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

Liesel M. FitzGerald[1,*], **Akash Kumar**[2,*], **Evan A. Boyle**[2], **Yuzheng Zhang**[1], **Laura M. McIntosh**[1], **Suzanne Kolb**[1], **Marni Stott-Miller**[1], **Tiffany Smith**[3], **Danielle M. Karyadi**[3], **Elaine A. Ostrander**[3], **Li Hsu**[1], **Jay Shendure**[2], and **Janet L. Stanford**[1,4,**]

[1] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109-1024, U.S.A.

[2] Department of Genome Sciences, University of Washington, Seattle, WA, 98195-5065, U.S.A.

[3] Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD, 20892-8000, U.S.A.

[4] Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, 98195-7236, U.S.A.

Abstract

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

Methods—To identify rare mutations for PCa, we performed 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 ($n = 819$ PCa 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 PCa 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 PCa 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 PCa susceptibility gene.

[**]Corresponding author: Janet L. Stanford, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., M4-B874, Seattle, WA 98109-1024 (jstanfor@fhcrc.org; phone: +1 206-667-2715; fax: +1 206-667-2717).
[*]These authors contributed jointly to this work.

Conflict of Interest

There are no conflicts of interest.

Introduction

PCa 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 PCa have identified over 70 common lowpenetrance single nucleotide polymorphisms (SNPs) that are confirmed to be associated with weak to modest alterations (average per allele ORs = 1.1–1.3) in disease risk (2, 3), and which taken together may explain up to 30% of the genetic risk for PCa. In addition, genome-wide linkage studies of HPC families have searched for genomic regions that harbor rare, moderate- to high-penetrance mutations. These linkage studies have discovered more than two 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 PCa 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 four 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 PCa cases ascertained from the general population (12-14).

In order to find novel germline mutations for PCa, we completed one of the first 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 five or more affected men with at least three 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 two to six affected men ($n = 80$) and, where possible, one older, unaffected, PSA 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 (IBD), the majority of affected men selected are 2nd- or 3rd-degree relatives. The eleven 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 (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 PCa associated with genetic variants ($n = 2$) with supportive evidence from analyses of HPC families. Participants are from two population-based studies of PCa conducted 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.

Whole-Exome Sequencing (WES) in 19 HPC Families

A total of 10μg of genomic DNA per subject was sent to the Center for Inherited Disease Research (CIDR) for sequencing. For quality control and inheritance checks, all samples were first run on the OmniExpress Array (Illumina, Inc.). Once initial quality control checks were completed, 3μg 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 de-multiplexed at CIDR and fastq files were created for each sample. The Burrows-Wheeler Aligner (BWA) (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 (SNVs), 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) (21, 22).

SNVs were filtered on a family-level basis using four 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 NHLBI/NIH Exome Sequencing Project (ESP). The four filtering approaches were as follows: 1) MAF <0.02, present in all affecteds, not present in the unaffected male relative if available, and not present in any other unaffected males; 2) 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; 3) MAF <0.01, present in all affecteds, present in the unaffected male relative if available, and not present in any other unaffected males; and, 4) 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 Figure 1) 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 GERP (5) (23), PolyPhen (probably damaging) (24), SIFT (0.00–0.10) (25), Grantham (151) (26), PhyloP (3) (23), likelihood ratio test (damaging) (27), and BLOSUM62 (–2 to –4) (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 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 four 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 Figure 1) were then prioritized according to the following information: type (frameshift < UTR < 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 et al. (31). Briefly, 70 bp oligonucleotide inversion probes (Integrated DNA Technologies, Inc.) were designed against 196 candidate SNVs and indels (Supplementary Tables 1 and 2). These oligonucleotides were 5' phosphorylated and added to ~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 non-circularized 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 8x 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 SNV 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 $^2 P$-value by two for candidate variants that were observed more frequently in men with, compared to without, PCa, and one minus the $^2 P$-value divided by two for candidate variants that were observed less frequently in affected men than in unaffected men. A P-value < 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, Foster City, CA) was used to genotype the two *BTNL2* candidate variants (rs41441651 and rs28362675) on the ABIPrism 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 odds ratio (OR) and 95% confidence interval (CI) as a measure of association between the two *BTNL2* candidate variants and PCa (33), as implemented in STATA version 11.0 (Stata Corp). Potential confounding factors, including age at reference date, PCa screening history, and first-degree family history of PCa, were examined to see if 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 SNV genotypes in men stratified by family history (yes vs. no). A product term between SNV 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 SNV genotypes differ 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 70x (range: 20x to 132x) coverage of the target was achieved (Agilent SureSelect Human All Exon 50 Mb, Illumina HiSeq $2 \times$ 75 bp) with ~88% of target bases having at least 8x coverage (Supplementary Table 3). 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 1 and 2) 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% (non-reference concordance was 82.5%), and 99.9% (non-reference 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 > 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 two 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 (LD) was not possible, however they are located within a single haplotype block of eight Kb. This and the fact that all but two of the 22 affected carriers with data were concordant for both variants (two men had poor coverage for rs28362675 and thus have unknown carrier status) suggest that these variants are in strong LD. The smallest P-value observed was for rs41441651 ($P = 0.0032$).

