Application of Weighted Gene Co-expression Network Analysis to Identify Key Modules and Hub Genes in Systemic Juvenile Idiopathic Arthritis

Mi Zhou
Shanghai Jiao Tong University School of Medicine https://orcid.org/0000-0001-6183-738X

Ruru Guo
Department of Rheumatology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University

Yongfei Wang
Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong

Wanling Yang
Department of Paediatrics and Adolescent Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong

Rongxiu Li
State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dong Chuan Road, Shanghai 200240, China.

Liangjing Lu (lu_liangjing@163.com)
Department of Rheumatology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, 145 Middle of Shandong Road, Huangpu District, Shanghai, 200001, People's Republic of China.

Research

Keywords: sJIA, WGCNA, microarray, enrichment analysis, erythrocyte differentiation, natural killer cell

Posted Date: March 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-254358/v1

Abstract

Background: Systemic juvenile idiopathic arthritis (sJIA) is a severe autoinflammatory disorder whose molecular mechanism is still not clearly defined. To better understand the disease using scattered datasets from public domains, we performed a weighted gene co-expression network analysis (WGCNA) to identify key modules and hub genes underlying sJIA pathogenesis.

Methods: Two gene expression datasets, GSE7753 and GSE13501, were used to construct WGCNA. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were applied to the entirety of genes and the hub genes in the sJIA modules. Cytoscape was used to screen and visualize the hub genes. We further compared the hub genes with the GWAS genes and used a consensus WGCNA analysis to prove that our conclusions are conservative and reproducible across multiple independent data sets.

Results: A total of 5414 genes were obtained for WGCNA, from which highly correlated genes were divided into 17 modules. The red module demonstrated the highest correlation with the sJIA module ($r=0.8$, $p=3e-29$), while the green-yellow module was found to be closely related to the non-sJIA module ($r=0.62$, $p=1e-14$). Functional enrichment analysis demonstrated that the red module was largely enriched in activation of immune responses, infection, nucleosome and erythrocyte, the green-yellow module was mostly enriched in immune responses and inflammation. Additionally, the hub genes in the red module were highly enriched in erythrocyte differentiation, including ALAS2, AHSP, TRIM10, TRIM58 and KLF1. The hub genes from the green-yellow module were mainly associated with immune responses, exemplified by genes such as KLRB1, KLRF1, CD160, KIRs etc.

Conclusion: We identified sJIA-related modules and several hub genes that might be associated with the development of sJIA. The two modules may help understand the mechanisms of sJIA and the hub genes may become biomarkers and therapeutic targets of sJIA in the future.

1. Introduction

Systemic juvenile idiopathic arthritis (sJIA) is a serious immune-inflammatory pediatric disorder, with significant differences from other JIA subtypes in many aspects. Symptoms for patients with sJIA range from fever, rash, serositis, lymphadenectasis, liver and/or spleen enlargement[1], to more severe ones such as macrophage activation syndrome (MAS) that is potentially life-threatening. It is estimated that about 10% of sJIA patients will develop overt MAS, and more than 50% of the patients may have subclinical MAS[2, 3]. sJIA shares many characteristics with classic autoinflammatory diseases, most notably the response to IL-1 inhibition[4]. It is associated with other inflammatory cytokines such as IL-6 and IL-18, and S100 alarm protein[5]. Besides, NK cell dysfunction has been suggested to represent a common pathway in patients with sJIA, MAS, and HLH (hemophagocytic lymphohistiocytosis, closely resembles MAS)[6, 7]. However, there are still many questions about the molecular basis of sJIA immune

dysfunction and the etiology of sJIA and MAS[8], as many genetic and genomic investigations on sJIA are limited by small samples studied.

Weighted gene co-expression network analysis (WGCNA) is a network-based approach focusing on sets of genes instead of individual genes from gene expression data. By transforming data of gene expression into co-expression modules, WGCNA provides insights into key genes and signaling networks that could play critical roles in the progression of diseases[9–11]. It has extensively used in biological research, such as cancer[12], COPD (chronic obstructive pulmonary disease)[13], neuropsychiatric disorder[14], and it is a powerful tool for screening candidate biomarkers or therapeutic targets. In this study, based on the integrated microarray datasets, we used the WGCNA method to identify sJIA-related and non-sJIA co-expression modules and analyzed the hub genes in the modules. The biological function and pathway of the two modules were also identified and analyzed. This is the first study applying WGCNA method on multiple datasets to understand the molecular mechanism of sJIA.

2. Materials And Methods

2.1. Data Collection

The gene profiles were downloaded from GEO database according to the following keywords: (1) Systemic juvenile idiopathic arthritis, (2) Homo sapiens, (3) Peripheral blood tissue, and (4) Absence of drug intervention. Finally, we confirmed these two datasets: GSE7753[15] and GSE13501[16] by the same Platforms Affymetrix Human Genome U133 Plus 2.0 Array (GPL570). The GSE7753 contained 17 sJIA samples and 30 normal samples, and the GSE13501 included 21 sJIA samples and 59 normal samples. The raw data of GSE7753[15] and GSE13501[16] were downloaded from the GEO database. Altogether 127 samples (38 sJIA and 89 healthy controls) were used in the analysis. The overall research process is presented in Fig. 1.

The affy package (R environment, version 3.6.1) was used to normalize (RMA normalization) and preprocess the raw data[17]. The parameters were set as RMA (for background correction) and impute (for supplemental missing value). The expression profiles were log2 transformed, and batch normalization was done using “sva” and “combat” functions in SVA R package[18], in order to avoid a possible bias of the two separate microarray datasets. Probes with more than one gene were eliminated, and the max value was selected from these probes after the probe annotation. A series matrix file was preprocessed to identify differentially expressed genes based on variance analysis, and the top 25%[19] (5414 genes) were obtained for subsequent analysis.

2.2. Construction of Weighted Co-expression Network

The “WGCNA ” package in R software was used for the network construction[11]. The expression values of the 5414 genes were imported into WGCNA to construct co-expression modules by the automatic network construction. By the pickSoftThreshold function calculating the scale-free topology fit index for 1 to 20 powers, soft threshold power six was chosen as the most appropriate one for network construction.

Then we perform automatic block-wise module detection by the function blockwiseModules. The function first pre-clusters nodes into large clusters. Then hierarchical clustering is applied to each block, and modules are defined as branches of the resulting dendrogram. An automatic module merging step is performed to merges modules whose eigengenes are highly correlated (maxBlockSize = 6000, TOMType = 'unsigned', minModuleSize = 40, mergeCutHeight = 0.25). Thus the genes with similar expression profiles were separated into the same module.

2.3. Construction of Consensus Weighted Co-expression Network

In order to verify the reliability and the stability of the previous results and the module, we applied the consensus WGCNA method. The GSE7753 and the GSE13501 were named as sJIA1 datasets and sJIA2 datasets for subsequent analysis. It is also important to choose the soft thresholding power β to construct a consensus weighted gene network. Approximate scale-free topology is attained around the soft-thresholding power of 6 for both sets. We chose the soft thresholding power 6, minimum module size 40, cut height for merging of modules 0.25 (implying that modules whose eigengenes are correlated above 1 − 0.25 = 0.75 will be merged), and nineteen distinct gene co-expression modules were constructed and shown in different colors. We also related consensus modules to external microarray sample information, sJIA patients and healthy people. Besides, we compared the correspondence among individual dataset modules, merged dataset modules, and consensus modules.

2.4. Identification of Co-expression Modules related to Normal or sJIA

The associations between the module and trait were estimated by the correlation between the module eigengene and the clinic traits, namely normal or sJIA. Here gene significance (GS) is defined as the (the absolute value of) the correlation between the gene and the trait, and module membership (MM) represent the correlation of the genes with each module eigengene and clinical feature, respectively. Furthermore, module importance (MS) is defined as the correlation of the module eigengene and the gene expression profile. Among all co-expression modules, the module with the absolute MS ranking first is regarded as a module related to clinical traits (normal or sJIA). The gene modules of the highest correlation of normal and sJIA were chosen for subsequent studies.

