

Article

Not peer-reviewed version

Scissor–CIBERSORTx Deconvolution Reveals Functional Heterogeneity of CTAL/aTAL Cells and Associated Biomarkers in Renal Fibrosis

Hengping Wang, [Yuan Zhang](#), Jiale Li, [Ying Fu](#)^{*}, [Huiyan Wang](#)^{*}

Posted Date: 26 January 2026

doi: 10.20944/preprints202601.1831.v1

Keywords: renal fibrosis; single-cell RNA sequencing analysis; adaptive thick ascending limb cell; cortical thick ascending limb cell; Scissor; CIBERSORTx



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Scissor–CIBERSORTx Deconvolution Reveals Functional Heterogeneity of CTAL/aTAL Cells and Associated Biomarkers in Renal Fibrosis

Hengping Wang, Yuan Zhang, Jiale Li, Ying Fu * and Huiyan Wang *

Jilin Medical University, Jilin Collaborative Innovation Center for Antibody Engineering, Jilin 132013, China

* Correspondence: fuying94@126.com (Y.F.); zswwhy518@163.com (H.W.)

Abstract

Renal fibrosis (RF) represents a major pathological outcome of chronic kidney disease, currently accompanied by extremely limited therapeutic strategies. To decipher key cellular and molecular drivers, we integrated single-cell and bulk transcriptomic profiles for comprehensive analysis. Based on the RF-related single-cell and bulk transcriptomic data, key cell subtypes was identified through Scissor analysis, custom signature matrix construction via CIBERSORTx, and Weighted gene co-expression network analysis (WGCNA). Subsequently, key subtypes-related biomarkers were identified through the expression analysis, and functional enrichment analysis for biomarkers was conducted to elucidate the potential mechanisms by which biomarkers regulate RF. Through comprehensive profiling, thick ascending limb (TAL) cells were predominant and displayed marked heterogeneity in renal fibrosis (RF), with cortical TAL (CTAL) and adaptive TAL (aTAL) identified as principal subtypes. Five candidate biomarkers—STAT1, PARP8, HS6ST2, PTGER3, and TMEM207—were identified. Quantitative polymerase chain reaction (qPCR) validation in mouse models confirmed aberrant expression of these biomarkers, with STAT1 and PARP8 upregulated and HS6ST2, PTGER3, and TMEM207 downregulated in RF. Furthermore, functional enrichment analyses indicated that these biomarkers were associated with pathways underlying metabolic reprogramming and immune perturbation. Our study implicates CTAL and aTAL as central cellular players in RF and identified their associated biomarkers (STAT1, PARP8, HS6ST2, PTGER3, and TMEM207), providing experimentally supported novel biomarkers and repurposing drugs for therapeutic intervention.

Keywords: renal fibrosis; single-cell RNA sequencing analysis; adaptive thick ascending limb cell; cortical thick ascending limb cell; Scissor; CIBERSORTx

1. Introduction

Chronic kidney disease (CKD) is defined as abnormalities in kidney structure or function persisting for more than three months, with significant morbidity and mortality [1]. Renal fibrosis (RF) represents a key pathological feature and common endpoint of CKD [2], which can be initiated by various factors, including infections, ischemia, diabetes, autoimmune diseases, urinary tract obstruction, and toxic exposures [3,4]. Characteristic morphological changes of RF include glomerulosclerosis, tubular atrophy, chronic interstitial inflammation, fibrosis, and capillary rarefaction [1]. As RF progresses, the gradual loss of functional nephrons ultimately leads to end-stage renal disease [3]. Early detection of RF is therefore critical for timely intervention and improved patient prognosis. Currently, diagnosis of RF relies primarily on renal biopsy—an invasive procedure associated with risks such as bleeding [5]. Given this, exploring novel anti-fibrotic targets and therapeutic strategies represents a critical direction in current research.

Single-cell RNA sequencing (scRNA-seq) can reveal the heterogeneity of cells in tissues and the mechanism of disease [6,7], providing a new perspective for analyzing cell dynamics and interactions

during RF. However, despite its advantages over bulk RNA-seq in capturing cellular diversity within complex tissues, there are several limitations, including limited sample sizes, tissue dissociation bias, and insufficient sensitivity for detecting low-expression genes [8]. To overcome the limitations of the aforementioned methods, this study integrates CIBERSORTx with the Scissor algorithm to enable collaborative analysis of single-cell and batch transcriptomic data. Specifically, the Scissor algorithm correlates single-cell transcriptomes with clinical phenotypes to identify key cell subpopulations associated with disease states [9]. Meanwhile, CIBERSORTx constructs tissue-specific feature matrices from single-cell data and employs deconvolution analysis to resolve cell type composition and proportions from batch transcriptomic data. This combined approach systematically reveals phenotypically relevant cell subpopulations and their molecular characteristics during RF progression. [7,10,11]. Through this integrated analytical framework, key cellular subtypes and biomarkers implicated in RF progression were systematically delineated, thereby providing a conceptual framework and putative molecular targets for early diagnosis and precision therapeutic intervention. Furthermore, bioinformatically predicted biomarkers were validated by quantitative polymerase chain reaction (qPCR) in mouse models of renal injury and fibrosis, confirming their expression patterns and reinforcing their translational relevance. The study workflow is presented in Figure 1.

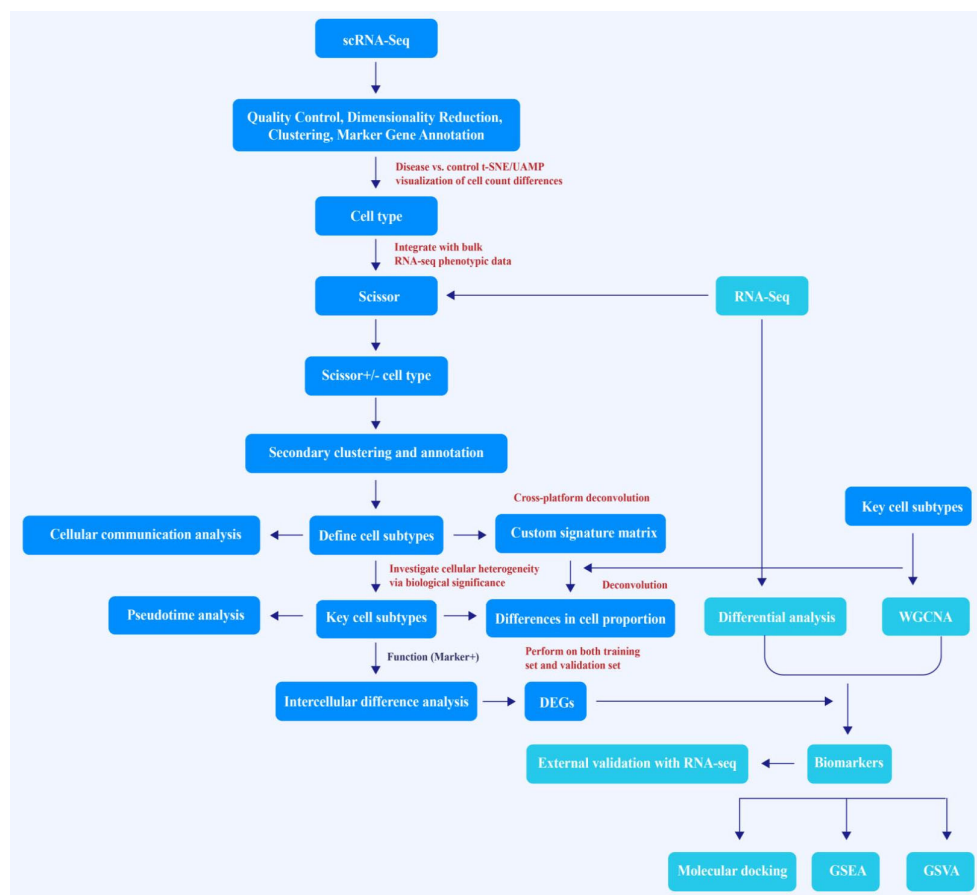


Figure 1. Schematic diagram of the workflow in this study. scRNA-seq, single-cell RNA sequencing; t-SNE, t-distributed Stochastic Neighbor Embedding; UMAP, Uniform Manifold Approximation and Projection; Scissor, Single-cell identification of subsets of interest via enrichment of regulators; DEGs, differentially expressed genes; WGCNA, Weighted Gene Co-expression Network Analysis; GSEA, Gene Set Enrichment Analysis; GSVA, Gene Set Variation Analysis.

