Fine-mapping the 2q37 and 17q11.2-q22 loci for novel genes and sequence variants associated with a genetic predisposition to prostate cancer

Virpi H. Laitinen[1], Tommi Rantapero[1], Daniel Fischer[2], Elisa M. Vuorinen[1], Teuvo L.J. Tammela[3], PRACTICAL Consortium, Tiina Wahlfors[1] and Johanna Schleutker[1,4]

[1] BioMediTech, University of Tampere and Fimlab Laboratories, FI-33520, Tampere, Finland
[2] School of Health Sciences, University of Tampere, FI-33014 Tampere, Finland
[3] Department of Urology, Tampere University Hospital and Medical School, University of Tampere, FI-33520 Tampere, Finland
[4] Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, FI-20014 Turku, Finland

Cancer Genetics

The 2q37 and 17q12-q22 loci are linked to an increased prostate cancer (PrCa) risk. No candidate gene has been localized at 2q37 and the *HOXB13* variant G84E only partially explains the linkage to 17q21-q22 observed in Finland. We screened these regions by targeted DNA sequencing to search for cancer-associated variants. Altogether, four novel susceptibility alleles were identified. Two *ZNF652* (17q21.3) variants, rs116890317 and rs79670217, increased the risk of both sporadic and hereditary PrCa (rs116890317: OR = 3.3–7.8, $p = 0.003$–3.3×10^{-5}; rs79670217: OR = 1.6–1.9, $p = 0.002$–0.009). The *HDAC4* (2q37.2) variant rs73000144 (OR = 14.6, $p = 0.018$) and the *EFCAB13* (17q21.3) variant rs118004742 (OR = 1.8, $p = 0.048$) were over-represented in patients with familial PrCa. To map the variants within 2q37 and 17q11.2-q22 that may regulate PrCa-associated genes, we combined DNA sequencing results with transcriptome data obtained by RNA sequencing. This expression quantitative trait locus (eQTL) analysis identified 272 single-nucleotide polymorphisms (SNPs) possibly regulating six genes that were differentially expressed between cases and controls. In a modified approach, prefiltered PrCa-associated SNPs were exploited and interestingly, a novel eQTL targeting *ZNF652* was identified. The novel variants identified in this study could be utilized for PrCa risk assessment, and they further validate the suggested role of *ZNF652* as a PrCa candidate gene. The regulatory regions discovered by eQTL mapping increase our understanding of the relationship between regulation of gene expression and susceptibility to PrCa and provide a valuable starting point for future functional research.

What's new?

Prostate cancer runs in families, but its heritability isn't completely explained by the genetic variants identified to date. In this paper, the authors delve deeper into two loci that have been linked to prostate cancer. Sequencing data revealed four new alleles within these loci that correlate with increased prostate cancer risk. The authors then used the eQTL mapping technique to identify six genes which may be regulated by variants within these two loci, genes which had not previously been associated with prostate cancer.

A large proportion of familial prostate cancer (PrCa) cases can be explained by genetic risk factors.[1] Despite extensive research, the identification of these factors has proven challenging. In Finland, mutations in hereditary prostate cancer (HPC) risk genes are relatively rare, with the exception of the *HOXB13* G84E mutation,[2] which is present in 8.4% of familial PrCa cases and has been significantly associated with an increased PrCa risk in unselected cases.[3]

The involvement of chromosomal regions 2q37 and 17q12-q22 with PrCa has been previously reported in numerous linkage[4–6] and genome-wide association studies (GWASs).[7,8] Cropp *et al.*[9] performed a genome-wide linkage scan of 69 Finnish high-risk HPC families and in the dominant model, the loci on 2q37.3 and 17q21-q22 exhibited the strongest linkage signals. No known PrCa candidate gene resides on 2q37.3, and as demonstrated in our earlier study, the *HOXB13* G84E mutation only partially explains the observed linkage to 17q21-q22.[3]

Here, we performed targeted resequencing that covered the linkage peaks on 2q37 and 17q11.2-q22. The sequence data were filtered to identify the variants within genes predicted to be involved in PrCa predisposition. These variants were validated in Finnish HPC families and in unselected PrCa patients by Sequenom genotyping, and several novel variants were discovered that were significantly associated with PrCa. To study the impact of single-nucleotide polymorphisms (SNPs) on the regulation of gene expression within the two linked regions, we performed transcriptome sequencing followed by expression quantitative trait loci (eQTL) mapping. eQTLs are known to modify the penetrance of rare

Cancer Genetics

deleterious variants and therefore likely contribute to genetic predisposition to complex diseases. New information was obtained on several genes as well as their regulatory elements that generated fresh insights into PrCa susceptibility, especially in HPC.

Material and Methods

All of the subjects were of Finnish origin. The samples were collected with written and signed informed consent. The cancer diagnoses were confirmed using medical records and the annual update from the Finnish Cancer Registry. The project was approved by the local research ethics committee at Pirkanmaa Hospital District and by the National Supervisory Authority for Welfare and Health.

Targeted resequencing of 2q37 and 17q11.2-q22

Based on the linkage analysis results from Cropp *et al.*,[9] 63 PrCa patients and five unaffected individuals belonging to 21 Finnish high-risk HPC families[10] were selected for targeted resequencing of the 2q37 and 17q11.2-q22 regions (Supporting Information Table S1). Each family had at least three first- or second-degree relatives diagnosed with PrCa. Paired-end next generation sequencing was performed at the Technology Centre, Institute for Molecular Medicine Finland (FIMM), University of Helsinki. The sequenced fragments spanned approximately 6.8 Mb for chromosome 2q and 21.6 Mb for 17q. The target regions were captured using SeqCap EZ Choice array probes (Roche NimbleGen, Madison, WI) and were sequenced on a Genome Analyzer IIx (Illumina, San Diego, CA) following the manufacturer's protocol. The read alignment and variant calling were performed according to FIMM's Variant-Calling Pipeline (VCP).[11]

