A meta-analysis of genome-wide association studies to identify prostate cancer susceptibility loci associated with aggressive and non-aggressive disease


†These authors jointly directed this work.
‡These authors contributed jointly to this work.

© The Author 2012. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
Norihiko Tsuchiya$^{49}$, Shintaro Narita$^{49}$, Guang-Wen Cao$^{57}$, Chavdar Slavov$^{80}$, Vanio Mitev$^{56}$, The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons’ Section of Oncology$^4$, The UK ProtecT Study Collaborators$^6$, The Australian Prostate Cancer Bioresource$^1$, The PRACTICAL Consortium$^6$, Stephen Chanock$^6,7$, Henrik Gronberg$^1,8$, Christopher A. Haiman$^3,7$, Peter Kraft$^{15,9}$, Douglas F. Easton$^1,8$ and Rosalind A. Eeles$^{2,83,*,†}$

$^1$Strangeways Laboratory, Centre for Cancer Genetic Epidemiology, Worts Causeway, Cambridge CB1 8RN, UK, $^2$The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK, $^3$Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Noirs Comprehensive Cancer Centre, Los Angeles, CA, USA, $^4$Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm SE-171 77, Sweden, $^5$Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA, $^6$Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, 723 Swanston Street, Carlton, VIC 3053, Australia, $^7$Department of Urology, Aarhus University Hospital, Skejby, Denmark, $^8$Australian Prostate and Biostatistics, School of Public Health, Imperial College London, UK, $^9$Lyon Cancer Research Center, INSERM Department of Medicine, Brigham and Women's Hospital, University of Cambridge, Box 279, Hills Road, Cambridge, UK, $^{10}$Li Ka Shing Centre, Cancer Research UK Cambridge Research Institute, Cambridge CB2 2QQ, UK, $^{11}$Nuffield Department of Surgery and $^{12}$Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK, $^{13}$Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK, $^{14}$School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK, $^{15}$Program in Molecular and Genetic Epidemiology, Department of Epidemiology and $^{16}$Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA, $^{17}$Epidemiology Research Program, American Cancer Society, Atlanta, GA 30303, USA, $^{18}$Massachusetts Veterans Epidemiology and Research Information Center (MAVERIC) and Geriatric Research, Education, and Clinical Center (GRECC), Boston Veterans Affairs Healthcare System, Boston, MA 02114, USA, $^{19}$Division of Aging, Department of Medicine, Brigham and Women’s Hospital, Boston, MA 02215, USA, $^{20}$Department of Genomics of Common Disease, School of Public Health, Imperial College, London SW7 2AZ, UK, $^{21}$Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK, $^{22}$Lyon Cancer Research Center, INSERM U1052, Lyon, France, $^{23}$Genomic Epidemiology Group, $^{24}$Division of Clinical Epidemiology and Aging Research and $^{25}$Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany, $^{26}$Division of Urologic Surgery, Washington University School of Medicine, St Louis, MO, USA, $^{27}$Division of Epidemiology, Department of Environmental Medicine, NYU Langone Medical Centre, NYU Cancer Institute, New York, NY 10016, USA, $^{28}$Institute of Biomedical Technology/BioMediTech, University of Tampere and $^{29}$Department of Pathology, Centre for Laboratory Medicine, Tampere University Hospital, Tampere, Finland, $^{30}$Department of Epidemiology, School of Health Sciences and $^{31}$Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland, $^{32}$Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland, $^{33}$Department of Clinical Biochemistry and $^{34}$Department of Urology, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, Herlev DK-2730, Denmark, $^{35}$Division of Public Health Sciences, Fred Hutchinson Cancer Research Centre, Seattle, WA, USA, $^{36}$Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA, $^{37}$National Human Genome Research Institute, National Institutes of Health, 50 South Drive, Room 5351, Bethesda, MD, USA, $^{38}$Department of Genetics and Pathology, International Hereditary Cancer Center, Pomeranian Medical University, Szczecin, Poland, $^{39}$Mayo Clinic, Rochester, MN, USA, $^{40}$Department of Molecular Medicine and $^{41}$Department of Urology, Aarhus University Hospital, Skejby, Denmark, $^{42}$Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation, Queensland University of Technology,

$^{*}$To whom correspondence should be addressed at: Oncogenetics Team, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. Tel: +44 2086613642; Fax: +44 2087701489; Email: rosalind.eeles@icr.ac.uk

$^1$UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons’ Section of Oncology—membership list provided in Supplementary Material, Notes.

