

# Proteome-wide association study of prostate cancer risk across populations

Received: 28 April 2025

Accepted: 4 November 2025

Published online: 06 December 2025

 Check for updates

Hua Zhong<sup>1,2,17</sup>, Jingjing Zhu<sup>2,3,17</sup>, Shuai Liu<sup>1,2</sup>, Chong Wu<sup>4</sup>, Liang Wang<sup>5</sup>, Seamus P. Whelton<sup>6</sup>, Catherine H. Marshall<sup>7</sup>, Michael J. Blaha<sup>8</sup>, Peter Durda<sup>9</sup>, Xiuqing Guo<sup>10</sup>, Craig W. Johnson<sup>11</sup>, Henry J. Lin<sup>10</sup>, Kent D. Taylor<sup>10</sup>, Russell P. Tracy<sup>9</sup>, Ronit I. Yarden<sup>12</sup>, Ani W. Manichaikul<sup>13</sup>, Stephen S. Rich<sup>13</sup>, Jerome I. Rotter<sup>10</sup>, Rajat Deo<sup>14</sup>, Ruth F. Dubin<sup>15</sup>, Peter Ganz<sup>16</sup> & Lang Wu<sup>1,2</sup> ✉

There is insufficient understanding of the molecular basis of prostate cancer (PCa) across different populations. We perform a large-scale proteome-wide association study (PWAS) to identify proteins with genetically regulated expression in plasma to be associated with PCa risk across populations. We develop genetic prediction models for expression of 1578, 1993, 1218, and 1390 proteins for African (n = 450), European (n = 758), Asian (n = 289), and Hispanic/Latino (n = 474) males, respectively, and evaluate associations of genetically regulated protein expression with PCa risk in 19,391 PCa cases and 61,608 controls of African population, 122,188 cases and 604,640 controls of European population, 10,809 cases and 95,790 controls of Asian population, and 3931 cases and 26,405 controls of Hispanic/Latino population. We identify three, four, 15, and 73 PCa-associated proteins in African, Hispanic/Latino, Asian, and European populations, respectively, and 83 in trans-population meta-analysis. There are both pan-population and population-specific associations. Our findings provide valuable insights into etiology of PCa.

Prostate cancer (PCa) is the most frequently diagnosed cancer among men in 118 countries/territories and the leading cause of cancer death in 52 countries<sup>1</sup>. There is a critical need to better understand its etiology for developing improved therapeutic and risk assessment strategies. Genetic factors have been demonstrated to play an important role in PCa etiology. It is well established that there is a significantly increased risk of PCa among men with a family history of the disease<sup>2</sup>. Genome-wide association studies (GWAS) have contributed significantly to our understanding of the heritability and familial risk of PCa<sup>3,4</sup>. Numerous genomic loci associated with PCa risk have been identified in GWAS<sup>3,4</sup>, suggesting a multigenic model of prostate tumorigenesis. On the other hand, the underlying biological mechanisms for a majority of the GWAS-identified risk loci remain unclear. Given the constraints inherent in conventional single-marker-based GWAS<sup>5</sup>, there has been a significant shift towards elucidating the functional implications underlying risk loci in the post-GWAS era. This

emphasis aims to accurately delineate the exact biological mechanisms, particularly in post-translational processes through which the identified single nucleotide polymorphism (SNP)s and target genes exert their effects<sup>6–11</sup>.

It is known that the proteome is commonly dysregulated during the development of diseases. Several proteins have been reported to be associated with PCa risk for their measured levels in blood, such as KLK11 (kallikrein related peptidase 11)<sup>12</sup>, IL-6 (interleukin 6)<sup>13</sup>, and EPCA (early PCa antigen)<sup>14</sup>. However, the findings are not always consistent<sup>15,16</sup>. Conventional epidemiological studies could confer several limitations, such as selection bias, potential confounding, and reverse causation, which may account for some of the inconsistent results observed. One alternative strategy is to evaluate the associations between genetically predicted protein levels and PCa risk. To identify proteins potentially relevant to PCa, we have comprehensively evaluated the relationship between genetically predicted protein

A full list of affiliations appears at the end of the paper. ✉ e-mail: [lwu3@lsuhsc.edu](mailto:lwu3@lsuhsc.edu)

levels and PCa risk in European population<sup>17,18</sup> and identified a potential link between several protein biomarkers and PCa risk, including Laminin, IL-21, and HOXB13 (homeobox B13)<sup>18,19</sup>.

Racial and ethnic background is a well-established risk factor for PCa<sup>20,21</sup>. PCa incidence rates vary significantly by regions, with the lowest rates observed in South Central Asia (6.4 age-standardized rate per 100,000), and markedly higher rates in Northern Europe (82.8 age-standardized rate per 100,000)<sup>1</sup>. Meanwhile, PCa mortality rates among men of African population are two to four times higher than those observed in men of other racial and ethnic groups<sup>1,22</sup>. Biological differences in PCa have also been observed across populations. For instance, the prostate-specific Ca<sup>2+</sup>-dependent chloride ion channel protein<sup>23</sup>, ANO7 (anoctamin 7), has been identified as a population-relevant protein-altering PCa-risk locus<sup>24,25</sup>. Two variants in the ANO7 gene, rs74804606(p.Ile740Leu) and rs60985508 (p.Ser914\*), have been identified to be associated with PCa in men of African ancestry<sup>25</sup>. In contrast, four European-specific PCa-risk variants in ANO7, rs77559646, rs2074840, p.Ala759Thr (rs76832527), and p.Glu226Lys/p.Glu226\* (rs77482050), were largely excluded from analyses in African, East Asian, and Hispanic populations due to low frequency. Additionally, African-American men are known to exhibit higher expression of gene sets involved in the immune response, apoptosis, hypoxia, and reactive oxygen species production<sup>26</sup>. Despite these well-established differences, a comprehensive characterization of population-specific proteomic heterogeneity in PCa remains lacking, as most existing studies have focused primarily on men of European descent<sup>19,27,28</sup>. Although a previous study has examined population-specific exosomal proteins, its findings were limited by a very small sample size ( $n = 12$  patients and nine controls) and a lack of adjustment for multiple comparisons<sup>29</sup>.

In this work, we leverage a very large and comprehensive reference dataset of 7367 plasma proteins measured across four diverse populations – 450 African, 758 European, 289 Asian, and 474 Hispanic/Latino men – to establish population-specific genetic prediction models. We further analyze GWAS summary statistics for PCa risk, involving men in the African population (19,391 cases and 61,608

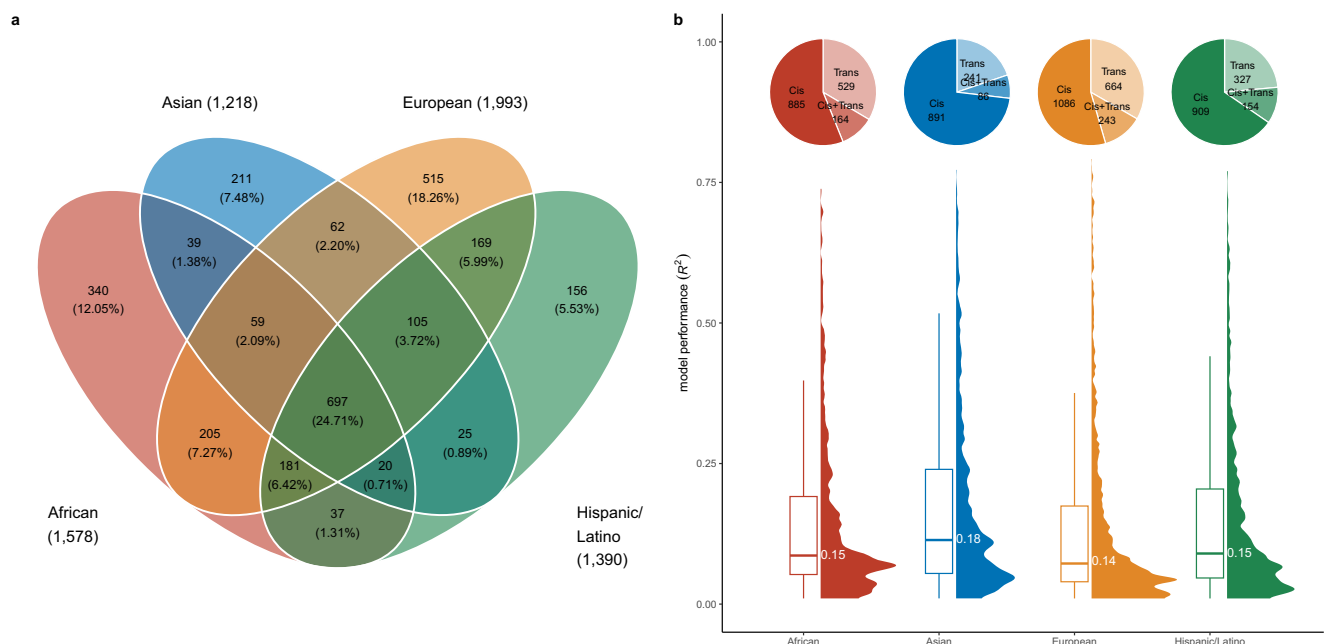
controls), European population (122,188 cases and 604,640 controls), Asian population (10,809 cases and 95,790 controls), and Hispanic/Latino population (3931 cases and 26,405 controls) to identify proteins whose genetically predicted levels are associated with PCa risk. Our study demonstrates the value of large-scale, diverse omics data for understanding the etiology of PCa to reduce its health disparities.

