Title: Gene expression and chromatin accessibility comparison in iPSC-derived microglia in African, European, and Amerindian genomes in Alzheimer's patients and controls.

Authors:

Sofia Moura[1], Luciana Bertholim Nasciben[1], Aura M. Ramirez[1], Lauren Coombs[1], Joe Rivero[1], Derek J. Van Booven[1], Brooke A. DeRosa[1], Kara L. Hamilton-Nelson[1], Patrice L. Whitehead[1], Larry D. Adams[1], Takiyah D. Starks[2], Pedro R. Mena[1], Maryenela Illanes-Manrique[3,4,], Sergio Tejada[1], Goldie S. Byrd[2], Mario R. Cornejo-Olivas[3,4], Briseida E. Feliciano-Astacio[5], Karen Nuytemans[1,6], Liyong Wang[1,6], Margaret A. Pericak-Vance[1,6], Derek M. Dykxhoorn[1,6], Farid Rajabli[1,6,] Anthony J. Griswold[1,6], Juan I. Young[1,6], Jeffery M. Vance[1,6,*].

[1] John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida, USA.

[2] Maya Angelou Center for Health Equity, Wake Forest University, Winston-Salem, North Carolina, USA.

[3] Neurogenetics Working Group, Universidad Científica del Sur, Lima, Peru.

[4] Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, Lima, Peru.

[5] Universidad Central del Caribe, Bayamon, Puerto Rico, USA.

[6] Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA

* Corresponding author.

Abstract

Alzheimer's disease (AD) risk differs between population groups, with African Americans and Hispanics being the most affected groups compared to non-Hispanic Whites. Genetic factors contribute significant risk to AD, but the genetic regulatory architectures (GRA) have primarily been studied in Europeans. Many AD genes are expressed in microglia; thus, we explored the impact of genetic ancestry (Amerindian (AI), African (AF), and European (EU)) on the GRA in iPSC-derived microglia from 13 individuals (~4 each with high global ancestry, AD and controls) through ATAC-seq and RNA-seq analyses. We identified several differentially accessible and expressed genes (2 and 10 AD-related, respectively) between ancestry groups. We also found a high correlation between the transcriptomes of iPSC-derived and brain microglia, supporting their use in human studies. This study provides valuable insights into genetically diverse microglia beyond the analysis of AD.

Introduction

Alzheimer's Disease (AD) affects millions of people worldwide with currently ~11% of the US population (65 and older) affected. It is predicted that over 150 million individuals will be affected by AD worldwide by 2050. Pathologically, AD is characterized by β-amyloid (Aβ) deposition as neuritic plaques and intracellular accumulation of hyperphosphorylated tau as neurofibrillary tangles, all of which lead to neurodegeneration and progressive cognitive impairment [1].

African American (AA) and Hispanic (HI) individuals have the highest risk of developing AD, followed by non-Hispanic White (NHW) individuals, likely due to a combination of environmental and genetic factors. Specifically, in the US, AD affects 19% of AA, 14% of Hispanics, and 10% of NHW [2]. Further, over the next 25 years, the greatest growth in AD will be in Africa and South America. Genetic diversity and admixture play important roles in disease risk. African American genomes are typically admixed between African and European ancestries while HI encompass a three-way admixture of European, Amerindian, and African ancestries [3]. Consistent with this, there are ancestry-related differences in the genetic architecture of AD [4]. Although there are gene variants consistently associated with AD risk across different populations, recent genome-wide association studies (GWAS) have identified several ancestry-specific risk variants, including variants in *ABCA7* [5–9], *MPDZ* [10], and *IGF1R* [5,10]. Thus, it is crucial to investigate ancestry-specific disease mechanisms to understand the differential disease susceptibility in different populations and to facilitate the move toward personalized medicine across ancestries.

Most AD-associated and GWAS [11–13] risk loci lie in non-coding, regulatory regions. However, the regulatory architecture of the genome has not been extensively analyzed in diverse populations, with most of the existing data derived from individuals of European ancestry. The different population risk profiles for AD of *APOEe4* carriers of different ancestry present a clear example of how differences in gene regulation can affect AD susceptibility. Rajabli *et al.* demonstrated that the lower risk for AD in carriers of *APOEe4* with African ancestry relative to European ancestry was due to differences in the local

genomic ancestry surrounding the *APOEe4* allele [14,15]. Subsequently, it was found that European local genomic ancestry carriers of *APOEe4* had higher *APOEe4* expression and more open chromatin accessibility than that of African local ancestry carriers [16,17], supporting the recent report that lower expression of *APOEe4* is tied to lower risk [18] and highlighting ancestral differences in gene regulatory networks.

Although much of AD pathogenesis research has focused primarily on neurons, studies suggest a critical role for microglia in the AD disease process. Autopsy studies found an elevated proportion of activated microglia significantly correlated with pathological AD [19], specifically the total Aβ load and number of neuritic plaques. Furthermore, a large number of reported AD GWAS genes are expressed in microglia [20,21], further supporting their role in AD pathology. Microglia are the resident immune cells of the central nervous system (CNS) and play key roles in brain development, synaptic pruning, homeostasis, and neuronal network maintenance, among other immune response processes [22]. Specifically, in the context of AD, microglia are particularly important for Aβ plaque clearance, neuroprotection, inflammatory responses, and synaptic homeostasis [23].

Here we report an examination of iPSC-derived microglia from African, European, and Amerindian ancestries, expanding on our previous studies of single nuclei RNA-seq and single nuclei ATAC-seq on postmortem microglia from the frontal cortex on African and European genomes [16,17]. Additionally, as iPSC-derived cells have become important models for human neurodegenerative research, we performed a comparison between our iPSC-derived microglia and autopsy samples to determine similarities and differences. While this study is focused on AD-GWAS genes, this data will be useful for all neurological genetic studies of African, European, and Amerindian populations, as well as admixed populations of African American and Hispanic individuals.

Results

Differentiation and validation of iPSC-derived microglia.

We differentiated thirteen iPSC-derived microglia (iMGL) lines from individuals of diverse ancestral backgrounds, AD cases and controls, males and females, all derived from individuals over 65 years of age (**Table 1**). Specifically, we differentiated 4 Amerindian (AI), 5 European (EU), and 4 African (AF) iMGL lines. Genotyping and whole genome sequencing were performed to 1) identify the global ancestry and 2) confirm the absence of known mutations in AD-related Mendelian genes (*APP, ABCA7, MAPT, PSEN1, PSEN2, SORL1,* and *TREM2*; **Supplementary Table 1**) that could affect the GRA.

Sample	Global Ancestry		Age	Sex	*APOE*	Clinical Diagnosis
1	AI	96.3%	86	Male	3/3	Control
2	AI	95.5%	86	Male	3/3	Control
3	AI	100%	71	Female	4/4	AD
4	AI	92.0%	86	Female	3/3	Control
5	EU	100.0%	88	Male	4/4	AD
6	EU	88.6%	76	Female	4/4	AD
7	EU	99.7%	65	Female	3/3	Control
8	EU	93.8%	67	Female	3/3	Control
9	EU	99.5%	72	Female	4/4	AD
10	AF	94.4%	70	Female	4/4	AD
11	AF	96.4%	75	Female	3/3	Control
12	AF	91.5%	84	Female	3/3	AD
13	AF	93.5%	90	Female	3/3	MCI

Table 1: iPSC-derived Microglia cell line information. AI: Amerindian. EU: European. AF: African. AD: Alzheimer's disease. MCI: Mild Cognitive Impairment.

All thirteen iMGL cell lines were further validated with microglia cell-specific lineage markers using immunocytochemistry (ICC) (*PU.1* (*SPI1*), *TMEM119, TREM2, and P2RY12*; **Supplementary Figure 1**). All microglia cell lines expressed these cell-type specific markers, and their transcriptomic profiles correlated well (r=0.83) when compared to previously published iMGL using the same differentiation approach [24]. In addition, we also verified that these cells did not express markers for other brain cell types (astrocytes, oligodendrocytes, and neurons; **Supplementary Figure 2 and Supplementary Table 2**).

Brain Microglia vs iPSC-derived Microglia.

We compared the transcriptomic profiles of our iMGL to both Fetal Brain and Adult Brain cell types [17,25]. In both comparisons, we observed the highest correlation between iMGL and Fetal Brain Microglia (ρ= 0.711), and Brain Microglia (ρ= 0.637) compared to other brain cell types (**Table 2**). This data suggests these iMGL recapitulate well the transcriptomic profiles observed in brain microglia and are a good study model.

Cell type		rho	Source
Fetal Brain (Cerebrum)	**Microglia**	**0.711**	Cao, J. *et al,* (2020).
	Astrocytes	0.580	
	Excitatory Neurons	0.599	
	Inhibitory Neurons	0.573	
	Oligodendrocytes	0.563	
	Vascular Endothelial	0.678	
Adult Brain	**Microglia**	**0.637**	Griswold, A. and Celis, K. *et al,* (2021).
	Astrocytes	0.507	
	Excitatory Neurons	0.503	
	Inhibitory Neurons	0.495	
	Oligodendrocytes	0.529	
	OPC	0.509	
	VLMC	0.553	
	Endothelial	0.586	

Table 2: Correlation analysis between iMGL from our study and other cell types. Note that all thirteen iMGL lines were included for these comparisons and the p-value was below $2.2x10^{-16}$ for all comparisons. The Adult Brain data is derived from both African and European ancestry.

Gene expression profiles across ancestries.

