Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Table of Contents

International Consortium for Prostate Cancer Genetics 2

ICPCG Aggressive Prostate Cancer Criteria 2

Stage-One Cases 2

Stage-One Controls 3

Stage-One Auxiliary Data 3

Bioinformatics processing, quality control and filtering 4

- Stage-One Bioinformatics 4
- Stage-One Quality Control 4
- Stage-One QC of Capture Efficiency 5
- Stage-Two Bioinformatics 6
- Stage-Two Pooling for Custom-Capture 6
- Stage-Two Quality Control 7
- Number of Variants per Tier for Stage-One and Stage-Two Analyses 7

Statistical Analysis Methods 8

- Variant Weights for Analyses 8
- Stage-One Covariate Evaluation 9
- Stage-Two Selection of Variants for Analyses 10
- Stage-Two Covariate Evaluation 10

Results: Stage-One: Screening by WES 11

Results: Stage-Two Significant Gene-Level Analyses 11

Supplemental Tables 12

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

International Consortium for Prostate Cancer Genetics

The 15 groups contributing to the study included: Australia (Cancer Council Victoria, Victoria, Australia); BC/CA/HI (Stanford University School of Medicine, Stanford, CA); CeRePP (Centre for Research on Prostatic Diseases, Tenon Hospital, Paris, France); Fred Hutchinson Cancer Research Center (Seattle, WA); Finland (University of Turku, Turku, Finland); Germany (University of Ulm, Ulm, Germany); Institute of Cancer Research (United Kingdom); John Hopkins University (Baltimore, Maryland); Louisiana State University Health Sciences Center (New Orleans, LA); University of Michigan (University of Michigan Medical School, Ann Arbor, MI); Northwestern University Feinberg School of Medicine (Chicago, IL); Karolinska Institutet (Stockholm, Sweden); Translational Genomics Research Institute (Phoenix, AZ); the University of Utah School of Medicine (Salt Lake City, UT); and the Mayo Clinic (Rochester, MN).

ICPCG Aggressive Prostate Cancer Criteria

Our ICPCG aggressive criteria were based on regional or distant stage (stage T3, T4, N1, or M1, based on pathology if radical prostatectomy was done; otherwise, clinical stage), tumor Gleason score at least seven, poorly differentiated grade (if Gleason grade was unavailable), a diagnostic pretreatment PSA at least 20 ng/ml, or death from metastatic PCa before age 65 years [1].

Stage-One Cases

We preferentially selected affected men from families that also had affected third-degree relatives (first cousins) to maximize identification of variants likely to be risk candidates,

although not all families met this criterion. We also preferentially selected individuals who had the most aggressive disease status, based upon tumor histologic grade, tumor stage, prostate-specific antigen (PSA) level and course after treatment, including PSA recurrence, development of metastases, and PCa-specific mortality, within the family.

Stage-One Controls

These included 494 de-identified samples from multiple studies, including 89 samples from the Mayo Clinic Community Biobank [2], 355 samples from studies of cardiovascular phenotypes, and 50 samples from studies of neuropathy. These samples had similar library preparation and sequencing conditions as the cases. All controls were presumed to be free of PCa at the time of collection and were of European ancestry.

Stage-One Auxiliary Data

The additional 140 FPC cases were from ICPCG member groups and required to meet the same selection criteria as the primary samples. The additional controls were 592 men from the NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (ARIC) study, downloaded from dbGaP. Secondary analyses included the cases and controls used in the primary analyses, pooled with 105 qc-passed PCa case and 272 qc-passed European ancestry controls from the auxiliary data.

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Bioinformatics processing, quality control and filtering

Stage-One Bioinformatics

Fastq files were obtained for all samples sequenced outside of Mayo Clinic. Bioinformatics processing of all whole exome sequencing (WES) samples was performed using GenomeGPS (formerly TREAT), a comprehensive analysis pipeline developed at Mayo Clinic [3]. Briefly, alignment to GRCh37 (hg19) reference was performed using Novoalign (v.07.13) with duplicates removed using PICARD and realignment and recalibration using the Genome Analysis Toolkit (GATK, v3.3). HaplotypeCaller was run in single-sample mode to call germline single nucleotide variants and small insertion/deletion variants following GATK's best practices [4-6]. Joint genotyping was performed including all cases and controls using GATK GenotypeGVCF. Variant Quality Score Recalibrator (VQSR) was applied separately to SNV and Indels following GATK best practices v3.4. Edit distance (ED) annotation was calculated for each variant by BLATing the 100bp sequence surrounding each variant (50bp upstream and downstream) against the GRCh37 reference and recording the number of matches.

Stage-One Quality Control

Quality control (QC) analysis was conducted using a Mayo Clinic-developed comprehensive, highly flexible QC workflow that produces sample and population-level QC metrics to evaluate sample quality, population stratification, sample relatedness, and potential batch effects. Metrics produced include (1) total and mapped reads, (2) coverage of the target capture region, (3) transition/transversion (TiTv) ratio, (4) known and novel variant counts, (5) sample heterozygosity, (6) sex chromosome call rate and homozygosity, (7) sample contamination estimate (FREEMIX) using VerifyBamID [G. Jun, 2012], (8) identity-by-descent

(IBD) probabilities and kinship coefficient using Pedigree Relationship EStimation Test (PREST-plus) [7]and KING-robust[8] as implemented in the SNPRelate R package, (9) population ancestry using Structure[9], (10) concordance with external array or a sample ID verification (SIDV) panel.

Population QC analyses, including relationship checking, Structure, and principal component analyses (PCA) were conducted using a pruned variant set in which high-quality variants were further filtered to exclude highly influential genomic regions (2q21, HLA, 8p23, and 17q21.31). Additional pruning to remove variants in high linkage disequilibrium (LD) was performed prior to running PREST-plus, Structure, and PCA. KING-robust results are not impacted by LD, thus no additional pruning was performed for these analyses. We used Structure to evaluate population admixture using 359 samples from the 1000 Genomes Project as population anchors[9].

Samples with low coverage (<90% of the exome capture region covered at 10X or <50% of capture covered at 40X), low on-capture call rate (<95%), low concordance with array genotypes (<99% concordance considering only on-capture variants, or <90% of array variants discovered by WES), sex errors, unexpected duplicates, relationship errors, cryptic relationships, moderate contamination (FREEMIX>0.03), and non-European ancestry were excluded (Table S3).

Stage-One QC of Capture Efficiency

All WES samples were sequenced using Illumina sequencers with 100bp paired end reads. However, different sequencing conditions and capture kits exhibited different capture

efficiency throughout the exome, resulting in systematic coverage and base quality differences between samples. In order to ensure consistent variant calling and reduce bias resulting from different capture efficiency, we defined an empirical coverage region for both primary and secondary analyses. The GATK walker CoveredByNSamplesSites was used to identify physical positions where at least 95% of samples had a sequencing depth of ≥5X. Two empirical coverage regions were constructed; one for the primary analysis and one for secondary analysis including auxiliary cases and controls. Due to more heterogeneity in capture kits included in the secondary analysis, including kits from Agilent, Illumina and Nimblegen (Supplementary Table S1), the empirical coverage region for the secondary analysis was smaller than that for the primary analysis (832k positions and 1M positions, respectively), resulting in fewer genes in the secondary analysis. In addition, we required variants to have VQSR PASS, >95% call rate, and Hardy-Weinberg equilibrium (HWE) test p-value > 1e-8. Finally, variants with ED > 4 were excluded.

Stage-Two Bioinformatics

Bioinformatics analysis of custom capture sequenced samples was performed using GenomeGPS v4.0 as described previously with alignment to the GRCh38 reference genome using BWA-MEM (v0.7.10) and following GATK's (v3.4-46) best practices.

Stage-Two Pooling for Custom-Capture

Two levels of pooling were performed for the custom-capture sequencing with 24 samples pooled prior to sequencing capture (pre-capture). Next, two pre-capture pools were combined for sequencing on a single flow cell lane, for a total of 48 samples per lane and 384 per

flow cell. Randomization was performed to ensure each pre-capture pool was balanced with respect to case-control status and source of the DNA (ICPCG contributing group). Pre-capture pools were then randomly assigned to a flow cell and lane. All samples underwent library preparation at the same time and were sequenced concurrently on the Illumina HiSeq 4000.

Stage-Two Quality Control

Quality control (QC) analysis was conducted as described above. Samples with low coverage (<90% of the capture region covered at 10X or <50% of capture covered at 40X), low on-capture call rate (<95%), two or more discordant genotypes in a 29 marker "finger-print" panel (either 0 mismatches and 2 null genotypes or 1 mismatch and 1 null genotype were allowed), contamination estimate > 0.03, relationship errors including unexpected duplicates and cryptic relateds, and females were excluded. In addition, we excluded samples with a high fraction of non-European ancestry. Variants with VQSR PASS or in the first tranche (threshold 99%), ED < 5, and >95% call rate, and HWE p-value > 1e-8 were eligible for analysis.

Number of Variants per Tier for Stage-One and Stage-Two Analyses

The number of variants within each tier for stage-one and stage-two analyses are provided below.

- Tier-1 variants: likely to cause protein truncation
- Tier-2 variants: nonsynonymous coding variants and in-frame indels
- Tier-3: all other variants

Stage-One:

Primary analysis:

- Tier 1: 10,581
- Tier 2: 151,531

- Tier 3: 263,228

Secondary analysis:

- Tier 1: 11,495
- Tier 2: 167,214
- Tier 3: 263,600

Stage-Two:

- Tier 1: 3,163
- Tier 2: 47,819
- Tier 3: 155,772

Statistical Analysis Methods

For both stage-one and stage-two analyses, we excluded variants that: 1) were outside of the capture region (empirical coverage region for stage-one; custom capture region for stage-two), 2) were monomorphic; 3) had call rate < 0.95; 4) had a Hardy-Weinberg test for equilibrium p-value < 10e-8; and 5) had flanking 100bp sequence that mapped to more than four places in the genome. We retained variants that passed quality score recalibration (described above).

All variants with minor allele frequency (MAF) > 0.001 were included in single variant association analyses to screen for common variants. In contrast, to screen for less common variants, all variants with MAF ≤ 5% were included in gene-based tests. Common exonic variants were used to compute principal components.

Variant Weights for Analyses

Variants were classified into tiers based on their likely functional effect determined by SNPeffect[10] for stage-one and CAVA[11] for stage-two. Tier-1 were those likely to truncate protein expression (i.e., nonsense, stop-loss, splice, and frameshift variants). Tier-2 were nonsynonymous coding variants and in-frame indels; all other variants were classified as Tier-3.

All Tier-1 variants were assigned a weight of 1. Tier-2 single nucleotide variants (SNVs) were assigned weights using scores obtained from the composite pathogenicity prediction algorithm called REVEL[12], while Tier-2 in-frame indels were assigned weights using SIFT-indel scores[13]. Finally, all other noncoding (Tier-3) variants were assigned weights using FIRE (Functional Inference of Regulators of Expression), a tool to score noncoding SNVs based on their cis-eQTL predictive potential[14]. A small number of variants with missing weights were excluded from analyses (122 of 415,340 in primary analysis). Because it was not obvious how to best combine Tier-2 indels and Tier-3 SNVs, we applied different weights to the latter two categories (Supplemental Table S2). This resulted in seven different gene-based analyses with these different weights.

Stage-One Covariate Evaluation

Principal component analysis was conducted using genotypes selected from complete WES including all European QC-passed samples. To determine which principal components might be important covariates to adjust for population stratification in association tests, we first identified principal components with Tracy-Widom (TW) test p-value < 0.05. Those principal components deemed significant were evaluated for association with affection status. For our primary analysis, five principal components were considered important and included as covariates in the association analyses. In addition, capture kit (V3 vs. V4+UTR) was also included as a covariate. For the secondary analysis including auxiliary cases and ARIC controls, 12 principal components were included as covariates.

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Stage-Two Selection of Variants for Analyses

Single-variant analyses evaluated all variants regardless of Tier. For gene-level analyses, the focus was on less-common variants so the analysis was restricted to variants with a MAF ≤ 5%. For gene-level analyses, only Tier-1 SNVs and indels and Tier-2 SNVs were included in the analyses. Variants that were not within any of our target genes were removed. If a variant mapped to multiple genes, it was included in the analysis for each of the genes. Both gene-burden and SKAT-optimal [15] analyses were conducted for each gene.

