












Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes

Gloria Novikova ¹, Manav Kapoor^{1,2}, Julia TCW ^{1,2}, Edsel M. Abud^{3,4}, Anastasia G. Efthymiou¹, Steven X. Chen ^{5,6}, Haoxiang Cheng ², John F. Fullard^{2,7}, Jaroslav Bendl ^{2,7}, Yiyuan Liu¹, Panos Roussos ^{2,7}, Johan LM Björkegren ^{2,8}, Yunlong Liu^{5,6}, Wayne W. Poon ⁹, Ke Hao ², Edoardo Marcora ^{1,2,10} ✉ & Alison M. Goate ^{1,2,10} ✉

Genome-wide association studies (GWAS) have identified more than 40 loci associated with Alzheimer's disease (AD), but the causal variants, regulatory elements, genes and pathways remain largely unknown, impeding a mechanistic understanding of AD pathogenesis. Previously, we showed that AD risk alleles are enriched in myeloid-specific epigenomic annotations. Here, we show that they are specifically enriched in active enhancers of monocytes, macrophages and microglia. We integrated AD GWAS with myeloid epigenomic and transcriptomic datasets using analytical approaches to link myeloid enhancer activity to target gene expression regulation and AD risk modification. We identify AD risk enhancers and nominate candidate causal genes among their likely targets (including *AP4E1*, *AP4M1*, *APBB3*, *BIN1*, *MS4A4A*, *MS4A6A*, *PILRA*, *RABEP1*, *SPI1*, *TP53INP1*, and *ZYX*) in twenty loci. Fine-mapping of these enhancers nominates candidate functional variants that likely modify AD risk by regulating gene expression in myeloid cells. In the MS4A locus we identified a single candidate functional variant and validated it in human induced pluripotent stem cell (hiPSC)-derived microglia and brain. Taken together, this study integrates AD GWAS with multiple myeloid genomic datasets to investigate the mechanisms of AD risk alleles and nominates candidate functional variants, regulatory elements and genes that likely modulate disease susceptibility.

¹Ronald M. Loeb Center for Alzheimer's Disease, Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ²Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ³Department of Neurobiology & Behavior, University of California Irvine, Irvine, CA, USA. ⁴Sue and Bill Gross Stem Cell Research Center, University of California Irvine, Irvine, CA, USA. ⁵Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. ⁶Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA. ⁷Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. ⁸Integrated Cardio Metabolic Centre, Department of Medicine, Karolinska Institutet, Karolinska Universitetssjukhuset, Huddinge, Sweden. ⁹Institute for Memory Impairments and Neurological Disorders, University of California Irvine, Irvine, CA, USA. ¹⁰These authors contributed equally: Edoardo Marcora, Alison M. Goate. ✉email: edoardo.marcora@mssm.edu; alison.goate@mssm.edu

Alzheimer's disease (AD) is the most common type of dementia with a global burden of approximately 50 million people and no disease-modifying treatments available¹. Several lines of genetic evidence implicate myeloid cells in the etiology of AD². Whole-exome sequencing and microarray studies have identified rare coding variants associated with AD in genes (e.g., *TREM2*³, *SORL1*⁴, *AB13*⁵, *PLCG2*⁵ and *ABCA7*⁶) that play important roles in myeloid cells of the brain (microglia) and peripheral tissues (e.g., monocytes and macrophages) and have high relative expression levels in microglia compared to other brain cell types⁷. Genome-wide association studies (GWAS) have identified common non-coding variants associated with AD in more than 40 loci⁸, but the identification of the functional variants and causal genes underlying these statistical associations has been lacking. Earlier studies have focused on mapping candidate causal genes to AD risk loci using whole-blood and brain expression quantitative trait loci (eQTL) datasets^{9–11}. However, using tissue-level data poses obstacles to identifying myeloid-specific signals, because myeloid cells (microglia and monocytes) represent small fractions (~10%) of the total cell population in their respective tissues (brain and peripheral blood). More importantly, given the strong enrichment of AD risk alleles in myeloid-specific epigenomic annotations and expressed genes^{12,13}, it is imperative to investigate their impact on myeloid epigenomes and transcriptomes in the modulation of AD susceptibility.

Here, we show that AD risk alleles are specifically enriched in active enhancers of monocytes, macrophages and microglia and identify transcription factor binding motifs (TFBMs) over-represented within these regulatory elements. We further identify myeloid transcription factors (TFs) whose binding sites at active enhancers are likely burdened by AD risk variants. Given the selective enrichment of AD risk alleles in myeloid active enhancers, we sought to link the activity of myeloid enhancers that contain AD risk variants to target gene expression regulation and AD risk modification. To accomplish this we use two complementary approaches. First, we map myeloid active enhancers that contain AD risk alleles (AD risk enhancers) to their target genes by integrating chromatin interactions (promoter-capture Hi-C) and eQTL datasets from monocytes and macrophages. This approach allows us to nominate candidate causal genes in eleven genome-wide significant and five suggestive AD risk loci, including *TP53INP1*, *APBB3*, *RABEP1*, and *SPPL2A*. In our second approach, we use Summary data-based Mendelian Randomization (SMR)¹⁴ to investigate the causal relationship between chromatin activity, target gene expression, and AD risk modification. This approach allows us to identify specific active chromatin regions that likely modify AD risk by regulating the expression of one or more of their target genes in 12 loci. Importantly, the target genes of the myeloid active enhancers identified by these two analytical approaches are highly consistent and implicate the endolysosomal system of myeloid cells in the etiology of AD. We further fine-map AD risk enhancers to identify candidate functional variants that likely affect TF binding and regulate gene expression in seven loci, and validate one of these variants in the *MS4A* locus in human induced pluripotent stem cell (hiPSC)-derived microglia and brain.

Results

AD risk alleles are specifically enriched in active enhancers of monocytes, macrophages, and microglia. Our earlier analyses showed a significant enrichment of AD risk alleles in several myeloid-specific epigenomic annotations, but not in brain or other tissues/cell types (with the exception of liver and B-lymphoid cells)¹². To further dissect this enrichment, we used ChIP-Seq profiles of histone modifications that define the chromatin signatures of regulatory elements (H3K27ac for active enhancers and promoters,

H3K4me1 for enhancers, and H3K4me2 for enhancers and promoters and H3K4me3 for promoters) from monocytes, macrophages, and microglia to annotate the genome with myeloid active enhancers (AE), active promoters (AP), primed enhancers (PE) and primed promoters (PP) (see Methods)¹⁵. We identified 37246, 48242, and 34014 active enhancers, 7871, 13979, and 8284 active promoters, 11534, 34623, and 52360 primed enhancers and 3107, 4028, and 3112 primed promoters in monocytes, macrophages, and microglia, respectively. To identify which of these myeloid regulatory elements are enriched for AD risk alleles, we performed stratified LD score regression (LDSC)¹⁶ of AD single nucleotide polymorphism (SNP) heritability partitioned by the aforementioned epigenomic annotations using the International Genomics of Alzheimer's Project (IGAP) AD GWAS dataset¹⁷. This analysis revealed selective enrichment of AD risk alleles in active enhancers of monocytes, macrophages, and microglia (Fig. 1a). In contrast, schizophrenia SNP heritability (using the Psychiatric Genomics Consortium SCZ GWAS dataset as control¹⁸) was not enriched in any of these myeloid regulatory elements (Fig. 1a).

To identify TFs that likely regulate the activity of myeloid enhancers, we performed de novo motif analysis¹⁹ in open chromatin regions (identified by ATAC-Seq) that overlap with active enhancers in all three cell types (Supplementary Data 1). The binding motif for PU.1 (a transcription factor critical for myeloid and B-lymphoid cell development and function and an AD risk gene (*SPI1*)¹²) was the best match for the most highly overrepresented sequence motif in active enhancers across all three cell types, followed by AP-1, C/EBP, CTCF, and RUNX binding motifs. The binding motif for MEF2 family TFs (which includes *MEF2C* in another AD risk locus¹⁷) was the best match for highly overrepresented sequence motif in active enhancers of human microglia, consistent with findings in mouse microglia²⁰. To test whether the binding sites of TFs that likely regulate active myeloid enhancers are enriched for AD risk variants, we stratified ATAC-Seq regions in all three cell types by the presence of the binding motifs of the TFs that were found to be overrepresented in active myeloid enhancers and expressed in monocytes, macrophages and microglia (TPM \geq 1), and applied LDSC to quantify the enrichment of AD SNP heritability partitioned by these subsets of ATAC-Seq regions (Fig. 1b). ATAC-Seq regions overlapping with active enhancers that were positive for the PU.1 binding motif in all three cell types were enriched for AD risk alleles. MAF binding motif-positive ATAC-Seq regions were enriched for AD risk alleles in macrophage and microglial active enhancers. SMAD, USF, and SP binding motif-positive ATAC-Seq regions were enriched for AD risk alleles only in microglial active enhancers. Interestingly, a study comparing two mouse strains reported that genetic variants in *Mafb*, *Smad3*, and *Usf1* binding sites affected PU.1 binding specifically in microglia, suggesting that these TFs could be binding partners of PU.1 in microglia²¹. These results show that AD risk alleles are specifically enriched in active enhancers of monocytes, macrophages, and microglia, and nominate shared and cell-type specific TFs that likely regulate the activity of these regulatory elements. Additionally, these results implicate TFs whose binding to myeloid active enhancers is likely to be affected by AD risk alleles. These results support our hypothesis that TF binding sites might be altered by AD risk variants to affect myeloid enhancer activity and gene expression, which in turn modulate disease susceptibility by altering the biology of myeloid cells.

Integration of AD GWAS signals with myeloid epigenomic annotations, chromatin interactions (promoter-capture Hi-C), and eQTL datasets identifies candidate causal genes in sixteen AD risk loci. Promoter-enhancer interactions constitute one of

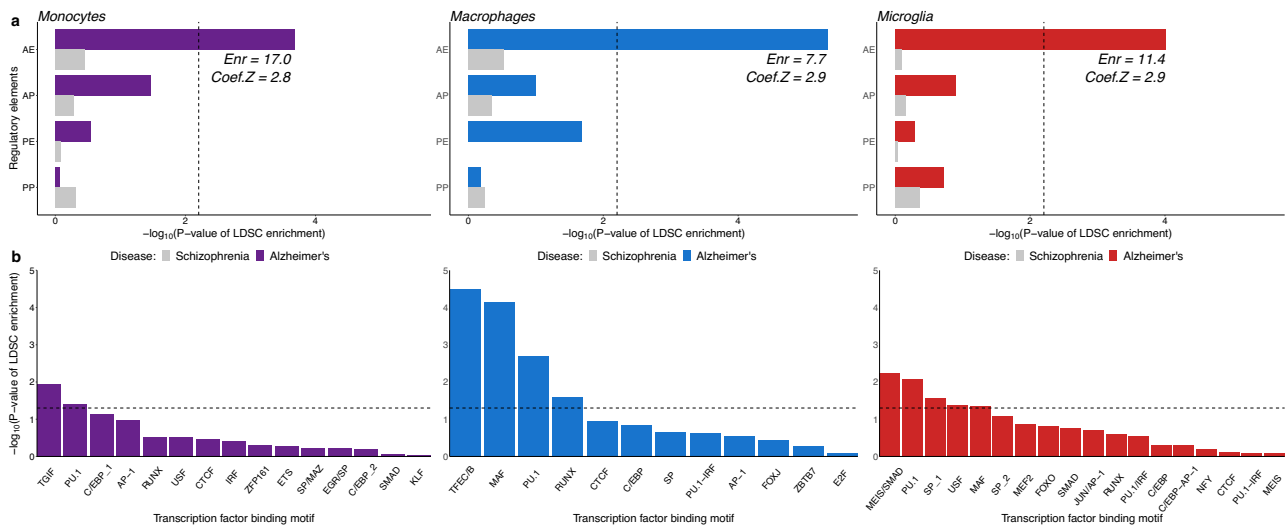


Fig. 1 AD risk alleles are specifically enriched in myeloid active enhancers and in putative transcription factor binding sites located in these enhancers.

a -Log₁₀ of enrichment *P*-values obtained from stratified LD Score Regression (LDSC) analysis of AD SNP heritability partitioned by active enhancer (AE), active promoter (AP), primed enhancer (PE) and primed promoter (PP) annotations in monocytes, macrophages, and microglia. Enr = Enrichment of AD SNP heritability partitioned by active enhancer annotations. Dashed line indicates the Bonferroni-corrected significance threshold. The enrichment standard errors for active enhancers are 3.8, 1.3, and 2.6 for monocytes, macrophages, and microglia, respectively. **b** -Log₁₀ of enrichment *P*-values obtained from stratified LD Score Regression (LDSC) analysis of AD SNP heritability partitioned by ATAC-Seq subsets. The subsets were obtained by stratifying ATAC-Seq regions in monocytes, macrophages, and microglia by the presence of the binding motif of TFs (listed on the x-axis) that were found to be overrepresented in active myeloid enhancers and expressed in monocytes, macrophages, and microglia, respectively (TPM ≥ 1)¹⁵. Dashed line indicates the nominal significance threshold.

the most fundamental mechanisms of gene expression regulation, where enhancer elements are brought into close proximity to cognate promoters to stimulate transcription of their target genes¹⁹. Given the observed enrichment of AD risk alleles in myeloid active enhancers, we reasoned that harnessing information about the spatial organization of chromatin and integrating it with epigenomic annotations and eQTLs in myeloid cells would facilitate the identification of candidate causal genes regulated by these elements in AD risk loci.

