Identification of cuproptosis-related genes related to the progression of ankylosing spondylitis by integrated bioinformatics analysis

Junyi Fan, MM[a], Qihua Liu, DR[b], Ting Chen, MM[c], Yongbin Chen, MM[b], Junzhe Wu, MM[d,*]

Abstract

Ankylosing spondylitis (AS) is an autoimmune disease, and the relationship between copper death and AS is not clear. The aim of this study was to analyze and identify potential cuprosis-related genes associated with the onset of AS by bioinformatics methods. We obtained the AS gene expression profile GSE25101 from the Gene Expression Omnibus (GEO) database, which consists of blood samples from 16 active AS patients and 16 sex-and age-matched controls. After analyzing the data, we utilized the WGCNA method to identify genes that exhibited significant differential expression. In order to assess the prognostic and predictive power of these genes, we constructed receiver operating characteristic (ROC) curves. To further validate our predictions, we employed nomograms, calibration curves, decision curve analysis, and external datasets. Lastly, we conducted an analysis on immune infiltration and explored the correlation between key genes and immune response. Three genes, namely INPP5E, CYB5R1, and HGD, have been identified through analysis to be associated with AS. The diagnosis of patients using these genes has been found to possess a high level of accuracy. The area under the ROC curve is reported to be 0.816 for INPP5E, 0.879 for CYB5R1, and also 0.879 for HGD. Furthermore, the nomogram demonstrates an excellent predictive power, and it has been calibrated using a Calibration curve. Its clinical usefulness and net benefit have been thoroughly analyzed and estimated through the use of a DCA curve. Moreover, INPP5E, CYB5R1, and HGD are found to be associated with various types of immune cells. In conclusion, the systematic analysis of cuprosis-related genes may aid in the identification of mechanisms related to copper-induced cell death in AS and offer valuable biomarkers for the diagnosis and treatment of AS.

Abbreviations: AS = ankylosing spondylitis, AUC = area under curve, CRGs = cuprosis-related genes, DCA = decision curve analysis, DEGs = differentially expressed genes, FLS = fibroblast-like synoviocytes, GEO = Gene Expression Omnibus, GSEA = Gene Set Enrichment Analysis, HGD = homogentisic acid 1,2-dioxygenase, INPP5E = inositol polyphosphate-5-phosphatase, LASSO = Least Absolute Shrinkage and Selection Operator, ME = module eigengene, PPI = protein-protein interaction, RA = rheumatoid arthritis, ROC = receiver operating characteristic, TCA = tricarboxylic acid, Th 17/23 = T helper 17/23, URGs = up-regulated genes, WGCNA = weighted gene co-expression network analysis.

Keywords: ankylosing spondylitis, bioinformatics, cuproptosis, GEO datasets, inflammations

1. Introduction

Ankylosing spondylitis (AS) is a long-term autoimmune inflammatory condition that primarily affects the spinal and sacroiliac joints. This disease can lead to fibrosis of the intervertebral discs, calcification of nearby connective tissues, and the formation of bony ankylosis. As a result, individuals with AS may experience chronic back pain, restricted spinal movement, and a significant decline in their overall quality of life.[1]

The global prevalence of AS is reportedly between 0.1% to 1.4%, with a male to female ratio of approximately 2 to 4:1. AS typically manifests before the age of 40. The specific underlying causes of AS remain unknown, although genetic, environmental, immunological, and metabolic factors are believed to contribute significantly to its onset and progression. Diagnosis of AS primarily relies on the recognition of characteristic clinical symptoms and signs (such as post-activity improved inflammatory

* Correspondence: Junzhe Wu, Orthopaedics Dept. 1 Hospital of Traditional Chinese Medicine of Zhongshan, Zhongshan 528400, China (e-mail: 2640625235@qq.com).

lumbago and backache, sacroiliac pain), imaging evaluations of the pelvis and spine, and serological tests. Notably, Human Leukocyte Antigen B27 is closely associated with AS susceptibility and exhibits a marked familial clustering pattern, accounting for approximately 30% of the overall genetic risk of AS.[2,3] Recent reports suggest a correlation between interleukin-23 receptor and T helper (Th) 17/23 genes with the genetic susceptibility to AS.[4] The identification of AS-related pathway genes holds promise for enhancing the diagnosis and treatment of AS. Unfortunately, no definitive cure for AS exists at present. Current management strategies primarily comprise non-drug interventions, such as optimizing posture and engaging in regular physical exercise, as well as pharmacological interventions involving the use of nonsteroidal anti-inflammatory drugs and sulfasalazine to alleviate pain and improve inflammatory biomarkers.[5,6] Currently, the most effective drugs for blocking tumor necrosis factor α, such as etanercept, infliximab, and adalimumab, are biological agents. It is recommended to use these drugs early in the disease progression, specifically when the inflammatory burden is at its peak. However, the efficacy of these drugs remains a matter of debate.[7] Consequently, future research on AS should prioritize investigating its underlying mechanisms and the development of drugs targeting these mechanisms.

In 2022, Tsvetkov et al published a groundbreaking study in Science, unveiling a novel mechanism of cell death called cuproptosis. This study revealed that an excess of copper ions leads to the direct binding of these ions to lipoylated dihydrolipoamide S-acetyltransferase lipoylproteins within the tricarboxylic acid (TCA) cycle. Consequently, the aberrant aggregation of lipoylproteins occurs, impairing the normal functioning of the TCA cycle and triggering a proteotoxic stress response. Ultimately, this cascade of events culminates in the demise of the affected cells.[8] Several recent studies have highlighted the association between cuproptosis and various multi-system diseases, including neurodegenerative diseases such as Alzheimer's disease, cancer, cardiovascular diseases like atherosclerosis, and metabolic diseases.[9–12] However, there has been a lack of bioinformatic analysis on cuproptosis-related genes (CRGs) and their role in the progression of AS. A study as early as 1978 by P. Aiginger et al found that patients with AS and RA (rheumatoid arthritis) had significantly higher serum copper levels compared to normal patients.[13] Our aim was to build upon the aforementioned findings by identifying CRGs and elucidating their potential associations with AS. This will serve as a foundational basis for further exploration and innovation in clinical treatment regimens for AS. One notable manifestation of AS is erosive synovitis,[14] which refers to the excessive and detrimental proliferation of fibroblast-like synoviocytes (FLS). FLS possess several tumor-like characteristics and can thrive and proliferate excessively within a microenvironment resembling a tumor. The abnormal proliferation of FLS is partially attributed to the inhibition of programmed cell death, and AS is characterized by the imbalance between cell survival and cell death in multiple cell types.[15] The metabolic mechanism potentially related to copper poisoning may be associated with these specific cell populations.

Bioinformatics plays a crucial role in unraveling the molecular mechanisms underlying diseases. It is an interdisciplinary field that combines biology and computer science. In recent times, bioinformatics has witnessed significant advancements, leading to the development of numerous public databases. These databases serve as valuable resources for researchers, aiding them in identifying disease diagnoses, prognoses, and potential treatment targets. In certain studies, researchers have employed gene expression datasets downloaded from the Gene Expression Omnibus (GEO) to screen for hub genes. This approach helps elucidate the onset and progression mechanisms of AS.[16,17] However, no study has been conducted to apply bioinformatics for the analysis of the correlation between CRGs and the progression of AS. We conducted an analysis of the whole blood RNA data from AS patients in the GEO database. Our aim was to construct a prediction model that can identify the risk of copper death. Through this analysis, we identified genes that are associated with copper death and its onset in AS. Our findings highlight 3 specific genes that are highly correlated with the onset of AS and prognosis of copper death. These genes may serve as potential biomarkers for accurately predicting the onset of AS.

This study offers a theoretical basis and novel insights into the investigation of copper demise in AS, providing a fresh perspective for genetic diagnosis and treatment of AS.

2. Materials and methods

The information for this study was exclusively sourced from public databases, eliminating the need for ethical approval.

2.1. Data collection

We downloaded gene expression profiles GSE25101, GSE73754, and GSE18781 of AS from the GEO database (http://www.ncbi.nlm.nih.gov/geo). GSE25101 analyzed blood samples from 16 AS patients in the active stage and 16 patients in the control group that were matched for gender and age. GSE73754 analyzed blood samples from 94 patients (53 males and 41 females) and 70 healthy control subjects (30 males and 40 females) using a gender-specific analysis. GSE18781 analyzed peripheral blood samples from 2 independent groups, one with 11 AS patients and 12 healthy control subjects, and another with 7 AS patients and 13 healthy control subjects.

2.2. Analysis of differentially expressed genes

The data in GSE25101 underwent standardization and normalization using the normalized inter-array function in the "limma" package of the R software. The "limma" R software package was utilized to identify differentially expressed genes (DEGs) between the AS group and the control group. The screening criteria for DEGs were a logFC > 1 and an adj.PVal < .05. A heatmap was generated to display the top 30 genes with the most significant differences based on the logFC value. To visually represent the up-regulated genes (URGs) and down-regulated genes, a volcano plot and heatmap were created using the ggplot2 package and the "pheatmap" package of the R software. Lastly, the selected DEGs underwent Gene Set Enrichment Analysis (GSEA).

2.3. Co-expression network analysis

The Weighted Gene Co-expression Network Analysis (WGCNA) algorithm is used to cluster genes into distinct modules, uncovering the relationships between these modules and disease characteristics.[18] This co-expression network analysis allows for a deeper understanding of the connections between genes and diseases. To comprehensively investigate the pathogenesis of AS, we utilized the "WGCNA" package in the R software to construct a co-expression network for DEGs. Only genes with a maximum variance of 25% in the GSE25101 dataset were included in the construction of the co-expression network. The modules with a threshold of 0.25 were combined using the Dynamic Tree Cut method. Data integrity was verified using the "goodSampleGenes" function. The optimal soft threshold (b) was selected and validated with the "PickSoftThreshold" function. To identify modules based on topological overlap, the matrix data were transformed into an adjacency matrix for clustering. Subsequently, the module eigengene (ME) was calculated, and relevant modules in the ME-based tree were merged, resulting in the generation of a cluster dendrogram. By assessing the importance of genes and clinical data based on phenotypic information and modules, we examined the relationship between the model and modules. The power exponent was set at 1:20 for modular clustering genes, with 40 modules serving as functional identification points.

2.4. Enrichment analysis

To identify overlapping genes, we performed an intersection analysis between the differentially expressed genes (DEGs) in the AS group and the control group, and the hub module genes obtained from the WGCNA. The resulting genes represent the DEGs among the hub module genes. To further explore the biological functions and pathways associated with these genes, we conducted GSEA using the "clusterProfiler" package. This analysis included Disease Ontology and Kyoto Encyclopedia of Genes and Genomes analysis. Additionally, we constructed a protein–protein interaction (PPI) network for these genes to investigate their potential interactions and functional relationships.

2.5. Validation of hub genes using the LASSO logistic analysis

We identified the intersected genes by performing an intersection analysis of the differentially expressed genes (DEGs) from the training set and the hub genes from the WGCNA. To further screen out hub feature genes, we employed the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm. Logistic regression analysis was conducted using the "glmnet" package from the R software. The optimal value was determined by defining the minimum lambda. LASSO, as a regression-based method, allows the inclusion of numerous covariates in a model, while also shrinking the absolute value of the regression coefficient.

