

Ψ



THE INDIANA UNIVERSITY O'BRIEN CENTER
PRESENTS

**FROM BASIC INFORMATICS TO
MACHINE LEARNING:**

Sponsors

- NIH O'Brien Kidney Resource Alliance
 - U54 DK137328 (Dagher)
- IU Center for Computational Biology and Bioinformatics
- OSU Department of Biomedical Informatics
- NIDDK / NSF
 - R01 DK138504 (Wang / Ma / Eadon)



**Indiana O'Brien Center for
Advanced Renal Microscopy
and Molecular Imaging**
NIDDK U54 DK137328

IU Center for Computational Biology and Bioinformatics



OSU Department of Biomedical Informatics



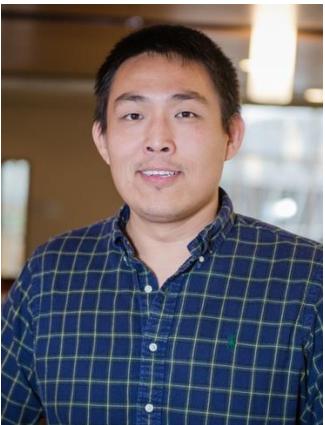
Organizing Committee



Juexin Wang, PhD
Indiana Univ



Ricardo Melo Ferreira, PhD
Indiana Univ



Qin Ma, PhD
Ohio State Univ



Michael Eadon, MD
Indiana Univ

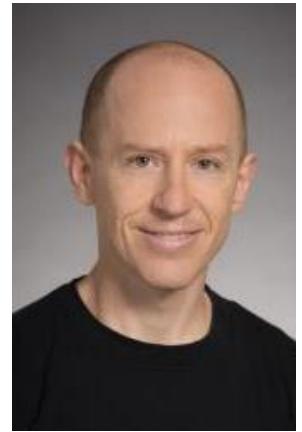
O'Brien Leadership



Pierre Dagher, MD
Indiana Univ



Tarek Ashkar, MD
Indiana Univ



Josh Vaughn, PhD
Univ Washington



Sharon Moe, MD
Indiana Univ



Amanda Anderson, PhD
Univ Alabama Birmingham

Guest Speakers



KEYNOTE SPEAKER

Mingyao Li, PhD

Unlocking the power of Spatial Omics with AI

Professor of Biostatistics in Biostatistics and Epidemiology

Senior Scholar, Center for Clinical Epidemiology and Biostatistics

Director of Biostatistics

Director, Statistical Center for Single-Cell and Spatial Genomics

University of Pennsylvania Perelman School of Medicine

Tuesday 11:05 am ET



Dong Xu, PhD

Using ChatGPT and Machine learning in single cell multi-omics studies and in kidney research

Curators' Distinguished Professor

Paul K. and Dianne Shumaker Professor

University of Missouri

Monday 11:50 pm ET

Graduate Student Tutorials



Mauminah Raina

PhD Candidate

Indiana University

Monday - Introduction to biological analyzing methods

Tuesday - Spatial transcriptomics dataset acquisition, model fitting, and analysis



Mahla Asghari

PhD Candidate

Indiana University

Monday - Spatial localization of cell types



Yi Jiang

PhD Candidate

Ohio State University

Tuesday - Single-cell multi-omics dataset acquisition, model training and analysis

GitHub Access and Tutorial Help Questions

- OKRA Tutorial 2024 - https://github.com/juexinwang/OKRA_Tutorial2024
- scGNN - <https://github.com/juexinwang/scGNN>
- DeepMAPS - <https://bmblx.bmi.osumc.edu/>
- BSP - <https://github.com/juexinwang/BSP>
- Niche - https://github.com/rimelof/intro_spatial_localization_workshop/
- Send questions to Joshua Kuhn
- We will send a link to the Microsoft Teams breakout room for 1:1 help



From Basic Informatics to Machine Learning

Michael Eadon, MD

Indiana University School of Medicine

12-2-24

Disclosures

- NIH funding

Contents

1. Data Sources
2. Quality considerations
3. Analysis approaches
4. Histologic Annotation with AI

Data Sources

KIDNEY PRECISION MEDICINE PROJECT

www.kpmp.org



TRANSCRIPTOMICS

- Single Cell RNA-seq
- Single Nucleus RNA-seq
- Regional Transcriptomics
- Spatial Transcriptomics
- Slide-seq Spatial Transcriptomics

EPIGENETICS

- DNA Methyl-seq •
- microRNA-seq •
- CUT & RUN •
- ATAC-seq •



INTEGRATED FUNCTIONAL KIDNEY MAP

IMAGING

- 3D Tissue Imaging •
- and Cytometry
- mlFISH •
- CODEX •
- Imaging Mass Cytometry •
- DART-FISH •



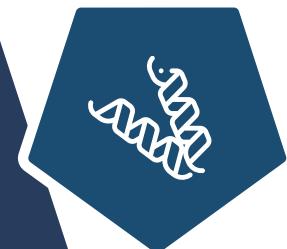
TRANSCRIPTOMICS

- Single Cell RNA-seq
- Single Nucleus RNA-seq
- Regional Transcriptomics
- Spatial Transcriptomics
- Slide-seq Spatial Transcriptomics



PROTEOMICS

- Regional Proteomics
- Spatial Proteomics



METABOLOMICS

- Spatial Metabolomics
- Spatial Lipidomics
- Spatial N-glycomics



Altruistically donated biopsy specimens in kidney disease

HuBMAP

The Human BioMolecular Atlas Program

HuBMAP

About ▾ Data ▾ News & Events ▾ Member Services ▾ Data Portal

What is HuBMAP?

The Human BioMolecular Atlas Program is working to catalyze the development of a framework for mapping the human body at single cell resolution.

[Find Out More](#)



1039 1013

Datasets

Samples



Explore our Data

86

Donors

21

Organs

Goal to map healthy cells in the human body across multiple organs.

KPMP data atlas

- <https://atlas.kpmp.org/>

Go forward one page (Alt+Right Arrow)
Right-click or pull down to show history as [Dashboard \(Home\)](#) [Explorer](#) [Repository](#) [Spatial Viewer](#) [Help ▾](#)

KIDNEY PRECISION MEDICINE PROJECT

Kidney Tissue Atlas

The Kidney Tissue Atlas is a set of interactive tools built to promote retrieval, exploration, discovery, and analysis of the KPMP data by the greater research community.
[Learn more about our data types and methodologies](#)

Explorer
Search for markers or cell types of interest and view summary data visualizations across the various KPMP 'omics' technologies.

[Go to Explorer](#)

Atlas Data Summary

IMPORTANT: Please follow this [citation guideline](#) when presenting or publishing data from the Kidney Tissue Atlas.

 **PARTICIPANTS**

281	70	19	125
CKD	AKI	DM-R	REFERENCE

Participants by -omics type
A subset of the raw data from the Data Repository has been analyzed and made available for interactive mining in Explorer and Spatial Viewer.

OMICS TYPE	REFERENCE	CKD	AKI	DM-R	ALL
Explorer					

Repository
Download raw gene data generated from KPMP participant biopsies and reference tissue samples.

[Go to Repository](#)

Spatial Viewer
View and interact with spatial data from various KPMP

HuBMAP Data Repository

<https://portal.hubmapconsortium.org>

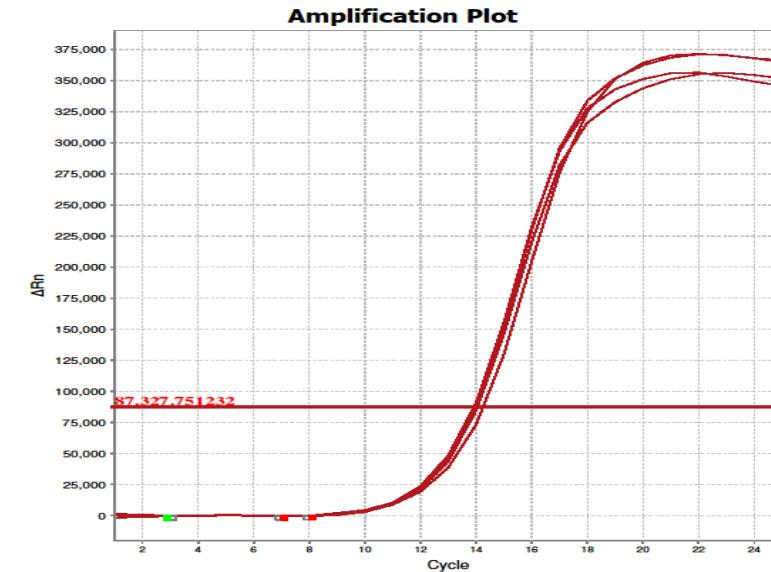
The screenshot shows the HuBMAP Data Portal interface. At the top, there is a navigation bar with links for 'HuBMAP', 'Data', and 'Resources'. Below the navigation bar, the title 'Datasets' is displayed next to a small icon. A modal window titled 'Getting Started' is open, welcoming users to the portal and providing a link to 'Begin The Dataset Search Tutorial'. The main content area features a search bar with a magnifying glass icon and a 'Metadata' dropdown menu. Below the search bar, several filters are applied: 'Dataset Type: Slideseq', 'Dataset Type: Slideseq [Salmon]', 'Organ: Kidney (Right)', 'Organ: Kidney (Left)', and 'Analyte Class: RNA'. On the left side, a sidebar displays filter counts for 'Dataset Type' (RNAseq: 248, Slide-seq: 84), 'Organ' (Kidney (Left): 46, Kidney (Right): 38), 'Analyte Class' (RNA: 84), and 'Sample Category' (section: 46, block: 38). The main table lists dataset metadata for each entry, including HuBMAP ID, Group, Data Types, Organ, Status, and Last Modified date. The first few entries are: HBM232.MBNR.586 (Group: Broad Institute RTI, Data Types: Slideseq [Salmon], Organ: Kidney (Right), Status: Published, Last Modified: 2023-11-07 00:34:27); HBM986.KFWG.239 (Group: Broad Institute RTI, Data Types: Slideseq [Salmon], Organ: Kidney (Right), Status: Published, Last Modified: 2023-11-07 00:32:26); and HBM532.KKRC.477 (Group: Broad Institute RTI, Data Types: Slideseq [Salmon], Organ: Kidney (Right), Status: Published, Last Modified: 2023-11-07 00:30:25).

Dataset Metadata	HuBMAP ID	Group	Data Types	Organ	Status	Last Modified
Dataset Type: Slideseq	HBM232.MBNR.586	Broad Institute RTI	Slideseq [Salmon]	Kidney (Right)	Published	2023-11-07 00:34:27
Dataset Type: Slideseq [Salmon]	HBM986.KFWG.239	Broad Institute RTI	Slideseq [Salmon]	Kidney (Right)	Published	2023-11-07 00:32:26
Organ: Kidney (Right)	HBM532.KKRC.477	Broad Institute RTI	Slideseq [Salmon]	Kidney (Right)	Published	2023-11-07 00:30:25
Organ: Kidney (Left)	HBM846.KVCF.674	Broad Institute RTI	Slideseq [Salmon]	Kidney (Right)	Published	2023-11-07 00:28:25
Analyte Class: RNA	HBM766.NZWP.682	Broad Institute RTI	Slideseq [Salmon]	Kidney (Right)	Published	2023-11-07 00:26:24

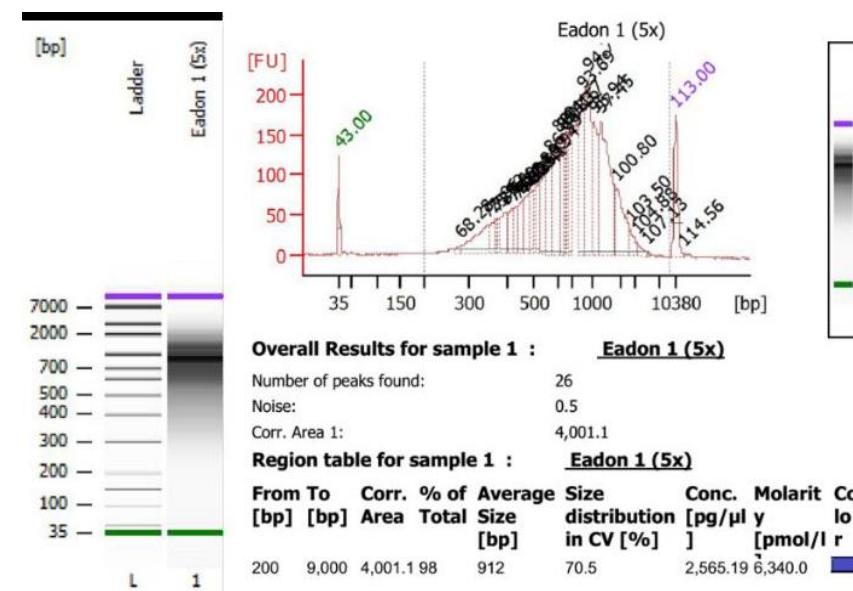
Quality Assurance Considerations

Quality Control Metrics

- QC metrics:
 - Bulk RNA quality (minimum RIN or DV200 by bioanalyzer)
 - Tissue affixment without folds or overlap
 - Microscope image visual inspection
 - Post-permeabilization RNA quantity and quality
 - cDNA quality
 - Sequencing Q30
 - Transcript mapping under spots/cells and to exons
 - Expected differential expression of marker genes



RNA assessment after permeabilization and isolation using rtPCR



cDNA quality assessed after amplification

Sequencing Output Metrics

3,007

Number of Spots Under Tissue

71,192

Mean Reads per Spot

2,970

Median Genes per Spot

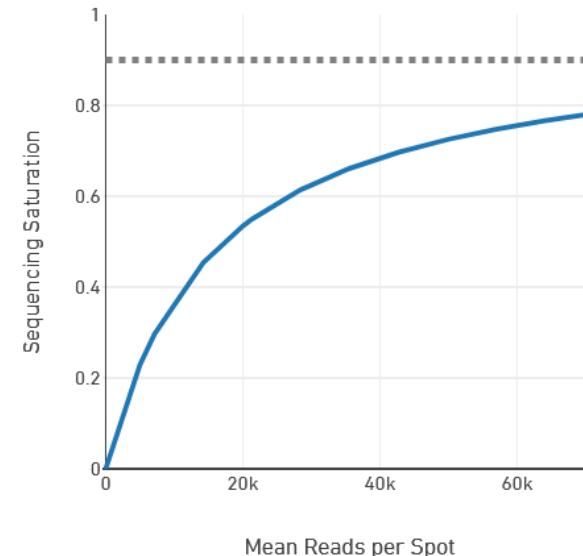
Sequencing ?

Number of Reads	214,074,559
Valid Barcodes	97.2%
Valid UMIs	99.9%
Sequencing Saturation	78.2%
Q30 Bases in Barcode	96.4%
Q30 Bases in RNA Read	93.2%
Q30 Bases in UMI	96.2%

Mapping ?

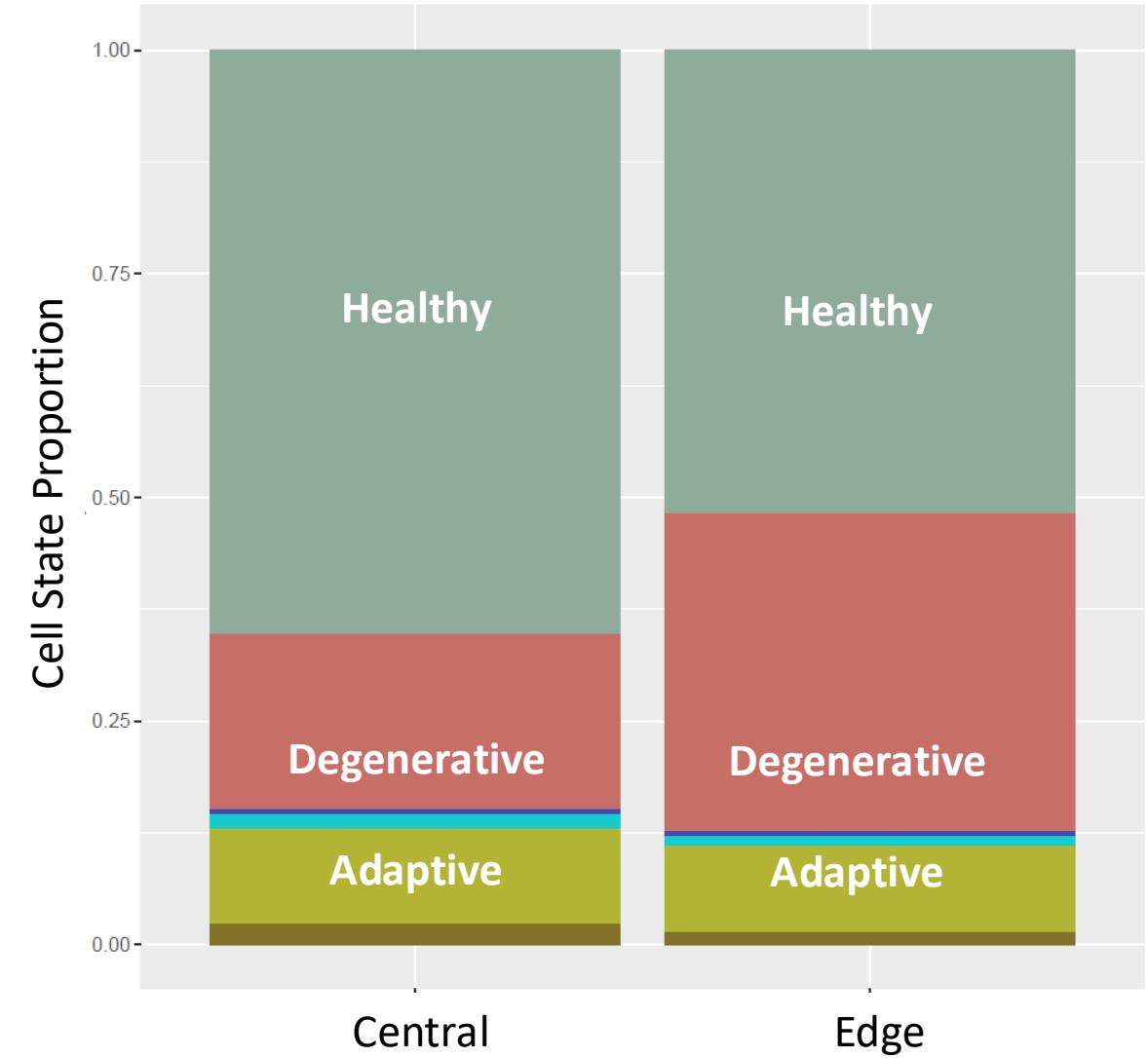
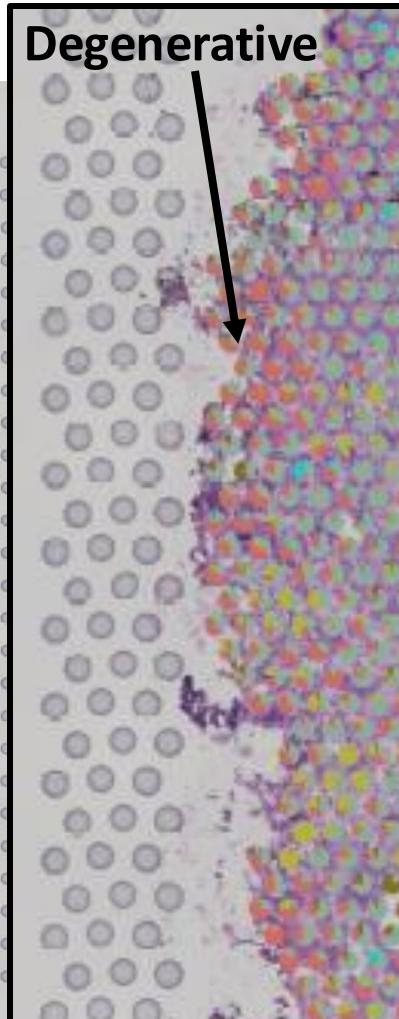
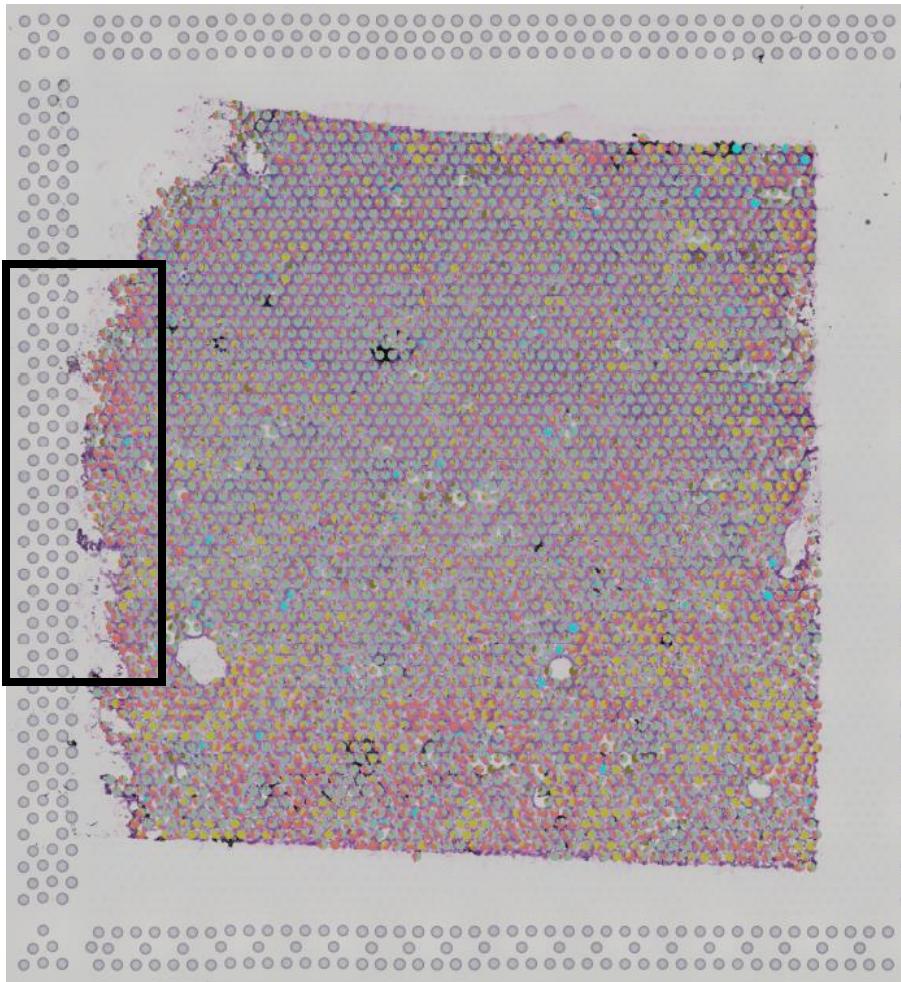
Reads Mapped to Genome	92.6%
Reads Mapped Confidently to Genome	87.9%
Reads Mapped Confidently to Intergenic Regions	4.9%
Reads Mapped Confidently to Intronic Regions	2.5%
Reads Mapped Confidently to Exonic Regions	80.5%
Reads Mapped Confidently to Transcriptome	75.9%
Reads Mapped Antisense to Gene	0.8%
Fraction Reads in Spots Under Tissue	87.2%
Mean Reads per Spot	71,192
Median Genes per Spot	2,970
Total Genes Detected	21,721
Median UMI Counts per Spot	8,349

