Genetic Variation in Prostate-Specific Antigen–Detected Prostate Cancer and the Effect of Control Selection on Genetic Association Studies

Duleeka W. Knipe1, David M. Evans1,2, John P. Kemp1,2, Rosalind Eeles3,4, Douglas F. Easton5, Zsőfia Kote-Jarai3,4, Ali Amin Al Olama5, Sara Benlloch5, Jenny L. Donovan1, Freddie C. Hamdy7, David E. Neal4,6, George Davey Smith1,2, Mark Lathrop8,9, and Richard M. Martin1,2

Abstract

**Background:** Only a minority of the genetic components of prostate cancer risk have been explained. Some observed associations of SNPs with prostate cancer might arise from associations of these SNPs with circulating prostate-specific antigen (PSA) because PSA values are used to select controls.

**Methods:** We undertook a genome-wide association study (GWAS) of screen-detected prostate cancer (ProtecT: 1,146 cases and 1,804 controls); meta-analyzed the results with those from the previously published UK Genetic Prostate Cancer Study (1,854 cases and 1,437 controls); investigated associations of SNPs with prostate cancer using either "low" (PSA < 0.5 ng/mL) or "high" (PSA ≥ 3 ng/mL, biopsy negative) PSA controls; and investigated associations of SNPs with PSA.

**Results:** The ProtecT GWAS confirmed previously reported associations of prostate cancer at three loci: 10q11.23, 17q24.3, and 19q13.33. The meta-analysis confirmed associations of prostate cancer with SNPs near four previously identified loci (8q24.21, 10q11.23, 17q24.3, and 19q13.33). When comparing prostate cancer cases with low PSA controls, alleles at genetic markers rs1512268, rs445114, rs10788160, rs11199874, rs17632542, rs266849, and rs2735839 were associated with an increased risk of prostate cancer, but the effect-estimates were attenuated to the null when using high PSA controls ($P_{het}$ in effect-estimates < 0.04). We found a novel inverse association of rs9311171-T with circulating PSA.

**Conclusions:** Differences in effect-estimates for prostate cancer observed when comparing low versus high PSA controls may be explained by associations of these SNPs with PSA.

**Impact:** These findings highlight the need for inferences from genetic studies of prostate cancer risk to carefully consider the influence of control selection criteria. Cancer Epidemiol Biomarkers Prev; 23(7); 1356–65. ©2014 AACR.

Introduction

Prostate cancer is the leading cause of cancer in men in developed countries, accounting for 25% of new cancer cases. Prostate-specific antigen (PSA) testing is currently the most widely used screening test for prostate cancer, but it is limited because current thresholds defining raised levels (typically circulating PSA levels >3 or 4 ng/mL) do not distinguish clinically important from indolent cancer and lower levels do not exclude prostate cancer (1).

Epidemiologic studies suggest a significant genetic component in the etiology of the disease, but only 30% of the estimated heritability has so far been explained (2), highlighting the need for more studies to identify the full genetic profile of prostate cancer. The majority of genome-wide association studies (GWAS) of prostate cancer have been based on clinically detected cases (3–11), with some studies specifically focusing on young age at onset and/or familial or aggressive prostate cancer (i.e., pathologic stage T3/T4, N+, M+, Gleason score ≥ 7, grade III, metastases,
androgen ablation therapy, or PSA at least 10–50 ng/mL; refs. 4–6, 8–10, 12). Most studies have compared cases to controls that have been selected on the basis of PSA levels under a certain threshold (5, 6, 10, 12–17), including some studies that have used controls with extremely low PSA levels (<0.5 ng/mL; “supernormal” controls; refs. 3, 5, 6, 12, 15). Such control selection strategies can make it difficult to interpret associations between a genetic variant and disease, as such associations may result from a relationship between the variant and PSA levels rather than prostate cancer. A limited number of GWAS have attempted to clarify this relationship by analyzing associations of PSA levels with genetic markers identified using PSA screened controls (2, 5, 6, 15, 18, 19). It has been shown that SNPs at 13 known prostate cancer susceptibility loci (rs6869841, rs1270884, rs17632542, rs2242652, rs6983561, rs620861, rs10090154, rs12504426, rs7127900, rs10993994, rs2659056, rs2735839, rs5945619) are also associated with PSA concentration in blood in controls (2, 5, 6, 15, 19). As controls were defined by a PSA cutoff, however, associations between genotypes and the full distribution of PSA in men without prostate cancer could not be assessed. Only one study specifically investigated the relationship between genetic variants and circulating PSA level in men without detected prostate cancer (19). The Prostate Testing for cancer and Treatment study (ProtecT) contributed replication data to this study for the significant SNPs identified in the discovery GWAS. Genome-wide significant associations between PSA levels and markers at 6 different loci were reported (19).

Our investigation had 3 aims. First, to conduct a GWAS in a population-based screen detected prostate cancer population (ProtecT; ref. 20). Second, to conduct a meta-analysis to combine the results of the ProtecT GWAS with the previously published UK Genetic Prostate Cancer Study (UKGPCS; ref. 6). Third, to investigate whether previously identified associations of genetic variants with prostate cancer could be explained by associations with PSA levels, rather than cancer per se. For the third aim, (i) we identified previously published prostate cancer risk SNPs and compared their association with prostate cancer in 2 case–control studies nested within ProtecT; the first compared prostate cancer cases (diagnosed following a PSA ≥ 3 ng/mL and a positive biopsy) with “low” PSA controls (PSA < 0.5 ng/mL) and the second compared prostate cancer cases with “high” PSA controls (PSA ≥ 3 ng/mL and a negative biopsy); our hypothesis was that if the genetic marker is associated with PSA level and not prostate cancer, then the effect-estimates would be greatest when using the low PSA controls and close to the null when using the high PSA controls; and (ii) we examined the relationship of each of the identified genetic markers with PSA level as a continuous variable in controls selected independently of PSA level (“unrestricted” controls), including those with PSA levels ≥ 3 ng/mL who were biopsy negative, as well as those with levels below 3 ng/mL.