2.6. Identification of cuproptosis-related differentially expressed genes

To model the disease, we intersected the disease feature genes obtained from LASSO with the CRGs retrieved from the literature.[8] Furthermore, we performed a GSEA on these genes to determine the relevant signaling pathways and biological processes associated with them.

2.7. Evaluation of disease prediction efficiency of hub genes

I constructed a nomogram model to predict the onset of AS based on disease-related hub CRGs. To accomplish this, I utilized the "rms package" from the R software. The model employed Logistic regression to identify the prediction model associated with AS onset and obtain the nomoscore formula. Each gene's score was represented by "points" in the nomogram model, while the "Total points" indicated the cumulative score of all the aforementioned genes. To evaluate the calibration of the nomogram, a calibration curve was generated. Additionally, a decision curve analysis (DCA) curve was plotted to analyze and estimate the clinical usability and net benefits. Subsequently, a receiver operating characteristic (ROC) curve was constructed using the "pROC" package of the R software, followed by calculating the area under curve (AUC) in RStudio to determine the diagnostic value of the screened hub disease feature genes and nomogram model. Genes with an AUC greater than 0.7 and a P value less than .05 were deemed to possess high diagnostic accuracy for AS.

2.8. Immune cells infiltration analysis

The correlation between AS and Hallmark pathway/immune infiltration was analyzed using the ssGSEA method in the "GSVA" package of the R software. Spearman Correlation Analysis was employed to assess the correlation between immune cells, immune functions, and the expressions of hub genes.

2.9. Analytical statistics

All statistical analyses were conducted using the R package of "R v4.2.2" framework. To examine differences in continuous variables that followed a normal distribution, the Student t test was employed. On the other hand, the Wilcoxon rank sum test was utilized for continuous variables that did not conform to a normal distribution. Additionally, ROC curves were employed to assess the diagnostic accuracy of genes. A significance level of $P < .05$ was adopted for all tests to determine statistical significance.

3. Results

3.1. Screening of DEGs

Figure 1 illustrates the flow chart of this study. To investigate genes associated with the onset of AS, we performed a retrospective analysis on blood samples from 16 AS patients at the active stage using data from the GSE25101 dataset. We also included 16 patients from the control group who were matched in terms of gender and age. Raw data from these samples were homogenized (Fig. 2A). A total of 12,923 DEGs were identified. Applying the criterion adj.$Pval < 0.05$, we obtained 344 DEGs, consisting of 147 URGs and 197 down-regulated genes. These findings are visually represented through the heatmap (Fig. 2B) and the volcano plot (Fig. 2C).

3.2. Functional enrichment analysis of DEGs

To gain a deeper understanding of the biological processes and signaling pathways of AS-related differentially expressed genes (DEGs), we conducted an enrichment analysis on the DEGs between the AS group and the control group. The results of the study revealed that the DEGs were up-regulated in several pathways, including Cardiac muscle contraction, Chemical carcinogenesis – DNA adducts, Oxidative phosphorylation, Renin-angiotensin system, and Ribosome, as depicted in Figure 2D. On the other hand, the DEGs were found to be down-regulated in pathways such as 2-Oxocarboxylic acid metabolism, Adherens junction, Citrate cycle (TCA cycle), GnRH secretion, and Primary immunodeficiency, as shown in Figure 2E. In order to visualize the gene set enrichment (GSE) functions of the DEGs, a ridgeline plot was generated. Each color curve on the plot represented a distinct function or pathway. The peak value of the upward or downward curve indicated the level of positive or negative regulation of each DEG, which is referred to as the enrichment score. Notably, the results indicated that these DEGs were down-regulated when enriched in pathways associated with Thermogenesis, non-alcoholic fatty liver disease, and Chemical carcinogenesis - DNA adducts, as clearly demonstrated in Figure 2F.

3.3. Co-expression network construction and module detection

The analysis of WGCNA was performed to examine the mRNA expressions of 344 DEGs. Subsequently, cluster analysis and comparison of grouping differences were conducted on the samples. As a result, a cluster dendrogram and heatmap of the samples were generated (Fig. 3A). To determine the appropriate soft threshold, a value of 10 was selected and applied to obtain the approximate scale-free topological fitting index of the network (Fig. 3B). Utilizing a height of 0.25 as the cutting line, another cluster dendrogram was produced (Fig. 3C), with the modules above the cutting line identified for further study. By setting minModuleSize = 30 and mergeCutH8 = 0.25 as the clustering criteria, a total of 29 gene modules were discerned. To describe the paired relationship among genes, the topological overlap matrix was employed (Fig. 3D). Each row and column in the topological overlap matrix corresponded to one gene, with lighter colors indicating a lower
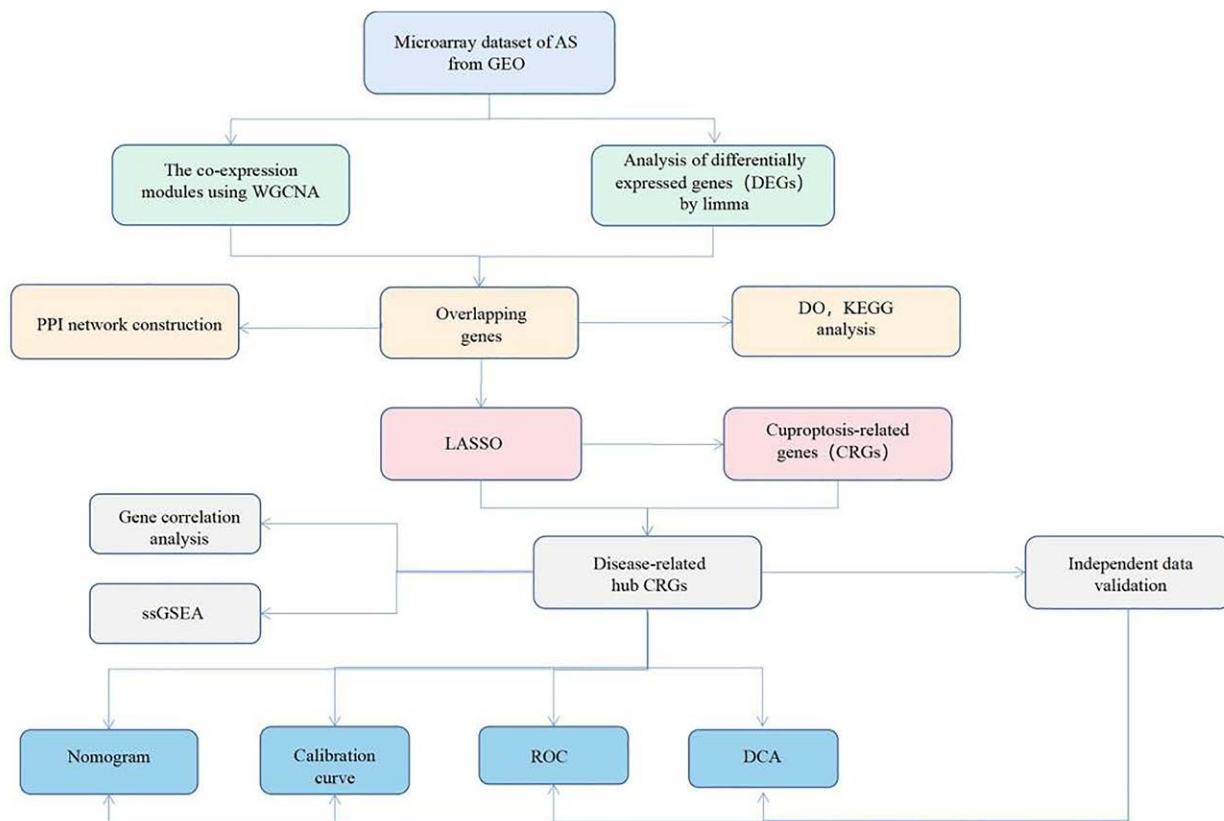
Figure 1. Study flow chart of present work. CRGs = cuproptosis-related genes, DCA = decision curve analysis, DEGs = differentially expressed genes, DO = disease ontology, GSEA = Gene Set Enrichment Analysis, KEGG = Kyoto Encyclopedia of Genes and Genomes, LASSO = least absolute shrinkage and selection operator, PPI = protein-protein interaction, ROC = receiver operating characteristic curve, ssGSEA = single sample gene set enrichment analysis, WGCNA = weighted gene co-expression network analysis.

degree of topological overlap and darker colors indicating a higher degree of topological overlap. Subsequently, the co-expression similarity and adjacency of genes between the AS group and control group were analyzed using the genes within the 29 color modules. The outcomes from the ME collinear heatmap illustrated that 6 gene modules, namely, MEblue, MElightcyan1, MEmidnightblue, MEbrown4, MEviolet, and MEdarkred, exhibited significant associations with the onset of AS (Fig. 3E). These 6 selected modules were subjected to subsequent analysis. The scatter plot (Fig. 3F–K) demonstrated the significant associations between the 6 gene modules of different colors and the onset of AS.

3.4. DEGs in hub module genes and functional enrichment analysis

By intersecting the 3325 hub module genes in the 6 hub gene modules obtained from WGCNA with the 344 previously screened DEGs, a total of 310 hub module DEGs were identified, as depicted in Figure 4A of the Venn diagram. The results of the differential expression (DE) analysis revealed their enrichment in ischemia, primary immunodeficiency disease, and neuropathy, as depicted in Figure 4B. Furthermore, the Kyoto Encyclopedia of Genes and Genomes analysis demonstrated their enrichment in Parkinson's disease, Cardiac muscle contraction, and 2-Oxocarboxylic acid metabolism, as shown in Figure 4C and D.

3.5. Protein-protein interaction networks

The gene list mentioned above was imported into the STRING database to retrieve the PPI Networks. Subsequently, the obtained PPI network was imported into the Cytoscape software to utilize the cytoHubba plugin for calculating the top 20 genes based on the MCC algorithm. The resulting PPI network is depicted in Figure 5.

3.6. Screening of disease feature genes

Disease feature genes were identified through LASSO regression analysis based on 310 hub module DEGs (Fig. 6A and B). The GSE25101 dataset was used for screening disease feature genes using the LASSO regression model, as shown in Figure 6A. The changes in coefficient λ of different genes are displayed in Figure 6B. By intersecting the 15 disease feature genes obtained from LASSO analysis and the 2978 CRGs obtained from references, 3 disease-related hub CRGs were finally identified, as illustrated in the Venn diagram (Fig. 6C). Enrichment analysis results for the 3 genes revealed that CYB5R1 was up-regulated in pathways such as Asthma, Chemical carcinogenesis-DNA adducts, Glycosphingolipid biosynthesis – lacto and neolacto series, Renin-angiotensin system, and Ribosome. In contrast, CYB5R1 was down-regulated in pathways including 12-Oxocarboxylic acid metabolism, Adherens junction, Alanine, aspartate and glutamate metabolism, DNA replication, and Primary immunodeficiency, as depicted in Figure 6D and E.