Sequencing Saturation ?

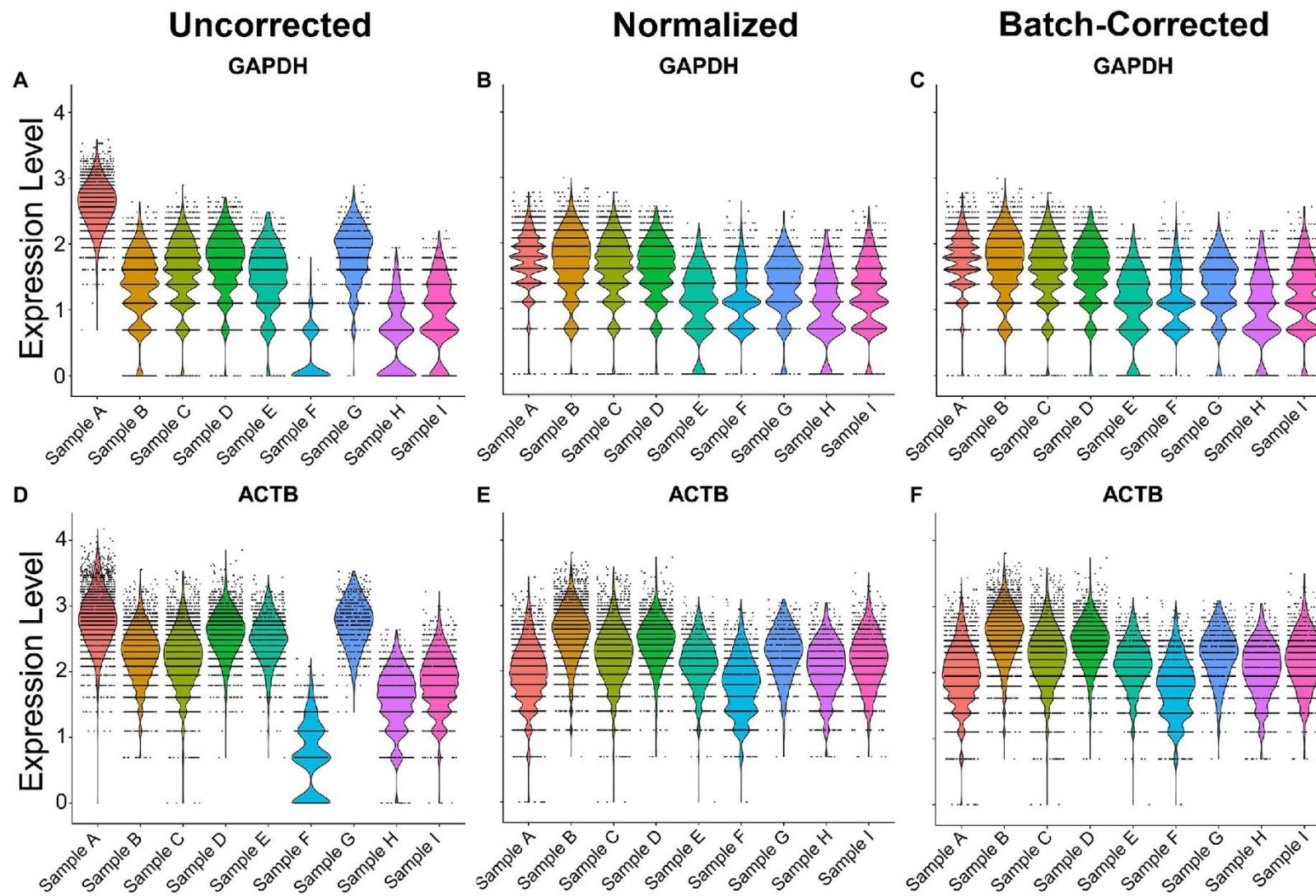


Artifactual Cell State Distribution

- We remove the outer border of cells or spots

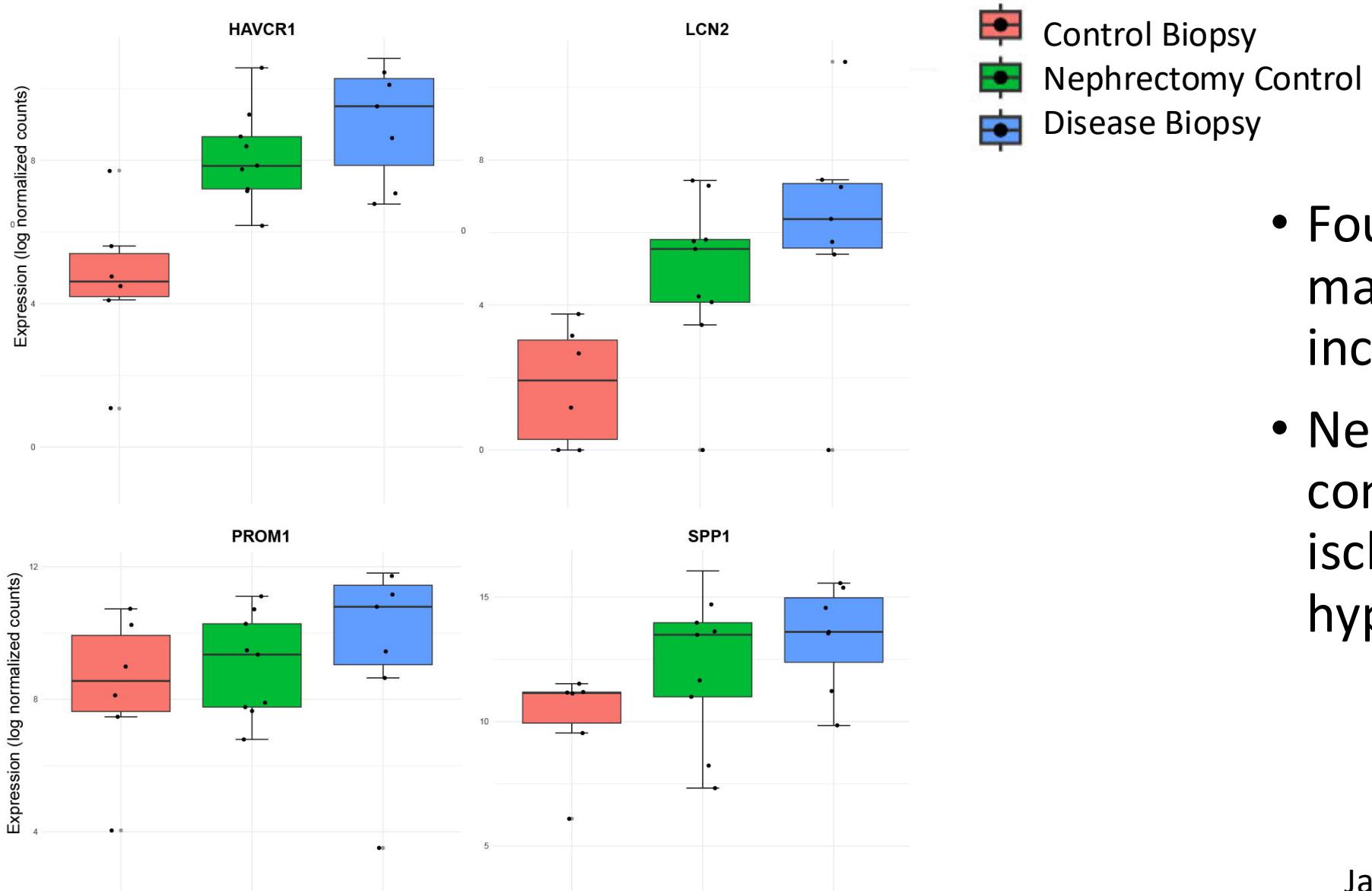


Normalization and Batch Correction



- Normalized *GAPDH* expression with SCTransform
- Normalization and batch correction were performed by adding a batch as a variable in SCTransform.

Not all controls are equal



- Four kidney injury markers are increased in disease
- Nephrectomy controls have cold ischemia time (i.e. hypoxic injury)

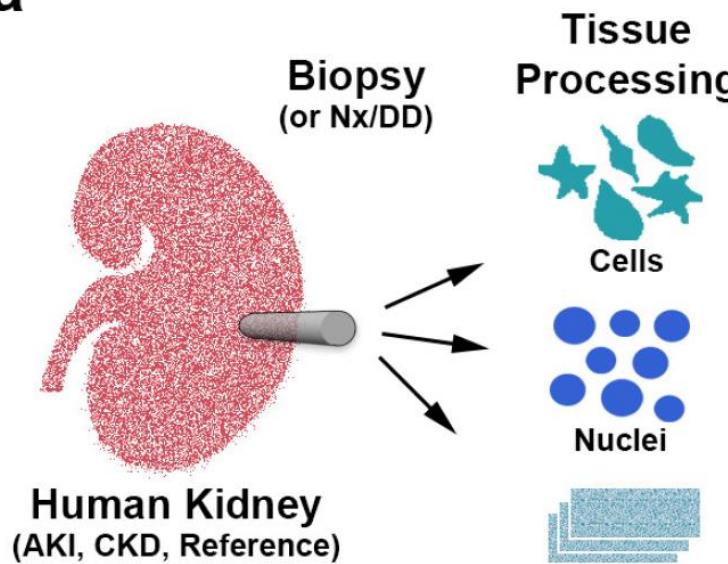
Analysis Approaches

- Label Transfer
- Deconvolution
- Neighborhood Analysis
- Spatial Variable Gene Analysis
- Segmentation

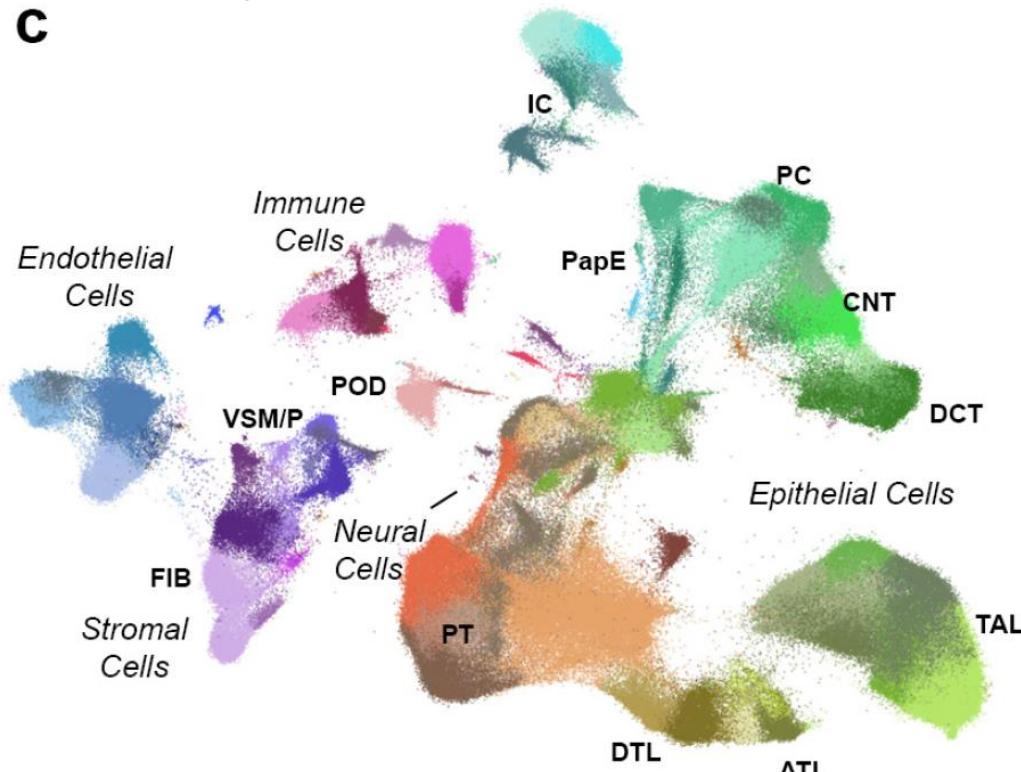
Kidney Cell Atlas Overview: Technologies

100 cell types defined, 6 cell states, 122 samples

a

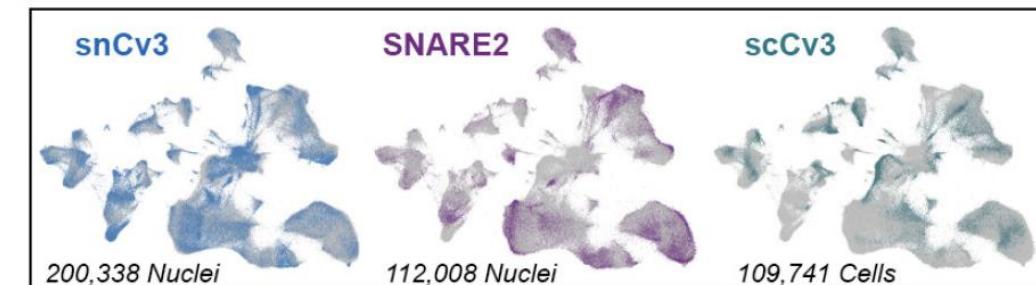
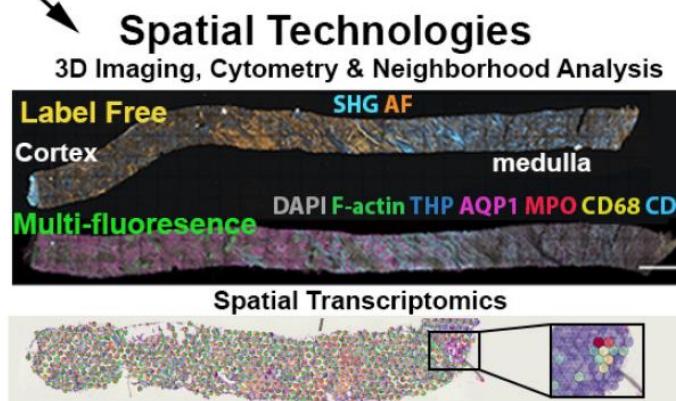


c

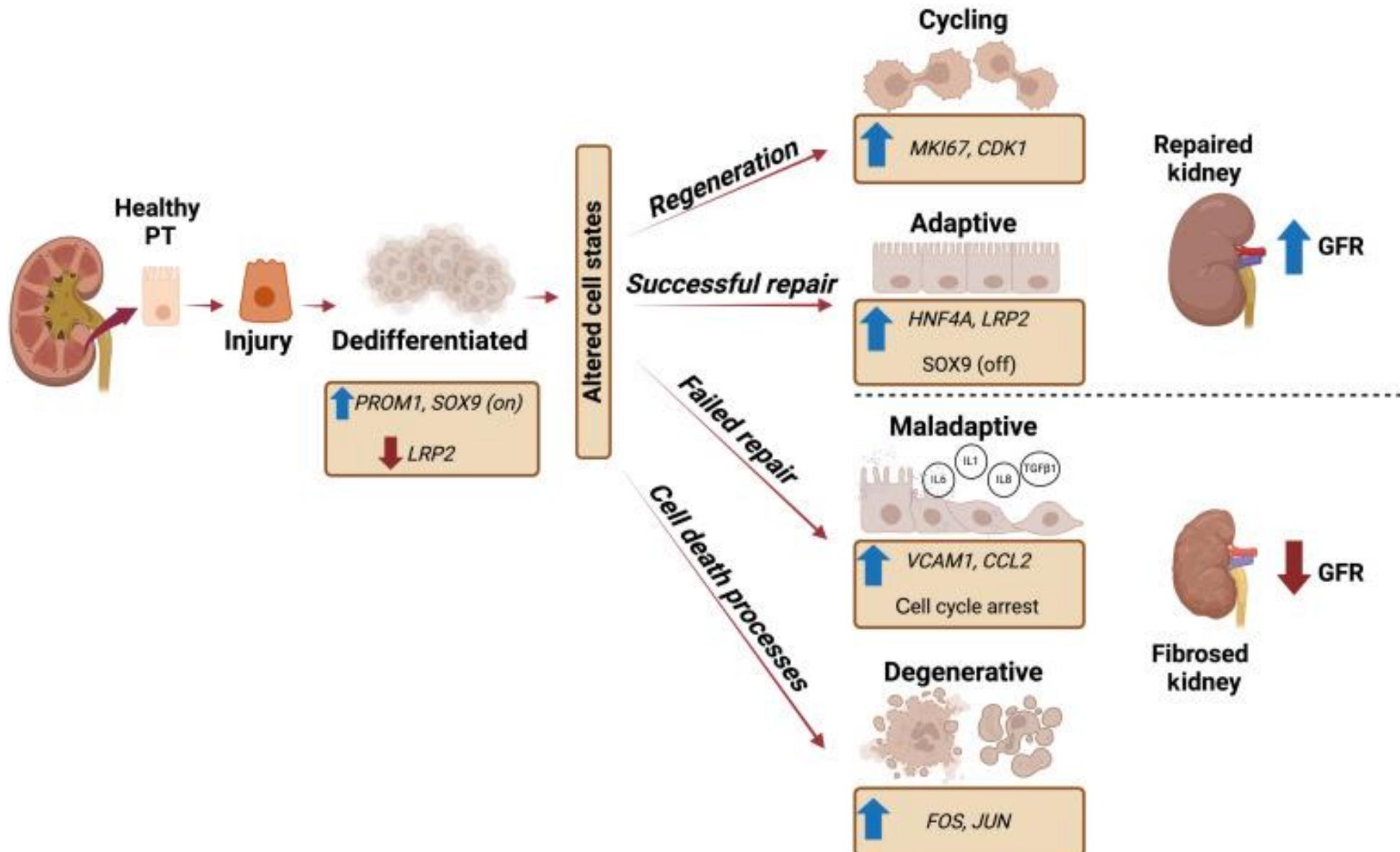


b

Assay	Sex		Condition		
	♂	♀	CKD	AKI	Ref
10X snRNA-Seq	19	17	13	10	13
10X scRNA-Seq	21	24	15	12	18
SNARE-Seq2	3	4	-	-	7
3D Imaging	6	9	9	6	-
Slide-seq2	5	1	-	-	6
Visium	12	10	10	6	6

