Genetics of early onset Parkinson's disease in Finland: exome sequencing and genome-wide association study

SUPPLEMENTARY MATERIAL:

Contents

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services [project Z01 AG000958-09]. The study was supported by grants from the Council for Health Research of the Academy of Finland [project 127764]; Sigrid Juselius Foundation; Asla-Fulbright Foundation; Orion-Farmos Foundation; The Finnish Parkinson Foundation; Medical Research Center Oulu Doctoral Programme; The Finnish Medical Foundation; Oulun Duodecim-seura; University of Oulu Graduate School; and Oulun Lääketieteellinen Tutkimussäätiö. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. We also thank NIH GWAS Data Repository for providing control samples to this study.

The FUSION study was conducted by the FUSION Investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The data and samples from the FUSION study reported here were supplied by the NIDDK Central Repositories. This manuscript was not prepared in collaboration with Investigators of the FUSION study and does not necessarily reflect the opinions or views of the FUSION study, the NIDDK Central Repositories, or the NIDDK.

The NFBC1966 Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, UCLA, University of Oulu, and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 Study and does not necessarily reflect the opinions or views of the NFBC1966 Study Investigators, Broad Institute, UCLA, University of Oulu, National Institute for Health and Welfare in Finland and the NHLBI.

Michael A. Nalls' participation is supported by a consulting contract between Kelly Services and the National Institute on Aging, NIH, Bethesda, MD, USA.

Subjects and methods

Patients and controls

In Finland, patients with PD are entitled to reimbursement of medicine expenses by the Social Insurance Institute (Kela). We have previously identified 1090 patients with EOPD nationwide in the Kela reimbursement database, and obtained a blood sample from 441 patients, who volunteered to participate in the study (Ylikotila et al., 2015). The study was approved by the Ethics Committee of the Turku University Hospital.

The STAMPEED dataset includes samples from the Northern Finland Birth Cohort 1966 (NFBC1966). In this cohort whole exome sequencing has been performed on 586 samples by using the Illumina platform. The FUSION dataset includes samples from the Finland-United States Investigation of NIDDM Genetics study. In this dataset there are 1161 patients with type 2 diabetes and 1174 subjects with normal glucose tolerance that were genotyped on HumanHap300v1.1 array. Permission to use these data was requested from dbGaP (phs000276.v2.p1 and phs000100.v4.p1) (Tryka et al., 2014).

Sample preparation and pre-processing

DNA was extracted from blood and DNA libraries were prepared using either the Nextera Rapid Capture Expanded Exome Kit (first set) or the TruSeq Exome Library Prep Kit, according to the manufacturer's protocol (Illumina, San Diego, CA, USA). Indexed and pooled libraries of 12 samples each were paired-end sequenced on three lanes using HiSeq2000 runs (Illumina, San Diego, CA, USA).

Raw data was processed to FastQ files using Illumina's BaseSpace utility. Files were prepared from Fastq files to BAM using Human genome assembly 19, Burrows-Wheeler aligner (BWA)(Li and Durbin, 2009) and to analysis-ready variants by following Genome Analysis Toolkit (GATK, ver. 3.1) best practices pipeline (DePristo et al., 2011, Van der Auwera et al., 2002). The 10 x depth of the contigs was at least 90 % and the 30 x depth of the contigs was an average of approximately 70 %. The variants were filtered by using 99.9% tranche. Variants that passed the filter were selected for subsequent analysis. Data was processed using Plink version 1.90 (Chang et al., 2015) after group variant sets were called and converted to binary format.

Genome-wide association study (GWAS)

Sample preparation and genotyping of the cases has been published elsewhere (Hernandez et al., 2012).

Quality control

Quality control exclusion of the variants was as follows: SNPs with missingness less than 95% by genotyping call rate, missingness by individual less than 90%, minor allele frequency (MAF) less than 5%, Hardy-Weinberg equilibrium (HWE) p value less than 1E-5. SNPs ambiguous to strand (A/T and G/C) were removed. Samples with wrong gender-information were removed. Missingness in cases compared to controls p value (from chi-squared test) <1E-5 and nonrandom missingness by haplotype (from chi-squared test) <1E-5. The case-control series then underwent the calculation of pairwise identity by descent, excluding any samples sharing greater than a 0.185 proportion of alleles

(Phi_hat > 0.185) to remove probands from related pairs, excluding all first, second and third degree relatives. Heterozygous outliers (deviate more than 3*SD from mean observed heterozygosity rate) were removed. Duplications were removed. Population outliers were removed. 403 cases and 1650 controls passed quality checks. Cases and controls were merged using Plink/Seq tool (https://atgu.mgh.harvard.edu/plinkseq/).

Imputation

Files were prepared for imputation as follows: To validate vcf files, HRC preparation checking Tool suggested by Sanger Institute's imputation service was used. Genomic inflation factor before imputation was 1.027. The inflation factor for an equivalent study of 1000 cases and 1000 controls is 1.041681. Before imputation there were 256,546 variants.

Imputation was done using Sanger Institute's imputation service with 1000 Genomes Phase 3 reference panel. Imputation server uses SHAPEIT2 (Delaneau et al., 2013) for pre-phasing variants and PBWT (Durbin, 2014) algorithm for imputation.

After imputation dataset variants with alternative allele count less than 1, imputation quality score (estimated haploid dosage r^2 from imputation) less than 0.5 and MAF less than 0.1% were filtered. Genomic inflation factor after imputation was 1.00 and quantile-quantile plot of p values did not reveal any systemic biases. After quality control 12,954,715 variants were left for the analysis.

Analysis

Single variant association (SVA) analysis was done using 'Efficient and parallelizable association container toolbox' (EPACTS) (http://genome.sph.umich.edu/wiki/EPACTS*) with* EMMAX (Kang et al., 2010) test. Variants with minor allele frequency (MAF) < 0.1% were excluded. We used *LocusZoom (Pruim et al., 2010)* to generate regional gene sets from the GWAS. A genome wide significant p value was set at p<5E-08.

Gene-based tests were carried out by using EPACTS with default settings. Kinship information from a variance component model was used to handle possible distant family relationships within the population. The EmmaxCMC (burden test) and sequence kernel association test (SKAT, performed with SKAT-O) were used. Gene-based analysis used variants with MAF less than 5% and with assumed functional significance (nonsynonymous, Essential Splice

Site, Normal Splice Site, Start Loss, Stop Loss, Stop Gain). The first 20 principal components (based on variance-standardized relationship matrix), age at sampling year, sex, and type 2 diabetes diagnosis in controls were used as covariates for both the SVA and gene-based tests.

Exome sequencing

Quality control

Supplementary tables S1 and S2 present the quality control steps made for WES discovery dataset. Before Quality control (QC) there was 238 cases and 563 controls with total of 411,343 variants. Insertion to deletion ratio was 1.02 for all, 0.87 for known (dbsnp 138 excluding sites after 129) and 1.08 for novel indels. Ti/Tv ratio was 2.26 for all, 2.62 for known (dbsnp 138 excluding sites after 129) and 2.09 for novel. QC exclusion of the variants was as follows: SNPs with missingness less than 95% by genotyping call rate, missingness by individual less than 90%, minor allele frequency (MAF) less than 10%, Hardy-Weinberg equilibrium (HWE) p value less than 1E-5. Samples with wrong gender-information were removed. Missingness in cases compared to controls p value <1E-5 (from chi-squared test) and non-random missingness by haplotype <1E-5 (from chi-squared test). The case-control series then underwent the calculation of pairwise identity by descent, excluding any samples sharing greater than a 0.185 proportion of alleles (Phi_hat > 0.185) to extract probands from related pairs, effectively excluding all first, second and third degree relatives. Heterozygous outliers (deviate more than 3*SD from mean observed heterozygosity rate) were removed. Duplications were removed. Population outliers were removed (samples that deviated more than 2*SD from mean of the observed two first principal components when comparing to aggregate european reference samples). After this QC there were 216 cases, 558 controls and 120,579 variants.

In the second QC step individuals that deviated more than +/- 2*SD from the following statistics were removed (calculated using PlinkSeq, https://atgu.mgh.harvard.edu/plinkseq/index.shtml, individual statistics): Number of non-reference genotypes, Number of genotypes with a minor allele, Number of heterozygous genotypes for individual, Total number of called variants for individual, Genotyping rate for individual, Number of singletons individual has and mean Ti/Tv for variants. After removing individuals failing previous statistics, we removed variants with missingness less than 99% by genotyping call rate. The final dataset had 185 cases and 480 controls; 337 males and 328 females with 112,838 variants after the second QC. Genotype rate was 0.999. Ti/Tv ratio was 2.77 for all, 3.16

for known (dbsnp 138 excluding sites after 129) and 2.47 for novel variants. Selecting only exonic intervals (UCSC Genes) Ti/Tv ratio was 2.98, 3.4 and 2.65, respectively. Genomic inflation factor was 0.921.

Table S1 Stage I quality control (QC) steps in WES discovery dataset.

Before QC: Cases, N=238; Controls, N=563; Variants = 411,343
1. Removed non-random missingness between cases and controls, cutoff P < 1E-4
2. Removed samples with wrong sex-information (reported vs genetically derived)
3. Removed variants with missingness by call rate more than 5%
4. Removed samples with variant missingness by individual more than 10%
5. Hardy-Weinberg fail threshold 1E-4
6. Removed heterozygous outliers; deviation from mean observed heterozygosity rate < 3*SD
7. Removed samples that fail IBD check (relatedness), Phi-hat > 0.185
8. Minor allele frequency less than 10%
9. Removed ancestry outliers comparing to aggregate European reference samples; samples that deviate more than 2SD from observed mean in PC1 & PC2
After QC: Cases, N=216; Controls, N=558; Variants = 120,579

Key: IBD, identity by descent; Phi-hat, proportion of IBD; PC, principal component; SD, standard deviation.

Table S2 Stage II quality control steps in WES discovery dataset.

Before QC: Cases, N=216; Controls, N=558; Variants = 120,579
• **Removed samples that deviate more than 2SD from observed mean in each statistics:**
➢ Number of non-reference genotypes
➢ Number of genotypes with a minor allele
➢ Number of heterozygous genotypes
➢ Total number of called variants
➢ Genotyping rate
➢ Number of singletons
➢ Mean Ti/Tv
• **Removed variants with missingness > 1%**
After QC: Cases, N=185; Controls, N=480; Men, N=337; women, N=328; Variants = 112,838; Genotype rate 0.999

Key: Ti/Tv, transition transversion ratio.

Analysis

Kinship information from the variance component model was used to handle possible distant family relationships within the population. Single variant association (SVA) test was calculated using EMMAX test. Variants with MAF > 0.1% were used in the SVA test. . Gene-based analysis used variants with MAF < 5% and with assumed functional significance (nonsynonymous, Essential Splice Site, Normal Splice Site, Start Loss, Stop Loss, Stop Gain).

The first 20 principal components (based on variance-standardized relationship matrix), the first five components of the multidimensional scaling (based on raw Hamming distances) (Plink 1.9), sex, and exome sequencing batch information were used for both the SVA and gene-based tests as covariates. The batch information included the

identification code of the pooled library and the sequencing run the sample was in, and information if the sample was split between flow cells or not.

Using genetic data from cases only, genetic contributions to age at onset (AAO, based on age at first symptoms) were analyzed. No significant associations were found at a genome wide level of significance (data not shown). Analysis was done using Raremetal and AAO trait was inverse normalized.

Functional annotations were done using Annovar (Wang et al., 2010), SNPEff (Cingolani et al., 2012b) and SNPSift (Cingolani et al., 2012a). Plink version 1.9 was used to transform file format and perform several QC steps. Candidate genes were ranked by mutation accumulation rates using recently published data (Shyr et al., 2014), in order to exclude highly mutated genes.

Monogenic characterization and characterization of risk loci

Characterization of known PD genes and risk loci were done in the WES discovery dataset. Genes within 1 Mbp upstream or downstream of the loci identified in a meta-analysis (Nalls et al., 2014) were selected (supplementary table S3). In total, 372 genes fulfilled these criteria. R (Ihaka and Gentleman, 1996) library SKAT package's (http://cran.r-project.org/web/packages/SKAT) power calculator (Power_Logistic function) was used post-hoc to estimate the power of the dataset. Analyses for genes in gene-based tests were done either with all variants, or selecting variants that alter amino acid sequence of a protein, defined here as nonsynonymous variants (nonsynonymous, Essential_Splice_Site, Normal_Splice_Site, Start_Loss, Stop_Loss, Stop_Gain).

Power analysis

In order to generate haplotypes for the samples, ShapeIt2 with default settings was used. R library SKAT package's (http://cran.r-project.org/web/packages/SKAT) power calculator (Power_Logistic function) was used to estimate the power of our dataset post-hoc. The following settings were used:

1) SubRegion.Length was set as the length of the (longest) target gene transcript.
2) N.Sample.ALL = number of samples in the dataset (N=665).
3) Prevalence=0.02
4) Case.Prop=0.28 (Cases, N=185 cases / Total, N=665 samples).
5) Causal.Percent=100

6) Causal.MAF.Cutoff=0.03
7) N.Sim=1000
8) Weight.Param=c(1,25)
9) MaxOR=5

Variant ratio equation

Variant ratio is a value used to compare the ratio of variants per gene in cases versus controls. It is a division (MR^{cases} / $MR^{controls}$). Formula of MR = Total number of variants divided by transcript length and number of samples and multiplied by 10,000,000 (bp). Canonical transcripts and transcript lengths are used in equations.

Results

Genome wide association study

The GWAS compared 403 cases and 1650 controls (1132 men, 921 women). In the single variant association test, 12 loci had genome-wide associations to PD ($P < 5 \times 10^{-8}$) (Supplementary figure S1, supplementary table S4). None of the 24 previously identified loci (Nalls et al., 2014) showed significant association in the current data after Bonferroni correction. However, we found variants with close to significant associations with PD in similar but not identical loci near GBA (chr1:155820734), SIPA1L2 (chr1:233279538) and SNCA (chr4:90646501). Loci near ITGA8 (chr10:14376577) and NMD3 (chr3:160099190), which were not replicated in meta-analysis, had also variants with close to significant associations to PD.

GWAS SVA test revealed significant variants in a locus near the CEL gene. CEL was also significant ($p<2.5E-6$) in gene-based burden test in WES discovery dataset, although the effect of the gene variation to PD was very small with odds ratio of 0.92 (95% confidence interval (CI); 0.88-0.96).

Exome sequencing

All the candidate variants are nonsynonymous and result in amino acid change in the protein. The *GPR126* variant chr6:142758 results in an amino acid change (aa-change) that is predicted to be deleterious by SIFT (Kumar et al., 2009), Polyphen2 (Adzhubei et al., 2010), LRT (Chun and Fay, 2009) and MutationTaster (Schwarz et al., 2010). *TAS2R19* variant rs12424373 aa-change was predicted to be deleterious by SIFT. *SERPINA1* variant rs141620200 aa-change was predicted to be deleterious by SIFT, FATHMM (Shihab et al., 2013) and MetaLR (Dong et al., 2015). SERPINA variant rs17580 aa-change was predicted to be deleterious with SIFT, Polyphen2, LRT, MutationTaster, MutationAssessor (Reva et al., 2011) and FATHMM. Variants rs201600563 (ANKRD3), rs12424373 (*TAS2R19*) and rs61730995 (*ZNF519*) are located in a segmental duplication region. Also, all these candidate genes in table 1 have low FLAG (frequently mutated genes in public exomes) value indicating low general mutation rate.

Monogenic characterization and a characterization of risk loci

Characteristics of genes currently associated to Parkinson's disease or parkinsonism are summarized in supplementary tables S9-S12. None of the p values of the burden test in known PD genes was significant (supplementary table S9**)**, but variant count proportional to length of the gene's transcript is high with GRN and

PARK7 genes, when comparing cases to controls (supplementary table S10**,** where supplementary table S11 shows the putative impact classes used). Statistical post-hoc power of the burden tests in these known PD genes is generally low in our WES discovery dataset, ranging from 0.00015-0.43% with 1E-6 significance level in nonsynonymous variants and 0.00198-11.7% with 1E-6 significance level when calculated for all variants (supplementary table S12**)**.

Characterization of genes with possible association to PD

CEL

Carboxyl-ester ligase (encoded as *CEL*, MIM114840) has a physiological role in cholesterol and lipid-soluble vitamin ester hydrolysis and absorption. It may also play role in platelet activation and thrombin formation (Panicot-Dubois et al., 2007). Gene has a highly polymorphic region of insertion/deletion type (Taylor et al., 1991) which may indicate that the association is a false positive finding in our study. The Diseases associated with *CEL* include maturity-onset diabetes of the young, type viii, and type 2 diabetes (Bengtsson-Ellmark et al., 2004, Raeder et al., 2006).