Homogentisic acid 1,2-dioxygenase (HGD) exhibited up-regulation in various pathways, including Complement and coagulation cascades, Fat digestion and absorption, Glycosphingolipid biosynthesis – lacto and neolacto series, Renin-angiotensin system, and Staphylococcus aureus infection. Conversely, HGD showed down-regulation in pathways such as Adherens junction, Citrate cycle (TCA cycle), Graft-versus-host disease, Primary immunodeficiency, and Viral myocarditis, as indicated in (Fig. 6F and G).
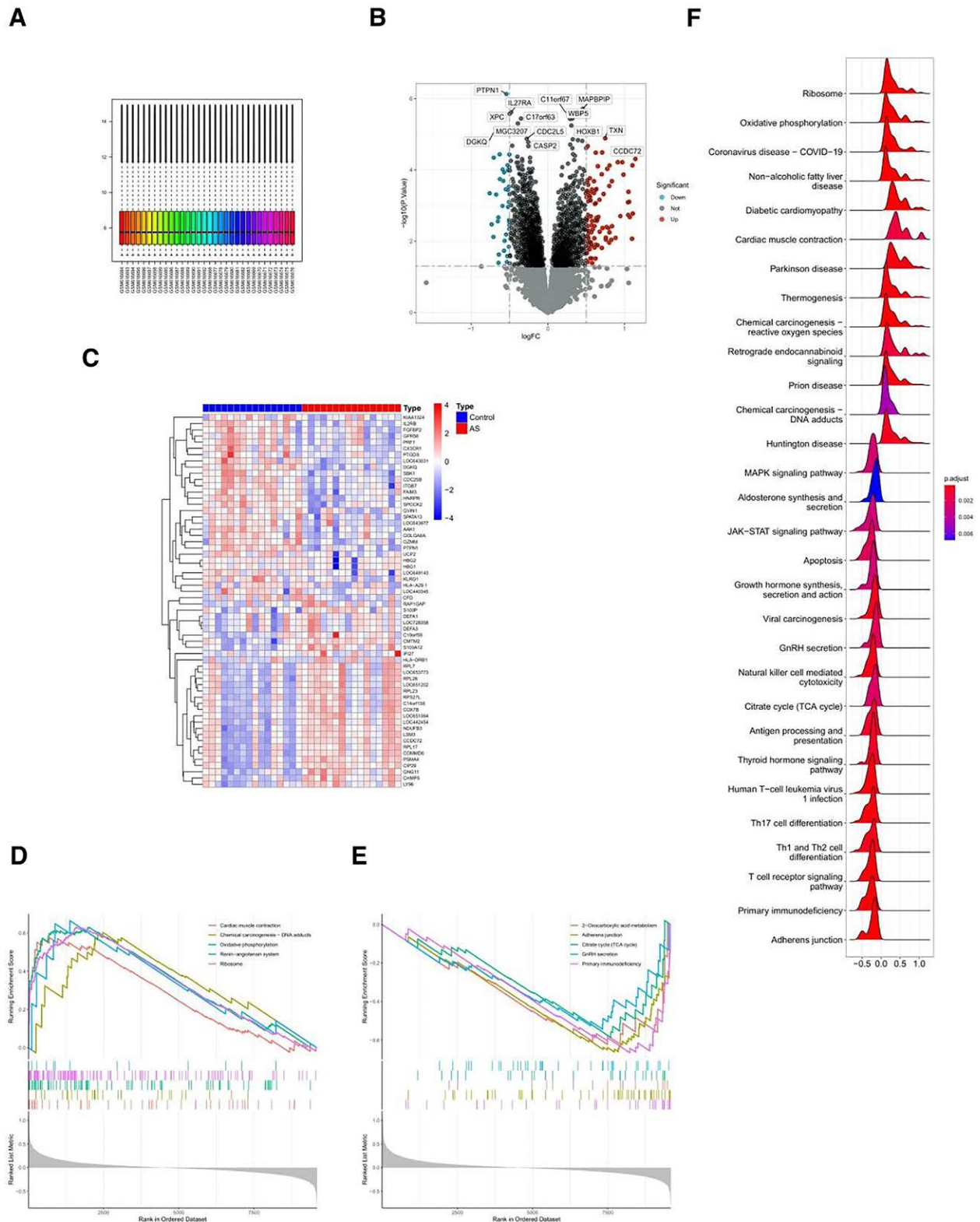
Figure 2. Screening and enrichment analysis of DEGs. The homogenization of raw data of samples was shown in (A) Box Plot. The expression patterns of DEGs were shown in (B) Heatmap and (C) Volcano Plot. Enrichment analysis showed the results of pathways where DEGs were up-regulated (D) and down-regulated (E). The enrichment of DEGs in pathways and diseases was shown in (F) Ridgeline Plot. AS = ankylosing spondylitis, DEGs = differentially expressed genes.

Inositol polyphosphate-5-phosphatase (INPP5E) exhibited up-regulation in several pathways, including Homologous recombination, Nucleocytoplasmic transport, Primary immuno-deficiency, Ribosome biogenesis in eukaryotes, and Th1 and Th2 cell differentiation. Conversely, it demonstrated down-regulation in pathways such as Ferroptosis, Mineral absorption, Porphyrin metabolism, Proteasome, and Renin-angiotensin system, as depicted in Figure 6H and I.

We analyzed the correlation between the 3 core genes. The correlation analysis in the verification set revealed that HGD
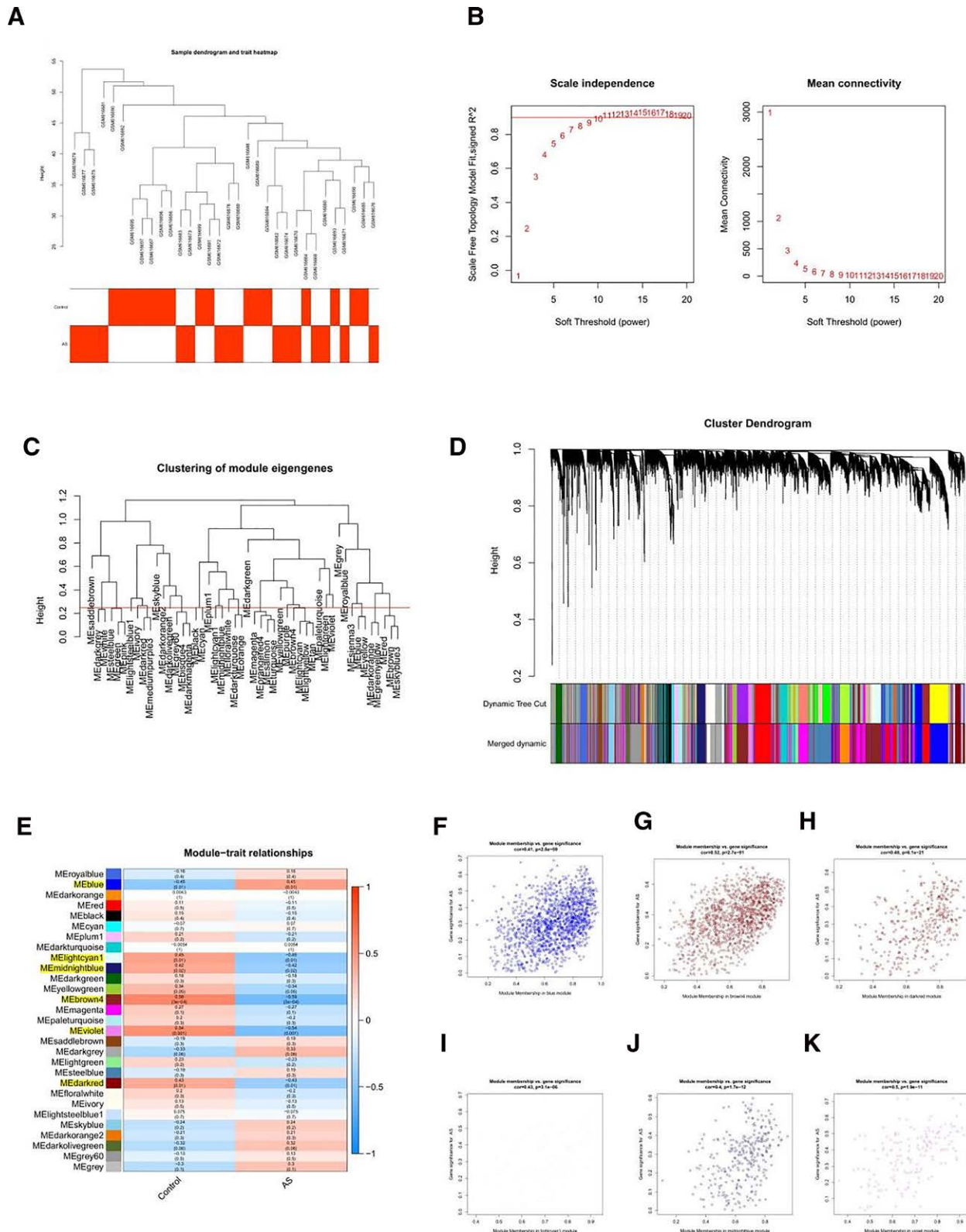
Figure 3. Co-expression network of DEGs in AS group. (A) Dendrogram presented by sample clustering. (B) Selection of soft threshold and scale-free topological fitting index. (C) Dendrogram cut into clusters at the height of 0.25 to present selected MEs. (D) Cluster dendrogram of co-expression modules to present raw data modules and combined modules, with different colors representing different co-expression modules. (E) Collinear heatmap for the screening of MEs. Scatter diagram between the membership of blue (F), light cyan 1 (G), midnight blue (H), brown 4 (I), violet (J), and dark red (K): scatter plots to present the module membership and gene significance of the AS group. AS = ankylosing spondylitis, DEGs = differentially expressed genes, ME = module eigengene.
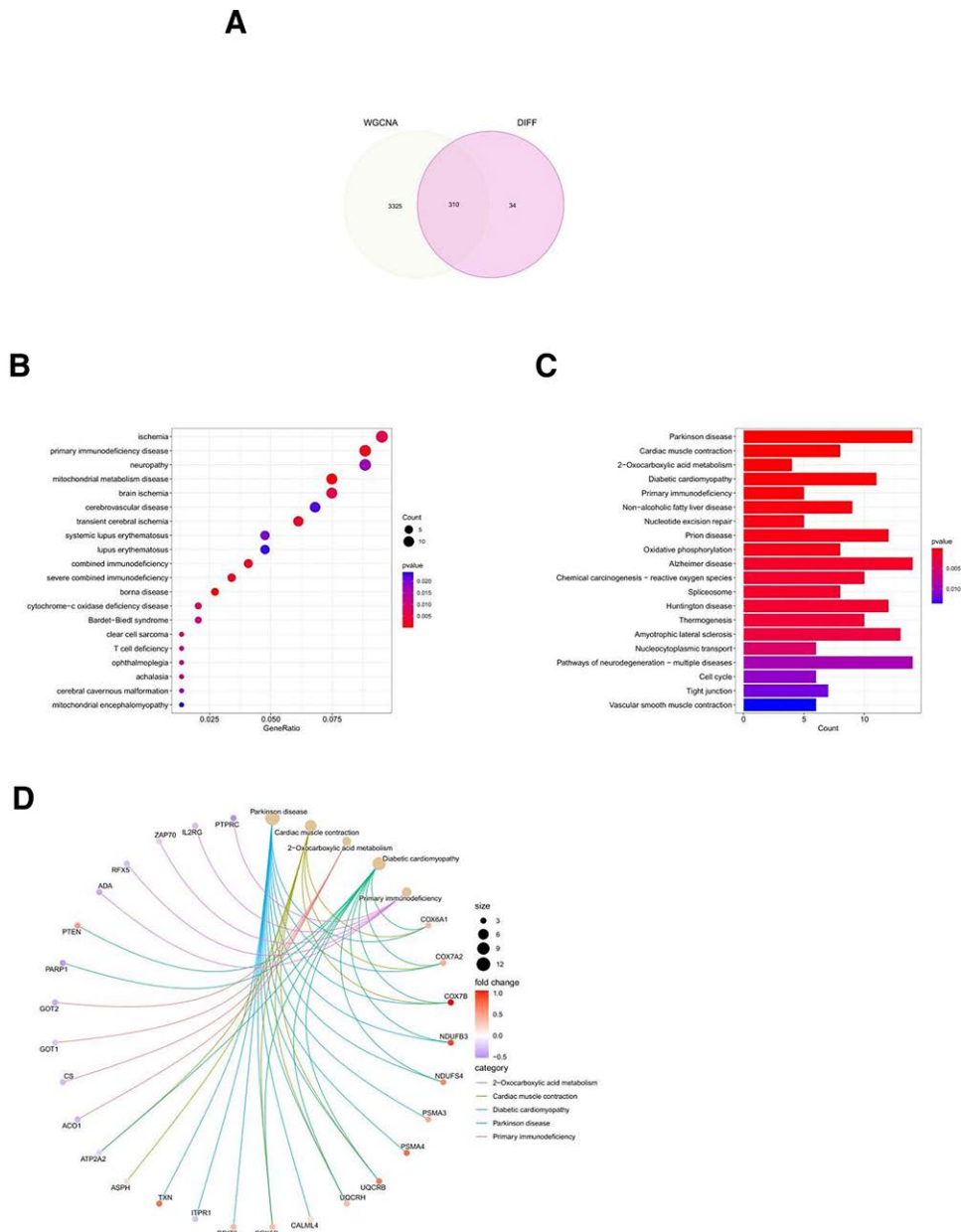
Figure 4. DEGs in Hub Module Genes and Functional Enrichment Analysis. (A) Venn diagram, presenting DEGs of 310 hub module genes obtained from taking intersections between 3325 hub module genes and 344 DEGs. (B) DO term analysis of DEGs. (C and D) KEGG term analysis of DEGs. DEGs = differentially expressed genes, KEGG = Kyoto Encyclopedia of Genes and Genomes.

