

Genome-wide copy-number variation analysis identifies common genetic variants at 20p13 associated with aggressiveness of prostate cancer

Guangfu Jin^{1,2,†}, Jishan Sun^{1,2,†}, Wennuan Liu^{1,2},
Zheng Zhang^{1,2}, Lisa W.Chu¹, Seong-Tae Kim^{1,2},
Jielin Sun^{1,2}, Junjie Feng^{1,2}, David Duggan³,
John D.Carpten³, Fredrik Wiklund⁴, Henrik Grönberg⁴,
William B.Isaacs⁵, S.Lilly Zheng^{1,2} and Jianfeng Xu^{1,2,*}

¹Center for Cancer Genomics and ²Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA, ³Translational Genomics Research Institute, Phoenix, AZ 85004, USA, ⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden and ⁵Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD 21205, USA

*To whom correspondence should be addressed. Tel: +1 336 713 7500;
Fax: +1 336 713 7566;
Email: jxu@wfubmc.edu

The genetic determinants for aggressiveness of prostate cancer (PCa) are poorly understood. Copy-number variations (CNVs) are one of the major sources for genetic diversity and critically modulate cellular biology and human diseases. We hypothesized that CNVs may be associated with PCa aggressiveness. To test this hypothesis, we conducted a genome-wide common CNVs analysis in 448 aggressive and 500 nonaggressive PCa cases recruited from Johns Hopkins Hospital (JHH1) using Affymetrix 6.0 arrays. Suggestive associations were further confirmed using single-nucleotide polymorphisms (SNPs) that tagged the CNVs of interest in an additional 2895 aggressive and 3094 nonaggressive cases, including those from the remaining case subjects of the JHH study (JHH2), the NCI Cancer Genetic Markers of Susceptibility (CGEMS) Study, and the CAncer of the Prostate in Sweden (CAPS) Study. We found that CNP2454, a 32.3 kb deletion polymorphism at 20p13, was significantly associated with aggressiveness of PCa in JHH1 [odds ratio (OR) = 1.30, 95% confidence interval (CI): 1.01–1.68; $P = 0.045$]. The best-tagging SNP for CNP2454, rs2209313, was used to confirm this finding in both JHH1 ($P = 0.045$) and all confirmation study populations combined ($P = 1.77 \times 10^{-3}$). Pooled analysis using all 3353 aggressive and 3584 nonaggressive cases showed the T allele of rs2209313 was significantly associated with an increased risk of aggressive PCa (OR = 1.17, 95% CI: 1.07–1.27; $P = 2.75 \times 10^{-4}$). Our results indicate that genetic variations at 20p13 may be responsible for the progression of PCa.

Introduction

Over the past few years, genome-wide association studies have had great success in identifying common genetic variants that are associated with complex diseases (1). To date, >30 susceptibility loci related to prostate cancer (PCa) have been identified (2–13). In combination with family history, these markers have shown potential to improve the accuracy of PCa risk prediction (14). In contrast, few markers have been identified that can distinguish more aggressive PCa from indolent PCa (15), which is critical for more informed treatment

Abbreviations: CAPS, CAncer of the Prostate in Sweden; CI, confidence interval; CNP, copy-number polymorphism; CNV, copy-number variation; JHH, Johns Hopkins Hospital; LD, linkage disequilibrium; OR, odds ratio; PCa, prostate cancer; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; SNP, single-nucleotide polymorphism.

[†]These authors contributed equally to this work.

decisions. Recently, we identified a single-nucleotide polymorphism (SNP) at 17p12 (rs4054823) that was able to distinguish between the risk for more versus less aggressive forms of PCa (16). Although that SNP discovery represented an important starting point, additional risk predictors will be required before application in clinical settings can be considered. The identification of additional genetic variants that are relevant to the aggressiveness of PCa may provide an opportunity to more reliably predict the aggressiveness of PCa.

In addition to SNPs, copy-number variations (CNVs) are considered another major source of genetic diversity, which has been shown to modulate cellular biological functions such as adhesion, recognition and communication (17). Emerging evidence suggests a relationship of common CNVs with complex diseases including Crohn's disease (18), psoriasis (19), neuroblastoma (20) and schizophrenia (21). About 80% of high frequency CNVs [minor allele frequency (MAF) > 0.1] with two or three classes (diallelic CNVs) are in strong linkage disequilibrium (LD) with surrounding SNP(s) (22). Taking into account, the fact that the alteration of genomic structure derived from a CNV is much higher than SNP(s) within the same LD region (17), CNVs are more probably to be causal variants (23).

Based on Affymetrix Genome-wide SNP array 5.0, we identified a common CNV (deletion) at 2p24.3 that was significantly associated with PCa risk (24). Although we also observed that the frequency of this deletion was higher in aggressive than in nonaggressive cases, we did not find a statistically significant association between this CNV and PCa aggressiveness (24).

