Identifying proteomic risk factors for overall, aggressive, and early onset prostate cancer using Mendelian Randomisation and tumour spatial transcriptomics


Published in:
EBioMedicine

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2024 The Authors.
This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback
Identifying proteomic risk factors for overall, aggressive, and early onset prostate cancer using Mendelian Randomisation and tumour spatial transcriptomics


Summary

Background Understanding the role of circulating proteins in prostate cancer risk can reveal key biological pathways and identify novel targets for cancer prevention.

Methods We investigated the association of 2002 genetically predicted circulating protein levels with risk of prostate cancer overall, and of aggressive and early onset disease, using cis-pQTL Mendelian randomisation (MR) and colocalisation. Findings for proteins with support from both MR, after correction for multiple-testing, and colocalisation were replicated using two independent cancer GWAS, one of European and one of African ancestry. Proteins with evidence of prostate-specific tissue expression were additionally investigated using spatial transcriptomic data in prostate tumour tissue to assess their role in tumour aggressiveness. Finally, we mapped risk proteins to drug and ongoing clinical trials targets.

Findings We identified 20 proteins genetically linked to prostate cancer risk (14 for overall [8 specific], 7 for aggressive [3 specific], and 8 for early onset disease [2 specific]), of which the majority replicated where data were available. Among these were proteins associated with aggressive disease, such as PPA2 [Odds Ratio (OR) per 1 SD increment = 2.13, 95% CI: 1.54–2.93], PYY [OR = 1.87, 95% CI: 1.43–2.44] and PRSS3 [OR = 0.80, 95% CI: 0.73–0.89], and those associated with early onset disease, including HEBP1 [OR = 2.89, 95% CI: 1.99–4.21], POGLUT3 [OR = 0.76, 95% CI: 0.67–0.86] and TPM3 [OR = 0.47, 95% CI: 0.34–0.64]. We confirmed an inverse association of MSMB with prostate cancer overall [OR = 0.81, 95% CI: 0.80–0.82], and also found an inverse association with both aggressive [OR = 0.84, 95% CI: 0.82–0.86] and early onset disease [OR = 0.71, 95% CI: 0.68–0.74]. Using spatial transcriptomics data, we identified MSMB as the genome-wide top-most predictive gene to distinguish benign regions from high grade cancer regions that comparatively had five-fold lower MSMB expression. Additionally, ten proteins that were associated with prostate cancer risk also mapped to existing therapeutic interventions.
Interpretation  Our findings emphasise the importance of proteomics for improving our understanding of prostate cancer aetiology and of opportunities for novel therapeutic interventions. Additionally, we demonstrate the added benefit of in-depth functional analyses to triangulate the role of risk proteins in the clinical aggressiveness of prostate tumours. Using these integrated methods, we identify a subset of risk proteins associated with aggressive and early onset disease as priorities for investigation for the future prevention and treatment of prostate cancer.

Funding  This work was supported by Cancer Research UK (grant no. C8221/A29017).

Copyright  © 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Proteins; Proteomics; -Omics; Mendelian randomisation; Cancer; Genetic epidemiology; Spatial transcriptomics

Introduction  Prostate cancer is a heterogeneous disease with a high survival rate for those diagnosed with indolent or low-stage disease, but a less than 50% 5-year survival rate for those diagnosed with aggressive or metastatic cancer.¹ The proportion of these clinically aggressive cases is higher among men younger than 55 years (early onset disease), which contributes to premature death among these men.² However, few risk factors for prostate cancer have been established. These include: advanced age, African ancestry, family history of the disease, circulating levels of insulin-like growth factor I and microsemminoprotein-beta (MSMB), with little evidence for successful strategies for prevention.²³

Recent advances in multiplexed and high throughput platforms as well as the widespread availability of genotypic arrays have identified genetic variants that determine circulating levels of thousands of proteins, known as protein-quantitative trait loci (pQTL). PQTL, in particular those lying in or near a protein’s cognate gene (referred to as cis-pQTL), can be leveraged to identify candidate aetiologic proteins for cancer risk through Mendelian randomisation (MR) analyses, an approach that can limit the impact of reverse causality when relevant assumptions are met.³⁴ In the context of cis-pQTL, these assumptions require that the cis-pQTL is robustly associated with levels of its encoding protein (relevance assumption), that the...
cis-pQTL is associated with the outcome of interest only through the association with its encoding protein (exclusion restriction assumption), and that the cis-pQTL shares no common causes with the outcome (independence assumption).11 MR can also be complemented with colocalisation analyses to further exclude confounding by linkage disequilibrium (LD).11 Candidate aetiologic proteins for cancer risk identified using these methods can provide a valuable starting point for further analyses using more resource-intensive methods, such as spatial transcriptomics, where their functional importance at the tissue level can be directly interrogated to triangulate their role in aetiology.12,13

Using an integrated cis-pQTL MR and colocalisation pipeline, we analysed the associations of 2002 unique proteins with overall, aggressive, and early onset prostate cancer and replicated and mapped those with significant findings to drug targets. Additionally, we investigated the spatial distribution and gene expression profiles of a subset of these proteins in prostate tumour tissue using spatial transcriptomics. In doing so, we demonstrated the value of protein MR and colocalisation analyses to identify proteins that may have a causal role in the tumour aggressiveness.

Methods

Overall study design

We extracted cis genetic instruments for circulating protein levels from publicly available datasets, and harmonised these cis-pQTL with the GWAS results from an international prostate cancer consortium, including aggressive and early onset subtypes (Supplementary Figure S1). We subsequently estimated risk associations for protein levels using cis-pQTL MR against each of these three prostate cancer endpoints. All associations passing a multiple testing threshold in MR analyses were then followed up with colocalisation analyses. Where data were available, we performed replication analyses in an external prostate cancer GWAS in European and African ancestry (using African ancestry specific cis-pQTLs—see below) populations for proteins with evidence from MR and colocalisation analyses. For proteins identified as risk factors for prostate cancer with evidence of specific expression in the prostate tissue, we additionally performed analyses using spatial transcriptomics to gain insights into the spatial distribution and gene expression patterns of these proteins in prostate tumour samples. Finally, we conducted an exploratory analysis restricting to cis-pQTL whose cognate genes are established drug targets.

