the Profile Study (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature Genetics, 50(7), 928-936. [8]. https://doi.org/10.1038/s41588-018-0142-8
TITLE:
Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci

AUTHOR LIST & AFFILIATIONS:

1 Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio, USA
2 Seidman Cancer Center, University Hospitals, Cleveland, Ohio, USA
3 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK.
4 University of Cambridge, Department of Clinical Neurosciences, Cambridge, UK
5 Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, USA
6 The Institute of Cancer Research, London, UK
7 Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, USA
8 Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain
9 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Cambridge, UK
10 Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK
11 Warwick Medical School, University of Warwick, Coventry, UK
12 Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA
13 SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
14 CHRISTUS Santa Rosa Hospital - Medical Center, San Antonio, Texas, USA
15 Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
16 Australian Prostate Cancer Research Centre-Qld, Translational Research Institute, Brisbane, Queensland, Australia
17 Menzies Health Institute of Queensland, Griffith University, Queensland, Australia
18 Cancer Council Queensland, Fortitude Valley, QLD, Australia
19 Chris O'Brien Lifehouse (COBLH), Camperdown, NSW, Australia
20 Garvan Institute of Medical Research, Sydney, NSW, Australia
21 Dame Roma Mitchell Cancer Research Centre, University of Adelaide, Adelaide, South Australia, Australia
22 Monash Biomedicine Discovery Institute Cancer Program, Prostate Cancer Research Program, Department of Anatomy and Developmental Biology, Monash University, Victoria, Australia
Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, Australia
Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
Department of Urology, Karolinska University Hospital, Stockholm, Sweden
Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
Department of Clinical Sciences at Danderyds Hospital, Karolinska Institutet, Stockholm, Sweden
Department of Applied Health Research, University College London, London, UK
Institute of Biomedicine, University of Turku, Finland
Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Finland
Department of Urology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Finland
Department of Epidemiology, School of Health Sciences, University of Tampere, Tampere, Finland
Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Sweden
Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, and The Christie NHS Foundation Trust, Manchester, UK
University of Cambridge, Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
Washington University School of Medicine, St. Louis, Missouri, USA
UPMC Sorbonne Universités, GRC No° 5 ONCOTYPE-URO, Tenon Hospital, Paris, France
CeRePP, Tenon Hospital, Paris, France
Department of Molecular Medicine, Aarhus University Hospital, Denmark
Department of Clinical Medicine, Aarhus University, Denmark
Department of Urology, Aarhus University Hospital, Denmark
Department of Medical Genetics, Oslo University Hospital, Oslo, Norway
University of Cambridge, Department of Oncology, Addenbrooke’s Hospital, Cambridge, UK
Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK
Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK
School of Social and Community Medicine, University of Bristol, Bristol, UK
Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
Department of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Canada
78 Biostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, Louisiana, USA
79 Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
80 Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, USA
81 International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
82 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA
83 Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
84 Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
85 Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
86 Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
87 Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
88 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
89 Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
90 Saarland Cancer Registry, Saarbrücken, Germany
91 Institute for Human Genetics, University Hospital Ulm, Ulm, Germany
92 Department of Urology, University Hospital Ulm, Germany
93 The University of Texas - MD Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston, Texas, USA
94 Cancer Prevention Institute of California, Fremont, California, USA
95 Department of Health Research & Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
96 Department of Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal
97 Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal
98 Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, California, USA
99 Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, Ghent, Belgium
100 Ghent University Hospital, Department of Radiotherapy, Ghent, Belgium
101 Department of Surgery, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
102 Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Selangor, Malaysia
103 Department of Urology, University of Washington, Seattle, Washington, USA
104 Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
Division of Medical Oncology, Urogenital Unit, Department of Oncology, University Hospital Centre Zagreb, Zagreb, Croatia

Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine, Zagreb, Croatia

Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, Bulgaria

Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta, Canada

Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta, Canada

Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, Sofia, Bulgaria

Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium

Department of Urology, University Hospitals Leuven, Leuven, Belgium

The University of Southampton, Southampton General Hospital, Southampton, UK

Division of Cancer Sciences, Manchester Cancer Research Centre, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Center, NIHR Manchester Biomedical Research Centre, Health Innovation Manchester, University of Manchester, Manchester, UK

The University of Surrey, Guildford, Surrey, UK

Genomic Medicine Group, Galician Foundation Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, Santiago De Compostela, Spain

University of California San Diego, Moores Cancer Center, La Jolla, California, USA

Genetic Oncology Unit, CHUVI Hospital, Complexo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), Vigo SERGAS, Spain

University of California San Diego, Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, California, USA

Department of Urology, Erasmus University Medical Center, Rotterdam, the Netherlands

Department of Clinical Chemistry, Erasmus University Medical Center, Rotterdam, the Netherlands

Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France


Program for Personalized Cancer Care, NorthShore University HealthSystem, Evanston, Illinois, USA
125 Clinical Gerontology Unit, University of Cambridge, Cambridge, UK
126 Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA
127 George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah, USA
128 Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA
129 Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, USA
130 Department of Epidemiology, Health Sciences Building, University of Washington, Seattle, Washington, USA
131 Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
132 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
133 Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square, London, UK
134 Genomic Epidemiology Group, German Cancer Research Center, (DKFZ), Heidelberg, Germany
135 Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA
136 A list of members and affiliations appears in the Supplementary Note
137 Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA
138 Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA
139 Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA
140 Royal Marsden NHS Foundation Trust, London, UK
141 In memorium
142 These authors contributed equally to this work
143 These authors jointly directed this work

Correspondence should be addressed to F.R.S. (frs2@case.edu), A. Al Olama (aa461@medschl.cam.ac.uk) and R.A.E. (ros.eeles@icr.ac.uk)
ABSTRACT:
Currently genome-wide association studies (GWAS) and fine-mapping efforts have identified over 100 prostate cancer (PrCa) susceptibility loci. We meta-analyzed genotype data from a custom high-density array of 46,939 PrCa cases and 27,910 controls of European ancestry with previously genotyped data of 32,255 PrCa cases and 33,202 controls of European ancestry. Our analysis identified 62 novel loci associated ($P<5.0\times10^{-8}$) with PrCa, and a locus significantly associated with early-onset PrCa ($\leq 55$ years). Our findings include missense variants rs1800057 (OR=1.16; $P=8.2\times10^{-9}$; G>C [Pro1054Arg]) in ATM and rs2066827 (OR=1.06; $P=2.3\times10^{-9}$; T>G [Val109Gly]) in CDKN1B. The combination of all loci captures 28.4% of the PrCa familial relative risk and a polygenic risk score confers an elevated PrCa risk for men in the 90-99%-ile (RR=2.69; 95%CI: 2.55-2.82) and 1%-ile (RR=5.71; 95%CI: 5.04-6.48) risk stratum compared to the population average. These findings improve risk prediction, enhance fine-mapping, and provide insight into the underlying biology of PrCa. 

MAIN TEXT:
Although prostate cancer (PrCa) is the most common non-cutaneous cancer among men in the Western world and 1 in 7 men will be diagnosed during their lifetime, very few modifiable risk factors have been established. Epidemiological studies have identified age, a positive family history and race/ethnicity as the most prominent risk factors for PrCa. PrCa incidence is highest among men of African ancestry, followed by men of European and Asian ancestry. These ancestral differences of PrCa risk, in conjunction with studies demonstrating the influence of family history, highlight the contribution of genetics in PrCa etiology. Our previous work, utilizing a multiplicative model, estimated over 1,800 common single nucleotide polymorphisms (SNPs) independently contribute to PrCa risk among populations of European ancestry. Genome-wide association studies (GWAS) have reported over 100 of these PrCa variants across multi-ethnic populations, with the vast majority being identified in populations of European ancestry.