Stage-Two Covariate Evaluation

Principal component analysis was conducted using common genotypes selected from captured exonic regions, including all European QC-passed samples. Principal components with TW test p-value < 0.05 were evaluated for association with affection status. Other factors considered as covariates include ICPCG member group and the design factors: randomization date, library preparation group, plate, sequencing flow cell and lane, and DNA Qubit concentration. No sequencing design factor was significantly associated with case status (all p-values > 0.20). Three of the 12 ICPCG member groups did not contribute controls; thus, each of these three groups was combined with another group that was similar with respect to clustering based on principal components and missing call rate. In addition to this modified group variable, the log(Qubit) concentration and log(missing call rate) were the only covariates associated with cases/control status. Therefore, our final association model included the modified ICPCG group, log(Qubit) concentration, and log(missing call rate).

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Results: Stage-One: Screening by WES

After extensive quality control (QC) and limiting to European ancestry (see Table S3 for details), the samples used for stage-one primary analyses included 491 cases (253 singleton cases and 238 cases from 80 families: six families with two cases; 70 with three cases; four with four cases), and 429 controls (see Table 1 for clinical characteristics of cases). The auxiliary samples resulted in 105 PCa cases and 272 controls. Hence, our secondary analyses that combined primary data with auxiliary data included 596 cases and 701 controls. A total of 415,340 variants in 20,043 genes passed QC and were included in the primary analysis. For secondary analyses, 442,309 variants in 19,022 genes were identified following QC.

Results: Stage-Two Significant Gene-Level Analyses

For the xix genes that achieved statistical significance (p-value < 5e-5) based on gene-level analyses, details on the SKAT-O and SKAT-burden results are shown in Table S6. The burden score was the sum of the rare variants within each gene, and this burden was analyzed by logistic regression, regressing status on the burden while adjusting for covariates. These six genes with significant gene-level results were further explored by performing single-variant analyses for any variants that had an allele frequency of at least 0.001. Results for comparisons of all cases and familial cases with controls for all variants within these genes are summarized in Supplemental Table 7. Results for comparisons of aggressive cases with controls or with non-aggressive cases for all variants within the six genes are summarized in Supplemental Table S8. Two variants in the *FAM111A* gene were more frequent among aggressive cases than controls

(OR ~ 3.6). In addition, the frequency of the alternate allele for the variant rs112966887 in the *PABPC1* gene was more frequent among aggressive cases than controls (OR=5.43; 95%CI: 2.60, 11.35), while 13 other variants in this gene had much less frequent alternate alleles in aggressive cases compared with controls. Finally, two *QKI* variants were found to be much less frequent among aggressive cases than controls (OR ~ 0.46)

Supplemental Tables

Supplemental Table S1. Stage-One Exome Sequencing capture kits for QC-passed Mayo cases and controls and external auxiliary cases and controls.

	Primary Cases	Primary Controls	Auxiliary Cases	Auxiliary Controls
Sample Size	491	429	105	272
Exome Capture sequencing kit				
Agilent Human All Exon 50Mb	52	109	39	
Agilent Human All Exon V4_UTR	439	320		
Illumina TruSeq Exome			14	
Illumina 64Mb			52	
Broad - Agilent				138
WashU - Nimblegen				134

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Supplemental Table S2. Stage-One: Weights for different gene-based analyses

Analysis	Tier-1 SNV & INDELS	Tier-2 SNV	Tier-2 INDELS	Tier-3 SNV
1	1	REVEL	0	0
2	1	REVEL	0.25 x SIFT	0
3	1	REVEL	0.5 x SIFT	0
4	1	REVEL	1.0 x SIFT	0
5	1	REVEL	SIFT	0.25 x FIRE
6	1	REVEL	SIFT	0.5 x FIRE
7	1	REVEL	SIFT	1.0 x FIRE

REVEL (Rare Exome Variant Ensemble Learner)[12],

SIFT (Sorting Intolerant from Tolerant)[13]

FIRE (Functional Inference of Regulators of Expression)[14].

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Supplemental Table S3. Quality Control Exclusions for Stage-One and Stage-Two Data.

	Stage-1 Primary Cases	Stage-1 Primary Controls	Stage-1 Auxiliary Cases	Stage-1 Auxilliary Controls	Strage-2 Cases	Stage-2 Controls
Total Sample Number	539	494	140	592	3,105	2,156
Families (families with multiple cases)	366 (84)	---	84 (49)	589 (3)	3,105	---
Singletons	282	---	35	586	3,105	---
Low Coverage <90% at 10x and/or <50% at 40x	4	18	27	72	29	9
<99% Concordance with Genotyping Array	3	0	0	0	8	11
<90% Genotyping Array Discovery Rate	1	0	0	0	0	0
<95% Call Rate	0	0	6	1	1	0
Mis-specified Relationships	7	11	2	4	60	23
Sample Contamination/Poor Quality	0	12	0	5	20	41
Sex Errors	1	2	0	14	5	55
Truncated BAM files	0	2	0	0	0	0
Race Misspecification/Non-Caucasian	32	20	0	224	65	118
TOTAL UNIQUE SAMPLE EXCLUSIONS	48	65	35	320	188	257
Final Sample Count	491	429	105	272	2,917	1,899
N Cases	491	---	105	---	2,917	---
Families (families with multiple cases)	333 (80)	---	93 (45)	---	2,917	---
N Singletons	253	---	12	---	2,917	---

Supplemental Material: Two-Stage Study of Familial Prostate Cancer by Whole-Exome Sequencing and Custom Capture Identifies 10 Novel Genes Associated with Risk of Prostate Cancer

Supplemental Table S4. QC Passed European Samples by Site

	Stage-1				Stage-2	
	Primary Samples		Auxiliary Samples			
Site	Cases	Controls	Cases	Controls	Cases	Controls
ARIC	0	0	0	272	0	0
AUSTRALIA	15	0	14	0	263	53
BCCAHI	9	0	0	0	111	0
MAYO BIOBANK	0	0	0	0	0	492
CEREPP	15	0	0	0	271	50
FHCRC	70	0	0	0	215	106
FINLAND	41	0	0	0	86	67
GERMANY	10	0	52	0	157	127
ICR	1	0	39	0	566	470
JHU	74	0	0	0	167	91
LSU	1	0	0	0	15	8
MAYO	66	0	0	0	344	0
MAYO_CONTROL	0	429	0	0	0	0
MICHIGAN	52	0	0	0	188	0
NORTHWESTERN	5	0	0	0	167	157
QUEBEC	3	0	0	0	0	0
SWEDEN	8	0	0	0	146	94
UTAH	121	0	0	0	221	184
Total	491	429	105	272	2917	1899

Supplemental Table S5 . Methods to choose genes for stage-2 evaluation (1=yes, 0=no).

Gene	Method for Stage-1 Selection			Stage-1	ICPCG
	Gene-level	Variant-level	Misc[(1)]	Selected	Nominated
A1BG-AS1	0	0	1	1	0
AAAS	0	0	0	0	1
AACS	0	0	0	0	1
ABCA10	0	0	1	1	0
ABCA8	0	0	0	0	1
ABCB6	0	1	0	1	0
ABCD4	0	0	1	1	0
ABHD1	1	0	0	1	0
ABHD10	0	0	1	1	0
ABHD3	1	0	0	1	0
ABL1	0	0	0	0	1
ACACA	0	0	0	0	1
ACE2	0	0	1	1	0
ACOT4	1	0	0	1	0
ACSF2	1	0	0	1	0
ACSM3	1	0	0	1	0
ACTR6	0	0	1	1	0
ACVR1	0	0	0	0	1
ACVR1C	1	1	0	1	0
ACVRL1	0	0	0	0	1
ADAM29	0	1	0	1	0
ADAM32	1	0	0	1	0
ADAMTS18	0	0	0	0	1
ADAMTS5	0	0	1	1	0
ADH1B	1	0	0	1	0
ADM	1	1	0	1	0
ADSL	0	0	0	0	1
AGAP2	0	1	0	1	0
AGRN	0	1	0	1	0
AHCTF1	0	1	0	1	0
AIF1	0	0	1	1	0
AIFM2	0	1	0	1	0
AIMP1	1	0	0	1	0
AK3	1	0	0	1	0
AKNAD1	0	1	0	1	0

AKR1E2	1	0	0	1	0
AKT1	0	0	0	0	1
ALKBH2	1	1	0	1	1
ALMS1	0	1	0	1	0
ALOX12	1	1	0	1	0
ALOX15B	0	0	0	0	1
AMBN	0	1	0	1	0
AMDHD2	0	1	0	1	0
ANKDD1A	0	1	0	1	0
ANKFN1	0	0	0	0	1
ANKRD13C	0	0	1	1	0
ANKRD35	0	0	1	1	0
ANKRD54	1	1	0	1	0
ANO7	0	0	0	0	1
ANP32B	0	0	1	1	0
APC	0	0	0	0	1
APOBEC2	0	0	0	0	1
APOL1	1	1	0	1	0
APOL4	1	1	0	1	0
AQP8	0	1	0	1	0
AR	0	0	0	0	1
ARFGAP3	0	0	0	0	1
ARFRP1	0	0	1	1	0
ARHGAP27	0	0	0	0	1
ARL16	0	0	1	1	0
ARMC3	1	0	0	1	0
ARRDC1	0	0	1	1	0
ARVCF	0	0	0	0	1
AS3MT	0	0	0	0	1
ASB16	0	0	0	0	1
ASB8	0	1	0	1	0
ASCL2	0	0	0	0	1
ASNS	0	1	0	1	0
ASPRV1	0	1	0	1	0
ASTN2	1	0	0	1	0
ATCAY	0	0	1	1	0
ATF7IP	1	1	0	1	0
ATM	0	1	0	1	1
ATP5A1	1	0	0	1	0
ATP5G1	1	0	0	1	0

ATP6V1G1P	0	0	0	0	1
ATR	0	0	0	0	1
ATRIP	0	0	0	0	1
AVEN	0	0	0	0	1
AZI1	0	0	0	0	1
B4GALNT2	0	0	0	0	1
B4GALNT4	0	1	0	1	0
BAG6	0	0	1	1	0
BAI2	0	1	0	1	0
BAIAP3	0	0	0	0	1
BANK1	0	1	0	1	0
BARD1	0	0	0	0	1
BBS9	1	0	0	1	0
BCL6	1	1	0	1	0
BCLAF1	0	0	0	0	1
BHLHA15	0	0	0	0	1
BIN1	0	0	0	0	1
BLM	0	0	0	0	1
BMP2K	1	1	0	1	0
BMPR1A	0	0	0	0	1
BMPR1B	0	0	0	0	1
BNC2	0	1	0	1	0
BNIP3	1	0	0	1	0
BOD1	0	0	1	1	0
BOLA2	0	1	0	1	0
BOLA2B	0	1	0	1	0
BRAF	0	0	0	0	1
BRCA1	0	0	0	0	1
BRCA2	0	0	0	0	1
BRD2	0	0	0	0	1
BRD8	0	0	0	0	1
BRIP1	0	0	0	0	1
BST2	1	1	0	1	0
BTNL2	0	0	0	0	1
C10ORF32	0	0	0	0	1
C10ORF95	0	0	0	0	1
C12ORF10	0	0	0	0	1
C12ORF51	1	1	0	1	0
C14ORF39	0	0	0	0	1
C14ORF49	0	0	0	0	1

