

1 Transcriptome-Wide Association Analysis Identifies Novel Candidate Susceptibility Genes for
2 Prostate-Specific Antigen Levels in Men Without Prostate Cancer

3 **Author List:**

4 Dorothy M. Chen^{1*}, Ruocheng Dong^{2*}, Linda Kachuri^{2,3}, Thomas Hoffmann^{1,4}, Yu Jiang¹,
5 Sonja I. Berndt⁵, John P. Shelley⁶, Kerry R. Schaffer^{7,12}, Mitchell J. Machiela⁵, Neal D.
6 Freedman⁵, Wen-Yi Huang⁵, Shengchao A. Li⁵, Hans Lilja^{8,9}, Stephen K. Van Den Eeden¹⁰,
7 Stephen Chanock⁵, Christopher A. Haiman^{11,12}, David V. Conti^{11,12}, Robert J. Klein¹³, Jonathan
8 D. Mosley¹⁴, John S. Witte^{1,2,15}†, Rebecca E. Graff¹†

9 **Affiliations:**

10 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San
11 Francisco, CA, 94158, USA

12 2) Department of Epidemiology and Population Health, Stanford University, Stanford, CA,
13 94305, USA

14 3) Stanford Cancer Institute, Stanford University, Stanford, CA, 94305, USA

15 4) Institute for Human Genetics, University of California San Francisco, San Francisco, CA,
16 94143, USA

17 5) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD,
18 20814, USA

19 6) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN,
20 37232, USA

21 7) Division of Hematology and Oncology, Vanderbilt University Medical Center, Nashville, TN,
22 37232, USA

- 23 8) Departments of Pathology and Laboratory Medicine, Surgery, Medicine, Memorial Sloan
24 Kettering Cancer Center, New York, NY, 10065, USA
- 25 9) Department of Translational Medicine, Lund University, Malmö, 21428, Sweden
- 26 10) Division of Research, Kaiser Permanente Northern California, Oakland, CA, 94612, USA
- 27 11) Center for Genetic Epidemiology, Department of Population and Preventive Health Sciences,
28 Keck School of Medicine, University of Southern California, Los Angeles, CA, 90032, USA
- 29 12) Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern
30 California, Los Angeles, CA, 90033, USA
- 31 13) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai,
32 New York, NY, 10029, USA
- 33 14) Departments of Internal Medicine and Biomedical Informatics, Vanderbilt University
34 Medical Center, Nashville, TN, 37232, USA
- 35 15) Departments of Biomedical Data Science and Genetics (by courtesy), Stanford University,
36 Stanford, CA, 94305, USA
- 37 * These authors contributed equally to this work
- 38 † These authors jointly supervised this work
- 39
- 40 * Correspondence: jswithe@stanford.edu and Rebecca.Graff@ucsf.edu

41 **Abstract**

42 Deciphering the genetic basis of prostate-specific antigen (PSA) levels may improve their utility
43 to screen for prostate cancer (PCa). We thus conducted a transcriptome-wide association study
44 (TWAS) of PSA levels using genome-wide summary statistics from 95,768 PCa-free men, the
45 MetaXcan framework, and gene prediction models trained in Genotype-Tissue Expression
46 (GTEx) project data. Tissue-specific analyses identified 41 statistically significant ($p <$
47 $0.05/12,192 = 4.10e-6$) associations in whole blood and 39 statistically significant ($p <$
48 $0.05/13,844 = 3.61e-6$) associations in prostate tissue, with 18 genes associated in both tissues.
49 Cross-tissue analyses that combined associations across 45 tissues identified 155 genes that were
50 statistically significantly ($p < 0.05/22,249 = 2.25e-6$) associated with PSA levels. Based on
51 conditional analyses that assessed whether TWAS associations were attributable to a lead GWAS
52 variant, we found 20 novel genes (11 single-tissue, 9 cross-tissue) that were associated with PSA
53 levels in the TWAS. Of these novel genes, five showed evidence of colocalization
54 (colocalization probability > 0.5): *EXOSC9*, *CCNA2*, *HIST1H2BN*, *RP11-182L21.6*, and *RP11-*
55 *327J17.2*. Six of the 20 novel genes are not known to impact PCa risk. These findings yield new
56 hypotheses for genetic factors underlying PSA levels that should be further explored toward
57 improving our understanding of PSA biology.

58 **Introduction**

59 Prostate-specific antigen (PSA) is a serine protease of the human tissue kallikrein-related
60 (KLK) peptidase family. Serum levels are commonly used as a biomarker for detection,
61 monitoring, and risk stratification of prostate cancer (PCa).¹⁻³ The process by which a small
62 portion of the highly abundant PSA in the prostate is released into the blood and elevated in men
63 prostate cancer is incompletely understood, but disruption of the prostate gland architecture by
64 neoplastic transformation has been speculated as a possible mechanism.^{1,3,4} PSA levels can
65 additionally be influenced by age, race, body mass index (BMI), infection, prostate volume,
66 benign prostate hyperplasia (BPH), and germline genetics.^{5,6}

67 PSA screening has been used for over 25 years for the detection of PCa, which is the
68 second leading cause of cancer death among men in the United States.^{1,7-10} However, low test
69 specificity and discrimination have complicated the use and interpretation of PSA as a screening
70 tool.^{9,11} Long-term outcomes data from large population-based randomized PSA screening trials
71 show that it significantly reduces death from prostate cancer but also results in considerable
72 overdiagnosis of low risk disease.^{12,13} Screening might be improved were it to account for
73 variation in PSA levels that is attributable to germline genetics rather than PCa. Twin studies and
74 genome-wide association studies (GWAS) have estimated that 30-45% of serum PSA variation is
75 influenced by hereditary factors.^{14,15} A recent GWAS meta-analysis from our group identified
76 128 independent variants associated with serum PSA levels, which explained approximately 9%
77 of genetic variation in PSA.^{16,17}

78 Additional genetic variation in PSA levels may be determined by analyses of genes, as
79 opposed to individual variants. Such work could lead to the identification of regions or biological
80 pathways that affect PSA levels and provide clarity on mechanisms underlying constitutional

81 increases in PSA levels in the absence of carcinogenesis. Transcriptome-wide association studies
82 (TWAS) using expression quantitative trait loci (eQTL) allow for identifying genes whose cis-
83 regulated expression is associated with complex polygenic traits.¹⁸⁻²¹ We thus performed a
84 TWAS of PSA levels based on summary statistics from a GWAS meta-analysis of 95,768 men
85 without PCa to identify genes associated with PSA levels and prioritize them for functional
86 investigation.^{15,22}
87

88 **Methods**

89 Discovery Populations

90 A total of 95,768 PCa-free men from the following five study populations were included:
91 UK Biobank, Kaiser Permanente's Genetic Epidemiology Research on Adult Health and Aging,
92 Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, Vanderbilt University
93 Medical Center's BioVU, and Malmö Diet and Cancer Study. The men included in these
94 analyses were restricted to PCa-free individuals (i.e., without a PCa diagnosis or history of
95 prostate resection, where information was available). Additionally, analyses were restricted to
96 men with PSA values between 0.01 ng/mL and 10 ng/mL. Additional details about the study
97 cohorts are described in detail in Kachuri et al. and Table S1. Median PSA levels were used for
98 individuals with multiple PSA measurements available, with the exception of PLCO, which used
99 PSA levels at the start of follow-up of the trial.

100

101 Quality Control and GWAS Meta-analysis

102 Genotyping, imputation, study-specific quality control of genetic data, and the GWAS
103 meta-analysis have been previously described for Kachuri et al.¹⁷ Briefly, ancestry and study-
104 specific GWAS analyses used linear regression of log(PSA) on genetic variants, age, and the first
105 10 genetic ancestry principal components. After results across studies were meta-analyzed within
106 ancestral groups, ancestry-specific summary statistics from individuals of European (n=85,824),
107 African (n=3,509), East Asian (n=3,337), and Hispanic/Latino (n= 3,098) ancestry were meta-
108 analyzed to generate multi-ancestry summary statistics.

109

110 MetaXcan Transcriptome-wide Gene-based Analysis

111 We undertook a TWAS using the MetaXcan approach, which directly estimates Z-scores
112 for associations between gene expression levels and PSA levels using meta-analyzed GWAS
113 summary statistics with tissue-specific prediction models. Multivariate adaptive shrinkage
114 (mashr) prediction models were trained on GTEx (version 8) eQTL data for whole blood and,
115 separately, prostate tissue. The covariances and weights for models trained on individual tissue
116 types were from the PredictDB Data Repository (<https://predictdb.org/>).²³ We also undertook a
117 cross-tissue analysis across 45 GTEx tissues that excluded tissues found primarily or exclusively
118 in women to identify potential associations with genes that may have been missed in single-
119 tissue analyses. Tissue types found primarily or exclusively in females (mammary breast, ovary,
120 uterus, vagina) were excluded from the multiple tissue analysis. A Bonferroni correction was
121 implemented based on the number of genes tested (whole blood: $0.05/12,192 = 4.10e-6$; prostate:
122 $0.05/13,844 = 3.61e-6$; cross-tissue: $0.05/22,249 = 2.25e-6$).