To follow-up and expand on our previous findings, in this study, we conducted a case–case study of 448 aggressive and 500 nonaggressive cases recruited at Johns Hopkins Hospital (JHH1) to identify common CNVs that may be related to PCa aggressiveness based on a genome-wide survey using the Affymetrix 6.0 array. Positive associations were further confirmed using tagging SNPs in an additional 2895 aggressive and 3094 nonaggressive cases, including those from the remaining case subjects of the JHH study (JHH2; 977 aggressive and 987 nonaggressive), the NCI Cancer Genetic Markers of Susceptibility (CGEMS) Study (687 aggressive and 488 nonaggressive) and the CAncer of the Prostate in Sweden (CAPS) Study (1231 aggressive and 1619 nonaggressive).

Subjects and methods

Study subjects

Three groups of subjects were included in the study (Figure 1). The JHH subjects included PCa patients who underwent radical prostatectomy at JHH between January 1 1999 and December 31 2008, which was reported previously (16). In the current study, only individuals of European descent were included. Patients with a Gleason score of 7, with the most prevalent pattern being Gleason grade 4 or higher, stage T3b or higher and/or N+ or M+, were defined as having aggressive disease. Those with Gleason score of 7, with the most prevalent pattern being Gleason grade 3 or lower or no evidence of disease dissemination (pathologic stage T2/N0/M0), were defined as having nonaggressive disease. In the discovery stage, a total of 948 JHH patients (448 aggressive and 500 nonaggressive PCa cases; JHH1) were selected for genotyping on the Affymetrix 6.0 array platform (supplementary Table 4 is available at *Carcinogenesis* Online). The remaining patients (977 aggressive and 987 nonaggressive cases; JHH2) were genotyped on the Illumina 610K chip platform for the follow-up confirmation stage (supplementary Table 5 is available at *Carcinogenesis* Online).

A second group of subjects used for confirmation came from the National Cancer Institute (NCI) Cancer Genetic Markers of Susceptibility (CGEMS) Study. Details for this study are described elsewhere (4,8). For the current study, all publicly available data from the genome-wide scan ($n = 1175$ PCa patients of European descent) were included. Patients with a Gleason score of 7 or higher, or a stage of T3/T4, were classified as having aggressive disease ($n = 687$), whereas all other patients were classified as having nonaggressive disease ($n = 488$).

A third group of subjects used for confirmation included PCa patients from the CAPS study. Details of this study are described elsewhere (16). Briefly, PCa

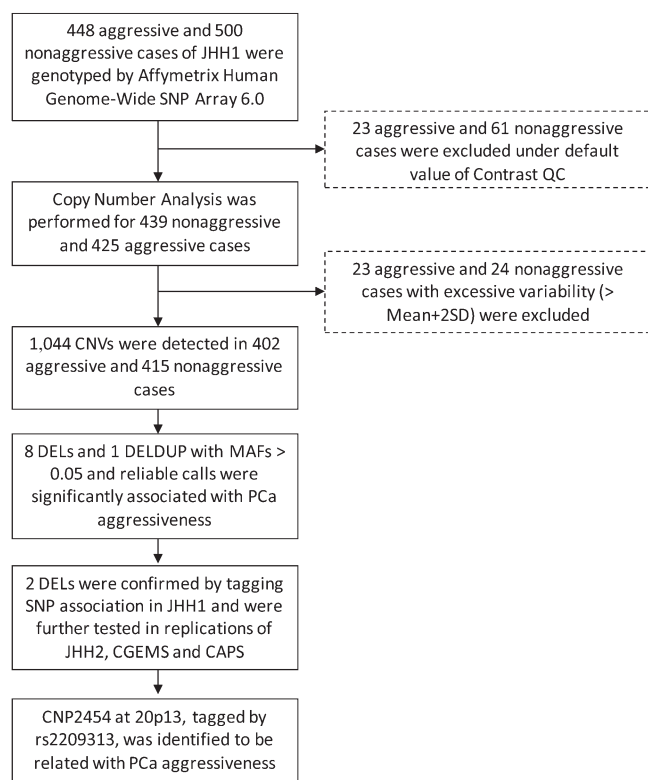


Fig. 1. Study design for genome-wide CNVs analysis and test of association with PCa aggressiveness in JHH1 population, followed by replications in JHH2, CGEMS and CAPS populations.

patients diagnosed between July 2001 and October 2003 were recruited from four regional cancer registries in Sweden. Patients were classified as having aggressive disease ($n = 1231$) if their tumors had a clinical stage of T3/T4, N+, M, Gleason score of 8 or higher or a serum prostate-specific antigen level of >50 ng/ml. Otherwise, the patients were classified as having nonaggressive disease ($n = 1619$) (supplementary Table 6 is available at *Carcinogenesis* Online).

This study was approved by the Institutional Review Board at Wake Forest University School of Medicine, Johns Hopkins University School of Medicine and the Karolinska Institutet (Sweden).

