Science Translational Medicine
AAAS

Supplementary Materials for

Analysis of the human kidney transcriptome and plasma proteome identifies markers of proximal tubule maladaptation to injury

Yumeng Wen *et al.*

Corresponding author: Chirag R. Parikh, chirag.parikh@jhmi.edu

Sci. Transl. Med. **15**, eade7287 (2023)
DOI: 10.1126/scitranslmed.ade7287

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Human snRNA-seq data library preparation and preprocessing

We used the Cell Ranger 7.0 pipeline to align snRNA-seq FASTQ files to the human hg38 reference genome, after which we removed ambient RNA contamination, removed doublets, excluded low-quality nuclei, and removed transcripts mapped to mitochondrial RNA from analysis. We then combined all samples using Seurat V4. We performed data integration using reciprocal principal component analysis on 2,000 highly variable genes across each sample after log-normalization and scaling to correct batch effects. We then performed a principal component analysis in the integrated dataset and chose the 15 principal components determined by using the ElbowPlot function in Seurat. We further performed dimension reduction to a uniform manifold approximation and projection (UMAP) plot and performed Louvain clustering using a resolution of 0.5 after k-nearest neighbor embedding. Using an approach similar to that of a recent snRNA-seq study of human AKI autopsy kidney tissues *(29)*, each cluster of a major kidney cell type was further integrated and clustered. Subclusters that expressed canonical markers of more than 2 distinct cell types, which may represent doublets, were removed. We repeated this integration-clustering step iteratively until no subcluster of doublets could be identified. The canonical markers of major kidney cell types used were proximal tubule: *CUBN* and *SLC5A12*; the thin limb of the loop of Henle: *SLC44A5*; the thick ascending limb of the loop of Henle: *UMOD* and *SLC12A1*; distal convoluted tubule: *SLC12A3*; connecting tubule: *CALB1*; principal cell: *AQP2* and *SCNN1G*; intercalated A cell: *SLC4A1*; intercalated B cell: *SLC26A4*; podocyte: *NPHS2*; endothelium: *FLT1*; fibroblast: *ACTA2* and *COL1A1*; immune cell: *CD163, IL7R, NKG7, MS4A1, MZB1, HLA-DQA1*, and *MS4A2*.

Sample collection, processing, and proteomic measurements in study cohorts

For the TRIBE-AKI adult and pediatric cardiac surgery cohorts, blood samples were collected preoperatively and postoperatively (within 6 hours) after cardiac surgery. For the TRIBE-AKI adult cohort, a subset of 54 participants had urine collected pre- and postoperatively for urine proteomic profiling. For the marathon cohort, blood samples were collected 24 hours before (prerace) and within 30 minutes after the marathon (postrace). Blood and urine samples were centrifuged and stored at -80 °C until measurement. Plasma and urine samples were shipped to SomaLogic (Boulder, CO) for identification and quantification of low-abundance proteins by SomaScan®, which uses easily quantifiable, chemically-modified oligonucleotides as binding reagents for proteins and protein complexes *(58)*. Protein analyte measurements underwent the SomaScan® data standardization and normalization process and were matched to their corresponding genes.

