Identification of the tubulointerstitial infiltrating immune cell landscape and immune marker related molecular patterns in lupus nephritis using bioinformatics analysis

Lu Zhang[1,2]^, Mengqin Zhang[3], Xing Chen[1,2], Yan He[3], Rongjuan Chen[3], Jun Zhang[1,2], Jiyi Huang[1,2], Chun Ouyang[4], Guixiu Shi[3]

[1]Department of Nephrology, The First Affiliated Hospital of Xiamen University, Xiamen, China; [2]The Fifth Hospital of Xiamen, Xiang'an Branch, The First Affiliated Hospital of Xiamen University, Xiamen, China; [3]Department of Rheumatology, The First Affiliated Hospital of Xiamen University, Xiamen, China; [4]Department of Nephrology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Contributions: (I) Conception and design: L Zhang, X Chen; (II) Administrative support: C Ouyang, G Shi; (III) Provision of study materials or patients: J Huang; (IV) Collection and assembly of data: L Zhang, M Zhang, H Yan, R Chen; (V) Data analysis and interpretation: J Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Chun Ouyang. Department of Nephrology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China. Email: c_ouyang@163.com; Guixiu Shi. Department of Rheumatology, The First Affiliated Hospital of Xiamen University, Xiamen, China. Email: gshi@xmu.edu.cn.

Background: Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease that commonly affects the kidneys. Research into markers that can predict the prognosis of tubulointerstitial lupus nephritis (LN) has been impeded by the lack of well-designed studies.

Methods: In this study, we selected and merged 3 sets of renal biopsy tubulointerstitial data from GSE32591, GSE69438, and GSE127797, including 95 LN and 15 living healthy donors. CIBERSORTx was utilized for differentially infiltrating immune cell (DIIC) analysis. Weighted Gene Co-Expression network analysis (WGCNA) was employed to explore differentially expressed gene (DEG) related modules. Combined WGCNA hub genes and protein-protein interaction (PPI) validation was used for immune marker identification. Lastly, unsupervised clustering was carried out to validate the correlation between these markers and clinical characteristics.

Results: Our findings unveiled TYROBP, C1QB, LAPTM5, CTSS, PTPRC as the 5 immune markers, which were negatively correlated with glomerular filtration rate (GFR). Specifically, the expression levels of TYROBP and C1QB were significantly different between proliferative LN (PLN) and membranous LN (MLN). Unsupervised clustering could aggregate LN by these immune marker expression spectrums.

Conclusions: This study is the first to identify infiltrating immune cells and associated molecular patterns in the tubulointerstitium of LN by utilizing bioinformatics methods. These findings contribute to a better understanding of the mechanisms behind LN, and promote more precise diagnosis.

Keywords: Lupus nephritis (LN); bioinformatic analysis; CIBERSORTx; weighted gene co-expression network analysis (WGCNA); Immune infiltration; Unsupervised clustering

---

^ ORCID: 0000-0002-1911-8696.

Introduction

Systemic lupus erythematosus (SLE) is heterogeneous disease with a broad range of clinical signs and symptoms including autoimmune manifestations, of which lupus nephritis (LN) is the most common cause of kidney disease in SLE. Approximately 50–60% patients have onset renal injury when SLE is diagnosed (1), and renal disease constitutes a major risk factor for SLE morbidity and mortality (1,2). Diagnostic pathological classification, as stipulated by the International Society of Nephrology/Renal Pathology Society (ISN/RPS) 2018 guidelines (3), relies on microscopic descriptions which mainly focus on glomeruli, therefore tubulointerstitial changes are rarely paid attention to. Although more and more nephrologists recognize that interstitial inflammatory cell infiltration and subsequent tubular cell necrosis can help predict poor prognosis (4-6), these processes are not fully understood. Extensive research has shown that SLE can be attributed to genetic factors and environmental triggers (1,2). Several risk alleles associated with SLE have been implicated in LN, but genetic studies or transcriptome studies that identify risk genes in LN but not in SLE are rare. These novel findings include apolipoprotein L1 (APOL1), platelet-derived growth factor receptor alpha (PDGFRA), and hyaluronan synthase 2 (HAS2) (1). The 2 risk alleles for APOL1 have more than 2.5-fold increased risk for renal failure compared with other markers (7,8). Genetic modifications and polymorphisms also identified HLA-DR3, HLA-DR4, HLA-DR11, and HLA-DR15, which have been proposed to protect or aggravate LN separately (9,10). Risk alleles alone only confer a small contribution to the development of SLE, or even LN and renal failure. Hence, larger studies or comprehensive analyses of the transcriptome are needed to better clarify the transcriptomic or translational change in LN. Previous studies have linked abnormal immunity to the pathogenesis of LN, as well as processes including aberrant podocyte apoptosis, autoantibody production, and immune complex (IC) deposition in the mesangial matrix under endothelial cells and epithelial cells following complement activation (11). So far, however, the activation of intrarenal immunity is still unclear, and the interaction with endogenous renal cells is also unknown. Apart from glomerular infiltration, the interstitial microenvironment (including tubules and interstitium) can produce clonally restricted autoantibodies as well (12), however, studies on this topic are lacking. Treat-to-target (T2T) strategies for SLE have been raised in last decade (13-15), and further research into LN mechanisms can help to treat patients more precisely and effectively (16). The classification system and pathological diagnosis can be refined by disease molecular patterns, such as B cell activation dominant (17,18), interferon (IFN)-opathy (IFN pathway dominant) (19,20), NETosis (neutrophil extracellular traps activation and release dominant) (21,22), and complement activation dominant (23), which can be treated by B-cell therapy such as anti-BAFF (B cell activating factor) (24), JAKi (JAK/STAT inhibitor), CTX/MMF (cyclophosphamide/ mycophenolate), and anti-C5 antibody separately (1). Hence, further studies which address molecular progression are required.

Bioinformatics analysis is a practical tool for deeply probing transcriptomic data. Our study set out to identify differentially expressed genes (DEGs) and differentially infiltrating immune cells (DIICs) which can help identify the key regulators of LN. Additionally, weighted gene co-expression network analysis (WGCNA) (25) and unsupervised clustering are powerful tools that can offer fresh perspectives for our project. Overall, we identified DIICs from 3 LN tubulointerstitium datasets and summarized the main immune markers by constructing protein-protein interaction (PPI) networks by DEGs and WGCNA. The overlapping gene sets of PPI and hub genes from WGCNA were recognized as immune markers for LN interstitial lesions. The final section validates the main findings of immune markers using unsupervised clustering. The selected immune markers distinguished LN samples into 5 clusters by their expression levels, and all of them were negatively correlated with glomerular filtration rate (GFR). In this paper, we identified 5 immune markers for LN tubulointerstitial immune infiltration, which are potential biomarkers and warrant further study to improve LN diagnosis and prognosis.

We present the following article in accordance with the MDAR reporting checklist (available at http://dx.doi. org/10.21037/atm-20-7507).

Methods

Microarray data acquisition and processing

Gene expression datasets from three previously published studies that compared LN with living healthy donors were identified from publicly available gene expression repositories, namely, the Gene Expression Omnibus (GEO) https://www.ncbi.nlm.nih.gov/geo/ (26). To begin this

process, the first step was to set "lupus nephritis" as the keyword for searching. The enrolment criteria had to obey the rules as follows: (I) the study object must be human renal biopsy microarray or RNA-seq data; (II) samples must be divided into glomeruli and tubulointerstitium from kidney biopsy samples instead of whole cortex tissue; (III) at least one of them must include LN renal biopsy pathological classification. In accordance with the above criteria, GSE32591 (27), GSE69438 (28), and GSE127797 (29) were finally enrolled. The first GSE32591 screened 94 samples in total (32 LN patients *vs.* 15 living donors, everyone had both glomeruli and tubulointerstitium tissues), and only tubulointerstitium data from both groups were extracted for further experiments. This dataset was performed on Affymetrix Genechip HG-U133A GPL14663. GSE69438 had only 16 LN tubulointerstitium samples that met our criteria. This microarray dataset was performed on Affymetrix Human Genome U133 plus 2.0 Array GPL11670. The GSE127797 had 88 samples, including 47 LN tubulointerstitium tissues and 41 glomerular tissues. Only tubulointerstitium tissue was extracted for further study, and the sequencing was performed on Affymetrix Human Transcriptome Array 2.0 GPL24299. Only this dataset covered patients' biopsy pathological classifications. The identified samples (95 LN tubulointerstitium tissues and 15 normal controls) from these resources were normalized by RMA and merged together. For the purposes of batch correction, the combat function of Sva on R software was employed (30). Subsequently, gene ENTREZID labels were transformed to gene SYMBOLs using the DAVID online database (https://david.ncifcrf.gov/). Finally, we obtained 11,802 genes from 3 GSE datasets, and 2 genes failed to be transformed. Following deletion of the 2 nonsense genes, 11,800 gene SYMBOLs were collected for subsequent analysis. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Evaluation of immune cell infiltration by CIBERSORTx

CIBERSORTx is a novel tool to analyze cell types and expression in tissue. It can also be applied to formalin fixed paraffin embedded (FFPE) samples instead of CIBERSORT. The gene expression datasets of the 3 GSEs were uploaded to the CIBERSORTx web portal (https://cibersortx.stanford.edu/). We chose the signature gene matrix file LM22, which is a validated leukocyte gene signature matrix that contains 547 genes distinguishing 22 humans hematopoietic cell phenotypes (31). We then set run mode as bulk-mode, disabled quantile normalization, and set 500 permutations for significance analysis. Once the computational analysis was finished, only P value <0.05 results were included for differentially infiltrating analysis. The Wilcoxon signed-rank test was conducted to find the significantly different infiltrating immune cell types.

