Title

Transcriptome-wide association analysis identifies candidate susceptibility genes for prostate-specific antigen levels in men without prostate cancer.

Journal

HGG Advances, 5(3)

Authors

Chen, Dorothy
Dong, Ruocheng
Kachuri, Linda
et al.

Publication Date

2024-07-18

DOI

10.1016/j.xhgg.2024.100315

Peer reviewed

HGG Advances

ARTICLE

Transcriptome-wide association analysis identifies candidate susceptibility genes for prostate-specific antigen levels in men without prostate cancer

Dorothy M. Chen,[1,17] Ruocheng Dong,[2,17] Linda Kachuri,[2,3] Thomas J. Hoffmann,[1,4] Yu Jiang,[1] Sonja I. Berndt,[5] John P. Shelley,[6] Kerry R. Schaffer,[7] Mitchell J. Machiela,[5] Neal D. Freedman,[5] Wen-Yi Huang,[5] Shengchao A. Li,[5] Hans Lilja,[8,9] Amy C. Justice,[10] Ravi K. Madduri,[10] Alex A. Rodriguez,[10] Stephen K. Van Den Eeden,[11] Stephen J. Chanock,[5] Christopher A. Haiman,[12,13] David V. Conti,[12,13] Robert J. Klein,[14] Jonathan D. Mosley,[15] John S. Witte,[1,2,16,18,*] and Rebecca E. Graff[1,18,19,*]

Summary

Deciphering the genetic basis of prostate-specific antigen (PSA) levels may improve their utility for prostate cancer (PCa) screening. Using genome-wide association study (GWAS) summary statistics from 95,768 PCa-free men, we conducted a transcriptome-wide association study (TWAS) to examine impacts of genetically predicted gene expression on PSA. Analyses identified 41 statistically significant ($p < 0.05/12{,}192 = 4.10 \times 10^{-6}$) associations in whole blood and 39 statistically significant ($p < 0.05/13{,}844 = 3.61 \times 10^{-6}$) associations in prostate tissue, with 18 genes associated in both tissues. Cross-tissue analyses identified 155 statistically significantly ($p < 0.05/22{,}249 = 2.25 \times 10^{-6}$) genes. Out of 173 unique PSA-associated genes across analyses, we replicated 151 (87.3%) in a TWAS of 209,318 PCa-free individuals from the Million Veteran Program. Based on conditional analyses, we found 20 genes (11 single tissue, nine cross-tissue) that were associated with PSA levels in the discovery TWAS that were not attributable to a lead variant from a GWAS. Ten of these 20 genes replicated, and two of the replicated genes had colocalization probability of >0.5: *CCNA2* and *HIST1H2BN*. Six of the 20 identified genes are not known to impact PCa risk. Fine-mapping based on whole blood and prostate tissue revealed five protein-coding genes with evidence of causal relationships with PSA levels. Of these five genes, four exhibited evidence of colocalization and one was conditionally independent of previous GWAS findings. These results yield hypotheses that should be further explored to improve understanding of genetic factors underlying PSA levels.

Introduction

Prostate-specific antigen (PSA) is a serine protease of the human tissue kallikrein-related (KLK) peptidase family.[1] Serum levels are commonly used as a biomarker for detection, monitoring, and risk stratification of prostate cancer (PCa).[2–5] A small portion of the highly abundant PSA in the prostate is released into blood and elevated in men with PCa. This process is incompletely understood, and disruption of prostate gland architecture by neoplastic transformation has been speculated as a possible mechanism.[2,4,6] PSA levels can additionally be influenced by age, ethnicity, body mass index, infection, prostate volume, benign prostatic hyperplasia (BPH), and germline genetic variation.[7,8]

PSA screening has been used for more than 30 years for the detection of PCa, which is the second leading cause of cancer death among men in the United States.[2,9–12] However, low test specificity and discrimination have complicated the use and interpretation of PSA as a screening tool.[11,13] Long-term outcomes data from large, population-based, randomized PSA screening trials show that PSA screening significantly reduces deaths from PCa and also results in the considerable overdiagnosis of low-risk disease.[14,15] PSA screening might be improved were it to account for variation in PSA levels that is attributable

[1]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158, USA; [2]Department of Epidemiology and Population Health, Stanford University, Stanford, CA 94305, USA; [3]Stanford Cancer Institute, Stanford University, Stanford, CA 94305, USA; [4]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA 94143, USA; [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20814, USA; [6]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [7]Department of Internal Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [8]Departments of Pathology and Laboratory Medicine, Surgery, Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; [9]Department of Translational Medicine, Lund University, 21428 Malmö, Sweden; [10]Argonne National Laboratory, Lemont, IL 60439, USA; [11]Division of Research, Kaiser Permanente Northern California, Oakland, CA 94612, USA; [12]Center for Genetic Epidemiology, Department of Population and Preventive Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA 90032, USA; [13]Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; [14]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [15]Departments of Internal Medicine and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37232, USA; [16]Departments of Biomedical Data Science and Genetics (by courtesy), Stanford University, Stanford, CA 94305, USA
[17]These authors contributed equally
[18]These authors contributed equally
[19]Lead contact
*Correspondence: jswitte@stanford.edu (J.S.W.), rebecca.graff@ucsf.edu (R.E.G.)
https://doi.org/10.1016/j.xhgg.2024.100315.

to germline genetics rather than PCa.[16–18] Twin studies and genome-wide association studies (GWASs) have estimated that 30%–45% of serum PSA variation is influenced by hereditary factors.[19,20] The most recent GWAS meta-analysis from our group identified 318 independent PSA-associated variants, which explained approximately 9% of genetic variation in PSA levels.[18]

Additional genetic variation in PSA levels may be determined by analyses of gene expression, as opposed to individual variants. Such work could lead to the identification of regions or biological pathways that affect PSA levels and point to mechanisms underlying constitutional increases in PSA levels in the absence of carcinogenesis. Transcriptome-wide association studies (TWASs) allow for the identification of genes whose *cis*-regulated expression is associated with complex polygenic traits.[21–24] To identify genes associated with PSA levels and prioritize them for functional investigation, we performed a TWAS of PSA levels based on summary statistics from a GWAS meta-analysis of 95,768 men without PCa.[16,25] In doing so, we aimed to identify genes located at loci where prior GWAS signals fell below genome-wide significance. In further implementing colocalization analyses, we were able to identify variants that influence both complex traits and gene expression.[26] We additionally implemented conditional analyses, fine-mapping, colocalization analyses, and pathway analyses to yield comprehensive insights into genes' relationships with PSA levels.

Subjects and methods

Ethical Considerations

Informed consent was obtained from all study participants. The UK Biobank received ethics approval from the Research Ethics Committee (reference: 11/NW/0382) in accordance with the UK Biobank Ethics and Governance Framework. The research was conducted with approved access to UK Biobank data under application number 14105. We used PSA GWAS results from Kaiser Permanente's Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort; these results were previously published by Hoffmann et al.[17] The original study was approved by the Kaiser Permanente Northern California Institutional Review Board and the University of California, San Francisco Human Research Protection Program Committee on Human Research. The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial was approved by the Institutional Review Board at each participating center and the National Cancer Institute. The informed consent document signed by PLCO study participants allows use of these data by investigators for discovery and hypothesis generation in the investigation of the genetic contributions to cancer and other adult diseases. Our study includes publicly posted genomic summary results from the PLCO Atlas. (add reference: Machiela MJ, et al. GWAS Explorer: an open-source tool to explore, visualize, and access GWAS summary statistics in the PLCO Atlas. Sci. Data. 2023;10:25.) No institutional review board review is required for PLCO summary data use. The Vanderbilt University Medical Center Institutional Review Board approved the BioVU study. The Malmö Diet and Cancer Study (MDCS) was approved by the local ethics committee. The Veteran Affairs Central Institutional Review Board approved the Million Veteran Program (MVP).

Discovery populations

Existing GWAS summary statistics[16] from 95,768 PCa-free men from the following five study populations were included: UK Biobank, GERA, PLCO, BioVU, and MDCS. The men included in these analyses were restricted to PCa-free individuals (i.e., without a PCa diagnosis or history of prostate resection, where information was available) with at least one PSA measurement. Analyses were restricted to men with PSA values between 0.01 ng/mL and 10 ng/mL to capture variations in PSA levels, while minimizing potential for including undiagnosed PCa cases. The upper threshold of 10 ng/mL corresponds to pre-treatment PSA levels for low-risk PCa under the validated D'Amico risk classification system for PCa.[27,28] The lower bound of 0.01 ng/mL ensured that all participants had a functioning prostate and corresponds to the analytical limit of detection for early generations of PSA assays, many of which were used in the studies included in this analysis.[29,30] Median PSA levels were used for individuals with multiple PSA measurements available, with the exception of PLCO, which used PSA levels at the start of follow-up of the trial. Additional details about the discovery cohorts are described in detail in Kachuri et al.[16] and Table S1.

Quality control and GWAS meta-analysis

Genotyping, imputation, study-specific quality control of genetic data, and the GWAS meta-analysis have been previously described in Kachuri et al.[16] Briefly, ancestry and study-specific GWAS analyses used linear regression of log(PSA) on genetic variants, age, and the first 10 genetic ancestry principal components. After results across studies were meta-analyzed within ancestral groups, ancestry-specific summary statistics from European ancestry (n = 85,824), African ancestry (n = 3,509), East Asian ancestry (n = 3,337), and Hispanic/Latino (n = 3,098) individuals were meta-analyzed to generate multi-ancestry summary statistics.

MetaXcan transcriptome-wide gene-based analysis

We undertook a TWAS using the MetaXcan approach, which directly estimates Z scores for associations between gene expression levels and PSA levels using meta-analyzed GWAS summary statistics with tissue-specific prediction models. Multivariate adaptive shrinkage (mashr) prediction models were trained on GTEx (version 8) expression quantitative trait loci (eQTL) data for whole blood and, separately, prostate tissue. The covariances and weights for models trained on individual tissue types were from the PredictDB Data Repository (https://predictdb.org/).[25] We also undertook a cross-tissue analysis across 45 GTEx tissues to identify potential associations with genes that may have been missed in single-tissue analyses. Tissue types found primarily or exclusively in females (mammary breast, ovary, uterus, and vagina) were excluded from the multiple tissue analysis. A Bonferroni correction was implemented based on the number of genes tested (whole blood: $0.05/12{,}192 = 4.10 \times 10^{-6}$; prostate: $0.05/13{,}844 = 3.61 \times 10^{-6}$; cross-tissue: $0.05/22{,}249 = 2.25 \times 10^{-6}$). Primary analyses were based on the multi-ancestry GWAS summary statistics. Sensitivity analyses restricted to men of European ancestry.

Replication analyses

We performed replication analyses of the findings from our discovery TWAS by implementing the MetaXcan approach in 209,318

PCa-free men of European ancestry enrolled in the MVP who had at least one PSA measurement.[18,31] The MVP has been previously described. Briefly, it is an expansive and diverse biobank that was initiated in 2011 and contains linkages to electronic health record data from more than 50 veterans affairs medical centers. Based on GWAS summary statistics, we estimated gene expression level associations using the aforementioned GTEx v8 mashr prediction models.[32] Replication analyses were conducted based on whole blood, prostate tissue, and cross-tissue. We used the following replication significance thresholds based on the number of significant genes from our discovery TWAS analyses: $0.05/41 = 1.21 \times 10^{-3}$ for whole blood, $0.05/39 = 1.29 \times 10^{-3}$ for prostate tissue, and $0.05/155 = 3.22 \times 10^{-4}$ for cross-tissue.

