

1 **Identifying proteomic risk factors for overall, aggressive and early onset prostate cancer**  
2 **using Mendelian randomization and tumor spatial transcriptomics**

3

4 Trishna A Desai<sup>1</sup>, Åsa K Hedman<sup>2,3</sup>, Marios Dimitriou<sup>2,3</sup>, Mine Koprulu<sup>4</sup>, Sandy Figiel<sup>5</sup>,  
5 Wencheng Yin<sup>5</sup>, Mattias Johansson<sup>6</sup>, Eleanor L Watts<sup>7</sup>, Joshua R Atkins<sup>1</sup>, Aleksandr V Sokolov<sup>8</sup>,  
6 Helgi B Schiöth<sup>8</sup>, Marc J Gunter<sup>6,9</sup>, Konstantinos K Tsilidis<sup>9,10</sup>, Richard M Martin<sup>11,12,13</sup>, Maik  
7 Pietzner<sup>4,14,15</sup>, Claudia Langenberg<sup>4,14,15</sup>, Ian G Mills<sup>5</sup>, Alastair D Lamb<sup>5</sup>, Anders Mälarstig<sup>2,3</sup>,  
8 Tim J Key<sup>1</sup>, The PRACTICAL Consortium, Ruth C Travis<sup>1\*</sup>, Karl Smith-Byrne<sup>1\*</sup>

9

10 <sup>1</sup>Cancer Epidemiology Unit, Oxford Population Health, University of Oxford, Oxford, United  
11 Kingdom,

12 <sup>2</sup>External Science and Innovation, Pfizer Worldwide Research, Development and Medical,  
13 Stockholm, Sweden,

14 <sup>3</sup>Department of Medicine, Department of Medicine, Stockholm, Sweden,

15 <sup>4</sup>MRC Epidemiology Unit, University of Cambridge, United Kingdom

16 <sup>5</sup>University of Oxford, Nuffield Department of Surgical Sciences, Oxford, United Kingdom

17 <sup>6</sup>Genomic Epidemiology Branch, International Agency for Research on Cancer (IARC-WHO),  
18 Lyon, France

19 <sup>7</sup>Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National  
20 Cancer Institute, Rockville, Maryland, United States of America

21 <sup>8</sup>Department of Surgical Sciences, Functional Pharmacology and Neuroscience Uppsala  
22 University, 75124 Uppsala, Sweden.

23 <sup>9</sup>Department of Epidemiology and Biostatistics, School of Public Health, Imperial College  
24 London, St Mary's Campus, Norfolk Place, London, W2 1PG, United Kingdom

25 <sup>10</sup>Department of Hygiene and Epidemiology, University of Ioannina School of Medicine,  
26 Ioannina, Greece

27 <sup>11</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United  
28 Kingdom

29 <sup>12</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom,

30 <sup>13</sup>NIHR Bristol Biomedical Research Centre, Hospitals Bristol and Weston NHS Foundation  
31 Trust and the University of Bristol, Bristol, United Kingdom

32 <sup>14</sup> Computational Medicine, Berlin Institute of HealthHealth (BIH) at Charité -  
33 Univeritätsmedizin– Universitätsmedizin Berlin, Berlin, Germany

34 <sup>15</sup> Precision Healthcare University Research Institute, Queen Mary University of London,  
35 London, United Kingdom

36

37 **\* These authors contributed equally.**

38 **Running Title:** Proteins and prostate cancer risk

39 **Keywords:** Proteomics, cis-pQTL, Prostate Cancer, plasma, protein

40

41 **Funding:** This work was supported by Cancer Research UK (grant no. C8221/A29017). TAD is  
42 supported by a Cancer Research UK studentship grant number (C8221/A30904). KSB is  
43 supported by Cancer Research UK (grant nos. C8221/A29017 and C16077/A29186) and UKRI  
44 grant no. 10063259. ELW is supported by the Intramural Research Program of the National  
45 Institutes of Health (NIH). RMM is a National Institute for Health Research Senior Investigator

46 (NIHR202411). RMM is supported by a Cancer Research UK 25 (C18281/A29019) program  
47 grant (the Integrative Cancer Epidemiology Program). RMM is also supported by the NIHR  
48 Bristol Biomedical Research Centre which is funded by the NIHR and is a partnership between  
49 University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol.  
50 RMM is affiliated with the Medical Research Council Integrative Epidemiology Unit at the  
51 University of Bristol which is supported by the Medical Research Council (MC\_UU\_00011/1,  
52 MC\_UU\_00011/3, MC\_UU\_00011/6, and MC\_UU\_00011/4) and the University of Bristol.  
53 Sandy Figiel is funded by the The Hanson Trust.

54

55 **Disclaimer:** Where authors are identified as personnel of the International Agency for  
56 Research on Cancer / World Health Organization, the authors alone are responsible for the  
57 views expressed in this article and they do not necessarily represent the decisions, policy, or  
58 views of the International Agency for Research on Cancer / World Health Organization. The  
59 findings and conclusions in this article are those of the authors and do not necessarily  
60 represent the official position of the National Institutes of Health. Department of Health and  
61 Social Care disclaimer: The views expressed are those of the author(s) and not necessarily  
62 those of the NHS, the NIHR or the Department of Health and Social Care. Anders Malarstig,  
63 Åsa Hedman, and Marios Dimitriou are employees of Pfizer Inc.

64

65 **Abstract (word count 375)**

66

## 67 **Background**

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

70

## 71 **Methods**

72 We investigated the association of 2,002 genetically predicted circulating protein levels with  
73 risk of prostate cancer overall, and of aggressive and early onset disease, using *cis*-pQTL  
74 Mendelian randomization (MR) and colocalization. Findings for proteins with support from  
75 both MR, after correction for multiple-testing, and colocalization were replicated using two  
76 independent cancer GWAS, one of European and one of African ancestry. Proteins with  
77 evidence of prostate-specific tissue expression were additionally investigated using spatial  
78 transcriptomic data in prostate tumor tissue to assess their role in tumor aggressiveness.  
79 Finally, we mapped risk proteins to drug and ongoing clinical trials targets.

80

## 81 **Results**

82 We identified 20 proteins genetically linked to prostate cancer risk (14 for overall [8 specific],  
83 7 for aggressive [3 specific], and 8 for early onset disease [2 specific]), of which a majority  
84 were novel and replicated. Among these were proteins associated with aggressive disease,  
85 such as PPA2 [Odds Ratio (OR) per 1 SD increment = 2.13, 95% CI: 1.54-2.93], PYY [OR = 1.87,  
86 95% CI: 1.43-2.44] and PRSS3 [OR = 0.80, 95% CI: 0.73-0.89], and those associated with early  
87 onset disease, including EHPB1 [OR = 2.89, 95% CI: 1.99-4.21], POGlut3 [OR = 0.76, 95% CI:  
88 0.67-0.86] and TPM3 [OR = 0.47, 95% CI: 0.34-0.64]. We confirm an inverse association of  
89 MSMB with prostate cancer overall [OR = 0.81, 95% CI: 0.80-0.82], and also find an inverse  
90 association with both aggressive [OR = 0.84, 95% CI: 0.82-0.86] and early onset disease [OR =  
91 0.71, 95% CI: 0.68-0.74]. Using spatial transcriptomics data, we identified MSMB as the  
92 genome-wide top-most predictive gene to distinguish benign regions from high grade cancer

93 regions that had five-fold lower MSMB expression. Additionally, ten proteins that were  
94 associated with prostate cancer risk mapped to existing therapeutic interventions.

95

## 96 **Conclusion**

97 Our findings emphasize the importance of proteomics for improving our understanding of  
98 prostate cancer etiology and of opportunities for novel therapeutic interventions.  
99 Additionally, we demonstrate the added benefit of in-depth functional analyses to triangulate  
100 the role of risk proteins in the clinical aggressiveness of prostate tumors. Using these  
101 integrated methods, we identify a subset of risk proteins associated with aggressive and early  
102 onset disease as priorities for investigation for the future prevention and treatment of  
103 prostate cancer.

104

## 105 **Introduction**

106

107 Prostate cancer is a heterogeneous disease with a high survival rate for those diagnosed with  
108 indolent or low-stage disease, but a less than 50% 5-year survival rate for those diagnosed  
109 with aggressive or metastatic cancer.<sup>1</sup> The proportion of these clinically aggressive cases is  
110 higher among men younger than 55 years (early onset disease), which contributes to  
111 premature death among these men.<sup>2,3</sup> However, few risk factors for prostate cancer have  
112 been established. These include: advanced age, African ancestry, family history of the disease,  
113 circulating levels of insulin-like growth factor I and microseminoprotein-beta (MSMB), with  
114 little evidence for successful strategies for prevention.<sup>4-7</sup>

115

116 Recent advances in multiplexed and high throughput platforms as well as the widespread  
117 availability of genotypic arrays have identified genetic variants that determine circulating  
118 levels of thousands of circulating proteins, known as protein-quantitative trait loci (pQTL).  
119 pQTL, in particular those lying in or near a protein's cognate gene (referred to as *cis*-pQTL),  
120 can be leveraged to identify candidate etiological proteins for cancer risk through Mendelian  
121 randomization (MR) analyses, an approach that can limit the impact of reverse causality.<sup>8,9</sup>  
122 MR can also be complemented with colocalization analyses to further exclude confounding  
123 by linkage disequilibrium (LD).<sup>10</sup> Candidate etiological proteins for cancer risk identified using  
124 these methods can provide a valuable starting point for further analyses using more resource-  
125 intense methods, such as spatial transcriptomics, where their functional importance at the  
126 tissue level can be directly interrogated to triangulate their role in etiology.<sup>11,12</sup>

127

128 Using an integrated *cis*-pQTL MR and colocalization pipeline, we analyzed the associations of  
129 2,002 unique proteins with overall, aggressive, and early onset prostate cancer and replicated  
130 and mapped those with significant findings to drug targets. Additionally, we investigated the  
131 spatial distribution and gene expression profiles of a subset of these proteins in prostate  
132 tumor tissue using spatial transcriptomics. In doing so, we demonstrate the value of protein  
133 MR and colocalization analyses to identify proteins that may have a causal role in the tumor  
134 aggressiveness.

135

136

137

138

139

## 140 **Methods**

141

### 142 *Overall study design*

143

144 We extracted *cis* genetic instruments for circulating protein levels from publicly available  
145 datasets, and harmonized these *cis*-pQTL with the GWAS results from an international  
146 prostate cancer consortium, including aggressive and early onset subtypes (Supplementary  
147 Figure 1). We subsequently estimated risk associations for protein levels using *cis*-pQTL MR  
148 against each of these three prostate cancer endpoints. All associations passing a multiple  
149 testing threshold in MR analyses were then followed up with colocalization analyses. Where  
150 data were available, we performed replication analyses in an external prostate cancer GWAS  
151 in European and African ancestry (using African ancestry specific *cis*-pQTLs – see below)  
152 populations for proteins with evidence from MR and colocalization analyses. For proteins  
153 identified as risk factors for prostate cancer with evidence of specific expression in the  
154 prostate tissue, we additionally performed analyses using spatial transcriptomics to gain  
155 insights into the spatial distribution and gene expression patterns of these proteins in  
156 prostate tumor samples. Finally, we conducted an exploratory analysis restricting to *cis*-pQTL  
157 whose cognate genes are established drug targets.

158

### 159 *Identification of cis-pQTL*

160

161 Genetic instruments for *cis*-pQTL were extracted from 4 publicly available protein GWAS at  $p$   
162  $< 5 \times 10^{-8}$  and clumped at  $R^2=0.01$  within their originating panel (instruments presented in  
163 Supplementary Table 1).<sup>13–16</sup> *Cis*-instruments were defined in the first instance as those that  
164 were genome-wide significant ( $p < 5 \times 10^{-8}$ ) within 1 Mb of the transcription start side of the  
165 measured protein encoded gene, or as the sentinel *cis*-pQTL for the measured protein  
166 depending on data availability. We additionally gathered data on *cis*-pQTL from published  
167 GWAS present on the OpenGWAS platform using a relaxed p-value threshold of  $5 \times 10^{-5}$  due  
168 to the high biological plausibility of identifying *cis*-pQTL at or near a protein's cognate  
169 gene.<sup>17,18</sup> Specifically, we extracted unreported *cis*-pQTL from the genomic region 1 megabase  
170 up and downstream of the cognate gene for a given protein GWAS (Supplementary Table 1).  
171 We subsequently extracted all instruments where no *cis*-pQTL was present at  $p < 5 \times 10^{-8}$  but  
172 at least one *cis*-pQTL was present at  $p < 5 \times 10^{-05}$ .

173

174 All instruments were mapped to Uniprot IDs, and *cis*-pQTL with weak instrument strength at  
175  $F_{\text{stat}} < 10$  [ $\beta^2/\sigma^2$ ] were excluded from the study. For *cis*-pQTL that were not present in the  
176 cancer outcome data, SNP proxies were selected at  $r^2_{\text{max}}$  where  $r^2 > 0.8$  in 1000 genomes CEU  
177 population with the index *cis*-pQTL. In total, 2,002 unique plasma proteins that fit these  
178 criteria were included in analyses.

