Supplementary Online Content

Tosto G, Vardarajan B, Sariya S, et al. Association of variants in *PINX1* and *TREM2* with late-onset Alzheimer disease per a gene-based, transethnic meta-analysis. *JAMA Neurol*. Published online May 6, 2019. doi:10.1001/jamaneurol.2019.1066

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethod 1. Whole-exome sequencing quality control.

Quality control (QC) procedures excluded monomorphic variants, VQSR non-"PASS" variants with call rates <80%, and variants with low average depth of data (DP) and genotype quality (GQ) (8<DP>500 reads and GQ <20, corresponding to a >1% likelihood of being an incorrect genotype call) [1]. "DP" values represented the number of reads passing QC used to calculate the genotype at a specific site in a specific sample, with higher values for DP generally leading to more accurate genotype calls. "GQ" is a Phred-scaled value representing the confidence that the called genotype is the true genotype. Again, higher values reflected more accurate genotype calls. These thresholds were chosen according to studies of concordance between sequencing experiments and genotyping arrays in order to achieve a 99% genotype likelihood. Because these simulations were focused on SNV only, for INDELS we additionally applied GATK-reccomanded hard filter (QualByDepth (QD) >2.0; FisherStrand (FS) <200.0; ReadPosRankSumTest (ReadPosRankSum) > 20.0) (https://software.broadinstitute.org/gatk/documentation/article?id=2806). We handled multi-allelic site by splitting the alternative alleles in multiple biallelic sites using *bcftools* utility [2]. After this step, we normalized variants by applying parsimonial representation (i.e. coding the variant in as few nucleotides as possible without reducing the length of any allele to 0) and left-alignment (i.e. shifting the start position of that variant to the left till it is no longer possible) [3]. Variants showing strong departure from Hardy Weinberg equilibrium (HWE, p<1-e07) in controls were also filtered out. Principal components (PCs) were estimated using WES data, and filtering out single nucleotide variant (SNV) with minor allele frequency (MAF) ≤5%, call rate <95%. We used the KING software to detect duplicates and cryptic relatedness, and output the first 10 PCs, separately for each ethnic group. We then excluded outliers that deviated more than six standard deviation from the mean.

eMethod 2. Replication cohorts' description.

1. Alzheimer Disease Sequencing Project (ADSP). Individuals were aged 60 years or older and met NINCDS-ADRDA criteria [4] for possible, probable or definite AD based on clinical assessment, or had presence of LOAD (moderate or high likelihood) upon neuropathology examination. Healthy controls were similar in age and either judged to be cognitively normal or did not meet pathological criteria for LOAD following brain autopsy [5]. We did not include the Caribbean Hispanic part of the ADSP dataset because they overlapped with those already present in our WHICAP WES data.
2. The ROS/MAP study. The Religious Orders Study (ROS) is a longitudinal cohort study of aging and Alzheimer's disease from Rush University recruiting Individuals from more than 40 groups of religious orders across the US. Subjects were included at baseline if dementia was not present. The Memory and Aging Project (MAP) is a longitudinal, epidemiologic clinical-pathologic cohort study based at Rush University of aging and risk of LOAD that began in 1997. This study was designed to complement the ROS study by enrolling individuals with a wider range of life experiences and socioeconomic status. The study enrolls older individuals without any signs of dementia, primarily recruiting from continuous care retirement communities throughout northeastern Illinois, USA. LOAD status was determined by an actuarial decision tree that incorporated a computer algorithm and clinical judgment made in series by a neuropsychologist and a clinician who ultimately reviewed all cases. Details of the ROS and MAP study has been described in detail elsewhere [6]. Participants were deemed "non-demented" (healthy or MCI) or LOAD. Other dementia types were excluded from analyses.
3. ADSP Family Study. Whole exome data from 67 families of Caribbean Hispanic ancestry (N=358 participants) were selected from the ADSP family dataset (Table 1). No individuals with known early-onset disease mutations (*APP*, *PSEN1*, *PSEN2*, *GRN*, or *MAPT*) were included. All probands were part of families with three or more affected individuals recruited as part of the Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA) study. Detailed description of this cohort has been published elsewhere [7].
4. We used single-marker summary statistics from the International Genomics of Alzheimer's Project (IGAP), a large two-stage genome-wide association study (GWAS) on individuals of European ancestry (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php - N=54,162) [8]. In stage 1, IGAP employed 7,055,881 genotyped and imputed single nucleotide polymorphisms to meta-analyses four previously-published GWAS datasets (The European Alzheimer's disease Initiative (EADI); the Alzheimer Disease Genetics Consortium (ADGC); The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium (CHARGE) The Genetic and Environmental Risk in AD consortium (GERAD)). In stage 2, top hits from Stage 1 were genotyped and tested for association in an independent set of 8,572 LOAD cases and 11,312 controls and then again meta-analyzed. We used stage 1 summary results in order to include all SNPs available within each gene.

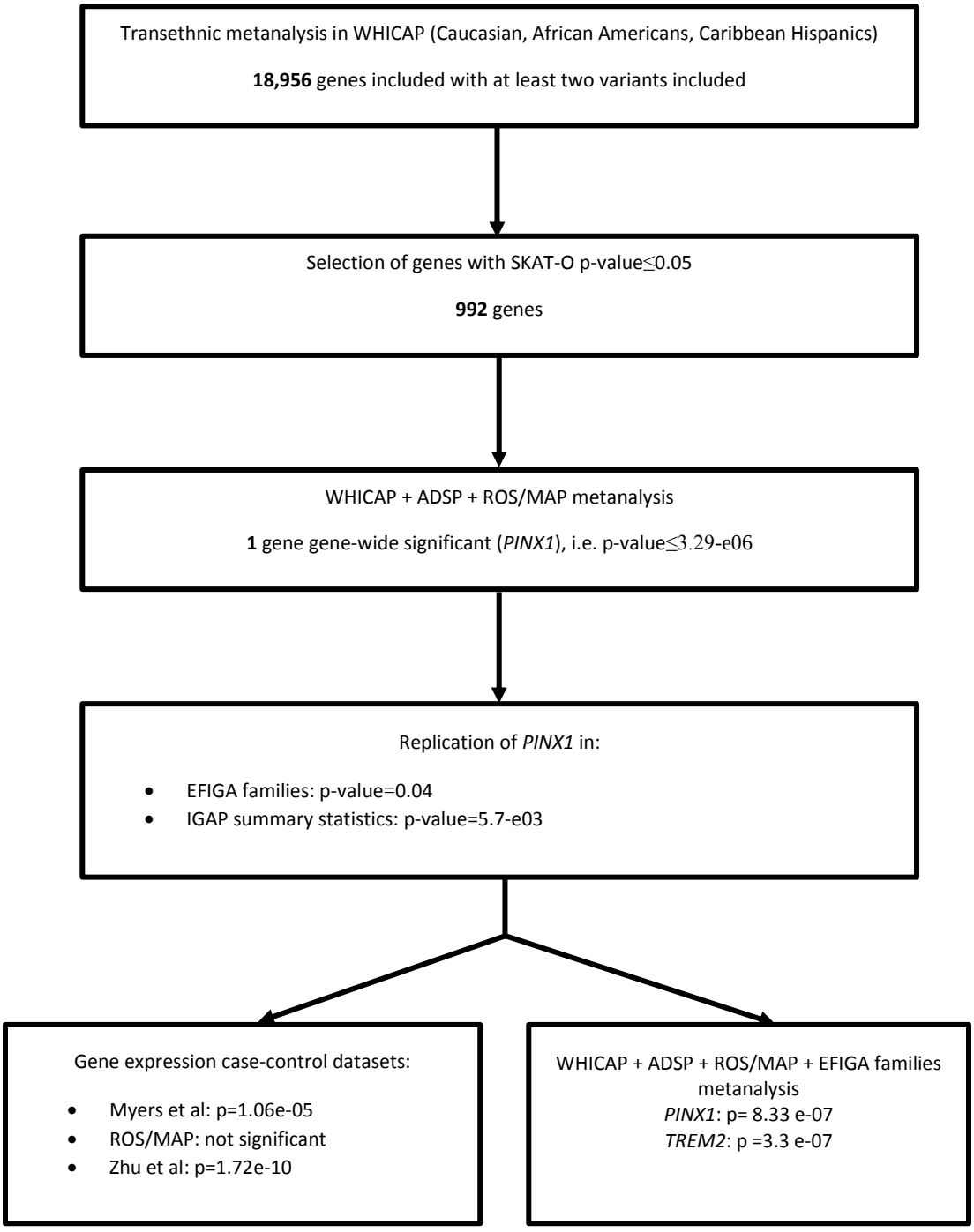
eMethod 3. Expression data description and analyses methods.

