CD36 identified by weighted gene co-expression network analysis as a hub candidate gene in lupus nephritis

Huiying Yang [1], **Hua Li** [Corresp. 1]

[1] Department of Nephrology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China

Corresponding Author: Hua Li
Email address: h_li@zju.edu.cn

Lupus nephritis (LN) is a severe manifestation of systemic lupus erythematosus (SLE), which often progresses to end-stage renal disease (ESRD) and ultimately leads to death. At present, there are no definitive therapies towards LN, so that illuminating the molecular mechanism behind the disease has become an urgent task for researchers. Bioinformatics has become a widely utilized method for exploring genes related to disease. This study set out to conduct weighted gene co-expression network analysis (WGCNA) and screen the hub gene of LN. We performed WGCNA on the microarray expression profile dataset of GSE104948 from Gene Expression Omnibus (GEO) database with 18 normal and 21 LN samples of glomerulus. A total of 5,942 genes were divided into 5 co-expression modules, one of which was significantly correlated to LN. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted on the LN-related module and the module was proved to be associated mainly with the activation of inflammation, immune response, cytokines, and immune cells. Genes in the most significant GO terms were extracted for sub-networks of WGNCA. We evaluated the centrality of genes in the sub-networks by Maximal Clique Centrality (MCC) method and *CD36* was ultimately screened out as a hub candidate gene of the pathogenesis of LN. The result was verified by its differentially expressed level between normal and LN in GSE104948 and the other three multi-microarray datasets of GEO. Moreover, we further demonstrated that the expression level of *CD36* is related to the WHO Lupus Nephritis Class of LN patients with the help of Nephroseq database. The current study proposed *CD36* as a vital candidate gene in LN for the first time and *CD36* may perform as a brand-new biomarker or therapeutic target of LN in the future.

CD36 identified by weighted gene co-expression network analysis as a hub candidate gene in lupus nephritis

Huiying Yang[1], Hua Li[1]

[1] Department of Nephrology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

Corresponding Author:
Hua Li1
3 Qingchun East Road, Hangzhou, Zhejiang province, China
Email address: h_li@zju.edu.cn

Abstract

Lupus nephritis (LN) is a severe manifestation of systemic lupus erythematosus (SLE), which often progresses to end-stage renal disease (ESRD) and ultimately leads to death. At present, there are no definitive therapies towards LN, so that illuminating the molecular mechanism behind the disease has become an urgent task for researchers. Bioinformatics has become a widely utilized method for exploring genes related to disease. This study set out to conduct weighted gene co-expression network analysis (WGCNA) and screen the hub gene of LN. We performed WGCNA on the microarray expression profile dataset of GSE104948 from Gene Expression Omnibus (GEO) database with 18 normal and 21 LN samples of glomerulus. A total of 5,942 genes were divided into 5 co-expression modules, one of which was significantly correlated to LN. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted on the LN-related module and the module was proved to be associated mainly with the activation of inflammation, immune response, cytokines, and immune cells. Genes in the most significant GO terms were extracted for sub-networks of WGNCA. We evaluated the centrality of genes in the sub-networks by Maximal Clique Centrality (MCC) method and *CD36* was ultimately screened out as a hub candidate gene of the pathogenesis of LN. The result was verified by its differentially expressed level between normal and LN in GSE104948 and the other three multi-microarray datasets of GEO. Moreover, we further demonstrated that the expression level of *CD36* is related to the WHO Lupus Nephritis Class of LN patients with the help of Nephroseq database. The current study proposed *CD36* as a vital candidate gene in LN for the first time and *CD36* may perform as a brand-new biomarker or therapeutic target of LN in the future.

Introduction

Systemic lupus erythematosus (SLE) is a chronic, systemic autoimmune disease characterized by autoantibody production, complement activation and immune complex deposition. The incidence of SLE ranges from 0.03‰ to 2.32‰ person-years worldwide (Rees et al. 2017).

Lupus nephritis (LN) is one of the most frequent and severe organ manifestations in patients with SLE, the hallmark of which is often glomerulonephritis. Approximately 50% of SLE patients develop clinically evident renal disease, up to 11% of whom develop end-stage renal disease (ESRD) at 5 years (Tektonidou, Dasgupta and Ward 2016, Almaani, Meara and Rovin 2017). LN is an important cause of ESRD and mortality. The initial and subsequent therapy of LN mainly consists of immunosuppressants and glucocorticoids, which means there are little efficient and specific therapies. Thus, it is a pressing task to clarify molecular mechanisms involved in LN.

LN is characterized by its complicated physiopathologic mechanism. In LN patients, the formation of immune complexes in different glomerular compartments, the activation of innate immune signal pathways, the infiltration of immune cells and proinflammatory mediators can harm glomerular cells through various approaches (Devarapu et al. 2017). Although many studies have determined certain pathological mechanisms of LN, the pathogenesis is still far from clear.

With the development of gene microarray and high-throughput next-generation sequencing, bioinformatics analysis of gene expression profiling has been broadly applied to explore the mechanism underlying diseases and potential diagnostic biomarkers or treatment targets. Among diverse means aiming to investigate altered molecular elements based on comparison between groups of different states, weighted gene co-expression network analysis (WGCNA) is a powerful tool utilized for describing the correlation patterns among genes and exploring hub genes related to certain traits (van Dam et al. 2018, Langfelder and Horvath 2008). WGCNA constructs a co-expression network between genes, and then, genes are divided into several co-expression modules by clustering techniques. Genes in certain module are deemed to share similar biological function and biological process. At last, after relating modules to clinical traits, modules with high correlation to disease are further analyzed and hub genes of pivotal importance to disease are identified. WGCNA has been broadly used for studying of diseases such as cancer (Jardim-Perassi et al. 2019), neuropsychiatric disorder (Huggett and Stallings 2019), chronic disease (Morrow et al. 2015) and proved to be quite useful.

However, although researchers have conducted numerous bioinformatics studies about LN and have got many achievements (Arazi et al. 2019, Panousis et al. 2019), WGCNA has rarely been used for studies of LN (Sun et al. 2019). In our study, we constructed a co-expression network of the expression profile of glomerulus tissue by WGCNA and confirmed gene modules related to LN. After systematically analyzing the LN-related co-expression module by series of bioinformatics methods, a hub gene associated with LN was identified and verified. Depend on the potential roles of the hub gene in the pathogenesis of LN, we expect to propose novel clues of the diagnosis and treatment of LN.

Materials & Methods

Expression profile data collection

The overall procedures of our study are illustrated in the flow chart (Fig. 1). The gene expression profile of GSE104948 was selected and obtained from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). The raw data is available in the GEO database. The dataset consists of microarray-based gene expression profiles of 32 LN samples and 18 normal samples of glomerular tissues from SLE patients and living kidney transplant donors respectively. The glomerular tissues were microdissected and verified with glomerular-selective transcripts (Grayson et al. 2018). The expression values have already been log2 transformed.

Data preprocessing

The probe annotation was conducted under the Perl environment with the microarray platform file. Probes matching with multiple genes were removed, and for genes corresponding to multiple probes, the average values of probes were regarded as the expression values of the genes. A total of 11,884 genes were left after the probe annotation.

Since non-varying genes are usually regarded as background noise, we filtered the genes by variance and the top 50% (5,942 genes) with larger variance were chosen for subsequent analyses.

Weighted co-expression network construction and module division

Before WGCNA, we excluded the outlier samples by sample clustering with the hierarchical clustering method. The sample IDs and details of all samples included in our study are available in Table. S1. After that, we applied WGCNA with the expression profile by using the WGCNA package (Langfelder and Horvath 2008) in the R environment (version 3.5.3). Firstly, we calculated the Pearson`s correlation for all pair-wise genes and constructed a correlation matrix. Secondly, the correlation matrix was transformed to an adjacency matrix (also known as scale-free network) with an appropriate soft-thresholding value β. A reasonable β value would emphasize strong correlations between genes and penalize weak ones. We calculated the scale-free fit index and mean connectivity of each β value from 1 to 30 respectively, and when the scale-free fit index is up to 0.85, the β value with highest mean connectivity is deemed as the most appropriate one. Then, the adjacency matrix was converted to a topological overlap matrix (TOM) so that the indirect correlations between genes are concerned. Finally, we used average linkage hierarchical clustering according to the TOM-based dissimilarity measure to classify all genes into several co-expression modules with a minimum size of 30 genes, thereby the genes with similar expression patterns were divided into the same module. After defining the first principal component of a given module as eigengene, we calculated the Pearson`s correlations of the eigengenes, and merged modules whose eigengenes were highly correlated (with Pearson`s correlation higher than 0.75) into one module.

To verify the reliability of the division of modules, we plotted an adjacency heatmap of all the 5,942 genes analyzed by WGCNA. Besides, we completed a cluster analysis of module eigengenes and plotted an adjacency heatmap to find out the interactions among modules.

Identification of clinically significant modules

The clinical traits of our samples included normal and LN, we calculated the correlation between modules and traits. Modules of positive correlation with LN were considered as playing roles in the pathogenesis of the disease. On the other hand, genes in modules of positive correlation with normal trait are indispensable for maintaining normal biological functions. Thus, we extracted gene modules of highest correlation with LN and normal for subsequent studies.

Here, we introduced the definition of gene significance (GS) and module membership (MM), which represent the correlation of a given gene with clinical trait and module eigengene respectively. Genes in clinical-related modules should have high values and preferable correlations of GS and MM.

GO and KEGG pathway enrichment analyses

To explore the involved signal pathways and biological characteristics of genes in clinical-related modules, we conducted Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis and visualized the top 10 significant terms respectively with the clusterProfiler R package (Yu et al. 2012). For both of GO and KEGG, enrichment terms arrived the cut-off criterion of p-value < 0.01 and Benjamin-Hochberg adjusted p-value < 0.01 were considered as significant ones.

Differentially expressed genes analysis

To investigate the difference of the expression profiles between normal and LN samples of genes in clinical-related modules, we applied differentially expressed genes (DEGs) analysis based on Empirical Bayes test with the limma R package (Ritchie et al. 2015). The cut-off criterion was set as follow: |log2fold change (logFC)| > 1; p-value < 0.01; false discovery rate (*FDR*) < 0.001. The results were visualized with the heatmap R package (Galili et al. 2018).

