

Original Research

Identification of Biomarker *IL2R β* in Ankylosing Spondylitis via Multi-Chip Integration Analysis of Gene Differential Expression

Bin Xiao^{1,†}, Peng-Lei Cui^{1,†}, Hong-Chao Li², Chao Wang³, Yan-Zhuo Zhang³, Zhi-Min Wu³, Cheng-Ai Wu^{3,*} 

¹National Center for Orthopaedics, Department of Spine Surgery, Beijing Jishuitan Hospital, Capital Medical University, Beijing Research Institute of Traumatology and Orthopaedics, 100035 Beijing, China

²National Center for Orthopaedics, Department of Rheumatology and Immunology, Beijing Jishuitan Hospital, Capital Medical University, Beijing Research Institute of Traumatology and Orthopaedics, 100035 Beijing, China

³National Center for Orthopaedics, Department of Molecular Orthopaedics, Beijing Research Institute of Traumatology and Orthopaedics, Beijing Jishuitan Hospital, Capital Medical University, 100035 Beijing, China

*Correspondence: wuchengai@jst-hosp.com.cn (Cheng-Ai Wu)

†These authors contributed equally.

Academic Editor: Elisa Belluzzi

Submitted: 26 April 2023 Revised: 4 September 2023 Accepted: 8 September 2023 Published: 26 December 2023

Abstract

Background: Ankylosing spondylitis (AS) is a chronic inflammatory autoimmune disease that affects axial joints such as the spine. Early diagnosis is essential to improve treatment outcomes. The purpose of this study is to uncover underlying genetic diagnostic features of AS. **Methods:** We downloaded gene expression data from the Gene Expression Omnibus (GEO) database for three studies of groups of healthy and AS samples. After preprocessing and normalizing the data, we employed linear models to identify significant differentially expressed genes (DEGs) and further integrated the differential genes to acquire reliable differential transcriptional markers. Gene functional enrichment analysis was conducted to obtain enriched pathways and regulatory gene interactions were extracted from pathways to further elucidate pathway networks. Seventy-three reliably differentially expressed genes (DEGs) were integrated by differential analysis. Utilizing the regulatory relationships of the 21 Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway genes that were enriched in the analysis, a regulatory network of 622 genes was constructed and its topological properties were further analyzed. **Results:** Functional enrichment analysis found 73 DEGs that were strongly associated with immune pathways like Th17, Th1 and Th2 cell differentiation. Using KEGG combined with DEGs, six hub genes (*KLRD1*, *HLA-DRB3*, *HLA-DRB5*, *IL2R β* , *CD247*, and *CXCL10*) were suggested from the network. Of these, the *IL2R β* gene was significantly differentially expressed compared with the normal control. **Conclusion:** *IL2R β* (Interleukin-2 receptor beta) is strongly associated with the onset and progression of autoimmune joint diseases, and may be used as a potential biomarker of AS. This study offers new characteristics that can help in the diagnosis and individualized therapy of AS.

Keywords: ankylosing spondylitis; *IL2R β* ; crosstalk; biomarkers; support vector machines; pathway network

1. Introduction

Ankylosing spondylitis (AS) is a chronic inflammatory disease mediated by immunity that primarily impacts the axial joints like the spine [1]. AS can lead to severe stiffness, back pain and the formation of new bone, often resulting in progressive joint stiffness that decreases the quality of life [2]. Although the current evaluation of the disease state of AS measures disease activity, its progression and prognosis remain difficult to predict [3]. Furthermore, the potential molecular processes that promote AS progression remain largely unclear. Consequently, it is urgent to investigate the AS pathogenesis.

The rapid development of high-throughput transcriptomic profiling has made available massive repositories of data, enhancing efforts to improve the diagnosis of target diseases. In the past decade, several studies have used gene chips to determine candidate genes associated with AS [4–

7] focusing on a single data set of samples. Specifically, Pimentel-Santo *et al.* [4] reported a group of differentially expressed genes (DEGs) highly correlated with AS that also modulate the joint destruction and inflammatory process. Similarly, Zhao *et al.* [7] utilized a bioinformatics approach to predict genes related to AS, including *MRPL22*, *RPL17*, *PSMA4* and *PSMA6*. However, to date, most studies evaluating AS genetics have been concerned with individual genes or individual pathways, and multi-dataset integration analysis and multi-gene joint diagnosis are often overlooked.

A pathway contains a large amount of information about the interaction between genes and the synergy among them. Abnormality in some genes may lead to changes in other genes in the same pathway. As a result, establishing a pathway network and mining for disease markers has been an effective method in disease research.