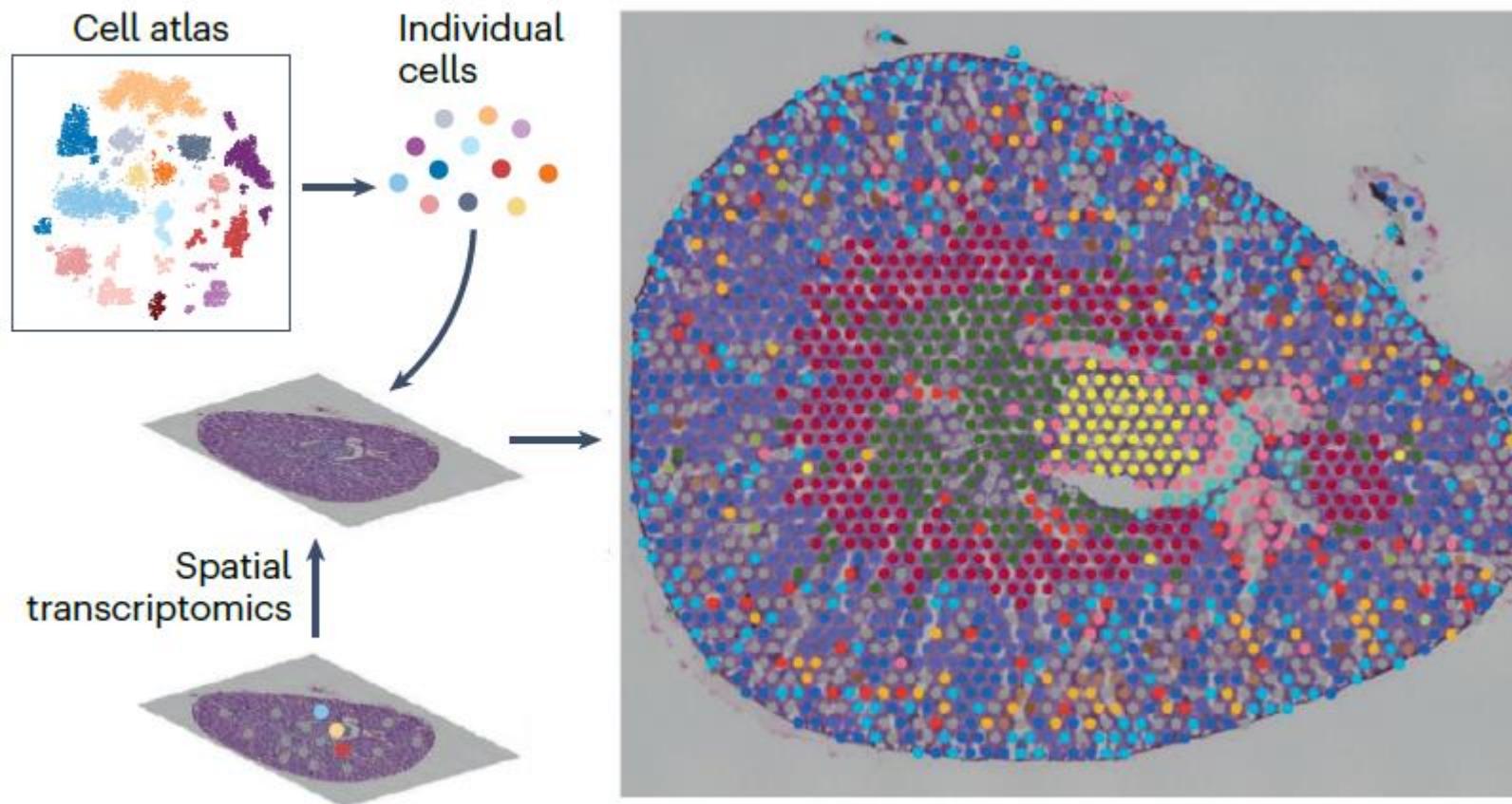


Defining Altered States in Injury



Label Transfer

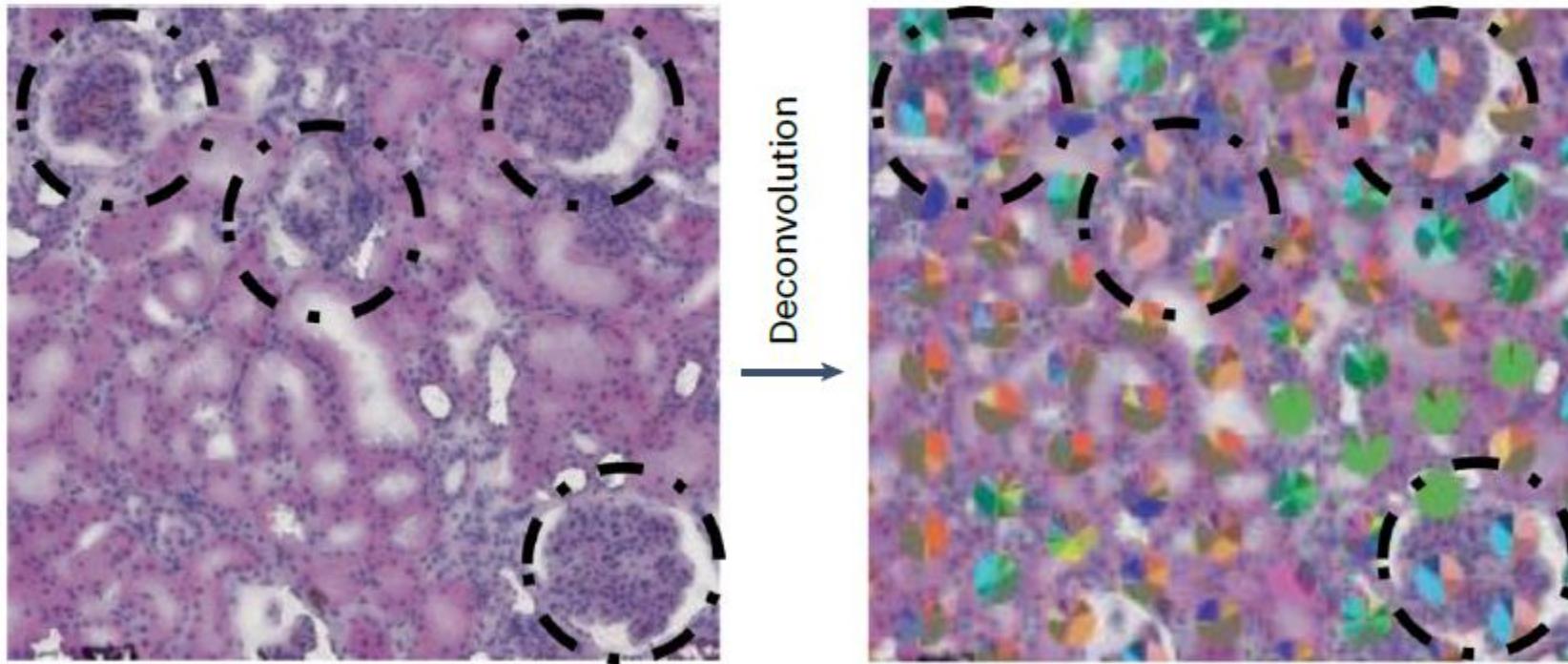
a Transfer of scRNA-seq cluster labels to ST cells or spots



Deconvolution

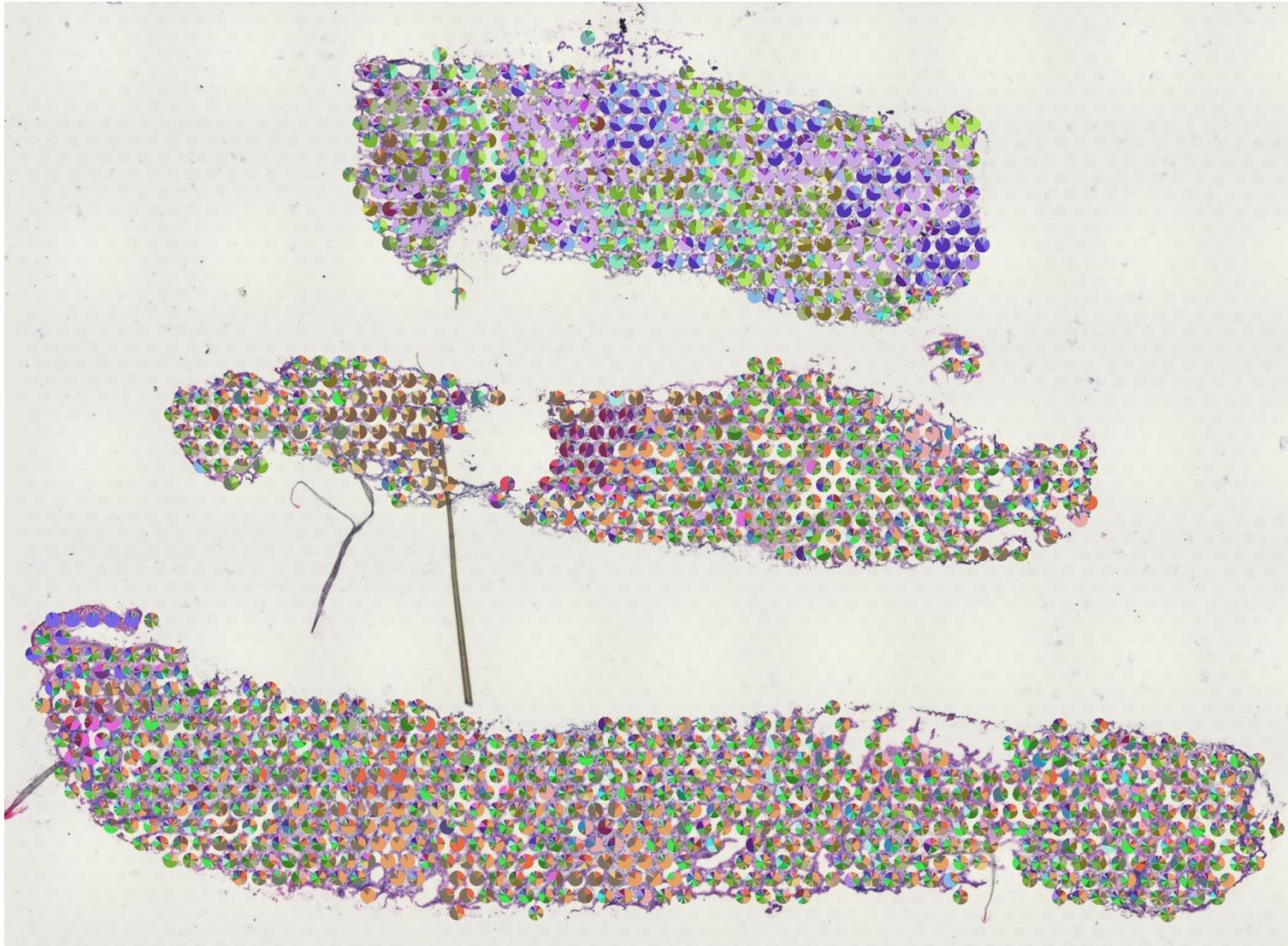
Useful when probes overly 2 or more cells

C Cell type deconvolution when indicated



Tips:

- There are an abundance of deconvolution methodologies
- Performance varies across tissue types
- Align your deconvolution with orthogonal validation (histology or protein)



Subclass

POD	M-TAL	tPC-IC	cycMYOF
dPOD	dM-TAL	IC-B	FIB
PEC	C-TAL	PapE	M-FIB
PT-S1	dC-TAL	EC-GC	dM-FIB
PT-S2	MD	EC-AEA	aFIB
PT-S3	DCT	EC-DVR	dFIB
aPT	dDCT	EC-PTC	B
cycPT	cycDCT	dEC-PTC	PL
dPT	CNT	EC-AVR	T
dPT/DTL	dCNT	dEC	NKT
DTL2	C-PC	cycEC	MAST
DTL1	M-PC	EC-LYM	MAC-M2
DTL3	dM-PC	MC	MDC
dDTL3	IMCD	REN	cDC
ATL	dIMCD	VSMC	ncMON
dATL	C-IC-A	VSMC/P	N
aTAL1	dC-IC-A	dVSMC	SC/NEU
aTAL2	M-IC-A	MYOF	

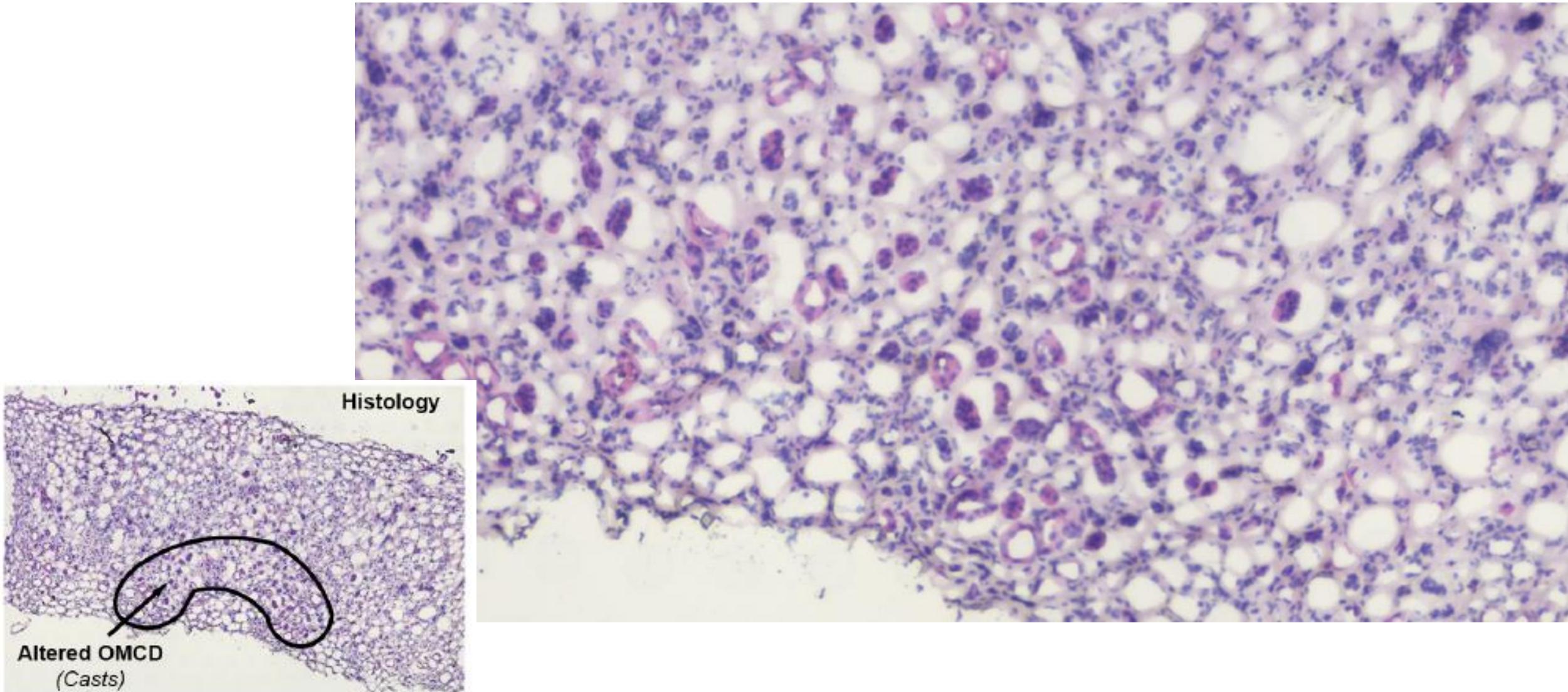
71 cell types mapped within tissue

Lake B...Melo Ferreira R...Eadon MT...Jain S, et. *Nature* 2023

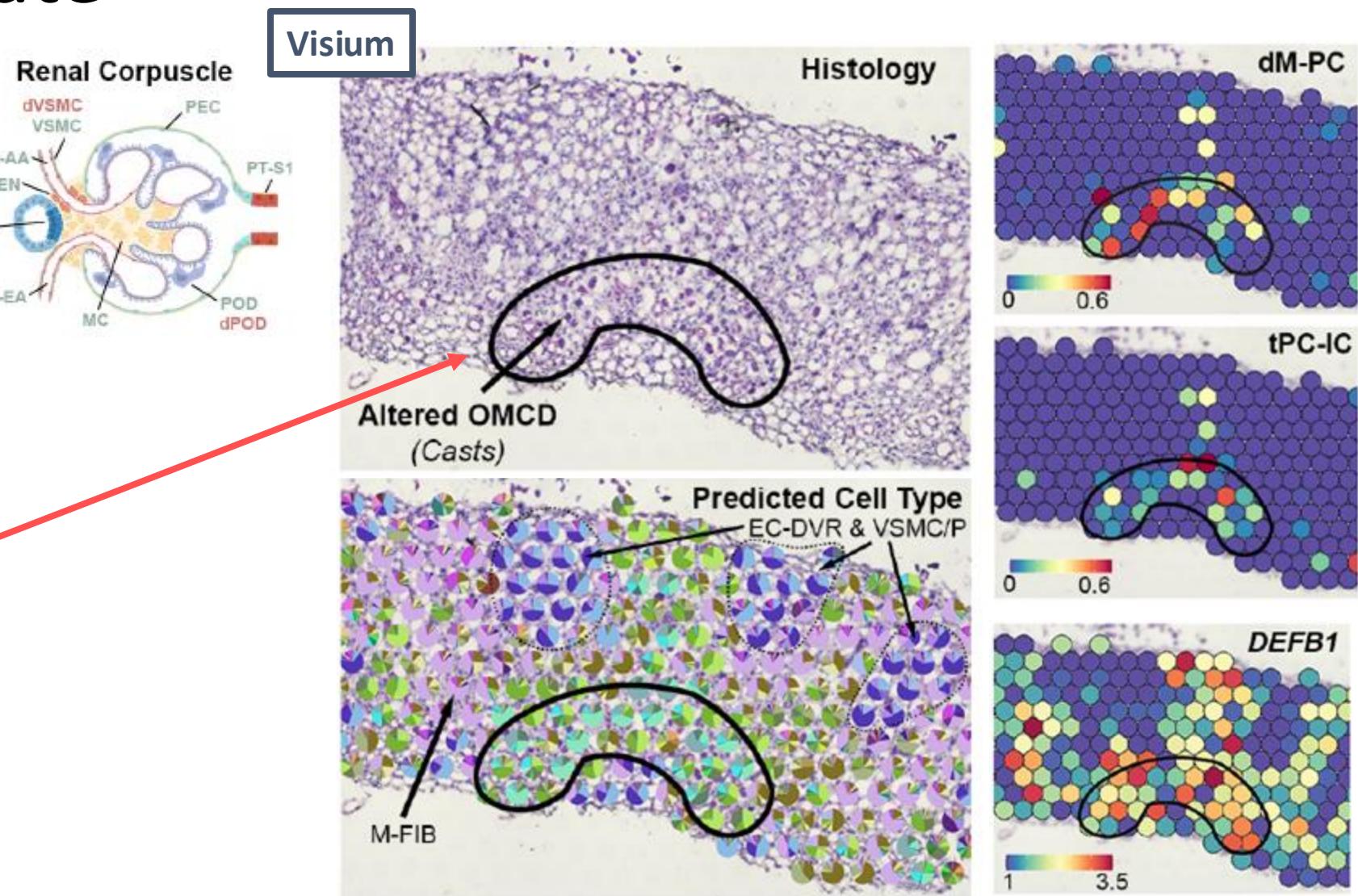
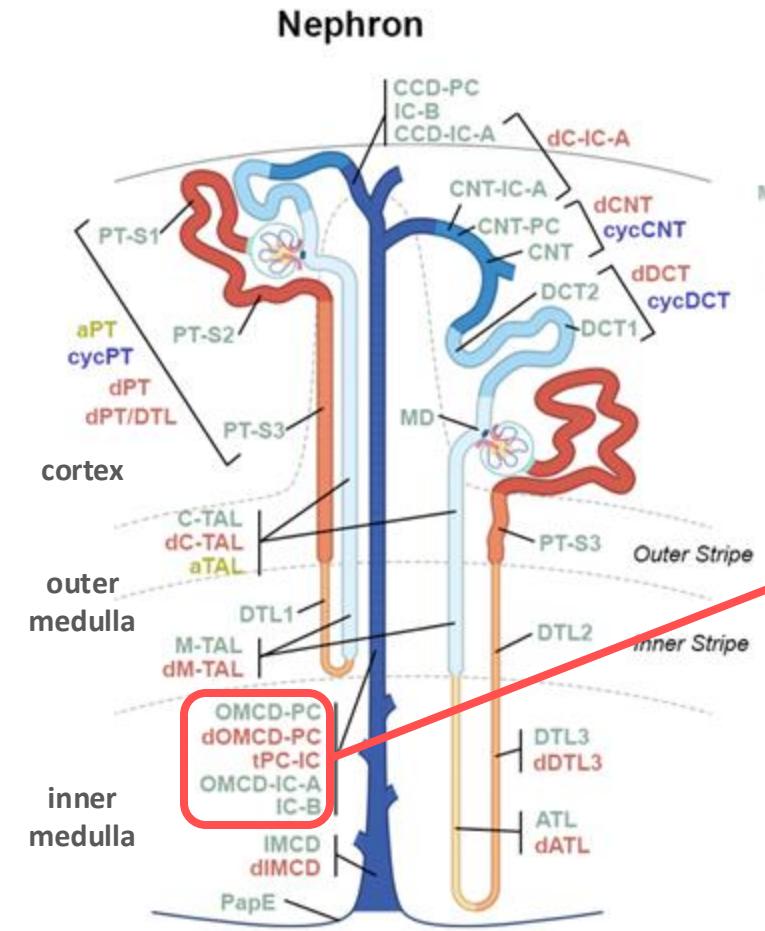


