**A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease**

Douglas P. Wightman, Iris E. Jansen, Jeanne E. Savage, Alexey A. Shadrin, Shahram Bahrami, Dominic Holland, Arvid Rongve, Sigrid Børte, Bendik S. Winsvold, Ole Kristian Drange, Amy E. Martinsen, Anne Heidi Skogholt, Cristen Willer, Geir Bråthen, Ingunn Bosnes, Jonas Bille Nielsen, Lars G. Fritsche, Laurent F. Thomas, Linda M. Pedersen, Maiken E. Gabrielsen, Marianne Bakke Johnsen, Tore Wergeland Meisingset, Wei Zhou, Petroula Proitsi, Angela Hodges, Richard Dobson, Latha Velayudhan, 23andMe Research Team, Julia M. Sealock, Lea K. Davis, Nancy L. Pedersen, Chandra A. Reynolds, Ida K. Karlsson, Sigurdur Magnusson, Hreinn Stefansson, Steinunn Thordardottir, Palmi V. Jonsson, Jon Snaedal, Anna Zettergren, Ingmar Skoog, Silke Kern, Margda Waern, Henrik Zetterberg, Kaj Blennow, Eystein Stordal, Kristian Hveem, John-Anker Zwart, Lavinia Athanasiu, Per Selnes, Ingvild Saltvedt, Sigrid B. Sando, Ingun Ulstein, Srdjan Djurovic, Tormod Fladby, Dag Aarsland, Geir Selbæk, Stephan Ripke, Kari Stefansson, Ole A. Andreassen & Danielle Posthuma

**Abstract**

Late-onset Alzheimer's disease is a prevalent age-related polygenic disease that accounts for 50–70% of dementia cases. Currently, only a fraction of the genetic variants underlying Alzheimer's disease have been identified. Here we show that increased sample sizes allowed identification of seven previously unidentified genetic loci contributing to Alzheimer's disease. This study highlights microglia, immune cells and protein catabolism as relevant to late-onset Alzheimer's disease, while identifying and prioritizing previously unidentified genes of potential interest. We anticipate that these results can be included in larger meta-analyses of Alzheimer's disease to identify further genetic variants that contribute to Alzheimer's pathology.

**Main**

Dementia has an age- and sex-standardized prevalence of ~7.1% in Europeans[1], with Alzheimer's disease (AD) being the most common form of dementia (50–70% of cases)[2]. AD is pathologically characterized by the presence of amyloid-beta plaques and tau neurofibrillary tangles in the brain[3]. Most patients are diagnosed with AD after the age of 65, termed late-onset AD (LOAD), while only 1% of AD cases have an early onset (before the age of 65)[3]. On the basis of twin studies, the heritability of LOAD is estimated to be ~60–80% (refs. 4,5), suggesting that a large proportion of individual differences in LOAD risk is driven by genetics. The heritability of LOAD is spread across many genetic variants; however, Zhang et al.[6] suggested that LOAD is more of an oligogenic than a polygenic disorder due to the large effects of APOE variants. Zhang et al.[6] and Holland et al.[7] predicted there to be ~100–10,000 causal variants contributing to LOAD; however, only a fraction have been identified. Increasing the sample size of genome-wide association studies (GWAS) will improve the statistical power to identify the missing causal variants and may highlight additional disease mechanisms. In combination with increasing the number of samples, it is beneficial to use different approaches to identify rare and private variation to help identify additional causal variants and increase understanding of disease mechanisms; however, we deem this to be out of the scope of the current analysis.

The largest previous GWAS of LOAD, identified 29 risk loci from 71,880 (46,613 proxy) cases and 383,378 (318,246 proxy) controls[8]. Our current study expands this to include 90,338 (46,613 proxy) cases and 1,036,225 (318,246 proxy) controls. The recruitment of LOAD cases can be difficult due to the late age of onset, so proxy cases can allow for the inclusion of younger individuals by estimating their risk of LOAD using parental status. Proxy cases and controls were defined on the basis of known parental LOAD status weighted by parental age (Supplementary Information). In the current study, we identified 38 loci, including seven loci that have not been reported previously. Functional follow-up analyses implicated tissues, cell types and genes of interest through tissue and cell type enrichment, colocalization and statistical fine-mapping. This study highlights microglia, immune cells and protein catabolism as relevant to LOAD, while identifying previously unidentified genes of potential interest.

**Results**

Genome-wide inferences

We performed meta-analysis on data from 13 cohorts, totaling 1,126,563 individuals (Supplementary Table 1). The inflation factors and linkage disequilibrium score (LDSC) regression[9] intercepts of each dataset are reported in Supplementary Table 2. The liability scale SNP heritability was estimated by LDSC regression[9] to be 0.031 (s.e. = 0.0062) given a population prevalence of 0.05 (UK Biobank (UKB) data excluded). This estimate is low but similar to the estimates obtained in previous GWAS meta-analyses (Jansen et al.[8]: liability scale heritability = 0.055, s.e. = 0.0099; Lambert et al.[10]: observed scale heritability = 0.069, s.e. = 0.013). The LDSC intercept was 1.022 (s.e. = 0.013), the inflation factor ($\lambda$) for the meta-analysis was 1.11 and the sample size-adjusted inflation factor ($\lambda_{1,000}$)[11] was 1.007. The genetic correlation[12] between proxy LOAD and case–control LOAD was 0.83 (s.e. = 0.21, $P = 6.61 \times 10^{-5}$). Separate Manhattan plots for the LOAD proxy data and the case–control LOAD data are available in Supplementary Figs. 1 and 2. Across 855 external phenotypes in LD Hub[13], two significant genetic correlations with the meta-analysis results were observed, both of which were identified in previous studies of LOAD (Supplementary Information and Supplementary Table 3).

