ORIGINAL ARTICLE

Cumulative Association of Five Genetic Variants with Prostate Cancer

S. Lilly Zheng, M.D., Jielin Sun, Ph.D., Fredrik Wiklund, Ph.D., Shelly Smith, M.S., Pär Stattin, M.D., Ph.D., Ge Li, M.D., Hans-Olov Adami, M.D., Ph.D., Fang-Chi Hsu, Ph.D., Yi Zhu, B.S., Katarina Bälter, Ph.D., A. Karim Kader, M.D., Ph.D., Aubrey R. Turner, M.S., Wennuan Liu, Ph.D., Eugene R. Bleecker, M.D., Deborah A. Meyers, Ph.D., David Duggan, Ph.D., John D. Carpten, Ph.D., Bao-Li Chang, Ph.D., William B. Isaacs, Ph.D., Jianfeng Xu, M.D., D.P.H., and Henrik Grönberg, M.D., Ph.D.

ABSTRACT

BACKGROUND

Single-nucleotide polymorphisms (SNPs) in five chromosomal regions — three at 8q24 and one each at 17q12 and 17q24.3 — have been associated with prostate cancer. Each SNP has only a moderate association, but when SNPs are combined, the association may be stronger.

METHODS

We evaluated 16 SNPs from five chromosomal regions in a Swedish population (2893 subjects with prostate cancer and 1781 control subjects) and assessed the individual and combined association of the SNPs with prostate cancer.

RESULTS

Multiple SNPs in each of the five regions were associated with prostate cancer in single SNP analysis. When the most significant SNP from each of the five regions was selected and included in a multivariate analysis, each SNP remained significant after adjustment for other SNPs and family history. Together, the five SNPs and family history were estimated to account for 46% of the cases of prostate cancer in the Swedish men we studied. The five SNPs plus family history had a cumulative association with prostate cancer (P for trend, 3.93×10^{-28}). In men who had any five or more of these factors associated with prostate cancer, the odds ratio for prostate cancer was 9.46 ($P=1.29\times10^{-8}$), as compared with men without any of the factors. The cumulative effect of these variants and family history was independent of serum levels of prostate-specific antigen at diagnosis.

CONCLUSIONS

SNPs in five chromosomal regions plus a family history of prostate cancer have a cumulative and significant association with prostate cancer.

From the Center for Human Genomics (S.L.Z., J.S., S.S., G.L., F.-C.H., Y.Z., A.R.T., W.L., E.R.B., D.A.M., B.-L.C., J.X.) and the Departments of Biostatistical Sciences (F.-C.H.) and Urology (A.K.K.), Wake Forest University School of Medicine, Winston-Salem, NC; the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm (F.W., H.-O.A., K.B., H.G.); the Department of Urology, Umeå University Hospital, Umeå, Sweden (P.S.); the Department of Epidemiology, Harvard School of Public Health, Boston (H.-O.A.); Translational Genomics Research Institute, Phoenix, AZ (D.D., J.D.C.); and Johns Hopkins Medical Institutions, Baltimore (W.B.I.). Address reprint requests to Dr. Xu at the Center for Human Genomics, Medical Center Blvd., Winston-Salem, NC 27157, or at jxu@wfubmc.edu; or to Dr. Isaacs at Marburg 115, Johns Hopkins Hospital, 600 N. Wolfe St., Baltimore, MD 21287, or at wisaacs@jhmi.edu.

This article (10.1056/NEJMoa075819) was published at www.nejm.org on January 16, 2008.

N Engl J Med 2008;358:910-9.

Genomewide association studies of complex diseases have identified sequence variants that are consistently associated with the risk of such diseases.[1] Often such variants have limited use in the assessment of disease risk in an individual patient, since most of them confer a relatively small risk. Whether combinations of individual variants confer larger, more clinically useful associations with increased risk remains to be shown.

Age, race, and family history are three factors that have a consistent association with the risk of prostate cancer.[2] A meta-analysis showed a pooled odds ratio of 2.5 for men who had a first-degree relative with the disease.[3] Recently, genomewide analysis has identified variants in five chromosomal regions that are significantly associated with a risk of prostate cancer. These variants occur in three independent regions at 8q24[4-7] and in one region at 17q12 and another at 17q24.3.[8] These five regions probably harbor genes that confer susceptibility to prostate cancer or regulate factors affecting critical genes, but the specific genes in these regions have not been identified.

Individually, single-nucleotide polymorphisms (SNPs) in each of the five chromosomal regions were shown to have only a moderate association with prostate cancer in previous studies. In our study, we investigated whether a combination of SNPs would have a stronger association with prostate cancer than any individual SNP. For this purpose, we assessed the joint associations of SNPs in the five chromosomal regions with prostate cancer in a large-scale study of Swedish men.

