Journal of the American Heart Association

ORIGINAL RESEARCH

# Integration of Genome-Wide Association Studies With Single-Cell and Bulk Expression Quantitative Trait Locus to Identify Stroke Susceptibility Genes

Yijie He, MSc*; Tao Zhang, MM*; Ping Zhu, PhD; Shan Gao, PhD; Shiyang Wu, BS; Fengzhen Liu, MSc; Yan Chen, PhD; Yingshui Yao, PhD; Guiyou Liu, PhD

**BACKGROUND:** Previous studies have integrated genome-wide association studies with expression quantitative trait locus (eQTL) data from bulk tissues to identify stroke susceptibility genes. However, eQTL data exhibit high cell-type specificity, and genetic variants may have distinct effects across stroke subtypes.

**METHODS:** We applied the summary-data-based Mendelian randomization (MR) method to integrate eQTL data from 7 brain cell types with genome-wide association studies data for 5 stroke phenotypes (stroke, ischemic stroke, cardioembolic stroke, large artery stroke, and small vessel stroke). Results were compared with summary-data-based MR using eQTL data from 49 tissues in the Genotype-Tissue Expression project. Robustness of significant single-cell summary-data-based MR associations was assessed via MR and colocalization analyses. Further evaluations included single-cell RNA-seq differential expression, protein–protein interaction, druggability, and phenome-wide association studies.

**RESULTS:** Single-cell summary-data-based MR identified many novel significant genes not detected using bulk tissue eQTL data. Validated associations revealed 2 stroke risk genes (*LRCH1*, *ICA1L*), 3 stroke protective genes (*AHI1*, *LYRM9*, *CENPQ*), 2 large artery stroke risk genes (*LIPA*, *ELL*), and 1 ischemic stroke protective gene (*CENPQ*). Single-cell RNA-seq showed significantly increased *LIPA* expression in mouse stroke samples compared with controls. Protein–protein interaction and druggability analyses, along with phenome-wide association studies, prioritized *LIPA* and *LRCH1* as potential therapeutic targets for stroke while indicating possible adverse effects.

**CONCLUSIONS:** Integrating single-cell eQTL with stroke-subtype genome-wide association studies uncovers novel cell-type-specific causal genes and highlights promising therapeutic targets, advancing understanding of stroke pathogenesis.

**Key Words:** expression quantitative trait loci ■ genome-wide association studies ■ Mendelian randomization ■ single-cell analysis ■ stroke

Correspondence to: Guiyou Liu, PhD, Beijing Institute for Brain Disorders, Capital Medical University, Room 713, Morphology Building, No. 10, Xitoutiao, You'an Men Wai, Fengtai District, Beijing 100069, China. Email: liuguiyou1981@163.com and Yan Chen, PhD, Department of Epidemiology and Biostatistics, School of Public Health, Wannan Medical College, No. 22, Wenchang Road, Wuhu 241002, Anhui, China. Email: chenyan2010@wnmc.edu.cn and Yingshui Yao, PhD, Department of Epidemiology and Biostatistics, School of Public Health, Wannan Medical College, No. 22, Wenchang Road, Wuhu 241002, Anhui, China. Email: yaoyingshui@wnmc.edu.cn

*Y. He and T. Zhang contributed equally.

This article was sent to Jacquelyn Y. Taylor, PhD, PNP-BC, RN, FAHA, FAAN, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.125.046208

For Sources of Funding and Disclosures, see page 14.



*JAHA* is available at: www.ahajournals.org/journal/jaha

## RESEARCH PERSPECTIVE

### What New Question Does This Study Raise?

- This study integrates single-cell expression quantitative trait loci data from 7 human brain cell types with genome-wide association studies data for stroke and 4 subtype phenotypes using summary-data-based Mendelian randomization, identifying causal genes that are not detectable using bulk tissue expression quantitative trait loci data alone and prioritizing high-confidence stroke risk genes (*LRCH1*, *ICA1L*, *LIPA*, and *ELL*) as well as protective genes (*AHI1*, *LYRM9*, and *CENPQ*) through comprehensive validation including Mendelian randomization, colocalization, differential expression analysis, and functional annotation.
- Multiomics evidence further demonstrates that *LIPA* and *LRCH1* exhibit strong biological relevance, druggability, and pleiotropic profiles, highlighting their potential as promising therapeutic targets for stroke.

### What Question Should be Addressed Next?

- Future research should integrate perturbation-based and single-cell multiomics approaches to elucidate the cell-type–specific mechanisms by which microglial and neuronal regulatory architectures influence distinct stroke subtypes, and to assess whether targeting *LIPA*, *LRCH1*, or their interacting pathways can modify clinical outcomes in stroke.

### Nonstandard Abbreviations and Acronyms

| | |
|---|---|
| **AIS** | any ischemic stroke |
| **BH** | Benjamini–Hochberg |
| **CES** | cardioembolic stroke |
| **eQTL** | expression quantitative trait locus |
| **Exc** | excitatory neurons |
| **GTEx** | Genotype-Tissue Expression |
| **HEIDI** | heterogeneity in dependent instruments |
| **Inh** | inhibitory neurons |
| **LAS** | large artery stroke |
| **LD** | linkage disequilibrium |
| **Mic** | microglia |
| **MR** | Mendelian randomization |
| **OPCs** | oligodendroglial progenitor cells |
| **PheWAS** | phenome-wide association studies |
| **PPI** | protein–protein interaction |
| **SMR** | summary-data-based Mendelian randomization |
| **SVS** | small vessel stroke |

Stroke remains a devastating global health burden, ranking as the second leading cause of death worldwide, responsible for ~5 million fatalities annually.[1,2] Beyond mortality, it imposes long-term disability and socioeconomic challenges.[3,4] Although numerous risk factors—such as hypertension, diabetes, and lifestyle habits—are well established,[2,5] the identification of causative genes and their molecular mechanisms underlying stroke pathogenesis remains incomplete. Elucidating these genetic drivers is critical for advancing targeted therapeutic strategies and understanding the cause of stroke.

Recent efforts have integrated bulk tissue expression quantitative trait locus (eQTL) with stroke genome-wide association studies (GWAS) to prioritize candidate genes, primarily focusing on broad stroke phenotypes without resolving subtypes or cell-type contexts.[6–11] However, accumulating evidence suggests that single-cell transcriptomic approaches can reveal regulatory heterogeneity obscured in bulk-tissue analyses.[12] These studies analyzed GWAS data for stroke (broad definition), ischemic stroke, or intracerebral hemorrhage but uniformly relied on bulk tissue eQTL—ignoring both cellular heterogeneity and subtype-specific genetic architectures. This strategy nonetheless faces key limitations. First, genetic heterogeneity exists across stroke subtypes, suggesting distinct pathogenic pathways.[13] Second, different cell types typically exhibit distinct eQTL patterns, and bulk tissue eQTL may obscure cell-type-specific regulatory effects.[14,15] Consequently, integrating single-cell eQTL data with subtype-specific GWAS signals could enhance the precision of causal gene discovery, enabling the identification of stroke risk genes across different stroke subtypes and cell type combinations.

To address these limitations, we employed a transcriptome-wide Mendelian randomization (MR) framework through summary-data-based Mendelian randomization (SMR) analysis to integrate stroke GWAS data (including 5 stroke phenotypes) with single-cell eQTL data from 7 human brain cell types.[16,17] This approach uses cis-eQTL as instrumental variables to test whether genetically predicted gene expression is causally associated with stroke risk, helping to prioritize specific genes whose products may mediate genetic risk beyond mere genomic localization. Candidate genes were rigorously validated via heterogeneity in dependent instruments (HEIDI) tests, MR, and colocalization

analysis to exclude pleiotropic effects. Downstream functional evaluations included single-cell transcriptomic profiling, protein–protein interaction (PPI) networks, drug-target assessments, and phenome-wide association studies (PheWAS) to explore biological pathways, therapeutic potential, and pleiotropic risks. Our multiomics framework not only identifies robust stroke-associated genes but also provides insights into their roles in stroke pathophysiology and druggability, bridging genetic discoveries to translational applications.

# METHODS

## Data Availability

Single-cell eQTL summary data were obtained from https://doi.org/10.7303/syn52335732. Stroke and stroke subtypes GWAS data are available from the GWAS Catalog (GCST90104539-GCST90104543). EQTL files in binomial effect size display (BESD) format from 49 tissues in the Genotype-Tissue Expression (GTEx) project can be downloaded from http://cnsgenomics.com/software/smr/#DataResource. European linkage disequilibrium (LD) reference files from the 1000 Genomes can be obtained from https://kauwelab.byu.edu/Portals/22/adgc_combined_1000G_09192014.pdf.

## Data Source

We obtained summary statistics of GWAS for individuals of European ancestry for 5 distinct stroke phenotypes. These included stroke (GCST90104539; 73 652 cases and 1 234 808 controls), which represents a broad definition encompassing all stroke types; any ischemic stroke (AIS, GCST90104540; 62 100 cases and 1 234 808 controls), which includes all ischemic strokes regardless of pathogenetic subtype; cardioembolic stroke (CES, GCST90104541; 10 804 cases and 1 234 808 controls); large artery stroke (LAS, GCST90104542; 6399 cases and 1 234 808 controls); and small vessel stroke (SVS, GCST90104543; 6811 cases and 1 234 808 controls).[16] Herein, CES, LAS, and SVS are the 3 principal pathogenetic subtypes of ischemic stroke. Single-cell eQTL data were derived from 7 cell types in the neocortex of 424 individuals, including astrocytes, endothelial cells, excitatory neurons (Exc), inhibitory neurons (Inh), microglia (Mic), oligodendrocytes, and oligodendroglial progenitor cells (OPCs).[17] In addition, bulk tissue eQTL data from 838 individuals across 49 tissues (n=73–670) were obtained from the GTEx project.[18]

## Study Design

Figure 1 illustrates the overall study design. We first performed SMR analysis using eQTL data from 7 brain cell types (n=424) and GWAS data on 5 stroke phenotypes. Next, SMR analysis was conducted on eQTL data from 49 GTEx tissues using the same GWAS data sets, allowing comparison with single-cell eQTL SMR results. We then performed 2-sample MR and colocalization analysis not only to mitigate potential low power in HEIDI (eg, due to sparse cis-eQTL signals) but also as a supplement and robustness check for SMR, providing orthogonal evaluations of shared causal variants versus LD confounding. Importantly, although both SMR and MR used the same summary-level data, they employed distinct instrumental variable selection strategies: SMR used the top cis-eQTL SNV per gene, whereas MR employed multiple independent SNVs ($P$<1e-5) as instruments. Finally, we performed functional enrichment analysis for SMR-significant genes, followed by single-cell analysis of validated genes to assess their expression across cell types and stroke conditions, as well as PPI analysis, druggability evaluation, and PheWAS analysis to identify potential therapeutic targets and assess clinical applicability.

