

MS4A4A modifies the risk of Alzheimer disease by regulating lipid metabolism and immune response in a unique microglia state

Shih-Feng You¹, Logan Brase^{1#}, Fabia Filipello^{1#}, Abhirami K. Iyer^{1#}, Jorge Del-Aguila¹, June He¹, Ricardo D'Oliveira Albanus¹, John Budde¹, Joanne Norton¹, Jen Gentsch¹, Nina M. Dräger², Sydney M. Sattler², Martin Kampmann², Laura Piccio^{1,3}, John C. Morris^{4,5}, Richard J. Perrin^{4,5,6}, Eric McDade⁴, Dominantly Inherited Alzheimer Network, Steven M. Paul¹, Anil G. Cashikar¹, Bruno A. Benitez⁷, Oscar Harari^{1,5*}, Celeste M. Karch^{1,5,8*}

¹Department of Psychiatry, Washington University in St. Louis School of Medicine, USA

²Institute for Neurodegenerative Diseases, Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA

³Charles Perkins Centre and Brain and Mind Centre, School of Medical Sciences, University of Sydney, Sydney, NSW, Australia

⁴Department of Neurology, Washington University in St. Louis School of Medicine, USA

⁵The Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, Missouri, USA

⁶Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri, USA.

⁷Department of Neurology, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

⁸NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, Missouri, USA.

^{*}, Authors contributed equally

[#], Authors contributed equally

Corresponding Author:

Celeste M. Karch, Department of Psychiatry, Washington University in St Louis, 425 S. Euclid Ave, Campus Box 8134, St Louis, MO, 63110, phone: 314-747-3161, fax: 314-747-2983, email: karchc@wustl.edu

Abstract

Genome-wide association studies (GWAS) have identified many modifiers of Alzheimer disease (AD) risk enriched in microglia. Two of these modifiers are common variants in the *MS4A* locus (rs1582763: protective and rs6591561: risk) and serve as major regulators of CSF sTREM2 levels. To understand their functional impact on AD, we used single nucleus transcriptomics to profile brains from carriers of these variants. We discovered a “chemokine” microglial subpopulation that is altered in *MS4A* variant carriers and for which *MS4A4A* is the major regulator. The protective variant increases *MS4A4A* expression and shifts the chemokine microglia subpopulation to an interferon state, while the risk variant suppresses *MS4A4A* expression and reduces this subpopulation of microglia. Our findings provide a mechanistic explanation for the AD variants in the *MS4A* locus. Further, they pave the way for future mechanistic studies of AD variants and potential therapeutic strategies for enhancing microglia resilience in AD pathogenesis.

Main

Alzheimer disease (AD) is driven by complex genetic architecture that can be leveraged to understand key pathways that may promote or suppress disease processes. Genome-wide association studies (GWAS) have revealed more than 70 loci associated with AD risk, many of which harbor genes that are enriched in microglia¹⁻⁵. Microglia, the innate immune cells of the brain, produce inflammatory signals in response to environmental stressors. Dynamic changes in microglia function have been reported during AD progression⁶; thus, understanding the factors that influence these states is critical to therapeutically harnessing microglia function.

The microglia-enriched *MS4A* gene family encodes a class of tetraspanin proteins that have been implicated in AD^{1,2,7}. Common variants in the *MS4A* locus are associated with AD susceptibility and regulation of soluble TREM2 (sTREM2) levels in cerebrospinal fluid (CSF)^{1,2,8-10}. Two independent signals in the *MS4A* locus are associated with altered CSF sTREM2 levels^{8,10}. The most significant signal occurs in an intergenic region near *MS4A4A* (rs1582763), which is associated with elevated CSF sTREM2 levels as well as reduced AD risk^{2,11} and delayed age-at-onset (e.g., protective)¹². A second, independent signal occurs in the coding region of *MS4A4A* (rs6591561; p.M159V) and is associated with lower CSF sTREM2 levels as well as increased AD risk^{2,11} and accelerated age-at-onset¹². However, the impact of these AD-modifying variants on microglia function and AD pathophysiology is unknown.

To date, with a few exceptions^{5,13-16}, the functional consequences of variants identified in AD risk loci are unknown. We now build upon emerging single nucleus RNA sequencing (snRNA-seq) technologies to look beyond disease status and cell-type proportion to establish a blueprint to pinpoint the function of common variants that drive neurodegenerative disease.

Here, we combined genomic and snRNA-seq technologies to identify the function of AD variants in the *MS4A* gene locus. We discovered that variants in the *MS4A* locus regulate a subpopulation of microglia that are enriched in chemokine function. The AD protective variant drives elevated *MS4A4A* expression and shifts the chemokine microglia subpopulation towards an interferon state, while the AD risk variant suppresses *MS4A4A* expression and the chemokine microglia subpopulation. Together, variants in the *MS4A* locus illustrate the power of microglia to promote or suppress AD pathophysiology.

Results

MS4A4A protective variant rs1582763 reduces the inflammatory response and enhances lipid metabolism pathways

To understand the impact of the AD protective variant in the *MS4A* locus (rs1582763), we evaluated gene expression in rs1582763 variant carrier brains (GG, GA, AA; Figure 1A-B). Meta-analysis of trans-eQTLs was performed using three independent AD and control brain cohorts (n=579 brains from the Knight ADRC, Mayo RNAseq Study (Mayo), and the Religious Orders Study and Memory and Aging Project (ROSMAP); Figure 1B and 2A; Supplemental

Table 1). After correcting for age, sex, and AD status, we identified 1,071 differentially expressed protein-coding genes (FDR<0.05; Figure 2A; Supplemental Table 2). These genes were enriched in pathways associated with reduced pro-inflammatory responses (NF- κ B [$p = 6.29 \times 10^{-11}$], Cytokine [$p = 1.12 \times 10^{-10}$], IFN- γ [$p = 4.59 \times 10^{-10}$], Complement [$p = 8.21 \times 10^{-8}$]); increased anti-inflammatory response (Type I-IFN [$p = 1.7 \times 10^{-4}$]); elevated lipid metabolism (cholesterol efflux [$p = 6.54 \times 10^{-5}$]); and metabolism (mTORC1 [$p = 4.05 \times 10^{-3}$]) (Figure 2B; Supplemental Table 3). Several AD risk genes were significantly increased in the presence of the minor allele of rs1582763: *PILRA*, *MS4A14*, *STAG3*, *C7orf61*, *EFHD1*, *CCP110*, *ATXN3*, *ESCO2* (Supplemental Table 4).

Many genes differentially expressed based on rs1582763 genotype were known to be enriched in microglia, including *TREM1*, *CH25H*, *CXCR4*, *HLA* genes, *CD68*, among others. To determine whether these genes were statistically enriched in microglia expression, we analyzed transcriptomic data from Fluorescence-Activated Cell Sorting (FACS)-isolated microglia derived from human brains ($n = 10$; ROSMAP)^{17,18}. Of the 1069 differentially expressed genes, 1067 were detected in the human brain-isolated microglia (Supplemental Table 1; Figure 2C). Together, these results suggest that rs1582763 primarily drives transcriptional changes in microglia.

Parallel trans-eQTL analyses were run within each of the brain cohorts. These more stringent analyses revealed an association between rs1582763 and reduced levels of *CH25H* ($\beta = -0.37$, $p = 1.22 \times 10^{-4}$), *DCDC1* ($\beta = -0.26$, $p = 3.16 \times 10^{-5}$), *PTX3* ($\beta = -0.41$, $p = 5.82 \times 10^{-5}$), and *FOSB* ($\beta = -0.49$, $p = 4.38 \times 10^{-6}$) (Figure 2D; Supplemental Table 5). These genes are involved in cholesterol metabolism and immune response and are highly related to one another by physical interaction, co-expression, and shared protein domains (Figure 2E). *DCDC1* is expressed in microglia¹⁹ and helps regulate microtubule polymerization and dynein-dependent transportation²⁰. *PTX3* is predominately induced by pro-inflammatory cytokines and regulates complement activities in neuroinflammatory conditions such as multiple sclerosis (MS) and experimental autoimmune encephalomyelitis (EAE)²¹⁻²⁴. *FOSB* was found to be expressed in activated microglia and plays a role in complement regulation²⁵. *CH25H* encodes the microglia-specific enzyme, cholesterol 25-hydroxylase, and is involved in many of the key pathways altered by rs1582763. *CH25H* converts cholesterol into 25-hydroxysterol (25HC) which is considered an inflammatory mediator in microglial recruitment that promotes IL1B and NLRP3 inflammasome-mediated pro-inflammatory response²⁶. Given that rs1582763 also results in elevated sTREM2 in human CSF¹⁰, we sought to investigate whether *CH25H* modulates sTREM2 levels. Microglia isolated from *Ch25h* KO mice (postnatal day 2) produced significantly elevated sTrem2 compared to littermate controls ($p = 1.0 \times 10^{-3}$; Figure 2F). Thus, rs1582763, in the *MS4A* locus, regulates microglia-enriched genes, whose expression may impact sTREM2 levels.

The protective variant impacts a specific population of microglia in human brains that regulates cytokines, chemokines, and the Type-I interferon pathway

Microglia are highly dynamic and context-dependent²⁷⁻³⁰. To further investigate the impact of rs1582763 on microglia states, we analyzed a cohort of AD and control brains with genomic and snRNA-seq data ($n = 66$; 39% GG, 42% AG, 19% AA, Figure 1C; Figure 3A; Supplemental Table 1). Microglia were identified and annotated by cell-type specific

markers ($n=17,089$)³¹ and grouped into nine distinct clusters (Mic.0-Mic.8) by performing unbiased clustering as previously described^{32,33}. We then compared the microglial states associated with the minor and major alleles of rs1582763. The minor allele of rs1582763 was associated with a significant increase in the proportion of nuclei within the Mic.3 cluster ($p=1.65\times 10^{-3}$; Figure 3B; Supplemental Figure 1; Supplemental Table 6). Within Mic.3, rs1582763 was associated with *MS4A4A* expression, whereby minor allele carriers exhibit significantly increased *MS4A4A* levels ($p=1.11\times 10^{-2}$; Kruskal-Wallis test; Figure 3C). *MS4A4A* and *TREM2* expression was concentrated in Mic.3, while *MS4A6A* and *MS4A4E* expression patterns differed (Supplemental Figure 2). We observed the highest number of differentially expressed genes in Mic.3 (107 genes, $FDR<0.05$; Figure 3D). Thus, the primary molecular effect of rs1582763 occurred within a specific microglia population: Mic.3.

Given the impact of rs1582763 on gene expression specifically in Mic.3, we sought to define the genes that were changed as a function of rs1582763 within this subpopulation of microglia. We identified 259 genes that were significantly differentially expressed within Mic.3 according to rs1582763 genotype (Figure 3E). Many of these genes influence pathways related to immune function (TYROBP [$p=2.1\times 10^{-5}$], Chemokine [$p=8.39\times 10^{-3}$], TNF- α [$p=3.65\times 10^{-3}$], NLRP3 inflammasome [4.54×10^{-2}]) and are associated with an increased Type-I interferon response (IFN- β [$p=1.11\times 10^{-4}$]). Other genes are also enriched in pathways related to enhanced lipid metabolism ($p=2.39\times 10^{-2}$) enhanced protein degradation (mTOR signaling [$p=2.09\times 10^{-2}$], lysosome [$p=9.11\times 10^{-3}$]), and decreased calcium regulation ($p=4.31\times 10^{-3}$) (Figure 3F; Supplemental Table 8).

