Multiomic analysis of human kidney disease identifies a tractable inflammatory, pro-fibrotic tubular cell phenotype

Bryan Conway

bryan.conway@ed.ac.uk

University of Edinburgh

Maximilian Reck

University of Edinburgh

David Baird

University of Edinburgh

Callum Sutherland

University of Edinburgh

Rachel Bell

University of Edinburgh

Heeyoun Hur

University of Edinburgh

Carolynn Cairns

University of Edinburgh

Ross Campbell

University of Edinburgh

Andy Nam

NanoString, Technologies

Wei Yang

NanoString, Technologies

Nathan Schurman

NanoString, Technologies

Claire Williams

NanoString, Technologies

Stefan Veizades

University of Edinburgh

Eoin O'Sullivan

University of Queensland

Andrea Corsinotti

University of Edinburgh
Christopher Bellamy
University of Edinburgh
Jeremy Hughes
University of Edinburgh
Alexander Laird
University of Edinburgh
Laura Denby
University of Edinburgh
Tamir Chandra
University of Edinburgh
David Ferenbach
University of Edinburgh

Article

Keywords:

Posted Date: February 20th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3964901/v1

Additional Declarations: Yes there is potential Competing Interest. WY and CW are employees and shareholders of NanoString Technologies, Inc

---

Abstract

Maladaptive proximal tubular cells have been implicated in failure of repair following renal injury in rodent models, however whether this translates to human kidney disease is unknown. Hence, we integrated snRNA-ATAC-seq with single-cell molecular imaging to generate a multiomic atlas of human kidney disease. In injured kidneys, a subset of tubular epithelial cells acquired an inflammatory phenotype, enriched with pro-fibrotic and senescence markers, analogous to maladaptive cells in mice. Cell neighborhood analysis positioned the inflammatory phenotype adjacent to leucocytes and myofibroblasts and ligand-receptor analysis highlighted paracrine signaling from inflammatory tubular cells to mediate leucocyte recruitment and myofibroblast activation. Loss of an HNF4α-driven gene regulatory network and activation of NF-κβ and AP-1 transcription factors epigenetically imprinted the inflammatory phenotype. Targeting these inflammatory tubular cells by administration of an AP-1 inhibitor or a senolytic agent ameliorated inflammation, expression of senescence-associated transcripts and fibrosis in murine models of kidney injury suggesting these as therapies for human kidney disease.

Introduction

Human kidneys comprise multiple different cell types in a complex anatomical arrangement that is optimal for their key functions including removal of waste, regulation of fluid and electrolyte homeostasis and production of hormones. In chronic kidney disease (CKD), this arrangement is disrupted with de-differentiation and atrophy of epithelial cells, recruitment of leucocytes and activation of myofibroblasts(1, 2). An improved understanding of the spatial arrangement of these cells and their intercellular signaling pathways could lead to development of new therapies.

Recent advances in single-cell technologies including single nuclear RNA-sequencing (snRNA-seq) and assay for transposase-accessible chromatin sequencing (ATAC-seq) have facilitated the development of atlases of the human and murine kidney in health and disease(3, 4, 5, 6, 7, 8). These studies have characterized maladaptive tubular cell phenotypes characterized by expression of *VCAM1* that are present in normal kidney, but are expanded in acute kidney injury (AKI) and CKD(4, 8, 9, 10, 11). In murine studies, a subset of the *Vcam1 +* proximal tubular cells express *Ccl2* and other pro-inflammatory, pro-fibrotic and senescence-associated genes and persist after ischaemic renal injury(12, 13). However, the relevance of this *Ccl2*$^{+}$ subset to human disease and whether it may be targeted to prevent maladaptive repair progressing to CKD is unknown.

To this end, we performed simultaneous snRNA-seq and ATAC-seq in non-tumour bearing tissue from tumour nephrectomy specimens from patients with otherwise healthy kidneys and those in which the tumour obstructed the ureter leading to inflammatory fibrosis. Specifically in obstructed kidneys, a discrete subset of tubular epithelial cells adopted an inflammatory phenotype resembling the *Ccl2*$^{+}$ tubular cells in mice including expression of adhesion molecules, chemokines and genes consistent with activation of a senescence program and additionally the senescence marker, *CDKN1A*. High-plex single-cell molecular imaging of biopsies from patients with CKD localized this inflammatory phenotype to the

inflamed, fibrotic niche. Integration of gene expression and ATAC data implicated HNF4-α in maintaining proximal tubular cell health and AP-1 signaling in promoting the inflammatory tubular phenotype. Finally, administration of a small molecule inhibitor of AP-1 or the senolytic agent ABT-263 reduced expression of genes consistent with an inflammatory tubular cell phenotype and ameliorated inflammation and fibrosis in murine models of transition from acute kidney injury (AKI) to CKD.

Results

Generation of an atlas of healthy and obstructed kidneys

To characterize the cellular and molecular landscape in the healthy and injured human kidney, we performed simultaneous snRNA-seq and ATAC-seq on tissue from the non-tumorous pole of nephrectomies from 5 patients with urothelial carcinoma causing unilateral ureteric obstruction (UUO) as a paradigm of tubular injury, with 7 patients who had renal cell carcinoma without urinary obstruction constituting our controls. The clinical and pathological characteristics of the patients are given in **Tables S1,2** and **Fig.S1**. All controls had normal renal function (eGFR > 75ml/min/m^2) at baseline and exhibited a decline in kidney function post-operatively. Conversely, in 4 of the 5 obstructed patients, the renal function remained stable post-operatively indicating the resected kidney was non-functional due to complete ureteric obstruction (**Fig.S1A**), while in the remaining case there was a decline in kidney function post-operatively suggesting incomplete ureteric obstruction. In both the controls and obstructed specimens there was mild glomerulosclerosis, commensurate with the cohort age (mean 71 years); however, the obstructed samples exhibited significantly greater tubulointerstitial fibrosis and atrophy (**Table S1, Fig.S1B**).

To mitigate against batch effects, we isolated nuclei from snap-frozen wedge biopsies from each patient and constructed 6 pools, each comprising nuclei from 3–5 patients, with most patients represented in more than one pool (**Fig.S2A,B, Table S2**). We performed combined snRNA-seq and ATAC-seq using the Chromium Multiome Assay (**Fig. 1A, S2B-G**) and pooled cells were then assigned to each donor by computational SNP-based deconvolution confirmed by genotyping a surplus piece of tissue (**Fig.S2A**). There was high (98.68%) agreement between donor assignment based on SNP deconvolution by snRNA or ATAC-seq, and between inferred genotypes assigned to the same donor across different pools, validating the accuracy of our approach (**Fig.S2H-K**). The pooling approach facilitated detection of nuclei doublets, identified by the presence of SNPs characteristic of different donors. Indeed, not all doublets would have been excluded by standard methods such as applying a maximum threshold to the gene or peak count (**Fig. S2G**).

Following quality control (**Fig.S3**), we analyzed joint transcriptomes and chromatin state of 46,957 nuclei, identifying 59 discrete clusters (**Fig. 1B, Table S3**) which were annotated using the differentially expressed genes and linked differentially accessible chromatin regions in each cluster (**Fig.S4, S5, Table S4**). To assess how our clustering compared with existing datasets, we projected snRNA-seq data from the Kidney Precision Medicine Project (KPMP) Atlas onto our dataset, confirming high correlation

between cells assigned to comparable clusters across both datasets (**Fig.S6A-B**). Within our tubular cell clusters, cells from obstructed kidneys segregated from controls (**Fig. 1C**) and were characterized by a core injury gene signature including *PROM1*, *DCDC2*, *SPP1*, *ITGB6 and ITGB8* (**Fig. 1D)** consistent with that of cells annotated as 'adaptive' in the KPMP atlas(4). Hence following ureteric obstruction, injured tubules adopt an injury response that is conserved across the ischaemic, diabetic and hypertensive nephropathies assessed in the KPMP atlas. The obstructed kidneys contained proportionally fewer healthy tubular cells than unobstructed controls and were enriched for this injured tubular cell phenotype (**Fig. 1E**). In addition, there were fewer peritubular capillary endothelial cells in obstructed kidneys consistent with rarefaction of the peri-tubular capillary network (**Fig. 1E**). Conversely, in obstructed kidneys there was a higher proportion of myeloid cells, T-cells and B-cells and the $CD56^{Bright}$ subset of natural killer cells, in keeping with prior studies implicating this subset in progression of CKD(14). In addition, there was evidence of a phenotypic switch in stromal cells with fewer resident fibroblasts and macrophages and an increase in myofibroblasts and activated macrophages (**Fig. 1E**).

Identification of an inflammatory subset of tubular cells specific to injured kidneys

To explore the phenotypes of the tubular cell clusters in more detail, we segregated and re-clustered each tubular cell phenotype. Proximal tubular (PT) cells comprised healthy cells, corresponding to the S1-3 segments; injured cells, which in addition to the core injury gene set above also expressed *HAVCR1* and *VCAM1* as observed previously(10); and an additional subset that was almost exclusively detected in obstructed kidneys and which was characterized by expression of chemokines, pro-fibrotic factors and adhesion molecules, hence we designated as inflammatory PT cells (**Fig. 2A-C, Table S5**). When projected onto the KPMP Atlas, both the injured and inflammatory cells mapped to the same cluster denoted as 'adaptive' PT cells (**Fig.S6C-F**), and hence the inflammatory PT cells represent a novel subset of the adaptive phenotype that has not previously been characterized in detail.

Trajectory analysis demonstrated that following injury PT cells down-regulated genes encoding key tubular cell functions such as ion transport and metabolism (**Figs. 2D-F, Table S6**) and expressed the core tubular injury signature (**Figs. 2C,E**). RNA velocity suggested that cells in the initial stages of the injury trajectory may retain the potential to revert towards a healthy phenotype, while cells in a more advanced injury state transition towards an inflammatory phenotype (**Fig. 2A**). To determine whether the trajectory analysis correlated with real-time dynamic modelling in pre-clinical models, we compared the human PT cell phenotypes with scRNA-seq datasets from two murine models – ischaemic reperfusion injury (IRI) (9) and reversible unilateral ureteric obstruction (R-UUO)(15), observing conserved injury and inflammatory PT cell phenotypes across species (**Fig.S7A-C**). Following injury, there was a rapid increase in the proportion of injured tubular cells, with the inflammatory cell phenotype appearing later (**Fig. 2G, S7D**). In both models, the proportion of injured PT cells fell progressively during the recovery period while the proportion of healthy PT cells increased, suggesting reversion of injured to healthy tubules. In contrast, the inflammatory PT cells persisted at a higher level for longer after injury.

Taken together, the human and murine data indicate a 'point-of-no-return' beyond which injured PT cells adopt an inflammatory phenotype and are incapable of reverting towards a healthy phenotype, thereby persisting despite cessation of injury. These features are present in senescent renal epithelia(16) and we noted that the inflammatory PT cells also expressed senescent marker genes including *CDKN1A* and *HDAC9*(17), which has recently been shown to mediate renal fibrosis(18). To assess this further, we induced senescence by irradiating renal proximal tubular epithelial cells (RPTECs) and performed bulk RNA-seq to identify genes that were > 2-fold up-regulated in senescent compared with healthy RPTECs (**Fig. 2H,I, S8**). The signature score of these genes was markedly higher in the inflammatory PT cells (**Fig. 2H**), consistent with consistent with the transcriptional activation of a senescence program.

Finally, we sought to determine whether similar inflammatory phenotypes were observed for other tubular epithelial cells in obstructed kidneys. We observed a subset of injured cells in the thick ascending limb of the Loop of Henle that expressed multiple chemokine, cytokine and adhesion genes comparable to the inflammatory PT cells (**Fig. 2J, S9A-D, Tables S5,S7**). Other injured tubular cell types, including connecting tubules and principal cells expressed a modest number of inflammatory transcripts, though this was less marked in distal convoluted tubular cells (**Fig. 2J, S9E-G, S10, Table S8**).

Inflammatory tubular cells co-localise with monocytes and myofibroblasts in the fibrotic niche

A key limitation of snRNA-seq is the loss of spatial orientation, therefore to localize our cellular phenotypes in patients with kidney disease we performed high-plex spatial molecular single-cell imaging on the CosMx platform, utilizing renal biopsy tissue from patients with minimal change disease (n = 3) or IgA nephropathy (n = 6) and nephrectomy specimens from patients with inflammatory fibrosis due to recurrent pyelonephritis (n = 4, **Fig. 3A, Table S9**). ~250,000 cells were allocated into 25 clusters which were annotated using differentially expressed genes (**Fig. 3B, Fig. S11,12, Tables S10,11**) to generate a spatial atlas of the nephrectomy and biopsy samples (**Fig. 3C**). We found discrete injured and inflammatory phenotypes for most tubular cell types, which exhibited similar differentially expressed genes as in the snRNA/ATAC-seq analysis (**Fig. 3D, Table S12**). We performed neighbourhood network analysis to algorithmically compartmentalise the tissue into niches enriched with specific cell types (**Fig. 3E, S13**). In the fibrotic niche there was over-representation of myofibroblasts, leucocytes and inflammatory tubular epithelial cells and related transcripts and the proportion of the tissue occupied by fibrotic niches correlated with the extent of tubulointerstitial fibrosis and inversely with estimated glomerular filtration rate (eGFR, **Fig. 3E-G, S13,14A,B**). Conversely, in the niche enriched with injured tubular epithelial cells there was no increase in leucocytes and myofibroblasts.

To identify cell phenotypes that were preferentially co-located, for each cell type we determined the log-fold enrichment ratio as the observed frequency of all other cell types located within a 25 μm radius compared to that expected in a random topographical distribution. We observed co-location of glomerular cell types, juxtaposition of endothelial and smooth muscle cells and that cell types from the same region

of the nephron were adjacent (e.g. intercalated cells and principal cells), thus validating our approach (**Fig. 3H, Table S13**). We observed monocyte recruitment and enrichment of myofibroblasts, macrophages and dendritic cells (**Fig. 3H,I**) adjacent to the inflammatory but not injured PT cells. Similarly, myofibroblasts, dendritic cells and monocytes, but not tissue resident macrophages were enriched adjacent to inflammatory but not the injured loop of Henle/distal convoluted tubular cells. In addition, we assessed whether individual gene transcripts co-localised with inflammatory tubular cells. There was a progressive decline in expression of inflammatory transcripts with distance from the center of inflammatory but not healthy or injured tubular cells (**Fig. 3J**). Transcripts characteristic of myofibroblasts, dendritic cells, monocytes, and lymphocytes but not tissue macrophages were enriched specifically adjacent to inflammatory tubular cells (**Fig. 3J, S14C**). Crucially, of all the tubular cell phenotypes, the inflammatory PT cells were most strongly correlated with fibrosis and inversely correlated with eGFR (**Fig. 3K, Fig. S14D-F, Table S14**). Hence, our data implicate inflammatory PT and LoH/distal tubular cells in recruitment of monocytes, dendritic cells and lymphocytes and activation of myofibroblasts to promote inflammation and fibrosis in patients with CKD.

Inter-cellular signaling pathways between inflammatory tubules, leucocytes and myofibroblasts

We next sought to identify signalling pathways by which inflammatory tubular cells communicate with adjacent cell types. Ligand-receptor analysis of the snRNA-seq data demonstrated that inflammatory PT cells expressed chemokines, adhesion molecules and complement whereas the corresponding receptors were found primarily on monocytes, dendritic cells, macrophages and lymphocytes (**Fig. 4A, Table S15**). These leucocytes were located adjacent to the inflammatory tubules (**Fig. 4B**), implicating these ligand-receptor pairs in leucocyte recruitment consistent with previous studies(19).

In trajectory analysis, recruited monocytes differentiated towards 2 phenotypes, activated macrophages or conventional type 2 dendritic cells (cDC2) (**Fig. 4C,D**). *TNF* and *TNFSF14* ligand-receptor pairs were expressed in inflammatory PT cells and monocytes/activated macrophages, respectively (**Fig. 4A**), implicating TNF signaling from inflammatory PT cells to skew monocytes towards an inflammatory macrophage phenotype as observed previously(20, 21, 22). Conversely, we identified *CCL20-CCR6* signaling from inflammatory PT cells to cDC2 cells and lymphocytes (**Fig. 4A**), consistent with prior work implicating CCL20-CCR6 binding in DC recruitment to inflamed epithelia(23, 24).

