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

Joke Beuten[1,2,*], Jonathan A.L.Gelfond[3], Margarita
L.Martinez-Fierro[4], Korri S.Weldon[2], AnaLisa
C.Crandall[2], Augusto Rojas-Martinez[4], Ian M.Thompson[5]
and Robin J.Leach[1,2,5]

[1]Department of Pediatrics, [2]Department of Cellular and Structural Biology
and [3]Department of Epidemiology and Biostatistics, University of Texas
Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900,
USA, [4]Department of Biochemistry and Molecular Medicine, School of
Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico and
[5]Department 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) = 0.48, 95% confidence interval (CI) = 0.35–0.65, $P = 2.74 \times 10^{-6}$]. 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, $P = 4.33 \times 10^{-4}$, OR = 2.09, 95% CI = 1.40–3.12, $P = 3.13 \times 10^{-4}$, 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, conferred protection against PCa. We found high linkage disequilibrium between significant SNPs within the region and SNPs within the CUB and Sushi Multiple Domains 1 gene (*CSMD1*), on the short arm of chromosome 8 in both ethnicities. These data suggest that multiple interacting SNPs within 8q24, as well as different regions on chromosome 8 far beyond this 8q24 candidate region, may confer increased risk of PCa. This is the first report to investigate the involvement of 8q24 variants in the susceptibility for PCa in Hispanic men.

Introduction

The distal long arm of chromosome 8 has been implicated by genome-wide association studies in several cancers, including colorectal, prostate, ovarian and smoking-related carcinogenesis. Multiple independent studies have demonstrated compelling evidence that genetic variations in at least three regions of 8q24 independently influence the risk of prostate cancer (PCa) [meta-analysis by Cheng *et al.* (1)]. Some of the associations within 8q24 were found to be population specific, in particular the most centromeric region appears to play a more significant role in non-European populations (2). Associations of 8q24 variants with aggressive PCa and/or increased tumor grade have been reported but are yet to be confirmed (1).

Despite the evidence of the importance of the 8q24 region in PCa risk, none of the associated single-nucleotide polymorphisms (SNPs) appear to cause functional changes. Furthermore, the region is known to be gene poor and so far, only one gene has been reported

Abbreviations: CI, confidence interval; HWE, Hardy–Weinberg equilibrium; IFN, interferon; LD, linkage disequilibrium; LOD, log of odds; OR, odds ratio; PCa, prostate cancer; SNP, single-nucleotide polymorphism; *TMEM75*, transmembrane protein 75.

to be located in the associated region, the pseudogene *POU5F1P1*, a retrotransposed copy of the POU-domain transcription factor gene *POU5F1* (3). *POU5F1*, located at 6p21, has been shown to promote tumor growth and to play a role in maintaining stem cell pluripotency, self-renewal and chromatin structure (4,5). An increased number of cells expressing Oct4A, a splice variant of the *POU5F1* (*Oct3/4*) gene, have been found in PCa (6). This 8q24 region also contains several annotated genes: the family with sequence similarity 84 member B gene (*FAM84B*), the oncogene *MYC* and the transmembrane protein 75 (*TMEM75*). *FAM84B* (alias *BCMP101*) is involved in the formation of DNA repair complex (7,8) and is over-expressed in breast cancer (9). The *MYC* proto-oncogene regulates expression of numerous target genes that control key cellular functions, including cell growth and cell cycle progression. *MYC* also has a critical role in DNA replication. Deregulated *MYC* expression resulting from various types of genetic alterations leads to constitutive *MYC* activity in several cancers and promotes oncogenesis (10). However, studies suggest that the 8q24 risk alleles do not affect *MYC* expression (11,12).