2.5. Function enrichment analysis

Then functional enrichment analysis was performed on the genes in the sJIA module and non-sJIA module. The module genes' information was submitted to DAVID (Database for Annotation, Visualization, and Integrated Discovery) online tool[20] to perform functional annotation based on Gene Ontology (GO) [21] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. We used the packages anRichment and anRichmentMethods to do GOenrichment analysis in the whole modules (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/GeneAnnotation/Tutorials/). The packages were written by the writer of WGCNA, which helped evaluates the enrichment of the gene modules in the collection of GO terms (threshold = 1e-4, thresholdType = "Bonferroni"), and select the top GO results in each module to draw a bar graph (Supplementary_Material 1).

2.6. Identification of Hub Genes

Hub genes have been considered as functionally significant because of their high connectivity with other genes within a module. In this study, the 30 top genes with the highest levels of intramodular connectivity in the two modules were selected as the candidates for further analysis by DAVID and visualized by Cytoscape. Subsequently, we utilized the GWAS catalogue(https://www.ebi.ac.uk/gwas/studies/GCST004025) to obtain the disease susceptibility genes identified by a previously published GWAS[22]. The protein-protein interaction(PPI) network of the module hub genes and genome-wide-associated genes were analyzed by STRING database (confidence score ≥ 0.4) and was visualized using Cytoscape. The comparative analyses of the functional enrichment among the module hub genes and genome-wide-associated genes were performed by the online bioinformatics database Metascape (http://metascape.org/gp/index.html#/main/step1)[23].

3. Results

3.1. Construction of weighted co-expression network

After the data preprocessing, a total of 5414 genes were selected for WGCNA(Supplementary Figure S1). First, it's an essential step to choose the appropriate soft thresholding power to ensure a scale-free network. By the network topology analysis function, the power value 6 was selected (Supplementary Figure S2) and used to construct the co-expression module. Seventeen distinct gene co-expression modules were constructed and shown in different colors (Fig. 2A). The number of genes in the 17 modules was shown in Supplementary Table 1. Figure 2B demonstrated the topological overlap matrix (TOM) of the 5,414 genes, indicating that each module and the genes expression in each module were relatively independent. Furthermore, we also plotted the clustering dendrogram, according to the module correlation and the heatmap according to adjacencies (Fig. 2C), meaning that those modules were largely divided into two clusters.

3.2. Construction of Consensus Weighted Co-expression Network

Since the overall connectivity index generally drops sharply with the increase of the soft-thresholding power, it is advantageous to select the lowest power that meets the approximate scale-free topology standard. Supplementary Figure S3 shows that approximate scale-free topology is obtained around the soft-thresholding power of 6 for both sets.

Nineteen different gene co-expression modules were constructed, shown in different colors in SSS 1C, and related to external microarray sample information, sJIA patients and healthy people. In each of the two sets, consensus module eigengenes would be related to the traits, respectively. To summarizes the two sets into one measure of module-trait relationships, we took the correlation that had the lower absolute value in the two sets if the two correlations had the same sign, and zero relationship if the two correlations had opposite signs (Fig. 3A). We checked these genes in the modules related to clinical features, which were basically in line with our previous results

Then we compared the correspondence among individual dataset modules, merged dataset modules, and consensus modules. Figure 3B1-B6 demonstrates the two data sets are indeed very similar. The preservation heatmap and barplots indicate that most relationships are very highly preserved and the overall preservation of the two eigengene networks is 0.89. Figure 3B7 showed the number of genes overlapping between the merged dataset modules (our previous method) and consensus modules was very high, and the hub genes obtained in the previous WGCNA modules were all in overlapping genes.

3.3. Identification of Co-expression Modules related to Normal or sJIA

The module-trait correlation coefficients in Fig. 2D illuminated that the red module and the green-yellow module revealed a high correlation with disease status. The red module was best positively correlated with the sJIA-related module ($r = 0.8$, $p = 3e - 29$), while the green-yellow module was negatively related to the sJIA ($r = 0.62$, $p = 1e - 14$). The scatterplots in Fig. 4A and 4B also showed that the gene significance (GS) and module membership (MM) value were of high correlation in the red module (cor = 0.85, $p = 8.8e - 86$) and the green-yellow module (cor = -0.59, $p = 5.3e - 16$). The results suggested that the genes in the two modules were probably related to the disease status, which were suitable for further analyses and mining of the hub gene.

3.4. Function enrichment analysis

Function enrichment analysis conducted by DAVID was performed on the genes in the two constructed modules. There was a significant difference in the biological processes of genes in the sJIA and non-sJIA modules. The detailed information was displayed in Fig. 4G and 4H.

For the red module, GO biological process (BP) annotation showed that the gene products mainly enriched in activation of immune response, infection, nucleosome and erythrocyte. As to GO molecular function (MF) annotation, protein heterodimerization and oxygen transporter were the most enriched terms. And enriched GO-CC terms were mainly involved in extracellular exosome, nucleosome, hemoglobin complex, and extracellular space. The results of KEGG enrichment showed the module was similar to Systemic lupus erythematosus (gene count = 11, $p = 3.1e - 5$). For the green-yellow module, the GO-BP annotation was mainly enriched in the immune response and inflammation. And receptor activity was the top enriched GO-MF terms, with plasma membrane enriched in GO-CC terms. Similarly, the KEGG terms were mainly about “Antigen processing and presentation” and “Natural killer cell mediated cytotoxicity” (Fig. 4C)..

We also used the packages anRichment and anRichmentMethods to do GO enrichment analysis in the whole modules and select the top GO term in each module to draw a bar graph (Fig. 5A& Fig. 5F). We further analyzed the functional enrichment of genes in several other relatively important modules: yellow, salmon, purple, cyan. As the Fig. 5B-E showed, the cyan module is mainly related to response to external stimulus, the purple module is mostly about the function of platelet alpha granule, involving some pathways such as wound healing, coagulation, and hemostasis etc, the salmon module may play an important role in the cell cycle process, and the yellow module is associated to transcriptional regulation.

3.5. Identification of Hub Genes

The 30 top-ranked hub genes in the two modules were shown in Cytoscape (Fig. 4D and 4E). By enrichment analysis shown in Fig. 4F, the hub genes from the red module were largely related to erythrocyte differentiation (*ALAS2, AHSP, KLF1, TRIM10, TRIM58*), and the hub genes from the green-yellow module were largely involved to immune responses, exemplified by genes such as *KLRB1, KLRF1, CD160, KIR2DL1, KIR2DL2, KIR2DL3, KIR3DL1, SH2D1B, GZMA, TGFBR3*.

Only one GWA study was previously conducted in sJIA[22], and we obtain the disease susceptibility genes from this manuscript and the GWAS catalogue. In order to further compare the hub genes with the GWAS genes, we adopted the PPI protein network analysis and functional enrichment. The Fig. 6A showed, the sJIA-susceptible genes (*HLA-DRA, TRIM58, LDB2, TAPT1*) may be related to the hub genes from the red module, with *TRIM58* was also a hub gene of the red module, and those (*LGMN, JPH3*) may be related to the hub genes from the greenyellow module. The Fig. 6B-C demonstrates that the comparative analyses of the functional enrichment, such as tissue morphogenesis related to GWAS genes and the hub gene s of the red module, endothelial cell migration associated to GWAS genes and the hub genes of the greenyellow module.

