Received: 26 September 2024 | Revised: 14 January 2025 | Accepted: 29 January 2025

DOI: 10.1002/alz.70031

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Comparing Alzheimer's genes in African, European, and Amerindian induced pluripotent stem cell–derived microglia

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, Villa EL Salvador, Peru

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

[5]Universidad Central del Caribe, Bayamón, Puerto Rico

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

Correspondence
Jeffery M. Vance, John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida, USA.
Email: jvance@miami.edu

Funding information
National Institute on Aging, Grant/Award Numbers: U01-AG072579, RF1- AG059018, U01-AG066767, U01-AG052410, R56-AG072547, R01-AG070864; Alzheimer Association, Zenith Award

Abstract

INTRODUCTION: Genome-wide association studies (GWAS) studies in Alzheimer's disease (AD) demonstrate ancestry-specific loci. Previous studies in the regulatory architecture have only been conducted in Europeans (EUs), thus studies in additional ancestries are needed. Given the prevalence of AD genes expressed in microglia, we initiated our studies in induced pluripotent stem cell (iPSC) -derived microglia.

METHODS: We created iPSC-derived microglia from 13 individuals of either high Amerindian (AI), African (AF), or EU global ancestry, including both AD and controls. RNA-seq, ATAC-seq, and pathway analyses were compared between ancestries in both AD and non-AD genes.

RESULTS: Twelve AD genes were differentially expressed genes (DEGs) and/or accessible between ancestries, including *ABI3*, *CTSB*, and *MS4A6A*. A total of 5% of all genes had differential ancestral expression, but differences in accessibility were less than 1%. The DEGs were enriched in known AD pathways.

DISCUSSION: This resource will be valuable in evaluating AD in admixed populations and other neurological disorders and understanding the AD risk differences between populations.

KEYWORDS
Alzheimer's disease, ATAC-seq, diversity, genetic ancestry, genetic regulatory architecture, iPSC-derived microglia, RNA-seq

Highlights

- First comparison of the genomics of AI, AF, and EU microglia.
- Report differences in expression and accessibility of AD genes between ancestries.
- Ancestral expression differences are greater than differences in accessibility.
- Good transcriptome correlation was seen between brain and iPSC-derived microglia.
- Differentially expressed AD genes were in known AD pathways.

1 | BACKGROUND

Alzheimer's disease (AD) affects millions of people worldwide with currently ~11% of the U.S. population (65 and older) affected. It is predicted that over 150 million individuals will be affected by AD worldwide by 2050. 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 United States, AD affects 19% of AA, 14% of HI, and 10% of NHW individuals.[1] Further, over the next 25 years, the greatest growth in AD will be in African and South American populations. Genetic diversity and admixture play important roles in disease risk. AA genomes are typically admixed between African (AF) and European (EU) ancestries while HI ancestry encompasses a three-way admixture of EU, Amerindian (AI), and AF ancestries.[2] Consistent with this, there are ancestry-related differences in the genetic architecture of AD.[3] 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*,[4–8] *MPDZ*,[9] and *IGF1R*.[4,9] 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 genome-side association studies (GWAS)[10–12] 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 EU ancestry. The different population risk profiles for AD of apolipoprotein E (*APOE*) *e4* 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 AF ancestry relative to EU ancestry was due to differences in the local genomic ancestry surrounding the *APOEe4* allele.[13,14] Subsequently, it was found that EU local genomic ancestry carriers of *APOEe4* had higher *APOEe4* expression and more open chromatin accessibility than that of AF local ancestry (LA) carriers,[15,16] supporting the recent report that lower expression of *APOEe4* is tied to lower risk[17] 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 (MGL) in the AD disease process. Autopsy studies found an elevated proportion of activated microglia significantly correlated with pathological AD,[18] specifically the total Aβ load and number of neuritic plaques. Additionally, a large number of reported AD GWAS genes are expressed in microglia,[19,20] further supporting their role in AD pathology. MGL 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.[21] Specifically, in the context of AD, MGL are particularly important for Aβ plaque clearance, neuroprotection, inflammatory responses, and synaptic homeostasis.[22]

Here, we report an examination of induced pluripotent stem cells (iPSC)-derived MGL from AF, EU, and AI ancestries, expanding on our previous studies of single nuclei RNA-seq and single nuclei ATAC-seq on postmortem MGL from the frontal cortex on AF and EU genomes.[15,16] Additionally, as iPSC-derived cells have become important models for human neurodegenerative research, we performed a comparison between our iPSC-derived MGL 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 AF, EU, and AI populations, as well as admixed populations of AA and HI individuals.

2 | METHODS

2.1 | 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 iPSCs 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).

2.2 | Global ancestry ascertainment

We calculated the admixture proportions using a model-based clustering algorithm, as implemented in the ADMIXTURE software.[23] A

supervised ADMIXTURE analysis was performed at $K = 4$, incorporating four reference populations: 104 AF, 84 EU, 108 AI, and 102 East Asian individuals from the Human Genome Diversity Project reference populations.

2.3 | LA ascertainment

To infer LA, we first merged our dataset with the Human Genome Diversity Project reference panel, including EU, AF, and AI reference populations.[24] Next, we phased the combined data using SHAPEIT4 with default settings, referencing the 1000 Genomes Phase 3 reference panel.[25,26] Finally, we estimated LA at each genomic locus using RFMix v2 software.[27]

2.4 | Whole genome sequencing

DNA was extracted from all individual cell lines using the QIAamp DNA Blood Kit (QIAGEN, #51104) according to the manufacturer's instructions. A total of 1.5 μg of DNA was submitted for whole genome sequencing (WGS) at the Center for Genome Technology (CGT) Sequencing Core at the HIHG using standard Illumina polymerase chain reaction (PCR)-free library prep and sequencing protocols on the NovaSeq6000 followed by a bioinformatics pipeline incorporating the GATK Best Practices analysis recommendations.[28] 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 10 AD genes that had differential gene expression (Table S1).

2.5 | iPSC 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 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 (ICC), STR profiling, karyotyping, and mycoplasma testing as previously described.[29] PBMC isolation and reprogramming was performed at the HIHG iPSC Core at the University of Miami. Validation analyses were performed by the HIHG-iPSC Core and WiCell.

2.6 | Differentiation of iPSCs to MGL

iPSCs were differentiated into hematopoietic progenitor cells (HPCs) and subsequently into MGL as previously described[30] with minor modifications.

RESEARCH IN CONTEXT

1. **Systematic review**: The mechanisms by which different ancestry groups have different Alzheimer's disease (AD) risk are not understood but might be due to ancestry-specific genetic risk factors. We reviewed the literature regarding ancestry-specific risk factors that could explain the different disease risk.
2. **Interpretation**: Our study provides novel findings on the regulatory architectures of these ancestry groups, specifically Amerindian (AI), African (AF), and European (EU), in the context of AD pathology. We highlight potential ancestry-specific risk factors and pathology relevant mechanistic differences in these ancestry groups. Moreover, we demonstrate the use of induced pluripotent stem cell (iPSC)-derived microglia to study human disease beyond AD.
3. **Future directions**: Further functional validation and investigation of the relationship between these ancestry-specific genetic risk factors and their role in AD pathology will be critical to expand upon this work.