exhibited a negative correlation with INPP5E, and a positive correlation with CYB5R1. Additionally, INPP5E showed a negative correlation with CYB5R1. These findings suggest that these hub disease feature genes share similar functions.

3.7. Expressions of disease-related hub CRGs in training set and verification set

In the AS samples of the training set (GSE25101), we observed a significant up-regulation of CYB5R1 and HGD, whereas INPP5E showed a significant down-regulation (Fig. 7A–C) compared to normal samples. To further validate the expression patterns of these 3 genes in AS subjects, we conducted additional analyses on the verification sets GSE73754 and GSE18781. In the verification set GSE73754, there was no significant difference in the expression levels of the aforementioned genes between normal and AS samples (Fig. 7D–F). However, in the verification set

GSE18781, INPP5E exhibited a significant down-regulation, no significant difference was observed in the expression of HGD, and CYB5R1 was significantly down-regulated in the AS samples in comparison to the normal samples. Interestingly, these findings contradicted the results obtained from the training set (Fig. 7G–I).

3.8. Construction of nomogram model

A nomogram was created using the training set GSE25101, incorporating the 3 disease-related hub CRGs through logistic regression analysis. Figure 8A illustrates the prediction model diagram for the onset of AS based on the 3 CRGs. The independent factors in the model are INPP5E, CYB5R1, and HGD. By referring to the expression levels of these genes, one can calculate the nomoscore and obtain the total score, which represents the cumulative score of relevant predictive factors. For instance,
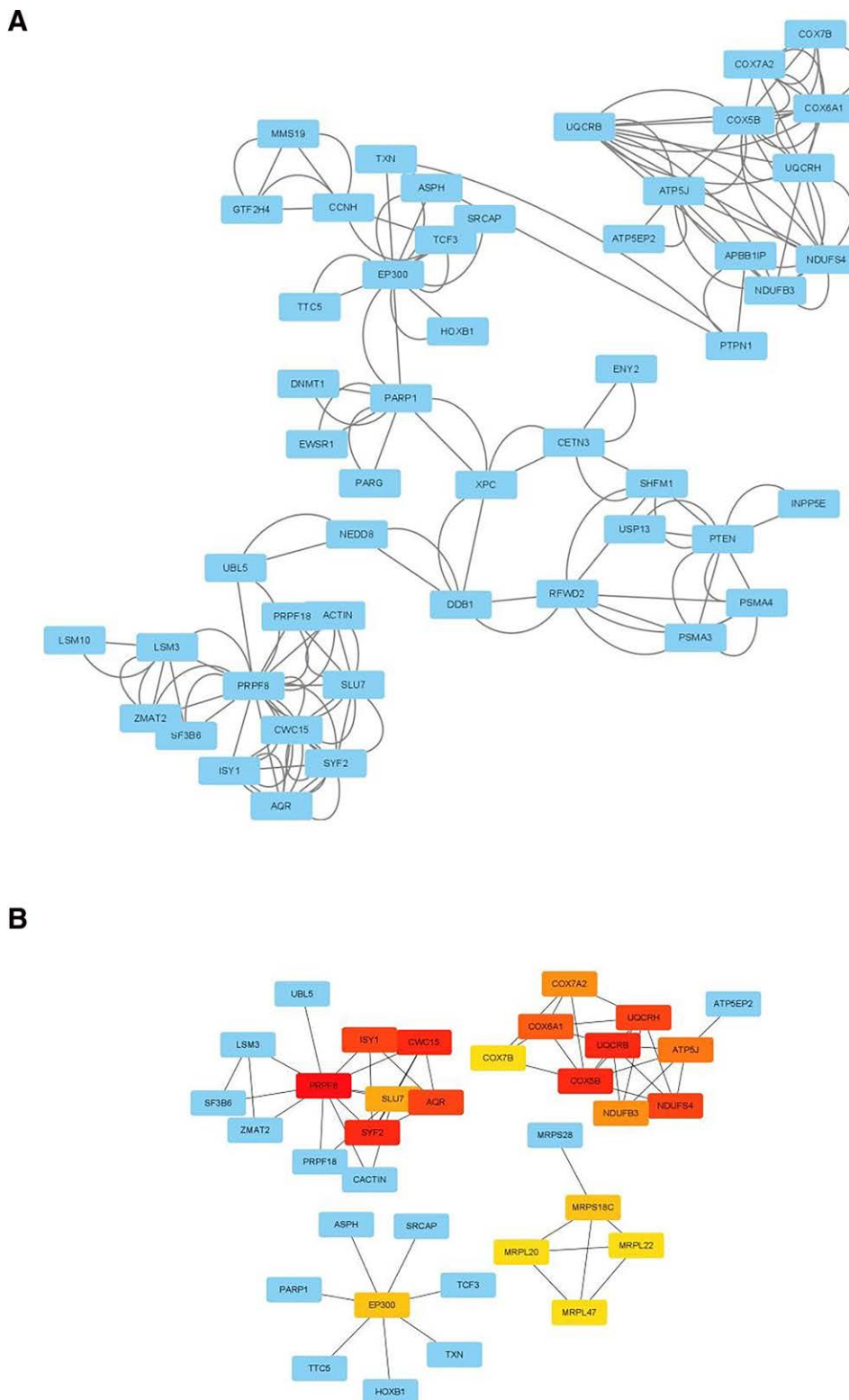
Figure 5. PPI network analyzed by cytoHubba app of Sytoscape. (A and B) Top 20 hub genes in DEGs of hub module genes. DEGs = differentially expressed genes, PPI = protein-protein interaction.

when the expression levels of INPP5E, HGD, and CYB5R1 are 8.2, 7.5, and 8.4, respectively, the total score is 172 points, indicating an onset risk of AS of approximately 0.96. To assess the diagnostic value of the 3 genes, CYB5R1, HGD, and INPP5E, in the diagnosis of AS, we analyzed their respective AUC values in the ROC curve analysis. The AUC values for CYB5R1, HGD, and INPP5E were 0.879, 0.879, and 0.816, respectively. Notably, the AUC value based on the nomoscore reached 0.969, suggesting that these genes exhibit a high diagnostic value for AS (Fig. 8B–E). The calibration chart demonstrated excellent alignment between the calibration prediction curve and the standard curve, indicating good consistency between the actual
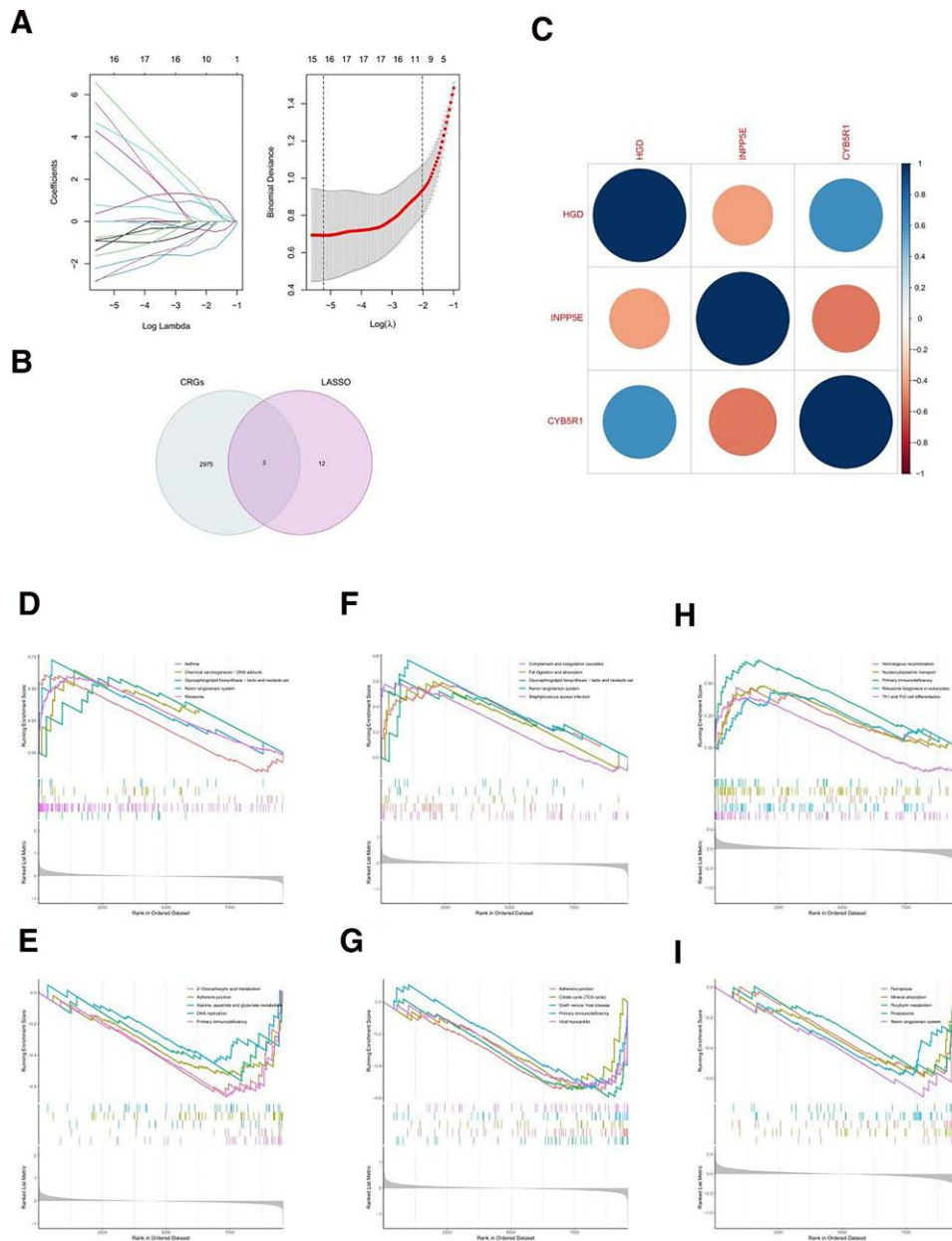
Figure 6. Screening of disease-related hub CRGs. (A) LASSO regression analysis was conducted to screen out disease feature genes and the logarithm (lambda) sequence was used to construct a coefficient distribution diagram. The optimal parameter (lambda) of the LASSO model was selected. (B) Venn diagram, presenting the 3 disease-related hub CRGs obtained from the intersection between 15 disease feature genes and 2978 CRGs. (C) Correlation analysis of 3 hub genes. The enrichment analysis showed the up-regulation (D) pathway and down-regulation (E) pathway of CYB5R1, up-regulation (F) pathway and down-regulation (G) pathway of HGD, and up-regulation (H) pathway and down-regulation (I) pathway of INPP5E. CRGs = cuprosis-related genes, HGD = Homogentisic acid 1,2-dioxygenase.