## Statistical Analysis

Using summary data from GWAS of various stroke subtypes, eQTL information from diverse bulk tissues, and single-cell eQTL studies, we applied the SMR (version 1.3.1) software to integrate the data from these sources to explore the susceptibility genes of stroke.[19] In the study, the SMR method, resembling the MR strategy, employed the top eQTL as an instrumental variable to estimate the impact of gene expression on the trait. The HEIDI filtering test within the SMR workflow used multiple single-nucleotide variants (SNVs) in cis-eQTL regions to eliminate significant SMR associations attributed to LD between disease-related SNVs and eQTL SNVs. Specifically, HEIDI tests the null hypothesis of no heterogeneity in SNV effects (consistent with pleiotropy or causality at a shared variant); a low $P$ value (rejection of null) indicates heterogeneity likely due to LD. However, HEIDI's power to detect such heterogeneity is limited in regions with few independent cis-eQTL SNVs (eg, <10), potentially leading to false retention of LD-attributable associations. We applied a threshold of HEIDI $P$>0.05 to retain associations consistent with the null hypothesis.

For bulk tissue, we acquired eQTL data in BESD format for 49 tissues from the SMR website. For single-cell eQTL data across 7 cell types, we followed the SMR online tutorial to convert them into the necessary BESD format using the "make-besd" function in the SMR software. Gene and SNV information files were updated based on positions from Ensembl version GRCh38.99 (Homo_sapiens.GRCh38.99.gtf) and SNV data from the eQTL files using the "--update-esi" and "--update-epi" functions.[20] Subsequently, SMR analysis was conducted using the "smr" function with bulk tissues BESD files, single-cell BESD files, LD reference files from the 1000 Genomes Phase 3,[21] and GWAS data for various

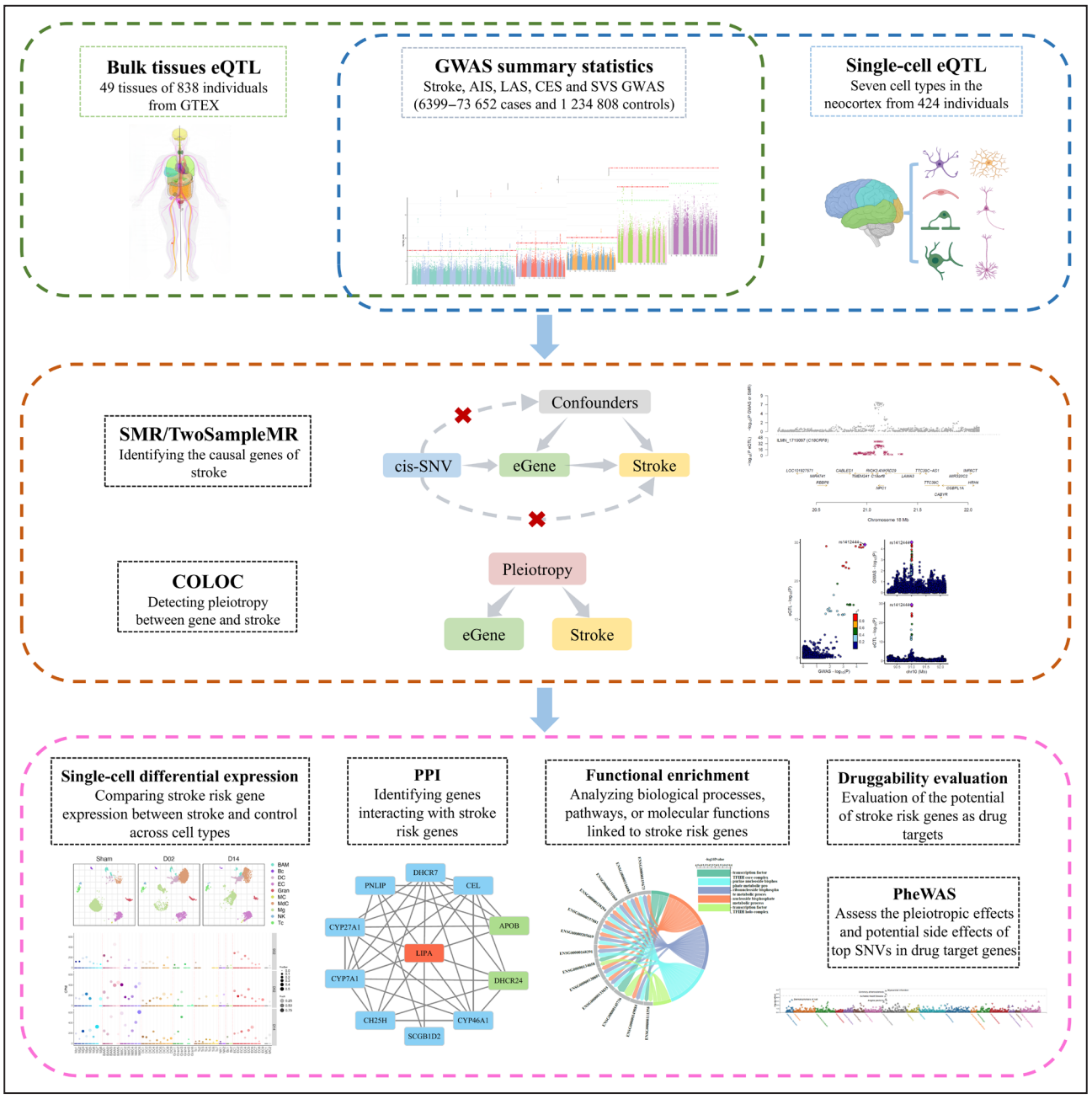


**Figure 1. Flow chart of the study design.**
AIS indicates any ischemic stroke regardless of subtype; CES, cardioembolic stroke; eQTL, expression quantitative trait locus; GTEx, Genotype-Tissue Expression; GWAS, genome-wide association studies; LAS, large-artery atherosclerotic stroke; MR, Mendelian randomization; PheWAS, phenome-wide association studies; PPI, protein–protein interaction; SMR, summary-data-based Mendelian randomization; SNV, single-nucleotide variant; and SVS, small-vessel stroke.

stroke subtypes. Finally, Benjamini–Hochberg (BH) correction was applied to adjust the *P* values derived from the SMR analysis.[22] Throughout this article, the term '*q*-value' refers to this BH-adjusted *P* value.

## Functional Enrichment Analyses

We performed Gene Ontology enrichment analysis using the clusterProfiler R package to characterize functional profiles of single-cell SMR significant genes ($P<0.05$).[23] Enrichment analysis was performed for disease-cell combinations with >10 significant SMR genes ($P<0.05$). This conservative threshold was applied because, as highlighted by Huang et al., insufficient gene set size compromises the statistical validity of enrichment analysis.[24] Therefore, the >10 cutoff balances inclusiveness with analytical reliability while mitigating overinterpretation of underpowered results. To ensure the robustness

of the enrichment results, a stringent BH-adjusted *P* value threshold of $q<0.05$ was applied to identify significantly enriched Gene Ontology terms. This 2-step approach—using a nominal *P* value threshold for gene selection followed by BH correction at the pathway level—is an established strategy to maintain statistical rigor while preserving power for biological hypothesis generation. Subsequently, results with $q<0.05$ were selected from the enrichment analysis.

## Mendelian Randomization Analysis

To assess the robustness of the putative causal relationships identified by SMR using a different instrumental variant selection strategy, we performed an MR analysis on the top candidate genes (with SMR $q<0.05$ and HEIDI $P>0.05$). This MR approach provides a more robust framework that accommodates multiple instrumental variables and explicitly tests for assumptions like pleiotropy. For significantly associated genes identified by SMR, we conducted MR analysis as a robustness check using the "TwoSampleMR" R package (version 0.6.3).[25] Initially, we performed instrumental variant selection by filtering variants with *P* values <1e to 5 in the exposure, namely the eQTL data. Subsequently, leveraging the 1000 Genomes European panel, we employed the 'clumping' method in PLINK (clump_kb=100, clump_r2=0.1) to select variations with low LD.[26] Following this, to minimize the potential for weak instrument bias, we filtered instrumental variables based on a commonly applied threshold of F-statistic >10.[27] For genes with multiple instrumental variants, we used the inverse-variance weighting, weighted median, and MR-Egger methods for MR analysis. Genes with 2 instrumental variants were analyzed using the inverse-variance weighting method, and those with a single instrumental variant were assessed using the Wald ratio method. Moreover, for genes with >2 instrumental variants, we conducted Cochran's Q and MR-Egger intercept tests to evaluate horizontal pleiotropy and heterogeneity, respectively.[28,29]

## Colocalization Analysis

We performed Bayesian colocalization analysis using the "coloc" R package on eQTL data for the cell type of interest and GWAS data for the stroke subtype to determine whether SMR-significant genes and the stroke subtype share the same causal variants.[30] We evaluated 5 hypotheses within coloc: H0: no association, H1: association with gene expression only, H2: association with stroke subtype only, H3: independent associations, H4: shared causal variant for gene expression and stroke subtype. A posterior probability of hypothesis 4 ≥0.75 is considered strong colocalization, and genes showing strong colocalization are used for downstream analysis.