Together, we showed that rs1582763 mediates reduced inflammatory gene expression and enhanced cholesterol metabolism gene expression in a specific population of microglia in the human brain. These pathways were also highly related to *TREM2* function, and in primary microglia, targeting one of these genes, *CH25H*, was sufficient to shift s*TREM2* levels in a manner consistent with the protective variant.

AD risk variant MS4A4A rs6591561 promotes inflammasome response and increases intracellular cholesterol storage

To understand the impact of the variant in *MS4A4A* that is associated with increased AD risk (rs6591561), we evaluated gene expression patterns in AD brains among rs6591561 variant carriers (AA, GA, GG; Figure 1A-B). Meta-analysis of three brain cohorts ($n=579$ brains in Knight ADRC, Mayo, and ROSMAP; Figure 4A; Supplemental Table 1) revealed 1597 differentially expressed genes as a function of the rs6591561 genotype ($FDR<0.05$; Supplemental Table 9) including *MS4A4A* ($p=3.8\times 10^{-2}$), which showed reduced expression. These genes were also enriched in sorted microglia from human brains (1593/1597; Figure 4B). Genes altered by rs6591561 are involved in pathways associated with elevated pro-inflammatory responses (NF- κ B [$p=1.04\times 10^{-6}$], cytokine-mediated signaling [$p=2.75\times 10^{-7}$], IFN- γ [$p=1.78\times 10^{-6}$]); reduced Type-I interferon response (IL-4 [$p=4.61\times 10^{-7}$], IL-10 [$p=1.92\times 10^{-3}$], IFN- α/β [$p=1.59\times 10^{-6}$]); and reduced lipid metabolism (cholesterol efflux [$p=5\times 10^{-3}$]) (Figure 4C; Supplemental Table 10). Interestingly, many of these pathways were shared with the protective variant but with opposing effects. Several genes in the AD risk loci, including *MS4A4A*, were significantly reduced among minor allele carriers: *HLA-DRA*,

HLA-DQA1, MS4A4A, HLA-DPB1, CFD, C2, PVRIG, PCOLCE, TREM1, HLA-DMA, MIDN, AQP9, MBLAC1, HLA-DOA, KCNJ13, AZU1, and CYP27C1 (Supplemental Table 11).

Next, we examined the impact of rs6591561 on specific microglia states using snRNA-seq data from human brains. We detected a significant increase in total microglia nuclei among carriers of the minor allele in rs6591561 ($p=1.11\times 10^{-2}$; Figure 4D). Among microglia subclusters, nuclei in Mic.3 were significantly increased in heterozygous carriers of rs6591561 in comparison to those homozygous for the major allele ($p=4.52\times 10^{-2}$; Figure 4E). Consistent with the eQTL between rs6591561 and *MS4A4A* observed in brain homogenates (Figure 4A), we detected a significant eQTL between rs6591561 and *MS4A4A* in Mic.3 (Figure 4F; Supplemental Figure 1). Mic.3 also exhibited the highest number of differentially expressed genes as a function of the rs6591561 genotype (90 genes, $FDR<0.05$; Figure 4G; Supplemental Table 12). Thus, rs6591561 drives molecular changes in a specific microglia state: Mic.3.

Genes differentially expressed in Mic.3 due to rs6591561 were associated with immune response (inflammasomes [$p=3.53\times 10^{-3}$], NF- κ B [$p=3.33\times 10^{-5}$], cytokine [$p=1.17\times 10^{-3}$], IL-2/STAT5 [$p=4.01\times 10^{-3}$]), cholesterol metabolism (lipid storage [$p=5.37\times 10^{-3}$]), protein degradation (Lysosome [$p=4.65\times 10^{-3}$]), and calcium regulation (negative regulation of Ca^{2+} transport [$p=3.10\times 10^{-3}$], Figure 4H-I; Supplemental Table 13). Calcium homeostasis is critical to the function of many of the pathways identified. To test whether *MS4A4A* expression impacts intracellular calcium levels, we silenced *MS4A4A* with RNAi in THP-1 myeloid lineage cells. Calcium levels were monitored by Fluo-4 AM calcium indicator (Figure 4J). Consistent with the snRNA-seq findings, silencing *MS4A4A* expression produced significantly higher intracellular Ca^{2+} levels at baseline compared with scrambled controls ($p=0.01$; Figure 4L). Additionally, increased release of intracellular Ca^{2+} stores with ionomycin led to a significant increase in Ca^{2+} levels when *MS4A4A* was silenced (Figure 4M). In the presence of a calcium effector, CXCL12 (SDF-1 α , a ligand for CXCR4), silencing *MS4A4A* led to elevated intracellular Ca^{2+} kinetics compared to scrambled-treated cells (Figure 4N). These data suggest a role for *MS4A4A* in regulation of calcium homeostasis in microglia. Together, we find that rs6591561 induced molecular and functional effects in microglia in human brains.

Microglia cluster 3 is a chemokine/cytokine-specific cluster

Microglia exhibit tremendous transcriptional diversity in human brains, which may reflect dynamic changes due to aging, environment, and disease states³⁴. Microglia subclusters were defined using unsupervised clustering³². Transcriptionally, Mic.0 was enriched in genes associated with homeostatic states (*SELPLG, MED12L, BIN1, CX3CR1, P2RY13, TMEM119*), while Mic.1 was enriched in genes found in activated and disease-associated microglia states (*CD83, TNFAIP3, CSAR1, CD68, GPNMB, ABCA1*). However, the identity of Mic.3, which was impacted by both protective and risk variants, was unclear.

To begin to understand the identity of this microglia state, we compared gene expression in Mic.3 to other the microglia subclusters in the brain ($p < 0.05$; Figure 1D; Figure 5A). We identified more than 150 genes within Mic.3 that were significantly upregulated, including several AD risk genes: *BIN1* ($p=8.58\times 10^{-12}$), *HLA-DRB1* ($p=1.27\times 10^{-2}$), *SLC12A9* ($p=2.67\times 10^{-4}$)

(Figure 5A-5B; Supplemental Table 14). Mic.3-specific genes are enriched in pathways associated with elevated immune response, such as cytokine and chemokine pathways (IL-2 [$p=1.18 \times 10^{-12}$], NF- κ B [$p=1.4 \times 10^{-2}$], CXCR4 [$p=1.16 \times 10^{-2}$], IFN- γ [$p=8.85 \times 10^{-6}$]), and other pathways like TLR [$p=1.33 \times 10^{-6}$], TNF- α [$p=8.64 \times 10^{-6}$], and TGF- β [$p=1.33 \times 10^{-2}$] (Figure 5B; Supplemental Table 15). Additionally, while *CH25H* was significantly elevated, other lipid metabolism genes were reduced ($p=3.28 \times 10^{-2}$) (Figure 5B; Supplemental Table 15).

Due to the pleiotropic nature of cytokines and chemokines, the expression patterns of these genes and pathways alone were not sufficient to clarify the identity of Mic.3. To refine the identity of Mic.3, we leveraged a recently described dataset from human microglia-like cells (induced transcription factor (iTF)-microglia) as an *in vitro* reference. iTF-microglia were transformed into a spectrum of transcriptionally diverse microglia via CRISPRi-based editing³⁵. The top 50 differentially expressed genes that define Mic.3 were set as a gene module and visualized by calculating the average expression score of the iTF-microglia dataset on a single-cell level. Genes defining Mic.3 were enriched in iTF-microglia cluster 8 and cluster 9, which were associated with chemokine and C-C motif chemokine ligand 13 (CCL13+) chemokine functions, respectively (Figure 5C).

Mic.3 contains anti-inflammatory microglia that are regulated by MS4A4A expression

Microglia exhibit multiple activation states in response to distinct contexts. Microglia transition to a TREM2-dependent disease-associated microglia (DAM) activation state that can be further sub-classified into pro-inflammatory DAM (*CD44+*) and anti-inflammatory DAM (*CXCR4+*) states³⁶ (Figure 5D). Mic.3 shares many activation markers with Mic.1, suggesting Mic.1 and Mic.3 may both represent sub-types of activated microglia in the human brain³².

To investigate this possibility, we evaluated whether homeostatic microglia (Mic.0) were equally likely to transition to the populations of activated microglia (Mic.1 and Mic.3). We employed pseudo-time analytical frameworks, including Monocle3³⁷ and velocyto³⁸, to predict the transcriptomic trajectories across Mic.0, Mic.1, and Mic.3 (Figure 5E). We found that Mic.0 points outward, then diverges in opposite directions toward both Mic.1 and Mic.3 (Supplemental Figure 3). Thus, we hypothesized that Mic.1 and Mic.3 represent pro-inflammatory (*CD44+*) and anti-inflammatory (*CXCR4+*) DAM states. High-dimensional geometry diffusion analysis was performed on *CD44/CXCR4* expression levels in each cluster using Markov Affinity-based Graph Imputation of Cells (MAGIC)³⁹. Mic.0 expressed *CD44* and *CXCR4* at low levels, consistent with a homeostatic state. Mic.1 exhibited high *CD44* and *CXCR4*, suggesting equal proportions of pro- and anti-inflammatory DAMs. Interestingly, Mic.3 nuclei were high in *CXCR4*, while *CD44* remained low, consistent with an anti-inflammatory-enriched pool (Figure 5F; Supplemental Figure 4). *CXCR4+* anti-inflammatory DAM has been reported to exhibit enhanced phagocytosis, lipid metabolism, and cholesterol efflux³⁶.

Many of these features are also found in rs1582763 carriers, leading us to hypothesize that *MS4A4A* expression regulates microglial chemokine/cytokine function and promotes an anti-inflammatory DAM cell state. To test this hypothesis, human peripheral blood mononuclear cells (PBMCs) were transduced with Lentiviral particles containing *MS4A4A* or

GFP, mimicking the effect of rs1582763 on *MS4A4A*. Overexpression of *MS4A4A* led to a reduction in chemokine, inflammasome, and lysosome genes (Figure 5G) along with an increase in sTREM2 levels (Figure 5H). This result was consistent with the observation that rs1582763 carriers have higher sTREM2 levels in CSF¹⁰ and reduced expression of genes associated with chemokine, inflammasome, and lysosome pathways in human brain tissue samples (Figure 3F).

***MS4A4A* protective variant rs1582763 modulates microglial behavior by inhibiting chemokine response and promoting interferon response**

It was surprising to identify a single microglia state that was altered by variants with opposing effects on AD risk in the *MS4A* gene locus. To test how protective and risk variants impact the microglia population, Mic.3, we leveraged the CROP-seq data from iTF-microglia, which produced a spectrum of microglial phenotypes from “interferon” to “chemokine” by silencing genes with a druggable sgRNA library³⁵. Interestingly, *MS4A4A* expression was enriched in the interferon cluster (Figure 6A), and inversely correlated with the expression of the genes enriched in Mic.3, which mapped to the chemokine clusters in the iTF-microglia (Figure 6A).

To determine how the protective variant impacts microglia identity, we mapped the genes that were associated with rs1582763 genotype in Mic.3 on the iTF-microglia (Figure 6B). We found that the upregulated genes, including *MS4A4A*, were enriched in iTF-microglia clusters associated with interferon signaling (Figure 6B), while the downregulated genes were localized to the chemokine clusters. Thus, the AD-protective variant may regulate Mic.3 by promoting interferon function and inhibiting chemokine function.

Interestingly, genes that were dysregulated as a function of the rs6591561 genotype in Mic.3 were mapped to the chemokine cluster in the iTF-microglia (Figure 6C). The mapping of rs6591561-associated genes suggested an overall repression of the chemokine clusters that define Mic.3 in conjunction with a failure to promote the interferon function (Figure 6C).