C16ORF79	0	0	1	1	0
C17ORF70	0	1	0	1	0
C17orf82	0	0	1	1	0
C19ORF57	1	1	0	1	0
C1ORF151	0	0	0	0	1
C1ORF192	1	0	0	1	0
C1QTNF8	0	1	0	1	0
C20ORF194	0	1	0	1	0
C22ORF23	0	0	0	0	1
C22orf31	0	0	1	1	0
C22ORF46	1	0	0	1	0
C2CD2	1	0	0	1	0
C2ORF43	0	0	0	0	1
C3ORF20	1	0	0	1	0
C5ORF43	1	0	0	1	0
C6ORF163	0	0	0	0	1
C6ORF226	0	0	1	1	0
C6ORF97	1	0	0	1	0
C7ORF34	0	0	1	1	0
C7ORF70	0	1	0	1	0
C8ORF46	0	0	1	1	0
C9ORF96	1	0	0	1	0
CABP1	0	0	1	1	0
CACNA1G	1	0	0	1	0
CACNG1	1	0	0	1	0
CACNG5	0	0	1	1	0
CALR	1	0	0	1	0
CAMKK2	0	0	0	0	1
CAPN3	0	1	0	1	0
CAPN5	0	1	0	1	0
CAPSL	0	0	1	1	0
CARD11	0	1	0	1	0
CASP1	0	0	0	0	1
CASP8	0	0	0	0	1
CASQ2	1	0	0	1	0
CATSPERG	0	0	0	0	1
CBX2	0	1	0	1	0
CCDC132	0	0	1	1	0
CCDC30	0	0	1	1	0
CCDC59	1	0	0	1	0

CCDC60	0	1	0	1	0
CCDC68	1	1	0	1	0
CCDC77	0	0	0	0	1
CCNB1	0	0	0	0	1
CCND1	0	0	0	0	1
CCNDBP1	0	0	1	1	0
CCNE1	0	0	0	0	1
CCR6	0	1	0	1	0
CD109	0	1	0	1	1
CD300LB	0	1	0	1	0
CDC25A	0	0	0	0	1
CDC25C	0	0	0	0	1
CDC27	0	0	0	0	1
CDC42EP4	1	0	0	1	0
CDH1	0	0	0	0	1
CDH13	1	0	0	1	1
CDH16	0	0	0	0	1
CDK1	0	0	0	0	1
CDK2	0	0	0	0	1
CDK4	0	0	0	0	1
CDK5RAP1	0	0	1	1	0
CDK5RAP3	0	0	1	1	0
CDK7	0	0	0	0	1
CDK9	1	0	0	1	0
CDKN1A	0	0	0	0	1
CDKN1B	0	0	0	0	1
CDKN2A	0	0	0	0	1
CEACAM21	0	0	0	0	1
CELSR1	0	1	0	1	0
CENPE	0	1	0	1	0
CEP72	1	1	0	1	0
CERCAM	0	1	0	1	0
CFH	1	0	0	1	0
CHAD	0	0	0	0	1
CHADL	1	0	0	1	0
CHEK1	0	0	0	0	1
CHEK2	0	0	0	0	1
CHMP2B	0	0	0	0	1
CHST10	1	0	0	1	0
CHSY1	0	1	0	1	0

CHTF8	1	1	0	1	0
CKAP2L	0	1	0	1	0
CLCN3	0	1	0	1	0
CLIP2	1	0	0	1	0
CLK1	0	0	1	1	0
CLPB	0	0	1	1	0
CMTM7	1	0	0	1	0
CNGA3	0	1	0	1	0
CNP	1	0	0	1	0
CNTNAP5	0	1	0	1	0
CNTROB	0	0	0	0	1
COASY	0	0	0	0	1
COG1	0	1	0	1	0
COL11A2	0	1	0	1	0
COL14A1	0	1	0	1	0
COL18A1	0	0	0	0	1
COL2A1	1	0	0	1	1
COL4A2	1	1	0	1	0
COL4A3BP	0	0	1	1	0
COL5A3	0	1	0	1	0
COMMD10	0	0	0	0	1
COMT	0	0	0	0	1
COPS6	0	0	0	0	1
CORO1B	0	1	0	1	0
CPNE4	1	0	0	1	0
CPSF3	0	0	1	1	0
CPSF4	0	0	1	1	0
CREB3L1	0	0	1	1	0
CRELD1	0	0	1	1	0
CRISP3	1	0	0	1	0
CRLS1	0	0	1	1	0
CROCCP2	1	0	0	1	0
CRTC2	0	0	1	1	0
CSDC2	1	0	0	1	0
CSMD3	0	0	0	0	1
CSPP1	0	0	0	0	1
CTBP1	0	1	0	1	0
CTBP2	1	1	0	1	1
CTNNA3	0	0	0	0	1
CTNNB1	0	0	0	0	1

CTSH	0	0	1	1	0
CTSS	0	0	0	0	1
CUBN	0	0	1	1	0
CUL4B	0	0	1	1	0
CUX2	1	1	0	1	0
CXCL12	1	0	0	1	0
CXCL14	0	0	0	0	1
CYB5R2	0	1	0	1	0
CYCS	1	0	0	1	0
CYP11A1	0	1	0	1	0
CYP17A1	0	0	0	0	1
CYP26A1	0	0	0	0	1
CYP2C8	0	1	0	1	0
CYP3A43	0	0	0	0	1
CYP4Z2P	0	0	1	1	0
D2HGDH	0	0	0	0	1
DAB2IP	0	0	0	0	1
DAPK3	1	0	0	1	0
DBIL5P	0	0	0	0	1
DCBLD2	1	0	0	1	0
DCLK2	1	0	0	1	0
DDB2	0	0	0	0	1
DDO	0	0	0	0	1
DEFA4	0	0	0	0	1
DEFB119	0	0	0	0	1
DENND4B	0	0	1	1	0
DEPDC7	0	1	0	1	0
DHX8	0	1	0	1	0
DIEXF	0	0	1	1	0
DISC1	0	1	0	1	0
DLGAP1	0	0	1	1	0
DLGAP5	0	0	0	0	1
DLX1	1	0	0	1	1
DLX2	0	0	0	0	1
DMRT3	0	0	0	0	1
DNAH11	0	1	0	1	0
DNAH2	0	0	0	0	1
DNAJC3	0	0	1	1	0
DNAJC5	1	0	0	1	0
DNAJC6	0	1	0	1	0

DNASE1L2	0	0	0	0	1
DOCK6	0	0	1	1	0
DOCK9	0	0	0	0	1
DPH3P1	1	0	0	1	0
DPPA3	0	1	0	1	0
DPYSL2	1	0	0	1	0
DSE	0	1	0	1	0
DTNBP1	0	0	1	1	0
DTX4	0	0	0	0	1
DUOX2	0	1	0	1	0
DUS1L	1	0	0	1	0
DUSP22	1	0	0	1	0
E2F1	0	0	0	0	1
E2F4	0	0	0	0	1
EARS2	1	1	0	1	0
EEPD1	1	0	0	1	0
EFCAB13	0	0	0	0	1
EFCAB6	0	0	0	0	1
EGFR	0	0	0	0	1
EHMT2	0	0	0	0	1
EIF2AK2	1	0	0	1	0
EIF2B2	0	0	1	1	0
EIF2C4	1	0	0	1	0
EIF2S1	0	0	0	0	1
EIF3I	1	1	0	1	0
EIF3M	1	0	0	1	0
ELAC2	0	0	0	0	1
ELF5	0	0	1	1	0
ELK3	0	0	1	1	0
ELK4	0	0	0	0	1
ELMO3	0	0	0	0	1
ELOVL6	1	0	0	1	0
EME1	0	0	1	1	0
EME2	0	1	0	1	0
EMSY	0	0	0	0	1
EN2	0	0	0	0	1
ENOX1	0	0	0	0	1
EP300	0	0	0	0	1
EPC1	1	1	0	1	0
EPHA3	1	0	0	1	0

EPHA4	1	0	0	1	0
EPHA8	0	0	0	0	1
EPHB2	0	0	0	0	1
EPHB4	0	0	0	0	1
EPHX2	0	0	0	0	1
EPRS	1	1	0	1	0
ERBB2IP	0	0	0	0	1
ERCC1	0	0	0	0	1
ERCC5	0	0	0	0	1
ERCC8	0	0	0	0	1
ERG	0	0	0	0	1
ERI2	1	0	0	1	0
ERICH1	0	0	0	0	1
ERMAP	0	0	1	1	0
ESCO1	0	0	0	0	1
ESPL1	0	0	1	1	0
ESPNL	0	1	0	1	0
ESR1	0	0	0	0	1
ESR2	0	0	0	0	1
ETV4	0	0	0	0	1
ETV6	0	1	0	1	0
EXD2	1	0	0	1	0
EXOSC3	0	0	1	1	0
FA2H	1	1	0	1	0
FAAH2	0	0	1	1	0
FAM109B	0	0	0	0	1
FAM111A	0	0	0	0	1
FAM114A1	0	1	0	1	0
FAM117A	0	0	0	0	1
FAM136A	0	1	0	1	0
FAM160B2	0	1	0	1	0
FAM171A2	1	0	0	1	0
FAM198B	0	1	0	1	0
FAM46B	0	1	0	1	0
FAM57A	0	0	0	0	1
FAM63B	0	1	0	1	0
FAM65B	0	0	1	1	0
FAM82A2	0	1	0	1	0
FAM83H	0	1	0	1	0
FAN1	0	0	0	0	1

FANCA	0	0	0	0	1
FANCB	0	0	0	0	1
FANCC	0	0	0	0	1
FANCD2	0	0	0	0	1
FANCE	0	0	0	0	1
FANCF	0	0	0	0	1
FANCG	0	0	0	0	1
FANCI	0	0	0	0	1
FANCL	0	0	0	0	1
FANCM	0	0	0	0	1
FASTK	1	1	0	1	0
FAT2	0	1	0	1	0
FAT3	0	0	1	1	0
FBN1	1	0	0	1	0
FCGRT	1	1	0	1	0
FCRL1	0	1	0	1	0
FCRL2	0	1	0	1	0
FEN1	0	0	0	0	1
FGF23	1	0	0	1	0
FGFR3	0	0	0	0	1
FHIT	0	0	0	0	1
FHOD1	0	0	0	0	1
FLII	0	0	0	0	1
FLJ35024	1	0	0	1	0
FLT1	0	0	0	0	1
FMN1	0	0	0	0	1
FOLR3	0	1	0	1	0
FOXA1	0	0	0	0	1
FOXB2	1	0	0	1	0
FOXO1	0	0	0	0	1
FOXP4	0	0	0	0	1
FPGT-TNNI3	1	0	0	1	0
FRAS1	0	1	0	1	0
FRAT1	1	0	0	1	0
FREM2	0	1	0	1	0
FTX	1	0	0	1	0
FURIN	0	1	0	1	0
FUT4	1	1	0	1	0
FXYD3	0	0	1	1	0
GAL3ST2	0	1	0	1	0

GAS2L1	1	1	0	1	0
GATAD2B	0	0	1	1	0
GBAS	0	0	1	1	0
GDF2	0	0	1	1	0
GDF6	1	0	0	1	0
GDF7	0	0	0	0	1
GEMIN2	0	0	0	0	1
GEMIN4	0	0	0	0	1
GEN1	0	0	0	0	1
GGA3	0	1	0	1	0
GGCX	0	0	0	0	1
GGN	0	1	0	1	0
GIMAP2	1	0	0	1	0
GIMAP8	0	1	0	1	0
GJB1	0	0	0	0	1
GJB4	0	1	0	1	0
GJD3	1	0	0	1	0
GKAP1	1	0	0	1	0
GKN1	1	0	0	1	0
GMCL1	1	0	0	1	0
GNE	0	1	0	1	0
GNL2	0	0	1	1	0
GOSR2	1	0	0	1	0
GPATCH2	0	0	0	0	1
GPATCH8	0	1	0	1	0
GPC5	0	1	0	1	0
GPC6	0	1	0	1	0
GPR110	0	0	1	1	0
GPR119	1	0	0	1	0
GPR143	0	0	0	0	1
GPR20	0	0	1	1	0
GPR98	0	1	0	1	0
GRM2	0	1	0	1	0
GSDMC	0	0	0	0	1
GSTM1	0	0	0	0	1
GTF2H5	0	1	0	1	0
GTSF1	0	0	0	0	1
GZMM	0	0	0	0	1
H2AFX	0	0	0	0	1
H3F3A	1	0	0	1	0

