

The PRACTICAL CONSORTIUM

Information about the consortium and/or how to access the data used in this model can be found at <http://practical.ccge.medschl.cam.ac.uk/>

Members of the consortium (in addition to those named in the author list):

Margaret Cook¹
Michelle Guy²
Koveela Govindasami²
Daniel Leongamornlert²
Emma J. Sawyer²
Rosemary Wilkinson²
Edward J. Saunders²
Malgorzata Tymrakiewicz²
Tokhir Dadaev²
Angela Morgan²
Cyril Fisher²
Steve Hazel²
Naomi Livni²
Artitaya Lophatananon^{3,4}
Robert Szulkin⁵
Jan Adolfsson^{6,7}
Pär Stattin^{8,9}
Jan-Erik Johansson^{5,10}
Carin Cavalli-Bjoerkman⁵
Ami Karlsson⁵
Michael Broms⁵
Anssi Auvinen¹¹
Paula Kujala¹²
Kirsi Talala¹³
Teemu Murtola^{14,15}
Kimmo Taari¹⁶
Peter Klarskov¹⁷
Hans Wallinder¹⁸
Sven Gustafsson¹⁸
Angela Cox¹⁹
Paul Brown²⁰
Anne George²⁰
Gemma Marsden²⁰
Athene Lane²⁰
Michael Davis²⁰
Wei Zheng²¹
Lisa B. Signorello²²
William J. Blot^{23,24}
Lori Tillmans²⁵
Shaun Riska²⁵
Liang Wang²⁵
Antje Rinckleb²⁶
Jan Lubinski²⁷
Christa Stegmaier²⁸
Julio Pow-Sang²⁹
Hyun Park²⁹
Selina Radlein²⁹
Maria Rincon²⁹
James Haley²⁹
Babu Zachariah²⁹
Darina Kachakova³⁰

Elenko Popov³¹
 Atanaska Mitkova³⁰
 Aleksandrina Vlahova³²
 Tihomir Dikov³²
 Svetlana Christova³²
 Peter Heathcote³³
 Glen Wood³³
 Greg Malone³³
 Pamela Saunders³³
 Allison Eckert³³
 Trina Yeadon³³
 Kris Kerr³³
 Angus Collins³³
 Megan Turner³³
 Srilakshmi Srinivasan³³
 Mary-Anne Kedda³³
 Kimberly Alexander³³
 Tracy Omara³³
 Huihai Wu³⁴
 Rui Henrique³⁵
 Pedro Pinto³⁵
 Joana Santos³⁵
 Joao Barros-Silva³⁵
 Mohamed El Tibi³⁶
 Graham G. Giles^{37,38}
 Melissa C. Southey³⁹
 Liesel M. Fitzgerald⁴⁰
 John Pedersen⁴¹
 John L. Hopper⁴²
 Robert MacInnis^{37,42}
 Brian E. Henderson⁴³
 Fredrick Schumacher⁴⁴
 Christopher A. Haiman⁴³
 Janet L. Stanford^{45,46}
 Susanne Kolb⁴⁵
 Yong-Jie Lu⁴⁷
 Hong-Wei Zhang⁴⁸

¹Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK

²The Institute of Cancer Research, London, SM2 5NG, UK

³Institute of Population Health, University of Manchester, Manchester, UK

⁴Warwick Medical School, University of Warwick, Coventry, UK

⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

⁶Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden

⁷Swedish Agency for Health Technology Assessment and Assessment of Social Services, Stockholm, Sweden

⁸Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Umeå, Sweden

⁹Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

¹⁰Department of Urology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

¹¹Department of Epidemiology, School of Health Sciences, University of Tampere, Tampere, Finland

¹²Fimlab Laboratories, Tampere University Hospital, Tampere, Finland

¹³Finnish Cancer Registry, Helsinki, Finland

¹⁴School of Medicine, University of Tampere, Tampere, Finland

¹⁵Department of Urology, Tampere University Hospital, Tampere, Finland

¹⁶Department of Urology, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland

- ¹⁷Department of Urology, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-230 Herlev, Denmark.
- ¹⁸Department of Epidemiology and Biostatistics, School of Public Health, Imperial College, London, United Kingdom
- ¹⁹CR-UK/YCR Sheffield Cancer Research Centre, University of Sheffield, Sheffield, UK
- ²⁰School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK
- ²¹Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN 37232 USA.
- ²²National Cancer Institute, NIH, 9609 Medical Center Drive, Suite 2W-172, MSC 9712, Bethesda, MD, 20892-9712 (mail), Rockville, MD 20850 (FedEx/Courier), USA
- ²³International Epidemiology Institute, 1555 Research Blvd., Suite 550, Rockville, MD 20850, USA
- ²⁴Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, US
- ²⁵Mayo Clinic, Rochester, Minnesota, USA
- ²⁶Department of Urology, University Hospital Ulm, Germany
- ²⁷International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
- ²⁸Saarland Cancer Registry, 66119 Saarbrücken, Germany
- ²⁹Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA
- ³⁰Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria
- ³¹Department of Urology and Alexandrovska University Hospital, Medical University, Sofia, Bulgaria
- ³²Department of General and Clinical Pathology
- ³³Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia
- ³⁴The University of Surrey, Guildford, Surrey, GU2 7XH
- ³⁵Department of Genetics, Portuguese Oncology Institute, Porto, Portugal
- ³⁶University Hospital "Tsaritsa Yoanna", Medical University - Sofia, Sofia, †Bulgaria
- ³⁷Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria, 3004, Australia
- ³⁸Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria 3010, Australia
- ³⁹Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Grattan Street, Parkville, Victoria 3010, Australia
- ⁴⁰Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria, Australia
- ⁴¹Tissupath Pty Ltd., Melbourne, Victoria 3122, Australia
- ⁴²Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
- ⁴³Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, California, US
- ⁴⁴Case Western Reserve University, School of Medicine, 10900 Euclid Ave., Cleveland, OH, 44106-4945, USA
- ⁴⁵Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109-1024, USA
- ⁴⁶Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, US
- ⁴⁷Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London, EC1M 6BQ, UK
- ⁴⁸Second Military Medical University, 800 Xiangyin Rd., Shanghai 200433, P. R. China

