ARTICLES
https://doi.org/10.1038/s41588-021-00921-z

nature genetics

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

Douglas P. Wightman[1], Iris E. Jansen[1], Jeanne E. Savage[1], Alexey A. Shadrin[2,3], Shahram Bahrami[2,3,4], Dominic Holland[5], Arvid Rongve[6,7], Sigrid Børte[3,8,9], Bendik S. Winsvold[9,10,11], Ole Kristian Drange[12,13], Amy E. Martinsen[3,9,10], Anne Heidi Skogholt[9,14], Cristen Willer[15], Geir Bråthen[16,17,18], Ingunn Bosnes[12,19], Jonas Bille Nielsen[9,15,20], Lars G. Fritsche[21], Laurent F. Thomas[9,14], Linda M. Pedersen[10], Maiken E. Gabrielsen[9], Marianne Bakke Johnsen[3,8,9], Tore Wergeland Meisingset[16,17], Wei Zhou[22,23], Petroula Proitsi[24], Angela Hodges[24], Richard Dobson[25,26,27,28,29], Latha Velayudhan[24], Karl Heilbron[30], Adam Auton[30], 23andMe Research Team*, Julia M. Sealock[31,32], Lea K. Davis[31,32], Nancy L. Pedersen[33], Chandra A. Reynolds[34], Ida K. Karlsson[33,35], Sigurdur Magnusson[36], Hreinn Stefansson[36], Steinunn Thordardottir[37], Palmi V. Jonsson[37,38], Jon Snaedal[37], Anna Zettergren[39], Ingmar Skoog[39,40], Silke Kern[39,40], Margda Waern[39,41], Henrik Zetterberg[42,43,44,45], Kaj Blennow[44,45], Eystein Stordal[12,19], Kristian Hveem[9,46], John-Anker Zwart[3,9,10], Lavinia Athanasiu[2,4], Per Selnes[47], Ingvild Saltvedt[16,18], Sigrid B. Sando[16,17], Ingun Ulstein[48], Srdjan Djurovic[49,50], Tormod Fladby[3,47], Dag Aarsland[24,51], Geir Selbæk[3,48,52], Stephan Ripke[23,53,54], Kari Stefansson[36], Ole A. Andreassen[2,3,4,56] and Danielle Posthuma[1,55,56] ✉

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.

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

A full list of affiliations appears at the end of the paper.

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 (λ) 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).

Tissue type, cell type and gene-set enrichment. MAGMA tissue specificity analysis[15] identified spleen ($P_{\text{Bonferroni}} = 0.034$) 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 ($P_{\text{Bonferroni}} = 0.054$) 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

Table 1 | The 38 genomic risk loci identified from 90,338 (46,613 proxy) cases and 1,036,225 (318,246 proxy) controls

