Journal of Biosciences and Medicines, 2026, 14(5), 340-359
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

Identification of Immune-Related Diagnostic Biomarkers and Construction of a Combined Predictive Model for Intracranial Aneurysm Based on Bioinformatics Analysis

Xiaochi Jiao[1,2], Yingying Huang[2], Xiang Qin[2], Shengliang Wei[2], Huadong Huang[1*], Yixia Yin[3*]

[1]Department of Neurosurgery, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China
[2]Graduate School, Youjiang Medical University for Nationalities, Baise, China
[3]Department of Gastroenterology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China
Email: *15577687387@163.com

How to cite this paper: Jiao, X.C., Huang, Y.Y., Qin, X., Wei, S.L., Huang, H.D. and Yin, Y.X. (2026) Identification of Immune-Related Diagnostic Biomarkers and Construction of a Combined Predictive Model for Intracranial Aneurysm Based on Bioinformatics Analysis. *Journal of Biosciences and Medicines*, **14**, 340-359.
https://doi.org/10.4236/jbm.2026.145024

Received: April 12, 2026
Accepted: May 23, 2026
Published: May 26, 2026

Abstract

Objective: This study aimed to systematically identify core immune-related genes involved in the development and progression of intracranial aneurysm (IA) through bioinformatics analysis, validate their differential expression in an independent dataset, and construct a precise diagnostic prediction model. The findings are intended to provide a theoretical basis for the early non-invasive diagnosis of IA, enrich the understanding of its immune-inflammatory pathological mechanisms, and offer potential intervention targets and novel research directions for clinical diagnosis and treatment. **Methods:** The GSE75436 dataset was downloaded from the Gene Expression Omnibus (GEO) database as the training set. Differentially expressed genes (DEGs) were screened using the "limma" package in R software. Immune-related differentially expressed genes (IM-DEGs) were identified by cross-referencing with the ImmPort database. Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed to elucidate the biological functions and signaling pathways associated with the IM-DEGs. To visualize protein-protein interactions (PPI), STRING, Cytoscape, and the CytoHubba plugin were utilized to construct a PPI network and identify hub genes. The GSE54083 dataset was then downloaded as an independent validation set to verify the expression of the core genes. Genes exhibiting statistically significant differential expression in both the training and validation sets were selected to construct an IA diagnostic model based on a linear prediction algorithm. The diagnostic performance of the model was assessed using receiver operating characteristic (ROC) curves in both datasets,

*Co-corresponding authors.

and correlation analysis of the core genes was conducted to reveal their synergistic regulatory relationships. **Results:** A total of 173 IM-DEGs were identified from the intersection of 757 DEGs in the training set GSE75436 and the immune gene list. These genes were predominantly enriched in biological processes such as leukocyte migration, leukocyte-mediated immunity, cell chemotaxis, and positive regulation of cytokine production. Analysis using Cytoscape software identified the top 10 hub genes with the highest scores: IL1B, TNF, CXCL8, CD8A, TYROBP, CCR1, CXCL10, CCL4, CCR5, and CCL20. Validation using the independent dataset GSE54083 confirmed that three hub genes (CXCL8, TNF, and CD8A) exhibited strong diagnostic potential. Expression validation revealed significant differences in the relative expression of these three genes between the control and aneurysm groups ($P < 0.05$). Specifically, CXCL8 was highly expressed in the aneurysm group ($P < 0.001$), whereas TNF and CD8A were expressed at low levels ($P < 0.001$ and $P < 0.05$, respectively), suggesting their potential role in the immune-inflammatory mechanisms of IA. ROC curve analysis demonstrated that the combined diagnostic model based on CXCL8, TNF, and CD8A achieved an area under the curve (AUC) of 0.947 (95% CI: 0.862 - 1.000) in the training set and an AUC of 0.754 (95% CI: 0.546 - 0.962) in the validation set, indicating favorable diagnostic performance. **Conclusion:** By integrating GEO data mining with independent dataset cross-validation, this study identified CXCL8, TNF, and CD8A as key genes associated with the immune-inflammatory pathogenesis of intracranial aneurysm. A combined prediction model constructed from these genes demonstrated robust diagnostic efficacy, thereby providing a novel combination of biomarkers and a theoretical foundation for the early non-invasive diagnosis and immune-targeted therapy of IA.

Keywords

Intracranial Aneurysm, Bioinformatics Analysis, Immune-Related Genes, Diagnostic Model

1. Introduction

Intracranial aneurysm (IA) is a prevalent cerebrovascular disorder characterized by localized pathological dilation of the cerebral arterial wall. Although the majority of aneurysms remain clinically silent prior to rupture, the consequent aneurysmal subarachnoid hemorrhage (aSAH) represents a catastrophic event, carrying a mortality rate of 30% - 50% and a substantial risk of long-term neurological disability. This imposes a considerable burden on patients, their families, and healthcare infrastructure worldwide [1]. Current epidemiological estimates indicate a global IA prevalence of approximately 1% - 2%, with an annual rupture incidence of roughly 16.4 cases per 100,000 individuals, underscoring its significance as a pressing public health concern [2]. Consequently, elucidating the underlying pathogenesis and developing robust strategies for early detection and risk stratification are of paramount clinical importance for improving patient outcomes and advancing precision medicine.

At present, the clinical diagnosis of IA relies heavily on advanced neuroimaging modalities, including digital subtraction angiography (DSA), computed tomography angiography (CTA), and magnetic resonance angiography (MRA). More recently, the application of artificial intelligence (AI) in medical image analysis has opened new avenues for automated aneurysm detection and segmentation. Several deep learning-based algorithms have demonstrated diagnostic performance comparable to that of expert human readers in assisting interpretation [3] [4]. Nevertheless, these techniques fundamentally focus on the morphological identification of established aneurysms and are inherently limited in their capacity to provide molecular-level warnings before aneurysm formation or a critical elevation in rupture risk occurs. Furthermore, current therapeutic interventions—primarily surgical clipping and endovascular coiling—are designed to manage existing lesions rather than prevent disease initiation or progression. This gap has driven increasing research interest toward the molecular underpinnings of IA, with the goal of identifying novel biomarkers suitable for early warning and individualized risk assessment.

Advances in high-throughput technologies, particularly gene microarray and next-generation sequencing platforms, have propelled molecular investigations across a broad spectrum of diseases, including cerebrovascular disorders, malignancies, and autoimmune conditions. These methodologies have not only deepened our mechanistic understanding of disease pathogenesis but also furnished efficient tools and analytical frameworks for screening core genetic signatures, constructing diagnostic prediction models, and evaluating prognosis [5].

The integration of bioinformatics analyses with independent dataset cross-validation has emerged as a robust and reliable strategy for identifying pivotal disease-associated genes while mitigating confounding effects arising from platform heterogeneity and sample variation [6]. By mining publicly accessible gene expression repositories and strategically focusing on immune-related gene subsets, researchers can more precisely delineate the regulatory influence of the immune-inflammatory microenvironment on disease evolution.

In the present study, we performed a systematic bioinformatics analysis of IA-associated gene expression microarray data retrieved from the Gene Expression Omnibus (GEO) repository. We constructed a protein-protein interaction (PPI) network to pinpoint potential core immune-related molecular targets and subsequently validated their differential expression and diagnostic utility using an independent dataset. Moreover, a combined diagnostic model incorporating the verified core genes was established to evaluate its discriminatory capacity between IA patients and healthy controls. This investigation seeks to elucidate the immune-inflammatory mechanisms driving IA pathogenesis, identify promising non-invasive diagnostic biomarker panels, and provide a theoretical foundation for exploring novel therapeutic avenues.

2. Materials and Methods

2.1. Data Acquisition

The GEO database was queried using the search terms “intracranial aneurysm” and

"Homo sapiens" to retrieve publicly available gene expression profiles. The dataset GSE75436, generated on the GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array), was designated as the training cohort. This dataset comprised gene expression array data from 10 intracranial aneurysm tissue specimens and 10 matched superficial temporal artery control samples. An independent validation cohort, GSE54083, was acquired from the GPL4133 platform (Agilent-014850 Whole Human Genome Microarray) and contained peripheral blood mononuclear cell (PBMC) expression profiles from 13 patients with intracranial aneurysm (disease group) and 13 healthy control individuals (control group). A comprehensive catalog of 3118 human immune-related genes was obtained from the ImmPort database.