Initially, the sdY parameter for eQTL data was set to 1 because gene expression had already undergone quantile normalization, following a standard normal distribution. Subsequently, the "coloc.abf" function was applied to conduct the coloc analysis with default parameter settings. Finally, the "locuscomparer" R package (version 1.0.0) was used to visualize regional results.[31]

## Differential Expression Analysis Across Cell Types and Between Stroke and Control

To independently assess the functional relevance of our SMR-prioritized candidate genes and investigate their expression dynamics within a pathological stroke context, we used a publicly available single-cell transcriptomic data set (https://anratherlab.shinyapps.io/strokevis/) from a mouse model of ischemic stroke.[32] This data set was chosen because the transient middle cerebral artery occlusion model recapitulates human ischemic stroke, and it provided transcriptomic measurements in immune/inflammatory cells at 2 days post stroke, allowing us to evaluate acute expression changes. It is important to emphasize that this analysis aimed to evaluate the poststroke expression response of these stroke susceptibility genes as supporting evidence for their biological relevance, not to identify the initial disease cause. Differential expression analysis and significance calculations (including log2FC and BH-adjusted *P* value) were performed by the strokevis platform's standardized pipeline. Initially, gene symbols were queried within the database. Subsequently, in the "Brain" module's "Expression Level" submodule, bubble plots depicting expression levels of different cell types pre and poststroke were downloaded, and uniform manifold approximation and projection plots illustrating gene expression across different cell types were obtained from the "UMAP Plot" submodule. Finally, in the "DiffExpression (clusters)" and "DiffExpression (time point)" submodules, we explored the differential expression of the gene across different cell types and between "2-Day Sham" and "2-Day Stroke" conditions.

## Protein–Protein Interaction and Druggability Evaluation

We conducted PPI analysis and druggability evaluation to identify potential drug targets. Initially, to comprehensively predict potential drug target genes, we used the "Protein by Name" module in the PPI database (https://string-db.org/) to explore genes interacting with target genes that have passed all validations.[33] Then, we searched the DGIdb database (https://dgidb.org/) for gene–drug interactions to investigate the potential of all these genes as therapeutic targets for stroke.[34]

## Phenome-Wide Association Analysis

Using the PheWeb database (https://pheweb.org/),[35] a PheWAS analysis was conducted to evaluate the pleiotropic effects and potential adverse reactions of the top SNVs for the drug target genes. The database selection included the "UKBiobank TOPMed-imputed" data sets, which encompass 1400 electronic health record-derived broad PheWAS codes in 400 000 White British individuals. Following the SNV query, locus zoom plots were saved, along with phenotypes and their associated information, with *P* values <0.05.

## Ethics Approval

The use of data from all GWAS was approved by the respective institutional review boards, and all participants provided written informed consent.

# RESULTS

## SMR and HEIDI Tests Between Single-Cell eQTL, Bulk eQTL, and Stroke GWAS Data

We integrated eQTL data from 7 cell types in the neocortex with GWAS data on 5 stroke phenotypes using SMR to explore the association between gene expression in different brain cells and stroke. Similarly, we also integrated eQTL data from 49 tissues from GTEx with stroke GWAS data and compared the results with single-cell SMR results. Initially, we conducted a statistical analysis of the single-cell SMR results, identifying 5069 stroke subtype–cell type–gene combinations with SMR *P* values <0.05, out of a total of 73 294 tested combinations (Table S1). This initial, more inclusive threshold was used for comparative and descriptive purposes (Figure 2), before stringent multiple-testing correction. Due to some genes appearing in multiple cell types or stroke subtypes simultaneously, after deduplication, 2288 unique genes with SMR *P* values <0.05 were identified, out of 7793 genes tested in total. The cellular distribution of genes with SMR *P* values <0.05 was visualized (Figure 2B), revealing that Exc had the highest number of SMR significant genes, followed by Inh. Among all combinations, AIS and Exc showed the highest number of SMR significant genes, totaling 457. Similarly, visualizing the distribution by disease (Figure 2C) showed that AIS had the highest number of SMR significant genes, followed by stroke. Furthermore, a statistical analysis (Table S2) and visualization (Figure 2A) were conducted on genes with SMR *P* values <0.05 in bulk tissues, revealing a total of 9926 significant SMR genes, with the highest number of significant genes found in the nerve tibial, thyroid, and testis tissues. We note that this ranking is primarily influenced by differences in sample size across GTEx tissues (nerve tibial n=532, thyroid n=574, testis n=322) and the exceptionally high eQTL diversity in testis, and may partly reflect chance variation in the absence of formal statistical testing for tissue enrichment.[18] To explore potential associations between GTEx brain tissue and cell types, we compared 2631 significant SMR genes in GTEx brain tissue with single-cell SMR results. After selecting the intersection of significant genes between each cell type and brain tissue, we statistically analyzed and visualized the proportion of intersection genes within brain tissue SMR significant genes (Figure 2D) and single-cell SMR significant genes (Figure 2E). It was observed that the intersection genes between Exc and brain tissues had the highest proportion among brain tissue SMR significant genes, aligning with the earlier finding of the high number of significant genes in Exc. The cerebellar tissues showed the highest proportion of overlapping genes with single-cell SMR results, followed by the cerebellar hemisphere and cortex. Of the nominally significant single-cell SMR associations (*P*<0.05), >75% (Figure 2E) were not detected in the corresponding GTEx brain tissue SMR analyses. This pattern is consistent with possible cell-type-specific regulation that is diluted in bulk tissue but may equally reflect differences in statistical power, false positives in either data set, or chance variation. This highlights the necessity of integrating single-cell eQTL data with stroke GWAS—an approach fundamentally distinct from bulk tissue analysis. Then, further filtering of the single-cell SMR results identified 15 genes with SMR *q*-values <0.05 across various cell types and stroke subtypes and displayed their chromosomal positions (Figure 3). This set of high-confidence genes, which passed multiple testing correction, was carried forward for all subsequent validation analyses. Among 35 cell-stroke combinations, 11 combinations had genes with SMR *q*-values <0.05. Lastly, HEIDI tests were performed on 15 genes with SMR *q* value <0.05, and 12 genes with HEIDI *P* value >0.05 were screened out (corresponding to 15 disease-cell-gene combinations). Among these, the HEIDI test used 9 SNVs for the LAS-Mic-*ELL* combination, 19 SNVs for the AIS-astrocytes-*HECTD4* combination, and 20 SNVs for each of the remaining 13 combinations. We note that although these genes passed HEIDI, the test's reliability varies by locus-specific power, which we addressed through subsequent MR and colocalization filters.

## Functional Enrichment Analyses of Single-Cell SMR Significant Genes

By summarizing the single-cell SMR results, we identified that 30 out of 35 disease-cell combinations contained >10 SMR significant genes (*P*<0.05). Functional enrichment analysis was then performed for these 30 combinations, and the results were visualized using bar plots, where the *x*-axis represents the $-\log_{10}$(*P*value)

