Nephrol Dial Transplant, 2024, 39, 1044–1047
https://doi.org/10.1093/ndt/gfae030
Advance access publication date: 6 February 2024

RESEARCH LETTER

A comparative scRNAseq data analysis to match mouse models with human kidney disease at the molecular level

Kathrien Abdank[1,*], Sena Zeynep Cetin[1,*], Amin Abedini[2,3], Katalin Susztak[2,3], Kai-Uwe Eckardt[1] and Michael S. Balzer[1,4]

[1]Department of Nephrology and Medical Intensive Care, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Renal, Electrolyte, and Hypertension Division, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[3]Institute for Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[4]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Biomedical Innovation Academy, BIH Charité Clinician Scientist Program, Berlin, Germany
Correspondence to: Michael S. Balzer; E-mail: michael-soeren.balzer@charite.de; Twitter: @ms_balzer; GitHub: https://github.com/ms-balzer
*Equal contribution

Watch the video of this contribution at https://academic.oup.com/ndt/pages/author_videos

To the Editor,

The advent of single-cell RNA-sequencing (scRNA-seq) has unveiled novel cell states in preclinical kidney disease models. Notably, recent ischemia reperfusion injury (IRI) studies spotlighted maladaptive repair cell states in specific tubular epithelial cells distinguished by *Vcam1* or *Gsdmd* expression [1, 2]. However, the clinical relevance of these states across the acute kidney injury (AKI) to chronic kidney disease (CKD) spectrum remains uncertain. Furthermore, the absence of a standardized scRNA-seq analysis approach contributes to methodological diversity among research labs [3]. Despite varied analytical approaches, assuming authors accurately extract the true biology from their models, a comparative examination of distilled results (e.g., lists of differentially expressed genes, DEGs) promises to address key questions: (i) Where on the AKI-to-CKD spectrum does a rodent model most accurately map pathology? (ii) How faithfully do murine studies mirror human AKI and CKD phenotypes? (iii) Are there predictive parameters in scRNA-seq studies for improved or diminished mapping? (iv) What prerequisites should readers bear in mind when interpreting scRNA-seq studies?

To answer these questions, we devised a straightforward analysis framework: We assigned the phenotypes health, AKI, recovery, or CKD to respective sample groups of six recent scRNA-seq studies representing 641,212 kidney cells (three human [4–6], three mouse [1, 2, 7]) based on authors' annotations (Table 1, Fig. 1a). Subsequently, we used the provided reduced biological output data (DEG lists defining cells from these four phenotypes) as input for our analyses. After a 1:1 mapping of mouse genes to orthologous human genes (Data S1, see online supplementary material), we examined commonalities and specificities at both the DEG and pathway levels. We assessed phenotypic similarity through Jaccard indices of DEG lists and hierarchical replicability analysis on shared core genes as an orthogonal method. Finally, we further condensed original DEG lists to pathways by calculating their enrichment levels individually for each phenotype (health, AKI, recovery, CKD) within a given dataset, elucidating commonalities and specificities across species (for methodological details, see Data S1, in online supplementary material).

As anticipated in the comparative analysis of data across different laboratories, the number of significantly upregulated DEGs representing health, recovery, AKI, and CKD phenotypes exhibited substantial heterogeneity across the six datasets (ranging from 110 to 1,321, Fig. 1b). Of note, the method for DEG computation did not influence the results substantially, as both Wilcoxon and MAST tests produced highly congruent results (Fig. S1, see online supplementary material) with identical lists of top 100 DEGs. The top 100 DEGs displayed generally low intra-phenotype overlap, underscoring high dataset specificity (Fig. S2, see online supplementary material). Assessing phenotype-specific similarity of DEG lists through Jaccard indices revealed significant overlap between AKI and CKD states from both murine and human datasets across various labs (Fig. S3a, see online supplementary material). However, there was no clear demarcation of health/recovery/AKI/CKD phenotypes, once again emphasizing high dataset specificity at the DEG level.