2.2. Data Preprocessing

Given the inherent differences in microarray platforms between the training set (GSE75436, Affymetrix) and the validation set (GSE54083, Agilent), preprocessing workflows were adapted to accommodate platform-specific characteristics while maintaining consistency in core analytical steps. All preprocessing procedures were executed using R software (version 4.3.1) for both the GSE75436 and GSE54083 datasets. The workflow comprised the following sequential steps:

Data Import: For the training set GSE75436, raw CEL files were imported utilizing the "affy" package in R to obtain the raw expression matrix, along with corresponding sample and gene expression information. For the validation set GSE54083, the processed data matrix provided by the platform was directly imported using the "GEOquery" package; this matrix had already undergone background correction via Agilent Feature Extraction software.

Background Adjustment: For the training set GSE75436, the Robust Multi-array Average (RMA) algorithm was applied to perform background correction, thereby minimizing noise introduced during chip hybridization. For the validation set GSE54083, background correction had been pre-performed by the data provider; therefore, this step was not repeated.

Quantile Normalization: The "limma" package was employed to conduct quantile normalization on the corrected or processed expression matrices for both datasets separately. This step standardized the distribution of gene expression values across all samples, mitigating systematic biases and ensuring inter-sample comparability [7].

Probe Annotation and Handling of Multi-Probe Mappings: Microarray probe identifiers were mapped to official gene symbols. Probes lacking a corresponding gene symbol or matching multiple genes were discarded. In instances where multiple probes corresponded to the same gene symbol, the median expression value across those probes was calculated and assigned as the representative expression value for that gene. Additionally, genes exhibiting expression levels within the lowest 25th percentile across all samples were filtered out to eliminate low-abundance transcripts that could introduce analytical noise.

Sample Quality Control: Principal component analysis (PCA) was utilized to detect potential outliers. Any sample displaying substantial deviation from other

samples within the same experimental group was flagged and removed to uphold data integrity. Following PCA-based quality assessment, no significant outlier samples were identified in either the training set GSE75436 or the validation set GSE54083; therefore, all samples were retained for subsequent analyses.

2.3. Identification of Differentially Expressed Genes

The "limma" package was applied to screen for differentially expressed genes (DEGs) between the IA and control groups. Statistical significance was defined by an adjusted P-value (P.adj) < 0.05 coupled with an absolute $\log_2$ fold change ($|\log_2 FC| \geq 1$. To visualize the distribution and expression patterns of DEGs, volcano plots and heatmaps were generated using the "ggplot2" and "pheatmap" R packages, respectively. Subsequently, the set of identified DEGs was intersected with the immune-related gene list retrieved from ImmPort to obtain immune-related differentially expressed genes (IM-DEGs) for downstream investigation.

2.4. Functional and Pathway Enrichment Analysis of Candidate Genes

Functional annotation and pathway enrichment analysis of IM-DEGs were performed using the "enrichGO" and "enrichKEGG" functions within the "clusterProfiler" R package. Gene Ontology (GO) analysis was conducted to categorize gene product functions across three domains: biological process (BP), molecular function (MF), and cellular component (CC) [8]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was carried out to explore the involvement of IM-DEGs in pertinent signaling cascades and higher-order biological functions [9]. The most significantly enriched GO terms and KEGG pathways were prioritized and presented, with the significance threshold established at an adjusted P-value < 0.05.

2.5. Construction of Protein-Protein Interaction Network and Hub Gene Identification

To investigate potential interactions among the encoded proteins of IM-DEGs, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (version 11.5) was utilized to construct a protein-protein interaction (PPI) network. A high-confidence interaction score threshold of ≥0.700 was applied, and disconnected nodes were concealed from the network view. The resulting network was subsequently visualized and analyzed using Cytoscape software (version 3.10.1). Hub genes were identified employing the CytoHubba plugin within Cytoscape, wherein the Maximal Clique Centrality (MCC) algorithm was applied to rank and select the top 10 pivotal genes [10].

2.6. Validation of Hub Genes and Development of the Diagnostic Model

The independent dataset GSE54083 served as the validation cohort for verifying the expression profiles of the identified hub genes. Violin plots illustrating expres-

sion distributions between control and IA groups were constructed using the "ggplot2" package. Statistical significance of intergroup expression differences was assessed according to the following procedure: First, the Shapiro-Wilk test was employed to evaluate the normality assumption of gene expression data distributions. For data conforming to a normal distribution, a two-tailed independent Student's *t*-test was applied; for data deviating from normality, the Wilcoxon rank-sum test was utilized. A P-value < 0.05 is considered indicative of a significant difference. Hub genes demonstrating consistent and statistically significant differential expression in the validation set were retained for model construction. A linear predictive joint diagnostic model was developed using the "glmnet" package, with parameter optimization achieved through cross-validation. Receiver operating characteristic (ROC) curves were generated using the "pROC" package to evaluate diagnostic performance in both the training and validation cohorts. Performance metrics, including area under the curve (AUC), 95% confidence intervals (CI), sensitivity, and specificity, were computed.

2.7. Correlation Analysis of Core Genes

Pearson correlation analysis was conducted among the core genes to assess their co-expression patterns using the "corrplot" R package. Correlation coefficients (Cor) were calculated, and the results were visualized via a correlation heatmap. The magnitude of the correlation coefficient reflects the strength of the association, with values closer to 1 indicating stronger correlations. Positive values denote a positive correlation, whereas negative values indicate an inverse relationship. Statistical significance for correlations was set at $P < 0.01$. This analysis aimed to elucidate potential synergistic regulatory relationships among core genes within the immune-inflammatory milieu of IA.

2.8. Statistical Considerations

All data processing and statistical analyses were executed using R software (version 4.3.1) and associated bioinformatics packages. Quantitative variables are expressed as mean ± standard deviation. Prior to intergroup comparisons, the Shapiro-Wilk test was applied to assess data normality: the two-tailed independent Student's *t*-test was used for normally distributed data, whereas the Wilcoxon rank-sum test was employed for non-normally distributed data. Significance thresholds were defined as follows: for DEG screening, $|\log_2 FC| > 1$ and adjusted $P < 0.05$; for functional enrichment, adjusted $P < 0.05$; and for gene correlation analysis, $P < 0.01$. Diagnostic model efficacy was assessed via ROC curve analysis, with AUC and corresponding 95% CI employed to quantify accuracy. All statistical tests were two-sided, and a P-value < 0.05 was uniformly considered statistically significant.

3. Results

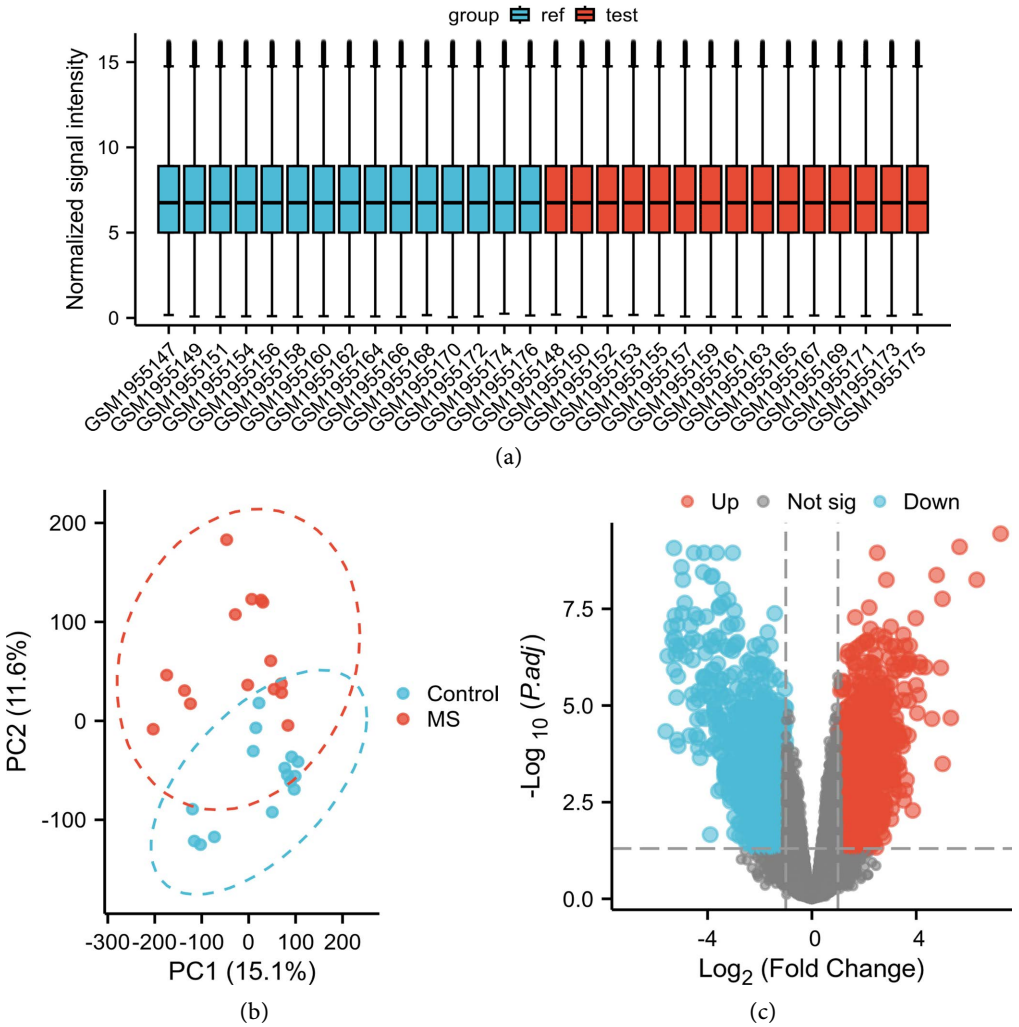
3.1. Identification of Differentially Expressed Genes

