

Causal Genetic Mechanisms of Hyperhomocysteinemia in Ischemic Stroke: Integrating Mendelian Randomization, Single-Cell Transcriptomics, and Machine Learning

Qing-hua Huang

Ningxia Medical University General Hospital

Nan Zhang

Ningxia Hui Autonomous Region Peoples Hospital

Jing Wang

Ningxia Medical University

Feng-chen Bi

Ningxia Medical University

Xi-bo Yang

Ningxia Medical University

Bin Wu

Xi'an Jiaotong University

Yan Li

Ningxia Medical University

Li-jun Ma

MLJ2026@hotmail.com

Ningxia Medical University

Research Article

Keywords: Ischemic stroke, Homocysteine, Mendelian randomization, Single-cell RNA sequencing

Posted Date: September 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10770596/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations:

No competing interests reported.

Figures are available in the Supplementary Files section.

Causal Genetic Mechanisms of Hyperhomocysteinemia in Ischemic Stroke: Integrating Mendelian Randomization, Single-Cell Transcriptomics, and Machine Learning

Qing-hua Huang¹, Nan Zhang², Jing Wang³, Feng-chen Bi³, Xi-bo Yang³, Bin Wu⁴, Yan Li^{3*}, Li-jun Ma^{3*}

1. Emergency Department, General Hospital of Ningxia Medical University(The First Clinical Medical College of Ningxia Medical University), Yinchuan 750004, China

2. Neurosurgery Department, Ningxia Hui Autonomous Region People's Hospital, Yinchuan 750011, China

3. Department of Human Anatomy, Histology and Embryology, School of Basic Medicine, Ningxia Medical University, Yinchuan 750004, China

4. Department of Clinical Medical College of Xi'an Jiao tong University, Xi'an, 710061, China

* Correspondence:

Corresponding author: Yan Li

E-mail addresses:20090015@nxmu.edu.cn

*Corresponding Author : Lijun Ma

E-mail addresses:MLJ2026@hotmail.com

Abstract

Background: Hyperhomocysteinemia has been widely linked to the development of cerebrovascular diseases, yet the cell type-specific genetic mechanisms remain elusive. This study aimed to identify hyperhomocysteinemia-driven genes and their functional roles in specific cell populations underlying ischemic stroke (IS) pathogenesis.

Methods: Data were obtained from publicly available databases. First, Mendelian randomization (MR) analysis was performed to investigate the causal relationship between homocysteine (Hcy) and IS. Next, a transcriptome-wide association study

(TWAS) was conducted to identify trait-related genes in IS. Key cell types were then determined using single-cell RNA sequencing (scRNA-seq) analysis. Feature genes were subsequently identified through a combination of machine learning (ML) and key gene was identified by expression analysis. Finally, the functions of these genes were explored using pseudotime analysis and receiver operating characteristic (ROC) analysis to elucidate potential molecular mechanisms underlying IS. Finally, Western blot analysis and qPCR were performed.

Results: MR analysis revealed that elevated Hcy is associated with an increased risk of IS (OR = 1.06, 95% Confidence Interval [CI] = 1.02-1.12, P = 0.0098). Subsequently, TWAS identified 103 trait-related genes enriched in chemokine signaling. scRNA-seq revealed macrophages as the key Hcy-active cell type, with HHcy (high HcyScore) macrophages accounting for 51.1% in IS versus 36.1% in controls. Gene set variation analysis (GSVA) indicated that HHcy macrophages were enriched in pathways related to oxidative phosphorylation. ML identified Gusb, Mrpl36, and Rab2a as feature genes, with each exhibiting correlations with HcyScore and favorable diagnostic potential (AUC > 0.6). Experimental verification confirmed significant expression differences in Mrpl36 and Rab2a between Hypoxia and control samples.

Conclusion: Macrophages were identified as the key cell type, and Gusb, Rab2a, and Mrpl36 emerged as critical genes, providing a valuable framework linking genetic susceptibility to macrophage-mediated neuroinflammation in IS.

Keywords: Ischemic stroke, Homocysteine, Mendelian randomization, Single-cell RNA sequencing

1. Introduction

Stroke, commonly known as a cerebrovascular accident (CVA), is an acute cerebrovascular disease that primarily results from the sudden rupture or blockage of blood vessels in the brain, leading to a lack of blood flow and subsequent damage to brain tissue[1]. Among the different types of stroke, ischemic stroke (IS) is the most common, accounting for approximately 87% of all stroke cases[2, 3]. Following the onset of IS, a series of complex pathophysiological changes occur rapidly, including intracellular and extracellular ion imbalance, neuroinflammatory responses, and abnormal activation of immune cells. These changes may ultimately lead to neuronal death, impairing brain function[4]. Currently, the key strategy for reducing brain injury after stroke remains the rapid restoration of cerebral blood flow[5]. Intravenous thrombolysis and mechanical thrombectomy are the two primary treatment modalities for IS[6, 7]. However, the high cost of treatment and the potential for complications still pose significant challenges for the management of IS. Given the current limitations of IS treatment, there is an urgent need to explore new therapeutic strategies.

Homocysteine (Hcy) is a sulfur-containing amino acid and a key intermediate in the metabolism of methionine[8]. Elevated serum Hcy levels can be caused by various factors, including congenital enzyme deficiencies, chronic liver and kidney dysfunction, and specific drug interventions. Hyperhomocysteinemia (HHcy), as an independent risk factor, is closely associated with the development of hypertension, myocardial infarction, peripheral arterial obstructive disease, venous thrombosis, and both hemorrhagic and ischemic strokes[9]. Mechanistically, Hcy and its related

75 metabolic products can regulate epigenetic mechanisms, disrupt basic cellular
76 processes such as the mTOR signaling pathway and autophagy, and induce
77 endothelial dysfunction, thereby promoting the progression of cardiovascular diseases
78 and stroke[10]. Clinically, folic acid supplementation is the primary strategy to lower
79 Hcy levels. Recent meta-analyses of randomized controlled trials suggest that
80 Hcy-lowering interventions can reduce the risk of first-time stroke by 10%[11, 12].
81 Although existing studies have preliminarily revealed the association between Hcy
82 and IS, the exact pathophysiological mechanisms remain incompletely understood.
83 Therefore, a deeper exploration of the complex molecular relationship between Hcy
84 and IS is of significant theoretical and clinical importance for the development of
85 more accurate diagnostic biomarkers and therapeutic targets.

86 Single-cell RNA sequencing (scRNA-seq) technology enables the analysis of
87 gene expression at the single-cell level, revealing the heterogeneity of cell types and
88 cellular states[13, 14]. In IS research, scRNA-seq has effectively unveiled the
89 complex cellular heterogeneity associated with IS, providing a crucial theoretical
90 foundation for a deeper understanding of its molecular mechanisms[15, 16]. Our
91 study utilized data from the Gene Expression Omnibus (GEO,
92 <https://www.ncbi.nlm.nih.gov/geo/>) and the Integrative Epidemiology Unit (IEU)
93 Open GWAS database (<https://gwas.mrcieu.ac.uk/>), conducting a series of
94 bioinformatics analyses to identify key genes associated with Hcy in IS. Specifically,
95 we performed Mendelian randomization (MR) analysis to investigate the causal
96 relationship between Hcy and IS. Following this, we identified candidate genes using

transcriptome-wide association study (TWAS) screening, single-cell analysis, and machine learning approaches, further narrowing down key genes based on their expression in critical cell types. In summary, our study provides novel insights into the molecular mechanisms underlying IS and offers valuable guidance for the development of personalized therapeutic strategies.

