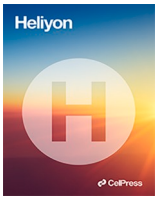
Heliyon 10 (2024) e38424

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Integrated bioinformatics analysis for identifying fibroblast-associated biomarkers and molecular subtypes in human membranous nephropathy

Chuying Gui [*,1], Sidi Liu [1], Zhike Fu, Huijie Li, Di Zhang, Yueyi Deng [**]

The Department of Nephrology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, 200032, China

ARTICLE INFO

Keywords:
Membranous nephropathy
Fibroblasts
Biomarkers
Molecular subtypes
Integrated bioinformatics

ABSTRACT

Background: Membranous nephropathy (MN) is characterized by immune complex deposition in the glomerular basement membrane, leading to proteinuria and potentially progressive renal dysfunction. Fibroblasts have been implicated in the pathogenesis of MN through their involvement in tissue remodeling and immune modulation.

Methods: We employed integrated bioinformatics analyses to identify fibroblast-associated biomarkers and molecular subtypes in MN. The xCell algorithm was used to assess fibroblast infiltration, and weighted gene co-expression network analysis (WGCNA) identified fibroblast-related gene modules. Differentially expressed fibroblast-associated genes (DEFAGs) were screened between MN and healthy controls (HC) using differential expression analysis and protein-protein interaction (PPI) network construction. Consensus clustering categorized MN patients into distinct subtypes based on DEFAG expression profiles.

Results: Fibroblast scores were a significant elevation in MN compared to HC, indicating increased fibroblast involvement in MN pathogenesis. WGCNA identified 13 fibroblast-related gene modules, with the brown and turquoise modules showing strong correlation with fibroblast scores (correlation coefficient = 0.79 and 0.75, respectively, $p < 0.01$). DEFAG analysis revealed 308 genes overlapping between WGCNA and differentially expressed genes (DEGs) in MN. Consensus clustering identified two molecular subtypes (C1 and C2) based on DEFAG expression patterns, with differential gene expression enriching pathways related to immune response and extracellular matrix remodeling. Core biomarker analysis highlighted COL3A1 and TGFB1 as central genes associated with MN, with elevated expression validated across multiple datasets.

Conclusion: Integrated bioinformatics analysis identified fibroblast-associated molecular subtypes in MN, revealing distinct immune profiles and biomarkers. COL3A1 emerged as a potential diagnostic and therapeutic target, implicating its role in immune regulation and disease progression in MN.

* Corresponding author. The Department of Nephrology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, NO.725, South Wanping Road, Xuhui District, Shanghai, 200032, China.

** Corresponding author. The Department of Nephrology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, NO.725, South Wanping Road, Xuhui District, Shanghai, 200032, China.

E-mail addresses: guichuy0231@aliyun.com (C. Gui), sh_dengyueyi@163.com (Y. Deng).

[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38424
Received 15 July 2024; Received in revised form 22 September 2024; Accepted 24 September 2024
Available online 25 September 2024

1. Introduction

Membranous nephropathy (MN) is the primary cause of the adult nephrotic syndrome. It is characterized by the formation of glomerular deposits that contain immunoglobulin components, which are typical for this condition [1,2]. The incidence of MN globally is approximately 1 in 100,000, with 80 % of patients being diagnosed with primary membranous nephropathy [3]. Currently, available therapies are ineffective in treating MN, and even in MN cases where complete remission is achieved, 25%–30 % of patients still experience recurrent MN [4]. Moreover, the heterogeneity in patient outcomes, with 20%–30 % eventually developing end-stage renal disease, highlights the limitations in current therapeutic approaches and the need for more precise biomarkers and treatment strategies [5]. Previous studies have often focused on isolated aspects of MN pathology without integrating comprehensive data on immune infiltration and fibroblast activity. The identification and development of MN-related biomarkers hold significant clinical value in diagnosing, prognosis, and treating of MN patients [6]. Therefore, it is imperative to investigate potential biomarkers and biological mechanisms associated with MN to enhance the effectiveness of its diagnosis and treatment.

MN is progressive inflammation and immune-related disease of the kidney. This condition is primarily triggered by an autoimmune response, which leads to changes in podocyte morphology, glomerular basement membrane thickness, and immune complex deposition. These changes ultimately result in proteinuria [7]. The majority of findings suggest that the origin of MN is closely linked to the functions of adaptive and innate immune cells in the kidney [8,9]. However, the definite mechanisms underlying MN pathogenesis remain unclear, as do the specific roles of various cellular components in disease progression. Studies have shown a clear link between experimental MN and the gradual deterioration of kidney function, characterized by interstitial inflammation, fibrosis, and an increase in the size of glomeruli [10]. While the pathogenesis of MN involves the generation of an immune complex, the definite mechanism behind this process remains unclear. Fibroblasts are vital components of stromal cells and are responsible for maintaining structural integrity [11]. They also play a crucial role in regulating immune responses during infection, chronic inflammation, and cancer by controlling the behavior of local immune cells within the tissue microenvironment. Fibroblasts are known to be heterogeneous cells, consisting of different populations that perform various functions. In addition to their diverse immunological properties, fibroblasts have been found to maintain an effective inflammatory environment in cases of chronic inflammation, promote immunosuppression in tumors [12]. A previous report has revealed that inflammation-related fibroblasts are involved in chronic persistent inflammation. This is due to the abnormal generation of extracellular matrix, chemokines, and cytokines [13]. In addition, a recent study has demonstrated that understanding the heterogeneity of fibroblasts can aid researchers in gaining deeper disease pathology, including aging, fibrotic disease, and wound healing [11]. Nonetheless, the potential association between fibroblasts and MN, particularly the role of fibroblasts in immune cell infiltration during MN progression, remains underexplored in existing studies. As a result, a more accurate MN classification is necessary in terms of disease definition and patient subdivision. The primary aim of this study was to identify molecular subtypes of MN to enhance the potential diagnostic markers.

In recent years, the combination of transcriptome sequencing and bioinformatics analysis has been utilized by researchers to identify potential biomarkers and functional pathways involved in various diseases. However, many studies have lacked comprehensive integration of genetic data with immune cell profiling and fibroblast activity, making it challenging to understand the full scope of disease mechanisms [14–18]. Our study aims to identify subpopulations with distinct pathological mechanisms and signature genes by integrating genetic data, immune cell infiltration, and phenotypic data. We downloaded the gene expression profiles of MN from the Gene Expression Omnibus (GEO) databases and assessed the level of fibroblast infiltration in the MN and healthy controls (HC) groups using the xCell algorithm. Additionally, we used weighted gene correlation network analysis (WGCNA) to identify modules and genes related to fibroblasts. Then, differentially expressed fibroblast-associated genes (DEFAGs) were identified via differentially expressed analysis. Based on the gene expression profiles of DEFAGs, the MN dataset was clustered to differentiate MN patients. Subsequently, GSVA, GSEA, enrichment function analysis, and immune cell infiltration were performed to explore the differential molecular biological functions between the different subtypes. Finally, the core genes were identified and validated in multiple datasets, and the correlation between core genes and immune cell infiltration was also assessed. Our findings could provide

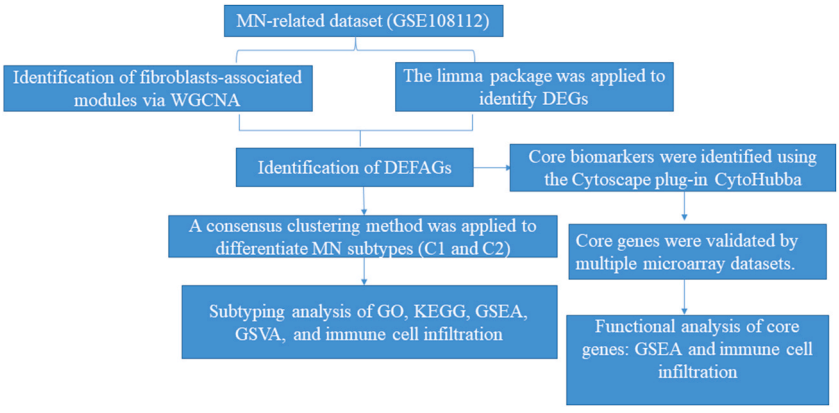


Fig. 1. The flowchart for the identification of potential molecular subtypes and core biomarkers for MN patients.

novel insight for MN diagnosis and management.

2. Materials and methods

2.1. Dataset information and data preprocessing

The flowchart of the present study is shown in Fig. 1. Our study exclusively focuses on datasets that satisfy the following conditions: 1) The MN group must consist of at least 10 participants; 2) The research must be conducted exclusively on Homo sapiens; 3) The raw or processed data must be available to the public; 4) The research must primarily aim to investigate gene expression profiles; 5) The study must incorporate renal samples from both MN patients and HC. We downloaded six eligible microarray datasets from the GEO database, including GSE108112, GSE104948, GSE108109, GSE115857, GSE104954, and GSE133288. GSE108112 was used as the discovery cohort, while the others were used as the validation cohorts. The details of these datasets are presented in Table 1. Prior to data analysis, we preprocessed the datasets using the AFFY package in R software to extract the original matrix data [19]. The annotation files were used to annotate the mRNA probes, which were converted to gene symbols. We used the robust multi-array averaging algorithm to perform background correction and normalization [20].

2.2. Identification and analysis of DEFAGs in MN

xCell is a gene signature-based method that infers the presence of different cell types in a heterogeneous sample [21]. We ran the xCell algorithm on the normalized gene expression data to estimate the enrichment scores for fibroblasts. MCPCounter is an alternative computational tool designed to quantify the abundance of various immune and stromal cell populations in gene expression data [22]. Similarly, we applied MCPCounter to the same GSE108112 dataset to independently calculate fibroblast enrichment scores. Subsequently, we performed the WGCNA using the "WGCNA" package in R to identify modules and genes associated with fibroblasts [23]. The "hclust" function was utilized to perform the hierarchical clustering analysis with a minimum module size of 30. The "pickSoftThreshold" function was employed to determine the soft thresholding power value, selecting a power of 2 based on the scale-free topology criterion. Additionally, the "blockwiseModules" function was used to build the co-expression network related to fibroblasts, with a deepSplit value of 3 and a mergeCutHeight of 0.25. The “limma” package (v.3.22.7) of R was applied to screen the differentially expressed genes (DEGs) between the HC and MN groups, the criteria used for analyzing DEGs included a p-value threshold of less than 0.05 and an absolute log fold change (FC) of at least 0.3 [24]. The “ggplot2” (v.3.3.6) and “heatmap” (v.2.13.1) packages were used to generate a volcano plot and heatmap of the DEGs, respectively.