Conditional analyses ascertaining independence of TWAS and previous GWAS findings

Methods that assess the conditional fit of eQTLs on GWAS hits are lacking. Thus, to identify the set of TWAS genes for PSA that were not previously identified by GWAS, we first limited consideration to significant TWAS genes that did not contain genome-wide significant ($p < 5 \times 10^{-8}$) variants from the Kachuri et al.[16] GWAS (i.e., within exact gene boundaries). Second, for the remaining genes, we performed conditional analyses using genome-wide complex trait analysis conditional and joint multiple-SNP analysis (COJO)[33] that simultaneously modeled eQTLs for a given TWAS-identified gene and GWAS results. For individual tissue analyses, eQTLs used in the prediction models of transcriptome-wide significant genes were selected from the tissue-specific PredictDB TWAS eQTL weight files for each gene. For the cross-tissue analyses, the maximum noncollinear eQTLs (variance inflation factor of <8 within a window size of 50 kb) were extracted from all 45 GTEx tissues (i.e., to remove variants in linkage disequilibrium [LD]). COJO was performed using LD reference panels from 10,000 European ancestry UK Biobank participants (as 89.6% of our population was European ancestry).[16] GWAS summary statistics were then conditioned on the eQTLs for one gene at a time. If the set of genome-wide significant variants was reduced after conditioning on eQTLs used to predict gene expression, then the eQTLs and genome-wide significant variants were assumed to be in LD and thus not independent. However, if the set of genome-wide significant variants remained unchanged after conditional analyses, then we considered the TWAS gene to be independent of GWAS variants from Kachuri et al.[16]

Probabilistic fine-mapping

Similar to GWASs, marginal TWAS associations cannot be interpreted as causal. Identifying the most likely causal genes requires accounting for LD and eQTL sharing among genes in *cis*-regions. We performed genome-wide statistical fine-mapping for whole blood and prostate tissue using causal TWAS (cTWAS),[34] which jointly models the effects of imputed gene expression and genetic variants to derive posterior inclusion probabilities (PIPs). Cross-tissue fine-mapping is currently unavailable within the cTWAS analysis framework. Standard cTWAS pre-processing included restriction to protein-coding genes. LD was estimated from UKB reference data in individuals of European ancestry. Prior parameters were estimated for gene and variant effects in each genomic block using an empirical Bayesian approach, and Sum of Single Effects regression with summary statistics fine-mapping was applied to all variables after parameter estimation.[35,36] A random subset of 10% of all variants was used during initial parameter estimation assuming at most five causal effects for each region, but regions with PIPs of 0.8 or greater were recomputed. PIPs were calculated for all genes, all variants in regions with strong gene signals, and 10% of variants in other regions. A standard threshold of PIP of more than 0.8 was used for declaring statistically significant genes.[34]

Colocalization analyses

To further investigate whether genes and variants exhibited shared genetic signals, we used fastENLOC software (version 2) to undertake colocalization analyses of eQTLs across multiple tissues for all genes with GTEx v8 prediction models and all variants investigated in the Kachuri et al. GWAS.[16,26,37] Colocalization was performed across tissues rather than complete individual-tissue colocalization analyses. Variant-specific enrichment priors were directly estimated by applying an established multiple imputation procedure to GWAS summary statistics and eQTL data.[37,38] Unlike traditional window-based approaches, fastENLOC uses Bayesian credible sets—clusters of variants in LD that share common underlying association signals—as regional LD blocks for colocalization analyses. These independent association signals, which were inferred from multi-variant fine-mapping analyses, were used to compute a colocalization probability for each LD region. For these analyses, LD blocks were based on 1000 Genome Project European ancestry individuals.[39] The GWAS summary statistics were colocalized for 49 GTEx version 8 tissues using pre-computed GTEx multiple-tissue annotations. Signal-level results returned regional-level colocalization probabilities (RCPs) between eQTL and GWAS signals, which sum up the variant-level colocalization probabilities of correlated variants within an LD block that harbors a single GWAS association signal. Gene results were based on gene-level RCPs (GRCPs), which represent the probability that a candidate gene contains at least one colocalized variant.[26] RCP and GRCP values of greater than 0.5 indicate strong evidence of shared genetic signals between eQTLs and GWAS variants.[26]

Pathway enrichment analysis

To explore the potential biological relevance of PSA-associated genes, we applied the Enrichr tool to all significant genes identified in the whole blood, prostate, and cross-tissue analyses to assess enrichment against three gene set libraries: Kyoto Encyclopedia of Genes and Genomes (KEGG) 2021 human; Gene Ontology (GO) biological process 2021; and GO molecular function 2021.[40,41] Enrichment was assessed by multiplying the p value from a Fisher's exact test with the Z score of the deviation from the expected rank.[42,43] Pathways with a Benjamini-Hochberg-corrected p value of <0.05 were considered statistically significantly enriched.

Results

Using MetaXcan and GWAS summary statistics based on 95,768 individuals (Table S1), TWASs were conducted for whole blood, prostate, and cross-tissue matrices. The sample was primarily European ancestry (n = 85,824), although African ancestry (n = 3,509), East Asian ancestry (n = 3,337), and Hispanic/Latino (n = 3,098) individuals were also included. The median PSA value across cohorts was 2.35 ng/mL.

Overall, 93% of the GTEx variants used to predict gene expression in individual tissue models were available in

Figure 1. Miami plot for the discovery TWAS of PSA levels based on predicted gene expression in whole blood
Each dot represents a gene, and the red dashed lines represent the significance threshold after Bonferroni correction (4.10×10^{-6}). All statistically-significant genes are annotated; the genes independent of results from the largest prior GWAS are highlighted in yellow. Genes on the top half of the plot are positively associated with PSA levels, and the genes on the lower half of the plot are inversely associated with PSA levels.

our multi-ancestry GWAS summary statistics that included 20,060,440 variants. GTEx gene expression models for whole blood were derived from $n = 670$ participants with genotype data, and samples from $n = 221$ individuals were used to train prostate tissue prediction models.[44] In the TWAS based on whole blood gene expression models, 41 of 12,192 genes were associated with PSA levels at the Bonferroni-corrected threshold (Figure 1 and Table S2). In the TWAS based on prostate tissue gene expression models, 39 of 13,844 genes were statistically significantly associated with PSA levels (Figure 2 and Table S3). Cross-tissue analyses identified 155 of 22,249 genes whose expression was associated with elevated PSA levels (Figure 3 and Table S4). In total, we replicated 151 of 173 unique associations across the three individual tissue analyses in the MVP cohort. Sensitivity analyses restricting to European ancestry individuals were largely comparable with our primary TWAS results (Tables S2–S4).

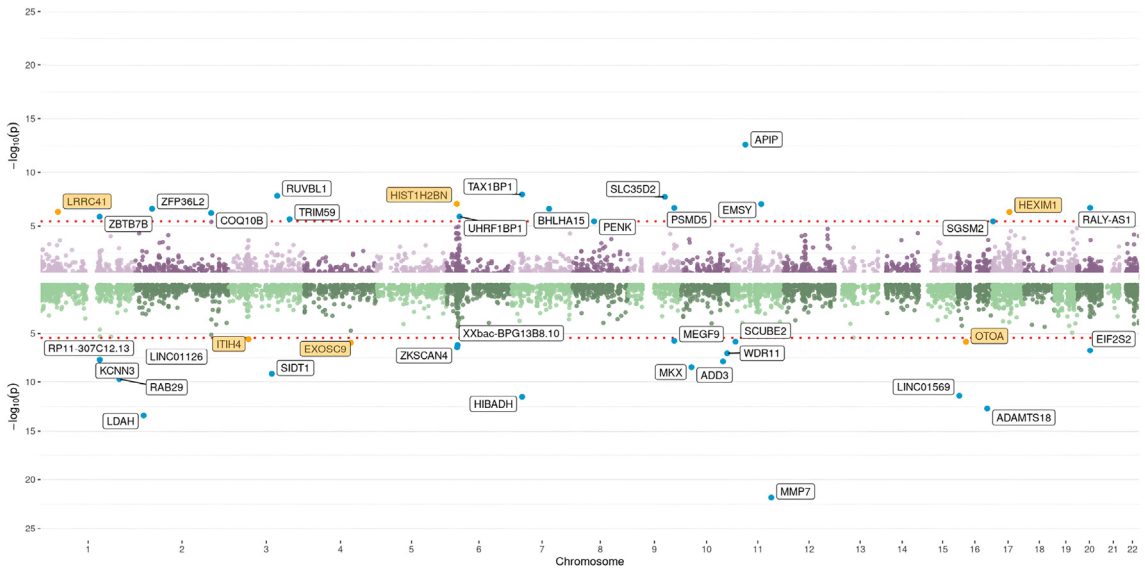

Figure 2. Miami plot for the discovery TWAS of PSA levels based on predicted gene expression in prostate tissue
Each dot represents a gene, and the red dashed lines represent the significance threshold after Bonferroni correction (3.61×10^{-6}). All statistically significant genes are annotated; the genes independent of results from the largest prior GWAS are highlighted in yellow. Genes on the top half of the plot are positively associated with PSA levels, and the genes on the lower half of the plot are inversely associated with PSA levels.

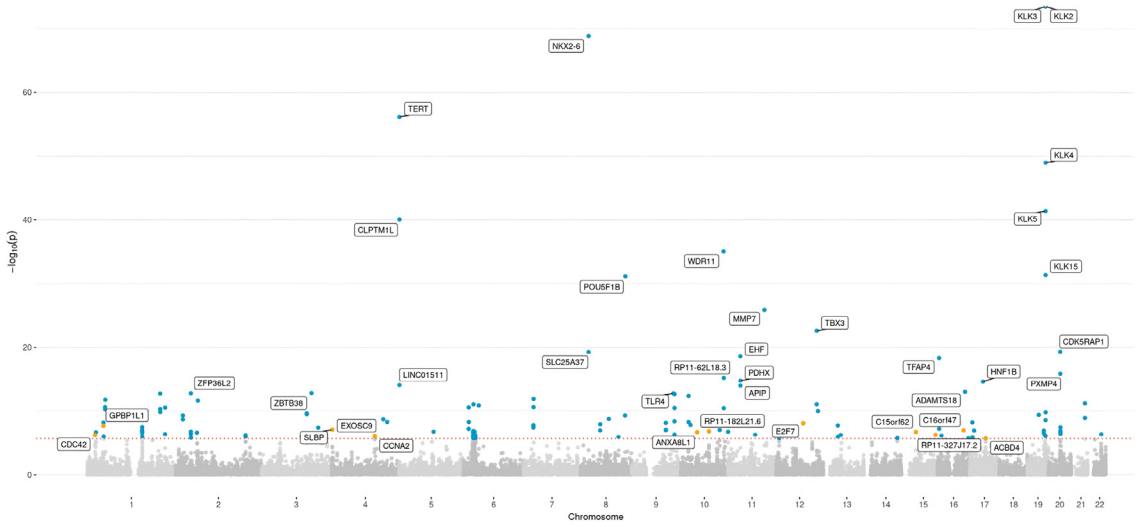

Figure 3. Manhattan plot for the discovery TWAS of PSA levels based on cross-tissue predicted gene expression
Each dot represents a gene, and the red dashed line represents the significance threshold after Bonferroni correction (2.25×10^{-6}). The top 30 genes with the smallest association p values are annotated and colored blue. Directions of association between genes and PSA levels are not available for cross-tissue analyses due to the joint tissue inference methodology used in S-MultiXcan.