It is good to acknowledge that control population data in our GWAS study has several individuals with type 2 diabetes (T2D). However, in GWAS this has been adjusted by using the T2D phenotype as covariate in SVA and gene-based tests. Also, in WES study the population controls are from different cohort (Stampeed) and in spite of that *CEL* gene has statistical significance in gene-based test.

Variants in *CEL* locus were associated with PD in GWAS and those in *CEL* gene in WES analysis. Eight EOPD cases had *CEL* variants in common in the GWAS and WES datasets and six of them had the *CEL* variant chr9:135955826 (human genome assembly 19) in the two datasets. The OR of the *CEL* variants was generally low in the analysis of the WES dataset. The *CEL* gene harbors a region rich in insertion/deletion polymorphisms (Taylor et al., 1991), which may point to the possibility that this is a false positive finding but, nevertheless, the role of this gene in PD warrants further investigation.

MPHOSPH10

M-phase phosphoprotein 10 (encoded as *MPHOSPH10*, MIM:605503), is involved in cell structural modifications during mitosis and is possibly a component of U3 snpRNP (Westendorf et al., 1998). Protein expression score is highest in brain tissue (Human proteome atlas (Uhlen et al., 2015)).

Three cases and one population control had rs143555311 variant in our study.

ANKRD36

Ankyrin repeat domain 36 (encoded as *ANKRD36*) is most highly transcribed in testis and bone marrow (Human proteome atlas). In the brain it is transcribed mainly in cerebral cortex.

Four cases and one control had rs201600563 variant in our study.

GPR126

G-protein-coupled receptor (encoded as *GPR126*, MIM: 616503) is involved in Swann cells in myelination (Monk et al., 2009). Variants in the gene are associated to arthrogryposis multiplex congenita (AMC) (Ravenscroft et al., 2015)which can overlap also with lethal congenital contracture syndrome 9, characterized by degeneration of anterior horn neurons, extreme skeletal muscle atrophy and congenital non-progressive joint contractures. Furthermore, *GPR126* has possible association to stature (MIM: 606255). Protein expression score is highest in the muscle, but relatively high also in the brain (Human proteome atlas).

Four cases had chr6:142758601 (human genome assembly 19) variant in our study. One case had familial PD.

TAS2R19

Taste receptor, type 2, member 19 (encoded as *TAS2R19*, MIM: 613961) is one of the bitter taste receptors (Fischer et al., 2005). There is evidence of impairment of taste in PD (see e.g. (Cecchini et al., 2015)), although the results are still inconclusive. Six cases and five controls had rs12424373 variant in our study.

SERPINA1

Serpin family A member 1 (encoded as *SERPINA1*, MIM: 107400) encodes a serine protease inhibitor protein, called alpha-1-antitrypsin (AAT) which is secreted outside the cell. Alpha 1-antitrypsin deficiency is suggested to be model for conformational diseases defined by aggregation of proteins due to abnormalities of various serpins (Carrell and Lomas, 2002). Alpha 1-antitrypsin inhibits transferrin binding to its receptor and is therefore involved in iron

metabolism (Graziadei et al., 1994). Alpha 1-antitrypsin allele PI*M3 has been associated to Alzheimer's disease (Kowalska et al., 1996).

The minor allele frequency of the rs141620200 variant in *SERPINA1* was significantly greater in the cases than in the controls, with 28 cases and 3 controls harboring the thymine allele. Intriguingly, it was found only in cases that were sequenced together suggesting a possible false positive finding because of e.g. batch effect. Furthermore, the allele coding differs between discovery and replication datasets, so that in the replication dataset, the reference allele is adenosine and alternate allele cytosine, whereas in the discovery dataset the reference allele is cytosine and alternate thymine or adenosine. It is noteworthy that seven of the 28 cases with the variant allele had family history of PD.

ZNF519

Not much is known about zinc finger protein 519 (encoded as *ZNF519*, NCBI gene: 162655). *ZNF519* has a KRAB domain and therefore the function may involve in transcriptional repression of euchromatic gene silencing (Schultz et al., 2002). Zinc finger genes that contain KRAB domain have been implicated to modulate retroelement transcription (Thomas and Schneider, 2011). In genetic studies, *ZNF519* has been previously implicated in ALS and autism (Cronin et al., 2009, Griswold et al., 2015).

Eleven cases and 10 population controls had rs61730995 variant. Age of PD onset was 35 years in one case.

PABPC1

The SVA analysis of the WES data identified several PABPC1 variants with possible associations with PD (supplementary table S5). These PABPC1 variants had generally low effect on disease risk, with SVA test OR varying from 1.10 to 1.13.

The poly(A)-binding protein (encoded as PABP, MIM: 604679) binds to 3´ end of eukaryotic mRNA and is involved in poly(A) shortening and translation initiation. Interestingly, *PABPC1* forms a complex with eukaryotic translation initiation factor 4-gamma (encoded as EIF4G1 MIM: 600495) (Tarun and Sachs, 1996), which have been suggested as a rare risk factor for PD (Huttenlocher et al., 2015). Furthermore, *PABPC1* is found in inclusions in amyotrophic lateral sclerosis patients (McGurk et al., 2014).

In our study, general mutation rate of the gene was estimated low by FLAG value. Quality value variant quality score log-odds in discovery dataset was relatively low in *PABPC1* variants, indicating possible false positive findings. Also the minor allele counts were rather high compared to allele frequencies in Exome aggregate consortium data of Finnish and European populations.

SUPPLEMENTARY TABLES AND FIGURES

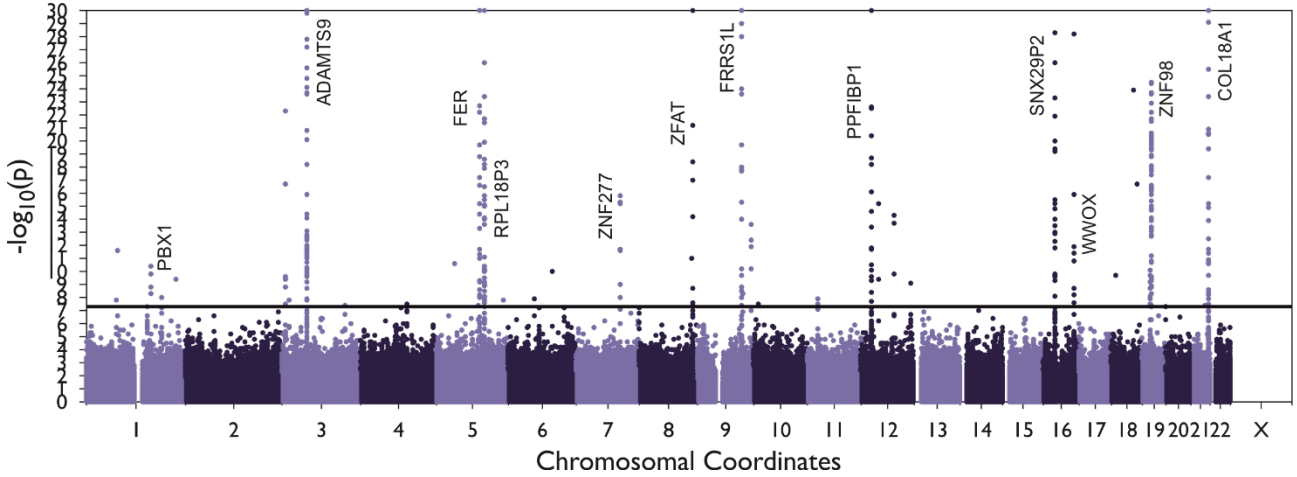


Figure S1 Manhattan plot of imputed GWAS single variant analysis. N=2053. Closest gene of the significant variants loci has been named with gene symbol.

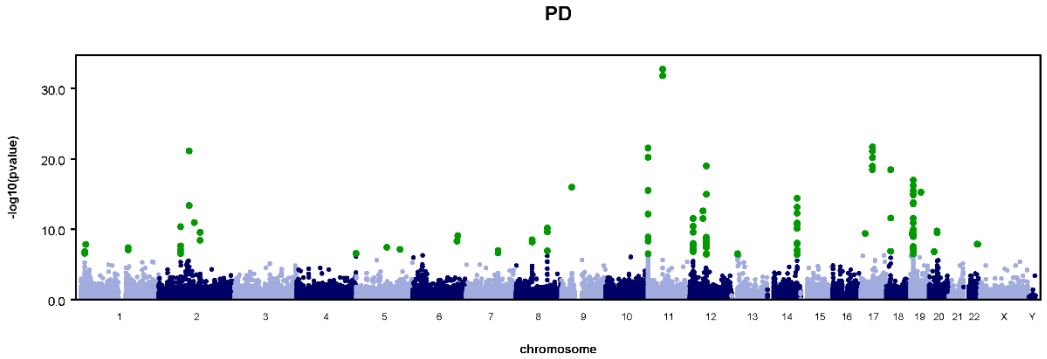
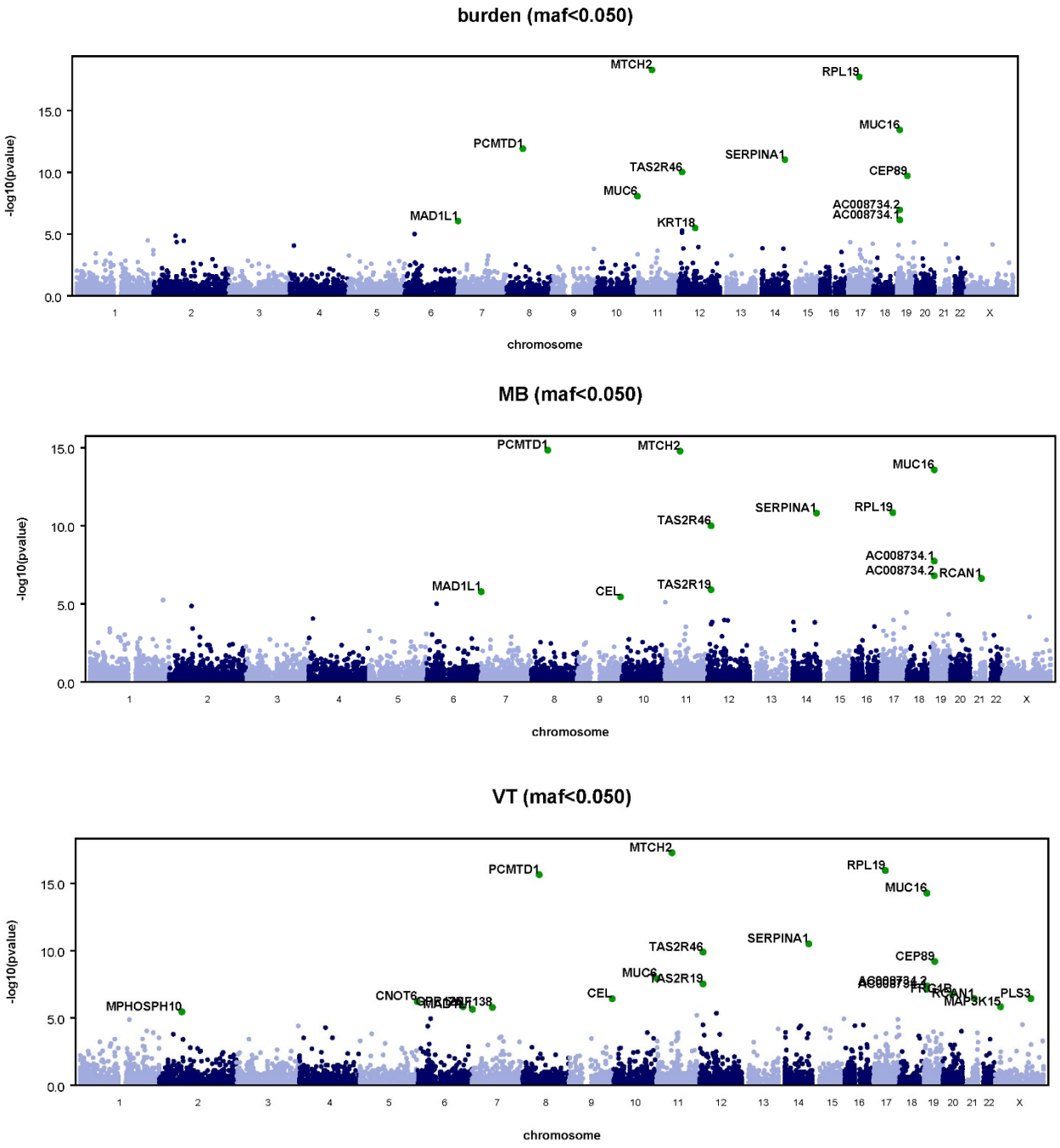


Figure S2 Manhattan plot for single variant associations in WES. Discovery dataset. Cases, N=185; Controls, N=480.


burden (maf<0.050)
MTCH2
RPL19
MUC16
PCMTD1
SERPINA1
TAS2R46
CEP89
MUC6
AC008734.2
AC008734.1
MAD1L1
KRT18
-log10(pvalue)
15.0
10.0
5.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
chromosome
MB (maf<0.050)
PCMTD1
MTCH2
MUC16
SERPINA1
RPL19
TAS2R46
AC008734.1
AC008734.2
RCAN1
MAD1L1
CEL
TAS2R19
-log10(pvalue)
15.0
10.0
5.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
chromosome
VT (maf<0.050)
MTCH2
PCMTD1
RPL19
MUC16
SERPINA1
TAS2R46
CEP89
MUC6
TAS2R19
CEL
PLS3
MAP3K15
MPHOSPH10
-log10(pvalue)
15.0
10.0
5.0
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
chromosome

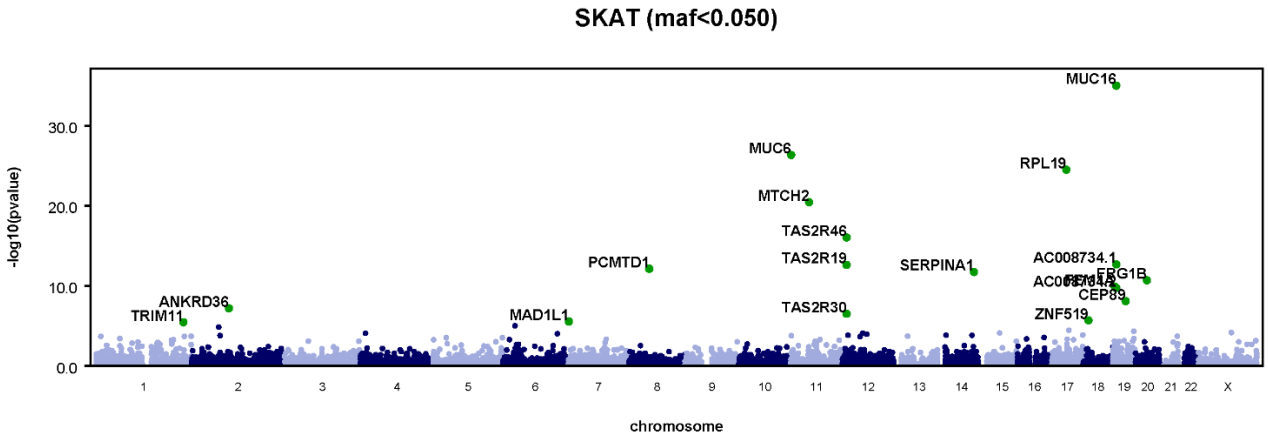


Figure S3 Manhattan plot for rare variant analysis in WES discovery dataset, using Burden, MB and SKAT tests with all variants with functional significance at MAF < 0.05.

Key: Burden, emmaxCMC test; SKAT, sequence kernel association test with SKAT-O; MB, Madsen-Browning weighted burden test; VT, variable threshold test.

Table S3 Genes within significant loci associated to PD in a recent meta-analysis.

GENE	TARGET REGION	GENE	TARGET REGION
SIPA1L2	chr1:231664611-233664611	GBA/SYT11	chr1:154135036-156135036
CCDC62	chr12:122303586-124303586	FGF20	chr8:15697091-17697091
GCH1	chr14:54348869-56348869	ACMSD/TMEM163	chr2:134539967-136539967
BST1	chr4:14737101-16737101	FAM47E/SCARB2	chr4:76198986-78198986
INPP5F	chr10:120536327-122536327	LRRK2	chr12:39614434-41614434
SREBF/RAI1	chr17:16715101-18715101	DDRGK1	chr20:2168166-4168166
RIT2	chr18:39673380-41673380	RAB7L1/NUCKS1	chr1:204723572-206723572
MCCC1	chr3:181762437-183762437	HLA-DQB1	chr6:31666660-33666660
BCKDK/STX1B	chr16:30121793-32121793	DLG2	chr11:81916055-86588314
STK39	chr2:168110394-170110394	ITGA8	chr10:14305947-17012334
MAPT	chr17:42994648-44994648	KRT8P25/APOOP2	chr3:86072000-88030857
GPNMB	chr7:22293746-24293746	MMP16	chr8:87799459-90589717
VPS13C	chr15:60994134-62994134	NMD3	chr3:159689098-162219795
MIR4697	chr11:132765367-134765367	SPPL2B	chr19:1363321-3363321
TMEM175/GAK/DGKQ	chr4:1948053-1951947	TMEM229B	chr14:66517653-68517653
SNCA	chr4:89626111-91626111	USP25	chr21:14542586-16542586

Target region = ± 1Mb area to search additional genes.