probability and predicted probability of the 3 hub genes in the nomogram model for the diagnosis of AS (Fig. 8F). Analyzing the DCA curve, we observed that the model performed well for most cases of the threshold probability, except for a small range where its performance was suboptimal. As the threshold probability increased, the net benefits of interventions based on the model's prediction results decreased. Importantly, interventions based on the model's prediction results resulted in positive net benefits, surpassing the standard value, when the threshold probability ranged from 0 to 0.9 (Fig. 8G).

Meanwhile, we have also generated a nomogram, ROC curve, calibration chart, and DCA curve using the verification sets GSE73754 and GSE18781 to analyze and validate the findings from the training set GSE25101. Figure 9A displays the nomogram generated based on the verification set GSE73754.

The ROC curve revealed that the AUC values for CYB5R1, HGD, and INPP5E were 0.493, 0.522, and 0.559, respectively. Additionally, the nomoscore was calculated to be 0.579, as depicted in Figure 9B–E. The calibration chart for the verification set GSE73754 demonstrated a misalignment between the calibration prediction curve and the standard curve, as illustrated in Figure 9F. The DCA curve indicated that under a threshold probability ranging from 0 to 0.73, interventions of AS showed mostly positive net benefits, closely resembling the standard value, as shown in Figure 9G. The nomogram generated based on the verification set GSE18781 is displayed in Figure 10A. In GSE18781, the AUC values for CYB5R1, HGD, and INPP5E were 0.752, 0.628, and 0.756, respectively. Furthermore, the calculated nomoscore was 0.620, as depicted in Figure 10B–E. The calibration chart for GSE18781
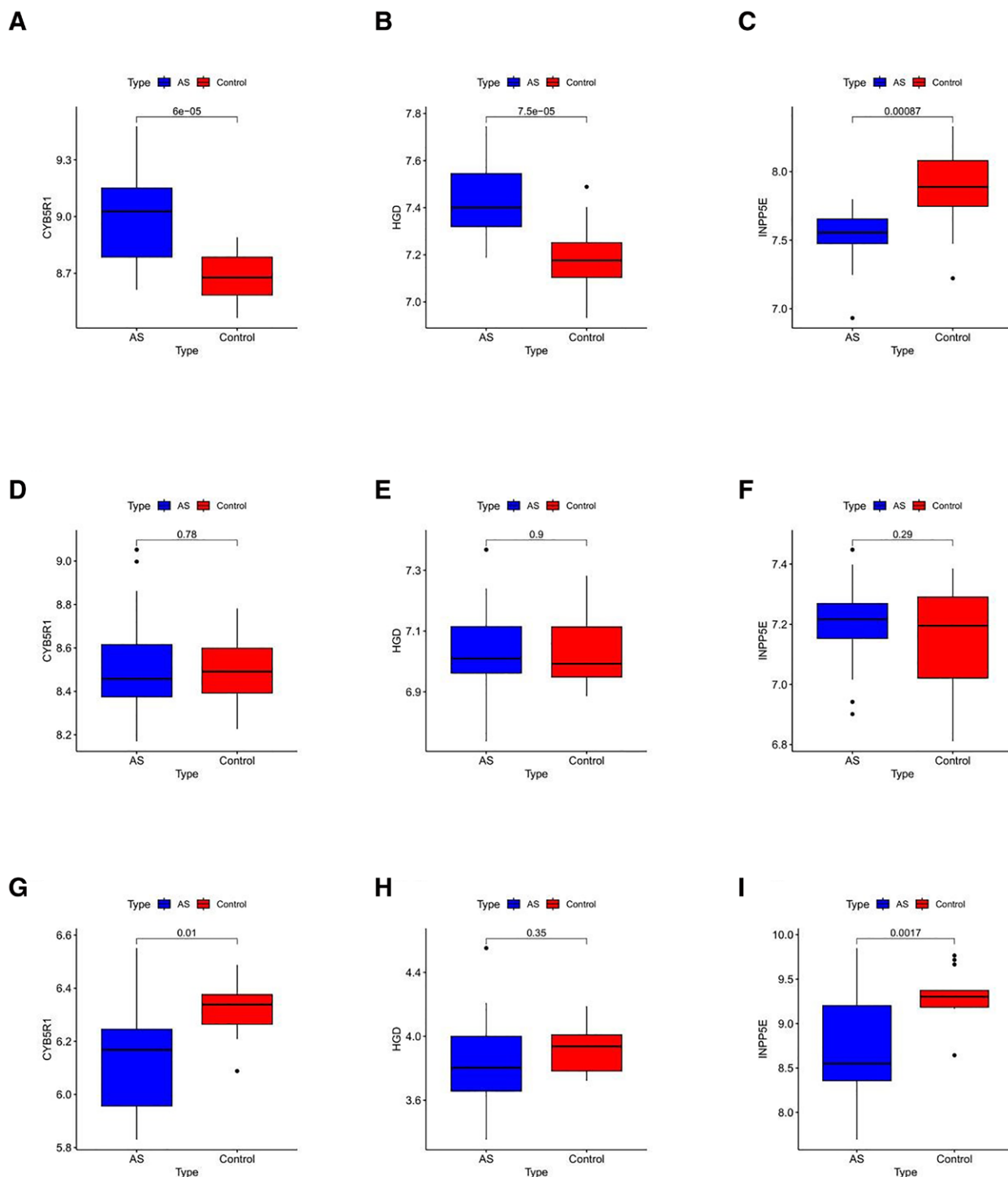
Figure 7. Expressions of disease-related hub CRGs in Training Set (GSE25101) and Verification Sets (GSE73754 and GSE18781). The comparison of expressions of CYB5R1 (A), HGD (B), and INPP5E (C) in samples collected from subjects in the AS group and control group in the training set GSE25101 is shown in box plots (A), (B), and (C). The comparison of expressions of CYB5R1 (D), HGD (E), and INPP5E (F) in samples collected from subjects in the AS group and control group in the verification set GSE73754 is shown in box plots (D), (E), and (F). The comparison of expressions of CYB5R1 (G), HGD (H), and INPP5E (I) in samples collected from subjects in the AS group and control group in the verification set GSE18781 is shown in box plots (G), (H), and (I). CRGs = cuprosis-related genes, HGD = Homogentisic acid 1,2-dioxygenase.

showcased a strong alignment between the calibration prediction curve and the standard curve, as shown in Figure 10F. The DCA curve illustrated that, with the exception of a small range exhibiting poor performance, interventions of AS showed mostly positive net benefits, slightly surpassing the standard value, under nearly all threshold probabilities, as depicted in Figure 10G.

3.9. Analysis of hallmark pathways based on ssGSEA algorithm

The ssGSEA algorithm was utilized to analyze the disparities in Hallmark pathways between the control group and AS group. Subsequently, a box plot was generated to visualize the findings. The box plot clearly demonstrated that there were substantial variations
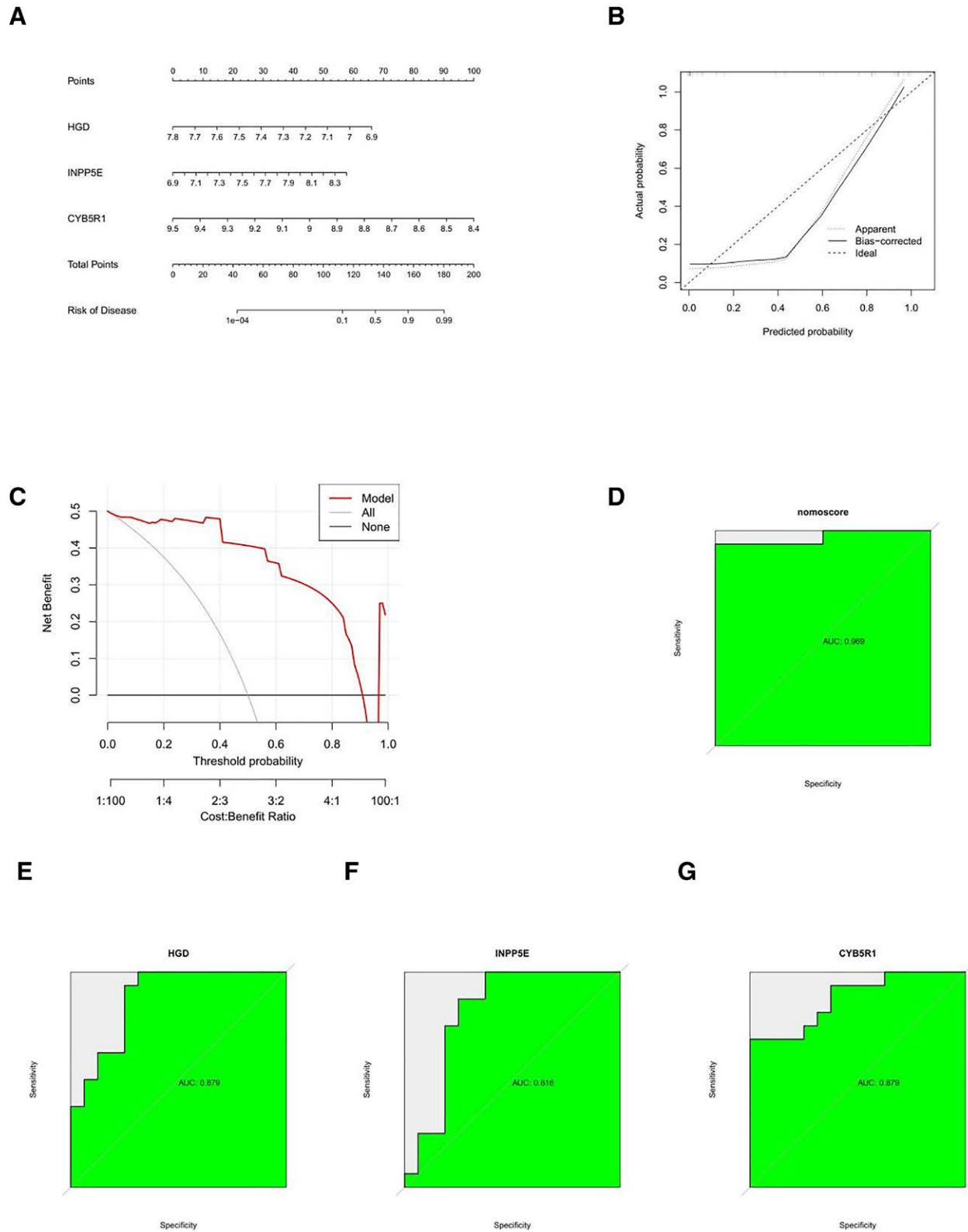
Figure 8. Construction of prediction curves based on 3 hub genes in GSE25101 for Risk and benefit assessment of AS. (A) Nomogram constructed based on 3 core genes. (B) Calibration curve for calibrating the nomogram. (C) DCA curve. (D) ROC curve based on the nomoscore of 3 core genes. (E) ROC curve of HGD. (F) ROC curve of INPP5E. (G) ROC curve of CYB5R1. AS = ankylosing spondylitis, DCA = decision curve analysis, HGD = Homogentisic acid 1,2-dioxygenase, INPP5E = inositol polyphosphate-5-phosphatase, ROC = receiver operating characteristic.