$^2$UK ProtecT Study Collaborators—membership list provided in Supplementary Material, Notes.

$^3$Full list of participants of The Australian Prostate Cancer BioResource is provided in Supplementary Material, Notes.

$^4$Full list of participants of The PRACTICAL Consortium is provided in the Supplementary Material, Notes.

$^5$These authors jointly directed this work.
Genome-wide association studies (GWAS) have identified multiple common genetic variants associated with an increased risk of prostate cancer (PrCa), but these explain less than one-third of the heritability. To identify further susceptibility alleles, we conducted a meta-analysis of four GWAS including 5953 cases of aggressive PrCa and 11,463 controls (men without PrCa). We computed association tests for approximately 2.6 million SNPs and followed up the most significant SNPs by genotyping 49,121 samples in 29 studies through the international PRACTICAL and BPC3 consortia. We not only confirmed the association of a PrCa susceptibility locus, rs11672691 on chromosome 19, but also showed an association with aggressive PrCa \[\text{odds ratio } 1.12 \pm 0.09\] (95% confidence interval 1.03–1.21), \[P = 1.4 \times 10^{-8}\]. This report describes a genetic variant which is associated with aggressive PrCa, which is a type of PrCa associated with a poorer prognosis.

INTRODUCTION

Genome-wide association studies (GWAS) have identified more than 50 common variants associated with susceptibility to prostate cancer (PrCa). However, these variants explain less than a third of the familial risk of the disease, indicating that further susceptibility loci remain to be identified. Moreover, few variants identified by GWAS have thus far been
shown to be associated with aggressive PrCa. Although the current set of SNPs contributes to overall PrCa risk prediction, overall they do not discriminate men who will develop aggressive disease, a clinically more relevant outcome. A recent GWAS of aggressive PrCa identified a novel susceptibility locus on 2q37.3, but the per-allele odds ratio (OR) did not differ between aggressive and non-aggressive cases in the replication stage (1). It has been also reported that a genetic variant in DAP2IP might be associated with the risk of aggressive PrCa (2). A recent GWAS and validation study of aggressive PrCa found an SNP on 15q13, rs6497287, to be uniquely associated with this disease trait \((P_{\text{replication}} = 0.004)\); however, lack of power due to small numbers and a non-significant test for heterogeneity between less and more aggressive PrCa warrants further investigation of this finding (3). Lin et al. (4) have reported an association of PrCa mortality with five germline SNPs from a candidate gene analysis.

RESULTS

In an attempt to identify susceptibility loci for aggressive PrCa, we conducted a meta-analysis of four GWAS (Table 1 and Supplementary Material, Notes). We also included data from the second stage of the UK study, that was genotyped for 43 671 SNPs showing evidence for association in stage 1 (5). These studies included, after quality control (QC) exclusions (see Materials and Methods), a total of 11 463 controls and 11 085 cases. For the present analysis, we included data from 5953 cases with aggressive disease defined as having a Gleason score of 8 or greater (with the exception of the BPC3 study, which also includes cases with tumor stage C or greater, and the CGEMS study, which also included cases with a Gleason score of 7) and all controls (it was ensured that there was no overlap between the studies). Following imputation using HapMap Phase II CEU as a reference, approximately 2.6 million genotyped and imputed SNPs were assessed in each GWAS study using a 1 df trend tests for association. Combined association tests were generated using a fixed effects meta-analysis (see Materials and Methods).