## Results

### Building population-specific protein genetic prediction models

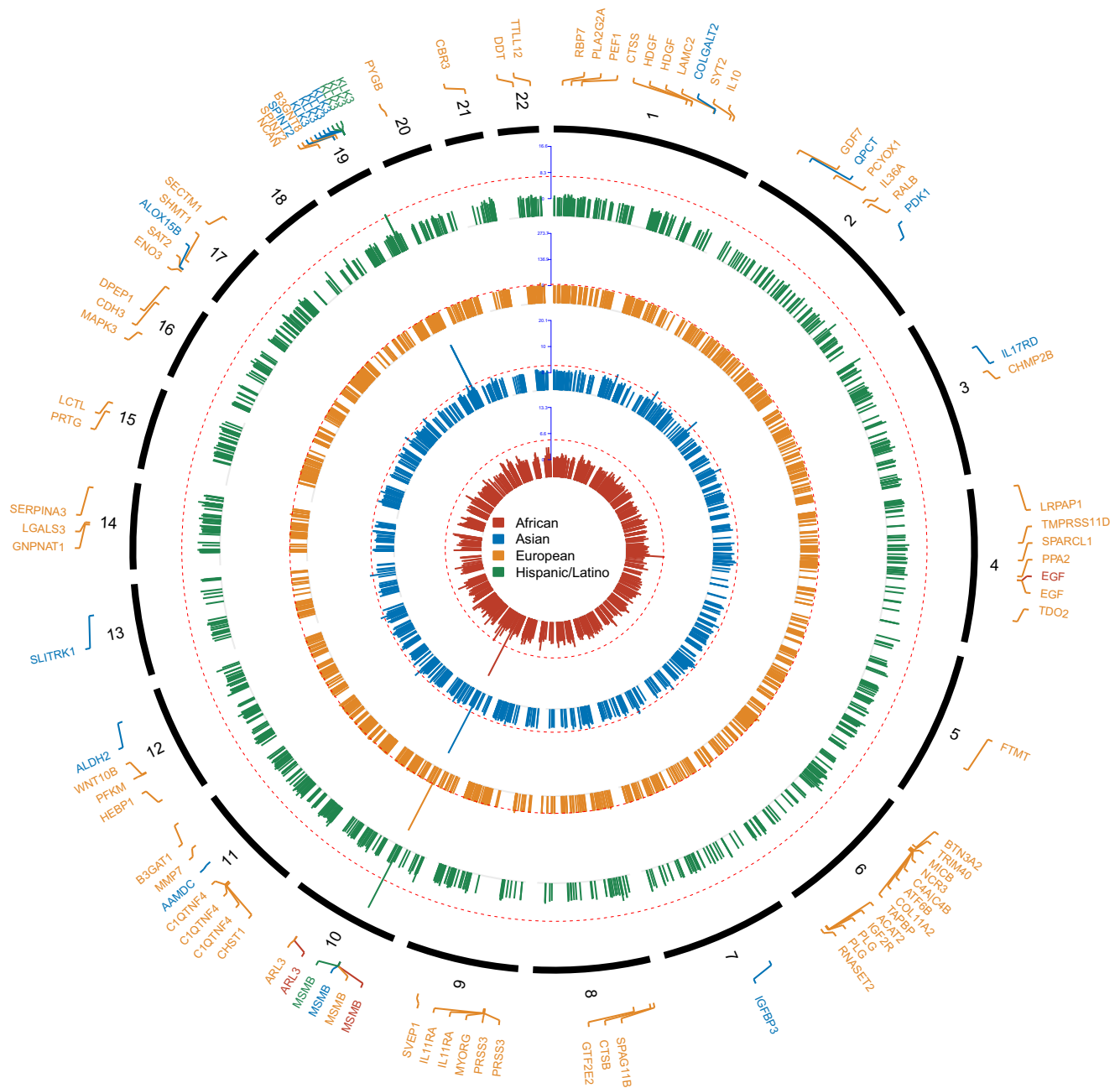
In this study, we first identified protein quantitative trait loci (pQTLs) for 450 African, 758 European, 289 Asian, and 474 Hispanic/Latino men residing in the USA in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort. Among the 6999 proteins examined, we identified at least one *cis*-pQTL (within  $\pm 100$  Kb of the transcription start site) associated at false discovery rate (FDR)  $< 0.05$  and/or one *trans*-pQTL associated at  $P < 5 \times 10^{-9}$  for 1609 proteins in the African population, 2058 in the European population, 1237 in the Asian population, and 1415 in the Hispanic/Latino population. We further established prediction models for 1578 proteins in the African population, 1993 in the European population, 1218 in the Asian population, and 1390 in the Hispanic/Latino population, each with an  $R^2 \geq 0.01$ , with 697 proteins overlapping across all genetic ancestries (Fig. 1a and Supplementary Data 1). Among the established prediction models, population-specific models were developed for 340 proteins in Africans, 515 in Europeans, 211 in Asians, and 156 in Hispanics/Latinos. The model performance  $R^2$  varies across populations, ranging from 0.01 to 0.74 in Africans, 0.01 to 0.79 in Europeans, 0.01 to 0.77 in Asians, and 0.01 to 0.77 in Hispanic/Latinos, with Asians exhibiting the highest mean  $R^2$  of 0.18 (Fig. 1b). Kruskal–Wallis test revealed a significant difference in  $R^2$  values across the four groups ( $\chi^2 = 98.39$ ,  $df = 3$ ,  $P < 2.20 \times 10^{-16}$ ). Post hoc Dunn's tests with Bonferroni correction confirmed that  $R^2$  was significantly higher in Asians compared with Africans ( $P = 2.57 \times 10^{-5}$ ), Europeans ( $P = 7.81 \times 10^{-22}$ ), and Hispanic/Latinos ( $P = 1.34 \times 10^{-5}$ ). It is worth noting that the Asian group was the smallest MESA population.

We further validated the European population prediction models using INTERVAL data in an external validation<sup>30</sup>. Due to differences



**Fig. 1 | Protein prediction models across populations. a** Venn plot of protein prediction models across populations. **b** Distribution and performance of protein prediction models in different populations. The pie charts show the proportion of modeled proteins categorized by the type of genetic variants used. The violin plot

illustrates the distribution of  $R^2$  values for protein models. The horizontal line within the box plot represents the mean value of  $R^2$ . The box boundaries define the interquartile range (IQR, 25th to 75th percentiles), and whiskers extend to the most extreme points within  $1.5 \times$  IQR. Source data are provided in the Source Data file.



**Fig. 2 | Manhattan plot of identified proteins associated with prostate cancer (PCa) risk.** Each dot represents the *P*-value for the association between genetically predicted protein abundances and PCa risk, plotted by genomic position on the x-axis. Red dashed lines indicated the significant threshold at False Discovery Rate

(FDR) < 0.05 for each population: African (19,391 PCa cases and 61,608 controls), European (122,188 cases and 604,640 controls), Asian (10,809 cases and 95,790 controls), and Hispanic/Latino (3931 cases and 26,405 controls). Source data are provided in the Source Data file.

between Somalogic’s SomaScan platform versions, 844 proteins were available for analysis. Of these testable models, 584 (69.19%) exhibited a prediction performance of  $R^2 \geq 0.01$  in the external validation (Supplementary Fig. 1).

**Identification of proteins associated with PCa risk**

To identify proteins associated with PCa risk in each population of interest, we applied our established prediction models to the large-scale PCa GWAS summary statistics generated from 156,391 PCa cases and 666,248 controls across diverse populations. We identified significant associations (FDR < 0.05) for three proteins in the African population, namely, microseminoprotein- $\beta$  (encoded by *MSMB*),

epidermal growth factor:extracellular domain (encoded by *EGF*), and ADP-ribosylation factor-like protein 3 (encoded by *ARL3*). In addition, 73 proteins in Europeans, 15 in Asians, and four in Hispanic/Latinos were significantly associated with PCa risk at FDR < 0.05 (Fig. 2 and Supplementary Data 2). Among them, microseminoprotein- $\beta$  was detected across all four populations, and two proteins (ADP-ribosylation factor-like protein 3 and epidermal growth factor:extracellular domain) were identified in both African and European populations. Two unique proteins, prostate-specific antigen (PSA) and benign prostate-specific antigen (BPSA) encoded by *KLK3* (kallikrein related peptidase 3) were shared between Asian and Hispanic/Latino populations. A total of 60 proteins were unique to European population

(FDR < 0.05 in European and raw  $p > 0.05$  in all other populations) and eight proteins tended to be Asian-specific (Supplementary Data 2).