2. Materials and Methods

2.1. Data Collection

Through the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>), the GSE195718 (platform: GPL24676), GSE76882 (platform: GPL13158), and GSE135327 (platform: GPL21290/GPL11154) datasets were acquired [12–14]. The GSE195718, a scRNA-seq dataset, comprises 3 control samples and 6 RF samples derived from normal and fibrotic kidney allografts, respectively. The bulk-RNA sequencing datasets GSE76882 (the training set) and GSE135327 (the validation set) contain 99 control and 135 RF samples, and 12 control and 18 RF samples, respectively.

2.2. Cell Annotation

The quality control was implemented in Seurat (v5.1.0) package on the GSE195718 dataset [15]. Data were filtered out under the following criteria: genes were detected in < 200 cells; cells with $nFeature_RNA \geq 7,000$; cells with $nCount_RNA \geq 30,000$; and mitochondrial content ($percent.mt$) $\geq 10\%$. Followed by logarithmic normalization, top 2,000 highly variable genes (HVGs) were identified via *vst* method for principal component analysis (PCA). Before PCA, the *ScaleData* function was applied to scaling the data. Based on the results of the scree plot and the PCA permutation test, the appropriate principal components (PCs) were selected for subsequent clustering analysis. Cell clustering was conducted applying the *FindNeighbors* and *FindClusters* functions ($resolution = 0.3$), followed by t-SNE-based visualization. Cell types were identified by comparing the expression of marker genes obtained from previous studies among cell clusters [12,16–20]. Furthermore, the *DoubletFinder* (v2.0.4) package was employed to detect and remove doublets, and the clustering situation of various types of cells and their proportion in the samples were further examined [21].

2.3. Scissor Analysis

The Scissor (v2.0.0) package was utilized to conduct association analysis in GSE195718 and GSE76882 datasets, to identify RF-associated cells types ($\alpha = 0.001$, $family = binomial$) [9]. Scissor+ denotes a positive association with the disease, Scissor- denotes a negative association, and a value of 0 indicates no statistically significant correlation. Building upon these identified RF-associated cell types, ROGUE (v1.0) package was applied to compute heterogeneity scores for these cells [22]. Lower scores reflect greater cellular heterogeneity. Further integrating the Scissor analysis results, we selected cell types with lower heterogeneity scores and performed secondary clustering analysis on them using the same analytical procedure [23,24]. For single-cell subclusters that could not be annotated to previously reported cells, the *FindMarkers* function was employed to identify differentially expressed genes (DEGs) with thresholds of $|\log_2 FoldChange (FC)| > 0.5$ and adjusted *P*-value < 0.05 between the non-annotated subclusters and other subclusters. Subsequent functional annotation was supported by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses performed on the identified DEGs. The proportion of each cell subtype within the sample was quantified and presented.

2.4. CIBERSORTx Analysis

In the GSE195718 dataset, the number of cells for each type was randomly downsampled through proportional sampling. Specifically, for cell types with more than 1,000 cells, a maximum of 1,000 cells were retained; and among these cells, 70% of the cells were resampled to meet the input requirements of the CIBERSORTx tool [10]. Subsequently, the *Create Signature Matrix* function was applied to generate deconvoluted single-cell data for constructing a custom signature matrix. Afterthat, utilizing the custom signature matrix, CIBERSORTx was applied to deconvolve bulk expression profiles from datasets GSE76882 and GSE135327. The inferred cell type proportions were then compared between the RF and control groups ($P < 0.05$).

2.5. WGCNA

WGCNA was performed via WGCNA (v1.71) package on the GSE76882 dataset to cluster samples and identify potential outliers for possible exclusion [25]. The soft threshold was selected to maximize the conformity of gene interactions to a scale-free distribution and to ensure that the mean connectivity tends toward zero ($R^2 > 0.85$). According to the standard of the dynamic tree cutting algorithm, the minimum module size was set to 200 genes to guide the cutting and merging of gene modules. The cell subtypes exhibiting significant differences in proportion between the RF and control groups were treated as traits to calculate correlations with gene co-expression modules via the Pearson algorithm. The module exhibiting the highest correlation with traits was selected as the key module. Genes in the key module were key module genes. Cell subtypes showing both the strongest correlations with the key module and statistically significant differences in proportion between groups in the GSE76882 and GSE135327 datasets, along with consistent trends, were identified as key cell subtypes.

2.6. Identification of Biomarkers Related to RF Progression in Key Cell Subtypes

Differential expression analysis was implemented using the limma (v3.52.4) package to identify notable DEGs between RF and control samples in the GSE76882 dataset ($|\log_2FC| > 1$ and adjusted $P < 0.05$) [26]. Meanwhile, within the custom signature matrix, the FindMarkers function was applied to perform differential expression analysis of key cell subtypes in comparison to all other subtypes, aiming to identify DEGs specific to these cell types. For each key cell type, the corresponding DEGs were intersected with both the DEGs between RF and control groups and the key module genes. Subsequently, expression levels of these intersected gene sets were analyzed in the GSE76882 and GSE135327 datasets. Genes exhibiting significant differential expression between the RF and the control groups, with consistent expression trends across both datasets, were designated as biomarkers. Expression levels of these were visualized in the GSE195718 dataset.

2.7. Cell Communication and Pseudotime Trajectory Analyses

The CellChat package (v1.6.1) was employed to conduct intercellular communication analysis in GSE195718 dataset, with particular emphasis on key cell subtypes. Additionally, pseudotime trajectory analysis was conducted adopting the monocle package (v2.26.0) on cell types to further investigate the functional dynamics of different cell types or states during RF progression [27]. Simultaneously, the expression of biomarkers during the pseudotime process was demonstrated.

2.8. Functional Enrichment Analysis

To investigate how varying biomarker expression levels influence pathway dynamics during RF progression, the clusterProfiler package (v4.7.1.001) was utilized for Gene Set Enrichment Analysis (GSEA) [28]. The gene-biomarker correlations in the GSE76882 dataset were computed, followed by gene ranking by correlation coefficient for GSEA ($|NES| > 1$ and adjusted $P < 0.05$). To further characterize pathway-level changes, Gene Set Variation Analysis (GSVA) was implemented. The GSVA package was utilized to quantify pathway activity scores in the GSE76882 dataset, and limma package was employed to conduct differential pathway analysis between the RF and control groups ($|t| > 2$ and $P < 0.05$). All analyses adopted the KEGG gene set (c2.cp.kegg.v7.0.symbols.gmt) from MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb>) as a reference set.

2.9. Animal Model

To validate bioinformatics findings related to renal injury and fibrosis, two well-established mouse models were employed in this study: the folic acid (FA)-induced acute kidney injury model and the unilateral ureteral obstruction (UUO)-induced chronic renal fibrosis model. These models were used to quantify the expression and temporal dynamics of five predicted biomarkers. Male C57BL/6 mice (6 weeks old, weighing 20–25 g) were purchased from Liaoning Changsheng

Biotechnology Co., Ltd. (Liaoning, China). All animals were housed in a specific pathogen-free (SPF) facility and randomly assigned into three groups: a wild-type normal control group (WT-normal group), a folic acid-induced acute kidney injury group (FA group), and a unilateral ureteral obstruction-induced chronic renal fibrosis group (UUO group). Mice in the FA group received a single intraperitoneal injection of folic acid (250 mg/kg), and samples were collected 48 hours later to establish the acute kidney injury model. Mice in the UUO group underwent unilateral ureteral ligation, and tissues were harvested after 14 days to induce chronic renal fibrosis and injury.

All animal experiments were conducted in strict accordance with ethical guidelines and were approved by the Animal Ethics Committee of Jilin Medical University (Approval No. JLYY-ETH-2024-007).

2.10. Quantitative PCR (qPCR)

Total RNA was extracted from each sample using TRIzol reagent (Thermo Fisher Scientific, California, USA). Approximately 2000 ng of total RNA was reverse-transcribed into complementary DNA (cDNA) using the PrimeScript RT reagent kit (Takara Bio, Japan). mRNA expression levels were quantified by real-time quantitative PCR with SYBR Green fluorescence dye on an Applied Biosystems real-time PCR system. All primers used in this study were synthesized by Sangon Biotech (Shanghai, China). The expression levels of target mRNAs were normalized to the internal control gene encoding ribosomal protein small subunit 16 (Rps16) for subsequent comparative analysis. Detailed primer sequences are listed in Table 1.

