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

Gloria Novikova1, Manav Kapoor1,2, Julia TCW1,2, Edsel M. Abud3,4, Anastasia G. Efthymiou1, Steven X. Chen5,6, Haoxiang Cheng2, John F. Fullard2,7, Jaroslav Bendl2,7, Yiyuan Liu1, Panos Roussos2,7, Johan LM Björkegren2,8, Yunlong Liu5,6, Wayne W. Poon9, Ke Hao2, Edoardo Marcora1,2,10 & Alison M. Goate1,2,10

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.
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 available\(^1\). Several lines of genetic evidence implicate myeloid cells in the etiology of AD\(^2\). Whole-exome sequencing and microarray studies have identified rare coding variants associated with AD in genes (e.g., TREM2\(^3\), SORL1\(^4\), AB\(^5\), PLCG2 and ABCAP4\(^6\)) 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 types\(^7\). Genome-wide association studies (GWAS) have identified common non-coding variants associated with AD in more than 40 loci\(^8\), 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 monocyes) 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 monocyes, 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 monocyes and macrophages. This approach allows us to nominate candidate causal genes in eleven genomewide significant and five suggestive AD risk loci, including TP53INP1, APBB3, RABEP1, and SPP1L2A. 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 MS4A4a locus in human induced pluripotent stem cell (hiPSC)-derived microglia and brain.

**Results**

**AD risk alleles are specifically enriched in active enhancers of monocyes, 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 monocyes, macrophages, and microglia to annotate the genome with myeloid active enhancers (AE), active promoters (AP), primed enhancers (PE) and primed promoters (PP) (see Methods)\(^3\). 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 monocyes, 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 dataset\(^17\). This analysis revealed selective enrichment of AD risk alleles in active enhancers of monocyes, macrophages, and microglia (Fig. 1a). In contrast, schizophrenia SNP heritability (using the Psychiatric Genomics Consortium SCZ GWAS dataset as control)\(^18\) 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 analysis\(^19\) 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, CTCEF, and RUNX binding motifs. The binding motif for MEF2 family TFs (which includes MEF2C in another AD risk locus\(^17\)) was the best match for highly overrepresented sequence motif in active enhancers of human microglia, consistent with findings in mouse microglia\(^20\). 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 monocyes, macrophages and microglia (TPM ≥ 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 microglia\(^21\). These results show that AD risk alleles are specifically enriched in active enhancers of monocyes, 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...
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 genes. 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 previously. We first identified active enhancers in monocytes and macrophages that contain AD risk alleles (P ≤ 1 × 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 Javiereve et al. promoter-capture Hi-C dataset and the Cardiogenic25, 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 (ABCA25) 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 SP11 locus, we identified AD risk enhancers shared by ACP2, MADD, MYBPC3, NR1H3, NUP160, PSMC3, and TP53INP1 in monocytes, and by NUP160, MYBPC3, and SP11 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 SP11 at the respective (previously CELFI) locus. Additionally, these analyses revealed regulatory landscapes that are shared across myeloid cell types or are cell type-specific. In the BIN1 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 BIN1 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.
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 variants that mediate these relationships to modulate disease susceptibility. 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 color29 to select those with evidence of independent or colocalized AD GWAS and hQTL signals (PP. H3.abf + PP.H4.abf ≥ 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 color 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 SPL2A in monocytes (Fig. 3c). In turn, expression of SPL2A is causally associated with AD risk. Furthermore, this chromatin region interacts with the promoter of SPL2A, providing converging evidence for regulation of SPL2A 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.