Genomic locus	Gene	Position (GRCh37)	Lead variant	A1	A1 frequency	*P* value	*N*
1	***AGRN***	**1:985,377**	**rs113020870**	**T**	**0.0041**	$\mathbf{3.83\times10^{-8}}$	**776,379**
2	*CR1*	1:207,750,568	rs679515	C	0.82	2.42×10^{-25}	762,176
3	*NCK2*	2:106,235,428	rs115186657	C	0.0035	1.33×10^{-8}	727,537
4	*BIN1*	2:127,891,427	rs4663105	C	0.41	3.92×10^{-58}	1,078,540
5	*INPPD5*	2:234,082,577	rs7597763	C	0.45	4.65×10^{-9}	819,541
6	*CLNK*	4:11,014,822	rs4504245	G	0.79	5.23×10^{-12}	1,080,458
7	***TNIP1***	**5:150,432,388**	**rs871269**	**T**	**0.32**	$\mathbf{1.37\times10^{-9}}$	**1,089,904**
8	***HAVCR2***	**5:156,526,331**	**rs6891966**	**G**	**0.77**	$\mathbf{7.91\times10^{-10}}$	**1,089,230**
9	*HLA-DRB1*	6:32,583,813	rs1846190	A	0.30	2.66×10^{-14}	754,040
10	*TREM2*	6:40,942,196	rs187370608	G	0.997	1.26×10^{-25}	791,668
11	*CD2AP*	6:47,552,180	rs9369716	T	0.27	1.70×10^{-17}	1,052,285
12	***TMEM106B***	**7:12,268,758**	**rs5011436**	**C**	**0.41**	$\mathbf{2.70\times10^{-9}}$	**1,123,678**
13	*ZCWPW1/NYAP1*	7:99,932,049	rs7384878	T	0.69	9.41×10^{-16}	1,084,138
14	*EPHA1-AS1*	7:143,104,331	rs3935067	G	0.62	4.69×10^{-11}	1,117,025
15	*CLU*	8:27,466,315	rs1532278	T	0.39	1.57×10^{-22}	1,126,563
16	*SHARPIN*	8:145,108,151	rs61732533	G	0.95	3.14×10^{-9}	1,122,653
17	*USP6NL/ECHDC3*	10:11,718,713	rs7912495	G	0.46	7.68×10^{-15}	1,120,367
18	*CCDC6*	10:61,738,152	rs7902657	T	0.54	3.68×10^{-8}	1,126,388
19	*MADD/SPI1*	11:47,380,340	rs3740688	T	0.54	8.78×10^{-9}	1,123,185
20	*MS4A4A*	11:60,021,948	rs1582763	G	0.62	3.40×10^{-33}	1,125,804
21	*PICALM*	11:85,800,279	rs561655	G	0.35	1.24×10^{-26}	1,126,563
22	*SORL1*	11:121,435,587	rs11218343	T	0.96	1.33×10^{-13}	1,125,100
23	*FERMT2*	14:53,298,853	rs7146179	G	0.89	6.99×10^{-11}	1,089,904
24	*RIN3*	14:92,938,855	rs12590654	G	0.67	6.63×10^{-17}	1,116,967
25	*ADAM10*	15:59,057,023	rs602602	T	0.70	6.22×10^{-15}	1,124,268
26	*APH1B*	15:63,569,902	rs117618017	T	0.13	7.00×10^{-12}	889,854
27	*SCIMP/RABEP1*	17:4,969,940	rs7209200	T	0.33	3.18×10^{-8}	1,125,637
28	***GRN***	**17:42,442,344**	**rs708382**	**T**	**0.61**	$\mathbf{1.98\times10^{-9}}$	**1,125,622**
29	*ABI3*	17:47,450,775	rs28394864	G	0.54	4.90×10^{-10}	1,084,218
30	*TSPOAP1-AS1*	17:56,409,089	rs2632516	G	0.54	7.46×10^{-10}	1,082,451
31	*ACE*	17:61,545,779	rs6504163	T	0.61	1.23×10^{-9}	1,083,145
32	*ABCA7*	19:1,050,874	rs12151021	G	0.68	2.81×10^{-15}	1,082,434
33	*APOE*	19:45,411,941	rs429358	T	0.84	$<1.0\times10^{-300}$	1,126,190
34	***NTN5***	**19:49,213,504**	**rs2452170**	**G**	**0.47**	$\mathbf{1.72\times10^{-8}}$	**1,088,626**
35	*CD33*	19:51,737,991	rs1354106	G	0.37	2.21×10^{-10}	716,038
36	***LILRB2***	**19:54,825,174**	**rs1761461**	**C**	**0.49**	$\mathbf{1.56\times10^{-9}}$	**1,116,336**
37	*CASS4*	20:54,995,699	rs6069737	T	0.083	6.73×10^{-16}	1,087,703
38	*APP*	21:27,520,931	rs2154482	T	0.44	7.66×10^{-10}	1,124,606

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 bold. The genes were assigned on the basis of colocalization results, fine-mapping results and previous literature. A1, tested allele; *N*, sample size.

