Transcriptional trajectories of human kidney injury progression

Pietro E. Cippà,[1,2] Bo Sun,[3] Jing Liu,[1] Liang Chen,[3] Maarten Naesens,[4] and Andrew P. McMahon[1]

[1]Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California (USC), Los Angeles, USA. [2]Division of Nephrology, Regional Hospital Lugano, Lugano, Switzerland. [3]Molecular and Computational Biology, USC, Los Angeles, USA. [4]Department of Microbiology and Immunology, KU Leuven, and Department of Nephrology and Kidney Transplantation, University Hospitals Leuven, Leuven, Belgium.

BACKGROUND. The molecular understanding of the progression from acute to chronic organ injury is limited. Ischemia/reperfusion injury (IRI) triggered during kidney transplantation can contribute to progressive allograft dysfunction.

METHODS. Protocol biopsies (n = 163) were obtained from 42 kidney allografts at 4 time points after transplantation. RNA sequencing–mediated (RNA-seq–mediated) transcriptional profiling and machine learning computational approaches were employed to analyze the molecular responses to IRI and to identify shared and divergent transcriptional trajectories associated with distinct clinical outcomes. The data were compared with the response to IRI in a mouse model of the acute to chronic kidney injury transition.

RESULTS. In the first hours after reperfusion, all patients exhibited a similar transcriptional program under the control of immediate-early response genes. In the following months, we identified 2 main transcriptional trajectories leading to kidney recovery or to sustained injury with associated fibrosis and renal dysfunction. The molecular map generated by this computational approach highlighted early markers of kidney disease progression and delineated transcriptional programs associated with the transition to chronic injury. The characterization of a similar process in a mouse IRI model extended the relevance of our findings beyond transplantation.

CONCLUSIONS. The integration of multiple transcriptomes from serial biopsies with advanced computational algorithms overcame the analytical hurdles related to variability between individuals and identified shared transcriptional elements of kidney disease progression in humans, which may prove as useful predictors of disease progression following kidney transplantation and kidney injury. This generally applicable approach opens the way for an unbiased analysis of human disease progression.

FUNDING. The study was supported by the California Institute for Regenerative Medicine and by the Swiss National Science Foundation.

Introduction

Genome-wide gene expression analysis has revolutionized biomedical research by providing a new level of understanding of human biology and by highlighting the molecular complexity of individual variability (1). The next step toward the goals of precision medicine is to decipher the molecular mechanisms determining the heterogeneity of multifactorial clinical conditions and to identify the best diagnostic and therapeutic targets to anticipate and prevent disease progression (2). Chronic inflammation and fibrosis are hallmarks of dysfunctional tissue repair leading to irreversible organ dysfunction, a process well exemplified by the progression of chronic kidney disease (3–6). Our previous studies in a mouse ischemia/reperfusion model demonstrated that the progression from acute kidney injury to fibrosis and irreversible damage follows a transcriptional program over months (7). The characterization of similar mechanisms in humans is hampered by the limited access to human tissue for time-course analyses and by the challenges related to the high variability of human biology and pathology (1, 8, 9). Despite additional elements of complexity specific to

Table 1. Clinical characteristics of the study population

Donor type (LD/DBD/DCD) (%)	9/60/31
Donor age (years)	47.8 ± 14.9
Donor sex (male/female/unknown) (%)	45/53/2
Donor creatinine (mg/dl)	0.75 ± 0.24
Recipient age (years)	52.3 ± 13.0
Recipient sex (male/female)	64/36
Recipient end-stage renal disease diagnosis (%)	
Diabetes	5
Hypertension	14
Glomerulonephritis/interstitial nephritis	31
Congenital/polycystic kidney disease	21
Other	12
Undefined	17
Previous renal transplant (%)	7
Cold ischemia time (h)	12.2 ± 5.3
Delayed graft function (%)	14
HLA-mismatches (mean)	2.9/6
Pretransplant HLA sensitization	14
Histology of the kidney at baseline (PRE), according to Banff criteria	
cg (0/1/>1) (%)	100/0/0
ci (0/1/>1) (%)	63/29/8
ct (0/1/>1) (%)	34/66/0
cv (0/1/>1) (%)	82/11/3
ATN score (0/1/>1) (%)	3/19/16
% sclerotic glomeruli	7.7
Induction therapy	
Anti-CD25 therapy (%)	36
Maintenance immunosuppression	
TAC-MMF-CS (%)	100

LD, living donor; DBD, donation after brain death; DCD, donation after cardiac death. Banff criteria: cg, chronic glomerular lesions; ci, chronic interstitial lesions; ct, chronic tubular lesions; cv, chronic vascular lesions. IFTA, interstitial fibrosis and tubular atrophy; ATN, acute tubular necrosis; TAC, tacrolimus; MMF, mycophenolate mofetil; CS, corticosteroids.

allograft transplantation, renal transplants offer a unique opportunity to study the response to tissue injury in humans. Every organ transplantation begins with ischemia/reperfusion injury (IRI) in well-defined conditions at the time of surgery, and protocol biopsies — performed in many transplant centers — provide access to renal tissue over time. The recent advances in computational biology offer new opportunities to analyze large datasets (10). Machine learning techniques originally developed to analyze single-cell gene expression are particularly effective in identifying similar transcriptomes in heterogeneous tissues and predicting transcriptional changes in intermediates in the transition between distinct biological states (11–14). The same computational approach might resolve the heterogeneity of clinical conditions with an unprecedented accuracy and facilitate the discovery of molecular mechanisms determining disease progression.

