

Germline Missense Variants in the *BTNL2* Gene Are Associated with Prostate Cancer Susceptibility

Liesel M. FitzGerald¹, Akash Kumar², Evan A. Boyle², Yuzheng Zhang¹, Laura M. McIntosh¹, Suzanne Kolb¹, Marni Stott-Miller¹, Tiffany Smith⁴, Danielle M. Karyadi⁴, Elaine A. Ostrander⁴, Li Hsu¹, Jay Shendure², and Janet L. Stanford^{1,3}

Abstract

Background: Rare, inherited mutations account for 5% to 10% of all prostate cancer cases. However, to date, few causative mutations have been identified.

Methods: To identify rare mutations for prostate cancer, we conducted whole-exome sequencing (WES) in multiple kindreds ($n = 91$) from 19 hereditary prostate cancer (HPC) families characterized by aggressive or early-onset phenotypes. Candidate variants ($n = 130$) identified through family- and bioinformatics-based filtering of WES data were then genotyped in an independent set of 270 HPC families ($n = 819$ prostate cancer cases; $n = 496$ unaffected relatives) for replication. Two variants with supportive evidence were subsequently genotyped in a population-based case-control study ($n = 1,155$ incident prostate cancer cases; $n = 1,060$ age-matched controls) for further confirmation. All participants were men of European ancestry.

Results: The strongest evidence was for two germline missense variants in the *butyrophilin-like 2* (*BTNL2*) gene (rs41441651, p.Asp336Asn and rs28362675, p.Gly454Cys) that segregated with affection status in two of the WES families. In the independent set of 270 HPC families, 1.5% (rs41441651; $P = 0.0032$) and 1.2% (rs28362675; $P = 0.0070$) of affected men, but no unaffected men, carried a variant. Both variants were associated with elevated prostate cancer risk in the population-based study (rs41441651: OR, 2.7; 95% CI, 1.27–5.87; $P = 0.010$; rs28362675: OR, 2.5; 95% CI, 1.16–5.46; $P = 0.019$).

Conclusions: Results indicate that rare *BTNL2* variants play a role in susceptibility to both familial and sporadic prostate cancer.

Impact: Results implicate *BTNL2* as a novel prostate cancer susceptibility gene. *Cancer Epidemiol Biomarkers Prev*; 22(9); 1520–8. ©2013 AACR.

Introduction

Prostate cancer is a complex and heterogeneous disease that has a strong genetic component to its etiology, with an estimated 42% of disease incidence attributed to heritable factors (1). Genome-wide association studies of prostate cancer have identified more than 70 common low-penetrance single-nucleotide polymorphisms (SNP) that are confirmed to be associated with weak to modest alterations (average per allele ORs, 1.1–1.3) in disease risk (2, 3)

Authors' Affiliations: ¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center; ²Department of Genome Sciences, School of Medicine; ³Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington; and ⁴Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland

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

L.M. FitzGerald and A. Kumar contributed equally to this work.

Corresponding Author: Janet L. Stanford, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., PO Box 19024, M4-B874, Seattle, WA 98109-1024. Phone: 206-667-2715; Fax: 206-667-4787; E-mail: jstanfor@fhcrc.org

doi: 10.1158/1055-9965.EPI-13-0345

©2013 American Association for Cancer Research.

and which taken together may explain up to 30% of the genetic risk for prostate cancer. In addition, genome-wide linkage studies of hereditary prostate cancer (HPC) families have searched for genomic regions that harbor rare, moderate- to high-penetrance mutations. These linkage studies have discovered more than 2 dozen putative susceptibility loci (4–6); but only a few candidate genes underlying these loci have been proposed and to date, even fewer rare, genetic mutations for prostate cancer have been confirmed (7–11). Recently, a targeted next-generation sequencing study of candidate genes across a linkage region on 17q21-22 identified a rare germline *HOXB13* mutation (G84E) in 4 HPC families of European descent (10). Subsequent studies confirmed that the mutation (rs138213197) was carried by 2.4% of affected members from 1,892-independent HPC families tested (11) and was present in about 1% of prostate cancer cases ascertained from the general population (12–14).

To find novel germline mutations for prostate cancer, we completed one of the first whole-exome sequencing (WES) studies of 19 HPC families in which multiple affected men per family with an aggressive or early-onset phenotype were selected for sequencing. Candidate variants were then genotyped in an independent set of 270

HPC families and a population-based, case-control study for further confirmation.

Materials and Methods

Study populations

Participants selected for WES are members of 19 selected families chosen from a larger dataset of 289 HPC families of European ancestry (15). Each of the 19 families has 5 or more affected men with at least 3 diagnosed with a more aggressive phenotype and/or early-onset prostate cancer based on the median age at diagnosis (i.e., 65 years) of cases from the 289 HPC families. From the 19 families, 2 to 6 affected men ($n = 80$) and, where possible, one older, unaffected, prostate-specific antigen screened negative male relative ($n = 11$) were sequenced (Table 1). The majority of affected men sequenced were diagnosed with more aggressive disease features (i.e., Gleason score 8–10 or regional/distant stage: $n = 43$ men) or at earlier ages (≤ 65 years: $n = 55$; mean age = 62 years), or both ($n = 23$). To decrease the likelihood of identifying false-positives due to inheritance identical-by-descent, the majority of affected men selected are second- or third-degree relatives. The 11 unaffected male relatives are older (mean age = 82 years) and thus are presumed less likely to develop HPC due to even moderately penetrant mutations. All 91 men sequenced were previously genotyped with the Illumina Linkage IVb panel (15).