METHODS

STUDY SUBJECTS

The study population has been described in detail elsewhere.[9] Briefly, we conducted a population-based, case–control study in Sweden, called CAPS (Cancer Prostate in Sweden). Subjects with prostate cancer were identified and recruited from four of the six regional cancer registries in Sweden. The inclusion criterion for case subjects was biopsy-confirmed or cytologically verified adenocarcinoma of the prostate, diagnosed between July 2001 and October 2003. Among 3648 identified subjects with prostate cancer, 3161 (87%) agreed to participate. DNA samples from blood, tumor–node–metastasis (TNM) stage, Gleason grade (as determined by biopsy), and levels of prostate-specific antigen (PSA) at diagnosis were available for 2893 subjects (92%). Case subjects were classified as having advanced disease if they met any of the following criteria: a grade 3 or 4 tumor, spread to nearby lymph nodes and metastasis, a Gleason score of 8 or more, or a PSA level of more than 50 ng per milliliter; otherwise, subjects were classified as having localized disease.

Control subjects, who were recruited concurrently with case subjects, were randomly selected from the Swedish Population Registry and matched according to the expected age distribution of cases (groups of 5-year intervals) and geographic region. A total of 2149 of 3153 control subjects (68%) who were invited subsequently agreed to participate in the study. DNA samples from blood were available for 1781 control subjects (83%). Serum PSA levels were measured for all control subjects but were not used as an exclusionary variable. A history of prostate cancer among first-degree relatives was obtained from a questionnaire for both case subjects and control subjects.

Table 1 presents the demographic and clinical characteristics of the study subjects. Recruitment of the study population was completed in two phases, each with a similar number of subjects; the first phase (CAPS-1) ended October 31, 2002, and the second phase (CAPS-2) ended November 1, 2002. Each subject provided written informed consent. The study received institutional approval from the Karolinska Institutet, Umeå University, and Wake Forest University School of Medicine.

SELECTION OF SNPs FOR GENOTYPING

We selected 16 SNPs from five chromosomal regions (three at 8q24 and one each at 17q12 and 17q24.3) that have been reported to be associated with prostate cancer.[6-8,10] Polymerase-chain-reaction (PCR) assays and extension primers for these SNPs were designed with the use of MassARRAY software, version 3.0 (Sequenom). (The primer information is available at www.wfubmc.edu/genomics.) PCR and extension reactions were performed according to the manufacturer's instructions, and extension product sizes were determined by mass spectrometry with the use of the iPLEX system (Sequenom). Duplicate test samples and two water samples (PCR-negative controls), of which the technician was unaware, were included in each 96-well plate. The rate of concordant results between duplicate samples was more than 99%.

STATISTICAL ANALYSIS

Tests for Hardy–Weinberg equilibrium were performed for each SNP separately among case subjects and control subjects with the use of Fisher's exact test. Pairwise linkage disequilibrium was tested for SNPs within each of the five chromosomal regions in control subjects with the use of SAS/Genetics software, version 9.0 (SAS Institute).

Differences in allele frequencies between case subjects and control subjects were tested for each

Table 1. Clinical and Demographic Characteristics of the Subjects.*

Characteristic	Aggressive Disease (N=1231)	Localized Disease (N=1619)	All Case Subjects (N=2893)	Control Subjects (N=1781)
Age — yr				
Mean age	68.0±7.3	65.1±6.7	66.4±7.1	67.2±7.4
Age at diagnosis — no. (%)				
≤65	514 (41.8)	926 (57.2)	1469 (50.8)	NA
>65	717 (58.2)	693 (42.8)	1424 (49.2)	NA
First-degree relative with prostate cancer — no. (%)				
No	1013 (82.3)	1295 (80.0)	2342 (81.0)	1565 (90.6)
Yes	218 (17.7)	324 (20.0)	551 (19.0)	163 (9.4)
Missing data	0	0	0	53
PSA level — no. (%)†				
No. of subjects	1221	1593	2814	1727
≤4.0 ng/ml	36 (2.9)	185 (11.6)	221 (7.9)	1439 (83.3)
4.1–9.9 ng/ml	171 (14.0)	755 (47.4)	926 (32.9)	233 (13.5)
10.0–19.9 ng/ml	216 (17.7)	438 (27.5)	654 (23.2)	38 (2.2)
20.0–49.9 ng/ml	252 (20.6)	215 (13.5)	467 (16.6)	14 (0.8)
50.0–99.9 ng/ml	229 (18.8)	0	229 (8.1)	2 (0.1)
≥100.0 ng/ml	317 (26.0)	0	317 (11.3)	1 (0.1)
Missing data	10	26	79	54
Tumor stage — no. (%)				
No. of subjects	1218	1602	2820	NA
T0	2 (0.2)	7 (0.4)	9 (0.3)	NA
T1	147 (12.1)	933 (58.2)	1080 (38.3)	NA
T2	242 (19.9)	662 (41.3)	904 (32.1)	NA
T3	724 (59.4)	0	724 (25.7)	NA
T4	103 (8.5)	0	103 (3.7)	NA
Could not be assessed	13	17	73	NA
Nodal stage — no. (%)				
No. of subjects	317	302	619	NA
N0	222 (70.0)	302 (100.0)	524 (84.7)	NA
N1	95 (30.0)	0	95 (15.3)	NA
Could not be assessed	914	1317	2274	NA
Metastasis stage — no. (%)				
No. of subjects	863	655	1518	
M0	589 (68.3)	655 (100.0)	1244 (81.9)	NA
M1	274 (31.7)	0	274 (18.1)	NA
Could not be assessed	368	964	1375	NA

Table 1. (Continued.)

