Evidence of Novel Susceptibility Variants for Prostate Cancer and a Multi-ancestry Polygenic Risk Score Associated with Aggressive Disease in Men of African Ancestry

Supplementary Materials

MATERIALS AND METHODS

Study participants in the GWAS and meta-analysis

The GWAS meta-analysis includes 19,378 prostate cancer cases and 61,620 controls of African ancestry from the AAPC Consortium (AAPC; 4,822 cases and 4,642 controls), ELLIPSE/PRACTICAL Onco-Array Consortium (ELLIPSE; 4,230 cases and 3,953 controls), Ghana Prostate Study (Ghana; 640 cases and 634 controls), ProHealth Kaiser GWAS (Kaiser; 601 cases and 1,650 controls), Electronic Medical Records and Genomics (eMERGE) Network (233 cases and 1,258 controls), BioVU Biobank (302 cases and 799 controls), BioMe Biobank (154 cases and 2,498 controls), California and Uganda Prostate Cancer Study (CA UG; 1,590 cases and 1,048 controls), VA Million Veteran Program (MVP; 6,353 cases and 44,637 controls), and the Maryland Prostate Cancer Case-Control Study (NCI-MD; 383 cases and 395 controls). Of all studies that contributed samples and/or summary statistics to this analysis, 9,011 cases and 50,634 controls from CA UG, eMERGE, BioVU, BioMe, NCI-MD, and MVP were not part of any previous prostate cancer GWAS. Prostate cancer cases in these studies were identified from cancer registries and/or health records, while controls were individuals without any indication of prostate cancer at the time of DNA sample collection or last known visit. An overview of each study is provided in **Table S1**. Informed consent was obtained from all participants, and study protocols were approved by respective Institutional Review Boards.

Genotyping and imputation in the GWAS

Genotyping arrays, quality control measures on samples and variants, imputation panels, and the statistical software used for each study or consortium are summarized in **Table S2**. Details for each study or consortium have been described elsewhere (see references in **Table S1**). In general, before imputation, samples and genotyped variants were excluded at the corresponding study-specific sample or genotyping call rate of 95%. Except for the Ghana and ELLIPSE studies that included all variants, most studies filtered variants with minor allele frequency (MAF) < 1%. Several studies excluded variants that were out of Hardy-Weinberg equilibrium at varying significance thresholds. Genotyping data were imputed in each study or consortium to either the 1000 Genomes phase 3 [1], the Haplotype Reference Consortium (HRC)[2], or the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium freeze 5[3] imputation panels using Minimac3/Minimac4 [4] or IMPUTE2 [5]. In most studies, genotyped and imputed variants with an imputation info score (R^2) ≥ 0.30 and MAF ≥ 1% were included in the association analyses. A lower MAF cutoff of 0.1% was applied in MVP while an info score ≥ 0.80 was used in BioMe. In most studies or consortia with samples from non-African countries, principal component analysis (PCA) was performed including reference populations from the 1000 Genomes Project to infer genetic ancestry based on the top principal components (PCs) [6]. Genetic ancestry for all MVP participants was assessed using HARE[7] based on the top 30 PCs and the African-ancestry GWAS was conducted in samples assigned to be of African ancestry. In both AAPC and ELLIPSE, individuals with estimated African global ancestry < 0.1 from a STRUCTURE[8] analysis were excluded from the African-ancestry GWAS.

Statistical analysis for GWAS and meta-analysis

Logistic regression analysis was carried out in each study or consortium to examine the association of variants with prostate cancer risk, adjusting for age, sub-study (if applicable), and up to ten PCs using PLINK [9], SNPtest [10], SAIGE [11] or R[12]. Per-allele odds ratios (ORs) and standard errors from individual studies were combined using a fixed-effects inverse-variance-weighted meta-analysis with METAL[13]. All statistical tests were two-sided. Variants were considered genome-wide significant if their marginal P-value from the meta-analysis was less than 5.0×10^{-8}.

We estimated the inflation factor (λ) from the P-values of all analyzed variants (N = 27,753,840). The inflation factor was then converted to an equivalent inflation factor for a study with 1,000 cases and 1,000 controls (λ_{1000}) following $\lambda_{1000} = 1 + \frac{500\,(\lambda-1)}{\sum_k(\frac{1}{n_k}+\frac{1}{m_k})^{-1}}$, where the n_k and m_k were the number of cases and number of controls for study k, respectively [14].

To determine the independency of genome-wide significant associations, conditional analysis was performed on variants in known risk regions. Among the 269 previously identified prostate cancer susceptibility loci [15], 178 known risk regions were defined and 51 contained more than one known risk variant. For genome-wide significant variants in known risk regions, we used Joint Analysis of Marginal summary statistics (JAM) to obtain conditional effects and P-values from multivariate models, conditioning on all known risk variants in the same region [15]. Associations with a conditional $P < 5.0\times10^{-8}$ were considered novel. Similar conditional analyses were carried out to determine if multiple independent associations exist in the ± 800 kb of the identified novel variant.

To identify credible set variants, we applied a modified JAM approach examining all variants within ± 800 kb of each identified novel variant (index variant). Based on the summary statistics from this African-ancestry meta-analysis, for each variant within the region, a posterior inclusion probability (PIP) was calculated considering both its marginal effect and its mediation effect after adjusting for the index variant. These scaled PIPs were then used to construct a 95% credible set by selecting the smallest set of variants accounting for a 95% PIP.

Study participants in the subsequent association analyses

Six GWAS studies (AAPC, ELLIPSE, Ghana, CA UG, NCI-MD, and MVP) in the meta-analysis and an independent sample from the Men of African Descent and Carcinoma of the Prostate (MADCaP) Network [16,17] were included in the subsequent association analyses, where we evaluated (1) the association of novel and known risk variants with disease aggressiveness, and (2) the association of polygenic risk scores (PRS) with prostate cancer risk, overall and stratified by age, geographic area, and disease aggressiveness (**Table S3**).

The MVP dataset in these association analyses included additional 8,725 controls with prostate-specific antigen (PSA) level ≥ 4 ng/mL that were excluded in the MVP GWAS. The MADCaP dataset included 405 prostate cancer cases and 396 controls from sub-Saharan Africa with a substantial proportion of cases diagnosed at late stages[16]. The MADCaP samples were genotyped on a customized array designed to capture genetic variation in diverse African populations [17] and were imputed using the 1000 Genomes Project Phase 3 reference panel (14).

In all seven studies, prostate cancer was considered aggressive if one or more of the following criteria was met: tumor stage T3/T4, regional lymph node involvement, metastatic disease (M1), Gleason score ≥ 8.0, PSA level ≥ 20 ng/mL or prostate cancer as the underlying cause of death. Non-aggressive prostate cancer was defined as men with no aggressive features meeting one or more of the following criteria: Gleason score ≤ 7.0, PSA < 20 ng/mL, and stage ≤ T2. After excluding subjects with missing information on these clinical features, a total of 5,469 aggressive cases and 11,710 non-aggressive cases were identified across these studies, including 6,399 cases with a Gleason score of 6; 6,758 with a Gleason score of 7; 3,099 with Gleason score of 8 or above; 10,497 cases with T1 or T2 tumor; 1,201 cases with T3 or T4 tumor; 769 cases with metastatic disease; and 304 fatal cases (prostate cancer as the cause of death; **Table S3**).

Association of risk variants with disease aggressiveness

For the novel and known risk variants with MAF > 1% in African populations (N = 255), we examined their prostate cancer risk-increasing alleles for association with disease aggressiveness in logistic regression analysis of aggressive cases versus controls, non-aggressive cases versus controls, and aggressive cases versus non-aggressive cases, adjusting for age, sub-study (if applicable) and up to ten principal components. Additional case-case analyses by tumor stage, Gleason score, metastatic or fatal disease were also performed (Gleason score of 8 versus Gleason score of 6, tumor stage T3/T4 versus tumor Stage T1/T2, metastatic cases versus non-aggressive cases, and fatal cases versus non-aggressive cases). In each analysis, results were meta-analyzed across individual studies using a fixed-effects inverse-variance-weighted method.

PRS construction and association with prostate cancer risk

The PRS was constructed by summing variant-specific weighted allelic dosages from 269 known prostate cancer risk variants and the nine novel variants. For the primary PRS association analysis, the multi-ancestry weights for all of the 278 variants were obtained from the previous trans-ancestry GWAS and used to calculate the PRS (**Table S7**)[15]. To assess the robustness of PRS associations, we also constructed a PRS using the African ancestry-specific effects estimated from men of African ancestry (10,367 cases and 10,986 controls; **Table S7**) and compared its association with prostate cancer risk to the multi-ancestry PRS.