CNVs genotyping

The Affymetrix Human Genome-Wide SNP Array 6.0 (Affymetrix, Santa Clara, CA) was used to perform genome-wide SNPs and CNVs genotyping at the Cancer Genomics, Wake Forest University. As shown in Figure 1, 23 aggressive and 61 nonaggressive samples from the discovery set were excluded from further analysis because they were below the default threshold of Contrast quality control as calculated by a genotyping call rate of at least 97% and an accuracy of at least 99%. Data processing and CNV calling in predefined genomic regions of CNVs were performed using Genotyping Console 4.0 (GTC 4.0) based on the CANARY algorithm developed by the Broad Institute (25). GTC 4.0 was used for CNV calls among the remaining 425 aggressive and 439 nonaggressive cases. GTC 4.0 uses a set of 1141 common CNV regions defined in HapMap samples (also named as copy-number polymorphism, CNP) (26), which are a subset of CNV that is filtered to ensure that each region is mapped by more than one smart probe set in order to reduce sample-to-sample variability. Each sample in a predefined region is assigned a copy number state of 0–4; CNV states of 5 or more are included in state 4. Subsequent to CNV calling, 23 aggressive and 24 nonaggressive samples were excluded from further analysis because of excessive variability, with the total number of CNVs that were greater than the mean plus two standard deviations (SD), resulting in 402 aggressive and 415 nonaggressive cases for the discovery phase analysis.

SNP genotyping and imputation

Genotyping in JHH2 subjects was performed using the Illumina 610K chip (Illumina, San Diego, CA) at Wake Forest University and the data cleaning included: call rate $<95\%$, MAF <0.01 and Hardy–Weinberg Equilibrium

(HWE) test P -value <0.001 . The genotyping data for CGEMS subjects were downloaded from the publicly accessible CGEMS website (<http://cgems.cancer.gov/>). Genotyping data for CAPS subjects were obtained using the MassARRAY iPLEX genotyping system (Sequenom, San Diego, CA) at Wake Forest University. Duplicate test samples and two water samples [polymerase chain reaction (PCR)-negative controls] that were blinded to the technician were included in each 96-well plate.

Using the IMPUTE computer program (27), genotypes for JHH1, JHH2 and CGEMS were imputed according to HapMap Phase II (www.hapmap.org) with a posterior probability of 0.9 as a threshold. The following quality control criteria were used to filter SNPs: MAF <0.01 , Hardy–Weinberg Equilibrium <0.001 and call rate <0.95 .

PCR and quantitative PCR

We examined CNV states using PCR with two sets of PCR primers: CNP2454 (5'-TAAAGCCAACGCCGTAGAGTCTGT-3', 5'-CCCAAGGATGTCAATTTGGCCAGT-3') for an amplicon of 141 bp, residing in CNP2454; and oligo 2422 (5'-ATGGCAGGGACCTGATTGACAGAA-3', 5'-ATGTG-GAGCAGGCATGAAGGTAGT-3') for an amplicon of 296 bp. The latter amplicon was used to demonstrate the integrity of the DNA. The absence of the 141 bp fragment indicated the homozygous deletion (zero copies) of CNP2454. To differentiate between two copies (normal) and one copy (hemizygous deletion) of CNP2454, we performed quantitative polymerase chain reaction (qPCR) using the same primers for CNP2454 and another pair of primers located in a diploid region (GAPDH) as a reference for normal copy number (5'-TCCTCATGCCTTCTTGCCTTGT-3', 5'-AGGCGCCAA-TACGACCAAATCTA-3'). Copy numbers were determined according to the $\Delta\Delta CT$ method; the $\Delta\Delta CT$ is defined as the difference between the threshold cycle of the CNP2454 amplicon and that of the reference GAPDH.

Statistical analysis

Logistic regression models were used to test for association of CNVs with PCa aggressiveness and to estimate the corresponding odds ratios (ORs) and 95% confidence interval (95% CI) assuming a log-additive genetic model with 1 d.f. for each set of subjects. Pooled OR were derived using ORs and 95% CIs of individual study populations in a Mantel–Haenszel analysis. The heterogeneity of ORs across the study populations was assessed using the Q -test.

To reduce false-positive associations due to low quality CNV calls, CNVtools (28) was used to estimate the number of copy-number classes at each significant CNV using principal components analysis (supplementary Figure 1 is available at *Carcinogenesis* Online). The CNV states of samples examined by qPCR were also assigned by using CNVtools based on values of $\Delta\Delta CT$.

Tagging SNPs for CNVs were identified using Haploview (version 4.2) on the basis of CNVs states and the genotypes of nearby SNPs in the JHH1 population. The SNPs with the highest r^2 value for the corresponding CNVs were selected as tagging SNPs for the respective CNVs.