Since the biological mechanisms underlying the 8q24 associations with cancer remains unclear and the fact that 8q24 is the most frequently gained chromosomal region in prostate tumors (13), the risk variants may predispose to PCa through either increased genome instability and/or could be markers for the true causal factors. Indeed, it has been shown that linkage disequilibrium (LD) stretches far beyond the interval of 8q24 where associations have been shown (14); as such, other regions of chromosome 8 might contain a causal variants and/or underlying genes. It has been shown that gain of 8q is often accompanied by the allelic loss of 8p in PCa (15,16). Begley *et al.* (17) further showed that transcription of genes in 8p or 8q were down-regulated in cells hemizygyous for 8p and upregulated in cells carrying three copies of 8q, respectively. In addition, deletions within 8p are the most common deletion event in the genome of prostate tumors (18). Sun *et al.* (16) found that 30% of PCas had a deletion at 8p21.3 and that the deleted region spans a large interval extending into 8p23.3 and 8p21.1 and contains many genes.

We genotyped 49 tagged SNPs and one microsatellite marker DG8S737 covering 615 kb of the 8q24 region previously shown to be involved in PCa risk in 2109 samples (797 cases and 1312 controls) of non-Hispanic Caucasian (Caucasians) or Hispanic Caucasian (Hispanics) origin. Our goals were to confirm previous associations and determine population specificity, in particular in Hispanics who have not been analyzed for this region. We also determined regions on chromosome 8p that are in LD with the region under study.

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 was determined. From this cohort, 197 incident cases (136 Caucasians and 61 Hispanics) were available. We also included 600 cases with a known history of PCa that are enrolled within the same time period in a parallel study of prevalent PCa. Institutional review board approval was obtained and informed consent from subjects in both studies. Cases had biopsy-confirmed PCa and controls consisted of male volunteers of at least 45 years old who had normal digital rectal examination and 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;

Table I. Clinical data of the study group

Subgroup	Cases (N = 797)	Controls (N = 1312)	
Ethnic background, *n* (%)			
Caucasian	601 (68.4)	840 (56.2)	
Hispanic	196 (22.3)	472 (31.6)	
Age, *n* (%)			
≤50	26 (3.3)	156 (11.9)	
51–60	182 (23.0)	473 (36.0)	
61–70	345 (43.6)	435 (33.2)	
>70	238 (30.1)	248 (18.9)	
Mean ± SD	66.0 ± 8.3	60.9 ± 8.8	$P < 0.0001$
Prostate-specific antigen (ng/ml)			
≤4.0	154	1312	
4.1–10.0	30	0	
10.1–20.0	2	0	
>20.0	2	0	
Mean ± SD	2.8 ± 3.3	0.8 ± 0.5	$P < 0.0001$
Gleason (N = 529)			
<7	300		
7	145		
>7	84		
Prognosis (N = 484)			
Poor prognosis	109		
Good prognosis	375		

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 r^2 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 from NCBI dbSNP Build 127 (www.hapmap.org). Genotyping was performed with the Golden Gate assay of the VeraCode technology using the BeadXpress Reader System according to the manufacturer's protocol (Illumina, San Diego, CA). Primers and probe sequences are available upon request. For the microsatellite marker DG8S737, the primers were described by Amundadottir *et al.* (19) and genotyping was performed as described previously by Wang *et al.* (20). To ensure reliability of the results, duplicate samples and/or known genotyped samples were included in the analysis as quality controls.

Statistics

Haploview version 4 beta 15 was used to check for Hardy–Weinberg equilibrium (HWE) for each SNP and to measure LD between the SNPs within the studied region for each race/ethnicity [(21), http://www.broad.mit.edu/mpg/haploview/].

The allele frequency for each SNP was determined in each ethnic group and the frequencies among the case–control groups were compared using the chi-square test. Association analyses were performed using R statistical software version 2.8.1 and were stratified by ethnicity. The odds ratio (OR) and its 95% confidence interval (CI) were estimated by unconditional logistic regression as a measure of the associations between genotypes and PCa risk. Associations by Gleason grade (Gleason score ≥7 versus <7) or prognosis (defined as Gleason score of 7 or higher or stage T3b or higher) were examined by logistic regression in case-only analyses. We used additive, dominant and recessive models in the test analysis and only considered results with a minimum of five individuals for a specific model. For the microsatellite DG8S737, we performed both the −8 allelic association (comparing allele −8 versus all other alleles) as well as a Wald test in which the allele count of all alleles were considered. To correct for multiple testing, we used the method of Storey *et al.* (22) based on the concept of false discovery rate. This estimation showed that for $P < 0.03$, the probability that the association is a true positive is >90% in the whole sample group. To test for the independent effect of a significant SNP while adjusting for other SNPs, we used a generalized linear model function from the R statistical package for which each SNP selected is entered into

a single multivariable logistic regression model. SNPs in the model were taken to have additive effects.