Identification of hub gene

Hub gene of the LN-related module should have high connectivity with the whole module and LN trait, which may play critical roles in the molecular mechanism of LN. For identifying the hub gene related with LN, we extracted gene clusters that enriched in certain GO terms from the WGCNA network to construct sub-networks after GO enrichment analysis of the LN-related module. Then, we utilized the Cytoscape software and its plug-in cytohubba to seek out the hub gene from sub-networks (Shannon et al. 2003). After calculating the Maximal Clique Centrality (MCC) value of each gene, those with high MCC values was regarded as hub genes (Chin et al. 2014). The results were exhibited with the Cytoscape software. We then surveyed the GS value, MM value and logFC value of the selected hub gene to validate its reasonability.

Validation of hub gene with the other GEO datasets

To further verify the differential expression level of the hub gene between normal and LN tissues, we analyzed the logFC value of the hub gene with data from the other three GEO datasets (GSE32591, GSE99339 and GSE113342). The GEO IDs and details of the datasets were given in Table. S1.

Validation of the clinical significance of hub gene by Nephroseq database

To assess the relationship between the expression level of the hub gene and the activity or grade of LN, we visited the Nephroseq database (http://v5.nephroseq.org), which provides unique

access to datasets from the Applied Systems Biology Core at the University of Michigan, incorporating clinical data which is often difficult to collect from public sources. We then analyzed the difference of the expression level of hub gene between patients in different WHO Lupus Nephritis Class (Weening et al. 2004) based on two datasets (the dataset of Peterson Lupus Glom and the dataset of Berthier Lupus Glom) from Nephroseq database (details available in Supplemental Document. 1). We performed unpaired t test for comparisons between groups and set the criterion of two-tailed value of $p < 0.05$ as statistically significant.

Results

Data preprocessing

After data preprocessing, 5,942 genes were selected for subsequent analyses. Sample clustering excluded the outlier samples and a total of 18 normal samples and 21 LN samples were left. The final result of sample clustering is shown in Fig. 2A, revealing satisfactory intra-group consistency and distinct difference between groups. The GEO IDs and details about the source of the samples are available in Table. S1.

Weighted co-expression network construction and module division

In the current study, taking both scale-free fit index and mean connectivity as reference, the soft-thresholding was determined as 10 (Fig. 2B and Fig. 2C). Accordingly, the correlation matrix was transformed to an adjacency matrix and then converted to a topological overlap matrix. Based on average linkage hierarchical clustering and module merging, the genes were divided into 6 modules and were displayed with different colors (Fig. 3A), including the black, blue, brown, magenta, pink, and grey modules, containing 220, 2,881, 777, 465, 770, and 829 genes, respectively. Genes in the grey module were those couldn`t be divided into any co-expression modules.

Fig. 3B depicts the topological overlap adjacency among all the 5,942 genes analyzed by WGCNA, indicating that most genes have higher correlation with genes in the same module and lower correlation with genes in other modules, which means the division of the modules was accurate. The clustering dendrogram and adjacency heatmap of eigengene are shown in Fig. 3C and Fig. 3D, meaning that the 5 modules were mainly separated into two clusters.

Identification of clinically significant modules

We calculated the module-trait correlation coefficients and showed the results in Fig. 4A. The results illuminated that the blue module displayed highest correlation with LN trait ($r = 0.91$, $p =$ 2e-15), while the brown module related best with normal trait ($r = 0.66$, $p =$ 4e-06). The GS and MM value of all member genes of the blue and brown modules were shown in the scatterplots (Fig. 4B and Fig. 4C). The GS and MM value were of high correlation in the two modules (cor = 0.89, $p <$ 1e-200, and cor = 0.73, $p =$ 3.1e-130 respectively), suggesting that the genes in the two modules were associate with their module eigengenes and clinical traits synchronously and thus suitable for further analyses and hub gene excavation. We then renamed the two modules as top LN module and top non-LN module respectively.

DEGs analysis of trait-related modules

We applied DEGs analysis for the two trait-related modules. For the top LN module, 203 DEGs were screened in LN compare with normal, including 195 up-regulated genes and 8 down-regulated ones. While, the top non-LN module contained 78 DEGs, all of which were down-regulated. The top 30 of up-regulated and down-regulated genes are displayed in Fig. 5 respectively. The logFC value, *p*-value, and *FDR* of DEGs are given in Table. S2.

GO and KEGG Enrichment analyses of trait-related modules

To confirm the biological themes of genes in the trait-related modules and find the underlying biological pathways behind LN, we performed GO and KEGG enrichment analyses towards the top LN and top non-LN modules and the top 10 significant terms of GO and KEGG are exhibited in Fig. 6 and Fig. 7 respectively. The complete results of GO and KEGG enrichment analyses were given in Table. S3.

For the top LN module, enriched GO-BP terms were mainly about activation of immune response, immune cells, cytokine production, and inflammation (Fig. 6A), such as “neutrophil activation” (gene count = 178, *p* = 7.26e-301.52E-20), “positive regulation of defense response” (gene count = 160, *p* = 9.14e-25), “activation of innate immune response” (gene count = 107, *p* = 3.45e-18). Enriched GO-MF terms were mainly about cytokine and their receptors (Fig. 6B), such as “cytokine binding” (gene count = 39, *p* = 2.47e-8), “cytokine receptor activity” (gene count = 34, *p* = 5.3e-7), “cytokine receptor binding” (gene count = 69, p = 2.98e-5). For GO-CC, enriched terms were mainly involved in membrane and endocytic vesicle (Fig. 6C), such as “vesicle lumen” (gene count = 107, *p* = 4.13e-15), “cytoplasmic vesicle lumen” (gene count = 106, *p* = 8.99e-15), “membrane region” (gene count = 95, *p* = 2.33e-12). The results of KEGG enrichment were similar to that of GO-BP, were mainly about immune and inflammation (Fig. 7A). The top KEGG terms included “Complement and coagulation cascades” (gene count = 31, *p* = 6.55e-6), “Fc gamma R-mediated phagocytosis” (gene count = 34, *p* = 1.90e-5), “Th1 and Th2 cell differentiation” (gene count = 33, *p* = 3.05e-5). Meanwhile, several well-known pathways involved in LN were also included, such as Th17 cell differentiation.

On the other hand, for the top non-LN module, the GO-BP results were mainly involved in the metabolism process of different kinds of molecule (Fig. 6D), for example, “small molecule catabolic process” (gene count = 93, *p* = 3.27e-41), “organic acid catabolic process” (gene count = 73, *p* = 3.97e-39), “carboxylic acid catabolic process” (gene count = 73, *p* = 3.97e-39). The top 10 terms of GO-MF and GO-CC are also displayed (Fig. 6E and Fig. 6F). Similarly, the top KEGG terms were mainly about metabolism process (Fig. 7B).

Besides, we found that the magenta module also had high correlation with LN trait. The results of enrichment analyses of magenta module are also given in Table. S3.

Identification and validation of hub gene

We extracted the genes enriched in two of the top 10 significant GO-BP terms in top LN module, namely “neutrophil activation” and “positive regulation of defense response”, and constructed two sub-networks of the weighted co-expression network respectively. Co-expression pairs with top 500 weighted correlations in the sub-networks were selected for hub gene excavation. After importing the gene co-expression pairs and their weighted correlations into Cytoscape, we

calculated the MCC value of genes and the most central genes of the sub-network were screened out as shown in Fig. 8. In both sub-networks, *CD36* had the maximum MCC value among all, and was therefore deemed as the hub gene under the pathogenesis of LN.

The GS and MM value of *CD36* were 0.772 and 0.833, revealing *CD36* was significant correlated with the top LN module and LN trait. The DEGs analysis showed that the expression level of *CD36* in LN was abnormally up-regulated compared with that of normal. The logFC of *CD36* was 2.298, which ranked 11th among all.

Validation of hub gene with the other GEO datasets

For verifying our conclusion in a broader range, we interrogated the GEO database for more datasets about LN. We downloaded the datasets of GSE32591, GSE99339 and GSE113342, and then analyzed the differentially expressed level of *CD36* between LN and normal (Fig. 9). In the four different datasets, the expression level of *CD36* was consistently up-regulated in LN samples, illustrating a satisfactory reliability of the result.

Validation of hub gene by Nephroseq database

We analyzed the expression level of *CD36* in glomerular tissues under different severity of LN evaluated by WHO Lupus Nephritis Class. The result indicates that *CD36* was significantly up-regulated with the aggravation of LN (Fig. 10), that is to say, *CD36* plays an important role in the development of LN.

Discussion

In the current study, we used the expression profile of GSE104948 to screen the hub gene involved in the pathogenesis of LN. We performed WGCNA and divided all genes into 5 co-expression modules. After relating the modules to clinical traits, we concluded that the blue module was of highest correlation with LN and was suitable for hub gene excavating. The brown module had the highest correlation with normal trait, and was also worthy of subsequent analyses. GO and KEGG enrichment illuminated that genes in the top LN module were mostly enriched in biological themes of the activation of inflammation, immune response, cytokine, and immune cells, and the top non-LN module was mainly about the metabolism process of various molecules. What`s more, DEGs analysis showed that almost all DEGs in top LN module were abnormally up-regulated, revealing an aberrant activated stage of inflammation and immune response in LN. On the other hand, all DEGs in the top non-LN module were down-regulated, meaning a reduced ability of material metabolism in LN. To achieve the ultimately purpose of finding out the hub gene, we extracted genes enriched in the GO terms of “neutrophil activation” and “positive regulation of defense response” and constructed sub-networks accordingly. Base on the MCC method, *CD36* with maximum value of MCC in both sub-networks was regarded as the hub gene behind the pathogenesis of LN. We investigated the GS, MM and logFC of *CD36* and validated its importance. We interrogated the GEO database and got more datasets of LN, the obvious overexpression of *CD36* was therefore further verified. Moreover, the association between *CD36* and WHO Lupus Nephritis Class showed directly that the expression level of *CD36* gradually up-regulates along with the development of LN evaluated by WHO Lupus

Nephritis Class, providing strong evidence that the abnormal over-expression of *CD36* is an important element in the pathogenesis of LN.

