

GLRX5 as a Differential Diagnosis Marker for Kashin-Beck Disease and Osteoarthritis and Its Clinical Significance

Qinglong Wang

The First Affiliated Hospital of Harbin Medical University

Zhe Zhao

The First Affiliated Hospital of Harbin Medical University

Wantao Wang

The First Affiliated Hospital of Harbin Medical University

Zhipeng Huang

The Second Affiliated Hospital of Air Force military Medical University

Wenbo Wang (✉ wenbowang1967@163.com)

The First Affiliated Hospital of Harbin Medical University

Research Article

Keywords: Kashin-Beck disease, Weighted gene correlation network analysis, Differentially expressed genes, Diagnostic biomarkers, Peripheral blood, LASSO, SVM-RFE, mutations

Posted Date: November 30th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-971437/v1>

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

[Read Full License](#)

Abstract

Background: Kashin-Beck disease (KBD) is currently an endemic form of osteoarthritis. In this study, we explored novel KBD diagnostic biomarkers.

Methods: The GSE59446 dataset was used to conduct Weighted Gene Co-expression Network Analysis (WGCNA) and differentially expressed genes (DEGs) analysis with peripheral blood samples of 100 healthy individuals and 100 KBD patients. As part of the gene ontology pathway enrichment analysis, genes related to SONFH and DEGs were selected from the extraction module. Then, central DEGs were selected for LASSO analysis, and, based on SVM-RFE and DEG results, overlapping genes were identified as key KBD genes. Next, we analyzed the correlations between the selected genes and age, gender, and other factors to eliminate their influences on gene expression. Finally, we evaluated the diagnostic value of key KBD genes using case information collected by us.

Results: Seven gene co-expression modules were created using WGCNA. The turquoise module was identified as a KBD key module since it showed the highest correlation to KBD. The functional enrichment analysis revealed that the genes associated with this key module were mainly involved in mitochondrial reactions, protein heterooligomerization, and negatively regulating cysteine-type endopeptidase-dependent apoptotic processes. Additionally, 12 key genes were identified using the LASSO analysis, 5 major genes using SVM-RFE analysis, and 36 DEGs were screened through the "limma" R package. The GLRX5 gene - pivotal in DEGs, LASSO, and SVM-RFE - was further aggregated as the key KBD gene. Correlation analyses confirmed the GLRX5 diagnostic value for KBD and that it was not related to age, gender, and other factors. Finally, data from our patients demonstrated that GLRX5 can be a KBD diagnostic biomarker.

Conclusions: We demonstrated that the target gene GLRX5 can be a KBD non-invasive diagnosis biomarker.

Introduction

Kashin-Beck disease (KBD) is an endemic form of osteochondrosis with irreversible clinical and pathological consequences. KBD's pathological manifestations comprehend extracellular matrix (ECM) degradation, chondrocyte necrosis, and apoptosis [1]. However, if KBD occurs at a young age some advanced patients can develop dwarfism [2]. Although KBD's incidence and prevalence have been declining, according to the 2016 Statistical Yearbook (<http://www.nhfpc.gov.cn>), approximately 567,600 patients were infected in China and 104 million residents live in endemic regions [3]. It is worth noting that, in Tibet, the newly diagnosed incidence rate of adolescent patients is 6.67%, and environmental risk factors still exist in some endemic areas such as selenium deficiency and toxins-contaminated foods [4]. At present, KBD diagnosis is usually based on medical history, clinical manifestations, and combined imaging technologies. In most cases, patients' conditions deteriorate without an early diagnosis.

Therefore, these patients can receive a poor prognosis and ineffective treatment options. To improve KBD patients' prognosis, biomarkers that can diagnose KBD earlier should be explored.

Many studies have screened differentially expressed genes (DEGs) to identify disease biomarkers. However, there are few reports on highly related genes that can have important effects on certain diseases. In this case, gene association patterns between samples can be examined using systems biology methods such as the Weighted Gene Co-expression Network Analysis (WGCNA). Based on the interconnection between gene sets, the WGCNA can identify highly coordinated gene sets. Also, a candidate biomarker gene, or therapeutic target, can be identified based on gene and gene set associations [5]. Finally, the modules corresponding to which features can be identified by analyzing the connectivity between them and clinical information. WGCNA has been extensively used for the identification of genes related to the different diseases' clinical features, such as Infantile Hemangioma [6], aortic aneurysm [7, 8], osteoarthritis [9], acute myocardial infarction [10], bladder cancer [11] and pancreatic cancer [12].

In this study, we first retrieved microarray KBD and control PBMC datasets from the Gene Expression Omnibus (GEO) database. After screening DEGs using the "limma" R package, we performed LASSO and SVM-RFE analyses to identify key KBD genes with the highest significant correlation levels. Then, based on the WGCNA screening of the GSE59446 dataset DEGs, we used the Cytoscape software to integrate DEGs into a protein-protein interaction network and define the overlapping genes between central genes in key modules and central genes in DEGs. Finally, the GLRX5 gene was identified as a new KBD diagnostic biomarker among the GSE59446 dataset key genes. Additionally, other potential KBD diagnostic biomarkers can be further identified from our study and used to gain new insights into their underlying mechanisms on KBD.

Methods

Subjects data collection and statement. Data from the Gene Expression Omnibus (GEO) database were retrieved from GSE59446, which including 100 controls samples and 100 KBD samples (Table S1). Clinically relevant samples were collected from the patient's peripheral blood and conducted in accordance with the declaration of Helsinki. Our study was approved by The Human Ethics Committee of the First Affiliated Hospital of Harbin Medical University (Harbin, China) and all subjects gave informed consent to be treated by our hospital (EC-HS-2020112). Patients were diagnosed according to the WS/T 207-2010 criteria, and those presenting any other osteoarthritis-related diseases were excluded. Collected blood samples from 5 healthy laboratory staff volunteers who had not been diagnosed with chronic inflammation or other diseases before, as a reference for the clinical cohort. Finally, a total of 10 subjects were included in the study, including 5 healthy controls and 5 KBD patients. All subjects provided written informed consent.

