Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci

Ali Amin Al Olama[1,56], Sara Benlloch[1,56], Antonis C. Antoniou[1,56], Graham G. Giles[2,3,55], Gianluca Severi*[2,55], David E. Neal[4,5,55], Freddie C. Hamdy[6,55], Jenny L. Donovan[7,55], Kenneth Muir[8,9,55], Johanna Schleutker[10,11,55], Brian E. Henderson[12,55], Christopher Haiman[12,55], Fredrick R. Schumacher[12,55], Nora Pashayan[1,13,55], Paul D.P. Pharoah[1,55], Elaine A. Ostrander[14,55], Janet L. Stanford[15,16,55], Jyotsna Batra[17,55], Judith A. Clements[17,55], Suzanne K. Chambers[18,19,20,55], Maren Weischer[21,55], Børge G. Nordestgaard[21,55], Sue A. Ingles[12,55], Karina D. Sorensen[22,55], Torben F. Orntoft[22,55], Jong Y. Park[23,55], Cezary Cybulski[24,55], Christiane Maier[25,55], Thilo Doerk[26,55], Joanne L. Dickinson[27,55], Lisa Cannon-Albright[28,29,55], Hermann Brenner[30,31,55], Timothy R. Rebbeck[32,55], Charnita Zeigler-Johnson[33], Tomonori Habuchi[34,55], Stephen N. Thibodeau[35,55], Kathleen A. Cooney[36,55], Pierre O. Chappuis[37,55], Pierre Hutter[38,55], Radka P. Kaneva[39,55], William D Foulkes[40,55], Maurice P. Zeegers[41,55], Yong-Jie Lu[42,55], Hong-Wei Zhang[43,55], Robert Stephenson[44], Angela Cox[45], Melissa C. Southey[46], Amanda B. Spurdle[47], Liesel FitzGerald[48], Daniel Leongamornlert[49], Edward Saunders[49], Malgorzata Tymrakiewicz[49], Michelle Guy[49], Tokhir Dadaev[49], Sarah J. Little[49], Koveela Govindasami[49], Emma Sawyer[49], Rosemary Wilkinson[49], Kathleen Herkommer[50], John L. Hopper[3], Aritaya Lophatonanon[9,10], Antje E. Rinckleb[25], Zsofia Kote-Jarai[49,56]; Rosalind A. Eeles[49,51,56] on behalf of The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology[52], on behalf of The UK ProtecT Study Collaborators[53] and on behalf of The PRACTICAL Consortium[54]; and Douglas F. Easton[1,56]

[1] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Laboratory, Worts Causeway, Cambridge, UK

[2] Cancer Epidemiology Centre, the Cancer Council Victoria, 1 Rathdowne Street, Carlton, Victoria, Australia

[3] Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Victoria, Australia

[4] University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Hills Road Cambridge CB2 0QQ

[5] Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, UK

[6] Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK

[7] School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK

[8] The University of Manchester, Centre for Epidemiology, Institute of Population Health, Oxford Road, Manchester, M13 9PL, United Kingdom

[9] University of Warwick, University House, Coventry, CV4 8UW, United Kingdom

[10] Institute of Biomedical Technology/BioMediTech, University of Tampere and FimLab Laboratories, Tampere, Finland.

[11] Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

[12] Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Centre, Los Angeles, California, US

[13] University College London, Department of Applied Health Research, 1-19 Torrington Place, London, WC1E 7HB

[14] National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Rm. 5351, Bethesda, Maryland, USA

[15] Division of Public Health Sciences, Fred Hutchinson Cancer Research Centre, Seattle, Washington, 98109-1024, USA

[16] Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington, US

[17] Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia

[18] Griffith Health Institute, Griffith University, Gold Coast, QLD, Australia

[19] Cancer Council Queensland, Brisbane, Australia

[20] Prostate Cancer Foundation of Australia, Sydney, Australia[21] Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, DK-2730 Herlev, Denmark

[22] Department of Molecular Medicine (MOMA), Aarhus University Hospital, Brendstrupgaardsvej 100, 8200 Aarhus N, Denmark

[23] Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Centre, 12902 Magnolia Drive, Tampa, FL 33612, USA

[24] International Hereditary Cancer Centre, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland

[25] Department of Urology, University Hospital Ulm, Germany

[26] Hannover Biomedical Research School, Carl-Neuberg Str. 1, D- 30625 Hannover, Germany

[27] University of Tasmania, Menzies Research Institute Tasmania, Medical Sciences Building 2, 17 Liverpool Street, Private Bag 23, Hobart, Tasmania, Australia 7001

[28] Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA

[29] George E. Wahlen Department of Veterans Affairs Medical Centre, Salt Lake City, Utah, US

[30] Division of Clinical Epidemiology and Ageing Research, German Cancer Research Centre (DKFZ), Heidelberg, Germany

[31] German Cancer Consortium (DKTK), Heidelberg, Germany

[32] University of Pennsylvania. 217 Blockley Hall. 423 Guardian Drive. Philadelphia, PA. 19104., USA

[33] Division of Population Sciences, Department of Medical Oncology, Thomas Jefferson University, 834 Chestnut Street, Suite 314, Philadelphia, PA 19107, USA

[34] Department of Urology, Akita University School of Medicine, 1-1-1 Hondo, Akita 010-8543, Japan

[35] Mayo Clinic, Rochester, Minnesota, USA

[36] Division of Haematology/Oncology University of Michigan Medical School, Associate Director for Faculty Affairs, UMCCC, Michigan, US

[37] Divisions of Oncology and Genetic Medicine, Geneva University Hospitals, 30 Bd de la Cluse, 1211 Geneva 14, Switzerland

[38] Hopital Cantonal Universitaire de Geneve, 24 rue Micheli-du-Crest, 1211, Geneve 14, Switzerland

[39] Department of Medical Chemistry and Biochemistry, Molecular Medicine Centre, Medical University – Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria

[40] McGill University, Montreal, Quebec, Canada

[41] Department of Complex Genetics, Cluster of Genetics and Cell Biology, NUTRIM School for Nutrition, Toxicology and Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands

[42] Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, Charterhouse Square, London EC1M 6BQ

[43] Second Military Medical University, 800 Xiangyin Rd., Shanghai 200433, P. R. China

[44] Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, Utah, USA

[45] CR-UK/YCR Sheffield Cancer Research Centre, University of Sheffield, Sheffield, UK

[46] Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Grattan Street, Parkville, Victoria 3010, Australia

[47] Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia

[48] Cancer Council Victoria, Cancer Epidemiology Centre, 615 St Kilda Rd, Melbourne, Australia

[49] The Institute of Cancer Research, Cotswold Road, Sutton, Surrey, SM2 5NG, UK

[50] Department of Urology, Technical University Munich, Germany

[51] Royal Marsden NHS Foundation Trust, Fulham and Sutton, London and Surrey, UK

[52] UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, membership list provided in Supplementary Note

[53] UK ProtecT Study Collaborators, membership list provided in Supplementary Note

[54] Full list of participants of The PRACTICAL Consortium is provided in the Supplementary Note

[55] These authors made a co-equal contribution in their respective positions in this paper

[56] These authors jointly directed this work

* Current address:

Gianluca Severi

Human Genetics Foundation, Via Nizza 52, 10126 Torino - Italy

Corresponding authors:

Prof Douglas F. Easton

Centre for Cancer Genetic Epidemiology,

Strangeways Laboratory,

Worts Causeway, Cambridge, UK

dfe20@medschl.cam.ac.uk

Tel: 0044 1223 748629

Fax: 0044 1223 748628

Dr Ali Amin Al Olama

Centre for Cancer Genetic Epidemiology,

Strangeways Laboratory,

Worts Causeway, Cambridge, UK

aa461@medschl.cam.ac.uk

Tel: 0044 1223 748638

Fax: 0044 1223 748628

Running Title: Risk Analysis of Prostate Cancer in PRACTICAL Consortium

Keywords: Prostate Cancer risk, Genetic and Molecular Epidemiology, Genitourinary Cancers: Prostate, Risk Assessment, Methodology, Modelling and biostatistics, Methodology for SNP data analysis, Statistical methods in Genetics

Financial Support: D F. Easton was recipient of the CR-UK grant C1287/A10118. R A. Eeles was recipient of the CR-UK grant C5047/A10692 and B E. Henderson was recipient of the NIH grant 1U19CA148537-01.

Conflict of Interest: There is no conflict of interest.

Word Count: 3993 words

Total Table/Figure Count: 5

Background:

Genome-wide association studies have identified multiple genetic variants associated with prostate cancer (PrCa) risk which explain a substantial proportion of familial relative risk. These variants can be used to stratify individuals by their risk of PrCa.

Methods:

We genotyped 25 PrCa susceptibility loci in 40,414 individuals and derived a polygenic risk score (PRS). We estimated empirical Odds Ratios for PrCa associated with different risk strata defined by PRS and derived age-specific absolute risks of developing PrCa by PRS stratum and family history.

Results:

The PrCa risk for men in the top 1% of the PRS distribution was 30.6 (95% CI 16.4-57.3) fold compared with men in the bottom 1%, and 4.2 (95% CI 3.2-5.5) fold compared with the median risk. The absolute risk of PrCa by age 85 was 65.8% for a man with family history in the top 1% of the PRS distribution, compared with 3.7% for a man in the bottom 1%. The PRS was only weakly correlated with serum PSA level (correlation=0.09).

Conclusions:

Risk profiling can identify men at substantially increased or reduced risk of PrCa. The effect size, measured by OR per unit PRS, was higher in men at younger ages and in men with family history of PrCa. Incorporating additional newly identified loci into a PRS should improve the predictive value of risk profiles.

Impact:

We demonstrate that the risk profiling based on SNPs can identify men at substantially increased or reduced risk that could have useful implications for targeted prevention and screening programs.

Introduction

Genome-wide association studies (GWAS) have identified multiple common genetic variants associated with prostate cancer (PrCa) risk. The risks associated with such variants are generally modest, but in combination their effects may be substantial, and may provide the basis of targeted prevention (1). However, since the risks associated with these variants are modest, large studies are required to estimate their risks precisely. To facilitate this estimation, we genotyped 25 PrCa susceptibility SNPs in studies from the PRACTICAL consortium. PRACTICAL is an international PrCa consortium that includes more than 78 studies, including men of European, Asian or African ancestry, and has a combined dataset of over 130,000 samples (http://practical.ccge.medschl.cam.ac.uk/). In the current analysis, we utilised data from 31,833 cases and controls from 24 studies in PRACTICAL and 8,581 samples from replication stage of a GWAS ("GWAS stage 3"). Sixteen out of the twenty five SNPs that we used in this study were identified through studies that included PRACTICAL (2-4) and nine SNPs were identified by other GWAS (5-10).

Materials and Methods

Samples

The current analysis was restricted to individuals of European ancestry, based on self-reported ethnicity, and thus we excluded samples with non-European ancestry. Data were contributed from 25 studies in PRACTICAL and GWAS stage 3. Twenty five SNPs were genotyped specifically for this analysis in 31,833 cases and controls in PRACTICAL phase III, unless the genotype data were already available. We also included four studies from the GWAS stage 3 conducted in the United Kingdom and Australia, comprising a further 8,581 cases and controls (11). In this replication

stage 1,536 SNPs were genotyped, including the 25 susceptibility SNPs analysed here. These two datasets were combined to give a total of 40,414 samples (20,288 cases and 20,126 controls). Three studies (MCCS, PFCS and UKGPCS) that were included in the GWAS stage 3 also contributed genotyping of additional samples for PRACTICAL phase III (Table1, Supplementary Table 1 and Supplementary Notes). Studies provided a minimum core dataset that included disease status, age at diagnosis/observation and ethnicity. Twenty two studies provided data on family history and eighteen studies provided data on Gleason score.

Where studied included more than one individual from the same family, only the index case was included, so that the analyses were based on unrelated men. For analyses of the polygenic risk score (PRS) we also excluded 5 studies (MAYO, PCFS, TASPRAC, ULM and UTAH) that oversampled cases with family history of PrCa. This reduced the total number of samples to 34,986 (16,643 cases and 18,343 controls). All studies were approved by the relevant ethics committees.

Eighty Nine percent (31,150) of the samples had information on age at diagnosis (interview/blood draw for controls). The mean age at diagnosis for the cases was 64 years, slightly higher than the mean age at interview/blood draw for the controls (58 years; Supplementary Table 2a). Family history information was available for 21,209 (60.6%) samples and among samples with family history information, 10.7% of controls and 18.2% of cases had a family history of PrCa. Before excluding studies with oversampled familial cases, these percentages were 12.9% and 22.6% respectively (Supplementary Table 2a and b).

Genotyping

Genotyping was performed in two experiments; these were subject to separate QC procedures appropriate to the platforms used, before the data were combined for statistical analysis. In PRACTICAL phase III, genotyping of samples from 2 studies was performed by Sequenom, while 22 study sites performed the 5'exonuclease assay (Taqman™) using the ABI Prism 7900HT sequence detection system according to the manufacturer's instructions. Primers and probes were supplied directly by Applied Biosystems as Assays-By-Design™. Assays at all sites included at least four negative controls and 2-5% duplicates on each 384-well plate. Quality control guidelines were followed by all the participating groups as previously described (4). In addition, all sites also genotyped 16 CEPH samples. We excluded individuals that were not typed for at least 80% of the SNPs attempted. Data on a given SNP for a given site were also excluded if they failed any of the following QC criteria: SNP call rate >95%, no deviation from Hardy-Weinberg equilibrium in controls at $P<.00001$; <2% discordance between genotypes in duplicate samples and in the CEPH control samples. Cluster plots for SNPs that were close to failing any of the QC criteria were re-examined centrally.

GWAS Stage 3 genotypes were generated using an Illumina Golden Gate Assay. All SNPs for this analysis passed the QC filters used for this experiment: call rate>95%, a minor allele frequency in controls of >1%, or genotype frequency in controls consistent with Hardy-Weinberg equilibrium at $p<0.00001$. Duplicate concordance was 99.99% (11).

Statistical methods

We used combined data across all studies for the analysis. We assessed the association between each SNP and PrCa using a 1-degree-of-freedom Cochran-Armitage trend test, stratified by studies. Odds ratios (OR) and 95% confidence intervals (95% CI) associated with each genotype and cancer risk, and genotypes for pairs of SNPs, were estimated using unconditional logistic regression, stratified by study as a covariate. Both per-allele ORs, and genotype-specific ORs, were estimated. Heterogeneity in the OR estimates among studies was evaluated using a likelihood ratio test, by comparing with a model in which separate ORs were estimated for each study.

