

Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer

UKGPCS collaborators (2021). Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. Prostate Cancer and Prostatic Diseases. Advance online publication. https://doi.org/10.1038/s41391-020-00311-2

Published in:

Prostate Cancer and Prostatic Diseases

Document Version: Peer reviewed version

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

Publisher rights Copyright 2021 Springer. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

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

Take down policy

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

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. - Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer

Roshan A. Karunamuni¹, Minh-Phuong Huynh-Le², Chun C. Fan³, Wesley Thompson⁴, Rosalind

- A. Eeles⁵, Zsofia Kote-Jarai⁵, Kenneth Muir⁶, Artitaya Lophatananon⁶, UKGPCS collaborators⁷, Johanna Schleutker⁸, Nora Pashayan⁹, Jyotsna Batra¹⁰, APCB BioResource (Australian Prostate Cancer BioResource)¹¹, Henrik Grönberg¹², Eleanor I. Walsh¹³, Emma L. Turner¹³, Athene Lane^{13,14}, Richard M. Martin^{13,14,15}, David E. Neal¹⁶, Jenny L. Donovan¹⁷, Freddie C. Hamdy^{16,18}, Børge G. Nordestgaard¹⁹, Catherine M. Tangen²⁰, Robert J. MacInnis^{21,22}, Alicja
- 10 Wolk²³, Demetrius Albanes²⁴, Christopher A. Haiman²⁵, Ruth C. Travis²⁶, Janet L. Stanford²⁷, Lorelei A. Mucci²⁸, Catharine M. L. West²⁹, Sune F. Nielsen³⁰, Adam S. Kibel³¹, Fredrik Wiklund¹², Olivier Cussenot³², Sonja I. Berndt²⁴, Stella Koutros²⁴, Karina Dalsgaard Sørensen³³, Cezary Cybulski³⁴, Eli Marie Grindedal³⁵, Jong Y. Park³⁶, Sue A. Ingles³⁷, Christiane Maier³⁸, Robert J. Hamilton³⁹, Barry S. Rosenstein⁴⁰, Ana Vega⁴¹, The IMPACT Study Steering
- 15 Committee and Collaborators⁴², Manolis Kogevinas⁴³, Kathryn L. Penney⁴⁴, Manuel R. Teixeira⁴⁵, Hermann Brenner⁴⁶, Esther M. John⁴⁷, Radka Kaneva⁴⁸, Christopher J. Logothetis⁴⁹, Susan L. Neuhausen⁵⁰, Azad Razack⁵¹, Lisa F. Newcomb²⁷, Canary PASS Investigators²⁷, Marija Gamulin⁵², Nawaid Usmani⁵³, Frank Claessens⁵⁴, Manuela Gago-Dominguez⁵⁵, Paul A. Townsend⁵⁶, Monique J. Roobol⁵⁷, Wei Zheng⁵⁸, The Profile Study Steering Committee⁵⁹, Ian G.
- Mills⁶⁰, Ole A. Andreassen⁶¹, Anders M. Dale⁶², Tyler M. Seibert^{1,62,63}, The PRACTICAL
 Consortium⁶⁴

¹Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, CA, USA

²Radiation Oncology, George Washington University, Washington, DC

³Center for Human Development, University of California San Diego, La Jolla, CA, USA
 ⁴Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA

⁵The Institute of Cancer Research, London, SM2 5NG, UK

⁶Division of Population Health, Health Services Research and Primary Care, University of 30 Manchester, Oxford Road, Manchester, M13 9PL, UK ⁷http://www.icr.ac.uk/our-research/research-divisions/division-of-genetics-andepidemiology/oncogenetics/research-projects/ukgpcs/ukgpcs-collaborators

⁸Institute of Biomedicine, University of Turku, Finland

⁹Department of Applied Health Research, University College London, London, WC1E 7HB, UK

35 ¹⁰Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane QLD 4059, Australia

¹¹Australian 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

¹²Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171
 77 Stockholm, Sweden

¹³Bristol Medical School, Department of Population Health Sciences, University of Bristol, Bristol, United Kingdom

¹⁴MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom

- ¹⁵National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, United Kingdom ¹⁶Nuffield Department of Surgical Sciences, University of Oxford, Room 6603, Level 6, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK ¹⁷School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom
- ¹⁵ ¹⁸Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom

¹⁹Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

²⁰SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

20 ²¹Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Road, Melbourne, VIC 3004, Australia

²²Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health,
 The University of Melbourne, Grattan Street, Parkville, VIC 3010, Australia
 ²³Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine,

 Karolinska Institutet, SE-171 77 Stockholm, Sweden
 ²⁴Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, 20892, USA
 ²⁵Center for Canadia Epidemiology, Department of Preventive Medicine, Keck School of

²⁵Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles,

30 CA 90015, USA

²⁶Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK

²⁷Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109-1024, USA

35 ²⁸Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

²⁹Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester, M13 9PL UK

40 ³⁰Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2200 Copenhagen, Denmark

³¹Division of Urologic Surgery, Brigham and Womens Hospital, 75 Francis Street, Boston, MA 02115, USA

³²Sorbonne Universite, GRC n°5, AP-HP, Tenon Hospital, 4 rue de la Chine, F-75020 Paris,

France

5

³³Department of Molecular Medicine, Aarhus University Hospital, Palle Juul-Jensen Boulevard 99, 8200 Aarhus N, Denmark

³⁴International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-115 Szczecin, Poland

³⁵Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway
 ³⁶Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA

³⁷Department of Preventive Medicine, Keck School of Medicine, University of Southern

 California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA
 ³⁸Humangenetik Tuebingen, Paul-Ehrlich-Str 23, D-72076 Tuebingen, Germany
 ³⁹Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto ON M5G 2M9, Canada

⁴⁰Department of Radiation Oncology and Department of Genetics and Genomic Sciences, Box

15 1236, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

⁴¹Fundación Pública Galega Medicina Xenómica, Santiago de Compostela, 15706, Spain. ⁴²http://impact.icr.ac.uk

⁴³ISGlobal, Barcelona, Spain

 ⁴⁴Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02184, USA
 ⁴⁵Department of Genetics, Portuguese Oncology Institute of Porto (IPO-Porto), 4200-072 Porto, Portugal

⁴⁶Division of Clinical Epidemiology and Aging Research, German Cancer Research Center

- (DKFZ), D-69120, Heidelberg, Germany
 ⁴⁷Departments of Epidemiology & Population Health and of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94304 USA
 ⁴⁸Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria
- ⁴⁹The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, 1515 Holcombe Blvd., Houston, TX 77030, USA
 ⁵⁰Department of Population Sciences, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte, CA 91010, 626-256-HOPE (4673)
 ⁵¹Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur,
- 35 Malaysia ⁵²Division of Medical Oncology, Urogenital Unit, Department of Oncology, University Hospital Centre Zagreb, University of Zagreb, School of Medicine, 10 000 Zagreb, Croatia ⁵³Department of Oncology, Cross Cancer Institute, University of Alberta, 11560 University Avenue, Edmonton, Alberta, Canada T6G 1Z2
- 40 ⁵⁴Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, BE-3000, Belgium

⁵⁵Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, 15706, Santiago de Compostela, Spai ⁵⁶Division 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

5 ⁵⁷Department of Urology, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands

⁵⁸Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN 37232 USA.

- ⁵⁹http://www.cancerresearchuk.org/about-cancer/find-a-clinical-trial/a-study-find-out-looking gene-changes-would-be-useful-in-screening-for-prostate-cancer-profile-pilot
 ⁶⁰Center for Cancer Research and Cell Biology, Queen's University of Belfast, Belfast, UK
 ⁶¹NORMENT, KG Jebsen Centre, Oslo University Hospital and University of Oslo, Oslo, Norway
 ⁶²Department of Radiology, University of California San Diego, La Jolla, CA, USA
 ⁶³Department of Bioengineering, University of California San Diego, La Jolla, CA, USA
- ⁶⁴Institute of Cancer Research, Sutton, SW7 3RPm UK

* Additional members from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium (PRACTICAL,http://practical.icr.ac.uk/) and

20 associated funding are provided in the Supplementary Data Appendix 1 and 2.

Corresponding Authors:

Roshan Karunamuni, PhD

University of California, San Diego

25 Department of Radiation Medicine and Applied Sciences
 3960 Health Sciences Dr, Mail Code 0865
 La Jolla, CA 92093
 rakarunamuni@health.ucsd.edu

- Tyler M. Seibert, MD, PhD
 University of California, San Diego
 Department of Radiation Medicine and Applied Sciences
 9500 Gilman Dr, Mail Code 0861
 La Jolla, CA 92093
- 35 <u>tseibert@ucsd.edu</u>

Running Title

Adding SNPs improves prostate cancer polygenic score

Funding

This study was funded in part by a grant from the United States National Institute of

5 Health/National Institute of Biomedical Imaging and Bioengineering (#K08EB026503), the University of California Cancer Research Coordinating Committee (#C21CR2060), the Research Council of Norway (#223273), KG Jebsen Stiftelsen, and South East Norway Health Authority.

Abstract

<u>Background:</u> Polygenic hazard scores (PHS) can identify individuals with increased risk of prostate cancer. We estimated the benefit of additional SNPs on performance of a previously validated PHS (PHS46).

- 5 <u>Materials and Method:</u> 180 SNPs, shown to be previously associated with prostate cancer, were used to develop a PHS model in men with European ancestry. A machine-learning approach, LASSO-regularized Cox regression, was used to select SNPs and to estimate their coefficients in the training set (75,596 men). Performance of the resulting model was evaluated in the testing/validation set (6,411 men) with two metrics: (1) hazard ratios (HRs) and (2) positive
- 10 predictive value (PPV) of prostate-specific antigen (PSA) testing. HRs were estimated between individuals with PHS in the top 5% to those in the middle 40% (HR95/50), top 20% to bottom 20% (HR80/20), and bottom 20% to middle 40% (HR20/50). PPV was calculated for the top 20% (PPV80) and top 5% (PPV95) of PHS as the fraction of individuals with elevated PSA that were diagnosed with clinically significant prostate cancer on biopsy.
- <u>Results:</u> 166 SNPs had non-zero coefficients in the Cox model (PHS166). All HR metrics showed significant improvements for PHS166 compared to PHS46: HR95/50 increased from 3.72 to 5.09, HR80/20 increased from 6.12 to 9.45, and HR20/50 decreased from 0.41 to 0.34. By contrast, no significant differences were observed in PPV of PSA testing for clinically significant prostate cancer.
- 20 <u>Conclusion:</u> Incorporating 120 additional SNPs (PHS166 vs PHS46) significantly improved HRs for prostate cancer, while PPV of PSA testing remained the same.

Introduction

Optimal prostate cancer screening strategies seek to strike a balance between identifying clinically significant and potentially lethal cases that require treatment, while minimizing overdiagnosis of indolent, lower-risk cases that do not need radical treatment^{1–3}.

- 5 Genetic risk models have emerged as potentially useful tools that identify individuals with greater risk for being diagnosied with prostate cancer^{4,5}, and so help inform if and when to initiate screening for an individual. A subset of these models called polygenic hazard scores (PHS) seeks to directly identify associations between common genetic variants and the age of diagnosis of prostate cancer by utilizing the framework of time-to-event analyses^{1,6}.
- 10 We have previously reported on a PHS model for prostate cancer, PHS46, that demonstrated excellent performance in an independent test set of men from varied genetic ancestries⁶. The model incorporates genetic data of 46 unique single nucleotide polymorphisms (SNPs), and was identified through a systematic search of European men genotyped on the iCOGS chipset (Illumina, San Diego, CA). With an ever-increasing list of loci associated with prostate cancer in the literature^{7–9}, we sought to determine what effect, if any, the incorporation of additional SNPs would have on the performance of PHS46.