Results

Kidney allograft transcriptome variability across individuals and time. We evaluated the transcriptome of 163 protocol biopsy samples from 42 kidney transplant recipients at 4 time points: before implantation (PRE), shortly after the restoration of blood flow in the graft (POST), and 3 and 12 months after transplantation. The clinical characteristics of the study population are presented in Table 1. The gene expression correlation analysis including the most variable genes in the whole dataset highlighted clusters of genes involved in a variety of biological processes from the immediate-early response after injury to organ fibrosis, with genes related to fibrosis and adaptive immunity forming a distinct cluster of highly correlated genes (Figure 1A and Supplemental Figure 1A; supplemental material available online with this
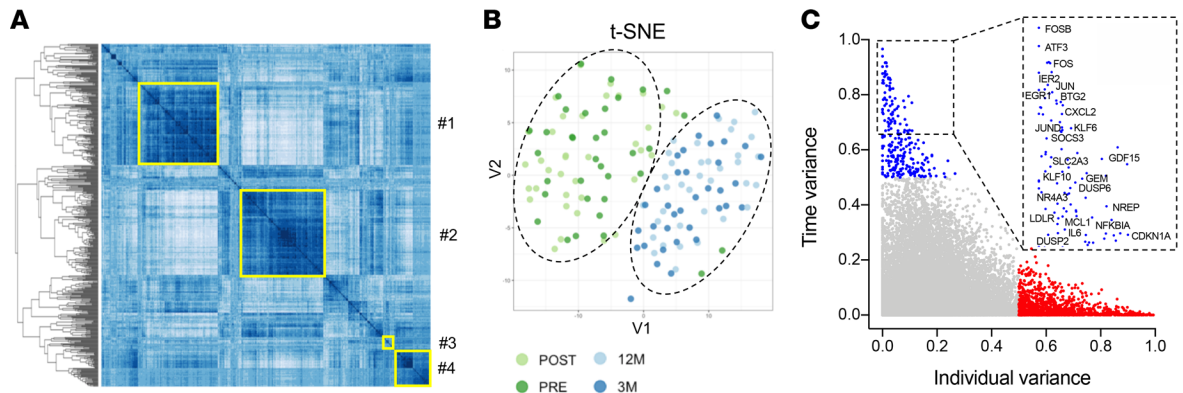
Figure 1. The kidney allograft transcriptome across individuals and time. (**A**) Gene expression correlation analysis including the 500 most variable genes in RNA-seq data (RPKM values) from 163 protocol biopsies obtained from 42 kidney allografts at 4 time points. Clusters of interest are highlighted in yellow (#1 renal physiology, #2 response to kidney injury, #3 fibrosis, #4 adaptive immunity; gene ontology analysis is presented in Supplemental Figure 1A). (**B**) T-distributed stochastic neighbor embedding (t-SNE) analysis on RNA-seq data, including all samples and showing the separation of the transcriptomes in 2 major clusters: early phase (green; PRE, before implantation; POST, after implantation), late phase (blue; 3M, 3 months after transplantation; 12M, 12 months after transplantation). (**C**) Gene expression variance decomposition analysis in linear mixed models showing the contribution of individual and time to gene expression variation. Genes showing an individual-driven variance are shown in red; genes with a time-drive variance are shown in blue, and some relevant examples are specified.

article; https://doi.org/10.1172/jci.insight.123151DS1). Dimension reduction analysis by t-distributed stochastic neighbor embedding (t-SNE) differentiated early (PRE and POST) and late phases (3 and 12 months, Figure 1B). At either the early or late phase, an individual's biopsies tended to cluster together (Supplemental Figure 1, B and C). The predominant role of interindividual variability was confirmed by the gene expression variance decomposition analysis in linear mixed models (LMMs): 21% of the total variance in gene expression reflected variability among the patient pool, while 9% of the total variance was attributed to the time of tissue sampling (Figure 1C). Genes displaying a high level of variability among individuals were associated with fundamental cell biology (transcription, RNA processing, mitochondrial function) and included genes previously associated with high interindividual variability (such as *RPL9* and *GSTM1)* and with sex-biased expression (such as *XIST* and *DDX3Y*) (1). Thus, despite the inherent variability related to the clinical setting and analysis of limited tissue samples in the context of a biological process not necessarily homogenously distributed in the tissue, the substantial level of intraindividual transcriptional constancy in repetitive biopsies from the same kidney indicated that the transcriptomic profiles were reproducible, reflecting biological differences among patients.

Early transcriptional response to ischemia/reperfusion. The gene expression variance decomposition highlighted a small number of genes characterized by a low level of variability among patients but a high level of variability across time points (blue dots in Figure 1C). This subset of genes was primarily involved in the immediate injury response (Supplemental Figure 2), suggesting a low level of interindividual heterogeneity in the initial response to ischemia (primarily triggered at the time point of tissue reperfusion) (15). The reperfusion time (i.e., the time between restoration of blood flow and POST biopsy collection) was determined by the duration of the surgical procedure and was therefore variable among patients (range 26–88 minutes, Supplemental Figure 3A). We considered this unsynchronized dataset as a time-course analysis of a biological transition, with each POST transcriptome representing an intermediate state along this process. Because of the analogy of this experimental setting with studies on cell state transitions investigated with single-cell transcription analysis, we used a similar computational approach to order PRE and POST biopsies on a pseudotime trajectory (Figure 2A) (13, 16).