1. Description. The Myers and colleagues [9] neocortical transcriptome data. Briefly, a set of 188 individuals without dementia and 176 autopsy confirmed LOAD. Data were downloaded from the NCBI GEO archive (ID: GSE15222). All data were generated using the Illumina HumanRef-8 expression BeadChip (GPL2700) v2 Rev0. Methods. Data were analyzed in R using residual corrected profiles for each individual and for each transcript. Expression measures were corrected for sex, *APOE* status, age at death, cortical region, day of expression hybridization, study cohort, post-mortem interval and transcript detection rate. One-way ANOVA compared expression profile between affected brain sand normal control brains for genes prioritized in genetic analyses.

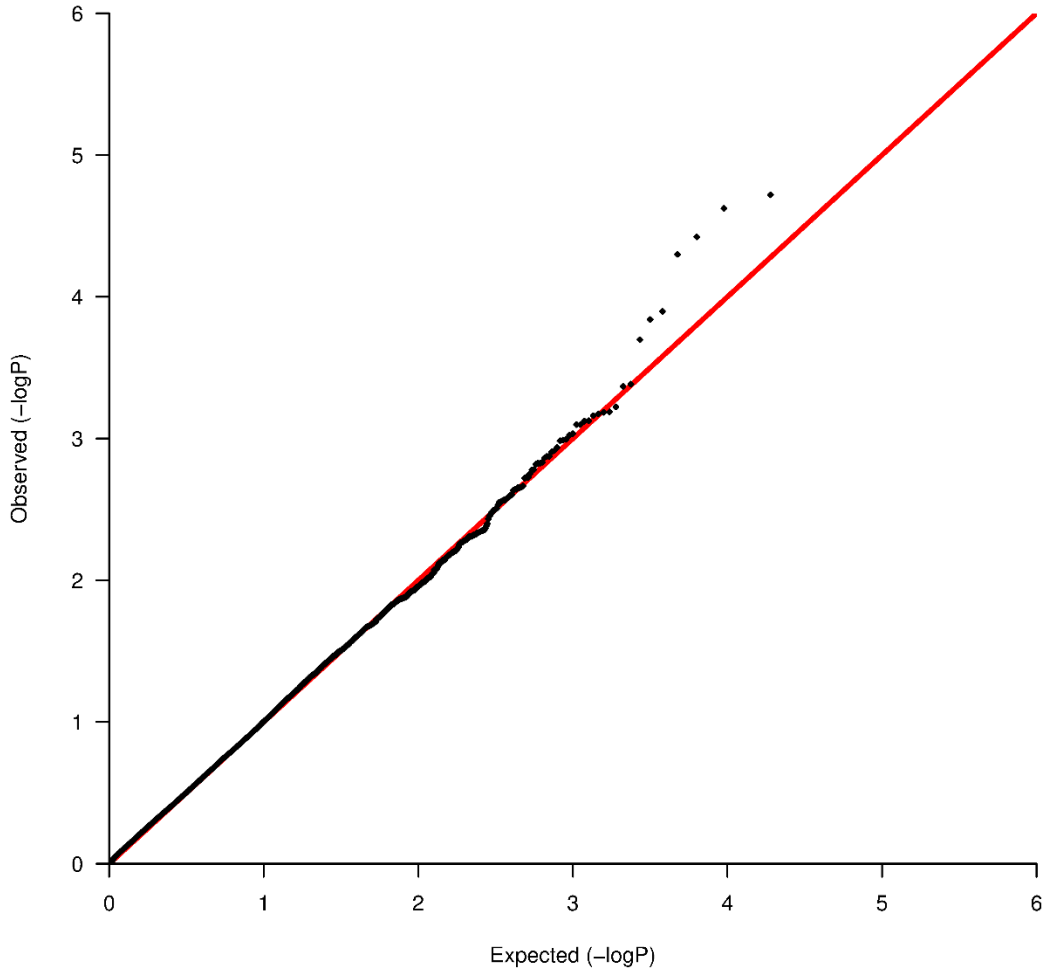
2. Description. ROS/MAP Next-generation RNA sequencing (RNA-seq) data. Briefly, RNA was purified from frozen dorsolateral prefrontal cortex tissue of ROS-MAP participants with the miRNeasy Mini Kit and RNase-Free DNase Set (Qiagen, Germantown, MD). RNA concentration and quality were measured with a Nanodrop (Thermo Fisher Scientific, Wilmington, DE) and Bioanalyzer (Agilent Technologies, Santa Clara, CA). A RIN score of >5 was required for library construction, which was assembled using the strand-specific dUTP method. Sequencing was performed with Illumina HiSeq with 101 base pair paired-end reads and a goal coverage of >85 million paired-end reads as previously described [10]. After QC, 508 subjects were available to analyze. Fragments per kb of transcript per million fragments mapped (FPKM) were quantile-normalized, correcting for batch effect with Combat [11]. Methods. Linear regressions were applied with neuropathological measure as the dependent variables, Combat-adjusted FPKM values as the independent variable, and technical factors as covariates (RNA integrity score, log2[total aligned reads], postmortem interval, and number of ribosomal bases) as detailed elsewhere [12].

3. Description. The Narayanan and colleagues dataset[12] comprised DLPFC (BA9) brain tissues of 624 individual (AD patients, HD patients and non-demented controls samples), obtained from Harvard Brain tissue resource center (HBTRC). The HBTRC samples were primarily of Caucasian ancestry, as only eight non-Caucasian outliers were identified, and therefore excluded for further analysis. Post-mortem interval (PMI) was 17.8+8.3 hours (mean ± standard deviation), sample pH was 6.4±0.3 and RNA integrity number (RIN) was 6.8±0.8 for the average sample in the overall cohort. Tissues were profiled on a custom-made Agilent 44K array (GPL4372). 310 LOAD cases and 157 controls were included in the analyses.