The remaining independent set of 270 HPC families (described in ref. 15) was used to determine the frequency and distribution of candidate variants ($n = 130$) discovered in the 19 families in a larger representative group of HPC families. A total of 869 affected men and 519 unaffected male relatives with DNA samples are included in the confirmation genotyping effort.

The population-based, case-control study was used to estimate risk of prostate cancer associated with genetic variants ($n = 2$) with supportive evidence from analyses of HPC families. Participants are from 2 population-based studies of prostate cancer carried out in residents of King County, Washington (16, 17). For this genotyping effort, only men of European ancestry with DNA available are included ($n = 1,155$ incident cases; $n = 1,060$ age-matched controls).

This study was approved by the Fred Hutchinson Cancer Research Center's Institutional Review Board and informed consent was obtained from all study participants. Genotyping of the case-control study samples was also approved by the Institutional Review Board of the National Human Genome Research Institute.

WES in 19 HPC families

A total of 10 μ g of genomic DNA per subject was sent to the Center for Inherited Disease Research (CIDR) for sequencing. For quality control and inheritance checks,

Table 1. Characteristics of 19 HPC families with WES data

Family ID	No. of HPC cases	Mean age at HPC diagnosis	No. of WES cases with aggressive HPC ^a	No. of WES cases with early-onset HPC ^a	No. of WES cases per family	No. of WES unaffected men per family
1	9	64.4	2	3	4	1
2	7	62.2	2	3	4	
3	7	69.9	2	2	5	1
4	9	67.3	4	2	5	1
5	8	68.0	3	1	4	1
6	6	60.6	1	5	5	
7	11	64.6	2	2	3	1
8	5	54.0	2	3	3 ^b	1
9	7	57.2	3	2	3	
10	6	66.0	3	2	4	
11	7	59.0	5	3	5 ^b	
12	9	68.4	2	1	4	1
13	9	60.0	1	5	5	
14	10	65.1	2	6	6	1
15	7	61.9	3	3	4	
16	8	63.4	0	4	4	
17	9	66.2	1	2	2	1
18	10	65.6	2	3	5	1
19	7	67.8	3	3	5	1
Total	151	63.8	43	55	80	11

^aA total of 23 cases had both aggressive and early-onset prostate cancer.

^bWES failed or was of low quality for one of the affected men in these families.

all samples were first run on the OmniExpress Array (Illumina, Inc.). Once initial quality control checks were completed, 3 μ g of DNA per subject was sheared, underwent library construction, and was hybridized to the SureSelect Human All Exon 50Mb Array (Agilent). The captured library was PCR amplified, indexed, and loaded on the HiSeq 2000 (Illumina, Inc.) for 75 bp paired-end sequencing.

WES data quality control and analyses

Sequencing reads were demultiplexed at CIDR and fastq files were created for each sample. The Burrows–Wheeler Aligner (BWA; ref. 18) was used to align reads to the hg19 reference genome and GATK (19) was used for local realignment. Molecular duplicates were marked using Picard and SAMtools (20) was used to sort, index, and generate pileup files for variant calling. Sequencing coverage statistics, bases on target, transition/transversion ratios (Ti/Tv), variant/reference base ratios for heterozygous single nucleotide variants (SNV), and concordance between the OmniExpress and sequencing data were calculated. Variant files containing SNVs and insertions or deletions (indels) were annotated using SeattleSeq and ANNOVAR, respectively, after filtering using the SAMtools.pl varFilter (all defaults except for minimum coverage of 8-fold and $D = 20,000$; refs. 21, 22).

SNVs were filtered on a family-level basis using 4 different methods, some allowing for incomplete penetrance and phenocopies. In all instances, the minor allele frequency (MAF) of SNVs was determined using a subset of exomes sequenced as part of the National Heart, Lung, and Blood Institute (NHLBI)/NIH exome-sequencing project (ESP). The 4 filtering approaches were as follows: (i) MAF < 0.02 , present in all affecteds, not present in the unaffected male relative if available, and not present in any other unaffected males; (ii) MAF < 0.02 , present in all but one of the affecteds, not present in the unaffected male relative if available, and not present in any other unaffected males; (iii) MAF < 0.01 , present in all affecteds, present in the unaffected male relative if available, and not present in any other unaffected males; and (iv) MAF < 0.01 , present in all but one of the affecteds, present in the unaffected male relative if available, and not present in any other unaffected males. Filtering methods 2 and 4 allowed for phenocopies and methods 3 and 4 allowed for incomplete penetrance. These filters highlighted 1,459 SNVs (Supplementary Fig. S1) that were further prioritized according to the following information: type (nonsense $<$ splice site $<$ missense); prediction scores based on the evolutionary conservation of the reference base and the impact the variant would have on the resulting amino acid change using Genomic Evolutionary Rate Profiling (GERP) (≥ 5 ; ref. 23), PolyPhen (probably damaging; ref. 24), SIFT (0.00–0.10; ref. 25), Grantham (≥ 151 ; ref. 26), PhyloP (≥ 3 ; ref. 23), likelihood ratio test (damaging; ref. 27), and BLOSUM62 (-2 to -4 ; ref. 28); gene information contained in the UCSC Genome Browser (29) and PubMed; and presence within a previously identified

linkage region [i.e., a dominant or recessive logarithm of the odds (LOD) ≥ 1.86] from an earlier genome-wide linkage scan (15).

