

Supplementary Information

Multimodal spatial transcriptomic characterization of mouse kidney injury and repair

Qiao Xuanyuan, Haojia Wu, Hemalatha Sundaramoorthi, Pierre Isnard, Changfeng Chen, Waleed Rahmani and Benjamin D. Humphreys

Table of Contents

Supplementary Notes	2
Supplementary Note 1. Benchmarking segmentation algorithms for Xenium.....	2
Supplementary Note 2. Benchmarking cellular deconvolution algorithms for Visium.....	2
Supplementary Table	4
Supplementary Table 1. Xenium 300 gene panel	4
Supplementary Figures	5
Supplementary References	26

Supplementary Notes

Supplementary Note 1. Benchmarking segmentation algorithms for Xenium

In imaging-based spatial transcriptomics, each transcript is detected and assigned a spatial coordinate within the tissue. To group transcripts into individual cells, a segmentation algorithm is used to delineate cell boundaries. Most commonly, segmentation methods leverage nuclei staining (e.g., DAPI) and/or transcript distributions to infer cell boundaries. We chose the Baysor algorithm for cell segmentation in this study because it integrates both transcript distribution and the DAPI signal to optimize cell boundaries. Unlike methods that rely predominantly on DAPI segmentation (e.g., Cellpose, StarDist, and DeepCell), Baysor infers cell outlines based on the joint likelihood of transcriptional composition and nuclear morphology. It also incorporates a prior segmentation mask (generated from DAPI nuclei segmentation), with an adjustable hyperparameter (prior-segmentation-confidence) that controls how closely the algorithm adheres to the prior segmentation.

We benchmarked Baysor¹, Cellpose², StarDist³, and DeepCell⁴ on a region of interest in our kidney dataset. The 10x Xenium platform provided a default segmentation pipeline that used nuclei (DAPI) masks and expands them to approximate cell boundaries. Since no absolute ground truth segmentation is available, we compared these tools based on the resulting transcript and cell counts. Because Cellpose, StarDist, and DeepCell rely solely on the DAPI channel for nuclear detection, all three algorithms yielded significantly reduced transcript and cell counts after segmentation (Supplementary Fig. 15). This reduction in transcripts could have several drawbacks, including potentially loss of molecular signals within the cell plasma region and lower accuracy in defining cell identities. In contrast, Baysor reduced transcript loss while maintaining clear separation between cells.

Supplementary Note 2. Benchmarking cellular deconvolution algorithms for Visium

We conducted a systematic benchmarking of five deconvolution methods leveraging our adjacent Xenium and Visium dataset. After aligning the data points from two platforms, we linked each Visium spot with the Xenium cells whose centroids were located within the 55-micron spot circle. Visium spot cell composition was calculated based on the frequency of the Xenium cell types aligned to each spot. This Xenium-derived cell composition was used as the ground truth for each Visium spot. We then deconvoluted the cell type abundance using the following algorithms: Tacco⁵, Tangram⁶, Cell2location⁷, CARD⁸, and RCTD⁹. All of these algorithms require single-cell reference dataset to estimate the cell type proportion within the low-resolution spatial transcriptomic spots. To assess the performance of these algorithms, we applied three metrics: root-mean-square error (RMSE), Pearson correlation coefficient (PCC), and Jensen–Shannon divergence (JSD). The formulas of these metrics are provided below.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - b_i)^2} \quad (1)$$

For RMSE, a_i and b_i represent the predicted and actual proportions of cell type i , respectively. A lower RMSE indicates better prediction accuracy.

$$\text{PCC} = \frac{\sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^n (a_i - \bar{a})^2 \sum_{i=1}^n (b_i - \bar{b})^2}} \quad (2)$$

For PCC, $\bar{a}$ and $\bar{b}$ are the mean predicted cell proportions across all cell types. High values indicate better alignment between predicted and actual proportions.

$$\text{JSD}(a, b) = \sqrt{\frac{1}{2} (\text{KLD}(a \parallel m) + \text{KLD}(b \parallel m))} \quad (3)$$

JSD measures the similarity between two probability distributions of cell type proportions. A lower value indicating more similarity between computational method derived cell proportions and the Xenium-referenced cell type compositions in individual Visium spot.

Our results showed that RCTD outperformed the other tools, demonstrating the highest PCC and the lowest RMSE and JSD values (Supplementary Fig. 16A). However, the performance of these algorithms varied based on the abundance and spatial distribution of different cell types. For example, TAL cells were the most abundant cell type, and showed the highest deconvolution accuracy across all IRI time points (Supplementary Fig. 16B). Moreover, different algorithms exhibited varying prediction performances when analyzing the same cell type. Cell2location yielded the highest PCC for intercalated cell type B (ICB), whereas Tacco was more effective in deconvolving proximal tubule cells (Supplementary Fig. 16B). Notably, some algorithms showed systematic biases. We compared the deconvoluted fibroblast frequencies obtained from the Visium data using different deconvolution algorithms. We found that none of the Visium-deconvoluted fibroblast frequencies showed a significant correlation with the fibrosis score quantified on the Sirius Red-stained adjacent section (Supplementary Fig. 16C). In addition, we observed that Cell2location and CARD tended to overestimate the abundance of immune cells (Supplementary Fig. 17). This overestimation could lead to inaccurate conclusions about the extent of inflammation, potentially skewing the interpretation of the disease state. These findings highlight the importance of selecting the appropriate deconvolution tool for accurate cellular compositional analysis. Given these considerations, to infer the structural organization in Visium data in this study, we opted to aggregate Xenium-derived cell information directly, rather than relying on deconvolution algorithms.

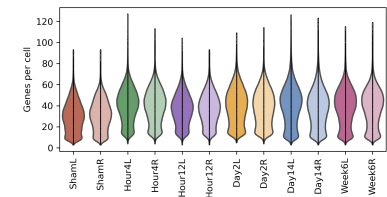
Supplementary Table

Supplementary Table 1. Xenium 300 gene panel

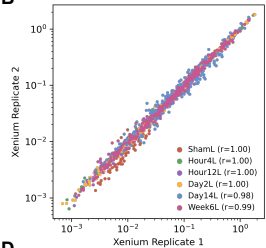
Abcc3	Bdh1	Cndp2	Egfem1	Glp1r	Kcnt1	Ngf	Proser2	Slc22a19	Tgfb1
Acmsd	Bmp4	Cntn5	Egfl7	Gm17268	Kdr	Nkx3-1	Ptger1	Slc22a22	Tgfb2
Acot7	Bmp6	Col18a1	Ehd3	Gpc6	Kl	Nlgn1	Ptger3	Slc22a6	Tgfb3
Acox1	Bmpr1a	Col1a1	Elf5	Gprc5a	Klf2	Nos3	Ptgfr	Slc22a8	Thsd7
Acox2	Bmpr1b	Col1a2	Eln	Grap	Klf5	Notch1	Ptgs2	Slc26a4	Tiam1
Acsn1	Bmpr2	Col5a2	Emcn	Grem2	Kmo	Notch3	Ptpro	Slc2a9	Timp3
Acsn3	Bst1	Col6a1	Emx1	Gsta1	Kng1	Nphs2	Qprt	Slc4a1	Timp4
Acta2	C3	Coro2b	Enox1	Haoa	Krt14	Npr1	Rbm20	Slc4a9	Tmem207
Adamts16	Cacna1c	Cpt2	Epcam	Havcr1	Krt15	Nrg1	Ren1	Slc5a12	Tmigd1
Adgrl3	Cacna1e	Csf1	Epha4	Hgf	Krt19	Nrip3	Rhbg	Slc5a2	Tnf
Adgrv1	Calb1	Csf1r	Epha6	Hmgcll1	Krt20	Nrp1	Rims2	Slc5a3	Tnfrsf11a
Adh1	Calm1	Ctnna3	Ephb2	Hmgcs2	Krt5	Nrp2	Rnf24	Slc5a8	Tpm1
Agrp	Calm2	Cwh43	Ephb4	Hnf4a	Kynu	Nrxn3	Runx1	Slc6a15	Trpm6
Agt	Calr	Cxcl1	Epo	Hyal2	Lair1	Ntng1	S1pr1	Slc6a19	Trpv5
Agtr1a	Casr	Cxcl12	Epor	Icam2	Lingo2	Osm	Sall3	Slc7a12	Tspan2
Akap12	Ccl7	Cyp24a1	Erbp4	Ido2	Lrp6	Osmr	Scara5	Slc7a13	Umod
Akr1b3	Ccl8	Cyp2e1	Fam189a1	Il2	Mapk10	Palmd	Scnn1g	Slc7a8	Upk1b
Aldh1a3	Ccr1	Cyp4b1	Fam221a	Il2ra	Mcam	Pappa2	Sdc2	Slc7a9	Utp20
Aldh1l2	Ccr4	Cyp7b1	Fbln1	Il2rb	Mgp	Papss2	Sema3d	Slc9a3r2	Vcam1
Amph	Ccl2	Cyr61	Fgf1	Il34	Mki67	Pde3a	Sema3g	Slco4a1	Vegfa
Ano1	Cd24	Dcl1	Fgf13	Inha	Mmp2	Pgf	Sema6a	Smoc2	Vegfb
Ano4	Cd44	Ddn	Fgfr4	Inhba	Mrc2	Pi16	Serpine1	Sorcs3	Vegfc
Aplnr	Cd53	Dkk2	Flt4	Inmt	Mro	Plau	Serpine2	Sost	Vegfd
Aqp1	Cd74	Dll3	Fosl1	Jag1	Msln	Plcx3	Sfrp1	Spink1	Vmn1r18
Aqp2	Cdkn1c	Dll4	Foxi1	Kcnc3	Muc20	Plin2	Sgc2	Spon1	Vnn1
Aqp4	Cgref1	Dmrt2	Fst	Kcnd2	Myct1	Plxnd1	Shd	Spp1	Vtn
Aqp6	Cldn16	Dpp4	Fyb	Kcnip1	Myh11	Podxl	Slc12a1	Spp2	Vwf
Atp6v1c2	Clec14a	Efnb1	Gabrb3	Kcnma1	Myl9	Ppara	Slc12a3	Stap1	Wif1
Avpr2	Clec1a	Efnb2	Gimap5	Kcnmb2	Myom2	Prkcg	Slc13a3	Synpo	Wnk1
Bcat1	Clnk	Egf	Gja5	Kcnq3	Nalcn	Prkch	Slc14a2	Tagln	Wnt16