Given the evidence of the regulatory potential of rs1582763 and rs6591561, we sought to determine whether genes enriched in Mic.3 are targets of specific transcription factors using ChEA3, a web-based transcription factor enrichment analysis application⁴⁰. Mic.3 genes were regulated by transcription factors enriched for microglial (*SPI1*, *MEF2C*), chemokine (*NFKB1*, *NFKB2*), interferon (*STAT1*, *STAT3*, *STAT6*, *IRF7*), and lipid metabolism (*SREBF1*, *SREBF2*). The AD protective variant, rs1582763, altered genes associated transcription factors that drive the upregulation of interferon (*IRF7* and *STAT6*), downregulation of chemokine (*NFKB1*), and downregulation of cholesterol synthesis (*SREBF2*) (Figure 6B, right). Conversely, the AD risk variant, rs6591561, was associated the upregulation of microglial transcription factors (*SPI1*) and downregulation of lipid metabolism transcription factors (*SREBF1*) (Figure 6C, right). *SPI1* has been reported to be an AD risk factor^{1,41,42}.

Leveraging FDA-approved drugs to promote protective effects of *MS4A4A*

To better understand the impact of the *MS4A4A* variants and leverage their therapeutic potential, we sought to identify the compounds that mimic the gene expression changes observed in the protective and risk variants. To achieve this, we used Connectivity Map (CMap)^{43,44} to search for drug candidates with effects that mimic the molecular signatures associated with each variant. The CMap database contains over one million gene expression signatures from multiple cell lines treated with more than 80K perturbagens. We focused on findings from hemopoietic cell lines (BJAB, CD34, HAP1, HBL1, HL60, JURKAT, K562, KMS34, MINO, NALM6, NOMO1, OCILY10, OCILY19, OCILY3, THP1, TMD8, U937) treated with compound perturbagens. We identified 46 compounds ($\text{norm_cs} > 1.35$) that uniquely mimic the protective variant gene signature (Figure 6D). These compounds were enriched in HDAC inhibitors (11%). We also identified 18 compounds that reverse the gene signature created by the risk variant. These compounds were enriched in KIT inhibitors (9%) and PDGFR inhibitors (9%) (Figure 6D). In parallel, we identified 14 drug candidates that reverse the molecular signatures produced by the protective variant and 75 drug candidates that mimic the gene signatures produced by the risk variant. (Supplemental Figure 5). Thus, we identified compounds that promote the protective effects of rs1582763 and counteract the detrimental effects of rs6591561 for consideration in therapeutic design.

Discussion

Variants in the *MS4A4A* locus are associated with age-at-onset of AD and AD risk; however, the role of *MS4A4A* in AD pathogenesis remains poorly understood¹⁻⁵. Our study shows that the AD protective variant drives elevated *MS4A4A* expression and shifts the chemokine microglia subpopulation to a more interferon state. In contrast, the AD risk variant suppresses *MS4A4A* expression and diminishes the chemokine microglia subpopulation with an inability to promote the interferon state (Figure 1E). We also show that we can leverage these findings for their therapeutic potential by predicting drugs that recapitulate the protective variant molecular signatures or reverse the risk variant molecular signatures.

Chemokines, cytokines, and their receptors modulate cellular events such as chemotaxis, phagocytosis, cytokine secretion, cell activation, proliferation, survival, and death. As a result, the dysregulation of chemokines or cytokines play a critical role in neuroinflammation and neurodegeneration. Under steady-state conditions, homeostatic chemokines and cytokines are involved in immune surveillance and modulation of immune cell differentiation during hematopoiesis. Conversely, proinflammatory chemokines and cytokines activate and recruit microglia and release more inflammatory mediators in the disease state. Occasionally, persistently activated microglia lead to chronic neuroinflammation, which may exacerbate AD. *MS4A4A* is reportedly a surface marker for the anti-inflammatory M2 macrophage⁴⁵. Here, we show that rs1582763 increases *MS4A4A* expression and decreases the inflammatory cytokine and chemokine responses in both bulk and snRNA-seq. Specifically, pro-inflammatory chemokines like CCL3 and CCL4 decrease in bulk and snRNA-seq, while anti-inflammatory IL4 receptors are upregulated. Furthermore, carrying the minor allele of rs1582763 directly correlates with the emergence of a specific microglia cluster (Mic.3), which is a *MS4A4A*-dominant cluster that persists in an anti-inflammatory DAM state (CXCR4 high/CD44 low)³⁶. Conversely, rs6591561 results in decreased *MS4A4A* expression and elevated inflammasome and inflammatory responses.

These results implicate a direct link between the increased *MS4A4A* expression and the decreased inflammatory response in microglia.

Cholesterol dyshomeostasis is a major risk factor for AD⁴⁶. Beyond *APOE4*⁴⁷⁻⁴⁹, growing evidence supports the role of intracellular cholesterol metabolism in microgliosis. Specifically, the accumulation of intracellular cholesterol esters (CE) impedes physiological immune surveillance and scavenger function, promotes chronic inflammatory activation, and ultimately leads to inflammasome activation in microglia⁵⁰. TREM2 also plays a critical role in cholesterol metabolism. TREM2 is required for DAM activation²⁸. TREM2 shifts the energy state by promoting glycolysis and mTOR signaling^{51,52}. TREM2 also activates the LXR transcriptional factor and initiates ABC transporters to promote cholesterol efflux, thus preventing lysosomal dysfunction⁵³. Our data suggest that *MS4A4A* variants rs1582763 and rs6591561 regulate lipid metabolism, such that rs1582763 promotes ABC transporter expression and cholesterol efflux, while rs6591561 downregulates the ABC transporters and promotes cholesterol storage. The AD protective variant, rs1582763, upregulates *LRP* (ApoE, Tau, and A β receptor), *PLTP* (converting VLDL to HDL), and *SORT1* (VLDL transportation). Alternatively, the AD risk variant, rs6591561, upregulates *PCSK9* (LDLR degradation) and *SOAT* (promoting CE storage). CSF PCSK9 is elevated in AD patients, and PCSK9 expression is associated with AD pathogenesis^{54,55}. A PCSK9 inhibitor reduces neuronal inflammation and A β aggregation in rodents⁵⁶. These findings suggest *MS4A4A* expression stymies AD progression by sustaining cholesterol homeostasis in the brain. *MS4A4A* and TREM2 co-localize in lipid rafts⁵⁷. This result, to our knowledge, is the first evidence that connects *MS4A4A* expression with cholesterol metabolism in microglia, suggesting that *MS4A4A* facilitates cholesterol efflux in microglia.

CH25H was found to be significantly downregulated among protective variant carriers (rs1582763), suggesting a linkage between elevated *MS4A4A* expression and reduced *CH25H* expression. Furthermore, cultured microglia with reduced *Ch25h* expression shed more sTrem2 into culture media. The product of *CH25H*, 25-hydroxycholesterol (25-HC), is one of the major microglia-secreted inflammatory mediators in the brain⁵⁸⁻⁶¹. Previous studies have shown that *CH25H* plays a pivotal role in immune response and cholesterol metabolism. For instance, *CH25H* is one of the downstream members of TREM2-mediated lipid metabolism⁵³ and takes part in cholesterol biosynthesis via the SREBP signaling pathway⁶². The product of *Ch25h* enzyme activity, 25-hydrocholesterol (25-HC), is a well-studied suppressor of cholesterol biosynthesis via the SREBP2 pathway^{63,64}. Both anti-inflammatory^{65,66} as well as proinflammatory^{26,59} activities have been reported for 25-HC. In the CNS, 25-HC produced by activated microglia appears to have proinflammatory effects^{67,68}. Our data show that decreased *Ch25h* levels leads to increased sTrem2 levels and an interferon response phenotype of microglia that may be beneficial. These results point to *CH25H* as a mediator by which *MS4A4A* regulates immune response and cholesterol metabolism in AD pathophysiology.

The transcriptional effects of *MS4A4A* variants, rs1582763 and rs6591561, were divergent; however, their effects were localized to the same microglia subpopulation. Microglia subcluster 3 (Mic.3) was altered specifically by both variants. This cluster may be highly relevant to AD pathophysiology as several top upregulated genes, such as *BIN1*, *HLA-DRB1*, and *SLC12A9*, were positively correlated with AD susceptibility⁴. Mic.3 likely plays a role in

chemokine/chemokine CCL13+ function and is functionally distinct from the homeostatic and activated microglia that also exist in the AD brains³³. Mic.3 likely acts as an alternative activation state that is enriched in Type-I interferon function.

In summary, we reported bidirectional biological effects among two variants located in the *MS4A4A* locus. The AD protective variant, rs1582763, featured higher CSF sTREM2 and lower AD risk, was found to have increased *MS4A4A* expression, lower chemokine and cytokine immune responses, and higher cholesterol efflux (Figure 1E). The AD risk variant, rs6591561, featured lower CSF sTREM2 and higher AD risk, was found to have decreased *MS4A4A* expression, and greater expression of genes associated higher inflammatory cytokines, chemokines, and the inflammasome responses, as well as higher cholesterol storage (Figure 1E). Combining the genotypic and brain transcriptomic data from large AD and control cohorts, our study provides a possible mechanistic explanation for how variants in *MS4A* loci alter the AD risk and the underlying causes for how *MS4A4A* modifies immune response and cholesterol metabolism in AD pathophysiology.

Online Methods

Bulk and single nucleus RNA-seq datasets

Bulk RNA-seq datasets were obtained from ROSMAP^{69,70}, Mayo⁷¹, and DIAN/Knight ADRC⁷². The ROSMAP (Synapse ID = Syn3219045) and Mayo (Synapse ID = Syn5550404) datasets were downloaded from AMP-AD Knowledge Portal (<https://adknowledgeportal.org>). Briefly, 579 samples from neuropathologically confirmed AD (N = 400) and neuropathology-free controls (N = 179) were selected for differential expression and composition analyses. The cohort sample size, brain region, average age, sex percentage, allele frequency for the risk and protective variants, and disease status are summarized in Supplemental Table 1.

Bulk RNA-seq datasets for FACS-isolated microglia from ROSMAP were downloaded from AMP-AD Knowledge Portal and processed as previously described³². Ten samples, AD (N = 4) and healthy control (N = 6), were included in the analysis. Microglia were collected and sequenced. The sample size, brain region, age, sex, postmortem interval (PMI) and disease status are summarized in Supplemental Table 1.

The snRNA-seq dataset was obtained from Knight ADRC and processed as previously described³². Briefly, snRNA-seq of microglia from 66 samples with neuropathologically confirmed AD or neuropathology-free controls were used in the differential expression and compositional analyses. The sample size, brain region, age, sex, PMI, allele frequency for the risk and protective variants, and the disease status description are summarized in Supplemental Table 1.

Genetic data obtained from each study was used to determine the genotype of the risk and protective variants. The meta-data acquired from each study, for example, age, sex, PMI, RIN, TIN, and diagnosis/symptom, were used as covariates to define differential gene expression and identify compositional changes.