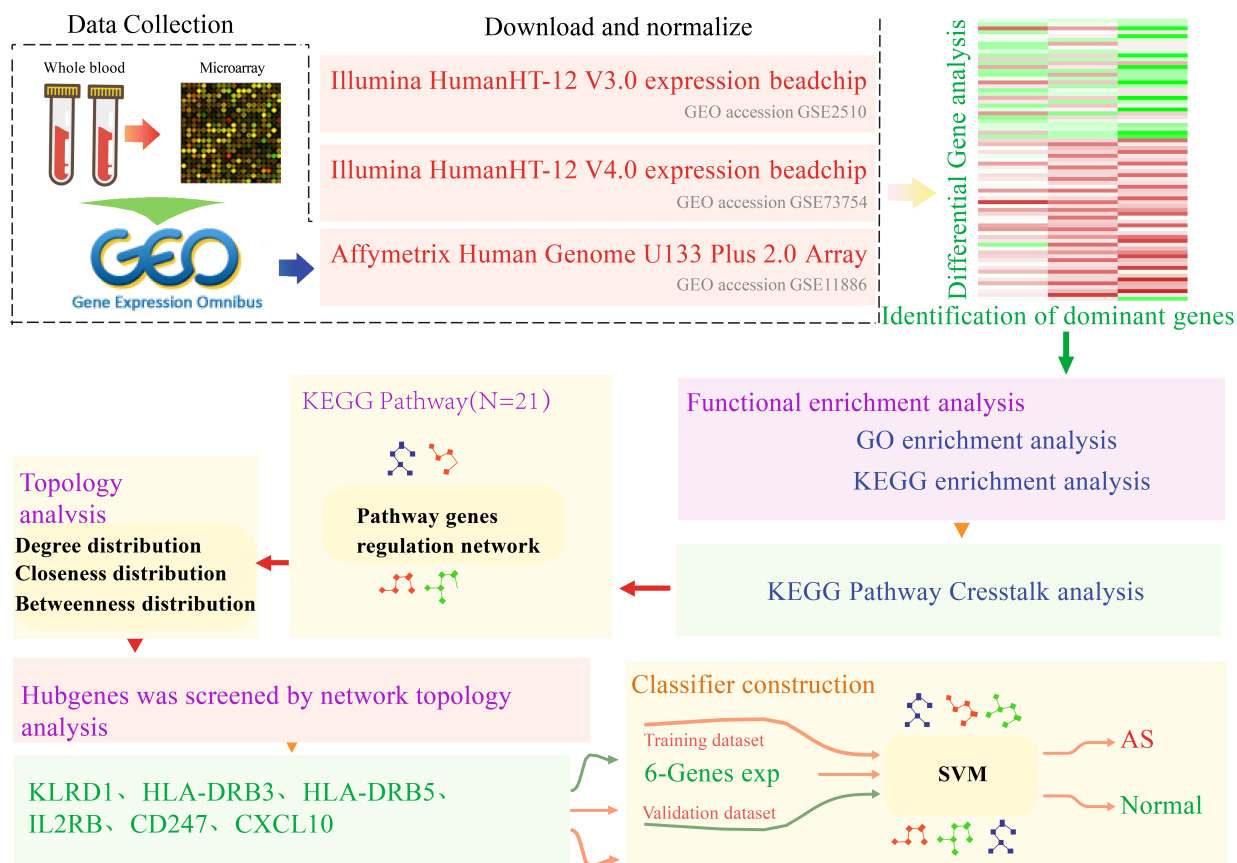


Fig. 1. Workflow chart. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; SVM, Support Vector Machine; GEO, Gene Expression Omnibus; AS, Ankylosing spondylitis.

This study integrates multi-platform high-throughput expression profiling data to (1) study altered gene expression patterns between AS and healthy women, (2) pre-screen diagnostic markers for AS, and (3) analyze these genes for their functions. These AS candidate biomarkers were further investigated by building Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway networks and an AS diagnostic classifier was established to assist clinical diagnosis.

2. Materials and Methods

Our method consisted of the following steps: collection of data, standardization, differential gene integration analysis, enrichment analysis, pathway integration analysis, pathway network construction, feature selection, and the construction and verification of the classifier. Fig. 1 summarizes the workflow.

2.1 Data Collection

We searched the Gene Expression Omnibus (GEO) database for studies on patients with AS and expression profiles from human blood samples were obtained. Three studies were identified and from them a dataset created of 76 cases and 45 healthy controls. Details are in Table 1.

2.2 Data Processing

For the two sets of gene expression data from the Illumina platform (<http://www.ncbi.nlm.nih.gov/geo/>), we downloaded the expression data after quantile normalization. Since there is no quantile normalization available for the GSE11886 dataset (<http://www.ncbi.nlm.nih.gov/geo/>), we downloaded the raw data and processed the data using the R package affy (version 4.2.0) [8] to acquire the probe expression profiles, which were further normalized using the RMA approach [9]. The result was an expression matrix of 54,675 gene probes for the 17 samples. The probes were annotated with R package hgu133plus2.db; probes matching multiple genes were deleted, and where multiple probes were matched to one gene, the median of the probes was used as the gene expression value. Subsequently, 23,313 gene expression profiles were acquired.

2.3 Integrated Analysis of Microarray Datasets

The R software package limma [10] was used to log₂-transform and normalize the matrix data in each GEO dataset. DEGs between normal and AS samples were then detected using the limma package. The DEGs recognized from the three datasets were genetically integrated with the R package “RobustRankAggreg” [11], which is

Table 1. Studies and data included in this analysis.

Study	GEO accession	Sample size		Sample source	Platform
		AS case	Control		
1	GSE2510	16	16	Blood	GPL6947 Illumina HumanHT-12 V3.0 expression BeadChip
2	GSE73754	52	20	Blood	GPL10558 Illumina HumanHT-12 V4.0 expression BeadChip
3	GSE11886	8	9	Blood	GPL570 Affymetrix Human Genome U133 Plus 2.0 Array

based on the Robust Rank Aggregation (RRA) approach. Those genes that consistently ranked superior to expectations based on the null hypothesis of input irrelevance could be screened with the RRA approach. Therefore, an integration of gene expression values for samples from various datasets was not performed. In common with many who have utilized the RobustRankAggreg package [12,13], we did not correct for batch effects. $p < 0.05$ together with 1.2-fold change in DEGs were deemed to be statistically significant.

2.4 Functional Enrichment Analyses

R package clusterProfiler [14] was exploited for analyzing KEGG and Gene Ontology (GO) pathway enrichment to determine GO terms that appeared disproportionately in three categories (cellular component, molecular function, and biological processes). KEGG pathways were also analyzed by visualization with R package DOSE [15]. In both analyses, q -values (adjusted p -values for multiple tests) less than 0.05 indicate statistical significance.

2.5 Development of KEGG Pathway Gene Interaction Network

We downloaded the XML file of each KEGG pathway that was enriched in the DEGs from the KEGG website [16]. The R package XML was employed to extract entries, relationships, and group relationships from these XML files. The information on gene interactions was subsequently extracted with a combination of scripts to build the KEGG gene interaction network to visualize the network on Cytoscape [17], and examine the topological characteristics of the cytohubba network [18].

2.6 Network Topological Property Analysis to Screen AS Key Diagnostic Genes

Visualization was performed using Cytoscape [17], with the degree and closeness of the network computed through the use of the Cytoscape [17] plug-in cytohubba [18]. Genes of degree >20 and closeness >5 were selected and intersected with the differential genes to identify the final characteristic genes.

2.7 Development of Predictive Models for AS Diagnosis and Assessment of Model Predictive Capability

A support vector machine (SVM) [19] was used to build a diagnostic predictive model using genes to predict healthy versus AS individuals. As a supervised learn-

Table 2. Characteristics of patients with ankylosing spondylitis and healthy controls.

Characteristics	AS patients	Healthy controls	p -value
Age (years)	35.73 $\pm$ 14.29	33.64 $\pm$ 12.23	0.6687
Male (%)	73%	64%	1.000
Weight (kg)	72.82 $\pm$ 19.06	69.82 $\pm$ 15.73	0.7053
Height (cm)	167 $\pm$ 11.01	168.7 $\pm$ 7.254	0.5568
BMI (kg/m ²)	28.09 $\pm$ 7.354	24.45 $\pm$ 5.507	0.2315
Disease duration (years)	11.7 (1, 20)	—	—
HLA-B27 (+)	55%	—	—

Values are presented as mean $\pm$ standard deviation (SD), median (IQR), or number (%). Abbreviations: BMI, body mass index; HLA-B27, human leucocyte antigen-B27.

Table 3. Primer sequences utilized in quantitative Real-Time PCR (qRT-PCR).

