

OPEN

Exploration of effective biomarkers for venous thrombosis embolism in Behçet's disease based on comprehensive bioinformatics analysis

Chunjiang Liu^{1,5}, Yuan Wang^{2,5}, Zhifeng Wu^{1,3,5}, Xiaoqi Tang¹, Guohua Wang¹ & Jiajia Wang⁴✉

Behçet's disease (BD) is a multifaceted autoimmune disorder affecting multiple organ systems. Vascular complications, such as venous thromboembolism (VTE), are highly prevalent, affecting around 50% of individuals diagnosed with BD. This study aimed to identify potential biomarkers for VTE in BD patients. Three microarray datasets (GSE209567, GSE48000, GSE19151) were retrieved for analysis. Differentially expressed genes (DEGs) associated with VTE in BD were identified using the Limma package and weighted gene co-expression network analysis (WGCNA). Subsequently, potential diagnostic genes were explored through protein–protein interaction (PPI) network analysis and machine learning algorithms. A receiver operating characteristic (ROC) curve and a nomogram were constructed to evaluate the diagnostic performance for VTE in BD patients. Furthermore, immune cell infiltration analyses and single-sample gene set enrichment analysis (ssGSEA) were performed to investigate potential underlying mechanisms. Finally, the efficacy of listed drugs was assessed based on the identified signature genes. The limma package and WGCNA identified 117 DEGs related to VTE in BD. A PPI network analysis then selected 23 candidate hub genes. Four DEGs (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*) were identified by intersecting gene sets from three machine learning algorithms. ROC analysis and nomogram construction demonstrated high diagnostic accuracy for these four genes (AUC: 0.816, 95% CI: 0.723–0.909). Immune cell infiltration analysis revealed a positive correlation between dysregulated immune cells and the four hub genes. ssGSEA provided insights into potential mechanisms underlying VTE development and progression in BD patients. Additionally, therapeutic agent screening identified potential drugs targeting the four hub genes. This study employed a systematic approach to identify four potential hub genes (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*) and construct a nomogram for VTE diagnosis in BD. Immune cell infiltration analysis revealed dysregulation, suggesting potential macrophage involvement in VTE development. ssGSEA provided insights into potential mechanisms underlying BD-induced VTE, and potential therapeutic agents were identified.

Keywords Behçet's disease, Venous thrombosis embolism, Diagnostic biomarker, Bioinformatics analysis, Machine learning

Behçet's disease (BD) is a chronic autoimmune disorder characterized by widespread vasculitis that can affect any bodily system¹. While the prognosis for most BD patients is favorable, cardiovascular involvement significantly worsens the outlook. It is now understood that the mortality rate is directly correlated with cardiovascular system involvement, with venous thromboembolism (VTE) occurring frequently in approximately 50% of BD

¹Division of Vascular Surgery, Department of General Surgery, Shaoxing People's Hospital, Shaoxing 312000, China. ²Department of Intervention Vascular, Hefei Hospital of Anhui Medical University, Hefei 230000, China. ³Medical College, Shaoxing University, Shaoxing 312000, China. ⁴Department of Rheumatology, Shaoxing People's Hospital, 568# Zhongxing North Road, Shaoxing 312000, China. ⁵These authors contributed equally: Chunjiang Liu, Yuan Wang and Zhifeng Wu. ✉email: wangjiajia8911@163.com

cases². Inflammation-driven endothelial cell dysfunction is thought to be the main cause of thromboembolism in patients with BD³. In the absence of traditional cardiovascular risk factors, compromised immune-inflammatory responses, along with neutrophil activation, endothelial dysfunction, and coagulation abnormalities, can directly contribute to thrombus formation in BD patients^{4,5}. Current evidence suggests that the likelihood of venous thrombosis is particularly heightened during the active phase of the disease, likely due to the complex coagulation mechanisms activated during this period⁶.

VTE, encompassing deep vein thrombosis (DVT) and pulmonary embolism (PE), represents the third most common cardiovascular disease after acute coronary heart disease and stroke⁷. This global health concern is responsible for a significant number of deaths annually and imposes substantial economic and psychological burdens on individuals and society⁸. Notably, over 50% of VTE patients lack clear symptoms and signs, leading to a high rate of misdiagnosis in clinical practice⁹. The resemblance of VTE symptoms to common conditions like dyspnea or post-exertional discomfort further complicates diagnosis. Pulmonary embolism, a potentially fatal complication in BD patients, can manifest with non-specific symptoms like chest pain and breathing difficulties. In some cases, extensive PE may even be the sole presenting feature of sudden death¹⁰. Therefore, prompt detection and treatment of VTE in BD are crucial for preventing VTE-related complications and mortality. Biomarkers offer significant potential for early identification and intervention in conditions like BD and asymptomatic VTE, which often present with subtle clinical manifestations. Timely diagnosis and treatment of VTE in BD patients is essential for mitigating adverse outcomes and improving patient prognosis.

Despite extensive global research on the pathogenesis of venous thrombosis in BD, the specific mechanisms underlying its occurrence remain elusive⁴. Microarray-based gene expression profiling has emerged as a prominent tool in biomedical research, enabling the identification of potential biomarkers. Studies have identified candidate biomarkers for both BD and VTE, including *IL10*, *FCRL3*, *MASPI*, *NF2*, *FAM3B*, and *MGMT* for BD¹¹, and *ISG15*, *RPS15A*, *MRPL13*, *ICT1*, *MRPL15*, and *RPLP0* for VTE¹². Additionally, research suggests that circulating microRNAs (miR-20a-3p, miR-134, miR-522-3p, and miR-874) may regulate genes encoding exosome components (*TLN1*, *LAMTOR3*, *STX7*, and *IMPA1*), potentially influencing exosome involvement in VTE pathophysiology¹³. While the precise etiology of BD remains unknown, it is likely a complex interplay of factors, including infections, genetics, epigenetics, and immunological components¹⁴. Emerging evidence points towards a growing link between venous thrombosis and the immune system, with inflammatory factors playing a confirmed role. The intricate interplay between enzymes and cellular processes in the endothelium, platelets, and white blood cells can create an inflammatory state, ultimately leading to acute vein thrombosis^{15,16}. Biomarkers related to immune function may, therefore, hold promise as valuable predictors of VTE susceptibility in BD patients and could aid in treatment decisions. However, the literature on the specific genetic mechanisms underlying BD-induced VTE is scarce, highlighting the need for further investigation.

Bioinformatics techniques have revolutionized disease research by facilitating the exploration of underlying mechanisms and the identification of potential biomarkers. Extensive utilization of various analytical approaches, including Limma analysis, weighted gene co-expression network analysis (WGCNA), protein–protein interaction (PPI) analysis, machine learning algorithms, and nomogram evaluation, can lead to a deeper understanding of diseases and offer novel insights and approaches for disease prevention, early diagnosis, and treatment^{17,18}. The advent of high-throughput technologies has significantly expanded the application of genomic data in biomarker discovery. However, a common challenge arises when the number of features in biomedical data exceeds the number of samples, potentially hindering the reproducibility of traditional feature selection methods^{19,20}. Machine learning algorithms, such as Support Vector Machine-Recursive Feature Elimination (SVM-RFE), Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest, have emerged as crucial tools for identifying promising biomarkers. These methodologies expedite the exploration of underlying biological processes and aid in the identification of reliable biomarkers. Recent studies suggest that machine learning techniques, known for their predictive flexibility, outperform traditional regression methods in terms of accuracy²¹. This study addresses a critical gap in our understanding of the genetic mechanisms underlying BD-induced VTE. We employed a combination of bioinformatics and machine learning methodologies to uncover potential disease mechanisms and identify potential biomarkers.

Methods

Microarray data

All gene expression datasets employed in this analysis are publicly available through the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Dataset GSE209567 encompasses gene expression data from 29 BD patients and 15 healthy controls. Additionally, datasets GSE48000 (comprising 107 VTE patients and 25 healthy controls) and GSE19151 (including 70 VTE patients and 63 healthy controls) encompass data from VTE patients and their corresponding controls. The genes within these datasets were converted into probes and linked to their corresponding annotation symbols. Table 1 provides a comprehensive overview of the datasets used in this study.

Data processing and differential expression analysis

Preprocessing and standardization are essential steps for ensuring the quality and reliability of microarray data. These procedures eliminate biases and inconsistencies inherent to raw biological data, promoting data integrity. In this study, we employed various techniques, including background correction, normalization, and probe data visualization. The “afy” package in R was used to implement these methods. Additionally, the “surrogate variable analysis” package in R was utilized to mitigate the confounding effects of batch variations. Limma²², a popular R package for differential gene expression analysis, was employed to identify differentially expressed genes (DEGs) between the BD and control groups in the GSE209567 dataset. Likewise, the Limma package was

ID	GSE series	Disease	Samples	Source types	Platform	Group
1	GSE209567	BD	29 BD patients and 15 normal controls	Peripheral blood	GPL570	Discovery cohort
2	GSE48000	VTE	107 VTE patients and 25 normal controls	Peripheral blood	GPL10558	Discovery cohort
3	GSE19151	VTE	70 VTE patients and 63 normal controls	Peripheral blood	GPL571	Validation cohort

Table 1. Detailed information of the GEO datasets in the study.

used to identify DEGs between the VTE and control groups within the GSE48000 dataset. Specifically, utilizing the obtained expression profile dataset, the lmFit function was applied to perform multiple linear regression. The eBayes function was then employed to compute modeled t-statistics, moderated F-statistics, and log-odds of differential expression, with empirical Bayes moderation of the standard errors towards a common value, to ascertain the significance of gene expression differences. Additionally, heat maps and volcano plots for DEGs were generated using the ggplot2 R package. A false discovery rate (FDR) below 0.05 and a fold change (FC) exceeding 1.2 were set as the criteria for selecting DEGs.

Weighted gene co-expression network analysis (WGCNA)

WGCNA, a widely employed technique, was utilized to construct co-expression networks and identify modules of genes exhibiting strong correlations with BD²³. The analysis commenced with the calculation of the median absolute deviation (MAD) for each gene. Subsequently, the 50% of genes with the lowest MAD values were excluded. To construct a scale-free co-expression network, the expression matrix of DEGs was filtered using the goodSamplesGenes function within WGCNA. The pick-Soft-Threshold function was then employed to determine the adjacency between genes. This function utilizes co-expression similarity to compute the soft-thresholding power (β). The resulting adjacency relationships were transformed into a topological overlap matrix (TOM). Subsequently, the gene dissimilarity and its corresponding ratio were calculated. Hierarchical clustering and dynamic tree-cutting algorithms were then implemented to identify gene modules. Finally, a truncation value was chosen for the module tree diagram. By merging specific modules based on the dissimilarity of the estimated module feature genes, the co-expression network of these feature genes was visualized.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis was employed to investigate the functional roles of DEGs in BD and VTE. This analysis aimed to identify enriched biological processes (BP), molecular functions (MF), and cellular components (CC) associated with the DEGs. Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed to explore enriched pathways potentially involved in BD and VTE²⁴. To achieve comprehensive functional enrichment analysis, we utilized the clusterProfiler package in R, integrating both GO and KEGG analyses²⁵. For visualization purposes, the top 10 enriched GO terms in each category were selected based on a significance threshold of FDR < 0.05 and p-value < 0.05.

