Chromosome 17q12 Variants Contribute to Risk of Early-Onset Prostate Cancer

Albert M. Levin,1 Mitchell J. Machiela,2 Kimberly A. Zuhlke,2 Anna M. Ray,2 Kathleen A. Cooney,2,3 and Julie A. Douglas1

Departments of1 Human Genetics, 2Internal Medicine, and 3Urology, University of Michigan Medical School, Ann Arbor, Michigan

Abstract
In a recent genome-wide association study by Gudmundsson and colleagues, two prostate cancer susceptibility loci were identified on chromosome 17q. The first locus, at 17q12, was distinguished by two intronic single-nucleotide polymorphisms (SNPs) in the TCF2 gene (rs4430796 and rs7501939). The second locus was in a gene-poor region of 17q24, where the strongest evidence of association was for SNP rs1859962. To determine if these loci were also associated with hereditary prostate cancer, we genotyped them in a family-based association sample of 403 non-Hispanic white families, including 1,015 men with and without prostate cancer. SNPs rs4430796 and rs7501939, which were in strong linkage disequilibrium ($r^2 = 0.68$), showed the strongest evidence of prostate cancer association. Using a family-based association test, the A allele of SNP rs4430796 was overtransmitted to affected men ($P = 0.006$), with an odds ratio of 1.40 (95% confidence interval, 1.09–1.81) under an additive genetic model. Notably, rs4430796 was significantly associated with prostate cancer among men diagnosed at an early (<50 years) but not later age ($P = 0.006$ versus $P = 0.118$). Our results confirm the prostate cancer association with SNPs on chromosome 17q12 initially reported by Gudmundsson and colleagues. In addition, our results suggest that the increased risk associated with these SNPs is approximately doubled in individuals predisposed to develop early-onset disease. Importantly, these SNPs do not account for a significant portion of our prior prostate cancer linkage evidence on chromosome 17. Thus, there likely exist one or more additional independent prostate cancer susceptibility loci in this region. [Cancer Res 2008;68(16):6492–5]

Introduction
The identification of prostate cancer susceptibility loci has traditionally been challenging. Linkage analyses have identified several regions of interest, but subsequent studies have not always been consistent (1, 2). Likewise, the candidate gene approach has yielded results that have been equally difficult to validate (3). In contrast, recent genome-wide association studies of prostate cancer have identified genetic variants that have now been validated across a number of studies. In particular, multiple studies have implicated chromosome 8q24 as a region harboring several single-nucleotide polymorphisms (SNPs) that independently predict prostate cancer risk (4–6). As follow-up to one of these genome-wide association studies, Gudmundsson and colleagues (7) recently identified two regions on chromosome 17q as harboring additional independent prostate cancer susceptibility loci. Specifically, two intronic SNPs in the TCF2 gene (rs4430796 and rs7501939) at 17q12 and a third SNP (rs1859962) at 17q24 were associated with the risk of sporadic prostate cancer.

To determine if these three 17q SNPs also predict prostate cancer risk among individuals who may have a particularly high genetic susceptibility to the disease, we genotyped them in our family-based association sample of early-onset and familial prostate cancer. Given that our strongest signal for prostate cancer linkage in a previous genome-wide scan was on chromosome 17q21 (8), we also genotyped these SNPs in our genome-wide scan linkage families to evaluate whether these SNPs could account for a portion of our linkage signal on chromosome 17q.

Materials and Methods

Study subjects. The details of the University of Michigan Prostate Cancer Genetics Program have been described elsewhere (9). Briefly, enrollment into the Prostate Cancer Genetics Program is restricted to (a) families with two or more living members with prostate cancer in a first- or second-degree relationship or (b) men diagnosed with prostate cancer at ≤55 y of age without a family history of the disease. For the present study, 421 families were identified in which DNA was available from at least one pair of brothers discordant for prostate cancer, the majority of whom self-identified as non-Hispanic white ($n = 403$). The remaining 18 families were African American ($n = 16$) and Asian ($n = 2$). Results below were restricted to non-Hispanic white families because the number of African American and Asian families was too small to make meaningful inferences about prostate cancer risk in these minority groups.

