Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2009 July ; 18(7): 2132–2136. doi:10.1158/1055-9965.EPI-08-1221.

Fine mapping and family-based association analyses of prostate cancer risk variants at Xp11

Lingyi Lu^{1,2}, Jielin Sun^{1,3}, Sarah D. Isaacs⁴, Kathleen E. Wiley⁴, Shelly Smith^{1,3}, Kristen Pruett^{1,3}, Yi Zhu^{1,3}, Zheng Zhang^{1,3}, Fredrik Wiklund⁵, Henrik Grönberg⁵, Patrick C. Walsh⁴, Bao-Li Chang^{1,3}, S. Lilly Zheng^{1,3}, William B. Isaacs^{4,†}, and Jianfeng Xu^{1,3}

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

²Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC

³Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC

⁴Johns Hopkins Medical Institutions, Baltimore, MD

⁵Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Abstract

Two SNPs (rs5945572 and rs5945619) at Xp11 were recently implicated in two genome-wide association studies (GWAS) of prostate cancer. Using a family-based association test for these two SNPs in 168 prostate cancer families, we showed in this study that the risk alleles of the two reported SNPs were over-transmitted to affected offspring, $P = 0.009$ for rs5945372 and $P = 0.03$ for rs5945619, which suggested the observed association in case-control studies were not driven by potential population stratification. We also performed a fine mapping study in ~800 kb region at Xp11 among two independent case-control studies, including 1,527 cases and 482 controls from Johns Hopkins Hospital and 1,172 cases and 1,157 controls from the Prostate, Lung, Colon and Ovarian (PLCO) Cancer Screening Trial. The strongest association was found with SNPs in the haplotype block where the two initial reported SNPs were located, although many SNPs in ~140 kb region were highly significant in the combined allelic tests ($P = 10^{-5}$ – 10^{-6}). The second strongest association was observed with SNPs in ~286 kb region at another haplotype block ($P = 10^{-4}$ – 10^{-5}), ~94 kb centromeric to the first region. The significance of SNPs in the second region decreased considerably after adjusting for SNPs at the first region, although P remained < 0.05 . Additional studies are warranted to test independent prostate cancer associations at these two regions.

Keywords

prostate cancer; population-based association; family-based association; Xp11; fine mapping

Introduction

Two SNPs, rs5945572 and rs5945619 at Xp11 were independently discovered in two GWAS of prostate cancer (PCa) (1–2) and confirmed in a large worldwide consortium of 13 case-control studies (3). These two SNPs were only 12 kb apart, in strong linkage disequilibrium (LD) ($D' = 1.00$ and $r^2 = 0.91$ in the HapMap CEU population), being located in the same

[†]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, Email: wisaacs@jhmi.edu.

haplotype block, and thus represented a single PCa risk locus at Xp11. Although the association at Xp11 far exceeded genome-wide significance, several important issues remained. First, because the reported association studies were uniformly based on case-control study designs, the association could be influenced by population stratification; i.e. the significantly different allele frequencies between cases and controls could in large or small part due to differences between the two groups in terms of race, ethnicity, or geographic regions (4). Family-based association tests which are not susceptible to population stratification can be used to address this issue (5). Second, the significant SNPs identified in GWAS may be indirectly associated with PCa risk via LD with other functional variants in the flanking regions. Fine mapping association tests that systematically evaluate association of SNPs within the haplotype block of these two SNPs may reveal the SNPs that have the strongest association, assisting in defining causal relationships. Finally, additional independent risk associated SNPs may exist at Xp11, as empirically demonstrated at other PCa risk loci such as 8q24 (6–9), 17q12 (10–11), and 11q13 (2,12,13) where second independent loci were subsequently discovered within a broader region in the vicinity of the initial locus identified by GWAS. A fine mapping study in a broader region that includes additional haplotype blocks at Xp11 may provide insight into this question.

In this study, we report results from a FBAT in PCa families and fine mapping studies in two case-control studies.

Methods

Study subjects

Family based association test was performed in hereditary prostate cancer (HPC) families collected and studied at the Brady Urology Institute at Johns Hopkins Hospital (JHH) as described previously (14). The working criterion for HPC in this study was PCa patients who have at least two additional first degree of relatives diagnosed with PCa. PCa diagnosis was verified by medical records for each affected male studied. Currently, 168 HPC families of European descent were informative for this analysis.