Supplementary Figures

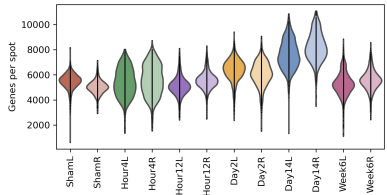
A



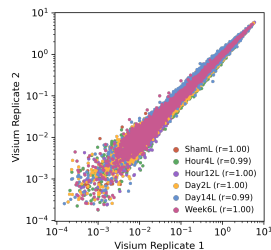
B



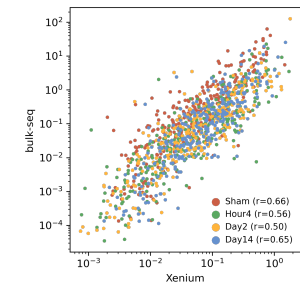
C



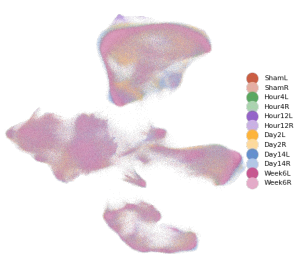
D



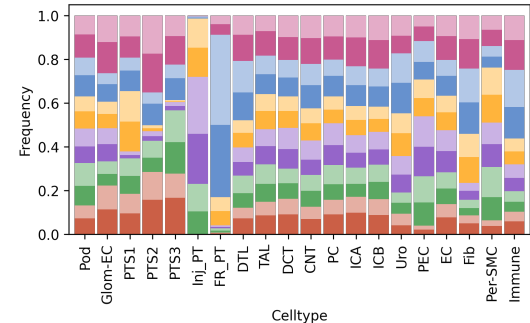
E



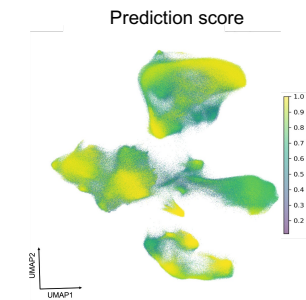
F



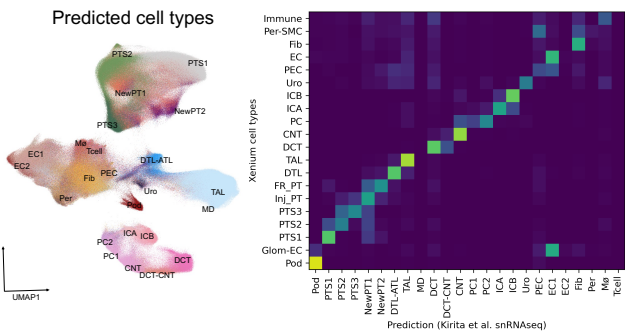
G



H

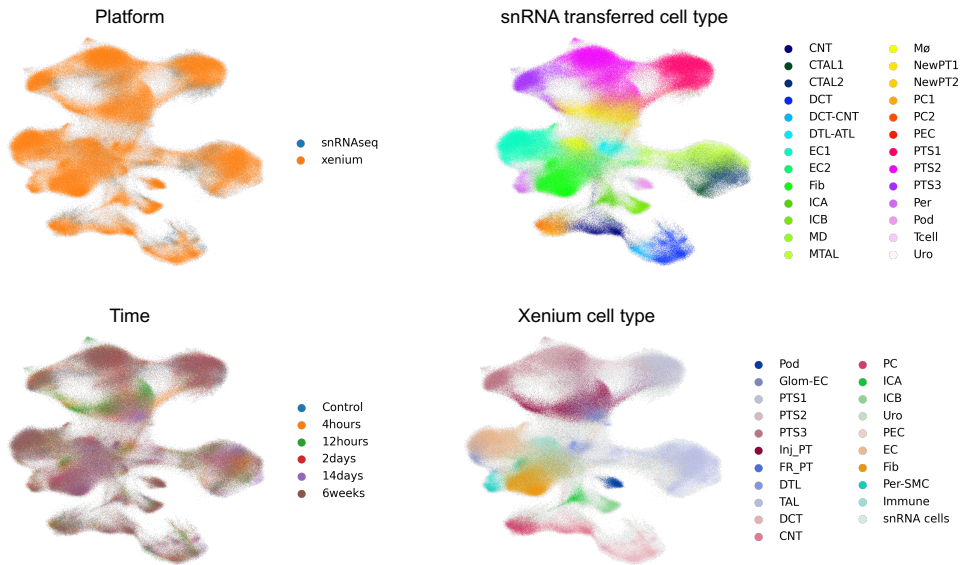
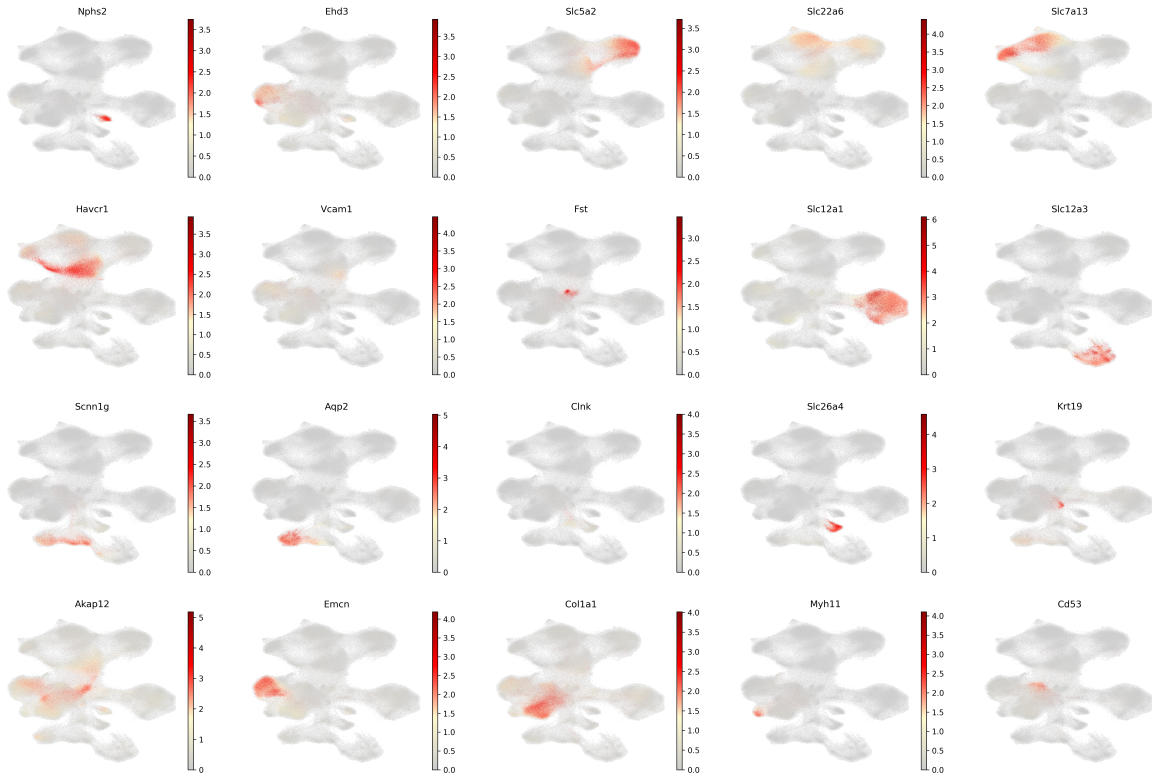


I



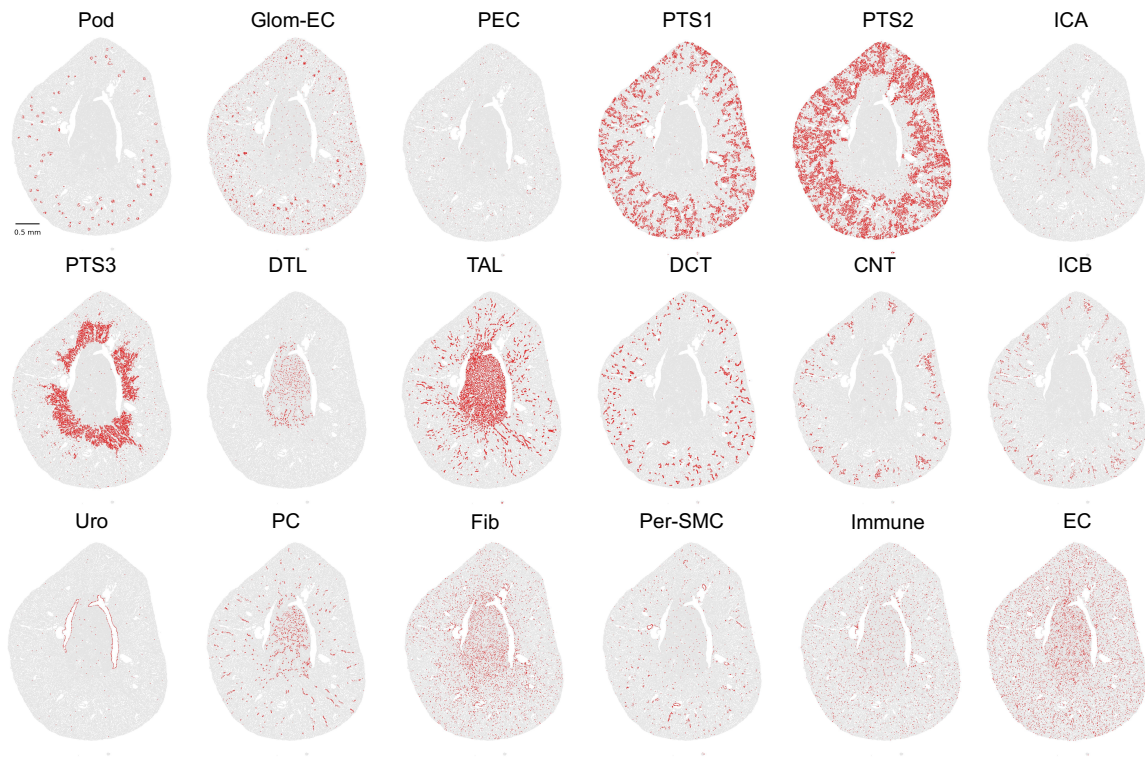
Supplementary Fig. 1 Xenium and Visium quality control metrics.

