# Comparative gene regulatory networks modulating *APOE* expression in microglia and astrocytes

Logan Brase[1,2], Yanbo Yu[1,2], Eric McDade[3], Dominantly Inherited Alzheimer Network (DIAN)°, Oscar Harari[1,2,4,¶], Bruno A. Benitez[5,¶,*]

Affiliations:

1. Department of Psychiatry, Washington University, Saint Louis, St. Louis, Missouri, United States of America
2. The Charles F. and Joanne Knight Alzheimer Disease Research Center, Washington University, St. Louis, Missouri, United States of America
3. Department of Neurology, Washington University School of Medicine in St. Louis, St. Louis, MO, USA
4. Department of Neurology, Division of Neurogenetics, The Ohio State University, Columbus, OH, United States of America
5. Department of Neurology and Neuroscience, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts, United States of America

° A list of authors and their affiliations appears at the end of the paper.

¶ Equally contributing senior author

* Corresponding author

E-mail: bbenitez@bidmac.harvard.edu

## Abstract

**Background**: Single-cell technologies have unveiled various transcriptional states in different brain cell types. Transcription factors (TFs) regulate the expression of related gene sets, thereby controlling these diverse expression states. Apolipoprotein E (*APOE*), a pivotal risk-modifying gene in Alzheimer's disease (AD), is expressed in specific glial transcriptional states associated with AD. However, it is still unknown whether the upstream regulatory programs that modulate its expression are shared across brain cell types or specific to microglia and astrocytes.

**Methods**: We used pySCENIC to construct state-specific gene regulatory networks (GRNs) for resting and activated cell states within microglia and astrocytes based on single-nucleus RNA sequencing data from AD patients' cortices from the Knight ADRC-DIAN cohort. We then identified replicating TF using data from the ROSMAP cohort. We identified sets of genes co-regulated with *APOE* by clustering the GRN target genes and identifying genes differentially expressed after the virtual knockout of TFs regulating *APOE*. We performed enrichment analyses on these gene sets and evaluated their overlap with genes found in AD GWAS loci.

**Results**: We identified an average of 96 replicating regulators for each microglial and astrocyte cell state. Our analysis identified the CEBP, JUN, FOS, and FOXO TF families as key regulators of microglial APOE expression. The steroid/thyroid hormone receptor families, including the THR TF family, consistently regulated APOE across astrocyte states, while CEBP and JUN TF families were also involved in resting astrocytes. AD GWAS-associated genes (*PGRN*, *FCGR3A*, *CTSH*, *ABCA1*, *MARCKS*, *CTSB*, *SQSTM1*, *TSC22D4*, *FCER1G*, and HLA genes) are co-regulated with APOE. We also uncovered that APOE-regulating TFs were linked to circadian rhythm (*BHLHE40*, *DBP*, *XBP1*, *CREM*, *SREBF1*, *FOXO3*, and *NR2F1*).

**Conclusions**: Our findings reveal a novel perspective on the transcriptional regulation of *APOE* in the human brain. We found a comprehensive and cell-type-specific regulatory landscape for *APOE*, revealing distinct and shared regulatory mechanisms across microglia and astrocytes, underscoring the complexity of *APOE* regulation. *APOE*-co-regulated genes might also affect AD risk. Furthermore, our study uncovers a potential link between circadian rhythm disruption and *APOE* regulation, shedding new light on the pathogenesis of AD.

Keywords: Alzheimer Disease, APOE, single-nucleus RNA-seq, gene regulatory networks, microglia, astrocyte

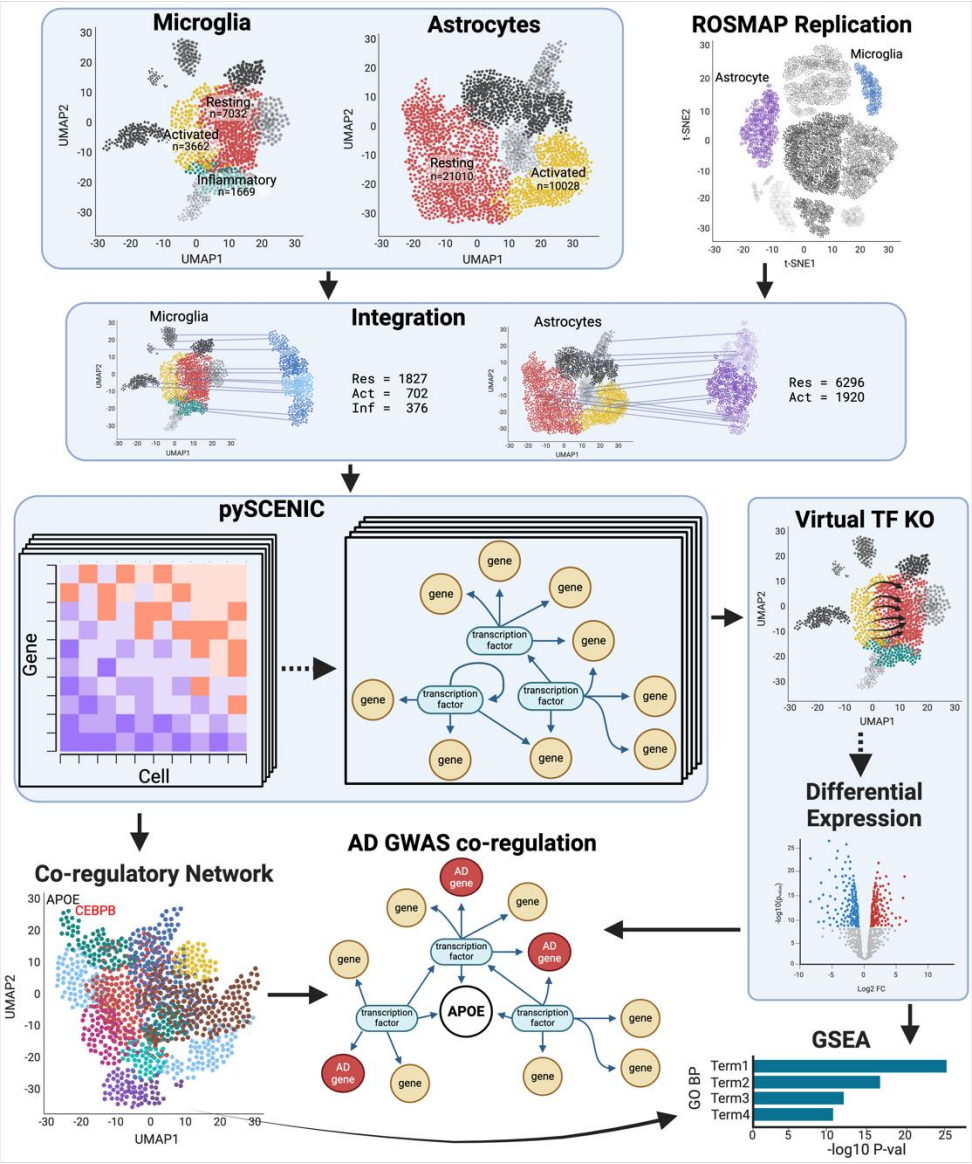
# Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disorder that affects millions of people worldwide. It is characterized by cognitive impairment, memory loss, and behavioral changes. The strongest genetic risk modifiers are variants in the gene *APOE*. It is involved in lipid transport and metabolism in the brain and has two primary variants that influence disease risk, e2 (protective) and e4 (risk). *APOE* is constitutively expressed in astrocytes and activated microglia[1], cell types that are key players in neuroinflammation and neurodegenerative processes[2].

When activated, microglial cells can significantly upregulate *APOE* expression, a process that is crucial for the phagocytic clearance of amyloid-beta and other debris[3]. Astrocytes, utilize *APOE* to support the lipid and energy demands of the surrounding neurons[4]. The ε4 allele, in particular, has been associated with a detrimental effect on microglial activation and astrocytic function, exacerbating neuroinflammatory and neurodegenerative processes[2–4]. These expression states are not static; they evolve in response to the brain's pathological environment, influencing disease progression and severity. Understanding the differences in *APOE* expression regulation between microglial and astrocytic cell states can provide valuable insights into the molecular mechanisms underlying *APOE*'s involvement in AD and potentially other neurodegenerative disorders.

In this study, we aimed to further investigate the differences in the regulation of *APOE* expression between various microglial and astrocytic cell states using single-nucleus RNA-seq (snRNA-seq). To this end, we performed state-specific gene regulatory network (GRN) analysis using pySCENIC on specific cell states (Figure 1). We then used two approaches to identify gene sets co-regulated with *APOE* in these networks: co-regulatory network clusters and transcription factor knockout (TF-KO) differentially expressed genes (DEGs). We assessed the relevance of these gene sets to AD by testing their overlap with AD genome-wide association study (GWAS) genes and their enrichment in AD-related pathways. By delineating the GRNs in these critical cell states, we seek to elucidate the distinct transcriptional regulatory mechanisms that govern *APOE* expression, which is crucial for unraveling the complex interplay between *APOE* regulation and AD pathogenesis. This knowledge may lead to the development of targeted therapeutic

85 interventions to modulate *APOE*-related biology in specific cell types to mitigate disease
86 progression.




