Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptomic and machine learning analyses identify entosis-related key genes and potential regulatory networks in ischemic stroke

Le Chang

changle0811@163.com

Xi'an Central Hospital

Weili Cao

Xi'an Central Hospital

Xurong Zhu

Xi'an Central Hospital

Naibing Gu

Xi'an Central Hospital

Weiping Zhang

Xi'an Central Hospital

Zhiqing Liu

Xi'an Central Hospital

Zhengli Di

Xi'an Central Hospital

Research Article

Keywords: Ischemic stroke, Entosis, Drug prediction, Molecular docking, Key genes

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11361689/v1

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

Additional Declarations: No competing interests reported.

Transcriptomic and machine learning analyses identify entosis-related key genes and potential regulatory networks in ischemic stroke

Le Chang[1*],Weili Cao[1],Xurong Zhu[1],Naibing Gu[1],Weiping Zhang[1],Zhiqing Liu[1],Zhengli Di[1*]

[1]Department of Neurology, The Affiliated Xi'an Central Hospital of Xi'an Jiaotong University, No.161 Xiwu Road, Xincheng District, Xi'an City, Shaanxi Province, China

*** Correspondence:**

Le Chang and Zhengli Di contributed equally as corresponding authors.

E-mail addresses: changle0811@163.com(Le Chang) , dizhengli@163.com (Zhengli Di)

ORCID: 0000-0001-9279-9988(Le Chang), 0009-0005-4446-9622 (Zhengli Di)

Abstract

Background: Ischemic stroke (IS) remains a major cause of mortality and long-term disability worldwide, while effective therapeutic targets remain limited. Entosis is a non-apoptotic form of cell-in-cell death, but its involvement in IS is unknown. This study sought to identify entosis-related key genes and delineate their potential regulatory networks in IS.

Methods: Bulk-transcriptome profiles were obtained for the training dataset GSE58294 and the validation dataset GSE122709, and entosis-related genes were compiled from previously published studies. Weighted gene co-expression network analysis and differential expression analysis were integrated with machine-learning selection and cross-dataset expression verification to determine key genes. Subsequent analyses included functional enrichment, immune infiltration, construction of molecular regulatory networks, drug prediction, and molecular docking.

Results: Two key genes, NOL9 and WDR3, were identified and were significantly upregulated in both transcriptomic datasets. The nomogram incorporating these genes demonstrated favorable apparent performance in the training cohort. Eight immune cell subtypes differed significantly in the infiltration analysis. M0 macrophages showed inverse correlations with NOL9 (correlation (cor) = -0.60, $P < 0.001$) and WDR3 (cor = -0.57, $P < 0.001$), whereas naive B cells correlated positively with NOL9 (cor = 0.46, $P < 0.001$) and WDR3 (cor = 0.35, $P < 0.001$). Genes associated with NOL9 were enriched in ribosome-biogenesis pathways, while those associated with WDR3 were enriched in spliceosome-related pathways. Construction of a competing endogenous

RNA network yielded six microRNAs predicted to target NOL9 and WDR3; hsa-miR-499a-5p was additionally connected with three long non-coding RNAs (XIST, MALAT1, NEAT1). The analysis also predicted 398 transcription factors for NOL9 (e.g., SPI1, CTCF) and 331 for WDR3 (e.g., BRD4, CBFB). Digoxigenin emerged as an exploratory compound, with predicted binding energies of -9.82 kcal/mol for NOL9 and -10.49 kcal/mol for WDR3. In RT-qPCR experiments, NOL9 and WDR3 were significantly downregulated in IS patients relative to controls; because this direction contrasted with the transcriptomic datasets, additional validation is required.

Conclusions: The analyses nominate NOL9 and WDR3 as exploratory entosis-related biomarker candidates for IS. Further studies are needed to resolve their expression direction and to establish their biological functions and therapeutic relevance.

Keywords: Ischemic stroke, Entosis, Drug prediction, Molecular docking, Key genes

1. Background

Stroke continues to impose a substantial worldwide burden through both mortality and disability, affecting not only patients but also caregivers and healthcare systems [1]. In the Global Burden of Disease 2021 study, stroke was the third leading cause of death and the fourth largest contributor to DALYs, with ischemic stroke accounting for the majority of cases [2]. Ischemic stroke occurs when arterial blood flow to the central nervous system (CNS) is interrupted, abruptly reducing the supply of oxygen and glucose to the involved brain tissue. The resulting ischemic environment initiates a sequence of pathological changes that includes failure of cellular energy metabolism, oxidative injury, loss of ionic homeostasis, and eventual neuronal death through

necrotic and apoptotic pathways [3]. Several vascular mechanisms can produce this interruption in cerebral perfusion. Common causes include large-artery atherothrombotic occlusion, cerebral embolism, non-thrombotic occlusion of small deep cerebral arteries, and proximal stenosis with hypotension that reduces perfusion in the corresponding arterial territories [4].

Although thrombolysis and mechanical thrombectomy are established clinical treatments, their use is limited by a narrow therapeutic window, making the identification of additional therapeutic targets and strategies an important objective in IS research [5]. Entosis was initially described as a non-apoptotic death process in which one living cell is engulfed by another. A defining feature is the active entry of a viable cell into a neighboring homologous cell, producing a “cell-in-cell” or “cannibalism” morphology; the internalized cell is subsequently degraded and cleared through the lysosomal system of the host cell [6]. Entotic cell-in-cell formation depends on calcium-sensitive cadherin-mediated adhesion and Rho-associated kinase (ROCK) activity, after which non-canonical autophagic and lysosomal processes contribute to entotic cell death. Other reported regulators include AMP-activated protein kinase (AMPK), the stress-responsive kinases c-Jun N-terminal kinase (JNK) and p38, the oncogenes Kirsten rat sarcoma 2 viral oncogene homolog (KRAS) and cellular Myc (c-Myc), the p53 tumor suppressor protein, and additional proteins [7]. In an immunosuppressed patient with a cerebral small-vessel inflammatory lesion, immunohistochemistry and multiplex immunofluorescence clearly demonstrated a typical entosis event, indicating that CNS infection may be accompanied by an

abnormal phagocytic mode of cell death [8]. Nevertheless, most available evidence on entosis originates from tumor and neurodevelopmental settings, and it remains uncertain whether classical cell-in-cell structures develop after cerebral ischemia [7,9,10]. Reported stimuli capable of promoting entosis in different experimental systems include glucose starvation, UV irradiation, disruption of cell polarity, and P53 activation [11]. Glucose deprivation, in particular, has been shown to trigger entosis in adherent cells, where ingestion of a neighboring cell can provide a means of nutrient acquisition during metabolic stress [12]. By contrast, the contribution of entosis to ischemic pathogenesis and to oxygen-glucose deprivation injury in neural cells has not been established. Direct experimental evidence demonstrating classical cell-in-cell structures in brain tissue from IS is currently lacking. Accordingly, examining how entosis may relate to neuronal death and functional repair after IS could offer another perspective on ischemic pathogenesis through the concept of intracellular phagocytic injury promotion and may provide a theoretical basis for developing new approaches to brain protection.

This study integrated public IS transcriptomic data with weighted gene co-expression network analysis (WGCNA), differential expression, PPI network, and machine-learning methods to systematically pinpoint key genes linked to entosis-mediated cell-in-cell death. The selected genes were then assessed across independent datasets and evaluated for diagnostic performance, followed by analyses of functional enrichment, relationships with immune infiltration, and potential post-transcriptional regulatory networks. Drug prediction and molecular docking were

further used to explore candidate small-molecule compounds. Through this analytical framework, we aimed to investigate the molecular regulatory features of IS from the perspective of entosis-related cell-in-cell death and to provide a theoretical and experimental basis for future work on early diagnosis, molecular typing, and targeted treatment of the disease.

2. Methods

2.1 Data collection

As the training set, GSE58294 (GPL570) was downloaded from GEO [13] and included 69 IS whole-blood samples plus 23 controls. For validation, we used GSE122709 (GPL20795), which provided PBMC specimens from 5 IS patients and 5 healthy controls. In addition, 95 entosis-related genes (ERGs) were assembled by taking the union of gene sets reported in two previous publications [14,15] (**Additional file 1**).

2.2 Identification of differentially expressed genes (DEGs)

Differential expression in GSE58294 (IS vs. controls) was analysed using limma (v3.54.0) [16], defining DEGs as $|\log_2 \text{fold change (FC)}| > 0.5$ and adjusted $p < 0.05$. A volcano plot (ggplot2 v3.5.1) [17] displayed global distribution, with the top 10 up/down genes annotated by $\log_2$FC; a heatmap (ComplexHeatmap v2.16.0) [18] showed expression patterns of the top 20 DEGs.

2.3 WGCNA

Single-sample gene set enrichment analysis (ssGSEA) was first carried out with the "GSVA" package (v 1.53.28) [19] to quantify sample-level variation in ERG

enrichment. An ERG score was obtained for every sample, and scores were compared between the IS and control groups ($P < 0.05$). WGCNA was then applied to all samples in the training set with the "WGCNA" package (v 1.73) [20] to identify modules associated with the ERG score. Potential outliers were examined by sample clustering using the "hclust" function. The soft-thresholding power was selected via pickSoftThreshold to achieve approximate scale-free topology ($R^2 > 0.8$). An adjacency matrix was built, converted to a TOM, and genes were hierarchically clustered; co-expression modules were then defined using dynamic tree-cut. Module–ERG score correlations and p-values were computed using WGCNA's cor function (v1.73). Gene significance (GS) and module membership (MM) were computed for the key modules, and all genes within them were designated ERG-associated module genes for downstream analysis.

2.4 Identification, functional enrichment and protein–protein interaction (PPI) network of candidate genes

Candidate genes were obtained by intersecting the DEGs identified in GSE58294 with the ERG-associated module genes derived from WGCNA, and this overlap was illustrated using the "VennDiagram" package (v 1.7.1) [21]. To characterize the biological processes and pathways represented by these candidates, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment were performed using clusterProfiler (v4.15.0.3) [22] with adjusted $p < 0.05$ as the significance threshold. Protein-level relationships among the candidate genes were examined by querying the STRING [23] database with a confidence score > 0.7. The

resulting protein–protein interaction (PPI) network was displayed in Cytoscape software (v 3.10.2) [24]. For prioritization before machine-learning analysis, the maximum clique centrality (MCC) algorithm in the cytoHubba plugin was used to select the top 30 genes.

2.5 Machine learning for feature gene selection

Three machine-learning approaches were applied to the candidate genes in the GSE58294 training set to select features for downstream analysis.

The first approach used Boruta, a random-forest-based feature-selection algorithm, implemented with the "Boruta" package (v 8.0.0) [25]. Random permutation of the original variables was used to generate shadow features. Across repeated iterations, the importance assigned to each candidate gene was compared with the importance of the shadow variables, and genes whose importance was significantly greater than that of the shadow features were retained as Boruta feature genes.

The second approach was SVM-RFE, performed with the "e1071" package (v 1.7-16) [26]. Classification performance for alternative feature subsets was assessed by 5-fold cross-validation. The feature combination that provided the optimal balance according to accuracy and error rate was selected, and the genes contained in that subset were designated as SVM-RFE feature genes.

For the third approach, random forest (RF) analysis was implemented using the "randomForest" package (v 4.7-1.2) [27]. Model performance across different variable subsets was compared with 5-fold cross-validation on the basis of accuracy, area under the curve, and mean squared error. The top 10 genes by importance ranking were

retained as RF feature genes.

Feature genes selected independently by Boruta, SVM-RFE, and RF were then intersected. Genes shared by all three methods constituted the feature-gene set used in the subsequent analyses.

2.6 Receiver operating characteristic (ROC) analysis and expression validation

Diagnostic performance for each feature gene was examined by ROC curve analysis using the "pROC" package (v 1.18.5) [28]. The area under the curve (AUC) was calculated for both GSE58294 (training) and GSE122709 (validation); genes with AUC > 0.7 in both datasets were retained as candidate key genes. We assessed expression differences (IS vs. controls) in the training (GSE58294) and validation (GSE122709) cohorts ($p < 0.05$). Genes showing significant and concordant directional changes in both datasets were considered key genes and advanced to further study.

2.7 Nomogram construction and evaluation

Using the GSE58294 training set, a nomogram for estimating ischemic stroke risk was developed from the key genes with the "rms" package (v 6.8-2) [29]. Each key gene was assigned a point value according to its expression level. The individual values were summed to obtain a total score, which was then mapped to the estimated probability of IS.