Indels were also filtered on a family-level basis, but different methods were used to prioritize candidates. The 4 filters applied to the data allowed for none, one, two, or three phenocopies, respectively. Indels were removed if they were observed in any of the unaffected men. The 2,510 filtered indels (Supplementary Fig. S1) were then prioritized according to the following information: type (frameshift $<$ untranslated region $<$ non-frameshift); location (exonic $<$ intronic); number of families in which the indel was observed, with greater weight placed on those that were seen in fewer families; gene information contained in the UCSC Genome Browser and PubMed; and presence within a previously identified linkage region (i.e., a dominant or recessive LOD ≥ 1.86) from an earlier genome-wide linkage scan (15).

Genotyping of candidate variants in HPC families

The molecular inversion probe (MIP) assay (30) was used to genotype candidate variants identified from the WES analyses. The protocol used was similar to that of O’Roak and colleagues (31). Briefly, 70 bp oligonucleotide inversion probes (Integrated DNA Technologies, Inc.) were designed against 196 candidate SNVs and indels (Supplementary Tables S1 and S2). These oligonucleotides were 5’ phosphorylated and added to approximately 200 ng of germline DNA at a ratio of 200:1 MIPs to template. The probe/DNA mixture was incubated with ligase, polymerase, and nucleotides for 48 hours, resulting in targeted regions being “captured” within single-stranded circular DNA. After exonuclease removal of noncircularized DNA, captured products were amplified using PCR with barcoded primers containing adaptor sequences. The amplified products were pooled and sequenced on the HiSeq Illumina platform. Sequencing data were aligned to the human hg19 reference genome using BWA (18) and variant calls were made using SAMtools (20). A position was considered to possess a variant if it was covered to a minimum $8\times$ depth and had at least 20% of reads supporting the variant allele.

Genotyping data quality control

To identify variants with a high probability of being artifacts, a comparison of the MIP and WES data was undertaken in members of the 19 sequencing families with both types of data. A second method of identifying probable artifacts looked for significant differences in variant allele frequencies in the 5,379 ESP exomes from individuals of European ancestry ($P < 0.05$ based on a binomial distribution). From the total of 196 candidate SNVs ($n = 174$) and indels ($n = 22$) selected for follow-up genotyping, 66 were excluded for the following reasons: MIP design failure ($n = 6$); low-call rate within genotyped subjects ($> 15\%$ missing, $n = 55$); and probable artifacts discovered through a comparison of MIP and WES data ($< 95\%$ concordance, $n = 5$).

Genotyping data analyses in HPC families

The PedGenie program (32) was used to assess the association of 130 candidate variants with affection status in the 270 independent HPC families. This program can handle pedigrees of arbitrary size and structure and provides valid statistical inference by gene dropping to generate the null distribution. Each SNP or indel was coded as a binary variable with 0 and 1 indicating the absence/presence of the candidate variant, respectively. Statistical significance was determined by the Monte Carlo approach to account for potential correlation of genotypes within a family and rarity of the SNVs and indels. A total of 100,000 simulated datasets were generated to form the null distribution. A one-sided P value was calculated by dividing the χ^2 P value by 2 for candidate variants that were observed more frequently in men with, compared to men without prostate cancer and one minus the χ^2 P value divided by 2 for candidate variants that were observed less frequently in affected men than in unaffected men. A P value less than 0.05, was considered statistically significant in testing for confirmation of candidate variants.

Genotyping of *BTNL2* candidate variants in case-control samples

A custom designed TaqMan SNP genotyping assay (Applied Biosystems) was used to genotype the 2 *BTNL2* candidate variants (rs41441651 and rs28362675) on the ABI Prism 7900HT sequence detection system according to the manufacturer's instructions.

Genotyping data analysis in the case-control study

Unconditional logistic regression was used to estimate the OR and 95% confidence interval (CI) as a measure of association between the 2 *BTNL2* candidate variants and prostate cancer (33), as implemented in STATA version 11.0 (Stata Corp.). Potential confounding factors—including age at reference date, prostate cancer screening history, and first-degree family history of prostate cancer, were examined to see whether such factors changed the risk estimates by 10% or more. After these analyses, only age at reference date was included in the final models. Regression models were also used to generate ORs and 95% CIs for the association between SNP genotypes in men stratified by family history (yes vs. no). A product term between SNP genotypes and family history was included in logistic regression models, and a log-likelihood ratio test was used to compare logistic models with and without the product term to test whether the effects of SNP genotypes differed by family history.