Whole blood TWAS

Of the 41 genes whose whole blood expression was significantly associated with PSA levels (Figure 1 and Table S2), the expression of 20 genes was positively associated with elevated PSA and the expression of 21 genes was inversely associated. Replication analyses in the MVP cohort confirmed 33 of 41 genes in whole blood, and all genes demonstrated consistent direction of effect. Two of the significant (and replicated) genes are located at 19q13.33, which contains *KLK3*, the gene most strongly linked to serum PSA levels.[45] While a prediction model for *KLK3* was not available in whole blood, a member of the same gene family at 19q13.33, *KLK2*, was the gene most strongly associated with PSA levels ($p = 1.28 \times 10^{-62}$).[45] Notably, *KLK2* and *KLK3* are located at the same genomic locus and are known to have high LD.[7] Twenty-four of the 41 genes did not contain genome-wide significant variants and had not been annotated in previous PSA GWASs. Conditional analyses identified 7 of the 41 TWAS genes to be independent of previously published PSA GWAS findings (Table 1).[16] Of these, increased expression of six genes was associated with elevated PSA levels in PCa-free men: *GPBP1L1* (1p34.1, $p = 2.21 \times 10^{-7}$); *TMEM69* (1p34.1, $p = 2.21 \times 10^{-7}$); *UQCRH* (1p33, $p = 4.76 \times 10^{-7}$); *ACTRT3* (3q26.2, $p = 5.90 \times 10^{-7}$); *EXOSC9* (4q27, $p = 1.80 \times 10^{-6}$); and *CCNA2* (4q27, $p = 7.80 \times 10^{-7}$). Decreased expression of *ITH4* was associated with increased PSA levels (3p21.1, $p = 3.65 \times 10^{-6}$).

Fine-mapping analyses in whole blood were based on 10,106 protein-coding genes. We identified 14 genes with a PIP of 0.8 or greater (Table S5), of which two overlapped with Bonferroni-significant genes in the whole blood TWAS discovery analyses: *PMVK* (1q21.3) and *A2M* (12p13.31).

Prostate tissue TWAS

In prostate tissue, increased expression of 18 and decreased expression of 21 genes was associated with elevated PSA levels (Figure 2 and Table S3). Thirty-one of 39 genes in prostate tissue replicated in the MVP cohort, and effect sizes for all were in the same direction as those reported in the discovery analyses. No gene prediction models were available for *KLK3* or any other genes in the *KLK* family in prostate tissue. Instead, *MMP7* on 11q22 was most strongly associated with PSA levels ($p = 2.78 \times 10^{-18}$). We identified 18 PSA-associated TWAS genes containing known genome-wide significant variants. Of the remaining 21 significant genes, six were conditionally independent of previous GWAS results after conditioning on the eQTL variants used in gene prediction models (Table 1). Of these six, increased expression of three genes was associated with elevated PSA levels: *LRRC41* (1p33, $p = 4.76 \times 10^{-7}$), *HIST1H2BN* (6p22.1, $p = 8.86 \times 10^{-8}$), and *HEXIM1* (17q21.31, $p = 4.95 \times 10^{-7}$). Decreased expression of the remaining three genes was associated with increased PSA levels: *ITIH4* (3p21.1, $p = 2.69 \times 10^{-6}$), *EXOSC9* (4q27 $p = 1.18 \times 10^{-6}$), and *OTOA* (16p12.2, $p = 1.44 \times 10^{-6}$).

Using 10,806 protein-coding gene expression models in prostate tissue, cTWAS identified 16 genes with a PIP of 0.8 or greater (Table S6), three of which were Bonferroni-significant in main analyses: *ADD3* (10q25.1), *MMP7* (11q22.2), and *OTOA* (16p12.2). As specified above, *OTOA* was also identified by conditional analyses.

Cross-tissue TWAS

In cross-tissue analyses, 136 of 155 genes had expression levels that were associated with elevated PSA levels in our replication cohort. One additional gene, *TMEM69*, which was discovered in the whole blood analyses and did not reach statistical significance in the cross-tissue discovery

Table 1. Conditionally independent genes associated with prostate-specific antigen levels in transcriptome-wide association studies based on whole blood, prostate tissue, and cross-tissue predicted gene expression

Ensembl gene ID[a]	Gene name	Chr	Band[b]	Analysis[c]	Z	Beta	Disc *p* value	Rep *p* value
ENSG00000132128	*LRRC41*	1	p33	prostate	5.04	0.29	4.76×10^{-7}	9.39×10^{-3}
ENSG00000159592	*GPBP1L1*	1	p34.1	whole blood	5.18	17.92	2.21×10^{-7}	1.81×10^{-3}
				cross-tissue[d]	−1.92	N/A	2.08×10^{-8}	4.28×10^{-8}
ENSG00000159596	*TMEM69*	1	p34.1	whole blood	5.18	0.28	2.21×10^{-7}	1.81×10^{-3}
ENSG00000173660	*UQCRH*	1	p33	whole blood	5.04	8.04	4.76×10^{-7}	9.39×10^{-3}
ENSG00000070831	*CDC42*	1	p36.12	cross-tissue	−1.55	N/A	6.24×10^{-7}	1.01×10^{-3}
ENSG00000184378	*ACTRT3*	3	q26.2	whole blood[d]	4.99	0.25	5.90×10^{-7}	7.81×10^{-11}
ENSG00000055955	*ITIH4*	3	p21.1	whole blood[d]	−4.63	−0.03	3.65×10^{-6}	2.75×10^{-8}
				prostate[d]	−4.69	−0.03	2.69×10^{-6}	8.25×10^{-8}
ENSG00000163950	*SLBP*	4	p16.3	cross-tissue	−2.15	N/A	8.40×10^{-8}	4.23×10^{-3}
ENSG00000145386	*CCNA2*	4	q27	whole blood[d]	4.94	0.61	7.80×10^{-7}	3.93×10^{-7}
				cross-tissue[d]	−1.81	N/A	1.02×10^{-6}	6.89×10^{-6}
ENSG00000123737	*EXOSC9*	4	q27	whole blood	4.77	0.09	1.80×10^{-6}	2.07×10^{-2}
				prostate	−4.86	−0.42	1.18×10^{-6}	1.87×10^{-2}
				cross-tissue	0.36	N/A	8.60×10^{-7}	4.04×10^{-5}
ENSG00000233822	*HIST1H2BN*	6	p22.1	prostate[d]	5.35	0.24	8.86×10^{-8}	3.58×10^{-18}
ENSG00000272447	*RP11-182L21.6*	10	q22.3	cross-tissue	1.25	N/A	1.54×10^{-7}	5.19×10^{-2}
ENSG00000264230	*ANXA8L1*	10	q11.22	cross-tissue	−0.46	N/A	2.17×10^{-7}	1.50×10^{-2}
ENSG00000165891	*E2F7*	12	q21.2	cross-tissue[d]	1.45	N/A	8.54×10^{-9}	1.05×10^{-13}
ENSG00000188277	*C15orf62*	15	q15.1	cross-tissue	−0.85	N/A	2.01×10^{-7}	3.56×10^{-1}
ENSG00000259359	*RP11-327J17.2*	15	q26.2	cross-tissue	−0.16	N/A	5.82×10^{-7}	3.71×10^{-2}
ENSG00000155719	*OTOA*	16	p12.2	prostate	−4.82	−5.36	1.44×10^{-6}	2.01×10^{-1}
ENSG00000197445	*C16orf47*	16	q22.3	cross-tissue[d]	0.44	N/A	1.11×10^{-7}	1.24×10^{-12}
ENSG00000141052	*MYOCD*	17	p12	cross-tissue[d]	−3.03	N/A	1.20×10^{-7}	8.40×10^{-7}
ENSG00000181513	*ACBD4*	17	q21.31	cross-tissue[d]	−1.05	N/A	2.17×10^{-6}	6.57×10^{-7}
ENSG00000186834	*HEXIM1*	17	q21.31	prostate[d]	5.03	0.12	4.95×10^{-7}	3.80×10^{-3}

Chr, chromosome; Disc, discovery; Rep, replication.
[a]The genes included in this table were significant (Bonferroni-adjusted $p < 0.05$) in at least one TWAS and, based on conditional analyses, independent from the largest prior GWAS of PSA levels.
[b]GRCh38.
[c]Indicates the tissue(s) in which predicted gene expression was significantly associated with PSA levels.
[d]Reached Bonferroni significance in tissue-specific replication analyses.

analyses, did not replicate in whole blood but was significantly associated with elevated PSA levels in cross-tissue replication analyses ($p = 1.948 \times 10^{-8}$). Cross-tissue analyses do not yield directionality of effects, so we were unable to compare effect directions between discovery and replication analyses. Among the 155 statistically significant genes from cross-tissue analyses, eight, all of which replicated, are located on 19q13.33–19q13.41 and are part of the expanded human *KLK* gene family (Figure 3 and Table S4).

We found 71 genes that do not contain genome-wide significant PSA variants within their gene boundaries. Conditional analyses identified 13 genes from this set that were independent of known genome-wide significant variants: *CDC42* (1p36.12, $p = 6.24 \times 10^{-7}$), *GPBP1L1* (1p34.1, $p = 2.08 \times 10^{-8}$), *SLBP* (4p16.3, $p = 8.40 \times 10^{-8}$), *EXOSC9* (4q27, $p = 8.60 \times 10^{-7}$), *CCNA2* (4q27, $p = 1.02 \times 10^{-6}$), *ANXA8L1* (10q11.22, $p = 2.17 \times 10^{-7}$), *RP11-182L21.6* (10q22.3, $p = 1.54 \times 10^{-7}$), *E2F7* (12q21.2, $p = 8.54 \times 10^{-9}$), *C15orf62* (15q15.1, $p = 2.01 \times 10^{-7}$), *RP11-327J17.2* (15q26.2, $p = 5.82 \times 10^{-7}$), *C16orf47* (16q22.3, $p = 1.11 \times 10^{-7}$), *MYOCD* (17p12, $p = 1.20 \times 10^{-7}$), and *ABCD4* (17q21.31, $p = 2.17 \times 10^{-6}$) (Table 1).