87

88 *Figure 1 Analysis workflow*
89 *We extracted microglial and astrocytic cell states and ran them individually through pySCENIC to create state-specific GRNs. We*
90 *replicated the putative pySCENIC-TFs by integrating microglia and astrocytes from the ROSMAP cohort and running pySCENIC on*
91 *them. We clustered the target genes into co-regulatory networks and virtually knocked out the transcription factors (TF KO) that*
92 *regulated APOE. We identified genes co-regulated with APOE using co-regulatory network clusters and TF-KO DEGs. Both sources*
93 *of gene sets were tested for inclusion of AD GWAS genes and gene set enrichment.*

94

# Results

## Cell-state-specific transcription factors modulating APOE expression

To identify the state-specific TF regulators of *APOE*, we utilized snRNA-seq data from the Knight ADRC and DIAN cohorts (parietal, n=67)[5], and replicated the TFs in snRNA-seq from the ROSMAP cohort (dorsolateral prefrontal cortex, n=32)[6]. Focusing on microglia and astrocytes, the principal *APOE* expressors[1], we isolated the resting and activated states for each cell type: mic-resting (n=7,032), mic-activated (n=3,662), mic-inflammatory (n=1,669), astro-resting (n=21,010), and astro-activated (n=10,028). We leveraged pySCENIC[7] to build a GRN for each state, and detected an average of 117 and 109 reproducible TF regulators for Knight and ROSMAP cohorts respectively (Supplementary Table 1). We compared the putative TFs between the cohorts and identified an average of 96 replicating regulators for each cell state (Figure 2, Supplementary Table 1). Among these, 35 TFs were shared by all five states, including *ARID3A* (*ABCA7* locus), *GABPA* (*APP* locus), and *MAF* (*WWOX* locus), which are within AD GWAS loci[8] (Supplementary Table 2). Moreover, 59 TFs were shared by the microglia states, including *SPI1*, *IKZF1*, *TCF3* (*KLF16* locus), and *GABPB1* (*SPPL2A* locus), which are within AD GWAS loci[8,9], and 69 TFs were shared by the astrocyte states including *SREBF1* (*MYO15A* locus) and *ZNF518A* (*BLNK* locus).

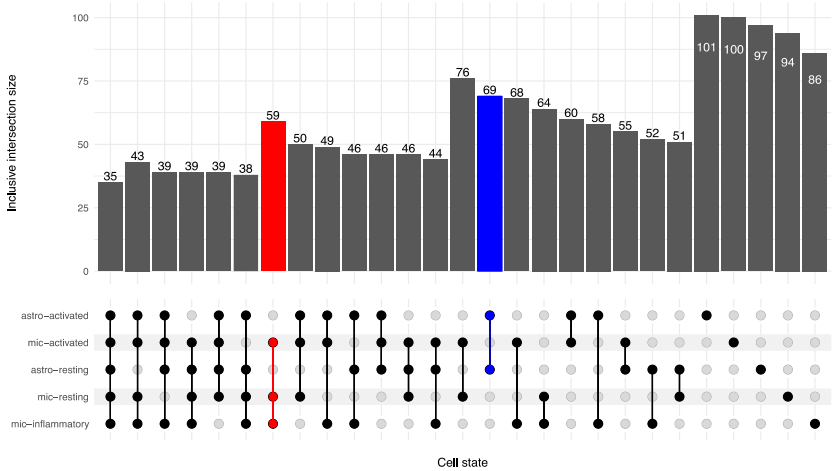



*Figure 2 Inclusive intersection of state-specific TF regulators*
*TF sets were identified by individually running pySCENIC on the Knight/DIAN and ROSMAP cell states and isolating the intersecting TFs between the discovery and replication runs (see Methods). The TF sets were then compared in an upset plot. The inclusive intersection indicates that all subsets are counted in the superset. For example, all 35 TFs shared between all cell states (the lowest subset) are counted in all other sets. The red column indicates the TFs that are common between all microglia states. The blue column indicates that TFs that are common between both astrocyte cell states.*

**Different TF families regulate APOE between microglia and astrocytes**

Next, we pinpointed the TFs regulating *APOE* in each state. In the mic-resting GRN, APOE was regulated by *CEBPA*, *CEBPB*, *CEBPD*, *IRF7*, and *JUND* TFs. In the mic-activated and mic-inflammatory GRNs, APOE was regulated by *CEBPB*, *CEBPD*, and *JUNB*, with *IRF7* also involved in the mic-activated state. In the astrocytic GRNs, *APOE* was regulated by *JUN* and *THRA* in the resting state and by *THRB* in the activated state.

These results highlight the CCAAT/enhancer-binding protein (CEBP) family of TFs, which form homo- and heterodimers to precisely control the expression of target genes that are primarily immune-related[10]. In mouse microglia cell cultures, *CEBPA* was shown to regulate homeostatic and anti-inflammatory DAM genes[11], and in mouse models, *CEBPB* modulates *APOE* expression[12]. These findings validate pySCENIC's ability to capture known biology and lends credence to the other findings. In addition, *CEBPD* was shown to boost the expression of activation genes, including *NOS2*, *C6*, *IL1B*, and *IL6* in a BV2 microglia cell line[13]. Intriguingly, LPS treatment upregulates *CEBPB* and *CEBPD* while downregulating *CEBPA*, implying contrasting regulatory functions. This could account for *CEBPA*'s role as an *APOE* regulator only in mic-resting, but not in mic-activated and mic-inflammatory.

IRF plays a crucial role in the immune response. *IRF7* regulates type I interferon genes and *IL6*[14]. The JUN family proteins are known to be involved in various cellular processes, including proliferation, apoptosis, and immune response[15]. Their role in regulating *APOE* expression suggests a potential link between *APOE* and these cellular processes.

Thyroid hormone receptor family proteins, including *THRA* and *THRB*, bind to thyroid hormone, which plays a crucial role in metabolism, growth, and development. Using cell lines, *THRB* was shown to upregulate *APOE* expression in astrocytes by forming a complex with *RXRA* and binding to the multienhancer ME.2 which activates the *APOE* promoter[16]. Once again, this shows that pySCENIC identifies true regulatory relationships.

Our findings suggest that *CEBPB* and *THRB* regulate *APOE* expression in the human brain, in concordance with studies in model systems. The GRNs from pySCENIC suggest that the CEBP family primarily regulates *APOE* in microglia, while the THR family chiefly regulates expression in astrocytes. This is in line with CEPB regulating immunity genes and microglia being the primary

148 immune cells in the brain, as well as THR influencing metabolism and astrocytes being the
149 primary support cells for neurons. This does not exclude the possibility that both families may
150 influence *APOE* expression in both cell types.
151
152 **APOE-associated target genes were regulated by TFs linked to the circadian rhythm**
153 We aimed to explore those genes co-regulated with *APOE*. Thus, we built a complex co-
154 regulatory network of genes, including all the putative TFs predicted by pySCENIC and their
155 respective regulated genes. Then, we clustered the target genes into a UMAP space using the
156 86-101 TFs signatures that defined each cell state. We identified an average of 27 gene clusters
157 in each expression state (Figure 3, Supplementary Table 1). The gene clusters containing *APOE*
158 comprised an average of 327 genes regulated by an average of 13 TFs for each cell state
159 (Supplementary Table 1).
160 To validate these gene clusters, we isolated the TFs that significantly upregulated the gene
161 clusters containing *APOE*, a known activated marker gene, and *CX3CR1* and *P2RY12*, known
162 homeostatic marker genes, in the three microglia clusters. We then performed a virtual
163 knockout (KO) of these TFs and observed the shift in the expression profile of the cells.
164 We adapted the virtual TF knockout method described in SCENIC+[17] to work with pySCENIC
165 outputs. Following the knockout of the TF, this method propagates the change through the rest
166 of the gene expression matrix (GEM) producing a perturbed expression profile and an
167 associated visualization of the shift in the reduced dimensionality space. The TFs regulating
168 these homeostatic and activated marker genes should have opposite influences on microglial
169 activation, thus opposing shifts in the expression profiles. In the mic-resting state, an individual
170 knockout of the TFs associated with the *APOE* gene cluster generally shifted cells toward the
171 mic-resting state, indicating these TFs are involved in maintaining an activated state. Conversely,
172 individual knockouts of seven of the 14 TFs associated with the *CX3CR1* and *P2RY12* clusters
173 generally pointed away from the mic-resting state, suggesting these TFs promote a resting state
174 (Supplementary Figure 1). A similar trend was seen within mic-activated, with KO of all 17 *APOE*
175 TFs pointing away from mic-activated and eight of the 19 TFs$^{CX3CR1/P2RY12+}$ pointing toward mic-
176 activated (Supplementary Figure 2). This indicates that the clustering groups mirror the

177 expected biology. The mic-inflammatory cluster had no TFs significantly upregulating the
178 *CX3CR1* or *P2RY12* clusters, so it was not explored for validation.
179 We then explored the TFs regulating *APOE* in each state. The mic-resting *APOE* cluster was
180 regulated by *BHLHE40*, *CEBPG*, *FOS*, *FOSL2*, *GABPA*, *ILF2*, *JUN*, *JUNB*, *KMT2B*, *REL*, and *XBP1*, in
181 addition to the TFs directly regulating *APOE* mentioned previously (Figure 3A, Supplementary
182 Figure 4A). *BHLHE40* regulates the circadian rhythm, negatively regulates itself, *DBP* (another
183 circadian gene), and represses RXR, a component of *APOE* activation through the THR family of
184 TFs[18–21]. BHLHE40 is located within an AD African-American GWAS locus is a key regulator for
185 clearing lipid-rich cellular debris by lipid-associated macrophages (LAMs)[8,22]. *XBP1* becomes a
186 functional TF in response to endoplasmic reticulum stress induced by unfolded proteins and is
187 modulated by the circadian rhythm[23,24]. The mic-activated *APOE* cluster included *BHLHE40*,
188 *CEBPA*, *CEBPG*, *CHD1*, *CREM*, *ETS2*, *FOS*, *FOSL2*, *FOXO1*, *FOXO3*, *JUND*, *SP3*, *SREBF1*, and *USF2*,
189 but not *IRF7*, compared to *APOE* alone (Figure 3B, Supplementary Figure 4B). *CREM* and *SREBF1*
190 are related to circadian rhythm and *SREBF1* is additionally related to lipid metabolism[25,26].
191 Upregulation of *FOXO3* and two other TFs, was shown to move weakly inflammatory microglia
192 to a strongly inflammatory state in a cell line[27]. Additionally, *FOXO3* is deactivated by *SIRT1*,
193 another gene associated with circadian rhythm[28]. The mic-inflammatory *APOE* gene cluster
194 added *CEBPA*, *E2F4*, *FOSL2*, *GABPB1*, *HDX*, *ILF2*, *IRF7*, *JUN*, *JUND*, *KMT2B*, *MAF*, *MAFG*, *NFE2L1*,
195 *NR3C1*, *TCF7L2*, and *ZBTB40* to the *APOE* TF list (Figure 3C, Supplementary Figure 4C). *E2F4* halts
196 the cell cycle and has been linked to cell quiescence, regulation of AD-related gene networks,
197 and upregulation in AD brains[29]. Macrophage/microglia activation factor (*MAF*) expression has
198 been linked to changes in the endosome/lysosome membranes[30], which suggests a concerted
199 action between *APOE* and endo-/lysosomal function in activated microglia. Taken together, the
200 CEBP, FOS, and JUN families regulate the *APOE* gene clusters across microglial cell states.
201 The *APOE* clusters within astrocytes were regulated by fewer TFs than their microglial
202 counterparts. Interestingly, astro-resting did not include either of the THR-related TFs. However
203 it included another steroid/thyroid hormone family receptor *NR2F1*[31] (Figure 3D,
204 Supplementary Figure 4D), which is associated with Bosh-Boonstra-Schaaf optic atrophy
205 (BBSOA), optic nerve atrophy featured by the loss of retinal ganglion cells critical for the light