2.3. Identification of core DEFAGs

The mutual relationship among the DEFAGs was explored using the STRING website. The TSV format file was downloaded and imported into the Cytoscape software (v.3.7.1) to visualize the results of the protein-protein interaction network. We utilized the cytoHubba plug-in, employing 10 different algorithms (stress, radiality, MNC, MCC, EPC, EcCentricity, degree, closeness, bottleneck, and betweenness), to accurately identify the core genes [25,26]. For each algorithm, we determined the top ten genes through various topological analyses within cytoHubba. Following this, we utilized the R package ‘UpSet’ to find the common genes among the top ten lists generated by all ten algorithms, thus designating them as the core genes. To further evaluate the predictive precision of core genes, we conducted Receiver Operating Characteristic (ROC) analysis. Using the pROC package (v.1.18.0), we generated ROC curves for the core genes, taking into account their gene expression profile.

2.4. Consensus clustering analysis

We utilized the gene expression profile of DEFAGs to conduct consensus clustering analysis using the “ConsensusClusterPlus” package (v.1.66.0) in R [27]. The analysis was performed with a maximum K value of 10, 80 % resampling rate, and 10 iterations, applying the partitioning around medoids algorithm with a euclidean distance metric. This allowed us to classify MN samples into distinct subtypes.

Table 1
Characteristics of the GEO datasets.

GEO ID	Platform ID	Tissues	Sample size	Applications
GSE108112	GPL19983	Renal tubule	MN (n = 43), HC (n = 5)	Discovery
GSE104948	GPL24120	Glomerular	MN (n = 21), HC (n = 3)	Validation
GSE108109	GPL19983	Glomerular	MN (n = 44), HC (n = 6)	Validation
GSE115857	GPL14951	Renal biopsy	MN (n = 11), HC (n = 7)	Validation
GSE104954	GPL24120	Renal tubule	MN (n = 18), HC (n = 3)	Validation
GSE133288	GPL19983	Renal tubule	MN (n = 48), HC (n = 5)	Validation

Note: The organisms represented in the dataset are *Homo sapiens*.

2.5. Functional enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) pathway enrichment analyses were performed by using the "clusterProfiler" package (v.4.4.4) of R [28]. P-value <0.05 was considered significantly enriched. Gene set enrichment analysis (GSEA) was applied to identify the most significant pathways between the molecular subtypes [29]. The "clusterProfiler" package (v.4.4.4) was used to perform GSEA. The "c2.cp.v7.2.symbols.gmt" was selected as the reference gene set. Gene set variation analysis (GSVA) could detect subtle changes in signaling pathways between samples [30]. Additionally, we used the "GSVA" package (v.1.44.5) of R to carry out the GSVA between molecular subtypes, and "c2.cp.v7.2.symbols.gmt" was selected as the reference gene set [31]. The R package "limma" (v.3.22.7) was utilized to identify differential pathways between the C1 and C2 groups. The analysis was conducted using a p-value cutoff of less than 0.05, along with a requirement for an absolute log fold change of no less than 1.

2.6. Analysis of the immune cell infiltration

xCell is a robust tool that can convert gene expression profiles into immune and stroma cell infiltration levels across samples [32]. In the present study, we used xCell to access the immune score, stroma score, microenvironment score, and subsets of immune cells in the molecular subtypes. The "ggplot2" package (v.3.3.6) in R was used to visualize the results.

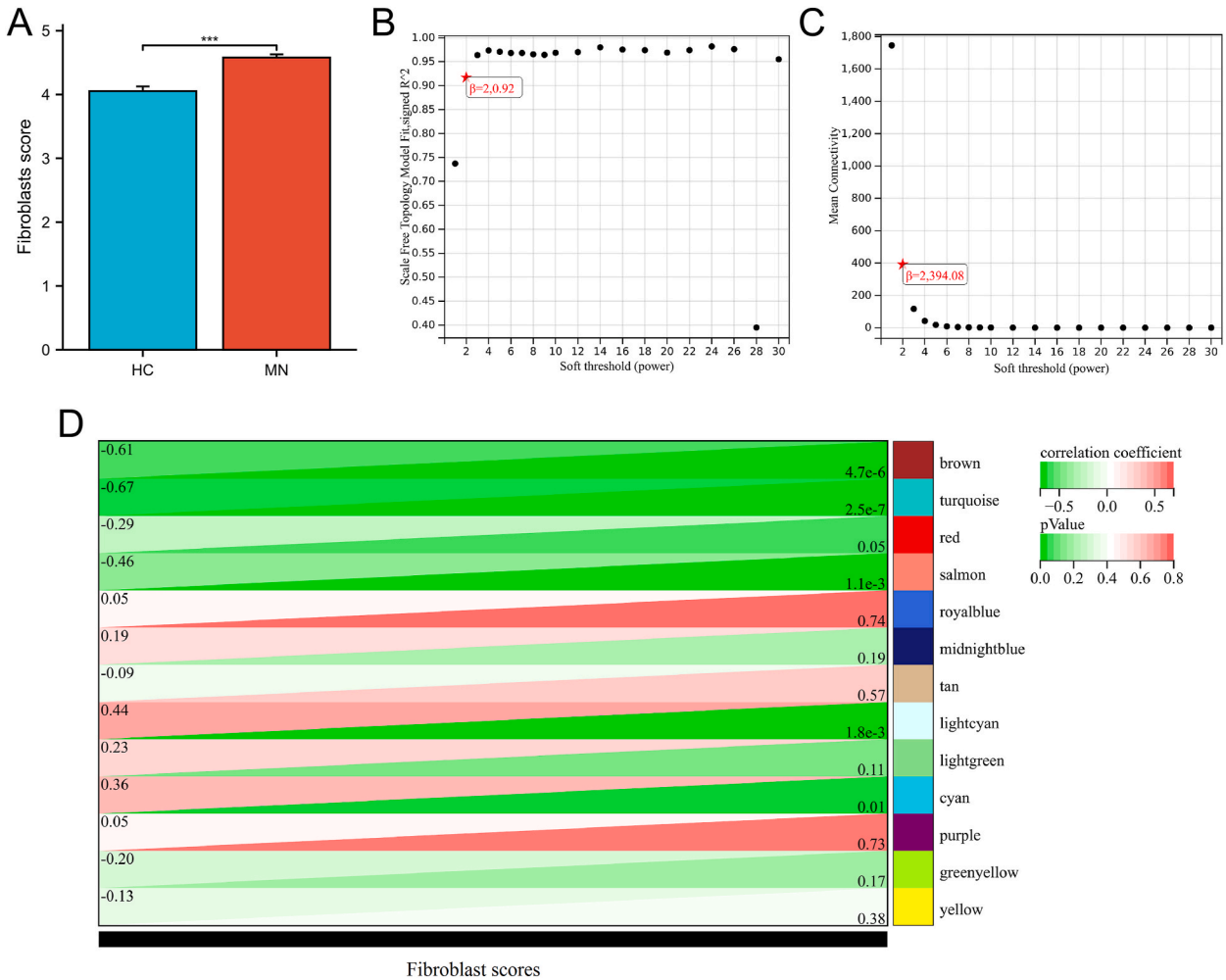


Fig. 2. Identification of fibroblast-related modules using WGCNA. (A) The level of fibroblast infiltration between the HC and MN groups was calculated using the xCell algorithm. ***p < 0.001. The method employed for statistical analysis is the Wilcoxon rank sum test. Scale-free fitting index (B) and mean connectivity (C) under different soft-threshold power. (D) Heatmap of the relationships between clinical traits and co-expression modules.

3. Results

3.1. Identification of fibroblasts-associated modules and genes in MN

First, we used the xCell and MCPCounter algorithms to evaluate the level of fibroblast infiltration between the MN and HC groups. As shown in Fig. 2A and Fig. S1, the fibroblast score in the MN group was higher than that in the HC group ($p < 0.001$), suggesting a