Table S4. Variants with the most significant p values in imputed GWAS study single variant analysis.

CHROM	POS	ALT	NS	AC	MAF	CTRLCNT	CASECNT	CLOSEST_GENE	PVALUE	OR	CI
21	46916033	A	2053	371	0.09036	1604/46/0	144/193/66	COL18A1	8.09E-240	1.719	1.673-1.768
3	64699445	T	2053	2298	0.44033	243/782/625	176/188/39	ADAMTS9-AS2	6.38E-44	1.78	1.707-1.857
12	27737074	C	2053	2316	0.43595	392/817/441	24/141/238	PPFIBP1	1.09E-33	1.4	1.359-1.442
16	29348809	G	2053	1625	0.39576	702/724/224	0/353/50	SNX29P2	4.35E-28	1.216	1.185-1.247
16	79360135	A	2053	38	0.00925	1650/0/0	366/36/1	WWOX	2.61E-27	0.852	0.834-0.871
8	135611154	A	2053	550	0.13395	1399/245/6	157/199/47	ZFAT	3.12E-96	1.152	1.127-1.177
19	22620820	T	2053	479	0.11666	1203/423/24	396/6/1	ZNF98	6.32E-25	1.213	1.176-1.253
7	111916329	T	2053	54	0.01315	1630/20/0	370/32/1	ZNF277	2.14E-16	1.151	1.123-1.179
5	108436958	G	2053	193	0.047	1641/9/0	248/126/29	FER	3.91E-133	1.945	1.734-2.182
5	120977240	A	2053	3141	0.23502	122/690/838	0/31/372	RPL18P3	2.51E-46	0.826	0.797-0.856
9	111929013	T	2053	590	0.14369	1298/337/15	210/163/30	FRRS1L	1.77E-31	1.517	1.375-1.673
1	164546863	T	2053	116	0.02825	1589/59/2	351/51/1	PBX1	1.49E-10	1.257	1.175-1.346
9	135955826	G	2053	14	0.00341	1650/0/0	389/14/0	CEL	1.19E-14	2.158	1.777-2.619

Cases, N=403; Controls, N=1650. Bonferroni multiple test correction cutoff value is 3,86E-09.

Key: CHR, chromosome; POS, Genome position (Grch37); ALT, Alternate allele; NS, number of samples; AC, minor allele count; AF, minor allele frequency; CTRLCNT & CASECNT, allele counts in controls and cases (respectively): reference/heterozygous alternate/homozygous alternate; CLOSEST_GENE, gene which is closest to variant locus; PVALUE, P value in single variant analysis; OR, odds ratio; CI, confidence interval in OR.

Table S5 Characterization of variants significant in WES discovery dataset SVA analysis that are also found in replication dataset.

#Discovery dataset I													#Replication dataset II	
GENE	ID	AF	ExAC FIN	ExAC All	RadialSVM_ pred	LR pred	ExonicFunction	Cases	Controls	P_BONF	OR	CI	AF	AC	AN
NMBR	rs41289829	0.0057	0.0043	0.000873	.	.	synonymous	0,6,6	0,3,3	8.02E-04	1.359	1.226-1.506	0.005	2	400
GPR126	rs146727650	0.0057	0.0051	0.001011	T	T	nonsynonymous	0,6,6	0,3,3	8.02E-04	1.359	1.226-1.506	0.005	2	400
PABPC1	rs78146983	0.0882	0.0003	0.00042	D	D	nonsynonymous	0,95,95	0,44,44	4.78E-09	1.131	1.096-1.167	0.0025	1	398
PABPC1	rs113574896	0.0952	0.0023	0.002347	T	T	nonsynonymous	0,99,99	0,51,51	4.66E-08	1.122	1.088-1.156	0.0051	2	392
PABPC1	rs112868101	0.0945	0.0048	0.007026	D	D	nonsynonymous	0,102,102	0,47,47	3.79E-05	1.103	1.07-1.136	0.0126	5	396
PABPC1	rs78407297	0.1022	0.0043	0.008401	T	T	nonsynonymous	0,106,106	0,55,55	1.72E-04	1.106	1.071-1.142	0.0102	4	394
PABPC1	rs140835766	0.0844	0.1801	0.198	D	D	nonsynonymous	0,90,90	0,43,43	1.28E-03	1.095	1.062-1.13	0.0026	1	388
PABPC1	rs72681440	0.0869	0.2228	0.226	T	D	nonsynonymous	0,88,88	0,49,49	1.33E-03	1.098	1.064-1.133	0.0077	3	392
SERPINA1	rs141620200	0.019,0.0006	0.0026	0.002339	T	D	nonsynonymous	0,28,28	0,3,3	1.49E-04	1.265	1.173-1.363	0.01	4	400

Discovery dataset Cases, N=185, controls, N=480. Replication dataset Cases, N=200.

Key: Gene, gene symbol where variant is located; ID, variant rs-number; AF, minor allele frequency; ExAC FIN, Exome aggregate consortium allele frequency for Finnish population samples; ExAC All, Exome aggregate consortium allele frequency for all populations; RadialSVM_pred, MetaSVM deleterious prediction score (Dong et al., 2015); LR_pred, MetaLR deleterious prediction score (Dong et al., 2015); Cases, AC count Homozygous/Heterozygous/Total; Controls, AC count Homozygous/Heterozygous/Total; P_Bonf, Bonferroni corrected P value; OR, estimated odds ratio for single variant; CI, Confidence interval for OR; AC, minor allele count; AN, Allele number.

Table S6 Top 20 variants in WES discovery and replication datasets not found from Finnish individuals in Exome Aggregation Consortium. Sorted by Odds ratio (OR).

GENE	ID	REF	ALT	AF	OR	CI	IMPACT
PAX8	rs78802229	T	C	0.13	1.63	1.202-2.197	modifier
TRPM6	rs11144075	C	T	0.03	1.49	0.858-2.576	modifier
ARPC5L	rs876663	A	G	0.45	1.46	1.19-1.791	modifier
NELFCD	rs1043219	C	G	0.11	1.46	1.60-2.00	modifier
SELL	rs909628	A	G	0.09	1.45	1.021-2.058	modifier
SLC12A2	rs10089	C	T	0.27	1.44	1.138-1.82	modifier
CCDC38	rs17369927	G	A	0.24	1.35	1.066-1.716	modifier
ATP6V1E1	rs3532	T	C	0.23	1.33	1.043-1.695	modifier
SMIM21	rs11660753	G	A	0.56	1.33	1.087-1.624	modifier
SCMH1	rs11555079	G	A	0.26	1.31	1.039-1.661	modifier
SH3BGRL3	rs7544	T	G	0.25	1.30	1.025-1.648	modifier
ATP6V1E1	rs5746448	A	C	0.25	1.28	1.012-1.622	modifier
UBXN11	rs12090258	C	T	0.25	1.26	0.998-1.598	modifier
SPOCK2	rs3088251	T	C	0.13	1.25	0.934-1.687	modifier
UGT2B4	rs1966151	A	G	0.32	1.25	1.004-1.564	modifier
AIM1L	rs17163868	C	T	0.25	1.25	0.985-1.576	modifier
.	rs9524568	T	A	0.36	1.24	1.001-1.538	.
PIFO	rs3738303	G	A	0.15	1.24	0.936-1.645	modifier
AKR1B10	rs2037004	G	A	0.59	1.24	1.017-1.505	modifier
ZNF417	rs10414704	G	T	0.41	1.23	0.998-1.514	modifier

Discovery dataset Cases, N=185, controls, N=480. Replication dataset Cases, N=200. In Minor allele frequency and odds ratio calculations cases from discovery and replication datasets are added together.

Key: Gene, gene symbol where variant is located; ID, variant rs-number; REF, reference allele; ALT, alternate allele; AF, minor allele frequency; OR, estimated odds ratio for single variant; CI, Confidence interval for OR; Impact, a subjective classification of the severity of the variant consequence; modifier, usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact.

Table S7 Burden, MB and SKAT test results in WES study. Using all nonsynonymous variants at MAF < 0.05

##METHOD=BURDEN								
GENE	**NUM_VAR**	**AVG_AF**	**MIN_AF**	**MAX_AF**	**EFFECT_SIZE**	**PVALUE**	**P_BONF**	**P_BH**
MTCH2	7	0.017546	0.001504	0.036145	0.078561	7.76E-19	9.94E-15	9.94E-15
RPL19	9	0.025653	0.003759	0.048872	0.04494	9.55E-17	1.22E-12	6.12E-13
MUC16	176	0.011085	0.000752	0.043609	0.010882	7.57E-14	9.70E-10	2.79E-10
PCMTD1	11	0.020096	0.001504	0.048872	0.053275	8.71E-14	1.12E-09	2.79E-10
SERPINA1	13	0.01637	0.006767	0.026316	0.023175	9.23E-11	1.18E-06	2.36E-07
CEP89	5	0.014806	0.000752	0.049219	-0.10159	2.32E-10	2.97E-06	4.95E-07
TAS2R46	18	0.013964	0.002262	0.034509	0.03013	1.00E-09	1.28E-05	1.83E-06
MUC6	99	0.012778	0.000752	0.048946	0.012889	3.77E-09	4.83E-05	6.04E-06
AC008734.2	29	0.014133	0.001504	0.043609	0.01913	1.44E-07	0.00184	0.000204
TAS2R19	26	0.008126	0.001504	0.029323	0.022546	8.53E-07	0.010931	0.001093
AC008734.1	44	0.011557	0.000752	0.029368	0.015441	9.94E-07	0.012733	0.001158
MAD1L1	2	0.004135	0.003008	0.005263	-0.19106	1.14E-06	0.014559	0.001213
KRT18	5	0.025058	0.002259	0.048485	0.035783	2.21E-06	0.02826	0.002174
MPHOSPH10	1	0.001504	0.001504	0.001504	0.465089	3.46E-06	0.044334	0.003167
TAS2R30	13	0.012608	0.002256	0.042857	0.037238	5.26E-06	0.067364	0.004491
CPXCR1	1	0.011278	0.011278	0.011278	0.11817	3.07E-05	0.393974	0.023981
KDM6B	7	0.007426	0.001504	0.018045	-0.07127	3.18E-05	0.407678	0.023981
RCAN1	2	0.005639	0.000752	0.010526	-0.151	3.90E-05	0.499893	0.027399
STK19	1	0.00458	0.00458	0.00458	0.239769	4.06E-05	0.520585	0.027399
COG2	2	0.012815	0.003759	0.02187	0.098783	6.17E-05	0.790162	0.039508
##METHOD=MB: ADDITIONAL GENES								
GENE	**NUM_VAR**	**AVG_AF**	**MIN_AF**	**MAX_AF**	**EFFECT_SIZE**	**PVALUE**	**P_BONF**	**P_BH**
CEL	4	0.016354	0.000753	0.041353	-0.00674	2.87E-06	0.036801	0.003067
KRTAP4-7	3	0.008779	0.002276	0.018045	0.009032	2.53E-05	0.323665	0.020229
CNOT6	2	0.00188	0.001504	0.002256	-0.01131	3.16E-05	0.404368	0.022465
FOXK2	3	0.010395	0.001527	0.01579	0.007004	4.08E-05	0.522144	0.026107

##METHOD=VT:
ADDITIONAL GENES

GENE	NUM_VAR	AVG_AF	MIN_AF	MAX_AF	EFFECT_SIZE	MAF_CUTOFF	PVALUE	P_BONF
FRG1B	16	0.004512	0.000752	0.012801	0.074994	0.012801	1.66E-07	0.002128
PLS3	1	0.000752	0.000752	0.000752	-0.73126	0.000752	3.61E-07	0.004625
CEL	1	0.000753	0.000753	0.000753	-0.73105	0.000753	3.68E-07	0.004719
CNOT6	1	0.001504	0.001504	0.001504	-0.5086	0.001504	5.90E-07	0.007561
GPR126	3	0.002758	0.001504	0.005263	0.220807	0.005263	1.42E-06	0.018178
MAP3K15	1	0.000752	0.000752	0.000752	-0.73126	0.000752	1.44E-06	0.018486
ZNF138	1	0.001504	0.001504	0.001504	-0.51083	0.001504	1.61E-06	0.020589

##METHOD=SKAT:
ADDITIONAL GENES

GENE	NUM_VAR	AVG_AF	MIN_AF	MAX_AF	PVALUE_LIU	P_BONF	P_BH
FRG1B	27	0.013924	0.000752	0.045113	7.87E-12	1.01E-07	1.12E-08
FEM1A	3	0.00604	0.003008	0.009848	6.81E-10	8.73E-06	7.28E-07
ANKRD36	33	0.009255	0.000752	0.045113	5.94E-07	0.007605	0.000543
ZNF519	4	0.006767	0.000752	0.010526	1.16E-06	0.014896	0.000931
TRIM11	2	0.000752	0.000752	0.000752	3.86E-06	0.04951	0.002606

Total samples, N=665; Cases, N=185; Controls, N=480. Using all variants with functional significance at MAF < 0.05.

Key: NUM_VAR, number of variants in the test. AVG_AF, Average allele frequency (AF). MIN_AF, Minimum AF.MAX_AF, Maxium AF. EFFECT SIZE, Effect size of the test.

PVALUE & PVALUE_LIU, P value of the test. P_BONF, Bonferroni corrected p value. P_BH, Benjamini & Hochberg corrected p value.

Table S8 Characterization of variants significant in WES discovery dataset in gene-based tests and found also in replication datasets.

#DISCOVERY DATASET I															REPLICATION DATASET II	
GENE	CHR	POS	ID	AF	ExAC FIN	ExAC All	AC	AN	Cases	Controls	Exonic Function	VQSLOD	FLAG	SISU_AF	SISU_Rep	AF	AC
MPHOSPH10	2	71361156	rs143555311	1.21E-03	0.0012	0.002	2	1654	0,1,1	0,1,1	nonsyn	2.84	48	0.0013	-	0.005	2
ANKRD36	2	97858622	rs201600563	1.21E-03	0.003	0.0014	2	1654	0,1,1	0,1,1	nonsyn	-6.042	24	0.0003	SEG_DUPL	0.0076	3
GPR126	6	142758601	.	1.21E-03	0.0006	4.14E-05	2	1652	0,2,2	0,0,0	nonsyn	1.77	82	0.0008	-	0.005	2
TAS2R19	12	11174795	rs12424373	4.23E-03	0.0048	0.05	7	1654	0,1,1	0,5,5	nonsyn	0.532	28	0.0039	SEG_DUPL	0.0125	5
SERPINA1	14	94845944	rs141620200	0.021,6.0e-04	0.0026	0.0023	31,1	1654	0,28,28	0,3,3	nonsyn	-0.7318	49	0.0019	-	0.01	4
SERPINA1	14	94847262	rs17580	7.86E-03	0.0089	0.02	13	1654	0,4,4	0,9,9	nonsyn	3.62	49	0.0083	-	0.0125	5
ZNF519	18	14105770	rs61730995	0.012	0.0044	0.0051	20	1652	1,6,8	0,10,10	nonsyn	3.01	42	0.0053	SEG_DUPL	0.01	4

Discovery dataset Cases, N=185, controls, N=480. Replication dataset Cases, N=200.

Key: Gene, gene symbol where variant is located; Chr, chromosome; Pos, Genome position (Grch37); ID, variant RS-number; SISU AF, SISU project (Lim et al., 2014) database allele frequencies; AF, minor allele frequency; AC, minor allele count; Cases, AC count Homozygous/Heterozygous/Total; Controls, AC count Homozygous/Heterozygous/Total; SISU Rep, Indicates if variant locates in segmental duplication region. ExAC FIN, Exome aggregate consortium allele frequency for Finnish population samples; ExAC All, Exome aggregate consortium allele frequency for all populations; FLAG, frequently mutated genes in public exomes -value: less is better; Gets values from 1 to 2659 (19721 genes, value of 100th gene = 253). VQSLOD, variant quality score log-odds in discovery dataset: QC value from Variant quality score recalibration in GATK.