Bioinformatics workflow for variant characterization

A schematic overview of our bioinformatics workflow is shown in Figure 1. Only those variants that were present in all the affected family members were selected for subsequent analysis. The variants were annotated using Ensembl V65 gene set retrieved from the UCSC Genome Browser.[12] The phenotypic effects of the variants were studied with three *in silico* pathogenicity prediction programs. MutationTaster[13] classifies single-nucleotide variants (SNVs) and small insertion/deletion polymorphisms (indels) as polymorphic or pathogenic. PolyPhen-2[14] and PON-P[15] only predict the effects of nonsynonymous SNVs that result in amino acid replacement. PolyPhen-2 classifies the variants as benign, possibly pathogenic or probably pathogenic, whereas PON-P defines them as neutral, unclassified or pathogenic. Variants categorized as pathogenic by at least one tolerance predictor were defined as pathogenic. In addition, minor allele frequencies (MAFs) were obtained from the dbSNP database and information on known PrCa-associated genes was retrieved from the COSMIC[16] and DDPC[17] databases. Pathway data were gathered from Pathway Commons,[18] KEGG[19] and WikiPathways[20] and Gene Ontology data were retrieved from
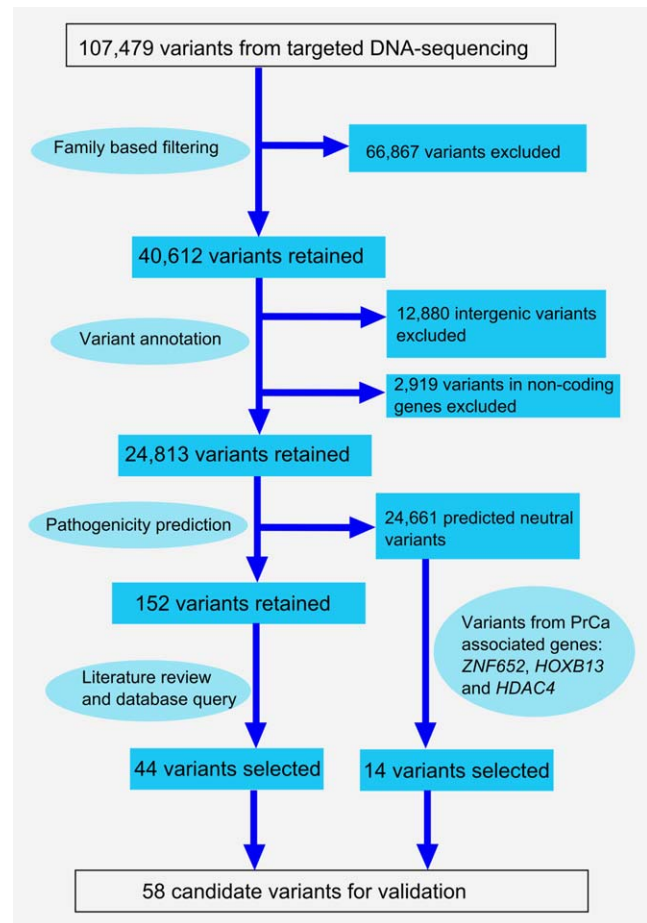
Figure 1. A flowchart describing the variant characterization pipeline. The targeted resequencing of 2q37 and 17q11.2-q22 from 68 Finnish HPC family members produced a total of 107,479 unique sequence variants. Family-based filtering excluded 66,867 variants that did not cosegregate with affection status. Annotation enabled the selection of 24,813 variants that were located within protein-coding genes. Pathogenicity predictions were performed *in silico* using MutationTaster, PolyPhen-2 and PON-P. As a result, the number of candidate variants was reduced to 152. The final filtering step exploited diverse information on genes and variants as well as gene ontology and pathway data stored in several public databases. In addition, select *HDAC4, ZNF652* and *HOXB13* variants, which were predicted to be nonpathogenic, were included in the validation because these genes have been associated with PrCa in previous studies.

Ensembl BioMart v.65.[21] Higher priority was assigned to rare variants (MAF <0.05), variants located in genes previously linked to PrCa, and variants located in genes functionally similar to PrCa-associated genes.

Validation of predicted PrCa-associated variants with Sequenom

After filtering, 58 variants in 35 target genes (listed in Supporting Information Tables S2–S4) were selected for validation which was performed on germline DNA from 2,216 subjects, including 1,293 cases and 923 population controls. The majority of the cases (1,105 individuals) represented

Cancer Genetics

unselected PrCa patients from the Pirkanmaa Hospital District, Tampere, Finland. In addition, 188 index cases from Finnish HPC families[10] were included in the study. The control DNA samples from anonymous male blood donors were provided by the Finnish Red Cross Blood Transfusion Service. Genotyping was performed at the Technology Centre, FIMM using the Sequenom MassARRAY system and iPLEX Gold assays (Sequenom, San Diego, CA). Genotyping reactions were performed with 20 ng of dried genomic DNA according to manufacturer's recommendations and with their reagents. The genotypes were called using TyperAnalyzer software (Sequenom). For quality control (QC) reasons, the genotype calls were also checked manually. Genotyping quality was examined using a detailed QC procedure that included success rate checks, duplicated samples and water controls.

Statistical and bioinformatic analyses of the validated variants

Association and Hardy-Weinberg equilibrium (HWE) tests were performed using PLINK.[22] The *p* value threshold for the HWE test was set to 0.05. Samples with low genotyping frequencies (<0.80) were excluded from the association analysis. The statistical significance of the association was evaluated using a two-sided Fisher's exact test. Odds ratios (OR) were calculated using PLINK with option — fisher. No further model adjustments for confounding factors were made. ENCODE information[23] for noncoding variants was retrieved from the Regulome database (RegulomeDB).[24] The linkage disequilibrium (LD) analysis of the statistically significant variants is described in Supplementary Methods.

Genotyping of the top four candidate variants in Finnish HPC families

Four variants were chosen for segregation analysis in Finnish HPC families based on a strong association with PrCa, a high OR value and/or predicted pathogenicity. The cosegregation of rs116890317 and rs79670217 in *ZNF652* (RefSeq NM_001145365), rs73000144 in *HDAC4* (RefSeq NM_006037) and rs118004742 in *EFCAB13* (RefSeq NM_152347) with affection status was determined in 41 families whose index cases were mutation-positive in the Sequenom validation. For these families, DNA samples were available from 243 PrCa cases and 204 healthy family members. The variants were genotyped in two to 17 (median: seven) individuals per family by Sanger sequencing.

RNA extraction and sequencing

Peripheral blood samples collected in PAXgene® Blood RNA Tubes (PreAnalytiX GmbH, Switzerland) were available from 84 PrCa patients and 15 healthy male relatives belonging to 31 Finnish HPC families. These included 11 families from the targeted resequencing step (Supporting Information Table S1) and additional 20 high-risk families.[10] Total RNA was purified with MagMAX™ for Stabilized Blood Tubes RNA

Isolation Kit (Ambion®/Life Technologies, Carlsbad, CA) and with a PAXgene Blood miRNA Kit (PreAnalytiX GmbH). RNA integrity and quality were analyzed using the Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA). The massively parallel paired-end RNA sequencing was performed at Beijing Genomics Institute (BGI Hong Kong Co., Tai Po, Hong Kong) using an Illumina HiSeq2000 sequencing platform (Illumina).

RNA sequencing data analysis

On average, RNA sequencing produced 45 million reads per sample. The QC check was performed using fastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc). The reads were aligned with Tophat2[25] using GRCh37/hg19 as the reference genome. The read counts for the genes were determined using HTSeq (http://www-huber.embl.de/users/anders/HTSeq/). The raw read counts were transformed into comparable expression values via normalization using the DESeq package for R[26] and the genes with very low or no expression (normalized read counts of <20) were removed. A differential gene expression analysis was then performed using a two-sided Mann–Whitney test with a *p* value cutoff of 0.05.

eQTL mapping and data analysis

The eQTL analysis was based on the RNA-seq data and on the SNP genotypes obtained from targeted DNA sequencing. This data existed for 19 samples at 2q37 and for 17 samples at 17q11.2-q22. In total, 54,919 SNPs (average 6,865 per gene, see Supporting Information Table S5 for details) were tested for association with their candidate target genes. Only genes with differential expression (DE) patterns between health status groups were included in the eQTL analysis, to increase the probability that found SNP-gene associations also link PrCa with a certain SNP genotype. The eQTL mapping was applied on 2q37 and 17q11.2-q22 to identify *cis*-regulated genes. SNPs associated in *cis* were defined as variants located within 1 Mb up- or downstream of the gene under study. The significance level for SNP-gene associations was set to $p \leq 0.005$. A multiple testing adjustment was omitted because of the large number of tested SNPs and the nature of the permutation type tests, acknowledging that this resulted in compromised resolution.

