Published in final edited form as:

Cancer Res. 2009 April 1; 69(7): 2720–2723. doi:10.1158/0008-5472.CAN-08-3347.

A novel prostate cancer susceptibility locus at 19q13

Fang-Chi Hsu^{1,2,3}, Jielin Sun^{1,2}, Fredrik Wiklund⁴, Sarah D. Isaacs⁵, Kathleen E. Wiley⁵, Lina D. Purcell^{1,2}, Zhengrong Gao^{1,2}, Pär Stattin⁶, Yi Zhu^{1,2}, Seong-Tae Kim^{1,2}, Zheng Zhang^{1,2}, Wennuan Liu^{1,2}, Bao-Li Chang^{1,2}, Patrick C. Walsh⁵, David Duggan⁷, John D. Carpten⁷, William B. Isaacs^{5,†}, Henrik Grönberg⁴, Jianfeng Xu^{1,2,†}, and S. Lilly Zheng^{1,2}

¹Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, NC

²Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC

³Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC

⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

⁵Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD

⁶Department of Surgical and Perioperative sciences, Urology and Andrology, Umeå University Hospital, Umeå, Sweden

⁷Translational Genomics Research Institute, Phoenix, AZ

Abstract

A two-stage genome-wide association study (GWAS) of the Cancer Genetic Markers of Susceptibility (CGEMS) initiative identified SNPs in 150 regions across the genome that may be associated with prostate cancer (PCa) risk. We filtered these results to identify 43 independent single nucleotide polymorphisms (SNPs) where the frequency of the risk allele was consistently higher in cases than in controls in each of the five CGEMS study populations. Genotype information for 22 of these 43 SNPs was obtained either directly by genotyping or indirectly by imputation in our PCa GWAS of 500 cases and 500 controls selected from a population-based case-control study in Sweden (CAPS). Two of these 22 SNPs were significantly associated with PCa risk ($P < 0.05$). We then genotyped these two SNPs in the remaining cases ($N = 2,393$) and controls ($N = 1,222$) from CAPS and found rs887391 at 19q13 was highly associated with PCa risk ($P = 9.4 \times 10^{-4}$). A similar trend of association was found for this SNP in a case-control study from Johns Hopkins Hospital, albeit the result was not statistically significant. Altogether, the frequency of the risk allele of rs887391 was consistently higher in cases than controls among each of seven study populations examined, with an overall $P = 3.2 \times 10^{-7}$ from a combined allelic test. A fine mapping study in a 110 Kb region at 19q13 among CAPS and JHH study populations revealed rs887391 was the most strongly associated SNP in the region. Additional confirmation studies of this region are warranted.

Keywords

prostate cancer; association; genetic; 19q13

†Address for correspondence: Dr. William B. Isaacs, Ph.D., Marburg 115, Johns Hopkins Hospital, 600 N. Wolfe Street, Baltimore, MD 21287, Phone: (410) 955-2518, Fax: (410) 955-0833, wisaacs@jhmi.edu.

Disclosure of Potential Conflicts of Interest

The authors declare that they have no potential conflicts of interest.

Introduction

GWAS has been an effective tool to identify genetic variants associated with disease risk without any presumption about their location or function. More than a dozen PCa risk associated variants have been identified from GWAS and consistently replicated in multiple independent study populations (1–10). These newly discovered PCa risk associated variants may provide novel insight into disease etiology. It is anticipated that results from GWAS will lead to better prediction of PCa risk for early detection and better understanding of the molecular mechanisms of this disease.

Using a two-stage design GWAS among a total of 5,113 PCa patients and 5,121 control subjects from five study populations, the CGEMS study identified 150 distinct regions that were potentially associated with PCa risk ($P < 10^{-3}$) (8). Among these 150 regions, five reached genome-wide significance ($P < 10^{-8}$), including two at 8q24 and one each at 17q12, 10q11, 11q13. The associations at these five regions have been reported in other GWAS (1–5,9). Two additional regions did not reach genome-wide significance but were highly significant, including 10q26 ($P = 10^{-7}$) and 7p15 ($P = 10^{-6}$). For these seven regions, risk alleles of SNPs were consistently more common in cases than controls among all five study populations. In the current study, we examined SNPs in the remaining 143 regions and found 43 SNPs had this same consistency. We then sequentially examined these 43 SNPs in two additional study populations from Sweden and the U.S.