PRE and POST transcriptomes were correctly separated. PRE samples were classified in 2 groups, reflecting the heterogeneity of the grafts before implantation. Differences in the percentages of kidney grafts obtained from donation after cardiac death (DCD) and in the expression of *TIMP1* were consistent with a different level of injury prior to reperfusion (Supplemental Figures 3, B–D). In contrast, all POST biopsies were ordered along the same transcriptional trajectory (Figure 2A). Consistent with the gene expression variance decomposition, the invariant sharp upregulation of the early response genes in all POST samples (Supplemental Figure 3E) and the absence of branching along the pseudotime axis suggested that all patients underwent a similar transcriptional program in the first hours after reperfusion. The type of donation was associated with the progression
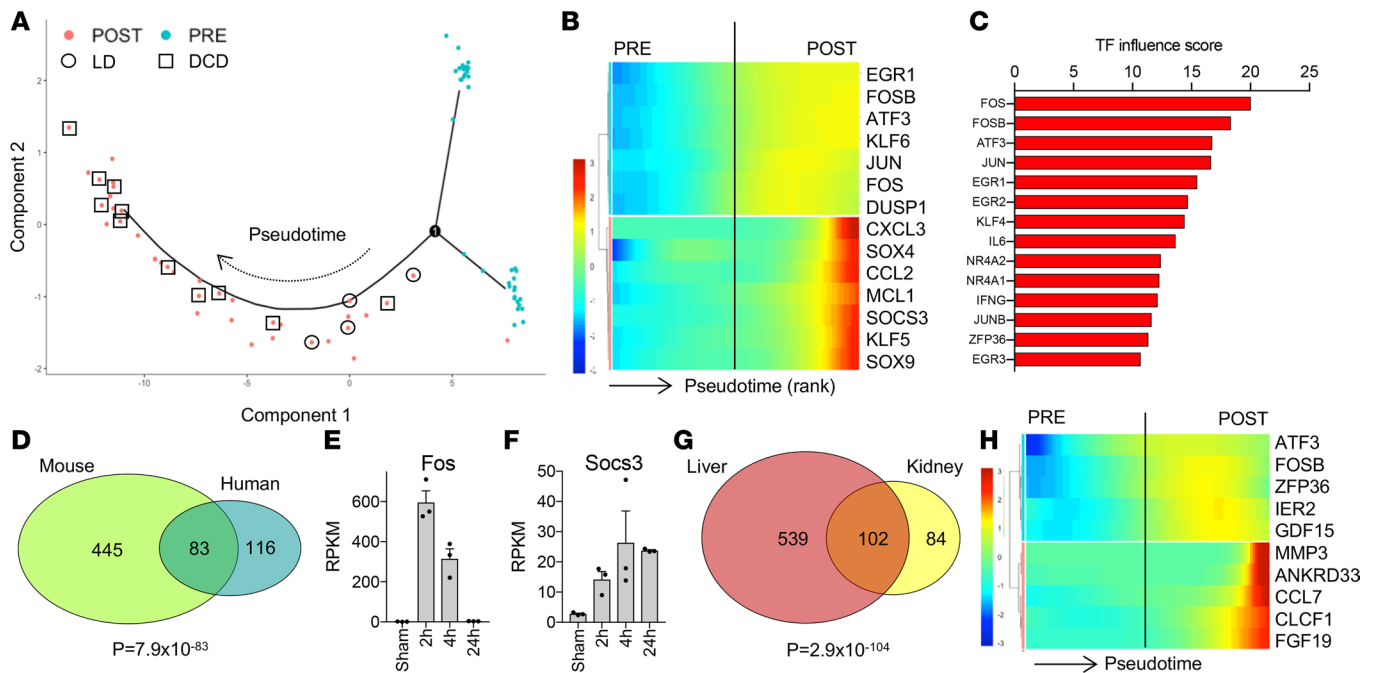
Figure 2. Early transcriptional response to ischemia/reperfusion. (A and **B)** Pseudotime analysis including samples collected before (PRE, *n* = 38) and after implantation (POST, *n* = 39). (**A)** Sample state ordering in the reduced dimensional space, as determined by the Monocle algorithm. PRE samples were classified in 2 groups and are shown in cyan; POST samples ordered along a pseudotime line from right to left and are shown in red. Among the POST samples, the circles mark samples from living donors (LD), and the black squares mark samples from donors after cardiac death (DCD). (**B)** Cluster analysis of representative genes differentially expressed along the pseudotime: samples are aligned from left to right according to the order shown in **A**. Genes are vertically aligned and classified in 2 clusters. The colors indicate the relative expression of the genes (log$_{10}$ scale). The complete list of genes is presented in Supplemental Figure 2 and Supplemental Table 1. (**C)** Influence score for the top 14 transcription factors as determined in the network analysis based on a modified Mogrify algorithm. (**D)** Venn diagram including genes differentially expressed in POST compared with PRE (human) and homologous mouse genes differentially expressed 2 hours after IRI compared with control. Significance of enrichment was determined by hypergeometric test. (**E** and **F)** Reads per kilobase per million mapped reads (RPKM) values along the early time-course analysis after IRI in mice (*n* = 3 for each time point). (**G)** Venn diagram including genes differentially expressed in POST compared with PRE in the human kidney and in the liver. Significance of enrichment was determined by hypergeometric test. (**H)** Cluster analysis of representative genes differentially expressed along the pseudotime in the liver.

along the trajectory: all samples obtained from living donation (LD) were classified in the first third of the trajectory, whereas samples from DCD were enriched in the final third (Figure 2A and Supplemental Figure 3F). Moreover, we found a significant correlation between the reperfusion time and the pseudotime (Supplemental Figure 3G), whereas the pseudotime did not correlate with cold ischemia time (Supplemental Figure 3H). Taken together, the initial transcriptional response to ischemia/reperfusion appears uniform among individuals, but the response was much less pronounced after LD. In DCD, the first response to ischemia is likely to start earlier (17); therefore the process was more advanced at the time the reperfusion biopsy was obtained.

The integrated analysis of the transcriptional changes over pseudotime displayed a dynamic gene expression regulation: the first cluster of genes (Figure 2B, Supplemental Figure 3I, and Supplemental Table 1) was characterized by a rapid upregulation of many immediate-early genes in stress responses (e.g., *FOS*, *JUN*, and *EGR1* in Supplemental Figure 3J). This initial response was swiftly followed by a second wave of genes showing marked upregulation, including the transcriptional regulators *SOX9* and *KLF5* (Figure 2B and Supplemental Figure 3, I and K) (18, 19). The regulation of other genes in the dataset can be interrogated online (https://lianglabusc.shinyapps.io/shinyapp). To identify the critical elements in the regulation of this coordinated process, we performed a network analysis, based on a modified Mogrify algorithm (12), which highlighted a major role for the transcription factors *FOS* in the context of a complex transcriptional network (Figure 2C, Supplemental Figure 4, and Supplemental Table 2).

Since data confirmation in human samples was hampered by limited tissue availability, we validated the findings in extensive data sets recently published by our group examining a mouse model of IRI (7). The comparison of differentially expressed genes in POST samples with homologous genes upregulated 2 hours after IRI in the mouse model revealed a highly significant overlap across the species (Figure 2D and

Supplemental Table 3). Moreover, the time-course analysis in mice confirmed the same pattern of regulation of homologous genes. Genes classified in the first wave in humans reached the maximal expression level 2 hours after reperfusion in the mouse model (e.g., Fos; Figure 2E), whereas the expression level of genes regulated in a second wave continued to rise at 4 hours (e.g., Socs3; Figure 2F).

The general applicability of this approach was investigated by comparing the early transcriptional response to ischemia/reperfusion across organs. We took advantage of RNA-seq data from liver transplants where liver biopsies were obtained before and after reperfusion (20). The comparison of the differentially expressed genes in PRE and POST samples in both organs highlighted a significant overlap (Figure 2G and Supplemental Table 4). The liver dataset consisted of only 15 samples for each time point, and the reperfusion time was less variable (about 2 hours). Nevertheless, pseudotime analysis delineated a biphasic transcriptional program (Supplemental Figure 5 and Supplemental Table 5): the first wave included the rapid upregulation of expression of mRNAs encoding transcriptional regulators governing the immediate-early responses. The overlap with the corresponding kidney dataset was less consistent in the second wave (12% vs. 37%, $P < 0.0001$), which was more organ specific, identifying transcripts encoding signaling factors previously characterized as regulators of liver injury and regeneration (e.g., *CCL7*, *FGF19*) (21, 22) (Figure 2H). Thus, as with previous experimental studies (23), the general pattern of transcriptional responses to ischemia/reperfusion was similar across organs: a highly conserved immediate-early response that likely initiates organ-specific mechanisms of injury and repair. More generally, an unsupervised computational approach generated a map of the early transcriptional response to IRI in a real clinical condition reminiscent of regulated transcriptional networks previously investigated in cell culture (24).