A modified *cis*-eQTL approach was also utilized, wherein a large genotype dataset from the iCOGS study[27] was used to preidentify possible PrCa-associated SNPs for 2,824 unselected Finnish PrCa patients and 2,440 controls. Here, Fisher's exact test with a modest significance level of 0.005 was used to study the association. Significant iCOGS variants that were also observed in the targeted DNA sequencing data were then selected for eQTL analysis, which was restricted to the fine-mapped regions. Additional details for the eQTL analysis are presented in Supplementary Methods.

RegulomeDB was used to annotate and assess the regulatory potential of the detected eQTLs.[24] The ENCODE

Cancer Genetics

Table 1. Variants significantly associated with prostate cancer based on a comparison of familial cases ($n = 186$) and controls ($n = 914$)

SNP Id	Function	Gene	Chr	Min/Maj	F_A/F_U (%)	*p* value	OR (95% CI)	Pathogenicity prediction
rs116890317	Intronic	*ZNF652*	17	A/T	2.96/0.39	**3.3×10^{-5}**	7.8 (3.0–20.3)	Polymorphism/−/−
rs79670217	Intronic	*ZNF652*	17	G/T	6.65/3.56	**0.009**	1.9 (1.2–3.1)	Polymorphism/−/−
rs10554930	Intronic	*HOXB3*	17	−ACA/ACA	27.5/21.3	**0.010**	1.4 (1.1–1.8)	Pathogenic/−/−
rs35384813	5′-UTR	*HOXB3*	17	+T/−	26.7/20.8	**0.013**	1.4 (1.1–1.8)	Pathogenic/−/−
rs73000144	Missense	*HDAC4*	2	T/C	0.80/0.06	**0.018**	14.6 (1.5–140.2)	Polymorphism/benign/neutral
rs13411615[1]	Near gene 5′	*MYEOV2*	2	C/A	52.1/45.6	**0.023**	1.3 (1.0–1.6)	Polymorphism/−/−
rs9899142	Intronic	*HOXB13*	17	T/C	11.2/15.6	**0.031**	0.7 (0.5–1.0)	Polymorphism/−/−
rs118004742	Nonsense	*EFCAB13*	17	G/T	4.79/2.73	**0.048**	1.8 (1.0–3.1)	Pathogenic/−/−
rs142044482	3′-UTR	*ZNF652*	17	+A/−	2.94/1.59	0.087	1.9 (0.9–3.8)	Polymorphism/−/−
rs140611363[1]	Near gene 5′	*ACACA*	17	−A/A	28.8/31.1	0.421	0.9 (0.7–1.1)	Pathogenic/−/−
rs72828246[1]	Near gene 5′	*ACACA*	17	G/A	28.8/30.9	0.459	0.9 (0.7–1.2)	Pathogenic/benign/neutral
rs13406410[1]	Near gene 5′	*MYEOV2*	2	C/T	47.6/46.8	0.817	1.0 (0.8–1.3)	Pathogenic/−/−
rs61752234	Synonymous	*HDAC4*	2	C/T	7.22/6.83	0.823	1.1 (0.7–1.6)	Polymorphism/−/−

Bold signifies $p < 0.05$.
[1]Variants are in linkage disequilibrium.
Abbreviations: Chr: chromosome; Min: minor allele; Maj: major allele; F_A: frequency of the minor allele in cases; F_U: frequency of the minor allele in controls; OR: odds ratio; CI: confidence interval; pathogenicity prediction results from: MutationTaster/PolyPhen-2/Pon-P.

datasets[23] were retrieved from the UCSC Genome Browser website for visualization purposes using the Table Browser tool.[12] As a general indicator of regulatory potential, we used the dataset that contained enriched DNase hypersensitive sites in 125 cell types. To highlight the regulatory potential of eQTLs in PrCa tissue, we used the LNCaP DNase (wgEncodeAwgDnaseUwDukeLncapUniPk) and LNCaP (Andr) DNase (wgEncodeAwgDnaseUwDukeLncapandrogenUniPk) datasets containing DNase hypersensitive sites in LNCaP cells under normal and androgen-induced conditions, respectively. Transcription factor (TF) binding site data were gathered from the Txn Fac ChIP V3 dataset, which contains ChIP-seq experimental data on 91 cell types and 189 TFs.

Results

Targeted DNA sequencing data analysis

The percentage of mapped reads was 95.0 and 95.7% for the samples sequenced for 2q37 and 17q11.2-q22, respectively. The target coverage was 99.8% for 2q37 and 99.5% for 17q11.2-q22. Correspondingly, the percentage of bases having coverage of $20\times$ or more was 79.9 and 63.4%. The total number of unique variants across all samples discovered by the utilized VCP was 107,479 (Fig. 1). Among the 41 predicted pathogenic variants in 2q37, there were 20 missense SNVs, 16 noncoding SNVs and five indels. Of all 111 predicted pathogenic variants in 17q11.2-q22, two variants were nonsense SNVs, 49 were missense SNVs, 36 were noncoding SNVs and 24 were indels.

PrCa-associated variants identified by Sequenom validation

Following prioritization, a total of 58 variants were selected for validation in a larger sample set (Supporting Information Table S2). In the QC analysis, four variants failed the HWE

test ($p < 0.05$), and 20 samples were omitted due to low genotyping frequencies (<0.80). In the case-control association analysis, a total of 13 variants in seven different genes were statistically significantly associated with PrCa ($p < 0.05$; Tables 1 and 2 and Supporting Information Tables S3 and S4). Three variants were located in the *ZNF652* gene at 17q21.3, and the *HDAC4* (2q37.2), *HOXB3* (17q21.3), *ACACA* (17q21) and *MYEOV2* (2q37.3) genes harbored two variants each. A single variant was identified in the *HOXB13* and *EFCAB13* genes at 17q21.3. Only three of these 13 PrCa-associated variants were located within exons, whereas the majority, 10 variants, resided in noncoding regions.

Four of the variants with a statistically significant association with PrCa were present in both the familial and the unselected sample sets. These were rs116890317 and rs79670217 in *ZNF652*, rs10554930 in *HOXB3* and rs13411615 in *MYEOV2*. The two *ZNF652* variants had the strongest association with an increased PrCa risk. rs116890317 had the most significant association with the familial cases (OR = 7.8, 95% CI 3.0–20.3, $p = 3.3 \times 10^{-5}$) and the same variant conferred the highest risk of 3.3 (95% CI 1.4–7.5, $p = 0.003$) among the unselected cases. rs79670217 had the most significant association with PrCa in the unselected sample set ($p = 0.002$) and was the second most significant variant in the familial PrCa patients (OR = 1.9, 95% CI 1.2–3.1, $p = 0.009$; Tables 1 and 2).