H3F3AP4	1	0	0	1	0
HAPLN4	1	0	0	1	0
HCAR3	0	1	0	1	0
HCG27	0	0	0	0	1
HCG4	0	0	0	0	1
HDAC4	0	0	0	0	1
HDAC6	0	0	0	0	1
HEATR5A	0	1	0	1	0
HERC2	0	0	1	1	0
HES7	1	0	0	1	0
HIRA	0	0	0	0	1
HIRIP3	1	1	0	1	0
HIST1H1A	0	0	0	0	1
HIST1H2AB	0	0	0	0	1
HIST1H3F	1	0	0	1	0
HMGXB4	1	0	0	1	0
HNF1A	0	0	0	0	1
HNF1B	0	0	0	0	1
HOTTIP	0	0	0	0	1
HOXA11	1	1	0	1	0
HOXA13	0	0	0	0	1
HOXA7	0	0	0	0	1
HOXA9	0	0	0	0	1
HOXB13	1	1	1	1	1
HOXB3	0	0	1	1	1
HOXB9	0	0	0	0	1
HOXB-AS1	0	0	1	1	0
HOXC10	0	0	0	0	1
HOXC6	0	0	0	0	1
HOXD13	0	0	0	0	1
HRH1	1	0	0	1	0
HRSP12	0	0	0	0	1
HSD17B13	0	0	0	0	1
HUS1	0	0	0	0	1
ICMT	1	0	0	1	0
IFIT5	0	0	1	1	0
IFRD2	0	1	0	1	0
IGF2	0	0	0	0	1
IGF2AS	0	0	0	0	1
IGF2BP1	0	0	1	1	0

IGFBP2	0	0	1	1	0
IGFL3	1	1	0	1	0
IGFN1	0	1	0	1	0
IGHMBP2	0	0	0	0	1
IL1R2	0	1	0	1	0
IL1RL2	0	0	0	0	1
IL4	0	0	0	0	1
ILK	0	0	0	0	1
IMPDH2	1	0	0	1	0
ING1	1	0	0	1	0
INHBB	0	0	0	0	1
INHBC	1	0	0	1	0
INHBE	0	0	0	0	1
IRS1	0	0	0	0	1
IRX4	0	0	0	0	1
ISYNA1	0	1	0	1	0
ITGA2	0	0	0	0	1
ITGA5	0	0	0	0	1
ITGA6	0	0	0	0	1
ITGAM	0	0	0	0	1
ITGAX	0	0	0	0	1
ITGB1BP1	1	0	0	1	0
ITPR1	0	1	0	1	0
ITPR2	0	0	0	0	1
JAK2	0	0	1	1	0
JAZF1	0	0	0	0	1
JMJD1C	0	1	0	1	0
JPX	1	0	0	1	0
KALRN	0	1	0	1	0
KAT7	0	0	0	0	1
KAZALD1	1	0	0	1	0
KBTBD11	0	0	0	0	1
KCNIP3	0	0	1	1	0
KCNK1	0	0	1	1	0
KCNK13	0	1	0	1	0
KCNMB1	1	0	0	1	0
KCNMB3	0	0	1	1	0
KCNQ1	0	0	1	1	0
KDM5B	0	0	0	0	1
KDM6B	0	0	0	0	1

KHDRBS2	0	0	0	0	1
KIAA0087	1	0	0	1	0
KIAA0100	0	1	0	1	0
KIAA0284	0	0	1	1	0
KIAA0494	1	0	0	1	0
KIAA0513	0	0	1	1	0
KIAA0664	1	1	0	1	0
KIAA1257	0	1	0	1	0
KIAA1683	0	1	0	1	0
KIAA1755	0	0	0	0	1
KIAA2018	1	0	1	1	0
KIF2B	0	0	0	0	1
KIF3C	1	0	0	1	0
KIFC2	0	1	0	1	0
KIR2DL3	0	0	0	0	1
KITLG	1	1	0	1	0
KLHDC8B	0	0	1	1	0
KLHL22	1	0	0	1	0
KLHL32	0	1	0	1	0
KLK1	0	0	1	1	0
KLK3	0	0	1	1	1
KMT2D	0	0	0	0	1
KRAS	0	0	0	0	1
KRI1	0	1	0	1	0
KRT18	0	0	0	0	1
KRT5	0	0	0	0	1
KRT6A	1	0	0	1	0
KRT6B	0	0	0	0	1
KRT76	0	0	0	0	1
KRT78	0	0	0	0	1
KRT8	0	0	0	0	1
KSP37	0	0	0	0	1
L1TD1	0	0	1	1	0
LAMA3	0	0	0	0	1
LAMC2	1	1	0	1	0
LAPTM4A	1	0	0	1	0
LARP7	1	1	0	1	0
LCAT	0	0	1	1	0
LCP2	0	0	1	1	0
LDHB	1	0	0	1	0

LDLRAD3	1	1	0	1	0
LEKR1	1	0	0	1	0
LEPR	0	0	0	0	1
LGALS13	0	0	1	1	0
LGI2	1	0	0	1	0
LGSN	0	0	0	0	1
LHFP	1	0	0	1	0
LHPP	1	0	0	1	0
LHX4	1	0	0	1	0
LIG1	0	0	0	0	1
LIG3	0	0	0	0	1
LIG4	0	0	0	0	1
LIME1	0	0	0	0	1
LINC00273	0	0	1	1	0
LIPH	1	0	0	1	0
LMTK2	0	0	0	0	1
LNP1	1	0	0	1	0
LNPEP	0	0	0	0	1
LOC1001285	0	0	1	1	0
LOC1001290	1	0	0	1	0
LOC1001301	1	0	0	1	0
LOC1001325	1	0	0	1	0
LOC1005054	0	0	0	0	1
LOC1005057	0	0	0	0	1
LOC1005067	1	1	0	1	0
LOC1005279	1	0	0	1	0
LOC284578	0	0	0	0	1
LOC284581	0	0	0	0	1
LOC286297	1	0	0	1	0
LOC389458	1	0	0	1	0
LOC440700	0	0	1	1	0
LOC442421	1	0	0	1	0
LOC606724	0	1	0	1	0
LOC93432	0	0	1	1	0
LONP1	0	1	0	1	0
LPCAT1	0	0	1	1	0
LPCAT2	0	0	0	0	1
LPCAT3	0	0	1	1	0
LPCAT4	0	0	1	1	0
LPGAT1	0	0	1	1	0

LPHN3	0	1	0	1	0
LPIN3	0	0	1	1	0
LRIG3	0	0	0	0	1
LRP1B	0	0	0	0	1
LRRC23	1	0	0	1	0
LRRC46	0	0	0	0	1
LRRC63	0	0	0	0	1
LRRC71	0	0	1	1	0
LSM12	1	0	0	1	0
LTBR	1	1	0	1	0
LTK	0	1	0	1	0
LY6G6C	0	0	1	1	0
LYRM4	1	0	0	1	0
MACROD1	0	0	1	1	0
MAK	0	0	0	0	1
MAP1LC3A	0	0	0	0	1
MAP2	0	1	0	1	0
MAP2K1	1	0	0	1	0
MAP2K6	0	0	0	0	1
MAP3K14	0	0	0	0	1
MARVELD3	0	0	0	0	1
MAZ	0	0	0	0	1
MBL2	0	0	0	0	1
MBOAT1	0	0	1	1	0
MBOAT2	0	0	1	1	0
MBTD1	1	0	0	1	0
MCM10	0	0	1	1	0
MDC1	0	0	0	0	1
MDH1	1	0	0	1	0
MDM2	0	0	0	0	1
MDM4	0	0	0	0	1
MED13L	1	0	0	1	0
MED28	1	0	0	1	0
MEF2A	0	0	1	1	0
MEI1	0	1	0	1	0
MESP1	0	0	0	0	1
METTL15	0	0	0	0	1
MICAL3	0	1	0	1	0
MIR205HG	1	0	0	1	0
MIR548N	0	1	0	1	0

MIR548Q	1	1	0	1	0
MIR658	1	1	0	1	0
MLH1	0	0	0	0	1
MLH3	0	0	0	0	1
MLL3	0	0	0	0	1
MLLT10	0	0	0	0	1
MLLT3	0	0	0	0	1
MLPH	0	0	0	0	1
MMP3	0	0	1	1	0
MMP7	0	0	0	0	1
MOGAT1	0	0	1	1	0
MOGS	0	1	0	1	0
MOK	0	1	0	1	0
MPZ	1	0	0	1	0
MRE11A	0	0	0	0	1
MRPL10	1	0	0	1	0
MRPS34	0	0	1	1	0
MRPS36	1	0	0	1	0
MRVI1	0	1	0	1	0
MSH2	0	0	0	0	1
MSH3	0	0	0	0	1
MSH4	0	0	0	0	1
MSH5	0	0	0	0	1
MSH6	0	0	0	0	1
MSMB	0	0	0	0	1
MSR1	0	1	0	1	1
MSS51	0	0	0	0	1
MTCH2	1	1	0	1	0
MTDH	0	1	0	1	0
MTHFD1L	0	1	0	1	0
MTOR	0	0	1	1	0
MTR	0	0	0	0	1
MTRR	0	0	0	0	1
MTUS2	0	0	0	0	1
MUC12	1	1	0	1	0
MUC16	1	1	0	1	0
MUC6	1	1	0	1	0
MUTYH	0	0	0	0	1
MXRA7	1	0	0	1	0
MYCBP2	0	1	0	1	0

MYEOV2	0	0	0	0	1
MYH4	0	1	0	1	0
MYLK	1	1	0	1	0
MYO10	0	0	0	0	1
MYO3B	0	0	0	0	1
MYO6	0	0	0	0	1
MYO7B	0	1	0	1	0
MYOF	0	0	0	0	1
MYOM1	0	0	0	0	1
NAALADL2	0	1	0	1	0
NAT8	0	1	0	1	0
NBL1	0	0	0	0	1
NBN	0	0	0	0	1
NCAPG	0	1	0	1	0
NCOA4	0	0	0	0	1
NCOR1	0	0	0	0	1
NCOR2	0	1	0	1	1
NDUFC2-KC	0	0	1	1	0
NDUFS4	1	0	0	1	0
NEDD4L	0	0	0	0	1
NEIL1	0	0	0	0	1
NEK10	0	1	0	1	0
NEK5	0	0	1	1	0
NEK8	0	1	0	1	0
NELL1	0	0	0	0	1
NEU2	1	0	0	1	0
NEUROD1	0	0	1	1	0
NGRN	1	0	0	1	0
NHP2L1	1	0	0	1	0
NKD2	0	1	0	1	0
NLK	1	1	0	1	0
NLRP9	1	1	0	1	0
NMB	0	0	1	1	0
NOL10	0	0	0	0	1
NOL3	0	0	0	0	1
NOL6	0	1	0	1	0
NOTCH4	0	0	0	0	1
NOXO1	0	1	0	1	0
NPAT	0	1	0	1	0
NPTN	1	0	0	1	0

NRG3	0	1	0	1	0
NRSN2	1	0	0	1	0
NSUN3	1	0	0	1	0
NT5C2	0	0	0	0	1
NT5C3L	1	0	0	1	0
NUCKS1	0	0	0	0	1
NUDCD1	1	0	0	1	0
NUDT11	0	0	0	0	1
NUDT7	0	0	0	0	1
NUP210	1	0	0	1	1
NUP93	1	0	0	1	0
NYNRIN	0	1	0	1	0
OCRL	1	0	0	1	0
OPCML	0	0	0	0	1
OR2L2	0	0	1	1	0
OR2L3	0	0	1	1	0
OR4B1	0	0	1	1	0
OR52K2	0	0	1	1	0
OR5H14	0	0	0	0	1
OR8U1	0	1	0	1	0
ORC4	0	1	0	1	0
ORC6	0	0	0	0	1
ORM1	1	0	0	1	0
OSBPL3	0	1	0	1	0
OXSR1	0	0	1	1	0
PABPC1	1	1	0	1	0
PABPC3	1	1	0	1	0
PACRG	0	0	0	0	1
PALB2	0	0	0	0	1
PANX3	0	1	0	1	0
PAPD7	0	0	0	0	1
PAQR4	0	1	0	1	0
PARP2	0	0	0	0	1
PARP3	0	0	0	0	1
PAX1	0	0	0	0	1
PCAT1	0	0	1	1	0
PCBP2	0	0	0	0	1
PCDHGA1	0	1	0	1	0
PCDHGA2	0	1	0	1	0
PCDHGA3	0	1	0	1	0