Results

WES data were available for 91 men from 19 HPC families (Table 1). In the 89 individuals for whom WES data passed quality control, an average of 70 \times (range: 20 \times to 132 \times) coverage of the target was achieved (Agilent

SureSelect Human All Exon 50 Mb, Illumina HiSeq 2 \times 75 bp) with approximately 88% of target bases having at least 8 \times coverage (Supplementary Table S3). Concordance between genotyping data from the OmniExpress array and WES data was 99.9%. Family- and bioinformatics-based filtering of the WES data prioritized 174 SNVs and 22 indels (Supplementary Tables S1 and S2) as candidate variants for follow-up genotyping in 270 independent HPC families. After quality control (see Materials and Methods), data for 130 of these candidates remained for analysis.

For the 130 candidate variants, the average concordance between 15 blind duplicates was 99.5% (nonreference concordance was 82.5%) and 99.9% (nonreference concordance was 99.3%) for individuals who had both WES and MIP data available. A total of 1,388 men from the 270 HPC confirmation families were genotyped; 73 individuals who were missing more than 15% of the 130 MIP genotypes were excluded, leaving 819 affected and 496 unaffected men in the analysis.

Family-based association analysis of the 270 independent HPC families provided evidence (i.e., higher MAF in affected vs. unaffected men, Monte-Carlo-based one-tailed $P < 0.05$) for 2 rare variants in *BTNL2* (Table 2). These missense variants, rs41441651 (D336N) and rs28362675 (G454C), are not present in HapMap and therefore a formal test for linkage disequilibrium was not possible, however, they are located within a single haplotype block of eight Kb. This and the fact that all but 2 of the 22 affected carriers with data were concordant for both variants (2 men had poor coverage for rs28362675 and thus have unknown carrier status) suggest that these variants are in strong linkage disequilibrium. The smallest P value observed was for rs41441651 ($P = 0.0032$).

The rs41441651 variant segregated with affection status in 2 of the 19 WES families (Fig. 1) and was present in 10 of 12 genotyped affected men. (Sanger sequencing was used to confirm the carrier status of the female in family 11). The affected carriers had an average age at diagnosis of 62.6 years and 60% had regional stage prostate cancer at diagnosis. Gleason scores ranged from 5 to 9 (average 6.5). Of the unaffected genotyped men in these 2 families, none carried either candidate variant.

In the 270 independent HPC families evaluated, 3.3% and 2.9% had one or more affected members who carried the rs41441651 or rs28362675 variant, respectively. In total 12 (1.47%) of 819 genotyped affected men, but none of the 496 genotyped unaffected men carried the rs41441651 candidate variant. However, in these families, which contained fewer affected men and were of smaller size than the 19 families selected for WES, there was not clear evidence of cosegregation with disease state.

The 2 *BTNL2* candidate variants genotyped in the case-control dataset had distributions among controls that were consistent with Hardy-Weinberg equilibrium ($P > 0.05$). For rs41441651, 26 (2.3%) cases and 9 (0.9%) controls carried the missense variant; for rs28362675, 24 (2.1%)

Table 2. Results for SNVs^a identified by WES and genotyped in 270 independent HPC families of European ancestry

Gene	Genomic position (hg19)	Variant and rs ID	Protein	Discovery			Validation			
				MAF in ESP	MAF in ClinSeq	No. of WES families with carriers (Aff/Unaff) ^b	No. (%) of 270 families with affected carriers	MAF in 819 genotyped affected men ^c	MAF in 496 genotyped unaffected men ^c	P ^d
<i>BTNL2</i>	Chr6: 32,363,888	C > T rs41441651	Missense: p.Asp336Asn	0.009	0.005	2 (10/0)	9 (3.33)	0.0073	0	0.0032
<i>BTNL2</i>	Chr6: 32,362,521	C > A rs28362675	Missense: p.Gly454Cys	0.008	0.005	2 (10/0)	8 (2.96)	0.0061	0	0.0070

^aTop ranked ($P < 0.05$) SNVs identified by WES of 19 HPC families.

^bNumber of affected carriers/number of unaffected carriers.

^cThe number of affected carriers for rs41441651 and rs28362675 is 12 and 10, respectively, in the 270 independent HPC families.

^dMonte Carlo-based one-sided P value from the PedGenie χ^2 test for association based on the 270 independent HPC families.

cases and 9 (0.9%) controls were carriers (Table 3). Both candidate variants were associated with statistically significant increases in the risk of prostate cancer (rs41441651: OR, 2.7; 95% CI, 1.27–5.87; $P = 0.010$; rs28362675: OR, 2.5; 95% CI, 1.16–5.46; $P = 0.019$). These risk estimates did not differ by family history of prostate cancer, but this subgroup analysis had limited power. The mean age at prostate cancer diagnosis was 59.0 years for carriers of one or both variants and was 59.8 years for noncarriers ($P = 0.6$).