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 |
|------------|------------------------------------------------|-----------------------------|
| BIN1       | BNI                                            | BNI                         |
| SPI1       | ACP2, FNB4, MADD, MYBPC3, MTC2, NRIH3, NUP160, PSMC3, SPI1 | CEF1, MTC2, MYBPC3, NUP160, PSMC3, SPI1 |
| ZYX        | ZYX                                            | ZYX                         |
| MS4A       | MS4A6A, MS4A6A*                                | AP4M1, GATS, PILRA, PILRB, ZKSCAN1, ZCWPW1 |
| PILRA      | AP4M1, GATS, PILRA, PILRB, ZKSCAN1, ZCWPW1     | AP4M1, GATS, LAMTOR4, MCM7, PILRA, PILRB, PVRIG, STAG3, TRIM4, ZKSCAN1 |
| TP53INP1   | CCNE2, INTS8, TP53INP1                         | INTS8, NDUFAF6, TP53INP1    |
| NDUFAF6    |                                                |                              |
| SPL2A      |                                                |                              |
| RIN3       | RIN3                                           | ABCA7, CEN2, CIRBP          |
| ABCA7      | ABCA7, CEN2, CIRBP                            |                              |
| APBB3      | APBB3, HPBF1                                   |                              |
| RABEP1     | NUP88, RABEP1                                  |                              |
| PTK2B      | PTK2B                                          |                              |
| CDSS       | CDSS                                           |                              |
| CASS4      | CASS4                                          |                              |
| TREM2      | TREM2                                          |                              |
| PICALM     |                                                |                              |

*Not expressed in microglia.

| Locus      | Monocytes                                      | Monocyte-derived macrophages |
|------------|------------------------------------------------|-----------------------------|
| BIN1       | BNI                                            | BNI                         |
| SPI1       | ACP2, FNB4, MADD, MYBPC3, MTC2, NRIH3, NUP160, PSMC3, SPI1 | CEF1, MTC2, MYBPC3, NUP160, PSMC3, SPI1 |
| ZYX        | ZYX                                            | ZYX                         |
| MS4A       | MS4A6A, MS4A6A*                                | AP4M1, GATS, PILRA, PILRB, ZKSCAN1, ZCWPW1 |
| PILRA      | AP4M1, GATS, PILRA, PILRB, ZKSCAN1, ZCWPW1     | AP4M1, GATS, LAMTOR4, MCM7, PILRA, PILRB, PVRIG, STAG3, TRIM4, ZKSCAN1 |
| TP53INP1   | CCNE2, INTS8, TP53INP1                         | INTS8, NDUFAF6, TP53INP1    |
| NDUFAF6    |                                                |                              |
| SPL2A      |                                                |                              |
| RIN3       | RIN3                                           | ABCA7, CEN2, CIRBP          |
| ABCA7      | ABCA7, CEN2, CIRBP                            |                              |
| APBB3      | APBB3, HPBF1                                   |                              |
| RABEP1     | NUP88, RABEP1                                  |                              |
| PTK2B      | PTK2B                                          |                              |
| CDSS       | CDSS                                           |                              |
| CASS4      | CASS4                                          |                              |
| TREM2      | TREM2                                          |                              |
| PICALM     |                                                |                              |

*Not expressed in microglia.
first selected variants in high to moderate LD (R² ≥ 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 (R² = 0.86, Alzheimer’s Disease Genetics Consortium case-control cohort (ADGC) reference panel was used to compute LD as described previously)12 and was previously shown to alter the ligand binding affinity of PILRA31. 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 annotations15,34–38, 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
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 algorithm\(^40\), conditional analyses\(^41\) and motif disruption/creation analyses\(^{39}\) 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 microglia\(^42\), 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

---

**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 |
|-------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| BIN1  | ACP2, CELF1, DDB2, MADD, MYBPC3, NRP1H3, NUP160, PSMC3                     | BIN1, CELF1, MADD, MYBPC3, NUP160, SPI1                                             |
| SPI1  | ZYX, GPR141                                                               | ZYX, GPR141                                                                          |
| CD2AP | CD2AP                                                                     | CD2AP                                                                               |
| ZYX   | (previously EPHA1)                                                       | ZYX                                                                                  |
| GPR141| (previously NME8)                                                        | GPR141                                                                               |
| MPS4A | INTS8, TP53INP1                                                           | INTS8, TP53INP1                                                                      |
| MPS4A | MS4A4A, MS4A6A                                                            | MS4A4A, MS4A6A                                                                       |
| RABEP1| NUP88, RABEP1, SCIMP                                                     | NUP88, RABEP1, SCIMP                                                                  |
| EF2B2 | ACYPI, EIF2B2, MLH3                                                      | EF2B2                                                                                |
| POLR2 | POLR2, RPP2                                                               | POLR2                                                                                |
| PILRA | AP4MI, TRIM4, PILRA, PILRB, ZKSCAN1                                      | PILRA, AP4MI, TRIM4, PILRA, PILRB, ZKSCAN1                                           |
| ZCWPW1| AP4MI, TRIM4, PILRA, PILRB, ZKSCAN1                                      | ZCWPW1                                                                               |
| SPPL2A| AP4MI, SPPL2A                                                             | SPPL2A                                                                               |