PCDHGA4	0	1	0	1	0
PCDHGA5	0	1	0	1	0
PCDHGA6	0	1	0	1	0
PCDHGA7	0	1	0	1	0
PCDHGA8	0	1	0	1	0
PCDHGB1	0	1	0	1	0
PCDHGB2	0	1	0	1	0
PCDHGB3	0	1	0	1	0
PCDHGB4	0	1	0	1	0
PCGF5	0	1	0	1	0
PCNA	0	0	0	0	1
PCSK4	0	1	0	1	0
PDCD11	0	1	0	1	0
PDE5A	0	1	0	1	0
PDE7A	1	0	0	1	0
PDGFRA	0	0	0	0	1
PDLIM5	0	0	0	0	1
PER1	0	0	0	0	1
PET112	0	0	1	1	0
PEX5L	1	0	0	1	0
PFKL	0	1	0	1	0
PGBD4	0	1	0	1	0
PGC	0	0	1	1	0
PGM3	1	1	0	1	0
PGR	0	0	0	0	1
PGS1	1	0	0	1	0
PHOSPHO1	1	0	0	1	0
PHTF2	0	0	1	1	0
PIGN	0	1	0	1	0
PIGW	0	1	0	1	0
PIGZ	0	1	0	1	0
PIK3CA	0	0	0	0	1
PIK3CG	0	1	0	1	0
PIM3	0	0	0	0	1
PKDREJ	0	1	0	1	0
PKP3	0	1	0	1	0
PLA2G12A	0	0	1	1	0
PLA2G16	0	0	1	1	0
PLA2G1B	0	0	1	1	0
PLA2G2A	0	0	1	1	0

PLA2G2D	1	0	0	1	0
PLA2G2E	0	0	1	1	0
PLA2G2F	0	0	1	1	0
PLA2G3	0	0	0	0	1
PLA2G4A	0	0	1	1	0
PLA2G4C	0	0	1	1	0
PLA2G4D	0	0	1	1	0
PLA2G4F	1	1	0	1	0
PLA2G5	0	0	1	1	0
PLA2G6	0	0	1	1	0
PLBD1	0	0	1	1	0
PLCB1	1	0	0	1	0
PLD3	0	1	0	1	0
PLEC	0	1	0	1	0
PLEKHG4	0	1	0	1	0
PLEKHM1	1	0	0	1	0
PLIN5	1	0	0	1	0
PLXNB1	0	0	0	0	1
PMS1	0	0	0	0	1
PNPLA8	0	0	1	1	0
PNRC1	0	0	0	0	1
POLG	0	0	0	0	1
POLI	0	0	0	0	1
POLM	1	1	0	1	0
POLN	0	0	0	0	1
POLQ	0	0	0	0	1
POLR2I	1	0	0	1	0
POLR3A	0	1	0	1	0
POLR3F	0	0	1	1	0
POMC	0	0	0	0	1
POU2F1	1	0	0	1	0
POU5F1B	1	0	1	1	0
PPAP2B	0	0	1	1	0
PPM1J	0	0	1	1	0
PPP1R13B	0	1	0	1	0
PPP1R14A	0	0	0	0	1
PPP1R3G	1	0	0	1	0
PPP2R1B	1	0	0	1	0
PPP2R5C	1	0	0	1	0
PPP6R2	0	0	0	0	1

PPT2	0	0	1	1	0
PPWD1	1	0	0	1	0
PRDM2	0	0	0	0	1
PRH1-PRR4	1	1	0	1	0
PRIM1	1	0	0	1	0
PRKACB	1	0	0	1	0
PRKCE	0	0	0	0	1
PRKCI	1	0	0	1	0
PRKCSH	1	0	0	1	0
PRKDC	0	0	0	0	1
PRKG2	0	0	1	1	0
PRMT3	0	0	1	1	0
PRMT7	0	0	0	0	1
PRR23C	1	1	0	1	0
PRRC2B	0	1	0	1	0
PRSS1	1	0	0	1	0
PRSS3	1	1	0	1	0
PRSS55	0	1	0	1	0
PSD4	0	0	1	1	0
PSEN2	0	0	1	1	0
PSMB11	0	1	0	1	0
PSMB7	1	0	0	1	0
PSMC1	1	0	0	1	0
PSMD5	0	0	0	0	1
PSORS1C1	0	0	0	0	1
PSORS1C3	0	0	0	0	1
PTCH2	0	0	0	0	1
PTCHD3	0	1	0	1	0
PTEN	0	0	0	0	1
PTK2B	1	1	0	1	0
PTN	0	0	0	0	1
PTPN7	0	1	0	1	0
PXDC1	1	0	0	1	0
PZP	0	1	0	1	0
QK1	0	0	0	0	1
RAB11FIP3	0	1	0	1	0
RAB17	0	0	0	0	1
RAB6A	1	0	0	1	0
RAB7L1	0	0	0	0	1
RAD17	0	0	0	0	1

RAD18	0	0	0	0	1
RAD23A	0	0	0	0	1
RAD23B	0	0	0	0	1
RAD50	0	0	0	0	1
RAD51	0	0	0	0	1
RAD51B	0	0	0	0	1
RAD51C	0	0	0	0	1
RAD51D	0	0	0	0	1
RAD52	0	0	0	0	1
RAD54B	0	0	0	0	1
RAD54L	0	0	0	0	1
RAD54L2	0	0	0	0	1
RAD9A	0	0	0	0	1
RAD9B	0	0	0	0	1
RAF1	0	0	0	0	1
RAG1	1	1	0	1	0
RAG2	1	1	0	1	0
RAPGEF4	1	1	0	1	0
RAPH1	0	1	0	1	0
RARG	0	0	0	0	1
RASSF7	0	0	0	0	1
RAVER2	0	0	1	1	0
RB1	0	0	0	0	1
RBBP6	0	1	1	1	0
RBBP8	0	1	0	1	0
RBM1	0	0	0	0	1
RCE1	1	0	0	1	0
RCOR2	0	0	1	1	0
RDH11	0	0	0	0	1
RECQL5	0	0	0	0	1
REPS2	0	0	1	1	0
RERE	0	0	1	1	0
RFC3	0	0	0	0	1
RGL3	0	0	1	1	0
RGS17	0	0	0	0	1
RHNO1	0	0	0	0	1
RHOBTB1	1	0	0	1	0
RIC8B	0	0	1	1	0
RIN1	0	0	0	0	1
RNASEH2B	0	0	1	1	0

RNASEL	0	0	0	0	1
RNF220	1	0	0	1	0
RNF41	0	0	0	0	1
RP1L1	0	1	0	1	0
RP9	1	0	0	1	0
RPAP3	1	1	0	1	0
RPS19BP1	1	0	0	1	0
RSBN1L	1	1	0	1	0
RTEL1	0	0	1	1	0
RTEL1-TNFF	0	0	1	1	0
RTKN	1	0	0	1	0
RTKN2	1	0	0	1	0
RTN4RL2	1	0	0	1	0
RUNX3	0	0	0	0	1
RUVBL1	0	0	0	0	1
RYR3	0	0	0	0	1
SAAL1	0	1	0	1	0
SAMD1	1	0	0	1	0
SAMD11	0	1	0	1	0
SBDS	0	0	0	0	1
SCAF1	1	0	0	1	0
SCGB1D1	1	0	0	1	0
SCGB3A1	0	0	0	0	1
SCN10A	0	1	0	1	0
SCN3B	1	0	0	1	0
SDC4	1	0	0	1	0
SDF2	1	0	0	1	0
SEC24C	0	1	0	1	0
SELT	1	0	0	1	0
SEMA3D	0	0	0	0	1
SEMA4C	0	1	0	1	0
SEMA6B	1	1	0	1	0
2-Sep	0	0	0	0	1
3-Sep	1	0	0	1	0
SERINC2	0	0	0	0	1
SERPINA6	1	0	0	1	0
SERPINA7	0	0	1	1	0
SESN1	0	0	0	0	1
SETMAR	0	1	0	1	0
SF1	0	0	0	0	1

SF3A3	1	1	0	1	0
SFXN2	1	0	0	1	1
SGSM2	0	1	0	1	0
SH2B1	1	1	0	1	0
SH3BP4	0	0	1	1	0
SH3TC2	0	1	0	1	0
SHANK3	1	0	0	1	0
SHROOM2	0	0	0	0	1
SIAE	1	0	0	1	0
SIK3	0	1	0	1	0
SKIV2L	0	1	0	1	0
SLC12A3	0	0	1	1	0
SLC15A1	0	1	0	1	0
SLC1A7	0	1	0	1	0
SLC22A2	0	0	1	1	1
SLC22A3	0	0	1	1	1
SLC23A1	0	0	1	1	0
SLC27A2	0	0	1	1	0
SLC2A5	1	1	0	1	0
SLC35G1	1	0	0	1	0
SLC36A1	1	1	0	1	0
SLC38A10	0	0	1	1	0
SLC39A7	1	0	0	1	0
SLC45A3	0	0	0	0	1
SLC48A1	1	0	0	1	0
SLC4A1	0	0	0	0	1
SLC5A4	0	1	0	1	0
SLC5A5	1	0	0	1	0
SLC6A6	0	0	1	1	0
SLC7A3	1	0	0	1	1
SLC9A10	0	1	0	1	0
SMARCD2	1	0	0	1	0
SMCR8	0	1	0	1	0
SMCY	0	0	0	0	1
SMPD1	1	0	0	1	0
SMPDL3A	1	1	0	1	0
SMUG1	0	0	0	0	1
SMURF2	0	0	1	1	0
SNAR-A3	0	1	0	1	0
SNORD62A	0	1	0	1	0

SNORD62B	0	1	0	1	0
SNORD97	1	0	0	1	0
SNX16	1	0	0	1	0
SNX9	0	1	0	1	0
SOCS2	0	0	0	0	1
SOLH	0	0	1	1	0
SORCS2	0	1	0	1	0
SOSTDC1	0	1	0	1	0
SOX17	0	0	0	0	1
SP2	0	0	0	0	1
SPATA13	0	1	0	1	0
SPATA45	0	0	1	1	0
SPG21	0	0	1	1	0
SPG7	1	0	0	1	0
SPHK2	0	0	0	0	1
SPINK7	1	0	0	1	0
SPINT2	0	0	0	0	1
SPOCK2	1	1	0	1	0
SPOP	0	0	0	0	1
SQRDL	0	0	1	1	0
SRCIN1	1	0	0	1	0
SRMS	0	1	0	1	0
SRSF7	1	0	0	1	0
STAB1	0	1	0	1	0
STAB2	0	1	0	1	0
STARD3	0	0	0	0	1
STAT4	0	1	0	1	0
STK11	0	0	0	0	1
STK11IP	0	1	0	1	0
STK36	0	0	0	0	1
STX3	1	0	0	1	0
STXBP2	1	0	0	1	0
STXBP6	1	0	0	1	0
SUGP1	0	0	1	1	0
SV2A	0	0	0	0	1
SVIL	0	0	0	0	1
SWSAP1	0	0	0	0	1
SYCP1	1	0	0	1	0
SYNGR1	0	0	0	0	1
SYNJ2BP	0	0	0	0	1

SYNM	0	1	0	1	0
TACC2	0	1	0	1	0
TACR3	1	0	0	1	0
TAF2	1	1	0	1	0
TANGO2	0	0	0	0	1
TAS2R19	1	1	0	1	0
TAS2R43	0	0	1	1	0
TBC1D2	0	0	0	0	1
TBC1D22A	1	1	0	1	0
TBCK	1	0	0	1	0
TBL2	1	0	0	1	0
TBRG1	1	0	0	1	0
TBX1	0	0	0	0	1
TBX22	0	0	1	1	0
TBX5	0	0	0	0	1
TCF19	1	0	0	1	0
TCOF1	0	1	0	1	0
TCONS_000	0	0	0	0	1
TCONS_000	0	0	0	0	1
TCONS_000	0	0	0	0	1
TCONS_000	0	0	0	0	1
TCONS_000	0	0	0	0	1
TDRD3	1	1	0	1	0
TEKT5	0	1	0	1	0
TERT	0	0	0	0	1
TET2	0	0	0	0	1
TFAM	1	0	0	1	0
TFAP2B	1	0	0	1	0
TFAP2C	0	0	1	1	0
TFAP2E	0	0	0	0	1
TFDP2	1	0	0	1	0
tgfb1	0	0	0	0	1
TGFBR1	1	0	0	1	0
TGOLN2	0	1	0	1	0
THADA	0	0	0	0	1
THAP3	1	1	0	1	0
THSD1	0	0	0	0	1
TIGD4	1	1	0	1	0
TIMP2	1	0	0	1	0
TIPARP	0	0	0	0	1