The *CD36* gene is located on band q11.2 of chromosome 7 (Fernandez-Ruiz et al. 1993). The protein encoded by *CD36* is a kind of transmembrane protein (also known as scavenger receptor B2) expresses on the surface of various kinds of cells. In the glomerulus of kidney, *CD36* expresses in podocyte (Hua et al. 2015), mesangial cells (Ruan et al. 1999) and interstitial macrophages (Kennedy et al. 2013). Meanwhile, *CD36* expresses in immunity correlating cells such as monocytes and macrophages (Collot-Teixeira et al. 2007). With multiple ligands, *CD36* involves in complex biological process such as lipid homeostasis, immune response and cell apoptosis.

There are no relative reports concerning the direct relationship between *CD36* and LN. Our results have shown that *CD36* mainly participates in the function of neutrophil, such as "neutrophil activation", "neutrophil activation involved in immune response", "neutrophil degranulation", "neutrophil mediated immunity", as well as the activation of immune response, such as "positive regulation of defense response", "positive regulation of innate immune response", "innate immune response-activating signal transduction", "positive regulation of immune effector process". Base on the results, we conclude that *CD36* performs important functions in the pathogenesis and development of LN through affecting the function of neutrophil and innate immune response. Studies have proved that deposited immune complexes (IC) could activate complement and attract neutrophils and potentiate their responses, which will lead to intense glomerulonephritis, release protease and Reactive Oxygen Species (ROS), and give rise to kidney involvement of SLE (Tsuboi et al. 2008). Besides, IC-induced activation of neutrophils can lead to the formation of neutrophil extracellular traps (NETs), which can be pathogenic and promote the release of type I interferon (Garcia-Romo et al. 2011). *CD36* may candidate in the pathogenesis of LN by the above-mentioned pathways.

Numerous studies have shown that *CD36* participates in the pathogenesis of several kinds of chronic kidney disease (CKD).

It has long been known that chronic inflammation is an important segment of the progression of CKD. Studies have proved that the ligands signal via *CD36* to promote inflammatory response and the recruitment and activation of macrophage in the glomerulus (Kennedy et al. 2013). A report of LN showed that renal macrophage is associated with onset of nephritis and indicates poor prognosis (Bethunaickan et al. 2011). Meanwhile, oxidant stress plays a critical role in glomerular dysfunction. Along with chronic inflammation, *CD36* may facilitate the development of oxidant stress in LN (Hua et al. 2015, Kennedy et al. 2013, Aliou et al. 2016).

Podocyte is most susceptible to injury among the component of the glomerulus and its injury leads to glomerular dysfunction in various renal diseases including LN. Here, we determined that podocyte functional markers were down-ragulated in LN glomerular tissue, including *WT1* (logFC = -0.273, $p = 0.009$) and *NPHS1* (logFC = -0.306, $p = 0.034$), indicating the fact of podocyte injury. It is reported that in primary nephrotic syndrome mouse, the overexpression of

CD36 in the podocyte promotes its apoptosis (Yang et al. 2018). There is probably similar pathogenesis in the progression of LN.

Ectopic lipid deposition in kidney may cause lipotoxicity and further affect the function of the kidney (Lin et al. 2019). *CD36* is a multifunctional protein function as a key molecule in the uptake of long-chain fatty acids, which is the main component of fatty acids uptake system in the kidney and plays a critical rule in the development of CKD (Gai et al. 2019). The expression level of *CD36* is higher in kidney with acute or chronic damage, and lipid disorders will stimulate the up-regulation of *CD36*. Furthermore, *CD36* can promote the uptake of lipid from plasma to tissue (Hua et al. 2015, Lin et al. 2019, Nosadini and Tonolo 2011, Yang et al. 2017). Among our results, the top 10 GO terms of non-LN module includes two terms about lipid metabolism: fatty acid catabolic process (p = 8.03e-22) and fatty acid metabolic process (p = 2.62e-20), implying that lipid disorders exist in the glomerular tissue, in which *CD36* may take part.

Regrettably, as most reports on the relationship between *CD36* and kidney diseases are about the disorders of kidney in metabolic diseases, there is no study about the role *CD36* plays in immune response, which is worthy of further study in LN.

Conclusions

In conclusion, through WGCNA and a series of comprehensive bioinformatics analyses, *CD36* was confirmed for the first time as a hub gene in the pathogenesis of LN. *CD36* is likely to become a new biomarker or therapeutic target of LN. Our work might provide a new insight for exploring the molecular mechanisms of LN.

References

Aliou, Y., M. C. Liao, X. P. Zhao, S. Y. Chang, I. Chenier, J. R. Ingelfinger & S. L. Zhang (2016) Post-weaning high-fat diet accelerates kidney injury, but not hypertension programmed by maternal diabetes. Pediatr Res, 79, 416-24.

Arazi, A., D. A. Rao, C. C. Berthier, A. Davidson, Y. Liu, P. J. Hoover, A. Chicoine, T. M. Eisenhaure, A. H. Jonsson, S. Li, D. J. Lieb, F. Zhang, K. Slowikowski, E. P. Browne, A. Noma, D. Sutherby, S. Steelman, D. E. Smilek, P. Tosta, W. Apruzzese, E. Massarotti, M. Dall'Era, M. Park, D. L. Kamen, R. A. Furie, F. Payan-Schober, W. F. Pendergraft, 3rd, E. A. McInnis, J. P. Buyon, M. A. Petri, C. Putterman, K. C. Kalunian, E. S. Woodle, J. A. Lederer, D. A. Hildeman, C. Nusbaum, S. Raychaudhuri, M. Kretzler, J. H. Anolik, M. B. Brenner, D. Wofsy, N. Hacohen & B. Diamond (2019) The immune cell landscape in kidneys of patients with lupus nephritis. Nat Immunol, 20, 902-914.

Almaani, S., A. Meara & B. H. Rovin (2017) Update on Lupus Nephritis. Clin J Am Soc Nephrol, 12, 825-835.

Bethunaickan, R., C. C. Berthier, M. Ramanujam, R. Sahu, W. Zhang, Y. Sun, E. P. Bottinger, L. Ivashkiv, M. Kretzler & A. Davidson (2011) A unique hybrid renal mononuclear phagocyte activation phenotype in murine systemic lupus erythematosus nephritis. J Immunol, 186, 4994-5003.

Chin, C. H., S. H. Chen, H. H. Wu, C. W. Ho, M. T. Ko & C. Y. Lin (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol, 8 Suppl 4, S11.

Collot-Teixeira, S., J. Martin, C. McDermott-Roe, R. Poston & J. L. McGregor (2007) CD36 and macrophages in atherosclerosis. Cardiovasc Res, 75, 468-77.

Devarapu, S. K., G. Lorenz, O. P. Kulkarni, H. J. Anders & S. R. Mulay (2017) Cellular and Molecular Mechanisms of Autoimmunity and Lupus Nephritis. Int Rev Cell Mol Biol, 332, 43-154.

Fernandez-Ruiz, E., A. L. Armesilla, F. Sanchez-Madrid & M. A. Vega (1993) Gene encoding the collagen type I and thrombospondin receptor CD36 is located on chromosome 7q11.2. Genomics, 17, 759-61.

Gai, Z., T. Wang, M. Visentin, G. A. Kullak-Ublick, X. Fu & Z. Wang (2019) Lipid Accumulation and Chronic Kidney Disease. Nutrients, 11.

Galili, T., A. O'Callaghan, J. Sidi & C. Sievert (2018) heatmaply: an R package for creating interactive cluster heatmaps for online publishing. Bioinformatics, 34, 1600-1602.

Garcia-Romo, G. S., S. Caielli, B. Vega, J. Connolly, F. Allantaz, Z. Xu, M. Punaro, J. Baisch, C. Guiducci, R. L. Coffman, F. J. Barrat, J. Banchereau & V. Pascual (2011) Netting neutrophils are major inducers of type I IFN production in pediatric systemic lupus erythematosus. Sci Transl Med, 3, 73ra20

Grayson, P. C., S. Eddy, J. N. Taroni, Y. L. Lightfoot, L. Mariani, H. Parikh, M. T. Lindenmeyer, W. Ju, C. S. Greene, B. Godfrey, C. D. Cohen, J. Krischer, M. Kretzler & P. A. Merkel (2018) Metabolic pathways and immunometabolism in rare kidney diseases. Ann Rheum Dis, 77, 1226-1233.

Hua, W., H. Z. Huang, L. T. Tan, J. M. Wan, H. B. Gui, L. Zhao, X. Z. Ruan, X. M. Chen & X. G. Du (2015) CD36 Mediated Fatty Acid-Induced Podocyte Apoptosis via Oxidative Stress. PLoS One, 10, e0127507.

Huggett, S. B. & M. C. Stallings (2019) Cocaine'omics: Genome-wide and transcriptome-wide analyses provide biological insight into cocaine use and dependence. Addict Biol.

Jardim-Perassi, B. V., P. A. Alexandre, N. M. Sonehara, R. de Paula-Junior, O. Reis Junior, H. Fukumasu, R. Chammas, L. L. Coutinho & D. Zuccari (2019) RNA-Seq transcriptome analysis shows anti-tumor actions of melatonin in a breast cancer xenograft model. Sci Rep, 9, 966.

Kennedy, D. J., Y. Chen, W. Huang, J. Viterna, J. Liu, K. Westfall, J. Tian, D. J. Bartlett, W. H. Tang, Z. Xie, J. I. Shapiro & R. L. Silverstein (2013) CD36 and Na/K-ATPase-alpha1 form a proinflammatory signaling loop in kidney. Hypertension, 61, 216-24.

Langfelder, P. & S. Horvath (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

Lin, Y. C., M. S. Wu, Y. F. Lin, C. R. Chen, C. Y. Chen, C. J. Chen, C. C. Shen, K. C. Chen & C. C. Peng (2019) Nifedipine Modulates Renal Lipogenesis via the AMPK-SREBP Transcriptional Pathway. Int J Mol Sci, 20.