Table 1. RT-qPCR primers.

Gene	Forward primer(5'-3')	Reverse primer(5'-3')
<i>STAT1</i>	TCACAGTGGTTCGAGCTTCAG	GCAAACGAGACATCATAGGCA
<i>PARP8</i>	TAAATCGCACAAACTTTTGGGC	TCTCCAGAACAAGATCGAGTCAA
<i>HS6ST2</i>	ACCGGGGAAGTCAGAAGCA	CTCTACGCTCCCTAIGTAGTCAT
<i>PTGER3</i>	CCGGAGCACTCTGCTGAAG	CCCCACTAAGTCGGTGAGC
<i>TMEM207</i>	TGCTCTCGGATCTATCCTGTG	ATTCCGCACCTTTTCAGCCA
<i>mRps16</i>	CGTGCTTGCTCGGAGCTA	GCTCCTTGCCCAGAAGCAAA

2.11. Histological Analysis

For morphological evaluation, renal tissues from normal mice, FA-induced acute kidney injury mice, and UUO-induced chronic kidney fibrosis mice (14 days post-operation) were paraffin-embedded and sectioned at a thickness of 5 μ m. Subsequently, the sections were deparaffinized, rehydrated, and subjected to sequential staining with hematoxylin and eosin (H&E) and periodic acid-Schiff (PAS).

2.12. Drug Prediction and Molecular Docking

Potential therapeutic drug compounds targeting the biomarkers were predicted utilizing the DSigDB (<https://dsigdb.tanlab.org/>). Molecular docking was then conducted to assess the binding affinity between these biomarkers and their candidate drugs. Drug structures were selected based on their significance (P-value). Protein structures for the biomarkers were obtained from UniProt (<https://www.uniprot.org/>), while 3D compound structures were sourced from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Water molecules and small-molecule ligands were removed through AutoDock software, followed by molecular docking analysis conducted employing AutoDock Vina [29]. Ligand-receptor binding energies were computed, with values ≤ -5.0 kcal/mol defined as indicative of favorable binding activity.

2.13. Statistical Analysis

R software was utilized for statistical analyses. Differences between groups were assessed utilizing Wilcoxon test, with a P -value of < 0.05 . Statistical analysis for qPCR was performed using GraphPad Prism 8.0 software (GraphPad Software, San Diego, CA, USA). Differences among experimental groups were compared by one-way analysis of variance (ANOVA). Data are presented as the mean \pm standard error of the mean (SEM). A P value < 0.05 was considered statistically significant.

3. Results

3.1. Single-Cell Atlas Reveal RF-Associated Renal Cell Heterogeneity and Functional Differentiation of Thick Ascending Limb (TAL) Cells

Following data filtering, 47,387 cells, comprising 27,263 cells from RF samples and 20,124 cells from control, as well as 30,767 genes, were retained (Figure S1A). PCA was performed on the top 2,000 HVGs, and the top 30 PCs were selected for subsequent analysis (Figure S1B-D). Subsequently, cell type annotation was conducted on the 17 clusters derived from the clustering analysis, resulting in the identification of 13 distinct cell types, namely podocytes (PODO), proximal tubule (PT)/proximal convoluted tubule (PCT) cells, principle cells (PC), type A intercalated cells (AIC), type B intercalated cells (BIC), mesangial cells (MES), distal convoluted tubule cells (DCT), TAL, parietal epithelial cells (PEC), type 2 endothelial cells (Endo2), type 1 endothelial cells (Endo1), monocytes, and T cells (Figure 2A-B, Figure S1E). Following the removal of 3,554 high-confidence doublets (7.5% of the dataset), cell boundaries became more distinct with significantly reduced impact of technical noise on biological signals (Figure S1F). Among these cell types, the proportions of TAL and PT/PCT in the samples were both relatively high (Figure 2C). The Scissor algorithm demonstrated that a substantial proportion of cells in PT/PCT, TAL, Endo, Monocytes, and T populations exhibited positive associations with RF. Notably, a considerable number of cells within the PT/PCT, TAL, and AIC also displayed negative correlations with RF (Figure 2D). These findings suggest that these cell types might exhibit a certain degree of heterogeneity. This hypothesis was further supported by the ROGUE analysis, which revealed relatively lower scores for PT/PCT, TAL, and AIC (Figure 2E). Based on the integrated analysis of overall cell proportion, heterogeneity, and disease-related correlations, TAL was selected for secondary clustering. Among the 7 cell subclusters, cluster 0 was annotated as medullary TAL (MTAL), cluster 1 as macula densa (MD), cluster 2 as ascending thin limb cell (ATL), cluster 3 as cortical TAL (CTAL), cluster 5 as adaptive TAL (aTAL). Cluster 4 was involved in protein targeted transport and specific subcellular structure localization and was named metabolically reprogrammed progenitor TAL (MRPTAL); Cluster 6 was enriched in amino acid metabolism and membrane transport-related biological processes and was named metabolically adapted transport TAL (MATTAL) (Figure 2F-G, Figure S1G, Table S1). As shown in Figure 2H, the annotation cell types and subtypes were summarized.

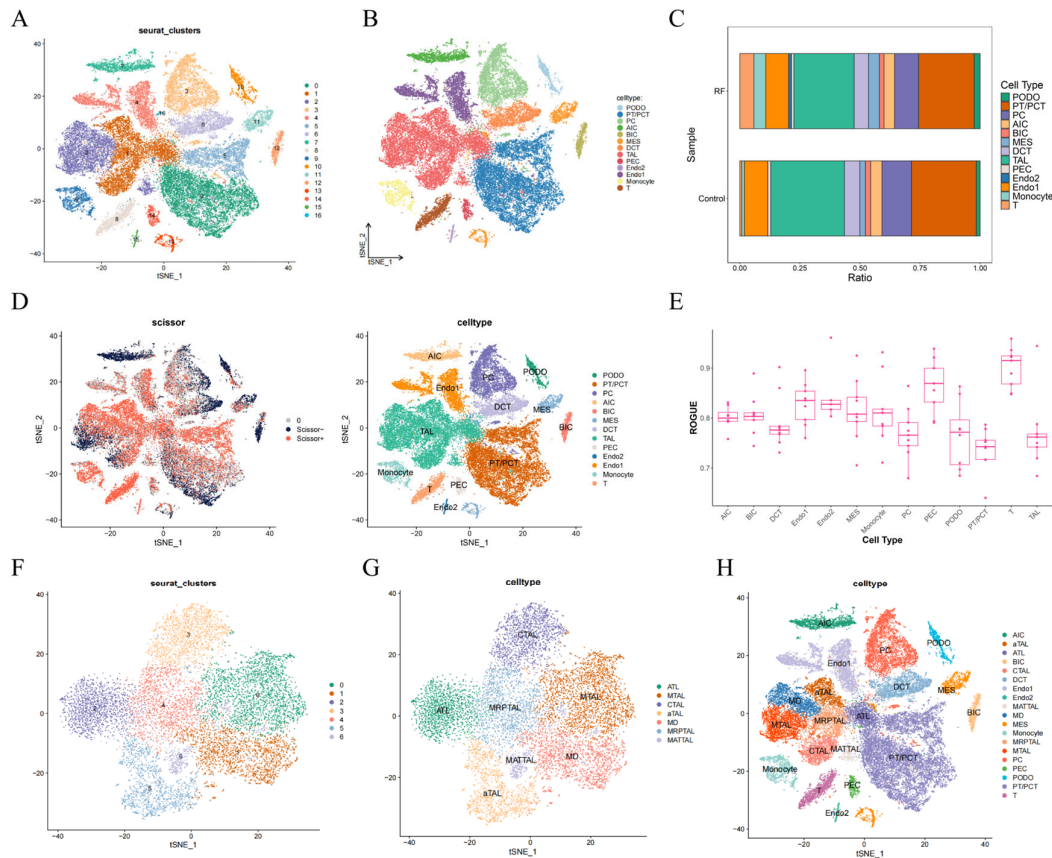


Figure 2. Identification of cell subtypes. (A) t-SNE clustering plot of 17 clusters in the GSE195718 dataset; (B) t-SNE plot illustrating the distribution of 13 cell types; (C) The cellular composition of each cell type in renal fibrosis (RF) and control groups; (D) t-SNE plots showing the correlation between cell types and RF; (E) The ROGUE scores for each cell type; (F) t-SNE plot illustrating the distribution of 7 subclusters of thick ascending limb (TAL); (G) t-SNE plot illustrating the distribution of 7 cell subtypes of TAL; (H) t-SNE plot illustrating the distribution of all cell types; Scissor+: positive correlation, Scissor-: negative correlation, 0: no significant correlation, ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

