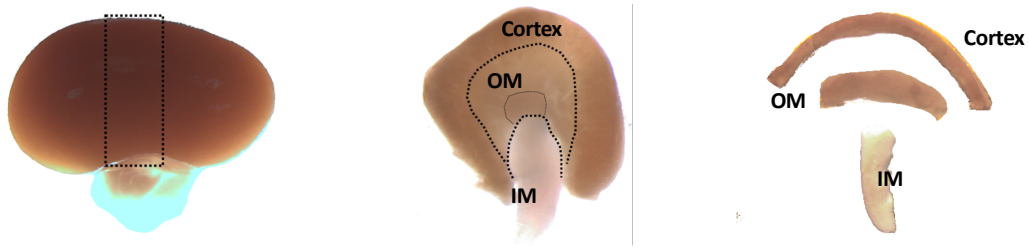
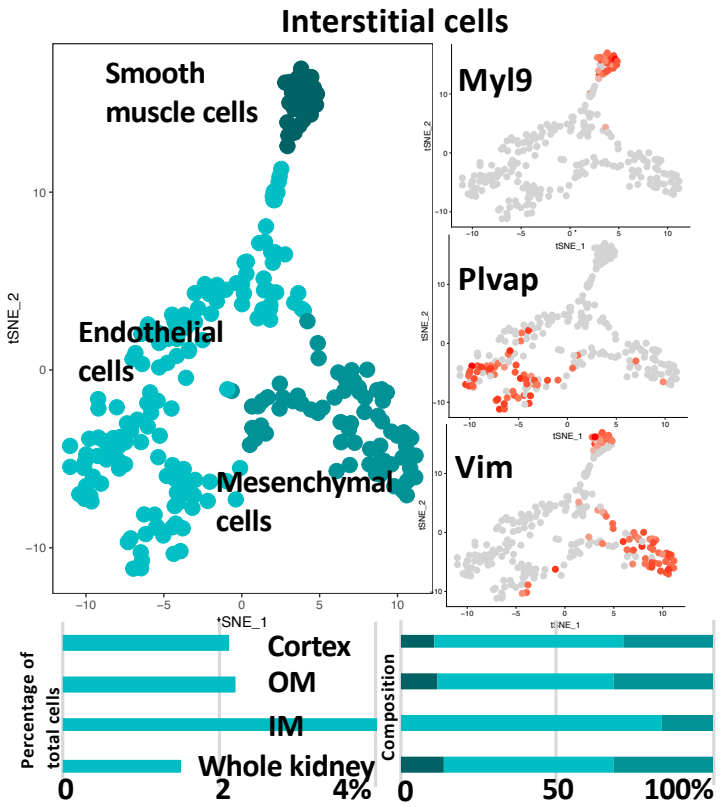
Supplemental Material for

Kidney single-cell transcriptomes predict spatial corticomedullary gene expression and tissue osmolality gradients

Supplemental material Table of Contents



Supplemental Figure S1: Preparation of dissected samples from whole kidneys. A slice as indicated by the dotted rectangle in the left picture was cut and dissected under the microscope. Individual regional samples as shown in the right picture were then subjected to scRNA-seq. For whole kidney samples, the whole organ was used.



Supplemental Figure S2: Clustering of interstitial cells from baseline kidney samples. Upper panels show tSNE plots and clustering results of sub clustering of cells assigned to interstitial cells in Fig. 1B. Using marker gene expression reveals distinct cell