Moving beyond the DEG level, we employed hierarchical replicability analysis as an orthogonal approach. The rationale was that if a disease phenotype's biological identity is rooted in the transcriptome, understanding its expression features in one dataset would facilitate the identification of similar phenotypes in other datasets [8]. Quantifying phenotype replicability with a neighbor voting approach across 13,012 core genes shared among all six datasets, we demonstrated generally high concordance between AKI and CKD (Fig. S3b, Data S2, see online supplementary material). Furthermore, healthy phenotypes clustered together consistently across species. However, we noted outliers representing CKD phenotypes from two studies that co-clustered with healthy phenotypes. To consolidate the results of these two orthogonal approaches (Jaccard index and hierarchical replicability analyses), we calculated correlation coefficients between both methods (Fig. 1c). The findings indicated (i) a generally high consensus between methodologies and (ii)

Received: December 12, 2023; Editorial decision: February 2, 2024

Table 1: Phenotype assignment from original studies as per the authors' annotations.

Study	Accession	Group	Assigned phenotype
Abedini *et al.* [4]	https://susztaklab.com	Control	Health
		CKD	CKD
		DKD	CKD
Balzer *et al.* [2]	GSE180420	Control	Health
		IRI short 1d	AKI
		IRI short 3d	AKI
		IRI short 14d	Recovery
		IRI long 1d	AKI
		IRI long 3d	AKI
		IRI long 14d	CKD
Doke *et al.* [7]	GSE182256	Control	Health
		UUO	CKD
Hinze *et al.* [5]	GSE210622	Control	Health
		COVID AKI	AKI
		Non-COVID AKI	AKI
Kirita *et al.* [1]	GSE139107	Control	Health
		IRI 4h	AKI
		IRI 12h	AKI
		IRI 2d	AKI
		IRI 14d	CKD
		IRI 6wks	CKD
Lake *et al.* [6]	https://atlas.kpmp.org	LD	Health
		AKI	AKI
		CKD	CKD

AKI, acute kidney injury; CKD, chronic kidney disease; DKD, diabetic kidney disease; IRI, ischemia reperfusion injury; LD, living donation; UUO, unilateral ureteral obstruction.

improved clustering of dataset-derived health/recovery/AKI/CKD phenotypes, suggesting alignment according to biology across species. Nevertheless, two outliers among CKD phenotypes (Kirita, Lake) were observed to cluster with health phenotypes (Fig. 1c). This observation could be attributed to their comparatively high cell fractions of tubular epithelial cells (proximal tubule, PT; loop of Henle, LOH; and distal nephron) (Fig. 1d), which were considerably higher than those of CKD phenotypes in other datasets and fell within the range of health phenotypes (Fig. 1d and e). Importantly, to account for potential influences of different pre-processing pipelines across individual datasets, cell annotations were validated with an orthogonal unified mapping approach using Azimuth [9] with a kidney tissue reference, which yielded similar cell fractions across phenotypes when compared to original authors' annotations (Figs S4–S5, see online supplementary material).

Finally, we analysed the alignment of phenotypes after further reduction of their respective DEG lists to the biological pathway level. Taking the top 100 DEGs per phenotype from each dataset, we uniformly computed enrichment scores across diverse curated databases, including Hallmark, Gene Ontology (GO) biological processes, BioCarta, KEGG, PID, Reactome, and WikiPathways (Data S3, see online supplementary material). Despite significant heterogeneity in the overall number of enriched pathways (Fig. S6a, see online supplementary material), mean enrichment was similar across datasets, phenotypes, and species, indicating absence of systematic bias (Fig. S6b, see online supplementary material). Importantly, enriched pathways exhibited strong commonalities across datasets and formed distinct clusters based on phenotypes (Fig. 1f). For example, all health phenotypes displayed specific enrichment of ion transport processes linked to proximal tubule transport of anions and sodium via solute carriers (SLC), featuring shared genes such as *SLC7A8*, *SLC22A6*, *SLC4A4*, *SLC34A1*, *SLC6A19*, *SLC5A2*, etc. Similarly, the recovery phenotype shared numerous pathways with both health and AKI phenotypes, signifying its transitional state. Early post-AKI time points enriched specifically for cell matrix adhesion, substrate junction organization, focal adhesion assembly, and epithelial-mesenchymal-transition (*EFNA5*, *VCL*, *SDC4*, *ROCK2*, *SMAD3*, *FERMT2*, *VIM*, *HES1*, *IL32*, *MMP7*, *ITGB8*), while later post-AKI time points enriched for antigen presentation, allograft rejection, and neutrophil degranulation, indicative of immune system activation. CKD phenotypes exhibited robust enrichment for hypoxia, TNFα/NFκB signaling, UV response, interleukin 18 (IL18) signaling, and viral defense mechanisms (*HIF1A*, *CITED2*, *NFKBIA*, *TNF*, *IL1B*, *CCL3*, *CEBPB*, *FOS*, *JUNB*, *PTGS2*). The sensitivity of our pathway analysis was underscored by the detection of pronounced hypoxia enrichment in health phenotype samples derived from living donor kidneys (Lake). This finding aligns with the logical context that control kidneys in this specific study underwent arterial clamping and organ removal from living donors.