We detected a total of 21,980 expressed genes across ancestries and performed differential expression pairwise comparisons between ancestries. In total, we observed 1,103 unique, differentially expressed genes (DEGs) between ancestries (FDR<0.05). Specifically, we identified 971 DEGs between Amerindian (AI) and AF, 320 between AI and EU ancestries, and 62 DEGs between African (AF) and Europeans (EU) (**Figure 1A and B; Supplementary Tables 3, 4, and 5**).

We focused on genes previously identified in AD GWAS studies [5,26–30]. Of the 121 AD GWAS genes (**Supplementary Table 6**), we identified 10 DEGs between AI and AF (*ABI3, CTSB, JAZF1, MS4A6A, PILRA, PLEKHA1, RASGEF1C, SORL1, TREM2,* and *TREML2*) and 3 DEGs between AI and EU (*JAZF1, MS4A6A,* and *SORL1*). Despite our recent report on brain microglia of European and African ancestries [17], we did not observe differential expression for AD risk-modifying genes between AF and EU in our iPSC-derived microglia. We observed significantly higher gene expression in AI compared to AF for *ABI3, JAZF1* (also compared to EU), and *RASGEF1C,* while AF had significantly higher expression of *CTSB, PLEKHA1, SORL1,* and *TREM2* compared to AI. Lastly, we observed that EU express significantly higher amounts of *SORL1* compared to AI (**Figure 1C**).

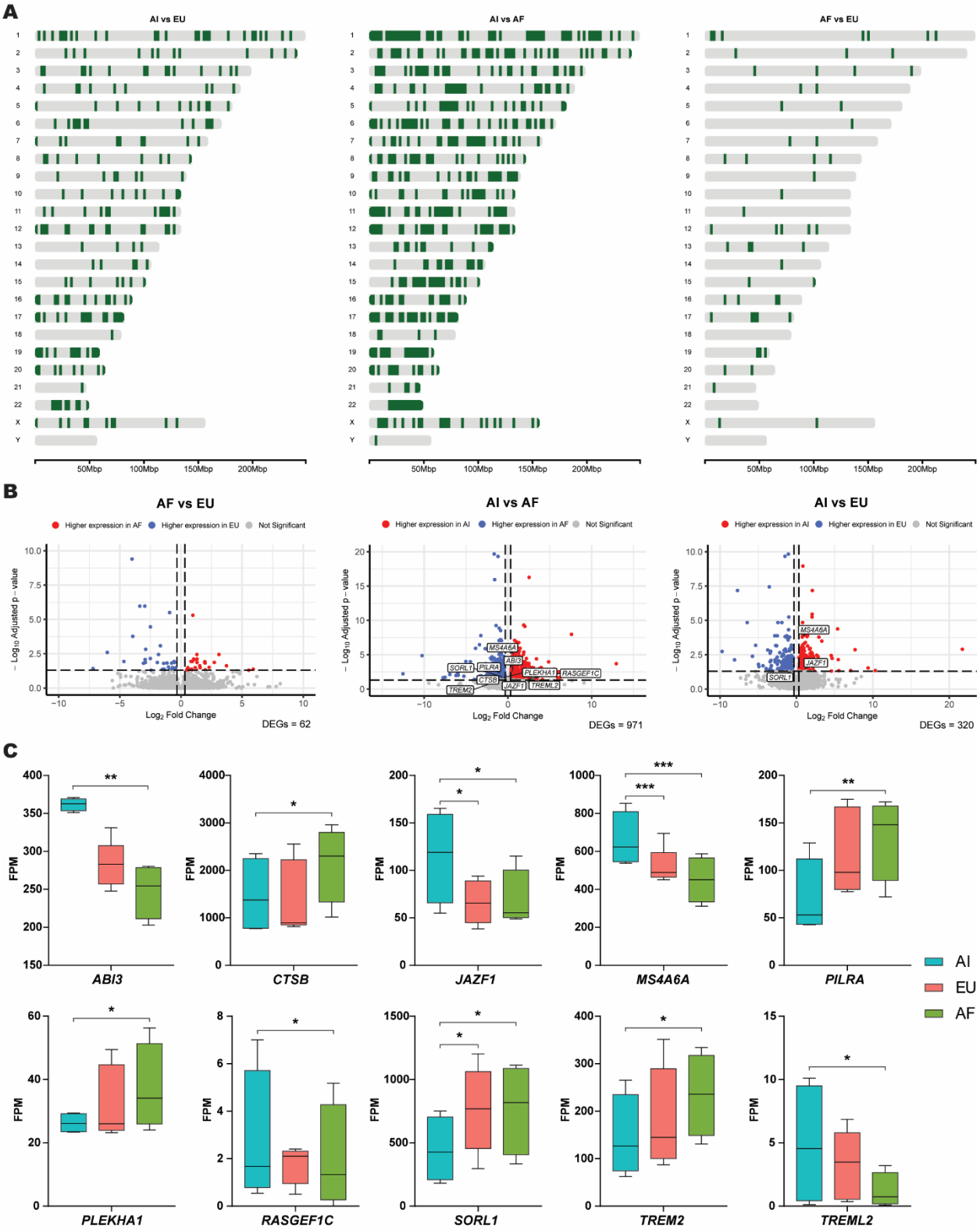

Figure 1: Gene expression across ancestries. **(A)** Chromosome Maps per pair-wise ancestral comparison demonstrating the distribution of differentially expressed genes (DEGs) genome-wide. The dark green color represents DEGs. **(B)** Volcano plots representing gene expression

(Log_2 Fold Change) per pair-wise comparison between ancestries (AF vs EU, AI vs AF, and AI vs EU). All 60,656 expressed variables are represented by the circles. The blue and red colored circles represent the genes that are differentially expressed (Fold Change cutoff of ± 1.25 and have an adjusted p-value (FDR)≤ 0.05). AD risk-modifying genes were highlighted in the white boxes. **(C)** Gene expression (FPM) of AD-related genes that were differentially expressed between ancestries. Box plots represent minimum to maximum FPM values and error bars denote the standard deviation. Asterisks denote adjusted p-value (FDR) with p≤ 0.05 (*), p≤ 0.01 (**), and p≤ 0.001 (***). FPM: Fragments per Million.

Chromatin accessibility across ancestries.

We measured a total of 171,929 ATAC peaks for all ancestries and performed differential accessibility analysis genome wide. Overall, we observed 225 differentially accessible peaks (DAPs) linked to 208 unique, differentially accessible genes (DAGs) between AI and AF, 57 DAPs (55 DAGs) between AF and EU ancestries, and 53 DAPs (52 DAGs) between AI and EU (**Figure 2; Supplementary Tables 7, 8, and 9**). We observed an enrichment in DAPs between AI and EU in chromosome 17 (12.28%, Chi-square p-value=0.038) and chromosome 13 (7.02%, Chi-square p-value=0.041), which contain only ~3% and ~4% of the genome, respectively. Between AI and AF, we observed a significant enrichment in DAPs in chromosome 17 (9.78%, Chi-square p-value=0.004). Lastly, we observed that DAPs between AI and EU were enriched in chromosome 7 (5.66%, Chi-square p-value=0.031; **Figure 2A and Supplementary Table 10**). Overall, we observed that among all DAPs between all three ancestral comparisons, the DAPs lie primarily in intronic regions (~28-39%) followed by distal intergenic (~16-32%) and promoter regions (~23-25%; **Figure 2B; Supplementary Table 11**). Interestingly, in the context of genes associated with AD, we only detected 2 DAGs (*PRDM7* and *SCIMP*) between AI and AF and 1 DAG between AI and EU (*PRDM7*) (**Figure 2C**).