The highest OR of 14.6 (95% CI 1.5–140.2, $p = 0.018$) was observed for the *HDAC4* variant rs73000144 (*c.958C>T, p.Val320Ile*) among the familial samples (Table 1). Only three familial PrCa patients (1.6%), seven unselected patients (0.6%) and one control individual (0.1%) carried the minor allele in a heterozygous state, and none of the genotyped individuals were homozygous. rs73000144 was predicted to

Table 2. Variants significantly associated with prostate cancer based on a comparison of unselected cases ($n = 1096$) and controls ($n = 914$)

SNP Id	Function	Gene	Chr	Min/Maj	F_A/F_U (%)	p value	OR (95% CI)	Pathogenicity prediction
rs79670217	Intronic	ZNF652	17	G/T	5.66/3.56	**0.002**	1.6 (1.2–2.2)	Polymorphism/−/−
rs116890317	Intronic	ZNF652	17	A/T	1.27/0.39	**0.003**	3.3 (1.4–7.5)	Polymorphism/−/−
rs13406410[1]	Near gene 5′	MYEOV2	2	C/T	51.5/46.8	**0.006**	1.2 (1.1–1.4)	Pathogenic/−/−
rs61752234	Synonymous	HDAC4	2	C/T	4.85/6.83	**0.008**	0.7 (0.5–0.9)	Polymorphism/−/−
rs142044482	3′-UTR	ZNF652	17	+A/-	0.68/1.59	**0.009**	0.4 (0.2–0.8)	Polymorphism/−/−
rs140611363[1]	Near gene 5′	ACACA	17	−A/A	27.9/31.1	**0.032**	0.9 (0.7–1.0)	Pathogenic/−/−
rs10554930	Intronic	HOXB3	17	−ACA/ACA	24.1/21.3	**0.034**	1.2 (1.0–1.4)	Pathogenic/−/−
rs13411615[1]	Near gene 5′	MYEOV2	2	C/A	49.0/45.6	**0.037**	1.1 (1.0–1.3)	Polymorphism/−/−
rs72828246[1]	Near gene 5′	ACACA	17	G/A	28.0/30.9	**0.044**	0.9 (0.8–1.0)	Pathogenic/benign/neutral
rs35384813	5′-UTR	HOXB3	17	+T/−	23.2/20.8	0.073	1.1 (1.0–1.3)	Pathogenic/−/−
rs73000144	Missense	HDAC4	2	T/C	0.33/0.06	0.078	5.9 (0.7–47.9)	Polymorphism/benign/neutral
rs118004742	Nonsense	EFCAB13	17	G/T	3.0/2.7	0.637	1.1 (0.8–1.6)	Pathogenic/−/−
rs9899142	Intronic	HOXB13	17	T/C	16.1/15.6	0.665	1.0 (0.9–1.2)	Polymorphism/−/−

Bold signifies $p < 0.05$.
[1]Variants are in linkage disequilibrium.
Abbreviations: Chr: chromosome; Min: minor allele; Maj: major allele; F_A: frequency of the minor allele in cases; F_U: frequency of the minor allele in controls; OR: odds ratio; CI: confidence interval; pathogenicity prediction results from: MutationTaster/PolyPhen-2/Pon-P.

be benign or neutral by all three *in silico* pathogenicity prediction algorithms (Supporting Information Table S2).

The rs118004742 nonsense mutation (*c.1638T>G, p.Tyr546Ter*) in the *EFCAB13* gene was predicted to be pathogenic by MutationTaster (Supporting Information Table S2). Three familial cases (1.6%) were homozygous for the minor allele. There were 12 heterozygotes among the familial index cases (6.5%) and 66 among the unselected cases (6.0%). A statistically significant association between rs118004742 and PrCa was only observed for the familial patients (Table 1). The OR of 1.8 (95% CI 1.0–3.1) suggested an increased risk of HPC. rs118004742 carriers in the unselected sample set did not have an increased cancer risk (OR = 1.1, 95% CI 0.8–1.6, $p = 0.637$; Supporting Information Table S4).

Two common noncoding variants in the *HOXB3* gene, rs10554930 and rs35384813, had a moderate effect on PrCa risk, with ORs ranging from 1.2 to 1.4 (Tables 1 and 2). MutationTaster predicted both of these variants to be pathogenic (Supporting Information Table S2). For five variants, the ORs were < 1.0, indicating a modulatory role in PrCa predisposition. These variants were located near or within the *ZNF652, HDAC4, HOXB13* and *ACACA* genes (Tables 1 and 2). According to the RegulomeDB, three of the 13 statistically significant variants were likely to affect protein binding: rs9899142 in *HOXB13* (Regulome score of 1f), rs13406410 in *MYEOV2* and rs72828246 in *ACACA* (both having Regulome score of 2b).

In case-case comparisons, none of the identified variants were significantly associated with Gleason score, average age or the serum prostate specific antigen (PSA) level at diagnosis (data not shown). The LD analysis (Supporting Information Fig. S1) revealed that none of our 13 statistically significant variants (Tables 1 and 2) were in linkage disequilibrium with previously reported PrCa-associated variants[27] (see Supplementary Results for details).

Segregation analysis of the top four candidate variants

Altogether, 41 familial index cases out of 188 genotyped by Sequenom carried at least one of the top four candidate variants. Segregation analysis was performed for these 41 HPC families. rs116890317, rs79670217 and rs118004742 were more common among PrCa patients than healthy family members and provided evidence for cosegregation with affection status in 20 families (Supporting Information Tables S6–S8). However, in 15 of these families, unaffected male mutation carriers were also observed. In seven families, all of the unaffected male carriers were young enough (<55 years) to develop PrCa later in life. rs116890317 segregated completely with affection status in one family (Supporting Information Fig. S2a), as did rs79670217 (Supporting Information Fig. S2b). Complete segregation of rs118004742 was observed in three families (Supporting Information Table S8). The *HDAC4* variant rs73000144 was detected in three families, and approximately one-third of the family members were identified as carriers, irrespective of their health status (Supporting Information Table S9).

Multiple variants were observed in 16 individuals from 14 families. Two families harbored rs116890317, rs79670217 and rs118004742, whereas one family was positive for rs79670217, rs73000144 and rs118004742. In the remaining families, the most common combination detected was rs79670217 together with rs118004742 (six families). Evidence for segregation with affection status was obtained for a maximum of one variant per family.