Chromatin interactions and eQTL datasets are currently not available for human microglia. However, our partitioned AD SNP heritability estimates suggest that active enhancers are enriched in monocytes, macrophages, and microglia to a similar extent, hence we used datasets from human peripheral blood monocytes and monocyte-derived macrophages as we did previously¹². We first identified active enhancers in monocytes and macrophages that contain AD risk alleles ($P \leq 1 \times 10^{-6}$, hereafter referred to as AD risk enhancers). Among these we then selected those that interact with at least one gene promoter and contain AD risk variants that are eQTLs for the same gene in monocytes and macrophages (FDR ≤ 5%) using the Javierre et al.¹⁹ promoter-capture Hi-C dataset and the Cardiogenics²², Fairfax et al. 2014²³ and STARNET²⁴ eQTL datasets. These analyses were performed within a single cell type: monocyte epigenomic marks were integrated with monocyte promoter capture Hi-C and 2 independent monocyte eQTL datasets (Cardiogenics and Fairfax). Similarly, macrophage epigenomic annotations were integrated with macrophage promoter capture Hi-C and 2 independent macrophage eQTL datasets (Cardiogenics and STARNET). Using this approach we nominate candidate causal genes in sixteen genome-wide significant and suggestive AD risk loci (Table 1). In some loci, this analysis identified genes that have known AD-associated coding variants (ABCA7²⁵) and genes that have been identified as most likely causal in previous studies (BIN1²⁶ and PTK2B²⁷). In other loci, we uncovered co-regulation of the expression of multiple target

genes by shared AD risk enhancers. For example, in the *SPI1* locus, we identified AD risk enhancers shared by *ACP2*, *MADD*, *MYBPC3*, *NR1H3*, *NUP160*, *PSMC3*, and *SPI1* in monocytes, and by *NUP160*, *MYBPC3*, and *SPI1* in macrophages. Similarly, in the *PILRA* locus (previously *ZCWPW1*), we identified AD risk enhancers shared by *AP4M1*, *PILRA*, *PILRB*, and *ZCWPW1* in monocytes, and by *AP4M1*, *MCM7*, *PILRA*, *PILRB*, *PVRIG*, and *STAG3* in macrophages. This could reflect the presence of either multiple causal genes at these loci or a single causal gene and several risk-neutral genes that show association by virtue of expression co-regulation. Additional evidence is necessary to distinguish between these two possibilities and prioritize one or more genes in the locus as we have shown for *SPI1* at the respective (previously *CELF1*) locus¹².

Additionally, these analyses revealed regulatory landscapes that are shared across myeloid cell types or are cell type-specific. In the *BINI* locus, we observed conserved AD risk enhancer-promoter chromatin interactions and similar eQTL signal profiles in monocytes and macrophages, suggesting that the AD risk regulome is similar in these two cell types and points to *BINI* as the strongest candidate causal gene at this locus (Fig. 2a). Conversely, in the *ZYX* (previously *EPHA1*) locus, we observed stronger chromatin interactions with a *ZYX* promoter in macrophages (mean interaction score 3.3 and 7.0 in monocytes and macrophages, respectively) and different eQTL signal profiles between monocytes and macrophages, suggesting that the AD risk regulome is different in these two cell types albeit pointing to the same candidate causal gene (Supplementary Figure 1). Finally, we identified candidate causal genes, such as *RABEP1* (Fig. 2b), *TP53INP1* (Supplementary Figure 2) and *APBB3* in suggestive loci. We also found that many of the genes prioritized through Hi-C in monocytes and macrophages are also associated with disease risk (SMR), including *ZYX*, *PILRA*, *AP4M1*, *RABEP1*, *APBB3* and *TP53INP1* (Table 1). In summary, this analytical approach allowed us to nominate candidate causal genes in sixteen AD risk loci.

Table 1 Candidate causal genes identified through integration of AD GWAS signals with myeloid active enhancer annotations, promoter-capture Hi-C, and eQTLs datasets.

Locus	Monocytes	Monocyte-derived macrophages
<i>BIN1</i> <i>SPI1</i> (previously <i>CELF1</i>)	<i>BIN1</i> <i>ACP2, FNBP4, MADD, MYBPC3, MTCH2, NR1H3, NUP160, PSMC3, SPI1</i>	<i>BIN1</i> <i>CELF1, MTCH2, MYBPC3, NUP160, PSMC3, SPI1</i>
<i>ZYX</i> (previously <i>EPHA1</i>) <i>MS4A</i> <i>PILRA</i> (previously <i>ZCWPW1</i>)	<i>ZYX</i> <i>MS4A6A</i> , <i>MS4A6E*</i> <i>AP4M1</i> , <i>GATS, PILRA, PILRB, ZKSCAN1, ZCWPW1</i>	<i>CLCN1*</i> , <i>EPHA1*</i> , <i>FAM131B, TAS2R41*</i> , <i>TAS2R60*</i> , <i>ZYX MS4A6A</i> <i>AP4M1</i> , <i>GATS, LAMTOR4, MCM7, PILRA, PILRB, PVRI, STAG3, TRIM4, ZKSCAN1</i> <i>INTS8, NDUFAF6, TP53INP1</i>
<i>TP53INP1</i> (previously <i>NDUFAF6</i>) <i>SPPL2A</i> <i>RIN3</i> (previously <i>SLC24A4</i>) <i>ABCA7</i> <i>APBB3</i> (previously <i>HBEGF</i>) <i>RABEP1</i> (previously <i>SCIMP</i>) <i>PTK2B</i> <i>CD55</i> (previously <i>CR1</i>) <i>CASS4</i> <i>TREM2</i> <i>PICALM</i>	<i>CCNE2, INTS8, TP53INP1</i> — <i>RIN3</i> <i>ABCA7, CNN2, CIRBP</i> — <i>NUP88, RABEP1</i> <i>PTK2B</i> — <i>AURKA</i> <i>NFYA</i> —	<i>AP4E1, TRPM7</i> — <i>ABCA7, CNN2, GPX4</i> <i>APBB3, PFDN1</i> <i>CHRNE, NUP88, RABEP1</i> <i>PTK2B</i> <i>CD55</i> — <i>TREM2*</i> <i>CCDC83*</i>

Genes highlighted in bold show a significant association between gene expression and disease risk.
*Not expressed in microglia.

Integration of AD GWAS signals with myeloid epigenomic annotations, chromatin activity (hQTL) and eQTL datasets identifies candidate causal genes in twelve AD risk loci.

Although chromatin interactions between active enhancers and gene promoters may suggest target gene expression regulation, inferring causal relationships between chromatin activity at enhancer elements and target gene expression can provide additional evidence for such regulation and help identify genetic variants that mediate these relationships to modulate disease susceptibility. We used SMR to explore the causal path that links chromatin activity to target gene expression and AD risk modification. To accomplish this, we used datasets from monocytes²⁸, since chromatin activity QTLs (hQTLs) are currently not available for human microglia or other macrophages. We first identified chromatin regions that contain AD risk alleles and overlap an active enhancer and used coloc²⁹ to select those with evidence of independent or colocalized AD GWAS and hQTL signals (PP.H3.abf + PP.H4.abf \geq 0.8) (Supplementary Data 2). To investigate the link between chromatin activity and target gene expression regulation, we used SMR to test for causal association between hQTL and eQTL effects in monocytes at the 26 regions selected using coloc as described above. We identified multiple genes that are likely regulated by the active enhancers in these regions (Fig. 3a, Table 2, Supplementary Data 3), including *BIN1*, *CD2AP*, *GPR141*, *MS4A4A*, *MS4A6A*, *RABEP1*, *SPI1*, *TP53INP1* and *ZYX*. We then used SMR to test for causal association between the expression of these genes and disease susceptibility. These analyses revealed specific active chromatin regions in monocytes, whose activity is causally associated with expression of their target genes, which in turn is causally associated with AD risk, including *BIN1*, *GPR141*, *MS4A4A*, *MS4A6A*, *RABEP1*, *SPI1*, *TP53INP1*, and *ZYX* (Fig. 3b, Supplementary Data 4). Seventeen of twenty-six genes nominated through causal associations between chromatin activity and gene expression and eight of fourteen genes nominated through causal associations between gene expression and disease susceptibility identified using the Cardiogenics monocyte eQTL dataset were replicated using the Fairfax monocyte eQTL dataset (Supplementary Data 5–6). Since the replication cohort is smaller, we expect that a larger number

of associations would replicate in a larger cohort, given the fact that almost all genes found through associations using the Fairfax dataset were significant in the main analysis using the Cardiogenics dataset. Additionally, in *MS4A*, *SPI1*, *TP53INP1* and *ZYX* loci both computational approaches pointed to the same candidate causal genes (albeit nominating different enhancers), while in *BIN1* and *RABEP1* loci both approaches pointed to the same AD risk enhancers and target gene (Fig. 2a). Hence, these results provide converging evidence for target gene expression regulation by active enhancers in these regions.

Although we observed a global enrichment of AD risk alleles in myeloid active enhancers across the human genome (Fig. 1a), we discovered a small subset of loci where the regulatory elements associated with causal gene expression regulation are either not active enhancers and/or do not themselves contain AD risk alleles. For example, we identified multiple primed enhancers in monocytes that do not contain AD risk alleles but whose hQTLs are causally associated with expression of *PILRA*, *AP4M1* and *ZKSCAN1*, which is in turn causally associated with AD risk (Fig. 3c). Moreover, we identified an active enhancer element whose activity is regulated by AD risk alleles located at a distance from it and which is strongly associated with expression of *AP4E1* and *SPPL2A* in monocytes (Fig. 3c). In turn, expression of *SPPL2A* is causally associated with AD risk. Furthermore, this chromatin region interacts with the promoter of *SPPL2A*, providing converging evidence for regulation of *SPPL2A* expression by this regulatory element. Therefore, it is possible that AD risk alleles indirectly affect the activity of this regulatory element by functional coupling through chromatin looping or another mechanism. In summary, this analytical approach allowed us to nominate candidate causal genes in twelve AD risk loci.