Discussion

WES of 91 men in 19 HPC families, followed by replication ($n = 130$ candidate variants) in an independent set of 270 HPC families and further testing of candidate variants with replication support ($n = 2$) in a population-based case-control study, provides compelling evidence that rare germline variants in *butyrophilin-like 2* (*BTNL2*) are associated with genetic susceptibility to prostate cancer. These rare missense variants, rs41441651 (exon 5; D336N) and rs28362675 (exon 6; G454C), occur in the same haplotype block on chromosome 6p21.32 and were observed to be in strong linkage disequilibrium among controls in the case-control dataset ($r^2 = 0.99$). This is the first WES study focused on aggressive or early-onset prostate cancer phenotypes and the first to implicate rare germline *BTNL2* variants as predisposing to familial and sporadic prostate cancer.

There is some prior suggestive evidence for a role of *BTNL2* in prostate cancer. A recent exome sequencing study of prostate tumor tissue from 50 patients with lethal prostate cancer identified one patient who had a somatic *BTNL2* mutation (c.709+9T>G; ref. 34). Also, Acevedo and colleagues (35) found that *BTNL2* protein was significantly overexpressed in advanced prostate cancer tumor tissue relative to normal prostate tissue in a mouse model of the disease (34). In comparing the protein-encoding transcriptomes of 79 different normal human tissues, Su and colleagues (36) found that *BTNL2* mRNA expression in the prostate was above the median expression level observed in other tissues.

Butyrophilin-like (BTNL) molecules are thought to play a role in immune regulation and have been functionally implicated in T-cell inhibition and modulation of epithelial cell-T-cell interactions (37). *In vitro* mouse studies indicate that the *BTNL2* protein is a negative regulator of T-cell proliferation and cytokine production (38, 39).

Genetic polymorphisms in *BTNL2* have been associated with several immunologic diseases. Studies have reported significant associations between *BTNL2* SNPs and the inflammatory autoimmune diseases sarcoidosis (40, 41) and rheumatoid arthritis (42), as well as inflammatory bowel disease and ulcerative colitis (43, 44). None of the affected carriers of *BTNL2* variants in our study population reported a history of these inflammatory conditions. In addition, no participants reported a family history of sarcoidosis or any other autoimmune diseases among close family members. Interestingly, one of the SNPs

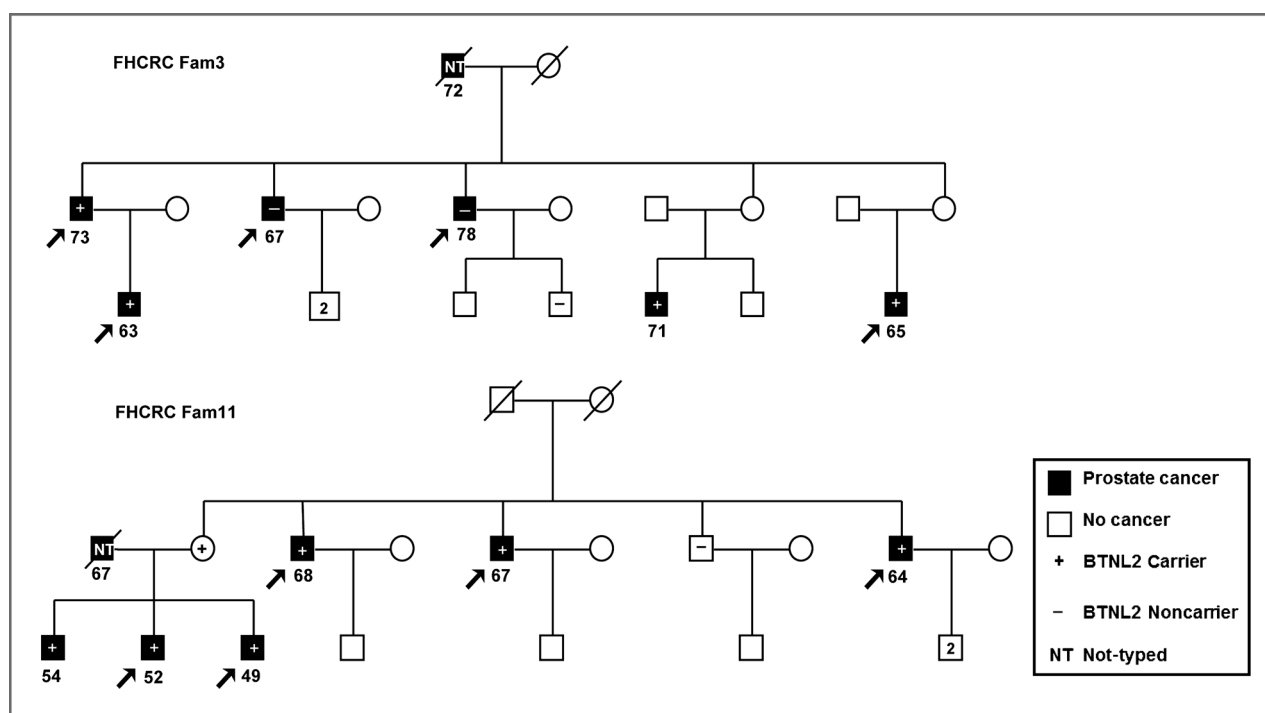


Figure 1. HPC pedigrees of 2 families with segregating *BTNL2* variants. Participants selected for WES in 19 HPC families are indicated by arrows. Affected men indicated by black shading. The remaining symbols are described in the key. Squares indicate males and circles indicate females. The age at diagnosis of men with prostate cancer is shown under the squares. A slash through the symbol indicates that the individual is deceased. The carrier status of the female in family 11 was confirmed by Sanger sequencing.

previously associated with ulcerative colitis, rs9268480 (43), is located only 44 bp from the *BTNL2* rs41441651 and is within the same haplotype block. Given the biologic activity of *BTNL2*, our results provide further support for the role of the inflammation pathway in the development of prostate cancer (45, 46).