Materials and Methods

Samples

Two study populations were used in this analysis. First, participants from the ProtecT study (20), and second the UK Genetic Prostate Cancer Study stage I (UKGPCS), both of European ancestry (6). The ProtecT study recruited all men aged 50 to 69 years between 2001 and 2009 from 337 general practices in 9 UK centers (Birmingham, Bristol, Cambridge, Cardiff, Edinburgh, Leeds, Leicester, Newcastle, and Sheffield). All men were invited to a prostate check clinic and were offered a PSA test; approximately 110,000 attended. For those with a PSA level ≥ 3 ng/mL, a transrectal ultrasound-guided prostate biopsy was conducted. All detected tumors were histologically confirmed and assigned a Gleason score by a specialist uropathologist following a standard proforma. Tumors were categorized as low (score < 7), mid (score = 7), or high (score > 7) grade. Clinical staging was assigned using the tumor–node–metastasis (TNM) system (21) as either localized (T1–T2) or advanced (T3–T4); although most in the latter category were locally advanced and few tumors had metastasized distally. Approximately 3,000 men were diagnosed with prostate cancer between 2001 and 2009. All men diagnosed with prostate cancer after attending a prostate check clinic before end November 2006 (and their matched controls) were eligible for inclusion in the current ProtecT GWAS (n = 1,215 cases).

Men with no evidence of prostate cancer were eligible for selection as controls (approximate n = 107,000); these were men with a PSA < 3 ng/mL or a PSA ≥ 3 ng/mL with the most recent biopsy being negative. Controls were stratified matched to the genotyped prostate cancer cases by age, GP practice, and calendar time period, in 2 distinct rounds of matching. In the first round, a random sample of 6 controls per case were selected from all eligible controls; as the prostate check clinics were completed in the same age and GP practice strata as the index case ("unrestricted" controls); as the prostate check clinics were completed in one GP practice before moving to the next, matching on GP practice also matched on calendar time period. The second round randomly selected 1 control with a PSA < 0.5 ng/mL (and with whole blood collected) per case from the same age and GP practice strata as the index case ("supernormal controls"). A total of 1,925 matched controls were eligible for inclusion in the current GWAS. The Trent Multicenter Research Ethics Committee (MREC) approved the ProtecT study (MREC/01/4/025) and the associated ProMPT study that collected biologic material (MREC/01/4/061), and written informed consent was obtained.

The UKGPCS (stage I) dataset has been previously described (6). Briefly the cases (n = 2,017) were clinically detected and selected if the man had a diagnosis at ≤ 60 years of age or a first-/second-degree family history of prostate cancer. Self-reported “non-white” men were excluded, as were those who were diagnosed through asymptomatic screening. UKGPCS controls (total n = 1,893) were solely selected from men in the ProtecT study
who had a PSA < 0.5 ng/mL. In the current meta-analysis, we excluded from the UKGPCS results those men who were in the ProtecT GWAS control series described above (n = 456) so that the populations in the pooled analysis were independent samples (Fig. 1).

**Genotyping in ProtecT**

The genotyping was performed at the Center National de Genotypage, Evry, France using the Illumina Human660W-Quad_v1_A array (Illumina, Inc.). Quality control measures included exclusions for sex-mismatches, minimal (<0.325) or excessive heterozygosity (>0.345), cryptic relatedness as estimated by proportion of loci identical by descent (IBD > 0.1), and disproportionate levels of individual missingness (>3%). Individuals were then checked for evidence of population stratification by multidimensional scaling analysis and compared with HapMap II (release 22) European descent (CEU), Han Chinese (CHB), Japanese (JPT), and Yoruba (YRI) reference populations. Individuals showing evidence of non-European ancestry were removed. SNPs with a minor allele frequency of below 1%, a call rate of <95%, or evidence for violation of Hardy–Weinberg equilibrium (P < 5 × 10⁻⁸) were discarded.

Imputation of autosomal genotypic data used Markov Chain Haplotyping software (MACH v.1.0.16; ref. 22) and phased haplotype data from CEU individuals (HapMap release 22, Phase II NCBI Build 36, dbSNP 126) on 514,432 autosomal SNPs. All SNPs with poor imputation quality (r² hat < 0.3) were removed. The final working dataset included 2,950 individuals (cases, n = 1,146; controls, n = 1,804). A total of 1,272 controls were included from the first round of “unrestricted” control selection (PSA < 0.5 ng/mL, n = 238; PSA ≥ 0.5 and PSA < 3 ng/mL, n = 941; PSA ≥ 3 ng/mL, n = 93) and a further 532 from the second round (“supernormal” control selection).

**UKGPCS genotyping**

This has been reported elsewhere (6). Briefly, genotypes were generated using the Illumina Infinium Human-Hap550 array, and we only used genotypes with a call rate of >97%. Related samples and those with Asian/African ancestry were excluded. After exclusions, 1,854 cases and 1,437 controls (independent of ProtecT—see Fig. 1) were used in the current meta-analysis.