Collectively, these findings indicate that DEG lists from individual scRNA-seq studies predominantly reflected reductionist, dataset-specific insights, lacking inherent potential for seamless cross-dataset or cross-species generalization *per se*. Conversely, a more comprehensive consideration of transcriptome-wide alterations and biological pathway enrichment enhanced the ability to capture the underlying biology of phenotypes. This, in turn, facilitated proper inter-dataset comparisons, as evidenced by the robust cross-species clustering of phenotypes.

It's important to acknowledge the limitations of our approach. We did not aim to systematically account for inherent dataset differences, whether on a technical level [sampling method (wedge vs. core biopsy), sampling location (cortex vs. medulla), cell preparation assay (droplet vs. well-based assay, 3′ vs. 5′ kits, cells vs. nuclei), sequencing depth, etc.], analytical level (ambient RNA correction, doublet removal, clustering resolution, etc.) or a biological level (appropriateness of murine models, cross-species immunological differences, cell fraction representation, sex disparities, etc.). Nevertheless, our approach underscores the utility of accepting the direct output of original studies at face value for comparative purposes, offering valuable insights into underlying biology. Even unbiased systematic approaches, such as the recently reported scRNA-seq integration of 18 commonly used mouse kidney disease models starting from raw data [10], may struggle to overcome inherent biological inter-dataset differences. Moreover, the availability of raw data is not consistently guaranteed. Lastly, striving to adjust for all variables may not always be desirable; for instance, including IRI kidneys from female mice could be counterproductive due to their known injury protection.

In conclusion, our straightforward analytical framework has proven effective in accurately positioning pre-clinical models along the human AKI-to-CKD spectrum, showcasing robust biological pathway concordance across species. Selecting appropriate pre-clinical models that closely mimic human disease enhances translational relevance. Notably, our approach identified problematic datasets and samples that deviated from anticipated patterns. Crucially, we identified cell fractions and their changes as pivotal parameters, underscoring their significance as essential prerequisites in interpreting the representativeness of kidney scRNA-seq datasets.