Modification of the ORs by disease aggressiveness and family history was assessed by using both family history (Yes *vs.* No) and Gleason score (<8 *vs.* ≥8) as binary variables. A test for association between SNP genotype at a locus and Gleason score as an ordinal variable was also performed, using polytomous regression. Modification of the ORs by age was assessed using a case-only analysis, assessing the association between age and SNP genotype in the cases using polytomous regression. The associations between SNP genotypes and PSA level were assessed using linear regression, after log-transformation of PSA level to correct for skewness.

Contribution to Familial Risk

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

$$\sum_k (\log \lambda_k) \Big/ (\log \lambda_0)$$

where λ_0 is the observed familial risk to first degree relatives of PrCa cases, assumed to be 2 (12), and λ_k is the familial relative risk due to locus k, given by:

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

where p_k is the frequency of the risk allele for locus k, $q_k = 1 - p_k$ and r_k is the estimated per-allele odds ratio (13).

To evaluate evidence for interactions between pairs of SNPs, we used a likelihood ratio test and evaluated the evidence for departures from a multiplicative model, by comparing models with and a model without the interaction term for each pair of SNPs. The interaction term was the product of the allele doses for the two SNPs, hence leading to a 1 degree of freedom test for an interaction. Based on the assumption of a log-additive model, we constructed a PRS from the summed genotypes weighted by the estimated per-allele log-odds ratios for each SNP, as estimated by logistic regression as above. Thus for each individual *j* we derived:

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

Where:

N : Number of SNPs (25)

g_{ij} : Allele dose at SNP i (0, 1, 2) for individual j

β_i : Per-allele log-odds ratio of SNP i

The missing genotypes for an individual were replaced with the mean genotype of each SNP separately for cases and controls. A sensitivity analysis, in which analyses were based on samples with complete genotype data, gave very similar results (data not shown). We then standardised the PRS by dividing by the overall standard deviation of PRS in the controls.

The risk of PrCa was estimated for the percentiles of the distribution of the PRS; <1%, 1-10%, 10-25%, 25-75% (defined here as “median risk”), 75-90%, 90-99%, >99%; and per standard deviation when fitted as a continuous covariate. We evaluated the fit of the combined risk score to a log-linear model by comparing the model with the PRS fit as a continuous covariate with a model in which separate parameters were estimated for percentiles of risk adjusted for age at diagnosis and family history, using a likelihood ratio test.

We used a likelihood ratio test to evaluate the evidence for interaction between PRS and age at diagnosis/observation, PRS and family history and also family history and age at diagnosis/observation by comparing models with and a model without an interaction term. Effect sizes by family history were compared using a case-only analysis. Analyses were performed using Stata 13.

The relative risk estimates were used to obtain estimates of the absolute risk of PrCa by PRS category and family history. Since we observed evidence for an interaction

between PRS and age, we used both models with and without PRS x age interaction term. Absolute risks were constrained such that the age-specific incidences, averaged over all categories of PRS and family history, were consistent with the age-specific incidences of PrCa for the UK population for 2012 (http://ci5.iarc.fr/CI5plus) (14). The model was adjusted for age at diagnosis (age <55, 55-59, 60-64, 65-70 and 70+). The procedure for deriving the age-specific incidences for each SNP profile category has been performed following the procedure explained by Antoniou et al. (15, 16), but adjusted to allow for competing causes of death.

For this purpose, we categorised PRS into seven risk groups (k=risk group 1 to 7), based on the percentile in the controls: <1%, 1-10%, 10-25%, 25-75%, 75-90%, 90-99% and >99%. We could not find any evidence for an interaction between PRS and family history of PrCa (*P*-value=0.49) and assumed that family history and PRS are independently predictive of PrCa risk. Under this model, the PrCa incidence $\lambda_k^h(t)$ at age *t* for an individual in risk group k and family history group h (h=1 with family history, h=0 no family history) was assumed to follow a model of the form: $\lambda_k^h(t) = \lambda_0(t)\exp(\beta_k^h)$ where $\lambda_0(t)$ is the baseline PrCa incidence and $\exp(\beta_k^h)$ is the risk ratio in the risk group k and family history group h, relative to the baseline category (*h*=0, *k*=1), approximated by the odds ratio estimates from the logistic regression analysis. To obtain the baseline incidence, $\lambda_0(t)$, we constrained the PrCa incidence averaged all risk groups to agree with the population age-specific PrCa incidences $\mu(t)$ (the incidence of PrCa at age t per 100,000 individuals in the UK (14)). The baseline incidence can be obtained for each age by:

$$\lambda_0(t) = \mu(t)\frac{p_0\sum_1^k f_k S_k^0(t-1) + p_1\sum_1^k f_k S_k^1(t-1)}{p_0\sum_1^k f_k S_k^0(t-1)\exp(\beta_k^0) + p_1\sum_1^k f_k S_k^1(t-1)\exp(\beta_k^1)}$$

Here p_0 is the probability of having no family history in the population (89.26% in the controls in this dataset) and $p_1 = 1 - p_0$ is the probability of having family history in the population (10.74% in the controls in this dataset). f_k is frequency of the SNP profile risk group k (f_1=0.01, f_2 =0.09, f_3 =0.15, f_4 =0.5, f_5 =0.15, f_6 =0.09, f_7 =0.01) and $S_k^h(t)$ is the probability of surviving PrCa by age (t) in the risk group k for samples in the family history group h, which can be derived from incidence rates $\lambda_k^h(t)$ for ages <t using the formula $S_k^h(t) = \exp(-\sum_0^t \lambda_k^h(t-1))$. Since definition $S_k^h(0) = 1$ for all k and h, it was possible to solve the above equation recursively, starting at age *t*=0, to obtain the baseline incidences $\lambda_0(t)$ and hence the age-specific PrCa incidences at age (t), $\lambda_k^h(t)$, for each group. We then computed the absolute risk by age t, adjusting for mortality from other causes, for each risk group, using the formula: $\sum_0^t S_k^h(t) \times \lambda_k^h(t) \times S_c(t)$

Where $S_c(t) = \exp(-\sum_0^t \mu_c(t-1))$ is the probability of not dying from another cause of death by age t, based on the age-specific mortality rates $\mu_c(t)$. The age-specific mortality rates, $\mu_c(t)$, was estimated by using all causes incidences of death per 100,000 individuals for England and Wales (http://www.ons.gov.uk/ons/index.html) and the PrCa death incidence per 100,000 individuals in UK in year 2012 (14).

Results

All 25 SNPs showed evidence of association with PrCa (P=0.02 to P=1.4x10^{-46}), with effect sizes that were consistent with previous reports. The largest per-allele OR estimate was 1.56 (95% CI 1.44-1.68) for rs16901979 on 8q24 (Table 2). For each of the 24 autosomal SNPs, the effect size was larger for rare homozygotes than for heterozygotes, and the estimates were consistent with a multiplicative (log-additive) model. There was no evidence for heterogeneity among studies (Table 2).

Gleason score was available for 15,107 (74.5%) of the cases used in the analyses; of these, 2,139 had a score of 8+ and 12,968 had a score less than 8. One SNP, rs1447295, on chromosome 8, showed a larger effect size with increasing grade (P=0.001), while four SNPs (rs17021918, rs1512268, rs7127900 and rs2735839) showed a larger effect sizes with decreasing grade (P<0.02; Supplementary Table 3).

Thirteen of the SNPs (rs1465618, rs7679673, rs10486567, rs1447295, rs6983267, rs16901979, rs10993994, rs7931342, rs7127900, rs4430796, rs11649743, rs1859962 and rs5759167) showed a higher per-allele OR for cases with a PrCa family history than those without (P<0.05), while no SNPs showed an effect in the opposite direction consistent with the predictions under a polygenic model (17) (Supplementary Table 3).

Data on serum PSA level were available for 3,922 controls from 6 studies. Six SNPs (rs1447295, rs6983267, rs1512268, rs10993994, rs7127900 and rs2735839)

showed association with PSA concentration levels significant at *P*-value < 0.03. rs1447295 showed an association with PSA in the opposite direction of the PrCa risk association but the rest of five SNPs showed an association with PSA in the same direction of the PrCa risk association (Supplementary Table 4).

Seven SNPs (rs1465618, rs12621278, rs10993994, rs7127900, rs1859962, rs2735839 and rs5945619) showed an evidence for a trend in the per-allele ORs with age; in each case the effect size was larger for cases diagnosed at younger ages (Supplementary Table 5).

The combined effect of all pairs of SNPs was evaluated through a logistic regression model that included each pair of SNPs and an interaction term. The interaction term was significant at *P*-value <0.05 level for 29 pairs (out of 300 possible pairs) compared with 15 expected by chance, and significant at the *P*-value <0.01 level for 12 pairs compared with 3 expected by chance. However, no pair was significant at the *P*-value <0.05 level after a bonferroni correction for the number of tests (nominal significance *P*-value=1.6×10^{-4}, Supplementary Table 6).

Under the assumption that these 25 SNPs combined approximately multiplicatively to alter the risk of PrCa, we constructed a PRS for 16,643 cases and 18,343 controls based on the estimated per-allele ORs of 25 SNPs, standardised by the standard deviation in controls. The standardised PRS had a mean=0.651 (range -3.81-5.36; SD=0.98) in cases and mean=0.104 (range -4.05-4.15; SD=1) in controls. The standardised PRS was strongly associated with disease risk (OR per unit PRS =1.74, 95%CI 1.70-1.78). The OR per unit increase of the standardised PRS

declined with age from 1.76 (95% CI 1.62-1.92) in cases diagnosed at age less than 55 to 1.48 (95% CI 1.37-1.60) in cases diagnosed at age 70+ (*P*-value= 2.6×10^{-4}, Supplementary Table 5).

The OR per unit increase of PRS was larger for men with PrCa family history (1.79 Vs 1.70; *P*-value= 1.8×10^{-4}, Supplementary Table 3). We found no evidence of an interaction between PRS and family history (*P*-value=0.49) or between age at diagnosis and family history (*P*-value=0.11) but there was some evidence for an interaction between PRS and age at diagnosis (*P*-value=0.003).

There was no evidence of a difference in the OR per unit PRS according to Gleason Score (OR=1.75, GS<8 Vs OR=1.65, GS 8+) after adjusting for age at diagnosis and family history (*P*=0.37; Supplementary Table 3). The correlation between PSA and the PRS was weak, both in controls (correlation=0.09) and in cases (correlation =0.02).

When PRS was categorised by percentile, the top 1% of the population had an estimated OR of 30.6 (16.4-57.3) compared with the bottom 1% of the population, and an OR of 4.2 (95%CI 3.2-5.5) compared with the median population risk (defined as the 25-75% risk group). The bottom 1% of the population had an estimated OR of 0.14 (95% CI 0.08-0.24) compared with the median risk (Table 3). After allowing for an interaction between PRS and age, the OR for the top 1% of the population, relative to the median risk group, decreased from 5.6, for men below age <55 years, to 3.8 for men aged 70+ years (Supplementary Table 7 & 8).There was no difference between fit of the model with a continuous covariate for PRS and the

model with separate parameters for percentiles of the PRS (*P*=0.24). In particular, the predicted ORs for the top 1% and the bottom 1% of the population, based on a log-linear model, did not differ from that observed.

To estimate the absolute risk of PrCa for different risk groups defined by the combined genotypes at the 25 PrCa susceptibility loci, we fitted a logistic regression model that It included parameters for PRS (in 7 categories) together with family history of PrCa once with (Supplementary Table 7) and once without a PRS x age at diagnosis interaction term (Table 3). We used both models (adjusted for age at diagnosis and family history) in order to estimate effect sizes for PRS. Then we used the UK age-specific incidences of PrCa (0 to 85+ years) (14) to estimate age-specific absolute risks of PrCa in the general population after considering competing causes of death for fourteen risk groups defined by PRS and family history (seven PRS risk groups and two family history, see methods). Based on this analysis, the absolute risk of PrCa by age 85 for a man in the top 1% of the risk distribution with family history of PrCa was 65.8% (67.1% in a model not allowing for interaction) and for a man in the lowest 1% was 3.65% (3.67% in a model not allowing for interaction). The absolute risk for a man in the top 1% of the risk distribution with no family history of PrCa was 35.0% (36.1% in a model not allowing for interaction) and 1.46% (1.47% in a model not allowing for interaction) for someone in the lowest 1%. By comparison, the estimated absolute risk for a man in the 25-75% category was 10.2% in the absence of a family history of PrCa, and 23.7% for a man with family history (Figure 1 & 2, Supplementary Figure 1 & 2).

Discussion

These results demonstrate that risk profiling based on SNPs can identify men at substantially increased or reduced risk of PrCa. We derived a PRS based on a sum of SNP genotypes, weighted by their per-allele log ORs. The estimated ORs for the highest and lowest 1% of the population (4.2 and 0.14, respectively) were consistent with those predicted under a simple polygenic model in which the log OR increases linearly with the PRS. We also showed that the effect size, measured by OR per unit PRS, was higher at younger ages. As expected, the majority of loci, and the PRS, showed a stronger effect for familial cases. In a logistic regression model, both PRS and family history were independently associated with PrCa risk. The OR due to family history was attenuated after adjustment for the PRS (from 2.63 to 2.50), as expected given that family history is, at least in part, a reflection of genetic susceptibility. However, the degree of attenuation (5% on a log-scale) was markedly less than 18%, the estimated contribution of these 25 loci to the familial risk of PrCa estimated based on their ORs and allele frequencies in this study (see methods). The reason for this difference is unclear but might reflect interactions between the known susceptibility loci summarised in the PRS and other factors influencing family history.

In order to investigate the added value of PRS, once we estimated the absolute risk for individuals with family history without fitting their PRS information and then repeated the same procedure after adding their PRS information. The absolute risk of PrCa for a man at age 85 with family history was estimated to be 26.5% when PRS information was ignored. When we incorporated PRS information, a man at age 85 ,depend on his PRS risk group, could have an absolute risk ranging from 3.67%

(if a man is in the bottom 1% of the risk distribution) to 67.1% (if a man is in the top 1% of the risk distribution, Supplementary Figure 1 and 3). These observations indicate that family history and the PRS independently influence risk and can be combined to provide stronger discrimination.