DEG screening and enriched GO and KEGG pathway analysis

Prior to further analysis of the significant DEGs, the limma R package based on R-3.6.2 was used, which can construct an analysis model between the disease group and the control group (32). P value <0.05 and |logFC|>0.5 were used as the limma cut-off standard. Immport (https://www.immport.org/) was used to emphasize the immune related genes among the DEGs. Following pathway enrichment analysis conducted using the clusterProfiler package (33), the DEGs were enriched and visualized by GO MF, GO biological process, GO cellular component, and KEGG ways. GO analysis was performed through the gseGO function, and KEGG pathway enrichment analysis was run by the gseKEGG function in the package. Q-value (the statistical test for P value) <0.05 was used as the cut-off criteria for pathway analysis.

WGCNA and immune cell related module investigation

WGCNA is a frequently utilized method which explores the complex relationships between genes and phenotypes (25). The predominant advantage of WGCNA is that it can transform gene expression to co-expression modules, providing insights into signaling networks that may be responsible for phenotypic traits of a disease. Based on the above DEGs, WGCNA installed in the R software package was employed to construct co-expression modules for 95 LN patient samples. Prior to commencing analysis, appropriate soft thresholds (power) were sought to construct a relatively scale-free network. To begin this process, the power value was screened out by the WGCNA algorithm, and the gradient method was conducted to test the scale independence and mean connectivity of different modules at different power values. The most appropriate power value (=14) was determined when the degree of independence was 0.85. Next, the expression matrix was transformed to

Page 4 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium

the adjacency matrix, which was further transformed to the topologic overlap matrix (TOM). To avoid bias and error, the minimum genes in each module was ruled as 30. After preparations were completed, flashcluster tools installed in R were used to perform the dynamic cluster analysis of samples based on the appropriate threshold value. Meanwhile, we also calculated every modules' eigengenes, attempted to analyze the clusters based on their eigengenes, and merged modules when their membership >0.8. In order to find the key drivers in the modules of interest, we also calculated the intramodular connectivity and estimated the most predominant key modules by differentially infiltrating immune cell. Finally, the corresponding gene information of the key module was output as hub genes (set 1).

PPI investigation

For the purposes of constructing a PPI network and looking for central nodes, the hub genes investigated from the key module by WGCNA were mapped to the search tool for the retrieval of interacting genes (STRING database; https://string-db.org/). PPI analysis results were output when combined score >0.6 and node connectivity degree >15 (set 2). Cytoscape was used to present the network (https://cytoscape.org/). Venn analysis was carried out to compare the candidates in the 2 sets of genes by using the online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/). The overlapping genes of set 1 and set 2 were recognized as immune markers.

Correlation between clinical characteristics and immune marker genes

The relationship between immune markers and renal function (GFR) was examined using the Nephroseq V5 tool (nephroseq.com). The expression data were downloaded from the online website and replotted using the ggplot2 of R package. The correlation coefficients were calculated by the geom function and labeled in the scatter plots. The research objects of GSE127797 included biopsy pathological classification, which contained classes I, III, IV, V, III + V, and IV + V. On completion of the calculation of 5 immune marker expression levels, the Wilcoxon rank test was used to determine any existing differences between PLN (class III, IV) vs. MLN (class V), and a one-way ANOVA test was performed for 3 group comparisons including the combined group PLN/MLN (class III + V, IV + V).

Unsupervised cluster analysis of novel molecular patterns for LN

Unsupervised cluster machine learning can find the relationships between samples by mining the intrinsic characteristics of the data, which can be utilized for clustering a new molecular pattern of the disease. Prior to clustering samples, we calculated the within group sum of squares. Once the model was fitted and the curve was visualized, we determined that the most appropriate number of clusters was 5. Next, k-means were employed for sample cluster analysis. Thereafter, the previously described 22 immune cell types and 5 freshly identified immune markers were validated by novel molecular clusters for estimation of the cluster efficiency.

Statistical analysis

Data are reported as mean ± SD. The Wilcoxon rank test was used for comparison between two groups and ANOVA followed by Tukey's multiple comparison test was used for comparison between three or more groups. GraphPad Prism software was used for statistical analyses. All experiments were repeated at least three times, and representative experiments are shown. Data were considered statistically significant when P<0.05.

Results

Acquisition and processing of transcriptomic datasets from the public domain

The research strategy is presented in *Figure 1*. Gene expression datasets from 3 previously published studies that compared LN tubulointerstitium to normal kidney tubulointerstitium were identified from publicly available gene expression repositories, namely, Gene Expression Omnibus (GEO). These datasets were GSE32591, GSE69438, and GSE127797. All of the data were normalized by RMA (Robust Multi-array Average). In total, we obtained 110 samples, including 95 LN patient tubulointerstitium samples and 15 normal controls (Table S1). Next, we merged all of the included data and performed batch correction using the ComBat function of sva on R software (Figure S1). Following correction, all of the data were evenly demonstrated in principal component analysis (PCA). Next, gene ENTREZID labels were transformed to SYMBOLs using the DAVID database. Subsequently, we obtained
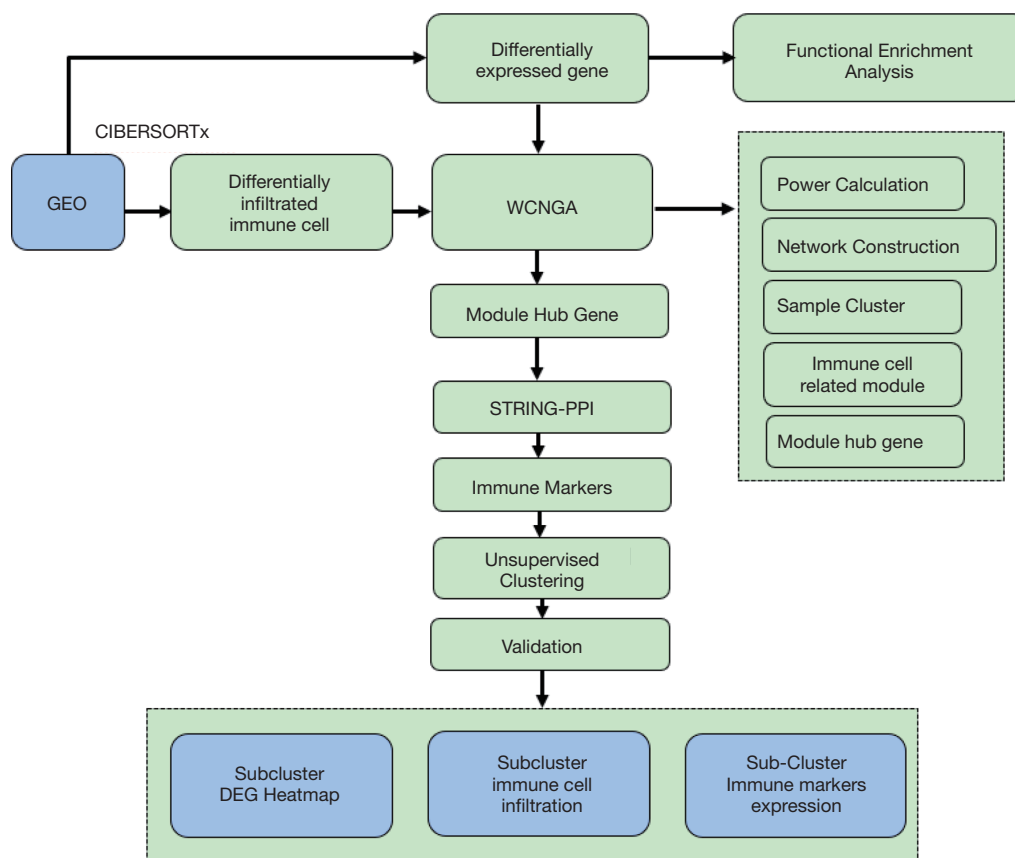
Figure 1 The workflow of the study.

11,802 genes, then excluded 2 genes (failed by searching computationally and manually). Following this deletion, a total of 11,800 gene SYMBOLs were used for further analysis.

Evaluation of the DIICs

Using CIBERSORTx, we analyzed the percentage of infiltrating immune cells based on the above RNA expression data. *Figure 2* provides an overview of the 22 kinds of immune cells infiltrating portion between the two groups. All of the differentially infiltrating immune cells were assessed by the Wilcoxon rank test. Only significantly different cell types and their statistical p values are listed in Table S2, which is sorted by infiltrating portion and cell type. The results, as shown in *Figure 2*, indicated that there were 10 types of DIICs between LN tubulointerstitium compared with normal controls. There was a clear trend of increasing B cells and plasma cells in LN, which suggests that humoral immunity plays a key role in the

mechanism of tubulointerstitial LN. T cells, especially CD8+ T cells, were exhausted and therefore lower in LN, and regulatory Treg cells were also inhibited in LN. Interestingly, there was a marked increase in the rates of both M1 and M2 macrophage subtypes, which could reflect the heterogeneously inflammatory character of LN. The percentage of eosinophils is higher in LN patients, while mast cells activated is over-exhausted by LN autoimmunity. The resting status immune cell or naïve cell may influence by their activated or differentiated status.