The association of PRS with prostate cancer risk was estimated using an indicator variable for the percentile categories of the PRS distribution: [0%-10%], (10%-20%], (20%-30%], (30%-40%], (40%-60%], (60%-70%], (70%-80%], (80%-90%], and (90%-100%]. An additional analysis was performed by splitting the top decile into two categories to obtain the PRS risk for the top 1%: (90%-99%], (99%-100%]. PRS thresholds were determined in the observed distribution among controls. Logistic regression was performed to estimate OR and 95% confidence intervals (CI) corresponding to each PRS category, adjusting for age, sub-study (if applicable), and up to ten principal components, with the (40%-60%] category as the reference. We also reported the OR and 95% CI for prostate cancer per one standard deviation (SD) increase in PRS, where the mean and SD were determined in the observed distribution among controls.

The association of PRS with prostate cancer risk was meta-analyzed across the six studies included in the GWAS ("Discovery Sample" in **Table 2**) and was evaluated for replication in the independent sample from MADCaP ("Replication Sample" in **Table 2**). In the comparison of multi-ancestry PRS and African-specific PRS, heterogeneity was assessed via a Q statistic between effect estimates with corresponding tests of significance using the R package *meta* [19]. Discriminative ability was evaluated in MVP by estimating the area under the curve (AUC) for logistic regression models of prostate cancer that included covariates only (age and four principal components of ancestry) and for models that additionally included the PRS.

PRS associations with prostate cancer risk stratified by age and geographic area

We investigated the association of the multi-ancestry PRS with prostate cancer risk stratified by age and geographic area. In the age-stratified analysis, cases and controls were both stratified into age groups (age ≤ 55 years vs. age > 55 years). In the geographic area stratified analysis, cases and controls were stratified based on being recruited from African countries (Uganda, Ghana, and the Democratic Republic of the Congo) and non-African countries (US, UK, Canada, and France). In each stratum, a logistic regression analysis was performed with prostate cancer status as the outcome and the PRS (categories or per one SD) as the independent predictors, adjusting for age, sub-study (if applicable), and up to ten principal components. Results within each stratum were meta-analyzed across studies, and the heterogeneity of associations between strata was assessed via a Q statistic with corresponding tests of significance[19].

PRS associations with disease aggressiveness

We assessed the association of the multi-ancestry PRS with disease aggressiveness in logistic regression analyses of aggressive cases versus controls and non-aggressive cases versus controls, with the specific prostate cancer phenotype as the outcome and the PRS (categories or per one SD) as the independent predictor, adjusting for age, sub-study (if applicable) and up to ten principal components. P for heterogeneity was determined using a Q statistic comparing the associations between the aggressive cases (vs. controls) and non-aggressive cases (vs. controls). Additional subgroup case-control analyses using similar models were performed for high-grade (Gleason score ≥ 8), low-grade (Gleason score of 6), advanced (tumor stage of T3 or T4), localized (tumor stage of T1 or T2), metastatic (tumor stage of M1), and fatal prostate cancer, comparing the subgroup cases versus controls. Results from each analysis were meta-analyzed using a fixed-effects inverse-variance-weighted method across studies with sufficient samples of the defined phenotypes (number of cases and controls > 100; **Table S3**). P for heterogeneity was estimated comparing the PRS association with high-grade prostate cancer to the association with low-grade prostate cancer, advanced prostate cancer to localized prostate cancer, metastatic prostate cancer to non-aggressive prostate cancer, and fatal prostate cancer to non-aggressive prostate cancer. A significantly (P-heterogeneity < 0.05) greater association with the aggressive phenotype than with the corresponding non-aggressive phenotype would support a PRS association with the aggressive phenotype.

The PRS association with disease aggressiveness (categories or per one SD) was also evaluated in case-case analyses comparing aggressive cases versus non-aggressive cases, advanced cases versus localized cases, high-grade cases versus low-grade cases, metastatic cases versus non-aggressive cases, and fatal cases versus non-aggressive cases. A statistically significant ($P < 0.05$) positive association ($OR > 1$) from the case-case analysis would support a PRS association with the aggressive phenotype.

For prostate cancer risk variants that were nominally and positively associated with disease aggressive ($P < 0.05$; $N = 14$), we assessed their contribution to the PRS association with disease aggressiveness in the analysis of aggressive cases vs. controls, non-aggressive cases vs. controls, and aggressive vs. non-aggressive cases, by separately removing each variant from the PRS. Associations with a specific prostate cancer phenotype were compared between the PRSs with and without the variant. A Q statistic was used to determine the significance of the observed differences in ORs[19].

Proportion of familial risk explained

The contribution of the 278 variants to the familial relative risk (FRR) of prostate cancer was estimated using the formula: $\frac{\sum_k(\log\lambda_k)}{(\log\lambda_0)}$, where λ_0 is the observed familial risk to first-degree relatives of prostate cancer cases, assumed to range from 2.0 to 3.0 [20–22], and λ_k is the FRR due to locus k, given by $\lambda = \frac{p_k r_k^2 + q_k}{(p_k r_k + q_k)^2}$, where $p_k$is the frequency of the risk allele for locus k, $q_k = 1 - p_k$ and r_k is the estimated per-allele OR [23,24]. The proportion of FRR explained was estimated using the per-allele OR from the previous trans-ancestry meta-analysis[15].

Estimation of global ancestry

The proportion of African ancestry (AFR%) for each individual in six prostate cancer studies (AAPC, ELLIPSE, CA UG, Ghana, NCI-MD, and MADCaP) was estimated using 20,047 genotyped common variants (MAF > 1%) from an unsupervised (K=2) ADMIXTURE analysis [25] which included the 1000 Genomes Project phase 3 European and African subjects [1]. We evaluated the relationship between allele frequency and AFR% for the nine novel risk variants and calculated the mean AFR% to indicate the level of admixture in each of the six prostate cancer studies.

Functional annotations

The nine novel variants and their 95% credible sets were annotated for putative evidence of biological functionality using publicly available datasets according to the framework described previously[15]. Briefly, variants were annotated for genomic context and proximity to genes using wANNOVAR. Annotation of variants against intersection with chromatin marks indicative of regulatory DNA regions was performed relative to peak data from publicly available datasets generated in the prostate-derived cell lines LNCaP, PC3, PrEC, and VCaP. Transcription factor binding site chromatin immunoprecipitation sequencing (ChIP-seq) peak data was obtained from the Androgen Receptor (GSM1274871, GSM1576447, and GSM1527834), CTCF (GSM1006874 and GSM2825574), ERG (GSM1193657 and GSM1328978), FOXA1 (GSM1274873, GSM1691142, and GSM2219863), GABPA

(GSM1193660), GATA2 (GSM941195 and GSM1600544), HOXB13 (GSM1716764 and GSM2537218), NKX3.1 (GSM989640) and POLR2A (GSM353623, GSM969566, GSM1059393, and GSM1059394).

REFERENCES

[1] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68–74. https://doi.org/10.1038/nature15393.

[2] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279–83. https://doi.org/10.1038/ng.3643.

[3] Kowalski MH, Qian H, Hou Z, Rosen JD, Tapia AL, Shan Y, et al. Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. PLOS Genet 2019;15:e1008500. https://doi.org/10.1371/journal.pgen.1008500.

[4] Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–7. https://doi.org/10.1038/ng.3656.

[5] Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. PLOS Genet 2009;5:e1000529. https://doi.org/10.1371/journal.pgen.1000529.

[6] Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9. https://doi.org/10.1038/ng1847.

[7] Fang H, Hui Q, Lynch J, Honerlaw J, Assimes TL, Huang J, et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. Am J Hum Genet 2019;105:763–72. https://doi.org/10.1016/j.ajhg.2019.08.012.

[8] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59. https://doi.org/10.1093/genetics/155.2.945.

[9] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet 2007;81:559–75.

[10] Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78. https://doi.org/10.1038/nature05911.

[11] Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335–41. https://doi.org/10.1038/s41588-018-0184-y.

[12] R: A language and environment for statistical computing. R Found Stat Comput Vienna Aust 2017.

[13] Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinforma Oxf Engl 2010;26:2190–1. https://doi.org/10.1093/bioinformatics/btq340.

[14] Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, et al. Assessing the impact of population stratification on genetic association studies. Nat Genet 2004;36:388–93. https://doi.org/10.1038/ng1333.

[15] Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat Genet 2021;53:65–75. https://doi.org/10.1038/s41588-020-00748-0.

[16] Andrews C, Fortier B, Hayward A, Lederman R, Petersen L, McBride J, 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 2018;4:1–14. https://doi.org/10.1200/JGO.18.00063.