Table S9 Variant count and burden test p values of the genes associated with PD, non-PD genes that may present with Parkinsonism and unconfirmed genes that may be associated with PD.

GENE	Snps	Rare_Snps	P_Burden	NonSyn_Snps	Rare_Nonsyn_snps	P_Burden_nonsyn
ATP13A2	16	9	0.4508	5	3	0.6602
ATXN2	5	5	0.642	3	3	0.4427
ATXN3	9	4	0.3301	2	1	0.5498
CSF1R	15	11	0.3355	6	6	0.574
DNAJC6	10	5	0.453	3	2	0.04895
EIF4G1	16	14	0.7365	6	5	0.9864
FBXO7	7	3	0.7476	4	2	0.2935
GBA	7	6	0.09467	4	5	0.2459
GCH1	5	3	0.8744	1	0	NA
GIGYF2	14	9	0.6204	2	0	NA
GRN	4	4	0.06568	2	2	0.03456
HTRA2	2	2	0.8221	1	1	0.786
LRRK2	21	8	0.07878	12	7	0.05942
MAPT	20	7	0.6435	7	2	0.8148
PARK2	6	5	0.2825	6	5	0.2825
PARK7	3	3	0.3488	1	1	0.9214
PINK1	9	5	0.7337	6	4	0.695
PLA2G6	9	9	0.675	2	2	0.9604
POLG	15	14	0.1513	11	11	0.1785
SPG11	16	15	0.04845	8	7	0.07389
TH	2	1	0.3972	2	1	0.3972
VPS35	3	3	0.1331	1	1	0.132

Key: GENE, gene symbol; Snps, Total variant count; Rare_Snps, Minor allele frequency < 5%; NonSyn_Snps, total nonsynonymous variants; Rare_Nonsyn_snps, rare nonsynonymous variants (MAF<5%); P_Burden_nonsyn, p value in burden test using nonsynonymous rare variants.

Table S10 Difference of variant counts and variant rates in WES discovery dataset in known PD genes.

GENE	HIGH	LOW	MODERATE	MODIFIER	Total variants	VariantRatio (MR^{case}:$MR^{control}$)
ATP13A2	0	-4	0	6	2	2.77353
ATXN2	0	-1	-1	2	0	2.5946
ATXN3	0	0	0	1	1	2.88288
CSF1R	0	-5	-1	6	0	2.5946
DNAJC6	0	1	1	-2	0	2.5946
EIF4G1	0	-3	0	1	-2	2.40927
FBXO7	0	-1	-1	2	0	2.5946
GCH1	0	0	0	0	0	2.5946
GIGYF2	0	-2	0	0	-2	2.16216
GRN	0	1	-1	1	1	3.24324
HTRA2	0	0	0	0	0	2.5946
LRRK2	0	-2	-1	2	-1	2.50513
MAPT	0	0	-1	1	0	2.5946
PARK2	0	0	0	0	0	2.5946
PARK7	0	0	1	0	1	3.89189
PINK1	0	-1	-1	0	-2	2.33514
PLA2G6	0	-2	-1	3	0	2.5946
POLG	0	0	-1	1	0	2.5946
SNCA	0	0	0	0	0	2.5946
SPG11	0	-3	1	2	0	2.5946
SPR	0	0	0	0	0	2.5946
TH	0	0	0	0	0	2.5946
UCHL1	0	0	0	0	0	2.5946
VPS35	0	0	-1	1	0	2.5946

Variant counts are results of substraction ($SNPS^{cases}$ – $SNPS^{controls}$). $Mean^{variantRatio}$=2.55669, $SD^{variantRatio}$=(2.297319, 2.816061). Mean and SD are calculated using all genes in discovery dataset.

Key: HIGH,LOW,MODERATE,MODIFIER, variant counts in putative impact classes (see supplementary **table S11**); Total variants, sum of all impact classes variant counts. Variant ratio, value defining variant ratio in cases vs controls (see methods).

Table S11 SNPEff putative impact classes.

Putative Impact	Sequence Ontology term
HIGH	chromosome_number_variation
	exon_loss_variant
	frameshift_variant
	rare_amino_acid_variant
	splice_acceptor_variant
	splice_donor_variant
	start_lost
	stop_gained
	stop_lost
	transcript_ablation
MODERATE	3_prime_UTR_truncation+exon_loss
	5_prime_UTR_truncation+exon_loss_variant
	coding_sequence_variant
	disruptive_inframe_deletion
	disruptive_inframe_insertion
	inframe_deletion
	inframe_insertion
	missense_variant
	regulatory_region_ablation
	splice_region_variant
	TFBS_ablation
LOW	5_prime_UTR_premature start_codon_gain_variant
	initiator_codon_variant
	splice_region_variant
	start_retained
	stop_retained_variant
	synonymous_variant
MODIFIER	3_prime_UTR_variant
	5_prime_UTR_variant
	coding_sequence_variant
	conserved_intergenic_variant

Table S12 Statistical post-hoc power of burden analysis in all variants in known PD genes.

Gene	post-hoc power			
	all snps		nonsyn snps	
	p=0.001	p=1.00E-06	p=0.001	p=1.00E-06
ATP13A2	0.592	0.11724	0.051	0.00071
ATXN2	0.035	0.0004	0.031	0.0003
ATXN3	0.004	1.98E-05	0.009	4.96E-05
CSF1R	0.227	0.00962	0.144	0.0043
DNAJC6	0.055	0.00071	NA	NA
EIF4G1	0.544	0.08763	0.036	0.00048
FBXO7	0.046	0.00063	0.033	0.00035
GCH1	0.011	7.71E-05	NA	NA
GIGYF2	0.113	0.00293	NA	NA
GRN	0.076	0.00151	NA	NA
LRRK2	0.096	0.00277	0.082	0.00205
MAPT	0.068	0.00137	0.004	2.23E-05
PARK7	0.074	0.00135	NA	NA
PINK1	0.091	0.00194	0.091	0.00195
PLA2G6	0.108	0.00285	NA	NA
POLG	0.186	0.00778	0.085	0.00189
SPG11	0.04	0.00049	0.002	1.50E-06
VPS35	0.099	0.00211	NA	NA

WES discovery dataset. Calculated with either all snps or nonsynonymous snps. Significance levels are 0.001 and 1E-6.

Key: Gene, Gene symbol; all snps, all rare variants; nonsyn snps, nonsynonymous variants. NA, not enough nonsynonymous variants to calculate the power.

Table S13 CEL variants in WES discovery dataset.

ID	AF_Cases	AF_Controls	POS	REF	ALT	AF	ExAC_FIN	ExAC_Freq	VQSLOD	AC	AN	ExonicFunction
rs113056079	0.00225	0	135940439	T	G	0.000622	0	0.005015	3.87	1	1530	nonsynon
rs150358550	0.01778	0.00446	135940488	C	G	0.009102	0.0085	0.00444	3.88	13	1570	synon
chr9:135940573	0.00222	0	135940573	TAC	T	0.000606	.	.	-2.324	1	1572	frameshift_deletion
chr9:135941981	0	0.01421	135941981	C	CG	0.01	0	0.0000662	4.37	16	1576	frameshift_insertion
chr9:135942572	0	0	135942572	T	C	0.000605	.	.	-3.867	0	1576	nonsynon
rs201255412	0.24215	0.27422	135944524	C	T	0.265	0.4323	0.165	4.31	408	1540	synon
chr9:135944553	0	0.00091	135944553	A	T	0.001259	0.0053	8.88E-05	0.975	1	1512	nonsynon
rs200119384	0.2644	0.2278	135944586	C	T	0.24	0.4926	0.127	2.05	337	1418	nonsynon
rs201074543	0	0.01421	135945988	A	G	0.009662	0	0.002216	-5.398	16	1576	nonsynon
rs201677850	0	0.01865	135945997	C	T	0.013	0	0.002282	-5.788	21	1576	nonsynon
rs77696629	0.00444	0.0524	135946015	T	C	0.038	0.0043	0.019	-3.78	61	1576	nonsynon
rs202171778	0.0067	0.00089	135946390	G	C,A	0.001208,0.001208	0	0.002478	-1.084	2,2	1576	nonsynon
chr9:135946507	0	0.00089	135946507	C	A	0.000605	0	5.79E-05	1.54	1	1574	nonsynon
rs488087	0.311	0.17557	135946599	C	T	0.225	0.4212	0.252	4.29	268	1204	synon
chr9:135946921	0	NA	135946921	CCCCCCCCCGTGCCGCCCACGGGTGACTCCGGCG	C	0.018	0.05	0.038	5.32	0	98	nonframeshift_deletion
chr9:135946957	0.00463	0	135946957	CCCCCCGTGCCGCCCACGGGTGACTCCGGGGCCCCCCCGTGACCCCCACGGGTGACTCCGAGACCG	C	0.003268	0	0.00079	4.57	1	274	nonframeshift_deletion
chr9:135947001	0	0.00192	135947001	CCCCACGGGTGACTCCGAGACCGCCCCCGTGCCG	C	0.001027	0.09	0.016	4.51	1	916	nonframeshift_deletion
chr9:135947018	0.00495	0	135947018	A	G	0.001718	0	5.30E-05	0.77	2	1098	nonsynon

chr9:135947020	0.00493	0	135947020	A	G	0.001709	0	4.41E-05	0.115	2	1104	nonsynon
chr9:135947023	0.00246	0	135947023	G	C	0.00084	0.0005	0.002629	1.51	1	1124	nonsynon
rs201133893	0	0.00209	135947051	G	A	0.001385	0.0004	0.000151	2.18	2	1378	nonsynon
rs202034862	0	0.0021	135947053	G	A	0.001391	0	1.68E-05	0.616	2	1372	nonsynon
chr9:135947055	0	0.00208	135947055	CCCCCCTGT	C	0.001389	0	5.03E-05	2.98	2	1376	frameshift_deletion
chr9:135947069	0	0.00196	135947069	CCACGGGTGACTCTGAGGCTGCCCCT	C	0.001319	0.0003	0.000108	3.79	2	1444	frameshift_deletion
rs200257295	0	0.00185	135947100	C	G	0.001267	0.0003	0.000124	0.471	2	1504	synon

Cases, N=185; Controls, N=480.

Key: ID, variant RS-number; AF_Cases, minor allele frequency in cases; AF_Controls, minor allele frequency in controls; POS, Genome position (Grch37); REF, reference allele; ALT, alternate allele; AF, minor allele frequency; ExAC FIN, Exome aggregate consortium allele frequency for Finnish population samples; ExAC Freq, Exome aggregate consortium allele frequency for all populations; AC, minor allele count; AN, Allele number; VQSLOD, variant quality score log-odds in discovery dataset: QC value from Variant quality score recalibration in GATK; Exonic Function, Functional effects caused by variant; nonsyn, nonsynonymous; synon, synonymous.

Table S14 *CEL* variants in WES replication dataset.

ID	POS	REF	ALT	AF	ExAC_FIN	ExAC_Freq	AC	AN	ExonicFunction
rs113056079	135940439	T	G	0	0	0.005015	0	400	nonsynonymous
rs150358550	135940488	C	G	0.0075	0.0085	0.00444	3	400	synonymous
chr9:135942035	135942035	C	T	0.0025	0.0005	0.0009177	1	400	synonymous
rs77696629	135946015	T	C	0.0025	0.0043	0.019	1	400	nonsynonymous
chr9:135946507	135946507	C	A	0.0025	0	5.79E-05	1	400	nonsynonymous
rs78256304	135946548	C	G	0	0	0.002525	0	400	synonymous
chr9:135946578	135946578	A	G	0	0.001	0.0008705	0	400	synonymous
chr9:135946581	135946581	G	T	0	0.001	0.0008707	0	400	synonymous
chr9:135946605	135946605	C	G	0	0.0017	0.0009902	0	400	synonymous
chr9:135946690	135946690	G	C	0.0075	.	0.036	3	400	nonsynonymous
chr9:135946953	135946953	C	G	0	0	0.001013	0	392	synonymous
chr9:135946954	135946954	G	C	0	.	.	0	394	nonsynonymous
chr9:135946995	135946995	C	T	0.0075377	0.0101	0.006498	3	398	synonymous
chr9:135947032	135947032	C	A	0.0025381	0	1.74E-05	1	394	nonsynonymous
chr9:135947034	135947034	G	C	0.0100503	0	0.0006173	4	398	synonymous
chr9:135947061	135947061	T	C	0.0025	0.0002	0.0001424	1	400	synonymous

ID, variant RS-number; POS, Genome position (Grch37); REF, reference allele; ALT, alternate allele; AF, minor allele frequency; ExAC FIN, Exome aggregate consortium allele frequency for Finnish population samples; ExAC Freq, Exome aggregate consortium allele frequency for all populations; AC, minor allele count; AN, Allele number; Exonic Function, Functional effects caused by variant.

Table S15 Genes closest to significant loci in GWAS analysis of the Finnish PD cohort.

FER	GON4L	EYS	chr12:126370875
COL18A1	PBX1	IRAK1BP1	TMEM117
ZFAT	chr1:189532844	ZNF277	chr12:84430416
chr3:64702684	chr1:226204970	chr8:132177139	chr16:79355947
chr5:120977240	chr3:20432637	chr8:135477079	chr18:11450936
PPFIBP1	SRGAP3	chr9:114064016	LOC643542
FRRS1L	chr4:117161575	CEL	chr19:22510418
SNX29P2	KCNIP1	RALGDS	chr19:22660286
ZNF98	chr6:113278451	chr10:14376577	chr20:699763
ALPK2	chr6:142312266	ANO3	LOC101928269

Table S16 Imputed GWAS study, gene-set analysis top hits with Burden test.

##Method=Burden										
GENE	**CHR**	**BEG**	**END**	**TOT_MARKERS**	**PASS_MARKERS**	**PVALUE**	**P_BONF**	**P_BH**	**OR**	**CI**
CELP	9	135961850	135961851	2	2	2.14E-13	3.11E-09	3.11E-09	2.14	1.747-2.614
TPTE2P6	13	25144872	25171521	3	1	1.66E-06	0.024	0.012	2.08	1.542-2.801
RAB11FIP2	10	119774577	119774598	2	2	4.35E-06	0.063	0.021	1.22	1.121-1.327
OR10G2	14	22102317	22102842	6	1	1.43E-05	0.207	0.052	1.77	1.368-2.287
GBA	1	155205043	155210498	3	3	3.27E-05	0.474	0.083	1.12	1.059-1.174
KRTAP10-7	21	46020720	46021546	10	5	3.44E-05	0.500	0.083	1.11	1.058-1.171
CTC-360P9.3	19	32528454	32528454	1	1	1.35E-04	1	0.279	2.04	1.416-2.94
CYP4F8	19	15730502	15739661	2	1	1.72E-04	1	0.311	1.65	1.271-2.141
COMMD5	8	146076611	146076708	2	1	4.35E-04	1	0.619	1.55	1.214-1.977
RP11-368J21.2	16	22546193	22546193	1	1	4.48E-04	1	0.619	2.15	1.404-3.304

Cases, N=403; Controls, N=1650.

Key: GENE, gene symbol of the closest gene in the variant loci; CHR, chromosome; BEG & END, Gene genome position (Grch37) beginning and end; TOT_MARKERS, total number of variants; PASS_MARKERS, variants that pass the filters; PVALUE, P value in burden variant analysis; P_BONF, Bonferroni corrected p value; P_BH, Benjamini & Hochberg corrected p value; OR, odds ratio; CI, confidence interval in OR.

Table S17 Imputed GWAS study, gene-set analysis top hits with SKAT test.

##Method=SKAT									
GENE	**CHR**	**BEGIN**	**END**	**NUM_ALL_VARS**	**NUM_PASS_VARS**	**NUM_SING_VARS**	**PVALUE**	**P_BONF**	**P_BH**
CELP	9	135961850	135961851	2	2	0	1.49E-11	2.16E-07	2.16E-07
COL18A1	21	46875586	46930147	22	18	2	3.66E-06	0.05312	0.02656
PTPN13	4	87610798	87724970	8	6	0	5.54E-06	0.080338	0.026779
OR10G2	14	22102317	22102842	6	1	0	1.07E-05	0.155387	0.038847
COX19	7	1015109	1015110	2	2	2	1.62E-05	0.234416	0.046883
TRIM32	9	119460579	119460579	1	1	0	2.61E-05	0.379019	0.06317
CYP4F8	19	15730502	15739661	2	1	0	3.16E-05	0.45915	0.065593
NETO1	18	70417396	70532471	3	2	0	4.66E-05	0.676019	0.072132
ANO3	11	26463582	26558947	2	2	0	4.90E-05	0.711868	0.072132
RAB11FIP2	10	119774577	119774598	2	2	0	4.97E-05	0.721317	0.072132
TPTE2P6	13	25144872	25171521	3	1	0	7.42E-05	1	0.097851
TEAD4	12	3131189	3147203	2	2	0	0.000111	1	0.134593
PRAMEF2	1	12918946	12921581	14	12	1	0.000133	1	0.135938
KRTAP10-7	21	46020720	46021546	10	5	1	0.000134	1	0.135938
GBA	1	155205043	155210498	3	3	0	0.00014	1	0.135938

Cases, N=403; Controls, N=1650.