DEGs and gene set enrichment analysis

The limma package was employed for the analysis of DEGs between 95 LN patients and 15 healthy controls (P value <0.05 and |logFC|>0.5 were the criteria), and a total of 530 DEGs were identified. The heatmap demonstrated the top 100 DEGs between the 2 groups according to ascending P values (*Figure 3A*). The white label in *Figure 3A* shows the immune related DEGs by the Immprot database in
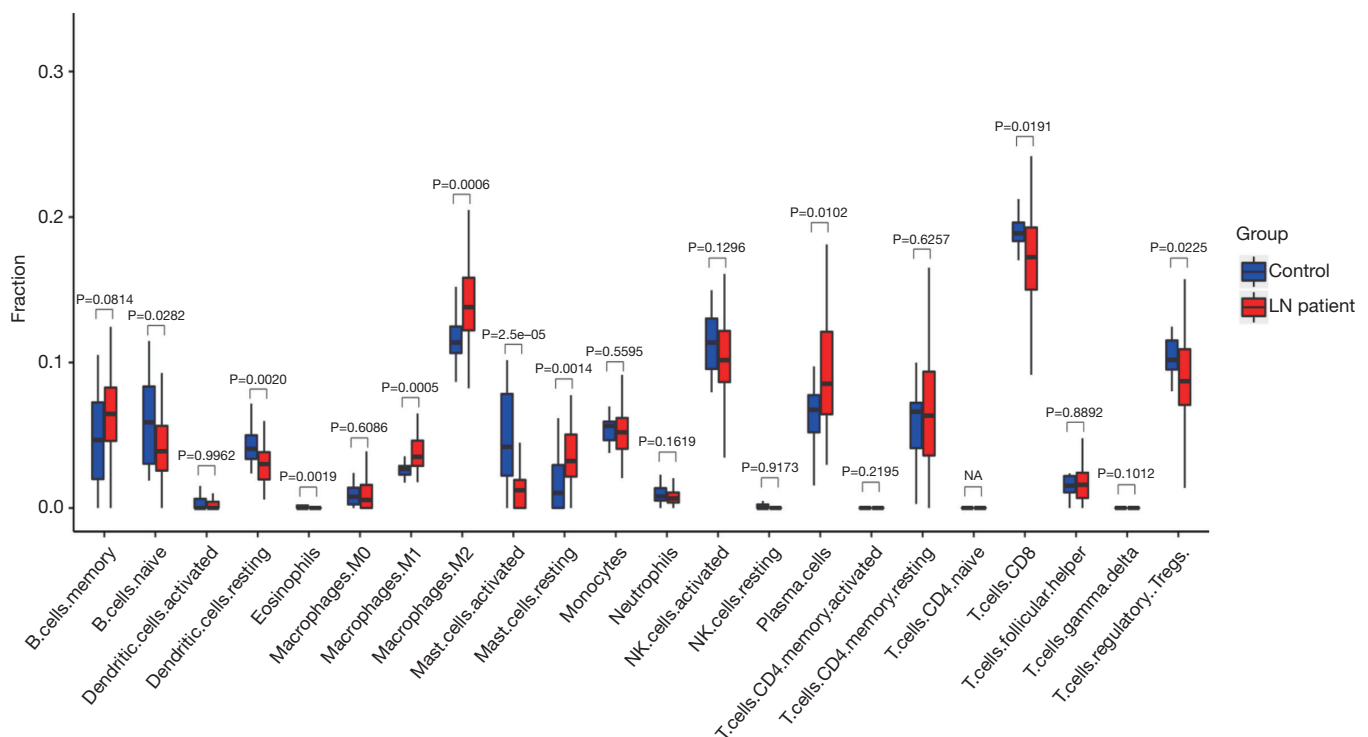
Page 6 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium

Figure 2 Differentially infiltrating immune cell type identification. The boxplot of 22 kinds of immune cell infiltrating fractions in lupus nephritis (LN) renal tubulointerstitium compared with living healthy donors. The red boxplot represents LN patients, and the blue boxplot represents healthy controls. P value <0.05 was recognized as a significant difference, and statistical analysis was performed using the Wilcoxon rank test. The fractions of Plasma.cells, Macrophages.M1, Macrophages.M2, Mast.cells.resting were higher in LN kidney tubulointerstitium. The fractions of T.cells. regulatory.Tregs, T.cells.CD8, B.cells.naïve, Dendritic.cells.resting, Eosinophils, and Mast.cells. activated were lower in LN patients.

volcanic plot (Figure S2). The function enrichment analysis was performed using the clusterProfiler package of R. A number of rules were considered when selecting pathways: every gene should at least be assigned into 6 pathways, and every pathway should be matched with 6 genes minimally in the whole cluster. Gene ontology (GO) recognized 807 biological process (BP), 75 cellular components (CC), and 50 molecular function (MF) pathways (difference at q<0.05). Kyoto Encyclopedia of Genes and Genomes (KEGG) recognized 54 pathways (q<0.05). *Figure 3B,C,D* displays the top 10 GO pathways separately, and *Figure 3E* shows the top 10 KEGG pathways. From the chart, it can be seen the greatest correlated BP terms of GO indicated that the DEGs were mainly related to the interferon-gamma and type I interferon (interferon-alpha and beta) pathways. The CC terms indicated that DEGs were related to cellular cytoplasmic lumen and particle transport. The MF terms emphasized enzyme and peptide binding and activity.

KEGG highlighted the correlation with interferons, anti-viral disease, and auto-immune disease (autoimmune thyroid disease, allograft rejection, type I diabetes mellitus). These immune related DEGs (Figure S2) and immune related pathway enrichment results shed substantial light on the feasibility of exploring novel immune markers.