Genes	F/R	Sequences (5'-3')
<i>KLRD1</i>	Forward	CAGGACCCAACATAGAACTCCA
	Reverse	GGAAATGAAGTAACAGTTGCACC
<i>HLA-DRB3</i>	Forward	CGGGGTTGGTGAGAGCTTC
	Reverse	AACCACCTGACTTCAATGCTG
<i>HLA-DRB5</i>	Forward	CGGGGTTGGTGAGAGCTTC
	Reverse	AACCACCTGACTTCAATGCTG
<i>IL2RB</i>	Forward	CTGCTTCACCAACCAGGGTTA
	Reverse	GGGGTCGTAAGTAAAGTACACCT
<i>CD247</i>	Forward	GCCAGAACCAGCTCTATAACG
	Reverse	GGCCACGTCTCTGTCCAA
<i>CXCL10</i>	Forward	GTGGCATTCAAGGAGTACCTC
	Reverse	TGATGGCCTTCGATTCTGGATT

ing model, SVM uses machine learning algorithms for data analysis and pattern recognition. SVM builds a hyperplane that can be applied to categorization and regression in infinite-dimensional or high-dimensional spaces. A group of training samples, each labeled as belonging to one of two classes is used by the SVM training algorithm to create a model, and new instances are then assigned to a category according to the model, thereby making it a non-probabilistic binary linear categorization. The three datasets were categorized into two validation datasets and one training dataset. The model was built with the training dataset, and its classification capability was confirmed using the 10-fold cross-validation approach. The constructed model was, in turn, used to predict the category of samples

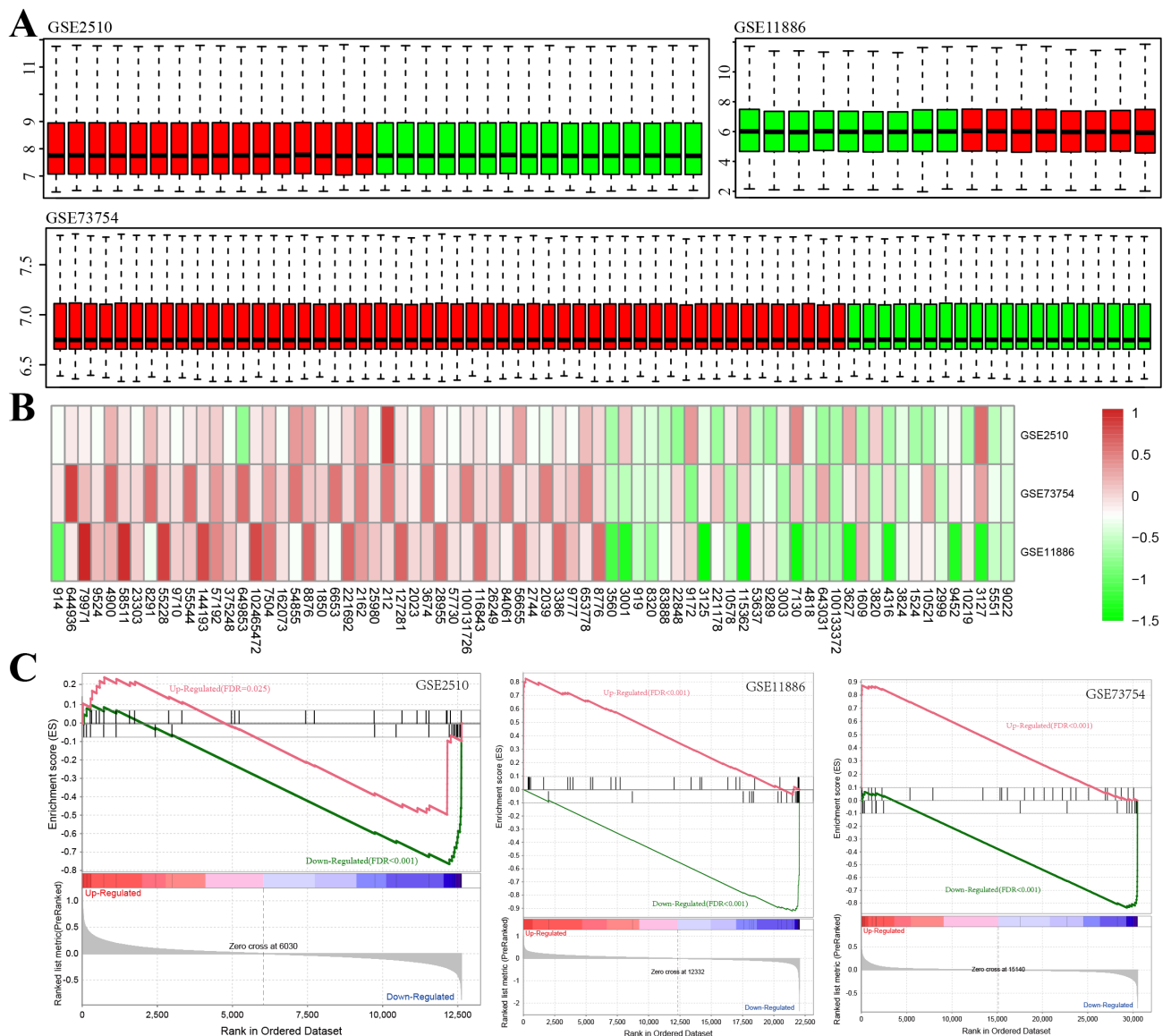


Fig. 2. Determination of differentially expressed genes (DEGs) between Healthy Control and AS Samples. (A) Box plot of the overall expression level of the normalized samples (green bars: normal samples, red bars: AS samples). (B) Heatmap of the down- and up-regulated DEGs from the integrated microarray analysis. Each column and row denote a dataset and a gene, separately. In each rectangle, the color denotes the log₂-FC value. The gradient from green to red goes from descending to ascending expression. (C) The enrichment results of log₂-FC in each dataset of differentially expressed genes. FC, fold change.

in the validation dataset. The model's predictive power was evaluated using the area under the receiver operating characteristic curve (ROC) curve (AUC), and the model was analyzed for its specificity and sensitivity in predicting AS.

2.8 Sample Preparation

Subjects were AS patients aged between 16 and 61 who were hospitalized in Jishuitan Hospital for the first time. The inclusion criteria were: (1) the diagnosis was in accordance with the New York classification revised in 1984; and (2) patients had not yet received medical treatment, including non-steroidal anti-inflammatory drugs or disease-modifying anti-rheumatic drugs. The exclusion cri-

teria were: (1) individuals with a history of severe infection, surgery, or tumor; (2) Individuals with severe heart, lung, liver, and kidney dysfunction; and (3) individuals with a history of autoimmune diseases such as rheumatoid arthritis. Included in the study was a control group consisting of healthy volunteers between 24 and 60 without acute or chronic diseases, such as rheumatoid arthritis, and no long-term use of any drugs. Table 2 shows the characteristics of the AS patients and healthy control subjects.