Key: GENE, gene symbol of the closest gene in the variant loci; CHR, chromosome; BEG & END, Gene genome position (Grch37) beginning and end; NUM_ALL_VARS, Number of all variants in gene locus; NUM_PASS_VARS, Variants that pass the filters; NUM_SING_VARS, Number of singleton variants; PVALUE, P value in burden variant analysis; P_BONF, Bonferroni corrected p value; P_BH, Benjamini & Hochberg corrected p value.

Table S18 Ten most significant p values in WES discovery dataset near (+/-1 Mbp) of the known PD risk loci.

MARKER_ID	TOT_MARKERS	PASS_MARKERS	P VALUE	FDR(BH)	Bonferroni
PPAPDC1A	3	1	0.0001064	0.0518168	0.0518168
STK19	2	2	0.0003437	0.08369095	0.1673819
HSPA1B	5	5	0.0007906	0.128340733	0.3850222
SPPL2C	22	8	0.001274	0.1551095	0.620438
PPT2	3	3	0.001898	0.1848652	0.924326
CORO1A	4	3	0.005918	0.480344333	1
MPHOSPH9	10	7	0.01132	0.689105	1
ZNF768	5	3	0.01054	0.689105	1
ACBD4	2	1	0.02869	0.85042375	1

Burden analysis with EMMAX test, using all the rare variants in target regions.

Key: TOT_MARKERS, total number of variants; PASS_MARKERS, variants that pass the filters; P VALUE, p value of the test; FDR(BH), Benjamini & Hochberg corrected p value; Bonferroni, Bonferroni corrected p value.

Table S19 Variant count and call rate in WES discovery dataset.

Gene	Snps	Rare_Snps	NonSyn_Snps	Rare_Nonsyn_snps	Callrate
ACMSD	1	1	1	1	1
BST1	7	5	4	3	0.989474
CCDC62	7	4	4	1	1
DDRGK1	4	2	2	0	1
FAM47E	9	5	7	4	0.998663
GAK	16	13	8	7	0.996335
GBA	7	6	4	3	0.996348
GCH1	5	3	1	0	0.992782
GPNMB	12	8	6	5	0.999749
HLA-DQB1	68	15	32	8	0.993764
INPP5F	9	7	3	2	1
LRRK2	21	8	12	7	0.99957
MAPT	20	7	7	2	0.996917
MCCC1	6	4	2	1	0.999499
RIT2	7	5	1	1	0.998067
SCARB2	6	6	1	1	0.992732
SIPA1L2	26	16	11	8	0.998554
STK39	2	2	1	1	1
SYT11	3	2	1	1	0.99599
TMEM175	4	2	2	1	1
VPS13C	32	26	19	15	0.999436

Selected genes are near loci with significant p values in meta-analysis.

Key: Gene, gene symbol; Snps, Total variant count; Rare_Snps, Minor allele frequency < 5%; NonSyn_Snps, total nonsynonymous variants; Rare_Nonsyn_snps, rare nonsynonymous variants (MAF<5%); Callrate, total variant missingness by call rate.

Table S20 Quality control values (GATK) of the WES discovery dataset top hits that were found also in replication dataset.

#CHROM	POS	ID	REF	ALT	GENE	BaseQ RankSum	Clipping RankSum	DP	FS	GQ_MEAN	GQ_STDDEV	Inbreeding Coeff	MLEAC	MLEAF	MQ	MQ0	MQ RankSum	NCC	QD	ReadPos RankSum	SOR	VQSLOD	culprit
2	71361156	rs143555311	C	G	MPHOSPH10	0.75	0.165	20171	3.785	68.14	149.75	-0.0016	2	1.21E-03	60	0	0.92	1	13.53	1.9	0.518	2.84	DP
2	97858622	rs201600563	C	A	ANKRD36	1.69	0.616	20554	45.101	62.16	12.01	-0.0018	2	1.21E-03	56.49	0	-2.9	1	1.02	1.31	7.275	-6.042	FS
6	142758601	.	T	G	GPR126	0.801	1.19	18767	3.21	62.19	23.84	-0.0039	2	1.21E-03	60	0	1	2	12.89	-0.607	1.06	1.77	DP
12	11174795	rs12424373	T	G	TAS2R19	7.45	0.068	77378	1.139	117.33	277.97	-0.0043	7	4.23E-03	59.33	0	1.91	1	14.84	1.03	0.58	0.532	MQRan
14	94845944	rs141620200	C	T,A	SERPINA1	-1.936	-0.156	19857	13.079	62.67	85.19	0.3235	30,1	0.021,6.046e-04	59.51	0	-0.49	1	7.61	-0.12	1.834	-0.7318	FS
14	94847262	rs17580	T	A	SERPINA1	-1.79	0.094	19885	3.25	77.45	139.66	-0.0082	13	7.86E-03	60	0	-0.053	1	13.63	1.45	0.914	3.62	DP
18	14105770	rs61730995	T	C	ZNF519	3.04	0.248	33735	0	103.18	251.6	0.0864	20	0.012	60	0	-0.128	2	14.78	1.94	0.658	3.01	FS

Table S21 Quality control values (GATK) of WES discovery dataset variants that were used in Gene-based analysis.

#CHR	POS	AC	AN	Cases	Controls	ExonicFunc.refGene	refGene	BaseQRankSum	ClippingRankSum	DP	FS	GQ_MEAN	GQ_STDDEV	Inbreeding Coeff	MLEAC	MLEAF	MQ	MQ0	MQRankSum	NCC	QD	ReadPosRankSum	SOR	VQSLOD	culprit
2	71361156	2	1654	0,1,1	0,1,1	nonsyn	MPHOSPH10	0.75	0.165	20171	3.785	68.14	149.75	-0.0016	2	1.21E-03	60	0	0.92	1	13.53	1.9	0.518	2.84	DP
2	97779601	8	1654	0,0,0	0,7,7	nonsyn	ANKRD36	-1.396	-0.195	21213	62.846	61.29	12.85	-0.0054	6	3.63E-03	57.25	0	-2.757	1	0.37	0.301	6.273	-8.954	FS
2	97779619	3	1654	0,0,0	0,3,3	nonsyn	ANKRD36	-2.526	-0.601	20212	0	62.29	29.8	-0.0022	3	1.81E-03	56.93	0	0.33	1	11.21	0.2	0.67	3.4	FS
2	97784177	4	1654	0,2,2	0,2,2	nonsyn	ANKRD36	2.42	-0.099	20575	1.125	67.94	63.02	-0.0043	4	2.42E-03	54.68	0	0.566	1	13.77	0.748	0.598	2.79	DP
2	97792829	2	1654	0,1,1	0,1,1	nonsyn	ANKRD36	3.11	-1	18702	17.532	61.85	9.83	-0.0014	2	1.21E-03	51.75	0	0.244	1	7.23	1.11	4.235	-1.427	FS
2	97808524	80	1654	0,57,57	0,16,16	nonsyn	ANKRD36	1.19	0.159	41646	21.099	89.79	84.18	-0.0545	78	0.047	58.6	0	-3.635	1	2.99	-2.261	5.435	-5.549	QD
2	97808562	9	1654	0,8,8	0,0,0	nonsyn	ANKRD36	2.86	-0.079	32269	6.486	74.15	26.32	-0.0084	7	4.23E-03	54.3	0	-2.898	1	0.73	0.031	1.781	-3.361	QD
2	97812240	2	1656	0,1,1	0,1,1	nonsyn	ANKRD36	0.278	0.896	22626	66.534	67.48	24.04	-0.0017	2	1.21E-03	53.35	0	-1.246	0	2.73	-0.597	4.231	-10.32	FS
2	97812249	2	1656	0,1,1	0,1,1	nonsyn	ANKRD36	5.05	1.2	22507	57.555	66.39	16.52	-0.0014	2	1.21E-03	54.6	0	-1.683	0	0.85	-1.203	6.184	-8.875	FS
2	97817670	69	1656	0,25,25	0,40,40	nonsyn	ANKRD36	1.79	0.012	142344	2.745	96.87	75.88	-0.0554	70	0.042	52.54	0	-3.613	0	1.51	-3.119	1.418	-7.829	QD
2	97818249	4	1648	0,2,2	0,2,2	nonsyn	ANKRD36	2.52	0.245	161009	53.737	108.77	57.47	-0.003	4	2.43E-03	37.42	0	3.11	4	3.61	1.06	2.127	-5.109	FS
2	97820422	7	1624	0,1,1	0,6,6	nonsyn	ANKRD36	0.318	-0.151	120383	1.767	88.98	39.53	-0.0137	3	1.85E-03	40.41	0	-0.635	16	0.88	-0.387	1.761	-2.527	QD
2	97823853	11	1654	0,3,3	0,8,8	nonsyn	ANKRD36	-1.123	0.463	22992	0	98	397.19	-0.0074	11	6.65E-03	59.21	0	0.771	1	13.63	0.279	0.727	3.26	FS
2	97823903	34	1654	0,32,32	0,0,0	nonsyn	ANKRD36	3.17	0.137	22402	36.784	58.11	22.41	-0.0341	29	0.018	56.51	0	-2.784	1	0.74	1.32	5.134	-3.211	FS
2	97827852	23	1656	0,7,7	0,16,16	nonsyn	ANKRD36	-1.505	-0.105	68813	1.596	101.49	32.35	-0.0155	21	0.013	44.84	0	0.067	0	0.88	-0.043	0.991	-2.214	QD
2	97830031	8	1652	0,5,5	0,3,3	stopgain	ANKRD36	-2.812	0.014	150652	0	104.58	29.38	-0.0062	4	2.42E-03	31.22	0	-1.472	2	0.13	0.502	1.006	-2.461	DP

2	97830177	27	1654	0,19,19	0,7,7	nonsyn	ANKRD36	-2.133	0.197	134601	0	104.86	31.32	-0.018	21	0.013	52.61	0	-3.766	1	0.69	-0.973	0.783	-3.807	MQRankSum
2	97847488	13	1654	0,5,5	0,8,8	nonsyn	ANKRD36	0.255	-0.162	23456	0	91.98	244.97	-0.0079	13	7.86E-03	44.31	0	-0.773	1	12.96	1.24	0.664	1.97	FS
2	97849319	3	1656	0,1,1	0,2,2	nonsyn	ANKRD36	-0.659	0.103	25849	0	75.2	150.36	-0.002	3	1.81E-03	55.24	0	1.23	0	13.34	0.373	0.699	2.66	FS
2	97851200	46	1654	0,35,35	0,7,7	nonsyn	ANKRD36	-1.274	-0.056	124476	27.061	67.43	62.27	-0.1	53	0.032	45.8	0	-3.859	1	1.03	-0.963	7.114	-4.44	FS
2	97851235	9	1650	0,8,8	0,1,1	nonsyn	ANKRD36	-0.898	0.274	110150	9.672	85.15	43.5	-0.019	5	3.03E-03	52.13	0	-1.734	3	0.31	-1.734	5.687	-4.238	QD
2	97852956	18	1656	0,5,5	0,13,13	nonsyn	ANKRD36	-0.755	0.019	27719	0.535	107.11	328.16	-0.0112	18	0.011	48.19	0	-0.417	0	11.62	1.32	0.778	1.92	FS
2	97858622	2	1654	0,1,1	0,1,1	nonsyn	ANKRD36	1.69	0.616	20554	45.101	62.16	12.01	-0.0018	2	1.21E-03	56.49	0	-2.9	1	1.02	1.31	7.275	-6.042	FS
2	97860471	18	1654	0,5,5	0,13,13	nonsyn	ANKRD36	3.95	0.589	46277	9.326	173.3	644.2	-0.011	18	0.011	41.58	0	0.045	1	13.81	1.32	0.682	-1.163	FS
2	97867947	8	1654	0,3,3	0,5,5	nonsyn	ANKRD36	1.09	0.135	35499	1.572	74.22	106.08	-0.0081	8	4.84E-03	58.49	0	-3.092	1	3.49	-0.65	0.599	-3.291	QD
2	97867956	7	1654	0,3,3	0,4,4	nonsyn	ANKRD36	-0.907	0.194	35405	6.237	72.3	75.46	-0.0073	7	4.23E-03	58.58	0	-2.49	1	3.04	-2.089	1.399	-5.386	QD
2	97867960	7	1654	0,3,3	0,4,4	nonsyn	ANKRD36	2.71	0.425	35318	6.653	71.43	65.47	-0.0073	7	4.23E-03	58.59	0	-2.49	1	2.41	-2.428	1.531	-6.246	QD
2	97868086	2	1656	0,1,1	0,1,1	nonsyn	ANKRD36	-2.253	1.08	29740	0	71.59	72.78	-0.0013	2	1.21E-03	55.29	0	0.398	0	4.44	1.99	0.739	0.873	FS
2	97869931	10	1654	0,10,10	0,0,0	nonsyn	ANKRD36	-2.994	0.162	28196	0	76.68	22.9	-0.0067	7	4.23E-03	56.93	0	-4.304	1	0.32	2.48	0.736	-5.401	MQRankSum
2	97875561	3	1656	0,0,0	0,3,3	nonsyn	ANKRD36	1.7	0.43	24549	1.175	76.94	201.89	-0.0019	3	1.81E-03	59.45	0	0.698	0	9.38	0.241	0.792	1.11	DP
2	97899493	14	1642	0,1,1	0,13,13	nonsyn	ANKRD36	1.21	0.192	17770	0	61.55	56.5	-0.0115	14	8.53E-03	32.14	0	0.694	7	4.91	0.64	0.34	1.87	FS
2	97909592	40	1654	0,35,35	0,2,2	nonsyn	ANKRD36	4.8	0.109	30566	15.798	69.06	28.08	-0.0302	36	0.022	33.91	0	-0.989	1	0.83	0.82	8.655	-3.127	FS
2	97909711	4	1652	0,4,4	0,0,0	nonsyn	ANKRD36	2.22	-0.786	105102	33.206	93.41	30.35	-0.0025	2	1.21E-03	29.7	0	0.029	2	0.19	1.41	4.775	-2.975	FS
2	97911105	7	1654	0,3,3	0,2,2	nonsyn	ANKRD36	-0.771	1.24	19787	27.765	63.89	48.16	-0.0044	7	4.23E-03	26.87	0	0.667	1	3.3	0.821	7.186	-3.508	FS
5	1.8E+08	3	1654	0,0,0	0,3,3	nonsyn	CNOT6	-2.073	-0.296	18995	1.158	72.51	277.99	-0.0034	3	1.81E-03	60	0	0.205	1	15.3	3.93	0.584	0.731	ReadPosRankSum
5	1.8E+08	3	1654	0,1,1	0,2,2	nonsyn	CNOT6	-2.529	-0.128	19871	1.651	63.67	92.42	-0.0023	3	1.81E-03	60	0	-1.006	1	12.85	1.54	0.576	1.83	DP
6	1.43E+08	2	1648	0,1,1	0,1,1	nonsyn	GPR126	3.89	1.31	16745	4.376	55.12	76.86	-0.0054	2	1.21E-03	60	0	1.1	4	14.34	1.5	0.786	1.42	DP
6	1.43E+08	9	1654	0,6,6	0,3,3	nonsyn	GPR126	-0.587	-0.129	19217	0	71.19	154.26	-0.0057	9	5.44E-03	60	0	-0.092	1	13.76	0.556	0.749	5.63	FS
6	1.43E+08	2	1652	0,2,2	0,0,0	nonsyn	GPR126	0.801	1.19	18767	3.21	62.19	23.84	-0.0039	2	1.21E-03	60	0	1	2	12.89	-0.607	1.06	1.77	DP
7	1855776	5	1654	0,0,0	0,5,5	nonsyn	MAD1L1	0.906	0.532	16462	8.432	53.55	60.36	-0.0088	5	3.02E-03	53.86	0	-0.086	1	14.03	0.775	0.396	1.24	FS
7	2265161	8	1656	1,2,4	0,4,4	nonsyn	MAD1L1	1.98	-0.1	19178	8.37	64.87	70.54	0.2263	8	4.83E-03	60	0	-0.211	0	16.14	0.742	0.349	2.08	DP
7	64292303	2	1654	0,2,2	0,0,0	nonsyn	ZNF138	-7.024	0.339	32377	1.789	75.09	101.38	-0.0017	2	1.21E-03	60	0	0.212	1	11.86	0.591	0.8	2.15	QD
8	52732958	31	1654	0,26,26	0,1,1	nonsyn	PCMTD1	-0.188	0.546	33179	34.837	81.51	27.44	-0.0209	29	0.018	54.98	0	-2.038	1	1.54	1.07	6.933	-4.213	FS
8	52732960	2	1654	0,2,2	0,0,0	nonsyn	PCMTD1	-0.279	-0.465	34059	5.991	85	24.26	-0.0014	1	6.05E-04	54.31	0	-2.265	1	1.35	0.093	4.111	-3.182	QD
8	52732961	31	1654	0,18,18	0,8,8	nonsyn	PCMTD1	-1.997	0.351	35923	23.001	78.33	36.12	-0.0267	29	0.018	54.8	0	-3.72	1	1.21	1.6	8.594	-3.837	QD
8	52732981	32	1654	0,22,22	0,5,5	nonsyn	PCMTD1	1.62	-0.211	35998	16.797	80.7	28.55	-0.0209	24	0.015	55.28	0	-2.483	1	0.82	-0.269	9.527	-3.183	FS
8	52733003	86	1654	0,52,52	0,27,27	nonsyn	PCMTD1	-1.491	0.138	43667	10.402	72.56	46.22	-0.0807	88	0.053	57.58	0	-3.683	1	1.59	-1.99	1.302	-4.937	QD
8	52733023	7	1654	0,7,7	0,0,0	nonsyn	PCMTD1	0.83	-0.38	37575	9.921	83.06	27.57	-0.0072	4	2.42E-03	59.26	0	-2.412	1	1.92	-1.924	4.29	-5.164	QD
8	52733079	4	1654	0,4,4	0,0,0	nonsyn	PCMTD1	2.66	-0.021	200108	1.934	65.04	51.4	-0.0832	2	1.21E-03	59.1	0	-1.59	1	0.43	1.99	0.233	-3.369	DP
8	52733110	28	1654	0,26,26	0,1,1	nonsyn	PCMTD1	-1.212	0.077	195411	2.391	91.51	42.17	-0.0275	21	0.013	53.89	0	-1.569	1	0.74	-0.258	1.58	-2.811	DP