Chatterjee et al. derived theoretical estimates for the predictive performance of polygenic models for ten complex traits or common diseases, including PrCa, using published estimates for individual SNPs (18). They estimated that ~7% of the population will be at two-fold risk or greater for PrCa. We estimated, empirically, that the (average) risk to men in the 90-99% category of the PRS was 2.41 fold, relative to the population median, or approximately 2 fold relative to the population mean. However, this is an average risk over the 90-99% category, so that the percentile of the PRS at which the risk exceeds 2 fold will be >90%. Based on the estimated log(OR) per standardised PRS, approximately 6% of men will have a risk of greater than twofold, very close to the estimate of Chatterjee et al (18).

These results show that genetic risk profiling using SNPs could be useful in defining men at high risk for the disease for targeted prevention and screening programs. The benefits of screening, relative to the costs, will be most favourable among men at higher risk. If, for example, the benefit-cost ratio is favourable for screening men at a greater than two-fold risk, the PRS provides an effective method for identifying such men.

While these analyses demonstrate the value of SNPs for risk prediction, a risk model could be improved in various ways. The analyses presented here are based on the

25 loci first identified to be associated with PrCa. Recently, however, additional loci have been identified (13, 19) and more than 100 common susceptibility loci are now known. In total, these loci increase the estimated proportion of the familial risk to 33% (19). Incorporating all known loci into a PRS should improve the predictive value of risk profiles.

Additionally, the analyses presented here consider family history as a binary (yes/no) covariate. It is known that the risk of PrCa is dependent on both the number of affected relatives and their ages. MacInnis et al. (12, 20) have shown using segregation analysis that the familial aggregation of PrCa can be modelled as the combined effect of a recessive allele and a polygenic component, and that the polygenic component can be further partitioned into a component due to measured SNPs and an unmeasured component. This approach should provide more powerful prediction, particularly in families with multiple cases of the disease. Finally, it is known that serum or urine PSA level is associated with PrCa risk, with the association persisting for several decades. Although some of the risks SNPs are also related to PSA level in the expected direction, the PSA level is only weakly correlated with PRS, indicating that incorporating PSA level and potentially other markers such as MSMB (21) into a risk algorithm should further improve the discrimination (22).

The absence of clear differences in the relative risk associated with SNPs by disease aggressiveness, even in this very large study, is striking. We did not find any convincing evidence for differences in the predictive values of the PRS by disease

aggressiveness. The effect size was higher for less aggressive disease, but the difference was still small (1.75 vs. 1.65). This result is in contrast to the clear differences in SNP associations by disease pathology seen in other diseases, for example in breast and ovarian cancer, and indicates that aggressive and non-aggressive disease, at least as measured by Gleason score, share these genetic risk factors as a common aetiology.

Analysis of pairwise combinations of SNPs did not identify any clear examples of departure from a multiplicative model, after adjusting for multiple testing. We did, however, find an excess of interactions at the $P<0.01$ level over the number that would be expected by chance. This suggests that interactions on this scale likely to exist, but their effect sizes are small and that very large sample sizes, exemplified by this collaborative study, will be required to identify and characterise them. If such interactions could be identified reliably, they may improve the predictive value of the risk profiling, and also provide insights into the biological interactions between the underlying risk variants.

Acknowledgments

We thank all the patients and control men who took part in this study. We also thank all the members from studies that they participated in this phase of PRACTICAL consortium: Esther John, Amit Joshi and Ahva Shahabi, Joanne L Dickinson, James R. Marthick, Mariana C. Stern, Roman Corral, David M A Wallace, Alan Doherty, R I Bhatt, K Subramonian, John Arrand, Louise Flanagan, Sita Ann Bradley, The UK Genetic Prostate Cancer Study Collaborators (www.icr.ac.uk/ukgpcs), Prasad Bollina, Sue

Bonnington, Lynne Bradshaw, James Catto, Debbie Cooper, Liz Down, Andrew Doble, Alan Doherty, Garrett Durkan, Emma Elliott, David Gillatt, Pippa Herbert, Peter Holding, Joanne Howson, Mandy Jones, Roger Kockelbergh, Rajeev Kumar, Peter Holding, Howard Kynaston, Athene Lane, Teresa Lennon, Norma Lyons, Hing Leung, Malcolm Mason, Hilary Moody, Philip Powell, Alan Paul, Stephen Prescott, Derek Rosario, Patricia O'Sullivan, Pauline Thompson, Lynne Bradshaw, Sarah Tidball, Paul M. Brown, Anne George, Gemma Marsden, Athene Lane, Michael Davis, Stephen Edwards, Cyril Fisher, Charles Jameson, Elizabeth Page, John Pedersen, Joanne Aitken, Robert A. Gardiner, Srilakshmi Srinivasan, Felicity Lose, Mary-Anne Kedda, Kimberly Alexander, Tracy O'Mara, Gail Risbridger, Wayne Tilley, Lisa Horvarth, Peter Heathcote, Glenn Wood, Greg Malone, Hema Samaratunga, Pamela Saunders, Allison Eckert, Trina Yeadon, Kris Kerr, Angus Collins, Megan Turner, Simon J. Foote, James R. Marthick, Andrea Polanowski, Rebekah M McWhirter, Terrence Dwyer, Christopher L. Blizzard, Elenko Popov, Darina Kachakova, Atanaska Mitkova, Teodora Goranova, Gergana Stancheva, Olga Beltcheva, Rumyana Dodova, Aleksandrina Vlahova, Tihomir Dikov, Svetlana Christova, Michael Borre, Peter Klarskov, Sune F Nielsen, Peter Iversen, Andreas Røder, Stig E Bojesen, Aida Karina Dieffenbach, Manuel Luedeke, Mark Schrader, Josef Hoegel, Walther Vogel, Liisa Määttänen, Teuvo Tammela, Anssi Auvinen, Lori Tillmans, Shaun Riska, Liang Wang, Dan Stram, Kolonel Laurence N., Julio Pow-Sang, Hyun Y. Park, Selina Radlein, Maria Rincon, Babu Zachariah (Supplementary Notes).

Reference List

1. 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;104:1656-63.
2. Amin Al Olama A, Kote-Jarai Z, Giles GG, Guy M, Morrison J, Severi G, et al. Multiple loci on 8q24 associated with prostate cancer susceptibility. Nat Genet. 2009;41:1058-60.
3. Eeles RA, Kote-Jarai Z, Amin Al Olama A, Giles GG, Guy M, Severi G, et al. Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. Nat Genet. 2009;41:1116-21.
4. Eeles RA, Kote-Jarai Z, Giles GG, Amin Al Olama A, Guy M, Jugurnauth SK, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet. 2008;40:316-21.
5. Amundadottir LT, Sulem P, Gudmundsson J, Helgason A, Baker A, Agnarsson BA, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet. 2006;38:652-8.
6. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet. 2007;39:631-7.
7. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, Gudbjartsson D, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet. 2008;40:281-3.
8. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet. 2007;39:977-83.
9. Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet. 2007;39:638-44.
10. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, Orr N, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet. 2008;40:310-5.
11. Kote-Jarai Z, Amin Al Olama A, Giles GG, Severi G, Schleutker J, Weischer M, et al. Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. Nat Genet. 2011;43:785-91.
12. MacInnis RJ, Antoniou AC, Eeles RA, Severi G, Guy M, McGuffog L, et al. Prostate cancer segregation analyses using 4390 families from UK and Australian population-based studies. Genet Epidemiol. 2010;34:42-50.
13. Eeles RA, Amin Al Olama A, 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;45:385-2.
14. Ferlay J, Soerjomataram II, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2014. Epub 2014/09/16.
15. Antoniou AC, Beesley J, McGuffog L, Sinilnikova OM, Healey S, Neuhausen SL, et al. Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. Cancer Res. 2010;70:9742-54.

16. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol. 2001;21(1):1-18.
17. Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol. 2003;25(3):190-202.
18. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nat Genet. 2013;45(4):400-5.
19. Amin Al Olama A, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, et al. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. 2014. PubMed PMID: 25217961.
20. MacInnis RJ, Antoniou AC, Eeles RA, Severi G, Amin Al Olama A, McGuffog L, et al. A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact. Genet Epidemiol. 2011.
21. Whitaker HC, Kote-Jarai Z, Ross-Adams H, Warren AY, Burge J, George A, et al. The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. PLoS One. 2010;5:e13363.
22. Aly M, Wiklund F, Xu J, Isaacs WB, Eklund M, D'Amato M, et al. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. Eur Urol. 2011;60:21-8.

Table 1: Total numbers of cases and controls used in the analyses

Study	Controls[a]	Cases[a]	Total[a]
GWAS Stage 3	4,076	4,505	8,581
PRACTICAL	16,050	15,783	31,833
Total	20,126	20,288	40,414
Total[b]	18,343	16,643	34,986

[a] Analyses were restricted to men of European ancestry (see text).

[b]Total after excluding 5 studies that oversampled cases with family history.

Table 2: Summary results of 25 SNPs using PRACTICAL and GWAS Stage 3 datasets in European.

Marker[a] Chr/Nearby Gene	Alleles[b] Position[c]	MAF[d]	Per allele[e] OR (95%CI)	Het OR [e,f] (95%CI)	Hom OR[e,g] (95%CI)	P-value[h]	P-value[i]
rs721048 2 / *EHBP1*	C/T 63131731	0.18	1.11 (1.07-1.16)	1.09 (1.04-1.15)	1.32 (1.17-1.48)	$9.8x10^{-8}$	0.13
rs1465618 2 / *THADA*	G/A 43553949	0.2150	1.07 (1.03-1.11)	1.08 (1.03-1.13)	1.14 (1.03-1.26)	$1.9x10^{-4}$	0.39
rs12621278 2 / *ITGA6*	A/G 173311553	0.06	.75 (.70-.80)	.76 (.71-.82)	.38 (.24-.58)	$4.9x10^{-17}$	0.57
rs2660753 3 / *Unknown*	G/A 87110674	0.10	1.12 (1.06-1.18)	1.12 (1.06-1.19)	1.32 (1.09-1.61)	$1.2x10^{-5}$	0.73
rs17021918 4 / *PDLIM5*	G/A 95562877	0.35	.88 (.85-.91)	.86 (.83-.90)	.80 (.74-.85)	$6.7x10^{-15}$	0.39
rs12500426 4 / *PDLIM5*	G/T 95514609	0.46	1.10 (1.06-1.13)	1.11 (1.06-1.18)	1.20 (1.12-1.28)	$4.8x10^{-8}$	0.54
rs7679673 4 / *TET2*	C/A 106061534	0.40	.88 (.85-.90)	.87 (.83-.91)	.77 (.72-.82)	$1.0x10^{-16}$	0.08
rs9364554 6 / *SLC22A3*	C/T 160833664	0.29	1.10 (1.06-1.14)	1.12 (1.07-1.18)	1.18 (1.09-1.27)	$4.8x10^{-8}$	0.85
rs10486567 7 / *JAZF1*	G/A 27976563	0.23	.85 (.82-.89)	.86 (.81-.91)	.72 (.63-.81)	$4.5x10^{-12}$	0.21
rs6465657 7 / *LMTK2*	A/G 97816327	0.46	1.10 (1.06-1.13)	1.09 (1.04-1.15)	1.21 (1.13-1.28)	$3.4x10^{-9}$	0.32
rs1447295 8 / *Unknown*	G/T 128485038	0.11	1.41 (1.35-1.48)	1.41 (1.34-1.49)	2.01 (1.69-2.41)	$1.4x10^{-46}$	0.50
rs6983267 8 / *Unknown*	C/A 128413305	0.49	.82 (.79-.85)	.80 (.76-.85)	.67 (.63-.72)	$2.3x10^{-35}$	0.61
rs16901979 8 / *Unknown*	G/T 128124916	0.03	1.56 (1.44-1.68)	1.55 (1.43-1.69)	2.39 (1.47-3.86)	$3.8x10^{-28}$	0.29
rs2928679 8 / *SLC25A37*	C/T 23438975	0.48	1.04 (1.01-1.07)	1.03 (.97-1.09)	1.08 (1.01-1.16)	.02	0.10
rs1512268 8 / *NKX3.1*	G/A 23526463	0.43	1.13 (1.10-1.17)	1.13 (1.08-1.19)	1.29 (1.21-1.37)	$2.6x10^{-16}$	0.19
rs4962416 10 / *CTBP2*	A/G 126696872	0.28	1.04 (1.01-1.08)	1.03 (.98-1.08)	1.11 (1.02-1.21)	.02	0.68
rs10993994 10 / *MSMB*	G/A 51549496	0.39	1.24 (1.20-1.28)	1.21 (1.15-1.27)	1.56 (1.46-1.66)	$7.9x10^{-41}$	0.36
rs7931342 11 / *Unknown*	C/A 68994497	0.50	.84 (.81-.86)	.86 (.82-.91)	.70 (.65-.74)	$4.8x10^{-27}$	0.86
rs7127900 11 / *Unknown*	G/A 2233574	0.19	1.23 (1.18-1.28)	1.24 (1.18-1.30)	1.47 (1.32-1.65)	$6.3x10^{-26}$	0.63
rs4430796 17 / *HNF1B*	A/G 36098040	0.48	.81 (.79-.84)	.81 (.77-.85)	.66 (.62-.71)	$2.7x10^{-38}$	0.79
rs11649743 17 / *HNF1B*	G/A 36074979	0.19	.88 (.85-.92)	.88 (.83-.92)	.79 (.70-.90)	$5.6x10^{-10}$	0.25
rs1859962 17 / *Unknown*	T/G 69108753	0.48	1.17 (1.14-1.21)	1.22 (1.15-1.28)	1.38 (1.30-1.47)	$3.7x10^{-24}$	0.19
rs2735839 19 / *KLK2/KLK3*	G/A 51364623	0.15	.81 (.77-.85)	.82 (.78-.86)	.62 (.53-.73)	$1.1x10^{-19}$	0.06
rs5759167 22 / *BIL/TTLL1*	G/T 43500212	0.50	.84 (.82-.87)	.83 (.79-.87)	.71 (.67-.76)	$3.4x10^{-28}$	0.87
rs5945619 X / *NUDT11*	T/C 51241672	0.36	1.13 (1.10-1.16)	-	1.28 (1.22-1.35)	$1.9x10^{-20}$	0.10

[a]dbSNP rs number, [b]Major/minor allele, based on the frequencies in controls in PRACTICAL III data, [c]Build 37 position, [d] MAF in controls in combined European dataset. [e]OR = odds ratio (minor allele) from a logistic regression using all European samples stratified by studies with no adjustment, [f]OR in heterozygotes, relative to major allele homozygotes, [g]OR in minor allele homozygotes, relative to major allele homozygotes, [h]Cochran-Armitage test for trend. [i] Heterogeneity *P*-value among studies

Table 3: Odds ratios for PrCa by percentile of the PRS and family history.

