Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men

Joke Beuten1,2,*, Jonathan A.L. Gelfond3, Margarita L. Martinez-Fierro4, Korri S. Weldon2, AnaLisa C. Crandall2, Augusto Rojas-Martinez4, Ian M. Thompson5 and Robin J. Leach1,2,5

1Department of Pediatrics, 2Department of Cellular and Structural Biology and 3Department of Epidemiology and Biostatistics, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA, 4Department of Biochemistry and Molecular Medicine, School of Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico and 5Department of Urology, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA

*To whom correspondence should be addressed. Tel: +1 210 567 6947; Fax: +1 210 567 6781; Email: beuten@uthscsa.edu

Genotyping of a 615 kb region within 8q24 with 49 haplotype-tagged single-nucleotide polymorphisms (SNPs) in 2109 samples (797 cases and 1312 controls) of two ethnic/racial groups found SNPs that are significantly associated with the risk for prostate cancer (PCa). The highest significance in Caucasian men was found for rs6983267; the AA genotype reduced the risk for PCa (odds ratio (OR) 0.65, 95% confidence interval (CI) 0.35–0.65, P = 2.74 × 10−4). This SNP also had a significant independent effect from other SNPs in the region in this group. In Hispanic men, rs7837328 and rs921146 showed independent effects (OR = 2.55, 95% CI = 1.51–4.31, P = 4.33 × 10−4, OR = 2.09, 95% CI = 1.40–3.12, P = 3.13 × 10−4, respectively). Significant synergist effects for increasing numbers of high-risk alleles were found in both ethnicities. Haplotype analysis revealed major haplotypes, containing the non-risk alleles, high-risk alleles were found in both ethnicities. Haplotype analytically. Significant synergist effects for increasing numbers of high-risk alleles were found in both ethnicities. Haplotype analysis revealed major haplotypes, containing the non-risk alleles, high-risk alleles were found in both ethnicities. Haplotype analytically.

Materials and methods

Subjects

Study subjects included men in the San Antonio Center for Biomarkers of Risk of Prostate Cancer (SABOR) cohort. SABOR is funded by the National Cancer Institute and has been prospectively enrolling healthy male volunteers since 2001. On each annual visit, a digital rectal examination was performed and serum prostate-specific antigen level ≤2.5 ng/ml on all study visits. Race/ethnicity was self-reported on a questionnaire completed at the time of enrollment. A total of 1441 Caucasians (601 cases and 840 controls) and 668 Hispanics (196 cases and 472 controls) were included in this analysis. Clinical characteristics of subjects are summarized in Table I. Study age among controls was the age at last follow-up, and age among cases was the age at PCa diagnosis;
controls were younger than PCa cases with a mean age (standard deviation, SD) of 60.9 (8.8) years and 66.0 (8.3) years, respectively (P < 0.0001).

SNP selection and genotyping
DNA was isolated from whole-blood cells using a QIAamp DNA Blood Maxi Kit (Qiagen, Valencia, CA). Forty-nine tagged SNPs covering the region 127992902–128608542 bp on chromosome 8 were selected using Haploview with the following criteria: (i) a minor allele frequency >0.05 in order to gain more statistical power; (ii) an r² threshold of 0.8 and a log of odds (LOD) threshold for multimarker testing of 3.0; (iii) a minimum distance between tags of 60 basepairs; (iv) SNPs for which an association with PCa has been reported were included and (v) we used the 2- and 3-marker haplotype tagging option (http://www.broad.mit.edu/mpg/haploview/). The selection was based on the information on the European population as provided by HapMap retrieved (http://www.broad.mit.edu/mpg/haploview/), the SNPMatrix tool in R [http://www.bioconductor.org/packages/2.3/bioc/html/ SNPMatrix.html, (24)]. Genotypes and frequency information for phases 1 and 2 of the HapMap chromosome 8 data were downloaded from NCBI build 36 (dbSNP build 126) coordinates for the European and Mexican populations (http://ftp.hapmap.org/ genotype/2008-10_phaseII/fixed Strand/non-redundant/). The 8q24 region under study was used as reference interval to which the remainder of chromosome 8 was analyzed for LD. A cutoff of LOD >3 was used to select regions on chromosome 8 in LD with the 8q24 region under study. For each ethnicity, we determined the LD across chromosome 8 for the SNPs that we found to be significant after correction for multiple testing from the single SNP analysis. (Note: rs7013278 and rs10094095 were not present in the HapMap data of Europeans and rs7013278 was not present in the HapMap data of Mexicans.) A hypothetical function for SNPs that were significant for single SNP analysis after corrections for multiple testing was assessed using silico analysis of transcription factor-binding sites: both possible alleles of each SNP were tested for their binding capability to human transcription factors using the web tool ‘transcription element search system’ [(25), http://www.cbil.upenn.edu/cgi-bin/tess/tess]. Options employed were 21 bases of genomic sequence around each SNP (10 bases on either side of the SNP) and string-based search query with default settings. A log-likelihood score of >16 for a pretty good match (deficit 1.0 or less) and >18 for a mismatch (deficit >1.0) was used as cutoff value for reporting.

