Research Article

Refining the Prostate Cancer Genetic Association within the JAZF1 Gene on Chromosome 7p15.2

Ludmila Prokunina-Olsson1,2, Yi-Ping Fu1,2, Wei Tang1,2, Kevin B. Jacobs3,4, Richard B. Hayes2,5, Peter Kraft6, Sonja I. Berndt2, Sholom Wacholder2, Kai Yu2, Amy Hutchinson2,3, Heather Spencer Feigelson5, Michael J. Thun9, W. Ryan Diver3, Demetrius Albanes2, Jarmo Virtamo10, Stephanie Weinstein2, Fredrick R. Schumacher6,7, Geraldine Cancel-Tassin11, Olivier Cussenot11, Antoine Valeri11, Gerald L. Andriole12, E. David Crawford13, Christopher A. Haiman14, Brian E. Henderson14, Laurence Kolonel15, Loic Le Marchand15, Afshan Siddiqu16, Elio Riboli16, Ruth Travis17, Rudolf Kaaks18, William B. Isaacs19, Sarah D. Isaacs19, Henrik Grönberg20, Fredrik Wiklund21, Jianfeng Xu21, Lars J. Vatten22, Kristian Hveem22, Merethe Kumle23, Margaret Tucker2, Robert N. Hoover2, Joseph F. Fraumeni, Jr.2, David J. Hunter2,6,7,8, Gilles Thomas2, Nilanjan Chatterjee2, Stephen J. Chanock1,2,3 and Meredith Yeager2,3

Abstract

Background: Genome-wide association studies have identified multiple genetic variants associated with susceptibility to prostate cancer (PrCa). In the two-stage Cancer Genetic Markers of Susceptibility prostate cancer scan, a single-nucleotide polymorphism (SNP), rs10486567, located within intron 2 of JAZF1 gene on chromosome 7p15.2, showed a promising association with PrCa overall \( (P = 2.14 \times 10^{-7}) \), with a suggestion of stronger association with aggressive disease \( (P = 1.2 \times 10^{-7}) \).

Methods: In the third stage of genome-wide association studies, we genotyped 106 JAZF1 SNPs in 10,286 PrCa cases and 9,135 controls of European ancestry.

Results: The strongest association was observed with the initial marker rs10486567, which now achieves genome-wide significance \( (P = 7.79 \times 10^{-11}; \text{OR}_{\text{HF}} = 1.19 \text{ (95\% confidence interval, 1.12-1.27); OR}_{\text{OM}} = 1.37 \text{ (95\% confidence interval, 1.20-1.56)}) \). We did not confirm a previous suggestion of a stronger association of rs10486567 with aggressive disease \( (P = 1.60 \times 10^{-8} \text{ for aggressive cancer; } n = 4,597; P = 3.25 \times 10^{-8} \text{ for non-aggressive cancer; } n = 4,514) \). Based on a multilocus model with adjustment for rs10486567, no additional independent signals were observed at chromosome 7p15.2. There was no association between PrCa risk and SNPs in JAZF1 previously associated with height \( (rs849140; P = 0.587) \), body stature \( (rs849141; P = 0.171) \), and risk of type 2 diabetes and systemic lupus erythematosus \( (rs864745, \text{tagged by } rs849141; P = 0.657) \).

Conclusion: rs10486567 remains the most significant marker for PrCa risk within JAZF1 in individuals of European ancestry.

Impact: Future studies should identify all variants in high linkage disequilibrium with rs10486567 and evaluate their functional significance for PrCa.

Cancer Epidemiol Biomarkers Prev; 19(5); OF1–7. ©2010 AACR.
Prostate cancer (PrCa) is the most common noncutaneous cancer in the developed world and the second leading cause of cancer death in men (1, 2). The disease is highly treatable when detected early, with an encouraging 5-year survival rate (3). The established risk factors for PrCa are age, ethnicity, and family history (3). Heritable factors have been estimated to explain 42% (29-50%) of PrCa risk in individuals of European ancestry (4). PrCa diagnostics based on the blood level of prostate-specific antigen (PSA) can result in 23% to 45% of overdiagnosis, that is, in detection of disease with mild or insignificant clinical manifestations that do not require treatment (5). Several genome-wide association studies (GWAS) have been performed to determine genetic factors that can identify individuals with increased risk of PrCa. More than 25 genomic regions that harbor genetic risk factors for PrCa have been identified to date (6-10).