Differentially expressed genes (DEG) screening. The GSE59446 dataset original data was read and the RMA algorithm was applied for background correction and data normalization. DEGs were filtered using

the "limma" R package [13]. The "ggplot2" R package was used to draw the DEGs' volcano map [14]. Also, the "gplots" R package was used to create a DEGs' heatmap [15]. DEGs with $p < 0.05$ and $|\log_2FC| > 1$ were considered statistically significant.

WGCNA. After ranking genes from the largest to the smallest according to the median absolute deviation, we used the "WGCNA" R package to list all genes in WGCNA [5, 16]. Then, we used the "pickSoftThreshold" function (WGCNA software package) to filter out power parameters from 1 to 20. We choose a minimum power value of 0.85 for the independence degree and a soft threshold value of 4. Finally, we selected 7 modules. To visualize the generated gene network, we used the difference between the topological overlap matrix and its cluster dendrogram.

Key modules identification. A heatmap was used to display the correlation between gene features in modules and clinical features, which allowed the identification of key modules that were significantly related to clinical features. The key modules of synchronous frequency hopping are closely related to the status of synchronous frequency hopping. Thus, to verify the exact relationship between modules and characters, we compared the correlation between general service and management. All correlation analyses were conducted using Pearson correlations by the 'WGCNA' R package [5].

Protein-protein interaction (PPI) network analysis. The Cytoscape software (v. 3.7.1) was used to construct the PPI network with the above key modules genes.

Key modules functional enrichment analyses. To evaluate the key modular genes and DEGs biological functions, the annotation, visualization, and comprehensive discovery database tool, DAVID (version 6.8) [17, 18], was used to perform gene ontology (GO) functional enrichment analyses. GO terms with $p < 0.05$ were considered significant. For the genes in key modules, the functional enrichment of the top 10 GO pathways, sorted by gene count, was visualized. For DEGs between control and KBD patients, all the important functionally enriched GO pathways were visualized.

Screening of diagnostic. To perform feature selection and filter OA diagnostic markers, we used the least absolute shrinkage and selection operator (LASSO) logic back (Regression Shrinkage and Selection Via the Lasso) and support vector machine recursive feature elimination (SVM-RFE) [19]. The GSE59446 dataset expression matrix was subjected to quality control to establish a new dataset. Then, an independent dataset was used to verify the diagnostic efficiency of the obtained markers. The LASSO algorithm was used in conjunction with the "glmnet" R package [20]. Additionally, SVM-RFE is a general machine learning technique based on support vector machines in which feature vectors generated by SVM are removed to find the best variables [21]. SVM modules were created using the "e1071" R package to better identify the diagnostic value of these biomarkers. Then, we combined the genes selected by LASSO and SVM-RFE algorithms to further analyze them. A double-sided $p < 0.05$ was considered statistically significant.

Key genes clinical relevance. Use the "ggpubr" software package to analyze the relationship between key genes and age, stage, gender, and experimental groupings.

Blood sampling and RNA isolation for RT-PCR. To analyze each subject's gene expression, 3 mL of peripheral blood was collected into heparinized vacuum tubes (Suzhou Bidi Medical Equipment Co., Ltd. China). An SYSMEXXE-2100 machine was used to determine the white blood cells count (Sysmex Corporation, Japan). Blood was centrifuged at 1,500 x g for 20 min to separate PBMC from plasma. Cell pellets were resuspended in Hanks Balanced Salt Solution (Solarbio). A 15 mL Falcon tube was filled with 5 mL Lympholyte-H (Solarbio) and centrifuged at 1,500 × g for 40 min. Then, they were rinsed with cold Hank balanced salt solution and stored in TRIZOL™ (ThermoFisher SCIENTIFIC). The RNA was extracted by the TRIZOL method. In the RT-PCR experiment, gene expression relative values were calculated using $2^{-\Delta\Delta ct}$.

Results

Data normalization processing. Before the GSE59446 dataset standardization, data distribution was uneven (Supplementary Table S1). After standardization, the data median was at the same level without significant differences.

WGCNA network construction. Following the GSE59446 dataset preprocessing and filtering, data were sorted from largest to smallest according to the median absolute deviation for further analyses. Samples' cluster analyses included in the WGCNA revealed no significant differences between them (Fig. 1A). By setting the power value (β) to 4, a scale-independent value was reached at 0.95 (Fig. 1B, C), and the co-expression network achieved scale-free topology requirements (Fig. 1D). The $\beta = 4$ was also used to generate a hierarchical clustering tree and to represent each module with different colors.

Key modules identification and visualization. A hierarchical clustering tree was generated by setting $\beta = 4$. Seven co-expression modules were identified and the merge threshold was 0.25 (Fig. 2A). After testing the obtained modules, the turquoise module, containing 50 genes, presented the best correlation to KBD status (Fig. 2B). All genes were included in the heatmap (Fig. 2C). Then, feature gene pedigree maps and heatmaps were used to identify related genome features. These results also showed that the turquoise module was significantly correlated with the synchronization frequency hopping state (Fig. 2C). A strong and significant correlation was found between the KBD in the turquoise module and the GS of MM, indicating that the genes in the turquoise module were highly related to KBD (Fig. 2D).

Turquoise module genes functional enrichment analyses. To identify potential biological processes related to KBD, the turquoise module genes were analyzed using GO enrichment analysis. Then, the top 10 functional enriched GO pathways, ranked by gene count, were visualized. The GO enrichment analysis results indicated that genes related to mitochondrial metabolism, protein heterooligomerization, negative regulation of cysteine-type endopeptidase activity involved in apoptotic processes, and alpha-beta T cell differentiation were significantly enriched (Fig. 3).

DEG analysis and identification. Subsequently, we screened DEGs between healthy individuals and KBD patients' peripheral blood, based on the GSE59446 dataset. Thus, 36 DEGs were detected, including 20

upregulated and 16 downregulated genes (Fig. 4A and B, and Supplementary Table S2). The modular biological gene interaction analysis in WGCNA (Fig. 4C) showed that GLRX5 was related to TNFSF11, BIRC3, ACSL6, ANAPC1, AF116618, HBA2, and *abl1 v-abl* Abelson genes.

