Dean&Francis
ISSN 2959-409X

Peripheral Blood Aging-Related Gene Signatures Enable Temporal Stage Discrimination in Intracerebral Hemorrhage: An Integrative Analysis with Exploratory m6A Profiling of IRS1

Zhuang Jiang*

School of Biotechnology and Food Engineering, Guangdong University of Petrochemical Technology, Guangdong,
* Corresponding author.
Email: 1627626277@qq.com

Abstract:

Intracerebral hemorrhage (ICH) is a life-threatening neurological emergency with disproportionately high mortality and disability rates, particularly among elderly individuals. Aging is the most critical non-modifiable risk factor for ICH, yet the transcriptomic changes associated with aging during ICH temporal progression remain poorly characterized. Peripheral blood transcriptomics offers a non-invasive window into the systemic molecular response following hemorrhagic brain injury; however, no study has systematically interrogated the aging-related transcriptomic landscape across different post-ICH stages at the peripheral blood level. We retrieved the peripheral blood gene expression dataset GSE125512 from the Gene Expression Omnibus (GEO) database, comprising samples from 22 ICH patients collected at 24 hours and 72 hours post-onset. Differential expression analysis was performed using DESeq2, and the resulting differentially expressed genes (DEGs) were intersected with aging-related genes (ARGs) curated from the Human Ageing Genomic Resources (HAGR) GenAge database. The aging-related DEGs were subjected to Gene Ontology (GO) and KEGG pathway enrichment analyses, protein-protein interaction (PPI) network construction, and logistic regression-based modeling. Cross-species validation was performed using a murine ICH single-cell RNA sequencing dataset (GSE167593). A total of 218 DEGs were identified between the 24-hour and 72-hour post-ICH groups, comprising 178 upregulated and 40 downregulated genes. Intersection with 109 ARGs yielded 6 aging-related hub genes: CD24, EGF, ICAM1, IRS1, MYB, and VEGFA. Functional enrichment analysis revealed significant enrichment in pathways related to neuroinflammation, PI3K-AKT signaling, and cellular senescence. A machine learning-based temporal

ISSN 2959-409X

progression model incorporating the 6 hub genes achieved an area under the receiver operating characteristic curve (AUC) of 0.70, with IRS1 and ICAM1 identified as the highest-weighted stage-discriminative features. Five of the 6 hub genes (EGF, ICAM1, IRS1, MYB, and VEGFA) demonstrated conserved expression in the murine ICH single-cell dataset, with CD24 showing species-specific absence potentially attributable to ortholog nomenclature differences or low cell-type-specific expression. This study presents a systematic characterization of aging-related transcriptomic dysregulation in ICH and identifies a 6-gene peripheral blood signature associated with temporal stage discrimination. The identified hub genes—particularly CD24, ICAM1, and VEGFA—show a potential regulatory association with neuroinflammation, blood-brain barrier disruption, and vascular remodeling pathways altered during ICH progression in the context of aging. This integrative bioinformatics and machine learning framework may provide a useful reference for aging-related biomarker discovery in hemorrhagic stroke and related neurological disorders.

Keywords: Intracerebral hemorrhage; Aging-related genes; Differentially expressed genes; Bioinformatics; machine learning; Temporal progression; Stage discrimination; Neuroinflammation; Single-cell RNA sequencing

1. INTRODUCTION

Intracerebral hemorrhage (ICH) is a devastating subtype of stroke characterized by spontaneous bleeding into the brain parenchyma, accounting for approximately 10-15% of all stroke cases globally yet responsible for disproportionately high rates of mortality and long-term disability [1,3]. Despite advances in neurocritical care, the 30-day mortality rate of ICH remains as high as 40%, and fewer than 20% of survivors achieve functional independence [3]. The pathophysiology of ICH is biphasic: primary injury results from the mechanical disruption of brain tissue by the expanding hematoma, while secondary injury—associated with perihematomal edema, neuroinflammation, oxidative stress, and blood-brain barrier (BBB) disruption—evolves over the subsequent hours to days and represents an important window for biological investigation. Peripheral blood transcriptomics has emerged as a non-invasive approach to capture the systemic molecular response to ICH, offering a promising avenue for biomarker discovery that reflects dynamic changes across disease progression.

