

# Analysis of individual patient pathway coordination in a cross-species single-cell kidney atlas

---

In the format provided by the  
authors and unedited

## **Supplementary Information Table of Content:**

I: Source Data Information

II: Supplementary Table Information

III: Supplementary Methods

IV: Supplementary Figures

## Source Data Information:

Source data Fig. 4

Source data Fig. 5

Source data Fig. 6

Source data Extended Data Fig. 4

Source data Extended Data Fig. 8

Source data Extended Data Fig. 10

## **Supplementary Tables:**

Supplementary Table 1: Multi-species dataset summary and sample information.

Supplementary Table 2: Conserved marker genes (MetaMarkers analysis).

Supplementary Table 3: Conserved fine-resolution clusters and marker genes.

Supplementary Table 4: Differentially expressed genes between species.

Supplementary Table 5: Conserved and species-specific cellular functions with a high prevalence of coordination changes.

Supplementary Table 6: Mouse KEGG Pathways (SISKA).

Supplementary Table 7: Extended human dataset summary and sample information.

Supplementary Table 8: Features of high prevalence across diseases (+ Quality Report example).

Supplementary Table 9: Features of high prevalence of coordination changes in ADPKD compared to a CKD reference.

Supplementary Table 10: Cancer hallmarks.



## Supplementary Methods:

### MetaMarkers analysis

We used the R package MetaMarkers<sup>23</sup> (v0.0.1) to identify conserved marker genes. To reduce the rate of false positive clusters, individual fine clusters were manually reclassified as “not conserved” when reported marker genes were inconsistent or non-specific. For all gene-based analysis we used objects with raw counts as input, followed by the tool’s default “cpm” normalization. Marker calculation was performed following default settings. Fine resolution cluster analysis considered level 1 annotations as higher hierarchical groups. We repeated the fine resolution MetaMarker analysis for each species independently, considering original datasets as batches. For the gene set (DEGO) analysis, since values were already based on normalized counts, no additional normalization was applied. To compute DEGO-based markers, the `make_meta_markers` function was set to `fc_threshold = 1.1`.

### MetaNeighbor analysis

We used the python implementation of MetaNeighbor<sup>82,29</sup> pyMN (v0.1.0). Using a mean score derived from the area under the receiver operating characteristic curve (AUROC), MetaNeighbor reflects the likelihood of correctly ranking a cell from the target group higher than a cell from a non-target group in terms of their gene expression correlation. We used pyMN to estimate the replicability of our cell types across datasets and species with project keys as batch information. Cell type replicability was estimated after gene selection with the `pymn.variableGenes()` function followed by the fast version of `pymn.MetaNeighborUS()`.

### Differential expression analysis between species

We used the `sc.tl.rank_genes_groups(method='wilcoxon')` function after normalization of counts per cell and log transformation. The analysis was performed on each cell type separately. Resulting differentially expressed genes with significantly higher expression in one species compared to the other two species were sorted based on the “scores” statistic.

### Differentially expressed gene sets

A gene ontology mouse database combining biological process, molecular function, and cellular compartment, was used as input to define functionally related gene sets<sup>35,94</sup>. One-to-one orthologous genes were converted to human gene symbols and non-one-to-one orthologous genes were removed. Analysis was performed on gene sets with at least 10 to 100 genes represented in our integrated atlas. Based on the modified Gene ontology database, the average expression of each gene set was calculated for each individual cell (Differentially expressed gene ontology terms; DEGOs). Thereby, the cell-gene matrix was transformed into a cell-gene set matrix. A similar approach was recently described<sup>73</sup>. We used log normalized counts for calculating the average gene expression. To improve the signal-to-noise ratio, this analysis was only performed on the initially defined 3000 highly variable genes. To improve the robustness of the approach, only gene sets containing more than 10 and up to 100 genes represented in the normalized variable gene matrix were considered. DEGOs were

visualized on spatial transcriptomics data of recently published 10x Genomics Visium transcriptomics datasets of mouse (downloaded from <https://www.10xgenomics.com/datasets/adult-mouse-kidney-ffpe-1-standard-1-3-0>) and rat<sup>36</sup>.

**Cell2location.** We used Cell2location<sup>95</sup> (v0.1), a Bayesian model for resolving cell types in spatial transcriptomics data. Unless stated otherwise, we used the default settings recommended for this analysis. We predicted cell type abundance in previously published 10X Genomics Visium transcriptomics datasets of human<sup>17</sup>, rat<sup>36</sup>, and mouse (see *Differentially expressed gene sets*). Visium datasets were translated into one-to-one orthologous human gene symbols after standard preprocessing in scanpy. The integrated multi-species atlas was used as a reference after subsetting to the respective species. The following set up was used before training on 250 epochs:

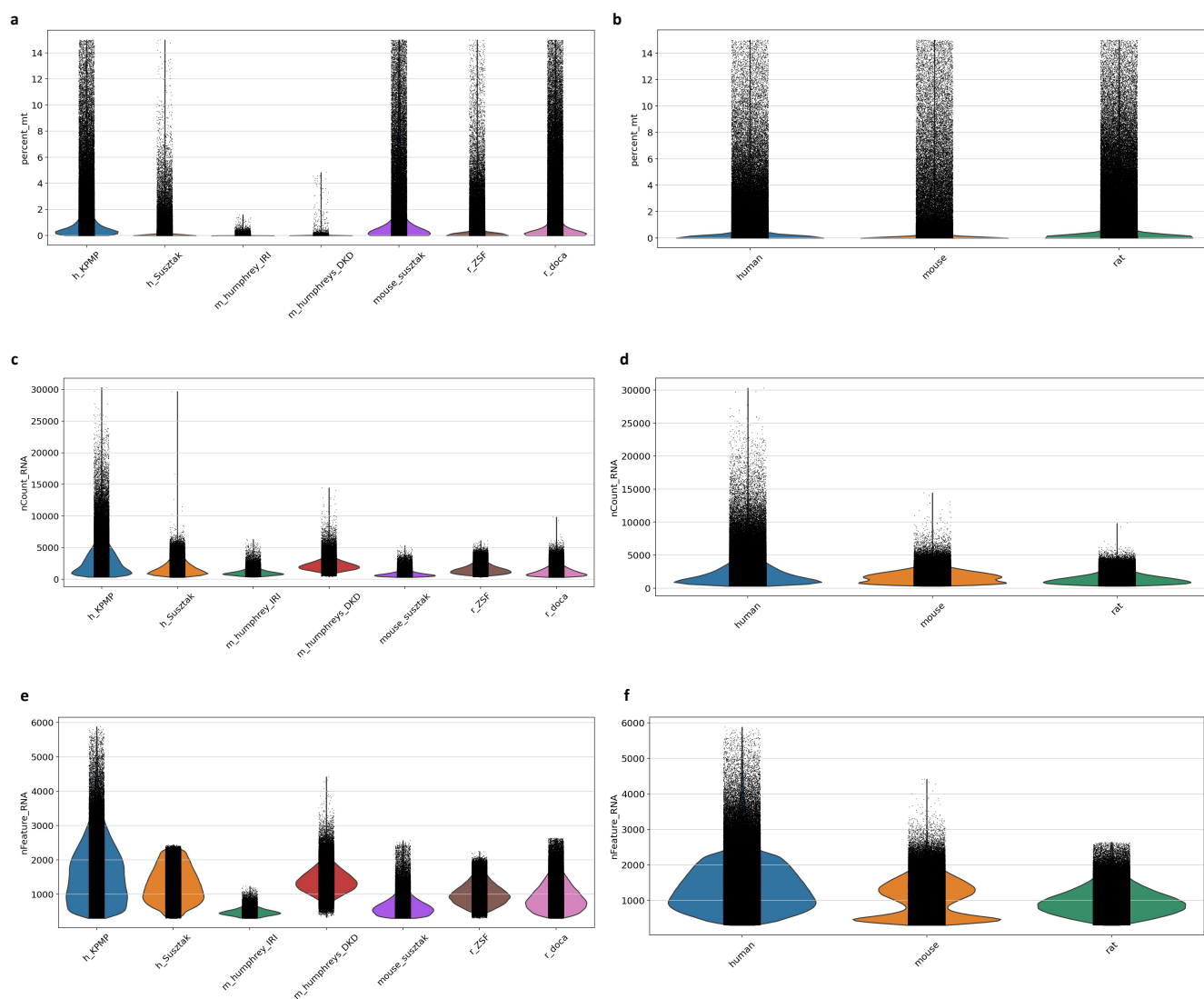
```
cell2location.models.RegressionModel.setup_anndata(adata=adata_ref,  
                                                    batch_key='orig_ident',  
                                                    labels_key='annotation_final_level1B',  
                                                    categorical_covariate_keys=['proj'])
```

Based on the intersection of visium data genes and the original highly variable atlas genes with  $N_{\text{cells\_per\_location}}=30$  and  $\text{detection\_alpha}=20$  we trained the final model for 30000 epochs on each visum slide separately. Cell abundance was plotted with a  $v_{\text{max}}$  of 99.2% quantile.

## NSCLC Analysis

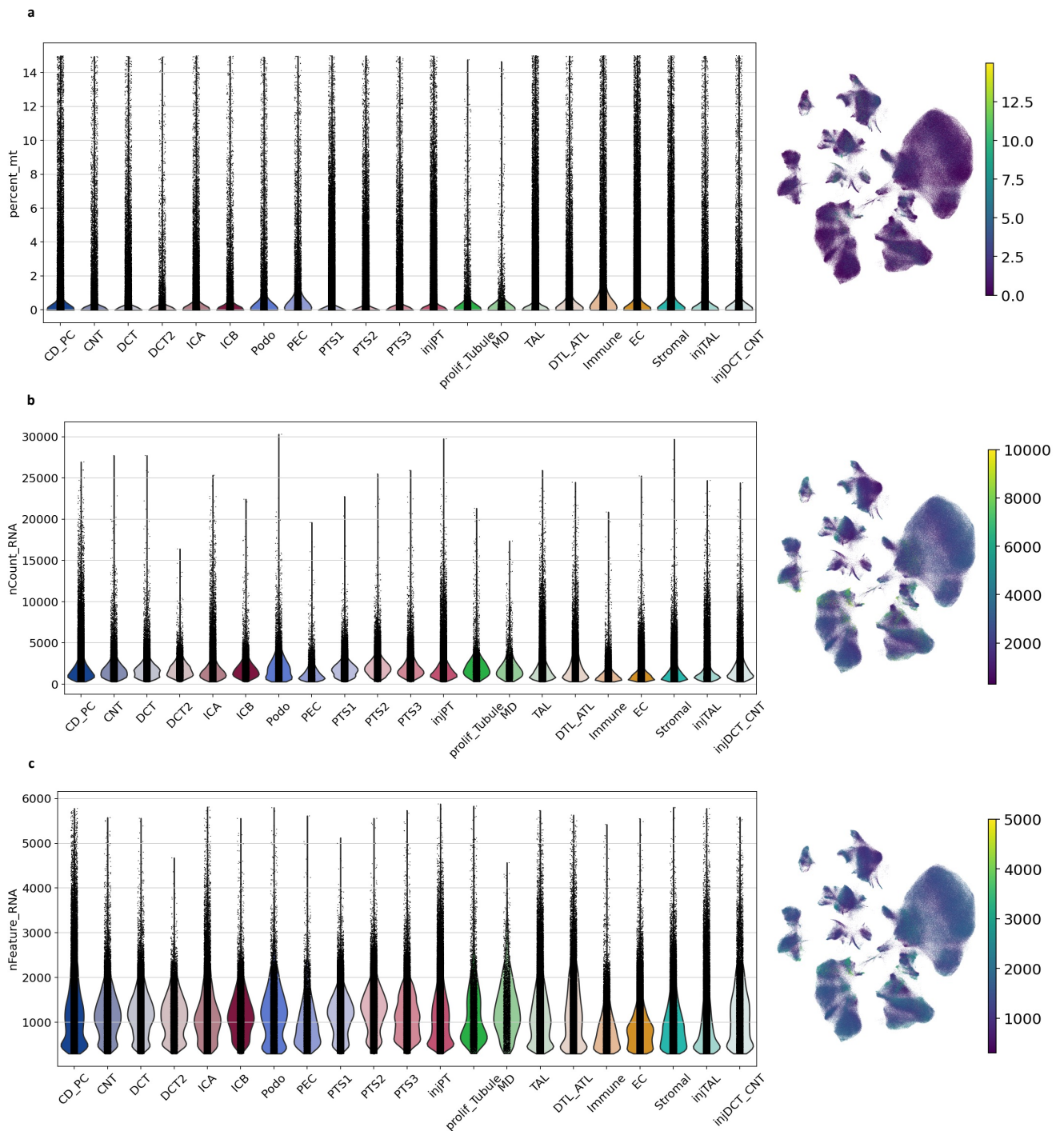
The lung core atlas<sup>46</sup> was downloaded in the anndata format from CELLxGENE<sup>96</sup> for further processing. Only datasets generated with the 10X Genomics platform were considered. Annotations "cell\_type\_major" were harmonized into one epithelial cell or cancer cluster. Anndata objects were converted to Seurat objects using the R package zellkonverter (V1.8.0). Only primary tumor samples were profiled. Samples from metastases were excluded from the analysis. The non-tumor reference consisted of non-tumor samples (healthy and COPD) or normal tumor adjacent samples. The tumor reference consisted of samples described as LUAD and uicc\_stage I based on the provided metadata. Since no tumor cells are present in the non-tumor reference, both tumor-derived cancer cells and non-malignant cells were profiled relative to non-tumor reference epithelial cells. As gene set input, we used the MSigDB C2 database<sup>44</sup>. CellSpectra was performed with a cell number threshold of 10, gene\_number\_threshold of 10, and using the filtering mode removing genes of zero expression.