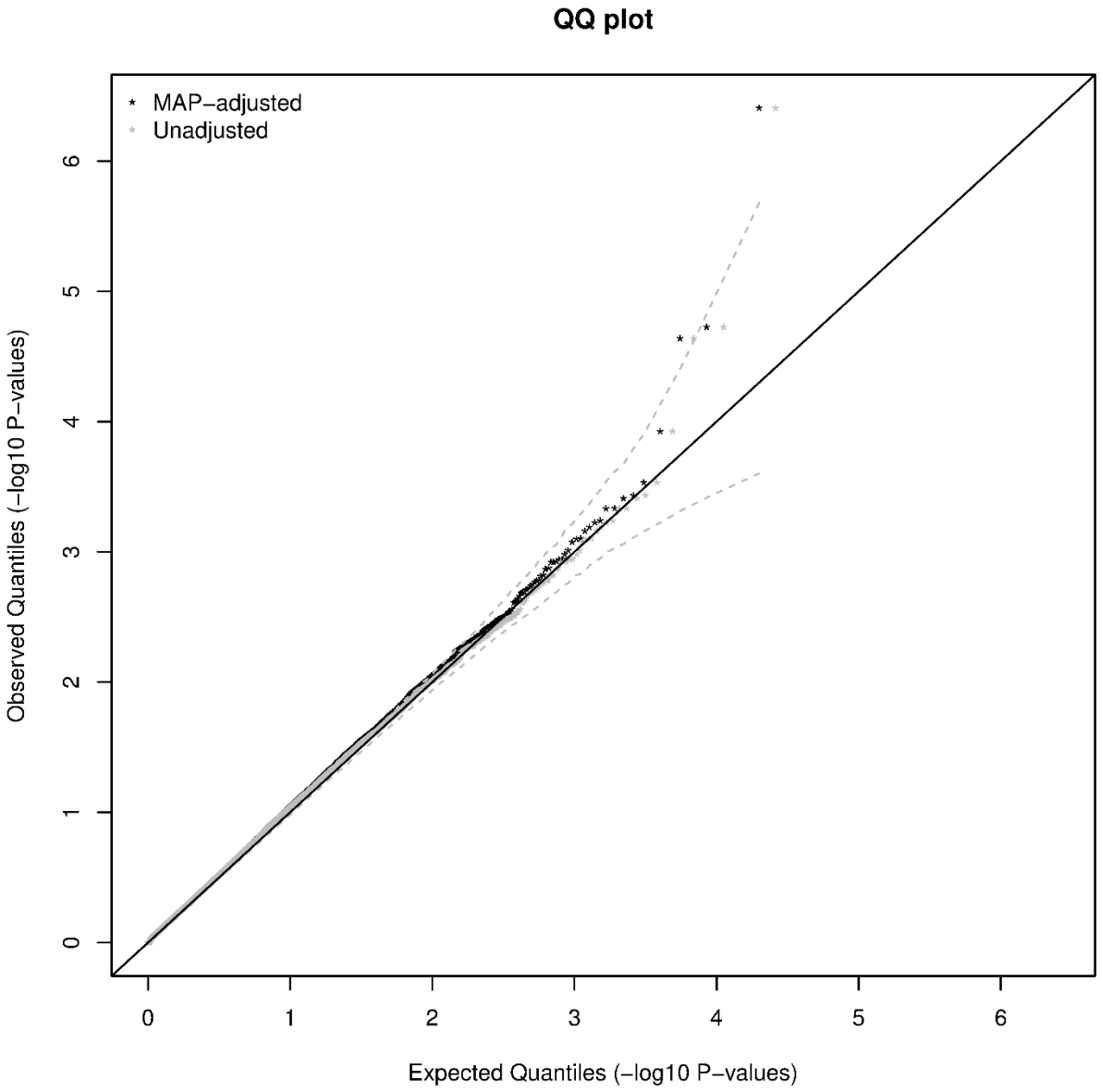
eFigure 1. Flowchart of the study for the MODERATE-HIGH VEP annotation model.



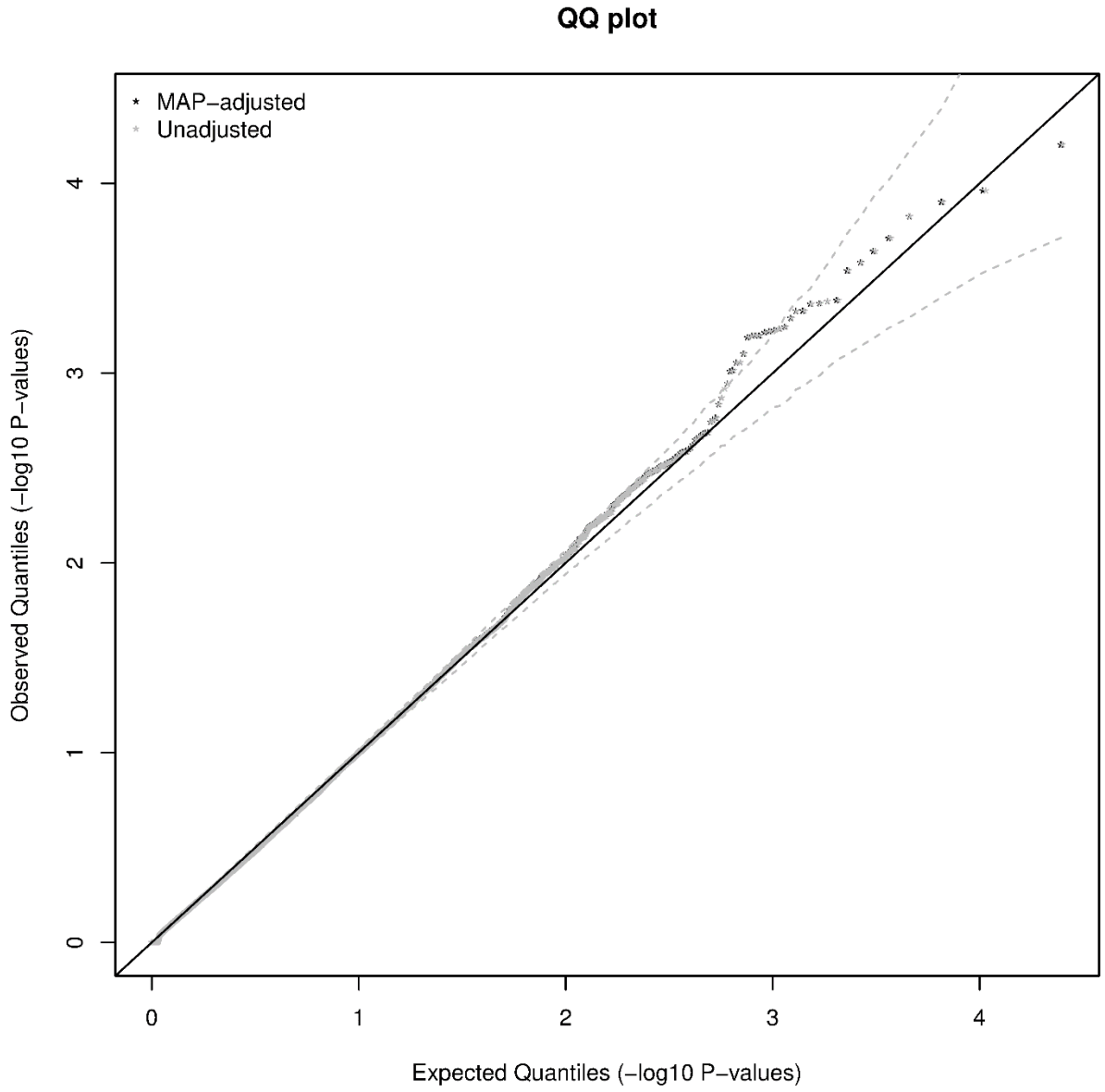
eFigure 2. Quantile-quantile plot for the moderate-high SKAT-O model in the WHICAP meta-analysis.



