Current Science
CS 4 (4), 680– 690 (2024)
Received 26 June 2024 | **Revised 25** July 2024 | **Accepted 20** August 2024 | **Online Available 18 September** 2024
https://doi.org/10.52845/CS/2024-4-5-3

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

ORIGINAL ARTICLE

Molecular Signatures and Shared Pathways in Cerebral Cavernous Malformations and Ischemic Stroke: A Comprehensive Analysis

Guohua Liu[1,†], Gaofei Yan[2,4†]Shufang Luo[3], Mingkai Xia[4], Qian Yu[4], Haiying Guo[3,*]

[1]Sihui City People's Hospital, Sihui, Guangdong, China

[2]Hunan University of Medicine, Huaihua, Hunan, China

[3]Department of Respiratory and Critical Care Medicine, the Fifth Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China

[4]Zhangjiajie Hospital Affiliated to Hunan Normal University, Zhangjiajie, Hunan, China

Corresponding Author: Haiying Guo

Abstract:
This study elucidates the distinct molecular signatures and shared pathways involved in cerebral cavernous malformations (CCM) and ischemic stroke (IS) through a comprehensive gene expression analysis. Using differential gene expression profiling, we identified a significant number of unique and shared differentially expressed genes (DEGs) between CCM and IS. Weighted Gene Co-Expression Network Analysis (WGCNA) further highlighted significant gene modules associated with each disease, with a subset shared between them, underscoring common molecular mechanisms. Integrating intersecting DEGs and WGCNA-identified genes into a unified analysis, we pinpointed 107 unique genes crucial to disease pathophysiology. Functional enrichment underscored roles in immunity and signaling, corroborated by protein-protein interaction networks identifying key regulatory genes such as CD8A, CD19, CCR7, and IL7R. Advanced statistical methods, including LASSO regression and Boruta algorithm, refined these findings, revealing potential diagnostic markers with high discriminatory power through ROC analysis. Immune cell infiltration assessments highlighted altered immune dynamics, offering insights into CCM and IS mechanistic underpinnings. This integrated analysis enhances understanding of the molecular landscapes of CCM and IS, suggesting novel therapeutic targets.

Introduction

Significant cerebrovascular disorders characterized by complex pathophysiological processes. Both conditions lead to disrupted neurological functions and pose substantial clinical burdens(C et al., 2022); however, they exhibit distinct clinical features and involve partially understood, disparate molecular mechanisms. Unraveling these molecular pathways is crucial for developing more effective diagnostic tools and therapeutic interventions.

Previous studies have independently identified genes associated with CCM and IS(Subhash et al., 2019), yet the comparative molecular landscapes and shared pathobiology remain unexplored. Understanding common molecular mechanisms could illuminate novel therapeutic targets and offer insights into disease mechanisms, potentially revealing shared pathways susceptible to modulation.

In this study, we integrate differential gene expression analysis with advanced bioinformatics approaches, including Weighted Gene Co-Expression Network Analysis (WGCNA), functional enrichment, and network analysis. We aim to decipher the complex molecular networks of CCM and IS, identify key regulatory genes, and explore the immune landscape alterations. By employing both LASSO regression and Boruta algorithm, we fine-tune the identification of pivotal genes, assessing their diagnostic utility via Receiver Operating Characteristic (ROC) curve analysis. Through this integrative approach, we strive to contribute a comprehensive understanding of the shared and unique molecular architectures of CCM and IS, ultimately advancing potential therapeutic strategies.This study elucidates the distinct molecular signatures and shared pathways involved in cerebral cavernous malformations (CCM) and ischemic stroke (IS) through a comprehensive gene expression analysis. Using differential gene expression profiling, we identified a significant number of unique and shared differentially expressed genes (DEGs) between CCM and IS. Weighted Gene Co-Expression Network Analysis (WGCNA) further highlighted significant gene modules associated with each disease, with a subset shared between them, underscoring common molecular mechanisms. Integrating intersecting DEGs and WGCNA-identified genes into a unified analysis, we pinpointed 107 unique genes crucial to disease pathophysiology. Functional enrichment underscored roles in immunity and signaling, corroborated by protein-protein interaction networks identifying key regulatory genes such as CD8A, CD19, CCR7, and IL7R. Advanced statistical methods, including LASSO regression and Boruta algorithm, refined these findings, revealing potential diagnostic markers with high discriminatory power through ROC analysis. Immune cell infiltration assessments highlighted altered immune dynamics, offering insights into CCM and IS mechanistic underpinnings. This integrated analysis enhances understanding of the molecular landscapes of CCM and IS, suggesting novel therapeutic targets.

Methods

Sample Collection and Data Preprocessing

Gene expression data for this study were sourced from the GEO under the accession numbers GSE123968(Koskimäki et al., 2019), GSE130174(Lyne et al., 2019), and GSE16561(Barr et al., 2010). These datasets include expression data relevant to CCM and IS. The raw data were meticulously preprocessed, involving initial quality checks and normalization processes using the R package ‘limma’(Ritchie et al., 2015). Batch effects, which could obscure true biological signals, were corrected using the ‘ComBat_seq’ function from the ‘sva’ package to ensure robust integration of data from different cohorts. Additionally, expression values were log-transformed and filtered to retain informative genes for downstream analyses.