The cumulative effect of combined genotypes on PCa risk was estimated by counting the number of genotypes associated with PCa, on the basis of the best-fitting genetic inheritance from single SNP analysis. ORs and their 95% CIs were calculated for men carrying any combination of one, two or more alleles associated with PCa as compared with men carrying none of the risk alleles using the unconditional logistic regression analysis. We selected SNPs that were not in LD with each other ($D' < 0.8$). If several SNPs presented higher LD values, we selected the most significant SNP.

Logistic regression was used to calculate the ORs of the haplotypes, using the method implemented in the haplo.ccs package (23). Only major haplotypes, with an estimated frequency of >5%, are considered in this report. The model was fit for each major haplotype so that the OR of each major haplotype was computed relative to a reference group consisting of all other haplotypes including rare haplotypes. Three genetic models (additive, dominant and recessive) were tested.

For all statistical analyses, age was used as covariate. Individuals with missing data for a particular analysis were removed from the analysis. All statistical tests were two sided and significance was set at $P < 0.05$. To test whether possible associations were true positives or due to confounding/admixture association in Hispanics, we adjusted for the proportion of ancestry based on the genotyping results of 64 ancestry informative markers (J.Beuten, I.Halder, K.S.Weldon, R.J.Leach, I.M.Thompson, M.Stern, D.M.Lehman, in preparation).

To measure LD between the markers across whole chromosome 8, we used 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 HapMap chromosome 8 data were downloaded from NCBI build 36 (dbSNP b126) coordinates for the European and Mexican populations (http://ftp.hapmap.org/genotypes/2008-10_phaseII/fwd_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 rs10094059 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 in 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, $P = 2.74 \times 10^{-6}$). 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, $P = 0.026$;

Table II. Genes, SNP selection, their location, MAF in cases and controls of each ethnicity