To facilitate additional novel discovery of PrCa genetic risk factors we developed a custom high-density genotyping array, the OncoArray, including a 260K SNP backbone designed to adequately tag most common genetic variants (MAF>5% in Europeans),
and 310K SNPs from the meta-analyses of five cancers (breast, colorectal, lung, ovarian, and prostate). Approximately 80,000 PrCa-specific markers derived from our previous multi-ethnic meta-analysis (including populations of European, African American, Japanese, and Latino ancestry), fine-mapping of known PrCa loci, and candidate SNPs nominated by study collaborators were included on the OncoArray. We assembled a new PrCa sample series from 52 studies to genotype with the OncoArray (Supplementary Tables 1 & 2). After applying rigorous quality control criteria and removing overlapping samples from previous studies, our OncoArray sample yielded 46,939 PrCa cases and 27,910 controls without a known diagnosis of PrCa of European ancestry for analysis (see Online Methods, Supplementary Table 3). Genotypes were phased and imputed to the cosmopolitan panel of the 1000 Genomes Project (1KGP; 2014 June release) using SHAPEIT31 and IMPUTEv232 software (Online Methods, Supplementary Table 3). We performed a fixed-effects meta-analysis combining the summary statistics from our OncoArray analysis, and seven previous PrCa GWAS or high-density SNP panels of European ancestry imputed to 1KGP. The final meta-analysis included 79,194 PrCa cases and 61,112 controls without a known diagnosis of PrCa (Figure 1).

Study- and consortia-specific meta-analyses were performed to identify novel PrCa loci. We established a P-value threshold of 5.0x10^{-8} to determine genome-wide significance. Our large sample size enabled several stratified meta-analyses focusing on key clinical and biological parameters (Online Methods, Supplementary Tables 4 & 5). All analyses used a likelihood ratio test to minimize bias from rare variants and a logistic regression framework was used for all analyses, except for Gleason score where linear regression was utilized. The genotype dosages were incorporated in an allelic genetic model. The average λ_{1000}, an inflation statistic calibrated to a sample size of 1000 cases and 1000 controls, across the eight GWAS studies was 1.02 (range: 0.98-1.09) and 1.00 for the overall meta-analysis (Supplementary Table 6). Our novel findings excluded variants within defined fine-mapped regions of previously reported PrCa loci (Supplementary Table 7).

After the exclusion of all known susceptibility regions (fine-mapping coordinates provided in Supplementary Table 7 & Supplementary Note) we identified 64 loci associated with overall PrCa susceptibility and one with early-onset (P<5.0x10^{-8}) in the meta-
analysis (Supplementary Figure 1), where 53 were imputed and 12 were genotyped using the OncoArray. The cluster plots for the genotyped makers are presented in Supplementary Figure 2. Although a majority of the imputed markers were of high quality with an average imputed $r^2$ greater than 0.80 for 61 of the 65 loci across all contributing GWAS (Supplementary Table 8), we closely examined four variants with a poor imputation quality score ($r^2 < 0.80$) in the OncoArray samples by inspecting linkage disequilibrium (LD) plots including only genotyped SNPs from the OncoArray and performing an imputation quality control assessment (Online Methods). After reviewing the LD plots and the imputation QC, we determined loci rs6602880 and rs144166867 are likely false positives due to imputation artifacts (Supplementary Figure 3; Supplementary Table 9). Overall, we identified 62 novel loci associated with overall PrCa risk and one novel loci associated with early-onset (Table 1). The consortia specific associations were consistent across the eight contributing GWAS studies (Supplementary Table 10).

We performed several stratified analyses defined by clinical and population parameters. We detected a novel variant, rs138004030, significantly associated with early-onset disease (Table 1), but only nominally significant for overall PrCa risk ($P=0.02$). In addition, we detected four markers significantly associated ($P<5\times10^{-8}$) with advanced PrCa and two markers associated with early-onset PrCa (Supplementary Table 11). However, the case-only analyses of these markers indicated marginal statistical significance ($P<1.0\times10^{-3}$). Additionally, these markers were in LD with nearby index markers associated with overall PrCa and not significantly associated with overall aggressive disease after adjusting for the index marker (Supplementary Table 11). A similar association pattern was observed for rs111599055, which was in LD with marker rs7295014 ($r^2=0.54$) associated with overall disease. The early-onset marker rs77777548 is independent of novel and known PrCa loci. However, the marker is relatively rare (EAF<0.02), indicated as monomorphic in 1KGP, and has a moderate imputation quality score (average $r^2=0.57$) hence we did not include it in further analyses.

Among the 63 novel associations, 38 variants are located within gene-rich regions (Supplementary Table 12): intronic (32 SNPs), missense (4 SNPs), and 3'-UTR (2 SNPs). eQTL analyses of the TCGA database identified statistically significant
associations \( (P<0.05; \text{Supplementary Table 12}) \) in normal PrCa tissue for 17 of the novel associations, including both 3'UTR SNPs and 11 of the 32 intronic SNPs. \textit{Cis} eQTL associations were identified for 3'UTR variant rs1048169 with \textit{HAUS6} (3'UTR) and intronic variants rs182314334 with \textit{MBNL1}, rs4976790 with \textit{COL23A1}, rs9469899 with \textit{UHRF1BP1}, rs878987 with \textit{B3GAT1}, rs11629412 with \textit{PAX9}, and rs11666569 with \textit{MYO9B}. The eQTL associations are consistent with the observed PrCa-SNP associations, given we assessed colocalization between the GWAS and eQTL SNPs. The TCGA data analysis failed to identify an eQTL association with any of the four missense SNPs.

We assessed the association of our newly discovered loci with prostate-specific antigen (PSA) levels using a series of disease-free controls \((N=9,090; \text{see Online Methods})\). Among the 48 available loci we observed a significant association for rs8093601 \( (P=5.0\times10^{-4}; \text{Supplementary Table 13}) \) after correcting for multiple testing \( (P=0.05/48=1.0\times10^{-3}) \). This marker lies near \textit{MBD2}, methyl-CpG binding domain protein 2, and has not been previously associated with either PrCa risk or PSA levels. The effect estimates of PrCa clinical features and overall PrCa did not differ (Supplementary Table 14). LD plots incorporating several functional annotation features for each of the 63 novel markers is presented in Supplementary Figure 4.

Several strong candidate genes were identified among the PrCa susceptibility loci, including \textit{ATM}, a key gene within the DNA damage response pathway, in which truncating variants contribute towards PrCa susceptibility and progression, particularly aggressive PrCa\(^{34,35}\). The index variant within this region is the missense variant rs1800057, exerting a modest increased risk of PrCa \( (\text{OR}=1.16; P=8.15\times10^{-9}; \text{G}>\text{C} \) [Pro1054Arg]; Figure 2, Panel A). Although rs1800057 is designated ‘benign’ by ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), it was previously suggested to be associated with a two-fold increased risk of early-onset PrCa in a small clinical series and was unassociated with morbidity following treatment\(^{36}\). In addition to the \textit{ATM} region, we identified missense variants in three separate loci: rs2066827 within the cyclin-dependent kinase inhibitor \textit{CDKN1B}, which controls cell cycle progression; rs33984059 within the transcription factor \textit{RFX7}; and rs2277283 within \textit{INCENP}, which encodes a centromere-interacting protein.
rs1048169 at 9p22 is located in the 3'UTR of HAUS6 (Figure 2, Panel B), a gene that encodes a subunit of augmin, a protein complex required for proper microtubule formation and chromosome segregation during cell division.\(^3\). rs1048169 is also an eQTL for HAUS6 expression. Interestingly, an additional lead SNP identified in this study, rs11666569 at 19p13, was an eQTL for two genes including HAUS8, which is another member of the augmin complex. These discoveries may implicate a potential role for augmin in PrCa susceptibility.