[17] Harlemon M, Ajayi O, Kachambwa P, Kim MS, Simonti CN, Quiver MH, et al. A Custom Genotyping Array Reveals Population-Level Heterogeneity for the Genetic Risks of Prostate Cancer and Other Cancers in Africa. Cancer Res 2020;80:2956–66. https://doi.org/10.1158/0008-5472.CAN-19-2165.

[18] Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature 2015;517:327–32. https://doi.org/10.1038/nature13997.

[19] Schwarzer G, Carpenter JR, Rücker G. Fixed Effect and Random Effects Meta-Analysis. In: Schwarzer G, Carpenter JR, Rücker G, editors. Meta-Anal. R, Cham: Springer International Publishing; 2015, p. 21–53. https://doi.org/10.1007/978-3-319-21416-0_2.

[20] Kiciński M, Vangronsveld J, Nawrot TS. An epidemiological reappraisal of the familial aggregation of prostate cancer: a meta-analysis. PloS One 2011;6:e27130. https://doi.org/10.1371/journal.pone.0027130.

[21] Nemesure B, Wu S-Y, Hennis A, Leske MC. Family history of prostate cancer in a black population. J Immigr Minor Health 2013;15:1107–12. https://doi.org/10.1007/s10903-012-9710-7.

[22] Mordukhovich I, Reiter PL, Backes DM, Family L, McCullough LE, O'Brien KM, et al. A review of African American-white differences in risk factors for cancer: prostate cancer. Cancer Causes Control 2011;22:341–57. https://doi.org/10.1007/s10552-010-9712-5.

[23] Wang K, Dickson SP, Stolle CA, Krantz ID, Goldstein DB, Hakonarson H. Interpretation of association signals and identification of causal variants from genome-wide association studies. Am J Hum Genet 2010;86:730–42. https://doi.org/10.1016/j.ajhg.2010.04.003.

[24] Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. Nat Rev Genet 2014;15:765–76. https://doi.org/10.1038/nrg3786.

[25] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–64. https://doi.org/10.1101/gr.094052.109.

SUPPLEMENTARY FIGURES

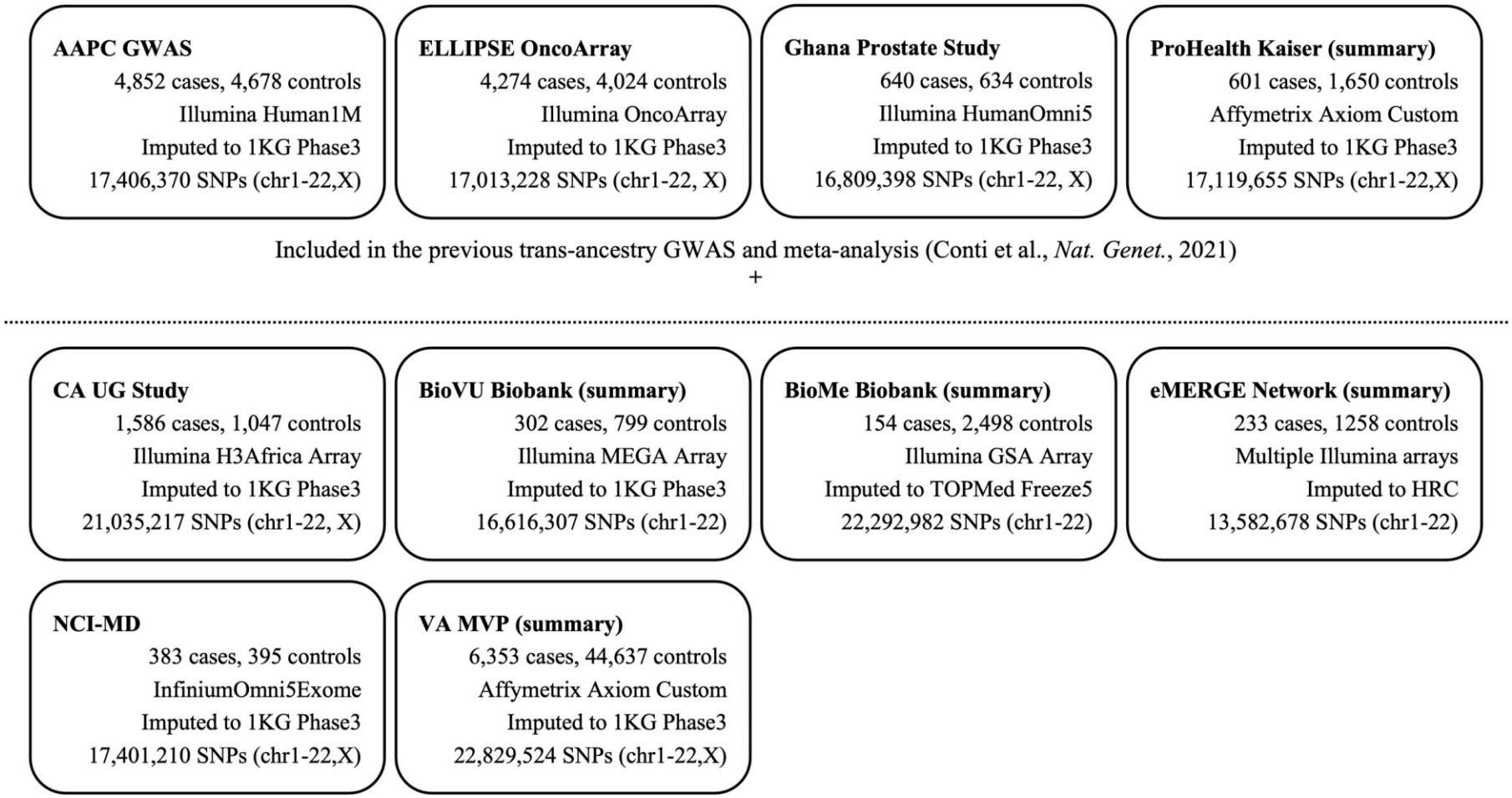


Figure S1 Summary of prostate cancer genome-wide association study (GWAS) included in this meta-analysis. The current meta-analysis in men of African ancestry included 19,378 prostate cancer cases and 61,620 controls from 10 GWAS studies, of which 9,011 cases and 50,634 controls from six GWAS were not part of the previous trans-ancestry GWAS (Conti, Darst et al., *Nature Genetics*, 2021). Studies that provided the summary statistics were indicated in the parenthesis. A total of 27,753,840 variants were tested for association with prostate cancer risk in the meta-analysis.