Identification of cis-pQTL

Genetic instruments for cis-pQTL were extracted from 4 publicly available protein GWAS at p < 5 × 10^{-8} and clumped at R^2 = 0.01 within their originating panel (instruments and study source characteristics presented in Supplementary Tables S1 and S2).14,15 Cis-instruments were defined in the first instance as those that were genome-wide significant (p < 5 × 10^{-8}) within 1 Mb of the transcription start side of the measured protein encoded gene, or as the sentinel cis-pQTL for the measured protein, depending on data availability. We additionally gathered data on cis-pQTL from published GWAS present on the OpenGWAS platform using a relaxed p-value threshold of 5 × 10^{-8} due to the high biological plausibility of identifying cis-pQTL at or near a protein’s cognate gene.16,17 Specifically, we extracted unreported cis-pQTL from the genomic region 1 megabase up and downstream of the cognate gene for a given protein GWAS (Supplementary Table S2). We subsequently extracted all instruments where no cis-pQTL was present at p < 5 × 10^{-8} but at least one cis-pQTL was present at p < 5 × 10^{-6}.

All instruments were mapped to Uniprot IDs, and cis-pQTL with weak instrument strength at F_stat < 10 (β^2/σ^2) were excluded from the study in order to mitigate weak instrument bias. For cis-pQTL that were not present in the cancer outcome data, SNP proxies were selected at r^2_max where r^2 > 0.8 in 1000 genomes CEU population with the index cis-pQTL. To ensure independence of instruments, SNPs were clamped at r^2 = 0.01 within each study source using the PLINK clumping method. In total, 2002 unique plasma proteins that fit these criteria were included in analyses.

Cancer outcome data

Genetic associations for overall, aggressive, and early onset prostate cancer were obtained from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium (Supplementary material).20 Full study characteristics have been described previously, but briefly, summary statistics for SNP associations with prostate cancer and subtypes were generated from the PRACTICAL consortium using 85,554 overall prostate cancer cases and 91,972 controls (database of Genotypes and Phenotypes [dbGaP] project #31553), 15,167 aggressive prostate cancer cases and 58,308 controls, and 6988 cases of early onset disease and 44,256 controls, all of European ancestry.16,20 Aggressive prostate cancer is defined in PRACTICAL as cases having metastatic disease or Gleason score ≥ 8 or PSA > 100 ng/mL or prostate cancer death. Early onset prostate cancer cases are defined as those diagnosed before the age of 55 years. Genotype information was imputed for samples using the 2014 release of the 1000 Genomes Project as a reference panel. Based on the median proportion of variance explained for cis-pQTL used in this study (0.01), we had 80% power to detect an odds ratio of 1.14 for overall prostate cancer, 1.29 for aggressive disease and 1.36 for early onset disease.
Statistics

Two-sample Mendelian randomisation

Cis-pQTL data were harmonised using the TwoSampleMR package and matched to each cancer outcome by rsID and by matching on effect and other allele, and oriented to the protein-increasing allele. All GWAS were on the hg19 build and there was no participant overlap between the protein GWAS and the cancer GWAS. Two-sample Mendelian randomisation (MR) was subsequently performed for each cis-pQTL on risk of overall, aggressive, and early onset prostate cancer using the Wald-ratio method (βcancer/βprotein). Resulting associations where the PWald passed a Bonferroni-corrected threshold of significance based on the total number of unique proteins assessed for each of the three prostate cancer outcomes were taken forward in analyses (PWald < 0.05/NProteins analysed per cancer outcome). Multiple independent [r² < 0.01] cis-pQTL that proxied the same protein and both associated after correction for multiple testing and colocalised with the same prostate cancer outcome, were combined using the inverse-variance weighted method (IVW). Odds ratio estimates are scaled per standard deviation increment in relative and normalised circulating protein concentrations.

Colocalisation analyses were performed to assess the probability that the protein and cancer instruments share a causal variant, fulfilling the important exchangeability assumption of MR. Specifically, single and conditional iterative colocalisation analyses were performed for all cis-pQTL MR results that passed a Bonferroni correction for multiple testing based on the number of unique proteins in the study (p < 0.05/NProteins), using all variants within a 75 kb region up- and downstream from the index cis-pQTL to assess confounding by linkage disequilibrium. To mitigate the chance of false-positive findings, we selected priors of P1: 1 × 10⁻³, P2: 1 × 10⁻⁴, and P12: 1 × 10⁻⁵, which roughly equate to a 0.1% prior belief in colocalisation (PP4). We defined a threshold PP4 in support of a shared association for a protein and cancer signal at 0.70 to take proteins forward for subsequent analysis and the highest PP4 of any method of colocalisation was recorded to assess confidence in the shared association for each SNP assessed.

Replication of robust proteins in European and African ancestry populations

We conducted a replication analysis of cis-pQTL MR associations passing both multiple testing correction and colocalisation threshold (referred to as robust proteins) using an external GWAS in a European ancestry population of overall prostate cancer risk. GWAS summary statistics represented a meta-analysis in FinnGen r9 and the UK Biobank (20,907 cases & 289,710 controls). Additionally, where possible, we also performed replication analyses using cis-pQTL identified in an African ancestry protein GWAS in the Atherosclerosis Risk in Communities study (4657 proteins in 1871 African-ancestry participants) and a GWAS of overall prostate cancer among African-ancestry populations obtained from dbGaP (project #31553) containing data from the AAPC GWAS, Ghana Prostate Study, ProHealth Kaiser GWAS, and ELLIPSE OncoArray (10,368 cases and 10,986 controls). GWAS for aggressive and early onset prostate cancer were unavailable to use as a replicate sample in either ancestry population. We considered a directionally concordant risk estimates and Wald ratio p < 0.05 using external data to positively indicate replication. No sample overlap was present between samples used to generate protein associations and those used to conduct replication analyses.

Drug target pQTL analyses

We restricted our MR results to those cis-pQTL that share a cognate gene that is an established drug target by reference to the DrugBank, Therapeutic Target Database, Pharos consortium, ClinicalTrials.gov or expert curation. As above, we defined robust associations as Wald p < 0.05/NProteins, where NProteins is the number of unique proteins analysed for a given cancer outcome that were identified as the cognate gene of a pharmaceutical target and PP4 > 0.7. Additionally, all proteins identified in overall and drug target analyses were queried in the Cortellis database (https://www.cortellis.com) to assess the highest current level of clinical development stage.