Variant rs7968403 (OR=1.06; \(P=3.38 \times 10^{-12}\); Figure 2, Panel C) is situated within the first intron of RASSF3. Members of the Ras association domain family (RASSF) are putative tumor suppressors implicated in a range of biological processes.\(^3\) RASSF3 is ubiquitously expressed across tissue types and has been observed to arrest the cell cycle in the G1 phase and induce apoptosis through the p53 pathway.\(^3\) A previously identified PrCa risk locus, ~100kb away, within the RASSF6 family member was previously identified.\(^1\) However, rs7968403 was also an eQTL for the distant WIF1 (WNT inhibitory factor 1) gene (Figure 2, Panel C). WIF1 inhibits Wnt signaling and is frequently down regulated in PrCa,\(^4\) whilst aberrant activation of Wnt signaling is common in many solid tumor types. Restoration of WIF1 expression has also been demonstrated to decrease cell motility and invasiveness in a metastatic PrCa cell-line and reduce tumor growth in a mouse xenograft model.\(^4\) Both RASSF3 and WIF1 therefore represent plausible mechanisms for the modulation of PrCa risk at this locus.

rs28441558 at 17p13 is the lead variant for a cluster of highly correlated SNPs centered on the CHD3 gene (Figure 2, Panel D). CHD3 is an ATPase that forms a component of the NuRD (nucleosome remodeling and deacetylase) histone deacetylase complex, involved in chromatin remodeling. NuRD plays an important role in regulating gene expression, both as a silencer and activator of transcription, in addition to maintenance of genomic integrity and the DNA damage response.\(^4\) Alterations to NuRD function have been implicated in several cancer types in a highly complex manner.\(^4\) Additionally however, rs28441558 was observed to be an eQTL for three genes; LOC284023, a currently uncharacterized non-coding RNA transcript, GUCY2D, a guanylate cyclase enzyme expressed predominantly in the retina and ALOX15B, a member of the lipoygenase family of enzymes that produce fatty acid hydroperoxides.
Although *CHD3* appears to represent the most biologically plausible candidate gene for this locus, we cannot exclude a role for any of these genes.

Our pathway analysis based on mapping each SNP to the nearest gene (see Online Methods) using the meta-analysis summary association statistic identified several pathways implicated in PrCa susceptibility. The top 53 pathways detected (enrichment score, ES>0.50) are provided (Supplementary Table 15). The most significant pathway detected was PD-1 signaling (ID: 389948), ES=0.74, as defined by the REACTOME database (Supplementary Figure 5). This pathway is intriguing given the therapeutic potential of several checkpoint inhibitors focusing on the PD-1 signaling pathway to enhance immune responses\(^{45}\).

In summary, we have identified 63 novel PrCa susceptibility variants, including strong candidate loci highlighting the DNA repair and cell cycle pathways. Previous studies likely overestimated the effect estimates of PrCa loci due to the “winner’s curse”, thus yielding a biased FRR and polygenic risk score (PRS). Here, we apply a weighted Bayesian correction approach and demonstrate our large sample size minimizes the “winner’s curse” bias (Online Methods; Supplementary Figure 6)\(^{46}\). We applied the beta estimates calculated in our overall meta-analysis to the OncoArray sample set to calculate the FRR and PRS risk models (Supplementary Table 16). Our prediction models included 85 previously reported PrCa loci replicating in our overall meta-analysis and our 62 novel loci associated with overall PrCa risk. Assuming a familial risk estimate of 2.5 for PrCa\(^{47,48}\), we demonstrate our 147 loci captures 28.4% of the FRR (Supplementary Table 17). The newly 62 identified PrCa loci increase the FRR by 4.4\%. On the assumption of a log-additive model, the estimated RR for PrCa relative to men in the 25-75% PRS percentile (baseline group) was 5.71 (95%CI: 5.04-6.48) for men in the top 1% of the polygenic risk score (PRS) distribution and 2.69 (95%CI: 2.55-2.82) for individuals in the 90-99%-ile of the PRS (Table 2). The PRS score was positively associated with overall PrCa compared to all controls (OR=1.86; 95%CI: 1.83-1.89; Supplementary Table 18). Our novel associations highlight several biological pathways that suggest further investigation is warranted. The increased PRS can be used to improve the identification of men at high risk of PrCa and therefore inform PSA guidelines for screening and management to reduce the burden of over testing.
ACKNOWLEDGMENTS:
The authors wish to pay tribute to Brian Henderson, who was a driving force behind the OncoArray project, for his vision and leadership, and sadly passed away before seeing its fruition.

We thank the individuals who participated in these studies enabling this work.

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 (F.R.S.).

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/A16692, 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.

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 Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust.

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.

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 GWAS study was supported by the Swedish Cancer Foundation (grant no 09-0677, 11-484, 12-823), the Cancer Risk Prediction Center (CRisP; www.crispcenter.org), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, Swedish Research Council (grant no K2010-70X-20430-04-3, 2014-2269)

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

For a full description of funding and acknowledgments see the Supplementary Note.

AUTHOR CONTRIBUTIONS:

COMPETING FINANCIAL INTERESTS STATEMENT:
The authors declare no competing financial interests.
REFERENCES:


41. Yee, D.S. *et al.* The Wnt inhibitory factor 1 restoration in prostate cancer cells was associated with reduced tumor growth, decreased capacity of cell migration and invasion and a reversal of epithelial to mesenchymal transition. *Mol Cancer* **9**, 162 (2010).


FIGURE LEGENDS:

Figure 1. ELLIPSE/PRACTICAL study overview of prostate cancer (PrCa) GWAS meta-analysis. The top section describes the PrCa GWAS meta-analysis published in 2014 (AA Olama et al, Nature Genetics 2014\textsuperscript{12}) where 23 novel variants were identified. The current PrCa GWAS meta-analysis incorporates an additional 46,939 PrCa cases and 27,910 controls independent of the meta-analyses. The current meta-analysis discovered 62 novel variants associated with overall PrCa and 1 novel variant associated with early-onset PrCa.