4. Discussion

Bioinformatic analysis of public gene expression data could shed new light on the pathogenesis sJIA. Our study is the first to integrate multiple datasets (GSE7753 and GSE13501) and construct WGCNA to identify the hub gene that may play an important role in sJIA. Among a total of 17 co-expression modules, the red module was positively related to sJIA, and the green-yellow was negatively related to sJIA. Moreover, we gained several hub genes related to the pathogenesis of sJIA. As the genes in the same module were considered to have similar functions, the analysis of biologically-relevant modules and hub genes may present new insights into the molecular mechanism of sJIA development. The functions of these two modules and their hub genes were as follows:

The red module was critical in the biological process and pathways such as antibacterial humoral response, innate immune response in mucosa, nucleosome assembly, defense response to Gram-positive bacterium, and erythrocyte differentiation. However, the functional enrichment of the top 30 hub genes in the red module was largely related to erythrocyte differentiation (*ALAS2, AHSP, TRIM10, TRIM58, KLF1*). In accordance with the present results, previous studies[24] have demonstrated that there were strong relations between the EDS (erythroid differentiation signature) and sJIA associated with the expansion of $CD34^+$ cells. The presence of the EDS was also seen in FHLH, infection, and PAH, which suggested that the increased recruitment of red blood cell might be a part of the systemic response to severe chronic local hypoxia[24, 25].

ALAS2 (Erythroid-specific 5-aminolevulinate synthase) is the first and rate-limiting enzyme in the erythroid heme biosynthetic pathway[26]. Mutations in *ALAS2* may be related to Porphyria and X-linked

sideroblastic anemia[27]. *AHSP* (Alpha hemoglobin–stabilizing protein) is also necessary for the proper assembly of nascent alpha-globin into hemoglobin-A [28]. And the altered expression or the function of *AHSP* might relate to the disease severity of thalassemia[29]. Recent research by Lechauve C et al also predicts that *AHSP* plays an important role in the physiological process of regulating vascular NO concentration[30]. *KLF1*(Erythroid Kruppel-like factor 1) is important in the function of the erythroid cell, such as red cell membrane stability and heme biosynthesis[31]. And *AHSP is also* the known *KLF1* target gene, the expression of which was significantly upregulated upon *KLF1* activation[31]. Further studies are needed to confirm and validate the function of these EDS genes (*ALAS2, AHSP, KLF1*) in the occurrence and development of diseases.

TRIM family proteins play an essential and unique role in several diseases classified like immunological disease, cancers and developmental disorders,and they may function as dual regulators of the immune response and carcinogenesis[32]. *TRIM10* was reported to participate in terminal red blood cell differentiation and survival[33]. However, recent research showed that *TRIM10* is related to share a genetic basis for Parkinson's Disease (PD) and other autoimmune diseases[34]. Silencing of TRIM10 can reduce apoptosis and reactive oxygen species levels in a cellular model of PD, which suggests a potential role of *TRIM10* in PD and other autoimmune diseases. An earlier study also revealed the role of *TRIM58* in the regulation of human erythrocyte traits[35]. Recent studies reported that *TRIM58* regulates EMT via the Wnt/β-catenin pathway[36]and may function as a tumor suppressor in some tumor cancers, such as colorectal cancer[36] and gastric cancer[37]. Another research[38] showed that *TRIM58* might protect against transduction of intestinal mucosal inflammation by inhibiting abnormal TLR2 signaling and might become a new therapeutic target in autoimmune diseases, such as ulcerative colitis. Furthermore, *TRIM58* was identified as a sJIA susceptibility gene by the previous GWAS in sJIA[22].

For the yellow-green module, function enrichment analysis mainly involves immune response and inflammation, and the results of the hub gene are similar. The hub genes from the green-yellow module were largely related to immune responses, exemplified by genes such as *KLRB1, KLRF1, CD160, KIR2DL1, KIR2DL2, KIR2DL3, KIR3DL1, SH2D1B, GZMA, TGFBR3*,which is in line with previous studies that NK cell dysfunction may be a common pathway in sJIA, MAS and HLH [6, 7]. Besides, inflammatory driver factors may be involved in the cytotoxic defects of NK cells in MAS and sJIA[7].

KLRB1(killer cell lectin-like receptor subfamily B member 1) is usually referred to as CD161 that is a type II transmembrane C-type lectin glycoprotein that appears to act as an inhibitory role in IFN-γ secretion[39] and human NK cells[40], while the function in T cells is not so clear, with reports suggesting both coactivating[39]and inhibitory[41] effects. *KLRB1* has been previously shown to be downregulated in rheumatoid arthritis[42] and SLE[43–45].

KLRF1 (killer cell lectin-like receptor F1) is an activated homodimeric C-type lectin-like receptor (CTLR) expressed on almost all NK cells, marking a critical step in human NK cells development[46]and stimulates cytotoxicity and cytokine release of the NK cells[47]. CD160 (a 27 kDa glycoprotein) tightly binds to peripheral blood NK cells and CD8 T lymphocytes and has cytolytic effect activity[48]. *KIRs*

(Killer cell immunoglobulin-like receptors) play the most important role in NK activation. *KIR2DL1, KIR2DL2, KIR2DL3, KIR3DL1* belong to the inhibition receptors in the *KIR* family[49]. However, a previous study [50] showed that sJIA, compared with poly and pauciarticular JIA, was related to the decreased NK cell function, with more IFN-γ, less TNF-α secretion of NK cell and lower *KIR2DS4* frequency. Further and larger studies about KIR gene are necessary.

The obvious advantage of our research is that this is the first study to apply WGCNA method to understand the molecular mechanism of sJIA. Due to the rarity of the disease, we are unable to obtain larger samples, and we have done our best to obtain available data. And we not only apply to merge multiple data sets, but also use a consensus WGCNA analysis to prove that our conclusions are conservative and reliable in multiple data sets. Previous research on the disease was mainly on blood leukocyte, such as the very innovative study by Cepika et al[51], which integrated the blood leukocyte responses to innate stimuli from multiple omics, and determined the gene set related to specific cytokine environment and activated leukocyte subsets in sJIA. However, our study found that the relatively novel mechanism of sJIA in red blood cell differentiation (Hinze et al., 2010) and NK cell disorder (Grom, 2004; Vandenhaute et al., 2019). Furthermore, by linking the susceptibility genes with the module-associated hub genes, we improved the understanding of biological processes in sJIA, and identified *TRIM58* both as an sJIA susceptibility gene and as a hub gene of the red module. There is still a big gap in the occurrence and development of sJIA disease. Therefore, our research may help to investigate the progress of sJIA, and the hub genes may become biomarkers and therapeutic targets of sJIA in the future.

5. Conclusion

To conclude, sJIA-associated key genes were identified, such as *ALAS2, AHSP, TRIM10, TRIM58, KLF1*, which were largely related to erythrocyte differentiation. Those genes might be related to anemia or MAS in sJIA. *KLRB1, KLRF1, CD160, KIRs etc* might be related to NK cell dysfunction, which was studied extensively but remains incompletely understood in the pathogenesis of sJIA. Our study may help to investigate the progress of sJIA, and the hub genes may become biomarkers and therapeutic targets of sJIA in the future.

6. Abbreviations List

sJIA: Systemic juvenile idiopathic arthritis; WGCNA: weighted gene co-expression network analysis; MAS: macrophage activation syndrome; HLH: hemophagocytic lymphohistiocytosis; COPD: chronic obstructive pulmonary disease; GS: gene significance; MM: module membership; MS: module importance; DAVID: Database for Annotation, Visualization, and Integrated Discovery; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: biological process; MF: molecular function; PPI: protein-protein interaction; *ALAS2*: Erythroid-specific 5-aminolevulinate synthase; *AHSP*: Alpha hemoglobin–stabilizing protein; *KLF1*: Kruppel-like factor 1; *KLRB1*: killer cell lectin-like receptor subfamily B member 1; *KLRF1: killer cell lectin-like receptor F1; KIRs:* Killer cell immunoglobulin-like receptors

7. Declarations

Acknowledgements

Not applicable.