Four of 7 functional and conservation prediction scores suggest that rs28362675 is damaging, although current evidence is equivocal for rs41441651; both variants change the encoded amino acid. The rs41441651 SNV seems to be present within a cluster of CpG dinucleotides that are

either heavily methylated or unmethylated according to the cell line assayed (47). This SNV may therefore disrupt methylation at this site. It is possible, however, that the *BTNL2* variants we describe here are not causative, rather they are in linkage disequilibrium with a yet undiscovered functional variant. This is a formal consideration as the haplotype block in which they are located extends into the 3' regulatory region of the gene, which had limited sequencing coverage.

Among the potential HPC variants highlighted by this study, the *BTNL2* variants were notable in that they were

Table 3. OR and 95% CI for prostate cancer associated with SNVs in *BTNL2* in European Americans

Genotype	Cases (n = 1,155) n (%)	Controls (n = 1,060) n (%)	OR ^a (95% CI)	P
rs41441651				
CC	1,129 (97.8)	1,051 (99.2)	1.00	
CT or TT ^b	26 (2.3)	9 (0.9)	2.73 (1.27–5.87)	0.010
rs28362675				
CC	1,131 (97.9)	1,051 (99.2)	1.00	
CA or AA ^b	24 (2.1)	9 (0.9)	2.52 (1.16–5.46)	0.019

^aAdjusted for age.

^bOne case is homozygous variant for both SNVs; 22 cases and 9 controls are heterozygous for both SNVs.

observed only in affected men in both the WES families and the 270 HPC family replication dataset. The variants segregated with disease in 2 of the WES families and although this was not the case in the 270 HPC family replication dataset, about 3% of the latter families had affected carriers of one or both variants. These variants were also observed in 2.3% (rs41441651) and 2.1% (rs28362675) of sporadic prostate cancer cases and 0.9% of the population-based controls. These observations are similar to those seen for the *HOXB13* mutation, rs138213197 (10); 2 unaffected males were observed to carry the *HOXB13* mutation in the 4 HPC-discovery families and a carrier frequency of 0.1% was observed in 1,401 controls. Furthermore, studies of both familial and case-control datasets have indicated that while rs138213197 is significantly associated with prostate cancer risk, it rarely segregates perfectly with disease in HPC families and it is seen at a low frequency in controls (11–14, 48).

Allele frequency data for the 2 *BTNL2* variants are available from several recent sequencing efforts. In the large NHLBI GO ESP (49) consisting of mixed race U.S.-based studies, the MAF for both variants is reported as 0.5% (chromosomes = 4,542–4,550). In the ClinSeq Project (50) consisting of individuals of European ancestry, the MAF for both variants is 1.5% (chromosomes = 1,310–1,323). Finally, in a pilot 1,000 Genomes Project (51) population of Chinese and Japanese, the MAF for both variants is 15% (chromosomes = 120). From previous analyses of linkage data, we confirmed that the 289 HPC families in this study are of European and not Asian or African descent. Therefore, the ClinSeq population is most representative of our HPC and case-control study populations and the average MAF of the case (2.2%) and control (0.9%) samples from our population-based dataset is similar to that of the ClinSeq study. Prostate cancer is a prevalent disease and the prostate cancer status of the ClinSeq male population is not publicly available, so it is possible that the MAF in the ClinSeq data is inflated due to the inclusion of affected men. Regardless, the MAF in our cases is higher than that in the population controls and the ClinSeq population.

There were a number of SNVs and indels highlighted in the 19 WES HPC families that were only observed once or not at all in the 270 replication HPC families. Because of the rarity of these potential variants, additional follow-up in a larger set of HPC families will be needed to confirm these associations and determine the proportion of HPC that may be attributable to these other rare variants. In addition, the 66 candidate SNVs and indels that were unable to be evaluated in the 270 independent HPC families require further study.