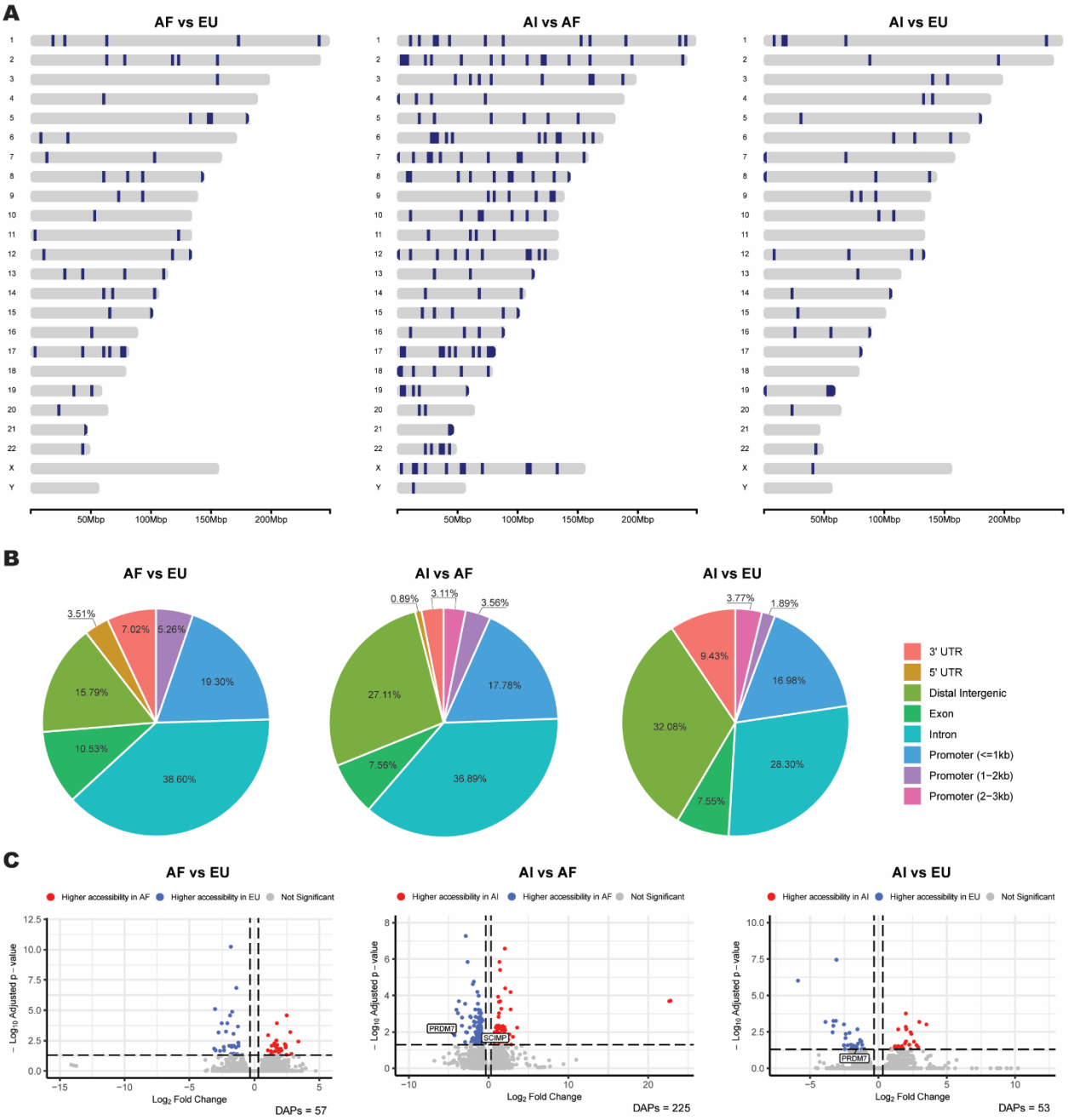

Figure 2: Chromatin accessibility across ancestries. **(A)** Chromosome Maps per pair-wise ancestral group comparison demonstrating the distribution of differentially accessible genes (DAGs) genome-wide. The dark blue color represents DAGs. **(B)** Pie Charts illustrate the regions of the genome in which the differentially accessible peaks lie for each of the ancestral comparisons. **(C)** Volcano plots representing chromatin accessible peaks (log_2 Fold change) per pair-wise comparison between ancestries (AF vs EU, AI vs AF, and AI vs EU). All 171,929 peaks are represented by the circles. The

blue and red colored circles represent the genes that are differentially accessible (Log_2 Fold Change cutoff of ±0.322 and adjusted p-value (FDR)≤ 0.05. AD risk-modifying genes were highlighted in the white boxes.

We observed two DAPs in *PRDM7*: one in the proximal enhancer (Peak 1) and another in a distal enhancer (Peak 2; **Figure 3A**), according to ENCODE classification. Specifically, we observed that compared to AI, AF have significantly higher chromatin accessibility in peak 1 while EU have significantly higher accessibility in peak 2. Interestingly, contrary to other samples of the same ancestry group, we observed that sample 4 (AI) has chromatin accessibility in peak 1 while sample 6 (EU) presents visibly less accessibility in both peaks 1 and 2 (**Supplementary Figure 3**). We performed local ancestry (LA) analyses surrounding the *PRDM7* locus (± 500kb) to further investigate whether it could explain the differences in chromatin accessibility (**Supplementary Table 12**). We observed that samples 1-3 of AI global ancestry, have homozygote Amerindian LA for the *PRDM7* locus while sample 4 has African LA for both haplotypes in this locus aligning with the chromatin accessibility observations within the African global ancestry group. While this data suggests that the African LA of sample 4 in the *PRDM7* locus plays a role in and promotes chromatin accessibility, we did not observe any LA differences in the European global ancestry samples (all homozygote EU LA for this locus).

In addition, we observed a DAP between AI and AF in a distal intergenic enhancer of *SCIMP* (~20kb; **Figure 3B**). We did not observe LA differences within the same global ancestry group for the *SCIMP* locus (**Supplementary Table 12**) which could explain chromatin accessibility differences seen between global ancestry groups in this region (**Supplementary Figure 4**).

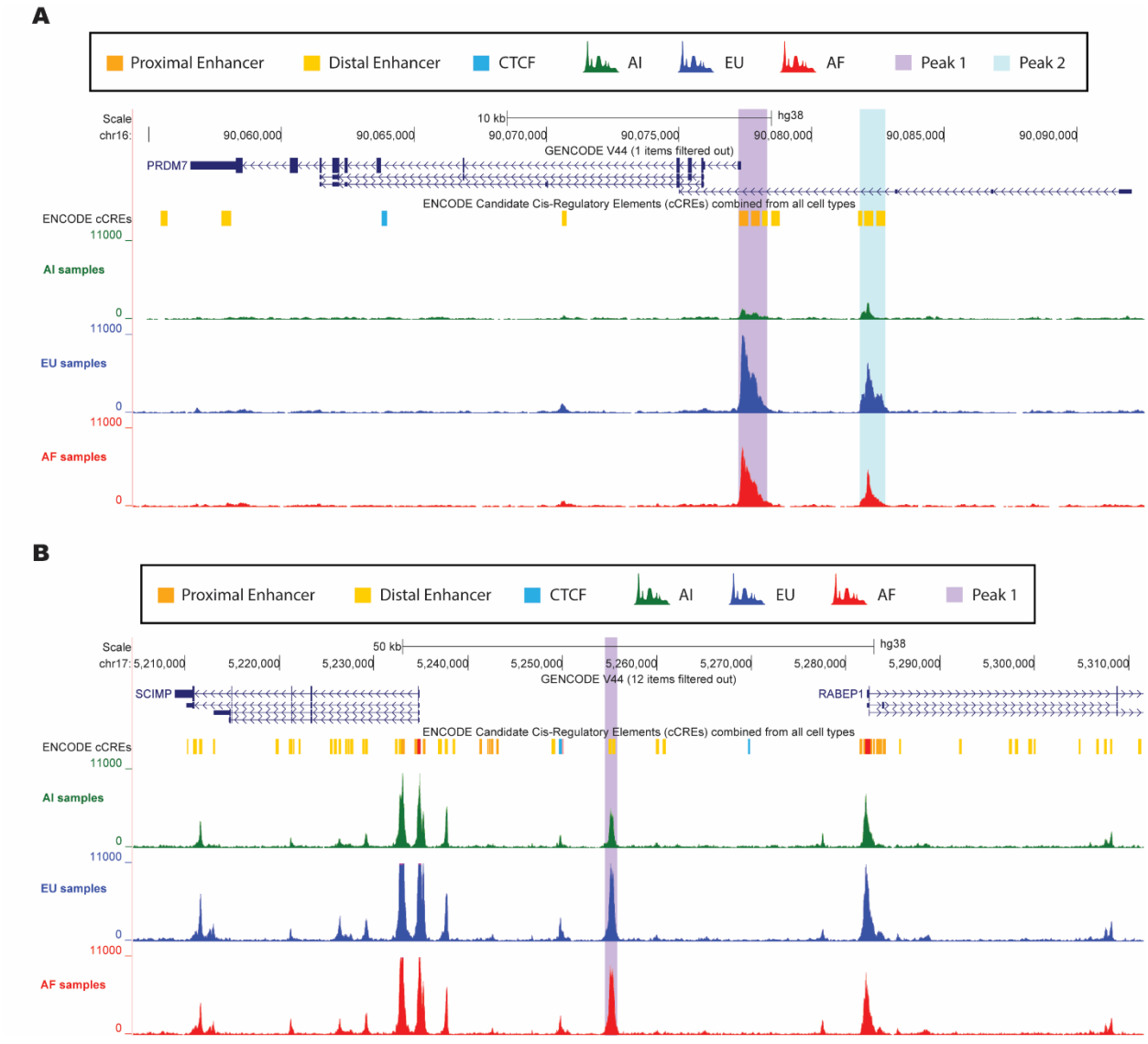

Figure 3: Differentially accessible peaks in AD-risk modifying genes across ancestries. **(A)** Differential chromatin accessible peaks in *PRDM7*. **(B)** Differential chromatin accessible peak in a distal intergenic enhancer of *SCIMP*. Note that the peaks represent merged data of all individuals within the same ancestry group.

Functional enrichment pathway analysis.

To understand the functional mechanisms that might contribute to the differential AD risk across ancestries, we performed functional enrichment pathway analysis between the three ancestral groups using the g:Profiler tool in R. As expected, given the smaller number of DEGs between EU and AF, we only observed two significant functionally

enriched pathways for these ancestries (**Supplementary Table 13**) and none have a known relation to AD. We observed that several DEGs across the other two ancestry group comparisons were involved in immune response, lysosomal activity, sterol and steroid biosynthesis and metabolism, cholesterol biosynthesis and metabolism, lipid transport and metabolism, and phagocytosis - all highly relevant processes in AD pathology (**Figure 4** and **Supplementary Tables 14 and 15**).

