bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Insights from Comparison of the Renal and Skin Single Cell Transcriptomes in Lupus Nephritis

Evan Der[#1], Hemant Suryawanshi[#2], Pavel Morozov[2], Manjunath Kustagi[2], Beatrice Goilav[3], Saritha Ranabathou[3], Peter Izmirly[4], Robert Clancy[4], H. Michael Belmont[4], Mordecai Koenigsberg[5], Michele Mokrzycki[6], Helen Rominieki[7], Jay A. Graham[7], Juan P. Rocca[7], Nicole Bornkamp[4], Nicole Jordan[1], Emma Schulte[1], Ming Wu[4], James Pullman[8], Kamil Slowikowski[9], Soumya Raychaudhuri[9], Joel Guthridge[10], Judith James[10], Accelerating Medicines Partnership in RA/SLE network, Jill Buyon[*4], Thomas Tuschl[*2], and Chaim Putterman[*1]

[#] Authors contributed equally

[*] Co-corresponding authors

[1]Albert Einstein College of Medicine, Bronx, New York, USA.

[2]Howard Hughes Medical Institute and Laboratory for RNA Molecular Biology, The Rockefeller University, New York, New York, USA.

[3]Pediatric Nephrology, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, New York, USA.

[4]New York University School of Medicine, New York, New York, USA.

[5]Department of Radiology, Montefiore Medical Center, Bronx, New York, USA.

[6]Division of Nephrology, Department of Medicine, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, New York, USA.

[7]Montefiore Einstein Center for Transplantation, Montefiore Medical Center, Bronx,

NY, USA.

[8]Clinical Pathology, Montefiore Medical Center, Bronx, NY, USA.

[9]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[10]Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Lupus nephritis (LN) occurs in up to 50% of patients with systemic lupus erythematosus (SLE), and is a major contributor to mortality and morbidity. LN presents as a highly heterogeneous disease both in histopathology and response to therapy. The molecular and cellular processes leading to renal damage and to the heterogeneity of the disease are not well understood. To elucidate the processes underpinning the heterogeneity of LN, we applied single-cell RNA-sequencing (scRNA-seq) to renal biopsies from LN patients. Skin biopsies were evaluated as a source of biomarkers for monitoring kidney disease. Type-I interferon (IFN) response signatures were identified in tubular cells and keratinocytes, differentiating LN patients from healthy controls. Non-responders associated with higher IFN signatures in both tissue compartments. Moreover, non-response was also associated with a fibrotic signature in the tubular cells. Receptor-ligand interaction analysis indicated that the fibrotic process is likely mediated by FGF receptors with the initiating signal originating from infiltrating leukocytes. Differential expression analysis of tubular cells between proliferative and membranous LN pointed to several fibrosis-relevant pathways, which may offer insight into their histological differences. In summary, scRNA-seq was applied to LN to deconstruct its heterogeneity and provide novel targets for personalized approaches to therapy.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Systemic lupus erythematosus (SLE) is a prototypical autoimmune disease that can affect multiple organs including the heart, brain, skin, lungs, and kidneys. SLE is characterized by the production of autoreactive antibodies against nuclear antigens such as ribonucleoproteins, dsDNA, and histones[1]. Lupus nephritis (LN) affects ~50% of patients with SLE and is a major contributor to mortality and morbidity[2]. Although the exact pathogenesis has yet to be fully characterized, immune complex deposition in and along the glomerular basement membrane and in the mesangial matrix, with secondary inflammation and proliferation of mesangial and endothelial cells, are hallmarks of the disease. Additionally, hypercellularity of mesangial and endothelial cells, as well as interstitial and glomerular fibrosis, are common features of chronicity and disease progression.

WITHDRAWN
see manuscript DOI for details

These immune, inflammatory, and parenchymal cell proliferative responses of LN have visible and heterogeneous histopathologic manifestations, which can be monitored by renal biopsy and evaluated according to the International Society of Nephrology/Renal Pathology Society (ISN/RPS) 2003 Lupus Nephritis Classification System[3]. The spectrum of glomerular pathology is variable not only between patients but frequently within the same patient. Moreover, neither initial clinical manifestations nor treatment responses uniformly correlate with the histologic class of glomerular injury. Thus, clinical findings and biopsy alone are insufficient for accurate prognosis and further measures need to be developed to improve treatment and prognostic decisions. Additionally, the molecular basis for the observed histopathology is not yet fully characterized and further heterogeneity may exist, which could explain the difficulty in accurately predicting response to treatment. For instance, fibrosis has been associated with poor response to treatment, but the underlying mechanisms initiating and promoting fibrosis are not

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fully understood. A further limitation within the ISN/RPS classification system is that histologic analysis is completely based on glomerular changes, despite a growing body of literature suggesting that the tubulointerstitial space is more predictive of response to therapy and prognosis, with infiltrates and fibrosis associated with poor renal outcome[4–6].

Other potential and more accessible tissue sites than the kidney could also be exploited to obtain tissue for biomarkers of SLE progression[7]. Discovery of signatures in readily accessible tissue, such as the skin, which has immune complex-mediated inflammation analogous to that seen in the kidney, would greatly facilitate early diagnosis and treatment decisions in a much less invasive manner. Previously, we demonstrated an interferon signature in the keratinocytes from biopsies of non-lesional non-sun exposed skin of patients with LN compared to healthy controls[8]. This provides a rationale for using skin as a potential surrogate of renal disease, which could be sampled serially to follow response.

WITHDRAWN
see manuscript DOI for details

Single-cell RNA-sequencing (scRNA-seq) is an emerging transcriptomic technology resolving cell type contributions in tissues[9,10]. This technique has been applied to a number of complex renal diseases including renal cell carcinoma[11] and recently to LN by our group[8]. When resolved at a cell type level, transcriptome analysis yields valuable information regarding intercellular signaling responses and cell-type-specific pathways involved in promoting and maintaining LN. When applying this technique to renal biopsies of LN patients, we found clinically relevant prognostic markers at the time of biopsy, associated with favorable responses to treatment. Additionally, our data revealed putative intercellular interactions, which may be

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associated with signaling cascades responsible for the progression of the disease in certain patients.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Samples and data acquisition

A total of 21 renal tissue samples were collected from LN patients undergoing a clinically indicated renal biopsy (Supplemental Table 1). Of these patients, 17 also had a skin punch biopsy performed at the time of the renal biopsy. In addition to LN patients, 3 biopsy pairs of control skin and renal tissue were obtained from healthy individuals undergoing a nephrectomy for kidney transplant donation. Cell suspensions from skin and kidney biopsies of the same patient were processed on a single chip capturing about 250 cells per tissue type (Figure 1A). The cells captured per chip were sequenced at an approximate depth of 200,000 reads/cell disregarding calibrator spike reads. A total of 19,200 wells were sequenced; however, only data originating from 6,158 wells confirmed by microscopy to contain single cells and resulting in a minimum read count of 20,000 were retained for downstream bioinformatics analysis.

WITHDRAWN
see manuscript DOI for details

Cell lineage determination

We first identified the cell lineage using principle component analysis (PCA). In an iterative process, we removed cells of abnormally high or low gene counts indicating doublet cell captures or poor-quality cells, respectively, resulting in 4,008 cells entering final analysis. Dispersion and mean expression values were calculated for each gene to identify highly variable genes, which were subjected to PCA, and resulted in 11 significant principle components. t-Distributed Stochastic Neighbor Embedding (tSNE) was used to collapse the principle

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

components into two dimensions and resulted in 6 distinct clusters of cell types (Figure 1B). Differential expression analysis identified mutually exclusive sets of genes, which were characteristic of the cell lineage and frequently included established markers of cell types. The top 30 most differentially expressed markers in each cluster are provided in Supplemental Table 2. For example, tubular cells uniquely expressed *UMOD* and *SLC12A1*, whereas keratinocytes uniquely expressed *KRT1* and *KRT10.* Fibroblasts expressed many extracellular matrix proteins including *DCN*, while endothelial cells distinctly expressed *FLT1* and *PECAM1.* Leukocytes expressed distinct myeloid, T-, and B-cells genes (*CD14*, *CD3G*, and *MZB1*, respectively) yet appeared as one cell type by tSNE analysis (Figure 1C-E). Although we did not capture all known types of glomerular cell types, mesangial cells were recovered as indicated by high expression of their unique marker *TAGLN.* As anticipated, skin and renal biopsies were predominated by keratinocytes and tubular cells, respectively. The residual cell types represented smaller percentages and their relative abundance varied widely across samples (Figure 1C).