entrainment of the circadian rhythm[32,33]. The *APOE* cluster in astro-resting also had TFs *DBP*, *CEBPB*, *ING4*, *JUND*, and *TBP* (Figure 3D, Supplementary Figure 4D). This supports the idea that the CEBP family of TFs influences *APOE* in astrocytes and microglia. As previously noted, *DBP* is related to the circadian rhythm. Astro-activated had *MAFG*, *NFIC*, *NR2F1*, *PBX1*, *POU3F2*, *RARB*, *RFX3*, and *TCF7L1* regulating the *APOE* gene cluster in addition to *THRB* which directly regulated *APOE* in this cell state (Figure 3E, Supplementary Figure 4E). *MAFG* has been implicated in astrocyte-driven inflammation linked to the chronic inflammatory disease, multiple sclerosis[34]. *TCF7L1* is important for astrocyte maturation, highlighting the central role that *APOE* plays in normal astrocyte function[35]. *RARB* is another steroid/thyroid hormone family receptor that forms heterodimers with RXR to facilitate transcriptional activation or repression. Despite the absence of the THR-related TFs in astro-resting, both *APOE* gene clusters were regulated by steroid/thyroid hormone receptors, suggesting expression regulation through thyroid hormone. Overall, we observed that circadian rhythm-related TFs were associated with regulating the *APOE* gene clusters across cell types and states. This implies that *APOE* participates in processes similar to those regulated by these TFs, even if they do not directly regulate *APOE*. Alternatively, these TFs may have a role in APOE regulation, but the pySCENIC analysis may lack the power to detect them.

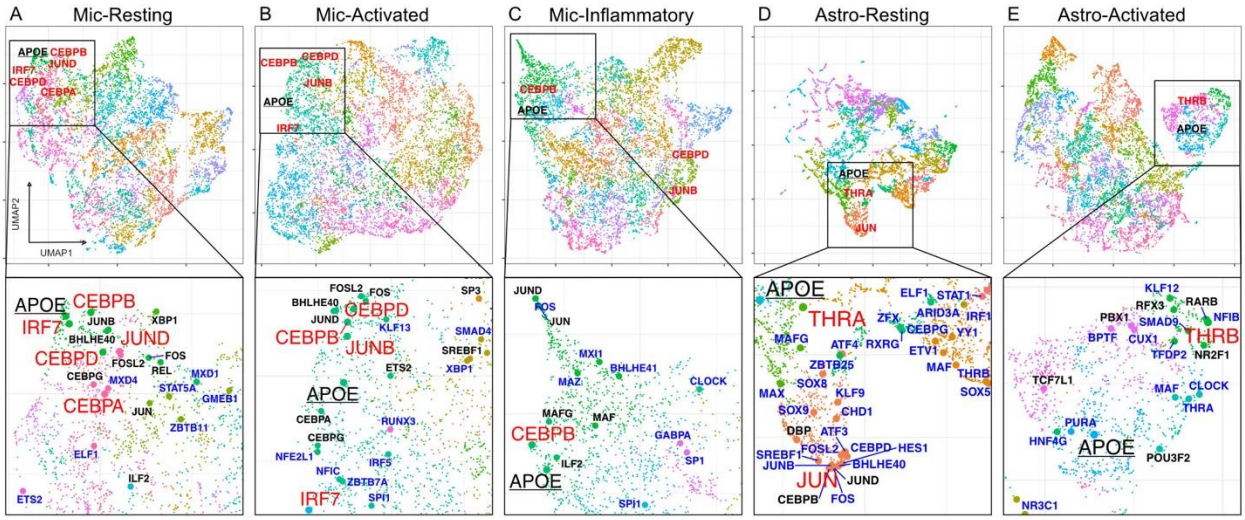



*Figure 3 UMAP representation of gene regulatory networks*
*A-E) A global UMAP visualization of each cell states' regulome along with a zoomed view highlighting APOE. Principle component analysis was performed on the TF-by-target gene matrix. The top five principal components were used to calculate the UMAP coordinates. APOE is underlined. The TFs in red directly regulate APOE. TFs in black significantly upregulate the APOE gene cluster. TFs in blue are the TFs that were replicated in the ROSMAP dataset. D) The global UMAP for Astro-resting is zoomed. The full figure can be found in Supplementary Figure 3.*

**Gene set enrichment of APOE gene clusters**

To investigate the pathways behind the genes within the *APOE* gene clusters from the co-regulatory networks, we performed an enrichment analysis using Gene Ontology (GO) biological process (BP) terms. Mic-resting showed an enrichment of pathways related to cytoplasmic translation, cytokine production, apoptotic process, integrated stress response, immunoglobulin-mediated immune response, response to lipid, IL-6 production, and response to LPS (Supplementary Figure 4F, Supplementary Table 3a). Mic-activated shared most of these pathways, but also the glycolytic process (Supplementary Figure 4F, Supplementary Table 3b), which is a metabolic feature of AD[36]. Mic-inflammatory had fewer pathways than the other two microglial states, but they included endocytosis, macroautophagy, and cotranslational protein targeting to the membrane (Supplementary Figure 4F, Supplementary Table 3c). These results imply that the genes in this cluster are related to the uptake and degradation of extracellular material and the replenishment of membrane proteins.

Astro-resting was enriched for genes in mitochondrial electron transport, NADH to ubiquinone and other aspects of the electron transport chain. It also had genes related to Golgi vesicle transport and synapse organization (Supplementary Figure 4F, Supplementary Table 3d). These results reflect the roles of astrocytes in energy metabolism and synaptic modulation. Astro-activated had many genes in its *APOE* cluster (n=409), which made the enrichment analysis less sensitive. None of the GO terms passed multiple testing correction, but the top hits (p-value < 0.01) included terms such as semaphorin-plexin signaling, axon guidance, synapse assembly, fatty-acyl-CoA biosynthetic process, glucan catabolic process, and amide metabolic process (Supplementary Table 3e). Semaphorin-plexin signaling, in conjunction with axon guidance and synapse assembly, suggests an active regulation of the surrounding synapses. Glucan catabolism indicates the cell is low on glucose and tapping into glycogen reserves. Fatty-acyl-CoA biosynthesis and amide metabolism both indicate that resources are being diverted from ATP production through cellular respiration toward the synthesis of lipids like phospholipids and cholesterol for membranes and neurotransmitters, which is consistent with previous findings in AD[37].

Overall, all microglia states were associated with immune, inflammation, and cytokine-related terms. We observed that mic-activated and astro-activated cells had genes related to known metabolic alterations in AD. We also noticed that mic-inflammatory and astro-activated cells had genes involved in endocytosis and autophagy, which are mechanisms for clearing amyloid-beta plaques and recycling other macromolecules.

**Gene set enrichment of TF target genes**

To complement the analysis of the *APOE* gene clusters, we took a deeper dive into the TFs directly regulating APOE. Specifically, we identified the target genes significantly perturbed by a virtual knockout of these TFs. As previously described, we implemented a modified version of the TF-KO method included in SCENIC+, producing a perturbed expression profile after TF-KO. We then performed a Mann-Whitney rank-sum analysis to identify differentially expressed genes between the original and TF-KO GEMs. We found 60 to 434 nominal DEGs for each TF knockout (Supplementary Table 1, Supplementary Table 4). Using the nominal DEGs, we performed a gene set enrichment analysis and visualized the results in a heatmap to better identify patterns in the regulation of biological processes.

*CEBPB*-regulated processes were common across the microglial states, including positive regulation of cytokine production, receptor-mediated endocytosis, and humoral immune response mediated by circulating immunoglobulin (Figure 4, Supplementary Table 5). These processes reflect the role of *CEBPB* in modulating the inflammatory response of microglia. However, *CEBPB* also regulated processes that were specific to certain microglial states. For example, in mic-inflammatory, *CEBPB*-regulated genes are involved in mRNA splicing, via spliceosome and cellular response to lipopolysaccharide (Figure 4, Supplementary Table 5). In contrast, in mic-activated, *CEBPB* lacked the terms related to positive regulation of protein phosphorylation, positive regulation of transferase activity, antigen processing and presentation of exogenous peptide antigen, and cellular response to cytokine stimulus (Figure 4, Supplementary Table 5).

*CEBPD* regulated shared processes in mic-resting and -activated, but had a distinct signature in mic-inflammatory. The main divergence between the primary *CEBPB* and *CEBPD* signatures was that *CEBPD* additionally regulated fat differentiation, lipoprotein biosynthesis process, and

positive regulation of small molecule metabolic process (Figure 4, Supplementary Table 5). Mic-activated *CEBPD* also regulates the lipid metabolic process. These results suggest that *CEBPD* has a more specialized role in regulating genes related to lipid processing than *CEBPB*.