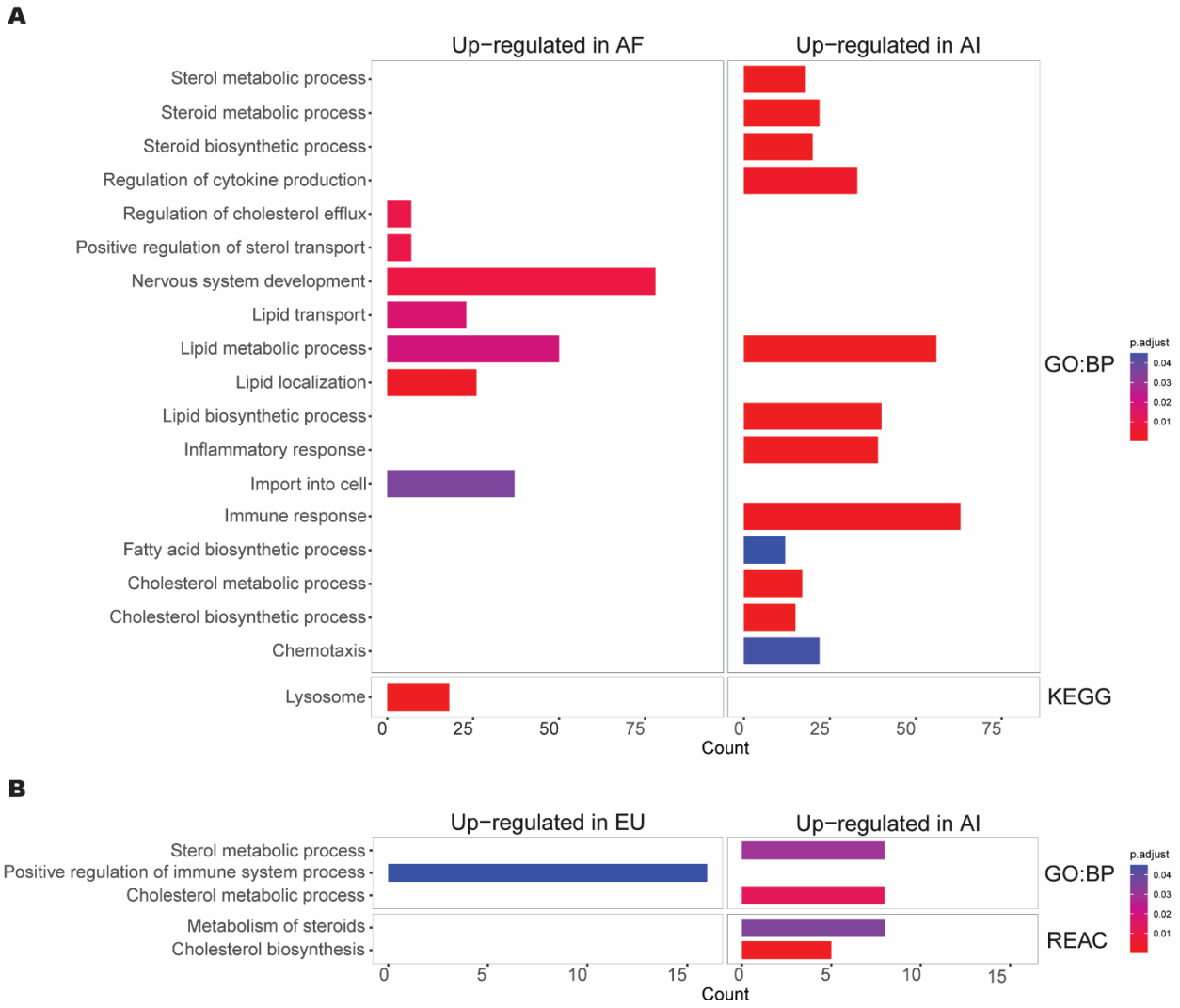

Figure 4: Functional enrichment pathway enrichment across ancestries relevant to AD. Pathway enrichment analyses between **(A)** AI and AF, and **(B)** AI and EU. See **Supplementary Tables 14 and 15**, respectively, for all significantly enriched pathways.

Regulatory architecture in iPSC-derived Microglia.

We studied the overlap between DAGs and DEGs to gain further insights into ancestry-specific regulatory mechanisms. Overall, we observed less than 2% shared DAGs and DEGs when comparing the ancestries (**Figure 5A** and **Supplementary Figure 5**). None of the overlapping DEGs and DAGs were from known AD GWAS genes. We observed that all overlapping DAGs and DEGs between AF and EU, and between AI and EU lay in promoter regions (**Supplementary Tables 16 and 17**, respectively) while there was a wider genomic distribution for those overlapping DAGs and DEGs between AI and AF (**Supplementary Table 18**).

However, despite the small overlap between DAGs and DEGs with p-value≤ 0.05, we still observed a correlation between expression and chromatin accessibility in the promoter peaks (r= 0.53 (AF vs EU); r= 0.57 (AI vs EU); r= 0.47 (AI vs AF); **Supplementary Figure 6**).

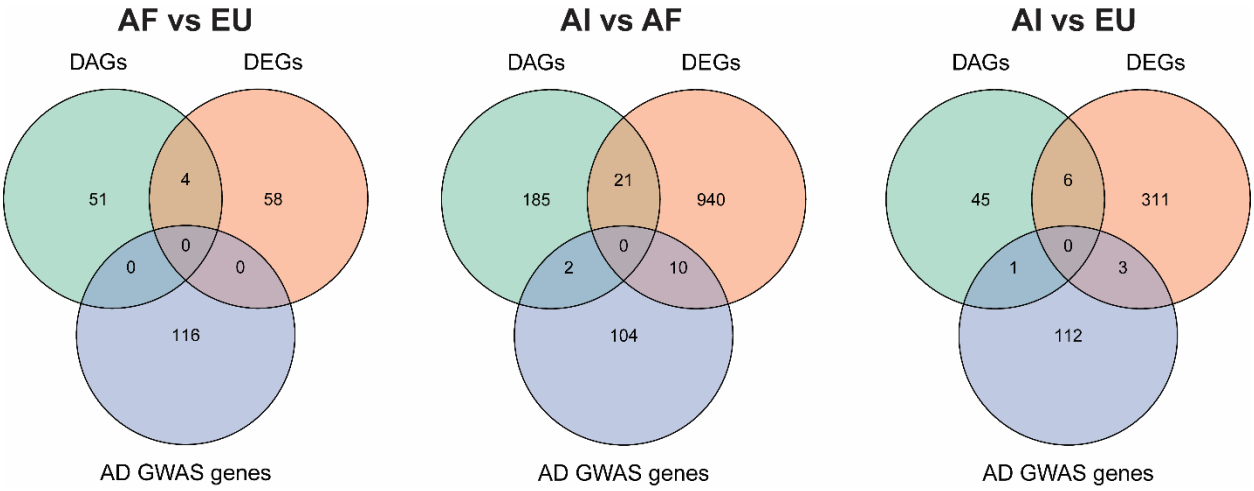

Figure 5: Overlap between differentially accessible ATAC-seq genes, differentially expressed RNA-seq genes, and AD GWAS genes between ancestry-group comparisons.

Regulatory differences specific to AD diagnosis, *APOE* genotype, and Sex.

Between AD cases and controls, we performed differential expression analysis for 12 samples (the MCI sample was excluded from this analysis) and observed a total of 7 DEGs between non-cognitively impaired individuals and AD samples (**Supplementary Table 19**). None were previously identified as AD risk-modifying genes. Differential expression analysis between *APOEe3* and *APOEe4* homozygote carriers revealed 7 DEGs (**Supplementary Table 20**). Between the two analyses, we only found one DEG in common, high mobility group AT-hook 2 (*HMGA2*), which was overexpressed in AD and *APOEe4* carriers as compared to controls and *APOEe3* carriers (**Supplementary Figure 7**). The sex comparison revealed a total of 116 DEGs between Males and Females (**Supplementary Table 21**), none of which were AD risk-modifying genes or overlapped with any of the DEGs from the two aforementioned analyses. On the chromatin accessibility level, we only observed three DAPs/DAGs between *APOEe3* and *APOEe4* carriers (**Supplementary Table 22**), one DAP/DAG between cases and controls (**Supplementary Table 23**), and 136 DAPs between Males and Females (90 DAGs; **Supplementary Table 24**). None of these peaks have been previously connected to either AD or *APOE* genotype. Lastly, we observed an overlap between eleven sex-specific DEGs and DAGs, most of which are located in chromosomes X and Y.

Ancestry-specific genetic regulatory architecture tool for other Neurological diseases.

Despite the lack of ancestry-specific studies for other neurological diseases, ancestry might affect disease risk as observed in AD pathology. To demonstrate the importance of this GRA resource for the study of other neurological diseases in diverse ancestries, we compared both DEGs and DAGs identified for each of the ancestry comparison groups in our study with GWAS genes identified for Autism Spectrum Disorder (ASD) [31–41], Schizophrenia (SZ) [42–57], Bipolar disorder (BP) [54,58–64], Parkinson's Disease (PD) [65,66], Multiple Sclerosis (MS) [67,68], Stroke [69], Coronary Artery Disease (CAD) [70–76], and Hyperlipidemia (HDL) [77,78] (**Figure 6**).

A

		GWAS								
		ASD	SZ	BP	PD	MS	Stroke	CAD	HLD	Total
DEGs	AF vs EU	0	4	0	0	2	0	1	1	62
	AI vs AF	4	57	13	5	22	3	28	7	971
	AI vs EU	1	25	5	2	10	2	8	2	320
	Total Queried	184	1989	482	86	424	87	449	106	

B

		GWAS								
		ASD	SZ	BP	PD	MS	Stroke	CAD	HLD	Total
DAGs	AF vs EU	0	7	0	0	0	0	3	0	55
	AI vs AF	2	19	5	1	3	0	4	0	208
	AI vs EU	0	1	1	0	0	0	1	0	52
	Total Queried	184	1989	482	86	424	87	449	106	

Figure 6: The genetic regulatory architecture in iMGL of diverse ancestries as a useful resource to study other neurological and associated diseases. We illustrate the overlap between ancestry-specific **(A)** DEGs and **(B)** DAGs from our study with previously identified GWAS genes for Autism Spectrum Disorder (ASD), Schizophrenia (SZ), Bipolar disorder (BP), Parkinson's Disease (PD), Stroke, Multiple Sclerosis (MS), Coronary Artery Disease (CAD), and Hyperlipidemia (HLD). Gray boxes represent the total number of genes queried.