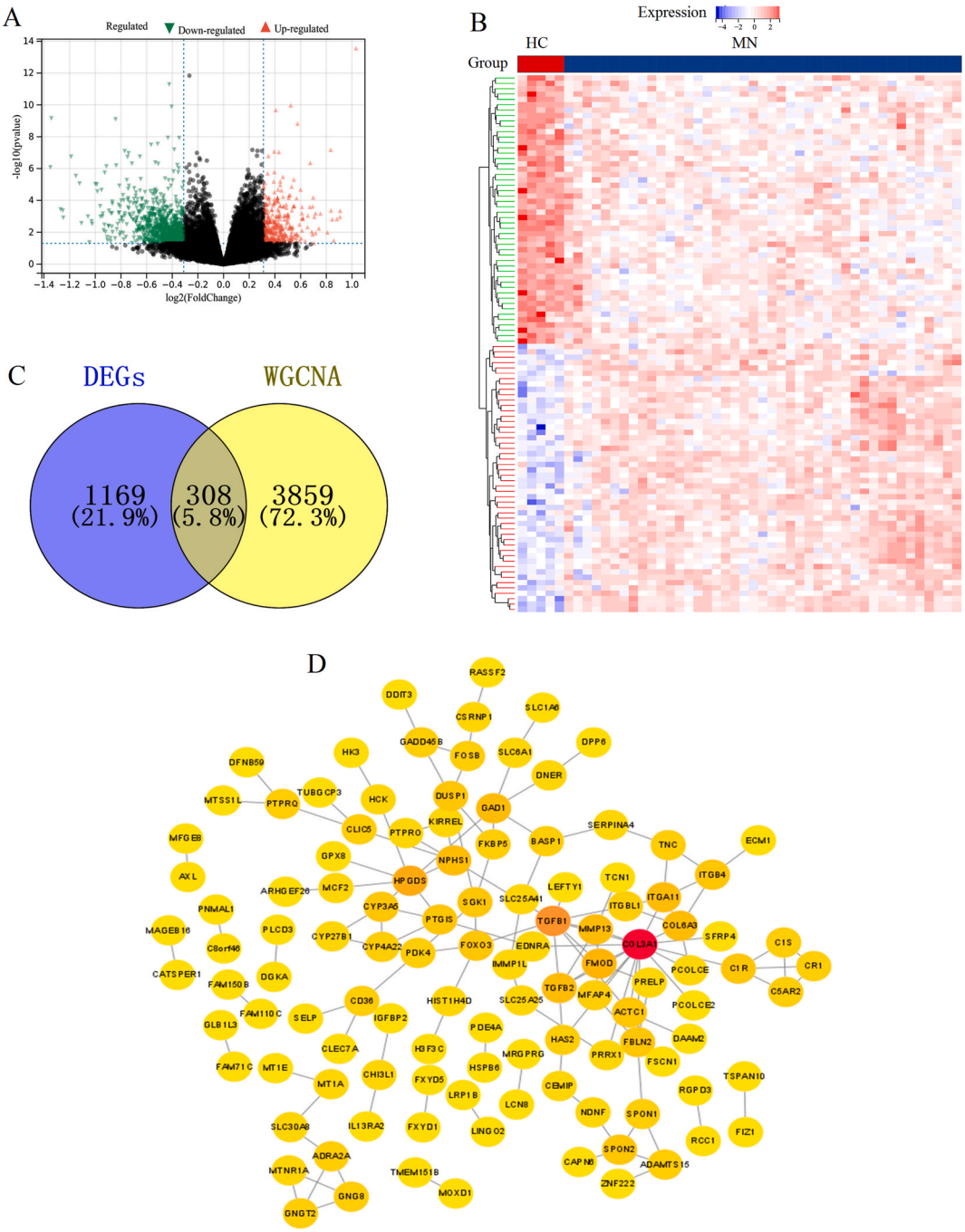


Fig. 3. Identification of DEFAGs in MN. (A) The volcano plot shows the DEGs in the MN and HC groups. The green dots represent down-regulated genes; the red dots represent up-regulated genes. (B) Heatmap shows the top 50 DEGs in the MN and HC groups. Light blue boxes represent gene down-regulation, while red boxes represent gene up-regulation. (C) The overlapped genes between the key modules and DEGs. (D) A PPI network of DEFAGs. The darker the color, the more important the gene in the PPI network.

potential involvement of fibroblasts in the development of MN. Then, WGCNA was used to screen fibroblast-related genes. As depicted in Fig. 2B and C, the power of $\beta = 2$ was selected as the soft-thresholding parameter to generate the scale-free network. Thirteen modules related to fibroblasts were identified through WGCNA analysis (Fig. 2D). Of these modules, the brown and turquoise modules exhibited a significant correlation with the fibroblast score (correlation coefficient = 0.79, $p < 0.01$ and correlation coefficient = 0.75, $p < 0.01$) (Figs. S2–S3). Therefore, we selected these two modules for further analysis.

3.2. Identification of DEFAGs in MN

As shown in Fig. 3A and B, it was found that there were 1477 DEGs between the MN and HC groups (Table S1). Out of these, 400 genes were up-regulated while 1077 were down-regulated. The WGCNA identified 4167 genes associated with fibroblasts (Table S2). Subsequently, a total of 308 intersection genes (DEFAGs) between the WGCNA and DEGs were obtained (Fig. 3C). Additionally, we used Cytoscape software to construct a PPI network of DEFAGs, which included 115 nodes and 134 edges (Fig. 3D).

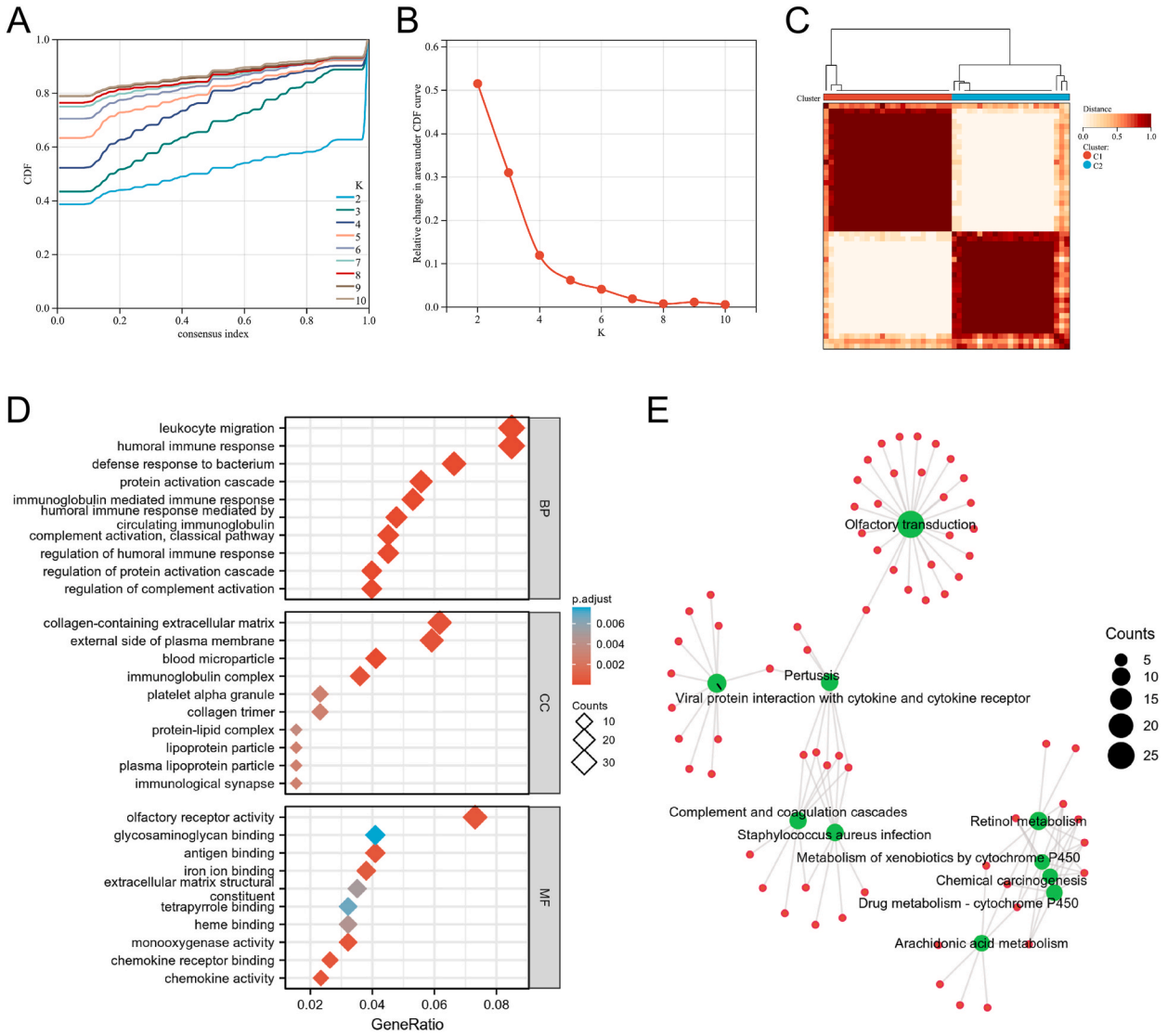


Fig. 4. Identification of differential molecular subtypes using consensus clustering analysis. (A) The cumulative distribution function (CDF) plots depict consensus distribution. (B) Relative changes in the area under the CDF curve. (C) The matrix heatmap was plotted when k = 2. (D) Enrichment analysis results of the top 10 BP, CC, and MF terms between C1 and C2 subgroups. (E) Enrichment analysis results of the top KEGG pathways between C1 and C2 subgroups. The target genes are represented by red circular nodes, while the KEGG pathways are represented by green circular nodes.

3.3. Consensus clustering analysis identifies fibroblasts-associated molecular subtypes

In this study, an unsupervised cluster analysis was performed on individuals with MN to categorize them into distinct fibroblast phenotypes. Based on the gene expression profile of DEFAGs, the MN patients were clustered into C1 and C2 subtypes through consensus clustering analysis (Fig. 4A–C). To validate the clustering results, we performed a principal component analysis. As shown in Fig. S4, subgroup C1 (blue dots) and subgroup C2 (red dots) exhibit a clear separation along the PC1 axis. The blue dashed line encloses