2. Materials and methods

2.1 Data collection

In this study, transcriptomic and scRNA-seq data related to IS were obtained from the GEO database. Specifically, the GSE32529 dataset (GPL1261 platform) included transcriptomic data from 58 IS and 54 control mouse brain tissue samples, while the GSE16561 dataset (GPL6883 platform) included whole blood samples from 24 patients with IS and 39 healthy controls. Additionally, a scRNA-seq dataset, GSE174574 (GPL21103 platform), was retrieved from GEO, comprising 3 IS and 3 control mouse brain tissue samples.

For MR analysis, data were obtained from the IEU Open GWAS database. Homocysteine was considered as the exposure factor, using dataset GCST009171. For IS as the outcome, dataset GCST006908 was employed, which included 4,881 European individuals (34,217 cases and 406,111 controls) and a total of 8,296,492 SNPs.

2.2 Data pre-processing for MR study

In the MR study, homocysteine was considered as exposure factor, with IS as the outcome. The classical MR analysis confirmed that three key assumptions were satisfied: (i) the independence assumption [instrumental variables (IVs) were not

associated with any confounders]; (ii) the association assumption (IVs directly affected the exposure); and (iii) the exclusivity assumption (IVs could only influence the outcome through the exposure factors and not via other pathways).

First, the "TwoSampleMR" package (v 0.5.6)[17] was used to read the data for exposure factor and filter IVs ($P < 1 \times 10^{-6}$). Next, SNPs in linkage disequilibrium (LD) were removed (`clump = TRUE`, $r^2 < 0.001$, $kb = 10,000$). SNPs associated with the outcome were excluded, while those associated with the exposure factors were retained. The effect alleles and effect sizes were then harmonized using the `harmonise_data` function from the "TwoSampleMR" package (v 0.5.6). Finally, the strength of each SNP was assessed using F-statistics, and those with an F-value below 10 were excluded. Only SNPs with three or more associations were retained in the final analysis.

2.3 MR study, sensitivity analysis, and Steiger test

After screening the IVs, `mr` function from the "TwoSampleMR" package (v 0.5.6) combined with five algorithms were employed for the MR study: Inverse Variance Weighted (IVW), MR Egger, Weighted median, Simple mode, and Weighted mode. Among these methods, IVW was considered the most critical. The screening criterion for the MR study was set at $P_{IVW} < 0.05$. Additionally, the odds ratio (OR) was calculated, with $OR > 1$ indicating a risk factor and $OR < 1$ indicating a protective factor. Next, scatter plot, forest plot, and funnel plot were created.

A series of sensitivity analyses were performed to ensure the robustness of the MR results. The heterogeneity test was conducted using the `mr_heterogeneity` function

based on Cochran's Q test. If $P > 0.05$, the fixed IVW method was applied; if $P < 0.05$, the random IVW method was used. Furthermore, a horizontal pleiotropy test ($P > 0.05$) was conducted using the MR-Egger and MR-PRESSO. A leave-one-out (LOO) analysis was also performed using the `mr_leaveoneout` function. Finally, the Steiger test was used to assess the directionality of the causal relationship, employing the `directionality_test` function from the "TwoSampleMR" package (v 0.5.6). The criteria for passing the Steiger test were a correct causal direction (TRUE) and $P < 0.05$.

2.4 The scRNA-seq analysis and identification of key cell types

The scRNA-seq dataset GSE174574 was analyzed using the "Seurat" package (v 5.0.1)[18]. Initially, quality control (QC) was performed to filter out low-quality cells and genes. Cells with fewer than 200 detected genes and genes expressed in fewer than 3 cells were excluded. Cells were retained if they met the following criteria: $200 < \text{nFeature_RNA (genes per cell)} < 7,000$, $\text{percent.mt (proportion of mitochondrial gene expression)} < 20\%$. After QC, the data were normalized using the `NormalizeData` function. Highly variable genes (HVGs) were identified using the `FindVariableFeatures` function. Principal component analysis (PCA) was performed on the HVGs with the `RunPCA` function. The number of principal components (PCs) to retain was determined using the `JackStrawPlot` function ($P < 0.05$). Unsupervised clustering analysis was performed using the `FindNeighbors` and `FindClusters` functions ($\text{resolution} = 0.60$). Cell annotation was conducted using three criteria: (1) significantly upregulated genes compared to other clusters, (2) unique gene expression patterns distinguishing the cell population, and (3) established canonical

markers reported in previous literature.

2.5 Whole transcriptome association study

Transcriptome-wide association study (TWAS) is an analytical approach that integrates GWAS and gene expression data to identify gene expression changes associated with specific phenotypes, including diseases or traits. This method can help reveal which gene expression levels are statistically correlated with a given disease or trait[19]]. In this study, the "MTWAS" package[20] was employed to perform TWAS analysis on homocysteine and to identify trait-associated genes ($P < 0.01$). Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to investigate the biological functions and pathways linked to these trait-associated genes, using the "clusterProfiler" package (v 4.6.2)[21] with a significance threshold of $P < 0.05$.

2.6 Evaluation of homocysteine activity

In the scRNA-seq dataset GSE174574, the HcyScore for each annotated cell type was calculated based on trait-associated genes using the AUCell, Ucell, scSE, singscore, and AddModuleScore algorithms from the "irGSEA" package (v 3.2.7)[22]. Differences in HcyScore between IS and control samples were assessed ($P < 0.05$). Cells exhibiting the highest HcyScores and showing significant differences between IS and control samples were identified as key cells. Based on HcyScore, key cells with scores below the median were classified as the low-homocysteine activity state (LHcy) group, whereas those with scores above the median were assigned to the high-homocysteine activity state (HHcy) group. The proportions of LHcy and HHcy cells in IS and control samples were compared ($P < 0.05$).

Subsequently, Slingshot and cytoTRACE analyses were performed to calculate cytoTRACE scores, and differences in cytoTRACE scores between LHcy and HHcy groups were evaluated ($P < 0.05$).

2.7 Cell-to-cell communication and Gene Set Variation Analysis (GSVA)

To gain a deeper understanding of the key cell subtypes (LHcy and HHcy cells), cell-to-cell communication analysis was performed on IS samples using the CellChat package (v 1.6.1)[23]. Communication networks were constructed and visualized to illustrate both the number and strength of interactions between key cell subtypes and other annotated cell types. Receptor-ligand pairs mediating these intercellular interactions were also analyzed using the same package.

To investigate functional differences between LHcy and HHcy cells, GSVA was conducted ($P < 0.05$). Furthermore, scMetabolism was employed to examine differences in metabolic pathways between these two key cell subtypes.