Differential Gene Expression Analysis

DEGs were identified separately for CCM and IS using the 'limma-voom' method (Ritchie et al., 2015). For CCM, the combined and batch-corrected dataset was used, while for IS, the GSE16561 dataset was analyzed independently. Genes with an adjusted p-value < 0.05 and absolute log2 fold change > 0.5 were considered significant. The resulting DEGs were visualized using volcano plots and heatmaps generated with the 'ggplot2' and 'pheatmap' packages, respectively.

Identification of Common DEGs

To identify shared molecular mechanisms between CCM and IS, we performed an intersection analysis of the DEGs identified in each condition. The overlap between CCM and IS DEGs was visualized using a Venn diagram created with the 'ggvenn' package.

WGCNA

We applied the 'WGCNA' R package to perform a weighted gene co-expression network analysis, aiming to identify clusters of co-expressed genes(Langfelder & Horvath, 2008). The scale-free topology criterion guided the selection of the soft thresholding power, which was used to construct the network. Modules were delineated through dynamic tree cutting, and their eigengenes were subsequently correlated with clinical traits to pinpoint modules significantly associated with these traits(Zhang & Horvath, 2005).

Gene Functional Annotation

Functional annotation of the identified DEIGs was carried out using the Database for Annotation, Visualization, and Integrated Discovery (DAVID)(Huang et al., 2009). Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analyzed to understand the biological processes and pathways involved(Ashburner et al., 2000; Kanehisa &

Goto, 2000).

Lasso Regression and Feature Selection

We conducted a feature selection process using Lasso regression on two datasets, CCM and IS, aiming to identify key gene predictors. Lasso regression is utilized due to its capability to perform variable selection and regularization to improve the prediction accuracy and interpretability of the statistical model it produces. The selected features were analyzed through the lambda error plot to determine the optimal lambda value, and the minimal gene set was saved for subsequent analysis.

Random Forest and Boruta Analysis

For both CCM and IS datasets, we applied Random Forest analysis combined with the Boruta algorithm to identify important gene features. Boruta is an all-relevant feature selection method that enhances the robustness of Random Forest to discern whether a feature is important. Genes confirmed by Boruta were considered significant and were saved as key gene features.

Immune Infiltration Analysis

To estimate immune cell infiltration proportions within the samples, we applied the CIBERSORT algorithm. This analysis utilized the LM22 signature matrix to deconvolute the expression data from both the CCM and IS datasets. The CIBERSORT results were then merged with the sample annotation data, allowing for a comparison between control and disease states. The outcomes were visualized using boxplots to depict the relative proportions of different immune cell types and statistically assessed using t-tests, differentiating between control and disease sample groups.

Statistical Analysis

Statistical analyses were performed using R software. P-values < 0.05 were considered statistically significant.

Results

Differential Gene Expression Analysis Reveals Distinct Molecular Signatures in CCM and IS

In this study, we analyzed the differential gene expression profiles in CCM and IS, compared against respective controls. Our CCM analysis (Figure 1A, Supplementary Table 1) identified a substantial number of DEGs. The heatmap (Figure 1C) further validates these findings by showing distinct clustering of CCM samples as compared to controls, revealing significant patterns of gene upregulation and downregulation. Similarly, in IS, a marked difference in gene expression profiles was observed compared to controls (Figure 1B and 1D, Supplementary Table 2). The corresponding heatmap (Figure 1D) confirms these observations, showcasing clear distinctions in gene expression that differentiate IS samples from controls.

To investigate potentially shared molecular mechanisms between CCM and IS, we compared DEGs identified in each condition. The Venn diagram (Figure 1E) depicts the overlap between CCM and IS DEGs, revealing that 3,378 DEGs (87.2%) are unique to CCM, while 394 DEGs (10.2%) are exclusive to IS. Notably, 102 DEGs (2.6%) are common to both conditions, indicating possible shared molecular pathways that may contribute to the pathogenesis of both diseases.