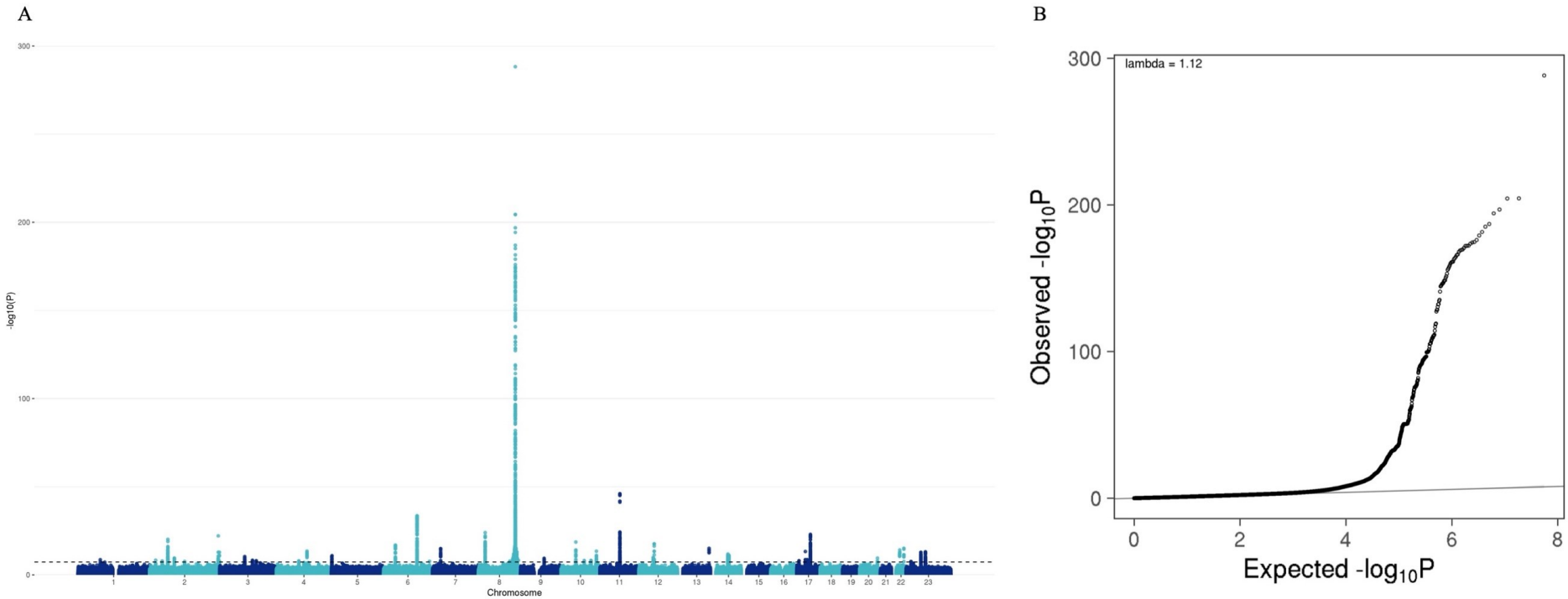
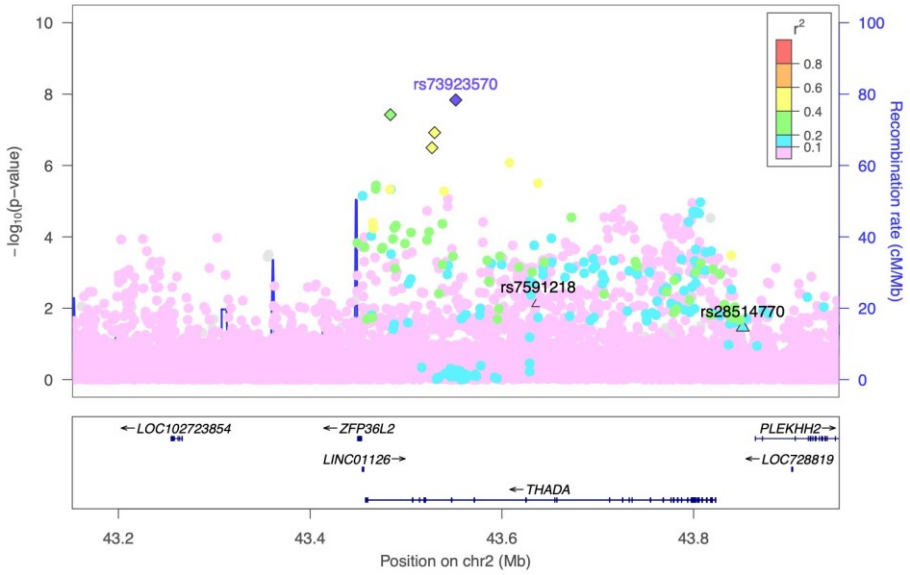
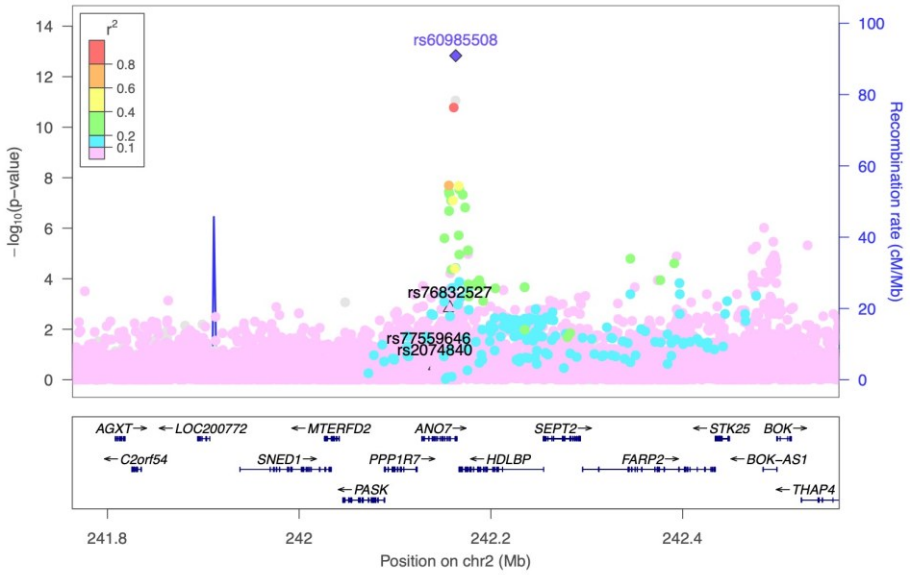
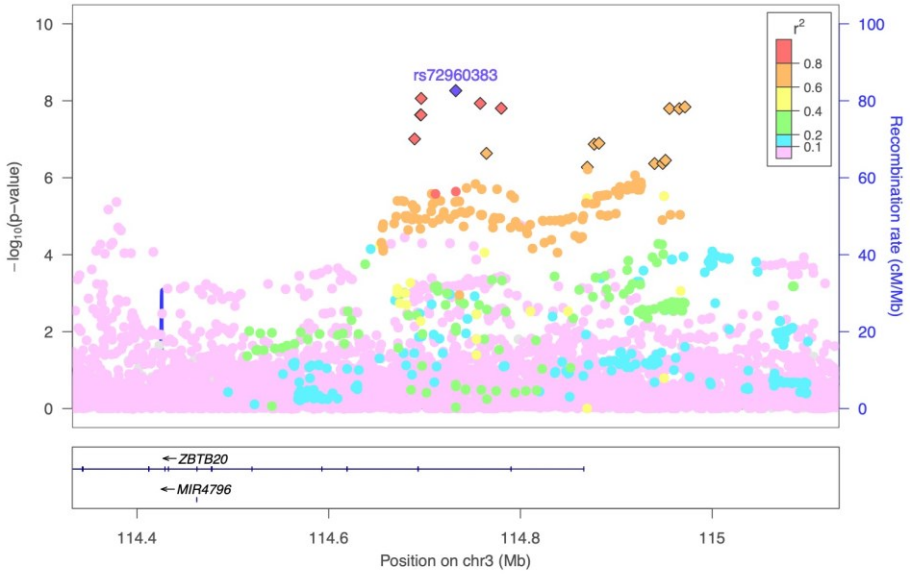
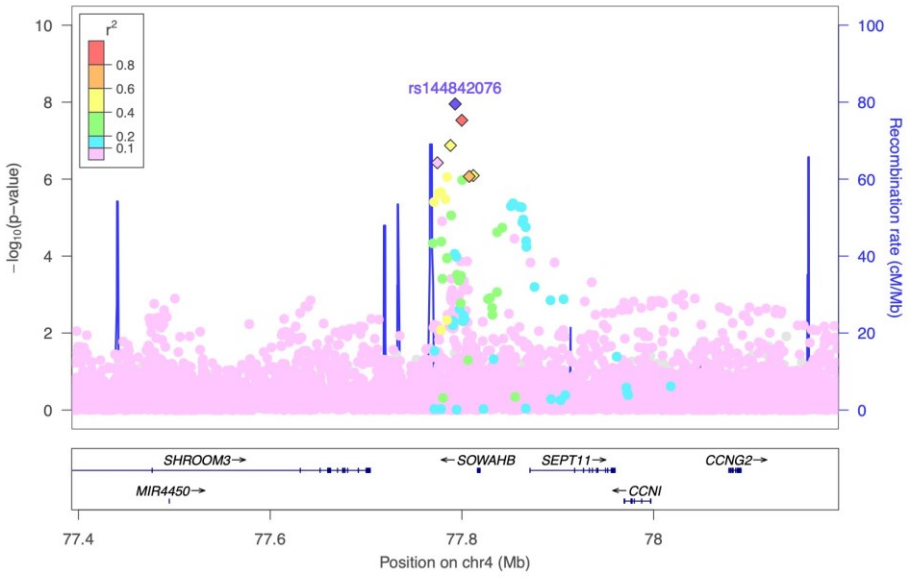


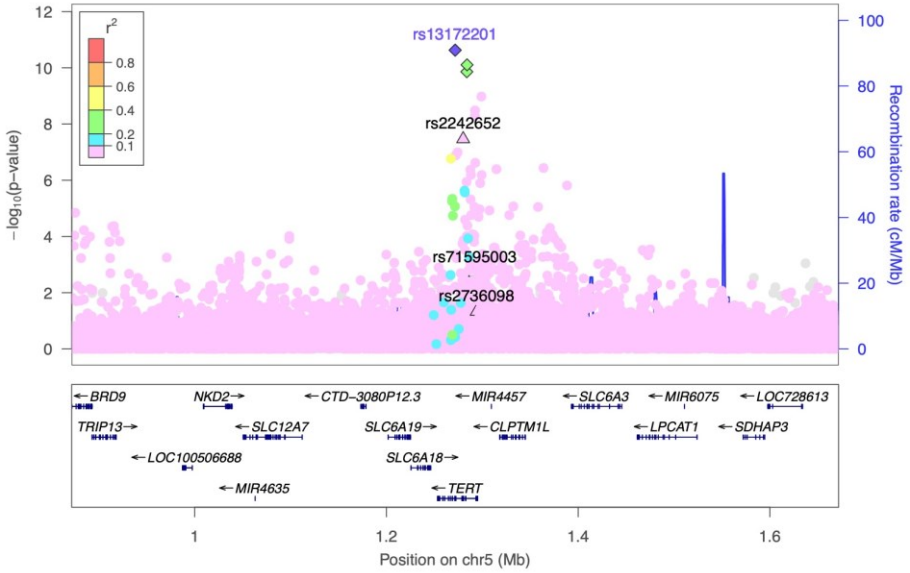
Figure S2 Manhattan plot (A) and quantile-quantile plot (B).