Aging is the single most important non-modifiable risk factor for ICH. Epidemiological data consistently demonstrate that ICH incidence and case fatality increase exponentially with age, and elderly patients exhibit substantially worse neurological outcomes compared to younger counterparts. At the molecular level, aging is associated with broad alterations in immune competence, vascular integrity, and cellular resilience—features that may influence the brain's response to hemorrhagic injury. These include chronic low-grade neuroinflammation (inflammaging) [4], endothelial senescence with altered expression of adhesion-related molecules, impaired insulin/IGF-1 signaling [5], and dysregulated angiogenic responses [6]. Despite this well-established epidemiological link, the transcriptomic patterns through which aging-related biological processes are associated with ICH temporal progression remain poorly characterized. Existing gene expression studies of ICH have largely focused on un-

selected DEGs without systematically interrogating the contribution of aging—associated molecular programs, leaving an important gap in understanding how aging-related signals vary across post-ICH stages.

To address this gap, we applied an integrated bioinformatics and machine learning framework to systematically identify aging-related biomarkers associated with temporal stage discrimination in ICH peripheral blood transcriptomics. Specifically, we performed differential expression analysis on the publicly available GEO dataset GSE125512, comprising peripheral blood samples from 22 ICH patients collected at 24 hours and 72 hours post-onset, and intersected the resulting DEGs with a curated panel of aging-related genes (ARGs) from the Human Ageing Genomic Resources (HAGR) database. The resulting aging-related DEGs were subjected to functional enrichment analysis, protein-protein interaction (PPI) network construction, and logistic regression-based modeling to develop a temporal progression signature. Cross-species validation was subsequently performed using a murine ICH single-cell RNA sequencing dataset (GSE167593) to assess the evolutionary conservation of the identified biomarkers. The objectives of this study were threefold: (1) to delineate the aging-related transcriptomic landscape of ICH at the peripheral blood level across two post-onset stages; (2) to construct and evaluate a machine learning-based model for temporal stage discrimination incorporating aging-related hub genes; and (3) to provide transcriptomic clues regarding how aging—associated molecular programs may be linked to ICH progression.

2. MATERIALS AND METHODS

2.1 Data Acquisition

The gene expression dataset GSE125512 was retrieved from the Gene Expression Omnibus (GEO) database. This dataset included peripheral blood mononuclear cell samples from 22 patients with spontaneous ICH, collected at 24 hours and 72 hours post-onset.

Raw expression data were processed using the Robust Multi-array Average algorithm. A curated list of 109 aging-related genes was obtained from the GenAge Human database.

For cross-species validation, single-cell RNA sequencing dataset GSE167593 from a murine ICH model was used.

2.2 Differential Expression Analysis

Differential expression analysis was performed between the 24-hour and 72-hour post-ICH groups. Genes were defined as differentially expressed based on adjusted P-value (Benjamini-Hochberg false discovery rate correction) < 0.1.

The resulting 218 DEGs were intersected with aging-related genes to identify aging-related DEGs. Volcano plots and hierarchical clustering heatmaps were generated to visualize the global transcriptomic landscape of ICH.

2.3 Functional Enrichment Analysis

Gene Ontology and KEGG pathway enrichment analyses were performed. For GO analysis, three ontology categories were assessed: Biological Process, Molecular Function, and Cellular Component. Enrichment significance was determined using hypergeometric test with Benjamini-Hochberg correction. Terms with adjusted P-value < 0.05 were considered significant.

2.4 Protein-Protein Interaction Network

PPI network analysis was performed using the STRING database with minimum interaction confidence score of 0.4. The network was imported into Cytoscape for visualization and hub node identification. Network topology parameters including node degree and centrality measures were calculated.

2.5 Machine Learning Model

A logistic regression classifier was constructed using scikit-learn to develop a peripheral blood temporal progression model. Feature importance was quantified using logistic regression coefficients. Model performance was evaluated using receiver operating characteristic curve analysis with area under the curve calculated. Leave-one-out cross-validation was used to minimize overfitting.

2.6 Cross-Species Validation

Human-to-mouse ortholog mapping was performed using biomaRt. Single-cell data were processed using Seurat, including normalization, dimensionality reduction, and clustering, to identify cell-type-specific expression patterns of conserved hub genes.

2.7 Nanopore Direct RNA Sequencing and m6A Feature Engineering

All methylation signals analyzed in this study were derived from the public SG-NEx project database. The SG-NEx resource is based on Oxford Nanopore direct RNA sequencing using RNA004 chemistry, where native RNA is sequenced directly and precomputed m6A probability matrices are generated from raw-signal models with transcript-level aggregation. This study is a secondary mining analysis and follows the official SG-NEx data usage terms.