**Statistical analysis**

The statistical analysis was undertaken in 4 stages (see Fig. 2). In the first stage, a genome-wide association analysis was performed on the ProtecT samples using the software package MACH2DAT (22, 23), using a logistic regression model based on an expected allelic dosage model for SNPs. Associations between each SNP and disease were assessed by a Wald test and P < 5 × 10⁻⁸ was considered genome-wide significant. All estimates are reported in the direction of the risk allele.

In the second stage, we meta-analyzed the top SNPs identified in ProtecT with the UKGPCS dataset. A total of 381 SNPs reaching a threshold for suggestive significance of 5 × 10⁻⁵ in the ProtecT GWAS were included in the meta-analysis. The threshold level of 5 × 10⁻⁵ was a pragmatic choice based on similar thresholds for suggestive significance cited in the literature, because of the relatively small sample size in ProtecT. Using a fixed-effect inverse variance model, the β-coefficients and SEs from ProtecT and UKGPCS were combined using the metan command in STATA. A P value of 5 × 10⁻⁸ was used to identify genome-wide significant SNPs in the
meta-analysis. Heterogeneity between the ProtecT and UKGPCS studies was evaluated using Cochran’s Q statistic and the I² value (24). The 2 studies had the same strand orientation and used the same effect alleles. Associations showing moderate-to-high levels of heterogeneity (I² > 50) were investigated further to determine whether the heterogeneity could be explained by the difference in control or case populations used in ProtecT and UKGPCS. We did this in 2 stages: first, we used only the low PSA (<0.5 ng/mL) controls in ProtecT to generate new effect-estimates and re-ran the meta-analysis for those SNPs showing moderate-to-high levels of heterogeneity in the original analysis; second, for those SNPs that still showed moderate-to-high levels of heterogeneity, we restricted the ProtecT cases to include only young cases (≤60 years of age) and /or those with a family history and re-ran the meta-analysis.

In stage 3, we investigated how the choice of controls in case–control GWAS studies may have impacted on associations of SNPs with prostate cancer, using 2 control populations from the ProtecT study: 770 low PSA controls (PSA < 0.5 ng/mL, from both the “unrestricted” control selection, n = 238, and "supernormal" control selection, n = 532); and 93 high PSA controls (PSA ≥ 3 ng/mL). We also did a sensitivity analysis relaxing the high PSA control threshold to a PSA level ≥ 2 ng/mL, resulting in 250 high PSA controls. We did this to increase power and move our high control population toward a more general population. We generated a list of SNPs to be tested by downloading all published SNPs associated with both prostate cancer risk and PSA level from the catalogue of Published Genome-wide Association Studies (http://www.genome.gov/gwastudies/: downloaded 21/11/2012). We supplemented this list with additional SNPs associated with prostate cancer in a recent meta-analysis (2), resulting in a total of 83 SNPs. Using this list, we extracted the individual genotypic data for all participants in our ProtecT dataset, although 2 SNPs were neither genotyped nor imputed (rs7210100 and rs16902094). Using the genotypic data, we generated effect-estimates for each SNP using multinomial logistic regression, where each control group (low and high) was compared with cases, separately (Fig. 2). To test the null hypothesis, that there was no difference between the 2 control groups when used to estimate the ORs for the associations of SNPs with prostate cancer, we carried out a separate logistic regression model that tested the association of each SNP with control type (low vs. high); this is the same
as testing whether the ratio of ORs using the 2 control groups were equivalent (i.e., \(-1\)), as the base case group would be cancelled out (Fig. 2).

Finally, we conducted a regression analysis of SNPs on PSA level within the “unrestricted” control population (controls selected independent of PSA level). PSA values were log-transformed to approximate a normal distribution. To account for the multiple comparisons between PSA levels and SNP markers, we only highlight findings that showed evidence of an association in both the stratified analysis and a complementary association in the regression analysis.

The anti-log of \( \beta \) was taken to estimate the percentage change in PSA level by increasing allele dose. To incorporate the uncertainty of the imputation process, expected allelic dosage data were used for the analysis. Genotypic dosages are the expected number of copies of an allele that ranges from 0 to 2. For genotyped SNPs, the exact number of alleles at the genetic markers rs1512268 (19q13.33), and rs17632542 was greatly attenuated when the analysis was repeated using the ProtecT effect-estimates from a restricted control population (low PSA controls < 0.5 ng/mL; Supplementary Table S3). The heterogeneity observed for rs1447295, however, remained high (\( I^2 = 83.1\% \), \( P_{\text{heterogeneity}} = 0.01 \)). When the analysis was repeated using ProtecT effect-estimates based on young and/or familial cases and low PSA controls, the heterogeneity was only slightly reduced and remained statistically significant (\( I^2 = 78\% \), \( P_{\text{heterogeneity}} = 0.03 \)).

Results

GWAS in ProtecT (stage 1)

The study characteristics of those individuals (cases: \( n = 1,146 \), controls: \( n = 1,804 \)) included in the ProtecT GWAS are presented in Supplementary Table S1. Supplementary Figure S1 shows the quantile–quantile (Q–Q) plot of the distribution of test statistics for comparison of genotype frequencies in cases versus controls. The lambda inflation factor was 1.025. In GWAS, 381 SNPs were associated with prostate cancer at \( P < 5 \times 10^{-8} \) (Supplementary Fig. S2 and Supplementary Table S2). We detected a genome-wide significant association (\( P < 5 \times 10^{-8} \)) between prostate cancer and SNPs at 3 previously identified loci: 10q11.23 (rs10993994), 17q24.3 (rs7222314), and 19q13.33 (rs1058205; Table 1 and locus zoom plots in Supplementary Figs. S3–S5).