in the majority of Hallmark pathways between the control group and AS group, which is illustrated in Figure 11A. Notably, the correlation analysis highlighted significant associations between several categories of Hallmark pathways and the expressions of the disease-related hub CRGs (INPP5E, CYB5R1, and HGD). Specifically, it was observed that most Hallmark pathways exhibited a positive correlation with CYB5R1 and HGD, whilst displaying a negative correlation with INPP5E, as illustrated in Figure 11B.
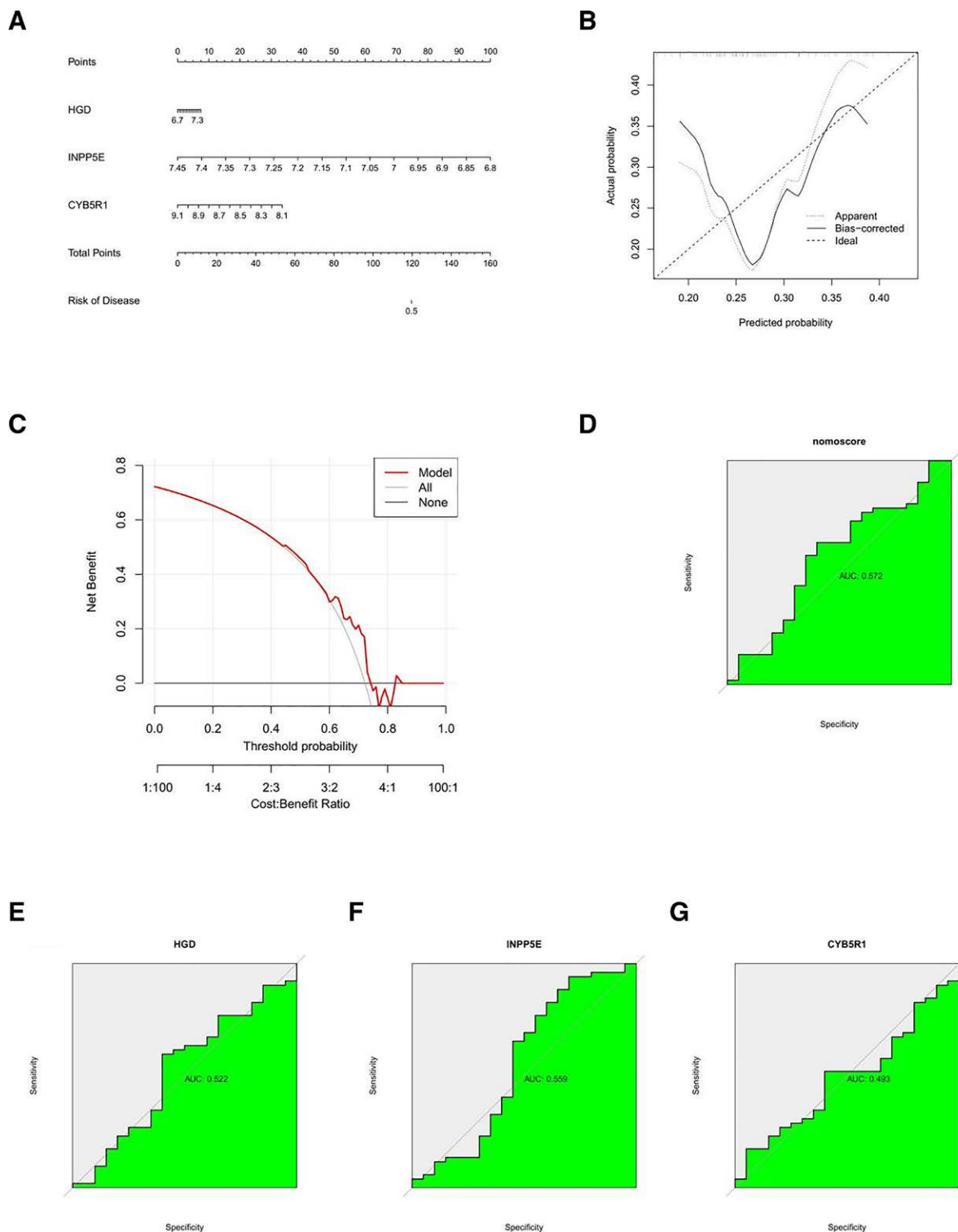
Figure 9. Construction of prediction curves based on 3 hub genes in GSE73754 for Risk and benefit assessment of AS. (A) Nomogram constructed based on 3 core genes. (B) Calibration curve for calibrating the nomogram. (C) DCA curve. (D) ROC curve based on the nomoscore of 3 core genes. (E) ROC curve of HGD. (F) ROC curve of INPP5E. (G) ROC curve of CYB5R1. AS = ankylosing spondylitis, DCA = decision curve analysis, HGD = Homogentisic acid 1,2-dioxygenase, INPP5E = inositol polyphosphate-5-phosphatase, ROC = receiver operating characteristic.

3.10. Analysis of immune cell infiltrations based on ssGSEA algorithm

The ssGSEA algorithm can be utilized to validate the distinction in immune cell infiltrations between the AS group and the control group in GSE25101, and to establish their correlation with diagnostic genes. The box plot results illustrated significant disparities in the expressions of the majority of immune cell infiltration clusters between the control group and AS group, as depicted in Figure 12A. Additionally, the correlation analysis unveiled significant associations between multiple types of immune cell infiltrations and the expressions of disease-related
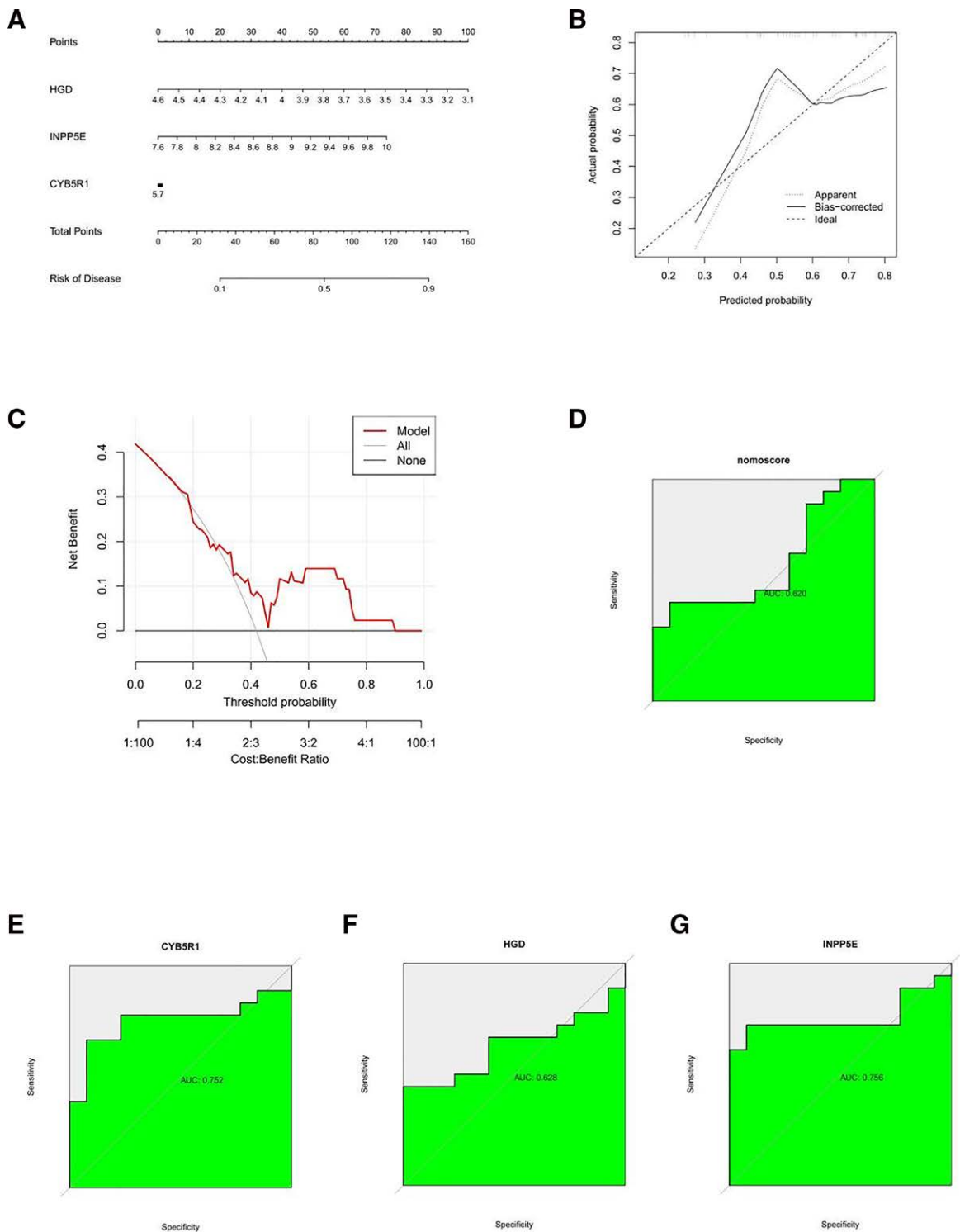
Figure 10. Construction of Prediction Curves Based on Three Hub Genes in GSE18781 for Risk and Benefit Assessment of AS. (A) Nomogram constructed based on 3 core genes. (B) Calibration curve for calibrating the nomogram. (C) DCA curve. (D) ROC curve based on the nomoscore of 3 core genes. (E) ROC curve of CYB5R1. (F) ROC curve of HGD. (G) ROC curve of INPP5E. AS = ankylosing spondylitis, DCA = decision curve analysis, HGD = Homogentisic acid 1,2-dioxygenase, INPP5E = inositol polyphosphate-5-phosphatase, ROC = receiver operating characteristic.

hub CRGs (INPP5E, CYB5R1, and HGD). Notably, most of the immune cell infiltrations displayed a negative correlation with CYB5R1, while exhibiting a positive correlation with HGD and INPP5E. Furthermore, the degree of positive correlation with INPP5E was remarkably higher than that with HGD, as exemplified in Figure 12B.