Cancer Genetics
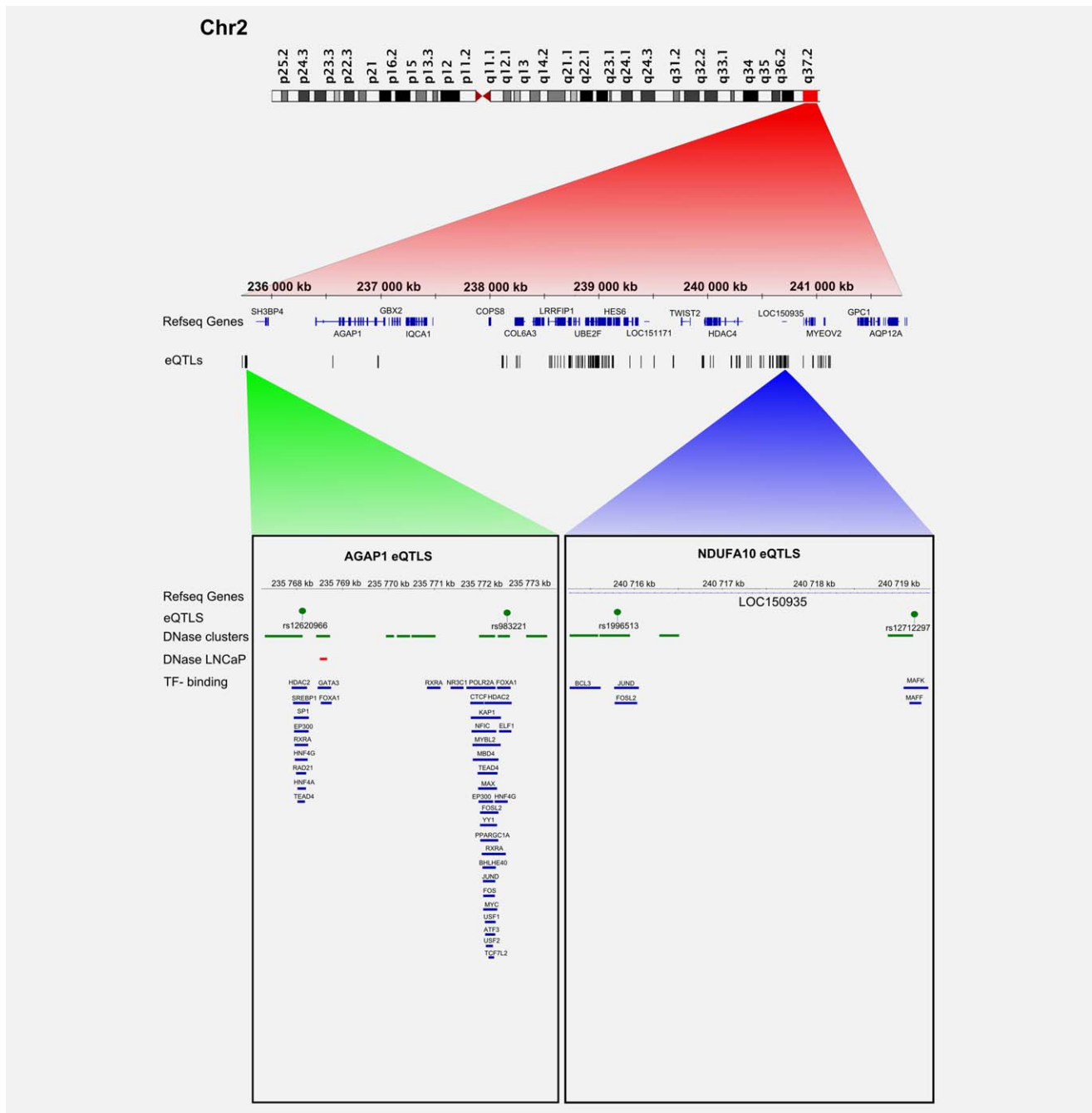
Figure 2. *Cis*-eQTLs targeting differentially expressed genes on chromosome 2. All statistically significant eQTLs are indicated with a track of black bars. Selected eQTLs, rs12620966 and rs983221 (targeting *AGAP1*) and rs1996513 and rs12712297 (targeting *NDUFA10*) are illustrated in more detail. DNaseI hypersensitive sites from the DNase cluster and LNCaP datasets are indicated with green and red rectangles, respectively. Blue rectangles denote TF binding sites.

eQTL mapping results

Differential gene expression analysis revealed three genes (of 173 tested) located at 2q37 and five genes (of 761 tested) at 17q11.2-q22 whose expression levels differed significantly between cases and controls ($p < 0.05$). In the targeted *cis*-eQTL analysis, SNPs within 2 Mb windows were tested for association with each of these eight DE genes (Supporting Information Table S5). Altogether, 272 candidate regulatory SNPs were identified for six DE genes only (Supporting Information Table S10). A vast majority, 237 candidate SNPs potentially regulate the expression of *AGAP1, SCLY* and *NDUFA10* at 2q37 (Fig. 2). The remaining 35 candidate SNPs possibly regulate *TBKBP1, PNPO* and *NAGS* at 17q11.2-q22 (Fig. 3). Based on the ENCODE data, the strongest evidence for regulatory potential was found for rs11650354 on chromosome 17, which targets the *TBKBP1*
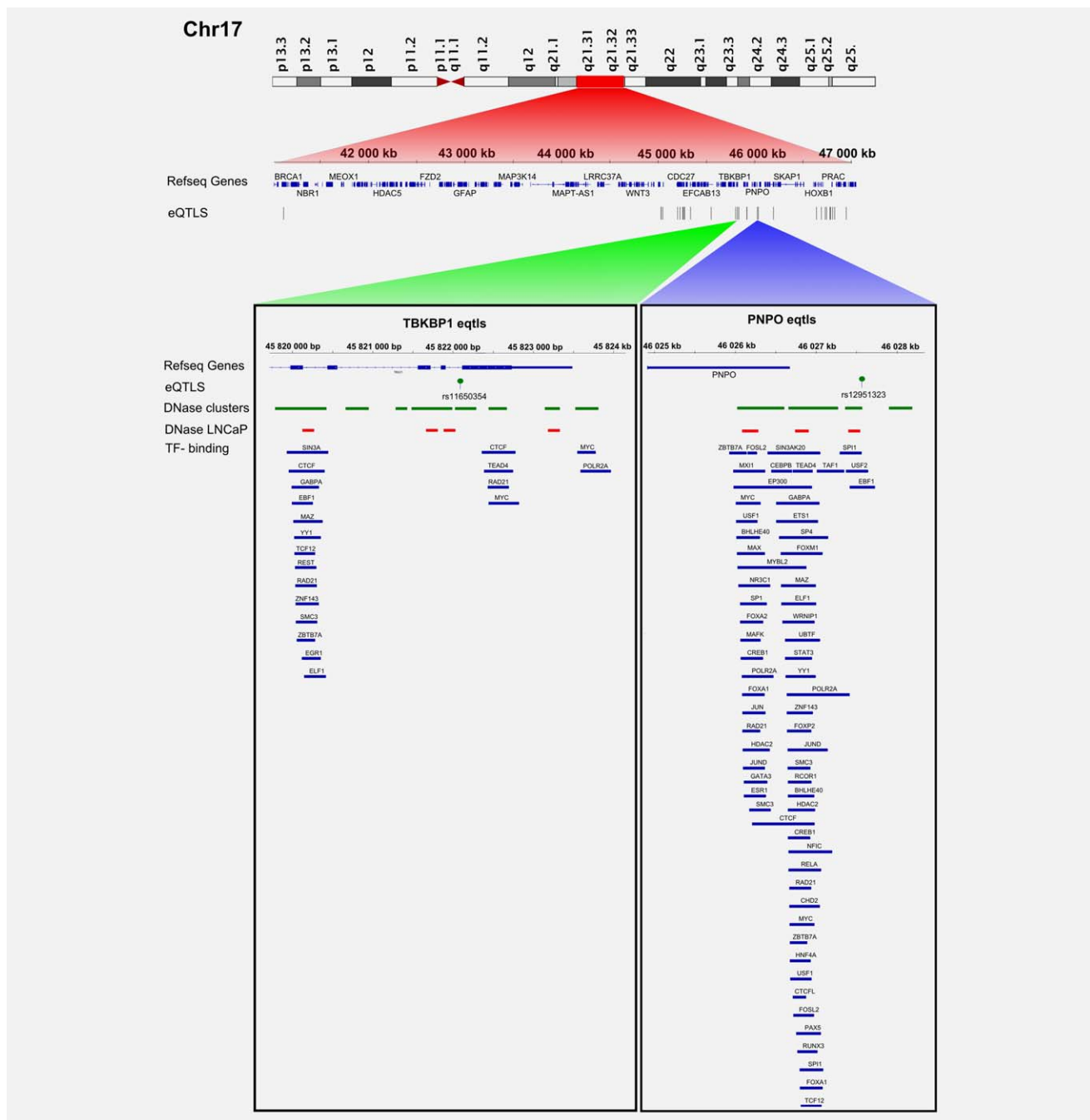
Figure 3. *Cis*-eQTLs targeting differentially expressed genes on chromosome 17. All statistically significant eQTLs are indicated with a track of black bars. Selected eQTLs, rs11650354 (targeting *TBKBP1*) and rs12951323 (targeting *PNPO*) are illustrated in more detail. DNaseI hypersensitive sites from the DNase cluster and LNCaP datasets are indicated with green and red rectangles, respectively. Blue rectangles denote TF binding sites.