Discussion

Recent studies have demonstrated that genetic disease associations differ in their strength and location between ancestries [5,27,28,30]. As the majority of genetic associations are in non-coding regions, it is important to gain insight into the regulatory architecture of other ancestries besides European. Given the key role of microglia in AD pathology, we report, for the first time, epigenetic and disease-relevant differences between these ancestries in iMGL. While we have focused on AD, the microglial regulatory architecture presented here will be applicable to any study of the CNS.

Several known AD genes demonstrated ancestral expression differences in the microglia. One of these genes was ABI family member 3 (*ABI3)*, differentially expressed between AI and AF in this study and which has been previously found to be associated with AD in African American individuals [79]. Studies have found that loss of *ABI3* function in mice was associated with Aβ-amyloidosis [80] and increased *ABI3* expression in microglia has been observed surrounding amyloid plaques in AD brain samples [81]. Both studies hypothesize that *ABI3* expression plays a role in microglia migration in the central nervous system and affects disease progression in the absence of a functioning protein. We find that AF have on average the lowest expression of *ABI3*, compared to AI, supporting *ABI3* as an AD risk factor specifically in AF.

Another known AD gene, Cathepsin B (*CTSB)*, identified here as differentially expressed with higher expression levels in AF compared to AI, has been implicated as a major contributor to cognitive dysfunction and neuropathological changes, such as lysosomal dysfunction, cell death, and inflammatory responses [82,83]. Interestingly, increased CTSB protein expression has been reported in AD patients compared to controls [84–86]. It was also previously reported that *APOEe4* carriers of AF local ancestry expressed higher *CTSB* in brain microglia compared to those of EU local ancestry surrounding the *APOE* locus [17], similar to the trend observed in our dataset between AF and EU (**Figure 1C**). Again, this could suggest a larger role in AD risk for *CTSB* lying on AF local ancestry in African American individuals. Both of these differences were seen between AF and AI samples, which displayed the largest genomic differences between the three ancestries

examined in this study. These are the two populations at either end of the migration spectrum for humans, implying these genetic ancestries had the longest time to evolve independently, creating ancestries who are the least related genetically.

In addition, even for genes without significant ancestral differences, the expression and accessibility data here can be useful for further understanding of the locus across population groups. For example, another AD-risk-modifying gene that showed differential gene expression is *MS4A6A.* This gene has been shown to be highly expressed in microglia [87] and it was previously reported that brain microglia of AF ancestry express less *MS4A6A* compared to those of EU ancestry [17]. Despite not reaching significance, we did observe a similar trend towards less *MS4A6A* expression in AF iMGL compared to EU iMGL. *TREM2*, another well-known AD-GWAS gene, is primarily expressed in microglia and has been heavily implicated in AD progression [88–91]. Interestingly, we found that AI cells express the lowest amount of *TREM2*. Data show that *TREM2* mRNA levels are associated with amyloid burden in cortical regions [92] and loss-of-function *TREM2* variants are associated with dementia [93–95], implying that the lower expression in AI microglia might impact AD risk in this ancestry due to reduced microglia functionality (Aβ-plaque clearance, *APOE*-mediated functions, immune modulation, and cell survival).

The iMGL lines used here varied not only in their genetic ancestry, but also in other variables such as sex, *APOE* genotype, and disease status which could complicate the interpretation of results. Therefore, we also performed differential expression analysis between Males and Females, AD vs controls, and *APOEe3* vs *APOEe4* carriers. Most of our AD patients were *APOEe4* homozygotes as at least 60% of AD patients carry the *APOEe4* allele. Despite observing a small number of DEGs between AD vs Controls and *APOE e3* vs *e4* carriers, we observed that *HMGA2,* a high-mobility protein that modulates transcription and chromatin condensation, was differentially expressed in both comparisons. Specifically, we observed higher gene expression in AD individuals and *APOEe4* carriers. Interestingly, silencing of *HMGA2* has been reported to lead to increased expression of the PI3K/AKT signaling pathway and improved memory and learning ability, reduced brain injury, and decreased oxidative stress and inflammatory reactions in mice [96]. It was also recently reported that downregulation of *HMGA2* in AD

patients was associated with increased lifespan [97]. Thus, together with these findings, our results also suggest and support that increased *HMGA2* expression is a risk factor for AD.

We are often taught that chromatin accessibility is a key factor controlling gene expression. Comparing the significantly different changes in gene expression and chromatin accessibility between ancestries provides one opportunity to examine this relationship. Our differential analysis between ancestries revealed greater differences in gene expression (DEG) (approximately 0.3-4.4% of genes depending on the paired comparison) than in chromatin accessibility (DAP/DAG) (0.03-0.13%). This supports the growing understanding of the complexity of our cells in regulating gene expression and that transcription is a much more complex mechanism and higher accessibility is only one factor that could affect gene expression. For example, DNA sequence variability both at binding sites and distal eQTLs can complicate interpretation of the (dis)concordance between gene expression and chromatin accessibility changes. However, as expected, when expanding our sample size by using all our expression and accessibility data, we do find the expected moderate correlation between chromatin accessibility and expression (r=0.47 to 0.57).

iPSCs and derived cells have become important models for human brain disorders. We demonstrated that their transcriptome has a strong correlation with brain single nuclei RNAseq results [17]. These iPSC-derived microglial cells were grown in the absence of other cell types and with a lack of environmental stressors. The complex gene regulatory networks operating in brain cells reflect the interplay of mostly invariable genetic factors with a dynamic exposome that includes chemical exposures, diet, and diverse stressors across the life course. One could postulate that microglia co-cultured with other CNS cell types or 3D organoids would feature cell-cell interactions that would provide an even stronger correlation with the brain transcriptome.

We did not observe any of the currently known African-specific AD GWAS genes [5] to be differentially expressed or accessible in the AF ancestry iMGL compared to the AI or EU ancestries. This could be explained by the fact that some of these genes were not expressed in iMGL and others had heterogenous expression levels between the limited

355 number of individuals. The relatively small number of individuals included is the main
356 limitation of this study. This is a general limitation of iPSC-derived cell studies which are
357 expensive and time-consuming. Some of the differential findings reported here may
358 reflect individual heterogeneity rather than ancestry generalizations. Additional iPSC-
359 derived cell lines are needed to fully explore the regulatory architecture and to capture
360 individual variability. Further genomic studies such as Hi-C will enhance these
361 comparisons, particularly for specific genes of interest.

Conclusions

Overall, we provide novel insights into the genetic regulatory architecture of microglia from three ancestry groups: Amerindian, African, and European. Transcriptional and architectural similarity was the most common finding, which is reassuring for future therapeutic interventions. We found a good correlation between the transcriptome of our iMGL and reported brain transcriptomes, as well as concordance for previously reported AD risk genes, supporting ancestral differences. These findings support the role of iMGL as a valuable model for human disease. Our data also supports a role for *HMGA2* expression in *APOEe4* carriers and AD risk. Lastly, this study provides a useful resource for the research community as it provides novel data on genome-wide regulatory architectures of diverse, understudied, genetic groups that could be applied to the study of other brain diseases, particularly those with high microglia involvement.

Methods

Sample collection.

All samples of AI, EU, and AI cases and controls selected for this study were obtained from the John P. Hussman Institute for Human Genomics (HIHG) at the University of Miami Miller School of Medicine with the exception of the induced pluripotent stem cells derived from samples 7-9 which were obtained through ADRC from the University of California Irvine (UCI). All participants were ascertained using a protocol approved by the appropriate Institutional Review Board. This study received ethical approval from the University of Miami Institutional Review Board (approved protocol #20070307).

Global ancestry ascertainment.

We calculated the admixture proportions using a model-based clustering algorithm, as implemented in the ADMIXTURE software [98]. A supervised ADMIXTURE analysis was performed at K = 4, incorporating four reference populations: 104 African, 84 European, 108 Amerindian, and 102 East Asian individuals from the Human Genome Diversity Project reference populations.

Local ancestry ascertainment.

To infer local ancestry, we first merged our dataset with the Human Genome Diversity Project reference panel, including European, African, and Amerindian reference populations [99]. Next, we phased the combined data using SHAPEIT4 with default settings, referencing the 1000 Genomes Phase 3 reference panel [100,101]. Finally, we estimated local ancestry at each genomic locus using RFMix v2 software [102].

Whole Genome Sequencing (WGS).

DNA was extracted from all individual cell lines using the QIAamp DNA Blood Kit (QIAGEN, #51104) according to the manufacturer's instructions. 1.5µg of DNA was submitted for WGS at the Center for Genome Technology (CGT) Sequencing Core at the HIHG using standard Illumina PCR-free library prep and sequencing protocols on the NovaSeq6000 followed by a bioinformatics pipeline incorporating the GATK Best Practices analysis recommendations [103]. Individuals were screened for rare coding

variants in seven AD-related genes nominated as likely causative by the ADSP Gene Verification Committee and variants in the promoter regions of the ten AD genes that had differential gene expression (**Supplementary Table 1**).

Induced pluripotent stem cell generation.