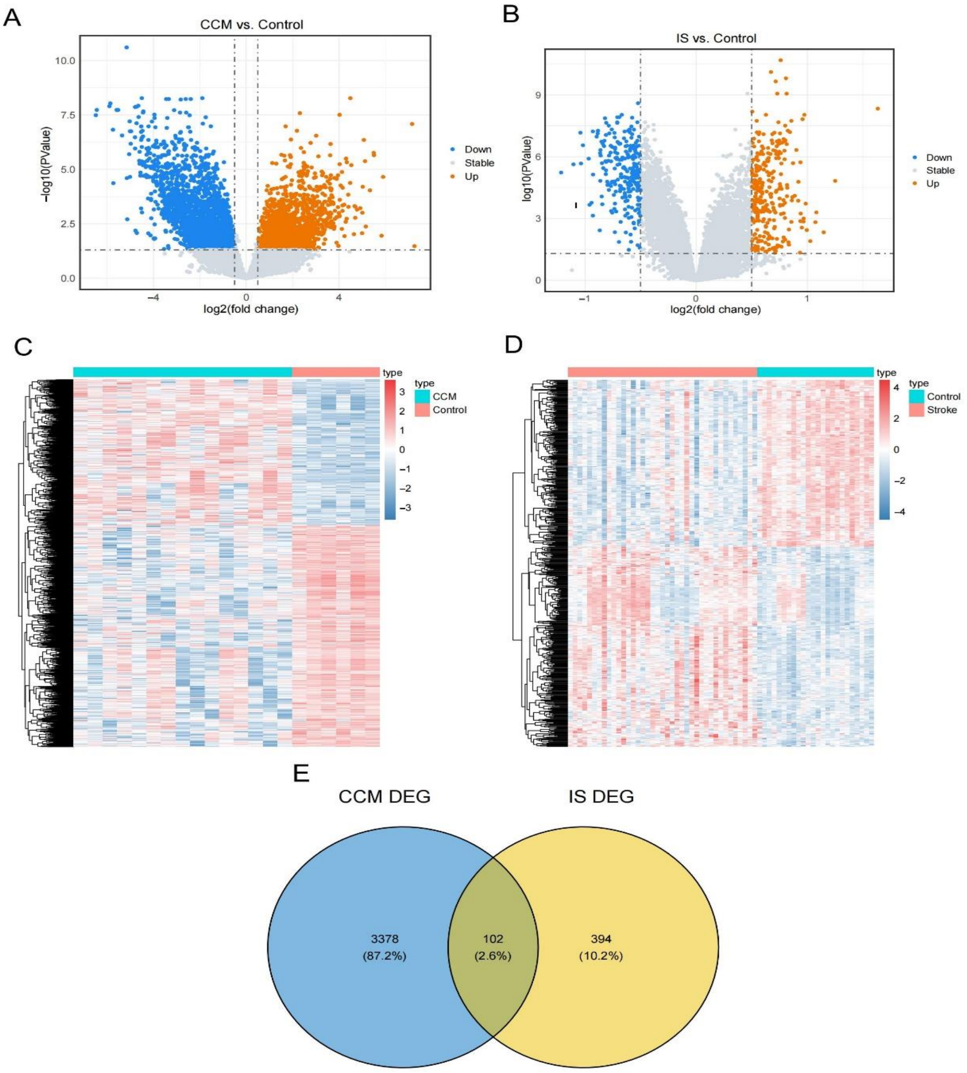

Figure 1. Differential Gene Expression Profiles in CCM and IS

A: Volcano plot of differentially expressed genes in CCM compared to controls. Orange dots represent significantly upregulated genes, while blue dots represent significantly downregulated genes. The x-axis represents log2 fold change, and the y-axis represents -log10(P-value).

B: Volcano plot of differentially expressed genes in IS compared to controls. Orange dots represent significantly upregulated genes, while blue dots represent significantly downregulated genes. The x-axis represents log2 fold change, and the y-axis represents -log10(P-value).

C: Heatmap showing gene expression differences between CCM samples and controls. Red indicates upregulated genes, while blue indicates downregulated genes.

D: Heatmap showing gene expression differences between IS samples and controls. Red indicates upregulated genes, while blue indicates downregulated genes.

E: Venn diagram showing the overlap of differentially expressed genes between CCM and IS. The blue circle represents genes unique to CCM, the yellow circle represents genes unique to IS, and the overlap indicates genes common to both conditions.

Weighted Gene Co-Expression Network Analysis Identifies Key Gene Modules in CCM and IS

We employed WGCNA to further explore the molecular mechanisms in CCM and IS, aiming to identify disease-associated gene modules. WGCNA revealed gene co-expression networks for both CCM (Figure 2A) and IS (Figure 2B). Dendrograms illustrated gene divisions into distinct modules, each represented by unique colors; substantial differences in gene co-expression patterns between the two diseases were evident.

Further analysis related gene modules to disease states. In CCM, the dark cyan, dark turquoise, and light green modules were significantly negatively correlated with disease status, while magenta and blue modules showed positive correlations (Figure 2C). Conversely, in IS, dark red, blue, and magenta modules were positively correlated, whereas grey60 and yellow modules were negatively correlated (Figure 2D). The biological relevance of these modules was validated by assessing the correlation between module membership and gene significance. A strong correlation of 0.94 (p=1.3e-149) was found in the yellow module of CCM, and a correlation of 0.71 (p=1.2e-13) in the green-yellow module of IS. These findings suggest consistent gene expression patterns within these modules under disease conditions.

A comparative analysis was performed to identify shared gene modules between CCM and IS. The Venn diagram (Figure 2G) illustrated that 626 modules (47.9%) are unique to CCM, 665 modules (50.9%) are unique to IS, and 15 modules (1.1%) are common to both diseases, potentially representing shared molecular mechanisms.