Characteristic	Aggressive Disease (N=1231)	Localized Disease (N=1619)	All Case Subjects (N=2893)	Control Subjects (N=1781)
Gleason score for biopsy — no. (%)‡				
No. of subjects	1087	1551	2638	
≤4	9 (0.8)	98 (6.3)	107 (4.1)	NA
5	43 (4.0)	247 (15.9)	290 (11.0)	NA
6	153 (14.1)	832 (53.6)	985 (37.3)	NA
7	414 (38.1)	374 (24.1)	788 (29.9)	NA
8	258 (23.7)	0	258 (9.8)	NA
9	185 (17.0)	0	185 (7.0)	NA
10	25 (2.3)	0	25 (0.9)	NA
Missing data	144	68	255	NA

* Plus–minus values are means ±SD. Because of missing phenotyping results, 43 subjects could not be classified as having either aggressive or localized disease, including 29 subjects who were 65 years of age or younger and 14 subjects who were over the age of 65. NA denotes not applicable.
† Prostate-specific antigen (PSA) levels were obtained at the time of diagnosis for case subjects and at the time of study enrollment for control subjects.
‡ The Gleason score ranges from 2 to 10, with higher scores indicating more aggressive disease.

SNP with the use of a chi-square test with 1 degree of freedom. Allelic odds ratios and 95% confidence intervals were estimated on the basis of a multiplicative model. For genotypes, a series of tests assuming an additive, dominant, or recessive genetic model were performed for each of the five SNPs with the use of unconditional logistic regression with adjustment for age and geographic region; the model that had the highest likelihood was considered to be the best-fitting genetic model for the respective SNP.

We tested the independent effect of each of the five previously implicated regions by including the most significant SNP from each of the five regions in a logistic-regression model with the use of a backward-selection procedure. Multiplicative interactions were tested for each pair of SNPs by including both main effects and an interaction term (a product of two main effects) in a logistic-regression model. We tested the cumulative effects of the five SNPs on prostate cancer by counting the number of genotypes associated with prostate cancer (on the basis of the best-fitting genetic model from single-SNP analysis) for these five SNPs in each subject. The odds ratio for prostate cancer for men carrying any combination of one, two, three, or four or more genotypes associated with prostate cancer was estimated by comparing them with men carrying none of the prostate-cancer–associated genotypes with the use of logistic-regression analysis. We also performed tests for the cumulative effect on prostate-cancer association, which included five SNPs and family history.

Population attributable risk (PAR) was estimated for SNPs that remained significant after adjustment for other covariates with the use of the following equation:

$$\text{PAR\%} = 100 \times \text{p(odds ratio} - 1) \div [\text{p(odds ratio} - 1) + 1].$$

In this equation, p is the prevalence of genotypes associated with prostate cancer among control subjects.[11] The joint PAR was calculated on the basis of the individual PAR of each associated SNP, assuming no multiplicative interaction among the SNPs, with the use of the following equation:

$$1 - [\prod_{i=1}^{5}(1 - PAR_i)].$$

In this equation, PAR_i is the individual PAR for each associated SNP calculated under the full model. For the model that included five SNPs and a family history of prostate cancer, the joint PAR for the associated factors was calculated in a similar manner.

Associations of these five SNPs with TNM stages, aggressiveness of prostate cancer (advanced or localized), and family history (yes or no) were tested only among case subjects with the use of a chi-square test of a 2×K table, in which K is the number of possible categories within each variable. A test for trend was used to assess the proportion of genotypes associated with prostate cancer with each increasing Gleason score, from 4 or less to 10. Associations of SNPs with the mean age at diagnosis were tested only among case subjects with the use of a two-sample t-test. Because serum PSA levels were not normally distributed, a nonparametric analysis (Wilcoxon rank-sum test) was used to assess the association between SNPs and preoperative serum PSA levels in case subjects or PSA levels at the time of sampling in control subjects. All reported P values are based on a two-sided test.

RESULTS

Sixteen SNPs in five chromosomal regions (three at 8q24 and two at 17q), which were previously implicated in harboring genes that confer susceptibility to prostate cancer, were evaluated. In the control group, each SNP was in Hardy–Weinberg equilibrium ($P \geq 0.05$). Significant pairwise linkage disequilibrium ($P<0.05$) was observed for the SNPs within each region.