179

### 180 *Cancer outcome data*

181

182 Genetic associations for overall, aggressive, and early onset prostate cancer were obtained  
183 from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in  
184 the Genome (PRACTICAL) consortium (Supplementary material).<sup>19</sup> Full study characteristics  
185 have been described previously, but briefly, summary statistics for SNP associations with  
186 prostate cancer and subtypes were generated from the PRACTICAL consortium using 85,554

187 overall prostate cancer cases and 91,972 controls (database of Genotypes and Phenotypes  
188 [dbGaP] project #31553), 15,167 aggressive PC cases and 58,308 controls, and 6,988 cases of  
189 early onset disease and 44,256 controls, all of European ancestry.<sup>20,21</sup> Aggressive prostate  
190 cancer is defined in PRACTICAL as cases having metastatic disease or Gleason score  $\geq 8$  or  
191 PSA  $>100$  ng/mL or prostate cancer death. Early onset PC cases are defined as those diagnosed  
192 before the age of 55 years. Genotype information was imputed for samples using the 2014  
193 release of the 1000 Genomes Project as a reference panel.

194

#### 195 *Two-sample Mendelian randomization*

196

197 *Cis*-pQTL data were harmonized to each cancer outcome by rsID and oriented to the protein-  
198 increasing allele. Two-sample Mendelian randomization (MR) was subsequently performed  
199 for each *cis*-pQTL on risk of overall, aggressive, and early onset prostate cancer using the  
200 Wald-ratio method ( $\beta_{cancer}/\beta_{protein}$ ). Resulting associations where the  $p_{Wald}$  passed a  
201 Bonferroni-corrected threshold of significance based on the total number of unique proteins  
202 assessed for each of the three prostate cancer outcomes were taken forward in analyses  
203 ( $p_{Wald} < 0.05/N_{Proteins}$  analyzed per cancer outcome).<sup>22</sup> Multiple independent [ $r^2 < 0.01$ ] *cis*-pQTL that  
204 proxied the same protein and were both associated after correction for multiple testing and  
205 that colocalised with the same prostate cancer outcome were combined using the inverse-  
206 variance weighted method (IVW). Odds ratio estimates are scaled per standard deviation  
207 increment in relative and normalized circulating protein concentrations.

208

#### 209 *Colocalization*

210

211 Colocalization was performed to assess the probability that the protein and cancer  
212 instruments share a causal variant, fulfilling an important instrumental variable assumption  
213 of MR.<sup>9,10</sup> Specifically, single and conditional iterative colocalization analysis were performed  
214 for all *cis*-pQTL MR results that passed a Bonferroni correction for multiple testing based on  
215 the number of unique proteins in the study ( $p < 0.05/N_{Proteins}$ ), using all variants within a 75kb  
216 region up- and downstream from the index *cis*-pQTL to assess confounding by linkage  
217 disequilibrium.<sup>10,23</sup> To mitigate the chance of false-positive findings, we selected priors of P1:  
218  $1 \times 10^{-3}$ , P2:  $1 \times 10^{-4}$ , and P12:  $1 \times 10^{-5}$ , which roughly equate to a 0.1% prior belief in  
219 colocalization (PP4).<sup>24</sup> We defined a threshold PP4 in support of a shared association for a  
220 protein and cancer signal at 0.70 to take proteins forward for subsequent analysis and the  
221 highest PP4 of any method of colocalization was recorded to assess confidence in the shared  
222 association for each SNP assessed.

223

#### 224 *Replication of robust proteins in European and African ancestry populations*

225

226 We conducted a replication analysis of *cis*-pQTL MR associations passing multiple testing  
227 correction and that colocalized (referred to as *robust* proteins) using an external GWAS in a  
228 European ancestry population of overall prostate cancer risk. GWAS summary statistics  
229 represented a meta-analysis in FinnGen r9 and the UK Biobank (20,907 cases & 289,710  
230 controls).<sup>25,26</sup> Additionally, where possible, we performed replication analyses using *cis*-pQTL  
231 identified in an African ancestry protein GWAS in the Atherosclerosis Risk in Communities  
232 study (4,657 proteins in 467 African-ancestry participants) and a GWAS of overall prostate  
233 cancer among African-ancestry populations obtained from dbGaP (project #31553) containing

234 data from the AAPC GWAS, Ghana Prostate Study, ProHealth Kaiser GWAS, and ELLIPSE  
235 OncoArray (10,368 cases and 10,986 controls).<sup>21,27</sup> GWAS for aggressive and early onset  
236 prostate cancer were unavailable to use as a replicate sample in either ancestry population.  
237 We considered a directionally concordant risk estimates and Wald ratio  $p < 0.05$  using  
238 external data to indicate replication. No sample overlap was present between samples used  
239 to generate protein associations and used to conduct replication analyses.

240

#### 241 *Drug target pQTL analyses*

242

243 We restricted our MR results to those *cis*-pQTL that share a cognate gene that is an  
244 established drug target by reference to the DrugBank, Therapeutic Target Database, Pharos  
245 consortium, ClinicalTrials.gov or expert curation.<sup>28–30</sup> As above, we defined *robust*  
246 associations as Wald  $p < 0.05/N_{\text{Proteins}}$ , where  $N_{\text{Proteins}}$  is the number of unique proteins  
247 analyzed for a given cancer outcome that were identified as the cognate gene of a  
248 pharmaceutical target and  $PP4 > 0.7$ . Additionally, all proteins identified in overall and drug  
249 target analyses were queried the Cortellis database (<https://www.cortellis.com>) to assess the  
250 highest current level of clinical development stage.

251

252 Statistical analysis was performed in R version 4.1 and all tests of significance were two-sided,  
253 where  $P$  values  $< 0.05$  were considered statistically significant. MR analyses were performed  
254 using the *TwoSampleMR* R package and colocalization analyses were performed using the  
255 *coloc* R package.<sup>17,23</sup>

256

#### 257 *Gene expression analysis using spatial transcriptomics*

258

259 Spatial transcriptomics provides a spatial map of gene expression within the target tissue.  
260 This spatial information can be used to investigate the variation in gene expression by healthy  
261 tissue and tumor type intratumorally, and as a result, it can provide valuable insights into  
262 tumorigenesis and inform causal inference in this molecularly heterogeneous disease.<sup>31,32</sup>  
263 Spatial transcriptomic analysis was performed for those proteins passing multiple testing  
264 correction and that colocalized and also showed high expression in the prostate epithelium.<sup>33</sup>  
265 Data for spatial transcriptomics were obtained from our previously published dataset derived  
266 from radical prostatectomy tissue taken from a patient with multifocal prostate cancer.<sup>34</sup> Our  
267 analysis focused on eight distinct tissue sections, which collectively comprised 32,156 spots,  
268 some of which contained regions of cancer as well as histo-pathologically benign prostate  
269 tissue, and some of which did not contain cancer. To ensure data quality, samples with less  
270 than 500 Unique Molecular Identifier (UMI) counts were excluded from the analysis. The  
271 initial fastq files were processed using the 10x Visium Spaceranger software, enabling the  
272 conversion of the files into gene expression data. Subsequently, the data underwent  
273 SCTransform normalization and variance reduction procedures. A consensus pathology  
274 approach was employed involving two pathologists who independently annotated each  
275 spatial transcriptomics spot, with the aim to include those that predominantly contained  
276 epithelial cells, which comprised approximately 1-15 cells. Violin plots were generated using  
277 Graphpad Prism (version 10).

278

279

280

## 281 *Iterative random forest network using spatial transcriptomics*

282

283 We used the iterative random forest (iRF) method to investigate gene interactions.<sup>35</sup> With  
284 this method, we randomly selected genes and constructed random forests with other genes  
285 as branches to identify the most robust gene expression network. The analysis specifically  
286 focused on comparing the gene interactions between benign and Gleason grade group 4  
287 histology status. The criteria we used to select credible random forest model was stability >  
288 0.8 and precision > 0.8. The resulting gene network was visualized using Gephi 0.99.

289

## 290 **Results**

291

292 We investigated the associations of 2,002 unique proteins using 4,592 *cis*-pQTL that  
293 harmonized with the GWAS summary statistics for at least one of overall (1,999 proteins;  
294 4,582 *cis*-pQTL), aggressive (1,986 proteins; 4,543 *cis*-pQTL), or early onset prostate cancer  
295 (1,984 proteins; 4,534 *cis*-pQTL) (Figure 1). From these analyses we identified 20 proteins that  
296 were associated, after correction for multiple testing, with at least one of overall (14  
297 proteins), aggressive (7 proteins), or early-onset (8 proteins) prostate cancer and with support  
298 from colocalization analyses (Figure 2, Table 1). Of the 20 proteins associated with any  
299 prostate cancer outcome, several showed robust associations in only one outcome, including  
300 seven that appeared specific to overall prostate cancer (5NTC, CREBL1, INFA14, ISLR2, MMP7,  
301 SERPINA1, TNSFRS10B), three that appeared specific to aggressive disease (C4A, C2,  
302 TNFRSF6B), and two that appeared specific to early onset disease (SERPINA3, PYY).

303

304 The most statistically significant associations per standard deviation increase in protein level  
305 with evidence of colocalization were seen for MSMB (a protein that is specifically expressed  
306 in the prostate) with a lower risk of all prostate cancer endpoints [OR<sub>Overall</sub> = 0.81, 95% CI: 0.79-  
307 0.82, PP4: 100%; OR<sub>Aggressive</sub> = 0.84, 95% CI: 0.82-0.86, PP4: 0.99; OR<sub>Early Onset</sub> = 0.71, 95% CI:  
308 0.68-0.74, PP4: 1.0, Table 1, Figure 2]. TPM3 was the only other protein that had a colocalised  
309 association with risk of all outcomes [OR<sub>Overall</sub> = 0.64, 95% CI: 0.57-0.73, PP4: 0.94; OR<sub>Aggressive</sub> =  
310 0.61, 95% CI: 0.49-0.76, PP4: 0.96; OR<sub>Early Onset</sub> = 0.47, 95% CI: 0.34-0.64, PP4: 0.95, Figure 2].

311

312 We also reported proteins with a colocalized association for one outcome and little evidence  
313 for an association with others after correction for multiple testing, such as IFNA14, ISLR2,  
314 MMP7, and TNSFRS10B which were associated exclusively with overall prostate cancer  
315 [OR<sub>IFNA14</sub> = 0.74, 95% CI: 0.70-0.78; OR<sub>ISLR2</sub> = 0.82, 95% CI: 0.75-0.90; OR<sub>MMP7</sub> = 0.67, 95% CI:  
316 0.58-0.76; OR<sub>TNSFRS10B</sub> = 0.87, 95% CI: 0.83-0.92, Figure 2]. Similarly, PYY and SERPINA3  
317 associated with an increased risk of early onset prostate cancer only [OR<sub>PYY</sub> = 1.87, 95% CI:  
318 1.43-2.44; OR<sub>SERPINA3</sub> = 2.08, 95% CI: 1.51-2.88, Figure 2] while C2 associated with aggressive  
319 prostate cancer only [OR<sub>C2</sub> = 0.67, 95% CI: 0.56-0.81, Figure 2].

320

321 We additionally identified proteins with evidence for a directionally concordant colocalized  
322 association with some but not all prostate cancer outcomes, including TNFRSF6B that had an  
323 inverse association with all outcomes but only showed evidence in favor of colocalization for  
324 aggressive disease [OR<sub>Overall</sub> = 0.53, 95% CI: 0.46-0.61, PP4: 0.00; OR<sub>Aggressive</sub> = 0.48, 95% CI:  
325 0.037-0.63, PP4: 0.90; OR<sub>Early Onset</sub> = 0.43, 95% CI: 0.29-0.63, PP4: 0.14, Figure 2,  
326 Supplementary Table 1]. Likewise PPA2 was associated with an increased risk of both  
327 aggressive and early onset disease but lacked support from colocalization analyses for

328 prostate cancer risk overall [OR<sub>Overall</sub> = 1.84, 95% CI: 1.52-2.22, PP4: 0.01; OR<sub>Aggressive</sub> = 2.13,  
329 95% CI: 1.54-2.93, PP4: 0.99; OR<sub>Early Onset</sub> = 2.70, 95% CI: 1.71-4.27, PP4: 0.93, Figure 2,  
330 Supplementary Table 1].

331

332

333

334 *Replication of robust proteins in European and African ancestry populations*

335

336 We replicated the association for all 14 proteins that were robustly associated with overall  
337 prostate cancer (5NTC, CREBL1, CREB3L4, EHBP1, INFA14, ISLR2, MMP7, MSMB, PRSS3, PLG,  
338 POGLUT3, SERPINA1, TNSFRF10B, TPM3) using an independent meta-analysis of European  
339 ancestry participants in the UK Biobank and FinnGen cohorts (Table 1). Among these, the most  
340 statistically significant association was for MSMB [OR<sub>Overall</sub> = 0.83, 95% CI: 0.81 to 0.85, Table  
341 1] and the largest effect size was for PLG [OR<sub>Overall</sub> = 0.46, 95% CI: 0.07 to 0.84, Table 1]. We  
342 additionally identified African ancestry-specific *cis*-pQTL for six of the 14 proteins (ISLR2,  
343 MMP7, MSMB, POGLUT3, PRSS3, SERPINA1; Table 1). Of these, three proteins associations  
344 with risk of prostate cancer overall were replicated in men of African ancestry: MSMB  
345 [OR<sub>Overall</sub> = 0.85, 95% CI: 0.80 to 0.91], MMP7 [OR<sub>Overall</sub> = 0.83, 95% CI: 0.73 to 0.95], and ISLR2  
346 [OR<sub>Overall</sub> = 0.85, 95% CI: 0.73 to 0.99] (Figure 3, Table 1).

