Short Communication

Fine-Mapping and Family-Based Association Analyses of Prostate Cancer Risk Variants at Xp11

Lingyi Lu,[1,2] Jielin Sun,[1,3] Sarah D. Isaacs,[4] Kathleen E. Wiley,[4] Shelly Smith,[1,3] Kristen Pruett,[1,3] Yi Zhu,[1,3] Zheng Zhang,[1,3] Fredrik Wiklund,[5] Henrik Grönberg,[5] Patrick C. Walsh,[4] Bao-Li Chang,[1,3] S. Lilly Zheng,[1,3] William B. Isaacs,[4] and Jianfeng Xu[1,3]

[1]Center for Cancer Genomics, [2]Department of Biostatistical Sciences, [3]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina; [4]Johns Hopkins Medical Institutions, Baltimore, Maryland; and [5]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Abstract

Two single nucleotide polymorphisms (SNP; rs5945572 and rs5945619) at Xp11 were recently implicated in two genome-wide association studies of prostate cancer. Using a family-based association test for these two SNPs in 168 families with prostate cancer, we showed in this study that the risk alleles of the two reported SNPs were overtransmitted to the affected offspring ($P = 0.009$ for rs5945372 and $P = 0.03$ for rs5945619), which suggested that the observed association in case-control studies were not driven by potential population stratification. We also did a fine-mapping study in the ~800 kb region at Xp11 between 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 Cancer screening trial. The strongest association was found with SNPs in the haplotype block in which the two initial reported SNPs were located, although many SNPs in the ~140 kb region were highly significant in the combined allelic tests ($P = 10^{-5}$ to 10^{-6}). The second strongest association was observed with SNPs in the ~286 kb region at another haplotype block ($P = 10^{-4}$ to 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 values remained at <0.05. Additional studies are warranted to test independent prostate cancer associations at these two regions. (Cancer Epidemiol Biomarkers Prev 2009;18(7):2132–6)

Introduction

Two single nucleotide polymorphisms (SNP), rs5945572 and rs5945619 at Xp11, were independently discovered in two genome-wide association studies (GWAS) of prostate cancer (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 (D' = 1.00 and r^2 = 0.91 in the HapMap CEU population), being located in the same haplotype block, and thus represented a single prostate cancer 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 be 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 prostate cancer risk via linkage disequilibrium with other functional variants in the flanking regions. Fine-mapping association tests that systematically evaluate the 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 shown empirically at other prostate cancer risk loci such as 8q24 (6-9), 17q12 (10, 11), and 11q13 (2, 12, 13) in which 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.

Received 1/7/09; revised 3/27/09; accepted 4/23/09; published OnlineFirst 6/23/09.

Grant support: National Cancer Institute grants CA129684, CA106523, and CA95052 (J. Xu); CA112517 and CA58236 (W.B. Isaacs), and Department of Defense grant PC051264 (J. Xu).

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers and Prevention Online (http://cebp.aacrjournals.org/).

Requests for reprints: William B. Isaacs, Johns Hopkins Hospital, Marburg 115, 600 North Wolfe Street, Baltimore, MD 21287. Phone: 410-955-2518; Fax: 410-955-0833. E-mail: wisaacs@jhmi.edu

doi:10.1158/1055-9965.EPI-08-1221

Table 1. Association between prostate cancer risk and SNPs at Xp11

CHR	SNP	BP	Alleles	Risk allele	Block*	JHH					CGEMS				
						Frequency		Allelic	*P*	Type	Frequency		Allelic	*P*	Type
						Case	Controls	OR (95% CI)			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)	1.71×02	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)	1.05×02	Genotyped
X	rs1327301	51,043,093	T/C	T	3	0.39	0.33	1.32 (1.07-1.64)	1.13×02	Genotyped	0.40	0.34	1.33 (1.12-1.58)	1.10×03	Genotyped
X	rs5987421	51,059,526	G/T	G	3	0.40	0.34	1.31 (1.06-1.62)	1.38×02	Genotyped	0.41	0.34	1.33 (1.12-1.57)	1.16×03	Imputed
X	rs5945572	51,062,719	A/G	A	3	0.40	0.33	1.32 (1.06-1.63)	1.26×02	Genotyped	0.41	0.34	1.33 (1.12-1.57)	1.23×03	Genotyped
X	rs5945619	51,074,708	C/T	C	3	0.40	0.34	1.31 (1.06-1.63)	1.28×02	Genotyped	0.41	0.33	1.40 (1.18-1.66)	1.11×04	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)	3.33×03	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)	6.99×03	Genotyped	0.40	0.33	1.37 (1.15-1.63)	3.98×04	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)	9.52×03	Genotyped	0.37	0.32	1.24 (1.05-1.48)	1.41×02	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)	5.46×03	Genotyped	0.14	0.10	1.46 (1.12-1.88)	4.22×03	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	

Abbreviations: CHR, chromosome; BP, 95% base pair CI, 95% confidence interval; NA, not applicable.
*Indicates which haplotype block the SNPs falls in.