The discovery component of the Cancer Genetic Markers of Susceptibility project (CGEMS) has reported a two-stage GWAS for PrCa (6). In the first stage, 1,172 individuals with PrCa and 1,157 controls of European ancestry were genotyped for 527,869 single-nucleotide polymorphisms (SNP). In the second stage, 26,958 SNPs, selected based on their association in stage I (P < 0.068), were genotyped in a total of 4,020 cases and 4,028 controls (6). The second stage of CGEMS identified SNPs on chromosome 17q12 (12). As a part of an expansive analysis, the 8q24 region (10), chromosome 11q13 (11), and the 11q13 region to be significantly associated with PrCa (6). An observed association for markers in the COOH-terminal binding protein 2 (CTBP2) gene, as well as in the juxtaposed with another zinc finger protein 1 (JAZF1), did not reach the threshold of genome-wide significance (P = 1.7 × 10^{-7} for CTBP2 and P = 2.14 × 10^{-6} for JAZF1). For both genes, there was a suggestion that the signal was more strongly associated with the risk of aggressive cancer, defined as Gleason score >7 or disease stage III (P = 2.7 × 10^{-8} for CTBP2 and P = 1.2 × 10^{-7} for JAZF1; ref. 6).

The goals of the third stage of CGEMS GWAS have expanded to include testing for the presence of independent signals within established candidate regions and conducting a first-order mapping of regions identified in CGEMS stage II. A similar approach has resulted in identification of independent signals within the 8q24 region (10), chromosome 11q13 (11), and HNF1B gene on chromosome 17q12 (12). As a part of stage III, we genotyped 106 JAZF1 tag SNPs in 10,286 PrCa patients and 9,135 controls of European ancestry. We aimed to confirm the suggestive association observed in stage II in a larger data set and to test for presence of stronger or independent association signals within JAZF1. We tested for association of rs10486567 with age of cancer diagnosis, family history, and aggressiveness of disease. Additionally, we tested whether the variants of JAZF1 associated with other traits were also associated with risk of PrCa.

Subjects. Subjects were PrCa patients and controls drawn from 10 studies conducted in the United States and Europe: Prostate, Lung, Colon and Ovarian (PLCO; ref. 13); American Cancer Society Cancer Prevention Study II Nutrition Cohort (CPS-II; ref. 6); Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC; ref. 6); CeRePP French Prostate Case-Control Study (CeRePP; ref. 6); Health Professionals Follow-up Study (HPFS; ref. 6); The European Prospective Investigation into Cancer and Nutrition (EPIC; ref. 14); Cohort of Norway (CONOR; ref. 15); The Multiethnic Cohort Study (MEC; ref. 16); Johns Hopkins University (JHU; ref. 17); and The Cancer Prostate in Sweden study (CAPS; ref. 18). The counts and characteristics of each study are presented in Supplementary Table S1. The two main PrCa subtypes were defined as nonaggressive (Gleason score <7 and stage <III) and aggressive (Gleason score ≥7 or stage ≥III) cancer at the time of diagnosis.

SNP selection. The design, SNP selection, and analysis of the study were described in detail in Supplementary Material of Yeager et al. (10). In brief, 7,034 SNPs were selected based on the previous GWAS results (6), and 115 of these SNPs were from the JAZF1 gene. The JAZF1 targeted region was selected based on a 0.2-cM HapMap recombination map centered on rs10486567 (chr7: 27290918-27826082). A preliminary set of 7 tags was selected using rs10486567 as an obligate-include and tagging the region at D’ > 0.6 based on genotypes from HapMap CEU (European ancestry; ref. 19). A final set of 127 tags was chosen by using rs10486567 as an obligate-include and by tagging the preliminary list of tags at r^2 > 0.8 in CEU (European), YRI (African), and CHB/JPT (Asian) HapMap samples (see Supplementary Table S2A for list of all SNPs). All 7,034 SNPs selected for stage III were genotyped using a custom Illumina iSelect assay chip in 9,135 controls and 10,286 PrCa cases. Of 7,034 selected SNPs, 6,313 SNPs passed the manufacturer’s QC and provided >90% genotype calls. Of 115 JAZF1 SNPs, 9 SNPs failed to provide genotype data or resulted in low genotype call rates (<90%) and were excluded from the analysis. Genotype quality control, assessment of call rates, assessment of unique subjects, analysis of duplicate DNA samples, fitness for Hardy-Weinberg proportion in control DNA, and subject exclusions are described in detail in Supplementary Methods of Yeager et al. (10).
African, and Asian origins, respectively. The origin of the study samples was left unspecified. A total of 372 subjects (1.8%) were estimated to have less than 80% European ancestry and were excluded from analysis. All individuals that had greater than 80% European ancestry were retained for the study, regardless of their self-reported origin.