To this end, we employed a machine-learning approach, LASSO-regularized Cox regression,^{10,11} to select SNPs from a list that included the 46 used in PHS46, as well as over 100 SNPs identified in previous analyses as having genome-wide significance for association

- 20 with prostate cancer⁷. LASSO-regularized regression is an established variable selection technique in datasets with a large number of predictors and has been previously implemented as a SNP selection tool for a breast cancer polygenic risk score¹². Performance metrics describing statistical model goodness-of-fit and clinically actionable screening utility of the LASSO-regularized PHS model for prostate cancer were compared with those achieved with
- 25 PHS46 to determine the potential benefit of incorporating additional SNPs in polygenic hazard models.

Material and Methods

Study dataset

We obtained genotype and phenotype data from the PRACTICAL¹³ consortium for this analysis. Genotyping was performed previously on either OncoArray¹³ or iCOGS⁹ chips, and

- 5 these data were previously imputed using the 1000 Genomes reference panel¹⁴. Missing SNP calls were replaced with the mean of the genotyped data for that SNP in the training set^{1,15}. In total, data from 82,007 men with European genetic ancestry (Supplementary Table 1)^{13,16} were available for this analysis. A testing set consisting of 6,411 men (4,828 controls and 1,583 cases) enrolled in the ProtecT clinical trial was set aside for estimating the performance of the
- 10 final PHS models. The data from ProtecT were chosen as the testing set because they are well characterized and were previously used for validation of PHS46¹, allowing us to directly benchmark the performance of the updated model against previous iterations. The ProtecT trial also included biopsies of participants with elevated prostate-specific antigen (PSA) level, which permits analysis of the positive predictive value of the current clinical standard for screening,
- 15 PSA testing. The remaining 75,596 individuals (25,127 controls and 50,469 cases) were used for training of the model. This first analysis was limited to men of European descent because of much greater data availability in that population, but our previous work has shown that development in Europeans can inform careful future work to assess and improve performance in other ancestries¹⁷.
- 20

Model development using LASSO regularization

A list of published SNPs previously identified^{1,7} to be associated with prostate cancer was compiled. In total, 180 unique SNPs were considered for estimation within the PHS model framework. An initial screening was conducted to identify pairs of SNPs that were highly

25 correlated (R² > 0.95). For each pair of highly correlated SNPs, a univariable Cox proportional hazards model using age of diagnosis of prostate cancer as the time to event was calculated for

each SNP in the pair, and the one with the larger p-value was discarded. The remaining SNPs were included as candidates for the new PHS model. The R (v.4.0.1) package 'glmnet' was used to estimate a LASSO-regularized Cox-proportional hazards model^{10,11} using age of diagnosis of prostate cancer as the time to event. The genetic data of candidate SNPs and first

5 four European ancestry principal components were included as predictors. Controls were censored at age of last follow-up. The hyper-parameter of the LASSO-regularized model, lambda, was selected using 10-fold cross-validation^{10,11}. The final form of the LASSO model was estimated at the value of lambda that minimized the mean cross-validated error.

10 Characterization of LASSO-regularized PHS model

The PHS score for each of the individuals in the training and testing set was estimated as the weighted sum of the genetic counts of each of the SNPs in the PHS model, using the LASSO model coefficients as weights. Distributions of the new PHS score were compared qualitatively between training and testing groups to confirm that the model was appropriately

We also sought to assess how the LASSO-regularized PHS score compared to family history in explaining the variation in age at diagnosis of prostate cancer. A multivariable Cox proportional hazards model was estimated using the age at diagnosis of any prostate cancer as the time to event, and the PHS score and family history as predictors in both training and testing

- 20 sets, separately. The family history variable was coded as a binary variable: "None" or "One or more affected first-degree relatives". Observations with missing family history values were removed from the analysis. The explained relative risk¹⁸ (ERR) of each of the covariables as well as the full model were estimated using the "clinfun" software package in R, and provided a quantifiable measure for the importance of each variable in the model. Empirical confidence
- 25 intervals for ERR were estimated using 1000 bootstrapped iterations.

¹⁵ calibrated for use in the testing set.

Performance comparison between PHS46 and LASSO-regularized PHS

Performance in the testing set was assessed using hazard ratios (HRs) and positive predictive value (PPV), as described below. In each case, performance metrics were generated for the newly developed LASSO PHS model and for PHS46. Model coefficients for PHS46 were

5 obtained from the literature¹⁷. For each performance metric, one thousand bootstrap samples of the testing set were used to generate empirical 95% confidence intervals for LASSO PHS and for PHS46. In addition, bootstrapped 95% confidence intervals were generated for the percentage change of each performance metric between the two models, using PHS46 as the reference. Percent changes were deemed statistically significant if the bootstrapped 95%

10 confidence interval did not include 0.

HR performance

Calibration Cox proportional hazards models were fit to the bootstrapped testing data using the PHS score as the sole predictor and the age-of-diagnosis of prostate cancer as the

- 15 dependent variable. The model coefficient of this Cox regression model is referred to as the calibration factor. Next, the hazard ratio between two PHS groups, such as those in the top 5% to the middle 40% (HR95/50), is estimated as the exponential of the product of the calibration factor and the difference in mean PHS scores of each group. Hazard ratios between the top 20% to the bottom 20% (HR80/20) and the bottom 20% to the middle 40% (HR20/50) were
- 20 similarly calculated. The PHS cutoffs used to define these groups were determined from the distribution of PHS in the training set controls under 70 years of age^{1,15}.

A similar strategy was used to estimate the HR performance for clinically significant prostate cancer. The criteria for clinical significance were any of: Gleason score >=7, stage T3-T4, PSA concentration >= 10ng/mL, pelvic lymph nodal metastasis, or distant metastasis¹⁹. In

25 this analysis, controls and low-risk (i.e., not clinically significant) cancers were censored at age of last follow-up and age of diagnosis, respectively. HRs are reported after sample-weight

correction^{1,17,20} using the total number of cases and controls in the ProtecT trial to generate weighting factors.

Sample-weight corrected HR values were also generated using the age at diagnosis of non-clinically significant prostate cancer. Individuals with clinically significant prostate cancer were removed from this secondary analysis.

PPV performance

5

10

One indicator of clinical utility of a risk-stratification approach like PHS is whether it can be used to improve the PPV of the standard clinical screening test, prostate-specific antigen (PSA). As a population-based screening study, ProtecT provides biopsy results of both cases

- and controls with a positive PSA result (i.e., ≥3 ng/mL). PPV performance of each model was estimated by randomly sampling individuals within the testing set with positive PSA results, while maintaining the case to control ratio of the ProtecT study (1:2). PPV is calculated as the fraction of positive PSA individuals in the top 20% (PPV80) or top 5% (PPV95) of PHS scores
- 15 that had clinically significant prostate cancer.

Cumulative incidence curves for LASSO-PHS in United Kingdom

To illustrate the utility of the LASSO PHS model in informing prostate cancer screening, cumulative incidence curves for various PHS risk groups were estimated, as described

- 20 previously²¹. The age-specific general cumulative incidence curve for prostate cancer was estimated for the United Kingdom population, aged 40 to 70, using data from Cancer Research UK 2015-2017²². The proportion of clinically significant and non-clinically significant prostate cancer at each age was estimated using data from the Cluster Randomized Trial of PSA Testing for Prostate Cancer (CAP) trial²³. Disease-specific cumulative incidence curves for clinically
- 25 significant and non-clinically-significant prostate cancer were estimated by multiplying the general cumulative incidence curve by their respective proportions. The risk-adjusted incidence

curves for individuals in the upper 5th percentile and upper 20th percentile were estimated by multiplying the disease-specific cumulative incidence curves by the mean value of HR95/50 and HR80/50 in the testing set, respectively. Hazard ratios were obtained using the age of diagnosis of clinically significant prostate cancer as the time-to-event and after sample-weight correction.

5

10

Results

SNP screening and PHS model training

Of the 180 SNPs originally considered for this study, 6 SNPs were discarded in the initial screening process of removing highly correlated SNPs. Of the 174 remaining candidate SNPs (Supplementary Table 2), 166 had non-zero LASSO model coefficients and were selected for the final PHS model (PHS166).

The majority of the 166 variants (ninety-seven, 53%) used in PHS166 were classified as intron variants (Supplementary Table 3). Of the genes associated with variants from PHS166,

HNF1B on chromosome 17 was associated with the greatest number of variants (4). Additional genes that were associated with multiple variants included ITGA6(x2), LINC00506(x2),
 PDLIM5(x2), TERT(x2), CTD-2194D22.4(x2), RGS17(x2), LOC105375751(x2), and CASC8(x3).
 Two of the SNPs used in PHS166 (rs721048 and rs10993994) were designated as 'pathogenic' by ClinVar²⁴ and associated with hereditary prostate cancer.

20

25

PHS166 model characterization

Distributions of PHS166 score were visually consistent between training and testing sets (Supplementary Figure 1). The 20th, 30th, 70th, 80th, and 98th percentiles of the reference PHS risk scores (controls in training set) were estimated as -0.411, -0.307, 0.048, 0.154, and 0.557, respectively.

PHS166 contributed roughly 80 to 90 percent of the total explained relative risk (Supplementary Table 4) of a Cox proportional hazards model containing both family history and PHS166. Family history was not found to be statistically significantly associated with age at diagnosis of prostate cancer in the testing set¹.

5

Performance comparison – PHS46 vs. PHS166

All PHS166 HR-based performance metrics showed statistically significant improvements compared to PHS46 (Table 1), for both any and clinically significant prostate cancer. The mean HR95/50 and HR80/20 values for PHS166 were roughly 36 to 55% greater

- 10 than those for PHS46. For example, HR80/20 for clinically significant prostate cancer increased from 6.12 to 9.45. Similarly, HR20/50 for PHS166 was, on average, 18% lower than that for PHS46. Similar trends were observed for non-clinically significant prostate cancer (Supplementary Table 5). No significant differences between models were observed in either of the PPV-based performance metrics (Table 2). Among individuals in the top 20% of risk scores
- 15 with a positive PSA test, the estimated mean PPV for clinically significant prostate cancer was roughly 0.19 irrespective of the model used – indicating approximately 19% of positive PSA tests in this risk group yielded a diagnosis of clinically significant prostate cancer. By comparison, approximately 13% of all positive PSA tests resulted in a diagnosis of clinically significant prostate cancer.
- 20

25

Cumulative incidence curves for PHS166 in United Kingdom

Cumulative incidence curves for clinically significant and non-clinically significant prostate cancer for the upper 5th percentile (>95th percentile) and upper 20th percentile (>80th percentile) of PHS166 scores in the United Kingdom demonstrated expected stratification of prostate cancer risk (Figure 1).

Discussion

Using a machine-learning, LASSO-regularized Cox framework, we identified 166 SNPs to be included in a polygenic hazard model (PHS166) for association with age of diagnosis of prostate cancer in men of European genetic ancestry. Variants used in PHS166 were associated with

- 5 several genes, including those encoding for hepatocyte nuclear factor-1 beta (HNF1B), cancer susceptibility 8 (CASC8), and telomerase (TERT). PHS166 also explained a much larger percentage of the total explained relative risk compared to family history, suggesting that the former is important for stratifying patients' risk. When compared to the original PHS, consisting of 46 SNPs, PHS166 demonstrated substantially improved HR performance. For example, the
- 10 HR for clinically significant prostate cancer comparing the upper and lower quintiles of genetic risk increased by 56% when using PHS166. No significant improvements were found in the PPV of PSA testing when using PHS to stratify risk.