Inflammatory PT cells expressed multiple ligands for fibroblast PDGF, TGF and FGF receptors (**Fig. 4A**), which have been implicated in fibroblast proliferation and activation to myofibroblasts(25, 26, 27). Indeed, in our spatial analysis, multiple PDGF transcripts were identified in inflammatory tubules adjacent to fibroblasts expressing *PDGFRA* (**Fig. 4E**). In addition, *FGFR1* was detected in activated macrophages, which is of interest as FGF receptor inhibition ameliorates LPS-induced macrophage activation, inflammation and fibrosis in the kidney(28, 29).

Inflammatory PT cells may also receive signaling from stromal cells as they expressed receptors for FAS and TWEAK (**Fig. 4F**), while the ligands were expressed by NK cells (FAS) and monocytes (TWEAK), with both these cell types being implicated in kidney disease(19, 30). While these are classically considered pro-apoptotic factors, they can also promote inflammation via NF-κβ signaling(31, 32) and hence enable bi-directional cross talk between leucocytes and inflammatory PT cells to mediate renal inflammation. Our data also implicate TNF superfamily members in autocrine and paracrine tubular signaling as *TNF* and *TNFSF10* (TRAIL) and their receptors *TNFRSF1A* and *TNFRSF10A,B,D* are all highly expressed in inflammatory PT cells (**Fig. 4E**). This may initiate a feed-forward loop to propagate the inflammatory phenotype along the nephron resulting in clustering of inflammatory tubular cells within the same nephron profile (**Fig. 3E-H, 4E**).

In summary, our integrated ligand-receptor and spatial analysis identifies multiple mechanisms by which inflammatory tubular cells signal to adjacent cells to orchestrate the fibrotic niche.

Loss of HNF4A network may promote irreversible loss of PT cell identity

We next sought to characterize the epigenetic programming that mediates the transition from healthy tubular cells to this inflammatory phenotype. In order to infer transcription factor (TF) dynamics, we leveraged the joint snRNA-seq and ATAC-seq modalities to identify correlations between chromatin accessibility at TF binding sites located within cis-regulatory elements (CREs) and adjacent gene expression and construct gene regulatory networks (**Fig. 5A,B, S15,16, Tables S16-18**).

During transition from health to the inflammatory state, PT cells down-regulated genes integral to tubular cell function such as *CUBN* and *SLC34A1* (**Fig. 2C**); therefore, we asked which TFs govern expression of these down-regulated genes. We linked these genes to their regulatory elements by identifying adjacent CREs with reduced accessibility correlating with lower gene expression. TF binding sites present within these CREs included those previously implicated in maintaining PT cell health, such as HNF1-α and HNF4-α (**Fig. 5C, E, Table S18**)(10, 33). In order to infer how these TFs co-operate to regulate gene expression, we compared their target genes, observing that TFs broadly segregated into two modules (**Fig. 5D)**, with the target gene expression in module A markedly down-regulated in PT cells during the trajectory from health towards an inflammatory phenotype (**Fig. 5F**). All other transcription factors in module A, but not module B, had linked CREs containing HNF4-α motifs (**Fig. 5G**), suggesting that HNF4-α may act as a master-regulator of the network. Hence, loss of HNF4-α activity, as confirmed by progressive reduction of physical occupancy at HNF4-α binding sites during transition from healthy to inflammatory PT cells, (**Fig. 5H**) may also reduce gene expression of its co-TFs in module A.

To understand the dynamics of HNF-α network disruption following kidney injury, we examined the expression of HNF4-α target genes in PT cells in a time course following murine ischaemia-reperfusion injury (**Fig. 5I**). HNF4-α target genes were down-regulated within 4 hours of ischaemic injury and were restored within 48 hours in PT cells that recovered to a healthy phenotype. Conversely, HNF4-α target

genes remained persistently low in inflammatory PT cells through to 6 weeks after injury, suggesting failure of the HNF4-α network to recover.

Finally, we addressed why the HNF4-α network remained persistently disrupted after injury. We predicted that HNF4-α bound CREs adjacent to its own gene locus (**Fig. 5G**) and this was confirmed by previous CUT&RUN analysis showing HNF4-α occupancy at multiple CREs at the *HNF4A* locus in the renal cortex of healthy individuals (**Fig. 5J**)(34). During the trajectory from health towards the inflammatory phenotype there was progressive loss of chromatin accessibility at these HNF4-α binding sites coinciding with loss of *HNF4A* gene expression (**Fig. 5J**). Hence, HNF4-α regulates its own expression in a feed-forward loop, which may be disrupted after depletion of *HNF4A* mRNA and protein in the nucleus after sustained injury. Collectively, these data were consistent with disruption of the HNF4-α network being a major mechanism accounting for loss of PT identify (**Fig. 5K**).

AP-1 promotes transition towards inflammatory tubular cell phenotype

Next, we sought to determine which transcription factors facilitate re-programming of PT cells towards an inflammatory phenotype by activating the injury-related gene expression program (**Fig. 2E**). To this end, we sought TFs for which there was an increased TF score (signature score of target CRE accessibility and target gene expression) combined with increased expression of the TF gene in injured and inflammatory PT cells (**Fig. 6A**). Amongst the most up-regulated TFs were SOX4, KLF6, ELF3 which have previously been implicated in promoting an injured tubular phenotype(7) as well as motifs related to the AP-1 and NF-κβ family of TFs. (**Fig. 6B**). Next, we asked what the regulatory impact each TF has on gene expression, grouping TF families with similar motifs or which bind as heterodimers (e.g. AP-1 family). Amongst the enriched TFs in the modules gaining accessibility mid-to-late in the trajectory were AP-1 family TFs with motif binding sites located in ~ 20% of accessible CREs and NF-κβ1 which was more strongly enriched in late CREs (**Fig. 6C**). Supporting this is a progressively increased TF score in the trajectory from healthy to inflammatory PTs (**Fig. 6D**) and the highest level of motif occupancy in inflammatory PT as determined by footprinting analysis (**Fig. 6E)**. To confirm these *in silico* findings experimentally, we stimulated primary renal epithelial cells (RPTECs) with TNF, which is central to CKD progression(20, 35) and which activates both the AP-1 and NF-κβ pathways(36, 37) and performed a CUT&RUN assay to determine binding sites of NF-κβ1 and JUN (**Fig. 6F, S17A-C**). We overlapped CUT&RUN peaks with the open chromatin regions in healthy and obstructed kidneys, observing an increase of accessibility during transition from healthy to inflammatory PT cells in a majority of the overlapping regions (**Fig. 6G**). Furthermore, we found AP-1 and NF-κβ enriched in the respective peak sets confirming increased accessibility at binding sites and physical occupancy of both TFs in inflamed PT cells (**Fig.S17D-F**).

Representative plots of chromatin accessibility at genomic loci of archetypal inflammatory PT cell markers demonstrate highly gene-specific activation patterns (**Fig. 6H**). For example, accessibility at multiple NF-κβ1 binding sites at the *CCL2* locus is gained by both injured and inflammatory PT cells,

while accessibility at the *TNF* locus is relatively restricted to inflammatory PT cells at a single NF-κβ1 binding site. AP-1 has previously been implicated with senescence induction(38) and we observed JUN directly binding to an enhancer element in the *CDKN1A* gene 20kb upstream of the transcription start site (TSS) which is predominantly accessible in inflammatory PT cells, while the remaining regulatory elements show no changes in accessibility (**Fig. 6H**). Similarly, four CREs with highest accessibility in inflammatory PTs are observed in proximity to the *HDAC9* locus, two of which bind JUN. Looking at the gene regulatory network more broadly, we predict both distinct and overlapping target genes for NF-κβ1 and JUN (**Fig. 6I**). For example, in addition to key senescence genes, AP-1 is predicted to activate injury markers such as *HAVCR1* and *DCDC2*. In contrast, NF-κβ1 is essential to activate many pro-inflammatory genes including a multiple chemokines including *CCL2* and *CXCL1*, suggesting both AP-1 and NF-κβ1 have intertwined but distinct regulatory roles in promoting the inflammatory phenotype.

Inhibition of AP-1 activity ameliorates fibrosis following ischaemia-reperfusion injury in mice

Given we identified multiple parallel intercellular signaling pathways between inflammatory tubular epithelial cells, leucocytes and myofibroblasts (**Fig. 4**), we postulated that redundancy may reduce the efficacy of targeting a single ligand-receptor pathway. Therefore, we assessed whether inhibiting core transcriptional regulators of the inflammatory tubular phenotype, such as AP-1, could reduce the severity of renal inflammation and fibrosis. In a murine model of AKI to CKD transition, we targeted AP-1 using T5224, which inhibits binding of c-Fos and c-Jun to DNA thereby preventing transcription factor activity(39). We administered 10 mg/kg of T5224 or DMSO vehicle by daily gavage for 2 weeks beginning 3 days following unilateral ischaemia-reperfusion injury (IRI) or sham surgery (**Fig. 7A**). Administration of T5224 ameliorated the reduction in renal mass following IRI (**Fig. 7B**), reduced the severity of inflammation and fibrosis (**Fig. 7C,D)** and the renal cortical expression of genes implicated in renal injury (*Havcr1*), macrophage accumulation (*Cd68*) and extracellular matrix production (*Col1a1*, **Fig. 6E**). T5224 also reduced expression of *Arg1*, which we previously identified as a marker of infiltrating monocytes transitioning to a pro-inflammatory phenotype(15). There was also a trend for reduced *Tgfb1* expression with T5224 administration, consistent with prior *in vitro* studies in which T5224 inhibited *TGFB1* expression in RPTECs(40). The expression of genes characteristic of inflammatory tubular cells including *Icam1*, *Vcam1* and *Pdgfa* was also reduced in animals treated with T5224 highlighting potential mechanisms for the reduced leucocyte and myofibroblast accumulation (**Fig. 6E).** This is in keeping with CUT&RUN analysis in TNF-treated RPTECs, which demonstrated that JUN bound to CREs adjacent to the *HAVCR1*, *VCAM1*, *ICAM1* and *PDGFA* genes (**Fig. 6F**). Importantly, these CREs were located in chromatin regions that were progressively more accessible in PT cells during transition from health to the inflammatory phenotype in human kidney disease. Taken together, the data confirm that inhibition of AP-1 activity using T5224 prevented transition from acute ischaemic kidney injury to renal fibrosis.

Senolytic therapies promote depletion of the inflammatory tubular cell phenotype

We had previously identified that inflammatory PT cells express *CKDN1A* and a gene signature consistent with activation of a senescence program and persisted despite cessation of injury, suggesting that they had adopted a senescent phenotype (**Fig. 2I-J**). JUN bound to AP-1 in inflamed RPTECs (**Fig. 6H**) and AP-1 inhibition reduced *Ckdn1a* expression (**Fig. 7E**), suggesting that AP-1 may promote transition to this senescent phenotype. However, we also wished to assess whether depletion of cells that had already become senescent was associated with reduction in the inflammatory tubular cell phenotype. We determined that expression of the anti-apoptotic gene *BCL2* was increased in inflammatory tubular cells, which may render them resistant to apoptosis and enable them to persist despite cessation of injury (**Fig. 7G**). Hence, we assessed whether depletion of senescent cells using the BCL2/w/xL inhibitor ABT-263, a senolytic compound known to promote renal repair(41, 42), also reduced expression of genes consistent with an inflammatory tubular cell phenotype (**Fig. 7H**). In the reversible ureteric obstruction model, archetypal inflammatory tubular cell genes (**Fig. 2C, J**) were up-regulated in the renal cortex one week following reversal of obstruction (**Fig. 7I**). In mice that received vehicle, these genes remained up-regulated 5 weeks following reversal of obstruction, indicating persistence of the inflammatory tubular phenotype, whereas administration of ABT-263 markedly reduced their expression (**Fig. 7I**) and inhibited development of renal fibrosis(42). Hence, senolytic therapies may represent an alternative approach to deplete inflammatory tubular cells that have developed following kidney injury.

Discussion

In the current studies, we identified an inflammatory, pro-fibrotic phenotype common to both PT and loop of Henle cells, that represents a small subset of the 'adaptive' tubular cells identified in the KPMP and is consistent with *Vcam1*$^{+}$ *Ccl2*$^{+}$ cells previously observed following ischaemic renal injury in mice(12, 13). Our study is among the first to apply high-plex spatial molecular imaging using the CosMx platform to human kidney disease and this determined that it is specifically the inflammatory tubular epithelial cells that localize to the fibrotic niche and correlate with the severity of fibrosis. Our ligand-receptor analysis identified multiple signaling pathways by which the inflammatory tubular cells mediate leucocyte chemotaxis and fibroblast-to-myofibroblast activation and our spatial analysis confirmed that the ligands and receptors are in close proximity to enable effective signaling. Inhibition of several of these ligand-receptor pairs is efficacious in murine kidney disease(13, 29, 32, 43, 44) and our data suggest that they offer new therapeutic opportunities in human disease.

Our study represents the first use of simultaneous single-cell and ATAC-sequencing in diseased human kidney. We leveraged these data to characterise the key transcription factors that promote tubular cell integrity and identified AP-1 as a key mediator of transition to the inflammatory state. While AP-1 inhibition has previously been shown to prevent endotoxin-injured kidney injury in pre-clinical models(45), our study is the first to suggest that it may prevent transition from AKI to chronic inflammation and fibrosis. The inflammatory tubular cells expressed *CKDN1A* and other genes consistent with a senescence phenotype. Furthermore, depletion of senescent cells after kidney injury reduced expression of the inflammatory tubular cell signature, suggesting that senolytic therapies may be an alternative treatment

for patients where the inflammatory tubular phenotype is well established, which is important as many patients present with advanced CKD.

In summary, we have generated a comprehensive multiomic atlas of the cellular heterogeneity in human kidney disease and identified druggable signaling pathways by which an inflammatory tubular cell subset may mediate tubulointerstitial inflammation and fibrosis. Furthermore, we confirmed that inhibition of AP-1 activity and senolytic therapies reduce progression of inflammation and fibrosis following experimental AKI and warrant further research as a treatment to prevent AKI to CKD transition and progression of CKD in patients with kidney disease.

Materials and methods

Patient recruitment

For the multiome analysis, we recruited patients undergoing radical nephro-ureterectomy or radical nephrectomy at the Western General Hospital, Edinburgh. We recruited patients with renal cell carcinoma (n = 6) or oncocytoma (n = 1) and normal renal function (eGFR > 75ml/min/1.73m^2) as controls and in addition we recruited 5 patients with transitional cell carcinoma of the ureter leading to ureteric obstruction as determined by the presence of hydronephrosis on CT imaging.

For spatial transcriptomics, we utilised surplus formalin-fixed, paraffin-embedded tissue from kidney biopsies from patients with IgA nephropathy (n = 6) or minimal change disease (n = 3) that were performed for clinical purposes. In addition, we employed 4 wedge biopsies from nephrectomy specimens removed due to advanced pyelonephritis.

Renal function at the time of nephrectomy or biopsy and during follow-up was estimated from serum creatinine results provided by the clinical lab by applying the CKD-EPI equation without correction for race(46). Urinary protein:creatinine results were provided by the clinical lab.

Use of both the nephrectomy and biopsy samples was approved by the steering committee of the National Research Scotland Lothian Bioresource (REC 20/ES/0061, studies SR1651,SR1887).

Animal experiment

Male C57BL/6JCrl mice were purchased from Charles River Laboratories and housed in a pathogen-free environment at the University of Edinburgh animal facility. All procedures had prior approval of the Animal Ethics Committee, University of Edinburgh and were conducted in accordance with the United Kingdom Animals Scientific Procedures Act 1986 and the ARRIVE guidelines.

IRI surgery was performed as previously described(47). Briefly, a posterior flank incision was made and the left renal pedicle identified and clipped using an atraumatic clamp for 15 minutes. During the ischemic period, body temperature was maintained at 37°C using a heating blanket with homeostatic control (Harvard Apparatus) via a rectal temperature probe. The clamp was then removed, the peritoneum

closed with 5/0 suture, and the skin closed with clips. In sham surgery, the mice underwent flank incision, however the renal pedicle was not clamped.

T-5224 was reconstituted in dimethyl sulfoxide (DMSO) and diluted 1:20 in sunflower oil. 3 days post-IRI mice in the respective group were given 10mg/kg T-5224 (APExBIO, B4664) by gavage daily for 12 days. Animals were maintained for 15 days post-IRI before tissue harvest. IRI or respective control kidneys were weighed, and one third was fixed for 24 hours in 10% neutral buffered formalin with the remainder snap-frozen in dry ice.