Transcriptional trajectories of kidney injury progression. To identify groups of patients with a similar transcriptional profile at later stages after renal transplant, we first generated a minimum-spanning tree, based on global gene expression profiles at 3 and 12 months. The last 10 POST samples along the pseudotime presented in Figure 2A were included as a control to benchmark the starting point of the late phase. We identified densely connected networks of samples (communities) by using random walks as implemented by the Walktrap community finding algorithm (25). Among the major communities identified with this approach, community A presented a transcriptional signature related to kidney injury, fibrosis, and chronic inflammation (Figures 3, A and B, and Supplemental Table 6). In analogy to the analytic approach applied for the early response, we considered the progression to fibrosis as a transitional process from acute injury (POST) to fibrosis (community A). We performed a pseudotime analysis including all POST, 3-month (3M), and 12M samples to cover this transition. The computational model correctly positioned the POST samples at one end of the pseudotime line (Figure 3C). The late-phase samples separated into 2 branches, with all transcriptomes previously classified in the community A forming the latest stage of transition on the longer branch and the samples previously classified in community B separated among the 2 branches (Figure 3D). In 86% of the patients, both the 3- and 12-month biopsies were positioned on the same branch or moved to the branch depicting the progression to fibrosis over time, whereas only 14% of the cases initially classified along the progression to fibrosis at 3 months were positioned on the opposite branch at 12 months. Early response genes displayed distinct characteristics in the long-term: a subset of genes (e.g., *ATF3*, *IER2*, *FOSB*) were restricted to the early response, while other genes underwent a secondary increase at late stages of chronic injury (e.g., *SOX9*, *GDF15*; Supplemental Figure 6). Of particular interest was the late activation of *HIF1A*, consistent with an oxygen supply-demand mismatch in the fibrotic kidney (26), and *JUN*, a gene associated with the pathogenesis of organ fibrosis (Supplemental Figure 6) (27).

To specifically characterize the sequence of transcriptional events during the transition to chronic kidney injury, we performed a pseudotime analysis including only samples collected at 3 and 12 months and based on genes differentially expressed in communities A and B (Figure 3E). Among the most significantly differentially expressed genes, we found genes involved in cell death (e.g., *DAD1*, *ANXA5*) and complement regulation (e.g., *CD59*, *SERPING1*), which progressively increased with pseudotime (28). The upregulation of genes related to fibrosis (e.g., *COL1A2*, *DCN*, *MMP2*; Figure 3, F and G) was a late event, coincident with the activation of genes regulating lymphocyte trafficking (e.g., *CCL19*, *CCL21*). Genes involved in the response to acute kidney injury (e.g., *HAVCR1*, *LCN2*) and innate immunity (e.g., *TRAF6*, *TLR3*) were upregulated earlier (Figure 3, F and G). The cluster of genes downregulated with pseudotime included mostly mitochondrial genes (Figure 3F; further characterized below). The complete list of genes and their distribution in each cluster is presented in Supplemental Tables 7–8. All differentially expressed genes along this pseudotime can be interrogated online (https://lianglabusc.shinyapps.io/shinyapp).
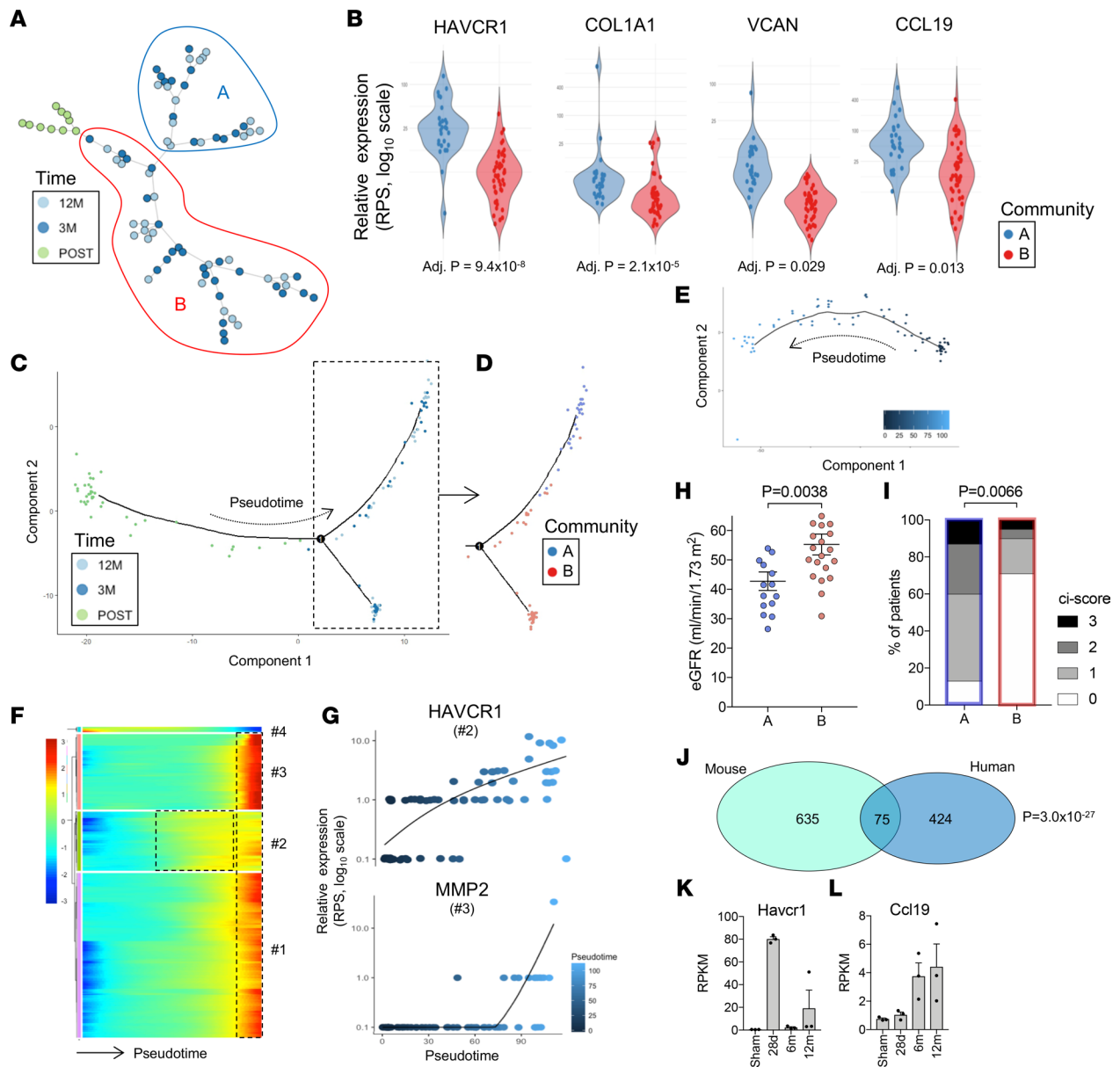
Figure 3. Transcriptional trajectory of transition from acute to chronic kidney injury. (**A**) Minimum spanning tree of kidney biopsy transcriptomes at 3 (3M, dark blue, n = 38) and 12 (12M, light blue, n = 39) months after transplantation, and 10 samples collected after implantation (POST, green). Community A (marked in blue) separated from the rest of the study population (community B, in red). (**B**) Violin plots showing mRNA per sample (RPS) values of representative genes selected as markers for kidney injury (HAVCR1, VCAN), fibrosis (COL1A1), and chronic inflammation (CCL19). Adjusted P values are reported (Benjamini-Hochberg). (**C** and **D**) Pseudotime analysis including samples collected after implantation (POST) and 3 and 12 months after transplantation. Sample state ordering in the reduced dimensional space is shown. The colors indicate the time point of biopsy collection in **C** and the classification in communities A or B in **D**. (**E**) Similar analysis including only 3- and 12-month samples based on genes differentially expressed in communities A and B. The color of the dots indicate the progression along the pseudotime, as indicated. (**F**) Cluster analysis of the top 2,000 genes differentially expressed along the pseudotime shown in **E**. Genes are vertically aligned and classified in 4 clusters (the complete list of genes is reported in the Supplemental Table 8). (**G**) Representative example of 1 gene expressed early (HAVCR1) and late (MMP2) in the transition to chronic injury. The colors indicate the pseudotime, as indicated. The numbers indicate the cluster, according to **F**. (**H**) Glomerular filtration rate at 12 months, estimated by CKD-EPI equation (eGFR) in communities A and B. P value was calculated by Mann-Whitney U test. (**I**) Histogram of the degree of fibrosis 12 months after transplantation, quantified by ci-score on conventional histology. The percentage of patients in each fibrosis category in communities A and B is reported. The groups were compared by χ^2 test. (**J**) Venn diagram including genes differentially expressed in community A compared with community B (human) and homologous mouse genes differentially expressed 12 months after IRI compared with control. Significance of enrichment was determined by hypergeometric test. (**K** and **L**) RPKM values along the late time-course analysis after IRI in mice (n = 3 for each time point).