*Association reported with a distal enhancer that does not contain AD risk variants, but whose hQTLs colocalize with AD risk alleles.

*Not expressed in microglia.
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 monocytes. One of the prioritized candidate causal variants in the MS4A locus, the rs636317-T AD risk-increasing allele (11:60019150:CT 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 expression. 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 activity. 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 GM12878 has a 97.6% probability of colocalization with 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).

(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. 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 prediction. 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 epigenome. 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 expression. 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).

Another 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 monocytes. One of the prioritized candidate causal variants in the MS4A locus, the rs636317-T AD risk-increasing allele (11:60019150:CT 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 expression. 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 activity. 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 GM12878 has a 97.6% probability of colocalization with 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).

(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. 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 prediction. 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 epigenome. 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 expression. 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).
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.

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 and our own earlier study showed an
enrichment of AD SNP heritability in myeloid-specific epigenomic annotations including the PU.1 cistrome. 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, 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 (BIN1, MS4A4A, SPI1) 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, MS4A6A, SPI1, as well as novel candidate causal genes AP4E1, AP4M1, PILRA, RABEP1, SPP12A, 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. BIN1, MS4A, SPI1, RABEP1, TP53INP1, and ZYX). Remarkably, when a BIN1 enhancer prioritized through our approaches was deleted in hiPSC-derived microglia, neurons and astrocytes, BIN1 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 findings.

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 macrophages and drives their intracellular movement by assembling actin filaments and cargo fission and fusion. BIN1 (Ras And Rab Interactor 3) encodes a member of the RIN family of RAS and RAB effectors that interacts and localizes with BIN1 to early endosomes. Other RIN family members, RIN3 has guanine nucleotide exchange factor (GEF) activity for RAB5 GTPases, which are required for early endosome and phagosome biogenesis and fusion. Interestingly, RABEP1 (Rab-GTPase binding effector protein 1) also encodes a RAB5 effector protein that is required for early endosome membrane fusion and trafficking. 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 endosomes. Interestingly, APPB3 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. Another novel candidate AD risk gene that we nominate in this study, SPP12A, encodes a transmembrane aspartyl protease that localizes to late endosomes and lysosomes and cleaves substrates involved in immunity and neurodegeneration. Finally, TP53INP1 regulates the stability and transcriptional activity of p53, and has been implicated in the phagocytic clearance of apoptotic cells (efferocytosis), 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 brain. Taken together, our findings implicate dysfunction of the endolysosomal system in myeloid cells (as opposed to neurons) 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 AD. Furthermore, coding variants that increase risk for AD have been identified in SORL1, 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 brain and plays important roles in the endolysosomal system and APP processing.

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 (BIN1, MS4A, and ZYX),...
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, SPI1, 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 and that the activation state of human macrophages does not have a major impact on AD heritability enrichment 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 microglia. 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 underscores 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)15,16,35,37. We selected studies that contained H3K4me1 (monocytes and macrophages), H3K4me2 (monocytes and microglia) and/or H3K4me3 (macrophages and microglia) 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 Bowtie2 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 MAQC score and duplicates were removed using samtools. MACS2 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 annotated_peak_pre". Histone modifications ChIP-Seq peaks were called using the following command: "callpeak -t case.sam -c input.sam -f SAM -g hs -q 0.01 -n file_name--outdir output_dir". Pu.1 ChIP-Seq peaks were called using the following command: "callpeak -t file -f BAM --nomodel -shift -37 -extsize 73 -g hs -q 0.01 -n file_name--outdir output_dir".

