Replication of five prostate cancer loci identified in an Asian population – Results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3)

Sara Lindstrom[1,2], **Fredrick R. Schumacher**[3], **Daniele Campa**[4], **Demetrius Albanes**[5], **Gerald Andriole**[6], **Sonja I. Berndt**[5], **H. Bas Bueno-de-Mesquita**[7,8], **Stephen J. Chanock**[5], **W. Ryan Diver**[9], **J. Michael Ganziano**[10,11], **Susan M. Gapstur**[9], **Edward Giovannucci**[12], **Christopher A. Haiman**[3], **Brian Henderson**[3], **David J Hunter**[1,2,13], **Mattias Johansson**[14,15], **Laurence N. Kolonel**[16], **Loic Le Marchand**[16], **Jing Ma**[13], **Meir Stampfer**[2,12,13], **Victoria L. Stevens**[9], **Dimitrios Trichopoulos**[2,17], **Jarmo Virtamo**[18], **Walter C. Willett**[12,13], **Meredith Yeager**[5], **Ann W. Hsing**[5], and **Peter Kraft**[1,2,19]

[1] Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, MA, USA [2] Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA [3] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [4] Genomic Epidemiology Group, German Cancer Research Center DKFZ, Heidelberg, Germany [5] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [6] Division of Urologic Surgery, Washington University School of Medicine, St Louis, MO, USA [7] National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands [8] Department of Gastroenterology and Hepatology, University Medical Centre Utrecht (UMCU), Utrecht, The Netherlands [9] Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA [10] Massachusetts Veterans Epidemiology and Research Information Center (MAVERIC) and Geriatric Research, Education, and Clinical Center (GRECC), Boston Veterans Affairs Healthcare System, Boston, MA, USA [11] Division of Aging, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA [12] Department of Nutrition, Harvard School of Public Health, Boston, MA, USA [13] Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA [14] International Agency for Research on Cancer, Lyon, France [15] Department of Surgical and Perioperative Sciences, Urology and Andrology, Umeå University, Sweden [16] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA. [17] Bureau of Epidemiologic Research, Academy of Athens, Athens, Greece [18] Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [19] Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

Abstract

Background—A recent Genome-Wide Association Study (GWAS) of prostate cancer in a Japanese population identified five novel regions not previously discovered in other ethnicities. In this study, we attempt to replicate these five loci in a series of nested prostate cancer case-control studies of European ancestry.

Methods—We genotyped five SNPs: rs13385191 (chromosome 2p24), rs12653946 (5p15), rs1983891 (6p21), rs339331 (6p22) and rs9600079 (13q22), in 7,956 prostate cancer cases and 8,148 controls from a series of nested case-control studies within the NCI Breast and Prostate

Corresponding Author: Sara Lindstrom, Ph.D. Department of Epidemiology Harvard School of Public Health 655 Huntington Avenue Building 2, Room 249B Boston, MA 02115 Phone: 617 432 7681 Fax: 617 432 1722 slindstr@hsph.harvard.edu.

Cancer Cohort Consortium (BPC3). We tested each SNP for association with prostate cancer risk and assessed if associations differed with respect to disease severity and age of onset.

Results—Four SNPs (rs13385191, rs12653946, rs1983891 and rs339331) were significantly associated with prostate cancer risk (p-values ranging from 0.01 to 1.1×10^{-5}). Allele frequencies and odds ratios were overall lower in our population of European descent compared to the discovery Asian population. SNP rs13385191 (*C2orf43)* was only associated with low-stage disease (p=0.009, case-only test). No other SNP showed association with disease severity or age of onset. We did not replicate the 13q22 SNP, rs9600079 (p=0.62).

Conclusions—Four SNPs associated with prostate cancer risk in an Asian population are also associated with prostate cancer risk in men of European descent.

Impact—This study illustrates the importance of evaluation of prostate cancer risk markers across ethnic groups.

INTRODUCTION

Ethnicity is a well-established but poorly understood risk factor for prostate cancer. African American men experience the highest risk, followed by European and Asian men (1). The difference in incidence across ethnicities has been attributed to both genetic and lifestyle factors, but the specific underlying mechanisms are unknown.