(A) Violin plot showing the number of genes detected per cell and the number of transcripts detected per cell in the Xenium dataset across different samples. **(B)** Scatterplot showing the correlation of average detected transcripts per cell for each gene between two replicates across each time-point Xenium sample. Replicates from different time points are encoded with different colors in the legend. **(C)** Violin plot showing the number of genes detected per spot and the number of unique molecular identifier (UMI) per spot in the Visium dataset across different samples. **(D)** Scatterplot showing the correlation of average UMI counts per spot for each gene between two replicates across each time-point Visium sample. **(E)** Scatterplot showing the correlation of average expression for each gene between Xenium and bulk RNA sequencing data from Liu, Jing, et al (GSE98622). A total of 293 shared genes were used in the comparison. **(F)** UMAP embedding of cells color by sample identity. **(G)** Bar plot showing the frequency distribution of each sample within each cell type. **(H)** UMAP projection of cell type labels predicted from Kirita et al. 2020 (GSE139107). A high prediction score indicates strong confidence in predicted cell types. **(I)** The heatmap displays the consistency between manually annotated cell type labels from Xenium and those predicted using reference snRNA-seq data.

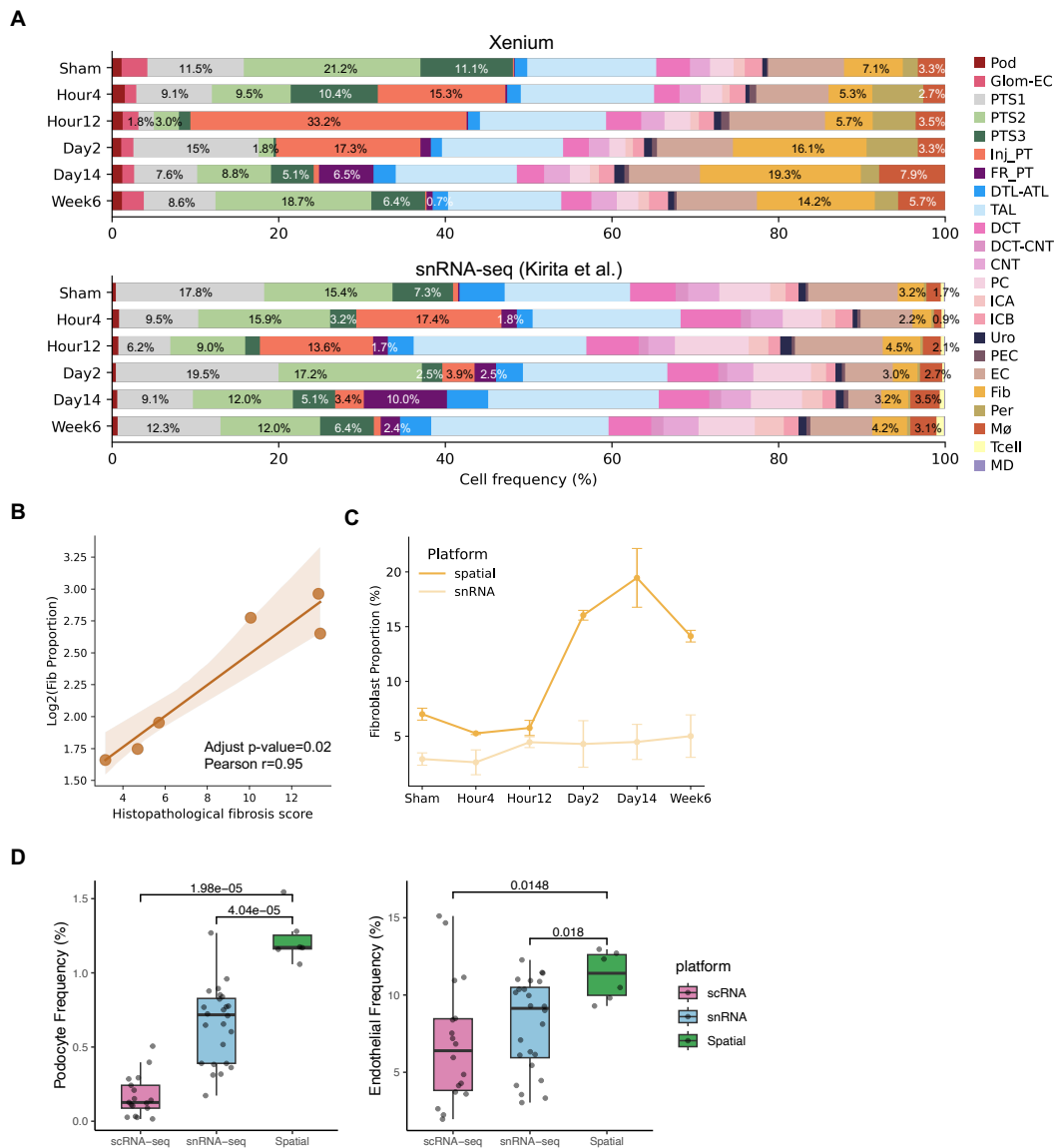
A**B**

Supplementary Fig. 2 Joint embedding of snRNA-seq data and Xenium spatial data. (A). UMAP embedding of snRNA-seq data and Xenium data. **(B)** Cell type-specific marker gene expression visualized on the UMAP.

A

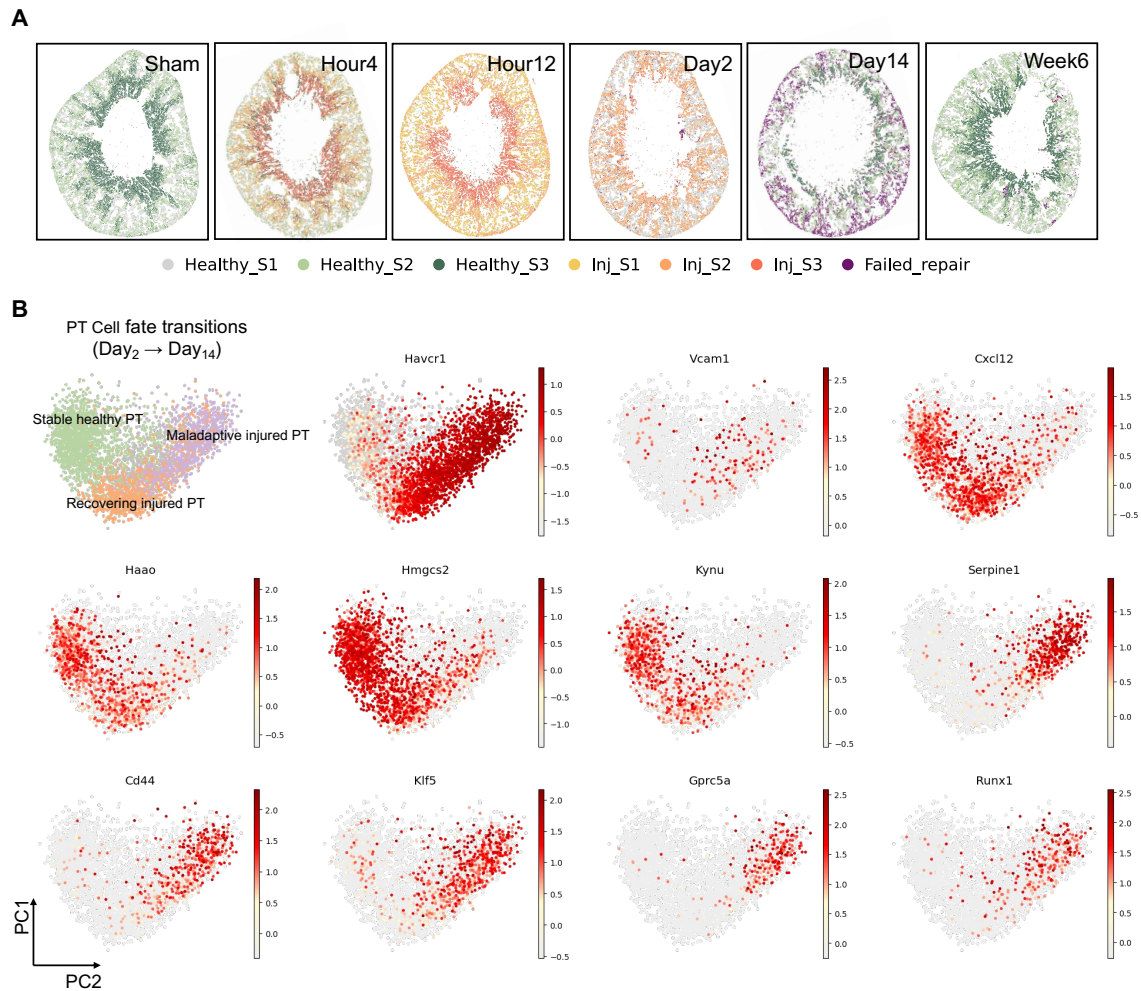


Supplementary Fig. 3 Spatial distribution of individual cell types in healthy kidney. Scale bar: 0.5 mm.



