

Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis

Evan Der, Saritha Ranabothu, Hemant Suryawanshi, Kemal M. Akat, Robert Clancy, Pavel Morozov, Manjunath Kustagi, Mareike Czuppa, Peter Izmirly, H. Michael Belmont, Tao Wang, Nicole Jordan, Nicole Bornkamp, Janet Nwaukoni, July Martinez, Beatrice Goilav, Jill P. Buyon, Thomas Tuschl, Chaim Putterman

JCI Insight. 2017;2(9):e93009. <https://doi.org/10.1172/jci.insight.93009>.

Research Article

Nephrology

Lupus nephritis is a leading cause of mortality among systemic lupus erythematosus (SLE) patients, and its heterogeneous nature poses a significant challenge to the development of effective diagnostics and treatments. Single cell RNA sequencing (scRNA-seq) offers a potential solution to dissect the heterogeneity of the disease and enables the study of similar cell types distant from the site of renal injury to identify novel biomarkers. We applied scRNA-seq to human renal and skin biopsy tissues and demonstrated that scRNA-seq can be performed on samples obtained during routine care. Chronicity index, IgG deposition, and quantity of proteinuria correlated with a transcriptomic-based score composed of IFN-inducible genes in renal tubular cells. Furthermore, analysis of cumulative expression profiles of single cell keratinocytes dissociated from nonlesional, non-sun-exposed skin of patients with lupus nephritis also revealed upregulation of IFN-inducible genes compared with keratinocytes isolated from healthy controls. This indicates the possible use of scRNA-seq analysis of skin biopsies as a biomarker of renal disease. These data support the potential utility of scRNA-seq to provide new insights into the pathogenesis of lupus nephritis and pave the way for exploiting a readily accessible tissue to reflect injury in the kidney.

Find the latest version:

<https://jci.me/93009/pdf>



Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis

Evan Der,¹ Saritha Ranabothu,² Hemant Suryawanshi,³ Kemal M. Akat,³ Robert Clancy,⁴ Pavel Morozov,³ Manjunath Kustagi,³ Mareike Czuppa,³ Peter Izmirly,⁴ H. Michael Belmont,⁴ Tao Wang,¹ Nicole Jordan,¹ Nicole Bornkamp,⁴ Janet Nwaukoni,⁴ July Martinez,¹ Beatrice Goilav,² Jill P. Buyon,⁴ Thomas Tuschl,³ and Chaim Putterman¹

¹Albert Einstein College of Medicine, Bronx, New York, USA. ²Pediatric Nephrology, Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, New York, USA. ³Howard Hughes Medical Institute and Laboratory for RNA Molecular Biology, The Rockefeller University, New York, New York, USA. ⁴New York University School of Medicine, New York, New York, USA.

Lupus nephritis is a leading cause of mortality among systemic lupus erythematosus (SLE) patients, and its heterogeneous nature poses a significant challenge to the development of effective diagnostics and treatments. Single cell RNA sequencing (scRNA-seq) offers a potential solution to dissect the heterogeneity of the disease and enables the study of similar cell types distant from the site of renal injury to identify novel biomarkers. We applied scRNA-seq to human renal and skin biopsy tissues and demonstrated that scRNA-seq can be performed on samples obtained during routine care. Chronicity index, IgG deposition, and quantity of proteinuria correlated with a transcriptomic-based score composed of IFN-inducible genes in renal tubular cells. Furthermore, analysis of cumulative expression profiles of single cell keratinocytes dissociated from nonlesional, non-sun-exposed skin of patients with lupus nephritis also revealed upregulation of IFN-inducible genes compared with keratinocytes isolated from healthy controls. This indicates the possible use of scRNA-seq analysis of skin biopsies as a biomarker of renal disease. These data support the potential utility of scRNA-seq to provide new insights into the pathogenesis of lupus nephritis and pave the way for exploiting a readily accessible tissue to reflect injury in the kidney.

Introduction

Systemic lupus erythematosus (SLE) is a prototypic autoimmune disease variously affecting the skin, kidneys, joints, serosal linings, CNS, and hematopoietic cells. Characteristic of this disease is the production of autoantibodies to nuclear antigens such as ribonucleoproteins, dsDNA, and histones. Autoantibody production by self-reactive B cells is further amplified by the activation of the innate immune system and secretion of type I IFN encoded by the *IFNA* and *IFNB* gene families (1–3). Renal involvement in SLE occurs in about half of the patients and is one of the leading causes of morbidity and a significant contributor to mortality (1). While the pathogenesis is not yet fully characterized, antibody and complement protein deposition in the kidney accompanying and/or triggering inflammation are hallmarks of lupus nephritis (LN) (2).

Renal injury in LN does not manifest as one uniform entity. Based on histologic analysis of renal core biopsies, the International Society of Nephrology/Renal Pathology Society (ISN/RPS) classification system categorizes a spectrum of glomerular pathology ranging from classes I to V (4). Since the extent of glomerular involvement can frequently vary even within the same biopsy, averages are used to arrive at a specific diagnosis (4), which may in part be the reason not all patients with the same ISN/RPS class respond similarly to treatment. Other histological indices including the NIH activity index (AI) and chronicity index (CI) (5) have been developed in an attempt to rectify this, but they also may have poor prognostic associations. While a higher NIH CI (6) is associated with a greater risk for the progression to end-stage renal disease, it is unclear why patients with low chronicity can still have a poor response to conventional therapy (7). The NIH AI is likewise not uniformly predictive of renal progression (6), justifying the continued search for candidate biomarkers. While much effort has been made to predict outcomes of therapies and general prognosis based on histological classification, these prognostic features based on histology often suffer from modest reproducibility and low sensitivity (6, 8). Despite these shortcomings, renal

Conflict of interest: The authors have declared that no conflict of interest exists.

Submitted: January 23, 2017

Accepted: March 29, 2017

Published: May 4, 2017

Reference information:

JCI Insight. 2017;2(9):e93009. <https://doi.org/10.1172/jci.insight.93009>.

Table 1. Demographics of SLE patients and healthy controls

	SLE skin biopsy (n = 12)	SLE kidney biopsy (n = 16)	Healthy skin biopsy (n = 5)
Age, years	30	29.3	39.3
Ethnicity, %			
Hispanic	25	31.3	0
Nonhispanic	75	68.7	100
Race, %			
Caucasian	25	18.8	40
Asian	16.7	25	0
African American	58.3	56.2	60
Sex, %			
Female	83.3	93.8	80
Male	16.7	6.2	20
Disease duration, years	6.4	4.7	
SELENA-SLEDAI score	13.1	12.7	

biopsy remains the “gold standard” upon which many therapeutic decisions are made. Although there have been intense efforts to identify biomarkers in serum and urine from LN patients, these have not yet yielded sufficiently robust markers to replace the renal biopsy. Previous studies have shown that endothelial changes such as increased protein levels of membrane protein C receptor (encoded by *PROCR*) in patients with LN predicts poor responses to therapy (9). Furthermore, increased levels were noted in biopsies of nonlesional, non-sun-exposed skin of LN patients, suggesting that alterations in the microvasculature are widespread and extend to the dermal vasculature (9). Thus, analysis of this more readily accessible tissue, even distant from the primary affected organ, may provide an opportunity to explore surrogates for renal tissue analyses (10).