Data analyses

Functional analysis of bulk RNA-seq

Primary Differential Expression Gene (DEG) discovery in ROSMAP, Mayo, and Knight ADRC were identified using *DESeq2*⁷³. To find DEGs for AD risk and protective variants, the age, sex, and disease status of the cohorts were used as covariates, and an additive model in the Likelihood Ratio Test (LRT) method was used for the variants. The overlap of DEGs (threshold was set at p -value <0.05) between ROSMAP, Mayo, and Knight ADRC were cross-compared and shown in a Venn Diagram. The p -value and Log2 Fold Change (Log2FC) listed in the table were obtained from *DESeq2*. Four DEGs with a common direction across three datasets were put into GeneMania⁷⁴ to define associated genes. The following formula design was used in *DESeq2*:

$$\text{Expression} \sim \text{age} + \text{sex} + \text{AD status} + \text{risk/protective SNP}$$

Meta-analysis was applied to combine the three independent brain cohorts using *MetaVolcanoR*. The Random Effects Model (REM) *rem_mv* was used to integrate p -values from *DESeq2*. Integrated p -values for candidate genes were ranked in ascending order, and the threshold was set at 5%. The top 5% ranked genes were identified as significant DEGs and plotted in a volcano plot.

DEGs identified by the meta-analysis were explored in FACS-isolated microglia from human brains. The transcripts per million (TPM) of sequenced genes from 10 brain samples were calculated and averaged. The average TPM for queried DEGs were summarized. Genes with an average TPM value greater than zero were considered expressed. The main figures show protein coding DEGs.

Pathway analysis was applied to DEGs using *EnrichR*⁷⁵. Pseudogenes and non-coding transcripts were identified by ToppGene⁷⁶ and removed before the query. Pathway enrichments were collected from “BioPlanet,” “KEGG human,” and “WikiPathway” gene sets.

The KEGG-based pathway integration and visualization tool *Pathview*⁷⁷ was applied. The DEGs and REM-estimated summary fold-change (randomSummary) derived from the meta-analysis were diagrammed in the pathway ‘Cytokine-cytokine receptor interaction’ and ‘Cholesterol metabolism’ (Supplemental Figure 6).

Functional analysis of single-nucleus RNA-seq

The processed, identified, and clustered single nuclei from microglia were used in this analysis. The detailed procedure can be found in a previous publication³². Briefly, microglia were clustered into nine expression states. For each expression state, the linear mixed model was applied to identify differentially upregulated genes, and those genes were used in a pathway analysis. Each expression state was associated with 12 previously described microglia functional states by hypergeometric test, to identify feature cell states like ‘homeostatic’ and ‘disease-associated’ microglia clusters.

Each microglia cluster was tested for DEGs associated with the risk or protective variants. The nuclei expressing 80% of all tested genes and the candidate genes expressed in 80% of nuclei were tested in each microglia cluster. The generalized linear mixed model (GLMM) *glmmTMB*⁷⁸ was used to test the DEGs for risk and protective variants, and age, sex, disease status. The alternative risk/protective variant was included among covariates. In each microglia cluster, the numbers of DEGs that passed the *p*-value and BH-FDR<0.05 threshold were recorded and plotted into a histogram. The following formula design was used:

$$\text{Expression} \sim \text{age} + \text{sex} + \text{AD status} + \text{risk SNP} + \text{protective SNP} + (1|\text{subjs}), \\ \text{ziformula}=\sim 1, \text{family}=\text{nbinom2}$$

The DEGs for the risk/protective variant in Microglia cluster 3 (Mic.3) were then plotted in a volcano plot and examined more thoroughly for pathway analyses. Pathway analyses were applied to DEGs with *EnrichR*⁷⁵. DEGs with a *p*-value < 0.05 were analyzed. Pseudogenes and non-coding genes were excluded before analysis. Pathway enrichments were collected from "BioPlanet_2019", "WikiPathway_2021_Human", "KEGG_2021_Human", "GO_Biological_Process_2021", "GO_Molecular_Function_2021", "MSigDB_Hallmark_2020", and "Jensen_COMPARTMENTS" databases.

The KEGG-based pathway integration and visualization tool *Pathview*⁷⁷ was applied. The *p*-value and fold-change derived from GLMM *glmmTMB*() for DEGs in Mic.3 were diagramed in the KEGG pathway 'Cytokine-cytokine receptor interaction' and 'Cholesterol metabolism' (Supplemental Figure 6).

Compositional analysis of bulk RNA-seq

Digital deconvolution was performed using ROSMAP, Mayo, and the Knight ADRC datasets was previously described^{72,79}. Briefly, gene markers were selected for major brain cell-types: neurons, astrocytes, microglia, and oligodendrocytes. The *CellMix*⁸⁰ was used to estimate the relative proportion of four cell types in each sample.

AD and controls in ROSMAP, Mayo, and Knight ADRC were selected for this analysis. Subjects with outliers (defined by Tukey's fences) in any of the four cell types were removed from each study before analysis. The effect of risk and protective variants on cell-type proportion was calculated by the generalized linear model (GLM) *glm*(). The model also listed age, sex, AD status, PC1 and PC2 derived from GWAS data, TIN, PMI, and the remaining risk/protective variant as covariates. A threshold was set as *p*-value < 0.05. The following model was used:

$$\text{Cell-type Proportion} \sim \text{age} + \text{sex} + \text{AD status} + \text{PC1} + \text{PC2} + \text{TIN} + \text{PMI} + \text{risk SNP} + \text{protective SNP}$$

Compositional analysis of single-nucleus RNA-seq

Microglia proportion was calculated relative to total cell count in each donor. The effect of risk and protective variants on microglia proportion was evaluated by GLM *glm*(). The model listed age, sex, AD status, PC1 and PC2 derived from GWAS data, PMI and the alternative

risk/protective variant as covariates. A threshold was set as p -value < 0.05. The following model was used:

$$\text{Microglia Proportion} \sim \text{age} + \text{sex} + \text{AD status} + \text{PC1} + \text{PC2} + \text{PMI} + \text{risk SNP} + \text{protective SNP}$$

Microglia cluster proportion was calculated relative to total microglia cell count in each donor. The effect of risk and protective variant on microglia cluster proportion was evaluated by GLM *glm()*. The model listed age, sex, AD status, PC1 and PC2 derived from GWAS data, PMI, and the remaining risk/protective variant as covariates. A threshold was set as p -value < 0.05. The following model was used:

$$\text{Cluster Proportion} \sim \text{age} + \text{sex} + \text{AD status} + \text{PC1} + \text{PC2} + \text{PMI} + \text{risk SNP} + \text{protective SNP}$$

To verify that the increment of Mic.3 in risk variant is the driving force for overall microglia increase. Microglia proportion amongst all cell types was recalculated in GLM *glm()* after dropping Mic.3 from all samples (Supplemental Figure 7).

Integration analysis with induced transcription factor microglia-like cells (iTF-Microglia)

The single cell RNA-seq dataset for iTF-microglia was downloaded from a publicly available database³⁵. To integrate the human microglia with iTF-microglia, ADAD samples were removed from the human dataset. Both datasets were normalized using *SCTransform* to regress out the variance caused by the total number of the molecules detected (nCount_RNA) and the gene detected (nFeature_RNA) in each cell.

Dataset integration was performed following Seurat's best practices workflow⁸¹. Human microglia from the Knight ADRC snRNA-seq were set as the reference by *FindIntegrationAnchors* and *IntegrateData*. Integration features were set at 3000 using *SelectIntegrationFeatures* and *PrepSCTIntegration*. The Mutual Nearest Neighbours (MNN) *RunFastMNN*⁸² was used as a graphic reduction method. The UMAP and integrated clusters are defined by *FindNeighbors* (PC = 1:20) and *FindClusters* (resolution = 0.2). The integrated map showed the relative location overlap between human Knight ADRC microglia and iTF-microglia. The top 50 DEGs for Microglia cluster 3 (Mic.3) were selected and visualized to display the gene expression pattern between clusters in integrated human brain microglia and iTF-microglia (Supplemental Figure 8).

Cell culture and quantitative assays

Primary mouse microglia culture

Primary mouse microglia cultures were obtained from mixed cultures prepared from the hippocampi and cerebral cortices of WT and *Ch25h* KO mice on postnatal day (P) 1-3. Microglial cells were isolated by shaking flasks for 45 min at 230 rpm on day ten after plating. Cells were then seeded on poly-L-ornithine (Sigma) pre-coated wells in DMEM containing 20% heat-inactivated fetal bovine serum (FBS).

Human macrophage culture and lentivirus transduction

The culture methods for human macrophage and lentivirus vectors used in this experiment were previously described¹⁰. Briefly, peripheral blood mononuclear cells (PBMCs) were purified from human blood on Ficoll-Paque PLUS density gradient (Amersham Biosciences, Piscataway, NJ). PBMCs were cultured in six-well culture plates (3×10^6 cells per well) in RPMI 1640 without FBS to generate macrophages. After 2 hours of culture, PBMCs were washed twice with 1X PBS and cultured in RPMI supplemented with macrophage colony-stimulating factor (MCSF) (50 ng/ml) for seven days at 37°C with 5% CO₂. Supernatants from macrophage cultures were collected, centrifuged at 21,000g for 15 min, and filtered through a 0.22- μ m filter to remove all cells and membrane debris. Macrophage supernatants were frozen in aliquots at -80°C until used in the ELISA for human sTREM2. Lentiviral vectors encoding GFP (PS100071) and human *MS4A4A* WT (NM_148975) in Myc-DDK-tagged vector (RC204646L1) were obtained from OriGene. All constructs were verified by Sanger sequencing. The viral particles were transfected and harvested from the human embryonic kidney (HEK) 293T cells. Cultured monocyte-derived macrophages were transduced with lentivirus vector for 24 hours and then rinsed, and fresh media was added. Cells and media were collected 48 hours later for sTREM2 ELISA and quantitative polymerase chain reaction (qPCR) analyses.

Mouse sTrem2 measurement

Mouse sTrem2 was measured as following. Briefly, an anti-mouse Trem2 antibody (MAB1729, R&D Systems, [2mg/ml]) was used as a capture antibody and coated overnight at 4°C on MaxiSorp 96-well plates (Nalge Nunc International, Rochester, NY) in sodium bicarbonate coating buffer [0.015 M Na₂CO₃ and 0.035 M NaHCO₃ (pH 9.6)]. All the washes between the different steps were done four times with PBS/0.05% Tween 20 (Sigma-Aldrich). After washing, wells were blocked for 4 hours at room temperature (RT) with phosphate-buffered saline (PBS)/3% BSA. Freshly thawed supernatant and recombinant mouse Trem2 standards (msTrem2-Fc (R&D Systems #1729-T2) were incubated in duplicate overnight at 4°C. For detection, wells were incubated with a Trem2 biotinylated detection antibody (R&D Systems, BAF1729, [0.25 μ g/ml]) for 1 hour at RT on an orbital shaker. After washing, wells were incubated with horseradish peroxidase-labeled streptavidin (BD Biosciences, San Jose, CA) for 1 hour at RT with orbital shaking. Horseradish peroxidase visualization was performed with 3,3',5,5' tetramethyl-benzidine (Sigma-Aldrich, St. Louis, MO) added to each well for 10 min at RT in the dark. Color development was stopped by adding an equal volume of 2.5N H₂SO₄. Optical density of each well was determined at 450 nm.

Quantitative PCR

The effect of overexpression was measured by real-time qPCR analysis. Macrophages transduced with the lentiviral particles were collected in TRIzol reagent (Thermo Fisher Scientific), and total RNA was extracted using the RNeasy Mini Kit (Qiagen). cDNA was prepared from the total RNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Gene expression levels were analyzed by real-time PCR using TaqMan assays for *MS4A4A* (Hs00254780_m1), *TREM2* (Hs00219132_m1), *CH25H* (Hs02379634_s1), *ATP6V0D2* (Hs00403032_m1), *LAMP2* (Hs00174474_m1), *CASP1*

(Hs00354836_m1), *NLRP1* (Hs00248187_m1), *CCL2* (Hs00234140_m1), and *GAPDH* (Hs02758991_g1) on a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific). To avoid amplification interference, expression assays were run in separate wells from the housekeeping gene *GAPDH*. Real-time data were analyzed by the comparative C_T method. Average C_T values for each sample were normalized to the average C_T values for the housekeeping gene *GAPDH*.