8	52733143	31	1652	0,21,21	0,5,5	nonsyn	PCMTD1	0.812	-0.043	176383	0	94.92	40.05	-0.0248	21	0.013	49.79	0	-1.68	2	0.57	1.21	0.395	-2.684	DP
8	52733144	77	1652	0,68,68	0,3,3	nonsyn	PCMTD1	-2.063	0.102	176298	1.651	91.6	43.64	-0.0592	62	0.038	49.3	0	-1.379	2	0.95	0.974	1.291	-2.464	QD
8	52733164	59	1652	1,39,41	0,15,15	nonsyn	PCMTD1	-1.913	0.042	111501	0	90.6	45.68	-0.0188	46	0.028	48.85	0	-1.14	2	0.71	0.256	1.026	-1.113	QD
9	1.36E+08	16	1656	0,0,0	0,16,16	nonsyn	CEL	-1.291	0.913	20642	19.575	59.25	17.03	-0.0113	13	7.85E-03	58.53	0	-2.777	0	0.9	-1.671	8.575	-5.398	FS
9	1.36E+08	21	1656	0,0,0	0,21,21	nonsyn	CEL	-1.06	0.344	21309	34.505	58.57	23.06	-0.0184	17	0.01	57.88	0	-3.317	0	1.19	-1.363	6.163	-5.788	FS
9	1.36E+08	63	1654	0,2,2	0,59,59	nonsyn	CEL	1.28	0.222	23641	21.579	56.25	57.35	0.3311	65	0.039	57.77	0	-3.338	1	2.15	-0.273	2.435	-3.78	QD
9	1.36E+08	2,2	1656	1,1,3	0,1,1	nonsyn	CEL	5.33	1.18	16792	1.134	55.24	72.14	0.2585	2,2	1.208e-03,1.208e-03	44.63	0	1.2	0	25.57	0	0.588	-1.084	DP
11	476442619	2	1654	0,2,2	0,0,0	nonsyn	MTCH2	-5.004	1.43	91259	3.899	96.69	54.78	-0.0013	2	1.21E-03	59.72	0	1.41	1	9.18	2.32	0.438	-1.196	ReadPosRankSum
11	476472378	72	1654	0,64,64	0,0,0	nonsyn	MTCH2	0.49	0.11	83788	11.955	49.26	51.15	-0.1344	96	0.058	58.85	0	-1.733	1	2.8	-1.386	2.181	-1.901	QD
11	476472655	66	1656	0,60,60	0,0,0	nonsyn	MTCH2	1.59	0.06	84812	15.426	52.61	50.41	-0.1184	73	0.044	59.04	0	-1.733	0	2.89	0.318	2.818	-1.951	QD
11	476603552	42	1644	0,15,15	0,24,24	nonsyn	MTCH2	-0.521	0	38649	7.763	44.63	86.49	-0.093	48	0.029	59.63	0	-0.722	6	2.92	-1.486	0.57	-4.046	QD
11	476603544	38	1644	0,15,15	0,21,21	nonsyn	MTCH2	0.603	-0.116	38101	9.214	42.14	78.14	-0.0975	46	0.028	59.63	0	-1.313	6	2.99	-1.571	0.771	-4.209	QD
11	476603577	7	1654	0,6,6	0,1,1	nonsyn	MTCH2	-0.29	-0.238	35676	9.242	71.95	31.91	-0.0053	6	3.63E-03	60	0	-0.276	1	2.33	-1.936	2.227	-5.158	QD
11	476605619	3	1652	0,1,1	0,1,1	nonsyn	MTCH2	2.52	0.747	24361	0	56.69	25.4	-0.0253	3	1.82E-03	60	0	0.33	2	6.35	-1.809	0.708	-0.8817	ReadPosRankSum
12	111744978	7	1654	0,3,3	0,4,4	nonsyn	TAS2R19	8.85	0.968	153777	1.114	127.63	389.66	-0.0048	7	4.23E-03	59.52	0	1.26	1	14.92	0.352	0.595	-0.7866	DP
12	111745324	16	1654	0,2,2	0,13,13	nonsyn	TAS2R19	2.53	-0.202	146208	11.695	95.03	31.51	-0.012	13	7.86E-03	58.74	0	-7.132	1	0.44	3.17	9.177	-8.169	MQRankSum
12	111745422	12	1654	0,2,2	0,9,9	nonsyn	TAS2R19	3.65	-0.136	145785	20.726	96.23	29.31	-0.0089	11	6.65E-03	58.7	0	-3.384	1	0.38	2.27	10.03	-4.626	FS
12	111745423	2	1654	0,1,1	0,1,1	nonsyn	TAS2R19	3.38	0.451	145650	6.164	104.35	141.88	-0.0012	2	1.21E-03	59.03	0	1.36	1	12.89	0.187	0.462	-1.029	DP
12	111745488	8	1654	0,1,1	0,6,6	nonsyn	TAS2R19	-1.133	-0.427	145730	15.104	97.09	28.42	-0.0062	7	4.23E-03	58.6	0	-3.765	1	0.55	1.96	9.08	-4.123	MQRankSum
12	111745666	7	1654	0,0,0	0,6,6	nonsyn	TAS2R19	0.137	0.193	145559	31.859	97.54	32.9	0.5581	7	4.23E-03	58.43	0	-4.06	1	0.97	-1.17	7.704	-5.425	FS
12	111745766	5	1654	0,0,0	0,4,4	nonsyn	TAS2R19	3.46	0.613	145549	35.005	97.97	30.66	-0.0033	4	2.42E-03	58.78	0	-3.36	1	1.1	-1.605	7.04	-5.967	FS
12	111745999	5	1654	0,4,4	0,0,0	nonsyn	TAS2R19	-0.976	-0.538	145774	11.869	97.15	26.07	-0.004	2	1.21E-03	59.05	0	-3.719	1	0.09	-0.538	6.715	-3.578	MQRankSum
12	111746045	5	1654	0,4,4	0,0,0	nonsyn	TAS2R19	-0.896	-0.498	145837	17.192	97.33	25.78	-0.0045	5	3.02E-03	58.85	0	-4.047	1	0.39	-0.538	7.132	-3.873	MQRankSum
12	111746111	10	1654	0,7,7	0,1,1	nonsyn	TAS2R19	1.85	0.458	146657	31.329	97.15	27.44	-0.0081	8	4.84E-03	58.69	0	-5.452	1	0.53	-1.55	7.848	-6.719	FS
12	111746144	9	1654	0,7,7	0,0,0	nonsyn	TAS2R19	3.65	1.06	146668	31.056	98.81	25.47	-0.0055	9	5.44E-03	59.28	0	-2.41	1	0.77	-0.418	8.852	-3.949	FS
12	111746189	9	1654	0,7,7	0,0,0	nonsyn	TAS2R19	1.52	0.685	146945	33.159	96.86	29.12	-1.8392	9	5.44E-03	58.28	0	-5.886	1	0.93	-1.442	8.884	-7.001	FS
12	111746326	12	1656	0,10,10	0,0,0	nonsyn	TAS2R19	-4.544	-0.061	146930	29.659	98	31.27	-0.0097	12	7.25E-03	58.28	0	-6.545	0	1.14	1.34	6.725	-6.32	FS
12	111746422	13	1656	0,11,11	0,0,0	nonsyn	TAS2R19	0.15	-0.345	146740	29.053	97.63	29.84	-0.0106	13	7.85E-03	58.3	0	-6.415	0	0.95	1.84	6.709	-6.652	FS
12	111746444	13	1656	0,11,11	0,0,0	nonsyn	TAS2R19	-4.157	0.371	146131	26.09	98.61	29.66	0.0136	13	7.85E-03	58.24	0	-6.225	0	1.1	1.97	6.227	-6.359	MQRankSum
12	111746722	11	1656	0,11,11	0,0,0	nonsyn	TAS2R19	-4.124	0.125	145751	29.053	96.54	26.83	-0.0077	9	5.44E-03	57.35	0	-5.12	0	0.16	-0.459	6.664	-5.191	FS
12	111747155	10	1654	0,6,6	0,4,4	nonsyn	TAS2R19	2.12	0.772	144838	2.512	123.3	339.17	-0.0061	7	4.23E-03	57.05	0	-2.914	1	12.69	0.281	0.535	-1.457	MQRankSum
12	111747635	3	1654	0,2,2	0,0,0	nonsyn	TAS2R19	-2.038	-0.047	78631	0	93.7	30.05	-0.0035	2	1.21E-03	57.97	0	-2.408	1	0.12	1.14	3.379	-2.566	MQRankSum
12	111747886	8	1654	0,6,6	0,0,0	nonsyn	TAS2R19	0.178	-0.116	78872	2.857	93.49	30.25	-0.0063	7	4.23E-03	58.56	0	-2.862	1	0.64	1.24	2.194	-2.6	MQRankSum

1	11174795	7	1654	0,1,1	0,5,5	nonsyn	TAS2R19	7.45	0.068	77378	1.139	117.33	277.97	-0.0043	7	4.23E-03	59.33	0	1.91	1	14.84	1.03	0.58	0.532	MQRankSum
1	11174845	43	1654	0,12,12	0,29,29	nonsyn	TAS2R19	-1.253	-0.062	76870	0.531	190.98	488.22	-0.0267	43	0.026	59.87	0	0.295	1	12.31	0.729	0.775	1.97	FS
1	11175041	5	1654	0,4,4	0,1,1	nonsyn	TAS2R19	-1.737	-0.026	73140	0	89.54	33.71	-0.006	4	2.42E-03	59.7	0	-2.005	1	1.38	-0.712	0.659	-2.84	QD
1	11175055	36	1654	0,25,25	0,5,5	nonsyn	TAS2R19	-1.627	0.04	74572	0	91.54	35.26	-0.024	31	0.019	59.2	0	-2.372	1	1.53	-0.625	0.687	-2.801	QD
1	11175060	36	1654	0,25,25	0,5,5	nonsyn	TAS2R19	-1.832	0.058	75044	0	91.73	35.57	-0.0249	33	0.02	59.34	0	-2.398	1	1.42	-0.347	0.594	-2.597	QD
1	11175073	46	1656	0,35,35	0,7,7	nonsyn	TAS2R19	-1.732	-0.087	75439	0	92.53	35.68	-0.029	36	0.022	59.1	0	-2.377	0	1.28	0.941	0.743	-2.357	QD
1	11175083	28	1656	0,17,17	0,7,7	nonsyn	TAS2R19	-1.732	0.088	75952	0	94.19	31.55	-0.0172	22	0.013	58.85	0	-2.367	0	0.72	1.25	0.795	-2.54	QD
1	11214001	19	1654	0,15,15	0,1,1	nonsyn	TAS2R46	-2.124	-0.227	27074	2.198	69.67	23.71	-0.0138	19	0.011	58.7	0	-4.353	1	1.47	-0.032	3.604	-3.731	QD
1	11214005	29	1654	0,22,22	0,3,3	nonsyn	TAS2R46	-2.64	-0.167	30364	9.335	64.97	28.71	-0.0318	28	0.017	58.42	0	-6.017	1	0.92	-0.732	3.396	-4.098	MQRankSum
1	11214006	32	1654	0,23,23	0,5,5	nonsyn	TAS2R46	-2.026	-0.308	31067	11.823	65.13	29.02	-0.0335	29	0.018	58.67	0	-6.023	1	0.79	-0.944	3.815	-4.295	MQRankSum
1	11214025	58	1654	0,28,28	0,24,24	nonsyn	TAS2R46	-0.848	-0.162	52357	14.685	66.68	50.01	-0.0646	62	0.037	58.39	0	-8.22	1	0.75	-0.206	3.179	-5.115	MQRankSum
1	11214182	13	1654	0,13,13	0,0,0	nonsyn	TAS2R46	-1.369	0.179	147365	9.14	96.52	33.46	0.4332	11	6.65E-03	58.87	0	-3.322	1	1.13	-0.769	3.888	-3.466	QD
1	11214191	31	1654	0,26,26	0,2,2	nonsyn	TAS2R46	-0.502	0.118	147396	14.905	95.09	36.61	-0.0219	26	0.016	58.95	0	-3.767	1	1.02	-0.976	3.638	-4.012	QD
1	11214199	36	1654	0,28,28	0,4,4	nonsyn	TAS2R46	-0.946	0.195	147190	17.459	94.59	34.09	-0.0238	31	0.019	59.06	0	-3.738	1	0.66	-1.286	3.871	-4.511	MQRankSum
1	11214214	5	1650	0,4,4	0,0,0	nonsyn	TAS2R46	-1.75	0.702	166278	3.027	104.86	24.44	-0.0031	3	1.82E-03	59.35	0	-2.419	3	0.19	0.11	4.796	-2.789	DP
1	11214231	8	1654	0,8,8	0,0,0	nonsyn	TAS2R46	1.03	-0.223	167365	0	103.42	26.75	-0.0063	7	4.23E-03	59.08	0	-2.892	1	0.32	-0.075	0.666	-3.005	DP
1	11214232	8	1654	0,7,7	0,0,0	nonsyn	TAS2R46	1.11	0.281	167458	0	103.54	26.44	-0.0066	7	4.23E-03	59.08	0	-2.892	1	0.52	-0.469	0.688	-3.221	DP
1	11214264	34	1654	0,24,24	0,8,8	nonsyn	TAS2R46	-1.095	0.152	168864	2.406	107.92	42.02	-0.0209	29	0.018	58.17	0	-4.008	1	0.89	0.273	1.203	-3.231	MQRankSum
1	11214498	57	1624	0,7,7	0,49,49	nonsyn	TAS2R46	2.33	0.038	53693	10.928	74.21	71.55	-0.0469	52	0.032	58.94	0	-4.172	16	0.96	-2.848	3.302	-6.773	MQRankSum
1	11214500	22	1624	0,2,2	0,20,20	nonsyn	TAS2R46	0.779	-0.045	38377	13.801	72.72	45.03	-0.0161	20	0.012	58.96	0	-3.103	16	1.02	-2.433	3.705	-6.29	QD
1	11214509	19	1654	0,5,5	0,13,13	nonsyn	TAS2R46	-0.397	-0.047	36051	17.94	66.23	27.72	-0.0205	19	0.011	58.03	0	-3.962	1	0.69	-2.423	7.201	-6.47	MQRankSum
1	11214535	36	1654	0,32,32	0,0,0	nonsyn	TAS2R46	0.066	0.391	33145	24.773	61.4	25.03	-0.0357	37	0.022	52.51	0	-3.663	1	0.94	0.284	3.037	-3.541	FS
1	11214601	19	1654	0,17,17	0,1,1	nonsyn	TAS2R46	0.3	0.346	27749	8.65	62.08	16.04	-0.0181	17	0.01	55.44	0	-3.528	1	0.59	-1.858	1.787	-5.64	MQRankSum
1	11214796	4	1654	0,4,4	0,0,0	nonsyn	TAS2R46	1	-0.456	23191	16.209	63.16	12.62	-0.0035	3	1.81E-03	57.77	0	-2.793	1	0.56	0.175	3.098	-3.45	FS
1	11214857	29	1652	0,28,28	0,0,0	nonsyn	TAS2R46	2.58	0.417	22766	20.548	58.5	17.27	-0.0241	28	0.017	55.09	0	-3.048	2	1.33	0.242	4.569	-3.566	QD
1	53343158	87	1628	0,64,64	0,17,17	nonsyn	KRT18	1.19	0.134	18787	64.64	65.25	144.36	-0.1109	102	0.063	57.7	0	-2.747	14	4.96	0.274	6.563	-9.414	FS
1	53343318	53	1650	0,42,42	0,5,5	nonsyn	KRT18	-0.964	0.206	12946	13.012	31.38	26.93	-0.0657	58	0.035	56.74	0	-2.784	3	2.22	0.257	3.955	-3.246	QD
1	53343322	52	1648	0,41,41	0,5,5	nonsyn	KRT18	-2.308	0.087	12497	13.108	30.84	26.15	-0.0635	55	0.033	57.01	0	-2.793	4	2.27	0.351	3.669	-3.209	QD
1	53343325	51	1648	0,40,40	0,5,5	nonsyn	KRT18	-0.799	0.079	12279	10.867	29.58	25.79	-0.0662	55	0.033	56.74	0	-2.775	4	2.23	0.026	3.411	-3.297	QD
1	53343352	5	1646	0,3,3	0,1,1	nonsyn	KRT18	0.694	-0.417	10928	0	30.66	18.96	-0.02	3	1.82E-03	57.51	0	-2.302	5	0.33	-1.826	0.498	-3.95	MQRankSum
1	94844947	37	1656	0,10,10	1,23,25	nonsyn	SERPINA1	3.69	0.35	30612	1.155	165.52	608.15	0.0312	37	0.022	60	0	0.388	0	15.1	0.752	0.829	3.93	DP
1	94845850	49	1654	0,33,33	0,13,13	nonsyn	SERPINA1	-0.544	-0.02	22816	3.807	67.61	57.52	-0.0392	49	0.03	59.87	0	-0.068	1	5.52	-0.38	0.375	-0.6819	DP
1	94845851	15	1654	0,9,9	0,4,4	nonsyn	SERPINA1	-0.479	0.675	22760	2.12	65.79	27.28	-0.0102	13	7.86E-03	59.85	0	-0.049	1	3.01	0.289	0.472	-0.5322	QD

