variant identification.

Genome-wide significant associations were defined as variants with $P < 5 \times 10^{-8}$ in the multi-ancestry meta-analysis. To identify independent index risk variants in the newly identified and previously known risk regions, we implemented a forward-selection conditional analysis approach using a multi-population Joint Analysis of Marginal summary statistic (mJAM). Within each region, the forward selection process started with a model containing the variants with the most significant multi-ancestry marginal P value, and additional variants were added if they were independent of the selected variants ($LD R^2 < 0.1$ in all four populations). Variants with a conditional multi-ancestry $P < 5 \times 10^{-8}$ were retained in the model. Imputation quality scores of all individual studies were checked for all selected risk variants (Supplementary Table 5).

Genome-wide significant variants were considered “novel” if they were not in LD with any previously known risk variants in any of the four populations and remained genome-wide significant after conditioning on nearby known risk variants. Previously known variants were 1) dropped if their marginal P values were below the genome-wide significance threshold, 2) replaced by a correlated new lead variant with a more significant conditional P value, or 3) not replaced.

GRS construction.

We constructed a GRS from the summed risk allelic dosages weighted by the per-allele log-odds ratios in the marginal model for independent variants and in the conditional model for the variants in the same region. GRS was constructed for the 451 risk variants, and also for risk variant sets reported in previous prostate cancer GWAS meta-analyses: (1) $N=269$ variants reported in a multi-ancestry study (107,247 cases / 127,006 controls)², (2) $N=181$ variants reported in European (25,723 cases / 26,274 controls)³⁵, African (10,202 cases / 10,810 controls)⁴⁷ and Asian (3,000 cases / 4,394 controls)⁶ ancestry-specific studies, respectively, and (3) $N=100$ variants reported in a multi-ancestry study (43,303 cases / 43,737 controls)⁸.

Discriminative improvement of GRS.

To visualize the improvement of predictive ability of prostate cancer GRS over time with the increasing number of risk variants included, we categorized the distributions of previous GRS (GRS_{100} , GRS_{181} , GRS_{269}) and the current GRS (GRS_{451}) into quintiles ([0–20%], (20–40%], (40–60%], (60–80%], and (80–100%]) based on the distribution of the score in controls for each study or consortium. We used Sankey diagrams to visualize the change in risk categorization from the previous GRS to the subsequent GRS among controls and cases, respectively.

To quantify the discriminative ability improvement by inclusion of additional risk variants, we calculated continuous-based NRI in our GWAS discovery sample³⁶. For each study, we calculated NRI comparing a risk model with age only (adjusted for sub-studies and top 10 principal components) to risk models with additional inclusion of GRS_{100} , GRS_{181} , GRS_{269} , and GRS_{451} , respectively. Additionally, we calculated NRI comparing the GRS_{451} model to the GRS_{269} model to show the discriminative ability improvement of the current GRS relative to last GRS. The 95% CIs for NRI were estimated using 1,000 bootstrap replications.

GRS association analysis.

The risk of prostate cancer was estimated for the per SD GRS change and for each percentile category of the GRS: [0–10%], (10–20%], (20–30%], (30–40%], (40–60%], (60–70%], (70–80%], (80–90%], and (90–100%]. Additional analysis was performed to obtain the risk of prostate cancer for the top 1% ((99–100%]). We reported the GRS associations using the median quintile (40–60%] category (Supplementary Table 13) as well as the bottom decile [0%–10%] category as the reference groups (Supplementary Table 14), respectively. The mean and SD, and the GRS categories were determined by the observed distribution among controls for each study or consortium. We applied the conditional multi-ancestry

effect estimates from the overall meta-analysis to calculate GRS for individuals from studies mentioned above. In each study, logistic regression was performed to estimate the OR and 95%CI corresponding to per SD change of GRS or each GRS category, adjusted for age, sub-study (if applicable), and up to 10 principal components. Within each population, the associations of GRS with prostate cancer risk were meta-analyzed across individual studies using a fixed-effect inverse-variance-weighted method.

GRS association in replication and overall samples.

We validated the GRS performance in independent samples that were not part of the GWAS discovery, including the Michigan Genomics Initiative⁴⁸ (MGI; European: 3,244 cases, 10,537 controls; African: 189 cases, 450 controls), Mass General Brigham Biobank^{49,50} (MGB; European: 1868 cases, 10,980 controls; African: 85 cases, 471 controls), Men of African Descent and Carcinoma of the Prostate⁵¹ (MADCaP; African: 2,505 cases, 2,160 controls), and Estonian Biobank⁵² (EstBB; European: 2,352 cases, 28,546 controls). Details of study population, genotyping and imputation were described in Supplementary Tables 1 and 2. GRS₄₅₁ and GRS₂₆₉ were constructed and weighted by the multi-ancestry conditional weights. ORs per SD and for each decile were estimated within study population using logistic regression adjusted for age, sub-study (if applicable), and up to 10 principal components.

Genome-wide PRS.

We compared our GRS₄₅₁ to a recent genome-wide PRS approach PRS-CSx⁵³, an extension of the Bayesian PRS-CS approach⁵⁴ that integrates GWAS summary statistics from multiple ancestry groups to improve cross-population polygenic modeling. We previously found that PRS-CSx was more predictive of prostate cancer risk relative to several other genome-wide PRS approaches in both European and African ancestry men³⁸. PRS-CSx was evaluated with the fully Bayesian approach to identify the optimal global shrinkage parameter ϕ , as recommended for large GWAS training data. PRS-CSx was trained on the population-specific (European, African, East Asian, and Hispanic populations) marginal GWAS summary statistics from the current investigation, using the meta=TRUE option to generate a multi-ancestry genome-wide PRS. Variants included were the 1.1 million HapMap3 panel variants⁵⁵. Populations from the 1000 Genomes Project⁵⁶ were used for LD reference panels. The resulting genome-wide PRS was evaluated in independent studies of European ancestry men from MGI and African ancestry men from MADCaP. Performance metrics included ORs calculated for the continuous standardized genome-wide PRS, adjusting for age, sub-study (if applicable), and up to 10 principal components.

GRS by Age and Disease Aggressiveness.

We investigated the association of GRS with prostate cancer risk stratified by age and its association with disease aggressiveness. In age-stratified analysis, cases and controls were both stratified into two age groups (age ≤ 55 vs. age >55 years). prostate cancer was defined as aggressive if one or more of the following criteria were met: tumor stage T3/T4, regional lymph node involvement, metastatic disease (M1), Gleason score ≥ 8 , PSA level ≥ 20 ng/mL, or prostate cancer as the underlying cause of death. Non-aggressive prostate cancer was defined as prostate cancer without aggressive features and meeting one or

more of the following criteria: Gleason score ≥ 7.0 , PSA < 20 ng/mL, and stage $\geq T2$. Logistic regressions were performed with prostate cancer status (non-aggressive vs. control, aggressive vs. control, or aggressive vs. non-aggressive) as the outcome and per SD GRS or GRS categories as the independent predictors, adjusting for age, sub-study (if applicable), and up to 10 principal components. Ancestry-specific GRS estimates were obtained via an inverse-variance weighted fixed effects meta-analysis performed within each population. Heterogeneity between stratum was assessed via a Q-statistic between effect estimates with corresponding tests of significance.

Impact of PSA screening on prostate cancer GWAS.

We compared the 128 PSA variant reported in the latest PSA GWAS⁵⁷ to the 451 prostate cancer risk variants and found 50 overlapping variants (in high LD ($R^2 > 0.8$) or identical index variant; supplementary Table 7). Three of the variants (2 of which overlapped with the PSA variants) are near the KLK3 gene, which encodes the PSA protein and are very strongly associated with PSA level. For the 48 overlapping variants (removing KLK3), it is currently difficult to differentiate whether they are prostate cancer risk variants, PSA variants or both. To better understand the likelihood of these variants being identified as the result of altering PSA levels, leading to biopsy and a prostate cancer diagnosis, we examined their aggregate effect on disease aggressiveness in our GWAS discovery samples. Additionally, we removed the 48 potential PSA variants (and 3 KLK3 variants) from the prostate cancer GRS (with 400 variants) and examine the association with aggressive versus non-aggressive prostate cancer in the multi-ancestry sample.

To account for the multiple comparisons being made in our sub-group analyses described above (in total 20 independent tests), we applied Bonferroni correction to the significance level ($0.05/20=0.0025$).

Age-specific absolute risk estimation.

Absolute risk for a given age for each GRS percentile and each population has been described previously^{2,58-1}. The approach constrains the GRS-specific absolute risks for a given age to be equivalent to the age-specific incidence for the entire population while accounting for competing causes of death. For each ancestry group, absolute risks by age t were calculated using age-specific prostate cancer incidence, $\mu(t)$, and age-specific mortality rates, $\mu_d(t)$, from the Surveillance, Epidemiology, and End Results (SEER) Program (2014–2018)^{62,63}.

Variant annotation.

Lead variants were annotated for indicators of functionality according to a framework described previously², and incorporating additional datasets. Gene-based information was obtained using wANNOVAR⁶⁴. Chromatin Immunoprecipitation Sequencing peaks were obtained from the Cistrome Data Browser⁶⁵ for the prostate cancer cell-lines LNCaP, PC3 and VCaP and prostate epithelium cell-line PrEC⁶⁶. Peak data were obtained for open chromatin (DNase-Seq and ATAC-seq), histone modifications (H3K27Ac, H3K9Ac, H3K4me1, H3K4me2 and H3K4me3), and transcription factor binding. A list of datasets included is provided in Supplementary Table 19.

Data for significant variant-gene pairs for differential gene expression (eQTLs) in three prostate tissue cohorts (GTEx v8⁶⁷, normal prostate tissue, n=221; TCGA PRAD⁶⁸, prostate adenocarcinoma, n=359; MAYO³⁰, tumor-adjacent normal prostate tissue, n=471) were obtained as described previously². All significantly associated genes at False Discovery Rate (FDR) 0.05 identified were reported for each lead variant.