## Supplementary Figures:



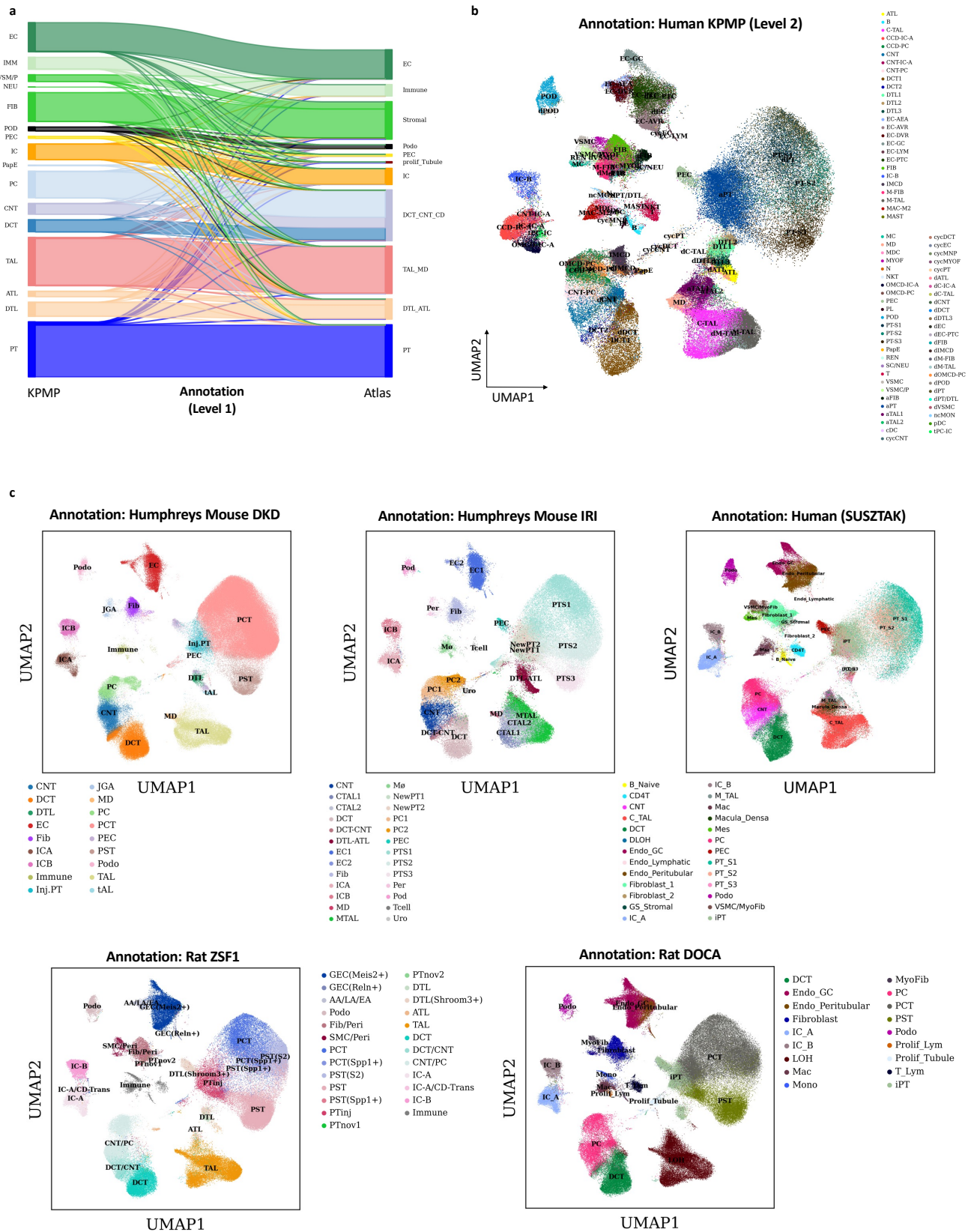
### Supplemental Figure 1: Multi-species dataset quality comparison.

Violin plots. The x-axis shows the sample (data source) while the y-axis shows different QC metrics including: percentage of mitochondrial genes (`percent_mt`), total number of counts (`nCount_RNA`), and number of detected features (`nFeature_RNA`) across different projects (**a, c, e**) and species (**b, d, f**), respectively.



**Supplemental Figure 2: Quality comparison across cell types within SISK 1.0.**

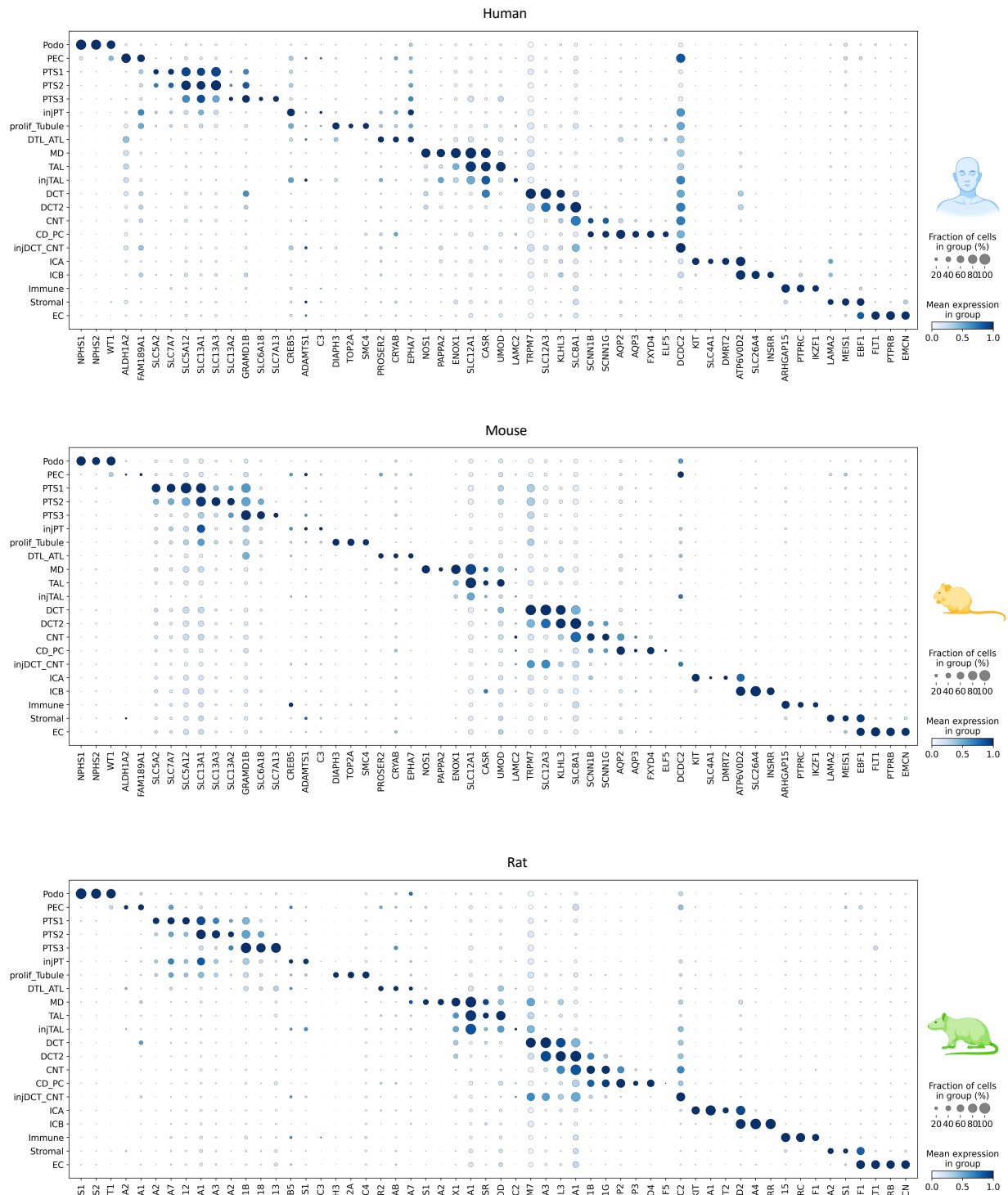
Violin plots and UMAP feature plots display quality markers for individual cells across cell types. The percentage of mitochondrial genes (a), total number of RNA counts (b), and number of detected features (c) for each cell type (x-axis) and all cells (UMAP). Values of the quality metrics are indicated on the y-axis (violin) or scale (UMAP). Principal cells of the collecting duct (CD\_PC), connective tubule (CNT), distal convoluted tubule (DCT and DCT2), intercalated cells (ICA and ICB), podocytes (Podo), parietal epithelial cells (PEC), proximal tubule (PTS1, PTS2, PTS3), injured proximal tubule (injPT), proliferating tubule (prolif\_Tubule), macula densa (MD), thick ascending limb (TAL), descending and ascending thin limb (DTL\_ATL), immune cells, endothelial cells (EC), stromal cells, injured TAL, and thick ascending and distal injured tubule (injTAL and injDCT\_CNT).