The functional relevance of this computational model was confirmed by reduced renal function, as determined by estimated glomerular filtration rate (eGFR), and higher levels of fibrosis, as quantified by the ci-score according to Banff classification on conventional histology, observed 1 year after transplantation in community A patients (Figure 3, H and I). However, the separation in communities A and B was not just reflecting fibrosis, with 60% of the patients in community A displaying low levels of fibrosis at 12 months (ci-score 0 or 1; Figure 3I). Notably, eGFR and ci-score were similar among the 2 groups 3 months after transplantation. Furthermore, drawing on an analogy to the interspecies validation of the early response, we compared genes differentially expressed in communities A and B with homologous mouse genes upregulated 12 months after IRI (Figure 3J). Among the conserved genes across species in the late-phase response, we identified genes associated with fibrosis (e.g., *COL1A1*, *MMP2*), adaptive immunity (e.g., *CD2*, *IL2RG*), and vascular biology (e.g., *VCAM1*, *VWF*) (Supplemental Table 9). Human-specific genes were mostly related to fundamental cellular functions indicating an additional human-specific complexity in the response. No evidence was observed for a specific regulation of genes related to adaptive immunity in humans, indicating that alloreactivity — exclusively present in human dataset — was not a predominant element in our computational model. Similar genes were upregulated in the late phase of the mouse as were observed in the pseudotime analysis of human samples (e.g., the early expression of *Havcr1* and the late upregulation of *Ccl19*; Figure 3, K and L). Together, these outcomes demonstrate that the analysis identified transcriptional trajectories reflecting dynamic biological processes associated with a progression to chronic allograft dysfunction.

Early discrimination of kidney injury progression. To identify early factors determining the alternative fate toward recovery or progression to fibrosis, we compared the transcriptomes classified right after the branching point of the pseudotime trajectories (Figure 4A). The first major difference among the groups was related to mitochondrial homeostasis: the expression of mitochondrial genes in group 1 was similar to normal renal tissue (defined by PRE samples obtained from living donors), but it was significantly lower in group 2 and decreased further in group 3 (Figure 4B). Additional evidence for a critical role of mitochondrial dysfunction and impaired ATP production in association with the separation of the 2 trajectories was provided by an analysis of key regulators of mitochondrial homeostasis (e.g. *NRF1*, *PPARG*, *DNM1L*) and glycolysis (e.g. *PARP1*) in the kidney (Figure 4C) (29, 30). Conversely, the expression level of genes associated with innate immunity and extracellular matrix organization was higher in group 2, with the key regulator of kidney fibrosis *MMP7* showing the strongest upregulation (Figure 4D and Supplemental Table 10) (31). Among the genes displaying the highest influence score in the network analysis based on the comparison between groups 1 and 2 were *TP53* and *EP300* (Figure 4E and Supplemental Table 11). *EP300* was upregulated very early, according to the pseudotime analysis presented in Figure 3, E and F, and highlighted in Figure 4F. In line with the network analysis, the 33 genes most highly correlated with *EP300* were sufficient for a classification of the transcriptomes in groups 1 or 2 with a sensitivity and a specificity >90% ($P < 0.0001$; Figure 4G and Supplemental Table 12). Interestingly, binding sites for *EP300* were recently associated with renal function in epigenome-wide association studies (32).

Discussion

We investigated a complex clinical condition applying an unsupervised computational strategy, which integrates genome-wide expression analysis in heterogeneous groups of patients to identify and characterize shared trajectories of disease progression. One main innovation of this study is the application of computational approaches originally developed to characterize single-cell transcriptomics in heterogeneous tissues, which proved to be powerful tools to recognize and classify single patient transcriptomes in heterogeneous populations. The application of manifold learning-based methods is justified only if the data vary "smoothly" in the local neighborhood (33). The transcriptional transition in the first hours after reperfusion represented an ideal experimental setting to verify the potentially novel application of this approach. The findings confirmed the manifold assumption from a computational perspective and generated data consistent with the biological perspective, which correlated with clinical data, detailed analyses of a mouse IRI model, and other published material. Applying the same strategy to the long-term dataset also produced predictions that could be secondarily validated. While the manifold assumption was not necessarily satisfied a priori in this analysis, the results confirmed the smooth variation in the transcriptional profile and biomedical expectations for the dataset.