The majority of Prostate Cancer Genetics Program families were recruited directly from the University of Michigan Comprehensive Cancer Center. Prostate cancer diagnoses were confirmed by review of pathology reports or medical records, and age at diagnosis was calculated from the date of the first positive biopsy. Cases were classified as clinically aggressive if they met at least one of the following criteria: (a) pathologic Gleason sum $>7$; (b) pathologic stage T3a (pT3a) tumor (indicating seminal vesicle involvement) or pT4 or N1 (positive regional lymph nodes); (c) pathologic Gleason sum of 7 and a positive margin; or (d) preoperative serum prostate-specific antigen (PSA) value $>15$ ng/mL, a biopsy Gleason score $>7$, or a serum PSA level $>10$ ng/mL and a biopsy Gleason score $>6$. Using data from D’Amico and colleagues (10), these criteria were developed by the Southwest Oncology Group (protocol 9921) to identify men at intermediate to high risk of clinical recurrence after primary therapy. Disease status of the unaffected brothers was confirmed through serum PSA testing whenever possible. The Institutional Review Board at the University of Michigan Medical School approved all aspects of the protocol, and all participants gave written informed consent including permission to release their medical records.

Genotyping assays. We genotyped three SNPs (rs4430796, rs7501939, and rs1859962) using TaqMan SNP assays (Applied Biosystems), and we used the ABI Prism 7900HT Sequence Detection System and SDS version 2.1 software (Applied Biosystems) to distinguish alleles as previously described.
Statistical methods. Observed genotype distributions were tested for departure from Hardy-Weinberg equilibrium in a subset of the oldest, unrelated, unaffected men from each family. Haplotype frequencies were estimated using the expectation-maximization algorithm and were used to calculate the linkage disequilibrium measure \( r^2 \) between each pair of markers. For association testing, we used the family-based association method (implemented in the FBAT software, version 1.7.3; ref. 11) to test for association between single SNPs and prostate cancer. To maximize power, we analyzed the combined set of affected and unaffected men using the offset option to test the null hypothesis of no association and no linkage. To account for the possible misclassification of unaffected men, we analyzed affected men only using the empirical variance estimate to test the null hypothesis of no association in the presence of linkage. In parallel, we used conditional logistic regression, coupled with a robust variance estimate that incorporates familial correlations (12), to generate odds ratios (OR) and 95% confidence intervals (95% CI). For both FBATs and conditional logistic regression analyses, we carried out adding additive, dominant, and recessive genetic models. In addition, we also examined a genotype (two degrees of freedom) model for conditional logistic regression and affected-only FBAT analyses. Because SNPs rs4430796 and rs7501939 were in strong linkage disequilibrium \( (r^2 = 0.68) \), the association between this two-SNP haplotype and prostate cancer was tested using the haplotype FBAT (HBAT) method (13).

We genotyped all three SNPs in 154 of our original 157 non-Hispanic white families from our genome-wide linkage scan (8) to determine if they could explain our prior linkage evidence on chromosome 17q (14). These 154 families included 411 affected and 72 unaffected men for whom we had sufficient DNA. We then used the genotype-identity-by-decent sharing test of Li and colleagues (15) and implemented in version 0.3 of their software to determine if these SNPs could explain the linkage signal. This method tests for a positive correlation between family-specific weights (based on the genotypes of affected family members and a specified genetic model) and family-specific identity-by-decent sharing, as represented by the nonparametric linkage score. The family-specific nonparametric linkage scores were estimated using the "pairs" scoring statistic, the exponential, the quartic model, and equal weights for each family (14).

All statistical tests were two-sided, and \( P < 0.05 \) was considered statistically significant. Conditional logistic regression was conducted using version 8.2 of the Statistical Analysis System programming language (SAS institute). All analyses were done with and without adjustment for age. Because the results were unaffected by adjusting for age, unadjusted results are presented below. All remaining analyses (except where noted above) were conducted using the R language (version 2.6.0).

Results

For this analysis, 542 affected and 473 unaffected men were available from 403 non-Hispanic white families, with at least one discordant sibling pair per family. Of these 403 families, 386 (96%) contributed a single sibship of men, and the remaining 17 contributed multiple sibships (e.g., sibships related as first cousins). A majority of the sibships (72%) had a single discordant sibling pair. In total, the sample consisted of 624 discordant sibling pairs from 421 sibships. With regard to our enrollment criteria, 310 (77%) families had two or more living members with prostate cancer in a first- or second-degree relationship, and 91 families (23%) included men diagnosed with prostate cancer at \( \leq 55 \) years of age without a family history of the disease. Two additional families, each with a single discordant sibling pair, were also included.