In the combined analysis, two loci, rs11672691 on 19q13 \((P\text{-value} = 3.8 \times 10^{-7})\) and rs11704416 on 22q13 \((P = 7.0 \times 10^{-6})\), showed strong evidence for association. rs11672691 is in the same region as rs887391 \((r^2 = 0.9)\) that it was previously reported to be associated with PrCa by Hsu et al. (6), but it did not reach GWAS significance level in that report. These two SNPs were selected for further replication analysis in two international consortia, PRACTICAL and BPC3. The present analysis was restricted to 24 395 cases (2008 aggressive) and 24 726 controls (17 445 controls and 11 085 cases). For the present analysis, we included data from 5953 cases with aggressive disease defined as having a Gleason score of 8 or greater, and the CGEMS study, which also included cases with tumor stage C or greater, and the BPC3 study, which also includes cases with tumor stage C or greater, and the CGEMS study, which also included cases with a Gleason score of 7) and all controls (it was ensured that there was no overlap between the studies). Following imputation using HapMap Phase II CEU as a reference, approximately 2.6 million genotyped and imputed SNPs were assessed in each GWAS study using a 1 df trend tests for association. Combined association tests were generated using a fixed effects meta-analysis (see Materials and Methods).

In the combined analysis, two loci, rs11672691 on 19q13 \((P\text{-value} = 3.8 \times 10^{-7})\) and rs11704416 on 22q13 \((P = 7.0 \times 10^{-6})\), showed strong evidence for association. rs11672691 is in the same region as rs887391 \((r^2 = 0.9)\) that it was previously reported to be associated with PrCa by Hsu et al. (6), but it did not reach GWAS significance level in that report. These two SNPs were selected for further replication analysis in two international consortia, PRACTICAL and BPC3. The present analysis was restricted to 24 395 cases (2008 aggressive) and 24 726 controls (17 445 controls and 17 445 controls in aggressive disease analysis) from 26 studies from European populations (Table 1, Supplementary Material, Table S1 and Supplementary Material, Notes show all 29 studies, 26 of which are European).

SNP rs11672691 showed evidence of replication \((P = 0.006)\) with a genome-wide significance of \(P = 1.4 \times 10^{-8}\) in a combined analysis across all stages (Table 2) for aggressive PrCa. When data from non-aggressive cases were also included, the overall evidence for association was stronger \((P = 2.2 \times 10^{-12}, \text{overall})\). The per-allele OR for aggressive PrCa in the replication stage \([1.12, 95\% \text{ confidence interval (CI)} 1.03–1.21; P = 0.006]\) was higher than that for non-aggressive cases \((OR 1.08, 95\% \text{ CI} 1.05–1.12; P = 8.2 \times 10^{-7})\); however, the difference was not statistically significant \((P\text{-value} = 0.18)\). SNP rs11704416 showed evidence of replication for all PrCa \((P = 0.002)\), but did not quite reach genome-wide significance overall \((P = 3.7 \times 10^{-7})\). The evidence for association with aggressive disease was weaker \((P = 0.16 \text{ in the replication, } P = 4.0 \times 10^{-6} \text{ overall})\). There was no evidence that either locus was associated with serum PSA (based on 1578 control samples; Supplementary Material, Table S2). SNP rs11672691 showed stronger effect \((P = 0.02)\) when we compared cases with a family history of PrCa \((OR 1.14, 95\% \text{ CI} 1.06–1.22)\) with those with no family history \((OR 1.06, 95\% \text{ CI} 1.02–1.10)\). The per-allele ORs did not differ significantly by ages (Supplementary Material, Table S3). Considering the estimated ORs in the replication stage, rs11672691 and rs11704416 together explain \(~0.16\%\) of the familial risk of PrCa.