In brief, feeder-free iPSCs were cultured and expanded in StemFlex medium (Gibco, #A3349401) on vitronectin (10 μg/ml, Gibco, #A31804) coated cell culture-treated plates. On day -1, iPSCs were passaged with 0.5 M ethylenediaminetetraacetic acid (EDTA) onto Matrigel-coated (Corning, #354277) 12-well plates at a density of 10-20 aggregates/cm^2 (> 50 μm in size). On day 0, if 4–10 colonies/cm^2 adhered, the StemFlex medium was replaced with 1 mL/well of HPC medium A (Basal medium with supplement A (1:200), STEMCELL Technologies, #05310). Half-medium change was carried out 48 h 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 MGL differentiation (day 12 of HPC differentiation), HPCs were plated at 22,000 cells/cm^2 onto a Matrigel-coated six-well plate containing 2 mL 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 min. The cell pellet was resuspended in 2 mL/well of fresh MGL differentiation medium and transferred to a fresh Matrigel-coated six-well plate. Cells were supplemented with 1 mL of media every second day until day 22. MGL cells were collected, resuspended in 2 mL of MGL maturation medium (Basal Medium with supplement 1 (1:9), 2, and 3 (1:225); STEMCELL Technologies, #100-0020), and re-plated for assays onto fresh Matrigel-coated six-well plates. Lastly, on day 26, MGL were harvested for ICC, bulk RNA-, and ATAC-sequencing.

2.7 | 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 min at 300 x *g*. 600 µl of RLT buffer (including β-Mercaptoethanol at 1/100) was used to collect semi-attached MGL and subsequently resuspend the cell pellet from the previous step. Cells were briefly vortexed for 1 min and homogenized by loading the lysate into a QIAshredder spin column (QIAGEN, #79656) and centrifuging for 2 min 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 s at 8000 x *g*. 350 µl of Buffer RW1 was added to the same spin column and centrifuged for 15 s at 8000 x *g*. Following this, 80 µl of DNAse I incubation mix (70 µl of RDD buffer and 10 µl of DNAse I, QIAGEN, #79254) were added to the spin column and incubated at room temperature (RT) for 15 min. Buffer RW2 (350 µl) was transferred to the spin column and centrifuged for 15 s at 8000 x *g*. 500 µl of RPE buffer were loaded into the column followed by a centrifugation step of 30 s at 8000 x *g*. The previous step was repeated once again but centrifuged for 2 min at 8000 x *g* to ensure all residual ethanol was removed. The RNeasy spin column was transferred to a new 1.5 mL collection tube and 30 µl of RNAse-free water were added to the column to elute the bound RNA. Lastly, the spin column was centrifuged at 8000 x *g* for 1 min 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).

2.8 | Bulk RNA sequencing

RNA libraries were prepared at the John P. HIHG CGT (University of Miami, FL, USA) 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 60 ng via QuBit and 16 PCR cycles. The normalized libraries were sequenced as paired end 100 bp reactions targeting 30 million reads/sample on the Illumina NovaSeq 6000 (Illumina, CA, USA). 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),[31] and gene counts quantified against the GENCODEv35 gene annotation release using the GeneCounts module implemented in STAR.

2.9 | Bulk ATAC-sequencing

Cultured cells were treated with DNase I (200 U/mL; QIAGEN, #79254) at 37°C for 30 min. The treated cells were then harvested and pelleted at 400 x *g* for 5 min at 4°C. The cell pellet was carefully washed in cold 1x phosphate buffered saline (PBS). The cells were re-pelleted as described before and then lysed in 100 µl of lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20, and 0.01% Digitonin) on ice for 5 min. Next, the lysed MGL were washed in 1 mL of wash buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% Tween-20), and 100,000 nuclei were pelleted at 500 x *g* for 10 min at 4°C. The nuclei were incubated at 37°C for 30 min at 1000 rpm in 100 µl of Transposition mix (2x Tagment DNA Buffer, 1x PBS, 0.1% v/v Tween-20, 0.01% v/v Digitonin, and 5 µl of Tagment DNA Enzyme 1). The transposed DNA was purified using the MinElute PCR Purification kit (QIAGEN, #28004) and eluted in 10 µl of Elution Buffer. The purified transposed DNA was combined with 25 µM of Custom Adapter 1 (no primer mix), 25 µM 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 min, 98°C for 30 s, and 5 cycles of 98°C for 30 s, 63°C for 30 s, and 72°C for 1 min. The additional number of cycles required was determined as described in Ref.32 and run with the same conditions abovementioned. The amplified libraries were purified with the MinElute PCR Purification kit and eluted in 20 µl 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 Table S2 for full adapter sequences. Libraries were sequenced in paired end 100 bp 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),[33] duplicates marked with Picard (v2.1.1) (https://broadinstitute.github.io/picard/), and peaks called using MACS2 (v2.2.7.1).[33] 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.

2.10 | Differential expression and accessibility analyses

Differential expression and accessibility analyses were carried out across the different ancestral populations using DESeq2 (version 3.17) package[34] in R language environment (version 4.2.1). We used DESeq2 default parameters and controlled for batch differences (design = ~batch + ancestry). Three comparisons were run: AF versus EU, AI versus AF, and AI versus EU. Genes that were significantly expressed and/or accessible were identified with a false discovery rate (FDR) adjusted *p*-value of < 0.05.

2.11 | Functional enrichment pathway analysis

Functional enrichment analysis was done with the R library gprofiler2.[35] We extracted gene symbols of differentially expressed genes (DEGs) 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 Tables S3–S5.

2.12 | ATAC peak annotation

The function *annotatePeak* from Chipseeker R library[36] 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".

2.13 | ICC and fluorescence imaging

Cultured MGL cells were fixed with 4% formaldehyde for 15 min at RT and washed with 1x PBS. Cells were permeabilized for 10 min with PBS-T solution (0.1% Triton X and 1x PBS). The MGL cells were then incubated in blocking buffer (1x PBS and 5% normal donkey serum) for 1 h 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 specific 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 h 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 Table S6 for details on all antibodies used for ICC analysis.

2.14 | 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. Additionally, principal component analyses (PCA) were carried out using DESEq2 on regularized log-transformed data after batch effect correction using ComBat-seq function of the *sva* package[37] to determine the individual level variation in gene expression and chromatin accessibility across all samples.

2.15 | Correlation analyses between iPSC-derived MGL and other cell types

Correlation analyses between iPSC-derived MGL (iMGL) and brain cell types were performed using Spearman correlation analyses. Specifically, we calculated the average expression of all 13 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.

3 | RESULTS

3.1 | Differentiation and validation of iMGL

We differentiated 13 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 AI, 5 EU, and 4 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

TABLE 1 iMGL cell line information.

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

Abbreviations: AD, Alzheimer's disease; AF, African; AI, Amerindian; EU, European; iMGL, induced pluripotent stem cell-derived microglia; MCI, mild cognitive impairment.

TABLE 2 Correlation analysis between iMGL from our study and other cell types

Cell type	rho	Source
Fetal brain (Cerebrum)		
MGL	0.711	Cao et al., (2020).
Astrocytes	0.580	
Excitatory neurons	0.599	
Inhibitory neurons	0.573	
Oligodendrocytes	0.563	
Vascular endothelial	0.678	
Adult brain		
MGL	0.637	Griswold 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	

Note: Note that all 13 iMGL lines were included for these comparisons and the *p*-value was below 2.2×10^{-16} for all comparisons. The adult brain data are derived from Brodmann Area 9 of both AF and EU ancestry.
Abbreviations: AF, African; EU, European; iMGL, induced pluripotent stem cell-derived microglia; MGL, microglia; OPC, oligodendrocyte precursor cells; VLMC, vascular leptomeningeal cells.

genes (*APP*, *ABCA7*, *MAPT*, *PSEN1*, *PSEN2*, *SORL1*, and *TREM2*; Table S1) that could affect the genetic regulatory architecture (GRA).