Statistical analyses were performed in R version 4.1 and all tests of significance were two-sided, where p-values < 0.05 were considered statistically significant. MR analyses were performed using the TwoSampleMR R package and colocalisation analyses were performed using the coloc R package.

Gene expression analysis using spatial transcriptomics

Spatial transcriptomics provides a spatial map of gene expression within the target tissue. This spatial information can be used to investigate the variation in gene expression by healthy tissue and tumour type intra-tumourally, and as a result, it can provide valuable insights into tumorigenesis and inform causal inference in this molecularly heterogeneous disease. Spatial transcriptomic analysis was performed for those proteins that passed both the multiple testing correction and colocalisation threshold, and that also showed high expression in the prostate epithelium. Data for spatial transcriptomics were obtained from our previously published dataset derived from radical prostatectomy tissue taken from a patient with multifocal prostate cancer. Our analysis focused on eight distinct tissue sections, which collectively comprised of 32,156 spots,
some of which contained regions of cancer as well as histo-pathologically benign prostate tissue, and some of which did not contain cancer. To ensure data quality, samples with less than 500 Unique Molecular Identifier (UMI) counts were excluded from the analysis. The initial fastq files were processed using the 10x Visium Spaceranger software. Spaceranger outputs were then aggregated and normalized (using aggr) to correct for batch effects, enabling the conversion of the files into comparable gene expression data and ensuring variance reduction between samples and within cells. A consensus pathology approach was employed involving two pathologists who independently annotated each spatial transcriptomics spot, with the aim to include those that predominantly contained epithelial cells, which comprised approximately 1–15 cells. Violin plots were generated using Graphpad Prism (version 10).

Iterative random forest network using spatial transcriptomics

We used the iterative random forest (iRF) method to identify the most robust gene expression network discriminating between benign and highest Gleason grade group histology, after normalising data using SCTransform.36 The criteria we used to select a credible random forest model was stability >0.8 and precision >0.8 and the resulting gene network was visualised using Gephi 0.99.

Ethics statement

Genetic analyses were performed using summary statistics and did not require individual level data. All studies contributing to these analyses obtained ethics approval from institutional review boards in their country and all participants included in underlying data provided informed consent.

Role of funders

The funding institutions had no role in the design and conduct of the study including data collection, analysis, and interpretation of results, or the preparation, review and decision to submit the manuscript for publication.

Results

We investigated the associations of 2002 unique proteins using 4592 cis-pQTL that harmonised with the GWAS summary statistics for at least one of overall (1999 proteins; 4582 cis-pQTL), aggressive (1986 proteins; 4543 cis-pQTL), or early onset prostate cancer (1984 proteins; 4534 cis-pQTL) (Fig. 1). From these analyses we identified 20 proteins that were associated, after correction for multiple testing, with at least one outcome of overall (14 proteins), aggressive (7 proteins), or early-onset (8 proteins) prostate cancer and that also had support from colocalisation analyses (Fig. 2, Table 1). Of the 20 proteins associated with any prostate cancer outcome, several showed robust associations in only one outcome, including seven that appeared specific to overall prostate cancer (SNTC, CREBL1, INFA14, ISLR2, MMP7, SERPINA1, TNSFRS10B), three that appeared specific to aggressive disease (C4A, C2, TNFRSF6B), and two that appeared specific to early onset disease (SERPINA3, PYY).

The most statistically significant associations per standard deviation increase in protein level, with evidence of colocalisation were seen for MSMB (a protein that is specifically expressed in the prostate) with a lower risk of all prostate cancer endpoints [OROverall = 0.81, 95% CI: 0.79–0.82, PP4: 100%; ORAggressive = 0.84, 95% CI: 0.82–0.86, PP4: 0.99; OREarly Onset = 0.71, 95% CI: 0.68–0.74, PP4: 1.0, Table 1, Fig. 2]. TPM3 was the only other protein that had a significant, colocalised association with risk of all outcomes [OROverall = 0.64, 95% CI: 0.57–0.73, PP4: 0.94; ORAggressive = 0.61, 95% CI: 0.49–0.76, PP4: 0.96; OREarly Onset = 0.47, 95% CI: 0.34–0.64, PP4: 0.95, Fig. 2].

We also reported proteins with a colocalised association for one outcome and little evidence for an association with others after correction for multiple testing, such as IFNA14, ISLR2, MMP7, and TNSFRS10B which were associated exclusively with overall prostate cancer [ORIFNA14 = 0.74, 95% CI: 0.70–0.78; ORISLR2 = 0.82, 95% CI: 0.75–0.90; ORMMP7 = 0.67, 95% CI: 0.58–0.76; ORTNSFRS10B = 0.87, 95% CI: 0.83–0.92, Fig. 2]. Similarly, PYY and SERPINA3 associated with an increased risk of early onset prostate cancer only [ORPYY = 1.87, 95% CI: 1.43–2.44; ORSERPINA3 = 2.08, 95% CI: 1.51–2.88, Fig. 2] while C2 associated with aggressive prostate cancer only [ORC2 = 0.67, 95% CI: 0.56–0.81, Fig. 2].

We additionally identified proteins with evidence for a directionally concordant and colocalised association with some, but not all, prostate cancer outcomes, including TNFRSF6B, which had an inverse association with all outcomes but only showed evidence in favour of colocalisation for aggressive disease [OROverall = 0.53, 95% CI: 0.46–0.61, PP4: 0.00; ORAggressive = 0.48, 95% CI: 0.37–0.63, PP4: 0.90; OREarly Onset = 0.43, 95% CI: 0.29–0.63, PP4: 0.14, Fig. 2, Supplementary Table S3]. Likewise, PPA2 was associated with an increased risk of both aggressive and early onset disease, but lacked support from colocalisation analyses for prostate cancer risk overall [OROverall = 1.84, 95% CI: 1.52–2.22, PP4: 0.01; ORAggressive = 2.13, 95% CI: 1.54–2.93, PP4: 0.99; OREarly Onset = 2.70, 95% CI: 1.71–4.27, PP4: 0.93, Fig. 2, Supplementary Table S3].

