



## OPEN ACCESS

## EDITED BY

Wei Wang,  
Zhejiang University, China

## REVIEWED BY

Guangyu Guo,  
First Affiliated Hospital of Zhengzhou  
University, China  
Jakob Körbelin,  
University Medical Center  
Hamburg-Eppendorf, Germany  
Yubo Wang,  
The Scripps Research Institute, United States  
Zhengzhao Lu,  
Capital Medical University, China

## \*CORRESPONDENCE

Qi Pan  
✉ seven627390@163.com

RECEIVED 14 June 2025

ACCEPTED 19 August 2025

PUBLISHED 02 September 2025

## CITATION

Deng H, Weng Q-X, Zeng M, Li Y and  
Pan Q (2025) Integrating scRNA-seq and  
GWAS data reveals potentially critical  
endothelial cells in large artery atherosclerotic  
stroke.  
*Front. Neurosci.* 19:1646993.  
doi: 10.3389/fnins.2025.1646993

## COPYRIGHT

© 2025 Deng, Weng, Zeng, Li and Pan. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Integrating scRNA-seq and GWAS data reveals potentially critical endothelial cells in large artery atherosclerotic stroke

Hong Deng, Qiu-Xue Weng, Mei Zeng, Yan Li and Qi Pan\*

Department of Neurology, Affiliated Hospital of Guizhou Medical University, Guiyang, China

**Introduction:** Stroke is a devastating cerebrovascular disease with limited treatment options. Structural and functional abnormalities in the cerebral vasculature contribute significantly to the development of stroke. Specifically, the blood–brain barrier formed by a combination of endothelial cells, smooth muscle cells, pericytes, and glial cells is proposed to play a critical role in stroke pathogenesis. However, the primary causative cell types in stroke are not well-defined.

**Methods:** Here we integrated single-cell RNA-sequencing data from cerebrovascular zonation structures with genome-wide association study (GWAS) data of various types of strokes to analyze stroke risk SNP loci enrichment in various cell types through heritability partitioning. Fourteen vascular cell clusters were identified and profiled using cell-type expression-specific integration for complex traits (CELLEX) to assess gene expression specificity. Heritability enrichment was analyzed across three GWAS datasets: general stroke, large-artery atherosclerotic stroke (Stroke\_Large\_AS), and small-vessel ischemic stroke. Gene set enrichment analysis (GSEA) and pseudotime trajectory modeling (Monocle2) were performed to characterize the functional and developmental dynamics of critical endothelial subtypes. Transcriptional regulatory networks were further interrogated using SCENIC.

**Results:** We found two specific endothelial cell subtypes play a potentially critical role in large artery atherosclerotic stroke. These subtypes demonstrated distinct expression profiles and pathway enrichments, with aEC\_cluster associated with atherosclerosis and inflammatory signaling, and EC2\_cluster with epithelial cell regulation and muscle tissue development. Pseudotime analysis revealed that EC2\_cluster cells represent an earlier endothelial state that differentiates into the aEC\_cluster, the terminal pathogenic state. Gene regulatory network analysis identified transcription factors Gata2, Irf7, and Jun as central regulators in aEC\_cluster differentiation, orchestrating the upregulation of key molecules such as Fos and Nr4a1.

**Conclusion:** Our study provides insights into the molecular mechanisms in stroke and suggests new therapeutic strategies for this debilitating disease.

## KEYWORDS

stroke, endothelial cells, scRNA-seq, GWAS, S-LDSC

## Introduction

Stroke is the leading cause of disability and the second leading cause of death globally (GBD 2019 Stroke Collaborators, 2021). Over 80% of strokes are ischemic events, which can cause significant neurological, morphological, and molecular abnormalities (Shehjar et al., 2023). To date, effective treatments for ischemic stroke remain lacking (Datta et al., 2020; Wen-Juan et al., 2023). Therefore, further investigation into the causes of ischemic stroke onset is crucial to prevent its occurrence and reduce the disease burden (Beylerli et al., 2024).

The health of the brain is dependent on the health of its vasculature, with the blood–brain barrier (BBB) being a specialized brain vascular structure that supports cerebral blood flow and neurological function under normal conditions (Mei-Yun et al., 2022; Profaci et al., 2020). Comprising various cell types, including endothelial cells, smooth muscle cells, perivascular cells, and glial cells, the BBB functions collectively to maintain its essential properties (Jonas et al., 2024; Veltkamp and Gill, 2016). Abnormalities within the BBB components, such as damage to vascular endothelial cells, can result in thrombosis and subsequent blockage of blood vessels, ultimately leading to the development of ischemic stroke (Edwards and Bix, 2019). Recent advancements in single-cell technology have enabled the exploration cellular heterogeneity within brain vascular structures (He et al., 2018; Pan et al., 2022; Yang et al., 2022). However, it remains unclear whether specific cell types within the brain vasculature contribute to the pathogenesis of ischemic.

Genome-wide association analysis (GWAS) has emerged as a powerful tool to identify potential causal loci for diseases by associating complex phenotypes with genotypes (Barbeira et al., 2019). Recent GWAS studies have successfully identified several risk loci for stroke (Malik et al., 2018; Kilarski et al., 2014; Holliday et al., 2012). Integration of expression data from single-cell RNA sequencing (scRNA-seq) with GWAS statistics using stratified linkage disequilibrium score regression (S-LDSC) enables identification of possible pathogenic cell types of a disease through heritability enrichment (Gareev et al., 2024). In this study, we utilized an algorithm called cell-type expression-specific integration for complex traits (CELLECT) (Timshel et al., 2020) to integrate single-cell profiles of the mouse cerebral vasculature with GWAS statistics from stroke, large-artery atherosclerotic stroke, and small-vessel stroke. Our analysis revealed the presence of two specific endothelial cell types, namely aEC\_cluster and EC2\_cluster, which are potential associated with large atherosclerotic stroke. Furthermore, gene set enrichment analysis (GSEA) revealed that these two cell subtypes are specifically associated with atherosclerosis and epithelial cell proliferation, respectively, providing further evidence of their crucial role in the development of large artery atherosclerotic stroke. Although our conclusions are based solely on existing scRNA-seq and GWAS data, analysis of transcriptional regulatory networks provides potential targets for the prevention and treatment of stroke in clinical settings.

## Materials and methods

### Data collection

The open GWAS database<sup>1</sup> was utilized to obtain GWAS statistics for different types of stroke, including ebi-a-GCST006906 for stroke, ebi-a-GCST005840 for large artery atherosclerosis stroke, and ebi-a-GCST006909 for small vessel stroke (Malik et al., 2018). To ensure adequate overlap between the genes obtained from single-cell data and the GWAS risk loci, we integrated a brain vascular single-cell atlas (Vanlandewijck et al., 2018), which was obtained using SMART-seq2-based technology. Specifically, we analyzed the brain vascular single-cell dataset from the Gene Expression Omnibus database (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) using accession numbers GSE98816 (Vanlandewijck et al., 2018).

