Published in final edited form as:

Urol Oncol. 2014 January ; 32(1): 44.e15–44.e20. doi:10.1016/j.urolonc.2013.06.004.

Validation of copy number variants associated with prostate cancer risk and prognosis

August Blackburn^{#a}, Desiree Wilson^{#a}, Jonathan Gelfond, MD, PhD^{b,e}, Li Yao, MD, PhD^c, Javier Hernandez, MD^{d,e}, Ian M Thompson, MD^{d,e}, Robin J Leach, PhD^{a,d,e}, and Donna M Lehman, PhD^{c,e}

^a University of Texas Health Science Center at San Antonio, Department of Cellular and Structural Biology

^b University of Texas Health Science Center at San Antonio, Department of Epidemiology and Biostatistics

^c University of Texas Health Science Center at San Antonio, Department of Medicine – Division of Clinical Epidemiology

^d University of Texas Health Science Center at San Antonio, Department of Urology

^e University of Texas Health Science Center at San Antonio, Cancer Therapy and Research Center

[#] These authors contributed equally to this work.

Abstract

Objective—Two recent studies have reported novel heritable copy number variants (CNVs) on chromosomes 2p, 15q, and 12q to be associated with prostate cancer risk in non Hispanic Caucasians. The goal of the current study was to determine whether these findings could be independently confirmed in the Caucasian population from the South Texas area.

Methods and Materials—The study subjects consisted of participants of the San Antonio Biomarkers Of Risk for prostate cancer (SABOR) cohort and additional cases ascertained in the same metropolitan area. We genotyped all 7 of the reported copy number variants using real time quantitative PCR in 1536 (317 cases, 1219 controls) non Hispanic Caucasian men, and additionally genotyped 632 (191 cases, 441 controls) Hispanic Caucasian men for one of these variants, a deletion on 2p24.3.

Results—Association of the deletion on 2p24.3 with overall prostate cancer risk did not meet our significance criteria but was consistent with previous reports [odds ratio (OR), 1.40; 95%

Corresponding author: Donna Lehman, Ph.D. Department of Medicine – Clinical Epidemiology UTHSCSA 7702 Floyd Curl Drive San Antonio, TX 78229 Office phone: (210) 567 6714 lehman@uthscsa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All other authors can be reached through correspondence with the corresponding author.

There are no personal or other conflicts of interest that need to be disclosed.

confidence interval (95% CI), 0.99 2.00; $P = 0.06$]. Among Hispanic Caucasians, this deletion is much less prevalent (MAF of 0.059 and 0.024 in non Hispanic and Hispanic Caucasians respectively) and did not show evidence of association with risk for prostate cancer. Interestingly, among non Hispanic Caucasians, carrying a homozygous deletion of 2p24.3 was significantly associated with high grade prostate cancer as defined by a Gleason sum ≥ 8 [OR, 27.99; 95% CI, 1.99 392.6; $P = 0.007$ (Fischer's Exact)]. The remaining six copy number variable regions were either not polymorphic in our cohort of non Hispanic Caucasians, or showed no evidence of association.

Conclusions—Our findings are consistent with the reported observation that a heritable deletion on 2p24.3 is associated with prostate cancer risk in non Hispanic Caucasians. Additionally, our observations indicate that the 2p24.3 variant is associated with risk for high grade prostate cancer in a recessive manner. We were unable to replicate any association with prostate cancer for the variants on chromosomes 15q and 12q which may be explained by regional population differences in low frequency variants and disease heterogeneity.

Keywords

Prostate; cancer; risk; deletion; prognosis

Introduction

Prostate cancer is the most common non skin cancer in American men. The majority of prostate cancer cases, however, is indolent and may not require treatments that are associated with significant rates of voiding and sexual function complications. There is great impetus to identify markers that distinguish indolent from aggressive disease. Given that risk for prostate cancer may be attributed to a heritable component in as many as 42% of cases (1), it may be possible to identify genetic polymorphisms that can be used to better predict risk and prognosis. Multiple studies have attempted to identify genetic variants that are associated with risk for prostate cancer.