Nomogram performance was evaluated by several complementary approaches. A calibration curve was generated with the "regplot" package (v 0.3.0) [30] to compare predicted probabilities with actual probabilities, and calibration was additionally assessed with the Hosmer-Lemeshow test. Clinical utility was examined by decision

curve analysis (DCA) using the "ggDCA" package (v 1.1) [31]. A net benefit > 0 was interpreted as evidence of a favorable predictive effect.

2.8 GeneMANIA analysis

GeneMANIA [32] was used to investigate functional connections and biological processes shared by the key genes. The key genes were entered as the query and the analysis was run with default settings. The resulting co-expression network was used to depict relationships among the queried genes and additional genes with related functions.

2.9 Chromosomal localization analysis

The chromosomal positions of the key genes were examined with the "RCircos" package (v 1.2.2) [33]. Their locations on human chromosomes were presented in a circular visualization in which individual chromosomes were distinguished by colour or texture, and the exact sites of the key genes were marked on the corresponding chromosome segments.

2.10 Gene set enrichment analysis (GSEA)

For each key gene, GSEA was conducted with the GSE58294 training dataset to examine associated biological pathways. Spearman correlations between the key gene and all other genes were calculated using psych (v2.4.6.26) [34], and genes were then ranked by descending correlation coefficient. The resulting pre-ranked list was analyzed with "clusterProfiler" (v 4.2.2) against the "h.all.v7.4.symbols.gmt" (Hallmarks) reference gene set obtained from the MSigDB [35] (|normalised enrichment score (NES)| > 1, false discovery rate (FDR) < 0.25, P < 0.05). For

visualization, the 5 most significantly enriched pathways associated with each key gene were retained.

2.11 Immune infiltration analysis

Immune-cell composition in IS was estimated in the GSE58294 training set by applying the CIBERSORT algorithm with its default parameters. Relative fractions of 22 immune cell subtypes [36] were calculated for each sample, and samples with a CIBERSORT output $P > 0.05$ were removed before subsequent analyses. In the remaining samples, the Wilcoxon rank-sum test ($p < 0.05$) compared immune-cell infiltration between IS and controls for each subtype. Associations between key genes and infiltration were then assessed in the training set using psych (v2.2.9) [34]. We correlated key-gene expression levels with the relative abundances of intergroup-differential immune subtypes, applying significance thresholds of $p < 0.05$ and $|\text{correlation coefficient (cor)}| > 0.3$.

2.12 Construction of molecular regulatory networks

Potential post-transcriptional regulation of the key genes was examined by first predicting targeting miRNAs with TargetScan and the miRDB. Only miRNAs identified by both resources were retained as candidate miRNAs. lncRNAs interacting upstream with these candidate miRNAs were then retrieved from starBase, and interactions were kept only when supported by more than four CLIP-seq experiments (clipExpNum > 4). These relationships were assembled into an lncRNA-miRNA-mRNA ceRNA network and visualised in Cytoscape software (v 3.10.2). lncRNA subcellular localisation was queried via lncATLAS [37], while

transcription factors (TFs) for the key genes were retrieved from hTFtarget [38]. A TF–mRNA regulatory network was then built and visualised in Cytoscape (v3.10.2).

2.13 Drug prediction

The Drug Signatures Database (DsigDB) [39] was queried to predict compounds that might target the key genes. Predicted drug-gene relationships were organized into a drug-gene interaction network and visualised in Cytoscape software (v 3.10.2). Candidate drugs were prioritized through iterative threshold filtering according to three network centrality measures: betweenness centrality (BC), closeness centrality (CC), and degree centrality (DC). The thresholds were progressively increased over successive filtering rounds. The compound remaining after application of the final criteria ($DC \geq 2$, $BC \geq 0.01$, $CC \geq 0.5$) was designated the top candidate drug.

2.14 Molecular docking

Molecular docking assessed binding affinities between selected compounds and key genes. Receptor protein structures were obtained from UniProt [40] (sequences) and AlphaFold [41] (3D models); ligand 3D structures were downloaded from PubChem [42]. Docking was performed with AutoDock software (v 4.2.6) [43] together with MGLTools (v 1.5.7), and the receptor-ligand binding energy was calculated for each pair. Binding energies below -5 kcal/mol were considered indicative of a stable interaction.

2.15 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

Whole-blood samples were obtained from 5 patients with IS and 5 controls and were subsequently used for RT-qPCR analysis. The study was approved by the Ethics

Committee of Xi'an Central Hospital Affiliated to Xi'an Jiaotong University (approval no. LW-2026-002), and written informed consent was provided by every participant. RT-qPCR was used to validate NOL9 and WDR3 expression. Total RNA from whole blood was extracted with TRIzol (R401-01, Vazyme), reverse-transcribed using Hifair® III cDNA Synthesis SuperMix (Yeasen), and amplified with 2× Universal Blue SYBR Green Master Mix (Servicebio). Primers are listed in **Table 1**; GAPDH served as the reference. Relative expression ($2^{-\Delta\Delta Ct}$) was compared between IS and controls by unpaired two-tailed t-test ($p < 0.05$).

Table 1. Primer sequences used for RT-qPCR.

Primers	Sequences
NOL9 F	TGAGAGTACCCTTTCAGCCCT
NOL9 R	AGAAATGCAACCAGGAGGGG
WDR3 F	CAGGCTTGGGGGTTAGAGG
WDR3 R	TGGTTTGGGGCTTGACCATA
GAPDH F	ATGGGCAGCCGTTAGGAAAG
GAPDH R	AGGAAAAGCATCACCCGGAG

2.16 Statistical analysis

All bioinformatics analyses used R (v4.2.2); group differences were tested by Wilcoxon rank-sum, with $p < 0.05$ considered significant.

3. Results

3.1 Identification of entosis-related module genes by WGCNA

The ERG score calculated by ssGSEA was significantly elevated in IS samples relative

to controls ($P < 0.05$) (**Figure 1a**). WGCNA was then conducted using all samples in the GSE58294 training set. Hierarchical clustering of samples detected no outliers, so no samples were removed (**Figure 1b**). A soft-thresholding power of 14 was chosen because the scale-free topology criterion reached $R^2 > 0.8$ (**Figure 1c**). Based on hierarchical clustering of the TOM, 12 co-expression modules were obtained (**Figure 1d**). When correlating module eigengenes with the ERG score, the pink module showed the strongest positive correlation (cor = 0.78, $p < 0.001$), while the purple module gave the strongest negative correlation (cor = -0.68, $p < 0.001$) (**Figure 1e**). The pink and purple modules were therefore retained as key modules, yielding 1,712 genes that were designated ERG-associated module genes (**Additional file 2**). In both modules, scatter plots of GS against MM demonstrated positive relationships, indicating that genes more closely related to the ERG score also tended to be more highly connected within their respective modules (**Figure 1f**). Together, these findings showed a global increase in entosis-related gene activity in IS and identified co-expression modules closely linked to the ERG score for subsequent analyses.

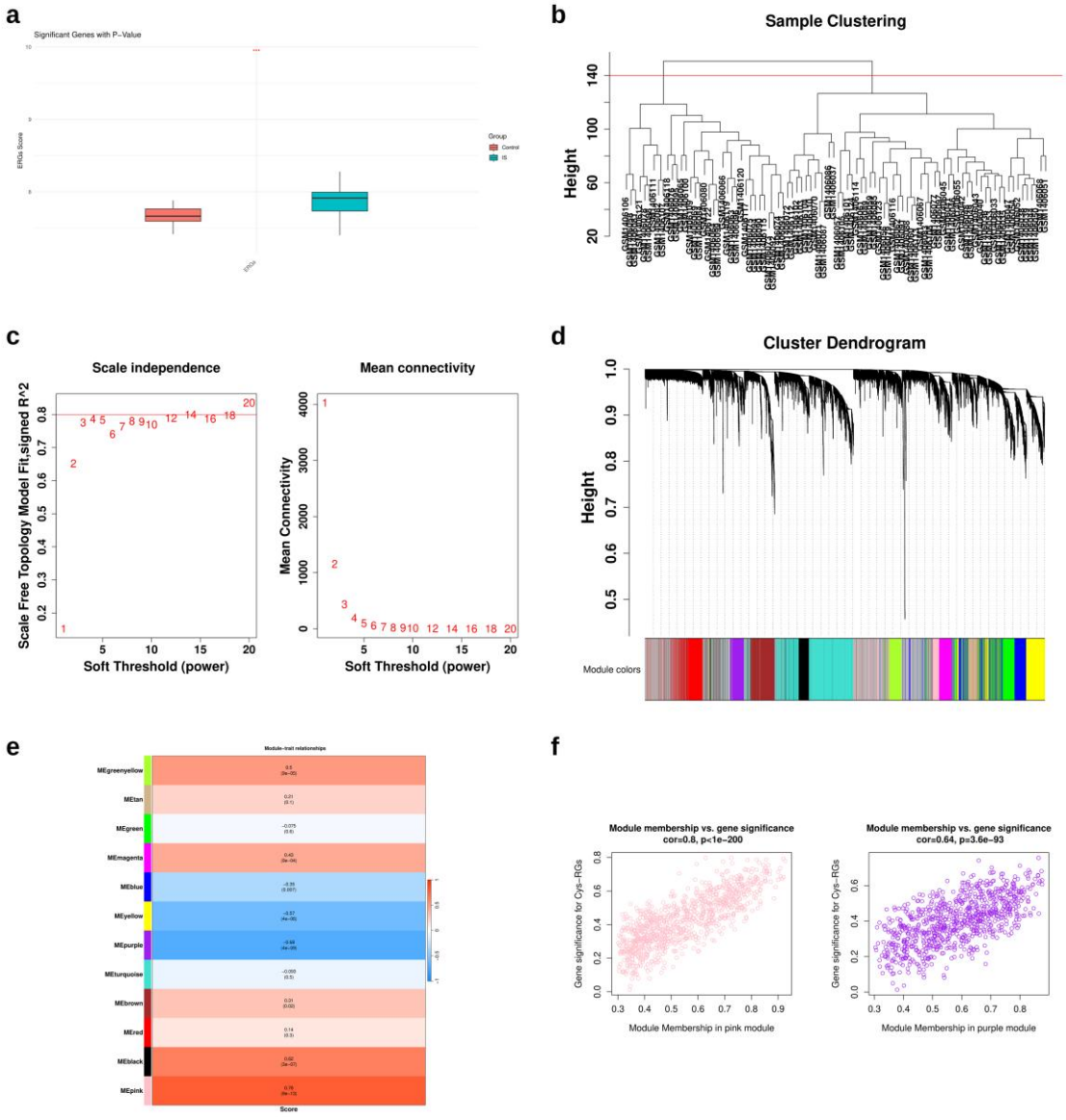


Figure 1. Identification of entosis-related module genes by WGCNA.

(a) ERG score was significantly higher in IS than in controls (Wilcoxon rank-sum test, $P < 0.05$).

(b) Sample clustering dendrogram showed no outlier samples.

(c) Soft-thresholding power of 14 was selected (scale-free $R^2 > 0.8$).

(d) Hierarchical clustering identified 12 co-expression modules.

(e) Correlation between modules and ERG score: pink module (cor = 0.78, $P < 0.001$), purple module (cor = -0.68, $P < 0.001$).

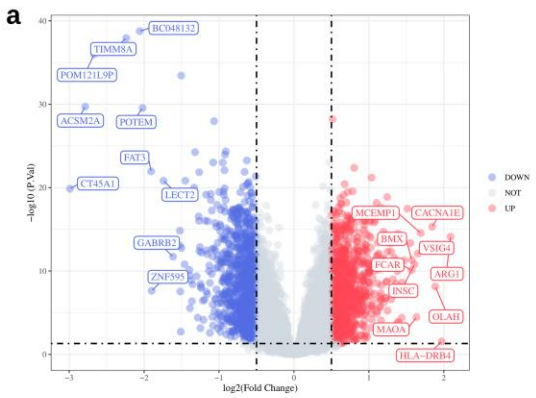
(f) Scatter plots of gene significance (GS) vs. module membership (MM) showed positive correlations in both key modules ($P < 0.001$).