RNA sequencing (RNA-seq) is an important method defining the transcriptomes from total RNA extracted from whole tissues, cell populations, and more recently even single cells, providing detailed gene expression and sequence variation information (11). Recent advances in rapid yet gentle tissue dissociation and microfluidics have improved standardization and commercialization of the single cell RNA-seq (scRNA-seq) technology (12). This new methodology has been applied in many fields, including in cancer where the analysis of the heterogeneity of tumor cells suggested potentially more effective therapeutic targeting approaches specifically directed at pathogenic tumor lineages (13, 14). Recently, scRNA-seq has been applied to renal carcinoma cells to differentiate between primary and metastatic tumor cells (15). Similarly, we applied this technology to LN tissue in order to characterize gene expression at the single cell level and resolve differential responses by cell type, in order to understand how activated and quiescent, resident (i.e., tissue-constituting), and/or infiltrating cellular subpopulations contribute to disease initiation, maintenance, and progression. We demonstrate for the first time to our knowledge that scRNA-seq can be applied to human renal biopsy tissue obtained for clinical purposes; provide insights into how to dissect the heterogeneity of the disease; and identify biomarkers and pathways driving disease pathogenesis.

Results

Single cell dissociation of human tissue biopsies of skin and kidney. A total of 5 and 12 skin samples from healthy controls and patients with LN, respectively, and 16 kidney samples from LN patients (Table 1) were processed. Four of the kidney samples were from the same patients as four of the skin samples. Tissue samples were collected at 2 clinical sites and were transported in Tyrode's solution or HypoThermosol on ice to a central technical site within 2 hours. Samples were then enzymatically and mechanically disaggregated, yielding approximately 5,000 viable cells/mg tissue from skin and 15,000 viable cells/mg from kidney. A small subset of tissue samples was first cryopreserved and then processed, but no differences were seen in the number of viable cells obtained per biopsy. Additionally, a sample of viable CD4⁺/CD14⁺ peripheral blood mononuclear cells (PBMCs) from a healthy control individual was collected by flow cytometry.

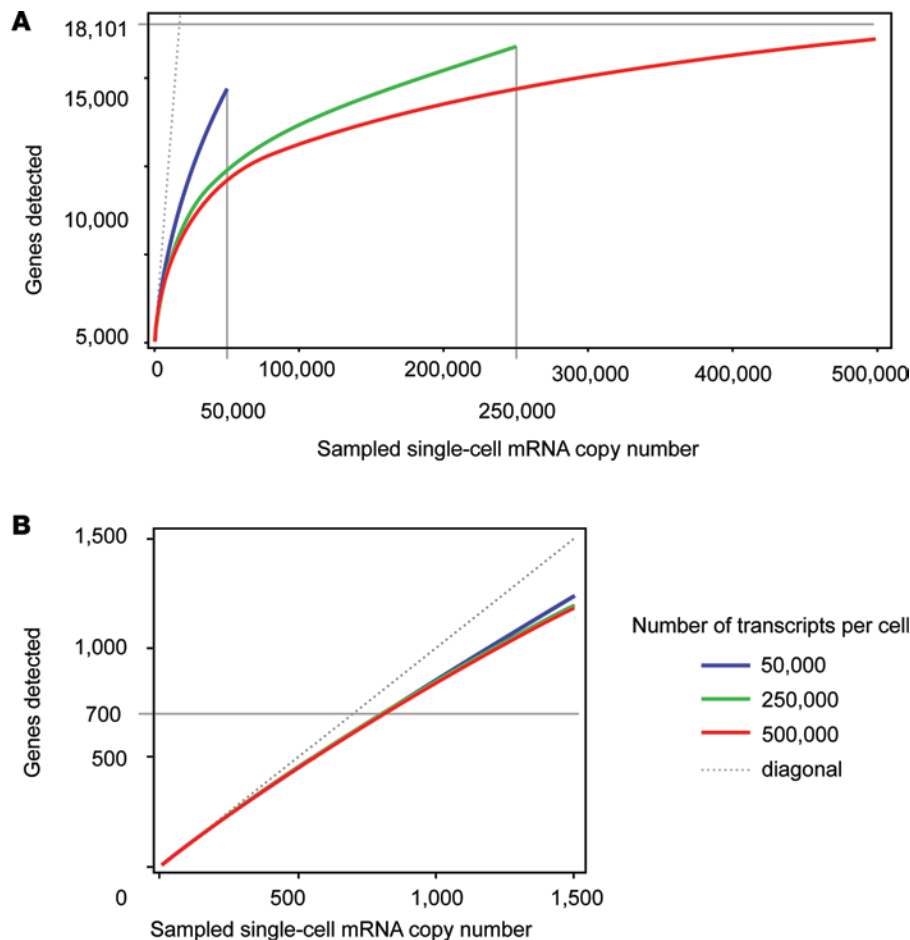


Figure 1. Simulated average number of detected genes by the number of mRNA transcripts sampled from simulated single cell transcriptomes. Three sizes of a single cell transcriptome were simulated: 50,000 (blue); 250,000 (green); and 500,000 (red) mRNA transcripts. The simulation was based on gene frequencies from bulk HEK293 polyA RNA-seq data with 18,101 distinctive genes with 100 iterations for each point. **(A)** Complete graph for sampled simulated single cell transcriptome sizes from 1–500,000. Because simulated single cell transcriptomes almost never have all 18,101 genes detected in bulk RNA-seq, the average number of detected genes is asymptotically approaching the maximal number of genes with an increase of the size of simulated single cell transcriptome. **(B)** Enlarged fragment of near-linear part of **A** corresponding to the approximate number of genes detected in individual cells (700 genes). The gray dotted diagonal ($y = x$) represents a hypothetical linear relationship between number of transcripts sampled and number of genes detected.

scRNA-seq analysis and single cell data simulation. Cell suspensions dissociated from skin and kidney, or presorted from PBMCs, were each loaded into a Fluidigm C1 Autoprep system for capture of up to 96 single cells followed by cell lysis and conversion into single cell cDNA libraries by oligoT reverse transcription. On average, we captured 50 single cell transcriptomes per skin and kidney biopsy, and sequenced to a depth of approximately 0.5 million sequence reads/cell. We sequenced a total of 1584 single cell libraries, of which 899 passed our quality criteria after read alignments and quantification, as described below in Methods.

In the 899 single cell transcriptomes that passed our quality filter, we were able to call approximately 700 genes per cell as expressed after a read count threshold was imposed. To determine how the number of detected genes was related to the transcript capture rate, we simulated scRNA-seq from a dataset of conventionally bulk-sequenced HEK293 cells. The simulated expression profiles showed that the experimental values fell within the range where the number of detected genes was independent of the size of estimated single cell transcriptomes and the relation of captured transcripts to number of detected genes was linear. An average number of 700 detected genes corresponded to about 850 captured transcripts (Figure 1, A and B). When all single cell expressed transcripts were summed up across all cells, overall expression of 15,322 genes was evident. Averaging of all single cell renal biopsy expression data and comparing with a bulk-tissue conventionally sequenced renal biopsy yielded a Pearson's correlation coefficient of 0.72 ($P < 0.00001$), indicating that the capture frequencies of single cells mirrored the overall tissue composition (Figure 2).