The resulting molecular map of the transcriptional program following ischemia/reperfusion in the human kidney correlated with 2 major clinical outcomes: progression to chronic kidney injury or recovery (Figure 4H). Many genes involved in this process have been previously characterized, serving to validate
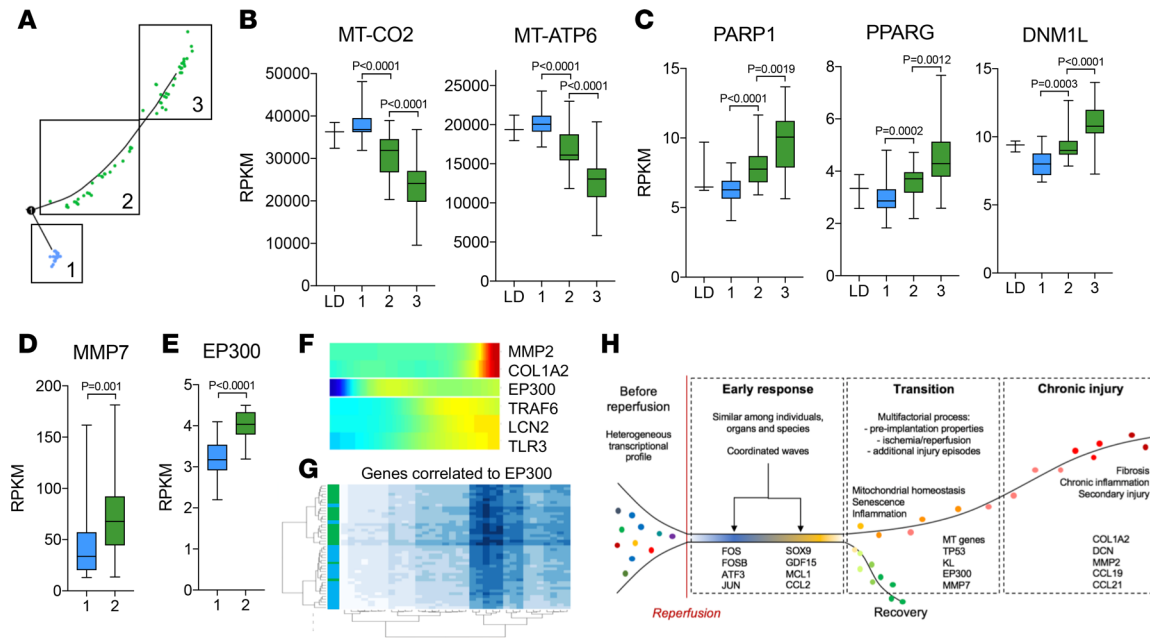
Figure 4. Early markers of transition to kidney fibrosis. (**A**) Schematic representation of the definition of groups along the divergent branches of the pseudotime analysis presented in Figure 3C. (**B–E**) Box plots of RPKM values of the indicated genes in the group defined in **A** and PRE biopsies obtained from living donor (LD), as a surrogate of normal renal tissue. The groups were compared by Mann-Whitney *U* test. (**F**) Expression profile of EP300 along the pseudotime presented in Figure 3E in comparison with other genes, reflecting the very early upregulation of EP300 along the transition to chronic kidney injury. (**G**) Samples cluster analysis based on the 33 genes highly correlated with EP300 in groups 1 and 2 (Pearson's correlation > 0.88). The division of the samples in 2 clusters is indicated by the blue (group 1) and green (group 2) bars on the left. (**H**) Conceptual summary of the study highlighting the common early injury response after ischemia/reperfusion followed by a multifactorial transition phase with divergent long-term outcomes: recovery versus the initiation of a chronic injury signature. Some of the critical genes involved in each phase are shown.

the potentially novel approach, lending confidence to a first systematic insight into the molecular mechanisms of individual human kidney disease progression. Genes revealed through these studies have potential diagnostic and therapeutic relevance. The early response to ischemia/reperfusion in the human kidney and liver biopsy samples identified similar waves of transcriptional regulation to those reported in time-course studies in vitro (24). Moreover, we identified early markers of disease progression, which represent potential targets to anticipate and prevent chronic kidney disease. Our data suggest the presence of common mechanisms in the response to kidney injury and in the transition to chronic kidney damage, despite the different factors contributing to potentially different types of kidney damage after transplantation in a heterogenous population, including the initial IRI, rejection episodes, and drug toxicity. Thus, the molecular process characterized here reflects the fundamental response of the kidney to injury and the transition to organ fibrosis in humans. These fundamental processes might have a prognostic relevance and need to be considered for any further application of diagnostic strategies based on transcriptional analyses (including single-cell transcriptomics). However, this study was not powered to predict clinical outcome. Previous clinical studies provided evidence for the potential clinical relevance of early transcriptional analysis after kidney transplantation to predict outcome (34); our detailed characterization of the transcriptional program associated with distinct clinical outcomes represents a blueprint to identify genes of interest for future clinical trials, including a larger cohort of patients and a longer follow-up.

The comparison with the mouse model not only validated the validity of the RNA-seq data in modeling human biology. The strong concordance with the mouse IRI datasets indicates that many of the injury-invoked gene regulatory responses are conserved across species, and the overlap highlights those gene sets that are likely best-suited for interventional mouse-directed human disease modeling. Moreover, the presence of common elements in a nontransplant model expands the relevance of the human model beyond transplantation, indicating that the molecular processes described here are mostly independent of the peculiarities related to transplantation. However, multiple pharmacological effects of immunosuppressive drugs might influence the transcriptional response to kidney injury (35) and should be considered before the presented data can be generalized to a nontransplant setting.

Chronic, progressive diseases substantially contribute to global morbidity and mortality (36). Our computational approach can be applied to other organs and clinical conditions (as shown in the liver), and it overcomes major limitations related to the need for repetitive tissue analyses over time to understand disease progression. In consideration of the level of accuracy obtained here even with relatively small numbers of samples, our approach might represent a powerful strategy to advance the molecular understanding of progressive clinical conditions.

Methods

Study design. We performed a genome-wide gene expression profile by RNA-seq in kidney transplant recipients, randomly selected among patients with a full set of 4 available biopsies from a database at the University of Leuven. Renal biopsies were performed at the University Hospitals of Leuven at following time points: before implantation (kidney flushed and stored on ice), after reperfusion (at the end of the surgical procedure), and 3 months and 12 months after transplantation (protocol biopsies). Additional biopsies performed in the same patients for a clinical indication were not considered for this study. For histological evaluation, kidney sections were stained with H&E, periodic acid-Schiff (PAS), and silver methenamine (Jones). All biopsies were centrally scored by pathologists dedicated to transplant pathology following the same standard procedures. The severity of chronic histological lesions was semiquantitatively scored according to the Banff categories for mesangial matrix expansion ("mm"), tubular atrophy ("ct"), vascular intimal thickening/arteriosclerosis ("cv"), interstitial fibrosis ("ci"), arteriolar hyalinosis ("ah"), and transplant glomerulopathy ("cg"). The team involved in the computational analysis was not informed about any clinical information until the end of the study, when computational and clinical data were matched. The histological evaluation was independent of the computational analysis, and the pathologist was not informed about the results of the transcriptional analysis.