Funding for the CRUK study 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, 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).

COGS acknowledgement:

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

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.

Additional funding and acknowledgments from studies in PRACTICAL:

The Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden was supported by the Cancer Risk Prediction Center (CRiSP; www.crispcenter.org), a Linneus Centre (Contract ID 70867902) financed by the Swedish Research Council, Swedish Research Council (grant no K2010-70X-20430-04-3), the Swedish Cancer Foundation (grant no 09-0677), the Hedlund Foundation, the Soederberg Foundation, the Enqvist Foundation, ALF funds from the Stockholm County Council. Stiftelsen Johanna Hagstrand och Sigfrid Linner's Minne, Karlsson's Fund for urological and surgical research. We thank and acknowledge all of the participants in the Stockholm-1 study. We thank Carin Cavalli-Bjoerkman and Ami Roennberg 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. Hans Wallinder at Aleris Medilab and Sven Gustafsson at Karolinska University Laboratory are thanked for their good cooperation in providing historical laboratory results.

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>.

The ESTHER study was supported by a grant from the Baden Württemberg Ministry of Science, Research and Arts. Additional cases were recruited in the context of the VERDI study, which was supported by a grant from the German Cancer Aid (Deutsche Krebshilfe). The ESTHER group would like to thank Hartwig Ziegler, Sonja Wolf, Volker Hermann, Katja Butterbach.

The FHCRC studies were supported by grants RO1CA056678, RO1CA082664, and RO1CA092579 from the US National Cancer Institute, National Institutes of Health, with additional support from the Fred Hutchinson Cancer Research Center.

The IPO-Porto study was in part funded by Liga Portuguesa Contra o Cancro.

The Mayo group was supported by the US National Cancer Institute (R01CA72818)

The Prostate Cancer Program of Cancer Council Victoria also acknowledge grant support from The National Health and Medical Research Council, Australia (126402, 209057, 251533, , 396414, 450104, 504700, 504702, 504715, 623204, 940394, 614296,), VicHealth, Cancer Council Victoria, The Prostate Cancer Foundation of Australia, The Whitten Foundation, PricewaterhouseCoopers, and Tattersall's. EAO, DMK, and EMK acknowledge the Intramural Program of the National Human Genome Research Institute for their support.

The MEC was supported by NIH grants CA63464, CA54281 and CA098758.

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

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).

ProtecT would like to acknowledge the support of The University of Cambridge, Cancer Research UK. Cancer Research UK grants [C8197/A10123] and [C8197/A10865] supported the genotyping team. We would also like to acknowledge the support of the National Institute for Health Research which funds the Cambridge Bio-medical Research Centre, Cambridge, UK. We would also like to acknowledge the support of the National Cancer Research Prostate Cancer: Mechanisms of Progression and Treatment (PROMPT) collaborative (grant code G0500966/75466) which has funded tissue and urine collections in Cambridge. We are grateful to staff at the Wellcome Trust Clinical Research Facility, Addenbrooke's Clinical Research Centre, Cambridge, UK for their help in conducting the ProtecT study. We also acknowledge the support of the NIHR Cambridge Biomedical Research Centre, the DOH HTA (ProtecT grant) and the NCRI / MRC (ProMPT grant) for help with the bio-repository. The UK Department of Health funded the ProtecT study through the NIHR Health Technology Assessment Programme (projects 96/20/06, 96/20/99). The ProtecT trial and its linked ProMPT and CAP (Comparison Arm for ProtecT) studies are supported by Department of Health, England; Cancer Research UK grant number C522/A8649, Medical Research Council of England grant number G0500966, ID 75466 and The NCRI, UK. The epidemiological data for ProtecT were generated though funding from the Southwest National Health Service Research and Development. DNA extraction in ProtecT was supported by USA Dept of Defense award W81XWH-04-1-0280, Yorkshire Cancer Research and Cancer Research UK. The authors would like to acknowledge the contribution of all members of the ProtecT study research group. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the Department of Health of England. The bio-repository from ProtecT is supported by the NCRI (ProMPT) Prostate Cancer Collaborative and the Cambridge BMRC grant from NIHR.

The QLD research is supported by The National Health and Medical Research Council, Australia Project Grant [390130, 1009458] and Enabling Grant [614296 to APCB]; the Prostate Cancer Foundation of Australia (Project Grant [PG7] and Research Grant [to APCB]).