Genome-wide association studies (GWAS) have identified multiple common genetic variants associated with prostate cancer risk. However, current GWAS of prostate cancer have predominantly been conducted in European populations and to date, only two GWAS have been completed in non-European ethnicities. Takata and colleagues conducted a GWAS in a Japanese population and identified five novel loci with allele frequencies ranging between 0.36 and 0.44 (2). Haiman and colleagues carried out a GWAS in an African American population and identified a novel locus at chromosome 17q21 (3). The prevalence of the risk allele was 5% in men of African descent, but rare in other populations (<1%), possibly explaining some of the differences in risk between ethnicities.

Attempts to generalize known genetic associations across ethnicities have had mixed results. Of 31 loci reported in Europeans, 19 were replicated (p 0.05) in an Asian study of 4,584 prostate cancer cases and 8,801 controls (2). Using the same significance level, an African American prostate cancer case-control study of 3,425 prostate cancer cases and 3,290 controls replicated about half of 49 risk variants identified in men of European and Asian descent (4). These results highlight genetic differences across ancestral populations and call for further investigation.

The aim of this study is to assess five prostate cancer SNPs identified in men of Asian descent in a series of nested case-control studies within the NCI Breast and Prostate Cancer Cohort Consortium (BPC3) including a total of 7,956 prostate cancer cases and 8,148 controls, of European ancestry. We also investigated whether the associations differed with disease severity and age of onset.

MATERIAL AND METHODS

Study Population

The BPC3 has been described in detail elsewhere (5). In brief, the consortium combines resources from eight well-established cohort studies with blood samples collected as follows: the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study (6), American Cancer Society Cancer Prevention Study II (CPS-II) (7), the European

Prospective Investigation into Cancer and Nutrition Cohort (EPIC – comprised of cohorts from Denmark, Great Britain, Germany, Greece, Italy, the Netherlands, Spain, and Sweden) (8), the Health Professionals Follow-up Study (HPFS) (9), the Multiethnic Cohort (MEC) (10), the Melbourne Collaborative Cohort Study (MCCS) (11), the Physicians' Health Study (PHS) (12) and the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial (13). Together, these eight cohorts collectively include over 265,000 men who provided a blood sample. The current study was restricted to individuals who self-reported as being Caucasian. We had genotype data for a total of 7,956 prostate cancer cases and 8,148 controls; ATBC (862 cases and 742 controls), CPS-II (1,765 cases and 1,998 controls), EPIC (938 cases and 1,248 controls), HPFS (1,285 cases and 1,249 controls), MEC (666 cases and 735 controls), PHS (1,366 cases and 1,281 controls) and PLCO (1,074 cases and 895 controls).

Prostate cancer cases were identified through population-based cancer registries or self-reports confirmed by medical records, including pathology reports. BPC3 consists of a series of matched nested case-control studies within each cohort; controls were matched to cases on a number of potential confounding factors, such as age, ethnicity, and region of recruitment, depending on the cohort. Informed consent was obtained from all subjects and each study was approved by the Institutional Review Boards at their respective institutions. Data on disease stage and grade at time of diagnosis were collected from each cohort, wherever possible. A total of 1,254 cases were classified as high-stage (defined as stage C or D at diagnosis) and 985 cases were classified as high-grade (defined as Gleason grade 8 or equivalent, i.e. coded as poorly differentiated).

SNP selection and genotyping

We genotyped 5 SNPs identified by Takata et al. (rs13385191, rs12653946, rs1983891, rs339331, rs9600079). For rs1983891, we genotyped either rs1983891or a surrogate, rs9381080 (r^2=1.00 in HapMap CEU population and r^2=0.93 in HapMap JPT+CHB population).

Genotyping was performed using the TaqMan assay (Applied Biosystems, Foster City, CA) in four different genotyping laboratories: Core Genotyping Facility at National Cancer Institute, Harvard School of Public Health, University of South California and DKFZ. Average success rate was 0.97 (0.92-1.00). Blinded duplicated samples indicated high reproducibility (100%). For each SNP and study, we tested for fitness for Hardy Weinberg Equilibrium proportion, commonly referred to as Hardy-Weinberg Equilibrium (HWE) in the controls. rs1983891 did not conform to HWE in MEC (p=0.002) and PHS (p=0.001). We reviewed the cluster plots for rs1983891 in MEC and PHS and as these were satisfactory, the genotype data for those cohorts were included in the analysis. All other SNPs were in HWE (P>0.01).