Morrow, J. D., W. Qiu, D. Chhabra, S. I. Rennard, P. Belloni, A. Belousov, S. G. Pillai & C. P. Hersh (2015) Identifying a gene expression signature of frequent COPD exacerbations in peripheral blood using network methods. BMC Med Genomics, 8, 1.

Nosadini, R. & G. Tonolo (2011) Role of oxidized low density lipoproteins and free fatty acids in the pathogenesis of glomerulopathy and tubulointerstitial lesions in type 2 diabetes. Nutr Metab Cardiovasc Dis, 21, 79-85.

Panousis, N. I., G. K. Bertsias, H. Ongen, I. Gergianaki, M. G. Tektonidou, M. Trachana, L. Romano-Palumbo, D. Bielser, C. Howald, C. Pamfil, A. Fanouriakis, D. Kosmara, A. Repa, P. Sidiropoulos, E. T. Dermitzakis & D. T. Boumpas (2019) Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. Ann Rheum Dis, 78, 1079-1089.

Rees, F., M. Doherty, M. J. Grainge, P. Lanyon & W. Zhang (2017) The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford), 56, 1945-1961.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi & G. K. Smyth (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 43, e47.

Ruan, X. Z., Z. Varghese, S. H. Powis & J. F. Moorhead (1999) Human mesangial cells express inducible macrophage scavenger receptor. Kidney Int, 56, 440-51.

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski & T. Ideker (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res, 13, 2498-504.

Sun, G., P. Zhu, Y. Dai & W. Chen (2019) Bioinformatics Analysis of the Core Genes Related to Lupus Nephritis Through a Network and Pathway-Based Approach. DNA Cell Biol, 38, 639-650.

Tektonidou, M. G., A. Dasgupta & M. M. Ward (2016) Risk of End-Stage Renal Disease in Patients With Lupus Nephritis, 1971-2015: A Systematic Review and Bayesian Meta-Analysis. Arthritis Rheumatol, 68, 1432-41.

Tsuboi, N., K. Asano, M. Lauterbach & T. N. Mayadas (2008) Human neutrophil Fcgamma receptors initiate and play specialized nonredundant roles in antibody-mediated inflammatory diseases. Immunity, 28, 833-46.

van Dam, S., U. Vosa, A. van der Graaf, L. Franke & J. P. de Magalhaes (2018) Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform, 19, 575-592.

Weening, J. J., V. D. D'Agati, M. M. Schwartz, S. V. Seshan, C. E. Alpers, G. B. Appel, J. E. Balow, J. A. Bruijn, T. Cook, F. Ferrario, A. B. Fogo, E. M. Ginzler, L. Hebert, G. Hill, P. Hill, J. C. Jennette, N. C. Kong, P. Lesavre, M. Lockshin, L. M. Looi, H. Makino, L. A. Moura & M. Nagata (2004) The classification of glomerulonephritis in systemic lupus erythematosus revisited. J Am Soc Nephrol, 15, 241-50.

Weening, J. J., V. D. D'Agati, M. M. Schwartz, S. V. Seshan, C. E. Alpers, G. B. Appel, J. E. Balow, J. A. Bruijn, T. Cook, F. Ferrario, A. B. Fogo, E. M. Ginzler, L. Hebert, G. Hill, P. Hill, J. C. Jennette, N. C. Kong, P. Lesavre, M. Lockshin, L. M. Looi, H. Makino, L. A. Moura & M. Nagata (2004) The classification of glomerulonephritis in systemic lupus erythematosus revisited. J Am Soc Nephrol, 15, 241-50.

Yang, X., D. M. Okamura, X. Lu, Y. Chen, J. Moorhead, Z. Varghese & X. Z. Ruan (2017) CD36 in chronic kidney disease: novel insights and therapeutic opportunities. Nat Rev Nephrol, 13, 769-781.

Yang, X., Y. Wu, Q. Li, G. Zhang, M. Wang, H. Yang & Q. Li (2018) CD36 Promotes Podocyte Apoptosis by Activating the Pyrin Domain-Containing-3 (NLRP3) Inflammasome in Primary Nephrotic Syndrome. Med Sci Monit, 24, 6832-6839.

Yu, G., L. G. Wang, Y. Han & Q. Y. He (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. Omics, 16, 284-7.

Figure 1

Flow chart of the whole procedures in this study.

Data processing, analyses, hub gene identification and validation.

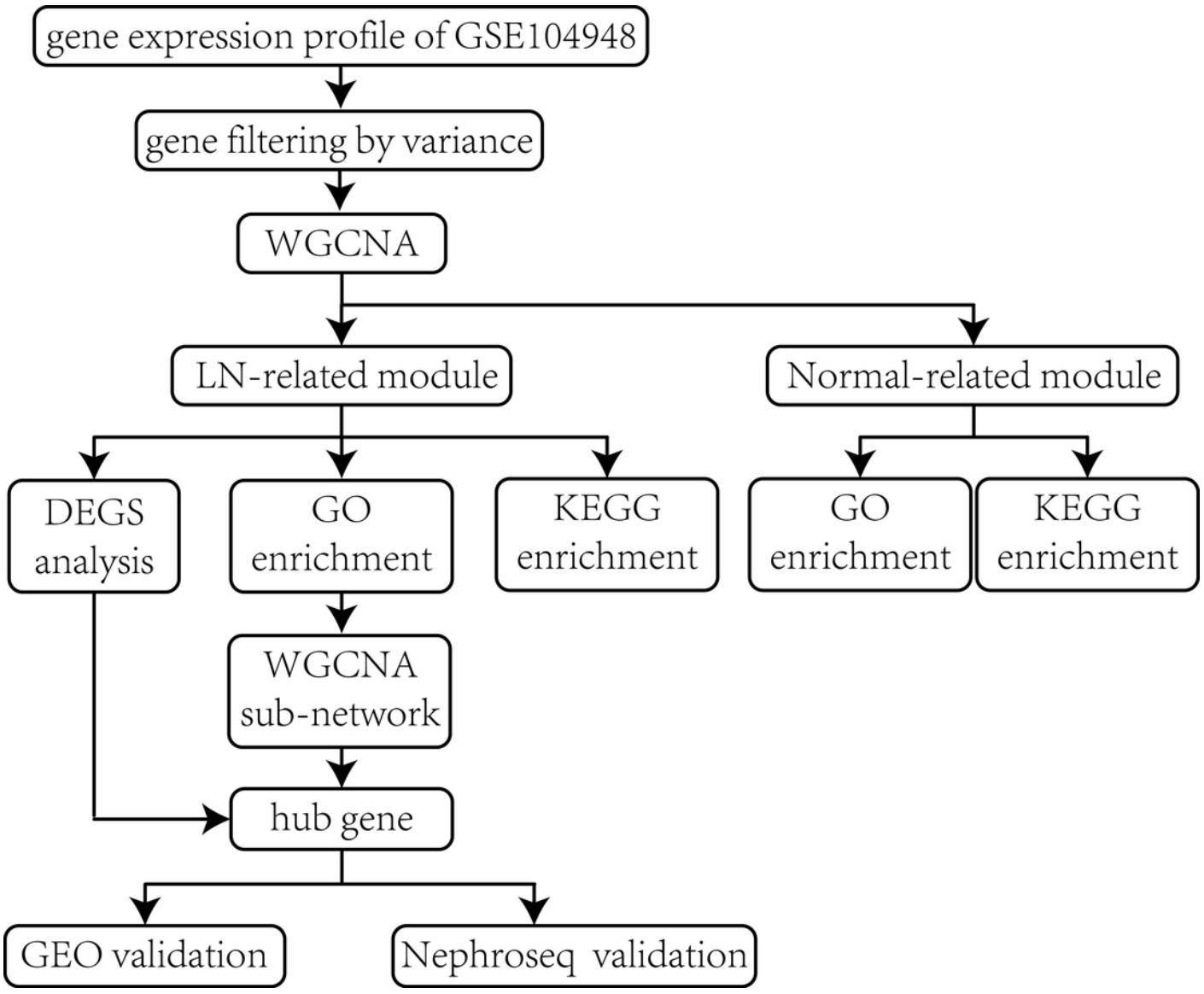


Figure 2

Sample cluster dendrogram and soft-thresholding values (β) estimation.

(A) Sample cluster dendrogram and clinical trait heatmap of 18 normal samples and 21 LN samples based on their expression profile. (B) Analysis of scale-free fit index of each β value from 1 to 30. (C) Analysis of mean connectivity of each β value from 1 to 30. β = 10 was chosen for subsequent analyses as it has the biggest mean connectivity when the scale-free fit index is up to 0.85.