Data for significant variant-gene pairs for differential gene splicing (sQTLs) were obtained for two prostate tissue cohorts. sQTLs for GTEx v8 normal prostate tissue (n=221) were downloaded from the GTEx portal. sQTLs for TCGA PRAD (n=485) were obtained from the CancerSplicingQTL database⁶⁹. All genes significantly associated with alternative splicing in the respective datasets were reported for each lead variant.

Functional enrichment permutations.

To quantify the extent to which the prostate cancer risk variants are enriched with regulatory activity compared to the genome-wide background, we performed a permutation test based on simulations. Briefly, we sought to sample 439 autosomal variants from the genomic background and compare the number of functional annotations observed with those observed in the original 439 autosomal prostate cancer risk variants. We first estimated the deciles of MAF and LD scores among the 439 prostate cancer risk variants using the combined Human Genome Diversity Project (HGDP)⁷⁰ and 1000 Genomes Project⁵⁶ datasets as reference. For a given simulation, we sampled 439 variants from the genomic background, after stratifying by the number of variants observed in the MAF and LD deciles. For a given functional category C , let $C(S)$ denote the number of variants in set S with annotation C . We computed a permutation P value as $p(C) = \frac{1}{1001} + \frac{1}{1001} \sum_s C(S) \geq C(R)$, where R denotes the 439 prostate cancer risk variants. The additional $1/1001$ term is the result of R acting as an “identity” permutation of the data and to prevent permutation P values of 0. Similarly, we computed enrichment as $e(C) = \frac{C(R)}{\bar{C}(S)}$ where $\bar{C}(S) = \frac{1}{1000} \sum_s C(S)$ represents the average number of annotated variants in the genomic background. We performed this procedure using genomic annotations from prostate eQTL and sQTL in GTEx v8⁶⁷, tumor prostate eQTL in TCGA PRAD⁶⁸, and cis-regulatory elements (CRE) in prostate samples using EnTEX/ENCODE annotations⁷¹.

Fitting prediction models of gene expression in prostate tissues.

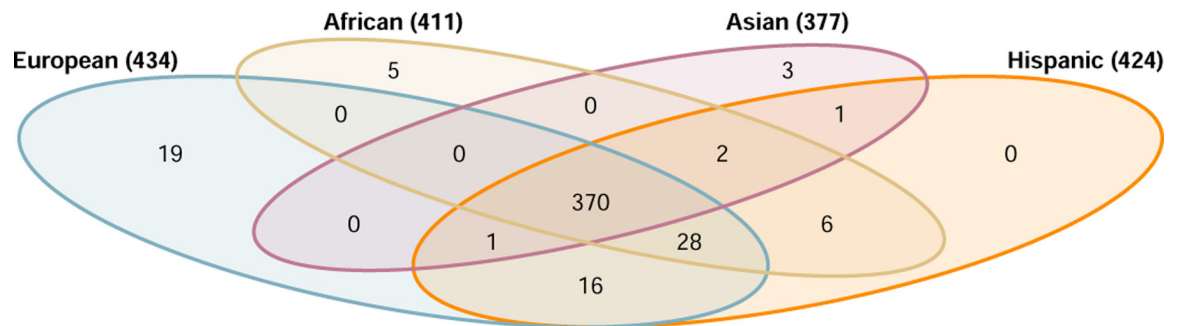
To perform a TWAS, we fitted predictive models using genotype and mRNA measurements from samples of normal prostate in GTEx v8 (n=221)²⁹ and histologically normal prostate in refZ (n=471)³⁰. We performed quality control (QC) on genotype data and kept only biallelic SNPs with MAF > 0.01, HWE $P > 5 \times 10^{-5}$, imputation quality score > 0.6, and were annotated in HapMap3. Using the FUSION pipeline, we estimated cis-h2g using QC'd genotypes within 1 Mb flanking the gene body (i.e., ± 500 Kb transcription start and stop sites)²⁷. For GTEx expression data, we adjusted expression models using eQTL covariates described in reference²⁹, which included 5 principal components, 30 PEER factors⁷², and two binary indicators for sequencing protocol and platform. For expression data in refZ³⁰, we adjusted expression models for histologic characteristics, percent lymphocytic population, percent epithelium present, and 14 gene expression principal components, which were defined in

refZ. We limited downstream model fitting to genes whose expression levels exhibited evidence of genetic control by testing for non-zero cis-heritability ($P < 0.01$) using GCTA⁷³. To build prediction models of expression, we fit penalized linear models using a modified version of the FUSION software which included SuSiE⁷⁴.

TWAS and PWAS using predicted gene and protein expression levels.

To perform downstream TWAS, we used the FUSION software²⁷ to integrate our fitted prostate expression models together with the current multi-ancestry GWAS summary statistics. In addition to our fitted models of prostate expression, we also downloaded prediction models of gene expression in prostate adenocarcinoma samples from TCGA ($n=468$)³¹. To test the association between genetically predicted levels of protein expression in plasma with prostate cancer risk, we downloaded prediction models fitted using the INTERVAL study ($n=3301$)³². In total, we performed $m=19,352$ association tests ($m_{\text{GTEx}}=5063$, $m_{\text{refZ}}=8632$, $m_{\text{TCGA}}=4664$, $m_{\text{INTERVAL}}=993$). We used a per-reference panel Bonferroni adjustment to determine transcriptome- or proteome-wide significance (TWAS $P < 0.05 / m_{\text{study}}$). To quantify the extent to which novel risk regions identify from TWAS replicate in larger GWAS, we also performed TWAS and PWAS using a smaller, previously published meta-analyzed GWAS summary statistics of prostate cancer ($N=234,253$)². A region exhibiting TWAS/PWAS significant signal was determined to be novel if it did fall within 250Kb of a lead GWAS variant.

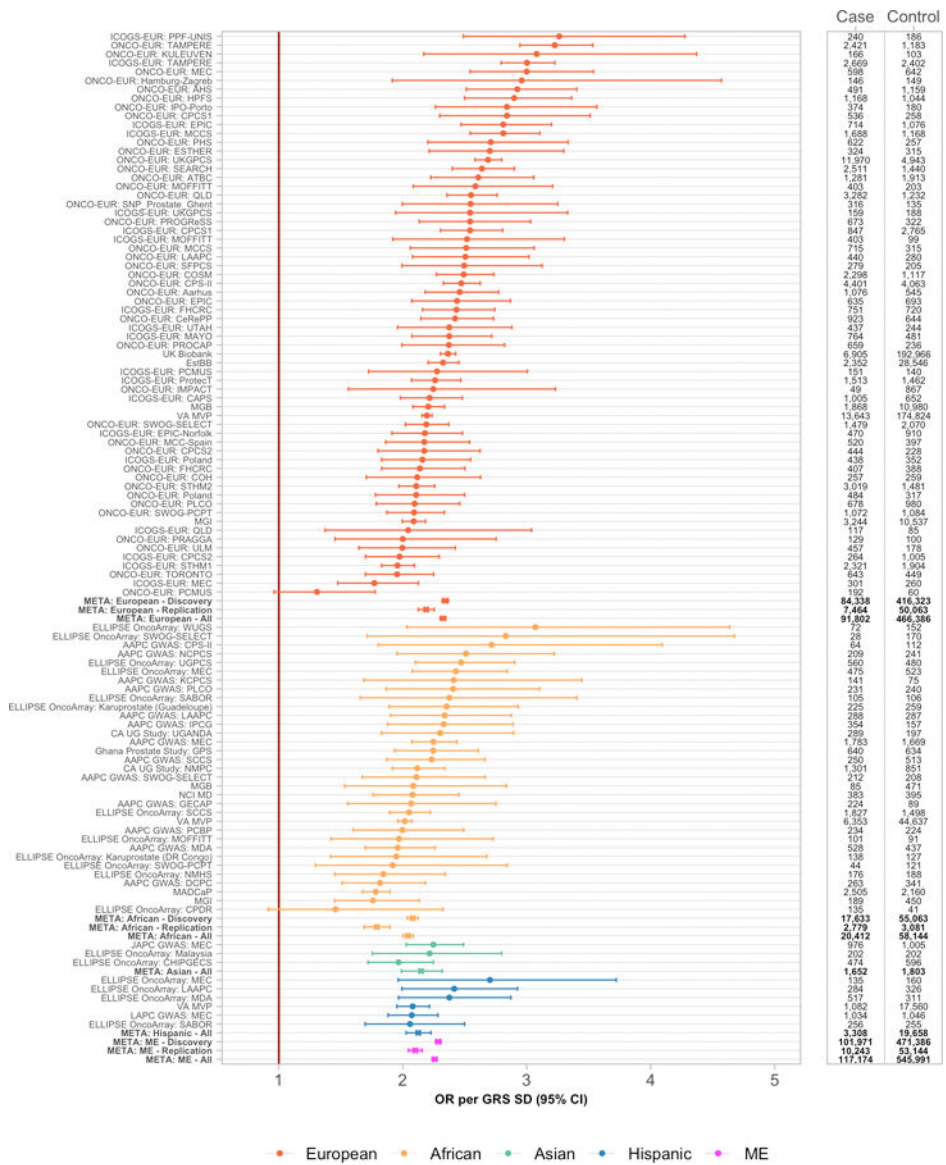
Extended Data



Extended Data Fig. 1.

Venn diagram of prostate cancer risk variants common ($MAF > 1\%$) among European, African, Asian and Hispanic populations.

The plot illustrates the distribution of 451 prostate cancer risk variants, highlighting the number of variants that are either unique to or shared among European, African, Asian, and Hispanic populations. Five variants with a minor allele frequency (MAF) of 1% across all populations are specifically included under the European population, where they have the highest MAF. Numbers in parentheses denote the total count of variants common to each respective population.