Increased separation in hazard rates between PHS risk groups may allow for more nuance in clinical decision making in certain scenarios. Accurate identification of low, intermediate, and

- 15 high PHS risk groups in prostate cancer may help in decisions of when (or if) to initiate screening as well as possibly improving the interpretation of the disease screens²⁵. Targeting screening to men in the upper percentiles of polygenic risk as opposed to those in the lowest risk group may reduce the proportion of overdiagnosed indolent cancers from 43% to 19%^{26,27}. Risk stratification achieved here by PHS166 is similar or better than commonly used clinical
- tools for diseases such as breast cancer, diabetes, and cardiovascular disease^{25,28–30}. Clinically meaningful risk stratification is illustrated by the estimated cumulative incidence curves in Figure
 This effect is particularly pronounced for clinically significant disease because of the increased proportion of clinically significant cases observed at older ages^{2,21,23}.

The lack of improvement in PPV in this study may suggest a "performance plateau" when using PHS to define broad risk categories for certain clinical applications. A similar effect has been previously described for prostate cancer polygenic models, in the context of using risk

scores to discriminate prostate biopsy outcomes³¹. Some of the precision in a score may also be diluted in broad clinical applications. The PPV analysis here is applied to participants in the ProtecT trial, which enrolled men aged 50 to 69 years, and screening in the trial was offered irrespective of underlying genetic risk². Further investigation is needed to learn whether timing

5 screening according to genetic risk might better leverage the superior HR performance of PHS166 risk score to improve the PPV of PSA testing.

LASSO frameworks have been used to identify SNPs for polygenic risk scores of several phenotypes, including fracture risk³², type 2 diabetes³³, and breast cancer¹². In this work, we have extended the application of LASSO to select SNPs in a polygenic hazard model of

- 10 prostate cancer from a list of candidates previously identified through logistic and time-to-event analysis. Simulation studies¹¹ have suggested that LASSO provides more robust estimates than stepwise selection in cases with both a few large effects, as well as many small effects. As new prostate cancer associated variants are discovered, this framework can be easily implemented to develop updated polygenic hazard models.
- 15 One limitation of PHS166 is that it was entirely developed and tested in European men. However, a well-vetted, well-tested PHS model for men of European genetic ancestry can be used as a starting block for developing models for other genetic ancestries, where large-scale databases are often more scarce, as has been shown for PHS46^{17,34}. Furthermore, some of the SNPs selected for incorporation into PHS166 were originally discovered in analyses that
- 20 included men from the ProtecT testing set. Therefore, the improvements in HRs observed for PHS166 may be somewhat overestimated. However, this bias is likely small, given that the testing set was only a small fraction (less than 5%) of the data used in prior discovery analyses, and the ProtecT data were not used to calculate SNP weights in PHS166. The LASSOregularized Cox framework was also used to minimize any potential for over-fitting³⁵ by
- 25 introducing penalties for large effect sizes. In addition, this study uses age of diagnosis as the time-to-event variable, and any preceding period of undiagnosed disease is unknown.

Hypothetical perfect measurement of age of onset would likely further improve performance of the PHS model.

In conclusion, we applied a machine-learning, LASSO-regularized Cox regression framework to develop a larger PHS that includes 166 previously discovered SNPs. When

5 comparing the performance of PHS166 to the original model, PHS46, we found that incorporating 120 more SNPs significantly improved HRs for clinically significant prostate cancer. However, incorporating more SNPs did not improve on the ability of PHS46 to inform the PPV of PSA testing in the ProtecT dataset, perhaps illustrating a plateau effect and/or dilution of risk stratification in a broad clinical application.

Ethics Statement

All contributing studies were approved by the relevant ethics committees and performed in accordance with the Declaration of Helsinki; written informed consent was obtained from the study participants. The present analyses used de-identified data from the PRACTICAL

5 consortium and have been approved by the review board at the corresponding authors' institution.

Conflict of Interest:

All authors declare no personal or financial conflicts of interest for the submitted work except as follows. CCF is a scientific consultant for CorTechs Labs, Inc. RE reports honorarium as a speaker for GU-ASCO meeting in San Francisco Jan 2016, support from Janssen, and

- 5 honorarium as speaker for RMH-FR meeting Nov 2017. She reports honorarium as a speaker at the University of Chicago invited talk May 2018, and an educational honorarium by Bayer & Ipsen to attend GU Connect "Treatment sequencing for mCRPC patients within the changing landscape of mHSPC" at ESMO Barcelona, Sep 2019. She reports member of external Expert Committee on the Prostate Dx Advisory Panel. OAA received speaker's honorarium from
- 10 Lundbeck, and is a consultant for Healthlytix. AMD reports that he was a founder and holds equity in CorTechs Labs Inc., and serves on its Scientific Advisory Board. He is a member of the Scientific Advisory Board of Human Longevity, Inc., and the Mohn Medical Imaging and Visualization Centre. He received funding through research grants from GE Healthcare to UCSD. The terms of these arrangements have been reviewed by and approved by UCSD in
- 15 accordance with its conflict of interest policies. TMS reports honoraria, outside of the present work, from: University of Rochester, Varian Medical Systems, Multimodal Imaging Servcies Corporation; and WebMD. He reports research funding from NIH/NBIB, U.S. Department of Defense, Radiological Society of North America, American Society for Radiation Oncology, and Varian Medical Systems.

20

Data Availability Statement

The data used in this work were obtained from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium. Readers who are interested in accessing the data must first submit a proposal to the Data

- 5 Access Committee. If the reader is not a member of the consortium, their concept form must be sponsored by a principal investigator (PI) of one of the PRACTICAL consortium member studies. If approved by the Data Access Committee, PIs within the consortium, each of whom retains ownership of their data submitted to the consortium, can then choose to participate in the specific proposal. In addition, portions of the data are available for request from dbGaP
- 10 (database of Genotypes and Phenotypes) which is maintained by the National Center forBiotechnology Information (NCBI):

https://www.ncbi.nlm.nih.gov/gap/?term=lcogs+prostatehttps://www.ncbi.nlm.nih.gov/gap/?term =lcogs+prostate.

Anyone can apply to join the consortium. The eligibility requirements are listed here:

- 15 http://practical.icr.ac.uk/blog/?page_id=9. Joining the consortium would not guarantee access, as a proposal for access would still be submitted to the Data Access Committee, but there would be no need for a separate member sponsor. Readers may find information about application by using the contact information below:
- 20 Rosalind Eeles Principal Investigator for PRACTICAL Professor of Oncogenetics Institute of Cancer Research (ICR) Sutton, UK
- 25 Email: PRACTICAL@icr.ac.uk URL: http://practical.icr.ac.uk Tel: ++44 (0)20 8722 4094

References

10

- Seibert TM, Fan CC, Wang Y, Zuber V, Karunamuni R, Parsons JK *et al.* Polygenic hazard score to guide screening for aggressive prostate cancer: Development and validation in large scale cohorts. *BMJ* 2018; **360**: 1–7.
- Huynh-Le MP, Myklebust TÅ, Feng CH, Karunamuni R, Johannesen TB, Dale AM *et al.* Age dependence of modern clinical risk groups for localized prostate cancer—A
 population-based study. *Cancer* 2020; **126**: 1691–1699.
 - Pashayan N, Duffy SW, Chowdhury S, Dent T, Burton H, Neal DE *et al.* Polygenic
 susceptibility to prostate and breast cancer: Implications for personalised screening. *Br J Cancer* 2011; **104**: 1656–1663.
 - 4 Witte JS. Personalized prostate cancer screening: Improving PSA tests with genomic information. *Sci Transl Med* 2010; **2**: 1–5.
 - 5 Chen H, Liu X, Brendler CB, Ankerst DP, Leach RJ, Goodman PJ *et al.* Adding genetic risk score to family history identifies twice as many high-risk men for prostate cancer:
- 15 Results from the prostate cancer prevention trial. *Prostate* 2016; **76**: 1120–1129.
 - Huynh-Le M-P, Fan CC, Karunamuni R, Thompson WK, Martinez ME, Eeles RA *et al.* Polygenic hazard score is associated with prostate cancer in multi-ethnic populations.
 medRxiv 2020; : 1–34.
 - 7 Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ et al.
- 20 Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet* 2018; **50**: 928–936.
 - Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, Ingles SA *et al.* Multiple novel prostate cancer predisposition loci confirmed by an international study: The
 PRACTICAL consortium. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 2052–2061.
- 25 9 Eeles RA, Olama AA Al, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom

genotyping array. Nat Genet 2013; 45: 385-391.

- Tibshiranit BR. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc B* 1996; :
 267–288.
- 11 Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* 1997;16: 385–395.
- Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A *et al.* Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet* 2019; **104**: 21–34.
- 13 Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA *et al.* The oncoarray
- consortium: A network for understanding the genetic architecture of common cancers.
 Cancer Epidemiol Biomarkers Prev 2017; 26: 126–135.
 - Eeles R. Prostate cancer genome-wide association study from 89,000 men using the
 OncoArray chip to identify novel prostate cancer susceptibility loci. *J Clin Oncol* 2016; 34: 1525.
- 15 Karunamuni RA, Huynh-Le MP, Fan CC, Eeles RA, Easton DF, Kote-Jarai ZsS *et al.* The effect of sample size on polygenic hazard models for prostate cancer. *Eur J Hum Genet* 2020. doi:10.1038/s41431-020-0664-2.
 - Li Y, Byun J, Cai G, Xiao X, Han Y, Cornelis O *et al.* FastPop: A rapid principal component derived method to infer intercontinental ancestry using genetic data. *BMC*
- 20 *Bioinformatics* 2016; **17**: 1–8.

5

- Huynh-Le M-P, Chieh Fan C, Karunamuni R, Martinez ME, Eeles RA, Kote-Jarai Z *et al.* Polygenic hazard score is associated with prostate cancer in multi-ethnic populations.
 medRxiv 2019. doi:https://doi.org/10.1101/19012237.
- Heller G. A measure of explained risk in the proportional hazards model. *Biostatistics*2012; **13**: 315–325.
 - 19 NCCN Clinical Practice Guidelines in Oncology. Prostate Cancer. Version 1.2019.

- 20 Therneau TM, Li H. Computing the Cox Model for Case Cohort Designs. *Lifetime Data Anal* 1999; **5**: 99–112.
- 21 Huynh-Le M-P, Fan CC, Karunamuni R, Walsh EI, Turner EL, Lane JA *et al.* A genetic risk score to personalize prostate cancer screening, applied to population data. *Cancer*

5 Epidemiol Biomarkers Prev 2020; : cebp.1527.2019.

- 22 Prostate cancer incidence statistics | Cancer Research UK. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-bycancer-type/prostate-cancer/incidence#heading-One (accessed 23 Jul2020).
- 23 Martin RM, Donovan JL, Turner EL, Metcalfe C, Young GJ, Walsh El et al. Effect of a
- 10 low-intensity PSA-based screening intervention on prostate cancer mortality: The CAP randomized clinical trial. *JAMA J Am Med Assoc* 2018; **319**: 883–895.
 - Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM *et al.* ClinVar:
 Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014; **42**: 980–985.
- 15 25 Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018; **19**: 581–590.
 - 26 Pashayan N, Duffy SW, Neal DE, Hamdy FC, Donovan JL, Martin RM *et al.* Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* 2015; **17**: 789–795.

20 27 Pashayan N, Pharoah PDP, Schleutker J, Talala K, Tammela TLJ, Määttänen L *et al.* Reducing overdiagnosis by polygenic risk-stratified screening: Findings from the Finnish section of the ERSPC. *Br J Cancer* 2015; **113**: 1086–1093.

- Wang TJ, Larson MG, Levy D, Benjamin EJ, Leip EP, Omland T *et al.* Plasma Natriuretic
 Peptide Levels and the Risk of Cardiovascular Events and Death. *N Engl J Med* 2004;
 350: 655–663.
- 29 Yang X, Leslie G, Gentry-Maharaj A, Ryan A, Intermaggio M, Lee A et al. Evaluation of

25

polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J Med Genet* 2018; **55**: 546–554.