Figure 2. Locus Explorer plots depicting the statistical association with PrCa and biological context of variants from four of the novel prostate cancer loci identified (N=74,849 biologically independent samples). For each panel (a-d), top panels depict Manhattan plots of variant -log10 $P$ values (y-axis), with the index SNP labeled. Variants that were directly genotyped by the OncoArray are represented as triangles and imputed variants are represented as circles. Variants in linkage disequilibrium with the index SNP are denoted by color (red = $r^2$ >0.8, orange = $r^2$ 0.6-0.8, yellow = $r^2$ 0.4-0.6, green = $r^2$ 0.2-0.4, blue = $r^2$ ≤0.2). Middle panels depict the relative locations of selected biological annotations; histone marks within 7 cell lines from the ENCODE project; genes for which the index SNP is an eQTL in the TCGA prostate adenocarcinoma dataset; chromatin state annotation by ChromHMM in PrEC cells; conserved elements within the genome and DNAsel hypersensitivity sites in ENCODE prostate cell lines. The lower panel denotes the position of genes within the region, with genes on the positive and negative strands marked in green and purple, respectively. The horizontal axis represents genomic co-ordinates in the hg19 reference genome. (a) rs1800057 (chr11:107643000-108644000) - The index variant is a non-synonymous SNP in the ATM gene. (b) rs1048160 (chr9:18556000-19557000) - The index variant is located within the 3'UTR of the HAUS6 gene and is an eQTL for HAUS6. (c) rs7968403 (chr12:64513000-65514000) - The signal is centered on the RASSF3 gene, with the index variant located within the first intron. This SNP is also situated within a region annotated for multiple regulatory markers and is an eQTL for the more distant WIF1 gene. (d) rs28441558 (chr17:7303000-8304000) - The signal implicates a cluster of highly correlated variants centered upon the CHD3 gene. The index SNP is also an eQTL for three other more distantly located genes.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Reference RAF</th>
<th>Band</th>
<th>Position</th>
<th>Nearest Gene</th>
<th>Alleles</th>
<th>RAF</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs56391074</td>
<td>0.329</td>
<td>1p22.3</td>
<td>88210715</td>
<td>RP11-60A14.1</td>
<td>AT/A</td>
<td>0.38</td>
<td>1.05</td>
<td>1.03-1.06</td>
<td>1.7E-08</td>
</tr>
<tr>
<td>rs34579442</td>
<td>0.316</td>
<td>1q21.3</td>
<td>153899900</td>
<td>DENND4B</td>
<td>C/CT</td>
<td>0.34</td>
<td>1.07</td>
<td>1.05-1.09</td>
<td>4.5E-14</td>
</tr>
<tr>
<td>rs62106670</td>
<td>0.400</td>
<td>2p25.1</td>
<td>8597123</td>
<td>AC011747.3</td>
<td>T/C</td>
<td>0.38</td>
<td>1.05</td>
<td>1.04-1.07</td>
<td>7.1E-09</td>
</tr>
<tr>
<td>rs74702681</td>
<td>0.024</td>
<td>2p14</td>
<td>66652885</td>
<td>MEIS1-AS3</td>
<td>T/C</td>
<td>0.02</td>
<td>1.17</td>
<td>1.11-1.23</td>
<td>2.0E-09</td>
</tr>
<tr>
<td>rs11691517</td>
<td>0.750</td>
<td>2q13</td>
<td>111893096</td>
<td>BCL2L11</td>
<td>T/G</td>
<td>0.74</td>
<td>1.07</td>
<td>1.05-1.08</td>
<td>3.5E-12</td>
</tr>
<tr>
<td>rs34925593</td>
<td>0.481</td>
<td>2q31.1</td>
<td>174234547</td>
<td>CDC47</td>
<td>C/T</td>
<td>0.48</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>2.8E-08</td>
</tr>
<tr>
<td>rs59308963</td>
<td>0.726</td>
<td>3q13.12</td>
<td>106962521</td>
<td>RNU6-456P</td>
<td>T/TATTCTGTC</td>
<td>0.73</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>2.4E-08</td>
</tr>
<tr>
<td>rs1283104</td>
<td>0.407</td>
<td>3q25.1</td>
<td>152004202</td>
<td>MBLN1</td>
<td>T/C</td>
<td>0.90</td>
<td>1.09</td>
<td>1.06-1.12</td>
<td>4.1E-11</td>
</tr>
<tr>
<td>rs182314334</td>
<td>0.012</td>
<td>3q26.2</td>
<td>169093100</td>
<td>MECOM</td>
<td>G/A</td>
<td>0.01</td>
<td>1.25</td>
<td>1.16-1.34</td>
<td>4.7E-09</td>
</tr>
<tr>
<td>rs10793821</td>
<td>0.580</td>
<td>5q31.1</td>
<td>133836209</td>
<td>RNU6-456P</td>
<td>T/C</td>
<td>0.57</td>
<td>1.05</td>
<td>1.04-1.07</td>
<td>5.4E-11</td>
</tr>
<tr>
<td>rs76551843</td>
<td>0.991</td>
<td>5q35.1</td>
<td>169172133</td>
<td>DOCK2</td>
<td>A/G</td>
<td>0.99</td>
<td>1.31</td>
<td>1.19-1.44</td>
<td>1.7E-08</td>
</tr>
<tr>
<td>rs4976790</td>
<td>0.096</td>
<td>5q35.3</td>
<td>177968915</td>
<td>COL23A1</td>
<td>T/G</td>
<td>0.11</td>
<td>1.08</td>
<td>1.05-1.10</td>
<td>6.7E-09</td>
</tr>
<tr>
<td>rs12665339</td>
<td>0.148</td>
<td>6p21.33</td>
<td>30601232</td>
<td>ATAT1</td>
<td>G/A</td>
<td>0.17</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>5.6E-09</td>
</tr>
<tr>
<td>rs9296068</td>
<td>0.645</td>
<td>6p21.32</td>
<td>32988695</td>
<td>HLA-DOA</td>
<td>T/G</td>
<td>0.65</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>1.3E-08</td>
</tr>
<tr>
<td>rs9469899</td>
<td>0.356</td>
<td>6p21.31</td>
<td>34793124</td>
<td>UHRF1BP1</td>
<td>A/G</td>
<td>0.36</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>5.3E-09</td>
</tr>
<tr>
<td>rs4711748</td>
<td>0.232</td>
<td>6p21.1</td>
<td>43694598</td>
<td>RP1-261G23.5</td>
<td>T/C</td>
<td>0.23</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>3.4E-08</td>
</tr>
<tr>
<td>rs527510716</td>
<td>0.251</td>
<td>7p22.3</td>
<td>1944537</td>
<td>MAD1L1</td>
<td>C/G</td>
<td>0.24</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>4.9E-08</td>
</tr>
<tr>
<td>rs11452686</td>
<td>0.567</td>
<td>7p21.1</td>
<td>20414110</td>
<td>ITGB8</td>
<td>T/TA</td>
<td>0.56</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>7.8E-09</td>
</tr>
<tr>
<td>rs17621345</td>
<td>0.758</td>
<td>7p14.1</td>
<td>40875192</td>
<td>SUGCT</td>
<td>A/C</td>
<td>0.74</td>
<td>1.07</td>
<td>1.05-1.09</td>
<td>6.7E-14</td>
</tr>
<tr>
<td>rs1048169</td>
<td>0.367</td>
<td>9p22.1</td>
<td>19055965</td>
<td>HAUS6</td>
<td>C/T</td>
<td>0.38</td>
<td>1.06</td>
<td>1.05-1.08</td>
<td>6.5E-14</td>
</tr>
<tr>
<td>rs10122495</td>
<td>0.296</td>
<td>9p13.3</td>
<td>34049779</td>
<td>RN7SKP114</td>
<td>T/A</td>
<td>0.31</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>1.3E-08</td>
</tr>
<tr>
<td>SNP</td>
<td>T-Stat</td>
<td>Chromosome</td>
<td>Position</td>
<td>Gene</td>
<td>Mutation</td>
<td>p-Value</td>
<td>OR</td>
<td>95% CI</td>
<td>Minor Allele</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>-----</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>rs1182</td>
<td>0.258</td>
<td>9q34.11</td>
<td>132576060</td>
<td>TOR1A</td>
<td>A/C</td>
<td>0.22</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>1.1E-09</td>
</tr>
<tr>
<td>rs141536087</td>
<td>0.166</td>
<td>10p15.3</td>
<td>854691</td>
<td>LARP4B</td>
<td>G/C</td>
<td>0.15</td>
<td>1.08</td>
<td>1.06-1.11</td>
<td>9.0E-13</td>
</tr>
<tr>
<td>rs1935581</td>
<td>0.605</td>
<td>10q23.31</td>
<td>90195149</td>
<td>RNLS</td>