14	94845863	47	1654	0,33,33	0,11,11	nonsyn	SERPINA1	-0.648	0.019	22632	0.793	67.22	52.23	-0.0364	47	0.028	59.72	0	-0.047	1	4.86	-1.479	0.566	-1.555	ReadPosRankSum
14	94845891	33	1654	0,27,27	0,2,2	nonsyn	SERPINA1	-0.743	-0.248	217866	4.176	64.47	46.73	-0.0261	32	0.019	59.95	0	-0.659	1	5.57	0.818	1.267	-0.6081	DP
14	94845900	35	1654	0,28,28	0,2,2	nonsyn	SERPINA1	0.441	-0.331	21244	12.882	65.25	42.24	-0.025	35	0.021	59.62	0	-0.829	1	5.07	1.07	2.147	-2.231	FS
14	94845902	35	1654	0,28,28	0,2,2	nonsyn	SERPINA1	-1.459	-0.118	21208	15.983	64.49	39.23	-0.0258	35	0.021	59.66	0	-0.65	1	4.59	0.693	2.577	-2.685	FS
14	94845914	23	1654	0,18,18	0,3,3	nonsyn	SERPINA1	0.981	0	20858	20.781	59.27	18.68	-0.0189	20	0.012	59.95	0	-0.6	1	2.27	-0.066	4.895	-3.609	QD
14	94845930	36	1654	0,29,29	0,3,3	nonsyn	SERPINA1	1.33	-0.075	20481	40.998	62.01	37.11	-0.0278	35	0.021	59.51	0	-1.104	1	5.75	0.06	6.161	-4.257	FS
14	94845940	34	1654	0,28,28	0,2,2	nonsyn	SERPINA1	-1.66	0.057	19950	41.033	60.73	38.99	-0.027	34	0.021	59.33	0	-0.65	1	6.41	-0.233	6.786	-4.047	FS
14	94845943	34	1654	0,28,28	0,2,2	nonsyn	SERPINA1	-1.552	-1.335	19756	43.719	60.32	39.51	-0.0274	34	0.021	59.28	0	-0.38	1	6.58	-0.193	6.827	-4.287	FS
14	94845944	34,1	1654	0,28,28	0,3,3	nonsyn	SERPINA1	-1.936	-0.156	19857	13.079	62.67	85.19	0.3235	34,1	0.021,6.046e-04	59.51	0	-0.49	1	7.61	-0.12	1.834	-0.7318	FS
14	94847262	13	1654	0,4,4	0,9,9	nonsyn	SERPINA1	-1.79	0.094	19885	3.25	77.45	139.66	-0.0082	13	7.86E-03	60	0	-0.053	1	13.63	1.45	0.914	3.62	DP
17	37357494	26,2	1654	0,20,20	0,8,8	nonsyn	RPL19	1.53	0.162	24273	0	57.69	20.36	-0.0108	22,2	0.013,1.209e-03	57.5	0	-2.884	1	1.17	0.919	0.688	-3.835	QD
17	37358581	7	1654	0,0,0	0,7,7	nonsyn	RPL19	0.453	0	18736	0	54.54	19.35	-0.0095	3	1.81E-03	58.17	0	-1.732	1	1.09	-1.36	1.633	-3.608	QD
17	37358612	38	1656	0,2,2	0,35,35	nonsyn	RPL19	1.12	-0.031	23018	0	59.8	32.14	-0.0345	29	0.018	54.79	0	-2.101	0	1.48	-0.413	0.258	-3.63	QD
17	37358654	11	1656	0,1,1	0,10,10	nonsyn	RPL19	-0.682	0.042	25569	0	66.14	22.34	-0.007	5	3.02E-03	53.68	0	-1.733	0	0.48	0.866	0.745	-3.571	QD
17	37358685	8	1656	0,1,1	0,7,7	nonsyn	RPL19	0.048	-0.134	24518	0	62.72	23.02	-0.008	3	1.81E-03	56.86	0	-1.733	0	0.58	-1.329	0.262	-4.122	QD
17	37360364	88	1654	0,62,62	0,20,20	nonsyn	RPL19	-3.217	0.067	23905	44.237	75.8	76.6	-0.0575	87	0.053	56.77	0	-4.587	1	3.35	1.02	3.433	-6.08	FS
17	37360398	85	1654	0,64,64	0,15,15	nonsyn	RPL19	-1.197	0.153	24437	39.604	72.96	76.58	-0.056	84	0.051	56.77	0	-4.34	1	3.01	1.46	3.466	-5.699	FS
17	37360404	85	1654	0,63,63	0,15,15	nonsyn	RPL19	-2.155	-0.204	24141	37.409	72.52	74.98	-0.0547	83	0.05	56.71	0	-4.343	1	2.86	0.487	3.177	-4.934	FS
17	37360424	57	1654	0,49,49	0,5,5	nonsyn	RPL19	0.691	-0.107	22690	37.515	62.78	38.73	-0.0369	52	0.031	56.61	0	-3.728	1	2.1	-2.041	3.378	-7.318	QD
18	14104957	3	1654	0,1,1	0,2,2	nonsyn	ZNF519	0.051	0.32	20823	1.132	70.97	170.58	-0.0023	3	1.81E-03	60	0	0.138	1	15.16	1.26	0.826	2.59	DP
18	14105770	20	1652	1,6,8	0,10,10	nonsyn	ZNF519	3.04	0.248	33735	0	103.18	251.6	0.0864	20	0.012	60	0	-0.128	2	14.78	1.94	0.658	3.01	FS
18	14124411	12	1656	0,11,11	0,0,0	nonsyn	ZNF519	0.595	0.38	23857	60.358	60.35	20.2	-0.0121	12	7.25E-03	58.09	0	-2.175	0	1.42	1.53	8.254	-9.01	FS
18	14124459	19	1656	0,17,17	0,1,1	nonsyn	ZNF519	-1.315	-0.057	20403	31.411	51.02	27.85	-0.2289	18	0.011	57.89	0	-2.823	0	1.95	-0.376	7.442	-4.453	QD
19	33372814	27	1656	0,6,6	0,21,21	nonsyn	CEP89	1.78	0.724	19672	0	78.82	110.47	-0.0167	27	0.016	60	0	0.086	0	13.16	1.2	0.718	5.47	FS
19	33378696	6	1656	0,3,3	0,3,3	nonsyn	CEP89	2.82	0.767	20217	1.746	78.69	238.25	-0.004	6	3.62E-03	60	0	0.627	0	13.87	1.69	0.547	2.76	DP
19	33406311	4	1654	0,1,1	0,3,3	nonsyn	CEP89	1.18	1.53	60903	1.801	112.78	251.08	-0.0027	4	2.42E-03	60	0	1.18	1	12.21	0.844	0.576	0.167	MQRankSum
19	33445588	68	1598	0,4,4	0,64,64	nonsyn	CEP89	1.6	0.34	137358	3.355	76.58	79.5	-0.1163	73	0.046	60	0	-0.063	29	4.63	-3.172	0.915	-7.18	ReadPosRankSum
19	33450873	2	1656	0,2,2	0,0,0	nonsyn	CEP89	0.144	0.914	18691	2.135	61.33	23.49	-0.0016	2	1.21E-03	60	0	1.49	0	11.87	2.13	0.982	2.08	DP
20	29623215	20	1654	0,12,12	0,7,7	.	FRG1B	2.93	0.045	31738	0	93.48	45	-0.0492	18	0.011	41.07	0	-8.616	1	0.27	-0.362	0.652	-7.784	MQRankSum
20	29623217	18	1654	0,8,8	0,10,10	.	FRG1B	-3.86	-0.125	31720	0	108.05	35.47	-0.0156	13	7.86E-03	41.33	0	-1.404	1	0.34	0.497	0.906	-3.053	DP
20	29623218	4	1654	0,2,2	0,1,1	.	FRG1B	-1.917	0.065	31706	0	112.09	20.78	-0.0032	2	1.21E-03	41.09	0	-2.524	1	0.08	1.24	0.71	-3.677	DP
20	29623231	14	1654	0,4,4	0,8,8	.	FRG1B	-1.075	-0.097	26845	1.222	113.84	29.01	-0.0086	12	7.26E-03	41.06	0	2.06	1	0.36	-0.748	0.868	-3.247	DP
20	29625934	20	1652	0,17,17	0,2,2	.	FRG1B	-0.243	0.575	121186	0	9.04	31.89	-0.2995	101	0.061	51.49	0	-2.23	2	1.04	-0.918	1.072	-3.256	QD

20	29625941	3	1654	0,2,2	0,0,0	.	FRG1B	0.107	0.095	112948	2.992	32.88	43.52	-0.1613	2	1.21E-03	44.49	0	-1.826	1	0.62	0.49	1.682	-2.301	QD
20	29625956	3	1654	0,1,1	0,2,2	.	FRG1B	-1.504	0.142	106803	0	99.22	34.23	-0.0125	2	1.21E-03	48.57	0	-1.506	1	0.41	0.09	0.459	-2.284	QD
20	29625961	4	1654	0,3,3	0,0,0	.	FRG1B	-1.555	0.473	106686	2.541	104.25	28.58	-0.0063	2	1.21E-03	53.32	0	-1.678	1	1.01	-0.673	1.706	-2.808	QD
20	29625971	6	1656	0,2,2	0,3,3	.	FRG1B	-1.149	0.195	110867	1.583	80.94	46.42	-0.0433	5	3.02E-03	48.5	0	-1.328	0	0.52	1.92	0.381	-2.923	QD
20	29628233	4	1656	0,2,2	0,1,1	.	FRG1B	3.02	-0.339	369523	1.608	114.17	17.04	-0.0024	2	1.21E-03	46.86	0	-0.016	0	0.33	0.843	1.006	-2.982	DP
20	29628236	3	1656	0,2,2	0,0,0	.	FRG1B	0.588	0.558	369516	2.081	114.1	15.69	0.2581	2	1.21E-03	49.12	0	-0.241	0	0.24	-0.073	1.204	-3.243	DP
20	29628263	3	1656	0,0,0	0,3,3	.	FRG1B	1.91	0.043	334314	0	117.65	16.19	-0.0018	3	1.81E-03	45.01	0	0.317	0	0.35	-0.181	0.586	-2.969	DP
20	29628282	2	1656	0,2,2	0,0,0	.	FRG1B	-1.187	-0.481	161147	2.309	112.73	16.61	-0.0014	2	1.21E-03	43.63	0	-1.491	0	0.38	0.668	1.535	-2.447	DP
20	29628299	5	1656	0,2,2	0,2,2	.	FRG1B	-0.328	-0.201	159594	0	109.69	21.89	-0.0037	5	3.02E-03	41.33	0	-0.458	0	0.06	-0.442	1.758	-2.747	DP
20	29631562	2	1656	0,0,0	0,2,2	.	FRG1B	2.45	0.765	20446	5.01	67.07	42.02	-0.0014	2	1.21E-03	42.42	0	-2.058	0	11.18	1.86	1.043	-0.0967	MQRankSum
20	29632641	17	1656	0,4,4	0,12,12	.	FRG1B	2.09	0.087	139942	0	109.72	40.52	-0.0163	14	8.45E-03	49.05	0	0.149	0	0.69	-0.952	0.503	-1.438	QD
21	35893925	2	1654	0,2,2	0,0,0	nonsyn	RCAN1	1.9	-0.211	19922	1.421	59.97	13.77	-0.0023	2	1.21E-03	60	0	0.549	1	17.11	0.634	0.426	2.29	DP
21	35893950	16	1652	0,7,7	0,8,8	nonsyn	RCAN1	-2.371	0	19594	1.381	68.03	117.21	-0.012	16	9.69E-03	60	0	0.058	2	12.78	0.151	0.872	3.22	DP
X	19482454	3	1654	0,1,1	1,0,2	nonsyn	MAP3K15	-1.732	0.192	19661	2.362	64.16	22.14	0.4168	3	1.81E-03	60	0	1.23	1	23.36	0.898	0.924	0.418	DP
X	1.15E+08	2	1656	0,2,2	0,0,0	nonsyn	PLS3	1.52	1.33	19045	8.117	61.58	27.71	-0.0016	2	1.21E-03	60	0	-0.067	0	12	1.82	0.706	1.12	DP

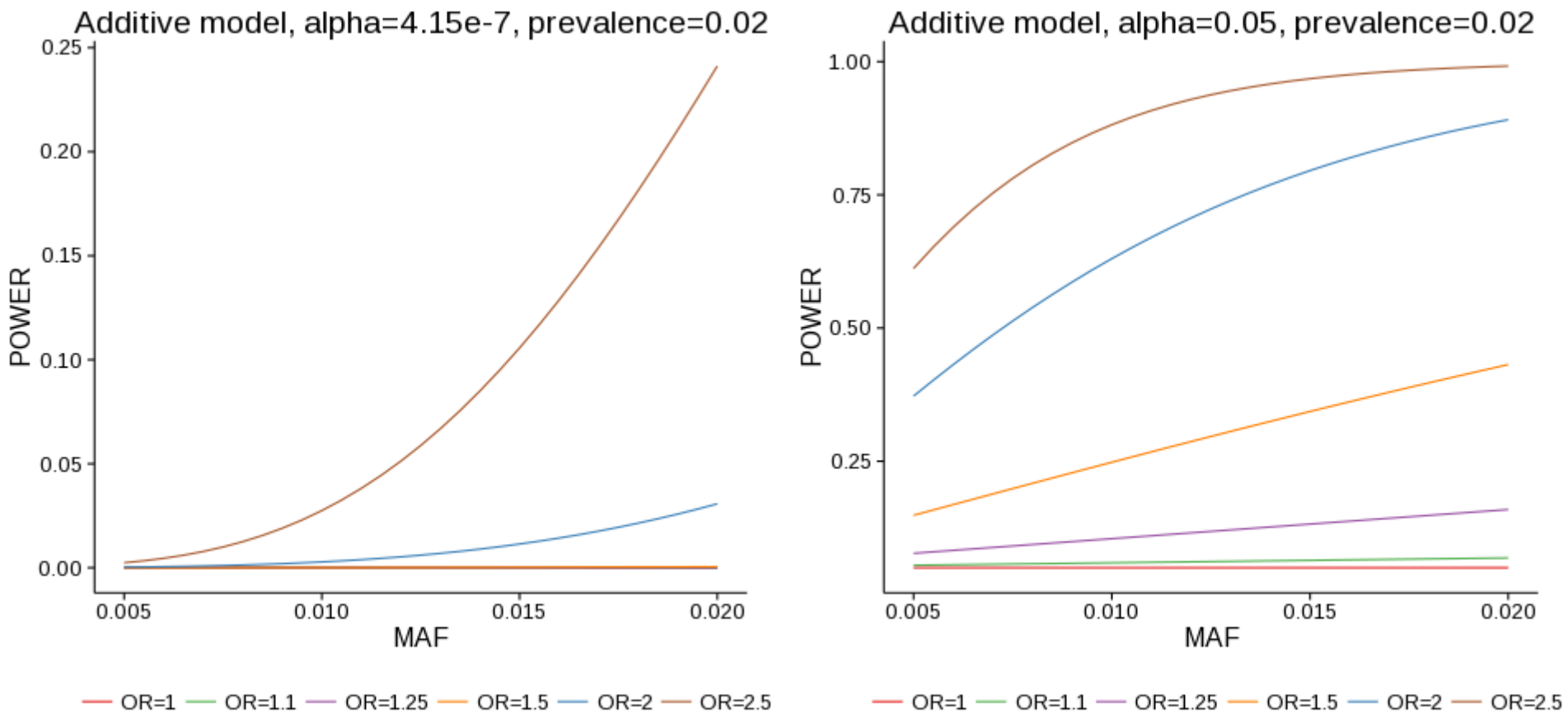


Figure S5 Statistical power of single variant given known allele frequency in WES discovery dataset, using different odds ratio values.