Our experiments were in line with the principles of the Declaration of Helsinki and were approved by the ethics committee of Beijing Jishuitan Hospital (ethics code: 201611-03). From all participants, fasting whole blood was

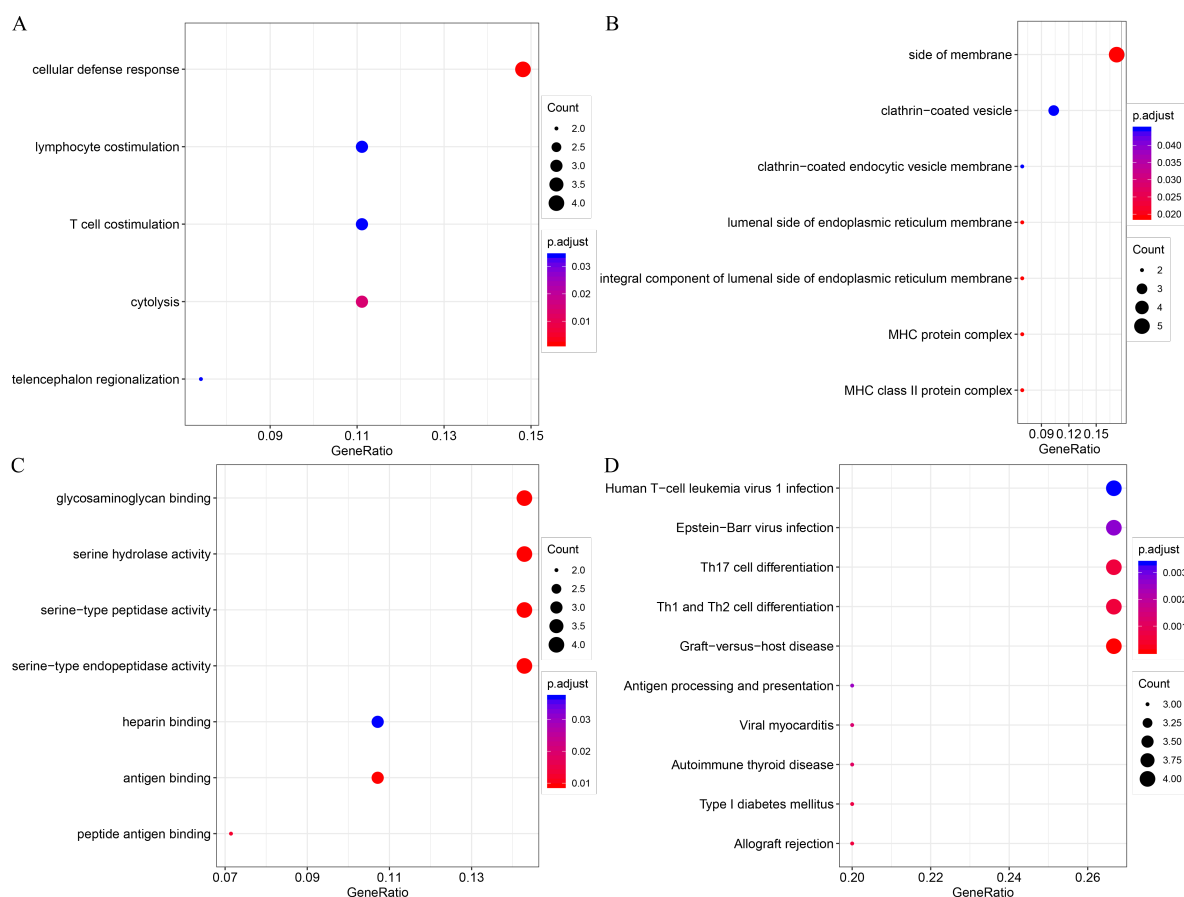


Fig. 3. Functional enrichment analysis of 31 downregulated DEGs. (A) Enriched GO terms in the “biological process” category. (B) Enriched GO terms in the “cellular component” category. (C) Enriched GO terms in the “molecular function” category. (D) Enriched KEGG biological pathways. The x-axis indicates the ratio of DEGs, the y-axis shows the various categories, the different colors denote levels of significance, and the size represents the number of differentially expressed genes. All enrichment results are provided in the **Supplementary Table 2**.

collected in sodium heparin-coated tubes (Greiner Bio-one, Chonburi, Thailand) and stored at -80°C until further analysis. All RNA samples had an A260/A280 ratio between 1.66 and 2.58 and concentrations ranging from 10.9–143 ng/mL and thus met the criteria for gene expression analysis.

2.9 Quantitative Real-Time PCR (qRT-PCR)

Extraction of total RNA from blood samples was implemented with Trizol reagent (Tiangen Biochemical, Beijing, China) and later extracted RNA was reverse transcribed to cDNA using FastQuant RT Kit containing gDNase (Tiangen Biochemical). Amplification was performed on an ABI7500 real-time fluorescence quantitative PCR instrument utilizing the Power SYBR® Green PCR Kit (Applied Biosystems, Foster City, CA, USA). Table 3 gives the primer sequences. The procedure followed was: pre-denaturation 95°C for 10 min; 95°C for 15 s; and 60°C for 60 s for 40 cycles. Three replicates were employed. Relative expression levels were determined via the $2^{-\Delta\Delta\text{CT}}$ method.

2.10 Statistical Analysis

All data are expressed as mean $\pm$ standard deviation (SD) calculated using the GraphPad Prism 9.0 software (GraphPad Software Inc., San Diego, CA, USA). An unpaired two-tailed Student’s *t*-test was performed to compare the differences between healthy and AS groups. $p < 0.05$ was considered statistically significant.

3. Results

3.1 Determination of DEGs between Healthy Control and AS Samples

Gene chip data from three datasets, GSE2510, GSE73754, and GSE1188655, were acquired from the GEO database. Following standardization, the gene expression distribution of each sample appeared similar (Fig. 2A). Genes with differential expression between AS and healthy samples from all three data sets were then further integrated and analyzed using limma. Altogether, 73 DEGs were identified (**Supplementary Table 1**), with 31 and 42 genes increased in the healthy and AS group, respectively, as indi-