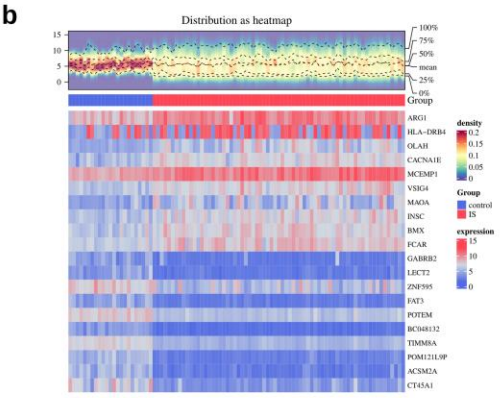
3.2 Identification of candidate genes and functional enrichment

Differential expression analysis yielded 1,987 DEGs (951 up-regulated, 1,036 down-regulated) at adjusted $p < 0.05$ and $|\log_2 FC| > 0.5$ (**Figure 2a-b, Additional file 3**). Intersecting these 1,987 DEGs with the 1,712 ERG-associated module genes from WGCNA produced 429 candidate genes (**Figure 2c, Additional file 4**).

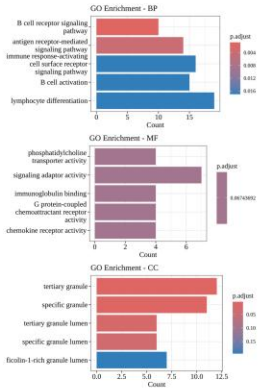
GO enrichment (adjusted $p < 0.05$) of the 429 candidate genes yielded 3,017 BP, 316 CC, and 527 MF terms. The top five BP terms were "B cell receptor signalling", "antigen receptor-mediated signalling", "immune response-activating cell surface receptor signalling", "B cell activation", and "lymphocyte differentiation" (**Additional file 5**). For CC, the top five terms were "tertiary granule", "specific granule", "tertiary granule lumen", "specific granule lumen", and "ficolin-1-rich granule lumen" (**Additional file 6**). The five leading MF terms consisted of "phosphatidylcholine transporter activity", "signaling adaptor activity", "immunoglobulin binding", "G protein-coupled chemoattractant receptor activity", and "chemokine receptor activity" (**Figure 2d, Additional file 7**). KEGG analysis identified 239 significantly enriched pathways (adjusted $p < 0.05$). The top ten were "B cell receptor signalling", "axon guidance", "haematopoietic cell lineage", "primary immunodeficiency", "cell adhesion molecules", "nucleotide metabolism", "purine metabolism", "signalling pathways regulating stem cell pluripotency", "Fc gamma R-mediated phagocytosis", and "viral protein–cytokine/cytokine receptor interaction" (**Figure 2e, Additional file 8**).

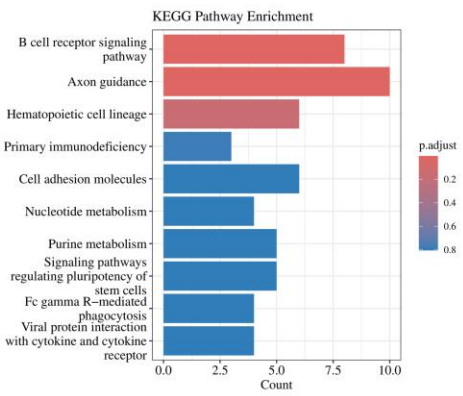
The PPI analysis generated a network containing 74 nodes and 101 edges at confidence > 0.7 (**Figure 2f**). Genes were ranked by the MCC algorithm in cytoHubba to select those for later analysis. Although the initial cutoff was the top 30 genes, the 31st-ranked gene had an MCC value identical to that of the 30th-ranked gene and was therefore also retained, giving 31 candidate genes in total (**Figure 2g, Additional file 9**). Overall, the enrichment patterns linked these candidate genes mainly to immune-associated pathways and cellular activation, providing a basis for further examination of the relationship between ERG-associated transcriptional features and immune responses in IS.

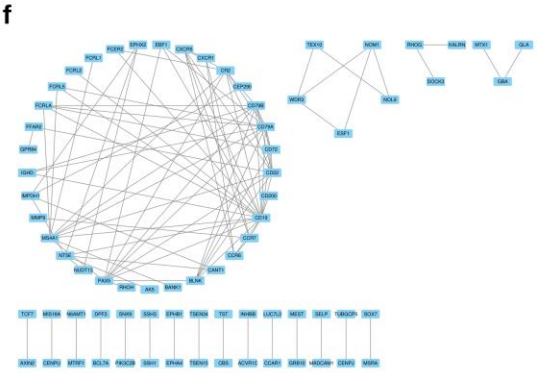

a
BC048132
TIMM8A
POM121L9P
ACSM2A
POTEM
FAT3
CT45A1
LECT2
GABRB2
ZNF595
MCEMP1
CACNA1E
BMX
VSIG4
FCAR
ARG1
INSC
OLAH
MAOA
HLA−DRB4
DOWN
NOT
UP
−log10 (P.Val)
log2(Fold Change)

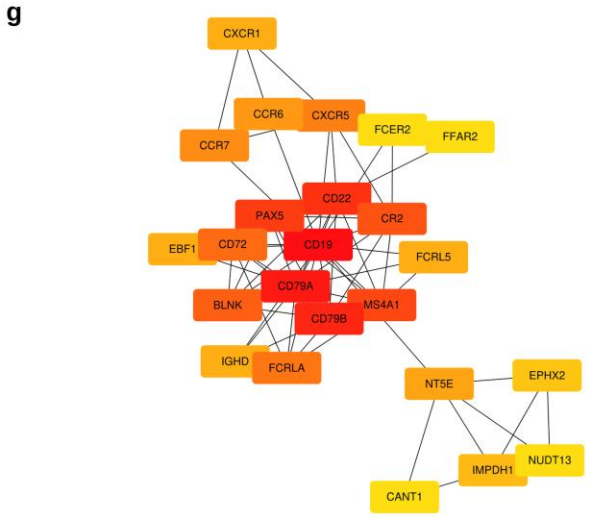

b
Distribution as heatmap
100%
75%
50%
mean
25%
0%
Group
ARG1
HLA−DRB4
OLAH
CACNA1E
MCEMP1
VSIG4
MAOA
INSC
BMX
FCAR
GABRB2
LECT2
ZNF595
FAT3
POTEM
BC048132
TIMM8A
POM121L9P
ACSM2A
CT45A1
density
Group
control
IS
expression


c
ERGs
DEG
1283
429
1558


d
GO Enrichment - BP
B cell receptor signaling pathway
antigen receptor-mediated signaling pathway
immune response-activating cell surface receptor signaling pathway
B cell activation
lymphocyte differentiation
p.adjust
Count
GO Enrichment - MF
phosphatidylcholine transporter activity
signaling adaptor activity
immunoglobulin binding
G protein-coupled chemoattractant receptor activity
chemokine receptor activity
GO Enrichment - CC
tertiary granule
specific granule
tertiary granule lumen
specific granule lumen
ficolin-1-rich granule lumen


e
KEGG Pathway Enrichment
B cell receptor signaling pathway
Axon guidance
Hematopoietic cell lineage
Primary immunodeficiency
Cell adhesion molecules
Nucleotide metabolism
Purine metabolism
Signaling pathways regulating pluripotency of stem cells
Fc gamma R−mediated phagocytosis
Viral protein interaction with cytokine and cytokine receptor
p.adjust
Count


f


g
CXCR1
CCR6
CXCR5
FCER2
FFAR2
CCR7
CD22
PAX5
CR2
EBF1
CD72
CD19
FCRL5
CD79A
BLNK
MS4A1
CD79B
IGHD
FCRLA
NT5E
EPHX2
NUDT13
IMPDH1
CANT1
NOM1
RHOG
GBA
ESF1
WDR3
NOL9

334

Figure 2. Identification of candidate genes and functional enrichment.

(a) Volcano plot of DEGs ($|\log_2 FC| > 0.5$, adjusted $P < 0.05$).

(b) Heatmap of top 20 DEGs.

(c) Venn diagram showing 429 candidate genes intersecting 1,987 DEGs and 1,712 ERG-associated module genes.

(d) Top 5 GO terms (adjusted $P < 0.05$).

(e) Top 10 KEGG pathways (adjusted $P < 0.05$).

(f) PPI network of candidate proteins (STRING confidence > 0.7).

(g) Top 31 candidate genes selected by MCC algorithm (cytoHubba).

3.3 Identification of eight feature genes by machine learning

Machine-learning selection was next performed in GSE58294 using the 31 candidate genes prioritized from the PPI network. Boruta retained 22 feature genes (**Figure 3a, Additional file 10**). With 5-fold cross-validation, SVM-RFE achieved an accuracy of 0.755 and an error rate of 0.245 and selected 24 feature genes (**Figure 3b-c, Additional file 11**). After tuning the random forest parameters ntree and mtry, we retained the 10 genes with the highest importance scores (**Figure 3d-e, Additional file 12**). Intersecting the three feature sets identified 8 genes shared by Boruta, SVM-RFE, and random forest: CXCR1, RHOG, CANT1, NOL9, CCR7, WDR3, GBA, and IMPDH1 (**Figure 3f**).

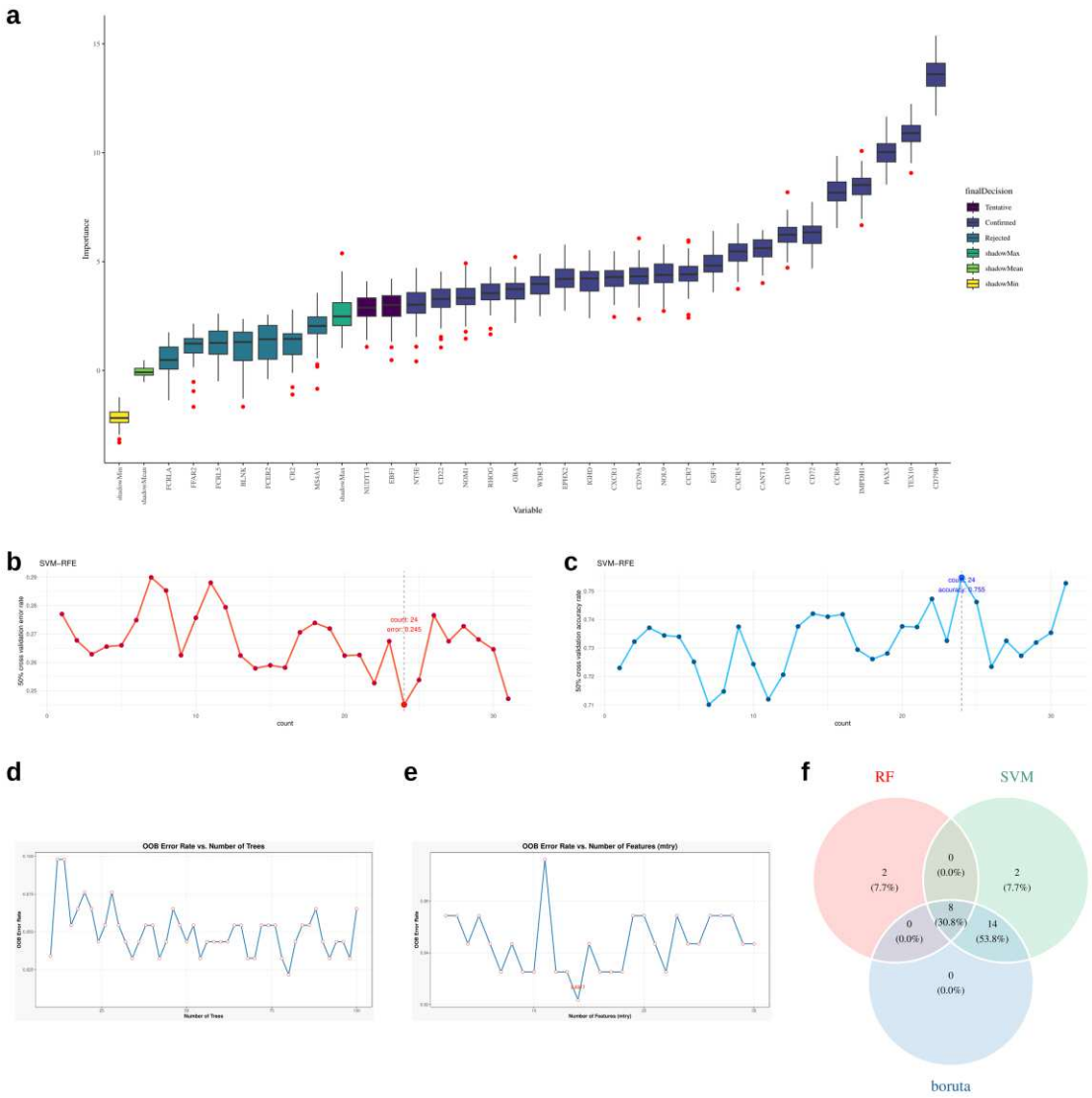


Figure 3. Identification of eight feature genes by machine learning.