R package 'CaTS' v1.02 (Skol et al., 2006) was used with following parameters: prevalence=0.02; additive model; pimarkers=1; ncases=185; ncontrols=563; alpha=4.15e-7 or 0.05.

Key: MAF, minor allele frequency; OR, odds ratio

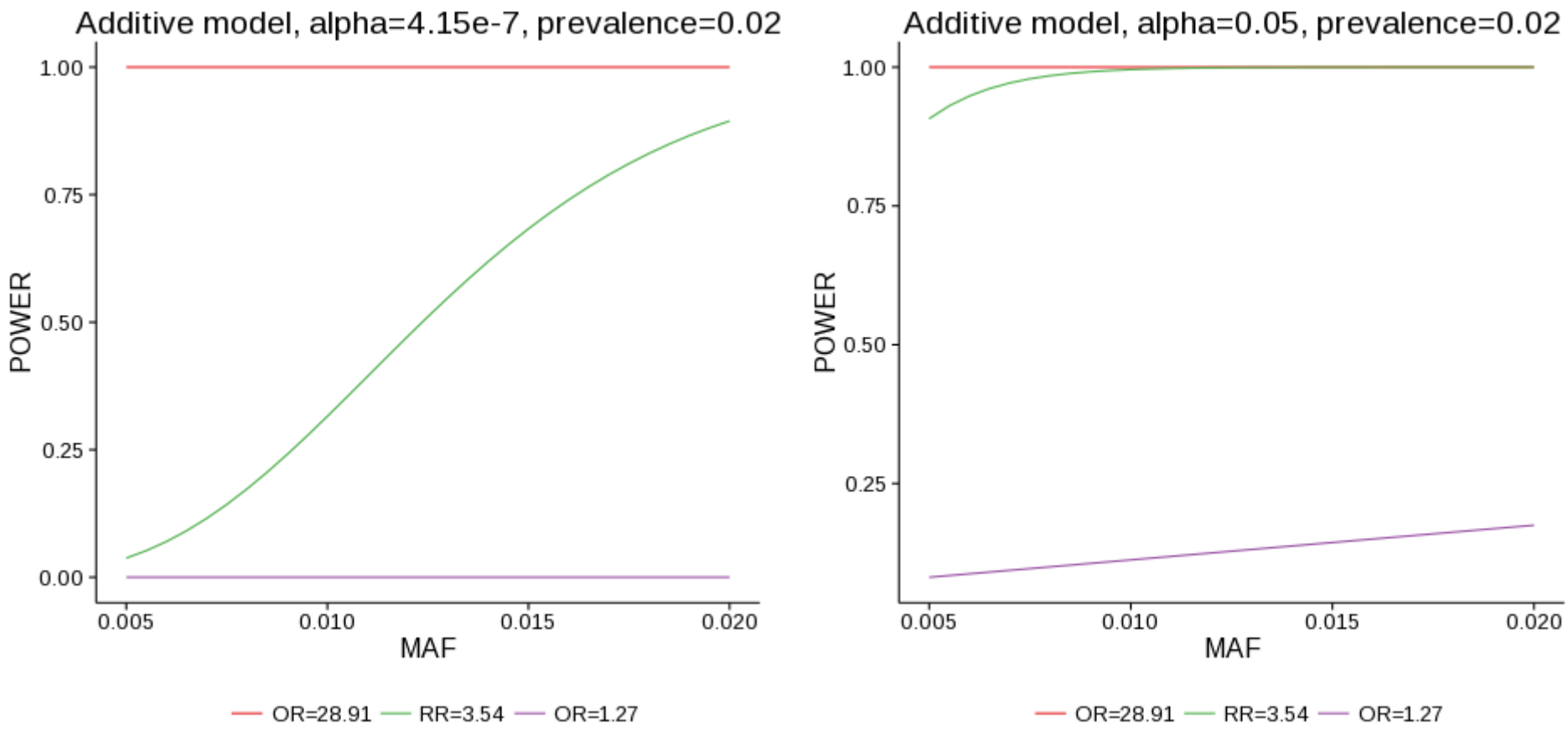


Figure S6 Statistical power of single variant rs141620200 given known allele frequency in WES discovery dataset, using different odds ratio and risk ratio values.

OR=28.91 and RR=3.54 were calculated using allele frequencies, and OR=1.27 was estimated from beta coefficient in q.emmax test. R package 'CaTS' v1.02 (Skol et al., 2006) was used with following parameters: prevalence=0.02; additive model; pimarkers=1; ncases=185; ncontrols=563; alpha=4.15e-7 or 0.05.

Key: maf, minor allele frequency; OR, odds ratio; RR, risk ratio.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat.Methods. 7, 248-249.

Bengtsson-Ellmark, S.H., Nilsson, J., Orho-Melander, M., Dahlenborg, K., Groop, L., Bjursell, G., 2004. Association between a polymorphism in the carboxyl ester lipase gene and serum cholesterol profile. Eur.J.Hum.Genet. 12, 627-632.

Carrell, R.W., Lomas, D.A., 2002. Alpha1-antitrypsin deficiency--a model for conformational diseases. N.Engl.J.Med. 346, 45-53.

Cecchini, M.P., Fasano, A., Boschi, F., Osculati, F., Tinazzi, M., 2015. Taste in Parkinson's disease. J.Neurol. 262, 806-813.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4, 7-015-0047-8. eCollection 2015.

Chun, S., Fay, J.C., 2009. Identification of deleterious mutations within three human genomes. Genome Res. 19, 1553-1561.

Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012a. Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Front.Genet. 3, 35.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012b. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6, 80-92.

Cronin, S., Tomik, B., Bradley, D.G., Slowik, A., Hardiman, O., 2009. Screening for replication of genome-wide SNP associations in sporadic ALS. Eur.J.Hum.Genet. 17, 213-218.

Delaneau, O., Howie, B., Cox, A.J., Zagury, J.F., Marchini, J., 2013. Haplotype estimation using sequencing reads. Am.J.Hum.Genet. 93, 687-696.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat.Genet. 43, 491-498.

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., Liu, X., 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum.Mol.Genet. 24, 2125-2137.

Durbin, R., 2014. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. 30, 1266-1272.

Graziadei, I., Gaggl, S., Kaserbacher, R., Braunsteiner, H., Vogel, W., 1994. The acute-phase protein alpha 1-antitrypsin inhibits growth and proliferation of human early erythroid progenitor cells (burst-forming units-erythroid) and of human erythroleukemic cells (K562) in vitro by interfering with transferrin iron uptake. Blood. 83, 260-268.

Griswold, A.J., Dueker, N.D., Van Booven, D., Rantus, J.A., Jaworski, J.M., Slifer, S.H., Schmidt, M.A., Hulme, W., Konidari, I., Whitehead, P.L., Cuccaro, M.L., Martin, E.R., Haines, J.L., Gilbert, J.R., Hussman, J.P., Pericak-Vance, M.A., 2015. Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. Mol.Autism. 6, 43-015-0034-z. eCollection 2015.

Halbgebauer, S., Nagl, M., Klafki, H., Haussmann, U., Steinacker, P., Oeckl, P., Kassubek, J., Pinkhardt, E., Ludolph, A.C., Soininen, H., Herukka, S.K., Wiltfang, J., Otto, M., 2016. Modified serpinA1 as risk marker for Parkinson's disease dementia: Analysis of baseline data. Sci.Rep. 6, 26145.

Hernandez, D.G., Nalls, M.A., Ylikotila, P., Keller, M., Hardy, J.A., Majamaa, K., Singleton, A.B., 2012. Genome wide assessment of young onset Parkinson's disease from Finland. PLoS One. 7, e41859.

Huttenlocher, J., Kruger, R., Capetian, P., Lohmann, K., Brockmann, K., Csoti, I., Klein, C., Berg, D., Gasser, T., Bonin, M., Riess, O., Bauer, P., 2015. EIF4G1 is neither a strong nor a common risk factor for Parkinson's disease: evidence from large European cohorts. J.Med.Genet. 52, 37-41.

Ihaka, R., Gentleman, R., 1996. R: a language for data analysis and graphics. 5, 299-314.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., Eskin, E., 2010. Variance component model to account for sample structure in genome-wide association studies. Nat.Genet. 42, 348-354.

Kowalska, A., Danker-Hopfe, H., Wender, M., Florczak, J., Walter, H., 1996. Association between the PI*M3 allele of alpha 1-antitrypsin and Alzheimer's disease? A preliminary report. Hum.Genet. 98, 744-746.

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat.Protoc. 4, 1073-1081.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25, 1754-1760.

Lim, E.T., Wurtz, P., Havulinna, A.S., Palta, P., Tukiainen, T., Rehnstrom, K., Esko, T., Magi, R., Inouye, M., Lappalainen, T., Chan, Y., Salem, R.M., Lek, M., Flannick, J., Sim, X., Manning, A., Ladenvall, C., Bumpstead, S., Hamalainen, E., Aalto, K., Maksimow, M., Salmi, M., Blankenberg, S., Ardissino, D., Shah, S., Horne, B., McPherson, R., Hovingh, G.K., Reilly, M.P., Watkins, H., Goel, A., Farrall, M., Girelli, D., Reiner, A.P., Stitziel, N.O., Kathiresan, S., Gabriel, S., Barrett, J.C., Lehtimaki, T., Laakso, M., Groop, L., Kaprio, J., Perola, M., McCarthy, M.I., Boehnke, M., Altshuler, D.M., Lindgren, C.M., Hirschhorn, J.N., Metspalu, A., Freimer, N.B., Zeller, T., Jalkanen, S., Koskinen, S., Raitakari, O., Durbin, R., MacArthur, D.G., Salomaa, V., Ripatti, S., Daly, M.J., Palotie, A., Sequencing Initiative Suomi (SISu) Project, 2014. Distribution and medical impact of loss-of-function variants in the Finnish founder population. PLoS Genet. 10, e1004494.

McGurk, L., Lee, V.M., Trojanowksi, J.Q., Van Deerlin, V.M., Lee, E.B., Bonini, N.M., 2014. Poly-A binding protein-1 localization to a subset of TDP-43 inclusions in amyotrophic lateral sclerosis occurs more frequently in patients harboring an expansion in C9orf72. J.Neuropathol.Exp.Neurol. 73, 837-845.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M.F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J.H., Cheng, R., International Parkinson's Disease Genomics Consortium (IPDGC), Parkinson's Study Group (PSG) Parkinson's Research: The Organized GENetics Initiative (PROGENI), 23andMe, GenePD, NeuroGenetics Research Consortium (NGRC), Hussman Institute of Human Genomics (HIHG), Ashkenazi Jewish Dataset Investigator, Cohorts for Health and Aging Research in Genetic Epidemiology (CHARGE), North American Brain Expression Consortium (NABEC), United Kingdom Brain Expression Consortium (UKBEC), Greek Parkinson's Disease Consortium, Alzheimer Genetic Analysis Group, Ikram, M.A., Ioannidis, J.P., Hadjigeorgiou, G.M., Bis, J.C., Martinez, M., Perlmutter, J.S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R.H., Clark, L.N., Stefansson, K., Hardy, J.A., Heutink, P., Chen, H., Wood, N.W., Houlden, H., Payami, H., Brice, A., Scott, W.K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., Singleton, A.B., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat.Genet. 46, 989-993.

Panicot-Dubois, L., Thomas, G.M., Furie, B.C., Furie, B., Lombardo, D., Dubois, C., 2007. Bile salt-dependent lipase interacts with platelet CXCR4 and modulates thrombus formation in mice and humans. J.Clin.Invest. 117, 3708-3719.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 26, 2336-2337.

Raeder, H., Johansson, S., Holm, P.I., Haldorsen, I.S., Mas, E., Sbarra, V., Nermoen, I., Eide, S.A., Grevle, L., Bjorkhaug, L., Sagen, J.V., Aksnes, L., Sovik, O., Lombardo, D., Molven, A., Njolstad, P.R., 2006. Mutations in the CEL VNTR cause a syndrome of diabetes and pancreatic exocrine dysfunction. Nat.Genet. 38, 54-62.

Ravenscroft, G., Nolent, F., Rajagopalan, S., Meireles, A.M., Paavola, K.J., Gaillard, D., Alanio, E., Buckland, M., Arbuckle, S., Krivanek, M., Maluenda, J., Pannell, S., Gooding, R., Ong, R.W., Allcock, R.J., Carvalho, E.D., Carvalho, M.D., Kok, F., Talbot, W.S., Melki, J., Laing, N.G., 2015. Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. Am.J.Hum.Genet. 96, 955-961.

Reva, B., Antipin, Y., Sander, C., 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 39, e118.

Schwarz, J.M., Rodelsperger, C., Schuelke, M., Seelow, D., 2010. MutationTaster evaluates disease-causing potential of sequence alterations. Nat.Methods. 7, 575-576.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N., Gaunt, T.R., 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum.Mutat. 34, 57-65.

Shyr, C., Tarailo-Graovac, M., Gottlieb, M., Lee, J.J., van Karnebeek, C., Wasserman, W.W., 2014. FLAGS, frequently mutated genes in public exomes. BMC Med.Genomics. 7, 64-014-0064-y.

Skol, A.D., Scott, L.J., Abecasis, G.R., Boehnke, M., 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat.Genet. 38, 209-213.

Tarun, S.Z.,Jr, Sachs, A.B., 1996. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. EMBO J. 15, 7168-7177.

Taylor, A.K., Zambaux, J.L., Klisak, I., Mohandas, T., Sparkes, R.S., Schotz, M.C., Lusis, A.J., 1991. Carboxyl ester lipase: a highly polymorphic locus on human chromosome 9qter. Genomics. 10, 425-431.

Thomas, J.H., Schneider, S., 2011. Coevolution of retroelements and tandem zinc finger genes. Genome Res. 21, 1800-1812.

Tryka, K.A., Hao, L., Sturcke, A., Jin, Y., Wang, Z.Y., Ziyabari, L., Lee, M., Popova, N., Sharopova, N., Kimura, M., Feolo, M., 2014. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 42, D975-9.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Ponten, F., 2015. Proteomics. Tissue-based map of the human proteome. Science. 347, 1260419.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2002. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline, in: Anonymous Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164.

Westendorf, J.M., Konstantinov, K.N., Wormsley, S., Shu, M.D., Matsumoto-Taniura, N., Pirollet, F., Klier, F.G., Gerace, L., Baserga, S.J., 1998. M phase phosphoprotein 10 is a human U3 small nucleolar ribonucleoprotein component. Mol.Biol.Cell. 9, 437-449.

Ylikotila, P., Tiirikka, T., Moilanen, J.S., Kaariainen, H., Marttila, R., Majamaa, K., 2015. Epidemiology of early-onset Parkinson's disease in Finland. Parkinsonism Relat.Disord. 21, 938-942.