Extended Data Fig. 2.

The associations of GRS₄₅₁ and total prostate cancer risk in GWAS discovery and replication sub-studies and meta-analysis by ancestry. Odds ratios and 95% confidence intervals for one SD increase in GRS₄₅₁ and total prostate cancer risk were calculated from logistic regression. The columns ‘case’ and ‘control’ show the case and control sample sizes, respectively. ‘META’ refers to the meta-analyzed results using the inverse-variance weighted method. The y-axis shows each individual sub-studies (details of each sub-studies are available in Supplemental Table 1 and 2) and their corresponding meta-analyzed results by ancestry and study phase (GWAS discovery or replication), as well as overall meta-analyzed results.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Anqi Wang¹, Jiayi Shen¹, Alex A. Rodriguez², Edward J. Saunders³, Fei Chen¹, Rohini Janivara⁴, Burcu F. Darst^{1,5}, Xin Sheng¹, Yili Xu¹, Alisha J. Chou¹, Sara Benlloch⁶, Tokhir Dadaev³, Mark N. Brook³, Anna Plym^{7,8}, Ali Sahimi⁹, Thomas J. Hoffman^{10,11}, Atushi Takahashi^{12,13}, Koichi Matsuda¹⁴, Yukihide Momozawa¹⁵, Masashi Fujita¹⁶, Triin Laisk¹⁷, Jéssica Figuerêdo¹⁷, Kenneth Muir^{18,19}, Shuji Ito^{20,21}, Xiaoxi Liu²², The Biobank Japan Project*, Yuji Uchio²⁰, Michiaki Kubo²¹, Yoichiro Kamatani^{22,23}, Artitaya Lophatananon²⁴, Peggy Wan⁹, Caroline Andrews²⁵, Adriana Lori²⁶, Parichoy P. Choudhury²⁷, Johanna Schleutker^{28,29}, Teuvo L.J. Tammela³⁰, Csilla Sipeky²⁸, Anssi Auvinen³¹, Graham G. Giles^{32,33,34}, Melissa C. Southey³⁴, Robert J. MacInnis^{32,33}, Cezary Cybulski³⁵, Dominika Wokolorczyk³⁵, Jan Lubinski³⁵, Christopher T. Rentsch^{36,37,38}, Kelly Cho^{39,40}, Benjamin H. McMahon⁴¹, David E. Neal^{42,43,44}, Jenny L. Donovan⁴⁵, Freddie C. Hamdy^{46,47}, Richard M. Martin^{48,49,50}, Borge G. Nordestgaard^{51,52}, Sune F. Nielsen^{51,52}, Maren Weischer⁵², Stig E. Bojesen^{51,52}, Andreas Røder^{53,54}, Hein V. Stroomberg⁵³, Jyotsna Batra^{55,56}, Suzanne Chambers⁵⁷, Lisa Horvath^{58,59}, Judith A. Clements^{55,56}, Wayne Tilly⁶⁰, Gail P. Risbridger^{61,62}, Henrik Gronberg⁶³, Markus Aly^{63,64,65}, Robert Szulkin^{63,66}, Martin Eklund⁶³, Tobias Nordstrom^{63,67}, Nora Pashayan^{68,69,70}, Alison M. Dunning⁶⁹, Maya Ghousaini⁷¹, Ruth C. Travis⁷², Tim J. Key⁷², Elio Riboli⁷³, Jong Y. Park⁷⁴, Thomas A. Sellers⁷⁴, Hui-Yi Lin⁷⁵, Demetrius Albanes⁷⁶, Stephanie Weinstein⁷⁶, Michael B. Cook⁷⁷, Lorelei A. Mucci⁷⁸, Edward Giovannucci⁷⁸, Sara Lindstrom⁷⁹, Peter Kraft⁸⁰, David J. Hunter⁸¹, Kathryn L. Penney⁸², Constance Turman⁸⁰, Catherine M. Tangen⁸³, Phyllis J. Goodman⁸³, Ian M. Thompson Jr.⁸⁴, Robert J. Hamilton^{85,86}, Neil E. Fleshner⁸⁵, Antonio Finelli⁸⁷, Marie-Élise Parent⁸⁸, Janet L. Stanford⁵, Elaine A. Ostrander⁸⁹, Stella Koutros⁷⁶, Laura E. Beane Freeman⁷⁶, Meir Stampfer⁸², Alicja Wolk^{90,91}, Niclas Håkansson⁹⁰, Gerald L. Andriole⁹², Robert N. Hoover⁷⁶, Mitchell J. Machiela⁷⁶, Karina Dalsgaard Sørensen^{93,94}, Michael Borre^{95,94}, William J. Blot^{96,97}, Wei Zheng⁹⁶, Edward D. Yeboah⁹⁸, James E. Mensah^{99,98}, Yong-Jie Lu¹⁰⁰, Hong-Wei Zhang¹⁰¹, Ninghan Feng¹⁰², Xueying Mao¹⁰⁰, Yudong Wu¹⁰³, Shan-Chao Zhao¹⁰⁴, Zan Sun¹⁰⁵, Stephen N. Thibodeau¹⁰⁶, Shannon K. McDonnell¹⁰⁷, Daniel J. Schaid¹⁰⁷, Catharine M.L. West¹⁰⁸, Gill Barnett¹⁰⁹, Christiane Maier¹¹⁰, Thomas Schnoeller¹¹¹, Manuel Luedeke¹¹⁰, Adam S. Kibel¹¹², Bettina F. Drake¹¹³, Olivier Cussenot^{114,115}, Geraldine Cancel-Tassin^{114,115}, Florence Menegaux¹¹⁶, Thérèse Truong¹¹⁶, Yves Akoli Koudou¹¹⁷, Esther M. John¹¹⁸, Eli Marie Grindedal¹¹⁹, Lovise Maehle¹¹⁹, Kay-Tee Khaw¹²⁰, Sue A. Ingles¹²¹, Mariana C Stern¹²¹, Ana Vega^{122,123,124}, Antonio Gómez-Caamaño¹²⁵, Laura Fachal^{126,127,123,124}, Barry S. Rosenstein^{128,129}, Sarah L. Kerns¹³⁰, Harry Ostrer¹³¹, Manuel R. Teixeira^{132,133,134}, Paula Paulo¹³³, Andreia Brandão¹³³, Stephen Watya¹³⁵, Alexander Lubwama¹³⁵, Jeannette T. Bensen^{136,137}, Ebonee N. Butler¹³⁶, James L. Mohler^{138,139}, Jack A. Taylor^{140,141}, Manolis Kogevinas^{142,143,144,145}, Trinidad Dierssen-Sotos^{146,145}, Gemma Castaño-Vinyals^{142,143,144,145}, Lisa Cannon-Albright^{147,148}, Craig C. Teerlink^{147,148}, Chad D. Huff¹⁴⁹, Patrick Pilie¹⁵⁰, Yao Yu¹⁴⁹, Ryan J. Bohlender¹⁴⁹, Jian Gu¹⁴⁹, Sara

S. Strom¹⁵¹, Luc Multigner¹⁵², Pascal Blanchet¹⁵³, Laurent Brureau¹⁵³, Radka Kaneva¹⁵⁴, Chavdar Slavov¹⁵⁵, Vanio Mitev¹⁵⁴, Robin J. Leach¹⁵⁶, Hermann Brenner^{157,158,159}, Xuechen Chen¹⁵⁷, Bernd Holleczeck¹⁶⁰, Ben Schöttker¹⁵⁷, Eric A. Klein^{161,162}, Ann W. Hsing¹⁶³, Rick A. Kittles¹⁶⁴, Adam B. Murphy¹⁶⁵, Christopher J. Logothetis¹⁶⁶, Jeri Kim¹⁶⁶, Susan L. Neuhausen¹⁶⁷, Linda Steele¹⁶⁷, Yuan Chun Ding¹⁶⁷, William B. Isaacs¹⁶⁸, Barbara Nemesure¹⁶⁹, Anselm J.M. Hennis^{169,170}, John Carpten¹⁷¹, Hardev Pandha¹⁷², Agnieszka Michael¹⁷², Kim De Ruyck¹⁷³, Gert De Meerleer¹⁷⁴, Piet Ost¹⁷⁴, Jianfeng Xu¹⁷⁵, Azad Razack¹⁷⁶, Jasmine Lim¹⁷⁶, Soo-Hwang Teo¹⁷⁷, Lisa F. Newcomb^{178,179}, Daniel W. Lin^{178,179}, Jay H. Fowke¹⁸⁰, Christine M. Neslund-Dudas¹⁸¹, Benjamin A. Rybicki¹⁸¹, Marija Gamulin¹⁸², Davor Lessel¹⁸³, Tomislav Kulis¹⁸⁴, Nawaid Usmani^{185,186}, Aswin Abraham^{185,186}, Sandeep Singhal¹⁸⁵, Matthew Parliament^{185,186}, Frank Claessens¹⁸⁷, Steven Joniau¹⁸⁸, Thomas Van den Broeck^{187,188}, Manuela Gago-Dominguez^{189,190}, Jose Esteban Castelao¹⁹¹, Maria Elena Martinez¹⁹², Samantha Larkin¹⁹³, Paul A. Townsend¹⁹⁴, Claire Aukim-Hastie¹⁹⁵, William S. Bush¹⁹⁶, Melinda C. Aldrich¹⁹⁷, Dana C. Crawford¹⁹⁶, Shiv Srivastava¹⁹⁸, Jennifer Cullen^{199,196}, Gyorgy Petrovics¹⁹⁹, Graham Casey²⁰⁰, Ying Wang²⁶, Yao Tettey^{201,99}, Joseph Lachance⁴, Wei Tang²⁰², Richard B. Biritwum⁹⁹, Andrew A. Adjei²⁰³, Evelyn Tay⁹⁹, Ann Truelove²⁰⁴, Shelley Niwa²⁰⁴, Kosj Yamoah^{205,74}, Koveela Govindasami³, Anand P. Chokkalingam²⁰⁶, Jacob M. Keaton^{96,207}, Jacklyn N. Hellwege^{96,208}, Peter E. Clark²⁰⁹, Mohamed Jalloh²¹⁰, Serigne M. Gueye²¹⁰, Lamine Niang²¹⁰, Olufemi Ogunbiyi²¹¹, Olayiwola Shittu²¹², Olukemi Amodu²¹³, Akindele O. Adebisi²¹⁴, Oseremen I. Aisuodionoe-Shadrach²¹⁵, Hafees O. Ajibola²¹⁵, Mustapha A. Jamda²¹⁵, Olabode P. Oluwole²¹⁵, Maxwell Nwegbu²¹⁵, Ben Adusei²¹⁶, Sunny Mante²¹⁶, Afua Darkwa-Abrahams⁹⁹, Halimatou Diop²¹⁷, Susan M. Gundell¹, Monique J. Roobol²¹⁸, Guido Jenster²¹⁸, Ron H.N. van Schaik²¹⁹, Jennifer J. Hu²²⁰, Maureen Sanderson²²¹, Linda Kachuri²²², Rohit Varma²²³, Roberta McKean-Cowdin⁹, Mina Torres²²³, Michael H. Preuss²²⁴, Ruth J.F. Loos²²⁴, Matthew Zawistowski^{225,226}, Sebastian Zöllner^{225,226,227}, Zeyun Lu⁹, Stephen K. Van Den Eeden²²⁸, Douglas F. Easton²²⁹, Stefan Ambs²⁰², Todd L. Edwards⁹⁶, Reedik Mägi¹⁷, Timothy R. Rebbeck²³⁰, Lars Fritsche²³¹, Stephen J. Chanock²³², Sonja I. Berndt²³², Fredrik Wiklund⁷, Hidewaki Nakagawa²³³, John S. Witte^{222,234}, J. Michael Gaziano^{39,40}, Amy C. Justice³⁶, Nick Mancuso¹, Chikashi Terao^{235,236,237}, Rosalind A. Eeles^{3,238}, Zsafia Kote-Jarai³, Ravi K. Madduri², David V. Conti¹, Christopher A. Haiman¹