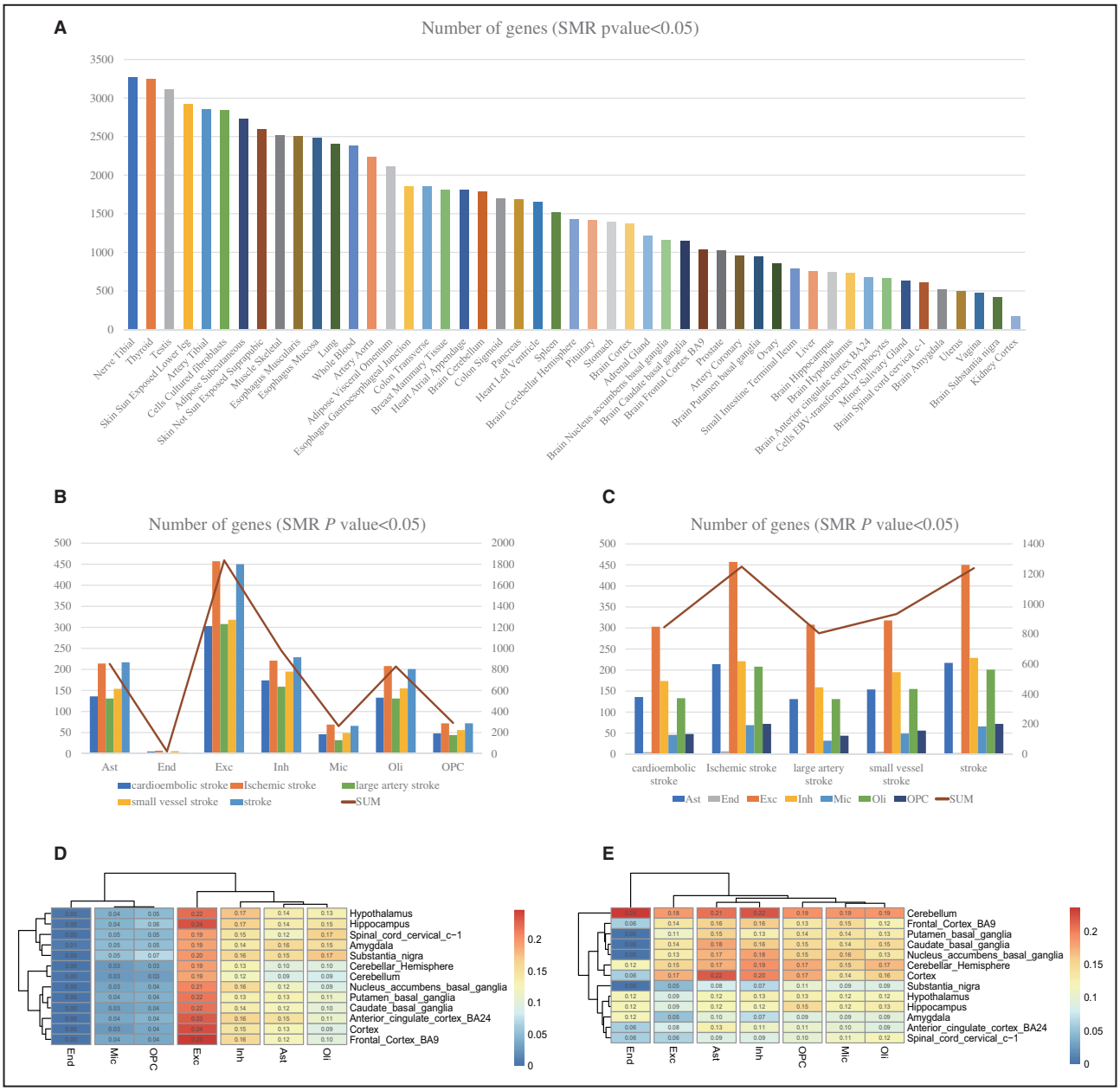


**Figure 2. Visualization of SMR-significant genes across cell types, stroke subtypes, and GTEx tissues.**
**A**, Distribution of SMR-significant genes (SMR *P*<0.05) across GTEx tissues. The top 3 tissues with the highest number of significant genes are the nerve tibial, thyroid, and testis. **B**, Distribution of SMR-significant genes across different neocortical cell types. Excitatory neurons show the highest number of significant genes, followed by inhibitory neurons. The bar height (left *y*-axis) indicates the number of significant genes for each stroke subtype-cell type combination. The connected SUM line (right *y* axis) represents the total number of unique significant genes aggregated across all stroke subtypes for each cell type. **C**, Distribution of SMR-significant genes across stroke and stroke subtypes. Stroke has the highest number of significant genes, followed by ischemic stroke. The bar height (left *y* axis) indicates the number of significant genes for each stroke subtype-cell type combination. The connected SUM line (right *y*-axis) represents the total number of unique significant genes aggregated across all cell types for each stroke subtype. **D**, Intersection ratio from bulk tissue perspective: Percentage of GTEx brain tissue SMR-significant genes also detected in single-cell SMR for each cell type. For example, 0.13 indicates 13% of cortex SMR genes overlap with Exc SMR genes. **E**, Intersection ratio from single-cell perspective: Percentage of single-cell SMR-significant genes also detected in GTEx brain tissue SMR for each cell type. For example, 0.15 indicates 15% of Mic SMR genes overlap with cortex SMR genes. Ast indicates astrocytes; EBV, Epstein–Barr virus; End, endothelial cells; Exc, excitatory neurons; GTEx, Genotype-Tissue Expression; Inh, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPCs, oligodendrocyte progenitor cells; and SMR, summary-data-based Mendelian randomization.

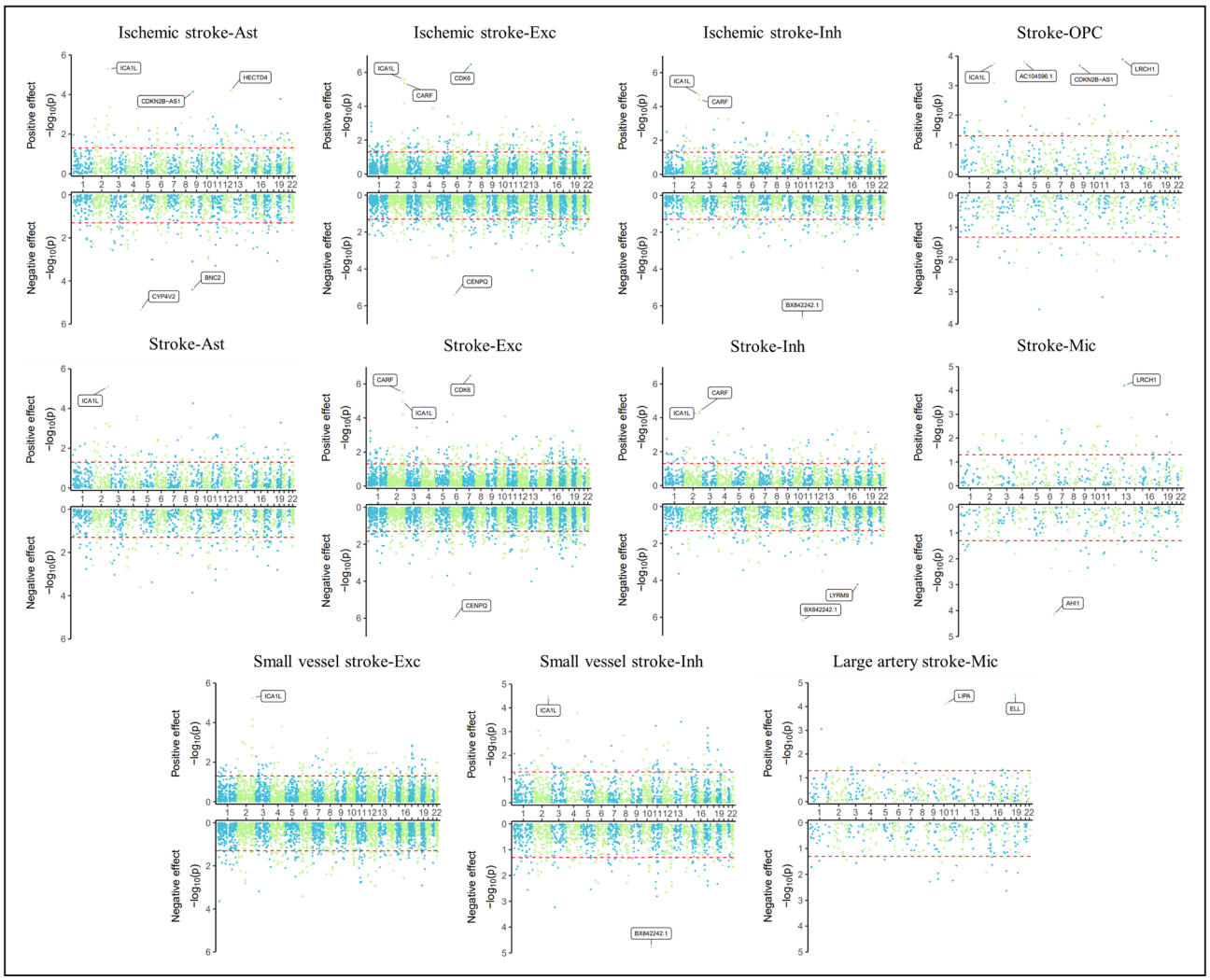


**Figure 3. Chromosomal distribution of single-cell SMR significant genes across stroke subtypes and cell types.**
Fifteen genes with SMR *q*-value <0.05 (BH adjusted) were identified through the integration of single-cell eQTL data from 7 brain cell types and GWAS data for stroke and its subtypes. Eleven cell-subtype combinations (out of 35 total) containing genes with *q*-value <0.05 were visualized. The plot displays genes based on chromosomal position, *P* value, and effect direction, with labels for genes meeting the *q*-value threshold of 0.05. The red dashed line indicates the *P* value threshold of 0.05. Ast indicates astrocytes; BH, Benjamini–Hochberg; End, endothelial cells; eQTL, expression quantitative trait locus; Exc, excitatory neurons; GWAS, genome-wide association studies; Inh, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; OPCs, oligodendrocyte progenitor cells; and SMR, summary-data-based Mendelian randomization.

of each enriched term. Statistical significance after multiple-testing correction was indicated by the *q*-value, with a threshold of $q<0.05$ used to define significantly enriched terms. Functional enrichment analysis revealed that 15 of the 30 combinations contained significantly enriched terms ($q<0.05$), among which 3 combinations exhibited ≥10 significantly enriched terms: CES-Mic, SVS-Mic, CES-OPC (Figure 4; Table S3).

For CES-Mic, the enriched functions are primarily involved in cytoskeletal dynamics and cell motility (eg, regulation of actin nucleation, actin cytoskeleton reorganization, lamellipodium assembly), immune cell function and polarization (eg, establishment of T-cell polarity, positive regulation of macrophage migration, phagocytosis), cellular signaling and membrane organization (eg, purinergic nucleotide receptor signaling pathway, regulation of plasma membrane organization), and muscle function adaptation (eg, muscle atrophy, relaxation of cardiac muscle, muscle adaptation). For SVS-Mic, the enriched functions are mainly associated with cellular adhesion and structural specializations (eg, focal adhesion, cell-substrate junction, uropod) and transcriptional machinery (eg, TFIIH (transcription factor II H) core complex, carboxy-terminal domain protein kinase complex); For CES-OPC, the enriched functions are predominantly related to intercellular communication and vascular regulation (eg, cell communication by electrical coupling, vascular process in circulatory