Fine-mapping using myeloid epigenomic annotations identifies candidate causal variants in seven AD risk loci. To prioritize candidate causal variants in myeloid enhancers we selected loci where we discovered significant associations between chromatin activity, gene expression and AD risk (i.e. *BIN1*, *GPR141*, *MS4A*, *PILRA*, *RABEP1*, *SPI1*, *SPPL2A*, *TP53INP1*, and *ZYX*). We

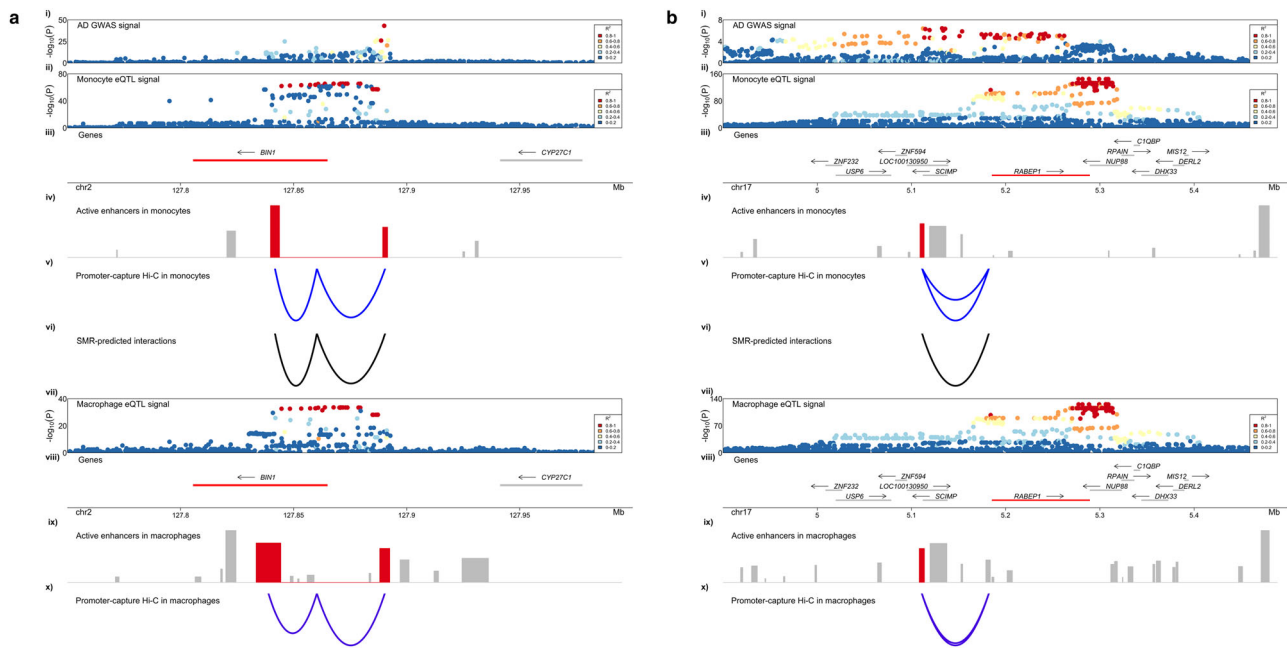


Fig. 2 AD risk enhancers spatially interact with the promoters of *BIN1* and *RABEP1* and regulate their expression in myeloid cells. **a.** (i) AD GWAS association signal in the *BIN1* locus. (ii) eQTL signal for *BIN1* in monocytes obtained from the Cardiogenics study. (iii) Genes that reside in the locus are plotted. Likely target genes of the highlighted AD risk enhancers are shown in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. (iv) Active enhancers in monocytes are plotted. The height of the bar is proportional to the strength of the epigenomic signal. AD risk enhancers that are prioritized through both Hi-C and SMR approaches are highlighted in red. (v) Promoter-capture Hi-C interactions between the *BIN1* promoter and the highlighted AD risk enhancers in monocytes. The depth of the arc is proportional to the strength of the interaction. (vi) AD risk enhancer-target gene interactions predicted by SMR analysis of causal associations between chromatin activity and *BIN1* expression in monocytes. The depth of the arc is proportional to the strength of the association. (vii) eQTL signal for *BIN1* in macrophages obtained from the Cardiogenics study. (viii) Genes that reside in the locus are plotted. Likely target genes of the highlighted AD risk enhancers are shown in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. (ix) Active enhancer elements in macrophages are plotted. AD risk enhancers that interact with the gene promoter are highlighted in red. (x) Promoter-capture Hi-C interactions between the *BIN1* promoter and the highlighted AD risk enhancers in macrophages. The depth of the arc is proportional to the strength of the interaction. Both Hi-C and SMR-predicted interactions are anchored at the AD risk enhancer highlighted. **b.** (i) AD GWAS association signal in the *RABEP1* locus. (ii) eQTL signal for *RABEP1* in monocytes obtained from the Cardiogenics study. (iii) Genes that reside in the locus are plotted. Likely target genes of the highlighted AD risk enhancers are shown in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. (iv) Active enhancers in monocytes are plotted. The height of the bar is proportional to the strength of the epigenomic signal. AD risk enhancers that are prioritized through both Hi-C and SMR approaches are highlighted in red. (v) Promoter-capture Hi-C interactions between the *RABEP1* promoter and the highlighted AD risk enhancers in monocytes. The depth of the arc is proportional to the strength of the interaction. (vi) AD risk enhancer-target gene interactions predicted by SMR analysis of causal associations between chromatin activity and *RABEP1* expression in monocytes. The depth of the arc is proportional to the strength of the association. (vii) eQTL signal for *RABEP1* in macrophages obtained from the Cardiogenics study. (viii) Genes that reside in the locus are plotted. Target genes of the highlighted AD risk enhancers are shown in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. (ix) Active enhancer elements in macrophages are plotted. AD risk enhancers that interact with the gene promoter are highlighted in red. (x) Promoter-capture Hi-C interactions between the *RABEP1* promoter and the highlighted AD risk enhancers in macrophages. The depth of the arc is proportional to the strength of the interaction. Hi-C and SMR-predicted interactions are anchored at the AD risk enhancers highlighted.

first selected variants in high to moderate LD ($R^2 \geq 0.8$) with the tagging variant in each locus and queried them in Haploreg³⁰ to identify coding variants. We identified a missense variant (rs1859788-G) in *PILRA* that is in high LD with the tagging variant ($R^2 = 0.86$, Alzheimer's Disease Genetics Consortium case-control cohort (ADGC) reference panel was used to compute LD as described previously¹² and was previously shown to alter the ligand binding affinity of *PILRA*³¹. Conditioning on this variant eliminates the AD GWAS signal at this locus (Supplementary Figure 3). The other eight AD risk loci did not contain coding variants in high LD with the tagging variant, prompting us to proceed with fine-mapping to prioritize candidate non-coding functional variants. To accomplish this we used PAINTOR, a Bayesian fine-mapping method that allows for integration of epigenomic annotations³². Due to the inflation of posterior probabilities when individuals in GWAS and LD reference panel are not well matched³³, we used GWAS and LD statistics

calculated using the ADGC cohort¹². Although this approach reduces the number of loci that can be statistically fine-mapped due to the smaller sample size in ADGC, the results are more stable. We obtained and reprocessed 38 myeloid epigenomic annotations^{15,34–38}, selected the ones that overlapped with active enhancers in myeloid cells and quantified their enrichment in each locus (Supplementary Figure 4). We then used PAINTOR with significantly enriched annotations (see Methods) to prioritize candidate causal variants and selected those with posterior probabilities of at least 0.1. To probe the likely effects of these variants on transcription factor binding, we screened for disruption or creation of binding motifs for TFs expressed in monocytes, macrophage and/or human microglia (TPM ≥ 1)¹⁵ using motifbreakR³⁹.

We identified candidate non-coding functional variants in the *BIN1*, *MS4A* and *ZYX* loci and proposed their likely mechanism of action (Supplementary Data 7). Additionally, we employed an

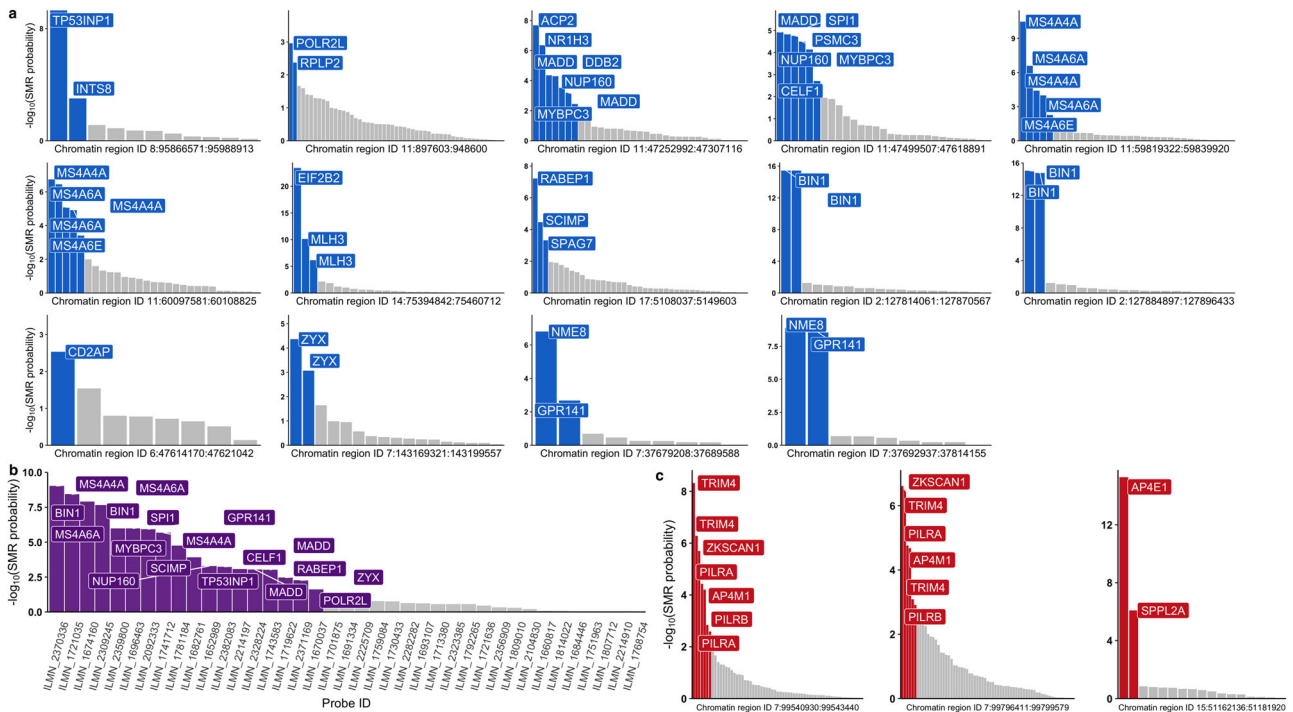


Fig. 3 Putative causal associations between chromatin activity, target gene expression regulation and AD risk modification point to candidate causal genes in myeloid cells. **a** $-\log_{10}$ of causal association probabilities between chromatin activity and gene expression in monocytes obtained through SMR analysis for each probe are plotted for each chromatin region. Probes (labeled by the respective gene) in blue indicate significant associations, while grey bars indicate non-significant associations based on a 5% FDR threshold. **b** $-\log_{10}$ of causal association probabilities between gene expression and AD risk. Probes (labeled by their respective gene) in purple indicate significant associations, while grey bars indicate non-significant associations based on a 5% FDR threshold. **c** $-\log_{10}$ of causal association probabilities between activity of two active chromatin regions in the PILRA locus and one active chromatin region in the SPPL2A locus and gene expression in monocytes obtained through SMR analysis for each probe are plotted. Probes (labeled by the respective gene) in red indicate significant associations, while grey bars indicate non-significant associations based on a 5% FDR threshold.

Table 2 Candidate causal genes identified through integration of AD GWAS signals with myeloid active enhancer annotations, hQTL, and eQTL datasets.