- 30 Yeh HC, Duncan BB, Schmidt MI, Wang NY, Brancati FL. Smoking, smoking cessation, and risk for type 2 diabetes mellitus: A cohort study. *Ann Intern Med* 2010; **152**: 10–17.
- S 31 Ren S, Xu J, Zhou T, Jiang H, Chen H, Liu F *et al.* Plateau effect of prostate cancer risk-associated SNPs in discriminating prostate biopsy outcomes. *Prostate* 2013; **73**: 1824–1835.
 - Forgetta V, Keller-baruch J, Forest M, Durand A, Bhatnagar S, Kemp JP *et al.* Development of a polygenic risk score to improve screening for fracture risk : A genetic risk prediction study. *PLoS Med* 2020; : 1–19.
 - 33 Chen T-H, Chatterjee N, Landi MT, Shi J. A penalized regression framework for building polygenic risk models based on summary statistics from genome-wide association studies and incorporating external information. *J Am Stat Assoc* 2020; **1459**: 1–19.
 - 34 Karunamuni R, Huynh-Le M-P, Fan CC, Thompson W, Eeles RA, Kote-Jarai Z et al.
- 15 African-specific improvement of a polygenic hazard score for age at diagnosis of prostate cancer. *medRxiv* 2020; : 1–32.
 - 35 McNeish DM. Using Lasso for Predictor Selection and to Assuage Overfitting: A Method Long Overlooked in Behavioral Sciences. *Multivariate Behav Res* 2015; **50**: 471–484.

20

10

Figure Legends

Figure 1. Cumulative incidence curves for PHS166. Risk-adjusted cumulative incidence curves for the upper 5th percentile (>95th percentile) and upper 20th percentile (>80th percentile) of PHS166 scores for clinically significant and non-clinically-significant prostate cancer.

5 Reference curves representing the population average cumulative incidence (i.e., unadjusted for genetic risk).

Table 1. HR performance in testing set. Sample-weight-corrected hazard ratios areestimated for PHS166 and PHS46 in the testing set, using age-of-onset of any or clinicallysignificant prostate cancer. The percent change for each metric is calculated using the value ofPHS46 as the reference. Mean values and 95% confidence intervals are reported.

Type of cancer	HR	PHS46	PHS166	Change (%)
	HR95/50	3.29 [2.73,3.77]	4.45 [3.68,5.06]	36 [18,53]
Any	HR80/20	5.15 [3.92,6.18]	7.85 [6.04,9.33]	53 [25,78]
	HR20/50	0.44 [0.40,0.49]	0.37 [0.33,0.40]	-18 [-25,-10]
Clinically	HR95/50	3.72 [2.89,4.43]	5.09 [3.84,6.05]	37 [13,59]
Significant	HR80/20	6.12 [4.18,7.67]	9.45 [6.17,11.79]	55 [17,88]
	HR20/50	0.41 [0.35,0.47]	0.34 [0.29,0.39]	-18 [-28,-9]

Table 2. PPV performance in testing set. Positive predictive value (PPV) of PSA testing for clinically significant prostate cancer using top 5% (PPV95) and top 20% (PPV80) cutoffs of PHS166 and PHS46 risk scores. The percent change for each metric is calculated using the value of PHS46 as the reference.

PPV	PHS46	PHS166	Change (%)
PPV95	0.227 [0.159,0.292]	0.239 [0.171,0.305]	6.3 [-25.5,32.1]
PPV80	0.192 [0.155,0.231]	0.187 [0.150,0.222]	-2.8 [-16.3,9.9]



Table of Contents

_

Supplementary Table 1. Contributing studies	2
Supplementary Table 2. SNP characteristics	6
Supplementary Table 3. Annotations for PHS166 variants	11
Supplementary Figure 1. Distribution of PHS166 score.	15
Supplementary Table 4. Explained relative risk (ERR) comparison	16
Supplementary Table 5. Testing performance for non-clinically significant prostate	17
cancer	17
Appendix 1. Additional members of the PRACTICAL Consortium	16
Appendix 2. Funding sources and Acknowledgements for the PRACTICAL consortium	21

Study Group Acronym	Study Group Name	cases	controls	Average age of cases
	Training Set			
AHS	The Agricultural Health Study	491	1159	67.6
ATBC	Alpha-Tocopherol Beta- Carotene (BPC3)	1281	1913	72.2
Aarhus	Aarhus Prostate Cancer Study	1076	545	64.0
CCI	Cross Cancer Institute Prostate Brachytherapy Cohort	266	0	63.7
СОН	City Of Hope	257	259	60.4
COSM	The Cohort Of Swedish Men	2298	1117	70.6
CPCS1	Copenhagen Prostate Cancer Study 1	536	258	68.2
CPCS2	Copenhagen Prostate Cancer Study 2	444	228	63.8
Canary PASS	Canary Prostate Active Surveillance Study (PASS)	364	0	62.1
CeRePP	French Prostate Case- Control Study	923	644	65.8
EPIC	European Prospective Investigation Into Cancer and Nutrition (BPC3)	635	693	66.7
ERSPC	European Randomised study of Screening for Prostate Cancer	71	65	71.2
ESTHER	Epidemiological investigations of the chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population	324	315	64.8
FHCRC	Fred Hutchinson Prostate Cancer Studies	407	388	60.4
Gene-PARE	Adverse Radiotherapy Response	242	0	66.2

Supplementary Table 1.Contributing Studies. Descriptions of contributing studies to training and testing sets.

HPFS	Health Professionals Follow-up Study (BPC3)	1168	1044	69.8
Hamburg-Zagreb	· · · · · · · · · · · · · · · · · · ·	146	149	68.1
IMPACT	Identification of Men with a genetic predisposition to ProstAte Cancer: Targeted screening in men at a higher genetic risk and controls	49	867	63.8
IPO-Porto	Portuguese Oncology Institute of Porto	374	180	56.3
KULEUVEN	Katholieke Universiteit Leuven	166	103	65.8
LAAPC	University of Southern California – Los Angeles Prostate Cancer Study	440	280	67.3
MCC-Spain	Multi Case Control Study- Spain	520	397	66.8
MCCS	Melbourne Collaborative Cohort Study MD Anderson Cancer	715	315	69.6
MDACC_AS	Center, Active surveillance trial	501	0	64.7
MEC	Multiethnic Cohort Study (BPC3)	598	642	69.9
MOFFITT	The Moffitt Group	403	203	64.7
Malaysia	Prostate cancer study in Malaysia	1	0	78.4
Oslo	COhort of NORway (CONOR)	1443	0	72.2
PCMUS	Prostate Cancer study Medical University Sofia	192	60	68.2
PHS	Physicians Health Study (BPC3)	622	257	68.7
PLCO	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (BPC3)	678	980	72.4
PRAGGA	PRostate cAncer Genetics in Galicia	129	100	68.1
PROCAP	PROgression in CAncer of the Prostate	659	236	64.3
PROFILE	correlation with targeted prostate cancer screening	13	21	59.4

	and treatment – The Pilot			
	Profile Sludy			
PROGReSS	Prostate Cancer Group,	673	322	69.7
Deland	Sanuago, Spain	404	047	60.4
Poland	The Poland Group	484	317	69.1
	Prostate cancer :	000	40	04 5
PromPT	Mechanisms of progression	839	12	64.5
	and Treatment			
	Supportive Care and	0000	4000	00 5
QLD	Patient Outcomes Project	3282	1232	62.5
	(ProsCan), The Qidivien			
	and the Red Cross study			
	Radiogenomics:			
	Assessment of	1011	0	70.0
RAPPER	Polymorphisms for	1914	0	70.2
	Predicting the Effects of			
	Radiotherapy			
0545011	Study of Epidemiology and	0544		00 4
SEARCH	Risk factors in Cancer	2511	1440	63.1
	Heredity			
05500	San Francisco Bay Area			
SFPCS	Prostate Cancer Study	279	205	64.8
	(former NC_CCPC)			
SNP Prostate Ghent		316	135	65.3
	Serum Proteomic analysis			
	for biomarkers of			
SPAG	Aggressive prostate	39	169	65.3
	disease in the Guernsey			
	population	/ -		
STHM2	Stockholm 2	3019	1481	65.3
SWOG-PCPT	Prostate Cancer	1072	1084	69.9
	Prevention Trial			
SWOG-SELECT	Selenium and Vitamin E	1479	2070	67.4
	Cancer Prevention Trial			
	Finnish Genetic			
TAMPERE	Predisposition to Prostate	2421	1183	67.2
	Cancer Study			
TORONTO	Princess Margaret Biopsy	644	449	64 6
	Database	011	110	0.110
	U.K. Genetic Prostate			
UKGPCS	Cancer Study and The	11939	1432	61 1
	Prostate Cancer Research			01.1
	Foundation Study			

ULM	Familial Prostate Cancer Study Germany	457	178	64.3
WUGS	Washington University Genetics Study	669	0	61.2
	Testing Set			
ProtecT	Prostate Testing for Cancer and Treatment	1553	1464	62.8
UKGPCS	U.K. Genetic Prostate Cancer Study and The Prostate Cancer Research Foundation Study	30	3364	61.0

Supplementary Table 2. SNP characteristics. RS-ID, Chromosome (Chr), Effect allele, Reference Allele (Ref), Base pair position (version 37), and LASSO-derived PHS coefficient for each of the 174 SNPs considered for this study. SNPs highlighted in gray (n=8) were not included in the final model – PHS166.

RS-ID	Chr	Effect	Ref	Position	PHS coefficient
rs56391074	1	AT	А	88210715	0.019
rs17599629	1	G	А	150658287	0.028
rs34579442	1	С	СТ	153899900	0.033
rs1218582	1	G	А	154834183	0.023
rs4245739	1	С	А	204518842	-0.039
rs62106670	2	Т	С	8597123	0.023
rs9287719	2	С	Т	10710730	0.037
rs9306895	2	С	Т	20878153	0.013
rs1465618	2	Т	С	43553949	0.030
rs721048	2	А	G	63131731	0.002
rs6545977	2	А	G	63301164	-0.049
rs74702681	2	Т	С	66652885	0.054
rs10187424	2	С	Т	85794297	-0.035
rs11691517	2	G	Т	111893096	-0.032
rs12621278	2	G	А	173311553	-0.116
rs16860513	2	А	Т	173342367	-0.004
rs34925593	2	С	Т	174234547	0.018
rs59308963	2	TATTCTGTC	Т	202123479	-0.017
rs2292884	2	G	А	238443226	0.036
rs3771570	2	Т	С	242382864	0.044
rs2660753	3	Т	С	87110674	0.000
rs75219487	3	А	Т	87147922	0.057
rs6788616	3	G	А	87205079	0.050
rs7611694	3	С	А	113275624	-0.041
rs4857841	3	А	G	128046643	0.048
rs6763931	3	А	G	141102833	0.016
rs182314334	3	С	Т	152004202	-0.041
rs142436749	3	G	А	169093100	0.083
rs78416326	3	С	G	170074517	-0.091
rs10936632	3	С	А	170130102	-0.009
rs10009409	4	Т	С	73855253	0.008
rs1894292	4	А	G	74349158	-0.026
rs12500426	4	А	С	95514609	0.000
rs6853490	4	G	А	95544718	0.020
rs17021918	4	Т	С	95562877	-0.036
rs7679673	4	А	С	106061534	-0.073
rs2242652	5	А	G	1280028	-0.030