Statistical methods

We tested the association between prostate cancer risk and each SNP with a likelihood ratio test based on unconditional logistic regression. We adjusted all analyses for study and age at diagnosis or selection as a control (in five year intervals) using indicator variables. All odds ratios are calculated per copy of minor allele (0,1,2) carried. For each SNP, we used Cochran's Q statistic to test for heterogeneity between studies. To estimate odds ratios for high-grade or low-grade disease, we performed multinomial regression with an outcome variable coded as 0 (control), 1 (low-grade) or 2 (high-grade). To test for differential SNP associations between low-grade and high-grade disease, we used a likelihood ratio test based on the case-only analysis. We repeated these analyses for high-stage/low-stage disease. We tested for interaction between SNPs and age of onset (65 years/ >65 years) by conducting

a one degree-of-freedom likelihood ratio test of a single interaction term (SNPxAge) as implemented in unconditional logistic regression. We tested for dominance deviation from an additive model by including an additional SNP covariate coded as (0,1,0) for (homozygote common allele, heterozygote, homozygote rare allele) respectively. Based on unconditional regression, we performed a one degree-of-freedom likelihood ratio test where the full model (with two SNP covariates coded as (0,1,0) and (0,1,2)), was tested against a model only including the SNP covariate with additive coding (0,1,2) as described above. All reported p-values are two-sided and uncorrected for multiple hypothesis testing. Analyses were conducted in R (14), Quanto (15) and SAS version 9.1.

RESULTS

For all association tests between SNP and prostate cancer risk, we observed no evidence of heterogeneity among studies (all p 0.01, data not shown). Compared to the Japanese study, all SNPs except the 13q22 SNP rs9600079, were less common in our population of European descent (average difference in minor allele frequency (MAF): -0.096, Table 1). We replicated four of the five SNPs at the 0.05 significance level (Table 1). All associations were in the same direction as in the Japanese study; however, the allele-specific odds ratios were lower in our population, possibly reflecting the "winners curse". In agreement with the Japanese GWAS, we observed the strongest association for rs12653946, with an allele-specific odds ratio in BPC3 of 1.10 (95% CI: 1.06-1.16, $p=1.12\times10^{-5}$), compared to the Japanese GWAS (OR: 1.26 (95% CI: 1.20-1.33). The only SNP not associated with prostate cancer risk in BPC3 was 13q22 SNP rs9600079 (p=0.62).

In agreement with the Japanese report, associations generally did not differ with Gleason grade or tumor stage (Table 2). SNP rs13385191 (*C2orf43)* was only associated with low-stage disease (p=0.009, case-only test). However, as these analyses are of exploratory nature, we cannot rule out chance findings. There were no significant interactions with age of onset (Table 3) and restricting to young cases (<60 years, n=705) did not alter the results (data not shown). SNP rs1983891 showed strong evidence of departure from an additive inheritance ($p=3.6x10^{-4}$), and instead conformed to a dominant inheritance model (OR: 1.19; 95% CI: 1.11-1.27 for C/T carriers and OR: 1.07; 95% CI: 0.96-1.20 for T/T carriers). No other SNP showed evidence of departure from additivity (p=0.06-0.51).

DISCUSSION

In this study, five SNPs identified in a GWAS of prostate cancer among men of Asian descent were assessed in a series of nested case-control studies of men with European descent. We replicated four SNPs: rs13385191, rs12653946, rs1983891 and rs339331 at p<0.05. All associations were in the same direction as observed in the Japanese population but minor allele frequencies were in general lower in our study population of European descent.

The association with rs9600079, a SNP located in a gene desert on chromosome 13q22 did not replicate. Using HapMap3 data, we compared the linkage disequilibrium (LD) pattern in European Americans (CEU) and Japanese (JPT) in a 100kb region around rs9600079. By visual inspection, we divided the region in two main haplotype blocks split by a recombination hotspot for both JPT and CEU. For JPT, rs9600079 is located in a 58 kb block and the 'T' allele (risk allele) was found on one haplotype with frequency above 0.05. For CEU, rs9600079 is located in a 44 kb block and the 'T' allele was found on three haplotypes each with frequency above 0.05. Thus, it is possible that we fail to detect the association in Europeans due to different LD patterns. If the true causal allele is located on the single haplotype that was identified in JPT, this signal would be diminished in a

population of European descent as the LD with the causal allele would be lower. In agreement with our study, an African American study failed to replicate rs9600079 (p=0.53) (4). Of note, the risk allele was more common in both Europeans (0.45) and African Americans (0.52) compared to Japanese (0.38). Extensive genotyping in this region would be required to investigate this finding further. The lack of association in Europeans could also be due to a possible GxE interaction involving an environmental exposure only present or more common in Asian populations.