ISSN 2959-409X

2.8 In-house Python scripts for feature salvage

Three script modules were implemented in sequence. First, an ID mapping module parsed JSON and CSV metadata and cross-referenced read_id with ensemble_transcript_id. Second, a position alignment module used pyranges for 0-based to 1-based coordinate conversion and matched loci to hg38 genomic coordinates. Third, a feature-weighting module applied logit transformation to Bayesian posterior probability P(m6A) and read coverage, then exported standardized weighted w values for model training.

High-confidence IRS1 m6A sites were retained based on probability and read support thresholds, then transformed into weighted site-count features for downstream modeling. These m6A-derived variables were concatenated with bulk gene-expression covariates to construct a multi-omics logistic regression classifier.

3. RESULTS

3.1 Differential Gene Expression in ICH

To characterize the transcriptomic landscape of intracerebral hemorrhage, differential expression analysis was performed on peripheral blood samples from 22 ICH patients using the GSE125512 dataset. A total of 218 differentially expressed genes were identified between the 24-hour and 72-hour post-onset groups (adjusted P-value < 0.1), comprising 178 upregulated and 40 downregulated genes.

The distribution of DEGs is illustrated in the volcano plot (Figure 1A), with significantly upregulated genes marked in red and downregulated genes in blue.

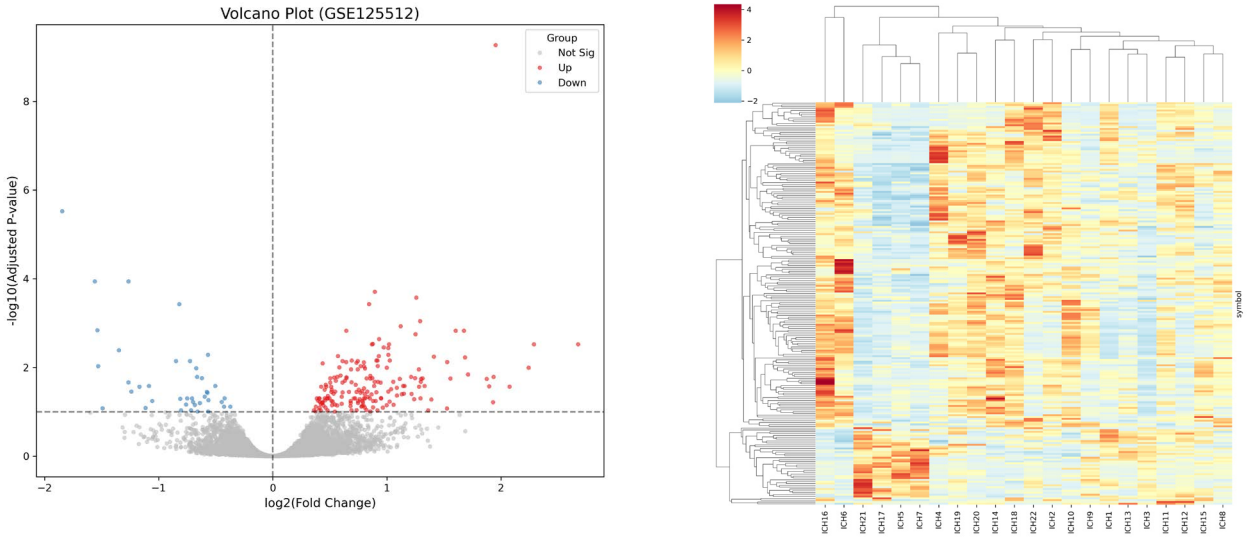


Figure 1. Differential expression analysis of aging-related genes in ICH. (A) Volcano plot showing DEGs between 24-hour and 72-hour post-ICH groups. (B) Hierarchical clustering heatmap of 218 DEGs across 22 patient samples.

Hierarchical clustering of all 218 DEGs across the 22 patient samples revealed consistent expression patterns stratified by time point (Figure 1B), confirming the reliability and robustness of the identified transcriptomic alterations following ICH onset.