(a) Boruta algorithm identified 22 feature genes (importance > shadow features).

(b-c) SVM-RFE with 5-fold cross-validation achieved accuracy 0.755 and error rate 0.245, selecting 24 feature genes.

(d-e) Random forest selected the top 10 feature genes by importance (5-fold cross-validation).

(f) Intersection of the three algorithms yielded 8 feature genes (CXCR1, RHOG, CANT1, NOL9, CCR7, WDR3, GBA, IMPDH1).

3.4 Identification and validation of NOL9 and WDR3 as key genes

ROC analysis of the 8 feature genes in both GSE58294 (training) and GSE122709 (validation) showed that CANT1, CCR7, CXCR1, GBA, IMPDH1, NOL9, RHOG, and WDR3 all achieved AUC > 0.7 in the training set (**Figure 4a**). In the validation set (GSE122709), AUC values above 0.7 were obtained for five genes: CANT1, CXCR1, IMPDH1, NOL9, and WDR3 (**Figure 4b**). Expression of these five candidate key genes was then assessed in the training set (**Figure 4c**) and independently in the validation set (**Figure 4d**). Only NOL9 and WDR3 showed significant differential expression in the same direction between IS and control samples in both datasets ($P < 0.05$), and they were consequently retained as the key genes for further analysis. Chromosomal mapping placed both NOL9 and WDR3 on chromosome 1 (**Figure 4e**). Thus, the two genes displayed consistent upregulation and robust diagnostic performance across the two independent transcriptomic datasets.

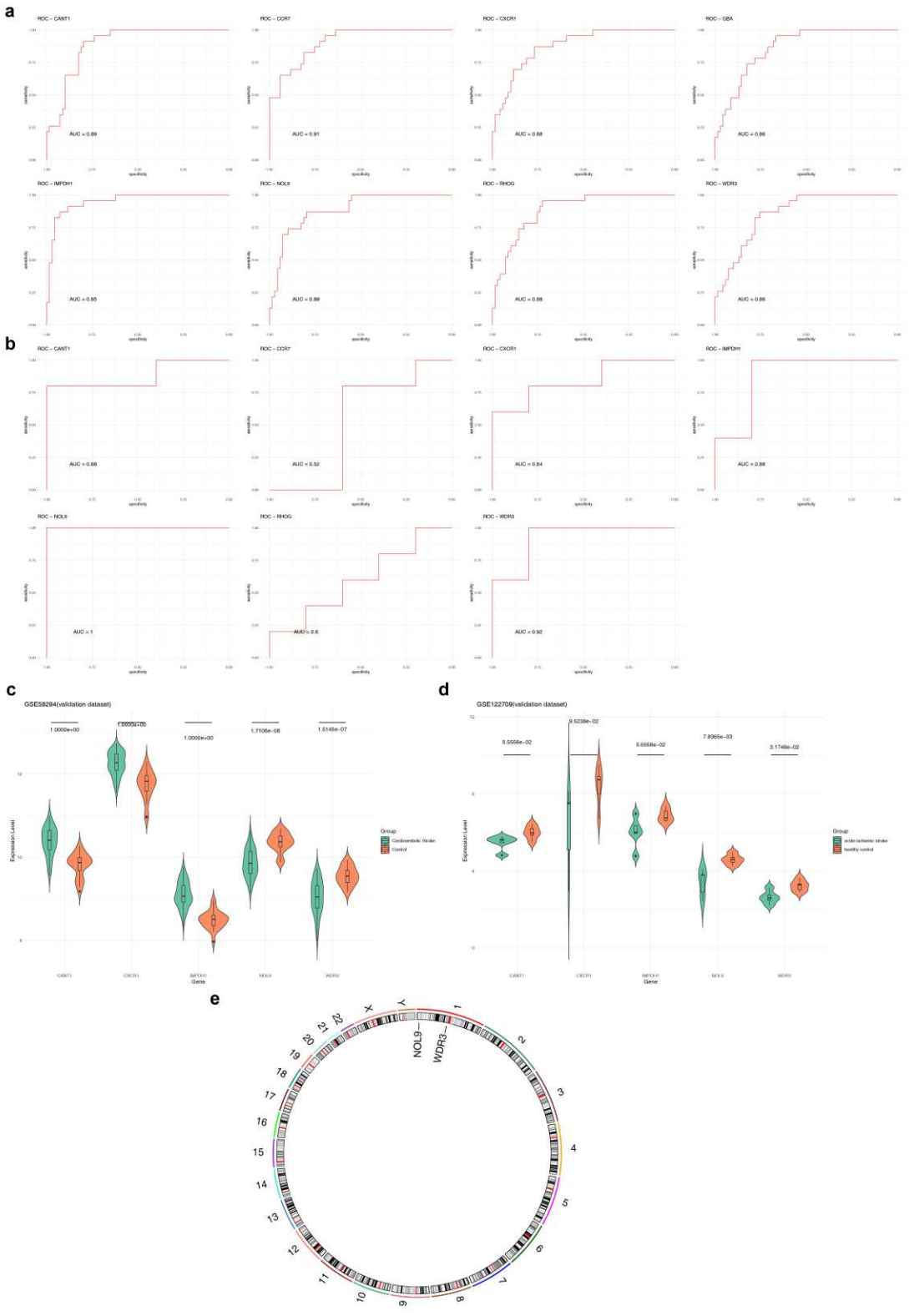


Figure 4. Identification and validation of NOL9 and WDR3 as key genes.

(a) ROC curves in the training set (GSE58294): 8 genes with AUC > 0.7.

(b) ROC curves in the validation set (GSE122709): 7 genes with AUC > 0.7.

(c) Expression validation in the training set GSE58294 (Wilcoxon rank-sum test; $P < 0.05$ for NOL9 and WDR3).

(d) Expression validation in the validation set GSE122709 ($P < 0.05$ for NOL9 and WDR3).

(e) Chromosomal localization: both NOL9 and WDR3 are located on chromosome 1.

3.5 Pathway enrichment and functional network analysis of NOL9 and WDR3

GSEA against the Hallmark gene set (h.all.v7.4.symbols.gmt) with |NES|>1, FDR<0.25, and p<0.05 identified 91 significant pathways for NOL9; the top five were "ribosome biogenesis in eukaryotes", "neuroactive ligand–receptor interaction", "olfactory transduction", "keratinization", and "ribosome" (**Figure 5a, Additional file 13**). WDR3 was associated with 107 significantly enriched pathways, with "ribosome", "neuroactive ligand-receptor interaction", "neuroactive ligand signaling", "spliceosome", and "keratinization" forming the top five (**Figure 5b, Additional file 14**).

Functional relationships involving NOL9 and WDR3 were additionally examined with GeneMANIA. The generated network connected both key genes with multiple genes participating in RNA metabolism and ribosome assembly. NOL9, WDR3, and LAS1L appeared as prominent nodes within the network, consistent with their involvement in ribosome biogenesis and rRNA processing (**Figure 5c**). Taken together, the pathway and network results associated NOL9 primarily with ribosome-biogenesis processes and WDR3 with spliceosome-related processes, suggesting distinct links of the two genes to RNA metabolism.

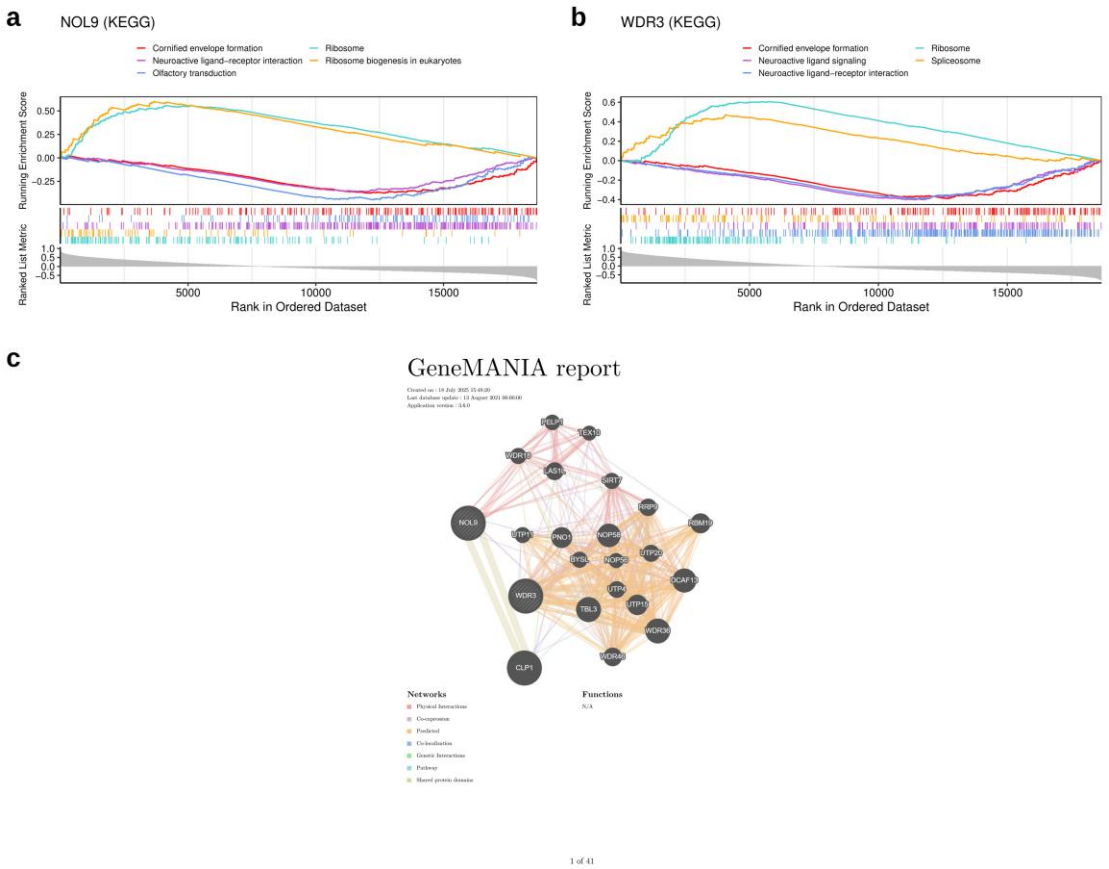


Figure 5. Pathway enrichment and functional network analysis of NOL9 and WDR3.

(a) GSEA for NOL9: top 5 significantly enriched pathways (|NES| > 1, FDR < 0.25, P < 0.05).

(b) GSEA for WDR3: top 5 significantly enriched pathways using the same thresholds.

(c) GeneMANIA functional network (default parameters; co-expression and shared pathways shown).

3.6 Nomogram construction and evaluation

A nomogram incorporating NOL9 and WDR3 was generated from the GSE58294 training set to estimate the risk of ischemic stroke. Each gene contributed a point value according to its expression, and the summed score was converted to the predicted probability of IS. As an example, 45.1 total points corresponded to an estimated IS risk of 57% (**Figure 6a**). Model calibration was evaluated by plotting predicted against

observed probabilities. The Hosmer-Lemeshow test returned P = 0.653, showing no statistically significant discrepancy between predicted and observed outcomes, while the calibration curve demonstrated close agreement between predicted and actual probabilities (P > 0.05) (**Figure 6b**). DCA further showed a net benefit above zero over a broad interval of threshold probabilities, and the nomogram curve remained above the two extreme-reference curves, indicating favorable clinical predictive performance (**Figure 6c**). Accordingly, within the training dataset, the NOL9/WDR3 nomogram displayed favorable apparent calibration and net benefit.