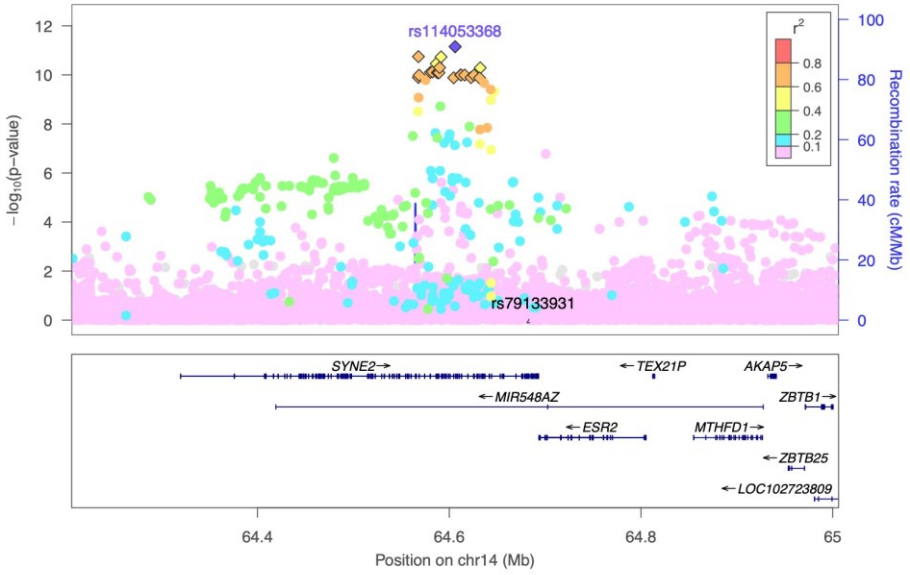

rs73923570
rs7591218
rs28514770
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
LOC102723854
ZFP36L2
PLEKHH2
LINC01126
LOC728819
THADA
43.2
43.4
43.6
43.8
Position on chr2 (Mb)

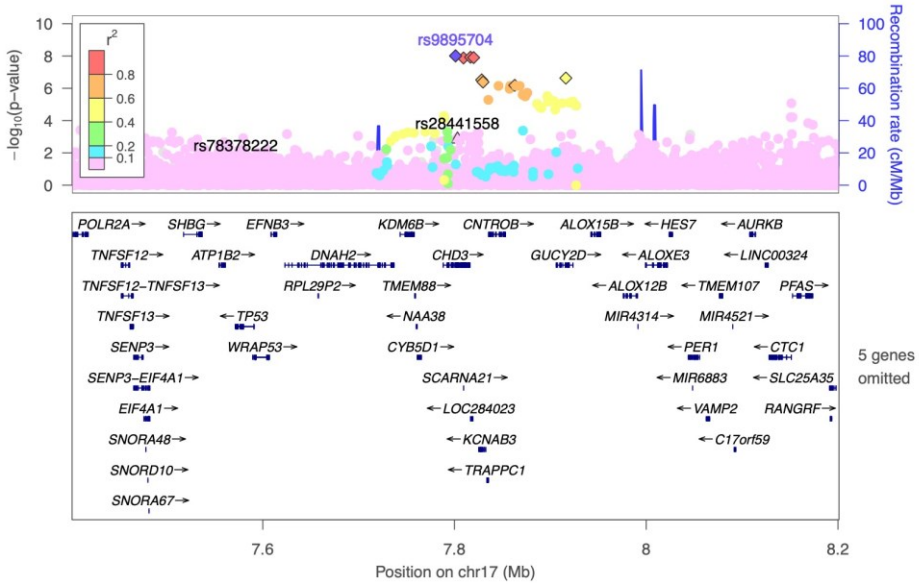

rs60985508
rs76832527
rs77559646
rs2074840
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
AGXT
LOC200772
MTERFD2
ANO7
SEPT2
STK25
BOK
C2orf54
SNED1
PPP1R7
HDLBP
FARP2
BOK−AS1
PASK
THAP4
241.8
242
242.2
242.4
Position on chr2 (Mb)

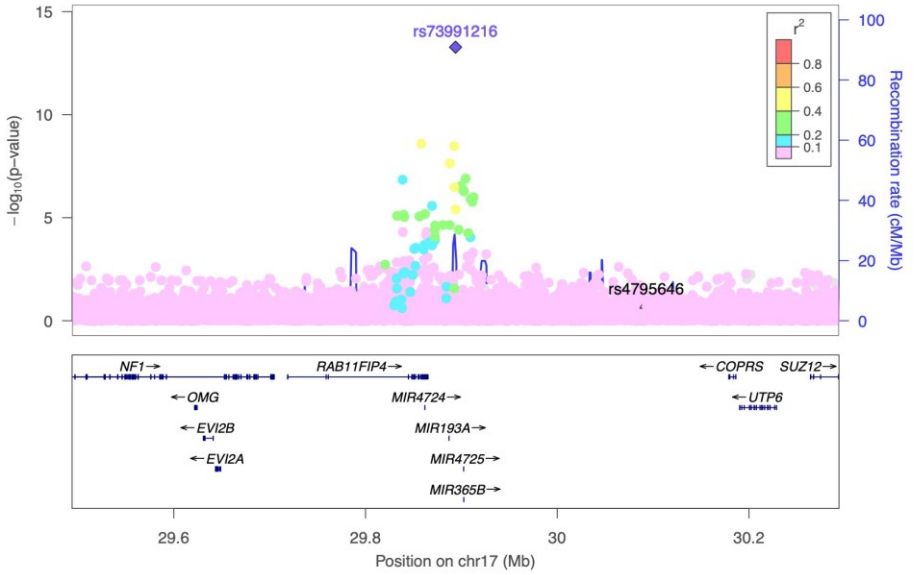

rs72960383
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
ZBTB20
MIR4796
114.4
114.6
114.8
115
Position on chr3 (Mb)


rs144842076
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
SHROOM3
SOWAHB
SEPT11
CCNG2
MIR4450
CCNI
77.4
77.6
77.8
78
Position on chr4 (Mb)


rs13172201
rs2242652
rs71595003
rs2736098
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
BRD9
NKD2
CTD−3080P12.3
MIR4457
SLC6A3
MIR6075
LOC728613
TRIP13
SLC12A7
SLC6A19
CLPTM1L
LPCAT1
SDHAP3
LOC100506688
SLC6A18
MIR4635
TERT
1
1.2
1.4
1.6
Position on chr5 (Mb)


rs114053368
rs79133931
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
SYNE2
TEX21P
AKAP5
MIR548AZ
ZBTB1
ESR2
MTHFD1
ZBTB25
LOC102723809
64.4
64.6
64.8
65
Position on chr14 (Mb)


rs9895704
rs28441558
rs78378222
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
POLR2A→ SHBG→ EFNB3→ KDM6B→ CNTROB→ ALOX15B→ ←HES7 ←AURKB
TNFSF12→ ATP1B2→ DNAH2→ CHD3→ GUCY2D→ ←ALOXE3 ←LINC00324
TNFSF12−TNFSF13→ RPL29P2→ TMEM88→ ←ALOX12B ←TMEM107 PFAS→
TNFSF13→ ←TP53 ←NAA38 MIR4314→ MIR4521→
SENP3→ WRAP53→ CYB5D1→ ←PER1 ←CTC1
SENP3−EIF4A1→ SCARNA21→ ←MIR6883 ←SLC25A35
EIF4A1→ ←LOC284023 ←VAMP2 RANGRF→
SNORA48→ ←KCNAB3 ←C17orf59
SNORD10→ ←TRAPPC1
SNORA67→
5 genes omitted
7.6
7.8
8
8.2
Position on chr17 (Mb)


rs73991216
rs4795646
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
NF1→ RAB11FIP4→ ←COPRS SUZ12→
←OMG MIR4724→ ←UTP6
←EVI2B MIR193A→
←EVI2A MIR4725→
MIR365B→
29.6
29.8
30
30.2
Position on chr17 (Mb)