Statification 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 epigenetic marks around the TSS37,38, 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 ≤ 1 × 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%).

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 putative sequences in these regions: findMotifTopGenome.PEaks.bed hg19. -size given. To identify regions that contained our motifs of interest, we used the following commands: findMotifTopGenome.PEaks.bed hg19 -find motif -motif -size given and annotatePeaks.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 parameters39. We used coloc in the following manner: coloc(abf,dataset1,dataset2, p = 1e-04, p2 = 1e-04, p1 = 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 datasets41. 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.

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. After controlling for the 53 functional annotation categories of the full baseline model, GWAS summary statistics for AD42 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 values of enrichment were reported, the p-values for annotations that had negative enrichments were not displayed on the figures.

Causal interactions analysis. We used the following command: "smr --bfile reference_file --bqtl-summary Exposure_besd_file_prefix --bqtl-summary Outcome_besd_file_prefix --out-output_prefix". The results were filtered for FDR of 5% calculated using the padjust function from R. To conduct SMR for targeted SMR analyses to make sure that they do indeed tag the majority of the GWAS signal in the annotation enrichments the following command was used: "smr --bfile reference_file --gwas-summary gwas_summary_file --bqtl-summary eQTL_besd_file_prefix --target-snp rs12345 --out-output_prefix".

Conditional and haplotype analyses. We used GCTA-COJO44 to conduct conditional analyses using iGAP GWAS summary statistics data and ADGC GWAS cohort individual-level genotype data using plink45. To conduct standard SMR analysis, we ran the following command: "gcta4 --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-LD46 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 functional alternate allele information for each SNP. We then ran the CLQ algorithm provided within the Big-LD package for SNP clustering and BigLD for haplotype block construction. We used the following command: "CLQD(geno = genotype_data, SNPinfo = locus.snp_information, hTtype = "fast", CLQucut = 0.5) and Big_LD(genotype_data, SNPinfo = locus.snp_information, startBp = start_basepair, endBp = end_basepair, appendRare = TRUE)."

Prioritization of candidate causal variants. For each locus we constructed LD blocks using Big-LD package. We then selected variants that reside in active enhancers in monocytes, macrophages and/or microglia (with the exception of the SPPI2A 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)35. 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, ZFY and BIN1 loci. PAINTOR is a Bayesian fine-mapping method that leverages functional annotations through an Empirical Bayes prior31. 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 -Lname ld -enumerate max number of causal_ variants -name name indir -dir dir outdir. The 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 binding38. We used HOCO motifbreakR with the following command: motifbreakR -p file -o output file -t threshold 5 × 10−5. We used the following command to do so: motifbreakRsnpList = variant_list, pwmList = hocomo, filterp = TRUE, threshold = 5e-5.

For conditional analyses using iGAP GWAS summary statistics data and ADGC GWAS cohort individual-level genotype data using plink. To conduct standard SMR analysis, we ran the following command: "gcta4 --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-LD 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 functional alternate allele information for each SNP. We then ran the CLQ algorithm provided within the Big-LD package for SNP clustering and BigLD for haplotype block construction. We used the following command: "CLQD(geno = genotype_data, SNPinfo = locus.snp_information, hTtype = "fast", CLQucut = 0.5) and Big_LD(genotype_data, SNPinfo = locus.snp_information, startBp = start_basepair, endBp = end_basepair, appendRare = TRUE)."
Technologies). DAPI (4 Signaling, 2266) and anti-CX3CR1 (Bio-Rad, AHP1589). Secondary antibodies were for 10 min. Cells were permeabilized with 1.0% Triton in PBS at room temperature to visualize nuclei. Images were acquired using a Leica Fluorescence Microscope at 40× magnification.

**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 overlaying the variant of interest using rmaipup command in samtools. 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 0.1% 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 μg/ml anti-TREM2 (R&D, AF1828), 1:1000 anti-P2RY12 (Sigma, HPA014518), 1:100 anti-PU.1 (Cell Signaling, #4352) and anti-CX3CR1 (Bio-Rad, AB1809). 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 μg/ml) was used to visualize nuclei. Images were acquired using a Leica Fluorescence Microscope at 40× magnification.