SCCS is funded by NIH grant R01 CA092447, and SCCS sample preparation was conducted at the Epidemiology Biospecimen Core Lab that is supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA68485). Data on SCCS cancer cases used in this publication were provided by the Alabama Statewide Cancer Registry; Kentucky Cancer Registry, Lexington, KY; Tennessee Department of Health, Office of Cancer Surveillance; Florida Cancer Data System; North Carolina Central Cancer Registry, North Carolina Division of Public Health; Georgia Comprehensive Cancer Registry; Louisiana Tumor Registry; Mississippi Cancer Registry; South Carolina Central Cancer Registry; Virginia Department of Health, Virginia Cancer Registry; Arkansas Department of Health, Cancer Registry, 4815 W. Markham, Little Rock, AR 72205. The Arkansas Central Cancer Registry is fully funded by a grant from National Program of Cancer Registries, Centers for Disease Control and Prevention (CDC). Data on SCCS cancer cases from Mississippi were collected by the Mississippi Cancer Registry which participates in the National Program of Cancer Registries (NPCR) of the Centers for Disease Control and Prevention (CDC). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the CDC or the Mississippi Cancer Registry.

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 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 (9N069 and X51003). The PSA screening samples were collected by the Finnish part of ERSPC (European

Study of Screening for Prostate Cancer). Riina Liikanen is thanked for technical assistance. Riitta Vaalavuo and Liisa Maeaettaenen are thanked for their work with databases.

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 study. UKGPCS would like to thank all urologists and other persons involved in the planning, coordination, and data collection of the CAPS study.

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

The Keith and Susan Warshaw Fund, C. S. Watkins Urologic Cancer Fund and The Tennity Family Fund supported the Utah study. The project was supported by Award Number P30CA042014 from the National Cancer Institute

Supplementary Methods

Description of ProtecT Cohort (Validation Set) Selection

As part of the ProtecT study, genotyping with the iCOGS custom Illumina array was performed on cases diagnosed by PSA screening¹. After quality control steps described previously, there were 1,558 cases available for analysis¹. Controls with normal (<3 ng/ml) or elevated (≥ 3 ng/ml) PSA were selected using the same 5-year age band as the cases and from the same GP register (1,464 analyzed after quality control; 739 with normal PSA, 725 with elevated PSA)¹. Additionally, genotyping was performed for the iCOGS project on ProtecT trial participants who were selected as geographically matched controls for the UK Genetic Prostate Cancer Study (UKGPCS)^{1,2}. This category comprised 3,395 men from ProtecT; 31 of these subsequently developed PCa after initial selection as controls and are therefore analyzed as cases in the present study.

PHS Model SNP Selection and Model Generation

Because prostate cancer risk increases with age³ and anticipated age of developing prostate cancer is highly relevant to clinical management, we applied PHS for deriving both predicted absolute risk and potential age at PCa onset⁴. A univariate trend test was applied to the entire Development Set (31,747 patients x 201,043 SNPs) to assess association with case or control status. All SNPs with resulting p -values $< 10^{-6}$ in the trend test were then entered in a forward, stepwise, greedy algorithm, to select the most predictive SNPs. In each step, logistic regression was used first to improve computational efficiency. SNPs were selected for the model only if they improved prediction of case-control status. After forward, stepwise selection, coefficients for selected SNPs were estimated using a Cox proportional hazard model to predict age at diagnosis with PCa.

Evaluation of Proportional Hazards Assumption

The proportionality of each selected SNP was checked by correlating their Schoenfeld residuals and PCa-free survival. In addition, Kaplan-Meier curves and the predicted values from Cox regression were overlaid on a single plot to assess for overlap that would suggest that the proportionality assumption held for the final PHS model.

Accounting for Potential Sampling Bias

The PHS method includes Cox proportional hazards modeling, a method ideally applied to a cohort design with unbiased samples. The Development Set here has the essential advantage of being large enough to support inquiries into modest single-SNP associations, but the contributing studies include case-control and other designs with a net effect of over-representing cases compared to the general population. This disproportionate number of cases in the Development Set would tend to overestimate the general risk of PCa and therefore underestimate the risk (among cases) attributable to a given SNP. Overall, this means our method yields a conservative estimate of SNP effect sizes in the general population⁵.

A Cox model was also used to test PHS prediction of age of PCa onset in the Validation Set. Here, we have the advantage of ProtecT's cohort design, and the Validation Set can be treated as a nested case-control design, with known sampling rates. The sampling weights for cases and controls were determined from the overall ProtecT numbers⁶, and adjustments to the Cox model were made according to previously published and validated methods⁷ using the R 'survival' package (R version 3.2.2)^{8,9}. Results from the adjusted model were compared to results from the simple model to see whether accounting for potential sampling bias affected PHS performance in the Validation Set.