Author contributions

LL and RL designed the experiments. MZ obtained data from GEO. YW, WY analyzed the data. MZ and RG wrote the manuscript. Finally, all authors read and approved the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2017YFC0909002), the National Natural Science Foundation of China(81974251), shanghai Municipal Education Commission—Gaofeng Clinical Medicine Grant Support (20152214), shanghai Hospital Development Center, joint research of new advanced technology project(SHDC12018106), the pilot project construction of clinical collaboration between Chinese traditional medicine and Western medicine in Shanghai (ZY(2018-2020)-FWTX-1009), the pilot project construction of clinical collaboration between Chinese traditional medicine and Western medicine in Shanghai (the pilot project construction of clinical collaboration between Chinese traditional medicine and Western medicine in Shanghai (ZY(2018-2020)-FWTX-1010), the National key research and development program of China (2017YFC0909002), the National Science & Technology Major Projects of China (Major Infectious Diseases 2017ZX10201301-003-004).

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be downloaded from here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7753 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13501.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, He X, Maldonado-Cocco J, Orozco-Alcala J, Prieur AM, et al: **International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001.** *J Rheumatol* 2004, **31:**390-392.
2. Minoia F, Davi S, Horne A, Demirkaya E, Bovis F, Li C, Lehmberg K, Weitzman S, Insalaco A, Wouters C, et al: **Clinical features, treatment, and outcome of macrophage activation syndrome complicating systemic juvenile idiopathic arthritis: a multinational, multicenter study of 362 patients.** *Arthritis Rheumatol* 2014, **66:**3160-3169.
3. Behrens EM, Beukelman T, Paessler M, Cron RQ: **Occult macrophage activation syndrome in patients with systemic juvenile idiopathic arthritis.** *J Rheumatol* 2007, **34:**1133-1138.
4. Mellins ED, Macaubas C, Grom AA: **Pathogenesis of systemic juvenile idiopathic arthritis: some answers, more questions.** *Nat Rev Rheumatol* 2011, **7:**416-426.
5. Shimizu M, Nakagishi Y, Yachie A: **Distinct subsets of patients with systemic juvenile idiopathic arthritis based on their cytokine profiles.** *Cytokine* 2013, **61:**345-348.
6. Grom AA: **Natural killer cell dysfunction: A common pathway in systemic-onset juvenile rheumatoid arthritis, macrophage activation syndrome, and hemophagocytic lymphohistiocytosis?** *Arthritis Rheum* 2004, **50:**689-698.
7. Vandenhaute J, Wouters CH, Matthys P: **Natural Killer Cells in Systemic Autoinflammatory Diseases: A Focus on Systemic Juvenile Idiopathic Arthritis and Macrophage Activation Syndrome.** *Front Immunol* 2019, **10:**3089.
8. Yasin S, Schulert GS: **Systemic juvenile idiopathic arthritis and macrophage activation syndrome: update on pathogenesis and treatment.** *Current Opinion in Rheumatology* 2018, **30:**514-520.
9. Stuart JM, Segal E, Koller D, Kim SK: **A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules.** *Science* 2003, **302:**249-255.
10. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Stat Appl Genet Mol Biol* 2005, **4:**Article17.
11. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9:**559.
12. Jardim-Perassi BV, Alexandre PA, Sonehara NM, de Paula-Junior R, Reis Júnior O, Fukumasu H, Chammas R, Coutinho LL, Zuccari DAPdC: **RNA-Seq transcriptome analysis shows anti-tumor actions of melatonin in a breast cancer xenograft model.** *Scientific Reports* 2019, **9:**966.
13. Morrow JD, Qiu W, Chhabra D, Rennard SI, Belloni P, Belousov A, Pillai SG, Hersh CP: **Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods.** *BMC Medical Genomics* 2015, **8:**1.
14. Huggett SB, Stallings MC: **Cocaine'omics: Genome-wide and transcriptome-wide analyses provide biological insight into cocaine use and dependence.** *Addiction Biology* 2020, **25:**e12719.
15. N F, M B, S T, L L, J O, NT I, BS G, T G, DD S, S T, et al: **Gene expression profiling of peripheral blood from patients with untreated new-onset systemic juvenile idiopathic arthritis reveals molecular**

heterogeneity that may predict macrophage activation syndrome. *Arthritis and rheumatism* 2007, **56:**3793-3804.

16. MG B, AA G, SD T, TA G, P P, L I, N F, DP S, CH H, BJ A, et al: **Subtype-specific peripheral blood gene expression profiles in recent-onset juvenile idiopathic arthritis.** *Arthritis and rheumatism* 2009, **60:**2102-2112.
17. L G, L C, BM B, RA I: **affy--analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics (Oxford, England)* 2004, **20:**307-315.
18. JT L, WE J, HS P, AE J, JD S: **The sva package for removing batch effects and other unwanted variation in high-throughput experiments.** *Bioinformatics (Oxford, England)* 2012, **28:**882-883.
19. H Z, Y H, L C, H L, X M: **Exploration and validation of the effects of robust co-expressed immune-related genes on immune infiltration patterns and prognosis in laryngeal cancer.** *International immunopharmacology* 2020, **85:**106622.
20. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4:**P3.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25-29.
22. Ombrello MJ, Arthur VL, Remmers EF, Hinks A, Tachmazidou I, Grom AA, Foell D, Martini A, Gattorno M, Özen S, et al: **Genetic architecture distinguishes systemic juvenile idiopathic arthritis from other forms of juvenile idiopathic arthritis: clinical and therapeutic implications.** *Ann Rheum Dis* 2017, **76:**906-913.
23. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK: **Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.** *Nat Commun* 2019, **10:**1523.
24. Hinze CH, Fall N, Thornton S, Mo JQ, Aronow BJ, Layh-Schmitt G, Griffin TA, Thompson SD, Colbert RA, Glass DN, et al: **Immature cell populations and an erythropoiesis gene-expression signature in systemic juvenile idiopathic arthritis: implications for pathogenesis.** *Arthritis Res Ther* 2010, **12:**R123.
25. Jaklitsch MT, Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, MacMahon H, Mulshine JL, Munden RF, Salgia R, et al: **The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups.** *J Thorac Cardiovasc Surg* 2012, **144:**33-38.
26. Bishop DF, Henderson AS, Astrin KH: **Human delta-aminolevulinate synthase: assignment of the housekeeping gene to 3p21 and the erythroid-specific gene to the X chromosome.** *Genomics* 1990, **7:**207-214.
27. Tchaikovskii V, Desnick RJ, Bishop DF: **Molecular expression, characterization and mechanism of ALAS2 gain-of-function mutants.** *Mol Med* 2019, **25:**4.
28. Yu X, Mollan TL, Butler A, Gow AJ, Olson JS, Weiss MJ: **Analysis of human alpha globin gene mutations that impair binding to the alpha hemoglobin stabilizing protein.** *Blood* 2009, **113:**5961-

5969.