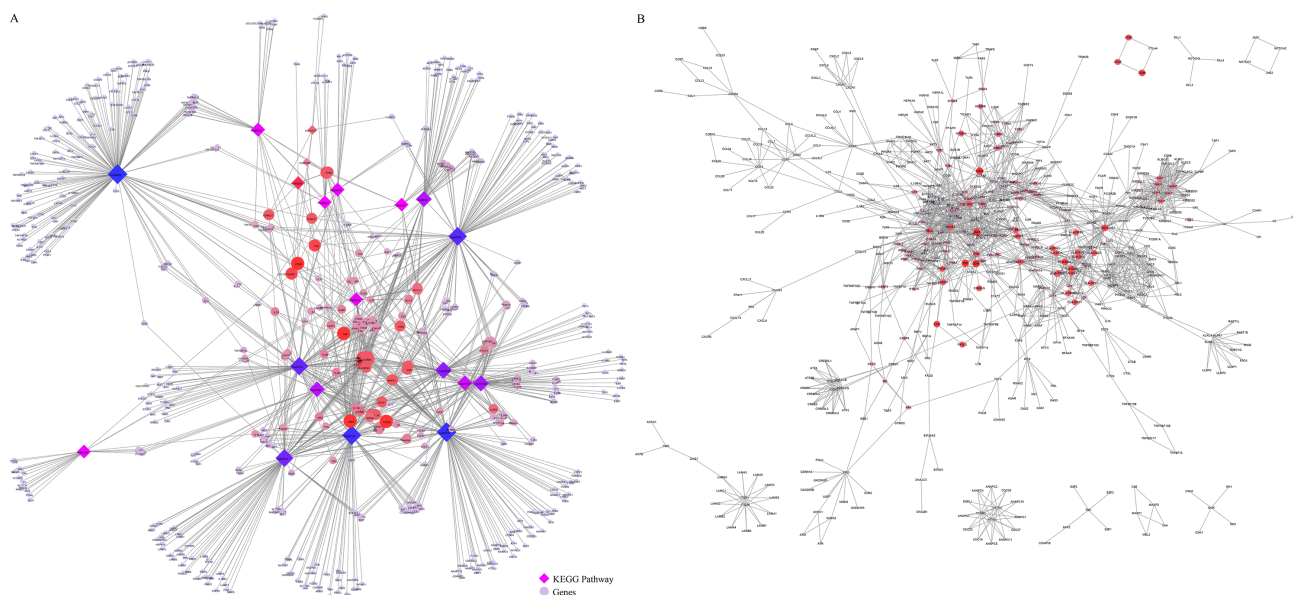


Fig. 4. KEGG pathway gene interaction network analysis. (A) Crosstalk among genes; color, and size between KEGG pathways indicate the number of pathways in which a particular gene is involved. (B) The interaction network; color and size of KEGG pathway genes indicate the number of pathways in which genes participate.

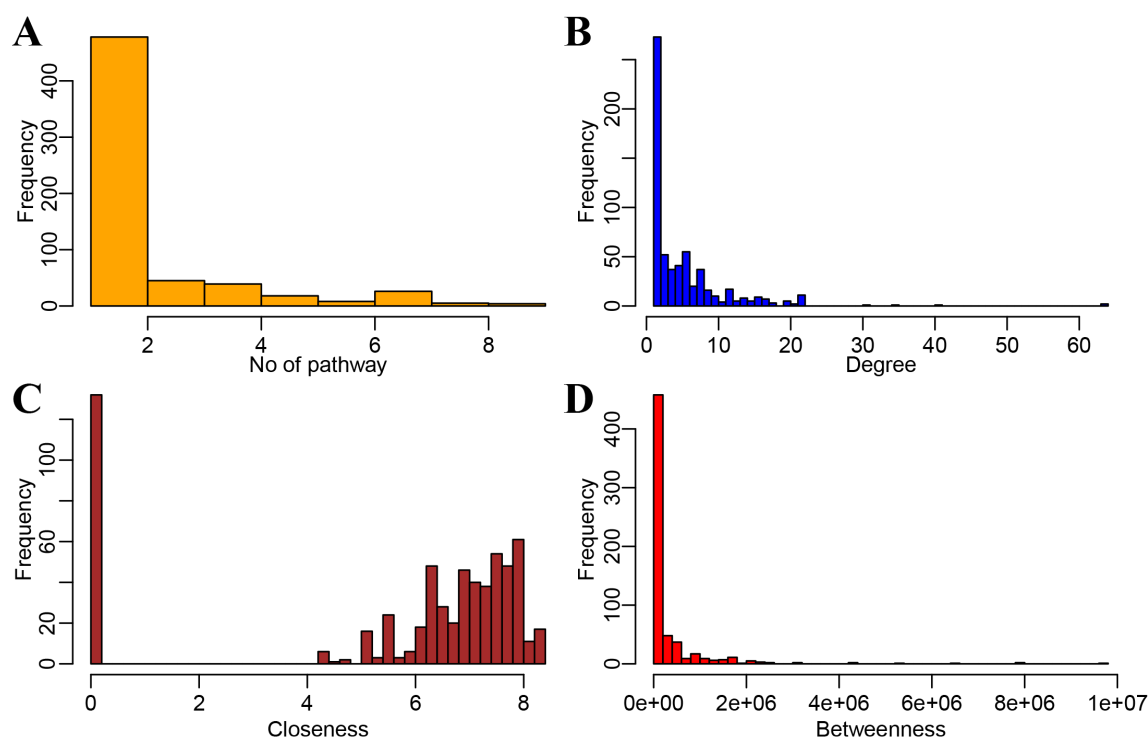


Fig. 5. Network topological property analysis to screen AS key genes. (A) Distribution of the number of pathways involved in the pathway gene network. (B) Degree distribution of the network. (C) Closeness distribution of the network. (D) Betweenness distribution of the network.

cated by the differential multiple heatmaps in each dataset (Fig. 2B). The enrichment results of the differential multiple GSEA [20] in all data sets were analyzed (Fig. 2C). In each data set, up- and downregulated genes were predominantly enriched in groups with high and low differential

multiples, respectively. The only exception was observed for the upregulated genes in data set GSE2510 where they were enriched in the group with lower differential multiplicity, which may be caused by sampling error.