Calculation of Confidence Intervals for Cox prediction

Based on the variance in genotypes, X , in the Development Set and the uncertainty of the Cox parameter estimates, $\hat{\beta}$, we calculated 95% confidence intervals for the Cox prediction, applicable to Δ Age and Prostate Cancer-Risk (PCaR). Assuming the genotypes distribute independently with the effect sizes on the trait of interest, we can estimate the variance of $\hat{X}\hat{\beta}$:

$$\text{Var}(\hat{X}\hat{\beta}) = \text{Var}(\hat{\beta})\text{Var}(\bar{X}) + \text{E}(\hat{\beta})^2\text{Var}(\bar{X}) + \text{Var}(\hat{\beta})\text{E}(\bar{X})^2$$

The 95% confidence interval of $\hat{X}\hat{\beta}$ can then be derived accordingly, such that the confidence interval of instantaneous hazard at a given age T is:

$$\lambda_0(T) \exp(95\%CI)$$

where λ_0 is the baseline hazard.

Calculation of Positive Predictive Value in Validation Set

In the Validation Set, 2,555 patients had positive PSA: 1,580 were then diagnosed with PCa, while 975 were designated controls without PCa. Because genotype information was collected in more cases than controls, we matched the overall ProtecT control:case ratio⁶ by taking a random sample of 471 cases with the 975 controls and calculating the positive predictive value of PSA testing without regard to PHS, as well as in subsets based on PHS percentile thresholds of <20th, >50th, >80th, and >95th. This process was repeated for a total of 1,000 random samples of 471 cases.

Polygenic Risk Score Analysis using Previously Reported SNPs from GWAS

Traditional GWAS have revealed a number of SNPs associated with prostate cancer. In the present study, the PHS model was built without prior assumptions on which SNPs would be most useful and then optimized parameter estimates for prediction of age of PCa onset. However, it may also be of interest to consider the performance of a traditional polygenic risk score (PRS), built with previously published SNPs and their corresponding odds ratios (OR). We therefore conducted a post-hoc analysis, reported here.

Two recent papers together published a total of 99 SNPs associated with PCa, along with ORs^{14,15}. Genotype data were available for 63 of those SNPs in our Validation Set. A PRS model was constructed using the log odds ratios (from published ORs) for these SNPs and the allele counts in the 6,411 men from the Validation Set. The resulting PRS was used as the sole predictor in a Cox proportional hazards model, analogous to what was done for PHS in the main manuscript. As before, statistical significance was set at alpha of 0.01.

Supplementary Results

Evaluation of Proportional Hazards Assumption

Supplementary Figure S1 shows the correlation of Schoenfeld residuals and PCa-free survival. Additionally, Figure 1 demonstrates reasonable overlap of the Kaplan-Meier and Cox regression estimates of PCa-free survival in the Development Set.

Accounting for Potential Sampling Bias

After accounting for sampling weights in an adjusted Cox model⁷, PHS showed similar performance, with highly significant prediction of age of onset of aggressive PCa ($z=21.7, p<10^{-16}$). The hazard ratio for high PHS men (>98th percentile) compared to average risk was 4.6 [95% CI: 4.0, 5.2]. Overall, these results confirm that sampling bias in the main results leads to a conservative estimate of PHS predictive power.

Positive Predictive Value in Validation Set

As PHS is predictive of PCa risk, we expected it to modulate the PPV of PSA testing. Indeed, risk-stratification with PHS had considerable impact on PPV in the Validation Set. In terms of any PCa (which is what the PSA biopsy threshold was set for in ProtecT), only 18% of those with low PHS were true positives, whereas over half of those with high PHS had PCa (Supplementary Figure S2). A similar pattern was seen for aggressive PCa, though the absolute numbers are much lower, as is to be expected (Figure 2).

Polygenic Risk Score Analysis using Previously Reported SNPs from GWAS

The PRS calculated from 63 previously published SNPs^{14,15} was predictive of age of aggressive PCa onset in the Validation Set ($z=9.2$, $p<10^{-16}$, HR=1.4 [95% CI: 1.3, 1.4]), though its performance was not as good as that of PHS ($z=11.2$, $p<10^{-16}$, HR=2.9 [2.4, 3.4]).