In order to determine the number of expected kidney-specific expressed genes, we computed the Shannon's entropy of genes using expression profiles from publicly available Genotype-Tissue Expression (GTEx) project bulk tissue RNA-seq data (16). We found 238 kidney-specific genes with an entropy score greater than 0.3 and an average expression value greater than 20 transcripts per million (TPM). Of the kidney-specific genes, 12 were never identified in our scRNA-seq libraries (Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/jci.insight.93009DS1>). Several of

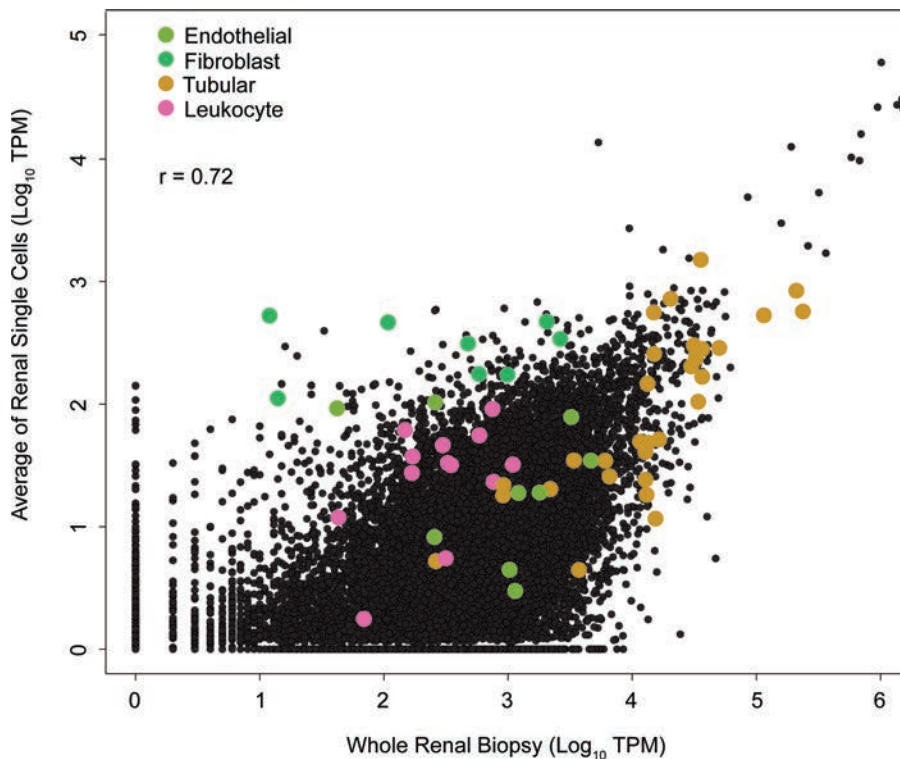


Figure 2. Pearson's correlation between the average kidney single cell expression and a bulk sequenced renal biopsy. Kidney single cells ($n = 361$) were averaged into a single "reconstructed" biopsy and correlated with a conventionally sequenced biopsy. Each dot represents a gene. Cell lineage markers for endothelial cells (light green: *SELE*, *PECAM1*, *FLT1*, *LYVE1*, *VWF*, *MCAM*, *FLK1*, *CDH5*, *ARHGDI1B*, *A2M*, *PTPRB*), fibroblasts (dark green: *MFAP4*, *MFAP5*, *COL3A1*, *COL1A2*, *PRG4*, *COL1A1*, *PLA2G2A*, *APOD*), tubular cells (yellow: *UMOD*, *SLC12A1*, *SPP1*, *CA12*, *ALDOB*, *CALB1*, *PDZK1IP1*, *NAT8*, *SLC22A6*, *SLC22A8*, *AQP1*, *SLC34A1*, *SLC12A3*, *SCNN1B*, *CLCN5*, *CLDN16*, *GPX3*, *DEFB1*, *KCNJ1*, *KNG*, *SCNN1A*, *SLC22A2*, *PAX8*, *SLC23A3*, *KCNJ15*, *MT1G*, *SLC12A6*, *BHMT*, *ALDOB*, *SLC13A1*), and leukocytes (pink: *CD14*, *LYZ*, *FCER1G*, *CD4*, *MS4A6A*, *PTPRC*, *ITGAX*, *MRC1*, *CD247*, *FCGR2A*, *LTB*, *MARCO*, *EMR1*, *IL10RA*) are highlighted.

these genes are known to have expression restricted to rare and difficult-to-dissociate glomerular cell types, such as juxtaglomerular cells, podocytes, and mesangial cells, while the remainder contained unusually GC-rich 3' UTRs potentially difficult to reverse transcribe. The percentage of Gs and Cs in the 3' UTRs of these genes are listed in Supplemental Table 1, along with the length of the 3' UTR.

Single cell lineage determination. In order to classify the single cell transcriptomes based on their stochastically composed scRNA-seq expression profiles, we performed principal component analysis (PCA). The mean and a dispersion measure (variance/mean) for genes expressed in at least 3 cells were calculated across all cells. Highly variable genes were loaded into the PCA, resulting in 9 significant principal components (PCs), which were collapsed into 2 dimensions using t-distributed stochastic neighbor embedding (t-SNE) (17) followed by a density clustering approach, which resulted in 6 clusters of cells (Figure 3, A and B). Only 2 of the 6 cell clusters were composed of skin- or kidney-specific cell types, while 3 clusters represented cell types common to both the kidney and skin, with the final cluster also including the sub-sorted PBMCs. Cell clusters were assessed to determine if clustering is driven by patient-specific or sample batch-specific effects; however, clusters comprised cells from each patient, and therefore, no clear batch effects were noted. Differential gene expression analysis of the resulting clusters revealed specific gene sets defining each cluster (Figure 3C).

Next, cell lineages were inferred based on mutually exclusive expression of multiple established cell lineage markers; for example, the presence of skin keratins *KRT1* and *KRT10* in cluster 5 identifies this cell type as keratinocytes, while the presence of various solute carriers such as *SLC12A1* is characteristic of kidney tubular cells. Fibroblasts were identified by expression of the type I collagen gene *COL1A2*, as well as other extracellular matrix proteins such as *FBN* and *DCN*. Endothelial cells expressed the cell adhesion molecule *PECAM1* and growth factor receptor *FLT1*. T cells and myeloid cells were assigned by expression of the T cell receptor protein *CD3G* and the antibacterial enzyme *LYZ*, respectively (Figure 3D). While the expression of only a single gene is shown, each cluster was defined by the expression of many genes, which are listed with their average abundance in Supplemental Table 2. Alternative to the PCA approach, conventional gene expression analysis using unsupervised clustering, by selecting the union of the top 10 most abundant nuclear encoded genes of each cell, was also able to differentiate cell types and group them according to clusters of cell-type-specifically expressed genes (Supplemental Figure 1).