Colocalization analysis of cross-tissue association signals

Cross-tissue colocalization analyses identified 372 signals with an RCP of more than 0.5 at the variant level

Table 2. Gene level colocalization results for significant genes from the transcriptome-wide association study of prostate-specific antigen levels based on cross-tissue predicted gene expression

Ensembl gene ID	Gene name	Chr	Cytoband[a]	GRCP	GLCP
ENSG00000069275	*NUCKS1*	1	q32.1	0.97	2.55
ENSG00000117280	*RAB29*	1	q32.1	1.13	0.52
ENSG00000117602	*RCAN3*	1	p36.11	2.13	−24.54
ENSG00000143603	*KCNN3*	1	q21.3	1.42	−136.50
ENSG00000158715	*SLC45A3*	1	q32.1	0.59	15.17
ENSG00000163344	*PMVK*	1	q21.3	1.97	−50.77
ENSG00000163346	*PBXIP1*	1	q21.3	1.99	−14.21
ENSG00000270361	*RP11-307C12.13*	1	q21.3	1.08	−0.32
ENSG00000115507	*OTX1*	2	p15	4.91	−7.33
ENSG00000143869	*GDF7*	2	p24.1	1.00	620.30
ENSG00000068885	*IFT80*	3	q25.33	2.41	24.05
ENSG00000072858	*SIDT1*	3	q13.2	0.62	−211.30
ENSG00000113810	*SMC4*	3	q25.33	2.18	−52.64
ENSG00000132394	*EEFSEC*	3	q21.3	1.34	−160.40
ENSG00000175792	*RUVBL1*	3	q21.3	1.24	−72.12
ENSG00000177311	*ZBTB38*	3	q23	1.22	−39.66
ENSG00000186432	*KPNA4*	3	q25.33	1.27	−19.18
ENSG00000213186	*TRIM59*	3	q25.33	1.54	−38.26
ENSG00000123737	*EXOSC9*[b]	4	q27	2.24	−108.20
ENSG00000145386	*CCNA2*[b]	4	q27	1.39	−54.29
ENSG00000049656	*CLPTM1L*	5	p15.33	0.71	−27.67
ENSG00000164362	*TERT*	5	p15.33	0.72	4.54
ENSG00000184357	*HIST1H1B*	6	p22.1	0.92	2.70
ENSG00000189298	*ZKSCAN3*	6	p22.1	0.84	1.28
ENSG00000204713	*TRIM27*	6	p22.1	0.76	−12.84
ENSG00000233822	*HIST1H2BN*[b]	6	p22.1	1.08	−9.88
ENSG00000281831	*HCP5B*	6	p22.1	0.88	1.25
ENSG00000164684	*ZNF704*	8	q21.13	1.00	−94.45
ENSG00000167034	*NKX3-1*	8	p21.2	0.96	−293.10
ENSG00000056558	*TRAF1*	9	q33.2	0.78	5,137.00
ENSG00000095261	*PSMD5*	9	q33.2	0.61	−14.25
ENSG00000119403	*PHF19*	9	q33.2	0.97	−23.73
ENSG00000130956	*HABP4*	9	q22.32	0.82	30.20
ENSG00000130958	*SLC35D2*	9	q22.32	1.22	−11.44
ENSG00000165244	*ZNF367*	9	q22.32	0.60	−10.13
ENSG00000148700	*ADD3*	10	q25.1	1.21	−118.60
ENSG00000272447	*RP11-182L21.6*[b]	10	q22.3	0.99	5.94
ENSG00000137673	*MMP7*	11	q22.2	0.96	183.70
ENSG00000158636	*EMSY*	11	q13.5	2.64	−9.56
ENSG00000177951	*BET1L*	11	p15.5	0.78	−63.92
ENSG00000111707	*SUDS3*	12	q24.23	1.33	−225.60

(*Continued on next page*)

Table 2. ***Continued***

Ensembl gene ID	Gene name	Chr	Cytoband[a]	GRCP	GLCP
ENSG00000175899	*A2M*	12	p13.31	0.64	−156.60
ENSG00000172766	*NAA16*	13	q14.11	1.25	−14.31
ENSG00000259359	*RP11-327J17.2*[b]	15	q26.2	1.10	2.93
ENSG00000141052	*MYOCD*	17	p12	0.62	−1.94
ENSG00000275410	*HNF1B*	17	q12	0.81	2.32
ENSG00000176920	*FUT2*	19	q13.33	3.04	320.20
ENSG00000177045	*SIX5*	19	q13.32	1.87	−15.39
ENSG00000185800	*DMWD*	19	q13.32	1.00	461.40
ENSG00000185022	*MAFF*	22	q13.1	3.63	3.06

Chr, chromosome; GLCP, gene-level colocalization probability; GRCP, gene-level variant colocalization probability.
[a]GRCh38.
[b]These genes were independent from the largest prior GWAS of PSA levels based on COJO analyses.

(Table S7) and 307 genes with a GRCP of more than 0.5 at the gene level (Table S8). The top 10 genes with the largest GRCP were *OTX1* (2p15), *MAFF* (22q13.1), *FUT2* (19q13.33), *EMSY* (11q13.5), *IFT80* (3q25.33), *EXOSC9* (4q27), *SMC4* (3q25.33), *RCAN3* (1p36.11), *PBXIP1* (1q21.3), and *PMVK* (1q21.3).

Among 155 significant genes from the cross-tissue TWAS, locus-level results identified 50 genes with strong evidence of colocalization (Table 2). Signal-level results identified 66 regions in 51 genes (Table S9). Colocalization analyses identified 50 genes from cross-tissue analyses that shared genetic signals with GWAS variants, five of which were significantly associated with PSA levels in cross-tissue TWAS analyses.

When comparing the overlap between transcriptome-wide significant genes and results from colocalization analyses, we identified five genes (*EXOSC9*, *CCNA2*, *HIST1H2BN*, *RP11-182L21.6*, and *RP11-327J17.2*) with GRCP and RCP values of greater than 0.5 that were not attributable to a known lead variant from GWAS. Two of these genes are long non-coding RNA, and the remaining three are protein-coding genes. *CCNA2* and *HIST1H2BN* had expression levels that were associated with PSA levels in replication analyses. Four genes identified through tissue-specific fine-mapping analyses also exhibited strong evidence of colocalization: *A2M* and *PVMK* in whole blood and *MMP7* and *ADD3* in prostate tissue.

Significant genes overlapping across TWAS

Of the 41 genes that were significantly associated with PSA levels in the whole blood TWAS, 34 were also imputed in the prostate tissue TWAS. Thirty of 39 significantly associated genes from the prostate tissue TWAS were also evaluated in whole blood analyses. Of the 155 significantly associated genes detected in the cross-tissue analysis, 89 and 105 genes had transcriptome prediction models available for whole blood and prostate tissue, respectively.

We identified 13 genes whose predicted expression was associated with elevated PSA levels across all three TWASs: *RAB29* (1q32.1), *LDAH* (2p24.1), *ZFP36L2* (2p21), *SIDT1* (3q13.2), *TRIM59* (3q25.33), *EXOSC9* (4q27), *HIBADH* (7p15.2), *TAX1BP1* (7p15.2), *MEGF9* (9q33.2), *ADD3* (10q25.1), *APIP* (11p13), *LINC01569* (16p13.3), and *EIF2S2* (20q11.22) (Figure 4).

Of these genes, only *EXOSC9* was conditionally independent after accounting for prior loci and demonstrated evidence of colocalization (described above). Five additional associations were not statistically significant in the cross-tissue analysis, but demonstrated significant signals in whole blood and prostate tissue. Seventeen significant genes were shared by only the whole blood and cross-tissue TWASs, six of which did not have prediction models in prostate tissue. Similarly, the prostate tissue and cross-tissue TWAS shared 14 significant genes, of which seven did not have prediction models in whole blood.

Pathway enrichment analysis

Using significantly associated signals from our main analyses, TWAS genes were significantly enriched in 18 pathways underlying molecular activities and biological processes present in KEGG and GO gene set databases (Table 3; Figure 5). Within the whole blood gene set, analyses detected nominally significant enrichment of the renin-angiotensin-aldosterone system in KEGG, which works to regulate arterial blood pressure. The observed signal was primarily driven by two genes in the *KLK* family and was only nominally significant. Transcriptional dysregulation in cancer was significantly enriched in KEGG for cross-tissue results. No pathways were significantly enriched in prostate gene sets in the KEGG catalog.

Enrichment analyses within the GO biological process and molecular function gene set libraries identified significant enrichments only in the cross-tissue gene set, with 16 pathways. No pathways were significantly enriched in GO for whole blood or prostate tissue TWAS sets. We also

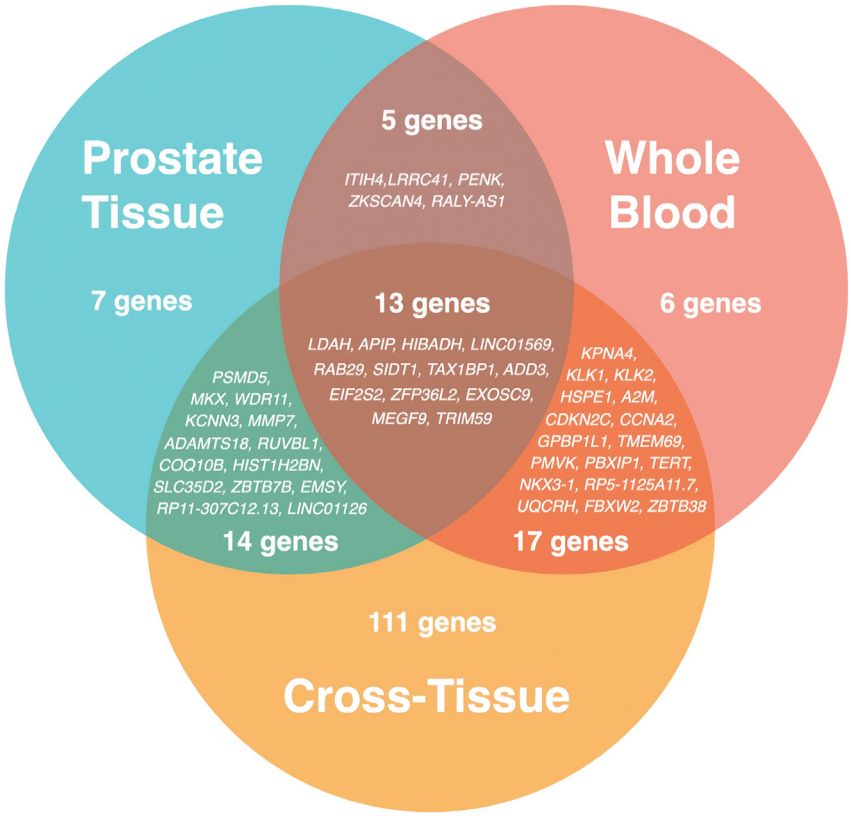

Figure 4. Statistically significant genes overlapping across transcriptome-wide association studies based on whole blood, prostate tissue, and cross-tissue predicted gene expression

Of the significant gene sets from whole blood (41 genes), prostate tissue (39 genes), and cross-tissue analyses (155 genes), 13 genes were significant across all three analyses, 36 genes were significant in exactly two analyses, and 124 genes were significant in a single analysis.

observed enrichment in four pathways involving both single and double stranded DNA binding mechanisms and regulation by transcription factors, including RNA polymerase II, a core component of the DNA transcription machinery. Moreover, two pathways underlying extracellular matrix and cellular component disassembly were enriched in the cross-tissue gene set. We detected significant enrichment in GO biological pathways underlying the regulation of systemic arterial blood pressure, which was also observed for whole blood TWAS using KEGG.

Discussion

This TWAS of 95,768 PCa-free men identified 173 unique genetically predicted transcripts whose expression levels were significantly associated with PSA. We replicated 151 unique gene expression signals in analyses across all three tissues, and all associations had effect estimates that were in the same direction as those in the discovery analyses. Conditional analyses identified 20 unique susceptibility genes for PSA that did not correspond with a known GWAS signal, 10 of which replicated in the original tissue. Fine-mapping analyses for whole blood and prostate tissue identified five protein-coding genes for PSA levels, four of which had strong evidence of colocalization across multiple tissues: *MMP7*, *ADD3*, *A2M*, and *PVMK*. Colocalization analyses further highlighted five of the 20 genes from conditional analyses: *EXOSC9* and *CCNA2* at 4q27, *HIST1H2BN* at 6p22.1, *RP11-182L21.6* at 10q22.3, and *RP11-327J17.2* at 15q26.2. Two of these genes (*CCNA2* and *HIST1H2BN*) replicated in the same tissue matrix as the discovery analysis in which they were originally identified. Pathway enrichment analyses across three catalogs implicated regulatory pathways related to transcription, cell signaling, and disassembly of cellular and noncellular components.