Affiliations

¹Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

²Argonne National Laboratory, Lemont, IL, USA

³The Institute of Cancer Research, London, UK

⁴School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

- ⁵Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
- ⁶Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK
- ⁷Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
- ⁸Urology Division, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- ⁹Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
- ¹⁰Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA
- ¹¹Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA
- ¹²Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
- ¹³Department of Genomic Medicine, National Cerebral and Cardiovascular Center Research Institute, Suita, Japan
- ¹⁴Department of Computational Biology and Medical Sciences, Laboratory of Clinical Genome Sequencing, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan
- ¹⁵Laboratory for Genotyping Development, RIKEN Center of Integrative Medical Sciences, Yokohama, Japan
- ¹⁶Laboratory for Cancer Genomics, RIKEN Center of Integrative Medical Sciences, Yokohama, Japan
- ¹⁷Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
- ¹⁸Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
- ¹⁹Warwick Medical School, University of Warwick, Coventry, UK
- ²⁰Department of Orthopaedics, Shimane University, Izumo, Shimane, Japan
- ²¹RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
- ²²Laboratory for Statistical and Translational Genetics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan
- ²³Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

- ²⁴Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester, UK
- ²⁵Harvard TH Chan School of Public Health and Division of Population Sciences, Dana Farber Cancer Institute, Boston, MA, USA
- ²⁶Department of Population Science, American Cancer Society, Kennesaw, GA, USA
- ²⁷Surveillance & Health Science Department, American Cancer Society, Kennesaw, GA, USA
- ²⁸Institute of Biomedicine, University of Turku, Turku, Finland
- ²⁹Department of Medical Genetics, Genomics, Laboratory Division, Turku University Hospital, Turku, Finland
- ³⁰Department of Urology, Tampere University Hospital, Tampere, Finland
- ³¹Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland
- ³²Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Australia
- ³³Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia
- ³⁴Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
- ³⁵International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
- ³⁶Yale School of Medicine, New Haven, CT, USA
- ³⁷Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK
- ³⁸VA Connecticut Healthcare System, West Haven, CT, USA
- ³⁹Division of Aging, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
- ⁴⁰VA Boston Healthcare System, Boston, MA, USA
- ⁴¹Theoretical Biology Division, Los Alamos National Lab, Los Alamos, NM, USA
- ⁴²Nuffield Department of Surgical Sciences, University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK
- ⁴³University of Cambridge, Department of Oncology, Addenbrooke's Hospital, Cambridge, UK
- ⁴⁴Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

- ⁴⁵Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
- ⁴⁶Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
- ⁴⁷Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK
- ⁴⁸Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
- ⁴⁹NIHR Bristol Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol, Bristol, UK
- ⁵⁰Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK
- ⁵¹Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
- ⁵²Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Copenhagen, Denmark
- ⁵³Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
- ⁵⁴Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
- ⁵⁵Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia
- ⁵⁶Translational Research Institute, Brisbane, Queensland, Australia
- ⁵⁷University of Technology, Sydney, Australia
- ⁵⁸Chris O'Brien Lifehouse (COBLH), Camperdown, Sydney, NSW, Australia, Sydney, Australia
- ⁵⁹Garvan Institute of Medical Research, Sydney, Australia
- ⁶⁰Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide, Adelaide, Australia
- ⁶¹Department of Anatomy and Developmental Biology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia
- ⁶²Prostate Cancer Translational Research Program, Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia
- ⁶³Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
- ⁶⁴Department of Molecular Medicine and Surgery, Karolinska Institutet, and Department of Urology, Karolinska University Hospital, Solna, Stockholm, Sweden

- ⁶⁵Department of Urology, Karolinska University Hospital, Stockholm, Sweden
- ⁶⁶SDS Life Sciences, Stockholm, Sweden
- ⁶⁷Department of Clinical Sciences at Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden
- ⁶⁸University College London, Department of Applied Health Research, London, UK
- ⁶⁹Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Cambridge, UK
- ⁷⁰Department of Applied Health Research, University College London, London, UK
- ⁷¹Open Targets, Wellcome Sanger Institute, Hinxton, Saffron Walden, Hinxton, UK
- ⁷²Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
- ⁷³Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
- ⁷⁴Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA
- ⁷⁵School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA, USA
- ⁷⁶Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA
- ⁷⁷Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH,, Bethesda, MD, USA
- ⁷⁸Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA
- ⁷⁹Department of Epidemiology, University of Washington, Seattle, WA, USA
- ⁸⁰Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA
- ⁸¹Nuffield Department of Population Health, University of Oxford, Oxford, UK
- ⁸²Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA, USA
- ⁸³SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
- ⁸⁴CHRISTUS Santa Rosa Hospital – Medical Center, San Antonio, TX, USA
- ⁸⁵Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Canada
- ⁸⁶Dept. of Surgery (Urology), University of Toronto, Toronto, Canada
- ⁸⁷Division of Urology, Princess Margaret Cancer Centre, Toronto, Canada

- ⁸⁸Epidemiology and Biostatistics Unit, Centre Armand-Frappier Santé Biotechnologie, Laval, QC, Canada
- ⁸⁹National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
- ⁹⁰Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
- ⁹¹Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
- ⁹²Brady Urological Institute in National Capital Region, Johns Hopkins University, Baltimore, MD, USA
- ⁹³Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark
- ⁹⁴Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
- ⁹⁵Department of Urology, Aarhus University Hospital, Aarhus, Denmark
- ⁹⁶Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
- ⁹⁷International Epidemiology Institute, Rockville, MD, USA
- ⁹⁸University of Ghana Medical School, Accra, Ghana
- ⁹⁹Korle Bu Teaching Hospital, Accra, Ghana
- ¹⁰⁰Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London, UK
- ¹⁰¹Second Military Medical University, Shanghai, P.R.China
- ¹⁰²Wuxi Second Hospital, Nanjing Medical University, Wuxi, Jiangzhu Province, China
- ¹⁰³Department of Urology, First Affiliated Hospital, The Academy of Medical Sciences, Zhengzhou University, Zhengzhou, China
- ¹⁰⁴Department of Urology, Nanfang Hospital, Southern Medical University, Guangzhou, China
- ¹⁰⁵The People's Hospital of Liaoning Proviouce, The People's Hospital of China Medical University, Shenyang, China, Shenyang, China
- ¹⁰⁶Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA
- ¹⁰⁷Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA
- ¹⁰⁸Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, UK
- ¹⁰⁹University of Cambridge Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

- ¹¹⁰Humangenetik Tuebingen, Tuebingen, Germany
- ¹¹¹Department of Urology, University Hospital Ulm, Ulm, Germany
- ¹¹²Division of Urologic Surgery, Brigham and Womens Hospital, Boston, MA, USA
- ¹¹³Washington University School of Medicine, St. Louis, MO, USA
- ¹¹⁴GRC 5 Predictive Onco-Urology, Sorbonne Université, Paris, France
- ¹¹⁵CeRePP, Paris, France
- ¹¹⁶Exposome and Heredity, CESP (UMR 1018), Paris-Saclay Medical School, Paris-Saclay University, Inserm, Gustave Roussy, Villejuif, France
- ¹¹⁷Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif Cédex, France
- ¹¹⁸Department of Medicine, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
- ¹¹⁹Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
- ¹²⁰Clinical Gerontology Unit, University of Cambridge, Cambridge, UK
- ¹²¹Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA
- ¹²²Fundación Pública Galega Medicina Xenómica, Santiago De Compostela, Spain
- ¹²³Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela, Spain
- ¹²⁴Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain
- ¹²⁵Department of Radiation Oncology, Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain
- ¹²⁶Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK
- ¹²⁷Fundación Pública Galega Medicina Xenómica, Santiago de Compostela, Spain
- ¹²⁸Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- ¹²⁹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
- ¹³⁰Department of Radiation Oncology, Medical College of Wisconsin, Milwaukee, WI, USA
- ¹³¹Professor of Pathology and Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA

¹³²Department of Laboratory Genetics, Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal

¹³³Cancer Genetics Group, IPO Porto Research Center (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal

¹³⁴School of Medicine and Biomedical Sciences (ICBAS), University of Porto, Porto, Portugal

¹³⁵Uro Care, Kampala, Uganda

¹³⁶Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹³⁷Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹³⁸Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹³⁹Department of Urology, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

¹⁴⁰Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

¹⁴¹Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

¹⁴²ISGlobal, Barcelona, Spain

¹⁴³IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

¹⁴⁴Universitat Pompeu Fabra (UPF), Barcelona, Spain

¹⁴⁵CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

¹⁴⁶University of Cantabria-IDIVAL, Santander, Spain

¹⁴⁷Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA

¹⁴⁸George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT, USA

¹⁴⁹Department of Epidemiology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

¹⁵⁰Department of Genitourinary Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

¹⁵¹The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

¹⁵²Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), Rennes, France

- ¹⁵³CHU de Pointe-à-Pitre, Univ Antilles, Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail), Pointe-à-Pitre, France
- ¹⁵⁴Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, Bulgaria
- ¹⁵⁵Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, Sofia, Bulgaria
- ¹⁵⁶Department of Cell Systems and Anatomy and Mays Cancer Center, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
- ¹⁵⁷Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
- ¹⁵⁸German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
- ¹⁵⁹Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
- ¹⁶⁰Saarland Cancer Registry, Saarbrücken, Germany
- ¹⁶¹Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA
- ¹⁶²Glickman Urological & Kidney Institute, Cleveland Clinic, Cleveland, OH, USA
- ¹⁶³Department of Medicine and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
- ¹⁶⁴City of Hope Comprehensive Cancer Center, Duarte, CA, USA
- ¹⁶⁵Department of Urology, Northwestern University, Chicago, IL, USA
- ¹⁶⁶The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, TX, USA
- ¹⁶⁷Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA, USA
- ¹⁶⁸James Buchanan Brady Urological Institute, Johns Hopkins Hospital and Medical Institution, Baltimore, MD, USA
- ¹⁶⁹Department of Family, Population and Preventive Medicine, Stony Brook University, Stony Brook, NY, USA
- ¹⁷⁰Chronic Disease Research Centre and Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados
- ¹⁷¹Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
- ¹⁷²The University of Surrey, Guildford, Surrey, UK
- ¹⁷³Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Ghent, Belgium

- ¹⁷⁴Ghent University Hospital, Department of Radiotherapy, Ghent, Belgium
- ¹⁷⁵Program for Personalized Cancer Care and Department of Surgery, NorthShore University HealthSystem, Evanston, IL, USA
- ¹⁷⁶Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
- ¹⁷⁷Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Subang Jaya, Selangor, Malaysia
- ¹⁷⁸Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
- ¹⁷⁹Department of Urology, University of Washington, Seattle, WA, USA
- ¹⁸⁰Department of Preventive Medicine, Division of Epidemiology, The University of Tennessee Health Science Center, Memphis, TN, USA
- ¹⁸¹Department of Public Health Sciences, Henry Ford Hospital, Detroit, Detroit, MI, USA
- ¹⁸²Division of Medical Oncology, Urogenital Unit, Department of Oncology, University Hospital Centre Zagreb, Zagreb, Croatia
- ¹⁸³Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
- ¹⁸⁴Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine, Zagreb, Croatia
- ¹⁸⁵Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada
- ¹⁸⁶Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada
- ¹⁸⁷Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, Leuven, Belgium
- ¹⁸⁸Department of Urology, University Hospitals Leuven, Leuven, Belgium
- ¹⁸⁹Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago de Compostela, Spain
- ¹⁹⁰University of California San Diego, Moores Cancer Center, La Jolla, CA, USA
- ¹⁹¹Genetic Oncology Unit, CHUVI Hospital, Complejo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), Vigo (Pontevedra), Spain
- ¹⁹²University of California San Diego, Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA

- ¹⁹³Scientific Education Support, Thames Ditton, Surrey, Formerly Cancer Sciences, University of Southampton, Southampton, UK
- ¹⁹⁴School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Surrey, UK
- ¹⁹⁵Scientific Education Support, Thames Ditton, Surrey, Surrey, UK
- ¹⁹⁶Department of Population and Quantitative Health Sciences, Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA
- ¹⁹⁷Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
- ¹⁹⁸Department of Biochemistry and Molecular & Cellular Biology, Georgetown University Medical Center, Washington, DC, USA
- ¹⁹⁹Department of Surgery, Center for Prostate Disease Research, Uniformed Services University of the Health Sciences, Bethesda, MD, USA
- ²⁰⁰Department of Public Health Science, Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA
- ²⁰¹Department of Pathology, University of Ghana, Accra, Ghana
- ²⁰²Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA
- ²⁰³Department of Pathology, University of Ghana Medical School, Accra, Ghana
- ²⁰⁴Westat, Rockville, MD, USA
- ²⁰⁵Department of Radiation Oncology, Moffitt Cancer Center, Tampa, FL, USA
- ²⁰⁶School of Public Health, University of California, Berkeley, Berkeley, CA, USA
- ²⁰⁷Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
- ²⁰⁸Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Nashville, TN, USA
- ²⁰⁹Atrium Health/Levine Cancer Institute, Charlotte, NC, USA
- ²¹⁰Hôpital Général Idrissa Pouye, Dakar, Senegal
- ²¹¹Department of Pathology, College of Medicine, University of Ibadan, Ibadan, Nigeria
- ²¹²Department of Surgery, College of Medicine, University of Ibadan and University College Hospital, Ibadan, Nigeria
- ²¹³Institute of Child Health, College of Medicine, University of Ibadan and University College Hospital, Ibadan, Nigeria
- ²¹⁴Clinical Epidemiology Unit, Department of Community Medicine, College of Medicine, University of Ibadan, Ibadan, Nigeria

²¹⁵College of Health Sciences, University of Abuja, University of Abuja Teaching Hospital and Cancer Science Center, Abuja, Nigeria

²¹⁶37 Military Hospital, Accra, Ghana

²¹⁷Laboratoires Bacteriologie et Virologie, Hôpital Aristide Le Dantec, Dakar, Senegal

²¹⁸Department of Urology, Erasmus University Medical Center, Rotterdam, Netherlands

²¹⁹Department of Clinical Chemistry, Erasmus University Medical Center, Rotterdam, Netherlands

²²⁰The University of Miami School of Medicine, Sylvester Comprehensive Cancer Center, Miami, FL, USA

²²¹Department of Family and Community Medicine, Meharry Medical College, Nashville, TN, USA

²²²Department of Epidemiology and Population Health, Stanford Cancer Institute, Stanford, CA, USA

²²³Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, CA, USA

²²⁴The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

²²⁵Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI, USA

²²⁶Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA

²²⁷Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

²²⁸Division of Research, Kaiser Permanente, Northern California, Oakland, CA, USA

²²⁹Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, Cambridge, UK

²³⁰Harvard TH Chan School of Public Health and Division of Population Sciences, Dana Farber Cancer Institute, Boston, MA, USA

²³¹Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

²³²Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA

²³³Laboratory for Cancer Genomics, RIKEN Center of Integrative Medical Sciences, Yokohama, Japan

²³⁴Departments of Biomedical Data Science, Stanford University, Stanford, CA, USA

²³⁵Laboratory for Statistical and Translational Genetics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan

²³⁶Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan

²³⁷The Department of Applied Genetics, School of Pharmaceutical Sciences, Shizuoka, Japan

²³⁸Royal Marsden NHS Foundation Trust, London, UK

Acknowledgements

This project was supported by the US National Institutes of Health (NIH) grants R01CA257328 (CAH), U19CA214253 (CAH), U01CA261339 (DVC), P01CA196569 (DVC), R00CA246063 (BFDarst), and the Prostate Cancer Foundation grants 20CHAS03 (CAH) and 21YOUN11 (BFDarst). We acknowledge support from The National Institute of Health Research to The Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This research has been conducted using the UK Biobank Resource under application number 42195. This research is based on data from the Million Veteran Program, Office of Research and Development, and the Veterans Health Administration. This publication does not represent the views of the Department of Veteran Affairs or the United States Government. A full description of funding and acknowledgements for each of the contributing studies can be found in the Supplementary Note.

Data Availability

The full summary statistics resulting from this investigation are available in the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) under accession codes as follows: cross-ancestry (GCST90274713), European (GCST90274714), African (GCST90274715), Asian (GCST90274716), and Hispanic (GCST90274717). Genotype and covariate data used in this study are deposited in dbGaP under accession codes phs001391.v1.p1, phs000306.v4.p1, phs001120.v2.p2, phs001221.v1.p1, phs000812.v1.p1, and phs000838.v1.p1. The variants and weights for the GRS₂₆₉ and GRS₄₅₁ are available on the PGS Catalog under accession codes PGP000122 and PGP000488, respectively (<https://www.pgscatalog.org/>). Publicly available data described in this manuscript can be found from the following websites: 1000 Genomes Project (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/>); Human Genome Diversity Project (<https://www.internationalgenome.org/data-portal/data-collection/hgdp>); SEER (<https://seer.cancer.gov/>); National Center for Health Statistics, CDC (<https://www.cdc.gov/nchs/index.htm>); Cistrome Data Browser (<http://cistrome.org/db/>); MAYO refZ (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000985.v1.p1); GTEx (<https://gtexportal.org/home/datasets>); TCGA (<https://portal.gdc.cancer.gov>); CancerSplicingQTL database (<http://www.cancersplicingqtl-hust.com/>); and EnTEX/ENCODE (<http://entex.encodeproject.org/>).