2.8 Machine learning

In our study, feature genes were identified based on trait-associated genes using five machine learning algorithms: Least Absolute Shrinkage and Selection Operator (LASSO), Boruta (Confirmed), XGBoost, Random Forest (RF), and Gradient Boosting Machine (GBM). Model validation was performed using random sampling with bootstrap ($n=2000$). Genes that were consistently identified across all five algorithms were considered as the final feature genes.

2.9 Performance evaluation of feature genes

The expression of feature genes was analyzed in the scRNA-seq dataset GSE174574, Wilcoxon test was utilized to compare the expression of feature genes in each annotated cells type between IS and control samples ($P < 0.05$). Receiver operating

characteristic (ROC) curves were plotted to explore the diagnosis performance of feature genes, and the area under the curve (AUC) scores for each gene were calculated ($AUC > 0.6$). Meanwhile, to validate the cross-species and cross-tissue stability of the feature genes, the ROC curves of feature genes were also plotted in GSE32529 and GSE6561 datasets.

2.10 Analysis of feature genes in cell level

To investigate the expression of feature genes in key cellular differentiation processes, pseudotime analysis was conducted using the Monocle2 package[24]. Based on the expression of each feature gene, key cells were categorized into two groups: feature gene⁺ (high expression) and feature gene⁻ (no expression). To further explore the mechanisms by which these three feature genes regulate key cells, trajectory analysis was performed using CytoTRACE to assess the transcriptional heterogeneity of key cells.

The differences in CytoTRACE scores between feature gene⁺ and feature gene⁻ cells were analyzed, and feature genes showing significant differences in CytoTRACE scores ($P < 0.05$) were identified as key genes. Cell-to-cell communication analysis was then conducted to examine the interactions between cell subtypes (key gene⁺ and key gene⁻ cells) and other cell types. Finally, GO and KEGG analyses were performed to explore the functional differences between feature gene⁺ and feature gene⁻ cells ($P < 0.05$).

2.11 Establishment of Oxygen-Glucose Deprivation (OGD) Cell Model

BV2 mouse microglial cells were selected to establish the oxygen-glucose deprivation

injury model. Cells in the logarithmic growth phase were cultured to 70%–80% confluence and randomly divided into the normal control group and oxygen-glucose deprivation (OGD) model group. The original complete medium of cells in the model group was discarded and replaced with glucose-free DMEM basal medium. The cell culture dishes were placed into sterile sealed bags, together with Mitsubishi Chemical carbon dioxide anaerobic gas-generating packs. The bags were quickly sealed to establish a closed hypercapnic and hypoxic microenvironment, maintaining a carbon dioxide concentration higher than 95% and an oxygen concentration lower than 5% in the sealed system. All sealed samples were incubated in a constant temperature incubator at 37°C for 8 hours. Cell samples were collected immediately after incubation for subsequent experimental detection. Cells in the normal control group were cultured in conventional glucose-containing complete DMEM medium and incubated synchronously for 8 hours under standard culture conditions with consistent other experimental parameters.

2.12 Western blot analysis

Cells were lysed using RIPA buffer, and total protein was recovered and quantified by the BCA assay. The quantified protein was then mixed with SDS-PAGE loading dye and heated at 100°C for 10 minutes to induce denaturation. Following this, proteins were separated using SDS-PAGE at a concentration of 100 g/L, and the separated proteins were transferred onto a PVDF membrane (pore size: 0.45µm) via electrotransfer. After the transfer, the PVDF membrane was incubated in blocking buffer (5% non-fat milk in TBST) for 1 hour at room temperature on a shaker at 60 rpm to minimize non-specific antibody binding. The membrane was washed with TBST, and then incubated overnight at 4°C with primary antibodies targeting specific

genes at a dilution of 1:1000. After washing the membrane three times with TBST, it was incubated with HRP-conjugated goat anti-mouse IgG at a dilution of 1:5000 for 2 hours at room temperature. The membranes were then washed three times with TBST again, exposed to enhanced chemiluminescence reagent for development, and protein levels were analyzed using ImageJ software.

2.13 Real-time quantitative polymerase chain reaction

Real-time quantitative polymerase chain reaction (RT-qPCR) analysis of the transcript levels of ACTB, Gusb, Mrpl36 and Rab2a. Total RNA was isolated using TRIzol reagent and reverse-transcribed into complementary DNA with the PrimeScript™ RT reagent Kit (TaKaRa Bio, Inc., Shiga, Japan) in accordance with the manufacturer's instructions. Subsequent amplifications were carried out with SYBR™ Green Universal Master Mix (Thermo Fisher Scientific, Waltham, MA USA). The reaction parameters included an initial denaturation step at 95°C for 5 min and amplification at 95°C for 10 s, followed by 40 cycles at 60°C for 30 s. The mRNA levels Gusb, Mrpl36 and Rab2a were calculated using the $2^{-\Delta\Delta C_t}$ method against GAPDH as a housekeeping gene. The following primers (forward/reverse) were employed for RT-qPCR analysis:

ACTB(GTGCTATGTTGCTCTAGACTTCG/ATGCCACAGGATTCCATACC);Gusb(GCCGCTGCTGGTAAAGGATTTC/CTAGGTGGTGCCGAAGTGACTC);Mrpl36(GAGTCTTCTCCTCGTTCCTGCTG/CCCACGCCTCTTCACCTTG TAG);Rab2a(AAGTACATCATCATCGGCGACAC/ACTCCTGCCCTGCTGTATCCC).

2.14 Statistical analysis

Statistical analysis was performed using R (v 4.3.1). For comparisons between two groups, the Wilcoxon test and student t test were employed ($P < 0.05$). Specifically, the levels of significance were denoted as follows: **** for $p < 0.0001$, *** for $p <$

0.001, ** for $p < 0.01$, * for $p < 0.05$, and ns for $p > 0.05$ (indicating no significant differences).

3. Results

3.1 The homocysteine had significant causal relationship with IS

The MR analysis revealed that homocysteine is associated with an increased risk of IS (OR = 1.06, 95% Confidence Interval [CI] = 1.02-1.12, $P = 0.0098$) (**Fig. 1A**). The results from five different methods are shown in **Supplementary Fig. S1A**.

The heterogeneity test indicated that the P-value for homocysteine was greater than 0.05, suggesting no significant heterogeneity (**Supplementary Table S1**). Additionally, the horizontal pleiotropy test also showed a P-value greater than 0.05 for homocysteine, indicating the absence of horizontal pleiotropy (**Supplementary Table S1**). LOO analysis confirmed the robustness of the MR results, with no significant deviations (**Supplementary Fig. S1B**). Finally, the Steiger test confirmed that the causal direction of homocysteine was correct ($p < 0.05$), supporting the reliability of the MR findings (**Supplementary Table S1**).

3.2 Recognition and functional investigation of trait-related genes

The TWAS analysis identified 103 trait-related genes ($p < 0.01$) (**Fig. 1B**, **Supplementary Table S2**). Subsequent GO analysis revealed that these genes were significantly enriched in biological processes such as "viral release from host cell", "dendrite extension", and "lipid droplet formation" ($p < 0.05$) (**Fig. 1C**). Furthermore, KEGG pathway analysis showed that these genes were significantly enriched in pathways like the "Chemokine signaling pathway", "Cytokine-cytokine receptor interaction", and "Taste transduction" (**Fig. 1D**). Collectively, these findings offer a

comprehensive understanding of the potential roles of the identified trait-related genes in IS.