Protein–protein interaction network construction

The Search Tool for the Retrieval of Interacting Genes (STRING) database (version 12.0) (<http://string-db.org>) was employed to construct a PPI network of DEGs. The STRING database provides interaction scores, with interactions exceeding a combined score of 0.4 considered significant for this analysis. Cytoscape software (version 3.9.1) was then utilized to visualize the interactive network data. Furthermore, to identify hub genes within the network, the CytoHubba plug-in for Cytoscape was employed. Three algorithms (Betweenness Centrality, Closeness Centrality, and Degree Centrality) were implemented within CytoHubba. Each algorithm assigned a score to each DEG based on its network properties, enabling subsequent ranking of the genes according to their respective scores. The top 30 DEGs identified by each algorithm were designated as node DEGs. Finally, a Venn diagram was constructed to visually represent the overlap of node DEGs identified by all three algorithms, facilitating further analysis.

Machine learning algorithms

To identify potential biomarkers for VTE in BD patients, three machine learning algorithms were employed: LASSO regression, SVM-RFE, and Random Forest. LASSO regression analysis was implemented to mitigate potential overfitting by shrinking less relevant coefficients towards zero, effectively performing variable selection²⁶. SVM-RFE is particularly well-suited for datasets with a limited number of samples as it iteratively eliminates redundant features while retaining only those most relevant for classification²⁷. Random Forest analysis offers advantages in handling high-dimensional datasets, constructing robust predictive models, and assessing the variable importance²⁸. The intersection of genes identified by all three algorithms was considered a set of putative biomarkers for further investigation.

Evaluation of receiver operating characteristic curve (ROC) and nomogram

The expression levels of candidate genes were compared between the VTE and control groups, as well as between the BD and control groups, using Student's t-test. This analysis aimed to identify statistically significant differences in gene expression between the groups. The diagnostic performance of candidate gene expression levels was subsequently evaluated using ROC curves. The area under the curve (AUC) was calculated using the “pROC” package in R to assess the discriminatory power of each gene²⁹. To develop a multi-gene prognostic model for

predicting VTE incidence in BD patients, a nomogram was constructed using the R package “RMS”³⁰. This model integrates risk scores derived from the expression levels of the identified genes. The accuracy of the nomogram model was further validated using ROC curves.

qRT-PCR of the hub genes and evaluation of the predictive model

To validate the identified hub genes *in vivo*, a retrospective study was conducted at Shaoxing People’s Hospital. Peripheral blood samples were collected from patients diagnosed with BD (n=8) and those diagnosed with both BD and VTE (n=6) between June 1, 2022, and June 1, 2024. All diagnoses of BD adhered to the established International Criteria³¹. VTE was confirmed through duplex sonography of the lower limbs or computed tomographic pulmonary angiography (CTPA). Detailed clinical characteristics of these patients are presented in Supplementary Table S1. This study was performed in accordance with the ethical standards of the Declaration of Helsinki. The hospital’s Ethics Committee approved the specimen collection protocol (Approval number: IEC-K-AF-016-1.2), and informed consent was obtained from all patients. Following collection, total RNA was extracted from peripheral blood using the MolPure® Blood RNA Kit according to the manufacturer’s instructions. RNA concentration was determined using a Nanodrop instrument (ThermoFisher, USA). cDNA synthesis was subsequently performed using the Hifair® II 1st Strand cDNA Synthesis Kit (11121ES60, Yeasen, Shanghai, China). The specific primers utilized in this analysis are listed in Supplementary Table S2. The expression levels of the hub genes were then compared between the two patient groups. Furthermore, a predictive nomogram was developed to discriminate between BD patients with VTE and those without VTE.

Immune infiltration analysis

The cellular composition of infiltrating immune cells within the tissue samples was deconvoluted from the normalized gene expression matrix using the CIBERSORT algorithm³². This algorithm estimates the relative abundances of 22 distinct immune cell types within the samples. The R package “CIBERSORT” was employed to quantify the proportions of these immune cell types between the VTE and control groups. To visualize potential correlations between different immune cell populations in the context of VTE development, a heatmap was generated using the R package “corrplot”. Furthermore, single-sample gene set enrichment analysis (ssGSEA) was performed to investigate the associations between immune cell infiltration and characteristic gene expression profiles. The “ggcorrplot” package was utilized to generate graphical representations of these associations.

ssGSEA and therapeutic agents screening

To elucidate potential connections between the identified biomarkers and hallmark gene sets, we employed ssGSEA. The hallmark gene sets were obtained from the Molecular Signatures Database (MSigDB) (<https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp>), a comprehensive resource encompassing 50 well-defined biological states and processes. We utilized the GSVA software to perform ssGSEA and evaluate the correlations between the potential biomarkers and these hallmark gene sets.

Following ssGSEA, we conducted an enrichment analysis using Enrichr (<https://maayanlab.cloud/Enrichr/>). This analysis, employing a significance threshold of $p < 0.05$, aimed to identify potential therapeutic agents that specifically interact with the hub genes.

Statistical analysis

Statistical analyses were performed using R software (version 4.2.1) with relevant packages, GraphPad Prism (version 9.4.0), and SPSS (version 26.0) software packages. For continuous variables, comparisons between two groups were conducted using Student’s t-test, with a p-value threshold of less than 0.05 considered statistically significant.

Ethical approval statement

This study was performed in accordance with the ethical standards of the Declaration of Helsinki. The specimen collection protocol was approved by the Ethics Committee of Shaoxing People’s Hospital (Approval number: IEC-K-AF-016-1.2), and informed consent was obtained from all patients.

Results

Identification of DEGs for BD and VTE using Limma

Figure 1 illustrates the study flowchart. A total of 1015 genes exhibiting differential expression were identified in the comparison between the BD and control groups. Among them, 477 genes showed up-regulation, while 538 genes displayed down-regulation. Likewise, the comparison between the VTE and control groups revealed a total of 5734 genes exhibiting differential expression. Among these, 3747 genes were up-regulated, and 1987 genes were down-regulated. To visually represent these findings, a heatmap was generated to display the top 20 significantly differentially expressed genes in both up-regulated and down-regulated regulation (BD: Fig. 2A, VTE: Fig. 2C). Additionally, a volcano plot was utilized to illustrate the differentially expressed genes (BD: Fig. 2B, VTE: Fig. 2D). A total of 160 genes (Fig. 2E) were found to be overlapped between BD-related and VTE-related DEGs.

Identification of significant module genes in BD via WGCNA

Significant module genes associated with BD were identified using WGCNA. The gray module, commonly referred to as the “junk module”, failed to cluster genes considered irrelevant or uninformative. The darkturquoise ($r=0.43$, $p=3.4 \times 10^{-3}$), midnightblue ($r=0.43$, $p=3.4 \times 10^{-3}$), and pink ($r=-0.47$, $p=1.3 \times 10^{-3}$) modules showed

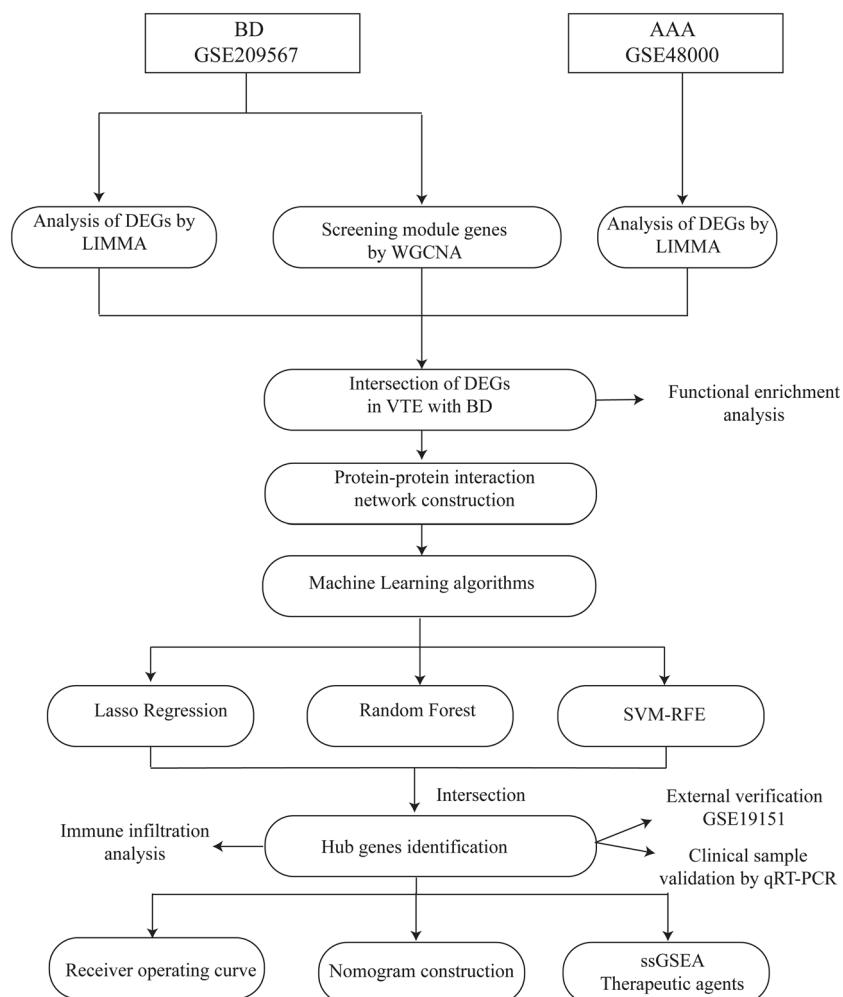


Figure 1. Flowchart of study.

the highest correlation with BD (Fig. 3A). Figure 3B–D depict the relationship between module membership and gene significance in the darkturquoise/midnightblue/pink modules of BD. A total of 3552 genes were identified to be associated with BD. The intersection of these genes with the 160 DEGs yielded 117 DEGs associated with VTE in BD (Fig. 3E). Supplementary Fig. S1 displayed the selection of the soft threshold and the gene cluster tree.

Functional enrichment analysis

The analysis of functional pathways in the 117 DEGs (Fig. 4A, Supplementary Table S3) identified significant enrichment in various pathways, such as “Cellular senescence”, “Th1 and Th2 cell differentiation”, and “Platelet activation”. The BP analysis of the 117 DEGs demonstrated significant enrichment in GO terms, including “Negative regulation of cellular metabolic process”, “Leukocyte migration”, and “Regulation of biological quality”. Additionally, the CC analysis revealed enrichment in terms such as “Endomembrane system”, “Nucleoplasm”, and “Plasma membrane part”. MF analysis of the DEGs demonstrated significant associations with terms like “Enzyme binding”, “Kinase binding”, and “Protein kinase binding” (Fig. 4B–D, Supplementary Table S3).