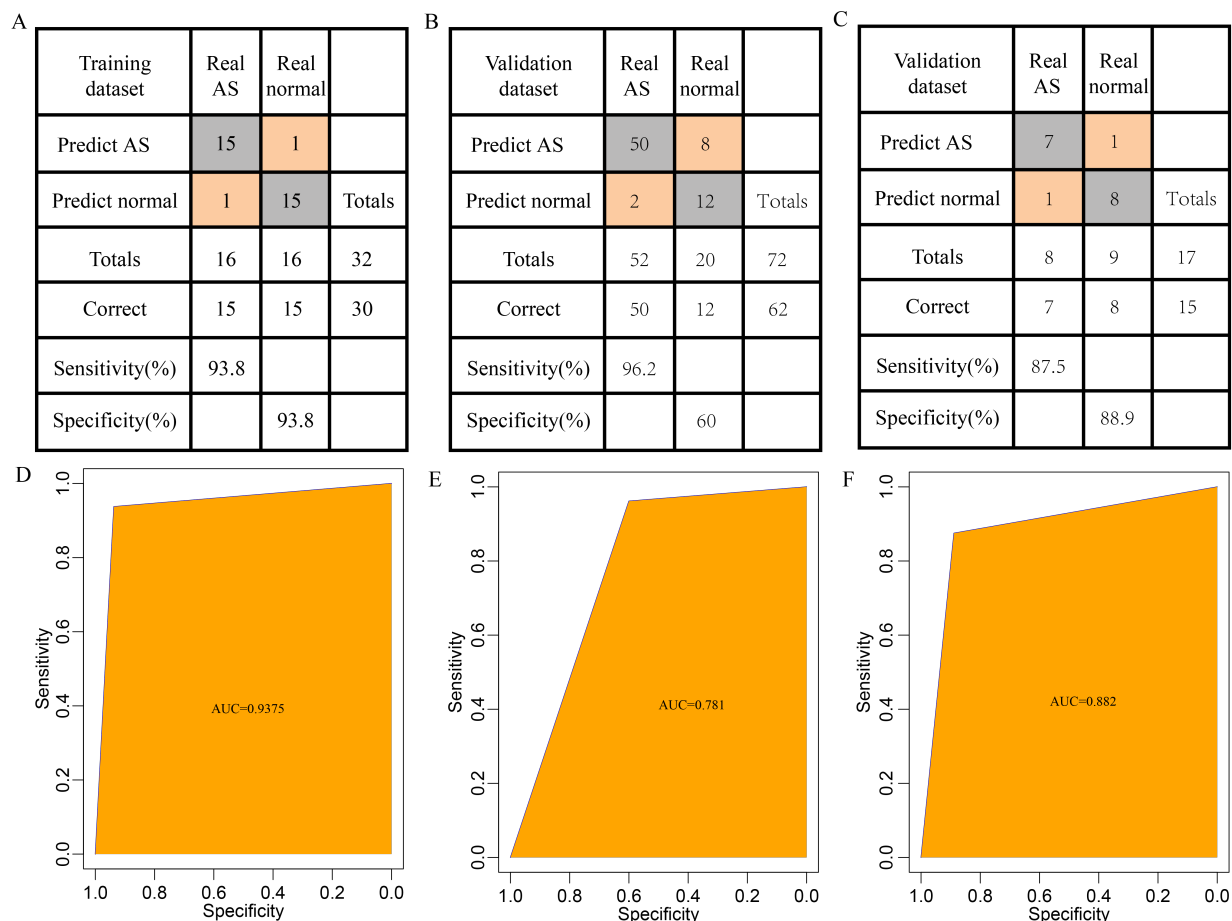


Fig. 6. Establishment and Testing of the Diagnostic Model. (A) Categorization results of the diagnostic model for the GSE2510 training dataset. (B) Categorization results of the diagnostic model for the GSE73754 validation dataset. (C) Categorization results of the diagnostic model for the GSE11886 validation dataset. (D) Receiver operating characteristic curve (ROC) curve of the diagnostic model for the training dataset (area under the curve (AUC) = 0.9375). (E) ROC curve of the diagnostic model for the GSE73754 validation dataset (AUC = 0.781). (F) ROC curve of the diagnostic model for the GSE11886 validation dataset (AUC = 0.882).

3.2 Functional Enrichment Analysis of DEGs

To gain insight into the functional implications of these 73 DEGs, GO and KEGG functional enrichment analyses were implemented on the 31 upregulated and 42 downregulated DEGs. While 42 increased genes were not enriched in any KEGG pathways and GO terms, 31 decreased genes were identified as enriched in multiple KEGG pathways and GO terms. In the bioprocess group, five enriched GO terms were noted, including “cellular defense response”, “cytolysis”, and “T cell costimulation” (Fig. 3A), suggesting that these genes are associated with the cellular defense response. Classification by cellular component displayed seven enriched GO terms, for example, “lumen side of endoplasmic reticulum membrane” and “MHC protein complex” (Fig. 3B). Seven molecular functions were found to be mediated by these genes and were primarily related to “heparin-binding” and “antigen-binding” (Fig. 3C). Furthermore, 21 KEGG pathways were identified including “Type I diabetes mellitus”, “Th17 cell differentiation”, and “Th1 and Th2 cell differentiation” (Fig. 3D). Previous

studies have also found a close relationship between “Th17 cell differentiation” and AS [21–23]. Together these data suggest that the downregulated DEGs are closely related to immunity.

3.3 KEGG Pathway Gene Interaction Network

For the 21 KEGG pathways identified, the XML files were downloaded from the KEGG website, and genetic interaction information was extracted using the R package XML. Specifically, gene ID was transformed into a gene symbol, a KEGG pathway gene interaction network was built and 622 genes were derived. Genetic analysis of the crosstalk between these pathways indicates that a close relationship exists among most pathways, and some genes are involved in multiple pathways (Fig. 4A). A total of 1663 genetic interactions were obtained based on the regulatory information of genes in the KEGG pathway (Fig. 4B). In addition, most of the genes involved in multiple pathways are closely related to other genes in the network.

3.4 Network Topological Property Analysis to Screen AS Key Genes

We analyzed the pathways involved in the KEGG gene interaction network and found that in total, 248 (39.8%) genes were involved in only one pathway (Fig. 5A). A small number of genes are involved in multiple pathways. We also observed a power-law degree distribution in our network (Fig. 5B), where the degree of most genes is less than 20. Additionally, closure of the network revealed that the majority of the nodes have lower closeness and only a few have high closeness (Fig. 5C). The network distribution of betweenness was also computed and the majority of the nodes exhibit a lower betweenness (Fig. 5D). Nodes with high degree, high betweenness or high closeness are viewed as significant in a network.

We considered a node to be the hub gene of the access network when it met the criteria of degree >20, closeness >5, and betweenness >0. Six genes met this criteria (*KLRD1*, *HLA-DRB3*, *HLA-DRB5*, *IL2RB*, *CD247*, *CXCL10*). Among these genes, *HLA-DRB3* and *HLA-DRB5* are HLA class II histocompatibility antigens. Major histocompatibility complex (MHC) is an important risk factor for AS; *IL2RB* is an important regulator of cytokine signaling and RET signaling in the immune system and has been reported as a potential transcriptional marker for AS [24]; *CXCL10* is a discriminant marker of autoimmune joint disease [25,26]. In sum, these six genes are strongly linked to the development and progression of autoimmune joint diseases, and can potentially be used as markers for AS.

3.5 Establishment and Testing of the Diagnostic Model

To construct a diagnostic model, GSE2510 was used as the training dataset ($n = 32$, AS = 16) while GSE73754 ($n = 72$, AS = 52) and GSE11886 ($n = 17$, AS = 8) were used as validation datasets. Six previously chosen hub genes were utilized as features in the training dataset to acquire their corresponding expression profiles. After the z-score, the support vector machine classification model was established. A 10-fold cross-validation indicated the classification accuracy was 93.8%, with 30 of the 32 samples categorized correctly (Fig. 6A). The model's specificity and sensitivity for AS were both 93.8%. The AUC was 0.94 (Fig. 6D). Validation of the model was then conducted with the GSE73754 dataset; 62 out of 72 samples were categorized correctly for an 86.1% classification accuracy. The model's specificity and sensitivity for AS in this dataset were 60% and 96.2%, respectively (Fig. 6B) and the AUC was 0.78 (Fig. 6E). Finally, for the GSE11886 dataset, 15 of 17 samples were correctly categorized with 88.2% classification accuracy. The model's specificity and sensitivity for AS were 88.9% and 87.5%, separately (Fig. 6C) and the AUC was 0.88 (Fig. 6F).