The meta-analysis identified 3,915 significant ($P < 5 \times 10^{-8}$) variants across 38 independent loci (Table 1 and Fig. 1). Of those 38 loci, seven have not shown associations with LOAD in previous GWAS, and five of those loci have not been associated with any form of dementia (AGRN, TNIP1, HAVCR2, NTN5, LILRB2). The lead variant effect estimates and significance values per dataset for each locus are reported in Supplementary Table 4. We largely replicated the loci identified in Jansen et al.[8], however, seven loci were not found to be genome-wide significant in this study, five of those were just below significance and two were driven by rare variants (largely) not included in this study (Supplementary Information and Supplementary Table 5). However, we successfully replicated all the significant loci in Kunkle et al.[14] (Supplementary Table 6).

Only variants with $P < 0.0005$ are displayed. The APOE region cannot be fully observed because the y axis is limited to the top variant in the second most significant locus, $-\log 10(1 \times 10^{-60})$, to display the less significant variants. The red line represents genome-wide significance ($5 \times 10^{-8}$). The P values were identified through a meta-analysis (two-sided test) of summary statistics generated by linear/logistic regressions (two-sided test) and were not adjusted for multiple testing. The previously unidentified loci are highlighted in green and indicated by the assigned gene name. The

TNIP1/HAVCR2 regions and the NTN5/LILRB2 regions are close enough together that they cannot be visually distinguished at this scale, but are different genomic risk loci.

MAGMA tissue specificity analysis[15] identified spleen ($P_{Bonferroni}$ = 0.034) as the only Genotype-Tissue Expression (GTEx) tissue, where expression of the MAGMA genes was significantly associated (Supplementary Fig. 3 and Supplementary Table 7). However, this tissue was slightly above the significance threshold ($P_{Bonferroni}$ = 0.054) when the larger APOE region (GRCh37: 19:40,000,000–50,000,000) was excluded (Supplementary Table 7). Spleen was also significant in the previous MAGMA tissue specificity analysis performed in Jansen et al.[8] and is a known contributor to immune function. To investigate enrichment at the cell type level, FUMA cell type analysis[16] was performed with a collection of cell types in mouse brain, human brain and human blood tissue. Six single-cell RNA-sequencing (scRNA-seq) datasets were significantly associated, after multiple testing correction, with the expression of LOAD-associated genes (Supplementary Fig. 4 and Supplementary Table 8). Microglia were the only significant cell type in all six independent scRNA-seq datasets. We confirm previously observed enrichment for nonhuman microglial cells[8], and report additional similar enrichments in human microglia. Four of these enrichments remained significant after exclusion of the larger APOE region, suggesting that genomic regions outside of the APOE region play a substantial role in the microglia finding. A combination of the cell type and tissue specificity results identifies microglia and immune tissues as potential experimental models for identifying the contribution of LOAD-associated genes towards LOAD pathogenesis.

MAGMA gene-set analysis[15] identified 25 gene ontology biological processes (Supplementary Table 9) that were significantly enriched, after multiple testing correction, for LOAD-associated variants. Subsequent conditional gene-set analyses confirmed independent association of four of these 25 gene sets, reflecting the role of LOAD-associated genes in amyloid and tau plaque formation, protein catabolism of plaques, immune cell recruitment and glial cells (Supplementary Table 9). The exclusion of the larger APOE region resulted in the loss of five significant gene sets related to amyloid-beta clearance, phospholipid efflux, cholesterol transport, protein lipid interactions and tau binding, and the gain of two significant gene sets related to tau degradation and astrocyte activation (Supplementary Table 9). Conditional gene-set analysis, with the larger APOE region excluded, identified four independent gene sets related to astrocyte activation, immune cell recruitment, amyloid catabolism and neurofibrillary tangles. The gene-set related to glial cells was still significant after removal of the APOE region, but was not identified as an independent gene set, which suggests that this association can be explained by the APOE region in addition to another significant independent gene set. Largely, the themes highlighted in the gene-set analysis are robust to the exclusion of the APOE region. Our gene-set analysis identified the same themes as Jansen et al.[8] and further identified significant gene sets involved in immune cell recruitment and neuronal cell types.

Gene prioritization

As expected, the genomic risk loci identified in this study were enriched for active chromatin and variant annotations relating to gene function (Supplementary Information). We performed functional follow-up (colocalization and fine-mapping) to further dissect the genomic risk loci to identify potential disease drivers. Functional mapping of variants to genes on the basis of position and expression quantitative trait loci (eQTL) information from brain and immune tissues/cells identified 989 genes that mapped to one of the 38 genomic risk loci (Supplementary Table 10). These mapped genes were annotated with the drugs that target them on the basis of information from DrugBank[17].

Due to linkage disequilibrium (LD) and the inability to distinguish true causal variants from variants in LD, many of the mapped genes may be functionally irrelevant to LOAD. To highlight potentially relevant genes, eQTL data from immune tissues, brain and microglia were colocalized with the genomic risk loci using Coloc[18]. We used the 19 successful colocalizations (Supplementary Table 11) for nine genes (TNIP1, MADD, APH1B, GRN, AC004687.2, ACE, NTN5, CD33 and CASS4) to prioritize genes in those loci. Statistical fine-mapping with susieR was additionally performed to narrow down the associated region (Supplementary Table 12). The statistical fine-mapping required an external reference panel, which limits the interpretation of the findings, so only high confidence variants (posterior inclusion probability (PIP) in a credible set >0.95) were considered in gene prioritization. Gene prioritization of the previously unidentified loci, and a description of colocalization and fine-mapping evidence for previously identified loci, is available in the Supplementary Information. Some of the most interesting findings for the previously unidentified loci are highlighted below.