The contribution of copy number variants (CNV) to risk for complex diseases have not yet been fully elucidated, partially due to the difficulty of identification and accurate genotyping (2). Efforts to characterize common and rare copy number variation in various human populations are still underway. Common CNVs have been shown to be in linkage disequilibrium with adjacent single nucleotide polymorphisms (SNPs) (3 5), which indicates that some common CNVs have already been indirectly assessed for association with traits in SNP based genome wide association studies (GWAS). However, recurrent variants and risk bearing alleles with low minor allele frequencies may not be well tagged by SNPs in GWAS, and these variants will likely require direct assessment.

Few studies have investigated risk for prostate cancer attributed to CNVs directly. Liu *et al.* published the first genome wide investigation of germ line CNVs and risk for prostate cancer, in which they report a germ line deletion spanning 5,947 base pairs at 2p24.3 associated with risk for prostate cancer and aggressive prostate cancer among men of European descent (6). Demichelis *et al.* reported 6 deletions associated with risk for prostate cancer with a false discovery rate <0.2 (7). Two of these variants, deletions on 15q and 12q,

were further supported in this same publication with evidence from bioinformatics analysis and functional assays (7).

Genome wide association studies for any genetic variant type are subject to a range of errors and biases, and therefore replication of association signals in independent samples are necessary to confirm true associations(8). One of the primary purposes of the San Antonio Center for Biomarkers of Risk of Prostate Cancer (SABOR), a Clinical and Validation Center of the Early Detection Research Network of the National Cancer Institute, is to independently confirm prostate cancer biomarkers, including genetic variants that are predictive of risk. With this motivation, we report the investigation of these reported variants in non Hispanic and Hispanic Caucasians from San Antonio, Texas.

Materials and Methods

Study subjects

Study subjects consisted of 1372 (153 cases, 1219 controls) non Hispanic Caucasian and 516 (75 cases, 441 controls) Hispanic Caucasian participants of SABOR. SABOR is a prospective longitudinal study that examines behavioral, genetic, and other markers of risk of prostate cancer. All study participants are screened annually for prostate cancer using prostate specific antigen (PSA) serum measurements, with exception to those with PSA <1, who are screened every other year. An additional 164 non Hispanic Caucasian and 116 Hispanic Caucasian men were recruited from a parallel cohort study of prevalent cases (PREF) in the same metropolitan area. This study consists of men who were diagnosed with prostate cancer (PCa) prior to enrollment in the study and were recruited using the same methods as SABOR (local advertisement). Age for this study was calculated in the following manner: for prevalent cases, self reported age of diagnosis was used; for incident cases, age of diagnosis was used; non cancer (control) participants were censored at their most recent SABOR examination age. Self reported ages of diagnosis for most of the prevalent cases were confirmed through medical records. Institutional Review Board approval was obtained from the UT Health Science Center at San Antonio. Informed consent was obtained from all participants of both cohorts. We refer to the total population as the SABOR cohort.

Liu *et al.* defined cancer aggressiveness as meeting any of the following criteria: T_{3/4}, N+, M+, Gleason score sum ≥ 8, or PSA >50 ng/mL(6). Although it would be ideal to use identical criterion to define aggressiveness, PSA serum levels at time of diagnosis as well as staging were not available for all participants of PREF, and thus these variables were not utilized in defining aggressiveness. Gleason scores were more broadly available, and thus, in this study aggressive cases in this study were defined by a Gleason score of ≥ 8.

CNV genotyping

DNA was isolated from whole blood using QIAamp DNA Blood Maxi Kit (Qiagen, Valencia, CA). For real time quantitative PCR, all primers and probes were purchased from Applied Biosystems (Valencia CA). The primer/probe pairs were either pre designed or

were designed for the regions of interest using Primer Express (Applied Biosystems, Valencia CA). Quantitative PCR primer/probe information is summarized in Table 1. .