4. Discussion

AS is a chronic autoimmune inflammatory disease that primarily affects the spine and sacroiliac joints. Extensive research has demonstrated that the development of this disease is the result of a intricate interplay between genetic, environmental, and immune factors. Consequently, it is of paramount importance
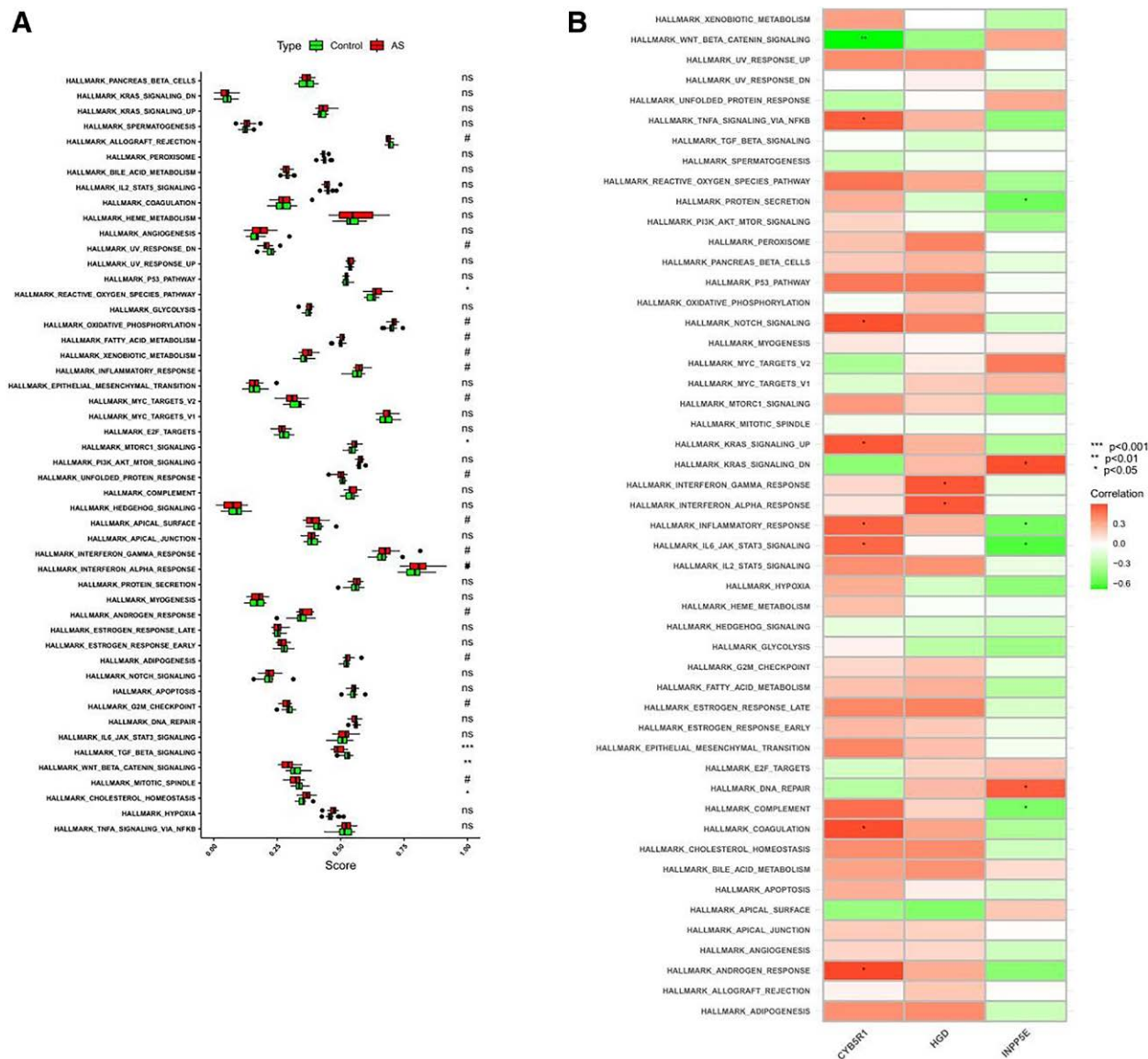
Figure 11. (A) Box plot for the differences in Hallmark pathways between the control group and AS group. (B) Heatmap for the correlation between the 3 core genes and Hallmark pathways in samples of the AS group. AS = ankylosing spondylitis.

to identify more effective targets for the diagnosis and treatment of AS. Nonetheless, the precise connection between copper death-related genes and AS remains poorly understood. The objective of our investigation is to shed light on the involvement of genes associated with copper death in AS phenotypes and immune microenvironments.

Copper death is a novel form of cell death characterized by an excessive intracellular buildup of copper ions. These copper ions directly bind to lipoylated dihydrolipoamide S-acetyltransferase in the tricarboxylic acid cycle, causing abnormal accumulation. This, in turn, leads to intracellular proteintoxic stress and ultimately cell death.[8] Research suggests that copper death may be involved in the occurrence and development of AS, as evidenced by its potential connection with key genes related to copper death. For instance, MTF1, a metal-binding transcription factor, is closely associated with copper homeostasis and is considered an important factor in the mechanism of copper death.[19,20] Furthermore, previous studies have found significantly higher serum copper levels in AS patients compared to normal individuals, and the excess copper ions in AS patients may indicate copper accumulation in the serum, consistent with the accumulation of copper ions observed in copper death.[21]

Additionally, there are other genes related to copper death that may be associated with AS, such as CDKN2A, LIPT1, PDHA1, and more. These genes are involved in the abnormal proliferation of fibroblast-like synovial cells (FLS) and the release of inflammatory factors during the development of AS.[22–24] These findings suggest that copper death may play a significant role in the progression of AS. In this study, we identified 3 CRGs (INPP5E, CYB 5R1, HGD) associated with AS development and validated them using bioinformatics. INPP5E is an important enzyme involved in tumor diseases and inflammation as an inositol polyphosphate 5-phosphatase. Studies have shown a significant correlation between autophagy and the onset of AS, where regulating autophagy levels can greatly impact the occurrence and development of this disease. Knockout of INPP5E leads to an accumulation of autophagy by disrupting its fusion with lysosomes. Therefore, we hypothesize that INPP5E may affect the occurrence and development of AS by regulating autophagy. Our study demonstrated that the expression levels of the INPP5E gene were decreased in AS samples, suggesting that down-regulation of INPP5E may increase the risk of AS. This finding was further validated in the validation set, with an area under the ROC curve greater than 0.7, indicating a high
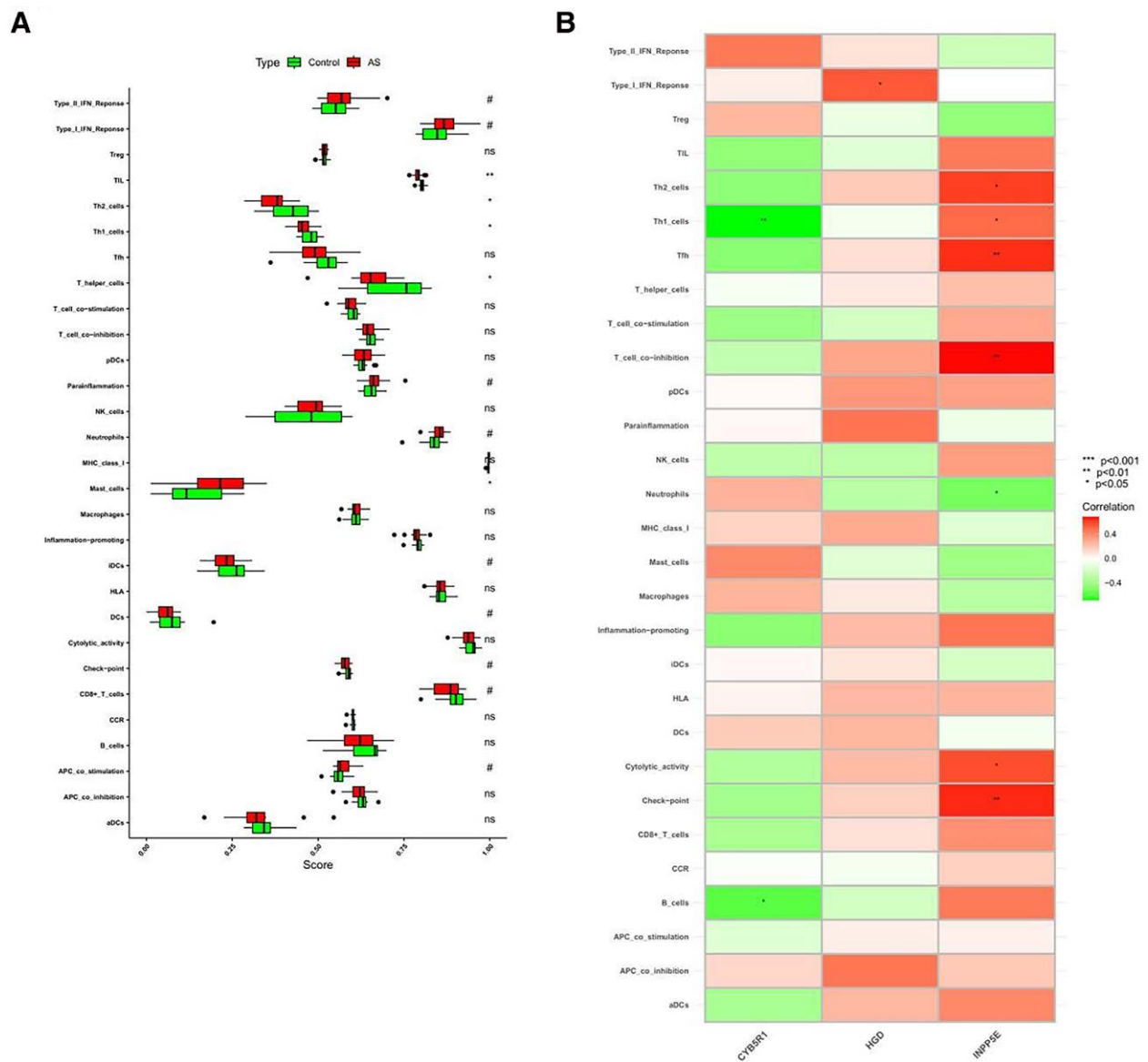
Figure 12. (A) Box plot for the differences in immune cells and functions between the control group and AS group. (B) Heatmap for the correlation between the 3 core genes and immune cells and functions in samples of the AS group. AS = ankylosing spondylitis.

diagnostic accuracy for AS patients[25] NADH-cytochrome b5 reductase 1 (CYB5R1) is an oxidoreductase that disrupts membrane integrity, leading to iron death in cells.[26] CYB5R1 has been shown to be essential for invasive tumor cell phenotypes and exhibits anti-apoptotic effects.[27] Considering the similarity between tumor invasion and the hyperproliferation of fibroblast-like synovial cells (FLS) in AS,[28] it suggests that CYB5R1 plays an important role in AS. Our study revealed that CYB5R1 was significantly up-regulated in AS samples compared to normal samples, indicating that up-regulation of CYB5R1 may increase the risk of AS. Moreover, the AUC value in ROC curve analysis was 0.879, highlighting the high diagnostic value of CYB5R1 for AS. Homogentisic acid 1,2-dioxygenase (HGD) is expressed in chondrocytes, synoviocytes, and osteoblasts. Deficiency of HGD leads to the accumulation of homogentisic acid in the urine, primarily affecting the spine and causing conditions such as spinal inflammation and spinal osteophyte proliferation.[29,30] However, contrary to expectations, our study discovered that HGD was significantly up-regulated in AS samples. This suggests that high expression of HGD may contribute to the progression of AS.