The clinical characteristics of men with prostate cancer are summarized in Table 1. The median age at diagnosis was 54 years, with 116 (21%) diagnosed before 50 years and 162 (30%) diagnosed with clinically aggressive disease. Approximately 80% of unaffected men reported their most recent PSA test result and/or had a medical record confirmation of their most recent value, and nearly 95% of these men had documented PSA levels <40 ng/mL.

In the sample of unrelated, unaffected men, all SNP genotype distributions were consistent with Hardy-Weinberg equilibrium (\( P > 0.05 \)). The base-pair position and major allele frequency for each SNP are presented in Table 2. SNPs rs7501939 and rs4430796, located on 17q12 within introns 1 and 2 of the TCF2 gene, respectively, exhibited strong linkage disequilibrium \( (r^2 = 0.68) \). In contrast, rs1859962, located \( \sim 33 \) Mb downstream from TCF2, was in weak linkage disequilibrium with both of the TCF2 SNPs (maximum pairwise \( r^2 = 0.002 \)). Ignoring family structure, we observed significant allele frequency differences between affected and unaffected men for SNPs rs4430796 \( (P = 0.01) \) but not for rs1859962 \( (P = 0.13) \).

Tables 2 and 3 summarize association results (under an additive genetic model) for all three SNPs for FBAT and conditional logistic regression analyses, respectively. For the FBAT results that follow, findings from the combined sample of affected and unaffected men are reported. The two TCF2 SNPs were significantly associated with prostate cancer in our sample. The \( A \) allele of SNP rs4430796 was overtransmitted to affected men (FBAT \( P = 0.006 \)), with an OR of 1.40 (95% CI, 1.09–1.81; \( P = 0.01 \)). As expected (given the strong linkage disequilibrium between rs4430796 and rs7501939), results for SNP rs7501939 were similar. Haplotype association analyses did not provide additional insight beyond single SNP analyses. For example, the two-SNP haplotype containing both risk alleles at rs4430796 and rs7501939 was significantly overtransmitted to affected men (HBAT \( P = 0.016 \)), and the haplotype containing neither risk allele was significantly undertransmitted to affected men (HBAT \( P = 0.008 \)). Whereas association results for SNP rs1859962 at 17q24 were not statistically significant, there was suggestive evidence that the \( G \) allele was overtransmitted to affected men (FBAT \( P = 0.06 \)), with an OR of 1.21 (95% CI, 0.94–1.56; \( P = 0.13 \)).
Stratified analyses revealed that men diagnosed with prostate cancer at an early age contributed disproportionately to the overall results for the TCF2 SNPs. For example, the A allele of rs4430796 was significantly overtransmitted to men diagnosed before age 50 years (FBAT \(P = 0.006\)), with an OR of 1.92 (95% CI, 1.15–3.18). In addition, homozygous carriers of the A allele had a 3.70-fold (95% CI, 1.33–10.29) increased risk of prostate cancer at an early age (<50 years) relative to noncarriers. In contrast, the A allele of rs430796 was not significantly overtransmitted to men diagnosed at or after the age of 50 years (FBAT \(P = 0.118\)), with an OR of 1.25 (95% CI, 0.93–1.67). Similar results were also observed for rs1859962. We found no significant evidence of an association between any of the SNPs and prostate cancer when the analyses were stratified by clinically aggressive disease.

To determine whether any of the three SNPs accounted for our prostate cancer linkage to chromosome 17q (14), we also genotyped them in 154 of our original genome-wide scan linkage families, all of non-Hispanic white descent. Based on the deCODE genetic map (16), our estimated linkage peak resided at \(81 \text{ to } 82\) cM, and using base-pair locations and interpolating between flanking microsatellite markers, the TCF2 SNPs and rs1859962 were placed at 66.19 and 104.00 cM, respectively. Using the genotype-identity-by-decent sharing test, we found no evidence that the risk allele at any of the SNPs was correlated with our linkage evidence on 17q (additive model \(P\) values of 0.69, 0.53, and 0.44 for rs4430796, rs7501939, and rs1859962, respectively). Similarly, there was no evidence that these SNPs accounted for linkage in the subset of families with an average age of prostate cancer diagnosis of <65 years or families with four or more confirmed cases of prostate cancer (data not shown).