The lead variant of locus 7 (rs871269; $P = 1.37 \times 10^{-9}$; minor allele frequency (MAF) = 0.34) is located in an intron of TNIP1 (Supplementary Fig. 5) and maps to GPX3, TNIP1 and SLC36A1 on the basis of eQTLs within blood tissue. The lead variant is supported by a few variants with suggestive signal (rs34294852; $P = 1.05 \times 10^{-6}$) but none of these variants are in LD ($R^2 > 0.1$) in the 1000 Genomes (1KG) European (EUR) population. However, these variants are in moderate/low LD with the lead variant ($R^2 = 0.2$–0.6) in the 1KG East Asian (EAS) and American populations. This suggests that the 1KG EUR reference panel does not accurately represent the LD structure of our data at this locus. The fine-mapping results from susieR identified the lead variant as the only variant with high PIP (>0.99). However, the association signal in this locus colocalized with a nearby suggestive variant (rs34294852; $R^2 = 0.29$ in 1KG EAS); this variant is an eQTL for TNIP1 in blood tissue (TwinsUK). Support from previous literature is sparse; however, TNIP1 has the most support of the three genes. TNIP1 contributes to hyperinflammation and has been previously identified in an autoimmune GWAS[19]. TNIP1 was included in a transcription module regulated by Bcl3 in mouse microglia[20], where this module was implicated in prolonged exposure to inflammation and aging of microglia. The gene encoding Bcl3 (BCL3) was found to be significantly associated with cerebrospinal fluid amyloid-beta1–42 peptide after conditioning for APOE[21] and was observed as upregulated postmortem in the brain of patients with LOAD[22]. Further investigation into this locus in nonEuropean populations may yield more support for the lead variant and improve the fine-mapping analysis.

The lead variant of locus 8 (rs6891966; $P = 7.91 \times 10^{-10}$) is located in an intron of HAVCR2 (Supplementary Fig. 6). HAVCR1 and TIMD4 also map to this region on the basis of brain eQTLs (PsychENCODE). HAVCR2 was significantly differentially expressed in bulk brain tissue of patients with LOAD compared to controls[23]. HAVCR2 is preferentially expressed in aged microglia[24], was included as one of the top 100 enriched transcripts in brain and microglia and was included in a cluster of transcripts that are involved in sensing endogenous ligands and microbes[25]. The protein encoded by HAVCR2 (Havcr2) has been suggested to bind to phosphatidylserine on cell surfaces to mediate apoptosis[26] and to interact with amyloid precursor protein[27]. TIMD4 is another gene in this region that encodes a protein (TIM-4) with a similar function to Havcr2; it binds to phosphatidylserine on cell surfaces to mediate apoptosis, and microglia without TIM-4 receptors have reduced apoptotic clearance[28]. Follow-up experimental work would be useful to determine the role that these genes play within LOAD.

Locus 12 and locus 28 have been previously associated with dementia[29] but not within a previous LOAD GWAS. The lead variant in locus 12 (rs5011436; $P = 2.7 \times 10^{-9}$) is an intron variant in TMEM106B (Supplementary Fig. 7). A nearby exonic variant (rs3173615; $R^2 = 0.976$ in 1KG EUR; $P = 6.61 \times 10^{-9}$) with a CADD score of 21.2 has been discussed as the association signal-driving

variant in frontotemporal dementia (FTD) by causing decreased transmembrane protein 106B (the protein encoded by TMEM106B) abundance through increased protein degradation30. TMEM106B was also found to be significantly differentially expressed in bulk brain tissue of patients with LOAD compared to controls23. The lead variant in locus 28 (rs708382; P = 1.98 × 10−9) is an upstream variant of FAM171A2 (Supplementary Fig. 8). Interestingly, the protein (integrin alpha-IIb) encoded by a nearby gene (ITGA2B) is a target for abciximab, an antibody that inhibits platelet aggregation and is used to estimate concentrations of coated platelets31. In patients with mild cognitive impairments, elevated coated-platelet levels are linked to increased risk of LOAD progression. However, the association signal in this locus colocalized with an eQTL for GRN in brain tissue (ROSMAP and BrainSeq) with the lead variant identified as the colocalized variant. GRN is also a known FTD gene32 and has the most evidence for being the causal gene in the region. The association signals in locus 12 and locus 28 do not appear to be primarily driven by the UKB data (Supplementary Information), which suggests that the associations of the known FTD genes are not driven by the proxy phenotype. These results suggest that TMEM106B and GRN are not solely contributing to FTD, but also contributing to LOAD, suggesting that their biological implications might be related to protein clearance mechanisms rather than involvement in specific disease-related protein aggregates.

The lead variant of locus 36 (rs1761461, P = 1.56 × 10−9) is an intergenic variant upstream of LILRA5 (Supplementary Fig. 9). The lead variant is an eQTL for LILRA5, LILRP2, LILRB1 and LILRA4 in GTEx whole blood. These genes encode a family of transmembrane glycoproteins that mediate immune activation33. LILRB5, LILRA5 and LILRB2 were significantly differentially expressed in bulk brain tissue of patients with LOAD compared to controls23. Interestingly, LILRB2 is a nearby gene in the same family and encodes a protein (leukocyte immunoglobulin-like receptor B2) that is known to inhibit axonal regeneration and to contribute to LOAD through amyloid binding33. The role of LILRB2 in LOAD has been investigated in mouse models and results suggest that drug-targeting this gene could be a beneficial treatment approach34. While prioritizing this region to a single gene is difficult, the LILR family appears to be the most likely candidate for explaining the association signal.


## Discussion

We performed a large GWAS for LOAD, including 1,126,563 individuals, and identified 38 LOAD-associated loci, including seven previously unidentified loci. The data included both clinical cases and proxy cases, defined on the basis of parental LOAD status, a strategy that was validated previously by ourselves8 and others35. Through gene-set analysis, tissue and single-cell specificity analysis, colocalization and fine-mapping, this study highlighted additional biological routes that connect genetic variants to LOAD pathology. These functional analyses all implicated immune cells and microglia as cells of interest, which provided genetic support to the current understanding of LOAD pathology36. The seven previously unidentified loci were functionally annotated and fine-mapped to help narrow down candidate causal genes. Two of the previously unidentified loci have been previously associated with FTD29. This signal is not driven by the nonmedically verified LOAD cases in the UKB proxy LOAD data (Supplementary Information), which suggests that this region is pleiotropic for FTD or contains separate causal variants within the same LD blocks.