**Supplemental Figure 3: Cell type label in SISKA1.0 compared to their original annotations.**

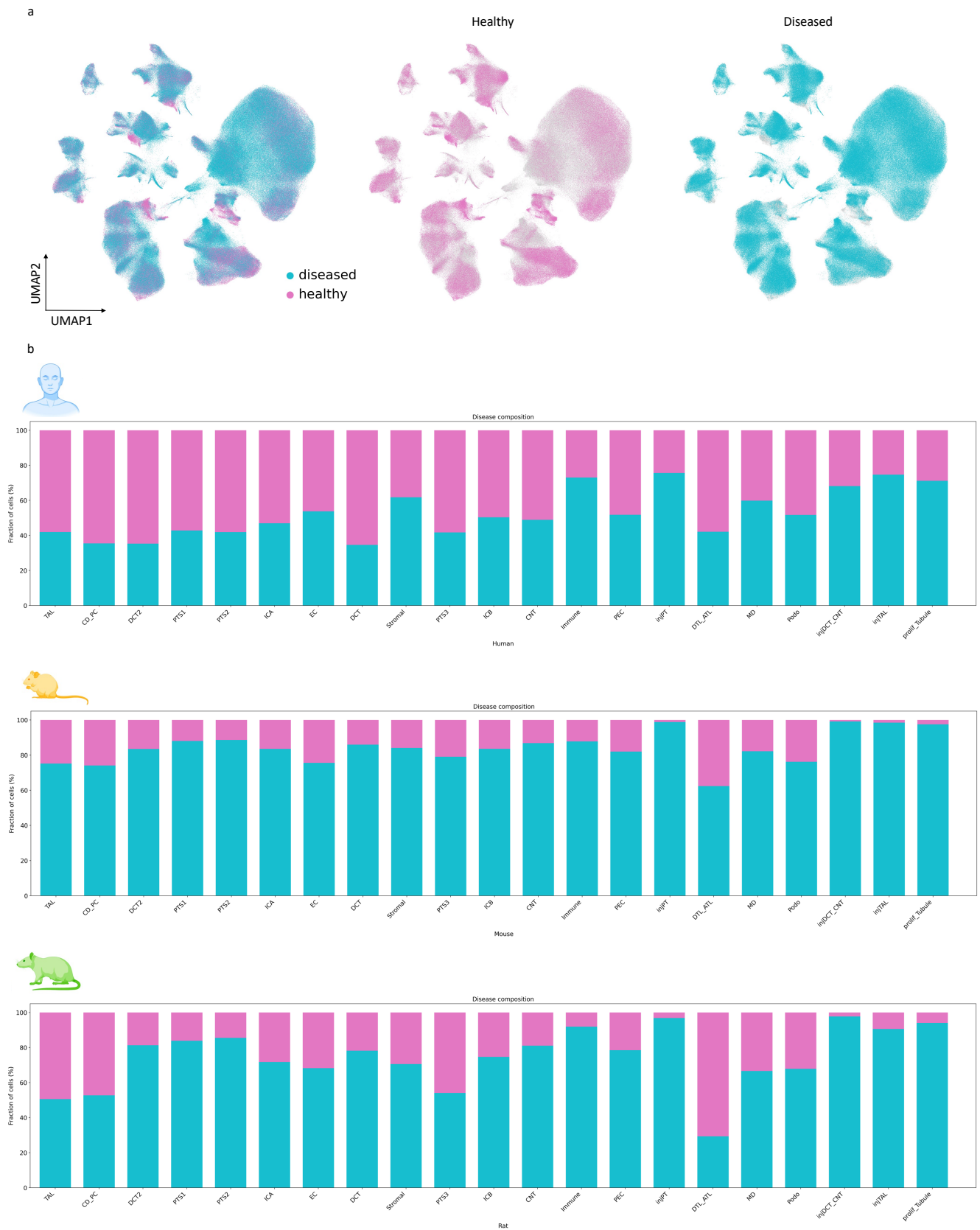
(a) A Sankey plot displaying the transition from original dataset (KPMP level 1 labels) on the left to new labels after integration on the right (SISKA1.0). (b) KPMP level 2 cell-type annotations, are visualized on the integrated SISKA UMAP embedding. (c) Similarly, the SISKA UMAP plots showing the original dataset annotations for rodent and human datasets.





**Supplemental Figure 4: Conserved marker gene expression across species within SISK1.0.**

Dot plot of conserved marker genes in human, mouse, and rat separately. Dot colors indicate the scaled mean expression per cell type, the size of the dot indicates the percentage of cells expressing the respective gene. Each row is a cell type while each column is an expression of one gene. *Principal cells of the collecting duct (CD\_PC), connective tubule (CNT), distal convoluted tubule (DCT and DCT2), intercalated cells (ICA and ICB), podocytes (Podo), parietal epithelial cells (PEC), proximal tubule (PTS1, PTS2, PTS3), injured proximal tubule (injPT), proliferating tubule (prolif\_Tubule), macula densa (MD), thick ascending limb (TAL), descending and ascending thin limb (DTL\_ATL), immune cells, endothelial cells (EC), stromal cells, injured TAL, and thick ascending and distal injured tubule (injTAL and injDCT\_CNT).*



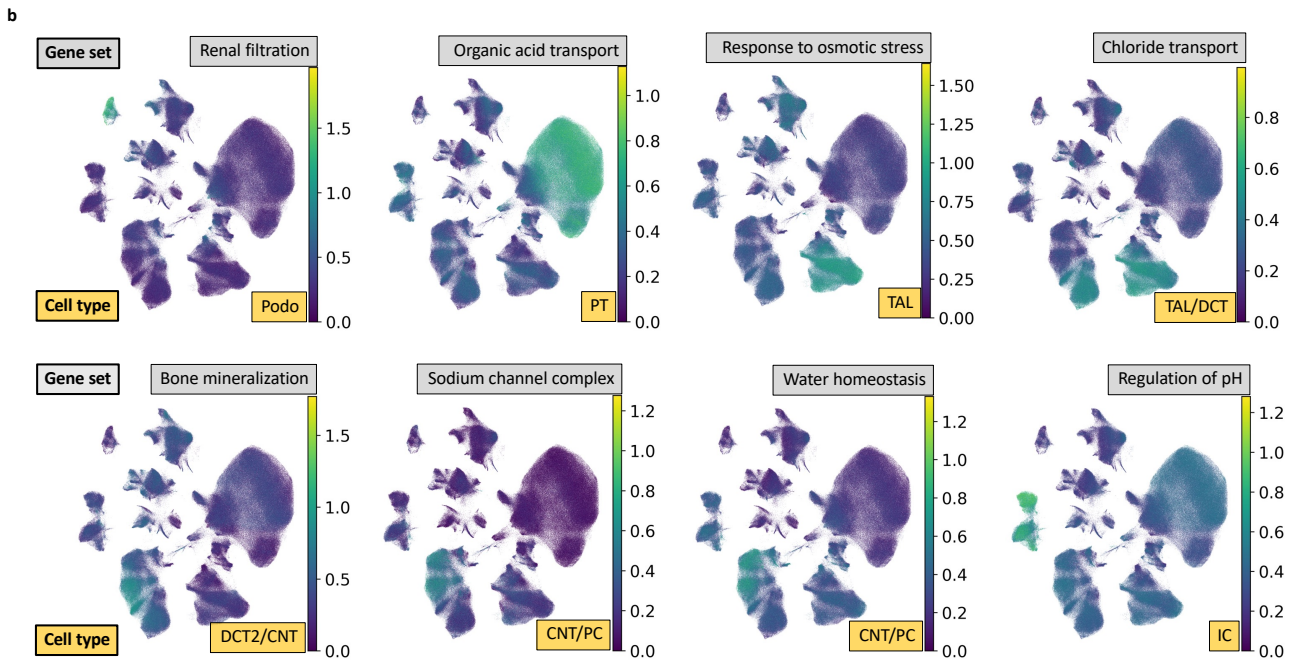
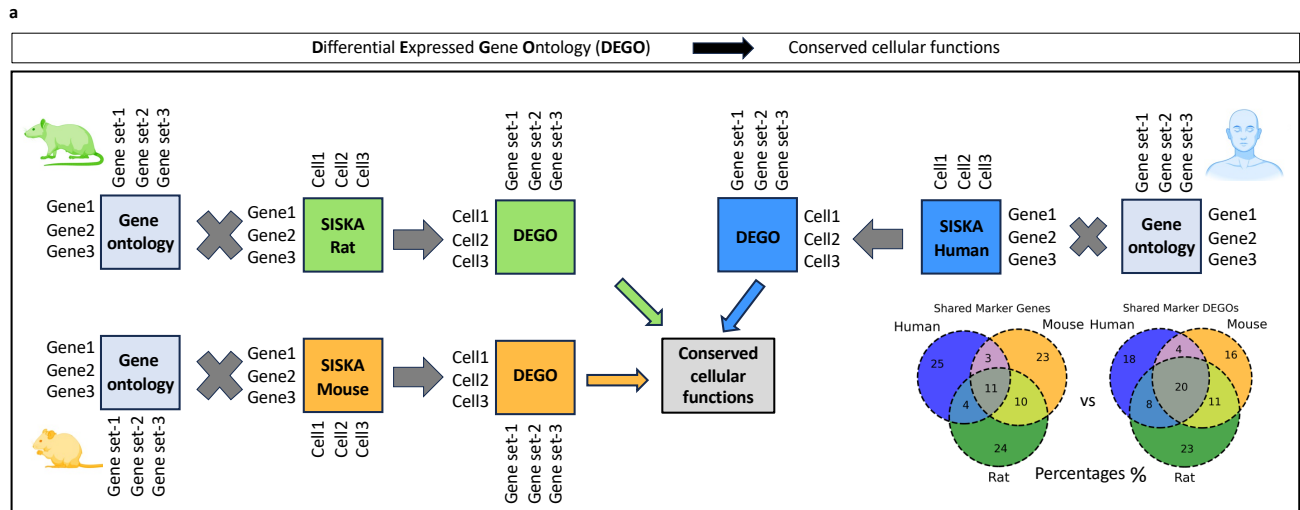
**Supplemental Figure 5: Relative fractions of healthy and disease within each cell type cluster.**

(a) Feature plot of healthy (pink) or diseased (blue) samples across species visualized in the UMAP embedding.  
 (b) Bar graphs of the percent of healthy and diseased cells (y-axis) within each cluster (x-axis). Each species is plotted separately.



Dot plots of the top 3 of significantly higher expressed genes (y-axis) within the same cell type in one species (x-axis, species order; human, mouse, rat). Dot colors indicate the scaled mean expression per cell type, the size of the dot indicates the percentage of cells expressing the respective gene. Genes were selected based on the scanpy.tl.rank\_genes\_groups results calculating DEGs of one respective species within the same cell type across all healthy and diseased samples. The expression of genes of interest across species can now be accessed on our website [www.susztaklab.com](http://www.susztaklab.com).