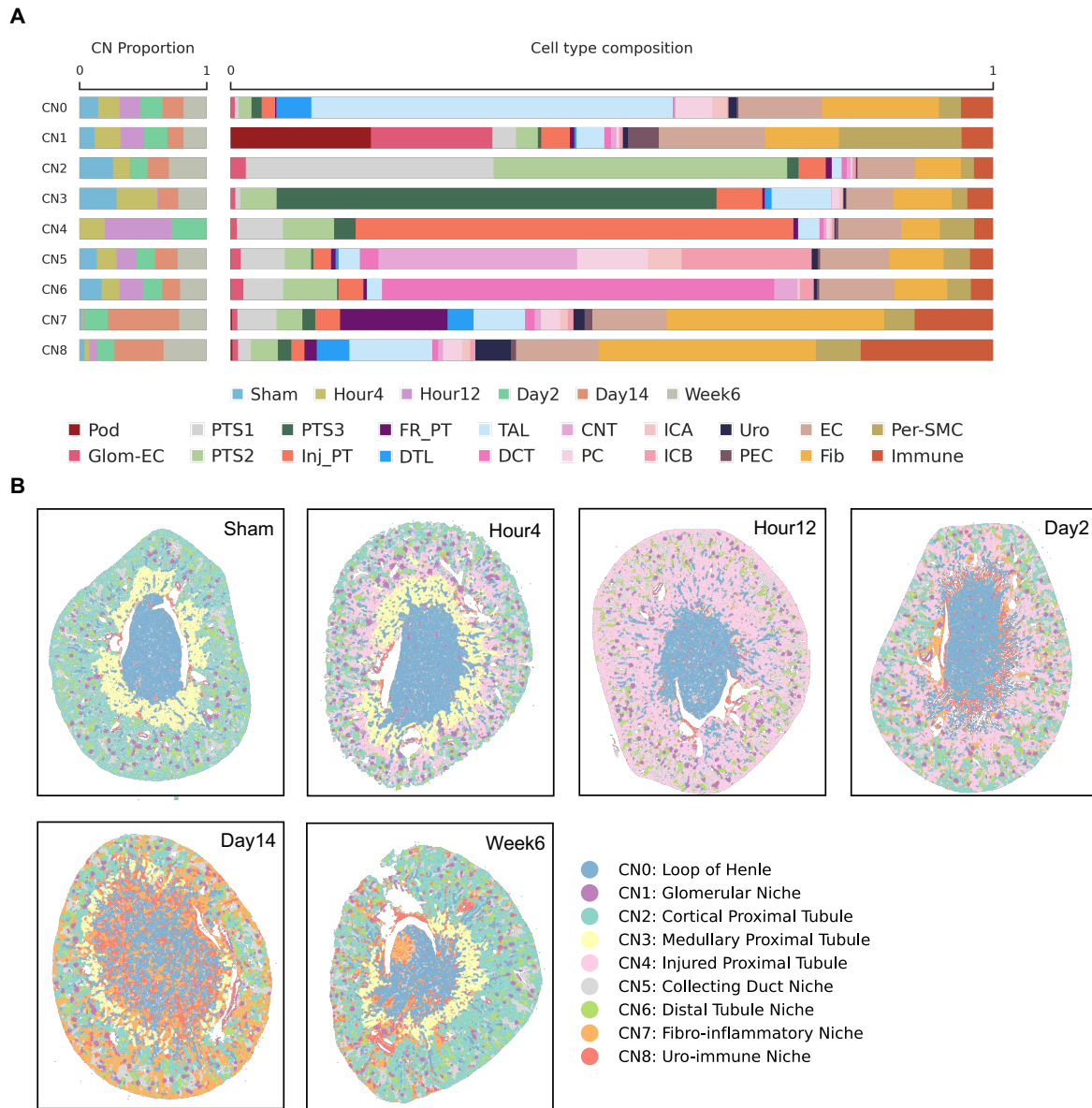
Supplementary Fig. 4 Comparison of cell type composition and proportions between spatial transcriptomics and single-cell RNA sequencing.

(A) Cell type composition in kidney tissues analyzed by Xenium spatial transcriptomics and single-nucleus RNA sequencing (Kirita et al., GSE139107) across the time course of ischemia-reperfusion injury (IRI). **(B)** Correlation between fibroblast abundance (log transformed) and the quantified fibrosis score from Sirius Red staining images. **(C)** Line plot comparing changes of fibroblast frequency over IRI time points between snRNA-seq and Xenium spatial dataset. Error bars indicate standard deviation of the sample replicates within corresponding time point. **(D)** Comparison of endothelial and podocyte cell type frequencies across single-cell RNA-seq (GSE180420), single-nucleus RAN-seq (GSE139107) and Xenium spatial transcriptomic platforms. P-values were calculated using Wilcoxon rank sum test. Source data provided in the Source Data file.



Supplementary Fig. 5 Spatiotemporal dynamics of proximal tubule cell subtypes.

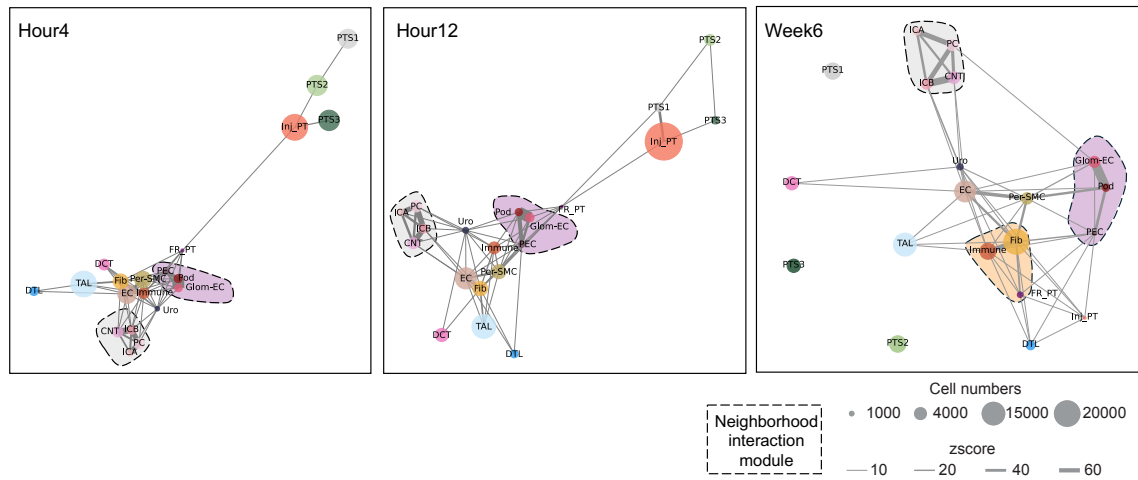
(A) Spatial distribution of PT cell subtypes. **(B)** Marker gene expression of day 2 PT cells with distinct cell fates at day 14.



Supplementary Fig. 6 Characterization of cell neighborhoods.