123

124 Conditional Analysis Ascertaining Independence of TWAS and Previous GWAS Findings

125 To identify the set of TWAS genes for PSA that were novel in comparison to previous
126 GWAS results, we first limited consideration to significant TWAS genes that do not contain
127 genome-wide significant ($p < 5e-8$) variants from the Kachuri et al GWAS (i.e., within exact
128 gene boundaries).¹⁷ Second, for the remaining genes, we performed conditional analyses using
129 COJO²⁴ that simultaneously modeled eQTLs for a given TWAS-identified gene and GWAS
130 results. For individual tissue analyses, eQTLs used in the prediction models of TWAS-significant
131 genes were selected from the tissue-specific PredictDB TWAS eQTL weight files for each gene.
132 For the cross-tissue analyses, the maximum noncollinear eQTLs (variant inflation factor < 8
133 within a window size of 50kb) were extracted from all 45 GTEx tissues (i.e., to remove variants

134 in linkage disequilibrium) . COJO was performed using LD reference panels from 10,000
135 European ancestry UK Biobank participants (as 89.6% of our population was European
136 ancestry).²⁴ GWAS summary statistics were then conditioned on the eQTLs for one gene at a
137 time. If the set of genome-wide significant variants was reduced after conditioning on eQTLs
138 used to predict gene expression, then the eQTLs and GWAS hits were assumed to be in LD and
139 thus not independent of one another. (Methods that assess the conditional fit of eQTLs on
140 GWAS hits are lacking.) However, if the set of genome-wide significant variants remained
141 unchanged after conditional analyses, then we considered the TWAS gene to be independent of
142 GWAS variants from Kachuri et al.¹⁷

143

144 Colocalization Analyses

145 To further investigate whether genes and variants exhibited shared genetic signals, we
146 used fastENLOC software (version 2) to undertake colocalization analyses of eQTLs for all
147 genes with prediction models across TWAS analyses and all variants investigated in the Kachuri
148 et al. GWAS.^{17,25,26} For these analyses, LD blocks were based on 1000 Genome Project
149 European ancestry individuals.²⁷ Z-scores were calculated by dividing the betas from GWAS
150 summary statistics by their standard errors and converting them to posterior inclusion
151 probabilities.²⁵ Finally, the GWAS summary statistics were colocalized for 49 GTEx version 8
152 tissues using pre-computed GTEx multiple-tissue annotations. Signal-level results returned
153 regional-level colocalization probabilities (RCP) between eQTL and GWAS signals, which sum
154 up the variant-level colocalization probabilities of correlated variants within an LD block that
155 harbors a single GWAS association signal. Gene-results were based on gene-level regional
156 colocalization probabilities (GRCP), which represent the probability that a candidate gene

157 contains at least one colocalized variant.²⁶ RCP and GRCP values > 0.5 indicate strong evidence
158 of shared genetic signals between eQTLs and GWAS variants.²⁶

159

160 Pathway Enrichment Analysis

161 To explore the potential biological relevance of PSA-associated genes, we applied the
162 Enrichr tool to all significant genes identified in the whole blood, prostate, and cross-tissue
163 analyses to assess enrichment against three gene set libraries: Kyoto Encyclopedia of Genes and
164 Genomes (KEGG) 2021 human; gene ontology (GO) biological process 2021; and GO molecular
165 function 2021. Enrichment was assessed by multiplying the p-value from a Fisher's exact test
166 with the z-score of the deviation from the expected rank.^{28,29} Pathways with a Benjamini-
167 Hochberg-corrected p-value < 0.05 were considered significantly enriched.

168 **Results**

169 Using MetaXcan and GWAS summary statistics based on 95,769 individuals (Table S1), TWAS
170 analyses were conducted for whole blood, prostate, and cross-tissue matrices. The sample was
171 primarily European (n=85,824), although individuals of African (n=3,509), East Asian
172 (n=3,337), and Hispanic/Latino (n= 3,098) ancestry were also included in analyses. The median
173 PSA value across all possible values across different cohorts was 2.35 ng/mL.

174

175 Whole Blood TWAS

176 In the TWAS based on whole blood gene expression models, 41 out of 12,192 genes
177 were associated with PSA levels at the Bonferroni corrected threshold (Figure 1, Table S2).
178 Expression of 20 genes was positively associated with elevated PSA, and expression of 21 genes
179 was inversely associated. Two of the significant genes are located at 19q13.33, which contains
180 *KLK3*, the gene encoding serum PSA.³⁰ While a prediction model for *KLK3* was not available in
181 whole blood, a member of the same gene family at 19q13.33, *KLK2*, was the gene most strongly
182 associated with PSA levels ($p = 1.28e-62$). Twenty-four of the 41 genes did not contain genome-
183 wide significant variants and had not been annotated in previous GWAS. Among them, 7 genes
184 were independent of previously published GWAS findings (Table 1).¹⁷ Of these, increased
185 expression of 6 genes was associated with elevated PSA levels in PCa-free men: *GPBP1L1*
186 (1p34.1, $p = 2.21e-7$); *TMEM69* (1p34.1, $p = 2.21e-7$); *UQCRH* (1p33, $p = 4.76e-7$); *ACTRT3*
187 (3q26.2, $p = 5.90e-7$); *EXOSC9* (4q27, $p = 1.80e-6$); and *CCNA2* (4q27, $p = 7.80e-7$). Decreased
188 expression of *ITH4* was associated with increased PSA levels (3p21.1, $p = 3.65e-6$).

189

190 Prostate Tissue TWAS

191 In TWAS based on prostate tissue gene expression models, 39 out of 13,884 genes were
192 statistically significantly associated with PSA levels (Figure 2, Table S3). Among them,
193 increased expression of 18 and decreased expression of 21 was associated with elevated PSA
194 levels. No prediction models were available for genes in the *KLK* family in prostate tissue.
195 Instead, *MMP7* on 11q22 was most strongly associated with PSA levels ($p = 2.78e-18$). *MMP7*
196 was among 18 PSA-associated TWAS genes containing known genome-wide significant
197 variants. Of the remaining 21 significant genes, six were conditionally independent of previous
198 GWAS results after conditioning on the eQTL variants used in gene prediction models (Table 1).
199 Of these, increased expression of 3 genes was associated with elevated PSA levels: *LRRC41*
200 (1p33, $p = 4.76e-7$), *HIST1H2BN* (6p22.1, $p = 8.86e-8$), and *HEXIMI* (17q21.31, $p = 4.95e-7$).
201 Decreased expression of the remaining 3 genes was associated with increased PSA levels: *ITIH4*
202 (3p21.1, $p = 2.69e-6$), *EXOSC9* (4q27 $p = 1.18e-6$), and *OTOA* (16p12.2, $p = 1.44e-6$).
203

204 Cross-Tissue TWAS

205 Using S-MultiXcan to integrate signals from 45 tissues excluding tissues primarily or
206 exclusively present in women, we identified 155 out of 22,249 genes whose predicted expression
207 was associated with elevated PSA levels in cross-tissue matrices (Figure 3). Among the 155
208 statistically significant genes, 8 are located on 19q13.33 and part of the expanded human *KLK*
209 gene family. We found 71 genes that do not contain genome-wide significant loci within their
210 gene boundaries. Conditional analyses determined that 13 genes from this set were independent
211 of known genome-wide significant variants: *CDC42* (1p36.12, $p = 6.24e-07$), *GPBP1L1* (1p34.1,
212 $p = 2.08e-08$), *SLBP* (4p16.3, $p = 8.40e-08$), *EXOSC9* (4q27, $p = 8.60e-07$), *CCNA2* (4q27, $p =$
213 $1.02e-06$), *ANXA8L1* (10q11.22, $p = 2.17e-07$), *RP11-182L21.6* (10q22.3, $p = 1.54e-07$), *E2F7*

214 (12q21.2, $p = 8.54e-09$), *C15orf62* (15q15.1, $p = 2.01e-07$), *RP11-327J17.2* (15q26.2, $p = 5.82e-$
215 07), *C16orf47* (16q22.3, $p = 1.11e-07$), *MYOCD* (17p12, $p=1.20E-07$), and *ABCD4* (17q21.31, p
216 $= 2.17e-06$) (Table 1).

217

218 Significant Genes Overlapping Across TWAS

219 Out of the 41 genes that were significantly associated with PSA levels in the whole blood
220 TWAS, 34 were also imputed in the prostate tissue TWAS. Thirty (30) out of 39 significantly
221 associated genes from the prostate tissue TWAS were also evaluated in whole blood analyses. Of
222 the 155 significantly associated genes detected in the cross-tissue analysis, 89 genes and 105
223 genes had transcriptome prediction models available for whole blood and prostate tissues,
224 respectively.

225 Across all TWAS in whole blood, prostate tissue, and cross-tissue models, we identified
226 13 genes whose predicted expression was associated with elevated PSA levels: *RAB29* (1q32.1),
227 *LDAH* (2p24.1), *ZFP36L2* (2p21), *SIDT1* (3q13.2), *TRIM59* (3q25.33), *EXOSC9* (4q27),
228 *HIBADH* (7p15.2), *TAX1BP1* (7p15.2), *MEGF9* (9q33.2), *ADD3* (10q25.1), *APIP* (11p13),
229 *LINC01569* (16p13.3), and *EIF2S2* (20q11.22) (Figure 4). Of these genes, only *EXOSC9* was
230 conditionally independent after accounting for prior loci and demonstrated evidence of
231 colocalization (described below). Five (5) associations were not statistically significant in the
232 cross-tissue analysis, but had significant tissue-specific signals in whole blood and/or prostate.
233 Seventeen significant genes were shared by only the whole blood and cross-tissue TWAS, six of
234 which did not have prediction models in prostate tissue. Similarly, the prostate tissue and cross-
235 tissue TWAS shared 14 significant genes; of these, seven genes did not have prediction models
236 in whole blood.

237

238 Colocalization Analysis

239 Colocalization identified 372 signals with RCP > 0.5 at the variant level (Table S6) and
240 307 genes with GRCP > 0.5 at the gene level (Table S7). The top 10 genes with the largest
241 GRCP were *OTX1* (2p15), *MAFF* (22q13.1), *FUT2* (19q13.33), *EMSY* (11q13.5), *IFT80*
242 (3q25.33), *EXOSC9* (4q27), *SMC4* (3q25.33), *RCAN3* (1p36.11), *PBXIP1* (1q21.3), and *PMVK*
243 (1q21.3). Of these genes, only *EXOSC9* was determined to be independent of previously
244 discovered GWAS loci.¹⁷

245 Among 155 significant genes from the cross-tissue TWAS, locus-level results identified
246 50 genes with strong evidence of colocalization (Table 2). Variant-level results identified 66
247 regions in 51 genes (Table S5). When comparing the overlap between transcriptome-wide
248 significant genes and results from colocalization analyses, we identified 5 novel genes (*EXOSC9*,
249 *CCNA2*, *HIST1H2BN*, *RP11-182L21.6*, and *RP11-327J17.2*) with GRCP/RCP > 0.5. Two of
250 these genes are long non-coding RNA while the remaining three are protein-coding genes.