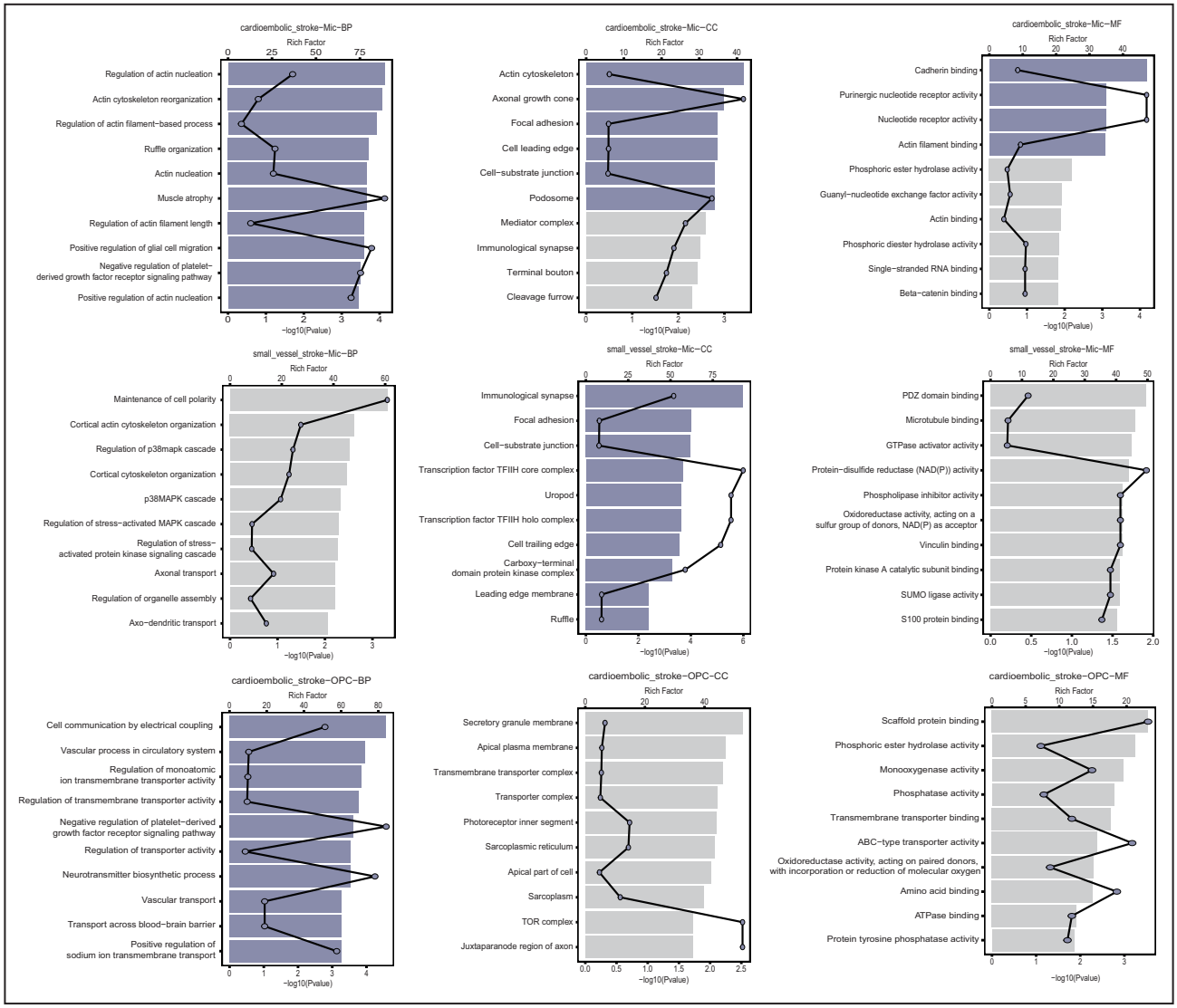


**Figure 4. Functional enrichment analysis of single-cell SMR significant genes.**
Enrichment analysis was conducted for 30 disease–cell combinations with >10 SMR significant genes (nominal $P<0.05$). Among these, only 3 combinations exhibited ≥10 significantly enriched terms (FDR $q<0.05$). The bar plot displays the enriched terms identified in these 3 combinations. The bottom *x*-axis corresponds to the bar plot and represents the $-\log_{10}(P \text{ value})$ from the enrichment analysis for each term. Bar colors indicate whether the term reached statistical significance after Benjamini–Hochberg correction: dark blue for terms with FDR $q<0.05$ and gray for terms with $q\geq0.05$. The top *x*-axis corresponds to the dot plot and represents the RichFactor, defined as the ratio of the foreground gene ratio to the background gene ratio (GeneRatio/BgRatio). Ast indicates astrocytes; BP, biological process; CC, cellular component; End, endothelial cells; Exc, excitatory neurons; FDR, false discovery rate; Inh, inhibitory neurons; MAPK, mitogen-activated protein kinase; MF, molecular function; Mic, microglia; NADP, nicotinamide adenine dinucleotide phosphate; OPCs, oligodendrocyte progenitor cells; SMR, summary-data-based Mendelian randomization; SUMO, small ubiquitin-like modifier; TFIIH, transcription factor II H; and TOR, target of rapamycin.

system, transport across blood–brain barrier) and ion transport homeostasis (eg, regulation of monoatomic ion transmembrane transporter activity, positive regulation of sodium ion transmembrane transport).

## Robustness Assessment of SMR Findings Using Mendelian Randomization

To corroborate the SMR results and test their robustness to a different analytical framework, we performed a 2-sample MR analysis using the "TwoSampleMR" R package on 12 genes from the single-cell SMR results with SMR *q*-values <0.05 and HEIDI *P* values >0.05. This MR analysis aimed to assess whether the eQTL-driven changes in gene expression may potentially influence stroke and its subtypes. In astrocytes, the upregulation of *BNC2* and *CYP4V2* reduces the risk of AIS (95% CI, 0.93–0.96, *P*=1.17E-09; 95% CI, 0.95–0.98, *P*=9.55E-07), the upregulation of *CDKN2B-AS1* and *HECTD4* increases the risk of AIS

| Cell type | Exposure | Outcome | Method | No.SNVs | *P* value | OR(95%CI) |
|---|---|---|---|---|---|---|
| Ast | BNC2 | Ischemic_stroke | MR Egger | 5 | 2.08e−01 | 0.96(0.92 to 1.01) |
| Ast | BNC2 | Ischemic_stroke | Weighted median | 5 | 2.54e−06 | 0.95(0.93 to 0.97) |
| Ast | BNC2 | Ischemic_stroke | Inverse variance weighted | 5 | 1.17e−09 | 0.95(0.93 to 0.96) |
| Ast | CDKN2B−AS1 | Ischemic_stroke | Wald ratio | 1 | 1.01e−05 | 1.07(1.04 to 1.1) |
| Ast | CYP4V2 | Ischemic_stroke | MR Egger | 5 | 4.53e−01 | 0.96(0.87 to 1.06) |
| Ast | CYP4V2 | Ischemic_stroke | Weighted median | 5 | 9.96e−07 | 0.96(0.94 to 0.98) |
| Ast | CYP4V2 | Ischemic_stroke | Inverse variance weighted | 5 | 9.55e−07 | 0.96(0.95 to 0.98) |
| Ast | HECTD4 | Ischemic_stroke | MR Egger | 8 | 3.72e−01 | 1.13(0.88 to 1.46) |
| Ast | HECTD4 | Ischemic_stroke | Weighted median | 8 | 6.46e−23 | 1.15(1.12 to 1.19) |
| Ast | HECTD4 | Ischemic_stroke | Inverse variance weighted | 8 | 4.13e−51 | 1.15(1.13 to 1.17) |
| Exc | CENPQ | Ischemic_stroke | MR Egger | 6 | 4.45e−01 | 0.97(0.9 to 1.04) |
| Exc | CENPQ | Ischemic_stroke | Weighted median | 6 | 4.96e−10 | 0.94(0.92 to 0.96) |
| Exc | CENPQ | Ischemic_stroke | Inverse variance weighted | 6 | 2.25e−19 | 0.93(0.92 to 0.95) |
| Exc | CENPQ | stroke | MR Egger | 6 | 6.36e−01 | 0.98(0.92 to 1.05) |
| Exc | CENPQ | stroke | Weighted median | 6 | 1.99e−09 | 0.94(0.92 to 0.96) |
| Exc | CENPQ | stroke | Inverse variance weighted | 6 | 2.52e−21 | 0.93(0.92 to 0.95) |
| Inh | LYRM9 | stroke | Wald ratio | 1 | 1.11e−05 | 0.9(0.86 to 0.95) |
| Mic | ELL | large_artery_stroke | Inverse variance weighted | 2 | 4.98e−06 | 1.37(1.2 to 1.56) |
| Mic | LIPA | large_artery_stroke | Inverse variance weighted | 2 | 2.53e−06 | 1.17(1.1 to 1.25) |
| Mic | AHI1 | stroke | MR Egger | 5 | 7.83e−01 | 0.99(0.93 to 1.06) |
| Mic | AHI1 | stroke | Weighted median | 5 | 6.66e−10 | 0.96(0.94 to 0.97) |
| Mic | AHI1 | stroke | Inverse variance weighted | 5 | 6.20e−15 | 0.96(0.95 to 0.97) |
| Mic | LRCH1 | stroke | MR Egger | 3 | 4.68e−01 | 1.1(0.93 to 1.3) |
| Mic | LRCH1 | stroke | Weighted median | 3 | 1.72e−07 | 1.09(1.06 to 1.13) |
| Mic | LRCH1 | stroke | Inverse variance weighted | 3 | 3.21e−11 | 1.09(1.06 to 1.12) |
| OPC | AC104596.1 | stroke | MR Egger | 4 | 3.92e−01 | 1.04(0.97 to 1.12) |
| OPC | AC104596.1 | stroke | Weighted median | 4 | 7.62e−08 | 1.06(1.04 to 1.08) |
| OPC | AC104596.1 | stroke | Inverse variance weighted | 4 | 7.24e−11 | 1.05(1.04 to 1.07) |
| OPC | CDKN2B−AS1 | stroke | Wald ratio | 1 | 7.57e−06 | 1.08(1.04 to 1.12) |
| OPC | ICA1L | stroke | Inverse variance weighted | 2 | 6.09e−15 | 1.13(1.1 to 1.17) |
| OPC | LRCH1 | stroke | Wald ratio | 1 | 4.68e−07 | 1.1(1.06 to 1.14) |

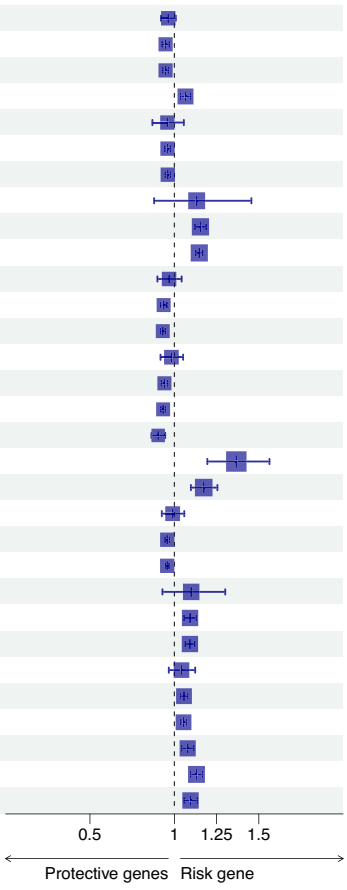


**Figure 5. Forest plots illustrating that all 12 SMR significant genes remained significant (*P*<0.05) in the MR analysis, with effect directions consistent with SMR results.**
Ast indicates astrocytes; Exc, excitatory neurons; Inh, inhibitory neurons; Mic, microglia; MR, Mendelian randomization; Oli, oligodendrocytes; OPCs, oligodendrocyte progenitor cells; OR, odds ratio; and SNV, single-nucleotide variant.