All 13 iMGL cell lines were further validated with MGL cell-specific lineage markers using ICC (*PU.1* [*SPI1*], *TMEM119*, *TREM2*, *and P2RY12*; Figure S1). All MGL 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.[30] In addition, we also verified that these cells did not express markers for other brain cell types (astrocytes, oligodendrocytes, and neurons; Figure S2 and Table S7).

3.2 | Brain MGL versus iMGL

We compared the transcriptomic profiles of our iMGL to both fetal brain and adult brain cell types.[16,38] In both comparisons, we observed the highest correlation between iMGL and fetal brain MGL ($\rho = 0.711$), and Adult brain MGL ($\rho = 0.637$) compared to other brain cell types (Table 2). These data suggest these iMGL recapitulate well the transcriptomic profiles observed in brain MGL and are a good study model.

3.3 | 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, DEGs between ancestries (FDR < 0.05). Specifically, we identified 971 DEGs between AI and AF, 320 between AI and EU ancestries, and 62 DEGs between AF and EU (Figure 1A and B; Tables S8–S10).

We focused on genes previously identified in AD GWAS studies.[4,39–43] Of the 121 AD GWAS genes (Table S11), 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 MGL of EU and AF ancestries,[16] we did not observe differential expression for AD risk-modifying genes between AF and EU in our iPSC-derived MGL. 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).

3.4 | 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; Tables S12–S14). We observed an enrichment in DAPs between AI and EU in chromosome 17 (12.28%, chi-squared *p*-value = 0.038) and chromosome 13 (7.02%, chi-squared *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 Table S15). 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; Table S16). Interestingly, in the context of genes associated with AD, we only detected two DAGs (*PRDM7* and *SCIMP*) between AI and AF and one DAG between AI and EU (*PRDM7*) (Figure 2C).

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 (Figure S3). We performed LA analyses surrounding the *PRDM7* locus (± 500 kb) to further investigate whether it could explain the differences in chromatin accessibility (Table S17). We observed that samples 1–3 of AI global ancestry, have homozygote AI LA for the *PRDM7* locus while sample 4

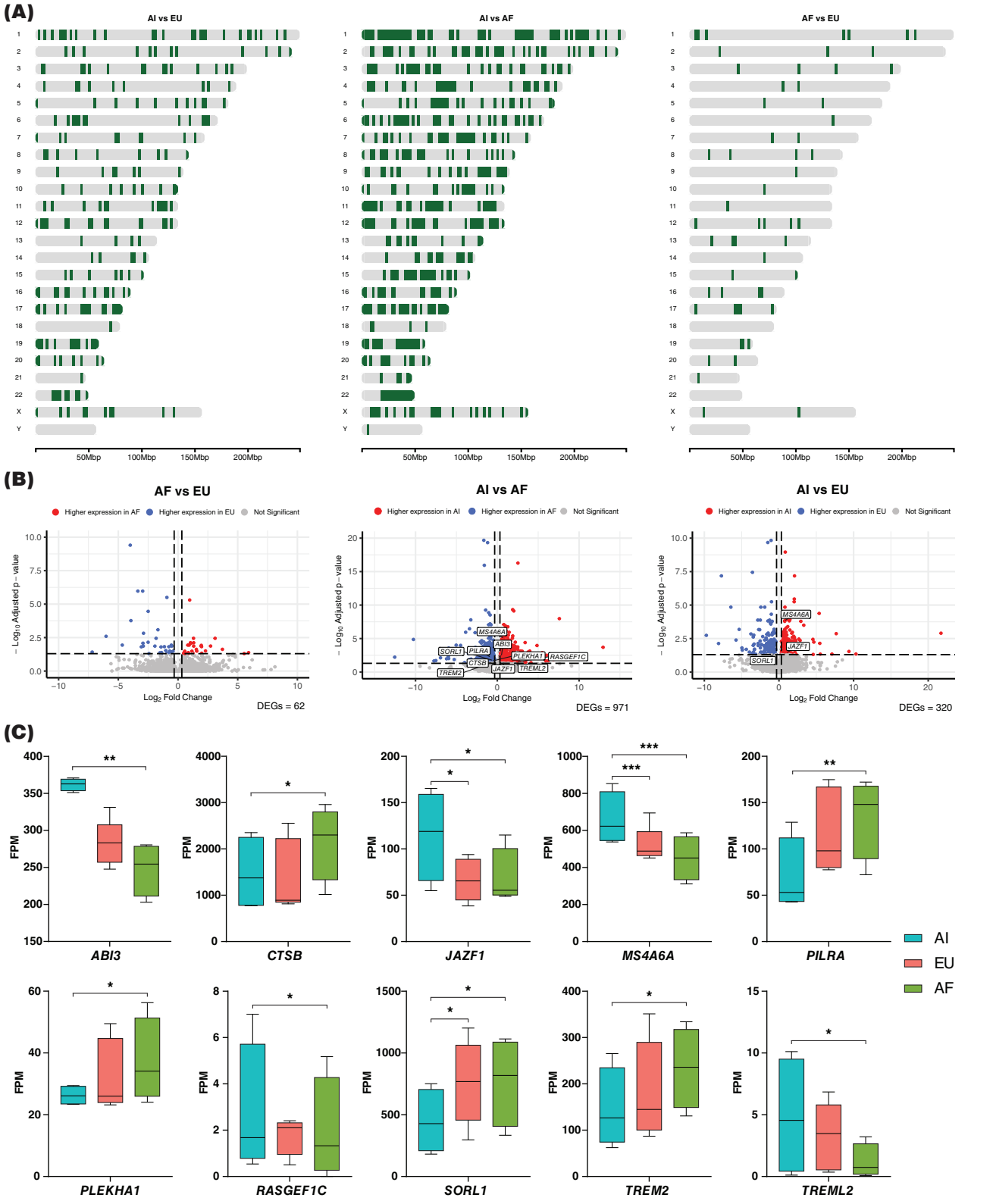

FIGURE 1 Gene expression across ancestries. (A) Chromosome maps per pair-wise ancestral comparison demonstrating the distribution of DEGs genome-wide. The dark green color represents DEGs. (B) Volcano plots representing gene expression (log_2 fold change) per pair-wise comparison between ancestries (AF 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 $p \leq 0.05$ (*), $p \leq 0.01$ (**), and $p \leq 0.001$ (***). AF, Africans; AI, Amerindian; DEGs, differentially expressed genes; EU, European; FDR, false discovery rate; FPM, fragments per million.