Diagnostic markers screening and verification. From DEGs, we identified 12 genes based on the LASSO logistic regression and 5 based on the SVM-RFE algorithm as KBD diagnostic markers (Fig. 5A and B, and C, respectively). Overlapping genes obtained by the two algorithms with DEGs are shown in Supplementary Table S3. Finally, a KBD diagnosis-related gene was obtained (Fig. 5D).

Identify and verify key genes through the database. Among the 36 central DEGs, GLRX5 was identified as a key KBD gene and was selected for subsequent analyses. We visualized the GLRX5 expression in the GSE59446 dataset. We found that the KBD group had a significantly lower GLRX5 expression than the healthy group ($p < 0.001$, Fig. 6A). Additionally, neither gender nor KBD stage significantly affected the GLRX5 expression ($p > 0.05$, Fig. 6B and C). Increasing age was related to GLRX5 gene expression decreases ($p < 0.001$, Fig. 6D). In another database, GSE32127, we also found that the GLRX5 relative expression decreased in controls vs. KBD in PBMCs cells. Then, we measured the GLRX5 gene expression in patients' blood. Compared to the control group, the KBD group showed significantly low levels of GLRX5 expression in PBMCs ($p < 0.001$, Fig. 6E).

Discussion

KBD is characterized by a localized cartilage necrosis area in the deep articular cartilage part and cartilage plate growth. The pathogenesis of KBD is multifactorial, with both genetic and environmental factors at play [22]. In clinical terms, KBD can be difficult to diagnose or predict since its early symptoms are not apparent. Therefore, some KBD patients have osteoarthropathy with a high permanent disability rate. Although current treatments presented only limited KBD curative effects, food replacement treatment can be effective at the early stages. Consequently, new KBD biomarkers are required for early diagnosis.

Advances in genomics and proteomics have made important contributions to the diagnosis and treatment of many rare diseases and pathological conditions [23]. In this study, 100 peripheral blood samples from KBD patients and 100 samples from healthy individuals were retrieved from the GSE59446 biochip. Machine learning methods based on support vector machines such as SVM-RFE were used to subtract feature vectors to find the best variables [24]. Another machine learning method, LASSO logistic regression was used to determine variables by identifying the lowest classification error value λ [19]. Using the two algorithms, main feature variables were screened and better classification models built. Therefore, we determined the GLRX5 gene using LASSO logistic regression and SVM-RFE methods as a KBD diagnostic marker. Each algorithm possesses its inherent characteristics and the GLRX5 selection using a LASSO and SVM-RFE combination proved to be reliable in further validations.

In the WGCNA analysis, we discovered that the KBD status was closely related to the turquoise module with 50 genes. These modular genes were highly enriched in mitochondrial responses, protein heterooligomerization, Cysteine-type endopeptidases that negatively regulate apoptotic processes, and

alpha-beta T cell differentiation. It has been shown that mitochondrial metabolism contributes significantly to KBD development. Several studies have demonstrated that in cartilage cells treated with T-2 toxin, mitochondrial dysfunction was related to mitochondrial swelling, mitochondrial outer membrane rupture, and apoptosis factors release (e.g. cytochrome c). First, cytochrome c is released from mitochondria and binds to Apaf-1 (apoptotic protease activator 1). Then, procaspase-9 is activated by this complex, which cleaves downstream caspase cascades, such as caspase-3 [25]. A recent study showed that cytochrome c is released from mitochondria into the cytoplasm, and caspase-9 and caspase-3 are activated in chondrocytes treated with T-2 toxin. Thus, the T-2 toxin can lead to human articular chondrocytes apoptosis by mitochondrial function inhibition [26].

KBD diagnosis can be easier with peripheral blood than cartilages since it can be more easily collected and result in fewer wounds. Additionally, peripheral blood changes can be rapidly triggered by disease states. Consequently, it could be useful for early KBD diagnosis optimization based on DEGs in peripheral blood. It is worth noting that we detected 36 DEGs between the peripheral blood samples of healthy individuals and KBD patients, and a differential gene, GLRX5, was determined by combining LASSO and SVM-RFE methods. This might help the study of KBD's early stages, diagnosis, and pathogenesis. Using correlation analyses, we were able to predict the GLRX5 diagnostic value based on bioinformatics analyses results. Overall, these results showed that the GLRX5 gene can be used as a KBD diagnostic marker.

In mitochondria, GLRX5 functions as a transfer protein for ISC to the iron regulatory protein (IRP), metaarabinidase, and iron chelatase [27]. GLRX5 mutations can have significant impacts on downstream ISC biosynthesis and maturation [28]. Also, IRP's binding activity to iron response elements (IREs) can be activated by GLRX5 genetic inhibition [29], resulting in iron starvation response upregulation and enhanced intracellular free iron, the lipid for iron death and a prerequisite for peroxidation. Additionally, this can lead to mitochondrial dysfunction and negatively regulates cysteine endopeptidase activity (essential for apoptosis).

Our study has several advantages. First, A relatively small number of studies have investigated KBD diagnostic biomarkers. For a comprehensive KBD understanding and identification of its diagnostic biomarkers, PBMCs' expression profile in peripheral blood samples of healthy individuals and KBD patients can be useful. Second, WGCNA was an excellent tool for processing gene expression data, as allowed connectivity analyses between modules and clinical features. Finally, we used LASSO regression and SVM-RFE algorithms combined with DEGs intersections to screen the target diagnostic gene, which proved our screening effectiveness. However, the current research also has limitations. Biomarkers seemingly helpful for diagnosis have previously been validated on real patients, but this study was limited by the lack of measured genes.

Conclusions

Overall, we used WGCNA and DEG analysis to screen KBD main pathways. Then, we used LASSO regression and SVM-RFE algorithms combined with DEGs intersections to detect target diagnostic marker genes. Finally, peripheral blood GLRX5 was presented as a potential KBD diagnostic biomarker.

Declarations

Data Availability

The data that support the findings of this study are openly available in:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59446>

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this article.

Funding Statement

This work was supported by Postgraduate Research & Practice Innovation Program of Harbin Medical University(grant number YJSKYCX2019-39HYD)

Acknowledgments

We are grateful for the support and help of the teachers in the rheumatology and immunology laboratory of the first affiliated hospital of Harbin medical university.