Histological analysis

After fixation and paraffin embedding, human and mouse tissues were sectioned at 4μm depth from FFPE embedded tissue blocks and adhered to Superfrost Plus slides (VWR, 631 – 0108). Sections were deparaffinized two times 5 minutes in xylene and rehydrated in 100%, 90% and 70% ethanol for 2 minutes each and placed in distilled water. PSR staining was performed using the Picro Sirius Red Stain Kit (Abcam, ab150681) following manufacturer guidelines. After rehydration of tissue sections, the tissue was covered picro sirius red solution for 60 min. Slides were rinsed in acetic acid solution and dehydrated in absolute ethanol before mounting. For macrophage quantification, murine kidney sections were incubated with the avidin/biotin blocking kit (Vector Laboratories, SP2001) and blocked with serum-free protein block (Dako, X0909) before incubation overnight at 4°C with IBA1 antibody (Invitrogen, PA5-27436) diluted 1:250 in antibody diluent (Dako UK Ltd, S202230). After washing, the sections were incubated with polyclonal goat anti-rabbit biotinylated secondary antibody (DAKO, E0432, 1:300 dilution) for 30 minutes at room temperature. Vectastain RTU ABC Reagent (Vector Laboratories, PK7100) was then applied, followed by incubation with the DAB1 Substrate Chromogen System (Dako, K3468), and then counterstaining with hematoxylin before dehydration and mounting with Pertex mounting medium (Histolab Products AB, 3808707E).

PSR stained sections were scanned with a Zeiss AxioScan Z1 slide scanner (Carl Zeiss). Images of human kidneys were reviewed by a consultant pathologist who confirmed the absence of tumour infiltration and provided a semi-quantitative assessment of glomerulosclerosis and tubulointerstitial inflammation and fibrosis. For mouse tissue, PSR staining of renal cortex was quantified in QuPath (v0.4.2). A threshold classifier was used to calculate cortex area with intensities above the thresholds denoting collagen fibres and this was assessed as a percentage of total tissue area. The percentage of total renal cortex that stained with IBA1 was quantified using a minimum threshold for positive staining on ImageJ analysis using a minimum of 5 x20-magnification images of renal cortex.

qPCR

Total RNA was extracted from kidney tissue using the miRNeasy Mini kit (Qiagen, 217004), following the kit instructions. cDNA for quantitative PCR was synthesised from the extracted RNA using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, 4368814). Quantitative PCR was performed using TaqMan Universal Master Mix II (Applied Biosystems, 4440040) and TaqMan Gene

Expression Assay-specific primers (**Table S19**) and normalised to Peptidylprolyl isomerase A (*Ppia*) expression.

DNA extraction and genotyping

DNA was extracted from approximately 25 mg of nephrectomy tissue using the DNeasy Blood & Tissue Kit (Qiagen, 69504) according to the manufacturer instructions. Briefly, tissue was cut into small pieces and incubated with proteinase K for 3 hours at 56°C. Genomic DNA was purified using a DNeasy Mini spin column, washed and eluted with TE buffer. 200 ng of genomic DNA per sample was used to genotype samples using an Infinium Global Screening Array (Illumina, 20030770) according to manufacturer instructions. Raw genotyping data was converted to variant call format (VCF) using GenomeStudio (Illumina, v2.0.3).

Single-nucleus multiome RNA and ATAC sequencing

Nuclei isolation, library preparation and sequencing

For single-nuclei multiome studies wedge biopsies were taken immediately after nephrectomy from the pole opposite the tumour in those with intra-renal tumours and samples snap-frozen in dry ice and a separate piece was fixed for 24 hours in 10% neutral buffered formalin before paraffin embedding for histological analysis. Nuclei were isolated from snap-frozen nephrectomy samples using an adapted 10x Genomics Demonstrated protocol (CG000375, Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing). Using a scalpel, the tissue was cut into approximately 25 mg of rice-grain sized pieces while avoiding thawing and all further steps were performed at 4°C. Initially, 150 µl lysis buffer containing 10 mM Tris-HCl (Sigma-Aldrich, T2194), 10 mM NaCl (Sigma-Aldrich, 59222C), 3 mM $MgCl_2$ (Sigma-Aldrich, M1028), 0.05% NP40 Substitute (Sigma-Aldrich, 74385), 1 mM DTT (Sigma-Aldrich, 646563) and 1 U/µl Protector RNase inhibitor (Sigma-Aldrich, 3335402001) in Nuclease-free Water (Invitrogen, 10429224) was added and tissue was homogenised using a sterile pestle (Fisher Scientific, 12-141-368). Once homogenised another 850 µl lysis buffer was added, and tubes were placed on ice for 5 minutes. The suspension was strained using a 40 µl filter (Cambridge Bioscience, 43-10040-60), centrifuged (500 g, 5 min, 4°C) and 1 ml 2% BSA (Miltenyi Biotec,130-091-376) in phosphate-buffered saline (PBS) (ThermoFisher, 10010023) was added without disrupting the pellet. After 5 minutes nuclei were resuspended, centrifuged (500 g, 5 min, 4°C) and finally resuspended in 300 µl of 2% BSA in PBS.

Nuclei were stained with 10 µl of 7-AAD viability staining solution (Thermo Fisher Scientific, 00-6993-50) and sorted on a BD FACSAria fusion (Becton Dickinson) cell sorter fitted with a 100 ml nozzle into a tube prepared with 200 µl of 10% BSA in PBS supplemented with 50 µl RNase inhibitor for a post-sorting volume of 2 ml. Between 3 and 5 samples from individual donors were pooled at pre-defined proportions based on counts of 7-AAD + nuclei (**Fig. S2B)** and a total of 500,000 nuclei were sorted per pool (**Fig. S2A**).

Sorted nuclei were centrifuged (500 g, 5 min, 4°C) and resuspended in 100 µl lysis buffer with 0.01% NP40 Substitute, 0.01% Tween-20 (Bio-Rad, 1662404), 0.001% Digitonin (Thermo Fisher Scientific, BN2006) and 2% BSA. After 2 minutes wash buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM $MgCl_2$, 2% BSA, 1 mM DTT, 1 U/µl RNase inhibitor, 0.1% Tween-20) was added and nuclei were centrifuged (500 g, 5 min, 4°C) and resuspended in 20 µl diluted nuclei buffer (20X Nuclei Buffer, 1 mM DTT, 1 U/µl RNase inhibitor in nuclease-free water).

Nuclei suspensions were counted using a LUNA-FX7 cell counter (Logos Biosystems) and processed immediately using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit (v1.0, 10x Genomics) according to manufacturer protocol (CG000338, User Guide Chromium Next GEM Single Cell Multiome ATAC + Gene Expression). Briefly, per library between 20,000 to 40,000 nuclei were input into the transposition reaction before partitioning into gel beads. After reverse transcription and pre-amplification, aliquots of the pool of transposed DNA and cDNA were used to generate separate ATAC and GEX libraries.

After quantification GEX and ATAC libraries were pooled at molarity ratio of 40:60 respectively and sequenced using an Illumina NovaSeq6000 flow cell (with the read lengths R1:151bp I1:10bp I2:24bp R2:151bp) for a target yield of approximately 50,000 reads/cell and 75,000 reads/cell for gene expression and ATAC libraries respectively. In total 6 libraries comprising a total of 7 individual control and 5 UUO individual samples were prepared and analysed.

Sequencing data processing and donor demultiplexing

BCL files from Illumina sequencing runs were demultiplexed for GEX and ATAC samples using cellranger-arc mkfastq (v2.0.2, 10x Genomics). The base mask Y28n*,I10n*,I10n*,Y151 and –filter-dual-index or Y100n*,I8n*,Y24,Y100n* and –filter-single-index was used to generate GEX or ATAC FASTQ files respectively. Reads were aligned to the precomputed GRCh38 (hg38) reference genome provided by 10x Genomics, and quantification as well as joint cell calling was performed using cellranger-arc count (v2.0.2).

Due to donor pooling, each library included cell barcodes from multiple different samples. To allocate barcodes to their original donor the cellsnp-lite (v1.2.2)(48) and vireo (v0.2.3)(49) packages were used. Barcodes passing cellranger-arc filters were genotyped independently for GEX and ATAC libraries by searching for SNPs present in sequencing reads using cellsnp-lite with the options –minMAF 0.1 (and –UMItag None for ATAC files) using a reference SNP list of 7.4 million SNPs from the 1000 Genomes Project, filtering for a minor allele frequency > 5% (provided by cellsnp-lite). To infer the original donor of each nuclei barcode using genetic variants that segregate between samples, a Bayesian model implemented by vireo was used (with option –N equal to the number of donors in the library). As donors were predicted independently for GEX and ATAC libraries, conflicts in donor assignments arising in a minority of nuclei barcodes were resolved using the following logic: i) barcodes assigned as doublet in either modality or unassigned in both modalities were removed (18.36% of total); ii) barcodes which were unassigned in one modality but not the other were assigned to the donor predicted by the successful

modality (1.55% of total); iii) barcodes with conflicting donor assignments were removed (98.68% matching, 1.32% conflicting; % of remaining after previous filters) (**Fig. S2H**).

To link predicted donors with physical samples two parallel approaches were used. Firstly, using the original tissue, genomic DNA was extracted and genotyped using an Illumina an Infinium Global Screening Array as described above. VCF files containing the predicted genotype of each donor-library combination were loaded in R (v4.3.1) and subset to SNPs commonly detected in all libraries. To intersect the predicted genotypes with the SNP array, both sets of SNPs were further reduced to a common set and a mean distance matrix was calculated based on the Hamming distance at each genomic location (with 0 being a homozygous reference allele, 2 a homozygous alternative allele and 1 being a heterozygous allele) (**Fig. S2J**). Secondly, a subset of samples was pooled in multiple libraries creating a distinctive pooling pattern which can be used to identify each sample (i.e. the same sample pooled in two different libraries is expected to have the same genotype, as predicted by vireo) (**Fig S2I**). VCF files were processed as previously and the mean Hamming distance between each donor-library pair was calculated, showing a lower distance between identical sample used in multiple libraries (**Fig. S2K**).

Joint dimensionality reduction, clustering and cell annotation

While SNP-based donor demultiplexing allows identification of doublets with distinct genotypes, doublets of nuclei from the same donor remain undetected. Therefore, we used an iterative approach to identify undetected doublets, initially retaining SNP doublets to aid identification. Seurat (v4.4.0)(50) and Signac (v1.11.0)(51) were used to load GEX data using Read10X_h5() and a Seurat object was created using the CreateSeuratObject function without filters. The data was normalised using SCTransform with options vst.flavor = "v2" and principal components were computed using RunPCA with npcs = 100. Different dimensionality reduction and clustering setting were used to compute different representations of the data with RunUMAP, FindNeighbors (with dims between 1 and 80) and FindClusters (resolution between 0.5 and 4). At each iteration presumed doublet subclusters were marked based on the criteria: i) enrichment of doublets identified based on SNP demultiplexing (presence of SNPs from different genetic backgrounds); ii) increased number of unique molecular identifiers (UMIs) or unique genes per barcode; iii) co-expression of multiple markers from distinct cell lineages (e.g. *PTPRC*, *FLT1*, *MME*, *SLC12A1*). Inversely, all barcodes labelled as doublet were separately clustered and previously misclassified doublets were retained.

After initial doublet identification, the full GEX and ATAC datasets were loaded, and doublets or barcodes not assigned to a donor were removed. Further, any gene expressed by less than 20 nuclei barcodes was removed. Additionally, barcodes with i) unique molecular identifiers (UMIs) < 300 or > 12,000; ii) unique genes > 4,000; iii) mitochondrial gene content > 10%; iv) number of ATAC fragments < 700 or > 100,000; v) number of unique peaks < 300 or > 40,000; vi) a nucleosomal signal score (calculated with NuceosomeSignal) < 2; vii) a TSS enrichment score (calculated with TSSEnrichment) < 2 were removed (**Fig. S2D-G, S3A,D**).

Next, dimensionality reductions were independently computed based on GEX and ATAC datasets. In both cases datasets were split by the donor sex using SplitObject. Data was normalised using SCTransform with option vst.flavor = "v2" as previously and integration anchors were computed using FindIntegrationAnchors with the top 3,000 shared variable features. The data was integrated using IntegrateData with dims = 1:80 and the UMAP embedding was computed with dims = 1:80. Similarly, the ATAC data was processed using FindTopFeatures, RunTFIDF, RunSVD with n = 100 before splitting by donor sex, computing anchors using all peaks, and integrating data with IntegrateData with dims = 1:80 and RunUMAP with dims = 2:80). Finally, the Seurat weighted nearest neighbor (WNN) analysis was used to compute a joint dimensionality reduction using FindMultiModalNeighbors with options dims.list = list(1:100, 2:100) and k.nn = 60) in the PCA and LSI embeddings of the GEX and ATAC datasets respectively. The joint UMAP was computed using the WNN graph with RunUMAP option nn.name="weighted.nn" and barcodes were clustered using FindClusters with options resolution = 4 and graph.name = "wsnn".

In order to refine the peak set and detect peaks present in small cell populations clusters were tentatively annotated (pending the final annotations) and the Signac CallPeaks wrapper function for MACS2 (v2.2.7.1)(52) was used to call cluster specific peaks. Peak sets for each cluster were further merged and pruned by removed peaks in scaffold and blacklisted regions. Finally, a peak by nuclei barcode count matrix was generated using the FeatureMatrix function with the new peaks regions as features. Following this, ATAC and WNN dimensionality reductions were recomputed as before, and barcodes were re-clustered (**Fig. S3B**). To determine final cluster annotations barcodes were first broadly clustered using the FindClusters function with lower resolution settings. Broad cluster (level 1 annotations) were further iteratively subclustered by re-computing dimensionality reductions and sub-clusters within a larger cluster. Marker genes for each cluster were determined using the FindMarkers function and remaining subclusters co-expressing signatures of distinct cell types were removed as doublets. Each subcluster was manually annotated (level 2 annotations) according to gene markers determined by prior human kidney atlases(4) (**Fig. S6A, Table S4**). We defined injured epithelial clusters based on a proportional enrichment in UUO samples compared to control samples as well increased expression of generic injury genes such as *PROM*, *DCDC2*, *SPP1*, *ITGB6*, *ITGB8* (**Fig. 1C-1E, Table S5,8**). Subclusters of PT and LOH nephron segments were identified as inflammatory cell states based on the expression of injury markers as well expression of chemokines, extracellular matrix remodelling and adhesion factors and markers of cell cycle arrest (**Fig. 2C**). Individual chemokines were identified in other injured nephron segments, however they lacked a consistent signature of multiple inflammatory markers and sufficient separation from injured counterparts (**Fig. 2J**). Agreement between GEX and chromatin profile of markers genes was assessed using the CoveragePlot function in genomic loci of marker genes determined by GEX (**Fig. S4, S5**).

All further basic graphs showing gene expression, chromatin state, and meta data features were generated using Seurat and Signac functions DimPlot, FeaturePlot, DotPlot, VlnPlot and CoveragePlot.

Integration with KPMP snCv3 dataset

The full (as of September 2023) The Kidney Precision Medicine Project (KPMP) snRNA-seq dataset along with associated clinical data were downloaded from the KPMP website (https://atlas.kpmp.org). Integration of the snCv3 dataset data was performed using Seurat with the multiome dataset as reference. Transfer anchors were computed with FindTransferAnchors using dims = 1:50 with the multiome pre-computed GEX principal components as reference. Subsequently, the snCv3 data along with prediction of respective multiome level 1 and level 2 cell annotations were projected on the WNN UMAP reduction using the MapQuery function (**Fig. S6A**). Agreement of original snCv3 annotations multiome annotations were assessed using sankeyNetwork graphs implemented by the networkD3 package (v0.4) (**Fig. S6B, H**). To assess the presence of inflammatory PT cells, the AddModuleScore_UCell function from the UCell package(53) was used to score each cell using a signature derived from inflammatory PT cells in the multiome dataset (**Fig. S6E**) and the predicted level 2 annotations were used to compare expression patterns of inflammatory PT markers in both datasets (**Fig. S6F**).

Differential abundance analysis

Abundance of different cell types in samples was calculated as proportion of total cells or proportion of total epithelia cells (non-Immune/Endothelial/Interstitial cells) to normalise for shifts in overall composition due to immune cell influx in UUO samples. Comparisons between cell populations in control and UUO samples were performed using a Wilcoxon signed-rank test using the R function wilcox.test with the option alternative = "two.sided" (**Fig. 1E, 2B, S9B, F**).