PPI network construction

A preliminary PPI network was constructed using the 117 DEGs to identify hub DEGs associated with BD and VTE. Among the DEGs, 57 genes were retained based on their interactions with other genes (Fig. 5A), while 60 DEGs lacking interaction were excluded. Furthermore, the CytoHubba plug-in was utilized, employing three distinct algorithms (degree, betweenness, and closeness) to identify the intersecting DEGs. The top 30 node genes determined by the betweenness, closeness, and degree algorithms are presented in Fig. 5B–D. The overlap of 30 genes identified by the three algorithms was visually represented in Fig. 5E, with 23 of these genes being identified as potential hub genes. Supplementary Table S4 provided comprehensive information on these 23 identified genes.

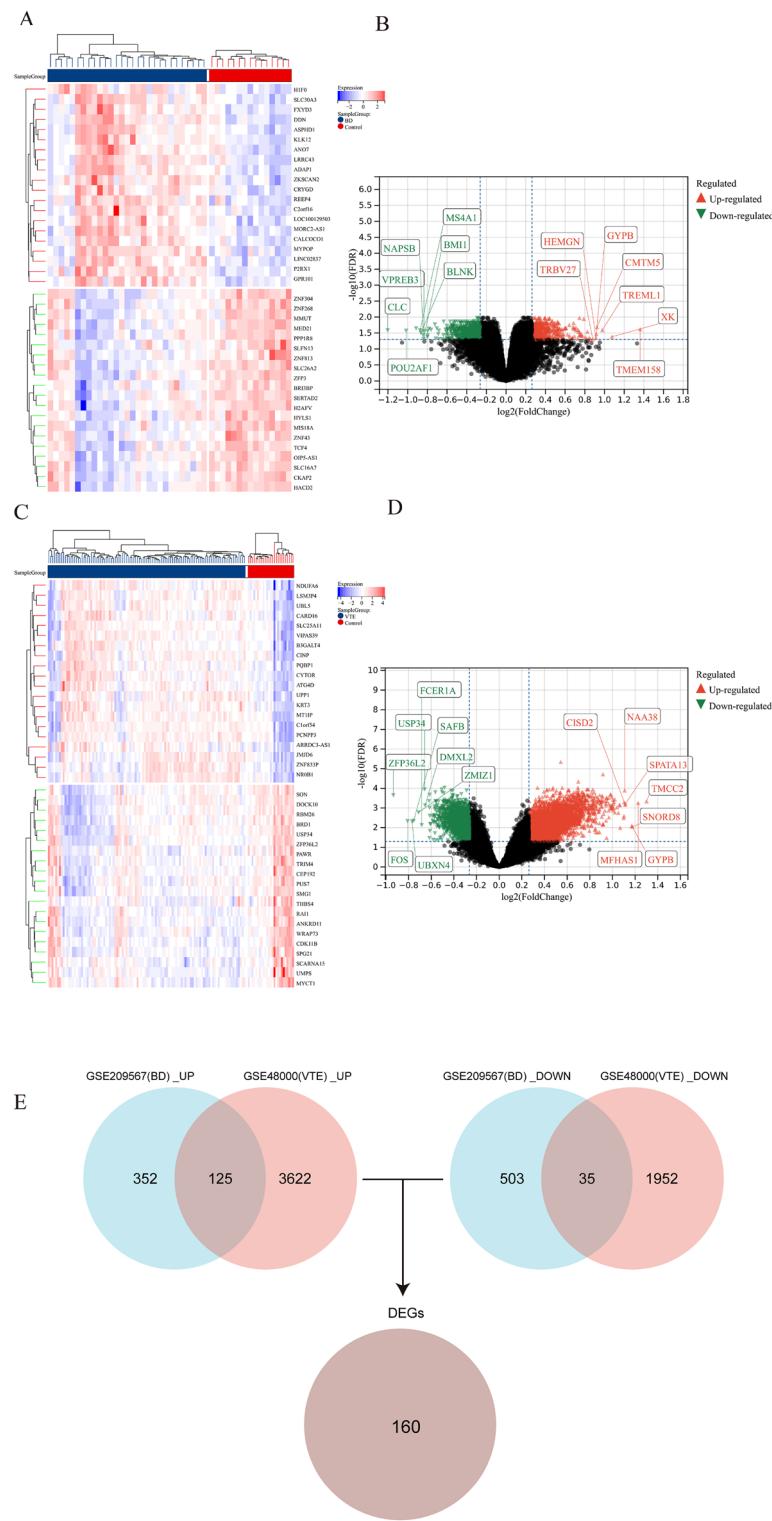
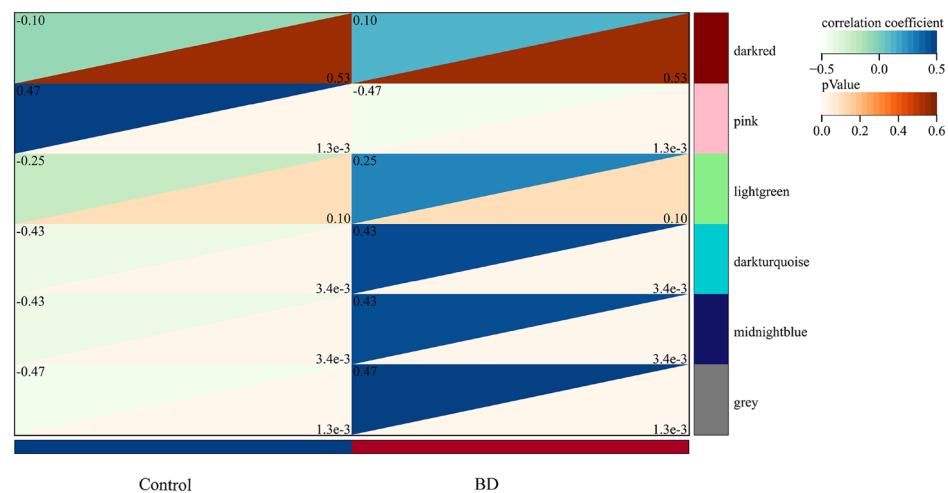
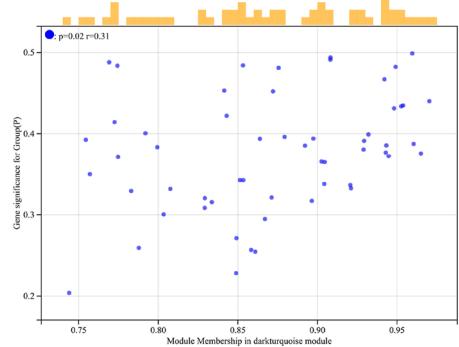


Figure 2. Heatmap and volcano plot of DEGs between BD and the control group, as well as between VTE and the control group. (A, B) Heatmap displaying the expression of the 20 most upregulated and downregulated DEGs. The red blocks represent upregulation, indicating increased gene expression in BD, while the blue blocks indicate downregulation, reflecting decreased gene expression. Volcano plot presenting all DEGs, with significantly upregulated genes labeled in red and downregulated genes labeled in green. (C, D) Heatmap and volcano plot of DEGs in VTE. The description remains consistent with the previous depiction. (E) The intersection of BD-related and VTE-related DEGs led to 160 DEGs.

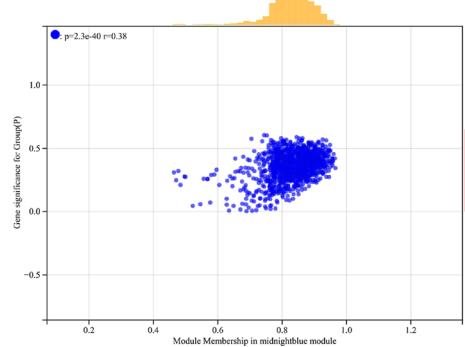
A



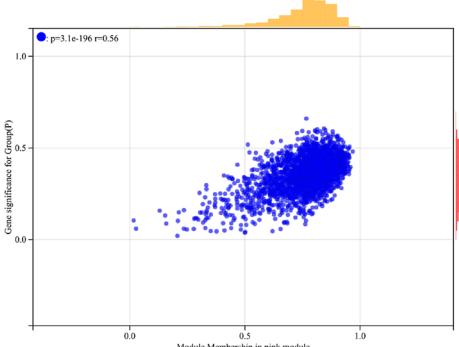
B



C



D



E

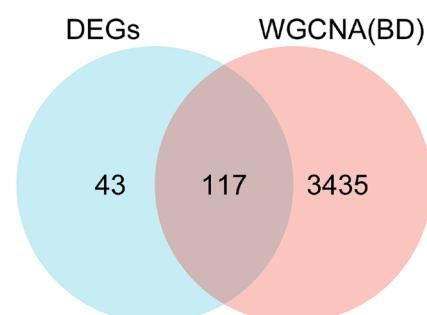
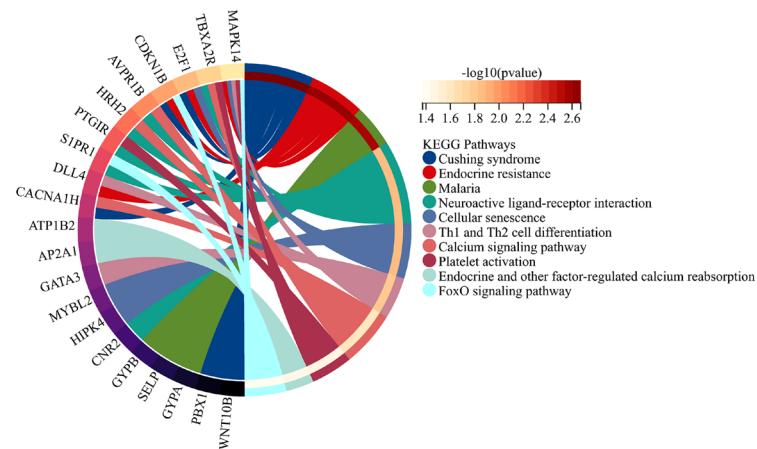
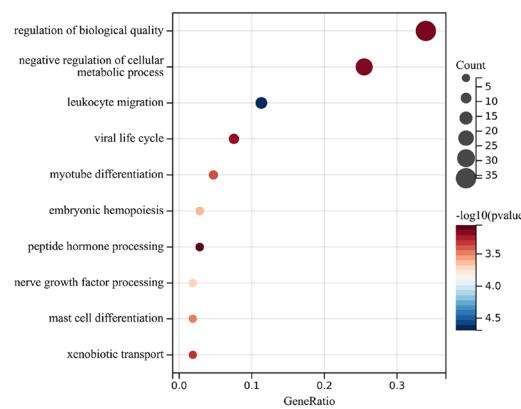


Figure 3. The recognition of module genes using WGCNA in BD. (A) Heatmap displays modules that are associated with BD. The correlation values between each module and BD are presented in the top left corner of the heatmap, while the p-values in the lower right corner indicate the statistical significance of these correlations. Among the modules within BD, the darkturquoise, midnightblue, and pink modules exhibit the strongest correlation. (B–D) The relationship between module membership and gene significance in the darkturquoise/midnightblue/pink modules of BD. (E) Venn plot illustrating the overlap of 160 DEGs and 3552 genes associated with BD modules, resulting in 117 DEGs associated with VTE in BD.