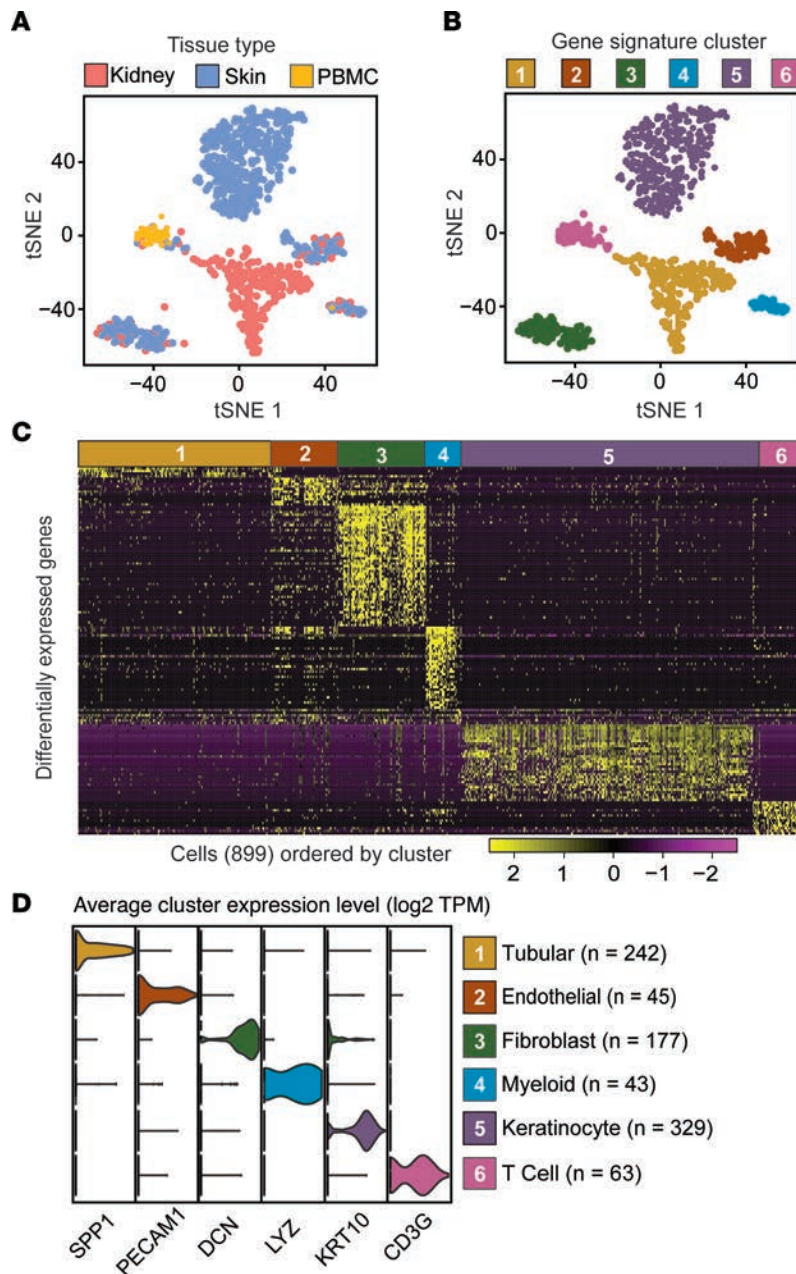


Figure 3. Lineage determination of single cells from skin, kidney, and peripheral blood mononuclear cells (PBMCs). (A) Clustering of cells ($n = 899$) by t-distributed stochastic neighbor embedding (t-SNE). Cells are colored based on tissue of origin from skin (blue), kidney cells (red), and PBMCs (yellow). (B) Six distinct clusters generated by t-SNE plotting. (C) Differentially expressed genes across 6 cell clusters. In this heat map, rows correspond to individual genes found to be selectively upregulated in individual clusters ($P < 0.01$). (D) Violin plots demonstrating expression of lineage markers that indicate the identity of the clusters generated by t-SNE plotting.

Tubular cell subpopulations determined. The expression of some specific tubular cell markers has previously been shown to depend on their specific location along the nephron (18). We therefore performed additional analysis of only tubular cells using PCA and t-SNE (19), which further subclustered them into 4 subtypes. These clusters were characterized by the expression of genes previously localized to the proximal convoluted tubule, loop of Henle, distal convoluted tubule, and collecting ducts (Figure 4). The full list of differentially expressed genes defining each tubular cell compartment is listed in Supplemental Table 2. Proximal convoluted tubular cells expressed *ALDOB* and *MIOX* (18), loop of Henle cells expressed *UMOD* and *SLC12A1* known to be restricted to the thick ascending limb (20, 21), distal convoluted tubular cells expressed *CDH16* (22), and collecting duct cells selectively expressed the transcription factor *KLF5* (23), as well as cell junction proteins *TJPI* and *CLDN4* (24).

Nonlesional non-sun exposed keratinocytes from patients with LN demonstrate an IFN response signature. To determine whether biopsies from nonlesional, non-sun-exposed skin from SLE patients have value as potential biomarkers for lupus in general or as surrogates for LN in particular, differential expression analysis was performed on keratinocytes from skin biopsies of LN patients and healthy controls. In this analysis, each keratinocyte was considered a biological replicate.

Twenty-eight genes were significantly upregulated, and 2 genes were significantly downregulated in the LN keratinocytes compared with keratinocytes from healthy controls (Figure 5A). The full list of gene fold-changes is provided in Supplemental Table 3. Interestingly, the 4 most significantly overexpressed genes are all known to be IFN inducible (Figure 5, A and B), indicating an upregulation of an IFN response pathway. To evaluate the IFN response, we first averaged the expression of LN and healthy control keratinocytes into a single LN or healthy keratinocyte profile. Then, using a list of experimentally identified IFN-inducible genes (25), we determined through a cumulative distribution function that LN keratinocytes significantly upregulated IFN-inducible genes as compared with healthy control keratinocytes (Figure 5C). When LN patients and healthy controls were blinded and randomly redistributed, IFN-inducible genes and ubiquitously expressed genes were expressed at similar levels. The limited number of nonkeratinocytes precluded determination of the IFN signature in other cell types present in the skin.

One major advantage of scRNA-seq technology, which we confirmed here, is its unique ability to resolve gene expression in a cell-type-dependent manner. Nevertheless, it would be valuable to understand whether the single cell method is more informative than bulk RNA sequencing in quantitating IFN-related gene expression.

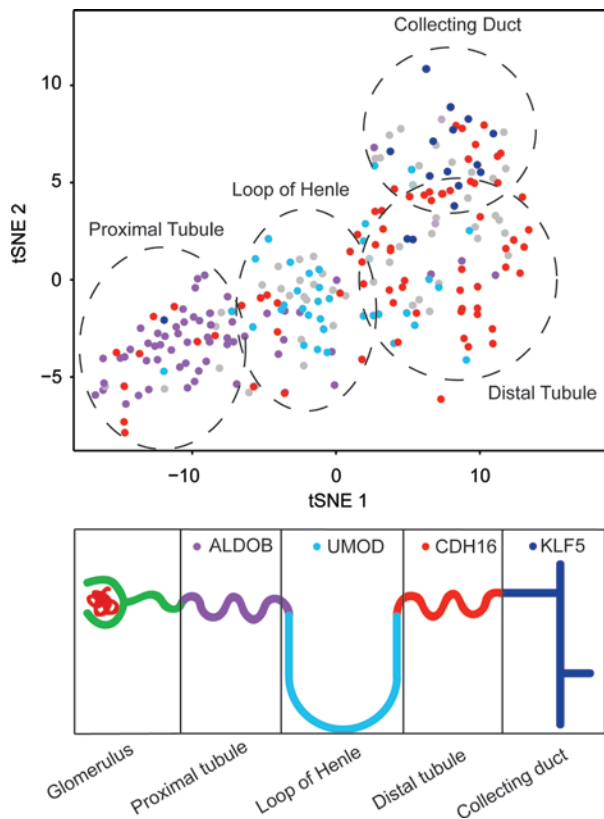


Figure 4. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) plot of tubular cell clusters from renal biopsies. Expression of markers of the different tubular compartments are overlaid, indicating the lineage of the cluster. *ALDOB* (purple), *UMOD* (cyan), *CDH16* (red), and *KLF5* (blue) for proximal tubules, loop of Henle, distal tubules, and collecting duct cells, respectively.

We were able to study 2 patients with lupus in whom there was sufficient tissue (within the same patient) to compare bulk and single cell sequencing. Interestingly, the IFN signature was also evident at the bulk level, but it was less intense than that seen in the single cells (data not shown). If confirmed in future studies, this finding would support a conclusion that only certain cell types in the skin of LN patients display an IFN signature.

IFN scores from tubular epithelial cells from LN patients correlate with renal clinical scores. To investigate if measuring the IFN response signatures in the tubular cells of LN patients is a useful metric in measuring LN disease activity, an IFN response score was created for each LN patient. First, tubular cells were combined into a single averaged tubular cell profile per patient for the subset of IFN-inducible genes (25) with a mean expression of over 100 TPMs (Supplemental Table 4). IFN response scores were then created for each patient tubular cell profile by averaging the expression of this subset of IFN-inducible genes. IFN scores varied from patient to patient. Restricting the analysis to only patients with at least 10 tubular cells, IFN response scores were significantly correlated with several clinical readouts of disease, including histologic evidence of chronicity, proteinuria, and glomerular IgG deposition (Figure 6A and Supplemental Table 5).