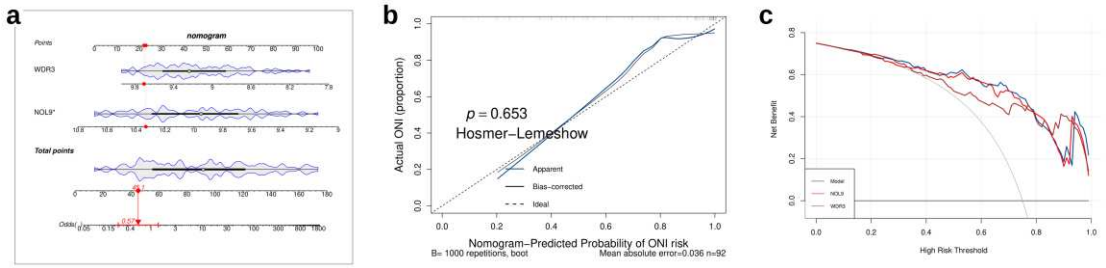


Figure 6. Nomogram construction and evaluation.

(a) Nomogram for predicting IS risk based on NOL9 and WDR3 expression.

(b) Calibration curve (Hosmer-Lemeshow test P = 0.653).

(c) Decision curve analysis showing net benefit above zero across a broad threshold range.

3.7 Immune infiltration analysis

CIBERSORT was used to estimate the relative abundance of 22 immune cell subtypes in the GSE58294 training cohort, and their distributions across IS and control samples were visualized in a stacked bar plot (**Figure 7a**). Eight immune subtypes differed significantly between groups ($p < 0.05$) (**Figure 7b**). IS samples showed lower naive B cells, naive CD4 T cells, resting CD4 memory T cells, and M2 macrophages, but higher

activated CD4 memory T cells, monocytes, M0 macrophages, and neutrophils. Spearman analysis of relationships among immune-cell subtypes identified a positive correlation between M0 macrophages and neutrophils (cor = 0.36, P < 0.001), an inverse correlation between M2 macrophages and monocytes (cor = -0.45, P < 0.001), and an inverse correlation between naive B cells and neutrophils (cor = -0.48, P < 0.001) (**Figure 7c, Additional file 15**). Correlating the two key genes with differential immune subtypes, M0 macrophages showed negative correlations with NOL9 (cor = -0.60, p < 0.001) and WDR3 (cor = -0.57, p < 0.001), while naive B cells correlated positively with NOL9 (cor = 0.46, p < 0.001) and WDR3 (cor = 0.35, p < 0.001) (**Figure 7d, Additional file 16**). These results identified associations of NOL9/WDR3 expression with the estimated proportions of M0 macrophages and naive B cells.

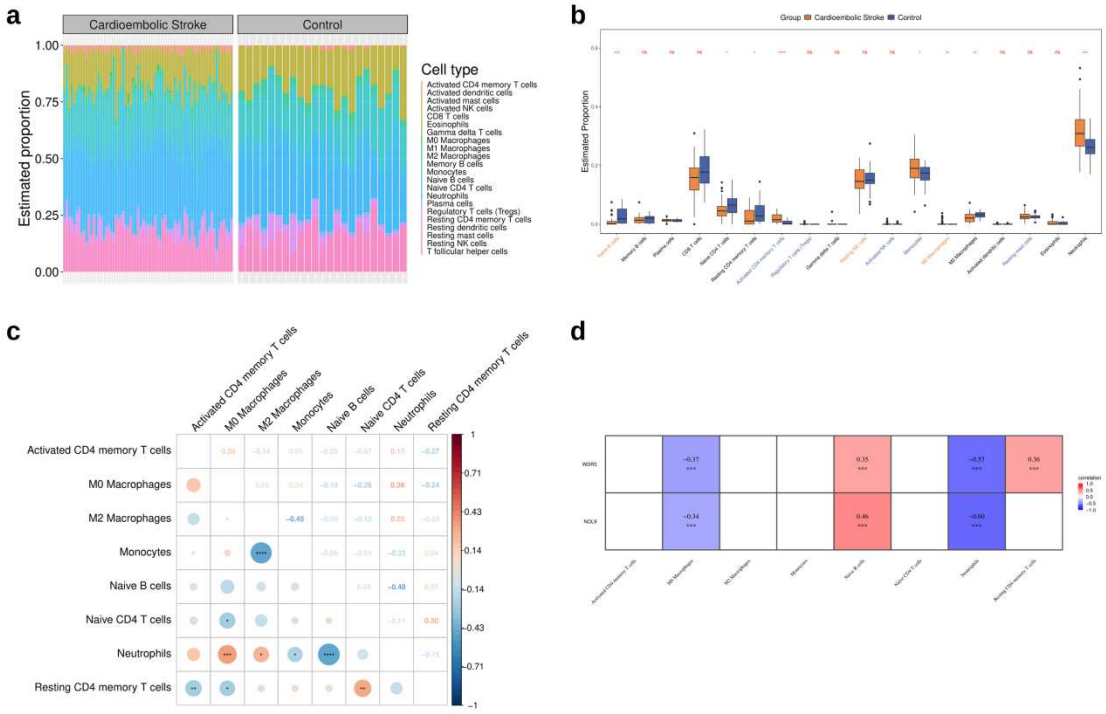


Figure 7. Immune infiltration analysis.

(a) Stacked bar plot of 22 immune cell subtypes (CIBERSORT; samples with P > 0.05

449 excluded).

450 (b) Eight immune cell subtypes significantly differed between IS and controls
451 (Wilcoxon rank-sum test, $P < 0.05$).

452 (c) Correlation matrix among differential immune subtypes (Spearman analysis).

453 (d) Correlations of NOL9 and WDR3 with differential immune subtypes, including M0
454 macrophages and naive B cells.

455 **3.8 Molecular regulatory network construction**

456 To characterize putative transcriptional and post-transcriptional regulation of NOL9
457 and WDR3 in IS, a multilayer molecular network was assembled. Comparison of the
458 two miRNA-prediction databases yielded six high-confidence miRNAs targeting the
459 key genes: hsa-miR-8485, hsa-miR-3155b, hsa-miR-3155a, hsa-miR-484,
460 hsa-miR-627-3p, and hsa-miR-499a-5p. Candidate upstream lncRNAs were obtained
461 from starBase and filtered at clipExpNum > 4. Under this criterion, only
462 hsa-miR-499a-5p was connected to three core lncRNAs, XIST, MALAT1, and NEAT1
463 (**Figure 8a**). lncATLAS analysis showed that XIST was predominantly localized in
464 GM12878, H1.hESC, MCF.7, SK.N.DZ, and SK.N.SH cell lines; MALAT1 was
465 mainly distributed in GM12878, H1.hESC, HepG2, HT1080, and HUVEC cell lines;
466 and NEAT1 was primarily localized in GM12878, H1.hESC, HeLa.S3, HepG2, and
467 HT1080 cell lines (**Figure 8b**). The TF analysis further yielded 398 TFs for NOL9,
468 including SPI1, CTCF, BCL3, and BRCA1, and 331 TFs for WDR3, including SPI1,
469 CTCF, BRD4, CBFB, and CDK9 (**Figure 8c, Additional file 17**). These predictions
470 together defined a putative ceRNA network containing NOL9, WDR3,

hsa-miR-499a-5p, XIST, MALAT1, NEAT1, and numerous transcription factors.

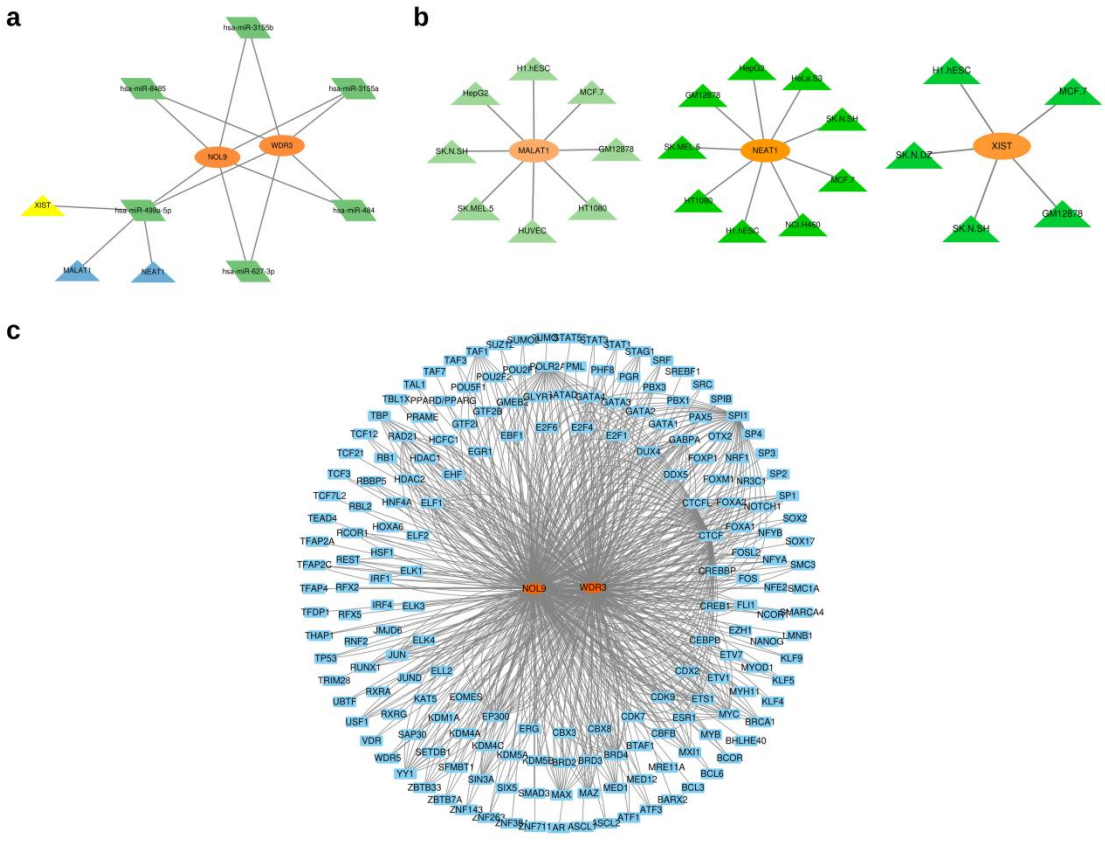


Figure 8. Molecular regulatory network construction.

(a) ceRNA network: six miRNAs targeting NOL9/WDR3; hsa-miR-499a-5p was linked to XIST, MALAT1, and NEAT1 (clipExpNum > 4).

(b) Subcellular localization of XIST, MALAT1, and NEAT1 in lncATLAS.

(c) Transcription factor prediction: 398 TFs for NOL9 and 331 TFs for WDR3 (hTFtarget).

3.9 Drug prediction and molecular docking

Visualization of predicted drug-gene relationships in Cytoscape identified 61 candidate compounds, including quinpirole, anisomycin, digoxigenin, and digitoxigenin (**Figure 9a, Additional file 18**). Molecular docking estimated a binding energy of -9.82 kcal/mol between digoxigenin and NOL9 (**Figure 9b**) and -10.49 kcal/mol between

digoxigenin and WDR3 (**Figure 9c**); both met the binding-energy screening criterion of ≤ -5.0 kcal/mol. These docking values indicated predicted affinity of digoxigenin for both key proteins, and digoxigenin was therefore prioritized as the top candidate compound.

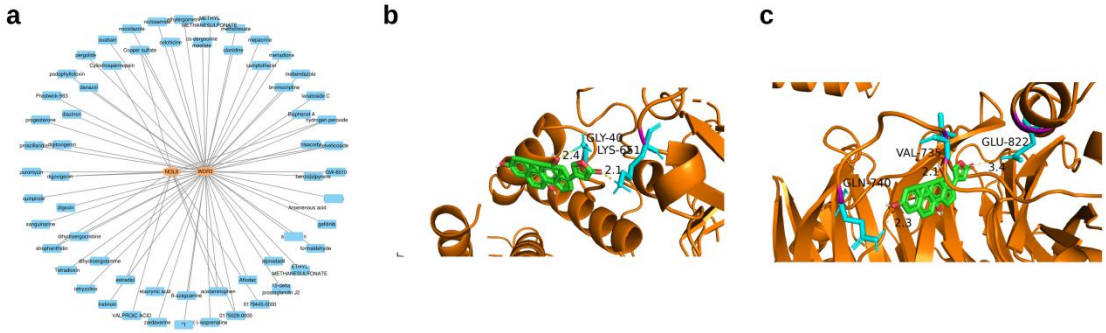


Figure 9. Drug prediction and molecular docking.

(a) Drug-gene interaction network containing 61 candidate compounds predicted by DsigDB.

(b) Molecular docking of digoxigenin with NOL9 (predicted binding energy -9.82 kcal/mol).