Results
Single SNP analysis
Forty-nine SNPs were genotyped in 2380 samples. All SNPs were in HWE (P > 0.01) in the controls of each ethnicity/race, except for rs7825118 and rs7017671 that showed deviation from HWE in Caucasian and Hispanic controls, respectively. Although the error rate was <0.2%, SNPs that were not in HWE were omitted for further statistical analyses in the respective study groups. Table II displays minor allele frequencies of the SNPs estimated in both ethnicities. Significant case–control differences of allele frequencies at a level <0.05 were observed for 16 polymorphisms in Caucasians and for 14 SNPs within Hispanics. When analyzed individually, 16 SNPs in Caucasians and 12 SNPs in Hispanics were significantly associated with PCa risk at the P < 0.05 level (supplementary Table I is available at Carcinogenesis Online). After correction for multiple testing, 14 SNPs remained significant in Caucasians (P values between 0.03–2.74 × 10⁻⁶). The most significant result was obtained for rs6983267 for which the AA genotype reduces the risk for PCa (OR = 0.48, 95% CI = 0.35–0.65, P = 2.74 × 10⁻⁶). In Hispanics, 11 SNPs remained significant after correction for multiple testing (P values 0.028–1.84 × 10⁻⁴). SNP rs921146 was more significantly associated with PCa risk than rs6983267 in this ethnic group and carriers of the CC genotype have a 3.84-fold increase in risk (95% CI = 1.17–12.61, P = 0.026;
Supplementary Table I is available at Carcinogenesis Online. Although less significant, the OR was even higher for rs4871799 (OR = 3.06, 95% CI = 1.57–9.94, P = 0.003). Including the proportion of Native American ancestral background as covariate in the Hispanic sample did not show a difference in outcome with results that were not conditioned on this variable. Significance for SNPs rs1985829, rs6983267, rs7013278, rs7837328, rs4871022 and rs1447293 was found in both Caucasians and Hispanics. Figure 1 shows a diagram of the single SNP results (panel A) as well as the LD across the region (panel B) for each ethnic/racial group. A plot of the LD of the significant SNPs after false discovery rate is shown in panel C.

Comparing associations of Gleason grade ≥7 versus Gleason grade <7 among PCa cases, trends toward greater significance for higher Gleason grade in Caucasians and lower Gleason grade in Hispanics were found, but data were not consistent for all SNPs (supplementary Table II is available at Carcinogenesis Online). A similar finding was seen when looking at prognosis among the PCa cases where a slight increase in significance in the Caucasians and a slight decrease in significance in Hispanics for bad prognosis was observed. Those outcomes were again not consistent for all SNPs and must be interpreted with caution due to small sample sizes (supplementary Table II is available at Carcinogenesis Online).

After conditioning on other significant SNPs not in LD with each other and thus testing whether the statistically significant associations with PCs were independent in our groups, rs6983267 showed a main effect independent of other significant SNPs in Caucasians.
(P = 0.0004), and both rs7837328 (P = 0.009) and rs921146 showed significant independent associations in Hispanics (P = 0.0009 and P = 0.003, respectively; data not shown).

**Cumulative effect of significant SNPs**

To evaluate cumulative effects of the risk alleles defined in the single SNP analysis, we performed an age-adjusted multivariate logistic regression on combinations of risk alleles compared with the combination with no risk allele as reference. In Caucasians, the combination of the five risk genotypes, not in LD with each other, showed a significant association with PCa (P_trend = 3.6 x 10^{-8}) and a 3.18-fold increase in risk (95% CI = 2.11–4.79) was observed for men with all five risk alleles as compared with men without any risk alleles (Table III). A similar observation was found in the Hispanics, where the significant association between the combination of the risk alleles of three SNPs and PCa increases the risk significantly (OR = 4.98, 95% CI = 2.52–9.85, P_trend = 3.84 x 10^{-6}).