3.3 Identification of macrophages as key cells

scRNA-seq identified 10 annotated cell types: endothelial cells, macrophages, microglial cells, oligodendrocytes, monocytes, astrocytes, vascular smooth muscle cells (VSMCs), Schwann cells, T cells, and NK cells (**Fig. 2A**). The marker genes for each cell type are shown in **Fig. 2B**.

HcyScore analysis revealed variation across the annotated cell types, with macrophages exhibiting the highest HcyScore (**Fig. 2C**). The distribution of HcyScore in annotated cells was further examined in IS and control samples (**Fig. 2D**). Notably, macrophages showed a significant difference in HcyScore ($p < 0.001$) (**Fig. 2E**), identifying macrophages as the pivotal effector cell type.

3.4 Analysis of macrophages subtypes

The distribution of HcyScore was analyzed, revealing that not all macrophages exhibited high HcyScore activity (**Fig. 3A**). Based on this analysis, macrophages were classified into two subtypes: HHcy (high HcyScore) and LHcy (low HcyScore) (**Fig. 3B**). The distribution of these two subtypes showed significant differences. In the IS samples, HHcy macrophages accounted for approximately 51.1% of all macrophages, whereas in the control group, the proportion was 36.1% (**Fig. 3C**). HHcy macrophages were predominantly located in the IS samples, while LHcy macrophages were more abundant in the control samples ($p < 0.05$) (**Fig. 3D**).

To further investigate the causes of HcyScore heterogeneity in macrophages, we performed Slingshot and CytoTRACE analyses. We found that macrophages with higher HcyScores also exhibited elevated CytoTRACE scores, showing significant co-localization (**Fig. 3E-G**). This suggests that the HHcy macrophage population may possess more pronounced stem-like properties and play a key role in initiating and driving the progression of IS.

3.5 Hcyscore activity drives macrophage behavior in IS

Cell-cell communication analysis revealed that HHcy macrophages in IS exhibited distinct intercellular communication pathways compared to other cell types (**Fig. 4A**).

In terms of signal transduction, HHcy macrophages showed significantly higher levels of both incoming and outgoing interaction strengths than other cell types (**Fig. 4B-C**).

Additionally, multiple cell types interacted with macrophages through receptor-ligand pairs. HHcy macrophages displayed a greater number of interaction pairs than LHcy macrophages, including interactions with monocytes (**Fig. 4D**), suggesting that HcyScore may play a crucial role in driving the biological processes of IS.

GSVA revealed that HHcy macrophages were enriched in pathways related to "oxidative phosphorylation", "MYC targets V1", and "protein secretion", among others. In contrast, LHcy macrophages were primarily enriched in pathways related to "angiogenesis" and the "inflammatory response" (**Fig. 4E**). Moreover, HHcy macrophages were found to drive the reprogramming of several metabolic pathways, including Nicotinate and nicotinamide metabolism, Linoleic acid metabolism, and Sphingolipid metabolism (**Fig. 4F**).

3.6 Detection of three feature genes

Based on 103 trait-related genes, multiple feature selection methods were applied: LASSO identified 68 genes, Boruta identified 48 genes, while XGBoost, RF, and GBM each identified the top 5 genes (**Fig. 5A-D**). By integrating these results, three key feature genes, *Gusb*, *Mrpl36*, and *Rab2a*, were obtained (**Fig. 5E**).

Expression analysis showed that, compared with other cell types, all three feature genes were upregulated in macrophages (**Fig. 6A**). Subsequently, we analyzed the expression levels of the three genes in macrophages in the ischemic stroke (IS) and control groups (**Fig. 6B**). ROC analysis demonstrated that these feature genes exhibited good diagnostic performance, with AUC values exceeding 0.6 (**Fig. 6C**). Furthermore, validation in the GSE16561 and GSE32529 datasets confirmed the robust performance of these feature genes (**Fig. 6D-E**).

3.7 Expression dynamics, cell-cell communication, and pathway enrichment analysis of key genes

In the HHcy group, *Gusb* expression gradually increased with disease progression, whereas *Rab2a* and *Mrpl36* initially decreased and then increased. In contrast, in the LHcy group, *Gusb* and *Mrpl36* expression levels remained relatively stable during disease development, while *Rab2a* showed a transient decrease followed by recovery, with final expression levels close to the initial values (**Fig. 7A**). Further analysis revealed significant differences in CytoTRACE scores between *Gusb*-positive (*Gusb*⁺) and *Gusb*-negative (*Gusb*⁻), *Mrpl36*-positive (*Mrpl36*⁺) and *Mrpl36*-negative (*Mrpl36*⁻), *Rab2a*-positive (*Rab2a*⁺) and *Rab2a*-negative (*Rab2a*⁻) macrophages (**Fig.**

7B-C). Correlation analysis showed that Hcyscore was significantly correlated with Gusb expression ($\text{cor} = 0.33, p < 0.05$), Mrpl36 expression ($\text{cor} = 0.3, p < 0.05$), Rab2a expression ($\text{cor} = 0.23, p < 0.05$) (**Fig. 7D**), indicating that Gusb, Mrpl36, and Rab2a are key genes, even though Rab2a's correlation score with HcyScore is lower than the other two genes.

Subsequently, we investigated cell-cell communication using Rab2a-positive and -negative macrophages. Rab2a⁺ macrophages exhibited higher responsiveness in both signal reception and transmission compared to Rab2a⁻ macrophages, suggesting a critical role for Rab2a in IS-related biological processes (**Fig. 7E**). Moreover, Rab2a⁺ macrophages demonstrated stronger interaction strengths in both incoming and outgoing pathways relative to Rab2a⁻ macrophages (**Fig. 7F**).

GO enrichment analysis revealed that Gusb⁺, Mrpl36⁺ and Rab2a⁺ macrophages were mainly enriched in "ribosomal biogenesis" and "ribosomal small subunit biogenesis", whereas Gusb⁻ and Mrpl36⁻ macrophages were primarily enriched in "cytokine-mediated signaling pathways" and "leukocyte cell-cell adhesion", among others. In contrast, Rab2a⁻ macrophages were primarily enriched in functions such as 'defense response to virus' and 'regulation of tumor necrosis factor production' (**Fig. 8A, C and E**).

KEGG pathway analysis further indicated that Gusb⁺, Mrpl36⁺ and Rab2a⁺ macrophages were enriched in "cell cycle" and "DNA replication" pathways. Additionally, Rab2a⁺ macrophages were enriched in the p53 signaling pathway, while Mrpl36⁺ macrophages showed enrichment in Fatty acid metabolism and Fatty acid

elongation pathways. While Gusb⁻ macrophages were enriched in the "NOD-like receptor signaling pathway", "lipid metabolism" pathways (**Fig. 8B**); Mrpl36⁻ macrophages were enriched in the "Lipid and atherosclerosis", "MAPK signaling pathway" pathways (**Fig. 8D**); Rab2a⁻ macrophages were enriched in the "Cell adhesion molecule (CAM) interaction", "Ferroptosis" pathways (**Fig. 8F**).