A recent study7 produced a power curve for LOAD using a model that accounts for large and small effect variants. This model was based on summary statistics from a previous GWAS of LOAD10. A sample size of 2.2 million is predicted to identify 80% of genetic variance on chromosome 19, and a sample size of 7.8 million is predicted to identify 80% genetic variance outside chromosome 19. The

effective sample size35 of our meta-analysis was ~169,608, so, on the basis of previous power estimates, our study was powered to explain ~6% of genetic variance outside chromosome 19 and 58.9% of genetic variance on chromosome 19 (Supplementary Fig. 10). We demonstrated that an increased sample size in a GWAS meta-analysis approach allowed for identification of previously unidentified loci; however, Holland et al.7 also predicted there to be approximately 300 large effect causal variants contributing to LOAD. These large effect variants (and small effect rare variants) are unlikely to be identified through traditional GWAS approaches focusing on common variants. Larger sample size GWAS approaches should be complemented with rare variant, copy number variant and private variant discovery to identify the remaining causal variants.

Future work focusing on fine-mapping, generating larger QTL databases in more specific cells types and incorporating other ancestries will improve the interpretability of associated loci. Our colocalization analysis identified a candidate causal gene in nine of the 38 loci and we expect that larger and more specific QTL datasets will improve the number of successful colocalizations. Yao et al.37 highlighted a need for higher sample size eQTL discovery and suggested that genes with smaller effect eQTLs are more likely to be causal for common traits. The identification of human microglia, but not bulk brain tissue, as a cell/tissue type of interest in this study supported a finding in a recent single-cell epigenomic study38, which showed that investigating individual cell types will be more fruitful than bulk brain tissue for understanding the route from variant to LOAD pathology.

One important goal for LOAD GWAS is the identification of medically actionable information that can help in diagnosis or treatment in all populations. This study was limited in the ability to identify causal genes and in the applicability to nonEuropean populations. Further study in nonEuropean populations will improve the equity of genetic information and also help with fine-mapping of associated regions. Larger sample sizes of GWAS, epigenomic studies and eQTL studies in all populations will improve identification and explanation of additional LOAD loci while increasing the applicability of these findings to a larger group of individuals. This could be accomplished by a push for facilitating data sharing and global collaboration within the field of Alzheimer's disease genetics. The current work provided genetic support for the role of immune cells and microglia in LOAD, identified previously unidentified LOAD-associated regions, prioritized causal genes of interest and highlighted the importance of collaboration to discern the biological process that mediates LOAD pathology.

**Methods**

Quality control and meta-analysis

The data from the participants in this study were obtained from freely available summary statistics and from genotype-level data. Additional cohorts have been obtained since our previous analysis8 (as well as an increased deCODE sample); these cohorts contain 12,968 additional cases and 488,616 additional controls. An overview of the cohorts is available in Supplementary Table 1. Informed consent was obtained from all participants and we complied with all relevant ethical regulations. A full description of each dataset, the quality control (QC) procedures and the analysis protocols are available in the Supplementary Information. In short, each dataset underwent initial QC, imputation, logistic/linear regression with at least sex and principal components as covariates and post-regression QC of the summary statistics using EasyQC39. If necessary, the data were converted to build GRCh37 before QC using the UCSC LiftOver tool40. During post-regression QC, each dataset was matched to the Haplotype Reference Consortium (HRC) or 1KG reference panel, and variants with absolute allele frequency differences >0.2 compared to the reference panel were removed. Variants

with an imputation quality score <0.8, minor allele count (MAC) <6, N <30 or absolute beta or s.e. >10 were removed. Low MAF variants were removed; low MAF41 was defined as <12×Nv. All datasets were meta-analyzed using mv-GWAMA (https://github.com/Kyoko-wtnb/mvGWAMA), a sample size weighted method previously developed in Jansen et al.8. The option to account for overlapping individuals was not utilized because no datasets were expected to contain overlapping samples and the estimates of overlapping samples (genetic covariance intercepts) were unreliable due to low heritability of the datasets. The effective sample size of the full meta-analysis for power estimates was calculated by assuming the individuals in the UKB proxy data with phenotype values <1 are controls and ≥1 are cases.

Genomic risk loci definition

We used FUMA v.1.3.6a (ref. 42) (http://fuma.ctglab.nl) to annotate and functionally map variants included in the meta-analysis. Genomic risk loci were defined around significant variants (<5 × 10−8); the genomic risk loci included all variants correlated (R2 > 0.6) with the most significant variant. The correlation estimates were defined using 1KG European reference information43. The 1KG European reference panel was chosen over the UKB44 10K reference panel because the meta-analysis included individuals from a range of European ancestries and this diversity would be better reflected in the 1KG European sample than the primarily British UKB sample. Genomic risk loci within 250 kilobases (kb) of each other are incorporated into the same locus. Previously unidentified genomic risk loci are loci that do not overlap with variants identified as significant in previous studies of LOAD8,10,45,46,47,48,49,50. Regional plots were generated using LocusZoom51 and 1KG reference information.

Heritability and genetic correlation

LDSC regression9 (https://github.com/bulik/ldsc) was used to estimate the liability scale heritability of the nonproxy LOAD meta-analysis (UKB data excluded). The nonproxy LOAD meta-analysis (43,725 cases and 717,979 controls) was performed in the same way as the full meta-analysis described above. The UKB data (N = 364,859) were excluded because LDSC liability scale heritability estimates are sensitive to sample prevalence and the UKB data were generated with a continuous phenotype and therefore a sample prevalence could not be perfectly estimated if the UKB data were included. Heritability estimates were converted to a liability scale using the LOAD population prevalence of 0.05 and a sample prevalence of 0.0574041885. LDSC12 was also used to determine the genetic correlation between a meta-analysis of the nonproxy LOAD datasets and the UKB proxy LOAD dataset. Precalculated LD scores for LDSC were derived from the 1KG European reference population (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2). Heritability and genetic correlation estimates were calculated using HapMap3 variants only. Further genetic correlations were determined using the full meta-analysis and LD Hub13 (http://ldsc.broadinstitute.org/), where all 855 traits were tested using the HapMap3 variants (http://ldsc.broadinstitute.org/static/media/w_hm3.noMHC.snplist.zip). The heritability estimate of Lambert et al.10 summary statistics was obtained from LD Hub.