### DISCUSSION

rs11672691 lies between \(ATP5SL\) and \(CEACAM21\) (Fig. 1A) and within a hypothetical locus, \(LOC100505349\), of a non-coding RNA. \(ATP5SL\) codes for an ATP synthase-like protein whose function is unknown; however, a variant in this gene has been associated with adult height (7). The carcinoma embryonic antigen (CEA) gene family belongs to the immunoglobulin super family of genes. Several CEA subgroup members possess cell adhesion properties and some seem to function in signal transduction or regulation of signal transduction, possibly in association with other CEA sub-family members (8). Several of these proteins show a complex expression pattern in normal and cancerous tissues. Both \(CEACAM5\) and \(CEACAM6\) have a role in cell adhesion, invasion and metastasis (9), and are known to be overexpressed in a majority of carcinomas, including those of the gastrointestinal tract, the respiratory and genitourinary systems and breast cancer. The closest gene, \(CEACAM21\), has been considered as a candidate gene for type 1 diabetes (10). A region on 19q13 (HPCQTL19) has been reported previously in a genetic linkage study to be a QTL for aggressive PrCa when the
Gleason score was used as a quantitative measure of tumor aggressiveness (11).

SNP rs11704416 lies upstream of TNRC6B on chromosome 22 (Fig. 1B). The TNRC6 (trinucleotide repeat containing 6) family of proteins have been shown to stably associate with argonaute (AGO) proteins. AGO proteins, through their association with small RNAs, perform a critical function in the effector step of RNA interference. TNRC6B protein has a role in translational inhibition through its binding to AGOs (12).

These results illustrate the value of combining GWAS to confirm candidate loci where the genome-wide significance threshold was not obtained, and improve power identifying susceptibility loci associated with sub-classifications of diseases. The original report by Hsu et al. (6) implicating the 19q13 region failed to reach genome-wide significance, whereas our findings verify a significant association. Although some samples overlap between the Hsu et al. report and our study, we expanded the discovery phase by incorporating the Stage 1 UK and Stage 2 UK/Melbourne participants and including additional samples in the replication stage. The identification of loci involved in PrCa aggressiveness has been hampered by relatively small sample sizes. The locus reported here is associated with both aggressive and non-aggressive diseases, and is therefore likely to be useful in determining those with clinically significant PrCa. Identification of such loci would aid the understanding of the biology of PrCa progression and targeted screening based on genetic risk profiling for aggressive disease.

### MATERIALS AND METHODS

#### Samples

The four GWAS data sets have been described previously (1,5,13,14) (Table 1). Analyses were based on the data sets following standard QC procedures as previously described (5). The replication stage included 25 072 cases (2160 cases with a Gleason score of 8+) and 25 536 controls (18 255 in aggressive disease analysis) from 29 PrCa case–control studies (Supplementary Material, Table S1 and Notes). All studies were approved by the appropriate ethics committees.

#### Genotyping

In BPC3 and PRACTICAL, genotyping of samples from 13 studies was performed by the KASPar assay.

### Table 2. Meta-analysis and replication results

<table>
<thead>
<tr>
<th>SNP chromosome allele position</th>
<th>Analysis</th>
<th>Study</th>
<th>OR^a (95% CI)</th>
<th>P-value</th>
<th>P-value combined</th>
<th>P-value combined all</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11672691</td>
<td>Aggressive disease cases</td>
<td>Stage 1 UK</td>
<td>1.20 (1.05–1.35)</td>
<td>0.02</td>
<td>3.8 × 10⁻⁷</td>
<td>1.4 × 10⁻⁸</td>
</tr>
<tr>
<td>rs11704416</td>
<td>Aggressive disease cases</td>
<td>Stage 1 UK</td>
<td>0.85 (0.69–1.02)</td>
<td>0.056</td>
<td>3.3 × 10⁻⁶</td>
<td>4.0 × 10⁻⁶</td>
</tr>
</tbody>
</table>

Cases not classified as aggressive were those without the features defined in the text for aggressive disease.

^aPer-allele OR for the first allele.

^bNA: imputation quality was poor (0.227) and this result was excluded.
Figure 1. Regional plots of the associated SNPs at (A) 19q13 and (B) 22q13. Circles and diamonds denote imputed and genotyped SNPs, respectively. Plots show the genomic regions associated with PrCa and $-\log_{10}$ association $P$-values of SNPs. Also shown are the SNP build 36/hg18 coordinates in kilo bases, recombination rates and genes in the region. The intensity of red shading indicates the strength of LD ($r^2$) with the index SNP. Plots were drawn with a modified R script from http://www.broadinstitute.org/mpg/snap/ldplot.php.
scores and the true genotypes (analysis if the estimated correlation between the genotype formed using MACH 1.0. The CAPS study used IMPUTE chromosome X. The imputation for the BPC3 study was performed using the HapMap phase 2 CEU population as a reference. UK Genotypes were imputed for approximately 2.6 million SNPs.