WITHDRAWN
see manuscript DOI for details

When averaged together across all renal cells, scRNA-seq expression resembled a bulk polyA-mRNA sequenced renal biopsy. Similarly, averaged skin single cells correlated with a bulk polyA-mRNA sequenced dissociated skin sample. Although averaged renal single cells also correlated with bulk sequenced skin and vice versa, they did so to a lesser extent than their originating tissue type (Figure S1).

When the keratinocyte subset identified by the first level tSNE analysis was once more subjected to tSNE analysis, the presence of a small number of sweat gland cells and melanocytes defined by *DCD* and *MLANA*, respectively, became apparent (Figure 2A-D)[12,13]. These cell types

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were excluded from participation in downstream comparative keratinocyte analysis. Similarly, the group of tubular cells identified by first level analysis was composed of various subtypes representative of the distinct nephron segments as previously reported (Figure 3A-D)[8].

LN skin and kidney epithelium indicate upregulation of type-I interferon response pathway genes

It has been shown that type-I interferons (IFN) are important in SLE in general and have been associated with disease flares in LN[14]. We previously demonstrated in a small cohort of patients that keratinocytes from LN patients show upregulation of IFN-responsive genes compared to healthy controls[8]. Here, through cumulative distribution function analysis we confirmed this observation in a separate and larger cohort of patients and further expanded this finding to tubular cells (Figure 4A). Type-I IFN pathway genes in tubular cells and keratinocytes from LN patients were significantly higher expressed than those of healthy controls as indicated by the right-shifted curve of established IFN-responsive genes compared to other genes (Figure 4A). Using the tubular expression of IFN-response genes we created an IFN response score for each patient and found that patients who did not respond to treatment had significantly higher ($p=0.04$) IFN response scores compared to those who were either partial (50% reduction in proteinuria at 6 months post biopsy) or complete responders (urine protein-to-creatinine ratio mg/mg < 0.5) (Figure 4B).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Patients non-responsive to treatment demonstrate higher expression of fibrotic extracellular matrix proteins as compared to responders

To explore pathways other than those reflective of IFN signaling in patients who did not respond to therapy, differential expression analysis was performed on the average tubular cell profiles created for each patient. This analysis identified 301 significantly ($p < 0.05$) differentially regulated genes (Figure 5A). Enrichment analysis revealed significant ($p < 0.001$) upregulation of extracellular matrix (ECM) proteins and proteins that interacted with the ECM, reflective of an active fibrotic pathway in patients that were unresponsive to therapy compared to those who responded. This expression pattern is consistent with the phenotypic change of tubular epithelial-myofibroblast transdifferentiation which is an important event that associates with progressive renal tubulointerstitial fibrosis[15,16]. Relevant to LN, tubulointerstitial fibrosis is a marker of poor prognosis[5,6] further supporting the finding of this expression in non-responders. Of clinical relevance, this gene signature may be predictive of a fibrotic response before it is measureable by standard histopathological assessment since the biopsies of some of these patients did not demonstrate fibrosis by typical scoring of tubulointerstitial damage. While it is acknowledged that ECM proteins are typically expressed by canonical fibroblasts, the cellular subset in this analysis expressed tubular cell but not fibroblast markers, supporting that this observation was not simply due to fibroblast contamination. Finally, although it is possible that fibroblasts may also play an important role in the fibrotic pathways leading to tubulointerstitial fibrosis and progressive renal insufficiency in LN, too few fibroblasts were captured to assess any potential differences in the contribution of fibroblasts between groups (data not shown).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly, two of the differentially expressed genes identified by pathway analysis as ECM interacting proteins, TIMP1 and SERPING, that were upregulated in tubular cells of patients who did not respond to treatment, have previously been shown to be pro-fibrotic and associated with renal fibrosis[17,18]. Similarly, upregulation of the complement and coagulation cascades including C1S and C1R were also noted in non-responders (Figure 5B)[19].

A similar analysis was applied to the keratinocytes of non-responders and responders to assess the possibility of monitoring pathways activated in the epithelium of a tissue distant from the site of inflammation. Pathway enrichment analysis on the differentially expressed genes from keratinocytes of patients who did not respond to treatment also demonstrated upregulation of extracellular matrix (Figure 5B). The full list of differentially expressed genes from each comparison can be found in Supplemental Table 3.

WITHDRAWN
see manuscript DOI for details

Using logistic regression analysis on fibrotic genes in the tubular cells, an equation predicting response to treatment at 6 months post biopsy was created using genes identified as fibrotic markers among the differentially expressed genes between responders and non-responders. Four genes, *COL1A1*, *COL14A1*, *COL1A2*, and *COL5A2* were found to significantly explain variance and predict response to treatment with a 92% accuracy and an area under the curve of 0.96 (Figure 5C). Correlations between response to treatment and patient demographics (race/ethnicity) were explored, but none were found (data not shown).

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fibrotic pathways in kidney may be initiated by infiltrating cell receptor-ligand interactions

Understanding the intercellular networks of communication can help elucidate potential targets for therapy in a cell-type-specific manner. scRNA-seq provides a unique starting point for deciphering ligand-receptor interactions by resolving gene expression according to cell type. Potential engagement of the highest expressed cognate receptors and ligands of cell types present in LN skin and kidney are indicated in Figure 6. Many cells in the kidney including tubular cells express various FGF receptors (FGFRs) such as *FGFR3* at high levels. FGFs and FGFRs have been implicated in fibrosis in many organs including the kidney[20]. While it has been reported that FGF can be produced by epithelium[21], in these studies *FGF13* was expressed at high levels by infiltrating leukocytes, but not other renal cell types. Additionally, tubular cells express high levels of the chemokine *CCL19* whose receptor *CCR7* is expressed within the leukocyte population. Tubular cells also expressed high levels of *TNFSF10*, potentially signaling to leukocytes which express the *TNFRSF10A* receptor.

WITHDRAWN
see manuscript DOI for details

Tubular cells and keratinocytes from patients with proliferative histologic classes compared to membranous class upregulate TNF and type-I IFN response pathways

The molecular basis for different histopathologies in LN is not completely understood. To determine if there are specific pathways involved we performed differential expression on tubular cells from LN patients with proliferative class disease (class III or class IV) and those with membranous disease (class V). This analysis excluded patients with mixed class III/V or IV/V disease. Pathway enrichment analysis of the upregulated genes in proliferative class disease

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

revealed increased type-I IFN and TNF family signaling compared with tubular cells from membranous class (Figure 7). Keratinocytes from patients with proliferative disease also showed an upregulation of several pathways including antigen presentation and response to type-I IFN compared with keratinocytes from membranous disease (Figure 7). The full list of differentially expressed genes can be found in Supplemental Table 3.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

scRNA-seq applied to kidney and skin biopsies from a cohort of LN patients and healthy controls identified clinically relevant cell-type-specific signatures associated with disease states. Small amounts of renal biopsy tissue not required for traditional histopathological evaluation by a nephropathologist thereby provided important adjunct diagnostic value. Skin biopsies were also obtained from LN patients and healthy controls, and the resulting sequencing data was mined for potential biomarkers that could be associated with clinically relevant parameters and disease states.

WITHDRAWN
see manuscript DOI for details

As previously reported we discovered an IFN response signature in both the tubular cells and keratinocytes from patients with LN compared to healthy controls, indicative of a systemic response to IFN measurable in multiple organ systems including the skin. We further found that the tubular IFN response score at the time of biopsy predicted patient response to treatment at 6 months post biopsy, and may therefore be a useful prognostic tool especially if this signature can be monitored in the skin over time.