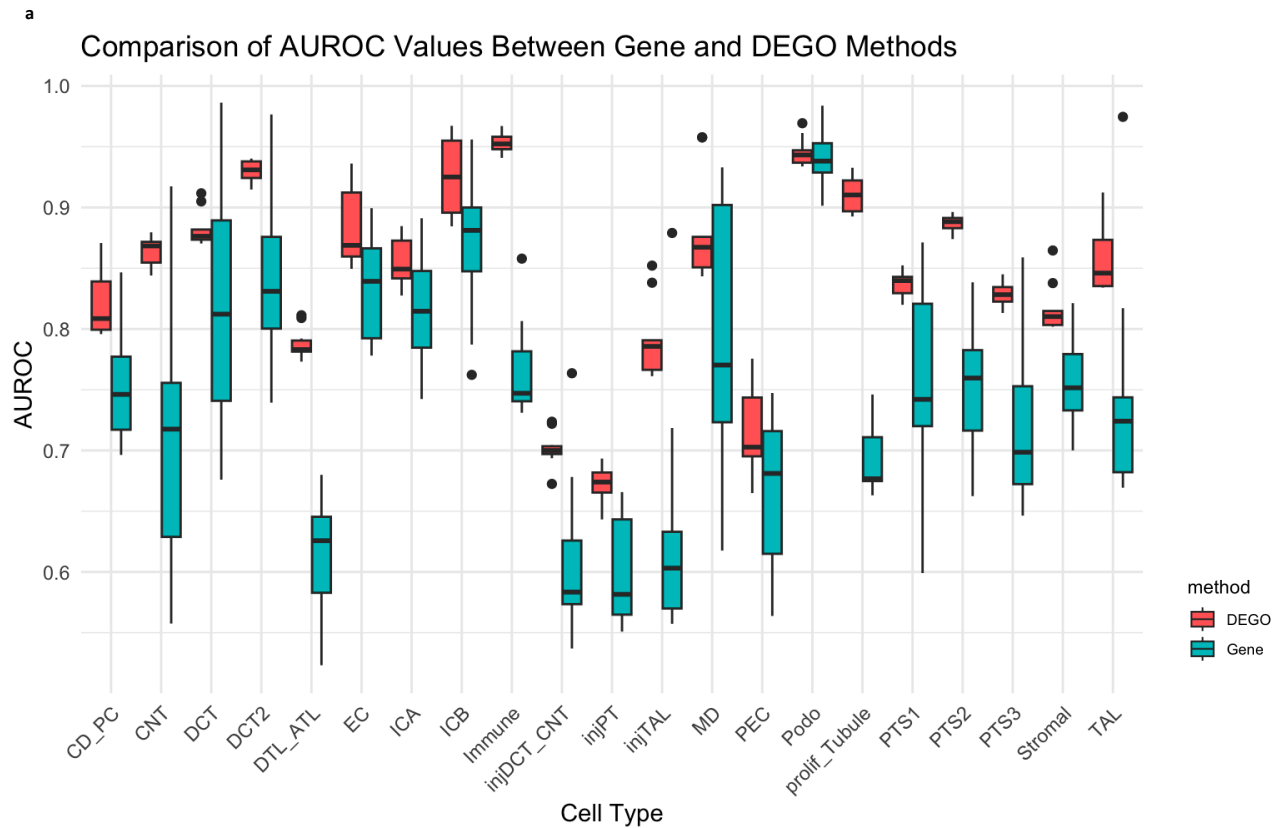




**Supplemental Figure 7: Gene-set centered analysis elucidates conservation of cellular functions.**

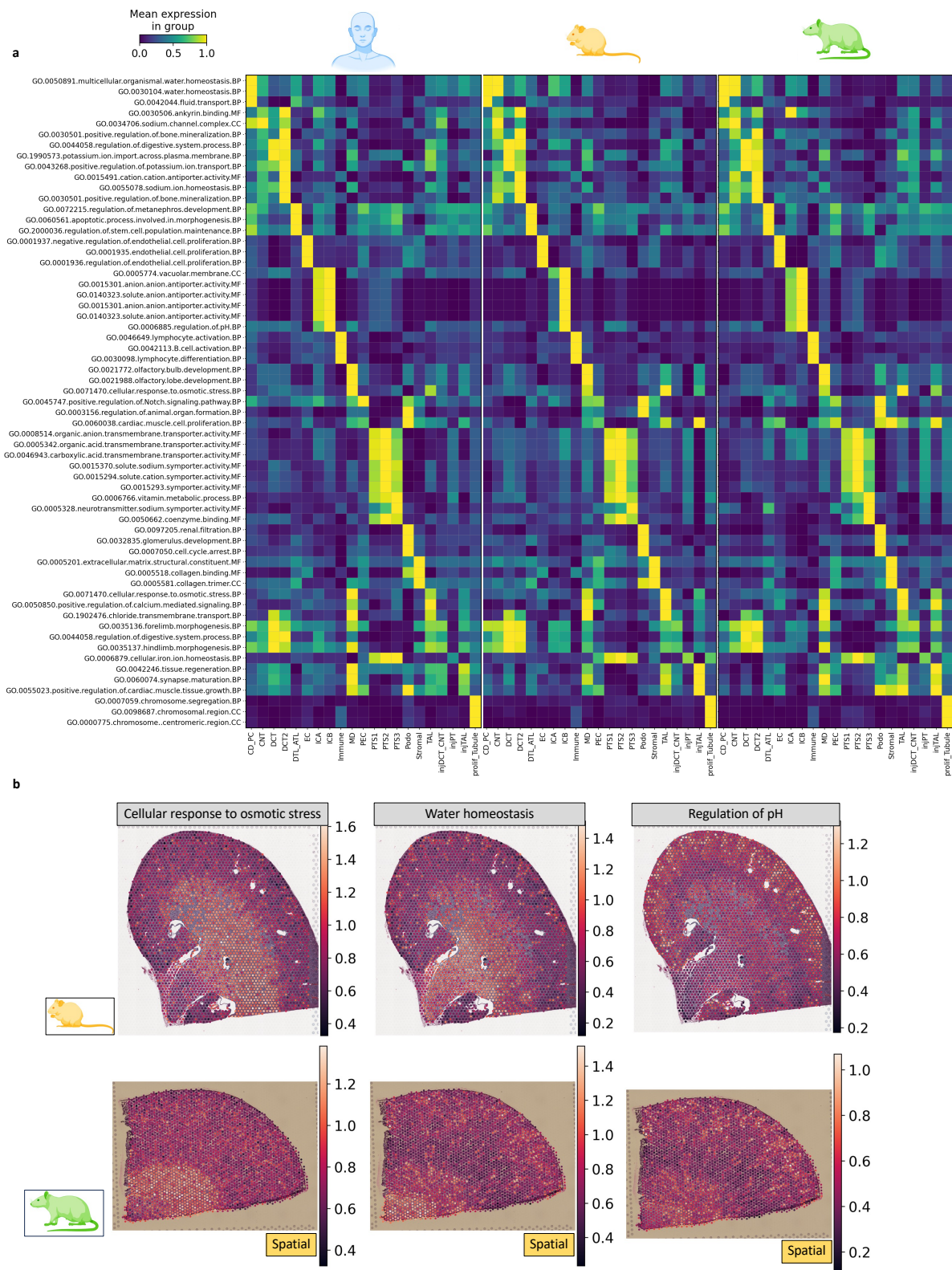
(a) Conceptual explanation of differentially expressed gene ontology (DEGO) to define conserved cellular functions. Gene ontology sets were used to functionally group genes, followed by the calculation of average gene expression, resulting in a matrix of single cells and gene sets for each species. MetaMarker was then used to identify conserved cellular functions with higher expression levels in specific cell types across species. Venn diagrams show the percentages of conserved marker genes (left) and marker gene sets (right) across species.

(b) Feature plots of selected conserved cellular functions plotted on the multispecies integrated UMAP in single-cell resolution. Cell types (bottom corner in yellow) and gene sets (top in grey) are highlighted. Green-Yellow indicates enrichment.



**Supplemental Figure 8: Comparison of AUROC scores for cell type-specific marker genes (Gene) and gene ontology descriptors (DEGO).**

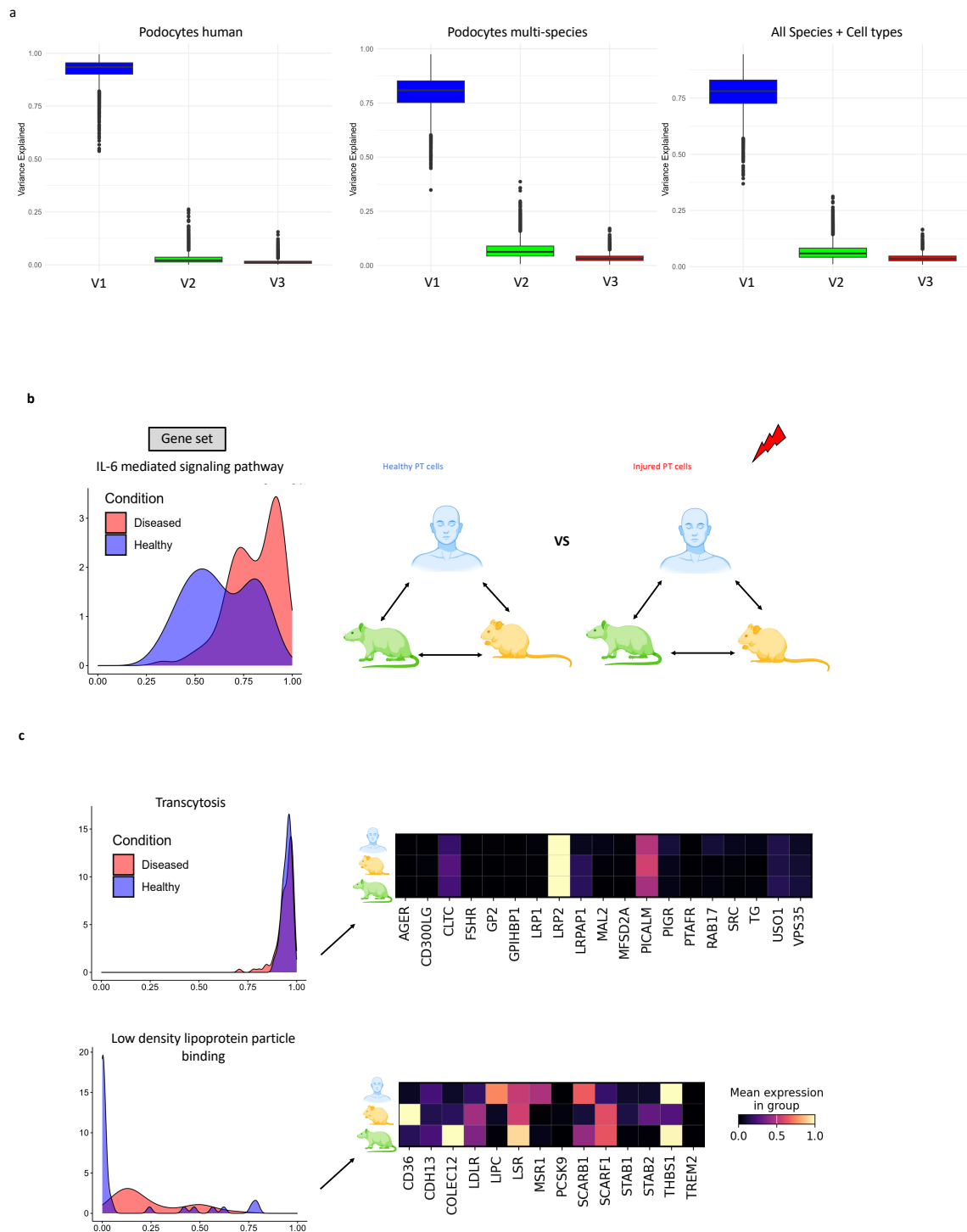
Boxplots display the area under the receiver operating characteristic curve (AUROC) values (y-axis) for the top 10 marker genes (blue) and marker gene sets (red, DEGO) across cell types (x-axis). Each boxplot represents the distribution of AUROC values derived from MetaMarker analyses conducted across species. The comparison highlights the performance of gene-based and gene ontology-based approaches in distinguishing cell types, with DEGO generally capturing broader biological contexts. Cell types include collecting duct principal cells (CD\_PC), connecting tubule cells (CNT), distal convoluted tubule cells (DCT and DCT2), endothelial cells (EC), intercalated cells (ICA and ICB), immune cells, injured proximal tubule cells (injPT), and others. Outliers are shown as individual points.

