Brief Original Contribution

A Replication Study Examining Novel Common Single Nucleotide Polymorphisms Identified Through a Prostate Cancer Genome-wide Association Study in a Japanese Population


* Correspondence to Dr. Jyotsna Batra, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Brisbane, Queensland 4059, Australia (e-mail: jyotsna.batra@qut.edu.au).

Initially submitted April 12, 2011; accepted for publication July 1, 2011.

Five novel prostate cancer risk loci were identified in a recent genome-wide association study (GWAS) of Japanese persons (Takata et al., Nat Genet. 2010;42(9):751–754). Those authors proposed that apart from population-specific linkage disequilibrium patterns, limitations of GWAS single nucleotide polymorphism (SNP) prioritization and/or study design could explain the lack of identification of these loci in GWAS previously conducted among Caucasians. Thus, the authors undertook a replication study in 1,357 prostate cancer patients and 1,403 healthy Australian males of European descent (2004–2008). The rs12653946 SNP at 5p15 was found to be significantly associated with prostate cancer risk (odds ratio = 1.20, 95% confidence interval: 1.07, 1.34; P = 0.002). On the basis of linkage disequilibrium calculations, the rs12653946 SNP represents an independent locus, distinct from the previously identified TERT-CLPTM1L cancer nexus region. Further, analysis from AceView (Thierry-Mieg and Thierry-Mieg, Genome Biol. 2006;7(suppl 1):S12) indicated that rs12653946 falls within the intron of a testis-expressed gene strongly predicted to translate a conceptual 8.1-kilodalton protein named tojy.aApr07. The authors’ findings suggest that follow-up of apparently ethnicity-specific risk associations are warranted in order to highlight risk-associated loci for experimental studies and for incorporation into future risk prediction models for prostate cancer.

Abbreviations: CLPTM1L, cisplatin resistance-related protein CRR9p; GWAS, genome-wide association study(ies); PSA, prostate-specific antigen; SNP, single nucleotide polymorphism; TERT, telomerase reverse transcriptase.

Genome-wide association studies (GWAS) have led to the identification of more than 30 common, low-penetrance loci for prostate cancer, principally through 4 GWAS carried out among persons of European ancestry (United Kingdom, Iceland, Sweden, and the United States) (1–6). Together these explain more than 20% of the familial risk (7). In replication studies, most (but not all) risk alleles have been shown to confer similar effects in different ethnic populations (8–10). Recently, Takata et al. (11) carried out a GWAS of prostate cancer in Japanese persons, comprising 1,583 cases and 3,386 controls in stage 1 and 3,001 cases and 5,415 controls in the replication stage. Stage 1 analysis confirmed associations with 8 independent loci previously identified in studies of Europeans, and an additional 5 novel loci were validated after the replication study. Plausible candidate genes were suggested for 3 of these 5 loci: GPRC6A (G protein-coupled receptor, family C, group 6, member A), C2orf43 (open reading frame 43 of chromosome 2), and FOXP4 (forkhead box P4), while the other 2 loci were reported to represent a gene desert region (11).

In previously conducted prostate cancer GWAS or follow-up and meta-analysis studies of GWAS-identified single nucleotide
polymorphisms (SNPs) carried out in European populations (12, 13), investigators have not found any association with the 5 risk-associated SNPs identified in the Japanese GWAS. Takata et al. (11) stated that the minor allele frequency for these SNPs was not markedly different between European and Japanese samples and speculated that the linkage disequilibrium patterns between unknown causative variants and marker SNPs differ between Japanese and European populations. An alternative explanation is that the sample sizes of the individual prostate cancer GWAS conducted to date have been too small at stage 1 and/or stage 2 to generate sufficient statistical power to prioritize all SNPs associated with a modest effect. In most of the prostate cancer GWAS described in the literature, researchers have selected only SNPs with \( P \) values at GWAS levels of significance \( (10^{-5}) \) for the follow-up analysis or the final publication reports; thus, a study with an average sample size of 1,500 patients and controls had only about 20%–70% power to prioritize SNPs of effect size 1.15–1.20. In addition, investigators in each of these GWAS have imposed criteria for selection of cases and/or controls at stage 1, such as serum prostate-specific antigen (PSA) levels and reported family history, which may have biased SNP prioritization.

We undertook a replication study on the 5 newly identified Japanese prostate cancer GWAS SNPs (rs13385191, rs12653946, rs1983891, rs339331, and rs9600079) in a large cohort of 1,357 prostate cancer patients and 1,403 healthy Australian males. We also used additional database searches to critically interpret our data and the results obtained in the Japanese GWAS (11).