INDIANA UNIVERSITY

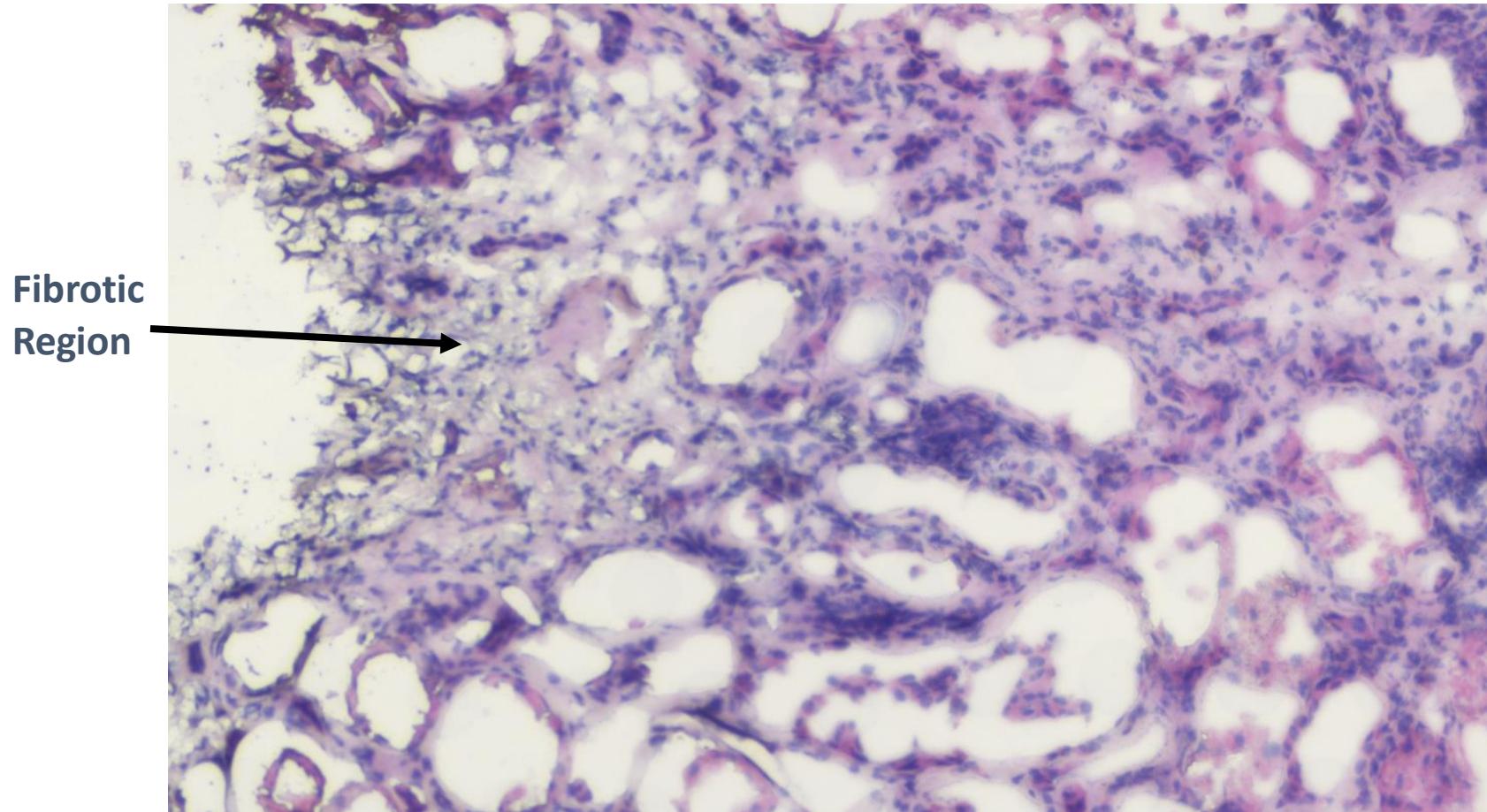
Region of casts in the medulla of a disease biopsy sample



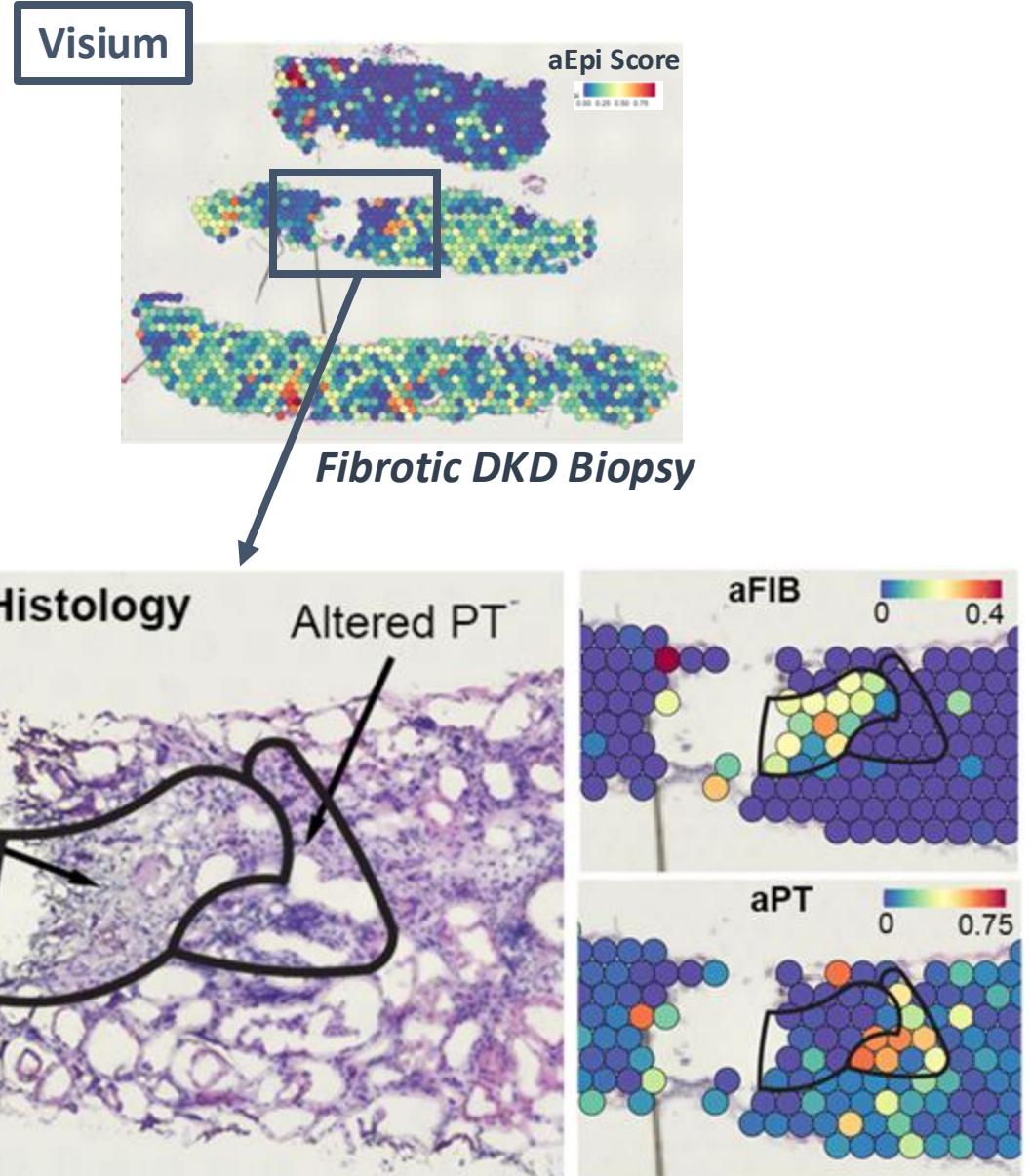
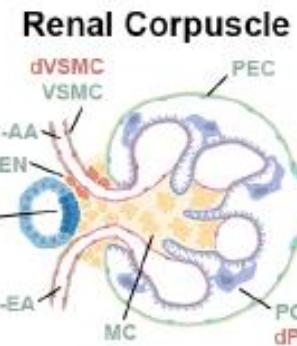
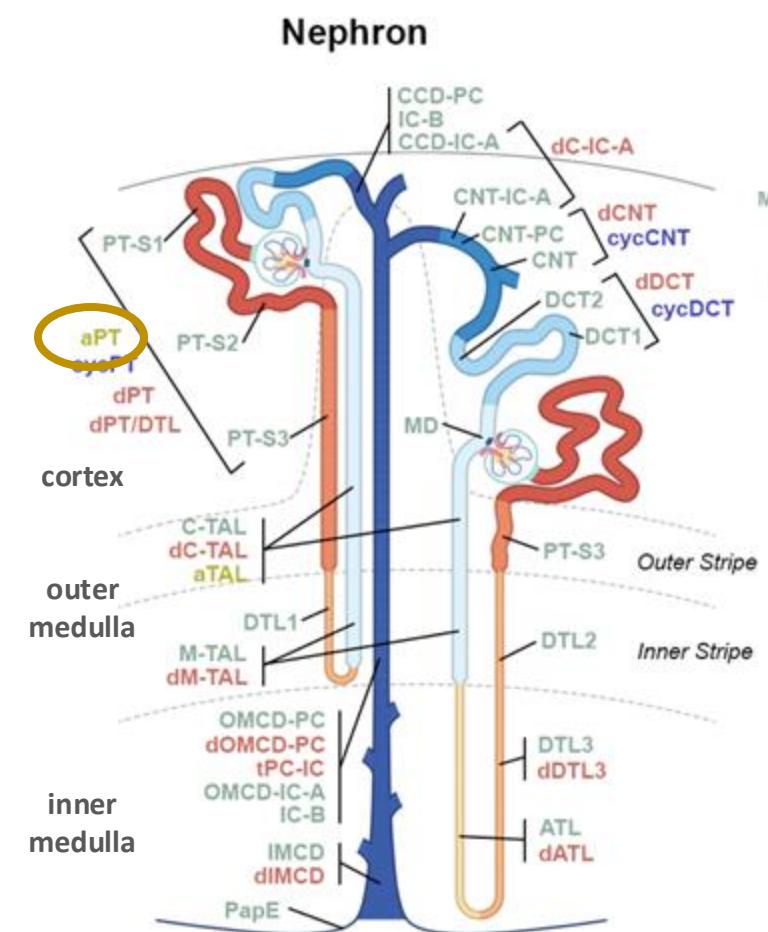
Degenerative State



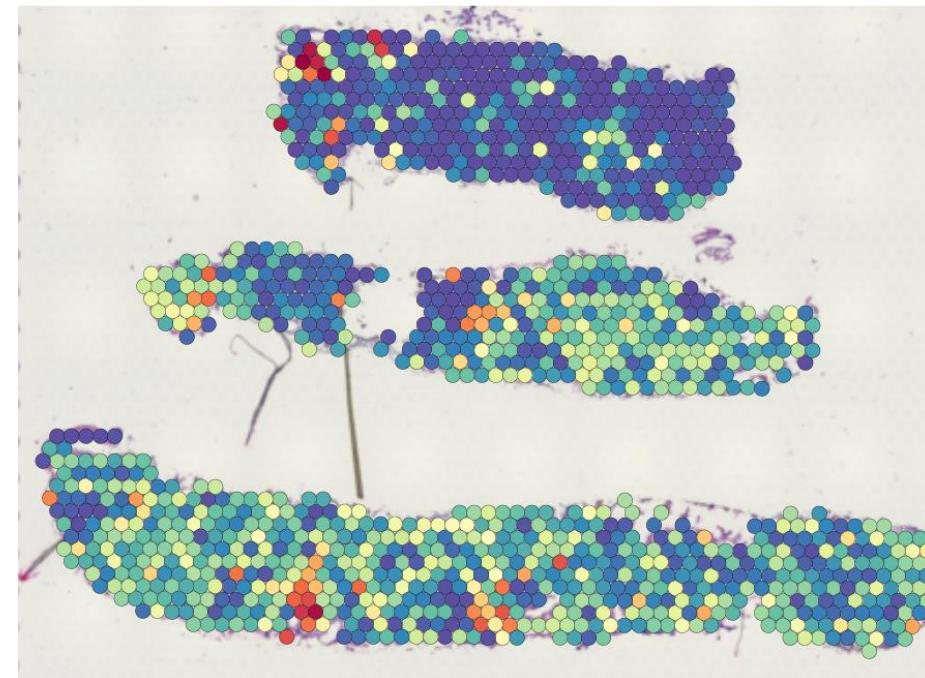
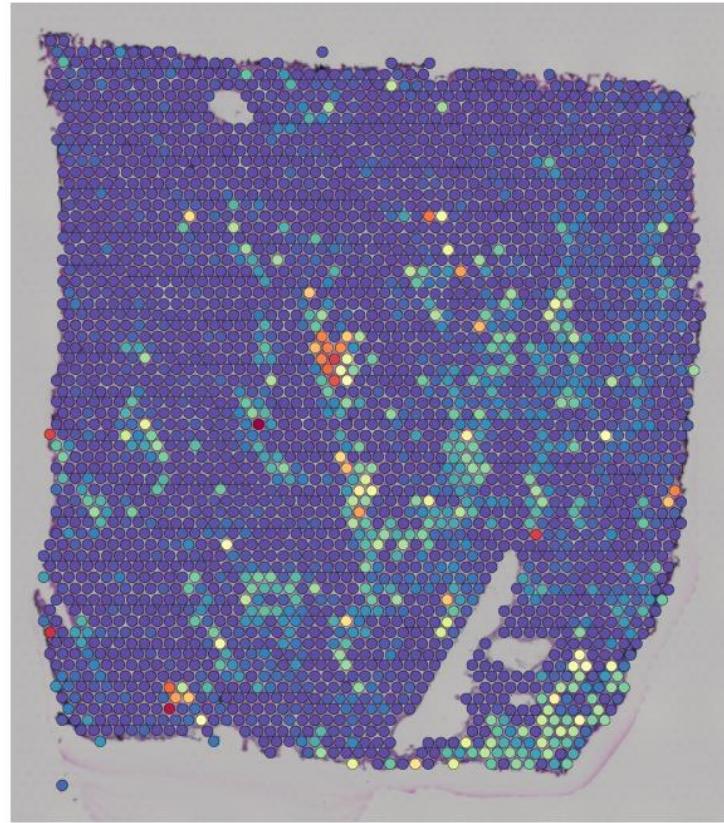
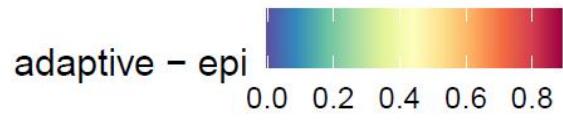
Fibrosis and injured tubules in the DKD sample



Adaptive Repair State

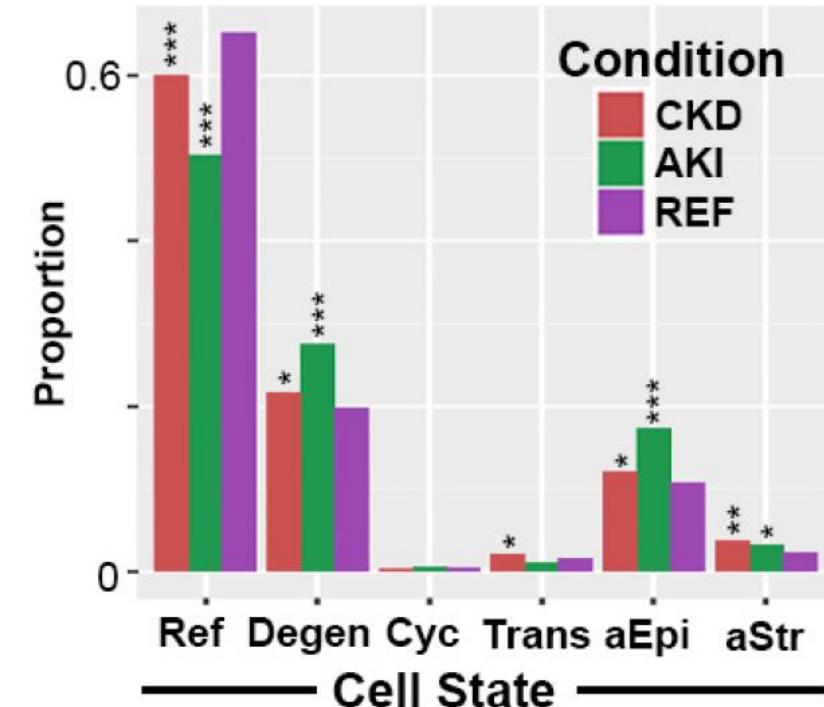


Cell State Mapping



Adaptive and degenerative cell states are more common in kidney disease samples (N = 23).