A trans-population meta-analysis was conducted to identify proteins associated with PCa risk. A total of 92 proteins encoded by 83 genes were significantly associated with PCa risk at FDR < 0.05. Among them, 45 demonstrated inverse protein-PCa associations and 47 showed associations between higher predicted levels and increased PCa risk. Eighteen proteins not detected in population-specific analyses were identified in this meta-analysis (Supplementary Data 2). Based on a heterogeneity  $p$ -value (HetPval) of < 0.05 or  $I^2$  heterogeneity statistic (HetISq) > 75%, we found significant heterogeneity across populations for 10 unique proteins, including microseminoprotein- $\beta$ , plasminogen, Charged multivesicular body protein 2b, Cathepsin S, Epidermal growth factor:Extracellular domain, tripartite motif-containing protein 40, Trypsin-3, Interleukin-36 alpha, PSA, Mth938 domain-containing protein, and UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 8 (Supplementary Data 2).

Through population-specific proteome-wide association study (PWAS) and trans-population meta-analysis, we finally identified 96 protein-PCa risk associations, corresponding to 104 SOMAscan protein IDs. We then compared these 96 associated proteins with independent risk loci previously reported in GWAS<sup>4,24</sup>. Of these, 92 proteins were located within 500 kb of previously reported lead SNPs. The remaining four proteins (SLIT and NTRK-like protein 1, toll-like receptor 3, synaptotagmin-2, and complement factor H-related protein 1) were located more than 500 kb away from known loci (Supplementary Data 2).

### Two-stage constrained maximum likelihood (2ScML) robustness analysis

To assess the robustness of our results, we applied the 2ScML method under relaxed instrumental variable assumptions to 95 population-specific associations spanning 79 unique proteins with 86 SOMAscan IDs (three in Africans, 73 in Europeans, 15 in Asians, and four in Hispanic/Latinos). As shown in Supplementary Data 3, approximately 44.21% (42 of 95) of the associations remained significant at FDR < 0.05 with consistent effect directions. Except for PSA, detected in both Asian and Hispanic/Latino populations, the remaining 40 proteins were population-specific. Of these, 36 proteins tend to be specific to the Europeans and four were uniquely linked to PCa risk in the Asian population (aldehyde dehydrogenase, mitochondrial, arachidonate 15-lipoxygenase B, insulin-like growth factor-binding protein 3, and SLIT and NTRK-like protein 1).

### Tissue expression, functional enrichment, and network analysis of identified proteins

We evaluated the tissue-specific expression of 94 genes coding the 96 PWAS-identified proteins using RNA data from the Human Protein Atlas (HPA). Among these, three genes (*KLK3*, *MSMB*, and *ALOX15B*) exhibited prostate-enriched expression. These well-known PCa markers<sup>31–33</sup> validate our gene selection strategy. An additional 56 genes showed higher expression in other tissues but remained detectable in the prostate, many previously implicated in PCa-related pathways, including growth factor signaling (e.g., *EGF*<sup>34</sup>, *IGFBP3*<sup>35</sup>), extracellular matrix remodeling and metastasis (e.g., *PRSS3*<sup>36</sup>, *MMP7*<sup>37</sup>), and inflammatory regulation (e.g., *SOCS3*<sup>38</sup>).

Our Gene Ontology (GO) enrichment analysis of the 94 candidate genes revealed significant overrepresentation in relevant biological processes, including *insulin-like growth factor receptor signaling pathway* (GO:0048009,  $P = 2.92 \times 10^{-5}$ ) and *extracellular matrix disassembly* (GO:0022617,  $P = 2.39 \times 10^{-4}$ ), both associated with tumor progression<sup>39,40</sup> and metastatic potential<sup>41</sup> in PCa. In the Molecular Function category, enriched terms included *serine-type endopeptidase activity* (GO:0004252,  $P = 2.70 \times 10^{-5}$ ), *serine-type peptidase activity* (GO:0008236,  $P = 4.90 \times 10^{-5}$ ), *growth factor receptor binding*

(GO:0070851,  $P = 7.63 \times 10^{-5}$ ), and *cytokine activity* (GO:0005125,  $P = 1.17 \times 10^{-3}$ ), all relevant to tumor proliferation and microenvironmental signaling<sup>42–44</sup>. Cellular Component analysis indicated predominant localization to the *collagen-containing extracellular matrix* (GO:0062023,  $P = 3.43 \times 10^{-5}$ ), highlighting roles in PCa development and metastasis<sup>45</sup> (Supplementary Data 4). Protein-protein interaction (PPI) analysis identified IL-10 (Interleukin-10) with the highest node degree of 18, followed by Angiostatin with a node degree of 14 (Supplementary Data 5), suggesting both topological centrality and functional significance in PCa.

Subsequent ingenuity pathway analysis (IPA) analysis was performed to explore potential regulatory mechanisms underlying the genes encoding the identified PCa-associated proteins (Supplementary Data 6). Kallikrein-related peptidase 3 (*KLK3*), (mitogen-activated protein kinase 3) (*MAPK3*), and ras-related protein (*RALB*) were significantly enriched in the PCa signaling pathway ( $P = 7.76 \times 10^{-3}$ ). Additionally, several PCa-related biofunctions were identified, involving *DPEP1* (dipeptidase 1), *EGF*, *IGF2R* (insulin-like growth factor 2 receptor), *KLK3*, *LGALS3* (galectin-3), *LHCGR* (luteinizing hormone/chorionic gonadotropin receptor), *MAPK3*, *MSMB*, *PLG* (plasminogen), *SOCS3* (suppressor of cytokine signaling 3), *TLR3* (toll-like receptor 3). These biofunctions encompassed proliferation ( $P = 8.42 \times 10^{-5}$ ), Cell cycle progression ( $P = 5.91 \times 10^{-4}$ ), apoptosis ( $P = 6.16 \times 10^{-4}$ ), binding ( $P = 1.27 \times 10^{-3}$ ), PCa pathogenesis ( $P = 2.94 \times 10^{-3}$ ), G0/G1 phase transition ( $P = 3.48 \times 10^{-3}$ ), atrophy of prostate gland ( $P = 3.48 \times 10^{-3}$ ), and hereditary PCa ( $P = 3.48 \times 10^{-3}$ ) (Supplementary Data 6).

### Candidate drugs targeting identified associated proteins

To explore potential drug repurposing opportunities for the 79 uniquely identified proteins, we curated the DrugBank database and found 31 proteins as targets of FDA-approved drugs for various human diseases (Supplementary Data 7). Of these, 25 proteins were further supported as relevant to PCa based on a positive *overallAssociationScore* from the OpenTargets platform (Supplementary Data 7). Notably, the lutropin-choriogonadotropic hormone receptor (LSHR) is targeted by goserelin (DrugBank ID: DB00014), a synthetic analog of luteinizing hormone-releasing hormone used clinically to treat both breast cancer and PCa by reducing the secretion of pituitary gonadotropins<sup>46,47</sup>. Additionally, aldehyde dehydrogenase, mitochondrial (ALDH-E2) is targeted by disulfiram (DrugBank ID: DB00822), an FDA-approved drug for treating chronic alcoholism. In preclinical studies, disulfiram, along with delivering copper, targets and eradicates copper-laden PCa cells in animal models, while leaving non-cancer cells intact<sup>48</sup>. Our findings suggest promising opportunities for repurposing these drugs as potential treatments for PCa (Supplementary Data 7).

### Discussion

In this study, we developed comprehensive protein prediction models by integrating genetic and protein data collected from 1971 independent males representing diverse racial and ethnic backgrounds, including African, European, Asian, and Hispanic/Latino populations. Interestingly, the models for the Asian population exhibited the highest average cross-validated  $R^2$  value of 0.18, despite this being the smallest subgroup within the MESA cohort. This suggests that, in this population, certain proteins are more strongly regulated by genetic variants, leading to improved predictive performance at specific loci. This hypothesis is further supported by the Kruskal–Wallis test, which revealed significant differences in heritability ( $H^2$ ) values across the four groups ( $\chi^2 = 156.89$ ,  $df = 3$ ,  $P < 2.20 \times 10^{-16}$ ). In post hoc Dunn's tests,  $H^2$  values were significantly higher in Asians compared with Africans ( $P = 0.02$ ), Europeans ( $P = 4.38 \times 10^{-29}$ ), and Hispanic/Latino participants ( $P = 2.69 \times 10^{-3}$ ). While these results are encouraging, they should be interpreted with caution given the limited sample size. Future studies leveraging larger and more diverse Asian

cohorts will be essential to validate and further investigate these observations.

These race- and ethnic-specific protein prediction models were then applied to GWAS summary statistics for PCa risk in each population of interest. Consequently, we identified three, 73, 15, and four proteins associated with PCa risk in African, European, Asian, and Hispanic/Latino populations, respectively. Furthermore, a trans-population meta-analysis detected 92 proteins that showed significant associations with PCa risk. To assess the robustness of causal inference, we applied the 2ScML method and found that over 40% of the identified population-specific associations remained statistically significant after accounting for potential pleiotropic effects of SNPs. Overall, the predicted abundances of a total of 104 proteins were associated with PCa risk in either population-specific or cross-population analyses. These proteins include targets such as cell adhesion molecules, protease inhibitors, receptors, enzymes, and cytokines, highlighting the multifaceted nature of the biological pathways involved in PCa development.