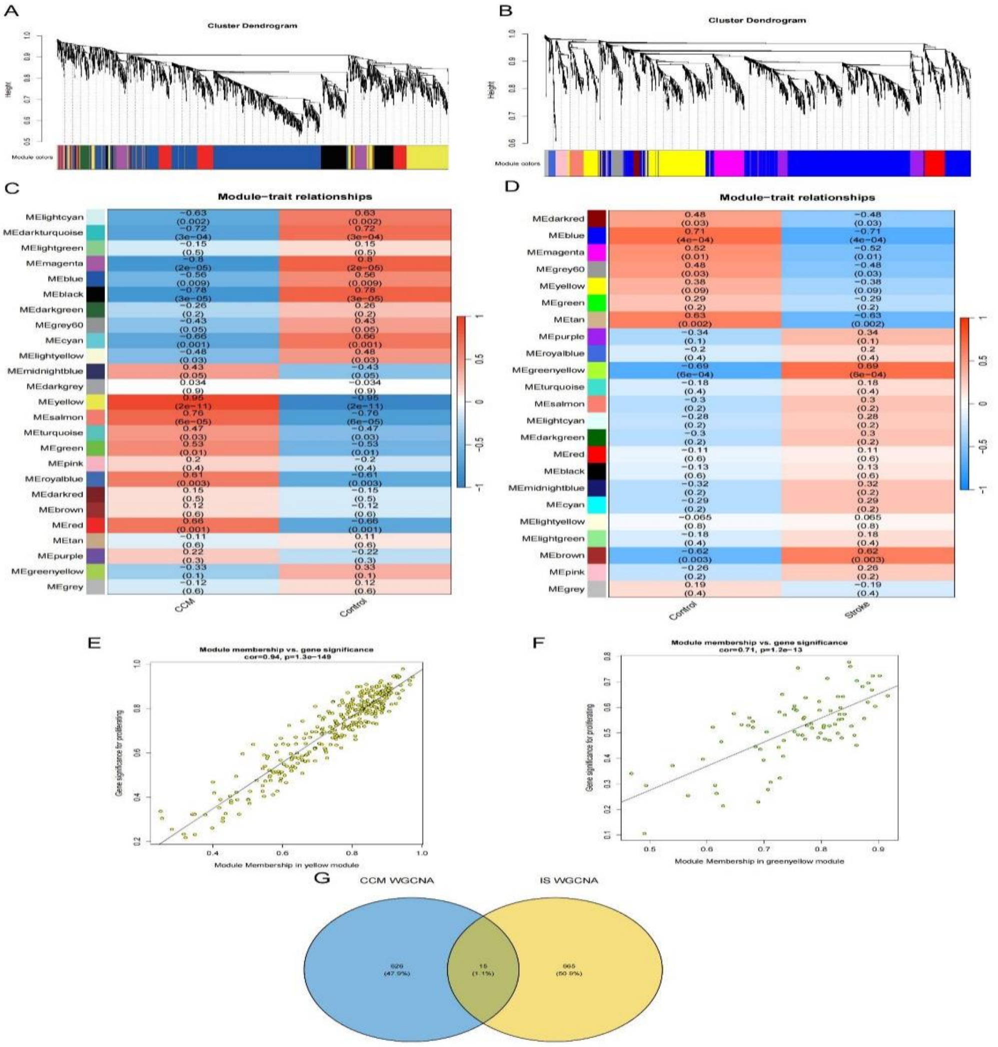

Figure 2: Co-Expression Network and Module Correlation Analysis in CCM and IS

A: Dendrogram of gene co-expression networks in CCM samples. Genes are organized into distinct modules, each represented by a different color.

B: Dendrogram of gene co-expression networks in IS samples. Genes are organized into distinct modules, each represented by a different color.

C: Heatmap of gene module correlations with disease status in CCM. Dark cyan, dark turquoise, and light green modules exhibit significant negative correlations, while magenta and blue modules show positive correlations.

D: Heatmap of gene module correlations with disease status in IS. Dark red, blue, and magenta modules exhibit significant positive correlations, while grey60 and yellow modules show negative correlations.

E: Scatter plot showing correlation between gene significance and module membership in the yellow module of CCM, with a correlation of 0.94 (p=1.3e-149).

F: Scatter plot showing correlation between gene significance and module membership in the green-yellow module of IS, with a correlation of 0.71 (p=1.2e-13).

G: Venn diagram illustrating shared gene modules between CCM and IS.

Functional Enrichment and Network Analysis of DEG and WGCNA Gene Unions

By integrating 15 intersecting WGCNA genes with 102 intersecting DEGs, we identified 107 unique genes, forming the DEG-WGCNA gene union. This union provides insights into potential key regulators of CCM and IS pathophysiology. Enriched GO terms indicated involvement in immune response regulation, signal transduction, and cell proliferation (Figure 3A, Supplementary Table 3). KEGG pathway analyses revealed pathways associated with the immune system and intracellular signaling cascades (Figure 3B, Supplementary Table 4). Further biological insights were derived by constructing a PPI network using STRING data (Figure 3C, Supplementary Table 5), which identified ten pivotal genes, including CD8A, CD19, CCR7, and IL7R, illustrating their key roles in protein interactions and cellular processes associated with disease.