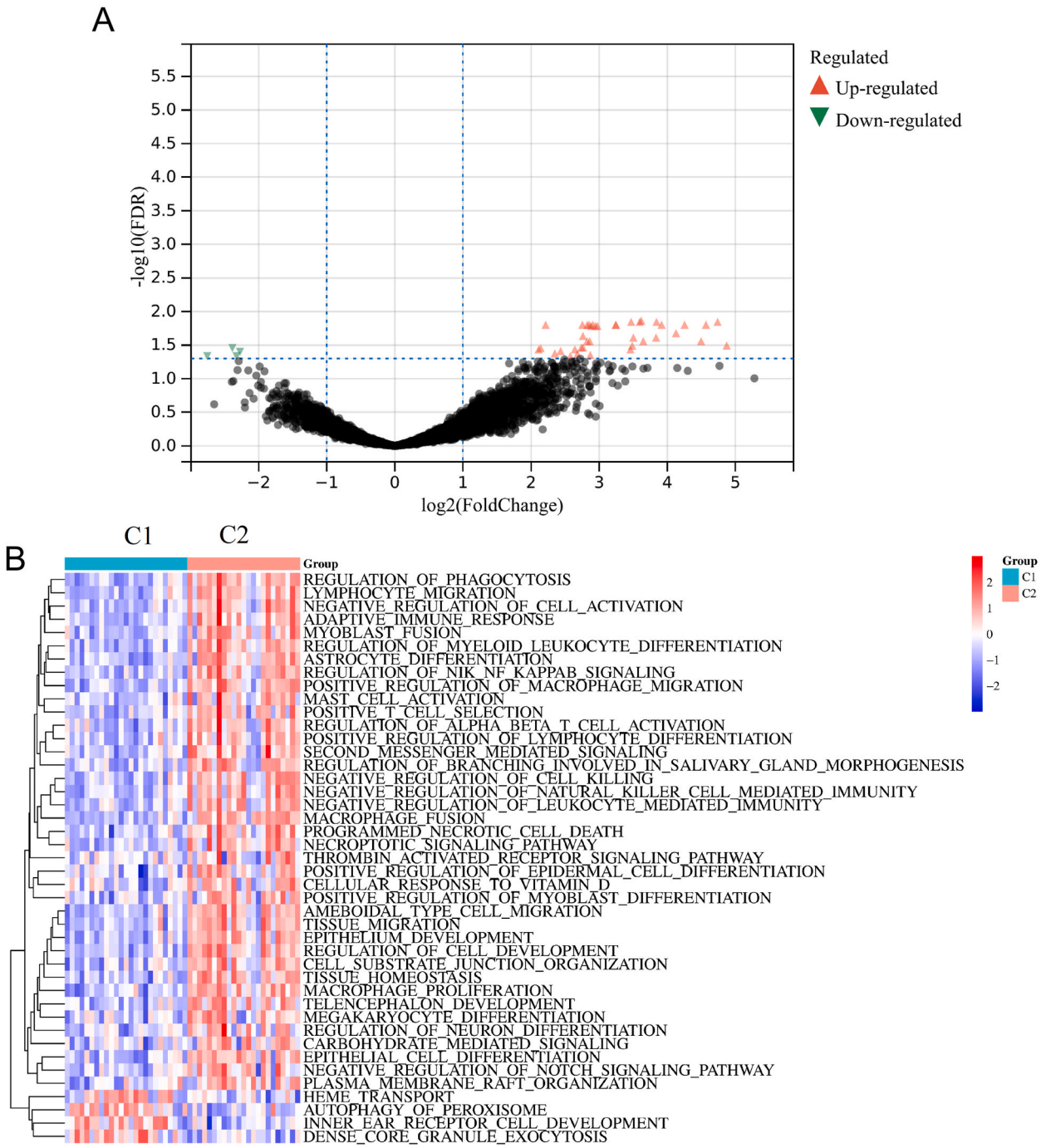


Fig. 5. GSVA analysis between the differential molecular subtypes. (A) The volcano plot describes the differentially expressed signaling pathways between the C1 and C2 subtypes. The green dots represent down-regulated pathways; the red dots represent up-regulated pathways. (B) Heatmap shows the differentially expressed signaling pathways between the C1 and C2 subtypes. Light blue boxes represent pathway inactivation, while red boxes represent pathway activation.

the C1 subgroup, while the red dashed line encloses the C2 subgroup, emphasizing distinct clustering patterns. The evident separation between the two subgroups along PC1 suggests that the clustering analysis effectively differentiates the underlying characteristics of MN patient samples into these two subgroups, corroborating the initial clustering results. A total of 606 DEGs were differentially expressed between the C1 and C2 subtypes, containing 253 down-regulated and 353 up-regulated genes (Fig. S5). We performed GO and KEGG pathway analysis of those DEGs. In terms of biological processes (BP), DEGs were mainly enriched in the leukocyte migration, humoral immune response, immunoglobulin-mediated immune response, humoral immune response mediated by circulating immunoglobulin, regulation of humoral immune response, etc. Cell components (CC), DEGs were mainly enriched in the collagen-containing extracellular matrix, external side of the plasma membrane, blood microparticle, immunoglobulin complex, etc. Molecular function (MF), DEGs were mainly enriched in olfactory receptor activity, glycosaminoglycan binding, antigen binding, iron ion binding, chemokine receptor binding, etc (Fig. 4D). KEGG, DEGs were mainly enriched in olfactory transduction, pertussis, viral protein interaction with cytokine and cytokine receptors, retinol metabolism, complement and coagulation cascades, chemical carcinogenesis, etc (Fig. 4E).

3.4. GSVA and GSEA

We performed GSVA and GSEA to further investigate the biological functions enriched in the fibroblast-associated subtypes. As shown in Fig. 5A and B, 43 differentially expressed signaling pathways were identified between the C1 and C2 groups. Among these

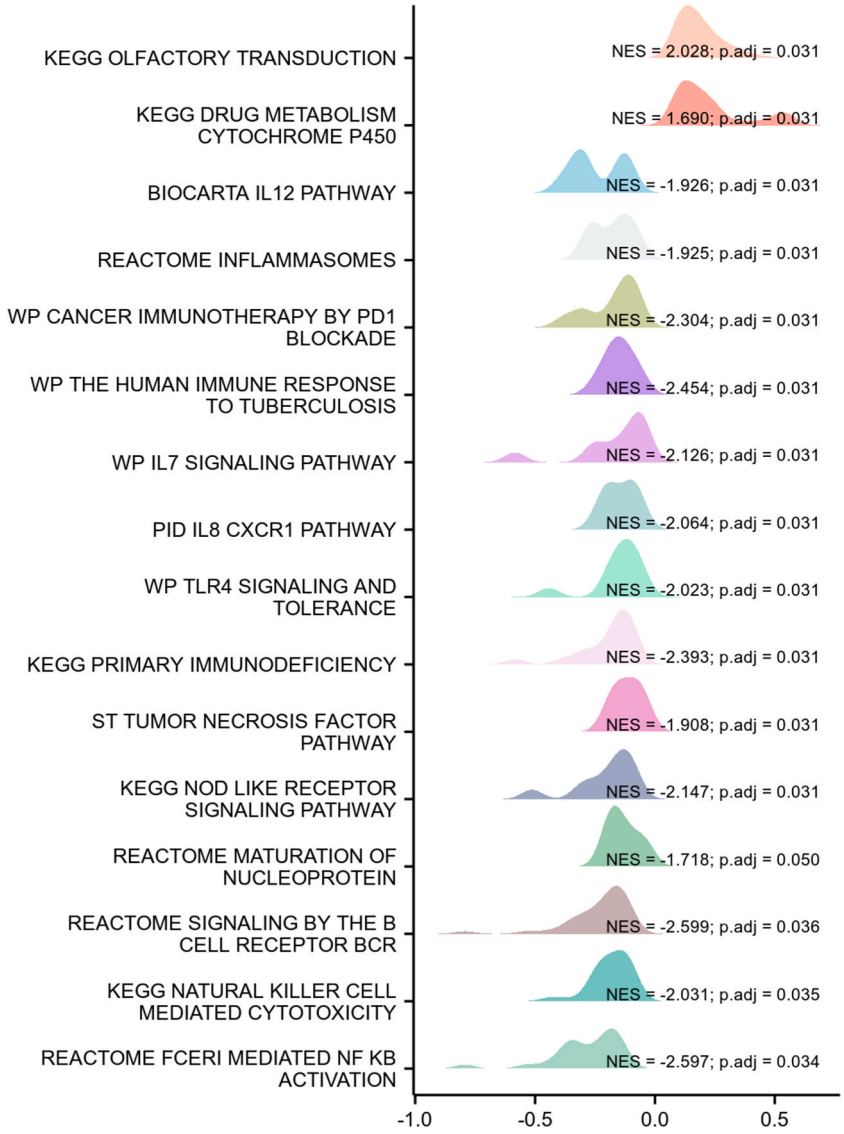


Fig. 6. Mutipeaked maps describe the results of GSEA analysis.

pathways, 4 were down-regulated and 39 were up-regulated. In addition, the GSVA findings revealed that immune-associated pathways were activated in C2 subtypes, including lymphocyte migration, adaptive immune response, regulation of myeloid leukocyte differentiation, positive regulation of macrophage migration, mast cell activation, positive T cell selection, regulation of alpha-beta T cell activation, positive regulation of lymphocyte differentiation, negative regulation of natural killer cell-mediated immunity, negative regulation of leukocyte mediated immunity, macrophage fusion, macrophage proliferation, etc.

Furthermore, GSEA results indicated that olfactory transduction, drug metabolism cytochrome P450, IL12 pathway, inflammasomes, cancer immunotherapy by PD1 blockade, the human immune response to tuberculosis, IL7 signaling pathway, IL8 CXCR1 pathway, TLR4 signaling and tolerance, primary immunodeficiency, NOD-like receptor signaling pathway, maturation of nucleoprotein, natural killer cell-mediated cytotoxicity, and fceri medicated NF KB activation were differentially enriched in the C1 and C2 subtypes (Fig. 6). These results revealed noteworthy disparities in the immune status between the two subgroups. In particular, the C2 subgroup exhibited a robust immune response.