3.2. CTAL and aTAL Were Identified as the Key Cell Subtypes

Through the CIBERSORTx tool, a unique custom signature matrix was constructed (Supplementary Table 2). After deconvolution, we observed that the proportions of CTAL and MTAL (downregulated), as well as aTAL and ATL (upregulated), showed significant differences between groups in both GSE76882 and GSE135327 datasets (Figure 3A-C). In the subsequent WGCNA, clustering results showed that sample grouping was consistent, indicating no need to exclude any samples (Figure 3D). Using a soft threshold of 8, nine distinct modules were identified (Figure 3E-F). The brown (1,775 genes) and blue (2,924 genes) modules had strong correlations with the proportions of TAL cell subtypes, except for MTAL, and were therefore identified as key modules, containing 4,699 key module genes (Figure 3G). ATAL and CTAL displayed strong correlations with these key modules ($|\text{cor}| > 0.53$). Because of their significant differences between groups, they were classified as key cell subtypes.

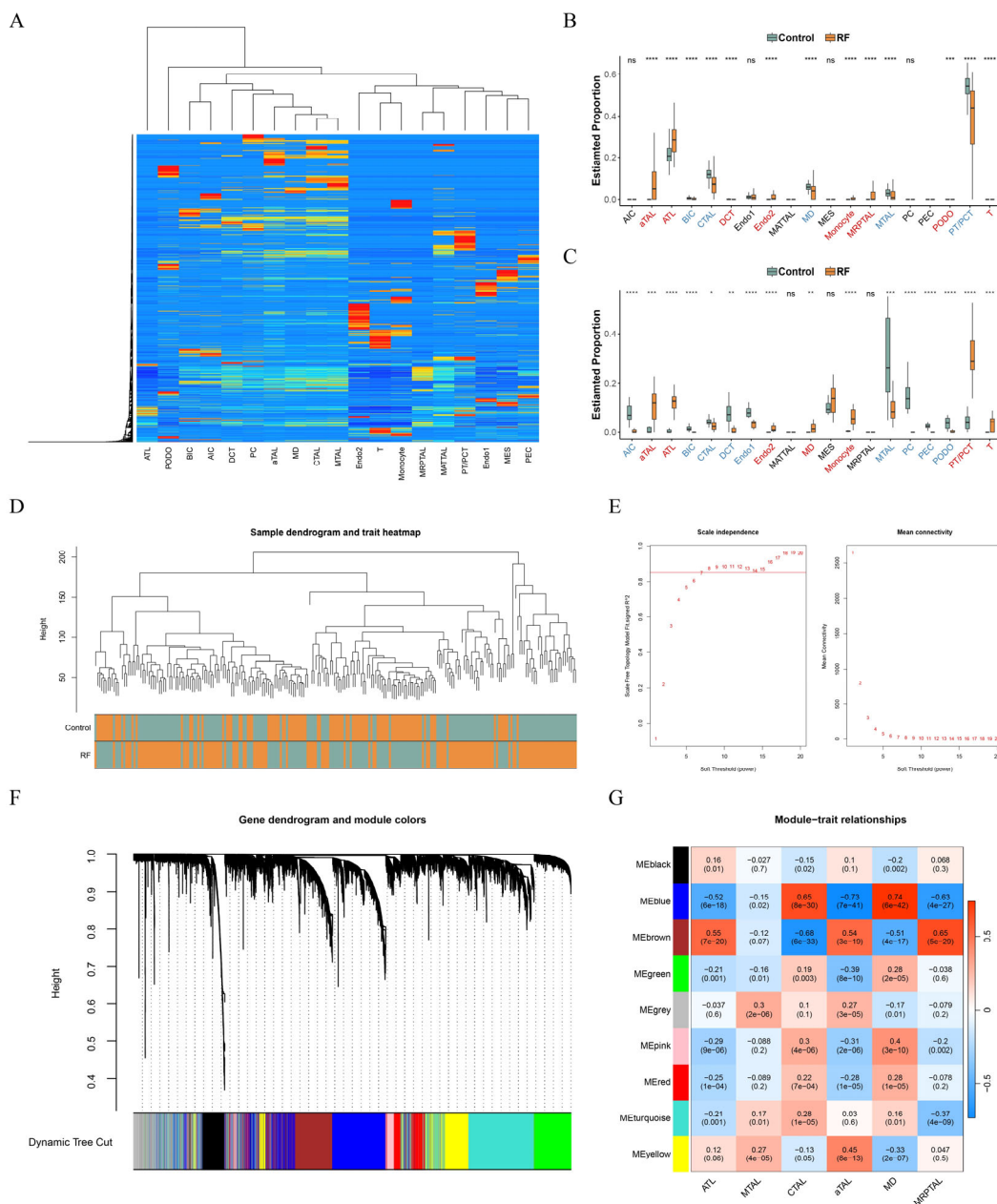


Figure 3. Identification of key cell subtypes from TAL. (A) Heatmap of CIBERSORTx deconvolution results. (B) Box plot showing the proportion of cells in the RF group and the control group in GSE76882 dataset; (C) Box plot showing the proportion of cells in the RF group and the control group in GSE135327 dataset; (D) Clustering results of WGCNA in GSE76882 dataset; (E) Selection of soft threshold; (F) Gene dendrogram and module colors; (G) Heat map of the correlation between modules and traits.

3.3. Cell Communication and Pseudotime Analysis of TAL Subtypes

Cell communication analysis revealed that the number and intensity of interactions involving aTAL did not show significant differences between the RF and control groups. In contrast, CTAL exhibited higher interaction counts and intensities in the control group compared to the RF group (Figure 4A-B). Moreover, given the critical role of TAL and its subtypes in RF, pseudotime analysis was performed to explore their dynamic changes during the progression. TAL differentiates into six

stages during the pseudotime development. CTAL was predominantly found in the early stage (stage 1), and aTAL mainly appeared in the middle stage (stage 4) (Figure 4C).

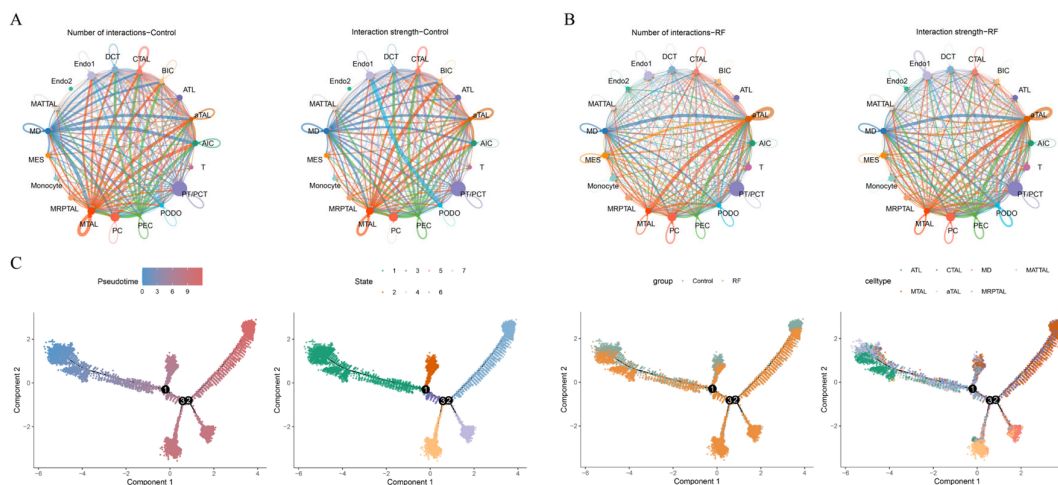


Figure 4. Cellular communication and pseudotime analysis. (A) Cellular communication network illustrating the number (left) and strength (right) of interactions among cell types in the control group; (B) Cellular communication network illustrating the number (left) and strength (right) of interactions among cell types in the RF group; (C) Time trajectory of TAL in pseudotime trajectory analysis (upper left); Cell state distribution of TAL in pseudotime trajectory analysis (upper right); Time trajectory of TAL in the RF group and the control group (lower left); Cell subtypes distribution of TAL in pseudotime trajectory analysis (lower right).