347

348 *Drug target analysis*

349

350 Out of the 2,002 unique proteins investigated, we identified 525 proteins that could be  
351 successfully mapped at the gene level to the target of a therapeutic intervention. Of these,  
352 ten (TPM3, PRSS3, PLG, MMP7, SERPINA1, SERPINA3, TNFRSF10B, C4A, HDGF, and LAYN)  
353 were associated with risk of at least one prostate cancer outcome after correction for  
354 investigating 525 drug target proteins and had evidence of colocalization (Supplementary  
355 Table 1). For example, PLG and C4A mapped to the clot dissolving class of fibrinolytics, TPM3  
356 mapped to phenethyl isothiocyanate, and MMP7 mapped to matrix metalloproteinase  
357 inhibitor, marimastat.

358

359 *Spatial transcriptomic analysis*

360

361 Of the 20 proteins that were associated with at least once prostate cancer outcome, we  
362 performed a targeted follow-up analysis for the two proteins with known high expression in  
363 the prostate tumor epithelium, MSMB and CREB3L4, using organ-wide spatial transcriptomic  
364 data on tissue obtained by radical prostatectomy from a patient with multifocal prostate  
365 cancer.<sup>34</sup> In analyzing epithelial-rich spots, we observed marked differences of MSMB  
366 expression between benign cells, where MSMB was highly abundant, and Gleason grade  
367 group 4 (GG4) cells, where MSMB was very low or absent in a majority of cells (log MSMB<sub>benign</sub>  
368 [median, interquartile range]: 2.73 [2.00-3.12] vs. log MSMB<sub>GG4</sub>: 0.48 [0.00-1.20], Figure 4.).  
369 A similarly, albeit more modestly, lower MSMB expression was observed in GG1 and GG2 cells  
370 (log MSMB<sub>GG1</sub>: 1.62 [0.97-1.88] and log MSMB<sub>GG2</sub>: 0.95 [0.69-1.32] compared to benign cells  
371 (Figure 4.). A lower expression of CREB3L4 was noted in GG2 and GG4 cells compared to  
372 benign cells (log CREB3L4<sub>benign</sub>: 0.60 [0.30-0.85] vs. log CREB3L4<sub>GG2</sub>: 0.31 [0.00-0.48] vs. log  
373 CREB3L4<sub>GG4</sub>: 0.48 [0.00-0.85], Supplementary Figure 2). Additional genome-wide random



374 forest analyses identified MSMB expression as the most important gene in terms of  
375 distinguishing between benign and GG4 cells (Figure 5).

376

## 377 **Discussion**

378

379 In this analysis, we conducted the largest study to date investigating the genetic associations  
380 of up to 2,002 unique proteins with the likely etiological risk of overall, early onset, and  
381 aggressive prostate cancer in up to 177,526 men using a MR and colocalization pipeline. In  
382 total, we found evidence supporting associations between 20 proteins and prostate cancer  
383 risk: 14 proteins for overall prostate cancer risk, seven for aggressive prostate cancer and  
384 eight for early onset prostate cancer. Among those 14 proteins that associated with prostate  
385 cancer risk overall, 14 were replicated in an external European ancestry population and three  
386 (out of six with available data) were replicated in an African ancestry population. A further  
387 half of the 20 proteins identified were also found to be the site of action for established drug  
388 targets with potential therapeutic implications. Finally, using spatial transcriptomics, we  
389 demonstrated a central role for the gene of our most robustly associated prostate cancer  
390 protein, MSMB, in distinguishing benign from undifferentiated, high-grade prostate cancer  
391 cells.

392

### 393 *MSMB*

394

395 MSMB is a secretory protein and member of the human immunoglobulin family that is  
396 released largely by luminal epithelial cells in the prostate epithelium and has a documented  
397 role in overall prostate cancer risk for both observational and genetic epidemiology.<sup>6,7,36</sup> In  
398 this study, we expand upon these previous findings by demonstrating that a 20% lower risk  
399 of overall prostate cancer is associated with higher levels of genetically predicted MSMB in  
400 two independent European ancestry cancer GWAS, and confirm for the first time its  
401 protective role in both aggressive and early onset disease etiology. We further successfully  
402 replicated this association with prostate cancer risk overall in an African-ancestry population,  
403 an ancestry group with an established higher risk of developing the disease. Subsequently, in  
404 reporting that MSMB gene expression is significantly depleted among high grade tumor when  
405 compared to expression on benign cells, we reiterate, through an independent line of  
406 evidence, that this gene may be particularly relevant to tumorigenesis and risk for aggressive  
407 disease.

408

409 Although the mechanism of action for MSMB in prostate cancer is not clear, MSMB has been  
410 shown to have a regulatory effect on cell growth, which may be lost during tumorigenesis  
411 while a MSMB-derived polypeptide was shown to induce prostate cell death.<sup>36,37</sup> Additionally,  
412 in a rodent model and in vitro, higher MSMB activity was found to suppress prostate tumor  
413 growth while a knockout of MSMB promoter/enhancer regions was characterized by tumor  
414 progression and metastases.<sup>38,39</sup> Given the integration of several compelling lines of evidence  
415 presented in this paper with existing literature, further research is warranted to understand  
416 the precise functional role of MSMB in prostate cancer tumorigenesis, identify environmental  
417 and lifestyle determinants, and explore potential clinical utility.

418

419

420

421 *Transcription Factors CREBL34 and CREBL1*

422

423 We identified novel proteins associated with overall, early onset, and aggressive prostate  
424 cancer etiology. These include the endoplasmic reticulum (ER) originating transcription  
425 factors CREB3L4 and CREBL1 (also known as ATF6B), which we find are associated with an  
426 increased risk of prostate cancer overall and that are expressed in the prostate epithelium.<sup>33</sup>  
427 These transcription factors form part of a transcriptional factor network that regulates the  
428 function of the endoplasmic reticulum (ER) and the activity of the unfolded protein response  
429 (UPR). There is an established role for the UPR and heat-shock proteins in maintaining AR  
430 stability and supporting AR-dependent tumorigenesis.<sup>40,41</sup> CREB3L4 has been shown to  
431 directly interact with AR in LNCaP cells to increase cellular proliferation and is abundantly  
432 expressed in prostate tumor tissue.<sup>42,43</sup> Furthermore, evidence suggests that disruptions in  
433 CREB3L4 contribute to ER stress downstream initiation of the unfolded protein response.<sup>43,44</sup>  
434 In a previous study of differential gene expression in prostate tissue, CREB3L4 was identified  
435 as a member of a co-expression gene cluster enriched for a previously described metabolic  
436 pathway (hsa05215) in prostate cancer that may regulate apoptosis and cell proliferation.<sup>45</sup>  
437 Interestingly, we found that CREB3L4 expression is lower in GG2 and GG4 cells as opposed to  
438 benign cells.

439

440 Previous studies have linked AR activation with members of the ATF6 family in LNCaP and PC3  
441 cells, however these experiments have mostly focused on ATF6A.<sup>46,47</sup> For example, a recent  
442 *in vivo* study showed that prostate cancer cells with ATF6A overexpression resisted cellular  
443 death by ferroptosis.<sup>46</sup> In parallel, a previous MR study reported a lowered risk of prostate  
444 cancer overall with genetically elevated circulating ATF6A levels from *trans*-pQTL.<sup>48</sup> Given the  
445 promising role of its paralog, and the increased risk we report here, targeted follow up of  
446 CREBL1 may prove valuable in characterizing the broader role of ER stress proteins and AR-  
447 dependent tumorigenesis.

448

449 *EHBP1*

450

451 We observed a more than two-fold increased risk of early-onset prostate cancer associated  
452 with higher EHBP1, an adaptor protein with a key role in vesicular trafficking and actin  
453 reorganization.<sup>49</sup> Variants in the EHBP1 intron have previously been associated with  
454 aggressive prostate cancer in a genetic association study the protein and is more highly  
455 expressed in prostate tumor tissue and may have a role in determining the invasiveness of  
456 PTEN-positive prostate cancer cells according to GWAS and expression data and in a cellular  
457 study.<sup>50-52</sup> While mechanisms that may link EHBP1 to prostate cancer risk are not yet fully  
458 described, it has role as an effector molecule for Rab8 family members that modulate  
459 polarized membrane transport via actin reorganization and may have a role in the mechanism  
460 of action for atorvastatin.<sup>50</sup>

461

462 *Other early onset and aggressive disease proteins*

463

464 We found several proteins that were associated with early onset and aggressive disease,  
465 including PYY, PRSS3, PPA2, C2, C4A, and SERPINA3. For example, PYY is a metabolic hormone  
466 involved in appetite regulation, and while there has been some hypothesized relationship  
467 between obesity and aggressive prostate cancer risk in the past, recent findings suggest that

468 obesity does not serve as a risk factor for disease itself, but may affect likelihood of  
469 diagnosis.<sup>53,54</sup> Additionally, one study found that mesotrypsin, a protease encoded by the  
470 PRSS3 gene, was essential for prostate cancer metastasis in vitro and mouse models, however  
471 the role of this protein has not been widely studied in humans.<sup>55</sup>

472

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

481

#### 482 *Drug target proteins*

483

484 We identified 10 proteins that were associated with both the risk of prostate cancer and that  
485 were the site of action for a known drug. These included TNFRSF10B, which is a receptor for  
486 the cytotoxic TRAIL ligand, and is essential for CASP8 and ER stress induced apoptosis.<sup>56</sup>  
487 Further, TNFRSF10B expression is lower in higher grade prostate tumors and a recent study  
488 of PARP inhibitors in prostate cancer cell lines suggested that TNFRSF10B may provide a  
489 mechanism by which the cancer drug Olaparib induces apoptosis.<sup>57</sup> The apparent protective  
490 association we observe with prostate cancer risk is in line with the results from multiple phase  
491 I/II trials of TNFRSF10B agonists that support their use for the treatment of multiple cancer  
492 endpoints, though not yet including prostate cancer.<sup>57-59</sup>

493

494 We also identified an inverse association of PLG, a serine protease targeted by tranexamic  
495 acids and several classes of thrombolytics, with prostate cancer risk overall and with early  
496 onset disease. Tranexamic acids are primarily prescribed to control excessive bleeding while  
497 thrombolytics are primarily used to dissolve blood clots and act via plasmin and fibrin  
498 pathways.<sup>60</sup> One molecular study found that PLG is generated by the cancer-mediated  
499 proteolysis of plasminogen which is released by human prostate carcinoma cells.<sup>61</sup> PLG in  
500 turn has been shown in many lab studies to inhibit angiogenesis which when unregulated can  
501 lead to the rapid formation of tumors.<sup>62,63</sup> Currently, several studies investigated combination  
502 therapy including plasminogen activation or inhibition for treatment of several cancer types,  
503 though not specifically for prostate cancer, in phase I/II trials.<sup>64,65</sup>

504

505 SERPINA1 maps to fazisiran, the treatment for alpha-1 antitrypsin deficiency and that is in  
506 phase II/III of drug trials. Previous findings have indicated that alpha-1 antitrypsin levels are  
507 often elevated in many carcinomas, including prostate.<sup>66</sup> However, no agents targeting  
508 SERPINA1 have been investigated in cancer trials thus far. MMP7 belongs to a class of matrix  
509 metalloproteinases that participate in wound healing, bone growth, and matrix remodeling.  
510 There are multiple lines of evidence that this protein is involved in many cancers, and agents  
511 targeting metalloproteinases, such as marimastat, are currently being investigated in clinical  
512 trials at various phases.<sup>67</sup> In prostate cancer, marimastat showed some efficacy in early trials,  
513 however has not yet progressed further.<sup>68</sup>

514

515 While we highlight proteins that may share the same target site for established drug targets  
516 that may have implications for therapeutic use, the suitability of these to act as preventative  
517 or remedial agents requires careful considerations including the site specificity, potential  
518 downstream effects, routes of administration and effectively capturing the population at  
519 risk.<sup>69</sup>

520

### 521 *Strengths and weaknesses*

522

523 This study offers several strengths including being the largest currently available GWAS of  
524 prostate cancer outcomes and the use of both aggressive and early onset endpoints with *cis*-  
525 pQTL covering up to 2,002 proteins. One previous MR study investigated the role of the  
526 circulating proteome in prostate cancer risk but did not stratify analyses by *cis* or *trans*- pQTL,  
527 and did not perform colocalization analyses, making it more challenging to infer causal  
528 relationships between individual proteins and cancer risk.<sup>48</sup> Additionally, by integrating gene  
529 expression data measured using spatial transcriptomics, the current paper introduces a novel  
530 translational approach to highlight biological enablers of prostate cancer. To our knowledge,  
531 this study provides the first demonstration that MR using *cis* instruments of plasma protein  
532 levels can be used to identify a risk protein that has both specific expression in the cell of  
533 cancer origin and is related to tumor aggressiveness – important features to consider when  
534 identifying candidate targets for therapeutic prevention.