In addition to the IFN response signature, we also identified significantly upregulated pathways associated with ECM proteins and ECM-interacting proteins indicative of a fibrotic response specifically expressed in the tubular cells and keratinocytes of patients who did not respond to treatment. Interestingly, while this signature was present in many non-responders, conventional histologic evaluation in 3 of these patients demonstrated none to only mild interstitial fibrosis. Tubular ECM protein expression has been linked to a process known as

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tubular epithelial-myofibroblast transdifferentiation during which tubular cells differentiate into myofibroblasts and begin secreting large amounts of ECM proteins[20]. This process has been linked to increased interstitial fibrosis and has implications for prognosis[22,23]. Since this pathway was found in tubular cells that did not express canonical markers of myofibroblast transformation other than *COL1A1* and *COL1A2*, we may have detected cells early in this differentiation pathway, or potentially a parallel fibrotic process amongst tubular cells that will not differentiate into myofibroblasts[24]. Accordingly, we demonstrated that our scRNA-seq could be used to create a model by which we could predict which patients would respond to treatment. Including such a diagnostic at the time of biopsy may predict which patients will need more aggressive therapy to control fibrotic scar formation leading to organ failure. Since this fibrotic signature was also present in the keratinocytes of non-responders, development of a system to monitor kidney disease using the skin as a surrogate, where repeat biopsies can be performed regularly, may prove to be a powerful prognostic tool. Pathway enrichment analysis in the kidney and skin was also able to differentiate between patients with membranous and proliferative nephritis. Proliferative nephritis had upregulation of more inflammatory pathways such as type-I IFN signaling in both skin and kidney and TNF signaling in the kidney.

WITHDRAWN
see manuscript DOI for details

By investigating the receptor ligand interactions among cell types in the skin and kidney, we identified several putative signaling interactions, which could be responsible for the association with clinical parameters. For instance, interactions were identified between infiltrating leukocytes and tubular cells through an FGF receptor, which is known to be involved in fibrotic processes and could be responsible for the upregulation of ECM and ECM-interacting proteins that were observed in the tubular cells of patients who did not respond to treatment.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Interestingly FGF receptors were highly expressed on all of the resident kidney cells, including fibroblasts, endothelial cells, and mesangial cells. Additionally, chemokines produced by resident renal cells including tubular cells, endothelial cells, and fibroblasts may be involved in the recruitment of inflammatory cells into the kidney. While validation of these interactions was not the focus of this study, such interactions provide potentially interesting and novel therapeutic targets, which may be useful in disease-state-specific treatment based on molecular diagnosis.

While our previous scRNA-seq study of renal biopsies yielded most of the dominant renal cell types, glomerular cells were absent from that analysis[8]. Using the 800-well platform markedly increased cell capture counts, and mesangial cell profiles from both healthy controls and LN patients were obtained. Podocytes, however, were not captured and a further increase in throughput by using the next-generation of droplet-based microfluidics may prove necessary to capture this rarer population of cells. Furthermore, although mesangial cells, endothelial cells, and fibroblasts were captured using this technology, their low abundance limited their differential expression analysis between patients and patient groups. Approaches increasing the number of each cell type captured would enable a similar type of analysis performed here between the major cell populations of the skin and kidney.

WITHDRAWN
see manuscript DOI for details

In summary, we have shown that scRNA-seq is a feasible and informative technique in the study of LN, despite the complexity and heterogeneity of the disease. scRNA-seq of LN tissues revealed molecular signatures clinically relevant to diagnostic and prognostic applications which could be used to meaningfully augment the current standard of care. Moreover, these

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

molecular signatures also begin to reveal some of the processes which may underlie the histologic heterogeneity of LN.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Online Methods

Procurement of clinical samples

Skin punch biopsies (1x2 mm) from non-lesional, non-sun exposed skin, and segments of 14 to 18 gauge renal needle core biopsies dispensable for clinical diagnosis (0.8x3 mm) were obtained from patients with SLE undergoing clinically indicated renal biopsies. The mean tissue mass of skin biopsies was 7 mg (2-12 mg) and that of renal biopsies was 3 mg (2-5 mg). Only renal biopsies with a pathology report indicating active LN (classes II to V, or a mixed class of III/V or IV/V), were included in this study. Comparable skin and kidney biopsies were collected from healthy control donors undergoing live kidney donation. Kidney biopsies are often standard of care for transplant departments to insure no overt pathology exists within the donor kidney. These patients also donated a small piece of skin tissue from the incision site. Both procedures are a minimal deviation from standard of care and present no significant risk to the patient. All SLE patients, healthy kidney donors, and respective recipients provided informed consent and the institutional review boards and ethics committees of Albert Einstein College of Medicine and New York University approved the sample collection. Renal biopsies were evaluated by a renal pathologist according to the ISN/RPS 2003 system for glomerular disease[3], and in addition, NIH activity and chronicity indices, which add evaluation of tubules and interstitium[25]. Patient demographics and clinical data are reported in Supplemental Table 1.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tissue dissociation and single-cell isolation

Tissue was collected at clinical sites (NYU and Einstein/Montefiore) and transported to a central technical site (Rockefeller University) within two hours of biopsy in either HypoThermasol FRS (BioLife Solutions) or Tyrode's solution. Tissue was then either immediately dissociated and processed, or if not intended for fresh processing, placed in 500 µl of Cryostor10 (StemCell) and frozen within an hour, during which samples were cooled on ice for 20 min before being placed into -80°C. Cryopreserved tissue was thawed on ice for 10-15 min directly before dissociation. Tissue was dissociated as previously described[8]. Briefly, renal and skin tissue biopsies were incubated for 15 min in a 37°C water bath in 450 µl of 0.25 mg/ml freshly prepared Liberase TL (Roche) in Tyrode's solution. Cells were collected through a 70 µm filter into FBS and stored on ice. Cells were collected by centrifugation in a 50 ml conical tube (BD) using an Eppendorf centrifuge 5804 and an A-4-44 rotor at 200 rcf for 5 min. The pelleted cells were resuspended in 100 µl of Tyrode's solution. The cell numbers and viability was determined using a Biorad TC20 automated cell counter and Trypan blue staining. The concentration of cells in suspension ranged from 20,000-1,000,000 cells/ml, but was typically near 300,000 cells/ml. If a large amount of debris was detected by microscopy and BioRad TC20 cell counting, the cell pellet was suspended in 1 ml of Tyrode's solution and recollected at 220 rcf for 3 min. Cell suspensions were either diluted or concentrated by centrifugation and subsequent resuspension in a smaller volume targeting a final concentration of 200,000 cells/ml in Tyrode's solution. A minimum of 10 µl of cell suspension was necessary to proceed with scRNA-seq and loading of one of the two 400-well partitions of the C1 HT microfluidic chip. Viability typically ranged between 20-60% at this step.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Single-cell capture, cDNA library preparation, barcoding, and sequencing

Single-cell suspensions at a concentration of 200,000 cells/ml and no less than 2,000 cells were loaded into a medium 10-17 µm diameter C1 HT 800-well integrated microfluidic chip (IFC) (Fluidigm) and processed according to the Fluidigm C1 HT protocol revision A using the recommended standard mRNA-seq reagents and program. The chip is divided into two 400-well sections allowing for loading of skin and kidney samples matched by individual on the same chip. Per manufacturer's recommendation, 7 µl of 10% PBS Tween was added to the valve fluid to reduce surface tension. The occupation of single-cell capture sites was verified using a Zeiss Axiovert 200 inverted microscope averaging 400 single cell captures per 800-well chip. The captured cells were lysed, polyA mRNA was reverse transcribed, and cDNA pre-amplified using the SMARTer Ultra Low RNA kit (Clontech) in the Fluidigm C1 Single-Cell Auto Prep system. To monitor cDNA library conversion, a cocktail of synthetic RNA spikes #1, #4, and #7 of the Ambion ArrayControl RNA Spikes (ThermoFisher) was prepared as described in the Fluidigm C1 protocol and added to the lysis reaction (Mix A) for each experiment. After the initial rounds of PCR on the microfluidic chip, products were harvested, exonuclease-treated, and further amplified according to the Fluidigm protocol.

WITHDRAWN
see manuscript DOI for details

Pre-amplified cDNA libraries were tagmented and barcoded using the Nextera XT Library Preparation Kit (Illumina) with indexing according to the Fluidigm C1 HT protocol revision A. An enrichment primer to select for the 3' ends was added during this step. PCR-products originating from up to 800 cells per chip were pooled together using the 20 barcodes

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

recommended by Fluidigm, and sequenced paired-end using the Illumina NextSeq500. Read 1 was sequenced 30 cycles and Read 2 120 cycles.