Principal components analysis. Principal components analysis was done using the same 1,399 SNPs included for population stratification. These results were based on the remaining subjects after removal of individuals with admixed ancestry as described above. A Wilcoxon rank test was done to check correlations with the case/control status for the top five eigenvectors. PC1 in JHU and top three PCs in CAPS studies showed significant differences between cases and controls. These principal components were used as covariates for association studies in JHU and CAPS sample sets.

Statistical software. Tag SNP selection was done using the GLU software package (20) using the HapMap CEU, JPT + CHB, and YRI data. Models were adjusted for study and significant principal components per study. Population stratification analysis was done with STRUCTURE software (21). Principal component analysis was done with EIGENSTRAT software (22). $\chi^2$ tests, Student's $t$ tests, and logistic regression were used to compare the basic characteristics between cases and controls using SAS/STAT system (SAS Institute, Inc.). For single-marker case-control analyses, logistic regression under an additive genetic model was done for each SNP adjusting for study site, and significant principle components using PLINK (23). A conditional analysis was also done under additive model to account for the effect of SNP rs10486567 (0, 1, or 2 risk alleles). The association and linkage disequilibrium (LD) plots were generated with snp.plotter, version 0.2 (24).

Results

We performed the third stage of CGEMS GWAS in 10,286 PrCa patients and 9,135 controls of European ancestry drawn from 10 studies conducted in Europe and the United States (ref. 10; also described in Materials and Methods and Supplementary Table S1). As part of this effort, we genotyped 106 tag JAZF1 SNPs chosen on the basis of a tiered tagging strategy, the details of which have been published separately (10). First, the target region was bounded by the 0.2-cM HapMap recombination map (chr7: 27290918-27826082). A preliminary set of 7 tags was selected using rs10486567 as an
obligate-include and tagging the region at $D^2 > 0.6$ based on genotypes from HapMap CEU (European ancestry; ref. 19). A final set of 127 tags was chosen by using rs10486567 as an obligate-include and by tagging the preliminary selected 7 tags at $r^2 > 0.8$ in CEU (European), YRI (African), and CHB/JPT (Asian) HapMap samples (see Supplementary Table S2A for list of all SNPs). Twenty-one of these selected SNPs failed design or showed poor performance. The successfully genotyped 106 SNPs covered a region of 275 kb, starting from 25 kb upstream of JAZF1 and well into intron 2 (Fig. 1).

Based on the genotype association test adjusted for study and significant principal components (see Supplementary Methods), rs10486567 was stronger for nonaggressive PrCa (Gleason score <7 or stage <III; $P = 1.60 \times 10^{-4}$ for aggressive cancer, $n = 4,597$ and $P = 3.25 \times 10^{-8}$ for nonaggressive cancer, $n = 4,514$; Table 2 and Supplementary Table S2B and C). Our sample set was enriched for individuals older than 60 years and without family history of PrCa, but the odds ratios (OR) for association of rs1086567 with PrCa were comparable in each of the subgroups (Table 3).

To test for presence of additional independent association signals within JAZF1, we performed a genotype association test adjusting for rs10486567. Only a weak association was observed for SNP rs3919460 ($P = 0.0053$; Supplementary Table S2A). We also examined whether the variants of JAZF1 associated with other traits would affect the genetic susceptibility to PrCa. In our samples, rs4205140, previously associated with height ($P = 5.3 \times 10^{-8}$; ref. 25), was not associated with PrCa ($P = 0.587$; rs4205141, previously reported to be associated with height and body stature ($P = 3.26 \times 10^{-11}$; ref. 26), was well tagged by rs4205136 ($r^2 = 1.0$ with rs4205141 in CEU HapMap), but was not associated with PrCa ($P = 0.171$); the type 2 diabetes (T2D)-associated SNP rs2943220 ($P = 1.4 \times 10^{-15}$; ref. 27) was tagged by rs4205136 ($r^2 = 1.0$ with rs2943220 in CEU HapMap), which was also associated with systemic lupus erythematosus (SLE; $P = 1.54 \times 10^{-10}$; ref. 28), but showed no association with PrCa ($P = 0.657$, Table 4).