We validated these gene expressions in clinical samples. The findings demonstrated a significant difference

in *IL2RB* expression versus normal controls, which agrees with the findings of our bioinformatics analysis (Fig. 7).

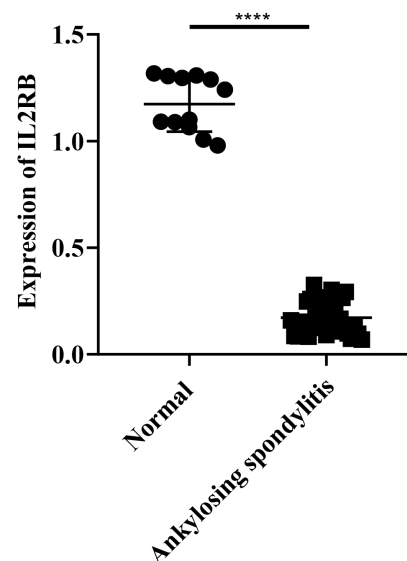


Fig. 7. Expressions of *IL2RB* in AS samples. The expression of *IL2RB* was lower in AS samples relative to controls. **** $p < 0.0001$.

These findings suggest that the diagnostic prediction model established in this research can differentiate effectively between healthy individuals and those with AS and that *IL2RB* can be used as a reliable biomarker for AS diagnosis.

4. Discussion

The comprehensive bioinformatics analysis focuses on the molecular screening of differential gene expression, and the network-based hub gene discovery has been extensively applied to determine underlying biomarkers relevant to the AS diagnosis and therapy. For example, Xu *et al.* [27] screened for hub genes by combining DEG screening, functional enrichment analysis, and PPI network construction, and further verified their outcomes by qRT-PCR. Ma *et al.* [28] used a set of GEO datasets to identify AS-related genes. Kanwal *et al.* [29] used protein interaction network analysis to find a *CD4*-centered AS anomaly-related regulatory network. Zhao *et al.* [7] used a set of GEO datasets for DEG screening, miRNA target gene analysis, and PPI network construction in efforts to screen for AS-related genes. Compared with previous studies, the current research not only integrates a relatively large sample size of microarray data from multiple GEO datasets but also constructs a pathway-based gene regulatory network in addition to an AS diagnostic model to identify potential diagnostic biomarkers for AS.

In our study, 73 consistent DEGs were identified, of which 42 were increased and 31 were decreased. The

genes downregulated were enriched in multiple GO terms and KEGG pathways. These pathways and GO terms were primarily correlated with immunological functions and the major histocompatibility complex (MHC), which is in keeping with prior findings that the immune pathway is affected in AS patients [6,30]. *CD2* is the most upregulated gene, a T-cell surface antigen that optimizes immune recognition by interacting with *LFA3* (*CD58*) on antigen-presenting cells and is linked to the onset and progress of various diseases [31–33]. The upregulated genes are not enriched in any of the KEGG pathways and GO terms, and thus their exact contribution to AS development is unknown. Further investigation is essential to clarify the roles of these genes as potential transcriptomic markers for AS. We identified six key genes, *KLRD1*, *HLA-DRB3*, *HLA-DRB5*, *IL2R β* , *CD247*, and *CXCL10* in the downregulated KEGG pathway gene regulatory network. Among these, *IL2RB* has been previously reported as a potential transcriptional marker for AS [24] and *CXCL10* is a distinguishing marker of autoimmune joint diseases [25,26]. All six genes are strongly correlated with the onset of autoimmune joint diseases.

In addition, as a common rheumatic disease, AS can cause inflammatory back pain, which greatly reduces quality of life [34]. Therefore, early diagnosis together with personalized medical interventions for AS are critical. In the past several years, gene expression profiling has been broadly applied to recognize biomarkers correlated with the disease [35–37]. Some of these identified biomarkers are functionally similar but have poor reproducibility, indicating they may not have accurate classification capabilities. In the current study, SVMs were employed to build and verify classifiers based on the expression profiles of six genes related to AS. The AUC of the training set reached 0.94 with our model, showing that these genes classify AS well. Additionally, in the verification datasets, we achieved AUCs of 0.88 and 0.78, which suggests these six genes could be applied as diagnostic markers for AS. Our results in clinical samples show that *IL2R β* was significantly differentially expressed in individuals with AS compared to healthy, which is consistent with our bioinformatics analysis. There were no significant differences in other genes, which may reflect the small sample size of the study. These findings reveal that the diagnostic prediction model built in this research can differentiate patients with AS from healthy controls effectively and that *IL2R β* can be a reliable biomarker for the diagnosis of AS. Moreover, *IL2R β* may be a potential target for the treatment of AS and/or a marker for monitoring disease activity and progression in AS patients.

Our work employed bioinformatics tools to identify underlying candidate genes involved in AS pathogenesis with large samples. However, limitations remain. First, the sample lacked other information about the clinic. Consequently, we did not account for the factors like whether the

patients had other diseases or health conditions. Second, the clinical sample size was small in validation experiments. Hence, further experimental and genetic studies with larger sample sizes and preclinical validation are required.

5. Conclusion

IL2R β is strongly associated with the onset and progression of autoimmune joint disease and may be a potential biomarker of AS. This study offers new genetic characteristics to help diagnosis and individualized therapy of AS. Our findings further provide potential targets and biomarkers for clinicians and researchers.

Abbreviations

AS, ankylosing spondylitis; GEO, Gene Expression Omnibus; DEGs, differentially expressed genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; SVM, support vector machine.

Availability of Data and Materials

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Author Contributions

CAW, BX and CW designed the study. BX, PLC, ZMW and YZZ carried out data acquisition and analysis. ZMW, YZZ and HCL carried out the verification experiment. BX and PLC wrote the manuscript. CW and CAW were revising the paper. HCL and CW contributed to preparing and making figures. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

All procedures for sample collection and use were approved by the Ethics Committee of Beijing Jishuitan Hospital, Beijing, China (ethics code: 201611-03). All authors approved to participate.

Acknowledgment

Not applicable.

Funding

This work was supported by the Beijing Municipal Health Commission (Grant No. BMHC-2021-6 and BJRITO-RDP-2023).

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2812343>.