The WGCNA co-expression network of tubulointerstitial LN

Based on the above 530 DEGs, the WGCNA R package was used for co-expression network construction. To begin this process, the appropriate soft thresholds were determined for constructing a scale-free network. The results showed when power =14 (*Figure 4A*), the scale free topology model fit curve was smooth and steady in plateau. In order to avoid similarity between each gene, the minimal gene number was 30 for every gene network module (*Table 1*). Thereafter, the flashclust package was used for dynamic
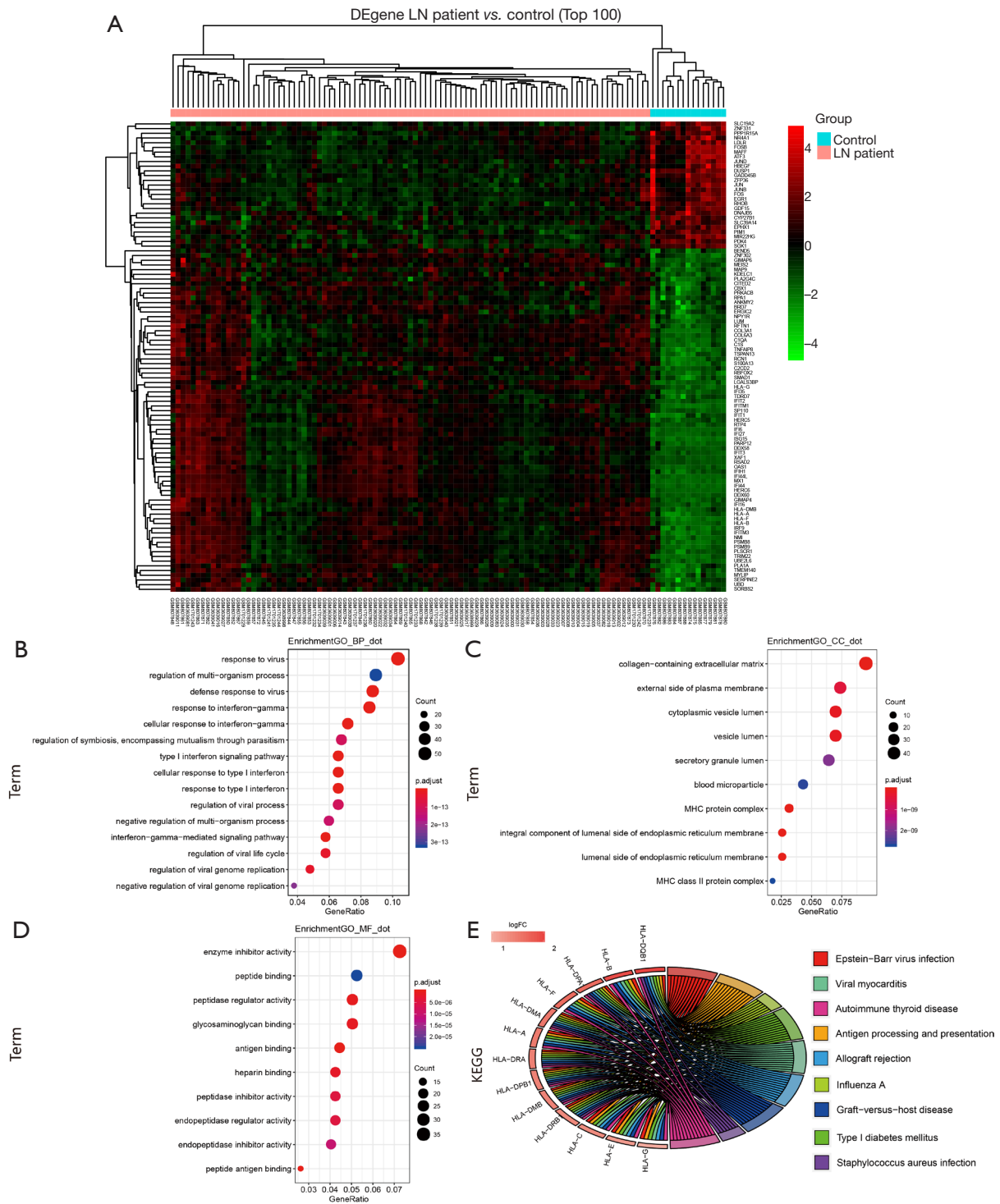
Figure 3 The top 100 identified differentially expressed genes (DEGs) in the cohort and pathway enrichment analysis by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). (A) Heatmap of DEGs demonstrating the top 100 among 530 DEGs recognized from the control and lupus nephritis (LN) groups according to P value. Red dots represent relatively upregulated genes and green dots represent downregulated genes. (B) Bubble plot of the top 15 GO pathways by biological process. (C) Bubble plot of top 10 GO pathways by cellular component. (D) Bubble plot of top 10 GO pathways by molecular function. (E) Top 10 KEGG enrichment pathways among the 54 KEGG pathways. All of the statistics were carried out using the Wilcoxon rank test, q value <0.05 as the differential pathways (q value is statistic test for P value).
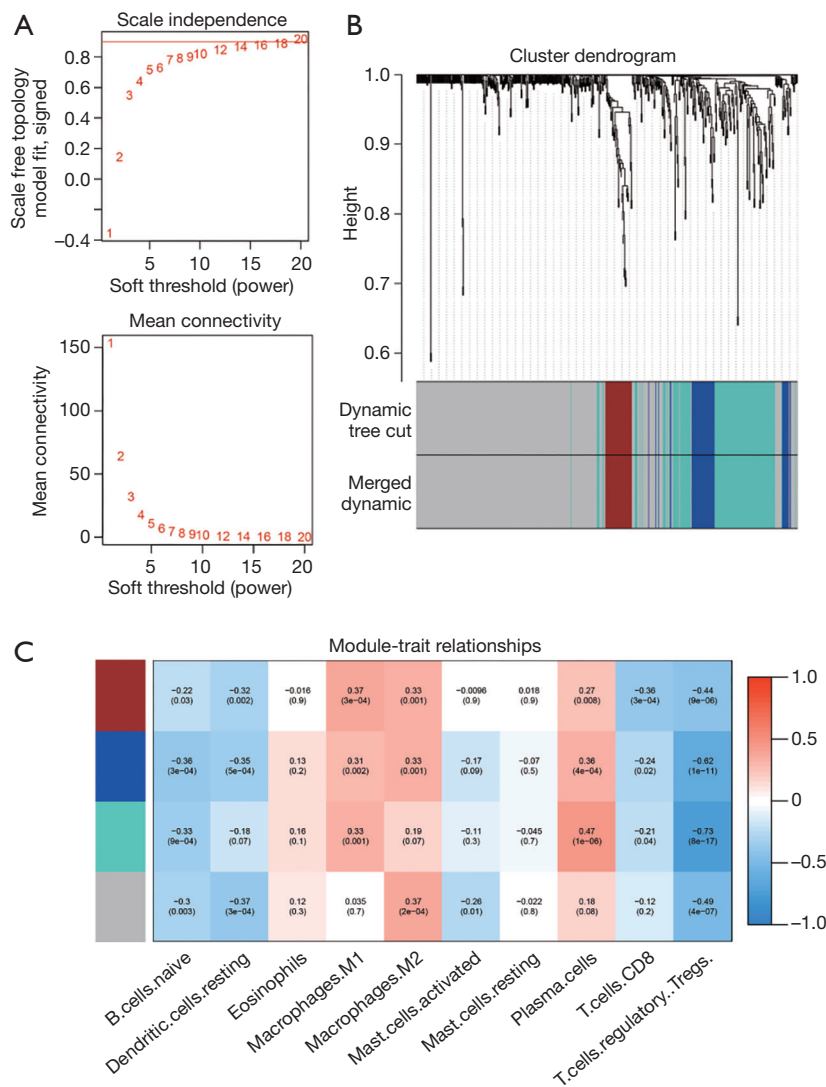
Page 8 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium

Figure 4 Weighted correlation network analysis (WGCNA) co-expression construction and key module identification. (A) Selection of the most appropriate power values to construct a hierarchical clustering. Analysis of the scale-free fit index of soft threshold power from 1–20, and analysis of the mean connectivity of 1–20 soft threshold power. (B) Genes were grouped into various modules by hierarchical clustering according to dynamic tree cut, and merged when the module's correlation >0.8. Different colors represent different modules. (C) Heatmap shows the correlation of module eigengenes with 10 kinds of differentially infiltrating immune cells in the tubulointerstitium of lupus nephritis (LN). The correlation parameter of R^2 and P value are labeled in the box.

cluster analysis. Firstly, a dendrogram was generated using dynamic hybrid cutting (*Figure 4B*). Each leaf on behalf of a single gene gathered together and formed a branch of the tree when those genes had close expression spectrum. The branch represented a gene module. We then also calculated eigengenes for every module to cluster the module in parallel, specifically once the correlation was above 0.8, those modules were merged together. As shown in the cluster dendrogram (*Figure 4B*), we obtained 3 modules excluding the nonsense one (grey). We calculated average linkage and Pearson's correlation value between modules and DIICs (Table S3). There was a significant negative correlation between turquoise and T. cells. regulatory. Tregs, with R^2=0.73 (*Figure 4C*). In summary, these results show MEturquoise can be recognized as key target module.

Table 1 The gene number in each module based on DEG

Module	MEbrown	MEblue	MEturquoise	MEgrey
Gene number	37	48	131	314

DEG, differentially expressed gene.

Identification of immune markers by PPI network

The highly connected genes of the module were recognized as potential key factors related to LN interstitial lesions. Therefore, using the STRING database, we input MEturquoise genes for PPI network analysis. Cut-offs of combined score >0.6 and connectivity >15 (node/edge) helped us to identify 6 hub genes (as set 2), and the network was visualized by Cytoscape v3.4.0 (*Figure 5A*, green node gene). Meanwhile, 29 hub genes were investigated by cut-offs standard (module-membership >0.8 and gene significance for T. cells. regulatory. Tregs. >0.6) from the key module (as set 1, *Figure 5B*, red box). In order to obtain the immune markers, we selected the overlapping set from the WGCNA key module hub gene (set 1) and PPI hub gene (set 2) (*Figure 5C*). The Venn diagram listed 5 overlapping genes, which were *TYROBP*, *C1QB*, *LAPTM5*, *CTSS*, and *PTPRC* separately.

Clinical correlation analysis with target immune markers

To determine the clinical applicability of these markers, we first queried the Nephroseq database to obtain the correlation between 5 immune markers and GFR. The scatter plot displays all of the immune markers that had a negative correlation with GFR (*Figure 6A*). The correlation values r were −0.601, −0.546, −0.58, −0.626, and −0.563 separately, and all of the results were significant at the p=0.001 level. We attempted to further correlate the pathological classification with the expression of each candidate immune marker in GSE127797, and all data were calculated using the Wilcoxon rank test (*Figure 6B*). The pathological class of GSE127797 is shown in *Table 2*. We grouped the pathological class by III and IV as proliferative LN (PLN), and V as membranous LN (MLN). The genes TYROBP and C1QB showed statistical differences between the two groups. When we set a combined group PLN/MLN as class III + V, IV + V with the above two single groups, the in group ANOVA analysis was different at the P=0.05 level (Table S4). The analysis verified that the candidate immune markers had a strong association with LN pathological diagnosis and prognosis.