Calcium Assay

THP-1 monocytes were cultured and passaged in T75 flask and differentiated into macrophages using two doses of 20ng/mL PMA treatment every two days. On the fourth day, mature THP-1 macrophages were harvested and seeded into a 96-well plate at a density of 50K cells/well. After allowing cells to adhere, cells were treated with 2 pmol/well of scrambled (ThermoFisher Cat. No. 4390843) or *MS4A4A* silencer s27981 (ThermoFisher Cat. No. 4392420) were treated following standard protocol provided by Lipofectamine RNAiMAX (Thermo Fisher Cat. No. 13778150). Media was removed and washed with 1X PBS three days after treatment. The macrophages were incubated at 37°C with 50μL Buffer + 50μL Fluo-4 (Fluo-4 Direct™ Calcium Assay Kit Cat. No. F10472) and incubated for 1 hour. After incubation, the baselines were read with a BioTek Synergy plate reader at 494 nm Excitation/516nm Emission. The kinetics of calcium stimulants were recorded immediately after adding SDF1α (1μg/mL; PeproTech Cat. No. AF-300-28A) or ionomycin (1μM; STEMCELL Technologies Cat. No. 73722) for 300 seconds. Analyses were performed with the proprietary software Gen5.

Acknowledgements

We would like to thank the research participants and their families who generously participated in this study. We thank Torri Ball and Grant Galasso for thoughtful discussions. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center, and the Departments of Neurology and Psychiatry at Washington University School of Medicine. Funding provided by the National Institutes of Health (AG067764, AG005681, AG062734, AG032438, AG058501, AG071706, NS118146 and NS127211), Hope Center for Neurological Disorders, Chan Zuckerberg Initiative (CMK, OH, MK), Thome Memorial Foundation, BrightFocus Foundation, and UL1TR002345, Alzheimer's Association ZEN-22-969903 (MK). The recruitment and clinical characterization of Knight ADRC research participants at Washington University were supported by NIH P30AG066444 (JCM), P01AG03991 (JCM), and P01AG026276 (JCM). Diagrams were generated using BioRender.com.

Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing,

targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. Genotype data: doi:10.1038/mp.2017.20. RNAseq: doi:10.1038/s41593-018-0154-9.

Data collection and sharing for this project was supported by The Dominantly Inherited Alzheimer Network (DIAN, U19AG032438) funded by the National Institute on Aging (NIA), the Alzheimer's Association (SG-20-690363-DIAN), the German Center for Neurodegenerative Diseases (DZNE), Raul Carrea Institute for Neurological Research (FLENI), Partial support by the Research and Development Grants for Dementia from Japan Agency for Medical Research and Development, AMED, and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), Spanish Institute of Health Carlos III (ISCIII), Canadian Institutes of Health Research (CIHR), Canadian Consortium of Neurodegeneration and Aging, Brain Canada Foundation, and Fonds de Recherche du Québec – Santé. This manuscript has been reviewed by DIAN Study investigators for scientific content and consistency of data interpretation with previous DIAN Study publications. We acknowledge the altruism of the participants and their families and contributions of the DIAN research and support staff at each of the participating sites for their contributions to this study.

Disclosures

M. Kampmann is an inventor on US Patent 11,254,933 related to CRISPRi and CRISPRa screening, serves on the Scientific Advisory Boards of Engine Biosciences, Casma Therapeutics, Cajal Neuroscience and Alector, and is an advisor to Modulo Bio and Recursion Therapeutics. L. Piccio received research grant funding from Alector.

Figure Legends

Figure 1. Study overview. A. MS4A4A and TREM2 reside on the plasma membrane in microglia. Prior work demonstrates that rs1582763 is associated with elevated CSF sTREM2 and reduced AD risk (termed: protective) and rs6591561 is associated with reduced CSF sTREM2 and increased AD risk (termed: risk) B. Trans-eQTL analyses for rs1582763 and rs6591561 were performed using data derived from tissue samples representing three cortical areas in three independent cohorts (ROSMAP, frontal; Knight ADRC, parietal; Mayo, temporal). C. Trans-eQTL analyses were performed in human microglia clusters from single nuclei RNA-seq. D. Functional genomics were applied to define the impact of rs1582763 and rs6591561 on microglia function using iTF-microglia. E. Summary of major findings.

Figure 2. Brain transcriptomics reveals that the AD protective variant rs1582763 reduces the inflammatory response and increases lipid metabolism pathways. A. Meta-analysis of the impact of rs1582763 on gene expression in 3 independent brain cohorts. Volcano plot of genes that are differentially expressed in the presence of rs1582763. B. Pathway analysis reveals genes enriched in pathways related to immune signaling and lipid metabolism. Red, up-regulated genes. Blue, down-regulated genes. Gray, all differentially expressed genes. C. Sorted brain microglia (ROSMAP; n=10 brains) were queried to determine whether identified DEGs were enriched in microglia. The waffle plot illustrates the number of genes present in FACS-isolated microglia. Red, up-regulated genes. Blue, down-regulated genes. Gray, absent in the microglia dataset. D. Venn diagram shows the number of differentially expressed genes within three independent cohort (FDR < 0.05). Covariates include age, sex and disease status. E. Gene network diagram reveals the relatedness of *CH25H*, *DCDC1*, *FOSB*, and *PTX3*. Network: Red, physical interaction. Purple, co-expression. Olive, shared protein domain. F. Schematic of primary microglia culture from WT and *Ch25h* KO mice. Primary mouse microglia from *Ch25h* KO mice display significantly higher sTrem2 levels in the media detected by ELISA ($p = 1.0 \times 10^{-3}$).

Figure 3. Single nucleus RNA-seq in human brains reveals that the AD protective variant rs1582763 alters a specific microglia population and drives the expression of genes associated with chemokine and lipid metabolism pathways. A. Schematic of single nucleus brain RNA-seq. B. UMAP reveals a significant increase in the proportion of a specific population of microglia (Mic.3) in the presence of the rs1582763 minor allele (A) (The fraction below indicates the proportion of Mic.3 in the total microglia of given genotype carriers). C. *MS4A4A* expression is significantly elevated in Mic.3 in the presence of the minor allele of rs1582763 (Kruskal-Wallis test, $p = 1.11 \times 10^{-2}$). D. Bar graph of the number of genes differentially expressed as a function of the rs1582763 genotype across microglia subclusters reveals that differentially expressed genes are enriched in Mic.3. Grey, $p < 0.05$. Black, FDR < 0.05. E. Volcano plot of genes differentially expressed based on rs1582763 genotype in Mic.3. F. Pathway analysis reveals that rs1582763 alters genes within Mic.3 that influence immune response, lipid metabolism, protein degradation and calcium regulation.

Figure 4. The AD risk variant rs6591561 is associated with upregulation of inflammasome and inflammatory cytokines and reduced lipid metabolism. A. Meta-analysis of the impact of rs6591561 on gene expression in three independent brain cohorts. Volcano plot of genes that are differentially expressed in the presence of rs6591561. B. Sorted brain microglia

(ROSMAP; n=10 brains) were queried to determine whether genes identified in (A) were enriched in microglia. The waffle plot illustrates the number of genes present in sorted microglia. Red, up-regulated genes. Blue, down-regulated genes. Gray, absent in the microglia dataset. C. Pathway analysis of genes from (A) reveals genes enriched in pathways related to immune signaling and lipid metabolism. Red, up-regulated genes. Blue, down-regulated genes. Gray, all differentially expressed genes. D. Single nucleus RNA-seq reveals that rs6591561 is associated with an increased microglia proportion (LR, $p = 1.11 \times 10^{-2}$). E. UMAP reveals a significant increase in the proportion of a specific microglia population of microglia (Mic.3) in the presence of the minor allele of rs6591561 (LR, $p = 4.52 \times 10^{-2}$). F. *MS4A4A* expression is reduced in minor allele carriers of rs6591561 in Mic.3 (ANOVA, $p = 7.91 \times 10^{-8}$). G. Bar graph of the number of genes differentially expressed as a function of the rs6591561 genotype across microglia subpopulations reveals that differentially expressed genes are enriched in Mic.3. Gray, $p < 0.05$. Black, $FDR < 0.05$. H. Volcano plot reveals genes differentially expressed based on the rs6591561 genotype in Mic.3. I. Pathway analysis reveals that rs6591561 alters genes involved in elevated inflammasome, inflammatory cytokines, enhanced lipid storage, and decreased lipid metabolism in Mic.3. J. Schematic representation of Fluo-4 calcium assay on THP-1 macrophage treated with *MS4A4A* siRNA (s27981). K. qPCR reveals *MS4A4A* is significantly reduced in the presence of *MS4A4A* siRNA. L. Baseline intracellular Ca^{2+} levels are elevated after treatment with *MS4A4A* siRNA s27981. M. Intracellular Ca^{2+} kinetics after treated with SDF1 α . N. Intracellular Ca^{2+} kinetics after treated with ionomycin ****, $p < 0.0001$.

Figure 5. Variants in the MS4A locus shift microglia identity. A. A volcano plot illustrates differentially expressed genes between Mic.3 and all microglia subclusters. The inset graphic shows the reference location of Mic.3. B. Pathway analysis reveals enrichment of genes associated with chemokines, immune/inflammatory response and lipid metabolism. Red, upregulated DEGs. Blue, downregulated DEGs. Gray, all differentially expressed genes. C. (Left) A graphic illustration shows CROP-seq from iTF-microglia. (Right) A dot plot shows the expression level of the top 50 Mic.3 differentially expressed genes in Knight ADRC brain microglia and iTF-microglia. Average expression indicated by color scale. Percent expressed illustrated by the size of the circle. D. An UMAP diagram shows homeostatic (Mic.0), activated (Mic.1) and *MS4A4A*-dependent (Mic.3) microglia clusters trajectory by Monocle 3. E. Scatterplots show imputed relative expression of *CXCR4* (anti-inflammatory)/*CD44* (pro-inflammatory) in homeostatic (Mic.0), activated (Mic.1), and *MS4A4A*-dependent (Mic.3) microglia cluster generated by MAGIC. F. A hypothesis proposes that disease-associated microglia (DAM) differentiate into the pro-inflammatory DAM and the anti-inflammatory DAM. G-H. Peripheral blood mononuclear cell (PBMC) derived macrophages were transduced with lentivirus particles carrying GFP or *MS4A4A* WT. G. RNA expression levels of *MS4A4A*, chemokine, inflammasome, and lysosomal genes were plotted. H. sTREM2 levels in media from GFP or *MS4A4A*-encoding lentiviral transduced macrophages. *, $p < 0.05$; **, $p < 0.001$.

Figure 6. The AD protective and risk variants in the MS4A locus alters the microglial state transcriptionally and functionally. A. Left, schematic of CROP-seq from iTF-microglia. Right, *MS4A4A* expression is enriched in the iTF-microglia cluster associated with interferon function (top). Right, genes differentially expressed in Mic.3 compared to all microglia

subclusters are enriched in iTF-microglia chemokine clusters (bottom). B. Molecular signature of the protective variant is enriched in interferon iTF-microglia (left) and alters transcription factor networks (right). C. Risk variant signature is dampened across interferon and chemokine iTF-microglia (left) and alters transcription factor networks (right). D. Compounds that mimic the molecular signature of the protective and reverse risk variant effects in Mic.3 are shown along with the target percentage and top 10 compounds.