### Discussion

Our study clearly confirms that common SNPs in JAZF1 are associated with risk for PrCa overall and establishes

<table>
<thead>
<tr>
<th>Trait</th>
<th>rs10486567 genotypes (no. of cases)</th>
<th>$P^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA, $n = 451$</td>
<td>AG, $n = 3423$</td>
</tr>
<tr>
<td>Age at diagnosis, y</td>
<td>66.56 ± 7.58</td>
<td>66.36 ± 7.50</td>
</tr>
<tr>
<td>Positive family history, %</td>
<td>17.15</td>
<td>17.18</td>
</tr>
<tr>
<td>Aggressive disease, %</td>
<td>52.1</td>
<td>51.32</td>
</tr>
</tbody>
</table>

* $P$ values are for $\chi^2$ test or ANOVA.
PrCa (\(=7.79 \times 10^{-5}\)) reported as showing a promising association (European ancestry. We show that rs10486567, previously this candidate gene for PrCa susceptibility in individuals of Europeans (0.73) and Africans (0.68), whereas it is rare
generation of their possible functional effects for PrCa.
rs10486567 to determine the optimal markers for investi-
gation all common and rare variants in strong LD with
provided by the 1000 Genomes project (29) will help to cat-
tor or miRNA binding sites. Further deep resequencing
to affect mRNA expression, splicing, or transcription fac-
ment did not reveal any independent
signal, indicating that rs10486567 is a marker represent-
the range of 1.1 to 1.25, as has been observed for
in an independent set of 1,725 cases and 35,392 controls of
rs10486567. This SNP has also been associated with PrCa
was previously suggested (6), or cases with
family history or early/late age of disease diagnosis.
Association analysis for 106 \(JAZF1\) SNPs with adjust-
ment for rs10486567 failed to reveal any independent
signal, indicating that rs10486567 is a marker represent-
ing a single common allele associated with PrCa risk
within \(JAZF1\).

Located in intron 2 of \(JAZF1\), rs10486567 is not predicted
to affect mRNA expression, splicing, or transcription factor or miRNA binding sites. Further deep resequencing

The risk allele \(G\) of rs10486567 is the major allele in
Europeans (0.73) and Africans (0.68), whereas it is rare
in East Asians (0.16) based on allele frequencies in Hap-
Map (19). Of note, PrCa is less frequent in individuals of
Asian ancestry (30). Despite differences in allele frequen-
cies, an association between rs10486567 and PrCa has
been noted in African Americans, Latinos, Japanese
Americans, and Native Hawaiians, but due to the small
sample sizes in non-Caucasian populations, the results of-
ten did not reach conclusive statistical significance (31). In
the first stage of GWAS for PrCa conducted by PRACTI-
CAL consortium that included 1,854 cases and 1,894 con-
trols of European ancestry, rs10486567 did not meet the
criteria \((P < 0.05 \text{ or } P_{\text{trend}} < 0.01)\) to be followed up in larger
sample sets (8). It is important to note that the first stage of
the GWAS mentioned above is not sufficiently powered to
detect most variants with modest effect size, nominally in
the range of 1.1 to 1.25, as has been observed for
rs10486567. This SNP has also been associated with PrCa
in an independent set of 1,725 cases and 35,392 controls of
European ancestry with OR 1.13 and \(P = 4.4 \times 10^{-3}\) (32).

Currently, there is no biological explanation for the
functional implications for \(JAZF1\) in prostate carcinogen-
esis. Intrachromosomal fusions between \(JAZF1\) and
\(SUZ12\) (33) or \(JAZF1\) and \(PHF\) (34) have been identified

| Table 3. Prostate cancer risk for rs10486567 of \(JAZF1\) stratified by potential risk factors |
|-------------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Stratified variable | No. controls/ cases | Risk allele freq. | \(P^*\) | Heterozygous | Homozygous |
|                   |                  | Controls | Cases | OR (95% CI) | \(P^\dagger\) | OR (95% CI) | \(P^\dagger\) |
| Age <60 y          | 1,853/1,761      | 0.767    | 0.799  | 0.07 | 1.12 (0.77-1.63) | 0.602 | 1.31 (0.91-1.89) | 0.549 |
| Age \(\geq60\) y    | 6,539/7,771      | 0.756    | 0.785  | \(2.74 \times 10^{-8}\) | 1.14 (0.98-1.34) | 1.37 (1.17-1.59) |
| No family history  | 6,480/7,087      | 0.758    | 0.787  | \(5.75 \times 10^{-8}\) | 1.11 (0.95-1.31) | 0.335 | 1.34 (1.14-1.57) | 0.384 |
| With family history| 642/1,521        | 0.759    | 0.794  | 0.041 | 1.39 (0.91-2.14) | 1.62 (1.07-2.45) |
| Nonaggressive cancer\(\dagger\) | 9,109/4,510      | 0.754    | 0.792  | \(3.25 \times 10^{-8}\) | 1.21 (1.12-1.31) | 0.528 | 1.41 (1.19-1.68) | 0.25 |
| Aggressive cancer\(\ddagger\) | 9,109/4,598      | 0.754    | 0.781  | \(1.60 \times 10^{-4}\) | 1.14 (1.05-1.23) | 1.29 (1.10-1.53) |