### Data processing of scRNA-seq data

We processed and integrated our raw single-cell count matrices using the Seurat package (version 4.0; <https://satijalab.org/seurat/>). To ensure data quality, we excluded low-quality cells which expressed  $\leq 200$  genes,  $\geq 20,000$  unique molecular identifiers, or  $\geq 20\%$  mitochondrial genes. After that, we performed normalization and dimensionality reduction using the t-Distributed Stochastic Neighbor Embedding (t-SNE) algorithm (Skrodzki et al., 2024) for non-linear dimensionality reduction. Clusters were identified and annotated using previously published data (He et al., 2018).

### Quantifying expression specificity in cell types using CELLEX

To identify a set of genes that are expressed in specific cell types and to quantify their expression specificity (ES), we used CELLEX (CELL type EXpression-specificity) (Timshel et al., 2020). This tool allowed us to measure the ES of each gene in the expression matrix of each cell type. To simplify the analysis, we calculated a single ES estimate ( $ES_{\mu}$ ) for each gene based on its expression in each cell type. This score represents the degree of specificity of the gene's expression to the cell type in which it is expressed.

### Identification of critical cell types in stroke through scRNA-seq and GWAS integration

To systematically prioritize cell types that may play etiological roles in complex diseases such as stroke, we employed an integrative analytical framework that combines scRNA-seq data with GWAS summary statistics. Specifically, we utilized the CELLECT tool

<sup>1</sup> <https://gwas.mrcieu.ac.uk/>

(Timshel et al., 2020), which incorporates gene expression specificity information across cell types to evaluate heritability enrichment using S-LDSC. The CELLECT workflow consists of the following key steps: (1) Calculation of cell type-specific expression profiles: For each cell type or cluster identified in the scRNA-seq dataset, CELLECT quantifies expression specificity by computing metrics such as the mean expression and ES $\mu$  score (expression specificity metric), identifying genes that are highly specific to each cell type. (2) Gene annotation to regulatory regions: Each cell type-specific gene is annotated to a genomic region that includes a  $\pm 100$  kilobase (kb) window surrounding the transcription start site, capturing nearby regulatory elements that may influence gene expression. (3) Linkage disequilibrium (LD) score computation: For each annotated gene region, CELLECT calculates LD scores for HapMap3 SNPs, quantifying the extent to which these variants are correlated with each other across the population. This step creates a cell type-specific annotation map that reflects the distribution of regulatory elements across the genome. (4) Heritability enrichment analysis: Finally, the framework assesses whether the SNPs associated with a given trait or disease (e.g., stroke or stroke subtypes) are significantly enriched in the gene sets specific to certain cell types. This is done through S-LDSC, which models the contribution of each annotated SNP set to the overall SNP-based heritability of the trait. A statistically significant enrichment indicates that the corresponding cell type may be critically involved in the pathogenesis of the disease. By integrating GWAS summary statistics for stroke and its subtypes with high-resolution scRNA-seq data from cerebrovascular cells, this approach enables the identification of cell types whose gene expression landscapes are disproportionately enriched for genetic risk variants, thereby pinpointing potential cellular origins of disease susceptibility.

## Differential expression analysis and functional enrichment analysis

To identify differentially expressed genes between cell subpopulations, we utilized the “Findmarker” function in the Seurat package. GSEA was performed using the “clusterProfiler” package.

## Pseudotime-trajectory analysis

We analyzed the developmental trajectory of cell types using the Monocle2 (v2.12.0) R package (Trapnell et al., 2014; Qiu et al., 2017). To identify gene ordering (qval < 0.01), we utilized the differential GeneTest function of the Monocle2 package. Subsequently, we utilized the reduce Dimension function and performed the order Cells function with default parameters to infer the trajectory.

## Network analysis for transcriptional regulators

pySCIENIC (Single-Cell rEgulatory Network Inference and Clustering)(version 0.11.0) was used to identify single-cell gene regulatory network (GRN) (Van de Sande et al., 2020). To construct the GRNs, we first generated an expression matrix and co-expression modules of transcription factors (TFs) and their potential target genes

based on co-expression analysis. We then performed TF motif analysis using RcisTarget to identify direct-binding regulons. To evaluate the activity of the regulons in each cell, we used AUCell (regulon\_AUC) to calculate the regulon activity score (RAS), which represents the average activity score across regulons. Finally, we summarized the GRNs and regulons for the endothelial cell types of mouse cerebrovascular cells.

## Results

### Data integration and profiling of single-cell data

We conducted a comprehensive analysis of a scRNA-seq dataset obtained from a recent study (He et al., 2018), which aimed to uncover the gene expression profiles of vascular and vessel-associated cells across defined cerebrovascular zonation in the mouse brain. Our analysis identified 14 distinct cell clusters using t-Distributed stochastic neighbor embedding (Figure 1A), and revealed the top genes that were specifically expressed in each cell cluster (Figure 1B). Notably, genes such as Cldn5, Myh11, Pla1a, Theme119, Plp1, and Lum, were consistently expressed at higher levels in their respective cell clusters (Figure 1C).

To quantitatively assess and rank gene expression specificity in the 14 cerebrovascular cell subtypes, we utilized the CELLEX toolkit. The ES $\mu$  score was utilized to represent the degree of gene expression specificity in each cell cluster. Our analysis identified the most specifically expressed genes in each cell cluster (Figure 1D), and their expression patterns were consistent with those observed in the dot plot of average expression. These findings indirectly validated the reliability of the ES $\mu$  score in assessing gene expression specificity.

### Two endothelial cell subtypes are specifically associated with risk loci for large artery atherosclerotic stroke

To identify vascular cell types associated with stroke, we employed the CELLECT algorithm to analyze the 14 vascular cell types using three different GWAS statistics, namely stroke, Stroke\_Large\_AS, and ischemic Stroke\_small\_vessel. Our analysis revealed that two of the EC clusters, namely the aEC cluster and EC2 cluster, showed significant enrichment for variants associated with Stroke\_large\_AS (Figure 2A). Moreover, we assessed the ES $\mu$  of seven stroke candidate genes obtained from a GWAS study (Bevan et al., 2012), which supported the reliability of our findings (Figure 2B). Notably, several of these candidate genes, including DDAH1, NOS3, and SGK1, are specifically enriched in either the aEC\_cluster or the EC2\_cluster, providing insights into the molecular mechanisms by which these cell types contribute to the pathogenesis of large artery atherosclerotic stroke.

### Enrichment of atherosclerosis-related pathways in the two potentially critical endothelial cell subtypes

To explore the relationship between stroke pathology and the two distinct subtypes of endothelial cells, we employed GSEA to analyze