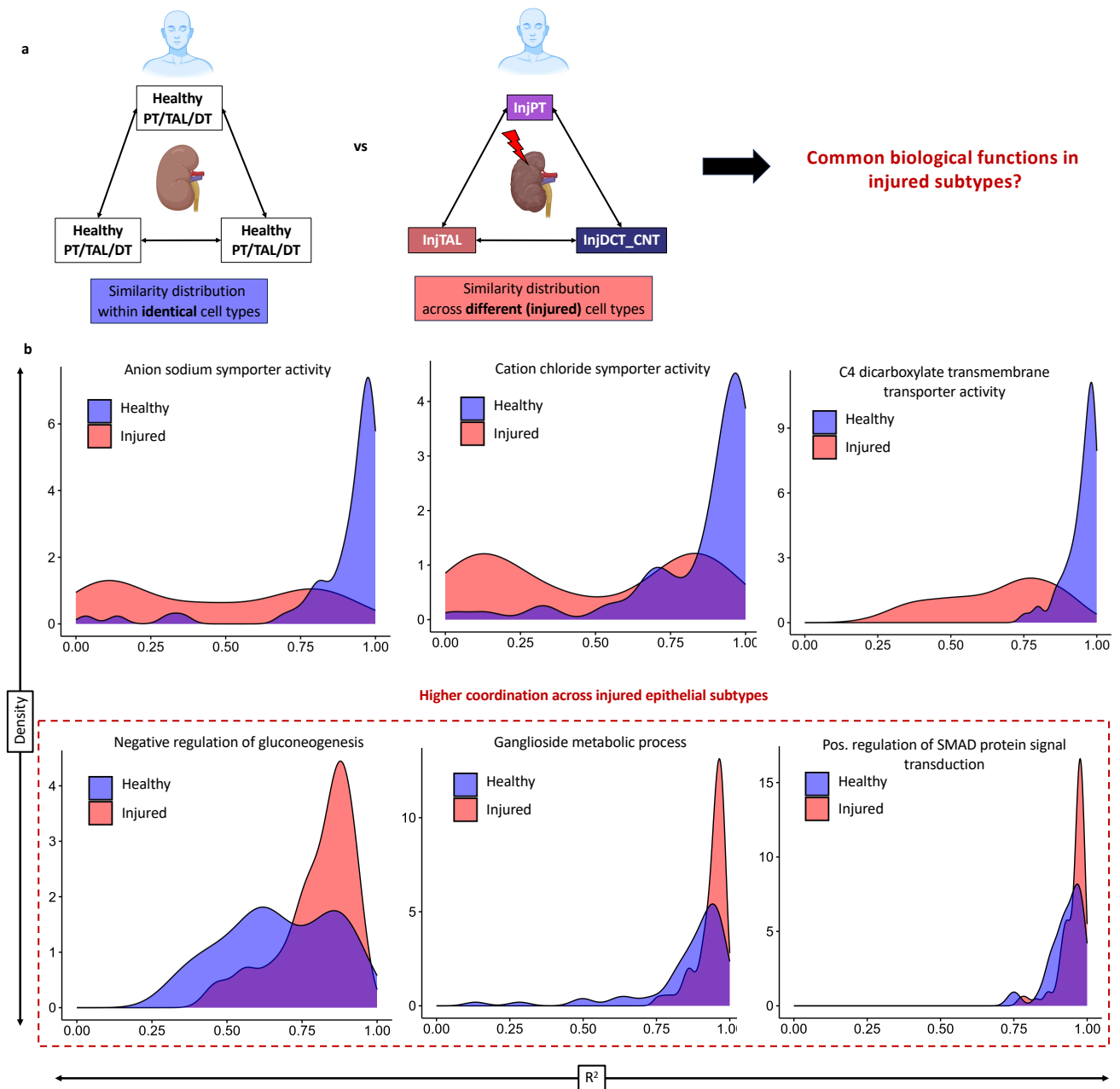


**Supplemental Figure 9: Cell function conservation within the kidney across species.**

(a) Top 3 conserved marker gene sets (DEGOs, MetaMarker rank) per cell type plotted in three separate matrix plots (human, mouse, rat) with scaling applied on each gene set. Each row is a gene ontology group, each column is a cell type.

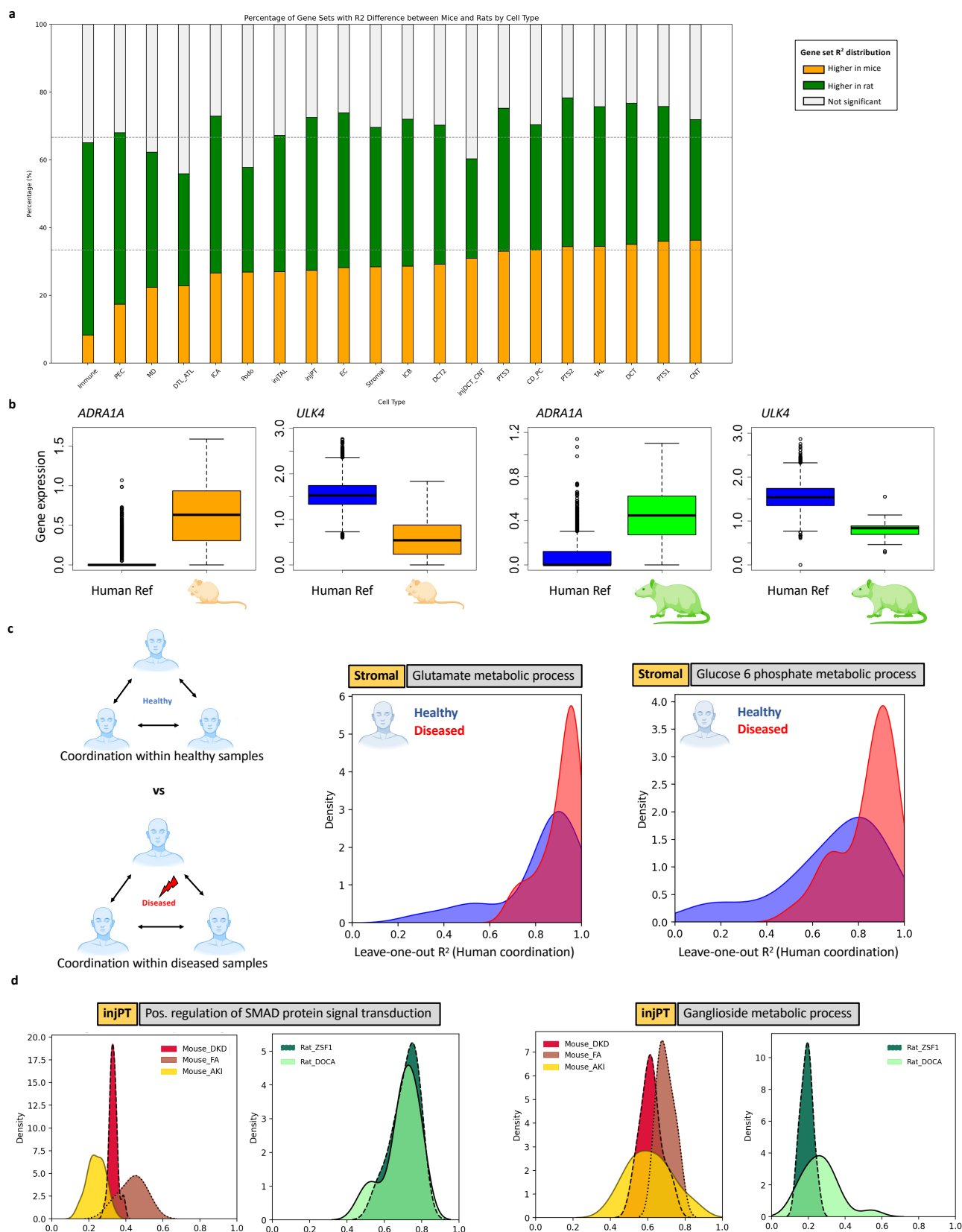
(b) Expression of selected conserved cellular functions on a mouse and rat Visium spatial transcriptomics slide (10X Genomics, Onoda et al. 2022). Visium dots are colored based on the average expression of the respective function, calculated by applying the DEGO approach to spatial transcriptomics data. Yellow indicates a higher expression. Top row shows a mouse kidney, while the bottom row shows the rat kidney, the top label indicates the plotted pathway.





**Supplemental Figure 11: Highly coordinated functions in human injured epithelial cell types.**

**(a)** Analysis overview. CellSpectra was used to assess coordination within sets of samples comprising different cell types to uncover common biological functions. Internal leave-one-out coordination analysis within mixed injured epithelial cell types compared to their healthy counterparts revealed several functions with significantly higher coordination within injured epithelial cells. Analyses were performed exclusively on human data. **(b)** Density plots of  $R^2$  similarity distributions comparing human healthy epithelial cells of the same cell type (blue) and mixed injured epithelial cell types (red). The y-axis represents a smoothed density estimate, reflecting the relative number of samples for each  $R^2$  value, and the x-axis shows  $R^2$  values. Separate plots depict pathways with significantly altered coordination in injured epithelial cells, showing either a loss or gain of coordination. Examples include “Anion sodium symporter activity,” “Cation chloride symporter activity,” “C4 dicarboxylate transmembrane transporter activity,” “Negative regulation of gluconeogenesis,” “Ganglioside metabolic process,” and “Positive regulation of SMAD protein signal transduction.” These findings highlight shared biological functions across injured epithelial subtypes.

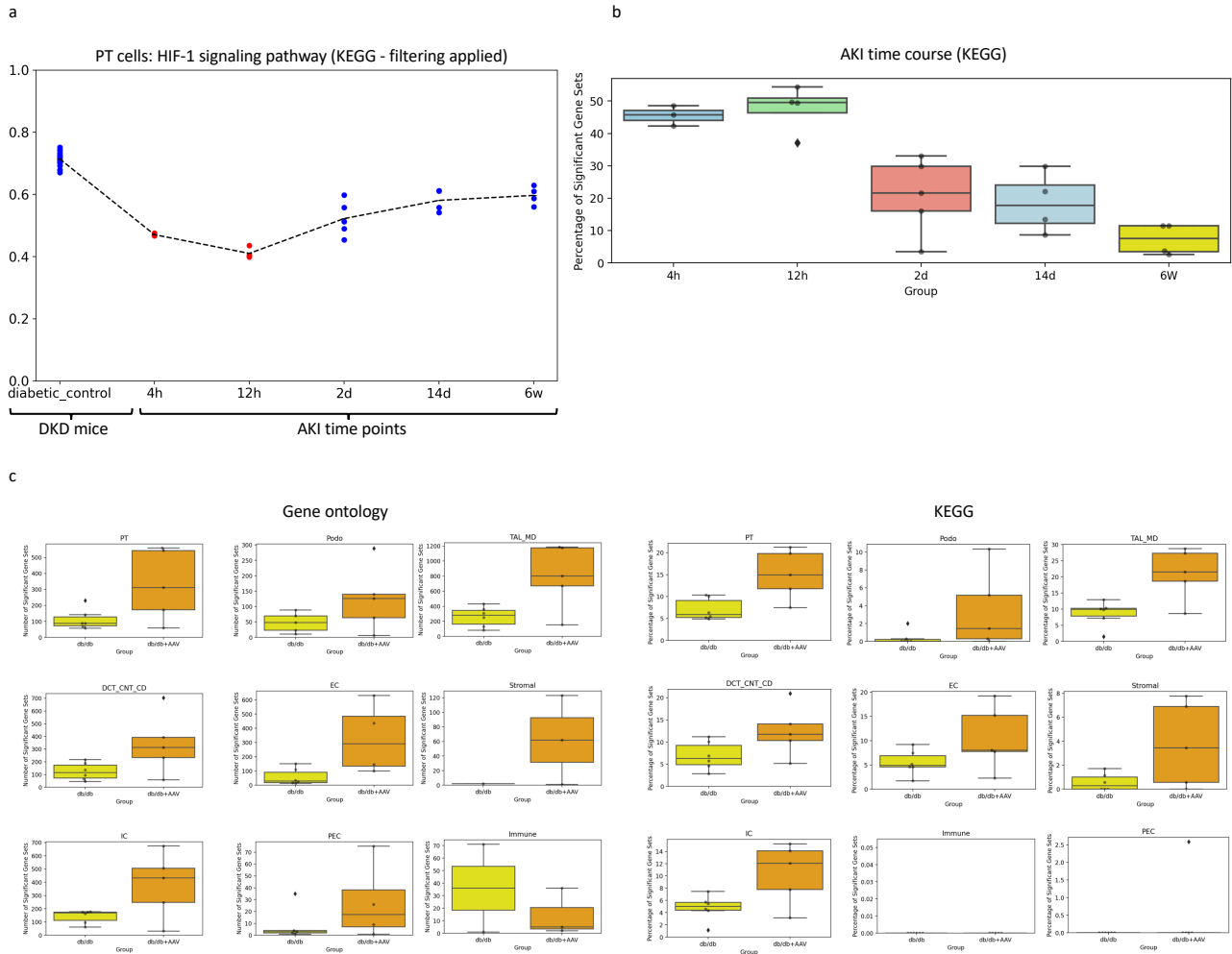


**Supplemental Figure 12: CellSpectra R<sup>2</sup> coordination within cell types and conditions.**

(a) Percentage of gene sets with significantly different R<sup>2</sup> distributions between mice and rats per cell type (padj < 0.05; Wilcoxon Rank-Sum Test) relative to the human diseased reference. Gene sets with significantly higher R<sup>2</sup> values in mice are indicated by orange bars, gene sets significantly higher in rats by green bars. Grey bars indicate pathways without significant R<sup>2</sup> differences between mice and rats. X-axis indicates the cell type, the y-axis the percentage of all compared gene sets. (b) Relative gene expression of *ADRA1A* and *ULK4* (y-axis) as examples of “underlying genes” of the observed coordination changes in PEC “Regulation of protein kinase C signaling” between humans and rodents (x-axis). Boxplots of the respective samples of the pooled reference (“Human Ref”, blue) and the respective rodent samples (X-axis). The y-axis

indicates the normalized gene expression. Mouse gene expression (left, orange) and rat gene expression (right, green) is shown separately. **(c)** Analysis overview (left). Internal coordination within healthy and diseased humans. Density plots of  $R^2$  (leave-one-out) distributions of healthy and diseased humans in stromal cells (right). Examples are shown for Stromal “Glutamate metabolic process” and Stromal “Glucose 6 phosphate metabolic process”.  $R^2$  distributions are visualized on the x-axis for healthy (blue), and diseased samples (red), the y-axis represents a smoothed density estimate, reflecting the relative number of samples for each  $R^2$  value. **(d)** Density plots of rodent model  $R^2$  coordination distributions relative to human patients (human diseased reference).  $R^2$  values of individual sample are plotted as distributions in a density plot, the y-axis represents a smoothed density estimate, reflecting the relative number of samples for each  $R^2$  value, and  $R^2$  values indicated on the x-axis. Mouse and rat models are plotted separately. Examples are shown for injured PT “Positive regulation of SMAD protein signal transduction” and “Ganglioside metabolic process”. Mouse diabetes (red), Mouse ischemia AKI (yellow), Mouse folic acid (brown). Rat diabetes (dark green), Rat hypertension (light green).