**Haplotype analysis of significant SNPs**

SNPs that were found to be significantly associated with PCa risk after correction for multiple testing and were not in LD with each other were included in haplotypes to examine the joint effect of variant alleles on risk in our population; these are the same SNPs as used for analyzing the cumulative effect. A major haplotype (28%) A-G-G-G-A for the SNPs rs6983267-rs6985419-rs7357486-rs10109622-rs1447293 showed a significant decrease in risk for PCa (OR = 0.68, 95% CI = 0.58–0.81, P = 1.65 x 10^{-5}; Table IV) in the Caucasians under the additive model. All alleles of the haplotype are the respective risk alleles as found in the single SNP analysis. In Hispanics, the major haplotype G-A-G (43%) for rs7837328-rs921146-rs6981321 is
significantly associated with disease risk with an OR of 0.71 (95% CI 0.55–0.91, P = 0.007) under the additive model and carries all three alleles associated with decreased risk in single SNP analysis.

Alterations of binding sites for transcription factors
To investigate possible functional implications of the significant SNPs, a search for transcription factor-binding sites using transcription element search system in non-coding genomic regions was performed. Two of 14 SNPs, rs7837328 and rs10094059 that show significant association with PCa in Caucasians, and three significant SNPs, rs7837328, rs921146 and rs6981321, in Hispanics have changes in transcription factor-binding properties related to allelic alterations (data not shown).

The presence of the G allele in rs7837328 creates a binding site for the repressor of the Interferon (IFN)-beta gene (26). The IFN-regulatory factor 1-binding site, which is a positive regulator of IFN-beta and IFN-induced genes, is present for the G allele in rs10094059 (27,28). A binding site for FOXJ2, a transcriptional activator, is present for the A allele in rs921146 (29). SNP rs6981321 has an albumin negative factor-binding site, which is a negative regulator, for the C allele (30).

LD of significant SNPs across chromosome 8
The pairwise LD calculations from the SNPMatrix tool in R indicated that several of the significant SNPs (after false discovery rate correction) within the 8q24 region studied in this report were in high LD (LOD > 3) with SNPs located within the CUB and Sushi Multiple Domains 1 gene (CSMD1) in both Caucasians and Hispanics. This gene is located on the short arm of chromosome 8, within the 8p23 region. In particular, rs11985829, rs6985419, rs1447293 and rs6981424 were in high LD with 10 SNPs within CSMD1 in Caucasians, and rs11985829, rs10956372 and rs6981321 were in high LD with nine SNPs within CSMD1 in Hispanics (Table V).

We also found high LD (LOD > 3) between significant SNPs and SNPs within the β-defensin-1 gene (DEFB1) located at 8p23 in Caucasians, within the CUB and Sushi multiple domains 3 gene (CSMD3) gene at 8q23 in Hispanics and within the pleckstrin and Sec7 domain containing 3 gene (PSD3) located at 8p22 in both Caucasians and Hispanics. High LD (LOD > 5) was also noticed between 8q24 SNPs and SNPs within the TMEM75 gene downstream of the region investigated in this study. In Caucasians, the highest LD was found for rs11985829 with rs7825794 and rs16903109 (LOD score of 5.36). A LOD score of 5.37 was found for LD between rs6981321 and rs2720672, with the latter located in TMEM75, in the Hispanics (data not shown).

Discussion
Genotyping of a 615 kb region within 8q24 with 49 haplotype-tagged SNPs in 2109 samples (797 cases and 1312 controls) of two ethnic/
racial groups found SNPs that are significantly associated with the risk for PCa. The highest significance in Caucasian men was found for rs6983267 for which the A allele is associated with a decreased risk for PCa. This SNP shows an independent effect from other significant SNPs in the studied region in this ethnic group. SNP rs6983267 is also significantly associated with a decreased risk for PCa in Hispanic men. However, rs7837328 and rs921146 reached higher significance and both SNPs showed independent effects, which was not found for rs6983267 in this ethnic group.

These results are consistent with previous reports showing significant association of rs6983267 in different ethnicities, including European American, Asian Indian and Japanese (31–35). In addition to confirming previous findings, these data now show this marker to play a role in the susceptibility of PCa in Hispanic men; to our knowledge, this is the first report of the involvement of this SNP in this ethnic group.