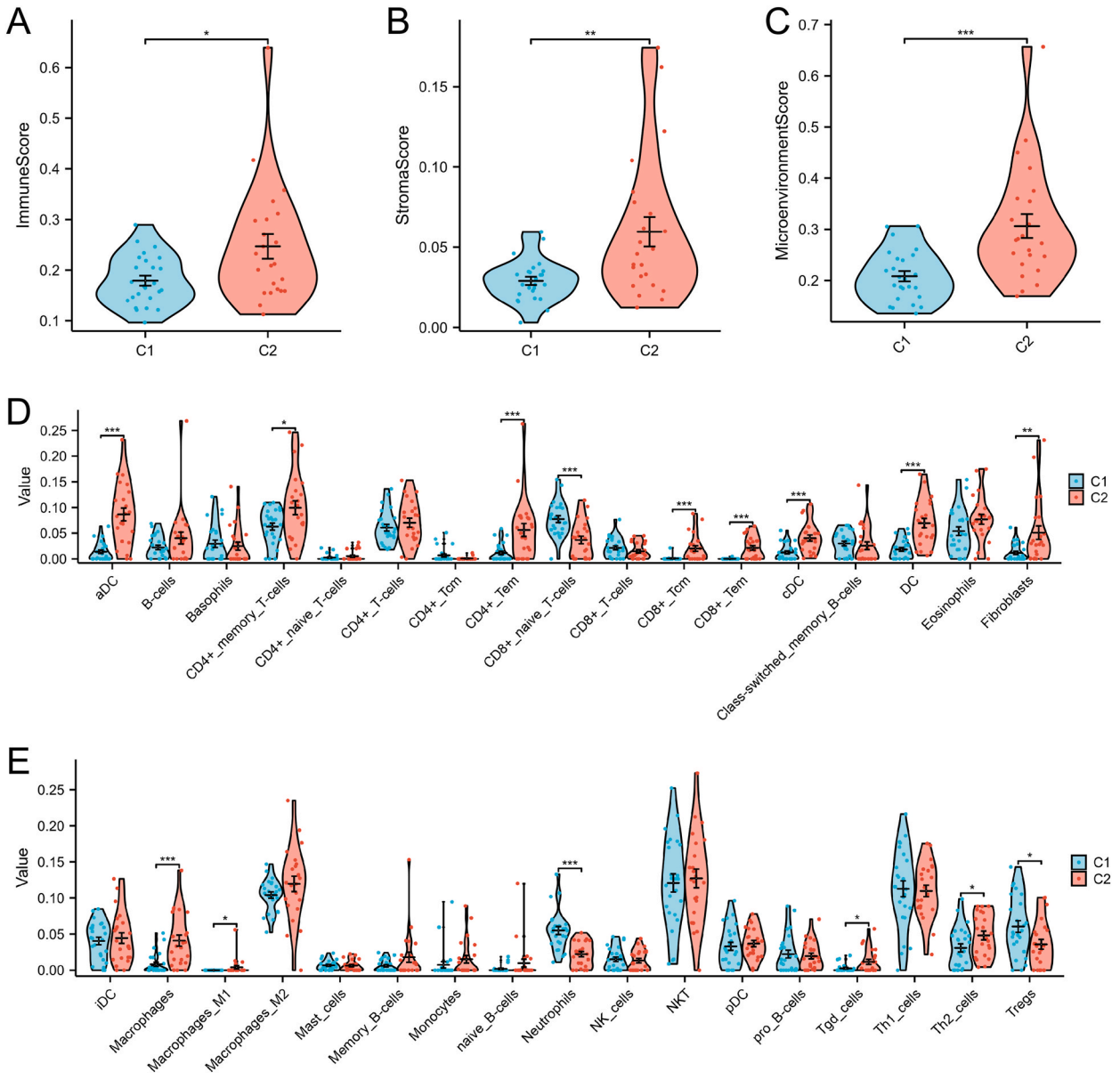


Fig. 7. Characterizing differential immunological features in the molecular subtypes. The comparisons of the immune score (A), stroma score (B), and microenvironment score (C) between C1 and C2 subtypes. (D–E) The comparisons of immune cell infiltration levels between the C1 and C2 subtypes. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. The method employed for statistical analysis is the Wilcoxon rank sum test.

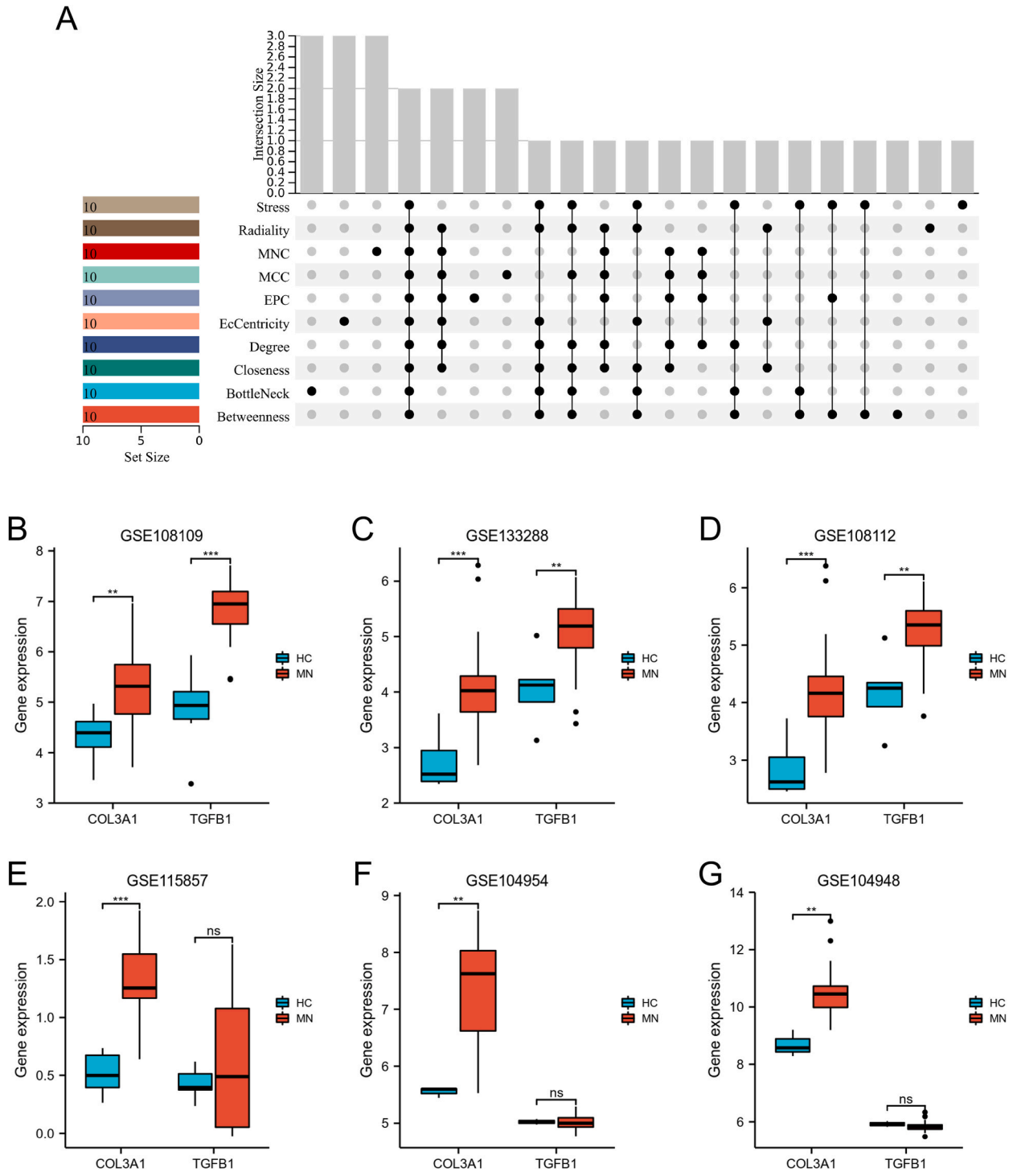


Fig. 8. Identification and validation of core genes for MN patients. (A) Ten algorithms (stress, radiality, MNC, MCC, EPC, EcCentricity, degree, closeness, bottleneck, and betweenness) in CytoHubba were used to identify core genes. (B–G) The comparisons of COL3A1 and TGFB1 expression levels in HC and MN groups using multiple microarray datasets. **p < 0.01, ***p < 0.001. The method employed for statistical analysis is the Wilcoxon rank sum test.

3.5. Characterization of immune cell infiltration landscapes

The results of xCell showed notable differences in the immune status between the C1 and C2 subtypes. Compared with the C1 subtype, the C2 subtype exhibited a higher immune score, stroma score, and microenvironment score (Fig. 7A–C). Additionally, the proportions of aDC, $CD4^+$ memory T cells, $CD4^+$ Tem, $CD8^+$ Tcm, $CD8^+$ Tem, cDC, fibroblasts, macrophages, macrophages M1, Tgd cells were significantly higher in the C2 subtype than those in the C1 subtype, whereas the proportions of $CD8^+$ naive T cells, neutrophils, and Tregs were lower in the C2 subtype than those in the C1 subtype (Fig. 7D and E). The results revealed notable disparities in immune cell infiltration between the two subgroups. Specifically, the C2 subgroup demonstrated elevated levels of immune cell infiltration, which is in line with the findings from the enrichment analysis.

3.6. Screening core biomarkers for MN patients

In the present study, two core genes (COL3A1 and TGFB1) were identified by the stress, radiality, MNC, MCC, EPC, EcCentricity, degree, closeness, bottleneck, and betweenness algorithm in CytoHubba (Fig. 8A). In addition, multiple microarray datasets were used to validate the gene expression of core genes. As shown in Fig. 8B–D, the expression levels of COL3A1 and TGFB1 were higher in the MN group than in the HC group significantly ($p < 0.05$). The expression level of COL3A1 was higher in the MN group than in the HC group significantly ($p < 0.05$). In the GSE115857, GSE104954, and GSE104948 datasets, there was no significant difference in the expression of TGFB1 (Fig. 8E–G). The diagnostic ROC results showed that the AUC value of COL3A1 was greater than 0.8 in all six datasets, indicating a good diagnostic efficacy of COL3A1 (Fig. 9A–F).

3.7. Analysis of the potential biological functions of COL3A1

We conducted GSEA to gain deeper insights into the biological functions associated with COL3A1. As shown in Fig. 10A, T cell receptor signaling pathway (ES = −0.53, p = 0), integrin-mediated signaling pathway (ES = −0.6079, p = 0), kappab phosphorylation (ES = −0.7416, p = 0), T cell-mediated cytotoxicity (ES = −0.6311, p = 0), myeloid cell differentiation (ES = −0.4417, p = 0), myeloid leukocyte activation (ES = −0.4673, p = 0), and kappab kinase NF kappab signaling (ES = −0.4367, p = 0) were enriched in the COL3A1 high-expressed phenotype.