Replication of robust proteins in European and African ancestry populations

We replicated the association for all 14 proteins that were robustly associated with overall prostate cancer (SNTC, CREBL1, CREB3, EHHB1, INFA14, ISLR2, MMP7, MSMB, PRSS3, PLG, POGLUT3, SERPINA1, and
TNSFRF10B, TPM3) using an independent meta-analysis of European ancestry participants in the UK Biobank and FinnGen cohorts (Table 1). Among these, the most statistically significant association was for MSMB [OROverall = 0.83, 95% CI: 0.81–0.85, Table 1] and the largest effect size was for PLG [OROverall = 0.46, 95% CI: 0.07–0.84, Table 1]. We additionally identified African ancestry-specific cis-pQTL for six of the 14 proteins (ISLR2, MMP7, MSMB, POGLUT3, PRSS3, SERPINA1; Table 1). Of these, three protein associations with risk of prostate cancer overall were replicated in men of African ancestry: MSMB [OROverall = 0.85, 95% CI: 0.80–0.91], MMP7 [OROverall = 0.83, 95% CI: 0.73–0.95], and ISLR2 [OROverall = 0.85, 95% CI: 0.73–0.99] (Fig. 3, Table 1).

Drug target analysis
Out of the 2002 unique proteins investigated, we identified 525 proteins that could be successfully mapped at the gene level to the target of a therapeutic intervention. Of these, ten (TPM3, PRSS3, PLG, MMP7, SERPINA1, SERPINA3, TNFRSF10B, C4A, HDGF, and LAYN) were associated with risk of at least one prostate cancer outcome after correction for multiple testing (525 proteins), and also showed evidence of colocalisation (Supplementary Table S4). For example, PLG and C4A mapped to the clot dissolving class of fibrinolytics, TPM3 mapped to phenethyl isothiocyanate, and MMP7 mapped to matrix metalloproteinase inhibitor, marimastat.

Spatial transcriptomic analysis
Of the 20 proteins that were associated with at least one prostate cancer outcome, we performed a targeted follow-up analysis for the two proteins with known high expression in the prostate tumour epithelium, MSMB and CREB3L4, using organ-wide spatial transcriptomic data on tissue obtained by radical prostatectomy from a patient with multifocal prostate cancer.35 In analysing epithelial-rich spots, we observed marked differences of MSMB expression between benign cells, where MSMB was highly abundant, and Gleason grade group 4 (GG4) cells, where MSMB was very low or absent in a majority of cells (RNA Count MSMB benign [median, interquartile range]: 532 [97–1310] vs. MSMBGG4: 1.0 [0.0–6.0], Fig. 4c). A similarly, albeit more modestly, lower MSMB expression was observed in GG2 and GG4 cells (MSMBGG1: 40.0 [8.0–75.0] and MSMBGG2: 9.0 [4.0–21.0] compared to benign cells (Fig. 4c). A lower expression of CREB3L4 was noted in GG2 and GG4 cells compared to benign cells CREB3L4benign: 3.0 [1.0–6.0] vs. CREB3L4GG2: 1.0 [0.0–2.0] vs. CREB3L4GG4:
Fig. 2: Odds ratios (95% confidence intervals) for genetically predicted protein levels and prostate cancer risk (for proteins with $p < $ Bonferroni threshold based on 0.05/number of proteins analysed). Odds ratio estimates are scaled per standard deviation increment in genetically predicted relative circulating protein concentrations. Filled circles represent Bonferroni-significant associations and asterisks indicate evidence for colocalisation (PP4 > 0.70).
<table>
<thead>
<tr>
<th>Prostate cancer outcome</th>
<th>Gene</th>
<th>SNP</th>
<th>Proportion variance explained (%)</th>
<th>Uniprot ID</th>
<th>Platform</th>
<th>Odds ratio (95% CI) practical</th>
<th>p-value (unadjusted)</th>
<th>Maximum PP4</th>
<th>Odds ratio (95% CI) UKBB/Finn gen</th>
<th>Odds ratio (95% CI) African ancestry</th>
<th>Drug target</th>
<th>Current drug trials</th>
<th>Current drug trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>MSMB</td>
<td>rs10993994</td>
<td>49.6</td>
<td>P08118</td>
<td>SomaScan</td>
<td>0.81 (0.8-0.82)</td>
<td>1.32E-165</td>
<td>1.00</td>
<td>0.83 (0.8-0.85)</td>
<td>0.85 (0.8-0.91)</td>
<td>Phase I (Prostate Cancer therapy; PCK-3145)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EHBPI</td>
<td>rs73934251</td>
<td>0.53%</td>
<td>Q8NDI1</td>
<td>SomaScan</td>
<td>1.89 (1.63-2.19)</td>
<td>8.09E-18</td>
<td>0.98</td>
<td>2.09 (1.66-2.65)</td>
<td>0.64 (0.53-0.77)</td>
<td>Phenethyl Isothiocyanate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td>rs72696208</td>
<td>0.68%</td>
<td>P06753</td>
<td>SomaScan</td>
<td>0.64 (0.57-0.72)</td>
<td>2.72E-13</td>
<td>0.94</td>
<td>0.95 (0.91-0.99)</td>
<td>0.95 (0.87-1.1)</td>
<td>Biological testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRSS3</td>
<td>rs2005617</td>
<td>14.5%</td>
<td>P35030</td>
<td>SomaScan</td>
<td>0.91 (0.89-0.94)</td>
<td>6.47E-11</td>
<td>1.00</td>
<td>0.95 (0.91-0.99)</td>
<td>0.95 (0.8-1.1)</td>
<td>Biological testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLG</td>
<td>rs982403</td>
<td>0.23%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.48 (0.38-0.60)</td>
<td>2.96E-10</td>
<td>1.00</td>
<td>0.46 (0.31-0.67)</td>
<td>0.59 (0.36-0.97)</td>
<td>Transaminic Acid</td>
<td>11 Drugs Launched</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs71751347</td>
<td>1.66%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.8 (0.75-0.86)</td>
<td>3.47E-10</td>
<td>0.83</td>
<td>0.84 (0.77-0.93)</td>
<td>0.60 (0.44-0.81)</td>
<td>Biological testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4252185</td>
<td>0.23%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.60 (0.49-0.74)</td>
<td>2.07E-06</td>
<td>0.99</td>
<td>0.60 (0.44-0.81)</td>
<td>0.78 (0.55-1.1)</td>
<td>Biological testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVW</td>
<td>rs14983</td>
<td>1.56%</td>
<td>P09237</td>
<td>SomaScan</td>
<td>0.75 (0.62-0.92)</td>
<td>5.80E-03</td>
<td>0.98</td>
<td>0.62 (0.51-0.74)</td>
<td>0.83 (0.73-0.95)</td>
<td>Marimastat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP7</td>
<td>rs14983</td>
<td>1.56%</td>
<td>P09237</td>
<td>SomaScan</td>
<td>0.67 (0.58-0.76)</td>
<td>8.84E-10</td>
<td>0.98</td>
<td>0.62 (0.51-0.74)</td>
<td>0.83 (0.73-0.95)</td>
<td>Marimastat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POGLUT3</td>
<td>rs7491261</td>
<td>9.89%</td>
<td>Q7Z4H8</td>
<td>SomaScan</td>
<td>0.90 (0.86-0.93)</td>
<td>9.99E-08</td>
<td>1.00</td>
<td>0.85 (0.8-0.91)</td>
<td>0.97 (0.87-1.1)</td>
<td>HGS-TR2; Lexatumumab</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2293400</td>
<td>4.30%</td>
<td>Q14763</td>
<td>OLINK</td>
<td>0.87 (0.83-0.92)</td>
<td>3.02E-07</td>
<td>1.00</td>
<td>0.73 (0.68-0.79)</td>
<td>1.06 (0.97-1.15)</td>
<td>Glassia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs28929474</td>
<td>16.10%</td>
<td>P01009</td>
<td>SomaScan</td>
<td>1.08 (1.05-1.11)</td>
<td>3.01E-07</td>
<td>1.00</td>
<td>1.05 (1.0-1.1)</td>
<td>1.06 (0.97-1.15)</td>
<td>Glassia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4919682</td>
<td>0.50%</td>
<td>P49902</td>
<td>SomaScan</td>
<td>0.70 (0.6-0.81)</td>
<td>1.43E-06</td>
<td>1.00</td>
<td>0.64 (0.49-0.84)</td>
<td>0.83 (0.64-1.1)</td>
<td>Biological Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs8111</td>
<td>1.31%</td>
<td>Q99941</td>
<td>SomaScan</td>
<td>1.21 (1.12-1.31)</td>
<td>4.92E-06</td>
<td>0.99</td>
<td>1.23 (1.03-1.48)</td>
<td>1.23 (1.13-1.34)</td>
<td>Biological Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs4845586</td>
<td>3.12%</td>
<td>Q8TEY5</td>
<td>SomaScan</td>
<td>1.14 (1.08-1.2)</td>
<td>7.13E-06</td>
<td>0.73</td>
<td>1.23 (1.13-1.34)</td>
<td>1.23 (1.13-1.34)</td>
<td>Biological Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs662463</td>
<td>0.558%</td>
<td>P01570</td>
<td>SomaScan</td>
<td>0.74 (0.65-0.85)</td>
<td>9.04E-06</td>
<td>0.95</td>
<td>0.80 (0.64-0.99)</td>
<td>0.85 (0.73-0.99)</td>
<td>Biological Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs751527</td>
<td>1.36%</td>
<td>Q6UXK2</td>
<td>SomaScan</td>
<td>0.82 (0.75-0.90)</td>
<td>1.78E-05</td>
<td>0.97</td>
<td>0.84 (0.73-0.97)</td>
<td>0.85 (0.73-0.99)</td>
<td>Biological Testing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Table 1 continues on next page)
### Table 1: Mendelian randomisation and colocalisation results for protein-cancer associations that passed Bonferroni correction (0.05/n proteins analyzed) and colocalised at PP4 >0.70.