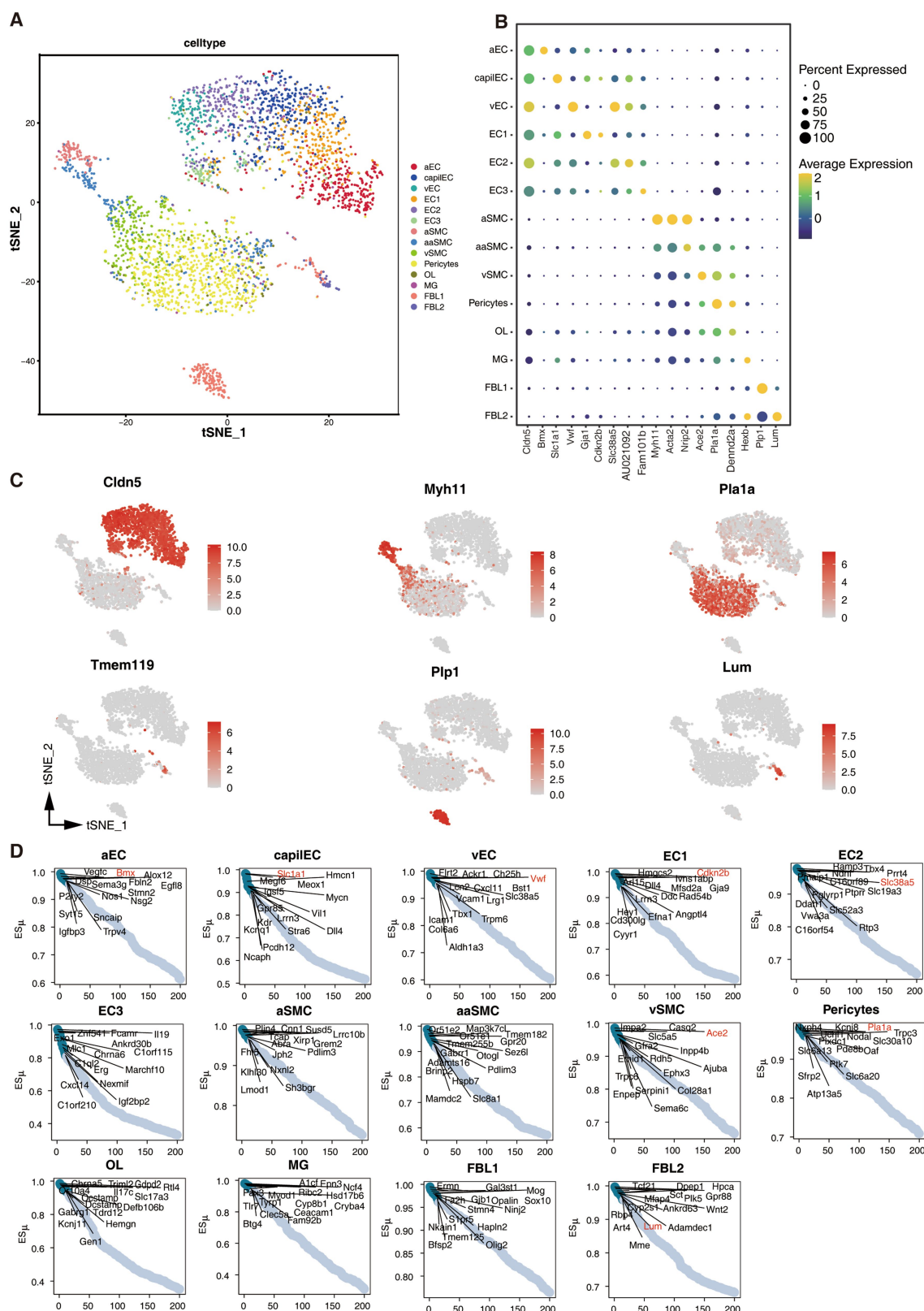
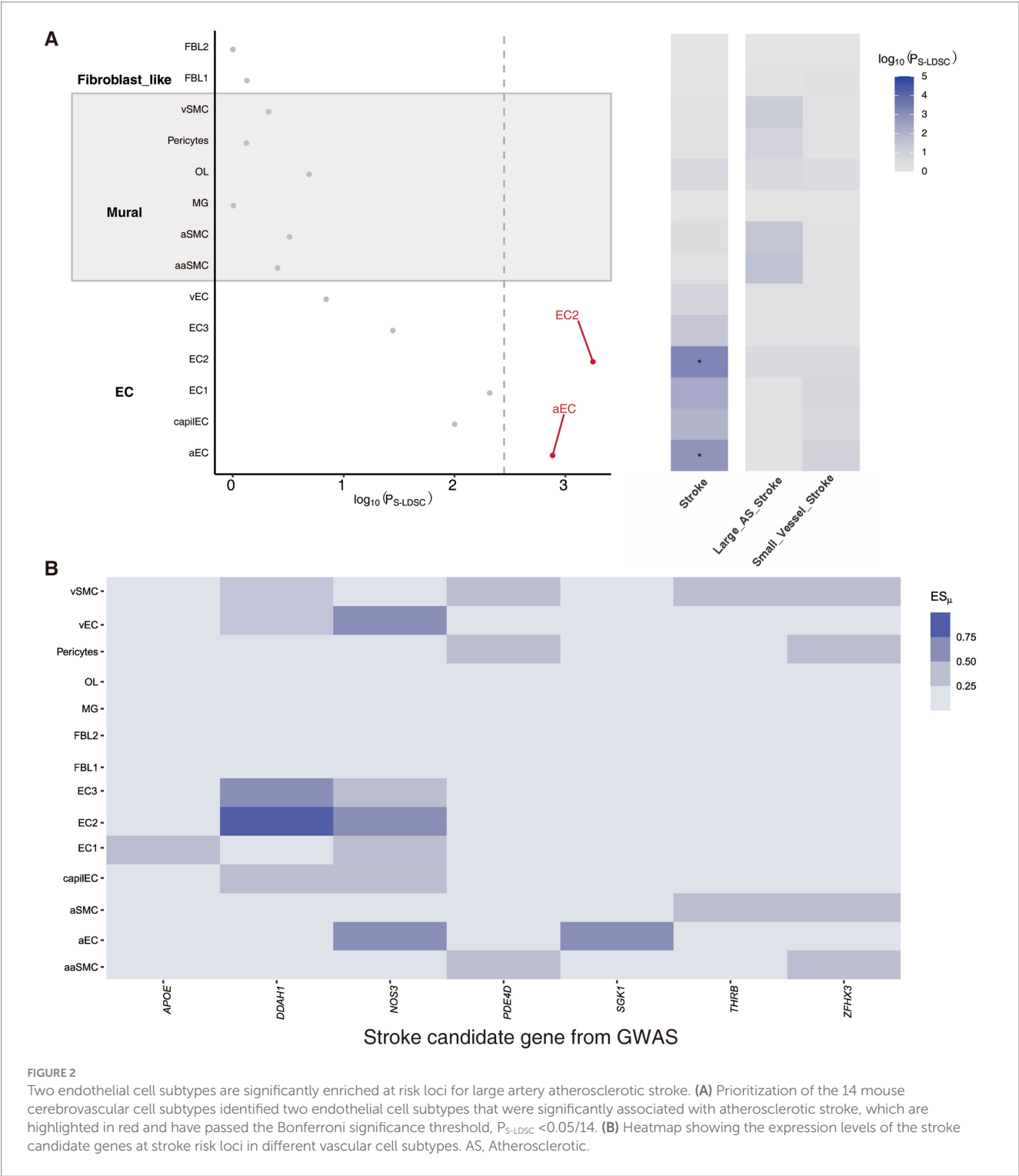


FIGURE 1  
Single-cell landscape of brain vasculature and cell subtype-specific gene expression. (A) t-Distributed stochastic neighbor embedding (t-SNE) plot showing the cell types of brain vasculature represented by diverse colors: aEC (arterial endothelial cell), capilEC (capillary endothelial cell), vEC (venous endothelial cell), EC1, EC2, EC3, aSMC (arterial smooth muscle cell), aaSMC (arterial smooth muscle cell), vSMC (venous smooth muscle cell), Pericytes, OL (oligodendrocyte-lineage cells), MG (Microglia), FBL1 (vascular Fibroblast-like cell 1), FBL2 (Vascular Fibroblast-like cell 2). (B) Dot plot