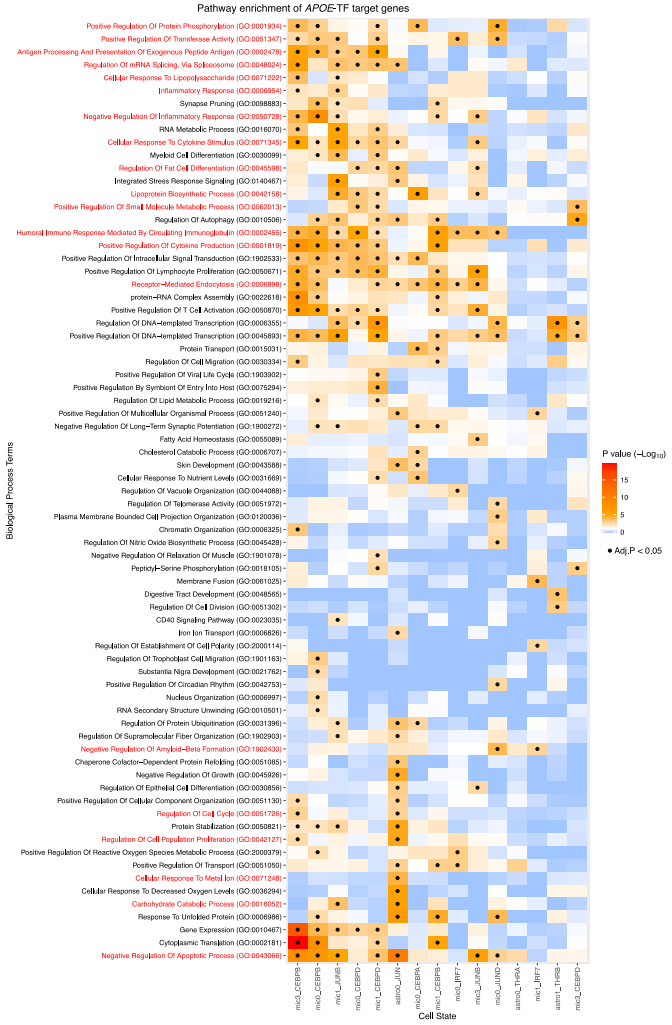



*Figure 4 Gene set enrichment analysis of the target genes for each cell-state-specific TF regulating APOE*
*Target genes for each TF were assessed for GO Biological Process terms. Terms were condensed using Revigo. Rows and columns were sorted using hierarchical clustering on Euclidean distance. Tiles with (.) indicate significance after multiple testing correction.*

*JUNB* regulated processes related to the immune response in mic-resting and mic-activated states, including negative regulation of the apoptotic process, negative regulation of inflammatory response, cellular response to cytokine stimulus, and humoral immune response mediated by circulating immunoglobulin (Figure 4, Supplementary Table 5). Within mic-activated, it regulated inflammatory response, cellular response to lipopolysaccharide, but also synapse pruning and regulation of mRNA splicing, via splicesome indicating that *JUNB* regulates

303 additional processes other than immune roles in microglia. *JUND* was the regulator in mic-
304 resting state that also regulated negative regulation of apoptotic process and humoral immune
305 response mediated by circulating immunoglobulin, but uniquely among the JUN family, it also
306 regulated negative regulation of amyloid-beta formation through *RIN1* and *RIN3*, which
307 negatively regulate *BACE1*. This implies that *JUND* may be beneficial in preventing amyloid-beta
308 accumulation in AD. Within astro-resting, *JUN* was the regulator and it had the most divergent
309 profile, regulating processes involved in negative regulation of growth, regulation of cell cycle,
310 regulation of cell population proliferation, cellular response to metal ion, and carbohydrate
311 catabolic process (Figure 4, Supplementary Table 5).
312 *THRA* target genes were not enriched for any particular GO term, but *THRB*-regulated genes
313 associated with transcription, negative regulation of inclusion body assembly, positive
314 regulation of neurogenesis, cellular response to oxidative stress, axonogenesis, and regulation
315 of cytokinesis (Figure 4, Supplementary Table 5). These processes suggest that *THRB* regulates
316 genes that prepare astrocytes for activation and inflammatory response. Overall, this alternative
317 method of examining the genes co-regulated with *APOE* revealed similar immune, lipid, and
318 energy metabolism pathways to those of the *APOE* gene clusters, strengthening the validity of
319 these biological pathways with *APOE*.
320
321 **Additional AD GWAS genes are co-regulated with *APOE***
322 The identified TFs that regulate *APOE* also regulate other genes in AD GWAS loci. Using the
323 same set of nominal DEGs identified after the TF KO, we isolated the DEGs that intersected the
324 list of genes within AD GWAS loci. Several of the genes are regulated by multiple *APOE* TFs
325 including *PGRN*, *FCGR3A*, *CTSH*, *ABCA1*, *MARCKS*, *CTSB*, *SQSTM1*, *TSC22D4*, *FCER1G*, and several
326 of the HLA genes (Figure 5A,B, Supplementary Figure 5).
327 *PGRN, CTSH,* and *CTSB* are all related to lysosome function and are co-regulated with *APOE* in
328 the microglia states. *CTSB* is specifically associated with the proteolytic processing of *APP*[38].
329 *FCER1G* and *FCGR3A* are in the *ADAMTS4* locus and encode Fc immunoglobulin receptors that
330 bind to IgE and IgG respectively. *CEBPB* regulates *FCER1G* in mic-resting and mic-inflammatory.
331 It represses AD risk genes through the Herpes simplex (HSV-1) escape strategy[39]. *FCGR3A*

exacerbates neurodegeneration and is a potential link behind the increased AD risk in chronic periodontitis patients[40,41]. It is regulated in mic-resting by *CEBPB* and *IRF7*; mic-activated by *CEBPB* and *CEBPD*; and mic-inflammatory by *CEBPB*, and *JUNB*. *ABCA1* is a transporter that transfers cholesterol and other lipids to *APOE* and there is a link between *ABCA1*, *APOE*, and Aβ levels[42]. *APOE* and *ABCA1* are co-regulated in both astrocyte and microglial states. *MARCKS* is associated with lipid rafts, dissociates from the membrane at decreased concentrations of cholesterol, causing PIP2 release, and also influences inflammation[43–45]. *SQSTM1* (p62), found in the *RASGEF1C* locus, is also found in both astrocytes and microglia and functions as a bridge between ubiquitinated proteins and autophagosomes[46]. *TSC22D4* is within the *NYAP1* locus and is regulated in astro-resting by both *APOE* TFs (*THRA* and *JUN*). *TSC22D4* participates in forming a complex that degrades *BRI2* and *BRI3*, inhibitors of Aβ production and aggregation.

Looking at the *APOE* gene clusters, we identified genes within the *ABCA7* and *SCIMP* loci across all five cell states and four out of the five states had *HLA-DRB1*, *LILRB2, SIGLEC11*, *KLF16*, *DOC2A*, and *NYAP1* (Figure 5D, Supplementary Figure 5, Supplementary Figure 6). Using *SCIMP* as an example, mic-resting had gene *PFN1* within the APOE gene cluster; mic-activated had *CXCL16* and *PFN1*; mic-inflammatory had *CXCL16*; astro-resting had *SLC25A11*; and astro-activated had *RABEP1* (Figure 5C). Interestingly, *PFN1* is related to amyotrophic lateral sclerosis (ALS), action polymerization, and binding of PIP2[47,48], which could connect it to *MARCKS* and, therefore membrane cholesterol concentrations as mentioned above. *CXCL16* induces a chemokine-induced chemotactic response, *SLC25A11* maintains the organization and morphology of the mitochondrial cristae[49], and *RABEP1* is involved in endocytic membrane fusion and recycling of endosomes, all of which connect to AD biology. Genes in AD loci *CD33* and *APH1B* were exclusively found in both astrocyte states (Figure 5D). *LACTB* (APH1B locus) was co-regulated with APOE in astro-resting (Figure 5C) and regulated mitochondrial lipid metabolism[50]. *TLN2* (*APH1B* locus) was co-regulated in astro-activated (Figure 5C) and is a component of the actin cytoskeleton linking to focal adhesion plaques, which have been shown to have higher concentrations of cholesterol than the surrounding membrane[51]. Even though the genes were not necessarily the same across the cell states, the consistency of genes within

these loci clustering with *APOE,* and many relating to lipid metabolism, suggests common regulatory mechanisms behind these loci and *APOE*.

These findings suggest that the regulatory networks involving *APOE* and the identified transcription factors extend beyond *APOE* itself and involve other genes associated with AD. Many of these AD-GWAS genes are related to critical immune, lysosome, lipid, and energy metabolism pathways highlighted by the previous TF and gene set enrichment analyses. It also supports the notion that some GWAS loci have many genes that can simultaneously influence AD risk.

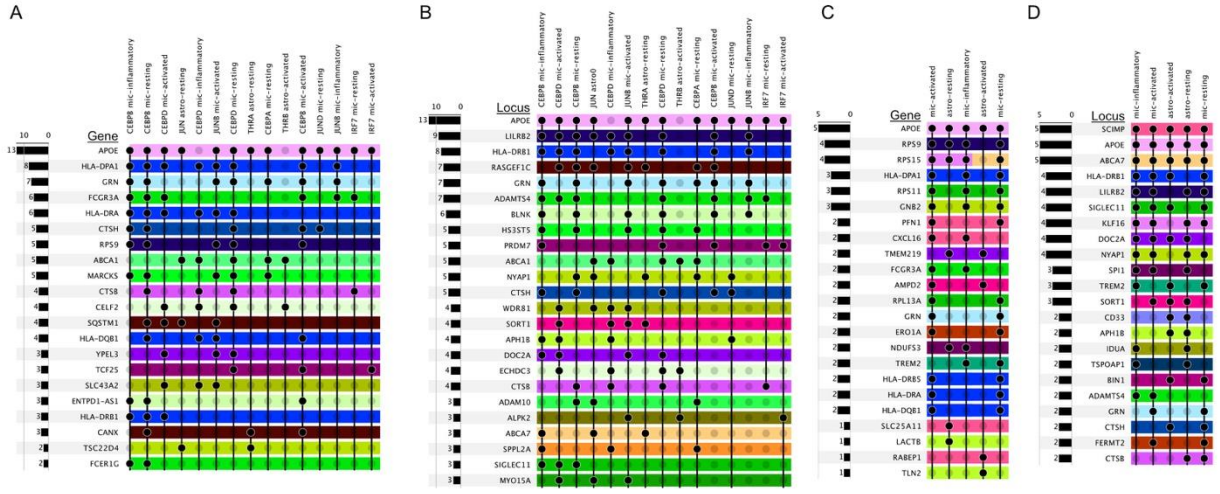



*Figure 5 Other AD GWAS genes are coregulated with APOE*
*(A,B) Genes (A) in AD GWAS loci (B) that are co-regulated by APOE-regulating TFs. Genes co-regulated by APOE-TFs were those differentially expressed (Mann-Whitney Rank U) between the original expression profiles and the perturbed values following virtual TF-KO. (C,D) Genes (C) in AD GWAS loci (D) that co-clustered with APOE in the co-regulation networks. Principle component analysis was performed on the TF-by-target gene matrix. The top five principal components were used to calculate the UMAP coordinates. Gene clusters were calculated from a shared-nearest-neighbor graph using the Louvain algorithm. Rows are colored by loci. Some genes are within two loci and therefore have two colors. These plots are truncated; the full plots can be found in Supplementary Figure 6.*

