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


Abstract

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

Methods: We genotyped 25 prostate cancer susceptibility loci in 40,414 individuals and derived a polygenic risk score (PRS). We estimated empirical odds ratios (OR) for prostate cancer associated with different risk strata defined by PRS and derived age-specific absolute risks of developing prostate cancer by PRS stratum and family history.

Results: The prostate cancer risk for men in the top 1% of the PRS distribution was 5.7 (95% CI, 1.6–5.7) 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 prostate cancer by age of 85 years 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 prostate cancer. The effect size, measured by OR per unit PRS, was higher in men at younger ages and in men with family history of prostate cancer. 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. Cancer Epidemiol Biomarkers Prev; 24(7); 1121–9. ©2015 AACR.

1Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom. 2Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge, United Kingdom. 3School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom. 4The University of Manchester, Centre for Epidemiology, Institute of Population Health, Manchester, United Kingdom. 5University of Warwick, University House, Coventry, United Kingdom. 6University of Tampere and FimLab Laboratories, Tampere, Finland. 7Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Centre, Los Angeles, California. 8University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom. 9School of Medical Biochemistry and Genetics, University of Turku, Turku, Finland. 10Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Centre, Los Angeles, California. 11Institute of Biomedical Technology/BioMediTech, University of Tampere and FimLab Laboratories, Tampere, Finland. 12Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Centre, Los Angeles, California. 13Faculty of Medical Science, University of Oxford, Oxford, United Kingdom.
Introduction

Genome-wide association studies (GWAS) have identified multiple common genetic variants associated with prostate cancer 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, because the risks associated with these variants are modest, large studies are required to estimate their risks precisely. To facilitate this estimation, we genotyped 25 prostate cancer susceptibility SNPs in studies from the PRACTICAL consortium. PRACTICAL is an international prostate cancer consortium that includes 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 used data from 31,833 cases and controls from 24 studies in PRACTICAL and 8,581 samples from replication stage of a GWAS (“GWAS stage III”). Sixteen out of the 25 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 III. 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 III 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 analyzed 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 UKGPC3) that were included in the GWAS stage III also contributed genotyping of additional samples for PRACTICAL phase III (Table 1; Supplementary Table S1 and Supplementary Data). 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 18 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 studies (MAYO, PCFS, TASPRA, UIM, and UTSAH) that oversampled cases with family history of prostate cancer. 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.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Authors: Ali Amin Al Olama, Sabine Benlloch, A.C. Antoniou, Z. Kote-Jarai, R.A. Eles and D.F. Easton jointly directed this work.

Current address for G. Severi: Human Genetics Foundation, Via Nizza 52, 10126 Torino, Italy.

Corresponding Authors: Ali Amin Al Olama, Centre for Cancer Genetic Epidemiology, Strangeways Research Laboratory, Worts Causeway, Cambridge, CB1 8RN, United Kingdom. Phone: 44-1223-748638; Fax: 44-1223-748628; E-mail: aa461@medschl.cam.ac.uk; and Douglas F. Easton, E-mail: dfe20@medschl.cam.ac.uk

doi: 10.1158/1055-9965.EPI-14-0317

©2015 American Association for Cancer Research.
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 (38 years; Supplementary Table S2A). 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 prostate cancer. Before excluding studies with oversampled familial cases, these percentages were 12.9% and 22.6%, respectively (Supplementary Table S2A and S2B).

Genotyping

Genotyping was performed in two experiments; these were subject to separate quality control procedures appropriate to the platforms used, before the data were combined for statistical analysis. In PRACTICAL phase III, genotyping of samples from 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 quality control criteria: SNP call rate >95%, no deviation from the Hardy–Weinberg equilibrium in controls at P < 0.00001; <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 quality control criteria were re-examined centrally.

GWAS Stage III genotypes were generated using an Illumina Golden Gate Assay. All SNPs for this analysis passed the quality control filters used for this experiment: call rate >95%, minor allele frequency in controls of >1%, or genotype frequency in controls consistent with the 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 prostate cancer 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 models with and without the interaction term for each pair of SNPs. The interaction test was the product of the allele doses for the two SNPs, hence leading to a 1-degree-of-freedom test for an interaction. On the basis of the assumption of a log-additive model, we constructed a PRS from the summed genotypes weighted by the estimated per-allele log-OR for each SNP, as estimated by logistic regression as above. Thus, for each individual j we derived:

\[
\text{Score}_j = \sum_{i=1}^{N} \beta_i g_{ij}
\]

Where:
- \(N\): number of SNPs (25)
- \(g_{ij}\): allele dose at SNP i (0, 1, 2) for individual j
- \(\beta_i\): per-allele log-OR 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 standardized the PRS by dividing by the overall standard deviation of PRS in the controls.