<table>
<thead>
<tr>
<th>Prostate cancer outcome</th>
<th>Gene</th>
<th>SNP</th>
<th>Proportion variance explained (%)</th>
<th>Uniprot ID</th>
<th>Platform</th>
<th>Odds ratio (95% CI) (practical)</th>
<th>p-value (unadjusted)</th>
<th>Maximum PP4</th>
<th>Odds ratio (95% CI) UKBB/FinnGen</th>
<th>Odds ratio (95% CI) African ancestry</th>
<th>Drug target</th>
<th>Current drug trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early onset</td>
<td>MSMB</td>
<td>rs10993994</td>
<td>49.6%</td>
<td>P08118</td>
<td>SomaScan</td>
<td>0.71 (0.68-0.74)</td>
<td>1.46E-05</td>
<td>1.00</td>
<td></td>
<td></td>
<td>Phenethyl Isothiocyanate</td>
<td>Phase I (Prostate Cancer therapy; PCK-3145)</td>
</tr>
<tr>
<td>PLG</td>
<td></td>
<td>rs11751347</td>
<td>3.57%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.63 (0.53-0.73)</td>
<td>9.38E-09</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td>11 Drugs Launched</td>
</tr>
<tr>
<td>PLG</td>
<td></td>
<td>rs11751347</td>
<td>1.66%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.61 (0.52-0.72)</td>
<td>9.38E-09</td>
<td>0.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLG</td>
<td></td>
<td>rs982403</td>
<td>0.23%</td>
<td>P00747</td>
<td>SomaScan</td>
<td>0.26 (0.14-0.46)</td>
<td>4.07E-06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IVW</td>
<td></td>
<td>rs73934251</td>
<td>0.53%</td>
<td>Q8NDI1</td>
<td>SomaScan</td>
<td>2.90 (2.0-4.2)</td>
<td>2.60E-08</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs2696208</td>
<td>0.68%</td>
<td>P06253</td>
<td>SomaScan</td>
<td>0.47 (0.34-0.64)</td>
<td>1.92E-06</td>
<td>0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs8074783</td>
<td>0.016%</td>
<td>P06815</td>
<td>SomaScan</td>
<td>1.87 (1.43-2.44)</td>
<td>4.36E-06</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs74911261</td>
<td>3.46%</td>
<td>Q74HI8</td>
<td>SomaScan</td>
<td>0.81 (0.73-0.89)</td>
<td>6.67E-06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs8023057</td>
<td>7.95%</td>
<td>P01011</td>
<td>SomaScan</td>
<td>2.08 (1.51-2.88)</td>
<td>8.04E-06</td>
<td>0.99</td>
<td></td>
<td>Zinc; Acetate, Chloride, Sulfate</td>
<td></td>
<td>Preclinical</td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs469179</td>
<td>0.378%</td>
<td>Q9H2U2</td>
<td>SomaScan</td>
<td>2.7 (1.71-4.27)</td>
<td>1.99E-05</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggressive</td>
<td>MSMB</td>
<td>rs10993994</td>
<td>49.6%</td>
<td>P08118</td>
<td>SomaScan</td>
<td>0.84 (0.82-0.86)</td>
<td>1.55E-35</td>
<td>1.00</td>
<td></td>
<td></td>
<td>Phenethyl Isothiocyanate</td>
<td>Phase I (Prostate Cancer therapy; PCK-3145)</td>
</tr>
<tr>
<td>TNRSF68</td>
<td></td>
<td>rs6011040</td>
<td>0.50%</td>
<td>Q95407</td>
<td>SomaScan</td>
<td>0.48 (0.37-0.62)</td>
<td>1.72E-07</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
<td>Preclinical</td>
</tr>
<tr>
<td>PPA2</td>
<td></td>
<td>rs469179</td>
<td>0.38%</td>
<td>Q9H2U2</td>
<td>SomaScan</td>
<td>2.13 (1.54-2.93)</td>
<td>3.73E-06</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td>Preclinical</td>
</tr>
<tr>
<td>C4A</td>
<td></td>
<td>rs2763982</td>
<td>22.0%</td>
<td>P0C0L4</td>
<td>SomaScan</td>
<td>0.91 (0.88-0.95)</td>
<td>8.60E-06</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td>Biological testing</td>
</tr>
<tr>
<td>C4A</td>
<td></td>
<td>rs2763982</td>
<td>22.0%</td>
<td>P0C0L4</td>
<td>SomaScan</td>
<td>0.91 (0.88-0.95)</td>
<td>8.60E-06</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td>Biological testing</td>
</tr>
<tr>
<td>TPM3</td>
<td></td>
<td>rs2696208</td>
<td>0.68%</td>
<td>P0653</td>
<td>SomaScan</td>
<td>0.61 (0.50-0.76)</td>
<td>8.56E-06</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
<td>Preclinical</td>
</tr>
<tr>
<td>FRSS3</td>
<td></td>
<td>rs6555900</td>
<td>3.52%</td>
<td>P35030</td>
<td>SomaScan</td>
<td>0.80 (0.73-0.89)</td>
<td>1.13E-05</td>
<td>0.99</td>
<td></td>
<td>4-(1,3,2-DIOXABOROLAN-2-YLOYX)BUTAN-1-AMIN; 4-HYDROXYBUTAN-1-AMIN; 4-(HYDROXYMETHYL)BENZAMIDINE; BENZAMIDINE; DIAMINO-N-[3-(1,3,2-DIOXABOROLAN-2-YLOYX)PROPYL] METHANMINUM; GUANIDINE-3-PROPANOL; [4-(1,3,2-DIOXABOROLAN-2-YLOYX)METHYL] BENZAMIDINE; 1,3,2-DIOXABOROLAN-2-OL</td>
<td>Phase III (Cancer immunotherapy; 99mTc-ior C5)</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td>rs3094662</td>
<td>0.75%</td>
<td>P06681</td>
<td>SomaScan</td>
<td>0.67 (0.56-0.81)</td>
<td>2.39E-05</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
<td>Preclinical</td>
</tr>
</tbody>
</table>