<td>C/T</td>
<td>0.63</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>6.5E-09</td>
</tr>
<tr>
<td>rs7094871</td>
<td>0.540</td>
<td>10q25.2</td>
<td>114712154</td>
<td>TCF7L2</td>
<td>G/C</td>
<td>0.54</td>
<td>1.04</td>
<td>1.03-1.06</td>
<td>4.8E-08</td>
</tr>
<tr>
<td>rs1881502</td>
<td>0.193</td>
<td>11p15.5</td>
<td>1507512</td>
<td>MOB2</td>
<td>T/C</td>
<td>0.19</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>7.4E-09</td>
</tr>
<tr>
<td>rs61890184</td>
<td>0.088</td>
<td>11p15.4</td>
<td>7547587</td>
<td>PPFIBP2</td>
<td>A/G</td>
<td>0.12</td>
<td>1.07</td>
<td>1.05-1.10</td>
<td>6.6E-09</td>
</tr>
<tr>
<td>rs547171081</td>
<td>0.468</td>
<td>11p11.2</td>
<td>47421962</td>
<td>RP11-750H9.5</td>
<td>C/G</td>
<td>0.47</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>3.4E-08</td>
</tr>
<tr>
<td>rs2277283</td>
<td>0.300</td>
<td>11q12.3</td>
<td>61908440</td>
<td>INCENP</td>
<td>C/T</td>
<td>0.31</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>3.0E-10</td>
</tr>
<tr>
<td>rs12785905</td>
<td>0.051</td>
<td>11q13.2</td>
<td>66951965</td>
<td>KDM2A</td>
<td>C/G</td>
<td>0.05</td>
<td>1.12</td>
<td>1.08-1.17</td>
<td>7.8E-09</td>
</tr>
<tr>
<td>rs11290954</td>
<td>0.688</td>
<td>11q13.5</td>
<td>76260543</td>
<td>C11orf30</td>
<td>A/C</td>
<td>0.68</td>
<td>1.06</td>
<td>1.05-1.08</td>
<td>7.4E-13</td>
</tr>
<tr>
<td>rs1800057</td>
<td>0.031</td>
<td>11q22.3</td>
<td>108143456</td>
<td>ATM</td>
<td>G/C</td>
<td>0.02</td>
<td>1.16</td>
<td>1.10-1.22</td>
<td>8.1E-09</td>
</tr>
<tr>
<td>rs138466039</td>
<td>0.009</td>
<td>11q24.2</td>
<td>125054793</td>
<td>PKNOX2</td>
<td>T/C</td>
<td>0.01</td>
<td>1.32</td>
<td>1.22-1.44</td>
<td>2.0E-11</td>
</tr>
<tr>
<td>rs878987</td>
<td>0.143</td>
<td>11q25</td>
<td>134266372</td>
<td>B3GAT1</td>
<td>C/T</td>
<td>0.15</td>
<td>1.07</td>
<td>1.04-1.09</td>
<td>4.8E-08</td>
</tr>
<tr>
<td>rs2066827</td>
<td>0.757</td>
<td>12p13.1</td>
<td>12871099</td>
<td>CDKN1B</td>
<td>C/T</td>
<td>0.76</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>2.3E-09</td>
</tr>
<tr>
<td>rs10845938</td>
<td>0.554</td>
<td>12p13.1</td>
<td>14416918</td>
<td>RNU6-491P</td>
<td>G/A</td>
<td>0.55</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>9.8E-13</td>
</tr>
<tr>
<td>rs7968403</td>
<td>0.655</td>
<td>12q14.2</td>
<td>65012824</td>
<td>RASSF3</td>
<td>T/C</td>
<td>0.64</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>3.4E-12</td>
</tr>
<tr>
<td>rs5799921</td>
<td>0.697</td>
<td>12q21.33</td>
<td>90160530</td>
<td>RNU6-148P</td>
<td>GA/G</td>
<td>0.68</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>7.0E-12</td>
</tr>
<tr>
<td>rs7295014</td>
<td>0.342</td>
<td>12q24.33</td>
<td>133067989</td>
<td>FBRSL1</td>
<td>G/A</td>
<td>0.35</td>
<td>1.05</td>
<td>1.04-1.07</td>
<td>9.5E-10</td>
</tr>
<tr>
<td>rs1004030</td>
<td>0.581</td>
<td>14q11.2</td>
<td>23305649</td>
<td>MMP14</td>
<td>C/T</td>
<td>0.58</td>
<td>1.05</td>
<td>1.03-1.06</td>
<td>1.5E-08</td>
</tr>
<tr>
<td>rs11629412</td>
<td>0.582</td>
<td>14q13.3</td>
<td>37138294</td>
<td>PAX9</td>
<td>C/G</td>
<td>0.58</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>2.3E-12</td>
</tr>
<tr>
<td>rs4924487</td>
<td>0.836</td>
<td>15q15.1</td>
<td>40922915</td>
<td>CASC5</td>
<td>C/G</td>
<td>0.81</td>
<td>1.06</td>
<td>1.04-1.09</td>
<td>1.3E-08</td>
</tr>
<tr>
<td>rs33984059</td>
<td>0.982</td>
<td>15q21.3</td>
<td>56385868</td>
<td>RFX7</td>
<td>A/G</td>
<td>0.98</td>
<td>1.19</td>
<td>1.12-1.27</td>
<td>1.1E-08</td>
</tr>
<tr>
<td>rs112293876</td>
<td>0.280</td>
<td>15q22.31</td>
<td>66764641</td>
<td>MAP2K1</td>
<td>C/CA</td>
<td>0.29</td>
<td>1.06</td>
<td>1.04-1.08</td>
<td>3.5E-10</td>
</tr>
<tr>
<td>rs11863709</td>
<td>0.945</td>
<td>16q21</td>
<td>57654576</td>
<td>GPR56</td>
<td>C/T</td>
<td>0.96</td>
<td>1.16</td>
<td>1.11-1.21</td>
<td>1.8E-11</td>
</tr>
<tr>
<td>rs201158093</td>
<td>0.435</td>
<td>16q23.3</td>
<td>82178893</td>
<td>RP11-510J16.5</td>
<td>TAA/TA</td>
<td>0.44</td>
<td>1.05</td>
<td>1.03-1.07</td>
<td>9.1E-09</td>
</tr>
<tr>
<td>rs28441558</td>
<td>0.050</td>
<td>17p13.1</td>
<td>7803118</td>
<td>CHD3</td>
<td>C/T</td>
<td>0.05</td>
<td>1.16</td>
<td>1.12-1.20</td>
<td>1.0E-16</td>
</tr>
<tr>
<td>rs142444269</td>
<td>0.798</td>
<td>17q11.2</td>
<td>30098749</td>
<td>RP11-805L22.3</td>
<td>C/T</td>
<td>0.78</td>
<td>1.07</td>
<td>1.05-1.09</td>
<td>3.2E-10</td>
</tr>
<tr>
<td>SNP</td>
<td>Risk allele freq</td>
<td>Chromosome</td>
<td>Position</td>
<td>Gene/Region</td>
<td>Risk Allele</td>
<td>Reference Allele</td>
<td>OR</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------------</td>
<td>------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>rs2680708</td>
<td>0.623</td>
<td>17q22</td>
<td>56456120</td>
<td>RNF43</td>
<td>G/A</td>
<td></td>
<td>0.61</td>
<td>1.05</td>
<td>1.03-1.06</td>
</tr>
<tr>
<td>rs8093601</td>
<td>0.459</td>
<td>18q21.2</td>
<td>51772473</td>
<td>MBD2</td>
<td>C/G</td>
<td></td>
<td>0.44</td>
<td>1.05</td>
<td>1.03-1.06</td>
</tr>
<tr>
<td>rs28607662</td>
<td>0.085</td>
<td>18q21.2</td>
<td>53230859</td>
<td>TCF4</td>
<td>C/T</td>
<td></td>
<td>0.10</td>
<td>1.08</td>
<td>1.05-1.11</td>
</tr>
<tr>
<td>rs12956892</td>
<td>0.300</td>
<td>18q21.32</td>
<td>56746315</td>
<td>OACYLP</td>
<td>T/G</td>
<td></td>
<td>0.30</td>
<td>1.05</td>
<td>1.03-1.07</td>
</tr>
<tr>
<td>rs533722308</td>
<td>0.390</td>
<td>18q21.33</td>
<td>60961193</td>
<td>BCL2</td>
<td>CT/C</td>
<td></td>
<td>0.42</td>
<td>1.05</td>
<td>1.03-1.07</td>
</tr>
<tr>
<td>rs10460109</td>
<td>0.414</td>
<td>18q22.3</td>
<td>73036165</td>
<td>TSHZ1</td>
<td>T/C</td>
<td></td>
<td>0.42</td>
<td>1.05</td>
<td>1.03-1.06</td>
</tr>
<tr>
<td>rs11666569</td>
<td>0.728</td>
<td>19p13.11</td>
<td>17214073</td>
<td>MYO9B</td>
<td>C/T</td>
<td></td>
<td>0.71</td>
<td>1.05</td>
<td>1.03-1.07</td>
</tr>
<tr>
<td>rs118005503</td>
<td>0.912</td>
<td>19q12</td>
<td>32167803</td>
<td>THEG5</td>
<td>G/C</td>
<td></td>
<td>0.91</td>
<td>1.09</td>
<td>1.06-1.13</td>
</tr>
<tr>
<td>rs61088131</td>
<td>0.848</td>
<td>19q13.2</td>
<td>42700947</td>
<td>POU2F2</td>
<td>T/C</td>
<td></td>
<td>0.82</td>
<td>1.06</td>
<td>1.04-1.09</td>
</tr>
<tr>
<td>rs6091758</td>
<td>0.465</td>
<td>20q11.21</td>
<td>31347512</td>
<td>DNMT3B</td>
<td>C/CA</td>
<td></td>
<td>0.60</td>
<td>1.05</td>
<td>1.03-1.06</td>
</tr>
<tr>
<td>rs9625483</td>
<td>0.026</td>
<td>22q12.1</td>
<td>28888939</td>
<td>TTC28</td>
<td>A/G</td>
<td></td>
<td>0.03</td>
<td>1.14</td>
<td>1.09-1.20</td>
</tr>
<tr>
<td>rs17321482</td>
<td>0.873</td>
<td>23p22.2</td>
<td>11482634</td>
<td>ARHGAP6</td>
<td>C/T</td>
<td></td>
<td>0.87</td>
<td>1.07</td>
<td>1.05-1.09</td>
</tr>
</tbody>
</table>