Bioinformatic analysis

Single FASTQ files corresponding to up to 800 cells were demultiplexed into 20 FASTQ files by separating reads based on the Illumina Nextera index primers. Each of the 20 FASTQ files represents a single column (up to 40 cells) on the Fluidigm C1 HT IFC and was further demultiplexed into single-cell FASTQ files using a Perl script provided by Fluidigm. Resulting FASTQ files were then trimmed using cutadapt (version 1.12) in nextseq mode followed by polyA trimming[26]. FASTQ file reads were aligned to the human reference genome GRCh38 downloaded from Ensembl using the STAR aligner (version 2.5.0a) allowing up to 2 mismatches to the reference sequence and keeping directionality[27]. The reference genome only contains the canonical chromosomes and non-chromosomal contigs; haplotypes were excluded. Uniquely mapped reads to the reference genome were counted using featureCounts (version 1.5.0) and the reads mapping to the human genome were collapsed on the gene level[28]. Transcripts from the Havana database were removed from the Ensembl 83 GTF as they frequently overlapped with older gene annotation leading to multimapping. Transcripts identified in both the Ensembl and Havana databases were kept in the GTF file and annotated as ensembl_havana. The pipeline was run on RedHat Linux or MacOS 10.10.3. A chip-dependent and low-frequency cross-well RNA or DNA contamination was encountered, requiring background subtraction to correct the count matrices. Briefly, the read counts observed in empty wells of each chip were averaged and

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subtracted from each well of the same chip. Additional details of this step are provided in supplementary methods.

PCA and tSNE analysis

Principal component analysis and t-Distributed Stochastic Neighbor Embedding (tSNE) were performed using the Seurat package (version 2.2.1) for R[29]. The count matrices were depth-normalized to 100,000 reads and used to identify the set of genes that was most variable across datasets. We used a z-score cutoff of 0.1 to identify 2099 highly variable genes. In this analysis, all genes were evaluated for variability. Highly expressed ubiquitous genes such as mitochondrially encoded or nuclear-genome-encoded ribosomal proteins were excluded for clustering. The highly variable genes discovered by this process were loaded into a principal component analysis (PCA), which yielded 11 significant principal components and provided the input for tSNE visualization[29].

WITHDRAWN
see manuscript DOI for details

Receptor/ligand analysis

Lists of potential ligand-receptor pairs were obtained and manually curated from the Database of Interacting Proteins (http://dip.doe-mbi.ucla.edu) and the IUPHAR/BPS guide to pharmacology (http://www.guidetopharmacology.org) as described previously[30]. The list of interactions was intersected with CPM-normalized expression of genes for each cell type. A ligand-receptor interaction was considered active if the receptor was above 45 TPM and the cognate ligand above 65 TPM expression thresholds. The thresholds were selected to capture up

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to 100 interactions. The identification of interacting pairs used custom Perl scripts, which averaged and normalized gene expression within each cell type and checked gene expression values against thresholds and the list of interaction pairs. Custom Perl and R scripts were used to draw interaction diagrams.

Interferon score

Interferon scores were calculated using a set of 212 experimentally derived type-I IFN responsive genes as previously described[8]. Briefly, average cell-type-specific expression profiles per patient were created and the IFN responsive genes were subsetted and averaged excluding genes with a 0 value across all patients.

WITHDRAWN
see manuscript DOI for details

Cumulative distribution analysis

To visualize IFN-response signatures, we used the ratios between LN and control for two gene sets of the IFN-responsive genes and ubiquitously expressed genes mentioned above. For each group of ratios, we calculated the cumulative distribution function (CDF) and estimated the statistical significance of the difference between two distributions using the Wilcoxon signed rank test using standard functions from the R statistical package.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical analysis

The differential expression analysis was performed using DESeq2 (version 1.10.1) and R (version 3.4.2)[31]. Patient- and cell-type-specific expression profiles were created by averaging the expression across all cells of the same cell type for each patient creating a pseudo-bulk RNA-seq expression matrix for differential expression analysis. Briefly, expression count matrices were fit to a generalized linear model per gene following a negative binomial distribution. Dispersion estimates for each gene within groups were shrunk using an empirical Bayesian approach using default DESeq2 parameters. Log_2 fold changes were compared between disease groups using the Wald test. Pathway enrichment analyses were performed by enrichR[32,33] using the Reactome 2016[34,35] and KEGG 2016[36–38] pathway databases. Logistic regression analysis was performed using R, fitting, and 'pscl' (version 1.5.2) and 'ROCR' (version 1.0-7) packages for accuracy and area under the curve analysis.

Cumulative distribution functions were compared with a Mann-Whitney *U* test. Differences between groups were compared using a two-tailed Student's t test with a p value less than 0.05 considered significant.

Code availability

All software packages and programs used are publically available and open source. Scripts used to analyze the data with these packages are available in the supplementary materials.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Data availability

Raw and processed data will be available from dbGAP, Accession number to be determined. Currently raw and processed data for review is available from Immport (https://aspera-immport.niaid.nih.gov/aspera/user/?B=%2FAMP_RA_SLE.Phase1%2FSLE%2FRNASeq_C1.Tuscl) (username: immport-upload15, password: 15@1mmp0rt).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Tsokos, G. C. Systemic lupus erythematosus. *N. Engl. J. Med.* **365,** 2110–2121 (2011).
2. Mohan, C. & Putterman, C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat. Rev. Nephrol.* **11,** 329–341 (2015).
3. Weening, J. J. *et al.* The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J. Am. Soc. Nephrol. JASN* **15,** 241–250 (2004).
4. Yu, F. *et al.* Tubulointerstitial lesions of patients with lupus nephritis classified by the 2003 International Society of Nephrology and Renal Pathology Society system. *Kidney Int.* **77,** 820–829 (2010).
5. Hsieh, C. *et al.* Predicting outcomes of lupus nephritis with tubulointerstitial inflammation and scarring. *Arthritis Care Res.* **63,** 865–874 (2011).
6. Alsuwaida, A. O. Interstitial inflammation and long-term renal outcomes in lupus nephritis. *Lupus* **22,** 1446–1454 (2013).
7. Misra, R. & Gupta, R. Biomarkers in lupus nephritis. *Int. J. Rheum. Dis.* **18,** 219–232 (2015).
8. Der, E. *et al.* Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight* **2,** pii: 93009. doi: 10.1172/jci.insight.93009 (2017).
9. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33,** 495–502 (2015).
10. Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C. & Yuan, J. Single-cell RNA sequencing: unraveling the brain one cell at a time. *Trends Mol. Med.* **23,** 563–576 (2017).
11. Kim, K.-T. *et al.* Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol.* **17,** 80 (2016).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Schittek, B. *et al.* Dermcidin: a novel human antibiotic peptide secreted by sweat glands. *Nat. Immunol.* **2,** 1133–1137 (2001).

13. Du, J. *et al.* MLANA/MART1 and SILV/PMEL17/GP100 are transcriptionally regulated by MITF in melanocytes and melanoma. *Am. J. Pathol.* **163,** 333–343 (2003).

14. Elkon, K. B. & Stone, V. V. Type I interferon and systemic lupus erythematosus. *J. Interferon Cytokine Res.* **31,** 803–812 (2011).

15. Lan, H. Y. Tubular epithelial-myofibroblast transdifferentiation mechanisms in proximal tubule cells. *Curr. Opin. Nephrol. Hypertens.* **12,** 25–29 (2003).

16. Ng, Y. Y. *et al.* Tubular epithelial-myofibroblast transdifferentiation in progressive tubulointerstitial fibrosis in 5/6 nephrectomized rats. *Kidney Int.* **54,** 864–876 (1998).

17. Zhang, X. *et al.* TIMP-1 promotes age-related renal fibrosis through upregulating ICAM-1 in human TIMP-1 transgenic mice. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61,** 1130–1143 (2006).

18. Ling, X. B. *et al.* Integrative urinary peptidomics in renal transplantation identifies biomarkers for acute rejection. *J. Am. Soc. Nephrol.* **21,** 646–653 (2010).

19. Su, Z. *et al.* Excessive activation of the alternative complement pathway in autosomal dominant polycystic kidney disease. *J. Intern. Med.* **276,** 470–485 (2014).

20. Strutz, F. *et al.* Basic fibroblast growth factor expression is increased in human renal fibrogenesis and may mediate autocrine fibroblast proliferation. *Kidney Int.* **57,** 1521–1538 (2000).

21. Smith, E. R., Tan, S.-J., Holt, S. G. & Hewitson, T. D. FGF23 is synthesised locally by renal tubules and activates injury-primed fibroblasts. *Sci. Rep.* **7,** 3345 (2017).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. Yang, J. & Liu, Y. Blockage of tubular epithelial to myofibroblast transition by hepatocyte growth factor prevents renal interstitial fibrosis. *J. Am. Soc. Nephrol. JASN* **13,** 96–107 (2002).

23. Meran, S. & Steadman, R. Fibroblasts and myofibroblasts in renal fibrosis. *Int. J. Exp. Pathol.* **92,** 158–167 (2011).