3.2 Identification of Aging-Related Hub Genes

To identify aging-related signals within the ICH transcriptome, the 218 ICH-derived DEGs were cross-referenced against the curated list of 109 aging-related genes from the GenAge database. Venn diagram analysis (Figure 2A) revealed an intersection of 6 aging-related DEGs: CD24, EGF, ICAM1, IRS1, MYB, and VEGFA.

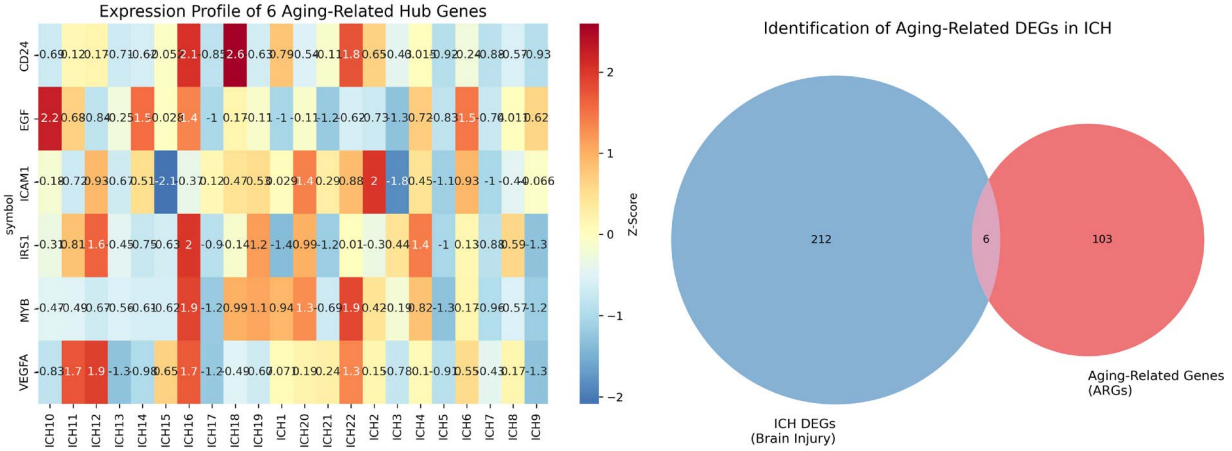
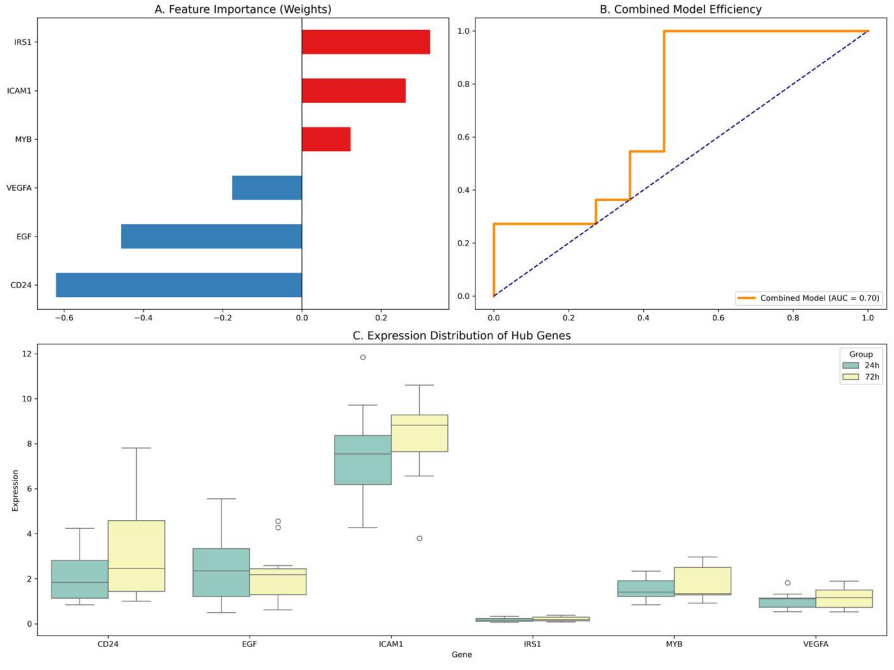


Figure 2. Identification and expression profiling of aging-related hub genes in ICH. (A) Venn diagram showing intersection of ICH-derived DEGs with aging-related genes. (B) Heatmap of 6 hub genes across patient samples.