(Continued)

FIGURE 1 (Continued)  
showing the percentage of featured genes expressed in each cluster and their average expression levels. (C) Feature plots demonstrating the expression of selective marker genes of cerebrovascular cells. (D) Scatter plots showing the ordering of specific genes expressed in each cluster obtained by the CELLEX algorithm.





their cellular functions (Figures 3A–D). Our findings revealed significant enrichment of specific pathways in each subtype. In the aEC\_cluster, we identified a significant upregulation of pathways associated with fluid shear stress and atherosclerosis, growth hormone synthesis, inflammatory mediator regulation of TRP channels, and TNF signaling (Figure 3C). Conversely, the EC2\_cluster exhibited a significant downregulation of pathways related to muscle organ development, muscle tissue development, and regulation of epithelial cell proliferation (Figure 3D). These findings suggest distinct functional profiles and underlying molecular processes associated with the two critical endothelial cell subtypes.

Moreover, we identified several key molecules, including Nr4a1, Fos, and Alox12, that exhibited significant changes in both aEC\_cluster and EC2\_cluster (Figure 3E). These molecules are known to be involved in crucial biological processes, highlighting their potential roles in mediating the pathogenesis of stroke within these endothelial cell subtypes.

## The two potentially critical endothelial cell subtypes are at distinct stages along the same differentiation trajectory

To gain insights into the differentiation trajectory of the two stroke-associated endothelial cell subtypes, we performed pseudotime analysis on all endothelial cells using Monocle2. Our analysis revealed a differentiation trajectory characterized by the bifurcation of endothelial cells into two states: state 2 cells differentiating toward either state 1 or state 3. Specifically, the aEC\_Cluster mainly belonged to state 1, while EC2\_Cluster predominantly belonged to state 2. As a result, the two pathogenic EC classes within the EC2\_Cluster gradually transitioned toward aEC\_Cluster differentiation (Figures 4A–C). Notably, the aEC\_Cluster represents the terminal stage of differentiation.

Furthermore, we performed unsupervised clustering of gene expression kinetic patterns and identified six expression modules shared by endothelial cells. Intriguingly, genes within modules 2, 3 and 5 showed a gradual elevation in expression as pseudotime advanced (Figure 4D). Consistent with our findings in Figure 3, differential expression and functional enrichment analysis revealed the gradual upregulation of genes associated with atherosclerosis, such as FOS, NR4A1, and ALOX12, along the differentiation trajectory from EC2\_Cluster to aEC\_Cluster.

## Key regulatory networks for endothelial cell differentiation into potentially critical subtypes

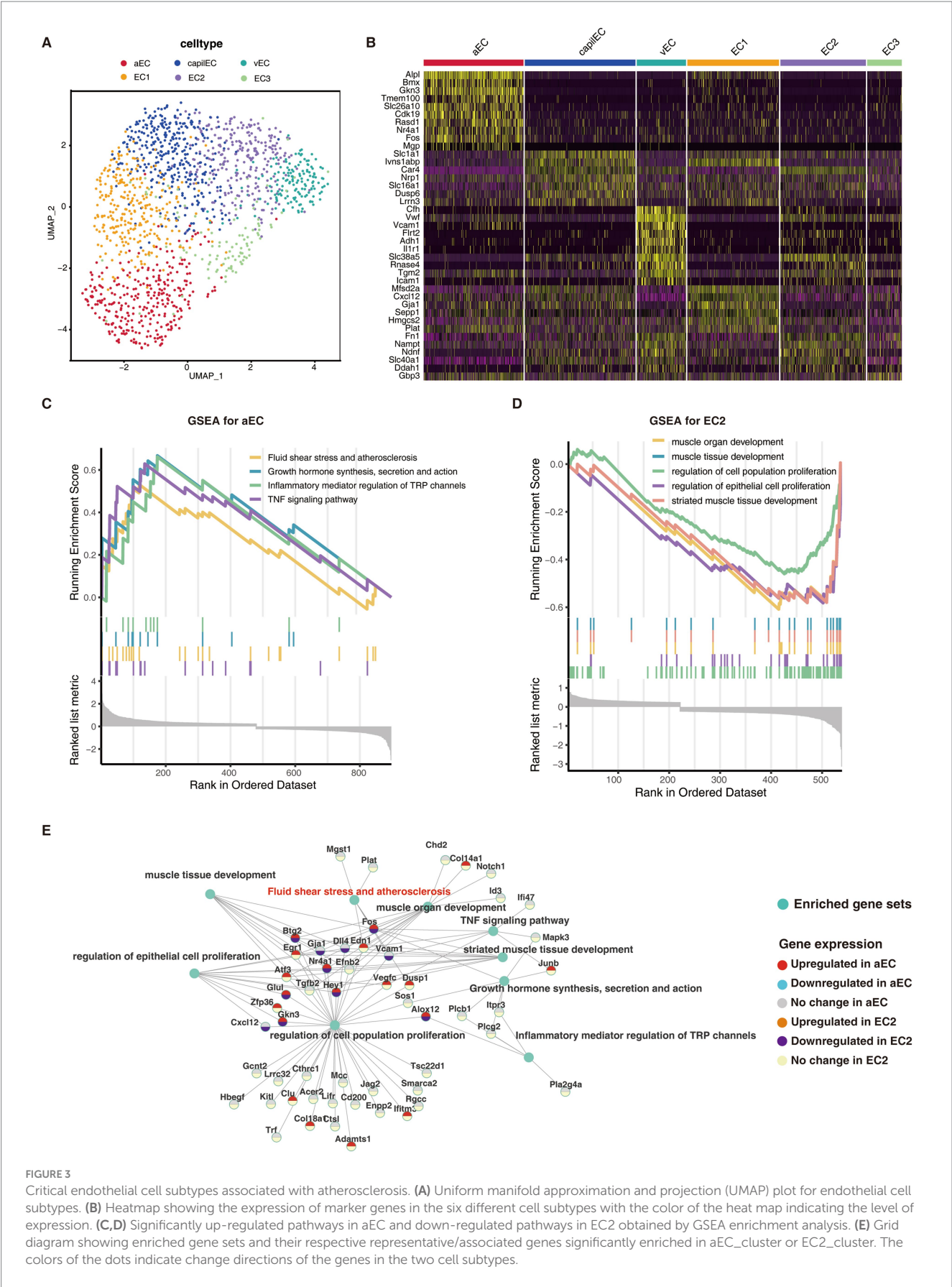
To investigate the upstream transcriptional regulatory events that play potential key roles in the differentiation of endothelial cells into these critical cell subtypes, we conducted a gene regulatory network analysis. Using SCENIC's integrated algorithm, we identified 10 transcriptional regulator modules of regulons (i.e., transcription factor with its target) and their co-expression modules in all the endothelial cells (Figure 5A; Supplementary Table 1). We further evaluated the scoring of each endothelial cell subtype in these modules and found that aEC\_cluster had the highest scores in modules M1 and M2, while