Indicators of activity or chronicity in the histopathologic analysis of LN do not always correlate with long-term clinical outcomes (26). Although the follow-up of our patients was not sufficiently extended to analyze the correlation of scRNA-seq data with important endpoints such as doubling of serum creatinine or the development of end-stage renal disease, we studied the correlation of tubular IFN scores with complete and partial responses in the 9 patients for which this data was available at the 12 month time point. It is intriguing that tubular IFN scores at baseline were significantly lower in patients that entered a complete response within 1 year (Figure 6B). However, this observation should be taken with caution, given the limited sample size, and confirmation will be needed in a larger cohort of patients.

Although there was insufficient material from clinically indicated kidney biopsies to perform both bulk and single cell RNA-seq in the same patient, we did compare the single cell tubular IFN signature to a “pseudo-bulk” signature. The latter was compiled considering the multiple single cell types that could be stratified from a single biopsy. Importantly, pseudo-bulk IFN scores no longer showed a statistically significant correlation with IgG deposition ($r = 0.07$, $P = 0.89$) or proteinuria ($r = 0.56$, $P = 0.1$).

Discussion

We have demonstrated that scRNA-seq of renal and skin biopsies from LN patients is a feasible and informative technique. Furthermore, this methodology can be applied to very small tissue samples; with regard to the kidney, all biopsies were done for clinical purposes with the immediate availability of extra tissue considered in excess of that needed for full clinical evaluation or obtained through an additional research core. While this study demonstrated feasibility in skin and kidney biopsies, the methodology is applicable to many other biopsy types where tissue is limited and often necessary for diagnostic purposes.

We have shown in this study the presence of active IFN signatures in the major tissue-constituting cell populations of kidney and skin. One possibility that needs to be further considered is that bulk biopsy RNA sequencing may yield similar classifiers, as long as the average cell type composition is comparable across biopsies. Nevertheless, our approach sets the stage for discovery of IFN signatures, or any other gene expression pathways, in a cell-type-dependent manner. Importantly, such signals may not be detectable, or would be less obvious, in bulk analysis if the particular pattern of gene expression is limited to a minor cell type (e.g., fibroblasts or endothelial cells).

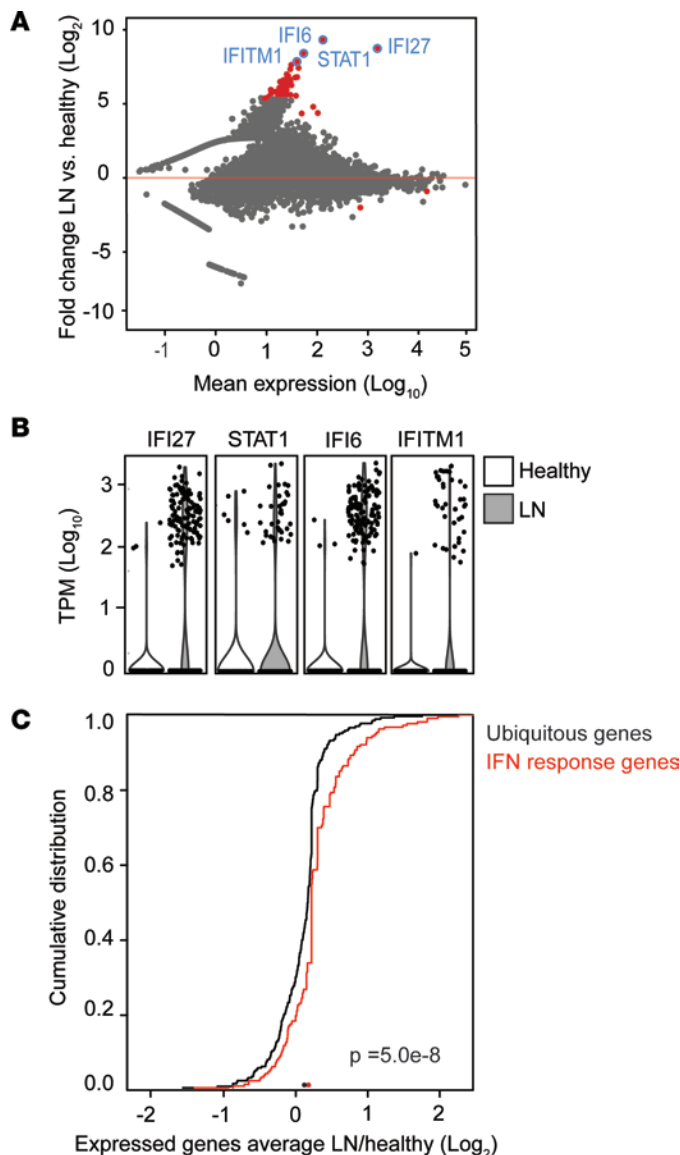


Figure 5. Keratinocytes in lupus nephritis (LN) patients demonstrate an increased IFN response signature as compared with healthy controls. (A) MA-plot comparing differential expression of genes between keratinocytes from LN patient skin biopsies ($n = 240$) and healthy controls ($n = 89$). Genes above the red line indicate increased expression in LN patients. Significantly differentially expressed genes as determined by the Wald test are colored red ($P < 0.0001$), and the 4 most highly significant genes are outlined and labeled in blue. (B) Violin plots of the 4 most significantly differentially expressed genes. (C) Cumulative distribution function (CDF) of the ratio of averaged patient to healthy control keratinocytes for IFN-inducible genes ($n = 212$) and ubiquitous genes ($n = 262$) compared using the Mann-Whitney U test.

Although we acknowledge that a relatively small number of patients were studied, the current work is innovative and highly novel in being the first, to our knowledge, to demonstrate the feasibility of scRNA-seq technology to study clinically obtained renal biopsy tissue. Based on these and other interesting results being obtained by other network investigators, plans are already in place to significantly scale up collection of patient data and the number of cells captured during the next phase of this research effort.

Since cells were sequenced agnostically and not preselected by cell sorting, it was important to demonstrate the ability to differentiate cell types on transcriptomic data alone. While we were able to detect many of the expected cell types, a limitation of the study was the inability to identify podocytes and mesangial cells residing in glomeruli from the kidney single cells since genes known to be expressed at the protein level by these cells were absent from our datasets. Similarly, B cells were not captured in our analysis, and only a modest number of T cells were identified. This is likely due to the relatively low presence of the latter cell types as compared with kidney tissue-resident cells in the clinically obtained samples and should be addressed by the higher throughput of the next generation of single cell sequencing techniques. Histopathologically, in the renal milieu of LN, kidney tissue-resident cells far outnumber the immune-infiltrating cells. Other network investigators are studying the application of flow cytometry-based presorting methodologies for targeted characterization of specific cell types, including infiltrating

immune cells, by scRNA-seq. When different digestion protocols were initially evaluated and established based on cell viability, there was no evidence that the methods used to dissociate the cells specifically affected the viability of the lymphocytes. We were, however, able to identify the different tubular segments of the nephron. Interestingly, many of the genes we report as enriched based on transcriptomic analysis have not yet been reported as characteristic for particular tubular subtypes. Further investigation will be necessary to confirm the specificity of these markers.