535

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

549

### 550 **Conclusion**

551

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

560

561

## 562 References

563

- 564 1. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA: A Cancer Journal*  
565 *for Clinicians* **68**, 7–30 (2018).
- 566 2. Salinas, C. A., Tsodikov, A., Ishak-Howard, M. & Cooney, K. A. Prostate Cancer in  
567 Young Men: An Important Clinical Entity. *Nat Rev Urol* **11**, 317–323 (2014).
- 568 3. Shih, H.-J., Fang, S.-C., An, L. & Shao, Y.-H. J. Early-onset prostate cancer is  
569 associated with increased risks of disease progression and cancer-specific mortality. *Prostate*  
570 **81**, 118–126 (2021).
- 571 4. Bergengren, O. *et al.* 2022 Update on Prostate Cancer Epidemiology and Risk  
572 Factors—A Systematic Review. *European Urology* **84**, 191–206 (2023).
- 573 5. Watts, E. L. *et al.* Circulating insulin-like growth factors and risks of overall,  
574 aggressive and early-onset prostate cancer: a collaborative analysis of 20 prospective studies  
575 and Mendelian randomization analysis. *International Journal of Epidemiology* **52**, 71–86  
576 (2023).
- 577 6. Smith Byrne, K. *et al.* The role of plasma microseminoprotein-beta in prostate cancer:  
578 an observational nested case-control and Mendelian randomization study in the European  
579 prospective investigation into cancer and nutrition. *Ann Oncol* **30**, 983–989 (2019).
- 580 7. Haiman, C. A. *et al.* Levels of Beta-Microseminoprotein in Blood and Risk of  
581 Prostate Cancer in Multiple Populations. *J Natl Cancer Inst* **105**, 237–243 (2013).
- 582 8. Swerdlow, D. I. *et al.* Selecting instruments for Mendelian randomization in the wake  
583 of genome-wide association studies. *Int J Epidemiol* **45**, 1600–1616 (2016).
- 584 9. Burgess, S., Foley, C. N. & Zuber, V. Inferring Causal Relationships Between Risk  
585 Factors and Outcomes from Genome-Wide Association Study Data. *Annual Review of*  
586 *Genomics and Human Genetics* **19**, 303–327 (2018).
- 587 10. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic  
588 Association Studies Using Summary Statistics. *PLoS Genet* **10**, e1004383 (2014).
- 589 11. Lawlor, D. A., Tilling, K. & Davey Smith, G. Triangulation in aetiological  
590 epidemiology. *Int. J. Epidemiol.* dyw314 (2017) doi:10.1093/ije/dyw314.
- 591 12. Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R. & Haque, A. An  
592 introduction to spatial transcriptomics for biomedical research. *Genome Medicine* **14**, 68  
593 (2022).
- 594 13. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics  
595 and disease. *Nat Genet* **53**, 1712–1721 (2021).
- 596 14. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the  
597 plasma proteome on complex diseases. *Nat Genet* **52**, 1122–1131 (2020).
- 598 15. Folkersen, L. *et al.* Genomic and drug target evaluation of 90 cardiovascular proteins  
599 in 30,931 individuals. *Nat Metab* **2**, 1135–1148 (2020).
- 600 16. Sun, X. *et al.* Genetically predicted levels of circulating cytokines and prostate cancer  
601 risk: A Mendelian randomization study. *Int. J. Cancer* **147**, 2469–2478 (2020).
- 602 17. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across  
603 the human phenome. *eLife* **7**, e34408 (2018).
- 604 18. Fauman, E. B. & Hyde, C. An optimal variant to gene distance window derived from  
605 an empirical definition of cis and trans protein QTLs. *BMC Bioinformatics* **23**, 169 (2022).
- 606 19. PRACTICAL. <http://practical.icr.ac.uk/>.
- 607 20. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63  
608 new prostate cancer susceptibility loci. *Nat Genet* **50**, 928–936 (2018).
- 609 21. Conti, D. V. *et al.* Trans-ancestry genome-wide association meta-analysis of prostate  
610 cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet* **53**,  
611 65–75 (2021).

- 612 22. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. M. A robust and  
613 efficient method for Mendelian randomization with hundreds of genetic variants. *Nat*  
614 *Commun* **11**, 376 (2020).
- 615 23. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple  
616 causal variants. *PLOS Genetics* **17**, e1009440 (2021).
- 617 24. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in  
618 colocalisation analyses. *PLoS Genet* **16**, e1008720 (2020).
- 619 25. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated  
620 population. *Nature* **613**, 508–518 (2023).
- 621 26. Hewitt, J., Walters, M., Padmanabhan, S. & Dawson, J. Cohort profile of the UK  
622 Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open* **6**, e009161  
623 (2016).
- 624 27. Zhang, J. *et al.* Plasma proteome analyses in individuals of European and African  
625 ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet* **54**,  
626 593–602 (2022).
- 627 28. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for  
628 2018. *Nucleic Acids Res* **46**, D1074–D1082 (2018).
- 629 29. Sheils, T. K. *et al.* TCRD and Pharos 2021: mining the human proteome for disease  
630 biology. *Nucleic Acids Res* **49**, D1334–D1346 (2021).
- 631 30. Zhou, Y. *et al.* Therapeutic target database update 2022: facilitating drug discovery  
632 with enriched comparative data of targeted agents. *Nucleic Acids Research* **50**, D1398–  
633 D1407 (2022).
- 634 31. Erickson, A. *et al.* A Systematic Review of Prostate Cancer Heterogeneity:  
635 Understanding the Clonal Ancestry of Multifocal Disease. *European Urology Oncology* **4**,  
636 358–369 (2021).
- 637 32. Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially  
638 resolved transcriptomics reveals the architecture of the tumor-microenvironment interface.  
639 *Nat Commun* **12**, 6278 (2021).
- 640 33. THE TABULA SAPIENS CONSORTIUM. The Tabula Sapiens: A multiple-organ,  
641 single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
- 642 34. Erickson, A. *et al.* Spatially resolved clonal copy number alterations in benign and  
643 malignant tissue. *Nature* **608**, 360–367 (2022).
- 644 35. Basu, S., Kumbier, K., Brown, J. B. & Yu, B. Iterative random forests to discover  
645 predictive and stable high-order interactions. *Proceedings of the National Academy of*  
646 *Sciences* **115**, 1943–1948 (2018).
- 647 36. Whitaker, H. C., Warren, A. Y., Eeles, R., Kote-Jarai, Z. & Neal, D. E. The potential  
648 value of microseminoprotein-beta as a prostate cancer biomarker and therapeutic target.  
649 *Prostate* **70**, 333–340 (2010).
- 650 37. Annabi, B. *et al.* Contribution of the 37-kDa laminin receptor precursor in the anti-  
651 metastatic PSP94-derived peptide PCK3145 cell surface binding. *Biochem Biophys Res*  
652 *Commun* **346**, 358–366 (2006).
- 653 38. Annabi, B. *et al.* A PSP94-derived peptide PCK3145 inhibits MMP-9 secretion and  
654 triggers CD44 cell surface shedding: implication in tumor metastasis. *Clin Exp Metastasis* **22**,  
655 429–439 (2005).
- 656 39. Lamy, S. *et al.* A prostate secretory protein94-derived synthetic peptide PCK3145  
657 inhibits VEGF signalling in endothelial cells: implication in tumor angiogenesis. *Int J Cancer*  
658 **118**, 2350–2358 (2006).
- 659 40. Jin, Y. & Saatcioglu, F. Targeting the Unfolded Protein Response in Hormone-  
660 Regulated Cancers. *Trends in Cancer* **6**, 160–171 (2020).

- 661 41. Centenera, M. M., Fitzpatrick, A. K., Tilley, W. D. & Butler, L. M. Hsp90: Still a  
662 viable target in prostate cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*  
663 **1835**, 211–218 (2013).
- 664 42. Labrie, C. *et al.* Androgen-regulated transcription factor AIBZIP in prostate cancer.  
665 *The Journal of Steroid Biochemistry and Molecular Biology* **108**, 237–244 (2008).
- 666 43. Kim, T.-H., Park, J.-M., Kim, M.-Y. & Ahn, Y.-H. The role of CREB3L4 in the  
667 proliferation of prostate cancer cells. *Sci Rep* **7**, 45300 (2017).
- 668 44. Pu, Q. *et al.* The Novel Transcription Factor CREB3L4 Contributes to the Progression  
669 of Human Breast Carcinoma. *J Mammary Gland Biol Neoplasia* **25**, 37–50 (2020).
- 670 45. Pidò, S., Ceddia, G. & Masseroli, M. Computational analysis of fused co-expression  
671 networks for the identification of candidate cancer gene biomarkers. *NPJ Syst Biol Appl* **7**, 17  
672 (2021).
- 673 46. Zhao, R. *et al.* ATF6 $\alpha$  promotes prostate cancer progression by enhancing PLA2G4A-  
674 mediated arachidonic acid metabolism and protecting tumor cells against ferroptosis.  
675 *Prostate* **82**, 617–629 (2022).
- 676 47. Zhou, H. *et al.* The functional implication of ATF6 $\alpha$  in castration-resistant prostate  
677 cancer cells. *The FASEB Journal* **37**, e22758 (2023).
- 678 48. Wu, L. *et al.* Analysis of over 140,000 European descendants identifies genetically-  
679 predicted blood protein biomarkers associated with prostate cancer risk. *Cancer Res* **79**,  
680 4592–4598 (2019).
- 681 49. Rai, A., Bleimling, N., Vetter, I. R. & Goody, R. S. The mechanism of activation of  
682 the actin binding protein EHBP1 by Rab8 family members. *Nat Commun* **11**, 4187 (2020).
- 683 50. Ghalali, A., Wiklund, F., Zheng, H., Stenius, U. & Högberg, J. Atorvastatin prevents  
684 ATP-driven invasiveness via P2X7 and EHBP1 signaling in PTEN-expressing prostate  
685 cancer cells. *Carcinogenesis* **35**, 1547–1555 (2014).
- 686 51. Gudmundsson, J. *et al.* Common sequence variants on 2p15 and Xp11.22 confer  
687 susceptibility to prostate cancer. *Nat Genet* **40**, 281–283 (2008).
- 688 52. Mamidi, T. K. K., Wu, J. & Hicks, C. Integrating germline and somatic variation  
689 information using genomic data for the discovery of biomarkers in prostate cancer. *BMC*  
690 *Cancer* **19**, 229 (2019).
- 691 53. Hurwitz, L. M., Dogbe, N., Barry, K. H., Koutros, S. & Berndt, S. I. Obesity and  
692 prostate cancer screening, incidence, and mortality in the Prostate, Lung, Colorectal, and  
693 Ovarian Cancer Screening Trial. *JNCI: Journal of the National Cancer Institute* djad113  
694 (2023) doi:10.1093/jnci/djad113.
- 695 54. Perez-Cornago, A., Dunneram, Y., Watts, E. L., Key, T. J. & Travis, R. C. *Adiposity*  
696 *and risk of prostate cancer death: a prospective analysis in UK Biobank and meta-analysis of*  
697 *published studies*. 2021.10.05.21264556  
698 <https://www.medrxiv.org/content/10.1101/2021.10.05.21264556v1> (2021)  
699 doi:10.1101/2021.10.05.21264556.
- 700 55. Hockla, A. *et al.* PRSS3/Mesotrypsin is a therapeutic target for metastatic prostate  
701 cancer. *Molecular cancer research : MCR* **10**, 1555 (2012).
- 702 56. Hagenlocher, C. *et al.* ER stress-induced cell death proceeds independently of the  
703 TRAIL-R2 signaling axis in pancreatic  $\beta$  cells. *Cell Death Discov.* **8**, 1–9 (2022).
- 704 57. Hernandez-Cueto, A. *et al.* Death receptor 5 expression is inversely correlated with  
705 prostate cancer progression. *Molecular Medicine Reports* **10**, 2279–2286 (2014).
- 706 58. Subbiah, V. *et al.* Preclinical Characterization and Phase I Trial Results of INBRX-  
707 109, A Third-Generation, Recombinant, Humanized, Death Receptor 5 Agonist Antibody, in  
708 Chondrosarcoma. *Clin Cancer Res* **29**, 2988–3003 (2023).
- 709 59. Forero-Torres, A. *et al.* TBCRC 019: An open label, randomized, phase II trial of  
710 nanoparticle albumin-bound paclitaxel (nab-PAC or Abraxane®) with or without the anti-