All SNPs had smaller odds ratios in our population (1.06-1.10) compared to the original GWAS (odds ratio ranging between 1.15-1.26). There are several potential explanations for this including the "winners curse", a phenomenon where the actual genetic effect is typically smaller than its original estimate (16). Another possible explanation is that the identified variants are in lower LD with the causal variant in Europeans, and would thus have not been detected in previous GWAS of European descent. The power to detect these SNPs at a p-value $<10^{-5}$ in a modest-sized GWAS (2,000 cases and 2,000 controls) assuming the same allele frequency and odds ratios as observed here ranged from 0.09%-1% indicating that previous GWAS in European populations had very limited power to detect these at the discovery stage. This stresses that GWAS results from alternative populations provide a way to identify candidate SNPs that may have been missed due to stringent p-value thresholds.

These five SNPs included in the analysis herein were also assessed in a population of African Americans (4). In agreement with our study, rs1983891 (p=0.02) and rs339331 (p=3.1×10^{-6}) were both associated with prostate cancer and for rs339331, a stronger signal (rs12202378) was identified. In contrast to our results, rs12653946 (p=0.15) and rs13385191 (p=0.90) did not replicate in African Americans. However, rs13385191 was rare in African Americans (MAF=0.05) and fine-mapping revealed association with the nearby SNP rs340623. Using HapMap data, we observed differences in LD between rs13385191 and rs340623 across ethnicities, with modest LD in Asians (r-sq=0.40) and Europeans (r-sq=0.27) but no LD in African Americans (r-sq=0) possibly explaining the discordant results.

Attempts to replicate GWAS findings across population have had mixed results, most likely reflecting different LD patterns across ethnicities resulting in differences in allele frequencies for specific SNP markers. It is believed that most index signals from GWAS tag a yet unknown variant directly associated with prostate cancer risk and failure to replicate across ethnicities supports this hypothesis. Indeed, fine-mapping efforts of known prostate cancer regions in African Americans identified a stronger associated marker in 12 out of 28 regions (4). Other reasons for failure in replication include false positives in the original report and false negatives in the replication report. Nonetheless, our study has shown that GWAS data from one ethnic group can identify markers associated with disease in a different ethnicity. Continued efforts to characterize susceptibility regions in various ethnicities will be instrumental for localizing the causal markers.

In summary, four SNPs identified by a prostate cancer GWAS in an Asian population were replicated in this large study of men of European ancestry. Our study confirms the association of these loci with prostate cancer across multiple ethnicities, and supports continued evaluation of markers across ethnic groups.

Acknowledgments

REFERENCES

1. Gronberg H. Prostate cancer epidemiology. Lancet. 2003; 361:859–864. [PubMed: 12642065]

2. Takata R, Akamatsu S, Kubo M, Takahashi A, Hosono N, Kawaguchi T, et al. Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. Nat Genet. 2010; 42:751–754. [PubMed: 20676098]

3. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. Nat Genet. 2011; 43:570–573. [PubMed: 21602798]

4. Haiman CA, Chen GK, Blot WJ, Strom SS, Berndt SI, Kittles RA, et al. Characterizing genetic risk at known prostate cancer susceptibility Loci in african americans. PLoS Genet. 2011; 7:e1001387. [PubMed: 21637779]

5. Hunter DJ, Riboli E, Haiman CA, Albanes D, Altshuler D, Chanock SJ, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer. 2005; 5:977–985. [PubMed: 16341085]

6. The Alpha-Tocopherol, Beta Carotene Cancer Prevention Study Group. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. N Engl J Med. 1994; 330:1029–1035. [PubMed: 8127329]

7. Calle EE, Rodriguez C, Jacobs EJ, Almon ML, Chao A, McCullough ML, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. Cancer. 2002; 94:2490–2501. [PubMed: 12015775]

8. Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002; 5:1113–1124. [PubMed: 12639222]

9. Giovannucci E, Pollak M, Liu Y, Platz EA, Majeed N, Rimm EB, et al. Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. Cancer Epidemiol Biomarkers Prev. 2003; 12:84–89. [PubMed: 12582016]

10. Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. Nat Rev Cancer. 2004; 4:519–527. [PubMed: 15229477]