Gene-based and gene-set analyses

Genome-wide gene association analysis was performed using MAGMA v.1.08 (ref. 15) (http://ctg.cncr.nl/software/magma). All variants in the GWAS outside the major histocompatibility complex (MHC) region (GRCh37: 6:28,477,797–33,448,354) that positionally map within one of the 19,019 protein coding genes were included to estimate the significance value of that gene. Genes were considered significant if the P value was <0.05 after Bonferroni correction for 19,019 genes. All MAGMA analyses utilized 1KG43 LD information. MAGMA gene-set analysis was performed where variants map to 15,496 gene sets from the MSigDB v.7.0 database52. Gene sets were considered significant if the P value was <0.05 after Bonferroni correction for the number of tested gene sets. Forward selection of significantly associated gene sets was performed using MAGMA v.1.08 conditional analysis53. Initially the most significant gene set was selected as a covariate and the remaining gene sets were analyzed. The most significant gene set from this conditional analysis was added as a covariate in addition to the previous gene set, and a new analysis was run. This process was repeated until no gene set met the significance threshold (PBonferroni < 0.05). MAGMA tissue specificity analysis was performed in FUMA using 30 general tissue type gene expression profiles (from GTEx v.8). Tissues were considered significant if the P value was <0.05 after Bonferroni correction for 30 tissues.

FUMA cell type specificity analysis16 utilizes the MAGMA gene association results to identify cell types enriched in expression of trait-associated genes. We focused on brain and immune-related cell types, with the inclusion of pancreas as a control, therefore selecting the following scRNA-seq datasets: Allen_Human_LGN_level1 (ref. 54), Allen_Human_LGN_level2 (ref. 54), Allen_Human_MTG_level1 (ref. 54), Allen_Human_MTG_level2 (ref. 54), DroNc_Human_Hippocampus55, DroNc_Mouse_Hippocampus55, GSE104276_Human_Prefrontal_cortex_all_ages56, GSE67835_Human_Cortex57, GSE81547_Human_Pancreas58, Linnarsson_GSE101601_Human_Temporal_cortex59, MouseCellAtlas_all60, PBMC_10x_68k61 and PsychENCODE_Adult62. Within-dataset corrected results were reported to indicate which single cells are most likely to be disease relevant. The gene-based and gene-set analyses were also performed without the larger APOE region (19:40,000,000–50,000,000).

Gene mapping

The individual genomic risk loci were mapped to genes using FUMA v.1.3.6a42 using positional mapping and eQTL mapping. For positional mapping, all variants within 10 kb of a gene in the genomic risk locus were assigned to that gene. For eQTL mapping, variants were mapped to genes on the basis of significant eQTL interactions in a collection of immune and brain tissues. Brain tissue eQTLs were used due to the importance of brain tissue in LOAD pathology, and immune tissue/cell eQTLs were used for gene mapping because MAGMA tissue specificity analysis highlighted immune tissues as tissues of interest. The brain and immune tissues eQTLs used for mapping were: Alasoo naive macrophage63, BLUEPRINT monocyte64, BLUEPRINT neutrophil64, BLUEPRINT T-cell64, BrainSeq Brain65, CEDAR B-cell66, CEDAR monocyte, CEDAR neutrophil66, CEDAR T-cell66, Fairfax B-cell67, Fairfax naive monocyte68, GENCORD T-cell69, Kasela CD4 T-cell70, Kasela CD8 T-cell70, Lepik Blood71, Naranbhai neutrophil72, Nedelec macrophage73, Quach monocyte74, Schwartzentruber sensory neuron75, TwinsUK blood76, PsychENCODE brain62, eQTLGen blood cis and trans77, BloodeQTL blood78, BIOS Blood79, xQTLServer blood80, CommonMind Consortium brain81, BRAINEAC brain82, GTEX v.8 lymphocytes, brain, spleen and whole blood. The genes that mapped to

previously unidentified loci were searched in a database (https://diegomscoelho.github.io/AD-IsoformSwitch/index.html)23 to identify whether they were differentially expressed in bulk brain tissue of patients with LOAD compared to controls.

Colocalization

All variants within 1.5 megabases (Mb) of the lead variant of each genomic risk loci were used in the colocalization analysis. The GWAS data and eQTL data were trimmed so that all variants overlap. Colocalization was performed per gene using coloc.abf from the Coloc R package18. Default priors were used for prior probability of association with the GWAS data and eQTL data. The prior probability of colocalization was set as $1 \times 10^{-6}$ as recommended83. Nominal P, sample size and MAF from the GWAS data and eQTL data were used in all the colocalization analyses. Colocalizations with a posterior probability >0.8 were considered successful colocalizations. eQTL data from all tissues except microglia were obtained from the eQTL catalog84. The microglia data were obtained from Young et al.85.

Fine-mapping

Fine-mapping was performed with susieR v.0.9.1 (ref. 86) on all variants within 1.5 Mb of the lead variant of each genomic risk loci. The APOE and HLA-DRB1 (MHC) regions were excluded from fine-mapping due to the complicated LD structure. The sample size of the fine-mapping reference panel should be proportional to the sample size of the data being fine-mapped. A good-sized reference panel is 10% to 20% of the sample size of the data87. UKB data were used as a reference panel for the fine-mapping because it had the largest sample size of the available reference panels and was the only available European reference panel to fulfill the criteria for a good-sized reference panel. The reference panel was ~10% of the size of the GWAS data. An LD matrix was generated using 100,000 individuals in R v.3.4.3 (ref. 88). The 100,000 individuals were chosen for each locus as the top 100,000 people with the most genotyped variants in the locus to maintain the highest number of variants in the fine-mapping. Only the top 100,000 were chosen for computational feasibility and to maintain as many variants as possible while having a large reference panel. The meta-analysis data were trimmed to match the variants included in the LD reference. The maximum number of causal variants in the region was set to 10. The susieR credible sets are reported in Supplementary Table 12. The allele frequency in the UKB data and meta-analysis data of all the variants in the fine-mapping analyses were compared to identify outliers. No variants included in the confidence set or credible set had an allele frequency difference >0.2.