## Discussion

We used GRNs to elucidate the cell-type- and -state-specific regulation of *APOE*. These analyses revealed that *APOE* and its co-regulated genes were modulated by several transcription factor (TF) families, including CEBP, JUN, FOS, FOXO, and THR. The CEBP (A/B/D/B), JUN (/B/D), and FOS (/L2) families were involved in *APOE* regulation in all microglial states, whereas FOXO (1/3) was specific to the mic-inflammatory state. Despite the consistency of the CEBP, JUN, and FOS families across microglial states, the individual members had different strengths and associations based on cell state suggesting individual TFs could more strongly promote one state over another. The CEBP family of TFs form homo- and heterodimers with each other, so the proportions of these dimers could influence cell state. Additional experimental studies are needed to better understand the interplay between members of these families and how they regulate the activation states of microglia. Interestingly, *CEBPB* and *JUND* also regulated the *APOE* gene cluster in the astro-resting state, suggesting that these TF families play a role in *APOE* regulation in both microglia and astrocytes. *APOE* in ~~a~~stro-activated was controlled by the known *APOE* regulator *THRB*[16], and *NR2F1* and *RARB*, two other members of the steroid/thyroid hormone receptor family. We previously reported an upregulation of *THRB* in the mic-activated state compared to the other microglia states, implying that the THR family of TFs could influence *APOE* expression in microglia as well, but to a lower degree[5]. Further studies are required to confirm this conclusion.

These GRNs identified associations with immunoglobulin binding, IL1B/IL6/cytokine production, glycolysis, endocytosis, macroautophagy, mitochondrial electron transport, fatty-acyl-CoA biosynthesis, glucan catabolism, and amide metabolism. Related to immunoglobulin binding, AD GWAS associated genes *FCER1G*, *FCGR2A*, and *FCGR3A* are Fc immunoglobulin receptors that bind to IgE and IgG and were coregulated with *APOE* in microglia. Usually, IgE and IgG cannot cross the blood-brain-barrier (BBB), but it has been suggested that in *APOE*e4 carriers, the vasculature and endothelial cells making up the BBB are compromised, and albumin, antibodies, cytokines, and other inflammatory molecules more easily cross into the brain[52,53]. This might account for the association between this biology and *APOE*. It might also explain how periodontitis, a peripheral infection, could increase AD risk in association with *FCGR3A*[41].

Additionally, when microglia and astrocytes encounter these inflammatory signals, they propagate the signal by producing their own *IL1B*, *IL6*, and other cytokines, as suggested by these GRNs[54–56]. Astrocytes also boost their cholesterol production in response to cytokine signaling[53]. Within astro-activated, we saw many *APOE*-related genes pointing to fatty-acyl-CoA biosynthesis, a precursor to cholesterol. *APOE* is one of the key molecules transporting this newly synthesized cholesterol to the surrounding cells. Evidence indicates that increases in cholesteryl ester, the storage molecule for excess cholesterol, increase the accumulation of phosphorylated tau (pTau) by reducing the proteasome protein units[57,58]. Studies have also suggested that this increased cholesterol enlarges the lipid rafts, influencing many associated receptors[53,59,60]. For instance, *APP*, *BACE1*, *TREM2*, *MS4A*, *TLR4*, Fc immunoglobulin receptors, and *GLUT1* are all associated with lipid rafts[53,59–61]. Our analysis shows that other AD GWAS genes are co-regulated with APOE; some are also associated with cholesterol levels in the membrane, including *MARCKS*, *PRN1*, and *TLN2*. In neurons, *APP* is upregulated and preferentially beta-cleaved when associated with lipid rafts which is amyloidogenic[59,60]. *TREM2*, *TLR4*, *IFNGR*, and *TNFA* are all receptors that rely upon clustering as a part of their activation[53], and increased cholesterol mediates this clustering, enhancing the activation of these cells, again highlighting the connection between cytokines and the *APOE*-related genes. Astrocytes are the primary support cells of the brain and are in contact with the vasculature to facilitate the uptake and delivery of essential molecules to the neurons and other cells. *GLUT1* transports glucose into astrocytes through the endfeet, but evidence suggests that *GLUT1* is less efficient when incorporated into large lipid rafts induced by increased cholesterol[62,63]. This might explain why the mitochondrial electron transport chain is associate with astro-resting, while glucan catabolism, or the breakdown of glycogen into glucose, is linked to astro-activated state. Other studies have reported a reduction in glucose uptake in AD[36]. Interestingly, while extra cholesterol decreases glucose uptake, it increases neuronal energy demand by promoting neurotransmitter release in presynapses and the endocytosis and internalization of receptors in postsynapses[64]. This energy demand stimulates the transport of lactate and ketones from astrocytes to neurons, enabling increased neural activation by bypassing the rate-limiting step of glycolysis in energy production[36,65]. Evidence shows that in young *APOE*e4 carriers, there is an

436 increase in cognitive performance and neuronal activity, suggesting this biology is
437 occurring[36,66,67]. With age we see the negative aspects of this increased neuronal activity.
438 Neuronal activity produces reactive oxygen species (ROS), so these hyperactive neurons
439 produce increased quantities of ROS which then peroxidate the fatty acids (pFA) found in the
440 cell and organelle membranes[68]. These pFA are toxic, so the neurons expel them with the help
441 of *APOE*. Over time, ROS and pFA build up in the neuron, causing neurotoxicity, especially in
442 APOEe4 carriers with less efficient binding and, therefore, export of lipids. The expelled pFA is
443 then picked up by microglia and astrocytes and stored in lipid droplets (LD), which is consistent
444 with our finding GO terms, TFs (*MAF* and *CEBPB*), and GWAS genes (*RABEP1*) related to
445 endocytosis. This is also concordant with our finding that *BHLHE40*, a regulator for the clearance
446 of lipid-rich cellular debris, is an *APOE* regulator. Excess cholesterol produced by astrocytes also
447 collects in the lysosome, where it causes lysosomal dysfunction, leading to the buildup of
448 autophagosomes and mitochondrial dysfunction[36]. This is consistent with many of the co-
449 regulated GWAS genes we identified, including *PGRN*, *CTSH*, *CTSB* (lysosome), *SQSTM1*
450 (autophagosomes), *SLC25A11*, and *LACTB* (mitochondria metabolism).
451 Our findings also show that many TFs regulating *APOE* and the *APOE* gene clusters are related to
452 circadian rhythm, including *BHLHE40*, *DBP*, *XBP1*, *CREM*, *SREBF1*, *FOXO3*, and *NR2F1*. This is
453 consistent with many studies that have reported an association between AD and disrupted
454 sleep due to circadian rhythm dysfunction[33,66,69–71]. These findings highlight the importance of
455 the circadian rhythm in regulating APOE expression and function. Interestingly, *NR2F1* is
456 associated with BBSOA, a neurodevelopmental disease characterized by the loss of retinal
457 ganglion cells[32]. Another study with *APOE*$^{-/-}$ reported the loss of melanopsin-expressing ganglion
458 cells in the suprachiasmic nucleus, resulting in the loss of light entrainment[33], the process by
459 which light synchronizes the biological clock with the environment. These studies suggest that
460 *NR2F1* and *APOE* are connected and involved in the circadian rhythm, as indicated by our GRNs.
461 Furthermore, the disruption of energy metabolism and the suggested increase in ketone
462 metabolism are modulated by *AMPK*[72]. *AMPK* also degrades the circadian proteins *Per* and *Cry*,
463 which are produced by the transcriptional activation of *Clock* and *Bmal1*[73]. *Per* and *Cry* form a
464 complex with *CK1E/D* and inhibit the transcription of *Clock*, *Bmal1*, *Per*, and *Cry*, creating a

negative feedback loop. The degradation of *Per* and *Cry* by *AMPK* breaks this loop and results in the overexpression of *Clock* and *Bmal1*, as well as their downstream targets, such as *NR1D1*, which encodes the Rev-ERBα protein. *NR1D1* represses *Cx3cr1*, a marker of homeostatic microglia, thus activating microglia and influencing lipid droplet formation[74,75]. Disruption of the circadian rhythm is associated with impaired function and integrity of the endothelial and vascular smooth muscle cells, causing inflammation and passing these inflammatory markers through a weakened BBB[73]. It is also related to energy metabolism during the fasting stage of sleep[26]. Overall, our results reemphasize the connection of AD to circadian rhythm and suggest that *APOE* and lipid biology are key contributors.

In general, the TF families primarily associated with microglia suggest a relationship between immunity and *APOE*, while the TF family associated with astrocytes suggests a link between *APOE* and lipid metabolism. This is consistent with the primary function of each of these cell types. In astrocytes, the evidence points to the primary purpose of *APOE* in lipid metabolism is to shuttle cholesterol and other lipids synthesized by the astrocytes to other cells. The primary purpose of *APOE* in immunity is not quite as clear in microglia. A promising possibility is the secretion of extracellular vesicles (EVs) that are induced upon microglial activation. The biology related to these EVs is closely tied with that of the endosomal-lysosomal system[76], highlighted in our GRNs. These EVs can contain RNA, protein, lipids, and cytokines for extracellular communication[77]. A study in *Apoe*$^{-/-}$ mice on a high-fat diet highlighted the role of *Apoe* in influencing the miRNA composition of EVs secreted by these macrophages and their impact on atherosclerosis[78]. Although most studies on EVs interrogate their miRNA composition, other studies suggest differences in lipid composition based on microglia state as well[79], including increased secretion of cholesterol in EVs in response to increased free cholesterol[80]. *APOE* could play a role in the transfer of lipids to these EVs.