29. Ray R, Kalantri SA, Bhattacharjee S, Biswas A, Shahab A, Biswas S, Bhattacharyya M: **Association of alpha hemoglobin-stabilizing protein (AHSP) gene mutation and disease severity among HbE-beta thalassemia patients.** *Ann Hematol* 2019, **98:**1827-1834.
30. Lechauve C, Butcher JT, Freiwan A, Biwer LA, Keith JM, Good ME, Ackerman H, Tillman HS, Kiger L, Isakson BE, Weiss MJ: **Endothelial cell alpha-globin and its molecular chaperone alpha-hemoglobin-stabilizing protein regulate arteriolar contractility.** *J Clin Invest* 2018, **128:**5073-5082.
31. Yang CT, Ma R, Axton RA, Jackson M, Taylor AH, Fidanza A, Marenah L, Frayne J, Mountford JC, Forrester LM: **Activation of KLF1 Enhances the Differentiation and Maturation of Red Blood Cells from Human Pluripotent Stem Cells.** *Stem Cells* 2017, **35:**886-897.
32. Hatakeyama S: **TRIM Family Proteins: Roles in Autophagy, Immunity, and Carcinogenesis.** *Trends Biochem Sci* 2017, **42:**297-311.
33. Blaybel R, Théoleyre O, Douablin A, Baklouti F: **Downregulation of the Spi-1/PU.1 oncogene induces the expression of TRIM10/HERF1, a key factor required for terminal erythroid cell differentiation and survival.** *Cell Research* 2008, **18:**834-845.
34. Witoelar A, Jansen IE, Wang Y, Desikan RS, Gibbs JR, Blauwendraat C, Thompson WK, Hernandez DG, Djurovic S, Schork AJ, et al: **Genome-wide Pleiotropy Between Parkinson Disease and Autoimmune Diseases.** *JAMA Neurology* 2017, **74:**780-792.
35. Thom CS, Traxler EA, Khandros E, Nickas JM, Zhou OY, Lazarus JE, Silva AP, Prabhu D, Yao Y, Aribeana C, et al: **Trim58 degrades Dynein and regulates terminal erythropoiesis.** *Dev Cell* 2014, **30:**688-700.
36. Liu M, Zhang X, Cai J, Li Y, Luo Q, Wu H, Yang Z, Wang L, Chen D: **Downregulation of TRIM58 expression is associated with a poor patient outcome and enhances colorectal cancer cell invasion.** *Oncol Rep* 2018, **40:**1251-1260.
37. Liu X, Long Z, Cai H, Yu S, Wu J: **TRIM58 suppresses the tumor growth in gastric cancer by inactivation of β-catenin signaling via ubiquitination.** *Cancer Biol Ther* 2020, **21:**203-212.
38. Eyking A, Ferber F, Köhler S, Reis H, Cario E: **TRIM58 Restrains Intestinal Mucosal Inflammation by Negatively Regulating TLR2 in Myeloid Cells.** *J Immunol* 2019, **203:**1636-1649.
39. Fergusson JR, Smith KE, Fleming VM, Rajoriya N, Newell EW, Simmons R, Marchi E, Björkander S, Kang YH, Swadling L, et al: **CD161 defines a transcriptional and functional phenotype across distinct human T cell lineages.** *Cell Rep* 2014, **9:**1075-1088.
40. Mesci A, Ljutic B, Makrigiannis AP, Carlyle JR: **NKR-P1 biology: from prototype to missing self.** *Immunol Res* 2006, **35:**13-26.
41. Le Bourhis L, Dusseaux M, Bohineust A, Bessoles S, Martin E, Premel V, Coré M, Sleurs D, Serriari NE, Treiner E, et al: **MAIT cells detect and efficiently lyse bacterially-infected epithelial cells.** *PLoS Pathog* 2013, **9:**e1003681.
42. Chalan P, Kroesen BJ, van der Geest KS, Huitema MG, Abdulahad WH, Bijzet J, Brouwer E, Boots AM: **Circulating CD4+CD161+ T lymphocytes are increased in seropositive arthralgia patients but**

decreased in patients with newly diagnosed rheumatoid arthritis. *PLoS One* 2013, **8:**e79370.

43. Villanueva E, Yalavarthi S, Berthier CC, Hodgin JB, Khandpur R, Lin AM, Rubin CJ, Zhao W, Olsen SH, Klinker M, et al: **Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus.** *J Immunol* 2011, **187:**538-552.
44. Lin YL, Lin SC: **Analysis of the CD161-expressing cell quantities and CD161 expression levels in peripheral blood natural killer and T cells of systemic lupus erythematosus patients.** *Clin Exp Med* 2017, **17:**101-109.
45. Crow MK, Wohlgemuth J: **Microarray analysis of gene expression in lupus.** *Arthritis Res Ther* 2003, **5:**279-287.
46. Freud AG, Keller KA, Scoville SD, Mundy-Bosse BL, Cheng S, Youssef Y, Hughes T, Zhang X, Mo X, Porcu P, et al: **NKp80 Defines a Critical Step during Human Natural Killer Cell Development.** *Cell Rep* 2016, **16:**379-391.
47. S K, S K, A K, G P, HG R, A S: **NKp80 defines and stimulates a reactive subset of CD8 T cells.** *Blood* 2009, **113:**358-369.
48. Liu W, Garrett SC, Fedorov EV, Ramagopal UA, Garforth SJ, Bonanno JB, Almo SC: **Structural Basis of CD160:HVEM Recognition.** *Structure* 2019, **27:**1286-1295.e1284.
49. Ashouri E, Dabbaghmanesh MH, Ranjbar Omrani G: **Presence of more activating KIR genes is associated with Hashimoto's thyroiditis.** *Endocrine* 2014, **46:**519-525.
50. Zhou J, Tang X, Ding Y, An Y, Zhao X: **Natural killer cell activity and frequency of killer cell immunoglobulin-like receptors in children with different forms of juvenile idiopathic arthritis.** *Pediatr Allergy Immunol* 2013, **24:**691-696.
51. Cepika AM, Banchereau R, Segura E, Ohouo M, Cantarel B, Goller K, Cantrell V, Ruchaud E, Gatewood E, Nguyen P, et al: **A multidimensional blood stimulation assay reveals immune alterations underlying systemic juvenile idiopathic arthritis.** *J Exp Med* 2017, **214:**3449-3466.

Figures

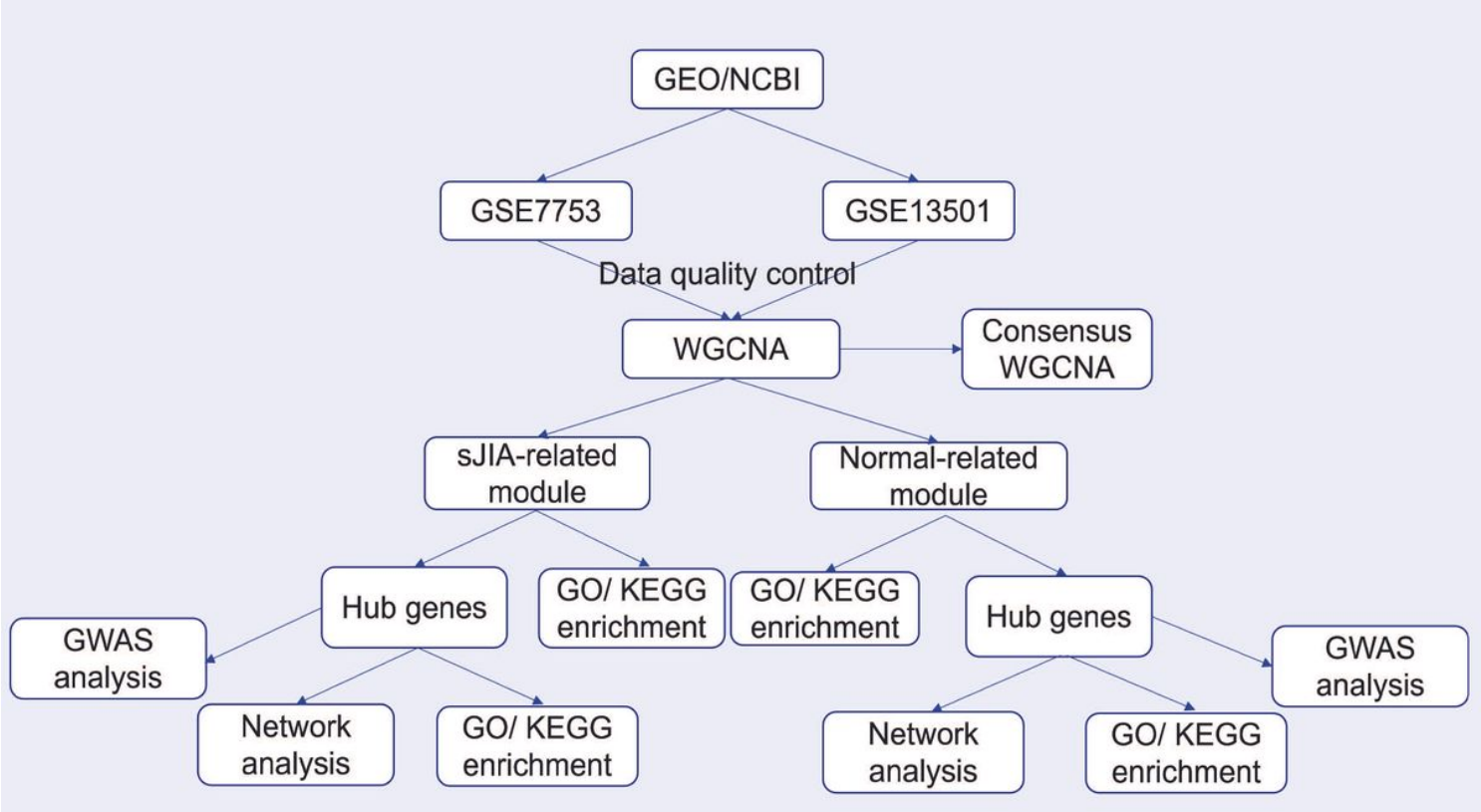

Figure 1

Flow chart of the whole procedures in this study.