Functional enrichment of significantly associated regions

All enrichment analyses were performed using a Fisher's exact test (fisher.test) implemented in R v.4.0.1 (ref. 88). The enrichment analyses compared all variants within the genomic risk loci (excluding the MHC region; GRCh37: 6:28,477,797–33,448,354) to all other variants present in the meta-analysis (excluding of the MHC region). Enrichment of active chromatin was performed using ROADMAP Core 15-state model annotation89 obtained from https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/all.mnemonics.bedFiles.tgz. For each of the 127 cell types, all variants within the analysis were annotated with one of the 15 states using the R package Genomic Ranges90. All

variants annotated with a state <8 were defined as being within active chromatin. The enrichment of active chromatin within the specified region was performed for each of the cell types and the resulting P values were corrected for 127 tests using Bonferroni correction. To perform enrichments of functional consequences, variants were annotated with ANNOVAR91 using ANNOVAR and FASTA sequences for all annotated transcripts in RefSeq Gene92. Enrichments were considered significant if the P value was <0.05 after Bonferroni correction for 11 functional consequences. The enrichment plots were generated using the R package ggplot2 (ref. 93).

## Statistics and reproducibility

No statistical method was used to predetermine sample size, all available datasets were included in the meta-analysis. Exclusion of data was predetermined and based on quality control procedures outlined in the Supplementary Information. Phenotype values were assigned on the basis of (parental) diagnoses so the experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. Scientific findings were compared to findings from previous LOAD meta-analyses. Replication of previously identified loci is reported in the text and in the Supplementary Information.

## Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Access to raw data can be requested via the Psychiatric Genomics Data Access portal (https://www.med.unc.edu/pgc/shared-methods/open-source-philosophy/), UK Biobank (www.ukbiobank.ac.uk), or 23andMe. Restriction of raw data is to protect the privacy of participants. Summary statistics from IGAP (https://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php) and Finngen (https://www.finngen.fi/en/access_results) can be obtained from their respective online portals. Summary statistics from the meta-analysis excluding 23andMe are available at https://ctg.cncr.nl/software/summary_statistics. Access to the full set, including 23andMe results, can be obtained after approval from 23andMe is presented to the corresponding author. Approval can be obtained by completion of a Data Transfer Agreement. The Data Transfer Agreement exists to protect the privacy of 23andMe participants. Please visit https://research.23andme.com/dataset-access/ to initiate a request. Summary statistics of the primary microglia eQTLs are also available from EGA (accession no.: EGAD00001005736). MSigDB gene sets are available online (https://www.gsea-msigdb.org/gsea/msigdb/) and integrated in FUMA (https://fuma.ctglab.nl/).

## Code availability

The code used to perform the analyses is available at https://github.com/dwightman/PGC-ALZ2. All software used in the analyses is freely available online.

References

1.

Bacigalupo, I. et al. A systematic review and meta-analysis on the prevalence of dementia in Europe: estimates from the highest-quality studies adopting the DSM IV diagnostic criteria. J. Alzheimers Dis. 66, 1471–1481 (2018).

PubMed

PubMed Central

Article

Google Scholar

2.

Winblad, B. et al. Defeating Alzheimer's disease and other dementias: a priority for European science and society. Lancet Neurol. 15, 455–532 (2016).

PubMed

Article

Google Scholar

3.

DeTure, M. A. & Dickson, D. W. The neuropathological diagnosis of Alzheimer's disease. Mol. Neurodegener. 14, 32 (2019).

PubMed

PubMed Central

Article

Google Scholar

4.

Gatz, M. et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. J. Gerontol. A Biol. Sci. Med. Sci. 52, M117–M125 (1997).

CAS

PubMed

Article

Google Scholar

5.

Gatz, M. et al. Role of genes and environments for explaining Alzheimer disease. Arch. Gen. Psychiatry 63, 168–174 (2006).

PubMed

Article

Google Scholar

6.

Zhang, Q. et al. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. Nat. Commun. 11, 4799 (2020).

CAS

PubMed

PubMed Central

Article

Google Scholar

7.

Holland, D. et al. The genetic architecture of human complex phenotypes is modulated by linkage disequilibrium and heterozygosity. Genetics 217, iyaa046 (2021).

PubMed

Article

PubMed Central

Google Scholar

8.

Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. 51, 404–413 (2019).

CAS

PubMed

PubMed Central

Article

Google Scholar

9.

Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015).

CAS

PubMed

PubMed Central

Article

Google Scholar

10.

Lambert, J.-C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458 (2013).

CAS

PubMed

PubMed Central

Article

Google Scholar

11.

de Bakker, P. I. W. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).

PubMed

PubMed Central

Article

CAS

Google Scholar

12.

Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).

CAS

PubMed

PubMed Central

Article

Google Scholar

13.

Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33, 272–279 (2016).

PubMed

PubMed Central

Article

CAS

Google Scholar

14.

Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat. Genet. 51, 414–430 (2019).

CAS

PubMed

PubMed Central

Article

Google Scholar


15.

de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015).


PubMed


PubMed Central


Article


CAS


Google Scholar


16.

Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P. & Posthuma, D. Genetic mapping of cell type specificity for complex traits. Nat. Commun. 10, 3222 (2019).


CAS


PubMed


PubMed Central


Article

Google Scholar

17.

Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018).

CAS

PubMed

Article

PubMed Central

Google Scholar

18.

Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).

PubMed

PubMed Central

Article

CAS

Google Scholar

19.