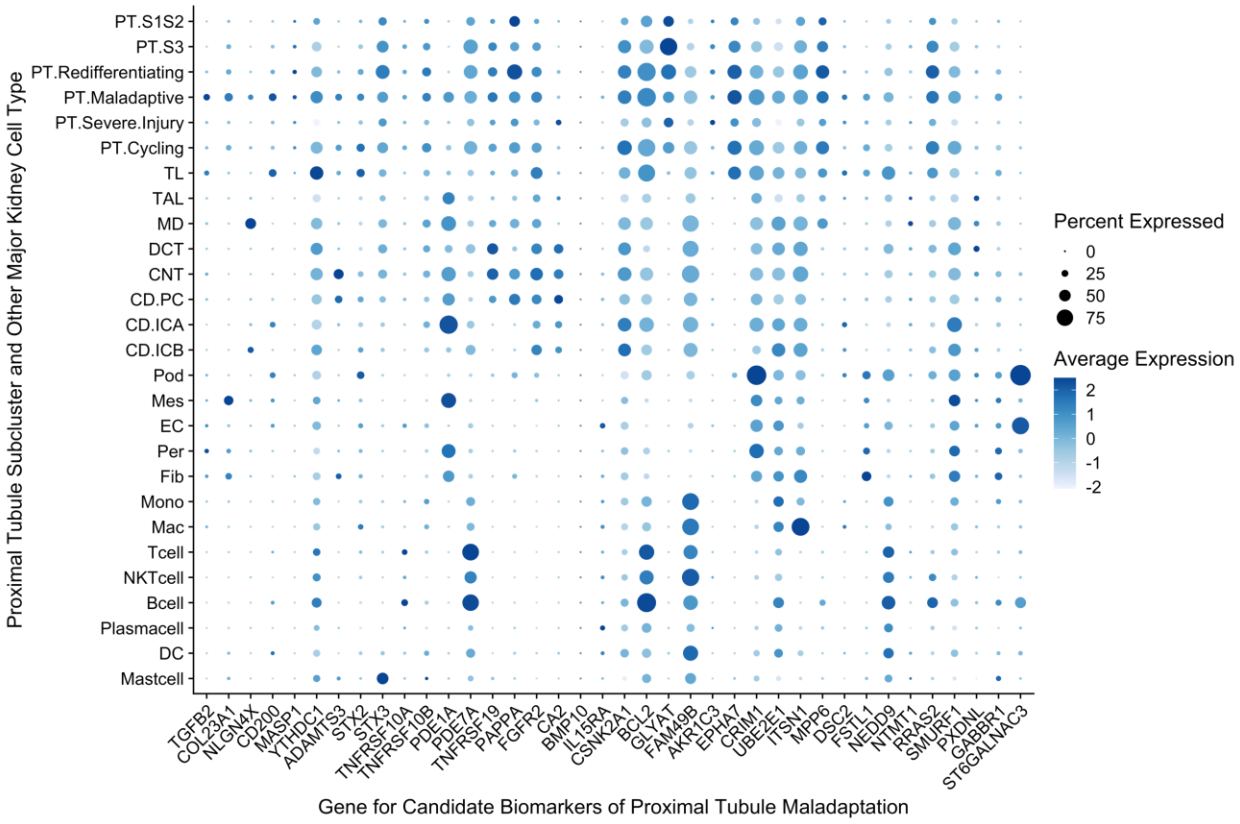

Fig S1. Tissue gene expression of 39 candidate markers of PT maladaptation identified from the integrated workflow in KPMP participants with AKI

Tissue gene expression of candidate markers of PT maladaptation in 17 participants with AKI from the KPMP cohort at single-nucleus resolution.