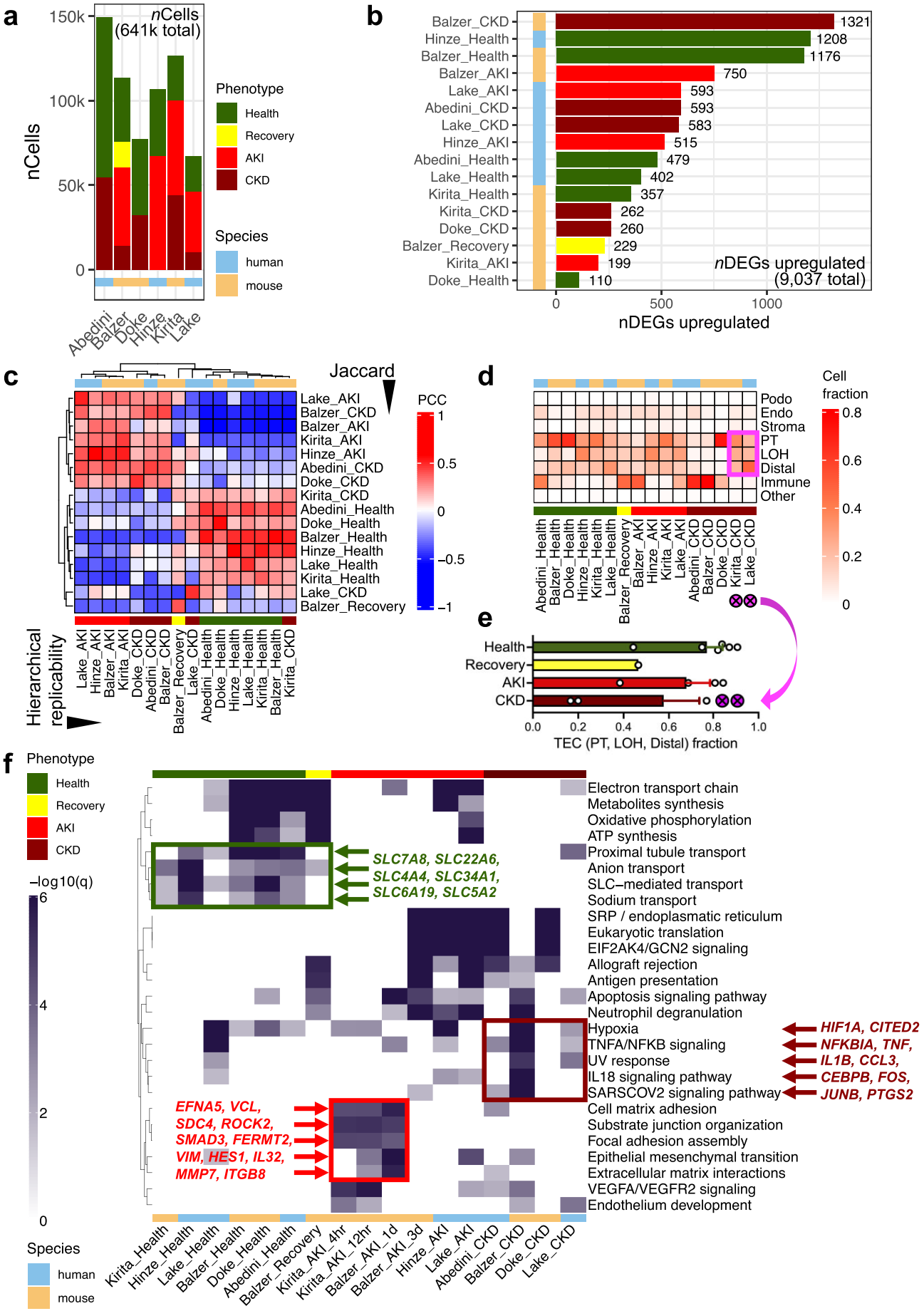

Figure 1: Cross-species and cross-scRNA-seq dataset commonalities and differences along the AKI-to-CKD spectrum. (**a**) Number of cells contributed by health, recovery, AKI, and CKD phenotypes, respectively, across six recent scRNA-seq datasets (three human, three mouse), representing >641k kidney cells; x axis denotes first authors of the original papers; color denotes phenotype assigned as per original authors' description (Table 1) and species, respectively. AKI, acute kidney injury; CKD, chronic kidney disease. (**b**) Number of significantly upregulated differentially expressed genes (DEGs) per phenotype across datasets; color annotation as in (a). (**c**) Pearson correlation coefficients (PCC) demonstrate generally good concordance of phenotypes across species upon correlation of two separate methods analysing DEG similarity (rows: Jaccard indexing; columns: hierarchical replicability analysis); color annotation as in (a). (**d**) Cell fractions of podocytes (Podo), endothelial (Endo), stroma (Stroma), proximal tubule (PT), loop of Henle (LOH), distal nephron tubular (Distal), immune (Immune), and other (Other) cells, respectively, across individual phenotypes and datasets; magenta highlights refer to two outlier CKD phenotypes with high tubular epithelial cell fractions, see (e); color annotation as in (a). (**e**) Tubular epithelial cell (TEC) fractions across health, recovery, AKI, and CKD phenotypes, respectively; bars with error bars represent means ± SEM; two outlier CKD phenotypes are highlighted in magenta; color annotation as in (a). (**f**) Biological pathway enrichment by phenotype across datasets and species, heatmap color refers to enrichment specificity (-log10(q) value) of top enriched pathways of Hallmark, Gene Ontology (GO) biological processes, BioCarta, KEGG, PID, Reactome, and WikiPathways curated databases; post-AKI time points were given individually where possible (Balzer, Kirita); exemplary representative genes shared among a specific phenotype across datasets and species are called out for health, AKI, and CKD phenotypes, respectively; color annotation as in (a).

SUPPLEMENTARY DATA

Supplementary data are available at *ndt* online.

ACKNOWLEDGEMENTS

Computation was performed on the HPC for Research cluster of the Berlin Institute of Health.