gene. This known eQTL overlaps with an open chromatin region (Mcf7 and Gm12892 cell lines) and its role in the regulation of *TBKBP1* expression has been confirmed in a previous study.[28] rs12620966 targeting *AGAP1* on chromosome 2 overlaps with several TF binding sites discovered by ChIP-seq (HepG2 cell line), position weight matrix (PWM) matching and digital DNaseI footprinting studies (Supporting Information Table S10). None of the coding variants that

were identified by targeted DNA sequencing and validated by Sequenom were statistically significant eQTLs (data not shown).

The modified *cis*-eQTL analysis was based on 12 SNPs at 2q37 and 22 SNPs at 17q11.2-q22 that were shared between the iCOGS dataset and our set of variants obtained by targeted resequencing. The regulatory potential of these 34 SNPs was evaluated for 144 genes at 2q37 and for 160 genes

Cancer Genetics

at 17q11.2-q22. The modified eQTL approach identified only one PrCa-associated candidate eQTL on chromosome 2 and 36 candidate eQTLs on chromosome 17. Selected examples of these eQTLs and their target genes are shown in Supporting Information Table S11. The ENCODE data from RegulomeDB indicated the strongest evidence of regulatory potential for two variants on chromosome 17, rs4796751 and rs4796616, which target the *DHX58, MLX* and *JUP* genes. Both variants have previously been reported as eQTLs targeting *MGC20781* and *NT5C3L*[29] and they overlap with open chromatin regions (in 16 and 17 cell lines, respectively). rs4796616 is also located within a TF binding site (U2OS cell line). Two additional chromosome 17 variants, rs4793943 and rs16941107 were defined as likely to affect gene expression. These variants target the *ZNF652* and *ARL17B* genes, respectively, and overlap with open chromatin regions (in 6 and 42 cell lines, respectively) as well as several TF binding sites (Supporting Information Table S11). Of particular interest was the chromosome 17 variant rs4793976 targeting the *SPOP* gene. Although no data for this eQTL was available in the RegulomeDB, the importance of *SPOP* in PrCa predisposition has been recognized.[30]

Discussion

Prior studies have identified a strong relationship between PrCa and linkage to chromosomal regions 2q37 and 17q11.2-q22. Inspired by the lack of candidate genes and mutations, we resequenced the linkage peaks and confirmed the sequencing results by validating select variants. As the number of variants provided by the VCP was high, their prioritization for validation was critical.

The variants that were statistically significantly associated with PrCa were clustered in two genes on chromosome 2q37, *HDAC4* and *MYEOV2*, and in five genes on chromosome 17q11.2-q22, *ZNF652, HOXB3, HOXB13, EFCAB13* and *ACACA* (Tables 1 and 2). Interestingly, four of these genes, *HDAC4, ZNF652, HOXB3* and *HOXB13* encode TFs. Transcriptional regulation plays an essential role in maintaining normal gene control, and mutations in genes coding for TFs have been identified in PrCa. Examples of commonly occurring alterations include the fusion of *TMPRSS2* with *ERG*, and mutations in genes coding for the forkhead-box family of TFs.[31]

The *ZNF652* gene at 17q21.3 codes for a DNA-binding transcriptional repressor protein with seven zinc finger motifs.[32] Highest expression levels have been detected in normal breast, prostate and pancreas, whereas in primary tumors and cancer cell lines, *ZNF652* expression is generally lower.[32] However, in PrCa, the coexpression of high levels of ZNF652 and the androgen receptor (AR) has been shown to increase the risk of PSA relapse.[33] In addition, the recently characterized ZNF652 DNA binding site was found in the promoters of several genes that are involved in PrCa development and progression.[34] ZNF652 also interacts with CBFA2T3, a puta-

tive breast cancer tumor suppressor, which has been shown to enhance the repressor activity of ZNF652.[32]

To date, only a single PrCa-associated risk variant has been identified in the *ZNF652* gene. rs7210100 has been reported to predispose men of African descent to PrCa. The risk allele is present at a low frequency (<1%) in non-African populations.[35] A possible European-specific risk variant, rs11650494, is located in a lincRNA just downstream of the *ZNF652* gene and was recently described by the PRACTICAL Consortium.[27] The present study identified two novel *ZNF652* gene variants, rs116890317 and rs79670217, which were significantly associated with PrCa in both familial and unselected cases. The risk association was particularly apparent in patients with a positive family history of the disease. Correspondingly, both variants showed evidence for at least partial cosegregation with affection status in a substantial portion of Finnish HPC families. Like rs7210100, these two novel variants are located in the first intron of the gene, suggesting that they may play a role in regulating *ZNF652* by affecting splicing events and/or tissue-specific expression.

The *HDAC4* gene at 2q37.2 encodes a well-characterized transcriptional repressor. HDAC4 has been reported to accumulate in the nucleus in hormone-refractory PrCa[36] and to bind to and inhibit the activity of AR by SUMOylation.[37] Here, we determined that the exonic *HDAC4* variant rs73000144 (*c.958C>T*) was significantly associated with familial PrCa (OR = 14.6, 95% CI 1.5–140.2, $p = 0.018$). The variant also had a high OR (=5.8, 95% CI 0.7–47.9) among the unselected cases (Supporting Information Table S4), suggesting an increased cancer risk, but this result was not statistically significant ($p = 0.078$). The pathogenicity of rs73000144 is uncertain. The resulting amino acid change, a substitution of isoleucine for valine (*p.Val320Ile*) is conservative and was not considered pathogenic by any of the *in silico* predictors used (Supporting Information Table S2). The strikingly high OR for the familial sample set, together with the observation that this variant was detected in only three out of 186 index cases from the Finnish HPC families, suggested that rs73000144 may be a private mutation. The importance of private mutations has been emphasized in many diseases, some of which are associated with specific ethnic groups.