Supplementary Table S1: Study names and participant numbers

Development Set	Country	Dates	Source ^b	Number of participants				Age - median (interquartile range)				PHS - median (range)
				All	Any PCa	Aggressive PCa	Control	All	Any PCa	Aggressive PCa	Control	
CAPS	Sweden	2001-2003	Population-based	1,817	1,153	792	664	66.3 (60.3-72.7)	65.7 (59.5-72.0)	67.0 (60.7-73.8)	68.5 (61.2-73.9)	0.16 (-1.30-1.18)
CPCS1	Denmark	2008-2011	Hospital recruitment	3,610	840	557	2,770	62.0 (51.0-71.0)	69.1 (63.7-75.0)	69.1 (64.0-74.7)	58.0 (46.0-68.0)	0.02 (-2.65-1.02)
CPCS2	Denmark	2010-2011	Hospital recruitment	1,273	264	161	1,009	60.7 (49.0-68.7)	64.5 (60.5-68.5)	64.5 (60.6-68.4)	58.0 (45.0-69.0)	0.00 (-0.99-1.01)
EPIC	EU	1992-2000	Population-based	1,801	722	137	1,079	61.1 (58.1-66.0)	65.2 (61.3-68.7)	65.9 (62.4-69.3)	60.0 (56.0-63.0)	0.08 (-1.01-1.08)
EPIC-Norfolk	UK	1992-2000	Population-based	1,401	484	28	917	73.2 (65.9-80.0)	72.8 (66.8-77.9)	71.3 (65.5-76.2)	73.7 (65.2-81.5)	0.01 (-3.79-1.20)
ESTHER	Germany	2000-2002	Population-based	631	313	175	318	66.0 (62.3-69.0)	66.1 (62.8-68.8)	66.2 (62.8-68.9)	66.0 (62.0-69.0)	0.08 (-0.99-1.14)
IPO-Porto	Portugal	1999-2011	Hospital recruitment	242	183	166	59	58.5 (51.9-62.5)	60.7 (56.9-63.0)	60.8 (57.0-63.0)	34.0 (25.0-47.5)	0.15 (-0.64-0.89)
MAYO	USA	1994-2007	Hospital recruitment	1,254	766	548	488	65.4 (60.0-70.0)	65.7 (61.3-69.7)	66.2 (61.9-70.0)	65.0 (59.0-71.5)	0.11 (-1.07-1.53)
MOFFITT	USA	2002-2009	Hospital recruitment	513	413	195	100	64.0 (59.0-71.0)	65.0 (59.8-71.0)	66.0 (61.0-73.0)	62.0 (57.0-67.0)	0.14 (-0.72-0.97)
PCMUS	Bulgaria	1993-2011	Hospital recruitment	291	151	122	140	68.0 (62.0-74.0)	69.3 (63.4-74.4)	69.9 (63.5-75.4)	67.0 (60.0-73.3)	0.07 (-2.61-0.84)
PPF-UNIS	UK	1993-2011	Hospital recruitment	433	245	151	188	68.3 (62.1-73.6)	69.4 (63.2-73.5)	70.9 (65.2-75.0)	67.2 (59.8-73.8)	0.12 (-2.28-1.10)
Poland	Poland	1999-2009	Hospital recruitment	790	438	259	352	67.0 (58.0-72.0)	68.0 (63.0-73.0)	69.0 (63.0-73.8)	62.0 (54.0-71.0)	0.12 (-0.78-0.93)
ProMPT	UK	2001-2009	Population-based	168	166	130	2	65.0 (61.5-72.0)	65.0 (61.4-72.0)	66.0 (62.2-72.0)	70.1 (65.0-75.2)	0.14 (-0.61-0.98)
QLD	Australia	2004-2011	Hospital recruitment	212	127	100	85	65.8 (59.5-69.0)	61.0 (57.0-66.0)	62.0 (58.0-67.5)	68.7 (66.4-72.5)	0.13 (-2.74-0.98)
SEARCH	UK	2005-2013	Population-based	2,613	1,371	565	1,242	60.0 (54.0-65.0)	64.0 (60.0-67.0)	64.0 (61.0-67.0)	55.0 (50.0-60.0)	0.12 (-2.78-1.24)
STHM1	Sweden	2005-2007	Population-based cohort	4,228	2,005	758	2,223	66.2 (62.1-71.5)	65.6 (61.4-71.2)	67.3 (62.5-73.2)	66.6 (62.7-71.6)	0.09 (-3.85-1.27)
TAMPERE	Finland	1993-2008	Population-based	2,754	2,754	1,642	-	67.5 (63.0-73.1)	67.5 (63.0-73.1)	68.7 (63.7-74.6)	-	0.19 (-0.64-1.05)
UKGPCS	UK	1993-2011	Hospital recruitment	5,287	4,497	3,083	790	60.3 (57.0-68.8)	62.9 (58.0-70.0)	63.8 (58.4-70.8)	56.0 (53.0-59.0)	0.18 (-2.31-1.35)
ULM	Germany	1998-2007	Hospital recruitment	800	592	406	208	63.1 (57.6-68.0)	63.8 (59.6-68.2)	64.1 (60.1-68.4)	58.0 (49.0-67.0)	0.16 (-0.86-1.30)
UTAH	USA	1991-2007	Population-based	685	440	68	245	64.0 (57.0-71.0)	63.0 (56.5-68.0)	64.0 (57.0-71.0)	68.0 (60.0-74.0)	0.16 (-0.83-1.07)
WUGS	USA	2004-2011	Hospital recruitment	944	944	592	-	61.0 (56.0-66.0)	61.0 (56.0-66.0)	62.0 (56.0-67.0)	-	0.29 (-0.62-2.43)
All				31,747	18,868	10,635	12,879	64.0 (58.2-70.1)	65.1 (59.9-70.5)	66.0 (60.1-71.3)	62.0 (55.0-69.6)	0.12 (-3.85-2.43)
Validation Set												
ProtecT ^c	UK	2001-2009	Population-based cohort	6,411	1,583	628	4,828	60.0 (55.7-64.4)	63.4 (59.0-67.0)	64.3 (60.2-67.5)	59.0 (55.0-63.0)	0.06 (-4.13-1.09)

^aIncludes the 31 cases and 3,364 controls who participated in both ProtecT and UKGPCS

^bCase-control design unless otherwise specified. More detailed descriptions of each study are provided in the supplementary material from the original iCOGS publication¹