RNA velocity and pseudotime analysis

Count matrices of spliced and unspliced reads were generated from cellranger GEX BAM files for each library using velocyto (v0.17.15)(54). The gene model from the 10X hg38 reference genome along with a mask for repetitive regions downloaded from the UCSC genome browser were passed to the velocyto run command along with BAM files and a whitelist of nuclei barcodes retained in the final analysis. RNA velocity streams were inferred independently for PT, other epithelial and myeloid cell subclusters using scVelo (v0.2.5)(55). Velocyto loom files were imported using the scv.read function and UMAP embeddings and cell annotations were imported to the AnnData object. The data was normalised according to the standard workflow and moments across 50 nearest neighbours in the first 50 principal components were computed. Splice kinetics were then modelled using a likelihood based dynamical models and visualised using pl.velocity_embedding_stream (**Fig. 2A, 4C, S9A, E**).

To infer pseudotime orderings for PT, LOH and myeloid cell subsets we used the slingshot R (v2.8.0) package(56). For PT trajectories we considered healthy PT S1-S3 as start points and PT Inflammatory cells as end point. Similarly, for LOH trajectories cTAL1, cTAL2 and mTAL clusters were considered as starting point and LOH Inflammatory as end point, while ignoring Macula Densa cells. Two trajectories towards cDC2 and Activated Macrophages each starting from CD14 monocytes were specified for myeloid trajectories. The getLineages with relevant start and end clusters was used to construct the minimum spanning tree and getCurves was used to get a smoothed ordering of cells along the

pseudotime trajectory. Gene expression and chromatin accessibility dynamics correlated with the inferred pseudotime values were identified using the tradeseq package, using the fitGAM function to fit regression models to counts and pseudotime values. Normalised count values of correlated genes or peaks were then extracted from the Seurat object and smoothed using the smoothing function (with strength = 0.2) from the modelbased R package (v0.8.6.3). Heatmap plots of smoothed dynamics were then visualised using the pheatmap package (v1.0.12) (**Fig. 2E, 4D, 6B, S9D**).

Gene set analysis

We defined differentially expressed genes between PT Injured/Inflammatory and PT S1-S3 cells using the FindMarkers Seurat function. Up- or downregulated genes with an absolute average log2-fold change > 0.25 and adjusted p-value < 0.05 were forwarded for over-representation analysis implemented by the clusterProfiler R package (v4.10.0)(57). Using the enrichGO function enrichment for GO biological process terms with < 600 genes were tested. To project the most significantly enriched non-redundant GO terms onto the pseudotime trajectory, genes associated with each GO term were extracted and cells were scored using the AddModuleScore_UCell function. A locally estimated scatterplot smoothing curve was fitted to scaled GO scores as implemented by geom_smooth with a span smoothing strength of 0.4 (**Fig. 2F**).

Gene regulatory network analysis

We used the SCENIC + python (v1.0.0) package(58) for gene regulatory network inference and TF analysis. For this, a custom cistarget database overlapping with ATAC peaks found in the dataset was created by generating a FASTQ file containing peak regions using the bedtools getfasta command. The create_cisTarget_databases.py script was then used with the SCENIC + motif collection (provided by SCENIC+) to score motif sequences in each ATAC peak.

Next, the Seurat dataset was subsampled to < = 1000 cells per level 2 annotation and exported to python (v3.8.0). The remaining SCENIC + workflow was followed with minor alterations. Briefly, GEX and ATAC were pre-processed with default settings, chromatin topic modelling was performed using pycisTopic, and 70 topics were selected for the dataset. PycisTarget was then used to find motif enrichments in accessible chromatin regions, enhancer to gene relationships and TF to gene relationships were calculated and the gene regulatory network was constructed using SCENIC + functions.

The derived regulons consisting of TF-CRE-gene links were further pruned by removing regulons with negative CRE to gene correlations and regulons with < 15 target genes. In total we detected filtered regulons for 474 unique TF motifs found in 109147 unique regions linked to 11567 genes. The filtered regulons were exported to R and cells were independently scored by calculating scores for TF gene targets and CRE targets of each TF using the AddModuleScore_UCell function. A combined TF score for each TF in each cell was calculated as mean of the gene target and CRE target scores. For all further analysis of gene regulatory networks in specific cell types (e.g. PT cells) only genes and CRE expressed or accessible in the given cell type were considered to generate cell-type specific regulons. Genes were considered expressed in the cluster if i) for genes with global expression level below expression by 10% of

cells the gene is expressed by at least 2% of cells in a subcluster ii) for genes with a global expression level above 10% it is expressed by an equal or higher proportion of cells than the global expression level iii) it is expressed by more than 20% of cells. Similarly, regions were considered accessible if the same thresholds were met. Other regions and genes were considered not expressed and removed from the analysis. For linkage analysis between TFs and target genes the linkages between TFs, CREs and target genes were derived from SCENIC + regulons (**Fig. 5, 6, Table S18**). Overlap between TF target genes was assessed by calculating the Jaccard index between lists of target genes of individual TFs (**Fig. 5D**). TF score dynamics along pseudotime trajectories were visualised by fitting a locally estimated scatterplot smoothing curve with a smoothing factor of 0.4 using the geom_smooth function (**Fig. 5F**).

Ligand-receptor analysis

Analysis of ligand-receptor interactions was performed using the CellChat (v2.1.0) package(59). We asked how different PT subclusters interact with immune cells and fibroblast, therefore the dataset was subset to these cell types. To infer interaction the CellChat workflow was followed with default setting, except adjusting the truncated mean gene filtering to 0.01 in the computeCommunProb function in order to account for low expression levels observed particularly by chemokine genes. For visualisation purposes ligands and receptors overexpressed in injured or inflammatory PT cells were extracted and plotted using the Seurat DotPlot function for the ligands and receptors identified by CellChat. Interactions were manually connected using the CellChat interaction database as reference, while omitting individual receptors with no expression or non-relevant expression patterns due space constrains (**Fig. 4A, 4F**).

Re-analysis of mouse sc/snRNA-seq datasets

Count matrices from previously published datasets were downloaded from gene expression omnibus under accession numbers GSE140023(15) and GSE139107(9). In both cases the count matrices were loaded in R and cells were subset to clusters identified as PT cells in the original publication. For both datasets Seurat RPCA integration was used to account for batch effects in experimental design. For the R-UUO datasets the data was split by the timepoint of sample collection (Control, UUO2, UUO7, R-UUO) and each subset was independently normalised using SCTransform with the option vst.flavor = "v2" and principal components were computed using RunPCA with npcs = 100. Integration anchors were computed using FindIntegrationAnchors with options reduction = "rpca" and k.anchor = 30 and subsets were integrated using IntegrateData with dims = 1:50. Subsequently the UMAP dimensionality reduction and clusters were computed using RunUMAP, FindNeighbors and FindClusters with the first 50 principal components. Analogously, the IRI data integrated using the replicate information provided in the GEO records.

In both cases cluster markers were computed using the FindAllMarkers function with options min.pct = 0.01 and logfc.threshold = 0.1. Similarly, we found clusters expressing injury markers such as *Havcr1* and *Vcam1* were enriched post-injury. We identified subclusters expressing markers of maladaptive repair(9), showing similar expression to inflammatory PT cells in human kidneys such as *Ccl2*, *Cxcl1*, *Icam1* and *Tgfb2*. Notably, the expression of *Vcam1* and *Dcdc2a* was more restricted to this cluster and not found in

generic injury clusters in both datasets, as opposed to humans where respective *VCAM1* and *DCDC2* genes were more widely expressed (**Fig. S7C**). Due to overall similarities, we followed naming conventions using for the human data. Finally, the proportion of each cell state was calculated as the proportion of total PT cells in each animal model at each timepoint and visualised as time series graph (**Fig. 2G**).

CosMx spatial molecular imaging

Sample processing

5 µm sections were cut from FFPE embedded tissue blocks and adhered to Superfrost Plus slides (VWR, 631 – 0108). Slides were deparaffinized with two 5 min Xylene washes and two 2 min 100% ethanol washes followed by proteinase K (3 µg/ml; Thermo Fisher, AM2546) digestion (40°C for 30 min) and heat induced epitope retrieval at 100°C for 15 min in ER1 buffer (Leica, AR9961). After rinsing with nuclease free water, slides were incubated with 1:1000 diluted fiducials (Bangs Laboratory) in 2X SSCT (2X saline sodium citrate, 0.001% Tween 20) for 5 min at room temperature. After PBS washes, slides were fixed with 10% neutral buffered formalin for 5 min at room temperature. After rinsing with Tris-glycine buffer (0.1 M glycine, 0.1 M Tris-base) and a 5 min wash with PBS, samples were blocked using 100 mM N-succinimidyl acetate (Thermo Fisher, 464750250) in NHS-acetate buffer (0.1 M NaP, 0.1% Tween pH 8) for 15 min at room temperature. Samples were rinsed with 2X saline sodium citrate (SSC) for 5 min and covered with an Adhesive SecureSeal Hybridization Chamber (Grace Bio-Labs).

The 980-plex ISH probe panel(60) (**Suppl Table 20**) was denatured at 95°C for 2 min and placing on ice. The ISH probe mix was prepared using 1nM ISH probes in 1X Buffer R and 0.1 U/µL SUPERaseIn RNase Inhibitor (Thermo Fisher, AM2696). Samples were then incubated with the probe mix at 37°C overnight. On the next day slides were washed with 50% formamide (VWR) in 2X SSC at 37°C for 25 min and rinsed twice with 2X SSC for 2 min at room temperature. Before antibody incubation slides were treated with blocking buffer containing DAPI nuclear stain for 15 min. For visualization of cell morphology antibodies against C298/B2M, PanCK, CD45 and CD3 from the CosMx Segmentation and Supplemental Markers Kit were chosen and incubated at room temperature for 1 hours. Finally, after washing with 2X SSC for 2 min at room temperature custom-made slide covers were attached to the sample slide to form a flow cell.

Image acquisition

RNA readouts were acquired according to previously described protocols(60). Briefly, the assembled flow cells were loaded onto the SMI instrument and washed with reporter buffer. The flow cell was washed and a median of 12 and total of 140 field of views (FOVs) were defined inside which RNA was detected. The imaging procedure was repeated with 16 reporter pools, with each imaging step repeated 8 times to increase RNA detection sensitivity. Each cycle was initiated by flowing 100 µL of the Reporter Pool into the flow cell and incubating for 15 min followed by flushing with 1ml of Reporter Wash Buffer. The wash buffer was replaced with imaging buffer and nine Z-stack images (0.8 µm step size) of each FOV were

acquired. Each cycle was completed with UV cleavage of fluorophores on the reporter probes and washing using Strip Wash buffer.

Following RNA readouts, staining for DAPI and the four morphology antibodies was acquired by capturing nine Z-stack images after washing with Reporter Washing Buffer.

Image processing and cell segmentation

Image processing and feature extraction followed the NanoString pipeline previously described(60) which extracted XYZ locations of individual transcripts. After image registration, feature detection and localisation the Z-stack images the nuclear and cell membrane markers were used define cell boundaries using a machine learning algorithm(60). Cell segmentation boundaries were used to assign transcripts to cells and subcellular compartments from which a count matrix was generated. Pipeline outputs included a cell by gene matrix, cell segmentation polygons, and cell meta data.

Data processing, clustering and dimensionality reduction

Data was initially imported into Seurat using the LoadNanostring() function. The number of transcripts and genes per cell were assessed for regional biases (**Fig. S11A,B**) and cells with < 20 transcripts were excluded from quantitative analysis but not from image plots. Batch effects between biopsy and nephrectomy samples were observed and therefore integrated using the harmony (v0.1) package(61).

For this, each group was independently normalised with SCTransform using vst.flavor = "v2", and principal components were computed with RunPCA with option npcs = 100. RunHarmony was then used to align principal components and the UMAP projection was computed on adjusted principal components. Cell types were defined in an iterative manner similarly to the multiome dataset by first defining broad clusters resembling distinct cell lineages, followed by sub-clustering individual clusters.

Cell type annotations were guided by marker genes previously identified in the multiome dataset. As the probe panel did not target kidney-specific genes similar clustering resolutions were not possible, however we could resolve broad nephron segment signatures based on distinct marker expression patterns. For example, *EGF* expression characteristic of LOH and DCT tubule cells was observed, but due to lack of additional markers discriminating these two groups robust sub-clusters could not be derived. Therefore, epithelial cells of the nephron were broadly divided into PT, LOH-DCT as well as collecting duct (CD, with resolved PC and IC) segments. Notably, no CNT and ATL/DTL markers were detected, and these are likely grouped together with other segments. Using the previously defined injury and inflammatory markers we next searched for patterns of epithelial cell activation by sub clustering the broader epithelial clusters. We observed expression of injury markers such *VCAM1*, *TPM1*, *VIM*, *SPP1*, *ITGB6* and ITGB8. Further, markers predominantly expressed by inflammatory cells such as *CCL2*, *CXCL1*, *ICAM1* and *MMP7* were co-expressed by subsets of epithelia, confirming expression patterns observed in the multiome dataset (**Fig. 3D**). Therefore, these cells were annotated following these conventions. We observed expression patterns clearly identifying other non-epithelial cell types such as endothelia, fibroblasts, smooth muscle

cells, mesangial cells, myeloid cells as well as T and B lymphocytes. These were further subclustered if discrete populations were observed, e.g. myofibroblasts are distinguished by expression of *COL1A1* and *COL3A1* in addition to shared fibroblast markers (**Fig. S12**). In addition to level 1 and level 2 annotation, for visualisation purposes we derived a further annotation summarizing healthy, injured, and inflammatory epithelia as well as broad non-epithelial cell types (**Fig. S11D**). Visualisations of cell segmentations and transcript locations were generated using Seurat's ImageDimPlot function. For epithelial cell clusters proportions were calculated as the number of cells relative to the total of that cell type, while for non-epithelial cell proportions were calculated relative to the total number of cells.

Niche analysis

To define cell niches, a cell x neighbouring cells matrix was constructed by counting neighbouring cell types within a 75 µm radius. To account for differences in cell density, each entry was subsequently divided by the number of observed neighbours. Niches were defined using the Mclust algorithm with an EII model. Initially, 20 clusters were defined and annotated according to the most prominent cell type or anatomical structure in the cluster. Similar clusters were collapsed until 7 niches were defined (**Fig. 3E**). Enrichment of cell types within niches was assessed by calculating a log2-fold enrichment ratio defined as the quotient between the observed and expected number of cells of cells (proportions of the cell type across the dataset) plus 1. After log-transformation the data was centred on 0 (equivalent to no enrichment) by subtracting 1 (**Fig. S13C**). Correlation between the proportion of cells in the fibrotic niche and eGFR or %fibrosis by PSR staining was assessed using a linear regression slope with Pearson correlation coefficient and associated p-value (**Fig. S13D**).

Cell neighbourhood analysis

We inferred enrichment of a given cell type (ct_{query}) in proximity to other cell types by first defining a search radius of 25 µm measured from the centroid (of each cell's boundary polygon) to the centroid of a neighbouring cell. Next, we generated a ct_{query} by $ct_{neighbour}$ count matrix by counting the number of encounters of each level 2 annotations within the search radii of each query cells. Observed counts were then normalised by the absolute number of cells of the respective cell type encountered in the tissue section giving the observed frequency f_{obs}.

Simultaneously, we calculated the expected frequency f_{exp} as the proportion a given cell type across the entire tissue sample multiplied by the total number of cells encountered within the search radius (equivalent to random tissue architecture). The enrichment ratio was then defined as the log2-fold ratio between (f_{obs} / f_{exp}) + 1 and calculated for each ct_{query} to $ct_{neighbour}$ pair.

To derive a symmetric matrix the log2-fold enrichment ratios of a given ct_{query} to $ct_{neighbour}$ pair were multiplied with each other. For visualisation purposes enrichment values were log2 transformed and centred on 0 by substracting 1 (with > 0 indicating positive enrichment). We calculated the enrichment ratio for each tissue sample individually and used the mean across all samples for visualisation as

heatmap (**Fig. 3H**). For individual pairs we assessed statistical significance of enrichments across samples using the Wilcoxon signed-rank test with the R function wilcox.test with alternative = "two.sided" (**Fig. 3I**).