Meta-analysis of ProtecT and UKGPCS studies (stage 2)

All SNPs at \( P < 5 \times 10^{-8} \) in ProtecT were meta-analyzed with results from the UKGPCS. The total number of individuals included was 6,241 (cases: ProtecT \( n = 1,146 \), UKGPCS \( n = 1,854 \); controls: ProtecT \( n = 1,804 \), UKGPCS \( n = 1,437 \)). A total of 175 SNPs reached genome-wide significance (\( P < 5 \times 10^{-8} \)) for an association with prostate cancer in the meta-analysis and were all located in or near genes across 4 previously identified loci (Table 2), with 6 independent signals. There was evidence of significant heterogeneity between the studies for the rs1447295 (8q24.21) marker (\( I^2 = 86.1\% \), \( P_{\text{heterogeneity}} = 0.01 \)). There was also evidence of moderate-to-high levels of heterogeneity for rs6983267, rs10993994, and rs17632542. The heterogeneity of rs6983267, rs10993994, and rs17632542 was greatly attenuated when the analysis was repeated using the ProtecT effect-estimates from a restricted control population (low PSA controls < 0.5 ng/mL; Supplementary Table S3). The heterogeneity observed for rs1447295, however, remained high (\( I^2 = 83.1\% \), \( P_{\text{heterogeneity}} = 0.01 \)). When the analysis was repeated using ProtecT effect-estimates based on young and/or familial cases and low PSA controls, the heterogeneity was only slightly reduced and remained statistically significant (\( I^2 = 78\% \), \( P_{\text{heterogeneity}} = 0.03 \)).

Stratified analysis using different control populations in ProtecT and PSA regressed in the unrestricted controls (stages 3 and 4)

Of the 81 SNPs examined, SNPs showing evidence of heterogeneity (\( P < 0.05 \)) in effect-estimates quantifying the difference in risk between cases and low or high PSA controls and/or SNPs that showed at least nominal association with PSA levels in unrestricted controls (\( P < 0.05 \)) are shown in Table 3. When comparing cases with low PSA controls, alleles at the genetic markers rs1512268 (8p21.2), rs445114 (8q24.21), rs10788160 (10q26.12), rs11199874 (10q26.12), rs17632542 (19q13.33), rs266849 (19q13.33), and rs2735839 (19q13.33) were strongly associated with an increased risk of prostate cancer (all \( P \leq 0.01 \)), with ORs ranging from 1.24 to 2.73. The effect-estimates, however, remained high (\( I^2 = 78\% \), \( P_{\text{heterogeneity}} = 0.03 \)). When the analysis was repeated using ProtecT effect-estimates based on young and/or familial cases and low PSA controls, the heterogeneity was only slightly reduced and remained statistically significant (\( I^2 = 78\% \), \( P_{\text{heterogeneity}} = 0.03 \)). There was also some evidence of associations of these SNPs with PSA levels in the “unrestricted” control.

Table 1. Results of independent SNPs reaching genome-wide significance (\( P < 5 \times 10^{-8} \)) in the ProtecT GWAS

<table>
<thead>
<tr>
<th>Loci</th>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>Alleles</th>
<th>( \beta^a )</th>
<th>SE</th>
<th>OR(^b)</th>
<th>( P )</th>
<th>EAF</th>
<th>RSQR(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10q11.23</td>
<td>rs10993994</td>
<td>10</td>
<td>51219502</td>
<td>T/C</td>
<td>0.33</td>
<td>0.06</td>
<td>1.39</td>
<td>1.58 ( \times 10^{-9} )</td>
<td>0.5983</td>
<td>0.9793</td>
</tr>
<tr>
<td>17q24.3</td>
<td>rs7222314</td>
<td>17</td>
<td>66616533</td>
<td>A/G</td>
<td>0.32</td>
<td>0.06</td>
<td>1.38</td>
<td>8.59 ( \times 10^{-9} )</td>
<td>0.4772</td>
<td>0.9295</td>
</tr>
<tr>
<td>19q13.33</td>
<td>rs1058205</td>
<td>19</td>
<td>56055210</td>
<td>T/C</td>
<td>0.40</td>
<td>0.07</td>
<td>1.49</td>
<td>4.63 ( \times 10^{-8} )</td>
<td>0.1828</td>
<td>1</td>
</tr>
</tbody>
</table>

Abbreviation: EAF, effect allele frequency.

\(^a\)Correlation coefficient.

\(^b\)Effect-estimate (OR).