References

1. Ning, Y., Wang, X., Wang, S. & Guo, X. Comparative analysis of signaling pathways in peripheral blood from patients with Kashin-Beck disease and osteoarthritis, *Experimental and therapeutic medicine*, 12 (2016)4077–4084
2. Wang, X., Ning, Y., Yang, L., Yu, F. & Guo, X. Zinc: the Other Suspected Environmental Factor in Kashin-Beck Disease in Addition to Selenium, *Biological trace element research*, 179 (2017)178–184
3. Ning, Y., Wang, X., Lammi, M. J. & Guo, X. Changes in the NF- κ B signaling pathway in juvenile and adult patients with Kashin-Beck disease. *Exp Cell Res*, **379**, 140–149 (2019).
4. Lei, R. *et al.* Prevalence of Selenium, T-2 Toxin, and Deoxynivalenol in Kashin-Beck Disease Areas in Qinghai Province, Northwest China, *Biological trace element research*, 171 (2016)34–40
5. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, **9**, 559 (2008).
6. Xu, M., Ouyang, T., Lv, K. & Ma, X. Integrated WGCNA and PPI Network to Screen Hub Genes Signatures for Infantile Hemangioma. *Frontiers in genetics*, **11**, 614195 (2020).

7. Chen, S. *et al.* Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *PeerJ*, **7**, e7873 (2019).
8. Chen, S. *et al.* Identification of crucial genes mediating abdominal aortic aneurysm pathogenesis based on gene expression profiling of perivascular adipose tissue by WGCNA. *Annals of translational medicine*, **9**, 52 (2021).
9. Zhu, N., Zhang, P., Du, L., Hou, J. & Xu, B. Identification of key genes and expression profiles in osteoarthritis by co-expressed network analysis. *Computational biology and chemistry*, **85**, 107225 (2020).
10. Liu, Z., Ma, C., Gu, J. & Yu, M. Potential biomarkers of acute myocardial infarction based on weighted gene co-expression network analysis. *Biomedical engineering online*, **18**, 9 (2019).
11. Giulietti, M. *et al.* Emerging Biomarkers in Bladder Cancer Identified by Network Analysis of Transcriptomic Data. *Frontiers in oncology*, **8**, 450 (2018).
12. Zhou, Y. Y. *et al.* Integrated transcriptomic analysis reveals hub genes involved in diagnosis and prognosis of pancreatic cancer, *Molecular medicine (Cambridge, Mass.)*, **25** (2019) 47
13. Ritchie, M. E. *et al.* Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, **43**, e47 (2015).
14. Villanueva, R. A. M. & Chen, Z. J. *ggplot2: Elegant graphics for data analysis* (Taylor & Francis, 2019).
15. Warnes, M. G. R., Bolker, B., Bonebakker, L. & Gentleman, R. W.J.V.R.p.t.f.p.d.Huber, Package 'gplots', (2016)
16. Pei, G. & Chen, L. W.J.M.i.e. Zhang, WGCNA application to proteomic and metabolomic data analysis, **585** (2017) 135-158
17. Balakrishnan, R., Harris, M. A., Huntley, R., Van Auken, K. & Cherry, J. M. A guide to best practices for Gene Ontology (GO) manual annotation, *Database: the journal of biological databases and curation*, **2013** (2013) bat054
18. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, **4**, 44–57 (2009).
19. Suykens, J. A. J.J.N.p.l. Vandewalle, Least squares support vector machine classifiers, **9** (1999)293–300
20. Friedman, J. & Hastie, T. R.J.J.o.s.s. Tibshirani, Regularization paths for generalized linear models via coordinate descent, **33** (2010)1
21. 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, *TheScientificWorldJournal*, **2014** (2014) 795624
22. Wang, S., Guo, X., Wu, X. M. & Lammi, M. J. Genome-wide gene expression analysis suggests an important role of suppressed immunity in pathogenesis of Kashin-Beck disease. *PLoS One*, **7**, e28439 (2012).

23. Ganau, L., Prisco, L., Ligarotti, G. K. I., Ambu, R. & Ganau, M. Understanding the Pathological Basis of Neurological Diseases Through Diagnostic Platforms Based on Innovations in Biomedical Engineering: New Concepts and Theranostics Perspectives, *Medicines* (Basel, Switzerland), 5 (2018)
24. Tibshirani, R. Regression shrinkage and selection via the lasso, 58 (1996) 267-288
25. Estaquier, J., Vallette, F., Vayssiere, J. L. & Mignotte, B. The mitochondrial pathways of apoptosis. *Adv Exp Med Biol*, **942**, 157–183 (2012).
26. Liu, J. *et al.* The role of mitochondria in T-2 toxin-induced human chondrocytes apoptosis. *PLoS One*, **9**, e108394 (2014).
27. Ciofi-Baffoni, S., Nasta, V. & Banci, L. Protein networks in the maturation of human iron-sulfur proteins, *Metallomics: integrated biometal science*, 10 (2018)49–72
28. Liu, G. *et al.* Functional Analysis of GLRX5 Mutants Reveals Distinct Functionalities of GLRX5 Protein. *Journal of cellular biochemistry*, **117**, 207–217 (2016).
29. Ye, H. *et al.* Glutaredoxin 5 deficiency causes sideroblastic anemia by specifically impairing heme biosynthesis and depleting cytosolic iron in human erythroblasts. *J Clin Invest*, **120**, 1749–1761 (2010).