(c) Molecular docking of digoxigenin with WDR3 (predicted binding energy -10.49 kcal/mol).

3.10 Experimental validation of NOL9 and WDR3 expression by RT-qPCR

RT-qPCR demonstrated significantly lower expression of NOL9 and WDR3 in patients with IS than in healthy controls (**Figure 10**). Because this experimental direction was opposite to the upregulation observed in GSE58294 and GSE122709, the RT-qPCR findings do not independently validate the transcriptomic expression pattern and should be replicated in larger cohorts matched for sampling time.

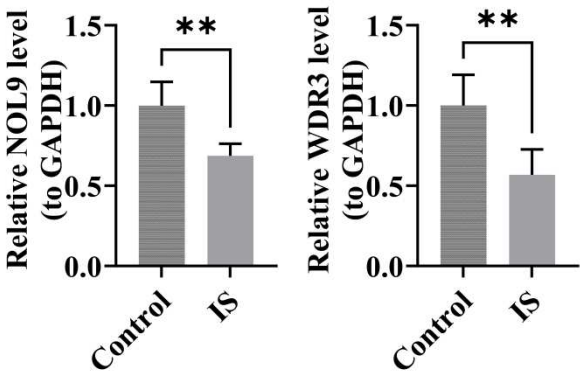


Figure 10. Experimental validation of NOL9 and WDR3 expression by RT-qPCR

4. Discussion

No previous study has directly established a role for entosis in IS. Although entosis-like events reported in neurodevelopmental settings make such a process biologically plausible, those observations cannot be considered evidence that classical entosis occurs in ischemic brain tissue [7,9]. Against this background, the present study used a series of complementary analyses to examine how entosis-related molecular features might be associated with IS. Gene-expression screening was combined with machine learning and expression validation to prioritize NOL9 and WDR3 as two IS-related key genes. We then examined the functional enrichment patterns linked to these genes to characterize biological processes with which they may be associated in the disease context. The immune-infiltration analysis further identified M0 macrophages as a cell subtype differing in IS and showed a significant inverse relationship between NOL9 and the estimated abundance of M0 macrophages. Drug-prediction analysis additionally yielded several compounds with potential relevance to IS, including quinpirole, anisomycin, and digoxigenin. Collectively, these analyses provide hypotheses and candidate molecular directions for further investigation of IS

pathogenesis and therapeutic targets, while the computational associations themselves do not establish direct biological causality.

NOL9 is a nucleolar protein that contributes to biogenesis of the large ribosomal subunit by participating in pre-rRNA processing and thereby supporting production of functional ribosomal subunits [44]. In the two public transcriptomic datasets analyzed here, NOL9 expression was increased in IS, whereas the RT-qPCR experiment showed lower expression in patients than in controls. The direction of NOL9 dysregulation in IS therefore remains unresolved. Increased NOL9 expression has previously been described in hepatocellular carcinoma, where it has been associated with tumor size, pathological grade, proliferative activity, and cell-cycle-related phenotypes [45]. Those observations are consistent with a relationship between NOL9 and increased biosynthetic demand, but the proliferative biology of a tumor differs substantially from the acute metabolic disturbance, inflammatory response, and multicellular stress that characterize ischemic stroke; the cancer findings therefore cannot be directly transferred to IS. A possible interpretation is that altered NOL9 expression accompanies stress-related remodeling of ribosome biogenesis, although the present results cannot determine whether such a response would be adaptive or harmful. Consistent with this possibility, GSEA linked NOL9-related genes to the “ribosome” and “ribosome biogenesis in eukaryotes” pathways. Existing work indicates that NOL9 forms a ribosome-processing complex with LAS1L and participates in nucleolar processing and post-transcriptional maturation of pre-rRNA, making it an important component of normal ribosome biosynthesis. Li et al. further reported that USP36

542 regulates the LAS1L-NOL9 endonuclease-kinase complex through sumoylation
543 modification, thereby influencing processing of pre-rRNA ITS2 and ribosome
544 production; this finding expands the regulatory framework connecting NOL9-LAS1L
545 activity with ribosomal RNA maturation [46]. On this basis, the increased NOL9
546 expression observed in the transcriptomic datasets could potentially support neuronal
547 repair after ischemia and hypoxia through enhanced rRNA processing efficiency and
548 ribosome production. However, given the opposite RT-qPCR result, this proposed
549 neuroprotective interpretation remains a hypothesis that requires direct experimental
550 testing.
551 WDR3 (also known as DIP2 and UTP12), a eukaryotic WD-repeat family member,
552 encodes a 943-aa nuclear protein with a 10-WD-repeat module, maps to 1p12-p13, and
553 is involved in ribosome biogenesis and p53-mediated cell-cycle control [47]. Loss of
554 WDR3 can markedly disrupt ribosomal biogenesis, producing pronounced defects in
555 18S rRNA processing, a transient reduction in precursor rRNA abundance, and
556 moderate suppression of RNA polymerase I activity [48]. To date, little evidence
557 directly addresses the relationship between WDR3 and ischemic stroke. In the current
558 analysis, WDR3 was significantly upregulated in peripheral-blood transcriptomic
559 profiles from IS patients, and the validation dataset showed the same
560 differential-expression direction. GSEA also associated WDR3 with the “Spliceosome”
561 signaling pathway in IS samples. Previous work in mice after ischemic brain injury has
562 reported increased expression of cytoplasmic ribosomal proteins together with proteins
563 of the spliceosome complex, suggesting that molecular changes involving both

translation and splicing occur after stroke [49]. These observations make WDR3 and RNA-processing pathways relevant candidates for further study in ischemic injury, but they do not demonstrate that WDR3 itself causally regulates spliceosome dysfunction. The spliceosome is a large, multimegadalton ribonucleoprotein (RNP) assembly composed of five small nuclear RNPs together with numerous accessory factors. Because most eukaryotic genes are initially transcribed as precursor mRNAs, splicing is required to generate mature transcripts, while alternative splicing increases the range of protein isoforms that can be produced [50]. Neurons may be particularly vulnerable to disruption of this process because maintenance of neuronal homeostasis depends on cell-type-specific transcript isoforms [51]. Alternative splicing is also involved in neuronal differentiation, migration, and synaptic maturation, and disturbed splicing has been linked with neurodegenerative disorders [52]. In addition, a recent protein-interaction study comparing ischemic stroke with inflammatory bowel disease identified spliceosome-related and ubiquitin-proteasome-related pathways among several shared network signals [53]. This type of network evidence is indirect, but it supports continued investigation into whether altered splicing contributes to neuronal injury following ischemia and whether WDR3 participates in that process.

Ribosome biogenesis begins with transcription of rRNA genes in the nucleolus and continues through processing, folding, and assembly of rRNA precursors to form functional ribosomal subunits; in this way, it helps determine the basal capacity for cellular protein synthesis [54]. A recent study identified a set of 12 ribosome-biogenesis-related genes with a key role in evaluating prognostic risk in

ischaemic stroke [55]. During severe acute stress such as cerebral ischemia, however, a compensatory increase in ribosome-biogenesis activity could place additional demands on ATP and nucleotide pools that are already limited by energy failure. Such an overload may intensify metabolic stress and can engage a p53-dependent apoptotic program through nucleolar-stress signaling [56]. In our GSEA results, both key genes, NOL9 and WDR3, were strongly associated with the “ribosomal biogenesis” pathway and were also upregulated in the IS transcriptomic samples. These enrichment findings establish an association with ribosome-related molecular programs, but whether nucleolar stress and p53 activation are actually involved in the observed pattern remains to be demonstrated experimentally.

Macrophages are highly plastic cells of the innate immune system whose functional state can change in response to the surrounding microenvironment. M0 macrophages represent an unpolarized state and, under steady-state conditions, are characterized by relatively low metabolic activity while contributing to maintenance of immune homeostasis [57]. Exposure to cytokines and other local mediators can drive macrophage polarization toward M1 or M2 phenotypes with different biological functions [58]. In experimental ischemic brain injury in mice, M1 macrophages can aggravate tissue damage and neuronal death through release of pro-inflammatory cytokines, proteases, reactive oxygen species, and reactive nitrogen species. By contrast, M2 macrophages can contribute to debris removal, angiogenesis, tissue remodeling, and repair. These observations have led to the proposal that modulation of transitions from the M0 state toward M1 or M2 phenotypes may represent a

608 neuroprotective strategy [59]. In the present study, however, the inverse correlations
609 involving NOL9/WDR3 and M0 macrophages should not be interpreted as evidence
610 that either gene drives macrophage polarization. The associations may instead reflect
611 differences in blood-cell composition or shared inflammatory programs accompanying
612 IS.
613 Drug prediction followed by molecular docking prioritized digoxigenin from the 61
614 candidate compounds. The calculated binding energies were -9.82 kcal/mol for NOL9
615 and -10.49 kcal/mol for WDR3, indicating favorable predicted affinity under the
616 docking criteria used in this study. Digoxigenin is a cardiac-glycoside compound and
617 represents the characteristic sterol aglycone of this drug class; its chemical framework
618 is highly homologous to the core aglycone present in digitalis glycoside and digoxin
619 [60]. CGS have been used for decades in the management of heart failure and
620 arrhythmia, whereas their potential neuroprotective effects in IS have received
621 increasing attention only more recently. In a mouse model of bilateral common carotid
622 artery occlusion/reperfusion injury, pretreatment with digoxigenin was reported to
623 reduce cerebral infarct volume and improve cognitive and motor dysfunction. The
624 reported protective effect was mainly linked to an increase in intracellular calcium
625 induced by digoxin, followed by activation of calcium-sensitive signal-transduction
626 pathways, suggesting that calcium signaling may contribute to the neuroprotective
627 actions of cardiac glycosides [61]. More recent work in a rat model of cerebral ischemia
628 showed that the semi-synthetic cardiotonic steroid gamma-benzylidene digoxin-15
629 (BD-15) improved motor capacity and exploratory behavior by enhancing

Na/K-ATPase（NKA） and Ca^{2+}-ATPase activity during oxidative stress and by regulating acetylcholinesterase function [62]. These observations provide context for the computational prioritization of digoxigenin, but the present docking analysis does not establish target engagement, brain exposure, or therapeutic efficacy in IS.

The present analysis also generated a putative ceRNA network connecting NOL9 and WDR3 with predicted upstream miRNAs and lncRNAs. Such an lncRNA-miRNA-mRNA framework can be used to formulate hypotheses about post-transcriptional regulation in IS, but the predicted interactions require experimental confirmation. Among the identified relationships, hsa-miR-499a-5p was connected with the three lncRNAs XIST, MALAT1, and NEAT1. Published findings concerning miR-499a-5p in IS have not been uniform. One report described lower plasma miR-499a-5p levels, an inverse association with neurological-deficit severity, and protective effects through PDCD4 targeting in experimental systems [63].In contrast, another study reported higher serum miR-499a and proposed that it suppresses astrocyte autophagy and viability through a Beclin 1-related mechanism [64]. Variation in sample source, disease phase, and experimental model may account for at least part of these divergent findings. XIST has likewise been reported to change over time after IS, with lower expression described during early stages and higher expression at later stages [65]. Experimental studies have connected XIST with cerebral microvascular injury as well as endothelial proliferation, migration, and angiogenesis through miR-92a/ITGA5- and KLF4-related pathways [66]. These observations support a vascular role for XIST but do not show that it regulates NOL9 or WDR3. MALAT1 has

652 also been implicated in experimental ischemia-reperfusion injury; in oxygen-glucose
653 deprivation/reoxygenation models, signaling involving MALAT1 and
654 miR-375/PDE4D has been associated with cellular injury, apoptosis, and release of
655 inflammatory factors [67]. NEAT1 has been reported as a hub molecule in IS-related
656 ceRNA networks in multiple independent studies. By sponging miR-377, NEAT1 can
657 increase SIRT1, VEGFA, and BCL-XL expression and thereby enhance survival and
658 angiogenic capacity in brain microvascular endothelial cells exposed to OGD [68].
659 NEAT1 has also been described as sponging miR-339-5p to increase HOXA1
660 expression and suppress neuronal apoptosis in neonatal mouse hypoxia-ischemia brain
661 injury [69]. Taken together, prior studies support involvement of these ncRNAs in
662 ischemic stress, but the specific ceRNA axes predicted here remain to be tested through
663 reporter assays, RIP, and targeted perturbation experiments [70].
664 Through an integrated bioinformatics workflow, this study identified NOL9 and
665 WDR3 as two entosis-associated key genes and examined their diagnostic performance
666 and expression consistency in an independent public validation cohort. Drug-prediction
667 and docking analyses further suggested that digoxigenin has predicted binding affinity
668 for both proteins. A miRNA-mRNA-lncRNA regulatory network was also assembled,
669 in which hsa-miR-499a-5p together with XIST, MALAT1, and NEAT1 may participate
670 in post-transcriptional regulation related to NOL9 and WDR3. Several limitations
671 remain important when interpreting these findings. GSE58294 contains repeated
672 time-point measurements from the same individuals with stroke, whereas GSE122709
673 contains both acute and subacute samples; therefore, treatment of within-subject

dependence and selection of time points require clarification. In addition, the direction of RT-qPCR expression differed from that observed in the transcriptomic datasets, the clinical validation sample was small, and no morphological or functional experiment directly demonstrated entosis. The estimated immune-cell proportions, ceRNA relationships, drug-target interactions, and therapeutic activity were also derived computationally. Confirmation in larger cohorts matched by sampling time, together with mechanistic experiments, will therefore be necessary.