Unsupervised clustering of the molecular patterns for LN

Based on well-studied immune marker expression spectrums, unsupervised clustering was carried out for validation. Prior to this process, the within groups sum of squares was calculated to choose the appropriate cluster number. In *Figure 7A*, the sum of squares sloped gently when the cluster number was 5 (*Figure 7A*, left). By applying k-means software, all of the samples were divided into 5 clusters (*Table 3*), and thereafter data were further dimension reduced by t-SNE and visualized by a heatmap (*Figure 7A*, right). As the clusters were performed on candidate immune markers, we found 5 kinds of clusters had significantly different populations in PCA. Meanwhile, the immune markers divided all samples by visible boundaries in the heatmap plot (*Figure 7B*), which confirmed the candidate immune markers could indicate a molecular pattern classification for LN. In the second part, we retrieved 22 kinds of immune cells by newly identified clusters for the purposes of subcluster DIIC validation (*Figure 8A*). The cell proportions distributed differently in each cluster from 1–5, specifically the in-group ANOVA indicated that T.cells. regulatory.Tregs., B.cells.naive, Plasma.cells, NK.cells. resting, Macrophages.M1, and Dendritic.cells.resting had significant differences among 5 clusters (Table S5). Comparing the 2 parts of the results, it could be seen that these immune markers directed molecular clusters could divide the samples not only by gene expression level, but also by immune cell infiltration rate. Lastly, the new clusters were further validated by the expression levels of the 5 immune markers (*Figure 8B*). All of the candidate genes showed different expression levels among the 5 clusters. Surprisingly, cluster 4, in which the plasma cell infiltrating proportion was the highest and Treg was the lowest, showed a relatively higher expression level of *TYROBP*, *C1QB*, *LAPTM4*, *CTSS*, and *PRPRC* than the other clusters. These results validated that our candidate immune markers and immune related cluster patterns had the power to distinguish the detailed molecular patterns of LN.
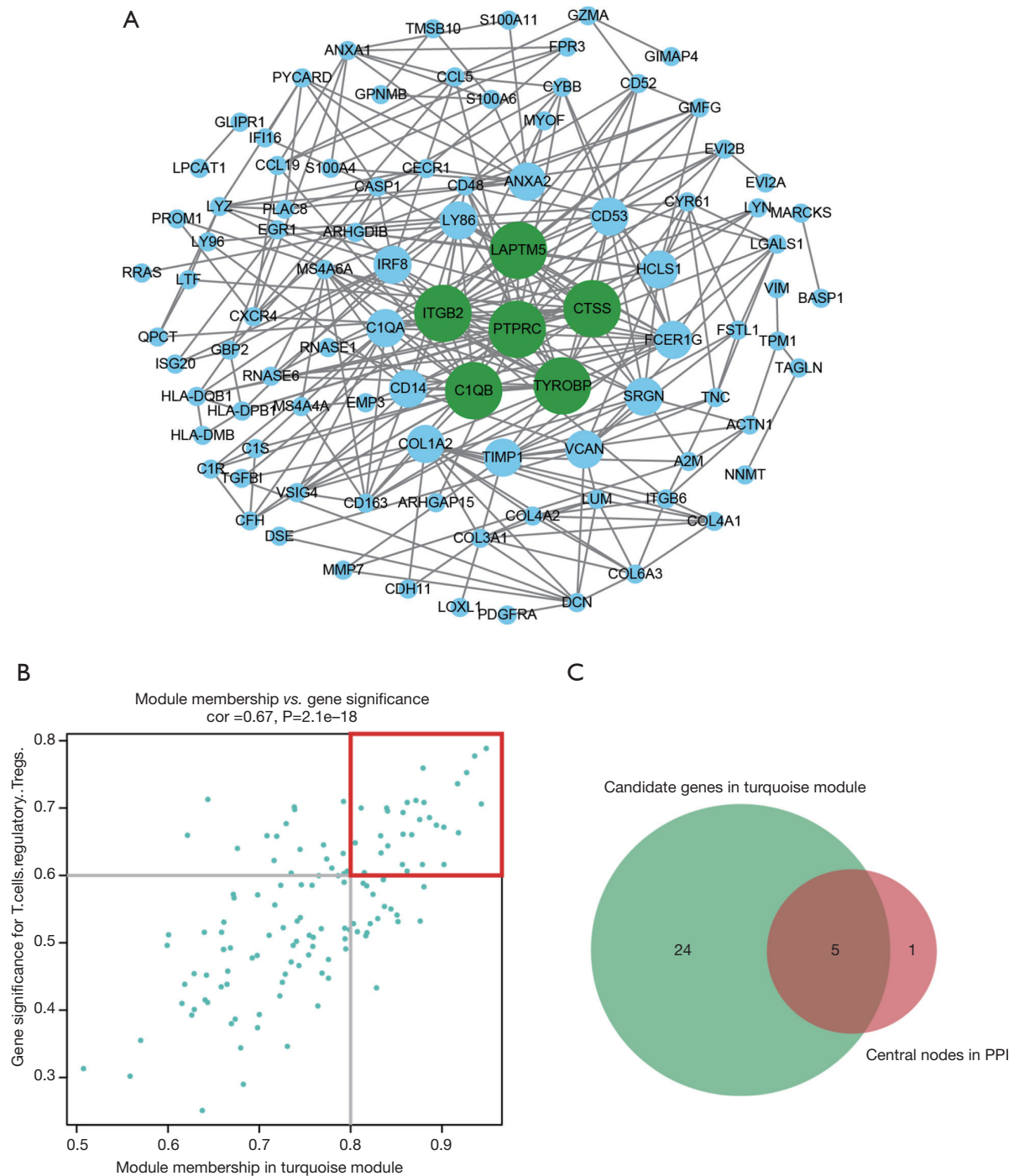
Page 10 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium

Figure 5 Identification of immune markers based on hub genes and protein-protein interaction (PPI) analysis. (A) PPI network of genes from the key turquoise module. The higher the number of connected nodes, the larger the size of the node. The green nodes were recognized as a central node with more than 15 connections (degree >15, set 2), and the blue node was recognized as 10≤ degree ≤15. (B) A scatter plot of genes in the turquoise module, and were Treg related. Each green dot represents a gene, and dots within the red box indicate these genes had module membership >0.8 in the turquoise module, meanwhile their gene significance for Treg >0.6 (set 1). (C) The 5 immune markers were visualized by a Venn diagram, shown as the overlap between central nodes in PPI (set 2) and hub genes in the turquoise module by weighted correlation network analysis (WGCNA) (set 1).
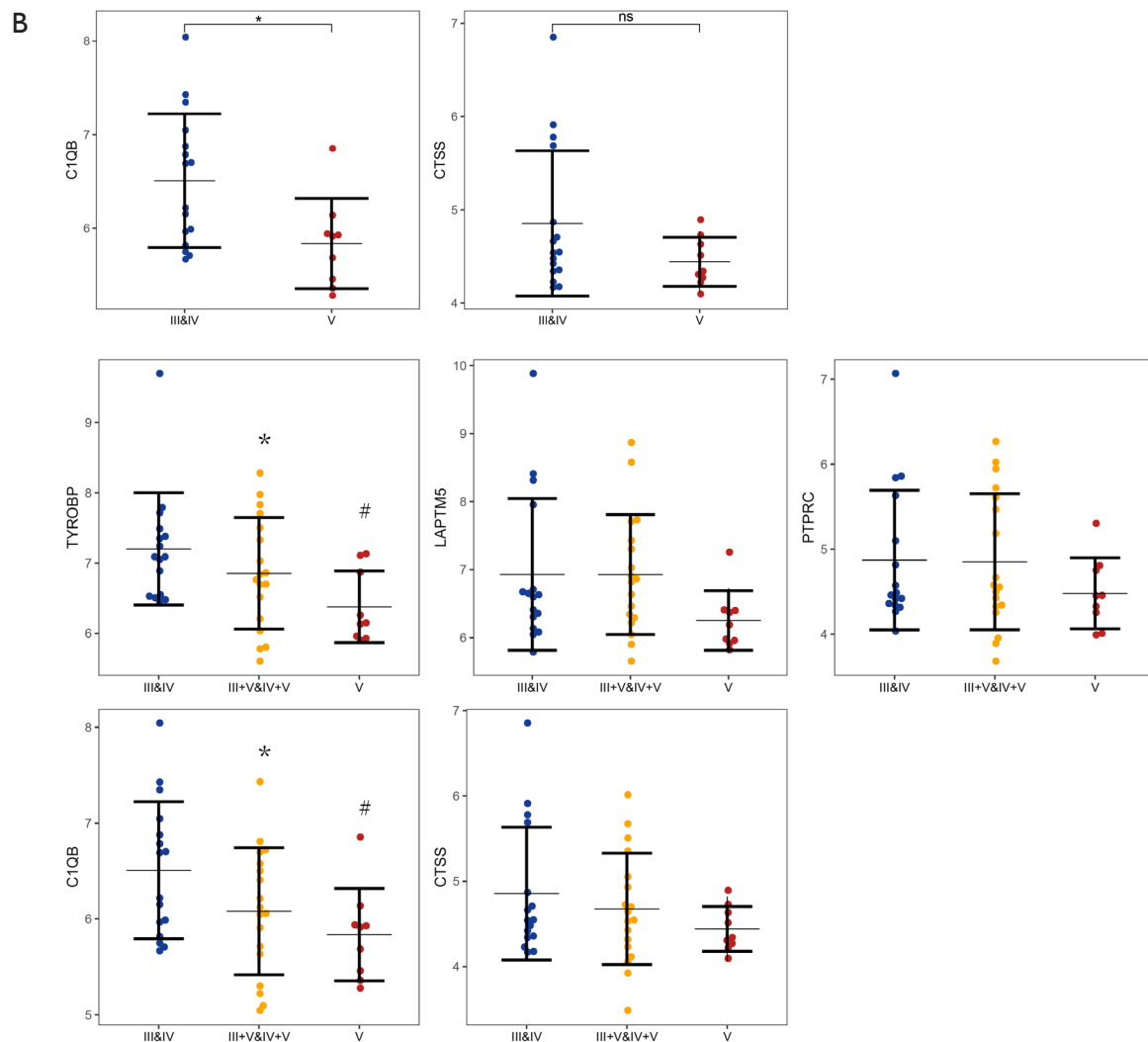
Page 12 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium
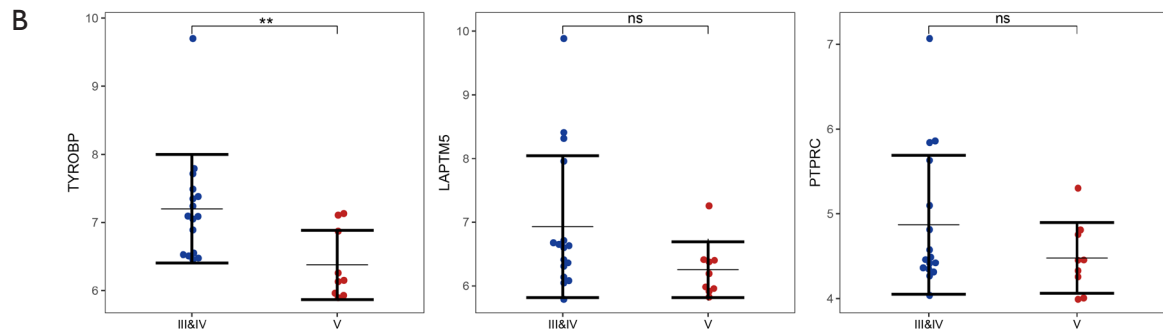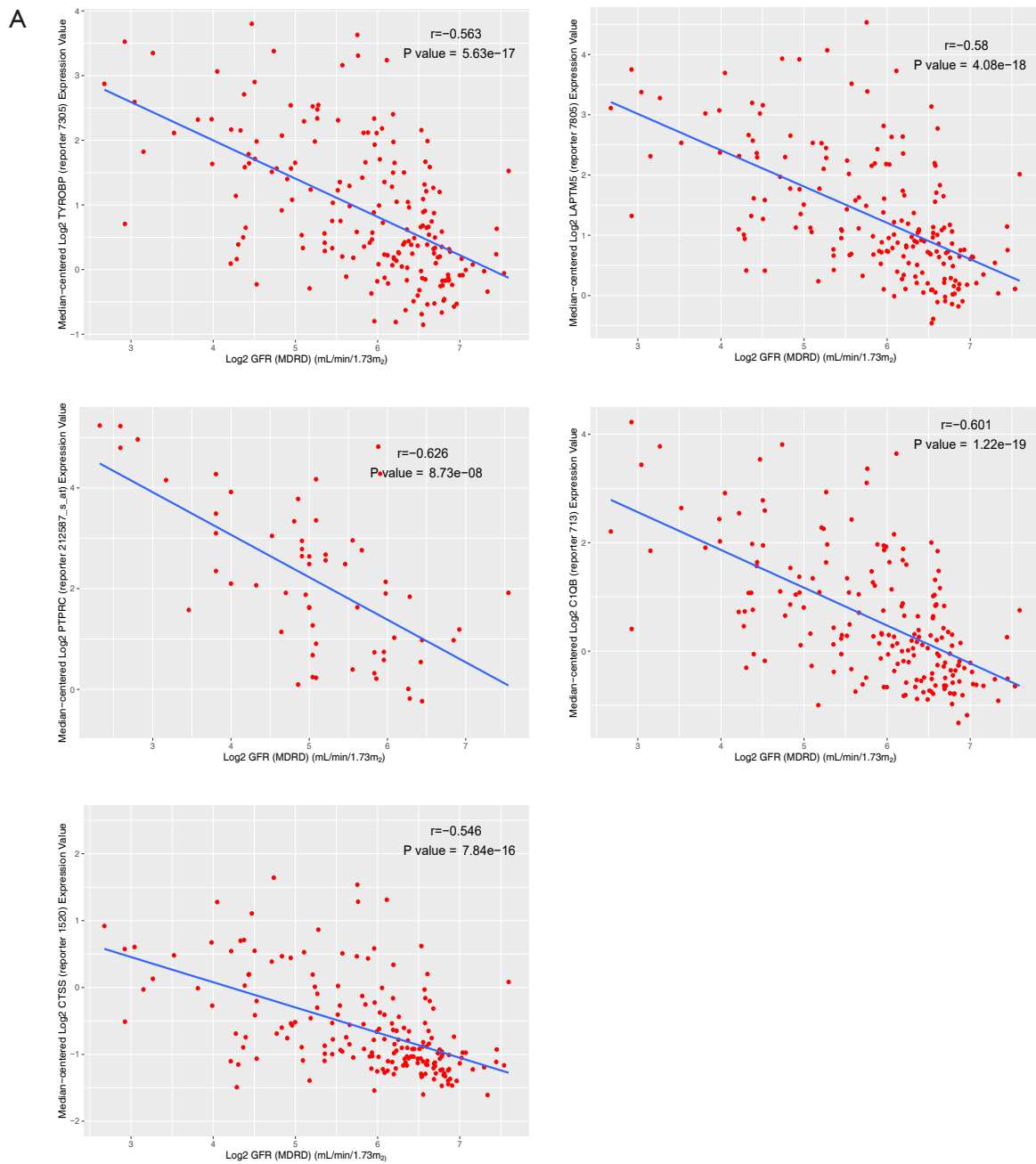
Figure 6 Clinical correlation between immune markers, renal function, and pathological classification. (A) Correlation between selected immune markers with renal function (glomerular filtration rate) from the Nephroseq database. (B) Dot plot of the immune marker expression levels between the PLN group (class III, IV) and MLN group (class V), as well as III + V, IV + V as a combined PLN/MLN group. The Wilcoxon rank test was run between PLN and MLN, and a one-way ANOVA test was performed to test the differences among all groups. *P<0.05, **P<0.01, PLN (proliferative LN class III, IV) *vs*. MLN (membranous LN class V). *P<0.05, compared with PLN and MLN group; #P<0.05,compared with PLN and combined PLN/MLN group.