Locus	Genes implicated through enhancer activity to gene expression associations	Genes implicated through enhancer activity to gene expression to disease risk associations
<i>BIN1</i>	<i>BIN1</i>	<i>BIN1</i>
<i>SPI1</i>	<i>ACP2, CELF1, DDB2, MADD, MYBPC3, NRIH3, NUP160, PSMC3</i>	<i>CELF1, MADD, MYBPC3, NUP160, SPI1</i>
<i>CD2AP</i>	<i>CD2AP</i>	—
<i>ZYX</i> (previously <i>EPHA1</i>)	<i>ZYX</i>	<i>ZYX</i>
<i>GPR141</i> (previously <i>NME8</i>)	<i>NME8, GPR141</i>	<i>GPR141</i>
<i>TP53INP1</i>	<i>INTS8, TP53INP1</i>	<i>INTS8, TP53INP1</i>
<i>MS4A</i>	<i>MS4A4A, MS4A6A, MS4A6E*</i>	<i>MS4A4A, MS4A6A</i>
<i>RABEP1</i> (previously <i>SCIMP</i>)	<i>NUP88, RABEP1, SCIMP, SPAG7</i>	<i>NUP88, RABEP1, SCIMP</i>
<i>EIF2B2</i>	<i>ACY1, EIF2B2, MLH3</i>	—
<i>POLR2L</i>	<i>POLR2L, RPLP2</i>	<i>POLR2L</i>
<i>PILRA</i> ^a (previously <i>ZCWPWT</i>)	<i>AP4M1, TRIM4, PILRA, PILRB, ZKSCAN1</i>	<i>AP4M1, PILRA, ZKSCAN1</i>
<i>SPPL2A</i> ^a	<i>AP4E1, SPPL2A</i>	<i>SPPL2A</i>

^aAssociation reported with a distal enhancer that does not contain AD risk variants, but whose hQTLs colocalize with AD risk alleles.
*Not expressed in microglia.

alternative strategy for fine-mapping for the aforementioned loci and the loci that were not significant in the ADGC GWAS (but were significant in the IGAP GWAS). Briefly, using a block partitioning algorithm⁴⁰, conditional analyses⁴¹ and motif disruption/creation analyses³⁹ as well as integration of active enhancer annotations and eQTL datasets in monocytes and macrophages we were able to prioritize variants with regulatory

potential in seven AD risk loci (Supplementary Data 7, see Methods). As an example, in the *BIN1* locus we identified two independent AD GWAS signals. One of these signals is associated with rs6733839-T that is an eQTL for *BIN1* in human microglia⁴², resides in a PU.1 binding site in microglia and creates a binding motif for the MEF2 transcription factor, likely acting as a binding partner for PU.1 at that site. The other variant

(rs13025717-T) also resides in a PU.1 binding site, is an eQTL for *BIN1* in monocytes and a binding QTL for PU.1 in a B-lymphoblastoid cell line (GM12878). This variant likely affects PU.1 binding by disrupting motifs of its binding partners, such as SP1 and KLF4^{43,44}. Both of these variants demonstrated a significant difference in open chromatin accessibility in the brain between homozygotes for reference and alternative alleles, suggesting functional impact of these variants on the microglial epigenome (Supplementary Figure 6). Our findings in this locus are also supported by a recent study that nominated both rs6733839 and rs13025717 as candidate causal variants in the *BIN1* locus through integration of single-cell epigenomics and a machine learning approach for variant effect prediction⁴⁵. Another recent preprint provided more promising independently derived data that demonstrated a significant allelic imbalance at rs6733839 in iPSC-derived macrophages, further supporting its functional impact on the myeloid epigenome⁴². Additionally, the microglial enhancer that harbors rs6733839 has been recently validated in the *BIN1* locus, where a CRISPR knockout of this regulatory region leads to a microglia-specific reduction in *BIN1* gene and protein expression⁴⁶. We performed conditional analyses using candidate functional variants as covariates and confirmed that they do indeed tag the majority of AD GWAS signal in their respective loci (Supplementary Figure 5). SNP-targeted SMR analyses also confirmed that the prioritized candidate functional variants drive the association between gene expression levels in myeloid cells and AD risk in their respective loci (Supplementary Data 8).

A candidate causal variant in the *MS4A* locus disrupts an anchor CTCF binding site and is associated with reduced chromatin accessibility and increased *MS4A6A* gene expression in myeloid cells. One of the prioritized candidate causal variants in the *MS4A* locus, the rs636317-T AD risk-increasing allele (11:60019150:C:T in GRCh37.p13 coordinates), resides in a CTCF binding site (Fig. 4b (ii)). CTCF binding sites serve as anchors for long-range chromatin loops and this protein plays a pivotal role in determining the spatial organization of chromatin to regulate gene expression⁴⁷. The CTCF motif is highly evolutionarily conserved, and previous studies have shown that single point mutations in this motif can lead to a dramatic reduction of CTCF binding and chromatin accessibility at the site as well as alteration of chromatin looping and activity⁴⁷. We further confirmed that rs636317-T not only resides in a CTCF ChIP-Seq peak in monocytes, but also breaks the CTCF binding consensus sequence (Fig. 4b (iii)) and is a binding QTL for CTCF in a B-lymphoblastoid cell line (GM12878). Additionally, the CTCF binding QTL signal in GM12878⁴⁸ has a 97.6% probability of colocalization with AD risk alleles at this locus. rs636317-T is a strong eQTL for *MS4A6A* in monocytes and macrophages, and the risk increasing T allele is associated with increased *MS4A6A* expression (Fig. 4g). Given that rs636317-T is predicted to disrupt a CTCF binding site, a likely scenario is that this SNP may destroy one of the two anchor CTCF binding sites in a chromatin loop, leading to altered chromatin architecture and activity in the locus, which in turn leads to upregulation of *MS4A6A* expression and increased AD risk. rs636317-T is an hQTL for multiple enhancers in monocytes and a strong eQTL for *MS4A6A* in monocytes and macrophages, reinforcing the hypothesis that rs636317-T causes epigenetic dysregulation in the locus, which in turn may lead to increased expression of *MS4A6A*. Examination of promoter-capture Hi-C interactions in this region in monocytes and macrophages identified chromatin loops that connect the *MS4A6A* promoter to regulatory elements approximately 360 kilobases away (Fig. 4a (vi)). Importantly, examination of ChIA-PET interactions for CTCF and

RAD21 (a component of the cohesin complex often colocalized with CTCF at anchor sites to form chromatin loops⁴⁷) in GM12878 identified a chromatin loop that contains the *MS4A6A* promoter and connects two CTCF/RAD21 anchor sites, one of which is likely disrupted by rs636317-T (Fig. 4a (vii-ix)). This arrangement suggests that rs636317-T may alter chromatin architecture in such a way that the promoter of *MS4A6A* may lose its interaction with the regulatory elements mentioned above and instead fall under the influence of other regulatory elements that may boost *MS4A6A* expression in myeloid cells. Another established role of CTCF is the separation of regions of inner condensed chromatin and outer open chromatin domains, marking repressed and active regions, respectively⁴⁷. Hence, we examined the density of epigenetic signals within and outside the CTCF/RAD21 loop boundaries in monocytes, macrophages and microglia (Fig. 4a (ii-iv)) and observed that chromatin activity within the loop is repressed. To gather additional experimental evidence in support of the epigenetic effects of this genetic polymorphism that we predicted based on computational analysis of experimental data obtained in a B-lymphoblastoid cell line or primary myeloid cells from peripheral blood, we investigated whether the C to T variation at rs636317 results in differential chromatin accessibility at this site in human microglia. To accomplish this, we generated hiPSC-derived microglia (Fig. 4c) from 3 subjects, performed ATAC-Seq and quantified the number of reads that correspond to the protective and risk-increasing alleles. We observed a significant difference in the number of normalized ATAC-Seq reads overlapping rs636317 with the protective allele (C) compared to the risk-increasing allele (T) (P -value = 0.007, paired one-sided t-test) (Fig. 4d). To test whether rs636317-T also leads to an increase in *MS4A6A* expression, we performed RNA sequencing in 4 hiPSC-derived microglia samples. We identified a single synonymous exonic variant in the *MS4A6A* gene (rs12453-C) that is in high LD with the risk variant in the CTCF binding site (rs636317-T, R^2 = 0.92, ADGC reference panel was used to compute LD). Allele specific expression analysis revealed a difference in the number of normalized reads aligned to the T allele versus the C allele that was trending to significance (P -value = 0.088, one-sided paired t-test) (Methods). The direction, however, was the opposite of what is predicted by our analyses using primary myeloid cells from peripheral blood. This phenomenon has also been observed in a recent study showing that in another AD risk locus (*PTK2B*) the direction of the eQTL effect is flipped in hiPSC-derived macrophages as compared to primary blood monocytes and brain microglia⁴². These observations suggest that hiPSC-derived microglia might not be the best model for in-depth studies of the effects of genetic variation on gene expression and chromatin architecture at the *MS4A* and other AD risk loci.

Since a recent single-cell ATAC-seq study in the brain revealed that rs636317 resides in a microglia specific ATAC-seq peak⁴⁹, we utilized brain ATAC-seq data from CommonMind⁵⁰ to test if the ATAC-seq imbalance that we observed in hiPSC-derived microglia can be replicated in primary brain microglia. Indeed, we saw a significant imbalance in normalized ATAC-seq reads consistent with our computational and experimental data (P -value = 0.006, one-sided paired t-test) (Fig. 4e). Since expression of *MS4A6A* is also highly specific to microglia in the brain¹⁵, we performed allele-specific gene expression analysis using brain RNA-seq data from CommonMind (Methods). We observed a significant allelic imbalance (P -value=0.002, one-sided paired t-test) that is consistent with the direction of effect that we predicted using primary myeloid cells from peripheral blood (Fig. 4f-g). We were able to replicate this effect in the Mount Sinai Brainbank (MSBB)⁵¹ RNA-seq dataset, where we also observed a significant allelic imbalance (P -value = 3.0×10^{-5} , one-sided paired t-test). These results are consistent with a model in which the presence of the