TLL2	1	1	0	1	0
TLN2	0	1	0	1	0
TMC1	1	1	0	1	0
TMEFF2	1	1	0	1	0
TMEM116	1	1	0	1	0
TMEM14B	0	1	0	1	0
TMEM180	0	0	0	0	1
TMEM63C	0	1	0	1	0
TMPRSS2	0	0	0	0	1
TNFRSF10C	0	1	0	1	0
TNK2	0	0	0	0	1
TNNI3K	1	0	0	1	0
TOE1	0	1	0	1	0
TOMM5	1	0	0	1	0
TP53	0	0	0	0	1
TP53BP1	0	0	0	0	1
TP53RK	1	0	0	1	0
TPH1	0	0	1	1	0
TPTE	0	0	1	1	0
TRAF4	1	0	0	1	0
TRAIP	1	0	0	1	0
TRIM17	0	0	1	1	0
TRIM38	0	1	0	1	0
TRIM54	0	0	1	1	0
TRIO	0	1	0	1	0
TRIOBP	0	0	0	0	1
TRPM1	0	0	1	1	0
TSPAN15	0	1	0	1	0
TSSC4	1	0	0	1	0
TTC21B	0	1	0	1	0
TTC8	0	0	1	1	0
TTC9C	1	0	0	1	0
TTLL10	0	1	0	1	0
TUBA1B	0	0	0	0	1
TUBA4A	0	0	1	1	0
TXNDC11	0	0	1	1	0
UBAP1	1	0	0	1	0
UBB	0	0	0	0	1
UBE3C	0	0	0	0	1
UBOX5	0	0	1	1	0

UCKL1	0	0	1	1	0
UCP3	1	0	0	1	0
UGT2B15	0	0	0	0	1
UGT2B17	0	0	0	0	1
ULK4	0	1	0	1	0
USP53	1	0	0	1	0
UTP18	0	0	0	0	1
VARS2	0	0	0	0	1
VCAM1	0	0	1	1	0
VCP	1	0	0	1	0
VEGFC	0	0	0	0	1
VEZF1	1	0	0	1	0
VGLL3	0	1	0	1	0
VLDLR	1	0	0	1	0
VPS39	1	0	0	1	0
VPS4A	1	0	0	1	0
VPS53	0	0	0	0	1
VRK3	1	0	0	1	0
VWA7	0	0	1	1	1
WBP11	1	0	0	1	0
WDR11	0	0	1	1	0
WDR52	0	0	0	0	1
WDR5B	1	1	0	1	0
WDR91	1	0	0	1	0
WNK1	0	0	0	0	1
WNK4	0	0	0	0	1
WNT3	0	0	0	0	1
WNT3A	1	1	0	1	0
WNT7B	1	0	0	1	0
WNT9B	0	0	0	0	1
WWOX	0	0	0	0	1
XPC	0	0	0	0	1
XPO7	1	0	0	1	0
XRCC1	0	0	0	0	1
XRCC2	0	0	0	0	1
XRCC3	0	0	0	0	1
XRCC4	0	0	0	0	1
XRCC5	0	0	0	0	1
XRCC6	0	0	0	0	1
YBX2	1	0	0	1	0

YEATS2	1	0	0	1	0
YIF1A	1	0	0	1	0
YIF1B	0	0	1	1	0
ZBP1	0	0	1	1	0
ZBTB32	0	0	1	1	0
ZCCHC14	0	0	1	1	0
ZFHX3	0	1	0	1	1
ZFP106	0	1	0	1	0
ZFY	0	0	0	0	1
ZFYVE27	1	1	0	1	0
ZGPAT	0	0	0	0	1
ZHX3	0	0	1	1	0
ZMIZ2	0	1	0	1	1
ZMYND19	1	0	0	1	0
ZNF142	1	1	0	1	0
ZNF177	0	1	0	1	0
ZNF197	1	1	0	1	0
ZNF222	0	1	0	1	0
ZNF232	0	0	1	1	0
ZNF318	0	0	0	0	1
ZNF385A	0	0	1	1	0
ZNF394	1	1	0	1	0
ZNF425	1	0	0	1	0
ZNF434	1	0	0	1	0
ZNF440	0	0	1	1	0
ZNF462	1	1	0	1	0
ZNF483	1	0	0	1	0
ZNF491	0	1	0	1	0
ZNF518B	0	1	0	1	0
ZNF519	0	1	0	1	0
ZNF550	0	1	0	1	0
ZNF559-ZNF	0	1	0	1	0
ZNF587	0	1	0	1	0
ZNF597	1	1	0	1	0
ZNF608	1	1	0	1	0
ZNF652	0	0	0	0	1
ZNF69	1	0	0	1	0
ZNF710	1	1	0	1	0
ZNF716	1	0	0	1	0
ZNF740	0	0	1	1	0

ZNF774	0	0	1	1	0
ZNF789	0	0	1	1	0
ZNF805	1	1	0	1	0
ZNF827	1	1	0	1	0
ZNFX1	0	0	1	1	0
ZSCAN1	1	0	0	1	0
ZSWIM3	0	0	1	1	0

1) Miscellaneous analyses included: pedigree linkage, gene-based recessive cosegregation, pathway analysis

Supplemental Table S6. Six genes that achieved statistical significance (p-value < 5e-5) based on gene-level analyses. Red font indicates p-value < 5e-5. The burden beta is the log odds ratio for burden score (sum of rare variants); 95%CI is the confidence interval on beta.

Gene	Cases vs Controls				Familial Cases vs Controls			
	SKAT-O	SKAT burden			SKAT-O	SKAT burden		
	p-value	p-value	beta	95%CI	p-value	p-value	beta	95%CI
ATM	1.73E-04	8.23E-05	0.62	0.31, 0.94	9.59E-05	***4.37E-05***	0.69	0.36, 1.03
BRCA2	1.22E-01	6.82E-02	0.27	-0.01 0.57	6.89E-01	5.75E-01	0.10	-0.23, 0.43
FAM111A	***4.00E-05***	2.55E-04	0.85	0.40, 1.34	4.02E-03	1.20E-02	0.63	0.14, 1.15
HOXB13	***5.43E-13***	***1.13E-13***	7.02	4.79, 10.19	***1.64E-13***	***2.34E-14***	7.08	4.85, 10.23
PABPC1	3.77E-02	1.05E-01	-0.10	-0.23, 0.05	5.43E-01	4.02E-01	-0.06	-0.21, 0.09
QK1	2.64E-02	2.78E-02	-0.12	-0.22, -0.01	6.80E-01	5.09E-01	1.21	-2.00, 5.67

	Aggressive Cases vs Controls				Aggressive Cases vs Non-Aggressive Cases			
	SKAT-O	SKAT burden			SKAT-O	SKAT burden		
	p-value	p-value	beta	95%CI	p-value	p-value	beta	95%CI
ATM	2.30E-02	1.28E-02	0.53	0.11, 0.95	1.49E-01	4.12E-01	-0.15	-0.53, 0.23
BRCA2	2.36E-03	1.40E-03	0.60	0.23, 0.96	***3.86E-05***	***2.42E-05***	0.84	0.46, 1.23
FAM111A	***3.28E-05***	2.09E-04	1.06	0.49, 1.65	2.45E-01	2.73E-01	0.29	-0.23, 0.82
HOXB13	***2.15E-10***	***4.13E-10***	7.16	4.69, 10.46	1.00E+00	9.56E-01	-0.03	-1.10, 1.02
PABPC1	***3.28E-08***	6.93E-02	-0.13	-0.32, 0.06	***2.45E-10***	6.19E-02	-0.19	-0.40 0.01
QK1	***3.10E-06***	***3.73E-06***	-0.40	-0.58, -0.23	***1.39E-07***	***1.50E-07***	-0.48	-0.67, -0.30

Supplemental Table S7. All cases and familial cases: gene-level analyses that detected significant associations, reduced to single variants with allele frequency at least 0.001 to perform single-variant analyses.