The risk of prostate cancer 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

---

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

<table>
<thead>
<tr>
<th>Study</th>
<th>Controls</th>
<th>Cases</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS stage III</td>
<td>4,076</td>
<td>4,505</td>
<td>8,581</td>
</tr>
<tr>
<td>PRACTICAL</td>
<td>16,050</td>
<td>15,783</td>
<td>31,833</td>
</tr>
<tr>
<td>Total</td>
<td>20,126</td>
<td>20,288</td>
<td>40,414</td>
</tr>
<tr>
<td>Total*</td>
<td>16,343</td>
<td>16,643</td>
<td>34,986</td>
</tr>
</tbody>
</table>

*Analyses were restricted to men of European ancestry (see the text).

*Total after excluding 5 studies that oversampled cases with family history.

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 prostate cancer, under a multiplicative model, was computed using the formula:

\[
\frac{\sum_{k} (\log \lambda_k)}{(\log \lambda_0)}
\]

where \(\lambda_0\) is the observed familial risk to first-degree relatives of prostate cancer cases, assumed to be 2 (12), and \(\lambda_k\) is the familial relative risk due to locus \(k\), given by:

\[
\lambda_k = \left( \frac{p_k f_k^2 + q_k}{p_k f_k^2 + q_k} \right)^7
\]

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 OR (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 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.
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 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 prostate cancer by PRS category and family history. Because we observed evidence for an interaction between PRS and age, we used both models with and without PRS × 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 prostate cancer for the U.K. population for 2012 (http://ci5.iarc.fr/CISplus; ref. 14). The model was adjusted for age at diagnosis (age <55, 55–59, 60–64, 65–70, and 70+ years). The procedure for deriving the age-specific incidences for each SNP profile category has been performed following the procedure explained by Antoniou and colleagues (15, 16), but adjusted to allow for competing causes of death.

For this purpose, we categorized 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 prostate cancer (P = 0.49) and assumed that family history and PRS are independently predictive of prostate cancer risk. Under this model, the prostate cancer incidence le(t) at age t for an individual in risk group k and family history group h, relative to the baseline category (h = 0, k = 1), approximated by the OR estimates from the logistic regression analysis. To obtain the baseline incidence, le(t), we constrained the prostate cancer incidence averaged all risk groups to agree with incidence rates, c incidence for each SNP pro

\[ l_a(t) = \mu(t) \]

Here, \( l_a(t) \) is the probability of having no family history in the population (89.26% in the controls in this dataset) and \( p_1 = 1 - l_o \) 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 (\( f_1 = 0.01, f_2 = 0.09, f_3 = 0.15, f_4 = 0.5, f_5 = 0.15, f_6 = 0.09, \) and \( f_7 = 0.01 \)) and \( S_k^i(t) \) is the probability of surviving prostate cancer by age t (in the risk group k for samples in the family history group h, which can be derived from incidence rates \( l^i_k(t) \) for ages < t using the formula \( S_k^i(t) = \exp \left( - \sum_{j=0}^{l^i_k(t) - 1} \right) \). Because definition \( S_k^i(0) = 1 \) for all k and h, it was possible to solve the above equation recursively, starting at age t = 0, to obtain the baseline incidences \( l_a(t) \) and hence the age-specific prostate cancer incidences at age t, \( l^i_k(t) \), 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:

\[ \sum_{m=1}^{55} \sum_{h} l^i_k(t) \times S_k^i(t) \]

Where \( S_k^i(t) = \exp \left( - \sum_{m=1}^{l^i_k(t) - 1} \mu_i(t) \right) \) is the probability of not dying from another cause of death by age t, based on the age-specific mortality rates \( \mu_i(t) \). The age-specific mortality rates, \( \mu_i(t) \), 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 prostate cancer death incidence per 100,000 individuals in the United Kingdom in year 2012 (14).

Results

All 25 SNPs showed evidence of association with prostate cancer (P = 0.02 to P = 1.4 × 10^{-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 rs6901979 on 8q24 (Table 2). For each of the 25 autosome 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 S3).