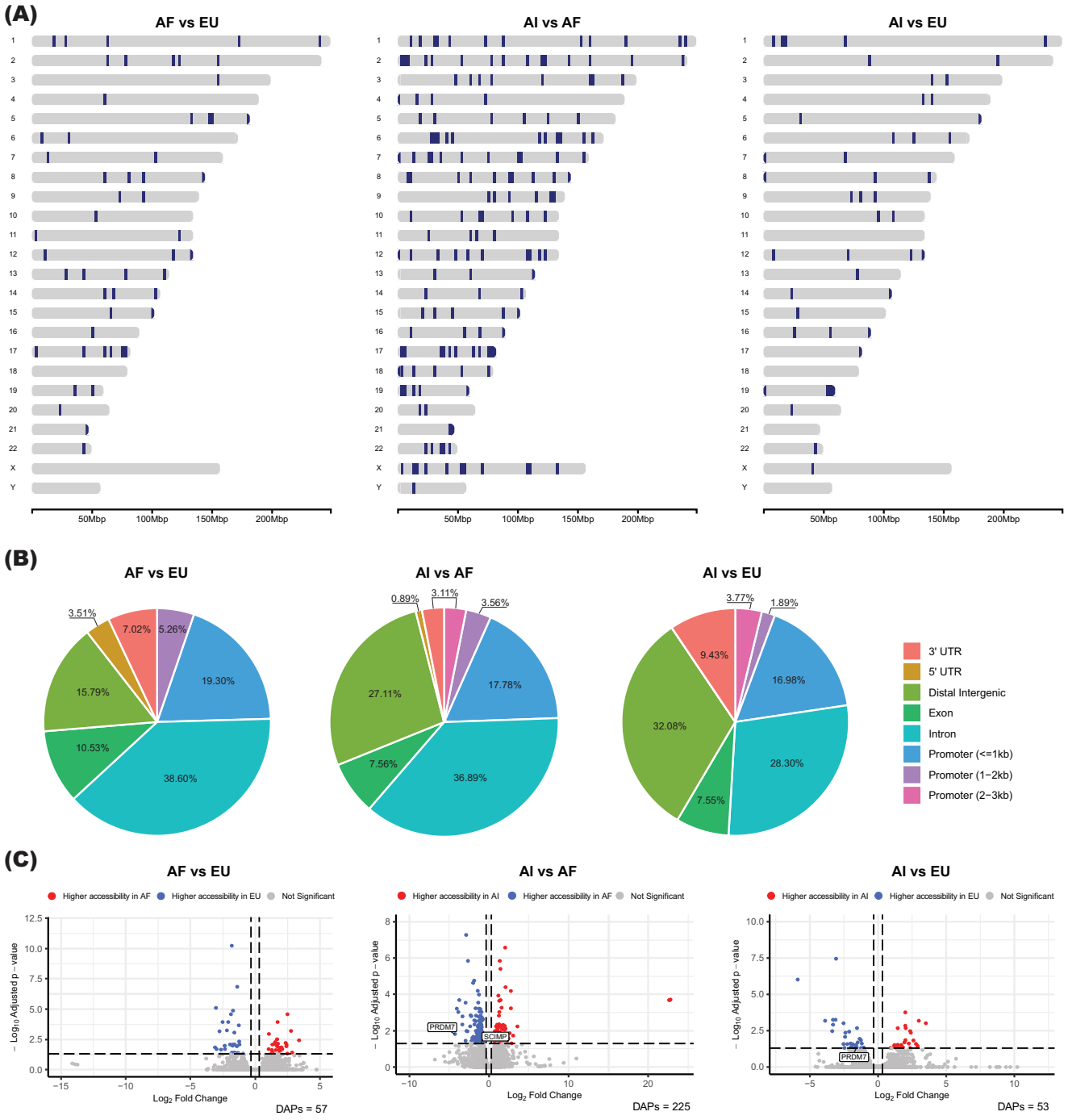

FIGURE 2 Chromatin accessibility across ancestries. (A) Chromosome maps per pair-wise ancestral group comparison demonstrating the distribution of DAGs genome-wide. The dark blue color represents DAGs. (B) Pie charts illustrate the regions of the genome in which the differentially accessible peaks lie 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. AD, Alzheimer's disease; AF, Africans; AI, Amerindian; DAGs, differentially accessible genes; EU, European; FDR, false discovery rate.

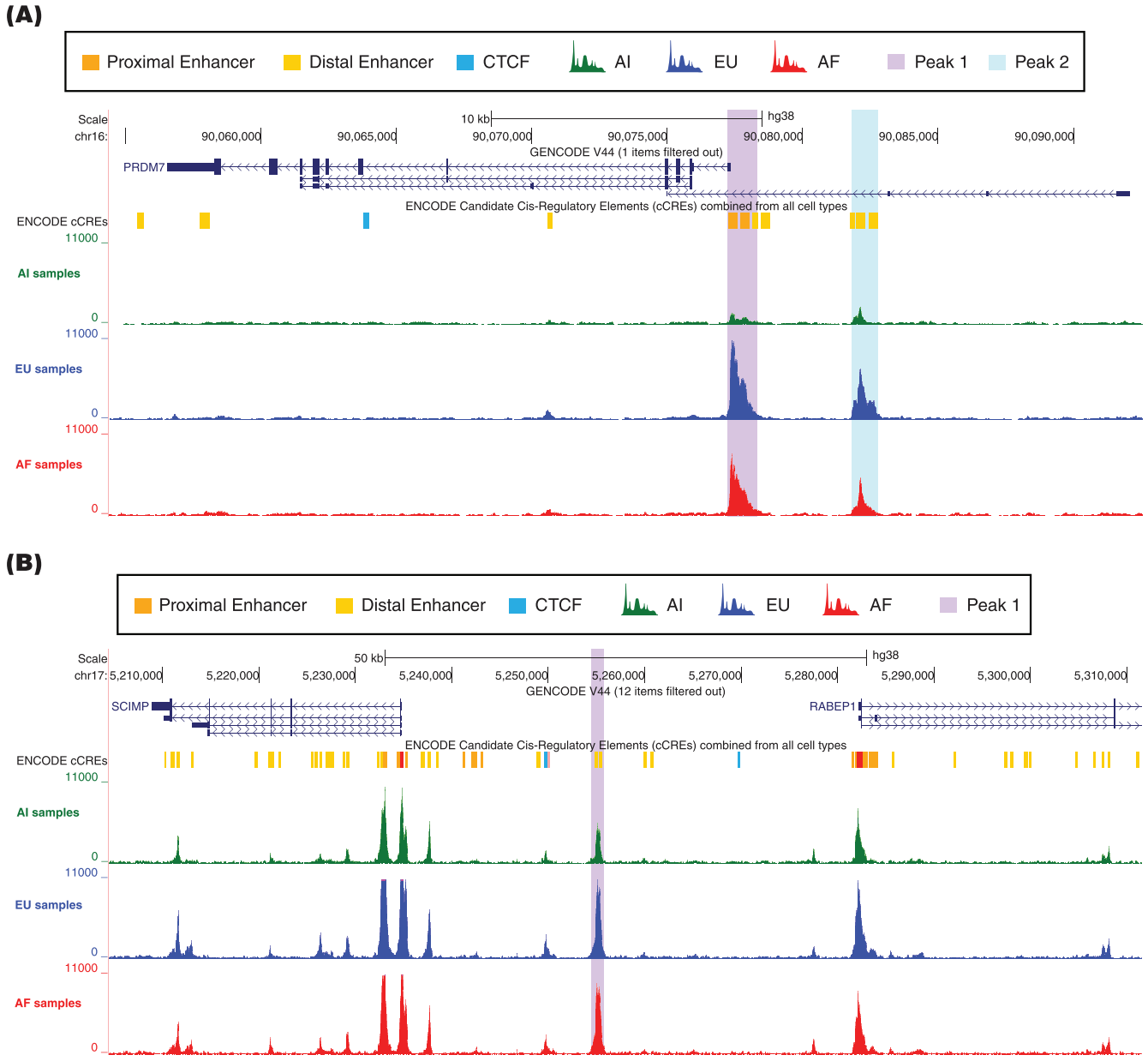

FIGURE 3 Differentially accessible peaks in AD-risk modifying genes across ancestries. (A) Differential chromatin accessible peaks in *PRDM7*. (B) Differential chromatin accessible peak in a distal intergenic enhancer of *SCIMP*. Note that the peaks represent merged data of all individuals within the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

has AF LA for both haplotypes in this locus aligning with the chromatin accessibility observations within the AF global ancestry group. While these data suggest that the AF 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 EU 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* (~20 kb; Figure 3B). We did not observe LA differences within the same global ancestry group for the *SCIMP* locus (Table S17) which could explain chromatin accessibility differences seen between global ancestry groups in this region (Figure S4).