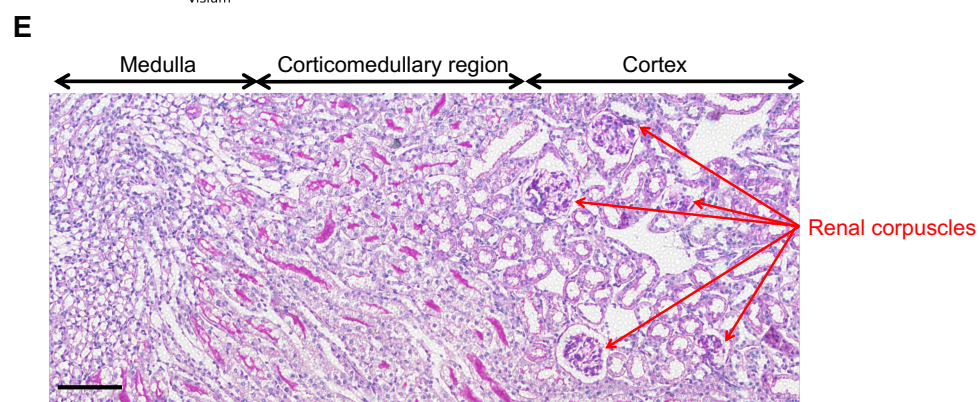
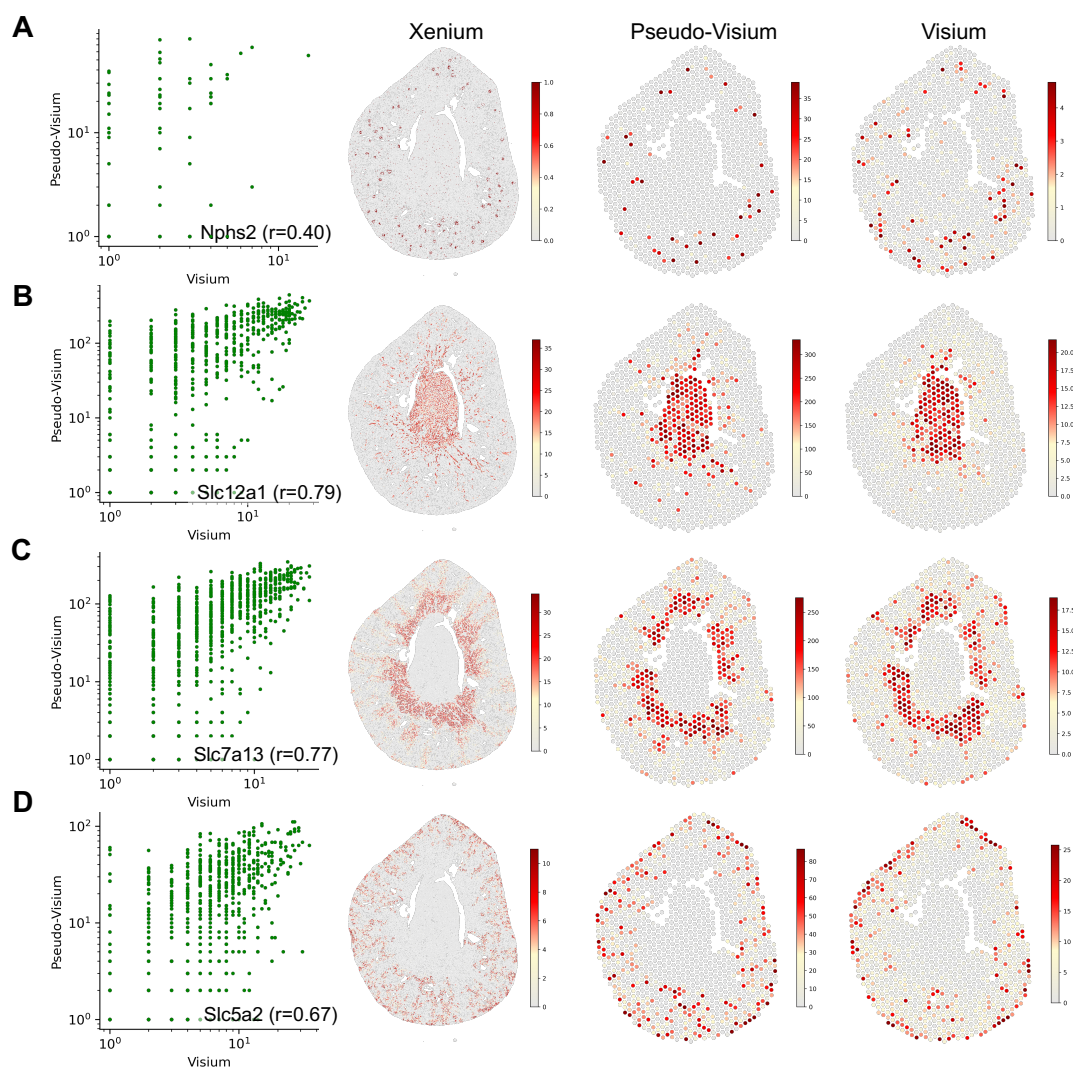
(A) Bar plots showing the time point identity and cell type composition within each identified cell neighborhood (CN). **(B)** Spatiotemporal dynamics of cell neighborhoods in whole kidney sections across different time points. A single replicate sample from each time point is visualized here.

A



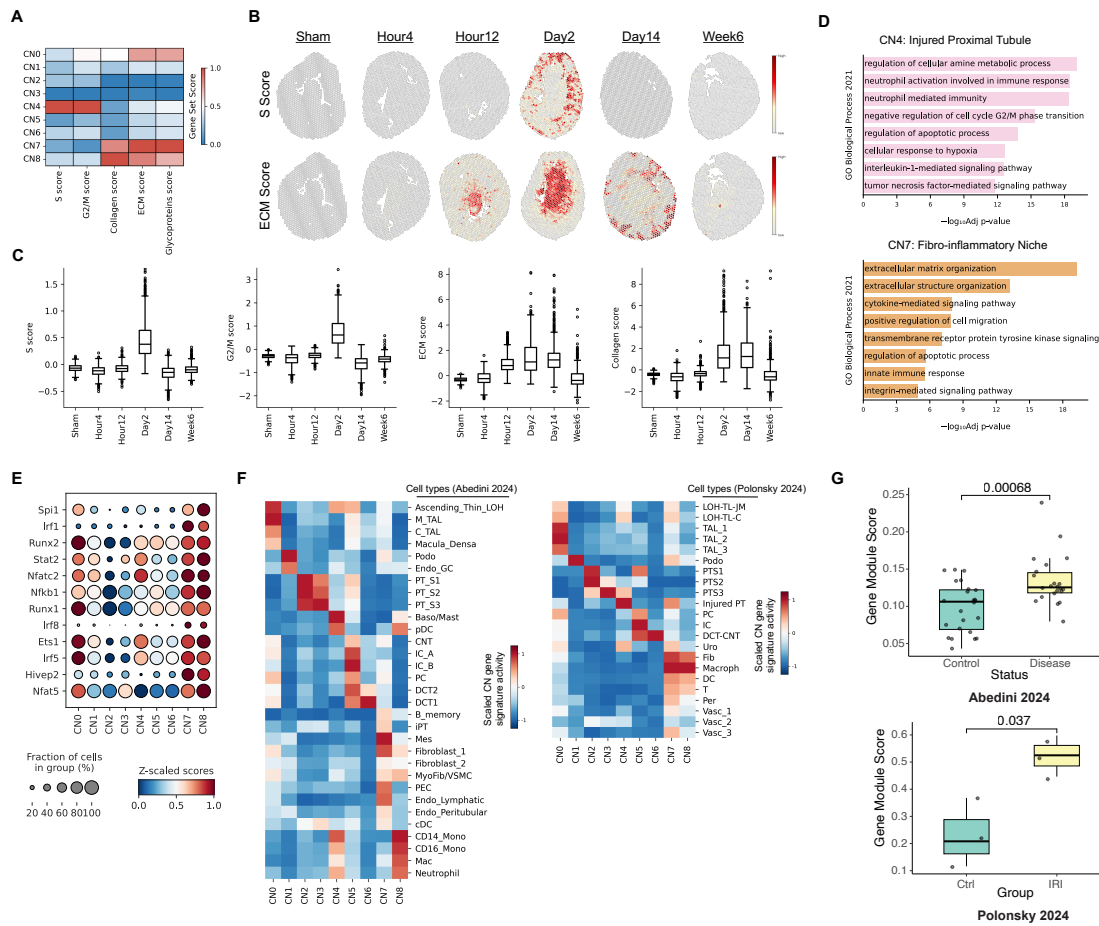
Supplementary Fig. 7 Cell type interaction dynamics.

(A) Graph networks showing interaction frequencies between cell type pairs. See also Figure 4D.



Supplementary Fig. 8 Cross-modality gene expression comparison.

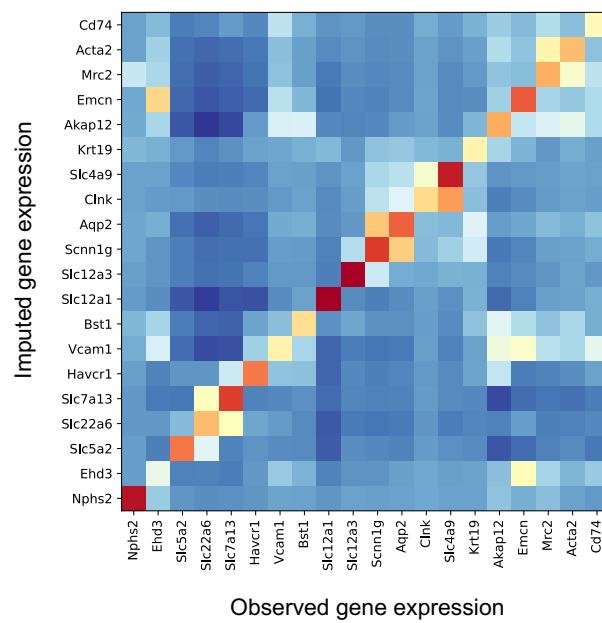
(A-D) Comparison of pseudo-Visium expression (aggregated from Xenium data) and Visium expression (counts per spot) for gene *Nphs2*, *Slc12a1*, *Aqp2* and *Slc5a2*. **(E)** PAS staining of the kidney section shown in Figure 5A. Scalebar: 100 μm .



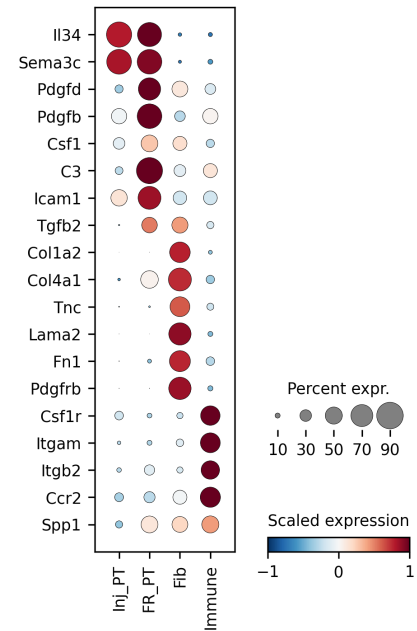
Supplementary Fig. 9 Molecular description of cell neighborhoods.

(A) Heatmap showing quantified gene module score. Source data are provided in Source Data file. **(B)** Spatiotemporal visualization of S score and ECM score in Visium data. **(C)** Boxplots showing the distribution of functional scores across IRI time points. **(D)** Gene Ontology analysis of differentially expressed gene sets from the injured tubule niche (CN4) and the fibro-inflammatory niche (CN7). **(E)** Dotplot showing top predicted transcription factor activities identified in Visium fibro-inflammatory niche (CN7). **(F)** Heatmap highlighting the activity of gene signatures derived from each cell neighborhood across cell types in the Abedini (human) and Polonsky (mouse) spatial datasets. CN0: Loop of Henle; CN1: Glomerular Niche; CN2: Cortical Proximal Tubule; CN3: Medullary Proximal Tubule; CN4: Injured Proximal Tubule; CN5: Collecting Duct Niche; CN6: Distal Tubule Niche; CN7: Fibro-inflammatory Niche; CN8: Uro-immune Niche. **(G)** Boxplot comparing the distribution of fibro-inflammatory niche (CN7) derived gene signatures between healthy and diseased kidney samples in mouse and human datasets. P-values were calculated using two-sided t-test.