SNP	Position	Minor allele	Caucasians			Hispanics		
			MAF cases	MAF controls	P value[a]	MAF cases	MAF controls	P value[a]
rs979200	127992902	A	0.357	0.351	0.742	0.351	0.363	0.705
rs1456310	128121615	A	0.398	0.392	0.744	0.571	0.532	0.217
rs6993569	128153279	A	0.051	0.058	0.426	0.141	0.149	0.735
rs6470494	128157086	A	0.295	0.275	0.242	0.337	0.321	0.581
rs16901979	128194098	T	0.039	0.033	0.362	0.062	0.036	**0.040**
rs13281615	128424800	G	0.396	0.39	0.746	0.658	0.647	0.702
rs16902124	128426400	A	0.047	0.049	0.811	0.257	0.263	0.825
rs16902126	128451539	C	0.429	0.436	0.706	0.238	0.234	0.889
rs10505476	128477298	A	0.286	0.259	0.108	0.197	0.159	0.100
rs11985829	128478414	A	0.333	0.282	**0.003**	0.285	0.215	**0.008**
rs17467139	128481192	G	0.063	0.065	0.844	0.046	0.044	0.882
rs6983267	128482487	A	0.429	0.513	**9.2×10^{-6}**	0.345	0.438	**0.003**
rs7013278	128484074	A	0.403	0.348	**0.003**	0.351	0.269	**0.004**
rs13248944	128489740	G	0.053	0.049	0.639	0.015	0.019	0.644
rs7837328	128492309	A	0.456	0.402	**0.004**	0.376	0.296	**0.007**
rs4871022	128496902	G	0.455	0.401	**0.004**	0.377	0.305	**0.014**
rs6985419	128498903	A	0.319	0.266	**0.006**	0.199	0.208	0.761
rs7842552	128500876	G	0.308	0.27	**0.028**	0.218	0.181	0.127
rs9297756	128509349	A	0.179	0.142	**0.008**	0.115	0.084	0.099
rs6995633	128509833	A	0.112	0.09	0.069	0.227	0.219	0.766
rs6999921	128510110	G	0.097	0.079	0.090	0.092	0.075	0.313
rs7357486	128510805	A	0.308	0.266	**0.013**	0.173	0.17	0.925
rs10090421	128522947	G	0.275	0.232	**0.010**	0.203	0.176	0.262
rs6981397	128524836	G	0.288	0.315	0.127	0.376	0.372	0.898
rs7012462	128526872	A	0.423	0.414	0.630	0.394	0.404	0.748
rs10109622	128527333	A	0.234	0.196	**0.014**	0.187	0.187	0.994
rs4871791	128527826	A	0.451	0.47	0.312	0.49	0.455	0.267
rs10094059	128530789	C	0.228	0.261	**0.044**	0.3	0.313	0.642
rs7841264	128535996	A	0.174	0.196	0.136	0.323	0.313	0.726
rs7017671	128536936	C	0.481	0.488	0.727	0.426[b]	0.432[b]	0.826[b]
rs10099905	128537116	A	0.207	0.212	0.754	0.277	0.3	0.418
rs10956372	128539438	T	0.327	0.305	0.202	0.35	0.276	**0.010**
rs7830412	128540223	A	0.045	0.045	0.993	0.115	0.092	0.200
rs10094871	128541151	G	0.165	0.18	0.319	0.131	0.117	0.508
rs1447293	128541502	G	0.373	0.354	0.311	0.454	0.373	**0.008**
rs921146	128544367	C	0.215	0.216	0.951	0.222	0.132	**1.0×10^{-4}**
rs7825118	128544999	A	0.272[b]	0.254[b]	0.274[b]	0.323	0.392	**0.022**
DG8S737	128545629	25	0.363	0.301	**0.004**	0.307	0.262	**0.030**
rs4871799	128551824	G	0.236	0.233	0.830	0.263	0.193	**0.007**
rs6981424	128552278	A	0.396	0.435	**0.038**	0.251	0.28	0.298
rs1447295	128554220	A	0.116	0.093	**0.045**	0.119	0.064	**0.002**
rs9643227	128565278	A	0.504	0.526	0.245	0.418	0.419	0.984
rs13258548	128566029	G	0.313	0.334	0.253	0.25	0.266	0.559
rs16902172	128567224	G	0.207	0.191	0.296	0.287	0.346	**0.043**
rs7831150	128568620	G	0.03	0.033	0.700	0.064	0.08	0.325
rs16902173	128573181	G	0.142	0.162	0.154	0.116	0.09	0.166
rs1562431	128576833	A	0.061	0.06	0.972	0.067	0.073	0.712
rs4078240	128580745	G	0.131	0.136	0.716	0.122	0.111	0.571
rs6981321	128582487	C	0.277	0.259	0.275	0.328	0.282	0.106
rs7837688	128608542	A	0.107	0.085	**0.047**	0.089	0.059	0.058

MAF, minimum allele frequency. Significant *P* values are in bold. For the repeat polymorphisms, the MAF are values below the median repeat length.
[a]Assumes HWE.
[b]Markers not in HWE (*P* < 0.01) in the controls.