Supplementary Table S2: SNPs in final PHS model

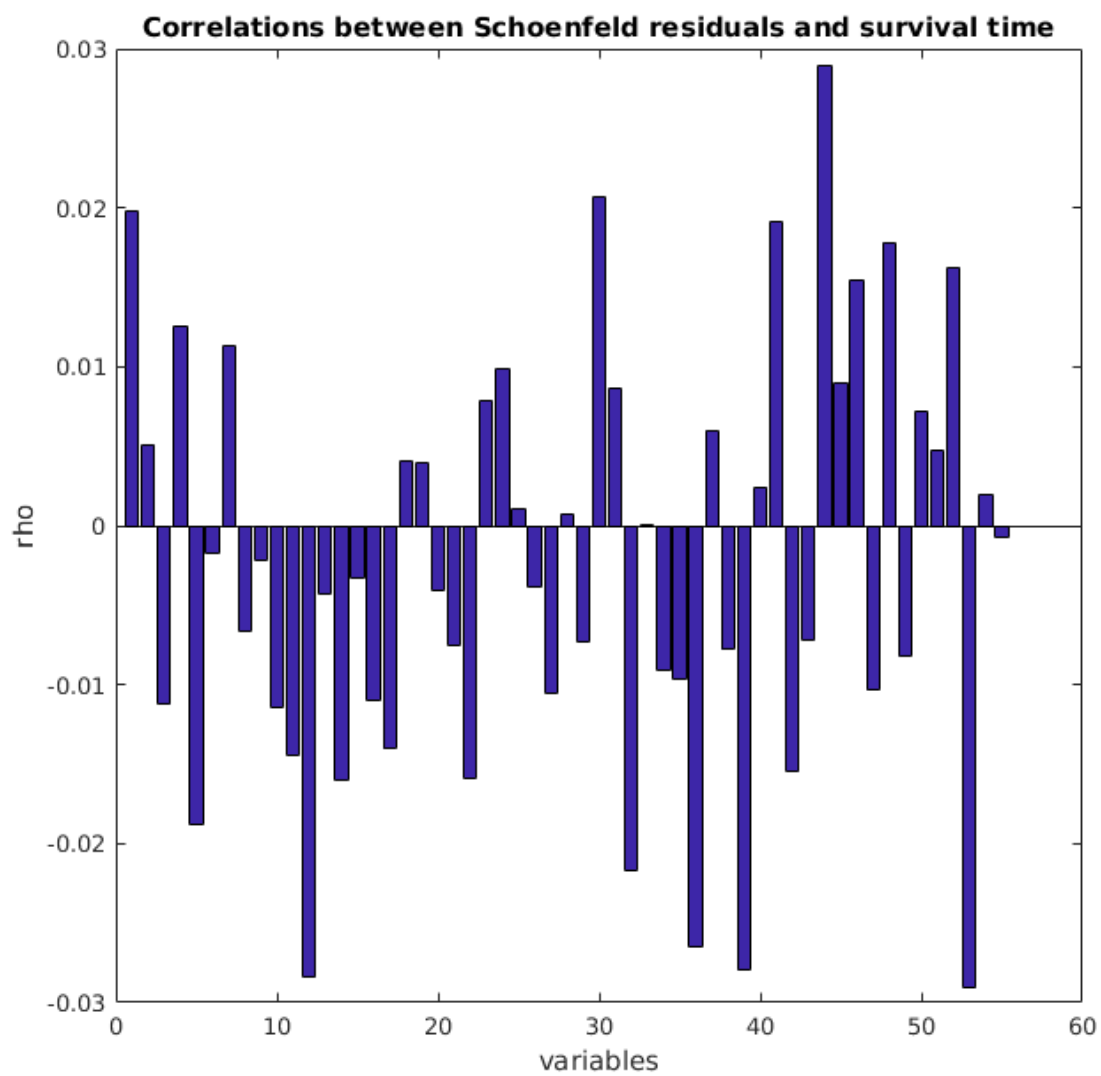
SNP name	log(p-value), univariate ^a	log(p-value), multivariate ^b	β from PHS
rs6983267 ^c	-53	-25	-0.095
c8_pos128146328	-48	-19	0.174
rs10993994 ^c	-48	-30	0.100
rs9297759	-42	-22	0.073
rs11651052	-37	-37	-0.093
rs12275055	-35	-9	-0.076
rs7929962	-32	-7	0.048
rs7679673 ^c	-27	-22	-0.066
rs7841060	-26	-21	-0.082
rs28556804	-25	-11	0.077
rs12549761	-25	-10	0.054
rs5945631	-24	-12	-0.192
rs9889335	-23	-22	0.077
c8_pos128389706	-22	-6	0.066
rs6545977	-22	-15	-0.066
rs13265330	-22	-12	-0.060
rs4907775	-21	-8	0.131
rs16860513	-20	-16	0.198
rs718961	-20	-6	-0.075
rs9297746	-19	-7	0.055
c17_pos44175675	-19	-10	0.142
rs17632542	-18	-9	0.140
rs232964	-18	-15	1.031
c11_pos2181240	-17	-14	0.068
rs7725218	-17	-16	-0.070
rs651164	-16	-8	-0.050
c3_pos171557211	-16	-13	0.073
rs6788616	-15	-7	-0.040
rs4643253	-14	-7	0.052
rs7769879	-14	-10	0.054
c10_pos8072007	-14	-11	-1.530
c3_pos87230612	-13	-7	-0.115
rs11672691 ^c	-13	-7	-0.059
rs2736108	-12	-12	0.050
rs6965016	-11	-9	-0.052
rs747745	-11	-5	0.044
rs3910736	-11	-9	-0.068