Materials and Methods

Study subjects

The CAncer of the Prostate in Sweden (CAPS) study has been described in detail (11), including 2,899 cases and 1,722 controls. Case subjects were classified as having aggressive disease if they met any of the following criteria: T3/4, N+, M+, Gleason score sum ≥ 8 , or PSA > 50 ng/ml; otherwise, they were classified as having non-aggressive disease (Supplementary Table 1a). We selected 500 aggressive PCa cases and 500 controls matching the age distribution of cases for a GWAS (6). The sample size for the GWAS was determined based on available funds and statistical power; we had 80% power at a genome-wide significance level ($P < 2.5 \times 10^{-8}$) to detect a risk allele with OR ≥ 1.9 and minor allele frequency (MAF) ≥ 0.2 . No evidence for potential population stratification in the GWAS samples was observed using the D statistic of the Kolmogorov-Smirnov test (6). The study received institutional approval at the Karolinska Institutet, Umeå University.

The Johns Hopkins Hospital (JHH) study population was described in detail elsewhere (12–14), including 1,527 cases and 482 controls of European descent (by self report). Tumors with a Gleason score of 7 or higher or stage pT3 or higher or N+ or M1 (i.e., either high-grade or non-organ-confined disease) were defined as more aggressive (Supplementary Table 1b). The study received institutional approval.

We also utilized the published data from the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS) study (4,8). Summary genotype information from the five study populations was included in this study. The five study populations are the Prostate, Lung, Colon and Ovarian (PLCO) Cancer Screening Trial, American Cancer Society Cancer Prevention Study II (CPS-II); the Health Professionals Follow-up Study (HPFS); CeRePP French Prostate Case-Control Study (FPCC); and Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study (ATBC).

Genotyping

Methods for the genome-wide association study in 500 cases and 500 controls were described in detail elsewhere (6). The average genotyping call rate (i.e., the number of SNPs being called by BRLMM algorithm/total number of SNPs) was 99.1%. Genotype concordance for the duplicated samples was > 99%. We found that 260,852 SNPs (53.23%) met the quality control criteria of minor allele frequency (MAF) ≥ 0.01 , HWE $> 10^{-4}$ in controls, and genotyping call rate > 95% in cases and controls. These SNPs were selected for further analysis and imputation.

For confirmation and fine mapping studies, SNPs were genotyped using iPLEX (Sequenom, Inc). The primer information is available upon request. The rate of concordant results between 100 duplicate samples was >99%.

Statistical methods

Tests for Hardy-Weinberg equilibrium were performed for each SNP separately among case patients and control subjects using Fisher's exact test. Haplotype blocks were estimated using a computer program Haploview (15), and a default Gabriel method (16) was used to define each haplotype block.

We imputed all of the known SNPs in the 110 Kb-region of interest at 19q13 based on the genotyped SNPs and haplotype information in the HapMap Phase II data (CEU) using a computer program, IMPUTE (17). A posterior probability of 0.9 was used as a threshold to call genotypes. Imputed SNPs (N = 32) that had a call rate higher than 90% in both CAPS and JHH were included in the following analysis.

Allele frequency differences between case patients and control subjects were tested for each SNP using a chi-square test with 1 degree of freedom. Allelic odds ratio (OR) and 95% confidence interval (95% CI) were estimated based on a multiplicative model.

Associations of SNP rs887391 with aggressiveness of PCa (advanced or localized), Gleason score (≤ 6 , 7, or ≥ 8), and family history (yes or no) were tested only among case subjects with the use of a chi-square test of a $3 \times K$ table with $2 \times (K-1)$ degrees of freedom, in which K is the number of possible categories within each variable. Serum PSA level was log-transformed in order to approximate the distributional assumption. A test for trend was used to assess the association between log-PSA level and the number of risk allele carriers (0, 1, and 2) using the linear regression model. Association of SNP rs887391 with the mean age at diagnosis was tested only among case subjects with the use of analysis of variance (ANOVA).

Results and Discussion

We filtered the SNPs in 150 regions identified from the two-stage GWAS of the CGEMS study using criteria that the direction of association be consistent among all five CGEMS study populations, i.e., the frequency of the risk allele was higher in cases than in controls in each of these study populations (Supplementary Figure 1 and Supplementary Table 2). This resulted in the identification of 43 SNPs for further study. As a first-stage confirmation, we crosschecked these 43 SNPs in an independent GWAS performed in the CAPS population. High quality genotyping data were available for 22 of these 43 SNPs (Supplementary Table 2, **top panel**), including two SNPs that were directly genotyped in the Affymetrix 500K SNP arrays and 20 SNPs that could be successfully imputed, with a missing call rate <10%. Two imputed SNPs from two distinct regions were significantly associated with PCa risk using a Chi-square test: rs887391 at 19q13 (nominal $P=0.03$) and rs6922172 at 6p12 (nominal $P=0.04$). The direction of association in both SNPs was consistent with that of the five CGEMS study populations.