Fine mapping association analyses was performed in a hospital-based case-control population at JHH (9), including 1,527 PCa cases and 482 controls of European descent (by self report) (Supplementary Methods). To increase the sample size, we also included another study population from the National Cancer Institute Cancer Genetic Markers of Susceptibility (CGEMS) GWAS, including 1,172 PCa case patients and 1,157 control subjects of European American background (White and non-Hispanic) who were selected from the PLCO using an incidence density sampling strategy (12). Data were downloaded from <http://cgems.cancer.gov/data/>.

Region of interest, tagging SNPs, and SNP genotyping

We identified a ~800 kb region of interest for the fine mapping study (51,000,000–51,800,000, Build 35 of NCBI) based on the previously reported studies (1–3), results of our analysis of SNPs at Xp11 in the CGEMS GWAS that are publicly available, inferred haplotype blocks at Xp11 based on the HapMap CEU population, and known genes in the region. Haplotype blocks were estimated using the Haploview (15) computer program, and a default Gabriel method (16) was used to define each haplotype block; i.e. a region in which all (or nearly all) pairs of markers are in “strong LD”, which is consistent with no historical recombination. A total of 20 tagging SNPs were identified to capture ($r^2 > 0.8$) all the SNPs with minor allele frequency (MAF) of 5% or higher in the region of interest based on the HapMap CEU population. The tagging SNPs were genotyped using iPLEX (Sequenom, Inc). The genotype call rates of these SNPs were > 98%, and the average concordant rate between 100 duplicate samples was 99.8%.

Statistical methods

Family-based association tests were performed using the Family-Based Association Test (FBAT) software package (5). FBAT utilizes data from nuclear families, sibships, or a combination of the two, to test for linkage and linkage disequilibrium (association) between traits and genotypes. We used the empirical variance estimator in FBAT to perform a valid test of association, accounting for the correlation of transmitted alleles among multiple affected individuals in the same family due to linkage. We imputed all of the known SNPs in the genome 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. Allele frequency differences between case patients and control subjects were tested for each SNP, using a chi-square test with 1 degree of freedom. The allelic odds ratio (OR) and 95% confidence interval (95% CI) were estimated based on a multiplicative model. Results from two case-control populations were combined using a Mantel-Haenszel model in which the populations were allowed to have different population frequencies for alleles but were assumed to have a common OR. The homogeneity of ORs among different study populations was tested using a Breslow-Day chi-square test. Independence of PCa associations of several SNPs was tested by including significant SNPs in a logistic regression model using a backward selection method and adjusted for study population and age (categorized by 5 year intervals).

Results

To test whether the two SNPs at Xp11 (rs5945572 and rs5945619) were associated with PCa risk in hereditary PCa families, we genotyped these two SNPs in 168 HPC families of European ancestry. These two SNPs were significantly over-transmitted from parents to affected offspring (more than expected 50% under a null hypothesis of no association), $P = 0.009$ for rs5945572 and $P = 0.03$ for rs5945619. These results confirmed the association of these two SNPs with PCa risk in PCa families. More importantly, they suggested that the observed association in case-control studies were not entirely driven by potential population stratification.

We then performed a fine mapping analysis in an 800 kb region at Xp11 to identify SNPs that have the strongest associations with PCa risk. Twenty tagging SNPs were selected and genotyped in 1,527 PCa cases and 482 controls at JHH (Table 1). The two previously reported SNPs (rs5945572 and rs5945619) were significantly associated with PCa risk ($P < 0.05$), as described in an initial report (1). Four additional SNPs within the same haplotype block (~190 kb) of these two initial reported SNPs were also associated with PCa risk ($P < 0.05$). Another SNP (rs1595679) that was ~336 kb to the centromeric side and in a different haplotype block was also associated with PCa risk, $P = 0.005$.

To confirm these findings, we examined the associations for these 20 SNPs among 1,172 PCa cases and 1,157 control subjects from the CGEMS study, 7 of these SNPs were directly genotyped in the GWAS and the remaining 13 SNPs were imputed (12). Similar to the findings from the JHH study, multiple SNPs in the haplotype block of rs5945572 and rs5945619 were significantly associated with PCa risk ($P < 0.05$). The most significant SNP was rs5945619 ($P = 0.0001$). Importantly, we confirmed the association of the SNP rs1595679 that was initially identified in the JHH study population, $P = 0.004$.