rs7725218	5	А	G	1282414	-0.040
rs2736108	5	Т	С	1297488	0.039
rs10866527	5	Т	С	1891800	0.026
rs12653946	5	Т	С	1895829	0.003
rs2121875	5	С	А	44365545	0.009
rs10793821	5	С	Т	133836209	-0.019
rs76551843	5	G	А	169172133	-0.138
rs4976790	5	Т	G	177968915	0.027
rs4713266	6	Т	С	11219030	-0.033
rs7767188	6	А	G	30073776	0.021
rs12665339	6	G	А	30601232	0.008
rs3096702	6	А	G	32192331	0.016
rs9296068	6	G	Т	32988695	-0.018
rs9469899	6	А	G	34793124	0.034
rs1983891	6	Т	С	41536427	0.034
rs4711748	6	Т	С	43694598	0.019
rs9443189	6	G	А	76495882	-0.012
rs2273669	6	G	А	109285189	0.032
rs339331	6	С	Т	117210052	-0.041
rs3910736	6	Т	С	153412476	-0.020
rs1933488	6	G	А	153441079	-0.022
	_	_	_	100000001	0.059
rs9364554	6	Т	С	160833664	0.056
rs9364554 rs527510716	6 7	T C	C G	160833664 1944537	0.058
rs9364554 rs527510716 rs11452686	6 7 7	T C TA	C G T	160833664 1944537 20414110	0.038 0.018 0.000
rs9364554 rs527510716 rs11452686 rs12155172	6 7 7 7 7	T C TA A	C G T G	160833664 1944537 20414110 20994491	0.058 0.018 0.000 0.048
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567	6 7 7 7 7 7	T C TA A A	C G T G G	160833664 1944537 20414110 20994491 27976563	0.038 0.018 0.000 0.048 -0.060
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345	6 7 7 7 7 7 7	T C TA A A C	C G T G G A	160833664 1944537 20414110 20994491 27976563 40875192	0.038 0.018 0.000 0.048 -0.060 -0.029
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506	6 7 7 7 7 7 7 7 7	T C TA A A C A	C G T G G A G	160833664 1944537 20414110 20994491 27976563 40875192 47437244	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016	6 7 7 7 7 7 7 7 7 7	T C TA A A C A C	C G T G A G A G	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.023 0.051
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679	6 7 7 7 7 7 7 7 7 8	T C TA A C A C A C A	C G T G A G A G A G	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.023 0.051 0.028
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388	6 7 7 7 7 7 7 7 8 8 8	T C TA A C A C A C A C	C G G G A G A G T	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.023 0.051 0.028 0.047
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910	6 7 7 7 7 7 7 7 8 8 8 8 8	T C TA A C A C A C A C T	C G T G A G A G T C	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746	6 7 7 7 7 7 7 8 8 8 8 8 8 8	T C TA A C A C A C T C	C G G A G A G T C T	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.023 0.051 0.028 0.047 0.043 -0.040
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663	6 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C A C T C	C G G G A G A G T C T A	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.023 0.051 0.028 0.047 0.043 -0.040 0.000
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908	6 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C	C G G G A G A G T C T A T	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804	6 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C A C T C C C C G	C G G G A G A G T C T C T A T A	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 -0.061
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs77541621	6 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C C G A	C G G G A G A G T C T C T A T A G	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315 128077146	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 -0.061 0.152
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs77541621 rs1016343	6 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C A C T C C C C G A T	C G G A G A G T C T A T A G C	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315 128077146 128093297	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 0.000 0.000
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs28556804 rs77541621 rs1016343 rs183373024	6 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C C C G A T G	C G G A G A G T C T A T A G C A	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315 128077146 128093297 128104117	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 0.000 0.000 0.000
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs77541621 rs1016343 rs183373024 rs16901979	6 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C C C G A T G A	C G G A G A G T C T A T A G C A C	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315 128077146 128093297 128104117 128124916	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.052 0.081 0.361 0.022
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs28556804 rs77541621 rs1016343 rs183373024 rs16901979 rs60163266	6 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C C C G A T G A T G A A	C G G G A G A G T C T A G C A C G G	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128014315 128014315 128077146 128093297 128104117 128124916 128323157	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 -0.061 0.152 0.081 0.361 0.022 0.025
rs9364554 rs527510716 rs11452686 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs12543663 rs10086908 rs28556804 rs28556804 rs77541621 rs1016343 rs183373024 rs183373024 rs16901979 rs60163266 rs620861	6 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8	T C TA A C A C A C T C C C C C G A T G A T G A A A A	C G G G A G A G T C T A T A G C A C G G G	160833664 1944537 20414110 20994491 27976563 40875192 47437244 97807882 23438975 23525358 25892142 127909361 127924659 128011937 128014315 128077146 128093297 128104117 128124916 128323157 128335673	0.038 0.018 0.000 0.048 -0.060 -0.029 0.023 0.051 0.028 0.047 0.043 -0.040 0.000 0.000 0.000 -0.061 0.152 0.081 0.361 0.361 0.022 0.025 -0.051

rs1447295	8	А	С	128485038	0.000
rs7812894	8	А	Т	128520479	0.135
rs12549761	8	G	С	128540776	-0.083
rs1048169	9	С	Т	19055965	0.021
rs17694493	9	G	С	22041998	0.022
rs10122495	9	Т	Α	34049779	0.004
rs1182	9	А	С	132576060	0.034
rs141536087	10	GCGCA	G	854691	0.058
rs76934034	10	С	Т	46082985	-0.002
rs10993994	10	Т	С	51549496	0.117
rs1935581	10	Т	С	90195149	-0.029
rs3850699	10	G	А	104414221	-0.024
rs7094871	10	С	G	114712154	-0.008
rs4962416	10	С	Т	126696872	0.037
rs1881502	11	Т	С	1507512	0.008
rs72853963	11	А	G	2224664	0.022
rs7127900	11	А	G	2233574	0.069
rs61890184	11	А	G	7547587	0.032
rs547171081	11	CGG	С	47421962	0.014
rs2277283	11	С	Т	61908440	0.033
rs12785905	11	С	G	66951965	0.033
rs12275055	11	G	Α	68981359	0.076
rs7929962	11	С	Т	68985583	-0.047
rs11290954	11	А	AC	76260543	-0.025
rs11568818	11	С	Т	102401661	-0.031
rs1800057	11	G	С	108143456	0.052
rs11214775	11	А	G	113807181	-0.040
rs138466039	11	Т	С	125054793	0.086
rs878987	11	G	Α	134266372	0.025
rs2066827	12	G	Т	12871099	-0.034
rs10845938	12	А	G	14416918	-0.034
rs80130819	12	С	Α	48419618	-0.051
rs10875943	12	С	Т	49676010	0.038
rs902774	12	А	G	53273904	0.008
rs55914512	12	Т	G	53282274	0.067
rs7968403	12	С	Т	65012824	-0.030
rs5799921	12	G	GA	90160530	-0.032
rs1270884	12	А	G	114685571	0.022
rs7295014	12	G	Α	133067989	0.026
rs1004030	14	С	Т	23305649	-0.014
rs11629412	14	G	С	37138294	-0.032
rs8008270	14	<u> </u>	С	53372330	-0.042
rs4643253	14	С	Т	69106108	0.000

rs7141529	14	Т	С	69126744	-0.022
rs8014671	14	А	G	71092256	-0.021
rs4924487	15	G	С	40922915	-0.032
rs33984059	15	G	А	56385868	-0.074
rs112293876	15	С	CA	66764641	0.034
rs11863709	16	Т	С	57654576	-0.061
rs201158093	16	TAA	TA	82178893	0.034
rs684232	17	С	Т	618965	0.056
rs28441558	17	С	Т	7803118	0.061
rs142444269	17	Т	С	30098749	-0.027
rs11649743	17	А	G	36074979	-0.050
rs718961	17	А	G	36077099	-0.016
rs4430796	17	G	А	36098040	-0.011
rs11651052	17	А	G	36102381	-0.085
rs117576373	17	Т	С	46820676	0.098
rs11650494	17	А	G	47345186	0.047
rs2680708	17	А	G	56456120	-0.024
rs1859962	17	G	Т	69108753	0.083
rs8093601	18	С	G	51772473	0.018
rs28607662	18	С	Т	53230859	0.026
rs12956892	18	Т	G	56746315	0.013
rs533722308	18	СТ	С	60961193	0.026
rs10460109	18	Т	С	73036165	0.022
rs7241993	18	Т	С	76773973	-0.048
rs11666569	19	Т	С	17214073	-0.032
rs118005503	19	С	G	32167803	-0.024
rs8102476	19	Т	С	38735613	-0.047
rs11672691	19	А	G	41985587	-0.049
rs61088131	19	С	Т	42700947	-0.009
rs17632542	19	С	Т	51361757	-0.197
rs2735839	19	А	G	51364623	-0.020
rs11480453	20	CA	С	31347512	-0.025
rs12480328	20	С	Т	49527922	-0.039
rs6091758	20	G	А	52455205	0.045
rs2427345	20	Т	С	61015611	-0.032
rs35897249	20	G	А	62233638	-0.017
rs6062509	20	G	Т	62362563	-0.017
rs1041449	21	G	А	42901421	0.028
rs9625483	22	А	G	28888939	0.059
rs58133635	22	Т	С	40471188	0.025
rs5759167	22	Т	G	43500212	-0.064
rs73179053	22	С	Т	43501620	-0.085
rs747745	22	С	Т	43503547	-0.007

rs2405942	23	G	А	9814135	-0.021
rs17321482	23	Т	С	11482634	-0.024
rs4907775	23	G	Α	51263200	0.053
rs2807031	23	С	Т	52896949	0.000
rs7888856	23	G	А	66751555	-0.034
rs5919432	23	С	Т	67021550	-0.005
rs11795627	23	Т	С	69957441	-0.020
rs6625711	23	А	Т	70139850	0.004

Supplementary Table 3. Annotations for PHS166 variants. Annotations and nearby genes

RS-ID Gene annotations			
rs56391074	None		
rs17599629	GOLPH3L / Intron Variant		
rs34579442	LOC101928059 / Intron Variant		
rs1218582	KCNN3 / Intron Variant		
rs4245739	MDM4 / Non Coding Transcript Variant		
rs62106670	None		
rs9287719	NOL10 / 500B Downstream Variant		
rs9306895	GDF7 / 3 Prime UTR Variant		
rs1465618	THADA / Intron Variant		
rs721048	EHBP1 / Intron Variant		
rs6545977	None		
rs74702681	MEIS1-AS3 / Non Coding Transcript Variant		
rs10187424	None		
rs11691517	BCL2L11 / Intron Variant		
rs12621278	ITGA6 / Intron Variant		
rs16860513	ITGA6 / Intron Variant		
rs34925593	None		
rs59308963	CASP8 / Intron Variant		
rs2292884	MLPH / Missense Variant		
rs3771570	FARP2 / Intron Variant		
rs75219487	LINC00506 / Intron Variant		
rs6788616	LINC00506 / Intron Variant		
rs7611694	SIDT1 / Intron Variant		
rs4857841	EEFSEC / Intron Variant		
rs6763931	ZBTB38 / Intron Variant		
rs182314334	MBNL1 / Intron Variant		
rs142436749	MECOM / Intron Variant		
rs78416326	SKIL / 2KB Upstream Variant		
rs10936632	None		
rs10009409	LOC105377273 / Intron Variant		
rs1894292	AFM / Intron Variant		
rs6853490	PDLIM5 / Intron Variant		
rs17021918	PDLIM5 / Intron Variant		
rs7679673	None		
rs2242652	TERT / Intron Variant		
rs7725218	TERT / Intron Variant		
rs2736108	None		
rs10866527	CTD-2194D22.4 / Intron Variant		
rs12653946	CTD-2194D22.4 / Intron Variant		

for 166 variants used in PHS166 were tabulated using publicly available data from dbSNP.