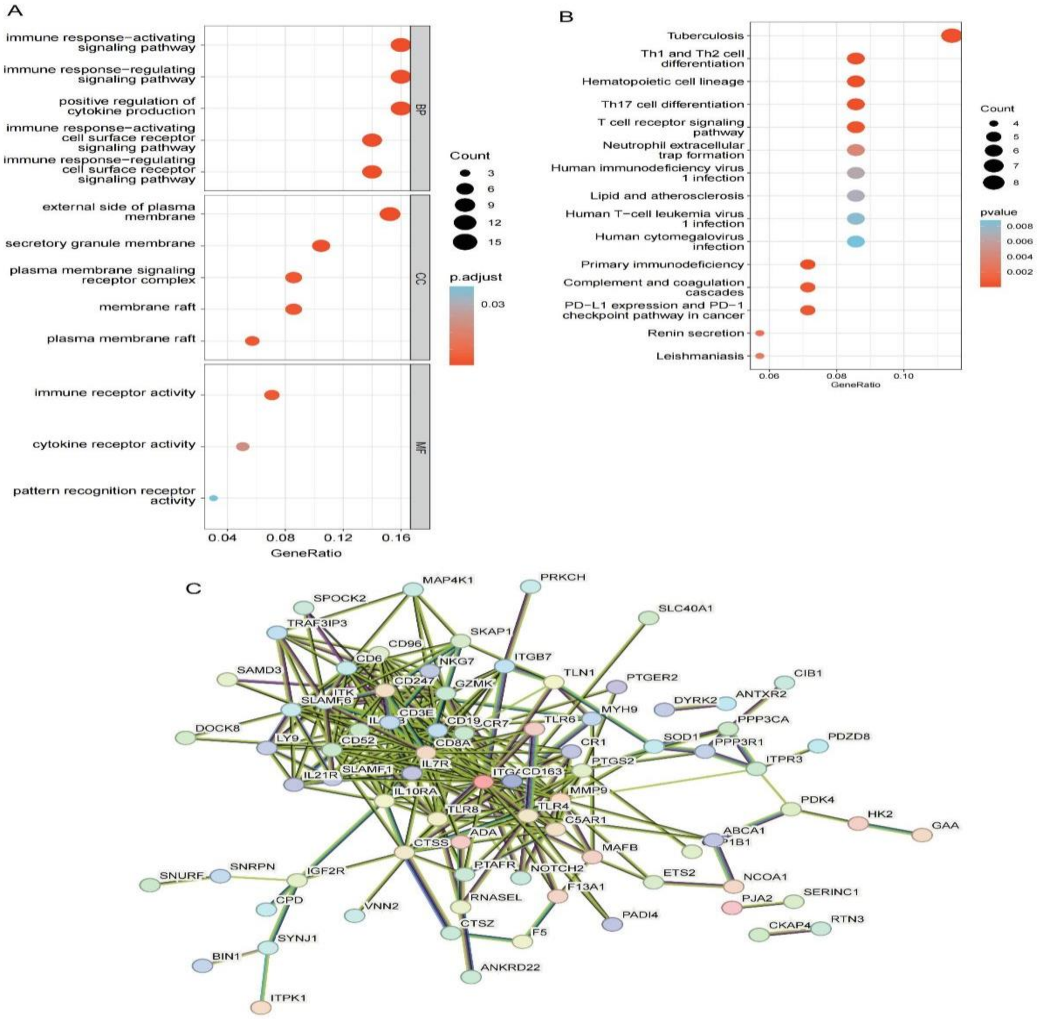

Figure 3:Functional Enrichment and Network Analysis of DEG and WGCNA Gene Union

A: Gene Ontology enrichment analysis of DEG and WGCNA gene union, indicating roles in immune response regulation, signal transduction, and cell proliferation.

B: KEGG pathway analysis of DEG and WGCNA gene union, highlighting pathways related to the immune system and intracellular signaling cascades.

C: PPI network constructed using STRING data for the DEG and WGCNA gene union. Ten pivotal genes, including CD8A, CD19, CCR7, and IL7R, are highlighted as hub nodes in the network.

Identification of Key Genes in CCM and IS Using LASSO and Boruta Algorithms

To identify key genes associated with CCM and IS, we applied LASSO regression and the Boruta algorithm. In CCM, LASSO analysis revealed optimal Lambda values, identifying critical genes such as CCR7, CD19, and IL2RB (Figures 4A and 4B). A similar analysis for IS identified key genes at an optimal Lambda of 0.04047 (Figures 4C and 4D). The Boruta algorithm corroborated these results, highlighting significant genes like CD8A, GZMK, and IL7R (Figures 4E and 4F). The Venn diagram (Figure 4G) demonstrated overlap in key genes identified by both methods; notably, two genes were consistently recognized across all analyses as key in both CCM and IS, suggesting their significant roles in the diseases' pathophysiology.