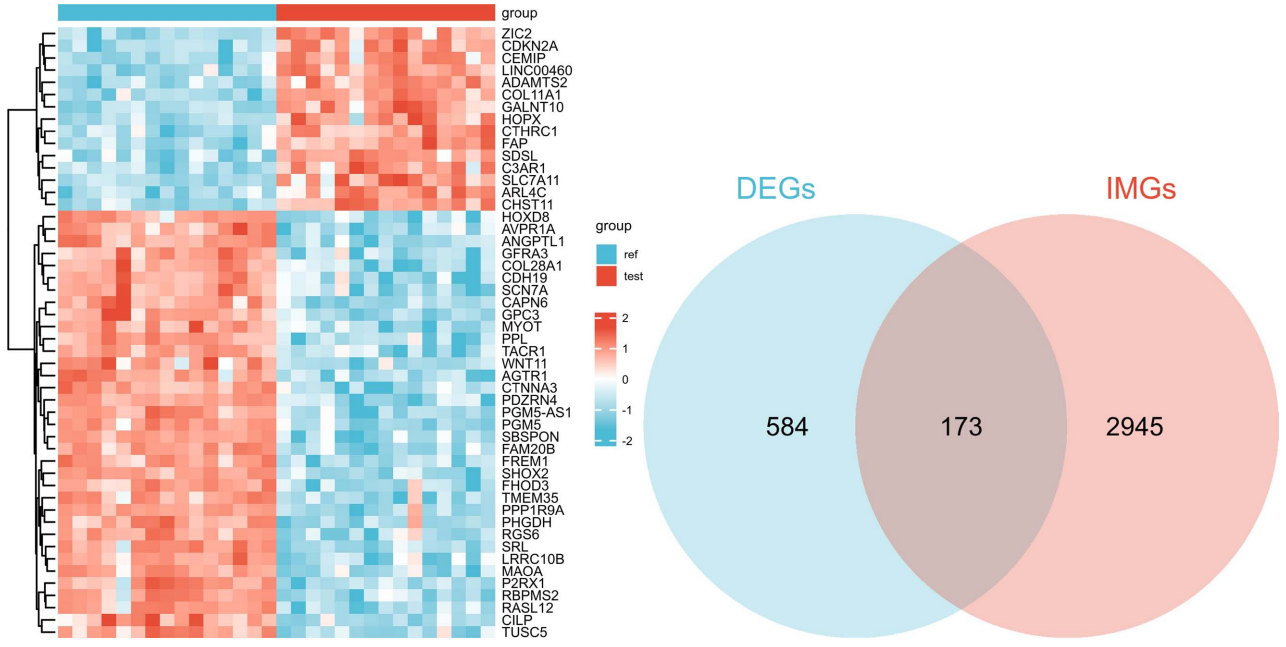
Following normalization of the training dataset GSE75436, boxplot visualization

confirmed consistent expression distributions across all samples, indicating satisfactory inter-group comparability (Figure 1(a)). Principal component analysis (PCA) revealed a distinct separation between IA specimens and control tissues along the principal component axes, signifying substantial divergence in global gene expression profiles between the two groups (Figure 1(b)). Applying the predefined thresholds of $|\log_2$ fold change (FC)$| \geq 1$ and adjusted P-value < 0.05, a total of 757 DEGs were identified, comprising 389 up-regulated and 368 down-regulated transcripts. The overall distribution of DEGs and the expression patterns of the top 50 most significantly altered genes are depicted via a volcano plot and a heatmap, respectively (Figure 1(c) and Figure 1(d)). Concurrently, a compendium of 3118 immune-related genes was retrieved from the ImmPort database. Intersection of the 757 DEGs with this immune gene set yielded 173 IM-DEGs for subsequent investigation (Figure 1(e)).

3.2. GO and KEGG Enrichment Analyses

GO enrichment analysis of the IM-DEGs was conducted using the "clusterProfiler" package in R to obtain standardized functional annotations across biological





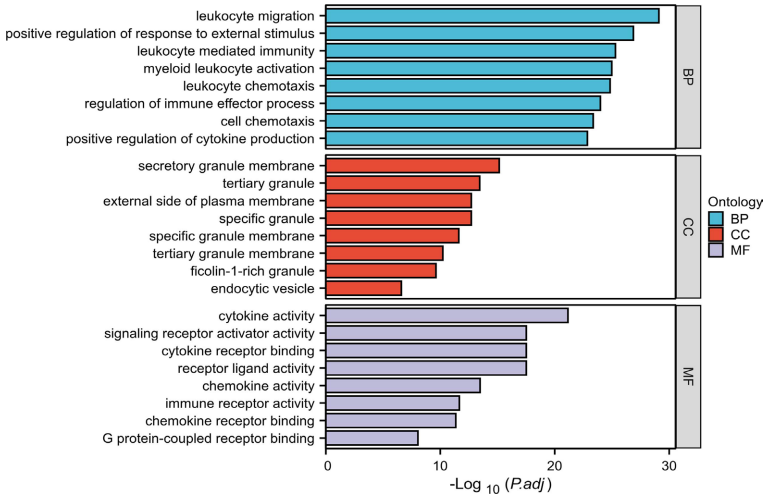
(d) (e)

Figure 1. Identification of differentially expressed genes. (a) Boxplot illustrating normalized expression distributions of the microarray dataset from IA patients. (b) Principal component analysis of the IA expression microarray dataset. (c) Volcano plot displaying the distribution of differentially expressed genes. (d) Heatmap depicting the expression profiles of the top 50 most significantly differentially expressed genes. (e) Venn diagram illustrating the overlap between DEGs and immune-related genes.

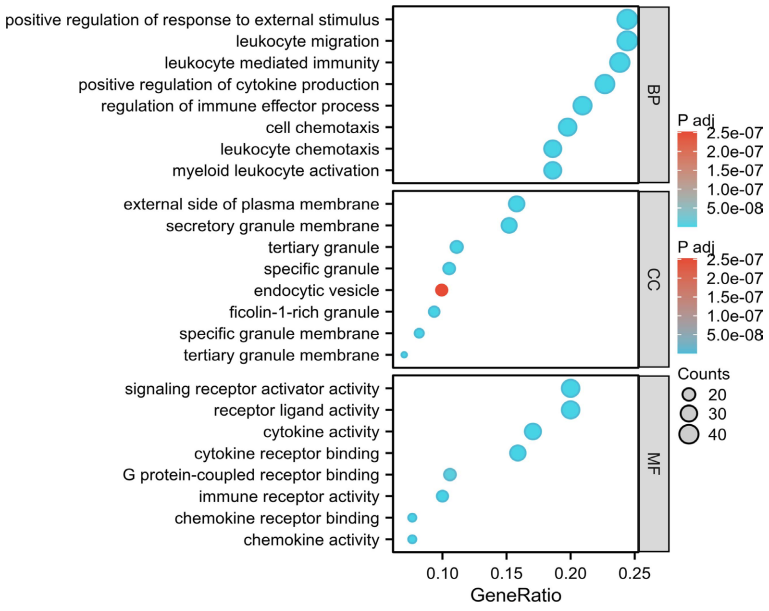
pathways, molecular functions, and cellular localizations. The top eight most significantly enriched terms within each GO category—biological process (BP), cellular component (CC), and molecular function (MF)—are presented in **Figure 2(a)** and **Figure 2(b)**. The GO analysis indicated that IM-DEGs were predominantly enriched in immune-inflammatory biological processes, including leukocyte migration and immune response pathways. Regarding cellular localization, these genes were primarily associated with structures such as secretory granule membranes. In terms of molecular function, significant enrichment was observed for cytokine activity and chemokine activity. KEGG pathway analysis further demonstrated that these genes predominantly participate in immune-inflammatory cascades, notably cytokine-cytokine receptor interaction, chemokine signaling, and viral protein interaction with cytokine and cytokine receptor pathways (**Figure 2(c)** and **Figure 2(d)**). Collectively, these findings suggest that IM-DEGs contribute to the pathological progression of IA through modulation of immune-inflammatory responses, thereby establishing a foundation for subsequent hub gene prioritization and mechanistic exploration.