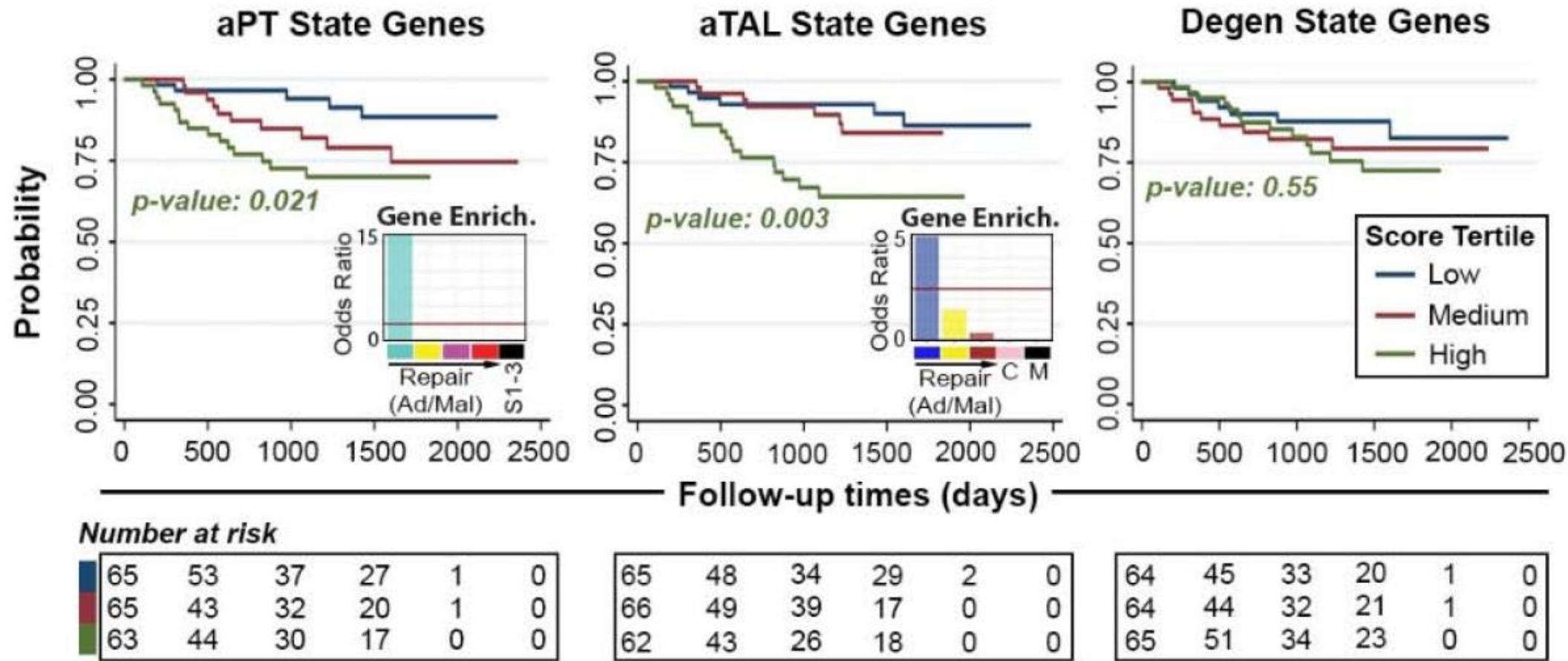
Visium Altered States



Outcomes

NEPTUNE Cohort Survival Curves

(Composite ESRD or 40% Decline in eGFR)

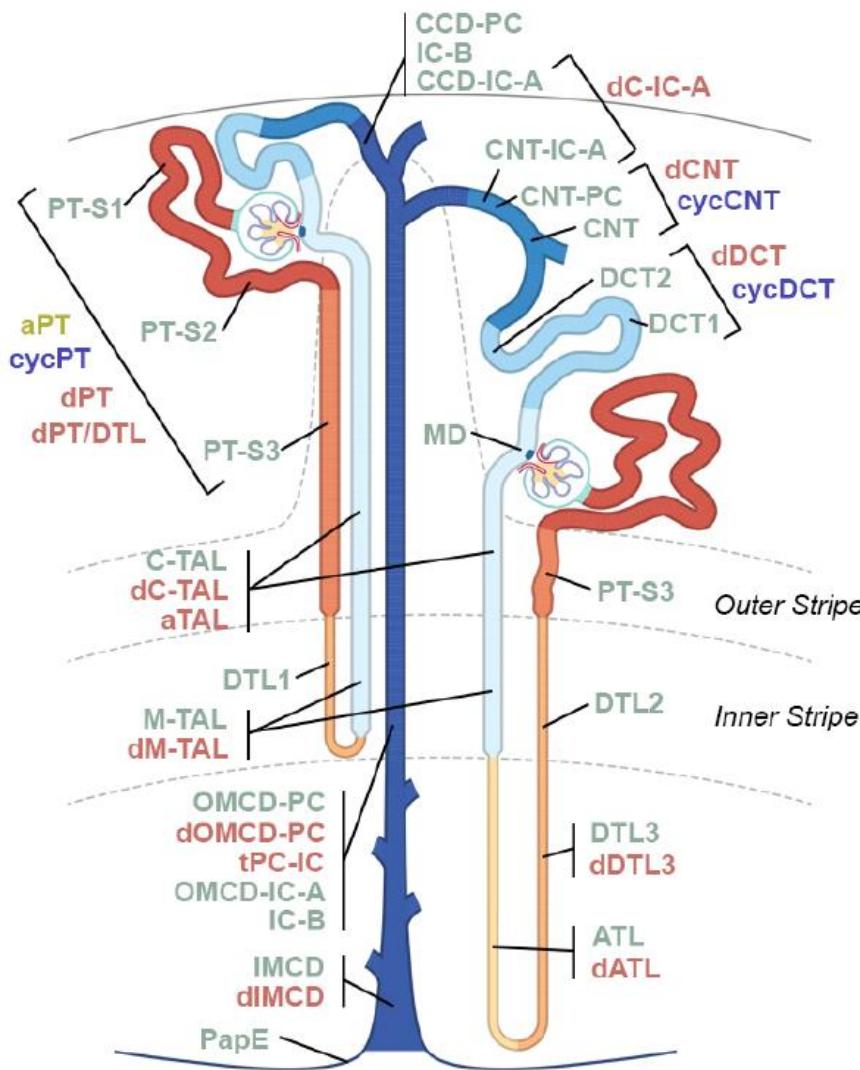


- In bulk kidney mRNASeq data of the NEPTUNE cohort, adaptive injury states associate with kidney disease progression.

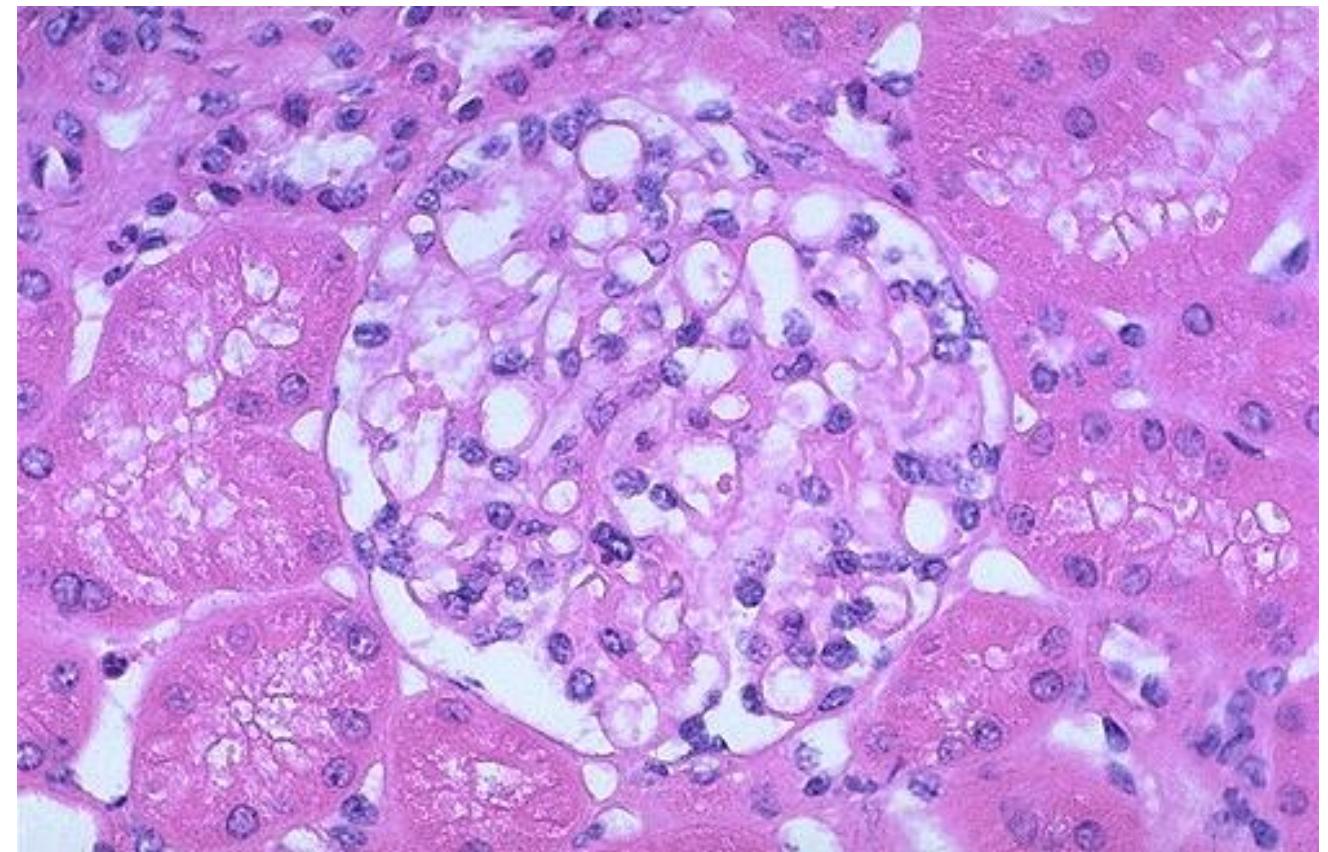
Neighborhoods

The Kidney

Nephron

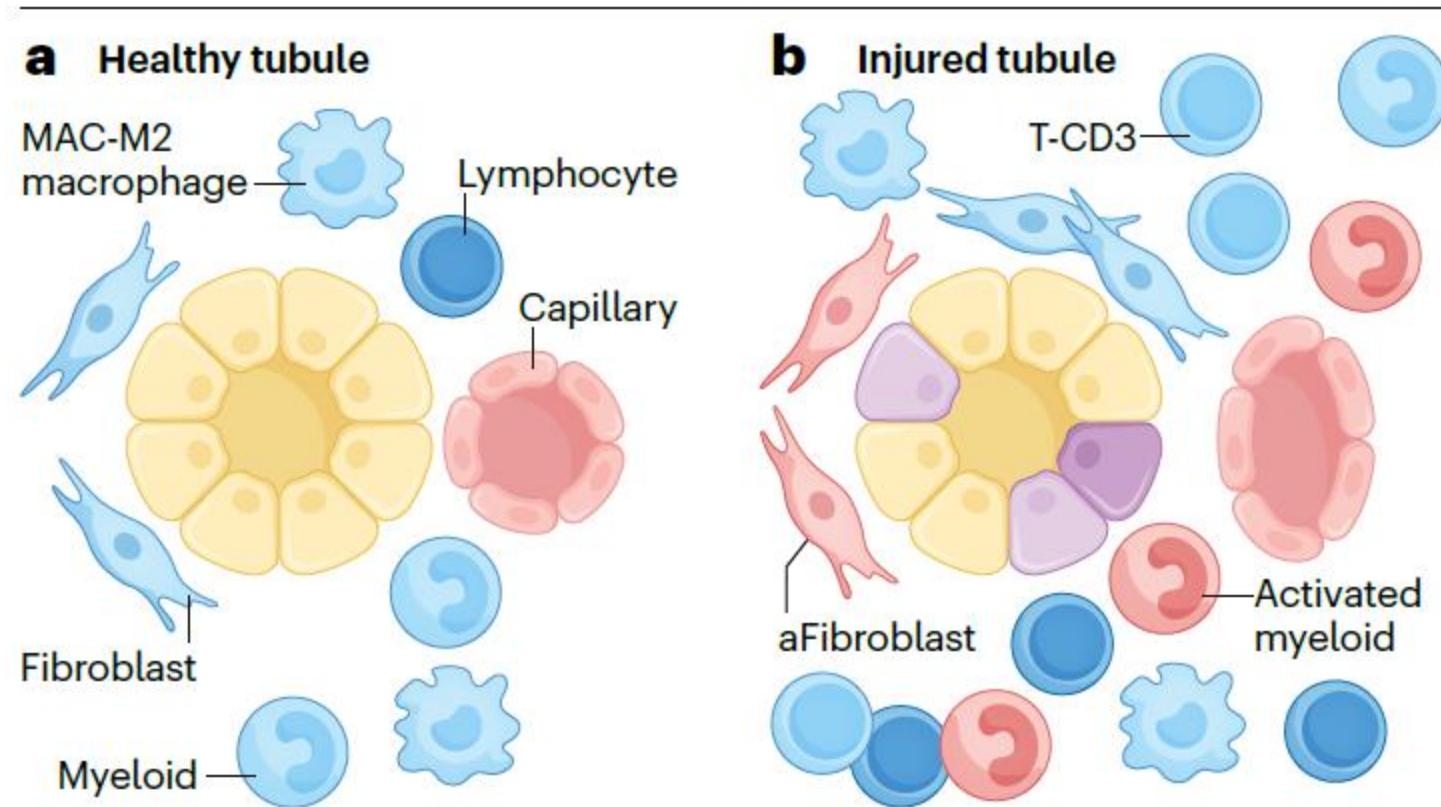


- Interface between blood and urine
- 1.25 million nephrons
- Mosaic like structure challenging for ST
- Over 100 cell types

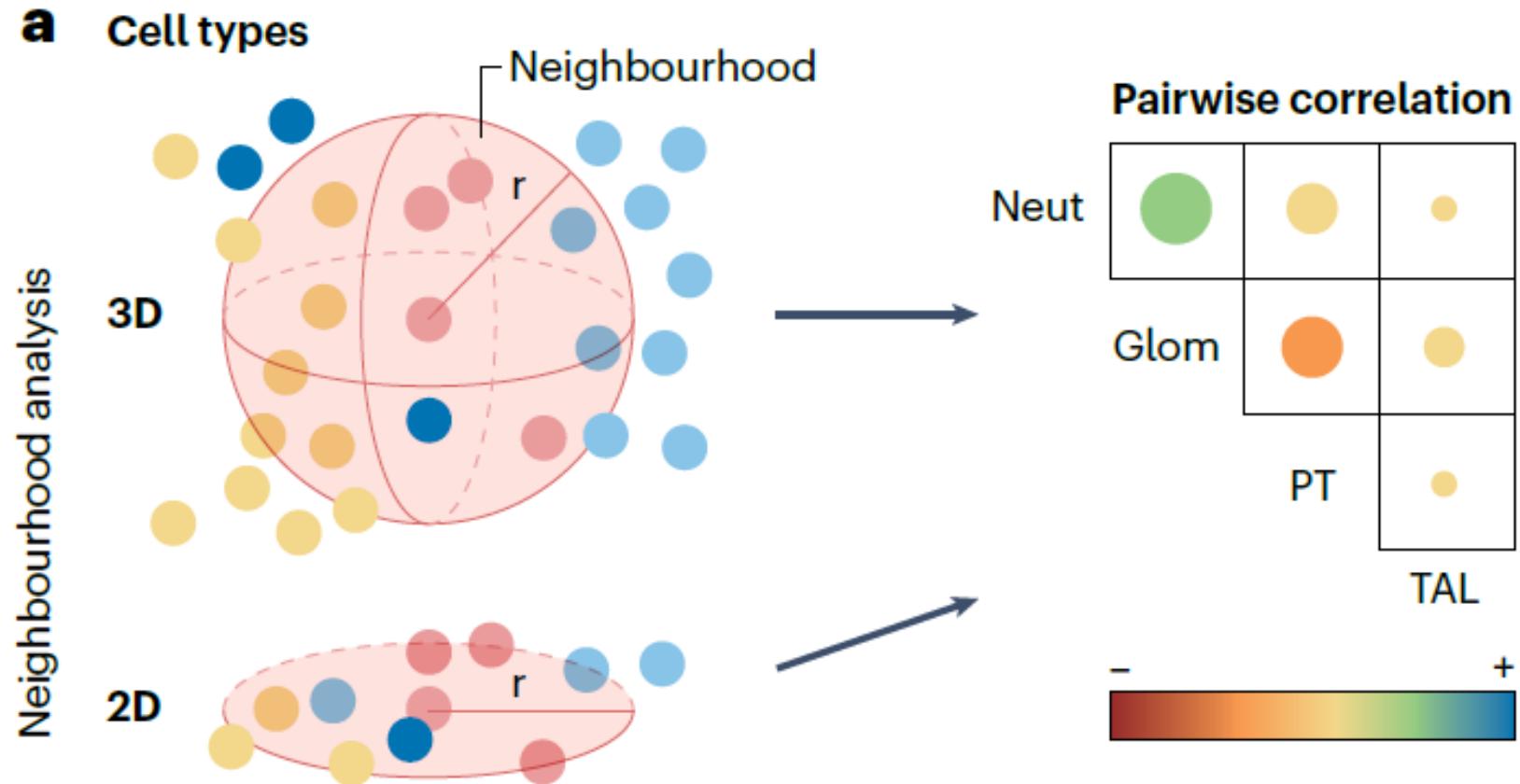


<https://webpath.med.utah.edu/RENAHTML/RENAL101.html>

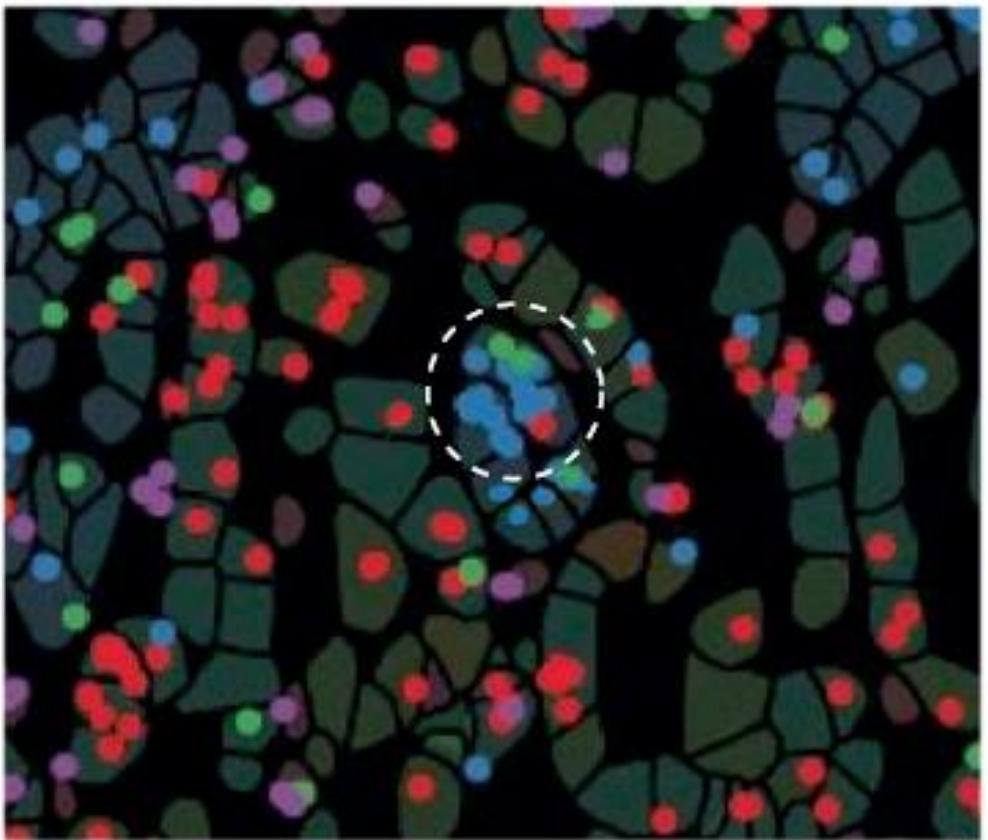
Neighborhood example – disease state



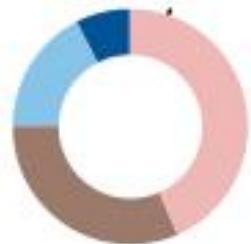
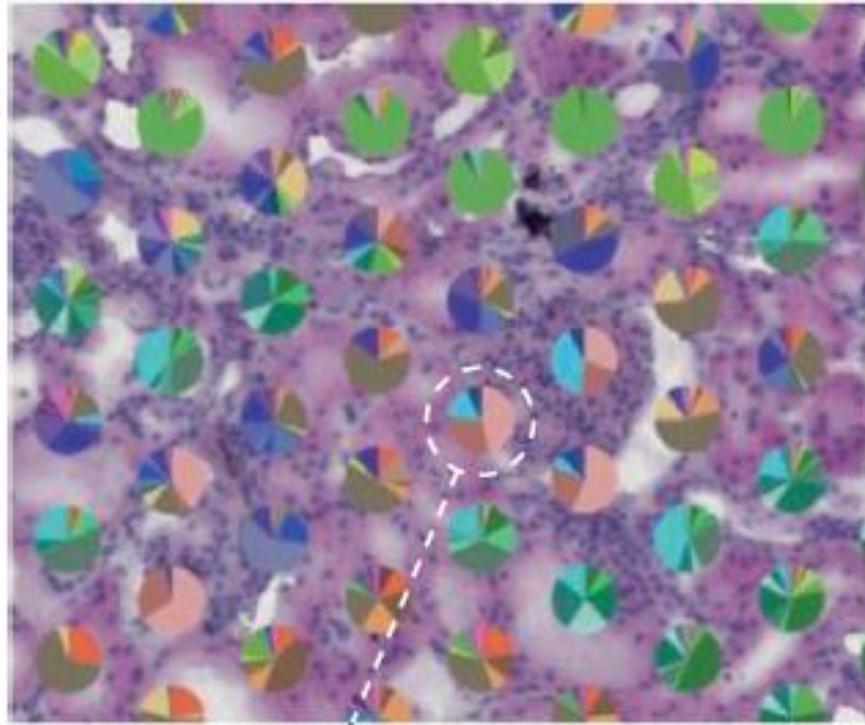
Neighborhood Analysis