5. Conclusions

The integrated analyses nominate NOL9 and WDR3 as exploratory entosis-related biomarker candidates for ischemic stroke. Their discordant transcriptomic and RT-qPCR expression directions require validation in larger, time-matched cohorts, and mechanistic studies are needed to establish their biological functions and therapeutic relevance.

List of abbreviations

Abbreviation	Definition
IS	Ischemic stroke
ERG	Entosis-related gene
DEG	Differentially expressed gene
WGCNA	Weighted gene co-expression network analysis
ssGSEA	Single-sample gene set enrichment analysis

GS	Gene significance
MM	Module membership
TOM	Topological overlap matrix
PPI	Protein–protein interaction
MCC	Maximum clique centrality
SVM-RFE	Support vector machine recursive feature elimination
RF	Random forest
ROC	Receiver operating characteristic
AUC	Area under the curve
DCA	Decision curve analysis
GSEA	Gene set enrichment analysis
NES	Normalized enrichment score
FDR	False discovery rate
miRNA	MicroRNA
lncRNA	Long non-coding RNA
ceRNA	Competing endogenous RNA
TF	Transcription factor
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
GEO	Gene Expression Omnibus
GO	Gene Ontology

KEGG	Kyoto Encyclopedia of Genes and Genomes
CNS	central nervous system
ROCK	Rho-associated kinase
AMPK	AMP-activated protein kinase
JNK	c-Jun N-terminal kinase

Declarations

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Xi'an Central Hospital Affiliated to Xi'an Jiaotong University (approval no. LW-2026-002). Written informed consent was obtained from all participants.

Consent for publication

Not applicable.

Data availability

The public transcriptomic datasets analysed during the current study are available in the Gene Expression Omnibus (GEO) under accession numbers GSE58294 and GSE122709. Supporting analysis results are included as Additional files 1-18. The RT-qPCR data generated in this study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Natural Science Basic Research Program of Shaanxi

Province (Program No. 2024JC-YBQN-0933) ,the Research Project of Xi'an Central Hospital (Grant No. 2024YB06) and Xi'an Science and Technology Plan Project (Program No. 25YXYJYB00087).

Authors' contributions

Le Chang: Conceptualization, Data curation, Validation, Visualization, Writing–original draft, Writing–review & editing.

Weili Cao: Data curation, Validation, Visualization.

Xurong Zhu: Validation, Visualization.

Naibing Gu: Supervision, Writing–review & editing.

Weiping Zhang: Project administration, Supervision.

Zhiqing Liu: Supervision, Writing–review & editing.

Zhengli Di: Conceptualization, Project administration, Supervision, Writing–review & editing.

This work currently described has not been published, is not being considered for publication elsewhere, and its publication was approved by all authors.

Acknowledgements

Not applicable

Reference

1. Herpich F, Rincon F: **Management of Acute Ischemic Stroke**. *Crit Care Med* 2020, **48**(11):1654–1663.

2. **Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy**

(HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* 2024, **403**(10440):2133–2161.

3. Pacinella G, Bona MM, Todaro F, Ciaccio AM, Daidone M, Tuttolomondo A: **Tracing Inflammation in Ischemic Stroke: Biomarkers and Clinical Insight**. *Int J Mol Sci* 2025, **26**(19).

4. Shademan B, Avci CB, Karamad V, Soureh GJ, Olia JBH, Esmaily F, Nourazarian A, Nikanfar M: **The Role of Mitochondrial Biogenesis in Ischemic Stroke**. *J Integr Neurosci* 2023, **22**(4):88.

5. Li X, Ding H, Qian S, Hou Z, Yang G, Ma Y, Gao Y, Zhang Y, Li T: **cGAMP Enhances Microglial/Macrophage Phagocytosis in Ischemic Stroke Via Activation of the TREM2-DAP10-PI3K Pathway**. *Inflammation* 2026, **49**(1):28.

6. Overholtzer M, Mailleux AA, Mouneimne G, Normand G, Schnitt SJ, King RW, Cibas ES, Brugge JS: **A nonapoptotic cell death process, entosis, that occurs by cell-in-cell invasion**. *Cell* 2007, **131**(5):966–979.

7. Kim S, Lee D, Kim SE, Overholtzer M: **Entosis: the core mechanism and crosstalk with other cell death programs**. *Exp Mol Med* 2024, **56**(4):870–876.

8. Gaggero G, Bertolotto M, Verzola D, Ferro J, Fiaschi P: **Cerebral small vessel vasculitis triggered by neurotoxoplasma-related ETosis: histological**

and immunofluorescence observation of a case. *J Rheum Dis* 2024, **31**(2):130–132.

9. Sterling NA, Park JY, Park R, Cho S-H, Kim S: **An entosis-like process induces mitotic disruption in Pals1 microcephaly pathogenesis**. *Nat Commun* 2023, **14**(1):82.

10. Sterling NA, Cho S-H, Kim S: **Entosis implicates a new role for P53 in microcephaly pathogenesis, beyond apoptosis**. *Bioessays* 2024, **46**(8):e2300245.

11. Sterling NA, Cho SH, Kim S: **Entosis implicates a new role for P53 in microcephaly pathogenesis, beyond apoptosis**. *Bioessays* 2024, **46**(8):e2300245.

12. Hamann JC, Surcel A, Chen R, Teragawa C, Albeck JG, Robinson DN, Overholtzer M: **Entosis Is Induced by Glucose Starvation**. *Cell Rep* 2017, **20**(1):201–210.

13. **Gene Expression Omnibus (GEO)** [https://www.ncbi.nlm.nih.gov/geo/].Accessed 12 July 2025.

14. Wu C, Fang S, Wu L, Mi Z, Yin Y, Liao Y, Zhao Y, Wang T, Na J: **Identification of the entosis-related prognostic signature and tumour microenvironment in hepatocellular carcinoma on the basis of bioinformatics analysis and experimental validation**. *Clin Exp Med* 2025, **25**(1):55.

15. Song J, Xu R, Zhang H, Xue X, Ruze R, Chen Y, Yin X, Wang C, Zhao Y: **Cell-in-Cell-Mediated Entosis Reveals a Progressive Mechanism in Pancreatic Cancer**. *Gastroenterology* 2023, **165**(6):1505–1521.e1520.

16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies**. *Nucleic Acids Res* 2015, **43**(7):e47.

17. Ito K, Murphy D: **Application of ggplot2 to Pharmacometric Graphics**. *CPT Pharmacometrics Syst Pharmacol* 2013, **2**(10):e79.

18. Gu Z: **Complex heatmap visualization**. *iMeta* 2022, **1**:e43.

19. Hänzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data**. *BMC Bioinformatics* 2013, **14**:7.

20. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC Bioinformatics* 2008, **9**:559.

21. Joshi S, Annida RM, Zuilhof H, van Beek TA, Nielen MWF: **Analysis of Mycotoxins in Beer Using a Portable Nanostructured Imaging Surface Plasmon Resonance Biosensor**. *J Agric Food Chem* 2016, **64**(43):8263–8271.

22. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *OMICS* 2012, **16**(5):284–287.

23. **STRING database** [https://string-db.org].Accessed 12 July 2025.

24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res* 2003, **13**(11):2498–2504.

25. Yan F, Chen X, Quan X, Wang L, Wei X, Zhu J: **Association between the stress hyperglycemia ratio and 28-day all-cause mortality in critically ill patients with sepsis: a retrospective cohort study and predictive model establishment based on machine learning**. *Cardiovasc Diabetol* 2024, **23**(1):163.

26. Shi H, Yuan X, Liu G, Fan W: **Identifying and Validating GSTM5 as an Immunogenic Gene in Diabetic Foot Ulcer Using Bioinformatics and Machine Learning**. *J Inflamm Res* 2023, **16**:6241–6256.

27. Alderden J, Pepper GA, Wilson A, Whitney JD, Richardson S, Butcher R, Jo Y, Cummins MR: **Predicting Pressure Injury in Critical Care Patients: A Machine-Learning Model**. *Am J Crit Care* 2018, **27**(6):461–468.

28. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M: **pROC: an open-source package for R and S+ to analyze and compare ROC curves**. *BMC Bioinformatics* 2011, **12**:77.

29. Akazawa M, Matsumoto T, Kumagai N: **Prevalence of problematic drinking among outpatients attending general hospitals in Tokyo**. *Nihon Arukoru Yakubutsu Igakkai Zasshi* 2013, **48**(5):300–313.

30. Rollo T, Blanchini F, Giordano G, Specogna R, Esseni D: **Stabilization of negative capacitance in ferroelectric capacitors with and without a metal interlayer**. *Nanoscale* 2020, **12**(10):6121–6129.

31. Xue M, Yu W, Zhang K, Chen Y, Zhang L, Zhang H, Lu S, Tao M, Yan H, Wei L *et al*: **GCDCA promotes hepatocellular carcinoma progression through S1PR2/PI3K/AKT-mediated polarization of M2-type macrophages**. *Front Immunol* 2026, **17**:1640450.

32. **GeneMANIA** [https://genemania.org/].Accessed 12 July 2025.

33. Guan Y, Yin X, Wang L, Diao Z, Huang H, Wang X: **Biomarkers of Arginine Methylation in Diabetic Nephropathy: Novel Insights from Bioinformatics Analysis**. *Diabetes Metab Syndr Obes* 2024, **17**:3399–3418.

34. Saidmamatov O, Jammatov J, Sousa C, Barros R, Vasconcelos O, Rodrigues P: **Translation and Adaptation of the Adult Developmental Coordination Disorder/Dyspraxia Checklist (ADC) into Asian Uzbekistan**. *Sports (Basel)* 2023, **11**(7):135.

35. **Molecular Signatures Database (MSigDB)** [https://www.gsea-msigdb.org/gsea/msigdb].Accessed 12 July 2025.

36. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles**. *Nat Methods* 2015, **12**(5):453–457.

37. **lncATLAS** [https://lncatlas.crg.eu/].Accessed 12 July 2025.

38. **hTFtarget** [https://guolab.wchscu.cn/hTFtarget/#!].Accessed 12 July 2025.

39. **Drug Signatures Database (DsigDB)** [http://dsigdb.tanlab.org/].Accessed 12 July 2025.

40. **UniProt** [https://www.uniprot.org/].Accessed 12 July 2025.

41. **AlphaFold Protein Structure Database** [https://alphafold.ebi.ac.uk/].Accessed 12 July 2025.

42. **PubChem** [https://pubchem.ncbi.nlm.nih.gov/].Accessed 12 July 2025.

43. Zhu S, Chan-Yeung M, Becker AB, Dimich-Ward H, Ferguson AC, Manfreda J, Watson WT, Paré PD, Sandford AJ: **Polymorphisms of the IL-4, TNF-alpha, and Fcepsilon RIbeta genes and the risk of allergic disorders in at-risk infants**. *Am J Respir Crit Care Med* 2000, **161**(5):1655–1659.

44. Heindl K, Martinez J: **Nol9 is a novel polynucleotide 5'-kinase involved in ribosomal RNA processing**. *Embo j* 2010, **29**(24):4161–4171.

45. Chen X, Song X, Zheng X, Qian T, Zhang B, Wu L, Lian Q, Chen J, Luo Q, Xu W *et al*: **Nucleolar NOL9 regulated by DNA methylation promotes hepatocellular carcinoma growth through activation of Wnt/β-catenin signaling pathway**. *Cell Death Dis* 2025, **16**(1):100.