3.4. *STAT1*, *PARP8*, *HS6ST2*, *PTGER3*, and *TMEM207* Were Biomarkers for RF

In the GSE76882 dataset, there were 421 DEGs, including 158 downregulated and 263 upregulated genes in RF samples (Figure 5A). The heatmap separately showed the expression patterns of the top 20 upregulated and downregulated DEGs (Figure 5B). There were 378 DEGs between aTAL and other TAL subgroups, while CTAL had 174 DEGs. Thereafter, 378 DEGs in aTAL, 421 DEGs in the GSE76882 dataset, and 4,699 key module genes were intersected, resulting in the 6 intersected genes (Figure 5C). For CTAL, there were 5 intersected genes (Figure 5D). *STAT1* and *PARP8* in aTAL were significantly upregulated in the RF group across GSE76882 and GSE135327 datasets, while *HS6ST2*, *PTGER3*, and *TMEM207* in CTAL were significantly downregulated (Figure 5E-F). Therefore, these 5 genes served as potential biomarkers. The expression of all biomarkers in TAL cell subtypes was displayed in Figure 5G. During the differentiation of TAL cells, it was worth noting that the expressions of *PTGER3* and *HS6ST2* increased at the terminal stage of differentiation (Figure 5H).

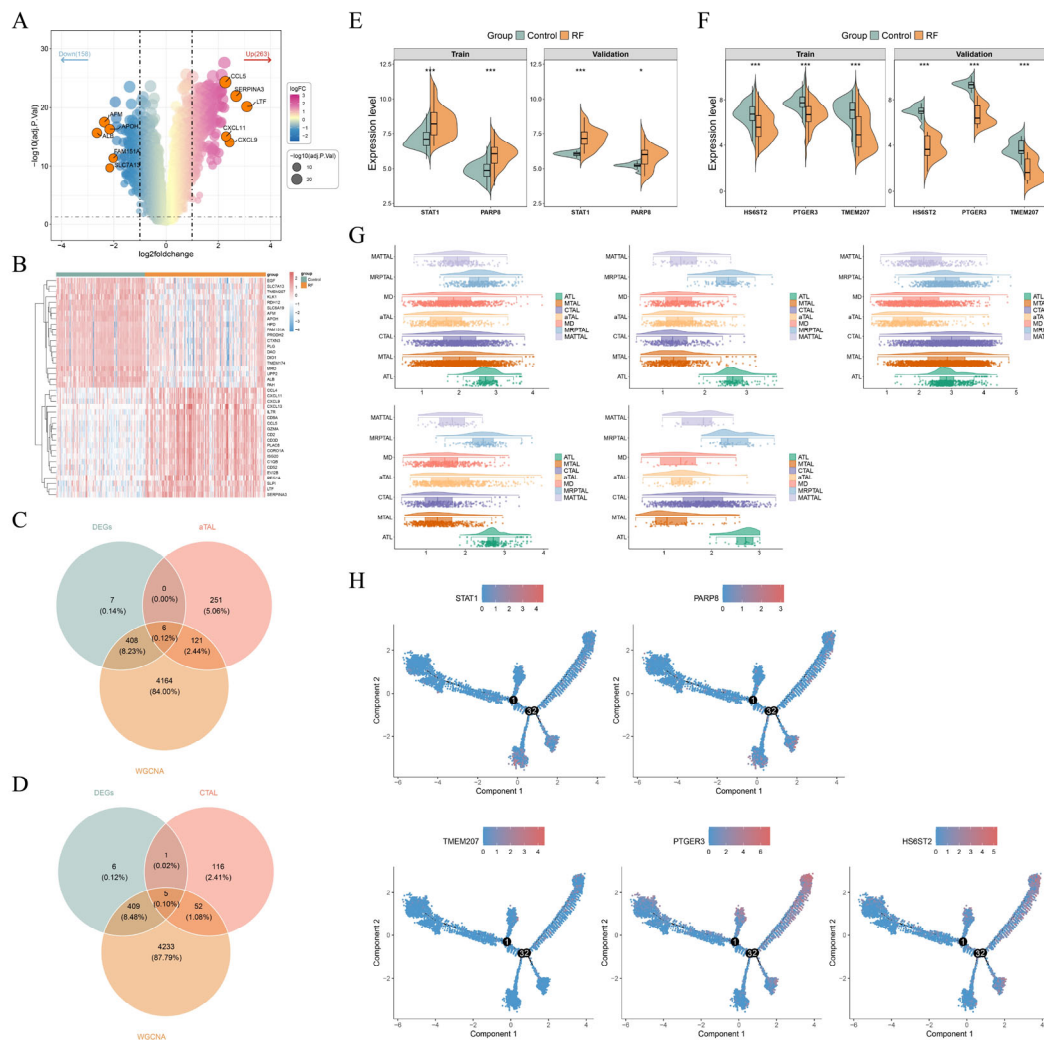


Figure 5. Identification of biomarkers. (A) Volcano plot of differentially expressed genes (DEGs) between the RF group and the control group in the GSE76882 dataset; (B) Heatmap of top 20 upregulated and downregulated DEGs in the GSE76882 dataset, ranked by $|\log_2FC|$; (C) Venn diagram of 378 DEGs in aTAL, 421 DEGs in GSE76882 dataset, and 4,699 key module genes; (D) Venn diagram of 174 DEGs in cTAL, 421 DEGs in GSE76882 dataset, and 4,699 key module genes; (E) Violin plot of STAT1 and PARP8 expression across samples in the GSE76882 and GSE135327 datasets; (F) Violin plot of HS6ST2, PTGER3, and TMEM207 expression across samples in the GSE76882 and GSE135327 datasets; (G) Expression of biomarkers in cell subtypes of TAL; (H) Biomarkers expression in pseudotime trajectory analysis. ns: not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

3.5. The Biomarkers Were Enriched in Metabolism and Immune Dysregulation Pathways

The function of the biomarker was analyzed, and the results showed STAT1, PARP8, HS6ST2, PTGER3, and TMEM207 were enriched in 73, 69, 67, 68, and 74 pathways, respectively (Figure 6A-E, Table S2). Their co-enrichment in energy metabolism pathways (e.g., oxidative phosphorylation, butanoate/propanoate metabolism) and immune dysregulation pathways (e.g., graft-versus-host disease, systemic lupus erythematosus) suggested dual roles in metabolic homeostasis and inflammatory injury responses. Notably, GSVA identified 155 pathways markedly different between the RF and control groups, including the aforementioned ones (Figure 6F, Table S3).

expansion, and fibrous tissue deposition. PAS staining further delineated structural alterations in the tubular basement membrane (TBM) and brush border. In the Normal group, the TBM was continuous and the brush border remained intact. The FA-2 d group exhibited disruption of the TBM and loss of the brush border, while the UUU-14 d group presented with fragmented TBM, marked tubular atrophy, and increased extracellular matrix deposition in the expanded interstitium. In summary, these histopathological findings confirm that the FA and UUU models successfully recapitulate the characteristic features of acute tubular injury and chronic renal fibrosis, respectively, and clearly distinguish the different stages of renal pathology (Figure 7A).

Real-time quantitative PCR analysis revealed differential expression of five core biomarkers in renal injury and fibrotic tissues compared with the normal control group (Fig 7B). Specifically, STAT1 and PARP8 were upregulated in disease models, whereas PTGER3 and TMEM207 were downregulated. Notably, HS6ST2 exhibited a temporospatial expression pattern: its expression was decreased in early acute kidney injury (FA model) but significantly increased during the chronic fibrotic phase (UUU model, 14 d). This expression profile aligned with the bioinformatics predictions. Previous temporal analysis indicated that HS6ST2 primarily functions in the CTAL, a segment susceptible to early damage. Consequently, HS6ST2 downregulation during the acute phase may reflect early CTAL damage, whereas its upregulation in chronic fibrosis may indicate compensatory and/or repair-associated responses. These findings are consistent with our earlier prediction that HS6ST2 expression is markedly elevated in the late stage of differentiation (Figure 5H). Collectively, the experimental results validate the biomarker expression patterns predicted by bioinformatics analysis.