Results are shown per SNP-cancer association, except for those proteins for which there were multiple SNPs that passed both multiple testing correction and colocalisation, and for which the summary estimate using the Inverse Variance Weighted (IVW) method is provided. Odds ratios (95% confidence intervals) are oriented per standard deviation increase in genetically predicted protein level. Maximum PP4 is reported as highest PP4 value from either single or conditional iterative colocalisation method assessing the probability of a shared causal locus. Odds ratios (95% confidence intervals) are reported for associations using external UK Biobank/FinnGen and for an African ancestry population where data existed. Drug targets and drug trials are annotated where they existed.
Additional genome-wide random forest analyses identified MSMB expression as the most important gene to distinguish between benign and GG4 cells (Fig. 5).

Discussion
In this analysis, we conducted the largest study to date investigating the genetic associations of up to 2002 unique proteins with the likely aetiological risk of prostate cancer.
In this study, we expand upon these previous findings by demonstrating that higher levels of genetically predicted MSMB are associated with a 20% lower risk of overall prostate cancer in two independent European ancestry cancer GWAS, and confirm its protective role in both aggressive and early onset disease aetiology. We further successfully replicated this association with prostate cancer risk overall in an African-ancestry population, an ancestry group with an established higher risk of developing the disease. Subsequently, in reporting that MSMB gene expression is significantly depleted among high grade tumour when compared to expression in benign cells, we reiterate, through an independent line of evidence, that this gene may be particularly relevant to tumourigenesis and risk for aggressive disease.

Although the mechanism of action for MSMB in prostate cancer is not clear, MSMB has been shown to have a regulatory effect on cell growth, which may be lost during tumourigenesis while a MSMB-derived polypeptide was shown to induce prostate cell death. \(^{37,38}\) Additionally, in a rodent model and in vitro, higher MSMB activity was found to suppress prostate tumour growth while a knockout of MSMB promoter/enhancer regions was characterized by tumour progression and metastases. \(^{39,40}\) Given the integration of several compelling lines of evidence presented in this paper with existing literature, further research is warranted to understand the precise functional role of MSMB and integrated evidence of prostate cancer risk

In this study, MSMB was one of only two proteins (including TPM3) to significantly associate and colocalise with all three prostate cancer outcomes. MSMB is a secretory protein and member of the human immunoglobulin family that is released largely by luminal epithelial cells in the prostate epithelium and has a documented role in overall prostate cancer risk for both observational and genetic epidemiology. \(^{6,7,37}\) In this overall, early onset, and aggressive prostate cancer in up to 177,526 men using a MR and colocalisation pipeline. In total, we found evidence supporting associations between 20 proteins and prostate cancer risk: 14 proteins for overall prostate cancer risk, seven for aggressive prostate cancer and eight for early onset disease. Among those 14 proteins that associated with prostate cancer risk overall, 14 were replicated in an external European ancestry population and three (out of six with available data) were replicated in an African ancestry population. A further half of the 20 proteins identified were also found to be the site of action for established drug targets with potential therapeutic implications. Finally, using spatial transcriptomics data, we demonstrated a central role for the gene of our most robustly associated prostate cancer protein, MSMB, in distinguishing benign from undifferentiated, high-grade prostate cancer cells.
MSMB in prostate cancer tumourigenesis, identify environmental and lifestyle determinants, and explore its potential clinical utility.