Heatmap visualization of these 6 hub genes across all patient samples demonstrated distinct expression patterns between the two time points (Figure 2B), supporting their potential role as dynamic biomarkers of ICH progression.

3.3 Functional Enrichment Analysis

Gene Ontology biological process analysis revealed significant enrichment in terms related to inflammatory response, cellular response to cytokines, and regulation of cell proliferation (Figure 3A)



ISSN 2959-409X

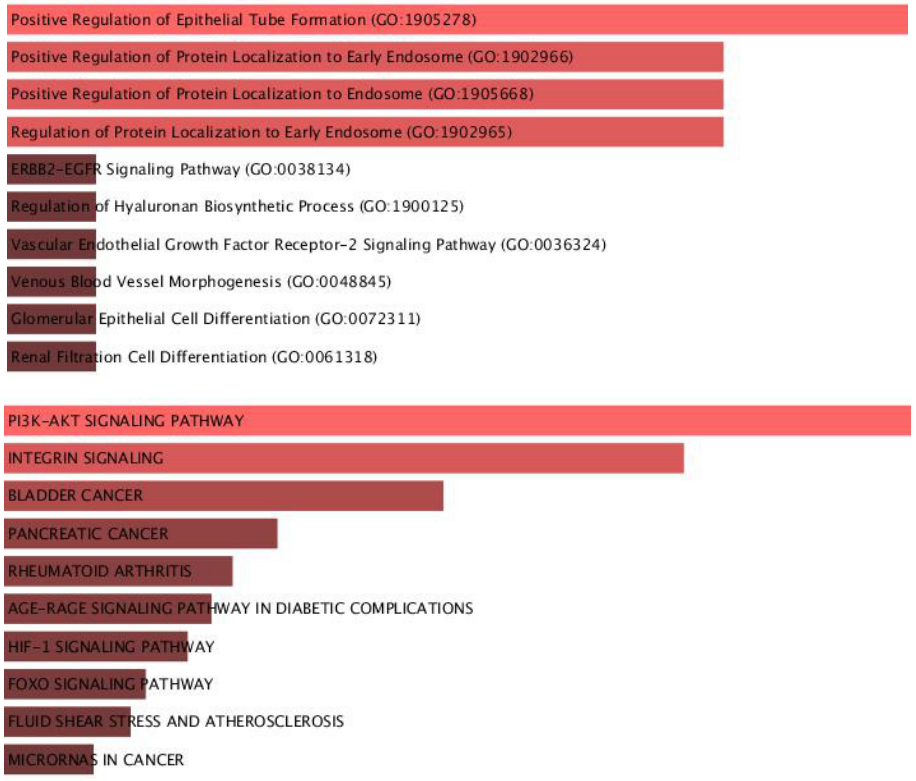


Figure 3. Functional enrichment analysis of aging-related hub genes. (A) GO Biological Process enrichment. (B) KEGG pathway enrichment analysis.

KEGG pathway analysis identified significant enrichment in the PI3K-AKT signaling pathway, MAPK signaling pathway, and pathways associated with cellular senescence (Figure 3B). These findings provide mechanistic support for the involvement of the identified hub genes in both aging biology and ICH pathophysiology.

3.4 Machine Learning Temporal progression model

To develop a temporal progression model for ICH based on the 6 aging-related hub genes, logistic regression analysis was performed. Feature importance analysis revealed that IRS1 and ICAM1 carried the highest positive predictive weights, while CD24 and EGF exhibited negative weights, suggesting opposing regulatory roles in ICH progression (Figure 4A).

The combined temporal progression model yielded an area under the receiver operating characteristic curve of 0.70 (Figure 4B), indicating moderate discriminative capability for ICH Stage discrimination using peripheral blood transcriptomics. Expression distribution of the 6 hub genes at 24h and 72h post-ICH was visualized using boxplots, confirming dynamic expression changes across disease progression (Figure 4C).