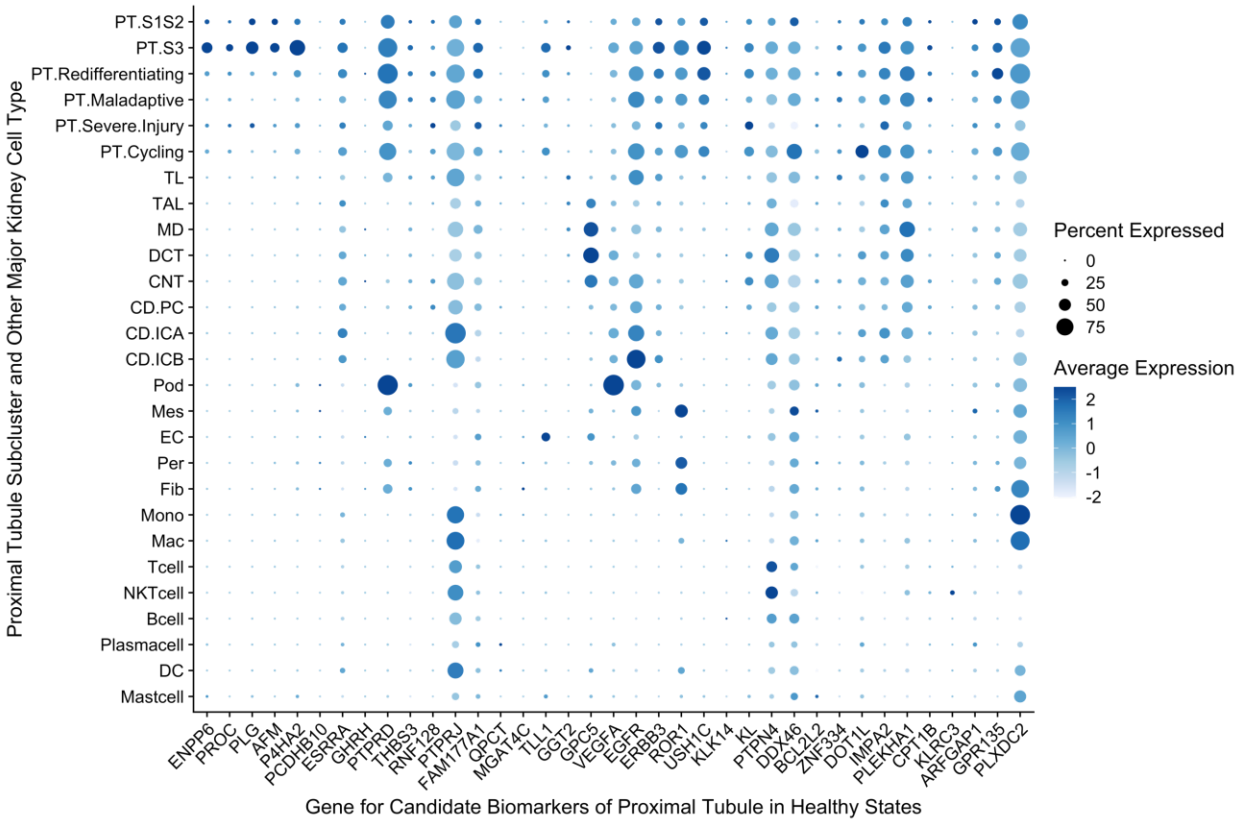

Fig S2. Tissue gene expression of 38 candidate markers of PT cells at healthy states identified from the integrated workflow in KPMP participants with AKI

Tissue gene expression of candidate markers of PT cells at healthy states in 17 participants with AKI from the KPMP cohort at single-nucleus resolution.

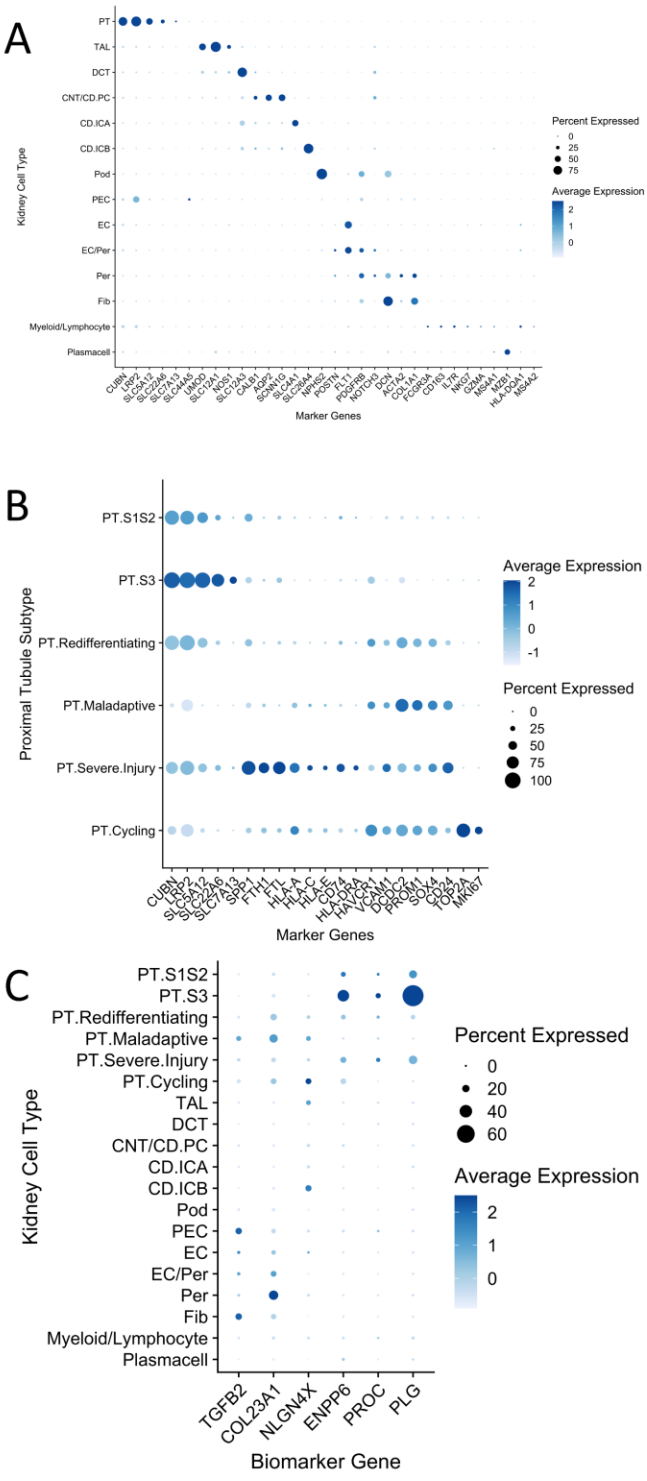

Fig S3. Tissue gene expression of markers of PT maladaptation and PT cells at healthy states in kidney biopsy tissues from 3 recently enrolled KPMP participants with AKI

(A) Marker gene expression of major kidney cell types in kidney biopsy tissues from 3 recently enrolled KPMP participants with AKI. **(B)** Marker gene expression of predicted PT subclusters using PT cells from 17 KPMP participants with AKI in the marker discovery phase. **(C)** Gene expression of markers of PT maladaptation and PT cells at healthy states in predicted PT subclusters and other major kidney cell types.