The colocalized genes introduce interesting hypotheses regarding the genetic mechanisms regulating PSA production. Of particular interest are the signals at the 4q27 locus (*CCNA2* and *EXOSC9*), a genomic region that was not detected in previous GWASs of PSA.[16,17] Researchers have previously investigated an autoimmune-related block on the 4q27 locus and did not find an association with PCa risk.[46–49] Signals at *CCNA2*, which were Bonferroni-significant in whole blood and cross-tissue matrices, replicated in additional TWAS analyses for both tissues. Literature documenting a relationship between *CCNA2* and PSA is limited, although *CCNA2* was previously tested as a candidate autoantibody signature marker for distinguishing PCa from BPH in patients with elevated serum PSA;[50] it was not determined to be a top antibody signature of any specific PCa targets. Gene co-expression network analyses have linked *CCNA2* to biochemical recurrence and survival in men with PCa.[51,52] Further exploration of *CCNA2* expression in individuals without PCa may unveil biological pathways that influence PSA levels. *EXOSC9*, which encodes a core protein involved in the RNA degradation machinery in humans, was also identified in conditional and colocalization analyses as a putative gene candidate for PSA levels. *EXOSC9* was identified across all three discovery TWAS analyses, though replication analyses did not yield any significant associations in individual tissue or cross-tissue matrices. Currently, there is no literature supporting an association between *EXOSC9* and PSA levels or PCa risk. Genetically predicted expression of *EXOSC9* in whole blood was positively associated with PSA levels, while the opposite direction of effect was observed for expression in prostate tissue. There exists limited tissue-specific

Table 3. Results from pathway analyses based on all genes that were analyzed in the transcriptome-wide association studies of prostate-specific antigen levels based on whole blood, prostate tissue, and cross-tissue predicted gene expression

Database	Pathway	Overlap[a]	*p* value	Genes
Whole blood				
KEGG	renin-angiotensin system	2/18	4.56×10^{-2}	*KLK1*, *KLK2*
Cross-tissue				
KEGG	transcriptional misregulation in cancer	8/48	1.95×10^{-3}	*SLC45A3*, *CDKN2C*, *TRAF1*, *TP53*, *HOXA10*, *HOXA9*, *H3C10*, *CCNA2*
GO molecular function	histone deacetylase binding	8/120	4.48×10^{-4}	*NKX3-1*, *TFAP4*, *SUDS3*, *TP53*, *HOXA10*, *MYOCD*, *ZBTB7B*, *H1-5*
GO molecular function	serine-type endopeptidase activity	9/120	6.58×10^{-4}	*KLK3*, *KLK2*, *KLK1*, *KLK11*, *KLK4*, *KLK5*, *KLK15*, *MMP7*, *KLK8*
GO molecular function	serine-type peptidase activity	9/120	8.65×10^{-4}	*KLK3*, *KLK2*, *KLK1*, *KLK11*, *KLK4*, *KLK5*, *KLK15*, *MMP7*, *KLK8*
GO molecular function	serine hydrolase activity	9/120	8.65×10^{-4}	*KLK3*, *KLK2*, *KLK1*, *KLK11*, *KLK4*, *KLK5*, *KLK15*, *MMP7*, *KLK8*
GO molecular function	DNA-binding transcription activator activity, RNA polymerase II specific	12/120	5.69×10^{-3}	*NKX2-6*, *TBX3*, *EHF*, *TFAP4*, *OTX1*, *TP53*, *HOXA10*, *HOXA9*, *GLIS2*, *SIX5*, *ZBTB7B*, *MAFF*
GO molecular function	DNA-binding transcription activator activity	12/120	5.69×10^{-3}	*NKX2-6*, *TBX3*, *EHF*, *TFAP4*, *OTX1*, *TP53*, *HOXA10*, *HOXA9*, *GLIS2*, *SIX5*, *ZBTB7B*, *MAFF*
GO molecular function	DNA-binding transcription factor binding	11/120	2.18×10^{-3}	*NKX3-1*, *TBX3*, *NUCKS1*, *FAF1*, *ETS2*, *TP53*, *SDR16C5*, *MYOCD*, *PBXIP1*, *HMGA1*, *EXOSC9*
GO molecular function	endopeptidase activity	10/120	2.18×10^{-2}	*KLK3*, *KLK2*, *KLK1*, *KLK11*, *KLK4*, *KLK5*, *KLK15*, *MMP7*, *ADAMTS18*, *KLK8*
GO molecular function	RNA polymerase II-specific DNA-binding transcription factor binding	9/120	3.17×10^{-2}	*NKX3-1*, *TBX3*, *FAF1*, *ETS2*, *TP53*, *SDR16C5*, *MYOCD*, *HMGA1*, *EXOSC9*
GO molecular function	structural constituent of chromatin	4/120	3.20×10^{-2}	*H3C10*, *H1-5*, *H2BC15*, *HMGA1*
GO biological process	stem cell differentiation	11/121	5.06×10^{-3}	*TBX3*, *HNF1B*, *ZFP36L2*, *PHF19*, *DMRTA2*, *JARID2*, *FGFR2*, *TP53*, *MYOCD*, *CDC42*, *A2M*
GO biological process	regulation of DNA replication	7/121	3.34×10^{-2}	*ZBTB38*, *NUCKS1*, *FAF1*, *RUVBL1*, *TP53*, *E2F7*, *CCNA2*
GO biological process	regionalization	11/121	3.34×10^{-2}	*NKX3-1*, *TBX3*, *HNF1B*, *OTX1*, *DMRTA2*, *FGFR2*, *ETS2*, *TP53*, *HOXA10*, *HOXA9*, *SETDB2*
GO biological process	negative regulation of cyclin-dependent protein serine/threonine kinase activity	4/121	3.34×10^{-2}	*CDK5RAP1*, *TFAP4*, *CDKN2C*, *MYOCD*
GO biological process	negative regulation of cyclin-dependent protein kinase activity	4/121	3.34×10^{-2}	*CDK5RAP1*, *TFAP4*, *CDKN2C*, *MYOCD*
GO biological process	positive regulation of histone modification	6/121	3.34×10^{-2}	*PHF19*, *JARID2*, *TP53*, *SDR16C5*, *ZBTB7B*, *H1-5*

[a]Number of imputed genes in our analysis / number of genes in pathway.

knowledge of the physiological and pathological function of *EXOSC9* in non-cancer cell lines.[48,49] It is worth noting that fine-mapping yielded PIP values lower than 0.8 at the 4q27 locus, suggesting that gene expression of *CCNA2* and *EXOSC9* in individual tissues may not be causally linked to PSA levels; the association between pooled gene expression and PSA levels is unknown.

Colocalized gene *HIST1H2BN* was identified in prostate tissue discovery and replication analyses, but not in fine-mapping results. *HIST1H2BN* encodes a component of a core nucleosome histone and has been linked to PCa cell growth and epithelial-mesenchymal transition through upregulated nuclear factor κB (NF-κB)/Rel expression.[53] Activation of the NF-κB pathway can induce activation of the PSA promoter enhancer, even in the absence of androgens, and NF-κB can directly bind to the PSA enhancer in PCa cell lines.[54] To our knowledge, no experiments have been conducted to investigate the relationship between NF-κB and PSA in PCa-free populations.

PSA is used for both PCa detection and monitoring of PCa progression, making it difficult to disentangle the mechanisms underlying our observed associations.

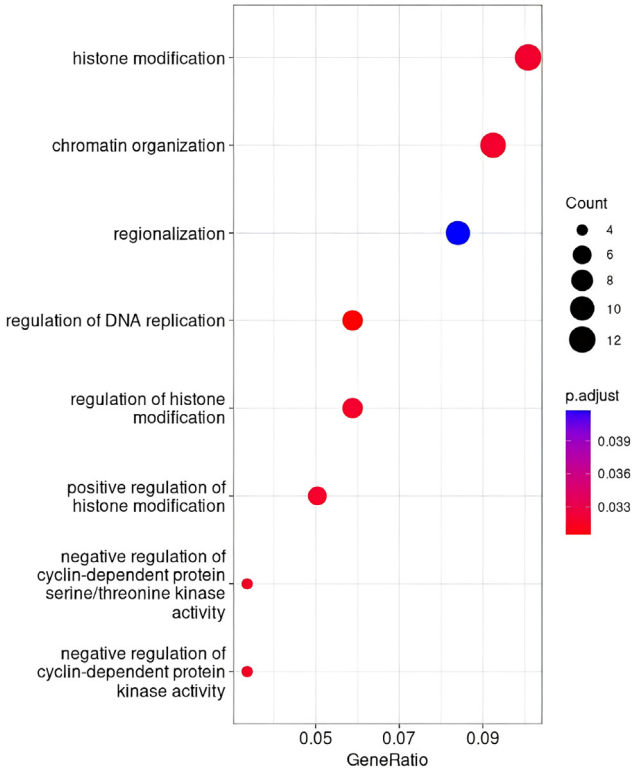

Figure 5. GO term enrichment results from the transcriptome-wide association study of prostate-specific antigen levels based on cross-tissue predicted gene expression

The GO molecular function and biological process data repositories were queried for pathway enrichment analyses across a background set of all genes used in our main analyses. The color gradient represents the magnitude of *p* values, with darker colors corresponding to smaller values. The size of the circles represents the number of genes in the pathway. "p.adjust" represents the *p* value adjusted for multiple testing.