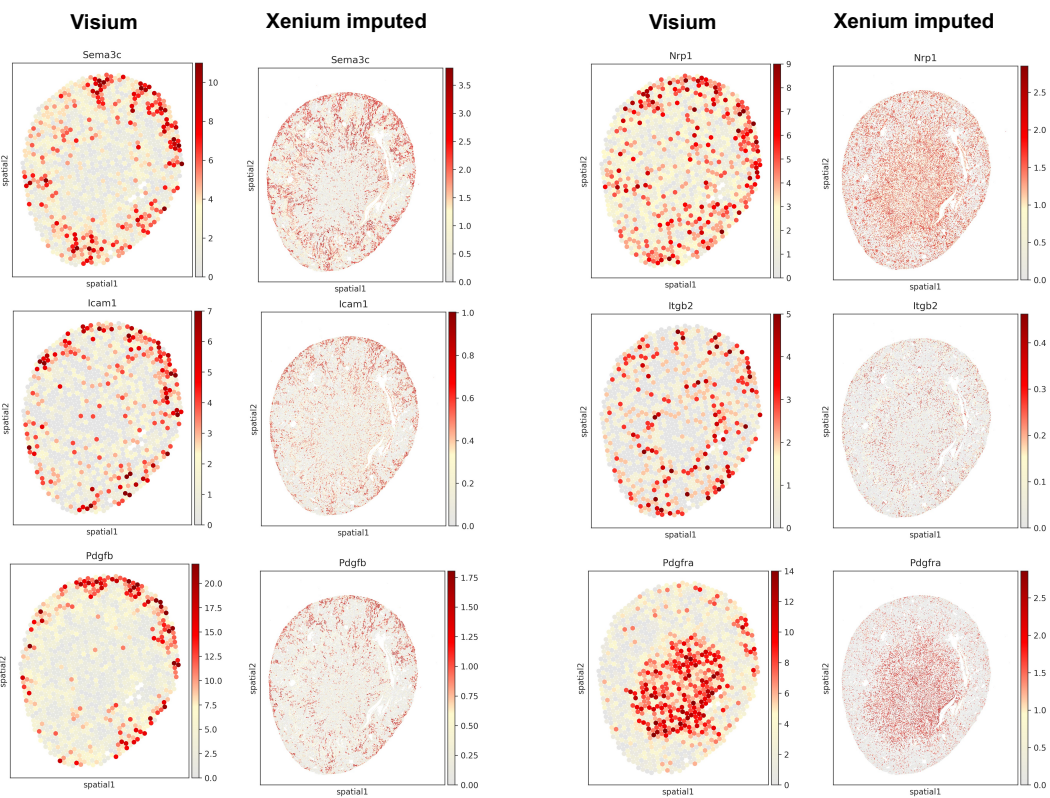
A



B

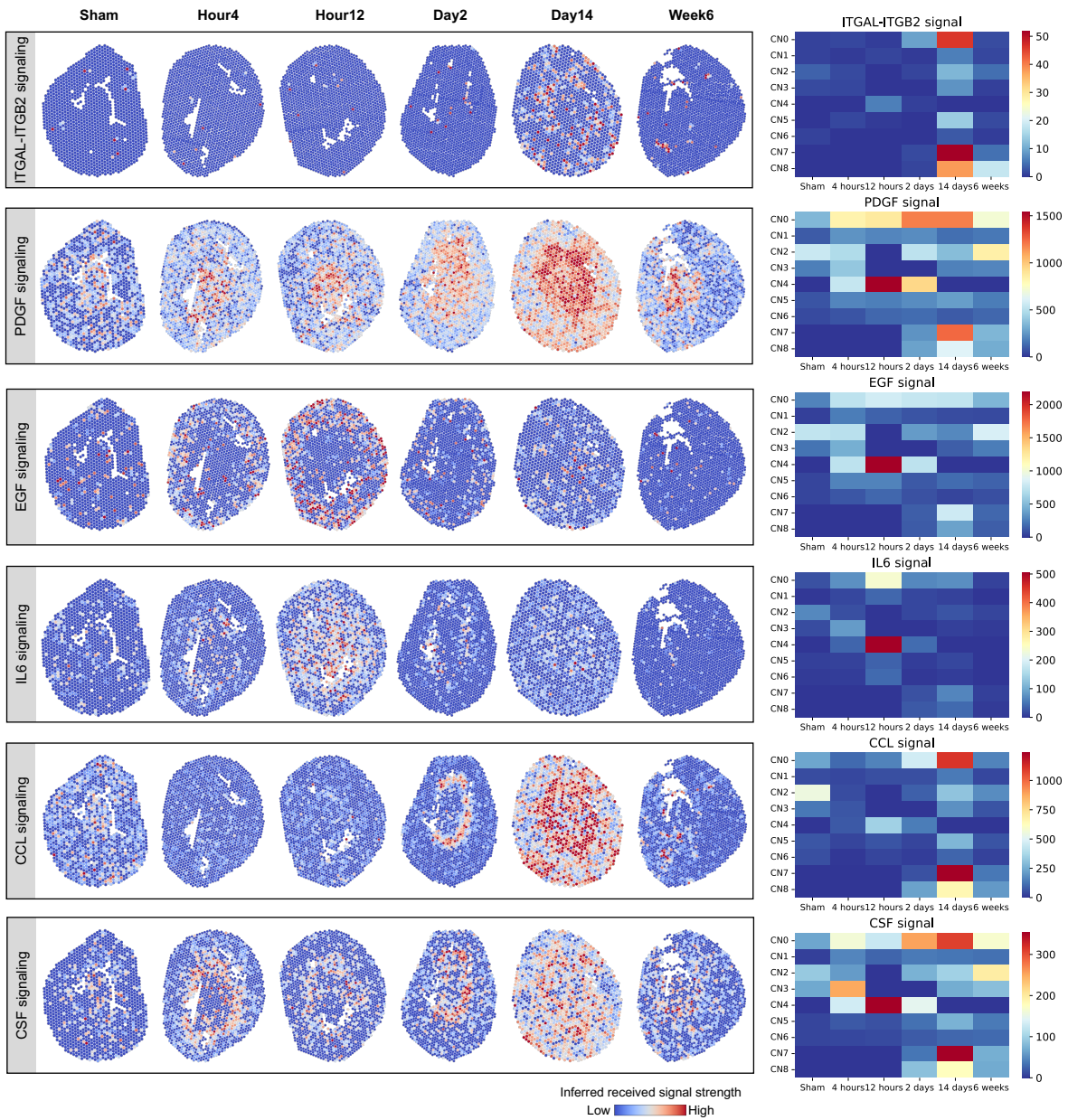


C



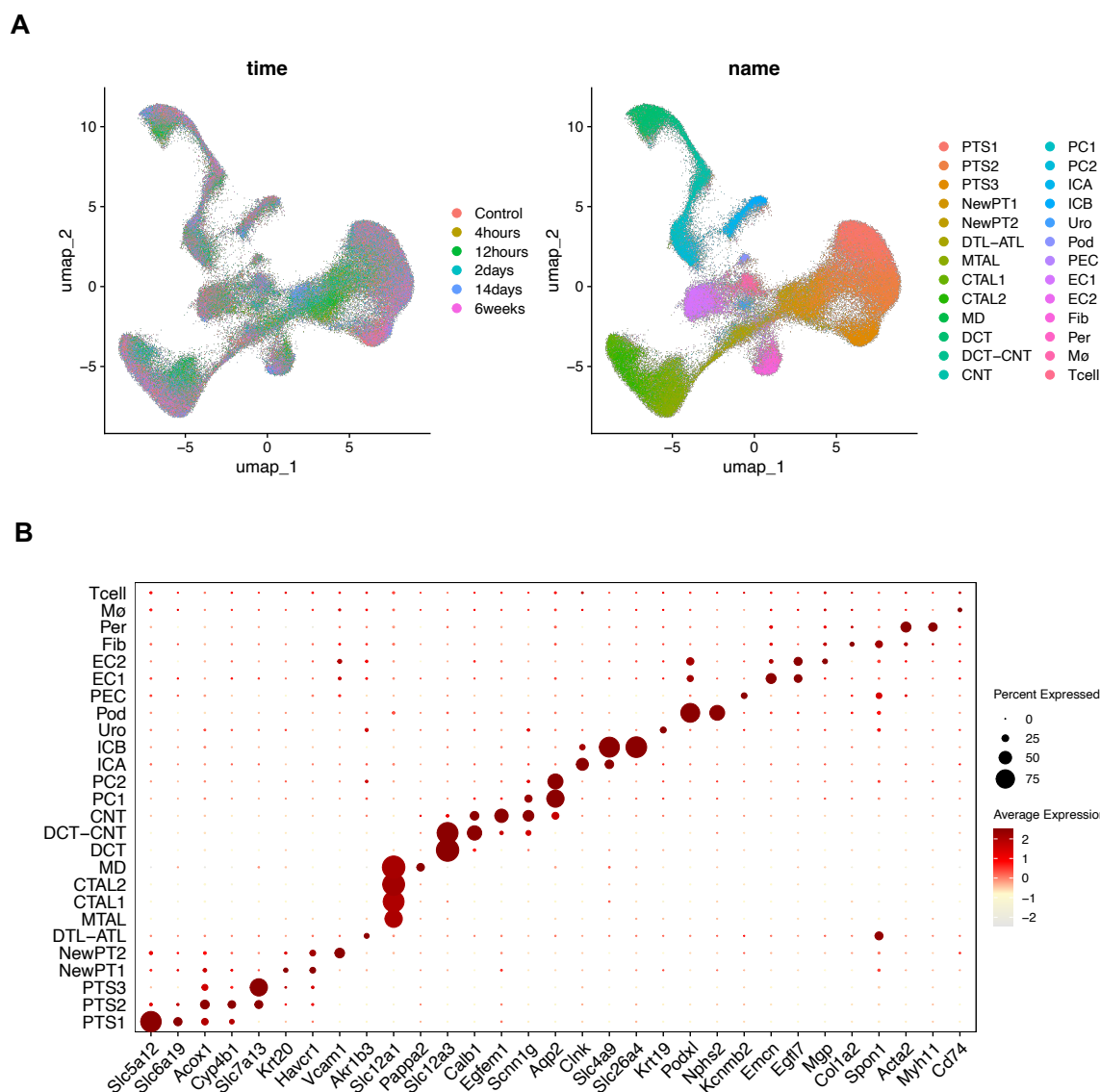
Supplementary Fig. 10 Xenium gene imputation.