For qPCR, primers and probes for a target sequence and reference sequence (RNAse P) were multiplexed in 384 well plates and all samples were run in duplicate. Fluorescence was detected using the 7900HT Real time PCR System (Applied Biosystems, Valencia CA). Real time PCR data was analyzed using the reference free $\Delta\Delta C_t$ method implemented in CopyCaller software (Applied Biosystems, Valencia CA). This approach estimates the mean ΔC_t for a copy number of 1, and subsequently uses this value to calculate $\Delta\Delta C_t$. We have found this method broadly consistent with the $\Delta\Delta C_t$ method using a reference sample. Discrete copy number calls were determined by plotting histograms of the raw calculated copy number values, which for polymorphic regions reveals non overlapping Gaussian distributions representing integer copy number states(9).

To confirm that CNV calls from qPCR accurately detected the 2p24.3 deletion identified by Liu *et al.* (6), we conducted PCR genotyping as reported by Liu *et al.* (6) for all homozygous deletions, 12 heterozygous deletions, and 14 wild type individuals, including 4 individuals on the Gaussian tails where the distributions approach each other in the non Hispanic Caucasian samples. We observed 100% concordance with the real time PCR results. An example of the genotyping analysis is shown in Figure 1.

Statistical Analysis

Power analyses for this study were conducted using Power for Genetic Association Analysis (PGA) (10) with the estimated disease frequency of 0.16, the minor allele frequency, estimated relative risk reported for each variant, and $\alpha = 0.05$.

Tests for deviation from Hardy Weinberg equilibrium were conducted separately for the non Hispanic Caucasian and Hispanic Caucasian samples. Frequency differences between cases and controls were tested using logistic regression with adjustment for age. Due to low counts, tests for association with aggressive prostate cancer were conducted using Fisher's exact test for a 2 by 2 contingency table under both dominant and recessive models.

Results

We genotyped 7 deletion variants previously reported to be associated with risk for prostate cancer using real time quantitative PCR in 1536 (317 cases, 1219 controls) non Hispanic Caucasian men, with an average of 1468 men genotyped per assay after applying strict genotyping criteria. The mean age of diagnosis was 66.7 ± 8.3 and the mean age of controls was 66.5 ± 9.5 . Using this dataset, we estimated power to detect association with 6 variants reported by Demichelis *et al.*(7) at >0.9 , and ≈ 0.37 for the 2p24.3 variant reported by Liu *et al.* (6) with an α of 0.05. Four of 6 CNVRs reported to be associated by Demichelis *et al.*(7) were not observed to be polymorphic. Variants on 12q21.31 and 19q13.12 were polymorphic (MAF ≈ 0.10 , 0.21 respectively), but showed no evidence for association with risk for prostate cancer (12q21.31: OR, 0.89; 95% CI, 0.28 2.77; P=0.84) (19q13.12: OR, 1.08; 95% CI, 0.87 1.34; P=0.49).

The deletion on 2p24.3 reported by Liu *et al.* (6) had a minor allele frequency (MAF) of 0.059 in non Hispanic Caucasians, and was not associated with risk of prostate cancer given our predetermined of 0.05 (OR, 1.40; 95% CI, 0.99 2.00; P = 0.06). However, these results are consistent with the reported MAF, effect size, and confidence intervals reported by Liu *et al* using a dominant model. Among SABOR participants, carrier frequency was not associated with high grade prostate cancer when compared to controls under a dominant model [OR, 1.28; 95% CI, 0.43 3.12; P = 0.62 (Fischer's Exact)], but was significantly associated with high grade prostate cancer under a recessive model (OR, 27.99; 95% CI, 1.99 392.6; P = 0.007). Two (4.44%) of 45 men with high grade prostate cancer carried homozygous deletions, compared to 2 (0.16%) men carrying homozygous deletions in 1219 controls. The 2 control carriers for the deletion were aged 43.4 and 55.35 years, which is relatively young for developing prostate cancer. Additionally, carrying a homozygous deletion is also significantly associated with prostate cancer grade with 2 out of 45 high grade cases carrying a homozygous deletion compared to 0 out of 272 low grade cases (OR, INF; 95% CI, 1.15 INF, P=0.02). The 2 case carriers had observed Gleason scores of 9 and 10. When considering controls and non aggressive cases together the enrichment for homozygous deletion in aggressive cases is strong (OR, 34.3; 95% CI, 2.43 477; P=0.005).