Results

The study design is shown in Figure 1. Among 1141 predefined CNV regions according to HapMap, 1044 CNVs were detected in our samples, of which 691 CNVs were deletions (DELs, diploid copy numbers ≤ 2), 184 CNVs were duplications (DUPs, diploid copy numbers ≥ 2) and 169 CNVs were both deletion and duplication (DELDUPS). There were 520 CNVs with MAF ≥ 0.01 and 235 CNVs with MAF ≥ 0.05 (Table I).

Considering the statistical power under the moderate sample size in the discovery stage, we restricted the association analysis to common CNVs (MAF ≥ 0.05). As shown in supplementary Table 1 is available at *Carcinogenesis* Online, we initially found 33 CNVs, including 26 DELs, 1 DUP and 6 DELDUPS, that were significantly associated with PCa aggressiveness ($P < 0.05$). Among these CNVs, eight DELs and one DELDUP were high-quality CNV calls that showed well separated copy number clusters in plots made by CNVtools as seen in Table II and supplementary Figure 1 is available at *Carcinogenesis* Online.

The feasibility of using SNPs to tag CNVs of interest was examined by performing LD analysis in the JHH1 population. We found that four of the nine CNVs could be tagged by at least one SNP with high LD ($r^2 > 0.90$, supplementary Table 2 is available at *Carcinogenesis* Online). When we then assessed these four tagging SNPs for association with aggressiveness of PCa in the JHH1 population (448 aggressive

and 500 nonaggressive cases), only rs1547024 (tagging CNP992; $r^2 = 0.98$) and rs2209313 (tagging CNP2454; $r^2 = 0.99$) were significantly associated with PCa aggressiveness ($P < 0.05$).

Next, we evaluated the association of SNPs rs1547024 and rs2209313 with PCa aggressiveness in JHH2, CGEMS and CAPS populations with an additional 2895 aggressive and 3094 nonaggressive cases to confirm the associations of CNP992 and CNP2454 with

PCa aggressiveness. As shown in Table III, only rs220913 remained consistently associated with PCa aggressiveness in confirmation studies ($P = 1.77 \times 10^{-3}$). After we combined data from the four populations (JHH1, JHH2, CGEMS and CAPS), the T allele of rs2209313, representing the presence of the CNP2454 segment, was significantly associated with a 1.17-fold (95% CI: 1.07–1.27) increased risk of aggressive disease for PCa with a P value of 2.75×10^{-4} . For SNP

Table I. Summary of CNV calls among PCa cases

MAF	All cases ^a				Aggressive ^a				Nonaggressive ^a			
	DEL	DUP	DELDUP	Total	DEL	DUP	DELDUP	Total	DEL	DUP	DELDUP	Total
<0.01	363	118	43	524	302	108	34	444	319	100	38	457
0.01–0.05	126	54	105	285	129	60	96	285	126	58	101	285
>0.05	202	12	21	235	202	12	21	235	202	12	21	235
Total	691	184	169	1044	634	180	150	964	647	171	159	977

^aCNVs were classed as deletion (DEL), duplication (DUP) and multiallelic CNV (DELDUP).

Table II. CNVs with reliable calls significantly associated with PCa aggressiveness in the JHH1 population

CNVs	Type ^a	Chr	Gene	Aggressive (copy number)			Nonaggressive (copy number)			MAF		P -value ^b	OR (95% CI) ^b
				2	1	0	2	1	0	Aggressive	Nonaggressive		
				CNP385 ^c	DELDUP	3	NA	327	71	4	364		
CNP422	DEL	3	NA	363	39	0	357	56	2	0.049	0.072	0.0431	0.65(0.43–0.99)
CNP668	DEL	4	<i>QRFPR</i>	368	33	1	362	51	2	0.044	0.066	0.0471	0.64(0.42–0.99)
CNP992	DEL	6	NA	104	205	93	136	196	83	0.486	0.436	0.0436	1.22(1.01–1.48)
CNP1294	DEL	8	NA	340	59	3	373	41	1	0.081	0.052	0.0198	1.61(1.08–2.39)
CNP2203	DEL	16	<i>WWOX</i>	97	200	105	122	212	81	0.510	0.451	0.0161	1.27(1.05–1.55)
CNP12610	DEL	18	<i>DOK6</i>	202	184	16	239	170	6	0.269	0.219	0.0107	1.39(1.08–1.78)
CNP2422	DEL	19	<i>PSG4</i>	341	55	6	372	39	4	0.083	0.057	0.0465	1.45(1.01–2.10)
CNP2454 ^d	DEL	20	<i>SIRPBI</i>	14	142	246	8	128	279	0.211	0.173	0.0447	1.30(1.01–1.68)

^aCNV types included deletion (DEL), duplication (DUP) and multiallelic CNV (DELDUP).

^b P -values and ORs were calculated by assuming a log-additive genetic model (1 d.f.).

^cThree samples with copy number of 3 were combined with the group of two copies.