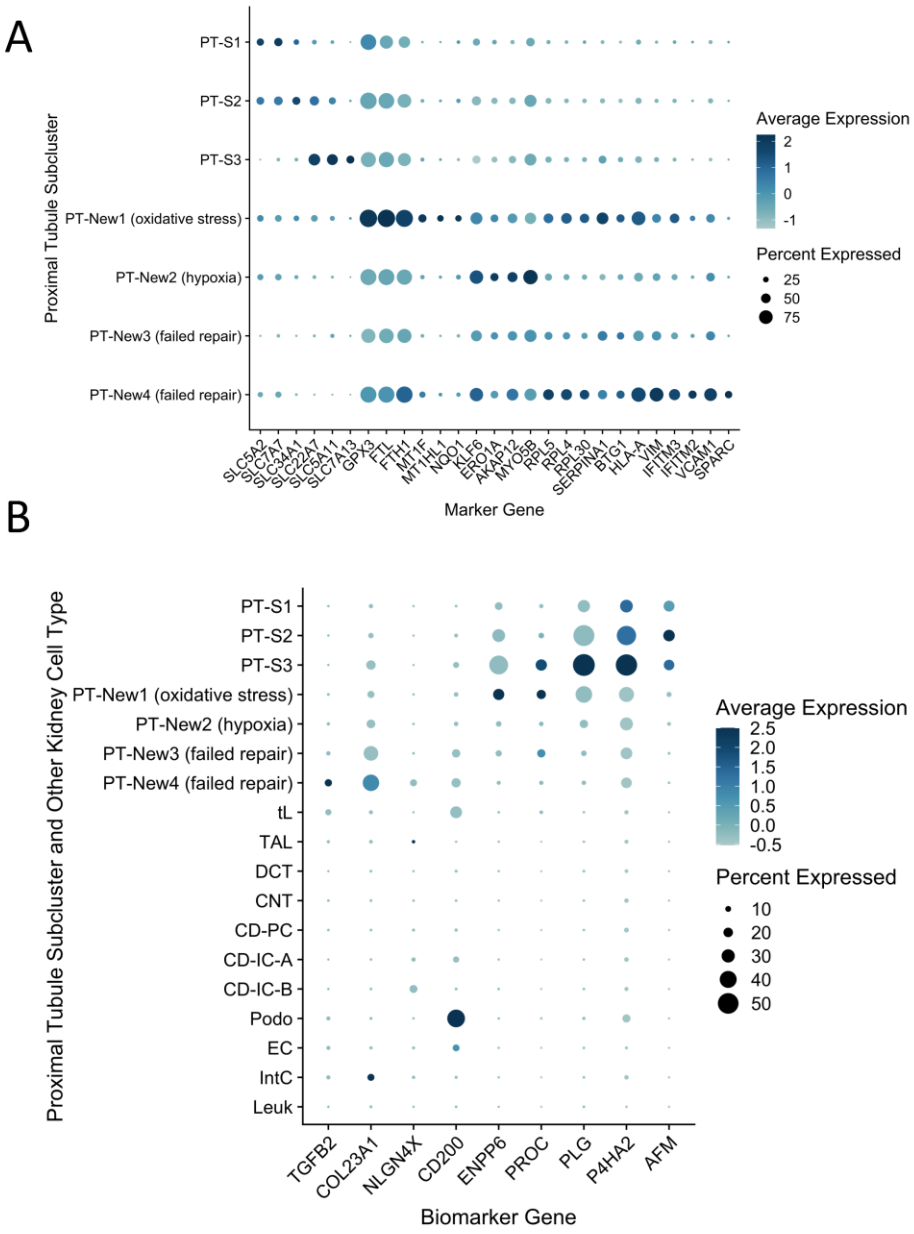

Fig S4. Tissue gene expression of markers of PT maladaptation and PT cells at healthy states in kidney autopsy tissues from an independent cohort of critically ill patients with AKI (A) Marker gene expression of PT subclusters classified using approaches described by Hinze et al *(33)*. **(B)** Gene expression of markers of PT maladaptation and PT cells at healthy states in PT subclusters and other kidney cell types.