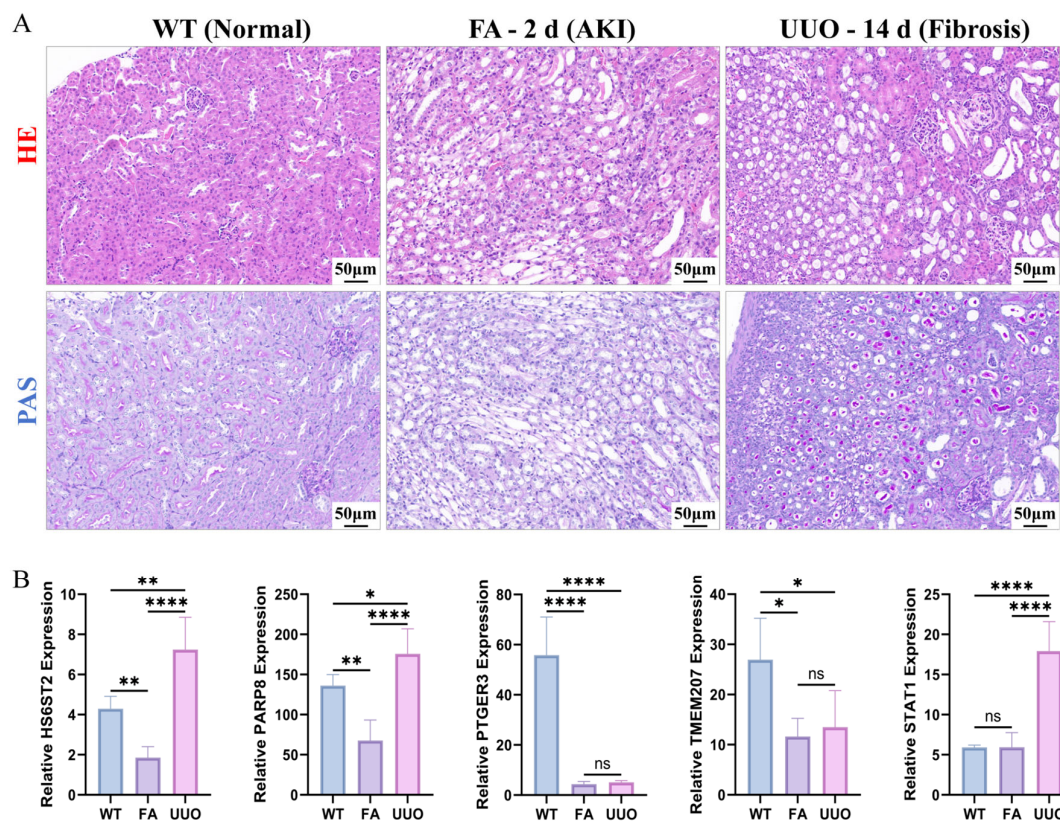


Figure 7. Histological examination and biomarker expression analysis in mouse renal tissues. (A) Representative images of hematoxylin & eosin (H&E) and periodic acid–Schiff (PAS) staining of kidney sections from the normal control group (WT), the folic acid-induced acute kidney injury group (FA, AKI), and the unilateral ureteral

obstruction-induced chronic renal fibrosis group (UUO, Fibrosis). Staining was used to evaluate renal tubular and glomerular structure and to assess pathological injury. Scale bar: 50 μm ; (B) mRNA expression levels of HS6ST2, PARP8, PTGER3, TMEM207, and STAT1 in kidney tissues of the WT, FA, and UUO groups were measured by quantitative real-time PCR (qPCR) and normalized to the reference gene Rps16 (ribosomal protein small subunit 16). Bar graphs show the relative expression of the five predicted biomarkers across the three groups. Data are presented as mean \pm SD; ns, not significant ($P > 0.05$); * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

3.7. The Biomarkers Exhibited Strong Binding Affinity with Their Targeted Drugs

Through DSigDB, 131 drugs targeted biomarkers were predicted. Among them, STAT1 was associated with 118 drugs, PARP8 with 3 drugs, and PTGER3 with 18 drugs. Unfortunately, there were no corresponding drugs for HS6ST2 and TMEM207 (Figure 8A). Amcinonide, zidovudine, and epigallocatechin gallate were selected for molecular docking based on their lowest P-values. The amino acid residues of STAT1 formed 3 hydrogen bonds with amcinonide, exhibiting a molecular binding energy of -7.13 kcal/mol (Figure 8B). PTGER3 established 5 hydrogen bonds with zidovudine, showing a binding energy of -6.64 kcal/mol (Figure 8C). PARP8 interacted with epigallocatechin gallate through 2 hydrogen bonds, with a binding energy of 7.57 kcal/mol (Figure 8D, Table 2).

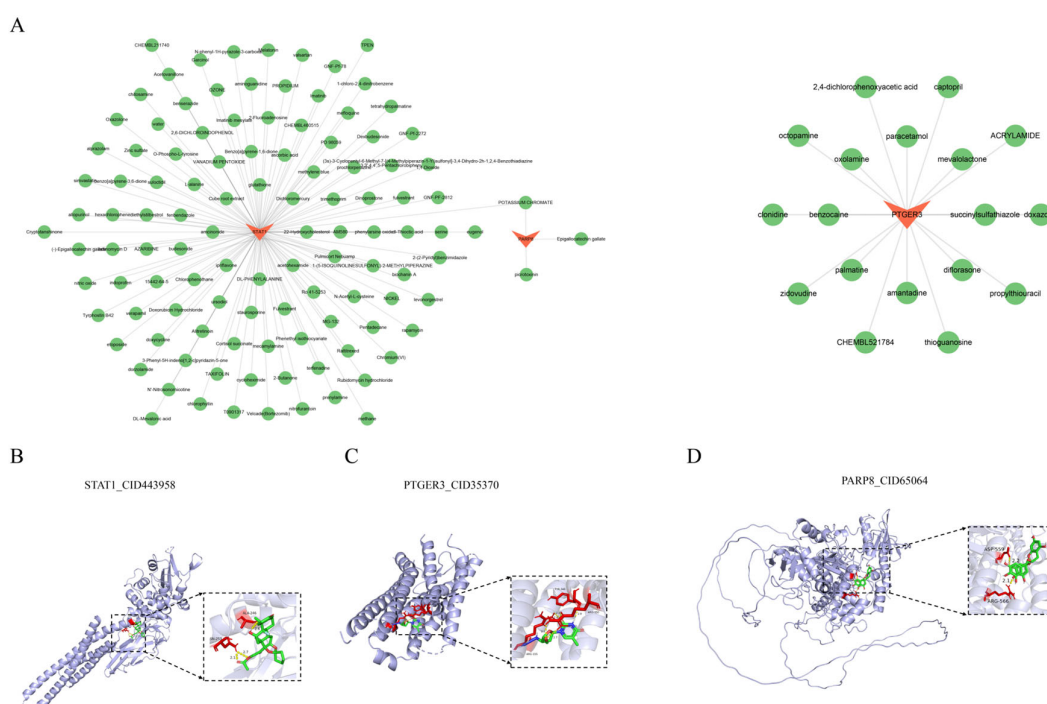


Figure 8. Drug prediction and molecular docking of biomarkers. (A) Drug-gene network of biomarkers. (B) Molecular docking between STAT1 and amcinonide. (C) Molecular docking between PTGER3 and zidovudine. (D) Molecular docking between PARP8 and epigallocatechin gallate.

Table 2. Binding energy of biomarkers with targeted compounds.