Table 2 The LN biopsy sample pathological class in GSE127797

LN pathologic class	I	III	IV	V	III + V	IV + V	NA
Samples	1	3	6	13	12	9	3

GSE127797 had 44 renal biopsy samples, the number of each classification has listed in the table. LN, lupus nephritis.
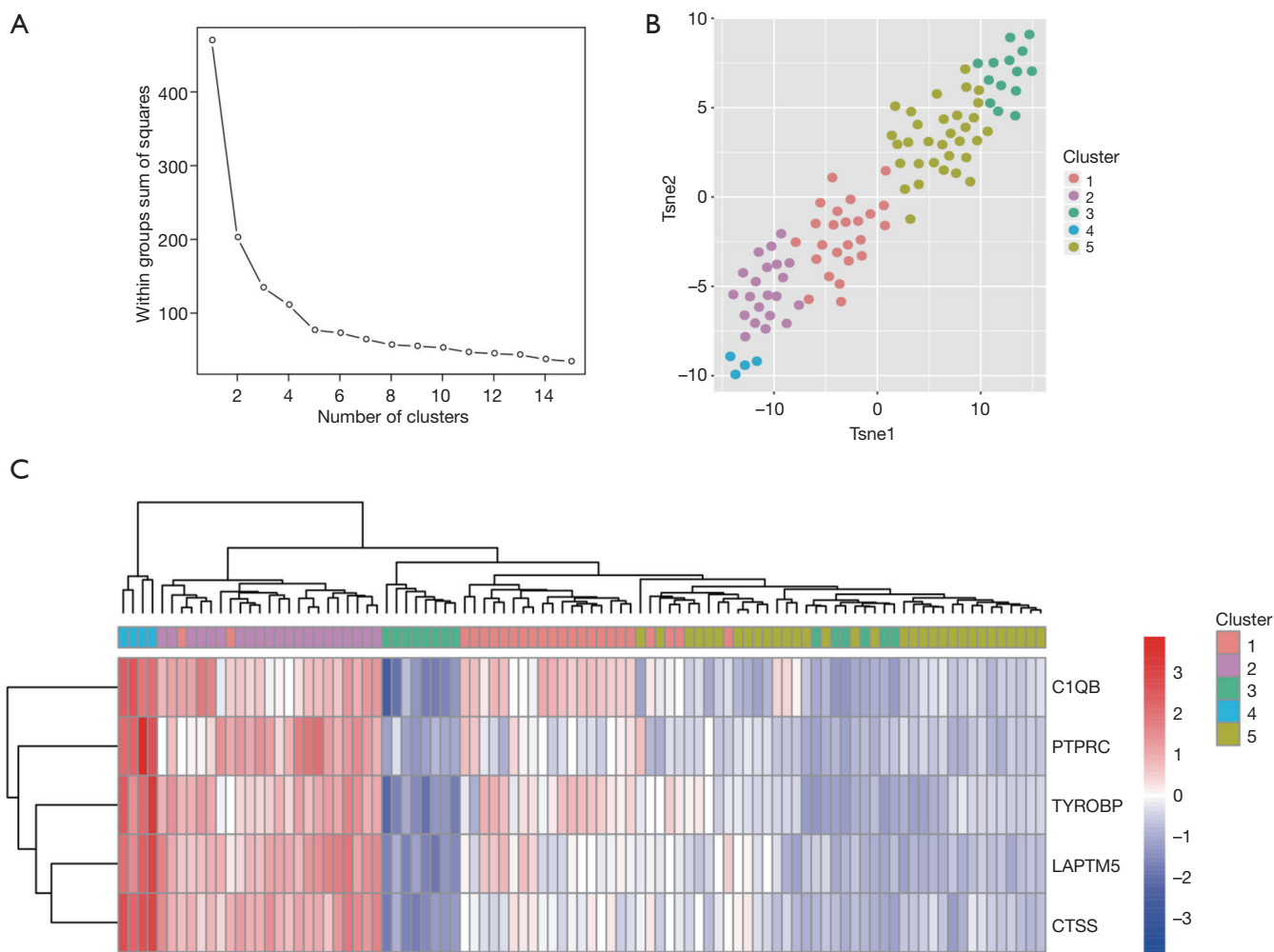
Figure 7 Unsupervised clustering for molecular pattern identification. (A) Elbow plot for determining the appropriate unsupervised cluster number by k-means. The within sum of squares (WSS) suddenly dropped as the number of clusters increased from 2 to 5. Therefore, the bend at k=5 provided stability for WSS. (B) Principal component analysis (PCA) by t-SNE demonstrated the presence of distinct expression patterns based on subclusters. (C) Heatmap showing 5 immune markers in each cluster.