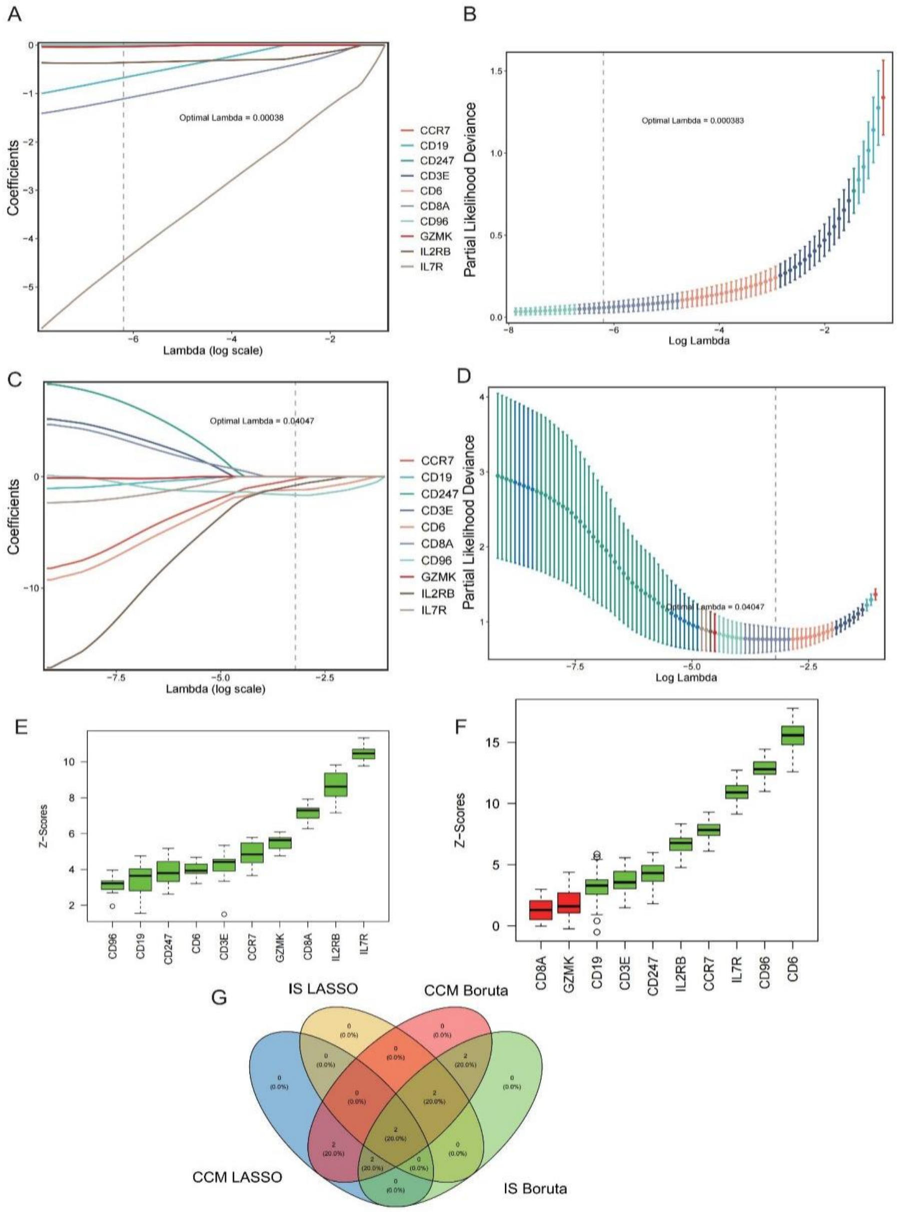

Figure 4: Key Gene Identification in CCM and IS Using LASSO and Boruta Methods

A: Coefficients of genes in CCM as a function of lambda during LASSO regression analysis. Optimal lambda was determined to be 0.00038.

B: Partial likelihood deviance versus log(lambda) in CCM, confirming the optimal lambda at 0.000383.

C: Coefficients of genes in IS as a function of lambda during LASSO regression analysis. Optimal lambda was determined to be 0.04047.

D: Partial likelihood deviance versus log(lambda) in IS, confirming the optimal lambda at 0.04047.

E: Z-scores of key genes identified in CCM using the Boruta algorithm.

F: Z-scores of key genes identified in IS using the Boruta algorithm.

G: Venn diagram showing overlap of key genes in CCM and IS identified through LASSO and Boruta methods.

ROC Curve and Gene Expression Analysis of Key Genes

The potential diagnostic value of identified key genes in CCM and IS was assessed using ROC curve analysis. Additionally, we examined the expression levels of these genes in various states to further understand their biological significance. For CCM diagnosis, the IL2RB gene demonstrated an AUC of 0.989, indicating nearly perfect discriminatory power. The CD6 gene, although slightly less effective, still showed substantial diagnostic capability with an AUC of 0.867 (Figure 5A). In IS diagnosis, CD6 and IL2RB genes exhibited high diagnostic efficacy with AUCs of 0.931 and 0.841, respectively (Figure 5B). Gene expression analysis further supported these findings, showing significantly higher expression levels of IL2RB and CD6 in the CCM group compared to controls (Figure 5C), as well as elevated expressions in the IS group (Figure 5D).

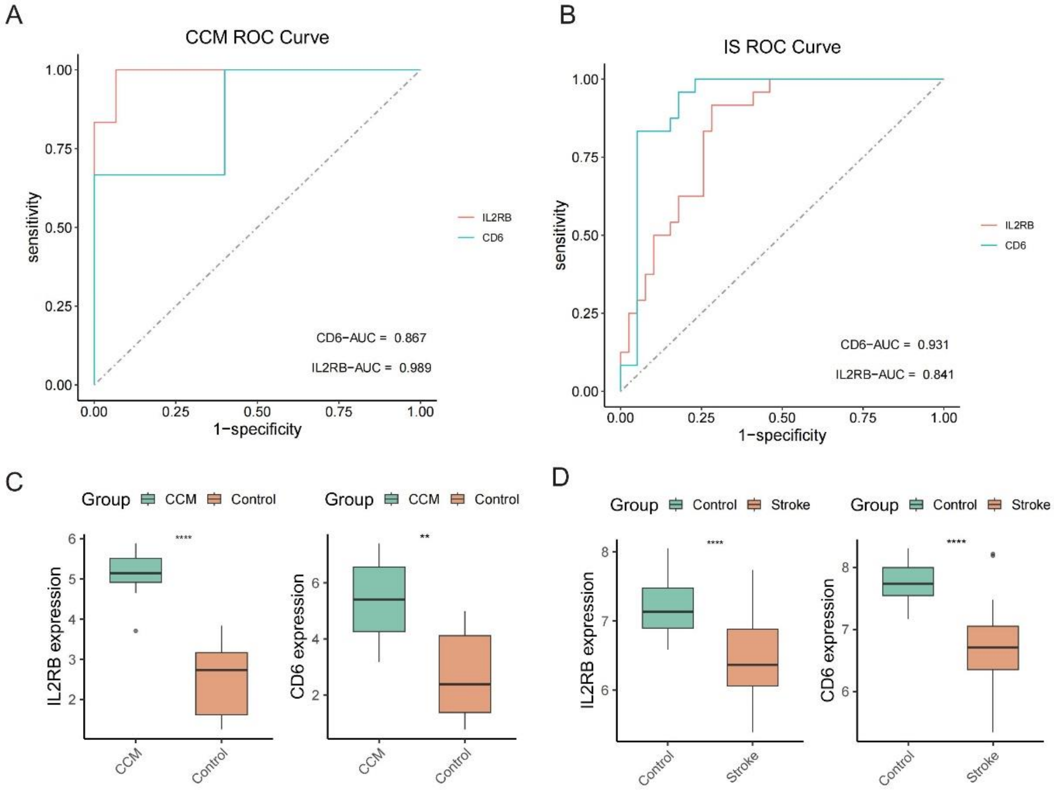

Figure 5: Diagnostic Potential and Expression Analysis of Key Genes in CCM and IS

A: ROC curve for IL2RB and CD6 in diagnosing CCM. The AUC for IL2RB is 0.989, and for CD6 is 0.867.