Among the proteins identified, several and/or their encoding genes have been previously reported to play crucial roles in PCa progression. For example, glucosamine-6 phosphate-N-acetyl Transferase (*GPNATI*) has been implicated in the pathogenesis of castration-resistant PCa<sup>49</sup> through the phosphatidylinositol 3-kinase/protein kinase B (PI3K/Akt) signaling pathway. Additionally, Shah et al. reported that enzalutamide-induced dysregulation of androgen receptor (AR) signaling increases exon-2 inclusion in *PLA2G2A* transcripts in PCa cells<sup>50</sup>. Moreover, PCa risk-associated SNPs rs9364554 and rs7629490 have been linked to the expression of nearby genes *IGF2R* and *CHMP2B*, respectively ( $\pm 500$  kb)<sup>51</sup>. We further compared our results to previously reported proteins associated with PCa risk in PWAS<sup>18,19,52–54</sup> and found 16 overlapping proteins (Supplementary Data 8). Of these, eight encoded by *ARL3*, *B3GNT8*, *CTSS*, *IGF2R*, *MSMB*, *MICB*, *PLG*, and *SPINT2* have been demonstrated to be associated with PCa risk in our earlier studies, which included over 140,000 cases and controls of European descent<sup>18,19</sup>. Among them, *MSMB* stands out as a well-established biomarker for PCa diagnosis and prognosis<sup>32</sup>. Two GWAS have reported the strongest association between PCa risk and rs10993994, located 2 base pairs upstream of the *MSMB* transcription start site<sup>55,56</sup>. This association has been validated in both European<sup>57–59</sup> and Asian populations<sup>60,61</sup>. Furthermore, two proteins identified in this study, encoded by *ATF6B* and *PFKM*, were previously implicated as potential causal genes at PCa susceptibility loci in our earlier transcriptome-wide association study<sup>11</sup>.

In addition to proteins previously reported to be associated with PCa risk, our study also identified several additional proteins that may play important roles in PCa development. While some of these proteins have been linked to other cancer types, their potential involvement in PCa has not been previously characterized. For instance, *LCTL* encoding Lactase-like protein has been reported as a prognostic biomarker in glioma, where it modulates immune responses<sup>62</sup>. Moreover, *SVEP1* expression is suppressed by miR-1269b, which activates the PI3K/Akt pathway and promotes recurrence and metastasis in hepatocellular carcinoma<sup>63</sup>. Our integrative analysis further identified PCa risk-associated proteins whose biological functions in prostate carcinogenesis remain largely unexplored. Functional annotation through IPA highlighted *DPEPI* as a potential biomarker for PCa. *DPEPI* has been shown to promote metastasis in colorectal cancer<sup>64</sup>, but inhibit invasiveness in pancreatic ductal adenocarcinoma cells<sup>65</sup>, suggesting a context-dependent role in tumor progression. These associations warrant further study to clarify their roles in PCa and may reveal new pathways or targets for biomarkers and therapy.

Through our drug repurposing analysis, we identified 31 proteins that are targets of approved drugs used to treat various human diseases, including PCa (treated by goserelin) (Supplementary Data 6). Among the list of implicated drugs, several drugs primarily used for

non-PCa indications (alcohol dependence, inflammation, leukemia) demonstrate significant anti-PCa activity in preclinical models and in some cases early clinical settings, highlighting their potential for repurposing in PCa treatment. For example, in vitro and in vivo studies demonstrate that disulfiram (DrugBank ID DBO0822), which targets ALDH-E2, has been shown to inhibit PCa cell growth, induces apoptosis, and demethylates DNA via DNMT1 inhibition<sup>66,67</sup>. Early-phase clinical trials in both recurrent<sup>68</sup> and metastatic PCa<sup>69</sup> have shown biological activity, although clinical benefit remains to be fully established. Additionally, mitogen-activated protein kinase 3 (ERK-1) is targeted by sulindac (DrugBank ID DBO0605) and arsenic trioxide (DrugBank ID DBO1169). Sulindac is a non-steroidal anti-inflammatory drug. Sulindac derivatives have demonstrated anti-proliferative and pro-apoptotic effects in PCa cell lines<sup>70,71</sup>. Arsenic trioxide, which is FDA-approved for the treatment of acute promyelocytic leukemia, has also shown promise in PCa. It induces apoptosis and cytotoxicity in PCa cell lines<sup>72</sup> and has demonstrated significant antitumor activity in an in vivo model of androgen-independent PCa<sup>73</sup>. Phase II clinical trial reported PSA reductions or disease stabilization in some patients with hormone-refractory PCa<sup>74</sup>. These findings underscore their promise for repurposing in PCa therapy and highlight the need for further clinical investigations to fully evaluate their efficacy and therapeutic value in this context.

One advantage of the current study is the use of genetic instruments rather than directly measured protein levels. Unlike traditional epidemiological approaches using measured protein levels for association estimation, our method minimized potential biases (including residual confounding, reverse causation, and selection bias) and also increased the statistical power. This improvement is achieved by applying comprehensive genetic prediction models to large-scale GWAS summary statistics of PCa risk, which involves a very large number of cases and controls. In our work, we also applied a design that leverages information from both *cis*- and *trans*-pQTLs to fully capture the genetically regulated components of protein levels. Although *trans*-regions exhibit more complex regulatory mechanisms than *cis*-regions, the majority of identified pQTLs are *trans*-acting instead of *cis*-acting<sup>30</sup>. Therefore, we integrated both *cis*- and *trans*-genetic instruments to develop our genetic prediction models for protein levels. The numbers of identified associations in African, Hispanic/Latino, and Asian populations were relatively smaller than that observed in European population. Future studies that increase the sample sizes of underrepresented groups including African, Hispanic/Latino, and Asian groups will be critical for enhancing our understanding of PCa etiology in these understudied populations and for reducing health disparity<sup>24</sup>. Another major strength of our study is the development of population-specific protein predicted models for males in African, Asian, and Hispanic/Latino groups. This method helps mitigate potential bias and loss of statistical power caused by varying linkage disequilibrium (LD) patterns across populations<sup>29</sup>. We leveraged a large blood proteome reference dataset to establish prediction models for nearly 7,000 proteins to examine their potential relationship with PCa.

Our study has several limitations that need to be acknowledged to appropriately interpret our findings. Firstly, we were unable to conduct external validation of the established prediction models for populations other than European population<sup>30</sup>, due to a lack of data from additional well-designed studies. However, our external validation results focusing on the European population suggest that our modeling strategy is effective. As reference data of proteomic and genomic resources expand to include more globally underrepresented populations, this limitation can be well addressed in the near future. Secondly, the SomaScan platform has inherent limitations. Its aptamer-based detection method may be more susceptible to binding variability and off-target effects compared with antibody-based assays, such as Olink in the UK Biobank Pharma Proteomics Project (UKB-

PPP)<sup>75</sup>. Moreover, SomaScan interrogates only a subset of plasma proteins, and may not accurately reflect protein levels in other tissues or specific cell types. Future work using more comprehensive proteome platforms, as they become available, and extending investigations to other disease-relevant solid tissues and/or cell types will be crucial to better characterize additional disease-relevant proteins. Thirdly, due to the scope of the current study, we were unable to functionally characterize the potential roles of the identified proteins in PCa development. Future functional studies are warranted to elucidate the specific function and mechanisms of these proteins, particularly those not previously reported in prostate tumorigenesis.