Table 2 lists allele frequencies of the 16 SNPs among case and control subjects and shows the results of allelic and genotypic tests. Significantly different frequencies ($P<0.05$) between case and control subjects were observed for SNPs in each of the five chromosomal regions. At 17q12, SNP rs4430796 had the strongest association with prostate cancer; the frequency of allele T (SNP rs4430796) was 0.61 in case subjects and 0.56 in control subjects ($P=6.0\times10^{-7}$). Of the four SNPs at 17q24.3, three were associated with prostate cancer, but only rs1859962 had a highly significant association ($P=2.1\times10^{-4}$). The results for 17q12 and 17q24.3 were similar to those that were reported previously.[8] For SNPs at 8q24, significant associations with prostate cancer were found for all SNPs examined across the three independent regions at 8q24. Of the 16 SNPs, 13 remained significant at $P<0.05$ after adjustment for 16 tests with the use of a Bonferroni correction.

Carriers of previously reported risk-associated alleles for SNPs at 17q12, 17q24.3, and 8q24 were significantly more likely to have prostate cancer than were control subjects (Table 2). When various genetic models were tested for SNPs at each region, a recessive model was the best-fitting genetic model for SNPs at 17q12 and 17q24.3, and a dominant model was the best-fitting genetic model for SNPs at regions 1, 2, and 3 of 8q24.

Strong genetic dependence (linkage disequilibrium) among SNPs within each region allowed for a combined analysis in which we were able to select one SNP (the most significant SNP from single SNP analysis) to represent each of the five regions in tests for an independent association with prostate cancer (Table 3). When these five SNPs were included in a multivariate logistic-regression model, each of the five remained significantly associated with prostate cancer after adjustment for other SNPs, and each continued to be highly significant when family history was included in the model. On the basis of adjusted odds ratios for each of these five SNPs and a positive family history, PARs were estimated to account for 4 to 21% of prostate-cancer cases in the Swedish population we studied. The estimated joint PAR for prostate cancer of the five associated SNPs plus family history was 46% in the studied population.

When multiplicative interaction was tested for each possible pair of these five SNPs with the use of an interaction term in logistic regression, none were significant at $P<0.05$. However, the five SNPs appeared to have a cumulative association with prostate cancer, after adjustment for age, geographic region, and family history (Table 4). Men who carried one, two, three, or four or more of the five SNPs had an increasing likelihood of having prostate cancer, as compared with men who did not carry any of the five SNPs (P for trend, 6.75×10^{-27}). When family history was included as another risk factor (coded as 0 or 1) for a total of six possible prostate-cancer associated factors, we observed a stronger cumulative effect after adjustment for age and geographic region (P for trend, 4.78×10^{-28}). For example, men who carried any five or more of these six factors had an odds ratio of 9.46 (95% confidence interval [CI], 3.62 to 24.72) for prostate cancer, as compared with men who carried none of the six factors ($P=1.29\times10^{-8}$). This cumulative effect was similarly observed in two subgroups of study subjects, with a P for trend of 1.36×10^{-10} in CAPS-1 and of 9.03×10^{-20} in CAPS-2 (data not shown).

We calculated the specificity and sensitivity of the regression model by constructing receiver-