The rs41441651 variant segregated with affection status in two of the 19 WES families (Figure 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 PCa at diagnosis.

Gleason scores ranged from 5–9 (average 6.5). Of the unaffected genotyped men in these two 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 co-segregation with disease state.

The two *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 %) cases and 9 (0.9%) controls were carriers (Table 3). Both candidate variants were associated with statistically significant increases in the risk of PCa (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 PCa, but this subgroup analysis had limited power. The mean age at PCa diagnosis was 59.0 years for carriers of one or both variants and was 59.8 years for non-carriers ($P = 0.6$).

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 PCa. 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 LD among controls in the case-control dataset ($r^2 = 0.99$). This is the first WES study focused on aggressive or early onset PCa phenotypes and the first to implicate rare germline *BTNL2* variants as predisposing to familial and sporadic PCa.

There is some prior suggestive evidence for a role of *BTNL2* in PCa. A recent exome sequencing study of prostate tumor tissue from 50 patients with lethal PCa identified one patient who had a somatic *BTNL2* mutation (c.709+9T>G) (34). Also, Acevedo et al. (35) found that BTNL2 protein was significantly over-expressed in advanced PCa 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 immunological 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

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 biological activity of BTNL2, our results provide further support for the role of the inflammation pathway in the development of PCa (45, 46).

Four of seven 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 appears 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 LD 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 observed only in affected men in both the WES families and the 270 HPC family replication dataset. The variants segregated with disease in two 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 PCa cases and 0.9% of the population-based controls. These observations are similar to those seen for the *HOXB13* mutation, rs138213197 (10); two unaffected males were observed to carry the *HOXB13* mutation in the four HPC discovery families and a carrier frequency of 0.1% was observed in 1,401 controls. Further, studies of both familial and case-control datasets have indicated that while rs138213197 is significantly associated with PCa 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 two *BTNL2* variants are available from several recent sequencing efforts. In the large NHLBI GO Exome Sequencing Project (49) consisting of mixed race U.S.-based studies, the MAF for both variants is reported as 0.5% (chromosomes = 4542–4550). In the ClinSeq Project (50) consisting of individuals of European ancestry, the minor allele frequency (MAF) for both variants is 1.5% (chromosomes = 1310–1323). Finally, in a pilot 1000 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. PCa is a prevalent disease and the PCa 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. Due to 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 PCa. This study is the first to demonstrate the value of WES in large multiplex HPC families characterized by aggressive or early onset PCa, with replication in an independent HPC family dataset and a population-based case-control dataset. We identified two rare *BTNL2* variants that segregate with disease in two HPC families with sequencing data and that are carried only by affected men, but no unaffected men, in eight (rs28362675) and nine (rs41441651) of the additional 270 HPC families tested. We also found that these two variants are associated with statistically significant 2.5- to 2.7-fold elevations in the relative risk of PCa in the general population, with slightly over 2% of incident sporadic PCa 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 PCa genetic susceptibility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
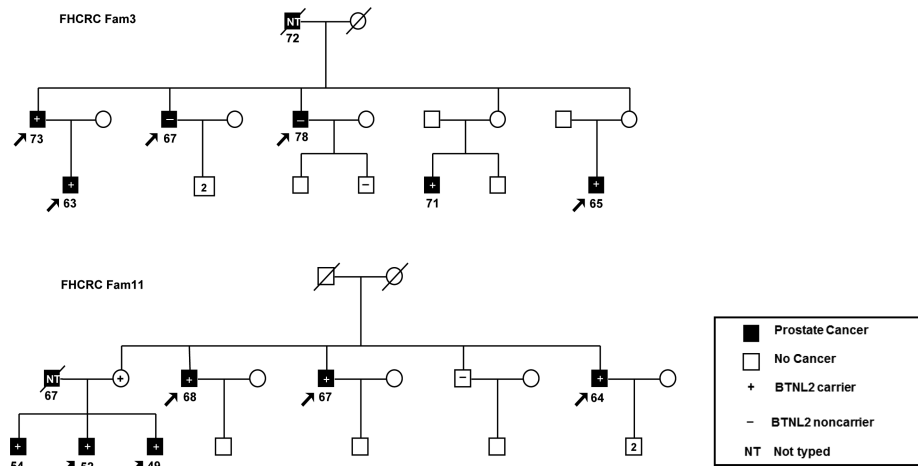
Acknowledgments

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. 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 Int Med. 2012; 271:353–65.