SNP rs6983267 has been shown to be associated with high Gleason and advanced PCa (1,34), although with inconsistent outcomes. Our results suggest that there might be a more pronounced effect of rs6983267 in Caucasian and Hispanic cases when measuring Gleason grade or prognosis. We found a trend toward increased risk of high Gleason grade and poor prognosis in Caucasians, whereas the opposite (decreased risk) was found in Hispanics. These data should be cautiously interpreted due to the small sample sizes; and thus, the association of rs6983267 with clinical features of PCa remains an open question.

One of two SNPs, rs7837328, that showed independent effect in Hispanics, has been shown to confer susceptibility of PCa in Europeans (32). However, rs921146 to date has not been reported to be significantly associated with PCa risk. Of interest is that this marker has a change in transcription binding for FOXJ2, a transcriptional activator. Further studies may identify the effect of this allele-specific change and its contribution to prostate carcinogenesis.

The frequency of the A allele of rs1447295, previously found to be a risk allele for PCa, among the Caucasian cases and controls in this study (11.6 and 9.3%, respectively) was similar to those reported by Amundadottir et al. (19). The allele frequency distribution of the microsatellite marker DG8S737 was similar to that found in previous reports (19,20,36). While the frequencies were consistent with previous reports, neither the 8 allele at microsatellite marker DG8S737 nor SNP rs1447295 was significantly associated with PCa risk in any of our ethnic groups.

Analysis of the significant SNPs in each ethnic group that are not in LD with each other further indicates a significant synergist effect for increasing numbers of potential high-risk genotypes in both ethnicities. In Caucasians, a >3-fold increase in the risk of PCa (OR = 3.18, P = 3.6 × 10⁻⁴) was found for carriers of all risk alleles of the five significant SNPs (rs6983267, rs6985419, rs7357486, rs10109622 and rs1447293). Carriers of all three risk alleles for rs7837328, rs921146 and rs6981321 had a 4.98 increased of PCa risk in Hispanics (P = 3.84 × 10⁻⁶). Analysis of the significant single SNPs in each ethnic group revealed major haplotypes containing the non-risk alleles, which were significantly protective against PCa. Both findings indicate that multiple interacting SNPs within 8q24 most probably confer increased risk of PCa.

The significant findings within 8q24 together with the fact that this region is gene poor have led many to speculate about hypotheses related to this region’s involvement in PCa risk (2,14). The results from SNP-matrix of the LD across chromosome 8 in this report emphasize the importance of a possible involvement of other regions far beyond the candidate 8q24 region. Indeed, we found that several SNPs in this region are in high LD with SNPs on the short arm of chromosome 8, a region often deleted in PCa (16,18,37). Of interest is that four SNPs found significant in Caucasians and three significant SNPs in Hispanics, showed high LD (LOD >3) with SNPs located within the CSMD1 gene on chromosome 8p23 when using ethnic-specific genotype data from HapMap. The CSMD1 gene, which contains multiple CUB and Sushi domains, encodes a large, type I transmembrane protein located on the surfaces of neuronal and epithelial cells; this protein’s function is unknown but is thought to participate in cell migration (38). No association studies between CSMD1 variants and risk of PCa have been reported so far. The gene has been shown to extend into the minimal regions of deletions within 8p23 and has been suggested as a candidate for a suppressor of multiple types of cancer (39,40). Decreased expression of CSMD1 is associated with advanced PCa (39). Both the sequence of the gene and the organization of the protein are highly conserved in the mouse. In light of these observations, further studies are warranted to clarify this interaction of CSMD1 with the 8q24 region.

Of note is that SNPs in other genes across chromosome 8 were found that showed high LD (LOD >3) with those studied in this report. In particular, SNPs within the β-defensin-1 gene (DEFB1) located at 8p23 in Caucasians, within the pleckstrin and Sec7 domain containing 3 gene (PSD3) located at 8p22 in both Caucasians and Hispanics and the CSMD3 gene at 8q23 in Hispanics were in strong LD with 8q24 SNPs. Cancer-specific loss of DEFB1, which plays an important role in the innate and adaptive immune response, has been found in prostatic carcinomas and is suggested to play a specific role in tumor suppression of advanced PCa via a pathway involving cMYC.