11. Severi G, Morris HA, MacInnis RJ, English DR, Tilley WD, Hopper JL, et al. Circulating insulin-like growth factor-I and binding protein-3 and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev. 2006; 15:1137–1141. [PubMed: 16775172]

12. Chan JM, Stampfer MJ, Ma J, Gann P, Gaziano JM, Pollak M, et al. Insulin-like growth factor-I (IGF-I) and IGF binding protein-3 as predictors of advanced-stage prostate cancer. J Natl Cancer Inst. 2002; 94:1099–1106. [PubMed: 12122101]

13. Hayes RB, Reding D, Kopp W, Subar AF, Bhat N, Rothman N, et al. Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. Control Clin Trials. 2000; 21:349S–355S. [PubMed: 11189687]

14. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2010.

15. Gauderman W; Morrison, J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies. 2006. http://hydra.usc.edu/gxe

16. Zollner S, Pritchard JK. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am J Hum Genet. 2007; 80:605–615. [PubMed: 17357068]

Table 1

Association between SNPs identified in the Asian GWAS (2) and prostate cancer risk in BPC3

SNP	Chr	Gene	Major Allele/Minor Allele	# Cases (MAF[1])	# Controls (MAF[1])	MAF Controls Takata et al.	OR[2] (95% CI) – BPC3	P-value	OR - Takata et al
rs13385191	2p24	*C2orf43*	A/G	7,808 (0.256)	8,017 (0.243)	0.558[3]	1.07 (1.02-1.12)	0.011	1.15[4]
rs12653946	5p15		C/T	7,777 (0.449)	7,974 (0.423)	0.443	1.10 (1.06-1.16)	1.12×10^{-5}	1.26
rs1983891	6p21	*FOXP4*	C/T	7,838 (0.296)	8,036 (0.277)	0.410	1.09 (1.04-1.15)	2.48×10^{-4}	1.15
rs339331	6p22	*GPRC6A/RFX6*	T/C	7,824 (0.276)	8,030 (0.291)	0.366	0.93 (0.88-0.97)	0.0019	0.82
rs9600079	13q22		G/T	7,803 (0.449)	8,017 (0.446)	0.382	1.01 (0.97-1.06)	0.62	1.18

[1] 'MAF=Minor allele frequency

[2] Odds Ratio per minor allele

[3] 'A' was the minor allele in Takata et al.

[4] Odds ratio per 'G' allele

Table 2

Association with Gleason grade and Tumor stage

SNP	Gleason <8 (n=6,691)	Gleason 8-10 (n=1,240)	P (case-only)	Stage AB (n=6,285)	Stage CD (n=1,624)	P (case-only)
rs13385191	1.08 (1.03-1.15)	1.04 (0.93-1.15)	0.69	1.09 (1.03-1.15)	0.96 (0.87-1.06)	0.009
rs12653946	1.09 (1.04-1.14)	1.14 (1.04-1.26)	0.23	1.12 (1.06-1.17)	1.08 (0.99-1.18)	0.58
rs1983891	1.09 (1.03-1.15)	1.15 (1.04-1.28)	0.17	1.11 (1.05-1.17)	1.06 (0.96-1.16)	0.42
rs339331	0.94 (0.89-0.99)	0.89 (0.80-0.99)	0.35	0.92 (0.87-0.97)	0.92 (0.84-1.02)	0.90
rs9600079	1.01 (0.96-1.06)	0.96 (0.87-1.06)	0.33	1.01 (0.96-1.07)	0.96 (0.88-1.04)	0.19

Table 3

Interaction with age of onset

SNP	Chr	Gene	#Cases/#Controls	Age of onset ≤65 years	#Cases/#Controls	Age of onset >65 years	$P_{Interaction}$
rs13385191	2p24	*C2orf43*	2,363/2,460	1.09 (1.00–1.20)	5,445/5,557	1.06 (1.00–1.12)	0.55
rs12653946	5p15		2,354/2,453	1.06 (0.98–1.15)	5,423/5,521	1.13 (1.07–1.19)	0.19
rs1983891	6p21	*FOXP4*	2,367/2,478	1.17 (1.07–1.27)	5,471/5,558	1.06 (1.01–1.13)	0.08
rs339331	6p22	*GPRC6A/RFX6*	2,366/2,470	0.92 (0.84–1.00)	5,458/5,560	0.93 (0.88–0.99)	0.91
rs9600079	13q22		2,366/2,474	1.03 (0.95–1.11)	5,437/5,558	1.01 (0.95–1.06)	0.64