251

252 Pathway Enrichment Analysis

253 TWAS genes were significantly enriched in 23 pathways underlying molecular activities
254 and biological processes present in KEGG and GO gene set databases (Table 3, Figure 5).
255 Enrichment analyses in the whole blood gene set identified the renin-angiotensin-aldosterone
256 system in KEGG, which works to regulate arterial blood pressure. No pathways were
257 significantly enriched by prostate and cross-tissue gene sets in the KEGG catalog.

258 Enrichment analyses within the GO biological process and molecular function gene set
259 libraries identified 23 pathways that were significantly enriched in the cross-tissue gene set. No

260 pathways were significantly enriched in GO for whole blood nor prostate tissue TWAS sets. Of
261 note, the biological process underlying the regulation of cyclin-dependent protein
262 serine/threonine kinase (CDK) activity in cell cycle progression is particularly relevant for the
263 function of *CCNA2*. We also observed enrichment in 13 pathways involving both single and
264 double stranded DNA binding mechanisms and regulation by transcription factors, including
265 RNA polymerase II (pol II), a core component of the DNA transcription machinery. Moreover,
266 two pathways underlying extracellular matrix and cellular component disassembly were enriched
267 in the cross-tissue gene set. Analogous to the findings in KEGG for the whole blood TWAS,
268 biological pathways underlying the regulation of systemic arterial blood pressure were also
269 identified in the GO biological process catalog.

270 Discussion

271 This TWAS of 95,768 PCa-free men identified 173 genes whose predicted expression
272 levels were significantly associated with PSA. Conditional analyses identified 20 novel candidate
273 susceptibility genes for PSA, and colocalization analyses highlighted five of them: *EXOSC9* and
274 *CCNA2* at 4q27, *HIST1H2BN* at 6p22.1, *RP11-182L21.6* at 10q22.3, and *RP11-327J17.2* at
275 15q26.2. Pathway enrichment analyses across three gene ontology catalogs implicated regulatory
276 pathways related to transcription, cell signaling, and disassembly of cellular and noncellular
277 components such as the extracellular matrix.

278 The five colocalized genes introduce new hypotheses and insights regarding the genetic
279 mechanisms regulating PSA production. Of particular interest are the signals at the 4q27 locus
280 (*EXOSC9* and *CCNA2*), a genomic region that was not detected in previous GWAS of PSA.^{16,17}
281 Researchers have previously investigated an autoimmune related block on the 4q27 locus and did
282 not find an association with PCa risk.^{32–35} *EXOSC9*, which encodes a core protein involved in
283 the RNA degradation machinery in humans, was identified across all three TWAS analyses.
284 Currently, there is no literature supporting an association between *EXOSC9* and PSA levels or
285 PCa risk. Genetically predicted expression of *EXOSC9* in whole blood was positively associated
286 with PSA levels, while the opposite direction of effect was observed for expression in prostate
287 tissue. There exists limited tissue-specific knowledge of the physiological and pathological
288 function of *EXOSC9* in non-cancer cell lines.^{34,35} Literature documenting the relationship
289 between *CCNA2* and PSA is also limited, though *CCNA2* was previously tested as a candidate
290 autoantibody signature marker for distinguishing PCa from BPH in patients with elevated serum
291 PSA;³⁶ it was not determined to be a top antibody signature of any specific PCa targets. Gene co-
292 expression network analyses have linked *CCNA2* to biochemical recurrence and survival in men

293 with PCa.^{37,38} Further exploration of *CCNA2* expression in individuals without PCa may unveil
294 biological pathways that influence PSA levels.

295 Pathway enrichment results also highlighted a number of interesting pathways involving
296 Pol II, which plays a critical role in the regulated synthesis of both protein-coding and noncoding
297 messenger RNA in eukaryotic genomes.³⁹⁻⁴¹ In addition to its numerous roles in DNA
298 transcription, Pol II is known to associate with androgen receptor binding regions of the PSA
299 enhancer and promoter to initiate androgen-dependent transcription.⁴² The two colocalized
300 lncRNA genes, *RP11-182L21.6* and *RP11-327J17.2*, could feasibly regulate epigenetic
301 modification through histone or DNA acetylation or methylation.⁴³ That histone deacetylase
302 binding was enriched in our pathway analyses suggests a possible mechanism through which
303 lncRNAs could affect constitutive PSA levels. Aberrant expression of various lncRNA products
304 has also been observed for various cancers, including PCa.⁴⁴

305 Because PSA is used for both PCa detection and monitoring of PCa progression, it is
306 difficult to disentangle the mechanisms underlying our observed associations. PSA-related
307 screening bias may account for a portion of the observed relationship between gene expression
308 and PSA levels. Genes may also exert pleiotropic effects on PCa and PSA through overlapping
309 biological mechanisms.^{1,9} Nevertheless, our restriction to PCa-free men for all analyses
310 minimizes the potential for reverse causation and bolsters confidence that the observed genetic
311 signals inform constitutive PSA levels. Fourteen of the 20 novel PSA-related genes identified
312 from conditional analyses, spanning 12 genomic regions not implicated by prior PSA GWAS,
313 have been associated with prostatic malignancies: *CDC42* (1p36.12), *GPBP1L1* and *TMEM69* (
314 1p34.1), *LRR41* (1p33), *ITIH4* (3p21.1), *SLBP* (4p16.3), *CCNA2* (4q27), *HIST1H2BN*
315 (6p22.1), *ANXA8L1* (10q11.2), *E2F7* (12q21.2), *C15orf62* (15q15.1), *OTOA* (16p12.2),

316 *C16orf47* (16q22.3), and *HEXIM1*(17q21.31).^{45–55} Colocalized gene *HIST1H2BN*, which
317 encodes a component of a core nucleosome histone, has been linked to PCa cell growth and
318 epithelial-mesenchymal transition through upregulated NF- κ B/Rel expression.⁴⁵ Activation of
319 the NF- κ B pathway can induce activation of the PSA promoter-enhancer, even in the absence of
320 androgens, and NF- κ B can directly bind to the PSA enhancer in prostate cancer cell lines.⁵⁶ To
321 our knowledge, no experiments have been conducted to investigate the relationship between NF-
322 κ B and PSA in PCa-free populations.

323 Conditional analyses further identified 6 genes that have not been implicated in PCa
324 susceptibility in gene function, experimental, or human population research: *UQCRH* (1p33),
325 *ACTRT3* (3q26.2), *EXOSC9* (4q27), *RP11-182L21.6* (10q22.3), *RP11-327J17.2* (15q26.2), and
326 *ACBD4* (17q21.31). *ACTRT3* is of particular interest, as it is critical in regulating sperm nucleus
327 cytomorphology upstream of the processing of spermatid into mature motile sperm.⁵⁷ Increased
328 expression of *ACTRT3*, which forms a testis-specific profilin III-*ACTRT3* complex that facilitates
329 male germ cell head cytomorphology and maintains sperm motility in animal models, was
330 associated with elevated PSA levels in the TWAS based on whole blood.^{57,58} Though there is no
331 documented link between *ACTRT3* and PSA levels or PCa risk, the pronounced role of *ACTRT3*
332 in mediating conformational changes in sperm nuclei suggests possible shared biological
333 pathways between PSA and the production and processing of male germ cells.^{5,57} Loci that map
334 to the multigenic region that contains *ACTRT3* have been linked to melanoma, colorectal cancer,
335 and lung cancer susceptibility.^{59–61} The functional role of *ACTRT3* is poorly characterized and
336 warrants further analysis.

337 Our study had several limitations. First, nearly 90% of men in the study population were
338 of primarily European ancestry, and 85% of the tissue samples used to train the TWAS models

339 were derived from European populations.⁶² Therefore, our TWAS may have missed ancestry-
340 specific signals, and our findings may not be generalizable to broader ancestral populations. To
341 better characterize the genetic mechanisms underlying circulating PSA levels, it will be
342 imperative for future analyses to expand to multi-ethnic populations; diversity in genetic studies
343 of PSA levels is critical for equitably improving PSA screening.⁶³ Second, our study assessed
344 only cis-eQTLs, so any trans-eQTL effects are not incorporated. Third, colocalization analyses
345 have a high type 2 error rate and may be underpowered to detect shared association signals.²⁶
346 Fourth, although we restricted our analyses to men who had not been diagnosed with PCa, we
347 cannot rule out the possibility of latent, undiagnosed disease or disease diagnosed at a later
348 timepoint. However, the prevalence of undiagnosed PCa in our population was likely to be low
349 on account of low prevalence at the time of the first PSA measurement or increased monitoring
350 and surveillance in longitudinal cohorts.^{7,9} Finally, many of the novel genes that we identified
351 were clustered at multi-gene loci, in part due to co-regulation by a shared set of eQTLs. Sentinel
352 genes at these loci should be interpreted with caution, as there may be correlated predicted
353 expression.⁶⁴

354 Our study also had several key strengths. The use of GWAS summary statistics from a
355 cohort of 95,768 men provided us with high statistical power to quantify PSA-associated genes.
356 In addition to conducting tissue-specific association analyses in tissues that are biologically
357 meaningful for PSA, we integrated association signals across 45 GTEx tissues to improve power
358 for genes with similar regulatory mechanisms across tissues. This method allowed us to capture
359 PSA-related genes with expression patterns that were significant for specific tissues, as well as
360 genes with expression that was similar across multiple tissues. Our study also used COJO
361 analysis to find novel genes conditionally independent from known GWAS variants.