**MATERIALS AND METHODS**

**Study subjects**

All cases \( (n = 1,357) \) had histopathologically confirmed prostate cancer, following an abnormal serum PSA level and/or lower urinary tract symptoms, and reported Caucasian ethnicity. Cases were ascertained consecutively during 2004–2008 in Queensland, Australia, as follows: 1) physician referrals from 3 hospitals in Brisbane \( (n = 155) \); age range, 51–87 years) made within 2 years of prostate cancer diagnosis (the Retrospective Queensland Study (14)); 2) 857 newly diagnosed prostate cancer patients (range of ages at diagnosis, 43–88 years) recruited from 26 private practices and 10 public hospitals in Brisbane and greater Queensland through clinician referrals to a longitudinal cohort study (the Prostate Cancer Supportive Care and Patient Outcomes Project (15)); and 3) 345 additional newly diagnosed prostate cancer patients (range of ages at diagnosis, 43–85 years) recruited through urologist referrals from greater Queensland into the Queensland node of an ongoing biobanking study (the Australian Prostate Cancer BioResource (http://www.apccbioresource.org.au/index.html)). Epidemiologic and clinical information was collected by questionnaire and from clinical reports, although cleaned data on these variables were not available for the Australian Prostate Cancer BioResource at the time of analysis.

Male controls \( (n = 1,403) \) with reported Caucasian ethnicity and no self-reported personal history of prostate cancer were ascertained in Queensland as follows: 1) 836 male blood donors (range of ages at interview, 18–75 years) were recruited through the Australian Red Cross Blood Service in Brisbane and 2) 567 men (range of ages at interview, 54–90 years) were randomly selected from the Australian electoral roll (voting is compulsory in Australia) and age- (in 5-year groups) and postcode-matched to cases from the Prostate Cancer Supportive Care and Patient Outcomes Project. Analyses excluded 67 persons with an age at interview under 43 years, the age of the youngest case.

The study protocol was approved by the human research ethics committees of the Queensland University of Technology, the Queensland Institute of Medical Research, the Mater Hospital (for Brisbane Private Hospital), the Royal Brisbane Hospital, Princess Alexandra Hospital, and the Cancer Council Queensland. All participants gave written informed consent.

**DNA isolation and genotyping**

Methods used for DNA preparation and genotyping have been described previously (16). Briefly, germ-line DNA was extracted from peripheral blood using the QIAGEN DNA isolation kit (QIAGEN GmbH, Hilden, Germany) for all men recruited into the study. Five SNPs (rs13385191, rs12653946, rs1983891, rs339331, and rs9600079) were genotyped using iPLEX Gold assays on the Sequenom MassARRAY platform (Sequenom, Inc., San Diego, California), as described previously (16). There were 4 negative (water) controls per 384-well plate, and quality control parameters included genotype call rates greater than 95%, a combination of cases and controls on each plate, and inclusion of 20 duplicate samples per 384-well plate (≥5% of samples), with ≥98% concordance between duplicates and Hardy-Weinberg equilibrium \( P \) values greater than 0.05.

**Statistical analysis**

Predictive Analytics Software (PASW) Statistics, version 17.0.2 (SPSS, Inc., Chicago, Illinois), was used for all analyses unless otherwise specified. Genotype and allele frequencies were calculated for the patient and control groups. Comparisons of allele and genotype distributions and their associations with prostate cancer susceptibility and clinical data were performed under a codominant model and a linear model, using logistic regression analysis. Prostate cancers with tumor Gleason scores \( \geq 7 \) were classified as aggressive. All analyses were adjusted for age (as a continuous variable).

**RESULTS**

Table 1 shows the clinical and epidemiologic characteristics of the Australian prostate cancer cases and healthy male controls recruited for the study. The median age of cases at diagnosis was 63 years (range, 43–88 years), similar to the age at interview for controls (62 years; range, 43–75 years). Aggressive cases (Gleason score \( \geq 7 \)) represented 72% of cases with available Gleason scores, and 23% of cases had self-reported serum PSA concentrations greater than 10 ng/mL, comparable to other studies (17). Information on prostate tumor stage was still being collated at the time of this study.
PSA screening of low-risk or asymptomatic men is not routinely undertaken in Australia (18), and PSA levels were not measured for the control subjects. There were significantly more cases with a family history of prostate cancer than controls \((P < 0.0001)\), consistent with the literature (18).