3.8 Expression analysis of the feature genes at the protein level

To assess the expression levels of Gusb, Mrpl36, and Rab2a in both hypoxia and control samples, Western blot experiments were conducted for verification. The results demonstrated a significant reduction in Mrpl36 expression in the hypoxia samples compared to the control. In contrast, the expression levels of Gusb and Rab2a showed no significant differences between hypoxia and control samples (**Fig. 9A-B**).

3.8 Expression analysis of the feature genes at the mRNA level

QPCR analysis revealed that RAB2A mRNA expression was significantly downregulated in the Hypoxia group compared with the Control group ($P < 0.05$). MRPL36 expression was also lower in the Hypoxia group, though the difference did not reach statistical significance (**Fig. 9C**).

4. Discussion

Hcy, a demethylation byproduct of methionine, is recognized as a toxic amino acid. It is typically cleared from the body through a remethylation process catalyzed by methionine synthase, which utilizes methylcobalamin as a cofactor and 5-methyl tetrahydrofolate as the methyl donor. In the brain, the catabolism of Hcy predominantly depends on its remethylation to methionine, a process facilitated by methylcobalamin and folate[25]. Research has demonstrated that elevated levels of

homocysteine are associated with an increased risk of vascular diseases, including stroke[26]. In this study, MR analysis also identified homocysteine as a significant risk factor for IS) with an odds ratio of 1.06 (95% CI: 1.02-1.12, P = 0.0098), thereby confirming its role in the susceptibility to IS.

In our study, we identified 103 genes related to Hcy traits, which exhibited significant enrichment in the chemokine signaling pathway. Chemokines are essential components of the immune system, as they regulate the directed migration and activation of leukocytes[27]. Additionally, these cytokines play a pivotal role in promoting angiogenesis and contributing to hemostasis[28]. Chemokines are also closely linked to ischemic cerebrovascular diseases[29]. Research has indicated that these molecules are critical in orchestrating the progression of atherosclerosis, serving as chemical attractants for various immune cells from the early stages of endothelial dysfunction, with mediators such as CXCL1 being particularly important[30]. Furthermore, both chemerin and homocysteine levels have been found to independently affect the occurrence and development of ischemic cerebrovascular disease[31].

Consistent with these findings, our study provides evidence that Hcy-related trait genes are significantly enriched in the chemokine signaling pathway. This highlights a potential mechanistic link between homocysteine metabolism and immune-mediated vascular processes.

Furthermore, this study found that macrophages represent the cell population with the highest HcyScore, and there is a significant difference in HcyScore levels between the

439 normal group and the disease group. This suggests that homocysteine levels may be
440 involved in the occurrence and development of IS by regulating macrophage function.
441 Macrophages are a type of multifunctional innate immune cells that can differentiate
442 into two major subtypes: M1 (classically activated macrophages) and M2
443 (alternatively activated macrophages), a process known as macrophage polarization.
444 M1 macrophages are primarily characterized by the secretion of pro-inflammatory
445 cytokines, whereas M2 macrophages are known for producing anti-inflammatory
446 cytokines[32]. Following cerebral ischemia, macrophages respond rapidly and can
447 infiltrate the stroke lesion in large numbers within 24 hours post-onset.
448 Macrophages play a dual role in IS[33, 34]. On one hand, they can exacerbate
449 neuroinflammation by releasing pro-inflammatory cytokines; on the other hand, they
450 contribute to tissue repair by secreting growth factors and anti-inflammatory
451 molecules. The results of this study indicate that macrophages with a high HcyScore
452 are predominantly enriched in IS lesions. These cells also exhibit elevated
453 CytoTRACE scores, demonstrating a notable co-localization characteristic. This
454 suggests that the high HcyScore macrophage population may possess more
455 pronounced stem cell-like traits and could play a pivotal role in the progression of IS.
456 Previous research has shown that Hcy can regulate macrophage polarization[35].
457 Specifically, Hcy can induce the polarization of macrophages towards the M1
458 pro-inflammatory phenotype, which is a critical mechanism by which it promotes the
459 onset and progression of disease[36, 37].

460 Further analysis in this study, using GSVA, revealed that macrophages in the high
461 HcyScore group are primarily enriched in the oxidative phosphorylation pathway.
462 Elevated oxidative stress levels are known to play a crucial role in autoimmune
463 diseases, exacerbating inflammatory responses and leading to tissue damage[38].
464 Under physiological conditions, mitochondria produce a small amount of reactive
465 oxygen species (ROS) through the electron transport chain during respiration.
466 However, under pathological conditions, factors such as inflammatory responses and
467 autoimmune attacks can impair mitochondrial function, resulting in excessive ROS
468 production. This, in turn, creates a vicious cycle of inflammation and tissue
469 damage[39]. Therefore, it can be concluded that high HcyScore macrophages may
470 primarily promote the progression of IS by enhancing inflammatory responses.

471 Furthermore, the three feature genes identified in this study - GUSB, RAB2A, and
472 MRPL36 were all downregulated under HHcy conditions. GUSB functions as a
473 lysosomal housekeeping enzyme involved in glycosaminoglycan degradation and
474 extracellular matrix remodeling[40]. Gusb⁺ macrophages were enriched in ribosomal
475 biogenesis and cell cycle pathways, whereas Gusb⁻ macrophages were associated with
476 NOD-like receptor signaling and lipid metabolism. RAB2A, a Rab GTPase, regulates
477 endoplasmic reticulum-to-Golgi vesicle trafficking and macroautophagy[41, 42], and
478 has been implicated in p53-mediated apoptotic signaling[43]. Rab2a⁺ macrophages
479 were enriched in the p53 signaling pathway and cell cycle, while Rab2a⁻ macrophages
480 were enriched in ferroptosis, cell adhesion molecule interactions, and tumor necrosis
481 factor production. MRPL36, a component of the mitochondrial ribosome large subunit,

is essential for oxidative phosphorylation protein translation and ATP generation[44, 45]. Mrpl36⁺ macrophages were enriched in fatty acid metabolism and fatty acid elongation, whereas Mrpl36⁻ macrophages were enriched in the MAPK signaling pathway and lipid and atherosclerosis-related pathways. Western blot and qPCR validation confirmed that Rab2a and Mrpl36 were significantly downregulated in the HHcy group compared with controls, with Gusb showing the same trend in WB. To our knowledge, this is the first study to implicate Gusb, Rab2a, and Mrpl36 in the context of Hcy-associated IS, providing novel molecular targets for elucidating the mechanisms underlying HHcy-driven, macrophage-mediated neuroinflammation.