- 711 death receptor 5 (DR5) monoclonal antibody tigatuzumab in patients with metastatic triple  
712 negative breast cancer. *Clin Cancer Res* **21**, 2722–2729 (2015).
- 713 60. Reed, M. R. & Woolley, L. T. Uses of tranexamic acid. *Continuing Education in*  
714 *Anaesthesia Critical Care & Pain* **15**, 32–37 (2015).
- 715 61. Didiasova, M., Wujak, L., Wygrecka, M. & Zakrzewicz, D. From Plasminogen to  
716 Plasmin: Role of Plasminogen Receptors in Human Cancer. *Int J Mol Sci* **15**, 21229–21252  
717 (2014).
- 718 62. Capello, M., Ferri-Borgogno, S., Cappello, P. & Novelli, F.  $\alpha$ -Enolase: a promising  
719 therapeutic and diagnostic tumor target. *FEBS J* **278**, 1064–1074 (2011).
- 720 63. Kwaan, H. C. & McMahon, B. The role of plasminogen-plasmin system in cancer.  
721 *Cancer Treat Res* **148**, 43–66 (2009).
- 722 64. University of Aarhus. *Perioperative Treatment With Tranexamic Acid in Melanoma;*  
723 *Prognostic and Treatment-related Impact of the Plasminogen-plasmin Pathway.*  
724 <https://clinicaltrials.gov/study/NCT05899465> (2023).
- 725 65. Northwestern University. *Phase I/II Trial of In Vivo Angiostatin Generation With*  
726 *Tissue Plasminogen Activator (tPA) and Captopril in Patients With Progressive, Metastatic*  
727 *Cancer.* <https://clinicaltrials.gov/study/NCT00086723> (2012).
- 728 66. El-Akawi, Z. J., Abu-awad, A. M. & Khouri, N. A. Alpha-1 Antitrypsin Blood Levels  
729 as Indicator for the Efficacy of Cancer Treatment. *World J Oncol* **4**, 83–86 (2013).
- 730 67. Liao, H.-Y., Da, C.-M., Liao, B. & Zhang, H.-H. Roles of matrix metalloproteinase-7  
731 (MMP-7) in cancer. *Clinical Biochemistry* **92**, 9–18 (2021).
- 732 68. Rosenbaum, E. *et al.* Marimastat in the treatment of patients with biochemically  
733 relapsed prostate cancer: a prospective randomized, double-blind, phase I/II trial. *Clin Cancer*  
734 *Res* **11**, 4437–4443 (2005).
- 735 69. Bull, S. C. & Doig, A. J. Properties of Protein Drug Target Classes. *PLoS One* **10**,  
736 e0117955 (2015).
- 737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759



760 **Table and Figure Legends**

761

762 **Table 1.** Mendelian randomization and colocalization results for protein-cancer associations  
763 that passed Bonferroni correction ( $0.05/n$  proteins analyzed) and colocalized at  $PP4 > 0.70$ .  
764 Results are shown per SNP-cancer association, except for those proteins for which there  
765 were multiple SNPs that passed both multiple testing correction and colocalization, and for  
766 which the summary estimate using the Inverse Variance Weighted (IVW) method is  
767 provided. Odds ratios are oriented per standard deviation increase in genetically predicted  
768 protein level. Maximum  $PP4$  is reported as highest  $PP4$  value from either single or  
769 conditional iterative colocalization method. Odds ratios are reported for associations using  
770 external UK Biobank/Finn Gen and for African ancestry population where data existed. Drug  
771 targets and drug trials are annotated if they existed.

772

773 **Figure 1.** Association of genetically predicted protein concentrations with prostate cancer  
774 risk presented as a Manhattan plot where position is given by *cis*-pQTL coordinate  
775 (chromosome and base-pair position) labelled with their association with cancer risk and the  
776 highest colocalization probability from single or conditional iterative methods ( $PP4$ ). Points  
777 highlighted as filled-in are those with evidence of a shared causal locus ( $PP4 > 0.7$ ) with  
778 point size reflecting  $PP4$  magnitude, which can vary between 0 and 1. Risk associations with  
779 MR  $p >$  Bonferroni correction threshold were not subject to colocalization analyses. The  
780 strongest protein-cancer association per chromosome is labelled and a zoomed-in plot for  
781 MSMB (rs10993994) on chromosome 10 is shown in the upper right-hand corner.

782

783 **Figure 2.** Odds ratios (95% confidence intervals) for genetically predicted protein levels and  
784 prostate cancer risk (for proteins with  $p <$  Bonferroni threshold based on  $0.05/\text{number of}$   
785  $\text{proteins analyzed}$ ). Odds ratio estimates are scaled per standard deviation increment in  
786 genetically predicted relative circulating protein concentrations. Filled circles represent  
787 Bonferroni-significant associations and asterisks indicate evidence for colocalization ( $PP4 >$   
788  $0.70$ ).

789

790 **Figure 3.** Odds ratios (95% confidence intervals) for genetically predicted protein levels and  
791 overall prostate cancer risk for proteins with  $p <$  Bonferroni threshold based on  $0.05/\text{number}$   
792  $\text{of proteins analyzed}$  in main analyses, and with data available to perform replication in an  
793 African ancestry and European ancestry population. Odds ratio estimates are scaled per  
794 standard deviation increment in genetically predicted circulating protein concentrations.

**Figure 4. A)** MSMB association with overall, early onset, and aggressive prostate cancer risk with replication in the FinnGen and UK Biobank populations and in an African ancestry population. Odds ratio (95% confidence interval) estimates are scaled per standard deviation increment in genetically predicted circulating MSMB concentrations **B)** Spatial visualization showing MSMB gene expression (top) and histology and tissue status (bottom) from organ-wide spatial transcriptomic data in two tumor sections (GG: Gleason grade group: GG1, Gleason score of 6 or lower; GG2, Gleason score of 3+4 = 7; GG4, Gleason score of 8). **C)** Violin plots representing gene expression in each spatial transcriptomics spot according to histological status. Statistical differences are indicated: \*\*\*\*  $p < 0.0001$  (Kruskal–Wallis; post-test: Dunn’s test).

**Figure 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 colored to demonstrate its central role in the network.

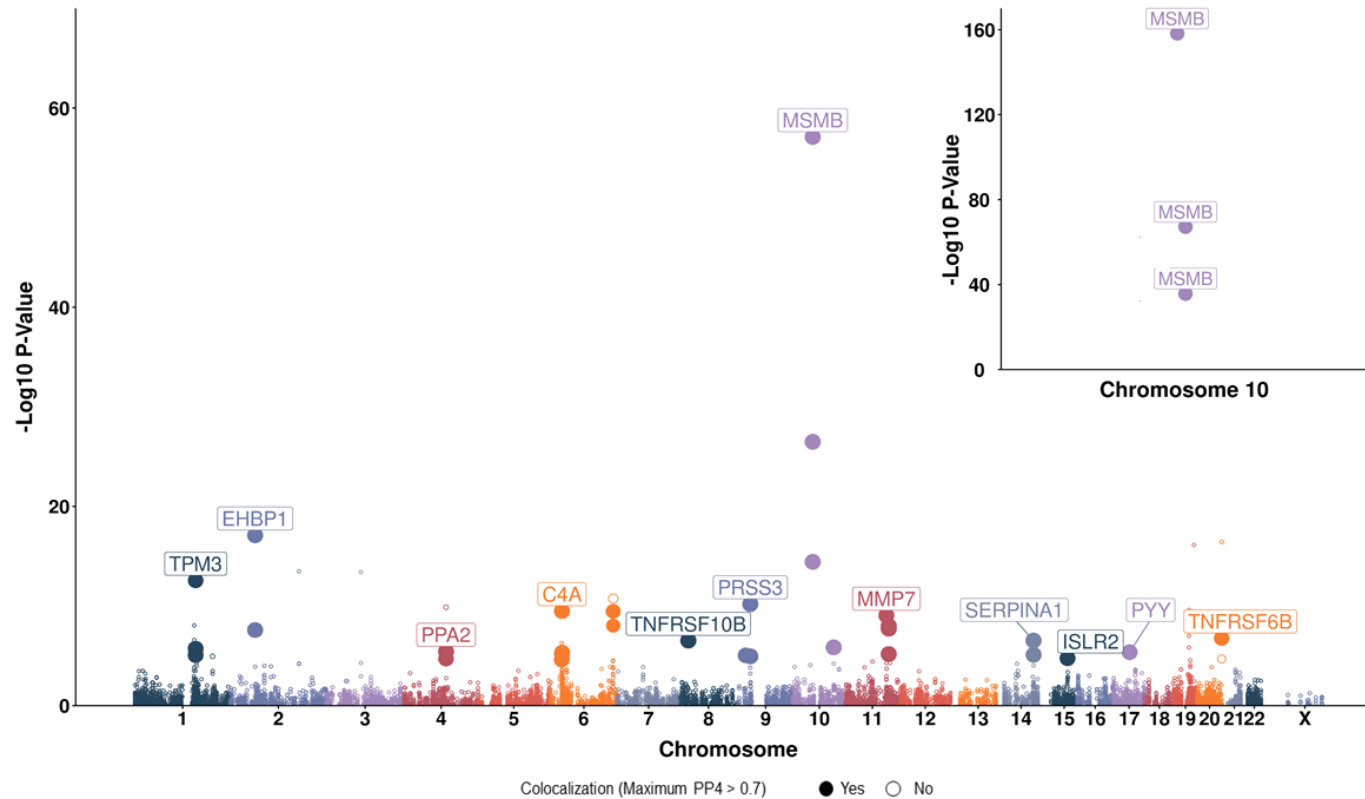
**Supplementary Figure 1. Flow-chart showing the overall study design.** PP4 = posterior probability of a shared causal locus. LD = Linkage disequilibrium.

**Supplementary Figure 2. A)** CREB3L4 association with overall, early onset, and aggressive prostate cancer risk with replication in the FinnGen and UK Biobank populations. Odds ratio (95% confidence interval) estimates are scaled per standard deviation increment in genetically predicted circulating CREB3L4 concentrations **B)** Spatial visualization showing CREB3L4 gene expression (top) and histology and tissue status (bottom) from organ-wide spatial transcriptomic data in two tumor sections (GG: Gleason grade group: GG1, Gleason score of 6 or lower; GG2, Gleason score of 3+4 = 7; GG4, Gleason score of 8). **C)** Violin plots representing gene expression in each spatial transcriptomics spot according to histological status. Statistical differences are indicated: \*\*\*\*  $p < 0.0001$  (Kruskal–Wallis; post-test: Dunn’s test).

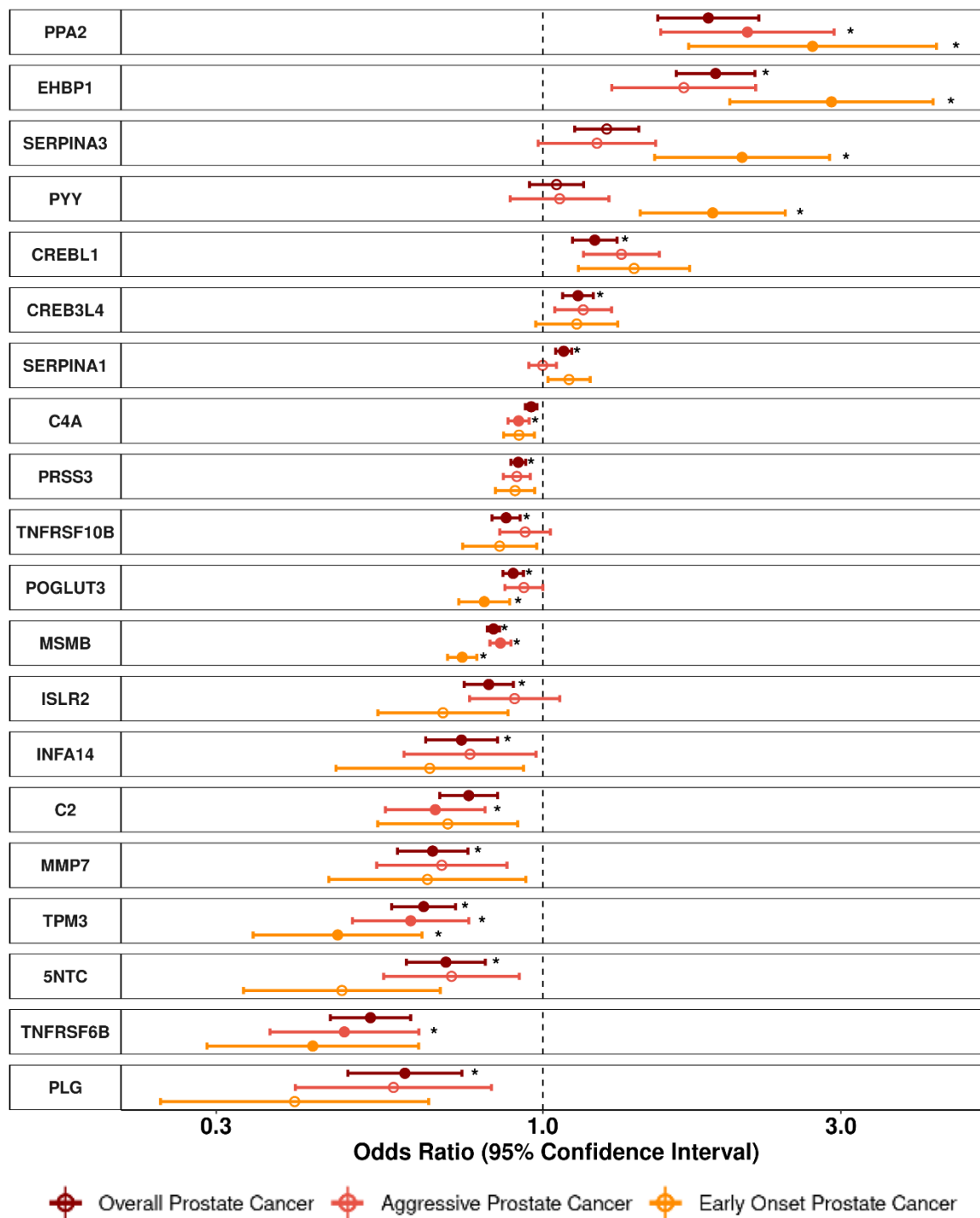
**Supplementary Table 1.** Full protein data characteristic data and Mendelian randomization and colocalization results for all proteins and cancer outcomes, where analyzed. Odds ratios (95% confidence intervals) are given per standard deviation increase in genetically predicted protein level, and maximum colocalization indicates the highest PP4 percentage from either single or conditional iterative colocalization methods.