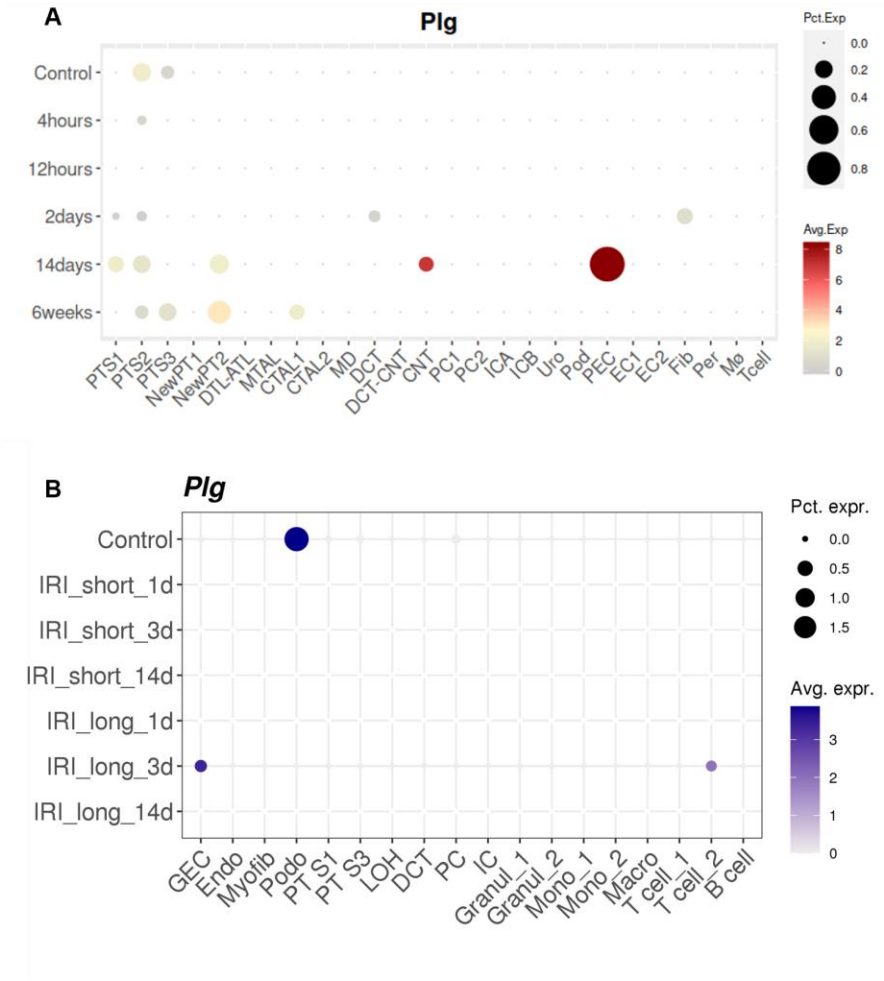

Fig S5. Low expression of Plg in publicly available single-cell and single-nucleus RNA sequencing datasets of mouse models of AKI (A) Low expression of *Plg* in snRNA-seq dataset of mouse models of AKI published by Kirita et al (*3*). **(B)** Low expression of *Plg* in single-cell RNA sequencing dataset of mouse models of AKI published by Balzer *et al* (*32*). Of note, the marker of PT maladaptation, *Nlgn4*, is not expressed in the mouse kidneys in these two datasets.

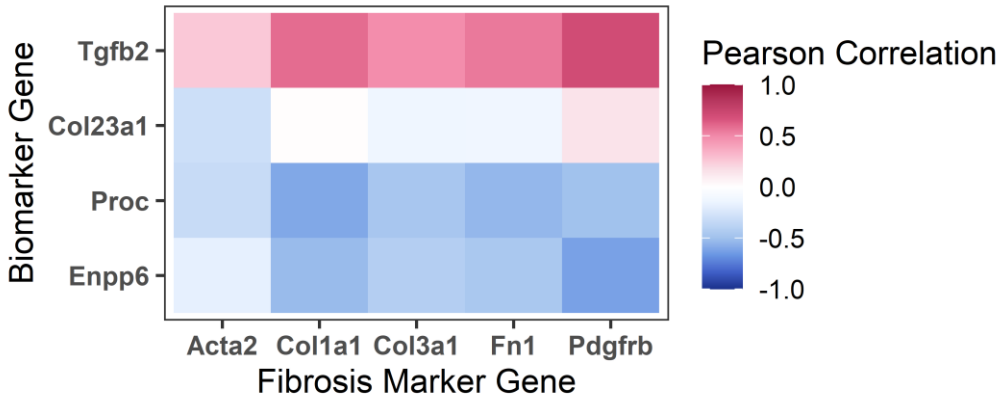

Fig S6. Correlation between gene expression of markers of PT maladaptation and PT cells at healthy states with fibrosis markers in the recovery phase of AKI from mouse models of IRI

In mouse models of IRI followed by repair or atrophy, the correlation in gene expression between markers of PT maladaptation, markers of PT cells at healthy states, and markers of fibrosis were determined.