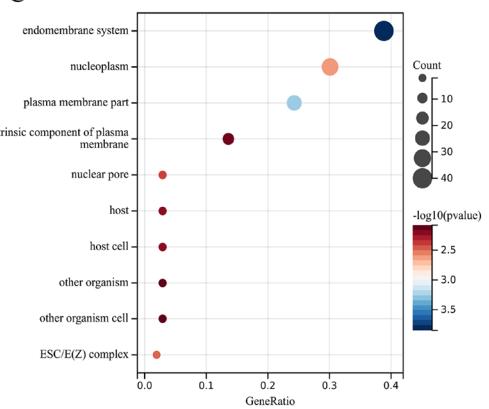
A



B



C



D

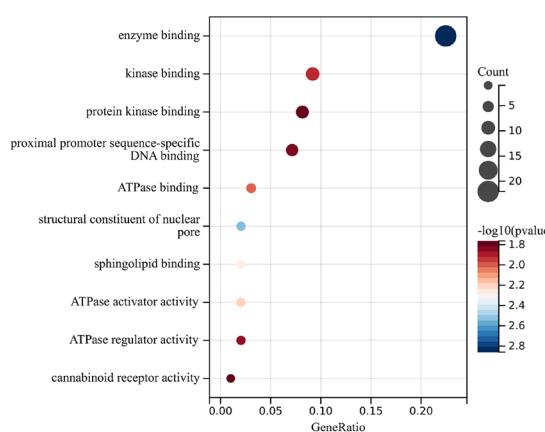
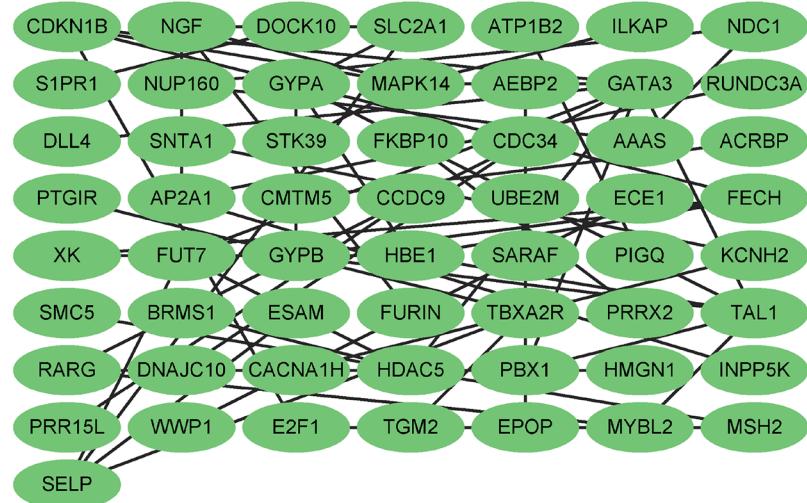


Figure 4. Functional enrichment analysis of DEGs associated with VTE in BD. (A) KEGG pathway analysis of DEGs associated with VTE in BD. (B–D) GO analysis (BP, CC, and MF) of DEGs associated with VTE in BD. The size of the circles corresponds to the number of genes, and the color indicates the significance level.

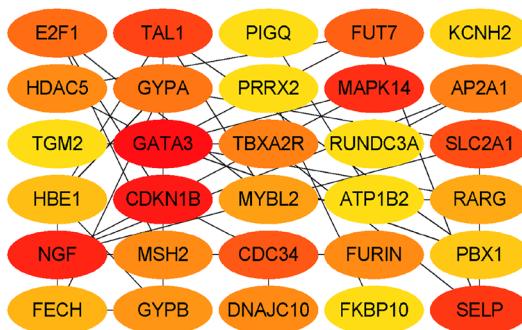
Machine learning

The LASSO regression algorithm was used to identify 6 candidate hub genes that serve as optimal biomarkers for diagnosing VTE in patients with BD. These genes correspond to the lowest point on the curve, as depicted in Fig. 6A. The random forest approach was utilized to evaluate the significance of DEGs, which successfully detected errors, as displayed in Fig. 6B. Figure 6C displays a compilation of the 15 most significant DEGs. Using SVM-RFE analysis, the top 18 DEGs with the lowest error and highest accuracy in diagnosing VTE in patients

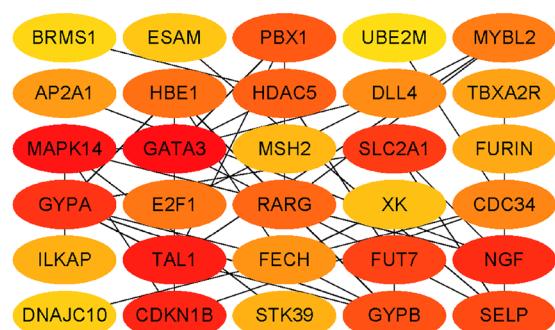
A



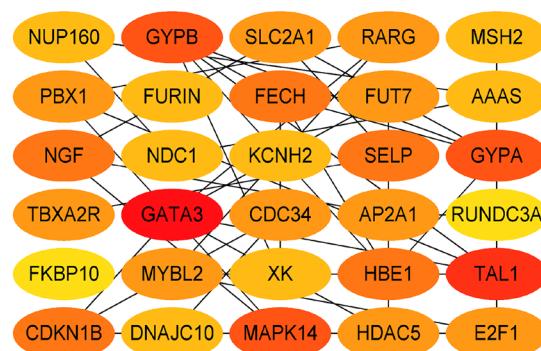
B



C



D



E

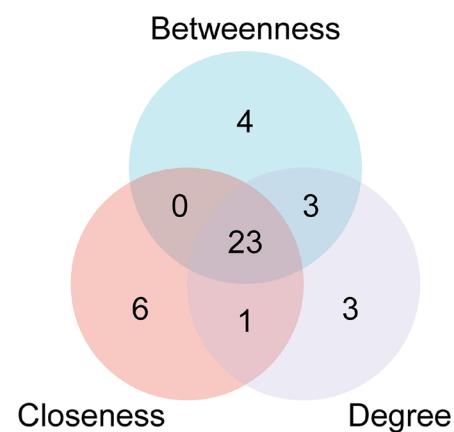


Figure 5. Construction of the PPI network and selection of hub genes. (A) Excluding 60 DEGs due to a lack of interaction, the preliminary PPI network was constructed using the remaining 57 DEGs. (B–D) Within the CytoHubba plug-in of Cytoscape, three distinct algorithms were utilized to select the top 30 DEGs. Panels (B–D) represent the application of betweenness, closeness, and degree algorithms, with darker colors indicating greater significance. (E) Through the intersection of genes identified by all three algorithms, a total of 23 DEGs were chosen.

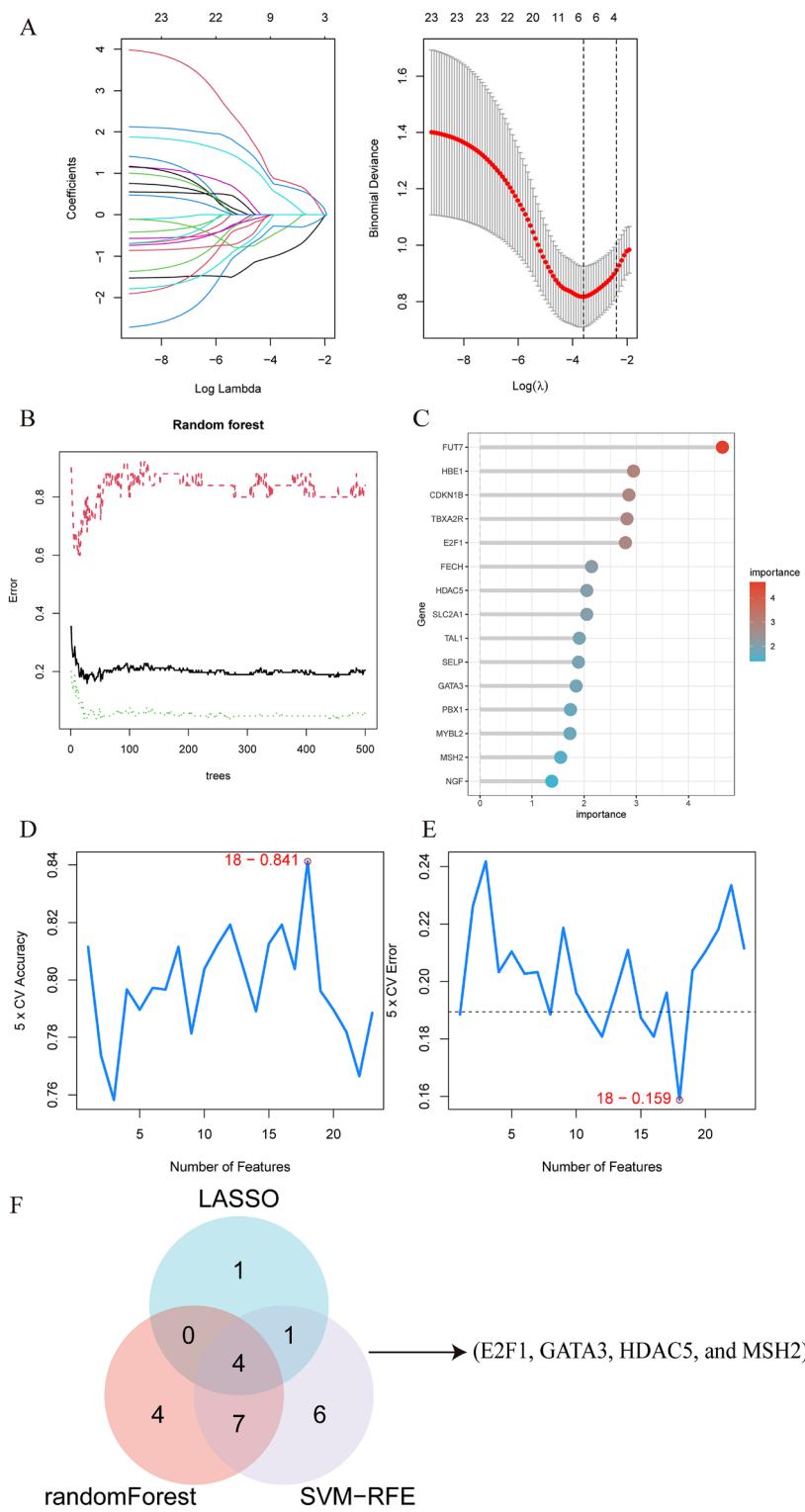


Figure 6. Identification of candidate diagnostic biomarkers through machine learning algorithms. **(A)** The Lasso regression algorithm identified six highly accurate biomarkers for diagnosing VTE with BD by precisely locating the curve's lowest point. **(B, C)** The random forest algorithm was employed to assess the error in VTE, leading to the ranking of the top 15 genes based on their importance scores. **(D, E)** The SVM-RFE algorithm selected 18 genes with the lowest error and highest accuracy. **(F)** Four hub genes (E2F1, GATA3, HDAC5, and MSH2) were identified as candidate biomarkers based on the intersection of the results from the three algorithms.

with BD were identified, as shown in Fig. 6D, E. Finally, four candidate hub genes (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*) were selected by intersecting the genes identified through the three algorithms, as depicted in Fig. 6F.