362 In summary, our TWAS identified gene expression profiles associated with PSA levels in
363 men without PCa. These findings provide several novel hypotheses for genes that affect
364 constitutive PSA. Further exploration of these results, including functional analyses of these
365 genes in in-vivo settings, will augment our understanding of the genetic etiology of PSA
366 variation. Transcriptomic analyses might also be vertically integrated with downstream -omic
367 approaches to uncover complete mechanisms through which genetics influence circulating PSA
368 levels. In addition, TWAS findings may be used to develop polygenic transcriptome risk scores⁶⁵
369 for PSA levels, which could be leveraged for improving PSA as a tool for PCa screening.

370 **Declaration of Interests:**

371 JSW is a non-employee, cofounder of Avail Bio. HL is named on a patent for assays to measure
372 intact prostate-specific antigen and a patent for a statistical method to detect prostate cancer
373 commercialized by OPKO Health (4KScore). HL receives royalties from sales of the assay and
374 has stock in OPKO Health. HL serves on the Scientific Advisory Board for Fujirebio Diagnostics
375 Inc and owns stock in Diaprost AB and Acousort AB.

376 **Data Availability:**

377 GWAS summary statistics used in this analysis are available for download from Kachuri et al are
378 available from the following Zenodo repository: <https://doi.org/10.5281/zenodo.7460134>.
379 Transcriptome prediction weights and models used for MetaXcan analyses are available from:
380 <https://predictdb.org>.

381 **Funding sources:**

382 The Precision PSA study is supported by funding from the National Institutes of Health (NIH)
383 and National Cancer Institute (NCI) under award number R01CA241410 (PI: JSW). REG is
384 supported by a Prostate Cancer Foundation Young Investigator Award. LK is supported by
385 funding from National Cancer Institute (R00CA246076). JPS is supported by funding from the
386 National Institute of General Medical Sciences (T32GM007347). HL is supported in part by
387 funding from NIH/NCI (P30-CA008748, U01-CA199338, R01CA244948) and the Swedish
388 Cancer Society (Cancerfonden 20 1354 PjF). RK is supported by funding from the NIH (R01
389 CA244948 and R01 CA175491).

390 The content is solely the responsibility of the authors and does not necessarily represent the
391 official views of the NIH.

392 **References**

- 393 1. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction,
394 detection and monitoring. *Nat Rev Cancer*. 2008;8(4):268-278. doi:10.1038/nrc2351
- 395 2. Cucchiara V, Cooperberg MR, Dall'Era M, et al. Genomic Markers in Prostate Cancer
396 Decision Making. *Eur Urol*. 2018;73(4):572-582. doi:10.1016/j.eururo.2017.10.036
- 397 3. LeBeau AM, Kostova M, Craik CS, Denmeade SR. Prostate-specific antigen: an
398 overlooked candidate for the targeted treatment and selective imaging of prostate cancer.
399 *Biol Chem*. 2010;391(4):333-343. doi:10.1515/BC.2010.044
- 400 4. Qiu SD, Young CY, Bilhartz DL, et al. In situ hybridization of prostate-specific antigen
401 mRNA in human prostate. *J Urol*. 1990;144(6):1550-1556. doi:10.1016/s0022-
402 5347(17)39797-5
- 403 5. Lilja H, Oldbring J, Rannevik G, Laurell CB. Seminal vesicle-secreted proteins and their
404 reactions during gelation and liquefaction of human semen. *J Clin Invest*. 1987;80(2):281-
405 285.
- 406 6. Pinsky PF, Kramer BS, Crawford ED, et al. Prostate volume and prostate-specific antigen
407 levels in men enrolled in a large screening trial. *Urology*. 2006;68(2):352-356.
408 doi:10.1016/j.urology.2006.02.026
- 409 7. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane*
410 *Database Syst Rev*. 2013;2013(1):CD004720. doi: 10.1002/14651858.CD004720.pub3.

- 411 8. Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-Specific Antigen-
412 Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US
413 Preventive Services Task Force. *JAMA*. 2018;319(18):1914-1931.
414 doi:10.1001/jama.2018.3712
- 415 9. Ilic D, Djulbegovic M, Jung JH, et al. Prostate cancer screening with prostate-specific
416 antigen (PSA) test: a systematic review and meta-analysis. *The BMJ*. 2018;362:k3519.
417 doi:10.1136/bmj.k3519
- 418 10. Potosky AL, Feuer EJ, Levin DL. Impact of screening on incidence and mortality of
419 prostate cancer in the United States. *Epidemiol Rev*. 2001;23(1):181-186.
420 doi:10.1093/oxfordjournals.epirev.a000787
- 421 11. Han PKJ, Kobrin S, Breen N, et al. National Evidence on the Use of Shared Decision
422 Making in Prostate-Specific Antigen Screening. *Ann Fam Med*. 2013;11(4):306-314.
423 doi:10.1370/afm.1539
- 424 12. Hugosson J, Roobol MJ, Månsson M, et al. A 16-yr Follow-up of the European
425 Randomized study of Screening for Prostate Cancer. *Eur Urol*. 2019;76(1):43-51.
426 doi:10.1016/j.eururo.2019.02.009
- 427 13. de Koning HJ, Gulati R, Moss SM, et al. The efficacy of prostate-specific antigen
428 screening: Impact of key components in the ERSPC and PLCO trials. *Cancer*.
429 2018;124(6):1197-1206. doi:10.1002/cncr.31178

- 430 14. Bansal A, Murray DK, Wu JT, Stephenson RA, Middleton RG, Meikle AW. Heritability of
431 Prostate-Specific Antigen and Relationship with Zonal Prostate Volumes in Aging Twins*.
432 *J Clin Endocrinol Metab.* 2000;85(3):1272-1276. doi:10.1210/jcem.85.3.6399
- 433 15. Pilia G, Chen WM, Scuteri A, et al. Heritability of Cardiovascular and Personality Traits in
434 6,148 Sardinians. *PLoS Genet.* 2006;2(8):e132. doi:10.1371/journal.pgen.0020132
- 435 16. Hoffmann TJ, Passarelli MN, Graff RE, et al. Genome-wide association study of prostate-
436 specific antigen levels identifies novel loci independent of prostate cancer. *Nat Commun.*
437 2017;8:14248. doi:10.1038/ncomms14248
- 438 17. Kachuri L, Hoffmann TJ, Jiang Y, et al. Incorporating Genetic Determinants of Prostate-
439 Specific Antigen Levels Improves Prostate Cancer Screening. *MedRxiv.*
440 doi:10.1101/2022.04.18.22273850
- 441 18. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are
442 more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.*
443 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888
- 444 19. Gamazon ER, Wheeler HE, Shah KP, et al. A gene-based association method for mapping
445 traits using reference transcriptome data. *Nat Genet.* 2015;47(9):1091-1098.
446 doi:10.1038/ng.3367
- 447 20. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide
448 association studies. *Nat Genet.* 2016;48(3):245-252. doi:10.1038/ng.3506

- 449 21. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association
450 statistics. *Nat Rev Genet.* 2017;18(2):117-127. doi:10.1038/nrg.2016.142
- 451 22. Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of
452 tissue specific gene expression variation inferred from GWAS summary statistics. *Nat*
453 *Commun.* 2018;9(1):1825. doi:10.1038/s41467-018-03621-1
- 454 23. Barbeira, A.N., Bonazzola, R., Gamazon, E.R. et al. Exploiting the GTEx resources to
455 decipher the mechanisms at GWAS loci. *Genome Biol.* 2021;22(49). doi: 10.1186/s13059-
456 020-02252-4
- 457 24. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS
458 summary statistics identifies additional variants influencing complex traits. *Nat Genet.*
459 2012;44(4):369-S3. doi:10.1038/ng.2213
- 460 25. Wen X, Pique-Regi R, Luca F. Integrating molecular QTL data into genome-wide genetic
461 association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS*
462 *Genet.* 2017;13(3):e1006646. doi:10.1371/journal.pgen.1006646
- 463 26. Hukku A, Sampson MG, Luca F, Pique-Regi R, Wen X. Analyzing and reconciling
464 colocalization and transcriptome-wide association studies from the perspective of
465 inferential reproducibility. *Am J Hum Genet.* 2022;109(5):825-837.
466 doi:10.1016/j.ajhg.2022.04.005
- 467 27. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human
468 populations. *Bioinformatics.* 2016;32(2):283-285. doi:10.1093/bioinformatics/btv546

- 469 28. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list
470 enrichment analysis tool. *BMC Bioinformatics*. 2013;14(1):128. doi:10.1186/1471-2105-14-
471 128
- 472 29. Xie Z, Bailey A, Kuleshov MV, et al. Gene Set Knowledge Discovery with Enrichr. *Curr*
473 *Protoc*. 2021;1(3):e90. doi:10.1002/cpz1.90
- 474 30. Balk SP, Ko YJ, Bubley GJ. Biology of Prostate-Specific Antigen. *J Clin Oncol*.
475 2003;21(2):383-391. doi:10.1200/JCO.2003.02.083
- 476 31. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted
477 transcriptome from multiple tissues improves association detection. *PLOS Genet*.
478 2019;15(1):e1007889. doi:10.1371/journal.pgen.1007889
- 479 32. Tindall EA, Hoang HN, Southey MC, et al. The 4q27 locus and prostate cancer risk. *BMC*
480 *Cancer*. 2010;10(1):69. doi:10.1186/1471-2407-10-69
- 481 33. Yoshino S, Hara T, Weng JS, Takahashi Y, Seiki M, Sakamoto T. Genetic Screening of
482 New Genes Responsible for Cellular Adaptation to Hypoxia Using a Genome-Wide shRNA
483 Library. *PLoS ONE*. 2012;7(4):e35590. doi:10.1371/journal.pone.0035590
- 484 34. Yoshino S, Matsui Y, Fukui Y, et al. EXOSC9 depletion attenuates P-body formation,
485 stress resistance, and tumorigenicity of cancer cells. *Sci Rep*. 2020;10(1):9275.
486 doi:10.1038/s41598-020-66455-2