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; $P=7.91\times10^{-10}$) 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

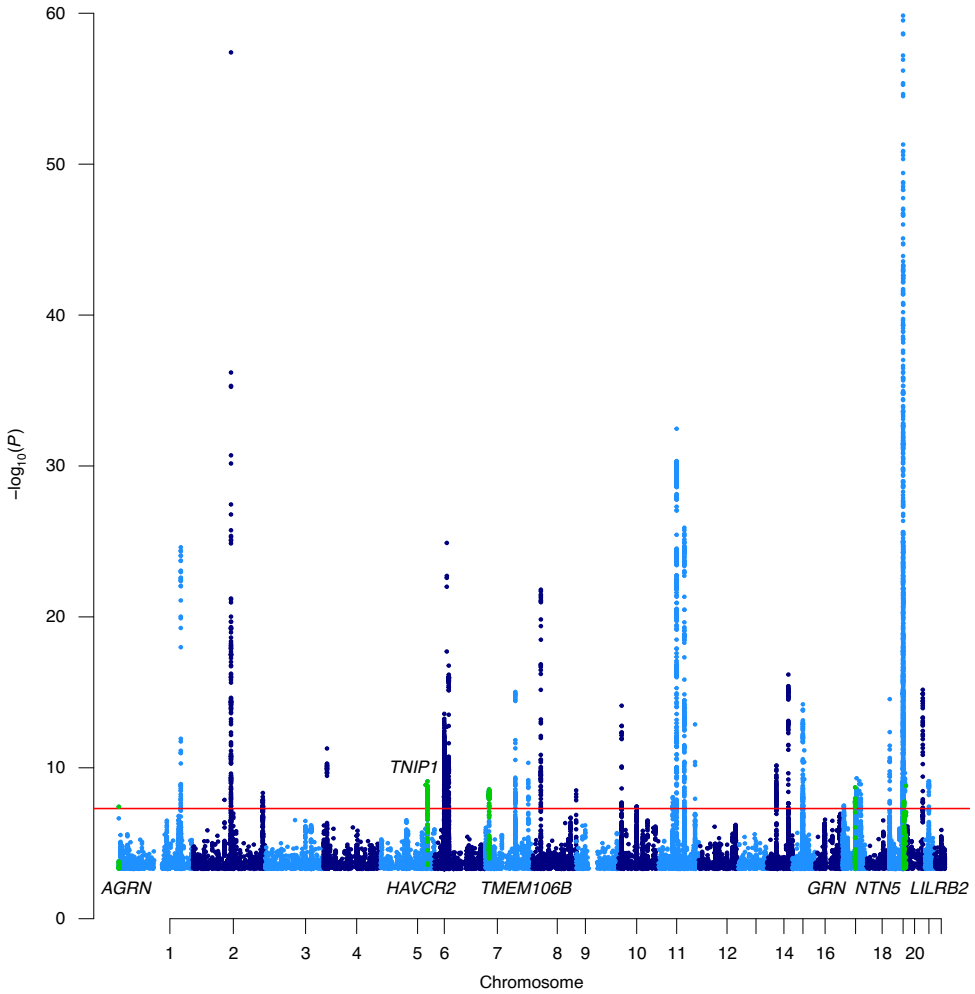

Fig. 1 | A Manhattan plot of the meta-analysis results highlighting 38 loci, including seven previously unidentified regions. Only variants with $P < 0.0005$ are displayed. The *APOE* region cannot be fully observed because the *y* axis is limited to the top variant in the second most significant locus, $-\log_{10}(1\times10^{-60})$, to display the less significant variants. The red line represents genome-wide significance (5×10^{-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.

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; $P=2.7\times10^{-9}$) is an intron variant in *TMEM106B* (Supplementary Fig. 7). A nearby exonic variant (rs3173615; $R^2=0.976$ in 1KG EUR; $P=6.61\times10^{-9}$) 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 degradation[30]. *TMEM106B* was also found to be significantly differentially expressed in bulk brain tissue of patients with LOAD compared to controls[23]. The lead variant in locus 28 (rs708382; $P=1.98\times10^{-9}$) 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 platelets[31]. 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 gene[32] 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, $P = 1.56 \times 10^{-9}$) 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 activation[33]. *LILRB5*, *LILRA5* and *LILRB2* were significantly differentially expressed in bulk brain tissue of patients with LOAD compared to controls[23]. 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 binding[33]. 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 approach[34]. 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 ourselves[8] and others[35]. 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 pathology[36]. 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 FTD[29]. 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 study[7] 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 LOAD[10]. 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 size[35] 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 study[38], 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.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00921-z.