3.3. PPI Network Construction and Hub Gene Identification

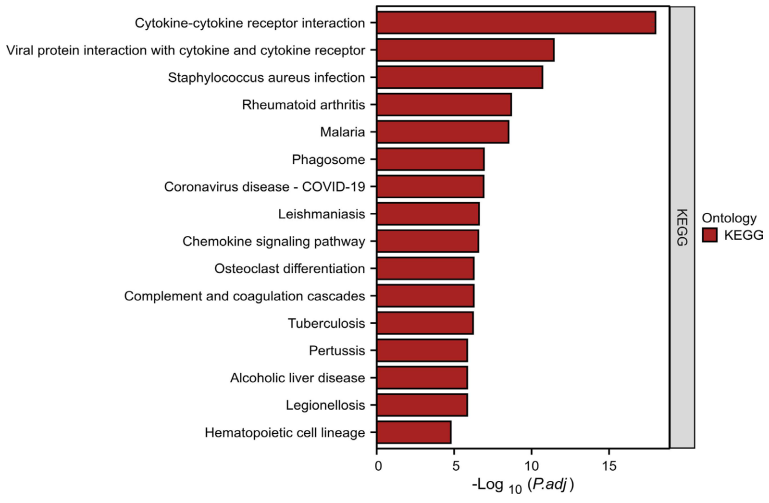
A protein-protein interaction (PPI) network encompassing the 173 IM-DEGs was constructed utilizing the STRING online resource. After concealing unconnected nodes, the resultant network comprised 156 nodes interconnected by 892 edges.


leukocyte migration
positive regulation of response to external stimulus
leukocyte mediated immunity
myeloid leukocyte activation
leukocyte chemotaxis
regulation of immune effector process
cell chemotaxis
positive regulation of cytokine production
secretory granule membrane
tertiary granule
external side of plasma membrane
specific granule
specific granule membrane
tertiary granule membrane
ficolin-1-rich granule
endocytic vesicle
cytokine activity
signaling receptor activator activity
cytokine receptor binding
receptor ligand activity
chemokine activity
immune receptor activity
chemokine receptor binding
G protein-coupled receptor binding
BP
CC
MF
Ontology
BP
CC
MF
0
10
20
30
-Log 10 (P.adj)

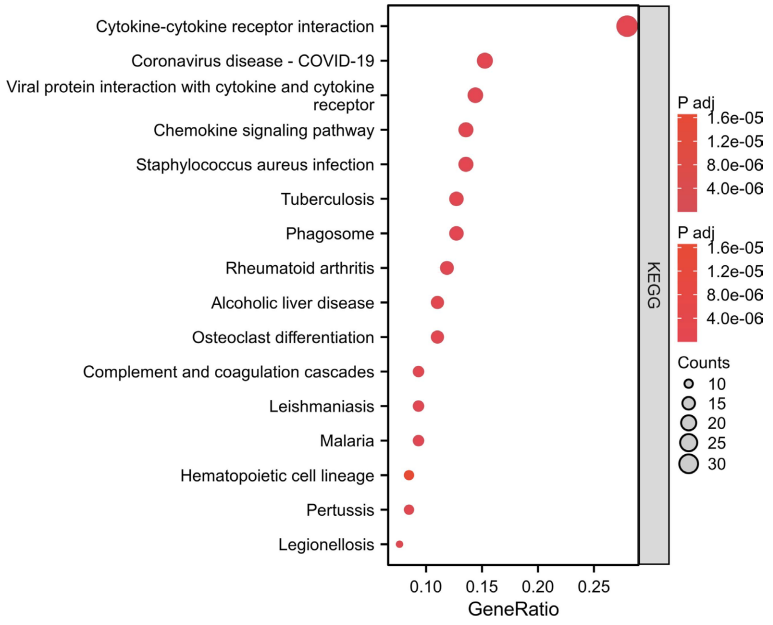
(a)


positive regulation of response to external stimulus
leukocyte migration
leukocyte mediated immunity
positive regulation of cytokine production
regulation of immune effector process
cell chemotaxis
leukocyte chemotaxis
myeloid leukocyte activation
external side of plasma membrane
secretory granule membrane
tertiary granule
specific granule
endocytic vesicle
ficolin-1-rich granule
specific granule membrane
tertiary granule membrane
signaling receptor activator activity
receptor ligand activity
cytokine activity
cytokine receptor binding
G protein-coupled receptor binding
immune receptor activity
chemokine receptor binding
chemokine activity
BP
CC
MF
P adj
2.5e-07
2.0e-07
1.5e-07
1.0e-07
5.0e-08
P adj
2.5e-07
2.0e-07
1.5e-07
1.0e-07
5.0e-08
Counts
20
30
40
0.10
0.15
0.20
0.25
GeneRatio

(b)


Cytokine-cytokine receptor interaction
Viral protein interaction with cytokine and cytokine receptor
Staphylococcus aureus infection
Rheumatoid arthritis
Malaria
Phagosome
Coronavirus disease - COVID-19
Leishmaniasis
Chemokine signaling pathway
Osteoclast differentiation
Complement and coagulation cascades
Tuberculosis
Pertussis
Alcoholic liver disease
Legionellosis
Hematopoietic cell lineage
KEGG
Ontology
KEGG
0
5
10
15
-Log 10 (P.adj)

(c)



(d)

Figure 2. GO and KEGG enrichment analyses of IM-DEGs. (a) Bar plot illustrating the top enriched GO terms for IM-DEGs. (b) Bubble plot depicting GO enrichment results for IM-DEGs. (c) Bar plot displaying the top enriched KEGG pathways for IM-DEGs. (d) Bubble plot showing KEGG pathway enrichment results for IM-DEGs.

Visualization of the network was performed using Cytoscape software. Application of the Maximal Clique Centrality (MCC) algorithm via the CytoHubba plugin identified the top 10 highest-scoring hub genes, namely IL1B, TNF, CXCL8, CD8A, TYROBP, CCR1, CXCL10, CCL4, CCR5, and CCL20. Node coloration intensity corresponds to the MCC score, with darker hues denoting higher centrality. Among these, IL1B, TNF, CXCL8, CD8A, and TYROBP occupied the most central positions within the network, exhibiting the highest connectivity (**Figure 3**).

3.4. Expression Validation of Core Genes in the Independent Dataset

Expression levels of the five identified core genes (IL1B, TNF, CXCL8, CD8A, and TYROBP) were validated using the independent dataset GSE54083. Violin plots were generated to compare expression distributions between the IA and control groups. The findings are detailed below:

CXCL8: Expression in the IA group (test) was markedly elevated relative to the control group (ref), a difference that achieved statistical significance ($P < 0.001$). The violin plot (**Figure 4(a)**) illustrates a clear upward shift in the expression distribution of the IA cohort with minimal overlap between groups, implying that heightened CXCL8 expression may facilitate IA initiation and progression, potentially acting as a disease-promoting factor.