**CREB3L4 and CREBL1 and risk of prostate cancer overall**

We identified 14 proteins associated with overall prostate cancer aetiology. These include the endoplasmic reticulum (ER) originating transcription factors CREB3L4 and CREBL1 (also known as ATF6B), which we find are associated with an 14% and 21% increased risk of prostate cancer overall, respectively, and that are expressed in the prostate epithelium. These transcription factors form part of a transcriptional factor network that regulates the function of the endoplasmic reticulum (ER) and the activity of the unfolded protein response (UPR). There is an established role for the UPR and heat-shock proteins in maintaining androgen receptor (AR) stability and supporting AR-dependent tumourigenesis. CREB3L4 has been shown to directly interact with AR in LNCaP cells to increase cellular proliferation and is abundantly expressed in prostate tumour tissue. Furthermore, evidence suggests that disruptions in CREB3L4 contribute to ER stress downstream and initiation of the unfolded protein response. In a previous study of differential gene expression in prostate tissue, CREB3L4 was identified as a member of a co-expression gene cluster enriched for a previously described metabolic pathway (hsa05215) in prostate cancer that may regulate apoptosis and cell proliferation. Interestingly, we found that CREB3L4 expression is lower in GG2 and GG4 cells as opposed to benign cells.

Previous studies have linked AR activation with members of the ATF6 family in LNCaP and PC3 cells, however these experiments have mostly focused on ATF6A. For example, a recent in vivo study showed that prostate cancer cells with ATF6A overexpression resisted cellular death by ferroptosis. In parallel, a previous MR study reported a lowered risk of prostate cancer overall with genetically elevated circulating ATF6A levels from trans-pQTL. Given the promising role of its paralog, and the increased risk we report here, targeted follow up of CREBL1 may prove valuable.

**Fig. 5:** Gene network from iterative random forests of the difference in gene expression between benign and GG4 prostate histology (Gleason Score = 8). Arrows indicate direction of influence and shape of the network. MSMB is coloured to demonstrate its central role in the network.
characterising the broader role of ER stress proteins and AR-dependent tumourigenesis.

Other aggressive and early onset prostate cancer proteins
We additionally found several proteins that were associated with risk for either aggressive or early onset disease, including PRSS3, PPA2, EHBPl, PYY, C2 and C4A. PR33S, a member of the trypsin family, was associated with a 20% reduced odds of aggressive disease in our study. While one previous study found that mesotrypsin, a protease encoded by the PRSS3 gene, was essential for prostate cancer metastasis in vitro and mouse models, the role of this protein and other isoforms have not been widely studied in humans. Likewise, we found PPA2, a serine/threonine phosphatase which has been shown to account for up to 50% of global cellular phosphatase activity, was associated with a two-fold increased odds of aggressive disease. While few studies have directly assessed the relationship between PPA2 and risk of prostate cancer prospectively, a proteomic tissue analysis comparing men with localised prostate cancer to men with lymph node metastatised prostate cancer found significantly higher levels of PPA2 in men with the more advanced disease.

We observed a more than two-fold increased risk of early-onset prostate cancer associated with higher EHBPl, an adaptor protein with a key role in vesicular trafficking and actin reorganization. Variants in the EHBPl intron have previously been associated with aggressive prostate cancer in a genetic association study of Europeans and North Americans. The protein was also found to be expressed in prostate tumour tissue and may have a role in determining the invasiveness of PTEN-positive prostate cancer cells, according to GWAS and expression data from The Cancer Genome Atlas, and in a cellular study. While mechanisms that may link EHBPl to prostate cancer risk are not yet fully described, it has a role as an effector molecule for Rab8 family members that modulate polarised membrane transport via actin reorganisation and may have a role in the mechanism of action for atorvastatin. Likewise, we found that PYY, a metabolic hormone involved in appetite regulation, was also associated with an increased odds of early onset disease in our study. While there has been some hypothesised relationship between obesity and aggressive prostate cancer risk in the past, recent findings suggest that obesity does not serve as a risk factor for disease itself, but rather may affect likelihood of diagnosis.

We also note that proteins associated with early onset disease were generally greater in magnitude when compared to their associations with overall or aggressive disease. Two of these proteins, the complement proteins C2 and C4A, sit on chromosome six, which contains a particularly dense genetic region including the MHC complex and is consequently particularly difficult to interpret. However, given the importance of addressing early onset and aggressive disease, future studies are needed to further investigate and replicate the associations with early onset disease to uncover potential subtype specific mechanisms of disease onset and progression.

Drug target proteins
We identified 10 proteins that were associated with both the risk of prostate cancer and that were the site of action for a known drug. These included TNFRSF10B, which is a receptor for the cytotoxic TRAIL ligand, and is essential for CASP8 and ER stress induced apoptosis. Further, TNFRSF10B expression is lower in higher grade prostate tumours and a recent study of PARP inhibitors in prostate cancer cell lines suggested that TNFRSF10B may provide a mechanism by which the cancer drug olaparib induces apoptosis. The apparent protective association we observe with prostate cancer risk is in line with the results from multiple phase I/II trials of TNFRSF10B agonists that support their use for the treatment of multiple cancer endpoints, though not yet including prostate cancer.

We also identified an inverse association of PLG, a serine protease targeted by tranexamic acids and several classes of thrombolitics, with prostate cancer risk overall and with early onset disease. Tranexamic acids are primarily prescribed to control excessive bleeding while thrombolitics are primarily used to dissolve blood clots and act via plasmin and fibrin pathways. One molecular study found that PLG is generated by the cancer-mediated proteolysis of plasminogen which is released by human prostate carcinoma cells. PLG in turn has been shown in many lab studies to inhibit angiogenesis which when unregulated can lead to the rapid formation of tumours. Currently, several studies are investigating combination therapy including plasminogen activation or inhibition for treatment of several cancer types, though not specifically for prostate cancer, in phase I/II trials.

SERPINA1 maps to fazisiran, the treatment for alpha-1 antitrypsin deficiency and that is in phase II/III of drug trials. Previous findings have indicated that alpha-1 antitrypsin levels are often elevated in many carcinomas, including prostate. However, no agents targeting SERPINA1 have been investigated in cancer trials thus far. MMP7 belongs to a class of matrix metalloproteinases that participate in wound healing, bone growth, and matrix remodelling. There are multiple lines of evidence that suggest this protein is involved in many cancers, and agents targeting metalloproteinases, such as marimastat, are currently being investigated in clinical trials at various phases. In prostate cancer, marimastat showed some efficacy in early trials, however has not yet progressed further.