Percentiles		OR[a,b]	OR[a,c]
PRS Group	< 1%	1 (baseline)	0.14 (0.08-0.24)
	1-10%	2.98 (1.66-5.35)	0.41 (0.36-0.47)
	10-25%	4.59 (2.58-8.17)	0.63 (0.57-0.70)
	25-75%	7.23 (4.08-12.80)	1 (baseline)
	75-90%	12.13 (6.83-21.54)	1.68 (1.54-1.83)
	90-99%	16.70 (9.38-29.72)	2.31 (2.09-2.56)
	>= 99%	30.63 (16.36-57.34)	4.24 (3.24-5.53)
Family History		2.52 (2.29-2.78)	2.52 (2.29-2.78)

[a]ORs obtained by fitting PRS group, family history and age at diagnosis jointly.
[b]ORs compared to men in the 1st percentile as baseline.
[c]ORs compared to men in the 25th-75th percentile as baseline.

Figure Legend

Figure 1: Absolute risk of PrCa by age in men with family history.

Figure 2: Absolute risk of PrCa by age in men with no family history.

Figure 1

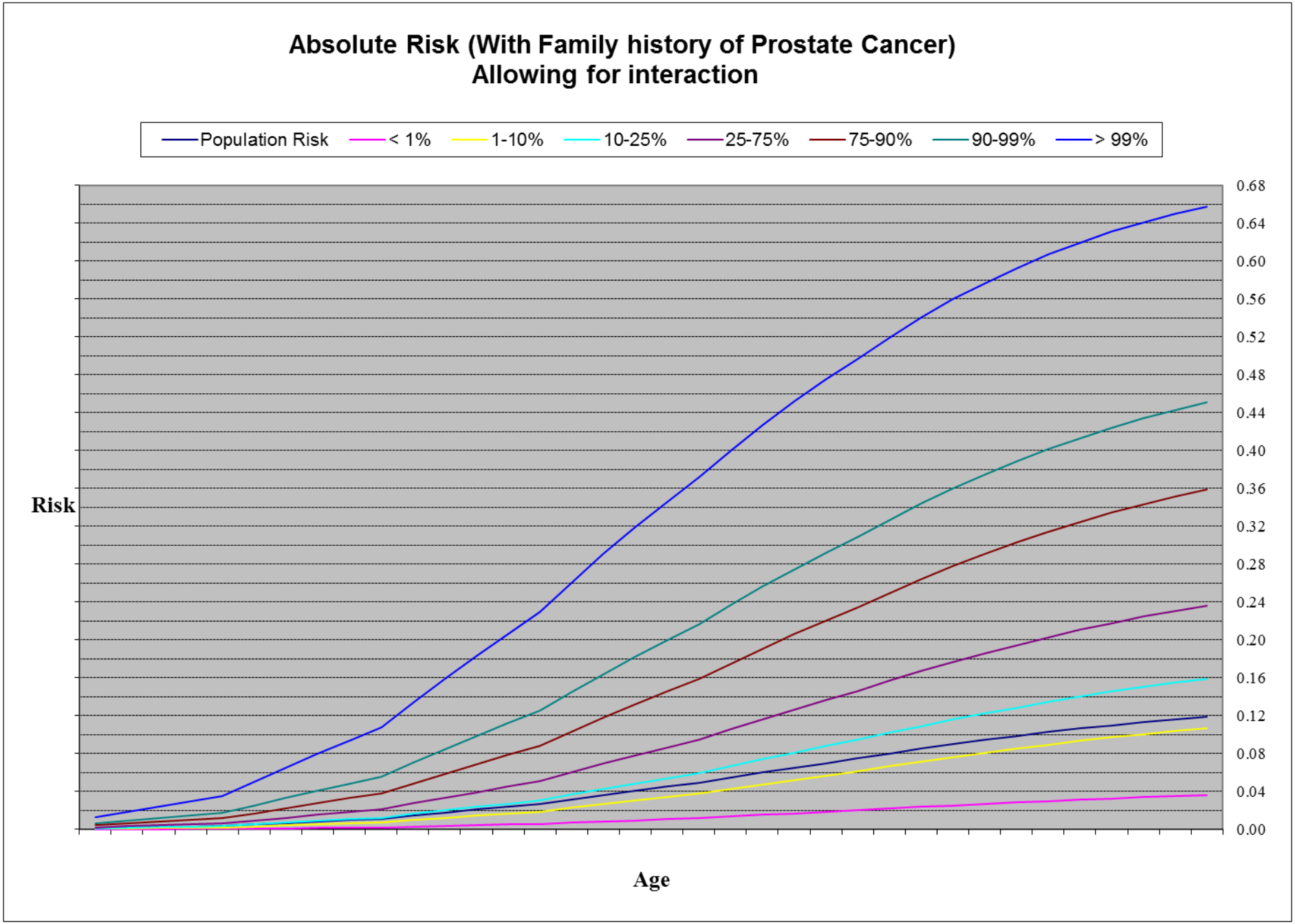
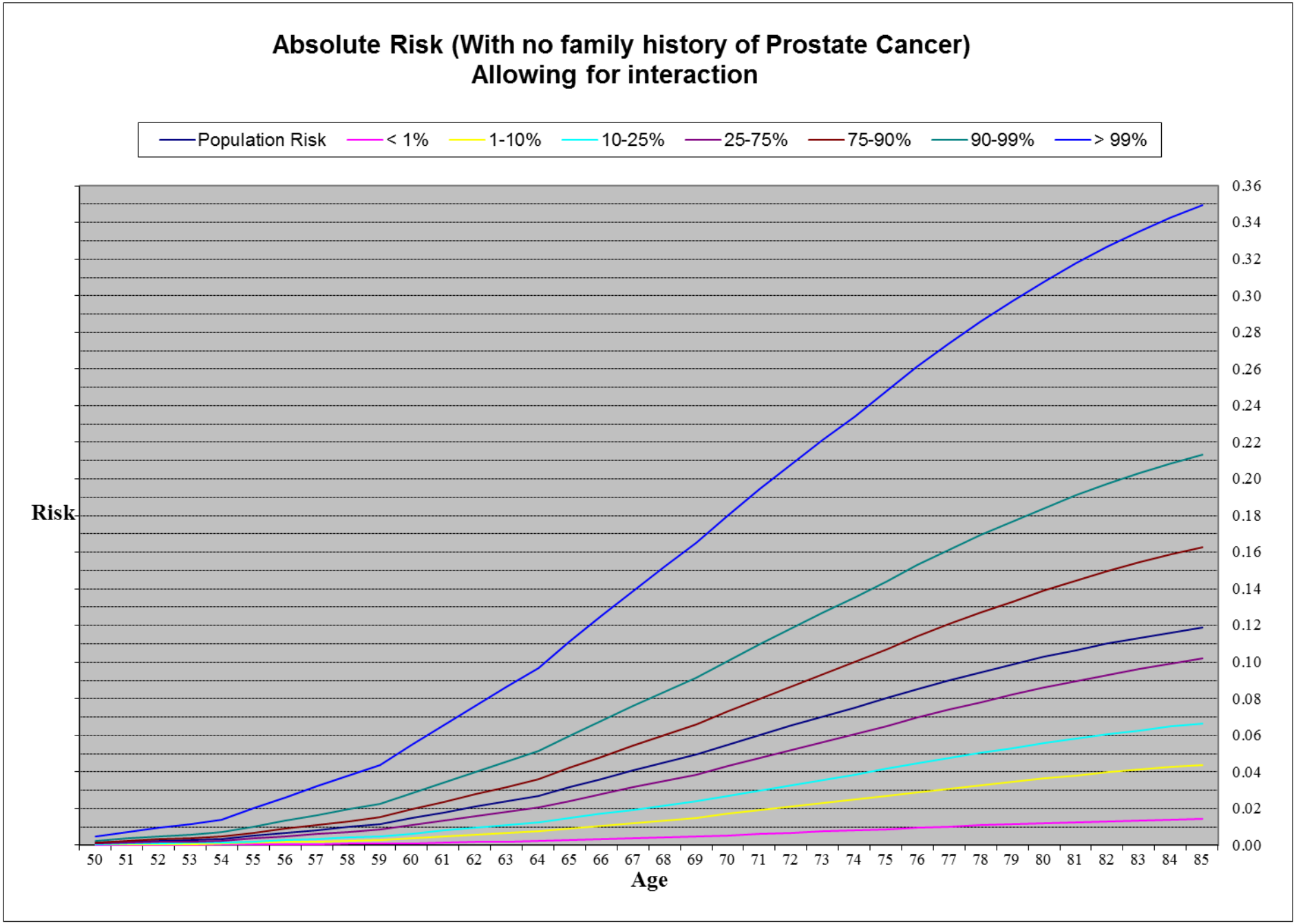

Absolute Risk (With Family history of Prostate Cancer)
Allowing for interaction
Population Risk
< 1%
1-10%
10-25%
25-75%
75-90%
90-99%
> 99%
Risk
0.68
0.64
0.60
0.56
0.52
0.48
0.44
0.40
0.36
0.32
0.28
0.24
0.20
0.16
0.12
0.08
0.04
0.00
Age

Figure 2


Absolute Risk (With no family history of Prostate Cancer)
Allowing for interaction
Population Risk
< 1%
1-10%
10-25%
25-75%
75-90%
90-99%
> 99%
Risk
0.36
0.34
0.32
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85
Age

Supplementary table 1: Total number of cases and controls in PRACTICAL III and GWAS stage 3 by population and study

Study	Ethnicity	No.	Study Acronym	Study Design	Controls[b]	Cases[b]	Total[b]	Age[c]
PRACTICAL PIII	European	1	Aarhus	Case-Control	592	650	1,242	64 (36-88)
		2	BiPAS	Case-Control	88	176	264	69 (51-85)
		3, 4	CPCS 1 & 2	Case-Control	1,474	413	1,887	69 (50-94)
		5	ESTHER	Case-Control	329	330	659	66 (51-75)
		6	FHCRC	Case-Control	1,265	1,309	2,574	60 (35-74)
		7	HaPCS	Case-Control	485	499	984	66 (41-81)
		8	MAYO [a]	Case-Control	527	884	1,411	65 (41-86)
		9	MCCS	Case-Control[d]	2,332	491	2,823	69 (50-85)
		10	MEC	Cohort Study[e]	396	448	844	59 (46-71)
		11	MOFFITT	Case-Control	372	702	1,074	69 (50-87)
		12	PCFS[a]	Case-Control[f]	82	76	158	65 (42-87)
		13	PCMUS	Case-Control	193	173	366	70 (47-90)
		14	Poland	Case-Control	572	574	1,146	68 (43-90)
		15	ProtecT	Case-Control[g]	2,194	2,204	4,398	63 (45-72)
		16	QLD	3 studies [h]	1,394	1,346	2,740	63 (40-89)
		17	SFPCS (NC-CCPC)	Case-Control	219	302	521	65 (44-80)
		18	TAMPERE	Case-Control	1,851	2,496	4,347	67 (37-95)
		19	TASPRAC[a]	Case-Control	337	492	829	62 (44-82)
		20	UKGPCS	Case-Control	212	392	604	60 (36-85)
		21	ULM [a]	Case-Control	379	702	1,081	64 (42-84)
		22	USC	Case-Control	295	488	783	68 (46-94)
		23	UTAH [a]	Case-Control	240	466	706	67 (42-87)
		24	Valais	Case-Control	222	170	392	65 (49-84)
Total					**16,050**	**15,783**	**31,833**	**64 (35-95)**
GWAS stage 3	European		MCCS	Case-Control[d]	548	327	875	69 (49-85)
			PCFS[b]	Case-Control[f]	218	1,025	1,243	57 (41-87)
		25	SEARCH	Case-Control	1,365	1,450	2,815	63 (44-73)
			UKGPCS	Case-Control	1,945	1,703	3,648	61 (36-85)
Total					**4,076**	**4,505**	**8,581**	**61 (36-87)**
Total European (PRACTICAL & GWAS stage 3)					**20,126**	**20,288**	**40,414**	**64 (35-95)**
Total European (PRACTICAL & GWAS stage 3) excluding familial studies					**18,343**	**16,643**	**34,986**	**64 (35-95)**

[a] Studies that oversampled cases with family history. [b] The number of samples given here are those that were available for this analysis. These may differ from the numbers for the complete studies given in the Supplementary Notes. [c] Mean age at diagnosis for cases (min-max). [d] Nested case-control. [e] Population-based prospective cohort study. [f] Population-based with cases from informative families. [g] PSA screening selected cases-controls. [h] One Cohort study, one hospital based and one prospective cohort study.

Supplementary Table 2: Data information for family history and age at diagnosis/observation

(a): All samples excluding studies with familial design			
Samples	Cases	Controls	Total
Without family history	9,740	8,305	18,045
With family history	2,165	999	3,164
Total having information	11,905	9,304	**21,209**
Missing family history information	4,738	9,039	13,777
Total	16,643	18,343	34,986
Family history information	71.53%	50.72%	**60.62%**
Percentage of samples with family history among those having family history information	**18.19%**	**10.74%**	**14.92%**
Age at diagnosis/recruitment	16,503	14,647	**31,150**
Mean age (range)	**64 (35-95)**	**58.3 (18-99)**	61.5 (18-99)
Missing age information	140	3,696	3,836
Total	16,643	18,343	34,986
Percentage of samples with age information	99.16%	79.85%	89.04%
(b): All samples			
Samples	Cases	Controls	Total
Without family history	11,940	9,451	21391
With family history	3,486	1,397	4,883
Total having information	15,426	10,848	**26,274**
Missing family history information	4,862	9,278	14,140
Total	20,288	20,126	40,414
Family history information	76.04%	53.90%	**65.01%**
Percentage of samples with family history among those having family history information	**22.60%**	**12.88%**	18.58%
Age at diagnosis/recruitment	20,137	16,262	**36,399**
Mean age (range)	64 (35-95)	58.7 (18-99)	61.6 (18-99)
Missing age information	151	3,864	4,015
Total	20,288	20,126	40,414
Percentage of samples with age information	99.26%	80.80%	**90.07%**

Supplementary Table 3: Grade-specific and family history-specific odds ratios.