As a second-stage confirmation, we genotyped these two SNPs in the remaining CAPS study subjects, including 2,393 PCa patients and 1,222 control subjects. SNP rs6922172 at 6p12 was not significant ($P=0.23$). However, a highly significant association was found for the SNP rs887391 at 19q13. The frequency of risk allele 'T' was significantly higher in cases (0.76) than in controls (0.73), $P=9.4 \times 10^{-4}$. As a third-stage confirmation, we genotyped this SNP in 1,527 PCa patients and 482 control subjects of European descent from Johns Hopkins Hospital (JHH). The risk allele 'T' was more common in cases (0.79) than controls (0.78), although the difference was not significant ($P=0.43$). Combining all the available data from the CAPS, JHH, and five populations of the CGEMS study using a Mantel-Haenszel method, the overall P -value of the allelic test was 3.2×10^{-7} for the SNP (Table 1). This P -value almost reached genome-wide significance of 9.5×10^{-8} for 5% Type I error of all tested SNPs in the genome. The odds ratio (OR) for allele 'T' was estimated to be 1.15, with a 95% confidence interval (CI) of 1.09–1.21. Notably, although the risk alleles were consistently higher in cases than controls in all examined populations, the difference was not significant in several individual study populations, likely due to limited statistical power to detect risk SNPs with moderate effect in a small study. Our study demonstrated the advantage of combining information from several small studies to detect such risk SNPs.

We next performed a fine mapping study in CAPS and JHH to assess associations of other SNPs at 19q13 with PCa risk. A 110 Kb region (46,630,000–46,740,000 bp) was identified based on the CAPS GWAS where SNPs with $P<0.05$ were aggregated. We selected 14 tagging SNPs to cover the fine mapping region based on HapMap Phase II data. These SNPs were genotyped among all CAPS and JHH study subjects. We also imputed 32 SNPs based on the HapMap Phase II data (CEU) (17). Allele frequency differences between cases and controls were tested using a chi-square test for these 46 SNPs in CAPS and JHH (Supplementary Table 3). A combined test was performed for each SNP using a Mantel-Haenszel method (Figure 1a). SNP rs887391 was the strongest PCa risk associated SNP in this region. SNPs associated with PCa risk at $P<0.01$ spanned ~62 kb, from 46,677,427–46,739,764, and were located in four haplotype blocks (16) (Figure 1b). A spliced transcript (DA869846) found in multiple cDNA libraries prepared from various tissues including the prostate is within the region (18).

SNP rs887391 is about 10 Mb centromeric to the *PSA* gene (*KLK3*) where a SNP near the 3' end (rs2735839) was reportedly associated with PCa risk (9). However, because the SNP rs2735839 was significantly associated with higher PSA levels in subjects without PCa (9), there was concern that the PCa association was confounded by PSA screening (19). Therefore, we tested the association of rs887391 with plasma PSA levels among 1,722 control subjects in CAPS. The mean PSA levels were 1.48, 1.55, and 1.57 ng/mL for men who had 0, 1, or 2 copies of the 'T' allele, respectively. The difference was not statistically significant assuming an additive model, $P=0.6$. The PCa association for rs887391 at 19q13 observed in this study is unlikely to be confounded by PSA screening.

We also tested the association of rs887391 with disease aggressiveness, Gleason score, family history, PSA at diagnosis, and age at diagnosis. No significant association was found (Supplementary Table 4). This finding is similar to most PCa risk variants identified from GWAS, where no association with clinical characteristics was found, including the SNPs at 8q24, 17q12, 17q24, 10q11, and 11q13. This observation, however, is not surprising because these SNPs were identified by comparing all PCa cases with controls. Study designs such as case-case studies may be needed to identify associations with aggressive PCa.

In summary, this three-stage confirmation study in CAPS and JHH identified a novel locus at 19q13 that is potentially associated with PCa risk. Because the statistical evidence did not reach genome-wide significance level, it could represent a chance finding and should be considered as suggestive. It is also important to note that our study is underpowered to evaluate many of

the 43 regions implicated in the CGEMS study because of the small sample size of our CAPS GWAS, the limited number of SNPs that we were able to examine, and the reliance on imputed SNPs for most of the SNPs examined. Additional studies are needed to further confirm the candidate regions discovered by the CGEMS study.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The study is supported by National Cancer Institute CA129684, CA105055, CA106523, and CA95052 to J.X., CA112517 and CA58236 to W.B.L., Swedish Cancer Society and Swedish Academy of Sciences to H.G.