\(^c\)Imputation accuracy.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome Position</th>
<th>Gene</th>
<th>Effect Allele</th>
<th>EAF</th>
<th>SE</th>
<th>P (95% CI)</th>
<th>Phet</th>
<th>EAFb</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs12682344</td>
<td>8</td>
<td>SRRM1P1</td>
<td>C0</td>
<td>0.79</td>
<td>0.14</td>
<td>9.66 x 10^-12</td>
<td>0.04</td>
<td>21.7%</td>
</tr>
<tr>
<td>rs12682344</td>
<td>8</td>
<td>SRRM1P1</td>
<td>C2</td>
<td>0.79</td>
<td>0.14</td>
<td>9.66 x 10^-12</td>
<td>0.04</td>
<td>21.7%</td>
</tr>
<tr>
<td>rs12682344</td>
<td>8</td>
<td>POU5F1B</td>
<td>C0</td>
<td>0.67</td>
<td>0.08</td>
<td>1.28 x 10^-10</td>
<td>0.04</td>
<td>12.2%</td>
</tr>
<tr>
<td>rs12682344</td>
<td>8</td>
<td>POU5F1B</td>
<td>C2</td>
<td>0.67</td>
<td>0.08</td>
<td>1.28 x 10^-10</td>
<td>0.04</td>
<td>12.2%</td>
</tr>
<tr>
<td>rs6983267</td>
<td>8</td>
<td>SRRM1P1</td>
<td>C0</td>
<td>0.29</td>
<td>0.06</td>
<td>2.30 x 10^-12</td>
<td>0.04</td>
<td>72.0%</td>
</tr>
<tr>
<td>rs6983267</td>
<td>8</td>
<td>SRRM1P1</td>
<td>C2</td>
<td>0.29</td>
<td>0.06</td>
<td>2.30 x 10^-12</td>
<td>0.04</td>
<td>72.0%</td>
</tr>
<tr>
<td>rs1447295</td>
<td>8</td>
<td>POU5F1B</td>
<td>C0</td>
<td>0.67</td>
<td>0.08</td>
<td>1.28 x 10^-10</td>
<td>0.04</td>
<td>12.2%</td>
</tr>
<tr>
<td>rs1447295</td>
<td>8</td>
<td>POU5F1B</td>
<td>C2</td>
<td>0.67</td>
<td>0.08</td>
<td>1.28 x 10^-10</td>
<td>0.04</td>
<td>12.2%</td>
</tr>
<tr>
<td>rs9311171</td>
<td>10</td>
<td>MYC</td>
<td>C0</td>
<td>0.46</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
<tr>
<td>rs9311171</td>
<td>10</td>
<td>MYC</td>
<td>C2</td>
<td>0.46</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
<tr>
<td>rs10993994</td>
<td>10</td>
<td>MSMB</td>
<td>C0</td>
<td>0.36</td>
<td>0.05</td>
<td>3.20 x 10^-12</td>
<td>0.04</td>
<td>7.3%</td>
</tr>
<tr>
<td>rs10993994</td>
<td>10</td>
<td>MSMB</td>
<td>C2</td>
<td>0.36</td>
<td>0.05</td>
<td>3.20 x 10^-12</td>
<td>0.04</td>
<td>7.3%</td>
</tr>
<tr>
<td>rs17632542</td>
<td>17</td>
<td>SOX9</td>
<td>C0</td>
<td>0.82</td>
<td>0.08</td>
<td>4.74 x 10^-10</td>
<td>0.04</td>
<td>62.7%</td>
</tr>
<tr>
<td>rs17632542</td>
<td>17</td>
<td>SOX9</td>
<td>C2</td>
<td>0.82</td>
<td>0.08</td>
<td>4.74 x 10^-10</td>
<td>0.04</td>
<td>62.7%</td>
</tr>
<tr>
<td>rs9311171</td>
<td>19</td>
<td>KLK3</td>
<td>C0</td>
<td>0.29</td>
<td>0.06</td>
<td>2.30 x 10^-12</td>
<td>0.04</td>
<td>7.3%</td>
</tr>
<tr>
<td>rs9311171</td>
<td>19</td>
<td>KLK3</td>
<td>C2</td>
<td>0.29</td>
<td>0.06</td>
<td>2.30 x 10^-12</td>
<td>0.04</td>
<td>7.3%</td>
</tr>
<tr>
<td>rs7611694</td>
<td>19</td>
<td>MSMB</td>
<td>C0</td>
<td>0.40</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
<tr>
<td>rs7611694</td>
<td>19</td>
<td>MSMB</td>
<td>C2</td>
<td>0.40</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
<tr>
<td>rs6869841</td>
<td>19</td>
<td>MSMB</td>
<td>C0</td>
<td>0.40</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
<tr>
<td>rs6869841</td>
<td>19</td>
<td>MSMB</td>
<td>C2</td>
<td>0.40</td>
<td>0.05</td>
<td>2.15 x 10^-10</td>
<td>0.04</td>
<td>10.0%</td>
</tr>
</tbody>
</table>

**Discussion**

Our GWAS and meta-analysis replicate previous findings for the following loci: 10q11.23 (rs10993994 - MSMB), 17q24.3 (rs222314 - CALM2P1 - SOX9), and 19q13.33 (rs1058205 - KLK3) in the ProtecT GWAS; and 8q24.21 (rs12682344, rs6983267, and rs1447295 - intergenic), 10q11.23 (rs10993994 - MSMB), 17q24.3 (rs7611694 - SOX9), and 19q13.33 (rs17632542 - KLK3) in the meta-analysis. We also demonstrated a novel association of one SNP with circulating PSA level: rs9311171-T allele only increased PSA levels by 1% (P = 0.09; Supplementary Table S4).

The T-allele at 3p22.2 (rs9311171) was associated with a more pronounced increase in risk of prostate cancer when cases were compared with high PSA controls (79% increase) as opposed to low PSA controls (7% increase; Phet = 0.01). There was also evidence of an association between the rs9311171-T allele and decreased PSA levels (P = 0.035). In the sensitivity analysis (high PSA controls ≥ 2 ng/mL), rs4775302 (15q21.1) was associated with a larger increase in risk of prostate cancer when cases were compared with the high PSA controls (29% increase), as opposed to the low PSA controls (2% increase; Phet = 0.03; Supplementary Table S4). This was supported by the regression analysis of PSA in the “unrestricted” control population, which showed an association in the opposite direction (5% decrease; P = 0.09).