3.5 | 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 (Table S3) 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

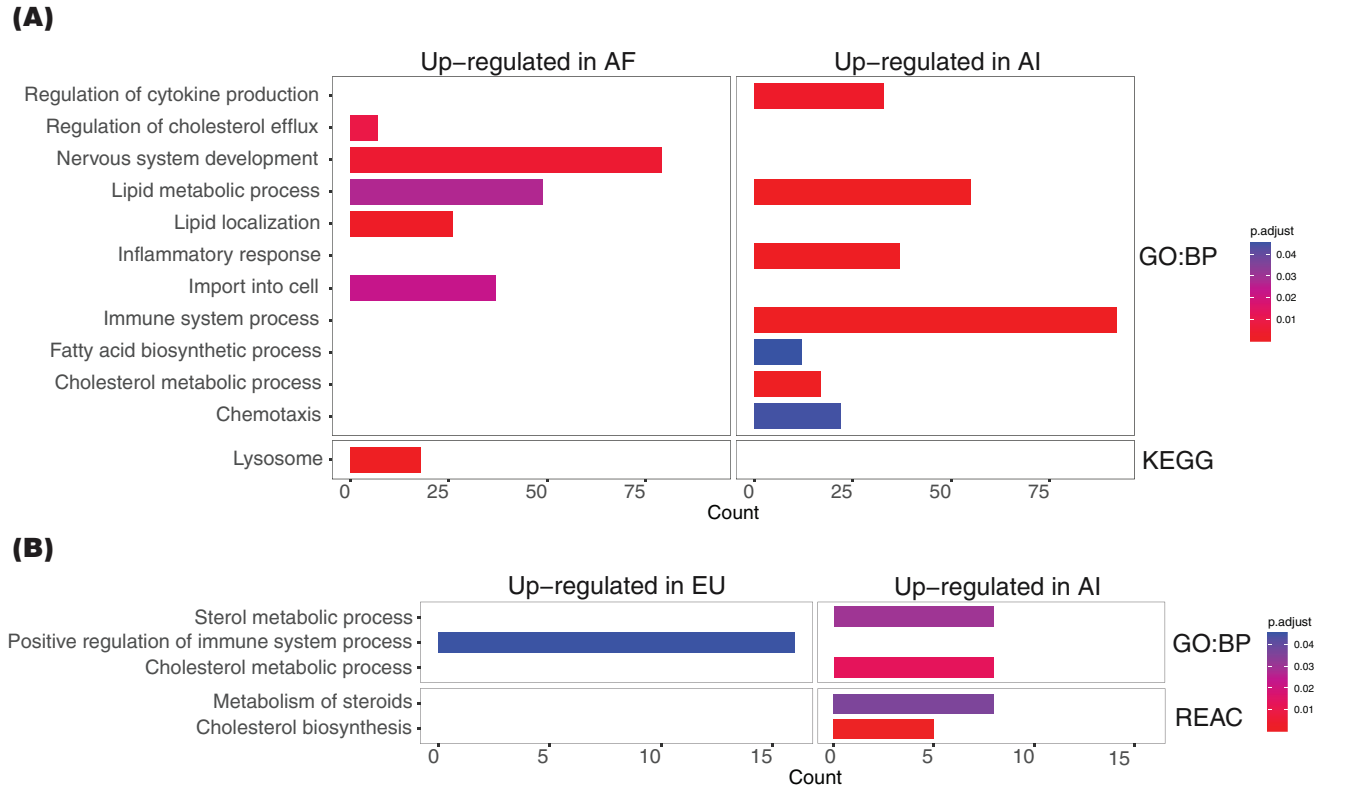

FIGURE 4 Functional enrichment pathway enrichment across ancestries relevant to AD. Pathway enrichment analyses between (A) AI and AF, and (B) AI and EU. See Tables S4 and S5, respectively, for all significantly enriched pathways. AD, Alzheimer's disease; AF, Africans; AI, Amerindian; EU, European.

and metabolism, cholesterol biosynthesis and metabolism, lipid transport and metabolism, and phagocytosis—all highly relevant processes in AD pathology (Figure 4 and Tables S4 and S5).

3.6 | Regulatory architecture in iMGL

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 Figure S5). None of the overlapping DEGs and DAGs were from known AD GWAS genes (Table S11). We observed that all overlapping DAGs and DEGs between AF and EU, and between AI and EU lay in promoter regions (Tables S18 and S19, respectively), while there was a wider genomic distribution for those overlapping DAGs and DEGs between AI and AF (Table S20).

Additionally, we screened nominated targets based on functional evidence available in the AGORA portal (https://agora.adknowledgeportal.org/genes/nominated-targets). We filtered gene names that had > 1 nomination, allowing us to screen a total of 163 nominated targets. Between AI and AF, we identified 13 AGORA genes to be differentially expressed (only 3 of those were in common with the AD GWAS genes; Table S8) and 3 AGORA genes to be differentially accessible (Table S12). Interestingly, when looking at the AF and EU comparison for which we had not identified any significant DEGs/DAGs, we also did not identify any differential gene expression/accessibility when screening the AGORA genes. Lastly, between AI and EU, we identified two DEGs in the AGORA nominated targets (one in common with the AD GWAS genes; Table S9) but no genes to be differentially accessible.

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]; Figure S6). Further, we performed PCA on our gene expression and chromatin accessibility data to determine individual variation across all samples. Overall, we observed greater variation in PC1 that is likely attributable to global ancestry genetic variation (Figure S7). For instances, when looking at samples with similar characteristics (e.g., sex, *APOE* genotype, clinical diagnosis, and so forth) such as samples 1 and 2 or samples 7 and 8, we observed a bigger variation in PC1 for samples 7 and 8 compared to samples 1 and 2 which are closer in global ancestry percentage.