Table 2. Association of SNPs at Five Chromosomal Regions with Prostate Cancer.*

SNP	Chromosomal Region	Position†	Alternative Alleles	Allelic Tests: Associated Allele§	Allelic Tests: Frequency, case subjects	Allelic Tests: Frequency, control subjects	Allelic Tests: Odds Ratio (95% CI)¶	Allelic Tests: P Value	Best-Fitting Genetic Model‡: Model	Best-Fitting Genetic Model‡: Genotype‖ reference	Best-Fitting Genetic Model‡: Genotype‖ associated	Best-Fitting Genetic Model‡: Odds Ratio (95% CI)	Best-Fitting Genetic Model‡: P Value**
rs4430796	17q12	33,172,153	T, C	T	0.61	0.56	1.24 (1.14–1.36)	6.0×10^{-7}	Recessive	CC or TC	TT	1.40 (1.23–1.59)	2.68×10^{-7}
rs7501939	17q12	33,175,269	G, A	G	0.66	0.62	1.22 (1.12–1.33)	9.0×10^{-6}	Recessive	AA or GA	GG	1.33 (1.17–1.50)	5.54×10^{-6}
rs3760511	17q12	33,180,426	A, C	C	0.41	0.38	1.17 (1.07–1.27)	5.0×10^{-4}	Recessive	AA or CA	CC	1.42 (1.20–1.68)	4.47×10^{-5}
rs1859962	17q24.3	66,620,348	G, T	G	0.54	0.50	1.17 (1.08–1.28)	2.1×10^{-4}	Recessive	GT or TT	GG	1.28 (1.12–1.46)	3.54×10^{-4}
rs7214479	17q24.3	66,702,544	C, T	T	0.50	0.48	1.08 (0.99–1.18)	0.07	Recessive	CC or CT	TT	1.15 (1.00–1.32)	0.06
rs6501455	17q24.3	66,713,406	A, G	A	0.56	0.54	1.09 (1.00–1.19)	0.05	Recessive	AG or GG	AA	1.13 (0.99–1.29)	0.06
rs983085	17q24.3	66,723,656	A, G	A	0.57	0.55	1.07 (0.98–1.16)	0.13	Recessive	GA or GG	AA	1.11 (0.97–1.26)	0.12
rs6983561	8q24 (region 2)	128,176,062	A, C	C	0.06	0.03	1.65 (1.33–2.05)	4.2×10^{-6}	Dominant	AA	CA or CC	1.60 (1.28–2.00)	2.14×10^{-5}
rs16901979	8q24 (region 2)	128,194,098	C, A	A	0.06	0.03	1.65 (1.33–2.05)	4.3×10^{-6}	Dominant	CC	AA or CA	1.60 (1.28–2.01)	2.14×10^{-5}
rs6983267	8q24 (region 3)	128,482,487	G, T	G	0.56	0.51	1.22 (1.12–1.33)	3.9×10^{-6}	Dominant	TT	GT or GG	1.38 (1.19–1.59)	1.74×10^{-5}
rs7000448	8q24 (region 3)	128,510,352	C, T	T	0.43	0.40	1.15 (1.06–1.25)	1.4×10^{-3}	Dominant	CC	CT or TT	1.18 (1.04–1.33)	1.21×10^{-2}
rs1447295	8q24 (region 1)	128,554,220	C, A	A	0.17	0.14	1.21 (1.07–1.36)	1.6×10^{-3}	Dominant	CC	CA or AA	1.26 (1.10–1.44)	8.27×10^{-4}
rs4242382	8q24 (region 1)	128,586,755	G, A	A	0.16	0.14	1.24 (1.10–1.39)	5.3×10^{-4}	Dominant	GG	AG or AA	1.29 (1.12–1.47)	2.53×10^{-4}
rs7017300	8q24 (region 1)	128,594,450	A, C	C	0.20	0.18	1.15 (1.03–1.28)	0.01	Dominant	AA	CA or CC	1.20 (1.05–1.36)	6.20×10^{-3}
rs10090154	8q24 (region 1)	128,601,319	C, T	T	0.16	0.13	1.26 (1.11–1.42)	2.0×10^{-4}	Dominant	CC	CT or TT	1.31 (1.14–1.50)	1.03×10^{-4}
rs7837688	8q24 (region 1)	128,608,542	G, T	T	0.15	0.13	1.17 (1.04–1.13)	9.6×10^{-3}	Dominant	GG	GT or TT	1.21 (1.06–1.39)	5.87×10^{-3}

* CI denotes confidence interval, and SNP single-nucleotide polymorphism.
† The position is based on the National Center for Biotechnology Information database, build 35.
‡ The best-fitting model for each SNP was determined after testing associations of a series of genetic models, including dominant and recessive models, with prostate cancer.
§ These alleles were reported to be associated with prostate cancer in studies published previously.[4-8,10]
¶ Allelic odds ratios are based on the multiplicative model.
‖ Reference genotypes and those associated with prostate cancer for each SNP were defined on the basis of the best-fitting genetic model.
** P values are two-sided and were calculated by the likelihood-ratio test with one degree of freedom, adjusted for age and geographic region.

Table 3. Adjusted Odds Ratios and Population Attributable Risks (PARs) for Representative SNPs at Five Chromosomal Regions and Family History.*

Variable or SNP†	Chromosomal Region	Alternative Alleles	Reference		Frequency of Associated Factors‡		Regression Coefficient	Odds Ratio (95% CI)	P Value§	PAR
					Case Subjects	Control Subjects				%
Age							0.01	1.01 (1.00–1.02)	0.02	
Geographic region							−0.77	0.46 (0.39–0.54)	<0.001	
Family history			No	Yes	0.19	0.09	0.80	2.22 (1.83–2.68)	1.15×10^{-17}	9.89
rs4430796	17q12	T, C	CC/TC	TT	0.38	0.30	0.32	1.38 (1.21–1.57)	1.62×10^{-6}	10.23
rs1859962	17q24.3	G, T	GT/TT	GG	0.30	0.25	0.24	1.28 (1.11–1.47)	5.49×10^{-4}	6.54
rs16901979	8q24 (region 2)	C, A	CC	AA/CA	0.10	0.07	0.42	1.53 (1.22–1.92)	1.83×10^{-4}	3.58
rs6983267	8q24 (region 3)	G, T	TT	GT/GG	0.82	0.77	0.32	1.37 (1.18–1.59)	3.44×10^{-5}	22.17
rs1447295	8q24 (region 1)	C, A	CC	CA/AA	0.31	0.26	0.19	1.22 (1.06–1.40)	5.31×10^{-3}	5.41
All five SNPs										40.45
All five SNPs and family history										46.34