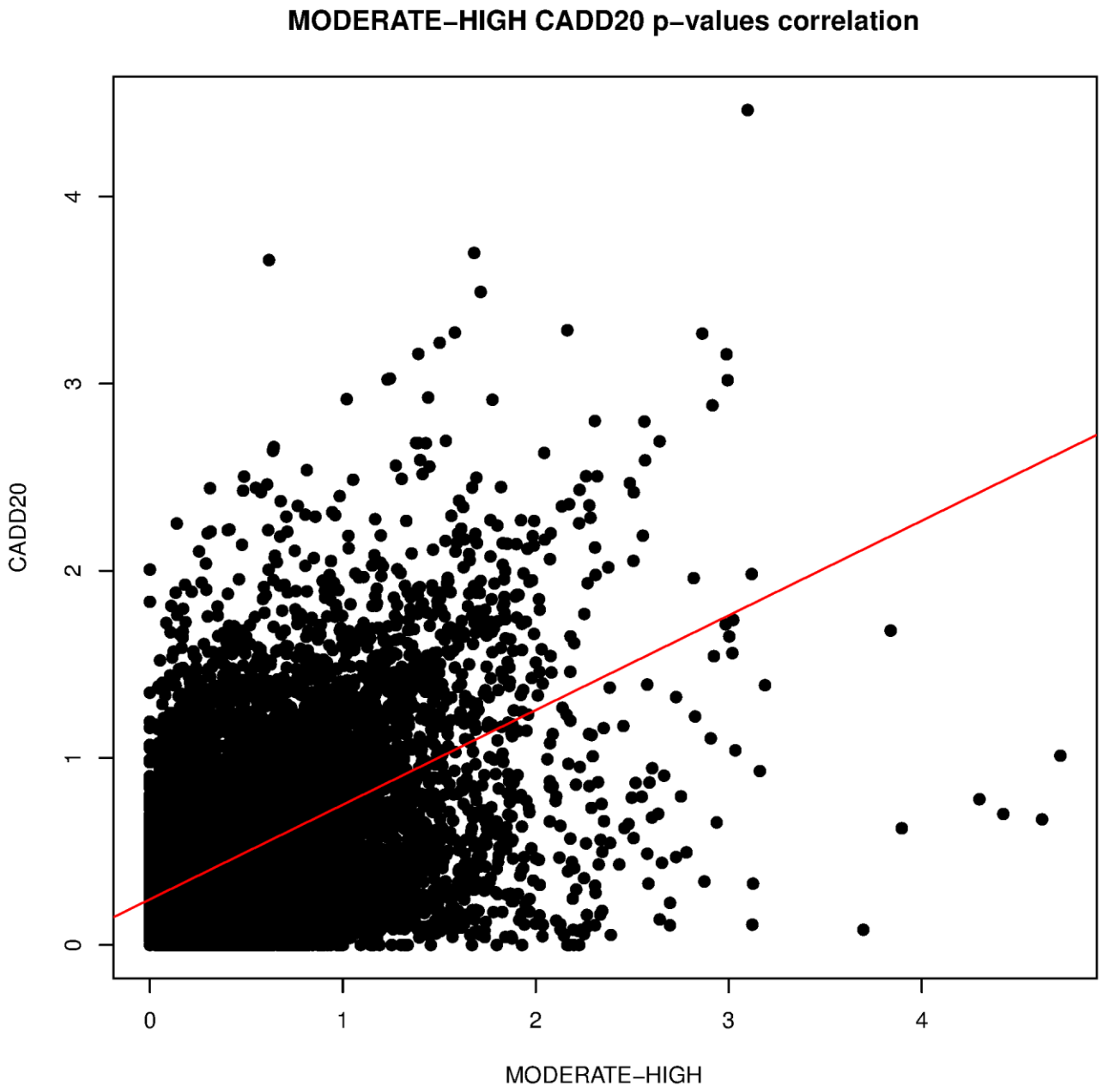
eFigure 3. Quantile-quantile plot for the moderate-high SKAT-O model in the ROS/MAP (unadjusted and adjusted by minimum achievable p-values – MAP).



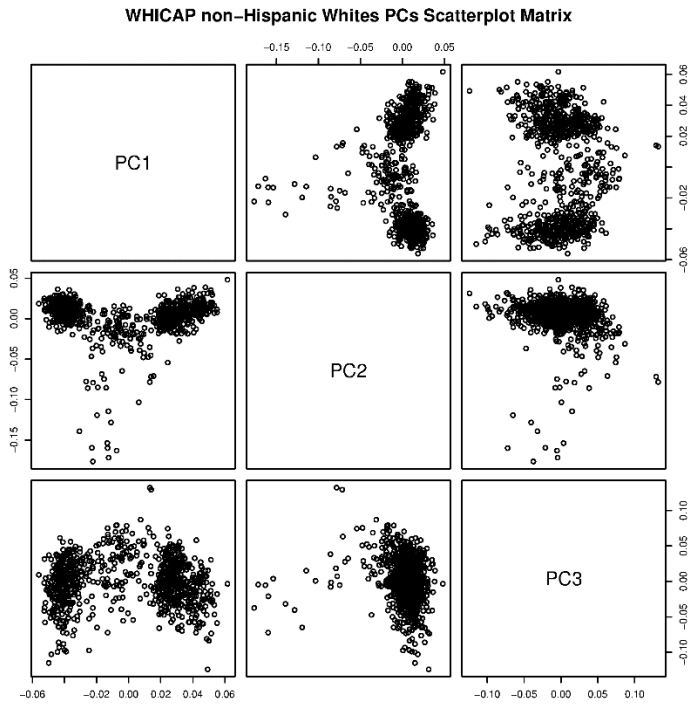
eFigure 4. Quantile-quantile plot for the moderate-high SKAT-O model in the ADSP (unadjusted and adjusted by minimum achievable p-values – MAP).