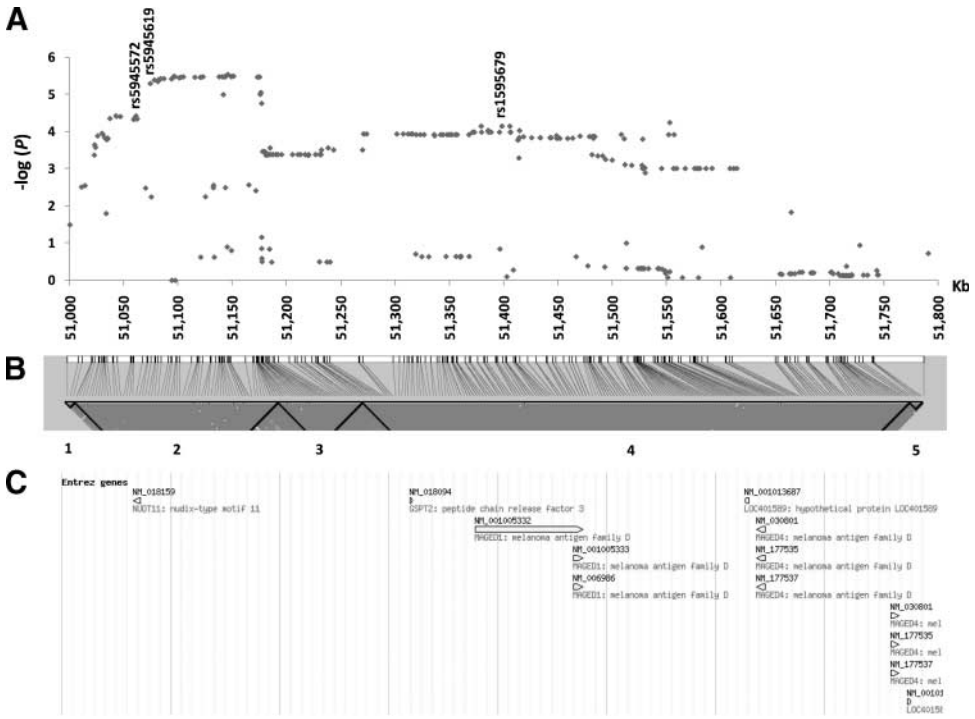

Figure 1. A schematic view of genetic association between SNPs at Xp11 and prostate cancer risk in JHH. **A.** Allelic tests for SNPs at Xp11 and prostate cancer risk in JHH. **B.** Inferred haplotype blocks of these SNPs were estimated from the control subjects in JHH using the HaploView computer program. **C.** Known genes at Xp11.

In this study, we report results from a family-based association test in families with prostate cancer and fine-mapping studies in two case-control studies.

Materials and Methods

Study Subjects. Family-based association tests were done in families with hereditary prostate cancer (HPC), and 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 patients with prostate cancer who have at least two additional first-degree relatives diagnosed with prostate cancer. Prostate cancer 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 done in a hospital-based case-control population at JHH (9), including 1,527 prostate cancer 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 prostate cancer case patients and 1,157 control subjects of European American background (White and non-Hispanic) who were selected from the Prostate, Lung, Colon and Ovarian Cancer screening trial using an incidence density sampling strategy (12). Data were downloaded from the National Cancer Institute CGEMS web site.[6]

Regions of Interest, Tagging SNPs, and SNP Genotyping. We identified an ~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). The results of our analysis of SNPs at Xp11 in the CGEMS GWAS, which 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 linkage disequilibrium'', 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 a minor allele frequency 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 done using the family-based association test (FBAT) software package (5). FBAT uses 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 χ^2 test with 1

[6] http://cgems.cancer.gov/data/

df. The allelic odds ratio (OR) and 95% confidence interval 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 χ^2 test. The independence of prostate cancer 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-y intervals).