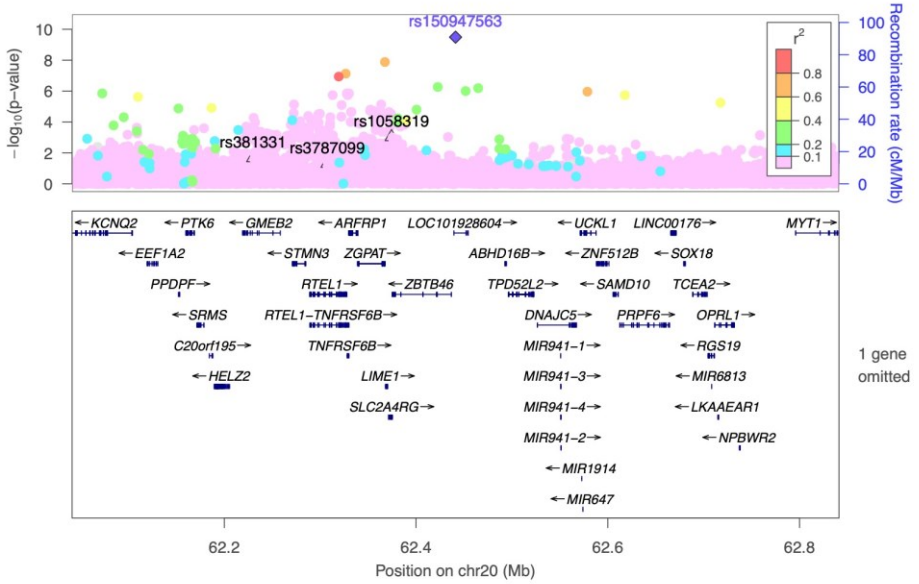
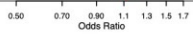

rs150947563
rs1058319
rs381331
rs3787099
r²
0.8
0.6
0.4
0.2
0.1
−log10(p−value)
Recombination rate (cM/Mb)
←KCNQ2 ←PTK6 ←GMEB2 ←ARFRP1 LOC101928604→ ←UCKL1 LINC00176→ MYT1→
←EEF1A2 ←STMN3 ZGPAT→ ABHD16B→ ←ZNF512B ←SOX18
PPDPF→ RTEL1→ ←ZBTB46 TPD52L2→ ←SAMD10 TCEA2→
←SRMS RTEL1−TNFRSF6B→ DNAJC5→ PRPF6→ OPRL1→
C20orf195→ TNFRSF6B→ MIR941−1→ ←RGS19
←HELZ2 LIME1→ MIR941−3→ ←MIR6813
SLC2A4RG→ MIR941−4→ ←LKAAEAR1
MIR941−2→ ←NPBWR2
←MIR1914
←MIR647
1 gene omitted
62.2
62.4
62.6
62.8
Position on chr20 (Mb)

Figure S3 LocusZoom plots of the nine novel risk variants. In each plot, the lead variant was represented by a purple diamond and the known risk variant(s) in the same region were represented by triangles. The color of each variant was based on their linkage disequilibrium (r^2) with the lead variant in the 1000 Genomes Project African populations. Variants in diamond shapes represent the 95% credible set identified in each risk region.

A **rs73923570**

Study	N	RAF(G)	Rsq	OR(95%CI)	P
NCI–MD	778	0.12	0.99	0.92 (0.68–1.26)	0.60
BioVU	1101	0.13	0.96	1.31 (0.98–1.75)	0.07
Ghana	1274	0.15	0.99	1.08 (0.84–1.39)	0.56
eMERGE	1491	0.11	0.98	0.80 (0.58–1.11)	0.19
ProHealth Kaiser	2251	0.10	0.98	1.15 (0.93–1.43)	0.21
CA UG	2633	0.10	0.98	1.39 (1.17–1.65)	2.40e–04
BioMe	2652	0.12	0.96	1.03 (0.71–1.48)	0.89
ELLIPSE	8298	0.12	0.98	1.05 (0.95–1.16)	0.33
AAPC	9530	0.12	0.99	1.14 (1.04–1.24)	4.30e–03
MVP	50990	0.11	0.89	1.13 (1.07–1.20)	3.75e–05
Meta–analysis	**80998**	**0.11**	**I^2 = 39**	**1.12 (1.08–1.17)**	**1.46e–08**



B **rs60985508**

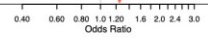
Study	N	RAF(T)	Rsq	OR(95%CI)	P
NCI–MD	778	0.32	0.97	1.07 (0.85–1.34)	0.59
BioVU	1101	0.30	0.93	1.32 (1.06–1.64)	0.01
Ghana	1274	0.40	0.98	0.96 (0.79–1.17)	0.72
eMERGE	1491	--	--	--	--
ProHealth Kaiser	2251	0.28	0.96	1.19 (1.02–1.39)	0.02
CA UG	2633	0.31	0.95	1.12 (0.99–1.26)	0.08
BioMe	2652	0.30	0.96	1.25 (0.96–1.62)	0.10
ELLIPSE	8298	0.31	0.92	1.00 (0.93–1.07)	0.97
AAPC	9530	0.30	0.97	1.11 (1.04–1.19)	1.02e–03
MVP	50990	0.31	0.95	1.15 (1.10–1.19)	5.34e–11
Meta–analysis	**80998**	**0.31**	**I^2 = 52**	**1.11 (1.08–1.15)**	**1.48e–13**

C **rs72960383**

Study	N	RAF(A)	Rsq	OR(95%CI)	P
NCI–MD	778	0.33	1.00	1.05 (0.84–1.32)	0.66
BioVU	1101	0.33	0.98	1.03 (0.84–1.26)	0.79
Ghana	1274	0.39	1.00	1.19 (0.98–1.43)	0.08
eMERGE	1491	0.33	0.97	1.10 (0.88–1.37)	0.42
ProHealth Kaiser	2251	0.31	0.96	0.99 (0.86–1.15)	0.92
CA UG	2633	0.32	1.00	1.18 (1.05–1.32)	7.18e–03
BioMe	2652	0.33	0.98	0.99 (0.77–1.28)	0.96
ELLIPSE	8298	0.34	0.94	1.07 (0.99–1.14)	0.07
AAPC	9530	0.33	1.00	1.06 (0.99–1.13)	0.08
MVP	50990	0.33	0.90	1.11 (1.06–1.15)	1.63e–06
Meta–analysis	**80998**	**0.33**	**I^2 = 0**	**1.09 (1.06–1.12)**	**5.46e–09**

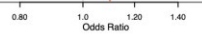
D **rs144842076**

Study	N	RAF(C)	Rsq	OR(95%CI)	P
NCI–MD	778	0.98	0.99	0.91 (0.44–1.85)	0.79
BioVU	1101	0.96	0.96	1.34 (0.80–2.24)	0.27
Ghana	1274	0.97	0.97	0.88 (0.49–1.58)	0.67
eMERGE	1491	0.96	0.96	1.71 (0.99–2.95)	0.06
ProHealth Kaiser	2251	0.96	0.99	1.37 (0.93–2.02)	0.11
CA UG	2633	0.96	0.99	1.26 (0.92–1.71)	0.15
BioMe	2652	0.97	0.95	1.34 (0.67–2.69)	0.40
ELLIPSE	8298	0.96	0.86	1.23 (1.03–1.47)	0.02
AAPC	9530	0.97	0.99	1.10 (0.94–1.30)	0.25
MVP	50990	0.97	0.95	1.34 (1.19–1.51)	1.69e–06
Meta–analysis	**80998**	**0.96**	**I^2 = 0**	**1.25 (1.16–1.35)**	**1.12e–08**



E **rs13172201**

Study	N	RAF(C)	Rsq	OR(95%CI)	P
NCI–MD	778	0.39	0.88	1.18 (0.94–1.47)	0.15
BioVU	1101	0.42	0.84	1.17 (0.95–1.44)	0.14
Ghana	1274	0.45	0.90	1.00 (0.82–1.21)	1.00
eMERGE	1491	0.36	0.73	1.12 (0.88–1.43)	0.36
ProHealth Kaiser	2251	0.41	0.84	1.01 (0.87–1.17)	0.86
CA UG	2633	0.40	0.88	1.11 (0.99–1.26)	0.08
BioMe	2652				
ELLIPSE	8298	0.40	0.96	1.15 (1.08–1.23)	3.18e–05
AAPC	9530	0.41	0.89	1.05 (0.98–1.12)	0.14
MVP	50990	0.40	0.75	1.12 (1.07–1.16)	8.62e–07
Meta–analysis	**80998**	**0.40**	**I^2 = 0**	**1.10 (1.07–1.13)**	**2.36e–11**