FUNDING

S.Z.C. is supported by German Academic Exchange Service (Deutscher Akademischer Austausch Dienst, DAAD) and Sonnenfeld Foundation grants; K.S. is supported by NIH grants DK076077, DK087635, DK105821, and DK132630; M.S.B. is supported by German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, BA 6205/2-1) and Else Kröner-Fresenius Foundation grants, as well as by the Berlin Institute of Health at Charité—Universitätsmedizin Berlin Clinician Scientist Program.

AUTHORS' CONTRIBUTIONS

M.S.B. designed and conceived the analytical framework. K.A. and S.Z.C. conducted single-cell analysis with input from A.A., K.S., and K.-U.E. M.S.B. supervised bioinformatics analysis. All authors discussed and commented on the results. M.S.B. wrote the manuscript and all authors edited and approved the final manuscript.

CONFLICT OF INTEREST STATEMENT

K.S. reports research support: AstraZeneca, Bayer, Boehringer Ingelheim, Calico, Genentech, Gilead, GSK, Jnana, Lilly, Maze, Merck, Novartis, Novo Nordisk, Regeneron, Variant Bio, and Ventus; advisory board membership: Jnana Therapeutics and Pfizer; consultancy: AstraZeneca, Bayer, GSK, Jnana Therapeutics, Maze, Novo Nordisk, Pfizer, and Ventus; patents: Jag1- and Notch-based targeting of chronic kidney disease; editorial board membership: *Cell Metabolism, eBioMedicine, Journal of the American Society of Nephrology, Journal of Clinical Investigation, Kidney International*, and *Med*; K.-U.E. reports research support: Amgen, Astra Zeneca Bayer, Evotec, Vifor; consultancy: Akebia, Astra Zeneka, Bayer, Otsuka; editoral board membership: *Kidney International*; M.S.B. reports consultancy: Boehringer Ingelheim; editorial board membership: *Journal of the American Society of Nephrology*. All other authors declare no competing interests.

REFERENCES

1. Kirita Y, Wu H, Uchimura K *et al.* Cell profiling of mouse acute kidney injury reveals conserved cellular responses to injury. *Proc Natl Acad Sci USA* 2020;**117**:15874–83. https://doi.org/10.1073/pnas.2005477117
2. Balzer MS, Doke T, Yang YW *et al.* Single-cell analysis highlights differences in druggable pathways underlying adaptive or fibrotic kidney regeneration. *Nat Commun* 2022;**13**:4018. https://doi.org/10.1038/s41467-022-31772-9
3. Balzer MS, Ma Z, Zhou J *et al.* How to get started with single cell RNA sequencing data analysis. *J Am Soc Nephrol* 2021;**32**:1279–92. https://doi.org/10.1681/ASN.2020121742
4. Abedini A, Ma Z, Frederick J *et al.* Spatially resolved human kidney multi-omics single cell atlas highlights the key role of the fibrotic microenvironment in kidney disease progression. *bioRxiv* 2022; 2022.2010.2024.513598, preprint: not peer reviewed.
5. Hinze C, Kocks C, Leiz J *et al.* Single-cell transcriptomics reveals common epithelial response patterns in human acute kidney injury. *Genome Med* 2022;**14**:103. https://doi.org/10.1186/s13073-022-01108-9
6. Lake BB, Menon R, Winfree S *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *Nature* 2023;**619**:585–94. https://doi.org/10.1038/s41586-023-05769-3
7. Doke T, Abedini A, Aldridge DL *et al.* Single-cell analysis identifies the interaction of altered renal tubules with basophils orchestrating kidney fibrosis. *Nat Immunol* 2022;**23**:947–59. https://doi.org/10.1038/s41590-022-01200-7
8. Crow M, Paul A, Ballouz S *et al.* Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* 2018;**9**:884. https://doi.org/10.1038/s41467-018-03282-0
9. Hao Y, Hao S, Andersen-Nissen E *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021;**184**:3573–87 e3529. https://doi.org/10.1016/j.cell.2021.04.048
10. Zhou J, Abedini A, Balzer M *et al.* Unified mouse and Human kidney single-cell expression atlas reveal commonalities and differences in disease states. *J Am Soc Nephrol* 2023;**34**:1843–62. https://doi.org/10.1681/ASN.0000000000000217

Received: December 12, 2023; Editorial decision: February 2, 2024