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using SepMate-50 tubes with Lymphoprep (STEMCELL Technologies, #85450 and #07801) through density-gradient centrifugation according to the manufacturer's instructions. PBMCs were reprogrammed into induced pluripotent stem cells (iPSCs) using CTS™ CytoTune™-iPS 2.1 Sendai Reprogramming Kit (Invitrogen, #A34546) according to the manufacturer's instructions. Reprogrammed cells were tested for Sendai Virus absence, trilineage differentiation capability, immunocytochemistry, STR profiling, karyotyping, and mycoplasma testing as previously described [104]. PBMC isolation and reprogramming was performed at the Hussman Institute for Human Genomics (HIHG) Induced Pluripotent Stem Cell (iPSC) Core at the University of Miami. Validation analyses were performed by the HIHG-iPSC Core and WiCell.

Differentiation of iPSCs to Microglia.

iPSCs were differentiated into hematopoietic progenitor cells (HPCs) and subsequently into Microglia (MGL) as previously described [24] with minor modifications.

In brief, feeder-free iPSCs were cultured and expanded in StemFlex medium (Gibco™, #A3349401) in vitronectin (10µg/ml, Gibco™, #A31804) coated cell culture-treated plates. On day -1, iPSCs were passaged with 0.5M EDTA onto Matrigel-coated (Corning, #354277) 12-well plates at a density of 10-20 aggregates/cm^2 (>50µm in size). On day 0, if 4-10 colonies/cm^2 adhered, the StemFlex medium was replaced with 1ml/well of HPC medium A (Basal medium with supplement A (1:200), STEMCELL Technologies, #05310). Half-medium change was carried out 48 hours later. On day 3, HPC medium A was replaced in full by medium B (Basal medium with supplement B at 1:200). Half-medium changes of medium B were performed on days 5, 7, and 10. HPCs were harvested on day 12.

On day 0 of microglia differentiation (day 12 of HPC differentiation), HPCs were plated at 22,000 cells/cm^2 onto a Matrigel-coated 6-well plate containing 2ml of Microglia differentiation medium (Basal Medium with supplement 1 and 2 at 1:9 and 1:225, respectively; STEMCELL Technologies, #100-0019). Cells were supplemented with fresh half-medium every other day from day 0 to day 10. On day 12, cells were collected and centrifuged at 300 x g for 5 minutes. The cell pellet was resuspended in 2ml/well of fresh Microglia differentiation medium and transferred to a freshly Matrigel-coated 6-well plate. Cells were supplemented with 1ml of media every second day until day 22. Microglia cells were collected, resuspended in 2ml of Microglia maturation medium (Basal Medium with supplement 1 (1:9), and 2 and 3 (1:225); STEMCELL Technologies, #100-0020), and re-plated for assays into new Matrigel-coated 6-well plates. Lastly, on day 26, microglia were harvested for immunocytochemistry (ICC), bulk RNA-, and ATAC-sequencing.

RNA isolation and sequencing.

Total RNA was isolated from 1 million microglial cells per cell line using the RNeasy Mini kit (QIAGEN, #74104) according to the manufacturer's instructions. Suspension cells were collected and centrifuged for 5 minutes at 300 x g. 600µl of RLT buffer (including β-Mercaptoethanol at 1/100) was used to collect semi-attached microglia and subsequently resuspend the cell pellet from the previous step. Cells were briefly vortexed for 1 minute and homogenized by loading the lysate into a QIAshredder spin column (QIAGEN, #79656) and centrifuging for 2 minutes at full speed. The homogenized lysate was resuspended in 1 volume of 70% ethanol and transferred to a RNeasy spin column and centrifuged for 30 seconds at 8,000 x g. 350µl of Buffer RW1 was added to the same spin column and centrifuged for 15 seconds at 8,000 x g. Following this, 80µl of DNAse I incubation mix (70µl of RDD buffer and 10µl of DNAse I, QIAGEN, #79254) were added to the spin column and incubated at RT for 15 minutes. Buffer RW2 (350µl) was transferred to the spin column and centrifuged for 15 seconds at 8,000 x g. 500µl of RPE buffer were loaded into the column followed by a centrifugation step of 30 seconds at 8,000 x g. The previous step was repeated once again but centrifuged for 2 minutes at 8,000 x g to ensure all residual ethanol was removed. The RNeasy spin column was transferred to a new 1.5ml collection tube and 30µl of RNAse-free water were added to

the column to elute the bound RNA. Lastly, the spin column was centrifuged at 8,000 x g for 1 minute and then stored at -80°C until further used. The RNA concentration and quality were assessed using the Agilent Tapestation (Agilent Technologies) to determine the RNA integrity number (RIN).

Bulk RNA sequencing.

RNA libraries were prepared at the John P. Hussman Institute for Human Genomics Center for Genome Technology (University of Miami, FL) from ribodepleted total RNA. In brief, total RNA was prepared with the TECAN Universal Plus Total RNA-seq with NuQuant® Human AnyDeplete according to the manufacturer's instructions, using 60ng via QuBit and 16 PCR cycles. The normalized libraries were sequenced as paired end 100bp reactions targeting 30 million reads/sample on the Illumina NovaSeq 6000 (Illumina, CA). The raw FASTQ files were processed through an in-house bioinformatics pipeline including adapter trimming by TrimGalore (v0.6.10) (https://github.com/FelixKrueger/TrimGalore), alignment to the GRCh38 human reference genome with STAR (v2.5.0a) [105], and gene counts quantified against the GENCODEv35 gene annotation release using the GeneCounts module implemented in STAR.

Bulk ATAC-sequencing.

Cultured cells were treated with DNase I (200U/mL; QIAGEN, #79254) at 37°C for 30 minutes. The treated cells were then harvested and pelleted at 400 x g for 5 minutes at 4°C. The cell pellet was carefully washed in cold 1x PBS. The cells were re-pelleted as described before and then lysed in 100µl of lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, and 0.01% Digitonin) on ice for 5 minutes. Next, the lysed microglia were washed in 1ml of wash buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, and 0.1% Tween-20) and 100,000 nuclei were pelleted at 500 x g for 10 minutes at 4°C. The nuclei were incubated at 37°C for 30 minutes at 1,000rpm in 100µl of Transposition mix (2x Tagment DNA Buffer, 1x PBS, 0.1% v/v Tween-20, 0.01% v/v Digitonin, and 5µl of Tagment DNA Enzyme 1). The transposed DNA was purified using the MinElute PCR Purification kit (QIAGEN, #28004) and eluted

in 10µl of Elution Buffer. The purified transposed DNA was combined with 25µM of Custom Adapter 1 (no primer mix), 25µM of Custom Adapter 2 (barcode), and NEBNext High-Fidelity 2x PCR Master Mix and ran on a thermocycler with the following conditions: 72°C for 3 minutes, 98°C for 30 seconds, and 5 cycles of 98°C for 30 seconds, 63°C for 30 seconds, and 72°C for 1 minute. The additional number of cycles required was determined as described in [106] and ran with the same conditions abovementioned. The amplified libraries were purified with the MinElute PCR Purification kit and eluted in 20µl of Nuclease-free water. Library traces were assessed by the Agilent Tapestation and when necessary, size selection purification was carried out using the AMPure XP beads (Beckman Coulter, #A63880) according to the manufacturer's instructions. See **Supplementary Table 26** for full adapter sequences. Libraries were sequenced in paired end 100bp reactions targeting 30 million reads/sample on the Illumina NovaSeq 6000. The ATAC-seq data were preprocessed (trimmed, aligned, filtered, and quality-controlled) and analyzed using an adapted version of the ENCODE ATAC-seq pipeline. In brief, adapters and poor-quality bases were trimmed using TrimGalore (v0.6.10) (https://github.com/FelixKrueger/TrimGalore). Reads were aligned to the CRCh38 human reference genome with bowtie (v2.2.2) [107], duplicates marked with Picard (v2.1.1) (https://broadinstitute.github.io/picard/), and peaks called using MACS2 (v2.2.7.1) [107]. Peaks were merged across all samples using an overlapping peak/union strategy to obtain a list of peaks across all samples. Counts per peak were calculated from individual aligned BAM files using htseq-count (v1.99.2) using the un-stranded option.

Differential expression and accessibility analyses.

Differential expression and accessibility analyses were carried out across the different ancestral populations using DESeq2 (version 3.17) package [108] in R language environment (version 4.2.1). We used DESeq2 default parameters and controlled for batch differences (design = ~batch + ancestry). Three contrasts were run: AF vs EU, AI vs AF, and AI vs EU. Genes that were significantly expressed and/or accessible were identified with an FDR adjusted p-value of <0.05.

Functional enrichment pathway analysis.

Functional enrichment analysis was done with the R library gprofiler2 [109]. We extracted gene symbols of DEG between ancestries (FDR adjusted p-value of <0.05), and the function *gost* was used to perform the gene set enrichment analysis for each ancestry comparison using the Gene Ontology, KEGG pathways, and REACTOME databases. Multiple comparison correction of enrichment scores was done with the 'gSCS' method. Pathways were considered significant if p-adj < 0.05. Results were manually curated to show known pathways related to AD pathogenesis, and the corresponding full lists of enriched terms are described in **Supplementary Tables 13, 14, and 15**.

ATAC peak annotation

The function *annotatePeak* from Chipseeker R library [110] was used to annotate peaks with the nearest gene and genomic region. The annotation was done at the transcript level using the GENCODE V44 database. The distance of ±3 kb from the transcription start sites (TSS) was used to assign a peak to a gene promoter-TSS, and the following priority was defined for annotation: "Promoter", "5UTR", "3UTR", "Exon", "Intron", "Downstream", "Intergenic".