We genotyped the 2p24.3 deletion in 632 (191 cases, 441 controls) Hispanic Caucasian men, for whom the average age of diagnosis was 63.0 ± 8.0 and the average age of controls was 60.0 ± 8.3 . Among Hispanic Caucasians, the deletion had a minor allele frequency of 0.023, and was not associated with prostate cancer risk (OR, 0.91; 95% CI, 0.36 2.31; P = 0.84). Out of 37 Hispanic Caucasians with high grade prostate cancer, we observed zero cases carrying the deletion. This would be expected based on the low minor allele frequency of this deletion in the Hispanic Caucasian population.

Discussion

In this study we have independently tested reported associations of germ line copy number variants with risk for prostate cancer. Of 6 variants reported by Demichelis *et al.* (7), 4 were not polymorphic in the non Hispanic Caucasian population from the SABOR cohort. This observation may be explained by regional population differences in low minor allele frequency variants. Alternatively, it is possible that although the primers used in this study fall within the reported boundaries of the variants, that they do not fall within the variant's true boundaries. To avoid this scenario, primers and probes were designed to fall within variants identified by the 1000 genomes project if applicable (11). Additionally, 3 separate primer probe sets, including those reported by Demichelis *et al.*, were used to genotype the deletion at 15q21.3. Taken together our data provides strong evidence that the non Hispanic Caucasian population from the SABOR cohort do not harbor these deletion variants.

For the 2p24.3 variant our observations are highly consistent with those reported by Liu *et al.*(6). Although we obtained a p value that is not strictly significant, the probability that these results would be observed by chance (P=0.06) is suggestive of an association with overall prostate cancer risk. It is worth noting that the estimated power to detect this association at $\alpha = 0.05$ was not strong (≈ 0.37). A comparison of results under dominant (as they reported it) and recessive models is illustrated in Figure 2. Interestingly, in Caucasians

we observed statistically significant enrichment of homozygous deletions in high grade prostate cancer cases. Although all three cohorts display overlapping confidence intervals, this observation appears to be somewhat discordant with the observations reported by Liu *et al*(6). However, the criteria that defined aggressiveness varied among the studies, which may explain different associations with aggressive disease. The definition of aggressiveness in this study did not include staging or PSA measurements because data was not available for all prevalent cases. However, the measurement used in this study, Gleason sum, is a robust predictor of prostate cancer prognosis (12). Follow up of this variant in larger cohorts with various definitions of aggressiveness is merited.

Conclusions

The initial discovery, and now independent supporting evidence, of the association between the 2p24.3 deletion reported by Liu *et al.* and prostate cancer support the involvement of copy number variation in the etiology of heritable disease. Further studies are merited to confirm the relationship between this deletion and risk for aggressive prostate cancer, and to investigate the relationship between this deletion, other linked variants, and nearby genes. The evidence from this study provides additional support that this copy number variant is a true risk allele for prostate cancer in non Hispanic Caucasians.

Acknowledgments

The participation and cooperation of all study subjects is gratefully acknowledged. The study could not have been accomplished without the skilled assistance of the SABOR/PREF clinical staff. We thank the expert technical assistance of Yasmin Ench.

Funding Support

This work was funded in part by DOD grant PC081025, by grant U01CA086402 from the Early Detection Research Network of the National Cancer Institute, and by the CTRC at UTHSCSA grant P30CA054174. Data was generated in the Genomics Core Shared Resource, which is supported by UTHSCSA, and NIH NCI P30CA054174 (CTRC of UTHSCSA).