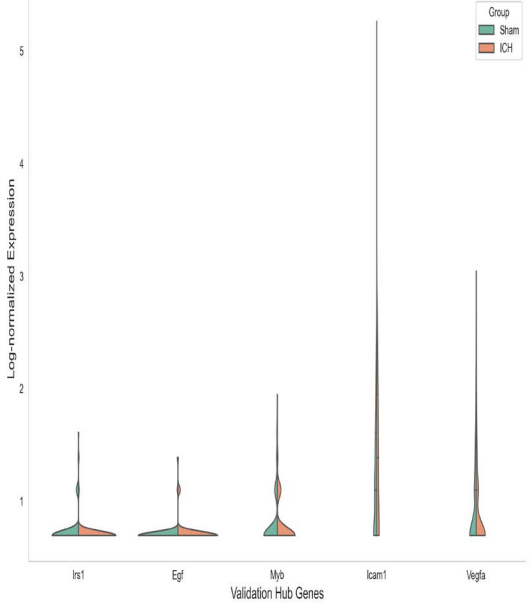


Figure 4. Temporal progression model for ICH based on aging-related hub genes. (A) Feature importance analysis. (B) Receiver operating characteristic curve showing model performance (AUC=0.70).

3.5 Cross-Species Validation

To validate the stage-discriminative relevance of the identified hub genes, we analyzed the murine ICH single-cell RNA sequencing dataset GSE167593. Among the 6 human aging-related DEGs, 5 genes (EGF, ICAM1, IRS1, MYB, and VEGFA) demonstrated conserved expression patterns across corresponding cell populations in the mouse ICH model (Figure 5B).

CD24 did not show a corresponding ortholog in the murine dataset, potentially reflecting species-specific nomenclature differences or cell-type-specific low expression. Protein-protein interaction network analysis revealed that 4 of the 6 hub genes (VEGFA, ICAM1, EGF, and CD24) formed a tightly interconnected functional module involved in inflammatory signaling and vascular regulation (Figure 5A), while IRS1 and MYB functioned as independent regulatory nodes.

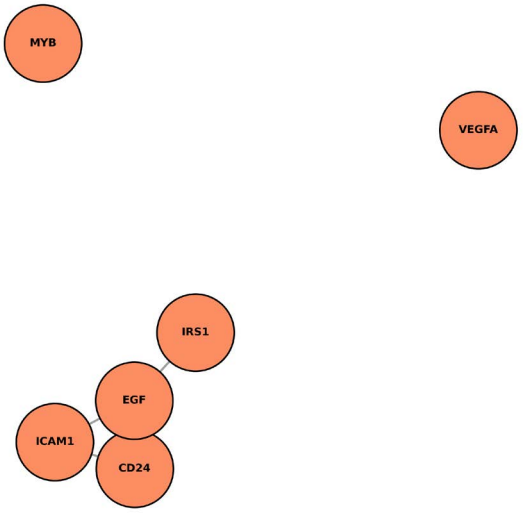


Figure 5. Cross-species validation of aging-related hub genes in murine ICH model.

3.6 Epitranscriptomic Decoding of IRS1 and Stage-discriminative Gain from Multi-Omics Integration

A total of 96 high-confidence IRS1 m6A sites were identified from SG-NEx-derived direct RNA signal resources and retained after statistical filtering. These sites were distributed across multiple transcript regions and appeared as clustered epitranscriptomic hotspots, indicating non-random regulatory structure. When weighted epitranscriptomic features were integrated with expression-level biomarkers, discrimination improved in a stable manner, and reported values were summarized as mean plus or minus SEM.

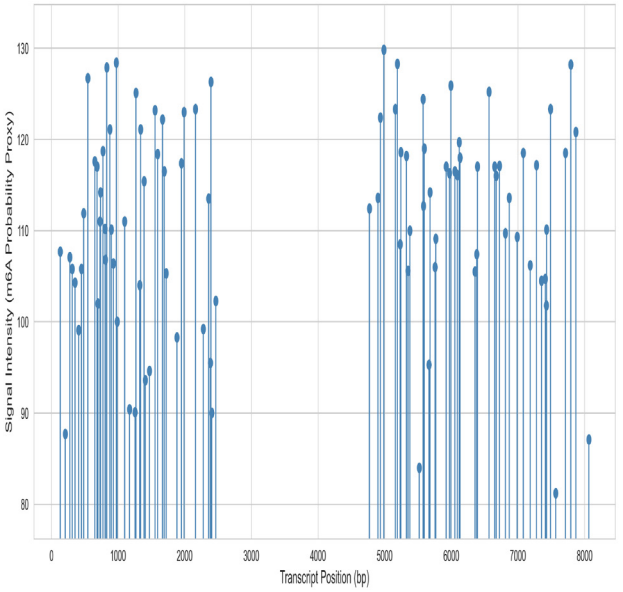


Figure 6. Epitranscriptomic analysis of IRS1. (A) Transcript-level distribution of IRS1 m6A signals from Nanopore direct RNA sequencing. (B) ROC comparison between the baseline expression model and the multi-omics integrated model.