rs11568818 ^c	-10	-9	0.041
rs17596465	-10	-7	0.114
c22_pos41831564	-10	-6	0.084
rs1010	-10	-9	0.050
rs2136486	-9	-2	0.024
rs4919763	-9	-11	-0.050
rs10866528	-9	-7	-0.045
rs3861106	-9	-7	-0.914
rs4809311	-8	-6	0.049
rs6853490	-8	-6	-0.054
rs13252265	-8	-7	-0.055
rs4857841	-8	-7	0.029
rs11795627	-8	-7	-0.042
rs7888856	-7	-7	0.049
rs684232 ^c	-7	-8	-0.039
rs10875943 ^c	-7	-7	-0.041
rs10051795	-7	-8	-1.501

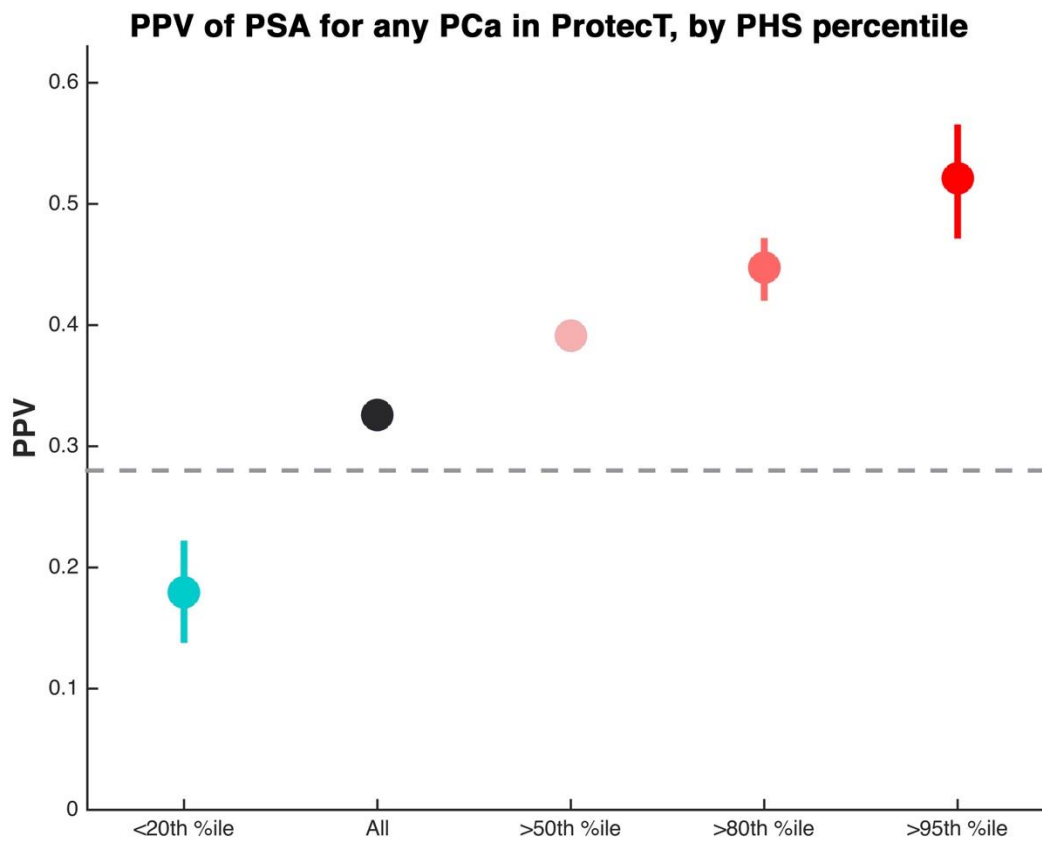
^aFrom trend test for this SNP only on Development Set case/control status.

^bFrom logistic regression for prediction of case/control with all SNPs in this table included as predictors, in addition to age and six principal components for European ancestry.