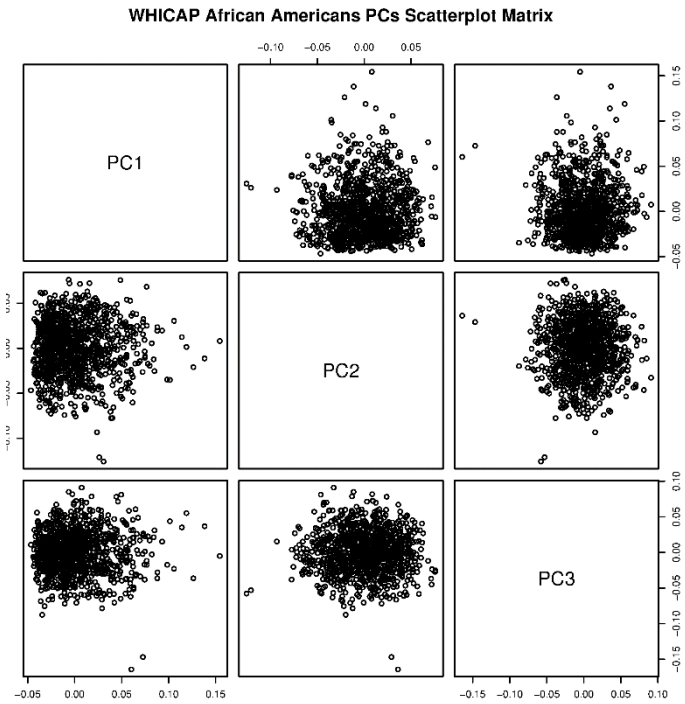


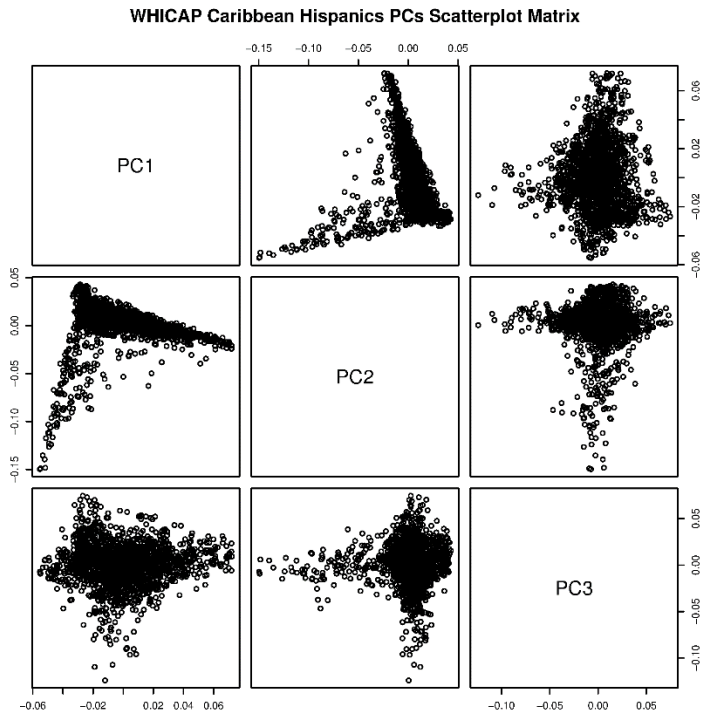
eFigure 5. Correlation between WHICAP meta-analysis SKAT-O p-values according to annotation models, i.e. VEP Moderate-high and CADD20 (Spearman coefficient was =0.49, p-values <0.001).



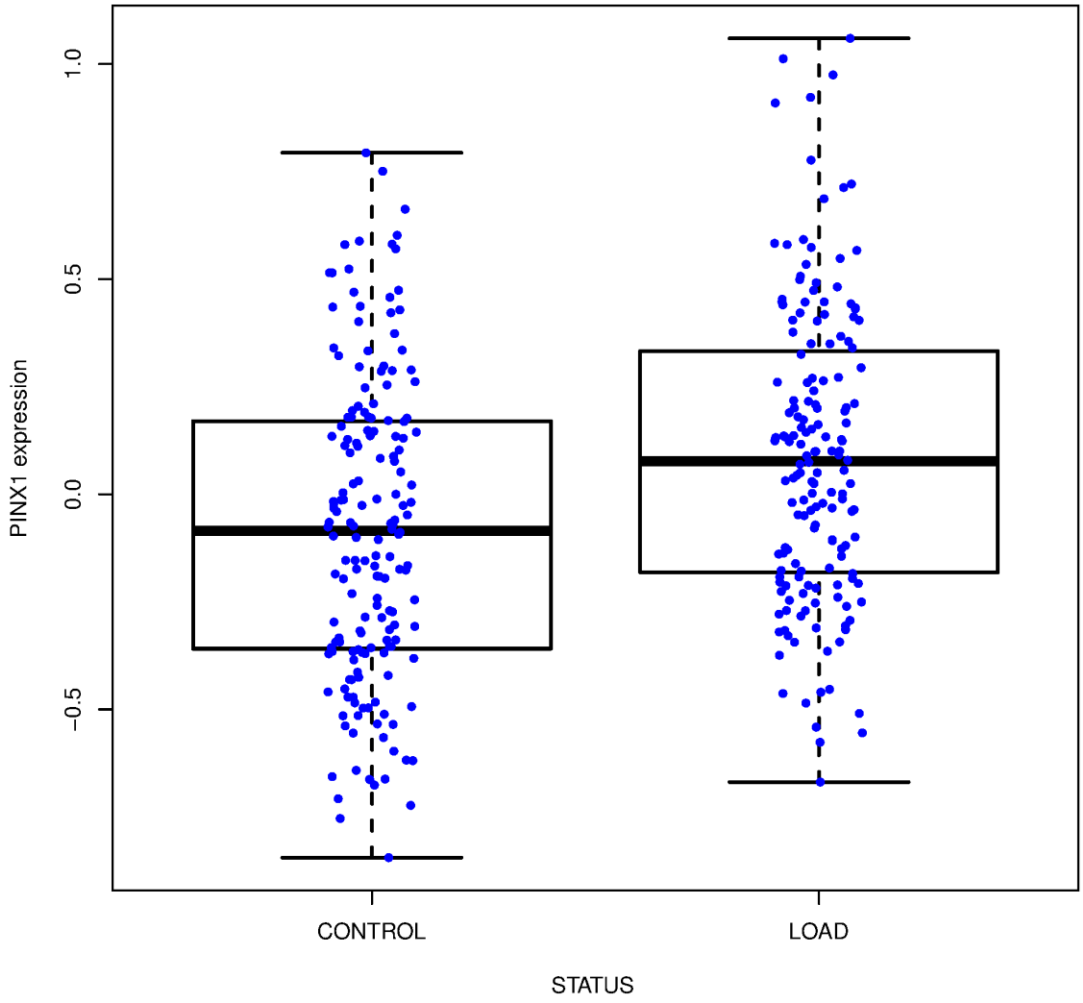
eFigure 6. Principal components (PC) scatterplot matrix for each ethnic group of the WHICAP dataset. We included the PCs #1,#2,#3 in each statistical model.







eFigure 7. Diffierential expression boxplot between LOAD cases and normal controls in Myers et al. dataset for *PINX1*. For LOAD brains, mean standardized pinx1 expression= 0.087; for control brains, mean standardized pinx1 expression= -0.085.



eTable 1. *PINX1* variants included in the VEP MODERATE-HIGH analyses. Table shows base pair (BP, minor allele frequency (MAF), missing rate, major and minor allele, p-value for single-marker analysis (in bold significant p-values) along with effect size (beta) and standard error (s.e.), CADD score. LoF variants are highlighted in red.