**Effect of PSA-Detected Controls on Genetic (GWAS) Studies**

The marker rs9311171 is located near a microRNA (MIR26A1) and a protein linked to oncogenesis (CTD5FL). This SNP (rs9311171) had a larger positive association with prostate cancer when cases were compared with high PSA, as opposed to low PSA controls. This suggests that this marker is directly associated with both PSA level and prostate cancer risk.
Table 3. Comparison of cases with supernormal or high PSA controls and PSA regressed on SNPs in unrestricted controls

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Putative gene</th>
<th>Risk allele</th>
<th>OR (SE)</th>
<th>P</th>
<th>OR (SE)</th>
<th>P</th>
<th>P_{het}^a</th>
<th>β</th>
<th>% PSA change</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>rs9311171</td>
<td>CTDSPL</td>
<td>T</td>
<td>1.07 (0.10)</td>
<td>0.49</td>
<td>1.79 (0.34)</td>
<td>0.002</td>
<td>0.01</td>
<td>-0.09</td>
<td>-9%</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>rs7611694</td>
<td>SIRT1</td>
<td>A</td>
<td>1.09 (0.63)</td>
<td>0.19</td>
<td>0.87 (0.19)</td>
<td>0.40</td>
<td>0.17</td>
<td>0.07</td>
<td>8%</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>rs6869841</td>
<td>FAM44B (BOD1)</td>
<td>T</td>
<td>1.35 (0.11)</td>
<td>3.44 × 10^{-4}</td>
<td>1.22 (0.24)</td>
<td>0.29</td>
<td>0.63</td>
<td>0.08</td>
<td>9%</td>
<td>0.03</td>
</tr>
<tr>
<td>6</td>
<td>rs1983891</td>
<td>FOXP4</td>
<td>T</td>
<td>1.10 (0.08)</td>
<td>0.22</td>
<td>0.96 (0.16)</td>
<td>0.80</td>
<td>0.43</td>
<td>0.08</td>
<td>8%</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>rs1512268</td>
<td>SLC25A37 - NKX3-1</td>
<td>T</td>
<td>1.30 (0.09)</td>
<td>9.51 × 10^{-5}</td>
<td>0.83 (0.13)</td>
<td>0.24</td>
<td>0.004</td>
<td>0.06</td>
<td>6%</td>
<td>0.07</td>
</tr>
<tr>
<td>8</td>
<td>rs445114</td>
<td>SRRM1P1 - POU5F1B</td>
<td>T</td>
<td>1.46 (0.10)</td>
<td>8.15 × 10^{-6}</td>
<td>0.92 (0.16)</td>
<td>0.60</td>
<td>0.01</td>
<td>0.08</td>
<td>8%</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>rs6983267</td>
<td>SRRM1P1 - POU5F1B</td>
<td>G</td>
<td>1.29 (0.05)</td>
<td>8.51 × 10^{-5}</td>
<td>1.15 (0.13)</td>
<td>0.36</td>
<td>0.44</td>
<td>0.08</td>
<td>9%</td>
<td>0.01</td>
</tr>
<tr>
<td>10</td>
<td>rs11199874</td>
<td>RPL19P16 - FGFR2</td>
<td>A</td>
<td>1.24 (0.06)</td>
<td>0.01</td>
<td>0.78 (0.21)</td>
<td>0.14</td>
<td>0.01</td>
<td>0.07</td>
<td>7%</td>
<td>0.07</td>
</tr>
<tr>
<td>10</td>
<td>rs10788160</td>
<td>RPL19P16 - FGFR2</td>
<td>A</td>
<td>1.24 (0.06)</td>
<td>0.01</td>
<td>0.80 (0.21)</td>
<td>0.18</td>
<td>0.01</td>
<td>0.06</td>
<td>6%</td>
<td>0.08</td>
</tr>
<tr>
<td>10</td>
<td>rs3850699</td>
<td>TRIM8</td>
<td>A</td>
<td>1.17 (0.06)</td>
<td>0.00</td>
<td>0.82 (0.22)</td>
<td>0.28</td>
<td>0.048</td>
<td>0.01</td>
<td>1%</td>
<td>0.73</td>
</tr>
<tr>
<td>10</td>
<td>rs10993994</td>
<td>MSMB</td>
<td>T</td>
<td>1.61 (0.11)</td>
<td>4.54 × 10^{-12}</td>
<td>1.45 (0.23)</td>
<td>0.02</td>
<td>0.51</td>
<td>0.08</td>
<td>8%</td>
<td>0.02</td>
</tr>
<tr>
<td>13</td>
<td>rs9600079</td>
<td>FABP5L1 - KLF12</td>
<td>G</td>
<td>1.07 (0.06)</td>
<td>0.33</td>
<td>0.99 (0.16)</td>
<td>0.97</td>
<td>0.65</td>
<td>0.06</td>
<td>6%</td>
<td>0.048</td>
</tr>
<tr>
<td>19</td>
<td>rs17632542</td>
<td>KLK3</td>
<td>G</td>
<td>2.73 (0.32)</td>
<td>1.40 × 10^{-17}</td>
<td>0.76 (0.27)</td>
<td>0.44</td>
<td>1.94 × 10^{-4}</td>
<td>0.28</td>
<td>32%</td>
<td>5.49 × 10^{-7}</td>
</tr>
<tr>
<td>19</td>
<td>rs2735839</td>
<td>KLK3 - KLK2</td>
<td>G</td>
<td>2.13 (0.20)</td>
<td>1.90 × 10^{-16}</td>
<td>1.29 (0.28)</td>
<td>0.26</td>
<td>0.02</td>
<td>0.21</td>
<td>23%</td>
<td>0.004</td>
</tr>
<tr>
<td>19</td>
<td>rs266849</td>
<td>KLK15 - KLK3</td>
<td>A</td>
<td>1.76 (0.05)</td>
<td>8.62 × 10^{-12}</td>
<td>1.15 (0.17)</td>
<td>0.49</td>
<td>0.03</td>
<td>0.12</td>
<td>13%</td>
<td>2.40 × 10^{-6}</td>
</tr>
</tbody>
</table>