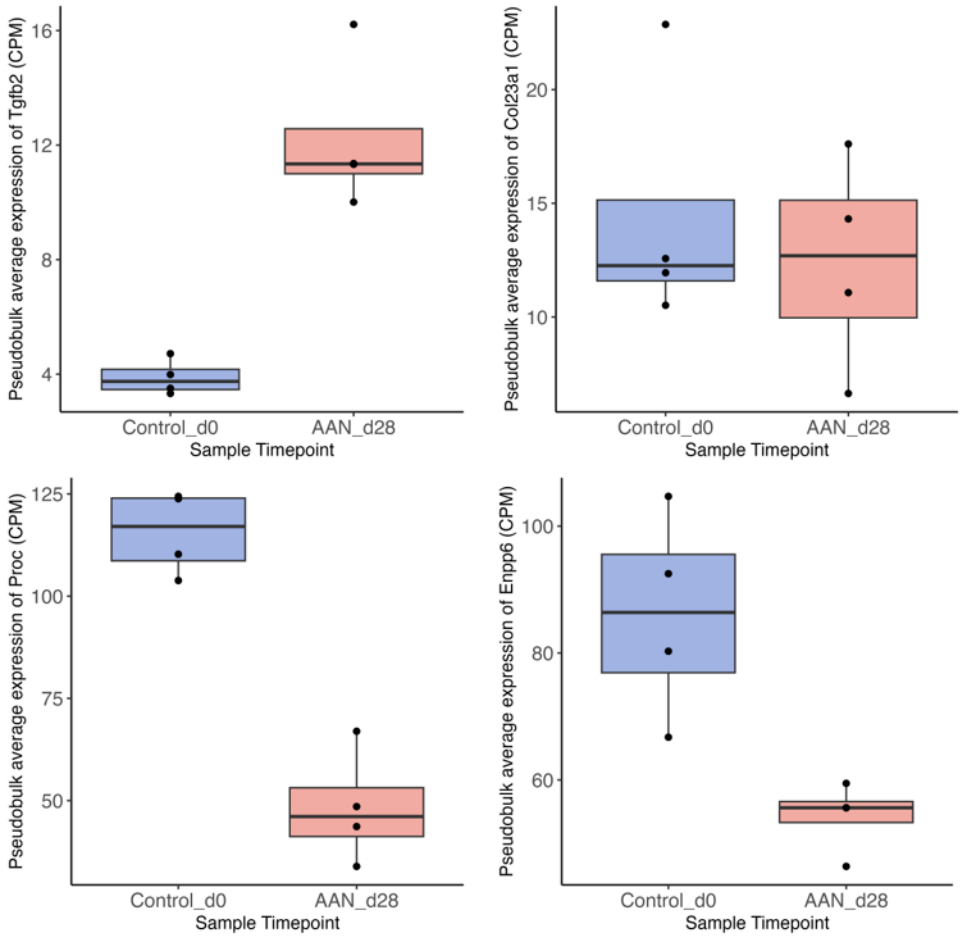

Fig S7. Gene expression of markers of PT maladaptation and PT cells at healthy states in a publicly available snRNA-seq dataset of an AAN model SnRNA-seq data published by Lu et al *(37)* was downloaded from the ArrayExpress database (accession code No. E-MTAB-9390). Gene expression of markers on day 0 and day 28 after aristolochic acid injection (4 doses of 2.5 mg/kg over 2 weeks) is visualized after converting snRNA-seq count data to counts per million.

Table S1. Protein concentrations in the TRIBE-AKI adult cohort and 2 independent validation cohorts * All proteins were presented by names of aptamers used in the SomaScan® assay. Values are presented as the median (IQR) of normalized aptamer measurements in relative fluorescence intensities.

Abbreviations: AFM, afamin; COL23A1, collagen type XXIII α 1 chain; ENPP6, ectonucleotide pyrophosphatase/phosphodiesterase 6; NLGN4X, neuregulin-4 X linked; P4HA2, prolyl 4-hydroxylase; PROC, protein C; PLG, plasminogen; PT, proximal tubule; TGFB2, transforming growth factor β-2.