SNP	BP	MAF	Missing Rate	MajorAllele	MinorAllele	#SNP in controls	#SNP in cases	beta	s.e.	single-marker pvalue	CADD score	snp138
8:10622923:TTTC:T	10622923	0.006896552	0.000255363	TTTC	T	35	12	-0.788816595	0.306351562	**0.010027668**	.	rs201660183
8:10622930:T:C	10622930	0.004473415	0.00102145	T	C	17	12	0.2291413	0.401678461	0.568366073	17.38	rs61746595
8:10623000:T:TC	10623000	0.000255428	0.000255363	T	TC	0	2	1.918055022	1.43784195	0.182209815	.	.
8:10623027:G:C	10623027	0.000127845	0.001276813	G	C	0	1	2.100124255	2.005695283	0.29506249	0.852	.
8:10623094:C:G	10623094	0.000511117	0.000766088	C	G	2	2	0.204283648	1.01167179	0.83997395	18.77	rs369418829
8:10623104:C:T	10623104	0.000127779	0.000766088	C	T	1	0	-1.678203154	2.045461306	0.411958389	0.047	rs150489215
8:10623134:G:C	10623134	0.000383436	0.00102145	G	C	3	0	-1.721853739	1.183360734	0.14565432	0.002	rs377260866
8:10623135:C:T	10623135	0.000255493	0.000510725	C	T	0	1	1.964770358	2.006240402	0.327417212	5.588	rs202202690
8:10623137:CTCT:C	10623137	0.000127714	0.000255363	CTCT	C	0	1	7.080972646	2.872535284	**0.01369902**	.	.
8:10623141:T:A	10623141	0.000127714	0.000255363	T	A	0	1	7.080972646	2.872535284	**0.01369902**	37	rs201016513
8:10623203:GC:G	10623203	0.000127714	0.000255363	GC	G	1	0	-1.879564594	2.020813157	0.352317703	.	.
8:10623240:T:C	10623240	0.038825032	0.000255363	T	C	192	96	-0.315669347	0.131495673	**0.016367934**	0.001	rs17711777
8:10623254:C:A	10623254	0.01073071	0.000510725	C	A	52	27	-0.380986489	0.235443248	0.105626617	24.2	rs17855458
8:10623280:C:G	10623280	0.009200102	0.000766088	C	G	33	34	0.294594541	0.257680527	0.252932749	0.007	rs35530857
8:10623301:C:A	10623301	0	0.000510725	C	A	0	0	NA	NA	NA	10.82	rs202176931
8:10623309:C:T	10623309	0.000127747	0.000510725	C	T	1	0	-1.373771675	2.262806759	0.543778189	5.326	.
8:10623344:C:T	10623344	0.000127714	0.000255363	C	T	0	1	1.860406497	2.0077784	0.354134506	23.8	rs199675528
8:10623361:C:G	10623361	0.00038324	0.000510725	C	G	2	0	-1.743443751	1.513589574	0.249378482	21.8	rs375581032
8:10623379:G:C	10623379	0.000127714	0.000255363	G	C	0	1	2.218315869	2.014700316	0.270868398	4.485	.
8:10623383:G:T	10623383	0.00038324	0.000510725	G	T	3	0	-1.519737169	1.229890354	0.21658169	13.62	rs202049952
8:10623385:C:T	10623385	0.000638733	0.000510725	C	T	2	2	0.359249756	1.034425968	0.728370585	12.68	rs374083715
8:10623386:G:A	10623386	0.000127714	0.000255363	G	A	1	0	-1.826653777	2.013811482	0.364373511	14.1	rs377311809
8:10623393:C:T	10623393	0.000127714	0.000255363	C	T	1	0	-1.373771675	2.262806759	0.543778189	4.312	rs369880397
8:10623396:TCTC:T	10623396	0.000510856	0.000255363	TCTC	T	3	1	0.143017387	1.219164593	0.906616228	.	.
8:10623407:G:C	10623407	0.000127714	0.000255363	G	C	0	1	3.858862262	2.2865729	**0.091484525**	0.119	.
8:10623423:C:T	10623423	0.001021711	0.000255363	C	T	6	2	-0.508481321	0.809915074	0.530121508	6.407	rs200616748
8:10677710:G:A	10677710	0.000127812	0.00102145	G	A	0	1	2.160866408	2.016170191	0.283824313	13.07	rs189167078
8:10677806:G:C	10677806	0.000255624	0.00102145	G	C	1	1	0.279738036	1.493422653	0.851414938	.	rs368049075
8:10683689:G:T	10683689	0.000127812	0.00102145	G	T	1	0	-1.711168477	2.038053008	0.401127423	32	.
8:10683721:C:G	10683721	0.000127812	0.00102145	C	G	1	0	-1.82932458	2.013645853	0.363633174	26.9	.
8:10683752:A:G	10683752	0.000127812	0.00102145	A	G	0	1	2.370250715	2.030440971	0.243065983	15.88	.
8:10689208:C:T	10689208	0.000127747	0.000510725	C	T	0	1	1.863226453	2.01129488	0.354247733	24.8	rs201784803
8:10689232:C:A	10689232	0	0.000766088	C	A	0	0	NA	NA	NA	32	.
8:10690414:G:GGTA	10690414	0.000127714	0.000255363	G	GGTA	0	1	2.450764209	2.036470665	0.228807245	.	.
8:10690422:C:T	10690422	0.000127714	0.000255363	C	T	1	0	-1.361704771	2.277752881	0.549954612	33	rs377715499
8:10690474:T:G	10690474	0.001405571	0.000766088	T	G	5	5	0.909309109	0.712155695	0.201658626	25.4	rs189959562
8:10690476:G:T	10690476	0.000127779	0.000766088	G	T	1	0	-2.377636724	2.027578635	0.240936855	29	.
8:10692193:C:T	10692193	0.000127747	0.000510725	C	T	1	0	-1.963378374	2.005330672	0.327540645	34	.
8:10692229:C:T	10692229	0.000127714	0.000255363	C	T	1	0	-1.740323036	2.025811988	0.39029952	32	rs142521930
8:10692232:CA:C	10692232	0	0.000255363	CA	C	0	0	NA	NA	NA	.	.
8:10692283:G:A	10692283	0.000127747	0.000510725	G	A	1	0	-1.789379059	2.018317319	0.375310592	26.4	.
8:10697250:T:A	10697250	0.000127747	0.000510725	T	A	0	1	3.19360382	2.1627751	0.139776678	32	.
8:10697260:T:C	10697260	0.000255493	0.000510725	T	C	1	1	-0.016001327	1.4200209	0.991009329	23	rs200886591
8:10697266:T:A	10697266	0.000255493	0.000510725	T	A	0	2	1.991418163	1.418068329	0.160224327	24.4	.