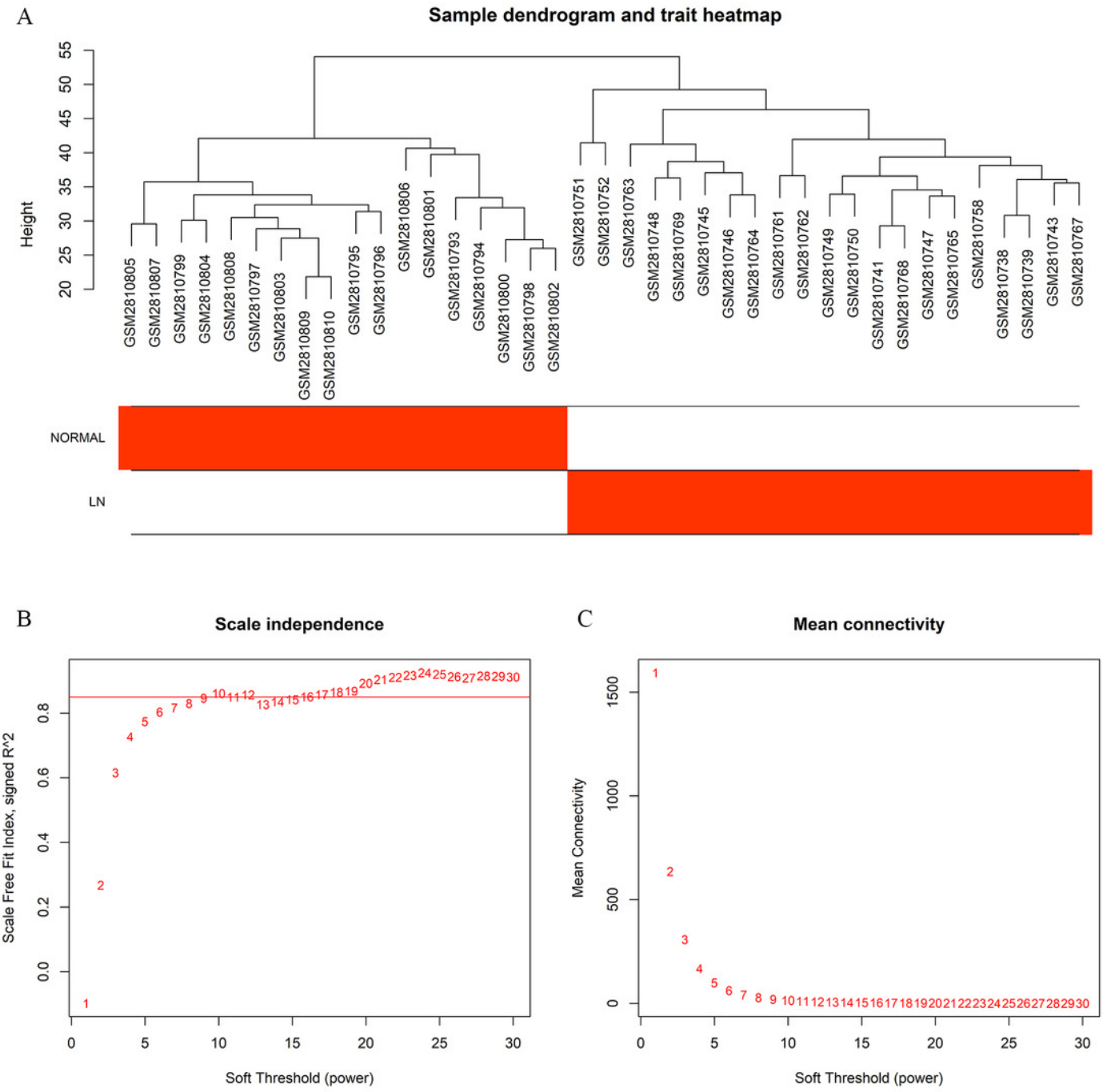

A
Sample dendrogram and trait heatmap
Height
20 25 30 35 40 45 50 55
GSM2810805 GSM2810807 GSM2810799 GSM2810804 GSM2810808 GSM2810797 GSM2810803 GSM2810809 GSM2810810 GSM2810795 GSM2810796 GSM2810806 GSM2810801 GSM2810793 GSM2810794 GSM2810800 GSM2810798 GSM2810802 GSM2810751 GSM2810752 GSM2810763 GSM2810748 GSM2810769 GSM2810745 GSM2810746 GSM2810764 GSM2810761 GSM2810762 GSM2810749 GSM2810750 GSM2810741 GSM2810768 GSM2810747 GSM2810765 GSM2810758 GSM2810738 GSM2810739 GSM2810743 GSM2810767
NORMAL
LN
B
Scale independence
Scale Free Fit Index, signed R^2
0.0 0.2 0.4 0.6 0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
0 5 10 15 20 25 30
Soft Threshold (power)
C
Mean connectivity
Mean Connectivity
0 500 1000 1500
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
0 5 10 15 20 25 30
Soft Threshold (power)

Figure 3

Division and validation of co-expression modules.

(A) Dendrogram of all genes divided into 6 modules base on a dissimilarity measure (1-TOM). The modules labeled by color are indicated below the dendrogram. The upper presents the original division by average linkage hierarchical clustering according to the TOM-based dissimilarity measure and the under presents the modules merged according to the Pearson`s correlation of eigengenes. (B) Adjacency heatmap of the 5,942 genes analyzed by WGCNA. The depth of the red color indicates the correlation between all pair-wise genes. The red color mainly distributes in the diagonal of the heatmap. (C) Clustering dendrogram of eigengenes. (D) Adjacency heatmap of eigengenes. Red represented high adjacency (positive correlation) and blue represented low adjacency (negative correlation).

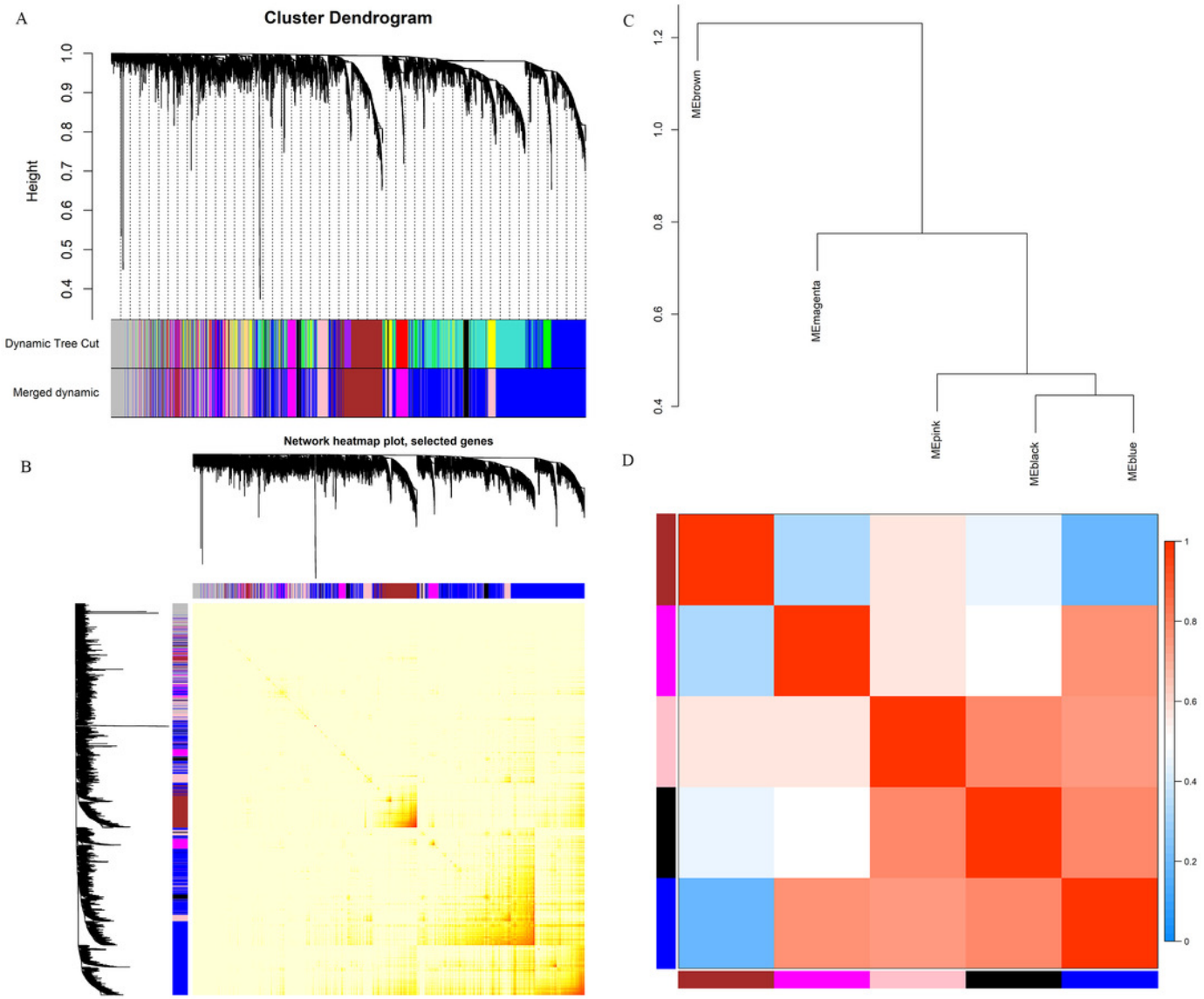
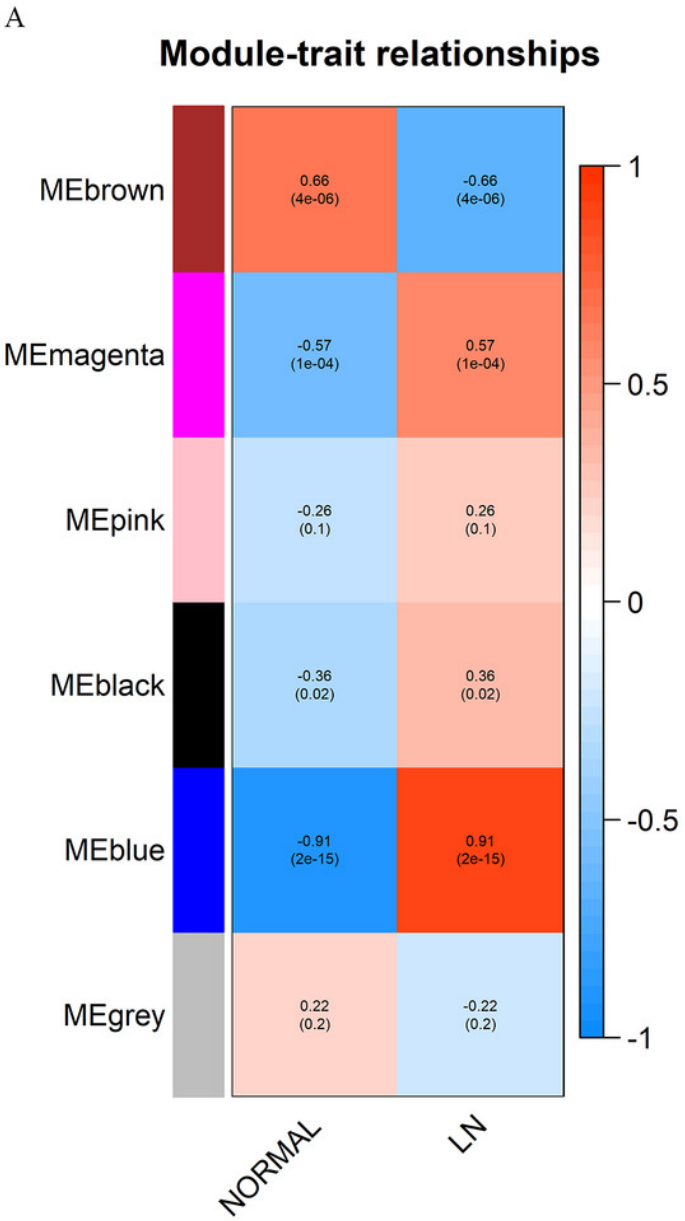
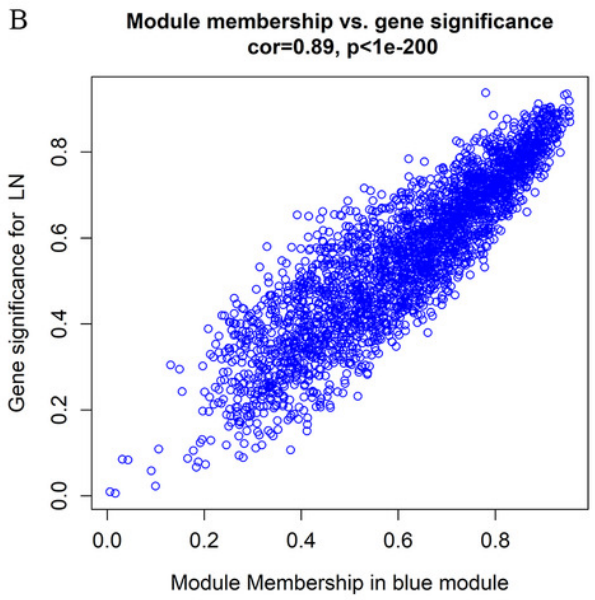