References

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343:78–85. [PubMed: 10891514]
2. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet.* 2011; 12:363–76. [PubMed: 21358748]
3. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature.* 2010; 464:704–12. [PubMed: 19812545]
4. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010; 467:52–8. [PubMed: 20811451]
5. Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature.* 2010; 464:713–20. [PubMed: 20360734]
6. Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, et al. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res.* 2009; 69:2176–9. [PubMed: 19258504]

7. Demichelis F, Setlur SR, Banerjee S, Chakravarty D, Chen JY, Chen CX, et al. Identification of functionally active, low frequency copy number variants at 15q21.3 and 12q21.31 associated with prostate cancer risk. *Proc Natl Acad Sci U S A*. 2012; 109:6686–91. [PubMed: 22496589]
8. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*. 2008; 9:356–69. [PubMed: 18398418]
9. Barnes C, Plagnol V, Fitzgerald T, Redon R, Marchini J, Clayton D, et al. A robust statistical method for case-control association testing with copy number variation. *Nat Genet*. 2008; 40:1245–52. [PubMed: 18776912]
10. Menashe I, Rosenberg PS, Chen BE. PGA: power calculator for case-control genetic association analyses. *BMC Genet*. 2008; 9:36. [PubMed: 18477402]
11. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–73. [PubMed: 20981092]
12. Egevad L. Recent trends in gleason grading of prostate cancer. II. Prognosis, reproducibility and reporting. *Anal Quant Cytol Histol*. 2008; 30:254–60. [PubMed: 18980156]

FIGURE 1. Genotyping of deletion in San Antonio Cohorts

A) Histograms of calculated copy number values in Caucasian and Hispanic individuals for a deletion on 2p24.3. Gaussian distributions are present representing measurement variation around integer copy number values of 0, 1, and 2. Discrete values were inferred based on these distributions. B) Gel electrophoresis of PCR products, as reported by Liu *et al.*, was used to confirm the inferred genotypes in a subset of individuals, including the individuals with calculated copy number values at the edges of these distributions. This gel displays the

genotyping results for a subset of the samples and displays all three genotypic states of this deletion. Negative control indicates no DNA.

FIGURE 2. Comparison between studies of estimated effects for a deletion on 2p24.3

A comparison of odds ratios and 95% confidence intervals for Caucasians in this study (SABOR), and for cohorts previously reported on by Liu *et al.* Total indicates the results when counts are combined between studies. **A)** The results under a dominant model (DX vs WW) and **B)** a recessive model (DD vs WX).

Table 1

Summary of quantitative PCR primer/probe pairs

Deletion region	Primer design	Pre-designed Assay ID or Pubmed ID	Probe location (GRCh37/hg19)	Forward primer	Reverse primer	Probe
2p24.3	Custom		Chr2:14707671	TTGGGTCAGGCTGTCTTCT	TTTCAGCAGAGTGGTGAAGGAA	TCTTTCCCAACTCATTACC
15q21.3	Pre-designed	Hs05369169_cn [/]	chr15:54199360			
15q21.3	Pre-designed	Hs05345872_cn [/]	chr15:54198190			
19q13.12	Pre-designed	Hs01617872_cn [/]	chr19:36829874			
8p23.2	Custom [/]		Chr8:3719510			
8q24.3	Pre-designed	Hs03668742_cn [/]	chr8:145686582			
7p22.2	Pre-designed	Hs01408227_cn [/]	chr7:4091360			
15q21.3	Previously reported	PMID:22496589				
12q21.31	Previously reported	PMID:22496589				

[/] Assays designed and ordered from Applied Biosystems and reported according to MIQE Guidelines.

Table 2

Genotyping counts in Caucasians

Deletion region	Pubmed ID	Cases				Controls			
		DD	DW	WW	MAF	DD	DW	WW	MAF
2p24.3	19258504	2	43	272	0.074	2	129	1088	0.055
15q21.3	22496589	0	0	310	0	0	0	1184	0
19q13.12	22496589	14	106	184	0.220	48	397	741	0.208
12q21.31	22496589	4	52	243	0.100	16	185	902	0.098
8p23.2	22496589	0	0	305	0	0	0	1173	0
8q24.3	22496589	0	0	297	0	0	0	1120	0
7p22.2	22496589	0	0	305	0	0	0	1154	0

Abbreviations: DD: homozygous deletion; DW: heterozygous deletion; WW: normal copy number; MAF: minor allele frequency