(95% CI, 1.04–1.1, *P*=1.01E-05; 95% CI, 1.13–1.17, *P*=4.13E-51); In Exc, the upregulation of *CENPQ* reduces the risk of AIS and stroke (95% CI, 0.92–0.95, *P*=2.25E-19; 95% CI, 0.92–0.95, *P*=2.52E-21); In Inh, the upregulation of *LYRM9* reduces the risk of stroke (95% CI, 0.86–0.95, *P*=1.11E-05); In Mic, the upregulation of *ELL* and *LIPA* increases the risk of LAS (95% CI,1.2–1.56, *P*=4.98E-06; 95% CI, =1.1–1.25, *P*=2.53E-06), the upregulation of *AHI1* reduces the risk of stroke (95% CI, 0.95–0.97, *P*=6.20E-15), the upregulation of *LRCH1* increases the risk of stroke (95% CI, 1.06–1.12, *P*=3.21E-11); In OPCs, the upregulation of *AC104596.1*, *CDKN2B-AS1*, *ICA1L*, and *LRCH1* increases the risk of stroke (95% CI, 1.04–1.07, *P*=7.24E-11; 95% CI, 1.04–1.12, *P*=7.57E-06; 95% CI, 1.1–1.17, *P*=6.09E-15; 95% CI, 1.06–1.14, *P*=4.68E-07) (Figure 5). Overall, all 12 significant genes from the single-cell SMR results remained significant (*P*<0.05) in the MR analysis, with effect directions fully consistent with those observed in the SMR results. Sensitivity analyses for genes with multiple instruments revealed no significant evidence of heterogeneity (Cochran's Q *P*>0.05) or horizontal pleiotropy (MR-Egger intercept *P*>0.05), supporting the robustness of these causal inferences (Tables S4 and S5).

## Colocalization Prioritizes High-Confidence Stroke and Stroke Subtype Risk Genes

To determine whether the significant SMR genes share a common underlying genetic signal (potentially a shared causal variant) with stroke and its subtypes, we conducted colocalization analysis. First, we found strong evidence of colocalization (posterior probability of hypothesis 4>0.75, representing >75% probability of a shared causal variant) for 7 out of the 12 significant single-cell SMR genes with stroke and its subtypes (Table S6). These 7 genes corresponded to 9 distinct stroke subtype-cell type-gene combinations. Regional associations for these genes with stroke and stroke subtype GWAS data were visualized in Figure 6, which demonstrates the critical advantage of integrating

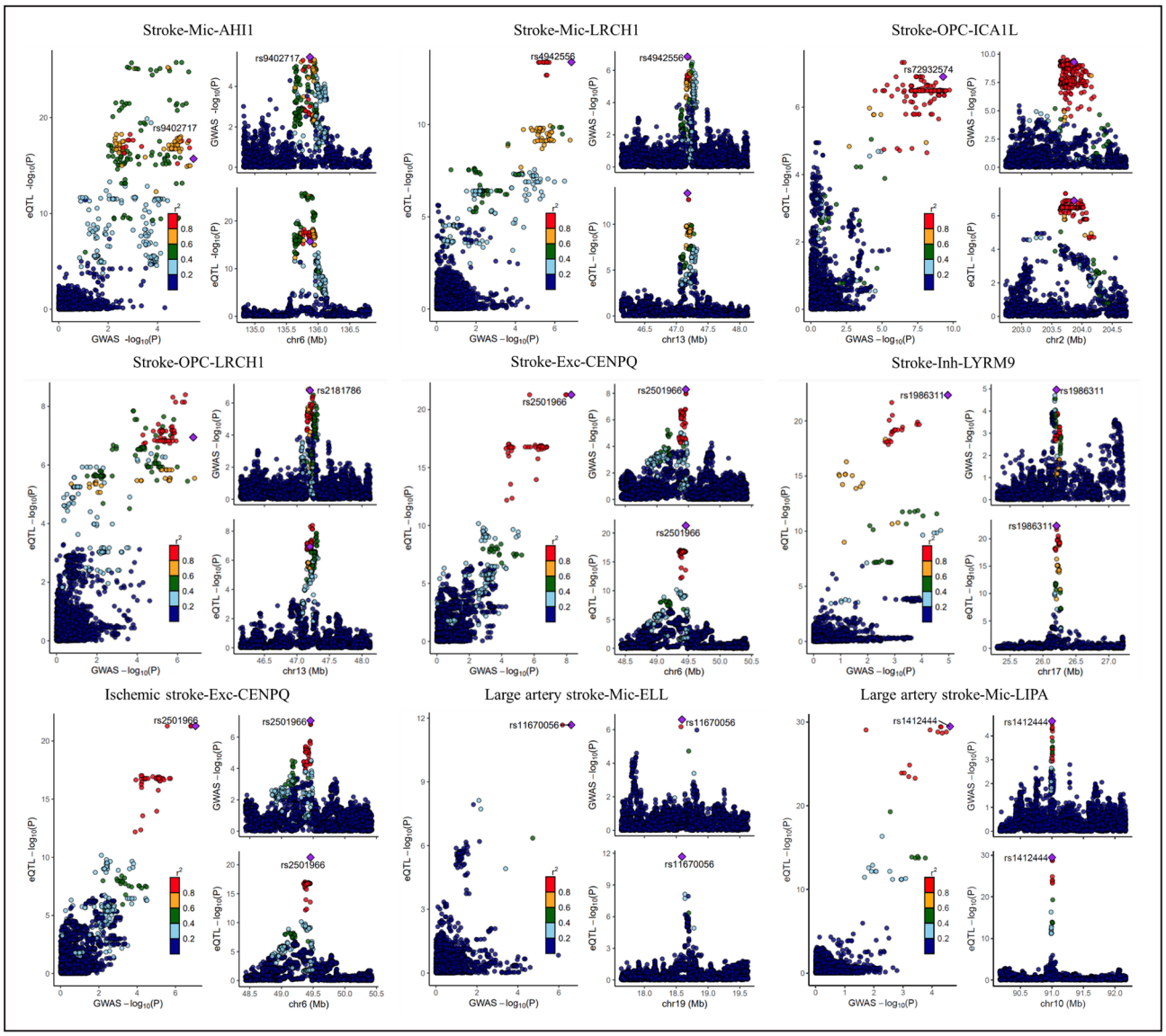


**Figure 6. Regional association plots for colocalized single-cell SMR significant genes.**
For the 7 genes that passed the colocalization analysis, the LocusZoom plot illustrates the GWAS associations for stroke and its subtypes alongside eQTL associations derived from single-cell data at their respective loci. For each validated gene, the **left** shows the $-\log_{10}(P)$ values of GWAS (*x* axis) against eQTL (*y* axis), the **top-right** displays the GWAS locus zoom plot, and the **bottom-right** shows the eQTL locus zoom plot. Variants are color-coded based on their $r^2$ values, with the risk variant distinctly highlighted in purple. Crucially, most GWAS signals shown do not reach genome-wide significance ($P<5\times10^{-8}$), demonstrating how eQTL data enables gene discovery beyond conventional GWAS thresholds. eQTL indicates expression quantitative trait locus; Exc, excitatory neurons; GWAS, genome-wide association studies; Inh, inhibitory neurons; Mic, microglia; OPCs, oligodendrocyte progenitor cells; and SMR, summary-data-based Mendelian randomization.

eQTL data with GWAS signals. Notably, in 7 of the 9 presented loci, the GWAS associations (top-right panels) do not reach genome-wide significance, meaning these genes would be overlooked in conventional GWAS analysis. The eQTL integration provides the necessary precision to identify these potentially causal genes that exist below standard GWAS thresholds. Finally, to assess variant-level concordance, we compared coloc's most probable causal SNVs with the top SMR SNVs. Only 3/9 colocalized SNVs matched the SMR top SNVs exactly, whereas 4 were in strong LD (eg, $r^2>0.8$) with them (Table S7). This partial overlap indicates both methods implicate the same genomic region for a shared causal mechanism but often prioritize different specific variants. This expected discrepancy arises from methodological differences (SMR's effect-size-driven selection versus coloc's Bayesian probability maximization) and local LD complexity. Consequently, although coloc strongly supports a regional shared causal signal, it does not definitively

identify a single shared causal variant nor guarantee the top SMR SNV is causal for both traits.

### Integration of Genetic Evidence, Single-Cell Expression, and Network Analyses Identifies Novel Stroke Risk Genes and Therapeutic Targets

First, 7 risk genes corresponded to 8 top SNVs. Among these 8 top SNVs, only rs1471539 in stroke showed significance at the genome-wide level ($P$=1.04E-08), whereas the other SNVs, including rs1471539 in AIS, did not exhibit genome-wide significance ($P$>5E-8). Therefore, all 7 risk genes were novel at the genome-wide level.

Next, to assess the behavior of our SMR-prioritized genes in a poststroke environment, we performed single-cell analysis. Single-cell analysis revealed that *LIPA* expression was significantly elevated in 2-Day Stroke samples compared with controls (logFC=1.65, false discovery rate [FDR]=4.66E-03). This supports the notion that *LIPA* not only acts as a prestroke risk factor but also may play a role in the pathological processes following a stroke. Additionally, *LIPA* expression was significantly lower in B cells, mural cells, natural killer cells, and T cells compared with other cell types (log2FC=−0.835, FDR=6.55E-11; log2FC=−1.033, FDR=6.50E-03; log2FC=−1.09, FDR=1.17E-12; log2FC=−0.893, FDR=7.96E-24). To visualize the 7 risk genes identified in the study, uniform manifold approximation and projection plots and bubble plots were used to illustrate their expression distribution across different cell types and between stroke and normal samples (Figures S1 through S7).