Figures

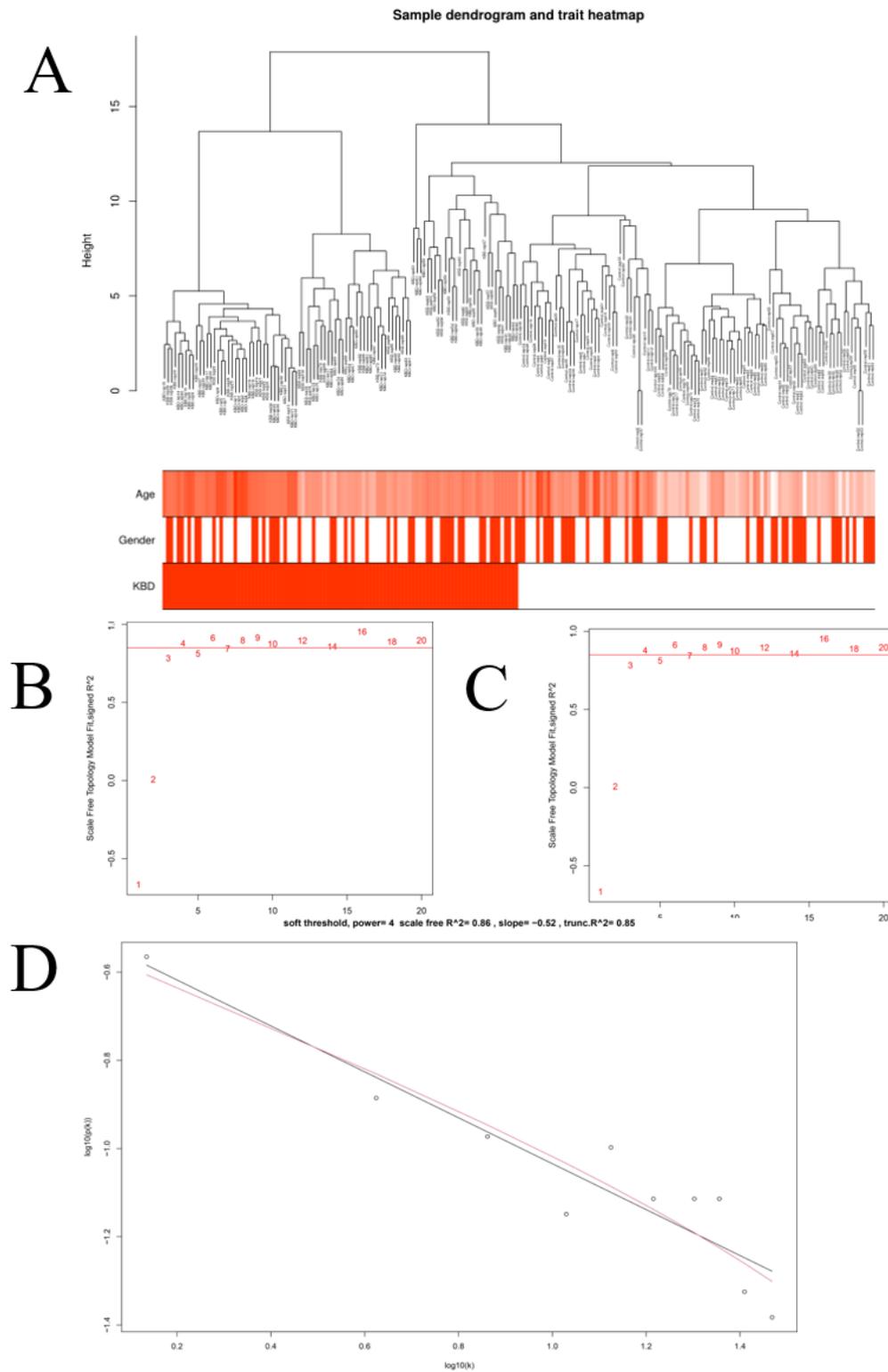


Figure 1

Samples clustering and soft threshold screening. (A) Samples were clustered to detect outliers. The feature heatmap shows the sample's characteristics. (B) Scale-free fitting index analysis of different soft threshold powers (β). (C) Average connectivity analysis of different soft threshold powers. (D) When $\beta = 4$, a scale-free topology was verified. WGCNA: Weighted Gene Co-expression Network Analysis.