Immunocytochemistry (ICC) and fluorescence imaging.

Cultured microglia cells were fixed with 4% formaldehyde for 15 minutes at RT and washed with 1x PBS. Cells were permeabilized for 10 mins with PBS-T solution (0.1% Triton X and 1x PBS). The microglia cells were then incubated in blocking buffer (1x PBS and 5% normal donkey serum) for 1 hour at RT. The blocking buffer was removed and incubated in the primary antibody solution (1% donkey serum, 0.1% Tween-20, 0.01% Sodium Azide, and target primary antibody) at 4°C overnight. The following day, the primary antibody solution was removed, and the cells were washed three times with 1x PBS. Following this, the secondary antibody solution (1% donkey serum, 0.1% Tween-20, 0.01% Sodium Azide, and secondary antibody) were added to each well, and cells were incubated for 1 hour at RT in the dark. Lastly, the secondary antibody solution was removed, and cells were washed thrice with PBS. The cells were washed with 1x PBS and incubated with DAPI (NucBlue Fixed Cell Stain). Images were acquired using a

Keyence Microscope BZ-X800. See Supplementary Table 25 for details on all antibodies used for ICC analysis.

Correlation analyses between differential expression and differential accessibility

Pearson correlation (r) was used to evaluate the relationship between gene expression and corresponding promoter accessibility. First, DEGs between ancestries with $|\log_2(\text{FoldChange})| \geq 1$ and adjusted p-values ≤ 0.1 were considered for the analysis. Then, promoter peaks (distance of ±3 kb from TSS) annotated to those DEGs were considered for correlation analysis.

Correlation analyses between iPSC-derived Microglia and other cell types.

Correlation analyses between iMGL and Brain cell types were performed using Spearman correlation analyses. Specifically, we calculated the average expression of all thirteen iPSC-derived Microglia (iMGL) cell lines included in this study for each gene. Note that genes with an expression value of 0 were excluded as well as sex-related (Chromosomes X and Y) and mitochondrial genes. Following this, genes were ranked in descending order by expression level for both iMGL and brain cell types, and only genes present in both comparison datasets were included in the Spearman correlation test.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. Sequencing files can be requested to the corresponding author.

Acknowledgments

This study was supported by the National Institute on Aging (grant numbers U01-AGO72579, RF1-AGO59018, U01-AG066767, U01-AG052410, R56-AG072547, R01-AG070864).

We acknowledge the Center for Genome Technology (CGT) from the John P. Hussman Institute for Human Genomics (HIHG) from the University of Miami, Miller School of Medicine for the genomic and data analyses. We thank Dr. Lily Wang for meaningful data discussions and express our gratitude to the numerous participants, researchers, and staff involved for their invaluable contributions to the present study.

Author contributions

S.M., D.M.D, A.J.G., J.I.Y., and J.M.V. conceptualized the project and planned experiments. S.M. performed the experiments. S.M., L.B.N., and A.J.G. analyzed the data. A.M.R., and L.C., contributed to the performance of experiments. B.DR. assisted with iPSC reprogramming. J.R. and F.R. performed local and global ancestry analyses. D.V.B. and L.B.N. performed bioinformatic analyses. K.H.-N., P.W., L.A., T.S., P.M., M.I.-M., S.T., G.B., M.C.-O., B.F.-A., and M.A.P.-V. contributed to sample collection and processing. S.M., L.B.N., A.M.R., B.DR., K.N., L.W., D.M.D., F.R., A.J.G., J.I.Y., and J.M.V. discussed the data results. S.M., L.B.N., F.R., A.J.G., D.M.D., J.I.Y, J.M.V. wrote the manuscript. All authors read and approved the final manuscript.

Materials and Correspondence

Correspondence to Jeffery M. Vance (jvance@miami.edu)

References

1. Knopman, D. S. *et al.* Alzheimer disease. *Nature Reviews Disease Primers 2021 7:1* **7**, 1–21 (2021).

2. Rajan, K. B. *et al.* Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). *Alzheimer's & Dementia* **17**, 1966–1975 (2021).

3. Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D. & Mountain, J. L. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet* **96**, 37 (2015).

4. Reitz, C., Pericak-Vance, M. A., Foroud, T. & Mayeux, R. A global view of the genetic basis of Alzheimer disease. *Nature Reviews Neurology 2023 19:5* **19**, 261–277 (2023).

5. 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).

6. Reitz, C. *et al.* Variants in the ATP-Binding Cassette Transporter (ABCA7), Apolipoprotein E ϵ4, and the Risk of Late-Onset Alzheimer Disease in African Americans. *JAMA* **309**, 1483–1492 (2013).

7. Cukier, H. N. *et al.* ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet* **2**, (2016).

8. Talebi, M. *et al.* ABCA7 and EphA1 Genes Polymorphisms in Late-Onset Alzheimer's Disease. *Journal of Molecular Neuroscience* **70**, 167–173 (2020).

9. Vardarajan, B. N. *et al.* Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann Neurol* **78**, 487–498 (2015).

10. Ray, N. R. *et al.* Extended genome-wide association study employing the African genome resources panel identifies novel susceptibility loci for Alzheimer's disease in individuals of African ancestry. *Alzheimers Dement* (2024) doi:10.1002/ALZ.13880.

11. Giral, H., Landmesser, U. & Kratzer, A. Into the Wild: GWAS Exploration of Non-coding RNAs. *Front Cardiovasc Med* **5**, 181 (2018).

12. Mirza, A. H., Kaur, S., Brorsson, C. A. & Pociot, F. Effects of GWAS-Associated Genetic Variants on lncRNAs within IBD and T1D Candidate Loci. *PLoS One* **9**, e105723 (2014).

13. Andrews, S. J. *et al.* The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMedicine* **90**, 104511 (2023).

14. Rajabli, F. *et al.* A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's Disease in African Ancestry. *PLoS Genet* **18**, e1009977 (2022).

15. Rajabli, F. *et al.* Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet* **14**, e1007791 (2018).

16. Celis, K. *et al.* Ancestry-related differences in chromatin accessibility and gene expression of APOE ε4 are associated with Alzheimer's disease risk. *Alzheimer's & Dementia* **19**, 3902–3915 (2023).

17. Griswold, A. J. *et al.* Increased APOE ε4 expression is associated with the difference in Alzheimer's disease risk from diverse ancestral backgrounds. *Alzheimer's & Dementia* **17**, 1179–1188 (2021).

18. Vance, J. M. *et al.* Report of the APOE4 National Institute on Aging/Alzheimer Disease Sequencing Project Consortium Working Group: Reducing APOE4 in Carriers is a Therapeutic Goal for Alzheimer's Disease. *Ann Neurol* (2024) doi:10.1002/ANA.26864.

19. Felsky, D. *et al.* Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nature Communications 2019 10:1* **10**, 1–12 (2019).

20. Nott, A. *et al.* Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. *Science (1979)* **366**, 1134–1139 (2019).

21. Novikova, G. *et al.* Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nature Communications 2021 12:1* **12**, 1–14 (2021).

22. Colonna, M. & Butovsky, O. Microglia Function in the Central Nervous System During Health and Neurodegeneration. *Annu Rev Immunol* **35**, 441 (2017).

23. Gao, C., Jiang, J., Tan, Y. & Chen, S. Microglia in neurodegenerative diseases: mechanism and potential therapeutic targets. *Signal Transduction and Targeted Therapy 2023 8:1* **8**, 1–37 (2023).

24. McQuade, A. *et al.* Development and validation of a simplified method to generate human microglia from pluripotent stem cells. *Mol Neurodegener* **13**, 1–13 (2018).

25. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science (1979)* **370**, (2020).

26. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452 (2013).

27. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics 2019 51:3* **51**, 414–430 (2019).

28. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics 2022 54:4* **54**, 412–436 (2022).

29. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nature Genetics 2021 53:9* **53**, 1276–1282 (2021).

30. Lake, J. *et al.* Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. *Mol Psychiatry* (2023) doi:10.1038/s41380-023-02089-w.

31. Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* **21**, 4781–4792 (2012).

32. Anney, R. *et al.* A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* **19**, 4072–4082 (2010).

33. Weiss, L. A. *et al.* A genome-wide linkage and association scan reveals novel loci for autism. *Nature 2009 461:7265* **461**, 802–808 (2009).

34. Ma, D. *et al.* A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1. *Ann Hum Genet* **73**, 263–273 (2009).

35. Matoba, N. *et al.* Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Translational Psychiatry 2020 10:1* **10**, 1–14 (2020).

36. Liu, X. *et al.* Genome-wide Association Study of Autism Spectrum Disorder in the East Asian Populations. *Autism Research* **9**, 340–349 (2016).

37. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism* **8**, 21 (2017).

38. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics 2019 51:3* **51**, 431–444 (2019).

39. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Translational Psychiatry 2020 10:1* **10**, 1–11 (2020).

40. Kuo, P. H. *et al.* Genome-Wide Association Study for Autism Spectrum Disorder in Taiwanese Han Population. *PLoS One* **10**, (2015).

41. Almandil, N. B. *et al.* Integration of Transcriptome and Exome Genotyping Identifies Significant Variants with Autism Spectrum Disorder. *Pharmaceuticals* **15**, (2022).

42. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics 2019 51:12* **51**, 1670–1678 (2019).

693 43. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci.
694 *Nature 2014 511:7510* **511**, 421–427 (2014).

695 44. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for
696 schizophrenia. *Nature Genetics 2013 45:10* **45**, 1150–1159 (2013).