^dThe OR and 95% CI were estimated based on the reference group of homodeletion (0 copies).

Table III. Tagging SNPs for CNP992 and CNP2454 and their associations with PCa aggressiveness in JHH1, JHH2, CGEMS and CAPS populations

Marker	Variation	Population	Minor allele	Aggressive (minor allele copy number)			Nonaggressive (minor allele copy number)			MAF		OR(95%CI) ^a	P^a
				0	1	2	0	1	2	Aggressive	Nonaggressive		
				CNP992 (rs1547024)	Deletion (T > C)	JHH1 ($n = 817$)	DEL	104	205	93	136		
		JHH1 ($n = 931$)	C	119	218	104	159	235	96	0.483	0.436	1.21(1.01–1.45)	0.043
		Confirmation											
		JHH2 ($n = 1849$)	C	268	450	209	279	446	197	0.468	0.456	1.05(0.93–1.19)	0.446
		CGEMS ($n = 1136$)	C	157	335	170	137	239	98	0.510	0.459	1.23(1.04–1.46)	0.016
		CAPS ($n = 2796$)	C	344	589	265	453	754	391	0.467	0.481	0.95(0.85–1.05)	0.323
		Subtotal										1.06(0.92–1.22)	0.435
		All population ^b										1.09(0.96–1.23)	0.190
CNP2454 (rs2209313)	Deletion (C > T)	JHH1 ($n = 817$)	nonDEL ^c	246	142	14	279	128	8	0.211	0.173	1.30(1.01–1.68)	0.045
		JHH1 ($n = 938$)	T	272	156	17	330	151	12	0.214	0.178	1.27(1.01–1.61)	0.045
		Confirmation											
		JHH2 ($n = 1896$)	T	564	344	42	600	300	46	0.225	0.207	1.11(0.95–1.30)	0.177
		CGEMS ($n = 1168$)	T	400	252	29	309	159	19	0.228	0.202	1.17(0.95–1.44)	0.136
		CAPS ($n = 2826$)	T	768	385	62	1074	484	53	0.210	0.183	1.18(1.03–1.34)	0.014
		Subtotal										1.15(1.05–1.26)	1.77E-03
		All population ^b										1.17(1.07–1.27)	2.75E-04

^a P -values and ORs were calculated by assuming a log-additive genetic model (1 d.f.).

^b P value and OR (95% CI) were derived from Mantel–Haenszel analysis.

^cAbsence of this segment was the minor allele for CNP2454.