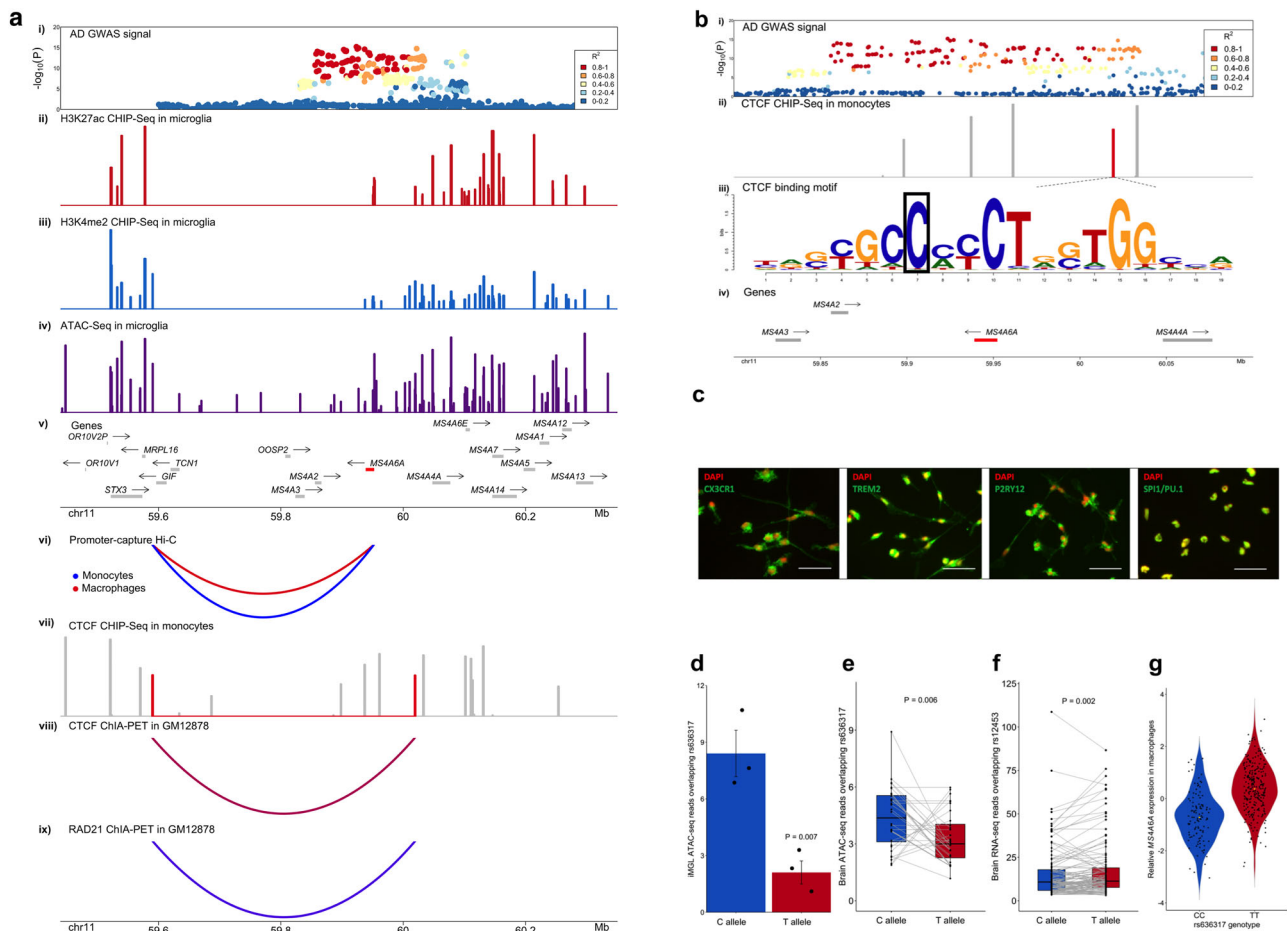


Fig. 4 A candidate causal variant in the *MS4A* locus disrupts an anchor CTCF binding site and is associated with reduced chromatin accessibility and increased *MS4A6A* gene expression in myeloid cells and in the brain. **a** (i) AD GWAS signal in the *MS4A* locus. (ii) H3K27ac peaks in microglia. (iii) H3K4me2 peaks in microglia. (iv) ATAC-seq peaks in microglia. (v) Genes that reside in the locus are plotted. Putative AD risk genes are highlighted in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. (vi) Strongest promoter-capture Hi-C interactions between the *MS4A6A* promoter and distal regulatory elements contained within the CTCF loop in monocytes (blue) and macrophages (red). (vii) CTCF ChIP-Seq peaks in monocytes. The peaks highlighted in red are anchor CTCF binding sites for the chromatin loop. (viii) CTCF ChIA-PET interactions in GM12878. (ix) RAD21 ChIA-PET interaction in GM12878. **b** (i) AD GWAS signal in the *MS4A* locus. (ii) CTCF ChIP-Seq peaks in monocytes. The peak highlighted in red is an anchor CTCF binding site for a chromatin loop and contains the candidate causal variant (rs636317-T). (iii) A CTCF binding motif resides in the CTCF ChIP peak highlighted in red in (ii). The candidate causal variant (rs636317-T) resides in position 7 (boxed) of this motif and is predicted to disrupt CTCF binding. (iv) Genes that reside in the locus are plotted. Putative AD risk genes are highlighted in red. The arrow indicates the direction of transcription, while the bar indicates the gene body. **c** Immunofluorescent images of microglial markers (CX3CR1, TREM2, P2RY12 and PU.1) confirming differentiation of hiPSC-derived microglia. Scale bar = 100 μ m. **d** Allelic imbalance of chromatin accessibility at the rs636317 site is observed in hiPSC-derived microglia. Mean normalized ATAC-Seq read counts are plotted for the protective (C) and risk-increasing (T) alleles; the dots represent each individual, centers for the error bars represent mean normalized ATAC-seq read counts and error bars represent standard errors. The protective allele (C) shows significantly more ATAC-Seq reads than the risk-increasing allele (T) (P -value = 0.007, paired one-sided t -test), which is consistent with the hypothesis that the presence of the rs636317 AD risk-increasing allele leads to disruption of CTCF binding. **e** Allelic imbalance of chromatin accessibility at the rs636317 site is observed in the brain. Each pair of dots connected by a grey line represent an individual. The protective allele (C) shows significantly more ATAC-Seq read counts than the risk-increasing allele (T) (P -value = 0.006, paired one-sided t -test, n = 32), which replicates our observations in hiPSC-derived microglia. **f** Allelic imbalance in normalized brain RNA-seq reads at rs12453 site. Each pair of dots connected by a grey line represent an individual. The protective allele (C) shows significantly less *MS4A6A* RNA-Seq read counts than the risk-increasing allele (T) (P -value = 0.002, paired one-sided t -test, n = 118), which is consistent with our hypothesis. **g** Relative expression of *MS4A6A* in macrophages increases in a rs636317-T allele dose-dependent manner. Each dot represents the relative expression level of *MS4A6A* in each individual, while the yellow dot represents the median. Horizontal lines in box plots depict 25%, 50%, and 75% quantiles; lower whisker = lower hinge - 1.5*IQR; upper whisker = upper hinge + 1.5*IQR.

rs636317-T AD risk-increasing allele leads to disruption of CTCF binding, decreased chromatin accessibility at this site, altered chromatin looping and activity in the locus, and increased expression of *MS4A6A* in microglia. Further investigation of the mechanistic details of this model will require better human microglia culture systems or the use of acutely isolated primary microglia from the brain of larger numbers of human subjects or human-mouse chimeras^{46,52,53}.

Discussion. In this study we report an integration of AD GWAS with epigenomic and transcriptomic datasets from myeloid cells to nominate candidate causal variants, regulatory elements, genes and pathways and thus inform a mechanistic understanding of AD genetics and pathobiology for the formulation of novel therapeutic hypotheses (Supplementary Figure 7). Previous studies have shown that myeloid cells are the most disease-relevant cell type for AD^{7,13} and our own earlier study showed an

enrichment of AD SNP heritability in myeloid-specific epigenomic annotations including the PU.1 cistrome¹². Here we have extended these observations to demonstrate that AD risk alleles are specifically enriched in active enhancers of monocytes, monocyte-derived macrophages and microglia. Concordant with previous studies^{15,20}, we show that PU.1, AP-1, C/EBP, CTCF, and RUNX binding motifs are overrepresented in open chromatin regions associated with active enhancers in all three myeloid cell types, while MEF2 transcription factor binding motifs are highly overrepresented in open chromatin regions associated with microglial active enhancers. To identify transcription factor binding sites burdened by AD risk variants, we stratified open chromatin regions that overlapped with myeloid active enhancers by the presence of cognate consensus motifs for the TFs mentioned above and quantified the enrichment of AD risk alleles in these subsets. A significant enrichment was observed in PU.1 binding motif-positive ATAC-Seq regions in all three myeloid cell types, while MAF binding motif-positive open chromatin regions were specifically enriched in macrophages and microglia. Furthermore, a significant enrichment of AD risk alleles was observed in SMAD, USF and SP binding motif-positive ATAC-Seq regions in microglia. These results suggest that AD risk variants are likely to modify disease susceptibility, at least in part, by modulating the binding of TFs to their cognate sequences in myeloid enhancers thus affecting their activity and in turn leading to target gene expression dysregulation. Although the global enrichment of AD risk alleles in active enhancers of myeloid cells narrows the search space for causal regulatory elements, identifying the target genes of these enhancers would directly point to candidate causal genes in AD risk loci.

In this study we used two complementary approaches to prioritize candidate causal target genes of myeloid active enhancers in AD risk loci. First, we mapped AD risk enhancers to their target genes in myeloid cells using chromatin interactions (Hi-C) and eQTL datasets from monocytes and macrophages. Using this approach, we identified previously nominated AD risk genes (*BINI*²⁶, *MS4A6A*¹², *SPII*¹²) as well as novel candidate causal genes including *AP4E1*, *APPB3*, *RIN3*, *TP53INP1*, and *ZYX* in sixteen loci. In a subset of AD risk loci we report shared active enhancers that interact with multiple target gene promoters to likely regulate their expression. This could reflect the presence of either multiple causal genes at these loci or a single causal gene and several risk neutral genes that show association by virtue of expression co-regulation. Additional evidence will be necessary to distinguish between these two possibilities and prioritize one or more genes at these loci. Second, we used SMR to test the causal relationships between activity at myeloid active chromatin regions with target gene expression regulation and AD risk modification. We sequentially studied the path linking active chromatin region activity with gene expression in myeloid cells using myeloid hQTLs as the exposure and myeloid eQTLs as the outcome, followed by myeloid eQTLs as the exposure and AD diagnosis as the outcome to identify regions that likely modulate AD risk by regulating the expression of one or more of their target genes in myeloid cells. Using this approach, we identified previously nominated AD risk genes *MS4A4A*¹², *MS4A6A*¹², *SPII*¹², as well as novel candidate causal genes *AP4E1*, *AP4M1*, *PILRA*, *RABEP1*, *SPPL2A*, *TP53INP1*, *ZKSCAN1*, and *ZYX* in twelve loci. Importantly, these two analytical approaches yielded largely overlapping results and led to the nomination of several candidate causal genes in twenty loci (Fig. 5). In all twenty loci we mapped candidate causal genes by identifying target genes of AD risk enhancers either through Hi-C interactions or chromatin activity to gene expression SMR associations. For those loci where the gene expression to disease risk association was significant, we were able to assign the directionality of AD risk gene expression

that is associated with increased disease susceptibility (blue for lower expression and red for higher expression). The genes that did not show a significant expression to disease risk association but were prioritized through Hi-C interactions or chromatin activity to gene expression SMR associations, are shown in gold, since a causal association and its directionality cannot be robustly inferred. Moreover, in some of these loci both analytical approaches pointed to the same candidate causal genes (i.e. *BINI*, *MS4A*, *SPII*, *RABEP1*, *TP53INP1*, and *ZYX*). Remarkably, when a *BINI* enhancer prioritized through our approaches was deleted in hiPSC-derived microglia, neurons and astrocytes, *BINI* expression and protein level dramatically decreased only in microglia, underpinning cell type-specific regulatory potential of a rather ubiquitously expressed gene and pointing to the robustness of our findings⁴⁶.

Notably, many of the candidate causal genes that we identified in myeloid cells are functionally related to the endolysosomal system. For example, *ZYX* encodes a zinc-binding phosphoprotein that localizes to early endosomes and phagosomes in IFN- γ -activated macrophages⁵⁴ and drives their intracellular movement by assembling actin filament rocket tails⁵⁵. *RIN3* (Ras And Rab Interactor 3) encodes a member of the RIN family of RAS and RAB effectors that interacts and localizes with *BINI* to early endosomes⁵⁶. Like other RIN family members, *RIN3* has guanine nucleotide exchange factor (GEF) activity for RAB5 GTPases⁵⁶, which are required for early endosome and phagosome biogenesis and function. Interestingly, *RABEP1* (Rab-GTPase binding effector protein 1) also encodes a RAB5 effector protein that is required for early endosome membrane fusion and trafficking⁵⁷. Two other novel candidate AD risk genes that we nominated in this study, *AP4E1* and *AP4M1*, encode two of the four subunits of the heterotetrameric adaptor protein complex 4 (AP-4), which is required for the sorting of transmembrane proteins like APP from the trans-Golgi network (TGN) to endosomes⁵⁸. Interestingly, *APBB3* has also been shown to bind to the intracellular domain of APP and is thought to play a role in the internalization of APP from the cell surface into endosomes where it is cleaved by membrane-embedded aspartyl proteases BACE1 and γ -secretase to generate the amyloid β peptide^{59,60}. Another novel candidate AD risk gene that we nominate in this study, *SPPL2A*, encodes a transmembrane aspartyl protease that localizes to late endosomes and lysosomes and cleaves substrates involved in immunity and neurodegeneration^{61–63}. Finally, *TP53INP1* regulates the stability and transcriptional activity of p53, and has been implicated in the phagocytic clearance of apoptotic cells (efferocytosis)^{64,65}, a hallmark function of macrophages for the maintenance of tissue homeostasis and immune tolerance, and the resolution of inflammation. All of these genes are highly or selectively expressed in microglia in the brain¹⁵. Taken together, our findings implicate dysfunction of the endolysosomal system in myeloid cells (as opposed to neurons⁶⁶) in the etiology of AD. Previous human genetic findings reinforce our conclusion. For example, a rare variant in the 3' UTR of *RAB10*, a member of the RAB family of small GTPases that are critical regulators of membrane trafficking and vesicular transport, confers resilience to AD⁶⁷. Furthermore, coding variants that increase risk for AD have been identified in *SORL1*^{4,68}, a member of the vacuolar protein sorting 10 (VPS10)-domain-containing receptor family and the low density lipoprotein receptor (LDLR) family of APOE receptors that is expressed primarily in microglia in the brain¹⁵ and plays important roles in the endolysosomal system and APP processing⁶⁶.