3.7 | 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 seven DEGs between non-cognitively

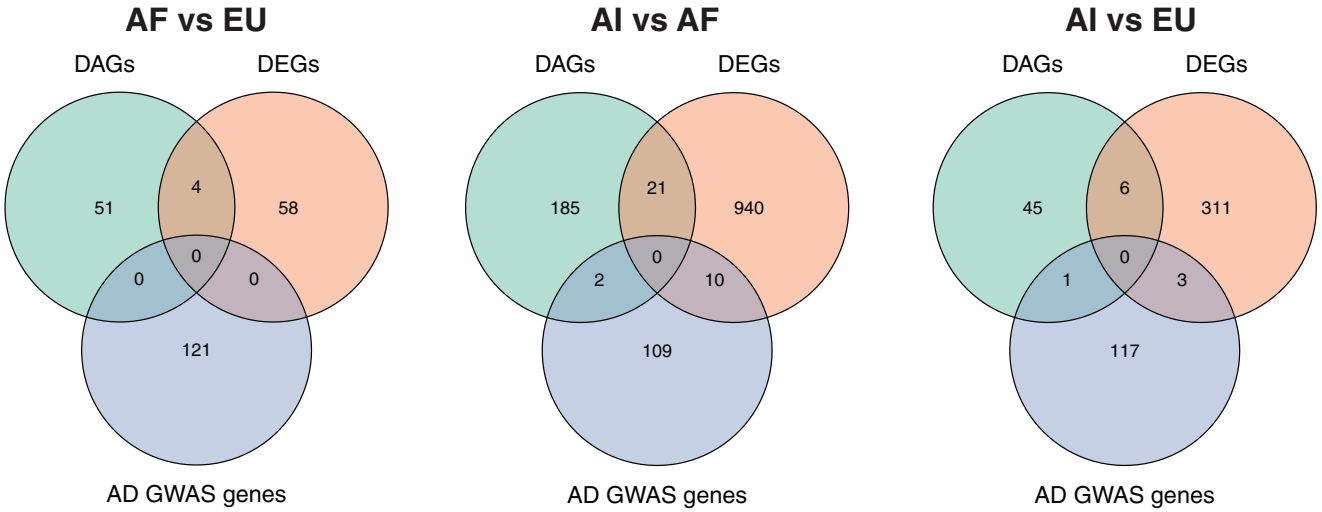

FIGURE 5 Overlap between differentially accessible ATAC-seq genes, differentially expressed RNA-seq genes, and AD GWAS genes between ancestry-group comparisons. AD, Alzheimer's disease; GWAS, genome-wide association studies.

impaired individuals and AD samples (Table S21). None were previously identified as AD risk-modifying genes. Differential expression analysis between *APOEe3* and *APOEe4* homozygote carriers revealed seven DEGs (Table S22). 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 (Figure S8). The sex comparison revealed a total of 116 DEGs between males and females (Table S23), 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 (Table S24), one DAP/DAG between cases and controls (Table S25), and 136 DAPs between males and females (90 DAGs; Table S26). None of these peaks have been previously connected to either AD or *APOE* genotype. Lastly, we observed an overlap between 11 sex-specific DEGs and DAGs, most of which are located in chromosomes X and Y.

3.8 | Ancestry-specific GRA 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),[44–54] schizophrenia (SZ),[55–70] bipolar disorder (BP),[52,67,71–76] Parkinson's disease (PD),[77,78] multiple sclerosis (MS),[79,80] stroke,[81] coronary artery disease (CAD),[82–88] and hyperlipidemia (HDL)[89,90] (Figure 6). When comparing our ancestry-specific DEGs with previously reported GWAS genes for other neurological disorders, we observed the most DEGs across all ancestry groups for SZ followed by CAD, MS, and BD (Figure 6A). Similarly, we observed the most DAGs in SZ, followed by CAD, BP, and MS (Figure 6B).

4 | DISCUSSION

Recent studies have demonstrated that genetic disease associations differ in their strength and location between ancestries.[4,40,41,43] 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 EU. Given the key role of MGL 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 broader study of the CNS.

Several known AD genes demonstrated ancestral expression differences in the MGL. 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 AA individuals.[91] Studies have found that loss of *ABI3* function in mice was associated with Aβ-amyloidosis[92] and increased *ABI3* expression in MGL has been observed surrounding amyloid plaques in AD brain samples.[93] Both studies hypothesize that *ABI3* expression plays a role in MGL migration in the CNS 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.[94,95] Interestingly, increased CTSB protein expression has been reported in AD patients compared to controls.[96–98] It was also previously reported that *APOEe4* carriers

(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 GRA 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 ASD, SZ, BP, PD, stroke, MS, CAD, and HLD. Gray boxes represent the total number of genes queried. ASD, autism spectrum disorder; BP, bipolar disorder; CAD, coronary artery disease; DAGs, differentially accessible genes; DEGs, differentially expressed genes; GRA, genetic regulatory architecture; GWAS, genome-wide association studies; HLD, hyperlipidemia; iMGL, induced pluripotent stem cell-derived microglia; MS, multiple sclerosis; PD, Parkinson's disease; SZ, schizophrenia.

of AF LA expressed higher *CTSB* in brain MGL compared to those of EU LA surrounding the *APOE* locus,[16] 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 LA in AA 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 and undergo different selective pressures, creating ancestries who are the least related genetically and present the most differences in their regulatory architectures.

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 MGL[99] and it was previously reported that brain MGL of AF ancestry express less *MS4A6A* compared to those of EU ancestry.[16] Despite not reaching significance, we did observe a similar trend toward lower *MS4A6A* expression in AF iMGL compared to EU iMGL. *TREM2*, another well-known AD-GWAS gene, is primarily expressed in MGL and has been heavily implicated in AD progression.[100–103] 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[104] and loss-of-function *TREM2* variants are associated with dementia,[105–107] implying that the lower expression in AI MGL might impact AD risk in this ancestry due to reduced MGL 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 versus controls, and *APOEe3* versus *APOEe4* carriers. Most of our AD patients were *APOEe4* homozygotes as at least 60% of AD patients carry at least one copy of the *APOEe4* allele. Despite observing a small number of DEGs between AD versus controls and *APOE e3* versus *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.[108] It was also recently reported that downregulation of *HMGA2* in AD patients was associated with increased lifespan.[109] 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 expression quantitative trait loci (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 ($r = 0.47$–0.57). While most AD GWAS genes were identified in EU ancestry populations, we did not find AD GWAS genes to be differentially expressed/accessible between AF and EU ancestries. It is important to realize that AD GWAS genes are identified through allelic association which may not be reflected in expression/accessibility changes. Additionally, the "named" GWAS genes may not be the actual genes involved in the association. Both these possibilities could contribute to our findings. As a result, we expanded our data analysis and screened AD nominated targets based on functional evidence from the AGORA portal. Interestingly, we did not observe any DEG/accessibility in the ancestry comparisons. This could be due to the fact these genes are often nominated on evidence comparing AD versus controls samples and do not take ancestry into consideration. Given that our study primarily focuses on the comparison of ancestries rather than disease status, it is possible there are no significant differences between ancestries but we should not disregard their potential importance and involvement in AD.

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.[16] These iMGL 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 MGL 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 AF-specific AD GWAS genes[4] 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 heterogeneous expression levels between the limited number of individuals sampled in these analyses. The relatively small number of individuals included is the main limitation of this study. This is a general limitation of iPSC-derived cell studies which are expensive and time-consuming. Some of the differential findings reported here may reflect individual heterogeneity rather than ancestry generalizations. Additional iPSC-derived cell lines are needed to fully explore the regulatory architecture and to capture individual variability. Further genomic studies such as Hi-C will enhance these comparisons, particularly for specific genes of interest.

Finally, defining the regulatory architecture of these ancestries will not only be useful for all CNS diseases and will provide new resources for AD researchers interested in diversity but will also facilitate the move toward personalized medicine in diverse ancestries as we further our understanding of the functional genomics of AD in these ancestry groups.

ACKNOWLEDGMENTS

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. This study was supported by the National Institute on Aging (grant numbers U01-AG072579, RF1- AG059018, U01-AG066767, U01-AG052410, R56-AG072547, R01-AG070864). and Alzheimer Association, Zenith Award (JMV)

CONFLICT OF INTEREST STATEMENT

The authors declare nothing to disclose or competing interests. Author disclosures are available in the Supporting Information.

DATA AVAILABILITY STATEMENT

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.

CONSENT STATEMENT

All human subjects provided informed consent. This study received ethical approval from the University of Miami Institutional Review Board (approved protocol #20070307).