TNF: Conversely, TNF expression was significantly higher in the control group (ref) compared with the IA group (test) ($P < 0.001$). As depicted in **Figure 4(b)**,

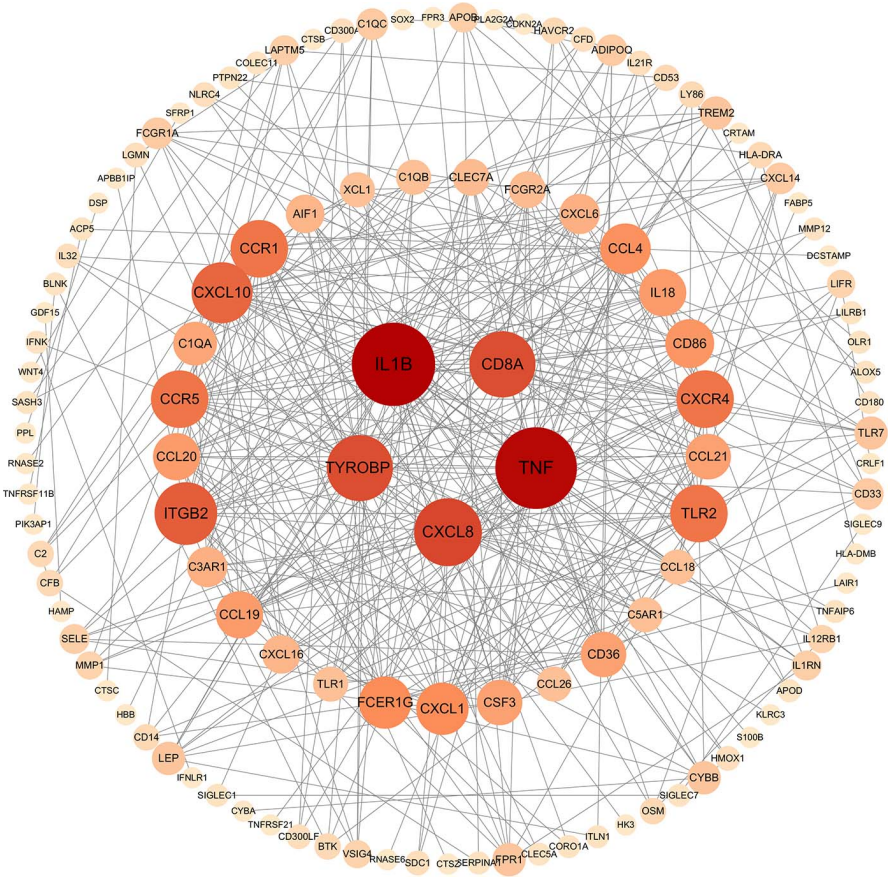


Figure 3. PPI network construction and hub gene identification. Visualization of the IMDEG PPI network and identification of core hub genes.

the expression distribution of the control cohort is globally elevated relative to that of the aneurysm cohort, suggesting that diminished TNF expression is closely associated with IA pathogenesis and may function as a potential protective or inhibitory factor.

CD8A: Expression of CD8A in the control group (ref) significantly surpassed that in the IA group (test) ($P < 0.05$). The violin plot (**Figure 4(c)**) reveals a higher overall expression distribution in controls, indicating that downregulation of CD8A may reflect aberrant T-cell immune functionality within aneurysmal tissue, thereby contributing to immune-inflammatory dysregulation.

IL1B: Although expression levels in the control group (ref) tended to be higher than those in the IA group (test), the difference did not reach statistical significance ($P > 0.05$) (**Figure 4(d)**). This suggests that the expression discrepancy may be influenced by limitations in validation cohort sample size, platform-specific variations, or sample heterogeneity. The precise role of IL1B in IA pathogenesis warrants further investigation.

TYROBP: Expression in the IA group (test) appeared elevated relative to controls (ref); however, this trend was not statistically significant ($P > 0.05$) (**Figure 4(e)**). Given its strong correlation with other core genes such as CXCL8 and IL1B,

TYROBP retains potential research value, and future studies with expanded sample sizes are needed to clarify its expression characteristics and functional relevance.

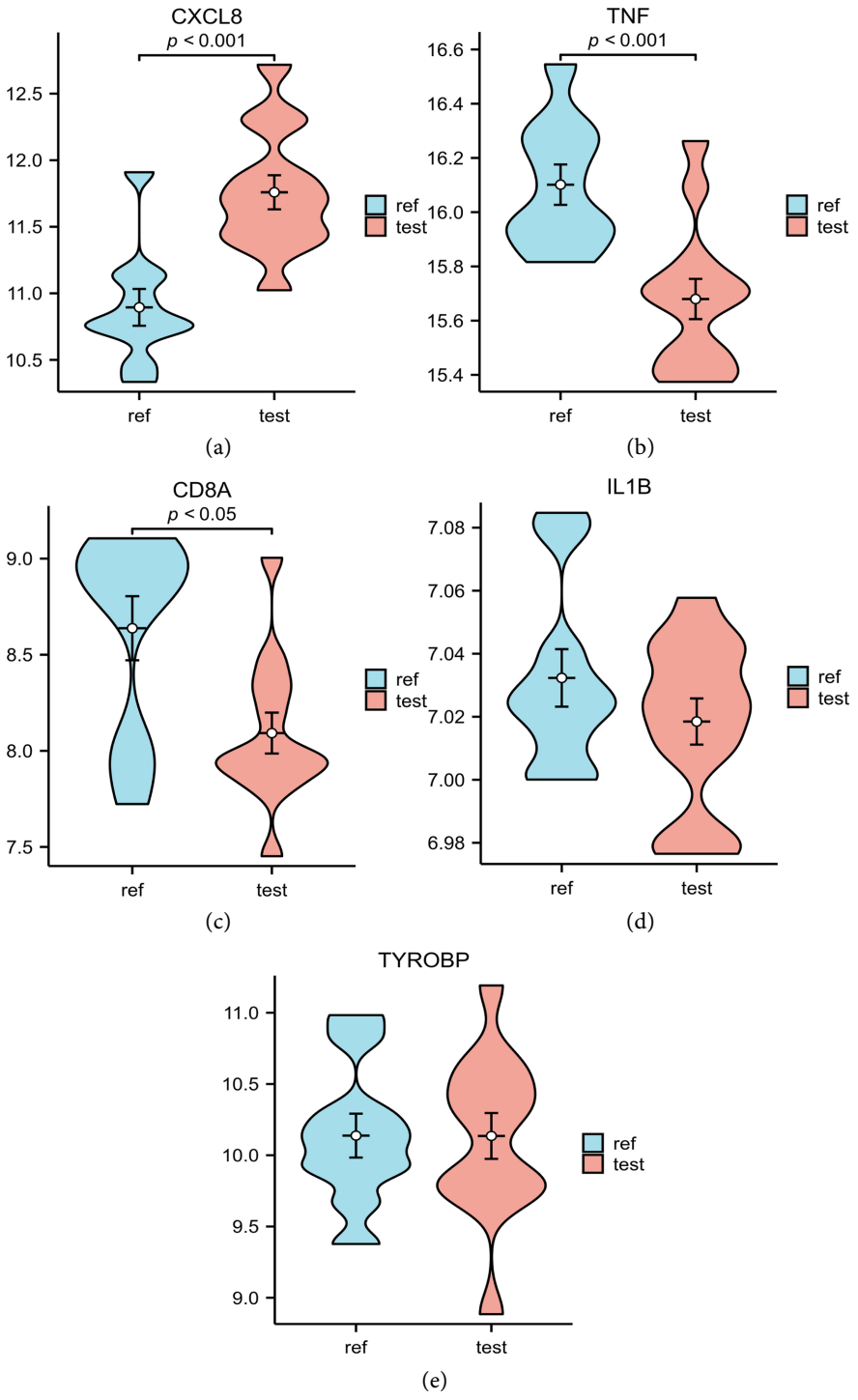


Figure 4. Expression analysis of core genes in the validation dataset. (a) Violin plot of CXCL8 expression in the validation cohort. (b) Violin plot of TNF expression in the validation cohort. (c) Violin plot of CD8A expression in the validation cohort. (d) Violin plot of IL1B expression in the validation cohort. (e) Violin plot of TYROBP expression in the validation cohort.

3.5. Correlation Analysis of Core Genes

Pearson correlation analysis was performed among the five core genes (CXCL8, TNF, CD8A, TYROBP, and IL1B), with the results visualized as a correlation heatmap (**Figure 5(a)** and **Figure 5(b)**). The analysis revealed robust positive correlations among all core genes ($P < 0.01$), with correlation coefficients ranging from 0.46 to 0.78. The strongest association was observed between CXCL8 and TYROBP (Cor = 0.78), followed by IL1B and TNF (Cor = 0.71). Additional notable correlations included CD8A with TNF (Cor = 0.65), CXCL8 with IL1B (Cor = 0.58), and CD8A with TYROBP (Cor = 0.46). These significant positive interrelationships suggest that these core genes may operate synergistically within the immune-inflammatory regulatory network of IA, collectively contributing to disease onset and evolution.