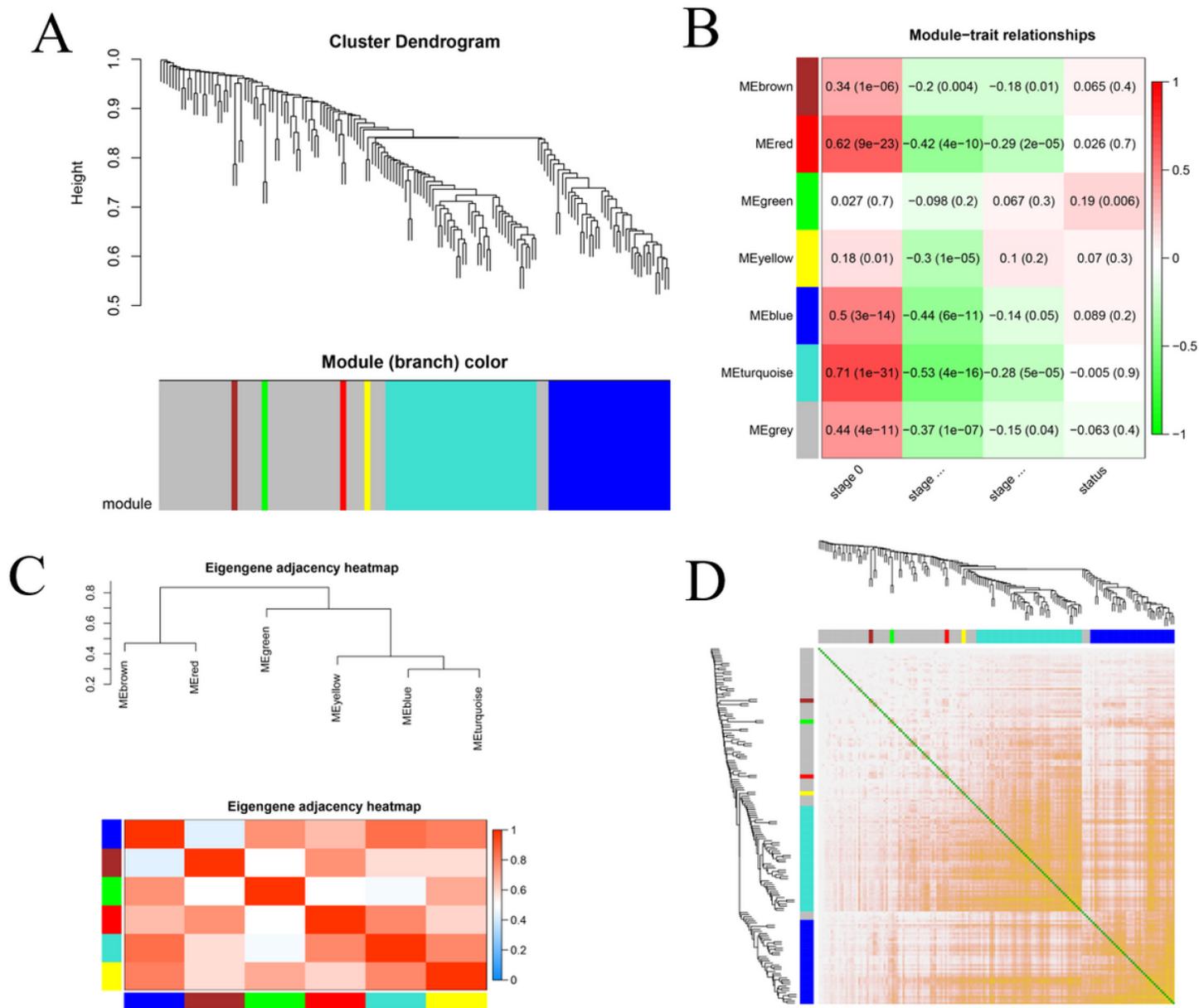


Figure 2

WGCNA sample. (A) Genes' cluster tree diagram. (B) Relationship between modules and features. Each pixel is composed of a corresponding correlation and p-value, which are color-coded by correlation according to the color legend. (C) Combination of characteristic gene dendrogram and heatmap. (D) Heatmap visualization of all genes in the network to describe the TOM between the genes under analysis. WGCNA: Weighted Gene Co-expression Network Analysis; TOM: Topological Overlap Matrix.

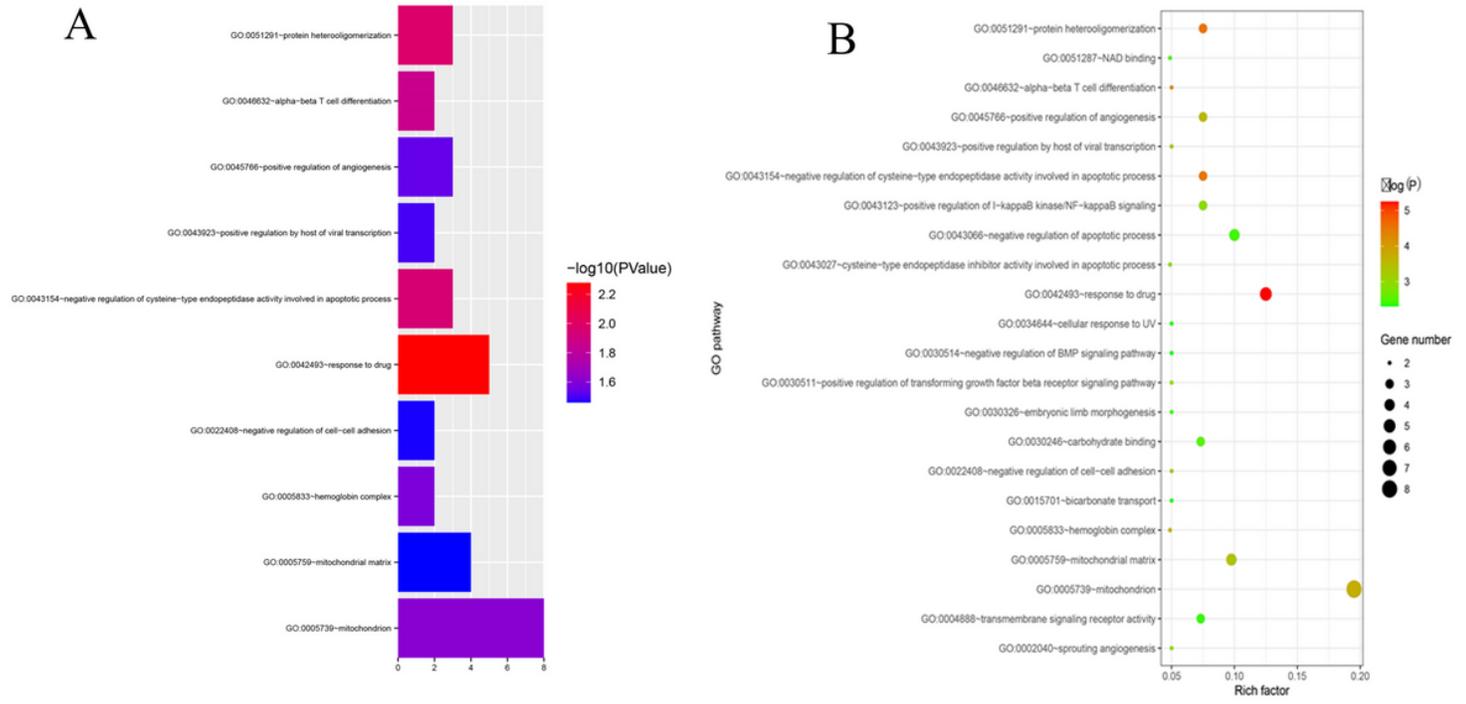


Figure 3

KBD-related genes functional enrichment analysis. (A) GO enrichment analysis bar chart. (B) KBD-related DEGs pathway enrichment analysis dot chart.