PSA-related screening bias may account for a portion of the observed relationship between gene expression and PSA levels. Genes may also exert pleiotropic effects on PCa and PSA through overlapping biological mechanisms.[2,11] Nevertheless, our restriction to PCa-free men for all analyses minimizes the potential for reverse causation and bolsters confidence that the observed genetic signals inform constitutive PSA levels. In addition, there was no overlap between genes reported in a recent PCa TWAS[55] that used prostate tissue gene expression models and the 20 genes we identified in our conditional analyses, nor for the five genes identified in our colocalization analyses. A comparison between the results of the PCa TWAS and our prostate tissue discovery analyses revealed that 14 of the 39 PSA-associated genes were also significantly associated with PCa risk. None of these genes were determined to be independent from previous GWAS findings identified in Kachuri et al.[16]

Fourteen of the 20 PSA-related genes identified from conditional analyses, spanning 12 genomic regions not implicated by prior PSA GWAS, have been associated with prostatic malignancies in prior literature: *CDC42* (1p36.12), *GPBP1L1* and *TMEM69* (1p34.1), *LRRC41* (1p33), *ITIH4* (3p21.1), *SLBP* (4p16.3), *CCNA2* (4q27), *HIST1H2BN* (6p22.1), *ANXA8L1* (10q11.2), *E2F7* (12q21.2), *C15orf62* (15q15.1), *OTOA* (16p12.2), *C16orf47* (16q22.3), and *HEXIM1* (17q21.31).[53,56–65] Prostate tissue fine-mapping also highlighted *OTOA* as a gene whose expression levels may be causal for PSA. *OTOA* has been identified as a testis-selective gene that shares sequence homology with cancer antigen mesothelin. The remaining six genes identified from conditional analyses have not been implicated in PCa susceptibility in gene function, experimental, or human population research: *UQCRH* (1p33), *ACTRT3* (3q26.2), *EXOSC9* (4q27), *RP11-182L21.6* (10q22.3), *RP11-327J17.2* (15q26.2), and *ACBD4* (17q21.31). *ACTRT3* and *ACBD4* were significantly associated with PSA levels in replication analyses, but were not causally associated with PSA level in fine-mapping analyses. *ACTRT3* is of particular interest, as it is critical in regulating sperm nucleus cytomorphology upstream of the processing of spermatid into mature motile sperm.[66] Increased expression of *ACTRT3*, which forms a testis-specific profilin III-ACTRT3 complex that facilitates male germ cell head cytomorphology and maintains sperm motility in animal models, was associated with elevated PSA levels in the TWAS based on whole blood.[66,67] Though there is no documented link between *ACTRT3* and PSA levels or PCa risk, the pronounced role of *ACTRT3* in mediating conformational changes in sperm nuclei suggests possible shared biological pathways between PSA and the production and processing of male germ cells.[66,68] Loci that map to the multigenic region that contains *ACTRT3* have been linked to colorectal and bladder cancer susceptibility.[69,70]

The function of *ABCD4* is not well understood, although the literature suggests that ATPase activity of the protein it encodes plays an essential role in vitamin B12 (cobalamin) metabolism.[71] *ABCD4* is a member of the ATP-binding cassette transporter family that contributes to intracellular transport of B12 through interactions with LMBD1.[71,72] Elevated plasma levels of circulating B12 have been associated with an up to three-fold increased PCa risk,[73,74] although an analysis of B12 and changes in post-diagnostic PSA levels was not suggestive of an association.[75] Associations between serum levels of B12 and PSA levels have not been evaluated in PCa-free men. It must be noted that 17q21.31 includes a ~900 kB chromosomal inversion polymorphism that is prevalent in approximately 20% of European ancestry individuals.[76] This chromosomal inversion is known to have a tissue-specific effect on the expression of multiple genes and is linked to gene expression differences in neurological disease and breast carcinogenesis.[77–79] As such, our results at this locus should be interpreted with caution.

Individual-tissue fine-mapping identified five protein-coding genes from our discovery TWAS that were causal for PSA levels. None of these genes have been evaluated in relation to PSA levels in PCa-free men, but *MMP7* and *A2M* have been previously linked to PSA overall. Serum

levels of *MMP7*, which were the strongest signals in our discovery prostate tissue TWAS, have been linked to large PSA declines in men with metastatic hormone-sensitive PCa who were undergoing chemotherapy treatment.[80] *MMP7* was also significantly associated with PCa risk in a recent TWAS.[55] *A2M*, which was significantly and causally associated with PSA in whole blood, is also inversely correlated with PSA levels in men with PCa.[81] The alpha-2-macroglobulin (A2M) protein functions as an enzymatic inhibitor of PSA in blood; PSA interacts with A2M and forms inactive PSA-A2M complexes.[82] The remaining three genes identified by fine-mapping have not been linked to PSA or PCa. *ADD3* encodes the cytoskeleton protein adducin and is located in a known tumor suppressor region for glioblastomas.[83] *PMVK* encodes the phosphomevalonate kinase protein, which catalyzes a key step in the mevalonate pathway, but its function with respect to cancer development and progression is unknown.[84,85] *OTOA* is primarily implicated in normal hearing and sound transduction.[86,87] Fusion gene annotations using FusionGDB 2.0[88] have found that *OTOA* can form a fusion gene with *NCOA4* (FusionGDB2 ID: 61935). *NCOA4* encodes an androgen receptor co-activator that potentially forms a tripartite complex with PSA and the androgen receptor, highlighting a potential mechanism through which *OTOA* can be linked to PSA levels.[89,90]

In general, fine-mapping analyses for whole blood and prostate tissue yielded fewer signals than our primary TWAS, and only five of 30 genes with a PIP of greater than 0.8 overlapped with TWAS hits. cTWAS does not currently allow for evaluation of results from cross-tissue analyses, whereas the majority of signals from our discovery TWAS were from cross-tissue analyses. It is also possible that pooled gene expression across multiple tissue matrices exerts an effect on PSA levels in PCa-free men. Moreover, cTWAS has comparatively low power to detect associations compared to other methods for integrating GWAS data and eQTL results. Associations that are causal may not be captured by cTWAS due to the conservative significance threshold, particularly in gene sets with a low proportion of variance explained.[34] Additionally, preprocessing using the cTWAS analytical pipeline restricted gene expression models to protein-coding genes. Regardless, our fine-mapping results suggest that results from TWAS must be interpreted with caution.

Our study had several limitations. First, nearly 90% of men in the study population were of primarily European ancestry,[16] and 85% of the tissue samples used to train the TWAS models were derived from European populations.[44] Investigations into ancestry differences in genetic regulatory effects suggest that for variants that are common (minor-allele frequency of >1%) across populations, the majority of eQTL signals tend to be shared with little evidence of heterogeneity in cross-population effect sizes.[44,91,92] However, our findings may not be generalizable to broader ancestral populations for rarer variants and further investigation in multi-ethnic populations are still necessary to equitably improve PSA screening. Second, our study assessed only *cis*-eQTLs, so any *trans*-eQTL effects were not incorporated.[93,94] Third, probabilistic colocalization methods have a high type two error rate and may be underpowered to detect shared association signals.[26] As such, a lack of colocalization signal at a particular locus does not rule out the presence of a true signal and may instead reflect analytical factors like LD mismatch.[38] Fourth, although we restricted our analyses to men who had not been diagnosed with PCa, we cannot rule out the possibility of latent, undiagnosed disease or disease diagnosed at a later time point. However, the prevalence of undiagnosed PCa in our population was likely to be low on account of increased monitoring and surveillance in longitudinal cohorts.[9,11] Fifth, we were unable to impute and run analyses conditional on the chromosomal inversion at 17q21.31 due to the summary-level nature of our available GWAS data. Finally, many of the genes that we identified were clustered at multi-gene loci, in part due to co-regulation by a shared set of eQTLs. Sentinel genes at these loci should be interpreted with caution, as there may be correlated predicted expression.[95]

Our study also had several key strengths. The use of GWAS summary statistics from a cohort of 95,768 men provided us with high statistical power to detect PSA-associated genes. In addition to conducting tissue-specific association analyses in tissues that are biologically meaningful for PSA, we integrated association signals across 45 GTEx tissues to improve power for genes with similar regulatory mechanisms across tissues. TWAS is one of several approaches for connecting GWAS variants to target genes. Although other approaches like locus-to-gene fine-mapping[96] and combined variant-to-gene heritability frameworks[97] are powerful for systematically integrating functional data for gene prioritization, they do not restrict to disease-specific tissues and cell types. Our analysis better captures gene expression levels for PSA-relevant tissue matrices and incorporates similar gene-prioritization methods via colocalization. Our study also used COJO analysis to find genes conditionally independent from known GWAS variants. Last, the study results were validated in an independent cohort of more than 200,000 MVP participants.

In summary, our TWASs identified gene expression profiles associated with PSA levels in men without PCa. These findings provide several hypotheses for genes that affect constitutive PSA. Further exploration of these results, including functional analyses of these genes in *in vivo* settings, will augment our understanding of the genetic etiology of PSA variation. Transcriptomic analyses might also be vertically integrated with downstream -omic approaches to uncover complete mechanisms through which genetics influence circulating PSA levels. In addition, TWAS findings may be used to develop polygenic transcriptome risk scores[98] for PSA levels, which could be leveraged for improving PSA as a screening tool for PCa.

Data and code availability

The GWAS summary statistics from Kachuri et al.[16] used in this analysis are available for download from the following Zenodo repository: https://doi.org/10.5281/zenodo.7460134. Transcriptome prediction weights and models used for MetaXcan analyses are available from: https://predictdb.org.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2024.100315.

Acknowledgments

The Precision PSA study is supported by funding from the National Institutes of Health (NIH) and National Cancer Institute (NCI) under award number R01CA241410 (PI: J.S.W.). This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing and Data at the Icahn School of Medicine at Mount Sinai and supported by the Clinical and Translational Science Awards (CTSA) grant UL1TR004419 from the National Center for Advancing Translational Sciences. Additionally, this research was supported by the Intramural Research Program of the National Cancer Institute. R.E.G. is supported by a Prostate Cancer Foundation Young Investigator Award. L.K. is supported by funding from the National Cancer Institute (R00CA246076). J.P.S. is supported by funding from the National Institute of General Medical Sciences (T32GM007347). K.R.S. is supported by funding from the NIH (2K12CA090625). H.L. is supported in part by funding from NIH/NCI(P30-CA008748, U01-CA199338, R01CA244948), the Swedish Cancer Society (Cancerfonden 23 3074 Pj 01 H), and the Swedish Prostate Cancer Foundation (Prostatacancerförbundet). R.K. is supported by funding from the NIH (R01 CA244948 and R01 CA175491). Research reported in this publication was also supported by the Office of Research Infrastructure of the National Institutes of Health under award numbers S10OD026880 and S10OD030463. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Declaration of interests

J.S.W. is a non-employee and cofounder of Avail Bio. H.L. is named on a patent for assays to measure intact PSA and a patent for a statistical method to detect prostate cancer commercialized by OPKO Health (4KScore). H.L. receives royalties from sales of the assay and has stock in OPKO Health. H.L. serves on the Scientific Advisory Board for Fujirebio Diagnostics Inc and owns stock in Diaprost AB and Acousort AB. R.E.G. consults for Hunton Andrews Kurth LLC on subject matter unrelated to this study.

Received: January 17, 2024
Accepted: June 3, 2024

Web resources

https://doi.org/10.5281/zenodo.7460134.
https://predictdb.org.