References

1. Novikova, G., *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun* **12**, 1610 (2021).
2. Wightman, D.P., *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**, 1276-1282 (2021).
3. Bellenguez, C., *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* **54**, 412-436 (2022).
4. Kunkle, B.W., *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
5. Nott, A., *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366**, 1134-1139 (2019).
6. Hansen, D.V., Hanson, J.E. & Sheng, M. Microglia in Alzheimer's disease. *J Cell Biol* **217**, 459-472 (2018).
7. Mattioli, I., Mantovani, A. & Locati, M. The tetraspan MS4A family in homeostasis, immunity, and disease. *Trends Immunol* **42**, 764-781 (2021).
8. Deming, Y., *et al.* The MS4A gene cluster is a key regulator of soluble TREM2 and Alzheimer disease risk. *bioRxiv* (2018).
9. Piccio, L., *et al.* Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta Neuropathol* **131**, 925-933 (2016).
10. Deming, Y., *et al.* The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci Transl Med* **11**(2019).
11. Lambert, J.C., *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* (2013).
12. Naj, A.C., *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature genetics* **43**, 436-441 (2011).
13. Gratuze, M., *et al.* Impact of TREM2R47H variant on tau pathology-induced gliosis and neurodegeneration. *J Clin Invest* **130**, 4954-4968 (2020).
14. Song, W.M., *et al.* Humanized TREM2 mice reveal microglia-intrinsic and -extrinsic effects of R47H polymorphism. *J Exp Med* **215**, 745-760 (2018).
15. Griciuc, A., *et al.* Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron* **78**, 631-643 (2013).
16. Lin, P.B., *et al.* INPP5D deficiency attenuates amyloid pathology in a mouse model of Alzheimer's disease. *Alzheimers Dement* (2022).
17. Olah, M., *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat Commun* **11**, 6129 (2020).
18. Olah, M., *et al.* A transcriptomic atlas of aged human microglia. *Nat Commun* **9**, 539 (2018).
19. Rajab, N., *et al.* An integrated analysis of human myeloid cells identifies gaps in in vitro models of in vivo biology. *Stem Cell Reports* **16**, 1629-1643 (2021).
20. Kaplan, A. & Reiner, O. Linking cytoplasmic dynein and transport of Rab8 vesicles to the midbody during cytokinesis by the doublecortin domain-containing 5 protein. *J Cell Sci* **124**, 3989-4000 (2011).
21. Garlanda, C., Bottazzi, B., Magrini, E., Inforzato, A. & Mantovani, A. PTX3, a Humoral Pattern Recognition Molecule, in Innate Immunity, Tissue Repair, and Cancer. *Physiol Rev* **98**, 623-639 (2018).

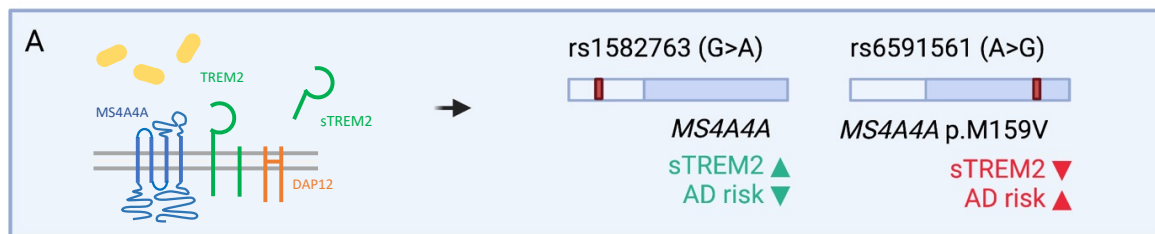
22. Ham, D.W., *et al.* Chronic *Toxoplasma gondii* Infection Alleviates Experimental Autoimmune Encephalomyelitis by the Immune Regulation Inducing Reduction in IL-17A/Th17 Via Upregulation of SOCS3. *Neurotherapeutics* **18**, 430-447 (2021).
23. Stravalaci, M., *et al.* Control of Complement Activation by the Long Pentraxin PTX3: Implications in Age-Related Macular Degeneration. *Front Pharmacol* **11**, 591908 (2020).
24. Ummenthum, K., *et al.* Pentraxin-3 is upregulated in the central nervous system during MS and EAE, but does not modulate experimental neurological disease. *Eur J Immunol* **46**, 701-711 (2016).
25. Nomaru, H., *et al.* Fosb gene products contribute to excitotoxic microglial activation by regulating the expression of complement C5a receptors in microglia. *Glia* **62**, 1284-1298 (2014).
26. Wong, M.Y., *et al.* 25-Hydroxycholesterol amplifies microglial IL-1beta production in an apoE isoform-dependent manner. *J Neuroinflammation* **17**, 192 (2020).
27. Hammond, T.R., *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253-271 e256 (2019).
28. Keren-Shaul, H., *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290 e1217 (2017).
29. Masuda, T., Sankowski, R., Staszewski, O. & Prinz, M. Microglia Heterogeneity in the Single-Cell Era. *Cell Rep* **30**, 1271-1281 (2020).
30. Mathys, H., *et al.* Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. *Cell Rep* **21**, 366-380 (2017).
31. Del-Aguila, J.L., *et al.* A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers Res Ther* **11**, 71 (2019).
32. Brase, L., *et al.* A landscape of the genetic and cellular heterogeneity in Alzheimer disease. *medRxiv*, 2021.2011.2030.21267072 (2021).
33. Da Mesquita, S., *et al.* Meningeal lymphatics affect microglia responses and anti-Abeta immunotherapy. *Nature* **593**, 255-260 (2021).
34. Paolicelli, R.C., *et al.* Microglia states and nomenclature: A field at its crossroads. *Neuron* **110**, 3458-3483 (2022).
35. Drager, N.M., *et al.* A CRISPRi/a platform in human iPSC-derived microglia uncovers regulators of disease states. *Nat Neurosci* **25**, 1149-1162 (2022).
36. Rangaraju, S., *et al.* Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. *Mol Neurodegener* **13**, 24 (2018).
37. Cao, J., *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
38. La Manno, G., *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
39. van Dijk, D., *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729 e727 (2018).
40. Keenan, A.B., *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res* **47**, W212-W224 (2019).
41. 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).
42. Pimenova, A.A., *et al.* Alzheimer's-associated PU.1 expression levels regulate microglial inflammatory response. *Neurobiol Dis* **148**, 105217 (2021).

43. Lamb, J., *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935 (2006).
44. Subramanian, A., *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417 (2017).
45. Sanyal, R., *et al.* MS4A4A: a novel cell surface marker for M2 macrophages and plasma cells. *Immunol Cell Biol* **95**, 611-619 (2017).
46. Feringa, F.M. & van der Kant, R. Cholesterol and Alzheimer's Disease; From Risk Genes to Pathological Effects. *Front Aging Neurosci* **13**, 690372 (2021).
47. Corder, E.H., *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923 (1993).
48. Saunders, A.M., *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* **43**, 1467-1472 (1993).
49. Strittmatter, W.J., *et al.* Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 8098-8102 (1993).
50. Cantuti-Castelvetri, L., *et al.* Defective cholesterol clearance limits remyelination in the aged central nervous system. *Science* **359**, 684-688 (2018).
51. Xing, J., Titus, A.R. & Humphrey, M.B. The TREM2-DAP12 signaling pathway in Nasu-Hakola disease: a molecular genetics perspective. *Res Rep Biochem* **5**, 89-100 (2015).
52. Ulland, T.K., *et al.* TREM2 Maintains Microglial Metabolic Fitness in Alzheimer's Disease. *Cell* **170**, 649-663 e613 (2017).
53. Nugent, A.A., *et al.* TREM2 Regulates Microglial Cholesterol Metabolism upon Chronic Phagocytic Challenge. *Neuron* **105**, 837-854 e839 (2020).
54. Adorni, M.P., Ruscica, M., Ferri, N., Bernini, F. & Zimetti, F. Proprotein Convertase Subtilisin/Kexin Type 9, Brain Cholesterol Homeostasis and Potential Implication for Alzheimer's Disease. *Front Aging Neurosci* **11**, 120 (2019).
55. Zimetti, F., *et al.* Increased PCSK9 Cerebrospinal Fluid Concentrations in Alzheimer's Disease. *J Alzheimers Dis* **55**, 315-320 (2017).
56. Apaijai, N., *et al.* Pretreatment With PCSK9 Inhibitor Protects the Brain Against Cardiac Ischemia/Reperfusion Injury Through a Reduction of Neuronal Inflammation and Amyloid Beta Aggregation. *J Am Heart Assoc* **8**, e010838 (2019).
57. Ewers, M., *et al.* Increased soluble TREM2 in cerebrospinal fluid is associated with reduced cognitive and clinical decline in Alzheimer's disease. *Sci Transl Med* **11**(2019).
58. Fu, H., *et al.* Interleukin-1 potently contributes to 25-hydroxycholesterol-induced synergistic cytokine production in smooth muscle cell-monocyte interactions. *Atherosclerosis* **237**, 443-452 (2014).
59. Gold, E.S., *et al.* 25-Hydroxycholesterol acts as an amplifier of inflammatory signaling. *Proc Natl Acad Sci U S A* **111**, 10666-10671 (2014).
60. Koarai, A., *et al.* 25-Hydroxycholesterol enhances cytokine release and Toll-like receptor 3 response in airway epithelial cells. *Respir Res* **13**, 63 (2012).
61. Rosklint, T., Ohlsson, B.G., Wiklund, O., Noren, K. & Hulten, L.M. Oxysterols induce interleukin-1beta production in human macrophages. *Eur J Clin Invest* **32**, 35-42 (2002).
62. Liu, Y., *et al.* 25-Hydroxycholesterol activates the expression of cholesterol 25-hydroxylase in an LXR-dependent mechanism. *J Lipid Res* **59**, 439-451 (2018).