- 487 35. Burns DT, Donkervoort S, Müller JS, et al. Variants in EXOSC9 Disrupt the RNA
488 Exosome and Result in Cerebellar Atrophy with Spinal Motor Neuronopathy. *Am J Hum*
489 *Genet.* 2018;102(5):858-873. doi:10.1016/j.ajhg.2018.03.011
- 490 36. O'Rourke DJ, DiJohnson DA, Caiazzo RJ, et al. Autoantibody Signatures as Biomarkers to
491 Distinguish Prostate Cancer from Benign Prostatic Hyperplasia in Patients with Increased
492 Serum Prostate Specific Antigen. *Clin Chim Acta Int J Clin Chem.* 2012;413(5-6):561-567.
493 doi:10.1016/j.cca.2011.11.027
- 494 37. Feng T, Wei D, Li Q, et al. Four Novel Prognostic Genes Related to Prostate Cancer
495 Identified Using Co-expression Structure Network Analysis. *Front Genet.* 2021;12. doi:
496 10.3389/fgene.2021.584164
- 497 38. Yang R, Du Y, Wang L, Chen Z, Liu X. Weighted gene co-expression network analysis
498 identifies CCNA2 as a treatment target of prostate cancer through inhibiting cell cycle. *J*
499 *Cancer.* 2020;11(5):1203-1211. doi:10.7150/jca.38173
- 500 39. Smale ST, Kadonaga JT. The RNA polymerase II core promoter. *Annu Rev Biochem.*
501 2003;72:449-479. doi:10.1146/annurev.biochem.72.121801.161520
- 502 40. Wagner SD, Yakovchuk P, Gilman B, et al. RNA polymerase II acts as an RNA-dependent
503 RNA polymerase to extend and destabilize a non-coding RNA. *EMBO J.* 2013;32(6):781-
504 790. doi:10.1038/emboj.2013.18
- 505 41. Schier AC, Taatjes DJ. Structure and mechanism of the RNA polymerase II transcription
506 machinery. *Genes Dev.* 2020;34(7-8):465-488. doi:10.1101/gad.335679.119

- 507 42. Louie MC, Yang HQ, Ma AH, et al. Androgen-induced recruitment of RNA polymerase II
508 to a nuclear receptor–p160 coactivator complex. *Proc Natl Acad Sci*. 2003;100(5):2226-
509 2230. doi:10.1073/pnas.0437824100
- 510 43. Zhang X, Wang W, Zhu W, et al. Mechanisms and Functions of Long Non-Coding RNAs
511 at Multiple Regulatory Levels. *Int J Mol Sci*. 2019;20(22):5573. doi:10.3390/ijms20225573
- 512 44. Misawa A, Takayama K, Inoue S. Long non-coding RNAs and prostate cancer. *Cancer Sci*.
513 2017;108(11):2107-2114. doi:10.1111/cas.13352
- 514 45. Zhang, J., Chang, Y., Xia, H. *et al.* HIST1H2BN induced cell proliferation and EMT
515 phenotype in prostate cancer via NF- κ B signal pathway. *Genes Genom*. 2021;43:1361-
516 1369. doi: 10.1007/s13258-021-01164-2
- 517 46. Liu Q, Reed M, Zhu H, et al. Epigenome-wide DNA methylation and transcriptome
518 profiling of localized and locally advanced prostate cancer: Uncovering new molecular
519 markers. *Genomics*. 2022;114(5):110474. doi:10.1016/j.ygeno.2022.110474
- 520 47. Pressinotti NC, Klocker H, Schäfer G, et al. Differential expression of apoptotic genes
521 PDIA3 and MAP3K5 distinguishes between low- and high-risk prostate cancer. *Mol*
522 *Cancer*. 2009;8:130. doi:10.1186/1476-4598-8-130
- 523 48. Mahajan NP, Liu Y, Majumder S, et al. Activated Cdc42-associated kinase Ack1 promotes
524 prostate cancer progression via androgen receptor tyrosine phosphorylation. *Proc Natl Acad*
525 *Sci U S A*. 2007;104(20):8438-8443. doi:10.1073/pnas.0700420104

- 526 49. Davaliev K, Kiprijanovska S, Komina S, Petrussevska G, Zografaska NC, Polenakovic M.
527 Proteomics analysis of urine reveals acute phase response proteins as candidate diagnostic
528 biomarkers for prostate cancer. *Proteome Sci.* 2015;13:2. doi:10.1186/s12953-014-0059-9
- 529 50. Krishnan N, Titus MA, Thapar R. The Prolyl Isomerase Pin1 Regulates mRNA Levels of
530 Genes with Short Half-Lives by Targeting Specific RNA Binding Proteins. *PLOS ONE.*
531 2014;9(1):e85427. doi:10.1371/journal.pone.0085427
- 532 51. Wang Y, Pei X, Xu P, et al. E2F7, regulated by miR-30c, inhibits apoptosis and promotes
533 cell cycle of prostate cancer cells. *Oncol Rep.* 2020;44(3):849-862.
534 doi:10.3892/or.2020.7659
- 535 52. Sheeba JDJ, Hegde S, Tamboli N, Nadig N, Keshavamurthy R, Ranganathan P. Gene
536 expression signature of castrate resistant prostate cancer. *BioRxiv.* 2022.03.16.484397.
537 doi:10.1101/2022.03.16.484397
- 538 53. Grasso CS, Wu YM, Robinson DR, et al. The mutational landscape of lethal castration-
539 resistant prostate cancer. *Nature.* 2012;487(7406):239-243. doi:10.1038/nature11125
- 540 54. Lew QJ, Chu KL, Chia YL, Cheong N, Chao SH. HEXIM1, a New Player in the p53
541 Pathway. *Cancers.* 2013;5(3):838-856. doi:10.3390/cancers5030838
- 542 55. Chen R, Yik JHN, Lew QJ, Chao SH. Brd4 and HEXIM1: Multiple Roles in P-TEFb
543 Regulation and Cancer. *BioMed Res Int.* 2014;2014:e232870. doi:10.1155/2014/232870

- 544 56. Chen CD, Sawyers CL. NF- κ B Activates Prostate-Specific Antigen Expression and Is
545 Upregulated in Androgen-Independent Prostate Cancer. *Mol Cell Biol.* 2002;22(8):2862-
546 2870. doi:10.1128/MCB.22.8.2862-2870.2002
- 547 57. Hara Y, Yamagata K, Oguchi K, Baba T. Nuclear localization of profilin III-ArpM1
548 complex in mouse spermiogenesis. *FEBS Lett.* 2008;582(20):2998-3004.
549 doi:10.1016/j.febslet.2008.07.058
- 550 58. Umer N, Arévalo L, Phadke S, et al. Loss of Profilin3 Impairs Spermiogenesis by Affecting
551 Acrosome Biogenesis, Autophagy, Manchette Development and Mitochondrial
552 Organization. *Front Cell Dev Biol.* 2021;9 9:749559. doi: 10.3389/fcell.2021.749559
- 553 59. Song F, Amos CI, Lee JE, et al. Identification of a melanoma susceptibility locus and
554 somatic mutation in TET2. *Carcinogenesis.* 2014;35(9):2097-2101.
555 doi:10.1093/carcin/bgu140
- 556 60. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al. Association analyses identify 31 new
557 risk loci for colorectal cancer susceptibility. *Nat Commun.* 2019;10(1):2154.
558 doi:10.1038/s41467-019-09775-w
- 559 61. McKay JD, Hung RJ, Han Y, et al. Large-scale association analysis identifies new lung
560 cancer susceptibility loci and heterogeneity in genetic susceptibility across histological
561 subtypes. *Nat Genet.* 2017;49(7):1126-1132. doi:10.1038/ng.3892
- 562 62. THE GTEx CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects
563 across human tissues. *Science.* 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776

- 564 63. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current
565 polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51(4):584-591.
566 doi:10.1038/s41588-019-0379-x
- 567 64. Wainberg M, Sinnott-Armstrong N, Mancuso N, et al. Opportunities and challenges for
568 transcriptome-wide association studies. *Nat Genet.* 2019;51(4):592-599.
569 doi:10.1038/s41588-019-0385-z
- 570 65. Liang, Y., Pividori, M., Manichaikul, A. *et al.* Polygenic transcriptome risk scores
571 (PTRS) can improve portability of polygenic risk scores across ancestries. *Genome*
572 *Biol* 2022;23(23). doi: 10.1186/s13059-021-02591-w

573 **Figure Titles and Legends:**

574 **Figures 1-3. TWAS results by chromosome**, as depicted by (1) Miami plot for whole blood
575 analyses, (2) Miami plot for prostate tissue analyses, and (3) Manhattan plot for cross-tissue
576 analyses. Each dot represents a gene, and the red dashed lines represent the significance
577 thresholds after Bonferroni correction (whole blood: $4.10e-6$, prostate: $3.61e-6$, cross-tissue:
578 $2.25e-6$). For the Miami plots in figures 1 and 2, all statistically significant genes are annotated,
579 and the genes independent of results from the largest prior GWAS are highlighted in yellow; the
580 genes in the upper half of the plots were positively associated with PSA levels, and the genes in
581 the lower half of the plots were inversely associated with PSA levels. For the Manhattan plot in
582 figure 3, the 30 genes with the smallest association p-values are annotated and colored blue, and
583 significant genes from the whole blood and prostate tissue analyses are annotated and colored
584 yellow. Directions of association between genes and PSA levels are not available for cross-tissue
585 analyses due to the joint tissue inference methodology used in S-MultiXcan.

586 **Figure 4. Statistically significant genes overlapping across TWAS analyses.** Of the
587 significant gene sets from whole blood (41 genes), prostate tissue (39 genes), and cross-tissue
588 analyses (155 genes), 13 genes were significant across all three analyses, 36 genes were
589 significant in exactly two analyses, and 124 genes were significant in a single analysis.