Transcript enrichment

In order to assess transcript enrichment in proximity of a given cell type we defined search radii up to a distance of 50 µm in 1 µm steps from the cell centroid. For a given cell type the number of detected transcripts of a given species in each interval were then counted and normalised by the area of the circle segment and the number cells in the area. Normalised transcript counts were averaged over all tissue sections and plotted as log2 + 1-fold enrichment over a sample of 10,000 random cells (independent of cell type) (**Fig. 3J, S14C**).

Cell culture

Primary renal proximal tubule epithelial cells (RPTEC/TERT1, ATCC CRL-4031) were cultured at 37°C in a 5% CO2 incubator using custom medium containing 25 ng/ml hydrocortisone (Merck, H0888), 3.5 µg/ml ascorbic acid (Sigma-Aldrich, A4403), 1% insulin-transferrin-selenium-ethanolamine (ITS-X) (Thermo-Fisher Scientific, 51500056), 6 pM Triiodo-L-Thyronine (Merck, T6397), 25 ng/ml Prostaglandin E1 (Merck, P8908), 0.8% v/v of 0.28 g/ml N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid (HEPES, Sigma-Aldrich, H4034) dissolved in 1M sodium hydroxide, 10 ng/ml recombinant human Epidermal Growth Factor (Promega, G5021) and 0.1 mg/ml geneticin (Gibco, 10131027) in Dulbecco's Modified Eagle Medium (GIBCO, 31331093). All experiments were performed with cells between passage 3 and 6.

For irradiation experiments, cells were plated in 6 well plates (200,000 cells in 2 ml media) then administered 10 Gy of irradiation 3 days later using a Caesium-137 gamma irradiator (CIS Bio International Gamma Irradiator type IBL 637). Control cells were 'mock' irradiated (taken to irradiator but not exposed to irradiation). Media was changed immediately following (+/- mock) irradiation and again 3 days later. SA-β-GAL staining (Cell Signalling Technology) was performed as per the manufacturer's instructions and imaged on the EVOS FL Auto 2 microscope. Images were analysed in in Image J (version 8). Separate macros were run for total cell counting and SA-β-GAL positive cell counting.

Samples were collected for RNA analysis 7 days after irradiation using 1 ml of TRIzol (Invitrogen, 15596026). The suspension was stored at − 80°C until RNA extraction. RNA was extracted using the RNeasy Plus micro kit (Qiagen) and included the optional DNase step. The RNA was eluted in 14 µl RNase-free water and concentration estimated using a Nanodrop One (Thermo- Fisher Scientific). RNA quality was assessed using the RNA 6000 Pico Kit (Agilent, 5067 − 1513) for RNA integrity numbers (RIN) > 7. Extracted RNA was sent to Genewiz where Poly-A enriched libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina following manufacturer's instructions (New England Biolabs, E7760).

For CUT&RUN experiments were plated at 0.8 × 10^6 cells/well into 6-well plates. The next day cells were treated with recombinant human TNF (20 ng/ml, Bio-Techne, 210-TA-020). After 24 hours cells were detached using 0.05% Trypsin and 0.02% EDTA in PBS for 8 min and resuspended in culture medium before nuclei isolation.

Bulk RNA-seq analysis

RNA-seq libraries from RPTEC irradiation experiments were sequenced on an NovaSeq 6000 platform (Illumina) with a mean output of 159 million 150bp paired-end reads per sample. After sample demultiplexing, adapter sequences were trimmed from FASTQ files using Trim-Galore (Cutadapt v2.8). Reads were aligned the GRCh38 (hg38) reference genome (GENCODE primary assembly) with a mean alignment rate of 84.3% using STAR (v2.7.9a)(62) with the option –quantMode GeneCounts. Gene counts were imported to R and analysed using DESeq2 (v3.18)(63). Genes with less than 200 overall counts were removed from the analysis. Differential expression analysis was performed using the DESeq function with default settings while controlling for variations due to replicate number in the model design. For visualisation purposes the count data were normalised using a variance stabilising transformation implemented by the DESeq2 vst function and scaled before plotting.

Re-analysis of ABT-263 RNA-seq data

Count matrices of previously published RNA-seq data were downloaded from gene expression omnibus under accession numbers GSE157866(42). Transcript abundances (Salmon transcript-level quantification) were imported into R and summarized to a gene level using the tximport function. Gene IDs were subsequently annotated to MGI nomenclature using BioMart functionality and the data was analysed using DESeq2 as described previously. Briefly, a DESeq2 object was created using DESeqDataSetFromMatrix with treatment and replicate as the design formula. Genes with less than 10 gene total gene counts were removed and the data was normalised using the DESeq and vst function with default setting. After Variance stabilizing transformation expression and gene-wise scaling, expression values of previously identified markers were visualised using pheatmap. Sample 4 from the ‘vehicle post R-UUO’ group was excluded from visualisation, due to behaving during outlier in principal component analysis and gene expression not behaving similarly to other samples in the group, leaving 5 remaining samples in the group.

CUT&RUN

Library preparation

CUT&RUN libraries from TNF-treated RPTECs were generated using the CUTANA ChIC/CUT&RUN kit (EpiCypher, 14–1048) with provided reagents unless indicated. After cell harvest samples nuclei were isolated. For this, cells were resuspended in 500 µl cold nuclei extraction buffer containing 2 mM HEPES pH 7.9, 10 mM KCl, 0.1% Triton X-100 (Merck, X100), 20% glycerol, 1X cOmplete protease inhibitor

cocktail (Roche, 11873580001), 0.5 mM spermidine in PBS. After incubation on ice for 5 min, nuclei were centrifuged (500 g, 5 min, 4°C) and the nuclei pellet was resuspended in nuclei extraction buffer without Triton X-100. Nuclei were counted using a LUNA-FX7 cell counter (Logos Biosystems) and 0.5×10^6 nuclei were used as input for the CUT&RUN assay.

The CUT&RUN assay was then performed following manufacturer instructions. Briefly, activated Concanavalin A (ConA)-conjugated paramagnetic beads were bound to nuclei. Using a magnetic rack bead-bound nuclei were washed and resuspended in antibody buffer. For control reactions 0.5 µg IgG negative control antibody (EpiCypher, 13–0042) was used while 0.5 µg of NFKB1 (Cell Signalling Technology, 13586S) or JUN (EpiCypher, 13-2019) were added for positive reactions. After overnight incubation and washes pAG-MNase was added and chromatin was digested by addition of $CaCl_2$ solution. Digestion was stopped by addition of the Stop Master Mix and the supernatant was transferred to a DNA clean-up column. After washes DNA was eluted in 12 µl of DNA Elution Buffer and quantified using the Qubit dsDNA Quantification kit (Thermo Fisher Scientific, Q32851).

For each reaction 5 ng purified DNA were used as input for library preparation using the CUTANA CUT&RUN Library Prep Kit (EpiCypher, 14–100) according to manufacturer instructions. Library was quantified using the Qubit dsDNA Quantification kit and fragment size was assessed using the Bioanalyzer High Sensitivity DNA Analysis kit (Agilent, 5067 – 4626).

CUT&RUN Analysis

Libraries were sequenced on a NextSeq 2000 platform (Illumina) with 100bp paired-end reads and a mean read depth of 10 million reads per library. After library demultiplexing remaining Illumina indices were removed using Trim Galore (Cutadapt v2.8) and aligned to the GRCh38 (hg38) reference genome with bowtie2 –local –very-sensitive –no-mixed –no-discordant –phred33 -I 10 -X 700 (v2.5.2)(64). Peak calling was performed using MACS2 (v2.2.7.1) with default settings. The peak significance threshold was determined using p-values derived from MACS2 with manual inspection of peak calling sensitivity at expected locations. A threshold of -log10(p-value) of 5 and 4 were used for JUN and NFKB1 peaks respectively (**Fig. S17A**). The BAM files were converted to BigWig format using the bamCoverage command from DeepTools (v3.5.4) with the normalisation option –normalizeUsing CPM and visualised using Signac's CoveragePlot() function. Overlap of ATAC peaks with CUT&RUN peaks was calculated requiring a 10% overlap in both peaks using the bedtools (v2.31.0) intersect function with parameters -wao -r -f 0.1. The overlapping ATAC peaks were used to infer differential accessibility in these regions using the RunPresto function with parameters logfc.threshold = 0, min.pct = 0.00 or find motifs using the Signac function FindMotifs with default settings (**Fig. 6F-G**). The fraction of ATAC reads in CUT&RUN peaks for each cell was calculated using Signac's FeatureMatrix() function using the original CUT&RUN peaks in BED format as the feature set to be quantified (**Fig. S17E-F**).

Data availability

Raw and processed sequencing data is available on Gene Expression Omninbus. Multiome snRNA/snATAC-seq data is available under the accession GSE254185. CosMx pipeline outputs are available under accession GSE253439. Bulk RNA-seq data of irradiated RPTECs and CUT&RUN data can be found under accessions GSE252584 and GSE254187 respectively.

R and python scripts employed in the analysis are available from the corresponding authors on request.

Statistics

Comparisons of clinical and histological characteristics between the unobstructed and obstructed groups were performed by students t-test or Chi-test for continuous or categorical data respectively. In the CosMx analysis, the relationship between abundance of cell types and eGFR or percentage fibrosis area was assessed using a linear regression slope with Pearson correlation. In the AP-1 inhibitor experiment, comparisons between the vehicle and T-5224 groups were assessed by students t-test.

Declarations

Acknowledgements

This work was funded by Kidney Research UK grants RP_024_20190306 and RP_013_20221129. MR was supported by a PhD studentship from the MRC Precision Medicine Programme, grant number MR/N013166/1. S.V. is supported by a pre-doctoral fellowship from the American Heart Association (967503) and a PhD studentship from University of Edinburgh College of Medicine and Veterinary Medicine Scholarship. WY and CW are employees and shareholders of NanoString Technologies, Inc.

We are grateful for the assistance of the flow cytometry core facility, University of Edinburgh for their assistance with nuclei isolation. We are also grateful to Mr Steve Leung and Mr Edward Mains, Consultant Urologists; Dr Marie O'Donnell, Consultant Uro-pathologist, Western General Hospital, Edinburgh and NHS Lothian Bioresource for assistance with sample selection and acquisition

References

1. Yu SM, Bonventre JV. Acute kidney injury and maladaptive tubular repair leading to renal fibrosis. Curr Opin Nephrol Hypertens. 2020;29(3):310-8.
2. Tanaka S, Portilla D, Okusa MD. Role of perivascular cells in kidney homeostasis, inflammation, repair and fibrosis. Nat Rev Nephrol. 2023;19(11):721-32.
3. Kuppe C, Ibrahim MM, Kranz J, Zhang X, Ziegler S, Perales-Paton J, et al. Decoding myofibroblast origins in human kidney fibrosis. Nature. 2021;589(7841):281-6.
4. Lake BB, Menon R, Winfree S, Hu Q, Ferreira RM, Kalhor K, et al. An atlas of healthy and injured cell states and niches in the human kidney. Nature. 2023;619(7970):585-94.

5. Park J, Shrestha R, Qiu C, Kondo A, Huang S, Werth M, et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science. 2018;360(6390):758-63.
6. Schreibing F, Kramann R. Mapping the human kidney using single-cell genomics. Nat Rev Nephrol. 2022;18(6):347-60.
7. Gisch DL, Brennan M, Lake BB, Basta J, Keller MS, Melo Ferreira R, et al. The chromatin landscape of healthy and injured cell types in the human kidney. Nat Commun. 2024;15(1):433.
8. Wilson PC, Muto Y, Wu H, Karihaloo A, Waikar SS, Humphreys BD. Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. Nat Commun. 2022;13(1):5253.
9. Kirita Y, Wu H, Uchimura K, Wilson PC, Humphreys BD. Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. Proc Natl Acad Sci U S A. 2020;117(27):15874-83.
10. Muto Y, Wilson PC, Ledru N, Wu H, Dimke H, Waikar SS, et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun. 2021;12(1):2190.
11. Wen Y, Su E, Xu L, Menez S, Moledina DG, Obeid W, et al. Analysis of the human kidney transcriptome and plasma proteome identifies markers of proximal tubule maladaptation to injury. Sci Transl Med. 2023;15(726):eade7287.
12. Gerhardt LMS, Liu J, Koppitch K, Cippa PE, McMahon AP. Single-nuclear transcriptomics reveals diversity of proximal tubule cell states in a dynamic response to acute kidney injury. Proc Natl Acad Sci U S A. 2021;118(27).
13. O'Sullivan ED, Mylonas KJ, Xin C, Baird DP, Carvalho C, Docherty MH, et al. Indian Hedgehog release from TNF-activated renal epithelia drives local and remote organ fibrosis. Sci Transl Med. 2023;15(698):eabn0736.
14. Law BMP, Wilkinson R, Wang X, Kildey K, Lindner M, Rist MJ, et al. Interferon-gamma production by tubulointerstitial human CD56(bright) natural killer cells contributes to renal fibrosis and chronic kidney disease progression. Kidney Int. 2017;92(1):79-88.
15. Conway BR, O'Sullivan ED, Cairns C, O'Sullivan J, Simpson DJ, Salzano A, et al. Kidney Single-Cell Atlas Reveals Myeloid Heterogeneity in Progression and Regression of Kidney Disease. J Am Soc Nephrol. 2020;31(12):2833-54.
16. Docherty MH, O'Sullivan ED, Bonventre JV, Ferenbach DA. Cellular Senescence in the Kidney. J Am Soc Nephrol. 2019;30(5):726-36.
17. Mason DX, Jackson TJ, Lin AW. Molecular signature of oncogenic ras-induced senescence. Oncogene. 2004;23(57):9238-46.
18. Zhang Y, Yang Y, Yang F, Liu X, Zhan P, Wu J, et al. HDAC9-mediated epithelial cell cycle arrest in G2/M contributes to kidney fibrosis in male mice. Nat Commun. 2023;14(1):3007.
19. Cormican S, Negi N, Naicker SD, Islam MN, Fazekas B, Power R, et al. Chronic Kidney Disease Is Characterized by Expansion of a Distinct Proinflammatory Intermediate Monocyte Subtype and by Increased Monocyte Adhesion to Endothelial Cells. J Am Soc Nephrol. 2023;34(5):793-808.

20. Mariani LH, Eddy S, AlAkwaa FM, McCown PJ, Harder JL, Nair V, et al. Precision nephrology identified tumor necrosis factor activation variability in minimal change disease and focal segmental glomerulosclerosis. Kidney Int. 2023;103(3):565-79.
21. Gao J, Wang D, Liu D, Liu M, Ge Y, Jiang M, et al. Tumor necrosis factor-related apoptosis-inducing ligand induces the expression of proinflammatory cytokines in macrophages and re-educates tumor-associated macrophages to an antitumor phenotype. Mol Biol Cell. 2015;26(18):3178-89.
22. Ware CF, Croft M, Neil GA. Realigning the LIGHT signaling network to control dysregulated inflammation. J Exp Med. 2022;219(7).
23. Vanbervliet B, Homey B, Durand I, Massacrier C, Ait-Yahia S, de Bouteiller O, et al. Sequential involvement of CCR2 and CCR6 ligands for immature dendritic cell recruitment: possible role at inflamed epithelial surfaces. Eur J Immunol. 2002;32(1):231-42.
24. Martina MG, Giorgio C, Allodi M, Palese S, Barocelli E, Ballabeni V, et al. Discovery of small-molecules targeting the CCL20/CCR6 axis as first-in-class inhibitors for inflammatory bowel diseases. Eur J Med Chem. 2022;243:114703.
25. Strutz F. The role of FGF-2 in renal fibrogenesis. Front Biosci (Schol Ed). 2009;1(1):125-31.
26. Ostendorf T, Boor P, van Roeyen CR, Floege J. Platelet-derived growth factors (PDGFs) in glomerular and tubulointerstitial fibrosis. Kidney Int Suppl (2011). 2014;4(1):65-9.
27. Meng XM, Nikolic-Paterson DJ, Lan HY. TGF-beta: the master regulator of fibrosis. Nat Rev Nephrol. 2016;12(6):325-38.
28. Huang Y, Wang F, Li H, Xu S, Xu W, Pan X, et al. Inhibition of Fibroblast Growth Factor Receptor by AZD4547 Protects Against Inflammation in Septic Mice. Inflammation. 2019;42(6):1957-67.
29. Chen X, Zhang X, Xu J, Zhao Y, Bao J, Zheng Z, et al. AZD4547 Attenuates Lipopolysaccharide-Induced Acute Kidney Injury by Inhibiting Inflammation: The Role of FGFR1 in Renal Tubular Epithelial Cells. Drug Des Devel Ther. 2020;14:833-44.
30. Turner JE, Rickassel C, Healy H, Kassianos AJ. Natural Killer Cells in Kidney Health and Disease. Front Immunol. 2019;10:587.
31. Poveda J, Tabara LC, Fernandez-Fernandez B, Martin-Cleary C, Sanz AB, Selgas R, et al. TWEAK/Fn14 and Non-Canonical NF-kappaB Signaling in Kidney Disease. Front Immunol. 2013;4:447.
32. Ko GJ, Jang HR, Huang Y, Womer KL, Liu M, Higbee E, et al. Blocking Fas ligand on leukocytes attenuates kidney ischemia-reperfusion injury. J Am Soc Nephrol. 2011;22(4):732-42.
33. Gerhardt LMS, Koppitch K, van Gestel J, Guo J, Cho S, Wu H, et al. Lineage Tracing and Single-Nucleus Multiomics Reveal Novel Features of Adaptive and Maladaptive Repair after Acute Kidney Injury. J Am Soc Nephrol. 2023;34(4):554-71.
34. Yoshimura Y, Muto Y, Omachi K, Miner JH, Humphreys BD. Elucidating the Proximal Tubule HNF4A Gene Regulatory Network in Human Kidney Organoids. J Am Soc Nephrol. 2023;34(10):1672-86.
35. Niewczas MA, Pavkov ME, Skupien J, Smiles A, Md Dom ZI, Wilson JM, et al. A signature of circulating inflammatory proteins and development of end-stage renal disease in diabetes. Nat Med.