### Discussion

In summary, we have confirmed that the prostate cancer–associated SNPs on chromosome 17q originally identified by Gudmundsson and colleagues are also associated with early-onset and familial prostate cancer. In our sample, the two significantly associated SNPs, rs4430796 and rs7501939, had the strongest evidence of association in the subset of families in which men were diagnosed with prostate cancer at an early age. To our knowledge, we are the first group to report an association between the 17q SNPs and hereditary prostate cancer and the first to report a significant association between early-onset disease and the TCF2 SNPs. Notably, we estimate that men with two risk alleles at rs4430796 are ~4 times more likely to develop early-onset prostate cancer than those with no risk alleles. Whereas this age-of-onset effect was suggested by Gudmundsson and colleagues (7), it was not statistically significant in their sample, which was primarily composed of men who developed prostate cancer at a comparatively later age (mean age at diagnosis, 70.8 years). In fact, the average age at prostate cancer diagnosis in our sample (56 years) was considerably lower than even the national average (67 years).

Notably, all three 17q SNPs have also been investigated in the Cancer Genetic Markers of Susceptibility genome-wide prostate cancer association studies (CGEMS), and the Cancer Genetic Markers of Susceptibility to prostate cancer (CGEMS) (8). In contrast to the results of our sample, all three SNPs showed no evidence of association with prostate cancer in the CGEMS sample (9).

### Table 2. Major allele frequencies and FBAT results (\(n = 403\) families)

<table>
<thead>
<tr>
<th>dbSNP ID</th>
<th>Base pair position</th>
<th>Gene</th>
<th>Location</th>
<th>Alleles major-minor</th>
<th>Major allele frequency</th>
<th>FBAT*</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4430796</td>
<td>33,172,153</td>
<td>TCF2</td>
<td>Intron 2</td>
<td>A&gt;G</td>
<td>0.59</td>
<td>210</td>
</tr>
<tr>
<td>rs7501939</td>
<td>33,175,269</td>
<td>TCF2</td>
<td>Intron 1</td>
<td>C&gt;T</td>
<td>0.67</td>
<td>198</td>
</tr>
<tr>
<td>rs1859962</td>
<td>66,620,384</td>
<td></td>
<td>Intergenic</td>
<td>G&gt;T</td>
<td>0.55</td>
<td>208</td>
</tr>
</tbody>
</table>

NOTE: rs4430796 and rs7501939 are in strong linkage disequilibrium (\(r^2 = 0.68\)).

*Based on the combined sample of men with and without prostate cancer and an additive genetic model for the major allele.

†Number of informative families.

### Table 3. Comparison of ORs for 17q SNPs tested in the University of Michigan, Gudmundsson et al., and the Cancer Genetic Markers of Susceptibility studies

<table>
<thead>
<tr>
<th>dbSNP ID</th>
<th>Risk allele</th>
<th>UM* OR (95% CI)</th>
<th>UM* (P)</th>
<th>Gudmundsson et al. † OR (95% CI)</th>
<th>Gudmundsson et al. † (P)</th>
<th>CGEMS ‡ OR (95% CI)</th>
<th>CGEMS ‡ (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4430796</td>
<td>A</td>
<td>1.40 (1.09–1.81)</td>
<td>0.009</td>
<td>1.20 (1.11–1.31)</td>
<td>&lt;0.001</td>
<td>1.17 (1.04–1.31)</td>
<td>0.009</td>
</tr>
<tr>
<td>rs7501939</td>
<td>C</td>
<td>1.44 (1.10–1.89)</td>
<td>0.008</td>
<td>1.17 (1.06–1.27)</td>
<td>&lt;0.001</td>
<td>1.19 (1.06–1.34)</td>
<td>0.004</td>
</tr>
<tr>
<td>rs1859962</td>
<td>G</td>
<td>1.21 (0.94–1.56)</td>
<td>0.132</td>
<td>1.16 (1.07–1.26)</td>
<td>&lt;0.001</td>
<td>1.19 (1.06–1.34)</td>
<td>0.003</td>
</tr>
</tbody>
</table>

NOTE: ORs were computed under an additive genetic model.

Abbreviations: UM, University of Michigan; CGEMS, Cancer Genetic Markers of Susceptibility.