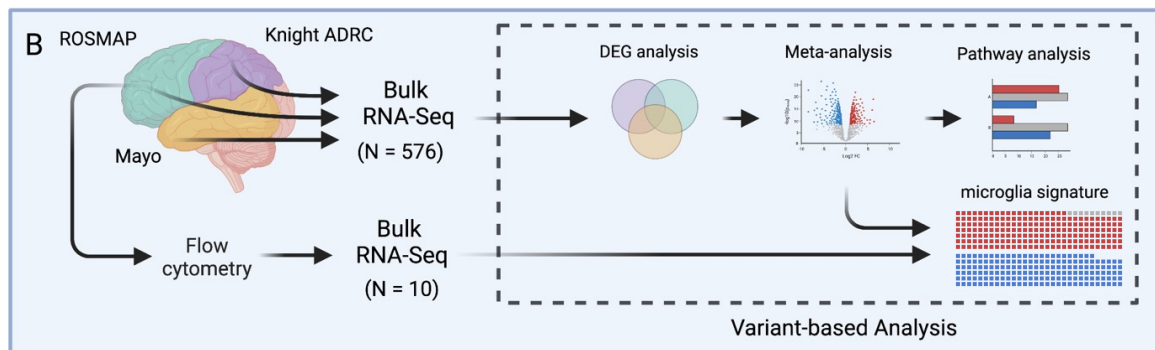
63. Adams, C.M., *et al.* Cholesterol and 25-hydroxycholesterol inhibit activation of SREBPs by different mechanisms, both involving SCAP and Insigs. *J Biol Chem* **279**, 52772-52780 (2004).
64. Cashikar, A.G., *et al.* Regulation of astrocyte lipid metabolism and ApoE secretion by the microglial oxysterol, 25-hydroxycholesterol. *bioRxiv*, 2022.2005.2003.490494 (2022).
65. Dang, E.V., McDonald, J.G., Russell, D.W. & Cyster, J.G. Oxysterol Restraint of Cholesterol Synthesis Prevents AIM2 Inflammasome Activation. *Cell* **171**, 1057-1071 e1011 (2017).
66. Reboldi, A., *et al.* Inflammation. 25-Hydroxycholesterol suppresses interleukin-1-driven inflammation downstream of type I interferon. *Science* **345**, 679-684 (2014).
67. Jang, J., *et al.* 25-hydroxycholesterol contributes to cerebral inflammation of X-linked adrenoleukodystrophy through activation of the NLRP3 inflammasome. *Nat Commun* **7**, 13129 (2016).
68. Chalmin, F., *et al.* Oxysterols regulate encephalitogenic CD4(+) T cell trafficking during central nervous system autoimmunity. *J Autoimmun* **56**, 45-55 (2015).
69. Bennett, D.A., Schneider, J.A., Arvanitakis, Z. & Wilson, R.S. Overview and findings from the religious orders study. *Curr Alzheimer Res* **9**, 628-645 (2012).
70. Bennett, D.A., *et al.* Overview and findings from the rush Memory and Aging Project. *Curr Alzheimer Res* **9**, 646-663 (2012).
71. Allen, M., *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data* **3**, 160089 (2016).
72. Li, Z., *et al.* Genetic variants associated with Alzheimer's disease confer different cerebral cortex cell-type population structure. *Genome Med* **10**, 43 (2018).
73. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
74. Franz, M., *et al.* GeneMANIA update 2018. *Nucleic Acids Res* **46**, W60-W64 (2018).
75. Kuleshov, M.V., *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).
76. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311 (2009).
77. Luo, W., Pant, G., Bhavnasi, Y.K., Blanchard, S.G., Jr. & Brouwer, C. Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res* **45**, W501-W508 (2017).
78. Mollie E. Brooks, K.K., Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Mächler and Benjamin M. Bolker. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal* **9**, 378-400 (2017).
79. Li, Z., *et al.* The TMEM106B FTLN-protective variant, rs1990621, is also associated with increased neuronal proportion. *Acta Neuropathol* **139**, 45-61 (2020).
80. Gaujoux, R. & Seoighe, C. CellMix: a comprehensive toolbox for gene expression deconvolution. *Bioinformatics* **29**, 2211-2212 (2013).
81. Luecken, M.D. & Theis, F.J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* **15**, e8746 (2019).

82. Haghverdi, L., Lun, A.T.L., Morgan, M.D. & Marioni, J.C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36**, 421-427 (2018).

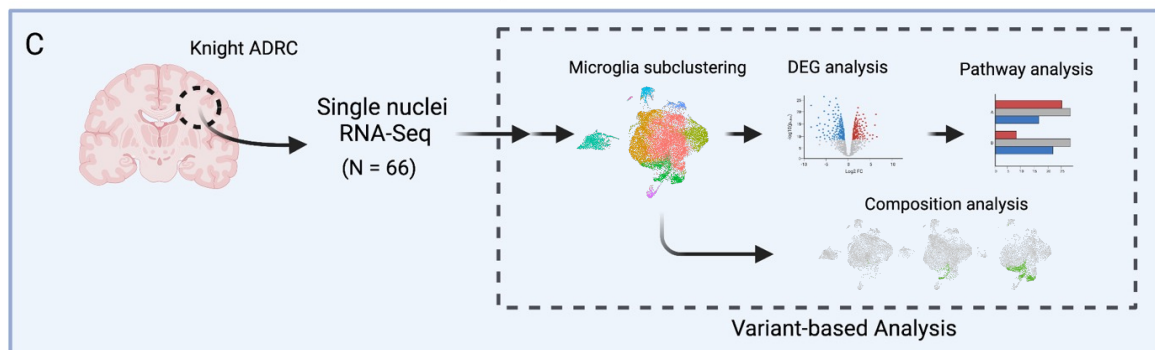
Figure 1



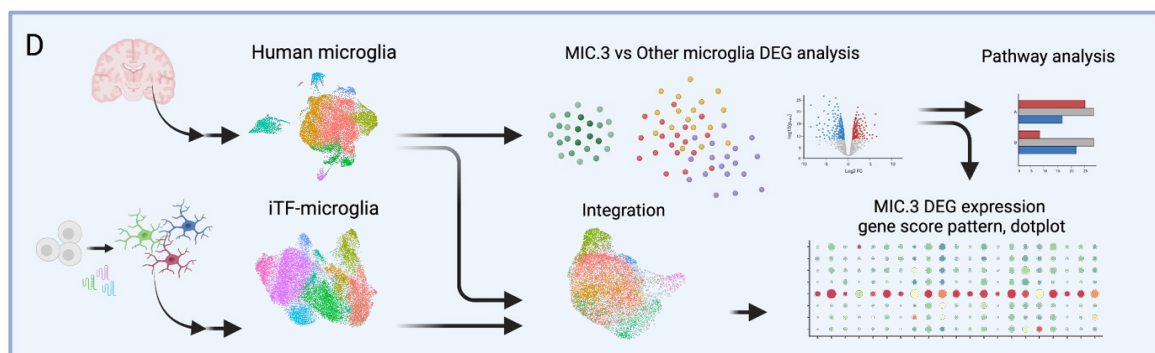
Variant-based differential expression analysis across large scale and multiple brain regions