While we highlight proteins that share the same target site for established drug targets that may have
implications for therapeutic use, the suitability of these to act as preventative or remedial agents requires careful considerations including: site specificity, potential downstream effects, routes of administration and effectively capturing the population at risk. 31

**Study strengths**

This study offers several strengths including being the largest currently available GWAS of prostate cancer outcomes and the use of both aggressive and early onset endpoints with cis-pQTL covering up to 2002 proteins. One previous MR study investigated the role of the circulating proteome in prostate cancer risk but did not stratify analyses by cis or trans-pQTL, and did not perform colocalisation analyses, making it more challenging to infer causal relationships between individual proteins and cancer risk. 49 Additionally, by integrating gene expression data measured using spatial transcriptomics data, the current paper introduces a translational approach to highlight biological enablers of prostate cancer. To our knowledge, this study provides the first demonstration that MR using cis instruments of plasma protein levels can be used to identify a risk protein that has both specific expression in the cell of cancer origin and is related to tumour aggressiveness—important features to consider when identifying candidate targets for therapeutic prevention.

**Caveats and limitations**

While we have analysed a wide array of proteins, we have not investigated the entire human plasma proteome (n ~ 20,000 protein-coding genes). As more protein GWAS data become available, it will become possible to use genetic methods to investigate more proteins. However, some blood proteins are unlikely to have a cis-pQTL due to the degree of evolutionary constraint for a protein’s cognate gene. Additional limitations include the more modest GWAS sample sizes for aggressive and early onset prostate cancer, which have lower power to discover novel protein associations. Finally, while we were able to perform additional analyses to replicate some, but not all, of our robust proteins in populations of African ancestry, we note as a limitation that GWAS sample sizes in this group are not yet sufficient to perform well-powered discovery analyses. Especially given the increased risk for prostate cancer among populations of African ancestry, it is essential that future studies identify risk proteins in more diverse populations and allow for the discovery of ancestry-specific markers of risk.

**Conclusion**

This paper provides a catalogue of 20 proteins with evidence of aetiological significance for prostate cancer. These proteins present an opportunity to direct further molecular and epidemiological investigations aimed at exploring the specific roles that the proteome plays in tumourigenesis and ultimately may inform future research into therapeutic prevention. In particular, converging evidence from population genetic and tumour sequencing analyses implicates MSMB as having an important protective role in prostate tumourigenesis, both in European and African-ancestry men, which is particularly marked for aggressive and early onset disease.

**Contributors**

All authors have read and approved the final version of the manuscript. TAD and KSB accessed and verified the underlying data. Authors contributed as following:

- **Conceptualization:** RCT, KSB, TJK, TAD, AKH, MD.
- **Data Curation:** PRACTICAL Consortium, KSB, AKH, JRT, CL, MP, MK, AM, ADL, SF, WY.
- **Formal Analysis:** TAD, SF, WY, MD, AVS.
- **Funding Acquisition:** RCT, TJK.
- **Methodology:** KSB, RCT, TD.
- **Supervision:** KSB, RCT, TJK, ELW.
- **Writing Original Draft:** TAD, KSB.

**Reviewing and Editing:** All authors: TAD, AKH, MD, SF, WY, ELW, JRT, AVS, HBS, KKT, MJG, RMM, MJ, MP, CL, IGM, ADL, AM, TJK, RCT, KSB.

Data used in this study were generated by the PRACTICAL Consortium and all affiliated members were invited to review the final manuscript.

**Data sharing statement**

PQTL data were obtained through publicly available GWAS. Summary statistics from Zheng et al. (2020) can be obtained from OpenGWAS (https://gwas.mrcieu.ac.uk/), from Folkesen et al. (2020) at http://www.scalarp-consortium.com, from Perkingedal et al. (2021) at https://www.decode.com/summarydata/, and from Pietzner et al. (2021) at https://omicscience.org. Summary statistics for overall prostate cancer were available through the PRACTICAL Consortium (http://practical.icr.ac.uk/?page_id=8164) and approval was received to obtain summary statistics for further cases using [dbGaP] project #31553, and for aggressive and early onset disease from the PRACTICAL Consortium (for enquiries contact PRACTICAL@icr.ac.uk).

**Declaration of interests**

This work was supported by Cancer Research UK (grant no. CR221/A29017). Anders Målarstig, Åsa Hedman, and Marios Dimitriou are employees of Pfizer Inc. Anders Målarstig declares stock options for Pfizer Inc. Alastair D. Lamb is Section Editor for Prostate Cancer and Web, British Journal of Urology International.

**Acknowledgements**

The authors would like to thank the participants of all studies who volunteer their time and efforts to make these analyses possible. Furthermore, the authors would also like to acknowledge the investigators and support staff of consortia who generated the data used in this study.

TAD is supported by a Cancer Research UK studentship grant number (CR221/A30904). KSB is supported by Cancer Research UK (grant nos. CR221/A29017 and C16077/A29186) and UKRI grant no. 10063259. RCT has support from Cancer Research UK (grant no. CR221/A29017). ELW is supported by the Intramural Research Program of the National Institutes of Health (NIH). RMM is a National Institute for Health Research Senior Investigator (NIHR202411). RMM is supported by a Cancer Research UK Research Centre which is funded by the NIHR and is a partnership between University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. RMM is affiliated with the Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is
supported by the Medical Research Council (MC_UU_00011/1, MC_UU_00011/3, MC_UU_00011/6, and MC_UU_00011/4) and the University of Bristol. SF is funded by the The Hanson Trust. ADL is supported by Cancer Research UK and the Hanson Trust (Grant nos. CS7899/A23812 and HJD00160, respectively), and also holds grants from National Institute for Health Research, Health Technology Assessment (HTA) (grant no. NIHR131233), John Black Chrest Foundation, and from the John Fell Foundation (grant no. 001782).

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. The findings and conclusions in this article are those of the authors and do not necessarily represent the official position of the National Institutes of Health. Department of Health and Social Care disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2024.105168.

References
8 Burgess S, Foley CN, Zuber V. Inferring causal relationships be-}


www.thelancet.com Vol 105 July, 2024 15