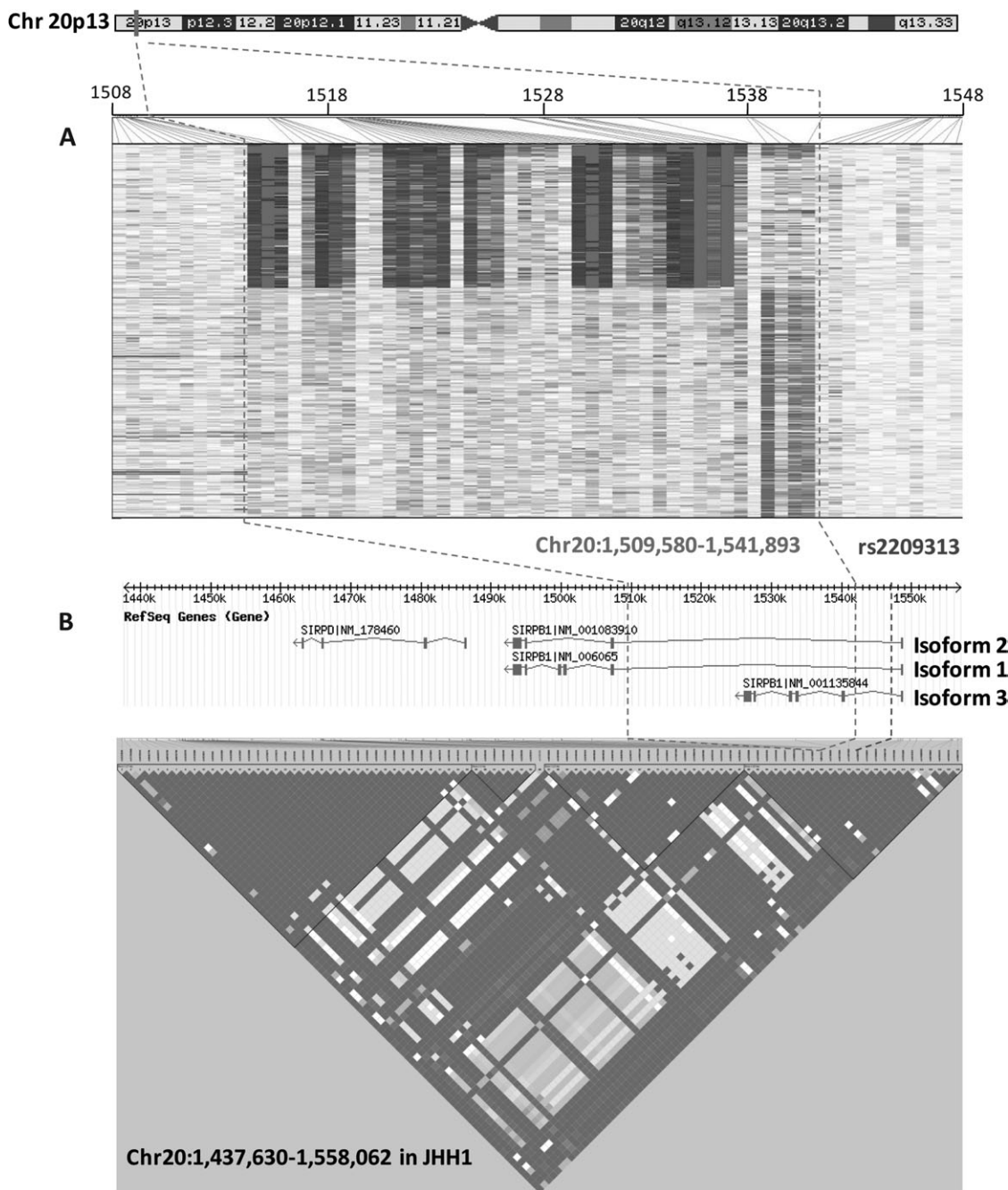


Fig. 2. CNP2454 and tagging SNP rs2209313 on chromosome 20p13. Illustration of CNP2454 calling by GTC4.0 (white: 0 copy; grey: 1 copy and dark grey: 2 copy) (Chart A); CNP2454 and its tagging SNP rs2209313 are located in the *SIRPB1* gene and map to a 120 kb LD region (Chart B).

rs1547024, we failed to observe evidence for replication ($P = 0.435$ for all confirmation subjects).

In addition, we also analyzed the association of rs2209313 with other specific clinicopathologic variables, including Gleason score and TNM stage, in JHH and CAPS populations. As shown in supplementary Table 3 is available at *Carcinogenesis* Online, rs2209313 was significantly associated with Gleason score ($P = 0.023$) in the JHH population but not with T or N stages. In CAPS, rs2209313 was associated with T (0.002) and N stages (0.027) but not with Gleason score.

To provide further evidence that rs2209313 can be used as a tagging SNP for CNP2454, we estimated the LD level between the rs2209313 genotype and CNV states of CNP2454 as determined by PCR and qPCR assays in 44 CAPS subjects (supplementary Figure 2 is available at *Carcinogenesis* Online). The sample numbers assigned for

states 0, 1 and 2 were 28, 12 and 4, respectively, and the r^2 value was 0.93 between CNP2454 and rs2209313.

Discussion

In this case–case study, we first investigated the relationship of common CNVs with aggressiveness of PCa in a cohort consisting of 448 aggressive and 500 nonaggressive PCa cases using Affymetrix’s 6.0 arrays. In the confirmation phase, we used a tagging SNP approach to replicate one candidate CNV (CNP2454) at 20p13 in three larger and independent populations. To our knowledge, this is the first report on the relationship of CNVs with aggressiveness of PCa.

Genome-wide association studies have identified >30 loci for PCa risk (2–13); however, until recently, none of these has been reproducibly related to PCa aggressiveness (29–31). This may reflect the

inability of case–control studies to uncover the genetic determinants of PCa progression. Using a case–case design, our group has recently identified an SNP that can consistently distinguish aggressive PCa from indolent cases in multiple populations (16), thus indicating the utility of case–case studies in the identification of PCa aggressiveness related loci. In the current study, we have used CNVs analyses to identify an additional common locus that is consistently associated with PCa aggressiveness. Our new findings suggest that CNP2454, or tagging SNP rs2209313, have potential utility in distinguishing the PCa patients at highest risk of aggressive PCa. In addition, our results lend additional support for the feasibility of applying case–case designs when searching for genetic variants that are responsible for PCa aggressiveness.

It is noteworthy that many common diallelic CNVs may be adequately represented by adjacent-tagging SNPs and that well-tagged CNPs can be detected indirectly via genotyping SNPs (25). For example, The Wellcome Trust Case Control Consortium (WTCCC) recently conducted a genome-wide association study of CNVs for eight common diseases and identified three CNVs that were associated with common diseases, including Crohn's disease, rheumatoid arthritis and diabetes (22). However, all of the three loci were previously identified in SNP-based studies. Using an SNP–CNP LD approach, Willer *et al.* (23) assessed 261 SNPs that tagged CNPs ($r^2 > 0.8$) for their association with body mass index and identified rs2815752, an SNP that tagged a 45 kb deletion polymorphism with perfect LD, near the *NEGR1* gene as a strongly associated with body mass index. Our results are consistent with the idea that many of the common CNVs can be tagged by surrounding SNPs such as rs2209313 that tags CNP2454 with an r^2 of 0.98. Thus, interrogating SNPs that tag common CNVs is a reliable, rapid and cost-effective approach for association studies involving common CNVs.