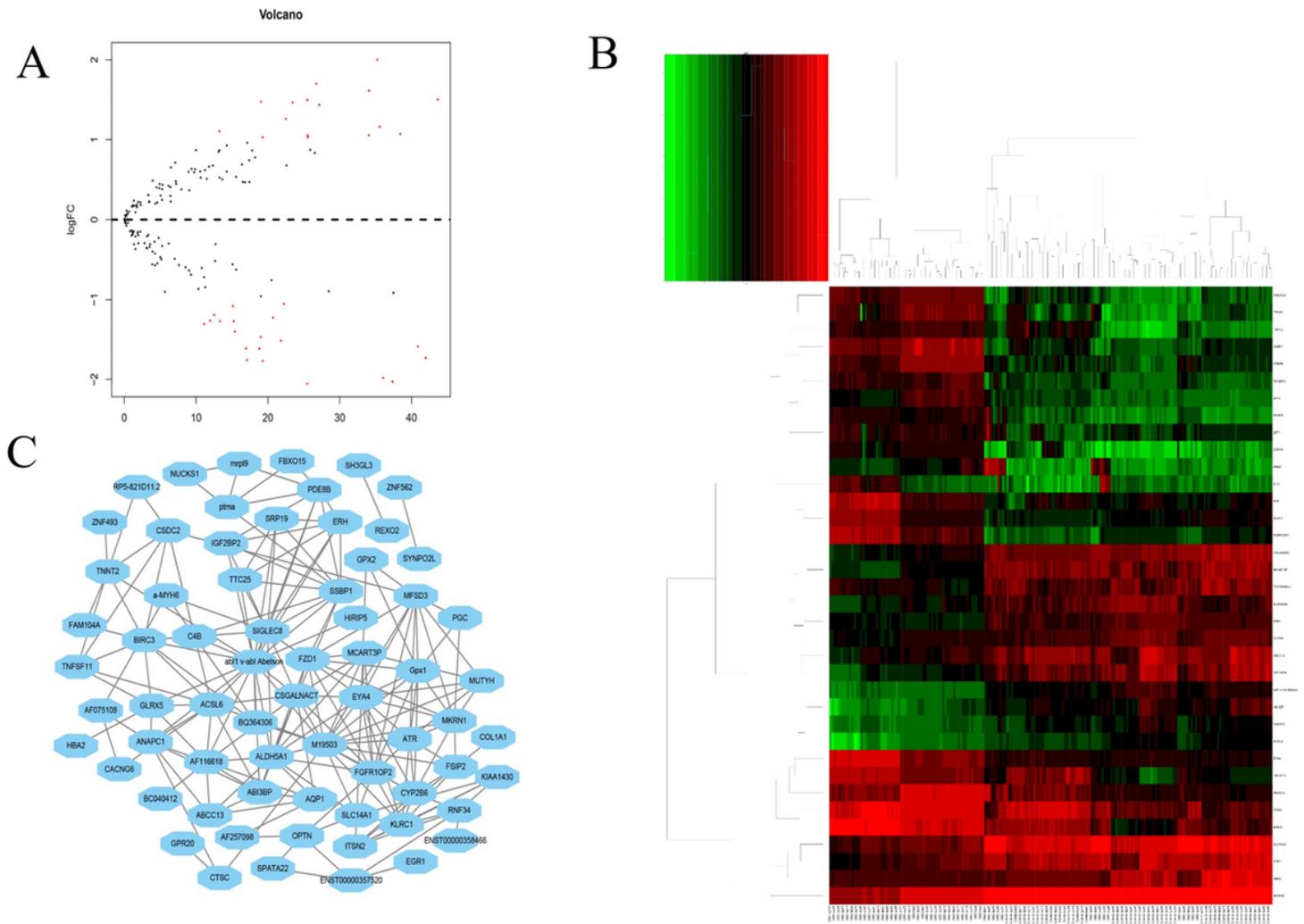


Figure 4

DEG analysis. (A) Volcano amp and (B) heatmap of DEGs between healthy individuals and KBD patients. An adjusted p-value < 0.05 and $|\log_2(\text{fold change})| > 1.5$ were used as cut-off thresholds. (C) DEG network analysis. DEGs: differentially expressed genes. KBD: Kashin-Beck disease.