24. Yang, J. & Liu, Y. Dissection of key events in tubular epithelial to myofibroblast transition and its implications in renal interstitial fibrosis. *Am. J. Pathol.* **159,** 1465–1475 (2001).

25. Austin, H. A., Boumpas, D. T., Vaughan, E. M. & Balow, J. E. Predicting renal outcomes in severe lupus nephritis: contributions of clinical and histologic data. *Kidney Int.* **45,** 544–550 (1994).

26. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

27. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma.* **29,** 15–21 (2013).

28. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma.* **30,** 923–930 (2014).

29. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161,** 1202–1214 (2015).

30. Pavličev, M. *et al.* Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. *Genome Res.* **27,** 349–361 (2017).

31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** (2014).

32. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

33. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44,** W90-97 (2016).

34. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42,** D472-477 (2014).

35. Fabregat, A. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **46,** D649–D655 (2018).

36. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

37. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45,** D353–D361 (2017).

38. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44,** D457-462 (2016).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

This work was supported by the Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus Network. AMP is a public-private partnership (AbbVie, Arthritis Foundation, Bristol-Myers Squibb, Foundation for the National Institutes of Health, Lupus Foundation of America, Lupus Research Alliance, Merck Sharp & Dohme, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer, Rheumatology Research Foundation, Sanofi, and Takeda Pharmaceuticals) created to develop new ways of identifying and validating promising biological targets for diagnostics and drug development. Funding was provided through grants from the National Institutes of Health (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-AR067685, UH2-AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-AR067694, and UM2-AR067678). We thank the Rockefeller University Genomics Resource Center for providing access to the Fluidigm C1 system and Illumina sequencing.

WITHDRAWN
see manuscript DOI for details

Competing interests

The authors declare no competing interests.

Author contributions

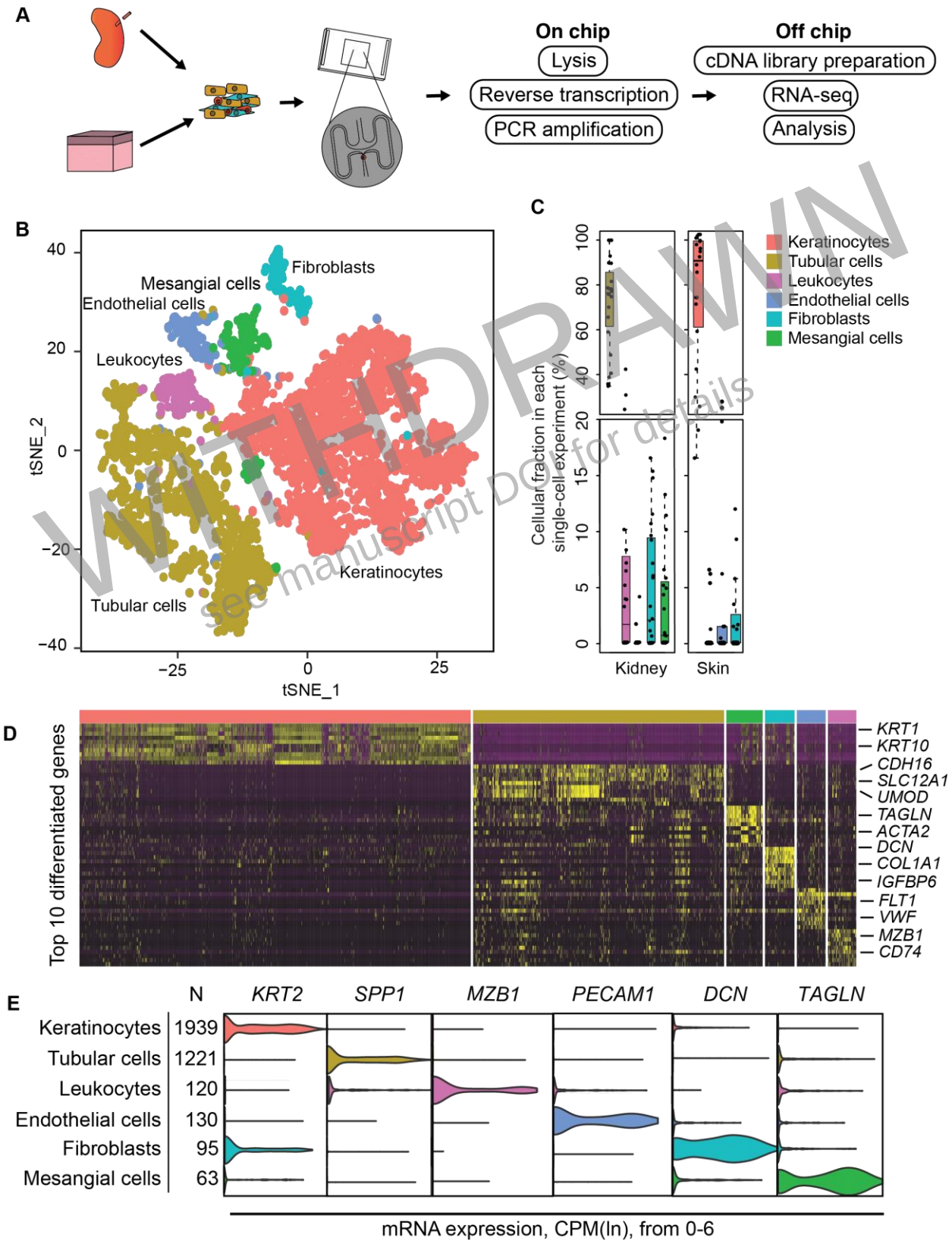
JB, TT, and CP conceived the study with help from SR, JJ and JG. Input regarding the skin came from RC and HMB. ED, HS, and SR performed all biopsy dissociations and single-cell

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

experiments. BG, PI, HMB, MK, MM, NJ, NB, and ES assisted with patient consent and sample acquisition of LN biopsies. HR, JR, JG assisted with patient consent and sample acquisition of live kidney donor tissue. Renal biopsy histology was evaluated by MW and JP. HMB and PI performed all skin biopsies. Analysis was performed by ED, HS, PM, and MK. ED, JB, TT and CP prepared and wrote the manuscript.

WITHDRAWN
see manuscript DOI for details

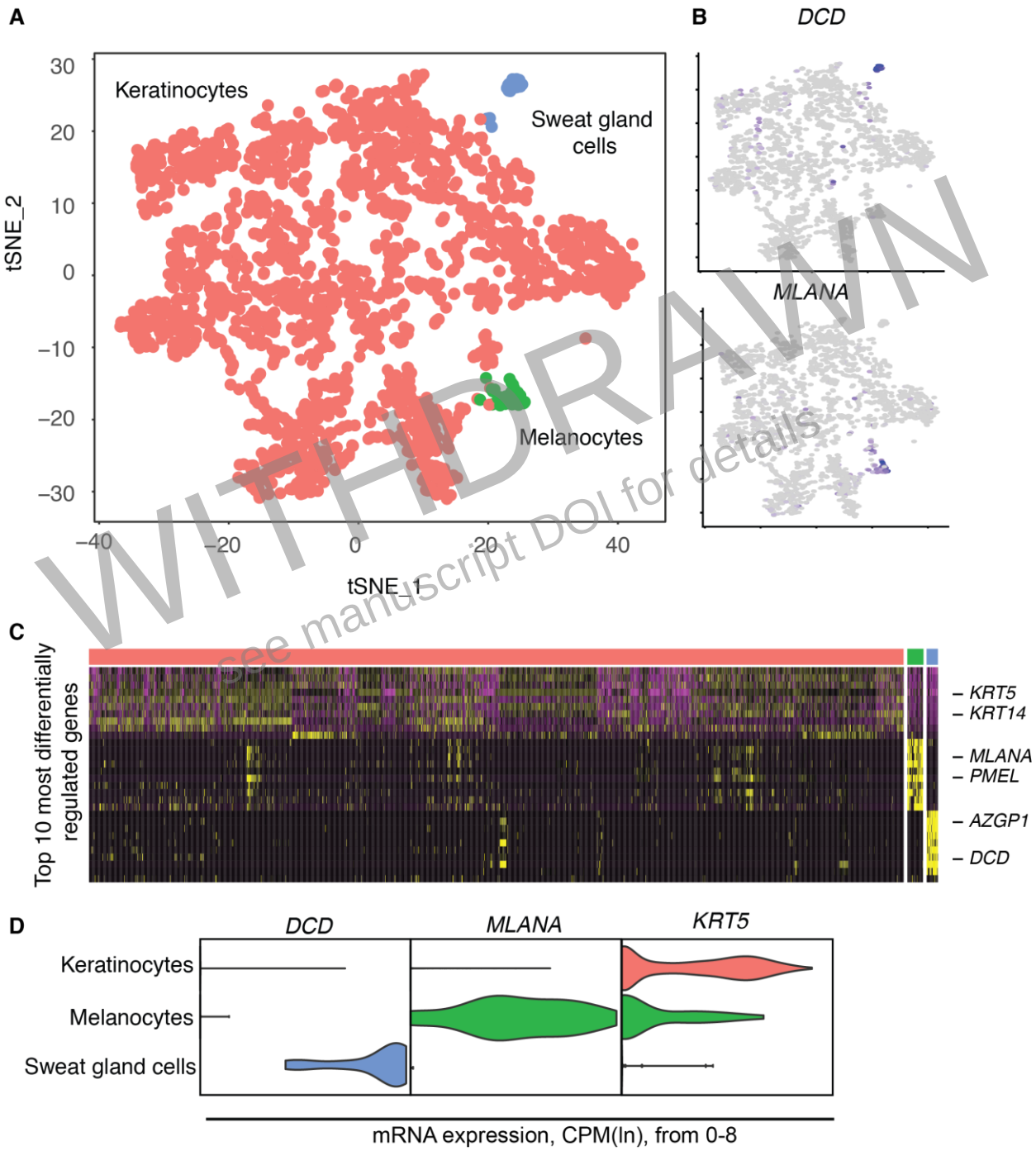
bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