A
Cluster Dendrogram
Height
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Dynamic Tree Cut
Merged dynamic
B
Network heatmap plot, selected genes
C
1.2
1.0
0.8
0.6
0.4
MEbrown
MEmagenta
MEpink
MEblack
MEblue
D
1
0.8
0.6
0.4
0.2
0

Figure 4

Identification and verification of clinical related modules.

(A) Heatmap of module-trait correlations. Each cell depicts the correlation coefficients and *P*-value. The cells are colored by the intensity of correlation according to the color legend (red for positive correlation and blue for negative correlation). The blue module (top LN module) and brown module (top non-LN module) were identified as trait-related modules. (B) Scatter plot for correlation between the Gene significance (GS) and Module Membership (MM) in the top LN module. Correlation coefficients and *P*-value is labeled at the top. (C) Scatter plot for correlation between the GS and MM in the top non-LN module.


A
Module-trait relationships
MEbrown
MEmagenta
MEpink
MEblack
MEblue
MEgrey
0.66
(4e-06)
-0.66
(4e-06)
-0.57
(1e-04)
0.57
(1e-04)
-0.26
(0.1)
0.26
(0.1)
-0.36
(0.02)
0.36
(0.02)
-0.91
(2e-15)
0.91
(2e-15)
0.22
(0.2)
-0.22
(0.2)
1
0.5
0
-0.5
-1
NORMAL
LN


B
Module membership vs. gene significance
cor=0.89, p<1e-200
Gene significance for LN
0.0
0.2
0.4
0.6
0.8
Module Membership in blue module


C
Module membership vs. gene significance
cor=0.73, p=3.1e-130
Gene significance for NORMAL
0.0
0.2
0.4
0.6
0.8
Module Membership in brown module

Figure 5

DEGs analysis of two trait-related modules.

The color of each cell represents the expression level of a gene in a sample (red for high level and blue for low level). Only DEGs with top 30 logFC values are displayed. (A) Heatmap of DEGs in the top LN module. (B) Heatmap of DEGs in the top non-LN module.

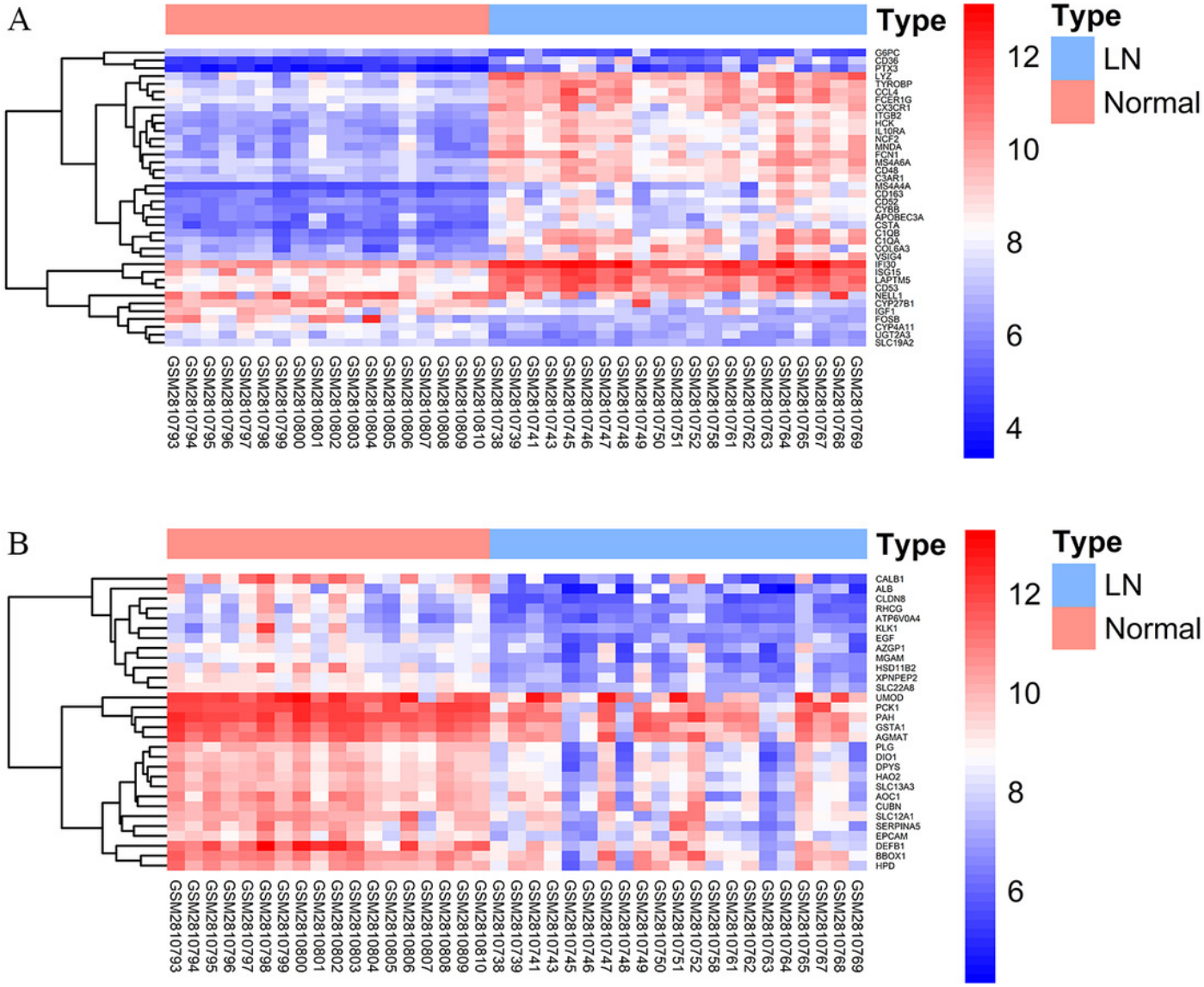


Figure 6

GO enrichment analyses of two trait-related modules.

The depth of color is corresponded to the enrichment significant of each term and the x-axis indicates the enriched gene count. (A) Top 10 significantly enriched GO Biological Process (BP) terms of top LN module. (B) Top 10 significantly enriched GO Molecular Function (MF) terms of top LN module. (C) Top 10 significantly enriched GO Cellular Component (CC) terms of top LN module. (D) Top 10 significantly enriched GO Biological Process (BP) terms of top non-LN module. (E) Top 10 significantly enriched GO Molecular Function (MF) terms of top non-LN module. (F) Top 10 significantly enriched GO Cellular Component (CC) terms of top non-LN module.