Thirteen of the SNPs (rs1465618, rs7679673, rs10486567, rs1447295, rs6983267, rs16901979, rs10993994, rs7931342, rs7127900, rs4430796, rs1649743, rs1859962, and rs5759167) showed a higher per-allele OR for cases with a prostate cancer family history than those without (P < 0.05), whereas no SNPs showed an effect in the opposite direction consistent with the predictions under a polygenic model (Supplementary Table S3; ref. 17).

Data on serum PSA level were available for 3,922 controls from six studies. Six SNPs (rs1447295, rs6983267, rs1512268, rs10993994, rs7127900, and rs2735839) showed association with PSA concentration levels significant at P < 0.03, rs1447295 showed an association with PSA in the opposite direction of the prostate cancer risk association but the rest of five SNPs showed an association with PSA in the same direction of the prostate cancer risk association (Supplementary Table S4).

Seven SNPs (rs1465618, rs12621278, rs10993994, rs7127900, rs1859962, rs7235839, and rs945619) 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 S5).

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 < 0.05 level for 29 pairs (out of 300 possible pairs) compared with 15 expected by chance, and significant at the P < 0.01 level for 12 pairs compared with three expected by chance. However, no pair was significant at the P < 0.05 level after a Bonferroni correction for the number of tests (nominal significance P = 1.6 × 10^{-4}; Supplementary Table S6).
The table below summarizes the results of 25 SNPs using PRACTICAL and GWAS stage III datasets in European populations. The table includes the marker name, chromosomal location, allele frequency, OR per unit increase of the PRS, and other relevant statistics. The data is analyzed with and without adjustment for age and family history.

### Table 3. ORs for prostate cancer by percentile of the PRS and family history

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>OR**&lt;br&gt;without adjustment</th>
<th>OR***&lt;br&gt;with adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1%</td>
<td>1 (baseline)</td>
<td>1 (baseline)</td>
</tr>
<tr>
<td>1%-10%</td>
<td>2.98 (1.66–5.35)</td>
<td>3.21 (1.93–5.28)</td>
</tr>
<tr>
<td>10%-20%</td>
<td>4.59 (2.58–8.17)</td>
<td>4.86 (2.67–8.88)</td>
</tr>
<tr>
<td>25%-50%</td>
<td>6.13 (3.65–10.24)</td>
<td>6.32 (3.59–11.12)</td>
</tr>
<tr>
<td>50%-75%</td>
<td>9.14 (5.13–16.23)</td>
<td>9.44 (5.25–16.92)</td>
</tr>
<tr>
<td>&gt;75%</td>
<td>17.58 (9.94–31.34)</td>
<td>17.93 (10.12–31.81)</td>
</tr>
</tbody>
</table>

**ORs obtained by fitting PRS group, family history, and age at diagnosis jointly.**

***ORs compared with men in the 1st percentile as baseline.**

**ORs compared with men in the 25th–75th percentile as baseline.**

---

Under the assumption that these 25 SNPs combined approximately multiplicatively to alter the risk of prostate cancer, we constructed a PRS for 16,643 cases and 18,343 controls based on the estimated per-allele ORs of 25 SNPs, standardized by the standard deviation in controls. The standardized PRS had a mean of 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 standardized PRS was strongly associated with disease risk (OR per unit PRS, 1.78; 95% CI, 1.70–1.83). The OR per unit increase of the standardized PRS declined with age from 1.48 (95% CI, 1.46–1.80) in cases diagnosed at age <55 years to 1.35 (95% CI, 1.20–1.52) in cases diagnosed at age >75+ (P = 2.0 × 10⁻⁴; Supplementary Table S5).

The OR per unit increase of PRS was larger for men with prostate cancer family history (1.79 vs. 1.70; P = 1.8 × 10⁻²; Supplementary Table S3). We found no evidence of an interaction between PRS and family history (P = 0.49) or between age at diagnosis and family history (P = 0.11), but there was some evidence for an interaction between PRS and age at diagnosis (P = 0.003).

There was no evidence of a difference in the OR per unit PRS according to Gleason score (OR, 1.75; 95% CI, 1.68–1.82) after adjusting for age at diagnosis and family history (P = 0.57; Supplementary Table S3). The correlation between PSA and the PRS was weak, both in controls (correlation = 0.09) and in cases (correlation = 0.02).

When PRS was categorized 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 Tables S7 and S8). 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.