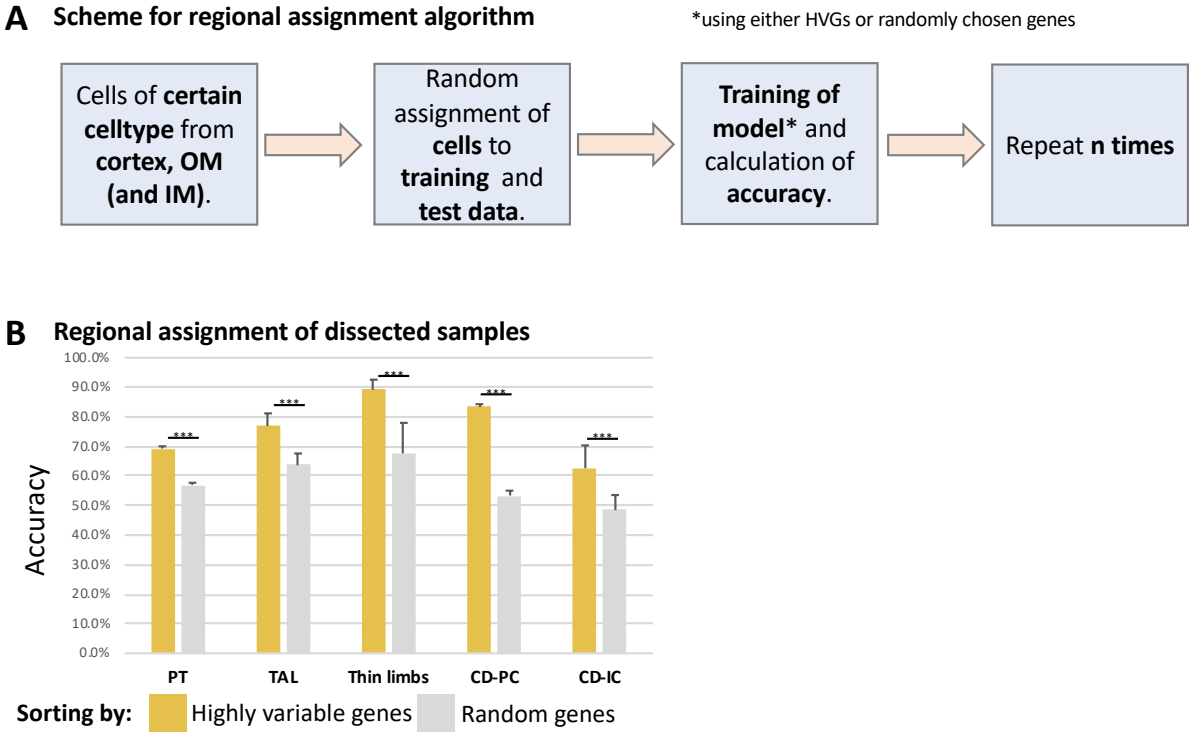
subpopulations. Abundances of the respective cell types are depicted in the lower panels. Interstitial cells comprise a heterogeneous group of cell types including smooth muscle, endothelial and vimentin-positive mesenchymal cells such as parietal epithelial and mesangial cells. Immune cells in our samples were mostly comprised of MHC II-positive cells which is consistent with the literature as most of the resident immune cells of the kidney are dendritic cells and macrophages (not shown).

See external Excel file: Suppl.Table.S1.marker.genes.xlsx

Supplemental Table S1: Cell type-specific marker genes.

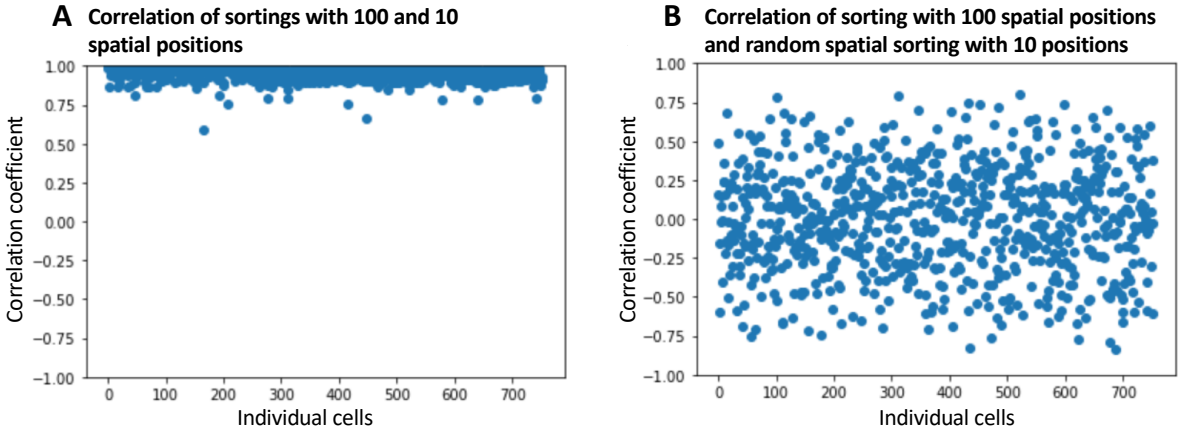
See external Excel file: Suppl.Table.S2.HV.genes.xlsx

Supplemental Table S2: Cell type-specific highly variable genes.