To systematically evaluate the associations in the entire 800 kb region in a larger sample, we imputed 258 SNPs in the region for subjects in the JHH study (based on the 20 tagging SNPs) and 251 SNPs in the CGEMS study (based on the 27 genotyped SNPs in the GWAS data) and performed a combined Mantel-Haenszel test analysis (No evidence for heterogeneity in OR was found for any of the SNPs between the two study populations, $P > 0.05$). Many SNPs in

the region were associated with PCa risk and congregated in haplotype blocks 2–4 (Fig 1). The strongest association in the 800 kb region was with SNPs in haplotype block 2 where the two initial reported SNPs (rs5945572 and rs5945619) were located. SNPs in a ~140kb region were highly significant ($P = 10^{-5}$ – 10^{-6}), including rs5945619. This genomic region includes a known gene, nudix-type motif 11 (*NUDT11*). The second strongest association was with SNPs at haplotype block 4 where the SNP rs1595679 is located. SNPs in a ~286 kb region were highly significant ($P = 10^{-4}$ – 10^{-5}). This region includes two known genes: peptide chain release factor 3 (*GSPT2*), and melanoma antigen family D (*MAGED1*). The two strongest associated regions were separated by ~94 kb. When one representative SNP from each region was both included in a logistic regression analysis, they were both significant; $P = 0.002$ for rs5945619 at region 1 and $P = 0.03$ for SNP rs1595679 at region 2.

Discussion

This study addressed two important issues following the discovery of a novel PCa risk locus at Xp11 from GWASs (1–2). Using a family-based association test, we showed that the risk alleles of the two reported SNPs were over-transmitted to affected offspring in PCa families and therefore suggested that they are associated with PCa risk. Because family-based association test does not depend on a comparison between cases and controls, it avoids potential problem of comparability in genetic background between the two groups in case-control studies (4–5). Therefore, the results of this family-based association study accomplished the first goal of the study (issue of potential population stratification) and provided an independent confirmation of the PCa risk locus at Xp11.

The fine mapping analysis of this study addressed the second issue regarding the location of potential functional variants in the region. By systematically evaluating association for SNPs in the ~800 kb region among two case-control studies, we found SNPs in an ~140kb region had the strongest association with PCa risk. On this basis, priority should be given to this region, especially the *NUDT11* gene, for future studies that intend to identify functional variants at Xp11. Unfortunately, due to strong LD among the SNPs in the region, the region that is highly associated with PCa risk remains broad.

NUDT11 codes for a member of the MutT or nudix family of nucleoside hydrolyzing enzymes (18), metabolizing the small signaling molecules, diphosphoinositol polyphosphates, IP7 and IP8, as well as diadenosine polyphosphates. Transfection of *NUDT11* into human embryonic kidney cells results in a reduction of IP7 and IP8 by 35% and 45% respectively (19). Turnover of diphosphoinositol polyphosphates have been implicated in a variety of physiologic functions including apoptosis, endocytosis, telomere length maintenance, and chemotaxis (20).

Our fine mapping study also revealed another region at Xp11, ~94 kb from the two initially reported SNPs, is highly associated with PCa. Although the significance of association at this new locus decreases considerably after adjusting for the original SNP, additional studies are needed to further test the independence between the two regions. If confirmed, this is another example of an independent locus in the flanking region of a locus initially identified from GWAS, as previously observed for PCa risk loci at 8q24 (6–9), 17q12 (10–11), and 11q13 (2,12,13). Two known genes are in this new region (*GSPT2* and *MAGED1*). *GSPT2* (G1 to S phase transition 2), also known as *eRF3b* (eukaryotic peptide chain release factor subunit 3b), may play a role in translation termination and in cell cycle regulation (21–22). *MAGED1* (melanoma antigen family D, 1) is a member of the melanoma antigen gene (MAGE) family and has been shown to be involved in cell cycle progression and apoptosis (23).