This study systematically investigated the potential causal relationship between Hcy and IS through a series of integrated analyses. We calculated the HcyScore at the single-cell level and successfully identified IS-associated macrophages with high HcyScores, followed by functional enrichment analysis of these high-HcyScore macrophages. Ultimately, we identified key genes associated with Hcy in the context of IS. Collectively, our findings reveal the potential biological mechanisms of homocysteine in IS from multiple dimensions, providing important theoretical foundations and novel research directions for the diagnosis and treatment of this disease. However, this study has several limitations. First, the relatively small sample size may affect the generalizability of our findings. Second, this study lacks experimental validation, such as animal model studies. In the future, we will continue to investigate the pathogenesis of IS and further validate and optimize our findings.

5. Conclusion

This study integrated Mendelian randomization, TWAS, single-cell sequencing, and machine learning to systematically elucidate the cell-specific molecular mechanisms by which HHcy drives IS. MR analysis confirmed homocysteine as an independent risk factor for IS (OR = 1.06, 95% CI: 1.02–1.12, P = 0.0098). Single-cell analysis identified macrophages as the key effector cells of HHcy, with the high-activity subtype (HHcy) significantly increased in IS (51.1% vs. 36.1%) and exhibiting stemness properties and oxidative phosphorylation pathway enrichment. Machine learning screened three feature genes—Gusb, Rab2a, and Mrpl36. Functional enrichment revealed that Gusb⁺, Rab2a⁺, and Mrpl36⁺ macrophages were predominantly enriched in ribosomal biogenesis and cell cycle pathways, whereas their negative counterparts were associated with pro-inflammatory and injury-related pathways, including NOD-like receptor signaling (Gusb⁻), ferroptosis and cell adhesion molecule interactions (Rab2a⁻), and MAPK signaling and lipid/atherosclerosis (Mrpl36⁻). Western blot and qPCR validation further confirmed that all three genes were downregulated under HHcy conditions, with Rab2a (qPCR, P < 0.05) and Mrpl36 (WB, P < 0.05) exhibiting statistically significant differences. This study establishes, for the first time, a "HHcy-macrophage-Gusb/Rab2a/Mrpl36" molecular regulatory axis, providing novel insights and a theoretical foundation for deciphering IS pathogenesis and developing targeted interventions.

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.

Author Contributions

Conceptualization, Li-jun Ma.; methodology, Feng-chen Bi and Xi-bo Yang.; software, Jing Wang.; validation, Yan Li. and Li-jun Ma; formal analysis, Jing Wang.; investigation, Li-jun Ma; resources, Qing-hua Huang; data curation, Nan Zhang.; writing—original draft preparation, Qing-hua Huang; writing—review and editing, Li-jun Ma.; visualization, Jing Wang.; supervision, Li-jun Ma. and Yan Li.; project administration, Li-jun Ma; funding acquisition, Li-jun Ma. and Yan Li. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the scientific research project of Ningxia Medical University (Special talent start project, NO. XT2023004) and the Open Research Fund of Key Laboratory of Endocrine Glucose & Lipids Metabolism and Brain Aging, Ministry of Education (Open Research Fund NO. 2024JYBZDSYS013).

Acknowledgments

The author is grateful to the participants and investigators of this study.

Data Availability Statement

Our study utilized data from the Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) and the Integrative Epidemiology Unit (IEU) Open GWAS database (<https://gwas.mrcieu.ac.uk/>).

Reference

1. Lyu Y, Li W, Tang T: **Prevalence Trends and Influencing Factors of Post-Stroke Depression: A Study Based on the National Health and Nutrition Examination Survey.** *Med Sci Monit* 2022, **28**:e933367.
2. Saini V, Guada L, Yavagal DR: **Global Epidemiology of Stroke and Access to Acute Ischemic Stroke Interventions.** *Neurology* 2021, **97**:S6-s16.

- 551 3. Kuriakose D, Xiao Z: **Pathophysiology and Treatment of Stroke: Present Status and**
552 **Future Perspectives.** *Int J Mol Sci* 2020, **21**.
- 553 4. Zhu H, Hu S, Li Y, Sun Y, Xiong X, Hu X, Chen J, Qiu S: **Interleukins and Ischemic Stroke.**
554 *Front Immunol* 2022, **13**:828447.
- 555 5. Torphy RJ, Fujiwara Y, Schulick RD: **Pancreatic cancer treatment: better, but a long way**
556 **to go.** *Surg Today* 2020, **50**:1117-1125.
- 557 6. Grotta JC: **Intravenous Thrombolysis for Acute Ischemic Stroke.** *Continuum (Minneap*
558 *Minn)* 2023, **29**:425-442.
- 559 7. Meyne J, Domschikowski M, Hensler J, Helmers AK, Berg D, Deuschl G, Paschen S:
560 **Intravenous thrombolysis and mechanical recanalization for acute ischemic stroke in**
561 **deep brain stimulation patients: a case series.** *J Neurol* 2023, **270**:4354-4359.
- 562 8. Lu W, Wen J: **Role and Relationship Between Homocysteine and H(2)S in Ischemic**
563 **Stroke.** *Mol Neurobiol* 2025, **62**:14613-14626.
- 564 9. Zhang T, Jiang Y, Zhang S, Tie T, Cheng Y, Su X, Man Z, Hou J, Sun L, Tian M, et al: **The**
565 **association between homocysteine and ischemic stroke subtypes in Chinese: A**
566 **meta-analysis.** *Medicine (Baltimore)* 2020, **99**:e19467.
- 567 10. Jakubowski H, Witucki Ł: **Homocysteine Metabolites, Endothelial Dysfunction, and**
568 **Cardiovascular Disease.** *Int J Mol Sci* 2025, **26**.
- 569 11. Shi M, Zheng J, Liu Y, Mao X, Wu X, Chu M, Niu H, Sun L, Chang X, He Y, et al: **Folate,**
570 **Homocysteine, and Adverse Outcomes After Ischemic Stroke.** *J Am Heart Assoc* 2024,
571 **13**:e036527.
- 572 12. Martí-Carvajal AJ, Solà I, Lathyris D, Dayer M: **Homocysteine-lowering interventions for**
573 **preventing cardiovascular events.** *Cochrane Database Syst Rev* 2017, **8**:Cd006612.
- 574 13. Cheng C, Chen W, Jin H, Chen X: **A Review of Single-Cell RNA-Seq Annotation,**
575 **Integration, and Cell-Cell Communication.** *Cells* 2023, **12**.
- 576 14. Hu X, Wang Z, Zhang H, Cui P, Li Y, Chen X, Kong C, Wang W, Lu S: **Single-cell**
577 **sequencing: New insights for intervertebral disc degeneration.** *Biomed Pharmacother*
578 2023, **165**:115224.
- 579 15. Qiu M, Zong JB, He QW, Liu YX, Wan Y, Li M, Zhou YF, Wu JH, Hu B: **Cell Heterogeneity**
580 **Uncovered by Single-Cell RNA Sequencing Offers Potential Therapeutic Targets for**
581 **Ischemic Stroke.** *Aging Dis* 2022, **13**:1436-1454.
- 582 16. Zheng K, Lin L, Jiang W, Chen L, Zhang X, Zhang Q, Ren Y, Hao J: **Single-cell RNA-seq**
583 **reveals the transcriptional landscape in ischemic stroke.** *J Cereb Blood Flow Metab* 2022,
584 **42**:56-73.
- 585 17. Zhou H, Zhang Y, Liu J, Yang Y, Fang W, Hong S, Chen G, Zhao S, Zhang Z, Shen J, et al:
586 **Education and lung cancer: a Mendelian randomization study.** *Int J Epidemiol* 2019,
587 **48**:743-750.
- 588 18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers**
589 **differential expression analyses for RNA-sequencing and microarray studies.** *Nucleic*
590 *Acids Res* 2015, **43**:e47.
- 591 19. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel
592 R, Ruusalepp A, Quertermous T, Hao K, et al: **Opportunities and challenges for**
593 **transcriptome-wide association studies.** *Nat Genet* 2019, **51**:592-599.
- 594 20. Song S, Wang L, Hou L, Liu JS: **Partitioning and aggregating cross-tissue and**

595 tissue-specific genetic effects to identify gene-trait associations. *Nat Commun* 2024,
596 15:5769.