Received: 12 April 2021; Accepted: 16 July 2021;
Published online: 7 September 2021

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).
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).
3. DeTure, M. A. & Dickson, D. W. The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* **14**, 32 (2019).
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).
5. Gatz, M. et al. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **63**, 168–174 (2006).
6. Zhang, Q. et al. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat. Commun.* **11**, 4799 (2020).
7. Holland, D. et al. The genetic architecture of human complex phenotypes is modulated by linkage disequilibrium and heterozygosity. *Genetics* **217**, iyaa046 (2021).
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).
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).
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).
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).
12. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
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).
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).
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).
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).

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).
18. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
19. Shamilov, R. & Aneskievich, B. J. TNIP1 in autoimmune diseases: regulation of Toll-like receptor signaling. *J. Immunol. Res.* **2018**, 3491269 (2018).
20. Cho, C. E. et al. A modular analysis of microglia gene expression, insights into the aged phenotype. *BMC Genomics* **20**, 164 (2019).
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).
22. Li, X. et al. Systematic analysis and biomarker study for Alzheimer's disease. *Sci. Rep.* **8**, 17394 (2018).
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).
24. Olah, M. et al. A transcriptomic atlas of aged human microglia. *Nat. Commun.* **9**, 539 (2018).
25. Hickman, S. E. et al. The microglial sensome revealed by direct RNA sequencing. *Nat. Neurosci.* **16**, 1896–1905 (2013).
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).
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).
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).
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).
30. Li, Z. et al. The *TMEM106B* FTLD-protective variant, rs1990621, is also associated with increased neuronal proportion. *Acta Neuropathol.* **139**, 45–61 (2020).
31. Prodan, C. I. et al. Coated-platelet levels and progression from mild cognitive impairment to Alzheimer disease. *Neurology* **76**, 247–252 (2011).
32. Greaves, C. V. & Rohrer, J. D. An update on genetic frontotemporal dementia. *J. Neurol.* **266**, 2075–2086 (2019).
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).
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).
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).
36. Schwabe, T., Srinivasan, K. & Rhinn, H. Shifting paradigms: the central role of microglia in Alzheimer's disease. *Neurobiol. Dis.* **143**, 104962 (2020).
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).
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).

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU Amsterdam, Amsterdam, the Netherlands. [2]NORMENT Centre, University of Oslo, Oslo, Norway. [3]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [4]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [5]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. [6]Department of Research and Innovation, Helse Fonna, Haugesund Hospital, Haugesund, Norway. [7]The University of Bergen, Institute of Clinical Medicine (K1), Bergen, Norway. [8]Research and Communication Unit for Musculoskeletal Health (FORMI), Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [9]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [10]Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway. [11]Department of Neurology, Oslo University Hospital, Oslo, Norway. [12]Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [13]Division of Mental Health Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [14]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway. [15]Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI, USA. [16]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway. [17]Department of Neurology and Clinical Neurophysiology, University Hospital of Trondheim, Trondheim, Norway. [18]Department of Geriatrics, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway. [19]Department of Psychiatry, Hospital Namsos, Nord-Trøndelag Health Trust, Namsos, Norway. [20]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [21]Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. [22]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [23]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [24]Institute of Psychiatry Psychology and Neurosciences, King's College London, London, UK. [25]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK. [26]NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, UK. [27]Health Data Research UK London, University College London, London, UK. [28]Institute of Health Informatics, University College London, London, UK. [29]NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, UK. [30]23andMe Inc., Sunnyvale, CA, USA. [31]Division of Genetic Medicine, Department of Medicine Vanderbilt University Medical Center Nashville, Nashville, TN, USA. [32]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. [33]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [34]Department of Psychology, University of California-Riverside, Riverside, CA, USA. [35]Institute of Gerontology and Aging Research Network – Jönköping (ARN-J), School of Health and Welfare, Jönköping University, Jönköping, Sweden. [36]deCODE Genetics/Amgen, Reykjavik, Iceland. [37]Department of Geriatric Medicine, Landspitali University Hospital, Reykjavik, Iceland. [38]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [39]Neuropsychiatric Epidemiology Unit, Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy, Centre for Ageing and Health (AGECAP) at the University of Gothenburg, Gothenburg, Sweden. [40]Region Västra Götaland, Sahlgrenska University Hospital, Psychiatry, Cognition and Old Age Psychiatry Clinic, Gothenburg, Sweden. [41]Region Västra Götaland, Sahlgrenska University Hospital, Psychosis Clinic, Gothenburg, Sweden. [42]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK. [43]UK Dementia Research Institute at UCL, London, UK. [44]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden. [45]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden. [46]HUNT Research Center, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [47]Department of Neurology, Akershus University Hospital, Lørenskog, Norway. [48]Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway. [49]Department of Medical Genetics, Oslo University Hospital, Oslo, Norway. [50]NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway. [51]Centre of Age-Related Medicine, Stavanger University Hospital, Stavanger, Norway. [52]Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway. [53]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [54]Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin, Berlin, Germany. [55]Department of Child and Adolescent Psychiatry and Pediatric Psychology, Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam University Medical Center, Amsterdam, the Netherlands. [56]These authors contributed equally: Ole A. Andreassen, Danielle Posthuma. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: d.posthuma@vu.nl