To fine-map the AD risk enhancers identified in this study and thus nominate candidate causal variants, we conducted Bayesian fine-mapping in the three loci that were significantly associated with AD risk in the ADGC GWAS (*BINI*, *MS4A*, and *ZYX*),

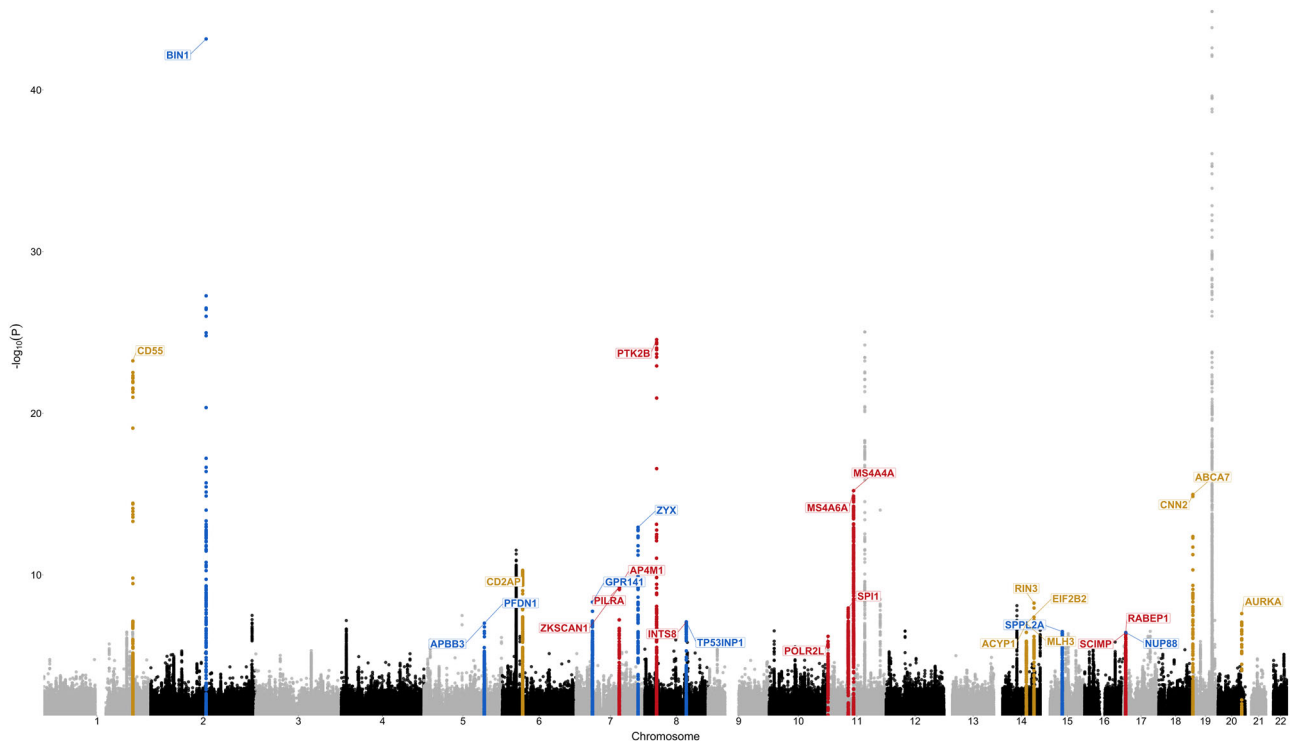


Fig. 5 Candidate causal genes nominated through both Hi-C and SMR approaches in twenty loci. The Manhattan plot depicts the IGAP GWAS signal with putative AD risk genes assigned to each locus through both Hi-C and SMR approaches. Red indicates that increased expression of the gene is predicted to increase risk for AD. Blue indicates that decreased expression of the gene is predicted to increase risk for AD. Gold indicates that the directionality of gene expression that is associated with increased disease susceptibility cannot be robustly inferred. These genes were prioritized if they either a) interact with an AD risk enhancer that contains an eQTL for this gene or b) were implicated in enhancer activity to gene expression association, but did not have significant expression to disease risk associations (SMR). ZYX and PTK2B showed opposite directions of expression associated with disease risk in monocytes and macrophages. Strongest associations are reported (macrophages in ZYX locus and monocytes in PTK2B locus). The TREM2 locus is not shown since a well replicated rare loss-of-function mutations were found in TREM2³. The PICALM locus is not shown since the prioritized gene is not expressed in microglia.

followed by functional in silico screening of the candidate causal variants for disruption/creation of TF binding motifs. We also fine-mapped the loci that did not reach significance in the ADGC GWAS (but were significant or suggestive in the IGAP GWAS) and identified candidate causal variants in the *GPR141*, *RABEP1*, *SP11*, and *SPPL2A* loci. Taken together, we have identified putative functional variants that tag the majority of AD GWAS signals at these loci, and likely affect disease risk by altering the DNA binding motifs of transcription factors that modulate the activity of enhancers which in turn regulate the expression of causal genes to ultimately steer myeloid cells like microglia toward neurotoxic and/or away from neuroprotective phenotypes. Finally, we experimentally validated one of these candidate functional variants in the *MS4A* locus by showing allelic imbalance in open chromatin in hiPSC-derived microglia as well as in open chromatin and *MS4A6A* mRNA levels in the brain. The epigenetic effects of this variant are likely mediated by the disruption of CTCF binding at one of two anchor sites of a repressive chromatin loop leading to increased *MS4A6A* expression and AD risk, although investigation of the mechanistic details of this model will require further experimentation.

Our analyses demonstrate that active enhancers in monocytes, macrophages and microglia are enriched significantly and to a similar extent. These results provide evidence that AD risk alleles burden regulatory sequences similarly across all three myeloid cell types and that the basal state is, at least in part, relevant to the study of regulatory variants that affect AD risk. Recent findings that TREM2 loss of function similarly impacts the response of

both central nervous system (CNS) and peripheral macrophages to lipid overload^{69–71} and that the activation state of human macrophages does not have a major impact on AD heritability enrichment⁷² could indicate that Alzheimer's disease-associated variants might regulate core functions of the macrophage lineage (e.g., the phagocytic clearance of apoptotic cells and other lipid-rich cellular debris). These results highlight the need to generate additional large-scale human microglial/myeloid epigenomic and transcriptomic datasets (e.g., in the context of immune and metabolic stress) which will enable identification of the most disease-relevant myeloid cell states and enable replication and extension of our findings.

The integrative genomic approaches presented here offer a framework to identify regulatory elements, genes and variants that are likely causal for AD. A potential limitation of our study is that integration of epigenomic and transcriptomic datasets from different studies using varying protocols for the isolation and preparation of monocytes and macrophages, might lead to false positive and negative results in some of our analyses. This highlights the need for paired epigenomic and transcriptomic datasets in myeloid cells to further validate and expand our findings. Further experimental validation of the variants and enhancers nominated in this manuscript will be needed to dissect the molecular mechanism of action as well as downstream effects in myeloid cells. Using our prediction as a guiding tool, CRISPR experiments can be performed to test the effects of a single variant or regulatory elements in isogenic lines on TF binding, gene expression and downstream myeloid cell biology, e.g

phagocytosis of lipid-rich debris. Additionally, recent studies have demonstrated that iPSC-derived microglia can be transplanted into the mouse brain while recapitulating expression profiles of human primary microglia⁵². These advances can be utilized to transplant iPSC-derived microglia lines with CRISPR induced alterations to study the effects of non-coding AD risk variants and regulatory elements in vivo.

In summary, this study reveals a link between chromatin activity, gene expression and AD risk in myeloid cells, proposes the molecular mechanism of action of candidate functional variants in several AD risk loci, identifies specific AD risk enhancers that are burdened by these variants and regulate target gene expression, which in turn most likely modulates disease susceptibility by altering the biology of myeloid cells. We highlight the coalescence of candidate causal genes in the endolysosomal system of myeloid cells and underscore its importance in the etiology of AD.

Methods

Processing of ChIP-Seq and ATAC-Seq data and peak calling. Relevant ChIP-Seq and ATAC-Seq studies were found through Gene Expression Omnibus (GEO)^{15,36,38,73}. We selected studies that contained H3K4me1 (monocytes and macrophages), H3K4me2 (monocytes and microglia) and/or H3K4me3 (macrophages) as well as H3K27ac (all cell types) and ATAC-Seq (all cell types) data for human monocytes, macrophages and microglia for our analyses. To generate the epigenomic annotations FASTQ files were obtained from Sequence Read Archive (SRA). Technical replicates were merged and Bowtie⁷⁴ was used for alignment for both single and paired-end files. FASTQC was used for quality control of the files. Resulting SAM files were filtered by MAPQ score and duplicates were removed using samtools⁷⁵. MACS2⁷⁶ was used to call peaks for ATAC-Seq and ChIP-Seq files. ATAC-Seq peaks were called using the following command: “callpeak -t file.sam -f SAM --nomodel --shift -37 --extsize 73 -g hs -q 0.01 -n filename --outdir output_dir/”. PU.1 ChIP-Seq peaks were called using the following command: callpeak -t case.sam -c input.sam -f SAM -g hs -q 0.01 -n filename --outdir output_dir/”. Histone modifications ChIP-Seq peaks were called using the following command: “callpeak -t case.sam -c input.sam -f SAM --broad --broad-cutoff 0.01 -g hs -q 0.01 -n filename --outdir output_dir/”.

Stratification into promoter and enhancer regions and overlap with GWAS and Hi-C data.

To identify optimal distance from TSS we used ChromHMM model of CD14 + monocytes from Roadmap Epigenomics project (see URLs) to visualize the distribution of active promoters around the TSS. We observed a bimodal distribution around the TSS and found that -500 base pairs to 1000 base pairs window captures more than 60% of active promoters. Based on previous studies that have demonstrated a bimodal distribution of promoter epigenomic marks around the TSS^{77,78}, we established that the boundary of -500, 1000 bp would appropriately mark active promoters, while also not misclassifying H3K4me1 positive regions (enhancers) that are in close proximity to the TSS. To annotate the peaks with distance from TSS we used HOMER. We then split the H3K4me1/2/3 peaks into distal and proximal. We then used bedmap to filter H3K4me1/2/3 peaks by the presence of H3K27ac peak such that proximal H3K4me2/3 peaks with H3K27ac were classified as active promoters, distal H3K4me1/2 peaks with H3K27ac were classified as active enhancers, proximal H3K4me2/3 peaks without H3K27ac were classified as primed promoters and distal H3K4me1/2 peaks without H3K27ac were classified as primed enhancers. AD risk enhancers were identified by overlapping active enhancers (including a 500-bp flanking region on each side) with AD risk alleles ($P \leq 1 \times 10^{-6}$). To identify likely targets of AD risk enhancers, enhancers (including a 3000-bp flanking region on each side) were overlapped with Hi-C target regions that showed evidence of regulatory effect (eQTL FDR 5%).