\(*P^* values are for additive logistic regression model adjusted for study and significant principle component per study.\)
\(\dagger\)Breslow-Day test for homogeneity.
\(\ddagger\)Compared with all controls.

| Table 4. Association results for SNPs of \(JAZF1\) previously associated with other traits |
|-------------------|------------------|------------------|------------------|-----------------|-----------------|-----------------|-----------------|
| SNP (minor allele)* | Location | MAF | \(\chi^2\) statistics | \(P^\dagger\) | Heterozygous | Homozygous | Adjusted for rs10486567 |
|                   |          |    |                      |                | OR (95% CI) | OR (95% CI) | \(P^\ddagger\) |
| rs849140 (T)       | 28,150,227 | 0.397 | 0.396 | 1.07 | 1.03 (0.96-1.09) | 0.99 (0.91-1.08) | 0.829 |
| rs849136 (A)       | 28,141,482 | 0.281 | 0.287 | 3.53 | 0.17 | 1.06 (0.99-1.13) | 1.03 (0.92-1.15) | 0.199 |
| rs849142 (G)       | 28,152,416 | 0.496 | 0.498 | 0.84 | 0.657 | 1.03 (0.96-1.11) | 1.01 (0.93-1.10) | 0.899 |

Abbreviation: MAF, minor allele frequency.

\(*rs849140 is associated with height; rs849136 is in \(r^2 = 1.0\) with rs849141 associated with body stature; rs849142 is in \(r^2 = 1.0\) with rs864745 associated with T2D and SLE.\)
\(\dagger\)P values are for genotype association test (additive model) adjusted for study and significant principal components.
\(\ddagger\)P values are from logistic additive genetic model adjusted for study and significant principal component per study; rs10486567 was also included as a covariate with 0, 1, and 2 allele dosage coding.
in endometrial cancer, but there are no reports on these or other fusions of JAZF1 in PrCa. It will be important to determine the molecular functions of JAZF1 that can be important for several traits with which genetic variants within JAZF1 have recently been associated. JAZF1 is a large gene of ~350 kb. Rs10486567, associated with risk of PrCa, is located in intron 2. Several SNPs located close to each other in another LD block within intron 1 and ~210 kb centromeric from rs10486567 have been associated with height (25), body stature (26), and increased risk of T2D (27) and SLE (28). We agnostically tested all these SNPs in our set of samples but observed no association with PrCa. An inverse correlation between the risk of PrCa and T2D has been observed in epidemiologic studies (35-37), a finding consistent with the results for genetic variants within the HNF1B (TCF2) gene (6, 38, 39). Although the inverse correlation between PrCa and T2D has also been suggested for variants in JAZF1 (40), our results did not show an effect of the T2D risk variant on susceptibility to PrCa.

In conclusion, our study has established rs10486567 within JAZF1 on chromosome 7p15.2 as a bona fide marker for association with susceptibility to PrCa in individuals of European ancestry. Here, we intentionally studied only tag SNPs from the region to test for presence of independent association signals. Future studies should be conducted to identify all common and uncommon variants by deep sequence analysis that are in strong LD with rs10486567 to nominate the optimal variants for evaluation of functional significance for susceptibility to PrCa.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Acknowledgments
We thank all individuals who participated in this study. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Grant Support
Federal funds from the National Cancer Institute, NIH, under contract no. HHSN261200800001E, and grants to the National Cancer Institute Breast and Prostate Cancer Cohort Consortium, U01-CA98233, U01-CA98710, U01-CA98216, and U01-CA98758. 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 11/19/2009; revised 03/03/2010; accepted 03/05/2010; published OnlineFirst 04/20/2010.

References
Cancer Epidemiology, Biomarkers & Prevention

Refining the Prostate Cancer Genetic Association within the JAZF1 Gene on Chromosome 7p15.2

Ludmila Prokunina-Olsson, Yi-Ping Fu, Wei Tang, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst April 20, 2010.

Updated version  Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-09-1181

Supplementary Material  Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2010/04/19/1055-9965.EPI-09-1181.DC1

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

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

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