Table 2. *TREM2* variants included in the CADD15/CADD20 analyses. Table shows base pair (BP, minor allele frequency (MAF), missing rate, major and minor allele, p-value for single-marker analysis along with effect size (beta) and standard error (s.e.), CADD score

SNP	MAF	Missing Rate	MajorAllele	MinorAllele	beta	se	single-marker pvalue	CADD score	ExAC_ALL	snp138
6:41127543:G:A	0.004984663	0.00102145	G	A	0.829210707	0.343249645	0.016	23.1	0.0036	rs2234255
6:41127561:C:T	0.000638733	0.000510725	C	T	0.271610307	0.994602461	0.78	23.2	0.0001	rs79011726
6:41129105:G:A	0.000894912	0.001276813	G	A	-0.070395503	0.884520054	0.94	24.3	0.0003	.
6:41130779:A:G	0.000255754	0.001532176	A	G	-2.145244892	1.430444422	0.13	23.6	.	.

eTable 3. Minor allele frequencies for *PINX1* variants in 1000G and ExAC databases ("_ALL"=cumulative, 1000g2015aug_afr= Africans from 1000 Genomes Project; 1000g2015aug_eur = Europeans from 1000 Genomes Project; ExAC_AFR= Africans/African Americans from Exome Aggregation Consortium; ExAC_NFE= Non-Finnish Europeans from Exome Aggregation Consortium).

VARIANT	1000g2015aug_all	1000g2015aug_eur	1000g2015aug_afr	ExAC_ALL	ExAC_AFR	ExAC_NFE
8:10622930:T:C	**0.000998403**	**.**	**0.0038**	**0.0012**	**0.0133**	**1.71E-05**
8:10623027:G:C	**.**	**.**	**.**	**.**	**.**	**.**
8:10623094:C:G	**0.000199681**	**.**	**0.0008**	**7.50E-05**	**0.0009**	**0**
8:10623104:C:T	**0.00259585**	**0.001**	**.**	**0.0024**	**0.0001**	**0.0011**
8:10623134:G:C	**.**	**.**	**.**	**3.33E-05**	**0.0004**	**0**
8:10623135:C:T	**.**	**.**	**.**	**5.83E-05**	**0.0006**	**0**
8:10623141:T:A	**.**	**.**	**.**	**0.0001**	**0.0001**	**0.0002**
8:10623240:T:C	**0.0319489**	**0.0487**	**0.003**	**0.0653**	**0.014**	**0.0624**
8:10623254:C:A	**0.0161741**	**0.0119**	**0.0015**	**0.0158**	**0.0034**	**0.0123**
8:10623280:C:G	**0.00738818**	**.**	**0.0272**	**0.002**	**0.023**	**1.50E-05**
8:10623301:C:A	**0.000199681**	**0.001**	**.**	**0.0001**	**0**	**0.0002**
8:10623309:C:T	**.**	**.**	**.**	**.**	**.**	**.**
8:10623344:C:T	**.**	**.**	**.**	**5.80E-05**	**0.0001**	**4.50E-05**
8:10623361:C:G	**.**	**.**	**.**	**4.97E-05**	**0.0003**	**0**
8:10623379:G:C	**.**	**.**	**.**	**4.14E-05**	**0**	**7.50E-05**
8:10623383:G:T	**0.000199681**	**.**	**0.0008**	**8.29E-05**	**0.001**	**0**
8:10623385:C:T	**.**	**.**	**.**	**0.0002**	**0.0021**	**1.50E-05**
8:10623386:G:A	**.**	**.**	**.**	**4.14E-05**	**0.0001**	**6.00E-05**
8:10623393:C:T	**.**	**.**	**.**	**1.66E-05**	**0.0002**	**0**
8:10623407:G:C	**.**	**.**	**.**	**.**	**.**	**.**
8:10623423:C:T	**0.000199681**	**.**	**.**	**0.0009**	**0.0001**	**0.0015**
8:10677710:G:A	**0.000399361**	**.**	**0.0015**	**6.96E-05**	**0.0007**	**0**
8:10677806:G:C	**.**	**.**	**.**	**3.84E-05**	**0.0004**	**0**
8:10683689:G:T	**.**	**.**	**.**	**2.10E-05**	**0**	**3.89E-05**
8:10683721:C:G	**.**	**.**	**.**	**1.05E-05**	**0**	**1.93E-05**
8:10683752:A:G	**.**	**.**	**.**	**1.25E-05**	**0.0001**	**0**
8:10689208:C:T	**0.000199681**	**.**	**.**	**8.82E-05**	**0**	**0.0001**
8:10689232:C:A	**.**	**.**	**.**	**3.30E-05**	**0.0002**	**3.11E-05**
8:10690422:C:T	**.**	**.**	**.**	**2.79E-05**	**0**	**3.29E-05**
8:10690474:T:G	**0.00159744**	**0.006**	**.**	**0.0054**	**0.0006**	**0.0084**
8:10690476:G:T	**.**	**.**	**.**	**.**	**.**	**.**
8:10692193:C:T	**.**	**.**	**.**	**.**	**.**	**.**
8:10692229:C:T	**0.000199681**	**0.001**	**.**	**1.66E-05**	**0**	**1.50E-05**
8:10692283:G:A	**.**	**.**	**.**	**3.40E-05**	**0**	**4.61E-05**
8:10697250:T:A	**.**	**.**	**.**	**1.64E-05**	**0**	**2.99E-05**
8:10697260:T:C	**0.000199681**	**.**	**.**	**6.27E-05**	**0.0004**	**2.85E-05**
8:10697266:T:A	**0.000199681**	**.**	**.**	**1.56E-05**	**0**	**0**
8:10622923:TTTC:T	**0.00778754**	**.**	**0.028**	**0.0026**	**0.0229**	**0.0007**
8:10623000:T:TC	**.**	**.**	**.**	**2.50E-05**	**0.0003**	**0**
8:10623137:CTCT:C	**.**	**.**	**.**	**0.0001**	**0.0001**	**0.0002**
8:10623203:GC:G	**.**	**.**	**.**	**.**	**.**	**.**
8:10623206:T:TAA	**.**	**.**	**.**	**.**	**.**	**.**
8:10623396:TCTC:T	**.**	**.**	**.**	**0.0001**	**0**	**7.50E-05**
8:10690414:G:GGTA	**.**	**.**	**.**	**.**	**.**	**.**
8:10692232:CA:C	**.**	**.**	**.**	**.**	**.**	**.**

eTable 4. LOAD known genes VEP MODERATE-HIGH SKAT-O results (Meta-analysis of WHICAP, ADSP, ROS/MAP). In bold genes that show a p-value<=0.05.