Variable/ Marker	GS < 8	GS 8 +	P^a	P^b	Family History (FH)		P^c
					No	Yes	
PRS[d]	1.75 (1.68-1.82)	1.65 (1.54-1.77)	.37	NA	1.70 (1.64-1.76)	1.79 (1.63-1.96)	$1.8x10^{-4}$
rs721048	1.13 (1.08-1.18)	1.12 (1.02-1.23)	.81	.96	1.09 (1.06-1.16)	1.11 (1.04-1.19)	.06
rs1465618	1.08 (.1.04-1.13)	1.00 (.92-1.09)	.09	.22	1.07 (1.03-1.12)	1.13 (1.05-1.21)	.02
rs12621278	.76 (.70-.82)	.75 (.64-.88)	.79	1.00	.72 (.66-.78)	.77 (.67-.88)	.21
rs2660753	1.12 (1.05-1.19)	1.05 (.94-1.18)	.29	.82	1.15 (1.08-1.22)	1.07 (.98-1.18)	.87
rs17021918	.88 (.85-.91)	.94 (.88-1.01)	.03	.007	.87 (.84-.90)	.92 (.86-.98)	.21
rs12500426	1.11 (1.07-1.15)	1.02 (.95-1.10)	.05	.29	1.10 (1.06-1.14)	1.10 (1.03-1.18)	.43
rs7679673	.86 (.83-.89)	.90 (.84-.97)	.16	.15	.90 (.87-.93)	.84 (.79-.89)	$2.9x10^{-5}$
rs9364554	1.11 (1.07-1.16)	1.03 (.95-1.12)	.16	.60	1.08 (1.03-1.12)	1.18 (1.11-1.26)	.15
rs10486567	.83 (.79-.88)	.81 (.73-.90)	.79	.90	.87 (.83-.92)	.80 (.73-.88)	.01
rs6465657	1.09 (1.05-1.13)	1.06 (.99-1.14)	.64	.21	1.08 (1.04-1.12)	1.14 (1.07-1.21)	.07
rs1447295	1.41 (1.34-1.49)	1.48 (1.33-1.64)	.64	.001	1.41 (1.33-1.49)	1.54 (1.42-1.68)	$1.4x10^{-6}$
rs6983267	.82 (.79-.85)	.84 (.79-.91)	.68	.85	.83(.80-.86)	.79 (.75-.84)	.003
rs16901979	1.48 (1.35-1.62)	1.77 (1.50-2.08)	.06	.13	1.58 (1.44-1.73)	1.63 (1.42-1.87)	.05
rs2928679	1.05 (1.01-1.09)	1.00 (.93-1.08)	.51	.33	1.05 (1.01-1.09)	1.08 (1.02-1.15)	.06
rs1512268	1.16 (1.12-1.20)	1.04 (.97-1.12)	.002	.02	1.13 (1.09-1.17)	1.19 (1.12-1.26)	.16
rs4962416	1.06 (1.02-1.11)	1.02 (.94-1.11)	.35	.80	1.04 (.99-1.08)	1.07 (1.00-1.14)	.13
rs10993994	1.23 (1.19-1.28)	1.21 (1.13-1.30)	.82	.80	1.24 (1.20-1.29)	1.32 (1.25-1.41)	.002
rs7931342	.85 (.82-.88)	.79 (.74-.86)	.19	.50	.85 (.82-.89)	.80 (.75-.85)	$9.6x10^{-5}$
rs7127900	1.28 (1.22-1.34)	1.14 (1.04-1.24)	.01	.003	1.21 (1.16-1.27)	1.27 (1.18-1.37)	.01
rs4430796	.80 (.77-.83)	.79 (.74-.85)	.98	.40	.82 (.79-.85)	.77 (.73-.82)	.004
rs11649743	.87 (.83-.91)	.87 (.79-.95)	.94	.30	.89 (.85-.94)	.83 (.77-.90)	.003
rs1859962	1.19 (1.15-1.24)	1.13 (1.05-1.21)	.20	.78	1.17 (1.13-1.21)	1.21 (1.14-1.28)	.007
rs2735839	.79 (.75-.83)	.93 (.83-1.03)	.002	.003	.81 (.76-.85)	.80 (.73-.88)	.31
rs5759167	.85 (.82-.88)	.85 (.80-.91)	.85	.58	.85 (.82-.88)	.82 (.77-.86)	.04
rs5945619	1.13 (1.10-1.17)	1.16 (1.09-1.23)	.54	.44	1.14 (1.10-1.18)	1.10 (1.04-1.16)	.81

[a] *P*-value for difference in per-allele OR between GS 8+ and GS <8

[b] *P*-value for trend in OR by Gleason score as an ordinal variable

[c] *P*-value for difference in per-allele OR between Family History Yes and No

[d] Polygenic Risk Score adjusted for age at diagnosis and family history. For Family history, PRS adjusted for age at diagnosis (OR is based on per unit increase of PRS).

Supplementary Table 4: PSA levels by genotype in controls.

Marker	Geometric mean PSA (95%CI)			*P*-value[a]
rs721048	CC	CT	TT	
	2.07 (1.63-2.50)	1.78 (1.39-2.16)	2.04 (.70-3.38)	.11
rs1465618	GG	GA	AA	
	1.42 (1.37-1.47)	1.40 (1.32-1.47)	1.30 (1.16-1.45)	.29
rs12621278	AA	AG	GG	
	1.94 (1.61-2.26)	2.03 (1.08-2.97)	1.68 (1.14-2.23)	.81
rs2660753	GG	GA	AA	
	1.90 (1.58-2.22)	2.06 (1.37-2.75)	7.01 (-4.66-18.67)	.21
rs17021918	GG	GA	AA	
	1.94 (1.48-2.39)	1.87 (1.42-2.31)	2.26 (1.21-3.30)	.34
rs12500426	GG	GT	TT	
	2.07 (1.43-2.71)	2.24 (1.65-2.83)	1.98 (1.07-2.88)	.19
rs7679673	CC	CA	AA	
	1.75 (1.34-2.16)	2.25 (1.69-2.81)	1.48 (1.34-1.63)	.38
rs9364554	CC	CT	TT	
	2.40 (1.79-3.00)	1.63 (1.36-1.90)	1.38 (1.27-1.49)	.25
rs10486567	GG	GA	AA	
	2.80 (1.90-3.70)	2.26 (1.36-3.16)	1.44 (1.17-1.71)	.12
rs6465657	AA	AG	GG	
	1.78 (1.35-2.22)	2.05 (1.57-2.53)	2.07 (1.38-2.76)	.83
rs1447295	GG	GT	TT	
	2.08 (1.70-2.46)	1.54 (1.25-1.84)	1.11 (.93-1.29)	.03
rs6983267	CC	CA	AA	
	2.04 (1.32-2.76)	2.12 (1.65-2.59)	1.61 (1.22-1.99)	.01
rs16901979	GG	GT	TT	
	1.94 (1.62-2.26)	1.56 (1.38-1.75)	1.30 (.69-1.91)	.18
rs2928679	CC	CT	TT	
	1.61 (1.31-1.92)	2.17 (1.63-2.70)	1.91 (1.26-2.57)	.27
rs1512268	GG	GA	AA	
	1.68 (1.30-2.06)	2.13 (1.59-2.66)	1.99 (1.33-2.64)	.004
rs4962416	AA	AG	GG	
	2.06 (1.59-2.53)	1.97 (1.50-2.45)	1.35 (1.21-1.49)	.73
rs10993994	GG	GA	AA	
	2.15 (1.54-2.77)	1.68 (1.38-1.98)	2.45 (1.50-3.40)	3.5×10^{-5}
rs7931342	CC	CA	AA	
	2.36 (1.62-3.09)	2.06 (1.57-2.56)	1.41 (1.33-1.49)	.11
rs7127900	GG	GA	AA	
	1.90 (1.54-2.25)	1.99 (1.35-2.64)	2.90 (.33-5.47)	.03
rs4430796	AA	AG	GG	
	1.54 (1.34-1.74)	2.08 (1.61-2.54)	2.22 (1.41-3.03)	.38
rs11649743	GG	GA	AA	
	2.13 (1.69-2.57)	1.57 (1.27-1.86)	2.59 (.44-4.75)	.84
rs1859962	TT	TG	GG	
	2.05 (1.49-2.60)	1.99 (1.50-2.48)	1.84 (1.34-2.33)	.17
rs2735839	GG	GA	AA	
	2.07 (1.70-2.43)	1.79 (1.15-2.42)	1.08 (.83-1.32)	2.8×10^{-13}
rs5759167	GG	GT	TT	
	1.86 (1.28-2.43)	2.16 (1.63-2.69)	1.59 (1.26-1.93)	.37
rs5945619	TT	TC	CC	
	2.63 (1.80-3.46)		2.24 (1.53-2.94)	.09

[a]Test for trend in log (PSA) by allele dose

Supplementary Table 5: Age-specific odds ratios.

Variable/ Marker	Age at diagnosis (years)					*P*trend[a]
	<55	55-59	60-64	65-69	70+	
PRS[b]	1.76 (1.62-1.92)	1.85 (1.72-1.98)	1.70 (1.59-1.82)	1.70 (1.57-1.83)	1.48 (1.37-1.60)	$2.6x10^{-4}$
rs721048	1.09 (.99-1.19)	1.13 (1.06-1.21)	1.12 (1.05-1.19)	1.13 (1.06-1.20)	1.07 (.99-1.15)	.86
rs1465618	1.14 (1.04-1.24)	1.07 (1.01-1.13)	1.09 (1.03-1.15)	1.09 (1.03-1.16)	1.01 (.94-1.07)	.03
rs12621278	.59 (.49-.71)	.76 (.67-.86)	.67 (.60-.76)	.79 (.71-.88)	.83 (.73-.94)	$1.9x10^{-5}$
rs2660753	1.16 (1.04-1.29)	1.10 (1.01-1.20)	1.18 (1.09-1.28)	1.13 (1.05-1.22)	1.04 (.95-1.14)	.12
rs17021918	.83 (.77-.90)	.86 (.81-.91)	.91 (.87-.96)	.89 (.85-.94)	.89 (.84-.94)	.06
rs12500426	1.17 (1.08-1.27)	1.16 (1.09-1.24)	1.09 (1.03-1.15)	1.05 (1.00-1.11)	1.07 (1.02-1.13)	.08
rs7679673	.87 (.81-.94)	.87 (.82-.92)	.90 (.85-.94)	.86 (.81-.90)	.89 (.85-.94)	.08
rs9364554	1.15 (1.06-1.24)	1.13 (1.07-1.20)	1.08 (1.02-1.14)	1.08 (1.03-1.14)	1.07 (1.01-1.14)	.07
rs10486567	.78 (.69-.87)	.82 (.76-.89)	.90 (.84-.97)	.87 (.81-.93)	.85 (.78-.92)	.19
rs6465657	1.14 (1.07-1.23)	1.11 (1.05-1.17)	1.11 (1.05-1.16)	1.06 (1.01-1.11)	1.09 (1.03-1.15)	.69
rs1447295	1.42 (1.28-1.57)	1.43 (1.33-1.55)	1.39 (1.29-1.49)	1.40 (1.30-1.51)	1.45 (1.33-1.58)	.75
rs6983267	.80 (.74-.85)	.80 (.75-.84)	.80 (.76-.84)	.82 (.78-.87)	.90 (.85-.95)	.06
rs16901979	1.54 (1.31-1.81)	1.53 (1.35-1.73)	1.53 (1.36-1.73)	1.56 (1.39-1.75)	1.60 (1.39-1.84)	.06
rs2928679	1.08 (1.01-1.17)	1.03 (.97-1.09)	1.02 (.96-1.07)	1.04 (.99-1.10)	1.04 (.98-1.11)	.09
rs1512268	1.11 (1.04-1.20)	1.19 (1.13-1.26)	1.15 (1.09-1.21)	1.13 (1.08-1.19)	1.10 (1.05-1.16)	.70
rs4962416	1.12 (1.03-1.21)	1.01 (.95-1.07)	1.08 (1.02-1.14)	1.01 (.96-1.07)	1.01 (.95-1.09)	.37
rs10993994	1.35 (1.26-1.45)	1.28 (1.22-1.36)	1.21 (1.15-1.28)	1.23 (1.17-1.29)	1.16 (1.10-1.23)	.009
rs7931342	.83 (.77-.89)	.84 (.79-.88)	.82 (.77-.86)	.87 (.82-.91)	.86 (.81-.92)	.22
rs7127900	1.39 (1.27-1.51)	1.24 (1.16-1.32)	1.22 (1.15-1.30)	1.24 (1.17-1.32)	1.15 (1.08-1.23)	.01
rs4430796	.80 (.75-.86)	.77 (.73-.82)	.81 (.77-.85)	.85 (.81-.89)	.83 (.79-.89)	.12
rs11649743	.91 (.83-.99)	.87 (.81-.93)	.85 (.79-.90)	.91 (.85-.97)	.90 (.83-.97)	.33
rs1859962	1.25 (1.17-1.34)	1.20 (1.14-1.27)	1.24 (1.18-1.30)	1.13 (1.07-1.18)	1.11 (1.05-1.18)	.04
rs2735839	.77 (.70-.86)	.72 (.67-.79)	.77 (.72-.83)	.85 (.79-.91)	.91 (.84-.99)	.001
rs5759167	.80 (.75-.86)	.84 (.79-.88)	.84 (.80-.88)	.84 (.80-.88)	.87 (.82-.91)	.11
rs5945619	1.14 (1.07-1.21)	1.19 (1.14-1.25)	1.16 (1.12-1.21)	1.11 (1.07-1.15)	1.07 (1.02-1.12)	.0003

[a] 1df trend test for trend in OR by age, using case only analysis.
[b] OR is based on per unit increase of PRS.

Supplementary Table 6: The results of 29 pair wise interaction of 25 SNPs significant at P <0.05 (29 out of 300 interactions; Bonferroni correction $0.05/300=1.67x10^{-4}$).