Novel loci associated with early-onset

<table>
<thead>
<tr>
<th>SNP</th>
<th>Risk allele freq</th>
<th>Chromosome</th>
<th>Position</th>
<th>Gene/Region</th>
<th>Risk Allele</th>
<th>Reference Allele</th>
<th>OR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs138004030</td>
<td>0.920</td>
<td>6q27</td>
<td>170475879</td>
<td>LOC154449</td>
<td>G/A</td>
<td></td>
<td>0.91</td>
<td>1.27</td>
<td>1.17-1.38</td>
</tr>
</tbody>
</table>

a Risk allele frequency - 1000 Genomes Project Europeans  
b Risk allele/Reference allele  
c Risk allele frequency  
d Odds ratio and confidence interval  
e Confidence interval  
f P-values are generated from a likelihood ratio test  
g Region previously reported by Wang et al (Nat Comm 2015), rs12791447; rs61890184-rs12791447 r2 (EUR)=0.41
Table 2. Polygenic Risk Score (PRS) estimation using 147 prostate cancer susceptibility variants.

<table>
<thead>
<tr>
<th>Risk Category Percentiles</th>
<th>RR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1%</td>
<td>0.15</td>
<td>0.11-0.20</td>
</tr>
<tr>
<td>1-10%</td>
<td>0.35</td>
<td>0.32-0.37</td>
</tr>
<tr>
<td>10-25%</td>
<td>0.54</td>
<td>0.51-0.57</td>
</tr>
<tr>
<td>25-75%</td>
<td>1.00</td>
<td>(Baseline)</td>
</tr>
<tr>
<td>75-90%</td>
<td>1.74</td>
<td>1.67-1.82</td>
</tr>
<tr>
<td>90-99%</td>
<td>2.69</td>
<td>2.55-2.82</td>
</tr>
<tr>
<td>≥99%</td>
<td>5.71</td>
<td>5.04-6.48</td>
</tr>
</tbody>
</table>

*Polygenetic Risk Score (PRS) percentiles based on the cumulative score distributed among controls. The beta coefficients computed from the European overall meta-analysis was applied to determine the PRS risk among individuals in the OncoArray study.*

*Relative risk and 95% confidence intervals.*
**Previous GWAS Meta-Analysis - 2014**

- **BPC3**
  - 2,068 cases
  - 2,993 controls
  - Illumina Infinium HumanHap
  - 525,766 SNPs

- **ICOGS**
  - 20,219 cases
  - 20,440 controls
  - Illumina Infinium (Custom)
  - 201,598 SNPs

- **CAPS 1**
  - 474 cases
  - 482 controls
  - Affymetrix GeneChip
  - 369,025 SNPs

- **CAPS 2**
  - 1,458 cases
  - 512 controls
  - Affymetrix GeneChip
  - 369,610 SNPs

**UK Stage 1**

- **Pegasus**
  - 4,600 cases
  - 2,941 controls
  - Illumina Omni HumanHap
  - 1,675,706 SNPs

- **UK Stage 2**
  - 3,650 cases
  - 3,940 controls
  - Illumina ISELECT
  - 43,671 SNPs

**23 Novel SNPs**

- AA Olama et al.
  - Nature Genetics
  - 2014

**Current GWAS Meta-Analysis - 2017**

- **ELLIPSE OncoArray**
  - 46,939 cases / 27,910 controls
  - Illumina Infinium (Custom)
  - 498,417 SNPs

**63 Novel SNPs**
ONLINE METHODS:

Study subjects

A brief overview and study details for participating prostate cancer (PrCa) studies in the newly genotyped OncoArray project are provided in Supplementary Table 1 for men of European ancestry. All studies were approved by the appropriate ethics committees and informed consent was obtained from all participants. Supplementary Table 2 summarizes the PrCa sample series of the Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) consortia contributing both newly obtained genotyping data for the OncoArray and previous genome-wide association studies (GWAS). The majority of the studies contributing to the OncoArray were case-control studies primarily based in either the United States or Europe. In total 52 new studies provided core data on disease status, age at diagnosis (age at observation or questionnaire for controls), family history of PrCa, and clinical factors for cases (e.g. PSA at diagnosis, Gleason score, etc.) for 48,455 PrCa cases and 28,321 disease-free controls. Previous GWAS contributed an additional 32,255 PrCa cases and 33,202 disease-free controls of European ancestry for the overall meta-analysis. Supplementary Table 3 provides quality control information by consortia (i.e. OncoArray project, UK GWAS, etc) for both samples and SNPs. After removing all overlapping samples the OncoArray contribution for newly genotyped samples was 46,939 PrCa cases and 27,910 disease-free controls.

Several strata-specific analyses were implemented to evaluate the impact of genetic variation in PrCa disease aggressiveness. Supplementary Table 4 describes the analysis title, outcome and reference groups, and the statistical model used. Several classification schemes (i.e. low aggressiveness, intermediate aggressiveness, etc.) were implemented to better assess the spectrum of genetic involvement. All classification schemes incorporated the diagnostic clinical features PSA, tumor stage and Gleason score. In order to compare to previous PrCa aggressive analyses by our research group, we included the ‘Advanced (plus death due to PrCa)’ classification. Contributing study groups missing clinical features were excluded (Supplementary Table 2). Individuals with missing or granular clinical information were excluded. The strata-specific sample sizes by PrCa GWAS consortium are
provided in **Supplementary Table 5**. Furthermore we analysed Gleason score as a continuous variable.