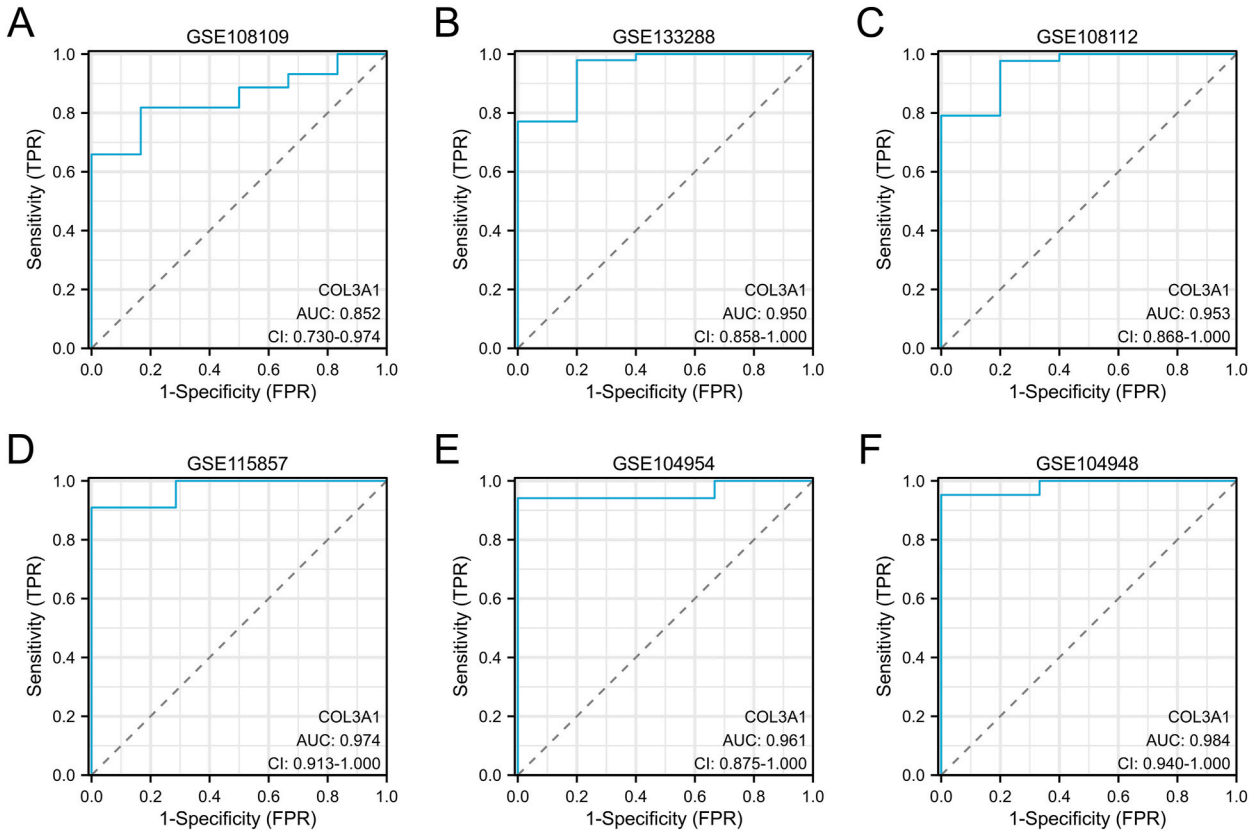


Fig. 9. ROC curves were used to assess the diagnostic efficacy of COL3A1. The GSE108109 (A), GSE133288 (B), GSE108112 (C), GSE115857 (D), GSE104954 (E), and GSE104948 (F) datasets were used to validate the diagnostic performance of COL3A1.

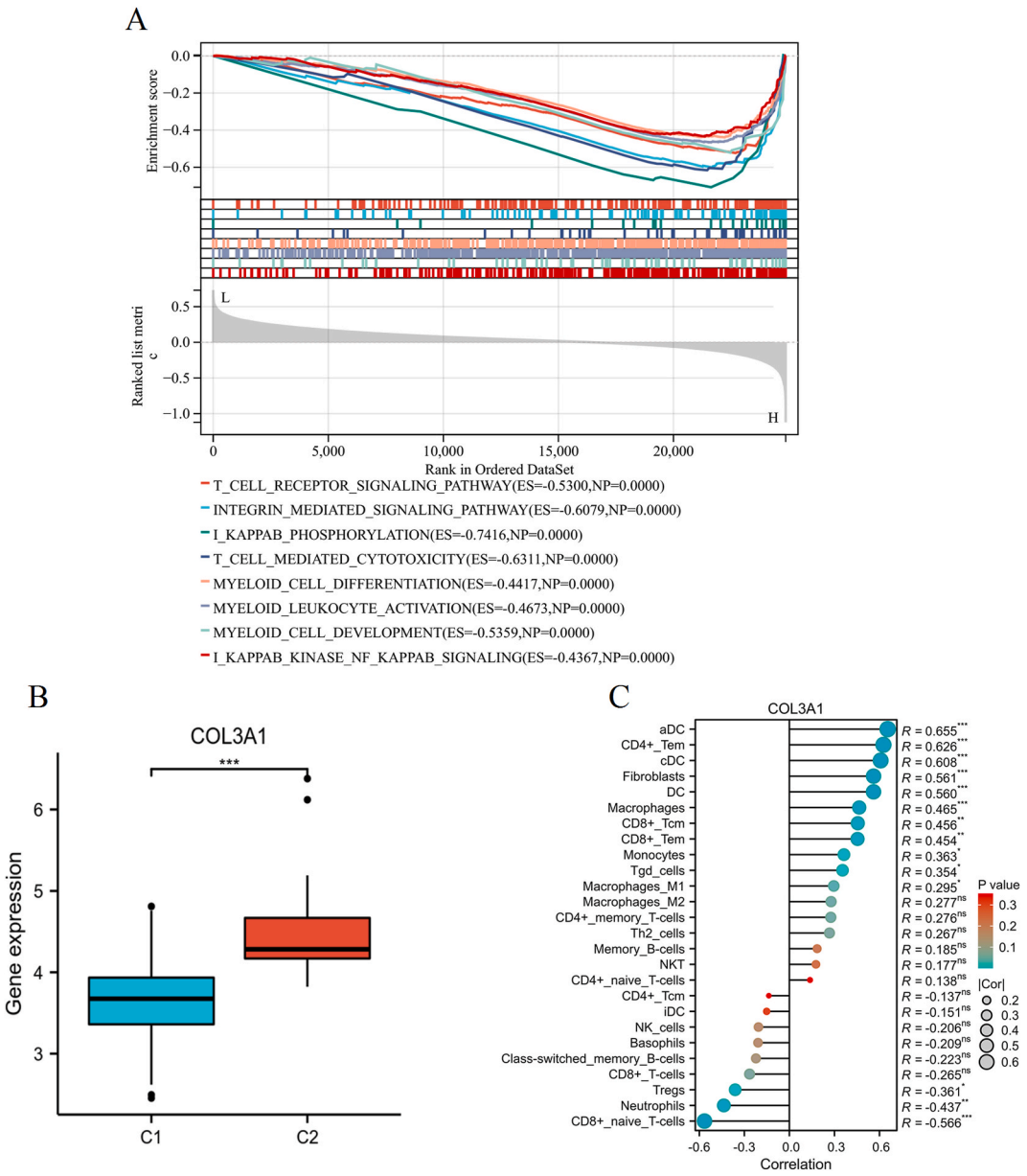


Fig. 10. Investigation of the potential biological functions of COL3A1. (A) Signaling pathways enriched in the high and low levels of COL3A1. (B) The comparisons of COL3A1 expression level in C1 and C2 subtypes. (C) Correlation analysis between COL3A1 expression and immune cell compositions. ***p < 0.001. The method employed for statistical analysis is the Wilcoxon rank sum test.

3.8. Correlation analysis between COL3A1 expression and immune cell infiltration

As shown in Fig. 10B, the expression level of COL3A1 was significantly up-regulated in the C2 subtype compared with that in the C1 subtype ($p < 0.05$), implying that COL3A1 may be involved in the progression of MN. In addition, the relationship between COL3A1 expression and immunological characteristics was investigated. As shown in Fig. 10C, COL3A1 expression was negatively correlated with $CD8^+$ naive T cells ($R = -0.56$, $p < 0.01$), neutrophils ($R = -0.43$, $p < 0.01$), and Tregs ($R = -0.36$, $p < 0.05$) and positively correlated with aDC ($R = 0.65$, $p < 0.01$), $CD4^+$ Tem ($R = 0.626$, $p < 0.01$), cDC ($R = 0.6$, $p < 0.01$), fibroblasts ($R = 0.56$, $p < 0.01$), dendritic cell (DC) ($R = 0.56$, $p < 0.01$), macrophages ($R = 0.46$, $p < 0.01$), $CD8^+$ Tcm ($R = 0.45$, $p < 0.01$), $CD8^+$ Tem ($R = 0.45$, $p < 0.01$), monocytes ($R = 0.36$, $p < 0.05$), and Tgd cells ($R = 0.35$, $p < 0.05$). Therefore, it is plausible that COL3A1 plays a significant role in the regulation of immunological processes during the development and progression of MN.

4. Discussion

Membranous nephropathy (MN) is characterized by podocyte damage, glomerular basement membrane thickening, and inflammatory immune response [8,33]. MN is a heterogeneous disease, and although there have been some studies on it, its specific pathogenesis remains unclear. Therefore, it is urgent to investigate the underlying pathological mechanism and develop effective biomarkers for the diagnosis and treatment of MN. As stromal fibroblasts interact with circulating immune cells at sites of inflammation, they play a significant role in driving local inflammation and mediating the inflammatory response [34]. To gain a better understanding of the potential molecular pathogenesis of MN, we downloaded MN-related datasets and used bioinformatics analysis methods to identify two fibroblast-related subtypes. Two fibroblast-associated subtypes (C1 and C2) were classified using the consensus clustering method. Our findings revealed distinct signaling pathways and immune statuses between these subtypes. Furthermore, the MN subtype demonstrated a strong correlation with specific immune and inflammatory-related pathways, indicating its potential clinical significance in the diagnosis and treatment of MN.