### Table V. LOD scores of LD between significant SNPs of 8q24 and SNPs within CSMD1

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position</th>
<th>Location in gene</th>
<th>rs11985829</th>
<th>rs6985419</th>
<th>rs1447293</th>
<th>rs6981424</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasians</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1026374</td>
<td>2641976</td>
<td>Flanking_3UTR</td>
<td><strong>3.00</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>rs7010867</td>
<td>2658917</td>
<td>Flanking_3UTR</td>
<td><strong>3.22</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>rs2962516</td>
<td>3306739</td>
<td>Intron</td>
<td><strong>3.62</strong></td>
<td>0.61</td>
<td>0.49</td>
<td><strong>3.18</strong></td>
</tr>
<tr>
<td>rs17066197</td>
<td>3513223</td>
<td>Intron</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.04</td>
</tr>
<tr>
<td>rs3849836</td>
<td>4027537</td>
<td>Intron</td>
<td><strong>3.30</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
</tr>
<tr>
<td>rs17069238</td>
<td>4029228</td>
<td>Intron</td>
<td><strong>3.30</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
</tr>
<tr>
<td>rs710261</td>
<td>4036818</td>
<td>Intron</td>
<td><strong>3.30</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
</tr>
<tr>
<td>rs14208460</td>
<td>5123201</td>
<td>Flanking_5UTR</td>
<td><strong>4.04</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>rs10112245</td>
<td>5124465</td>
<td>Flanking_5UTR</td>
<td><strong>4.04</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Hispanics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs78277228</td>
<td>2577400</td>
<td>Flanking_3UTR</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>3.92</strong></td>
<td></td>
</tr>
<tr>
<td>rs17755518</td>
<td>2583513</td>
<td>Flanking_3UTR</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>3.08</strong></td>
<td></td>
</tr>
<tr>
<td>rs1499693</td>
<td>3276858</td>
<td>Intron</td>
<td>0.00</td>
<td><strong>3.72</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2406295</td>
<td>3277924</td>
<td>Intron</td>
<td>0.00</td>
<td><strong>3.66</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7281211</td>
<td>3967725</td>
<td>Intron</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>3.96</strong></td>
<td></td>
</tr>
<tr>
<td>rs1658815</td>
<td>4262365</td>
<td>Intron</td>
<td>0.00</td>
<td>0.00</td>
<td><strong>3.09</strong></td>
<td></td>
</tr>
<tr>
<td>rs1038058</td>
<td>4513757</td>
<td>Intron</td>
<td><strong>3.01</strong></td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>rs2617020</td>
<td>4514625</td>
<td>Intron</td>
<td><strong>3.31</strong></td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>rs2617019</td>
<td>4514806</td>
<td>Intron</td>
<td><strong>3.01</strong></td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
and PA2 (41,42). A meta-analysis in breast cancer metastasis indicated that the PSD3 is significantly downregulated (43). The CSMD3 gene has previously been shown to have upregulated expression associated with chromosomal gains (44). Consistent with the findings of Camp et al. (14), LD was also found between the 8q24 region under study and the more telomeric located gene TMEM75.

A limitation of our study is that selection of the 49 tagged SNPs, covering a 615 kb region within the 8q24 candidate region, was based on HapMap data of the European population. Due to the ethnic-specific LD patterns, we might have missed some non-tagged SNPs in Hispanics and therefore the SNPs selected in the study may not fully represent all tagged variants in this ethnic group. In addition, our selection of SNPs differs from other reports and thus a complete comparison with several previously reported SNPs was not possible. However, it is clear that even with these weaknesses, our findings, coupled with data from previous reports, indicate that variants within 8q24 play a significant role in the susceptibility to PCA risk.

In summary, our results support the hypothesis that variants in the 8q24 region play an important role in susceptibility of PCA risk in several ethnic groups. This study is the first to confirm the importance of this region in Hispanic men. Our findings further suggest that multiple interacting SNPs within 8q24 but also in different regions on chromosome 8 far beyond this 8q24 candidate region most probably confer increased risk of PCA.

Supplementary material
Supplementary Tables I and II can be found at http://carcin.oxfordjournals.org/

Funding
Early Detection Research Network of the National Cancer Institute (U01 CA086402); American Cancer Society (TURSG-03-152-01-CCE), entitled ‘The Role of Genetic Variation in Prostate Cancer among Hispanics and Blacks’; Cancer Support Grant from the NCI number (P30 CA54174).

Acknowledgements
The participation of all study subjects in SABOR and in the prevalent prostate cancer studies at the University of Texas Health Science Center at San Antonio is gratefully acknowledged. The study could not have been accomplished without the skilled assistance of the SABOR clinical staff.

Conflict of Interest Statement: None declared.

References

Received May 4, 2009; revised June 3, 2009; accepted June 10, 2009