* CI denotes confidence interval, PAR population attributable risk, and SNP single-nucleotide polymorphism.
† A family history of prostate cancer and five SNPs were included in the multivariate logistic-regression model with adjustment for age and geographic region.
‡ For SNPs, the reference genotype and those associated with prostate cancer at each SNP were determined on the basis of the best-fitting model after testing associations of a series of genetic models with prostate cancer.
§ P values were calculated by the likelihood-ratio test.

operating-characteristic (ROC) curves and calculated statistics for the area under the curve (AUC) to estimate the ability of each of three models to distinguish case subjects from control subjects. The AUC was 57.7 (95% CI, 56.0 to 59.3) for model 1 (age and region alone), 60.8 (95% CI, 59.1 to 62.4) for model 2 (age, region, and family history), and 63.3 (95% CI, 61.7 to 65.0) for model 3 (age, region, family history, and the number of genotypes associated with prostate cancer at the five SNPs). The AUC was significantly higher for model 3 than for model 2 ($P=6.12\times10^{-6}$). It is important to note that overfitting could have influenced our results, and for this reason the models require verification in independent populations.

Table 5 shows that none of the five SNPs were significantly associated with the aggressiveness of prostate cancer, the Gleason score, the presence or absence of family history, the serum PSA level at diagnosis, or the age at diagnosis. Furthermore, no associations with these clinical variables were found when multiple SNPs associated with prostate cancer were considered simultaneously. For example, the 154 case subjects who carried four or more of the five SNPs were not significantly different from the 162 case subjects who had none of the SNPs with regard to the following clinical variables: positive family history (17% with four or more SNPs and 21% with no SNPs, P=0.39), the proportion with advanced disease (54% and 48%, respectively; P=0.33), and the median serum PSA level at diagnosis (15 ng and 14 ng per milliliter, respectively; P=0.27). A lack of association between the SNPs at 8q24 and clinical characteristics was also reported previously,[7,12-14] but in other studies a trend was found between 8q24 SNPs and a high Gleason grade, tumor stage, and aggressive disease.[4-6,15,16] Thus, the association of these SNPs with clinical features of prostate cancer remains an open question.

DISCUSSION

In genomewide studies, multiple chromosomal regions at 8q24 and 17q have been associated with prostate cancer.[4-8] All three regions at 8q24 have been replicated in all published studies,[10,12-16] but no study has yet replicated the associations in regions at 17q. The highly significant findings at 17q12 and 17q24.3 in our study independently confirm the association of these two regions with prostate cancer. In addition, we confirmed the association of SNPs at regions 1, 2, and 3 of 8q24 with prostate cancer. This independent confirmation of the association of these five chromosomal regions with prostate cancer supports the validity of genetic association studies in complex diseases.

Although each of the SNPs in the five chromosomal regions was only moderately associated with prostate cancer, we found that they had a strong cumulative association with the disease. We estimated that men who have five or more of the six factors associated with prostate cancer (specific genotypes at five SNPs and a positive family history for the disease) have an odds ratio of 9.46 for prostate cancer. The cumulative effect is highly significant in our overall study sample (P for trend, 4.78×10^{-28}) and consistent between the two subgroups in CAPS-1 and CAPS-2. It may be possible to use the combined information from the five SNPs and family history to assess an individual

Table 4. Cumulative Effect of Associated Factors on the Risk of Prostate Cancer.*

Variable	Case Subjects	Control Subjects	Regression Coefficient	Odds Ratio (95% CI)	P Value†	P Value for Trend‡
	no. of subjects (%)					
Genotypes at five SNPs§						
Age			0.01	1.01 (1.00–1.02)	0.02	
Geographic region			−0.76	0.46 (0.40–0.55)	<0.001	
Family history			0.8	2.22 (1.83–2.68)	7.73×10^{-18}	
No. of associated genotypes¶						
0	162 (5.6)	173 (10.1)	NA	1.00		
1	883 (30.8)	631 (36.8)	0.41	1.50 (1.18–1.92)	9.46×10^{-4}	
2	1123 (39.1)	618 (36.0)	0.67	1.96 (1.54–2.49)	4.19×10^{-8}	
3	548 (19.1)	255 (14.9)	0.79	2.21 (1.70–2.89)	4.33×10^{-9}	
≥4	154 (5.4)	38 (2.2)	1.5	4.47 (2.93–6.80)	1.20×10^{-13}	6.75×10^{-27}
Genotypes at five SNPs and family history‖						
Age			0.01	1.01 (1.00–1.02)	0.02	
Geographic region			−0.75	0.47 (0.40–0.55)	<0.001	
No. of associated factors**						
0	144 (5.0)	174 (10.1)	NA	1.00		
1	778 (26.9)	581 (33.6)	0.48	1.62 (1.27–2.08)	1.27×10^{-4}	
2	1053 (36.4)	622 (36.0)	0.73	2.07 (1.62–2.64)	5.86×10^{-9}	
3	642 (22.2)	286 (16.6)	0.99	2.71 (2.08–3.53)	9.54×10^{-14}	
4	236 (8.2)	60 (3.5)	1.56	4.76 (3.31–6.84)	9.17×10^{-19}	
≥5	40 (1.4)	5 (0.3)	2.24	9.46 (3.62–24.72)	1.29×10^{-8}	4.78×10^{-28}