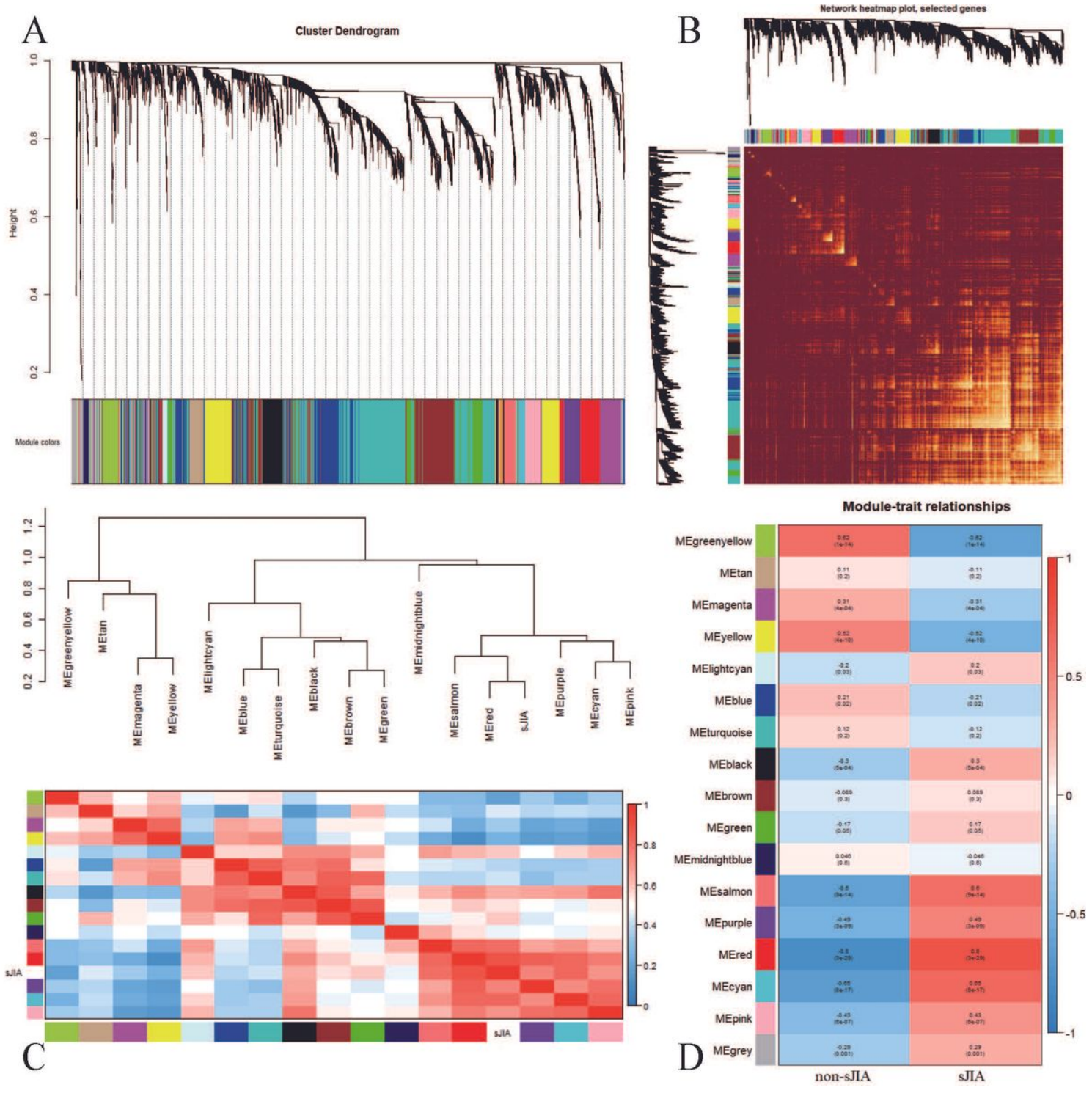

Figure 2

Division and validation of co-expression modules. (A). Dendrogram of all genes divided into 17 modules, with dissimilarity based on topological overlap, together with assigned module colors. The number of genes in each module were listed in Table1. (B). The heatmap depicts the Topological Overlap Matrix (TOM) among all genes in the analysis. The depth of the red color indicates the correlation between all pair-wise genes. (C). The upper part shows hierarchical clustering of the whole modules. Another is a

heatmap plot of the adjacencies in the hub gene network. (D). Heatmap of the correlation between module eigengenes and the Clinical Modules.