B: ROC curve for IL2RB and CD6 in diagnosing IS. The AUC for CD6 is 0.931, and for IL2RB is 0.841.

C: Box plots showing the expression levels of IL2RB and CD6 in CCM and control groups.

Both genes are significantly upregulated in CCM.

D: Box plots showing the expression levels of IL2RB and CD6 in IS and control groups. Both genes are significantly upregulated in IS.

Immune Cell Infiltration Analysis

To delve deeper into the roles of IL2RB and CD6 genes in CCM and IS, we analyzed immune cell infiltration in the two cohorts. In the CCM group, activated NK cells and CD8 T cells showed significant proportion increases (Figure 6A, Supplementary Table 6), with activated NK cells and CD8 T cells substantially more prominent than in the control group. Additionally, there was an increase in monocyte and neutrophil proportions, though not reaching statistical significance. In IS, monocyte and neutrophil proportions were significantly elevated (Figure 6B, Supplementary Table 7), alongside increased levels of activated NK cells and CD4 memory T cells. These findings suggest that IL2RB and CD6 may exert their effects on CCM and IS pathophysiology through modulation of specific immune cell activities. The increase of activated NK cells and CD8 T cells in CCM might be associated with the high expression of IL2RB and CD6, whereas the prominent monocyte and neutrophil presence in IS indicates these genes may play crucial roles in inflammatory responses and immune regulation.

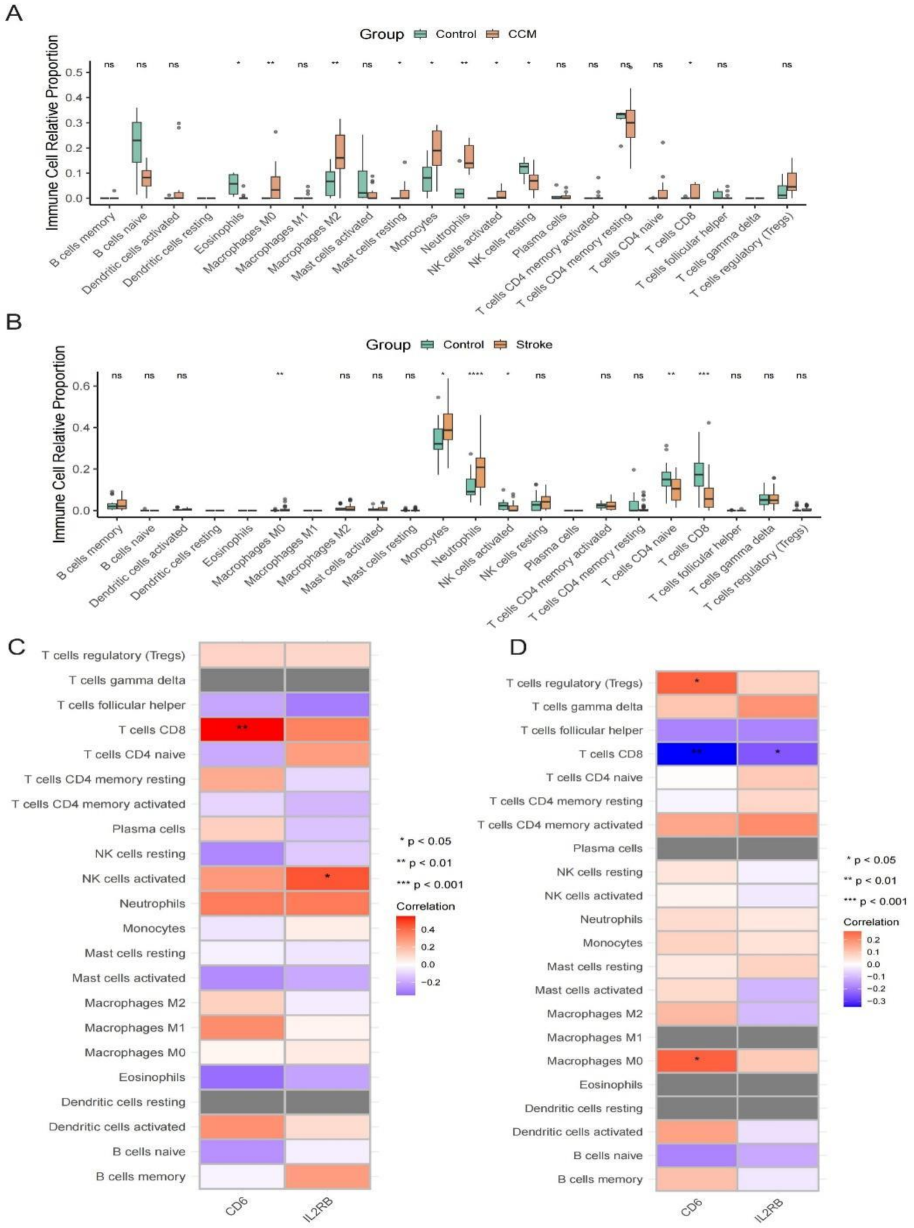

Figure 6: Immune Cell Infiltration and Correlation with Key Genes in CCM and IS

A: Relative proportions of different immune cell types in CCM. Activated NK cells and CD8 T cells show significant increases compared to controls.