rs2121875 rs10793821 rs76551843 rs4976790 rs4713266 rs7767188 rs12665339 rs3096702 rs9296068 rs9469899 rs1983891 rs4711748 rs9443189 rs2273669 rs339331 rs3910736 rs1933488 rs9364554 rs527510716 rs12155172 rs10486567 rs17621345 rs56232506 rs6965016 rs2928679 rs11782388 rs11135910 rs9297746 rs28556804 rs77541621 rs1016343 rs183373024 rs16901979 rs60163266 rs620861 rs6983267 rs7812894 rs12549761 rs1048169 rs17694493 rs10122495 rs1182 rs141536087

FGF10 / Intron Variant None DOCK2 / Intron Variant COL23A1 / Intron Variant NEDD9 / Intron Variant TRIM31 / Intron Variant ATAT1 / Intron Variant NOTCH4 / 2KB Upstream Variant None UHRF1BP1 / Intron Variant FOXP4 / Intron Variant None MYO6 / Intron Variant ARMC2 / Intron Variant **RFX6 / Intron Variant** RGS17 / Intron Variant RGS17 / Intron Variant SLC22A3 / Intron Variant MAD1L1 / Intron Variant LINC01162 / Intron Variant JAZF1 / Intron Variant SUGCT / Intron Variant TNS3 / Intron Variant LMTK2 / Intron Variant None LOC107986930 / Intron Variant EBF2 / Intron Variant LOC105375751 / Intron Variant LOC105375751 / Intron Variant None PCAT2 / Intron Variant PRNCR1 / Non Coding Transcript Variant None CASC8 / Intron Variant CASC8 / Intron Variant CASC8 / Intron Variant None None HAUS6 / 3 Prime UTR Variant CDKN2B-AS1 / Intron Variant UBAP2 / 2KB Upstream Variant TOR1A / Non Coding Transcript Variant LARP4B / 3 Prime UTR Variant

rs76934034	MARCHF8 / Intron Variant		
rs10993994	MSMB / 2KB Upstream Variant		
rs1935581	RNLS / Intron Variant		
rs3850699	TRIM8 / Intron Variant		
rs7094871	TCF7L2 / Intron Variant		
rs4962416	CTBP2 / Intron Variant		
rs1881502	MOB2 / Intron Variant		
rs72853963	None		
rs7127900	None		
rs61890184	PPFIBP2 / Intron Variant		
rs547171081	MIR4487 / 2KB Upstream Variant		
rs2277283	INCENP / Missense Variant		
rs12785905	KDM2A / Intron Variant		
rs12275055	None		
rs7929962	None		
rs11290954	EMSY / Intron Variant		
rs11568818	MMP7 / 2KB Upstream Variant		
rs1800057	ATM / Missense Variant		
rs11214775	HTR3B / Intron Variant		
rs138466039	None		
rs878987	B3GAT1 / Intron Variant		
rs2066827	CDKN1B / Missense Variant		
rs10845938	None		
rs80130819	LOC105369750 / Intron Variant		
rs10875943	None		
rs902774	None		
rs55914512	None		
rs7968403	RASSF3 / Intron Variant		
rs5799921	LOC107984543 / Intron Variant		
rs1270884	None		
rs7295014	FBRSL1 / Intron Variant		
rs1004030	MMP14 / 2KB Upstream Variant		
rs11629412	PAX9 / Intron Variant		
rs8008270	FERMT2 / Intron Variant		
rs7141529	RAD51B / Intron Variant		
rs8014671	LOC101928075 / Intron Variant		
rs4924487	KNL1 / Intron Variant		
rs33984059	RFX7 / Missense Variant		
rs112293876	None		
rs11863709	ADGRG1 / Intron Variant		
rs201158093	None		

rs684232	VPS53 / 2KB Upstream Variant		
rs28441558	CHD3 / Intron Variant		
rs142444269	None		
rs11649743	HNF1B / Intron Variant		
rs718961	HNF1B / Intron Variant		
rs4430796	HNF1B / Intron Variant		
rs11651052	HNF1B / Intron Variant		
rs117576373	LOC105371811 / Non Coding Transcript Variant		
rs11650494	None		
rs2680708	RNF43 / Intron Variant		
rs1859962	CASC17 / Intron Variant		
rs8093601	None		
rs28607662	TCF4 / Intron Variant		
rs12956892	None		
rs533722308	BCL2 / Intron Variant		
rs10460109	None		
rs7241993	LOC105372225 / Intron Variant		
rs11666569	MYO9B / Intron Variant		
rs118005503	None		
rs8102476	None		
rs11672691	PCAT19 / Intron Variant		
rs61088131	POU2F2 / Intron Variant		
rs17632542	KLK3 / Missense Variant		
rs2735839	None		
rs11480453	None		
rs12480328	ADNP / Intron Variant		
rs6091758	None		
rs2427345	LOC105372710 / Intron Variant		
rs35897249	GMEB2 / Intron Variant		
rs6062509	ZGPAT / Intron Variant		
rs1041449	None		
rs9625483	TTC28 / Intron Variant		
rs58133635	TNRC6B / Intron Variant		
rs5759167	None		
rs73179053	None		
rs747745	None		
rs2405942	SHROOM2 / Intron Variant		
rs17321482	ARHGAP6 / Intron Variant		
rs4907775	None		
rs7888856	None		
rs5919432	None		
rs11795627	TEX11 / Intron Variant		
rs6625711	None		





Supplementary Table 4. Explained relative risk (ERR) comparison. Values of ERR are tabulated for PHS166, Family History (None or >=1 affected relative), and the full model (PHS166 + Family History). PHS166 contributed 82% of the overall ERR in the training set, and 98% of the overall ERR in the testing set. z

			PHS166
dataset	PHS166	Family History	+
			Family History
training	0.116 [0.109,1.122]	0.016 [0.013, 0.018]	0.139 [0.132, 0.1460
testing	0.147 [0.116,0.182]	0.001 [0, 0.006]	0.15 [0.12,0.185]

Supplementary Table 5. Testing performance for non-clinically significant prostate

cancer. Sample-weight hazard ratios are estimated for PHS166 and PHS46 in the testing set using the age-of-onset of non-clinically significant prostate cancer.

HR	PHS46	PHS166	Change (%)
HR95/50	3.20 [2.59,3.78]	4.28 [3.49,4.96]	34 [16,51]
HR80/20	4.96 [3.61,6.14]	7.47 [5.49,9.07]	51 [23,77]
HR20/50	0.45 [0.39,0.50]	0.37 [0.33,0.42]	-17 [-25,-10]

Appendix 1. Additional members of the PRACTICAL Consortium

Teuvo L. J. Tammela^{1,2}, Csilla Sipeky³, Anssi Auvinen⁴, Alison M. Dunning⁵, Maya Ghoussaini⁶, Suzanne Chambers^{7,8}, Judith A. Clements^{9,10}, Lisa Horvath^{11,12}, Leire Moya^{9,10}, Gail P. Risbridger^{13,14}, Wayne Tilley¹⁵, Markus Aly^{16,17,18}, Tobias Nordström^{19,20}, Stig E. Bojesen^{21,22}, Peter Iversen²³, Phyllis J. Goodman²⁴, Ian M. Thompson Jr.²⁵, Melissa C. Southey^{26,27,28}, Graham G. Giles²⁷, Roger L. Milne^{27,29,26}, Niclas Håkansson³⁰, Stephanie Weinstein³¹, Fredrick R. Schumacher^{32,33}, Loic Le Marchand³⁴, Xin Sheng³⁵, Tim J. Key³⁶, Elaine A. Ostrander³⁷, Edward Giovannucci³⁸, Neil Burnet³⁹, Gill Barnett⁴⁰, Martin Andreas Røder²³, Bettina F. Drake⁴¹. Géraldine Cancel-Tassin^{42,43}, Stephen Chanock³¹, Gerald L. Andriole⁴¹, Robert N. Hoover³¹, Mitchell J. Machiela³¹, Laura E. Beane Freeman³¹, Michael Borre^{44,45}, Dominika Wokolorczyk⁴⁶, Jan Lubinski⁴⁶, Lovise Maehle⁴⁷, Hui-Yi Lin⁴⁸, Mariana C Stern⁴⁹, Manuel Luedeke⁵⁰, Thomas Schnoeller⁵¹, Neil E. Fleshner⁵², Antonio Finelli⁵³, Sarah L. Kerns⁵⁴, Harry Ostrer⁵⁵, Antonio Gómez-Caamaño⁵⁶, Laura Fachal^{57,58,59,60}, Javier Llorca^{61,62}, Gemma Castaño-Vinyals^{63,64,65,66}, Antonio Alcaraz⁶⁷, Robert Szulkin^{68,69}, Martin Eklund⁷⁰, Peter Kraft⁷¹, Sara Lindström⁷², Meir Stampfer⁷³, Paula Paulo⁷⁴, Andreia Brandão^{75,74}, Marta Cardoso⁷⁴, Maria P. Silva^{75,74}, Xin Gao⁷⁶, Bernd Holleczek⁷⁷, Ben Schöttker⁷⁶, Chavdar Slavov⁷⁸, Vanio Mitev⁷⁹, Jeri Kim⁸⁰, Yuan Chun Ding⁸¹, Linda Steele⁸¹, Jasmine Lim⁸², Soo-Hwang Teo⁸³, Daniel W. Lin^{84,85}, Davor Lessel⁸⁶, Zeljko Kastelan⁸⁷, Matthew Parliament^{88,89}, Aswin Abraham⁸⁸, Steven Joniau⁹⁰, Thomas Van den Broeck^{91,90}, Jose Esteban Castelao⁹², Maria Elena Martinez⁹³, Ron H.N. van Schaik⁹⁴, Guido Jenster⁹⁵, Maria Elena Martinez⁹⁸, Ron H.N. van Schaik⁹⁹, Guido Jenster¹⁰⁰

¹Department of Urology, Tampere University Hospital, FI-33521 Tampere, Finland ²Faculty of Medicine and Health Technology, Tampere University, FI-33100 Tampere, Finland ³Institute of Biomedicine, University of Turku, Finland

⁴Unit of Health Sciences, Faculty of Social Sciences, Tampere University, Tampere, Finland ⁵Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, CB1 8RN, UK

⁶Open Targets, Wellcome Sanger Institute, Hinxton, Saffron Walden, CB10 1SA, UK ⁷University of Technology, Sydney

⁸Cancer Council Queensland, Fortitude Valley, QLD 4006, Australia

⁹Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Sciences, Queensland University of Technology, Brisbane QLD 4059, Australia

¹⁰Translational Research Institute, Brisbane, Queensland 4102, Australia

¹¹Chris O'Brien Lifehouse (COBLH), Camperdown, Sydney, NSW 2010, Australia

¹²Garvan Institute of Medical Research, Sydney NSW 2010, Australia

¹³Department of Anatomy and Developmental Biology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria 3800, Australia

¹⁴Prostate Cancer Translational Research Program, Cancer Research Division, Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia

¹⁵Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide, Adelaide, South Australia, Australia

¹⁶Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171

77 Stockholm, Sweden

¹⁷Department of Molecular Medicine and Surgery, Karolinska Institutet

¹⁸Department of Urology, Karolinska University Hospital, Solna, 171 76 Stockholm

¹⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

²⁰Department of Clinical Sciences at Danderyds Hospital, Karolinska Institutet, Stockholm, Sweden

²¹Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark

²²Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, 2200 Copenhagen, Denmark

²³Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, DK-2730 Herlev, Copenhagen, Denmark

²⁴SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
 ²⁵CHRISTUS Santa Rosa Hospital – Medical Center, San Antonio, TX, USA

²⁶Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria 3168, Australia

²⁷Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Road, Melbourne, VIC 3004, Australia

²⁸Department of Clinical Pathology, The Melbourne Medical School, The University of Melbourne, Melbourne, VIC 3010, Australia.