Identifying HPC mutations has been challenging due to the genetic heterogeneity of the disease and the phenotypic complexity of prostate cancer. This study is the first to show the value of WES in large multiplex HPC families characterized by aggressive- or early-onset prostate cancer, with replication in an independent HPC family dataset and a population-based case-

control dataset. We identified 2 rare *BTNL2* variants that segregate with disease in 2 HPC families with sequencing data and that are carried only by affected men, but no unaffected men, in 8 (rs28362675) and 9 (rs41441651) of the additional 270 HPC families tested. We also found that these 2 variants are associated with statistically significant 2.5- to 2.7-fold elevations in the relative risk of prostate cancer in the general population, with slightly more than 2% of incident sporadic prostate cancer cases carrying at least one of these variants. Larger studies of densely affected HPC families (≥ 5 affected men) and case-control datasets are now needed to establish the significance of these novel *BTNL2* missense variants in further defining prostate cancer genetic susceptibility.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: L.M. FitzGerald, D.M. Karyadi, E.A. Ostrander, J. Shendure, J.L. Stanford

Development of methodology: L.M. FitzGerald, A. Kumar, J. Shendure, J.L. Stanford

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Kumar, E.A. Boyle, L.M. McIntosh, S. Kolb, T. Smith, D.M. Karyadi, J. Shendure, J.L. Stanford

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L.M. FitzGerald, A. Kumar, E.A. Boyle, Y. Zhang, M.S.-Miller, T. Smith, L. Hsu, J.L. Stanford

Writing, review, and/or revision of the manuscript: L.M. FitzGerald, A. Kumar, M.S.-Miller, E.A. Ostrander, L. Hsu, J. Shendure, J.L. Stanford

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Kumar, L.M. McIntosh, J.L. Stanford

Study supervision: L.M. FitzGerald, L.M. McIntosh, S. Kolb, J. Shendure, J.L. Stanford

Acknowledgments

The authors thank the members of prostate cancer families in the Prostate Cancer Genetic Research Study (PROGRESS) for their participation. The authors wish to acknowledge the support of the National Heart, Lung, and Blood Institute (NHLBI) and the contributions of the research institutions, study investigators, field staff, and study participants in creating the ESP resource for biomedical research.

Grant Support

This work was supported by grants from the U.S. National Cancer Institute, NIH (grant numbers RO1 CA080122 and P50 CA097186 to J.L. Stanford); with additional support from the Fred Hutchinson Cancer Research Center; and the Prostate Cancer Foundation. Sequencing services were provided by the Center for Inherited Disease Research, which is funded through a contract from NIH to The Johns Hopkins University (contract no. HHSN268200782096C). Funding for GO ESP was provided by NHLBI grants RC2 HL-103010 (HeartGO), RC2 HL-102923 (LungGO), and RC2 HL-102924 (WHISP). The exome sequencing was conducted through NHLBI grants RC2 HL-102925 (BroadGO) and RC2 HL-102926 (SeattleGO). A. Kumar is supported by an Achievement Award for College Scientists Fellowship; J. Shendure is supported by The Lowell Milken Prostate Cancer Foundation Young Investigator Award; E.A. Ostrander is supported by the Intramural Program of the National Human Genome Research Institute; J.L. Stanford is supported by a Prostate Cancer Foundation Creativity Award.

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.

Received April 5, 2013; revised June 12, 2013; accepted June 23, 2013; published OnlineFirst July 5, 2013.