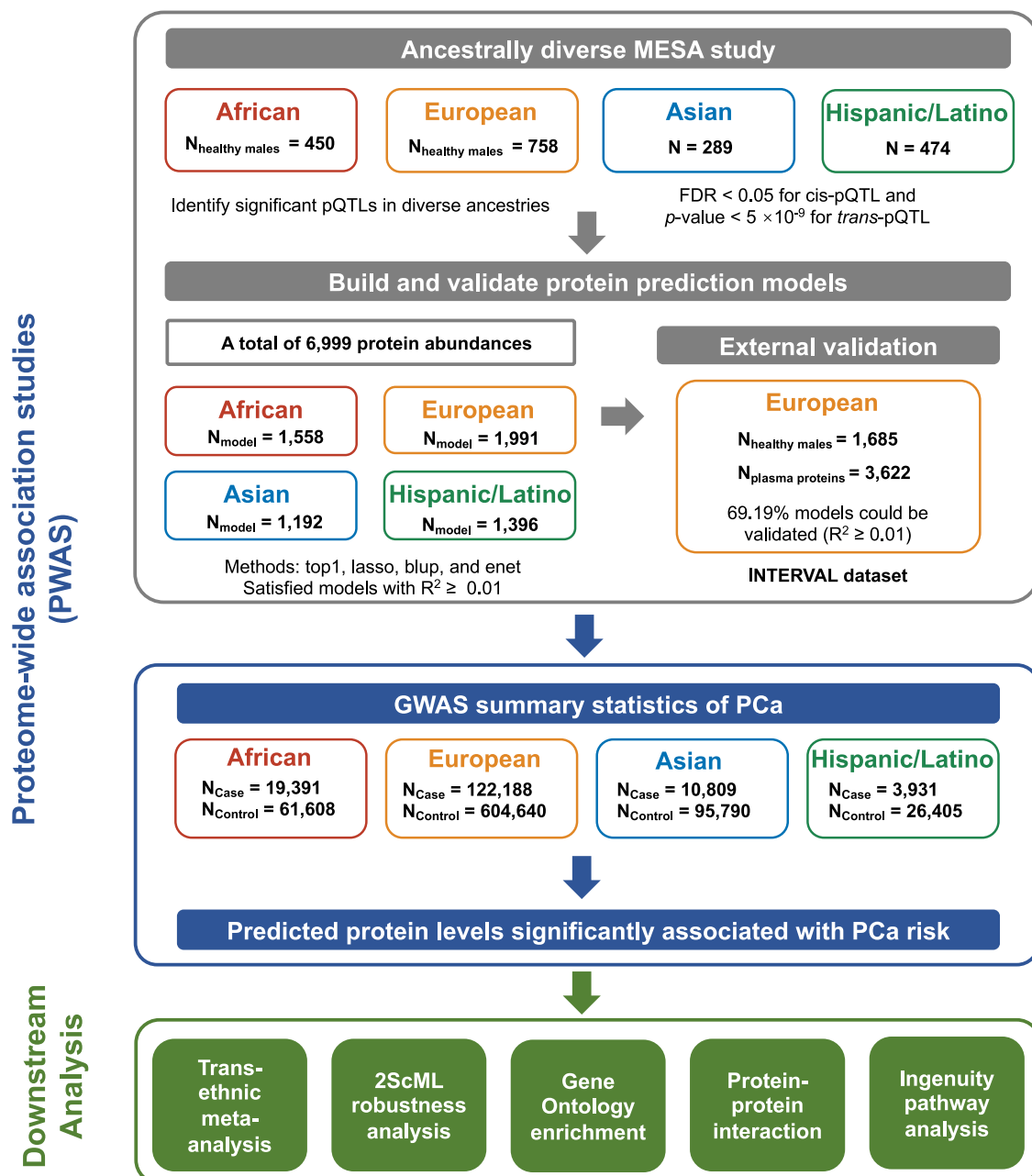
In conclusion, our comprehensive PWAS, conducted across diverse populations and encompassing a large number of proteins (~7000), identified multiple associations between genetically predicted circulating levels of proteins and PCa risk. By developing

population-specific genetic prediction models for protein abundances, we were able to elucidate distinct protein-PCa risk associations in each population of interest. These findings underscored the necessity of multi-population investigations to comprehensively elucidate the etiology of PCa and reduce health disparities.

## Methods

### Study population

The overall design of this study is illustrated in Fig. 3. The Multi-Ethnic Study of Atherosclerosis (MESA) included approximately 6500 men and women without known clinical cardiovascular disease at baseline, aged 45–84 years who were initially enrolled in 2000<sup>76</sup>. Baseline information and blood samples of the MESA participants were collected at their initial visit, which took place in six US states (New York, Maryland, Illinois, California, Minnesota, and North Carolina). Detailed



**Fig. 3 | The overall design of this study.** Key steps are illustrated with arrows connecting panels that indicate data processing, modeling, and analysis stages. Symbols and colors indicate different types of data and analytical procedures.

information on the MESA study design can be found elsewhere<sup>76</sup>. The MESA study protocol was reviewed and approved by the Institutional Review Boards (IRBs) of all participating institutions as well as by the National Heart, Lung, and Blood Institute (NHLBI). All participants provided written informed consent to participate in the parent study and received financial compensation.

In this study, we focused on 2013 male subjects who had no self-reported cancer diagnosis and no ICD-9 or ICD-10 cancer diagnoses in their hospitalization records or death certificates at baseline, including 453 African, 765 European, 296 Asian, and 499 Hispanic/Latino individuals residing in the USA. Population groups in our analysis were initially defined based on self-reported race/ethnicity as recorded in the original studies. All subjects had data available on blood protein levels (v1), genotype, and relevant covariates, including body mass index (BMI), sex, age, cigarette smoking status, and pack-years of cigarette smoking.

### Genotype data processing and quality control

The genotype data utilized in this study was generated using Affymetrix SNP 6.0, which was obtained from the MESA SNP Health Association Resource (SHARe) study (phs000420.v6.p3) and imputed on the Michigan imputation server (Minimac4.v1.0.0) using the 1000 Genomes reference panel<sup>77</sup>. For each population of interest, we excluded subjects who were identified as related via identity-by-descent (IBD) analysis, employing an in-house script that considered independent ( $R^2 < 0.2$ ) and common (minor allele frequency (MAF)  $\geq 5\%$ ) SNPs. This process resulted in a final sample of 1971 unrelated male subjects, including 450 of African background, 758 of European background, 289 of Asian background, and of 474 Hispanic/Latino background. Subsequently, SNPs were filtered based on pre-determined criteria including MAF  $> 0.05$ , genotyping missingness  $< 5\%$ , and adherence to Hardy-Weinberg equilibrium (HWE,  $P > 5 \times 10^{-6}$ ) using PLINK v1.9 software. After filtering for variants present in the 1000 Genomes reference panel, a total of 8,011,863, 5,814,519, 5,308,357, and 6,058,753 high-quality SNPs were retained for African, European, Asian, and Hispanic/Latino populations, respectively. The LD-pruned ( $R^2 < 0.2$ ), common (MAF  $\geq 5\%$ ), and genotyped variants within 200 base pair windows were used to calculate genetic principal components (PCs) using the EIGENSOFT software<sup>78</sup>.

### Proteomic data processing

Proteomic profiling was conducted using the aptamer-based SomaScan assay, which quantified 7289 human proteins. We excluded 10 SOMAmers whose target protein-encoding genes lacked positional information in the BioMart<sup>79</sup> database. An additional 280 SOMAmers targeting proteins encoded on sex chromosomes were removed to focus the analysis on plasma proteins or protein complexes encoded by autosomal genes. After these exclusions, 6999 proteins were retained for downstream analysis. The residuals of these 6999 protein abundances after adjusting for covariates were then transformed using a rank-based inverse normal transformation for model building.

### Building protein genetic prediction models

To identify informative SNPs for model building, we first conducted pQTL analyses adjusting for study site, BMI, sex, age, cigarette smoking status, pack-years of cigarette smoking, and top ten PCs. Significant *cis*-pQTLs were defined as SNPs within *cis*-regions associated with a protein at FDR  $< 0.05$ , while significant *trans*-pQTLs were defined as SNPs in *trans*-regions with  $P < 5 \times 10^{-9}$ . These significance thresholds were chosen to maximize the inclusion of potentially informative SNPs while minimizing excess noise<sup>80</sup>. We further extracted non-strand-ambiguous SNPs within 100 Kb of significant *cis*- and *trans*-pQTLs to serve as candidate predictors for each protein.

We used TWAS/FUSION framework<sup>81</sup> to construct subsequent genetic prediction models. Four methods were used for model

construction: best linear unbiased predictor (BLUP), least absolute shrinkage and selection operator (LASSO), elastic net, and top SNPs (top1). BLUP estimates the joint effect sizes of all SNPs using a single variance component<sup>82</sup>. LASSO is a penalized regression method utilizing  $L_1$  regularization techniques to produce sparse models<sup>83</sup>. As a generalization of the LASSO, elastic net linearly combines the  $L_1$ -penalty of LASSO and  $L_2$ -penalty of ridge regression, to select highly correlated variables together<sup>84</sup>. For each protein of interest, the prediction model with the most significant cross-validation  $P$ -value was retained. Only models with a cross-validation  $R^2 > 0.01$  (indicating they explain more than 1% of the variance, corresponding to a minimum -10% correlation between predicted and measured protein levels) were included in subsequent association analysis. This threshold is commonly applied in similar studies<sup>8,80,85-88</sup>. Cross-validation was performed using a five-fold scheme, where the dataset was randomly divided into five equal parts. In each fold, models were trained on 80% of the data and tested on the remaining 20%, rotating such that each subset served as a test set once. The final cross-validation performance was calculated by regressing observed protein levels against the predicted values aggregated from all test folds. This adjusted  $R^2$  accounts for model complexity and sample size, providing a conservative measure of the variance explained by the genetic predictors.

We compared  $R^2$  and  $H^2$  across African, European, Asian, and Hispanic/Latino populations using the Kruskal-Wallis test, given the non-normal distribution of  $R^2$  and  $H^2$  values. When significant, post hoc pairwise comparisons were performed using Dunn's test with Bonferroni correction. Analyses were conducted in R (version 4.1.2) using the 'FSA' package.