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 rs11985829, 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 PCa were independent in our groups, rs6983267 showed a main effect independent of other significant SNPs in Caucasians
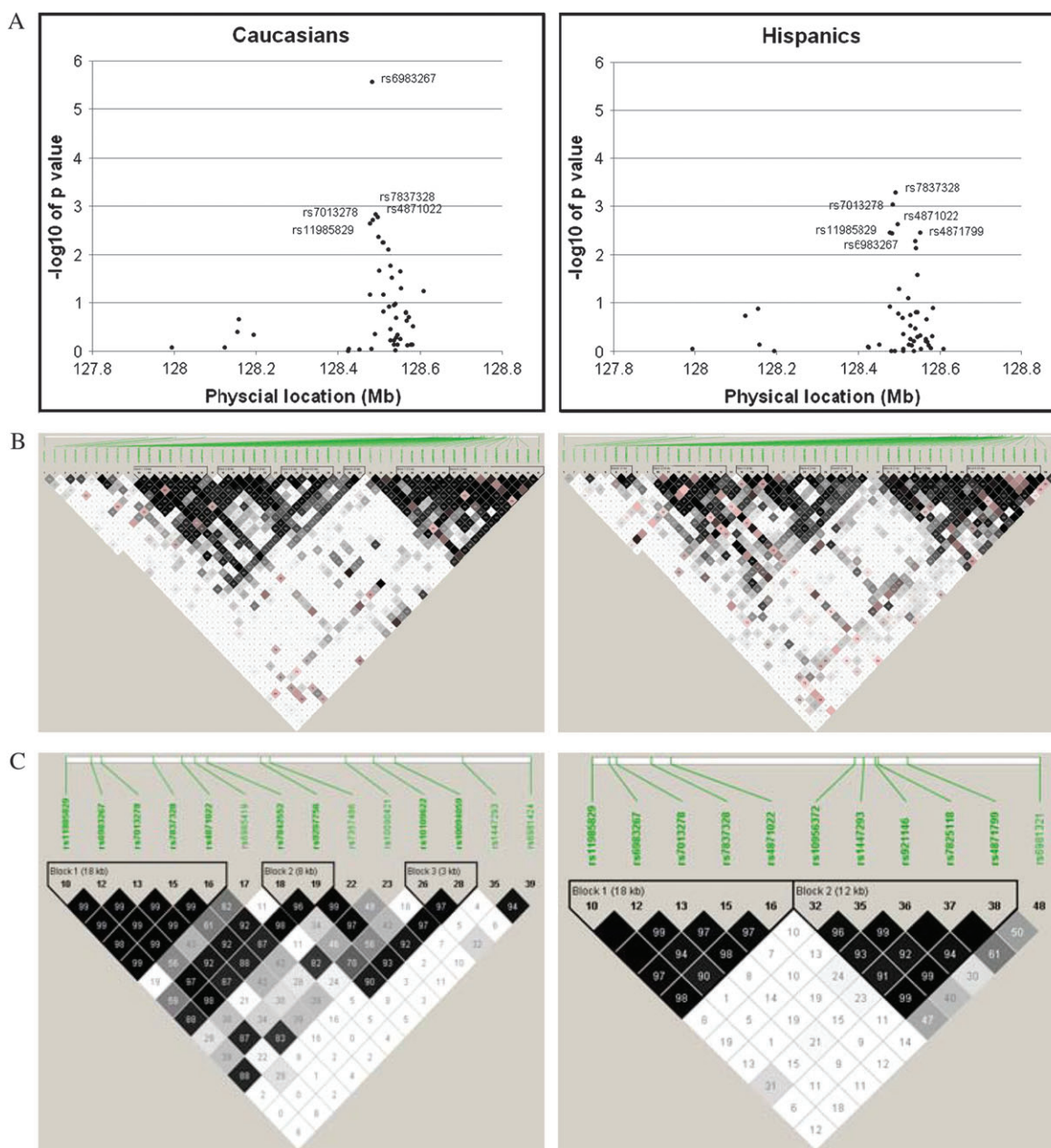
Fig. 1. Panel (**A**) Diagram of the effects of single SNPs on PCa risk. Results are for the additive model in Caucasians, and dominant model in Hispanics. Panel (**B**) shows the LD for the SNPs across the 8q24 interval studied in each ethnic group. Panel (**C**) shows the pairwise LD pattern for the significant SNPs from single SNP analysis with PCa in Caucasians (left) and Hispanics (right).