Consistent with the findings from the published studies (1,3,24), we did not observe significantly different allele frequencies between patients with aggressive or non-aggressive

disease for these SNPs at Xp11 in either the JHH or CGEMS study (data not shown). Lack of association with aggressiveness of PCa is also found for other PCa risk variants recently identified from GWAS, including at 8q24, 17q12, 17q24, 3p12, 7q21, 11q13, and 10q11 (3, 18). Other study designs, including those comparing aggressive with non-aggressive PCa may be more appropriate to discover risk variants for aggressive PCa.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank all the study subjects who participated in this study. The study is supported by National Cancer Institute CA129684, CA106523 and CA95052 to J.X., CA112517 and CA58236 to W.B.I., and Department of Defense grant PC051264 to J.X. The authors also thanks for the National Cancer Institute Cancer Genetic Markers of Susceptibility Initiative (CGEMS) for making the data available publicly.

References

1. 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]
2. 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]
3. Kote-Jarai Z, Easton DF, Stanford JL, et al. Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. *Cancer Epidemiol Biomarkers Prev* 2008;17:2052–2061. [PubMed: 18708398]
4. Lange EM, Sun J, Lange LA, et al. Family-based samples can play an important role in genetic association studies. *Cancer Epidemiol Biomarkers Prev* 2008;17:2208–2214. [PubMed: 18768484]
5. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol* 2000;19:S36–S42. [PubMed: 11055368]
6. 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]
7. 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]
8. 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]
9. Zheng SL, Sun J, Cheng Y, et al. Additive effects of two unlinked loci at 8q24 are associated with a considerable fraction of prostate cancer among European Americans. *JNCI* 2007;99:1525–1533. [PubMed: 17925536]
10. 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]
11. Sun J, Zheng SL, Wiklund F, et al. Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. *Nat Genet* 2008;40:1153–1155. [PubMed: 18758462]
12. 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]
13. Zheng SL, Stevens VL, Wiklund F, et al. Two independent prostate cancer risk associated loci at 11q13. *Cancer Epidemiol Biomarkers Prev*. 2009 in press.
14. Xu J, Zheng SL, Hawkins GA, et al. Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22–23. *Am J Hum Genet* 2001;69:341–350. [PubMed: 11443539]
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. Bessman MJ, Frick DN, O'Handley SF. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J Biol Chem* 1996;271:25059–25062. [PubMed: 8810257]
19. Hidaka K, Caffrey JJ, Hua L, et al. An adjacent pair of human NUDT genes on chromosome X are preferentially expressed in testis and encode two new isoforms of diphosphoinositol polyphosphate phosphohydrolase. *J Biol Chem* 2002;277:32730–32738. [PubMed: 12105228]
20. Bennett M, Onnebo SM, Azevedo C, Saiardi A. Inositol pyrophosphates: metabolism and signaling. *Cell Mol Life Sci* 2006;63:552–564. [PubMed: 16429326]
21. Chauvin C, Salhi S, Le Goff C, Viranaicken W, Diop D, Jean-Jean O. Involvement of human release factors eRF3a and eRF3b in translation termination and regulation of the termination complex formation. *Mol Cell Biol* 2005;25:5801–5811. [PubMed: 15987998]
22. Hoshino S, Imai M, Mizutani M, Kikuchi Y, Hanaoka F, Ui M, Katada T. Molecular cloning of a novel member of the eukaryotic polypeptide chain-releasing factors (eRF). Its identification as eRF3 interacting with eRF1. *J Biol Chem* 1998;273:22254–22259. [PubMed: 9712840]
23. Barker PA, Salehi A. The MAGE proteins: emerging roles in cell cycle progression, apoptosis, and neurogenetic disease. *J Neurosci Res* 2002;67:705–712. [PubMed: 11891783]
24. Xu J, Isaacs SD, Sun J, et al. Association of prostate cancer risk variants with clinicopathologic characteristics of the disease. *Clin Cancer Res* 2008;14:5819–5824. [PubMed: 18794092]

Figure 1. A schematic view of genetic association between SNPs at Xp11 and prostate cancer risk in JHH
(a) Allelic tests for SNPs at Xp11 and prostate cancer risk in JHH. (b) Inferred haplotype blocks of these SNPs were estimated from the control subjects in JHH using the Haploview computer program. (c) Known genes at Xp11.