Gene	rsID	Allele		All Cases vs. Controls				Familial Cases vs. Controls			ALT Allele Frequencies		
		REF	ALT	OR	95% CI	p-value	PRACTICAL p-value (2)	OR	lower95	p-value	All Cases	Familial Cases	Controls
ATM	rs1800054	C	G	1.21	0.81, 1.81	0.36	0.32	1.36	0.89, 2.07	0.15	0.01234	0.01430	0.01053
	rs2234997	T	A	0.77	0.23, 2.55	0.66	0.93	0.66	0.17, 2.54	0.54	0.00120	0.00100	0.00132
	rs28904919	C	T	0.73	0.28, 1.94	0.53	0.67	0.49	0.15, 1.65	0.25	0.00171	0.00100	0.00211
	rs56128736	T	C	1.86	0.75, 4.58	0.18	0.44	1.94	0.76, 4.95	0.17	0.00326	0.00351	0.00184
	rs2227922	C	T	1.32	0.53, 3.28	0.54	0.24	1.56	0.62, 3.90	0.35	0.00223	0.00326	0.00184
	rs4986761	T	C	1.27	0.84, 1.92	0.25	0.03	1.32	0.86, 2.04	0.21	0.01217	0.01279	0.01027
	rs1800056	T	C	1.18	0.82, 1.68	0.37	1.34E-06	1.40	0.97, 2.03	0.07	0.01543	0.01806	0.01422
	rs1800057	C	G	1.10	0.85, 1.43	0.48	8.15E-09	1.24	0.94, 1.64	0.13	0.02914	0.03061	0.02580
	rs149711770	G	A	0.99	0.29, 3.31	0.98	NA	1.22	0.36, 4.19	0.75	0.00120	0.00151	0.00132
	rs1800058	C	T	1.14	0.82, 1.57	0.44	0.10	1.03	0.73, 1.48	0.85	0.01885	0.01706	0.01659
	rs138327406	T	G	2.10	0.58, 7.62	0.26	0.53	2.70	0.72, 10.07	0.14	0.00154	0.00226	0.00079
	rs34640941	A	G	1.39	0.42, 4.62	0.59	0.94	1.38	0.37, 5.05	0.63	0.00171	0.00151	0.00105
	rs1800059	A	C	1.19	0.46, 3.09	0.72	0.85	1.19	0.43, 3.27	0.74	0.00223	0.00226	0.00184
	rs1801673	A	T	1.66	0.99, 2.79	0.06	3.23E-03	1.70	1.00, 2.89	0.05	0.00977	0.01129	0.00553
	rs11212587	G	A	1.11	0.53, 2.31	0.78	0.02	1.39	0.65, 2.99	0.40	0.00343	0.00401	0.00342
	rs56009889	C	T	NA	NA	NA	0.24	NA	NA	NA	0.00189	0.00226	0.00000
BRCA2	rs4987046	A	G	1.47	0.62, 3.49	0.38	0.94	1.68	0.69, 4.10	0.25	0.00309	0.00351	0.00211
	rs766173	A	C	0.89	0.70, 1.13	0.32	0.56	0.91	0.71, 1.18	0.48	0.03292	0.03362	0.03528
	rs28897706	C	A	1.83	0.45, 7.46	0.40	0.29	2.06	0.47, 9.08	0.34	0.00137	0.00126	0.00079
	rs41293475	C	T	1.27	0.39, 4.06	0.69	0.05	0.91	0.24, 3.47	0.89	0.00137	0.00100	0.00132
	rs28897710	A	G	1.37	0.56, 3.33	0.49	0.29	0.84	0.29, 2.38	0.74	0.00257	0.00176	0.00211
	rs28897716	G	A	1.36	0.37, 5.02	0.65	0.22	NA	0.15, 3.79	NA	0.00120	0.00075	0.00105
	rs1799944	A	G	0.90	0.71, 1.14	0.36	0.64	0.92	0.72, 1.19	0.54	0.03310	0.03387	0.03528
	rs28897727	G	T	0.65	0.40, 1.06	0.08	0.95	0.70	0.42, 1.19	0.19	0.00567	0.00627	0.00948
	rs4987117	C	T	0.99	0.76, 1.30	0.96	0.80	1.03	0.77, 1.36	0.86	0.02674	0.02760	0.02712
	rs1799954	C	T	0.98	0.54, 1.79	0.94	0.17	1.19	0.64, 2.23	0.58	0.00549	0.00602	0.00500
	rs28897747	G	T	0.61	0.20, 1.91	0.40	0.14	0.69	0.21, 2.26	0.54	0.00103	0.00125	0.00184
	rs28897749	G	A	1.56	0.81, 3.00	0.19	0.63	1.59	0.80, 3.15	0.19	0.00531	0.00577	0.00369
	rs11571747	A	C	1.23	0.42, 3.63	0.70	NA	1.52	0.52, 4.49	0.45	0.00223	0.00276	0.00132
	rs28897754	G	T	1.77	0.45, 6.96	0.42	0.55	1.52	0.35, 6.68	0.58	0.00137	0.00125	0.00079
	rs11571769	G	A	0.78	0.45, 1.35	0.37	0.30	0.76	0.42, 1.37	0.36	0.00497	0.00552	0.00685
	rs11571833	A	T	0.99	0.61, 1.59	0.96	0.60	0.80	0.47, 1.37	0.42	0.00874	0.00702	0.00790
	rs1801426	A	G	0.71	0.25, 2.03	0.52	0.14	0.62	0.19, 2.04	0.43	0.00154	0.00125	0.00184
FAM111A	rs533676902	G	GGCAGATACTT	2.91	1.71, 4.95	8.10E-05	1.21E-10	2.27	1.29, 4.00	4.71E-03	0.01251	0.01054	0.00474
	rs116918730	C	G	1.17	0.96, 1.42	0.12	2.19E-04	1.14	0.93, 1.41	0.21	NA	NA	0.04739
	rs117988338	A	G	2.07	0.72, 5.93	0.17	0.22	1.42	0.42, 4.81	0.57	0.00257	0.00151	0.00132
HOXB13	rs138213197	C	T	21.01	6.58, 67.15	2.78E-07	9.17E-63	22.44	6.99, 71.97	1.69E-07	0.01354	0.01606	0.00079
PABPC1	rs79986761	G	A	0.76	0.40, 1.41	0.38	NA	0.91	0.47, 1.78	0.79	0.00429	0.00452	0.00500
	rs749280353	G	A	1.03	0.32, 3.31	0.96	NA	1.12	0.33, 3.83	0.86	0.00137	0.00151	0.00132
	rs112966887	CT	C	2.02	1.01, 4.04	0.05	NA	0.74	0.30, 1.83	0.52	0.00754	0.00251	0.00290
	NA	C	CAACCTTTGCACTAGTG	1.75	0.53, 5.74	0.36	NA	0.52	0.09, 2.93	0.45	0.00171	0.00050	0.00105
	rs200538577	G	A	0.70	0.40, 1.22	0.21	NA	0.76	0.39, 1.46	0.40	0.00720	0.00477	0.00579
	rs201157005	C	T	0.65	0.39, 1.10	0.11	NA	0.68	0.36, 1.26	0.22	0.00788	0.00502	0.00685
	rs200479111	T	C	0.39	0.10, 1.55	0.18	NA	1.20	0.27, 5.35	0.81	0.00103	0.00100	0.00105
	rs201575415	C	T	0.61	0.35, 1.06	0.08	NA	0.57	0.29, 1.13	0.11	0.00651	0.00401	0.00632
	rs77943211	C	T	0.62	0.34, 1.10	0.10	0.73	0.57	0.28, 1.16	0.12	0.00566	0.00351	0.00579
	rs545344384	T	TC	0.61	0.35, 1.04	0.07	0.85	1.00	0.57, 1.77	1.00	0.00634	0.00702	0.00685
	rs113614781	G	C	0.59	0.33, 1.08	0.09	NA	0.99	0.53, 1.86	0.98	0.00514	0.00577	0.00553
	rs72681442	G	A	0.56	0.28, 1.14	0.11	NA	1.00	0.47, 2.11	1.00	0.00343	0.00401	0.00395
	rs72681443	T	C	0.54	0.26, 1.12	0.10	NA	0.95	0.43, 2.08	0.89	0.00309	0.00351	0.00369
	rs76261471	A	C	0.33	0.11, 1.01	0.05	NA	0.68	0.21, 2.18	0.52	0.00103	0.00125	0.00211
	rs771446357	C	T	0.45	0.14, 1.47	0.19	NA	0.80	0.23, 2.71	0.72	0.00103	0.00125	0.00158
QKI	rs1409696271	A	T	0.80	0.66, 0.98	0.04	NA	0.99	0.80, 1.22	0.93	0.04165	0.05469	0.05477
	rs1300041642	G	T	0.80	0.66, 0.9	0.03	NA	0.97	0.79, 1.19	0.76	0.04268	0.05544	0.05687

1) Genome Reference Consortium Human Build 38

2) PRACTICAL p-value and allele frequencies (among cases and controls of European descent) provided by Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium, based on genome-wide analyses of more than 140,000 men (Schumacher et al., Nature Genetics, 2018).

NA means not available because allele frequencies were too small to perform statistical comparisons, or PRACTICAL results were not available for praticular variants

Supplemental Table S8. Aggressive cases: gene-level analyses that detected significant associations, reduced to single variants with allele frequency at least 0.001 to perform single-variant analyses.

Gene	rsID	Allele		Aggressive Cases vs. Controls			Aggressive Cases vs. Non-aggressive Cases			ALT Allele Frequencies			
		REF	ALT	OR	lower95	p-value	OR	95% CI	p-value	All Cases	Familial Cases	Aggressive Cases	Controls
ATM	rs1800054	C	G	0.89	0.49, 1.64	0.71	0.61	0.34, 1.09	0.10	0.01234	0.01430	0.00874	0.01053
	rs2234997	T	A	1.00	0.22, 4.62	1.00	1.88	0.34, 10.35	0.47	0.00120	0.00100	0.00159	0.00132
	rs28904919	C	T	1.08	0.31, 3.71	0.90	2.25	0.53, 9.64	0.27	0.00171	0.00100	0.00239	0.00211
	rs56128736	T	C	2.25	0.71, 7.18	0.17	0.74	0.25, 2.18	0.58	0.00326	0.00351	0.00318	0.00184
	rs2227922	C	T	0.54	0.07, 4.53	0.57	0.30	0.04, 2.32	0.25	0.00223	0.00326	0.00040	0.00184
	rs4986761	T	C	1.22	0.69, 2.16	0.49	0.96	0.55, 1.66	0.87	0.01217	0.01279	0.01113	0.01027
	rs1800056	T	C	0.88	0.53, 1.48	0.64	0.57	0.34, 0.96	0.04	0.01543	0.01806	0.01113	0.01422
	rs1800057	C	G	0.91	0.63, 1.30	0.59	0.80	0.55, 1.17	0.25	0.02914	0.03061	0.02663	0.02580
	rs149711770	G	A	0.55	0.08, 3.89	0.55	0.75	0.09, 6.46	0.79	0.00120	0.00151	0.00079	0.00132
	rs1800058	C	T	1.28	0.84, 1.96	0.26	1.21	0.78, 1.89	0.39	0.01885	0.01706	0.02186	0.01659
	rs138327406	T	G	NA	NA	NA	NA	NA	NA	0.00154	0.00226	0.00000	0.00079
	rs34640941	A	G	2.12	0.50, 8.95	0.30	2.92	0.67, 12.75	0.15	0.00171	0.00151	0.00238	0.00105
	rs1800059	A	C	1.68	0.55, 5.10	0.36	1.46	0.43, 4.93	0.55	0.00223	0.00226	0.00278	0.00184
	rs1801673	A	T	1.32	0.60, 2.92	0.49	0.53	0.27, 1.03	0.06	0.00977	0.01129	0.00636	0.00553
	rs11212587	G	A	0.57	0.19, 1.74	0.33	0.47	0.15, 1.46	0.19	0.00343	0.00401	0.00238	0.00342
	rs56009889	C	T	NA	NA	NA	0.93	0.24, 3.69	0.92	0.00189	0.00226	0.00159	0.00000
BRCA2	rs4987046	A	G	0.78	0.22, 2.79	0.70	0.67	0.21, 2.07	0.48	0.00309	0.00351	0.00199	0.00211
	rs766173	A	C	0.85	0.61, 1.17	0.31	1.03	0.73, 1.46	0.85	0.03292	0.03362	0.03262	0.03528
	rs28897706	C	A	1.87	0.31, 11.35	0.49	0.98	0.19, 5.13	0.98	0.00137	0.00126	0.00159	0.00079
	rs41293475	C	T	1.17	0.26, 5.28	0.84	1.76	0.36, 8.63	0.49	0.00137	0.00100	0.00159	0.00132
	rs28897710	A	G	1.45	0.48, 4.35	0.51	1.65	0.47, 5.74	0.43	0.00257	0.00176	0.00318	0.00211
	rs28897716	G	A	NA	0.49, 11.20	NA	NA	0.28, 7.62	NA	0.00120	0.00075	0.00159	0.00105
	rs1799944	A	G	0.85	0.61, 1.17	0.31	1.02	0.73, 1.44	0.90	0.03310	0.03387	0.03264	0.03528
	rs28897727	G	T	0.43	0.20, 0.94	0.03	0.57	0.25, 1.34	0.20	0.00567	0.00627	0.00399	0.00948
	rs4987117	C	T	1.07	0.76, 1.52	0.68	1.29	0.89, 1.86	0.18	0.02674	0.02760	0.02901	0.02712
	rs1799954	C	T	0.85	0.38, 1.89	0.68	0.97	0.41, 2.32	0.95	0.00549	0.00602	0.00556	0.00500
	rs28897747	G	T	1.06	0.24, 4.62	0.94	2.15	0.39, 11.84	0.38	0.00103	0.00125	0.00119	0.00184
	rs28897749	G	A	1.16	0.47, 2.86	0.74	0.82	0.35, 1.91	0.64	0.00531	0.00577	0.00397	0.00369
	rs11571747	A	C	0.89	0.17, 4.75	0.89	0.91	0.25, 3.25	0.88	0.00223	0.00276	0.00159	0.00132
	rs28897754	G	T	2.44	0.48, 12.56	0.29	1.52	0.34, 6.80	0.58	0.00137	0.00125	0.00159	0.00079
	rs11571769	G	A	0.92	0.42, 2.01	0.83	1.39	0.58, 3.33	0.46	0.00497	0.00552	0.00477	0.00685
	rs11571833	A	T	1.09	0.59, 2.01	0.78	1.34	0.69, 2.60	0.39	0.00874	0.00702	0.01033	0.00790
	rs1801426	A	G	0.57	0.13, 2.54	0.46	0.86	0.14, 5.15	0.87	0.00154	0.00125	0.00159	0.00184
FAM111A	rs533676902	G	GGCAGATACTT	3.62	1.93, 6.81	6.44E-05	1.45	0.83, 2.53	0.19	0.01251	0.01054	0.01510	0.00474
	rs116918730	C	G	1.14	0.88, 1.48	0.34	0.90	0.69, 1.19	0.47	NA	NA	0.05048	0.04739
	rs117988338	A	G	3.67	1.12, 12.01	0.03	3.78	1.12, 12.78	0.03	0.00257	0.00151	0.00397	0.00132
HOXB13	rs138213197	C	T	23.36	6.75, 80.87	6.56E-07	0.94	0.56, 1.58	0.81	0.01354	0.01606	0.01192	0.00079
PABPC1	rs79986761	G	A	0.61	0.25, 1.50	0.28	0.52	0.20, 1.37	0.19	0.00429	0.00452	0.00397	0.00500
	rs749280353	G	A	0.50	0.08, 3.34	0.48	0.25	0.04, 1.65	0.15	0.00137	0.00151	0.00079	0.00132
	rs112966887	CT	C	5.43	2.60, 11.35	7.00E-06	11.92	3.96, 35.85	1.03E-05	0.00754	0.00251	0.01590	0.00290
	NA	C	CAACCTTTGCACTAGTG	5.79	1.61, 20.74	0.01	NA	NA	NA	0.00171	0.00050	0.00397	0.00105
	rs200538577	G	A	0.57	0.29, 1.12	0.10	0.73	0.34, 1.58	0.43	0.00720	0.00477	0.00954	0.00579
	rs201157005	C	T	0.53	0.28, 1.02	0.06	0.75	0.35, 1.58	0.45	0.00788	0.00502	0.01073	0.00685
	rs200479111	T	C	0.06	0.01, 0.41	0.00	0.05	0.01, 0.52	0.01	0.00103	0.00100	0.00079	0.00105
	rs201575415	C	T	0.60	0.30, 1.21	0.15	0.67	0.30, 1.54	0.35	0.00651	0.00401	0.00914	0.00632
	rs77943211	C	T	0.65	0.32, 1.34	0.24	0.80	0.33, 1.91	0.61	0.00566	0.00351	0.00795	0.00579
	rs545344384	T	TC	0.17	0.07, 0.41	1.02E-04	0.11	0.04, 0.29	4.80E-06	0.00634	0.00702	0.00358	0.00685
	rs113614781	G	C	0.19	0.07, 0.49	7.52E-04	0.10	0.04, 0.29	1.59E-05	0.00514	0.00577	0.00278	0.00553
	rs72681442	G	A	0.14	0.04, 0.48	1.73E-03	0.11	0.03, 0.39	7.33E-04	0.00343	0.00401	0.00159	0.00395
	rs72681443	T	C	0.15	0.04, 0.50	2.27E-03	0.12	0.03, 0.47	2.28E-03	0.00309	0.00351	0.00159	0.00369
	rs76261471	A	C	0.06	0.01, 0.56	0.01	0.10	0.01, 0.98	0.05	0.00103	0.00125	0.00040	0.00211
	rs771446357	C	T	0.12	0.01, 1.17	0.07	0.10	0.01, 0.98	0.05	0.00103	0.00125	0.00040	0.00158
QKI	rs1409696271	A	T	0.46	0.33, 0.64	4.62E-06	0.41	0.29, 0.58	5.67E-07	0.04165	0.05469	0.02186	0.05477
	rs1300041642	G	T	0.48	0.35, 0.67	1.48E-05	0.43	0.31, 0.61	1.57E-06	0.04268	0.05544	0.02266	0.05687