($P = 0.0004$), 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 ($P_{trend} = 3.6 \times 10^{-8}$) 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, $P_{trend} = 3.84 \times 10^{-6}$).

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, $P = 1.65 \times 10^{-5}$; 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

Table III. Cumulative effects of risk variants

Markers	Number of risk genotypes	Controls	Cases	OR (95% CI)[a]	P value
Caucasians					
rs6983267, rs6985419, rs7357486, rs10109622, rs1447293	0	71	24	Ref	—
	1	79	43	1.55 (0.85–2.82)	0.15
	2	133	83	1.85 (1.07–3.18)	0.03
	3	183	133	2.18 (1.30–3.67)	0.003
	4	113	99	2.52 (1.47–4.33)	0.0008
	5	91	113	3.81 (2.21–6.57)	1.57×10^{-6}
	Trend			3.18 (2.11–4.79)	3.6×10^{-8}
Hispanics					
rs7837328, rs921146, rs6981321	0	143	49	Ref	—
	1	157	63	1.29 (0.81–2.06)	0.28
	2	70	46	2.45 (1.43–4.19)	0.001
	3	9	17	6.95 (2.75–17.55)	4.15×10^{-5}
	Trend			4.98 (2.52–9.85)	3.84×10^{-6}

[a]Age adjusted.

Table IV. Association of common haplotypes with PCa risk in Caucasian and Hispanic men under the additive model

SNP combination	Frequency (%)	No. of haplotypes		OR[a]	95% CI	P
		Cases	Controls			
Caucasians, rs6983267-rs6985419-rs7357486-rs10109622-rs1447293						
A-G-G-G-A	28	211	349	**0.68**	**0.58–0.81**	**1.65×10^{-5}**
A-G-G-G-G	14	120	182	0.86	0.70–1.06	0.16
C-A-G-G-A	11	106	125	1.16	0.91–1.49	0.23
C-G-A-G-A	10	104	122	1.18	0.92–1.51	0.18
C-A-G-G-G	7	75	90	1.19	0.90–1.56	0.22
C-G-A-G-G	6	56	69	1.05	0.77–1.42	0.78
C-G-G-A-A	5	57	62	1.19	0.85–1.65	0.31
Hispanics, rs7837328-rs921146-rs6981321						
G-A-G	43	106	272	**0.71**	**0.55–0.91**	**0.007**
A-A-G	20	66	133	1.06	0.80–1.41	0.69
G-A-C	15	44	112	0.83	0.60–1.15	0.25
G-C-C	6	26	37	1.62	1.04–2.53	0.03
A-A-C	5	18	39	0.93	0.61–1.44	0.76

Significant results after Bonferroni correction are in bold ($P < 0.007$ and $P < 0.01$ in Caucasians and Hispanics, respectively). Only common haplotypes ($>5\%$) are shown.
[a]OR is age adjusted.

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 correc-

tion) 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 the *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-1gene (*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/

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

SNP	Position	Location in gene	rs11985829	rs6985419	rs1447293	rs6981424
Caucasians						
rs1026374	2641976	Flanking_3UTR	**3.00**	0.00	0.00	0.00
rs7010867	2658917	Flanking_3UTR	**3.22**	0.00	0.00	0.00
rs2924887	2662516	flanking_3UTR	**3.62**	0.00	0.00	0.00
rs17066197	3306739	Intron	0.61	0.49	**3.18**	0.64
rs41491246	3513223	Intron	0.00	0.00	1.04	**3.74**
rs3849836	4027537	Intron	**3.30**	0.00	0.00	0.58
rs17069238	4029228	Intron	**3.30**	0.00	0.00	0.58
rs710261	4036818	Intron	**3.30**	0.00	0.00	0.58
rs1420846	5123201	Flanking_5UTR	0.00	**4.04**	0.00	0.00
rs10112245	5124465	Flanking_5UTR	0.00	**4.04**	0.00	0.00
Hispanics			rs11985829	rs10956372	rs6981321	
rs7827228	2577400	Flanking_3UTR	0.00	0.00	**3.92**	
rs17755518	2583513	Flanking_3UTR	0.00	0.00	**3.08**	
rs1499693	3276858	Intron	0.00	**3.72**	0.00	
rs2406295	3277924	Intron	0.00	**3.66**	0.00	
rs7816211	3967725	Intron	0.00	0.00	**3.96**	
rs1658815	4262365	Intron	0.00	0.00	**3.09**	
rs1038058	4513757	Intron	**3.01**	0.00	0.00	
rs2617020	4514625	Intron	**3.31**	0.00	0.00	
rs2617019	4514806	Intron	**3.01**	0.00	0.00	