---

Under the assumption that these 25 SNPs combined approximately multiplicatively to alter the risk of prostate cancer, we constructed a PPRS for 16,643 cases and 18,343 controls based on the estimated per-allele ORs of 25 SNPs, standardized by the standard deviation in controls. The standardized PRS had a mean of 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 standardized PRS was strongly associated with disease risk (OR per unit PRS, 1.78; 95% CI, 1.70–1.83). The OR per unit increase of the standardized PRS declined with age from 1.48 (95% CI, 1.46–1.80) in cases diagnosed at age <55 years to 1.35 (95% CI, 1.20–1.52) in cases diagnosed at age >75+ (P = 2.0 × 10⁻⁴; Supplementary Table S5).

The OR per unit increase of PRS was larger for men with prostate cancer family history (1.79 vs. 1.70; P = 1.8 × 10⁻²; Supplementary Table S3). We found no evidence of an interaction between PRS and family history (P = 0.49) or between age at diagnosis and family history (P = 0.11), but there was some evidence for an interaction between PRS and age at diagnosis (P = 0.003).
To estimate the absolute risk of prostate cancer for different risk groups defined by the combined genotypes at the 25 prostate cancer susceptibility loci, we fitted a logistic regression model that it included parameters for PRS (in seven categories) together with family history of prostate cancer once with (Supplementary Table S7) and once without a PRS to model that it included parameters for PRS (in seven categories) and family history of prostate cancer once with family history of prostate cancer once with family history (seven PRS groups and two family history; see Materials and Methods). We used both models (adjusted for age at diagnosis and family history) in order to estimate effect sizes for PRS. Then, we used the U.K. age-specific incidences of prostate cancer (0 to 85+ years; ref. 14) to estimate age-specific absolute risks of prostate cancer in the general population after considering competing causes of death for 14 risk groups defined by PRS and family history (seven PRS risk groups and two family history; see Materials and Methods). On the basis of this analysis, the absolute risk of prostate cancer by the age of 85 years for a man in the top 1% of the risk distribution with family history of prostate cancer 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 prostate cancer was 35.0% (36.1% 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 reason for this difference is unclear but might reflect interactions between the known susceptibility loci summarized 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 prostate cancer for a man at the age of 85 years with family history was estimated to be 26.5% when PRS information was ignored. When we incorporated PRS information, a man at the age of 85 years, depending 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 Figs. S1 and S3). These observations indicate that family history and the PRS independently influence risk and can be combined to provide stronger discrimination.

Chatterjee and colleagues (18) derived theoretical estimates for the predictive performance of polygenic models for 10 complex traits or common diseases, including prostate cancer, using published estimates for individual SNPs. They estimated that

Discussion

These results demonstrate that risk profiling based on SNPs can identify men at substantially increased or reduced risk of prostate cancer. 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 have also shown 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 prostate cancer 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 prostate cancer estimated on the basis of their ORs and allele frequencies in this study (see Materials and Methods). The reason for this difference is unclear but might reflect interactions between the known susceptibility loci summarized in the PRS and other factors influencing family history.
Risk Analysis of Prostate Cancer in PRACTICAL Consortium

Figure 2.
Absolute risk of prostate cancer by age in men with no family history.

approximately 7% of the population will be at 2-fold risk or greater for prostate cancer. We estimated, empirically, that the (average) risk to men in the 90% to 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% to 99% category, so that the percentile of the PRS fold relative to the population mean. However, this is an average risk over the 90% to 99% category, so that the percentile of the PRS at which the risk exceeds 2-fold will be >90%. On the basis of the estimated log(OR) per standardized PRS, approximately 6% of men will have a risk of greater than 2-fold, very close to the estimate of Chatterjee and colleagues (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 favorable among men at higher risk. If, for example, the benefit:cost ratio is favorable for screening men at a greater than 2-fold risk, the PRS provides an effective method for identifying such men.

Although 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 prostate cancer. 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.

In addition, the analyses presented here consider family history as a binary (yes/no) covariate. It is known that the risk of prostate cancer is dependent on both the number of affected relatives and their ages. MacInnis and colleagues (12, 20) have shown using segregation analysis that the familial aggregation of prostate cancer can be modeled 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 prostate cancer 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 nonaggressive disease, at least as measured by Gleason score, share these genetic risk factors as a common etiology.

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 characterize them. If such interactions could be identified reliably, they may improve the predictive value of the risk profiling, and also provide insights into the biologic interactions between the underlying risk variants.

Disclosure of Potential Conflicts of Interest

R.A. Eeles has received speakers bureau honoraria from Succinct Communication and medical education support. No potential conflicts of interest were disclosed by the other authors.
Amin Al Olama et al.