Table 3 The gene number assigned in each cluster

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Gene number	24	21	14	4	32

Discussion

SLE is highly heterogeneous autoimmune disorder with many subsets. Scarce literature exists on the differences between the glomerular and tubulointerstitial transcriptome of LN. An initial objective of this study was to identify precise diagnostic markers and unveil novel molecular patterns for LN T2T. In the current study, we screened *GSE32591*, *GSE69438*, and *GSE127797* gene datasets of LN. As immune infiltration in glomeruli has been well-studied by several groups (34,35), we set out to assess the importance of the immune microenvironment in LN tubulointerstitial lesions. One single cell sequence study reported kidney infiltrating immune cells mainly
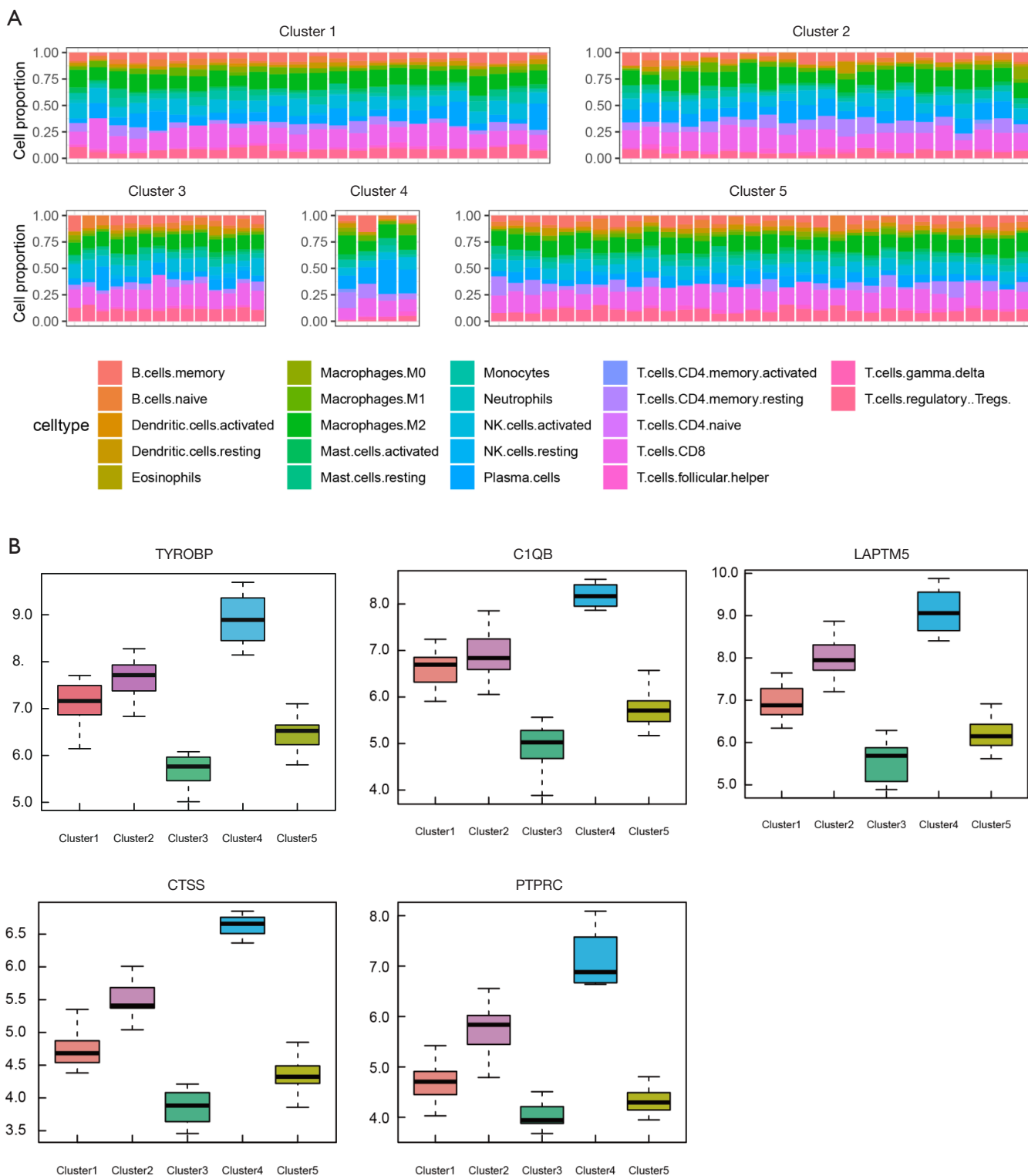
Figure 8 Immune cell infiltration and immune marker expression levels according to molecular pattern clusters. (A) Bar plot demonstrating the different fractions of immune cells in individual molecular pattern clusters. (B) Boxplot indicates 5 immune cell marker expression levels in individual molecular patterns.

included 3 categories: T cells (64.2%), B cells (10.4%), and macrophages (25.4%) (36). Kidney T and B cells are recruited from the peripheral immune system, but kidney-resident macrophages (KRM), a special type of macrophage, are characterized as CD45+CD11b+F4/80^{high}, which differentiated from embryonic macrophages (yolk sac) and monocytes (bone marrow macrophage) (37). Therefore, we believe that LN tubular and interstitial lesions have their own characteristics, and can affect LN prognosis by their crosstalk with immune cells. On the question of the LN infiltrating immune cell landscape, our study found that 10 kinds of immune cells had infiltration fraction differences between LN and normal controls, which is generally consistent with the abovementioned previous study (36). Treg cells, as the vital suppressive immune cell type, is downregulated in LN, and plasma cells, as autoantibody producers, are massively upregulated in LN. The reduction of CD8+T cells in LN could be attributed to functional exhaustion (36), and activation of both M1 and M2 macrophages might be explained by the mechanisms of lupus renal inflammation (M1 macrophages secrete TNF-α and ROS) and repair/fibrosis (M2 secrete TGF-β and IL10) processes (37). Mast cells are known to be SLE-related inflammation regulators, but their role in LN should be interpreted with caution (38), as well as eosinophils and resting dendritic cells (39).

Another important aspect of our study focused on DEGs and pathway enrichment analysis. As explained in the results, the pathway analysis based on DEGs mainly covered IFNs, viral response, collagen matrix deposition, and autoimmune diseases. These results corroborate the findings of multiple previous studies in SLE and LN (40). As our aim was to investigate immune markers of LN, we further utilized WGCNA to explore the key modules and hub genes from DEGs. The most obvious finding to emerge from the network construction analysis is that key modules (as a population of genes that have related functions) were highly correlated with Treg cells. Next, the immune markers were obtained from the overlapping set of key modules, and Treg positively correlated with both the gene sets and the PPI gene set. *TYROBP*, *C1QB*, *LAPTM5*, *CTSS*, and *PTPRC* were negatively correlated with GFR. Specifically, TYROBP and C1QB had significant differences between PLN (class III, IV) and MLN (class V). These results suggest that the immune markers have the ability to estimate prognosis by GFR and pathological class. Last but not least, the most clinically applicable finding was that immune markers could divide the enrolled LN samples

into 5 separate clusters, and had the ability to indicate the molecular pattern for LN by their expression level. These results further support the possibility that unsupervised clustering as a machine learning method can help to validate immune markers for clinical application. These findings will also help us to develop molecular diagnostic markers based on immune marker expression levels in large scale studies.

TYROBP (TYRO protein tyrosine kinase binding protein, also known as DAP12) is a transmembrane immune signaling adaptor. It is enriched in TREM-1 signaling by Ingenuity Pathway Analysis (IPA), and is highly correlated with proteinuria in a lupus mouse model (41). Recent single cell data identified it as a marker for the NK cell population in LN (16). C1QB stands for complement C1q B chain. Anti-C1q antibodies are associated with PLN, specifically class IV (42). Anti-ghB (globular head region of B chain) C1q antibody is strongly correlated with proteinuria and serum albumin (43). LAPTM5 (lysosomal-associated multi spanning membrane protein 5) positively regulates proinflammatory signaling pathways by facilitating NF-kB and MAPK signaling, as well as proinflammatory cytokine production in macrophages (44). CTSS (cathepsin S) is lysosomal cysteine proteinase that participates in the degradation of antigenic proteins to peptides for presentation on MHC class II molecules (45). Studies have shown that it can promote SLE and LN through regulating CD4 T cell and B cell priming (46), as well as modulating T$_{FH}$ cell repertoire (47). Finally, PTPRC (protein tyrosine phosphatase receptor type C, also known as CD45) is an essential regulator of T and B cell antigen receptor signaling (48). In accordance with the present results, previous studies have addressed the relationship between these proteins and LN, but the mechanism behind their involvement is still unclear.

Our study has some limitations. Of particular note, renal parenchymal cells also play a vital role in the mechanism of LN. In the present study, we did not further describe the crosstalk between parenchymal and infiltrating immune cells. These findings may be somewhat limited by the small-scale bioinformatics analysis. Next, the 3 GSE studies we enrolled did not cover pathological class. Hence, large-scale pathological class validation is urgently needed in future studies. Furthermore, the immune cluster based on 5 immune markers needs to be validated in a large population of LN patients to better evaluate their diagnostic ability. Although we tried to compare the glomeruli and tubulointerstitium data in parallel to produce more convincing and solid data, the batch effect could not

Page 16 of 18

Zhang et al. Immune cell infiltration in lupus nephritis tubulointerstitium

be corrected completely, so we only focused on lesions of interest in this study. Further investigations are needed to identify the differences and similarities between tubules and glomeruli, which can discover the tissue specific targets for LN. Lastly, further work is needed to distinguish the immune markers for LN with other immune related renal diseases like IgA nephropathy (IgAN) and membranous nephropathy (MN). Although bioinformatics can shed some light on the intrinsic mechanisms, molecular biology experiments and clinical analysis needs to be performed to further validate these findings.