Subsequently, to comprehensively identify potential therapeutic targets for stroke, we performed a PPI analysis and identified 70 genes interacting with the 7 risk genes. Druggability evaluation of the 77 risk and interacting genes revealed that *DHCR24*, an interacting partner of *LIPA*, is a target of erythropoietic agents; *LIPA*'s interacting partner *APOB* is a target of antihypertensive and antithrombotic agents; *LRCH1*'s interacting partner *CALM1* is a target of antihypertensive agents, antiarrhythmia agents, and immunosuppressants (Table S8).

Lastly, PheWAS analysis of the SMR top SNVs—rs1412444 for *LIPA* and rs842370 and rs1886964 for *LRCH1*—was conducted to explore potential side effects as stroke therapeutic targets. The results indicated significant associations of the top SNVs with myocardial infarction, coronary atherosclerosis, and hammer toe. Elevated *LIPA* expression may increase the risk of myocardial infarction and coronary atherosclerosis, while elevated *LRCH1* expression may increase the risk of hammer toe (Table S9; Figure S8). These findings underscore the importance of evaluating potential adverse effects when considering *LIPA* and *LRCH1* as therapeutic targets for stroke.

## DISCUSSION

Unlike bulk-tissue studies that reported broad stroke associations,[11] our cell-type-resolved approach dissects subtype-specific pathology by mapping causal genes to distinct cellular contexts. For instance, we identified microglia-specific targets for LAS, including *LIPA* (linked to lipid accumulation) and *ELL* (associated with transcriptional dysregulation), as well as an OPC-specific target for CES, *LRCH1* (involved in neuroinflammation modulation), forming "cell-subtype–target" mechanistic triads. Our integration of single-cell eQTL data with stroke-subtype GWAS advances the field by resolving 2 critical limitations of prior bulk-tissue approaches: (1) the conflation of genetically distinct stroke subtypes under a single pathological umbrella,[36] and (2) the masking of cell-type-specific regulatory effects in heterogenous brain tissues.[37] By mapping risk loci to cell-subtype-contextualized regulatory networks, we identified *LRCH1*, *ICA1L*, *AHI1*, *LYRM9*, *CENPQ*, *LIPA*, and *ELL* as novel stroke drivers—genes overlooked in bulk-brain SMR analyses (GTEx brain cortex: *ICA1L* q=0.48; *AHI1* q=0.92; *LYRM9* q=0.31; *CENPQ* q=0.48), *ELL* was not detected in the GTEx brain cortex eQTL file, and the minimum eQTL $P$-values of *LIPA* and *LRCH1* in the GTEx brain cortex eQTL file were >5e to 8, preventing SMR analysis. These 7 single-cell SMR significant genes are associated with Mic, OPCs, Exc, and Inh. These findings align with emerging evidence that microglial lipid metabolism dysregulation is closely linked to poststroke neuroinflammation and blood–brain barrier disruption.[38] Consistently, recent work highlights the dual and context-dependent roles of microglia in ischemic injury and repair.[39] Additionally, oligodendrocytes and OPCs play critical roles in stroke pathology by modulating blood–brain barrier damage and repair through inflammatory signaling.[40] Furthermore, the imbalance between increased excitatory input and decreased inhibitory input following stroke disrupts neural plasticity and impacts recovery, potentially involving Exc and Inh.[41]

Our gene enrichment analysis reveals distinct functional enrichments across stroke subtypes and cell types, highlighting the complexity of stroke pathology. In AIS, Exc are enriched in signal transduction and regulatory pathways, supporting the notion that excessive excitatory neuronal activation post stroke can lead to excitotoxicity and exacerbate neuronal damage.[42] In LAS, oligodendrocytes show enrichment in neural development pathways, consistent with their role in poststroke myelin regeneration.[43] Additionally, the enrichment of cell growth-related pathways in

oligodendrocytes during stroke suggests their critical involvement in neural repair during stroke recovery.[44]

Our downstream analyses, primarily focused on drug target evaluation, highlight the potential therapeutic relevance of *LIPA* and *LRCH1* in stroke while underscoring possible cardiovascular risks. Supporting this potential, data from the animal model indicate that *LIPA* undergoes expression changes in inflammatory cells following stroke, suggesting it resides on a continuous pathological axis: predisposing individuals to the disease at baseline and actively participating in the disease process after its onset, which enhances its attractiveness as a therapeutic target. Mechanistically, *LIPA*'s interaction with *APOB* suggests a link to lipid metabolism and vascular health, aligning with previous research on *APOB*-related lipid dysregulation in atherosclerosis.[45] Similarly, *LRCH1*'s interaction with *CALM1*, a target of antihypertensive and antiarrhythmic agents, supports its role in vascular integrity and inflammation modulation.[46] However, PheWAS analysis indicates that elevated *LIPA* expression may increase myocardial infarction and coronary atherosclerosis risk,[47] and *LRCH1* upregulation may elevate hammer toe susceptibility, suggesting a need for cautious therapeutic targeting. These PheWAS-documented risks demand a careful benefit–risk assessment. For *LIPA*, elevated expression is also associated with LAS risk, mirroring known lipid-targeting therapies such as statins, which increase diabetes risk but remain the first-line treatment for cardiovascular disease. Therefore, *LIPA* inhibition should be considered primarily in patients with recurrent LAS who do not have coronary disease and ideally coadministered with lipid-monitoring regimens to mitigate potential atherosclerotic side effects. In contrast, *LRCH1*'s association with hammer toe may reflect shared neurovascular developmental pathways rather than direct toxicity. Given its strong OPC-specific effect on CES, the minor orthopedic risk appears clinically acceptable. Future studies should explore how modulation of *LIPA* and *LRCH1* influences stroke recovery and comorbidity risk, with precision medicine strategies tailored to maximize benefit while minimizing unintended adverse effects.

Several limitations warrant consideration. First, our single-cell eQTL data originate exclusively from the neocortex. Although this design effectively captures cortical stroke mechanisms, its applicability to SVS may be limited given the predominant subcortical involvement of SVS pathology, particularly in regions such as basal ganglia, and other deep brain structures.[48] Future studies integrating subcortical single-cell eQTL resources will be essential to fully elucidate SVS genetics. Second, our Europe-centric cohort restricts generalizability; the absence of non-European stroke risk loci such as *FGF5* (East Asian specific) underscores the need for diverse cohorts.[16] Third, although MR analysis provides putative causal evidence, residual pleiotropy may persist despite HEIDI and MR-Egger adjustments. Future experimental validation (eg, CRISPR editing in disease models) is needed to confirm these associations. Such validation could leverage rodent ischemic stroke models, which remain indispensable for mechanistic and therapeutic translation.[49]

In conclusion, this study establishes a blueprint for cell-type-resolved stroke genetics. We link risk variants to their target cell types, moving beyond location-based inference to understanding their function in specific cellular contexts, as exemplified by *LIPA*'s microglial specificity. These insights not only refine stroke's molecular taxonomy but also illuminate cell-targeted therapeutic opportunities, bridging the gap between genetic association and actionable pathophysiology.

## CONCLUSIONS

Our integrative analysis of single-cell eQTL data and stroke-subtype GWAS identified novel potential causal genes (*LRCH1*, *ICA1L*, *AHI1*, *LYRM9*, *CENPQ*, *LIPA*, *ELL*) with subtype- and cell-type-specific effects. PPI and druggability analyses further supported *LIPA* and *LRCH1* as promising therapeutic targets for stroke. These findings underscore the power of integrating single-cell transcriptomic and stroke-subtype genomic data to elucidate stroke mechanisms and identify actionable targets, enhancing our mechanistic understanding of stroke pathogenesis.

## ARTICLE INFORMATION

Received September 11, 2025; accepted February 4, 2026.

### Affiliations

Laboratory of Brain Disorders, Collaborative Innovation Center for Brain Disorders, Beijing Institute of Brain Disorders, Ministry of Science and Technology, National Engineering Center of Internet Medical Diagnosis and Treatment Technology, Xuanwu Hospital, Laboratory of Cerebral Small Vessel Disease and Osteoporosis (Abnormal Bone Metabolism), Capital Medical University, Beijing, China (Y.H., P.Z., S.G., S.W., F.L., G.L.); Department of Intensive Care Unit, Tianjin Key Laboratory of Cerebral Vascular and Neurodegenerative Diseases, Tianjin Huanhu Hospital, Tianjin, China (T.Z.); Department of Epidemiology and Biostatistics, School of Public Health, Wannan Medical College, Wuhu, Anhui, China (Y.C., Y.Y., G.L.); Dongying Branch Center of Collaborative Innovation Center for Brain Disorders, Shengli Oilfield Central Hospital, Dongying, Shandong, China (G.L.); and Clinical Medicine Translational Research Institute, Chengdu Fifth People's Hospital, Geriatric Diseases Institute of Chengdu, The Second Clinical Medical College, Affiliated Fifth People's Hospital of Chengdu University of Traditional Chinese MedicineChengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China (G.L.).

### Acknowledgments

We acknowledge the provision of the stroke and stroke subtype GWAS data, the single-cell eQTL data, and the bulk tissue eQTL data from the GTEx consortium. Additionally, we sincerely appreciate the individuals who generously contributed samples to these studies. Guiyou Liu conceived and initiated the project. Guiyou Liu and Yijie He conducted data analysis and drafted the article. All authors contributed to result interpretation, critically revised the article for intellectual content, and approved the final version.

## Sources of Funding

This work was supported by funding from the National Key Research and Development Program of China (Grant No. 2023YFC3605200), Beijing Natural Science Foundation (Grant No. JQ21022), National Natural Science Foundation of China (Grant No. 82071212 and 82471449), and Anhui Province Youth Teacher Training Action Discipline (Professional) Leader Cultivation Project (Grant No. DTR2024031).

## Disclosures

None.