Another limitation of our study is that renal biopsy samples from patients without LN were not available for comparison. Nevertheless, comparing the spectrum of clinical parameters (histology and proteinuria) in relationship to the expression of IFN response genes (Figure 6) does, however, begin to address this important consideration. Future inclusion of patients with class II LN may also be revealing. Since renal tissue from lupus patients without nephritis is not likely to become available, efforts are being made to obtain normal kidney tissue (e.g., from kidney donors before transplant) as an additional control group for future studies. Finally, including patients with other indications for a kidney biopsy (e.g., vasculitis) was beyond the scope of this present study. Ultimately, our goal is to evaluate the renal transcriptome to gain insight into (i) the response to therapy and (ii) the pathogenesis of LN and the relationship to biopsy class, activity, and chronicity.

Type I IFNs, represented by evolutionarily similar IFN- α and IFN- β proteins, are important cytokines driving immune dysregulation in SLE (27). IFN- α has long been implicated in the pathobiology of SLE and was first

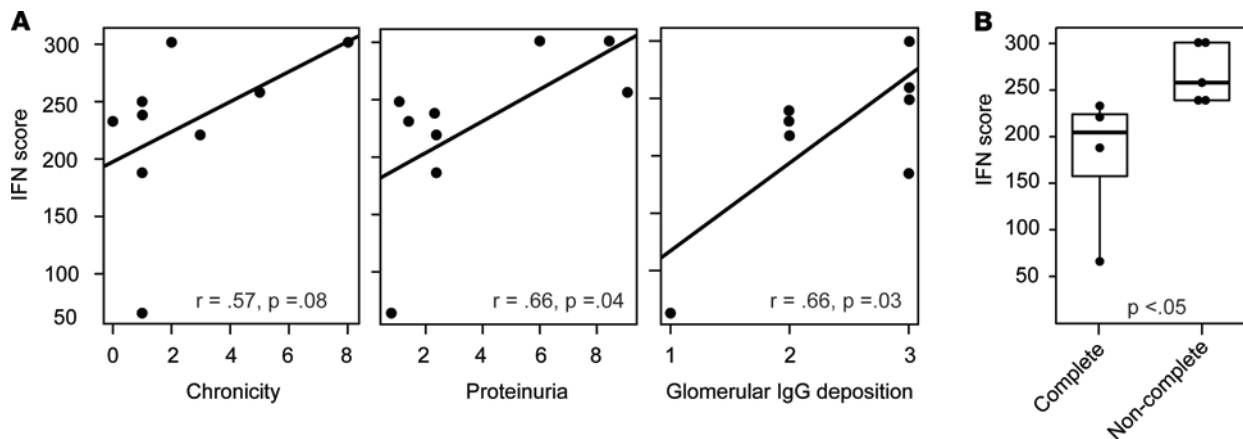


Figure 6. IFN scores of lupus nephritis (LN) tubular cells correlate with clinical scores and with the response to treatment. (A) Spearman's correlation between patient ($n = 8-9$) clinical scores of chronicity, proteinuria, and glomerular IgG deposition and tubular IFN scores. (B) Patients with complete response to treatment at 12 months after biopsy (urinary protein creatinine ratio of < 0.5 ; normal serum creatinine, or if abnormal, $\leq 125\%$ of baseline) ($n = 4$) had significantly lower IFN scores (2-tailed Student's t test $P < 0.05$) at the time of biopsy than patients who did not completely respond to treatment (the latter group included 4 partial responders [i.e., with decreased proteinuria but not to < 0.5] and 1 nonresponder). Only patients with at least 10 tubular cells were included.

documented as elevated in the serum of SLE patients in 1982 (28). More direct evidence for the role of IFN- α in the pathogenesis of SLE came from studies where IFN- α used to treat cancer caused induction of SLE-like symptoms, including elevated titers of antinuclear antibodies (29). Additional support for the importance of IFN- α in SLE was provided by microarray studies, which clearly demonstrated an IFN signature in the PBMC of many patients with SLE. Furthermore, the presence of an IFN signature was associated with increased severity of renal disease (30, 31). Mechanistic studies have since shown that plasmacytoid DCs are a major source of IFN- α in SLE in response to uptake of nucleic acids and stimulation of intracellular TLR7 and TLR9, and that IFN- α activates other immune cells that can also produce IFN- α (32). Finally, genetic studies have shown that several transcription factors such as IRF5 and IRF7 are lupus susceptibility loci and that healthy family members of SLE patients have increased circulating IFN- α as compared with healthy nonfamily members (33).

Independent of specific biopsy class, an IFN response signature was detectable in the tubular cells of every patient. However, tubular IFN response scores differed between patients and were found to correlate with chronicity scores, IgG deposition, degree of proteinuria, and perhaps the response to therapy. This suggests that, although IFN may be contributory in all cases of LN, quantitation of IFN response signatures at the mRNA level may be useful as biomarkers, especially since circulating IFN- α is not always measurable in all patients. Given the challenges in reproducibility of directly measuring circulating IFN- α at the protein level, monitoring the expression of a subset of genes known to be IFN inducible has been a standard approach in lupus research and can now also be based on RNA-seq data. The extent of IFN responsiveness may be dependent on cell types and access to the vasculature. As IFN- α has been a focus of attention in SLE for over a decade with new therapies directed at its inhibition, creating a reliable method of measuring IFN activity and its resulting gene expression profile in reactive cells could be used as a potential biomarker informing treatment.

An IFN response signature was also observed in the keratinocytes of LN patients. This is significant, especially because the skin biopsies were collected from nonlesional, non-sun-exposed skin that did not exhibit any clinical signs of inflammation. These findings may reflect exposure to systemic circulating levels of IFN- α or in situ production by resident or infiltrating cells. In support of our findings, IFN- α -producing cells as detected by IHC and in situ hybridization have been reported in nonlesional skin biopsies from lupus patients (34). The absence of cutaneous lesions, despite expression of IFN response signature, awaits explanation but could point to targetable protective candidates not present in injured organs, such as the kidneys, which are similarly exposed to IFN- α or other systemic circulating cytokines and suffer injurious consequences. Specificity related to LN, however, remains to be established since we have not yet examined nonlesional, non-sun-exposed skin from lupus patients without renal involvement.

We believe that it is likely that scRNA-seq analysis of kidney tissue from patients with LN will provide new biological insights beyond the IFN signature, as well. However, because the IFN signature was dominating, we were not confident at this time proposing additional pathways with lesser significance.

Moreover, studies of the IFN signature and its significance in other cell types beyond tubular cells and keratinocytes were limited by cell numbers but will remain a priority to be addressed when this maturing technology enables the study of larger cell counts.

In summary, scRNA-seq of biopsy specimens requires only a small amount of tissue from a single needle core or punch biopsy. Accordingly, the application of transcriptomics could be a powerful adjunct to histology, facilitating more informative treatment decisions for individualized patients. The technical optimization of single cell capture approaches enabling increased cell capture rates and numbers, and increased efficiency of reverse transcription, will contribute to this development.

Methods

Supplemental Methods are available online with this article.

Procurement of clinical samples. Skin punch biopsies (1 × 2 mm) from nonlesional, non-sun-exposed skin, and segments of a 18 gauge renal needle core biopsies that were dispensable for clinical diagnosis (0.8 × 3 mm) were obtained from patients with SLE undergoing clinically indicated renal biopsies. The mean tissue amount of skin biopsies was 8 mg (5–9 mg) and the mean amount of renal biopsies 3 mg (1–5 mg). Only renal biopsies with a pathology report indicating classes III–V, or a mixed class of III/V or IV/V, were included in the study. Comparable skin biopsies and blood were collected from healthy control donors. Renal biopsies were scored by a renal pathologist according to the International Society of Nephrology/Renal Pathology Society (ISN/RPS) guidelines and NIH activity and chronicity scales. IgG deposition was scored on a scale of 1 to 3+.