**Data availability.** The following studies obtained from GEO were used for the analyses presented in this paper: GSE29611 (monocyte CTCF, H3K27ac and H3K4me1/2/3), GSE85245 (macrophage ATAC-seq), GSE66594 (macrophage H3K4me1, H3K27ac and PU.1) and GSE98365 (macrophage ATAC-seq). Data generated in this study are available through accession numbers.

**References.**

1. Dementia statistics | Alzheimer's Disease International. https://www.alz.co.uk/research/statistics.
2. Elhymyouni, 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 SORLI and Alzheimer disease. Ann. Neurol. 77, 215–227 (2015).
5. Sims, R. et al. Rare coding variants in PLCG2, AB13, 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.
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. Finucane, H. K. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 413–419 (2019).
11. Wang, K.-L. et al. A common haplotype lowers P1U expression in myeloid cells and delays onset of Alzheimer’s disease. Nat. Neurosci. 20, 1052–1061 (2017).
12. Finucane, H. K. et al. Heteritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. 50, 621–629 (2018).
13. Zhu, Z. et al. Integration of summary data from GWAS and eQTLS studies predicts complex trait gene targets. Nat. Genet. 48, 481–487 (2016).
14. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. Science 356, 3222 (2017).
15. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228 (2015).
16. Lambert, J.-C. Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer’s Disease. Nat. Genet. 45, 1452–1458 (2013).
17. 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 Psychiatry 3, 1371–1379 (2016).
18. Javieer, 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).
19. Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 159, 1312–1326 (2014).
20. Gosselin, D. et al. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. Cell 159, 1327–1340 (2014).
21. Garnier, S. et al. Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes. PloS Genet. 9, e1003240 (2013).
22. 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.
23. 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 (ABC7), apolipoprotein E ε4, and the risk of late-onset Alzheimer disease in African Americans. JAMA Neurol. 1493–1492 (2013).

26. Chapuis, J. et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. Mol. Psychiatry. 12, 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 Cell 167, 1398–1414.e216 (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. Novakovíc, 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 1 interferons and the cytokine TNF cooperatively mediate the pathogenesis of Alzheimer disease. Dev. Cell 18, 425–436 (2010).

37. Díaz, F. et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer’s disease. Sci. Data 5, 180185 (2018).

38. Munoz, R. et al. Single-cell-derived human microglia transplanted in mouse brain to study human disease. Nat. Neurosci. 22, 2111–2116 (2019).

39. Hasselmann, J. et al. Development of a chimeric model to study and manipulate human microglia in vivo. Neuro. 103, 1016–1033.e10 (2019).

40. Peng, M. et al. The phagostimulatory and phagocytic properties of macrophages are regulated by actin polymerization and microtubule dynamics. Immunity 30, 143–154 (2009).

41. Southwick, F. S., Li, W., Zhang, F., Zeile, W. L. & Purich, D. L. Actin-based endosome and phagosome spinning in macrophages: activation by the secretagogue antagonists lanthanum and zinc. Cell Motil. Cytoskeleton 54, 143–155 (2003).

42. Kajijo, 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).

43. Stenmark, H., Vitále, 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).

44. Burgos, P. V. et al. Sorting of the Alzheimer's disease amyloid precursor 1505–1516 (2010). (2017) 12:1610 | https://doi.org/10.1038/s41467-021-21823-y | www.nature.com/naturecommunications

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

Acknowledgements We thank the Cardiogenics (European Project reference LISM-CT-2006-057593) project for providing summary statistics data for QTL analyses in monocytes and monocyte-derived macrophages. We thank the New York Genome Center for sequencing the hiPSC-derived microglia samples.
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.), R01AG054011 (A.M.G.), U01AG058635 (A.M.G.), NIA K01AG062683 (J.TCW.), AG016573 (W.W.P.), F31 AG059337-01 (A.G.E.), RO1AG050986 (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.M. 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.