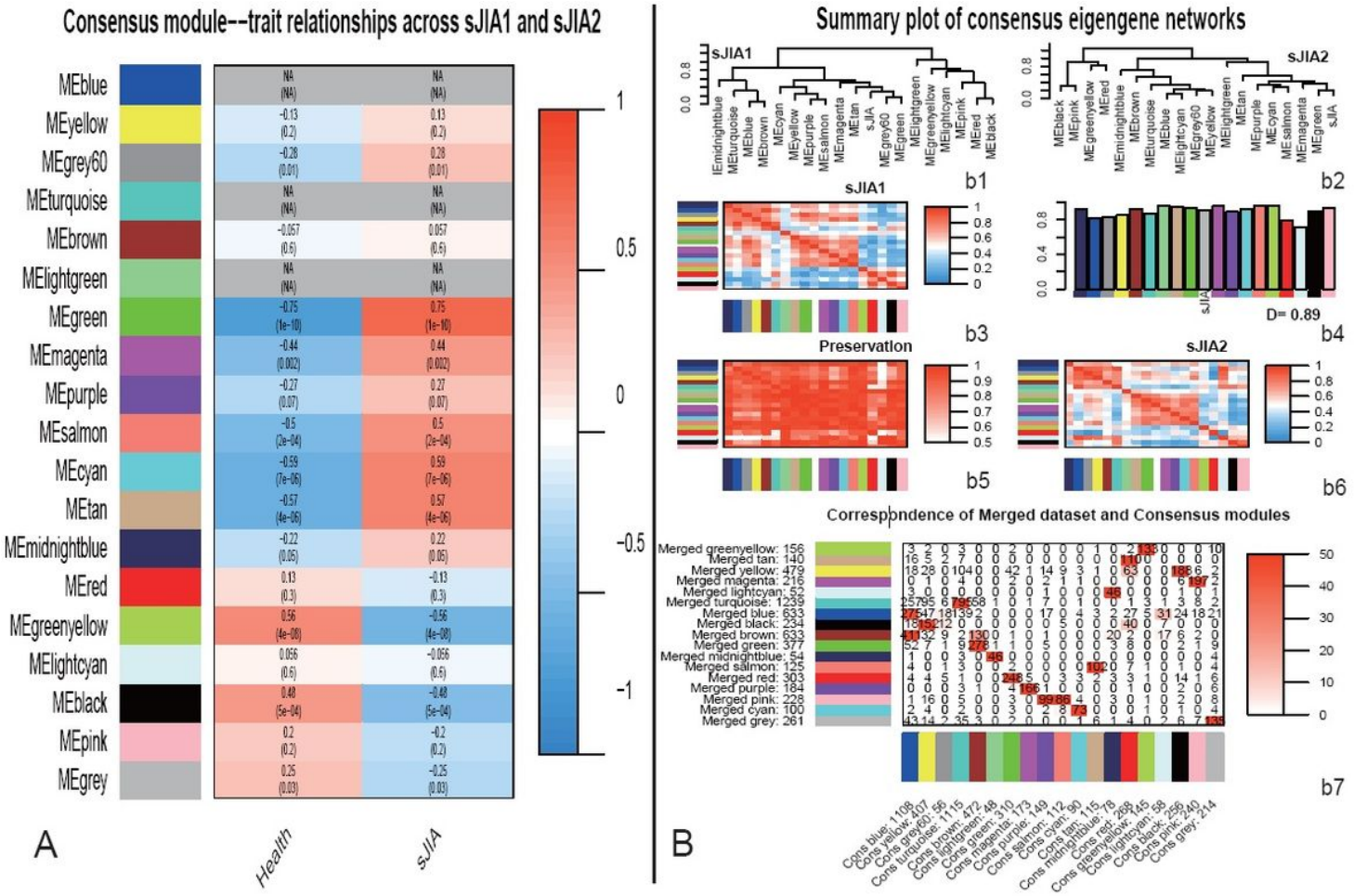

Figure 3

Correspondence among individual dataset modules, merged dataset modules, and consensus modules. (A). Heatmap of the correlation between module eigengenes and the Clinical Modules. Missing (NA) entries indicate that the correlations in the sJIA1 and sJIA2 data sets have opposite signs and no consensus can be formed. (B.b1-b6): Correspondence of the sJIA1 dataset modules and sJIA2 modules. The Preservation heatmap shows the preservation network, defined as one minus the absolute difference of the eigengene networks in the two data sets. The barplot shows the mean preservation of adjacency for each of the eigengenes to all other eigengenes. (B.b7): Correspondence of the merged dataset modules and consensus modules. Each row of the table corresponds to one sJIA2 set-specific module, and each column corresponds to one consensus module. Numbers in the table indicate gene counts in the intersection of the corresponding modules.

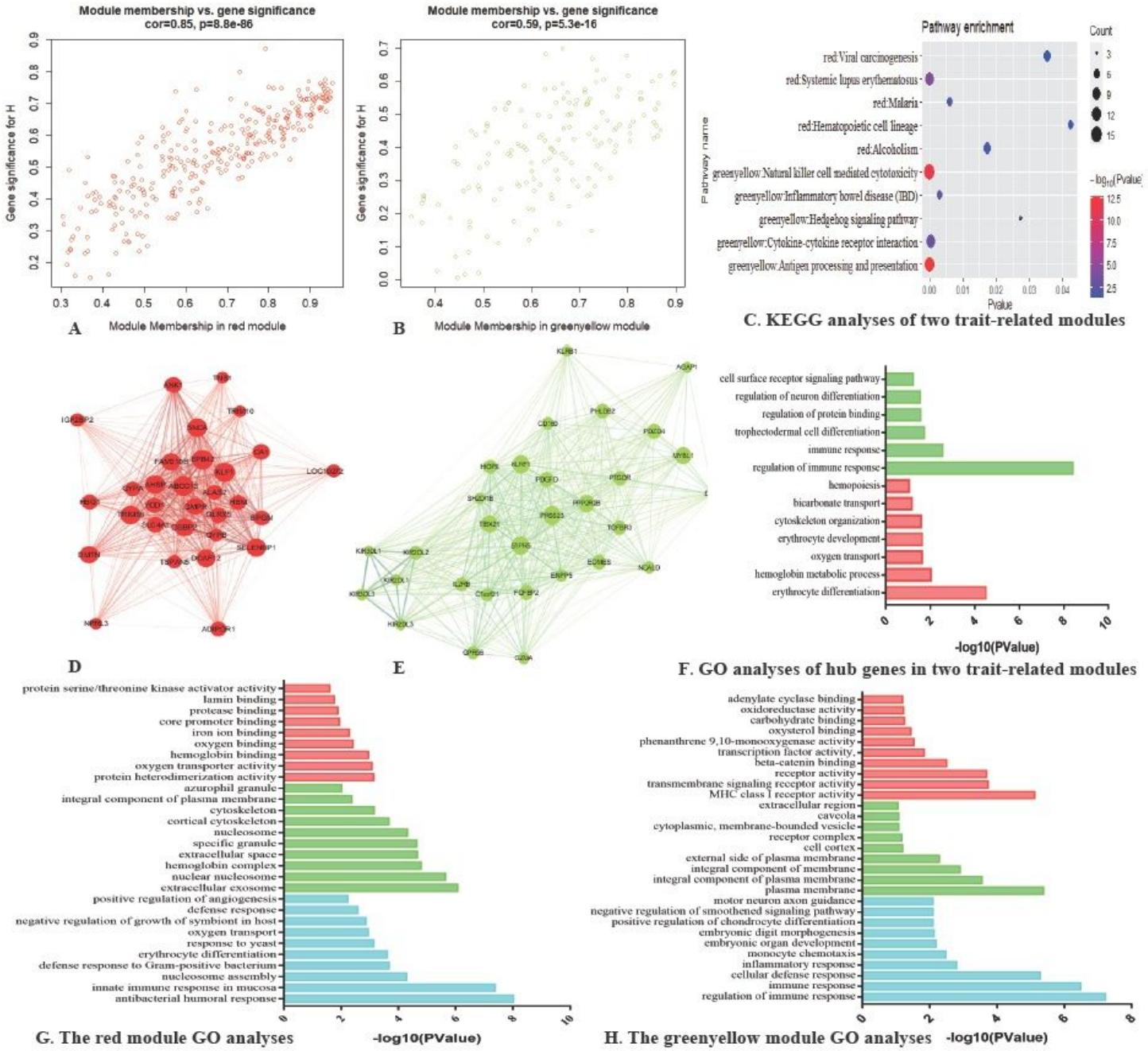

Figure 4

Functional enrichment analyses of two trait-related modules. (A) scatter plot of module eigengenes in the red module. (B) scatter plot of module eigengenes in the greenyellow module. (C) KEGG enrichment analyses of two trait-related modules. (D) The top 30 hub genes in the red module. (E) The top 30 hub genes in the greenyellow module. Nodes represent genes, and node size is correlated with connectivity of the gene. (F) GO enrichment analyses of the top 30 hub genes in the two trait-related modules. (G) GO enrichment analyses of the red module. (H) GO enrichment analyses of the greenyellow module.