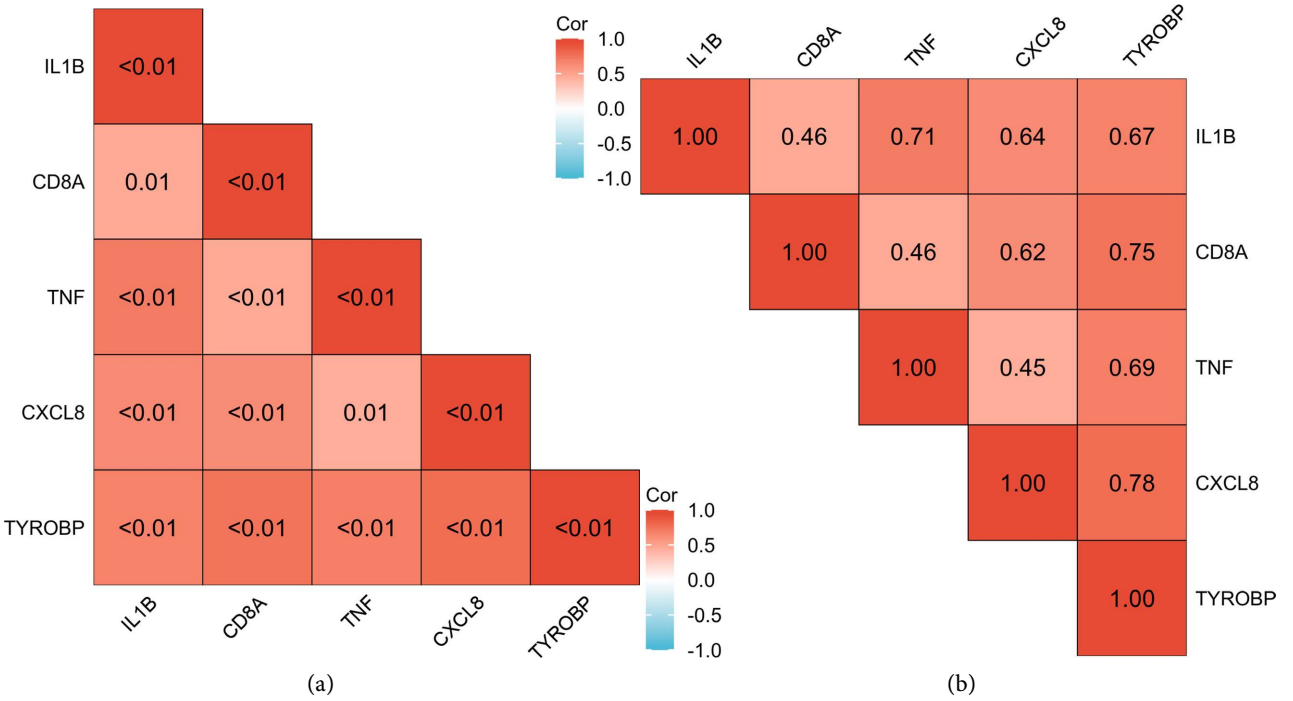


Figure 5. Correlation heatmap of core gene expression. (a) Heatmap displaying P-values for core gene correlation significance. (b) Heatmap depicting Pearson correlation coefficients among core genes.

3.6. Diagnostic Model Construction and Performance Evaluation

To translate the identified core immune-related genes into a clinically applicable individualized diagnostic tool, a combined diagnostic model was developed based on the expression profiles of the core genes. A multi-dimensional validation framework was employed to systematically assess its diagnostic accuracy, robustness, and potential clinical utility. The results are presented below.

3.6.1. Evaluation of Single-Gene Diagnostic Performance

To ascertain the independent diagnostic utility of each core gene for IA, receiver operating characteristic ROC curves were generated using the training cohort

GSE75436, and the area under the curve AUC was calculated to quantify discriminatory capacity. As illustrated in **Figure 6(a)**, all five candidate core genes exhibited robust diagnostic performance for IA. The AUC values were as follows: TYROBP, 0.960 (95% CI: 0.885 - 1.000); IL1B, 0.933 (95% CI: 0.836 - 1.000); TNF, 0.884 (95% CI: 0.765 - 1.000); CD8A, 0.849 (95% CI: 0.715 - 0.983); and CXCL8, 0.813 (95% CI: 0.668 - 0.958). Each of these values surpassed the threshold of 0.8, indicative of excellent discriminatory power. These findings confirm that the core genes identified in this study represent potential biomarkers for IA diagnosis and provide a solid molecular foundation for constructing a multi-gene joint diagnostic model.

3.6.2. Construction of a Nomogram Model and Variable Selection

To identify variables with independent incremental diagnostic value for IA and to mitigate model overfitting, logistic regression was employed for variable selection. TYROBP was ultimately excluded due to a lack of independent diagnostic gain within the model, while the remaining four core genes—IL1B, CD8A, TNF, and CXCL8—were retained for nomogram construction (**Figure 6(b)**). The scoring of core genes corresponds to their clinically detectable expression ranges: IL1B corresponds to an expression interval of 6 - 12; CD8A, 2 - 12; TNF, 2.5 - 8.5; and CXCL8 corresponds to its full detection range. The total point scale ranges from 0 to 160, corresponding to a linear predictor value interval of −8 to 8. The total points map to an estimated probability of IA risk, ranging from 0.2 to 0.8.

It is important to clarify that the four-gene nomogram described above was derived from variable selection using logistic regression within the training set and is intended to illustrate the capacity of multi-gene combinations for quantitative disease risk assessment. However, to ensure model robustness and reproducibility in external independent validation, an additional cross-platform expression stability screening step was implemented: only core genes demonstrating statistically significant intergroup differential expression ($P < 0.05$) in the validation set GSE54083 were retained. Following this screening, IL1B was excluded because its expression difference in the validation set did not reach statistical significance; TYROBP, despite exhibiting the highest single-gene AUC, was also excluded for the same reason. Consequently, the primary final diagnostic model of this study is the three-gene combined model based on CXCL8, TNF, and CD8A. This model balances internal goodness-of-fit with external generalizability and constitutes the core conclusion of this investigation. The four-gene nomogram is presented as an internal exploratory analysis result but is not recommended as the final model.

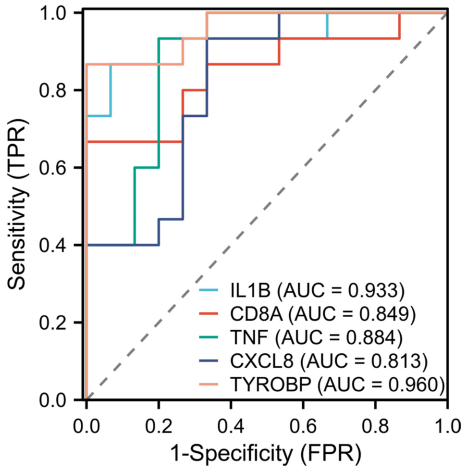
3.6.3. Diagnostic Model Construction with Internal and External Validation

Integrating the differential expression results from the validation cohort, only CXCL8, TNF, and CD8A demonstrated statistically significant inter-group differences in GSE54083, thereby exhibiting robust cross-cohort expression stability.

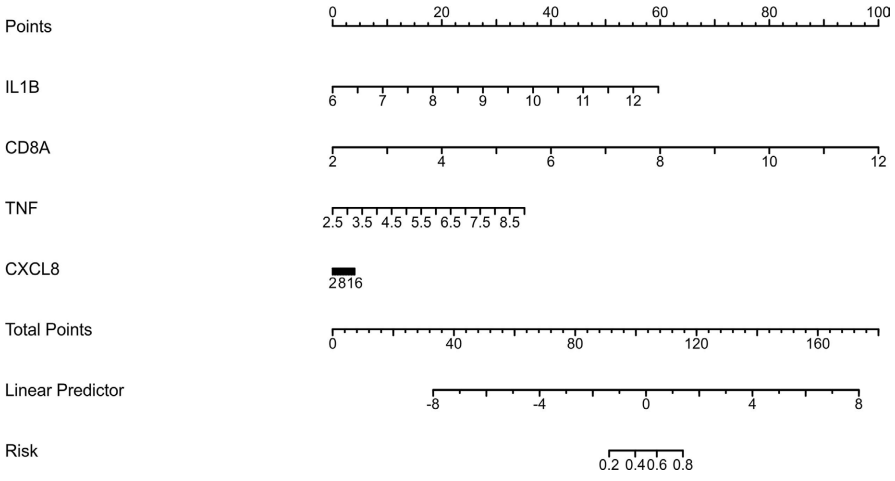
Consequently, these three genes were selected as core variables to construct a precise and reproducible joint diagnostic model, ensuring external reliability and potential for clinical translation. ROC curve analyses were performed in both the training set GSE75436 and the independent validation set GSE54083.