EC2\_cluster had the highest score in M4 (Supplementary Figure 1). To identify the key transcriptional regulatory network in the two potentially critical cell subtypes associated with stroke, we first scored the regulons in each subtype (Supplementary Table 2). From these results, the top regulons enriched in each subtype were identified (Figure 5B). Notably, the results revealed that the activity of the top five regulons, namely TF\_Mafk, TF\_Gata2, TF\_Irf7, TF\_Stat6 and TF\_Jund, was significantly elevated in aEC (Figure 5B). Further analysis revealed the extensive regulation of molecular networks in aEC by transcription factors Gata2, Irf7, and Jund. Remarkably, this regulation included molecules such as Fos and Nr4a1, which were displayed progressively elevation during aEC\_cluster differentiation in the pseudotime analysis (Figures 4D, 5C). To further validate the reliability of our transcriptional regulatory factor predictions, we incorporated scRNA-seq data from an external model of middle cerebral artery occlusion (MCAO) in the brain (Zhang et al., 2023). The results indicated that the levels of Irf7 and Fos in the MCAO model group were significantly higher than those in the control group (Figure 5D), thereby suggesting the reliability of our findings to some extent.

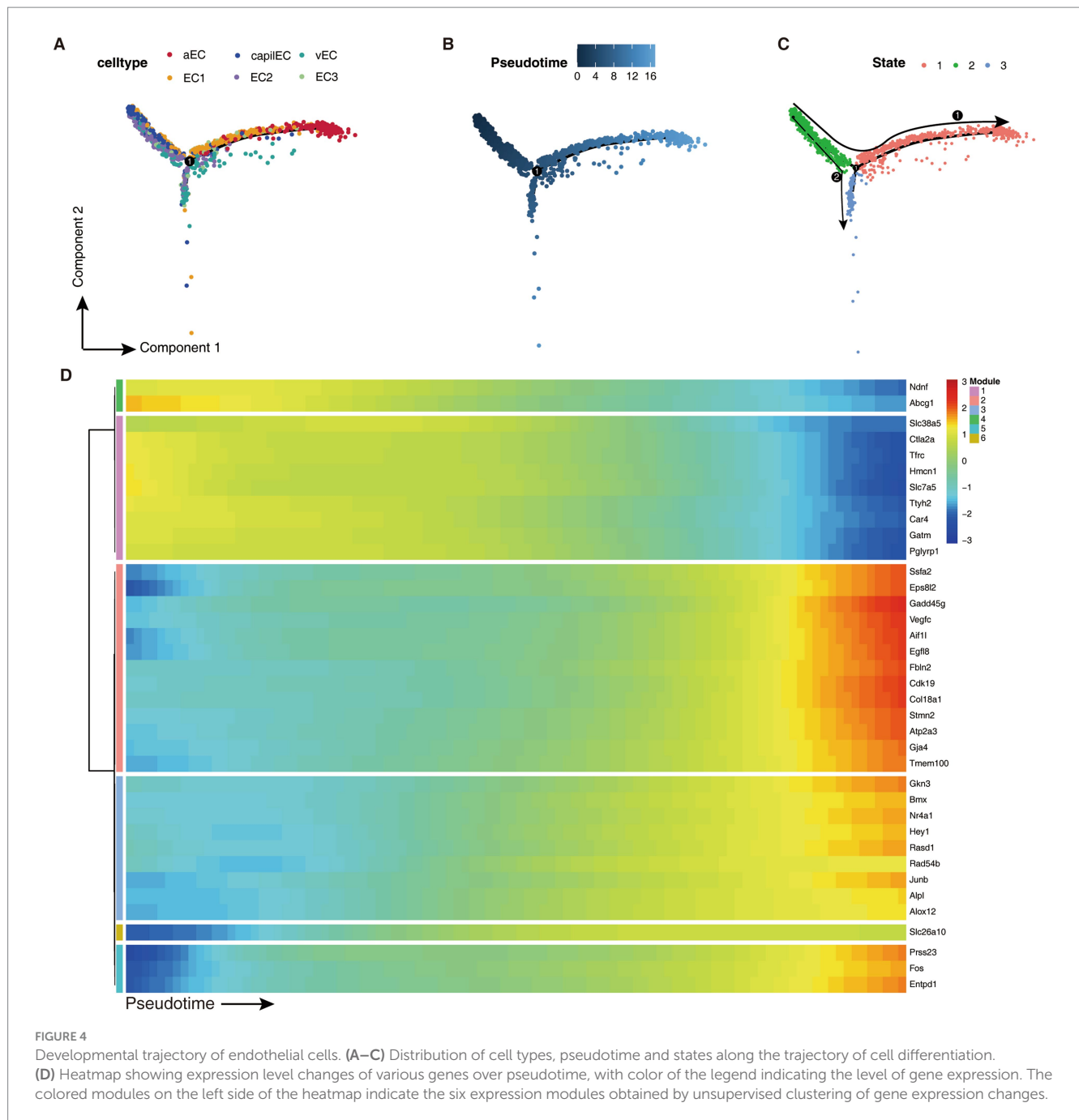
## Discussion

In this study, we aimed to infer the risk of stroke associated with cell types by integrating single-cell datasets of the cerebral vasculature with GWAS data from various stroke types. Using S\_LDSC heritability segmentation, we identified two endothelial cell subtypes, aEC\_cluster and EC2\_cluster, which displayed potential associations with risk loci for large atherosclerotic stroke. Our GSEA enrichment analysis further uncovered abnormalities in pathways related to fluid shear stress and smooth muscle development in these two cell subtypes, with key genes such as Fos and Nr4a1 playing potentially critical roles. Pseudotime analysis provided insights into the differentiation stages of these two potentially critical endothelial cell types, with aEC\_cluster representing the terminal stage of differentiation where the expression of Fos and Nr4a1 gradually increased over proposed time. Additionally, our transcriptional regulatory network analysis revealed the presence of regulon modules closely related to aEC\_cluster, in which Gata2 and Irf7 extensively regulated the molecular network involving Fos, Nr4a1, and other significant alterations in the aEC\_cluster. Collectively, our results suggest that these two endothelial cell subtypes play a potentially critical role in the pathogenesis of large atherosclerotic stroke and that the identified genes and pathways could be potential therapeutic targets for the treatment of stroke.

Under physiological conditions, endothelial cells play a crucial role in maintaining a dynamic balance between anticoagulation and fibrinolysis, and in vascular injury and repair (Krüger-Genge et al., 2019). Conversely, endothelial cell dysfunction is closely linked to cerebral small vessel disease, atherosclerosis, and stroke (Bai et al., 2022). For example, Wang et al. demonstrated that enhanced autophagy in endothelial cells contribute to increased formation of atherosclerosis (Wang et al., 2024). In our study, we investigated the genetic basis of single-cell expression profiles of endothelial cells in relation to stroke GWAS risk loci, and identified aEC\_cluster and EC2\_cluster as possible pathogenic cell subtypes in large artery atherosclerotic stroke. We found that the expression profile of aEC cells was significantly upregulated in pathways related to fluid shear



**FIGURE 3** Critical endothelial cell subtypes associated with atherosclerosis. **(A)** Uniform manifold approximation and projection (UMAP) plot for endothelial cell subtypes. **(B)** Heatmap showing the expression of marker genes in the six different cell subtypes with the color of the heat map indicating the level of expression. **(C,D)** Significantly up-regulated pathways in aEC and down-regulated pathways in EC2 obtained by GSEA enrichment analysis. **(E)** Grid diagram showing enriched gene sets and their respective representative/associated genes significantly enriched in aEC\_cluster or EC2\_cluster. The colors of the dots indicate change directions of the genes in the two cell subtypes.