Based on UCSC Genome Browser (32) Build 36 (March 2006), CNP2454 is located on human chromosome 20p13 and is a deletion polymorphism within the signal-regulatory protein beta-1 gene (*SIRPB1*) (Figure 2). *SIRPB1* has three isoforms, all of which are affected by CNP2454; CNP2454 is entirely within intron 1 of isoforms 1 and 2 but deletes the majority of isoform 3 from intron 1 to the end of the gene. CNP2454 is localized to a big LD region spanning ~120 kb (Figure 2). The tagging SNP for CNP2454, rs2209313, is located in intron 1 for all three isoforms of *SIRPB1*. Although we could not rule out the biologically functional relevance of the SNPs in this LD region, CNP2454 is more probably to impose a greater impact on the function of *SIRPB1* because it knocks out all exons except for exon 1 of isoform 3; for the other 2 isoforms, the effects are not as dramatic since the CNV removes only a 32 kb segment within intron 1. There is evidence that ionic interaction between SIRPβ1 and adapter protein DAP12 leads to phosphorylation of spleen tyrosine kinase (Syk) and mitogen-activated protein kinase and further promotes phagocytosis by macrophages and migration of neutrophils (33,34), although its direct role in tumorigenesis is unknown. Further functional assays are needed to identify the biological mechanism of CNP2454 and rs2209313 that can help explain their relevance to PCa aggressiveness.

Although rare CNVs have been suggested to play an important role in susceptibility for various complex diseases (35–38), relatively few associations have been reported for common CNVs and common diseases. This is due, in part, to lack of robust and cost-effective assays for large-scale CNVs genotyping. No single approach works well for typing every CNV (22). The objective in CNV calling at each CNV is to assign each assayed sample to a diploid copy member class. This step is analogous to, but typically considerably more challenging than, calling genotypes from SNP-chip data (22). To address concerns on the quality of CNV calls and the potential impact on statistical significance in association testing, we restricted our analysis to predefined regions in HapMap (26) and further confirmed candidate CNVs using another analytic tool, CNVtools. In the confirmation phase, we used tagging SNPs as surrogates for the CNVs of interest and then tested their association in independent populations with large numbers of samples. Finally, we validated

the signal of CNP2454 by PCR and qPCR and confirmed the LD pattern between CNP2454 and tagging SNP rs2209313. Together, these different approaches allowed us to critically assess and confirm our findings.

Limitations of this study should be noted. First, though we performed a genome-wide CNVs analysis, we only focused on common CNVs (MAF > 0.05) predefined in HapMap due to a moderate sample size in the discovery stage. This strategy misses rare CNVs and other common CNVs that were not included in the predefined regions. Secondly, although most CNVs are well tagged by adjacent SNPs (26), there are some CNVs that are insufficiently tagged by surrounding SNPs. For example, in our study, CNP2422 was best tagged by rs4803596 with an r^2 of only 0.42. This may have led to our inability to reproduce some of the promising results obtained in the discovery phase by tagging SNPs. Finally, in spite of the consistent findings in the multiple Caucasian populations from American and Sweden, the association of genetic variants at CNP2454 with aggressiveness of PCa did not reach the significance level ($0.05/253 = 1.98 \times 10^{-4}$) after Bonferroni correction. Sample size may be the primary factor that limits this *P* value. Furthermore, it has been argued that the appropriate *P* value thresholds depend on the prior probability of association at each locus but not the number of tests performed (39).

In summary, we identified a CNV, CNP2454 at 20p13 that was associated with PCa aggressiveness. The result was confirmed by analysis of a tagging SNP, rs2209313, in the original and three independent populations. This CNV is a 32.3 kb deletion polymorphism in the *SIRPB1* gene that may differentially affect the three different isoforms of the gene. Future studies are needed to evaluate the biological significance of our findings in PCa development.

Supplementary material

Supplementary Tables 1–6 and Figures 1 and 2 can be found at <http://carcin.oxfordjournals.org/>

Funding

National Institutes of Health (CA129684 to J.X., CA131338 to L.S.Z.); the Department of Defense Grant (W81XWH-07-1-0088 to J.X.); the Swedish Cancer Society (Cancerfonden) to H.G.; the Swedish Academy of Sciences (Vetenskapsrådet) to H.G.; the support of Kevin P. Jaffe to W.B.I. is gratefully acknowledged.

Acknowledgements

The authors thank all the study subjects who participated in the CAPS study and urologists, who included their patients in the CAPS study. 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, Alan W. Partin, and Patrick C. Walsh (for JHH).