Neighborhood Analysis



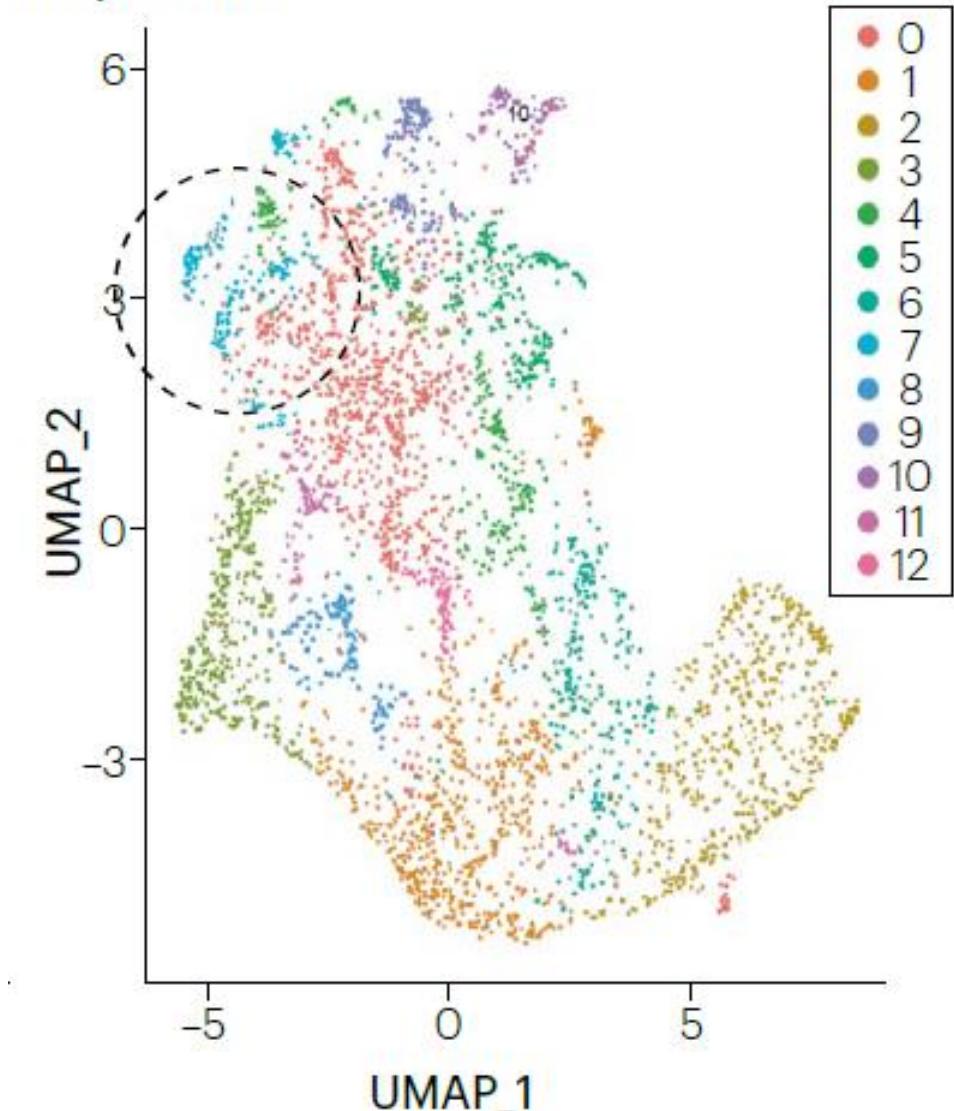
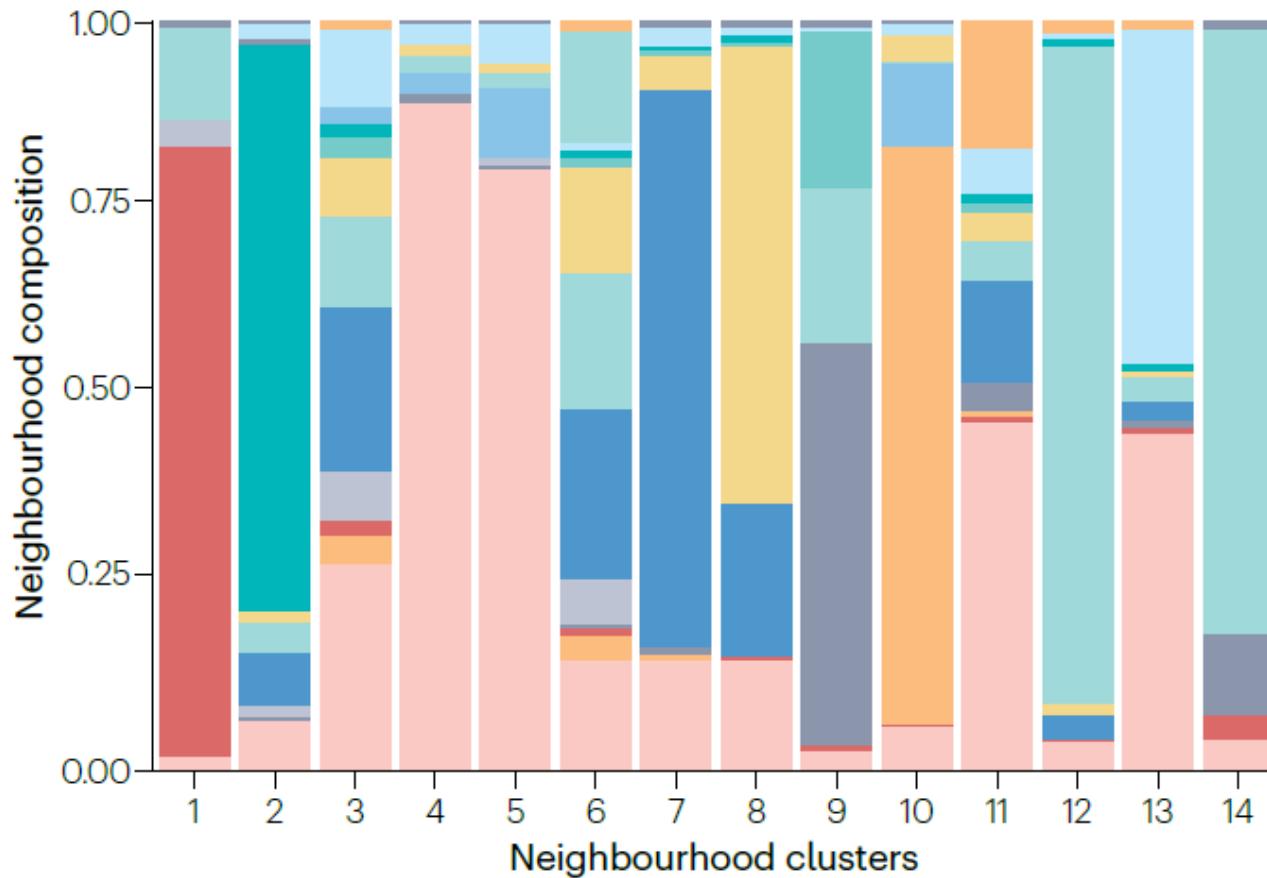
A neighbourhood defined
by a radius around a centroid



Components
of the spot
become the
neighbourhood

Neighborhood Components + Clustering

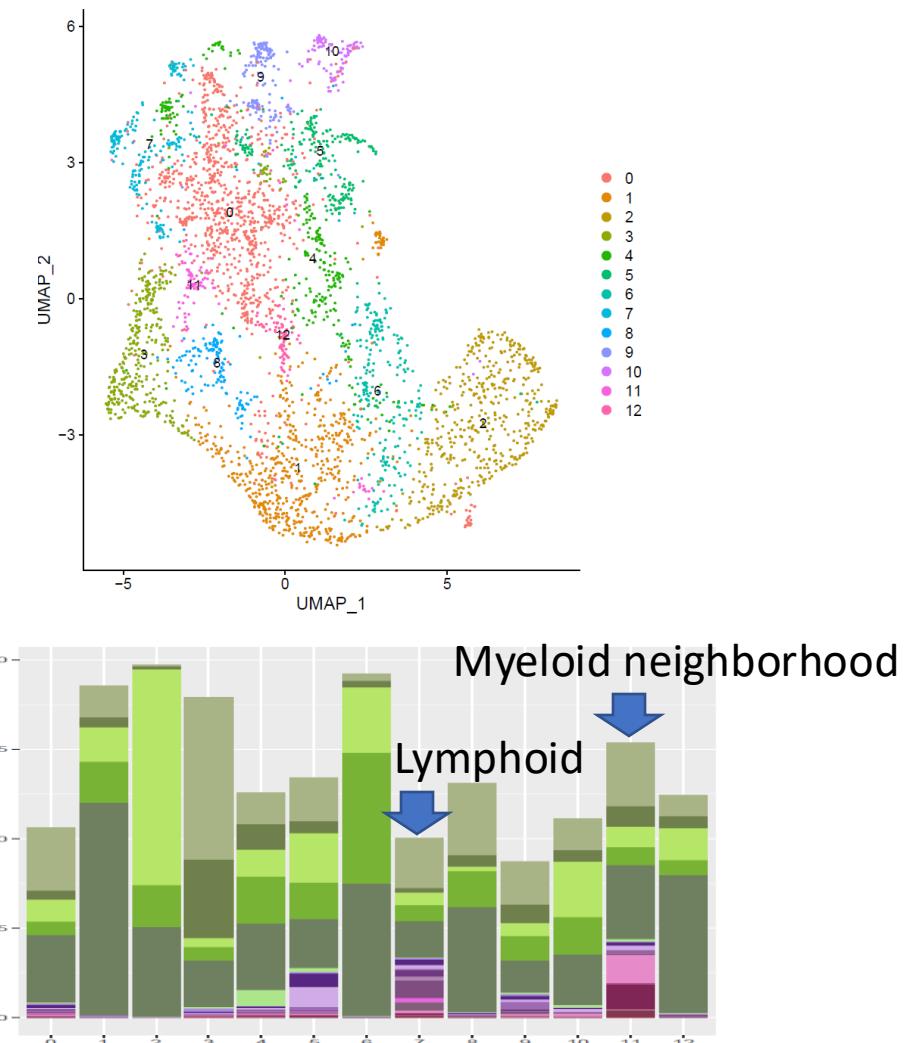
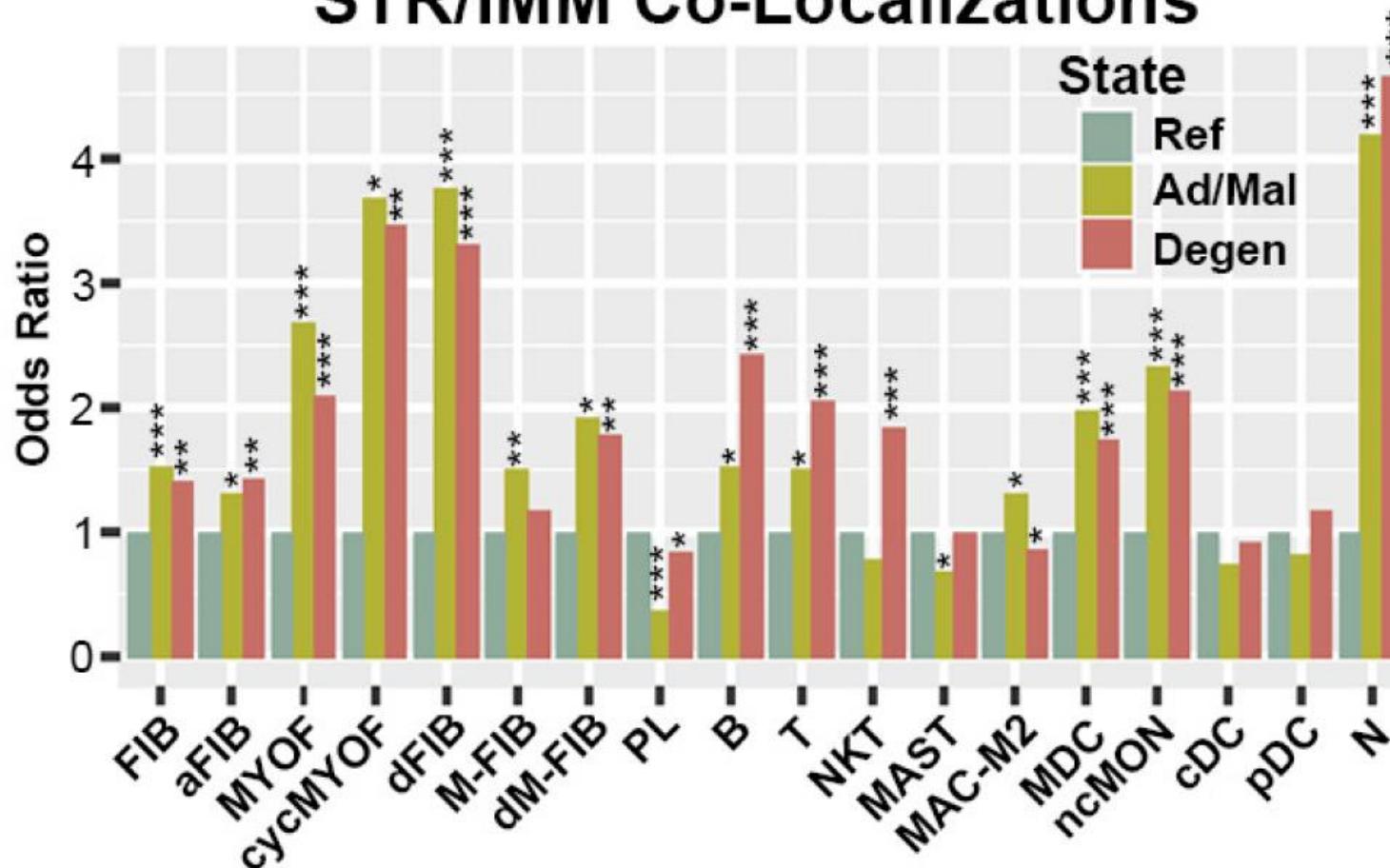
e Neighbourhood cluster composition



Epithelial cell co-localization with stromal and immune cells

C

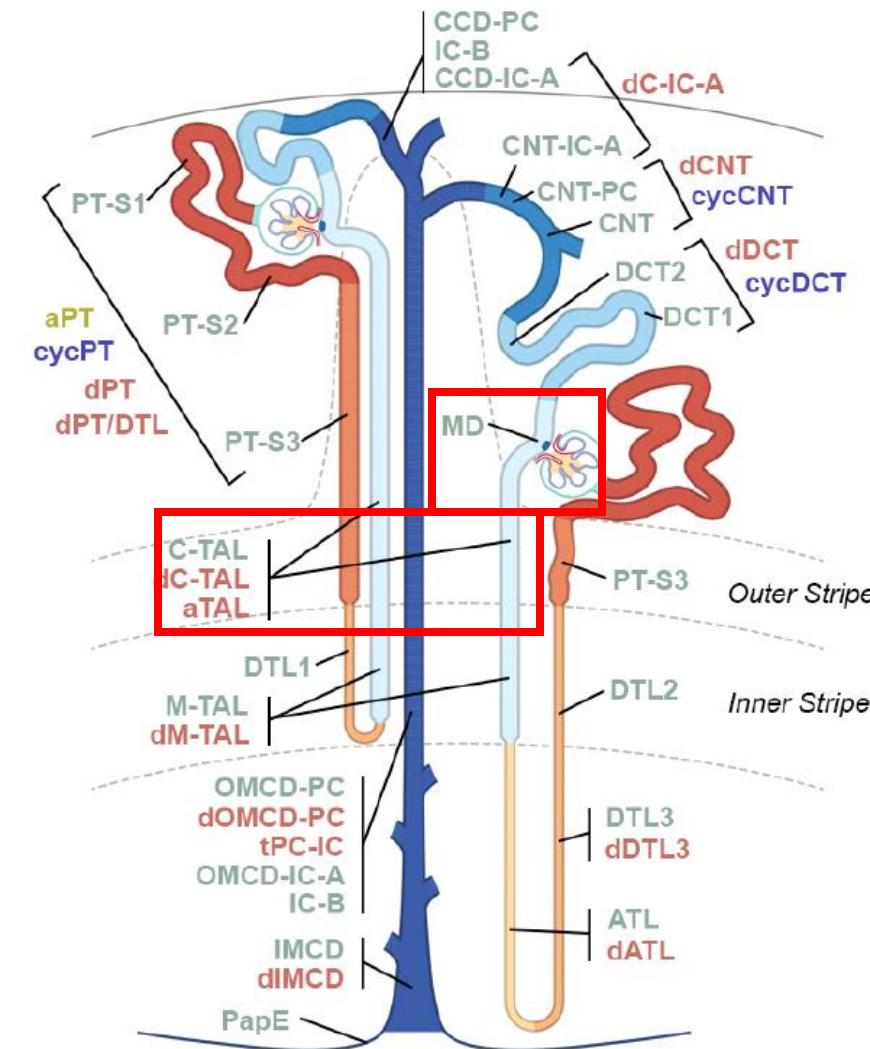
STR/IMM Co-Localizations



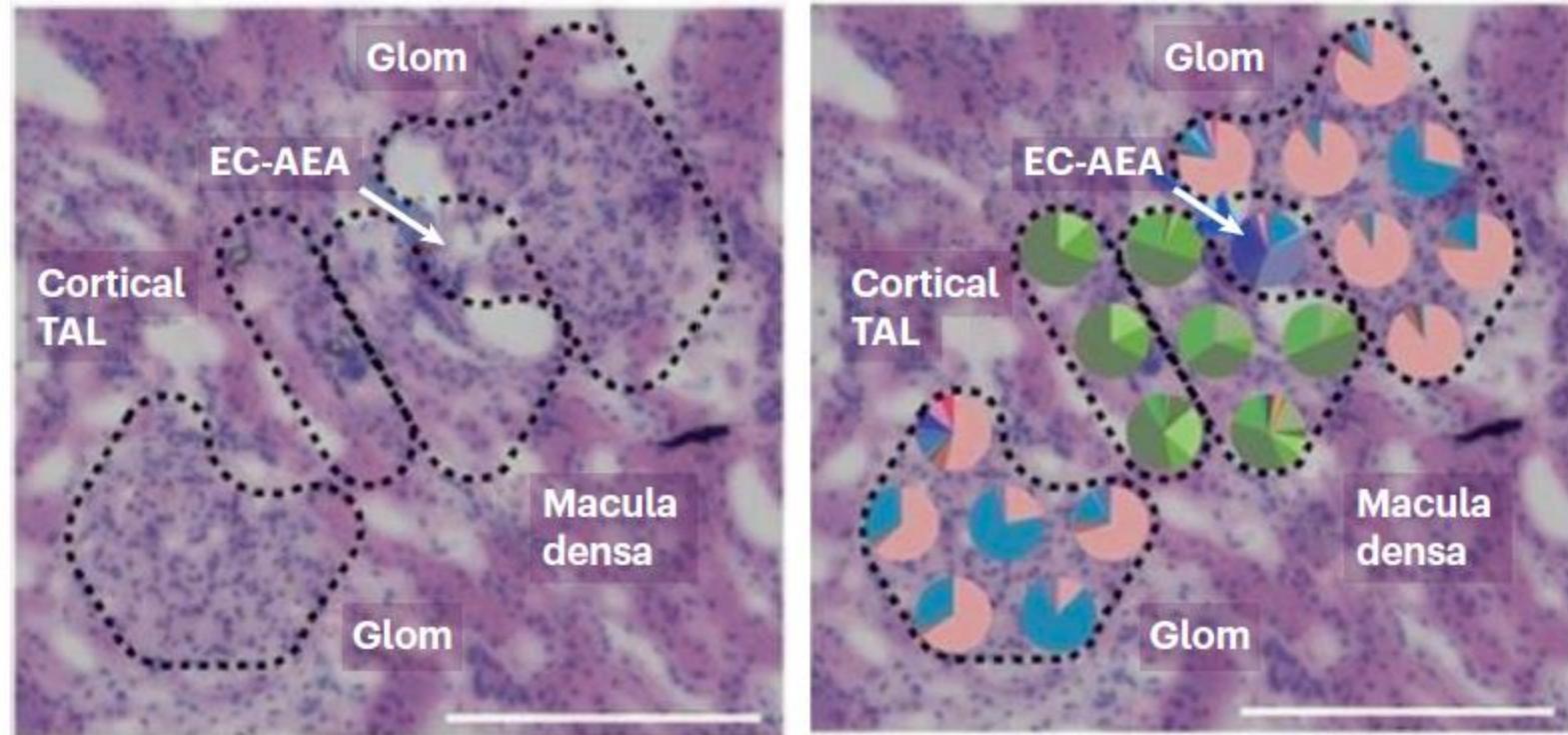
- ST reveals altered cell state signature colocalizes with immune cell infiltration. Each spot = a neighborhood.