**Table 1.** Mendelian randomization and colocalization results for protein-cancer associations that passed Bonferroni correction (0.05/n proteins analyzed) and colocalized at PP4 > 0.70. Results are shown per SNP-cancer association, except for those proteins for which there were multiple SNPs that passed both multiple testing correction and colocalization, 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 colocalization 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.

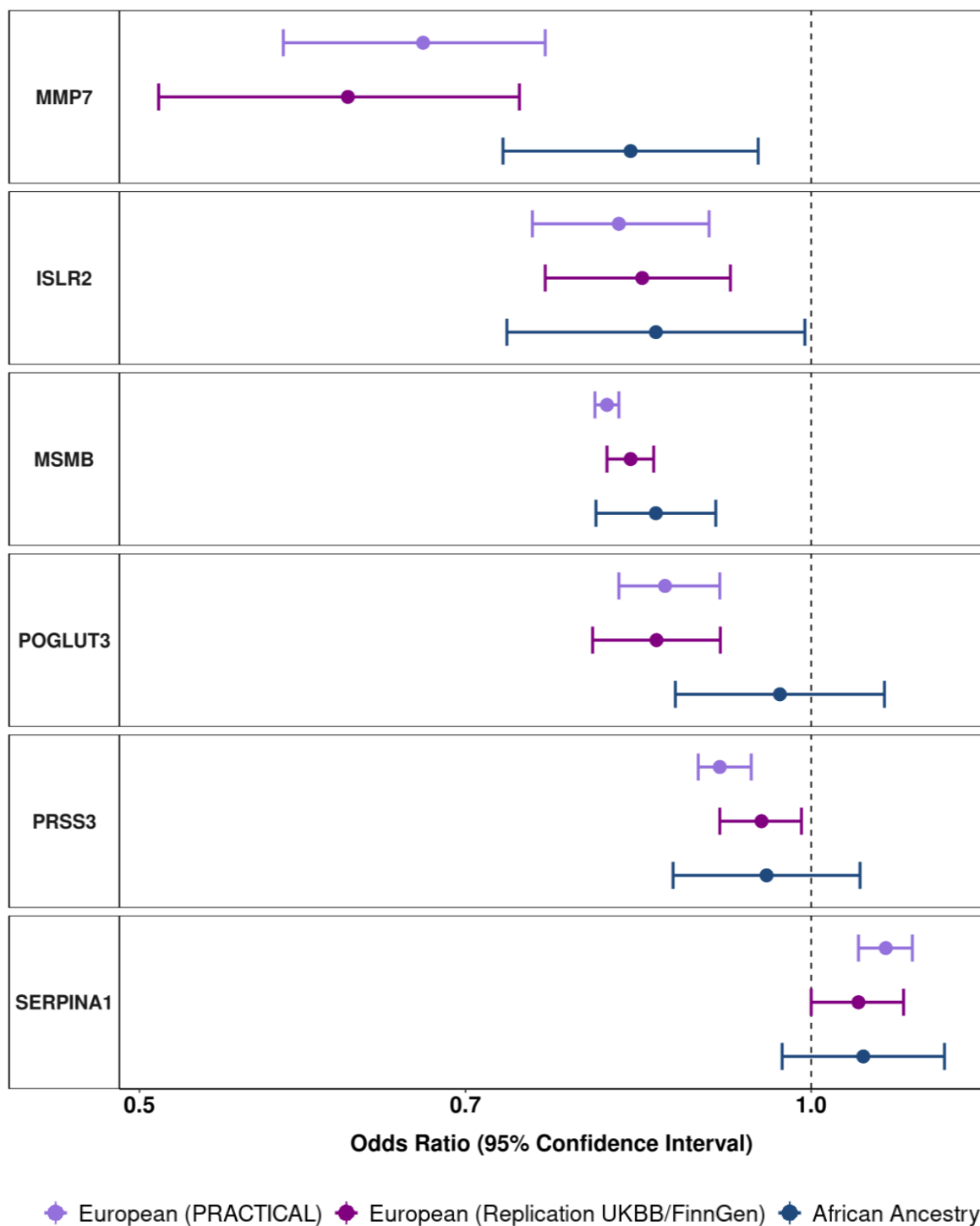
Prostate Cancer Outcome	Gene	SNP	Proportion Variance Explained (%)	Uniprot ID	Platform	Odds ratio (95% CI) PRACTICAL	P-value (Unadjusted)	Maximum PP4	Odds ratio (95% CI) UKBB/FinnGen	Odds ratio (95% CI) African Ancestry	Drug Target	Current Drug Trials
Overall	MSMB	rs10993994	49.60%	P08118	SomaScan	0.81 (0.8 -0.82)	1.32E-165	1.00	0.83 (0.81-0.85)	0.85 (0.80-0.91)		Phase I (Prostate Cancer therapy; PCK-3145)
	EHBP1	rs73934251	0.53%	Q8NDI1	SomaScan	1.89 (1.63 -2.19)	8.09E-18	0.98	2.09 (1.66-2.65)			
	TPM3	rs72696208	0.68%	P06753	SomaScan	0.64 (0.57 -0.72)	2.72E-13	0.94	0.64 (0.53-0.77)		Phenethyl Isothiocyanate	
	PRSS3	rs2005617	14.50%	P35030	SomaScan	0.91 (0.89 -0.94)	6.47E-11	1.00	0.95 (0.91-0.99)	0.95 (0.87-1.10)	4-(1,3,2-DIOXABOROLAN-2-YLO)	Biological testing
	PLG	rs982403	0.23%	P00747	SomaScan	0.48 (0.38 -0.6)	2.96E-10	1.00	0.46 (0.31-0.67)		Tranexamic Acid	11 Drugs Launched
		rs11751347	1.66%	P00747	SomaScan	0.8 (0.75 -0.86)	3.47E-10	0.83	0.84 (0.77-0.93)			
	IVW	rs4252185	0.23%	P00747	SomaScan	0.6 (0.49 -0.74)	2.07E-06	0.99	0.60 (0.44-0.81)			
		rs14983	2.11%	P00747	SomaScan	0.75 (0.62-0.92)	5.80E-03					
	MMP7	rs14983	1.56%	P09237	SomaScan	0.67 (0.58 -0.76)	8.84E-10	0.98	0.62 (0.51-0.74)	0.83 (0.73-0.95)	Marimastat	1 in Phase I/II (Colorectal Cancer therapy; IMA-910)
	POGLUT3	rs74911261	9.89%	Q7Z4H8	SomaScan	0.9 (0.86 -0.93)	9.99E-09	1.00	0.85 (0.80-0.91)	0.97 (0.87-1.10)		
	TNSFRS10B	rs2293400	4.30%	O14763	OLINK	0.87 (0.83 -0.92)	3.02E-07	1.00	0.73 (0.68-0.79)		HGS-TR2J; Lexatumumab	4 in Phase II (Therapy for multiple cancer types)
	SERPINA1	rs28929474	16.10%	P01009	SomaScan	1.08 (1.05 -1.11)	3.01E-07	1.00	1.10 (1.0-1.10)	1.06 (0.97-1.15)	Glassia	1 in Phase II/III (Alpha-1 antitrypsin deficiency; fazirsirar)
	SNTC	rs4919682	0.50%	P49902	SomaScan	0.7 (0.6 -0.81)	1.43E-06	1.00	0.64 (0.49-0.84)			1 in Phase I (Prostate Cancer therapy; FP-253)
	CREBL1	rs8111	1.31%	Q99941	SomaScan	1.21 (1.12 -1.31)	4.92E-06	0.99	1.23 (1.03-1.48)			
	CREB3L4	rs4845586	3.12%	Q8TEY5	SomaScan	1.14 (1.08 -1.2)	7.71E-06	0.73	1.23 (1.13-1.34)			
	INFA14	rs662463	0.56%	P01570	SomaScan	0.74 (0.65 -0.85)	9.04E-06	0.95	0.80 (0.64-0.99)			Biological Testing
	ISLR2	rs751527	1.36%	Q6UXK2	SomaScan	0.82 (0.75 -0.9)	1.78E-05	0.97	0.84 (0.73-0.97)	0.85 (0.73-0.99)		
Aggressive	MSMB	rs10993994	49.60%	P08118	SomaScan	0.84 (0.82-0.86)	1.55E-35	1.00				Phase I (Prostate Cancer therapy; PCK-3145)
	TNFRSF6B	rs6011040	0.50%	O95407	SomaScan	0.48 (0.37-0.63)	1.72E-07	0.90				
	PPA2	rs4699179	0.38%	Q9H2U2	SomaScan	0.76 (0.67-0.86)	3.73E-06	0.99				
	C4A	rs2763982	22.00%	P0C0L4	SomaScan	0.91 (0.88-0.95)	6.80E-06	0.99				Preclinical
	TPM3	rs72696208	0.68%	P06753	SomaScan	0.61 (0.5-0.76)	8.56E-06	0.96			Phenethyl Isothiocyanate	
	PRSS3	rs62555900	3.52%	P35030	SomaScan	0.8 (0.73-0.89)	1.13E-05	0.99			4-(1,3,2-DIOXABOROLAN-2-YLO)	Biological testing
	C2	rs3094662	0.75%	P06681	SomaScan	0.67 (0.56-0.81)	2.39E-05	0.96				Phase III (Cancer immunotherapy; 99mTc-ior C5)
Early Onset	MSMB	rs10993994	49.60%	P08118	SomaScan	0.71 (0.68-0.74)	1.46E-65	1.00				Phase I (Prostate Cancer therapy; PCK-3145)
	PLG	rs982403	3.57%	P00747	SomaScan	0.63 (0.53-0.73)	9.38E-09	0.72				11 Drugs Launched
		rs11751347	1.66%	P00747	SomaScan	0.61 (0.52-0.72)	9.38E-09	0.72				
	IVW	rs4252185	0.23%	P00747	SomaScan	0.26 (0.14-0.46)	4.07E-06	1.00				
		rs14983	2.11%	P00747	SomaScan	0.57 (0.36-0.91)	1.60E-02					
	EHBP1	rs73934251	0.53%	Q8NDI1	SomaScan	2.90 (2.0-4.2)	2.60E-08	0.98				
	TPM3	rs72696208	0.68%	P06753	SomaScan	0.47 (0.34-0.64)	1.92E-06	0.95			Phenethyl Isothiocyanate	1 Drug discontinued (Neurological Cancer therapy; anisi)
	PYY	rs8074783	0.00%	P10082	SomaScan	1.87 (1.43-2.44)	4.36E-06	0.98				Preclinical
	POGLUT3	rs74911261	3.46%	Q7Z4H8	SomaScan	0.81 (0.73-0.89)	6.67E-06	1.00				
	SERPINA3	rs8023057	7.95%	P01011	SomaScan	2.08 (1.51-2.88)	8.04E-06	0.99			Zinc; Acetate, Chloride, Sulfate	Preclinical
	PPA2	rs4699179	0.38%	Q9H2U2	SomaScan	2.7 (1.71-4.27)	1.99E-05	0.93				



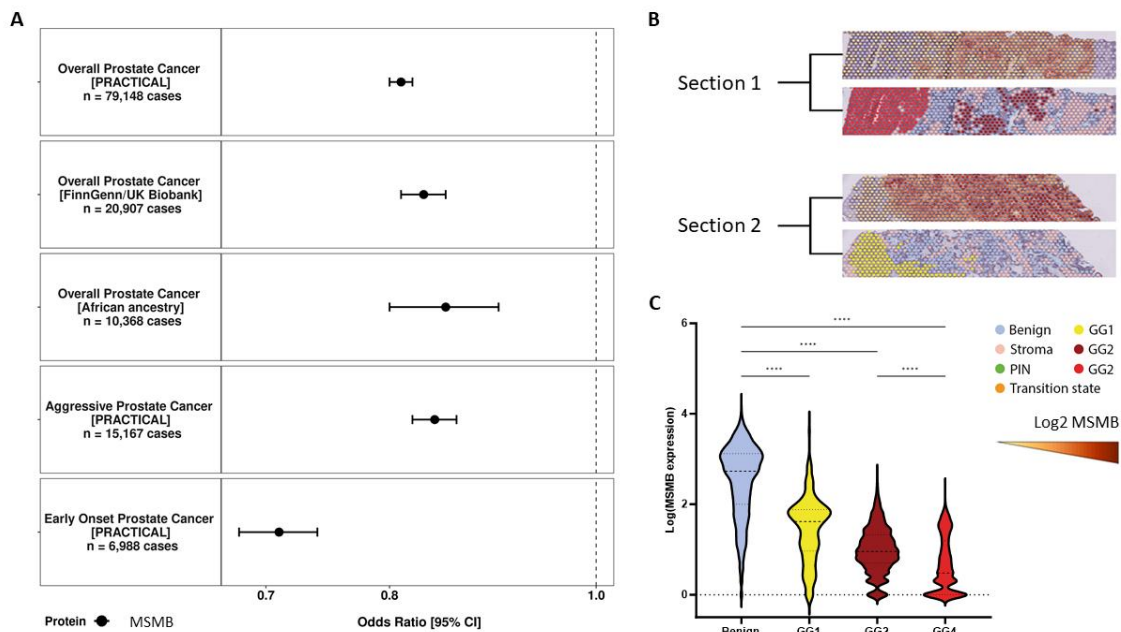
**Figure 1.** Association of genetically predicted protein concentrations with prostate cancer risk presented as a Manhattan plot where position is given by *cis*-pQTL coordinate (chromosome and base-pair position) labelled with their association with cancer risk and the highest colocalization probability from single or conditional iterative methods (PP4). Points highlighted as filled-in are those with evidence of a shared causal locus (PP4 > 0.70) with point size reflecting PP4 magnitude, which can vary between 0 and 1. Risk associations with MR  $p >$  Bonferroni correction threshold were not subject to colocalization analyses. The strongest protein-cancer association per chromosome is labelled and a zoomed-in plot for MSMB (rs10993994) on chromosome 10 is shown in the upper right-hand corner.