2019;25(5):805-13.
36. Park SK, Yang WS, Han NJ, Lee SK, Ahn H, Lee IK, et al. Dexamethasone regulates AP-1 to repress TNF-alpha induced MCP-1 production in human glomerular endothelial cells. Nephrol Dial Transplant. 2004;19(2):312-9.
37. Hayden MS, Ghosh S. Regulation of NF-kappaB by TNF family cytokines. Semin Immunol. 2014;26(3):253-66.
38. Martinez-Zamudio RI, Roux PF, de Freitas J, Robinson L, Dore G, Sun B, et al. AP-1 imprints a reversible transcriptional programme of senescent cells. Nat Cell Biol. 2020;22(7):842-55.
39. Aikawa Y, Morimoto K, Yamamoto T, Chaki H, Hashiramoto A, Narita H, et al. Treatment of arthritis with a selective inhibitor of c-Fos/activator protein-1. Nat Biotechnol. 2008;26(7):817-23.
40. Zhou P, Wan X, Zou Y, Chen Z, Zhong A. Transforming growth factor beta (TGF-beta) is activated by the CtBP2-p300-AP1 transcriptional complex in chronic renal failure. Int J Biol Sci. 2020;16(2):204-15.
41. Mylonas KJ, O'Sullivan ED, Humphries D, Baird DP, Docherty MH, Neely SA, et al. Cellular senescence inhibits renal regeneration after injury in mice, with senolytic treatment promoting repair. Sci Transl Med. 2021;13(594).
42. O'Sullivan ED, Mylonas KJ, Bell R, Carvalho C, Baird DP, Cairns C, et al. Single-cell analysis of senescent epithelia reveals targetable mechanisms promoting fibrosis. JCI Insight. 2022;7(22).
43. Adachi T, Sugiyama N, Gondai T, Yagita H, Yokoyama T. Blockade of Death Ligand TRAIL Inhibits Renal Ischemia Reperfusion Injury. Acta Histochem Cytochem. 2013;46(6):161-70.
44. Xu S, Yang X, Chen Q, Liu Z, Chen Y, Yao X, et al. Leukemia inhibitory factor is a therapeutic target for renal interstitial fibrosis. EBioMedicine. 2022;86:104312.
45. Ishida M, Ueki M, Morishita J, Ueno M, Shiozawa S, Maekawa N. T-5224, a selective inhibitor of c-Fos/activator protein-1, improves survival by inhibiting serum high mobility group box-1 in lethal lipopolysaccharide-induced acute kidney injury model. J Intensive Care. 2015;3:49.
46. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race. N Engl J Med. 2021;385(19):1737-49.
47. Ferenbach DA, Nkejabega NC, McKay J, Choudhary AK, Vernon MA, Beesley MF, et al. The induction of macrophage hemeoxygenase-1 is protective during acute kidney injury in aging mice. Kidney Int. 2011;79(9):966-76.
48. Huang X, Huang Y. Cellsnp-lite: an efficient tool for genotyping single cells. Bioinformatics. 2021;37(23):4569-71.
49. Huang Y, McCarthy DJ, Stegle O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. Genome Biol. 2019;20(1):273.
50. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87 e29.

51. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods. 2021;18(11):1333-41.
52. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.
53. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796-8.
54. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.
55. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38(12):1408-14.
56. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018;19(1):477.
57. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.
58. Bravo Gonzalez-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods. 2023;20(9):1355-67.
59. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.
60. He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvatana K, et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat Biotechnol. 2022;40(12):1794-806.
61. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96.
62. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.
63. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

Figures

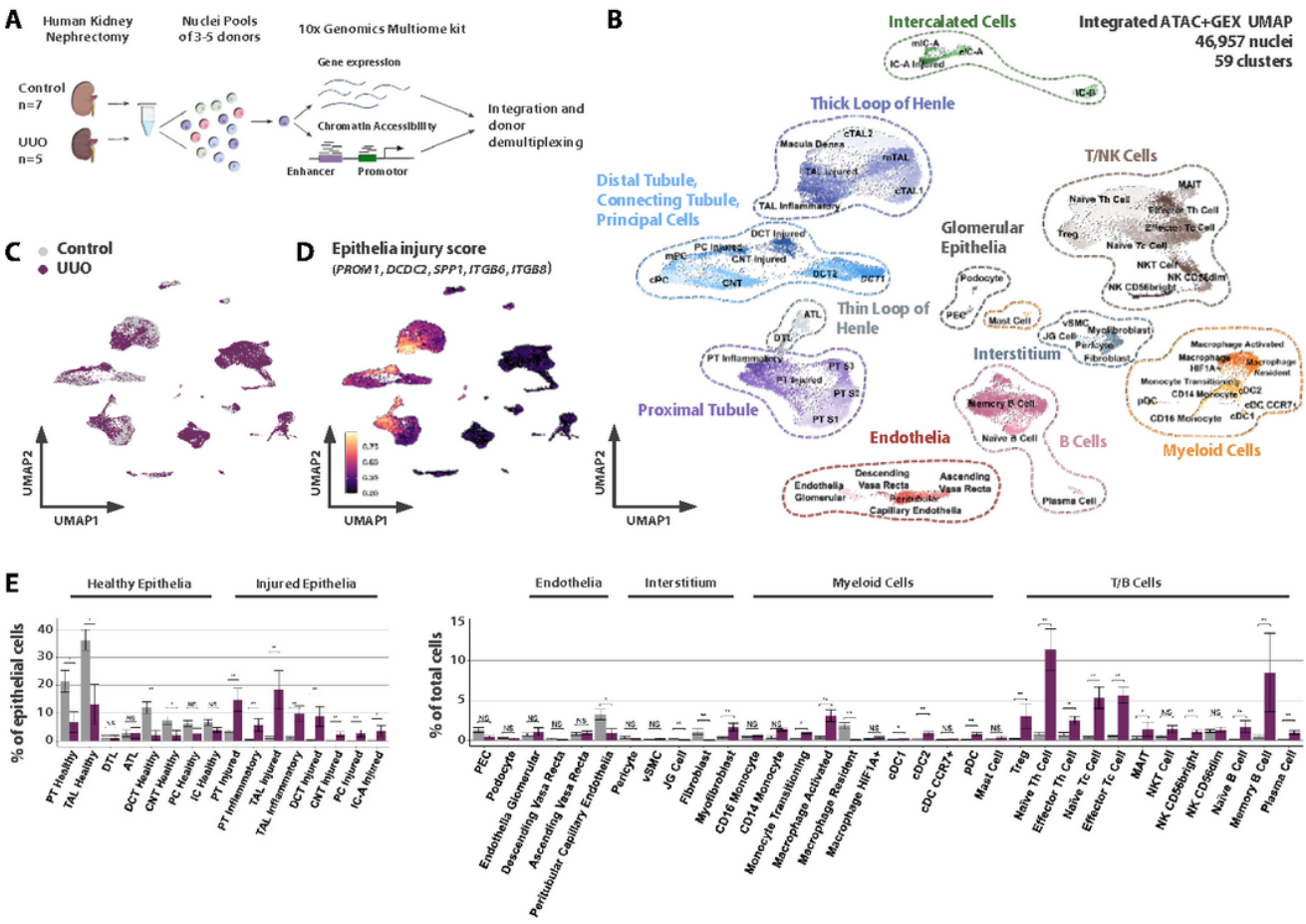

Figure 1

A multiomic single-nuclei atlas of kidney injury.

A) Overview of the single nuclear multiome Assay for Transposase Accessible Chromatin for sequencing (snATAC-seq) and snRNA-seq) workflow. Nuclei were isolated from human nephrectomy specimens from control (n = 7) or unilateral ureteric obstructed (UUO, n = 5) kidneys and processed in pools of three to five samples per library using the 10x Genomics Multiome kit. The original donor identity of each sequenced nucleus was inferred using single nucleotide polymorphisms present in sequencing reads of snRNA-seq and snATAC-seq data.

B)Weighted-nearest neighbour UMAP projection of joint transcriptome and open chromatin state of 46,957 nuclei from control (n = 7) and UUO (n = 5) samples with detailed cluster annotations. PEC, parietal epithelial cell; PT, proximal tubule; DTL, descending thin limb; ATL, ascending thin limb; TAL, thick ascending limb; DCT, distal convoluted tubule; CNT, connecting tubule; PC, principal cell; IC, intercalated cell; vSMC vascular smooth muscle cell; JG cell, juxtaglomerular cell; pDC, plasmacytoid dendritic cell; cDC, classical dendritic cell; Th, T helper cell; Tc, cytotoxic T cell.

C) UMAP projection annotating cells to control or UUO kidneys.

D) Gene expression score of epithelial injury genes universally expressed in injured nephron segments (*PROM1*, *DCDC2*, *SPP1*, *ITGB6*, *ITGB8*).

E) Percentage of nuclei assigned to each cell type as a proportion of all epithelial cells (for epithelial clusters, left) and of total cells (for non-epithelial clusters, right). Plots show means ± standard error of the mean (SEM). Wilcoxon rank-sum test. *p < 0.05; **p < 0.01.

<u>*A high resolution figure file is available in the Supplementary Files section of this preprint.</u>

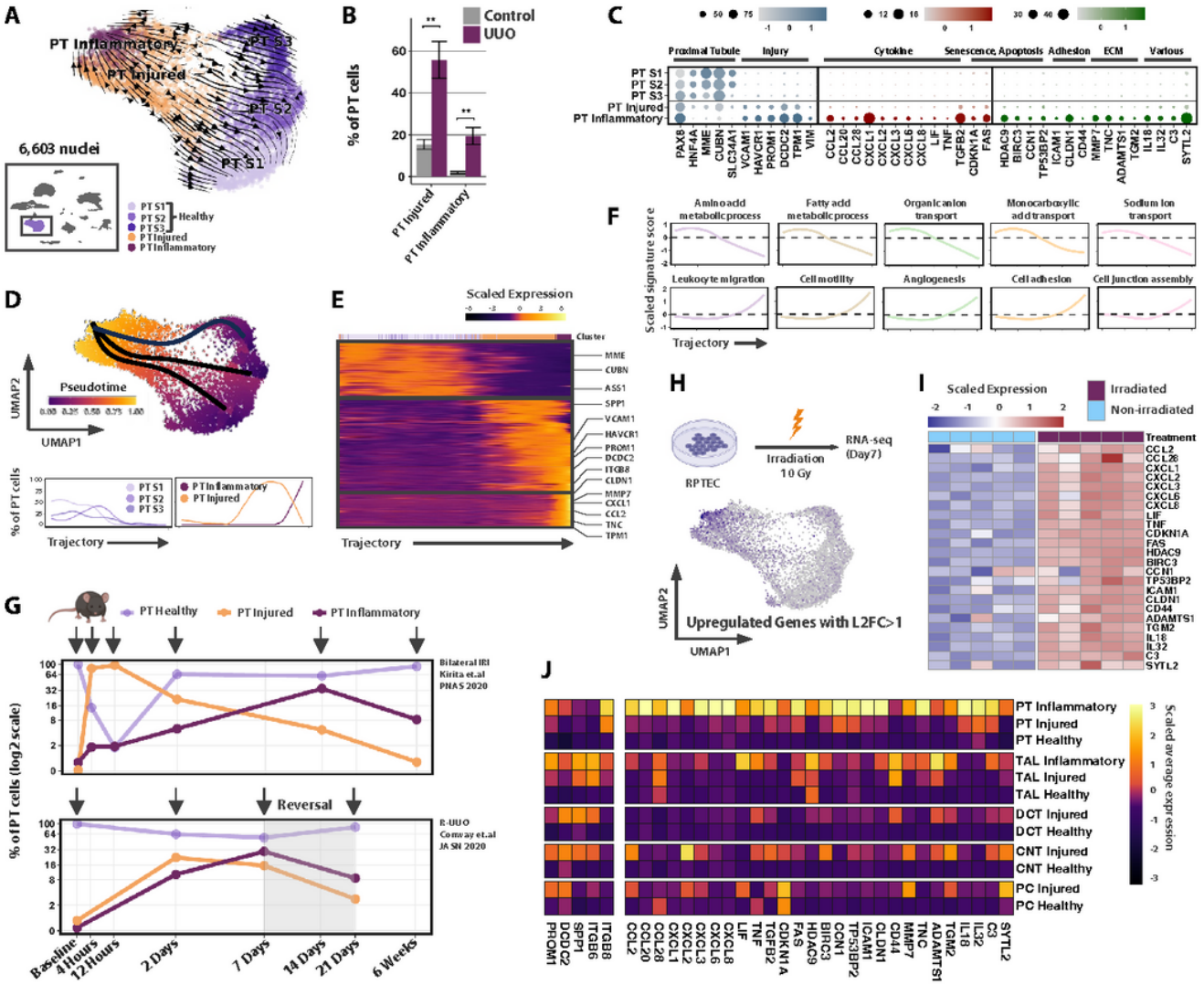

Figure 2

Injured epithelial cells can adopt a pro-inflammatory, pro-fibrotic phenotype.

A) Proximal tubule (PT) subset of the UMAP (**Fig. 1B**) with cells coloured by subcluster annotations and projected RNA velocities derived from the most variable genes.

B) Barplot showing the percentage of injured and inflammatory PT cells in control and UUO kidneys as a proportion of total PT cells. The plots show means ± SEM. Wilcoxon rank-sum test. **p < 0.01.

C) Dot plot of selected cytokines, senescence, apoptosis, adhesion and extracellular matrix remodelling associated genes upregulated by inflammatory PT cells. Dot colours show the averaged gene expression values (log scale) and size indicates proportion of cells expression each gene with scales indicated by the colour.

D) Top panel: UMAP (as in **Fig. 2A**) with cells ordered in pseudotime on a trajectory from healthy PT segments to the inflammatory cell state. Bottom panel: The percentage of cells of each subcluster along the pseudotime trajectory as a proportion of all PT cells at that trajectory point.

E) Heatmap of smoothed gene expression dynamics along the inferred pseudotime trajectory (**Fig. 2D**).

F) Smoothed dynamics of gene signature scores from genes associated with the top GO terms enriched in healthy (PT S1-S3) or injured and inflammatory PT cells.

G) Inflammatory PT cell phenotypes are observed in murine models and persist following injury. The time course shows the proportion of inflammatory, injured and healthy PT cells (as a proportion of total PT cells) at different timepoints in single-cell/nuclei datasets of ischaemia-reperfusion injury (IRI, top(9)) and reversible unilateral ureteric obstruction (R-UUO(15), bottom - grey shading indicates the period following reversal of ureteric obstruction).