Most studies conducted so far have primarily focused on the immune response in the pathogenesis of AS, considering it as an autoimmune disease. A bioinformatics study discovered a significant enrichment of T cell receptor signaling pathways and natural killer cell-mediated cytotoxicity in AS patients.[16] It has been established that CD4 + T cells that produce IL-17 play a critical role in the development of AS, and the inhibition of T cell co-stimulation pathways may improve AS by reducing the overactivation of CD4 + T cells and diminishing IL-17 production.[31,32] Similarly, the involvement of immune follicular Th (TFH) cells, an important subset of effector T lymphocytes, has been closely linked to the pathogenesis of AS. These cells express chemokine receptor 5, T cell co-stimulatory factors, and secrete IL-21.[33,34] Additionally, B cells, Th1 cells, and type I interferons have been found to be highly associated with AS.[35] These findings highlight the significant role of the immune system in AS. In this study, the analysis of immunocyte infiltration demonstrated a positive correlation between the expression of INPP5E and the T cell co-inhibition pathway, as well as the immune checkpoint signal. This suggests that by overexpressing INPP5E, the activation of T cell co-inhibitory and immune

checkpoint signaling pathways can potentially enhance the efficacy of AS treatment. Furthermore, the expression of INPP5E displayed a positive correlation with Tfh cells. In AS samples, both INPP5E expression and Tfh immune infiltration were reduced. Conversely, studies have shown that strategies aimed at reducing Tfh cell production can alleviate immune diseases by reducing IL-21 secretion,[36,37] indicating that further research could be focused on Tfh cells. On the other hand, CYB5R1 expression exhibited a negative correlation with B cells and Th1 cells. Therefore, inhibiting CYB5R1 expression may activate immune cells and potentially improve AS. Additionally, the expression of HGD demonstrated a positive correlation with type I interferon. In AS samples, both HGD expression and type I interferon infiltration were increased. Studies have suggested that interferons can activate inflammatory cells, release cytokines, and other inflammatory mediators, thus exacerbating the negative effects of AS,[38] which is consistent with the findings of this study.

However, our study has certain limitations. First, we only included data on gene expressions derived from blood samples, and data on diseased tissues from AS patients were not considered in the selected datasets. Second, further verification of the expression levels of CRGs in AS patients is necessary. For our future study, we plan to collect blood samples from clinical AS patients and use Quantitative Real-Time PCR to determine the expression levels of CRGs. We will also explore the correlations between the expression levels of CRGs and clinical parameters of AS, such as clinical stages, inflammatory indicators, and prognosis. On the other hand, we have not yet conducted in vivo and in vitro experiments to investigate the potential mechanism of these CRGs in AS patients. Additionally, it is unfortunate that the verification results from the other two datasets did not fully align with those from the training set. Given the tissue specificity and individual differences, it is important to further verify the results using more datasets or clinical studies.

To summarize, this study has successfully identified and validated 3 CRGs (INPP5E, CYB5R1, and HGD) associated with the advancement of AS through the utilization of bioinformatics. These findings hold promise in terms of potentially serving as novel biomarkers for the timely detection, early recognition, and continuous monitoring of AS.

Acknowledgments

Author contributions

Conceptualization: Junyi Fan, Qihua Liu, Ting Chen, Junzhe Wu.
Data curation: Ting Chen, Yongbin Chen.
Formal analysis: Junyi Fan, Qihua Liu, Ting Chen, Yongbin Chen.
Funding acquisition: Junzhe Wu.
Investigation: Qihua Liu, Ting Chen, Yongbin Chen.
Methodology: Junyi Fan, Qihua Liu, Yongbin Chen, Junzhe Wu.
Writing – original draft: Junyi Fan.
Writing – review & editing: Junzhe Wu.

References

[1] Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet. 2017;390:73–84.

[2] Mahmoudi M, Aslani S, Nicknam MH, Karami J, Jamshidi AR. New insights toward the pathogenesis of ankylosing spondylitis; genetic variations and epigenetic modifications. Mod Rheumatol. 2017;27:198–209.

[3] Khan MA, Yong SB, Wei JC. Ankylosing spondylitis: history, epidemiology, and HLA-B27. Int J Rheum Dis. 2023;26:413–4.

[4] Jethwa H, Bowness P. The interleukin (IL)-23/IL-17 axis in ankylosing spondylitis: new advances and potentials for treatment. Clin Exp Immunol. 2016;183:30–6.

[5] van der Heijde D, Ramiro S, Landewé R, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis. 2017;76:978–91.

[6] Ward MM, Deodhar A, Gensler LS, et al. 2019 update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Care Res (Hoboken). 2019;71:1285–99.

[7] Mohammadi H, Hemmatzadeh M, Babaie F, et al. MicroRNA implications in the etiopathogenesis of ankylosing spondylitis. J Cell Physiol. 2018;233:5564–73.

[8] Tsvetkov P, Coy S, Petrova B, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. Science. 2022;375:1254–61.

[9] Hu Y, Amoah AN, Zhang H, et al. Effect of ginger in the treatment of nausea and vomiting compared with vitamin B6 and placebo during pregnancy: a meta-analysis. J Matern Fetal Neonatal Med. 2022;35:187–96.

[10] Lai Y, Lin C, Lin X, Wu L, Zhao Y, Lin F. Identification and immunological characterization of cuproptosis-related molecular clusters in Alzheimer's disease. Front Aging Neurosci. 2022;14.

[11] Huo S, Wang Q, Shi W, et al. ATF3/SPI1/SLC31A1 signaling promotes cuproptosis induced by advanced glycosylation end products in diabetic myocardial injury. Int J Mol Sci. 2023;24:1667.

[12] Luo L, Li A, Fu S, et al. [Cuproptosis-related immune gene signature predicts clinical benefits from anti-PD-1/PD-L1 therapy in non-small-cell lung cancer. Immunol Res. 2023;71:213–28.

[13] Liu B, Wang Y, Liu S, et al. A randomized controlled study of preoperative oral carbohydrate loading versus fasting in patients undergoing elective craniotomy. Clin Nutr. 2019;38:2106–12.

[14] Xiong Y, Cai M, Xu Y, et al. Joint together: the etiology and pathogenesis of ankylosing spondylitis. Front Immunol. 2022;13.

[15] Dai Y, Zhao Y, Nie K. The antiemetic mechanisms of gingerols against chemotherapy-induced nausea and vomiting. Evid Based Complement Alternat Med. 2022;2022:1–14.

[16] Zheng Y, Cai B, Ren C, et al. Identification of immune related cells and crucial genes in the peripheral blood of ankylosing spondylitis by integrated bioinformatics analysis. PeerJ. 2021;9:e12125.

[17] Han Y, Zhou Y, Li H, et al. Identification of diagnostic mRNA biomarkers in whole blood for ankylosing spondylitis using WGCNA and machine learning feature selection. Front Immunol. 2022;13.

[18] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008;9.

[19] Balamurugan K, Schaffner W. Copper homeostasis in eukaryotes: teetering on a tightrope. Biochim Biophys Acta. 2006;1763:737–46.

[20] Murphy BJ, Sato BG, Dalton TP, Laderoute KR. The metal-responsive transcription factor-1 contributes to HIF-1 activation during hypoxic stress. Biochem Biophys Res Commun. 2005;337:860–7.

[21] Aiginger P, Kolarz G, Willvonseder R. Copper in ankylosing spondylitis and rheumatoid arthritis. Scand J Rheumatol. 1978;7:75–8.

[22] Ni M, Solmonson A, Pan C, et al. Functional assessment of lipoyltransferase-1 deficiency in cells, mice, and humans. Cell Rep. 2019;27:1376–86.e6.

[23] Lin HC, Chen YJ, Wei YH, et al. Lactic acid fermentation is required for NLRP3 inflammasome activation. Front Immunol. 2021;12630380.

[24] Goronzy JJ, Henel G, Sawai H, et al. Costimulatory pathways in rheumatoid synovitis and T-cell senescence. Ann N Y Acad Sci. 2005;1062:182–94.

[25] Nakamura S, Hasegawa J, Yoshimori T. Regulation of lysosomal phosphoinositide balance by INPP5E is essential for autophagosome-lysosome fusion. Autophagy. 2016;12:2500–1.

[26] Woischke C, Blaj C, Schmidt EM, et al. CYB5R1 links epithelial-mesenchymal transition and poor prognosis in colorectal cancer. Oncotarget. 2016;7:31350–60.

[27] Ai Y, Yan B, Wang X. The oxidoreductases POR and CYB5R1 catalyze lipid peroxidation to execute ferroptosis. Mol Cell Oncol. 2021;8:1881393.

[28] Zhao J, Guo S, Schrodi SJ, He D. Cuproptosis and cuproptosis-related genes in rheumatoid arthritis: implication, prospects, and perspectives. Front Immunol. 2022;13.

[29] Doganavsargil B, Pehlivanoglu B, Bicer EK, et al. Black joint and synovia: histopathological evaluation of degenerative joint disease due to Ochronosis. Pathol Res Pract. 2015;211:470–7.

[30] Laschi M, Tinti L, Braconi D, et al. Homogentisate 1,2 dioxygenase is expressed in human osteoarticular cells: implications in alkaptonuria. J Cell Physiol. 2012;227:3254–7.

[31] Schütz C, Baraliakos X. What do we know about co-stimulatory and co-inhibitory immune checkpoint signals in ankylosing spondylitis? Clin Exp Immunol. 2023;213:288–300.

[32] Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol. 2013;13:227–42.

[33] Yusuf I, Kageyama R, Monticelli L, et al. Germinal center T follicular helper cell IL-4 production is dependent on signaling lymphocytic activation molecule receptor (CD150). J Immunol. 2010;185:190–202.

[34] Wu S, Yang T, Pan F, et al. Increased frequency of circulating follicular helper T cells in patients with ankylosing spondylitis. Mod Rheumatol. 2015;25:110–5.

[35] Rezaiemanesh A, Abdolmaleki M, Abdolmohammadi K, et al. Immune cells involved in the pathogenesis of ankylosing spondylitis. Biomed Pharmacother. 2018;100:198–204.

[36] Ji YR, Kim HJ, Yu DH, et al. Enforced expression of roquin protein in T cells exacerbates the incidence and severity of experimental arthritis. J Biol Chem. 2012;287:42269–77.

[37] Linterman MA, Rigby RJ, Wong RK, et al. Follicular helper T cells are required for systemic autoimmunity. J Exp Med. 2009;206:561–76.

[38] Xiao F, Zhang HY, Liu YJ, Zhao D, Shan YX, Jiang YF. Higher frequency of peripheral blood interleukin 21 positive follicular helper T cells in patients with ankylosing spondylitis. J Rheumatol. 2013;40:2029–37.