46. Li Y, Yang Y, Sears RC, Dai M-S, Sun X-X: **USP36 SUMOylates Las1L and Promotes Its Function in Pre-Ribosomal RNA ITS2 Processing**. *Cancer Res Commun* 2024, **4**(10):2835–2845.

47. Li M, Li N, Fan Y, Zhang Z, Zhou L, Yu Y, Ni M, Tan M, Huang W, Zhu T: **WDR3 undergoes phase separation to mediate the therapeutic mechanism of Nilotinib against osteosarcoma**. *J Exp Clin Cancer Res* 2025, **44**(1):201.

855 48. McMahon M, Ayllón V, Panov KI, O'Connor R: **Ribosomal 18 S RNA**
856 **processing by the IGF-I-responsive WDR3 protein is integrated with p53**
857 **function in cancer cell proliferation**. *J Biol Chem* 2010, **285**(24):18309–
858 18318.
859 49. Agarwal A, Park S, Ha S, Kwon JS, Khan MR, Kang BG, Dawson TM,
860 Dawson VL, Andrabi SA, Kang SU: **Quantitative mass spectrometric**
861 **analysis of the mouse cerebral cortex after ischemic stroke**. *PLoS One*
862 2020, **15**(4):e0231978.
863 50. Will CL, Lührmann R: **Spliceosome structure and function**. *Cold Spring*
864 *Harb Perspect Biol* 2011, **3**(7).
865 51. Nik S, Bowman TV: **Splicing and neurodegeneration: Insights and**
866 **mechanisms**. *Wiley Interdiscip Rev RNA* 2019, **10**(4):e1532.
867 52. Li D, McIntosh CS, Mastaglia FL, Wilton SD, Aung-Htut MT:
868 **Neurodegenerative diseases: a hotbed for splicing defects and the**
869 **potential therapies**. *Transl Neurodegener* 2021, **10**(1):16.
870 53. Hu W, Li P, Zeng N, Tan S: **Exploring the hub mechanisms of ischemic**
871 **stroke based on protein-protein interaction networks related to ischemic**
872 **stroke and inflammatory bowel disease**. *Sci Rep* 2023, **13**(1):1741.
873 54. Dörner K, Ruggeri C, Zemp I, Kutay U: **Ribosome biogenesis factors-from**
874 **names to functions**. *EMBO J* 2023, **42**(7):e112699.

55. Wang X, Zhang X-Y, Liao N-Q, He Z-H, Chen Q-F: **Identification of ribosome biogenesis genes and subgroups in ischaemic stroke**. *Front Immunol* 2024, **15**:1449158.

56. McMahon M, Ayllón V, Panov KI, O'Connor R: **Ribosomal 18 S RNA processing by the IGF-I-responsive WDR3 protein is integrated with p53 function in cancer cell proliferation**. *J Biol Chem* 2010, **285**(24):18309–18318.

57. Wicks EE, Ran KR, Kim JE, Xu R, Lee RP, Jackson CM: **The Translational Potential of Microglia and Monocyte-Derived Macrophages in Ischemic Stroke**. *Front Immunol* 2022, **13**:897022.

58. Zhang K, Guo J, Yan W, Xu L: **Macrophage polarization in inflammatory bowel disease**. *Cell Commun Signal* 2023, **21**(1):367.

59. Han D, Liu H, Gao Y: **The role of peripheral monocytes and macrophages in ischemic stroke**. *Neurol Sci* 2020, **41**(12):3589–3607.

60. Green MR, Sambrook J: **Digoxigenin**. *Cold Spring Harb Protoc* 2022, **2022**(3):pdb.top100784.

61. Kaur S, Rehni AK, Singh N, Jaggi AS: **Studies on cerebral protection of digoxin against ischemia/reperfusion injury in mice**. *Yakugaku Zasshi* 2009, **129**(4):435–443.

62. de Souza Gonçalves B, Dos Santos CP, Machado MV, Toledo MM, Dos Santos HB, Thomé RG, Maia GAS, Tilelli CQ, de Carvalho LED, Santos HL *et al*: **Gamma-Benzylidene Digoxin Derivative Attenuates Neurotoxicity**

Response in a Murine Stroke Model. *Transl Stroke Res* 2025, **16**(6):2014–2025.

63. Shan W, Ge H, Chen B, Huang L, Zhu S, Zhou Y: **Upregulation of miR-499a-5p Decreases Cerebral Ischemia/Reperfusion Injury by Targeting PDCD4**. *Cell Mol Neurobiol* 2022, **42**(7):2157–2170.

64. Xian G, Song D, Wei Y, Huang F, Luo H: **Upregulation of miR-499a Targets Beclin1 to Modulate Astrocyte Autophagy and Promote Ischemic Stroke**. *Ann Clin Lab Sci* 2024, **54**(1):47–55.

65. Minhas G, Mathur D, Ragavendrasamy B, Sharma NK, Paanu V, Anand A: **Hypoxia in CNS Pathologies: Emerging Role of miRNA-Based Neurotherapeutics and Yoga Based Alternative Therapies**. *Front Neurosci* 2017, **11**:386.

66. Wang C, Dong J, Sun J, Huang S, Wu F, Zhang X, Pang D, Fu Y, Li L: **Silencing of lncRNA XIST impairs angiogenesis and exacerbates cerebral vascular injury after ischemic stroke**. *Mol Ther Nucleic Acids* 2021, **26**:148–160.

67. Zhang G, Wang Q, Su D, Xie Y: **Long Non-coding RNAMALAT1 Knockdown Alleviates Cerebral Ischemia/Reperfusion Injury of Rats Through Regulating the miR-375/PDE4D Axis**. *Front Neurol* 2020, **11**:578765.

68. Zhou Z-W, Zheng L-J, Ren X, Li A-P, Zhou W-S: **LncRNA NEAT1 facilitates survival and angiogenesis in oxygen-glucose deprivation**

(OGD)-induced brain microvascular endothelial cells (BMECs) via targeting miR-377 and upregulating SIRT1, VEGFA, and BCL-XL. *Brain Res* 2019, **1707**:90–98.

69. Zhao J, He L, Yin L: **lncRNA NEAT1 Binds to MiR-339-5p to Increase HOXA1 and Alleviate Ischemic Brain Damage in Neonatal Mice**. *Mol Ther Nucleic Acids* 2020, **20**:117–127.

70. Li S, Xu Z, Zhang S, Sun H, Qin X, Zhu L, Jiang T, Zhou J, Yan F, Deng Q: **Non-coding RNAs in acute ischemic stroke: from brain to periphery**. *Neural Regen Res* 2025, **20**(1):116–129.

Additional file legends

Additional file 1 List of 95 entosis-related genes (ERGs) obtained from two previous publications [5,6].

File name: Additional file 1

File format: .csv

Title of data: Entosis-related genes (ERGs).

Description of data: This file contains the list of 95 entosis-related genes (ERGs) obtained from two previous publications [5,6] and used for subsequent analyses.

Additional file 2 List of ERG-associated module genes identified by WGCNA.

File name: Additional file 2

File format: .csv

Title of data: ERG-associated module genes identified by WGCNA.

Description of data: This file contains the list of 1,712 ERG-associated module genes identified by WGCNA from the pink and purple modules.

Additional file 3 Complete list of differentially expressed genes (DEGs) between IS and control samples in the training set GSE58294.

File name: Additional file 3

File format: .csv

Title of data: Differentially expressed genes in GSE58294.

Description of data: This file contains the complete list of differentially expressed genes (DEGs) between IS and control samples in the training set GSE58294, including $\log_2$ fold change, adjusted P value, and regulation direction.

Additional file 4 List of candidate genes obtained by intersecting DEGs with ERG-associated module genes.

File name: Additional file 4

File format: .csv

Title of data: Candidate genes identified by intersection analysis.

Description of data: This file contains the list of 429 candidate genes obtained by intersecting the DEGs with the ERG-associated module genes.

Additional file 5 Complete GO enrichment results for biological process terms of the candidate genes.

File name: Additional file 5

File format: .csv

Title of data: GO biological process enrichment results.

Description of data: This file contains the complete Gene Ontology (GO) biological process enrichment results for the candidate genes at adjusted $P < 0.05$, including term ID, description, gene count, and P value.

Additional file 6 Complete GO enrichment results for cellular component terms of the candidate genes.

File name: Additional file 6

File format: .csv

Title of data: GO cellular component enrichment results.

Description of data: This file contains the complete Gene Ontology (GO) cellular component enrichment results for the candidate genes at adjusted $P < 0.05$, including term ID, description, gene count, and P value.

Additional file 7 Complete GO enrichment results for molecular function terms of the candidate genes.

File name: Additional file 7

File format: .csv

Title of data: GO molecular function enrichment results.

Description of data: This file contains the complete Gene Ontology (GO) molecular function enrichment results for the candidate genes at adjusted $P < 0.05$, including term ID, description, gene count, and P value.

Additional file 8 Complete KEGG pathway enrichment results of the candidate genes.

File name: Additional file 8

File format: .csv

Title of data: KEGG pathway enrichment results.

Description of data: This file contains the complete Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment results for the candidate genes at adjusted $P < 0.05$, including pathway ID, description, gene count, and P value.

Additional file 9 List of candidate genes selected from the PPI network by the MCC algorithm.

File name: Additional file 9

File format: .csv

Title of data: Candidate genes selected by the MCC algorithm.

Description of data: This file contains the list of 31 candidate genes selected from the protein–protein interaction (PPI) network using the maximum clique centrality (MCC) algorithm.

Additional file 10 List of feature genes identified by the Boruta algorithm.

File name: Additional file 10

File format: .csv

Title of data: Feature genes identified by Boruta.

Description of data: This file contains the list of 22 feature genes identified using the Boruta feature-selection algorithm.

Additional file 11 List of feature genes selected by the SVM-RFE algorithm.

File name: Additional file 11

File format: .csv

Title of data: Feature genes selected by SVM-RFE.

Description of data: This file contains the list of 24 feature genes selected using the support vector machine recursive feature elimination (SVM-RFE) algorithm.

Additional file 12 List of the top 10 feature genes ranked by random forest importance.

File name: Additional file 12

File format: .csv

Title of data: Top feature genes ranked by random forest importance.

Description of data: This file contains the top 10 feature genes ranked according to their importance scores in the random forest analysis.

Additional file 13 Complete GSEA results for NOL9.

File name: Additional file 13

File format: .csv

Title of data: GSEA results for NOL9.

Description of data: This file contains the complete gene set enrichment analysis (GSEA) results for NOL9, including pathway name, normalized enrichment score (NES), false discovery rate (FDR), and P value.

Additional file 14 Complete GSEA results for WDR3.

File name: Additional file 14

File format: .csv

Title of data: GSEA results for WDR3.

Description of data: This file contains the complete gene set enrichment analysis (GSEA) results for WDR3, including pathway name, normalized enrichment score (NES), false discovery rate (FDR), and P value.

Additional file 15 Spearman correlation matrix among the differential immune cell subtypes.

File name: Additional file 15

File format: .csv

Title of data: Correlations among differential immune cell subtypes.

Description of data: This file contains the Spearman correlation matrix among the differential immune cell subtypes, including correlation coefficients and corresponding P values.

Additional file 16 Spearman correlations between the key genes and differential immune cell subtypes.

File name: Additional file 16

File format: .csv

Title of data: Correlations between NOL9/WDR3 and differential immune cell subtypes.

Description of data: This file contains the Spearman correlations between the key genes NOL9 and WDR3 and the differential immune cell subtypes.

Additional file 17 List of transcription factors targeting NOL9 and WDR3 predicted by hTFtarget.

File name: Additional file 17

File format: .xlsx

Title of data: Transcription factors targeting NOL9 and WDR3.

Description of data: This file contains the transcription factors predicted by hTFtarget to target NOL9 and WDR3, including 398 transcription factors for NOL9 and 331 transcription factors for WDR3.

Additional file 18 List of candidate drugs predicted by DsigDB for NOL9 and WDR3.

File name: Additional file 18

File format: .csv

Title of data: Candidate drugs predicted for NOL9 and WDR3.

Description of data: This file contains the list of 61 candidate drugs predicted by DsigDB for NOL9 and WDR3, including degree centrality (DC), betweenness centrality (BC), and closeness centrality (CC).

Supplementary Files

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

- Table1.docx
- GraphicalAbstract.pptx