3. 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. [PubMed: 23535732]

4. Ostrander EA, Stanford JL. Genetics of prostate cancer: too many loci, too few genes. Am J Hum Genet. 2000; 67:1367–75. [PubMed: 11067781]

5. Easton DF, Schaid DJ, Whittemore AS, Isaacs WJ, ICPCG. Where are the prostate cancer genes?-- A summary of eight genome wide searches. Prostate. 2003; 57:261–9. [PubMed: 14601022]

6. Schaid DJ. The complex genetic epidemiology of prostate cancer. Hum Mol Genet. 2004; 13:R103–R21. [PubMed: 14749351]

7. 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. [PubMed: 12474142]

8. 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. [PubMed: 17700570]

9. 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. [PubMed: 21952622]

10. 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. [PubMed: 22236224]

11. 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. [PubMed: 23064873]

12. 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. [PubMed: 22781434]

13. 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. http://dx.doi.org/10.1016/j.eururo.2012.07.027

14. 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. [PubMed: 23129385]

15. 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. [PubMed: 19251732]

16. Stanford JL, Wicklund KG, McKnight B, Daling JR, Brawer MK. Vasectomy and risk of prostate cancer. Cancer Epidemiol Biomarkers Prev. 1999; 8:881–6. [PubMed: 10548316]

17. 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. [PubMed: 18556686]

18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009; 25:1754–60.

19. 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. [PubMed: 20644199]

20. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009; 25:2078–9.

21. 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. [PubMed: 19684571]

22. 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. [PubMed: 21572417]

23. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15:901–13. [PubMed: 15965027]

24. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. Hum Mol Genet. 2001; 10:591–7. [PubMed: 11230178]

25. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31:3812–4. [PubMed: 12824425]

26. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185:862–4. [PubMed: 4843792]

27. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009; 19:1553–61. [PubMed: 19602639]

28. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci USA. 1992; 89:10915–19. [PubMed: 1438297]

29. 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. [PubMed: 12045153]

30. 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. [PubMed: 19349981]

31. O'Roak BJ, Vives L, Fu W, Egertson J, Stanaway I, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012; 338:1619–22. [PubMed: 23160955]

32. 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. [PubMed: 16620382]

33. Breslow, NE.; Day, NE. Statistical Methods in Cancer Research, Volume 1-The Analysis of Case-Control Studies. International Agency for Research on Cancer; Lyon: 1980.

34. 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. [PubMed: 22722839]

35. Acevedo VD, Gangula RD, Freeman KW, Li R, Y. Z, Wang F, et al. Inducible FGFR-1 activation leads to irreversible prostate adenocarcinoma and an epithelialto-mesenchymal transition. Cancer Cell. 2007; 12:559–71. [PubMed: 18068632]

36. 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 USA. 2004; 101:6062–67. [PubMed: 15075390]

37. Abeler-Dorner L, Swamy M, Williams G, Hayday AC, Bas A. Butyrophilins: an emergine family of immune regulators. Trends in Immunol. 2012; 33:34–41. [PubMed: 22030238]

38. 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. [PubMed: 17237401]

39. 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. [PubMed: 16751379]

40. 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. [PubMed: 15735647]

41. 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. [PubMed: 16080124]

42. 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. [PubMed: 23364395]

43. 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. [PubMed: 18836448]

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. [PubMed: 19122664]

45. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002; 420:860–7. [PubMed: 12490959]

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. discussion S-2. [PubMed: 15535435]

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. [PubMed: 22714738]

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/

Figure 1.
HPC pedigrees of two families with segregating *BTNL2* variants. Participants selected for whole-exome sequencing 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.

Table 1

Characteristics of 19 hereditary prostate cancer (PCa) families with whole-exome sequencing (WES) data

Family ID	No. PCa cases	Mean age at PCa diagnosis	No. WES cases with aggressive PCa[a]	No. WES cases with early-onset PCa[a]	No. WES cases per family	No. 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

[a]A total of 23 cases had both aggressive and early-onset PCa.

[b]WES failed or was of low quality for one of the affected men in these families.

Table 2

Results for single nucleotide variants [a] identified by whole-exome sequencing and genotyped in 270 independent HPC families of European ancestry

Gene	Genomic Position (hg19)	Variant and rs ID	Protein	MAF in ESP	MAF in ClinSeq	Discovery No. of WES families with carriers (Aff/Unaff)[b]	No. (%) of 270 families with affected carriers	Validation MAF in 819 genotyped affected men[c]	MAF in 496 genotyped unaffected men[c]	P-value[d]
BTNL2	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
BTNL2	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

[a] Top ranked ($P < 0.05$) single nucleotide variants identified by whole-exome sequencing (WES) of 19 HPC families.

[b] Aff = number of affected carriers / Unaff = number of unaffected carriers.

[c] The number of affected carriers for rs41441651 and rs28362675 is 12 and 10, respectively, in the 270 independent HPC families.

[d] Monte-Carlo based one-sided P-value from the PedGenie chi-square test for association based on the 270 independent HPC families.

Table 3

Odds ratios (OR) and 95% confidence intervals (CI) for prostate cancer associated with single nucleotide variants in *BTNL2* in European Americans

Genotype	Cases (*n* = 1,155)		Controls (*n* = 1,060)		OR[a]	95% CI	*P*-value
	n	%	n	%			
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

[a]Adjusted for age.

[b]One case is homozygous variant for both SNVs; 22 cases and 9 controls are heterozygous for both SNVs.