* All comparisons are of case subjects with control subjects. CI denotes confidence interval, NA not applicable, and SNP single-nucleotide polymorphism.
† P values are two-sided and were calculated by the likelihood-ratio test.
‡ P values were calculated by the Cochran–Armitage test for trend.
§ Testing for the cumulative effect of five SNPs (rs4430796, rs1859962, rs16901979, rs6983267, and rs1447295) was adjusted for age, geographic region, and family history.
¶ Listed are the number of genotypes associated with prostate cancer at the five SNPs for 2870 case subjects and 1715 control subjects.
‖ Testing for cumulative effect of the five SNPs plus family history was adjusted for age and geographic region.
** Listed are the number of factors associated with prostate cancer (the five SNPs plus family history) for 2893 case subjects and 1728 control subjects.

Table 5. Association of Five SNPs with Clinical Characteristics.*

Variable	rs4430796 (17q12)			rs1859962 (17q24.3)			rs16901979 (8q24)			rs6983267 (8q24)			rs1447295 (8q24)		
	No. of Subjects	Reference	Associated	No. of Subjects	Reference	Associated	No. of Subjects	Reference	Associated	No. of Subjects	Reference	Associated	No. of Subjects	Reference	Associated
		CC/TC	TT		GT/TT	GG		CC	AA/CA		TT	GT/GG		CC	CA/AA
Aggressiveness of disease — no. (%)															
Localized	1608	1021 (63.5)	587 (36.5)	1604	1113 (69.4)	491 (30.6)	1603	1446 (90.2)	157 (9.8)	1597	294 (18.4)	1303 (81.6)	1603	1130 (70.5)	473 (29.5)
Aggressive	1217	748 (61.5)	469 (38.5)	1215	860 (70.8)	355 (29.2)	1214	1077 (88.7)	137 (11.3)	1213	243 (20.0)	970 (80.0)	1214	838 (69.0)	376 (31.0)
P value†			0.27			0.42			0.20			0.28			0.40
Gleason score — no. (%)‡															
≤4	105	69 (65.7)	36 (34.3)	105	69 (65.7)	36 (34.3)	105	96 (91.4)	9 (8.6)	105	22 (21.0)	83 (79.0)	105	80 (76.2)	25 (23.8)
5	289	182 (63.0)	107 (37.0)	288	200 (69.4)	88 (30.6)	286	256 (89.5)	30 (10.5)	287	61 (21.3)	226 (78.7)	287	198 (69.0)	89 (31.0)
6	978	619 (63.3)	359 (36.7)	976	675 (69.2)	301 (30.8)	976	881 (90.3)	95 (9.7)	972	170 (17.5)	802 (82.5)	977	697 (71.3)	280 (28.7)
7	781	497 (63.6)	284 (36.4)	778	554 (71.2)	224 (28.8)	778	701 (90.1)	77 (9.9)	776	161 (20.7)	615 (79.3)	776	536 (69.1)	240 (30.9)
8	255	152 (59.6)	103 (40.4)	255	184 (72.2)	71 (27.8)	255	215 (84.3)	40 (15.7)	254	47 (18.5)	207 (81.5)	255	179 (70.2)	76 (29.8)
9	184	106 (57.6)	78 (42.4)	184	126 (68.5)	58 (31.5)	184	165 (89.7)	19 (10.3)	184	32 (17.4)	152 (82.6)	184	128 (69.6)	56 (30.4)
10	25	13 (52.0)	12 (48.0)	25	19 (76.0)	6 (24.0)	25	24 (96.0)	1 (4.0)	25	8 (32.0)	17 (68.0)	25	18 (72.0)	7 (28.0)
P value§			0.08			0.30			0.28			0.97			0.43
Family history in first-degree relative — no. (%)															
No	2324	1466 (63.1)	858 (36.9)	2318	1623 (70.0)	695 (30.0)	2317	2066 (89.2)	251 (10.8)	2313	451 (19.5)	1862 (80.5)	2317	1628 (70.3)	689 (29.7)
Yes	544	331 (60.8)	213 (39.2)	544	380 (69.9)	164 (30.1)	543	491 (90.4)	52 (9.6)	540	94 (17.4)	446 (82.6)	543	370 (68.1)	173 (31.9)
P value†			0.33			0.94			0.39			0.27			0.33
PSA level at diagnosis															
Median — ng/ml		12.0	13.0		13.0	11.9		12.0	14.5		12.0	12.0		12.0	13.0
P value¶			0.83			0.66			0.16			0.17			0.07
Age at diagnosis															
Mean — yr		65.9	65.7		65.9	65.6		65.8	65.8		65.8	65.8		65.9	65.7
P value‖			0.63			0.22			0.91			0.90			0.54