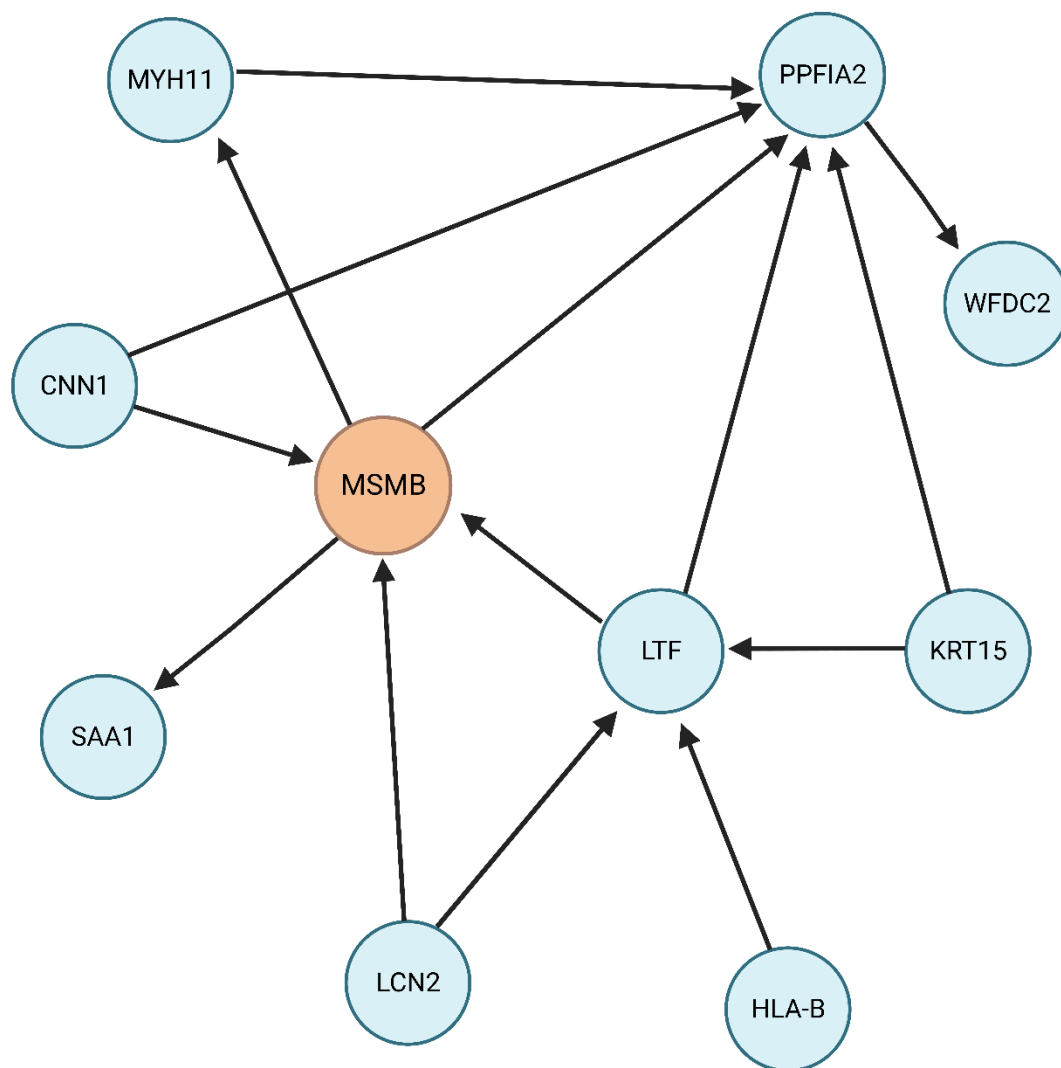
**Figure 2.** Odds ratios (95% confidence intervals) for genetically predicted protein levels and prostate cancer risk (for proteins with  $p <$  Bonferroni threshold based on  $0.05/\text{number of proteins analyzed}$ ). Odds ratio estimates are scaled per standard deviation increment in genetically predicted relative circulating protein concentrations. Filled circles represent Bonferroni-significant associations and asterisks indicate evidence for colocalization ( $PP4 > 0.70$ ).



**Figure 3.** Odds ratios (95% confidence intervals) for genetically predicted protein levels and overall prostate cancer risk for proteins with  $p <$  Bonferroni threshold based on  $0.05/\text{number}$  of proteins analyzed in main analyses, and with data available to perform replication in an African ancestry and European ancestry population (UK Biobank/FinnGen). Odds ratio estimates are scaled per standard deviation increment in genetically predicted circulating protein concentrations.

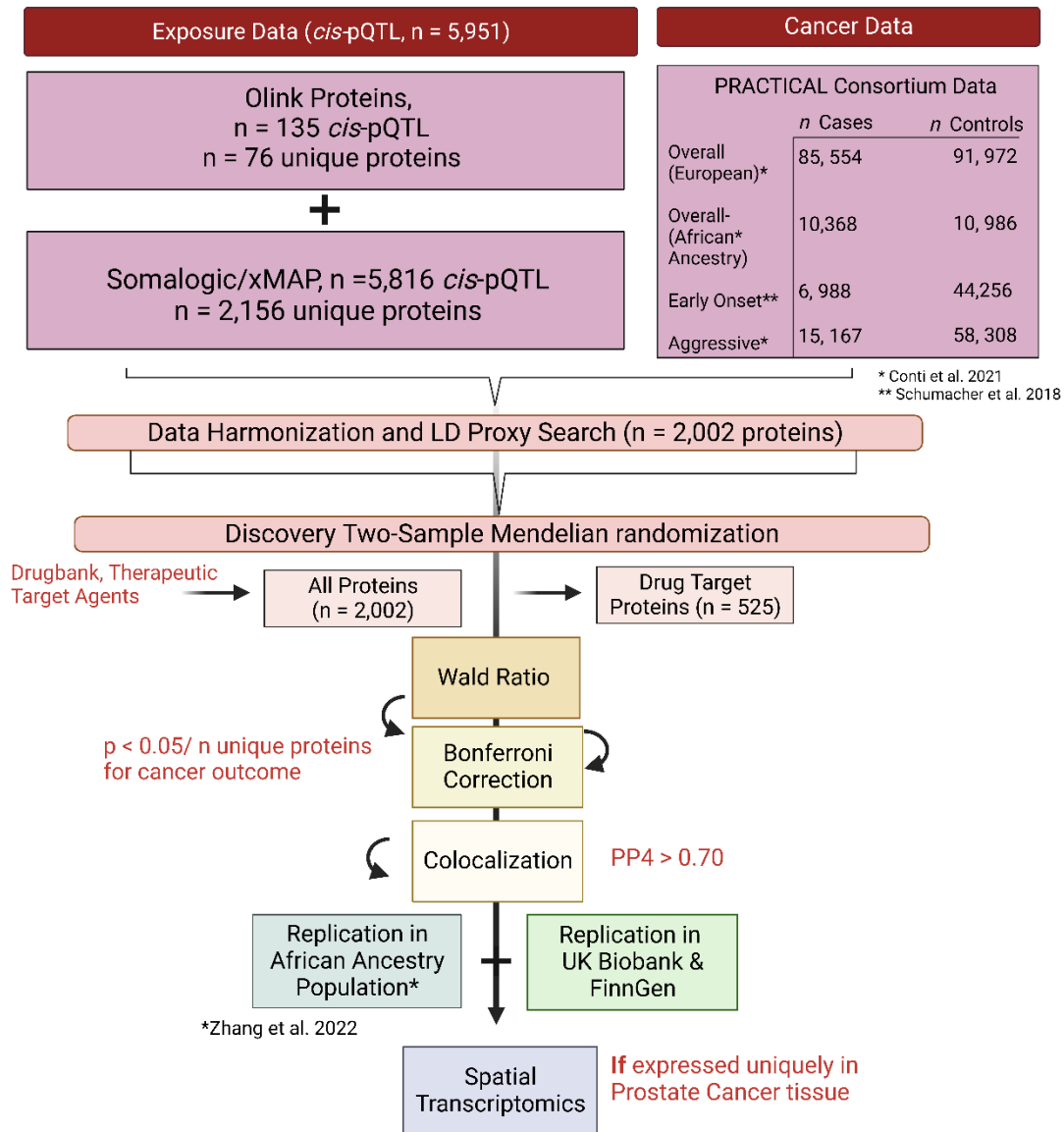


**Figure 4. A)** MSMB association with overall, early onset, and aggressive prostate cancer risk with replication in the FinnGen and UK Biobank populations and in an African ancestry population. Odds ratio (95% confidence interval) estimates are scaled per standard deviation increment in genetically predicted circulating MSMB concentrations **B)** Spatial visualization showing MSMB gene expression (top) and histology and tissue status (bottom) from organ-wide spatial transcriptomic data in two tumor sections (GG: Gleason grade group: GG1, Gleason score of 6 or lower; GG2, Gleason score of 3+4 = 7; GG4, Gleason score of 8). **C)** Violin plots representing gene expression in each spatial transcriptomics spot according to histological status. Statistical differences are indicated: \*\*\*\*  $p < 0.0001$  (Kruskal–Wallis; post-test: Dunn’s test).

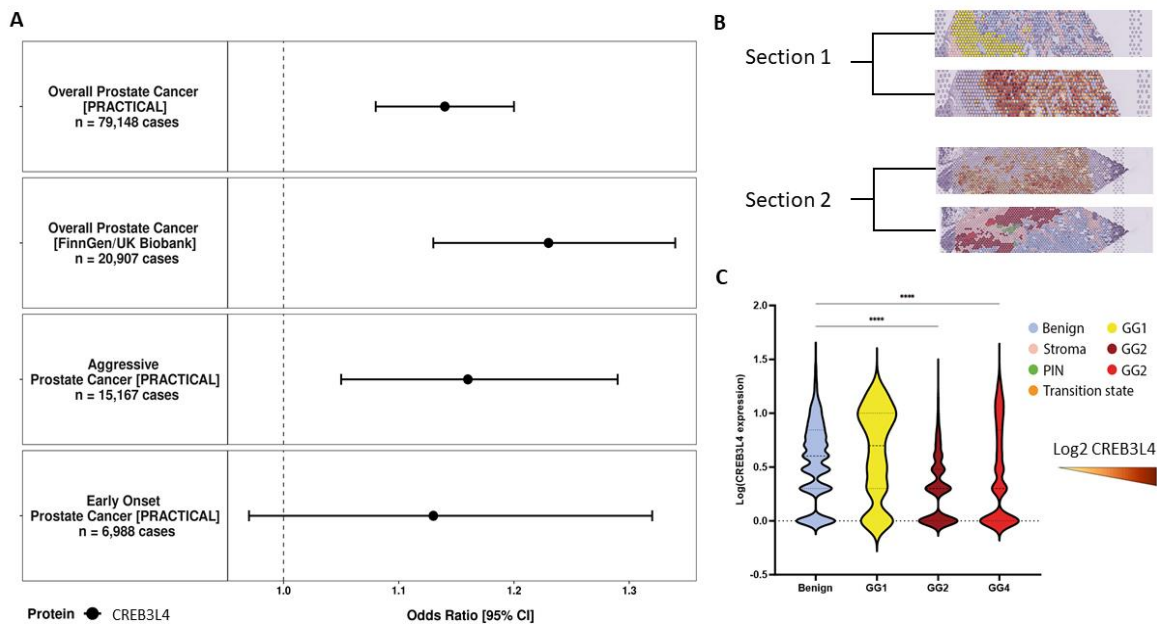


**Figure 5.** Gene network from iterative random forests of the difference in gene expression between benign and GG4 prostate histology (Gleason Score = 8). Arrows indicate direction of influence and shape of the network. MSMB is colored to demonstrate its central role in the network.





**Supplementary Figure 1. Flow-chart showing the overall study design. PP4 = posterior probability of a shared causal locus. LD = Linkage disequilibrium.**



**Supplementary Figure 2. A)** CREB3L4 association with overall, early onset, and aggressive prostate cancer risk with replication in the FinnGen and UK Biobank populations. Odds ratio (95% confidence interval) estimates are scaled per standard deviation increment in genetically predicted circulating CREB3L4 concentrations **B)** Spatial visualization showing CREB3L4 gene expression (top) and histology and tissue status (bottom) from organ-wide spatial transcriptomic data in two tumor sections (GG: Gleason grade group: GG1, Gleason score of 6 or lower; GG2, Gleason score of 3+4 = 7; GG4, Gleason score of 8). **C)** Violin plots representing gene expression in each spatial transcriptomics spot according to histological status. Statistical differences are indicated: \*\*\*\*  $p < 0.0001$  (Kruskal–Wallis; post-test: Dunn’s test).