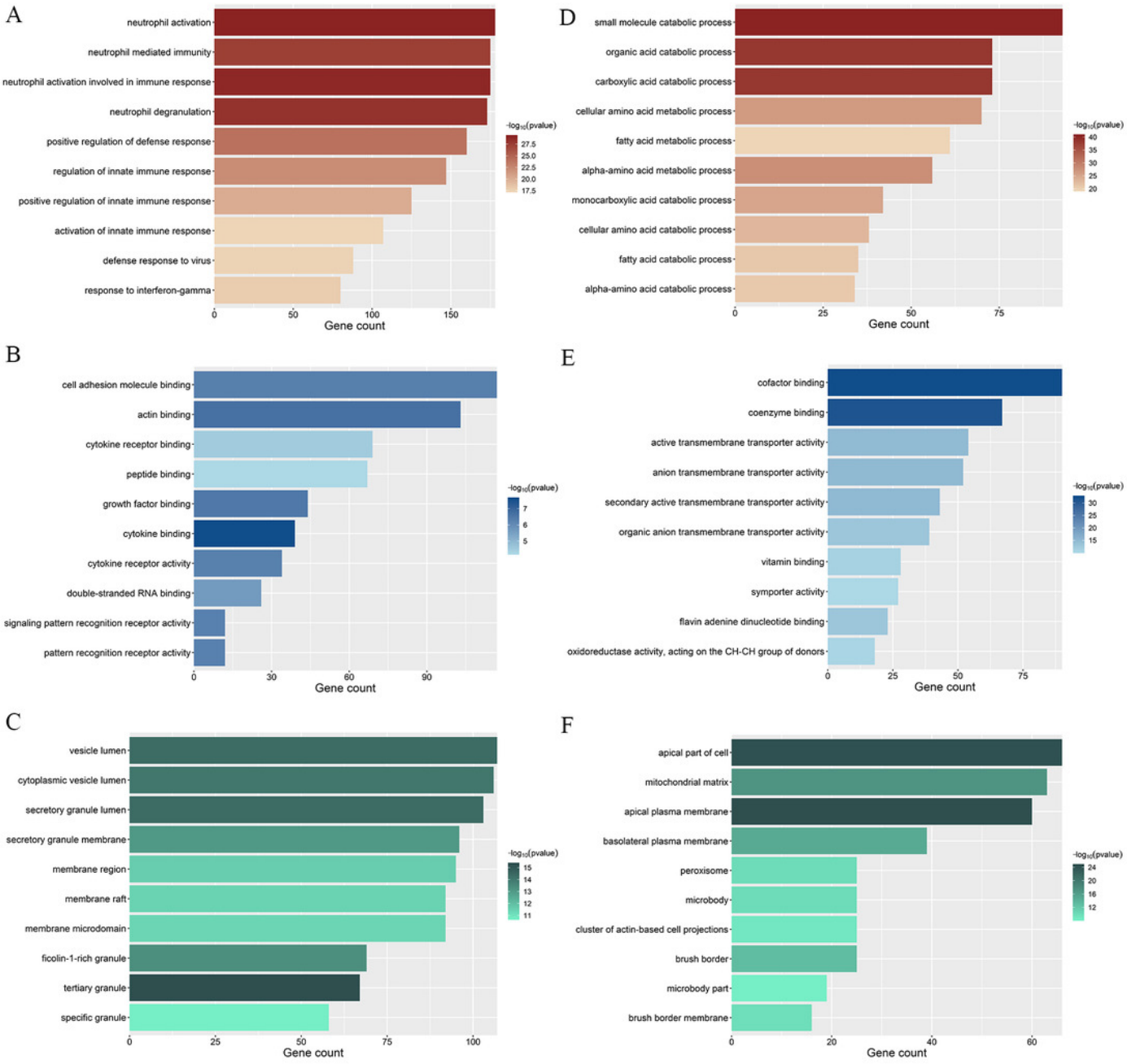
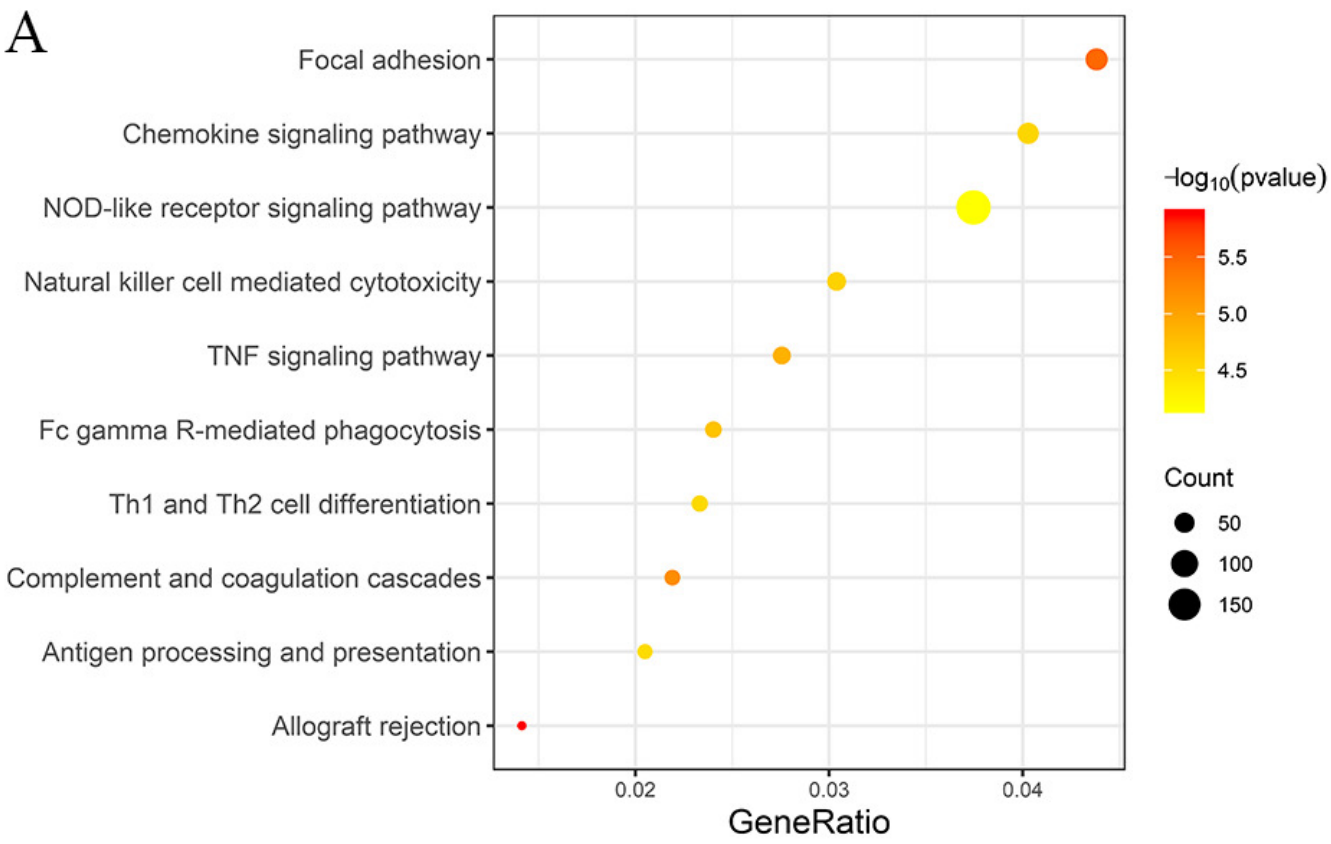

A
neutrophil activation
neutrophil mediated immunity
neutrophil activation involved in immune response
neutrophil degranulation
positive regulation of defense response
regulation of innate immune response
positive regulation of innate immune response
activation of innate immune response
defense response to virus
response to interferon-gamma
0
50
100
150
Gene count
-log10(pvalue)
27.5
25.0
22.5
20.0
17.5
D
small molecule catabolic process
organic acid catabolic process
carboxylic acid catabolic process
cellular amino acid metabolic process
fatty acid metabolic process
alpha-amino acid metabolic process
monocarboxylic acid catabolic process
cellular amino acid catabolic process
fatty acid catabolic process
alpha-amino acid catabolic process
0
25
50
75
Gene count
-log10(pvalue)
40
35
30
25
20
B
cell adhesion molecule binding
actin binding
cytokine receptor binding
peptide binding
growth factor binding
cytokine binding
cytokine receptor activity
double-stranded RNA binding
signaling pattern recognition receptor activity
pattern recognition receptor activity
0
30
60
90
Gene count
-log10(pvalue)
7
6
5
E
cofactor binding
coenzyme binding
active transmembrane transporter activity
anion transmembrane transporter activity
secondary active transmembrane transporter activity
organic anion transmembrane transporter activity
vitamin binding
symporter activity
flavin adenine dinucleotide binding
oxidoreductase activity, acting on the CH-CH group of donors
0
25
50
75
Gene count
-log10(pvalue)
30
25
20
15
C
vesicle lumen
cytoplasmic vesicle lumen
secretory granule lumen
secretory granule membrane
membrane region
membrane raft
membrane microdomain
ficolin-1-rich granule
tertiary granule
specific granule
0
25
50
75
100
Gene count
-log10(pvalue)
15
14
13
12
11
F
apical part of cell
mitochondrial matrix
apical plasma membrane
basolateral plasma membrane
peroxisome
microbody
cluster of actin-based cell projections
brush border
microbody part
brush border membrane
0
20
40
60
Gene count
-log10(pvalue)
24
20
16
12

Figure 7

KEGG enrichment analyses of two trait-related modules.

The depth of color is corresponded to the enrichment significant of each term and the size of the circle indicates the enriched gene count. (A) Top 10 significantly enriched KEGG terms of top LN module. (B) Top 10 significantly enriched KEGG terms of top non-LN module.


A
Focal adhesion
Chemokine signaling pathway
NOD-like receptor signaling pathway
Natural killer cell mediated cytotoxicity
TNF signaling pathway
Fc gamma R-mediated phagocytosis
Th1 and Th2 cell differentiation
Complement and coagulation cascades
Antigen processing and presentation
Allograft rejection
0.02
0.03
0.04
GeneRatio
-log10(pvalue)
5.5
5.0
4.5
Count
50
100
150