ROC, nomogram construction, and validation of the hub genes

In comparison to the control group, VTE demonstrated the up-regulation of genes *E2F1* and *HDAC5*, as well as the down-regulation of genes *GATA3* and *MSH2*, as indicated in Fig. 7A. Figure 7B presented the AUC and 95% CI for each gene: *E2F1* (AUC: 0.775, 95% CI: 0.678–0.871); *GATA3* (AUC: 0.764, 95% CI: 0.666–0.862); *HDAC5* (AUC: 0.739, 95% CI: 0.631–0.848); *MSH2* (AUC: 0.720, 95% CI: 0.607–0.832). The analysis of the ROC curves demonstrated that these four genes exhibited satisfactory diagnostic performance. Finally, the construction of the nomogram, as depicted in Fig. 7C, resulted in an AUC of 0.816 (95% CI: 0.723–0.909), indicating significant clinical diagnostic value, as demonstrated in Fig. 7D. Similarly, BD showed the up-regulation of genes *E2F1* and *HDAC5*, as well as the down-regulation of genes *GATA3* and *MSH2*, as shown in Supplementary Fig. S2A. Supplementary Fig. S2B presented the AUC and 95% CI for each gene. The analysis of the ROC curves also revealed satisfactory diagnostic performance for these four genes. Finally, the construction of the nomogram, as depicted in Supplementary Fig. S2C, resulted in an AUC of 0.823 (95% CI: 0.699–0.947), confirming significant clinical diagnostic value, as demonstrated in Supplementary Fig. S2D. To further validate its diagnostic potential, the validation dataset GSE19151 was utilized for ROC curve analysis. The analysis of the nomogram in the validation dataset showed an AUC of 0.902, confirming substantial clinical diagnostic value, as illustrated in Supplementary Fig. S2E. Additionally, as depicted in Fig. 7E, clinical validation of peripheral blood samples revealed an elevated expression of two DEGs (*E2F1* and *HDAC5*) in BD patients with coexisting VTE in contrast to the BD-only group. In contrast, the expression levels of two additional DEGs (*GATA3* and *MSH2*) were found to be lower under the same conditions. Moreover, to assist in the differentiation of BD patients with potential VTE, we established a predictive nomogram. This model, presented in Fig. S3A–C, achieved an AUC of 1, substantiating its significant predictive capability for VTE in individuals with BD.

Immune cell infiltration analysis

Three distinct immune cell types exhibited zero percent distribution across all samples and were therefore excluded from further analysis. Figure 8A presents a bar plot depicting the distribution of the remaining 19 immune cell types following the application of the CIBERSORT algorithm. Boxplot analysis revealed a significantly higher prevalence of macrophages and M0 macrophages in the VTE group compared to the control group. Conversely, VTE samples displayed a lower prevalence of neutrophils (Fig. 8B). Correlation analysis identified a positive correlation between M2 macrophages and activated dendritic cells ($r=0.39$). Regulatory T cells exhibited the strongest negative correlation with activated CD4 memory T cells ($r=-0.45$), as shown in Fig. 8C. These findings suggest that modulating macrophage populations may hold promise as a therapeutic strategy for VTE. Additionally, analysis of immune cell infiltration demonstrated a significant correlation with all three identified hub DEGs, as illustrated in Fig. 8D.

ssGSEA and therapeutic agent screening

Next, ssGSEA revealed significant associations between the four hub genes and diverse biological processes, including “hypoxia”, “heme metabolism”, “coagulation”, “cholesterol homeostasis”, and “angiogenesis”, as illustrated in Supplementary Fig. S4. These biological processes might be intricately connected to the onset and progression of VTE in individuals BD.

Enrichr database was utilized to screen therapeutic agents that target the four hub genes. The analysis of the predicted results indicated potential efficacy in targeting the four hub genes associated with VTE induced by BD. The identified therapeutic agents include chlorpromazine, dronabinol, trabectedin, vorinostat, and carmustine (Supplementary Fig. S5, Supplementary Table S5).

Discussion

Recent research has increasingly highlighted the link between VTE and the immune system, with confirmed associations to inflammatory factors. DVT arises from complex interactions between enzymes and cellular processes. The endothelium, platelets, and white blood cells collaborate to induce a pro-inflammatory state, ultimately culminating in thrombosis and acute DVT^{16,33}. VTE is associated with a significant mortality rate, particularly from PE. Untreated cases have a mortality rate of 30%. However, prompt diagnosis and treatment can reduce the mortality rate due to PE or treatment-related causes to less than 1%³⁴. VTE also carries the burden of long-term complications and substantial healthcare costs. Approximately 50% of DVT patients may experience post-thrombotic syndrome, a chronic and debilitating condition. Additionally, the expensive treatment of VTE further burdens the healthcare system³⁵. The presence of cardiovascular involvement worsens the prognosis of Budd-Chiari syndrome (BD). Therefore, considering these potential adverse consequences, early prevention, detection, and treatment of VTE are of paramount importance in patients with BD.

The precise mechanisms by which Budd-Chiari syndrome and VTE share transcriptional pathogenesis remain to be elucidated. To address this knowledge gap, we conducted a comprehensive analysis of microarray datasets, aiming to identify common biomarkers and biological pathways associated with the development of both diseases. Reliable biomarkers hold significant value in modern medicine. The application of bioinformatics and machine learning methodologies have significantly advanced the investigation of potential mechanisms and biomarker discovery. These methods offer a powerful approach for the accurate identification of disease-associated biomarkers, facilitating research into disease initiation, progression, and exploration of potential pathogenic mechanisms.

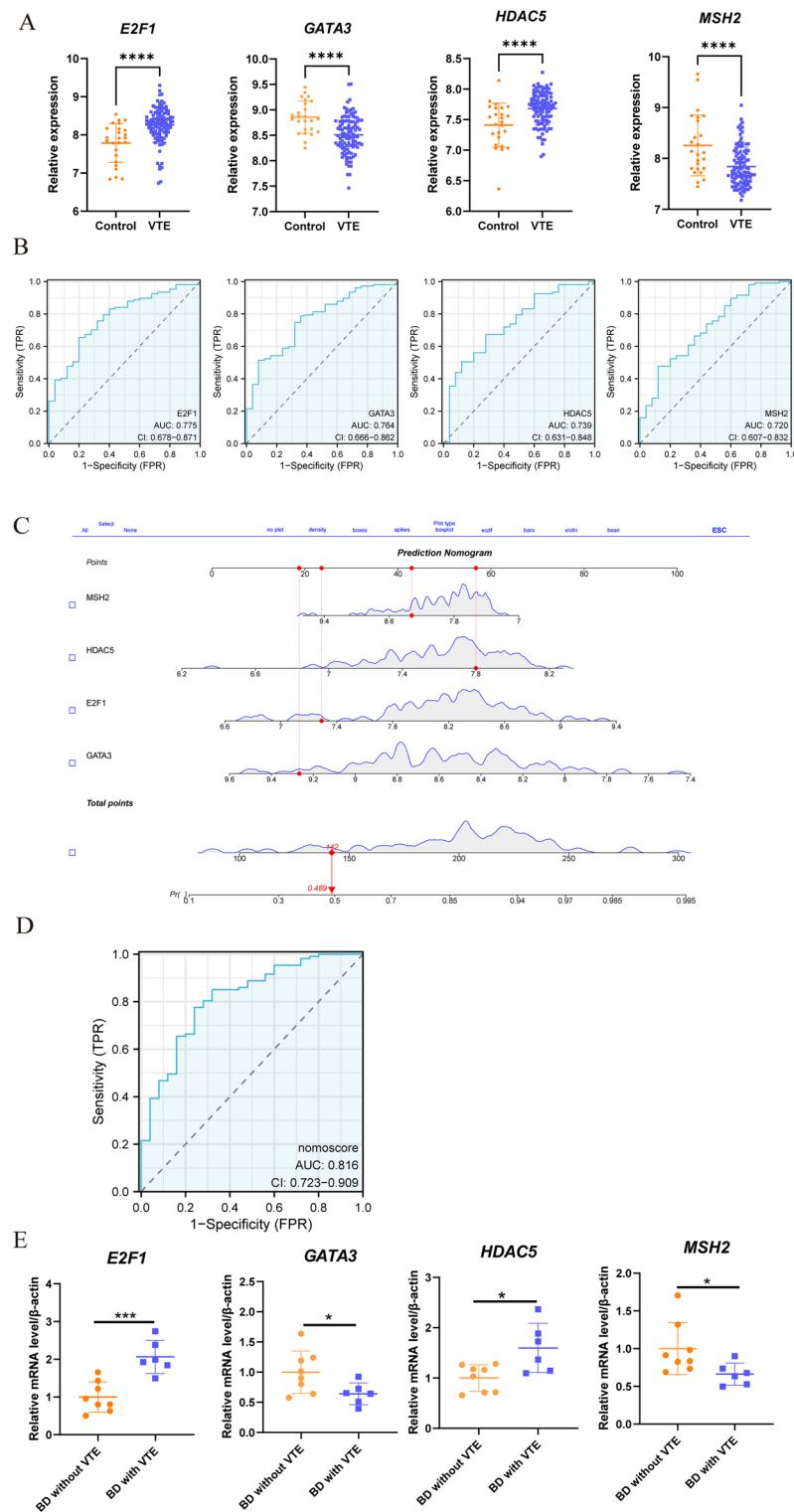


Figure 7. Evaluation of ROC and nomogram construction. (A) Differential expression is observed in four genes when comparing the VTE and control groups (***, $p < 0.0001$). (B) The ROC curve was utilized to evaluate the diagnostic efficacy of the four hub genes (E2F1, GATA3, HDAC5, and MSH2) in identifying VTE in patients with BD, demonstrating satisfactory diagnostic performance with AUC and 95% CI displayed in each panel. (C) The construction of a diagnostic nomogram employing the four genes aims to enhance the diagnosis of VTE in patients with BD. (D) The ROC curve for the nomogram is depicted in patients with BD presenting VTE. (E) Differential expression is observed in four genes via qRT-PCR when comparing the BD with VTE group and BD without VTE group (*, $p < 0.05$; ***, $p < 0.001$).