H)Radiation-induced senescence in human primary renal proximal tubule epithelial cells (RPTECs) promotes a gene signature comparable to that of inflammatory PT cells *in vivo* (**Fig 2C**). Top panel: Overview of experimental strategy. Bottom panel: Scaled gene signature score of upregulated genes (log2-fold change (L2FC) > 1, adjusted p-value < 0.05) in irradiation-induced senescent RPTECs mapped onto the UMAP of human PT cells (**Fig 2A**).

I) Heatmap of scaled gene expression in control and irradiated RPTECs, showing the gene expression changes of markers of inflammatory PT cells (**Fig 2C**).

J) Inflammatory PT markers are partially shared with other tubular segments. Heatmap showing the scaled average expression of markers of inflammatory PT cells (**Fig 2C**).

***A high resolution figure file is available in the Supplementary Files section of this preprint.**

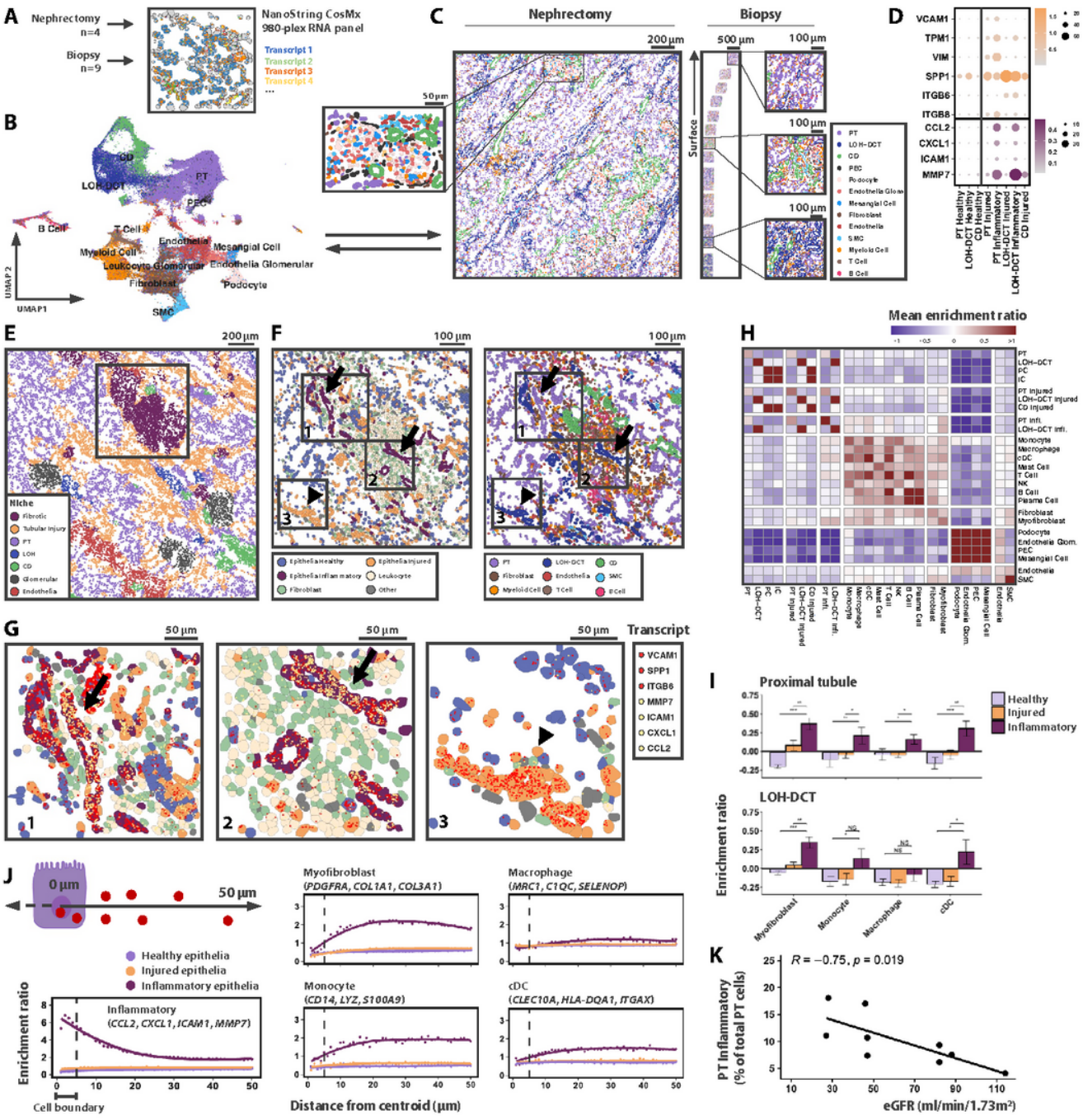

Figure 3

High-plex in situ imaging reveals the association of the fibrotic niche with inflammatory epithelia.

A) Overview of the CosMx spatial imaging workflow. Nephrectomy samples (n = 4) from patients with chronic pyelonephritis and needle biopsy samples (n = 9) with minimal change disease or IgA nephropathy were analysed using a 980-plex RNA panel.

B) UMAP visualisation of 246,055 cells captured in nephrectomy and biopsy samples coloured by broad cell annotations. PEC, parietal epithelial cell; PT, proximal tubular cell; LOH-DCT, Loop of Henle and distal convoluted tubule; CD, Collecting Duct; SMC, Smooth muscle cell.

C) Projection of cell annotation into 2D space recapitulates the kidney architecture. Representative images of a nephrectomy and a biopsy sample. Images show the cell segmentation boundaries in their 2D coordinates coloured by broad cell annotations (**Fig. 3B**).

D) Expression pattern of injury and inflammatory markers identified by snRNA-seq (**Fig. 2C**) by sub-clusters of PT, LOH-DCT and CD cells in the CosMx dataset. Dot colours show the averaged gene expression values (log scale) and size indicates proportion of cells in which they are expressed.

E) Representative image demonstrating the location of niches enriched for cells of a given cell type.

F) Higher resolution of area highlighted in **Fig. 3E** demonstrating that inflammatory epithelia (arrows), myofibroblasts and leucocytes localise to the fibrotic niche, whereas tubules enriched for injured cell phenotypes (arrowhead) are located in the injury niche, which contains fewer myofibroblasts and leucocytes. Cells are coloured according to their activation state and broad cell lineage (left), with a more granular annotation of tubular and interstitial cell types (right).

G) Higher resolution of the 3 areas highlighted in **Fig. 3F** with dots showing the location of selected injury (red) or inflammatory-related (yellow) transcripts. Inflammatory epithelia are enriched for both injury and pro-inflammatory related transcripts (arrows) while injured epithelia lack the latter (arrowhead). Cells are coloured according to the legend in **Fig.3E**

H) Cell neighbourhoods. Heatmap of mean log2-fold enrichment (mean of all samples) showing the probability of observing each cell type within 25 µm radius of the reference cell type. An enrichment of zero indicates a ratio equal to that expected in a random distribution of cells in the tissue. Values greater than zero indicate enrichment in the neighbourhood and value less than zero indicate relative depletion. cDC, classical dendritic cell; NK, natural killer cell

I) Bar plots showing the enrichment ratio of myofibroblasts and myeloid cell types in proximity to PT (top panel) or LOH-DCT (bottom panel) cell states. Plots show means ± standard error of the mean (SEM). Wilcoxon rank-sum test. *$p < 0.05$; **$p < 0.01$.

J) Transcripts associated with myofibroblasts, monocytes and cDCs are enriched in proximity to inflammatory cell states. Plots show the enrichment ratio (mean of the indicated transcripts) in 1 µm intervals relative to centroids of a given cell type.

K) Correlation between the percentage of inflammatory PT cells and estimated glomerular filtration rate (eGFR) in biopsy samples. Displaying a linear regression slope with Pearson correlation.

*A high resolution figure file is available in the Supplementary Files section of this preprint.

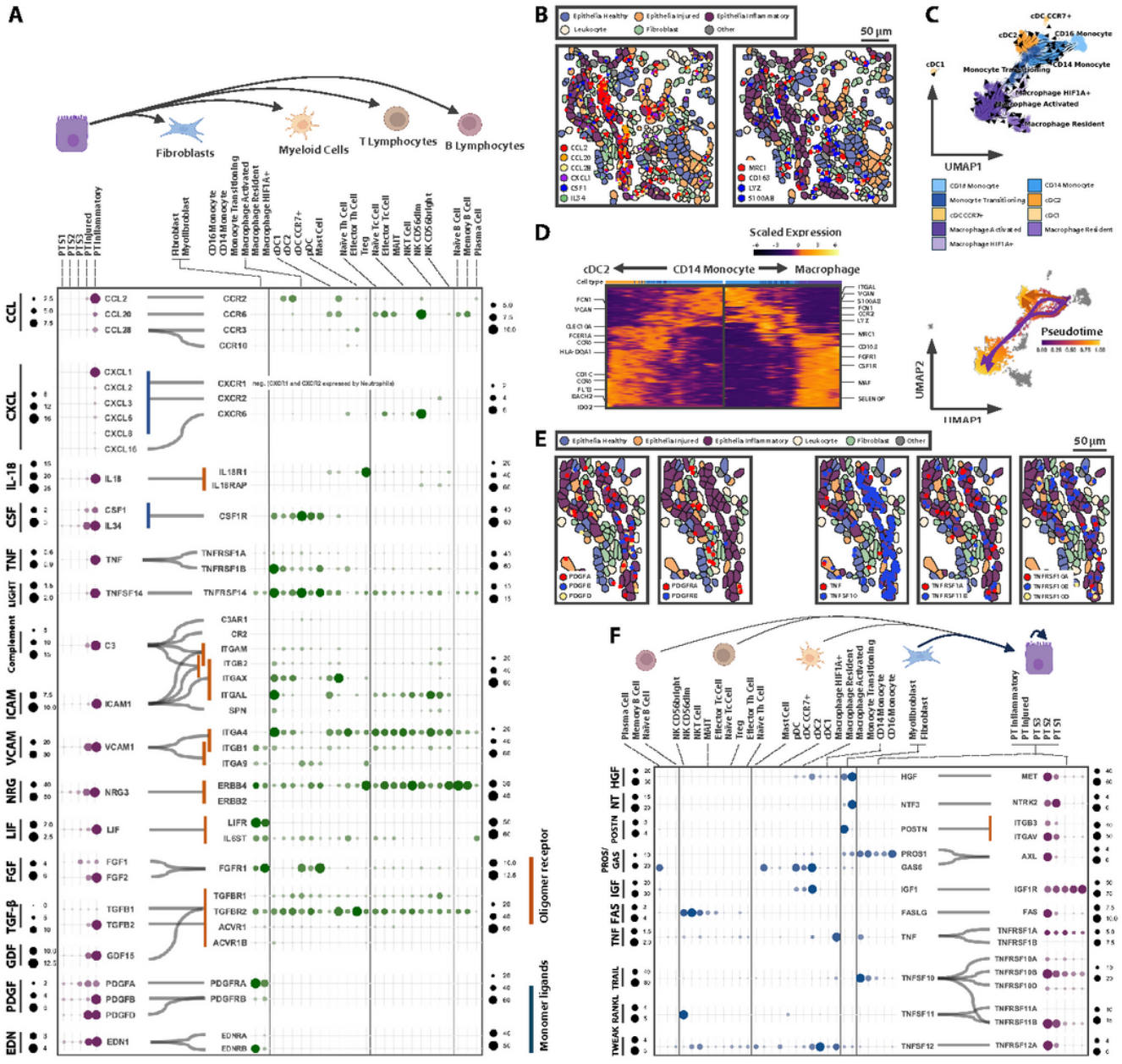

Figure 4

Intercellular signalling mediating immune recruitment and fibroblast activation.

A) Outgoing signalling patterns for ligands enriched in inflammatory PTs. Dot plots in the left columns show the expression of ligands in PT clusters and the right column the expression pattern of respective receptors in immune and fibroblast subclusters. Dot size is scaled by the fraction of cells expressing the gene and colour by the average (log-scale) scaled expression values.

B) Spatial expression patterns of identified ligands interacting with myeloid cells. The images show the spatial location of detected ligands (left) and localisation of monocytes (blue) and macrophages (red) in the same frame (right). Cell boundaries are coloured by cell lineage and epithelial activation state.

C) Myeloid differentiation trajectories. UMAPs of myeloid subclusters and projected RNA velocities using a dynamical model of the most variable genes.

D) Heatmap showing smoothed gene expression dynamics along the trajectories (**Fig. 4C**) to cDC2 cells (left panel) and activated macrophages (right panel).

E) Spatial expression patterns of ligand-receptor pairs implicated in fibroblast (left) and paracrine signalling in inflammatory tubules (right). Transcripts are represented by dots in 2D space and coloured by transcript species. Cell boundaries are coloured by cell lineage and epithelial activation state.

F) Incoming signalling patterns for receptors differentially expressed by Inflammatory PT cells. The left columns shows the expression patterns of ligands in interacting cell types and the right column the expression of respective receptors by PTs. The dot size is scaled by the fraction of cells expressing the gene and the colour by the average (log-scale) expression.

***A high resolution figure file is available in the Supplementary Files section of this preprint.**

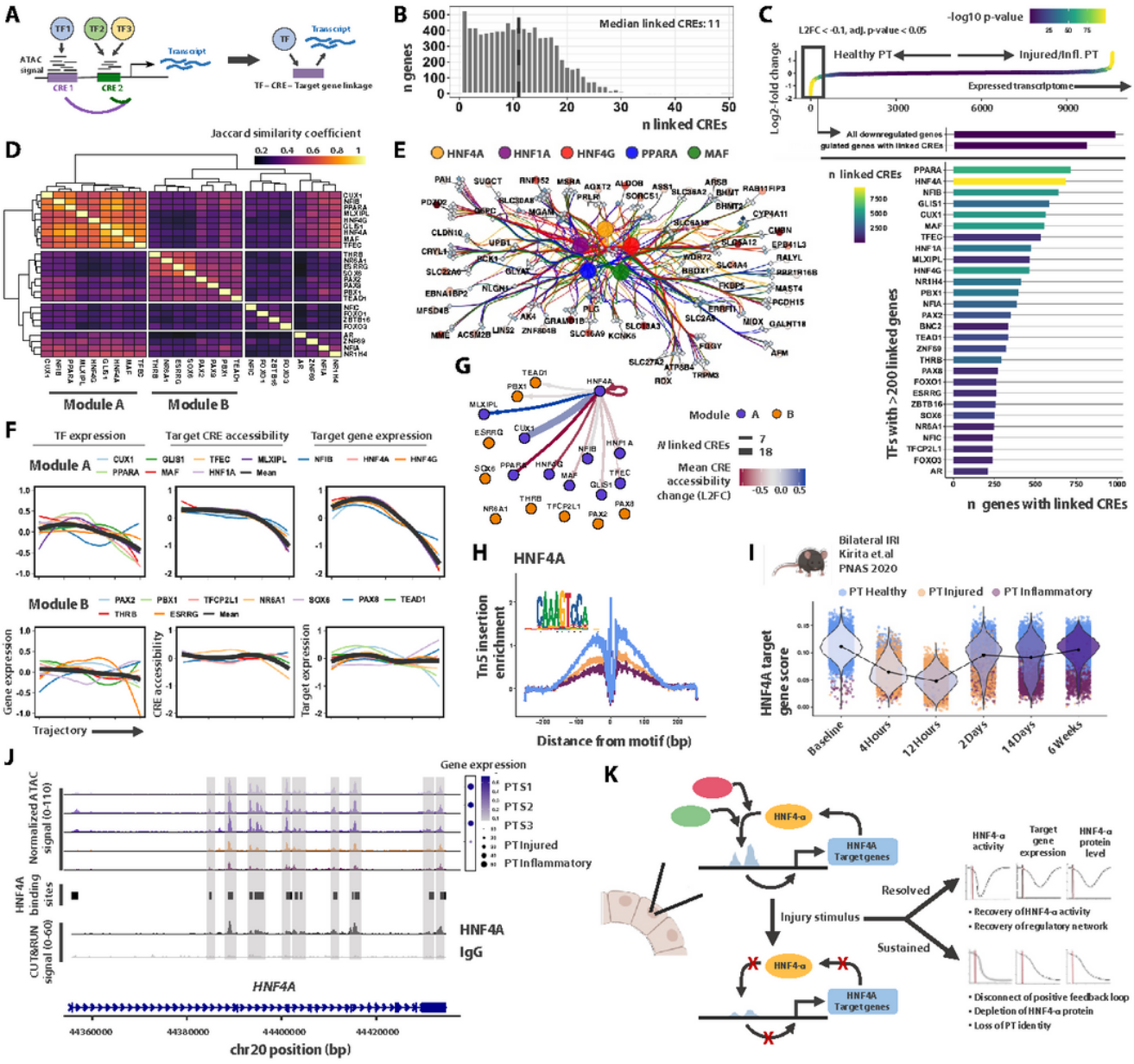

Figure 5

Collapse of an HNF4A-driven gene-regulatory network is linked to loss of proximal tubule identity.