697 45. Betcheva, E. T. *et al.* Whole-genome-wide association study in the Bulgarian
698 population reveals HHAT as schizophrenia susceptibility gene. *Psychiatr Genet*
699 **23**, 11–19 (2013).

700 46. Goes, F. S. *et al.* Genome-wide association study of schizophrenia in Ashkenazi
701 Jews. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*
702 **168**, 649–659 (2015).

703 47. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature*
704 *2009 460:7256* **460**, 744–747 (2009).

705 48. Aberg, K. A. *et al.* A Comprehensive Family-Based Replication Study of
706 Schizophrenia Genes. *JAMA Psychiatry* **70**, 573–581 (2013).

707 49. Levinson, D. F. *et al.* Genome-wide association study of multiplex schizophrenia
708 pedigrees. *American Journal of Psychiatry* **169**, 963–973 (2012).

709 50. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci
710 for schizophrenia. *Nature Genetics 2017 49:11* **49**, 1576–1583 (2017).

711 51. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology
712 in schizophrenia. *Nature 2022 604:7906* **604**, 502–508 (2022).

713 52. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-
714 intolerant genes and in regions under strong background selection. *Nature*
715 *Genetics 2018 50:3* **50**, 381–389 (2018).

716 53. Liu, J. *et al.* Genome-wide association study followed by trans-ancestry meta-
717 analysis identify 17 new risk loci for schizophrenia. *BMC Med* **19**, 1–15 (2021).

718 54. Blokland, G. A. M. *et al.* Sex-Dependent Shared and Nonshared Genetic
719 Architecture Across Mood and Psychotic Disorders. *Biol Psychiatry* **91**, 102–117
720 (2022).

721 55. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by
722 genome-wide association and follow-up. *Nature Genetics 2008 40:9* **40**, 1053–
723 1055 (2008).

724 56. Athanasiu, L. *et al.* Gene variants associated with schizophrenia in a Norwegian
725 genome-wide study are replicated in a large European cohort. *J Psychiatr Res* **44**,
726 748–753 (2010).

727 57. Kirov, G. *et al.* A genome-wide association study in 574 schizophrenia trios using
728 DNA pooling. *Molecular Psychiatry 2009 14:8* **14**, 796–803 (2008).

58. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics 2019 51:5* **51**, 793–803 (2019).

59. Wu, Y. *et al.* Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Translational Psychiatry 2020 10:1* **10**, 1–11 (2020).

60. Ikeda, M. *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Molecular Psychiatry 2018 23:3* **23**, 639–647 (2017).

61. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics 2021 53:6* **53**, 817–829 (2021).

62. Cichon, S. *et al.* Genome-wide Association Study Identifies Genetic Variation in Neurocan as a Susceptibility Factor for Bipolar Disorder. *The American Journal of Human Genetics* **88**, 372–381 (2011).

63. Li, H. J. *et al.* Novel Risk Loci Associated With Genetic Risk for Bipolar Disorder Among Han Chinese Individuals: A Genome-Wide Association Study and Meta-analysis. *JAMA Psychiatry* **78**, 320–330 (2021).

64. Bigdeli, T. B. *et al.* Genome-Wide Association Studies of Schizophrenia and Bipolar Disorder in a Diverse Cohort of US Veterans. *Schizophr Bull* **47**, 517–529 (2021).

65. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* **18**, 1091–1102 (2019).

66. Pan, H. *et al.* Genome-wide association study using whole-genome sequencing identifies risk loci for Parkinson's disease in Chinese population. *npj Parkinson's Disease 2023 9:1* **9**, 1–11 (2023).

67. Beecham, A. H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics 2013 45:11* **45**, 1353–1360 (2013).

68. Patsopoulos, N. A. *et al.* Multiple Sclerosis Genomic Map implicates peripheral immune cells & microglia in susceptibility. *Science* **365**, (2019).

69. Mishra, A. *et al.* Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature 2022 611:7934* **611**, 115–123 (2022).

70. Tcheandjieu, C. *et al.* Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nature Medicine 2022 28:8* **28**, 1679–1692 (2022).

71. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nature Genetics 2020 52:11* **52**, 1169–1177 (2020).

72. Klarin, D. *et al.* Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nature Genetics 2017 49:9* **49**, 1392–1397 (2017).

73. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics 2017 49:9* **49**, 1385–1391 (2017).

74. Howson, J. M. M. *et al.* Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nature Genetics 2017 49:7* **49**, 1113–1119 (2017).

75. Van Der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res* **122**, 433–443 (2018).

76. Verweij, N., Eppinga, R. N., Hagemeijer, Y. & Van Der Harst, P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Scientific Reports 2017 7:1* **7**, 1–9 (2017).

77. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature 2021 599:7886* **599**, 628–634 (2021).

78. Nam, K., Kim, J. & Lee, S. Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. *Cell Genomics* **2**, 100189 (2022).

79. Conway, O. J. *et al.* ABI3 and PLCG2 missense variants as risk factors for neurodegenerative diseases in Caucasians and African Americans. *Mol Neurodegener* **13**, 1–12 (2018).

80. Karahan, H. *et al.* The effect of Abi3 locus deletion on the progression of Alzheimer's disease-related pathologies. *Front Immunol* **14**, 1102530 (2023).

81. Satoh, J. I. *et al.* Microglia express ABI3 in the brains of Alzheimer's disease and Nasu-Hakola disease. *Intractable Rare Dis Res* **6**, 262 (2017).

82. Drobny, A. *et al.* The role of lysosomal cathepsins in neurodegeneration: Mechanistic insights, diagnostic potential and therapeutic approaches. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1869**, 119243 (2022).

83. Cermak, S. *et al.* Loss of Cathepsin B and L Leads to Lysosomal Dysfunction, NPC-Like Cholesterol Sequestration and Accumulation of the Key Alzheimer's Proteins. *PLoS One* **11**, e0167428 (2016).

84. Sundelöf, J. *et al.* Higher Cathepsin B Levels in Plasma in Alzheimer's Disease Compared to Healthy Controls. *Journal of Alzheimer's Disease* **22**, 1223–1230 (2010).

85. Sun, Y. *et al.* Translational Study of Alzheimer's Disease (AD) Biomarkers from Brain Tissues in AβPP/PS1 Mice and Serum of AD Patients. *Journal of Alzheimer's Disease* **45**, 269–282 (2015).

86. Morena, F. *et al.* A Comparison of Lysosomal Enzymes Expression Levels in Peripheral Blood of Mild- and Severe-Alzheimer's Disease and MCI Patients: Implications for Regenerative Medicine Approaches. *Int J Mol Sci* **18**, 1806 (2017).

87. Zhang, Y. *et al.* An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *The Journal of Neuroscience* **34**, 11929 (2014).

88. Ulland, T. K. & Colonna, M. TREM2 — a key player in microglial biology and Alzheimer disease. *Nature Reviews Neurology 2018 14:11* **14**, 667–675 (2018).

89. Zhao, Y. *et al.* TREM2 Is a Receptor for β-Amyloid that Mediates Microglial Function. *Neuron* **97**, 1023-1031.e7 (2018).

90. McQuade, A. *et al.* Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nature Communications 2020 11:1* **11**, 1–17 (2020).

91. Schmid, C. D. *et al.* Heterogeneous expression of the triggering receptor expressed on myeloid cells-2 on adult murine microglia. *J Neurochem* **83**, 1309–1320 (2002).

92. Winfree, R. L. *et al.* TREM2 gene expression associations with Alzheimer's disease neuropathology are region-specific: implications for cortical versus subcortical microglia. *Acta Neuropathol* **145**, 733–747 (2023).

93. Cady, J. *et al.* TREM2 Variant p.R47H as a Risk Factor for Sporadic Amyotrophic Lateral Sclerosis. *JAMA Neurol* **71**, 449–453 (2014).

94. Gratuze, M., Leyns, C. E. G. & Holtzman, D. M. New insights into the role of TREM2 in Alzheimer's disease. *Molecular Neurodegeneration 2018 13:1* **13**, 1–16 (2018).

95. Lill, C. M. *et al.* The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. *Alzheimer's & Dementia* **11**, 1407–1416 (2015).

96. Liu, X. *et al.* The protective role of miR-132 targeting HMGA2 through the PI3K/AKT pathway in mice with Alzheimer's disease. *Am J Transl Res* **13**, 4632 (2021).

97. Zhang, J., Li, X., Xiao, J., Xiang, Y. & Ye, F. Analysis of gene expression profiles in Alzheimer's disease patients with different lifespan: A bioinformatics study focusing on the disease heterogeneity. *Front Aging Neurosci* **15**, (2023).

98. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).

99. Cavalli-Sforza, L. L. Human evolution and its relevance for genetic epidemiology. *Annu Rev Genomics Hum Genet* **8**, 1–15 (2007).

100. Delaneau, O. *et al.* Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nature Communications 2014 5:1* **5**, 1–9 (2014).

101. Auton, A. *et al.* A global reference for human genetic variation. *Nature 2015 526:7571* **526**, 68–74 (2015).

102. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics* **93**, 278–288 (2013).

103. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics 2011 43:5* **43**, 491–498 (2011).

104. DeRosa, B. A. *et al.* Generation of two iPSC lines (UMi038-A & UMi039-A) from siblings bearing an Alzheimer's disease-associated variant in SORL1. *Stem Cell Res* **62**, 102823 (2022).

105. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

106. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **109**, 21.29.1 (2015).

107. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4* **9**, 357–359 (2012).

108. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 1–21 (2014).

109. Peterson, H., Kolberg, L., Raudvere, U., Kuzmin, I. & Vilo, J. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, (2020).

110. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

872