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), albeit 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 association 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) were 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 indicate a significant synergist effect for increasing numbers of potential high-risk genotypes in both ethnic-ities. In Caucasians, a >3-fold increase in the risk of PCa (OR = 3.18, $P = 3.6 \times 10^{-8}$) 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 in Hispanics ($P = 3.84 \times 10^{-6}$). 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 has led many to speculate about hypotheses related to this region's involvement in PCa risk (2,14). The results from SNPmatrix 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 c*MYC*

and *PAX2* (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

Acknowledgements

References

1. Cheng,I. *et al.* (2008) 8q24 and prostate cancer: association with advanced disease and meta-analysis. *Eur. J. Hum. Genet.*, **16**, 496–505.
2. Haiman,C.A. *et al.* (2007) Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.*, **39**, 638–644.
3. Suo,G. *et al.* (2005) Oct4 pseudogenes are transcribed in cancers. *Biochem. Biophys. Res. Commun.*, **337**, 1047–1051.
4. Gidekel,S. *et al.* (2003) Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell*, **4**, 361–370.
5. Campbell,P.A. *et al.* (2007) Oct4 targets regulatory nodes to modulate stem cell function. *PLoS ONE*, **2**, e553.
6. Sotomayor,P. *et al.* (2009) Oct4A is expressed by a subpopulation of prostate neuroendocrine cells. *Prostate*, **69**, 401–10.
7. McDonald,W.H. *et al.* (2003) Novel essential DNA repair proteins Nse1 and Nse2 are subunits of the fission yeast Smc5-Smc6 complex. *J. Biol. Chem.*, **278**, 45460–45467.
8. Pebernard,S. *et al.* (2004) Nse1, Nse2, and a novel subunit of the Smc5-Smc6 complex, Nse3, play a crucial role in meiosis. *Mol. Biol. Cell*, **15**, 4866–4876.
9. Adam,P.J. *et al.* (2003) Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer. *J. Biol. Chem.*, **278**, 6482–6489.
10. Dominguez-Sola,D. *et al.* (2007) Non-transcriptional control of DNA replication by c-Myc. *Nature*, **448**, 445–451.
11. Gudmundsson,J. *et al.* (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.*, **39**, 631–637.
12. Kiemeney,L.A. *et al.* (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, **40**, 1307–1312.
13. Baudis,M. *et al.* (2001) Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics*, **17**, 1228–1229.
14. Camp,N.J. *et al.* (2007) Statistical recombinant mapping in extended high-risk Utah pedigrees narrows the 8q24 prostate cancer locus to 2.0 Mb. *Prostate*, **67**, 1456–1464.
15. Das,K. *et al.* (2005) Chromosomal changes in prostate cancer: a fluorescence *in situ* hybridization study. *Clin. Genet.*, **68**, 40–47.
16. Sun,J. *et al.* (2007) DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. *Prostate*, **67**, 692–700.
17. Begley,L. *et al.* (2006) Concordant copy number and transcriptional activity of genes mapping to derivative chromosomes 8 during cellular immortalization *in vitro*. *Genes Chromosomes Cancer*, **45**, 136–146.
18. Dong,J.T. (2001) Chromosomal deletions and tumor suppressor genes in prostate cancer. *Cancer Metastasis Rev.*, **20**, 173–193.
19. Amundadottir,L.T. *et al.* (2006) A common variant associated with prostate cancer in European and African populations. *Nat. Genet.*, **38**, 652–658.
20. Wang,L. *et al.* (2007) Two common chromosome 8q24 variants are associated with increased risk for prostate cancer. *Cancer Res.*, **67**, 2944–2950.