The protein encoded by the *EFCAB13* (*EF-hand calcium binding domain 13*) gene at 17q21.3 contains a particular helix-loop-helix domain, the EF-hand, which is required for calcium ion binding. EF-hands are often found in calcium sensor and calcium signal modulator proteins. Ca^{2+} binding triggers a conformational change in the EF-hand motif, which leads to the activation or inactivation of target proteins. Currently, there is no evidence linking EFCAB13 with PrCa. The nonsense mutation rs118004742 in the *EFCAB13* gene introduces a premature stop codon, leading to a significant truncation of the nascent protein. Truncating mutations are generally considered deleterious and, as expected, rs118004742 was predicted pathogenic by MutationTaster (Supporting Information Table S2). The variant segregated

completely with affection status in three Finnish mutation-positive HPC families and showed evidence for partial cosegregation in four additional families. In these seven families, the variant was observed in all of the patients but in only half of the genotyped unaffected men (Supporting Information Table S8). It is possible that rs118004742 contributes to hereditary, but not sporadic, disease. Once a more detailed characterization of the EFCAB13 protein function is available, it will be possible to assess the indicative role of *EFCAB13* as a PrCa risk gene more accurately.

Considering the importance of the *HOXB13* variant G84E[2] in familial PrCa predisposition, we compared the families that were positive for the top four SNPs with the existing G84E genotyping data.[3] Interestingly, ten of the 11 families that were positive for the *ZNF652* variant rs116890317 also harbored G84E. In these ten families, 12/21 (57%) of PrCa patients carried both the rs116890317 variant and the *HOXB13* variant G84E. Cosegregation of the *ZNF652* variant rs79670217 (Supporting Information Table S7) and G84E was detected in 6/42 (14%) of affected individuals, and among the 31 PrCa patients carrying the *EFCAB13* variant rs118004742 (Supporting Information Table S8), G84E was identified in only 2 (6%) patients. In addition, one of the three PrCa patients carrying the *HDAC4* variant rs73000144 also carried G84E. The co-occurrence of the *ZNF652* variant rs116890317 with the *HOXB13* variant G84E suggests possible interaction between these two genomic regions and is an interesting issue for future research.

The *HOXB3* gene belongs to the same evolutionarily conserved *HOXB* gene family at 17q21-q22 as *HOXB13*. Recently, HOXB3 overexpression was observed in primary PrCa tissues, predicting poor survival.[38] In our study, two possibly pathogenic *HOXB3* variants were associated with a moderately increased PrCa risk, rs10554930 in both datasets and rs35384813 in the familial sample set only (Tables 1 and 2). rs10554930 is intronic, located ∼730 bp upstream of the *HOXB3* transcription start site (TSS), whereas rs35384813 is in the 5′-UTR of the gene. Most variants affecting the expression level of a particular gene are located near the TSS of that gene[29] making it possible that these two variants participate in the regulation of *HOXB3* gene expression.

The ENCODE data supported a possible regulatory role for three of the statistically significant noncoding variants validated by Sequenom. The intronic *HOXB13* variant rs9899142 likely affects the binding of ZNF263, a transcriptional repressor that participates in cell structure maintenance and proliferation.[39] This variant is also a known *cis*-eQTL that regulates the expression of the *SKAP1* gene which has been associated with PrCa-specific mortality.[40] The SNPs rs13406410 and rs72828246 are located near the 5′ ends of the *MYEOV2* and *ACACA* genes, respectively. Both of these variants likely affect the binding of E2F1. This TF plays a central role in DNA damage-induced apoptosis and DNA repair.[41] Recently, a strong correlation between E2F1 and increased expression of NuSAP, a protein that binds DNA to the mitotic spindle, was observed in recurrent PrCa.[42] The minor alleles of rs9899142, rs13406410 and rs72828246 had a low OR and were present at a high frequency in both cases and controls. Nevertheless, according to the common disease–common variant hypothesis, it is possible that the major alleles, rather than the minor alleles, explain a proportion of PrCa susceptibility.

The eQTL mapping enabled us to identify genomic regions that were likely to be regulated by variants in the 2q37 and 17q11.2-q22 loci. A drawback of the eQTL analysis was the use of peripheral blood for RNA-sequencing. However, fresh PrCa tissue is rarely available and, due to the multifocal nature of PrCa, the quality of prostate biopsies may be compromised. Postmortem material, on the other hand, represents expression profiles typical for end-stage disease, whereas our aim was to identify inherited mutations predisposing their carriers to PrCa. Therefore, we consider blood to be a valid starting point for expression profiling of the early changes in PrCa. It will be exciting to see whether future studies confirm our results in another, independent sample set, preferably a collection of PrCa tissue samples.

The traditional eQTL analysis identified six DE genes that were putatively regulated by eQTLs in *cis* (Figs. 2 and 3; Supporting Information Table S10). None of these genes has previously been associated with PrCa. The protein encoded by the *AGAP1* gene is involved in membrane trafficking and cytoskeleton dynamics.[43] SCLY and PNPO participate in metabolic processes, SCLY in the decomposition of L-selenocysteine[44] and PNPO in the biosynthesis of vitamin B6. The adaptor protein encoded by *TBKBP1* plays a role in the TNF-alpha/NF-kappa B signal transduction pathway.[45] NDUFA10 and NAGS are mitochondrial enzymes. NDUFA10, a member of the respiratory chain complex I, is responsible for electron transport.[46] NAGS catalyzes the formation of N-acetylglutamate, an activator of urea cycle enzyme CPSI.[47]

In the modified eQTL analysis, several *cis*-acting variants that were associated with altered gene expression were identified (Supporting Information Table S11). The most interesting finding was the association of rs4793943 with *ZNF652* expression. This interaction may alter the TF function of ZNF652, thereby modulating susceptibility to PrCa. Data from RegulomeDB suggest that rs4793943 may have a more generalized role in transcriptional regulation. It is located within the binding site of ZNF263[39] and it overlaps with HOXA9 and HOXB13 binding motifs. Both of these TFs have been connected with PrCa initiation and progression.[2,48] Furthermore, our data provided suggestive evidence that rs4793976 is an eQTL regulating the expression of SPOP (Supporting Information Table S11). *SPOP*, a putative tumor suppressor gene, is frequently mutated in localized and advanced prostate tumors.[30] *SPOP* mutations are regarded as driver lesions in prostate carcinogenesis[31] and the loss of SPOP expression may contribute to PrCa development.[49]

While interpreting the eQTL results, it is important to recall that the significant DE genes and SNP-gene

Cancer Genetics

associations could be identified merely by chance. The number of observed significant test results lies in the same magnitude as the number of expected significant test results, if the null hypothesis would hold for all performed tests. However, the risk of an excess of false positive results was accepted in favor of minimizing the risk of obtaining too many false negative results. Although several of the SNP-gene connections detected in this study achieved statistical significance, this does not necessarily indicate biological significance. Neither is the mechanism of interaction between the individual eQTLs and their target genes currently known. Further validation with independent datasets is required to confirm the significance of the SNP-gene associations identified here.