Tissue storage and RNA extraction. For each renal allograft biopsy included in this study, at least half a core was immediately stored on Allprotect Tissue Reagent (Qiagen Benelux BV) and, after incubation at 4°C for at least 24 hours and a maximum 72 hours, stored locally at –20°C, until shipment to the Laboratory of Nephrology of the KU Leuven. We performed RNA extraction using the Allprep DNA/RNA/miRNA Universal Kit (Qiagen Benelux BV) on a QIAcube instrument (Qiagen Benelux BV). The quantity (absorbance at 260 nm) and purity (ratio of the absorbance at 230, 260, and 280 nm) of the RNA isolated from the biopsies were measured using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Life Technologies Europe BV). Before library preparation, RNA integrity was verified by high sensitivity RNA ScreenTape analysis (Agilent Technologies). Five samples were discarded on the basis of suboptimal mRNA quality.

Library preparation, RNA-seq, and data processing. Library preparation and RNA-seq were performed in 2 batches. The first batch consisted of a pilot study with 3 patients (12 samples), and the second batch included the rest of the study population. To avoid biases related to batch effects, the samples included in the pilot study were excluded from the pseudotime analyses. The library was prepared with Clontech SMARTer technology at the Genome Technology Access Center of the Washington University (St. Louis, Missouri, USA). The sequencing was performed in the same laboratory using the HiSeq 3000 system on the Illumina platform, with a target of 30 M reads per sample. The reads were aligned to the Ensembl top-level assembly with STAR version 2.0.4b. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 1.4.5. Transcript counts were produced by Sailfish version 0.6.3. Sequencing performance was assessed for total number of aligned reads, total number of uniquely aligned reads, genes and transcripts detected, ribosomal fraction, known junction saturation, and read distribution over known gene models with RSeQC version 2.3.

LMM. To examine the gene expression variation that can be explained by the variables of interest, a LMM was implemented with the R package lme4. After excluding genes with a maximum RPKM < 0.1, RPKM values were transformed as $\log_{10}(\text{RPKM} + 0.01)$ before being used in the LMM model. Gene expression was modeled by variables of interest as random effects plus an intercept considered as fixed effects. In R syntax convention, the formula is: GeneExp (RPKM) = $(1 | \text{Variable}_1) + (1 | \text{Variable}_2) + \varepsilon$, with 1 representing the intercept and ε representing the residual term. Time (PRE, POST, 3 and 12 months) and individual were defined as random effects. Variances were estimated by restricted maximum likelihood (REML) estimators. Then, the variances explained by the 2 random effects were normalized by the total variance (sum of each variance explained by the random effect and the residual variance), which were used as the *x* axis and *y* axis coordinates.

Pseudotime analysis. All pseudotime analyses were performed using the Monocle workflow in R (13). First, genes with minimum RPKM greater than 0.1 and expressed in at least 40 samples were selected as expressed genes. RPKM values were then transformed into mRNA per sample (RPS) by Census (included in Monocle 2) (13). RPS was equivalent to mRNA per cell (RPC); we named it RPS since our data was bulk RNA-seq instead of single-cell RNA-seq. The minimum expression level for considering a gene in Monocle was 0.1; a negative binomial distribution was used. The differential expression analysis in Monocle was based on the time point of biopsy or on the classification in categories, as indicated in the text. Cell state ordering and gene expression pattern plots were generated with the codes included in the Monocle package.

Comparison of mouse and human data. The experimental model was previously characterized (7). Briefly, 10- to 12-week-old, 25–28g C57BL/6CN male mice, purchased from the Charles River Laboratory, were anesthetized with an i.p. injection of a ketamine/xylazine (105 mg ketamine/kg; 10 mg xylazine/kg). Body temperature was maintained at 36.5°C–37°C throughout the procedure. The kidneys were exposed by a midline abdominal incision, and both renal pedicles were clamped for 21 minutes using nontraumatic microaneurysm clips (Roboz Surgical Instrument Co.). Restoration of blood flow was monitored by the return of normal color after removal of the clamps. All the mice received i.p. 1 ml of normal saline at the end of the procedure. Sham-operated mice underwent the same procedure, except for clamping of the pedicles. RNA was extracted from whole renal tissue with an RNeasy kit (Qiagen) and was provided to the USC Epigenome Center's Data Production Core Facility for library construction and sequencing. Library construction was carried out using the Illumina TruSeq RNA Sample Prep kit v2 through polyA selection. Libraries were applied to an Illumina flow cell at a concentration of 16 pM on a version 3 flow cell and run on the Illumina HiSeq 2000. Final file formatting, demultiplexing, and fastq generation were carried out using CASAVA version 1.8.2. The sequencing data were aligned to mm10 genome assembly with STAR aligner. The interspecies comparison was based on orthologous genes between human and mouse. The orthology information was downloaded from Biomart (https://www.ensembl.org/biomart) by selecting human data set hg38 and adding mouse orthology information. The Genome Assembly version used were GRCh38.p12 for human and GRCm38.6 for mouse.

Comparison of kidney and liver. The liver dataset was downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87487) and was analyzed in analogy to the kidney data (20).

Software. t-SNE, differential gene expression analysis, gene correlation analyses, the Monocle workflow, the community finding analysis, and the network analysis were performed with the PIVOT platform, developed by the Kim Lab, University of Pennsylvania (http://kim.bio.upenn.edu/software/pivot.shtml). Violin plots were generated on the PIVOT platform and visualized on Plotly (https://plot.ly). Box plots and histograms were generated with Prism 7. Gene enrichment analyses were performed with ToppFun (https://toppgene.cchmc.org).

Webpage. An interactive website was developed for a convenient interrogation of how genes of interest changed along pseudotime in human kidney injury early response (PRE vs. POST) and how genes behaved in the long run — a pseudotime trajectory constructed from POST to 3- and 12-month samples.