Shamilov, R. & Aneskievich, B. J. TNIP1 in autoimmune diseases: regulation of Toll-like receptor signaling. J. Immunol. Res. 2018, 3491269 (2018).

PubMed

PubMed Central

Article

CAS

Google Scholar

20.

Cho, C. E. et al. A modular analysis of microglia gene expression, insights into the aged phenotype. BMC Genomics 20, 164 (2019).

PubMed

PubMed Central

Article

Google Scholar

21.

Nho, K. et al. Association analysis of rare variants near the APOE region with CSF and neuroimaging biomarkers of Alzheimer's disease. BMC Med. Genomics 10, 29 (2017).

PubMed

PubMed Central

Article

Google Scholar

22.
Li, X. et al. Systematic analysis and biomarker study for Alzheimer's disease. Sci. Rep. 8, 17394 (2018).

PubMed

PubMed Central

Article

CAS

Google Scholar

23.
Marques-Coelho, D. et al. Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. NPJ Aging Mech. Dis. 7, 2 (2021).

CAS

PubMed

PubMed Central

Article

Google Scholar

24.

Olah, M. et al. A transcriptomic atlas of aged human microglia. Nat. Commun. 9, 539 (2018).

PubMed

PubMed Central

Article

CAS

Google Scholar

25.

Hickman, S. E. et al. The microglial sensome revealed by direct RNA sequencing. Nat. Neurosci. 16, 1896–1905 (2013).

CAS

PubMed

PubMed Central

Article

Google Scholar

26.

Nam, K. N. et al. Effect of high fat diet on phenotype, brain transcriptome and lipidome in Alzheimer's model mice. Sci. Rep. 7, 4307 (2017).

PubMed

PubMed Central

Article

CAS

Google Scholar

27.

Oláh, J. et al. Interactions of pathological hallmark proteins: tubulin polymerization promoting protein/p25, beta-amyloid, and alpha-synuclein. J. Biol. Chem. 286, 34088–34100 (2011).

PubMed

PubMed Central

Article

CAS

Google Scholar


28.

Mazaheri, F. et al. Distinct roles for BAI1 and TIM-4 in the engulfment of dying neurons by microglia. Nat. Commun. 5, 4046 (2014).


CAS


PubMed


Article


Google Scholar


29.

Ciani, M., Benussi, L., Bonvicini, C. & Ghidoni, R. Genome wide association study and next generation sequencing: a glimmer of light toward new possible horizons in frontotemporal dementia research. Front. Neurosci. 13, 506 (2019).


PubMed


PubMed Central


Article


Google Scholar


30.

Li, Z. et al. The TMEM106B FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. Acta Neuropathol. 139, 45–61 (2020).

CAS

PubMed

Article

Google Scholar

31.

Prodan, C. I. et al. Coated-platelet levels and progression from mild cognitive impairment to Alzheimer disease. Neurology 76, 247–252 (2011).

CAS

PubMed

PubMed Central

Article

Google Scholar

32.

Greaves, C. V. & Rohrer, J. D. An update on genetic frontotemporal dementia. J. Neurol. 266, 2075–2086 (2019).

PubMed

PubMed Central

Article

Google Scholar

33.

Zhang, J. et al. Leukocyte immunoglobulin-like receptors in human diseases: an overview of their distribution, function, and potential application for immunotherapies. J. Leukoc. Biol. 102, 351–360 (2017).

CAS

PubMed

PubMed Central

Article

Google Scholar

34.

Cao, Q. et al. Inhibiting amyloid-β cytotoxicity through its interaction with the cell surface receptor LilrB2 by structure-based design. Nat. Chem. 10, 1213–1221 (2018).

CAS

PubMed

PubMed Central

Article

Google Scholar

35.

Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. Nat. Genet. 49, 325–331 (2017).

CAS

PubMed

Article

PubMed Central

Google Scholar

36.

Schwabe, T., Srinivasan, K. & Rhinn, H. Shifting paradigms: the central role of microglia in Alzheimer's disease. Neurobiol. Dis. 143, 104962 (2020).

CAS

PubMed

Article

PubMed Central

Google Scholar

37.

Yao, D. W., O'Connor, L. J., Price, A. L. & Gusev, A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat. Genet. 52, 626–633 (2020).

CAS

PubMed

PubMed Central

Article

Google Scholar

38.

Corces, M. R. et al. Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat. Genet. 52, 1158–1168 (2020).

CAS

PubMed

PubMed Central

Article

Google Scholar

39.

Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. Nat. Protoc. 9, 1192–1212 (2014).

PubMed

PubMed Central

Article

Google Scholar

40.

Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. Brief. Bioinform. 14, 144–161 (2013).

CAS

PubMed

Article

PubMed Central

Google Scholar

41.

Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D. & Lin, X. Sequence kernel association tests for the combined effect of rare and common variants. Am. J. Hum. Genet. 92, 841–853 (2013).

CAS

PubMed

PubMed Central

Article

Google Scholar


42.

Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826 (2017).

PubMed

PubMed Central

Article

CAS

Google Scholar


43.

Auton, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).

PubMed

Article

CAS

PubMed Central

Google Scholar

44.

Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 12, e1001779 (2015).

PubMed

PubMed Central

Article

Google Scholar

45.

Liu, P.-P., Xie, Y., Meng, X.-Y. & Kang, J.-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. Signal Transduct. Target. Ther. 4, 29 (2019).

PubMed

PubMed Central

Article

CAS

Google Scholar

46.

Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. Transl. Psychiatry 8, 99 (2018).

PubMed

PubMed Central

Article

Google Scholar

47.

Desikan, R. S. et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. PLoS Med. 14, e1002258 (2017).

PubMed

PubMed Central

Article

Google Scholar

48.

Jun, G. et al. A novel Alzheimer disease locus located near the gene encoding tau protein. Mol. Psychiatry 21, 108–117 (2016).

CAS

PubMed

Article

PubMed Central

Google Scholar

49.

de Rojas, I. et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. Nat. Commun. 12, 3417 (2021).