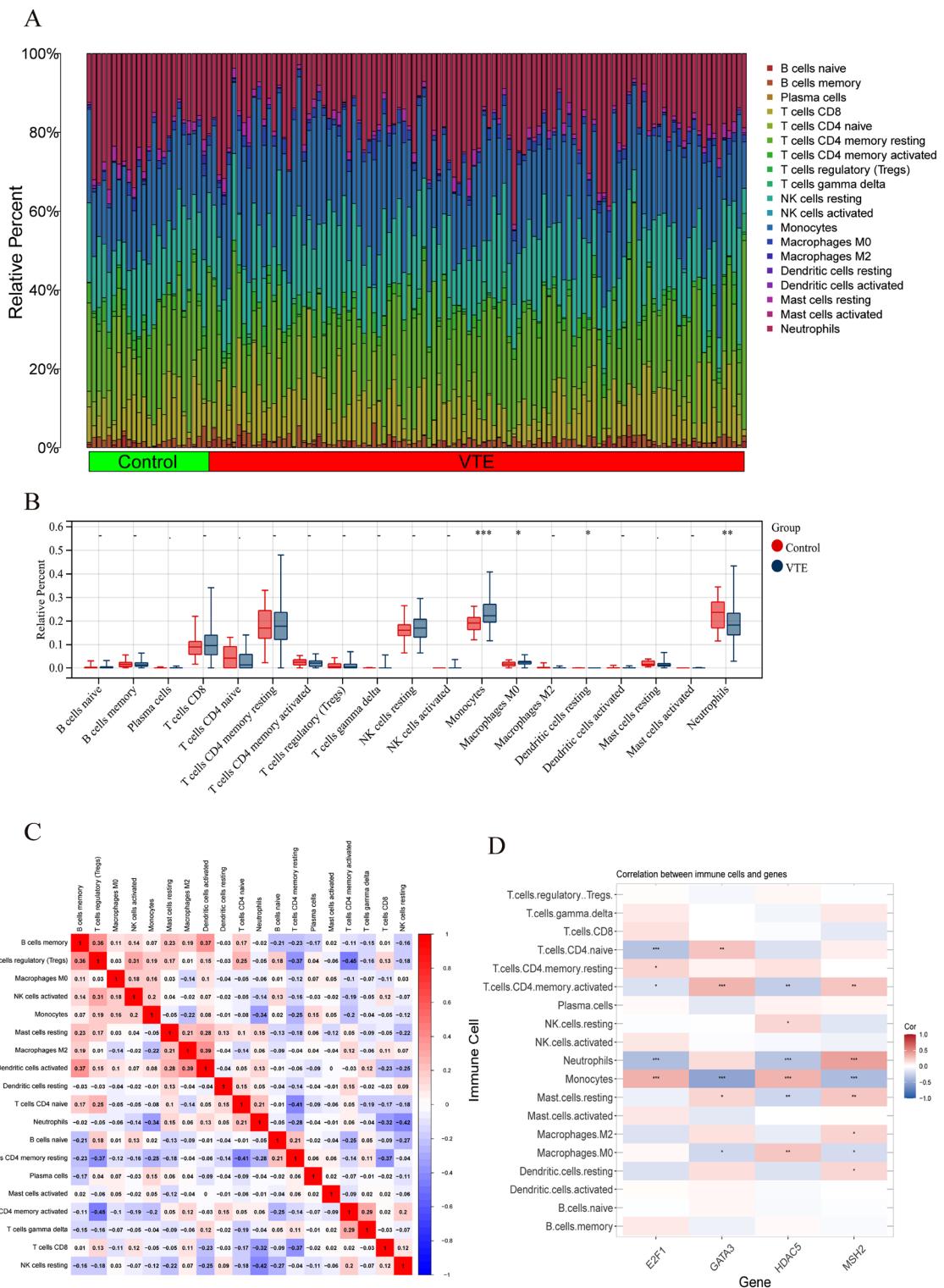


Figure 8. Immunological alterations between control and VTE groups, and correlations between hub DGEs and immunological features in VTE. (A) A bar plot representing the distribution of immune cells among different samples. (B) A boxplot comparing the expression of immune cells between the VTE and control groups, with statistical significance denoted as * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (C) A heatmap illustrating the correlations among various immune cells implicated in the pathogenesis of VTE. (D) Correlation analysis assessing the relationship between immune cell infiltrations and the four identified hub DGEs.

Assessment of hub gene expression levels in peripheral blood samples from patients diagnosed with BD and VTE offers valuable insights for predicting the risk of VTE development in this specific population. This approach, therefore, emerges as a potentially practical and efficacious clinical tool. Our nomogram facilitates the calculation of both cumulative and total scores for each gene. As a result, this nomogram holds significant promise for clinical application by aiding in the identification and early intervention of BD patients with elevated total scores. Early intervention has the potential to improve the prognosis of this patient population.

This study employed the Limma R package, a robust tool for gene expression data analysis, to identify DEGs associated with Budd-Chiari syndrome and VTE. Overlap analysis revealed 160 genes shared between BD- and VTE-associated DEGs. To further explore co-expressed genes potentially related to BD pathogenesis, a gene coexpression network was constructed using WGCNA. This analysis identified significant correlations between gene modules and phenotypes, leading to the discovery of 3552 genes associated with BD. Subsequent analysis of these genes alongside the 160 overlapping DEGs pinpointed 117 DEGs specifically associated with VTE in the context of BD. To elucidate the functional roles of these 117 DEGs, enrichment analyses were performed using the KEGG pathway and GO databases. The results indicated a primary association with cellular senescence, differentiation of Th1 and Th2 cells, platelet activation, negative regulation of cellular metabolism, and leukocyte migration. These findings suggest a strong link between immune system dysregulation, inflammatory responses, and the interplay between BD and VTE. Dysregulations in Th1/Th2 characterize BD's complex immune responses³⁶. Th1 cell-driven inflammation can compromise vascular endothelium, elevating thrombosis likelihood. Additionally, cellular senescence is intimately linked with both BD and VTE^{37,38}. The exacerbation of senescence-associated secretory phenotype (SASP) during cellular senescence contributes to platelet activation, a critical element in VTE development³⁹. Moreover, leukocyte migration significantly influences the pathology of both BD and VTE. In BD sufferers, aberrant leukocyte migration and activation may intensify vascular inflammation and thrombosis risk⁴⁰. These aspects support the view that there is a potential association between BD and VTE.

Furthermore, the 117 DEGs were mapped onto a PPI network to investigate potential interactions. This analysis identified 23 potential hub DEGs, which may play crucial roles in the underlying biological processes. Machine learning algorithms, including SVM-RFE, LASSO, and Random Forest, are increasingly employed for biomarker discovery due to their effectiveness in identifying promising candidates. These methodologies facilitate the exploration of complex biological processes and aid in the identification of reliable biomarkers. In this study, we utilized three machine learning methods alongside ROC curve analysis to evaluate the diagnostic accuracy of candidate gene expression. This approach led to the identification of four candidate hub genes with potential diagnostic value: *E2F1*, *GATA3*, *HDAC5*, and *MSH2*. To further validate the diagnostic potential of these genes, an additional dataset (GSE19151) and clinical validation using peripheral blood samples were employed. These analyses confirmed the significant clinical diagnostic value of the identified gene expression signature.

E2F1, the first identified member of the E2 promoter binding factor (E2F) family, is believed to play a critical role in cell cycle regulation, differentiation, apoptosis, and the DNA damage response⁴¹. Dysregulation of *E2F1* can impact downstream transcriptional targets, leading to DNA replication stress. This stress arises when persistent obstacles, both internal and external, hinder DNA replication and ultimately contribute to various diseases, including cancer⁴². A study suggested a potential link between *E2F1* and vascular diseases, as inhibiting its expression protected Human Umbilical Vein Endothelial Cells (HUVECs) from high-glucose-induced injury⁴¹. Overexpression of *E2F1* has been demonstrably associated with the development and progression of various malignant tumors⁴³. As established, active malignant tumors are a leading cause of VTE from an etiological perspective. Therefore, *E2F1* may play a role, albeit potentially limited, in VTE pathogenesis⁴⁴. VTE is a complex disease influenced by a multitude of factors, including environmental and genetic components. Numerous genes have been identified as conferring susceptibility to VTE. The established association between VTE and inflammation is further supported by the involvement of inflammatory cells like monocytes, platelet aggregation, and C-reactive protein⁴⁴. Consequently, immune gene mutations linked to inflammation may contribute to VTE pathogenesis. Interestingly, *E2F1* upregulation has been observed in both VTE and Behcet's disease, suggesting a potential connection between *E2F1*, VTE, and BD.

GATA3, a transcription factor belonging to the GATA family, plays a critical role in regulating cell differentiation and proliferation. Studies have shown a potential association between the immune response and low expression of *GATA3* (Th2) in patients with Neuro-Behcet's disease (NBD)⁴⁵. Interestingly, in active BD with skin lesions, the expression ratio of Th2/Th1 (*GATA3/Tbet*) in these lesions exhibits a tendency to increase. This suggests a potential bias towards the *GATA3/Th2* axis during T cell immune activation within BD lesions⁴⁶. These findings collectively indicate a close relationship between *GATA3* and BD. Additionally, *GATA3* has been demonstrated to possess tumor suppressive effects in various cancers, including osteosarcoma⁴⁷. Nevertheless, the relationship between *GATA3* and VTE remains uncertain.

Histone deacetylase 5 (*HDAC5*) is a member of the class IIa histone deacetylase family. Primarily localized within the nucleus of vascular smooth muscle cells (VSMCs), *HDAC5* functions as a transcriptional repressor, suppressing the expression of its target genes. Upon phosphorylation, *HDAC5* interacts with specific proteins, leading to the de-repression of genes associated with vascular diseases⁴⁸. Interestingly, *HDAC5* expression appears to be dynamically regulated in the endothelium. Studies suggest overexpression of *HDAC5* in endothelial cells during venous return, while normal endothelial venous flow conditions are associated with down-regulated *HDAC5* expression⁴⁹. Additionally, *HDAC5* has been implicated in impaired angiogenesis, a hallmark of scleroderma (SSc). *HDAC5* inhibits pro-angiogenic factors in endothelial cells, contributing to the disease. Silencing *HDAC5* in SSc-affected endothelial cells restores normal angiogenic function⁵⁰. This collective evidence strongly suggests a close link between *HDAC5* and vascular diseases. Additionally, research has shown that inhibiting *HDAC5* can enhance tissue-type plasminogen activator (t-PA) production, a molecule with antithrombotic properties, thereby ameliorating intravascular thrombosis⁵¹. While a study demonstrated that mutant forms of *HDAC5* (*HDAC5AA*) activate the mTOR pathway and promote the survival and regeneration of retinal ganglion

cells (RGCs)⁵², it is important to note that BD can significantly affect patients' eyes, establishing a connection between BD and vascular-related diseases. However, the precise relationship between BD and VTE remains to be elucidated.