**OncoArray SNP selection**

The NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) consortia (http://epi.grants.cancer.gov/gameon/) provided SNPs to be included on the Illumina OncoArray. Approximately 50% of the OncoArray was a compilation of SNP lists by the GAME-ON disease consortia of cancer (breast, colorectal, lung, ovarian, and prostate), a common set of variants for common risk regions, other related traits (i.e. BMI, age at menarche, etc), pharmacogenetics, and candidates. The remaining content of the OncoArray was selected as a “GWAS backbone” (Illumina HumanCore), which aimed to provide high coverage for the majority of common variants through imputation. Approximately 79k SNPs were selected specifically for their relevance to PrCa, based on prior evidence of association with overall or subtype-specific disease, fine-mapping of known PrCa regions, and candidate submissions (i.e. survival, exome sequencing, etc). In order to maximize efficiency of the array, cancer-specific candidate lists were merged to remove redundant genetic variation.

**Genotype calling and quality control**

Details of the genotype calling and quality control (QC) for the iCOGS and GWAS are described elsewhere.

Of the 568,712 variants selected for genotyping on OncoArray, 533,631 were successfully manufactured on the array (including 778 duplicate probes). OncoArray genotyping of ELLIPSE studies was conducted at five sites (Cambridge [UK], CIDR, Copenhagen, USC, NCI). Details of the genotyping calling for the OncoArray are described in more detail elsewhere. Briefly, we developed a single calling pipeline that was applied to more than 500,000 samples across the GAME-ON consortia. An initial cluster file was generated using 56,284 samples selected from all major genotyping centers and ethnicities, using the Gentrain2 algorithm. Variants likely to have problematic clusters were selected for manual inspection using the following criteria: call rate below 99%, minor allele frequency (MAF) <0.001, poor Illumina intensity and clustering.
metrics, deviation from the MAF observed in the 1000 Genomes Project (1KGP) using the criterion: 
\[ \frac{(|p_1 - p_0| - 0.01)^2}{(p_1 + p_0)(2 - p_1 - p_0)} > C, \]
where \( p_0 \) and \( p_1 \) are the minor frequencies in the 1KGP and OncoArray datasets, respectively, and \( C = 0.008 \). This resulted in manual adjustment of the cluster file for 3,964 variants, and the exclusion of 16,526 variants. The final cluster file was then applied to the full dataset.

Our quality control pipeline for ELLIPSE excluded SNPs with a call rate <95% by study, not in Hardy-Weinberg equilibrium (\( P < 10^{-7} \) in controls, or \( P < 10^{-12} \) in cases) or with concordance <98% among 11,260 duplicate pairs. In order to minimize imputation errors, we additionally excluded SNPs with a MAF<1% and a call rate <98% in any study, SNPs that could not be linked to the 1KGP reference, those with MAF for Europeans that differed from that for the 1KGP and a further 16,526 SNPs where the cluster plot was judged to be not ideal. Of the 533,631 manufactured SNPs on the OncoArray, we retained 498,417 SNPs among our samples of European ancestry following QC.

We excluded duplicate samples and first-degree relatives within each study, duplicates across studies, samples with a call rate <95%, and samples with extreme heterozygosity (>4.9 standard deviations from the mean for the reported ethnicity). We excluded duplicated samples as well as first-degree relatives across the GWAS studies CAPS1, CAPS2, UK Stage 1, UK Stage 2, and iCOGS. Duplicate and first-degree related samples were assessed across the BPC3 and Pegasus GWAS studies as well. Ancestry was computed using a principal component analysis using 2,318 informative markers on a subset of ~47,000 samples and projected onto the complete OncoArray dataset. The current analysis was restricted to men of European ancestry, defined as individuals with an estimated proportion of European ancestry >0.8, with reference to the HapMap populations, based on the first two principal components. Of the 78,182 samples genotyped (regardless of race/ethnicity), the final dataset consisted of 74,849 samples, of which 46,939 PrCa cases and 27,910 disease-free controls (Supplementary Table 3) after excluding overlap samples, were meta-analysed with previous studies.
Imputation

Genotypes for ~70M SNPs were imputed for all samples using the October 2014 (Phase 3) release of the 1KGP data as the reference panel. The OncoArray and GWAS datasets were imputed using a two-stage imputation approach, using SHAPEIT\textsuperscript{21} for phasing and IMPUTE\textsuperscript{22} for imputation. The imputation was performed in 5Mb non-overlapping intervals. All subjects were split into subsets of ~10,000 samples, with subjects from the same group in the subset. We imputed genotypes for all SNPs that were polymorphic (MAF>0.1%) in European samples. We excluded data for all monomorphic SNPs and those with an imputation \( r^2 < 0.3 \) leaving a total of 20,370,935 SNP across chromosomes 1-22 and chromosome X. Of the SNPs imputed, 49.3% had a MAF<1%, 15.2% had a MAF ranging between 1-5%, and 35.5% had a MAF≥5%.

Statistical analyses

Per-allele odds ratios and standard errors were generated for the OncoArray and each GWAS, adjusting for principal components and study relevant covariates using logistic regression. The OncoArray and iCOGS analyses were additionally stratified by country and study, respectively. We used the first seven principal components in our analysis of individuals of European ancestry, as additional components did not further reduce inflation in the test statistics.

Odds ratio (OR) estimates were derived using either SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) or an in-house C++ program (Supplementary Table 3). OR estimates and standard errors were combined by a fixed effects inverse variance meta-analysis using METAL\textsuperscript{23}. All statistical tests conducted were two-sided.

Our analyses included overall PrCa and several clinically relevant strata. These included: 1) high vs low aggressive PrCa; 2) high vs low/intermediate aggressive PrCa; 3) advanced vs non-advanced PrCa; 4) advanced PrCa vs controls; 5) early-onset PrCa (≤55 yrs) vs controls; and 6) Gleason score (Supplementary
Tables 4 & 5). We defined low aggressive as tumor stage ≤T1 and Gleason ≤6 and PSA <10 ng/mL, intermediate aggressive as tumor stage T2 or Gleason=7 or PSA 10-20 ng/mL, high aggressive as tumor stage T3/T4 or N1 or M1 or Gleason ≥8 or PSA >20 ng/mL, and advanced as either metastatic disease, Gleason ≥8, PSA>100 or PrCa-related deaths (Supplementary Table 4).

Definition of new hits
To search for novel loci, we assessed all SNPs excluding those within a known PrCa locus, defined by current fine-mapping assessments (Supplementary Table 7). SNPs that were associated with disease risk at $P<5 \times 10^{-8}$ in the meta-analysis (GWAS and OncoArray) were considered novel. The SNP with the lowest p-value in a region was considered the lead SNP. Imputation quality assessed by IMPUTE2 imputation $r^2$ in the OncoArray dataset (Supplementary Table 8).

For ten regions where the newly identified locus was near a previously known region, we reported a novel association if the pairwise $r^2$ between the new and the previously known SNP was less than 0.2. For novel PrCa associations where the variant was imputed in the OncoArray study samples series and had an imputed quality score less than 0.70, we assessed the quality of the imputation by masking the variant in a subset of the 1KGP European sample and calculating the concordance following re-imputation in the remaining 1KGP samples.