GENE	P.value (WHICAP-ADSP-ROS/MAP MODERATE-HIGH SKATO metanalysis)	symbol
ENSG00000095970	6.10E-05	*TREM2*
ENSG00000140090	0.00254	*SLC24A4*
ENSG00000166961	0.008017	*MS4A15*
ENSG00000137642	0.01175	*SORL1*
ENSG00000086288	0.01686	*NME8*
ENSG00000166928	0.01906	*MS4A14*
ENSG00000166927	0.02066	*MS4A7*
ENSG00000100599	0.0216	*RIN3*
ENSG00000108798	0.02523	*ABI3*
ENSG00000166926	0.06901	*MS4A6E*
ENSG00000130203	0.08161	*APOE*
ENSG00000110077	0.08168	*MS4A6A*
ENSG00000156738	0.09308	*MS4A1*
ENSG00000073921	0.09368	*PICALM*
ENSG00000116032	0.1076	*GRIN3B*
ENSG00000110079	0.1186	*MS4A4A*
ENSG00000105383	0.1314	*CD33*
ENSG00000203710	0.1514	*CR1*
ENSG00000064687	0.1788	*ABCA7*
ENSG00000172689	0.1795	*MS4A10*
ENSG00000183580	0.1854	*FBXL7*
ENSG00000149534	0.1936	*MS4A2*
ENSG00000046604	0.207	*DSG2*
ENSG00000120885	0.2243	*CLU*
ENSG00000073712	0.2399	*FERMT2*
ENSG00000198087	0.2796	*CD2AP*
ENSG00000081189	0.2809	*MEF2C*
ENSG00000197943	0.2825	*PLCG2*
ENSG00000182168	0.3067	*UNC5C*
ENSG00000168918	0.3277	*INPP5D*
ENSG00000166959	0.3386	*MS4A8*
ENSG00000204979	0.3583	*MS4A13*
ENSG00000149516	0.4002	*MS4A3*
ENSG00000149187	0.4011	*CELF1*
ENSG00000166930	0.4132	*MS4A5*
ENSG00000087589	0.4216	*CASS4*
ENSG00000071203	0.4274	*MS4A12*
ENSG00000120899	0.4524	*PTK2B*
ENSG00000078487	0.4723	*ZCWPW1*
ENSG00000136717	0.4788	*BIN1*
ENSG00000198502	0.526	*HLA-DRB5*
ENSG00000146904	0.6476	*EPHA1*
ENSG00000196126	0.7991	*HLA-DRB1*

eAppendix 4. Acknowledgment.

WHICAP and EFIGA
Data collection for this project was supported by the Washington Heights and Inwood Community Aging Project (WHICAP) funded by the National Institute on Aging (NIA), by the National Institutes of Health (NIH) (1RF1AG054023, R56AG051876, R21AG054832,), and the National Center for Advancing Translational Sciences, NIH through Grant Number TL1TR001875. We acknowledge the WHICAP study participants and the research and support staff for their contributions to this study.
The Alzheimer's Disease Sequencing Project (ADSP) is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Lindsay Farrer, Jonathan Haines, Richard Mayeux, Margaret Pericak-Vance, and Gerard Schellenberg); U01AG049505 to Dr. Sudha Seshadri; U01AG049506 to Dr. Eric Boerwinkle; U01AG049507 to Dr. Ellen Wijsman; and U01AG049508 to Dr.Alison Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage. Data generation and harmonization in the Follow-up Phases is supported by U54AG052427 to Drs. Schellenberg and Li-San Wang.
The ADGC cohorts include: Adult Changes in Thought (ACT supported by NIA grant U01AG006781 to Drs. Eiic Larson and Palu Crane), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP supported by NIA grant RF1AG054023 to Dr. Mayeux) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA supported by NIA grant RF1AG015473 to Dr. Mayeux), the University of Toronto (UT), and Genetic Differences (GD). Analysis of ADGC cohorts us supported by NIA grants R01AG048927 and RF1AG057519 to Dr. Farrer. Efforts of ADGC investigators were also supported by grants from the NIA (R03AG054936) and National Library of Medicine (R01LM012535).
The CHARGE cohorts are supported in part by National Heart, Lung, and Blood Institute (NHLBI) infrastructure grant HL105756 (Psaty), RC2HL102419 (Boerwinkle) and the neurology working group is supported by the National Institute on Aging (NIA) R01 grant AG033193. The CHARGE cohorts participating in the ADSP include the following: Austrian Stroke Prevention Study (ASPS), ASPS-Family study, and the Prospective Dementia Registry-Austria (ASPS/PRODEM-Aus), the Atherosclerosis Risk in Communities (ARIC) Study, the Cardiovascular Health Study (CHS), the Erasmus Rucphen Family Study (ERF), the Framingham Heart Study (FHS), and the Rotterdam Study (RS). ASPS is funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180 and the Medical University of Graz. The ASPS-Fam is funded by the Austrian Science Fund (FWF) project I904),the EU Joint Programme - Neurodegenerative Disease Research (JPND) in frame of the BRIDGET project (Austria, Ministry of Science) and the Medical University of Graz and the Steiermärkische Krankenanstalten Gesellschaft. PRODEM-Austria is supported by the Austrian Research Promotion agency (FFG) (Project No. 827462) and by the Austrian National Bank (Anniversary Fund, project 15435. ARIC research is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C). Neurocognitive data in ARIC is collected by U01 2U01HL096812, 2U01HL096814, 2U01HL096899, 2U01HL096902, 2U01HL096917 from the NIH (NHLBI, NINDS, NIA and NIDCD), and with previous brain MRI examinations funded by R01-HL70825 from the NHLBI. CHS research was supported by contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, and grants U01HL080295 and U01HL130114 from the NHLBI

with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided by R01AG023629, R01AG15928, and R01AG20098 from the NIA. FHS research is supported by NHLBI contracts N01-HC-25195 and HHSN268201500001I. This study was also supported by additional grants from the NIA (R01s AG054076, AG049607 and AG033040 and NINDS (R01 NS017950). The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by a joint grant from the Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the municipality of Rotterdam. Genetic data sets are also supported by the Netherlands Organization of Scientific Research NWO Investments (175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project 050-060-810. All studies are grateful to their participants, faculty and staff. The content of these manuscripts is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the U.S. Department of Health and Human Services.

The three LSACs are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079).

Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigator institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes (dbGaP) funded by NIH. This research was supported in part by the Intramural Research Program of the National Institutes of health, National Library of Medicine. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

ROSMAP. The Religious Orders Study and Rush Memory and Aging Project are supported by NIA grants P30AG10161, R01AG15819, R01AG17917, R0136836, and U01AG46152.

Biogen Inc. provided support for whole exome sequencing for the WHICAP cohort through a grant to David Goldstein, PhD. Individuals at Biogen were not involved in the collection of data, analysis or interpretation of the genetic data, nor in the production of this manuscript.

eReferences. REFERENCES.

1. Carson AR, Smith EN, Matsui H, et al. Effective filtering strategies to improve data quality from population-based whole exome sequencing studies. *BMC bioinformatics*. 2014;15(1):125.

2. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*. 2016;32(11):1749-1751.

3. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics*. 2015;31(13):2202-2204.

4. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939-939.

5. Beecham GW, Bis J, Martin E, et al. The Alzheimer's Disease Sequencing Project: Study design and sample selection. *Neurology Genetics*. 2017;3(5):e194.

6. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious Orders Study and Rush Memory and Aging Project. *Journal of Alzheimer's Disease*. 2018;(Preprint):1-28.

7. Vardarajan BN, Barral S, Jaworski J, et al. Whole genome sequencing of Caribbean Hispanic families with late-onset Alzheimer's disease. *Annals of clinical and translational neurology*. 2018;5(4):406-417.

8. Naj AC, Jun G, Beecham GW, et al. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature genetics*. 2011;43(5):436-441.

9. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. *Nature genetics*. 2007;39(12):1494.

10. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nature neuroscience*. 2018;21(6):811.

11. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127.

12. White CC, Yang H-S, Yu L, et al. Identification of genes associated with dissociation of cognitive performance and neuropathological burden: Multistep analysis of genetic, epigenetic, and transcriptional data. *PLoS medicine*. 2017;14(4):e1002287.