Neighborhood example 2: Anatomic

Nephron



C



Spatial Variable Gene Analysis

Spatial Variable Gene Analysis

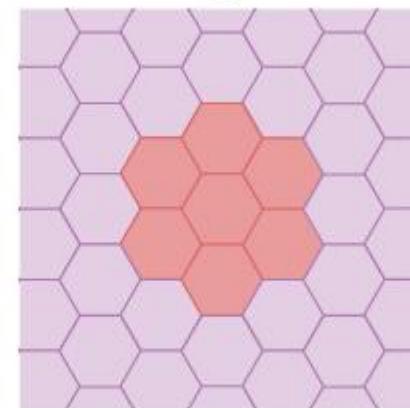
BSP: Useful for cross-cutting injury features like hypoxia

b Determination of spatially variable genes

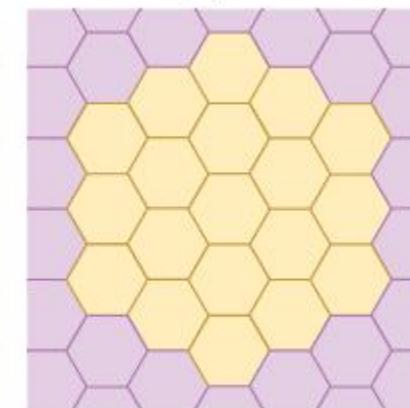


Focal region of injury with spatially variable expression determined by two sets of radii

Small patch



Big patch

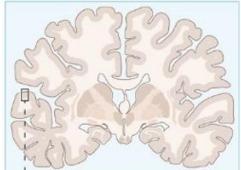


Juexin Wang

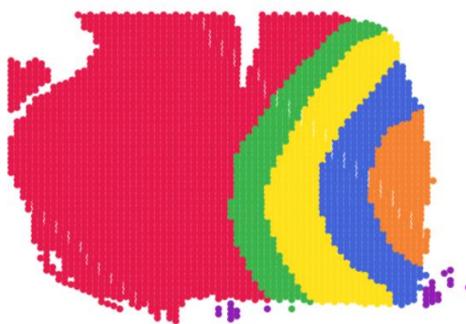
Relation Equivariant Graph Neural Networks

Brain

A



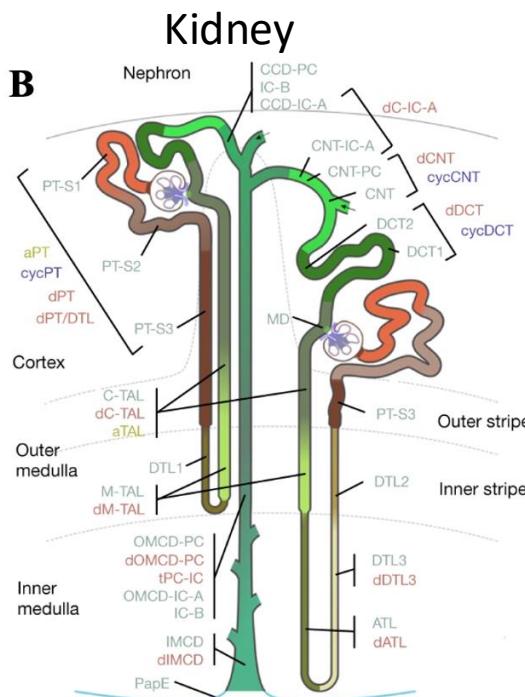
- I Molecular layer
- II External granular layer
- III External pyramidal layer
- IV Internal granular layer
- V Internal pyramidal layer
- VI Multiform layer



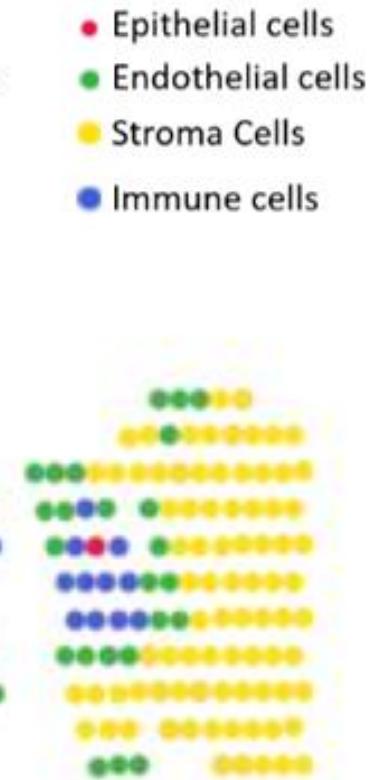
Large homogenous structures



Mosaic, stochastic structures

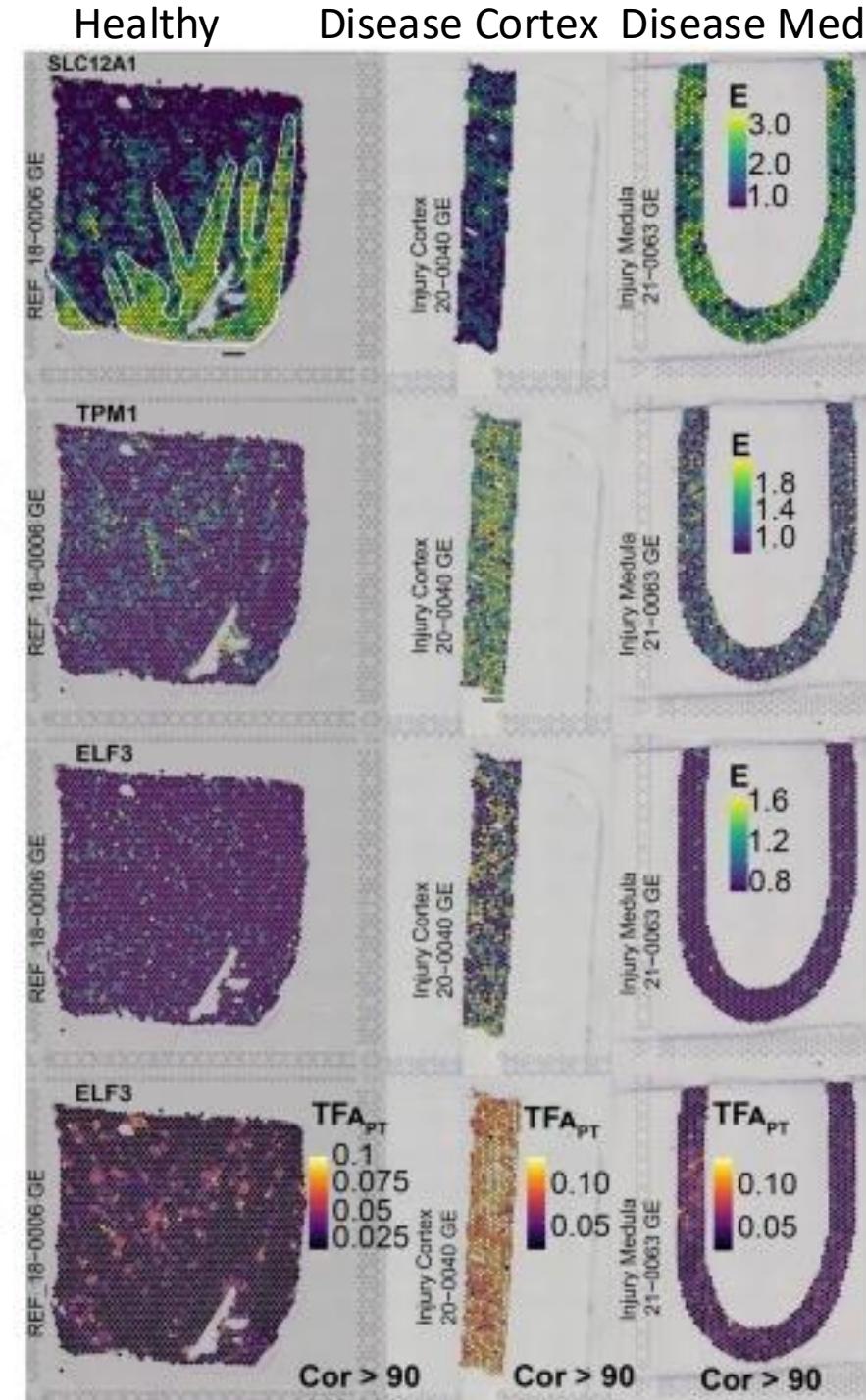
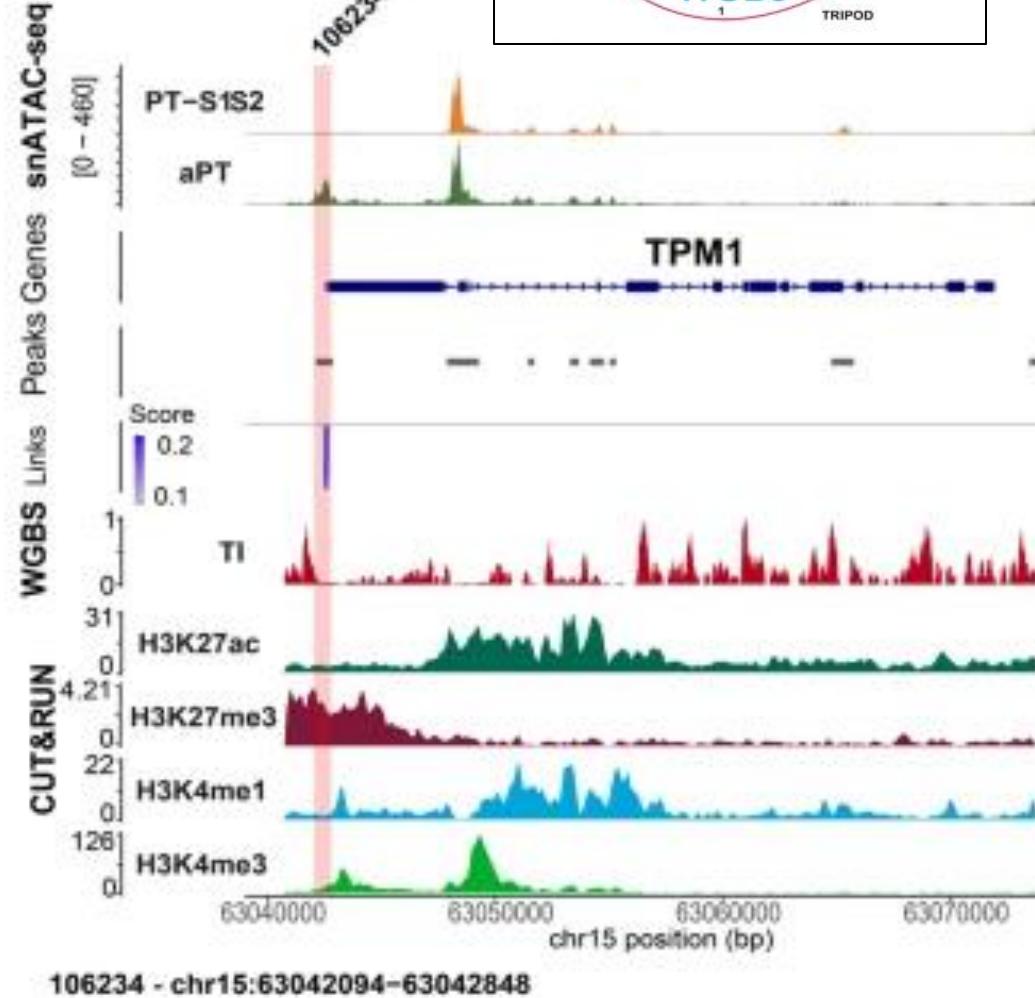


REGNN
ARI = 0.58

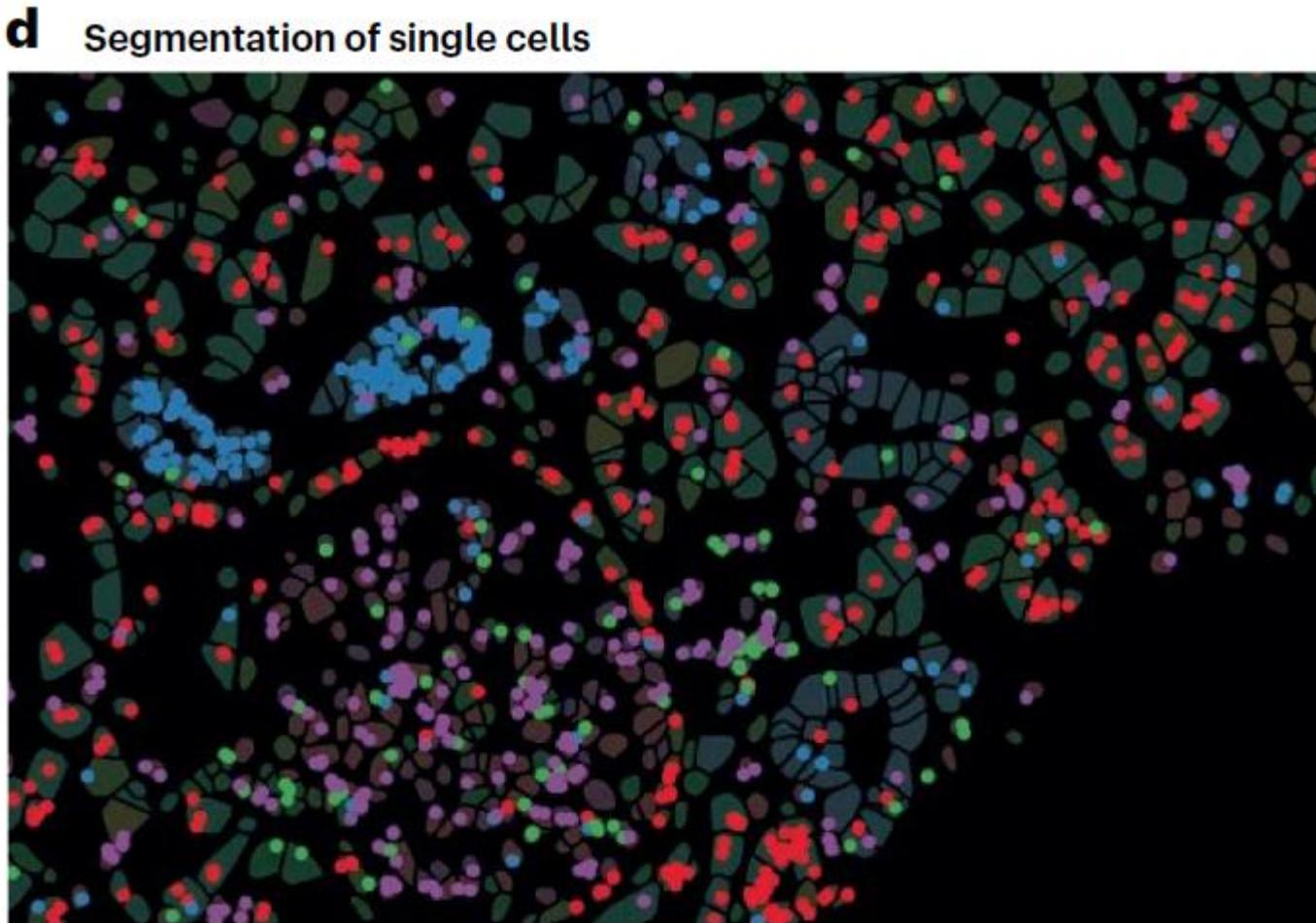


TF Network Mapping

Gene: TPM1 TF: ELF3



Segmentation



Tips:

- Cellpose is a leading approach
- Proprietary segmentation available in 10x platforms
- Use of cell membrane and cytoplasm markers improve accuracy.

Histologic Annotation with AI

Histologic Annotation using AI

- <https://github.com/spborder/FUSION/blob/main/README.md>
- <http://fusion.hubmapconsortium.org/>

FUSION
FUNCTIONAL UNIT STATE IDENTIFICATION AND NAVIGATION WITH WSI

User Survey Cell Cards Lab Website User Feedback

MENU REGISTERED USER LOGIN SIGN UP FOR USABILITY STUDY

Welcome: fusionguest

Welcome to FUSION

Click the Menu button above to see other pages



Or, Select a topic below to view more information

Getting Started: Select a category below to view tutorial slides

FUSION INTRODUCTION

- FUSION Introduction
- Preprocessing Steps

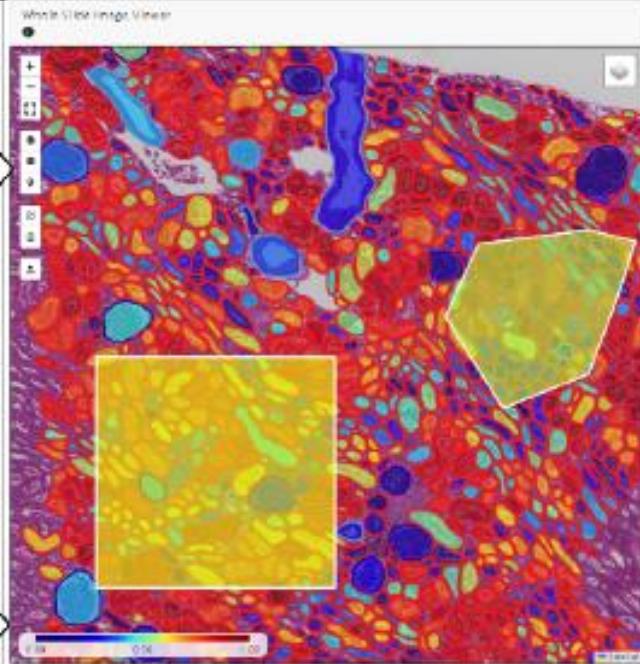
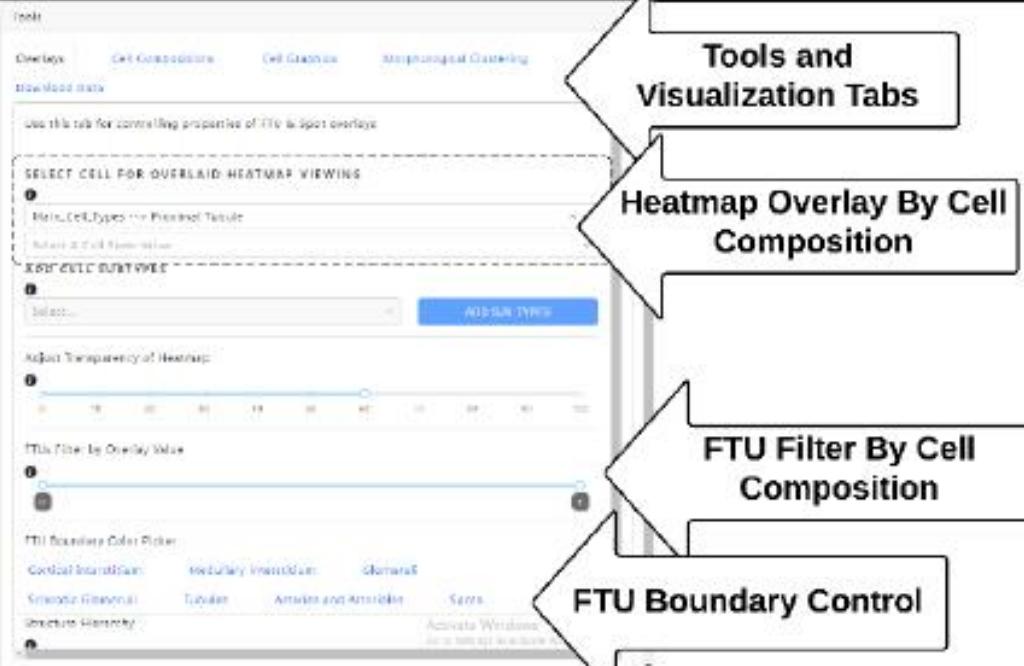