(A) Heatmap comparing the spatial gene expression correlation between true detected and imputed values using matched snRNA-seq data (Kirita et al., GSE139107). Those genes were held out as a validation set during the imputation process. **(B)** Dot plot of differentially expressed genes of failed repair PT (FR-PT), Injured PT, fibroblasts and immune cells using imputed dataset. **(C)** Comparison between Xenium imputed gene expression with Visium detection. Those genes were not included in the Xenium gene panel.



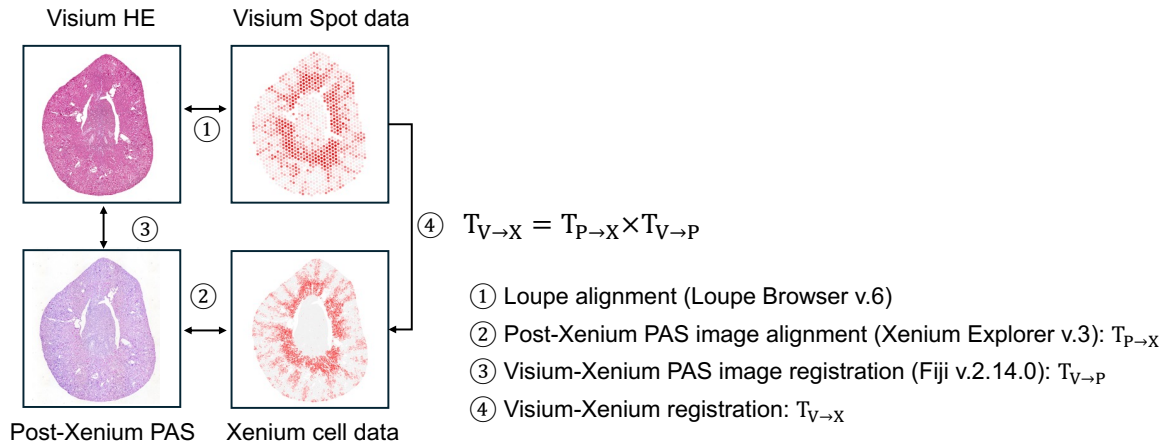
Supplementary Fig. 11 Spatiotemporal dynamics of signaling activities.

The left panel shows the spatiotemporal distribution of inferred signaling activity. The right panel displays the quantified total signal activity across each cell neighborhood. CN0: Loop of Henle; CN1: Glomerular Niche; CN2: Cortical Proximal Tubule; CN3: Medullary Proximal Tubule; CN4: Injured Proximal Tubule; CN5: Collecting Duct Niche; CN6: Distal Tubule Niche; CN7: Fibro-inflammatory Niche; CN8: Uro-immune Niche.



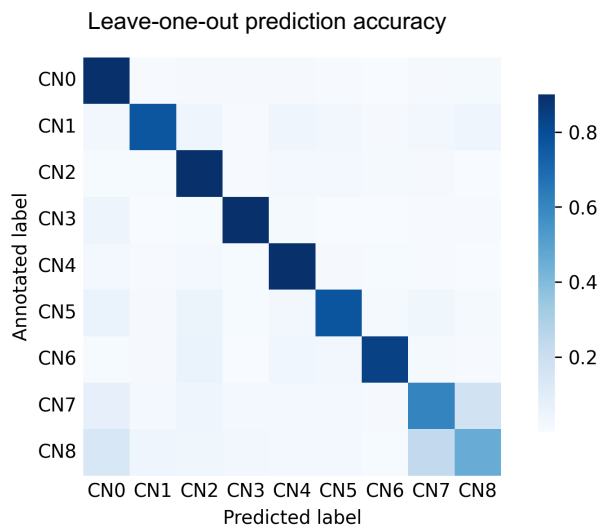
Supplementary Fig. 12 Xenium gene panel validation in snRNA-seq data.

(A) UMAP projection of snRNA-seq data clustered using 300 gene panel. **(B)** Dotplot showing the marker gene expression across cell types using reduced snRNA-seq data.



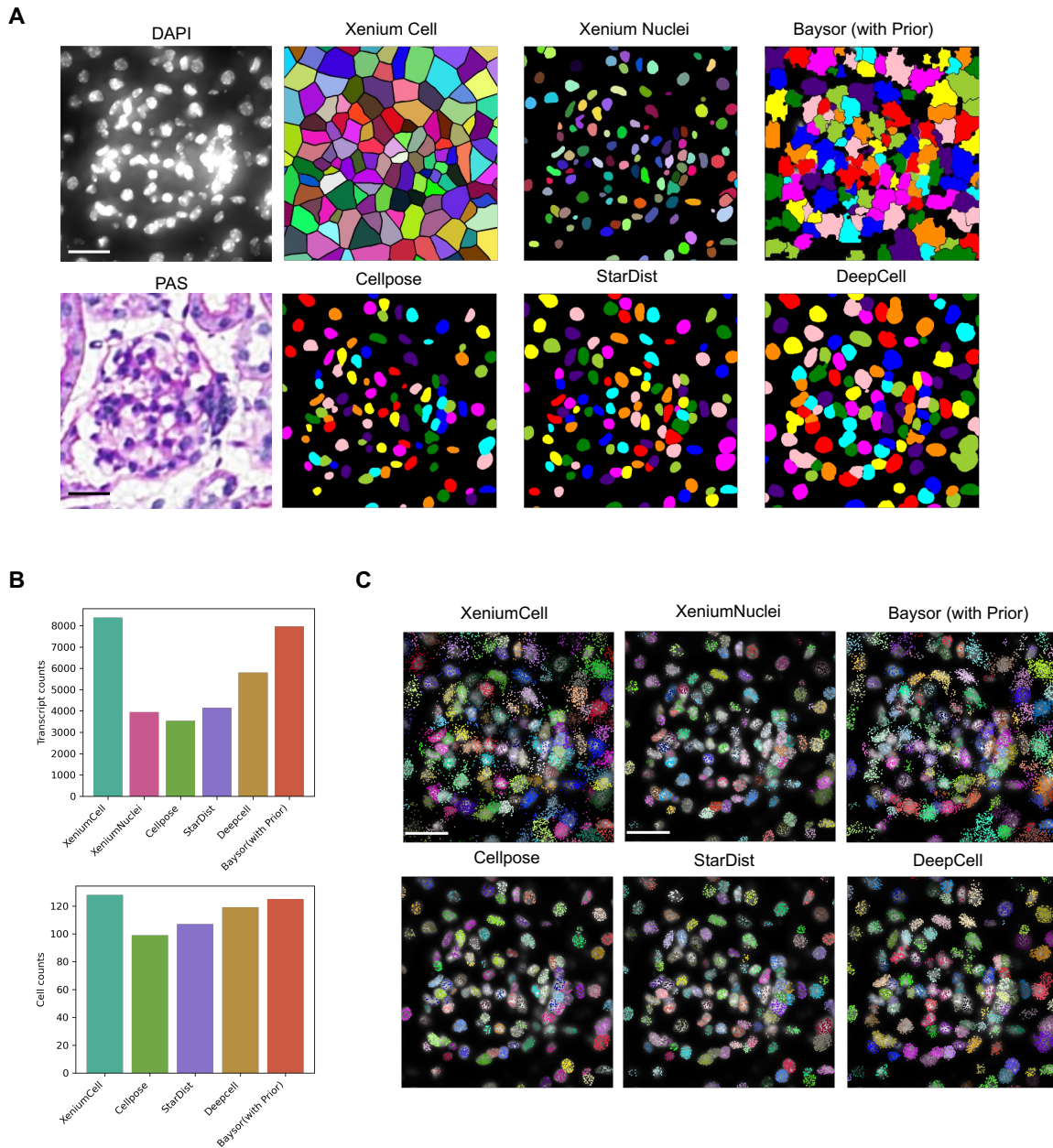
Supplementary Fig. 13 Workflow of Visium to Xenium registration.

(1). Loupe alignment was performed to align the high-resolution HE images to Visium spot coordinates (V). (2). The Xenium PAS image was aligned to the Xenium DAPI image using the Xenium Explorer software. The transformation matrix $T_{P \rightarrow X}$ maps the PAS image to the Xenium cell coordinates (X). (3). The Visium HE image was registered to the Xenium PAS image. This derived a transformation matrix ($T_{V \rightarrow P}$). (4). For the final registration, the transformation matrices from the two steps were combined to map Visium spots to Xenium coordinate system.