NOTE: SNPs showing evidence of heterogeneity (P < 0.05) in effect-estimates quantifying the difference in risk between cases and low or high PSA controls and/or SNPs that showed at least nominal association with PSA levels in unrestricted controls (P < 0.05).

^aTests the hypothesis that there is no difference between the effect-estimate of the cases versus low PSA controls and cases versus high PSA controls.

Knipe et al. Cancer Epidemiol Biomarkers Prev; 23(7) July 2014
We also found evidence that the association of prostate cancer risk of 7 other SNPs at loci 8p21.2, 10q26.12, and 19q13.33 could be confounded by the variant-PSA association (rs1512268-T, rs10788160-A, rs45114-T, rs11199874-A, rs17632542-T, rs2735839-G, and rs266849-A). The association between prostate cancer and the respective SNP allele was present in the comparison of cases with low PSA controls, but this association disappeared when compared with high PSA controls. There was also evidence of associations between these SNPs and PSA levels in the “unrestricted” control population. Associations of rs1512268-T, rs10788160-A, and rs11199874-A with PSA levels were only statistically significant \( P < 0.05 \) in the stratified analysis. While associations of SNPs with PSA levels were not statistically significant in the regression analysis, the direction of the effect provides some support for the conclusion that these SNP alleles are associated with increased circulating PSA level. Eeles and colleagues (5) and Gudmundsson and colleagues (19) previously observed a positive association of PSA with rs1512268-T (near NXX3.1) and rs10788160-A (FGFR2), respectively. The allele at the genetic marker rs445114 (intergenic) was previously shown to have a positive association with PSA level, albeit non-genome-wide significant \( P = 1.27 \times 10^{-2} \), ref. 19. This is consistent with the findings of Al Olama and colleagues (18), who found that an SNP (rs620861) in strong linkage disequilibrium with rs445114 was positively associated with PSA level \( P = 4.8 \times 10^{-7} \). While rs11199874-A has not previously been tested for a relationship with PSA, it is in strong linkage disequilibrium with rs10788160-A \( \left( r^2 = 0.94 \right) \) FGFR2, which has been shown to be associated with increasing PSA levels. Our analysis suggests that rs10788160-A is associated with higher PSA levels, whereas rs10788160-G is associated with lower levels. A previous study reported a positive association between rs11199874-G and prostate cancer (25). This association could be driven by the large proportion (50%) of high PSA \( (4–10 \text{ ng/mL}) \) controls used rather than reflecting a true association because this high PSA–enriched control sample is likely to have a higher proportion of individuals with the allele rs11199874-A; this could in turn account for the positive association of prostate cancer with rs11199874-G and prostate cancer (the opposite allele to rs11199874-A). Two markers in strong linkage disequilibrium (rs17632542-T and rs2735839-G), located near the kallikrein-related peptidase protein-coding genes, have previously been observed to be positively associated with PSA level (19). SNP rs266849 is also located near but is independent of this set of PSA-associated proteins; a previous study reported no association of rs266849 with PSA in controls with a PSA < 10 ng/mL (6). When we reduced our threshold for defining high PSA controls \( (PSA \geq 2 \text{ ng/mL}) \) in the stratified analysis, we observed similar findings in the same direction to those that were found to be significant using the higher PSA threshold \( (PSA \geq 3 \text{ ng/mL}) \). In addition, in this sensitivity analysis, 3 other SNPs were shown to be associated with PSA.

**Impact of control selection**

Our analysis suggests that the selection of controls for GWAS below a certain PSA threshold results in associations with prostate cancer, when the actual association is with PSA level (e.g., rs6869841). In addition, the selection of these controls can lead to the masking of a prostate cancer association (e.g., rs931171). To avoid these difficulties, an optimal design for a GWAS would be to select controls who have PSA levels at or above the threshold for biopsy and who are biopsy negative. These criteria ensure that the cases and controls are comparable in terms of PSA level and therefore associations of SNP markers with prostate cancer are not confounded by associations with PSA.

**Strengths and limitations**

This study was based on a population sample of men who underwent PSA testing, providing findings relevant to screen-detected cancer, and controls that were not selected according to PSA level. We also conducted a meta-analysis using 6,241 individuals (3,000 cases and 3,241 controls), but we were only able to confirm previous GWAS findings rather than show novel associations. There was significant heterogeneity between the ProtecT and UKGPCS studies for associations of markers at 8q24.21, 10q11.23, and 19q13.33 with prostate cancer. These differences were partly explained by the difference in control ascertainment between studies (Supplementary Table S3). We were, however, unable to explain the substantial heterogeneity \( (P = 78\%) \) for rs1447295. We had a relatively small number of high PSA controls \((n = 93)\) so our stratified analysis will have been underpowered, explaining why we were unable to detect some previously published associations of SNPs with PSA. To increase power, we carried out a sensitivity analysis by relaxing the high PSA control threshold to \( \geq 2 \text{ ng/mL} \) \((n = 250)\), which resulted in 3 additional SNPs being associated with PSA level \((rs6869841, rs10993994, \text{ and } rs4775302)\).