### Supplemental Figure 13: Consistency of analysis results between the Gene Ontology and KEGG pathway database.

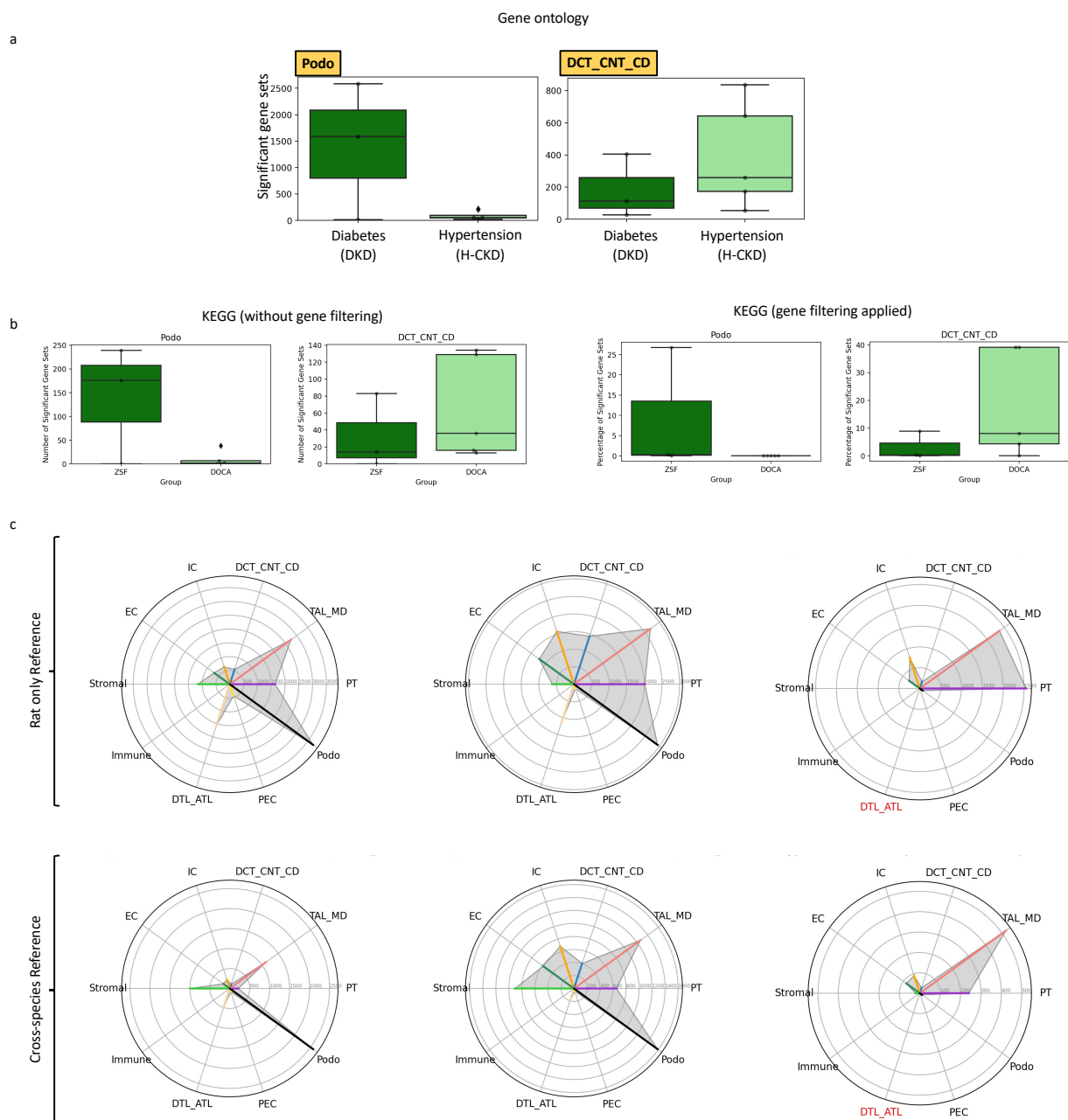
(a) KEGG HIF-1 signaling pathway gene coordination changes ( $R^2$  values in y-axis) in PT cells of individual mice (dots). The axis represents individual groups of DKD mice (diabetic kidney disease) and acute kidney injury (AKI) mice at different time points. Blue dots indicate a  $p_{adj} > 0.05$  in the CellSpectra analysis, while red dots indicate a sample  $p_{adj} < 0.05$  with significant coordination changes compared to the cross-species healthy reference.

(b) Percentages of significantly dyscoordinated pathways ( $p_{adj} < 0.05$ ) in PT cells per mouse AKI samples (y-axis) visualized in a boxplot per time point (x-axis,  $n = 3$  biological replicates per time point, with some having two libraries from the same sample). Data was generated considering the KEGG database of pathways.

(c) Boxplots showing the number of significant gene sets ( $p_{adj} < 0.05$ ) (y-axis) per sample in db/db (yellow,  $n = 6$  samples) and db/db + Renin AAV (orange,  $n = 5$  samples) mice across cell types considering the Gene Ontology (left) and KEGG database (right). Only samples with cell numbers above the threshold are shown. All boxplots show the median (center line), quartiles (box limits), and whiskers, which extend to the minimum and maximum values within  $1.5 \times IQR$ . Data points outside this range are plotted as outliers.

Principal cells of the collecting duct (CD\_PC), connective tubule (CNT), distal convoluted tubule (DCT and DCT2), intercalated cells (ICA and ICB), podocytes (Podo), parietal epithelial cells (PEC), proximal tubule (PTS1, PTS2, PTS3), injured proximal tubule (injPT), proliferating tubule (prolif\_Tubule), macula densa (MD), thick ascending limb (TAL), descending and ascending thin limb (DTL\_ATL), immune cells, endothelial cells (EC), stromal cells, injured TAL, and thick ascending and distal injured tubule (injTAL and injDCT\_CNT).

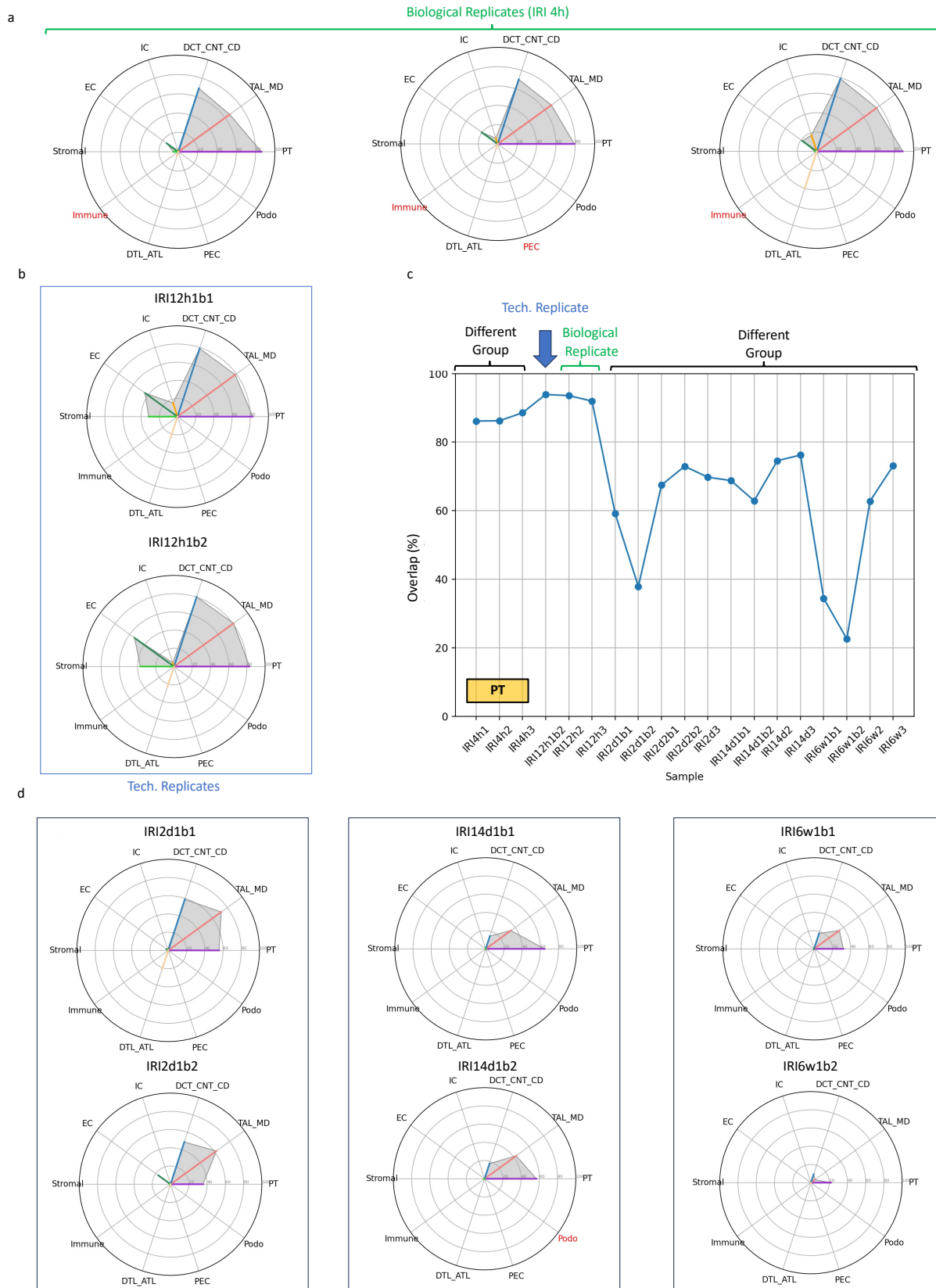




**Supplemental Figure 14: Cell type affection in diabetic versus hypertensive rats and results consistency for different references.**

(a) Boxplots showing the total number of dyscoordinated significant gene sets (Gene ontology) per sample on the y-axis in untreated diabetic ( $n = 3$ ) and hypertensive ( $n = 5$ ) rats (x-axis) for podocytes (left) and distal tubule clusters (DCT\_CNT\_CD, right). Only samples with cell numbers above the threshold are shown. (b) Boxplots showing the total number or percentage of significant dyscoordinated gene sets ( $p_{adj} < 0.05$ ) per sample in DKD (ZSF) and H-CKD (DOCA) rats. Comparisons are shown for podocytes and DCT\_CNT\_CD, using the KEGG database. The left panel represents results without applying the gene filtering mode (total number, y-axis), while the right panel shows results after gene filtering was applied (percentage, y-axis, Methods).

(c) Radar plots comparing results from rat-only (top row) and cross-species (bottom row) references for three untreated diabetic rat samples (columns). The x-axis labels indicate cell types: Immune, Stromal, PEC, DTL\_ATL, PT, TAL\_MD, Podo, IC, and DCT\_CNT\_CD. Circles represent the total number of significant gene sets per cell type. Red-highlighted cell types had cell counts below the threshold for analysis in the respective sample. Analyses with rat-only references (top) and cross-species healthy references (bottom) are displayed to highlight differences in results. These comparisons reveal how different reference datasets and treatments impact the identification of significant pathway coordination changes across cell types in DKD and H-CKD models.