1) Genome Reference Consortium Human Build 38

NA means not available because allele frequencies were too small to perform statistical comparisons, or PRACTICAL results were not available for praticular variants

Supplemental Table S9. Genes discovered in this study that are previously associated with PCa.

ATM	We detected associations of *ATM* with PCa by gene-level analyses, but not single-variant analyses, suggesting that an accumulation of rare alleles with relatively small effects (e.g., OR 1.1-2.1) is responsible for this observation. The *ATM* gene encodes a protein that regulates tumor suppressor proteins p53 and BRCA1, as well as checkpoint proteins CHK2, RAD17 and RAD9, and the DNA repair protein, NBS1. Mutations in the *ATM* gene have previously been associated with PCa onset and lethal disease[19,58]. Heterozygous carriers of *ATM* mutations have been reported to be at increased risk of breast cancer, and possibly stomach and colon cancer [59]. More general evidence suggests that the risk of PCa due to also *ATM* variants is similar to that of other DNA repair genes [15,20-24].
BRCA2	Similar to ATM, we detected associations of *BRCA2* with PCa by gene-level analyses but not single-variant analyses, perhaps due to an accumulation of small effects of rare alleles. *BRCA2* is also within the DNA repair pathways. *BRCA2* mutation carriers are reported to have a 2.6-fold increased risk for PCa compared to non-carriers, and BRCA2 mutations have been associated with higher PCa mortality [13,14,17,60].
EMSY	We detected a variant (rs200165356) associated with a reduced risk of aggressive PCa. Others have identified variants of *EMSY* that are associated with aggressive PCa [61,62]. The protein encoded by EMSY can silence *BRCA2* , and localizes to sites of repair following DNA damage, suggesting that *EMSY* and *BRCA2* interact in DNA repair processes, and this interaction is thought to lead to risk of sporadic breast and ovarian cancers [63].
FAM111A	Several variants found within the *FAM111A* gene were more frequent among cases than controls, and one (rs533676902) was also highly significant in the PRACTICAL study. Further, a common non-coding variant (rs1938781) near *FAM111A* was also associated with PCa in a Japanese population [64]. *FAM111A* has been identified as a factor involved in DNA replication and repair[65], and reduced expression of this gene has been associated with DNA replication defects[66].
HNF1B	The alternate allele of the variant rs3216929 within in the *HNF1B* gene was observed significantly less often among PCa cases than controls. The same variant has also been significantly associated with PCa risk within the PRACTICAL study. A recent meta-analysis demonstrated that multiple common variants in the gene *HNF1B* are associated with risk of PCa and endometrial cancers [67]. It would also be interesting to understand whether this gene links some of the observed associations between diabetes and PCa [69], or whether this gene plays a role in the reduced risk of PCa for men receiving metformin for diabetes [70].
KLK3	The *KLK3* gene is one of several kallikreins and encodes the serum protein prostate-specific antigen (PSA). Common variants in this gene, rs2735839 and rs17632542, have previously been associated with aggressive PC [70 71]. In our study we observed an association between rs17632542 and risk of familial PCa, but not with aggressive disease.
MSMB	The *MSMB* gene encodes a protein in the epithelial cells of the prostate gland and it is secreted into the seminal plasma. Expression of the encoded protein is decreased in PCa, [72] and common gene variants, including rs10993994 which was

	identified in our study, have previously been associated with PCa risk [73]. While variants in *MCMB* have previously been associated with PCa risk in the general population, our study is the first to show that variants in *MSMB* are also significantly associated with familial disease.
PCAT1	The *PCAT1* gene, located at 8q24.21, produces a long non-coding RNA that promotes cell proliferation and is up-regulated in prostate and colorectal cancers. It negatively regulates the BRCA2 tumor suppressor protein and positively regulates the Myc oncoprotein [74]. We found multiple variants in high linkage disequilibrium that are significantly associated with familial disease, and the same variants are also highly associated in the PRACTICAL dataset. Further evidence that *PCAT1* may play a role in PCa risk comes from a recent study demonstrating *PCAT1* overexpression promotes cell proliferation and inhibits apoptosis, and that a variant in this gene (rs1902432) is associated with PCa risk in a Chinese case-control cohort [75].
PRSS3	The *PRSS3* gene encodes a trypsinogen, which is expressed in the brain and pancreas. The expression of this gene has been associated with poor prognosis of pancreatic cancer[76] and ovarian cancer [77], and its expression is upregulated in metastatic PCa [78]. Interestingly, with respect to this latter finding, our study observed an association between the *PRSS3* variant rs764430438 and aggressive disease.
TERT	The *TERT* gene encodes one of the two proteins that makeup telomerase. The relevant locus on 5p15.33 has been previously associated with PCa, with strong evidence that variants in the *TERT* gene are associated with disease risk [79,80]. Our study identified an association between PCa risk and a variant upstream of *TERT*, rs7712562, which was strongest for cases with familial disease. This variant was also highly associated with PCa risk in the PRACTICAL dataset, providing compelling evidence for a role in PCa risk.

Supplemental Table 10. Features of genes not previously reported to be associated with PCa.

FAM114A1	We detected two variants that were significantly less frequent among aggressive PCa cases compared with controls. Little is known about this gene, although some evidences suggests that it encodes a protein that plays a role in neuronal cell development [81].
MUC6	We identified a variant in the *MUC6* gene that is associated with aggressive PCa. The gene encodes a protein produced by epithelial cells that provide a protective mucus layer. Mucin genes have characteristics that make them less likely to be critical for disease development, and caution must be taken not to over-interpret this result. [82]
MYCBP2	Two variants in the *MYCBP2* gene were associated with aggressive PCa. This gene encodes a protein involved in axon guidance and synapse formation in the developing nervous system.
PABPC1	The *PABPC1* gene encodes a protein that binds to poly(A), which promotes ribosome recruitment and translation initiation. We observed that the alternative allele of rs112966887 occurred much more frequently among aggressive PCa cases than controls. The protein encoded by *PABPC1* has been reported to induce carcinogenesis in gastric cancer [83], as well as local tumor progression and shorter survival for patients with esophageal cancer [84].
QKI	Two variants in the *QKI* gene had alternate alleles that were less frequent among aggressive PCa cases than among controls. *QKI* encodes an RNA-binding protein that regulates pre-mRNA splicing, protein translation, and mRNA stability. Frameshift mutations in this gene have been associated with gastric and colorectal cancers [85].
RAPGEF4	Three variants in the *RAPGEF4* gene were observed to have alternate alleles much less frequent among aggressive PCa compared with than controls. This gene encodes a protein, commonly referred to as EPAC2, which is expressed predominantly in brain, neuroendocrine, and endocrine tissues [86].
RNASEH2B	We found that the variant rs1172291060 within the *RNASEH2B* gene had an alternate allele that was much less frequent among aggressive PCa compared with controls. *RNASEH2B* appears to be a metastasis susceptibility gene in breast cancer [87]. Furthermore, *RNASEH2*-deficient cells exhibit elevated levels of DNA damage, and *RNASEH2* deficiency is frequently found in prostate adenocarcinomas.[88]
THAP3	Two variants upstream of the *THAP3* gene were associated with a significant reduction in aggressive PCa risk in both analyses for aggressive disease. *THAP3* is related to nucleic acid binding [89], and indirect evidence suggests it might be associated with oligodendrogliomas [90].
ULK4	The alternate allele of one variant in the *ULK4* gene occurred much less frequently among PCa cases compared to controls, and less frequently among aggressive cases versus controls. This gene encodes a protein that may be associated with neurodevelopmental, neuropsychiatric, and neurodegenerative diseases[91], and is associated with multiple myeloma risk[92] and acute aortic dissection[93]. The GTEx portal shows this gene to be more expressed in testis than other tissues[56].

XPO7	Two variants in the *XPO7* gene were observed with alternate alleles occurring significantly less frequently among aggressive PCa cases compared with controls. *XPO7* mediates the nuclear export of proteins[94], and has been reported to be a prognostic biomarker for serous epithelial ovarian cancer [95].

References

1. Schaid, D.J. *et al.* Pooled genome linkage scan of aggressive prostate cancer: results from the International Consortium for Prostate Cancer Genetics. *Hum Genet* **120**, 471-85 (2006).
2. Olson, J.E. *et al.* The Mayo Clinic Biobank: a building block for individualized medicine. *Mayo Clin Proc* **88**, 952-62 (2013).
3. Asmann, Y.W. *et al.* TREAT: a bioinformatics tool for variant annotations and visualizations in targeted and exome sequencing data. *Bioinformatics* **28**, 277-8 (2012).
4. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
5. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
6. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-33 (2013).
7. Sun, L. & Dimitromanolakis, A. PREST-plus identifies pedigree errors and cryptic relatedness in the GAW18 sample using genome-wide SNP data. *BMC Proc* **8**, S23 (2014).
8. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
9. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).
10. De Baets, G. *et al.* SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res* **40**, D935-9 (2012).
11. Munz, M. *et al.* CSN and CAVA: variant annotation tools for rapid, robust next-generation sequencing analysis in the clinical setting. *Genome Med* **7**, 76 (2015).
12. Ioannidis, N.M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* **99**, 877-885 (2016).
13. Sim, N.L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* **40**, W452-7 (2012).
14. Ioannidis, N.M. *et al.* FIRE: functional inference of genetic variants that regulate gene expression. *Bioinformatics* **33**, 3895-3901 (2017).
15. Lee, S., Wu, M.C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762-75 (2012).