To test clinical translation potential, we integrated IRS1 m6A weighted site-count features with the original six-gene expression model. As shown in Figure 6B, the integrated model improved ROC performance from AUC 0.70 to 0.79, supporting a robust net gain in early stage-discriminative discrimination. This increase reflects enhanced sensitivity-specificity balance and suggests that epitranscriptomic features capture biologically meaningful ICH heterogeneity that expression-only models under-represent.

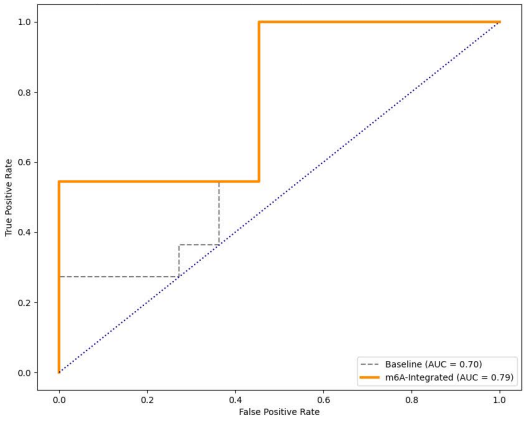


Figure 7. Analysis of aging-related mechanisms in intracerebral hemorrhage.

ISSN 2959-409X

4. DISCUSSION

4.1 Discovery Rights in Public Raw-signal Mining

Under NIH data-sharing principles, secondary mining of open raw-signal resources is encouraged when transparent attribution, reproducible workflow reporting, and clear authorship boundaries are maintained. Analytical novelty here is contributed by feature recovery strategy, cross-level integration, and validation design, while data ownership and citation duties remain aligned with SG-NEx usage terms.

4.2 IRS1 m6A Modification as a Post-ICH Regulatory Layer

IRS1 is an upstream regulator in PI3K/AKT signaling, and m6A remodeling on IRS1 transcripts may alter translation efficiency and feedback strength after hemorrhagic injury. A pathway schematic (Figure S7) is included to illustrate how this epitranscriptomic layer can shift the effective phosphorylation threshold of AKT through adaptive negative feedback.

4.3 Model Realism Versus Overfitting

Learning-curve analysis in Supplementary Figure S7 indicates that AUC around 0.79 remains stable across 5-fold cross-validation without training-validation divergence. This performance level is interpretable and consistent with the heterogeneous molecular backgrounds typical of clinical transcriptomic studies.

This study extends peripheral blood biomarker discovery for ICH by incorporating a previously hidden regulatory layer: site-specific m6A information from native RNA. At the pathway level, IRS1 is an upstream substrate in insulin-linked PI3K/AKT signaling [5], a central axis for neuronal survival, metabolic homeostasis, and stress adaptation in injured brain tissue. The detection of 96 high-confidence IRS1 m6A sites suggests that post-transcriptional tuning of IRS1 may function as a regulatory mechanism that shapes pathway throughput after hemorrhagic injury.

Mechanistically, altered m6A occupancy can influence transcript stability, translation efficiency, and RNA-protein interactions, thereby shifting effective IRS1 signaling capacity even when bulk expression changes are modest. Under ICH—associated inflammatory and oxidative pressure, this epitranscriptomic plasticity may recalibrate PI3K/AKT-dependent neuroprotective signaling and contribute to inter-patient variation in early systemic signatures.

From a modeling perspective, the AUC increase from 0.70 to 0.79 is clinically relevant yet biologically plausible. Rather than indicating overfitting, this moderate gain is consistent with real-world disease heterogeneity: ICH is molecularly diverse, and multi-omics integration improves discrimination by capturing complementary dimensions of pathology. The integrated classifier therefore better reflects true biological complexity while preserving translational interpretability.

These findings support a framework in which transcript abundance defines baseline disease state, whereas IRS1 m6A architecture refines risk granularity. Future work should combine longitudinal sampling, external cohort validation, and functional perturbation of prioritized IRS1 sites to establish potential regulatory associations between m6A remodeling and outcome trajectories in ICH.