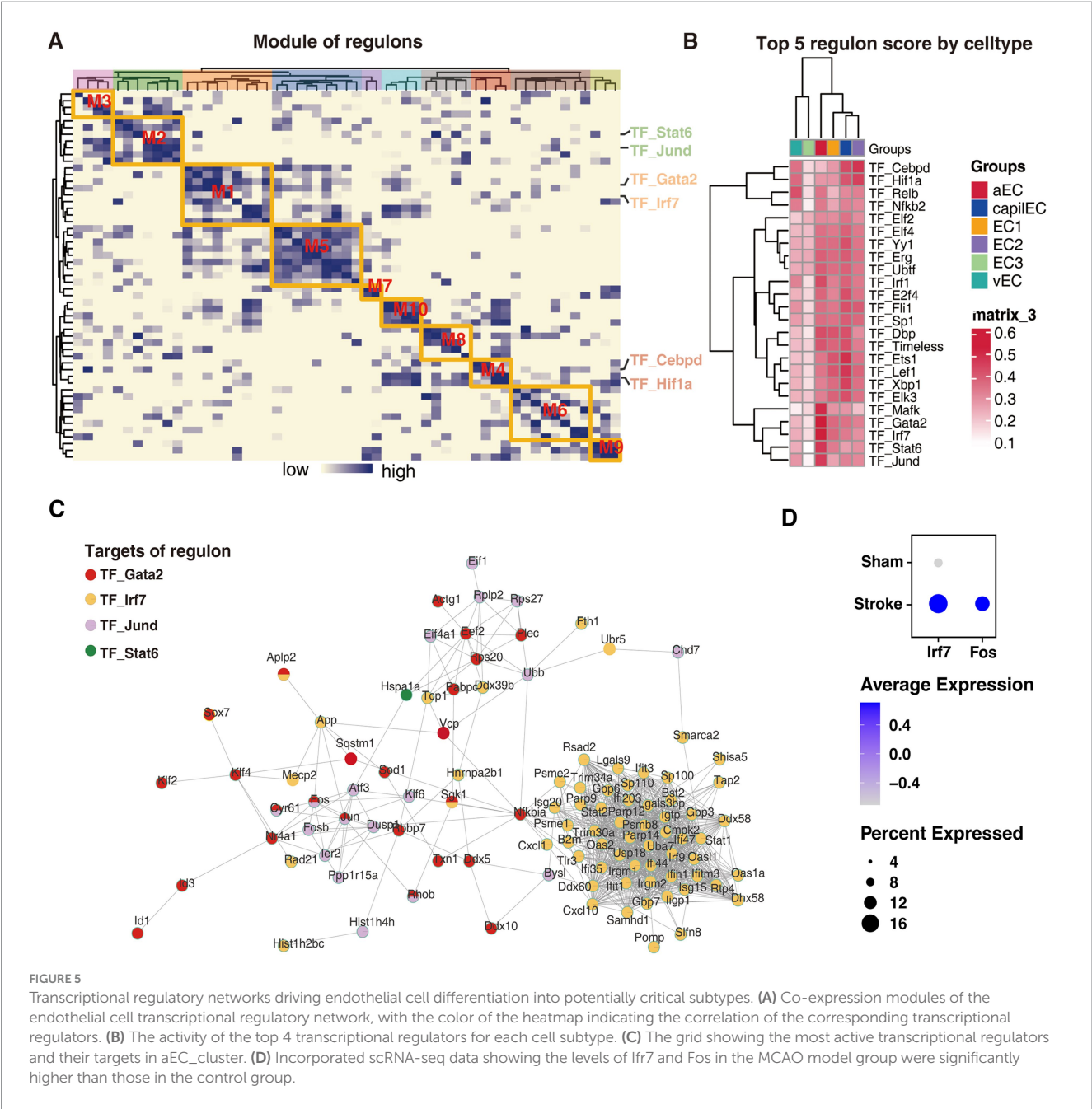
stress, atherosclerosis, and TNF signaling pathway, while EC2\_cluster showed downregulation in the muscle tissue development pathway. Both atherosclerosis and inflammation play crucial roles in the development of large artery atherosclerotic stroke (Jiang et al., 2022; Acosta et al., 2019). In addition, vascular smooth muscle is a crucial component of the vasculature, and an imbalance in its homeostasis can lead to the formation of atherosclerotic plaques (Grootaert et al., 2021). Therefore, blood flow shear factors may contribute to the shift of endothelial cells toward a stroke-causing subtype, leading to impaired regulation of smooth muscle cell homeostasis, atherosclerosis formation, and ultimately large-artery atherosclerotic stroke.

Among the differential genes significantly altered in the aEC\_cluster and EC2\_cluster, Fos plays a central role in pathways related to

fluid shear stress, atherosclerosis, and muscle tissue development. A previous study suggests that Fos plays a critical role in the development of atherosclerosis (Miao et al., 2022) and it has been demonstrated that TNF $\alpha$  can upregulate Fos levels in vascular lesions (Goetze et al., 2001). In our study, we found significant upregulation of the TNF signaling pathway in the aEC\_cluster (Figure 3), which partially supports a role for Fos in aEC\_cluster-associated atherosclerotic stroke.

Transcription factors play a crucial role in cell differentiation processes (Stadhouders et al., 2019). In this study, we investigated the transcription factors involved in the differentiation of endothelial cells into pathogenic subtypes by analyzing the transcriptional regulatory network. Our analysis revealed that the activities of Gata2, Irf7, and Junb were significantly elevated in the terminally differentiated





aEC\_cluster (Figure 5). Specifically, Gata2 and Jund together regulate the expression of Fos. GATA2 is a member of the GATA binding factors family that regulates the expression of multiple genes by interacting with other DNA motifs and transcription factors (Rodrigues et al., 2012). Previous research has shown that GATA2 can promote the progression of atherosclerosis by upregulating NRP2 (Luo et al., 2022; Yin et al., 2020). Jund, a member of the Jun family, forms the activated protein-1 complex (AP-1) by heterodimerizing with members of the Fos family, which plays a vital role in cell migration, proliferation, and apoptosis (Mechta-Grigoriou et al., 2001; Eferl and Wagner, 2003). Therefore, the upregulation of Fos through the actions of Gata2 and Jund may have mediated the differentiation of endothelial cells into the critical aEC\_cluster, therefore impacting the development of large atherosclerotic strokes. Additionally, our analysis identified numerous

targets of Irf7 regulation in the aEC\_cluster. Previous studies have demonstrated that Irf7 can exacerbate macrophage inflammation and disrupt cholesterol metabolism, thereby promoting atherosclerosis (Senatus et al., 2020). Therefore, Irf7 may represent a potential target for promoting the differentiation of endothelial cells into pathogenic subtypes.