The authors also thanks for the National Cancer Institute Cancer Genetic Markers of Susceptibility Initiative (CGEMS) for making the data available publicly.

Conflict of Interest Statement: None declared.

References

- Manolio, T.A. (2010) Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.*, **363**, 166–176.
- Amundadottir, L.T. *et al.* (2006) A common variant associated with prostate cancer in European and African populations. *Nat. Genet.*, **38**, 652–658.
- Gudmundsson, J. *et al.* (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.*, **39**, 631–637.
- Yeager, M. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, **39**, 645–649.
- Gudmundsson, J. *et al.* (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.*, **39**, 977–983.

6. Duggan,D. *et al.* (2007) Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene *DA-B2IP*. *J. Natl Cancer Inst.*, **99**, 1836–1844.
7. Thomas,G. *et al.* (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, **40**, 310–315.
8. Gudmundsson,J. *et al.* (2008) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat. Genet.*, **40**, 281–283.
9. Eeles,R.A. *et al.* (2008) Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.*, **40**, 316–321.
10. Yeager,M. *et al.* (2009) Identification of a new prostate cancer susceptibility locus on chromosome 8q24. *Nat. Genet.*, **41**, 1055–1057.
11. Gudmundsson,J. *et al.* (2009) Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat. Genet.*, **41**, 1122–1126.
12. Eeles,R.A. *et al.* (2009) Identification of seven new prostate cancer susceptibility loci through a genome-wide association study. *Nat. Genet.*, **41**, 1116–1121.
13. Takata,R. *et al.* (2010) Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population. *Nat. Genet.*, **42**, 751–754.
14. Xu,J. *et al.* (2009) Estimation of absolute risk for prostate cancer using genetic markers and family history. *Prostate*, **69**, 1565–1572.
15. Witte,J.S. (2009) Prostate cancer genomics: towards a new understanding. *Nat. Rev. Genet.*, **10**, 77–82.
16. Xu,J. *et al.* (2010) Inherited genetic variant predisposes to aggressive but not indolent prostate cancer. *Proc. Natl Acad. Sci. USA.*, **107**, 2136–2140.
17. Conrad,D.F. *et al.* (2010) Origins and functional impact of copy number variation in the human genome. *Nature*, **464**, 704–712.
18. McCarroll,S.A. *et al.* (2008) Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nat. Genet.*, **40**, 1107–1112.
19. de Cid,R. *et al.* (2009) Deletion of the late cornified envelope *LCE3B* and *LCE3C* genes as a susceptibility factor for psoriasis. *Nat. Genet.*, **41**, 211–215.
20. Diskin,S.J. *et al.* (2009) Copy number variation at 1q21.1 associated with neuroblastoma. *Nature*, **459**, 987–991.
21. Glessner,J.T. *et al.* (2010) Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl Acad. Sci. USA.*, **107**, 10584–10589.
22. Wellcome Trust Case Control Consortium *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, **464**, 713–720.
23. Willer,C.J. *et al.* (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.*, **41**, 25–34.
24. Liu,W. *et al.* (2009) Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res.*, **69**, 2176–2179.
25. Korn,J.M. *et al.* (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.*, **40**, 1253–1260.
26. McCarroll,S.A. *et al.* (2010) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*, **40**, 1166–1174.
27. Marchini,J. *et al.* (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
28. Barnes,C. *et al.* (2008) A robust statistical method for case-control association testing with copy number variation. *Nat. Genet.*, **40**, 1245–1252.
29. Kader,A.K. *et al.* (2009) Individual and cumulative effect of prostate cancer risk-associated variants on clinicopathologic variables in 5,895 prostate cancer patients. *Prostate*, **69**, 1195–1205.
30. Kote-Jarai,Z. *et al.* (2008) Multiple novel prostate cancer predisposition loci confirmed by an international study: the PRACTICAL Consortium. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 2052–2061.
31. Fitzgerald,L.M. *et al.* (2009) Analysis of recently identified prostate cancer susceptibility loci in a population-based study: associations with family history and clinical features. *Clin. Cancer Res.*, **15**, 3231–3237.
32. Kent,W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
33. Kharitonov,A. *et al.* (1997) A family of proteins that inhibit signalling through tyrosine kinase receptors. *Nature*, **386**, 181–186.
34. Barclay,A.N. *et al.* (2006) The SIRP family of receptors and immune regulation. *Nat. Rev. Immunol.*, **6**, 457–464.
35. Stefansson,H. *et al.* (2008) Large recurrent microdeletions associated with schizophrenia. *Nature*, **455**, 232–236.
36. Glessner,J.T. *et al.* (2009) Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*, **459**, 569–573.
37. Greenway,S.C. *et al.* (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat. Genet.*, **41**, 931–935.
38. Pinto,D. *et al.* (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, **466**, 368–372.
39. Wellcome Trust Case Control Consortium. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.

Received February 28, 2011; revised February 28, 2011;
accepted April 14, 2011