ORCID

Sofia Moura https://orcid.org/0000-0002-9000-2290
Luciana Bertholim Nasciben https://orcid.org/0009-0002-1518-1409
Aura M. Ramirez https://orcid.org/0009-0007-8892-5743
Kara L. Hamilton-Nelson https://orcid.org/0000-0003-3748-7193

Takiyah D. Starks https://orcid.org/0000-0001-7267-3480
Maryenela Illanes-Manrique https://orcid.org/0000-0002-3098-0981
Goldie S. Byrd https://orcid.org/0000-0003-1057-0261
Mario R. Cornejo-Olivas https://orcid.org/0000-0001-6313-5680
Briseida E. Feliciano-Astacio https://orcid.org/0000-0003-4747-8928
Karen Nuytemans https://orcid.org/0000-0003-2670-4914
Derek M. Dykxhoorn https://orcid.org/0000-0002-4998-9943
Farid Rajabli https://orcid.org/0000-0001-5229-7254
Jeffery M. Vance https://orcid.org/0000-0003-3815-8199

REFERENCES

1. Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). *Alzheimers Dement*. 2021;17:1966-1975.
2. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*. 2015;96:37.
3. Reitz C, Pericak-Vance MA, Foroud T, Mayeux R. A global view of the genetic basis of Alzheimer disease. *Nat Rev Neurol*. 2023;19(5):261-277.
4. Kunkle BW, Schmidt M, Klein H, et al. Novel Alzheimer disease risk loci and pathways in African American Individuals using the African genome resources panel: a meta-analysis. *JAMA Neurol*. 2021;78:102-113.
5. Reitz C, Jun G, Naj A, 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*. 2013;309:1483-1492.
6. Cukier HN, Kunkle BW, Vardarajan BN, et al. ABCA7 frameshift deletion associated with Alzheimer disease in African Americans. *Neurol Genet*. 2016;2(3):e79.
7. Talebi M, Delpak A, Khalaj-kondori M, et al. ABCA7 and EphA1 genes polymorphisms in late-onset Alzheimer's disease. *J Mol Neurosc*. 2020;70:167-173.
8. Vardarajan BN, Ghani M, Kahn A, et al. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann Neurol*. 2015;78:487-498.
9. Ray NR, Kunkle BW, Hamilton-Nelson K, 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;20(8):5247-5261. doi:10.1002/ALZ.13880
10. Giral H, Landmesser U, Kratzer A. Into the wild: GWAS exploration of non-coding RNAs. *Front Cardiovasc Med*. 2018;5:181.
11. Mirza AH, Kaur S, Brorsson CA, Pociot F. Effects of GWAS-associated genetic variants on lncRNAs within IBD and T1D candidate loci. *PLoS One*. 2014;9:e105723.
12. Andrews SJ, Renton AE, Fulton-Howard B, et al. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *EBioMedicine*. 2023;90:104511.
13. Rajabli F, Beecham GW, Hendrie HC, et al. A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's disease in African ancestry. *PLoS Genet*. 2022;18:e1009977.
14. Rajabli F, Feliciano BE, Celis K, et al. Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. *PLoS Genet*. 2018;14:e1007791.
15. Celis K, Moreno M, Rajabli F, et al. Ancestry-related differences in chromatin accessibility and gene expression of APOE ε4 are associated with Alzheimer's disease risk. *Alzheimers Dement*. 2023;19:3902-3915.
16. Griswold AJ, Celis K, Bussies PL, et al. Increased APOE ε4 expression is associated with the difference in Alzheimer's disease risk from diverse ancestral backgrounds. *Alzheimers Dement*. 2021;17:1179-1188.
17. Vance JM, Farrer LA, Huang Y, 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;95(4):625-634. doi:10.1002/ANA.26864
18. Felsky D, Roostaei T, Nho K, et al. Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nat Comm*. 2019;10(1):1-12.
19. Nott A, Holtman IR, Coufal NG, et al. Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. *Science*. 2019;366:1134-1139.
20. Novikova G, Kapoor M, TCW J, et al. Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Comm*. 2021;12(1):1-14.
21. Colonna M, Butovsky O. Microglia function in the central nervous system during health and neurodegeneration. *Annu Rev Immunol*. 2017;35:441.
22. Gao C, Jiang J, Tan Y, Chen S. Microglia in neurodegenerative diseases: mechanism and potential therapeutic targets. *Signal Transduct Target Ther*. 2023;8(1):1-37.
23. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655-1664.
24. Cavalli-Sforza LL. Human evolution and its relevance for genetic epidemiology. *Annu Rev Genomics Hum Genet*. 2007;8:1-15.
25. Delaneau O, Marchini J, McVean GA, et al. Integrating sequence and array data to create an improved 1000 genomes project haplotype reference panel. *Nat Comm*. 2014;5(1):1-9.
26. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
27. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013;93:278-288.
28. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.
29. DeRosa BA, Simon SA, Velez CA, Vance JM, Pericak-Vance MA, Dykxhoorn DM. Generation of two iPSC lines (UMi038-A & UMi039-A) from siblings bearing an Alzheimer's disease-associated variant in SORL1. *Stem Cell Res*. 2022;62:102823.
30. McQuade A, Coburn M, Tu CH, Hasselmann J, Davtyan H, Blurton-Jones M. Development and validation of a simplified method to generate human microglia from pluripotent stem cells. *Mol Neurodegener*. 2018;13:1-13.
31. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21.
32. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol*. 2015;109:21291.
33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth*. 2012;9(4):357-359.
34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1-21.
35. 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*. 2020;9:ELIXIR-709.

36. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382-2383.
37. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*. 2020;2:lqaa078.
38. Cao J, O'Day DR, Pliner HA, et al. A human cell atlas of fetal gene expression. *Science*. 2020;370(6518):eaba7721.
39. Lambert J, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452.
40. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Gen*. 2019;51(3):414-430.
41. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Gen*. 2022;54(4):412-436.
42. Wightman DP, Jansen IE, Savage JE, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Gen*. 2021;53(9):1276-1282.
43. Lake J, Warly Solsberg C, Kim JJ, et al. Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. *Mol Psychiatr*. 2023;28(7):3121-3132. doi:10.1038/s41380-023-02089-w
44. Anney R, Klei L, Pinto D, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet*. 2012;21:4781-4792.
45. Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010;19:4072-4082.
46. Weiss LA, Arking DE. A genome-wide linkage and association scan reveals novel loci for autism. *Nature*. 2009;461(7265):802-808.
47. Ma D, Salyakina D, Jaworski JM, et al. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet*. 2009;73:263-273.
48. Matoba N, Liang D, Sun H, et al. Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Trans Psychiatr*. 2020;10(1):1-14.
49. Liu X, Shimada T, Otowa T, et al. Genome-wide association study of autism spectrum disorder in the East Asian populations. *Autism Res*. 2016;9:340-349.
50. 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*. 2017;8:21.
51. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444.
52. Wu Y, Cao H, Baranova A, et al. Multi-trait analysis for genome-wide association study of five psychiatric disorders. *Transl Psychiatr*. 2020;10(1):1-11.
53. Kuo P, Chuang L, Su M, et al. Genome-wide association study for autism spectrum disorder in Taiwanese Han population. *PLoS One*. 2015;10(9):e0138695.
54. Almandil NB, AlSulaiman A, Aldakeel SA, et al. Integration of transcriptome and exome genotyping identifies significant variants with autism spectrum disorder. *Pharmaceuticals*. 2022;15(2):158.
55. Lam M, Chen C, Li Z, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51(12):1670-1678.
56. Ripke S, Neale BM, Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427.
57. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45(10):1150-1159.
58. Betcheva ET, Yosifova AG, Mushiroda T, et al. Whole-genome-wide association study in the Bulgarian population reveals HHAT as schizophrenia susceptibility gene. *Psychiatr Genet*. 2013;23:11-19.
59. Goes FS, McGrath J, Avramopoulos D, et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168:649-659.
60. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460(7256):744-747.
61. Aberg KA, Liu Y, Bukszár J, et al. A Comprehensive family-based replication study of schizophrenia genes. *JAMA Psychiatr*. 2013;70:573-581.
62. Levinson DF, Shi J, Wang K, et al. Genome-wide association study of multiplex schizophrenia pedigrees. *Am J Psychiatr*. 2012;169:963-973.
63. Li Z, Chen J, Yu H, et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat Genet*. 2017;49(11):1576-1583.
64. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-508.
65. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. 2018;50(3):381-389.
66. Liu J, Li S, Li X, et al. Genome-wide association study followed by trans-ancestry meta-analysis identify 17 new risk loci for schizophrenia. *BMC Med*. 2021;19:1-15.
67. Blokland GA, Grove J, Chen C, et al. Sex-dependent shared and non-shared genetic architecture across mood and psychotic disorders. *Biol Psychiatr*. 2022;91:102-117.
68. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40(9):1053-1055.
69. Athanasiu L, Mattingsdal M, Kähler AK, et al. Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J Psychiatr Res*. 2010;44:748-753.
70. Kirov G, Zaharieva I, Georgieva L, et al. A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatr*. 2008;14(8):796-803.
71. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. 2019;51(5):793-803.
72. Ikeda M, Takahashi A, Kamatani Y, et al. A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Mol Psychiatr*. 2017;23(3):639-647.
73. Mullins N, Forstner AJ, O'Connell KS, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet*. 2021;53(6):817-829.
74. Cichon S, Mühleisen TW, Degenhardt FA, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet*. 2011;88:372-381.
75. Li H, Zhang C, Hui L, 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 Psychiatr*. 2021;78:320-330.
76. Bigdeli TB, Fanous AH, Li Y, et al. Genome-wide association studies of schizophrenia and bipolar disorder in a diverse cohort of US veterans. *Schizophr Bull*. 2021;47:517-529.
77. Nalls MA, Blauwendraat C, Vallerga CL, 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*. 2019;18:1091-1102.
78. Pan H, Liu Z, Ma J, 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):1-11.
79. International Multiple Sclerosis Genetics Consortium (IMSGC), Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci

identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet*. 2013;45(11):1353-1360.
80. International Multiple Sclerosis Genetics Consortium. Multiple Sclerosis genomic map implicates peripheral immune cells & microglia in susceptibility. *Science*. 2019;365(6460):eaav7188.
81. Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115-123.
82. Tcheandjieu C, Zhu X, Hilliard AT, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med*. 2022;28(8):1679-1692.
83. Koyama S, Ito K, Terao C, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet*. 2020;52(11):1169-1177.
84. Klarin D, Zhu QM, Emdin CA, et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet*. 2017;49(9):1392-1397.
85. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49(9):1385-1391.
86. Howson JMM, Zhao W, Barnes DR, et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet*. 2017;49(7):1113-1119.
87. 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*. 2018;122:433-443.
88. Verweij N, Eppinga RN, 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. *Sci Rep*. 2017;7(1):1-9.
89. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599(7886):628-634.
90. Nam K, Kim J, Lee S. Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. *Cell Genom*. 2022;2:100189.
91. Conway OJ, Carrasquillo MM, Wang X, et al. ABI3 and PLCG2 missense variants as risk factors for neurodegenerative diseases in Caucasians and African Americans. *Mol Neurodegener*. 2018;13:1-12.
92. Karahan H, Smith DC, Kim B, et al. The effect of Abi3 locus deletion on the progression of Alzheimer's disease-related pathologies. *Front Immunol*. 2023;14:1102530.
93. Satoh J, Kino Y, Yanaizu M, et al. Microglia express ABI3 in the brains of Alzheimer's disease and Nasu-Hakola disease. *Intractable Rare Dis Res*. 2017;6:262.
94. Drobny A, Prieto Huarcaya S, Dobert J, et al. The role of lysosomal cathepsins in neurodegeneration: mechanistic insights, diagnostic potential and therapeutic approaches. *Biochim Biophys Acta Mol Cell Res*. 2022;1869:119243.
95. Cermak S, Kosicek M, Mladenovic-Djordjevic A, Smiljanic K, Kanazir S, Hecimovic S. Loss of cathepsin B and L leads to lysosomal dysfunction, NPC-like cholesterol sequestration and accumulation of the key Alzheimer's proteins. *PLoS One*. 2016;11:e0167428.
96. Sundelöf J, Sundström J, Hansson O, et al. Higher cathepsin B levels in plasma in Alzheimer's disease compared to healthy controls. *J Alzheimers Dis*. 2010;22:1223-1230.
97. Sun Y, Rong X, Lu W, et al. Translational study of Alzheimer's disease (AD) biomarkers from brain tissues in AβPP/PS1 mice and serum of AD patients. *J Alzheimers Dis*. 2015;45:269-282.
98. Morena F, Argentati C, Trotta R, 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*. 2017;18:1806.
99. Zhang Y, Chen K, Sloan SA, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosc*. 2014;34:11929.
100. Ulland TK, Colonna M. TREM2 — a key player in microglial biology and Alzheimer disease. *Nat Rev Neurol*. 2018;14(11):667-675.
101. Zhao Y, Wu X, Li X, et al. TREM2 is a receptor for β-amyloid that mediates microglial function. *Neuron*. 2018;97:1023-1031.e7.
102. McQuade A, Kang YJ, Hasselmann J, et al. Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nat Comm*. 2020;11(1):1-17.
103. Schmid CD, Sautkulis LN, Danielson PE, et al. Heterogeneous expression of the triggering receptor expressed on myeloid cells-2 on adult murine microglia. *J Neurochem*. 2002;83:1309-1320.
104. Winfree RL, Seto M, Dumitrescu L, et al. TREM2 gene expression associations with Alzheimer's disease neuropathology are region-specific: implications for cortical versus subcortical microglia. *Acta Neuropathol*. 2023;145:733-747.
105. Cady J, Koval ED, Benitez BA, et al. TREM2 variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis. *JAMA Neurol*. 2014;71:449-453.
106. Gratuze M, Leyns CEG, Holtzman DM. New insights into the role of TREM2 in Alzheimer's disease. *Mol Neurodegener*. 2018;13(1):1-16.
107. Lill CM, Rengmark A, Pihlstrøm L, 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. *Alzheimers Dement*. 2015;11:1407-1416.
108. Liu X, Wang H, Bei J, Zhao J, Jiang G, Liu X. The protective role of miR-132 targeting HMGA2 through the PI3K/AKT pathway in mice with Alzheimer's disease. *Am J Transl Res*. 2021;13:4632.
109. 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*. 2023;15:1072184.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Moura S, Nasciben LB, Ramirez AM, et al. Comparing Alzheimer's genes in African, European, and Amerindian induced pluripotent stem cell–derived microglia. *Alzheimer's Dement*. 2025;21:e70031.
https://doi.org/10.1002/alz.70031