590 **Figure 5. GO term enrichment results from the cross-tissue TWAS.** The GO molecular
591 function and biological process data repositories were queried for pathway enrichment analyses.
592 The color gradient represents the magnitude of p-values, with darker colors corresponding to
593 smaller values. The size of the circles represents the number of genes in the pathway.

594 **Table Titles**

595 **Table 1.** Novel genes associated with prostate-specific antigen (PSA) levels in transcriptome-
 596 wide association studies (TWAS) based on whole blood, prostate, and cross-tissue predicted
 597 gene expression*

Ensembl gene ID	Gene name	Chr	Cytoband [†]	Analysis [~]	Z- score	Beta	p	p-bonf
ENSG00000132128	<i>LRRC41</i>	1	p33	Prostate	5.04	0.29	4.76E-07	6.59E-03
				Whole Blood	5.18	17.92	2.21E-07	2.69E-03
ENSG00000159592	<i>GPBP1L1</i>	1	p34.1	Cross-Tissue	-1.92	-	2.08E-08	4.59E-04
ENSG00000159596	<i>TMEM69</i>	1	p34.1	Whole Blood	5.18	0.28	2.21E-07	2.69E-03
ENSG00000173660	<i>UQCRH</i>	1	p33	Whole Blood	5.04	8.04	4.76E-07	5.80E-03
ENSG00000070831	<i>CDC42</i>	1	p36.12	Cross-Tissue	-1.55	-	6.24E-07	1.37E-02
ENSG00000184378	<i>ACTRT3</i>	3	q26.2	Whole Blood	4.99	0.25	5.90E-07	7.19E-03
				Whole Blood	-4.63	-0.03	3.65E-06	4.45E-02
ENSG00000055955	<i>ITIH4</i>	3	p21.1	Prostate	-4.69	-0.03	2.69E-06	3.72E-02
ENSG00000163950	<i>SLBP</i>	4	p16.3	Cross-Tissue	-2.15	-	8.40E-08	1.85E-03
				Whole Blood	4.94	0.61	7.80E-07	9.51E-03
ENSG00000145386	<i>CCNA2</i>	4	q27	Cross-Tissue	-1.81	-	1.02E-06	2.25E-02
				Whole Blood	4.77	0.09	1.80E-06	2.19E-02
				Prostate	-4.86	-0.42	1.18E-06	1.63E-02
ENSG00000123737	<i>EXOSC9</i>	4	q27	Cross-Tissue	0.36	-	8.60E-07	1.89E-02
ENSG00000233822	<i>HIST1H2BN</i>	6	p22.1	Prostate	5.35	0.24	8.86E-08	1.23E-03
ENSG00000272447	<i>RP11-182L21.6</i>	10	q22.3	Cross-Tissue	1.25	-	1.54E-07	3.40E-03
ENSG00000264230	<i>ANXA8L1</i>	10	q11.22	Cross-Tissue	-0.46	-	2.17E-07	4.77E-03
ENSG00000165891	<i>E2F7</i>	12	q21.2	Cross-Tissue	1.45	-	8.54E-09	1.88E-04
ENSG00000188277	<i>C15orf62</i>	15	q15.1	Cross-Tissue	-0.85	-	2.01E-07	4.42E-03
ENSG00000259359	<i>RP11-327J17.2</i>	15	q26.2	Cross-Tissue	-0.16	-	5.82E-07	1.28E-02
ENSG00000155719	<i>OTOA</i>	16	p12.2	Prostate	-4.82	-5.36	1.44E-06	1.99E-02
ENSG00000197445	<i>C16orf47</i>	16	q22.3	Cross-Tissue	0.44	-	1.11E-07	2.47E-03
ENSG00000141052	<i>MYOCD</i>	17	p12	Cross-Tissue	-3.03	-	1.20E-07	2.66E-03
ENSG00000181513	<i>ACBD4</i>	17	q21.31	Cross-Tissue	-1.05	-	2.17E-06	4.77E-02
ENSG00000186834	<i>HEXIM1</i>	17	q21.31	Prostate	5.03	0.12	4.95E-07	6.85E-03

Abbreviations: Chr – chromosome; p-bonf – Bonferroni-adjusted p-value

* 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 genome-wide association study of PSA levels.

[†] GRCh38

[~] Indicates the tissue(s) in which predicted gene expression was significantly associated with PSA levels.

598 **Table 2.** Gene level colocalization results for significant genes from the cross-tissue
 599 transcriptome-wide association study*

Ensembl gene ID	Gene name	Chr	Cytoband[†]	GRCP	GLCP
ENSG0000069275	<i>NUCKS1</i>	1	q32.1	0.97	2.55
ENSG00000117280	<i>RAB29</i>	1	q32.1	1.13	0.52
ENSG00000117602	<i>RCAN3</i>	1	p36.11	2.13	-24.54
ENSG00000143603	<i>KCNN3</i>	1	q21.3	1.42	-136.50
ENSG00000158715	<i>SLC45A3</i>	1	q32.1	0.59	15.17
ENSG00000163344	<i>PMVK</i>	1	q21.3	1.97	-50.77
ENSG00000163346	<i>PBXIP1</i>	1	q21.3	1.99	-14.21
ENSG00000270361	<i>RP11-307C12.13</i>	1	q21.3	1.08	-0.32
ENSG00000115507	<i>OTX1</i>	2	p15	4.91	-7.33
ENSG00000143869	<i>GDF7</i>	2	p24.1	1.00	620.30
ENSG00000068885	<i>IFT80</i>	3	q25.33	2.41	24.05
ENSG00000072858	<i>SIDT1</i>	3	q13.2	0.62	-211.30
ENSG00000113810	<i>SMC4</i>	3	q25.33	2.18	-52.64
ENSG00000132394	<i>EEFSEC</i>	3	q21.3	1.34	-160.40
ENSG00000175792	<i>RUVBL1</i>	3	q21.3	1.24	-72.12
ENSG00000177311	<i>ZBTB38</i>	3	q23	1.22	-39.66
ENSG00000186432	<i>KPNA4</i>	3	q25.33	1.27	-19.18
ENSG00000213186	<i>TRIM59</i>	3	q25.33	1.54	-38.26
ENSG00000123737	<i>EXOSC9</i>	4	q27	2.24	-108.20
ENSG00000145386	<i>CCNA2</i>	4	q27	1.39	-54.29
ENSG00000049656	<i>CLPTM1L</i>	5	p15.33	0.71	-27.67
ENSG00000164362	<i>TERT</i>	5	p15.33	0.72	4.54
ENSG00000184357	<i>HIST1H1B</i>	6	p22.1	0.92	2.70
ENSG00000189298	<i>ZKSCAN3</i>	6	p22.1	0.84	1.28
ENSG00000204713	<i>TRIM27</i>	6	p22.1	0.76	-12.84
ENSG00000233822	<i>HIST1H2BN</i>	6	p22.1	1.08	-9.88
ENSG00000281831	<i>HCP5B</i>	6	p22.1	0.88	1.25
ENSG00000164684	<i>ZNF704</i>	8	q21.13	1.00	-94.45
ENSG00000167034	<i>NKX3-1</i>	8	p21.2	0.96	-293.10
ENSG00000056558	<i>TRAF1</i>	9	q33.2	0.78	5,137.00
ENSG00000095261	<i>PSMD5</i>	9	q33.2	0.61	-14.25
ENSG00000119403	<i>PHF19</i>	9	q33.2	0.97	-23.73
ENSG00000130956	<i>HABP4</i>	9	q22.32	0.82	30.20
ENSG00000130958	<i>SLC35D2</i>	9	q22.32	1.22	-11.44
ENSG00000165244	<i>ZNF367</i>	9	q22.32	0.60	-10.13
ENSG00000148700	<i>ADD3</i>	10	q25.1	1.21	-118.60
ENSG00000272447	<i>RP11-182L21.6</i>	10	q22.3	0.99	5.94
ENSG00000137673	<i>MMP7</i>	11	q22.2	0.96	183.70
ENSG00000158636	<i>EMSY</i>	11	q13.5	2.64	-9.56
ENSG00000177951	<i>BETIL</i>	11	p15.5	0.78	-63.92
ENSG00000111707	<i>SUDS3</i>	12	q24.23	1.33	-225.60
ENSG00000175899	<i>A2M</i>	12	p13.31	0.64	-156.60
ENSG00000172766	<i>NAA16</i>	13	q14.11	1.25	-14.31
ENSG00000259359	<i>RP11-327J17.2</i>	15	q26.2	1.10	2.93

ENSG00000141052	<i>MYOCD</i>	17	p12	0.62	-1.94
ENSG00000275410	<i>HNF1B</i>	17	q12	0.81	2.32
ENSG00000176920	<i>FUT2</i>	19	q13.33	3.04	320.20
ENSG00000177045	<i>SIX5</i>	19	q13.32	1.87	-15.39
ENSG00000185800	<i>DMWD</i>	19	q13.32	1.00	461.40
ENSG00000185022	<i>MAFF</i>	22	q13.1	3.63	3.06
Abbreviations: Chr – chromosome; GLCP - gene-level colocalization probability; GRCP - gene-level variant colocalization probability * The genes in bold were, based on COJO analyses, independent from the largest prior genome-wide association study of prostate-specific antigen levels. † GRCh38					

600 **Table 3.** Results from pathway analyses based on all significant genes identified in the whole blood,
 601 prostate tissue, and cross-tissue transcriptome-wide association studies