Figure 1.

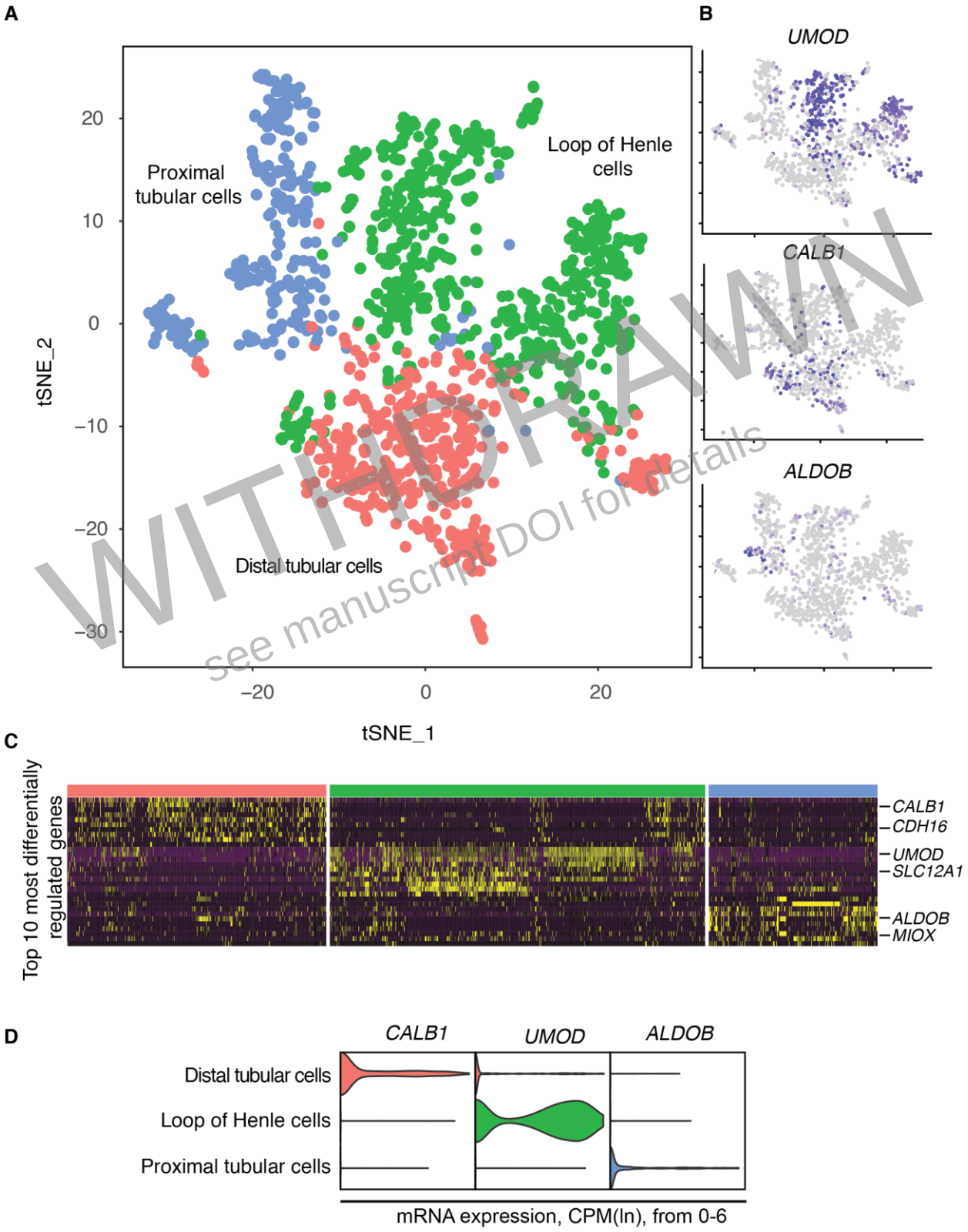
bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

Figure 2.

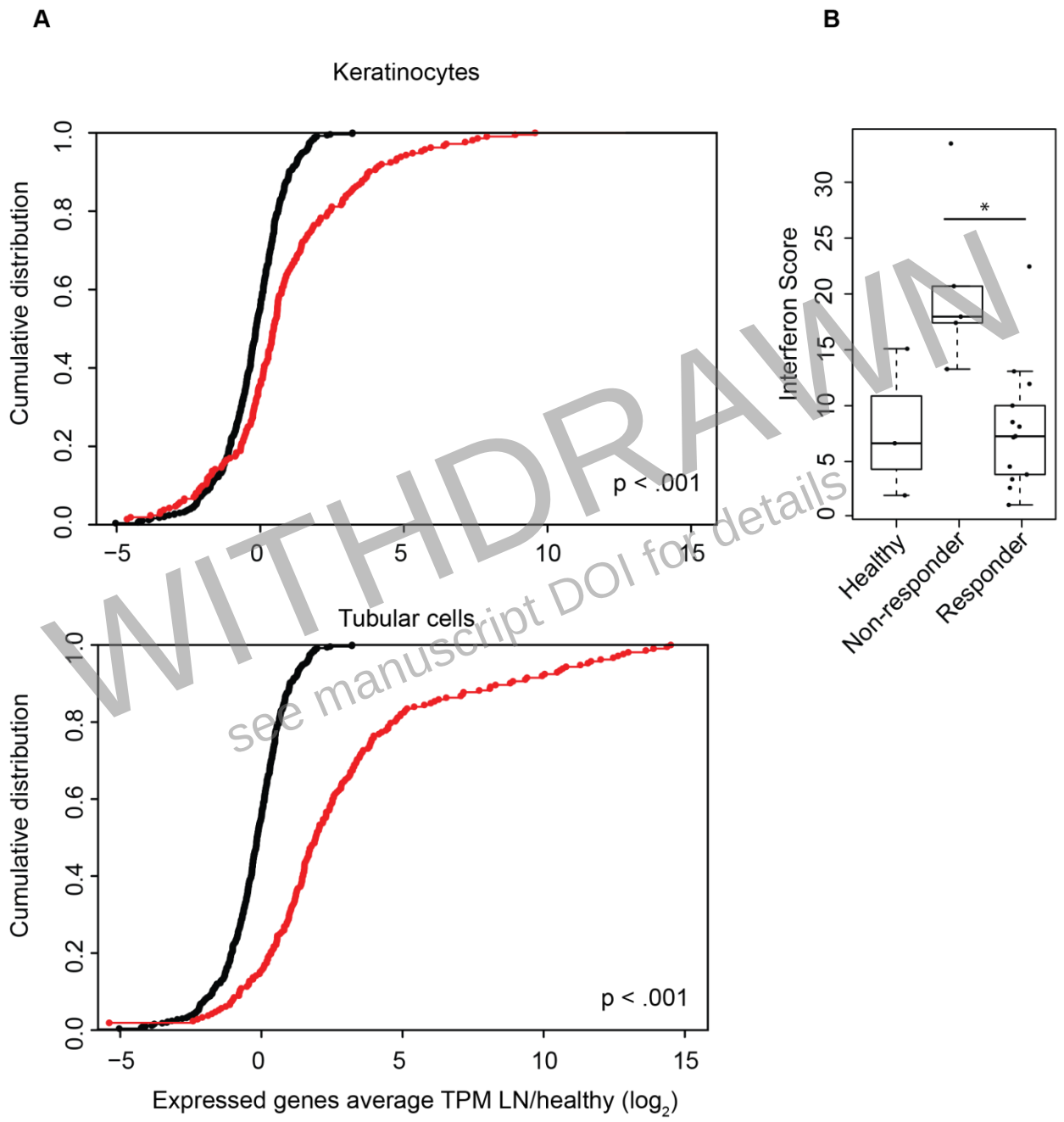
bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

Figure 3.