²⁹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Grattan Street, Parkville, VIC 3010, Australia

³⁰Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden

³¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland, 20892, USA

³²Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH 44106-7219, USA

³³Seidman Cancer Center, University Hospitals, Cleveland, OH 44106, USA.

³⁴Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA

³⁵Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of

Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA

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

³⁷National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Rm. 5351, Bethesda, MD 20892, USA

³⁸Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA
 ³⁹Division of Cancer Sciences, University of Manchester, Manchester Cancer Research Centre, Manchester Academic Health Science Centre, and The Christie NHS Foundation Trust, Manchester, UK

⁴⁰University of Cambridge, Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, CB2 0QQ, UK

⁴¹Washington University School of Medicine, 660 S. Euclid Avenue, Campus Box 8242, St.

Louis, MO 63110, USA

⁴²CeRePP, Tenon Hospital, F-75020 Paris, France.

⁴³Sorbonne Universite, GRC n°5 , AP-HP, Tenon Hospital, 4 rue de la Chine, F-75020 Paris, France

⁴⁴Department of Urology, Aarhus University Hospital, Palle Juul-Jensen Boulevard 99, 8200 Aarhus N, Denmark

⁴⁵Department of Clinical Medicine, Aarhus University, DK-8200 Aarhus N

⁴⁶International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-115 Szczecin, Poland

⁴⁷Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway
 ⁴⁸School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA
 70112, USA

⁴⁹Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90015, USA

⁵⁰Humangenetik Tuebingen, Paul-Ehrlich-Str 23, D-72076 Tuebingen, Germany ⁵¹Department of Urology, University Hospital Ulm, Germany

⁵²Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto ON M5G 2M9, Canada

 ⁵³Division of Urology, Princess Margaret Cancer Centre, Toronto ON M5G 2M9, Canada
 ⁵⁴Department of Radiation Oncology, University of Rochester Medical Center, Saunders Research Building, Rm 2.202, 265 Crittenden Boulevard, Rochester, NY 14620, USA
 ⁵⁵Professor of Pathology and Pediatrics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Ullman 817, Bronx, NY 10461, USA

⁵⁶Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela 15706, Spain

⁵⁷Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, CB2 0SR, UK.

⁵⁸Fundación Pública Galega Medicina Xenómica, Santiago De Compostela, 15706, Spain.
 ⁵⁹Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago De Compostela, 15706, Spain.

⁶⁰Centro de Investigación en Red de Enfermedades Raras (CIBERER), Spain

⁶¹University of Cantabria-IDIVAL, 39005 Santander, Spain

⁶²CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain⁶³ISGlobal, Barcelona, Spain

⁶⁴IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain

⁶⁵Universitat Pompeu Fabra (UPF), Barcelona, Spain

⁶⁶CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

⁶⁷Department and Laboratory of Urology. Hospital Clínic. Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona. Spain. C/Villarroel 170; 08036 Barcelona, Spain

⁶⁸Division of Family Medicine, Department of Neurobiology, Care Science and Society, Karolinska Institutet, Huddinge, Sweden

⁶⁹Scandinavian Development Services, Danderyd, Sweden

⁷⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institute, 171 77 Stockholm,

Sweden

⁷¹Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

⁷²Department of Epidemiology, Health Sciences Building, University of Washington

⁷³Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02184, USA

⁷⁴Cancer Genetics Group, IPO-Porto Research Center (CI-IPOP), Portuguese Oncology Institute of Porto (IPO-Porto), Porto, Portugal

⁷⁵Department of Genetics, Portuguese Oncology Institute of Porto (IPO-Porto), Porto, Portugal
 ⁷⁶Division of Clinical Epidemiology and Aging Research, German Cancer Research Center
 (DKFZ), D-69120, Heidelberg, Germany

⁷⁷Saarland Cancer Registry, 66119 Saarbrücken, Germany

⁷⁸Department of Urology and Alexandrovska University Hospital, Medical University of Sofia, 1431 Sofia, Bulgaria

⁷⁹Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University of Sofia, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria

⁸⁰The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, 1515 Holcombe Blvd., Houston, TX 77030, USA

⁸¹Department of Population Sciences, Beckman Research Institute of the City of Hope, 1500 East Duarte Road, Duarte, CA 91010, 626-256-HOPE (4673)

⁸²Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

⁸³Cancer Research Malaysia (CRM), Outpatient Centre, Subang Jaya Medical Centre, Subang Jaya, Selangor, Malaysia

⁸⁴Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109-1024, USA

⁸⁵Department of Urology, University of Washington, 1959 NE Pacific Street, Box 356510, Seattle, WA 98195, USA

⁸⁶Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany

⁸⁷Department of Urology, University Hospital Center Zagreb, University of Zagreb School of Medicine , Zagreb, Croatia

⁸⁸Department of Oncology, Cross Cancer Institute, University of Alberta, 11560 University Avenue, Edmonton, Alberta, Canada T6G 1Z2

⁸⁹Division of Radiation Oncology, Cross Cancer Institute, 11560 University Avenue, Edmonton, Alberta, Canada T6G 1Z2

⁹⁰Department of Urology, University Hospitals Leuven, Herestraat 49, Box 7003 41, BE-3000 Leuven, Belgium

⁹¹Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, BE-3000, Belgium

⁹²Genetic Oncology Unit, CHUVI Hospital, Complexo Hospitalario Universitario de Vigo,
 Instituto de Investigación Biomédica Galicia Sur (IISGS), 36204, Vigo (Pontevedra), Spain
 ⁹³University of California San Diego, Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA 92093-0012, USA

⁹⁴Department of Clinical Chemistry, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands

⁹⁵Department of Urology, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands

⁹⁶Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, BE-3000, Belgium

 ⁹⁷Genetic Oncology Unit, CHUVI Hospital, Complexo Hospitalario Universitario de Vigo, Instituto de Investigación Biomédica Galicia Sur (IISGS), 36204, Vigo (Pontevedra), Spain
 ⁹⁸University of California San Diego, Moores Cancer Center, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA 92093-0012, USA
 ⁹⁹Department of Clinical Chemistry, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands

¹⁰⁰Department of Urology, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands

Appendix 2. Funding sources and Acknowledgements for the PRACTICAL consortium

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 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.

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].

This study would not have been possible without the contributions of the following: Coordination team, bioinformatician and genotyping centers: Genotyping at CCGE, Cambridge: Caroline Baines and Don Conroy

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.

This study would not have been possible without the contributions of the following: Per Hall (COGS); Douglas F. Easton, Paul Pharoah, Kyriaki Michailidou, Manjeet K. Bolla, Qin Wang (BCAC), Andrew Berchuck (OCAC), Rosalind A. Eeles, Douglas F. Easton, Ali Amin Al Olama, Zsofia Kote-Jarai, Sara Benlloch (PRACTICAL), Georgia Chenevix-Trench, Antonis Antoniou, Lesley McGuffog, Fergus Couch and Ken Offit (CIMBA), Joe Dennis, Alison M. Dunning, Andrew Lee, and Ed Dicks, Craig Luccarini and the staff of the Centre for Genetic Epidemiology

Laboratory, Javier Benitez, Anna Gonzalez-Neira and the staff of the CNIO genotyping unit, Jacques Simard and Daniel C. Tessier, Francois Bacot, Daniel Vincent, Sylvie LaBoissière and Frederic Robidoux and the staff of the McGill University and Génome Québec Innovation Centre, Stig E. Bojesen, Sune F. Nielsen, Borge G. Nordestgaard, and the staff of the Copenhagen DNA laboratory, and Julie M. Cunningham, Sharon A. Windebank, Christopher A. Hilker, Jeffrey Meyer and the staff of Mayo Clinic Genotyping Core Facility

Additional funding and acknowledgments from studies in PRACTICAL:

Information of the PRACTICAL consortium can be found at http://practical.icr.ac.uk/

<u>Aarhus</u>

This study was supported by Innovation Fund Denmark, the Danish Cancer Society and The Velux Foundation (Veluxfonden).

The Danish Cancer Biobank (DCB) is acknowledged for biological material.

<u>AHS</u>

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics (Z01CP010119).

<u>ATBC</u>

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, and by U.S. Public Health Service contract HHSN261201500005C from the National Cancer Institute, Department of Health and Human Services.

Canary PASS

PASS was supported by Canary Foundation and the National Cancer Institute's Early Detection Research Network)U01 CA086402)

<u>CAP</u>

The CAP trial is funded by Cancer Research UK and the UK Department of Health (C11043/A4286, C18281/A8145, C18281/A11326, C18281/A15064, and C18281/A24432). RMM is supported by a Cancer Research UK (C18281/A19169) programme grant (the Integrative Cancer Epidemiology Programme) and by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

<u>CCI</u>

This work was awarded by Prostate Cancer Canada and is proudly funded by the Movember Foundation - Grant # D2013-36.

The CCI group would like to thank David Murray, Razmik Mirzayans, and April Scott for their contribution to this work.

<u>COH</u>

SLN is partially supported by the Morris and Horowitz Families Endowed Professorship

<u>COSM</u>

COSM is funded by The Swedish Research Council (grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research – SIMPLER), the Swedish Cancer Foundation.

CPCS1 / CPCS2

Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark

We thank participants and staff of the Copenhagen General Population Study for their important contributions.

<u>EPIC</u>

The coordination of EPIC was financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts (that recruited male participants) are supported by Danish Cancer Society (Denmark); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe. Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada; , PI13/01162 to EPIC-Murcia), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). For information on how to submit an application for gaining access to EPIC data and/or biospecimens, please follow the instructions at http://epic.iarc.fr/access/index.php

ERSPC

This study was supported by the DutchCancerSociety(KWF94-869,98-1657,2002-277,2006-3518, 2010-4800); The Netherlands Organisation for HealthResearch and Development (ZonMW-002822820,22000106,50-50110-98-311, 62300035), The Dutch Cancer Research Foundation(SWOP), and an unconditional grant from Beckman-Coulter-HybritechInc.

<u>ESTHER</u>

The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts.

The ESTHER group would like to thank Hartwig Ziegler, Sonja Wolf, Volker Hermann, Heiko Müller, Karina Dieffenbach, Katja Butterbach for valuable contributions to the study.

FHCRC

The FHCRC studies were supported by grants R01-CA080122, R01-CA056678, R01-CA082664, and R01-CA092579, and K05-CA175147 from the US National Cancer Institute, National Institutes of Health, with additional support from the Fred Hutchinson Cancer Research Center (P30-CA015704).

We thank all the individuals who participated in these studies.

Gene-PARE

The Gene-PARE study was supported by grants 1R01CA134444 from the U.S. National Institutes of Health , PC074201 and W81XWH-15-1-0680 from the Prostate Cancer Research Program of the Department of Defense and RSGT-05-200-01-CCE from the American Cancer Society. S.L.K. is supported by 1K07CA187546 from the U.S. National Cancer Institute.

<u>HPFS</u>

The Health Professionals Follow-up Study was supported by grants UM1CA167552, CA133891, CA141298, and P01CA055075.

We are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

IMPACT

The IMPACT study was funded by The Ronald and Rita McAulay Foundation, CR-UK Project grant (C5047/A1232), Cancer Australia, AICR Netherlands A10-0227, Cancer Australia and Cancer Council Tasmania, NIHR, EU Framework 6, Cancer Councils of Victorial and South Australia, Philanthropic donation to Northshore University Health System. We acknowledge support from the National Institute for Health Research (NIHR) to the Biomedical Research Centre at The Institute of Cancer Research and Royal Marsden Foundation NHS Trust. We acknowledge the IMPACT study steering committee, collaborating centres and participants.