Firstly, our comprehensive analysis of GO, KEGG, GSEA, and GSVA revealed a significant association between MN subtypes and processes related to leukocyte migration, humoral immune response, macrophage fusion, macrophage differentiation, negative regulation of leukocyte mediated immunity, positive T cell selection, natural killer cell-mediated cytotoxicity, inflammasomes, primary immunodeficiency, TLR4 signaling and tolerance, IL8 CXCR1 pathway, and chemokine activity, implying that immune response processes may be involved in the pathogenesis of MN. The macrophage migration inhibitory factor is an inflammatory factor generated by macrophages, and it plays a pathogenic role in glomerulonephritis [35]. The level of macrophage migration inhibitory factor in urine and plasma could be used to indicate the severity of MN [36]. A recent study showed that M2 macrophage subtypes are related to lgG subclass/complement deposition, which implies that M2 macrophage might have participated in the development of MN [37]. Activated plasma cells and T follicular helper cells may promote the progression of MN [38]. A high level of IL8 was closely related to poor prognosis for MN patients [39]. TLR4, TLR9, and IL6 genes were associated with idiopathic MN in Taiwan [40]. Overexpression of chemokines was observed in patients with severe MN [41]. Our findings reinforced the importance of diverse immune cells and molecules in the progression of MN's pathological process, aligning with previous studies.

As the C2 subtype showed enrichment in immune- and inflammatory-related pathways, we further investigated the association between the MN subtypes and immune cell infiltration. The xCell results revealed varying levels of immune cell infiltration between the C1 and C2 subtypes. The DC, aDC, $CD4^+$ memory T cells, $CD4^+$ Tem, $CD8^+$ Tcm, $CD8^+$ Tem, cDC, macrophages, macrophages M1, Tgd cells, and Th2 cells were highly expressed in C2 subtype compared with C1 subtype, while $CD8^+$ naïve T cells, neutrophils, and Tregs had low expression level in C2 subtype. A previous study revealed that activated kidney DC plays a vital role in renal disease development [42]. The number of glomerular macrophages was associated with the severity of proliferative glomerulonephritis [43,44]. Previous studies have indicated that activated macrophages could influence the generation of proinflammatory factors in the glomeruli and of adhesion molecules in the endothelium, making glomerular diseases more severe [45,46]. A recent study showed that neutrophils play an important role in the pathogenesis of glomerulonephritis [47]. In line with previous studies, our findings demonstrated significant differences in immune cell infiltration between the two subgroups. Notably, the C2 subgroup exhibited higher levels of immune cell infiltration, consistent with the results of the enrichment analysis. As a result, we hypothesized that the C2 subtype is more prone to developing advanced MN as compared with the C1 subtype. However, this result needs to be further substantiated by conducting cell or animal experiments.

Another important finding of our study was that the COL3A1 gene was the potential diagnostic biomarker for MN patients. Collagen type III alpha 1 (COL3A1) is an important extracellular matrix protein, that plays a crucial role in various physiological functions including inflammation, apoptosis, immune response, and bone development [48]. Previous reports have indicated that abnormal expression of COL3A1 happens in esophageal cancer, lung cancer, and head and neck cancer [49–51]. COL3A1 gene showed up-regulated expression in kidney tumors compared with normal tissue [52]. Our study discovered that the COL3A1 gene was up-regulated in MN patients. It was observed that the expression of COL3A1 was higher in the C2 subtype as opposed to the C1 subtype, suggesting that COL3A1 could potentially serve as a biomarker in distinguishing between different subtypes of MN. The GSEA results revealed that immune-related pathways were enriched in the phenotype with high expression of COL3A1. Furthermore, there was a correlation between the expression of COL3A1 and the composition of immune cells, particularly $CD8^+$ naive T cells, neutrophils, Tregs, aDC, $CD4^+$ Tem, cDC, DC, macrophages, $CD8^+$ Tcm, $CD8^+$ Tem, and monocytes. Our findings are consistent with previous studies that have suggested COL3A1 as a potential diagnostic biomarker for osteoarthritis and its close association with immune cell infiltration [53]. In addition, COL3A1 has the potential to serve as a molecular biomarker for predicting prognosis and immune infiltration in various types of cancer [54,55]. These results indicated that COL3A1 could potentially function as a reliable diagnostic biomarker for patients with MN in clinical practice, and it may have a role in modulating immunological processes throughout the progression of MN.

While this study provides valuable insights into fibroblast-associated biomarkers and molecular subtypes in MN, several limitations should be acknowledged. This study's small sample size, particularly in the control group, may limit the generalizability of our findings. Although we validated our results across five independent datasets, the variability among these datasets could introduce bias. Therefore, further validation in larger, more diverse cohorts is needed to confirm these findings. Additionally, the reliance on topological features within the PPI network may exclude genes that play critical roles outside of the identified network or in less central pathways. Future studies should consider broader approaches to capture a more comprehensive set of MN-associated genes.

5. Conclusions

This study identified fibroblast-associated biomarkers and molecular subtypes in MN through integrated bioinformatics. Elevated fibroblast infiltration and specific gene modules were linked to MN pathogenesis. We identified 308 key genes, leading to the classification of MN into two subtypes with distinct immune and extracellular matrix profiles. COL3A1 emerged as crucial biomarker, offering potential diagnostic and therapeutic target. These findings highlight the significant role of fibroblasts in MN and suggest avenues for targeted treatment strategies.

Data availability

All data used in the present study were available from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Funding

This work was supported by the National Key Research and Development Program for the Modernization of Traditional Chinese Medicine (2019YFC1709403).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

CRediT authorship contribution statement

Chuying Gui: Writing – original draft, Investigation, Conceptualization. **Sidi Liu:** Methodology, Formal analysis, Data curation. **Zhike Fu:** Validation, Software, Resources. **Huijie Li:** Visualization, Validation, Formal analysis. **Di Zhang:** Visualization, Resources, Project administration, Methodology. **Yueyi Deng:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Chuying Gui reports financial support was provided by the National Key Research and Development Program for the Modernization of Traditional Chinese Medicine (2019YFC1709403). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Not applicable.

Abbreviations

COL3A1	Collagen type III alpha 1
DC	Dendritic cell
DEFAGs	Differentially expressed fibroblast-associated genes
DEGs	differentially expressed genes
MN	membranous nephropathy
GEO	Gene Expression Omnibus
HC	healthy controls
KEGG	Kyoto Encyclopedia of Genes and Genomes
GO	Gene Ontology
GSEA	Gene set enrichment analysis
GSVA	Gene set variation analysis
ROC	Receiver Operating Characteristic
PPI	protein-protein interaction
WGCNA	weighted gene co-expression network analysis

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38424.