F **rs114053368**

Study	N	RAF(T)	Rsq	OR(95%CI)	P
NCI–MD	778	0.22	0.96	0.90 (0.70–1.15)	0.39
BioVU	1101	0.19	0.97	1.10 (0.85–1.41)	0.47
Ghana	1274	0.22	0.97	1.15 (0.92–1.44)	0.21
eMERGE	1491	0.18	0.98	0.98 (0.76–1.27)	0.90
ProHealth Kaiser	2251	0.21	0.96	0.91 (0.76–1.07)	0.26
CA UG	2633	0.22	0.96	1.01 (0.88–1.15)	0.93
BioMe	2652	0.19	0.97	1.06 (0.78–1.44)	0.69
ELLIPSE	8298	0.20	0.93	1.11 (1.03–1.20)	7.89e–03
AAPC	9530	0.19	0.97	1.13 (1.05–1.22)	1.09e–03
MVP	50990	0.20	0.93	1.16 (1.11–1.22)	2.57e–10
Meta–analysis	**80998**	**0.20**	**I^2 = 40**	**1.12 (1.08–1.16)**	**7.07e–12**

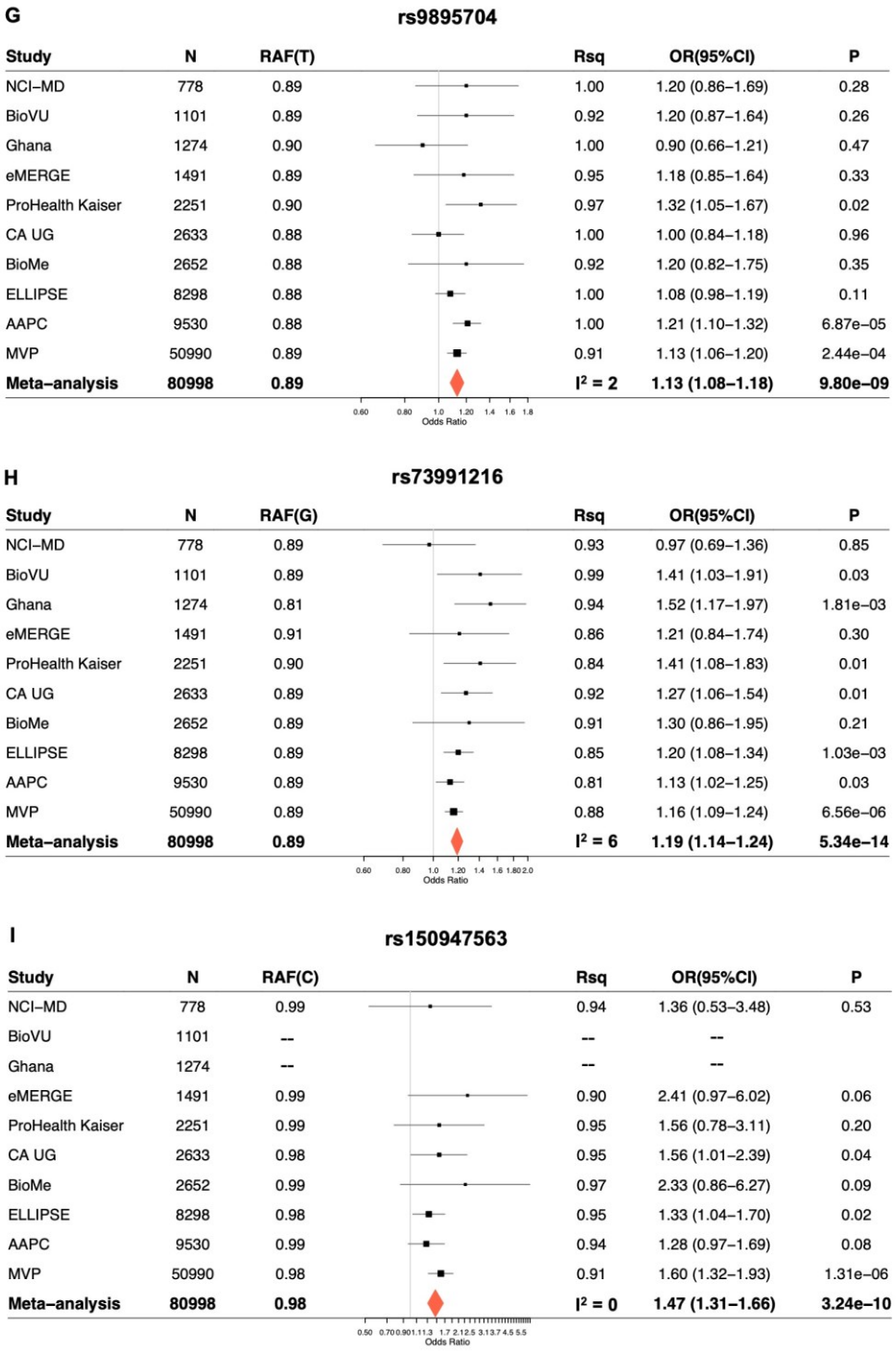


Figure S4 Associations of the novel variants with prostate cancer risk in individual studies of African ancestry. For each variant, the forest plot presented the risk allele frequency (RAF), imputation quality score (Rsq), odds ratio (OR), 95% confidence interval (CI), and P value from each study/consortium as well as the meta-analyzed results.

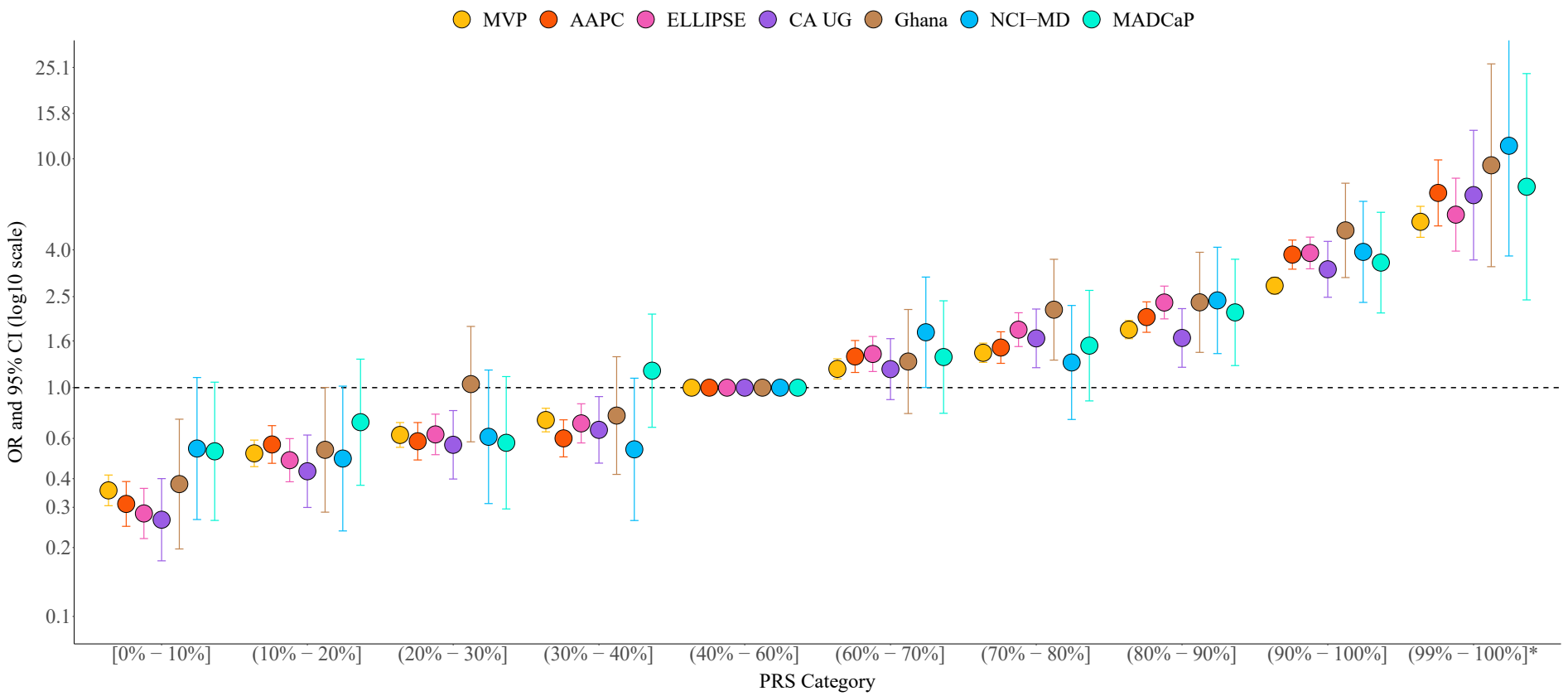


Figure S5 PRS associations with prostate cancer risk in individual studies of African ancestry. The association was assessed in MVP (6,353 cases and 53,362 controls), AAPC (4,822 cases and 4,642 controls), ELLIPSE (4,230 cases and 3,953 controls), CA UG (1,590 cases and 1,048 controls), Ghana (640 cases and 634 controls), NCI-MD (383 cases and 395 controls), and MADCaP Network (405 cases and 396 controls). The x-axis indicates the PRS category. The y-axis indicates the OR with error bars representing the 95% CIs for each PRS category compared to the 40%-60% PRS category. A separate analysis was performed to evaluate the PRS association with prostate cancer risk in men with extremely high genetic risk (*). The dotted horizontal line corresponds to an OR of 1.