MutS homolog 2 (MSH2) is a well-established key component of the mismatch repair (MMR) system. The MMR system plays a critical role in maintaining genomic stability by recognizing and repairing mismatched nucleotides during DNA replication. Mutations in genes encoding components of the MMR system, including MSH2, have been implicated in the pathogenesis of various autoimmune and autoinflammatory diseases. These mutations may disrupt the tightly regulated activity of pattern recognition receptors (PRRs) and their signaling pathways⁵³. Moreover, MSH2 has been demonstrably linked to the initiation and progression of cancer, potentially impacting tumorigenesis, development, and immune regulation⁵⁴. However, current research on MSH2 function primarily focuses on specific cancer types, with limited investigation into its role in VTE and BD.

BD is a chronic, systemic, and inflammatory vascular disorder characterized by vasculitis and endothelial injury. While the precise etiology of VTE associated with BD remains to be fully elucidated, the leading hypothesis suggests a complex interplay between neutrophil activation, endothelial dysfunction, and coagulation abnormalities in the pathogenesis of BD-induced VTE^{4,5}. Neutrophil activation leads to the release of high levels of neutrophil extracellular traps (NETs), which contribute to endothelial cell injury and alterations in coagulation function. Additionally, inflammation triggers the release of proinflammatory factors and chemokines, which in turn activate platelets, leukocytes, and endothelial cells. This activation cascade results in vascular endothelial dysfunction, ultimately promoting thrombosis. BD patients exhibit a pre-thrombotic state characterized by dysregulation of the hemostasis, coagulation, and anticoagulation systems. This dysregulation, potentially arising from alterations in the coagulation mechanism and platelet activation, ultimately contributes to the development of VTE in these patients. The significantly elevated risk of VTE in BD patients presents a substantial clinical challenge. A comprehensive understanding of the relationship between BD and VTE, alongside the identification of reliable disease markers, is crucial for facilitating early diagnosis and effective treatment, thereby improving the prognosis for BD patients.

Subsequent analysis of immune cell infiltration in VTE patients revealed a significantly higher prevalence of macrophages compared to the control group. Conversely, VTE samples displayed a lower prevalence of neutrophils. This finding appears to contradict previous literature reporting an increase in neutrophil numbers in VTE⁵⁵. The observed discrepancy may be attributed to the inclusion of recurrent and chronic VTE samples in the utilized GEO dataset. Neutrophils undergo activation during the acute phase of DVT development; however, their numbers are likely to decline in the chronic phase due to depletion. Existing literature supports the significant role of macrophages in the pathogenesis and progression of both BD and VTE^{56–58}. Inflammasome activation in macrophages leads to the release of tissue factor, a critical trigger for thrombosis⁵⁹. Our findings regarding macrophage prevalence align with these previous studies. Consequently, we investigated the potential interrelationship between the four identified hub genes (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*) and immune cell infiltration. This analysis revealed a positive correlation between immune-associated macrophages and the expression of all four hub genes (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*). In conclusion, these findings suggest that modulating macrophage populations may present a promising therapeutic strategy for managing VTE in patients with BD.

The ssGSEA analysis revealed significant associations between the four identified hub genes and various biological processes, including hypoxia, heme metabolism, coagulation, cholesterol homeostasis, and angiogenesis. These findings suggest that the four hub genes may be intricately linked to the initiation and progression of VTE in BD patients. Current treatment strategies for VTE in BD primarily focus on immunosuppressants, with anticoagulation therapy potentially unnecessary for managing DVT in this context. However, the high 5-year recurrence rate of 36.5%⁶⁰ necessitates the exploration of more effective therapeutic options. To address this need, we employed the Enrichr database to screen for potential therapeutic agents targeting the four hub genes. The analysis identified several candidate drugs, including chlorpromazine, dronabinol, trabectedin, vorinostat, and carmustine, which exhibited potential efficacy in targeting the hub genes associated with BD-induced VTE.

Our research elucidates potential disease mechanisms and identifies biomarkers related to VTE induced by BD, employing a synthesis of bioinformatics and machine learning methodologies. The nomogram model crafted in our study exhibits significant predictive value for VTE in BD patients. Additionally, this study provides some directions for future studies delving into the molecular mechanisms underlying the pathogenesis of VTE precipitated by BD. However, despite these advantages, the study has limitations. Firstly, despite utilizing validation datasets and clinical samples for diagnostic assessment, further experimental research is essential to validate and explore the underlying mechanisms of VTE induced by BD. Secondly, this study was based on bioinformatics analysis, and the verification of hub genes was constrained by relatively small validation samples due to sample acquisition challenges, so further validation using multicenter studies involving larger sample sizes should be performed. Thirdly, although predicted therapeutic agents have been screened, the absence of concrete experimental corroboration means that significant further research is required to evaluate the efficacy of drug treatments targeting these hub genes for BD with VTE patients.

Conclusion

This study employed a systematic approach to identify four candidate hub genes (*E2F1*, *GATA3*, *HDAC5*, and *MSH2*). A nomogram was subsequently constructed to facilitate the diagnosis of VTE in Budd-Chiari syndrome patients. Furthermore, analysis of immune cell infiltration revealed dysregulation in their relative proportions, suggesting a potential role for macrophages in VTE development within the context of BD. Importantly, ssGSEA provided insights into potential underlying mechanisms associated with BD-induced VTE. Additionally, the study identified potential therapeutic agents for further investigation.

Data availability

The datasets used in this study can be easily accessed and downloaded from the publicly accessible GEO database. GSE209567: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209567> (GPL570[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array). GSE48000: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48000> (GPL10558 Illumina HumanHT-12 V4.0 expression beadchip). GSE19151: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19151> (GPL571[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array). The study's codes have been uploaded to the GitHub repository (<https://github.com/wangjiajia-one/bioinformatics-R-code-1.0-git>).

Received: 7 April 2024; Accepted: 5 July 2024

Published online: 10 July 2024

References

- Rodolfi, S., Nasone, I., Folci, M., Selmi, C. & Brunetta, E. Autoinflammatory manifestations in adult patients. *Clin. Exp. Immunol.* **210**, 295–308. <https://doi.org/10.1093/cei/uxac098> (2022).
- Davatchi, F. *et al.* Behcet's disease: Epidemiology, clinical manifestations, and diagnosis. *Expert Rev. Clin. Immunol.* **13**, 57–65. <https://doi.org/10.1080/1744666x.2016.1205486> (2017).
- Pamuk, O. N. Behcet's Syndrome. *N. Engl. J. Med.* **390**, 1730. <https://doi.org/10.1056/NEJM2403905> (2024).
- Bettioli, A. *et al.* Vascular Behcet syndrome: From pathogenesis to treatment. *Nat. Rev. Rheumatol.* **19**, 111–126. <https://doi.org/10.1038/s41584-022-00880-7> (2023).
- Toledo-Samaniego, N. *et al.* Arterial and venous involvement in Behcet's syndrome: A narrative review. *J. Thromb. Thrombol.* **54**, 162–171. <https://doi.org/10.1007/s11239-022-02637-1> (2022).
- Fernández-Bello, I., López-Longo, F. J., Arias-Salgado, E. G., Jiménez-Yuste, V. & Butta, N. V. Behcet's disease: New insight into the relationship between procoagulant state, endothelial activation/damage and disease activity. *Orphanet. J. Rare Dis.* **8**, 81. <https://doi.org/10.1186/1750-1172-8-81> (2013).
- Oleksiuk-Bójko, M. & Lisowska, A. Venous thromboembolism: Why is it still a significant health problem?. *Adv. Med. Sci.* **68**, 10–20. <https://doi.org/10.1016/j.advms.2022.10.002> (2023).
- Wang, D. *et al.* Risk of venous thromboembolism in patients undergoing gastric cancer surgery: Protocol for a systematic review and meta-analysis. *BMJ Open* **10**, e033267. <https://doi.org/10.1136/bmjopen-2019-033267> (2020).
- Cunningham, M. S., Preston, R. J. & O'Donnell, J. S. Does antithrombotic therapy improve survival in cancer patients?. *Blood Rev.* **23**, 129–135. <https://doi.org/10.1016/j.blre.2008.10.002> (2009).
- Goldberg, J. B. *et al.* Surgical management and mechanical circulatory support in high-risk pulmonary embolisms: Historical context, current status, and future directions: A scientific statement from the american heart association. *Circulation* **147**, e628–e647. <https://doi.org/10.1161/cir.0000000000001117> (2023).
- Liu, H. *et al.* Identification of the immune-related biomarkers in Behcet's disease by plasma proteomic analysis. *Arthr. Res. Therapy* **25**, 92. <https://doi.org/10.1186/s13075-023-03074-y> (2023).
- Lu, S., Lijuan, R., Tang, Q. H., Liu, Q. L. & Xian-Lan, Z. Bioinformatics analysis and identification of genes and molecular pathways involved in venous thromboembolism (VTE). *Ann. Vasc. Surg.* **74**, 389–399. <https://doi.org/10.1016/j.avsg.2021.02.020> (2021).
- Chen, X., Cao, J., Ge, Z. & Xia, Z. Correlation and integration of circulating miRNA and peripheral whole blood gene expression profiles in patients with venous thromboembolism. *Bioengineered* **12**, 2352–2363. <https://doi.org/10.1080/21655979.2021.1935401> (2021).
- Al-Araji, A. & Kidd, D. P. Neuro-Behcet's disease: Epidemiology, clinical characteristics, and management. *Lancet Neurol.* **8**, 192–204. [https://doi.org/10.1016/s1474-4422\(09\)70015-8](https://doi.org/10.1016/s1474-4422(09)70015-8) (2009).
- Huang, S. L. *et al.* Recent advances on the molecular mechanism and clinical trials of venous thromboembolism. *J. Inflamm. Res.* **16**, 6167–6178. <https://doi.org/10.2147/jir.S439205> (2023).
- Momi, S. *et al.* Proline-rich tyrosine kinase Pyk2 regulates deep vein thrombosis. *Haematologica* **107**, 1374–1383. <https://doi.org/10.3324/haematol.2021.279703> (2022).
- Yang, Y. *et al.* Biomarkers prediction and immune landscape in ulcerative colitis: Findings based on bioinformatics and machine learning. *Comput. Biol. Med.* **168**, 107778. <https://doi.org/10.1016/j.combiomed.2023.107778> (2024).
- Xue, A. *et al.* Study on the neuroprotective effect of Zhimu-Huangbo extract on mitochondrial dysfunction in HT22 cells induced by D-galactose by promoting mitochondrial autophagy. *J. Ethnopharmacol.* **318**, 117012. <https://doi.org/10.1016/j.jep.2023.117012> (2024).
- Sun, T. H., Wang, C. C., Wu, Y. L., Hsu, K. C. & Lee, T. H. Machine learning approaches for biomarker discovery to predict large-artery atherosclerosis. *Sci. Rep.* **13**, 15139. <https://doi.org/10.1038/s41598-023-42338-0> (2023).
- Zhang, W. Y. *et al.* Analysis and validation of diagnostic biomarkers and immune cell infiltration characteristics in pediatric sepsis by integrating bioinformatics and machine learning. *World J. Pediatr.* **19**, 1094–1103. <https://doi.org/10.1007/s12519-023-00717-7> (2023).
- Xing, L. *et al.* Exploration of biomarkers of psoriasis through combined multiomics analysis. *Mediat. Inflamm.* **2022**, 7731082. <https://doi.org/10.1155/2022/7731082> (2022).
- Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47. <https://doi.org/10.1093/nar/gkv007> (2015).
- Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–d592. <https://doi.org/10.1093/nar/gkac963> (2023).
- Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287. <https://doi.org/10.1089/omi.2011.0118> (2012).
- Lancaster, S. M., Sanghi, A., Wu, S. & Snyder, M. P. A customizable analysis flow in integrative multi-omics. *Biomolecules* **10**, 1606. <https://doi.org/10.3390/biom10121606> (2020).
- Huang, M. L., Hung, Y. H., Lee, W. M., Li, R. K. & Jiang, B. R. SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *Sci. World J.* **2014**, 795624. <https://doi.org/10.1155/2014/795624> (2014).
- Dawkins, J. J. *et al.* Gut metabolites predict *Clostridioides difficile* recurrence. *Microbiome* **10**, 87. <https://doi.org/10.1186/s40168-022-01284-1> (2022).
- Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77. <https://doi.org/10.1186/1471-2105-12-77> (2011).
- Buciuc, M. *et al.* Utility of FDG-PET in diagnosis of Alzheimer-related TDP-43 proteinopathy. *Neurology* **95**, e23–e34. <https://doi.org/10.1212/wnl.0000000000009722> (2020).