References

- 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.
- Goh CL, Schumacher FR, Easton D, Muir K, Henderson B, Kote-Jarai Z, et al. Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *J Intern Med* 2012;271:353–65.
- Eeles R, Al Olama AA, Benlloch S, Saunders E, Leongamornlert D, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 2013;45:385–91.
- Ostrander EA, Stanford JL. Genetics of prostate cancer: too many loci, too few genes. *Am J Hum Genet* 2000;67:1367–75.
- Easton DF, Schaid DJ, Whittemore AS, Isaacs WJ, ICGC. Where are the prostate cancer genes?—A summary of eight genome wide searches. *Prostate* 2003;57:261–9.
- Schaid DJ. The complex genetic epidemiology of prostate cancer. *Hum Mol Genet* 2004;13:R103–R21.
- Edwards SM, Kote-Jarai Z, Meitz J, Hamoudi R, Hope Q, Osin P, et al. Two percent of men with early-onset prostate cancer harbor germline mutations in the *BRCA2* gene. *Am J Hum Genet* 2003;72:1–12.
- Agalliu I, Karlins E, Kwon EM, Iwasaki LM, Diamond A, Ostrander EA, et al. Rare germline mutations in the *BRCA2* gene are associated with early-onset prostate cancer. *Br J Cancer* 2007;97:826–31.
- Kote-Jarai Z, Leongamornlert D, Saunders E, Tymrakiewicz M, Castro E, Mahmud N, et al. *BRCA2* is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. *Br J Cancer* 2011;105:1230–4.
- Ewing CM, Ray AM, Lange EM, Zuhlke KA, Robbins CM, Tembe WD, et al. Germline mutations in *HOXB13* and prostate-cancer risk. *N Engl J Med* 2012;366:141–9.
- Xu J, Lange EM, Lu L, Zheng SL, Wang Z, Thibodeau SN, et al. *HOXB13* is a susceptibility gene for prostate cancer: results from the International Consortium for Prostate Cancer Genetics (ICPCG). *Hum Genet* 2013;132:5–14.
- Akbari MR, Trachtenberg J, Lee J, Tam S, Bristow R, Loblaw A, et al. Association between germline *HOXB13* G84E mutation and risk of prostate cancer. *J Natl Cancer Inst* 2012;104:1260–2.
- Karlsson R, Aly M, Clements M, Zheng L, Adolfsson J, Xu J, et al. A population-based assessment of germline *HOXB13* G84E mutation and prostate cancer risk. *Eur Urol* 2012.[Epub ahead of print].
- Stott-Miller M, Karyadi DM, King T, Kwon EM, Kolb S, Stanford JL, et al. *HOXB13* mutations in a population-based, case control study of prostate cancer. *Prostate* 2013;73:634–41.
- Stanford JL, FitzGerald LM, McDonnell SK, Carlson EE, McIntosh LM, Deutsch K, et al. Dense genome-wide SNP linkage scan in 301 hereditary prostate cancer families identifies multiple regions with suggestive evidence for linkage. *Hum Mol Genet* 2009;18:1839–48.
- Stanford JL, Wicklund KG, McKnight B, Daling JR, Brawer MK. Vasectomy and risk of prostate cancer. *Cancer Epidemiol Biomarkers Prev* 1999;8:881–6.
- Agalliu I, Salinas CA, Hansten PD, Ostrander EA, Stanford JL. Statin use and risk of prostate cancer: results from a population-based epidemiologic study. *Am J Epidemiol* 2008;168:250–60.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 2009;25:1754–60.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–6.
- O’Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011;43:585–9.
- Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglu S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005;15:901–13.
- Sunyaev S, Ramensky V, Koch I, Lathe W III, Kondrashov AS, Bork P. Prediction of deleterious human alleles. *Hum Mol Genet* 2001;10:591–7.
- Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–4.
- Grantham R. Amino acid difference formula to help explain protein evolution. *Science* 1974;185:862–4.
- Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res* 2009;19:1553–61.
- Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 1992;89:10915–19.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res* 2002;12:996–1006.
- Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods* 2009;6:315–6.
- O’Roak BJ, Vives L, Fu W, Egerton J, Stanaway I, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 2012;338:1619–22.
- Allen-Brady K, Wong J, Camp NJ. PedGenie: an analysis approach for genetic association testing in extended pedigrees and genealogies of arbitrary size. *BMC Bioinformatics* 2006;7:209.
- Breslow NE, Day NE. *Statistical methods in cancer research, volume 1—the analysis of case-control studies*. Lyon, France: International Agency for Research on Cancer; 1980.
- Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* 2012;487:239–43.
- Acevedo VD, Gangula RD, Freeman KW, Li R, Zhang Y, Wang F, et al. Inducible FGFR-1 activation leads to irreversible prostate adenocarcinoma and an epithelial-to-mesenchymal transition. *Cancer Cell* 2007;12:559–71.
- Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 2004;101:6062–67.
- Abeler-Dorner L, Swamy M, Williams G, Hayday AC, Bas A. Butyrophilins: an emergent family of immune regulators. *Trends Immunol* 2012;33:34–41.
- Arnett HA, Escobar SS, Gonzalez-Suarez E, Budelsky AL, Steffen LA, Boiani N, et al. *BTNL2*, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. *J Immunol* 2007;178:1523–33.
- Nguyen T, Liu XK, Zhang Y, Dong C. *BTNL2*, a butyrophilin-like molecule that functions to inhibit T cell activation. *J Immunol* 2006;176:7354–60.
- Valentonyte R, Hampe J, Huse K, Rosenstiel P, Albrecht M, Stenzel A, et al. Sarcoidosis is associated with a truncating splice site mutation in *BTNL2*. *Nat Genet* 2005;37:357–64.
- Rybicki BA, Walewski JL, Maliarik MJ, Kian H, Iannuzzi MC, Group AR. The *BTNL2* gene and sarcoidosis susceptibility in African Americans and whites. *Am J Hum Genet* 2005;77:491–9.
- Mitsunaga S, Hosomichi K, Okudaira Y, Nakaoka H, Kunii N, Suzuki Y, et al. Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the *BTNL2*. *J Human Genet* 2013;58:210–5.
- Franke A, Balschun T, Karlsen TH, Sventoraityte J, Nikolaus S, Mayr G, et al. Sequence variants in *IL10*, *ARPC2* and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 2008;40:1319–23.

44. Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* 2009;41:216–20.
45. Coussens LM, Werb Z. Inflammation and cancer. *Nature* 2002;420:860–7.
46. Nelson WG, De Marzo AM, DeWeese TL, Isaacs WB. The role of inflammation in the pathogenesis of prostate cancer. *J Urol* 2004;172:S6–11.
47. ENCODE. Available from: <http://genome.ucsc.edu/ENCODE/>.
48. Breyer JP, Avritt TG, McReynolds KM, Dupont WD, Smith JR. Confirmation of the *HOXB13* G84E germline mutation in familial prostate cancer. *Cancer Epidemiol Biomarkers Prev* 2012;21:1348–53.
49. Genome Variation Server. Available from: <http://pga.gs.washington.edu/>.
50. ClinSeq Project. Available from: <http://www.genome.gov/20519355>.
51. 1000 Genomes Project, 2008–2011. Available from: <http://www.1000genomes.org/>.