	TRIBE-AKI Adult (N = 322)		Pediatric Cardiac Surgery (N = 68)		Marathon (N = 39)	
Protein Name*	**Preoperative**	**Postoperative**	**Preoperative**	**Postoperative**	**Prerace**	**Postrace**
NLGN4X.5357.60	358 (334-391)	567 (482-680)	386 (356- 470)	625 (538- 837)	354 (335- 385)	404 (386- 434)
COL23A1.4543.65	703 (512-1,027)	120 (918-1,459)	556 (500- 684)	849 (728- 1,043)	264 (249- 285)	377 (334- 411)
TGFB2.4156.74	298 (261-380)	558 (439-718)	280 (255- 321)	430 (358- 518)	960 (836- 1,170)	1,018 (881- 1,258)
CD200.5112.73	1,662 (1,532- 1,872)	2,096 (1,836- 2,378)	1,744 (1,665- 1,891)	1,918 (1,824- 2,086)	1,521 (1,470- 1,597)	1,867 (1,751- 2,031)
ENPP6.15579.26	654 (616-719)	543 (499-590)	735 (680- 810)	567 (525- 618)	745 (683- 768)	800 (725- 857)
PLG.3710.49	23,389 (20,593-25,869)	18,045 (15,981-20,348)	20,384 (18,367- 23,113)	17,725 (16,018- 19,724)	23,013 (20,305- 26,357)	21,165 (19,355- 25,348)
PROC.2961.1	73,950 (63,166- 83,586)	68,505 (61,261- 79,560)	65,910 (54,156- 73,059)	58,245 (52,598- 66,119)	76,172 (71,758- 80,537)	64,882 (58,153- 72,615)
P4HA2.11348.132	1,106 (999-1,228)	1,048 (962-1,131)	935 (870- 1,028)	868 (814- 934)	1,119 (1,043- 1,275)	1,067 (948- 1,204)
AFM.4763.31	45,905 (40,285-51,472)	45,249 (39,436-51,053)	40,233 (36,576- 46,950)	46,759 (42,098- 50,751)	49,699 (44,046- 52,624)	48,302 (43,428- 51,216)

Table S2. Correlation of preoperative candidate markers of PT maladaptation and PT cells at healthy states with baseline eGFR in 322 TRIBE-AKI adult participants All proteins were presented by names of aptamers used in the SomaScan® assay. Abbreviations: AFM, afamin; COL23A1, collagen type XXIII α 1 chain; eGFR: estimated glomerular filtration rate; ENPP6, ectonucleotide pyrophosphatase/phosphodiesterase 6; NLGN4X, neuregulin-4 X linked; P4HA2, prolyl 4-hydroxylase; PROC, protein C; PLG, plasminogen; PT, proximal tubule; TGFB2, transforming growth factor β-2.

Protein Name	**Spearman Correlation Coefficient**	***P* Value**
NLGN4X.5357.60	-0.02	0.68
COL23A1.4543.65	0.03	0.61
TGFB2.4156.74	0.07	0.21
CD200.5112.73	-0.01	0.85
ENPP6.15579.26	0.06	0.3
PLG.3710.49	0.24	<0.001
PROC.2961.1	0.3	<0.001
P4HA2.11348.132	0.17	0.003
AFM.4763.31	0.27	<0.001

Table S3. Changes in urinary markers of PT maladaptation and PT cells at healthy states in 54 TRIBE-AKI participants * All proteins were presented by names of aptamers used in the SomaScan® assay. Values are presented as the median (IQR) of normalized aptamer measurements in relative fluorescence intensities with and without normalization by urine creatinine concentration (mg/dL).

#Fold change is based on comparing postoperative versus preoperative protein concentrations.

Abbreviations: AFM, afamin; COL23A1, collagen type XXIII α 1 chain; NLGN4X, neuregulin-4 X linked; PROC, protein C; PLG, plasminogen; PT, proximal tubule; TGFB2, transforming growth factor β-2; UCr, urine creatinine.

Protein Name*	Preoperative	Postoperative	Fold Change[#]	*P* Value
NLGN4X.5357.60	19.90 (15.83, 24.3)	52.15 (34.15, 83.8)	2.62	2.19E-08
COL23A1.4543.65	22.60 (16.55, 33.05)	61.00 (34.93, 111.85)	2.84	3.77E-07
TGFB2.4156.74	31.80 (20.7, 48.67)	25.60 (17.12, 36.05)	0.82	6.79E-02
CD200.5112.73	2,213.25 (1,326.02, 2,893.18)	1,364.05 (891.87, 2,186.9)	0.69	1.92E-02
PLG.3710.49	103.80 (26.83, 234.72)	136.1 (47.15, 356.48)	1.81	3.72E-02
PROC.2961.1	113.85 (59.35, 222.97)	135.8 (87.17, 305.02)	1.48	1.85E-01
AFM.4763.31	28,715.05 (14,351.73, 56,946.27)	64,010.65 (23,846.8, 98,098.25)	1.84	1.67E-03
NLGN4X.5357.60/UCr ratio	0.24 (0.13, 0.39)	0.89 (0.4, 1.8)	3.48	1.14E-08
COL23A1.4543.65/UCr ratio	0.26 (0.18, 0.44)	0.85 (0.53, 1.5)	3.23	8.95E-08
TGFB2.4156.74	0.36 (0.26, 0.55)	0.38 (0.2, 0.7)	0.82	8.3E-01
CD200.5112.73/UCr ratio	22.78 (17.39, 27.72)	20.5 (13.84, -28.75)	0.92	2.93E-01
PLG.3710.49/UCr ratio	1.09 (0.36, 3.04)	1.49 (0.67, 7.31)	2.61	3.61E-03
PROC.2961.1/UCr ratio	1.31 (0.55, 2.52)	2.26 (0.96, 4.92)	1.99	2.3E-03
AFM.4763.31/UCr ratio	325.03 (135.45, 672.37)	813.83 (289.24, 2,052.05)	2.32	7.75E-05