## Supplemental Material

Figures S1–S8
Tables S1–S9
Checklist

## REFERENCES

1. Johnson CO, Nguyen M, Roth GA, Nichols E, Alam T, Abate D, Abd-Allah F, Abdelalim A, Abraha HN, Abu-Rmeileh NME, et al. Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease study 2016. *Lancet Neurol*. 2019;18:439–458. doi: 10.1016/S1474-4422(19)30034-1
2. Feigin VL, Brainin M, Norrving B, Martins S, Sacco RL, Hacke W, Fisher M, Pandian J, Lindsay P. World Stroke Organization (WSO): global stroke fact sheet 2022. *Int J Stroke*. 2022;17:18–29. doi: 10.1177/17474930211065917
3. Gheorghe A, Griffiths U, Murphy A, Legido-Quigley H, Lamptey P, Perel P. The economic burden of cardiovascular disease and hypertension in low- and middle-income countries: a systematic review. *BMC Public Health*. 2018;18:975. doi: 10.1186/s12889-018-5806-x
4. Rajsic S, Gothe H, Borba HH, Sroczynski G, Vujicic J, Toell T, Siebert U. Economic burden of stroke: a systematic review on post-stroke care. *Eur J Health Econ*. 2019;20:107–134. doi: 10.1007/s10198-018-0984-0
5. Rahman AA, Amruta N, Pinteaux E, Bix GJ. Neurogenesis after stroke: a therapeutic perspective. *Transl Stroke Res*. 2021;12:1–14. doi: 10.1007/s12975-020-00841-w
6. Jung J, Lu Z, de Smith A, Mancuso N. Novel insight into the etiology of ischemic stroke gained by integrative multiome-wide association study. *Hum Mol Genet*. 2024;33:170–181. doi: 10.1093/hmg/ddad174
7. Islam T, Rahman MR, Khan A, Moni MA. Integration of Mendelian randomisation and systems biology models to identify novel blood-based biomarkers for stroke. *J Biomed Inform*. 2023;141:104345. doi: 10.1016/j.jbi.2023.104345
8. Yang L-Z, Yang Y, Hong C, Wu Q-Z, Shi X-J, Liu Y-L, Chen g-z. Systematic Mendelian randomization exploring druggable genes for hemorrhagic strokes. *Mol Neurobiol*. 2025;62:1359–1372. doi: 10.1007/s12035-024-04336-9
9. Song Z, Han Y, Li W, Xu Y, He Y, Wang Y. A cross-tissue transcriptome-wide association study identifies new key genes in ischemic stroke. *Gene*. 2025;941:149207. doi: 10.1016/j.gene.2024.149207
10. Muiño E, Carcel-Marquez J, Llucià-Carol L, Gallego-Fabrega C, Cullell N, Lledós M, Martín-Campos JM, Villatoro-González P, Sierra-Marcos A, Ros-Castelló V, et al. Identification of genetic loci associated with intracerebral hemorrhage using a multitrait analysis approach. *Neurology*. 2024;103:e209666. doi: 10.1212/WNL.0000000000209666
11. Zhang LY, Chu YH, You YF, Dong MH, Pang XW, Chen L, Zhu LF, Yang S, Zhou LQ, Shang K, et al. Systematic druggable genome-wide Mendelian randomization identifies therapeutic targets for functional outcome after ischemic stroke. *J Am Heart Assoc*. 2024;13:e034749. doi: 10.1161/JAHA.124.034749
12. Bakreen A, Jolkkonen J. The potential of a combined cell-based therapy and rehabilitation approach for stroke recovery. *Neuroprotection*. 2023;1:117–129. doi: 10.1002/nep3.27
13. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–537. doi: 10.1038/s41588-018-0058-3
14. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet*. 2013;9:e1003649. doi: 10.1371/journal.pgen.1003649
15. Raj T, Rothamel K, Mostafavi S, Ye C, Lee MN, Replogle JM, Feng T, Lee M, Asinovski N, Frohlich I, et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science*. 2014;344:519–523. doi: 10.1126/science.1249547
16. Mishra A, Malik R, Hachiya T, Jürgenson T, Namba S, Posner DC, Kamanu FK, Koido M, Le Grand Q, Shi M, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611:115–123. doi: 10.1038/s41586-022-05165-3
17. Fujita M, Gao Z, Zeng L, McCabe C, White CC, Ng B, Green GS, Rozenblatt-Rosen O, Phillips D, Amir-Zilberstein L, et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet*. 2024;56:605–614. doi: 10.1038/s41588-024-01685-y
18. Consortium TG, Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, Hadley K, Handsaker RE, Huang KH, et al. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369:1318–1330. doi: 10.1126/science.aaz1776
19. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48:481–487. doi: 10.1038/ng.3538
20. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, et al. Ensembl 2020. *Nucleic Acids Res*. 2019;48:D682–D688. doi: 10.1093/nar/gkz966
21. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. A global reference for human genetic variation. *Nature*. 2015;526:68–74. doi: 10.1038/nature15393
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol*. 1995;57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
23. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523. doi: 10.1038/s41467-019-09234-6
24. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2008;37:1–13. doi: 10.1093/nar/gkn923
25. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408. doi: 10.7554/eLife.34408
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795
27. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37:658–665. doi: 10.1002/gepi.21758
28. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet*. 2018;27:R195–R208. doi: 10.1093/hmg/ddy163
29. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44:512–525. doi: 10.1093/ije/dyv080
30. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10:e1004383. doi: 10.1371/journal.pgen.1004383
31. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet*. 2019;51:768–769. doi: 10.1038/s41588-019-0404-0
32. Garcia-Bonilla L, Shahanoor Z, Sciortino R, Nazarzoda O, Racchumi G, Iadecola C, Anrather J. Analysis of brain and blood single-cell transcriptomics in acute and subacute phases after experimental stroke. *Nat Immunol*. 2024;25:357–370. doi: 10.1038/s41590-023-01711-x
33. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2012;41:D808–D815. doi: 10.1093/nar/gks1094
34. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res*. 2018;46:D1068–D1073. doi: 10.1093/nar/gkx1143
35. Gagliano Taliun SA, VandeHaar P, Boughton AP, Welch RP, Taliun D, Schmidt EM, Zhou W, Nielsen JB, Willer CJ, Lee S, et al. Exploring

and visualizing large-scale genetic associations by using PheWeb. *Nat Genet*. 2020;52:550–552. doi: 10.1038/s41588-020-0622-5
36. Traylor M, Anderson CD, Rutten-Jacobs LCA, Falcone GJ, Comeau ME, Ay H, Sudlow CLM, Xu H, Mitchell BD, Cole JW, et al. Subtype specificity of genetic loci associated with stroke in 16 664 cases and 32 792 controls. *Circ Genom Precis Med*. 2019;12:e002338. doi: 10.1161/CIRCGEN.118.002338
37. Zheng K, Lin L, Jiang W, Chen L, Zhang X, Zhang Q, Ren Y, Hao J. Single-cell RNA-seq reveals the transcriptional landscape in ischemic stroke. *J Cereb Blood Flow Metab*. 2022;42:56–73. doi: 10.1177/0271678X211026770
38. Li H, Liu P, Deng S, Zhu L, Cao X, Bao X, Xia S, Xu Y, Zhang B. Pharmacological upregulation of microglial lipid droplet alleviates neuroinflammation and acute ischemic brain injury. *Inflammation*. 2023;46:1832–1848. doi: 10.1007/s10753-023-01844-z
39. Haupt M, Gerner ST, Doeppner TR. The dual role of microglia in ischemic stroke and its modulation via extracellular vesicles and stem cells. *Neuroprotection*. 2024;2:4–15. doi: 10.1002/nep3.39
40. Wang Y, Liu W, Geng P, Du W, Guo C, Wang Q, Zheng G-q, Jin X. Role of crosstalk between glial cells and immune cells in blood-brain barrier damage and protection after acute ischemic stroke. *Aging Dis*. 2023;15:2507. doi: 10.14336/AD.2023.1010
41. Carmichael ST. Brain excitability in stroke: the yin and yang of stroke progression. *Arch Neurol*. 2012;69:161–167. doi: 10.1001/archneurol.2011.1175
42. Wang F, Xie X, Xing X, Sun X. Excitatory synaptic transmission in ischemic stroke: a new outlet for classical neuroprotective strategies. *Int J Mol Sci*. 2022;23:9381. doi: 10.3390/ijms23169381
43. Li C, Li W, Wei W, Chen Q, Gao H, Zhao Y, Zhang L, Ling L, Shen H, Shen Y. Gene expression profiles of endothelium, microglia and oligodendrocytes in hippocampus of post-stroke depression rat at single cell resolution. *Mol Psychiatry*. 2025;30:1995–2008. doi: 10.1038/s41380-024-02810-3
44. Li W, He T, Shi R, Song Y, Wang L, Zhang Z, Tang Y, Yang GY, Wang Y. Oligodendrocyte precursor cells transplantation improves stroke recovery via oligodendrogenesis, neurite growth and synaptogenesis. *Aging Dis*. 2021;12:2096–2112. doi: 10.14336/ad.2021.0416
45. Morita S-y. Metabolism and modification of apolipoprotein B-containing lipoproteins involved in dyslipidemia and atherosclerosis. *Biol Pharm Bull*. 2016;39:1–24. doi: 10.1248/bpb.b15-00716
46. Liu T-t, Xu H-h, Liu Z-j, Zhang H-p, Zhou H-t, Zhu Z-x, Wang Z-q, Xue J-y, Li Q, Ma Y, et al. Downregulated calmodulin expression contributes to endothelial cell impairment in diabetes. *Acta Pharmacol Sin*. 2023;44:2492–2503. doi: 10.1038/s41401-023-01127-1
47. Wild PS, Zeller T, Schillert A, Szymczak S, Sinning CR, Deiseroth A, Schnabel RB, Lubos E, Keller T, Eleftheriadis MS, et al. A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circ Cardiovasc Genet*. 2011;4:403–412. doi: 10.1161/CIRCGENETICS.110.958728
48. Del Bene A, Makin SDJ, Doubal FN, Inzitari D, Wardlaw JM. Variation in risk factors for recent small subcortical infarcts with infarct size, shape, and location. *Stroke*. 2013;44:3000–3006. doi: 10.1161/STROKEAHA.113.002227
49. Thaysen M, Westi E, Clarkson AN, Wellendorph P, Kristensen M. Rodent ischemic stroke models and their relevance in preclinical research. *Neuroprotection*. 2024;2:296–309. doi: 10.1002/nep3.62