The Biobank Japan Project

Yuji Yamanashi²³⁹, Yoichi Furukawa²⁴⁰, Takayuki Morisaki²⁴¹, Yoshinori Murakami²⁴², Kaori Muto²⁴³, Akiko Nagai²⁴³, Wataru Obara²⁴⁴, Ken Yamaji²⁴⁵, Kazuhisa Takahashi²⁴⁶, Satoshi Asai^{247,248}, Yasuo Takahashi²⁴⁸, Takao Suzuki²⁴⁹, Nobuaki Sinozaki²⁴⁹, Hiroki Yamaguchi²⁵⁰, Shiro Minami²⁵¹, Shigeo Murayama²⁵², Kozo Yoshimori²⁵³, Satoshi Nagayama²⁵⁴, Daisuke Obata²⁵⁵, Masahiko Higashiyama²⁵⁶, Akihide Masumoto²⁵⁷, Yukihiro Koretsune²⁵⁸

²³⁹Division of Genetics, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²⁴⁰Division of Clinical Genome Research, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²⁴¹Division of Molecular Pathology IMSUT Hospital, Department of Internal Medicine Project Division of Genomic Medicine and Disease Prevention, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²⁴²Department of Cancer Biology, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²⁴³Department of Public Policy, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan, ²⁴⁴Department of Urology, Iwate Medical University, Iwate, Japan, ²⁴⁵Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan, ²⁴⁶Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan, ²⁴⁷Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan, ²⁴⁸Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University School of Medicine, Tokyo, Japan, ²⁴⁹Tokushukai Group, Tokyo, Japan, ²⁵⁰Department of Hematology, Nippon Medical School, Tokyo, Japan, ²⁵¹Department of Bioregulation, Nippon Medical School, Kawasaki, Japan, ²⁵²Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan, ²⁵³Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan, ²⁵⁴The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan, ²⁵⁵Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan, ²⁵⁶Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan, ²⁵⁷Iizuka Hospital, Fukuoka, Japan, ²⁵⁸National Hospital Organization Osaka National Hospital, Osaka, Japan

References

1. Bray F et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68, 394–424 (2018). [PubMed: 30207593]
2. Conti DV et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature Genetics* 53, 65–75 (2021). [PubMed: 33398198]
3. Schumacher FR et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 50, 928–936 (2018). [PubMed: 29892016]
4. Dadaev T et al. Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat Commun* 9, 2256 (2018). [PubMed: 29892050]
5. Conti DV et al. Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry. *J Natl Cancer Inst* 109(2017).
6. Wang M et al. Large-scale association analysis in Asians identifies new susceptibility loci for prostate cancer. *Nat Commun* 6, 8469 (2015). [PubMed: 26443449]
7. Hoffmann TJ et al. A large multiethnic genome-wide association study of prostate cancer identifies novel risk variants and substantial ethnic differences. *Cancer Discov* 5, 878–91 (2015). [PubMed: 26034056]
8. Al Olama AA et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 46, 1103–9 (2014). [PubMed: 25217961]
9. Eeles RA et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 45, 385–91, 391e1–2 (2013). [PubMed: 23535732]
10. Gudmundsson J et al. A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat Genet* 44, 1326–9 (2012). [PubMed: 23104005]

11. Takata R et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat Genet* 42, 751–4 (2010). [PubMed: 20676098]
12. Gudmundsson J et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 40, 281–3 (2008). [PubMed: 18264098]
13. Amundadottir LT et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 38, 652–8 (2006). [PubMed: 16682969]
14. Wojcik GL et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570, 514–518 (2019). [PubMed: 31217584]
15. Newcombe PJ, Conti DV & Richardson S JAM: A Scalable Bayesian Framework for Joint Analysis of Marginal SNP Effects. *Genet Epidemiol* 40, 188–201 (2016). [PubMed: 27027514]
16. Chen F et al. Evidence of Novel Susceptibility Variants for Prostate Cancer and a Multiancestry Polygenic Risk Score Associated with Aggressive Disease in Men of African Ancestry. *Eur Urol* 84, 13–21 (2023). [PubMed: 36872133]
17. Barfeld SJ, East P, Zuber V & Mills IG Meta-analysis of prostate cancer gene expression data identifies a novel discriminatory signature enriched for glycosylating enzymes. *BMC Med Genomics* 7, 513 (2014). [PubMed: 25551447]
18. Halvorsen OJ et al. Increased Expression of SIM2-s Protein Is a Novel Marker of Aggressive Prostate Cancer. *Clinical Cancer Research* 13, 892–897 (2007). [PubMed: 17289882]
19. Dhanasekaran SM et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 412, 822–826 (2001). [PubMed: 11518967]
20. Linn DE et al. Differential regulation of androgen receptor by PIM-1 kinases via phosphorylation-dependent recruitment of distinct ubiquitin E3 ligases. *J Biol Chem* 287, 22959–68 (2012). [PubMed: 22584579]
21. Wang J et al. Pim1 kinase synergizes with c-MYC to induce advanced prostate carcinoma. *Oncogene* 29, 2477–2487 (2010). [PubMed: 20140016]
22. Sun X et al. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet* 37, 407–12 (2005). [PubMed: 15750593]
23. Chandler JD, Williams ED, Slavin JL, Best JD & Rogers S Expression and localization of GLUT1 and GLUT12 in prostate carcinoma. *Cancer* 97, 2035–2042 (2003). [PubMed: 12673735]
24. White MA et al. GLUT12 promotes prostate cancer cell growth and is regulated by androgens and CaMKK2 signaling. *Endocr Relat Cancer* 25, 453–469 (2018). [PubMed: 29431615]
25. Mi Y et al. Down-regulation of Barx2 predicts poor survival in colorectal cancer. *Biochemical and Biophysical Research Communications* 478, 67–73 (2016). [PubMed: 27453340]
26. Zhang J et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet* 54, 593–602 (2022). [PubMed: 35501419]
27. Gusev A et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48, 245–52 (2016). [PubMed: 26854917]
28. Gamazon ER et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 47, 1091–8 (2015). [PubMed: 26258848]
29. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330 (2020). [PubMed: 32913098]
30. Thibodeau SN et al. Identification of candidate genes for prostate cancer-risk SNPs utilizing a normal prostate tissue eQTL data set. *Nat Commun* 6, 8653 (2015). [PubMed: 26611117]
31. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011–25 (2015). [PubMed: 26544944]
32. Sun BB et al. Genomic atlas of the human plasma proteome. *Nature* 558, 73–79 (2018). [PubMed: 29875488]
33. Mancuso N et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat Commun* 9, 4079 (2018). [PubMed: 30287866]
34. Liu D et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. *Int J Cancer* 150, 80–90 (2022). [PubMed: 34520569]

35. Amin AI Olama A et al. Multiple novel prostate cancer susceptibility signals identified by fine-mapping of known risk loci among Europeans. *Hum Mol Genet* 24, 5589–602 (2015). [PubMed: 26025378]
36. Pencina MJ, D'Agostino RB Sr. & Steyerberg EW Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 30, 11–21 (2011). [PubMed: 21204120]
37. Chen F et al. Validation of a multi-ancestry polygenic risk score and age-specific risks of prostate cancer: A meta-analysis within diverse populations. *eLife* 11, e78304 (2022). [PubMed: 35801699]
38. Darst BF et al. Evaluating approaches for constructing polygenic risk scores for prostate cancer in men of African and European ancestry. *Am J Hum Genet* 110, 1200–1206 (2023). [PubMed: 37311464]

Methods-only References

39. Das S et al. Next-generation genotype imputation service and methods. *Nat Genet* 48, 1284–1287 (2016). [PubMed: 27571263]
40. Howie BN, Donnelly P & Marchini J A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5, e1000529 (2009). [PubMed: 19543373]
41. Loh P-R et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* 48, 1443–1448 (2016). [PubMed: 27694958]
42. Kurki MI et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 613, 508–518 (2023). [PubMed: 36653562]
43. Auton A et al. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
44. Taliun D et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 590, 290–299 (2021). [PubMed: 33568819]
45. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209 (2018). [PubMed: 30305743]
46. de Bakker PI et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 17, R122–8 (2008). [PubMed: 18852200]
47. Conti DV et al. Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry. *J Natl Cancer Inst* 109(2017).
48. Zawistowski M et al. The Michigan Genomics Initiative: A biobank linking genotypes and electronic clinical records in Michigan Medicine patients. *Cell Genom* 3, 100257 (2023). [PubMed: 36819667]
49. Karlson EW, Boutin NT, Hoffnagle AG & Allen NL Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *J Pers Med* 6(2016).
50. Plym A et al. Evaluation of a Multiethnic Polygenic Risk Score Model for Prostate Cancer. *J Natl Cancer Inst* 114, 771–774 (2022). [PubMed: 33792693]
51. Andrews C et al. Development, Evaluation, and Implementation of a Pan-African Cancer Research Network: Men of African Descent and Carcinoma of the Prostate. *J Glob Oncol* 4, 1–14 (2018).
52. Leitsalu L et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* 44, 1137–47 (2015). [PubMed: 24518929]
53. Ruan Y et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 54, 573–580 (2022). [PubMed: 35513724]
54. Ge T, Chen CY, Ni Y, Feng YA & Smoller JW Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10, 1776 (2019). [PubMed: 30992449]
55. International HapMap C et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–8 (2010). [PubMed: 20811451]