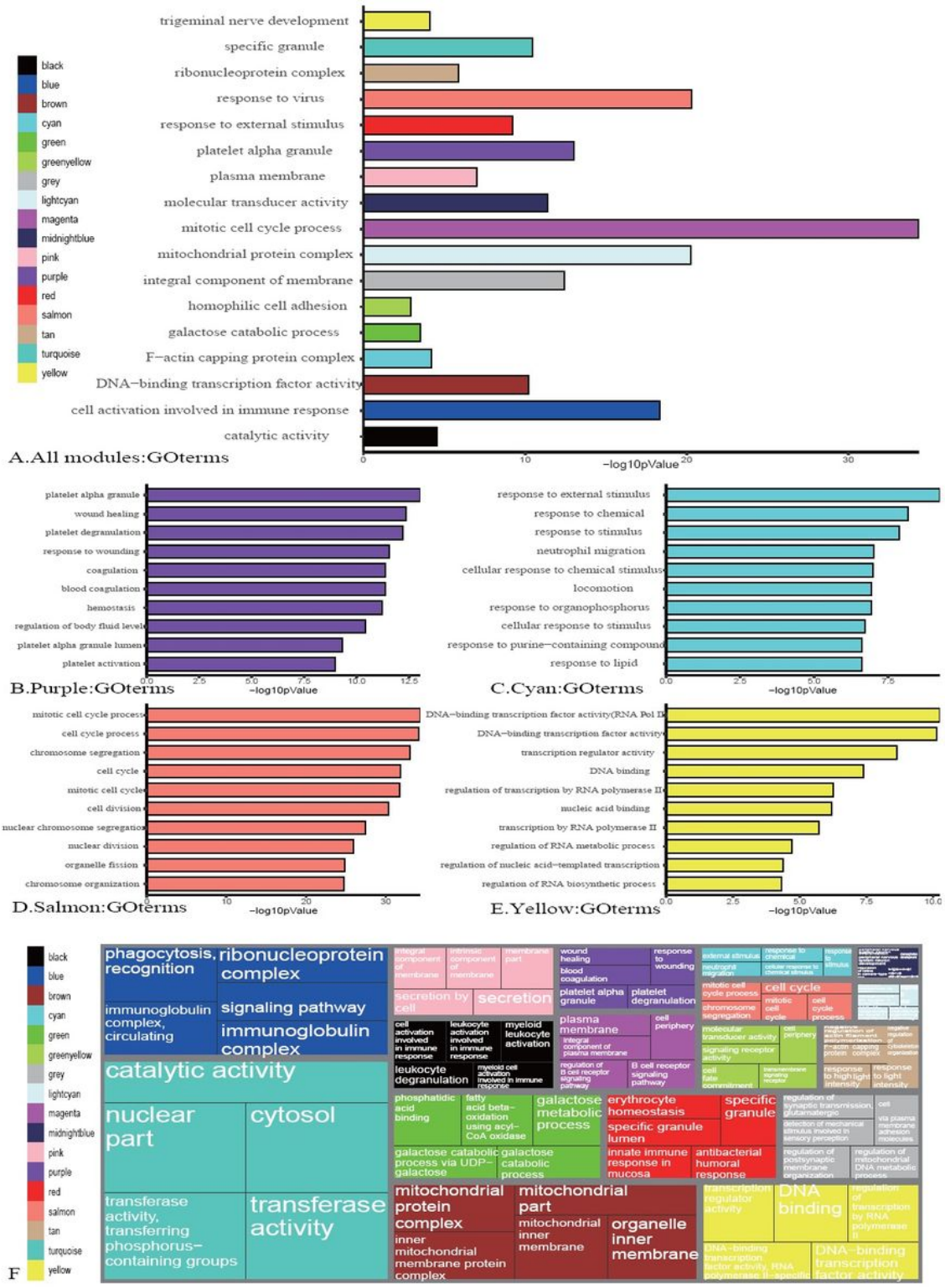

Figure 5

GO enrichment analysis of each module in WGCNA. (A) Barplot of the top GO terms of each module. The bars are marked with the corresponding module color, and the left side is the specific GO terms. (B-E) Barplot of GO enrichment analysis in other relatively important modules: yellow, salmon, purple, cyan. (F) Treemap of the top five GO terms of each module.

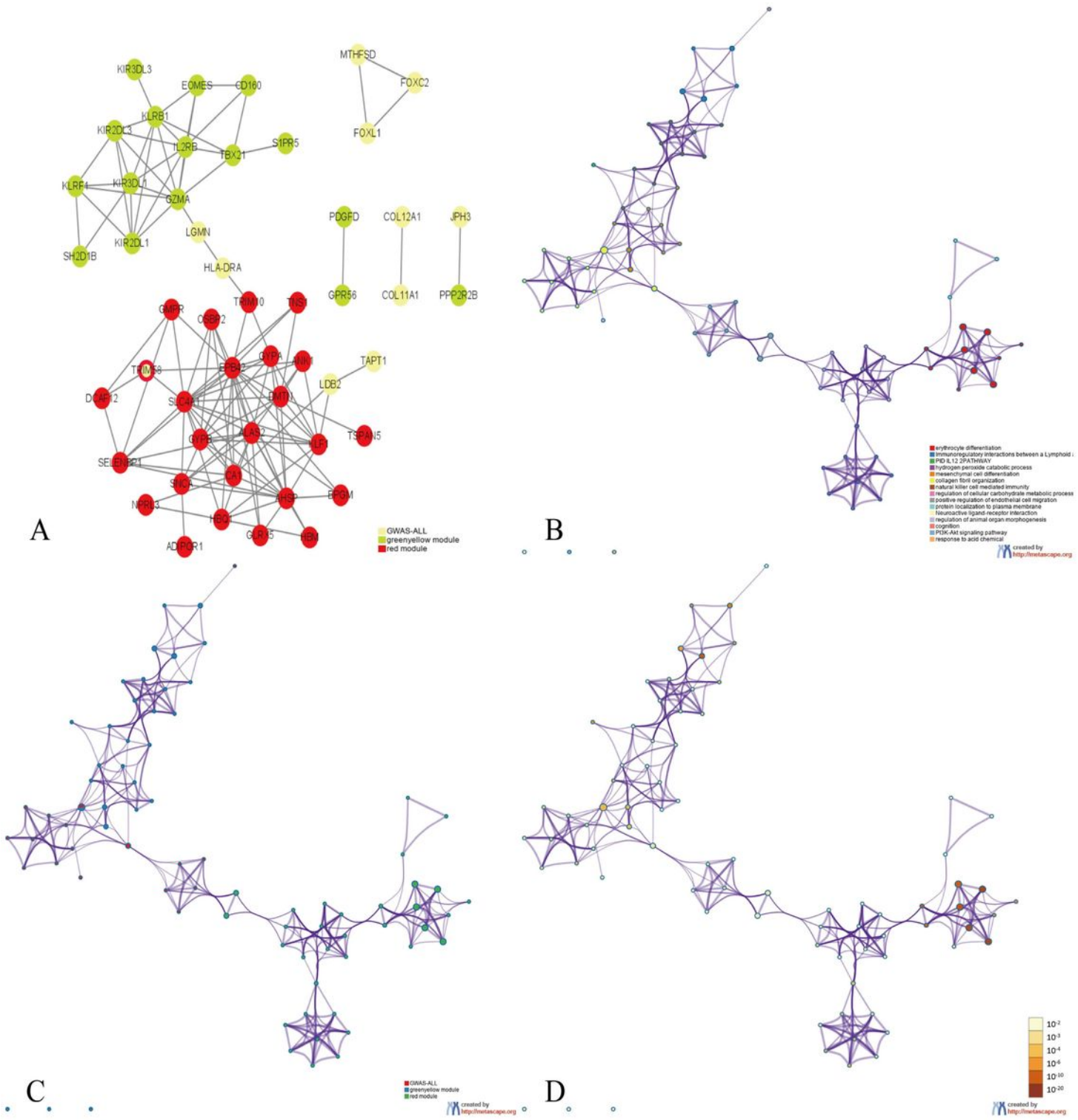

Figure 6

Interactions between module hub genes and genome-wide-associated genes. (A) Protein-Protein Interaction network of the module hub genes and genome-wide-associated genes. Common genes between GWAS and the “red” module are denoted in yellow nodes with red contour. (B) Enriched Ontology Clusters Colored by Cluster ID. Terms with a similarity score > 0.3 are linked by an edge (the thickness of the edge represents the similarity score). (C) Enriched Ontology Clusters Pied by Gene Counts Across

Studies. Each pie sector is proportional to the number of hits originated from a gene list. Color code for pie sector represents a gene list. (D) Enriched Ontology Clusters Colored by p-Value. The dark the color, the more statistically significant the node is.

Supplementary Files

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

- Supplementary1simplefigureandtable.docx
- Supplementary2complexexcel.xlsx
- Supplementary3codes.docx