597 21. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological**
598 **themes among gene clusters.** *Omics* 2012, 16:284-287.

599 22. Fan C, Chen F, Chen Y, Huang L, Wang M, Liu Y, Wang Y, Guo H, Zheng N, Liu Y, et al:
600 **irGSEA: the integration of single-cell rank-based gene set enrichment analysis.** *Brief*
601 *Bioinform* 2024, 25.

602 23. Luo J, Deng M, Zhang X, Sun X: **ESICCC as a systematic computational framework for**
603 **evaluation, selection, and integration of cell-cell communication inference methods.**
604 *Genome Res* 2023, 33:1788-1805.

605 24. Cao G, Xuan X, Li Y, Hu J, Zhang R, Jin H, Dong H: **Single-cell RNA sequencing reveals**
606 **the vascular smooth muscle cell phenotypic landscape in aortic aneurysm.** *Cell Commun*
607 *Signal* 2023, 21:113.

608 25. Rabelo NN, Telles JPM, Pipek LZ, Farias Vidigal Nascimento R, Gusmão RC, Teixeira MJ,
609 Figueiredo EG: **Homocysteine is associated with higher risks of ischemic stroke: A**
610 **systematic review and meta-analysis.** *PLoS One* 2022, 17:e0276087.

611 26. Pinzon RT, Wijaya VO, Veronica V: **The role of homocysteine levels as a risk factor of**
612 **ischemic stroke events: a systematic review and meta-analysis.** *Front Neurol* 2023,
613 14:1144584.

614 27. Cambier S, Gouwy M, Proost P: **The chemokines CXCL8 and CXCL12: molecular and**
615 **functional properties, role in disease and efforts towards pharmacological intervention.**
616 *Cell Mol Immunol* 2023, 20:217-251.

617 28. Sadeghi M, Dehnavi S, Asadirad A, Xu S, Majeed M, Jamialahmadi T, Johnston TP, Sahebkar
618 A: **Curcumin and chemokines: mechanism of action and therapeutic potential in**
619 **inflammatory diseases.** *Inflammopharmacology* 2023, 31:1069-1093.

620 29. Demir CF, Ataş İ N, Balgetir F, Artaş H, Gönen M, Aydin S: **Increased serum chemerin**
621 **levels associated with carotid intima-media thickness.** *Arq Neuropsiquiatr* 2021,
622 79:189-194.

623 30. Gencer S, Evans BR, van der Vorst EPC, Döring Y, Weber C: **Inflammatory Chemokines in**
624 **Atherosclerosis.** *Cells* 2021, 10.

625 31. Wang L, Jia J, Hong Z, Zhang L, Zhang J: **Effects of chemerin and homocysteine levels and**
626 **their associations with occurrence and development of ischemic cerebrovascular disease.**
627 *Lipids Health Dis* 2021, 20:108.

628 32. Bernier ED, Bartnicki E, Khanna KM: **Macrophages: sentinels, warriors, and healers.** *Hum*
629 *Mol Genet* 2025, 34:R110-r120.

630 33. Cai W, Dai X, Chen J, Zhao J, Xu M, Zhang L, Yang B, Zhang W, Rocha M, Nakao T, et al:
631 **STAT6/Arg1 promotes microglia/macrophage efferocytosis and inflammation resolution**
632 **in stroke mice.** *JCI Insight* 2019, 4.

633 34. Cai W, Liu S, Hu M, Sun X, Qiu W, Zheng S, Hu X, Lu Z: **Post-stroke DHA Treatment**
634 **Protects Against Acute Ischemic Brain Injury by Skewing Macrophage Polarity Toward**
635 **the M2 Phenotype.** *Transl Stroke Res* 2018, 9:669-680.

636 35. Liang Y, Wang Y, Tan J, Shu J, Xu Y, Wang M, Yang S, Ma L: **Homocysteine affects**
637 **macrophage polarization by altering m6A methylation of scavenger receptors CD209**
638 **and CD163L1.** *Epigenetics* 2024, 19:2437272.

36. Pang Y, Li Y, Lv Y, Sun L, Zhang S, Li Y, Wang Y, Liu G, Xu MJ, Wang X, Jiang C: **Intermedin Restores Hyperhomocysteinemia-induced Macrophage Polarization and Improves Insulin Resistance in Mice.** *J Biol Chem* 2016, **291**:12336-12345.
37. Gao S, Wang L, Liu W, Wu Y, Yuan Z: **The synergistic effect of homocysteine and lipopolysaccharide on the differentiation and conversion of raw264.7 macrophages.** *J Inflamm (Lond)* 2014, **11**:13.
38. Koutsaliaris IK, Moschonas IC, Pechlivani LM, Tsouka AN, Tselepis AD: **Inflammation, Oxidative Stress, Vascular Aging and Atherosclerotic Ischemic Stroke.** *Curr Med Chem* 2022, **29**:5496-5509.
39. Kurien BT, Ice JA, Wood RA, Pharaoh G, Cavett J, Lewis V, Bhaskaran S, Rasmussen A, Lessard CJ, Farris AD, et al: **Mitochondrial dysfunction and fatigue in Sjögren's disease.** *RMD Open* 2025, **11**.
40. Naz H, Islam A, Waheed A, Sly WS, Ahmad F, Hassan I: **Human β -glucuronidase: structure, function, and application in enzyme replacement therapy.** *Rejuvenation Res* 2013, **16**:352-363.
41. Tisdale EJ, Bourne JR, Khosravi-Far R, Der CJ, Balch WE: **GTP-binding mutants of rab1 and rab2 are potent inhibitors of vesicular transport from the endoplasmic reticulum to the Golgi complex.** *J Cell Biol* 1992, **119**:749-761.
42. Kajiho H, Kajiho Y, Frittoli E, Confalonieri S, Bertalot G, Viale G, Di Fiore PP, Oldani A, Garre M, Beznoussenko GV, et al: **RAB2A controls MT1-MMP endocytic and E-cadherin polarized Golgi trafficking to promote invasive breast cancer programs.** *EMBO Rep* 2016, **17**:1061-1080.
43. Cai H, Tian P, Ju J, Wang T, Chen X, Wang K, Wang F, Yu X, Wang S, Wang Y, et al: **Long noncoding RNA NONMMUT015745 inhibits doxorubicin-mediated cardiomyocyte apoptosis by regulating Rab2A-p53 axis.** *Cell Death Discov* 2022, **8**:364.
44. Prestele M, Vogel F, Reichert AS, Herrmann JM, Ott M: **Mrpl36 is important for generation of assembly competent proteins during mitochondrial translation.** *Mol Biol Cell* 2009, **20**:2615-2625.
45. Piao L, Li Y, Kim SJ, Byun HS, Huang SM, Hwang SK, Yang KJ, Park KA, Won M, Hong J, et al: **Association of LETM1 and MRPL36 contributes to the regulation of mitochondrial ATP production and necrotic cell death.** *Cancer Res* 2009, **69**:3397-3404.