Database	Pathway	Overlap	p	Score	Genes
Whole Blood					
KEGG	renin-angiotensin system	2/23	0.047	335.87	<i>KLK1;KLK2</i>
Cross-Tissue					
GO Molecular Function	RNA polymerase II transcription regulatory region sequence-specific DNA binding (GO:0000977)	26/1359	6.82E-04	30.85	<i>EHF;ZBTB4;ETS2;GLIS2;HOXA10;POU5F1B;PGBD1;HOXA9;SIX5;ZNF704;ZKSCAN3;OTX1;ZBTB7B;ZNF367;E2F7;NKX3-1;NKX2-6;BCL11A;ZBTB38;HNF1B;TBX3;DMRTA2;TFAP4;MAFF;MKX;TP53</i>
GO Molecular Function	sequence-specific double-stranded DNA binding (GO:1990837)	14/712	0.024	18.07	<i>ETS2;TBX3;GLIS2;HOXA10;DMRTA2;HOXA9;TFAP4;ZNF704;MAFF;OTX1;ZBTB7B;TP53;E2F7;NKX3-1</i>
GO Molecular Function	double-stranded DNA binding (GO:0003690)	14/651	0.013	22.45	<i>NUCKS1;ETS2;TBX3;GLIS2;HOXA10;DMRTA2;HOXA9;TFAP4;ZNF704;MAFF;OTX1;ZBTB7B;E2F7;NKX3-1</i>
GO Molecular Function	RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978)	26/1149	6.54E-05	47.24	<i>EHF;ZBTB4;ETS2;GLIS2;HOXA10;POU5F1B;PGBD1;HOXA9;SIX5;ZNF704;ZKSCAN3;OTX1;ZBTB7B;ZNF367;E2F7;NKX3-1;NKX2-6;BCL11A;ZBTB38;HNF1B;TBX3;DMRTA2;TFAP4;MAFF;MKX;TP53</i>
GO Molecular Function	cis-regulatory region sequence-specific DNA binding (GO:0000987)	27/1149	3.69E-05	53.95	<i>EHF;ZBTB4;ETS2;GLIS2;HOXA10;POU5F1B;PGBD1;HOXA9;SIX5;ZNF704;ZKSCAN3;OTX1;ZBTB7B;ZNF367;E2F7;NKX3-1;NKX2-6;BCL11A;ZBTB38;HMGA1;HNF1B;TBX3;DMRTA2;TFAP4;MAFF;MKX;TP53</i>
GO Molecular Function	sequence-specific DNA binding (GO:0043565)	17/707	0.0012	35.01	<i>ZBTB38;ZBTB4;ETS2;TBX3;GLIS2;HOXA10;DMRTA2;HOXA9;TFAP4;TERT;ZNF704;MAFF;ZKSCAN3;OTX1;ZBTB7B;E2F7;NKX3-1</i>
GO Molecular Function	endopeptidase activity (GO:0004175)	8/315	0.049	19.86	<i>MMP7;KLK1;ADAMTS18;KLK5;KLK3;KLK15;KLK2;KLK8</i>
GO Molecular Function	histone deacetylase binding (GO:0042826)	5/1993	0.016	53.49	<i>HOXA10;TFAP4;SUDS3;TP53;NKX3-1</i>
GO Molecular Function	serine-type endopeptidase activity (GO:0004252)	6/105	0.0047	69.79	<i>KLK1;KLK5;KLK3;KLK15;KLK2;KLK8</i>
GO Molecular Function	serine-type peptidase activity (GO:0008236)	8/125	3.30E-04	110.44	<i>KLK1;KLK5;KLK4;KLK3;KLK15;KLK2;KLK8;KLK11</i>
GO Molecular Function	thioesterase binding (GO:0031996)	2/11	0.049	166.1	<i>CDC42;TRAF1</i>
GO Biological Process	regulation of transcription by RNA polymerase II (GO:0006357)	36/2206	0.0064	28.19	<i>EHF;NUCKS1;ZBTB4;ETS2;GLIS2;HOXA10;POU5F1B;PGBD1;SDR16C5;HOXA9;ZNF827;SIX5;ZNF704;RUVBL1;ZKSCAN3;TRIM27;OTX1;SUDS3;ZBTB7B;ZNF367;E2F7;NKX3-1;MYOCD;NKX2-6;BCL11A;ZBTB38;HMGA1;HNF1B;TBX3;DMRTA2;TFAP4;MAFF;MKX;TP53;TLR4;FGFR2</i>
GO Biological	regulation of transcription,	38/2244	0.0028	33.66	<i>EHF;NUCKS1;PBXIP1;ZBTB4;ETS2;GLIS</i>

Process	DNA-templated (GO:0006355)				<i>2;HOXA10;POU5F1B;PGBD1;HOXA9;SIX5;ZNF704;RUVBL1;ZKSCAN3;GPBP1L1;OTX1;SUDS3;ZBTB7B;JARID2;ZNF367;E2F7;NKX3-1;MYOCD;NKX2-6;BCL11A;ZBTB38;SETDB2;HMGA1;HNF1B;TBX3;GDF7;DMRTA2;TFAP4;MAFF;NAA16;MKX;TP53;PHF19</i>
GO Biological Process	positive regulation of transcription, DNA-templated (GO:0045893)	22/1183	0.027	24.05	<i>EHF;MYOCD;NKX2-6;ZBTB38;HMGA1;HNF1B;ETS2;TBX3;GLIS2;GDF7;HOXA10;TFAP4;ZNF827;MAFF;ZKSCAN3;NAA16;OTX1;TP53;TLR4;FGFR2;E2F7;NKX3-1</i>
GO Biological Process	positive regulation of transcription by RNA polymerase II (GO:0045944)	18/908	0.032	23.3	<i>EHF;MYOCD;NKX2-6;ZBTB38;HMGA1;HNF1B;TBX3;GLIS2;HOXA10;TERT;ZNF827;MAFF;OTX1;TP53;TLR4;FGFR2;E2F7;NKX3-1</i>
GO Biological Process	negative regulation of transcription, DNA-templated (GO:0045892)	20/948	0.018	30.19	<i>BCL11A;ZBTB38;SETDB2;HMGA1;PBXIP1;ZBTB4;ETS2;TBX3;GLIS2;SDR16C5;TFAP4;ZKSCAN3;TRIM27;SUDS3;ZBTB7B;JARID2;TP53;FGFR2;E2F7;NKX3-1</i>
GO Biological Process	negative regulation of cellular macromolecule biosynthetic process (GO:2000113)	13/547	0.041	26.39	<i>ZBTB38;SETDB2;HMGA1;PBXIP1;ZBTB4;TBX3;GLIS2;TFAP4;ZKSCAN3;SUDS3;JARID2;TP53;NKX3-1</i>
GO Biological Process	negative regulation of nucleic acid-templated transcription (GO:1903507)	13/464	0.021	37.74	<i>ZBTB38;SETDB2;HMGA1;PBXIP1;ZBTB4;TBX3;GLIS2;TFAP4;ZKSCAN3;SUDS3;JARID2;TP53;NKX3-1</i>
GO Biological Process	regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0000079)	5/82	0.047	66.09	<i>CCNA2;MYOCD;CDKN2C;TFAP4;CDK5RAP1</i>
GO Biological Process	cellular component disassembly (GO:0022411)	5/66	0.027	94.49	<i>MMP7;KLK5;KLK4;KLK2;A2M</i>
GO Biological Process	extracellular matrix disassembly (GO:0022617)	5/66	0.027	94.49	<i>MMP7;KLK5;KLK4;KLK2;A2M</i>
GO Biological Process	negative regulation of cyclin-dependent protein serine/threonine kinase activity (GO:0045736)	3/20	0.047	176.28	<i>MYOCD;TFAP4;CDK5RAP1</i>
GO Biological Process	regulation of systemic arterial blood pressure (GO:0003073)	3/16	0.032	251.29	<i>KLK1;KLK3;KLK2</i>
GO Molecular Function	RNA polymerase II transcription regulatory region sequence-specific DNA binding (GO:0000977)	26/1359	6.82E-04	30.85	<i>EHF;ZBTB4;ETS2;GLIS2;HOXA10;POU5F1B;PGBD1;HOXA9;SIX5;ZNF704;ZKSCAN3;OTX1;ZBTB7B;ZNF367;E2F7;NKX3-1;NKX2-6;BCL11A;ZBTB38;HNF1B;TBX3;DMRTA2;TFAP4;MAFF;MKX;TP53</i>
Abbreviations: p – adjusted p-value; Score – combined score					