Statistics. All gene expression levels were normalized and quantified by number of reads per kilobase per million mapped reads (RPKM) and transformed into RPS (mRNA per sample) for pseudotime analyses. Dimension reduction analysis was performed with t-SNE by Rtsne package in R. Gene correlation analysis and feature correlation heatmaps were performed on PIVOT (16). Differential gene expression analyses were performed with EdgeR (37), by applying an exact test, with an FDR < 0.05. RPKM, RPS, and eGFR values were compared by Mann-Whitney U tests. Categorical analyses were performed with Fisher's exact or χ^2 tests. Venn diagrams were generated by the R package VennDiagram, and the significance of enrichment was calculated by hypergeometric tests. The community finding analysis was performed on PIVOT (25). The network analysis was performed based on genes differentially expressed in PRE and POST samples. The analysis was performed in PIVOT by using the STRING protein-to-protein interaction database and the Regnetwork regulatory network database (12). Gene Ontology (GO) analyses were performed on ToppGene (38).

Study approval. The patients received a kidney transplantation at the University of Leuven. All patients gave written informed consent, and the study was approved by the Ethical Review Board of the University Hospitals of Leuven (S53364 and S59572). Mouse handling and husbandry and all surgical procedures were performed according to guidelines issued by the IACUC at USC.

Author contributions

PEC, MN, and APM contributed to conception and experimental design. JL performed IRI surgeries. PEC and JL performed experimental data acquisition and analysis. MN performed human data acquisition. PEC, BS, and LC performed human data analysis. PEC and APM prepared the manuscript, incorporating comments from other authors.

1. Melé M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015;348(6235):660–665.
2. Letai A. Functional precision cancer medicine-moving beyond pure genomics. *Nat Med*. 2017;23(9):1028–1035.
3. Eming SA, Wynn TA, Martin P. Inflammation and metabolism in tissue repair and regeneration. *Science*. 2017;356(6342):1026–1030.
4. Netea MG, et al. A guiding map for inflammation. *Nat Immunol*. 2017;18(8):826–831.
5. Levey AS, Coresh J. Chronic kidney disease. *Lancet*. 2012;379(9811):165–180.
6. Forbes SJ, Rosenthal N. Preparing the ground for tissue regeneration: from mechanism to therapy. *Nat Med*. 2014;20(8):857–869.
7. Liu J, et al. Molecular characterization of the transition from acute to chronic kidney injury following ischemia/reperfusion. *JCI Insight*. 2017;2(18):e94716.
8. Turajlic S, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*. 2018;173(3):595–610.
9. Puchalski RB, et al. An anatomic transcriptional atlas of human glioblastoma. *Science*. 2018;360(6389):660–663.
10. Yu MK, Ma J, Fisher J, Kreisberg JF, Raphael BJ, Ideker T. Visible Machine Learning for Biomedicine. *Cell*. 2018;173(7):1562–1565.
11. Webb S. Deep learning for biology. *Nature*. 2018;554(7693):555–557.
12. Rackham OJ, et al. A predictive computational framework for direct reprogramming between human cell types. *Nat Genet*. 2016;48(3):331–335.
13. Trapnell C, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381–386.
14. Habib N, et al. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science*. 2016;353(6302):925–928.
15. Kalogeris T, Baines CP, Krenz M, Korthuis RJ. Cell biology of ischemia/reperfusion injury. *Int Rev Cell Mol Biol*. 2012;298:229–317.
16. Zhu Q, Fisher SA, Dueck H, Middleton S, Khaladkar M, Kim J. PIVOT: platform for interactive analysis and visualization of transcriptomics data. *BMC Bioinformatics*. 2018;19(1):6.
17. Damman J, et al. Hypoxia and Complement-and-Coagulation Pathways in the Deceased Organ Donor as the Major Target for Intervention to Improve Renal Allograft Outcome. *Transplantation*. 2015;99(6):1293–1300.
18. Kumar S, et al. Sox9 Activation Highlights a Cellular Pathway of Renal Repair in the Acutely Injured Mammalian Kidney. *Cell Rep*. 2015;12(8):1325–1338.
19. Fujiu 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.
20. Sosa RA, et al. Early cytokine signatures of ischemia/reperfusion injury in human orthotopic liver transplantation. *JCI Insight*. 2016;1(20):e89679.
21. Chen X, Xu C, Zhang F, Ma J. Comparative analysis of expression profiles of chemokines, chemokine receptors, and components of signaling pathways mediated by chemokines in eight cell types during rat liver regeneration. *Genome*. 2010;53(8):608–618.
22. Li Q, et al. The ileal FGF15/19 to hepatic FGFR4 axis regulates liver regeneration after partial hepatectomy in mice. *J Physiol Biochem*. 2018;74(2):247–260.
23. Chang NJ, et al. Genome-wide gene expression profiling of ischemia-reperfusion injury in rat kidney, intestine and skeletal muscle implicate a common involvement of MAPK signaling pathway. *Mol Med Rep*. 2015;11(5):3786–3793.
24. Arner E, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*. 2015;347(6225):1010–1014.
25. Darmanis S, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA*. 2015;112(23):7285–7290.
26. Thomas JL, et al. Hypoxia-inducible factor-1α activation improves renal oxygenation and mitochondrial function in early chronic kidney disease. *Am J Physiol Renal Physiol*. 2017;313(2):F282–F290.
27. Wernig G, et al. Unifying mechanism for different fibrotic diseases. *Proc Natl Acad Sci USA*. 2017;114(18):4757–4762.
28. Zhang S, et al. Effects of C1 inhibitor on endothelial cell activation in a rat hind limb ischemia-reperfusion injury model [published online ahead of print January 27, 2018]. *J Vasc Surg*. https://doi.org/10.1016/j.jvs.2017.10.072.
29. Bhargava P, Schnellmann RG. Mitochondrial energetics in the kidney. *Nat Rev Nephrol*. 2017;13(10):629–646.

30. Devalaraja-Narashimha K, Padanilam BJ. PARP-1 inhibits glycolysis in ischemic kidneys. *J Am Soc Nephrol*. 2009;20(1):95–103.

31. Ke B, Fan C, Yang L, Fang X. Matrix Metalloproteinases-7 and Kidney Fibrosis. *Front Physiol*. 2017;8:21.

32. Chu AY, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun*. 2017;8(1):1286.

33. Moon KR, Stanley JS III, Burkhardt D, van Dijk D, Wolf Krishnaswamy S. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Curr Opin Syst Biol*. 2018;(7):36–46.

34. O'Connell PJ, et al. Biopsy transcriptome expression profiling to identify kidney transplants at risk of chronic injury: a multi-centre, prospective study. *Lancet*. 2016;388(10048):983–993.

35. Lemoine S, et al. Dose and timing of injections for effective cyclosporine A pretreatment before renal ischemia reperfusion in mice. *PLoS One*. 2017;12(8):e0182358.

36. GBD 2015 Mortality Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459–1544.

37. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*. 2008;9(2):321–332.

38. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37(Web Server issue):W305–W311.