## Supplementary material. PRACTICAL Authors Information

### PIs from the PRACTICAL (<http://practical.icr.ac.uk/>), CRUK, BPC3, CAPS, PEGASUS consortia:

Rosalind A. Eeles<sup>1,2</sup>, Christopher A. Haiman<sup>3</sup>, Zsofia Kote-Jarai<sup>1</sup>, Fredrick R. Schumacher<sup>4,5</sup>, Sara Benlloch<sup>6,1</sup>, Ali Amin Al Olama<sup>6,7</sup>, Kenneth R. Muir<sup>8</sup>, Sonja I. Berndt<sup>9</sup>, David V. Conti<sup>3</sup>, Fredrik Wiklund<sup>10</sup>, Stephen Chanock<sup>9</sup>, Ying Wang<sup>11</sup>, Catherine M. Tangen<sup>12</sup>, Jyotsna Batra<sup>13,14,15</sup>, Judith A. Clements<sup>13,14,15</sup>, APCB BioResource (Australian Prostate Cancer BioResource)<sup>16,17</sup>, Henrik Grönberg<sup>10</sup>, Nora Pashayan<sup>18,19</sup>, Johanna Schleutker<sup>20,21</sup>, Demetrius Albanes<sup>9</sup>, Stephanie J. Weinstein<sup>9</sup>, Alicja Wolk<sup>22</sup>, Catharine M. L. West<sup>23</sup>, Lorelei A. Mucci<sup>24</sup>, Géraldine Cancel-Tassin<sup>25,26</sup>, Stella Koutros<sup>9</sup>, Karina Dalsgaard Sørensen<sup>27,28</sup>, Eli Marie Grindedal<sup>29</sup>, David E. Neal<sup>30,31,32</sup>, Freddie C. Hamdy<sup>33,34</sup>, Jenny L. Donovan<sup>35</sup>, Ruth C. Travis<sup>36</sup>, Robert J. Hamilton<sup>37,38</sup>, Sue Ann Ingles<sup>39</sup>, Barry S. Rosenstein<sup>40</sup>, Yong-Jie Lu<sup>41</sup>, Graham G. Giles<sup>42,43,44</sup>, Robert J. MacInnis<sup>42,43</sup>, Adam S. Kibel<sup>45</sup>, Ana Vega<sup>46,47,48</sup>, Manolis Kogevinas<sup>49,50,51,52</sup>, Kathryn L. Penney<sup>53</sup>, Jong Y. Park<sup>54</sup>, Janet L. Stanford<sup>55,56</sup>, Cezary Cybulski<sup>57</sup>, Børge G. Nordestgaard<sup>58,59</sup>, Sune F. Nielsen<sup>58,59</sup>, Hermann Brenner<sup>60,61,62</sup>, Christiane Maier<sup>63</sup>, Jeri Kim<sup>64</sup>, Esther M. John<sup>65</sup>, Manuel R. Teixeira<sup>66,67,68</sup>, Susan L. Neuhausen<sup>69</sup>, Kim De Ruyck<sup>70</sup>, Azad Razack<sup>71</sup>, Lisa F. Newcomb<sup>55,72</sup>, Davor Lessel<sup>73</sup>, Radka Kaneva<sup>74</sup>, Nawaid Usmani<sup>75,76</sup>, Frank Claessens<sup>77</sup>, Paul A. Townsend<sup>78,79</sup>, Jose Esteban Castela<sup>80</sup>, Monique J. Roobol<sup>81</sup>, Florence Menegaux<sup>82</sup>, Kay-Tee Khaw<sup>83</sup>, Lisa Cannon-Albright<sup>84,85</sup>, Hardev Pandha<sup>79</sup>, Stephen N. Thibodeau<sup>86</sup>, David J. Hunter<sup>87</sup>, Peter Kraft<sup>88</sup>, William J. Blot<sup>89,90</sup>, Elio Riboli<sup>91</sup>

<sup>1</sup>The Institute of Cancer Research, London, SM2 5NG, UK

<sup>2</sup>Royal Marsden NHS Foundation Trust, London, SW3 6JJ, UK

<sup>3</sup>Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA

<sup>4</sup>Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106-7219, USA

<sup>5</sup>Seidman Cancer Center, University Hospitals, Cleveland, OH 44106, USA.

<sup>6</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge CB1 8RN, UK

<sup>7</sup>University of Cambridge, Department of Clinical Neurosciences, Stroke Research Group, R3, Box 83, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK

<sup>8</sup>Division of Population Health, Health Services Research and Primary Care, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

<sup>9</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, 20892, USA

<sup>10</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171 77 Stockholm, Sweden

<sup>11</sup>Department of Population Science, American Cancer Society, 250 Williams Street, Atlanta, GA 30303, USA

<sup>12</sup>SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

<sup>13</sup>School of Biomedical Sciences, Faculty of Health, Queensland University of Technology,

Brisbane, Queensland, Australia

<sup>14</sup>Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, Queensland, Australia

<sup>15</sup>Translational Research Institute, QUT, Woolloongabba, Brisbane, Queensland, Australia

<sup>16</sup>Australian Prostate Cancer Research Centre-Qld, Queensland University of Technology, Brisbane; Prostate Cancer Research Program, Monash University, Melbourne; Dame Roma Mitchell Cancer Centre, University of Adelaide, Adelaide; Chris O'Brien Lifehouse and The Kinghorn Cancer Centre, Sydney, Australia

<sup>17</sup>Translational Research Institute, Brisbane, Queensland 4102, Australia

<sup>18</sup>Department of Applied Health Research, University College London, London, WC1E 7HB, UK

<sup>19</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK

<sup>20</sup>Institute of Biomedicine, University of Turku, Finland

<sup>21</sup>Department of Medical Genetics, Genomics, Laboratory Division, Turku University Hospital, PO Box 52, 20521 Turku, Finland

<sup>22</sup>Institute of Environmental Medicine, Karolinska Institutet, 177 77 Stockholm, Sweden

<sup>23</sup>Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, M13 9PL UK

<sup>24</sup>Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

<sup>25</sup>CeRePP, Tenon Hospital, F-75020 Paris, France.

<sup>26</sup>Sorbonne Universite, GRC n°5 , AP-HP, Tenon Hospital, 4 rue de la Chine, F-75020 Paris, France

<sup>27</sup>Department of Molecular Medicine, Aarhus University Hospital, Palle Juul-Jensen Boulevard 99, 8200 Aarhus N, Denmark

<sup>28</sup>Department of Clinical Medicine, Aarhus University, DK-8200 Aarhus N

<sup>29</sup>Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway

<sup>30</sup>Nuffield Department of Surgical Sciences, University of Oxford, Room 6603, Level 6, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK

<sup>31</sup>University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK

<sup>32</sup>Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge, CB2 0RE, UK

<sup>33</sup>Nuffield Department of Surgical Sciences, University of Oxford, Oxford, OX1 2JD, UK

<sup>34</sup>Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK

<sup>35</sup>Population Health Sciences, Bristol Medical School, University of Bristol, BS8 2PS, UK

<sup>36</sup>Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK

<sup>37</sup>Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto ON M5G 2M9, Canada

<sup>38</sup>Dept. of Surgery (Urology), University of Toronto, Canada

<sup>39</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA

<sup>40</sup>Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Box 1236, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York,

NY 10029, USA

<sup>41</sup>Centre for Cancer Biomarker and Biotherapeutics, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ, UK

<sup>42</sup>Cancer Epidemiology Division, Cancer Council Victoria, 200 Victoria Parade, East Melbourne, VIC, 3002, Australia

<sup>43</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Grattan Street, Parkville, VIC 3010, Australia

<sup>44</sup>Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria 3168, Australia

<sup>45</sup>Division of Urologic Surgery, Brigham and Womens Hospital, 75 Francis Street, Boston, MA 02115, USA

<sup>46</sup>Fundación Pública Galega Medicina Xenómica, Santiago de Compostela, 15706, Spain.

<sup>47</sup>Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, 15706, Spain.

<sup>48</sup>Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain

<sup>49</sup>ISGlobal, Barcelona, Spain

<sup>50</sup>IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

<sup>51</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain

<sup>52</sup>CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain

<sup>53</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02115, USA

<sup>54</sup>Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA

<sup>55</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109-1024, USA

<sup>56</sup>Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington 98195, USA

<sup>57</sup>International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-115 Szczecin, Poland

<sup>58</sup>Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

<sup>59</sup>Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2200 Copenhagen, Denmark

<sup>60</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), D-69120, Heidelberg, Germany

<sup>61</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), D-69120 Heidelberg, Germany

<sup>62</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany

<sup>63</sup>Humangenetik Tuebingen, Paul-Ehrlich-Str 23, D-72076 Tuebingen, Germany

<sup>64</sup>The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, 1515 Holcombe Blvd., Houston, TX 77030, USA

<sup>65</sup>Departments of Epidemiology & Population Health and of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94304 USA

<sup>66</sup>Department of Laboratory Genetics, Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal

<sup>67</sup>Cancer Genetics Group, IPO Porto Research Center (CI-IPOP) / RISE@CI-IPOP (Health Research Network), Portuguese Oncology Institute of Porto (IPO Porto) / Porto Comprehensive Cancer Center, Porto, Portugal

<sup>68</sup>School of Medicine and Biomedical Sciences (ICBAS), University of Porto, Porto, Portugal

<sup>69</sup>Department of Population Sciences, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte, CA 91010

<sup>70</sup>Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Proeftuinstraat 86, B-9000 Gent

<sup>71</sup>Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

<sup>72</sup>Department of Urology, University of Washington, 1959 NE Pacific Street, Box 356510, Seattle, WA 98195, USA

<sup>73</sup>Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany

<sup>74</sup>Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria

<sup>75</sup>Department of Oncology, Cross Cancer Institute, University of Alberta, 11560 University Avenue, Edmonton, Alberta, Canada T6G 1Z2

<sup>76</sup>Division of Radiation Oncology, Cross Cancer Institute, 11560 University Avenue, Edmonton, Alberta, Canada T6G 1Z2

<sup>77</sup>Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, BE-3000, Belgium

<sup>78</sup>Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, M13 9WL

<sup>79</sup>The University of Surrey, Guildford, Surrey, GU2 7XH, UK

<sup>80</sup>Genetic Oncology Unit, CHUVI Hospital, Complejo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), 36204, Vigo (Pontevedra), Spain

<sup>81</sup>Department of Urology, Erasmus University Medical Center, Cancer Institute, 3015 GD Rotterdam, The Netherlands

<sup>82</sup>"Exposome and Heredity", CESP (UMR 1018), Faculté de Médecine, Université Paris-Saclay, Inserm, Gustave Roussy, Villejuif

<sup>83</sup>Clinical Gerontology Unit, University of Cambridge, Cambridge, CB2 2QQ, UK

<sup>84</sup>Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA

<sup>85</sup>George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah 84148, USA

<sup>86</sup>Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905, USA

<sup>87</sup>Nuffield Department of Population Health, University of Oxford, United Kingdom

<sup>88</sup>Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

<sup>89</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN 37232 USA.

<sup>90</sup>International Epidemiology Institute, Rockville, MD 20850, USA

<sup>91</sup>Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, SW7 2AZ, UK

### CRUK and PRACTICAL consortium

This work was supported by the Canadian Institutes of Health Research, European Commission's Seventh Framework Programme grant agreement n° 223175 (HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative).

We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now PCUK), The Orchid Cancer Appeal, Rosetrees Trust, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK, and Prostate Cancer Canada. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research, The Royal Marsden NHS Foundation Trust, and Manchester NIHR Biomedical Research Centre. The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, , 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296,), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support.

Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) [U19 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I]. Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher).

Research reported in this publication also received support from the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA37429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Funding for the iCOGS infrastructure came from: the European Community's Seventh Framework Programme under grant agreement n° 223175 (HEALTH-F2-2009-223175) (COGS), Cancer Research UK (C1287/A10118, C1287/A 10710, C12292/A11174, C1281/A12014, C5047/A8384, C5047/A15007, C5047/A10692, C8197/A16565), the National Institutes of Health (CA128978) and Post-Cancer GWAS initiative (1U19 CA148537, 1U19 CA148065 and 1U19 CA148112 - the GAME-ON initiative), the Department of Defence (W81XWH-10-1-0341), the Canadian Institutes of Health Research (CIHR) for the CIHR Team in Familial Risks of Breast Cancer, Komen Foundation for the Cure, the Breast Cancer Research Foundation, and the Ovarian Cancer Research Fund

### BPC3

The BPC3 was supported by the U.S. National Institutes of Health, National Cancer Institute (cooperative agreements U01-CA98233 to D.J.H., U01-CA98710 to S.M.G., U01-CA98216 to E.R., and U01-CA98758 to B.E.H., and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics).

### CAPS

CAPS GWAS study was supported by the Cancer Risk Prediction Center (CRiSP; [www.crispcenter.org](http://www.crispcenter.org)), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, (grant no K2010-70X-20430-04-3), the Swedish Cancer Foundation (grant no 09-0677), the Hedlund Foundation, the Soederberg Foundation, the Enqvist Foundation, ALF funds from the Stockholm County Council. Stiftelsen Johanna Hagstrand och Sigfrid Linner's Minne, Karlsson's Fund for urological and surgical research.

### PEGASUS

PEGASUS was supported by the Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health.