Table 2 shows the allelic frequencies of the 5 SNPs (rs13385191, rs12653946, rs1983891, rs339331, and rs9600079) from our study, the Japanese GWAS (11), and HapMap data (International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/)) and their associations with prostate cancer in these Australian men. The minor allele frequencies for the analyzed SNPs were not markedly different between the European and Japanese samples, except for rs13385191, as indicated by FST scores from the HapMap data (Table 2). (The F_{ST} score was described by Ribas et al. as “a measure of the degree of population differentiation based on pairwise SNP frequency comparisons” (19, p. 254).) In the absence of knowledge regarding where a causative SNP might be located, we conducted a preliminary investigation to assess possible differences in linkage disequilibrium patterns between Japanese and European populations for a ±10-kilobase region around the 5 SNPs identified in the Japanese GWAS (see Web Table 1, which appears on the Journal’s Web site (http://aje.oxfordjournals.org/)). This extended analysis of F_{ST} values suggested, but did not prove, that heterogeneity may drive differences in linkage disequilibrium patterns between Japanese and European populations for rs13385191 and rs9600079 but is unlikely to do so for rs12653946, rs1983891, and rs339331.

We found rs12653946 at 5p15 to be significantly associated with risk of prostate cancer in Caucasian males (odds ratio = 1.20, 95% confidence interval: 1.07, 1.34; \(P = 0.002\)) (Table 2), with a per-allele effect size similar to that reported in Japanese men. The risk estimates did not differ markedly by Gleason score (<7 vs. ≥7) or by reported family history of prostate cancer (data not shown), similar to what was observed in the Japanese study (11).

None of the other 4 SNPs examined were found to be significantly associated with prostate cancer risk, although the directions of the estimated effects for rs339331 and rs9600079 were similar to those reported for the Japanese study (Table 2). This suggests that risk-associated loci identified in Japanese samples are worthy of prioritization for study in European men.

**DISCUSSION**

Our results indicate that the rs12653946 SNP at 5p15 is a prostate cancer risk locus that is not ethnicity-specific for an Asian population but is also associated with risk in a European population. It is notable that our study had 90% power to detect the risk estimate for rs12653946 reported in the Japanese GWAS (odds ratio = 1.2) at a \(P\) value of 0.001. Our study had lower power (44%–60%) to detect the risks
estimated for rs13385191, rs1983891, and rs339331 but sufficient power to detect the risk estimate reported for the rs9600079 SNP. This suggests that rs9600079 is unlikely to be associated with prostate cancer risk in Europeans, as is also supported by evidence for differences in linkage disequilibrium patterns. However, we cannot exclude that possibility for any of the other 3 SNPs. Another possible limitation is that our controls were not PSA-tested, and the existence of undetected cases in the control sample would have biased results towards the null. Thus, it would be interesting to validate results for all of the SNPs tested in additional cohorts with PSA-tested controls.

Interestingly, numerous SNPs at 5p15 have been associated with different cancers, including basal cell carcinoma, glioma, and lung, bladder, testicular, and prostate cancer (20). This “cancer nexus” region encompasses the telomerase reverse transcriptase (TERT) and cisplatin resistance-related protein CRR9p (CLPTM1L) genes, both of which are plausible candidates for a role in cancer. Two prostate cancer risk SNPs located at 5p15, rs401681 and rs2736098 (19), have also been reported to be associated with PSA levels \( P = 1.20 \times 10^{-10} \) and \( P = 2.84 \times 10^{-10} \) (respectively) (21). However, the TERT-CLPTM1L locus lies approximately 700 kilobases away from rs12653946, and our analysis of HapMap genotype data from Europeans showed no linkage disequilibrium \( r^2 < 0.01 \) between rs12653946 and the 2 previously reported 5p15 prostate cancer SNPs. Thus, we speculate that the rs12653946 SNP represents an independent locus associated with prostate cancer risk in European men as well as Japanese men.

Takata et al. (11) reported that the locus at 5p15 identified by means of rs12653946 represents a 20-kilobase block in Japanese samples and that it contains no known genes. Our in-depth analysis of the locus revealed that SNP rs12653946 falls within the intron of a gene predicted, with a high coding sequence identity of 10. Also, we provide evidence that a novel gene is likely to exist at this prostate cancer locus, although fine mapping and sequencing studies will be required to identify whether the likely causal SNP at this locus lies within the predicted gene tojy.