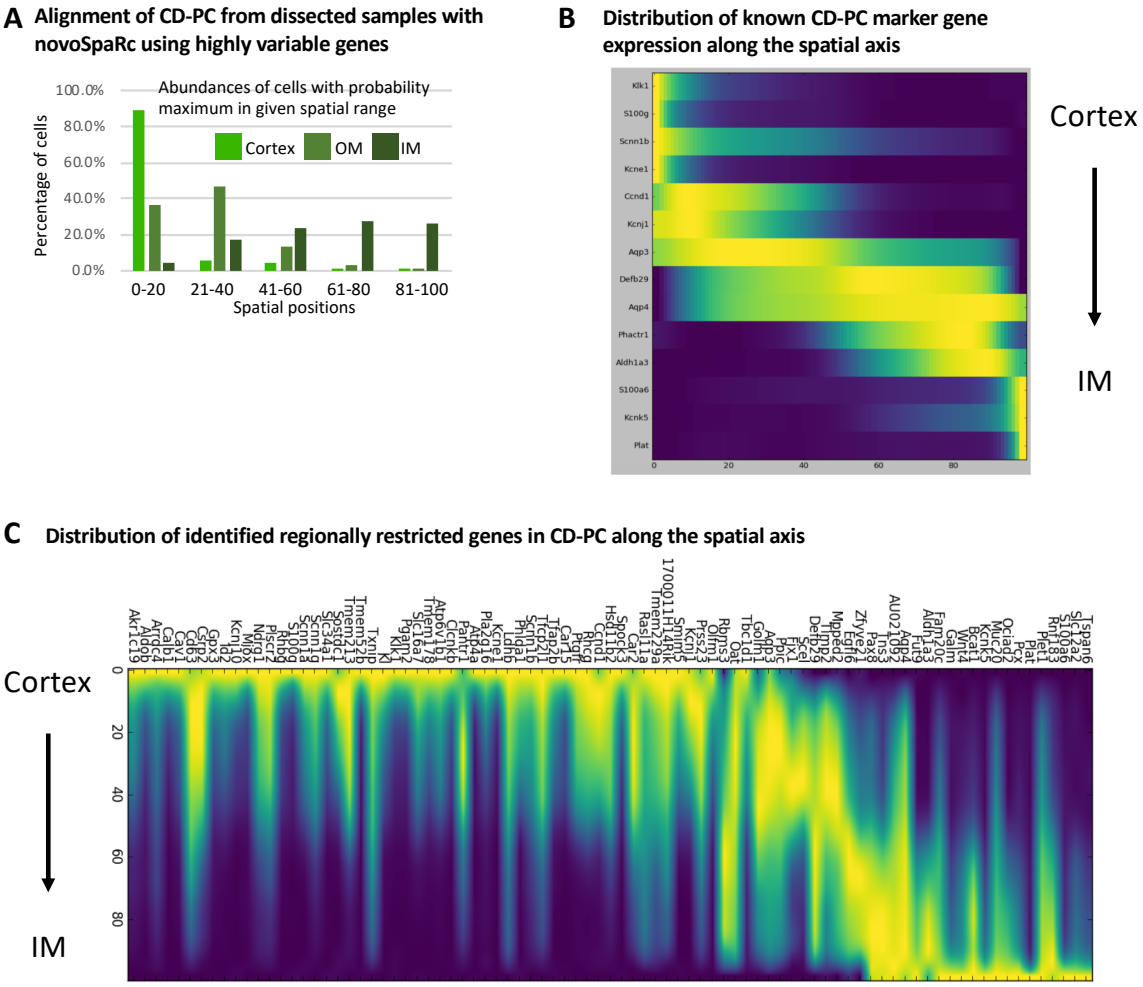


Supplemental Figure S3: Correct regional assignment of kidney tubule cells based on their transcriptomes is cell type-dependent. A. Scheme for regional assignment algorithm using a multinomial model. **B.** Accuracy of regional assignment (mean ± SD)

from dissected scRNA-seq data reveals cell type dependency. Highly variable genes always performed better than randomly chosen genes. In general, the discernibility of cortex from OM is difficult to achieve which is reflected by the low assignment accuracy for PT. Also, correct CD-IC regional assignment is low mainly due to the mostly cortical and OM origin of CD-IC. Random assignment to training or test data was performed 10 times for each cell type for statistical analyses. P-value ***<0.001.



Supplemental Figure S4: Comparison of spatial sortings with different numbers of provided positions. A. Scatter plot of correlation coefficients for sortings of all 753 CD-PCs from microdissected samples (x-axis) by providing 100 or 10 spatial positions (correlation was computed between probability distributions for both sortings comparing each CD-PC individually, correlation is Spearman's correlation, mean correlation coefficient is 0.95, standard deviation is 0.04). Only 9 CD-PCs showed correlation coefficients below 0.8, only one cell didn't show a significant correlation p-value. **B.** The same plot as depicted in A. for random sorting using 10 spatial positions. For this, probability distributions of cells were shuffled. Correlation coeefficients are significantly lower (p-value < 0.001) in B. when compared to coefficients in A.



Supplemental Figure S5: **Validation of spatial assignments**. **A-B.** Spatial assignment of CD-PC from a second dataset[1] reveals comparable sorting results regarding sample and marker gene distribution along the corticomedullary axis. CD-PCs were identified as described in the original publication[1]. **C.** Plot of identified regionally restricted genes as presented in Fig. 3E. Gene expression was min-max-normalized per gene for the heatmap plots.