*Before conducting the conditional logistic regression analyses, we excluded 38 men (from 10 families) who were not brothers of the index case, resulting in a reduced sample size of 977 men and 604 discordant sibling pairs.

†Logistic regression as reported by Gudmundsson et al. (7), all samples combined (1,501 cases and 11,289 controls).

‡Logistic regression, unadjusted for covariates (1,155 cases and 1,106 controls).
cancer association study. Similar to Gudmundsson and colleagues, the Cancer Genetic Markers of Susceptibility study enrolled Caucasian men who were diagnosed at a later age (≥55 years). These sample similarities likely explain the comparable ORs for all three SNPs in these two studies (Table 3). Whereas our ORs for the two TCF2 SNPs are not significantly different from these estimates, the increased magnitude of our estimates likely reflects the enhanced effect of these SNPs in our early-onset cases. In fact, in men diagnosed on or after the age of 50 years, our OR for SNP rs4430796 was 1.25 (95% CI 0.93–1.67), comparable to the other two studies. Together, these findings support a role for the TCF2 SNPs in both early-onset and sporadic prostate cancer.

Although rs4430796 and rs7501939 reside within the TCF2 gene, both are intronic, with no obvious effect on the TCF2 protein. It is possible that these SNPs may be in linkage disequilibrium with one or more genetic variants that directly increase prostate cancer risk. To explore this possibility, we used the Caucasian CEU sample from the International HapMap project (build 35) and computed the pairwise r² measure between each chromosome 17 SNP in HapMap and rs4430796 and rs7501939. Based on a threshold of r² > 0.5, four SNPs (rs2005705, rs757210, rs4239217, and rs7465696) were in strong linkage disequilibrium with rs4430796 and rs7501939, all of which were also located within introns in TCF2 and separated by <5 kb. However, it is difficult to resolve which of these SNP(s), if any, directly influences prostate cancer risk because our knowledge of genetic variation in the region is currently incomplete (i.e., many other untyped SNPs exist, including ones that may be in linkage disequilibrium with our associated SNPs).

The TCF2 finding is not only evidence of a locus on 17q predisposing to prostate cancer. After following up our strongest genome-wide signal on chromosome 17q21-22 (8), we recently identified two SNPs within the BRCA1 gene that were independently associated with early-onset (rs1799950) and hereditary (rs3737559) prostate cancer, with rs1799950 explaining some, but not all, of our original linkage signal. In contrast, the TCF2 SNPs, which are located ~15 cM (or ~15 Mb) upstream of this signal (14), did not explain a significant portion of linkage in our genome-wide scan families. Still, by virtue of using a family-based association test, we have shown that the TCF2 SNPs are both linked to and associated with prostate cancer in our sample of discordant sibling pair families, only 60 of which overlap with our genome-wide scan families. Notably, the TCF2 and BRCA1 genes are located ~5 Mb apart and are not in strong linkage disequilibrium with one another (i.e., maximum pairwise r² of 0.006 between the associated SNPs). Together, these results suggest that there likely exist one or more additional independent prostate cancer susceptibility loci in this region.

In conclusion, results from at least five studies (7, 17, 18), including the CGEM study (19), now indicate that genetic variation on chromosome 17q is associated with sporadic prostate cancer. Data from our family-based study, however, suggest that these associations also extend to hereditary prostate cancer in general and early-onset prostate cancer in particular. Moreover, results from our stratified analyses indicate that the genetic risk conferred by either SNP on 17q21 may be substantially increased, nearly 2-fold higher, in men predisposed to develop early-onset prostate cancer. Such findings hint at the potential for early genetic screening to identify a subset of men who are at greater risk of developing prostate cancer, even in the absence of a family history of disease.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

Received 1/30/2008; revised 5/21/2008; accepted 5/22/2008.

**Grant support:** The University of Michigan Prostate Cancer Genetics Project is made possible by funds from the National Cancer Institute Sponsored Program of Research Excellence in Prostate Cancer P50 CA69568 (J.A. Douglas and K.A. Cooney) and B01 CA79596 (K.A. Cooney).

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.

We thank all Prostate Cancer Genetics Program men and their families who generously volunteered their time to participate in our study. We also thank Joe Washburn and the University of Michigan Comprehensive Cancer Center DNA Microarray Facility (funded in part by NIH support grant P30 CA46592) for assistance with the genotyping assays.

**References**