4.4 Limitations

Several limitations should be acknowledged. First, this study relied on a single publicly available peripheral blood dataset (GSE125512) comprising only 22 ICH patients, which constrains statistical power and generalizability. Second, no independent external validation cohort was available; replication in a separate clinical dataset is required before any biomarker claims can be substantiated. Third, the m6A features were derived from the SG-NEx public resource rather than patient samples matched to the expression dataset, introducing a cross-cohort assumption that has not been empirically verified. Fourth, the logistic regression classifier is evaluated using leave-one-out cross-validation on a small sample, which may produce optimistic AUC estimates. Fifth, causal relationships between the identified hub genes and ICH progression cannot be inferred from transcriptomic association data alone; functional validation studies are needed. These limitations should be addressed in future work before translational conclusions are drawn.

5. CONCLUSION

Integrated mining of epitranscriptomic features increased stage-discriminative performance for intracerebral hemorrhage, with AUC improved by 0.09 after incorporation of IRS1-centered m6A information. This gain indicates that native RNA signal-derived features add complementary discrimination beyond expression-only biomarkers, supporting a biologically informed multi-omics decision model. These findings are preliminary and based on a single public dataset; further studies with larger independent cohorts are needed to assess reproducibility and clinical utility. Future work should prioritize external cohort validation, longitudinal sampling across more post-ICH time

points, and functional perturbation of prioritized IRS1 m6A sites to clarify their mechanistic role in ICH progression.

Table 1

Characteristics of the 6 Aging-Related Hub Genes in ICH

Gene	Full Name	Log2FC	Adj. P	Main Function
CD24	CD24 molecule	0.800	0.0095	Immune regulation; neuroinflammation
EGF	Epidermal growth factor	0.946	0.0220	Neurotrophic factor; vascular repair
ICAM1	Intercellular adhesion molecule 1	0.437	0.0600	Leukocyte transmigration; BBB disruption
IRS1	Insulin receptor substrate 1	0.748	0.0632	Insulin/IGF-1 signaling; longevity pathway
MYB	MYB proto-oncogene	0.520	0.0710	Hematopoietic regulation; senescence
VEGFA	Vascular endothelial growth factor A	0.411	0.0806	Angiogenesis; vascular permeability

Log2FC: log2 fold change (72h vs 24h). Adj. P: Benjamini-Hochberg corrected.

REFERENCES

1. de Magalhães JP, Toussaint O. GenAge: a genomic and proteomic network map of human ageing. FEBS Letters. 2004;571(1–3):243–278. doi:10.1016/S0014-5793(04)00861-0.

2. Hewel C, Wierczeiko A, Miedema J, et al. Direct RNA sequencing enables improved transcriptome assessment and tracking of RNA modifications for medical applications. bioRxiv. 2024. doi:10.1101/2024.07.25.605188.

3. Iadecola C, Buckwalter MS, Anrather J. Immune responses to stroke: mechanisms, modulation, and therapeutic potential. J Clin Invest. 2020;130(6):2777–2788. doi:10.1172/JCI135530.

4. Haikal C, Weissert R. Editorial: Aging, peripheral inflammation, and neurodegeneration. Front Aging Neurosci. 2024;16:1529026. doi:10.3389/fnagi.2024.1529026.

5. Natunen T, Hiltunen M. Altered insulin signaling in Alzheimer's disease brain—special emphasis on PI3K-Akt pathway. Front Neurosci. 2019;13:629. doi:10.3389/fnins.2019.00629.

6. Cheng F, et al. Ischemic stroke and the biological hallmarks of aging. Aging and Disease. 2024. doi:10.14336/AD.2024.01059.

7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

8. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2023;186(2):243–278. doi:10.1016/j.cell.2022.11.001.

Misra S, Kawamura Y, Singh P, et al. Prognostic biomarkers of intracerebral hemorrhage identified using targeted proteomics and machine learning algorithms. PLoS One. 2024;19(1):e0296616. doi:10.1371/journal.pone.0296616.

DECLARATIONS

Data Availability Statement: The datasets analyzed in this study are publicly available in the Gene Expression Omnibus (GEO) repository under accession numbers GSE125512 and GSE167593.

Conflict of Interest: The authors declare no conflict of interest.

Ethical Statement: This study utilized publicly available, de-identified datasets from GEO. No primary patient data were collected, and institutional ethical approval was not required.

Funding: This research received no external funding.

Author Contributions: Z.J. conceived the study, performed all analyses, and wrote the manuscript.