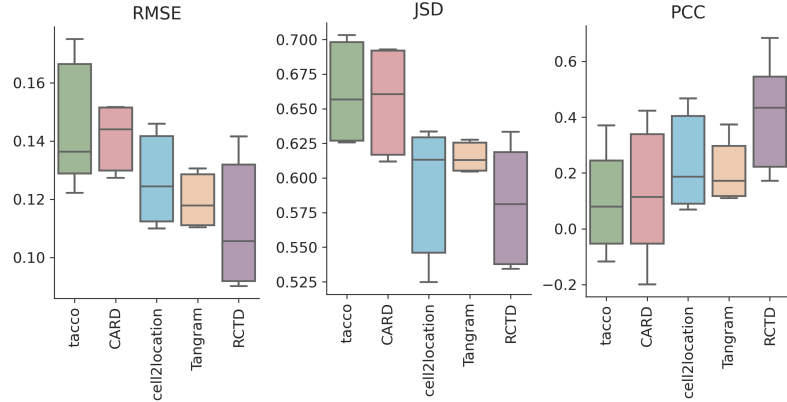
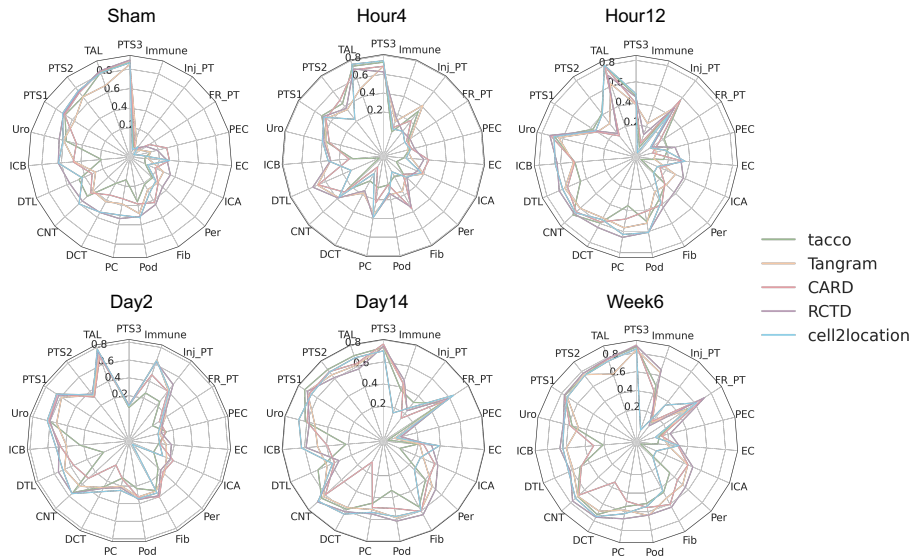
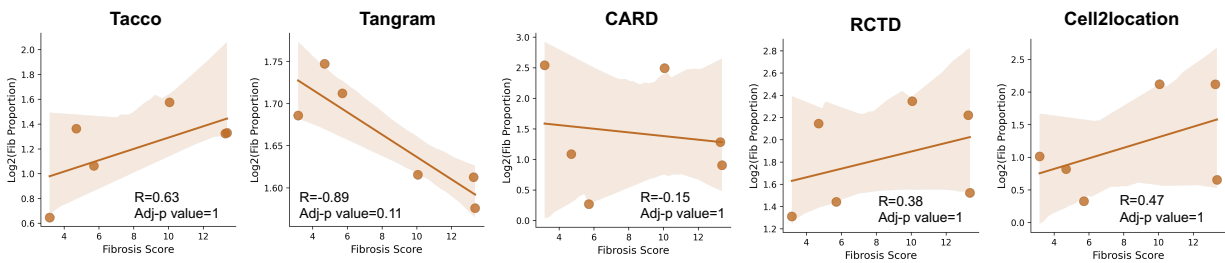
Supplementary Fig. 14 Xenium cell neighborhood prediction accuracy.

Leave-one-out cross-validation performance across cell neighborhood clusters between Xenium and Visium datasets.



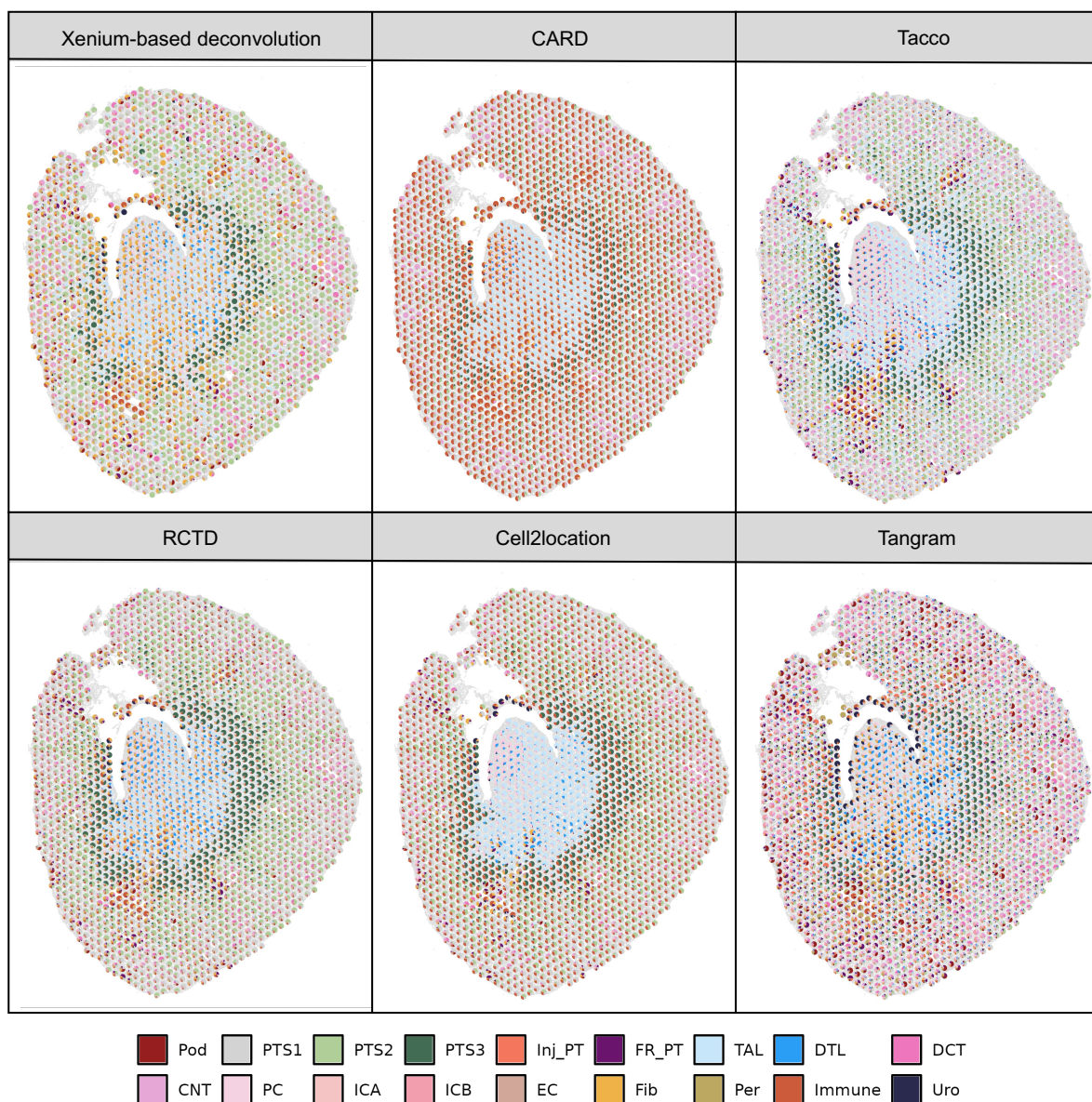
Supplementary Fig. 15 Benchmarking segmentation algorithm on Xenium data.

(A) Comparison of segmentation results. **(B)** Comparison of retrieved cell counts and transcript counts. **(C)** visualization of retrieved transcripts after remapping to segmented cells using different algorithms. Transcripts are colored by their corresponding segmented cell ID. XeniumCell: default Xenium nuclei segmentation with mask expansion; XeniumNuclei: default Xenium nuclei segmentation; Baysor (with Prior): Baysor segmentation using Xenium nuclei segmentation as a prior. Scale bar: 20 μ m.

A**B****C**

Supplementary Fig. 16 Benchmarking deconvolution algorithm on Visium data.

(A) Box plots comparing the root mean square error (RMSE), Pearson correlation coefficient (PCC) and Jensen-Shannon divergence (JSD) across different deconvolution methods used to predict cell type composition of the spots. (n=6) **(B)** Radar plot of PCC across deconvolution methods for each individual cell type and time point. Source data are provided in Source Data file. **(C)** Correlation between the deconvoluted fibroblast abundance and the fibrosis score (quantified from Sirius Red staining) across different deconvolution algorithms.



Supplementary Fig. 17 Cell type deconvolution algorithms evaluation in kidney tissue.

Visualization of cell type composition in kidney (week6 sample) estimated by xenium and five deconvolution algorithms: RCTD, CARD, Cell2location, Tacco and Tangram. Cell type abundance within each spot is visualized as a pie chart.

Supplementary References

1. Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.* **40**, 345–354 (2022).
2. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).
3. Weigert, M. & Schmidt, U. Nuclei instance segmentation and classification in histopathology images with stardist. in *2022 IEEE International Symposium on Biomedical Imaging Challenges (ISBIC)* 1–4 (IEEE, 2022).
4. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022).
5. Mages, S. *et al.* TACCO unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotechnol.* **41**, 1465–1473 (2023).
6. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
7. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
8. Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nat. Biotechnol.* **40**, 1349–1359 (2022).
9. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).