23andMe Research Team

Michelle Agee[30], Stella Aslibekyan[30], Elizabeth Babalola[30], Robert K. Bell[30], Jessica Bielenberg[30], Katarzyna Bryc[30], Emily Bullis[30], Briana Cameron[30], Daniella Coker[30], Gabriel Cuellar Partida[30], Devika Dhamija[30], Sayantan Das[30], Sarah L. Elson[30], Teresa Filshtein[30], Kipper Fletez-Brant[30], Pierre Fontanillas[30], Will Freyman[30], Pooja M. Gandhi[30], Barry Hicks[30], David A. Hinds[30], Karen E. Huber[30], Ethan M. Jewett[30], Yunxuan Jiang[30], Aaron Kleinman[30], Katelyn Kukar[30], Vanessa Lane[30], Keng-Han Lin[30], Maya Lowe[30], Marie K. Luff[30], Jey C. McCreight[30], Matthew H. McIntyre[30], Kimberly F. McManus[30], Steven J. Micheletti[30], Meghan E. Moreno[30], Joanna L. Mountain[30], Sahar V. Mozaffari[30], Priyanka Nandakumar[30], Elizabeth S. Noblin[30], Jared O'Connell[30], Aaron A. Petrakovitz[30], G. David Poznik[30], Morgan Schumacher[30], Anjali J. Shastri[30], Janie F. Shelton[30], Jingchunzi Shi[30], Suyash Shringarpure[30], Chao Tian[30], Vinh Tran[30], Joyce Y. Tung[30], Xin Wang[30], Wei Wang[30], Catherine H. Weldon[30] and Peter Wilton[30]

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 analysis[8] (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 EasyQC[39]. If necessary, the data were converted to build GRCh37 before QC using the UCSC LiftOver tool[40]. 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 MAF[41] was defined as $< \frac{1}{\sqrt{2 \times N}}$. 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 ($<5 \times 10^{-8}$); the genomic risk loci included all variants correlated ($R^2 > 0.6$) with the most significant variant. The correlation estimates were defined using 1KG European reference information[43]. The 1KG European reference panel was chosen over the UKB[44] 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 LOAD[8,10,45–50]. Regional plots were generated using LocusZoom[51] and 1KG reference information.

Heritability and genetic correlation. LDSC regression[9] (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. LDSC[12] 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 Hub[13] (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 1KG[43] LD information. MAGMA gene-set analysis was performed where variants map to 15,496 gene sets from the MSigDB v.7.0 database[52]. 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 analysis[53]. 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 ($P_{\text{Bonferroni}} < 0.05$).