Symbol	UniProt Accession	Molecule Name	CID	Affinity (kcal/mol)	hydrogen bonds
STAT1	P42224	Amcinonide	CID443958	-7.13	3
PTGER3	P43115	zidovudine	CID35370	-6.64	5
PARP8	Q8N3A8	Epigallocatechin Gallate	CID65064	-7.57	2

4. Discussion

The pronounced cellular heterogeneity and complex intercellular communication within the renal fibrotic microenvironment have long constituted major obstacles to elucidating disease mechanisms and developing targeted therapeutic strategies. [30,31]. Conventional bulk transcriptomics lacks the resolution to distinguish distinct cellular subtypes, whereas standalone single-cell analyses often fail to establish direct associations with clinical phenotypes. To overcome these limitations, this study innovatively integrated the Scissor algorithm with a customized CIBERSORTx signature matrix constructed from single-cell data. This strategy not only enabled high-resolution deconvolution of cell-type composition from bulk data but also allowed precise phenotypic association at the subpopulation level. Using this approach, we identified CTAL and aTAL cells as two key and functionally distinct TAL subtypes involved in RF progression, and further discovered five closely associated biomarkers: STAT1, PARP8, HS6ST2, PTGER3, and TMEM207.

The thick ascending limb plays a crucial role in maintaining water-electrolyte homeostasis and tubuloglomerular feedback, and its injury-induced cellular plasticity under chronic stress is considered a key driver of fibrosis [32–34]. Our findings indicate that CTAL is predominantly enriched in the early pseudotime stage of differentiation, with its proportion decreased and cell-cell communication significantly weakened in RF. This suggests that CTAL dysfunction may be an early event in RF, potentially initiating pathological processes by disrupting normal filtration feedback regulation [35–38]. In contrast, aTAL—a subtype emerging under chronic stress—is mainly present at intermediate pseudotime stages and shows a marked increase in proportion in RF. aTAL may represent a metabolically and functionally reprogrammed phenotype adopted by TAL cells to adapt to injury. Its sustained intercellular communication activity may reflect a compensatory attempt to preserve homeostasis; however, it may also inadvertently amplify profibrotic signaling [39].

At the molecular level, the five biomarkers form a network linked to TAL subtype functions. STAT1 (upregulated in aTAL) is a known pro-inflammatory and pro-fibrotic transcription factor whose activation exacerbates renal inflammation [40–42]. PARP8 (upregulated in aTAL), a member of the PARP family, is structurally related to PARP1, which has been shown to promote fibrosis through mediating inflammation, oxidative stress, and other pathways [43–49]. Conversely, HS6ST2, PTGER3, and TMEM207 (all downregulated in CTAL) are implicated in key physiological processes, including the maintenance of basement membrane integrity [50], anti-inflammatory and vasodilatory protection [51,52], and epithelial cell polarity/transport [53], respectively. Their collective downregulation likely compromises tubular structural integrity and defensive capacity.

Importantly, the expression patterns of these biomarkers were independently validated by qPCR in two well-established mouse models: FA-induced AKI and unilateral UUO-induced chronic renal fibrosis. The experimental results corroborated the bioinformatic predictions: STAT1 and PARP8 were significantly upregulated in disease groups; PTGER3 and TMEM207 were downregulated; and HS6ST2 exhibited a dynamic temporal profile, with initial downregulation during the acute phase followed by upregulation in the chronic phase, consistent with its predicted increase in late pseudotime (Figure 5H). This validation supports the robustness of the computational analyses and identifies HS6ST2 as a stage-specific biomarker whose dynamic expression may reflect the transition from early injury to late fibrotic remodeling. Gene set enrichment analysis further revealed that these biomarkers are collectively associated with metabolic reprogramming pathways (e.g., oxidative phosphorylation, butanoate metabolism) and immune dysregulation pathways (e.g., graft-versus-host disease, lupus). This outlines a complex interplay between energy metabolism disturbances and immune-inflammatory responses during RF progression. Building on this molecular network, our drug prediction efforts identified several candidate compounds with high binding affinity, such as amcinonide (targeting STAT1), zidovudine (targeting PTGER3), and epigallocatechin gallate (targeting PARP8), offering new avenues for drug repurposing or combination therapy.

Several limitations warrant consideration, including reliance on public datasets that may introduce batch-related confounding and the need for additional functional experiments to elucidate the molecular mechanisms governing CTAL/aTAL subtype transitions and biomarker activity.

Nevertheless, by integrating multi-omics computational analyses with in vivo experimental validation, this study systematically delineates the central contribution of TAL cellular heterogeneity to RF and identifies subtype-specific biomarkers and candidate therapeutic targets with translational relevance. Consequently, these findings provide a foundation for the future development of cell-resolved precision diagnostic and targeted therapeutic strategies for kidney disease.

5. Conclusions

In conclusion, this study systematically analyzed TAL cell heterogeneity under RF conditions and identified two principal subtypes (CTAL and aTAL) by integrating single-cell and bulk transcriptomic datasets. Based on this, five biomarkers (STAT1, PARP8, HS6ST2, PTGER3, TMEM207) closely associated with these subtypes were identified, and their enrichment patterns in metabolic reprogramming and immune dysregulation pathways were clarified. This provides a molecular mechanistic framework for elucidating the interactions among inflammation, metabolism, and fibrosis during RF progression. Future investigations should validate these findings in independent cohorts and conduct functional studies to define the regulatory mechanisms underlying TAL subtype state transitions and biomarker function, thereby advancing the development of cell-targeted therapeutic strategies for kidney disease.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Figure S1: Single-cell RNA sequencing (scRNA-seq) data processing; Table S1: The list of Gene Ontology (GO) items and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched for differentially expressed genes (DEGs) between clusters in thick ascending limb (TAL); Table S2: The list of pathways enriched for biomarkers in Gene Set Enrichment Analysis (GSEA); Table S3: The list of pathways in Gene Set Variation Analysis (GSVA).

Author Contributions: Conceptualization, Y.F. (Ying Fu) and H.Y.W. (Huiyan Wang); Methodology, Y.Z. (Yuan Zhang) and J.L. (Jiale Li); Software, J.L.; Validation, H.W. (Hengping Wang) and Y.Z.; Formal Analysis, Y.Z.; Investigation, H.W. and J.L.; Resources, Y.F. and H.Y.W.; Data Curation, H.W.; Writing—Original Draft Preparation, H.W.; Writing—Review and Editing, Y.F., H.Y.W. and Y.Z.; Visualization, J.L.; Supervision, Y.F. and H.Y.W.; Project Administration, Y. F. and H.Y. W.; Funding Acquisition, Y. F. and H.Y.W.

Funding: This project was supported by the Jilin Provincial Science and Technology Development Plan Project (Grant No. 20240404022ZP).

Institutional Review Board Statement: The animal study protocol were approved by the Animal Ethics Committee of Jilin Medical University (Approval No. JLYY-ETH-2024-007).

Informed Consent Statement: Not applicable.

Data Availability Statement: Relevant information is included in the article.

Acknowledgments: We would like to express our sincere gratitude to Professors Binghai Zhao and Hongzhi Li for their valuable technical support in the qPCR experiments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Huang, R., P. Fu, and L. Ma, *Kidney fibrosis: from mechanisms to therapeutic medicines*. Signal Transduct Target Ther, 2023. 8(1): p. 129.
2. Lv, W., et al., Inflammation and renal fibrosis: Recent developments on key signaling molecules as potential therapeutic targets. Eur J Pharmacol, 2018. 820: p. 65-76.
3. Ammirati, A.L., *Chronic Kidney Disease*. Rev Assoc Med Bras (1992), 2020. 66Suppl 1(Suppl 1): p. s03-s09.
4. Li, L., H. Fu, and Y. Liu, *The fibrogenic niche in kidney fibrosis: components and mechanisms*. Nat Rev Nephrol, 2022. 18(9): p. 545-557.