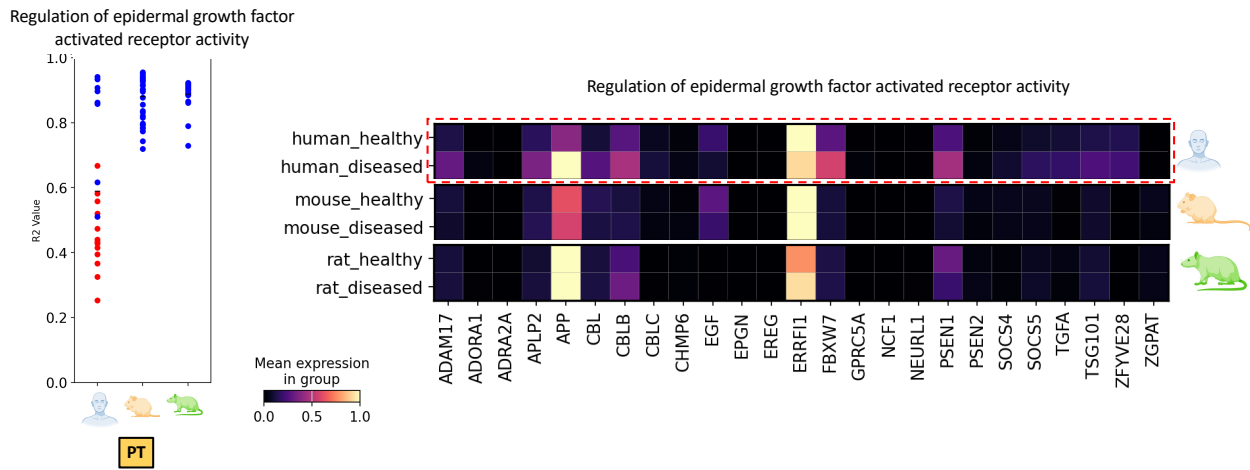
**Supplemental Figure 15: CellSpectra results demonstrate high consistency and replicability across biological and technical replicates.**

(a) Radar plots for biological replicates at the same time point (4 hours post-ischemia-reperfusion injury, mouse dataset) show the percentage of significant KEGG pathways for each cell type. The outer circle represents 100%. Cell types highlighted in red had cell counts below the threshold for analysis. Analyses used a mouse-specific reference to avoid bias from cross-species heterogeneity.

**(b)** Radar plots for two technical replicates of the same sample illustrate reproducibility in pathway significance across cell types.

**(c)** The percent of significant overlapping gene sets in PT cells across samples. Technical replicate 1 served as the reference, and the percentage of overlapping significant features in PT cells was calculated for all other samples. The y-axis indicates overlap percentage, while the x-axis lists samples and groups. The highest overlap was observed with the technical replicate (blue arrow), followed by biological replicates from the same group (green).

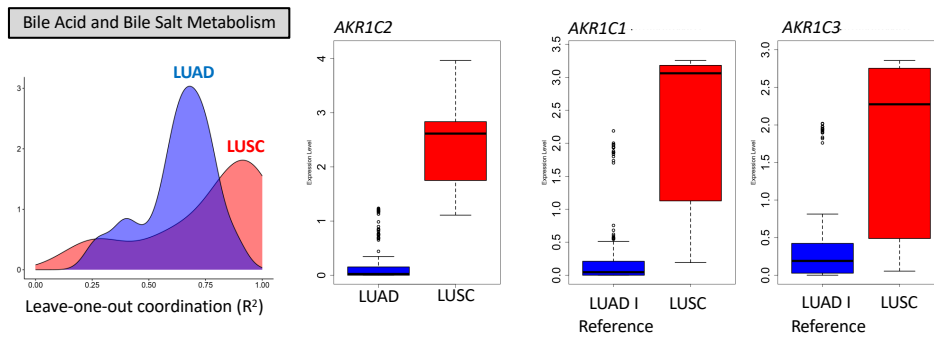
**(d)** Radar plots highlight the consistency of technical replicates (animal ID's listed on top) across different time points, demonstrating reproducibility in pathway significance over time. The axis represents cell types while the circles indicate the percentage of significant gene sets.



### Supplemental Figure 16: CellSpectra identifies human-specific changes in cellular functions.

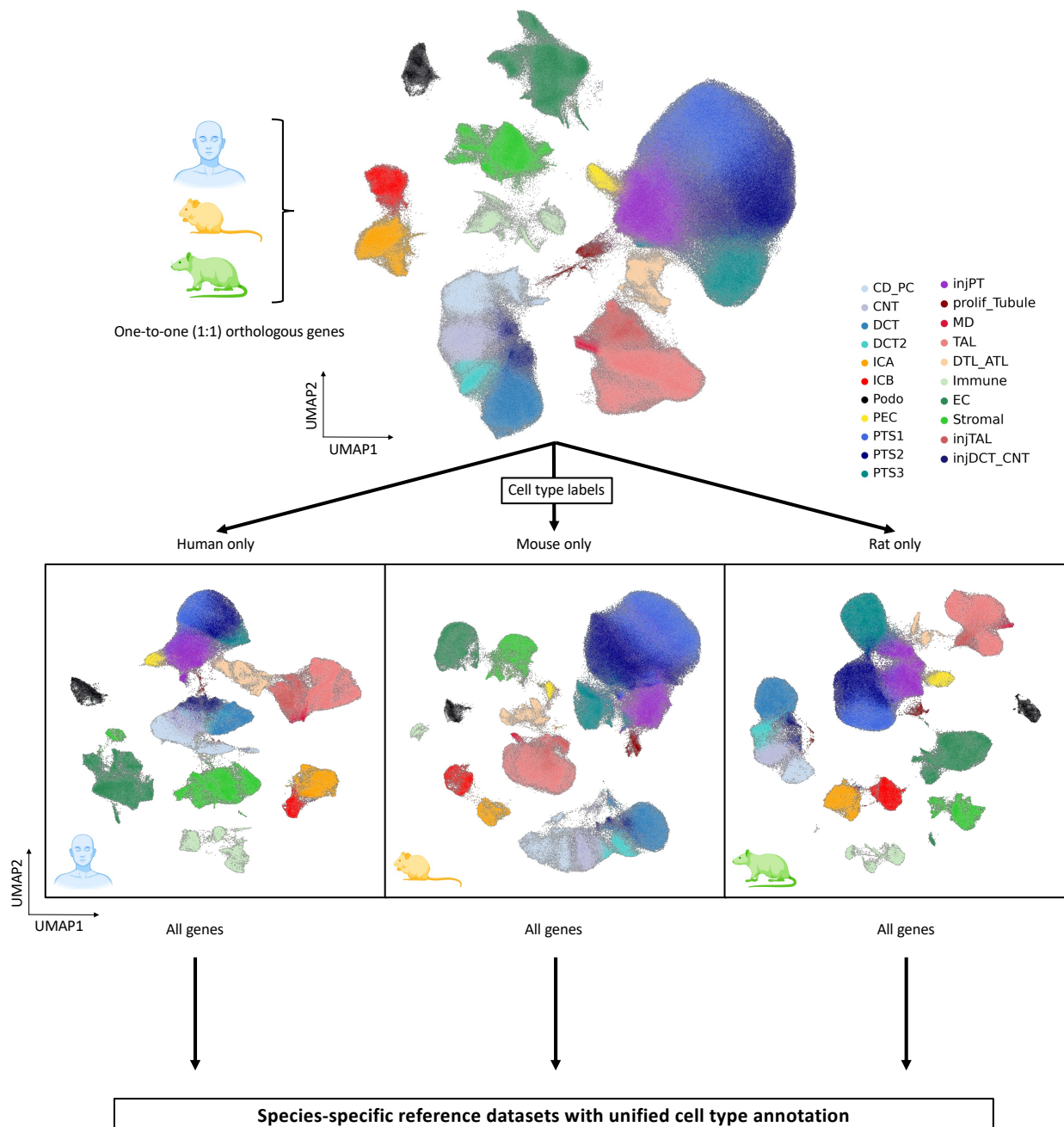
(Left panel) Scatter plot showing the coordination of the function “Regulation of epidermal growth factor-activated receptor activity” in TAL with species-specific changes. The y-axis indicates the  $R^2$  value, and the x-axis groups individual samples by species. Human ( $n = 21$ ), mouse ( $n = 49$ ), and rat ( $n = 22$ ). Blue dots represent samples with  $padj > 0.05$  (non-significant), while red dots indicate significant coordination changes ( $padj < 0.05$ ) compared to the cross-species healthy reference.

(Right panel) Matrix plot visualizing gene expression (columns) in human healthy and diseased samples compared to healthy and diseased samples in rodents (rows). Scaling was applied across genes within each group (rows). Yellow indicates higher expression of the respective gene relative to others. These analyses highlight human-specific changes in coordination for this function, which are not observed in rodents.



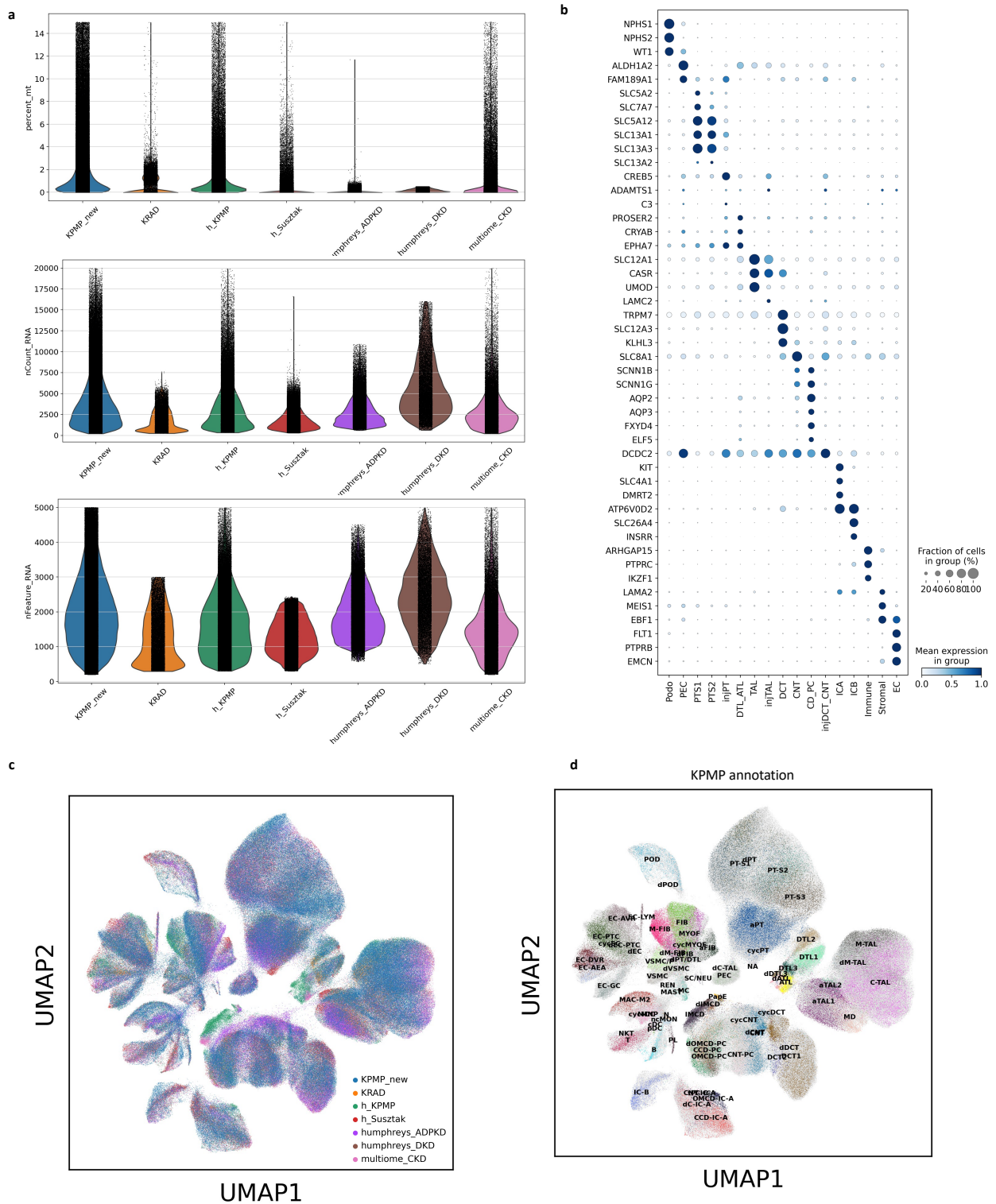
**Supplemental Figure 17: Gene expression coordination differences of LUAD and LUSC NSCLC samples.**

Left panel: Density plot of the internal leave-one-out coordination within LUAD (blue) and LUSC (red) samples. The example is shown for the pathway “Bile Acid and Bile Salt Metabolism”. Distributions of sample  $R^2$  values are shown (x-axis), while the y-axis represents a smoothed density estimate, reflecting the relative number of samples for each  $R^2$  value. Right panel: Boxplots visualizing the normalized sample gene expression of underlying genes (y-axis) for the LUAD reference (blue) and the LUSC query samples (red,  $n = 6$ ). Expression is shown for *AKR1C1*, *AKR1C2*, and *AKR1C3*, common underlying genes in LUSC samples for the “Bile Acid and Bile Salt Metabolism” pathway analysis. Reference expression was pooled from the “on the fly” references of all query samples (288 randomly permuted values equally derived from  $n = 16$  biological replicates). All boxplots show the median (center line), quartiles (box limits), and whiskers, which extend to the minimum and maximum values within  $1.5 \times \text{IQR}$ . Data points outside this range are plotted as outliers.