IPO-Porto

The IPO-Porto study was funded by Fundação para a Ciência e a Tecnologia (FCT; UID/DTP/00776/2013 and PTDC/DTP-PIC/1308/2014) and FEDER (POCI-01-0145-FEDER-028245). MC (SFRH/BD/116557/2016) and MPS (SFRH/BD/132441/2017) are research fellows from FCT.

We would like to express our gratitude to all patients and families who have participated in this study.

<u>KULEUVEN</u>

F.C. and S.J. are holders of grants from FWO Vlaanderen (G.0684.12N and G.0830.13N), the Belgian federal government (National Cancer Plan KPC_29_023), and a Concerted Research Action of the KU Leuven (GOA/15/017). TVDB is holder of a doctoral fellowship of the FWO.

LAAPC

This study was funded by grant R01CA84979 (to S.A. Ingles) from the National Cancer Institute, NIH.

<u>Malaysia</u>

The study was funded by the University Malaya High Impact Research Grant (HIR/MOHE/MED/35 to A.R).

We thank all associates in the Urology Unit, University of Malaya, Cancer Research Malaysia (CRM) and the Malaysian Men's Health Initiative (MMHI).

MCC-Spain

The study was partially funded by the ""Accion Transversal del Cancer"", approved on the Spanish Ministry Council on the 11th October 2007, by the Instituto de Salud Carlos III-FEDER (PI08/1770, PI09/00773-Cantabria, PI11/01889-FEDER, PI12/00265, PI12/01270, PI12/00715, PI15/00069), by the Fundación Marqués de Valdecilla (API 10/09), by the Spanish Association Against Cancer (AECC) Scientific Foundation and by the Catalan Government DURSI grant 2009SGR1489. Samples: Biological samples were stored at the Parc de Salut MAR Biobank (MARBiobanc; Barcelona) which is supported by Instituto de Salud Carlos III FEDER (RD09/0076/00036). Also sample collection was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncologia de Catalunya (XBTC). ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

We acknowledge the contribution from Esther Gracia-Lavedan in preparing the data. We thank all the subjects who participated in the study and all MCC-Spain collaborators.

<u>MCCS</u>

Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.

MEC

The MEC was supported by NIH grants CA063464, CA054281, CA098758, and CA164973.

<u>MOFFITT</u>

The Moffitt group was supported by the US National Cancer Institute (R01CA128813, PI: J.Y. Park).

<u>Oslo</u>

CONOR was supported by grants from the Nordic Cancer Union, the Swedish Cancer Society (2012/823) and the Swedish Research Council (2014/2269).

The authors wish to acknowledge the services of CONOR, the contributing research centres delivering data to CONOR, and all the study participants.

PCMUS

The PCMUS study was supported by the Bulgarian National Science Fund, Ministry of Education and Science (contract DOO-119/2009; DUNK01/2-2009; DFNI-B01/28/2012) with additional support from the Science Fund of Medical University - Sofia (contract 51/2009; 81/2009; 28/2010;).

<u>PHS</u>

The Physicians' Health Study was supported by grants CA34944, CA40360, CA097193, HL26490 and HL34595.

We are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-Up Study for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

<u>PLCO</u>

This PLCO study was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH

The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention at the National Cancer Institute, the screening center investigators and staff of the PLCO Cancer Screening Trial for their contributions to the PLCO Cancer Screening Trial. We thank Mr. Thomas Riley, Mr. Craig Williams, Mr. Matthew Moore, and Ms. Shannon Merkle at Information Management Services, Inc., for their management of the data and Ms. Barbara O'Brien and staff at Westat, Inc. for their contributions to the PLCO Cancer Screening Trial. We also thank the PLCO study participants for their contributions to making this study possible.

<u>PRAGGA</u>

PRAGGA was supported by Programa Grupos Emergentes, Cancer Genetics Unit, CHUVI Vigo Hospital, Instituto de Salud Carlos III, Spain.

PRAGGA wishes to thank Victor Muñoz Garzón, Manuel Enguix Castelo, Sara Miranda Ponte, Carmen M Redondo, Manuel Calaza, Francisco Gude Sampedro, Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigación Sanitaria Galicia Sur (IISGS), SERGAS, Vigo, Spain; Máximo Fraga, José Antúnez and the Biobank of University Hospital Complex of Santiago, Santiago de Compostela, Spain; and Maria Torres, Angel Carracedo and the Galician Foundation of Genomic Medicine.

PROCAP

PROCAP was supported by the Swedish Cancer Foundation (08-708, 09-0677).

We thank and acknowledge all of the participants in the PROCAP study. We thank Carin Cavalli-Björkman and Ami Rönnberg Karlsson for their dedicated work in the collection of data. Michael Broms is acknowledged for his skilful work with the databases. KI Biobank is acknowledged for handling the samples and for DNA extraction. We acknowledge The NPCR steering group: Pär Stattin (chair), Anders Widmark, Stefan Karlsson, Magnus Törnblom, Jan Adolfsson, Anna Bill-Axelson, Ove Andrén, David Robinson, Bill Pettersson, Jonas Hugosson, Jan-Erik Damber, Ola Bratt, Göran Ahlgren, Lars Egevad, and Roy Ehrnström.

PROFILE

We would like to acknowledge the support of the Ronald and Rita McAulay Foundation and Cancer Research UK. We also acknowledge support from the National Institute for Health Research (NIHR) to the Biomedical Research Centre at The Institute of Cancer Research and Royal Marsden Foundation NHS Trust

We acknowledge the Profile study steering committee and participants.

PROGReSS

This research was supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (INT15/00070, INT16/00154, INT17/00133; PI19/01424; PI16/00046; PI13/02030; PI10/00164), and through the Autonomous Government of Galicia (Consolidation and structuring program: IN607B) given to A.Vega. We would like to thank the patients for their contribution to the study

ProtecT/ProMPT

The ProtecT team acknowledges the support of The Universities of Oxford, Bristol and Cambridge, Cancer Research UK and NIHR. CR-UK grants (C8197/A10123; C8197/A10865) supported genotyping. We acknowledge the support of the National Institute for Health Research Cambridge Bio-medical Research Centre, Cambridge, UK, and Oxford Biomedical Research Centre, Oxford, UK. The national ProMPT (Prostate cancer Mechanisms of Progression and Treatment) collaborative (grant G0500966/75466) supported sample collections in Cambridge and Oxford. The University of Oxford sponsors ProMPT and ProtecT. We are grateful to staff at the Bristol Medical School, Bristol, UK, the Nuffield Department of Surgical Sciences at the University of Oxford, Oxford, UK, the Wellcome Trust Clinical Research Facility, Addenbrooke's Clinical Research Centre, Cambridge, UK, and all research groups in the nine ProtecT participating Clinical centres in the UK (Birmingham, Bristol, Cambridge, Cardiff, Edinburgh, Leeds, Leicester, Newcastle, Sheffield), for conducting the ProtecT study. The NIHR Health Technology Assessment Programme (HTA) funded the ProtecT study (projects 96/20/06, 96/20/99), and supported its linked ProMPT and CAP (Cluster Randomized Trial of PSA Testing for Prostate Cancer) studies, with Cancer Research UK (grants C522/A8649, C11043/A4286, C18281/A8145, C18281/A11326, and C18281/A15064), the Medical Research Council (grant G0500966, ID 75466) and NCRI, UK. The ProtecT epidemiological data were generated through funding from the Southwest National Health Service Research and Development. Protect DNA extraction was supported by the USA Dept of Defense award W81XWH-04-1-0280, Yorkshire Cancer Research and CR-UK. We thank the

National Institute for Health Research, Hutchison Whampoa Limited, the Human Research Tissue Bank (Addenbrooke's Hospital), the Oxford Human Tissue Bank, and Cancer Research UK for supporting the biorepository. The authors thank men with prostate cancer and all the participants who donated their time and samples to the Oxford and Cambridge Biorepositories used in this research. We acknowledge the support of research staff who so carefully curated the samples and follow-up data (Jo Burge, Marie Corcoran, Anne George, Sara Stearn, Rajeev Kumar, Michael Davis, Peter Holding). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the Department of Health and Social Care in England.

<u>QLD</u>

The QLD research is supported by The National Health and Medical Research Council (NHMRC) Australia Project Grants [390130, 1009458] and NHMRC Career Development Fellowship, Cancer Australia PdCCRS and Cancer Council Queensland funding to J Batra The QLD team would like to acknowledge and sincerely thank the urologists, pathologists, data managers and patient participants who have generously and altruistically supported the QLD cohort.

<u>RAPPER</u>

RAPPER is supported by Cancer Research UK [C1094/A11728; C1094/A18504] and Experimental Cancer Medicine Centre funding [C1467/A7286], and the NIHR Manchester Biomedical Research Centre.

The RAPPER group thank Dr. Holly Summersgill for project management.

<u>SEARCH</u>

SEARCH is funded by a programme grant from Cancer Research UK [C490/A10124] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support in respect of PP from the NHS in the East of England through the Clinical Academic Reserve.

<u>SFPCS</u>

SFPCS was funded by California Cancer Research Fund grant 99-00527V-10182

<u>SPAG</u>

Manchester Cancer Research Centre and MRC – Medical Research Council for PhD studentship work, MUG - Male Uprising in Guernsey, Hope for Guernsey and Wessex Medical Research for research support

STHM2

STHM2 was supported by grants from The Strategic Research Programme on Cancer (StratCan), Karolinska Institutet; the Linné Centre for Breast and Prostate Cancer (CRISP, number 70867901), Karolinska Institutet; The Swedish Research Council (number K2010-70X-20430-04-3) and The Swedish Cancer Society (numbers 11-0287 and 11-0624); Stiftelsen Johanna Hagstrand och Sigfrid Linnérs minne; Swedish Council for Working Life and Social

Research (FAS), number 2012-0073

The authors acknowledge the Karolinska University Laboratory, Aleris Medilab, Unilabs and the Regional Prostate Cancer Registry for performing analyses and help to retrieve data. Carin Cavalli–Björkman and Britt-Marie Hune for their enthusiastic work as research nurses. Astrid Björklund for skilful data management. We wish to thank the BBMRI.se biobank facility at Karolinska Institutet for biobank services.

SWOG-PCPT / SWOG-SELECT

PCPT is funded by Public Health Service grants U10CA37429 and 5UM1CA182883 from the National Cancer Institute.

The authors thank the site investigators and staff and, most importantly, the participants from PCPT / SELECT who donated their time to this trial.

TAMPERE

The Tampere (Finland) study was supported by the Academy of Finland (251074), The Finnish Cancer Organisations, Sigrid Juselius Foundation, and the Competitive Research Funding of the Tampere University Hospital (X51003). The PSA screening samples were collected by the Finnish part of ERSPC (European Study of Screening for Prostate Cancer).

TAMPERE would like to thank Riina Liikanen, Liisa Maeaettaenen and Kirsi Talala for their work on samples and databases.

<u>Toronto</u>

Prostate Cancer Canada Movember Discovery Grant (D2013-17) to RJH; Canadian Cancer Society Research Institute Career Development Award in Cancer Prevention (2013-702108) to RJH

<u>UKGPCS</u>

UKGPCS 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. UKGPCS should also like to acknowledge the NCRN nurses, data managers and Consultants for their work in the UKGPCS. UKGPCS would like to thank all urologists and other persons involved in the planning, coordination, and data collection of the study. KM and AL were in part supported from the NIHR Manchester Biomedical Research Centre

ULM

The UIm group received funds from the German Cancer Aid (Deutsche Krebshilfe).

<u>WUGS</u>

WUGS would like to thank the following for funding support: The Anthony DeNovi Fund, the

Donald C. McGraw Foundation, and the St. Louis Men's Group Against Cancer.