PH.D. STUDENT, BIOMEDICAL ENGINEERING
Samuel Border
Research: Computational Pathology

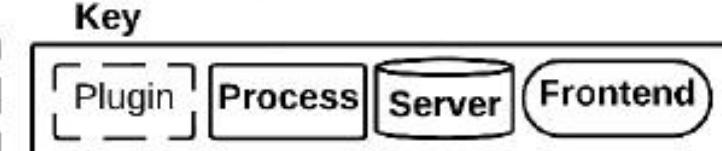
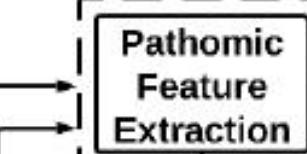
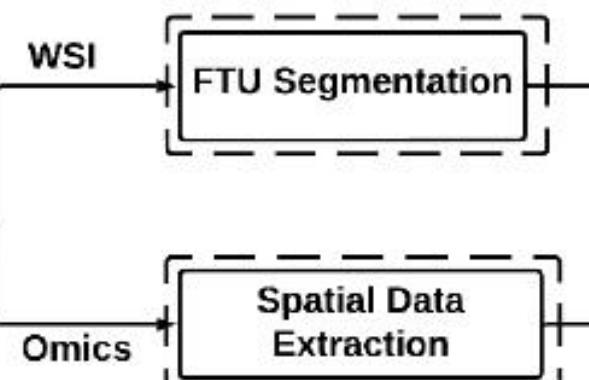


FUSION

FUNCTIONAL UNIT STATE IDENTIFICATION AND NAVIGATION WITH WSI

[User Survey](#)[Cell Cards](#)[Lab Website](#)[User Feedback](#)**A****Manual Annotation Control****Heatmap Overlays Colorbar****Tools and Visualization Tabs****Heatmap Overlay By Cell Composition****FTU Filter By Cell Composition****FTU Boundary Control****B****Inputs:**

- Whole Slide Image (WSI)
- Omics data

**User Annotations**

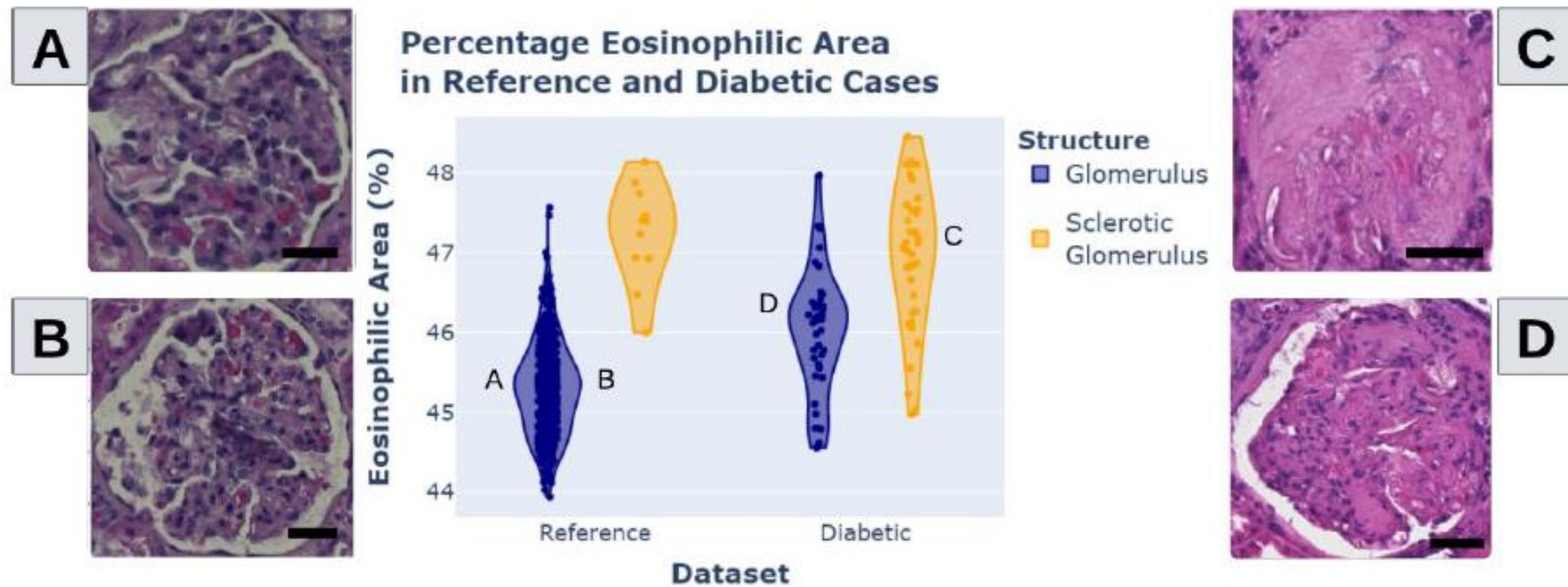
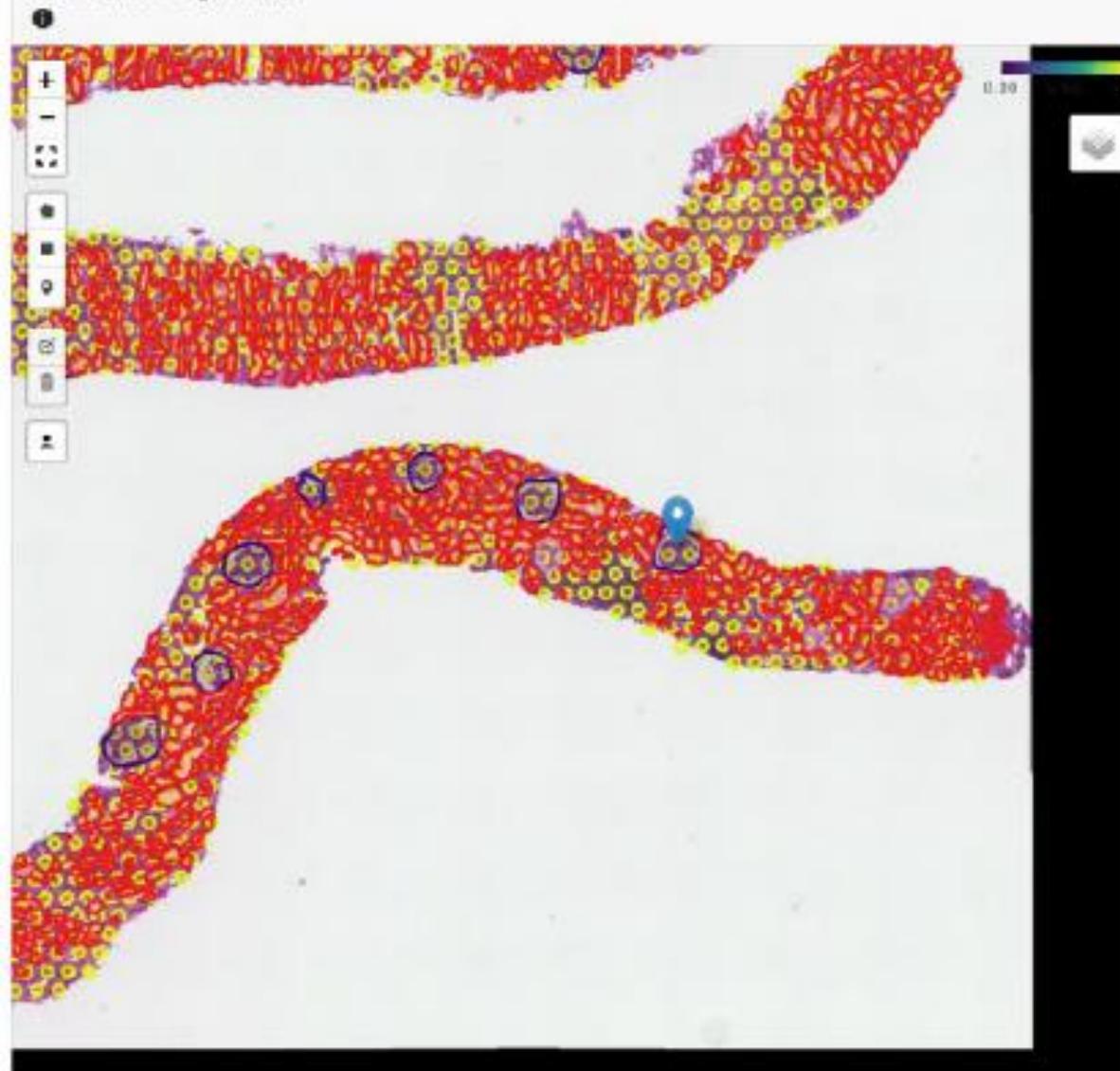
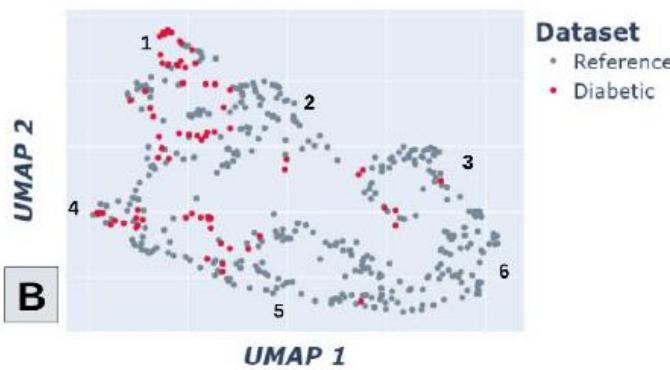


Fig. 3. Local comparison of digital histology images using a single feature. As an example, glomerular hypertrophy is compared in diabetes and reference. Percentage eosinophilic area between normal and sclerotic glomeruli from diabetic and reference kidney tissues. (A-D) Example glomeruli selected from the feature distribution where indicated. Dots represent the individual glomeruli. Sclerosis increases eosinophilic area by reducing open capillary lumens and Bowman's space. Diabetic kidney disease, in the absence of sclerosis, increases the eosinophilic area through hypertrophy and the expansion of the mesangium. Scalebar indicates 25 μ m.

Whole Slide Image Viewer



UMAP of Glomeruli
Reference and Diabetic



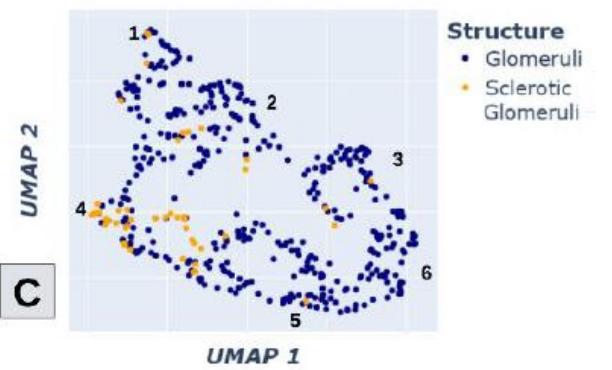
B

UMAP 1

UMAP 2

Dataset
• Reference
• Diabetic

UMAP of Glomeruli
Reference and Diabetic



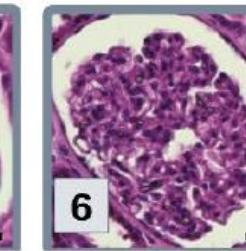
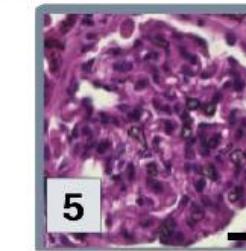
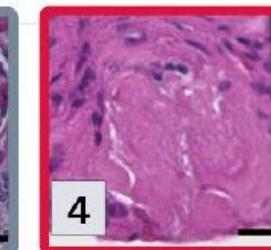
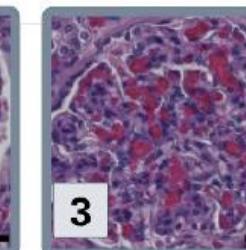
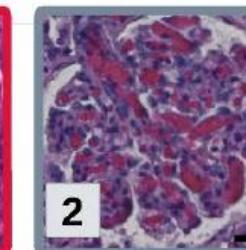
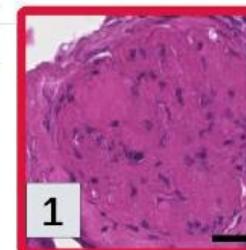
C

UMAP 1

UMAP 2

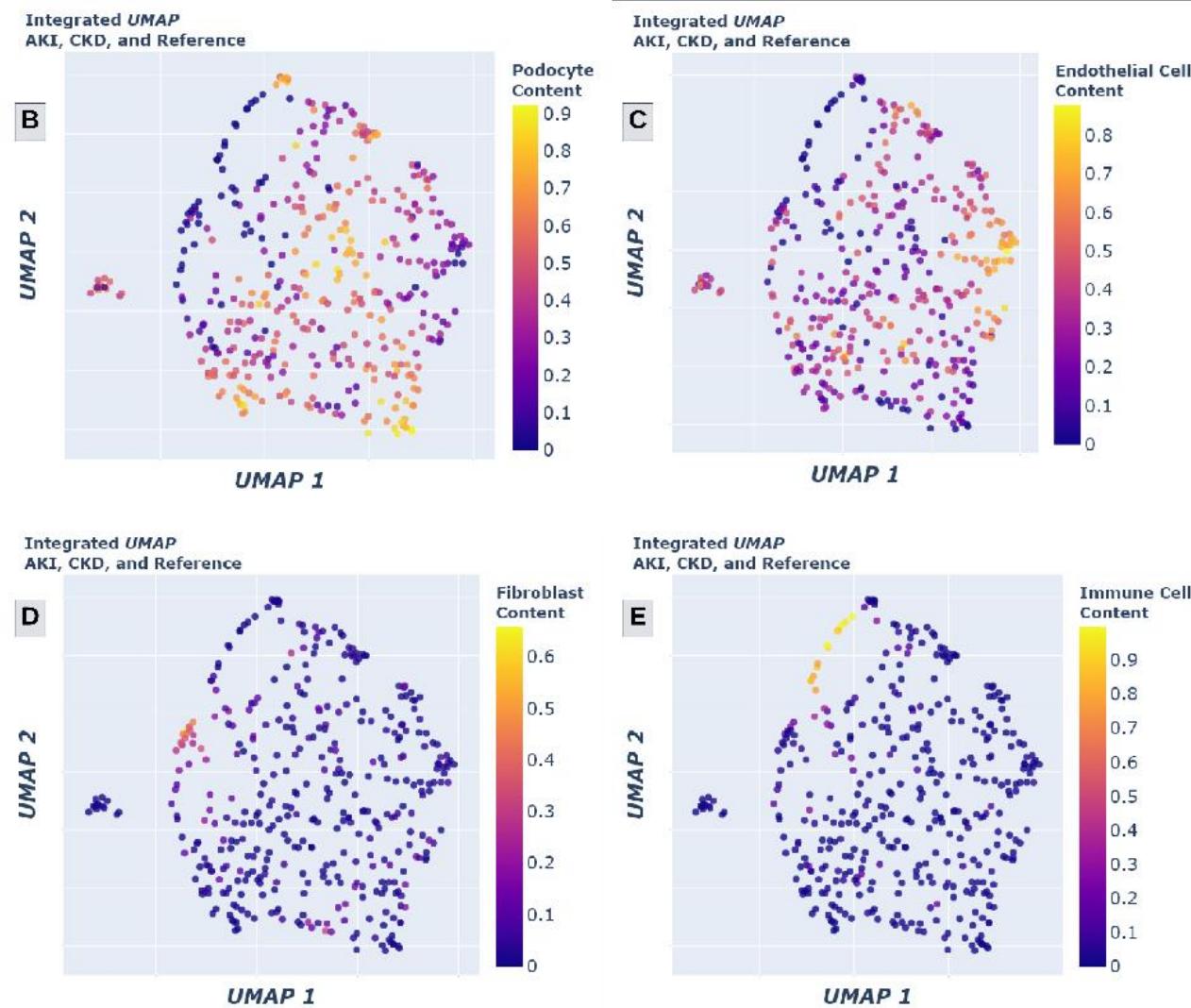
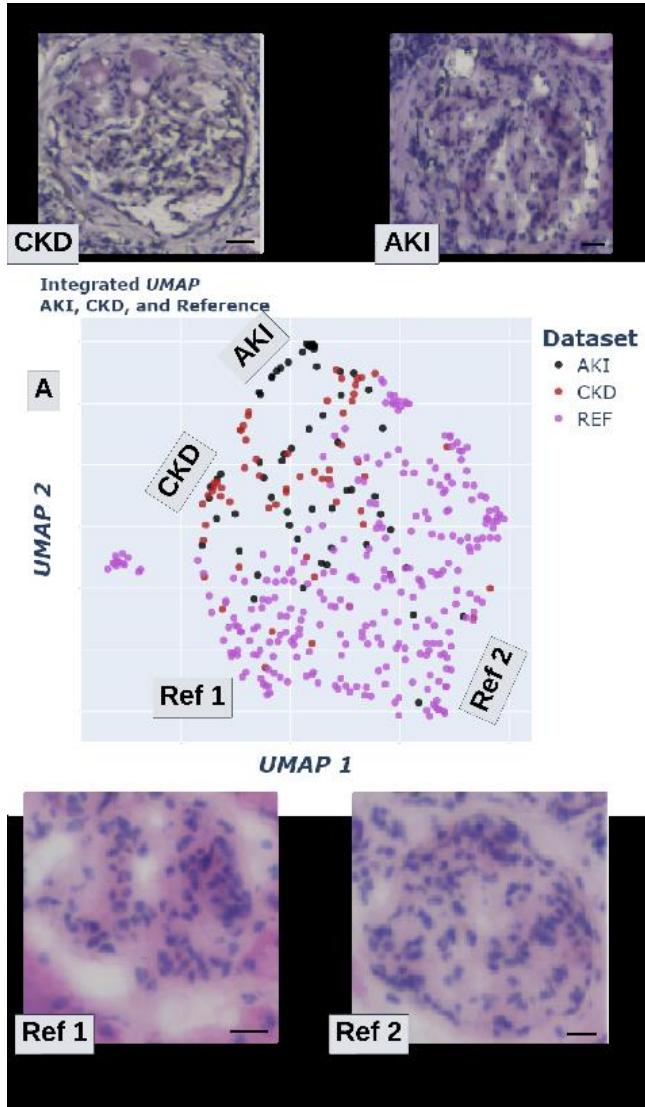
Structure
• Glomeruli
• Sclerotic
Glomeruli

D



- Clustering based on multiple pixel level features

Co-cluster Integrated Molecular and Pathologic features



- Integrates:
- Pixel features
- Molecular cell type identities

Acknowledgments

- Juixin Wang
- Qin Ma
- Ricardo Melo Ferreira
- Mauminah Raina
- Yi Jiang
- Mahla Asghari
- Pierre Dagher
- Tarek Ashkar
- Daria Barwinska
- Debora Gisch
- Yinghua Cheng
- Ken Dunn
- Joshua Vaughn
- Sanjay Jain



**Indiana O'Brien Center for
Advanced Renal Microscopy
and Molecular Imaging**

Funding:

R01 DK138504
U54 DK137328 (O'Brien)
NIDDK U01 DK114923 (KPMP)
NIH OT2OD033753 (HuBMAP HIVE)
NIDDK U54 DK134301 (HuBMAP TMC)
NIDDK P50 DK133943 (PCEN)
IU Grand Challenge Precision Health Initiative