Pair	*P*-value[a]	Pair	*P*-value[a]
rs721048 x rs17021918	0.05	rs9364554 x rs1859962	0.003
rs1465618 x rs17021918	0.004	rs10486567 x rs16901979	0.01
rs1465618 x rs12500426	0.001	rs10486567 x rs4962416	0.02
rs1465618 x rs7679673	0.03	rs10486567 x rs4430796	0.004
rs1465618 x rs7127900	0.02	rs1447295 x rs7127900	0.04
rs2660753 x rs12500426	0.02	rs16901979 x rs10993994	0.001
rs2660753 x rs16901979	0.04	rs4962416 x rs7931342	0.03
rs2660753 x rs2928679	0.05	rs4962416 x rs4430796	0.01
rs2660753 x rs5759167	0.04	rs4962416 x rs2735839	0.001
rs17021918 x rs7679673	0.01	rs10993994 x rs4430796	0.02
rs12500426 x rs7679673	0.02	rs7931342 x rs4430796	0.02
rs12500426 x rs11649743	0.04	rs7931342 x rs2735839	0.05
rs12500426 x rs5759167	0.04	rs7127900 x rs4430796	0.02
rs7679673 x rs5759167	0.01	rs4430796 x rs2735839	0.004
rs9364554 x rs16901979	0.001		

[a] Likelihood ratio test

Supplementary Table 7: Estimated odds ratios of PRS percentiles adjusted for age at diagnosis of PrCa (five categories) in a model allowing for an interaction between PRS and Age.

Covariate		Odds ratio (95% CI)	*P*-value	Odds ratio[a] (95% CI)	*P*-value
PRS Group	<1%	1		0.10 (0.05-0.18)	$2.4x10^{-13}$
	1-10%	3.43 (1.89-6.23)	$5.3x10^{-5}$	0.33 (0.27-0.41)	$3.7x10^{-26}$
	10-25%	5.90 (3.22-10.82)	$9.5x10^{-9}$	0.57 (0.50-0.65)	$2.5x10^{-18}$
	25-75%	10.33 (5.53-19.29)	$2.5x10^{-13}$	1	
	75-90%	19.12 (9.92-36.86)	$1.3x10^{-18}$	1.85 (1.66-2.07)	$6.6x10^{-28}$
	90-99%	28.92 (14.50-57.69)	$1.3x10^{-21}$	2.80 (2.38-3.30)	$1.5x10^{-34}$
	>=99%	58.11 (26.99-125.11)	$3.1x10^{-25}$	5.63 (4.05-7.82)	$6.4x10^{-25}$
Family History (Yes/No)		2.52 (2.29-2.78)	$4.7x10^{-77}$	2.52 (2.29-2.78)	$4.7x10^{-77}$
PRS x Age at Diagnosis interaction[b]		0.97 (0.95-0.99)	0.003	0.97 (0.95-0.99)	0.003

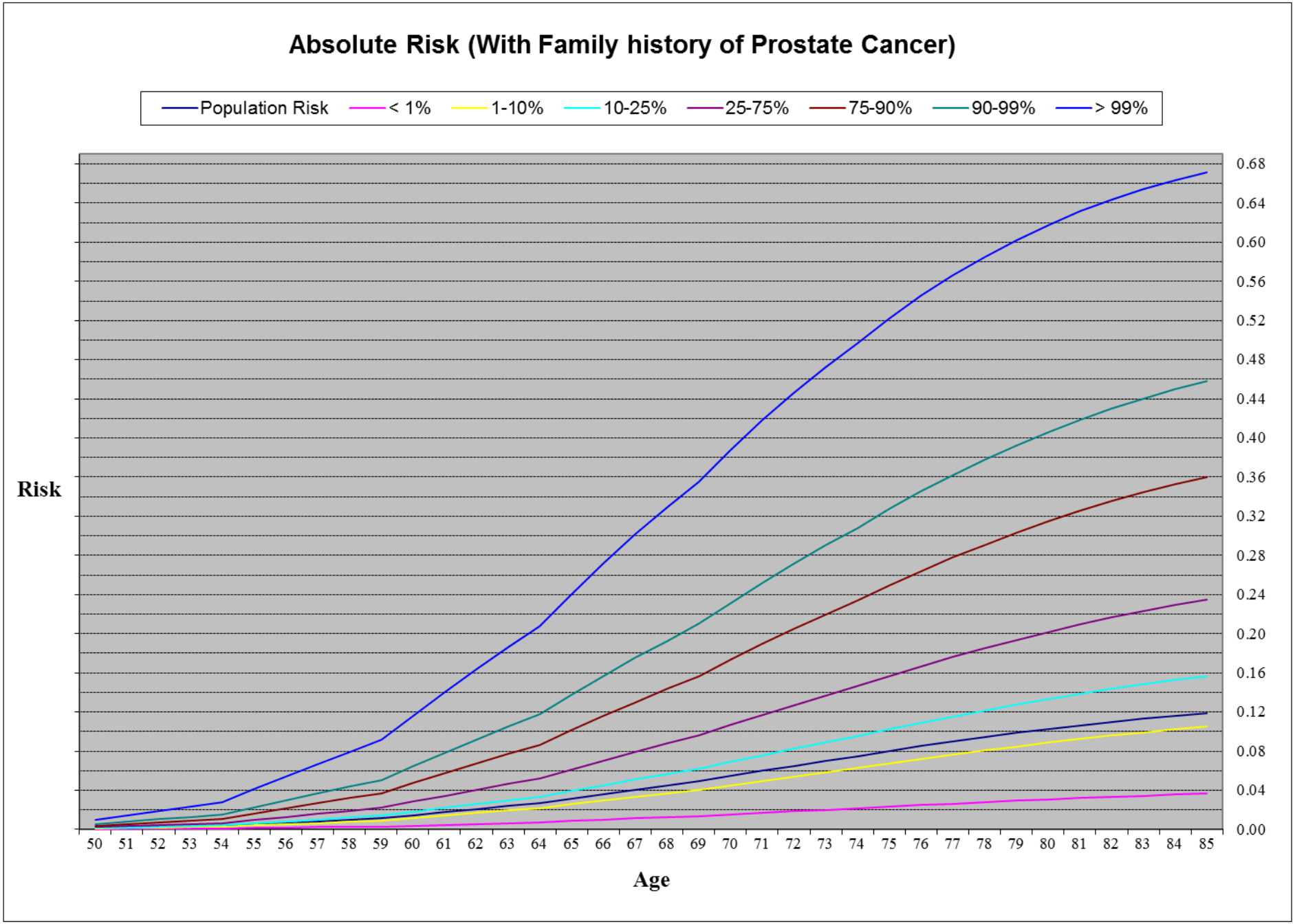
[a] Odds Ratios were compared to median risk group.
[b] PRS and age at diagnosis both as categorical variables.

Supplementary Table 8: Estimated odds ratios used for estimating absolute risk

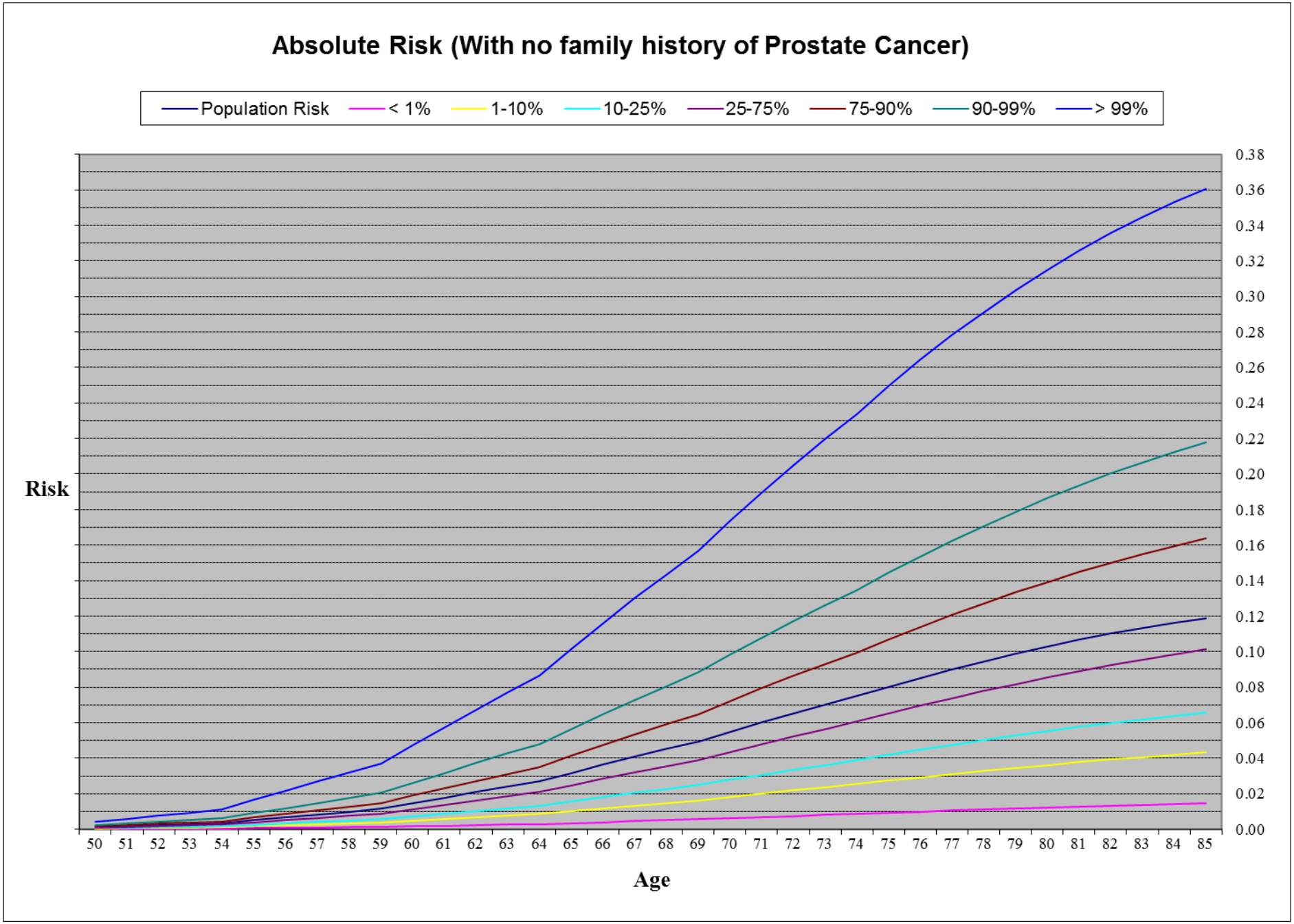
(a) Model without interaction							
Family History (FH)	Risk Group						
	<1%	1-10%	10-25%	25-75%	75-90%	90-99%	>=99%
No	0.14	0.41	0.63	1.00	1.68	2.31	4.24
Yes	0.35	1.04	1.60	2.52	4.24	5.83	10.70
(b) Model with interaction[a] (samples with no FH)							
Age group	<1%	1-10%	10-25%	25-75%	75-90%	90-99%	>=99%
<55	0.10	0.33	0.57	1.00	1.85	2.80	5.63
55-59	0.09	0.31	0.55	1.00	1.79	2.62	5.09
60-64	0.08	0.29	0.53	1.00	1.73	2.45	4.60
64-69	0.07	0.27	0.52	1.00	1.67	2.29	4.16
70+	0.06	0.25	0.50	1.00	1.62	2.14	3.76
(c) Model with interaction[a] (samples with FH)							
Age group	<1%	1-10%	10-25%	25-75%	75-90%	90-99%	>=99%
<55	0.24	0.84	1.44	2.52	4.67	7.07	14.21
55-59	0.22	0.78	1.39	2.52	4.52	6.61	12.84
60-64	0.20	0.73	1.35	2.52	4.37	6.18	11.60
64-69	0.18	0.68	1.30	2.52	4.22	5.78	10.49
70+	0.16	0.64	1.26	2.52	4.08	5.40	9.48

[a] Interaction between PRS and age at diagnosis both as categorical variables.

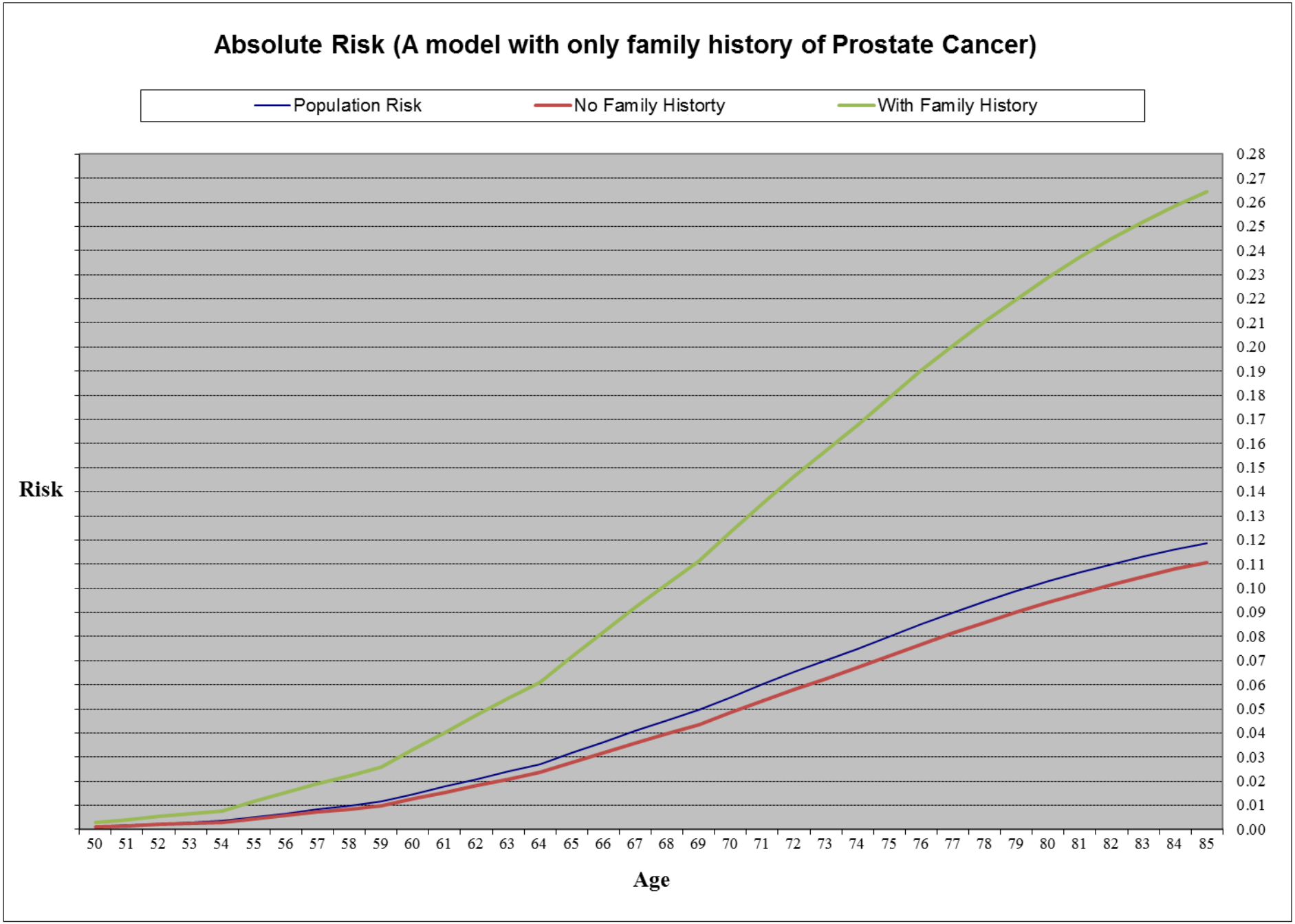
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



SUPPLEMENTARY NOTES: DESCRIPTION OF OPA AND THE PRACTICAL CONSORTIUM GROUPS

All studies were approved by the appropriate ethics committees. A list of the groups is in Supplementary Table 1.