Tissue dissociation and single cell isolation. Renal and skin tissue biopsies were incubated for 15 minutes in a 37°C water bath in a freshly prepared solution composed of 50 µl of 2 mg/ml Liberase TL (Roche Diagnostics) stock diluted in Tyrode's solution and 350 µl of Tyrode's solution. Subsequently, suspended cells and tissue fragments were removed from the residual biopsy core and strained through a 70-µm cell strainer (Falcon) and washed through with 300 µl of Tyrode's solution into 5 ml of FBS on ice. The residual biopsy was incubated once again with 500 µl fresh tissue dissociation solution, passed through the same mesh, and collected with the first batch of cells in FBS (Gibco). Finally, any remaining tissue was incubated for 10 minutes at 37°C in 400 µl of a solution composed of 50 µl of 0.25% trypsin (Gibco) and 350 µl of Tyrode's solution, and it was again strained through the mesh into the same FBS. Finally, cells were collected by centrifugation in a 50-ml conical tube (BD Biosciences) using an Eppendorf centrifuge 5804 with a A-4-44 rotor at 200 g for 5 min. Cells were then resuspended in 100 µl of Tyrode's solution in a fresh 1.5-ml microfuge tube. 10 µl of the cell suspension were removed, and the number of cells was determined using a Bio-Rad T10 automated cell counter. The concentration of cells in suspension ranged from 20,000–1,000,000 cells/ml but were typically approximately 300,000 cells/ml. Cell suspensions were either diluted or concentrated by centrifugation to arrive at a final concentration of 200,000 cells/ml in Tyrode's solution. Viability was assessed by trypan blue staining.

CD14⁺/CD4⁺ cell sorting. The isolation protocol achieved an enriched population of CD4⁺DAPI[−] and CD14⁺DAPI[−], which were obtained following sorting by FACSaria II (BD Biosciences) after using CD14 mouse anti-human PE conjugate (Invitrogen) and anti-CD4 APC-H7 (BD Pharmingen), respectively.

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 96-well microfluidic chip (Fluidigm) and processed according to the Fluidigm C1 protocol revision A using the recommended standard mRNA-seq reagents and program. The occupation of single cell capture sites was verified using a Zeiss Axiovert 200 inverted microscope averaging 50 single cell captures per experiment. The captured cells were lysed, polyA mRNA was reverse transcribed, and cDNA was preamplified using 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 Array-Control RNA Spikes (Thermo Fisher Scientific) was prepared as described in the Fluidigm C1 protocol and added to the lysis reaction (Mix A) for each experiment.

The cDNA concentration was determined using the Picogreen assay on a plate reader. Samples were typically measured in the 0.2–2 ng/µl range. cDNA from a subset of cells from each run was checked on an Agilent Bioanalyzer 2100 High Sensitivity DNA chip. Only wells with a visually confirmed single cell capture were selected for downstream library preparation and sequencing.

Preamplified cDNA libraries were tagged and barcoded using the Nextera XT Library Preparation Kit (Illumina) with dual indexing. Up to 96 cells from up to 3 distinct chips were pooled using these barcodes and sequenced single-end, 100-bp on Illumina HiSeq 2500 sequencers in Rapid Run mode.

Bioinformatic analysis. Single FASTQ files corresponding to up to 96 cells were demultiplexed into single cell FASTQ files by separating reads based on the Illumina Nextera index primers. The adaptors were then trimmed from the single cell FASTQ files and aligned to the human reference genome GRCh38 downloaded from Ensembl using the STAR aligner (version 2.5.0a) (35), allowing up to 2 mismatches to the reference sequence. The reference genome only contained the canonical chromosomes and nonchromosomal contigs; haplotypes were not included. The STAR run was done providing transcript annotation defined in a general transfer format (GTF) file from Ensembl release 83. Unaligned reads were further mapped against the sequences of the RNA spikes using the BWA aligner (versions 0.7.5-r404 and 0.7.12-r1039) (36), allowing up to 2 mismatches. Uniquely mapped reads to the reference genome or spike-ins were counted using featureCounts (version 1.5.0), and the reads mapping to the human genome were collapsed on the gene level (37). Transcripts from the Havana database (<http://www.ensembl.org/info/data/ftp/index.html>) were removed from the Ensembl 83 GTF, as they were frequently overlapping with older gene annotation, leading to multi-mapping. Transcripts from the ensembl_havana merge were kept, however. Demultiplexing and sequencing of barcoded single cell libraries is accompanied by a low frequency of misassigned reads across single cell transcriptomes, which causes an inflation of the number of cellular transcripts in scRNA-seq, if left uncorrected (38). To minimize false-positive expression reporting, we used an expression threshold to filter out low-frequency reads. First, the number of total reads was multiplied by the ratio of nonredundant reads to transcript length per gene in each cell. Expression of a gene was reset to 0 if the calculated value was less than 1. Nonredundant reads were extracted using Picard tools (version 2.7.1). Finally, single cell libraries with read evidence for less than 25 nuclear-encoded genes or less than 20,000 mapped reads were excluded from further analysis. The pipeline was run on RedHat Linux or MacOS 10.10.3.

Simulation of scRNA-seq. The relationship between mRNA capture rate and detectable number of expressed genes in scRNA-seq was estimated by a simulation technique. Simulated single cell transcriptomes were created using relative mRNA transcript frequencies obtained from stranded, polyA-selected bulk RNA-seq experiments of HEK293 Flp-In T-Rex cells. Simulations were run for 3 sizes of transcriptomes: 50,000; 250,000; and 500,000 mRNA transcripts.

At each step of the simulation, we created a single cell transcriptome of indicated size and performed random sampling without replacement of mRNA copies for each capture rate from 1 to the size of transcriptome. For each set of captured mRNA, we counted the number of genes they represented.

The number of transcripts N for each gene g , in a simulated single cell transcriptome was defined by:

$$N(g) = [T * F(g)] + R(\{T * F(g)\}),$$

where T is the size of the simulated transcriptome, $F(g)$ is the frequency of transcripts originating from gene g , $[T * F(g)]$ is the integer part of $T * F(g)$ and $\{T * F(g)\}$ is the fractional part of $T * F(g)$, and function $R(x)$ is the function for random choice with domain of $[0, 1]$ and defined by:

$$R(x) = 1, r \leq x \text{ and } R(x) = 0, r > x,$$

where r is a randomly generated number.

Simulated single cell transcriptomes of fixed size T had the same set of genes with $T * F(g) \geq 1$ and varied in the set of low expressed genes where $T * F(g) < 1$.

The simulation steps described above were performed 100 times, and median, mode, and variance were calculated using numbers of detected genes for each capture rate. Simulation scripts were written in Perl 5.0 and the R statistical language.

PCA and t-SNE analysis. PCA and t-SNE were performed using the Seurat package (version 1.2) for R as previously described (19). Briefly, the count matrix were depth normalized to 1 million reads and used to identify the set of genes that was most variable across datasets. To achieve this, we calculated the mean and the dispersion (variance/mean) for each gene across single cells, followed by z-normalization, to identify outlier genes whose expression values were highly variable even when compared with genes with similar average expression. We used a z-score cutoff of .75 to identify 2,304 highly variable genes. In this analysis, all genes were evaluated for variability. These highly variable genes were loaded into a PCA that yielded 9 significant PCs and then were used as the input for t-SNE with the perplexity parameter set to 15 (17).

Data and materials availability. RNA-seq data are available on the NCBI Short-Read Archive (SRA) under the accession number PRJNA379992.