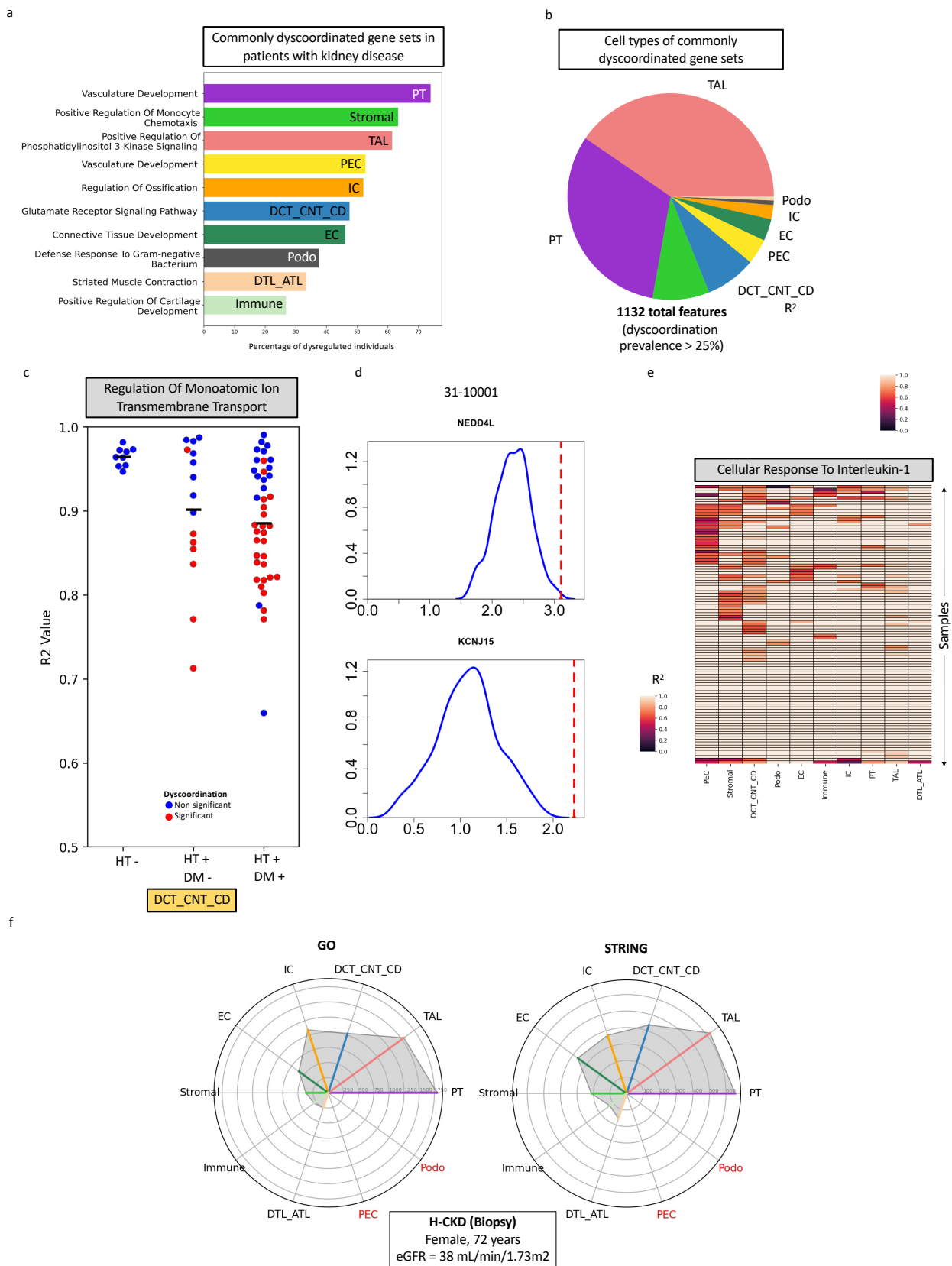
### Supplemental Figure 18: Multi-species atlas derived reference atlases for human, mouse, and rat.

Overview of species-specific atlases. We provide three separate integrated reference atlases with harmonized cell type labels based on our multi-species atlas (integrated using one-to-one orthologous genes). To make species comparable, the multi-species atlas only included one-to-one orthologous genes. By transferring cell type labels back to the species-specific data, we can further provide extended gene atlases containing all available genes. Each of these atlases were integrated separately (not considering any cell type labels). UMAPs (middle row) visualize the embedding of species-specific integrations and the SISK1.0 derived cell type labels.



**Supplemental Figure 19: Quality comparison, marker genes, and cell type consistency of the extended human kidney atlas.** (a) Violin plots of quality markers (y-axis): percentage of mitochondrial genes (percent\_mt), total number of counts (nCount\_RNA), and number of detected features (nFeature\_RNA) of the different datasets (x-axis). (b) Dot plot of conserved marker genes across cell types of scANVI-derived cell type labels transferred from the SISKa annotation. Dot colors indicate the scaled mean expression per cell type, the size of the dot indicates the percentage of cells expressing the respective gene. (c) UMAP of the scANVI-integrated embedding colored for the input datasets. (d) KPMP level 2 cell type annotations visualized on the integrated UMAP embedding.

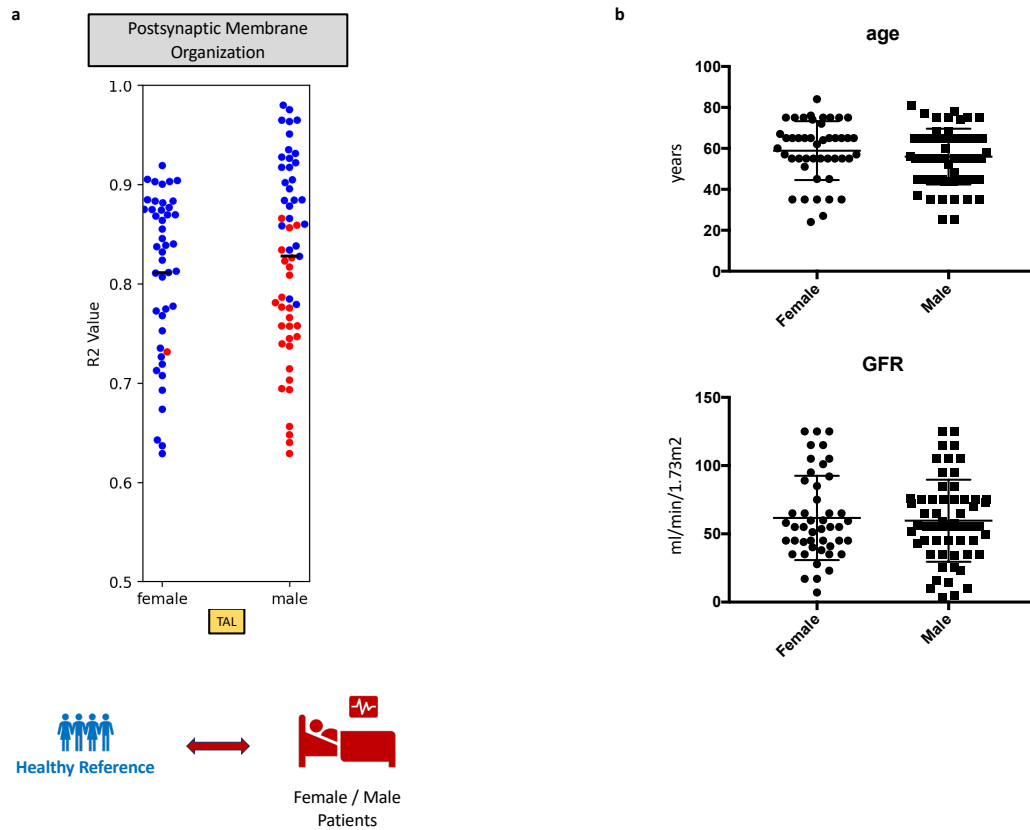




**Supplemental Figure 20: Highly prevalent pathway coordination changes in kidney disease patients.** (a) Cellular functions per cell type with the highest prevalence of changes in coordination. Each row is a cell type and the most commonly dysregulated pathway, the x-axis shows percent of individuals with dysregulation of the shown pathway.



**(b)** Pie chart of the cell types for the 1,132 highly prevalent cell functions (significance observed in more than 25% of samples,  $p_{adj} < 0.05$ ). The pie chart shows the relative frequency of the different cell types associated with these 1,132 functions. **(c)** Change in coordination in DCT\_CNT\_CD “Regulation of Monoatomic Ion Transmembrane Transport”. Scatterplots showing dyscoordination of individual samples ( $R^2$  indicated on the y-axis). Groups (on x-axis) include patients without hypertension (HT-,  $n = 9$ ), patients with hypertension but without diabetes (HT+ DM-,  $n = 15$ ), and patients with both hypertension and diabetes (HT+ DM+,  $n = 41$ ). Red dots indicate  $p_{adj} < 0.05$ . Bars display the mean  $R^2$  per group. **(d)** Examples of individual sample underlying genes for “Regulation of Monoatomic Ion Transmembrane Transport” in the distal cell cluster. The blue density line represents the expression distribution of the respective “on-the-fly” reference, while the red line indicates the query sample’s gene expression. X-axis indicates the respective gene expression, the y-axis represents a smoothed density estimate, reflecting the relative number of samples for each gene expression value. **(e)** Sample and cell-type heterogeneity of coordination changes in a targetable pathway (“Cellular Response to IL-1”). Individual sample  $R^2$  values are plotted, with darker red indicating lower values. Samples with  $p_{adj}$  greater than 0.05 or those below the cell number cut-off are assigned a fixed  $R^2$  value of 1. **(f)** Radar plots for the same sample using different gene set databases. Left: Gene Ontology. Right: STRING database. Both plots display the total numbers of significant pathways per cell type (axes).



**Supplemental Figure 21: CellSpectra identifies sex differences in gene expression coordination.**

**(a)** Scatter plot of “Postsynaptic Membrane Organization” coordination ( $R^2$ , y-axis) in TAL cells in females ( $n = 45$ ) and males ( $n = 56$ , x-axis). Red dots indicate  $p_{adj} < 0.05$ . Bars display the mean  $R^2$  per group.

**(b)** Scatter dot plots of sample age and glomerular filtration rate (GFR) of female ( $n = 46$ ) and male ( $n = 57$ ) samples. Each dot corresponds to an individual sample within the extended human atlas. Y-axis indicates years (age) or  $\text{ml/min/1.73m}^2$  (GFR). Scatter plots show mean and SD.