### Validation of protein genetic prediction models of the European population

We conducted external validation of the European population models using data from 1685 healthy European males of the INTERVAL study, which measured plasma concentrations of 3622 proteins<sup>30</sup>. Genotype data were obtained using the Affymetrix Axiom UK Biobank genotyping array, and variants were phased with SHAPEIT3 before being imputed with a reference panel comprising the 1000 Genomes Phase 3-UK10K dataset. Log-transformed protein levels were adjusted for age, sex, duration between blood draw and processing, and the top three PCs<sup>30</sup>. The rank-inverse normalized residuals from the linear regression were used to compare with the predicted protein abundances, which were generated by applying the established genetic prediction models to the INTERVAL genetic data. Models with a performance  $R^2$  value  $\geq 0.01$  (at least -10% correlation between predicted and measured protein levels) were considered to have successfully passed external validation<sup>18,80,86,89</sup>.

### Associations between genetically predicted circulating protein levels and PCa risk

We evaluated the associations between genetically predicted protein levels and PCa risk by leveraging the summary statistics of a large GWAS meta-analysis<sup>4</sup>. This meta-analysis included 156,319 PCa cases and 666,248 controls, comprising 19,391 PCa cases and 61,608 controls of African population, 122,188 cases and 604,640 controls of European population, 10,809 cases and 95,790 controls of Asian population, and 3931 cases and 26,405 controls of Hispanic/Latino population<sup>4</sup>. The TWAS/FUSION framework was used to evaluate the associations between predicted protein levels and PCa risk. The trans-population meta-analysis via an inverse variance fixed-effect approach was further implemented using METAL with fixed-effects model<sup>90</sup>. The protein-PCa risk associations were determined to be statistically significant with a threshold of FDR  $< 0.05$ . Evidence of heterogeneity of the associations across racial/ethnic groups was assessed using the  $I^2$  statistic in METAL, with high heterogeneity defined as  $I^2 > 75$  and a heterogeneity  $P$ -value (HetPVal)  $< 0.05$ <sup>91</sup>.

### Robustness test using 2ScML method

The protein prediction models in this study were built utilizing the TWAS/FUSION software, which allows for the incorporation of multiple correlated SNPs as predictors via methods such as LASSO, BLUP, or elastic net. In actual analyses, the assumption of PWAS might be violated, as it assumes that all employed SNPs serve as valid instrumental variables. To assess the robustness of our main results, we applied the 2ScML method to infer the likely causal effects of the associated proteins on PCa risk. This method relies on a plurality condition, which posits that the largest cluster of SNPs sharing the same causal effect comprises valid instrumental variables, thereby permitting more than 50% of the SNPs to be invalid<sup>92</sup>. By accounting for the horizontal pleiotropic effects of the SNPs used in the protein prediction models, 2ScML identifies valid instruments through constrained maximum likelihood estimation. This method is applicable when at least three predictors are present, enabling the selection of potentially pleiotropic SNPs. A threshold of  $FDR < 0.05$  was applied to determine significant associations using the 2ScML method.

### Tissue expression, functional enrichment, and network analysis

To evaluate tissue-specific expression, we cross-referenced 94 PCa-associated genes with RNA expression data from the HPA<sup>93</sup>. Genes were classified based on HPA criteria as: (1) Elevated in prostate, (2) Elevated in other but expressed in prostate, (3) Low tissue specificity but expressed in prostate, (4) Not detected in prostate, (5) Not detected in any tissue. Classifications were confirmed using prostate-specific RNA expression levels, enabling assessment of prostate relevance and transcriptional activity of each gene.

To investigate the biological functions of 96 unique proteins encoded by 94 genes identified in the PWAS, we performed GO enrichment analysis across three domains: Biological Process, Molecular Function, and Cellular Component. GO enrichment was conducted using ‘ClusterProfiler’ package<sup>94</sup>, with the human genome (org.Hs.eg.db) as background. Only terms with adjusted  $P$ -value  $< 0.05$  were considered significant.

Proteins operate within a complex network of intermolecular interactions rather than acting independently<sup>95</sup>. To understand the networks of PCa risk-associated proteins, we constructed a PPI network using the STRING database (<https://string-db.org/>, accessed on June 18, 2025). Additionally, we employed IPA (version 03-29-25) to perform an enrichment analysis of the genes encoding identified proteins, assessing their enrichment in canonical pathways, molecular and cellular functions, and networks. Detailed methodology for this tool has been described previously<sup>96</sup>.

### Drug repurposing analysis

To explore potential drug repurposing opportunities, we queried the DrugBank database<sup>97</sup> to investigate evidence of existing drugs targeting the identified associated proteins of interest. This analysis may reveal promising candidate drugs for further investigation of their potential efficacy in PCa treatment. We further assessed the potential relevance of these proteins to PCa using data from OpenTargets<sup>98</sup>. Specifically, proteins with a positive *overallAssociationScore* for PCa-related outcomes, including prostate-specific antigen levels, prostate carcinoma, prostate adenocarcinoma, and PCa, were retained for further consideration.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Specific genome, proteome, and covariate data of MESA<sup>76</sup> have been deposited to the database of Genotypes and Phenotypes (dbGaP) under accession code phs000209.v13.p3 ([https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000209.v13.p3](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000209.v13.p3)).

Additional data of MESA are available through a concept proposal application via MESA Genetics P and P Committee. These data are available under restricted access to protect participant privacy and comply with informed consent agreements. For data available through dbGaP, access is limited to qualified researchers who submit a Data Access Request (DAR) through dbGaP. Such requests must include a research use statement and a data use certification signed by the principal investigator and the institutional signing official. Requests will be reviewed by the relevant Data Access Committee (DAC) with a timeline developed by dbGaP. Once approved by dbGaP, access will be granted for one year. It requires an annual renewal via dbGaP to continue access beyond the initial 12-month period. Individual level data of genotype and proteomic data of INTERVAL<sup>30</sup> study are available under controlled access in European Genome-phenome Archive (EGA) under accession number EGAS00001002555. Access is restricted to protect participant privacy and is granted to qualified researchers following approval by the EGA Data Access Committee. Researchers interested in accessing these data can submit a request via the EGA data access portal, and access can be provided after the EGA Data Access Committee reviews and approves it. DAC Aim to respond to all initial requests in less than 2 weeks. The length of time you can access the data depends on the terms set by the DAC. The publicly available summary statistics of multi-population PCa GWAS<sup>4</sup> are available on the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>). These statistics are categorized by different racial/ethnicity groups with the following accession codes: European (GCST90274714, <https://www.ebi.ac.uk/gwas/studies/GCST90274714>)<sup>4</sup>, African (GCST90274715, <https://www.ebi.ac.uk/gwas/studies/GCST90274715>)<sup>4</sup>, Asian (GCST90274716, <https://www.ebi.ac.uk/gwas/studies/GCST90274716>)<sup>4</sup>, and Hispanic/Latino (GCST90274717, <https://www.ebi.ac.uk/gwas/studies/GCST90274717>)<sup>4</sup>. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

Additional data of MESA are available through a concept proposal application via MESA Genetics P and P Committee. These data are available under restricted access to protect participant privacy and comply with informed consent agreements. For data available through dbGaP, access is limited to qualified researchers who submit a Data Access Request (DAR) through dbGaP. Such requests must include a research use statement and a data use certification signed by the principal investigator and the institutional signing official. Requests will be reviewed by the relevant Data Access Committee (DAC) with a timeline developed by dbGaP. Once approved by dbGaP, access will be granted for one year. It requires an annual renewal via dbGaP to continue access beyond the initial 12-month period. Individual level data of genotype and proteomic data of INTERVAL<sup>30</sup> study are available under controlled access in European Genome-phenome Archive (EGA) under accession number EGAS00001002555. Access is restricted to protect participant privacy and is granted to qualified researchers following approval by the EGA Data Access Committee. Researchers interested in accessing these data can submit a request via the EGA data access portal, and access can be provided after the EGA Data Access Committee reviews and approves it. DAC Aim to respond to all initial requests in less than 2 weeks. The length of time you can access the data depends on the terms set by the DAC. The publicly available summary statistics of multi-population PCa GWAS<sup>4</sup> are available on the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>). These statistics are categorized by different racial/ethnicity groups with the following accession codes: European (GCST90274714, <https://www.ebi.ac.uk/gwas/studies/GCST90274714>)<sup>4</sup>, African (GCST90274715, <https://www.ebi.ac.uk/gwas/studies/GCST90274715>)<sup>4</sup>, Asian (GCST90274716, <https://www.ebi.ac.uk/gwas/studies/GCST90274716>)<sup>4</sup>, and Hispanic/Latino (GCST90274717, <https://www.ebi.ac.uk/gwas/studies/GCST90274717>)<sup>4</sup>. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

### Code availability

The code used to perform the analyses in this study is available at <https://github.com/HuaZ-bioinformatics/MESA-7K-PWAS-PCa> and Zenodo <https://zenodo.org/records/17280225><sup>99</sup>.