In conclusion, the present study demonstrated that next-generation sequencing is a valid and reliable approach for identifying novel disease-associated variants and mutations, especially those rare enough to escape the resolution of GWAS. In contrast to imputation and related prediction-based methods, next-generation sequencing methods provide true genotype data with a minimal error rate. The integrated analysis of rare and common variants with gene expression data generated unique knowledge of PrCa-associated variants with effects at the transcriptional level. This study provided a broader view of the causative factors in PrCa, implicating that regulatory variants co-operating with coding variants can modulate the inherited risk for the disease. The findings reported here encourage further research to elucidate the regulatory networks that control PrCa initiation and development.

References

1. Baker SG, Lichtenstein P, Kaprio J, et al. Genetic susceptibility to prostate, breast, and colorectal cancer among Nordic twins. *Biometrics* 2005;61: 55–63.

2. Ewing CM, Ray AM, Lange EM, et al. Germline mutations in HOXB13 and prostate-cancer risk. *N Engl J Med* 2012;366:141–9.

3. Laitinen VH, Wahlfors T, Saaristo L, et al. HOXB13 G84E mutation in Finland: population-based analysis of prostate, breast, and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev* 2013;22:452–60.

4. Xu J, Dimitrov L, Chang BL, et al. A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. *Am J Hum Genet* 2005;77:219–29.

5. Lange EM, Robbins CM, Gillanders EM, et al. Fine-mapping the putative chromosome 17q21–22 prostate cancer susceptibility gene to a 10 cM region based on linkage analysis. *Hum Genet* 2007;121:49–55.

6. Pierce BL, Friedrichsen-Karyadi DM, McIntosh L, et al. Genomic scan of 12 hereditary prostate cancer families having an occurrence of pancreas cancer. *Prostate* 2007;67:410–5.

7. Gudmundsson J, Sulem P, Steinthorsdottir V, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat Genet* 2007;39: 977–83.

8. Eeles RA, Kote-Jarai Z, Giles GG, et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* 2008;40:316–21.

9. Cropp CD, Simpson CL, Wahlfors T, et al. Genome-wide linkage scan for prostate cancer susceptibility in Finland: evidence for a novel locus on 2q37.3 and confirmation of signal on 17q21-q22. *Int J Cancer* 2011;129:2400–7.

10. Schleutker J, Matikainen M, Smith J, et al. A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease. *Clin Cancer Res* 2000;6:4810–5.

11. Sulonen AM, Ellonen P, Almusa H, et al. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol* 2011;12:R94,2011–12-9-r94.

12. Fujita PA, Rhead B, Zweig AS, et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research* 2010;.

13. Schwarz JM, Rodelsperger C, Schuelke M, et al. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7: 575–6.

14. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.

15. Olatubosun A, Valiaho J, Harkonen J, et al. PON-P: integrated predictor for pathogenicity of missense variants. *Hum Mutat* 2012;.

16. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* 2011;39:D945–50.

17. Maqungo M, Kaur M, Kwofie SK, et al. DDPC: dragon database of genes associated with prostate cancer. *Nucleic Acids Research* 2010;.

18. Cerami EG, Gross BE, Demir E, et al. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res* 2011;39:D685–90.

19. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28: 27–30.

20. Pico AR, Kelder T, van Iersel MP, et al. Wiki-Pathways: pathway editing for the people. *PLoS Biol* 2008;6:e184.

21. Flicek P, Ahmed I, Amode MR, et al. Ensembl 2013. *Nucleic Acids Res* 2013;41:D48–55.

22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.

23. ENCODE Project Consortium, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489: 57–74.

24. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7.

25. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105–11.

26. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010; 11:R106,2010–11-10-r106.

27. Eeles RA, Olama AA, Benlloch S, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 2013;45:385, 91, 391e1–2.

28. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond—The transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010;5: e10693.

29. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. *Nat Genet* 2007;39:1217–24.

30. Barbieri CE, Baca SC, Lawrence MS, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet* 2012;44:685–9.

31. Barbieri CE, Bangma CH, Bjartell A, et al. The mutational landscape of prostate cancer. *Eur Urol* 2013;64:567–76.

32. Kumar R, Manning J, Spendlove HE, et al. ZNF652, a novel zinc finger protein, interacts with the putative breast tumor suppressor CBFA2T3 to repress transcription. *Mol Cancer Res* 2006;4:655–65.

33. Callen DF, Ricciardelli C, Butler M, et al. Co-expression of the androgen receptor and the transcription factor ZNF652 is related to prostate cancer outcome. *Oncol Rep* 2010;23:1045–52.

34. Kumar R, Selth LA, Schulz RB, et al. Genome-wide mapping of ZNF652 promoter binding sites in breast cancer cells. *J Cell Biochem* 2011;112: 2742–7.

35. Haiman CA, Chen GK, Blot WJ, et al. Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21. *Nat Genet* 2011;43:570–3.

36. Halkidou K, Cook S, Leung HY, et al. Nuclear accumulation of histone deacetylase 4 (HDAC4) coincides with the loss of androgen sensitivity in hormone refractory cancer of the prostate. *Eur Urol* 2004;45:382, 9; author reply 389.

37. Yang Y, Tse AK, Li P, et al. Inhibition of androgen receptor activity by histone deacetylase 4 through receptor SUMOylation. *Oncogene* 2011;30:2207–18.

38. Chen J, Zhu S, Jiang N, et al. HoxB3 promotes prostate cancer cell progression by transactivating CDCA3. *Cancer Lett* 2013;330:217–24.

39. Frietze S, Lan X, Jin VX, et al. Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263. *J Biol Chem* 2010;285:1393–403.

40. Huang CN, Huang SP, Pao JB, et al. Genetic polymorphisms in androgen receptor-binding sites predict survival in prostate cancer patients receiving androgen-deprivation therapy. *Ann Oncol* 2012;23:707–13.

41. Biswas AK, Johnson DG. Transcriptional and nontranscriptional functions of E2F1 in response to DNA damage. *Cancer Res* 2012;72:13–7.

42. Gulzar ZG, McKenney JK, Brooks JD. Increased expression of NuSAP in recurrent prostate cancer is mediated by E2F1. *Oncogene* 2013;32:70–7.

43. Nie Z, Stanley KT, Stauffer S, et al. AGAP1, an endosome-associated, phosphoinositide-dependent ADP-ribosylation factor GTPase-activating protein that affects actin cytoskeleton. *J Biol Chem* 2002;277:48965–75.

44. Mihara H, Kurihara T, Watanabe T, et al. cDNA cloning, purification, and characterization of mouse liver selenocysteine lyase. Candidate for selenium delivery protein in selenoprotein synthesis. *J Biol Chem* 2000;275:6195–200.

45. Bouwmeester T, Bauch A, Ruffner H, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. *Nat Cell Biol* 2004;6:97–105.

46. Brandt U. Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* 2006;75:69–92.

47. Caldovic L, Morizono H, Gracia Panglao M, et al. Cloning and expression of the human N-acetylglutamate synthase gene. *Biochem Biophys Res Commun* 2002;299:581–6.

48. Chen JL, Li J, Kiriluk KJ, et al. Deregulation of a Hox protein regulatory network spanning prostate cancer initiation and progression. *Clin Cancer Res* 2012;18:4291–302.

49. Kim MS, Je EM, Oh JE, et al. Mutational and expressional analyses of SPOP, a candidate tumor suppressor gene, in prostate, gastric and colorectal cancers. *APMIS* 2013;121:626–33.

Cancer Genetics