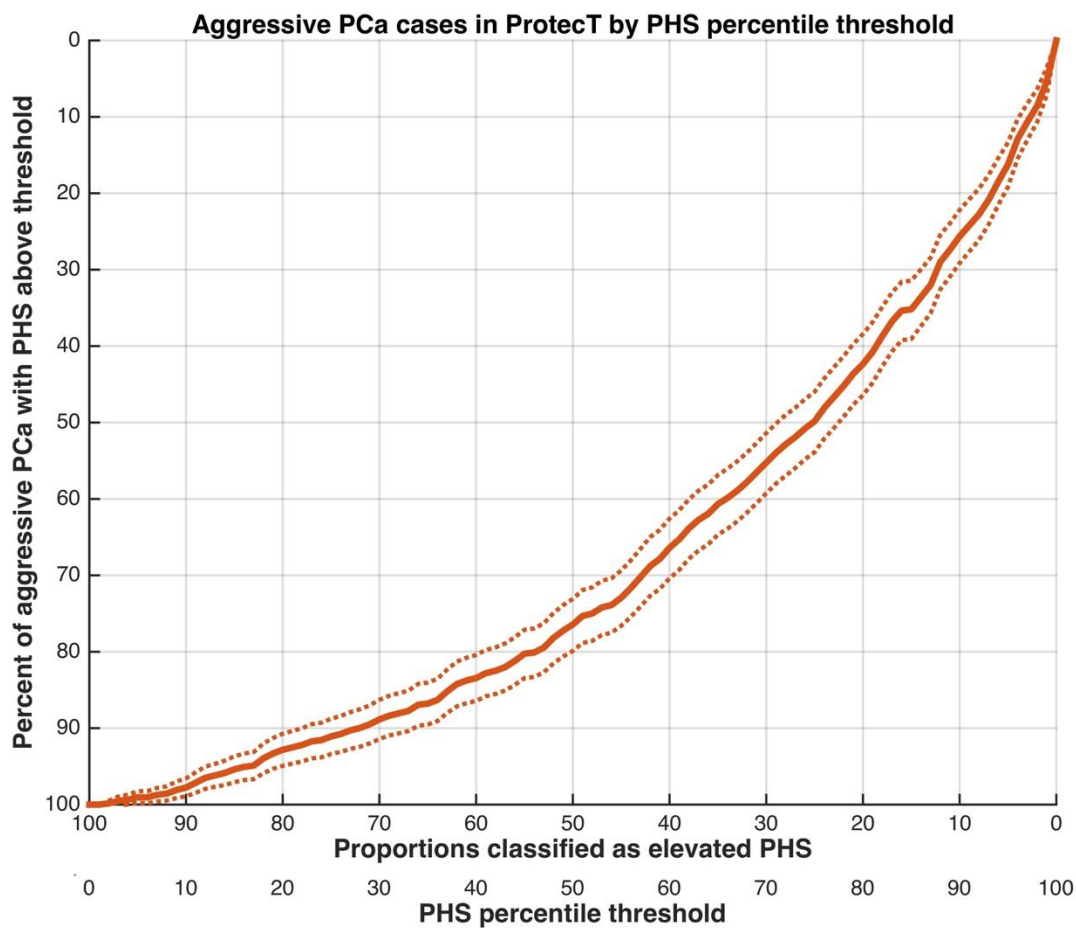
^cPreviously listed among 99 SNPs associated with prostate cancer in GWAS studies^{14,15}.



Supplementary Figure S1: Each column shows the rho value for Schoenfeld residuals for a single SNP (variable) in the final PHS model.¹⁶



Supplementary Figure S2: Positive predictive value (PPV) of PSA testing by PHS percentile thresholds for patients in the Validation Set. This is PPV for any PCa. Percentiles refer to the PHS distribution among young controls in the Development Set. Colored lines are 95% confidence intervals from random samples of cases in the Validation Set (see Methods). For reference, the expected PPV for PSA testing at this threshold is displayed as a gray, dashed line, based on a pooled analysis¹⁷.



Supplementary Figure S3: Lorenz curve to show the percent of the 632 aggressive PCa cases in the Validation Set (ProtecT) that were accounted for with various thresholds for PHS percentile. Dotted lines represent 95% confidence intervals calculated via 1,000 bootstrap samples of 632 aggressive cases. For example, the upper quintile of PHS (20 on upper x-axis, 80th PHS percentile) accounted for approximately 42% of all aggressive cases in the Validation Set.

References from Supplementary Material

- 1 Eeles RA, Olama AAA, Benlloch S, *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 2013; **45**: 385–91.
- 2 Eeles RA, Kote-Jarai Z, Al Olama AA, *et al.* Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat Genet* 2009; **41**: 1116–21.
- 3 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016; **66**: 7–30.
- 4 Desikan RS, Fan CC, Wang Y, *et al.* Personalized genetic assessment of age associated Alzheimers disease risk. *BioRxiv - Press PLOS Med* 2016; : 74864.
- 5 Borgan O, Goldstein L, Langholz B. Methods for the Analysis of Sampled Cohort Data in the Cox Proportional Hazards Model. *Ann Stat* 1995; **23**: 1749–78.
- 6 Lane JA, Donovan JL, Davis M, *et al.* Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. *Lancet Oncol* 2014; **15**: 1109–18.
- 7 Therneau TM, Li H. Computing the Cox Model for Case Cohort Designs. *Lifetime Data Anal*; **5**: 99–112.
- 8 R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2015 <https://www.R-project.org/>.
- 9 Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer, 2000.
- 10 Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011. *CA Cancer J Clin* 2011; **61**: 212–36.
- 11 Jemal A, Siegel R, Ward E, *et al.* Cancer Statistics, 2006. *CA Cancer J Clin* 2006; **56**: 106–30.
- 12 Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer Statistics, 2001. *CA Cancer J Clin* 2001; **51**: 15–36.
- 13 Arias E. United States life tables, 2008. Hyattsville, MD: National Center for Health Statistics, 2012 http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_03.pdf.
- 14 Eeles R, Goh C, Castro E, *et al.* The genetic epidemiology of prostate cancer and its clinical implications. *Nat Rev Urol* 2014; **11**: 18–31.
- 15 Al Olama AA, Kote-Jarai Z, Berndt SI, *et al.* A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 2014; **46**: 1103–9.
- 16 Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; **69**: 239–41.
- 17 Wolf AMD, Wender RC, Etzioni RB, *et al.* American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *CA Cancer J Clin* 2010; **60**: 70–98.