Internal validation (Training set GSE75436): The model achieved an AUC of 0.947 (95% CI: 0.862 - 1.000), demonstrating exceptional discriminatory capacity to effectively distinguish IA samples from normal controls. This indicates a high degree of model fit and diagnostic reliability within the training cohort (**Figure 6(c)**).

External validation (Independent validation set GSE54083): The model yielded an AUC of 0.754 (95% CI: 0.546 - 0.962). Although the external validation performance was moderately attenuated relative to internal validation—likely attributable to cross-platform technical variation and sample size constraints of the validation cohort—the model maintained a stable level of discriminatory ability. This corroborates the favorable cross-cohort robustness and generalizability of the diagnostic model constructed from these three core genes (**Figure 6(d)**).



(a)



(b)

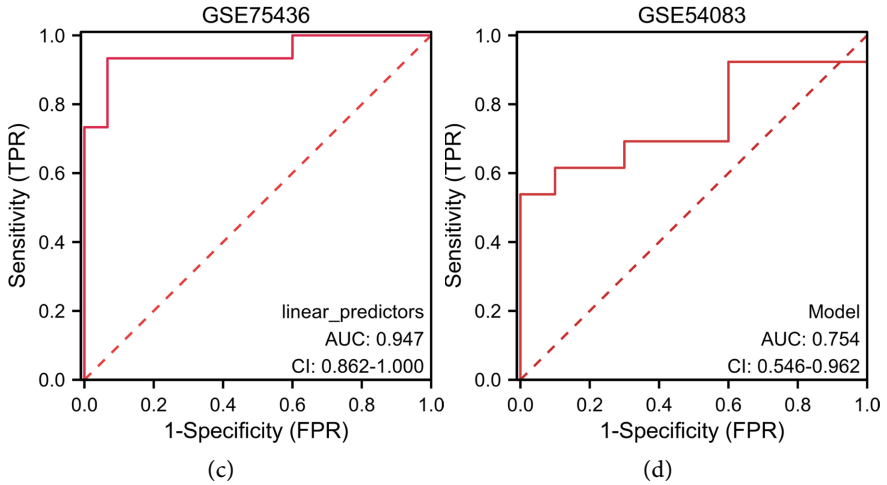


Figure 6. Diagnostic model construction with internal and external validation. (a) Single-gene ROC curve analysis for the five core genes. (b) Nomogram depicting individualized risk estimation for intracranial aneurysm. (c) ROC curve analysis of the three-gene combined diagnostic model in the GSE75436 training set. (d) External validation ROC curve of the three-gene combined diagnostic model in the independent GSE54083 dataset.

4. Discussion

IA represents a potentially life-threatening cerebrovascular disorder wherein rupture leads to subarachnoid hemorrhage—a catastrophic event associated with high mortality and long-term disability—thereby imposing a substantial burden on healthcare systems worldwide [11]. Currently, clinical diagnosis relies predominantly on advanced neuroimaging modalities, yet reliable molecular biomarkers for early warning and risk stratification remain conspicuously absent, posing formidable challenges for preventive intervention. Existing therapeutic strategies are largely directed toward managing radiographically confirmed aneurysms, offering limited efficacy in averting lesion formation or progression. Consequently, elucidating the underlying molecular mechanisms—particularly the contribution of immune-inflammatory processes—is of paramount and immediate clinical significance for the development of novel diagnostic and therapeutic paradigms.

Through the integration of comprehensive bioinformatics analysis with independent dataset validation, the present study systematically identified three immune-related genes—CXCL8, TNF, and CD8A—as pivotal molecular determinants of IA, and further constructed a combined predictive model exhibiting robust diagnostic performance. These findings not only furnish a novel perspective for comprehending the immunopathological underpinnings of IA, but also establish a molecular foundation for the advancement of non-invasive diagnostic tools. The ensuing discussion contextualizes our results within the framework of contemporary research advances.

Foremost, the enrichment analyses conducted herein compellingly corroborate that immune-inflammatory cascades constitute a central driving force in IA pathogenesis. Both GO and KEGG analyses revealed that IM-DEGs were prominently enriched in pathways governing “leukocyte migration”, “cell chemotaxis”, and

"cytokine-cytokine receptor interaction" [12]. This observation aligns seamlessly with the prevailing paradigm that endothelial injury initiates a coordinated cascade involving both innate and adaptive arms of the immune system [13]. Inflammatory mediators liberated by infiltrating immune cells disrupt vascular wall homeostasis, thereby fostering aneurysmal development.

Furthermore, the three core genes delineated in this investigation—CXCL8, TNF, and CD8A—precisely demarcate three critical, interconnected facets of the IA immune network. CXCL8 encodes a potent neutrophil chemoattractant. Its expression was consistently and significantly up-regulated in IA tissues across both training and validation cohorts, a finding congruent with the established role of neutrophil-driven inflammation in vascular pathology [13]. Neutrophil-derived proteases are capable of directly degrading the extracellular matrix of the vessel wall; thus, elevated CXCL8 may represent a key signaling axis propelling this destructive process [14]. In marked contrast to the unequivocal pro-inflammatory function of CXCL8, TNF expression was found to be significantly diminished in IA tissues. This ostensibly paradoxical observation may illuminate the intricate complexity of inflammatory regulation in chronic disease settings. As a master regulator of inflammation, TNF exerts pleiotropic effects, encompassing both pro-inflammatory activity and the induction of immunomodulatory feedback mechanisms. Its downregulation might reflect a localized compensatory response aimed at limiting excessive tissue injury within the IA microenvironment, yet could simultaneously signify a broader disruption of immune homeostasis [15]. Such perturbations in delicate regulatory balance may constitute a critical factor in sustained disease progression.

CD8A, encoding the CD8 alpha chain, serves as a canonical marker of cytotoxic T lymphocytes (CTLs). The observed downregulation of CD8A in this study suggests that CTL functionality may be suppressed or exhausted within the IA microenvironment. In the context of persistent inflammation, chronic antigenic stimulation can precipitate T-cell functional exhaustion [16]. Diminished CD8A expression may thus indicate attenuated immune surveillance against vascular wall-associated antigens, or alternatively, the establishment of an immunosuppressive milieu that permits aberrant reparative processes to persist. This insight offers a potential rationale for exploring immunomodulatory strategies—such as immune checkpoint modulation—that target T-cell function in IA [17].

Moreover, the combined diagnostic model constructed from CXCL8, TNF, and CD8A demonstrated stable discriminatory capacity in both the training cohort and the independent validation set, yielding area under the curve AUC values of 0.947 and 0.754, respectively. By integrating indicators of innate immunity (CXCL8), central inflammatory regulation (TNF), and adaptive cellular immunity (CD8A), this model captures a more comprehensive portrait of the IA immune signature [12]. It should be objectively acknowledged that the training set was derived from intracranial aneurysm tissue samples, whereas the validation set consisted of peripheral blood mononuclear cell (PBMC) samples—a fundamental dif-

ference in biological matrix. Thus, the attenuation in model performance observed in the validation cohort may reflect not only the impact of cross-platform technical variation but also, in part, tissue-specific differences in gene expression patterns. Although the differential expression trends of the core genes were partially conserved in the blood-based samples examined herein, caution is warranted when directly extrapolating tissue-derived molecular signatures to non-invasive peripheral blood testing [18]. This multi-gene panel constitutes a promising candidate combination for the development of non-invasive diagnostic assays predicated on peripheral blood gene expression profiling [19]; however, subsequent retraining and calibration in larger peripheral blood-based clinical cohorts are necessary to optimize its applicability and accuracy in non-invasive testing scenarios.

Notwithstanding these contributions, several limitations warrant acknowledgment. First, the study relied upon publicly available datasets with constrained sample sizes and a paucity of detailed clinicopathological annotations (e.g., aneurysm rupture status), thereby precluding more granular subgroup analyses [20]. Second, the bioinformatics analyses were primarily anchored in transcriptomic data; corresponding alterations at the protein level, as well as the precise cellular provenance of these gene expression changes (e.g., endothelial cells, vascular smooth muscle cells, or infiltrating leukocytes), remain to be delineated. Third, this investigation constitutes a retrospective association study. While robust correlations between gene expression and disease status were established, causal inferences cannot be drawn. Whether CXCL8, TNF, and CD8A function as drivers of IA pathogenesis or represent secondary epiphenomena necessitates further clarification through functional gain- and loss-of-function experiments [13]. Fourth, the external generalizability of the diagnostic model requires further corroboration and refinement within larger-scale, prospective, multicenter clinical cohorts.