Variant-based differential expression analysis in microglia-specific dataset



High throughput functional characterization for microglia cluster 3



Graphical Abstract

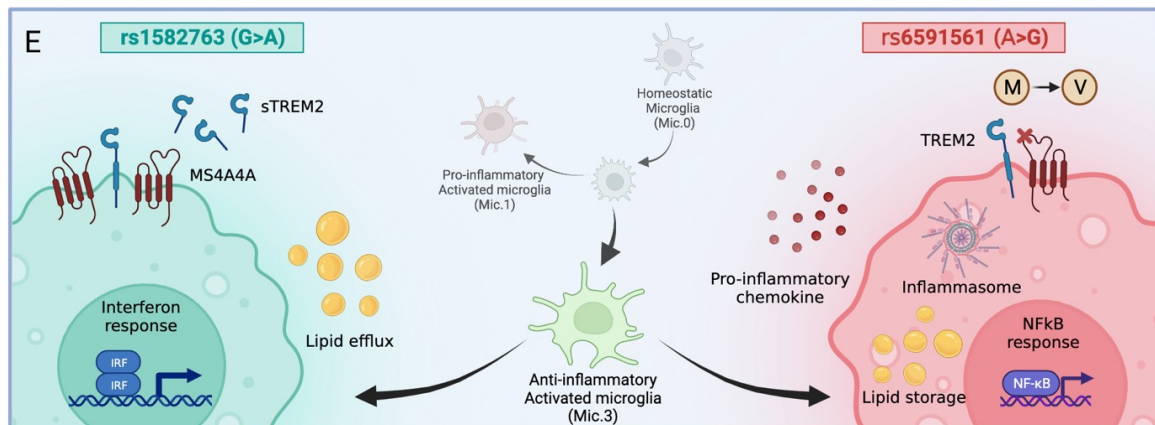


Figure 2

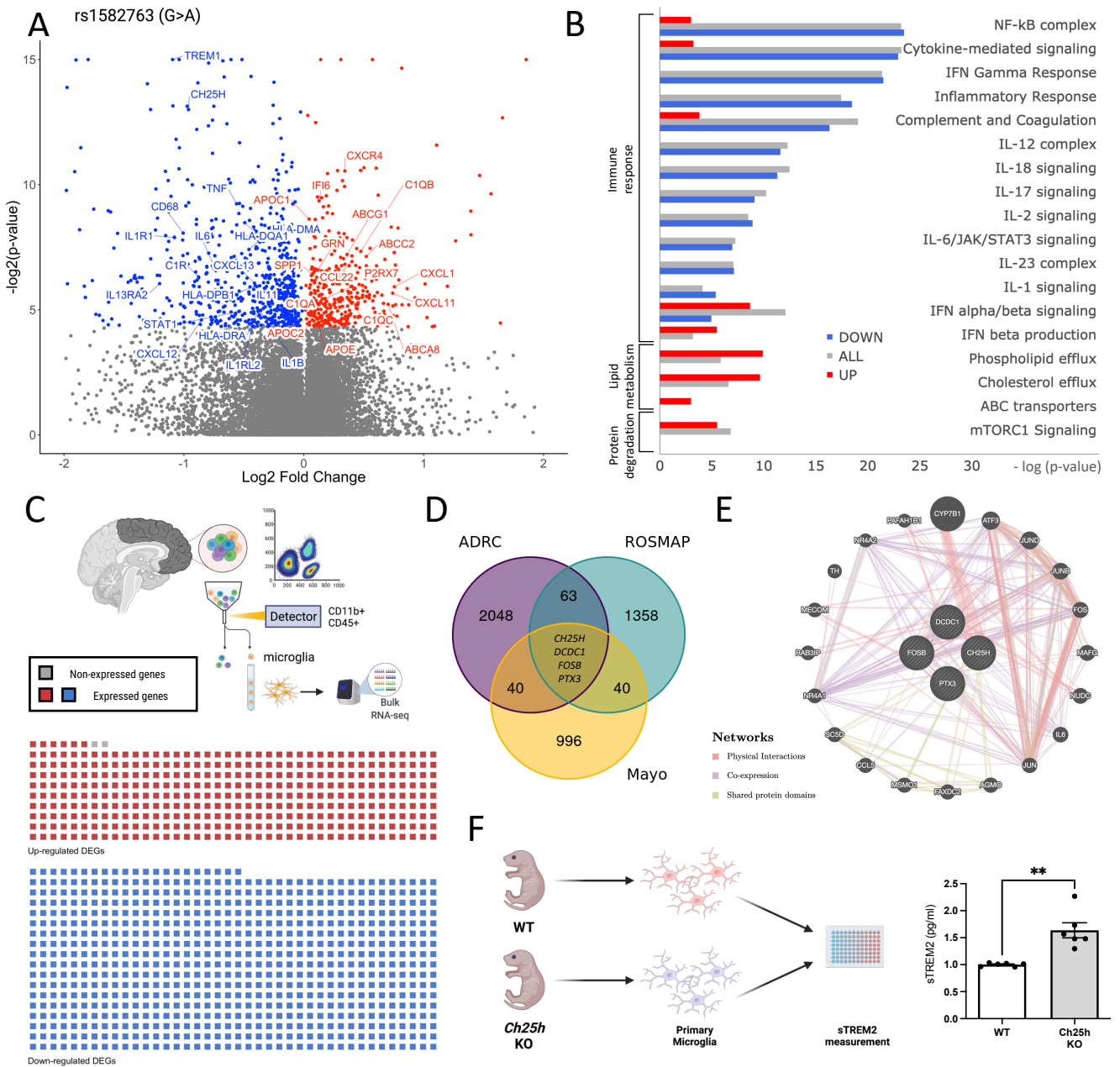


Figure 3

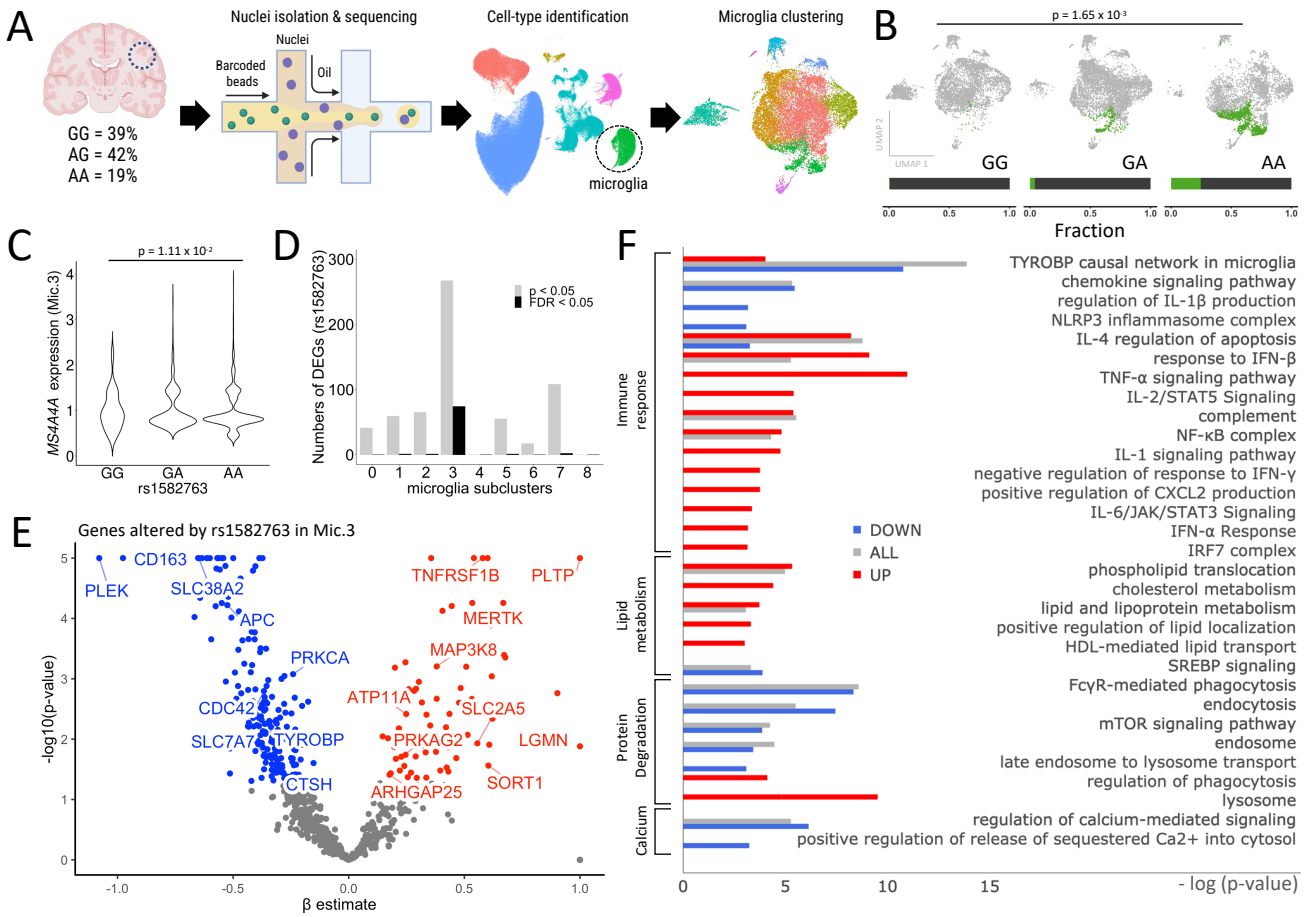


Figure 4

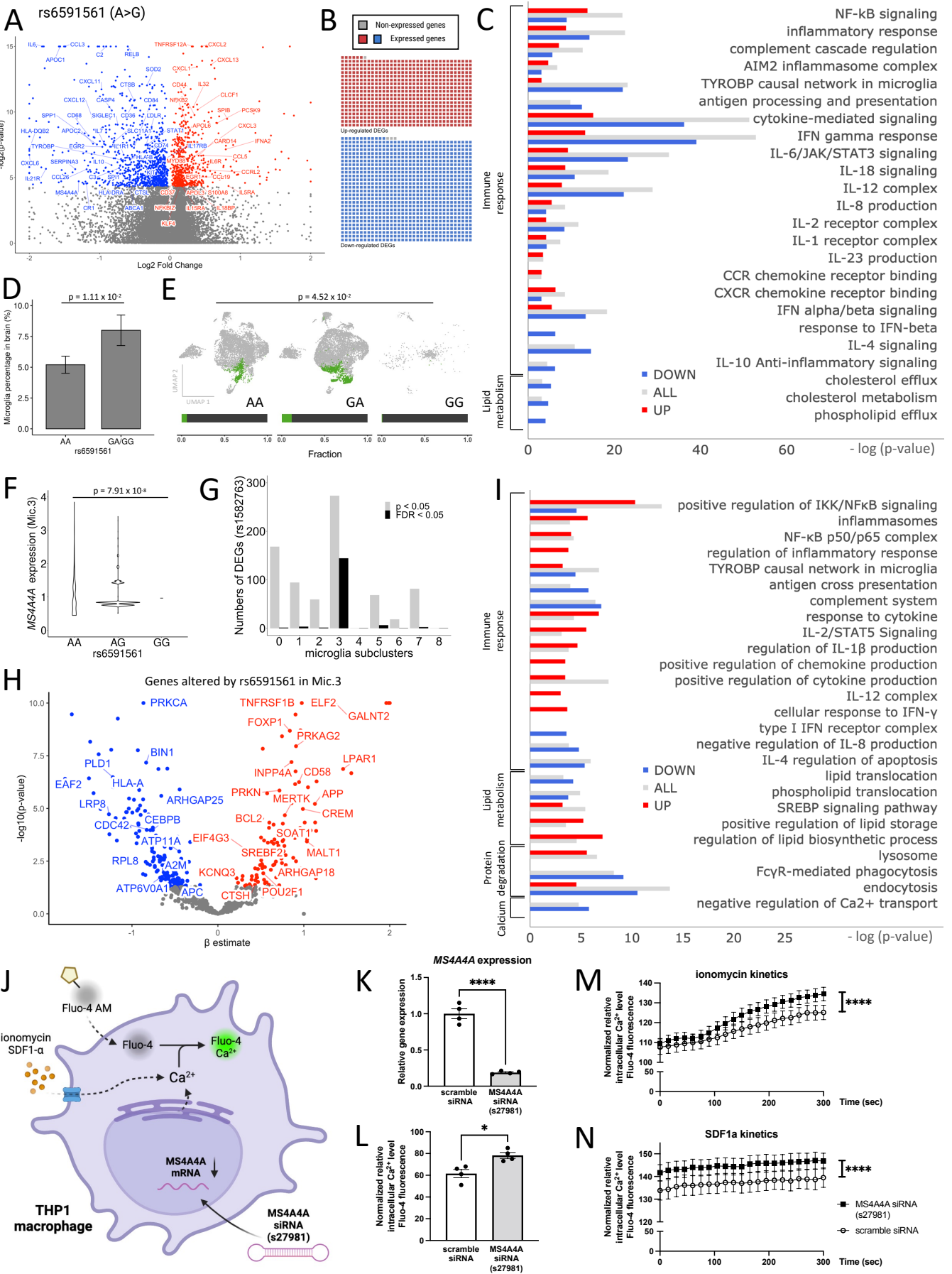


Figure 5

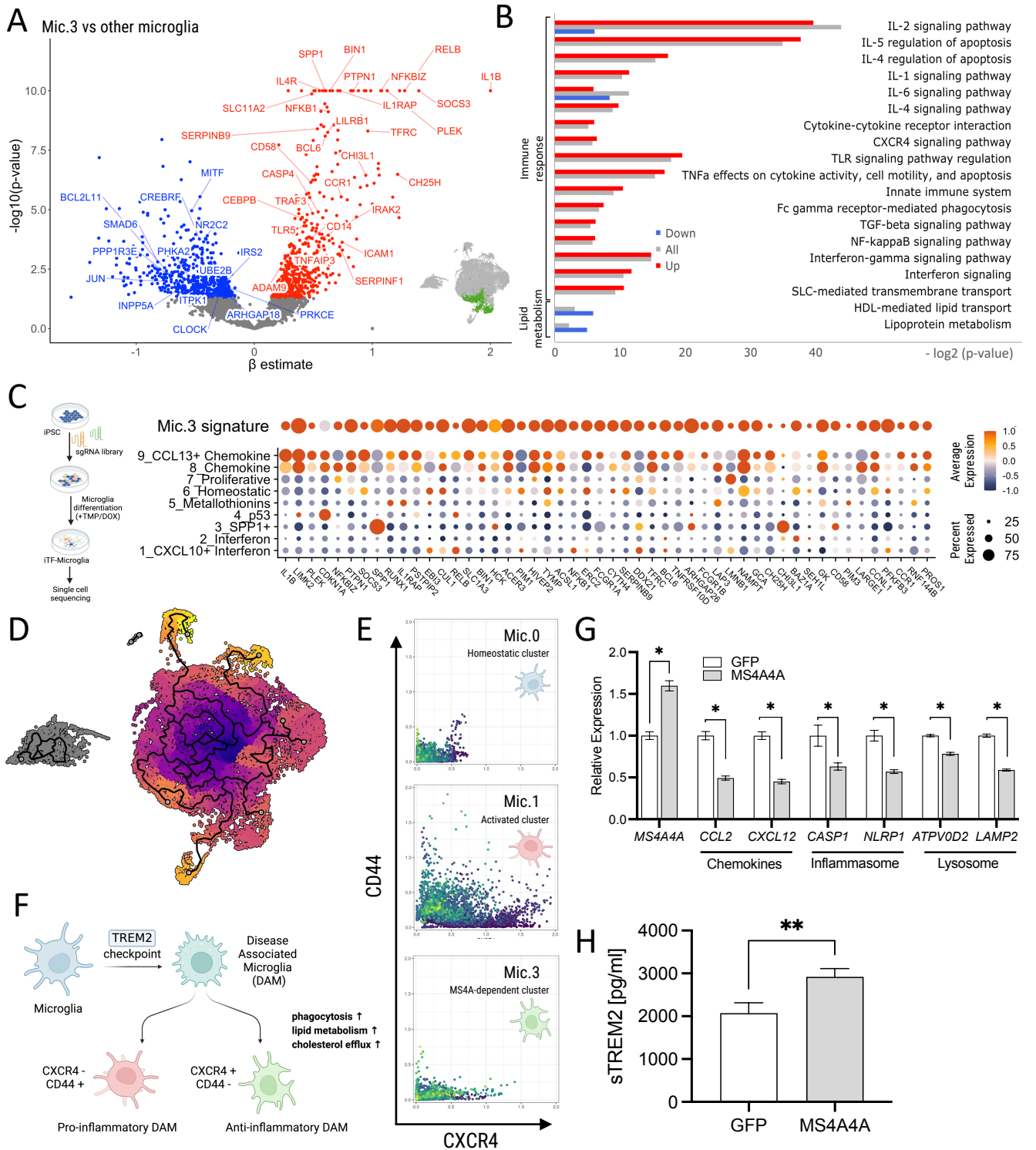


Figure 6