CPCS1+2:

The CPCS (Copenhagen Prostate Cancer Study) 1 + 2, Copenhagen, Denmark

The Copenhagen Prostate Cancer Study 1 included 872 unselected patients recruited with prostate cancer between 2008-2011 from Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark. The Copenhagen Prostate Cancer Study 2 included 306 unselected patients with prostate cancer recruited in 2010-2011 from Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. PSA is not routinely screened for in Denmark, and cases are therefore mainly clinically detected. Controls were 2,777 (CPCS1) and 798 (CPCS2) prostate cancer free men from the general population, whom participated in the Copenhagen City Heart Study. Diagnosis of prostate cancer was confirmed by fully trained pathologists. All participants were white and of Danish descent. Participants filled out questionnaires, gave blood samples for DNA extraction and gave written informed consent.

ESTHER

In the ESTHER study, patients with a first diagnosis of prostate cancer at age 50-75 years were recruited in hospitals and medical practices in Saarland, a state located in southwest Germany, from 2001 to 2003. Controls were selected from participants of a general health-check up within the same age range (and frequency matched to the cases by 5-year age groups) who were recruited in general practices in Saarland in 2000-2002. Cases and controls who were almost exclusively of European descent, filled out a detailed standardized questionnaire on life time history of potential risk factors and had a blood sample taken, and medical data were extracted from medical records.

FHCRC: Fred Hutchinson Cancer Research Centre, Seattle US

The study population consists of participants from two population-based case-control studies in Caucasian and African American residents of King County, Washington

(Study I and Study II), which have been previously described[7] Incident cases with histologically confirmed PrCa were ascertained from the Seattle-Puget Sound Surveillance, Epidemiology and End Results cancer registry. In Study I, cases were diagnosed between January 1, 1993, and December 31, 1996 and were 40-64 years of age at diagnosis. In Study II, cases were diagnosed between January 1, 2002, and December 31, 2005 and were 35-74 years of age at diagnosis. Overall, 2,244 eligible PrCa patients were identified and 1,754 (78%) were interviewed. Blood samples yielding sufficient DNA for genotyping were drawn from 1,457 (83%) cases who completed the study interview. A comparison group of controls without a history of PrCa, residing in King County, Washington, was identified for each study using random digit telephone dialling. Controls were frequency-matched to cases by five-year age groups and recruited evenly throughout each ascertainment period for cases. A total of 2,448 men were identified who met the eligibility criteria and 1,645 (67%) completed a study interview. Blood samples were drawn and DNA prepared from 1,352 (82%) interviewed controls.

MAYO, Rochester, Minnesota, US

The Mayo Clinic study consisted of clinic-based cases, including 476 affected men from 185 families with PrCa, 445 men with sporadic PrCa, 199 with aggressive (Gleason score > 7) PrCa, and 500 population-based controls. The controls (all males) were randomly selected from a sampling frame of Olmsted County, Minnesota, provided by the Rochester Epidemiology Project. The methods used to ascertain familial and sporadic PrCa patients, as well as controls, have been described previously[8]. All individuals from the Mayo Clinic study included in this report were of self-reported European descent.

Cancer Council Victoria Prostate Cancer Program, Melbourne

The Cancer Council Victoria's Prostate Cancer Program includes three studies: the Melbourne Collaborative Cohort Study (MCCS) and the Prostate Cancer Family Study (PCFS). Cases and controls (and informative families) from these studies have been used for several stages of this research effort, beginning with the UK and Melbourne stage 2 GWAS.

The MCCS is a prospective cohort study that includes 17,154 men who were aged 40 and 69 years when recruited between 1990 and 1994. MCCS participants are

regularly linked to the Victorian Cancer Registry and the Australian Cancer Database to ascertain incident cases (1582 by end of 2008) including men diagnosed in other states of Australia. A random sample of MCCS participants who were not diagnosed with prostate cancer during follow-up provides a control group.
The PCFS is a population-based family series of 1428 men diagnosed with prostate cancer before the age of 56 years and 256 men diagnosed after the age of 55 years who were recruited in Victoria between 1998 and 2010. Cases were ascertained using the population-based Victorian Cancer Registry, and family members were approached after gaining the consent of each case. Altogether, 77% of cases agreed to participate.

MEC: Multiethnic Cohort

The Multiethnic Cohort Study is a population-based prospective cohort study that was initiated between 1993 and 1996 and includes subjects from various ethnic groups -African-Americans and Latinos primarily from California (mainly Los Angeles) and Native Hawaiians, Japanese-Americans, and European Americans primarily from Hawaii. State drivers' license files were the primary sources used to identify study subjects in Hawaii and California. Additionally, in Hawaii, state voter's registration files were used, and, in California, Health Care Financing Administration (HCFA) files were used to identify additional African American men. All participants (n=215,251) returned a 26-page self-administered baseline questionnaire that obtained general demographic, medical and risk factor information. In the cohort, incident cancer cases are identified annually through cohort linkage to population-based cancer Surveillance, Epidemiology, and End Results (SEER) registries in Hawaii and Los Angeles County as well as to the California State cancer registry. Information on stage and grade of disease are also obtained through the SEER registries. Blood sample collection in the MEC began in 1994 and targeted incident PrCa cases and a random sample of study participants to serve as controls for genetic analyses. This nested PrCa case-control study in the MEC consists of 890 invasive PrCa cases and 895 controls. This study was approved by the Institutional Review Boards at the University of Southern California and at the University of Hawaii and informed consent was obtained from all study participants.

MOFFITT: Moffitt Study, Tampa, Florida, US

This is a hospital-based incident study of 638 patients with primary adenocarcinoma of the prostate. They were recruited from 2002 to 2009 at the H. Lee Moffitt Cancer Centre (Tampa, FL, US) and James A. Haley Veterans Affairs Hospital (Tampa, FL, US). Ninety-five percent of the case subjects who were asked to participate in the study agreed. All cancer cases were histologically confirmed by the Department of Pathology at each institution. The controls consisted of 147 subjects who were visiting the Lifetime Cancer Screening Centre, which is affiliated with the H. Lee Moffitt Cancer Centre or VA hospital. All control subjects were male and had had no previous diagnosis of cancer. The control subjects were frequency matched to the patients by age at diagnosis (± 5 years). Eighty-three percent of the control subjects who were asked to participate in the study consented. Non-genetic risk factor data for the present study were obtained through in-person interviews with the patients and controls at enrolment. The questionnaire covered demographic information, family history of cancer (i.e., whether they have one or more first-degree family member with PrCa), medical history, and detailed tobacco consumption. For the patients, data on cancer stage, Gleason score, and prostate specific antigen level were abstracted from the medical records. The subjects were asked to provide a blood or buccal sample after the interview as a source of genomic DNA.

PCMUS: Bulgaria

The Bulgarian sample of PrCa patients consist mainly of newly diagnosed cases, which are histopathologically confirmed. The patients (N=150, age range 39-93) are of Bulgarian origin. Transrectal biopsy was performed at the Urology Clinic, Alexandrovska University Hospital, mainly because of an elevated PSA. Some of the patients were referred from other centres to the tertiary university hospital after being previously diagnosed with PrCa. A small subset of patients had previously had definitive treatment (mainly radical prostatectomy) and they were called retrospectively with invitation to join the study. The control group is matched to the patients by sex, age, and ethnicity. It consists of two groups: (i) 72 healthy males, age range 54-87, presenting to our institution with lower urinary tract symptoms caused by benign prostatic hypertrophy (BPH) who had a PSA <3.5. The majority of them subsequently underwent surgical treatment with histological verification of the BPH; (ii) an additional healthy control group of 78 anonymous males matched to the PrCa patients by age and ethnicity, but with no PSA data.

POLAND

Polish case-control series included 458 men with PrCa, diagnosed in north-western Poland between 1999 and 2009 at the University Hospital in Szczecin. Study participants were unselected for age and family history. The mean age of PrCa diagnosis was 68 years (range 41–90 years). The control group included 476 cancer-free adult men from the same population (age range, 24–89 years; mean 63.1) taken from the healthy adult patients of five family doctors practicing in the Szczecin region. These individuals were selected randomly from the patient lists of the participating doctors.

ProtecT/ ProMPT, UK

The ProtecT[9] (Prostate testing for cancer and Treatment) trial is an NIHR-funded, UK-wide study of community-based PSA testing followed by a randomised controlled trial of PrCa treatment (radical surgery, radical conformal radiotherapy and active monitoring: ProMPT). Over 200,000 men between the ages of 50 and 69 years, ascertained through general practices in nine regions in the UK, were approached and over 100, 000 attended for PSA testing and, when PSA was 3.0ng/ml or more, for prostate cancer diagnosis. Over 95% of recruited men were of white ethnicity. For this study, after QC, 1563 cases identified by PSA screening within the ProtecT study were analysed. Controls with normal PSA levels (<3ng/ml) were selected from the same GP register and 5 year age band as the cases (n=1474 after QC were analysed).

QUEENSLAND (QLD): Australia

The Queensland cases included in this study were ascertained from three studies: (i) a longitudinal cohort study (Prostate Cancer Supportive Care and Patient Outcomes Project: ProsCan) being conducted through CCQ in Queensland, through which men newly diagnosed with prostate cancer from 26 private practices and 8 public hospitals were directly referred to ProsCan at the time of diagnosis by their treating clinician (N=780, age range 43-88 years) (ii) Caucasian patients accrued to date through the Queensland node of the Australian Prostate Cancer BioResource (APCB) where cases were recruited through local urologists at time of diagnosis (N=445) (iii) Prospective collection through various urologists (N=121). All cases had

histopathologically confirmed prostate cancer, following presentation with an abnormal serum PSA and/or lower urinary tract symptoms.. Controls, recruited through QUT and QIMR, comprised healthy male blood donors with no personal history of prostate cancer, from the (i) the Australian Red Cross Blood Services in Brisbane (N=865, age range 19-76 years) and (ii) the Australian Electoral Commission (N=529, age and post-code/area matched to ProsCan, age range 54-90 years)

SEARCH

Prostate cancer cases were identified via the Eastern Cancer Registration and Information Centre, East Anglia, UK. Incident cases <70 years at diagnosis are recruited. Controls are men attending general practice who are frequency- matched to cases by age and geographic region.

TAMPERE: Finland

Total of 8744 Finnish samples were sent to for typing. Of these, 2960 unselected cases and 165 controls (PSA < 4 µg/ml) were collected in Tampere, Finland and all are of Finnish origin. The mean age of diagnosis was 68.7 years (range 36-94). The patients were diagnosed with PrCa in 1993-2008 in the Tampere University Hospital, Department of Urology. Tampere University Hospital is a regional referral centre in the area for all patients with PrCa, which results in an unselected, population-based collection of patients. The other unselected set of samples were 5522 samples collected in the Finnish arm of The European Randomized Study of Screening for Prostate Cancer, which was initiated in the early 1990s to evaluate the effect of screening with prostate-specific–antigen (PSA) testing on death rates from PrCa. This sample set includes 1106 Finnish cancer cases and 4416 controls. These men were born in years 1933, 1937 and 1941 and were randomly assigned to a group that was offered PSA screening at an average of once every 4 years or to a control group that did not receive such screening. In addition to these two sporadic sample sets, 97 familial cancer cases (mean age at diagnosis 70 years) from Finnish PrCa families were genotyped.

UKGPCS

Blood DNA from PrCa cases was collected from cases throughout the UK aged ≤ 60 years at diagnosis and a systematic series from the PrCa clinic at The Royal Marsden NHS Foundation Trust. Diagnosis is confirmed from medical record or death certificate. 60% are clinically detected.

ULM: Germany

Cases were recruited in two different ways. Familial PrCa probands (index cases) were ascertained from all over Germany. They were advised by their attending physicians to contact the Clinic of Urology of Ulm. The positive family history was then verified by reviewing medical records or death certificates of family members. In each case, only one member of each family (e.g. the proband) was enrolled in the present study. Sporadic cases, who reported no relatives affected with PrCa, were almost exclusively collected at Ulm during their course of treatment (e.g. radical prostatectomy) in our Urology Clinic. The control group consists of 213 age-matched healthy men and 295 population controls of unknown disease status.

UTAH, US

All 455 prostate cancer cases were drawn from the set of sampled prostate cancer cases belonging to extended Utah high-risk pedigrees. All cases were selected to have kinship coefficients <= 0.0156 with any other case included from the high-risk pedigree set. The 256 controls were selected from other high-risk pedigree studies as: 1) not related to prostate kindred, 2) not having cancer, 3) not having a first degree relative with prostate cancer.