We acknowledge the contribution of multiple physicians and researchers in designing and recruiting study subjects, including Dr. Hans-Olov Adami (for CAPS) and Drs. Bruce J. Trock, and Alan W. Partin (for JHH).

The authors also thanks for the CGEMS for making the data available publicly.

References

1. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. *Nat Genet* 2006;38:652–658. [PubMed: 16682969]
2. 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 USA* 2006;103:14068–14073. [PubMed: 16945910]
3. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007;39:631–637. [PubMed: 17401366]
4. 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;39:645–649. [PubMed: 17401363]
5. 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;39:977–983. [PubMed: 17603485]
6. Duggan D, Zheng SL, Knowlton M, et al. Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. *J Natl Cancer Inst* 2007;99:1836–1844. [PubMed: 18073375]
7. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 2007;39:638–644. [PubMed: 17401364]
8. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet* 2008;40:310–315. [PubMed: 18264096]
9. Eeles RA, Kote-Jarai Z, Giles GG, et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008;40:316–321. [PubMed: 18264097]
10. Gudmundsson J, Sulem P, Rafnar T, et al. Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* 2008;40:281–283. [PubMed: 18264098]
11. Zheng SL, Sun J, Wiklund F, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008;358:910–919. [PubMed: 18199855]
12. Zheng SL, Sun J, Cheng Y, et al. Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. *JNCI* 2007;99:1525–1533. [PubMed: 17925536]
13. Epstein, JI.; Allsbrook, WC., Jr; Amin, MB.; Egevad, LL. ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma; *Am J Surg Pathol*; 2005. p. 1228-1242.
14. Hoedemaeker RF, Vis AN, Van Der Kwast TH. Staging prostate cancer. *Microsc Res Tech* 2000;51:423–429. [PubMed: 11074612]
15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–265. [PubMed: 15297300]

16. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229. [PubMed: 12029063]
17. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;39:906–913. [PubMed: 17572673]
18. Kimura K, Wakamatsu A, Suzuki Y, et al. Diversification of Transcriptional Modulation: Large-scale Identification and Characterization of Putative Alternative Promoters of Human Genes. *Genome Res* 2006;16:55–65. [PubMed: 16344560]
19. Ahn J, Berndt SI, Wacholder S, et al. Variation in KLK genes, prostate-specific antigen and risk of prostate cancer. *Nat Genet* 2008;40:1032–1034. [PubMed: 19165914]

Figure 1. A schematic view of genetic association between SNPs at 19q13 and prostate cancer risk (a) Combined allele tests for 46 SNPs at 19q13 and prostate cancer risk among a total of 4,456 prostate cancer cases and 2,357 controls from a population-based case-control study in CAPS and JHH. SNPs that were directly genotyped are indicated by a solid square and SNPs that were imputed are indicated by a cross. (b) Inferred haplotype blocks of these 46 SNPs were estimated from the control subjects in CAPS and JHH using the Haploview computer program¹³. The color of each square represents the pair-wise r^2 ; the darker, the stronger r^2 , with pure black representing $r^2=1$ and pure white representing $r^2=0$.

Table 1
Association of prostate cancer risk with SNP rs887391 (per study and combined)

Risk	Genotype counts				Allele frequency		Allelic OR					
	Allele	Cases		Controls		Cases	Controls	P [*]	OR	95% CI		
rs887391		CC	TC	TT	CC	TC	TT			T vs C		
CAPS	T	181	1024	1688	138	676	904	0.76	0.72	2.8E-05	1.23	1.12–1.35
JHH	T	64	509	942	29	154	294	0.79	0.78	0.43	1.07	0.90–1.28
ATBC	T	57	305	562	52	331	534	0.77	0.76	0.45	1.06	0.91–1.24
FPCC	T	33	210	413	31	241	384	0.79	0.77	0.20	1.13	0.94–1.36
HPFS	T	27	199	368	30	221	359	0.79	0.77	0.31	1.11	0.91–1.34
PLCO	T	58	370	747	58	404	638	0.79	0.76	0.02	1.19	1.03–1.37
CPS-II	T	86	584	1089	105	616	1053	0.79	0.77	0.07	1.11	0.99–1.24
ALL	T	506	3201	5809	443	2643	4166	0.78	0.76	3.2E-07	1.15	1.09–1.21

p^{*}: Based on allelic test assuming multiplicative model. The combined tests are based on Mantel-Haenszel test.

p[^]: Breslow-Day test for homogeneity