Statistics. The differential expression analysis was performed using DESeq2 (version 1.10.1) (39) and R (version 3.3.1). Briefly, count matrices were fit for a generalized linear model per gene following a

negative binomial distribution. Dispersion estimates for each gene within groups were shrunk using an empirical Bayes approach. Log fold-changes were compared between disease groups using the Wald test.

Correlation values either represent Spearman's ρ rank coefficients or Pearson's r as indicated. Cumulative distribution functions were compared with a Mann-Whitney U test. The differences in IFN scores and response to treatment were compared using a 2-tailed Student's t test, with a P value less than 0.05 considered significant.

Study approval. All patients and controls gave written informed consent prior to inclusion in the study. The IRBs and ethics committees of the Albert Einstein College of Medicine and New York University approved the sample collection.

Author contributions

CP, TT, and JPB conceived the study, with specific input from RC and HMB regarding the skin tissue. RC contributed to the dissociation protocol. ED, SR, MC, and HS performed all biopsy dissociations and single cell sequencing experiments. RC performed FACS sorting of PBMCs. BG, PI, HMB, NJ, NB, JN, and JM assisted with patient consent, sample acquisition, and transport. HMB and PI performed the skin biopsies. MK and PM assisted with GTEX specificity analysis and single cell simulations, respectively. KMA assisted with the cDNA library preparation and performed the read alignments. ED analyzed all data. TW aided with statistical analysis of clinical data. ED, JPB, CP, and TT wrote the manuscript.

Acknowledgments

This work was supported by the Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus Network. AMP is a public-private partnership created to develop new ways of identifying and validating promising biological targets for diagnostics and drug development (<http://fnih.org/what-we-do/current-research-programs/amp-ra-sle>). Funding was provided by grants from the NIH (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-AR067685, UH2-AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-AR067694, and UM2-AR067678). We would like to thank M. Yamaji for providing valuable comments during the preparation of the manuscript, and D. Schwartz, J. Pullman, and M. Wu for scoring the renal biopsy histology.

Address correspondence to: Chaim Putterman, Division of Rheumatology, Albert Einstein College of Medicine, F701N, 1300 Morris Park Avenue, Bronx, New York 10461, USA. Phone: 718.430.4266; E-mail: chaim.putterman@einstein.yu.edu.

1. Mohan C, Putterman C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat Rev Nephrol*. 2015;11(6):329–341.
2. Yung S, Chan TM. Anti-DNA antibodies in the pathogenesis of lupus nephritis—the emerging mechanisms. *Autoimmun Rev*. 2008;7(4):317–321.
3. Chan TM. Treatment of severe lupus nephritis: the new horizon. *Nat Rev Nephrol*. 2015;11(1):46–61.
4. Weening JJ, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol*. 2004;15(2):241–250.
5. Hill GS, et al. A new morphologic index for the evaluation of renal biopsies in lupus nephritis. *Kidney Int*. 2000;58(3):1160–1173.
6. Austin HA, et al. Prognostic factors in lupus nephritis. Contribution of renal histologic data. *Am J Med*. 1983;75(3):382–391.
7. Hsieh C, Chang A, Brandt D, Guttikonda R, Utset TO, Clark MR. Predicting outcomes of lupus nephritis with tubulointerstitial inflammation and scarring. *Arthritis Care Res (Hoboken)*. 2011;63(6):865–874.
8. Appel GB, Cohen DJ, Pirani CL, Meltzer JI, Estes D. Long-term follow-up of patients with lupus nephritis. A study based on the classification of the World Health Organization. *Am J Med*. 1987;83(5):877–885.
9. Izmirly PM, et al. Expression of endothelial protein C receptor in cortical peritubular capillaries associates with a poor clinical response in lupus nephritis. *Rheumatology (Oxford)*. 2009;48(5):513–519.
10. Izmirly PM, et al. Dysregulation of the microvasculature in nonlesional non-sun-exposed skin of patients with lupus nephritis. *J Rheumatol*. 2012;39(3):510–515.
11. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57–63.
12. Treutlein B, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*. 2014;509(7500):371–375.
13. Patel AP, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396–1401.
14. Navin NE. The first five years of single-cell cancer genomics and beyond. *Genome Res*. 2015;25(10):1499–1507.
15. Kim KT, et al. Application of single-cell RNA sequencing in optimizing a combinatorial therapeutic strategy in metastatic renal cell carcinoma. *Genome Biol*. 2016;17:80.

16. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–585.
17. Jamieson AR, Giger ML, Drukker K, Li H, Yuan Y, Bhooshan N. Exploring nonlinear feature space dimension reduction and data representation in breast Cdx with Laplacian eigenmaps and t-SNE. *Med Phys.* 2010;37(1):339–351.
18. Habuka M, et al. The kidney transcriptome and proteome defined by transcriptomics and antibody-based profiling. *PLoS One.* 2014;9(12):e116125.
19. Macosko EZ, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell.* 2015;161(5):1202–1214.
20. Gamba G, Friedman PA. Thick ascending limb: the Na(+):K (+):2Cl (-) co-transporter, NKCC2, and the calcium-sensing receptor, CaSR. *Pflugers Arch.* 2009;458(1):61–76.
21. Serafini-Cessi F, Malagolini N, Cavallone D. Tamm-Horsfall glycoprotein: biology and clinical relevance. *Am J Kidney Dis.* 2003;42(4):658–676.
22. Shen SS, Krishna B, Chirala R, Amato RJ, Truong LD. Kidney-specific cadherin, a specific marker for the distal portion of the nephron and related renal neoplasms. *Mod Pathol.* 2005;18(7):933–940.
23. Fujii K, Manabe I, Nagai R. Renal collecting duct epithelial cells regulate inflammation in tubulointerstitial damage in mice. *J Clin Invest.* 2011;121(9):3425–3441.
24. Kirk A, Campbell S, Bass P, Mason J, Collins J. Differential expression of claudin tight junction proteins in the human cortical nephron. *Nephrol Dial Transplant.* 2010;25(7):2107–2119.
25. Mostafavi S, et al. Parsing the Interferon Transcriptional Network and Its Disease Associations. *Cell.* 2016;164(3):564–578.
26. Parikh SV, Alvarado A, Malvar A, Rovin BH. The Kidney Biopsy in Lupus Nephritis: Past, Present, and Future. *Semin Nephrol.* 2015;35(5):465–477.
27. Elkon KB, Stone VV. Type I interferon and systemic lupus erythematosus. *J Interferon Cytokine Res.* 2011;31(11):803–812.
28. Preble OT, Black RJ, Friedman RM, Klippel JH, Vilcek J. Systemic lupus erythematosus: presence in human serum of an unusual acid-labile leukocyte interferon. *Science.* 1982;216(4544):429–431.
29. Rönnblom LE, Alm GV, Öberg KE. Possible induction of systemic lupus erythematosus by interferon-alpha treatment in a patient with a malignant carcinoid tumour. *J Intern Med.* 1990;227(3):207–210.
30. Bennett L, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med.* 2003;197(6):711–723.
31. Baechler EC, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA.* 2003;100(5):2610–2615.
32. Obermoser G, Pascual V. The interferon-alpha signature of systemic lupus erythematosus. *Lupus.* 2010;19(9):1012–1019.
33. Niewold TB, Clark DN, Salloum R, Poole BD. Interferon alpha in systemic lupus erythematosus. *J Biomed Biotechnol.* 2010;2010:948364.
34. Blomberg S, Eloranta ML, Cederblad B, Nordlin K, Alm GV, Rönnblom L. Presence of cutaneous interferon-alpha producing cells in patients with systemic lupus erythematosus. *Lupus.* 2001;10(7):484–490.
35. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
36. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–1760.
37. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–930.
38. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics.* 2015;31(5):770–772.
39. 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.