A limitation of this study is that pySCENIC utilizes only a single modality, co-expression, to identify relationships between TFs and target genes, unlike the newer SCENIC+, which also incorporates chromatin accessibility. Despite this, pySCENIC has successfully identified biologically relevant TF-target interactions. We presume that TFs regulating the *APOE* gene clusters within the co-regulatory networks influence APOE, even in the absence of direct TF-

494 *APOE* links in the pySCENIC results. Nevertheless, the findings underscore many known features
495 of AD and establish a link to *APOE* biology. Moreover, the discovery cohort includes carriers of
496 *APOE*, *TREM2*, *APP*, and *PSEN1* variants, all of which could influence the GRNs in ways not
497 accounted for. Additionally, the cohort's European ancestry may restrict the generalizability of
498 these results to other, less represented populations.

499 In conclusion, the TFs and pathways associated with *APOE* all support known AD pathology and
500 support the building evidence implicating lipids as a key driver of AD-related pathology.
501

# Methods

## Single nucleus RNA-seq acquisition and integration

Human snRNA-seq data of the parietal lobe from the Knight ADRC and DIAN was gathered at NIAGADs (accession number NG00108) and upon request from DIAN. These data include a total of 67 neuropathological controls, sporadic AD, and autosomal dominant AD participants. An independent human snRNA-seq dataset from ROSMAP (dorsolateral prefrontal cortex) was collected from Synapse (synapse ID syn21125841). The ROSMAP data has 11 sporadic AD, 11 *TREM2* R62H, and 10 control participants. The microglia from the ROSMAP dataset was previously integrated with the microglia from the Knight and DIAN data[5]. The ROSMAP astrocytes were integrated as a part of this study using the same process described by Brase, *et al*. Briefly, we isolated and normalized the astrocytes with Seurat's (v.4.3.0) *SCTransform*, setting "return.only.var.genes" to FALSE and regressing out "nCount_RNA" and "nFeature_RNA". We integrated the astrocytes with 3000 features in *SelectIntegrationFeatures*, *PrepSCTintegration*, our data as reference in *FindIntegrationAnchors*, and *IntegrateData*. We clustered integrated data with ten principal components in *FindNeighbors* and resolution 15 in *FindClusters*. We gave each cluster an 'original' identity by isolating our nuclei from clusters and finding the most common original ID. We transferred this ID to ROSMAP nuclei like a k-nearest neighbor classifier. We mapped cluster identities to pre-integrated normalized ROSMAP data.

## Gene regulatory network prediction by pySCENIC

We created cell-state-specific GRNs using the python implementation of the SCENIC[81] analysis method called pySCENIC[7] (version 0.12.1). We isolated the cell type of interest and filtered the genes to those that were expressed in ≥5% of the cell type population. Then we isolated the cell expression states of interest. We ran pySCENIC on each individual cell state 100 times (different seed for each run) as suggested by pySCENIC. We employed default values for all parameters and provided default reference data downloaded from https://resources.aertslab.org/cistarget/: Database (hg38_500bp_up_100bp_down_full_tx_v10_clust.genes_vs_motifs.rankings, hg38_10kbp_up_10kbp_down_full_tx_v10_clust.genes_vs_motifs.rankings), table (motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl), and transcription factor list (allTFs_hg38.txt). An

additional TF list was downloaded from http://humantfs.ccbr.utoronto.ca/download/v_1.01/TF_names_v_1.01.txt and the two lists were merged totaling 2093 TFs.

After the 100 runs were complete, we identified the transcription factors (TFs) that were present in >= 80% of the runs. We then ran pySCENIC an additional 100 times using only the >=80% TFs labeled as transcription factors. The other TFs were retained in the expression matrix, but not labelled as TFs. The TFs that appeared in >= 80% of the second round of runs were isolated as reproducible regulators.

The same resource files and parameters were used in the ROSMAP cohort replication. We filtered the genes to those intersecting those from the Knight ADRC microglia and astrocyte runs. We ran pySCENIC 100 times using the filtered list of TFs for each respective cell state (list used in second round of discovery runs). We isolated the TFs that were present in >= 50% of the runs. We intersected the final lists of discovery TFs with the lists of replication TFs to identify a set of replicable TFs for each cell state.

**Co-regulation network clusters**

Using the TF-target gene matrices for cell state, we identified co-regulation network clusters and visualized them in the UMAP space. The TF-target gene matrix had each of the replicable TFs as features, the target genes as elements, and the values were the number of times the target gene was regulated by the TF out of the 100 pySCENIC runs. These values were normalized by the number of times the TF was included in the pySCENIC networks. For example, if TF_1 was in the pySCENIC networks 85 times, all the values would be multiplied by (100/85) to get the normalized value. We then log transformed the values and ran them through a principal component analysis using *prcomp,* a native R (v4.2.2) function. We then passed the top five principal components (PCs) through the *umap* function from the umap library (v0.2.10.0) to create a two-dimensional representation of the space. The top five PCs were also passed to the *makeSNNGraph* function from the bluster package (v1.8.0) to create a shared nearest neighbor graph. The resulting graph was passed to the *cluster_louvain* function from the igraph package (v1.3.5) to identify clusters of similarly regulated target genes using a resolution of 2. All target

genes in the same cluster as APOE were considered co-regulated with APOE and were evaluated for additional AD GWAS genes and for gene set enrichment (Supplementary Table 6).

**TF regulators for gene clusters**

We identified the TFs significantly upregulating the genes in each gene cluster identified in the gene UMAPs. We utilized the normalized, but not log transformed, TF-target gene matrix described in the Gene UMAP clusters section. For each gene cluster at a time, we created a binary vector where 1 indicated the target gene was in the cluster. We then looped through each TF and modelled the TF-target gene values using the binary vector as a covariate. The native *glm* R function (v4.2.2) was used to train the model with a quasipoisson distribution and log link function. The hits for each cluster were then multiple testing corrected using the Benjamini-Hochberg (BH) method. TFs with a BH corrected p-value less than 0.05 and an estimate greater than zero were considered significant regulators for the gene cluster.

**In silico transcription factor knock-out**

SCENIC+[17] was recently released with the functionality to knockout a TF and determine the shift in the expression space. We implemented the algorithm proposed in SCENIC+ (v1.0.0, python v3.8.18), which was an adaption from the original method in CellOracle[82]. In summary, this method uses the cell by gene expression matrix to train prediction models for each gene. These models allow for the virtual knockout of a TF (TF-KO) by setting the expression of the TF to zero and then using the trained models to predict the shift in target gene expression. The expression shift is then transformed back onto the reduction space and depicted by arrows.

By default, SCENIC+ used the GradiantBoostingRegressor from sklearn (v1.3.2) with a learning rate of 0.01, 500 estimators, and max features set to 0.1 to train prediction models. We included the TFs that were regulating a target gene at least 5% of the time in the model that used TF expression to predict the target gene expression. To correct for the influence of sample proportion on each cell state, we also included the log transformed nuclei counts for the sample in the prediction model. SCENIC+ uses a default of five iterations to propagate the shift across the expression matrix. We extended SCENIC+'s implementation to work with the UMAP space rather than just the PCA space. We also switched the arrows to be drawn by the quiver function

as was found in the CellOracle implementation rather than the streamplot function used by SCENIC+.

To extend this method to the UMAP space, we trained a neural network to predict a cell's UMAP coordinates using as input the expression levels of all genes (python v3.11.0). We used the original gene expression matrix (GEM) and UMAP representation for the cell types which included all expression states not just the resting and activated states passed to pySCENIC. First, we log transformed the GEM. We then used the *train_test_split* function from sklearn (v1.3.2) to split the data into training and testing data (80% training, 20% testing) and stratified by cell state. We then split the training data into training and validation sets (90% training, 10% validation) and once again stratified by cell state. Using the *StandardScaler* function from sklearn (v1.3.2) we fit and transformed a scaler using the training GEM. This scaler was then used to transform the validation and test GEMs. The sequential neural network was constructed using tensorflow (v2.14.0) and consisted of three layers: two dense layers with the ReLU activation function and an output layer with the same dimensionality as the UMAP data (n=2). The model was compiled with the Adam optimizer and mean squared error loss function. Early stopping was implemented to prevent overfitting with monitoring set to 'val_loss' and a patience of 10. The model was then trained on the training data for a maximum of 1000 epochs. The trained model was evaluated on the test GEM. This trained model was used on the TF-KO perturbed matrices to transform the expression shift to the UMAP reduction space. This method was incorporated into a snakemake pipeline so that TF-KO could be performed consistently across TFs, cell states, and cell types.

**TF knockout differential expression**

We performed a differential expression analysis between the original GEM and perturbed GEM calculated after TF-KO. The original GEM and perturbed GEM were both log transformed. We then filtered out the cells that had zero expression of the TF in the original matrix as these cells have no changes in expression and would skew the results. We then performed a Mann-Whitney rank U test using the *mannwhitneyu* function from scipy (v1.2.2, python v3.11.0) to identify differentially expressed genes. We performed FDR multiple testing correction. Genes

with a corrected p-value less than 0.05 were considered co-regulated with APOE and were evaluated for additional AD GWAS genes and for gene set enrichment (Supplementary Table 6).

**Gene set enrichment of genes co-regulated with APOE**

Sets of co-regulated genes where run through enrichment analysis to better understand the related function. There were two sources of co-regulated genes, co-regulation network clusters and TF-targets, and they were analyzed independently. The gene sets identified by gene clusters were run through enrichment analysis using the *enrichr* function from the enrichR package (v3.1) in R (v4.2.2) and the TF-targets were run through the *enrichr* function from the gseapy package (v0.9.5) in python (v3.11.0). Both used the 2023 GO Biological Process (BP) terms and significant associations were those with adjusted p-values less than 0.05.