To avoid potential false-positive results arising from multiple comparisons, we only highlight findings that showed evidence of an association in both the stratified and regression analyses. A Bonferroni-corrected threshold significance level is \( P \leq 0.0006 \) \((0.05/81 \text{ tests})\). However, this correction assumes that the tests were independent of each other, which in this case may not be true because some SNPs were correlated. Hence, a rigid application of the Bonferroni-corrected \( P \) value could lead to an erroneous rejection of genuine loci (false-negatives). Nevertheless, we have presented exact \( P \) values, allowing comparison with the Bonferroni-corrected \( P \) values. While all controls with an elevated PSA level were biopsy negative, there is the possibility that some controls with a PSA level < 3 ng/mL had undetected prostate cancer, as evidenced by a study (1) showing that 14% men with a PSA level < 3 ng/mL had prostate cancer. The majority of controls in our study had a PSA level \( \leq 1 \text{ ng/mL} \), which was associated...
with less than a 9% chance of undetected prostate cancer prostate cancer (1). It is also possible that up to a quarter of men with an elevated PSA level (>3 ng/mL) and a negative biopsy had undetected prostate cancer (26), which could explain why we did not observe associations of SNPs with prostate cancer when using the high PSA control group.

Conclusion

We have confirmed associations of prostate cancer risk with 4 previously identified loci (8q24.21,10q11.23, 17q24.3, and 19q13.33) and associations of 7 markers with PSA level. We also found new evidence of an association of genetic variation at 3p22.2 with circulating PSA levels. We have highlighted that the method of selecting controls in case–control studies of associations of genetic variation with prostate cancer can influence the results, suggesting that inferences made in these studies should carefully consider control selection.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: D.W. Knipe, J.P. Kemp, J.L. Donovan, G. Davey Smith, R.M. Martin

Development of methodology: J.P. Kemp, M. Lathrop

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R. Eeles, D.F. Easton, Z. Kote-Jarai, S. Benlloch, J.L. Donovan, F.C. Hamdy, D.E. Neal

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): D.W. Knipe, D.M. Evans, J.P. Kemp, D.F. Easton, A.A. Al Olama, F.C. Hamdy, M. Lathrop, R.M. Martin

Writing, review, and/or revision of the manuscript: D.W. Knipe, D.M. Evans, J.P. Kemp, R. Eeles, D.F. Easton, A.A. Al Olama, J.L. Donovan, F.C. Hamdy, G. Davey Smith, M. Lathrop, R.M. Martin

Administrative, technical, or material support (i.e., reporting and organizing data, constructing databases): Z. Kote-Jarai, S. Benlloch

Study supervision: J.L. Donovan, R.M. Martin

Acknowledgments

The authors thank the tremendous contribution of all members of the ProtecT study research group, and especially the following who were involved in this research (Athene Lane, Prasad Bolina, Sue Bennington, Lynn Bradshaw, James Catto, Debbie Cooper, Michael Davis, Liz Down, Andrew Doble, Alan Doherty, Garrett Durkan, Emma Elliott, David Gittall, Pippa Herbert, Peter Holding, Joanne Howson, Mandy Jones, Roger Kockelbergh, Howard Kynaston, Teresa Lennon, Normal Lyons, Hing Leung, Malcolm Mason, Hilary Moody, Philip Powell, Alan Paul, Stephen Prescott, Derek Rosario, Patricia O’Sullivan, Pauline Thompson, Sarah Tidball). They also thank Emma Mansden, who processed the blood samples at the biorepository, and Dr. Chris Metcalfe for providing statistical advice and the Center National de Genotypage for genotyping the ProtecT samples.

Grant Support

The ProtecT study is funded by the U.K. Health Technology Assessment (HTA) Programme of the NIHR Research (HTA 96/20/99; ISRCTN20141297). D.W. Knipe is funded by the Wellcome Trust 4-year studentship (WT099874MA). D.M. Evans, J.P. Kemp, G. Davey-Smith, and R.M. Martin work in an MRC Unit that is supported by the UK Medical Research Council (MC_UU_12013/1-9) and the University of Bristol. D.F. Easton is supported by a Cancer Research UK Grant (C1297/A10118). R.A. Eeles and Z. Kote-Jarai are supported by Cancer Research UK Grant (C5047/A7357) and support from the NIHR to the Biomedical Research Centre at The Institute of Cancer Research and Royal Marsden NHS Foundation Trust. J.F.C. Handy, D.E. Neal, and J.L. Donovan are NIHR Senior Investigators. The authors thank the provision of the additional epidemiological data by the NHS R&D Directorate supported Prodigal study and the ProMPT (Prostate Mechanisms of Progression and Treatment) collaboration which is supported by the National Cancer Research Institute (NCRI) formed by the Department of Health, the Medical Research Council, and Cancer Research UK (G050966/75466).

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 August 29, 2013; revised March 5, 2014; accepted March 21, 2014; published OnlineFirst April 21, 2014.

References