In summary, we provide evidence that the rs12653946 SNP at 5p15 is associated with risk of prostate cancer in both Japanese and European populations and may be incorporated into future risk prediction models for prostate cancer. We also provide evidence that a novel gene may exist at this locus, and suggest that our findings should provide impetus

---

**Table 2.** Associations Between 5 Novel Japanese Prostate Cancer-Associated Single Nucleotide Polymorphisms and Risk of Prostate Cancer in Australian Men, Queensland, Australia, 2004–2008

<table>
<thead>
<tr>
<th>Reference SNP</th>
<th>Location*</th>
<th>Region and Nearby Gene</th>
<th>SNP and Risk Allele</th>
<th>Frequency of Designated Minor Alleleb</th>
<th>HapMap Population*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JPT (Japanese)d</td>
<td>CEU (Australians)d</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JPT</td>
<td>CEU</td>
</tr>
<tr>
<td>rs13385191</td>
<td>20751746</td>
<td>2p24</td>
<td>A &gt; G</td>
<td>0.44 (0.39)</td>
<td>0.25 (0.21)</td>
</tr>
<tr>
<td>rs12653946</td>
<td>1948829</td>
<td>5p15 tojy</td>
<td>C &gt; T</td>
<td>0.44 (0.43)</td>
<td>0.41 (0.42)</td>
</tr>
<tr>
<td>rs1983891</td>
<td>41644405</td>
<td>6p21 FOXP4</td>
<td>C &gt; T</td>
<td>0.41 (0.37)</td>
<td>0.28 (0.27)</td>
</tr>
<tr>
<td>rs339331</td>
<td>11731674</td>
<td>6q22 GPRC6A/RFX6</td>
<td>T &gt; C</td>
<td>0.36 (0.37)</td>
<td>0.31 (0.36)</td>
</tr>
<tr>
<td>rs9600079</td>
<td>72626140</td>
<td>13q22</td>
<td>G &gt; T</td>
<td>0.38 (0.35)</td>
<td>0.43 (0.47)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; OR, odds ratio; SNP, single nucleotide polymorphism.

* P < 0.05 (after multiple testing).

* Chromosomal location based on coordinates from the US National Center for Biotechnology Information human genome build 36.

* Numbers in parentheses represent the HapMap frequency.

* CEU, Utah residents with ancestry from northern and western Europe; JPT, Japanese persons from Tokyo. For details on HapMap population designations, refer to the HapMap guidelines (http://hapmap.ncbi.nlm.nih.gov/citinghapmap.html).

* Risk estimates obtained from a genome-wide association study combining 4,584 cases and 8,801 controls by Takata et al. (11).

* Age-adjusted allelic risk estimates obtained in the current study using a linear model.

* A measure of the degree of population differentiation based on pairwise SNP frequency comparisons (19). Genetic differentiation between the CEU and JPT populations from HapMap data (a value greater than 0.05 denotes a significant difference between the 2 samples).

* OR was inverted for comparison with the results of Takata et al. (11), from an observed OR of 0.91 (95% CI: 0.80, 1.03), using the common C allele as the reference.

* Gene desert region as reported by Takata et al. (11).
for future experimental studies confirming the nature of this gene and whether the putative protein is translated.

ACKNOWLEDGMENTS

Author affiliations: Australian Prostate Cancer Research Centre—Queensland, Queensland University of Technology, Brisbane, Queensland, Australia (Jyotsna Batra, Judith A. Clements); Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia (Jyotsna Batra, Judith A. Clements, Australian Prostate Cancer BioResource); Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia (Suzanne Chambers); Viertel Centre for Research in Cancer Control, Cancer Council Queensland, Brisbane, Queensland, Australia (Suzanne Chambers, Joanne Aitken); University of Queensland Centre for Clinical Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia (Robert A. Gardiner); and Brisbane Private Hospital, Brisbane, Queensland, Australia (John Yaxley).

This work was supported by the Australian National Health and Medical Research Council (NHMRC) (grants 390130, 290456, and 614296) and the Prostate Cancer Foundation of Australia (grant PG7); an NHMRC Early Career Fellowship and an Institute of Health and Biomedical Innovation Postdoctoral Fellowship (J. Batra); an NHMRC Senior Research Fellowship (A. B. Spurdele); an NHMRC Principal Research Fellowship (J. A. Clements); and an NHMRC Career Development Award and the Prostate Cancer Research Program of Cancer Council Queensland (S. K. Chambers).

The authors thank all staff members of the Australian Red Cross Blood Service, Cancer Council Queensland, the Urological Society of Australia and New Zealand, and numerous Australian Prostate Cancer BioResource researchers, including Drs. P. Heathcote, G. Wood, G. Malone, and D. Nicol for participant recruitment and Dr. M. A. Kedda, P. V. Bergh, P. Saunders, S. Srinivasan, K. Alexander, and L. Marquard for other contributions.

Drs. J. Batra and A. B. Spurdele had full access to all of the data in this study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of interest: none declared.

REFERENCES