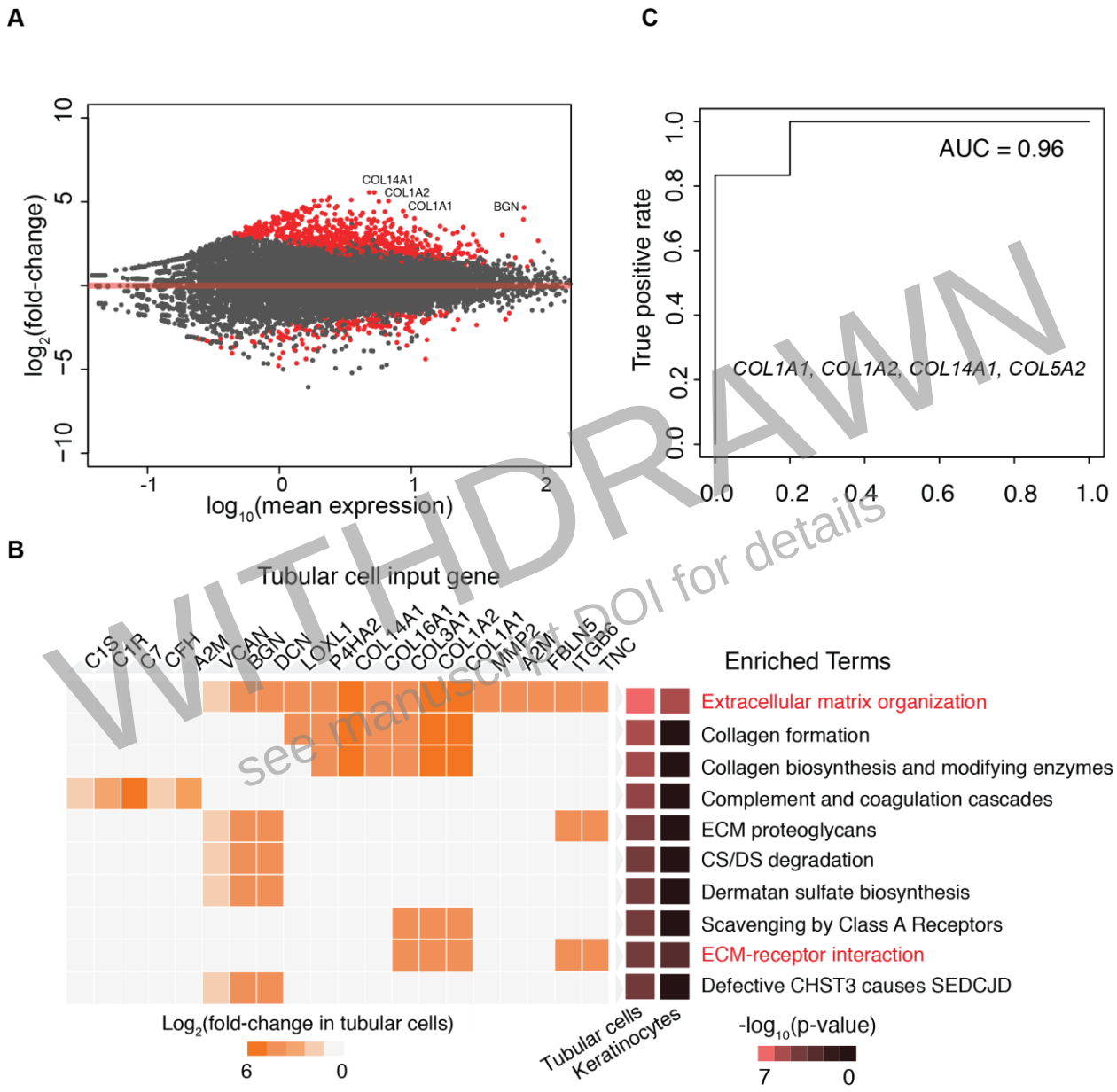
bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