* Reference genotypes or those associated with prostate cancer were determined on the basis of the best-fitting model at each single-nucleotide polymorphism (SNP). Percentages are the proportions of subjects with each genotype in the category. PSA denotes prostate-specific antigen.
† P values are two-sided and were calculated by Pearson's chi-square test.
‡ The Gleason score ranges from 2 to 10, with higher scores indicating more aggressive disease.
§ P values are two-sided and were calculated by the Cochran–Armitage test for trend.
¶ P values are two-sided and were calculated by the Wilcoxon rank-sum test.
‖ P values are two-sided and were calculated by the two-sample t-test.

patient's risk of prostate cancer, but this strategy will have to be tested in a prospective study before proceeding with any such risk assessments.

We found that the presence of the five prostate-cancer–associated SNPs was independent of PSA levels in both case subjects (Table 5) and control subjects (data not shown), which suggests that some men with low PSA levels may have an increased risk of prostate cancer if they carry one or more of the prostate-cancer–associated genotypes described here. However, this proposition also requires testing in a prospective trial, particularly one that uses PSA in combination with the associated SNPs and family history.

We do not know the mechanism by which the SNPs we analyzed could affect the risk of prostate cancer. Other than SNP rs4430796, which is located within the *TCF2* gene, the specific genes that are affected by the rest of the SNPs have not been identified. Since the five SNPs in our study appear to be associated with a risk of prostate cancer in general, rather than with a more or less aggressive form, we suspect that the genetic variants act at an early stage of carcinogenesis.

Our study is only a first step toward defining a genetic association with prostate cancer in populations. Future investigations will need to test the value of these findings in assessing the risk of prostate cancer in individual men.

Supported by grants (CA105055, CA106523, and CA95052, to Dr. Xu, and CA112517 and CA58236, to Dr. Isaacs) from the National Cancer Institute; a grant (PC051264, to Dr. Xu) from the Department of Defense; grants (to Dr. Grönberg) from the Swedish Cancer Society and the Swedish Academy of Sciences; an endowment from William T. Gerrard, Mario A. Duhon, and John and Jennifer Chalsty (to Dr. Isaacs); and a David H. Koch award (to Dr. Isaacs) from the Prostate Cancer Foundation.

A patent application has been filed by the Wake Forest University School of Medicine, Johns Hopkins University School of Medicine, and Dr. Henrik Grönberg at Karolinska Institutet, Stockholm, to preserve patent rights for the technology and results described in this study. No other potential conflict of interest relevant to this article was reported.

We thank all the study subjects who participated in the CAPS study and urologists who included their patients in the CAPS study, the Regional Cancer Registries, and the CAPS steering committee, including Drs. Jan Adolfsson, Jan-Erik Johansson, and Eberhart Varenhorst.

REFERENCES

1. Hunter DJ, Kraft P. Drinking from the fire hose — statistical issues in genomewide association studies. N Engl J Med 2007;357:436-9.
2. Grönberg H. Prostate cancer epidemiology. Lancet 2003;361:859-64.
3. Johns LE, Houlston RS. A systematic review and meta-analysis of familial prostate cancer risk. BJU Int 2003;91:789-94.
4. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006;38:652-8.
5. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39:631-7.
6. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 2007;39:638-44.
7. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007;39:645-9.
8. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007;39:977-83.
9. Lindström S, Wiklund F, Adami HO, Bälter KA, Adolfsson J, Grönberg H. Germline genetic variation in the key androgen-regulating genes androgen receptor, cytochrome P450, and steroid-5-alpha-reductase type 2 is important for prostate cancer development. Cancer Res 2006;66:11077-83.
10. Zheng SL, Sun J, Cheng Y, et al. Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. J Natl Cancer Inst 2007;99:1525-33.
11. Linkage disequilibrium. In: Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology. 2nd ed. New York: Oxford University Press, 1980:301-11.
12. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc Natl Acad Sci U S A 2006;103:14068-73.
13. Severi G, Hayes VM, Padilla EJ, et al. The common variant rs1447295 on chromosome 8q24 and prostate cancer risk: results from an Australian population-based case-control study. Cancer Epidemiol Biomarkers Prev 2007;16:610-2.
14. Schumacher FR, Feigelson HS, Cox DG, et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. Cancer Res 2007;67:2951-6.
15. Wang L, McDonnell SK, Slusser JP, et al. Two common chromosome 8q24 variants are associated with increased risk for prostate cancer. Cancer Res 2007;67:2944-50.
16. Suuriniemi M, Agalliu I, Schaid DJ, et al. Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24. Cancer Epidemiol Biomarkers Prev 2007;16:809-14.