**Statistical methods**

**Imputation**

Genotypes were imputed for approximately 2.6 million SNPs using the HapMap phase 2 CEU population as a reference. UK stages 1 and 2 and CGEMS were imputed using MACH 1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/) to impute genotypes of autosomal markers and IMPUTE v1 (15) for chromosome X. The imputation for the BPC3 study was performed using MACH 1.0. The CAPS study used IMPUTE v1. We included imputed data from an SNP in the combined analysis if the estimated correlation between the genotype scores and the true genotypes \( r^2 \) was >0.3 (MACH) or where the quality information was >0.3 (IMPUTE).

**Analyses**

For UK stages 1 and 2 and CGEMS, the imputed genotype probabilities were used to derive a 1 df association score statistic for each SNP, and its corresponding variance. The test statistic for UK stage 2 was stratified by population as described previously (5). In the BPC3 study, estimated betas and standard errors were calculated for each component study, including one principal component as a covariate to adjust for population structure, using ProbABEL (16), and the results were combined to generate overall betas and standard errors, using a fixed effects meta-analysis. CAPS used SNPTEST (https://mathgen.stats.ox.ac.uk/genetics_software/snpset/snpset.html) to estimate betas and standard errors. We converted the results from all studies into test scores and variances from each stage. SNPs were selected for validation on the basis of a significance level of \( P < 10^{-7} \) in a combined meta-analysis of UK stages 1 and 2, CGEMS, CAPS and BPC3, excluding SNPs that were correlated with known susceptibility SNPs (SNP rs1704416 was included since it reached \( P < 10^{-7} \) in an initial analysis). A total of 1921 subjects of non-European ancestry (Asian and African-American) were excluded from all analyses. Analyses were performed based on 2008 aggressive disease PrCa cases (out of 24395 cases) and 17445 controls (out of 24726 controls). OR and 95% CI were estimated using unconditional logistic regression, stratified by study. In the text, we have reported the combined tests of association over all stages in European populations, but have emphasized the OR estimates from the replication stage to minimize the effect of ‘winner’s curse’. ORs were computed separately for subsets of cases defined by family history, grade and age. Modification of the ORs by family history and grade was assessed using a case-only analysis, using the dichotomous variable as the endpoint (family history Yes versus No/Grade GS < 8 versus GS ≥ 8). Modification of the ORs by grade as a continuous covariate, and by age, was assessed using a case-only analysis, using polytomous regression with SNP genotype (scored 0, 1, 2) as the endpoint. The associations between SNP genotypes and PSA level were assessed using linear regression after log-transformation of PSA level to correct for skewness. We performed analyses for both all stages and only aggressive cases of PrCa. Analyses were performed in R principally using GenABEL (17), SNPTEST and ProbABEL (16) and Stata.

**Publication of GWAS data**

The U19, which provides funding for this work, plans to post summary data from this study onto a share point hosted by the NIH, by the end of 2012.

**URLS**


https://mathgen.stats.ox.ac.uk/genetics_software/snpset/snpset.html. (last accessed date 17 October, 2012)

http://www.sph.umich.edu/csg/abecasis/MACH/. (last accessed date 17 October, 2012)


**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.
ACKNOWLEDGEMENTS

Acknowledgements are detailed in Supplementary Material, Notes.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by Cancer Research UK grants (grant numbers: C5047/A7357, C1287/A10118, C5047/A3354, C5047/A10692, C16913/A6135) and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant (grant number: 1 U19 CA 148537-01 (the GAME-ON initiative)). Further support is detailed in Supplementary Material, Notes.

REFERENCES