References

[1] W.G. Couser, Primary membranous nephropathy, Clin. J. Am. Soc. Nephrol. : CJASN 12 (2017) 983–997.
[2] S. Akiyama, E. Imai, S. Maruyama, Immunology of membranous nephropathy, F1000Research 8 (2019).
[3] A. McGrogan, C.F. Franssen, C.S. de Vries, The incidence of primary glomerulonephritis worldwide: a systematic review of the literature, Nephrol. Dial. Transplant. (2011) 414–430, official publication of the European Dialysis and Transplant Association - European Renal Association 26.
[4] N. Polanco, E. Gutiérrez, A. Covarsí, F. Ariza, A. Carreño, A. Vigil, J. Baltar, G. Fernández-Fresnedo, C. Martín, S. Pons, D. Lorenzo, C. Bernis, P. Arrizabalaga, G. Fernández-Juárez, V. Barrio, M. Sierra, I. Castellanos, M. Espinosa, F. Rivera, A. Oliet, F. Fernández-Vega, M. Praga, Spontaneous remission of nephrotic syndrome in idiopathic membranous nephropathy, J. Am. Soc. Nephrol. : JASN (J. Am. Soc. Nephrol.) 21 (2010) 697–704.
[5] Y. Gu, H. Xu, D. Tang, Mechanisms of primary membranous nephropathy, Biomolecules 11 (2021).
[6] Q. Liu, J. Liu, B. Lin, Y. Zhang, M. Ma, M. Yang, X. Qin, Novel biomarkers in membranous nephropathy, Front. Immunol. 13 (2022) 845767.
[7] W. Liu, C. Gao, Z. Liu, H. Dai, Z. Feng, Z. Dong, Y. Zheng, Y. Gao, X. Tian, B. Liu, Idiopathic membranous nephropathy: glomerular pathological pattern caused by extrarenal immunity activity, Front. Immunol. 11 (2020) 1846.
[8] R. Motavalli, J. Etemadi, H. Kahroba, A. Mehdizadeh, M. Yousefi, Immune system-mediated cellular and molecular mechanisms in idiopathic membranous nephropathy pathogenesis and possible therapeutic targets, Life Sci. 238 (2019) 116923.
[9] W. Liu, C. Gao, H. Dai, Y. Zheng, Z. Dong, Y. Gao, F. Liu, Z. Zhang, Z. Liu, W. Liu, B. Liu, Q. Liu, J. Shi, Immunological pathogenesis of membranous nephropathy: focus on PLA2R1 and its role, Front. Immunol. 10 (2019) 1809.
[10] R.G. Bonegio, R. Fuhro, Z. Wang, C.R. Valeri, C. Andry, D.J. Salant, W. Lieberthal, Rapamycin ameliorates proteinuria-associated tubulointerstitial inflammation and fibrosis in experimental membranous nephropathy, J. Am. Soc. Nephrol. : JASN (J. Am. Soc. Nephrol.) 16 (2005) 2063–2072.
[11] M.D. Lynch, F.M. Watt, Fibroblast heterogeneity: implications for human disease, J. Clin. Invest. 128 (2018) 26–35.
[12] S. Davidson, M. Coles, T. Thomas, G. Kollias, B. Ludewig, S. Turley, M. Brenner, C.D. Buckley, Fibroblasts as immune regulators in infection, inflammation and cancer, Nat. Rev. Immunol. 21 (2021) 704–717.
[13] C.D. Buckley, D. Pilling, J.M. Lord, A.N. Akbar, D. Scheel-Toellner, M. Salmon, Fibroblasts regulate the switch from acute resolving to chronic persistent inflammation, Trends Immunol. 22 (2001) 199–204.
[14] Y. Wang, Y. Xu, Z. Yang, X. Liu, Q. Dai, Using recursive feature selection with random forest to improve protein structural class prediction for low-similarity sequences, Comput. Math. Methods Med. 2021 (2021) 5529389.
[15] Q. Dai, C. Bao, Y. Hai, S. Ma, T. Zhou, C. Wang, Y. Wang, W. Huo, X. Liu, Y. Yao, Z. Xuan, M. Chen, M.Q. Zhang, MTGIpick allows robust identification of genomic islands from a single genome, Briefings Bioinf. 19 (2018) 361–373.
[16] R. Kong, X. Xu, X. Liu, P. He, M.Q. Zhang, Q. Dai, 2SigFinder: the combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome, BMC Bioinf. 21 (2020) 159.
[17] S. Yang, Y. Wang, Y. Chen, Q. Dai, MASQC: next generation sequencing assists third generation sequencing for quality control in N6-methyladenine DNA identification, Front. Genet. 11 (2020) 269.
[18] Z. Yang, W. Yi, J. Tao, X. Liu, M.Q. Zhang, G. Chen, Q. Dai, HPVMD-C: a disease-based mutation database of human papillomavirus in China. Database : the journal of biological databases and curation 2022 (2022).
[19] L. Gautier, L. Cope, B.M. Bolstad, R.A. Irizarry, affy–analysis of Affymetrix GeneChip data at the probe level, Bioinformatics 20 (2004) 307–315.
[20] B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.
[21] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, Genome Biol. 18 (2017) 220.
[22] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, J. Selves, P. Laurent-Puig, C. Sautès-Fridman, W.H. Fridman, A. de Reyniès, Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (2016) 218.
[23] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinf. 9 (2008) 559.
[24] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic acids research 43 (2015) e47.
[25] C.H. Chin, S.H. Chen, H.H. Wu, C.W. Ho, M.T. Ko, C.Y. Lin, cytoHubba: identifying hub objects and sub-networks from complex interactome, BMC Syst. Biol. 8 (Suppl 4) (2014) S11.
[26] Y. Liu, D. Tan, H. Cui, J. Wang, Ganoderic acid C2 exerts the pharmacological effects against cyclophosphamide-induced immunosuppression: a study involving molecular docking and experimental validation, Sci. Rep. 13 (2023) 17745.
[27] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (2010) 1572–1573.
[28] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, OMICS A J. Integr. Biol. 16 (2012) 284–287.
[29] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15545–15550.
[30] S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, BMC Bioinf. 14 (2013) 7.
[31] M.R. Ferreira, G.A. Santos, C.A. Biagi, W.A. Silva Junior, W.F. Zambuzzi, GSVA score reveals molecular signatures from transcriptomes for biomaterials comparison, J. Biomed. Mater. Res., Part A 109 (2021) 1004–1014.
[32] D. Aran, Cell-type enrichment analysis of bulk transcriptomes using xCell, Methods Mol. Biol. 2120 (2020) 263–276.
[33] M. Cremoni, V. Brglez, S. Perez, F. Decoupigny, K. Zorzi, M. Andreani, A. Gérard, S. Boyer-Suavet, C. Ruetsch, S. Benzaken, V. Esnault, B. Seitz-Polski, Th17-Immune response in patients with membranous nephropathy is associated with thrombosis and relapses, Front. Immunol. 11 (2020) 574997.
[34] H. Farah, S.P. Young, C. Mauro, S.W. Jones, Metabolic dysfunction and inflammatory disease: the role of stromal fibroblasts, FEBS J. 288 (2021) 5555–5568.
[35] H.Y. Lan, N. Yang, D.J. Nikolic-Paterson, X.Q. Yu, W. Mu, N.M. Isbel, C.N. Metz, R. Bucala, R.C. Atkins, Expression of macrophage migration inhibitory factor in human glomerulonephritis, Kidney Int. 57 (2000) 499–509.
[36] N. Ding, P.L. Li, K.L. Wu, T.G. Lv, W.L. Yu, J. Hao, Macrophage migration inhibitory factor levels are associated with disease activity and possible complications in membranous nephropathy, Sci. Rep. 12 (2022) 18558.
[37] W. Hu, G. Li, J. Lin, W. Dong, F. Yu, W. Liu, Y. Wu, W. Hao, X. Liang, M2 macrophage subpopulations in glomeruli are associated with the deposition of IgG subclasses and complements in primary membranous nephropathy, Front. Med. 8 (2021) 657232.
[38] Z. Zhang, Y. Shi, K. Yang, R. Crew, H. Wang, Y. Jiang, Higher frequencies of circulating ICOS(+), IL-21(+) T follicular helper cells and plasma cells in patients with new-onset membranous nephropathy, Autoimmunity 50 (2017) 458–467.
[39] J. Chen, X. Fu, Y. Sun, S. Zhang, H. Xie, H. Lin, High urinary interleukin-8 levels is associated with poor prognosis in idiopathic membranous nephropathy, Intern. Med. J. 48 (2018) 207–209.
[40] S.Y. Chen, C.H. Chen, Y.C. Huang, C.J. Chan, D.C. Chen, F.J. Tsai, Genetic susceptibility to idiopathic membranous nephropathy in high-prevalence Area, Taiwan, Biomedicine 4 (2014) 9.

[41] S.A. Mezzano, M.A. Droguett, M.E. Burgos, L.G. Ardiles, C.A. Aros, I. Caorsi, J. Egido, Overexpression of chemokines, fibrogenic cytokines, and myofibroblasts in human membranous nephropathy, Kidney Int. 57 (2000) 147–158.
[42] F. Heymann, C. Meyer-Schwesinger, E.E. Hamilton-Williams, L. Hammerich, U. Panzer, S. Kaden, S.E. Quaggin, J. Floege, H.J. Gröne, C. Kurts, Kidney dendritic cell activation is required for progression of renal disease in a mouse model of glomerular injury, J. Clin. Invest. 119 (2009) 1286–1297.
[43] J. Li, Y.F. Yu, C.H. Liu, C.M. Wang, Significance of M2 macrophages in glomerulonephritis with crescents, Pathol. Res. Pract. 213 (2017) 1215–1220.
[44] S. Moll, A. Angeletti, L. Scapozza, A. Cavalli, G.M. Ghiggeri, M. Prunotto, Glomerular macrophages in human auto- and allo-immune nephritis, Cells 10 (2021).
[45] M.P. Rastaldi, F. Ferrario, A. Crippa, G. Dell'antonio, D. Casartelli, C. Grillo, G. D'Amico, Glomerular monocyte-macrophage features in ANCA-positive renal vasculitis and cryoglobulinemic nephritis, J. Am. Soc. Nephrol. : JASN (J. Am. Soc. Nephrol.) 11 (2000) 2036–2043.
[46] R.S. Bhardwaj, C. Zotz, G. Zwadlo-Klarwasser, J. Roth, M. Goebeler, K. Mahnke, M. Falk, G. Meinardus-Hager, C. Sorg, The calcium-binding proteins MRP8 and MRP14 form a membrane-associated heterodimer in a subset of monocytes/macrophages present in acute but absent in chronic inflammatory lesions, Eur. J. Immunol. 22 (1992) 1891–1897.
[47] M. Antonelou, R.D.R. Evans, S.R. Henderson, A.D. Salama, Neutrophils are key mediators in crescentic glomerulonephritis and targets for new therapeutic approaches, Nephrol. Dial. Transplant. (2022) 230–238, official publication of the European Dialysis and Transplant Association - European Renal Association 37.
[48] Y. Han, J. Wu, Z. Gong, Y. Zhou, H. Li, Y. Chen, Q. Qian, Identification and development of the novel 7-genes diagnostic signature by integrating multi cohorts based on osteoarthritis, Hereditas 159 (2022) 10.
[49] Y. Shen, X. Li, D. Wang, L. Zhang, X. Li, L. Su, X. Fan, X. Yang, COL3A1: potential prognostic predictor for head and neck cancer based on immune-microenvironment alternative splicing, Cancer Med. (2022).
[50] L. Wang, Y. Sun, Z. Guo, H. Liu, COL3A1 overexpression associates with poor prognosis and cisplatin resistance in lung cancer, Balkan Med. J. 39 (2022) 393–400.
[51] S.W. Zhang, N. Zhang, N. Wang, Role of COL3A1 and POSTN on pathologic stages of esophageal cancer, Technol. Cancer Res. Treat. 19 (2020) 1533033820977489.
[52] P. Ahluwalia, M. Ahluwalia, A.K. Mondal, N. Sahajpal, V. Kota, M.V. Rojiani, A.M. Rojiani, R. Kolhe, Prognostic and therapeutic implications of extracellular matrix associated gene signature in renal clear cell carcinoma, Sci. Rep. 11 (2021) 7561.
[53] S. Li, H. Wang, Y. Zhang, R. Qiao, P. Xia, Z. Kong, H. Zhao, L. Yin, COL3A1 and MMP9 serve as potential diagnostic biomarkers of osteoarthritis and are associated with immune cell infiltration, Front. Genet. 12 (2021) 721258.
[54] H. Zhang, C. Ding, Y. Li, C. Xing, S. Wang, Z. Yu, L. Chen, P. Li, M. Dai, Data mining-based study of collagen type III alpha 1 (COL3A1) prognostic value and immune exploration in pan-cancer, Bioengineered 12 (2021) 3634–3646.
[55] Y. Shen, X. Li, D. Wang, L. Zhang, X. Li, L. Su, X. Fan, X. Yang, COL3A1: potential prognostic predictor for head and neck cancer based on immune-microenvironment alternative splicing, Cancer Med. 12 (2023) 4882–4894.