References

- [1] El Maghraoui A. Extra-articular manifestations of ankylosing spondylitis: prevalence, characteristics and therapeutic implications. *European Journal of Internal Medicine*. 2011; 22: 554–560.
- [2] Reveille JD, Sims AM, Danoy P, Evans DM, Leo P, Pointon JJ, *et al.* Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nature Genetics*. 2010; 42: 123–127.
- [3] Machado P, Landewé R, Lie E, Kvien TK, Braun J, Baker D, *et al.* Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. *Annals of the Rheumatic Diseases*. 2011; 70: 47–53.
- [4] Pimentel-Santos FM, Ligeiro D, Matos M, Mourão AF, Costa J, Santos H, *et al.* Whole blood transcriptional profiling in ankylosing spondylitis identifies novel candidate genes that might contribute to the inflammatory and tissue-destructive disease aspects. *Arthritis Research & Therapy*. 2011; 13: R57.
- [5] Smith JA, Barnes MD, Hong D, DeLay ML, Inman RD, Colbert RA. Gene expression analysis of macrophages derived from ankylosing spondylitis patients reveals interferon-gamma dysregulation. *Arthritis and Rheumatism*. 2008; 58: 1640–1649.
- [6] Thomas GP, Duan R, Pettit AR, Weedon H, Kaur S, Smith M, *et al.* Expression profiling in spondyloarthropathy synovial biopsies highlights changes in expression of inflammatory genes in conjunction with tissue remodelling genes. *BMC Musculoskeletal Disorders*. 2013; 14: 354.
- [7] Zhao H, Wang D, Fu D, Xue L. Predicting the potential ankylosing spondylitis-related genes utilizing bioinformatics approaches. *Rheumatology International*. 2015; 35: 973–979.
- [8] Gautier L, Cope L, Bolstad BM, Irizarry RA. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004; 20: 307–315.
- [9] Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*. 2003; 31: e15.
- [10] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015; 43: e47.
- [11] Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012; 28: 573–580.
- [12] Yang J, Han S, Huang W, Chen T, Liu Y, Pan S, *et al.* A meta-analysis of microRNA expression in liver cancer. *PLoS ONE*. 2014; 9: e114533.
- [13] Shi KQ, Lin Z, Chen XJ, Song M, Wang YQ, Cai YJ, *et al.* Hepatocellular carcinoma associated microRNA expression signature: integrated bioinformatics analysis, experimental validation and clinical significance. *Oncotarget*. 2015; 6: 25093–25108.
- [14] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012; 16: 284–287.
- [15] Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics*. 2015; 31: 608–609.
- [16] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*. 2000; 28: 27–30.
- [17] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. 2003; 13: 2498–2504.
- [18] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*. 2014; 8: S11.
- [19] Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinformatics*. 2018; 19: 432.
- [20] Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*. 2007; 23: 3251–3253.
- [21] Gaston JSH, Jadon DR. Th17 cell responses in spondyloarthritis. *Best Practice & Research. Clinical Rheumatology*. 2017; 31: 777–796.
- [22] Ridley A, Hatano H, Wong-Baeza I, Shaw J, Matthews KK, Al-Mossawi H, *et al.* Activation-Induced Killer Cell Immunoglobulin-like Receptor 3DL2 Binding to HLA-B27 Licenses Pathogenic T Cell Differentiation in Spondyloarthritis. *Arthritis & Rheumatology*. 2016; 68: 901–914.
- [23] Ecoeur F, Weiss J, Kaupmann K, Hintermann S, Orain D, Guntermann C. Antagonizing Retinoic Acid-Related-Orphan Receptor Gamma Activity Blocks the T Helper 17/Interleukin-17 Pathway Leading to Attenuated Pro-inflammatory Human Keratinocyte and Skin Responses. *Frontiers in Immunology*. 2019; 10: 577.
- [24] Zhu ZQ, Tang JS, Cao XJ. Transcriptome network analysis reveals potential candidate genes for ankylosing spondylitis. *European Review for Medical and Pharmacological Sciences*. 2013; 17: 3178–3185.
- [25] Proost P, Struyf S, Loos T, Gouwy M, Schutyser E, Conings R, *et al.* Coexpression and interaction of CXCL10 and CD26 in mesenchymal cells by synergising inflammatory cytokines: CXCL8 and CXCL10 are discriminative markers for autoimmune arthropathies. *Arthritis Research & Therapy*. 2006; 8: R107.
- [26] Wang J, Zhao Q, Wang G, Yang C, Xu Y, Li Y, *et al.* Circulating levels of Th1 and Th2 chemokines in patients with ankylosing spondylitis. *Cytokine*. 2016; 81: 10–14.
- [27] Xu ZY, Zhou C, Zhang KF, Zheng YP. Identification of key genes in Ankylosing spondylitis. *Immunology Letters*. 2018; 204: 60–66.
- [28] Ma H, Xu D, Fu Q. Identification of ankylosing spondylitis-associated genes by expression profiling. *International Journal of Molecular Medicine*. 2012; 30: 693–696.
- [29] Kanwal A, Fazal S. Construction and analysis of protein-protein interaction network correlated with ankylosing spondylitis. *Gene*. 2018; 638: 41–51.
- [30] Duan R, Leo P, Bradbury L, Brown MA, Thomas G. Gene expression profiling reveals a downregulation in immune-associated genes in patients with AS. *Annals of the Rheumatic Diseases*. 2010; 69: 1724–1729.
- [31] Harcharik S, Bernardo S, Moskalenko M, Pan M, Sivendran M, Bell H, *et al.* Defining the role of CD2 in disease progression and overall survival among patients with completely resected stage-II to -III cutaneous melanoma. *Journal of the American Academy of Dermatology*. 2014; 70: 1036–1044.
- [32] Wolf G, Stahl RAK. CD2-associated protein and glomerular disease. *Lancet*. 2003; 362: 1746–1748.
- [33] Kim JM, Wu H, Green G, Winkler CA, Kopp JB, Miner JH, *et al.* CD2-associated protein haploinsufficiency is linked to glomerular disease susceptibility. *Science*. 2003; 300: 1298–1300.
- [34] Anderson JJ, Baron G, van der Heijde D, Felson DT, Dougados M. Ankylosing spondylitis assessment group preliminary definition of short-term improvement in ankylosing spondylitis. *Arthritis and Rheumatism*. 2001; 44: 1876–1886.

- [35] Bailey ST, Westerling T, Brown M. Loss of estrogen-regulated microRNA expression increases HER2 signaling and is prognostic of poor outcome in luminal breast cancer. *Cancer Research*. 2015; 75: 436–445.
- [36] Knowles MA, Hurst CD. Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nature Reviews. Cancer*. 2015; 15: 25–41.
- [37] Liu X, Wu J, Zhang D, Bing Z, Tian J, Ni M, *et al*. Identification of Potential Key Genes Associated With the Pathogenesis and Prognosis of Gastric Cancer Based on Integrated Bioinformatics Analysis. *Frontiers in Genetics*. 2018; 9: 265.