Authors’ Contributions
Development of methodology: A. Amin Al Olama, D.F. Easton, B.G. Nordestgaard, J.L. Hopper
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A. Amin Al Olama, A.C. Antoniou, N. Pashayan, K.D. Sorensen, H. Brenner, R. Stephenson, M.C. Southey, J.L. Hopper, D.F. Easton
Other (PI of a large study for which DNA samples and data are included as part of the PRACTICAL consortium): J.L. Stanford
Other (genotyping and phenotyping of participants): B.G. Nordestgaard
Other (generated laboratory data): J.L. Dickinson

Acknowledgments
The authors thank all the patients and control men who took part in this study. The authors also thank all the members from studies that they participated in this phase of PRACTICAL consortium: Esther John, Amit Joshi, Alva Shahabi, Joanne L. Dickinson, James R. Marthick, Mariciana C. Stern, Roman Corral, David M.A. Wallace, Alan Doherty, R.I. Bhatt, K. Subramanian, John Arndt, Louise Flanagan, Sita Ann Bradford, The UK Genetic Prostate Cancer Study Collaborators (www.iroc.ac.uk/ugkges), Prasad Bollina, Sue Bonnington, Lynne Bradshaw, James Catto, Debbie Cooper, Liz Down, Andrew Dobele, Alan Doherty, Garrett Durkan, Emma Elliott, David Gillatt, Pippa Herbert, Peter Holding, Joanne Howson, Mandy Jones, Roger Kockelbergh, Rajeev Kumar, Peter Holding, Howard Kynaston, Atheene Lane, Teresa Lennon, Norma Lyons, Hing Leung, Malcolm Mason, Hilary Moody, Philip Powell, Alan Paul, Stephen Prescott, Derek Rosario, Patricia O’Sullivan, Pauline Thompson, LynneBradshaw, Sarah Tidball, Paul M. Brown, Anne George, Gemma Marsden, Atheene Lane, Michael Davis, Stephen Edwards, Cyril Fisher, Charles Jameson, Elizabeth Page, John Pedersen, Joanne Aitken, Robert A. Gardiner, Srilakshmi Srinivasan, Felicity Lose, Mary-Anne Keldal, Kimberly Alexander, Tracy O’Mara, Gail Rishbridge, Wayne Tilley, Lisa Horvath, Peter Heathcoat, Glenn Wood, Greg Malone, Hema Samatarama, Pamela Saunders, Allison Eckert, Trina Yeadan, Kris Kerr, Angus Collins, Megan Turner, Simon J. Foote, James R. Marthick, Andrea Polanowski, Rebekah M. McWherter, Terrence Dwyer, Christopher L. Blizzard, Elenko Popov, Darina Kachakova, Atanaska Mitkova, Teodora Gordanova, Ergana Stancheva, Olga Belscheva, Rumyana Dodova, Aleksandra Vlahova, Tihomir Dikov, Svetlana Christova, Muhammed Beker, Peter Klarck, Sune F. Nielsen, Peter Iversen, Andreas Ruder, Stig E. Bojesen, Aida Karina Dieffenthaler, Manuel Lueneke, Mark Schrader, Josef Hoegel, Walther Vogel, Lisaa Maatianne, Teuvo Tammela, Anssi Auvinen, Lori Tillman, Shaun Riska, Liang Wang, Dan Stram, Kolene Laurence N., Julio Pow-Sang, Huy N. Park, Selina Radlein, Maria Rincon, and Babu Zachariah (Supplementary Data).

The Editor-in-Chief of Cancer Epidemiology, Biomarkers & Prevention is an author of this article. In keeping with the AACR’s Editorial Policy, the paper was peer reviewed and an AACR Journals’ Editor not affiliated with Cancer Epidemiology, Biomarkers & Prevention rendered the decision concerning acceptability.

Grant 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 1H11CA148537-01.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 16, 2014; revised January 15, 2015; accepted March 9, 2015; published onlineFirst April 2, 2015; DOI: 10.1158/1055-9965.EPI-14-0317

References


Cancer Epidemiology, Biomarkers & Prevention

Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci
Ali Amin Al Olama, Sara Benlloch, Antonis C. Antoniou, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-14-0317

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2015/06/11/1055-9965.EPI-14-0317.DC2

Cited articles
This article cites 22 articles, 1 of which you can access for free at:
http://cebp.aacrjournals.org/content/24/7/1121.full#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/24/7/1121.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.