A) Conceptual visualisation of the strategy used to infer the relationship of cis-regulatory elements (CREs) and transcription factors (TFs) with target genes. Links between genes and transcription factors are inferred when there is high correlation between gene expression and the extent of chromatin accessibility in CREs that harbour the relevant transcription factor binding motif.

B) Distribution of the number of linked CREs per gene (for all genes with detected linkages).

C) The number of genes linked to each TF for all genes downregulated in injured/inflammatory PT cells. Top panel: Gene expression changes between healthy and injured/inflammatory PT cells. All downregulated genes (L2FC < -0.1, adj. p-value < 0.05) were selected to quantify the fraction of genes linked to each TF. Bottom panel: The bar graph shows the number of downregulated genes which have one or more binding sites linked to a given transcription factor. The colour indicates the total number of CREs linked to downregulated genes bound by the TF.

D) Comparison of target genes suggests distinct regulatory networks. The heatmap shows the Jaccard similarity coefficient between pairs of transcription factors, with 1 indicating a full overlap and 0 no overlap of target genes.

E) Visualisation of the gene-regulatory network for a subset of module A TFs (showing TF-CRE-gene linkages with a L2FC < -0.6, CREs with a L2FC < -1.5, adj. p-value < 0.05). CREs are shown as diamond shaped nodes with darker blue colour indicating decreased accessibility. Target genes nodes are shown as circles with darker red colour indicating stronger down-regulation. TF-to-CRE edges are coloured by TF linked to the CRE.

F) Regulatory network activity during transition from health to an inflammatory phenotype. Left column: Changes of scaled gene expression for TF genes. Middle column: Scaled CREs accessibility score for TF target CREs. Right column: Scaled target gene expression score for TF target genes. Each panel shows the changes of the modality along the pseudotime trajectory form healthy to inflammatory PT cells (**Fig. 2D**) with the top row representing TFs in module A and the bottom row TFs in module B.

G) The network graph shows HNF4-α binding sites in CREs linked to TFs in module A and B. The edge width indicates the number of bound CREs and the edge colour the mean accessibility change in injured/inflammatory PT cells compared with healthy PT cells (negative values indicate reduced accessibility).

H) Footprinting analysis for the HNF4-α motif shows reduced binding site occupancy in injured and inflammatory PT.

I) The expression level of HNF4-α target genes does not recover in PT cells that adopt a persistent inflammatory phenotype following ischaemia-reperfusion injury (IRI) in mice (**Fig. 2G**). The graph shows the signature scores of HNF4-α target genes along the IRI time course. The violin plot shows the distribution of signature score in all cells, with dots representing individual PT cells coloured by their cell state.

J) The *HNF4A* locus demonstrates abundant self-regulation. Upper tracks: The chromatin accessibility profile in healthy and injured/inflammatory PT cells with associated *HNF4A* mRNA levels to the right (dot plot, dot size is scaled by the proportion of cells expressing the gene and coloured by the average (log-scale) expression). Middle track: The predicted HNF4-α binding sites based on presence of the HNF4-α motif. Lower track: HNF4-α CUT&RUN signal in adult human kidney (34).

K) Proposed mechanism. In health, HNF4-α binds its target CREs including in its own regulatory elements promoting a positive feedback loop along with a PT phenotype. Following injury, HNF4-α binding is disrupted, thereby uncoupling the positive feedback loop and reducing target gene expression. Sustained injury leads to depletion of *HNF4A*mRNA and protein along with chromatin remodelling leading to a loss of PT identity.

***A high resolution figure file is available in the Supplementary Files section of this preprint.**

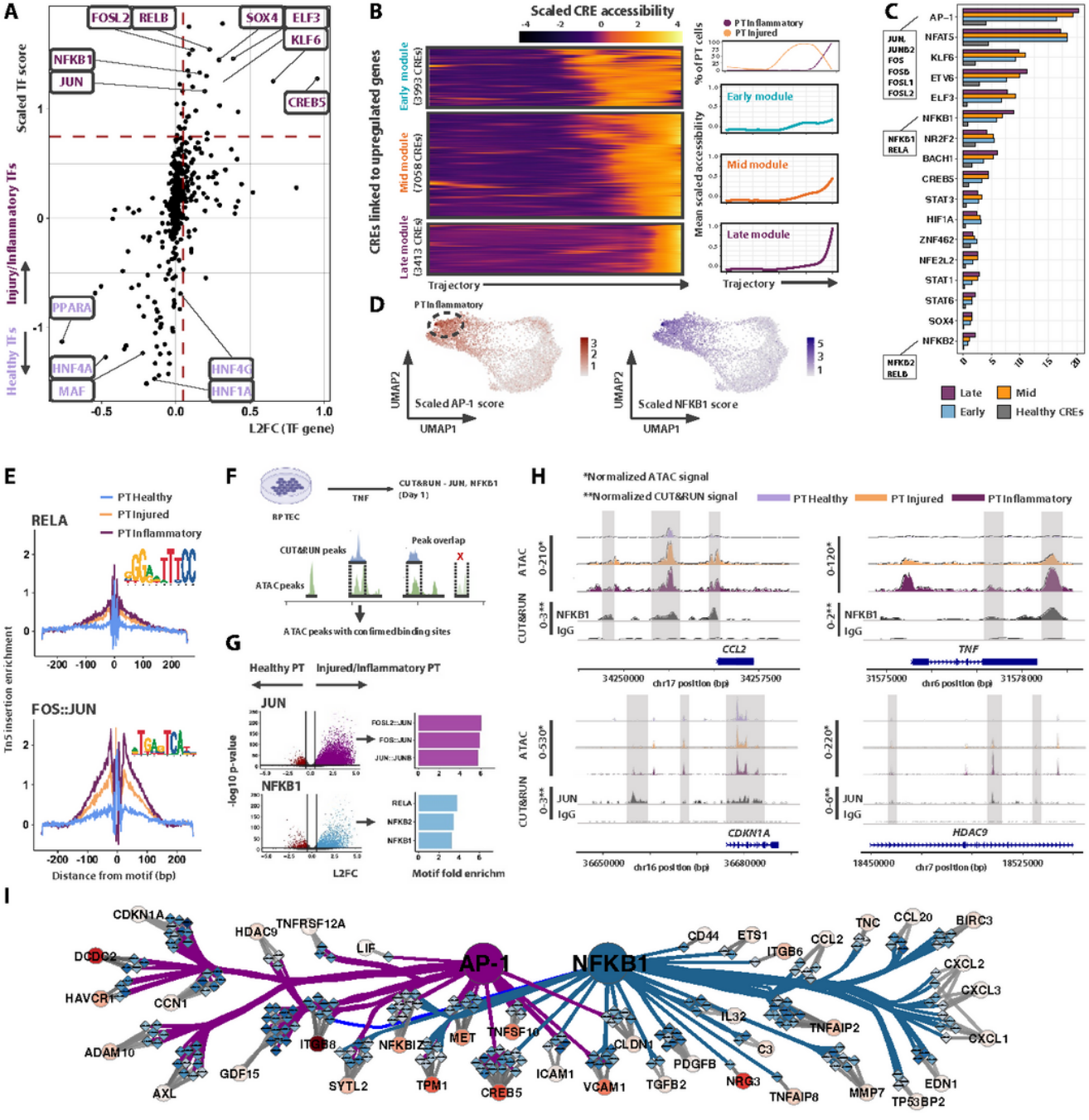

Figure 6

Transcription factors associated with inflammatory proximal tubule cells.

A) Identification of transcription factors enriched in injured and inflammatory PT cells. The L2FC in expression of the transcription factor gene plotted against the transcription factor score (mean of target gene and target CRE signature scores).

B) Chromatin accessibility dynamics along the trajectory from healthy to inflammatory PT cells (**Fig. 2D**). Left panel: CREs linked to upregulated genes (L2FC > 0.1, adj. p-value < 0.05) is plotted against the inferred pseudotime values. CREs are clustered into modules with accessibility changes early, intermediate and late in the trajectory. Right panel: Smoothed graphs summarizing the mean accessibility for CREs in each module are provided, with the percentage of injured and inflammatory PT cells along the same trajectory above.

C) The percentage of CREs with increased accessibility in each module that contain motifs for the respective TF family (for TFs falling within thresholds indicated in **Fig. 6A** and > 75 target genes) AP-1 and NF-κB family motifs are combined due to motif similarity and as they bind binding as heterodimers.

D) UMAP (**Fig 2A**) of the AP-1 and NF-κβ1 TF score.

E) TF footprint of RELA and JUN:FOS motifs show highest accessibility of binding sites in inflammatory PT cells.

F) Integration strategy for TF CUT&RUN (JUN and NF-κβ1) assay in renal proximal tubular cells (RPTECs) and chromatin accessibility. CUT&RUN peaks are intersected with ATAC peaks to identify areas of differentially accessible chromatin in inflammatory PT cells that bind JUN or NF-κβ1.

G) Left panel: Volcano plots showing accessibility changes of peaks overlapping with JUN or NF-κβ1 CUT&RUN peaks. For both JUN or NF-κβ1 there was an enrichment in the number of peaks mapping to regions of increased chromatin accessibility in injured/inflammatory PT cells; Right panel: Motif enrichment ratio in overlapping ATAC peaks for AP-1 or NF-κB family motifs.

H) Chromatin profile and CUT&RUN signal tracks at predicted archetypal inflammatory PT genes with NF-κβ1 (top row) or AP-1 (bottom row) target CREs. Upper tracks: The ATAC signal in healthy and injured/inflammatory PT cells. Lower track: The NF-κβ1 or JUN CUT&RUN peaks indicating the physical presence of the TF at genomic loci in TNF-treated RPTECs.

I) Visualisation of the predicted regulatory network of selected AP-1 and NF-κβ1 target genes. CREs are shown as diamond shaped nodes (darker blue colour indicating increased accessibility) and target genes as circles (darker red colour indicating greater up-regulation). TF to CRE edges are coloured by the TF bound to the CRE.

*A high resolution figure file is available in the Supplementary Files section of this preprint.

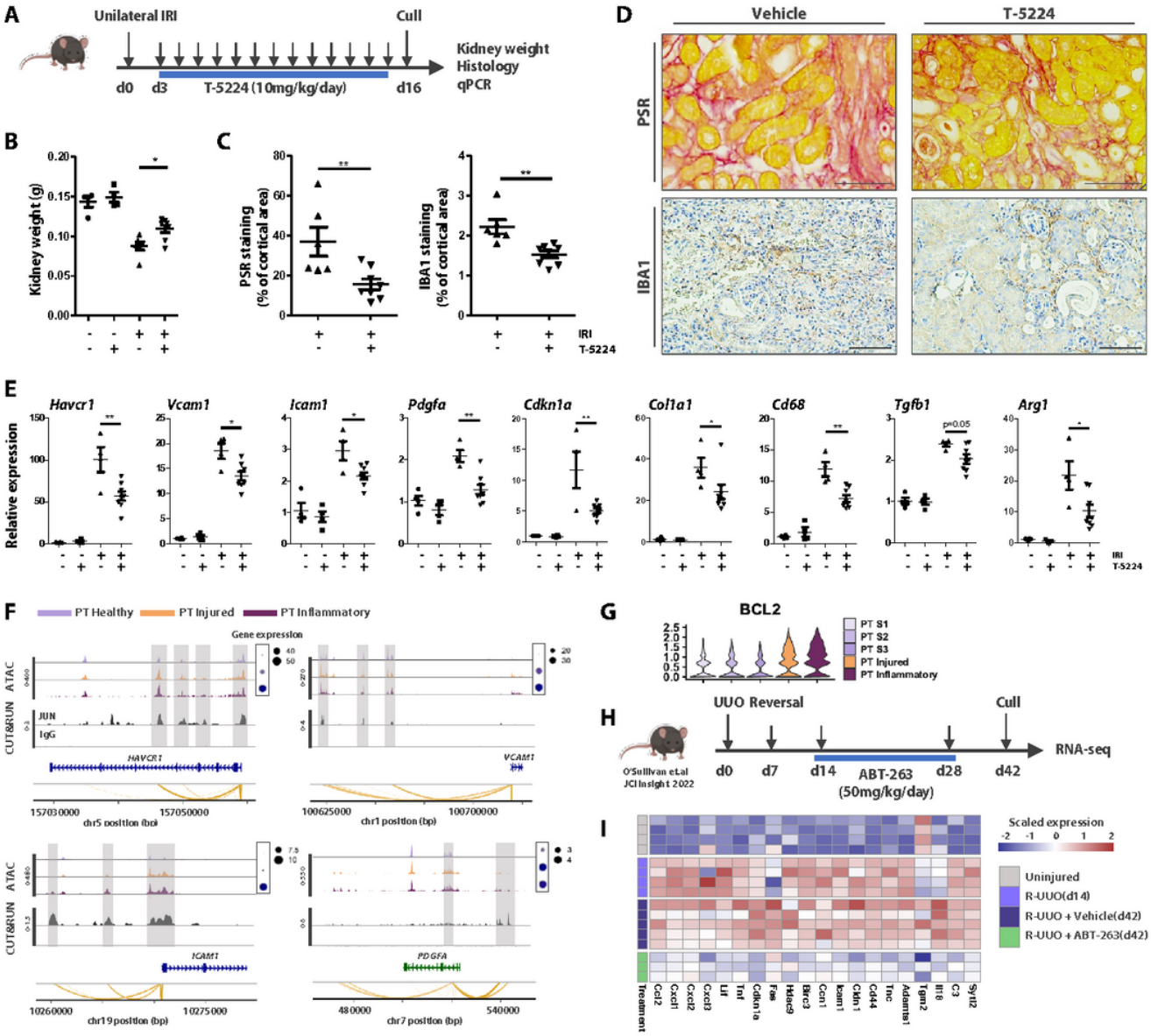

Figure 7

Targeting inflammatory PT cells ameliorates fibrosis in mouse models of AKI to CKD transition.

A) Schema of animal model. Mice underwent unilateral ischaemia-reperfusion (IRI) surgery, were given the AP-1 inhibitor T-5224 daily 3 days post-injury and mice sacrificed after 16 days. Kidney weight was recorded, and histology and qPCR were used to assess magnitude of fibrosis.

B) Mass of ischaemic or sham kidneys. Values are means ± SEM. Student's t test. *p < 0.05.

C) Quantification of percentage of renal cortex staining for picrosirius red (PSR, left) or IBA1 (macrophages, right). Values are means ± SEM. Student's t test. **p < 0.01.

D) Representative images of PSR staining (top) and IBA1 staining (bottom) in animals that underwent IRI and received either vehicle (left) or T-5224 (right).

E) Gene expression changes of genes regulated by AP-1, and injury or fibrosis markers, relative to *Ppia* expression. Values are means ± SEM. Student's t test. *p < 0.05; **p < 0.01.

F) Gene regulatory landscape of AP-1 target genes human PT cells. Top tracks: The chromatin accessibility profile in human PT cell subsets with corresponding gene expression indicated by dot plots next to chromatin tracks, with dots scaled by the proportion of cells expressing the gene. Bottom tracks: JUN CUT&RUN signal in TNF-induced inflamed RPTECs (**Fig. 6F**). Peak regions that correlate with gene expression are linked to the TSS below the gene annotation (yellow line, width scaled in proportion to the correlation coefficient).

G) Violin plot of quantification of *BCL2* expression in human PT subtypes (log scale).

H) Schema of reversible unilateral ureteric obstruction (R-UUO) model using the BCL2 inhibitor ABT-263(42).

I) Heatmap of scaled bulk level gene expression in control kidneys and 7 days post reversal of ureteric obstruction (R-UUO, d14) or 35 days after reversal following 1 week treatment with ABT-243 or inhibitor. Compared to vehicle treated mice, ABT-263 treated mice show a reduction of an inflammatory PT signature.

***A high resolution figure file is available in the Supplementary Files section of this preprint.**

Supplementary Files

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

- SupplementaryFigures.pdf
- SupplementaryTables.xlsx
- MainFigures.pdf