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 analysis[16] 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_Hippocampus[55], DroNc_Mouse_Hippocampus[55], GSE104276_Human_Prefrontal_cortex_all_ages[56], GSE67835_Human_Cortex[57], GSE81547_Human_Pancreas[58], Linnarsson_GSE101601_Human_Temporal_cortex[59], MouseCellAtlas_all[60], PBMC_10x_68k[61] and PsychENCODE_Adult[62]. 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.1.3.6a[42] 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 macrophage[63], BLUEPRINT monocyte[64], BLUEPRINT neutrophil[64], BLUEPRINT T-cell[64], BrainSeq Brain[65], CEDAR B-cell[66], CEDAR monocyte, CEDAR neutrophil[66], CEDAR T-cell[66], Fairfax B-cell[67], Fairfax naive monocyte[68], GENCORD T-cell[69], Kasela CD4 T-cell[70], Kasela CD8 T-cell[70], Lepik Blood[71], Naranbhai neutrophil[72], Nedelec macrophage[73], Quach monocyte[74], Schwartzentruber sensory neuron[75], TwinsUK blood[76], PsychENCODE brain[62], eQTLGen blood cis and trans[77], BloodeQTL blood[78], BIOS Blood[79], xQTLServer blood[80], CommonMind Consortium brain[81], BRAINEAC brain[82], 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 package[18]. Default priors were used for prior probability of association with the GWAS data and eQTL data. The prior probability of colocalization was set as 1×10^{-6} as recommended[83]. 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 catalog[84]. 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 data[87]. 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 annotation[89] 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 Ranges[90]. 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 ANNOVAR[91] using ANNOVAR and FASTA sequences for all annotated transcripts in RefSeq Gene[92]. 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

39. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
40. Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
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).
42. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
43. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
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).
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).
46. Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 99 (2018).
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).
48. Jun, G. et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108–117 (2016).
49. de Rojas, I. et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat. Commun.* **12**, 3417 (2021).
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).
51. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
52. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
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).
54. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
55. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
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).
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).
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).
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).
60. Han, X. et al. Mapping the mouse cell atlas by Microwell-seq. *Cell* **172**, 1091–1107.e17 (2018).
61. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
62. Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
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).
64. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016).
65. Jaffe, A. E. et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018).
66. Momozawa, Y. et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
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).
68. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
69. Gutierrez-Arcelus, M. et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* **2**, e00523 (2013).
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).
71. Lepik, K. et al. C-reactive protein upregulates the whole blood expression of CD59—an integrative analysis. *PLoS Comput. Biol.* **13**, e1005766 (2017).
72. Naranbhai, V. et al. Genomic modulators of gene expression in human neutrophils. *Nat. Commun.* **6**, 7545 (2015).
73. Nédélec, Y. et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* **167**, 657–669.e21 (2016).
74. Quach, H. et al. Genetic adaptation and Neandertal admixture shaped the immune system of human populations. *Cell* **167**, 643–656.e17 (2016).
75. Schwartzentruber, J. et al. Molecular and functional variation in iPSC-derived sensory neurons. *Nat. Genet.* **50**, 54–61 (2018).
76. Buil, A. et al. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat. Genet.* **47**, 88–91 (2015).
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).
79. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
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).
81. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).
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).
83. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet.* **16**, e1008720 (2020).
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).
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).
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).
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).
90. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
91. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
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).
93. Hadley, W. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, New York).