### References

1. Bray, F. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
2. Hemminki, K. Familial risk and familial survival in prostate cancer. *World J. Urol.* **30**, 143–148 (2012).
3. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
4. Wang, A. et al. Characterizing prostate cancer risk through multi-ancestry genome-wide discovery of 187 novel risk variants. *Nat. Genet.* **55**, 2065–2074 (2023).
5. Chimusa, E. R., Dalvie, S., Dandara, C., Wonkam, A. & Mazandu, G. K. Post genome-wide association analysis: dissecting computational pathway/network-based approaches. *Brief. Bioinform.* **20**, 690–700 (2019).
6. Freedman, M. L. et al. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.* **43**, 513–518 (2011).
7. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
8. Wu, L. et al. An integrative multi-omics analysis to identify candidate DNA methylation biomarkers related to prostate cancer risk. *Nat. Commun.* **11**, 1–11 (2020).

9. Liu, D. et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. *Int. J. Cancer* **150**, 80–90 (2022).
10. Zhong, H., Liu, S., Zhu, J. & Wu, L. Associations between genetically predicted levels of blood metabolites and pancreatic cancer risk. *Int. J. Cancer* **153**, 103–110 (2023).
11. Wu, L. et al. Identification of novel susceptibility loci and genes for prostate cancer risk: a transcriptome-wide association study in over 140,000 European Descendants. *Cancer Res.* **79**, 3192–3204 (2019).
12. Stephan, C. et al. Improved prostate cancer detection with a human kallikrein 11 and percentage free PSA-based artificial neural network. *Biol. Chem.* **387**, 801–805 (2006).
13. Nakashima, J. et al. Serum interleukin 6 as a prognostic factor in patients with prostate cancer. *Clin. Cancer Res.* **6**, 2702–2706 (2000).
14. Uetsuki, H. et al. Expression of a novel biomarker, EPCA, in adenocarcinomas and precancerous lesions in the prostate. *J. Urol.* **174**, 514–518 (2005).
15. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).
16. Balázs, K., Antal, L., Sáfrány, G. & Lumniczky, K. Blood-derived biomarkers of diagnosis, prognosis and therapy response in prostate cancer patients. *J. Personal. Med.* **11**, 296 (2021).
17. Zhang, J. et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
18. Zhong, H. et al. Identification of blood protein biomarkers associated with prostate cancer risk using genetic prediction models: analysis of over 140,000 subjects. *Hum. Mol. Genet.* **32**, 3181–3193 (2023).
19. Wu, L. et al. Analysis of Over 140,000 European Descendants identifies genetically predicted blood protein biomarkers associated with prostate cancer risk. *Cancer Res.* **79**, 4592–4598 (2019).
20. Farashi, S., Kryza, T., Clements, J. & Batra, J. Post-GWAS in prostate cancer: from genetic association to biological contribution. *Nat. Rev. Cancer* **19**, 46–59 (2019).
21. Wang, G., Zhao, D., Spring, D. J. & DePinho, R. A. Genetics and biology of prostate cancer. *Genes Dev.* **32**, 1105–1140 (2018).
22. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 17–48 (2023).
23. Guo, J. et al. ANO7: Insights into topology, function, and potential applications as a biomarker and immunotherapy target. *Tissue Cell* **72**, 101546 (2021).
24. Conti, D. V. et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat. Genet* **53**, 65–75 (2021).
25. Jiang, J. et al. ANO7 African-ancestral genomic diversity and advanced prostate cancer. *Prostate Cancer Prostatic Dis.* **27**, 558–565 (2024).
26. Rayford, W. et al. Comparative analysis of 1152 African-American and European-American men with prostate cancer identifies distinct genomic and immunological differences. *Commun. Biol.* **4**, 670 (2021).
27. Liu, D. et al. Associations between genetically predicted plasma N-glycans and prostate cancer risk: analysis of over 140,000 European descendants. *Pharmgenom. Pers. Med.* **14**, 1211 (2021).
28. Brandes, N., Linial, N. & Linial, M. Genetic association studies of alterations in protein function expose recessive effects on cancer predisposition. *Sci. Rep.* **11**, 1–16 (2021).
29. Turay, D. et al. Proteomic profiling of serum-derived exosomes from ethnically diverse prostate cancer patients. *Cancer Investig.* **34**, 1–11 (2016).
30. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
31. Lilja, H., Ulmert, D. & Vickers, A. J. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat. Rev. Cancer* **8**, 268–278 (2008).
32. Whitaker, H. C., Warren, A. Y., Eeles, R., Kote-Jarai, Z. & Neal, D. E. The potential value of microseminoprotein- $\beta$  as a prostate cancer biomarker and therapeutic target. *Prostate* **70**, 333–340 (2010).
33. Benatzy, Y., Palmer, M. A. & Brüne, B. Arachidonate 15-lipoxygenase type B: regulation, function, and its role in pathophysiology. *Front. Pharmacol.* **13**, 1042420 (2022).
34. Gregg, J. & Fraizer, G. Transcriptional regulation of EGFR1 by EGF and the ERK signaling pathway in prostate cancer cells. *Genes Cancer* **2**, 900–909 (2011).
35. Mehta, H. H. et al. IGFBP-3 is a metastasis suppression gene in prostate cancer. *Cancer Res.* **71**, 5154–5163 (2011).
36. Hockla, A. et al. PRSS3/mesotrypsin is a therapeutic target for metastatic prostate cancer. *Mol. Cancer Res.* **10**, 1555–1566 (2012).
37. Tregunna, R. Serum MMP7 levels could guide metastatic therapy for prostate cancer. *Nat. Rev. Urol.* **17**, 658–658 (2020).
38. Pierconti, F. et al. Epigenetic silencing of SOCS3 identifies a subset of prostate cancer with an aggressive behavior. *Prostate* **71**, (2011).
39. Kojima, S., Inahara, M., Suzuki, H., Ichikawa, T. & Furuya, Y. Implications of insulin-like growth factor-I for prostate cancer therapies: Review Article. *Int. J. Urol.* **16**, 161–167 (2009).
40. Wu, J. & Yu, E. Insulin-like growth factor receptor-1 (IGF-IR) as a target for prostate cancer therapy. *Cancer and Metastasis Reviews* **33**, 607–617 (2014).
41. Yang, H. et al. Matrix Metalloproteinase 11 Is a Potential Therapeutic Target in Lung Adenocarcinoma. *Mol. Ther. Oncolytics* **14**, 82–93 (2019).
42. Tagirasa, R. & Yoo, E. Role of Serine Proteases at the Tumor-Stroma Interface. *Front. Immunol.* **13**, 832418 (2022).
43. Barton, J., Blackledge, G. & Wakeling, A. *Growth factors and their receptors: new targets for prostate cancer therapy.* (2001).
44. Nguyen, D. P., Li, J. & Tewari, A. K. Inflammation and prostate cancer: The role of interleukin 6 (IL-6). *BJU Int.* **113**, 986–992 (2014).
45. Stewart, D. A., Cooper, C. R. & Sikes, R. A. Changes in extracellular matrix (ECM) and ECM-associated proteins in the metastatic progression of prostate cancer. *Reprod. Biol. Endocrinol.* **2**, 2 (2004).
46. Akaza, H. Future prospects for luteinizing hormone-releasing hormone analogues in prostate cancer treatment. *Pharmacology* **85**, 110–120 (2010).
47. Huerta-Reyes, M. et al. Treatment of breast cancer with gonadotropin-releasing hormone analogs. *Front. Oncol.* **9**, 943 (2019).
48. Prostate cancer cells tricked into self-destruction because of their hunger for copper. *Pharm. J.* <https://doi.org/10.1211/pj.2014.20066843> (2015).
49. Kaushik, A. K. et al. Inhibition of the hexosamine biosynthetic pathway promotes castration-resistant prostate cancer. *Nat. Commun.* **7**, 11612 (2016).
50. Shah, K. et al. Androgen receptor signaling regulates the transcriptome of prostate cancer cells by modulating global alternative splicing. *Oncogene* **39**, 6172–6189 (2020).
51. Penney, K. L. et al. Association of prostate cancer risk variants with gene expression in normal and tumor tissue. *Cancer Epidemiol. Biomark. Prevent.* **24**, 255–260 (2015).
52. Desai, T. A. et al. Identifying proteomic risk factors for overall, aggressive, and early onset prostate cancer using Mendelian Randomisation and tumour spatial transcriptomics. *EBioMedicine* **105**, 105168 (2024).