References

1. Lilja, H. (1985). A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. J. Clin. Invest. *76*, 1899–1903.
2. Lilja, H., Ulmert, D., and Vickers, A.J. (2008). Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat. Rev. Cancer *8*, 268–278.
3. Cucchiara, V., Cooperberg, M.R., Dall'Era, M., Lin, D.W., Montorsi, F., Schalken, J.A., and Evans, C.P. (2018). Genomic Markers in Prostate Cancer Decision Making. Eur. Urol. *73*, 572–582.
4. LeBeau, A.M., Kostova, M., Craik, C.S., and Denmeade, S.R. (2010). Prostate-specific antigen: an overlooked candidate for the targeted treatment and selective imaging of prostate cancer. Biol. Chem. *391*, 333–343.
5. Shi, W.X., and Bunney, B.S. (1990). A small volume chamber for electrical recording from submerged brain slices and a pulse-free medium supply system using a peristalic pump. J. Neurosci. Methods *35*, 235–240.
6. Qiu, S.D., Young, C.Y., Bilhartz, D.L., Prescott, J.L., Farrow, G.M., He, W.W., and Tindall, D.J. (1990). In situ hybridization of prostate-specific antigen mRNA in human prostate. J. Urol. *144*, 1550–1556.
7. Sävblom, C., Halldén, C., Cronin, A.M., Säll, T., Savage, C., Vertosick, E.A., Klein, R.J., Giwercman, A., and Lilja, H. (2014). Genetic variation in KLK2 and KLK3 is associated with concentrations of hK2 and PSA in serum and seminal plasma in young men. Clin. Chem. *60*, 490–499.
8. Pinsky, P.F., Kramer, B.S., Crawford, E.D., Grubb, R.L., Urban, D.A., Andriole, G.L., Chia, D., Levin, D.L., and Gohagan, J.K. (2006). Prostate volume and prostate-specific antigen levels in men enrolled in a large screening trial. Urology *68*, 352–356.
9. Ilic, D., Neuberger, M.M., Djulbegovic, M., and Dahm, P. (2013). Screening for prostate cancer. Cochrane Database Syst. Rev. *2013*, CD004720.
10. Fenton, J.J., Weyrich, M.S., Durbin, S., Liu, Y., Bang, H., and Melnikow, J. (2018). Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA *319*, 1914–1931.
11. Ilic, D., Djulbegovic, M., Jung, J.H., Hwang, E.C., Zhou, Q., Cleves, A., Agoritsas, T., and Dahm, P. (2018). Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. BMJ *362*, k3519.
12. Potosky, A.L., Feuer, E.J., and Levin, D.L. (2001). Impact of screening on incidence and mortality of prostate cancer in the United States. Epidemiol. Rev. *23*, 181–186.
13. Han, P.K.J., Kobrin, S., Breen, N., Joseph, D.A., Li, J., Frosch, D.L., and Klabunde, C.N. (2013). National Evidence on the Use of Shared Decision Making in Prostate-Specific Antigen Screening. Ann. Fam. Med. *11*, 306–314.
14. Hugosson, J., Roobol, M.J., Månsson, M., Tammela, T.L.J., Zappa, M., Nelen, V., Kwiatkowski, M., Lujan, M., Carlsson, S.V., Talala, K.M., et al. (2019). A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. Eur. Urol. *76*, 43–51.
15. de Koning, H.J., Gulati, R., Moss, S.M., Hugosson, J., Pinsky, P.F., Berg, C.D., Auvinen, A., Andriole, G.L., Roobol, M.J., Crawford, E.D., et al. (2018). The efficacy of prostate-specific antigen screening: Impact of key components in the ERSPC and PLCO trials. Cancer *124*, 1197–1206.

16. Kachuri, L., Hoffmann, T.J., Jiang, Y., Berndt, S.I., Shelley, J.P., Schaffer, K.R., Machiela, M.J., Freedman, N.D., Huang, W.Y., Li, S.A., et al. (2023). Genetically adjusted PSA levels for prostate cancer screening. Nat. Med. *29*, 1412–1423.
17. Hoffmann, T.J., Passarelli, M.N., Graff, R.E., Emami, N.C., Sakoda, L.C., Jorgenson, E., Habel, L.A., Shan, J., Ranatunga, D.K., Quesenberry, C.P., et al. (2017). Genome-wide association study of prostate-specific antigen levels identifies novel loci independent of prostate cancer. Nat. Commun. *8*, 14248.
18. Hoffmann, T.J., Graff, R.E., Madduri, R.K., Rodriguez, A.A., Cario, C.L., Feng, K., Jiang, Y., Wang, A., Klein, R.J., Pierce, B.L., et al. (2023). Genome-wide association study of prostate-specific antigen levels in 392,522 men identifies new loci and improves cross-ancestry prediction. Preprint at medRxiv. https://doi.org/10.1101/2023.10.27.23297676.
19. Bansal, A., Murray, D.K., Wu, J.T., Stephenson, R.A., Middleton, R.G., and Meikle, A.W. (2000). Heritability of Prostate-Specific Antigen and Relationship with Zonal Prostate Volumes in Aging Twins. J. Clin. Endocrinol. Metab. *85*, 1272–1276.
20. Pilia, G., Chen, W.M., Scuteri, A., Orrú, M., Albai, G., Dei, M., Lai, S., Usala, G., Lai, M., Loi, P., et al. (2006). Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. PLoS Genet. *2*, e132.
21. Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., and Cox, N.J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. *6*, e1000888.
22. Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., GTEx Consortium, and Nicolae, D.L., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. *47*, 1091–1098.
23. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
24. Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. *18*, 117–127.
25. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat. Commun. *9*, 1825.
26. Hukku, A., Sampson, M.G., Luca, F., Pique-Regi, R., and Wen, X. (2022). Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. Am. J. Hum. Genet. *109*, 825–837.
27. D'Amico, A.V., Whittington, R., Malkowicz, S.B., Schultz, D., Blank, K., Broderick, G.A., Tomaszewski, J.E., Renshaw, A.A., Kaplan, I., Beard, C.J., and Wein, A. (1998). Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. JAMA *280*, 969–974.
28. Hernandez, D.J., Nielsen, M.E., Han, M., and Partin, A.W. (2007). Contemporary evaluation of the D'amico risk classification of prostate cancer. Urology *70*, 931–935.
29. Vessella, R.L. (1993). Trends in immunoassays of prostate-specific antigen: serum complexes and ultrasensitivity. Clin. Chem. *39*, 2035–2039.
30. Junker, R., Brandt, B., Semjonow, A., Erren, M., Zechel, C., and Assmann, G. (1999). The biologic lower detection limit of six ultrasensitive PSA assays. Anticancer Res. *19*, 2625–2628.
31. Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., et al. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J. Clin. Epidemiol. *70*, 214–223.
32. Klarin, D., Damrauer, S.M., Cho, K., Sun, Y.V., Teslovich, T.M., Honerlaw, J., Gagnon, D.R., DuVall, S.L., Li, J., Peloso, G.M., et al. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. *50*, 1514–1523.
33. Yang, J., Ferreira, T., Morris, A.P., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics idntifies additional variants influencing complex traits. Nat. Genetics *44*, 369–375. S1-3.
34. Zhao, S., Crouse, W., Qian, S., Luo, K., Stephens, M., and He, X. (2024). Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. Nat. Genet. *56*, 336–347.
35. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. J. R. Stat. Soc. Ser B Stat. Methodol. *82*, 1273–1300.
36. Zou, Y., Carbonetto, P., Wang, G., and Stephens, M. (2022). Fine-mapping from summary data with the "Sum of Single Effects" model. PLoS Genet. *18*, e1010299.
37. Wen, X., Pique-Regi, R., and Luca, F. (2017). Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLoS Genet. *13*, e1006646.
38. Hukku, A., Pividori, M., Luca, F., Pique-Regi, R., Im, H.K., and Wen, X. (2021). Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. Am. J. Hum. Genet. *108*, 25–35.
39. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics *32*, 283–285.
40. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.
41. The Gene Ontology Consortium (2019). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. *47*, D330–D338.
42. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinf. *14*, 128.
43. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr. Protoc. *1*, e90.
44. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.
45. Balk, S.P., Ko, Y.J., and Bubley, G.J. (2003). Biology of Prostate-Specific Antigen. J. Clin. Oncol. *21*, 383–391.
46. Tindall, E.A., Hoang, H.N., Southey, M.C., English, D.R., Hopper, J.L., Giles, G.G., Severi, G., and Hayes, V.M. (2010). The 4q27 locus and prostate cancer risk. BMC Cancer *10*, 69.

47. Yoshino, S., Hara, T., Weng, J.S., Takahashi, Y., Seiki, M., and Sakamoto, T. (2012). Genetic Screening of New Genes Responsible for Cellular Adaptation to Hypoxia Using a Genome-Wide shRNA Library. PLoS One *7*, e35590.
48. Yoshino, S., Matsui, Y., Fukui, Y., Seki, M., Yamaguchi, K., Kanamori, A., Saitoh, Y., Shimamura, T., Suzuki, Y., Furukawa, Y., et al. (2020). EXOSC9 depletion attenuates P-body formation, stress resistance, and tumorigenicity of cancer cells. Sci. Rep. *10*, 9275.
49. Burns, D.T., Donkervoort, S., Müller, J.S., Knierim, E., Bharucha-Goebel, D., Faqeih, E.A., Bell, S.K., AlFaifi, A.Y., Monies, D., Millan, F., et al. (2018). Variants in EXOSC9 Disrupt the RNA Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. Am. J. Hum. Genet. *102*, 858–873.
50. O'Rourke, D.J., DiJohnson, D.A., Caiazzo, R.J., Nelson, J.C., Ure, D., O'Leary, M.P., Richie, J.P., and Liu, B.C.S. (2012). Autoantibody Signatures as Biomarkers to Distinguish Prostate Cancer from Benign Prostatic Hyperplasia in Patients with Increased Serum Prostate Specific Antigen. Clin. Chim. Acta *413*, 561–567.
51. Feng, T., Wei, D., Li, Q., Yang, X., Han, Y., Luo, Y., and Jiang, Y. (2021). Four Novel Prognostic Genes Related to Prostate Cancer Identified Using Co-expression Structure Network Analysis. Front. Genet. *12*, 584164.
52. Yang, R., Du, Y., Wang, L., Chen, Z., and Liu, X. (2020). Weighted gene co-expression network analysis identifies CCNA2 as a treatment target of prostate cancer through inhibiting cell cycle. J. Cancer *11*, 1203–1211.
53. Zhang, J., Chang, Y., Xia, H., Xu, L., and Wei, X. (2021). HIST1H2BN induced cell proliferation and EMT phenotype in prostate cancer via NF-κB signal pathway. Genes Genomics *43*, 1361–1369.
54. Chen, C.D., and Sawyers, C.L. (2002). NF-κB Activates Prostate-Specific Antigen Expression and Is Upregulated in Androgen-Independent Prostate Cancer. Mol. Cell Biol. *22*, 2862–2870.
55. Liu, D., Zhu, J., Zhou, D., Nikas, E.G., Mitanis, N.T., Sun, Y., Wu, C., Mancuso, N., Cox, N.J., Wang, L., et al. (2022). A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. Int. J. Cancer *150*, 80–90.
56. Liu, Q., Reed, M., Zhu, H., Cheng, Y., Almeida, J., Fruhbeck, G., Ribeiro, R., and Hu, P. (2022). Epigenome-wide DNA methylation and transcriptome profiling of localized and locally advanced prostate cancer: Uncovering new molecular markers. Genomics *114*, 110474.
57. Pressinotti, N.C., Klocker, H., Schäfer, G., Luu, V.D., Ruschhaupt, M., Kuner, R., Steiner, E., Poustka, A., Bartsch, G., and Sültmann, H. (2009). Differential expression of apoptotic genes PDIA3 and MAP3K5 distinguishes between low- and high-risk prostate cancer. Mol. Cancer *8*, 130.
58. Mahajan, N.P., Liu, Y., Majumder, S., Warren, M.R., Parker, C.E., Mohler, J.L., Earp, H.S., and Whang, Y.E. (2007). Activated Cdc42-associated kinase Ack1 promotes prostate cancer progression via androgen receptor tyrosine phosphorylation. Proc. Natl. Acad. Sci. USA *104*, 8438–8443.
59. Davalieva, K., Kiprijanovska, S., Komina, S., Petrusevska, G., Zografska, N.C., and Polenakovic, M. (2015). Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic biomarkers for prostate cancer. Proteome Sci. *13*, 2.
60. Krishnan, N., Titus, M.A., and Thapar, R. (2014). The Prolyl Isomerase Pin1 Regulates mRNA Levels of Genes with Short Half-Lives by Targeting Specific RNA Binding Proteins. PLoS One *9*, e85427.
61. Wang, Y., Pei, X., Xu, P., Tan, Z., Zhu, Z., Zhang, G., Jiang, Z., and Deng, Z. (2020). E2F7, regulated by miR-30c, inhibits apoptosis and promotes cell cycle of prostate cancer cells. Oncol. Rep. *44*, 849–862.
62. Sheeba, J.D.J., Hegde, S., Tamboli, N., Nadig, N., Keshavamurthy, R., and Ranganathan, P. (2024). Gene Expression Signature of Castrate Resistant Prostate Cancer. Gene *925*, 148603.
63. Grasso, C.S., Wu, Y.M., Robinson, D.R., Cao, X., Dhanasekaran, S.M., Khan, A.P., Quist, M.J., Jing, X., Lonigro, R.J., Brenner, J.C., et al. (2012). The mutational landscape of lethal castration-resistant prostate cancer. Nature *487*, 239–243.
64. Lew, Q.J., Chu, K.L., Chia, Y.L., Cheong, N., and Chao, S.H. (2013). HEXIM1, a New Player in the p53 Pathway. Cancers *5*, 838–856.
65. Chen, R., Yik, J.H.N., Lew, Q.J., and Chao, S.H. (2014). Brd4 and HEXIM1: Multiple Roles in P-TEFb Regulation and Cancer. BioMed Res. Int. *2014*, 1–11.
66. Hara, Y., Yamagata, K., Oguchi, K., and Baba, T. (2008). Nuclear localization of profilin III-ArpM1 complex in mouse spermiogenesis. FEBS Lett. *582*, 2998–3004.
67. Umer, N., Arévalo, L., Phadke, S., Lohanadan, K., Kirfel, G., Sons, D., Sofia, D., Witke, W., and Schorle, H. (2021). Loss of Profilin3 Impairs Spermiogenesis by Affecting Acrosome Biogenesis, Autophagy, Manchette Development and Mitochondrial Organization. Front. Cell Dev. Biol. *9*, 749559.
68. Lilja, H., Oldbring, J., Rannevik, G., and Laurell, C.B. (1987). Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. J. Clin. Invest. *80*, 281–285.
69. Law, P.J., Timofeeva, M., Fernandez-Rozadilla, C., Broderick, P., Studd, J., Fernandez-Tajes, J., Farrington, S., Svinti, V., Palles, C., Orlando, G., et al. (2019). Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat. Commun. *10*, 2154.
70. Figueroa, J.D., Ye, Y., Siddiq, A., Garcia-Closas, M., Chatterjee, N., Prokunina-Olsson, L., Cortessis, V.K., Kooperberg, C., Cussenot, O., Benhamou, S., et al. (2014). Genome-wide association study identifies multiple loci associated with bladder cancer risk. Hum. Mol. Genet. *23*, 1387–1398.
71. Coelho, D., Kim, J.C., Miousse, I.R., Fung, S., du Moulin, M., Buers, I., Suormala, T., Burda, P., Frapolli, M., Stucki, M., et al. (2012). Mutations in ABCD4 cause a new inborn error of vitamin B12 metabolism. Nat. Genet. *44*, 1152–1155.
72. Shani, N., Jimenez-Sanchez, G., Steel, G., Dean, M., and Valle, D. (1997). Identification of a Fourth Half ABC Transporter in the Human Peroxisomal Membrane. Hum. Mol. Genet. *6*, 1925–1931.
73. Collin, S.M., Metcalfe, C., Refsum, H., Lewis, S.J., Zuccolo, L., Smith, G.D., Chen, L., Harris, R., Davis, M., Marsden, G., et al. (2010). Circulating folate, vitamin B12, homocysteine, vitamin B12 transport proteins, and risk of prostate cancer: a case-control study, systematic review, and meta-analysis. Cancer Epidemiol. Biomarkers Prev. *19*, 1632–1642.
74. Hultdin, J., Van Guelpen, B., Bergh, A., Hallmans, G., and Stattin, P. (2005). Plasma folate, vitamin B12, and homocysteine and prostate cancer risk: a prospective study. Int. J. Cancer *113*, 819–824.
75. Collin, S.M., Metcalfe, C., Refsum, H., Lewis, S.J., Smith, G.D., Cox, A., Davis, M., Marsden, G., Johnston, C., Lane, J.A., et al. (2010). Associations of Folate, Vitamin B12, Homocysteine,