Table 1

Association between prostate cancer risk and SNPs at Xp11

JHH										CGEMS							
CHR	SNP	BP	Alleles	Risk allele	Block*	Frequency		Allelic	P	Type	Frequency		Allelic		P	Type	
						Cases	Controls				Cases	Controls	OR	95% CI			
X	rs972635	51,000,724	A/T	A	2	0.33	0.32	1.06	0.85-1.32	0.61	genotyped	0.33	0.28	1.25	1.04-1.51	1.71E-02	Imputed
X	rs11091727	51,034,123	A/C	A	2	0.21	0.20	1.09	0.84-1.41	0.50	genotyped	0.23	0.19	1.30	1.06-1.60	1.05E-02	genotyped
X	rs1327301	51,043,093	T/C	T	3	0.39	0.33	1.32	1.07-1.64	1.13E-02	genotyped	0.40	0.34	1.33	1.12-1.58	1.10E-03	genotyped
X	rs5987421	51,059,526	G/T	G	3	0.40	0.34	1.31	1.06-1.62	1.38E-02	genotyped	0.41	0.34	1.33	1.12-1.57	1.16E-03	Imputed
X	rs5945572	51,062,719	A/G	A	3	0.40	0.33	1.32	1.06-1.63	1.26E-02	genotyped	0.41	0.34	1.33	1.12-1.57	1.23E-03	genotyped
X	rs5945619	51,074,708	C/T	C	3	0.40	0.34	1.31	1.06-1.63	1.28E-02	genotyped	0.41	0.33	1.40	1.18-1.66	1.11E-04	genotyped
X	rs5945642	51,133,714	A/G	A	4	0.02	0.02	1.08	0.53-2.19	0.84	genotyped	0.04	0.03	1.37	0.85-2.22	0.20	genotyped
X	rs3131302	51,143,890	C/T	C	4	0.23	0.21	1.14	0.89-1.47	0.29	genotyped	0.25	0.20	1.34	1.10-1.64	3.33E-03	Imputed
X	rs5945645	51,145,627	A/G	A	4	0.05	0.04	1.22	0.73-2.05	0.44	genotyped	0.04	0.03	1.39	0.87-2.21	0.17	Imputed
X	rs3131301	51,149,503	G/A	G	4	0.04	0.03	1.34	0.76-2.38	0.31	genotyped	0.03	0.02	1.32	0.76-2.26	0.32	Imputed
X	rs1541238	51,176,671	G/C	G	4	0.39	0.32	1.35	1.09-1.67	6.99E-03	genotyped	0.40	0.33	1.37	1.15-1.63	3.98E-04	Imputed
X	rs5945652	51,177,331	A/G	A	4	0.27	0.29	0.92	0.73-1.15	0.47	genotyped	0.29	0.31	0.89	0.74-1.06	0.19	genotyped
X	rs2721997	51,177,728	A/C	A	4	0.03	0.02	1.33	0.68-2.59	0.40	genotyped	0.01	0.01	1.30	0.52-3.24	0.58	Imputed
X	rs2721998	51,184,563	C/A	C	4	0.01	0.00	4.45	0.58-33.96	0.11	genotyped	0.01	0.01	1.35	0.51-3.56	0.54	Imputed
X	rs1541242	51,186,589	G/A	G	4	0.20	0.20	1.04	0.81-1.35	0.76	genotyped	0.21	0.19	1.11	0.91-1.37	0.31	genotyped
X	rs7057039	51,195,647	G/A	G	4	0.36	0.30	1.34	1.07-1.67	9.52E-03	genotyped	0.37	0.32	1.24	1.05-1.48	1.41E-02	Imputed
X	rs12393443	51,361,128	G/T	G	4	0.19	0.18	1.09	0.84-1.42	0.53	genotyped	0.20	0.19	1.12	0.91-1.38	0.30	Imputed
X	rs5951081	51,396,816	C/T	C	4	0.06	0.07	0.88	0.58-1.32	0.53	genotyped	0.08	0.10	0.80	0.58-1.10	0.17	Imputed
X	rs1595679	51,398,670	G/A	G	4	0.15	0.10	1.60	1.15-2.23	5.46E-03	genotyped	0.14	0.10	1.46	1.12-1.88	4.22E-03	Imputed
X	rs12385318	51,403,149	A/G	A	4	0.03	0.03	0.97	0.52-1.83	0.93	genotyped	NA	NA	NA	NA	NA	NA

* indicates which haplotype block the SNPs falls in.