B: Relative proportions of different immune cell types in IS. Monocytes and neutrophils show significant increases compared to controls.

C: Heatmap showing the correlation between the expression of IL2RB and CD6 genes and various immune cell types in the CCM group.

D: Heatmap showing the correlation between the expression of IL2RB and CD6 genes and various immune cell types in the IS group.

Discussion

This study extensively characterizes the molecular landscapes of CCM and IS, highlighting both unique and shared features through comprehensive gene expression profiling and network analysis. The discovery of distinct DEGs across both conditions underscores the specific molecular environments and intrinsic pathophysiological processes underpinning these cerebrovascular diseases. These distinctions are crucial, as they offer potential insights into tailored therapeutic approaches specific to each condition.

However, the identification of shared DEGs points towards underlying common pathways that may influence similar clinical manifestations or contribute to shared risk factors. These overlapping pathways, particularly those involved in vascular integrity and immune response, suggest that some therapeutic strategies might benefit both CCM and IS, potentially offering dual benefits in disease management.

The use of WGCNA allowed for the identification of disease-associated gene modules, highlighting gene networks that exhibit coordinated changes in expression. These modules illustrate complex interactions and biological processes specific to each disease, while also pointing out shared motifs that could be crucial in understanding their pathophysiological overlaps.

Functional enrichment and PPI network analyses pinpoint immune regulation and signal transduction as pivotal elements in both diseases. Hub genes such as CD8A, CD19, CCR7, and IL7R emerge as central figures in these pathways, representing potential biomarkers for disease progression or targets for novel therapeutic interventions.

Utilization of LASSO and Boruta algorithms for refining key gene identification ensures robustness in our findings, with ROC analysis affirming the potential diagnostic power of genes like IL2RB and CD6. These genes not only aid in distinguishing between disease states but could also guide the development of predictive models for patient stratification.

The immune cell infiltration analysis sheds light on the altered immune landscape in CCM and IS, with increased NK and T cells observed in CCM, and a heightened presence of monocytes and neutrophils in IS. These findings emphasize the significant role of immune mechanisms, suggesting that targeting immune responses might offer therapeutic benefits.

Overall, our study advances the understanding of CCM and IS at a molecular level, providing a foundation for future research focused on functional validation and the exploration of shared and distinct pathways. This knowledge can pave the way for innovative therapeutic strategies, potentially improving clinical outcomes for individuals affected by these challenging cerebrovascular disorders.

Conclusions

Our integrative approach delineates the molecular landscapes of CCM and IS, emphasizing the critical roles of immune response and signaling mechanisms. Key genes and shared pathways identified offer insights into disease pathogenesis and potential therapeutic strategies. Further validation and exploration of these findings, particularly through functional studies, could lead to the development of targeted interventions, improving outcomes for CCM and IS patients.

References

1. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology:

Tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

2. Barr, T. L., Conley, Y., Ding, J., Dillman, A., Warach, S., Singleton, A., & Matarin, M. (2010). Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling. *Neurology*, *75*(11),1009 –1014. https://doi.org/ 10.1212/WNL.0b013e3181f2b37f
3. C, Q., S, Y., Yh, C., H, Z., Xw, P., L, C., Lq, Z., M, C., Ds, T., & W, W. (2022). Signaling pathways involved in ischemic stroke: Molecular mechanisms and therapeutic interventions. *Signal Transduction and Targeted Therapy*, *7*(1). https://doi.org/10.1038/41392-022-01064-1
4. Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211
5. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27
6. Koskimäki, J., Girard, R., Li, Y., Saadat, L., Zeineddine, H. A., Lightle, R., Moore, T., Lyne, S., Avner, K., Shenkar, R., Cao, Y., Shi, C., Polster, S. P., Zhang, D., Carrión-Penagos, J., Romanos, S., Fonseca, G., Lopez-Ramirez, M. A., Chapman, E. M., … Awad, I. A. (2019). Comprehensive transcriptome analysis of cerebral cavernous malformation across multiple species and genotypes. *JCI Insight*, *4*(3), e126167. https://doi.org/10.1172/jci.insight.126167
7. Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559. https://doi.org/10.1186/1471-2105-9-559
8. Lyne, S. B., Girard, R., Koskimäki, J., Zeineddine, H. A., Zhang, D., Cao, Y., Li, Y., Stadnik, A., Moore, T., Lightle, R., Shi, C., Shenkar, R., Carrión-Penagos, J., Polster, S. P., Romanos, S., Akers, A., Lopez-Ramirez, M., Whitehead, K. J., Kahn, M. L., … Awad, I. A. (2019). Biomarkers of cavernous angioma with symptomatic hemorrhage. *JCI Insight*, *4*(12), e128577, 128577. https://doi.org/10.1172/jci.insight.128577
9. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007
10. Subhash, S., Kalmbach, N., Wegner, F., Petri, S., Glomb, T., Dittrich-Breiholz, O., Huang, C., Bali, K. K., Kunz, W. S., Samii, A., Bertalanffy, H., Kanduri, C., & Kar, S. (2019). Transcriptome-wide Profiling of Cerebral Cavernous Malformations Patients Reveal Important Long noncoding RNA molecular signatures. *Scientific Reports*, *9*(1), 18203. https://doi.org/10.1038/s41598-019-54845-0
11. Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*, Article 17. https://doi.org/10.2202/1544-6115. 1128