Results

To test whether the two SNPs at Xp11 (rs5945572 and rs5945619) were associated with prostate cancer risk in families with HPC, we genotyped these two SNPs in 168 HPC families of European ancestry. These two SNPs were significantly overtransmitted from the parents to the affected offspring (more than the 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 and prostate cancer risk in families with prostate cancer. More importantly, they suggested that the observed association in case-control studies were not entirely driven by potential population stratification.

We then did a fine-mapping analysis in an 800 kb region at Xp11 to identify SNPs that have the strongest associations with prostate cancer risk. Twenty tagging SNPs were selected and genotyped in 1,527 prostate cancer cases and 482 controls at JHH (Table 1). The two previously reported SNPs (rs5945572 and rs5945619) were significantly associated with prostate cancer 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 prostate cancer 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 prostate cancer risk ($P = 0.005$).

To confirm these findings, we examined the associations for these 20 SNPs among 1,172 prostate cancer 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 prostate cancer 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 did 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 prostate cancer risk and congregated in haplotype blocks 2 to 4 (Fig. 1). The strongest association in the 800 kb region was with SNPs in haplotype block 2 in which the two initial reported SNPs (rs5945572 and rs5945619) were located. SNPs in a ~140 kb region were highly significant ($P = 10^{-5}$ to 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, in which SNP rs1595679 is located. SNPs in a ~286 kb region were highly significant ($P = 10^{-4}$ to 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 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 prostate cancer risk locus at Xp11 from GWAS (1, 2). Using FBAT, we showed that the risk alleles of the two reported SNPs were overtransmitted to affected offspring in families with prostate cancer and therefore suggested that they are associated with prostate cancer risk. Because FBAT does not depend on a comparison between cases and controls, it avoids the 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 (the issue of potential population stratification) and provided an independent confirmation of the prostate cancer 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 the two case-control studies, we found that SNPs in the ~140 kb region had the strongest association with prostate cancer 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 linkage disequilibrium among the SNPs in the region, the region that is highly associated with prostate cancer 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). The 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 that another region at Xp11, ~94 kb from the two initially reported SNPs, is highly associated with prostate cancer. 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 prostate cancer 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* (G_1 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 D1) 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 nonaggressive disease for these SNPs at Xp11 in either the JHH or CGEMS study (data not shown). A lack of association with the aggressiveness of prostate cancer was also found for other prostate cancer 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 nonaggressive prostate cancer may be more appropriate to discover risk variants for aggressive prostate cancer.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The authors thank all the study subjects who participated in this study, and the National Cancer Institute 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;40:281–3.
2. Eeles RA, Kote-Jarai Z, Giles GG, et al. Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 2008;40: 316–21.
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; 17:2052–61.
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;17:2208–14.
5. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol 2000;19 Suppl 1: S36–42.
6. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African populations. Nat Genet 2006;38:652–8.
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;39:631–7.
8. Haiman CA, Patterson N, Freedman ML, et al. Multiple regions within 8q24 independently affect risk for prostate cancer. Nat Genet 2007;39:638–44.
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. J Natl Cancer Inst 2007;99:1525–33.
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;39:977–83.
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;40:1153–5.
12. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008; 40:310–5.
13. Zheng SL, Stevens VL, Wiklund F, et al. Two independent prostate cancer risk associated loci at 11q13. Cancer Epidemiol Biomarkers Prev. In press 2009.
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;69:341–50.
15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21: 263–5.
16. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
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;39:906–13.
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;271:25059–62.
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;277:32730–8.
20. Bennett M, Onnebo SM, Azevedo C, Saiardi A. Inositol pyrophosphates: metabolism and signaling. Cell Mol Life Sci 2006;63:552–64.
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;25:5801–11.
22. Hoshino S, Imai M, Mizutani M, et al. 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: 22254–9.
23. Barker PA, Salehi A. The MAGE proteins: emerging roles in cell cycle progression, apoptosis, and neurogenetic disease. J Neurosci Res 2002;67:705–12.
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;14:5819–24.

Fine-Mapping and Family-Based Association Analyses of Prostate Cancer Risk Variants at Xp11

Lingyi Lu, Jielin Sun, Sarah D. Isaacs, et al.

Cancer Epidemiol Biomarkers Prev 2009;18:2132-2136.