CAS

PubMed

PubMed Central

Article

Google Scholar

50.

Schwartzentruber, J. et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. Nat. Genet. 53, 392–402 (2021).

CAS

PubMed

PubMed Central

Article

Google Scholar

51.

Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 26, 2336–2337 (2010).

CAS

PubMed

PubMed Central

Article

Google Scholar

52.

Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–425 (2015).

CAS

PubMed

PubMed Central

Article

Google Scholar

53.

de Leeuw, C. A., Stringer, S., Dekkers, I. A., Heskes, T. & Posthuma, D. Conditional and interaction gene-set analysis reveals novel functional pathways for blood pressure. Nat. Commun. 9, 3768 (2018).

PubMed

PubMed Central

Article

CAS

Google Scholar

54.

Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. Nature 573, 61–68 (2019).

CAS

PubMed

PubMed Central

Article

Google Scholar

55.

Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017).

CAS

PubMed

PubMed Central

Article

Google Scholar

56.

Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature 555, 524–528 (2018).

CAS

PubMed

Article

Google Scholar

57.

Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. Proc. Natl Acad. Sci. USA 112, 7285–7290 (2015).

CAS

PubMed

PubMed Central

Article

Google Scholar

58.

Enge, M. et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell 171, 321–330.e14 (2017).

CAS

PubMed

PubMed Central

Article

Google Scholar

59.

Hochgerner, H. et al. STRT-seq-2i: dual-index 5' single cell and nucleus RNA-seq on an addressable microwell array. Sci. Rep. 7, 16327 (2017).

PubMed

PubMed Central

Article

CAS

Google Scholar

60.

Han, X. et al. Mapping the mouse cell atlas by Microwell-seq. Cell 172, 1091–1107.e17 (2018).

CAS

PubMed

PubMed Central

Article

Google Scholar

61.

Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017).

CAS

PubMed

PubMed Central

Article

Google Scholar

62.

Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaat8464 (2018).

CAS

PubMed

PubMed Central

Article

Google Scholar

63.

Alasoo, K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. 50, 424–431 (2018).

CAS

PubMed

PubMed Central

Article

Google Scholar

64.

Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 167, 1398–1414.e24 (2016).

CAS

PubMed

PubMed Central

Article

Google Scholar

65.

Jaffe, A. E. et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nat. Neurosci. 21, 1117–1125 (2018).

CAS

PubMed

PubMed Central

Article

Google Scholar

66.

Momozawa, Y. et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat. Commun. 9, 2427 (2018).

PubMed

PubMed Central

Article

CAS

Google Scholar

67.

Fairfax, B. P. et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat. Genet. 44, 502–510 (2012).

CAS

PubMed

PubMed Central

Article

Google Scholar

68.

Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343, 1246949 (2014).

PubMed

PubMed Central

Article

CAS

Google Scholar

69.

Gutierrez-Arcelus, M. et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. Elife 2, e00523 (2013).

PubMed

PubMed Central

Article

CAS

Google Scholar

70.

Kasela, S. et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. PLoS Genet. 13, e1006643 (2017).

PubMed

PubMed Central

Article

CAS

Google Scholar

71.

Lepik, K. et al. C-reactive protein upregulates the whole blood expression of CD59—an integrative analysis. PLoS Comput. Biol. 13, e1005766 (2017).

PubMed

PubMed Central

Article

CAS

Google Scholar

72.

Naranbhai, V. et al. Genomic modulators of gene expression in human neutrophils. Nat. Commun. 6, 7545 (2015).

PubMed

Article

Google Scholar

73.

Nédélec, Y. et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell 167, 657–669.e21 (2016).

PubMed

Article

CAS

Google Scholar

74.

Quach, H. et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. Cell 167, 643–656.e17 (2016).

CAS

PubMed

PubMed Central

Article

Google Scholar

75.

Schwartzentruber, J. et al. Molecular and functional variation in iPSC-derived sensory neurons. Nat. Genet. 50, 54–61 (2018).

CAS

PubMed

Article

Google Scholar

76.

Buil, A. et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. 47, 88–91 (2015).

CAS

PubMed

Article

Google Scholar

77.

Võsa, U. et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at bioRxiv https://doi.org/10.1101/447367 (2018).

78.

Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013).

CAS

PubMed

PubMed Central

Article

Google Scholar

79.

Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. 49, 139–145 (2017).

CAS

PubMed

Article

PubMed Central

Google Scholar

80.
Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nat. Neurosci. 20, 1418–1426 (2017).

CAS

PubMed

PubMed Central

Article

Google Scholar

81.
Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. 19, 1442–1453 (2016).

CAS

PubMed

PubMed Central

Article

Google Scholar

82.

Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. 17, 1418–1428 (2014).

CAS

PubMed

PubMed Central

Article

Google Scholar

83.

Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. PLoS Genet. 16, e1008720 (2020).

CAS

PubMed

PubMed Central

Article

Google Scholar

84.

Kerimov, N. et al. eQTL Catalogue: a compendium of uniformly processed human gene expression and splicing QTLs. Preprint at bioRxiv https://doi.org/10.1101/2020.01.29.924266 (2020).

85.

Young, A. M. H. et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. Nat. Genet. 53, 861–868 (2021).

CAS

PubMed

PubMed Central

Article

Google Scholar

86.

Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. B Stat. Methodol. 82, 1273–1300 (2020).

Article

Google Scholar


87.

Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. Am. J. Hum. Genet. 101, 539–551 (2017).


CAS


PubMed


PubMed Central


Article


Google Scholar


88.

R Core Team. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2017).


89.

Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).


CAS


PubMed


PubMed Central

Article

Google Scholar

90.

Lawrence, M. et al. Software for computing and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118 (2013).

CAS

PubMed

PubMed Central

Article

Google Scholar

91.

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

PubMed

PubMed Central

Article

CAS

Google Scholar


92.

O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745 (2016).


PubMed


Article


CAS


Google Scholar


93.

Hadley, W. ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, New York).


Download references