Acknowledgements

We thank all the participants included in this study, including the participants from Finngen, GR@CE, IGAP, UKB, DemGene, TwinGene, STSA, the Gothenburg H70 Birth Cohort Studies and Clinical AD from Sweden, ANMerge, BioVU, 23andMe, HUNT, and deCODE. We thank the research participants from 23andMe who made this study possible. We thank the participants of the Norwegian Dementia Genetics Network (DemGene). This work was supported by BRAINSCAPES: A Roadmap from Neurogenetics to Neurobiology (grant no. 024.004.012), and a European Research Council advanced grant (grant no. ERC-2018-AdG GWAS2FUNC 834057 (to D.P.)). This work was supported by the Research Council of Norway (RCN: 248980, 248778, 223273), Norwegian Regional Health Authorities, Norwegian Health Association (22731, EU JPND: PMI-AD RCN 311993); and National Institutes of Health, National Institute on Aging R01 AG08724, R01 AG17561, R01 AG028555 and R01 AG060470. We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses. The investigators within IGAP contributed to the design and implementation of IGAP and/or provided data but did not participate in analysis or writing of this report. IGAP was made possible by the generous participation of the control subjects, the patients and their families. The iSelect chips were funded by the French National Foundation on Alzheimer's disease and related disorders. EADI was supported by the LABEX (laboratory of excellence program investment for the future) DISTALZ grant, Inserm, Institut Pasteur de Lille, Université de Lille 2 and the Lille University Hospital. GERAD/PERADES was supported by the Medical Research Council (grant no. 503480), Alzheimer's Research UK (grant no. 503176), the Wellcome Trust (grant no. 082604/2/07/Z) and German Federal Ministry of Education and Research (BMBF): Competence Network Dementia (CND) (grant nos. 01GI0102, 01GI0711, 01GI0420). CHARGE was partly supported by the NIH/NIA grant R01 AG033193 and the NIA AG081220 and AGES contract N01–AG–12100, the NHLBI grant R01 HL105756, the Icelandic Heart Association and the Erasmus Medical Center and Erasmus University. ADGC was supported by the NIH/NIA grants U01 AG032984, U24 AG021886, U01 AG016976 and the Alzheimer's Association grant ADGC–10–196728.

Author contributions

D.P. and O.A.A. conceived of the study. D.P.W. performed the meta-analysis and follow-up analyses. I.E.J., J.E.S., D.P. and O.A.A. supervised analyses. I.E.J. and J.E.S. generated the UKB data. S. Bahrami and A.A.S. helped with the study design. A.H.S., C.W., J.B.N., L.G.F., M.E.G., K.H., T.W.M. and M.B.J. contributed to the organization of the HUNT data. B.S.W., A.E.M., O.K.D., G.B., I.B., E.S., S. Børte, L.F.T., W.Z., J.-A.Z., S.B.S., G.S. and L.M.P. contributed to the methods and analysis of the HUNT data. D.A., E.S., O.A.A., A.R. and G.S. collected and analyzed the DemGene data. S.K., K.B., A.Z., I. Skoog, M.W. and H.Z. financed and collected the Gothenburg H70 Birth Cohort Studies and Clinical AD Sweden data, and A.Z. processed and coordinated the analysis. A.H. contributed to the IGAP and ANMerge data. P.P. provided ANMerge data. D.H. provided power estimates. R.D., L.V., the 23andMe Research Team, J.M.S., L.K.D., N.L.P., C.A.R., I.K.K., S.M., H.S., S.T., P.V.J., J.S., L.A., P.S., I. Saltvedt, I.U., S.D., T.F., S.R. and K.S. analyzed and provided data. D.P.W. wrote the first draft of the manuscript. All authors critically reviewed the paper

Competing interests

H.Z. has served at scientific advisory boards for Denali, Roche Diagnostics, Wave, Samumed, Siemens Healthineers, Pinteon Therapeutics and CogRx; has given lectures in symposia sponsored by Fujirebio, Alzecure and Biogen; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program (outside submitted work). K.B. has served as a consultant, at advisory boards, or at data monitoring committees for Abcam, Axon, Biogen, JOMDD/Shimadzu, Julius Clinical, Lilly, MagQu, Novartis, Roche Diagnostics and Siemens Healthineers and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. O.A.A. is a consultant to HealthLytix and received speaker's honorarium from Lundbeck and Sunovion. J.B.N. is employed by Regeneron Pharmaceuticals, Inc. T.W.M. has received speaker's honorarium from Roche. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00921-z.

Correspondence and requests for materials should be addressed to Danielle Posthuma.