and Folate-Pathway Polymorphisms with Prostate-Specific Antigen Velocity in Men with Localized Prostate Cancer. Cancer Epidemiol. Biomarkers Prev. *19*, 2833–2838.
76. Donnelly, M.P., Paschou, P., Grigorenko, E., Gurwitz, D., Mehdi, S.Q., Kajuna, S.L.B., Barta, C., Kungulilo, S., Karoma, N.J., Lu, R.B., et al. (2010). The Distribution and Most Recent Common Ancestor of the 17q21 Inversion in Humans. Am. J. Hum. Genet. *86*, 161–171.
77. de Jong, S., Chepelev, I., Janson, E., Strengman, E., van den Berg, L.H., Veldink, J.H., and Ophoff, R.A. (2012). Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. BMC Genom. *13*, 458.
78. Wang, H., Makowski, C., Zhang, Y., Qi, A., Kaufmann, T., Smeland, O.B., Fiecas, M., Yang, J., Visscher, P.M., and Chen, C.H. (2023). Chromosomal inversion polymorphisms shape human brain morphology. Cell Rep. *42*, 112896.
79. Cole, N., Lee, P., Schwarz, T., Zhang, P., Freedman, M.L., Gusev, A., Lindström, S., Gandal, M.J., Pasaniuc, B., Bhattacharya, A., et al. (2022). Distal gene regulation mediated by non-coding RNAs contributes to germline risk for breast and prostate cancer. Preprint at medRxiv. https://doi.org/10.1101/2022.02.08.22270601.
80. Szarvas, T., Csizmarik, A., Váradi, M., Fazekas, T., Hüttl, A., Nyirády, P., Hadaschik, B., Grünwald, V., Tschirdewahn, S., Shariat, S.F., et al. (2021). The prognostic value of serum MMP-7 levels in prostate cancer patients who received docetaxel, abiraterone, or enzalutamide therapy. Urol. Oncol. Semin. Orig. Investig. *39*, 296.e11–296.e19.
81. Kostova, M.B., Brennen, W.N., Lopez, D., Anthony, L., Wang, H., Platz, E., and Denmeade, S.R. (2018). PSA-alpha-2-macroglobulin complex is enzymatically active in the serum of patients with advanced prostate cancer and can degrade circulating peptide hormones. Prostate *78*, 819–829.
82. Lilja, H., Christensson, A., Dahlén, U., Matikainen, M.T., Nilsson, O., Pettersson, K., and Lövgren, T. (1991). Prostate-specific antigen in serum occurs predominantly in complex with alpha 1-antichymotrypsin. Clin. Chem. *37*, 1618–1625.
83. Kiang, K.M.Y., Sun, S., and Leung, G.K.K. (2021). ADD3 Deletion in Glioblastoma Predicts Disease Status and Survival. Front. Oncol. *11*, 717793.
84. Chen, Z., Zhou, X., Zhou, X., Tang, Y., Lu, M., Zhao, J., Tian, C., Wu, M., Liu, Y., Prochownik, E.V., et al. (2023). Phosphomevalonate Kinase Controls β-Catenin Signaling via the Metabolite 5-Diphosphomevalonate. Adv. Sci. *10*, 2204909.
85. Clendening, J.W., Pandyra, A., Boutros, P.C., El Ghamrasni, S., Khosravi, F., Trentin, G.A., Martirosyan, A., Hakem, A., Hakem, R., Jurisica, I., and Penn, L.Z. (2010). Dysregulation of the mevalonate pathway promotes transformation. Proc. Natl. Acad. Sci. USA *107*, 15051–15056.
86. Zazo Seco, C., Wesdorp, M., Feenstra, I., Pfundt, R., Hehir-Kwa, J.Y., Lelieveld, S.H., Castelein, S., Gilissen, C., de Wijs, I.J., Admiraal, R.J., et al. (2017). The diagnostic yield of whole-exome sequencing targeting a gene panel for hearing impairment in The Netherlands. Eur. J. Hum. Genet. *25*, 308–314.
87. Mehta, D., Noon, S.E., Schwartz, E., Wilkens, A., Bedoukian, E.C., Scarano, I., Crenshaw, E.B., 3rd, and Krantz, I.D. (2016). Outcomes of evaluation and testing of 660 individuals with hearing loss in a pediatric genetics of hearing loss clinic. Am. J. Med. Genet. *170*, 2523–2530.
88. Kim, P., Tan, H., Liu, J., Lee, H., Jung, H., Kumar, H., and Zhou, X. (2022). FusionGDB 2.0: fusion gene annotation updates aided by deep learning. Nucleic Acids Res. *50*, D1221–D1230.
89. Niu, Y., Yeh, S., Miyamoto, H., Li, G., Altuwaijri, S., Yuan, J., Han, R., Ma, T., Kuo, H.C., and Chang, C. (2008). Tissue prostate-specific antigen facilitates refractory prostate tumor progression via enhancing ARA70-regulated androgen receptor transactivation. Cancer Res. *68*, 7110–7119.
90. Kollara, A., and Brown, T.J. (2012). Expression and function of nuclear receptor co-activator 4: evidence of a potential role independent of co-activator activity. Cell. Mol. Life Sci. *69*, 3895–3909.
91. Gay, N.R., Gloudemans, M., Antonio, M.L., Abell, N.S., Balliu, B., Park, Y., Martin, A.R., Musharoff, S., Rao, A.S., Aguet, F., et al. (2020). Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. Genome Biol. *21*, 233.
92. Shang, L., Smith, J.A., Zhao, W., Kho, M., Turner, S.T., Mosley, T.H., Kardia, S.L.R., and Zhou, X. (2020). Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. Am. J. Hum. Genet. *106*, 496–512.
93. Luningham, J.M., Chen, J., Tang, S., De Jager, P.L., Bennett, D.A., Buchman, A.S., and Yang, J. (2020). Bayesian Genome-wide TWAS Method to Leverage both cis- and trans-eQTL Information through Summary Statistics. Am. J. Hum. Genet. *107*, 714–726.
94. Bhattacharya, A., Li, Y., and Love, M.I. (2021). MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. PLoS Genet. *17*, e1009398.
95. Wainberg, M., Sinnott-Armstrong, N., Mancuso, N., Barbeira, A.N., Knowles, D.A., Golan, D., Ermel, R., Ruusalepp, A., Quertermous, T., Hao, K., et al. (2019). Opportunities and challenges for transcriptome-wide association studies. Nat. Genet. *51*, 592–599.
96. Mountjoy, E., Schmidt, E.M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., Fumis, L., Hayhurst, J., Buniello, A., Karim, M.A., et al. (2021). An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat. Genet. *53*, 1527–1533.
97. Gazal, S., Weissbrod, O., Hormozdiari, F., Dey, K.K., Nasser, J., Jagadeesh, K.A., Weiner, D.J., Shi, H., Fulco, C.P., O'Connor, L.J., et al. (2022). Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. Nat. Genet. *54*, 827–836.
98. Liang, Y., Pividori, M., Manichaikul, A., Palmer, A.A., Cox, N.J., Wheeler, H.E., and Im, H.K. (2022). Polygenic transcriptome risk scores (PTRS) can improve portability of polygenic risk scores across ancestries. Genome Biol. *23*, 23.