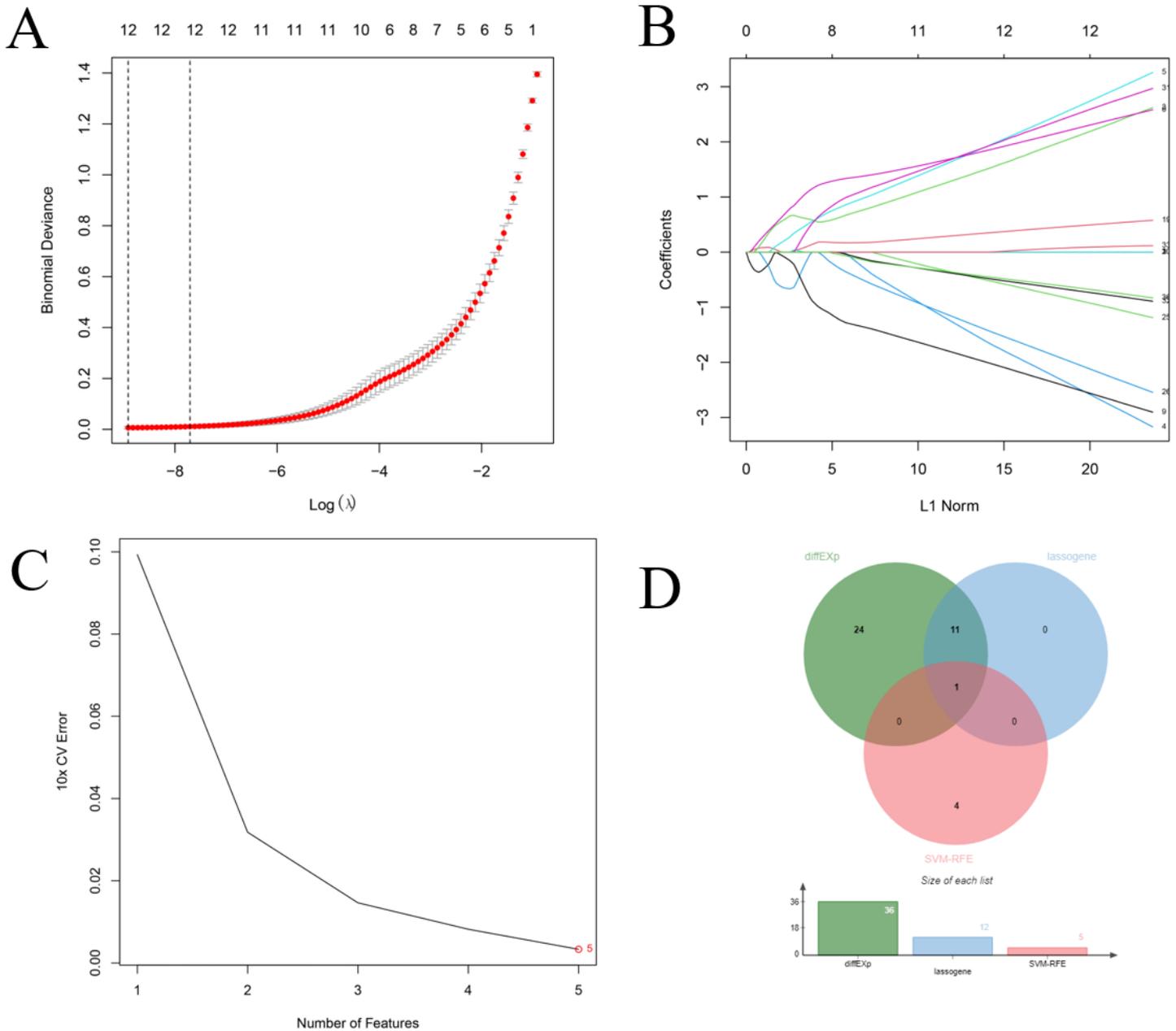


Figure 5

Diagnostic markers screening. (A) Screening of the smallest absolute contraction of absolute markers. (B) Selection operator (LASSO) logistic regression algorithm results. Each color represents a different gene. (C) Support vector machine recursive feature elimination (SVM-RFE) algorithm results. (D) The Venn diagram shows the intersection of the diagnostic markers derived from the three algorithms.

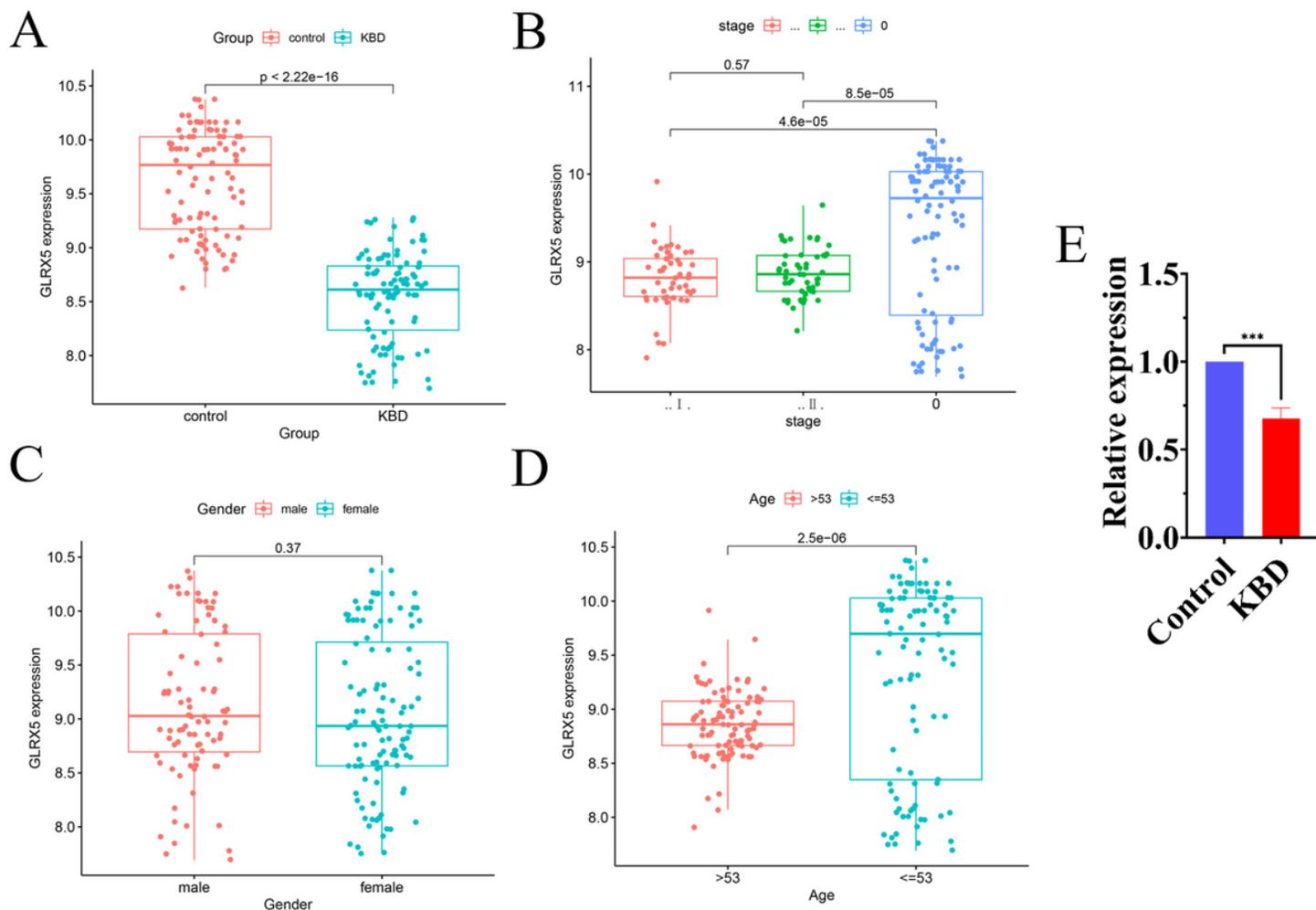


Figure 6

Key gene analysis and verification. (A) GLRX5 expression level in the KBD and Control groups in the GSE59446 dataset. (B) Relationship between GLRX5 expression in the GSE59446 dataset and the KBD stage. (C) Relationship between GLRX5 expression and gender in the GSE59446 data set. (D) Relationship between GLRX5 expression and age in the GSE59446 dataset. (E) Relative GLRX5 expression level in patients' blood PBMCs collected by us. A $p < 0.05$ was considered statistically significant.

Supplementary Files

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

- [TableS1.pdf](#)
- [TableS2.pdf](#)
- [TableS3.pdf](#)