5. Guo, Y., et al., Construction of a neural network diagnostic model for renal fibrosis and investigation of immune infiltration characteristics. *Front Immunol*, 2023. **14**: p. 1183088.
6. Jovic, D., et al., Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med*, 2022. **12**(3): p. e694.
7. Xie, S., et al., Single-cell transcriptome analysis reveals heterogeneity and convergence of the tumor microenvironment in colorectal cancer. *Front Immunol*, 2022. **13**: p. 1003419.
8. Valenzi, E., et al., Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. *Ann Rheum Dis*, 2019. **78**(10): p. 1379-1387.
9. Sun, D., et al., Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol*, 2022. **40**(4): p. 527-538.
10. Newman, A.M., et al., Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*, 2019. **37**(7): p. 773-782.
11. Zhu, H., et al., Fibroblast Subpopulations in Systemic Sclerosis: Functional Implications of Individual Subpopulations and Correlations with Clinical Features. *J Invest Dermatol*, 2024. **144**(6): p. 1251-1261.e13.
12. McDaniels, J.M., et al., The cellular landscape of the normal kidney allograft: Main players balancing the alloimmune response. *Front Transplant*, 2022. **1**: p. 988238.
13. Modena, B.D., et al., Gene Expression in Biopsies of Acute Rejection and Interstitial Fibrosis/Tubular Atrophy Reveals Highly Shared Mechanisms That Correlate With Worse Long-Term Outcomes. *Am J Transplant*, 2016. **16**(7): p. 1982-98.
14. He, X., et al., Myofibroblast YAP/TAZ activation is a key step in organ fibrogenesis. *JCI Insight*, 2022. **7**(4).
15. Hao, Y., et al., *Integrated analysis of multimodal single-cell data*. *Cell*, 2021. **184**(13): p. 3573-3587.e29.
16. Tsai, Y.C., et al., Single-cell transcriptomic profiles in the pathophysiology within the microenvironment of early diabetic kidney disease. *Cell Death Dis*, 2023. **14**(7): p. 442.
17. Wilson, P.C., et al., *The single-cell transcriptomic landscape of early human diabetic nephropathy*. *Proc Natl Acad Sci U S A*, 2019. **116**(39): p. 19619-19625.
18. Lu, Y.A., et al., Single-Nucleus RNA Sequencing Identifies New Classes of Proximal Tubular Epithelial Cells in Kidney Fibrosis. *J Am Soc Nephrol*, 2021. **32**(10): p. 2501-2516.
19. Chen, Z., et al., Single cell multi-omics of fibrotic kidney reveal epigenetic regulation of antioxidation and apoptosis within proximal tubule. *Cell Mol Life Sci*, 2024. **81**(1): p. 56.
20. Doke, T., et al., Single-cell analysis identifies the interaction of altered renal tubules with basophils orchestrating kidney fibrosis. *Nat Immunol*, 2022. **23**(6): p. 947-959.
21. McGinnis, C.S., L.M. Murrow, and Z.J. Gartner, DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*, 2019. **8**(4): p. 329-337.e4.
22. Liu, B., et al., An entropy-based metric for assessing the purity of single cell populations. *Nat Commun*, 2020. **11**(1): p. 3155.
23. Lake, B.B., et al., An atlas of healthy and injured cell states and niches in the human kidney. *Nature*, 2023. **619**(7970): p. 585-594.
24. Muto, Y., et al., Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. *Nat Commun*, 2021. **12**(1): p. 2190.
25. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. *BMC Bioinformatics*, 2008. **9**: p. 559.
26. Ritchie, M.E., et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, 2015. **43**(7): p. e47.
27. Trapnell, C., et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*, 2014. **32**(4): p. 381-386.
28. Yu, G., et al., clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*, 2012. **16**(5): p. 284-7.
29. Saikia, S. and M. Bordoloi, Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. *Curr Drug Targets*, 2019. **20**(5): p. 501-521.
30. Ouyang, Q., et al., Depleting profibrotic macrophages using bioactivated in vivo assembly peptides ameliorates kidney fibrosis. *Cell Mol Immunol*, 2024. **21**(8): p. 826-841.

31. Xia, W., et al., Long Non-coding RNA: An Emerging Contributor and Potential Therapeutic Target in Renal Fibrosis. *Front Genet*, 2021. **12**: p. 682904.
32. Mount, D.B., *Thick ascending limb of the loop of Henle*. *Clin J Am Soc Nephrol*, 2014. **9**(11): p. 1974-86.
33. Layton, A.T. and A. Edwards, Tubuloglomerular feedback signal transduction in a short loop of henle. *Bull Math Biol*, 2010. **72**(1): p. 34-62.
34. Simon, N. and A. Hertig, Alteration of Fatty Acid Oxidation in Tubular Epithelial Cells: From Acute Kidney Injury to Renal Fibrogenesis. *Front Med (Lausanne)*, 2015. **2**: p. 52.
35. Sipos, A., S. Vargas, and J. Peti-Peterdi, *Direct demonstration of tubular fluid flow sensing by macula densa cells*. *Am J Physiol Renal Physiol*, 2010. **299**(5): p. F1087-93.
36. Pihl, L., et al., Insulin induces the correlation between renal blood flow and glomerular filtration rate in diabetes: implications for mechanisms causing hyperfiltration. *Am J Physiol Regul Integr Comp Physiol*, 2012. **303**(1): p. R39-47.
37. Sällström, J., et al., High-protein-induced glomerular hyperfiltration is independent of the tubuloglomerular feedback mechanism and nitric oxide synthases. *Am J Physiol Regul Integr Comp Physiol*, 2010. **299**(5): p. R1263-8.
38. Wisman, M., et al., Dysregulated cross-talk between alveolar epithelial cells and stromal cells in idiopathic pulmonary fibrosis reduces epithelial regenerative capacity. *Front Med (Lausanne)*, 2023. **10**: p. 1182368.
39. Gonzalez-Vicente, A., et al., Thick Ascending Limb Sodium Transport in the Pathogenesis of Hypertension. *Physiol Rev*, 2019. **99**(1): p. 235-309.
40. Li, M., et al., USP22 deficiency in melanoma mediates resistance to T cells through IFN γ -JAK1-STAT1 signal axis. *Mol Ther*, 2021. **29**(6): p. 2108-2120.
41. Fu, Y., et al., The STAT1/HMGB1/NF- κ B pathway in chronic inflammation and kidney injury after cisplatin exposure. *Theranostics*, 2023. **13**(9): p. 2757-2773.
42. Liu, Y., et al., High salt aggravates renal inflammation via promoting pro-inflammatory macrophage in 5/6-nephrectomized rat. *Life Sci*, 2021. **274**: p. 119109.
43. Lee, J.S., J.Y. Lim, and J. Kim, Mechanical stretch induces angiotensinogen expression through PARP1 activation in kidney proximal tubular cells. *In Vitro Cell Dev Biol Anim*, 2015. **51**(1): p. 72-8.
44. Kim, J. and B.J. Padanilam, Loss of poly(ADP-ribose) polymerase 1 attenuates renal fibrosis and inflammation during unilateral ureteral obstruction. *Am J Physiol Renal Physiol*, 2011. **301**(2): p. F450-9.
45. Yoon, S.P. and J. Kim, Poly(ADP-ribose) polymerase 1 activation links ischemic acute kidney injury to interstitial fibrosis. *J Physiol Sci*, 2015. **65**(1): p. 105-11.
46. Cheng, L., et al., MiR-194 targets Runx1/Akt pathway to reduce renal fibrosis in mice with unilateral ureteral obstruction. *Int Urol Nephrol*, 2020. **52**(9): p. 1801-1808.
47. John, A., et al., GYY4137, a Hydrogen Sulfide Donor Modulates miR194-Dependent Collagen Realignment in Diabetic Kidney. *Sci Rep*, 2017. **7**(1): p. 10924.
48. Lucarini, L., et al., HYDAMTIQ, a selective PARP-1 inhibitor, improves bleomycin-induced lung fibrosis by dampening the TGF- β /SMAD signalling pathway. *J Cell Mol Med*, 2017. **21**(2): p. 324-335.
49. Huang, D., et al., Poly(ADP-ribose) polymerase 1 is indispensable for transforming growth factor- β induced Smad3 activation in vascular smooth muscle cell. *PLoS One*, 2011. **6**(10): p. e27123.
50. Stöcker, G., et al., Development of an enzyme immunoassay specific for a core protein epitope of a novel small basement membrane associated heparan sulphate proteoglycan from human kidney. *Eur J Clin Chem Clin Biochem*, 1997. **35**(2): p. 95-9.
51. Yu, Y., et al., PTGER3 and MMP-2 play potential roles in diabetic nephropathy via competing endogenous RNA mechanisms. *BMC Nephrol*, 2021. **22**(1): p. 27.

52. Nakamoto, S., et al., EP3 signaling in dendritic cells promotes liver repair by inducing IL-13-mediated macrophage differentiation in mice. *Faseb j*, 2020. **34**(4): p. 5610-5627.
53. Chen, S., et al., Knockdown of TMEM30A in renal tubular epithelial cells leads to reduced glucose absorption. *BMC Nephrol*, 2023. **24**(1): p. 250.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.