The terms from the TF-targets analysis were condensed using rrvgo (v1.6.0), an R package that implements the Revigo[83] tool for summarizing GO BP terms. The function *calculateSimMatrix* was used to calculate the relationships between the GO BP terms with variable inputs: orgdb = “org.Hs.eg.db”, ont = “BP”, method = “Rel”. The terms were then summarized using *reduceSimMatrix* and the following variables: score = “Rank”, threshold = .7, orgdb = “org.Hs.eg.db”. The summarized terms or “parentTerms” and their P values were then used to make a heatmap. Rows and columns were ordered using dist function method = “manhattan”.

**Identifying other AD GWAS genes co-regulated APOE**

There were two sources of co-regulated genes, co-regulation network clusters and TF-targets, and they were analyzed independently. These gene sets were compared against a list of genes within GWAS loci (+/- 500KB of the lead SNP, Supplementary Table 7) and can be found in Supplementary Table 6. The Loci for each source were identified and merged. This merged list was used to consistently color the genes and loci for each row in the upset plots in Figure 5.

## Data Availability

The single nucleus data from the Knight ADRC accessed in this study are found in the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) with accession number NG00108 [https://www.niagads.org/datasets/ng00108]. The raw single nucleus data

from the DIAN brain bank accessed in this study are available under restricted access to maintain individual and family confidentiality. These samples contain rare disease-causing variants that could be used to identify the participating individuals and families. Access can be obtained by request through the online resource request system on the DIAN Website: https://dian.wustl.edu/our-research/for-investigators/dian-observational-study-investigator-resources/. The ROSMAP single nucleus RNA sequencing data used in this study are available at Synapse under Synapse ID syn21125841 [https://www.synapse.org/#!Synapse:syn21125841/wiki/597278]. The pySCENIC default reference data was downloaded from https://resources.aertslab.org/cistarget/: Database [https://resources.aertslab.org/cistarget/databases/homo_sapiens/hg38/refseq_r80/mc_v10_clust/gene_based/hg38_500bp_up_100bp_down_full_tx_v10_clust.genes_vs_motifs.rankings.feather], [https://resources.aertslab.org/cistarget/databases/homo_sapiens/hg38/refseq_r80/mc_v10_clust/gene_based/hg38_10kbp_up_10kbp_down_full_tx_v10_clust.genes_vs_motifs.rankings.feather], table [https://resources.aertslab.org/cistarget/motif2tf/motifs-v10nr_clust-nr.hgnc-m0.001-o0.0.tbl], and transcription factor list [https://resources.aertslab.org/cistarget/tf_lists/allTFs_hg38.txt]. An additional TF list was downloaded from [http://humantfs.ccbr.utoronto.ca/download/v_1.01/TF_names_v_1.01.txt]. Source data are provided with this paper.

## Code Availability

Custom code used to analyze the snRNA-seq data and datasets generated and/or analyzed in the current study are available from the corresponding authors upon request or at https://github.com/HarariLab/APOE-GRN.

# References

1. Ulrich, J. D. *et al.* ApoE facilitates the microglial response to amyloid plaque pathology. *J. Exp. Med.* **215**, 1047–1058 (2018).
2. Singh, D. Astrocytic and microglial cells as the modulators of neuroinflammation in Alzheimer's disease. *J. Neuroinflammation* **19**, 206 (2022).
3. Eskandari-Sedighi, G. & Blurton-Jones, M. Microglial APOE4: more is less and less is more. *Mol. Neurodegener.* **18**, 99 (2023).
4. Fernandez, C. G., Hamby, M. E., McReynolds, M. L. & Ray, W. J. The Role of APOE4 in Disrupting the Homeostatic Functions of Astrocytes and Microglia in Aging and Alzheimer's Disease. *Front. Aging Neurosci.* **11**, (2019).
5. Brase, L. *et al.* Single-nucleus RNA-sequencing of autosomal dominant Alzheimer disease and risk variant carriers. *Nat. Commun.* **14**, 2314 (2023).
6. Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).
7. Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
8. Kunkle, B. W. *et al.* Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA Neurol.* **78**, 102–113 (2021).
9. Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* **53**, 392–402 (2021).
10. Lekstrom-Himes, J. & Xanthopoulos, K. G. Biological Role of the CCAAT/Enhancer-binding Protein Family of Transcription Factors*. *J. Biol. Chem.* **273**, 28545–28548 (1998).
11. Gao, T. *et al.* Transcriptional regulation of homeostatic and disease-associated-microglial genes by IRF1, LXRβ and CEBPα. *Glia* **67**, 1958–1975 (2019).
12. Xia, Y. *et al.* C/EBPβ is a key transcription factor for APOE and preferentially mediates ApoE4 expression in Alzheimer's disease. *Mol. Psychiatry* **26**, 6002–6022 (2021).

13. Ejarque-Ortiz, A. *et al.* CCAAT/enhancer binding protein delta in microglial activation. *J. Neurosci. Res.* **88**, 1113–1123 (2010).
14. Li, Z. *et al.* Interferon Regulatory Factor 7 Promoted Glioblastoma Progression and Stemness by Modulating IL-6 Expression in Microglia. *J. Cancer* **8**, 207–219 (2017).
15. Angel, P. & Schorpp-Kistner, M. Jun/Fos. in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* 928–935 (Springer, Berlin, Heidelberg, 2006). doi:10.1007/3-540-29623-9_4560.
16. Roman, C. *et al.* Thyroid hormones upregulate apolipoprotein E gene expression in astrocytes. *Biochem. Biophys. Res. Commun.* **468**, 190–195 (2015).
17. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
18. Honma, S. *et al.* Dec1 and Dec2 are regulators of the mammalian molecular clock. *Nature* **419**, 841–844 (2002).
19. Kawamoto, T. *et al.* A novel autofeedback loop of Dec1 transcription involved in circadian rhythm regulation. *Biochem. Biophys. Res. Commun.* **313**, 117–124 (2004).
20. Cho, Y. *et al.* The Basic Helix-Loop-Helix Proteins Differentiated Embryo Chondrocyte (DEC) 1 and DEC2 Function as Corepressors of Retinoid X Receptors. *Mol. Pharmacol.* **76**, 1360–1369 (2009).
21. Yoshitane, H. *et al.* Functional D-box sequences reset the circadian clock and drive mRNA rhythms. *Commun. Biol.* **2**, 1–10 (2019).
22. Podlesny-Drabiniok, A. *et al.* BHLHE40/41 regulate macrophage/microglia responses associated with Alzheimer's disease and other disorders of lipid-rich tissues. 2023.02.13.528372 Preprint at https://doi.org/10.1101/2023.02.13.528372 (2023).
23. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001).
24. Pan, Y. *et al.* 12-h clock regulation of genetic information flow by XBP1s. *PLOS Biol.* **18**, e3000580 (2020).

25. Foulkes, N. S., Borjigin, J., Snyder, S. H. & Sassone-Corsi, P. Transcriptional control of circadian hormone synthesis via the CREM feedback loop. *Proc. Natl. Acad. Sci.* **93**, 14140–14145 (1996).
26. Lech, K. *et al.* Dissecting Daily and Circadian Expression Rhythms of Clock-Controlled Genes in Human Blood. *J. Biol. Rhythms* **31**, 68–81 (2016).
27. Sun, N. *et al.* Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386-4403.e29 (2023).
28. Yan, D. *et al.* SIRT1/FOXO3-mediated autophagy signaling involved in manganese-induced neuroinflammation in microglia. *Ecotoxicol. Environ. Saf.* **256**, 114872 (2023).
29. López-Sánchez, N., Ramón-Landreau, M., Trujillo, C., Garrido-García, A. & Frade, J. M. A Mutant Variant of E2F4 Triggers Multifactorial Therapeutic Effects in 5xFAD Mice. *Mol. Neurobiol.* **59**, 3016–3039 (2022).
30. Bräuer, A. U., Nitsch, R. & Savaskan, N. E. Identification of macrophage/microglia activation factor (MAF) associated with late endosomes/lysosomes in microglial cells. *FEBS Lett.* **563**, 41–48 (2004).
31. Tocco, C., Bertacchi, M. & Studer, M. Structural and Functional Aspects of the Neurodevelopmental Gene NR2F1: From Animal Models to Human Pathology. *Front. Mol. Neurosci.* **14**, (2021).
32. Bertacchi, M. *et al.* Mouse Nr2f1 haploinsufficiency unveils new pathological mechanisms of a human optic atrophy syndrome. *EMBO Mol. Med.* **11**, e10291 (2019).
33. Zhou, L. *et al.* Degeneration and energy shortage in the suprachiasmatic nucleus underlies the circadian rhythm disturbance in ApoE−/− mice: implications for Alzheimer's disease. *Sci. Rep.* **6**, 36335 (2016).
34. Wheeler, M. A. *et al.* MAFG-driven astrocytes promote CNS inflammation. *Nature* **578**, 593–599 (2020).
35. Szewczyk, L. M. *et al.* Astrocytic β-catenin signaling via TCF7L2 regulates synapse development and social behavior. *Mol. Psychiatry* 1–17 (2023) doi:10.1038/s41380-023-02281-y.