Partitioned SNP-heritability analysis. We used LD Score regression to estimate AD SNP heritability partitioned by epigenomic annotations using GWAS summary statistics (excluding the APOE (chr19:45000000–45800000) and MHC/HLA (chr6:28477797–33448354) regions) in myeloid cells as described in the companion website (see URLs), while controlling for the 53 functional annotation categories of the full baseline model. GWAS summary statistics for AD¹⁷ and Schizophrenia¹⁸ (SCZ) were downloaded from the IGAP Consortium and Psychiatric Genomics Consortium websites respectively (see URLs). All epigenomic annotations were downloaded from SRA and processed as described in “Processing of ChIP-Seq and ATAC-Seq data and peak calling”. Negative log₁₀ *p*-values of enrichment were reported, the *p*-values for annotations that had negative enrichments were not displayed on the figures.

De novo motif discovery. We used HOMER to perform de novo motif discovery in ATAC-Seq regions that reside in active enhancers in monocytes, macrophages and microglia. The following command was used to identify enriched motif sequences in these regions: findMotifsGenome.pl Peaks.bed hg19. -size given. To identify regions that contained our motifs of interest, we used the following commands: findMotifsGenome.pl Peaks.bed hg19. -find motif.motif -size given and annotatePeaks.pl Peaks.bed hg19 -m motif.motif -size given.

Colocalization analysis. We used coloc (coloc.abf function) to perform colocalization analyses between IGAP GWAS and hQTLs with default parameters²⁹. We used coloc in the following manner: coloc.abf(dataset1, dataset2, p1 = 1e-04, p2 = 1e-04, p12 = 1e-05). We used a filter of PP.H3.abf + PP.H4.abf ≥ 0.8 to select chromatin regions with evidence of independent or colocalized AD GWAS and hQTL signals.

Causal association analysis. We used SMR to test for causal associations between IGAP GWAS and QTL datasets¹⁴. We converted the summary statistics for monocyte H3K4me1 hQTLs obtained from BLUEPRINT epigenome project website (see URLs) and monocyte eQTLs from the Cardiogenics and Fairfax studies into BESD format (epi/esi/besd) as described in the SMR manual (see URLs). Allele frequencies and LD were estimated from the ADGC GWAS cohort individual-level genotype data using plink⁷⁹. To conduct standard SMR analysis, we ran the following command: “smr --bfile reference_file --beqtl-summary Exposure_besd_file_prefix --beqtl-summary Outcome_besd_file_prefix --out output_prefix”. The results were filtered for FDR of 5% calculated using the p.adjust function in R. To conduct SNP-targeted SMR analysis, we ran the following command: “smr --bfile reference_file --gwas-summary gwas_summary_file --beqtl-summary eQTL_besd_file_prefix --target-snp rs12345 --out output_prefix”.

Conditional and haplotype analyses. We used GCTA-COJO⁴¹ to conduct conditional analyses using IGAP GWAS summary statistics data and ADGC GWAS cohort individual-level genotype data as a reference panel. To conduct the conditional analysis we ran the following command: “gcta64 --bfile reference_file --maf 0.05 --cojo-file GWAS_summary_statistics --cojo-cond list_of_snps --out output_prefix”. To construct haplotype blocks and examine SNP clustering, we used Big-LD⁴⁰ which is provided as an R package. We prepared the genotype file, which contained genotypes of individuals for each SNP, and the SNP information file that contained chromosome, position, reference and alternative allele information for each SNP. We then used the CLQ algorithm provided within the Big-LD package for SNP clustering and BigLD for haplotype block construction. We used the following commands: CLQD(geno = genotype_data, SNPinfo = locus_snp_information, hrstType = “fast”, CLQcut = 0.5) and Big_LD(genotype_data, SNPinfo = locus_snp_information, chrN = chromosome_number, startbp = start_basepair, endbp = end_basepair, appendRare = TRUE).

Prioritization of candidate causal variants. For each locus we constructed LD blocks using Big-LD package⁴⁰. We then selected variants that reside in active enhancers in monocytes, macrophages and/or microglia (with the exception of the *SPPL2A* locus, since these variants likely regulate a distal enhancer as reported in Fig. 3c). We also conducted a motif disruption/creation analysis on these variants and selected the variants that are predicted to strongly disrupt or create binding sites of transcription factors that are expressed in myeloid cells (TPM ≥ 1)¹⁵. We then screened the remaining variants for eQTLs in monocytes and macrophages from the Cardiogenics and Fairfax studies. We also used PAINTOR to conduct Bayesian fine-mapping in MS4A, ZYX and BIN1 loci. PAINTOR is a Bayesian fine-mapping method that leverages functional annotations through an Empirical Bayes prior³². The input files for PAINTOR (v3.1) were prepared as described on the PAINTOR website and ADGC GWAS summary statistics along with individual-level genotype data were used for fine-mapping (see URLs). The reprocessed epigenomic annotations were used to quantify enrichment at each locus. To quantify the annotation enrichments the following command was used: “python AnnotateLocus.py --input list_of_annotation_directories --locus locus_prefix --out output_prefix --chr chr --pos pos”. To classify the annotations as enriched or not, we computed the relative probability for a SNP to be causal given that it resides in the annotation as described in the companion website (URLs). We deemed the annotation to be significant if the relative probability of a SNP to be causal given that it is in the annotation was greater than 1. To quantify the posterior probabilities for variants to be causal, we used the following command: PAINTOR -input input.file -Zhead Zscore -LDname ld -enumerate max_number_of_causal_variants -annotations annotation_name -in in_dir -out out_dir. Once candidate causal variants were selected through both approaches, we conducted conditional analyses to make sure that they do indeed tag the majority of the GWAS signal in the locus.

TF binding motif disruption/creation analysis. We used motifbreakR to predict the impact of AD risk variants on transcription factor binding³⁹. We used HOCO MOCO to screen for TFBMs and a *P*-value significance threshold of 5×10^{-5} . We used the following command to do so: motifbreakR(snpList = variant_list, pwmList = hocomoco, filterp = TRUE, threshold = 5e-5).

Generation of hiPSC microglia for ATAC-Seq and RNA-seq analysis. hiPSC-derived microglia were generated from patient lines following the protocol as described⁸⁰. For the ATAC-Seq analysis, hiPSC-derived microglia (50 K cells) from each patient line were collected and processed as described⁸¹. Samples were either processed at New York Genome Center or at UCI's Genomics High-Throughput Facility and sequenced as 50 bp paired-end reads on a HiSeq 2500 and 100 bp paired-end reads on a HiSeq 4000, respectively. The consent for reprogramming patient somatic cells to hiPSC was carried out on protocol 2013-9561 (UCI), laboratory protocol 2017-1061 (UCI) and protocol ESCRO 19-04 (Mount Sinai). Microglia RNA was isolated using a standard RNA isolation kit (Qiagen) and RNA quality (RIN) assessed (Bioanalyzer 2100). PolyA-mRNA (200 ng) with a RIN score ≥ 9.5 was used to assemble libraries in which ERCC spike-ins (Ambion) were included for downstream normalization. RNA-seq libraries were quantified and normalized using a Library Quantification Kit (Kapa Biosystems) prior to sequencing (Illumina) by the UCI Genomics High Throughput Facility as 100 bp paired-end reads.

Human iPSC cell lines were either generated by the University of California, Irvine Alzheimer's Disease Research Center (UCI ADRC) Induced Pluripotent Stem Cell Core or by the Icahn School of Medicine at Mount Sinai Induced Pluripotent Stem Cell Core. The iPSC lines generated by University of California, Irvine and the Icahn School of Medicine at Mount Sinai were derived from subject fibroblasts from either the University of California, Irvine or Washington University in St. Louis, respectively, with approved Institutional Review Boards (IRB) and human Stem Cell Research Oversight (hSCRO) committee protocols. Informed consent was received by each of the participants who donated fibroblasts.

Allele specific expression and open chromatin analysis. In CommonMind and Mount Sinai Biobank (MSBB) datasets we selected the RNA-seq samples that contained at least 10 reads aligned to the SNP of interest. For CommonMind ATAC-seq samples, we required at least 5 reads aligned to the SNP of interest. To perform allele specific expression/open chromatin analyses, we have quantified the number of reads overlapping the variant of interest using mpileup command in samtools⁷⁵. The CommonMind and MSBB reads were normalized to the number of reads on chromosome 11 and were used to assess the significance of the allelic imbalance using a paired t-test.

Immunocytochemistry. Cells were fixed with 4% paraformaldehyde in PBS at 4 °C for 10 min. Cells were permeabilized with 1.0% Triton in PBS at room temperature for 15 min and blocked in 5% donkey serum with 0.1% Triton in PBS at room temperature for 30 min. Primary antibodies were used at 10 μ g/mL anti-TREM2 (R&D, AF1828), 1:1,000 anti-P2RY12 (Sigma, HPA014518), 1:100 anti-PU.1 (Cell Signaling, 2266) and anti-CX3CR1 (Bio-Rad, AHP1589). Secondary antibodies were used at 1:300 Alexa donkey 488 and 568 anti-rabbit, mouse, or chicken (Life Technologies). DAPI (4',6-diamidino-2-phenylindole, 0.5 μ g/mL) was used to visualize nuclei. Images were acquired using a Leica Fluorescence Microscope at 40 \times magnification.

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

Data availability

The following studies obtained from GEO were used for the analyses presented in this paper: [GSE29611](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29611) (monocyte CTCF, H3K27ac and H3K4me1/2/3), [GSE85245](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85245) (macrophage H3K27ac and H3K4me1/3), [GSE100380](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE100380) (monocyte and macrophage ATAC-seq), [GSE66594](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66594) (macrophage H3K4me1, H3K27ac and PU.1) and [GSE98365](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98365) (macrophage ATAC-seq). Data generated in this study are available through accession number [GSE164315](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164315). DbGAP accession study number for the human microglia dataset is phs001373.v2.p2. The genotype and phenotype data from ADGC are available under phs000372.v1.p1 dbGAP study accession number. The Cardiogenics dataset can be requested on EGA using accession number EGAS00001000411. DbGAP accession study number for the STARNET eQTL dataset is phs001203.v1.p1. Summary statistics for Fairfax eQTL data can be obtained from ArrayExpress using accession number E-MTAB-2232. All data supporting the findings of this study are provided within the paper and its supplementary information. All other relevant data are available from the authors upon reasonable request.

Code availability

We provide commands for the tools that were used for the analyses presented in this manuscript in the Methods section. Although we have used the software cited in this manuscript with default parameters or minor changes, code for these analyses is available upon request. Roadmap Epigenomics Project, <http://www.roadmappigenomics.org/> - The dataset was used to examine the distribution of active promoters around the TSS in the ChromHMM model of CD14+ monocytes to identify an appropriate window to stratify enhancers and promoters. LD Score Regression, <https://github.com/bulik/ldsc>; - LDSC was used to quantify the enrichment of AD heritability in myeloid epigenomic annotations. International Genomics of Alzheimer's Project (IGAP), http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php; - The summary statistics for AD GWAS were obtained from IGAP. Psychiatric Genomics Consortium, <https://www.med.unc.edu/pgc/>; - SCZ GWAS summary statistics were obtained from PGC. Blueprint Consortium, <http://www.blueprint-epigenome.eu/>; - Monocyte hQTLs were obtained from the Blueprint Consortium. SMR, <https://cns.genomics.com/software/smr/>; - SMR was used to test for candidate causal associations between chromatin activity, gene expression and disease risk. COJO, <https://cns.genomics.com/software/gcta/#COJO>; - COJO was used to conduct conditional analyses. PAINTOR, https://github.com/gkichaev/PAINTOR_V3.0; - PAINTOR was used to conduct Bayesian fine-mapping analyses. ADGC, <https://brightspotcdn.byu.edu/bf/be/75f2076b4241a30840eadda2c66c/adgc-combined-1000g-09192014.pdf>; - ADGC genotype data were used for fine-mapping and as an LD reference panel. HOMER, <http://homer.ucsd.edu/homer/>; - HOMER was used for de novo motif enrichment analyses.