Figures

Figure 1. Causal relationship between homocysteine and ischemic stroke (IS), and identification of trait-related genes. (A) Forest plot of Mendelian randomization analysis showing the causal effect of homocysteine on IS. (B) Manhattan plot displaying the transcriptome-wide association study (TWAS) results for trait-related

genes. (C) Gene Ontology (GO) enrichment analysis of the 103 identified trait-related genes, showing the top enriched biological processes. (D) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of the trait-related genes, highlighting key pathways involved. OR, odds ratio; CI, confidence interval.

Figure 2. Single-cell landscape and HcyScore-based identification of pivotal effector cells in IS. (A) Uniform manifold approximation and projection (UMAP) visualization of 10 annotated cell types in the scRNA-seq dataset. (B) Dot plot showing the expression of marker genes across the annotated cell types. Dot size represents the proportion of cells expressing the gene, and color intensity indicates the average expression level. (C) HcyScore distribution across the 10 annotated cell types. (D) HcyScore distribution in individual cells stratified by IS and control samples. (E) Violin plot comparing HcyScore in macrophages between IS and control groups. VSMCs, vascular smooth muscle cells; NK, natural killer; scRNA-seq, single-cell RNA sequencing; IS, ischemic stroke. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, ns $P > 0.05$.

Figure 3. Identification and characterization of macrophage subtypes based on HcyScore heterogeneity. (A) HcyScore distribution within the macrophage population. (B) Classification of macrophages into two subtypes based on HcyScore: HHcy (high HcyScore, red) and LHcy (low HcyScore, Green). (C) Proportional composition of HHcy and LHcy macrophages in IS and control samples. (D) Spatial distribution of HHcy and LHcy macrophages across IS and control samples. (E) Slingshot pseudotime trajectory analysis (F) UMAP displaying the CytoTRACE score

distribution (G) Boxline plot showing the differences of CytoTRACE scores between HHcy and LHcy macrophages. *P < 0.05, **P < 0.01, ***P < 0.001, ns P > 0.05.

Figure 4. HcyScore-driven intercellular communication and metabolic reprogramming in macrophage subtypes during ischemic stroke (IS). (A) Circle plot illustrating distinct intercellular communication patterns of HHcy macrophages with other cell types in IS. (B–C) Comparing incoming and outgoing interaction strengths across cell types. (D) Dot plot of receptor-ligand interactions between macrophage subtypes and other cell types. (E) Gene set variation analysis (GSVA) heatmap comparing pathway enrichment between HHcy and LHcy macrophages. (F) Metabolic pathway activity scores HHcy and LHcy macrophages.

Figure 5. Identification of feature genes through integrated machine learning approaches. (A) LASSO regression analysis and random forest (RF) algorithm. (B) Boruta algorithm. (C) XGBoost algorithm. (D) Gradient boosting machine (GBM) algorithm. (E) Venn diagram illustrating the intersection of feature genes identified by LASSO, Boruta, XGBoost, RF, and GBM, yielding three consensus feature genes. LASSO, least absolute shrinkage and selection operator; RF, random forest; GBM, gradient boosting machine.

Figure 6. Expression profiling and diagnostic validation of the three feature genes. (A) UMAP plot showing the expression of Gusb, Mrpl36, and Rab2a across all annotated cell types. (B) Violin plot comparing Gusb expression between ischemic stroke (IS) and control tissues in each annotated cell type. (C) Receiver operating characteristic (ROC) curves evaluating the diagnostic performance of the three feature

genes. (D–E) External validation of feature gene expression in the GSE16561 (D) and GSE32529 (E) datasets. *P < 0.05, **P < 0.01, ***P < 0.001, ns P > 0.05.

Figure 7. Dynamic expression of feature genes and functional characterization of Gusb⁺ macrophages in ischemic stroke. (A) Pseudotime trajectory analysis showing the dynamic expression patterns of Gusb, Rab2a, and Mrpl36 across disease progression in HHcy and LHcy macrophage subtypes. (B–C) Comparison of CytoTRACE scores between Gusb-positive (Gusb⁺) and Gusb-negative (Gusb⁻), Mrpl36-positive (Mrpl36⁺) and Mrpl36-negative (Mrpl36⁻), Rab2a-positive (Rab2a⁺) and Rab2a-negative (Rab2a⁻) macrophages. (D) Scatter plot showing a significant correlation between HcyScore and Gusb, Mrpl36 and Rab2a expression. (E) Cellular communication networks. (F) Incoming and outgoing interaction strengths. *P < 0.05, **P < 0.01, ***P < 0.001, ns P > 0.05.

Figure 8. Divergent functional profiles of positive and negative macrophages in ischemic stroke. (A, C, E) Gene Ontology (GO) enrichment analysis of Gusb⁺ and Gusb⁻, Mrpl36⁺ and Mrpl36⁻, Rab2a⁺ and Rab2a⁻ macrophages. (B, D, F) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of Gusb⁺ and Gusb⁻, Mrpl36⁺ and Mrpl36⁻, Rab2a⁺ and Rab2a⁻ macrophages.

Figure 9. Experimental verification of feature genes. (A) Representative Western blot images. (B) Results of Western blot for feature genes (P < 0.05). (C) Relative mRNA expression of RAB2A and MRPL36 in the Hypoxia group.

Additional files

Supplementary Figure 1. Sensitivity analyses of Mendelian randomization (MR) for homocysteine on ischemic stroke (IS). (A) Results of five MR methods estimating the causal effect of homocysteine on IS. (B) Leave-one-out (LOO) analysis demonstrating the robustness of the MR results. Each data point represents the pooled estimate when a single genetic instrument is excluded. Error bars indicate 95% confidence intervals.

Supplementary Table 1. Sensitivity and validity tests for Mendelian randomization analysis of homocysteine on ischemic stroke.

Supplementary Table 2. Trait-related genes identified by transcriptome-wide association study (TWAS).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable.xlsx](#)
- [SupplementaryFigure1.tif](#)
- [SupplementaryFigure2.tif](#)
- [supplementaryfile.tif](#)
- [Figure1.tif](#)
- [Figure2.tif](#)
- [Figure3.tif](#)
- [Figure4.tif](#)
- [Figure5.tif](#)
- [Figure6.tif](#)
- [Figure7.tif](#)
- [Figure8.tif](#)
- [Figure9.tif](#)