36. Lee, H. *et al.* ApoE4-dependent lysosomal cholesterol accumulation impairs mitochondrial homeostasis and oxidative phosphorylation in human astrocytes. *Cell Rep.* **42**, 113183 (2023).
37. Williams, H. C. *et al.* APOE alters glucose flux through central carbon pathways in astrocytes. *Neurobiol. Dis.* **136**, 104742 (2020).
38. Drobny, A. *et al.* The role of lysosomal cathepsins in neurodegeneration: Mechanistic insights, diagnostic potential and therapeutic approaches. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1869**, 119243 (2022).
39. Fu, L. *et al.* Enhanced expression of FCER1G predicts positive prognosis in multiple myeloma. *J. Cancer* **11**, 1182–1194 (2020).
40. Fuller, J. P., Stavenhagen, J. B. & Teeling, J. L. New roles for Fc receptors in neurodegeneration-the impact on Immunotherapy for Alzheimer's Disease. *Front. Neurosci.* **8**, 235 (2014).
41. Jin, J. *et al*. Shared Molecular Mechanisms between Alzheimer's Disease and Periodontitis Revealed by Transcriptomic Analysis. *BioMed Res. Int.* **2021**, 6633563 (2021).
42. Lanfranco, M. F., Ng, C. A. & Rebeck, G. W. ApoE Lipidation as a Therapeutic Target in Alzheimer's Disease. *Int. J. Mol. Sci.* **21**, 6336 (2020).
43. Martín, M. G., Pfrieger, F. & Dotti, C. G. Cholesterol in brain disease: sometimes determinant and frequently implicated. *EMBO Rep.* **15**, 1036–1052 (2014).
44. Iyer, D. N. *et al.* Pathophysiological roles of myristoylated alanine-rich C-kinase substrate (MARCKS) in hematological malignancies. *Biomark. Res.* **9**, 34 (2021).
45. Issara-Amphorn, J. *et al.* Myristoylated, alanine-rich C-kinase substrate (MARCKS) regulates toll-like receptor 4 signaling in macrophages. *Sci. Rep.* **13**, 19562 (2023).
46. Turco, E. *et al.* Reconstitution defines the roles of p62, NBR1 and TAX1BP1 in ubiquitin condensate formation and autophagy initiation. *Nat. Commun.* **12**, 5212 (2021).
47. Wu, C.-H. *et al.* Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature* **488**, 499–503 (2012).

784 48. Frantzi, M. *et al.* Silencing of Profilin-1 suppresses cell adhesion and tumor growth via
785 predicted alterations in integrin and Ca 2+ signaling in T24M-based bladder cancer models.
786 *Oncotarget* **7**, 70750–70768 (2016).
787 49. Gallo, M. *et al.* MISC-1/OGC links mitochondrial metabolism, apoptosis and insulin
788 secretion. *PloS One* **6**, e17827 (2011).
789 50. Keckesova, Z. *et al.* LACTB is a tumour suppressor that modulates lipid metabolism and cell
790 state. *Nature* **543**, 681–686 (2017).
791 51. Tachibana, H. *et al.* The plasma membrane of focal adhesions has a high content of
792 cholesterol and phosphatidylcholine with saturated acyl chains. *J. Cell Sci.* **136**, jcs260763
793 (2023).
794 52. Liu, C.-C. *et al.* Peripheral apoE4 enhances Alzheimer's pathology and impairs cognition by
795 compromising cerebrovascular function. *Nat. Neurosci.* **25**, 1020–1033 (2022).
796 53. Wang, H. *et al. Regulation of Neuroinflammation by Astrocyte-Derived Cholesterol*.
797 http://biorxiv.org/lookup/doi/10.1101/2022.12.12.520161 (2022)
798 doi:10.1101/2022.12.12.520161.
799 54. Hanisch, U.-K. Microglia as a source and target of cytokines. *Glia* **40**, 140–155 (2002).
800 55. Choi, S. S., Lee, H. J., Lim, I., Satoh, J. & Kim, S. U. Human Astrocytes: Secretome Profiles of
801 Cytokines and Chemokines. *PLOS ONE* **9**, e92325 (2014).
802 56. Sofroniew, M. V. Multiple Roles for Astrocytes as Effectors of Cytokines and Inflammatory
803 Mediators. *The Neuroscientist* **20**, 160–172 (2014).
804 57. van der Kant, R. *et al.* Cholesterol Metabolism Is a Druggable Axis that Independently
805 Regulates Tau and Amyloid-β in iPSC-Derived Alzheimer's Disease Neurons. *Cell Stem Cell*
806 **24**, 363-375.e9 (2019).
807 58. Wang, H. *et al.* Regulation of beta-amyloid production in neurons by astrocyte-derived
808 cholesterol. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2102191118 (2021).
809 59. Lee, S.-I. *et al.* APOE4-carrying human astrocytes oversupply cholesterol to promote
810 neuronal lipid raft expansion and Aβ generation. *Stem Cell Rep.* **16**, 2128–2137 (2021).

60. Ehehalt, R., Keller, P., Haass, C., Thiele, C. & Simons, K. Amyloidogenic processing of the Alzheimer β-amyloid precursor protein depends on lipid rafts. *J. Cell Biol.* **160**, 113–123 (2003).
61. Mattiola, I., Mantovani, A. & Locati, M. The tetraspan MS4A family in homeostasis, immunity, and disease. *Trends Immunol.* **42**, 764–781 (2021).
62. Koepsell, H. Glucose transporters in brain in health and disease. *Pflüg. Arch. - Eur. J. Physiol.* **472**, 1299–1343 (2020).
63. Yan, Q. *et al.* Mechanistic insights into GLUT1 activation and clustering revealed by super-resolution imaging. *Proc. Natl. Acad. Sci.* **115**, 7033–7038 (2018).
64. Li, D., Zhang, J. & Liu, Q. Brain cell type-specific cholesterol metabolism and implications for learning and memory. *Trends Neurosci.* **45**, 401–414 (2022).
65. Alessandri, B., Gugliotta, M., Levasseur, J. E. & Bullock, M. R. Lactate and glucose as energy substrates and their role in traumatic brain injury and therapy. *Future Neurol.* **4**, 209–228 (2009).
66. Graybeal, J. J. *et al.* Human ApoE ε4 Alters Circadian Rhythm Activity, IL-1β, and GFAP in CRND8 Mice. *J. Alzheimers Dis.* **43**, 823–834 (2014).
67. Wu, L., Zhang, X. & Zhao, L. Human ApoE Isoforms Differentially Modulate Brain Glucose and Ketone Body Metabolism: Implications for Alzheimer's Disease Risk Reduction and Early Intervention. *J. Neurosci.* **38**, 6665–6681 (2018).
68. Ioannou, M. S. *et al.* Neuron-Astrocyte Metabolic Coupling Protects against Activity-Induced Fatty Acid Toxicity. *Cell* **177**, 1522-1535.e14 (2019).
69. Ke, P. *et al.* Relationship between circadian genes and memory impairment caused by sleep deprivation. *PeerJ* **10**, e13165 (2022).
70. Niu, L. *et al.* Chronic sleep deprivation altered the expression of circadian clock genes and aggravated Alzheimer's disease neuropathology. *Brain Pathol.* **32**, e13028 (2022).
71. Ni, J. *et al.* An impaired intrinsic microglial clock system induces neuroinflammatory alterations in the early stage of amyloid precursor protein knock-in mouse brain. *J. Neuroinflammation* **16**, 173 (2019).

72. Guzmán, M. & Blázquez, C. Ketone body synthesis in the brain: possible neuroprotective effects. *Prostaglandins Leukot. Essent. Fatty Acids* **70**, 287–292 (2004).
73. Zhang, Z. *et al.* Circadian rhythm and atherosclerosis (Review). *Exp. Ther. Med.* **20**, 96 (2020).
74. Lam, M. T. Y. *et al.* Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**, 511–515 (2013).
75. Lee, J. *et al.* Microglial REV-ERBα regulates inflammation and lipid droplet formation to drive tauopathy in male mice. *Nat. Commun.* **14**, 5197 (2023).
76. Eitan, E., Suire, C., Zhang, S. & Mattson, M. P. Impact of lysosome status on extracellular vesicle content and release. *Ageing Res. Rev.* **32**, 65–74 (2016).
77. Nguyen, M.-A. *et al.* Extracellular Vesicles Secreted by Atherogenic Macrophages Transfer MicroRNA to Inhibit Cell Migration. *Arterioscler. Thromb. Vasc. Biol.* **38**, 49–63 (2018).
78. Bouchareychas, L. *et al.* Macrophage Exosomes Resolve Atherosclerosis by Regulating Hematopoiesis and Inflammation via MicroRNA Cargo. *Cell Rep.* **32**, 107881 (2020).
79. Gabrielli, M., Raffaele, S., Fumagalli, M. & Verderio, C. The multiple faces of extracellular vesicles released by microglia: Where are we 10 years after? *Front. Cell. Neurosci.* **16**, (2022).
80. Strauss, K. *et al.* Exosome Secretion Ameliorates Lysosomal Storage of Cholesterol in Niemann-Pick Type C Disease*. *J. Biol. Chem.* **285**, 26279–26288 (2010).
81. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
82. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).
83. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).

## Acknowledgments

Doctors Bruno A. Benitez, and Oscar Harari contributed equally to this work as co-senior authors. Data sharing for this project were 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. DIAN Study investigators have reviewed this manuscript for scientific content and consistency of data interpretation with previous DIAN Study publications. We acknowledge the altruism of the participants and their families and the contributions of the DIAN research and support staff at each of the participating sites for their contributions to this study.

The results published here are partly based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). 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.

This work was possible thanks to the following governmental grants from the National Institute of Health: NIA R01AG057777 (OH), R56AG067764 (OH), U01AG072464 (OH), R01AG074012 (OH), U19AG032438 (RJB), NINDS R01NS118146 (BAB), R21NS127211 (BAB), NIA T32AG058518 (LB), O.H. is an Archer Foundation Research Scientist.

This work was supported by access to equipment made possible by the Departments of Neurology and Psychiatry at Washington University School of Medicine.

The funders of the study had no role in the collection, analysis, or interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

Figure 1 was created with BioRender.com. Bing Copilot was used to improve grammar and flow of text.

## Author Contributions

L.B. performed the quality control, bioinformatic analyses, interpreted the results, created the figures, and wrote the manuscript. Y.Y. ran the pySCENIC analyses. E.M. was involved in the DIAN sample collection, profiling and provided a critical review of the manuscript. O.H. and B.A.B. were involved in experiment design, analysis, data interpretation, intellectual contribution, and manuscript writing and review.

## Competing Interests

E.M. receives Grant Funding from NIA; Institutional funding from Eli Lilly, Hoffmann-La Roche, Eisai. He is a DSMB member (paid directly) for Alector; Eli Lilly; a Scientific Advisory Board Member (paid directly) for Alzamend, Fondation Alzheimer. He acts as a Consultant/Advisor for Sage Therapeutics, Eli Lilly, Sanofi, AstraZeneca, Hoffmann La-Roche. The other authors have no conflicts of interest to disclose.

## Supplementary Information

Supplementary Information - .pdf file