31. Davatchi, F. *et al.* The International Criteria for Behcet's Disease (ICBD): A collaborative study of 27 countries on the sensitivity and specificity of the new criteria. *J. Eur. Acad. Dermatol. Venereol.* **28**, 338–347. <https://doi.org/10.1111/jdv.12107> (2014).
32. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457. <https://doi.org/10.1038/nmeth.3337> (2015).
33. Mukhopadhyay, S. *et al.* Fibrinolysis and inflammation in venous thrombus resolution. *Front. Immunol.* **10**, 1348. <https://doi.org/10.3389/fimmu.2019.01348> (2019).
34. Liederman, Z., Chan, N. & Bhagirath, V. Current challenges in diagnosis of venous thromboembolism. *J. Clin. Med.* **9**, 11. <https://doi.org/10.3390/jcm9113509> (2020).
35. Wang, J. *et al.* Current challenges in the prevention and management of post-thrombotic syndrome-towards improved prevention. *Int. J. Hematol.* **118**, 547–567. <https://doi.org/10.1007/s12185-023-03651-6> (2023).
36. Khoshbakht, S., Baškurt, D., Vural, A. & Vural, S. Behcet's disease: A comprehensive review on the role of HLA-B*51, antigen presentation, and inflammatory cascade. *Int. J. Mol. Sci.* **24**, 16382. <https://doi.org/10.3390/ijms242216382> (2023).
37. Yang, J. Y., Park, M. J., Park, S. & Lee, E. S. Increased senescent CD8+ T cells in the peripheral blood mononuclear cells of Behcet's disease patients. *Arch. Dermatol. Res.* **310**, 127–138. <https://doi.org/10.1007/s00403-017-1802-8> (2018).
38. Bochenek, M. L., Schütz, E. & Schäfer, K. Endothelial cell senescence and thrombosis: Ageing clots. *Thromb. Res.* **147**, 36–45. <https://doi.org/10.1016/j.thromres.2016.09.019> (2016).
39. Schmitt, C. A. *et al.* COVID-19 and cellular senescence. *Nat. Rev. Immunol.* **23**, 251–263. <https://doi.org/10.1038/s41577-022-00785-2> (2023).
40. Le Joncour, A., Cacoub, P., Boulaftali, Y. & Saadoun, D. Neutrophil, NETs and Behcet's disease: A review. *Clin. Immunol. Orlando Fla.* **250**, 109318. <https://doi.org/10.1016/j.clim.2023.109318> (2023).
41. Yuan, Y., Li, X. & Li, M. Overexpression of miR-17-5p protects against high glucose-induced endothelial cell injury by targeting E2F1-mediated suppression of autophagy and promotion of apoptosis. *Int. J. Mol. Med.* **42**, 1559–1568. <https://doi.org/10.3892/ijmm.2018.3697> (2018).
42. Fouad, S., Hauton, D. & D'Angiolella, V. E2F1: Cause and consequence of DNA replication stress. *Front. Mol. Biosci.* **7**, 599332. <https://doi.org/10.3389/fmolsb.2020.599332> (2020).
43. Dong, W. & Zhan, C. Bioinformatic-based mechanism identification of E2F1-related ceRNA and E2F1 immunoassays in hepatocellular carcinoma. *J. Gastrointest. Oncol.* **13**, 1915–1926. <https://doi.org/10.21037/jgo-22-674> (2022).
44. Gao, L.-N., Li, Q., Xie, J.-Q., Yang, W.-X. & You, C.-G. Immunological analysis and differential genes screening of venous thromboembolism. *Hereditas* **158**, 1. <https://doi.org/10.1186/s41065-020-00166-6> (2021).
45. Belghith, M. *et al.* Cerebrospinal fluid IL-10 as an early stage discriminative marker between multiple sclerosis and neuro-Behcet disease. *Cytokine* **108**, 160–167. <https://doi.org/10.1016/j.cyto.2018.03.039> (2018).
46. Kacem, O., Kaabachi, W., Dhifallah, I. B., Hamzaoui, A. & Hamzaoui, K. Elevated expression of TSLP and IL-33 in Behcet's disease skin lesions: IL-37 alleviate inflammatory effect of TSLP. *Clin. Immunol.* **192**, 14–19. <https://doi.org/10.1016/j.clim.2018.03.016> (2018).
47. Ma, L., Xue, W. & Ma, X. GATA3 is downregulated in osteosarcoma and facilitates EMT as well as migration through regulation of slug. *OncoTargets Therapy* **11**, 7579–7589. <https://doi.org/10.2147/ott.S176534> (2018).
48. Truong, V., Jain, A., Anand-Srivastava, M. B. & Srivastava, A. K. Angiotensin II-induced histone deacetylase 5 phosphorylation, nuclear export, and Egr-1 expression are mediated by Akt pathway in A10 vascular smooth muscle cells. *Am. J. Physiol. Heart Circ. Physiol.* **320**, H1554–H1565. <https://doi.org/10.1152/ajpheart.00683.2020> (2021).
49. Chang, S.-F. *et al.* Blood reflux-induced epigenetic factors HDACs and DNMTs are associated with the development of human chronic venous disease. *Int. J. Mol. Sci.* **23**, 20. <https://doi.org/10.3390/ijms232012536> (2022).
50. Tsou, P. S. *et al.* Histone deacetylase 5 is overexpressed in scleroderma endothelial cells and impairs angiogenesis via repression of proangiogenic factors. *Arthr. Rheumatol.* **68**, 2975–2985. <https://doi.org/10.1002/art.39828> (2016).
51. Larsson, P. *et al.* Role of histone acetylation in the stimulatory effect of valproic acid on vascular endothelial tissue-type plasminogen activator expression. *PLoS one* **7**, e31573. <https://doi.org/10.1371/journal.pone.0031573> (2012).
52. Pita-Thomas, W., Mahar, M., Joshi, A., Gan, D. & Cavalli, V. HDAC5 promotes optic nerve regeneration by activating the mTOR pathway. *Exp. Neurol.* **317**, 271–283. <https://doi.org/10.1016/j.exn.2019.03.011> (2019).
53. Kivanc, D. & Dasdemir, S. The relationship between defects in DNA repair genes and autoinflammatory diseases. *Rheumatol. Int.* **42**, 1–13. <https://doi.org/10.1007/s0296-021-04906-3> (2021).
54. Qiu, W. *et al.* Analysis of the expression and prognostic value of MSH2 in pan-cancer based on bioinformatics. *BioMed. Res. Int.* **2021**, 1–12. <https://doi.org/10.1155/2021/948573> (2021).
55. Wolberg, A. S. *et al.* Venous thrombosis. *Nat. Rev. Dis. Prim.* **1**, 15006. <https://doi.org/10.1038/nrdp.2015.6> (2015).
56. Zhan, H. *et al.* Novel insights into gene signatures and their correlation with immune infiltration of peripheral blood mononuclear cells in Behcet's disease. *Front. Immunol.* **12**, 794800. <https://doi.org/10.3389/fimmu.2021.794800> (2021).
57. Hu, D. & Guan, J. L. The roles of immune cells in Behcet's disease. *Adv. Rheumatol. Lond. Engl.* **63**, 49. <https://doi.org/10.1186/s42358-023-00328-w> (2023).
58. Abduhalike, R., Abduwayitay, A., Juan, S. & MaheMuti, A. Study on the mechanism of NLRP3/IL-1/ NF- κ B signaling pathway and macrophage polarization in the occurrence and development of VTE. *Ann. Vasc. Surg.* **89**, 280–292. <https://doi.org/10.1016/j.avsg.2022.09.056> (2023).
59. Zhang, Y. *et al.* Inflammasome activation promotes venous thrombosis through pyroptosis. *Blood Adv.* **5**, 2619–2623. <https://doi.org/10.1182/bloodadvances.2020003041> (2021).
60. Desbois, A. C. *et al.* Immunosuppressants reduce venous thrombosis relapse in Behcet's disease. *Arthr. Rheum.* **64**, 2753–2760. <https://doi.org/10.1002/art.34450> (2012).

Acknowledgements

We extend our sincere appreciation to SangerBox for generously providing us with the data analysis platform. Gratitude is also extended to the Home for Researchers editorial team for their language editing services.

Author contributions

Jiajia Wang and Chunjiang Liu designed and wrote the manuscript. During the analysis phase, Chunjiang Liu, Yuan Wang, and Zhifeng Wu conducted a comparative analysis using bioinformatics tools. Xiaoqi Tang and Guohua Wang participated in interpreting both the image data and numerical results. Chunjiang Liu, Zhifeng Wu and Jiajia Wang performed Clinical sample collection and experimental validation. The manuscript was collectively drafted and approved for submission by all authors.

Funding

This work was supported by research grants from the Key Program of the Scientific Foundation of Shaoxing People's Hospital (2023YA06).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-66973-3>.

Correspondence and requests for materials should be addressed to J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024