56. Genomes Project C et al. A global reference for human genetic variation. *Nature* 526, 68–74 (2015). [PubMed: 26432245]
57. Kachuri L et al. Genetically adjusted PSA levels for prostate cancer screening. *Nat Med* 29, 1412–1423 (2023). [PubMed: 37264206]
58. Amin Al Olama A et al. Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiol Biomarkers Prev* 24, 1121–9 (2015). [PubMed: 25837820]
59. Antoniou AC et al. Common Breast Cancer Susceptibility Alleles and the Risk of Breast Cancer for BRCA1 and BRCA2 Mutation Carriers: Implications for Risk Prediction. *Cancer research* 70, 9742–9754 (2010). [PubMed: 21118973]
60. Antoniou AC et al. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genetic epidemiology* 21, 1–18 (2001). [PubMed: 11443730]
61. Kuchenbaecker KB et al. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *J Natl Cancer Inst* 109(2017).
62. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Plus Data, 18 Registries, Nov 2020 Sub (2000–2018) - Linked To County Attributes - Total U.S., 1969–2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.
63. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality - All COD, Aggregated With State, Total U.S. (1990–2019) <Katrina/Rita Population Adjustment>, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021. Underlying mortality data provided by NCHS (www.cdc.gov/nchs).
64. Chang X & Wang K wANNOVAR: annotating genetic variants for personal genomes via the web. *J Med Genet* 49, 433–6 (2012). [PubMed: 22717648]
65. Mei S et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* 45, D658–d662 (2017). [PubMed: 27789702]
66. Lawrence M, Gentleman R & Carey V rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25, 1841–2 (2009). [PubMed: 19468054]
67. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45, 580–5 (2013). [PubMed: 23715323]
68. Gong J et al. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res* 46, D971–D976 (2018). [PubMed: 29036324]
69. Tian J et al. CancerSplicingQTL: a database for genome-wide identification of splicing QTLs in human cancer. *Nucleic Acids Research* 47, D909–D916 (2018).
70. Bergstrom A et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367, eaay5012 (2020). [PubMed: 32193295]
71. Rozowsky J et al. The EN-TEEx resource of multi-tissue personal epigenomes & variant-impact models. *Cell* 186, 1493–1511 e40 (2023). [PubMed: 37001506]
72. Stegle O, Parts L, Durbin R & Winn J A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol* 6, e1000770 (2010). [PubMed: 20463871]
73. Yang J, Lee SH, Goddard ME & Visscher PM GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76–82 (2011). [PubMed: 21167468]
74. Wang G, Sarkar A, Carbonetto P & Stephens M A simple new approach to variable selection in regression, with application to genetic fine mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82, 1273–1300 (2020). [PubMed: 37220626]

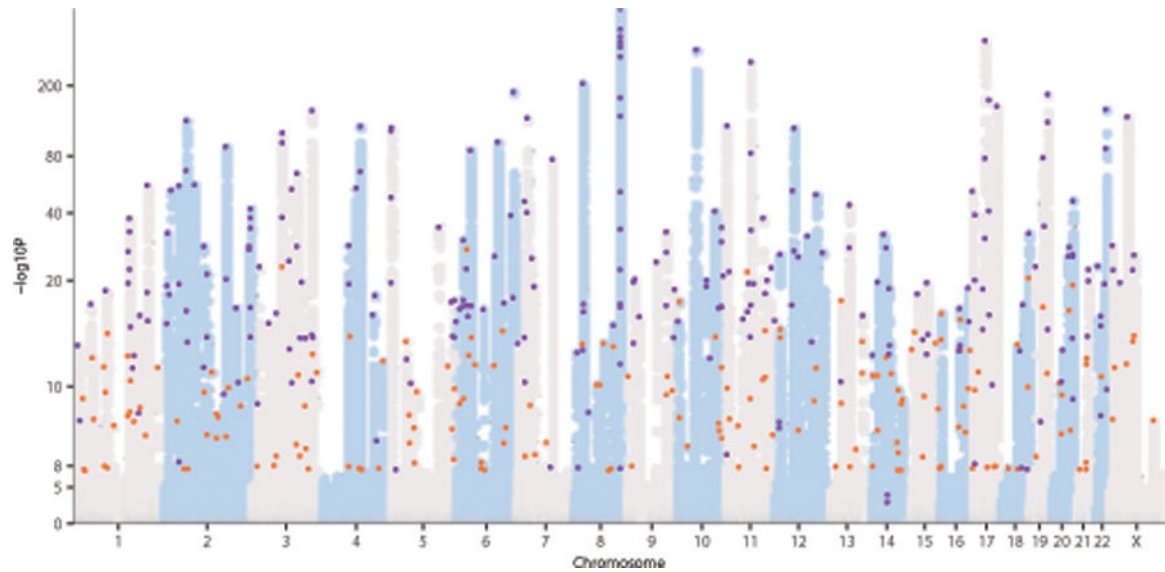


Figure 1.

Manhattan plot of results from the multi-ancestry prostate cancer meta-analysis.

Multi-ancestry meta-analysis (156,319 cases and 788,443 controls) was performed using an inverse-variance-weighted fixed-effects model. Nominal statistical significance is shown as $-\log_{10}P$ (two-sided) of z statistics on the y axis. Purple and orange circles indicate previously known or novel risk variants, respectively, that were genome-wide significant in multi-ancestry or ancestry-specific meta-analyses. The plot is truncated at $-\log_{10}P=600$.

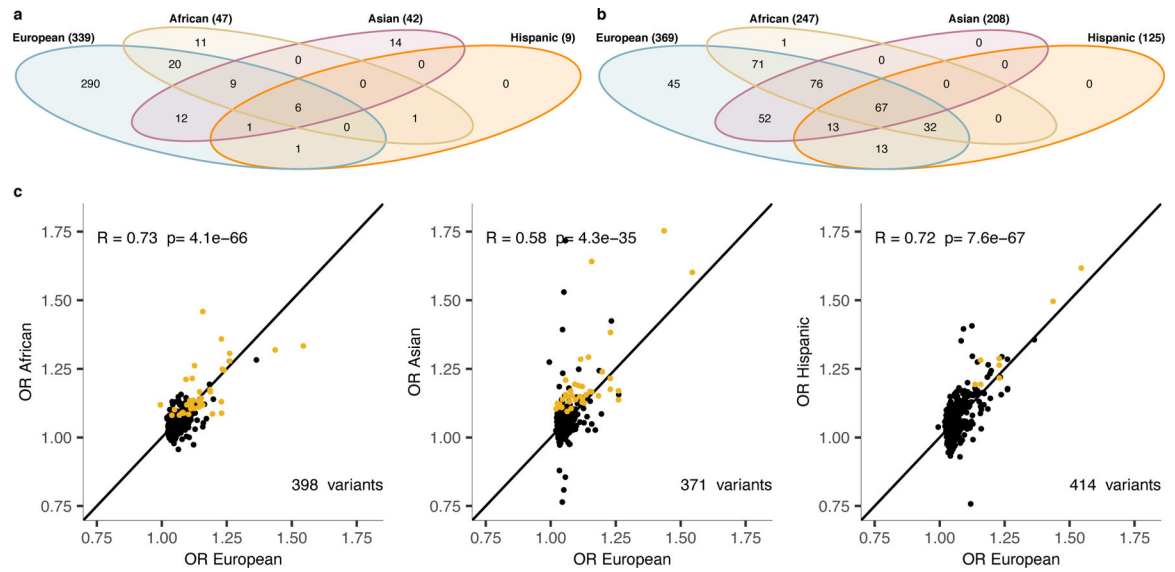


Figure 2.

Comparison of the ancestry-specific results of the 451 risk variants for prostate cancer.

(a) Venn diagram of genome-wide significant variants ($P < 5 \times 10^{-8}$) among European,

African, Asian, and Hispanic populations. (b) Venn diagram of nominally significant variants ($P < 0.05$) among European, African, Asian, and Hispanic populations. (c)

Comparison of ancestry-specific odds ratios (ORs) between European and African, Asian, and Hispanic populations, respectively. The number of variants is denoted in the lower right corner. Genome-wide significant variants among African, Asian, or Hispanic populations are highlighted in orange. Two-sided Pearson correlation tests were performed. The Pearson's correlation coefficient between effect size and corresponding P -value are denoted in the upper left in each sub-panel. Only common variants across all populations ($MAF > 1\%$, $n = 370$) were included in (a), (b), and (c).

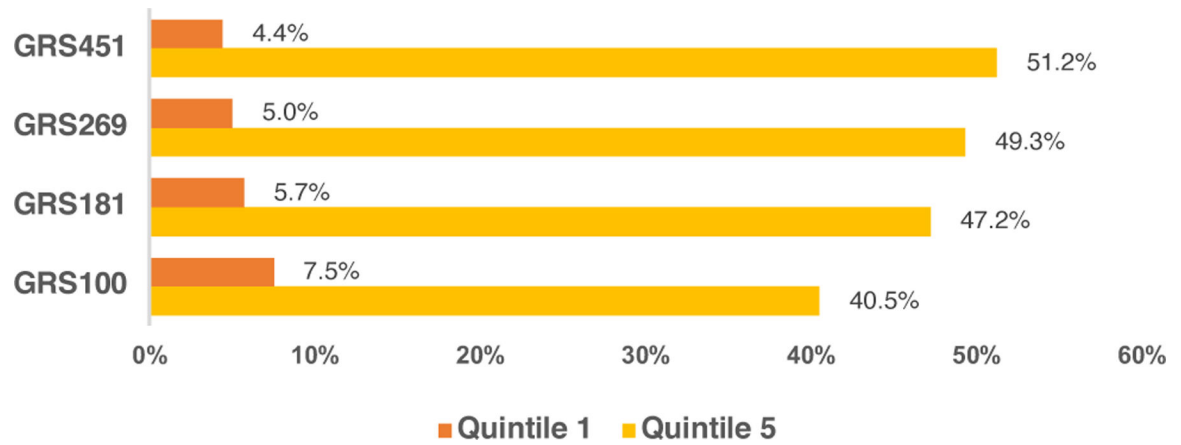


Figure 3.