Conclusions

This study set out to objectively analyze and assess LN tubulointerstitial immune cell infiltration, and seek potential diagnostic immune markers and molecular patterns using bioinformatics methods. Our findings unveiled the main infiltrating immune cells in the tubulointerstitium and their changes under LN. WGCNA recognized the T.cells. regulatory.Tregs related module as the key module, and 5 immune markers: TYROBP,C1QB,LAPTM5, CTSS, and PTPRC. All of them demonstrated a negative correlation with GFR, while TYROBP and C1QB demonstrated significant expression level differences between PLN and MLN. Unsupervised clustering validated that these immune markers could cluster LN by their expression level and immune cell infiltration rate. Our findings clearly indicate that candidate immune markers can help to distinguish the detailed molecular patterns of LN, and are highly related to prognosis.

Acknowledgments

Footnote

Reporting Checklist: Available at http://dx.doi.org/10.21037/atm-20-7507

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-7507). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Parikh SV, Almaani S, Brodsky S, et al. Update on Lupus Nephritis: Core Curriculum 2020. Am J Kidney Dis 2020;76:265-81.
2. Anders HJ. Re-biopsy in lupus nephritis. Ann Transl Med 2018;6:S41.
3. Bajema IM, Wilhelmus S, Alpers CE, et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int 2018;93:789-96.
4. Broder A, Mowrey WB, Khan HN, et al. Tubulointerstitial damage predicts end stage renal disease in lupus nephritis with preserved to moderately impaired renal function: A retrospective cohort study. Semin Arthritis Rheum 2018;47:545-51.
5. Rao DA, Arazi A, Wofsy D, et al. Design and application of single-cell RNA sequencing to study kidney immune cells in lupus nephritis. Nat Rev Nephrol 2020;16:238-50.
6. Wu CY, Chien HP, Yang HY, et al. Role of tubulointerstitial lesions in predicting renal outcome among pediatric onset lupus nephritis - A retrospective cohort study. J Microbiol Immunol Infect 2020;53:33-41.
7. Freedman BI, Langefeld CD, Andringa KK, et al. End-stage renal disease in African Americans with lupus nephritis is associated with APOL1. Arthritis Rheumatol 2014;66:390-6.
8. Larsen CP, Beggs ML, Saeed M, et al. Apolipoprotein L1 risk variants associate with systemic lupus erythematosus-associated collapsing glomerulopathy. J Am Soc Nephrol

2013;24:722-5.

9. Iwamoto T, Niewold TB. Genetics of human lupus nephritis. Clin Immunol 2017;185:32-9.

10. Munroe ME, James JA. Genetics of Lupus Nephritis: Clinical Implications. Semin Nephrol 2015;35:396-409.

11. Flores-Mendoza G, Sanson SP, Rodriguez-Castro S, et al. Mechanisms of Tissue Injury in Lupus Nephritis. Trends Mol Med 2018;24:364-78.

12. Espeli M, Bokers S, Giannico G, et al. Local renal autoantibody production in lupus nephritis. J Am Soc Nephrol 2011;22:296-305.

13. Aringer M, Leuchten N, Schneider M. Treat to Target in Systemic Lupus Erythematosus. Rheum Dis Clin North Am 2019;45:537-48.

14. Rios-Garces R, Espinosa G, van Vollenhoven R, et al. Treat-to-target in systemic lupus erythematosus: Where are we? Eur J Intern Med 2020;74:29-34.

15. Tsokos GC, Lo MS, Costa Reis P, et al. New insights into the immunopathogenesis of systemic lupus erythematosus. Nat Rev Rheumatol 2016;12:716-30.

16. Arazi A, Rao DA, Berthier CC, et al. The immune cell landscape in kidneys of patients with lupus nephritis. Nat Immunol 2019;20:902-14.

17. Almaani S, Rovin BH. B-cell therapy in lupus nephritis: an overview. Nephrol Dial Transplant 2019;34:22-9.

18. Gregersen JW, Jayne DR. B-cell depletion in the treatment of lupus nephritis. Nat Rev Nephrol 2012;8:505-14.

19. Ronnblom L, Alm GV, Eloranta ML. The type I interferon system in the development of lupus. Semin Immunol 2011;23:113-21.

20. Ronnblom L, Leonard D. Interferon pathway in SLE: one key to unlocking the mystery of the disease. Lupus Sci Med 2019;6:e000270.

21. Salemme R, Peralta LN, Meka SH, et al. The Role of NETosis in Systemic Lupus Erythematosus. J Cell Immunol 2019;1:33-42.

22. Yu Y, Su K. Neutrophil Extracellular Traps and Systemic Lupus Erythematosus. J Clin Cell Immunol 2013;4.

23. Sakuma Y, Nagai T, Yoshio T, et al. Differential activation mechanisms of serum C5a in lupus nephritis and neuropsychiatric systemic lupus erythematosus. Mod Rheumatol 2017;27:292-7.

24. Furie R, Rovin BH, Houssiau F, et al. Two-Year, Randomized, Controlled Trial of Belimumab in Lupus Nephritis. N Engl J Med 2020;383:1117-28.

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

26. Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 2013;41:D991-5.

27. Berthier CC, Bethunaickan R, Gonzalez-Rivera T, et al. Cross-species transcriptional network analysis defines shared inflammatory responses in murine and human lupus nephritis. J Immunol 2012;189:988-1001.

28. Ju W, Nair V, Smith S, et al. Tissue transcriptome-driven identification of epidermal growth factor as a chronic kidney disease biomarker. Sci Transl Med 2015;7:316ra193.

29. Almaani S, Prokopec SD, Zhang J, et al. Rethinking Lupus Nephritis Classification on a Molecular Level. J Clin Med 2019;8:1524.

30. Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28:882-3.

31. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods 2015;12:453-7.

32. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.

33. Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284-7.

34. Cao Y, Tang W, Tang W. Immune cell infiltration characteristics and related core genes in lupus nephritis: results from bioinformatic analysis. BMC Immunol 2019;20:37.

35. Cao Y, Mi X, Wang Z, et al. Bioinformatic analysis reveals that the OAS family may play an important role in lupus nephritis. J Natl Med Assoc 2020. [Epub ahead of print]. doi: 10.1016/j.jnma.2020.05.006.

36. Tilstra JS, Avery L, Menk AV, et al. Kidney-infiltrating T cells in murine lupus nephritis are metabolically and functionally exhausted. J Clin Invest 2018;128:4884-97.

37. Liu F, Dai S, Feng D, et al. Distinct fate, dynamics and niches of renal macrophages of bone marrow or embryonic origins. Nat Commun 2020;11:2280.

38. Kaczmarczyk K, Kosalka J, Soja J, et al. Renal interstitial mast cell counts differ across classes of proliferative lupus nephritis. Folia Histochem Cytobiol 2014;52:218-24.

39. Kaewraemruaen C, Ritprajak P, Hirankarn N. Dendritic cells as key players in systemic lupus erythematosus. Asian Pac J Allergy Immunol 2019. [Epub ahead of print]. doi: 10.12932/AP-070919-0639.

40. Der E, Suryawanshi H, Morozov P, et al. Tubular cell and

keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. Nat Immunol 2019;20:915-27.

41. Bethunaickan R, Berthier CC, Zhang W, et al. Identification of stage-specific genes associated with lupus nephritis and response to remission induction in (NZB x NZW)F1 and NZM2410 mice. Arthritis Rheumatol 2014;66:2246-58.

42. Tan Y, Song D, Wu LH, et al. Serum levels and renal deposition of C1q complement component and its antibodies reflect disease activity of lupus nephritis. BMC Nephrol 2013;14:63.

43. Radanova M, Vasilev V, Deliyska B, et al. Anti-C1q autoantibodies specific against the globular domain of the C1qB-chain from patient with lupus nephritis inhibit C1q binding to IgG and CRP. Immunobiology 2012;217:684-91.

44. Glowacka WK, Alberts P, Ouchida R, et al. LAPTM5 protein is a positive regulator of proinflammatory signaling pathways in macrophages. J Biol Chem

2012;287:27691-702.

45. Tato M, Kumar SV, Liu Y, et al. Cathepsin S inhibition combines control of systemic and peripheral pathomechanisms of autoimmune tissue injury. Sci Rep 2017;7:2775.

46. Rupanagudi KV, Kulkarni OP, Lichtnekert J, et al. Cathepsin S inhibition suppresses systemic lupus erythematosus and lupus nephritis because cathepsin S is essential for MHC class II-mediated CD4 T cell and B cell priming. Ann Rheum Dis 2015;74:452-63.

47. Kim SJ, Schatzle S, Ahmed SS, et al. Increased cathepsin S in Prdm1(-/-) dendritic cells alters the TFH cell repertoire and contributes to lupus. Nat Immunol 2017;18:1016-24.

48. Courtney AH, Shvets AA, Lu W, et al. CD45 functions as a signaling gatekeeper in T cells. Sci Signal 2019;12:eaaw8151.

(English Language Editor: C. Betlazar-Maseh)