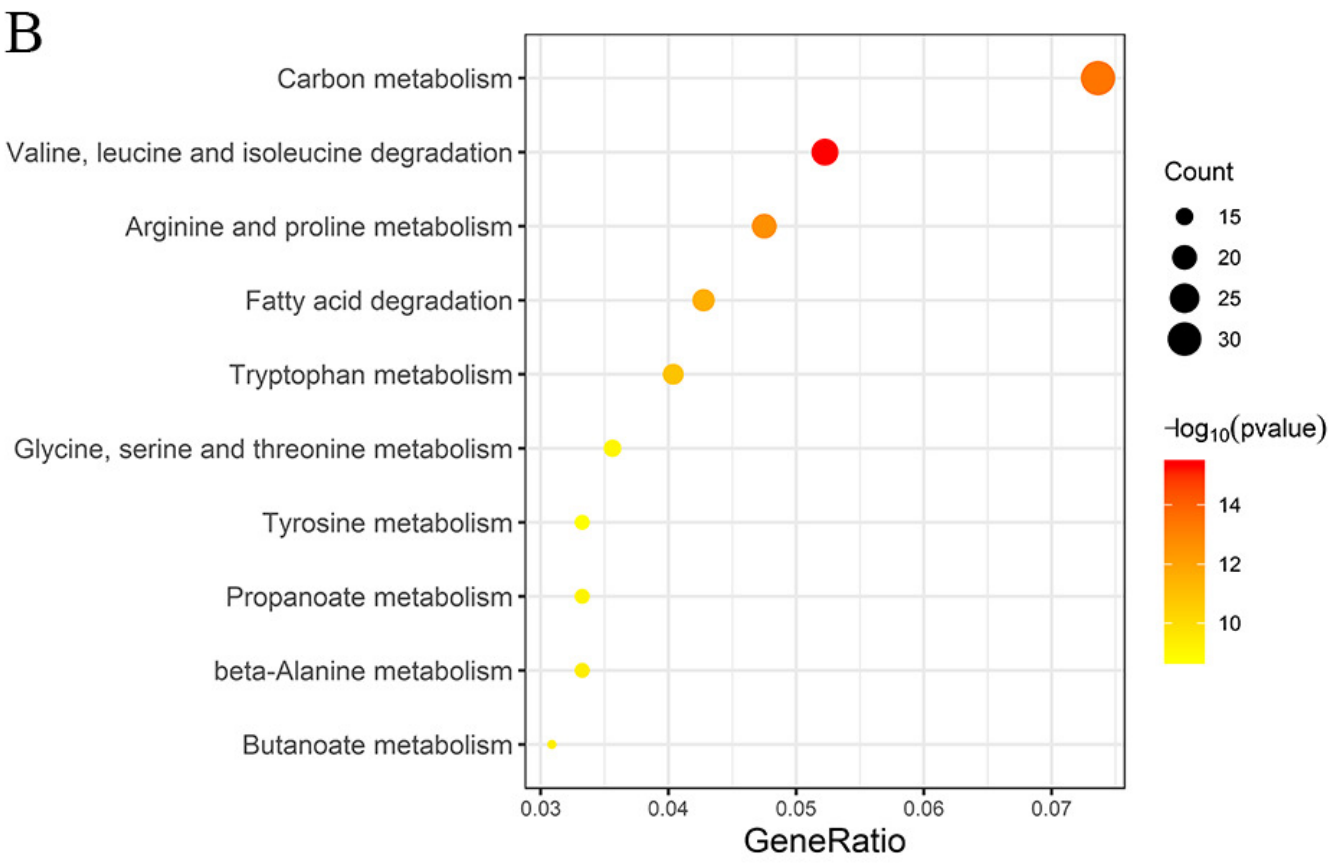

B
Carbon metabolism
Valine, leucine and isoleucine degradation
Arginine and proline metabolism
Fatty acid degradation
Tryptophan metabolism
Glycine, serine and threonine metabolism
Tyrosine metabolism
Propanoate metabolism
beta-Alanine metabolism
Butanoate metabolism
0.03
0.04
0.05
0.06
0.07
GeneRatio
Count
15
20
25
30
-log10(pvalue)
14
12
10

Figure 8

Sub-networks of WGCNA based extracted based on most significant GO terms.

The nodes represent the genes and the edges represent the weighted correlation. Only co-expression pairs with top 500 weighted correlations are included. The red and yellow nodes represent genes of top 10 MCC values (red for a higher MCC value and yellow for lower). (A) Sub-network of the GO term “neutrophil activation”. (B) Sub-network of the GO term “positive regulation of defense response”.

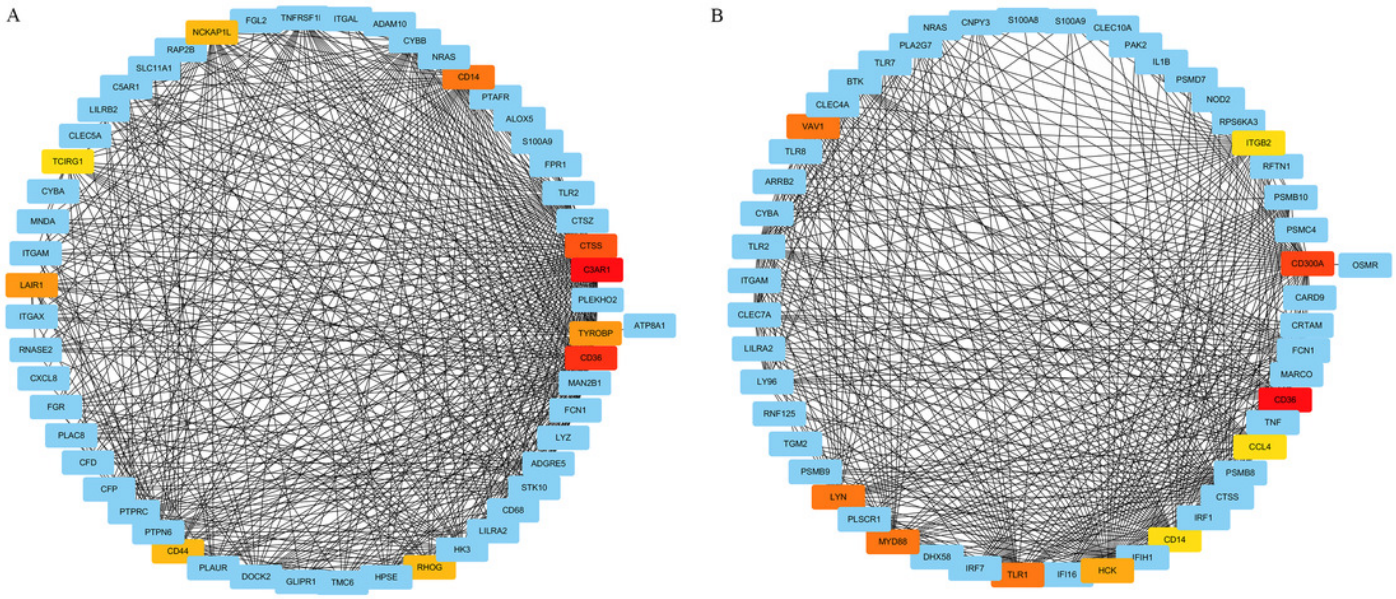


Figure 9

Differentially expressed level of *CD36* between LN and normal in different GEO datasets.

(A) Expression level of *CD36* in dataset of GSE104948. (B) Expression level of *CD36* in dataset of GSE32591. (C) Expression level of *CD36* in dataset of GSE99339. (D) Expression level of *CD36* in dataset of GSE113342.

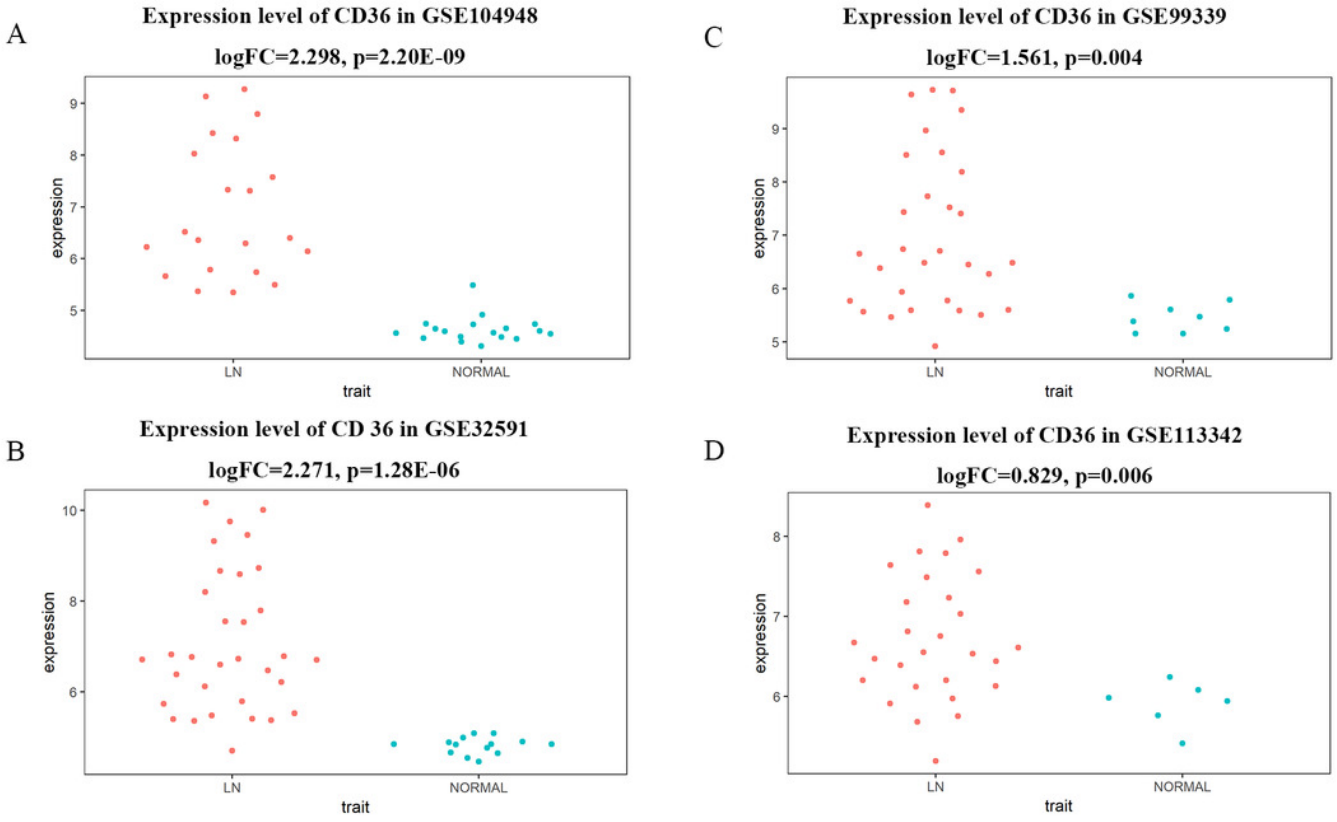


Figure 10

Differentially expressed level of *CD36* in glomerular tissues of different WHO Lupus Nephritis Class.

(A) Differentially expressed level of *CD36* in class II and class IV respectively (from the dataset of Berthier Lupus Glom). (B) Differentially expressed level of *CD36* in class III and class IV respectively (from the dataset of Peterson Lupus Glom).