Despite the valuable insights provided by our study, several limitations should be acknowledged. First, scRNA-seq data used for transcriptomic and regulatory analyzes were derived from mouse brain vasculature, which may not fully capture the complexity or exact cellular phenotypes present in the human cerebrovascular system. While murine models offer high-resolution cellular profiling and conserved vascular architecture, species-specific differences in gene expression, transcriptional regulation, and disease susceptibility may limit the direct

translatability of our findings. While several single-cell datasets exist for the cerebral vasculature, most are based on drop-seq technology, which has a significantly lower number of detected genes compared to the single-cell dataset obtained through SMART-seq2 (Ziegenhain et al., 2017; Svensson et al., 2018). Therefore, for the purpose of assessing the heritability of stroke risk loci, a higher number of genes detected is more favorable for inference. Hence, we selected the single-cell dataset for the cerebral vasculature obtained on the SMART-seq2 platform. Future studies incorporating human-derived single-cell or single-nucleus RNA-seq datasets, particularly from postmortem stroke patient samples or iPSC-derived vascular models, will be essential to validate and refine the identified endothelial subtypes and regulatory networks.

In addition, although transcription factors such as GATA2 and IRF7 were identified as potential key upstream regulators of pathogenic endothelial differentiation, further animal experiments are needed to validate this hypothesis. Moreover, direct pharmacological targeting of transcription factors remains challenging due to their intracellular localization and lack of defined binding pockets. Nonetheless, emerging therapeutic strategies—including antisense oligonucleotides (ASOs), RNA interference (RNAi), and CRISPR-based gene regulation tools—offer promising avenues to modulate transcription factor activity with higher specificity (Adachi et al., 2021). For example, ASO-based drugs have already shown clinical success in other neurological and cardiovascular conditions, suggesting that RNA-based therapeutics could provide feasible intervention strategies for modulating GATA2/IRF7-driven gene networks in stroke (Nitschke and Minassian, 2024; Dhuri et al., 2020).

In summary, our study highlights the significance of two potentially critical endothelial cell subtypes in the pathogenesis of large atherosclerotic stroke and identifies potential therapeutic targets. Gata2 and Irf7 emerge as theoretically critical upstream transcriptional regulators promoting endothelial cell differentiation toward critical subtypes, highlighting their potential as intervention targets for the prevention of large atherosclerotic stroke.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

## Author contributions

HD: Writing – original draft, Funding acquisition, Conceptualization, Software, Investigation, Writing – review & editing, Resources, Methodology, Formal analysis, Project administration, Supervision, Data curation. Q-XW: Investigation, Software, Data curation, Methodology, Writing – original draft, Formal analysis. MZ: Project administration, Writing – review & editing, Data curation. YL: Writing – review & editing, Supervision, Conceptualization. QP: Project administration, Supervision, Methodology, Formal analysis, Conceptualization, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Natural Science Foundation of Guizhou Province, China (Grant Qian Ke He Foundation (2022) No. 417); Guizhou Medical University (Grant 21NSFCP10); and Key Advantageous Discipline Construction Project of Guizhou Provincial Health Commission in 2023, China.

## Acknowledgments

We appreciate the GEO database and openGWAS database for making the data available.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2025.1646993/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Ranking of regulon scores across different endothelial cells. Each point in the scatter plot represents a cell type.

### SUPPLEMENTARY TABLE 1

Transcription factors and their associated modules.

### SUPPLEMENTARY TABLE 2

Regulon activity scores (AUC) across cell types.