Received: 16 September 2020; Accepted: 10 February 2021;
Published online: 12 March 2021

References

1. Dementia statistics | Alzheimer's Disease International. <https://www.alz.co.uk/research/statistics>.
2. Efthymiou, A. G. & Goate, A. M. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. *Mol. Neurodegener.* **12**, 43 (2017).
3. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
4. Vardarajan, B. N. et al. Coding mutations in SORL1 and Alzheimer disease. *Ann. Neurol.* **77**, 215–227 (2015).
5. Sims, R. et al. Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
6. Steinberg, S. et al. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. *Nat. Genet.* **47**, 445–447 (2015).
7. Hansen, D. V., Hanson, J. E. & Sheng, M. Microglia in Alzheimer's disease. *J. Cell Biol.* **217**, 459–472 (2018).
8. Andrews, S. J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol.* (2020) [https://doi.org/10.1016/S1474-4422\(19\)30435-1](https://doi.org/10.1016/S1474-4422(19)30435-1).
9. Kunkle, B. W. et al. Meta-analysis of genetic association with diagnosed Alzheimer's disease identifies novel risk loci and implicates Abeta, Tau, immunity and lipid processing. <https://doi.org/10.1101/294629>.
10. Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 99 (2018).
11. 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).
12. Huang, K.-L. et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat. Neurosci.* **20**, 1052–1061 (2017).
13. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
14. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
15. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, 3222 (2017).
16. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228 (2015).
17. Lambert, J.-C. Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease. *Nat. Genet.* **45**, 1452–1458 (2013).
18. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
19. Javierre, B. M. et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384.e19 (2016).
20. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
21. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
22. Garnier, S. et al. Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes. *PLoS Genet.* **9**, e1003240 (2013).
23. Fairfax, B. P. et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* (2014) <https://doi.org/10.1126/science.1246949>.
24. Franzen, O. et al. Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* **353**, 827–830 (2016).

25. Reitz, C. et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ 4, and the risk of late-onset Alzheimer disease in African Americans. *JAMA* **309**, 1483–1492 (2013).
26. Chapuis, J. et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* **18**, 1225–1234 (2013).
27. Raj, T. et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018).
28. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune. *Cells* **167**, 1398–1414.e24 (2016).
29. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
30. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
31. Rathore, N. et al. Paired Immunoglobulin-like Type 2 Receptor Alpha G78R variant alters ligand binding and confers protection to Alzheimer's disease. *PLoS Genet.* **14**, e1007427 (2018).
32. Kichaev, G. et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
33. 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).
34. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
35. Novakovic, B. et al. β -glucan reverses the epigenetic state of LPS-induced immunological tolerance. *Cell* **167**, 1354–1368.e14 (2016).
36. Park, S. H. et al. Type I interferons and the cytokine TNF cooperatively reprogram the macrophage epigenome to promote inflammatory activation. *Nat. Immunol.* **18**, 1104–1116 (2017).
37. Kang, K. et al. Interferon- γ represses M2 gene expression in human macrophages by disassembling enhancers bound by the transcription factor MAF. *Immunity* **47**, 235–250.e4 (2017).
38. Schmidt, S. V. et al. The transcriptional regulator network of human inflammatory macrophages is defined by open chromatin. *Cell Res.* **26**, 151–170 (2016).
39. Coetzee, S. G., Coetzee, G. A. & Hazelett, D. J. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. *Bioinformatics* **31**, 3847–3849 (2015).
40. Kim, S. A., Cho, C.-S., Kim, S.-R., Bull, S. B. & Yoo, Y. J. A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. *Bioinformatics* **34**, 388–397 (2018).
41. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). S1–3.
42. Young, A., Kumasaka, N., Calvert, F. & Hammond, T. R. A map of transcriptional heterogeneity and regulatory variation in human microglia. *bioRxiv* (2019).
43. Feng, X. et al. Sp1/Sp3 and PU.1 differentially regulate β 5 integrin gene expression in macrophages and osteoblasts. *J. Biol. Chem.* **275**, 8331–8340 (2000).
44. Feinberg, M. W. et al. The Kruppel-like factor KLF4 is a critical regulator of monocyte differentiation. *EMBO J.* **26**, 4138–4148 (2007).
45. 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.* (2020) <https://doi.org/10.1038/s41588-020-00721-x>.
46. Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134–1139 (2019).
47. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
48. Ding, Z. et al. Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. *PLoS Genet.* **10**, e1004798 (2014).
49. Corces, M. R. et al. Single-cell epigenomic identification of inherited risk loci in Alzheimer's and Parkinson's disease. <https://doi.org/10.1101/2020.01.06.896159>.
50. Hoffman, G. E. et al. CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and bipolar disorder. *Sci. Data* **6**, 180 (2019).
51. Wang, M. et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci. Data* **5**, 180185 (2018).
52. Mancuso, R. et al. Stem-cell-derived human microglia transplanted in mouse brain to study human disease. *Nat. Neurosci.* **22**, 2111–2116 (2019).
53. Hasselmann, J. et al. Development of a chimeric model to study and manipulate human microglia in vivo. *Neuron* **103**, 1016–1033.e10 (2019).
54. Trost, M. et al. The phagosomal proteome in interferon-gamma-activated macrophages. *Immunity* **30**, 143–154 (2009).
55. Southwick, F. S., Li, W., Zhang, F., Zeile, W. L. & Purich, D. L. Actin-based endosome and phagosome rocketing in macrophages: activation by the secretagogue antagonists lanthanum and zinc. *Cell Motil. Cytoskeleton* **54**, 41–55 (2003).
56. Kajih, H. et al. RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway. *J. Cell Sci.* **116**, 4159–4168 (2003).
57. Stenmark, H., Vitale, G., Ullrich, O. & Zerial, M. Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. *Cell* **83**, 423–432 (1995).
58. Burgos, P. V. et al. Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex. *Dev. Cell* **18**, 425–436 (2010).
59. Duilio, A., Faraonio, R., Minopoli, G., Zambrano, N. & Russo, T. Fe65L2: a new member of the Fe65 protein family interacting with the intracellular domain of the Alzheimer's beta-amyloid precursor protein. *Biochem. J.* **330**(Pt 1), 513–519 (1998).
60. Tanahashi, H. & Tabira, T. Molecular cloning of human Fe65L2 and its interaction with the Alzheimer's beta-amyloid precursor protein. *Neurosci. Lett.* **261**, 143–146 (1999).
61. Behnke, J. et al. Signal-peptide-peptidase-like 2a (SPPL2a) is targeted to lysosomes/late endosomes by a tyrosine motif in its C-terminal tail. *FEBS Lett.* **585**, 2951–2957 (2011).
62. Schneppenheim, J. et al. The intramembrane protease SPPL2a promotes B cell development and controls endosomal traffic by cleavage of the invariant chain. *J. Exp. Med.* **210**, 41–58 (2013).
63. Brady, O. A., Zhou, X. & Hu, F. Regulated intramembrane proteolysis of the frontotemporal lobar degeneration risk factor, TMEM106B, by signal peptide peptidase-like 2a (SPPL2a). *J. Biol. Chem.* **289**, 19670–19680 (2014).
64. Shahbazi, J., Lock, R. & Liu, T. Tumor protein 53-induced nuclear protein 1 enhances p53 function and represses tumorigenesis. *Front. Genet.* **4**, 80 (2013).
65. Yoon, K. W. et al. Control of signaling-mediated clearance of apoptotic cells by the tumor suppressor p53. *Science* **349**, 1261669 (2015).
66. Small, S. A., Simoes-Spassov, S., Mayeux, R. & Petsko, G. A. Endosomal traffic jams represent a pathogenic hub and therapeutic target in Alzheimer's disease. *Trends Neurosci.* **40**, 592–602 (2017).
67. Ridge, P. G. et al. Linkage, whole genome sequence, and biological data implicate variants in RAB10 in Alzheimer's disease resilience. *Genome Med.* **9** (2017).
68. Raghavan, N. S. et al. Whole-exome sequencing in 20,197 persons for rare variants in Alzheimer's disease. *Ann. Clin. Transl. Neurol.* **5**, 832–842 (2018).
69. Jaitin, D. A. et al. Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell* **178**, 686–698.e14 (2019).
70. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
71. Nugent, A. A. et al. TREM2 regulates microglial cholesterol metabolism upon chronic phagocytic challenge. *Neuron* **105**, 837–854.e9 (2020).
72. Soskic, B. et al. Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases. *Nat. Genet.* **51**, 1486–1493 (2019).
73. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
74. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
75. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
76. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
77. Sati, S., Ghosh, S., Jain, V., Scaria, V. & Sengupta, S. Genome-wide analysis reveals distinct patterns of epigenetic features in long non-coding RNA loci. *Nucleic Acids Res.* **40**, 10018–10031 (2012).
78. Zhang, Z. et al. H3K4 tri-methylation breadth at transcription start sites impacts the transcriptome of systemic lupus erythematosus. *Clin. Epigenetics* **8**, 14 (2016).
79. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
80. Abud, E. M. et al. iPSC-Derived Human Microglia-like Cells to Study Neurological Diseases. *Neuron* (2017) <https://doi.org/10.1016/j.neuron.2017.03.042>.
81. Buenostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

Acknowledgements

We thank the Cardiogenics (European Project reference LSHM-CT-2006-037593) project for providing summary statistics data for eQTL analyses in monocytes and monocyte-derived macrophages. We thank the New York Genome Center for sequencing the hiPSC-derived microglia samples.

Author contributions

A.M.G., E.M. and G.N. conceived and designed the experiments. G.N., M.K., S.X.C., H.C., K.H., and A.G.E. performed data analyses. J.T.C.W. and E.M.A. carried out the

hiPSC-derived microglia studies under the supervision of A.M.G. and W.W.P. P.R., J.F.F., and Ja.B. performed ATAC-seq experiments. Yi.L. analyzed RNA-seq data from hiPSC-derived microglia. E.M., J.B., Y.L., and A.M.G. supervised data analysis. G.N., E.M., and A.M.G. wrote and edited the manuscript. All authors read and approved the manuscript.

Competing interests

A.M.G. has consulted for Eisai, Biogen, Pfizer, AbbVie, Cognition Therapeutics and GSK. She also served on the SAB at Denali Therapeutics from 2015 to 2018. This work was funded by grants from the NIH: U01AG052411 (A.M.G.), RF1AG054011 (A.M.G.), U01AG058635 (A.M.G.), NIA K01AG062683 (J.TCW.), AG016573 (W.W.P.), F31 AG059337-01 (A.G.E.), R01AG050986 (P.R.), 1R01ES029212-01 (K.H.), R01HL125863 (J.L.M.B.), American Heart Association (J.L.M.B.), the Swedish Research Council (J.L.M.B.), Heart Lung Foundation (J.L.M.B.), and by Astra-Zeneca through ICMC, Karolinska Institutet (J.L.M.B.), The JPB Foundation, The Robert and Renee Belfer Foundation. E.M.A. and W.W.P. are named co-inventors of patent WO/2018/160496 related to the differentiation and use of human pluripotent stem cells and hematopoietic progenitors into microglia. K.H. receives financial compensation from Sema4 (an Icahn School of Medicine at Mount Sinai spin-off company). Sema4 is currently majority owned by the Icahn School of Medicine at Mount Sinai. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-21823-y>.

Correspondence and requests for materials should be addressed to E.M. or A.M.G.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021