SFPCS (NC-CCPC): San Francisco, California USA

The San Francisco Bay Area Prostate Cancer Study is a population-based case-control study of PrCa in non-Hispanic white and African-American men conducted in the San Francisco Bay Area (John et al 2005). Eligible cases with localized or advanced disease and controls completed an in-person interview; a blood or mouthwash sample was collected for advanced cases and controls only. Newly diagnosed cases aged 40-79 years wereidentified through the Greater Bay Area Cancer Registry, including on-Hispanic white cases diagnosed between July 1, 1997 and February 28, 2000, and African-American cases diagnosed between July 1, 1997 and December 31, 2000. Control men aged 40-79 years without a history of

prostate cancer were identified through random-digit dialing; Incontrols aged 65-79 years were also randomly selected from the rosters of beneficiaries of the Health Care Financing Administration (HCFA). Controls were frequency-matched to cases by five-year age group and race. This study included 389 advanced cases and 256 controls with DNA from a blood sample.
Esther John, Amit Joshi, Ahva Shahabi

TASPRAC: Tasmania, Australia

The Tasmanian cases were ascertained from two studies. The fifty-four familial PrCa cases included in this study were drawn from the Tasmanian Familial Prostate Cancer Study, which has recruited families with multiple individuals affected with PrCa. These families were identified using the records of the Tasmanian Cancer Registry (a register of all cancer diagnoses in Tasmania, Australia since 1978) and the genealogical records from the Menzies Research Institute Tasmania Genealogical Database; they comprise families with at least 2 affected first-degree relatives. One case per family was selected for inclusion. Blood samples, pathology specimens and pathology reports are available for these familial cases. The remaining 469 sporadic PrCa cases were drawn from the Tasmanian Prostate Cancer Case Control Study. Cases were again identified from the Tasmanian Cancer Registry, and eligible cases were men <70 years diagnosed with histologically confirmed PrCa diagnosed between 1996 and 2005. The 359 controls included here were randomly selected from the Tasmanian electoral roll (registration on the electoral roll is compulsory in Australia for individuals > 18 years). Eligible controls were sex and age-matched within 5-year age groups to the sporadic cases and self-reported as unaffected with PrCa. Blood samples, physical measures, dietary history, environmental exposure data and family history have been collected from participating individuals. Of note, controls diagnosed with PrCa subsequent to their recruitment have been removed from the control dataset. Notification of their subsequent diagnosis with PrCa was determined by their registration as a confirmed case by the Tasmanian Cancer Registry as of the end of 2009.
Joanne L Dickinson (PI), James R. Marthick

USC: Los Ángeles, Southern California, USA

Subjects were participants in a population-based case-control study of aggressive PrCa conducted in Los Angeles County. Cases were identified through the Los Angeles County Cancer Surveillance Program rapid case ascertainment system large hospitals are screened at least weekly and other sites at least monthly. Eligible cases included African American, Hispanic, and non-Hispanic white men diagnosed with a first primary PrCa between January 1, 1999 and December 31, 2003. Eligible cases also had either (1) prostatectomy with documented tumour extension outside the prostate, (2) metastatic PrCa in sites other than prostate, (3) needle biopsy of the prostate with Gleason grade $\geq$8 or (4) needle biopsy with Gleason grade 7 and tumour in more than 2/3 of the biopsy cores.

Eligible controls were men never diagnosed with PrCa, living in the same neighbourhood as a case, and were frequency matched to cases on age (±5 years) and race/ethnicity. Controls were identified by a neighbourhood walk algorithm which proceeds through an obligatory sequence of adjacent houses or residential units beginning at a specific residence that has a specific geographic relationship to the residence where the case lived at diagnosis.

Sue Ann Ingles, Mariana C. Stern, Roman Corral

BiPAS: The Birmingham Prostatic Neoplasms Association Study (BiPAS) – A Genetic and Environmental Case Control Study

The Birmingham Prostatic Neoplasms Association Study base consists of men living in the south Birmingham area, United Kingdom aged $\geq$50 years. The study recruited men with lower urinary tract symptoms (LUTS) and/or high serum prostate specific antigen (PSA) levels referred for prostate biopsies between March 2007 until October 2008. PrCa cases were recruited from the Queen Elizabeth Medical Centre, Birmingham. Cases are defined as men with histologically confirmed adenocarcinoma of the prostate. Controls were also recruited from the Queen Elizabeth Medical Centre and Selly Oak Hospital, Birmingham. Men with a normal repeat PSA and a negative biopsy were categorized as benign controls.

A blood sample from every hospital based subject was obtained using standard venepuncture methods, and collected in a 5ml tube containing EDTA. Samples were transported to the laboratory immediately in a cool bag with cool packs and stored at 4°C. DNA was extracted using the QIAGEN maxi blood kit.

Mr David M A Wallace - Queen Elizabeth Medical Centre, Edgbaston, Birmingham

Mr Alan Doherty - Queen Elizabeth Medical Centre, Edgbaston, Birmingham
Mr R I Bhatt - Queen Elizabeth Medical Centre, Edgbaston, Birmingham
Mr K Subramonian - Queen Elizabeth Medical Centre, Edgbaston, Birmingham
Dr John Arrand – University of Birmingham
Louise Flanagan – University of Birmingham
Sita Ann Bradley - Queen Elizabeth Medical Centre, Edgbaston, Birmingham

AARHUS, Denmark

The Aarhus Prostate Cancer Study included 661 cases and 601 controls. Cases were patients treated for prostate adenocarcinoma at Department of Urology, Aarhus University Hospital, Skejby (Aarhus, Denmark) from 1999-2008. Median age at diagnosis was 63 years. The vast majority of these patients (91%) were treated by radical prostatectomy for clinically localized PrCa, while the remaining patients (9%) underwent palliative treatment for disseminated disease (transurethral resection of the prostate and/or endocrine treatment). Controls were age-matched males treated for myocardial infarction or undergoing coronary angioplasty, but with no prostate cancer diagnosis based on information retrieved from the Danish Cancer Register and the Danish Cause of Death Register. For SNP genotyping analyses, genomic DNA was extracted from frozen blood samples using the automated Maxwell^{TM} 16 Instrument (Promega), according to the protocol provided by the manufacturer.

VALAIS: Switzerland in collaboration with Montreal, Canada

Between December 1, 2002 and January 31, 2007, all urologists in a relatively isolated alpine region of Switzerland (canton du Valais) invited their patients diagnosed with invasive PrCa (all stages) to participate to a research project on genetic factors involved in PrCa. Both parents were required to originate from the canton du Valais. A detailed family history on at least 3 generations was collected by a trained research nurse, as well as a blood sampling. A series of healthy men, without a self-reported family history of PrCa, originating from the same region, participated as controls (blood donors and elderly patients seen in private practice). Most of these men had regular PSA screening. This study has contributed DNA samples to molecular work in Montreal. For this report the genotyping was performed in the UK on DNA obtained from the Montreal laboratory which was sourced from the Valais sample set.

Hannover Prostate Cancer Study, Hannover, Germany (HaPCS)

A hospital-based series of 499 unselected Caucasian patients with prostate cancer, who were treated with brachytherapy between October 2000 and September 2007 at Hannover Medical School, were enrolled for this study. All patients had biopsy-proven adenocarcinoma of the prostate. Indication for permanent brachytherapy was clinically localized low-risk early prostate cancer (cT2a or less with a PSA serum level <10 ng/mL and a Gleason score <7) following the European Society for Therapeutic Radiology and Oncology/European Association of Urology/European Organization for Research and Treatment of Cancer recommendations. The median age at diagnosis was 67 y in this patient series (range, 42-82 y). For comparison, a series of 504 genomic DNA samples was established from ethnically matched adult male blood donors at Hannover Medical School in the period from 2006 to 2007.

UKGPCS, ProtecT and PRACTICAL co-authorship list

The UK Genetic Prostate Cancer Study Collaborators

Listed on UKGPC website: www.icr.ac.uk/ukgpcs

The UK ProtecT Study Collaborators

Prasad Bollina, Sue Bonnington, Lynne Bradshaw, James Catto, Debbie Cooper, Liz Down, Andrew Doble, Alan Doherty, Garrett Durkan, Emma Elliott, David Gillatt, Pippa Herbert, Peter Holding, Joanne Howson, Mandy Jones, Roger Kockelbergh, Rajeev Kumar, Peter Holding, Howard Kynaston, Athene Lane, Teresa Lennon, Norma Lyons, Hing Leung, Malcolm Mason, Hilary Moody, Philip Powell, Alan Paul, Stephen Prescott, Derek Rosario, Patricia O'Sullivan, Pauline Thompson, Lynne Bradshaw, Sarah Tidball.

UK Genetic Prostate Cancer Study and The Prostate Cancer Research Foundation Study

The Institute of Cancer Research & The Royal Marsden NHS Foundation Trust, Sutton UK

Stephen Edwards

Cyril Fisher

Charles Jameson

Elizabeth Page

The ProtecT Study

Paul M. Brown

Anne George

Gemma Marsden

Athene Lane

Michael Davis

The UK ProtecT Study Collaborators – as above

AUSTRALIA (Melbourne)

John Pedersen

AUSTRALIA (Queensland)

Joanne Aitken

Robert A. Gardiner

Srilakshmi Srinivasan

Felicity Lose

Mary-Anne Kedda

Kimberly Alexander

Tracy O'Mara

Australian Prostate Cancer BioResource:

Gail Risbridger

Wayne Tilley

Lisa Horvarth

Australian Prostate Cancer Bio Resource-QLD node:

Peter Heathcote

Glenn Wood

Greg Malone

Hema Samaratunga
Pamela Saunders
Allison Eckert
Trina Yeadon
Kris Kerr
Angus Collins
Megan Turner

TASPRAC: Tasmania, Australia
Simon J. Foote
James R. Marthick
Andrea Polanowski
Rebekah M McWhirter
Terrence Dwyer
Christopher L. Blizzard

BULGARIA PCMUS study
Medical University – Sofia, Department of Urology
Elenko Popov
Molecular Medicine Centre and Department of Chemistry and Biochemistry
Darina Kachakova
Atanaska Mitkova
Teodora Goranova
Gergana Stancheva
Olga Beltcheva
Rumyana Dodova
Department of General and Clinical Pathology
Aleksandrina Vlahova
Tihomir Dikov
Svetlana Christova

DENMARK

AARHUS

Department of Urology, Aarhus University Hospital, Aarhus, Denmark

Prof. DMSc Michael Borre

CPCS1

Department of Urology, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark

Dr. Peter Klarskov

Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark

Dr. Sune F Nielsen

CPCS2

Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

Prof. DMSc Peter Iversen and Dr. M. Andreas Røder

Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark

Dr. Stig E Bojesen

GERMANY

ESTHER

Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

German Cancer Consortium (DKTK), Heidelberg, Germany

Hermann Brenner

Aida Karina Dieffenbach

ULM

Department of Urology, University Hospital Ulm, Germany

Christiane Maier

Antje Rinckleb

Manuel Luedeke

Mark Schrader

Institute of Human Genetics, University Hospital Ulm, Germany

Josef Hoegel

Walther Vogel

Department of Urology, Klinikum rechts der Isar der Technischen Universiteat München, Munich, Germany
Kathleen Herkommer

FINLAND

TAMPERE
The Finnish Cancer Registry
Liisa Määttänen
Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland
Prof. Teuvo Tammela
Department of Epidemiology, School of Health Sciences, University of Tampere, Finland
Prof. Anssi Auvinen

MAYO
Lori Tillmans
Shaun Riska
Liang Wang

MEC-BPC3
Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
Dan Stram
Epidemiology Program, University of Hawaii Cancer Centre, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA
Kolonel Laurence N.

MOFFIT
Julio Pow-Sang
Hyun Y. Park
Selina Radlein

Maria Rincon
James A Haley VA Hospital, Tampa, FL, USA
Babu Zachariah

ACKNOWLEDGEMENTS

This work was supported by 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).

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 FHCRC, Mayo, MCCS, Tampere, UKGPCS and Ulm groups are part of the ICPCG supported by NIH Grant No.U01 CA089600-04.

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 QLD research is supported by The National Health and Medical Research Council, Australia Project Grants [390130, 1009458 and 10507742] and Enabling Grant [614296 to APCB]; the Prostate Cancer Foundation of Australia (Project

Grant [PG7] and Research infrastructure grant [to APCB]). A Spurdle is an NHMRC Senior Research Fellow, J. Clements is an NHMRC Principal Research Fellow, and J. Batra is an NHMRC Early Career Fellow.

TASPRAC study has been supported by Cancer Australia (PdCCRS ID544938), National Health and Medical Research Council Program Grant, Australian Cancer Research Foundation infrastructure grant, Cancer Council Tasmania, Royal Hobart Hospital Research Foundation, Max Bruce Trust, and the Mazda Foundation. We would also like to acknowledge the ongoing support of our Tasmanian clinical collaborators and the participants in our prostate cancer research.

We 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, and [C490/A10124] and [C490/A16561] supported SERACH study. The ProtecT study is supported by the UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme, HTA 96/20/99; ISRCTN20141297. 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 Welcome 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 Defence 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.

We should like to acknowledge the NCRN nurses, data managers and Consultants for their work in the UKGPCS study.

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

The MEC was support by NIH grants CA63464, CA54281, CA098758, R01-CA085074, P50-CA105641, and P60-MD006900. The Moffitt group was supported by the US National Cancer Institute (R01CA128813, PI: J.Y. Park).

The USC study was supported by the US National Cancer Institute (R01CA84979) and by the California Cancer Research Program (99-00524V-10258).

The San Francisco Bay Area Prostate Cancer Study was supported by the California Cancer Research Fund (99-00527V-10182).

The Tampere (Finland) study was supported by the Academy of Finland Grant 116437 and 126714, The Finnish Cancer Organisations, Sigrid Juselius Foundation, Reino Lahtikari Foundation and The Medical Research Fund of Tampere University Hospital (# 9L091). The PSA screening samples were collected by the Finnish part of ERSPC (European Study of Screening for Prostate Cancer). Linda Enroth is thanked for technical assistance**.** Riitta Vaalavuo and Liisa Määttänen are thanked for their work with databases.

The FHCRC studies were supported by grants RO1 CA056678, RO1 CA082664, and RO1 CA092579 from the National Cancer Institute, National Institutes of Health, with additional support from the Fred Hutchinson Cancer Research Centre.

Genotyping was supported by the Intramural Program of the National Human Genome Research Institute, National Institutes of Health.

The PCMUS study was supported by the Bulgarian National Science Fund, Ministry of Education, Youth and Science (contract DOO-119/2009) with additional support from the Science Fund of Medical University - Sofia (contract 51/2009; 8I/2009).

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.

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

The Aarhus Prostate Cancer study was supported by the Danish Council for Strategic Research, The John and Birthe Meyer Foundation, The Lundbeck Foundation, and Aarhus University. The Danish Cancer Biobank (DCB) is acknowledged for biological material and for information regarding handling and storage.

We would to thank all the patients and control men who took part in this study.