53. Ren, F., Jin, Q., Liu, T., Ren, X. & Zhan, Y. Proteome-wide mendelian randomization study implicates therapeutic targets in common cancers. *J. Transl. Med.* **21**, 646 (2023).
54. Wu, J. et al. Proteome-wide Mendelian randomization identifies causal plasma proteins in prostate cancer development. *Hum. Genom.* **19**, 17 (2025).
55. Thomas, G. et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
56. Eeles, R. A. et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
57. Lou, H. et al. Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. *Proc. Natl. Acad. Sci. USA* **106**, 7933–7938 (2009).
58. Camp, N. J. et al. Replication of the 10q11 and Xp11 prostate cancer risk variants: Results from a Utah pedigree-based study. *Cancer Epidemiol. Biomark. Prevent.* **18**, 1290–1294 (2009).
59. Chang, B. L. et al. Fine mapping association study and functional analysis implicate a SNP in MSMB at 10q11 as a causal variant for prostate cancer risk. *Hum. Mol. Genet.* **18**, (2009).
60. Mhatre, D. R. et al. The rs10993994 in the proximal MSMB promoter region is a functional polymorphism in Asian Indian subjects. *Springerplus* **4**, 380 (2015).
61. Xu, B. et al. A functional polymorphism in MSMB gene promoter is associated with prostate cancer risk and serum MSMB expression. *Prostate* **70**, 1146–1152 (2010).
62. Su, J. et al. LCTL is a prognostic biomarker and correlates with stromal and immune infiltration in gliomas. *Front. Oncol.* **9**, 1083 (2019).
63. Chen, L. et al. The novel miR-1269b-regulated protein SVEP1 induces hepatocellular carcinoma proliferation and metastasis likely through the PI3K/Akt pathway. *Cell Death Dis.* **11**, 320 (2020).
64. Zeng, C. et al. DPEP1 promotes drug resistance in colon cancer cells by forming a positive feedback loop with ASCL2. *Cancer Med.* **12**, 412–424 (2023).
65. Zhang, G. et al. DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS ONE* **7**, e31507 (2012).
66. Ijlin, K. et al. High-throughput cell-based screening of 4910 known drugs and drug-like small molecules identifies disulfiram as an inhibitor of prostate cancer cell growth. *Clin. Cancer Res.* **15**, 6070–6078 (2009).
67. Lin, J. et al. Disulfiram is a DNA demethylating agent and inhibits prostate cancer cell growth. *Prostate* **71**, 333–343 (2011).
68. Schweizer, M. T. et al. Pharmacodynamic study of disulfiram in men with non-metastatic recurrent prostate cancer. *Prostate Cancer Prostatic Dis.* **16**, 357–361 (2013).
69. Zhang, T. et al. Prospective clinical trial of disulfiram plus copper in men with metastatic castration-resistant prostate cancer. *Prostate* **82**, (2022).
70. Lim, J. T. E. et al. Sulindac derivatives inhibit growth and induce apoptosis in human prostate cancer cell lines. *Biochem. Pharmacol.* **58**, 1097–1107 (1999).
71. Nortcliffe, A. et al. Synthesis and biological evaluation of nitric oxide-donating analogues of sulindac for prostate cancer treatment. *Bioorg. Med. Chem.* **22**, 756–761 (2014).
72. Uslu, R. et al. Arsenic trioxide-mediated cytotoxicity and apoptosis in prostate and ovarian carcinoma cell lines. *Clin. Cancer Res.* **6**, 4957–4964 (2000).
73. Maeda, H. et al. Tumor growth inhibition by arsenic trioxide (As<sub>2</sub>O<sub>3</sub>) in the orthotopic metastasis model of androgen-independent prostate cancer. *Cancer Res.* **61**, 5432–5440 (2001).
74. Berry, W., Dakhil, S., Gecurich, M. A. & Asmar, L. Phase II trial of single-agent weekly docetaxel in hormone-refractory, symptomatic, metastatic carcinoma of the prostate. *Semin Oncol.* **28**, 8–15 (2001).
75. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
76. Bild, D. E. et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am. J. Epidemiol.* **156**, 871–881 (2002).
77. Mogil, L. S. et al. Genetic architecture of gene expression traits across diverse populations. *PLoS Genet.* **14**, e1007586 (2018).
78. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, 2074–2093 (2006).
79. Durinck, S. et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
80. Liu, S. et al. Identification of proteins associated with type 2 diabetes risk in diverse racial and ethnic populations. *Diabetologia* <https://doi.org/10.1007/S00125-024-06277-3> (2024).
81. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
82. Robinson, G. K. That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* **6**, 15–32 (1991).
83. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
84. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
85. Liu, S., Zhong, H., Zhu, J. & Wu, L. Identification of blood metabolites associated with risk of Alzheimer’s disease by integrating genomics and metabolomics data. *Mol. Psychiatry* **29**, 1153–1162 (2024).
86. Zhu, J. et al. Associations between genetically predicted plasma protein levels and Alzheimer’s disease risk: a study using genetic prediction models. *Alzheimers Res. Ther.* **16**, 8 (2024).
87. He, J. et al. Enhancing disease risk gene discovery by integrating transcription factor-linked trans-variants into transcriptome-wide association analyses. *Nucleic Acids Res.* **53**, gkae1035 (2025).
88. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
89. Zhu, J. et al. Proteome-wide association study and functional validation identify novel protein markers for pancreatic ductal adenocarcinoma. *Gigascience* **13**, giae012 (2024).
90. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
91. Graff, M. et al. Discovery and fine-mapping of height loci via high-density imputation of GWASs in individuals of African ancestry. *Am. J. Hum. Genet.* **108**, 564–582 (2021).
92. Xue, H., Shen, X. & Pan, W. Causal inference in transcriptome-wide association studies with invalid instruments and GWAS summary data. *J. Am. Stat. Assoc.* **118**, 1525–1537 (2023).
93. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
94. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
95. Lehne, B. & Schlitt, T. Protein-protein interaction databases: keeping up with growing interactomes. *Hum. Genom.* **3**, 1–7 (2009).
96. Krämer, A., Green, J., Pollard, J. Jr & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).
97. Wishart, D. S. et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).

98. Koscielny, G. et al. Open Targets: a platform for therapeutic target identification and Validation. *Nucleic Acids Res.* **45**, D985–D994 (2017).
99. Hua, Z. Proteome-wide association study of prostate cancer risk across populations. *Zenodo* <https://doi.org/10.5281/ZENODO.17280225> (2025).

## Acknowledgements

The authors also would like to thank all of the individuals for their participation in the parent studies and all the researchers, clinicians, technicians and administrative staff for their contribution to the studies. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Lang Wu is supported by National Human Genome Research Institute/National Institute on Minority Health and Health Disparities (NHGRI/NIMHD) U54 HG013243 and National Cancer Institute R01CA263494 and U01CA293883. Hua Zhong and Shuai Liu are partially supported under award number U24DK132746-01, UCLA LIFT-UP (Leveraging Institutional support for Talented, Upcoming Physicians and/or Scientists). Peter Ganz, Rajat Deo, and Ruth F Dubin are supported by R01HL159081. The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, DK063491, and R01HL105756. The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutes can be found at <http://www.mesa-nhlbi.org>.

## Author contributions

L.W. conceived the study and supervised the work. J.Z. contributed to the study design and conducted drug repurposing analysis. S.L., H.Z., and J.Z. performed statistical analyses. H.Z. performed PPI and IPA analyses. H.Z., J.Z., S.L., and L.W. wrote the initial draft manuscript. C.W., L.W., S.P.W., C.H.M., M.J.B., P.D., X.G., C.W.J., H.J.L., K.D.T., R.P.T., R.Y., A.W.M., S.S.R., J.I.R., R.D., R.F.D., and P.G. conducted the experiment, contributed materials, and provided valuable feedback for revising the

manuscript. All authors have reviewed and approved the final manuscript.

## Competing interests

Lang Wu provided consulting services to Pupil Bio Inc., Techspert, and Galiher DeRobertis & Waxman LLP, and reviewed manuscripts for *Gastroenterology Report*, not related to this study, and received honorarium. No potential conflicts of interest were disclosed by the other authors.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-66250-5>.

**Correspondence** and requests for materials should be addressed to Lang Wu.

**Peer review information** *Nature Communications* thanks Zhongming Zhao, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

<sup>1</sup>Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawai'i Cancer Center, University of Hawai'i at Mānoa, Honolulu, HI, USA. <sup>2</sup>Department of Interdisciplinary Oncology and Department of Genetics, LSU-LCMC Health Cancer Center, School of Medicine, Louisiana State University Health Sciences Center, New Orleans, LA, USA. <sup>3</sup>Department of Quantitative Health Sciences, John A. Burns School of Medicine, University of Hawai'i at Mānoa, Honolulu, HI, USA. <sup>4</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>5</sup>Department of Tumor Microenvironment and Metastasis, Moffitt Cancer Center, Tampa, FL, USA. <sup>6</sup>Johns Hopkins Ciccarone Center for the Prevention of Cardiovascular Disease, Johns Hopkins School of Medicine, Baltimore, MD, USA. <sup>7</sup>Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, USA. <sup>8</sup>Department of Medicine, Johns Hopkins Hospital, Baltimore, MD, USA. <sup>9</sup>Laboratory for Clinical Biochemistry Research, University of Vermont, Burlington, VT, USA. <sup>10</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>11</sup>Collaborative Health Studies Coordinating Center, University of Washington, Seattle, WA, USA. <sup>12</sup>Division of Cardiovascular Sciences, Epidemiology Branch, National Heart, Lung and Blood Institute, Bethesda, MD, USA. <sup>13</sup>Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA. <sup>14</sup>Division of Cardiovascular Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. <sup>15</sup>Department of Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>16</sup>Division of Cardiology, Zuckerberg San Francisco General Hospital and Department of Medicine, University of California San Francisco, San Francisco, CA, USA. <sup>17</sup>These authors contributed equally: Hua Zhong, Jingjing Zhu. ✉e-mail: [lwu3@lsuhsc.edu](mailto:lwu3@lsuhsc.edu)