## References

- Acosta, S. A., Lee, J. Y., Nguyen, H., Kaneko, Y., and Borlongan, C. V. (2019). Endothelial progenitor cells modulate inflammation-associated stroke. *Vasculome. Stem Cell Rev. Rep.* 15, 256–275. doi: 10.1007/s12015-019-9873-x
- Adachi, H., Hengesbach, M., Yu, Y. T., and Morais, P. (2021). From antisense Rna to Rna modification: therapeutic potential of Rna-based technologies. *Biomedicine* 9:550. doi: 10.3390/biomedicine9050550
- Bai, T., Yu, S., and Feng, J. (2022). Advances in the role of endothelial cells in cerebral small vessel disease. *Front. Neurol.* 13:861714. doi: 10.3389/fneur.2022.861714
- Barbeira, A. N., Pividori, M., Zheng, J., Wheeler, H. E., Nicolae, D. L., and Im, H. K. (2019). Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* 15:e1007889. doi: 10.1371/journal.pgen.1007889
- Bevan, S., Traylor, M., Adib-Samii, P., Malik, R., Paul, N. L., Jackson, C., et al. (2012). Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke* 43, 3161–3167. doi: 10.1161/STROKEAHA.112.665760
- Beylerli, O., Gareev, I., Beilerli, A., Ilyasova, T., Shi, H., and Sufianov, A. (2024). The function of astrocytes in cerebral infarction and potential therapeutic approaches. *Brain Hemorrhages* 5, 181–187. doi: 10.1016/j.hest.2024.03.003
- Datta, A., Sarmah, D., Kalia, K., Borah, A., Wang, X., Dave, K. R., et al. (2020). Advances in studies on stroke-induced secondary neurodegeneration (Snd) and its treatment. *Curr. Top. Med. Chem.* 20, 1154–1168. doi: 10.2174/1568026620666200416090820
- Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., et al. (2020). Antisense oligonucleotides: an emerging area in drug discovery and development. *J. Clin. Med.* 9:2004. doi: 10.3390/jcm9062004
- Edwards, D. N., and Bix, G. J. (2019). Roles of blood-brain barrier integrins and extracellular matrix in stroke. *Am. J. Physiol. Cell Physiol.* 316, C252–c263. doi: 10.1152/ajpcell.00151.2018
- Eferl, R., and Wagner, E. F. (2003). Ap-1: a double-edged sword in tumorigenesis. *Nat. Rev. Cancer* 3, 859–868. doi: 10.1038/nrc1209
- Gareev, I., Beylerli, O., Ilyasova, T., Mashkin, A., and Shi, H. (2024). The use of bioinformatic analysis to study intracerebral hemorrhage. *Brain Hemorrhages* 5, 188–196. doi: 10.1016/j.hest.2024.05.002
- GBD 2019 Stroke Collaborators (2021). Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet Neurol.* 20, 795–820. doi: 10.1016/S1474-4422(21)00252-0
- Goetze, S., Kintscher, U., Kaneshiro, K., Meehan, W. P., Collins, A., Fleck, E., et al. (2001). Tnfalpha induces expression of transcription factors c-fos, Egr-1, and Ets-1 in vascular lesions through extracellular signal-regulated kinases 1/2. *Atherosclerosis* 159, 93–101. doi: 10.1016/S0021-9150(01)00497-X
- Grootaert, M. O. J., Finigan, F., Figg, N. L., Uryga, A. K., Bennett, M. R. (2021). SIRT6 Protects Smooth Muscle Cells From Senescence and Reduces Atherosclerosis. *Circ Res* 128, 474–491. doi: 10.1161/CIRCRESAHA.120.318353
- He, L., Vanlandewijck, M., Mäe, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). Single-cell Rna sequencing of mouse brain and lung vascular and vessel-associated cell types. *Sci Data* 5:180160. doi: 10.1038/sdata.2018.160
- Holliday, E. G., Maguire, J. M., Evans, T. J., Koblar, S. A., Jannes, J., Sturm, J. W., et al. (2012). Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat. Genet.* 44, 1147–1151. doi: 10.1038/ng.2397
- Jiang, H., Ren, K., Li, T., Qian, C., Gong, S., Wang, T., et al. (2022). Correlation of the characteristics of symptomatic intracranial atherosclerotic plaques with stroke types and risk of stroke recurrence: a cohort study. *Ann. Transl. Med.* 10:658. doi: 10.21037/atm-22-2586
- Jonas, B., Anika, F., and Heinz, R. (2024). Aging and Parkinson's disease: a complex interplay of vulnerable neurons, the immune system and the blood-brain barrier. *Ageing Neurodegener. Dis.* 4:5. doi: 10.20517/and.2023.36
- Kilarski, L. L., Achterberg, S., Devan, W. J., Traylor, M., Malik, R., Lindgren, A., et al. (2014). Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology* 83, 678–685. doi: 10.1212/WNL.0000000000000707
- Krüger-Genge, A., Blocki, A., Franke, R. P., and Jung, F. (2019). Vascular endothelial cell biology: an update. *Int. J. Mol. Sci.* 20:4411. doi: 10.3390/ijms20184411
- Luo, S., Wang, F., Chen, S., Chen, A., Wang, Z., Gao, X., et al. (2022). Nrp2 promotes atherosclerosis by upregulating Parp1 expression and enhancing low shear stress-induced endothelial cell apoptosis. *FASEB J.* 36:e22079. doi: 10.1096/fj.202101250RR
- Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., et al. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* 50, 524–537. doi: 10.1038/s41588-018-0058-3
- Mehta-Grigoriou, F., Gerald, D., and Yaniv, M. (2001). The mammalian Jun proteins: redundancy and specificity. *Oncogene* 20, 2378–2389. doi: 10.1038/sj.onc.1204381
- Mei-Yun, T., Fredric, A. G., and Pamela, J. L. (2022). Review of evidence implicating the plasminogen activator system in blood-brain barrier dysfunction associated with Alzheimer's disease. *Ageing Neurodegener. Dis.* 2:2. doi: 10.20517/and.2022.05
- Miao, G., Zhao, X., Chan, S. L., Zhang, L., Li, Y., Zhang, Y., et al. (2022). Vascular smooth muscle cell c-Fos is critical for foam cell formation and atherosclerosis. *Metabolism* 132:155213. doi: 10.1016/j.metabol.2022.155213
- Nitschke, S., and Minassian, B. A. (2024). “Antisense oligonucleotide therapy for progressive myoclonus epilepsies” in Jasper's basic mechanisms of the epilepsies. eds. J. L. Noebels, M. Avoli, M. A. Rogawski, A. Vezzani and A. V. Delgado-Escueta (New York: Oxford University Press).
- Pan, Y., Cao, W., Mu, Y., and Zhu, Q. (2022). Microfluidics facilitates the development of single-cell Rna sequencing. *Biosensors (Basel)* 12:450. doi: 10.3390/bios12070450
- Profaci, C. P., Munji, R. N., Pulido, R. S., and Daneman, R. (2020). The blood-brain barrier in health and disease: important unanswered questions. *J. Exp. Med.* 217:e20190062. doi: 10.1084/jem.20190062
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402
- Rodrigues, N. P., Tipping, A. J., Wang, Z., and Enver, T. (2012). Gata-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *Int. J. Biochem. Cell Biol.* 44, 457–460. doi: 10.1016/j.biocel.2011.12.004
- Senatus, L., López-Díez, R., Egaña-Gorroño, L., Liu, J., Hu, J., Daffu, G., et al. (2020). Rage impairs murine diabetic atherosclerosis regression and implicates Irf7 in macrophage inflammation and cholesterol metabolism. *JCI Insight* 5:e137289. doi: 10.1172/jci.insight.137289
- Shehjar, F., Maktabi, B., Rahman, Z. A., Bahader, G. A., James, A. W., Naqvi, A., et al. (2023). Stroke: molecular mechanisms and therapies: update on recent developments. *Neurochem. Int.* 162:105458. doi: 10.1016/j.neuint.2022.105458
- Skrodzki, H., van Geffen, F. F., Chaves-de-Plaza, T., Höllt, E., and Eisemann, K. (2024). Hildebrandt, “Accelerating Hyperbolic t-SNE,” in IEEE Transactions on Visualization and Computer Graphics, 30, 4403–4415. doi: 10.1109/TVCG.2024.3364841
- Stadhouders, R., Fillion, G. J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*, 569, 345–354. doi: 10.1038/s41586-019-1182-7
- Svensson, V., Vento-Tormo, R., and Teichmann, S. A. (2018). Exponential scaling of single-cell Rna-seq in the past decade. *Nat. Protoc.* 13, 599–604. doi: 10.1038/nprot.2017.149
- Timshel, P. N., Thompson, J. J., and Pers, T. H. (2020). Genetic mapping of etiologic brain cell types for obesity. *eLife* 9:e55851. doi: 10.7554/eLife.55851
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386. doi: 10.1038/nbt.2859
- Van De Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., et al. (2020). A scalable Scenic workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 15, 2247–2276. doi: 10.1038/s41596-020-0336-2
- Vanlandewijck, M., He, L., Mäe, M. A., Andrae, J., Ando, K., Del Gaudio, F., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. *Nature* 554, 475–480. doi: 10.1038/nature25739
- Veltkamp, R., and Gill, D. (2016). Clinical trials of immunomodulation in ischemic stroke. *Neurotherapeutics* 13, 791–800. doi: 10.1007/s13311-016-0458-y
- Wang, C., Yang, X., Guo, Z., Zhu, G., and Fan, L. (2024). Circadian gene clock accelerates atherosclerosis by promoting endothelial autophagy. *Biotechnol. Genet. Eng. Rev.* 40, 1230–1245. doi: 10.1080/02648725.2023.2193061
- Wen-Juan, G., Hai-Ling, D., and Qi-Hui, Z. (2023). Potential effects of Chinese herbs on iron metabolism for the treatment of stroke. *Ageing Neurodegener. Dis.* 3:20. doi: 10.20517/and.2023.02
- Yang, A. C., Vest, R. T., Kern, F., Lee, D. P., Agam, M., Maat, C. A., et al. (2022). A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* 603, 885–892. doi: 10.1038/s41586-021-04369-3
- Yin, C., Vrieze, A. M., Rosoga, M., Akingbasote, J., Pawlak, E. N., Jacob, R. A., et al. (2020). Efferocytic defects in early atherosclerosis are driven by Gata2 overexpression in macrophages. *Front. Immunol.* 11:594136. doi: 10.3389/fimmu.2020.594136
- Zhang, Y., Guo, Y., Li, R., Huang, T., Li, Y., Xie, W., et al. (2023). Novel Ch25H(+) and Oasl(+) microglia subclusters play distinct roles in cerebral ischemic stroke. *J. Neuroinflammation* 20:115. doi: 10.1186/s12974-023-02799-6
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., et al. (2017). Comparative analysis of single-cell Rna sequencing methods. *Mol. Cell* 65, 631–643.e4. doi: 10.1016/j.molcel.2017.01.023