Table S4. Performance of markers of PT maladaptation and PT cells at healthy states in predicting severe AKI after cardiac surgery when added to known kidney disease markers Markers are presented in protein names. Predictive performance (AUC) was compared using bootstrap analysis of 1,000 samples. *$P < 0.05$; **$P < 0.01$. Abbreviations: AFM, afamin; COL23A1, collagen type XXIII α 1 chain; KIM1, kidney injury molecule-1; NGAL, neutrophil gelatinase associated lipocalin; NLGN4X, neuregulin-4 X linked; PROC, protein C; PLG, plasminogen; TGFB2, transforming growth factor β-2; suPAR, soluable urokinase-type plasminogen activator receptor.

Marker	AUC	Marker	AUC	Marker	AUC
KIM1	0.73 (0.64- 0.82)	NGAL	0.75 (0.66- 0.84)	suPAR	0.74 (0.65- 0.83)
+TGFB2	0.76 (0.68- 0.84)	+TGFB2	0.77 (0.69- 0.85)	+TGFB2	0.76 (0.68- 0.85)*
+COL23A1	0.77 (0.69- 0.84)*	+COL23A1	0.77 (0.69- 0.85)	+COL23A1	0.77 (0.69- 0.85)*
+NLGN4X	0.78 (0.7- 0.86)**	+NLGN4X	0.79 (0.7- 0.87)*	+NLGN4X	0.78 (0.7- 0.86)**
+CD200	0.75 (0.67- 0.83)	+CD200	0.77 (0.69- 0.85)	+CD200	0.76 (0.68- 0.84)
+PLG	0.77 (0.68- 0.85)*	+PLG	0.78 (0.69- 0.87)*	+PLG	0.78 (0.69- 0.86)*
+PROC	0.76 (0.67- 0.84)*	+PROC	0.77 (0.68- 0.85)	+PROC	0.77 (0.68- 0.85)*
+ENPP6	0.77 (0.69- 0.85)*	+ENPP6	0.78 (0.7- 0.86)*	+ENPP6	0.77 (0.69- 0.85)*
+P4HA2	0.77 (0.68- 0.85)*	+P4HA2	0.78 (0.69- 0.86)	+P4HA2	0.77 (0.68- 0.84)
+AFM	0.75 (0.66- 0.83)	+AFM	0.77 (0.68- 0.85)	+AFM	0.75 (0.66- 0.83)

Table S5. Correlation in gene expression between markers of PT maladaptation and PT cells at healthy states and fibrosis markers in mouse models of kidney atrophy and repair after IRI *Assessed using Pearson correlation.

Marker/ Fibrosis Gene	***Tgfb2***		***Col23a1***		***Enpp6***		***Proc***	
	Correlation*	*P* Value	Correlation *	*P* Value	Correlation *	*P* Value	Correlation *	*P* Value
Col1a1	0.59	9.41E-07	0.01	0.93	-0.53	2.11E-05	-0.60	6.95E-07
Col3a1	0.49	9.85E-05	-0.13	0.34	-0.41	1.27E-03	-0.47	2.04E-04
Fn1	0.55	6.50E-06	-0.12	0.37	-0.46	3.01E-04	-0.54	1.19E-05
Pdgfrb	0.71	4.63E-10	0.15	0.26	-0.62	1.64E-07	-0.50	6.94E-05
Acta2	0.26	0.051	-0.30	0.02	-0.18	0.16	-0.32	0.013

Data file S1. Demographic and clinical characteristics of participants with AKI and healthy controls in KPMP and HuBMAP consortia

Data file S2. Summary of snRNA-seq data library before and after quality control

Data file S3. Differential gene expression in subclusters of PT cells in 17 participants with AKI and 7 healthy controls from the KPMP and HubMAP consortia

Data file S4. Gene set enrichment analysis of subclusters of PT cells in 17 participants with AKI and seven healthy controls from the KPMP and HubMAP consortia

Data file S5. Regulon enrichment score in subclusters of PT cells in 17 participants with AKI from the KPMP consortium

Data file S6. List of transcription factors and target genes forming regulons in PT cells in 17 participants with AKI from the KPMP consortium