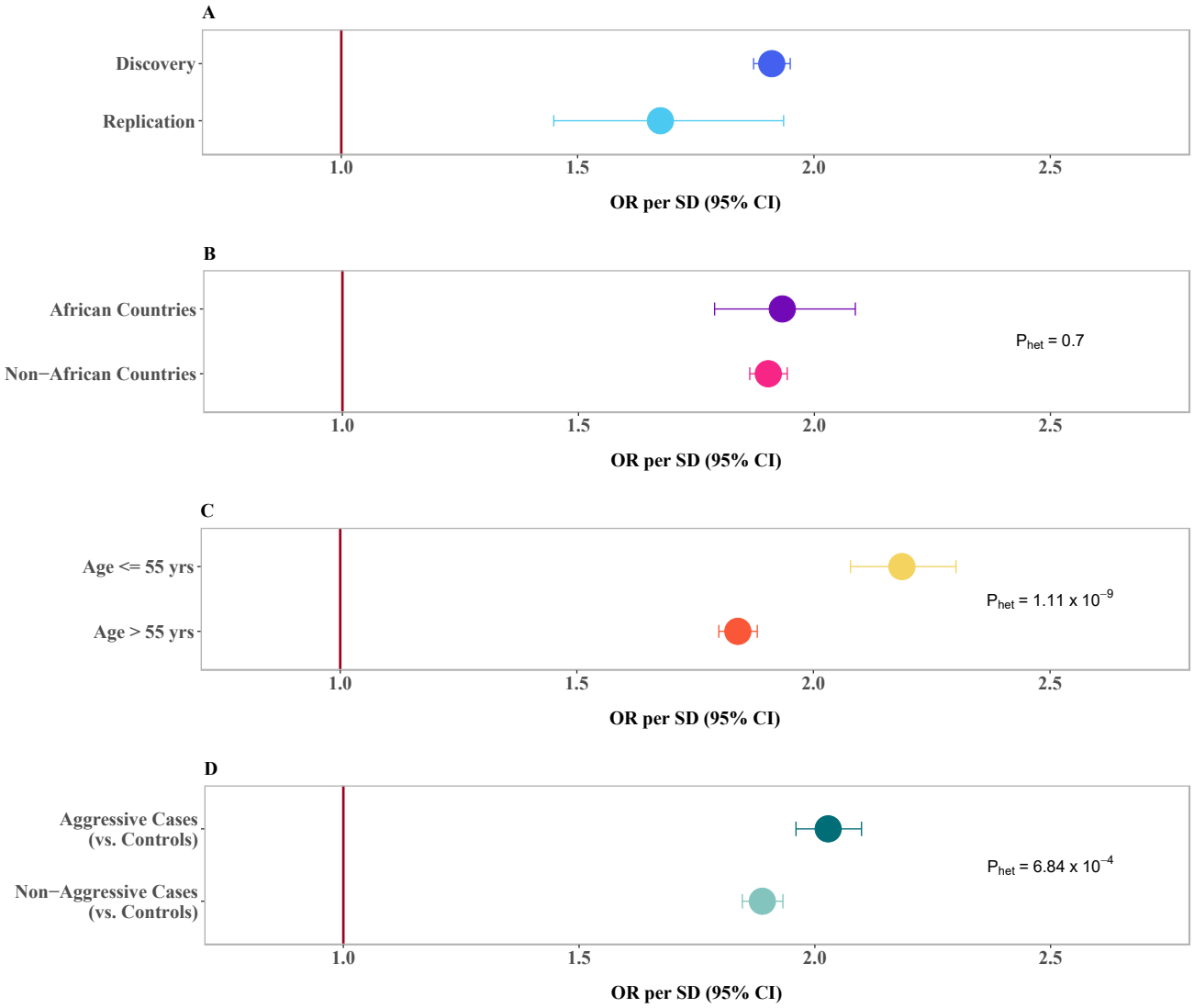


Figure S6 Association with prostate cancer risk per one SD increase of the multi-ancestry PRS. Associations with risk of total prostate cancer (A) in discovery and replication studies, (B) in studies from African countries and non-African countries, and (C) in men aged ≤ 55 years and > 55 years. Associations from the analyses of aggressive cases vs. controls and non-aggressive cases vs. controls are shown in (D). The x-axis indicates the OR with error bars representing the 95% CIs for per SD increase in PRS. The red vertical line corresponds to an OR of 1.

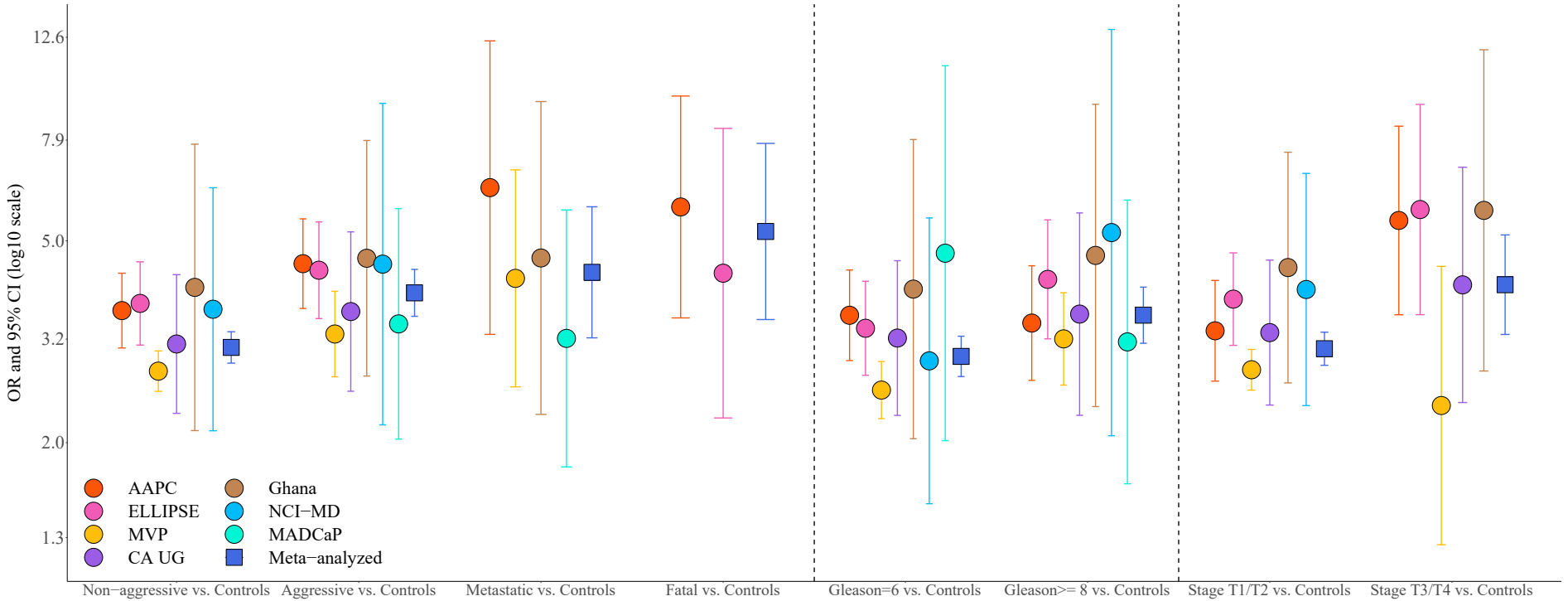


Figure S7 PRS associations (90% - 100% PRS category) with aggressive and non-aggressive forms of prostate cancer in individual studies of African ancestry. PRS associations were evaluated in each study separately and meta-analyzed across studies using a fixed-effects inverse-variance-weighted method. Studies with small number of cases (< 100) were not included in the meta-analysis (**Table S3**), and the meta-analyzed results are shown in **Table S15**. The x-axis indicates the specific case-control analysis. The y-axis indicates the OR with error bars representing the 95% CIs for the 90%-100% PRS category compared to the 40%-60% PRS category.