Peer review information *Nature Genetics* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

nature research

Corresponding author(s): Danielle Posthuma

Last updated by author(s): Jun 15, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: UCSC Liftover : Open-source software to liftover data from different builds
EasyQC v9.1: Open-source R package used to QC summary statistics

Data analysis	The following standard software packages were used for the analyses described in the current manuscript: MAGMA v1.08: In-house developed software (de Leeuw et al., 2015) was used to conduct gene and gene-set analysis. FUMA v1.3.6a: In-house developed online platform for functional annotation of GWAS results (Watanabe et al., 2017). PLINK v1.9: Open-source software (Chang et al., 2015) used for performing genome-wide association analysis. mvGWAMA: In-house developed software for GWAS meta-analysis (Jansen et al., 2019). LDSC: Open-source software (Bulik-Sullivan et al. 2015) for heritability and genetic correlation estimates. Coloc v4.0: Open-source software (Giambartolomei et al. 2014) for colocalization analysis. SusieR v0.9.1: Open-source R package (Wang et al. 2020) for fine-mapping. LDSTORE v2.0: Open-source software (Benner et al. 2017) for creating LD matrices. ANNOVAR v2019Oct24: Open-source software (Wang et al. 2010) for variant annotation. ggplot2: Open-source R package for plotting data. GAMBA: In-house web tool (Wei et al. 2019) for identifying brain regional gene expression. R 4.0.1: Statistical software used in enrichment analyses LDhub v1.9.3: Online tool (Zheng et al. 2017) for genetic correlation analysis LocusZoom: Online tool (Pruim et al. 2010) for plotting GWAS data METAL v2011-03-25: Meta-analysis tool (Willer et al. 2010) for generating summary statistics for polygenic risk scoring PRSice v2.3.3: Open-source tool (Euesden et al. 2017) for polygenic risk scoring

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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/), UKBiobank (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 will be made available upon publication from https://ctg.cncr.nl/software/summary_statistics. Access to the full set including 23andMe results can be obtained after the 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 ID: EGAD00001005736). MSigDB gene-sets are available online (https://www.gsea-msigdb.org/gsea/msigdb/) and integrated in FUMA (https://fuma.ctglab.nl/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all samples the sample size consists of all individuals that remain after quality control of the data and exclusion of withdrawn subjects. Detailed information on the samples used, as well as the exclusion/inclusion criteria, are provided in the Supplementary Note.
Data exclusions	See Supplementary Note. All individuals of non-European ancestry, individuals who failed quality control, or individuals who asked to withdraw from the study were removed. Non-European ancestry individuals were excluded based on pre-established criteria (principal component analysis results).
Replication	All available datasets were included in the primary meta-analysis. We explicitly examined the concordance of effects within the novel loci and reported how our results replicated previous GWAS. The concordance of the datasets for the lead variant of all loci is available in Supplementary Table 4. The results of the functional follow-up was compared to the results of previous Alzheimer's disease GWAS.
Randomization	Covariates were used to control for differences in age, ancestry, and sex.
Blinding	The case control status of individuals was assigned through clinical diagnosis and self-assessment. The analysts were not blinded to the status of the individuals because QC procedures require knowing case status.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants were of European ancestry, with an overrepresentation of older individuals (>65) and females. Sex was included as a covariate and age was included as a covariate for most analyses. The median ages of individuals included in analyses which did not include age as covariate were >70. The population characteristics of each of the datasets is available in the Supplementary Note.
Recruitment	Participants were recruited through local hospitals, as well as biobanks, and commercial genotyping companies (23andMe). The selection of individuals through biobanks and commercial genotyping companies can bias which individuals are included. UK Biobank participants are healthier than the general population, but since the data used in this study referred to parental diagnoses, the impact of selection bias should be minor. The concordance of results of significant loci across datasets in the meta-analysis are available in the Supplementary Tables.
Ethics oversight	All participants provided informed consent. Ethical approval was provided by the IRB of each institution that provided data (Supplementary Note). The deCODE dataset was approved by the National Bioethics Committee and the Icelandic Data Protection Authority. The UKB received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382). For the STSA data, the study was approved by the Regional Ethics Board in Stockholm and the Institutional Review Board at the University of Southern California. For the TwinGene data, the study was approved by the Regional Ethics Board in Stockholm.

Note that full information on the approval of the study protocol must also be provided in the manuscript.