Percentage of cases in the lowest and highest genetic risk score (GRS) quintiles based on GRS_{100} , GRS_{181} , GRS_{269} , and GRS_{451} in the multi-ancestry sample.

GRS risk quintiles were categorized based on GRS distributions among controls. Quintile 1 (orange bar) refers to the lowest quintile (0–20%), and quintile 5 (yellow bar) refers to the highest quintile (80–100%).

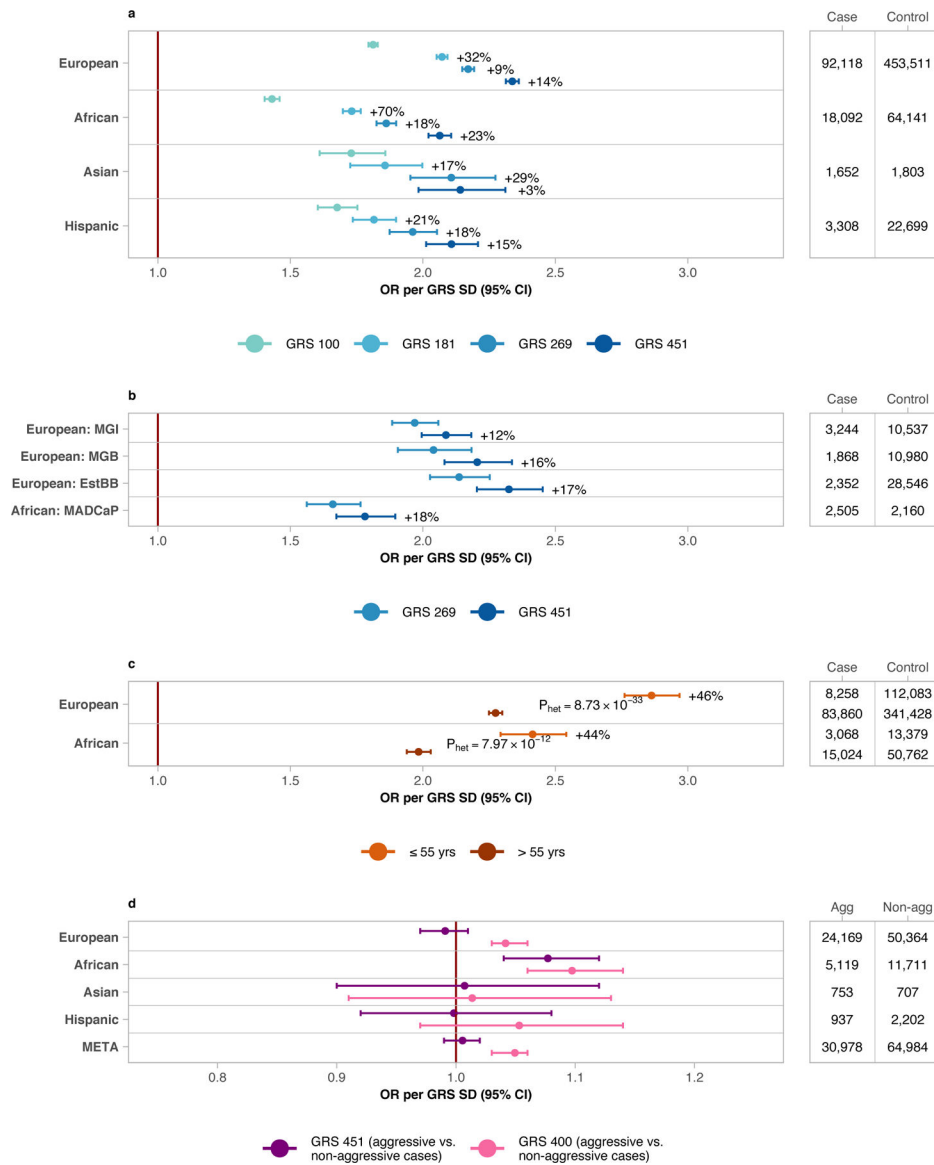


Figure 4.

The associations of GRS and prostate cancer risk in GWAS discovery and replication samples.

ORs and 95% Confidence Intervals (CIs) from logistic regression for one standard deviation (SD) increase in (a) GRS₁₀₀, GRS₁₈₁, GRS₂₆₉, and GRS₄₅₁ and total prostate cancer risk by ancestry in the GWAS discovery studies; (b) GRS₂₆₉ and GRS₄₅₁ and total prostate cancer risk in the replication studies: Michigan Genomics Initiative (MGI), Mass General Brigham Biobank (MGB), Estonian Biobank (EstBB), and Men of African Descent and Carcinoma of the Prostate (MADCaP); (c) GRS₄₅₁ and total prostate cancer risk by age; (d) GRS₄₅₁ and GRS₄₀₀ and prostate cancer aggressiveness among prostate cancer cases in the GWAS discovery studies. ‘META’ refers to the meta-analyzed results for all populations using the inverse-variance weighted method. Incremental percentage change of ORs were calculated for each comparison. The columns ‘case’ and ‘control’ show the case and control sample

sizes, and the columns 'agg' and 'non-agg' show the aggressive and non-aggressive cases sample sizes, respectively.

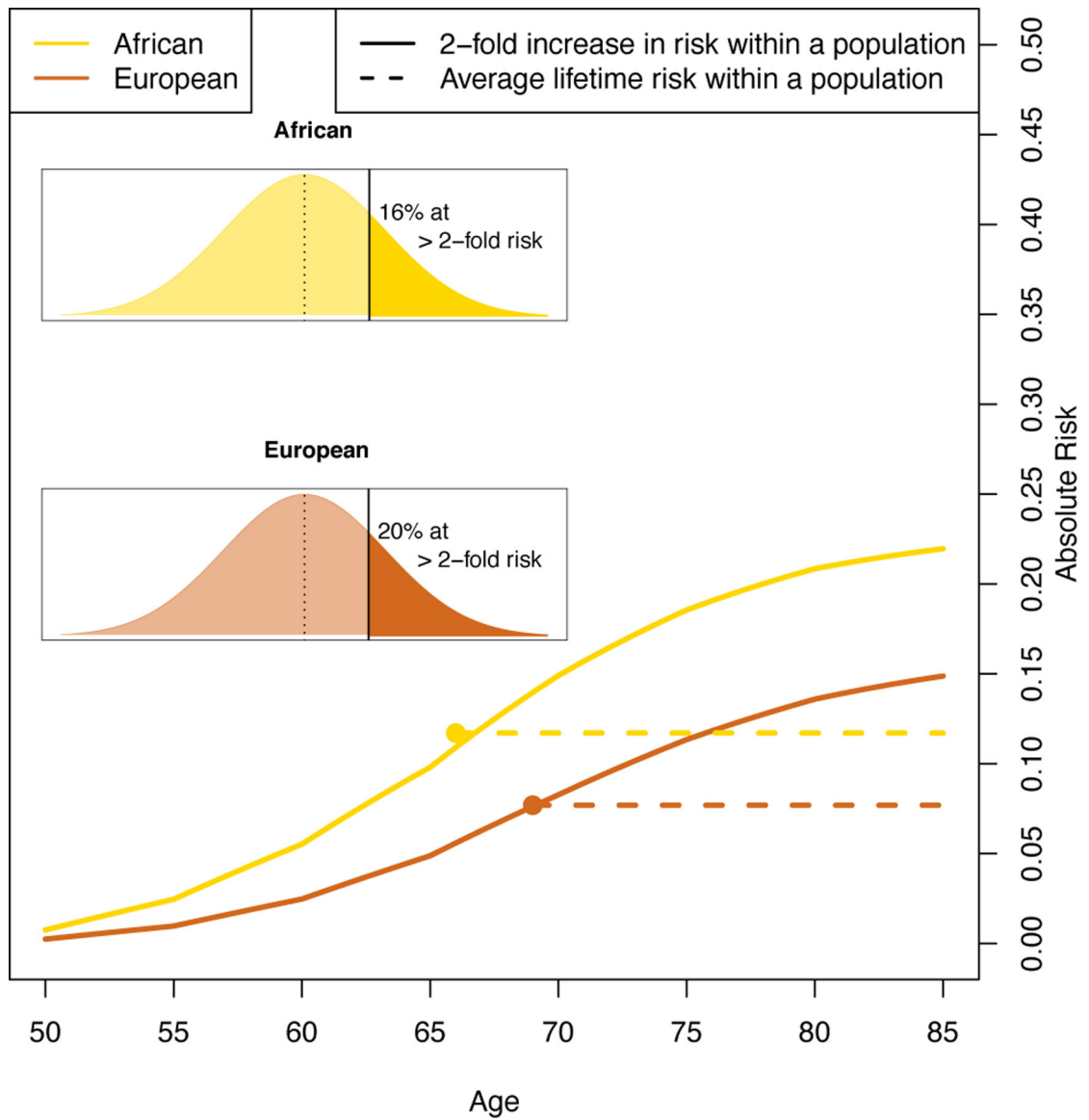


Figure 5.

Cumulative absolute risk by age.

Solid lines are the cumulative absolute risk for individuals in the top 16% GRS for African ancestry and top 20% for European ancestry. These GRS categories represent the percent of individuals in each population with at least a 2-fold increase in risk in comparison to the median GRS (as indicated in the inset distributions for African and European ancestries, respectively). Dashed horizontal lines indicate the lifetime absolute risk achieved at age 85 for the average (50% GRS) in African (11.6%) and European (7.8%) ancestry populations. Solid dots indicate the ages at which lifetime absolute risk levels are achieved for men of African ancestry in the top 16% GRS (age = 66 years) and men of European ancestry in the top 20% GRS (age = 69 years).