In summation, through systematic data mining and rigorous validation, the present study delineates the pivotal roles of CXCL8, TNF, and CD8A in the immunopathology of IA and establishes a promising multi-gene diagnostic model. These findings deepen our mechanistic understanding of IA, underscoring the centrality of chemokine signaling, inflammatory homeostasis, and T-cell functionality within the immune response. Future investigations should prioritize efforts to ascertain the cellular origins of these key molecules via technologies such as single-cell sequencing, validate their functional contributions in experimental animal models, and optimize and corroborate the diagnostic model within large prospective cohorts. Ultimately, such endeavors hold the potential to advance clinical translation, facilitating improved risk stratification and informing targeted therapeutic decision-making.

5. Conclusion

The findings of this investigation establish CXCL8, TNF, and CD8A as core genes orchestrating immune-inflammatory regulation in intracranial aneurysm, likely acting in concert through pivotal pathways such as cytokine-cytokine receptor in-

teraction and chemokine signaling to drive aneurysmal initiation and progression. The combined diagnostic model predicated upon these three genes demonstrated robust discriminatory capacity in tissue samples and exhibited preliminary cross-matrix consistency in the peripheral blood validation cohort, thereby providing a panel of biomarkers with exploratory value for the early non-invasive detection of IA. Furthermore, this study enriches the current understanding of the immune-inflammatory pathogenesis underlying IA and furnishes a critical theoretical framework and novel targets for immune-targeted therapeutic interventions. Collectively, these insights hold significant implications for optimizing clinical management strategies and improving patient outcomes in intracranial aneurysms. Future studies should prioritize re-training and rigorous validation of this model in prospective, peripheral blood-based cohorts.

Funding

2026 Innovation Project of Youjiang Medical University for Nationalities Graduate Education (YZCXJH2026010).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Wang, X., Wen, D., You, C., Tao, C. and Ma, L. (2023) Comprehensive Analysis of Immune Cell Infiltration and Role of MSR1 Expression in Aneurysmal Subarachnoid Haemorrhage. *Cell Proliferation*, **56**, e13379. https://doi.org/10.1111/cpr.13379

[2] Ma, Z., Wu, P., Abulizi, A., Yang, W., Maimaiti, A., Akram, P., *et al.* (2026) From Genes to Diagnosis: The Impact of UNC5B and DOK5 in Intracranial Aneurysm Detection and Pathogenesis. *PLOS ONE*, **21**, e0340496. https://doi.org/10.1371/journal.pone.0340496

[3] Joo, B., Choi, H.S., Ahn, S.S., Cha, J., Won, S.Y., Sohn, B., *et al.* (2021) A Deep Learning Model with High Standalone Performance for Diagnosis of Unruptured Intracranial Aneurysm. *Yonsei Medical Journal*, **62**, 1052-1061. https://doi.org/10.3349/ymj.2021.62.11.1052

[4] Wen, Z., Wang, Y., Zhong, Y., Hu, Y., Yang, C., Peng, Y., *et al.* (2024) Advances in Research and Application of Artificial Intelligence and Radiomic Predictive Models Based on Intracranial Aneurysm Images. *Frontiers in Neurology*, **15**, Article ID: 1391382. https://doi.org/10.3389/fneur.2024.1391382

[5] Zhang, M., Zhou, L., Zhao, Y., Wang, Y., Zhang, Z. and Liu, Z. (2025) Comprehensive Molecular Analyses of an Autoimmune-Related Gene Predictive Model and Immune Infiltrations Using Machine Learning Methods in Intracranial Aneurysma. *Frontiers in Immunology*, **16**, Article ID: 1531930. https://doi.org/10.3389/fimmu.2025.1531930

[6] Huang, J., Gao, S., Liu, S. and Liu, L. (2025) Exploration of Immune-Related Transcription Control/Regulation in Intracranial Aneurysm through KEGG Analysis and *In- Vivo* Validation. *Journal of Inflammation Research*, **18**, 6305-6317. https://doi.org/10.2147/jir.s506360

[7] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., *et al.* (2015) Limma

Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, **43**, e47. https://doi.org/10.1093/nar/gkv007

[8] Wilkerson, M.D. and Hayes, D.N. (2010) Consensusclusterplus: A Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics*, **26**, 1572-1573. https://doi.org/10.1093/bioinformatics/btq170

[9] Kanehisa, M. and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **28**, 27-30. https://doi.org/10.1093/nar/28.1.27

[10] Chin, C., Chen, S., Wu, H., Ho, C., Ko, M. and Lin, C. (2014) Cytohubba: Identifying Hub Objects and Sub-Networks from Complex Interactome. *BMC Systems Biology*, **8**, S11. https://doi.org/10.1186/1752-0509-8-s4-s11

[11] Li, S., Zhang, Q., Chen, Z., Huang, Z., Zhang, L. and Chen, F. (2022) Novel Insight into Ferroptosis-Related Genes, Molecular Subtypes, and Immune Characteristics in Intracranial Aneurysms. *Inflammation Research*, **71**, 1347-1364. https://doi.org/10.1007/s00011-022-01633-8

[12] Hou, J., Yue, Y., Wu, S., Li, D., Dai, W., Liu, Z., *et al.* (2025) Plastisphere Characterization and Ecological Implications of Plastic Pollution in Northeast Tiger and Leopard National Park, China. *Journal of Hazardous Materials*, **496**, Article ID: 139351. https://doi.org/10.1016/j.jhazmat.2025.139351

[13] Jung, D., Jung, S., Lee, T., Kim, Y., Lee, Y., Seo, I., *et al.* (2025) The Sesquiterpene Lactone Estafiatin Exerts an Anti-Inflammatory Effect against *Mycobacterium abscessus* Infection by Regulating Interleukin-1 Beta Production. *Phytomedicine*, **146**, Article ID: 157080. https://doi.org/10.1016/j.phymed.2025.157080

[14] Peng, Y., Li, R., Tang, Y., Liu, X., Zhang, Y., Gao, X., *et al.* (2025) Mechanistic Insights into Electric Field-Driven Specific Ion Effects on Nanoplastics Aggregation in Heavy Metal Co-Contaminated Aquatic Systems. *Journal of Hazardous Materials*, **496**, Article ID: 139347. https://doi.org/10.1016/j.jhazmat.2025.139347

[15] Pham, U.N.T., Chou, K., Ngo, T., Pan, W., Ruan, T. and Lin, Y. (2026) Targeted Immunomodulation with H2-Generating Nanostructures to Mitigate Obesity-Induced Inflammation and Metabolic Dysfunctions. *Biomaterials*, **325**, Article ID: 123570. https://doi.org/10.1016/j.biomaterials.2025.123570

[16] Gibson, J.S., Eberle, J.W., Werntz, A., Harrison, M.B., Manning, C.A., Yan, G., *et al.* (2025) Development of the Movement Disorders Interpretation Bias Scale and Psychometric Evaluation in Adults with Huntington's Disease. *Parkinsonism & Related Disorders*, **138**, Article ID: 107971. https://doi.org/10.1016/j.parkreldis.2025.107971

[17] Free, M.E., Ciavatta, D.J. and Falk, R.J. (2025) Can We Cure Vasculitis? *Current Opinion in Immunology*, **96**, Article ID: 102618. https://doi.org/10.1016/j.coi.2025.102618

[18] Zhou, W., Shao, J., Huang, J., Gan, J., Wang, Y., Chen, J., *et al.* (2025) Air Pollutants and Adverse Pregnancy Outcomes: A Multi-Method Integration of Mendelian Randomization, Meta-Analysis, and Animal Models. *Ecotoxicology and Environmental Safety*, **302**, Article ID: 118699. https://doi.org/10.1016/j.ecoenv.2025.118699

[19] Wang, S., Wu, M., Wei, J., Xia, W., Luo, Z. and Tian, L. (2025) Effectiveness of Mindfulness-Based Cognitive Therapy via E-Health on Anxiety and Depression in Adults: A Meta-Analysis. *General Hospital Psychiatry*, **96**, 211-222. https://doi.org/10.1016/j.genhosppsych.2025.07.017

[20] Tipiani, O. and Arce, E. (2025) Development of Z-Score-Based Models to Predict Perinatal Mortality in Pregnancies Complicated by Fetal Growth Restriction Managed at a Tertiary Care Hospital. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **313**, Article ID: 114612. https://doi.org/10.1016/j.ejogrb.2025.114612