Figure 4.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

Figure 5.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

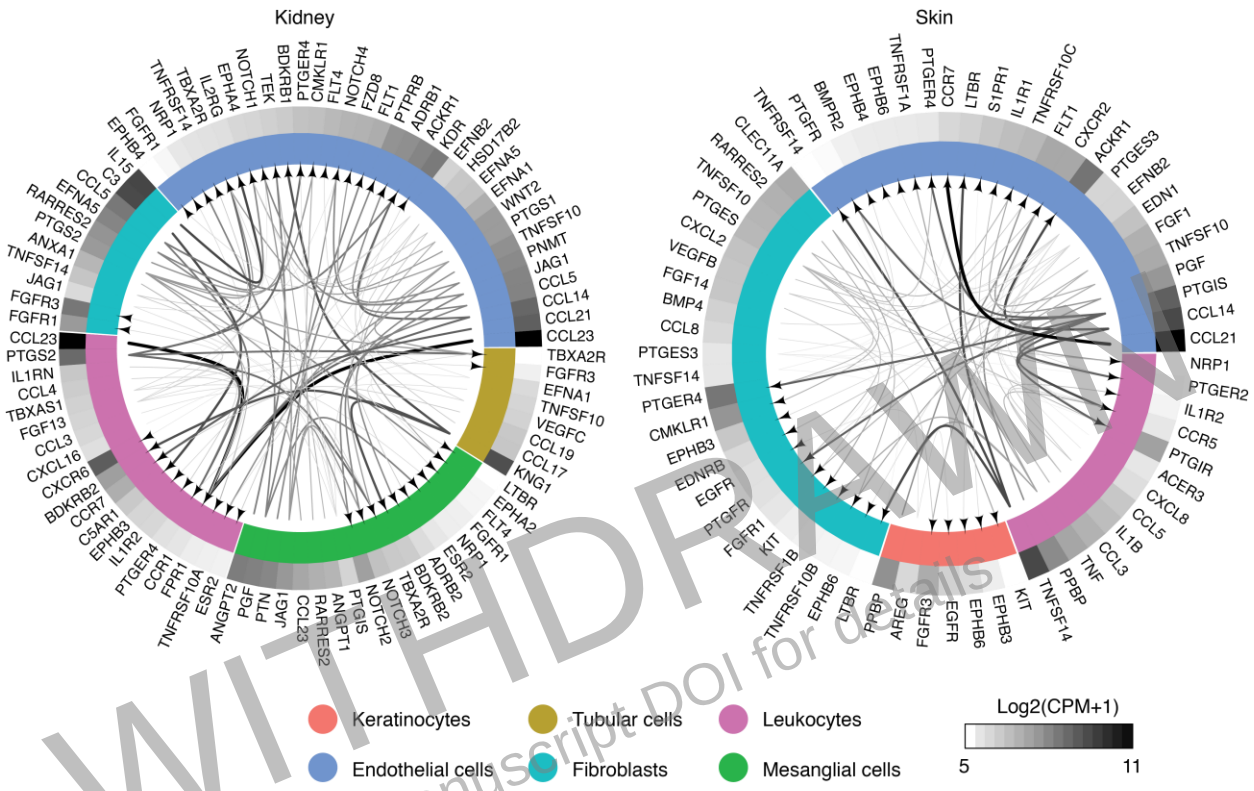
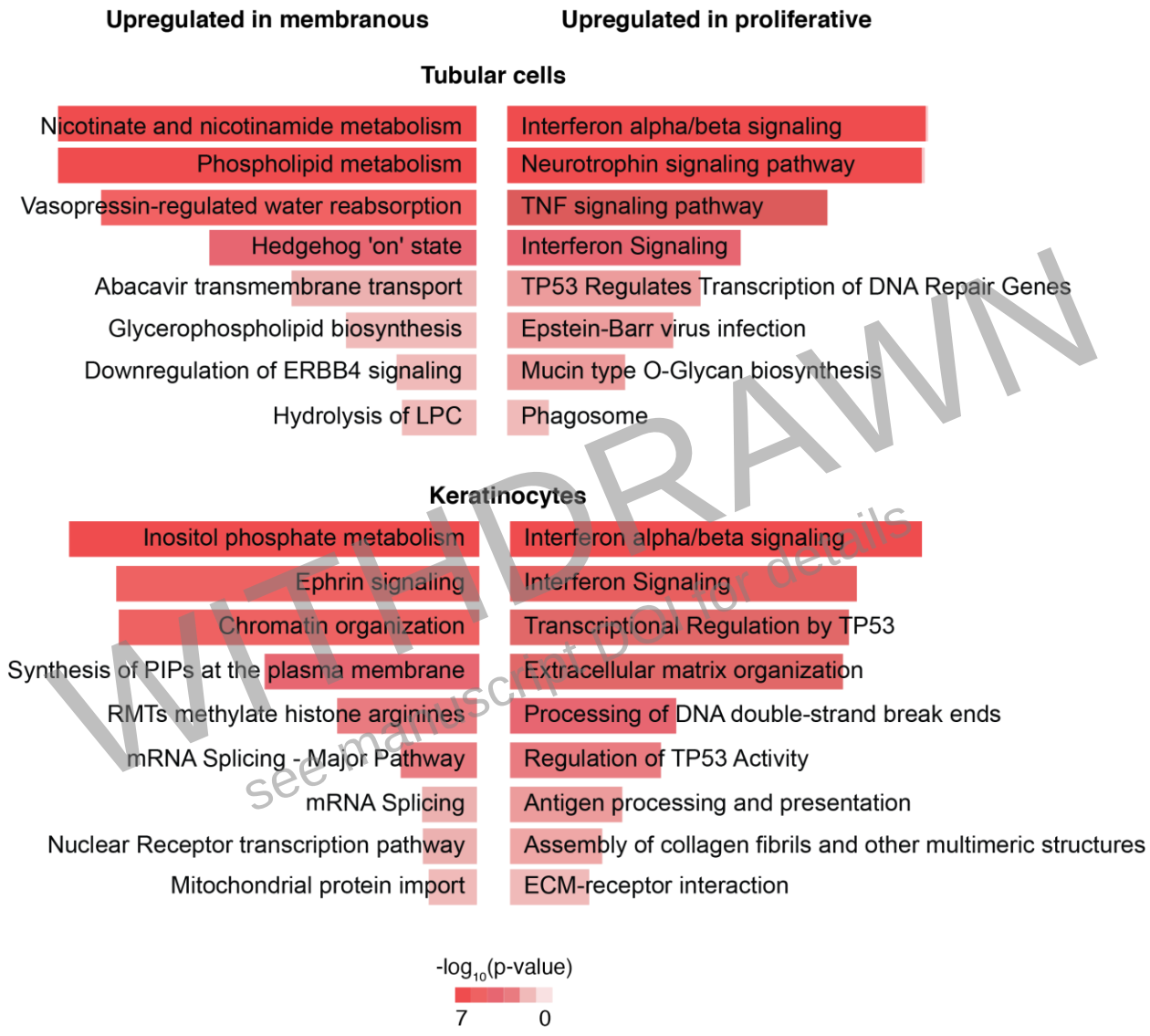


Figure 6.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript for details

Figure 7.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends

Figure 1. Cell lineage determination by dimensionality reduction analysis. A) Schematic of the scRNA-seq pipeline. Skin (n=17) and kidney (n=21) samples from LN patients or healthy controls (n=3) were collected at the time of clinically indicated renal biopsy or live kidney donation, respectively. Skin and kidney biopsies were enzymatically disaggregated into single cell suspensions and loaded onto a microfluidic device. Single cell captures were assessed via microscopy. B) t-Distributed Stochastic Neighbor Embedding (tSNE) clustering of 4,008 single cells. PCA identified six major clusters of cells from both skin and kidney biopsies. Cells are color-coded by an algorithm for determining expression clusters and cell types. C) Box plot of the percent contribution of each cluster from skin and kidney biopsies. Boxes are colored by cluster corresponding to Figure 1B. D) Heatmap of the top 10 most differentially expressed genes in each cluster to identify mutually exclusive gene sets used to determine the cell lineage of each cluster. Each row is a differentially expressed gene and each column is a single-cell organized by cluster identity. Transcript abundance ranges from low (purple) to high (yellow). E) Violin plot of selected markers indicating the expression level of canonical markers within each cluster.

WITHDRAWN
see manuscript DOI for details

Figure 2. Subclustering of keratinocytes reveals two rare skin-specific cell types. A) tSNE plotting of 1,939 keratinocytes identified in initial clustering analysis colored by cluster identifying algorithm with cell type labels next to each cluster. B) Expression of *DCD* and *MLANA*, markers of the new clusters, amongst clustered cells from low expression (grey) to high expression (dark blue). C) Heatmap of the differentially expressed genes between each identified cluster. D) Violin plot of *MLANA* and *DCD*, markers of melanocytes and sweat gland cells, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Subclustering of tubular cells identifies major tubular cell subtypes of the nephron. A) tSNE plotting of 1,221 tubular cells identified by initial clustering analysis. Three clusters of tubular cells are identified and colored by clustering algorithm with labels of putative cluster identity indicated next to each cluster. B) Expression of established tubular subtype marker within each cluster from low expression (grey) to high expression (dark blue). C) Heatmap of the top 10 most differentially expressed genes between each cluster. D) Violin plot of canonical tubular cell subset markers used to determine the tubular cell subsets.

WITHDRAWN
see manuscript DOI for details

Figure 4. IFN response signature differentiates LN patients from healthy controls and response to treatment. A) Cumulative distribution function (CDF) of the ratio of expression of 212 IFN responsive genes (red line) or ubiquitously expressed genes (black line) in both tubular cells (top) and keratinocytes (bottom) ($p<0.001$). B) Boxplot of IFN response scores in healthy controls (n=3), patients who responded to treatment (n=13), and patients who did not respond to treatment (n=5) ($p=0.0003$, $t=4.7234$). Box indicates the first quartile, median, and third quartile. Whiskers indicate the highest and lowest values.

Figure 5. A fibrotic gene signature as a potential prognostic marker for patients non-responsive to treatment. A) MA plot of differential expression analysis performed between tubular cells of patients responsive (n=13) or non-responsive to treatment (n=5). Significantly differentially expressed genes are colored in red. B). Pathway enrichment analysis of genes identified as upregulated in patients non-responsive to treatment in Figure 5A. $-Log_{10}$(p-value) of each pathway is shown for both keratinocytes and tubular cells colored from least significant (black) to most significant (red). Log_2 fold change in gene expression between patients non-

bioRxiv preprint doi: https://doi.org/10.1101/382846; this version posted August 1, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

responsive to treatment compared with patients responsive to treatment in each pathway are indicated for tubular cells from smallest (grey) to highest (orange). C) Receiver operating characteristic curve of the logistic regression equation of differentially expressed fibrotic genes, *COL1A2*, *COL1A1*, *COL14A1*, *COL5A2*, with area under the curve (AUC) indicated.

Figure 6. Putative receptor-ligand interactions between kidney and skin cells. Lines represent interactions between cell types in the skin and the kidney. Lines originate at the ligand and connect to its receptor as indicated by the arrowhead. Each cell type is color coded and represented by that color in each organ. Only receptors with expression above 45 CPM and ligand above 65 CPM within each cell type are shown. Receptors and ligand are arranged by expression strength clockwise from lowest expression to highest. Ligands or receptors without a cognate pair were excluded for visualization.

WITHDRAWN
see manuscript DOI for details

Figure 7. Differential expression and pathway enrichment analysis of tubular cells and keratinocytes between membranous and proliferative LN. Significantly enriched pathways in both tubular cells and keratinocytes in membranous (n=6) and proliferative (n=8) LN are indicated. Mixed class III/V or IV/V were excluded from this analysis. Color intensity and length of bar indicates higher -Log_{10}(p-value) from least significant (white) to most significant (red).