Reliability of Imputation
Novel SNPs with an IMPUTE2 $r^2<0.80$ among the OncoArray sample series (Supplementary Table 8) were flagged for further investigation to minimize the probability of a false positive. First, we examined linkage disequilibrium (LD) plots (http://locuszoom.org/) for poorly imputed SNPs (+/-500kb) including only genotyped SNPs within the region. The imputed index SNP was included in the plot to determine the strength of LD with nearby signals and assess a pattern of association. Furthermore, we performed an imputation experiment using the
We evaluated four SNPs where the IMPUTE2 $r^2$ was less than 0.80 in the OncoArray sample series: rs527510716 (Chr 7), rs6602880 (Chr 10), rs533722308 (Chr 18) and rs144166867 (Chr X). Supplementary Figure 3 includes the LD plots for three of the poorly imputed SNPs. The variant rs144166867 (Chr X) could not be plotted given no genotype SNPs were available +/- 500 KB on the OncoArray. Both LD plots for markers rs527510716 (Chr 7) and rs533722308 (Chr 18) showed significant associations ($P<1\times10^{-3}$) for several genotype markers with moderate LD of the index SNP. The Kappa coefficient for markers rs527510716 (Chr 7) and rs533722308 (Chr 18) was 0.911 and 0.931, respectively (Supplementary Table 9). The marker rs6602880 (Chr 10) had a Kappa coefficient of 0.812 and was the only significant variant in the LD plot. The Kappa coefficient for marker rs144166867 (Chr X) was 0.665 (Supplementary Table 9). The markers rs6602880 (Chr 10) and rs144166867 (Chr X) are most likely false positives due to poor imputation for these regions.

**Proportion of familial risk explained**

The contribution of the known SNPs to the familial risk of PrCa, under a multiplicative model, was computed using the formula

$$\frac{\left(\log \sum_{k} \ldots \right)}{\left(\log \ldots\right)}$$
where $\lambda_0$ is the observed familial risk to first degree relatives of PrCa cases$^{24,25}$, assumed to be 2.5, and $\lambda_k$ is the familial relative risk due to locus $k$, given by:

$$
\lambda_k = \frac{p_k r_k^2 + q_k}{(p_k r_k + q_k)^2}
$$

where $p_k$ is the frequency of the risk allele for locus $k$, $q_k = 1 - p_k$ and $r_k$ is the estimated per-allele odds ratio.

Based on the assumption of a log-additive model, we constructed a polygenic risk score (PRS) from the summed risk allelic dosages weighted by the per-allele log-odds ratios. Thus for each individual $j$ we derived:

$$
Score_j = \sum_{i=1}^{N} \beta_i g_{ij}
$$

Where:

- $N$: Number of SNPs
- $g_{ij}$: Allele dose at SNP $i$ for individual $j$
- $\beta_i$: Per-allele log-odds ratio of SNP $i$

The risk of PrCa was estimated for the percentile of the distribution of the PRS (<1%, 1-10%, 10-25%, 25-75%, 75-90%, 90-99%, >99% and <10%, 10-25%, 25-75%, 75-90%, >90%) where cumulative score thresholds were determined by the observed distribution among controls. We applied effect sizes and allele frequencies obtained from the overall meta-analysis of Europeans to estimate risk scores for individuals of European ancestry in the OncoArray study$^{26}$. A standardized PRS score was calculated by dividing the observed PRS score by the standard deviation of the PRS score among controls. A logistic regression framework was used to evaluate the percentile comparisons and determine the risk estimate. The models were adjusted for the first seven principal components to account for population stratification and stratified by country.

The FRR and PRS risk estimation was limited to the variants where our overall meta-analysis observed a statistically significant association. In total, we included
147 PrCa index SNPs in our risk score modelling, including 85 previously published associations and the 62 novel findings reported here. To correct for potential bias in effect estimation of newly discovered variants, we implemented a fully Bayesian version of a weighted correction given in Zhong and Prentice, Eq 3.4\textsuperscript{27}. Specifically, we place a normal prior distribution on MLE effect estimates of the form $\beta_m \sim N(\beta_{cor}, \tau^2)$. Here, $\beta_m$ is the log odds ratio from the overall meta-analysis; $\beta_{cor}$ is the bias corrected estimate calculated using the expectation-adjusted estimator from Eq 3.1 in Zhong and Prentice; and $\tau$ is a pre-specified variance of the effect distribution reflecting the bias and is defined as $\tau = |\hat{\beta}_m - \beta_{cor}|$.

**eQTL analyses**

Genotype and gene expression data were downloaded from The Cancer Genome Atlas (TCGA) for 494 samples with PrCa (https://gdc-portal.nci.nih.gov). Quality Control (QC) was performed on both these datasets as follows: on the genotype, we filtered out samples with high heterozygosity (mean heterozygosity +/- 2 standard deviation) and missing genotypes, duplicated or related samples. We then performed Principal Component Analysis on the 494 samples plus 2,506 samples from 1KGP to infer the ancestry of the TCGA samples; samples of non-European ancestry were removed. We also filtered out variants with missing call rate > 5%. For the expression data, samples from two plates had, on average, much higher expression values than the remaining samples, and these were excluded. We also filtered genes with mean expression across samples <= 6 counts. Finally, expression values were quantile-normalized by samples and rank-transformed by genes. After QC we used the data from 359 samples. For the eQTL analysis, 35 PEER factors from the top 10,000 expressed genes were used as covariates, plus three genotyping PCs (which explained 18% of total variation). eQTL analysis was performed using FastQTL with 1,000 permutations over the 85 regions. We used a window of 1 Megabases (upstream/downstream) from the transcription start site (TSS) of each gene.
Gene Set Enrichment Analyses

The file Human_GOBP_AllPathways_no_GO_iea_September_01_2016_symbol.gmt (http://baderlab.org/EM_GeneSets), from the GeneSets database, was used for all analyses. This database contains pathways from Reactome, NCI Pathway Interaction Database, GO (Gene Ontology) biological process, HumanCyc, MSigdb, NetPath and Panther. We manually corrected several pathways where the PDK1 gene was entered as PDPK1. GO pathways inferred from electronic annotation terms were excluded. The same pathway (e.g. apoptosis) may be defined in two or more databases with potentially different sets of genes, and all versions of these duplicate/overlapping pathways were included. Pathway size was determined by the total number of genes in the pathway to which SNPs in the imputed GWAS dataset could be mapped. To provide more biologically meaningful results, and reduce false positives, only pathways that contained between 10 and 200 genes were considered.

Gene information (hg19) was downloaded from the ANNOVAR website (http://www.openbioinformatics.org/annovar/). SNPs were mapped to the nearest gene within 500kb window; those that were further away from any gene were excluded. Gene significance was calculated by assigning the lowest p-value observed across all SNPs assigned to a gene, based on the combined European meta-analysis (previous GWAS and OncoArray).

The gene set enrichment analysis (GSEA) algorithm, as implemented in the GenGen package (http://gengen.openbioinformatics.org/en/latest/) was used to perform pathway analysis. Briefly, the algorithm calculates an enrichment score (ES) for each pathway based on a weighted Kolmogorov-Smirnov statistic. To calculate the ES we performed 100 permutations and averaged the final score. Pathways that have most of their genes at the top of the ranked list of genes obtain higher ES values. Only pathways with positive ES and at least one gene with P<5x10^{-8} were retained for subsequent analysis. An enrichment map
was created using the Enrichment Map (EM) v 2.1.0 app in Cytoscape v3.40, applying force directed layout, weighted mode. We restricted our pathway analysis to those with an ES ≥ 0.50 to ensure a true positive rate > 0.20 and a false positive rate < 0.15.
DATA AVAILABILITY:
The OncoArray genotype data and relevant covariate information (i.e. ethnicity, country, principal components, etc.) generated during this study will be deposited into dbGAP for access. In total 47 of the 52 OncoArray studies encompassing nearly 90% of the individual samples will be available (Supplementary Table 19). The previous meta-analysis summary results and genotype data currently are available in dbGAP (Accession #: phs001081.v1.p1). The complete meta-analysis summary associations statistics is publicly available at the PRACTICAL website (http://practical.icr.ac.uk/blog/).