Figure 1: Miami Plot for Whole Blood TWAS Analysis

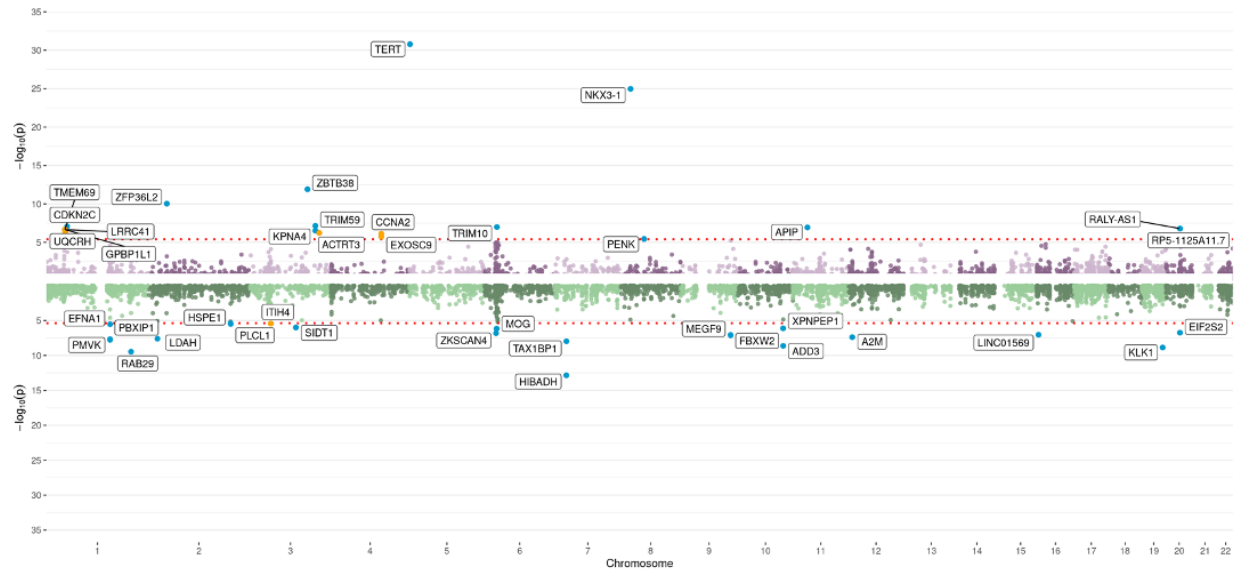


Figure 2: Miami Plot for Prostate Tissue TWAS Analysis

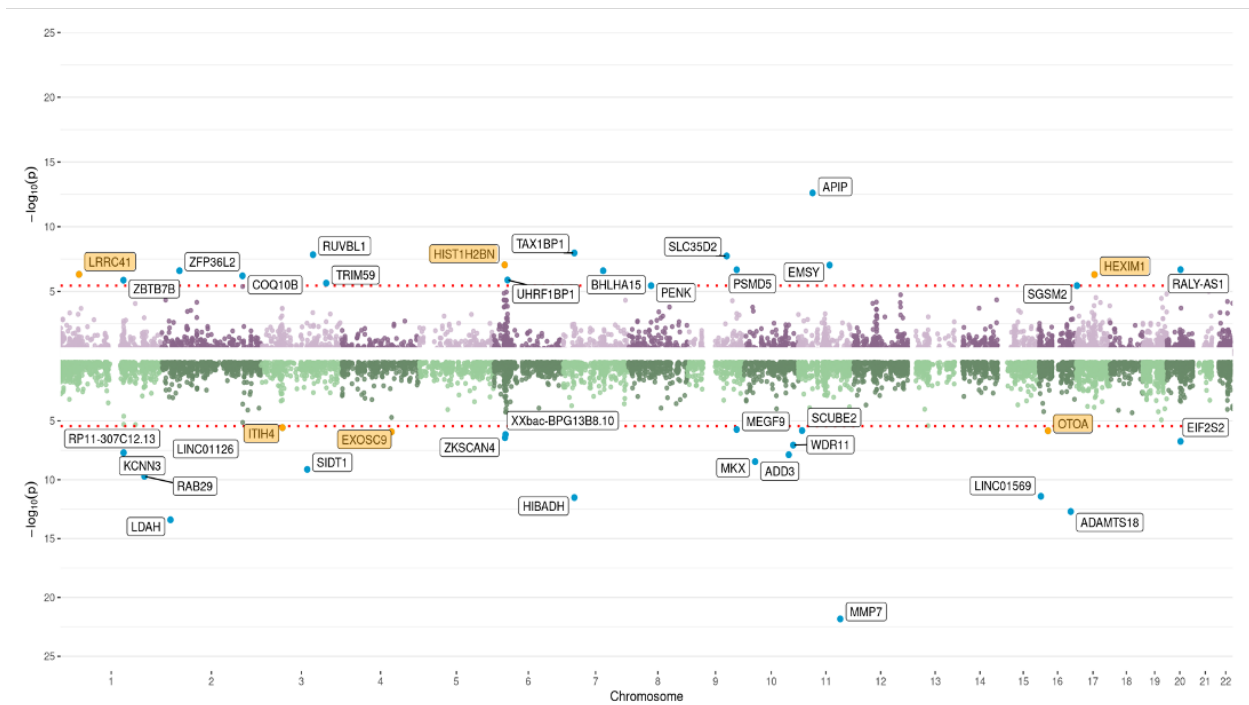


Figure 3: Manhattan plot for cross-tissue analyses

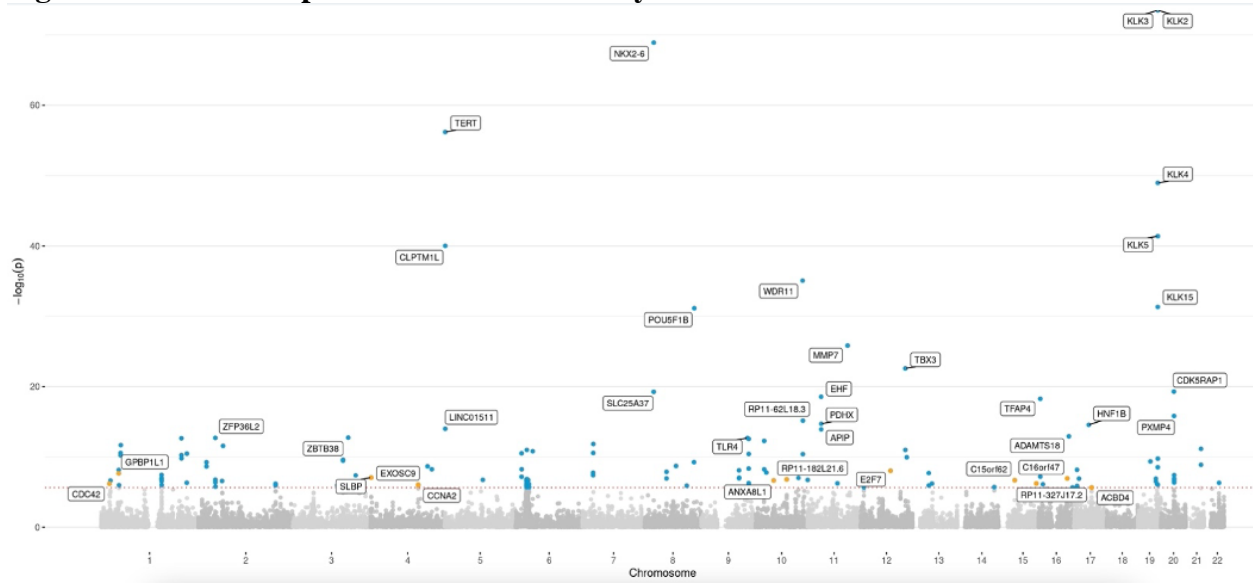


Figure 4. Statistically significant genes overlapping across TWAS analyses.

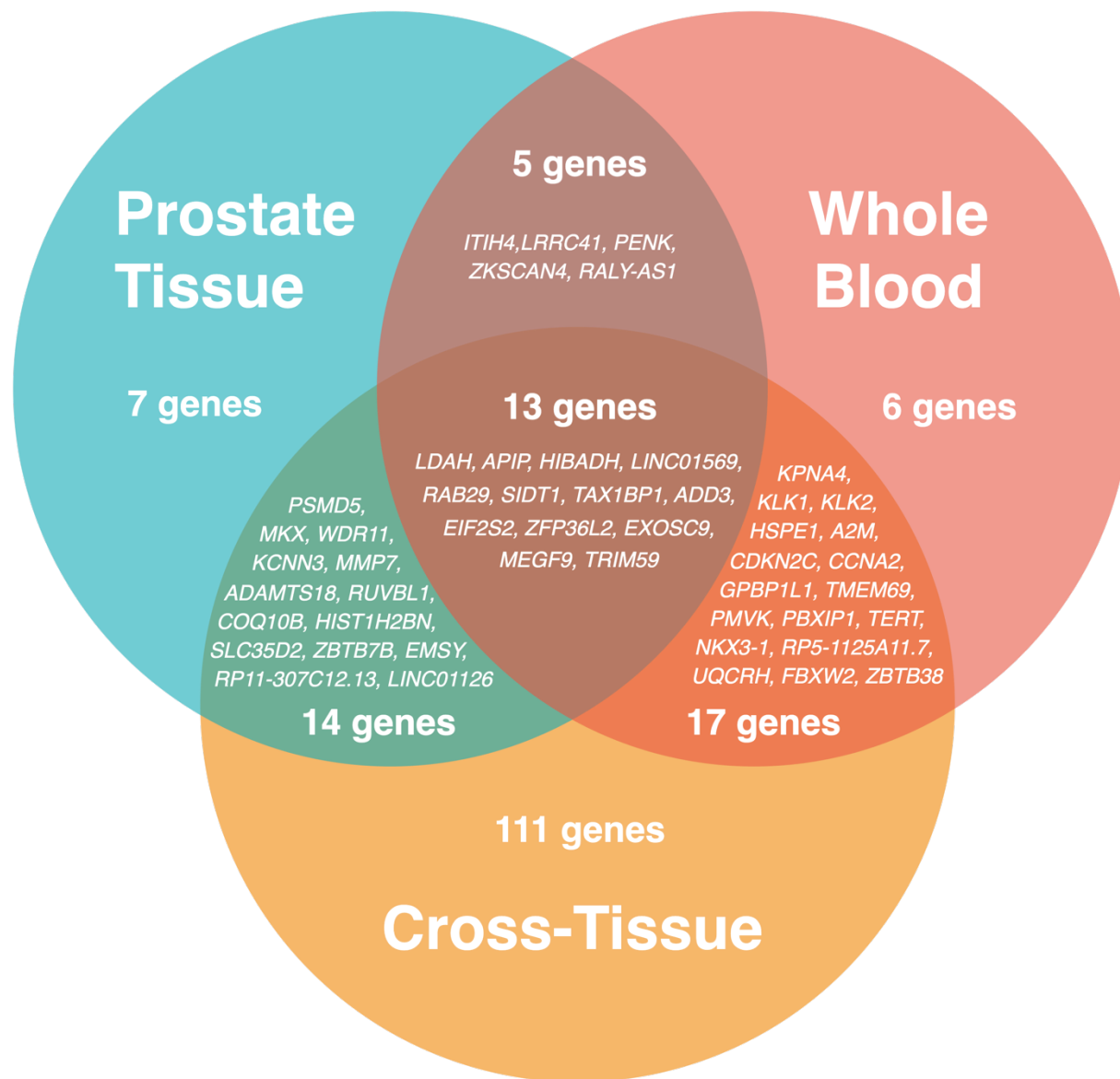


Figure 5: GO term enrichment results from the cross-tissue TWAS.