21. Barrett,J.C. *et al.* (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
22. Storey,J.D. *et al.* (2003) Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*, **100**, 9440–9445.
23. French,B. *et al.* (2006) Simple estimates of haplotype relative risks in case-control data. *Genet. Epidemiol.*, **30**, 485–494.
24. Clayton,D. *et al.* (2007) An R package for analysis of whole-genome association studies. *Hum. Hered.*, **64**, 45–51.
25. Schug,J. *et al.* (1997) *TESS: Transcription Element Search Software on the WWW. Technical report CBIL-TR-1997-1001-v0.0 Computational Biology and Informatics Laboratory School of Medicine*. University of Pennsylvania, PA, USA.
26. Keller,A.D. *et al.* (1991) Identification and characterization of a novel repressor of beta-interferon gene expression. *Genes Dev.*, **5**, 868–879.
27. Reis,L.F. *et al.* (1992) Critical role of a common transcription factor, IRF-1, in the regulation of IFN-beta and IFN-inducible genes. *EMBO J.*, **11**, 185–193.
28. Harada,H. *et al.* (1993) Anti-oncogenic and oncogenic potentials of interferon regulatory factors-1 and -2. *Science*, **259**, 971–974.
29. Perez-Sanchez,C. *et al.* (2000) FHX, a novel fork head factor with a dual DNA binding specificity. *J. Biol. Chem.*, **275**, 12909–12916.
30. Herbst,R.S. *et al.* (1990) The mouse albumin enhancer contains a negative regulatory element that interacts with a novel DNA-binding protein. *Mol. Cell. Biol.*, **10**, 3896–3905.
31. Zheng,S.L. *et al.* (2007) Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. *J. Natl Cancer Inst.*, **99**, 1525–1533.
32. Yeager,M. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, **39**, 645–649.
33. Salinas,C.A. *et al.* (2008) Multiple independent genetic variants in the 8q24 region are associated with prostate cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 1203–1213.
34. Tan,Y.C. *et al.* (2008) Common 8q24 sequence variations are associated with Asian Indian advanced prostate cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 2431–2435.
35. Terada,N. *et al.* (2008) Association of genetic polymorphisms at 8q24 with the risk of prostate cancer in a Japanese population. *Prostate*, **68**, 1689–1695.
36. Suuriniemi,M. *et al.* (2007) Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 809–814.
37. Chang,B.L. *et al.* (2007) Integration of somatic deletion analysis of prostate cancers and germline linkage analysis of prostate cancer families reveals two small consensus regions for prostate cancer genes at 8p. *Cancer Res.*, **67**, 4098–4103.
38. Kraus,D.M. *et al.* (2006) CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J. Immunol.*, **176**, 4419–4430.
39. Paris,P.L. *et al.* (2004) Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum. Mol. Genet.*, **13**, 1303–1313.

40. Sun,P.C. *et al*. (2001) Transcript map of the 8p23 putative tumor suppressor region. *Genomics*, **75**, 17–25.

41. Donald,C.D. *et al*. (2003) Cancer-specific loss of beta-defensin 1 in renal and prostatic carcinomas. *Lab. Invest.*, **83**, 501–505.

42. Bullard,R.S. *et al*. (2008) Functional analysis of the host defense peptide Human Beta Defensin-1: new insight into its potential role in cancer. *Mol. Immunol.*, **45**, 839–848.

43. Thomassen,M. *et al*. (2009) Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis. *Breast Cancer Res. Treat.*, **113**, 239–249.

44. Jiang,M. *et al*. (2008) Simultaneously detection of genomic and expression alterations in prostate cancer using cDNA microarray. *Prostate*, **68**, 1496–1509.