See external Excel files:

Suppl.Table.S3.decreasing.CD.PC.xlsx and Suppl.Table.S4.increasing.CD.PC.xlsx

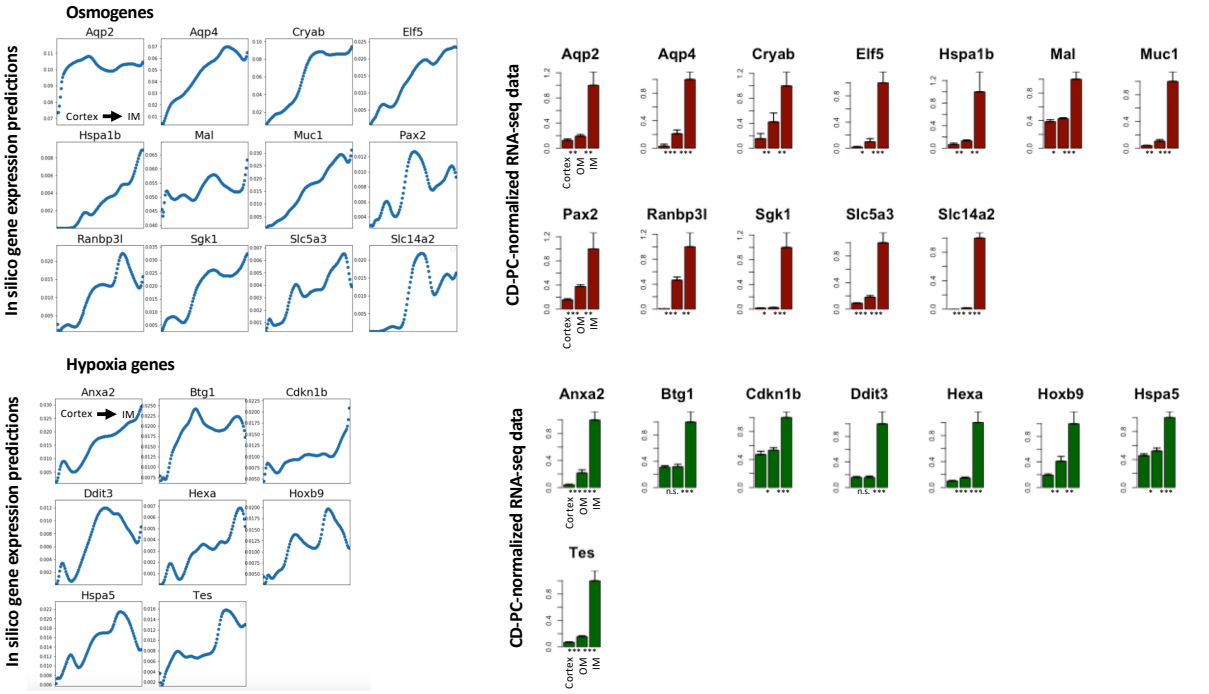
Supplemental Tables S3 and 4: Pathway enrichment analysis results and gene lists for genes with corticomedullary increasing and decreasing expression gradients.

Biomol Concepts . 2012 August ; 3(4): 345, 364.						
Experimental strategy: Review presenting results from various sources (see Table 1 in the original publication).						
Acan	Akr1b1	Aqp1	Aqp2	Aspn	Ccl2	Cd24
Col2a1	Cox2	Cryab	Cyp2e1	Cyp3a4	Cyp3a5	Cyp3a7
Cyr61	Enpp2	Hspa1b	Hspa4l	Igfbp5	Igfbp7	Il1b
Ltb	Muc5ac	Nfat5	Npr1	Ptprz1	S100a4	Sgk1
Slc14a2	Slc2a4	Slc38a2	Slc5a3	Slc6a12	Slc6a6	Sox9
Tnf	Vegfc					
Berl 2009: "How Do Kidney Cells Adapt to Survive in Hypertonic Inner Medulla?"						
Experimental strategy: Review of results from mouse renal inner medullary cell culture.						
Cldn4	Fxyd2	Mpdz	Nupp88			
Berry et al., 2017: "Renal Sodium Gradient Orchestrates a Dynamic Antibacterial Defense Zone"						
Experimental strategy: Exposure of HK2 cells to different surrounding osmolarities.						
Ccl2	Cx3cl1	Mapk13	Slc5a3			
Ho 2006: "Intracellular Water Homeostasis and the Mammalian Cellular Osmotic Stress Response"						
Experimental strategy: Review presenting results with upregulated genes from surrounding osmolarity in various also non-renal murine and human cell types from cell culture (see Table 2 in the original publication).						
Akr1b1	Aqp2	Hspa2	Hspa4l	Slc14a2	Slc38a2	Slc5a3
Slc6a12	Slc6a6					

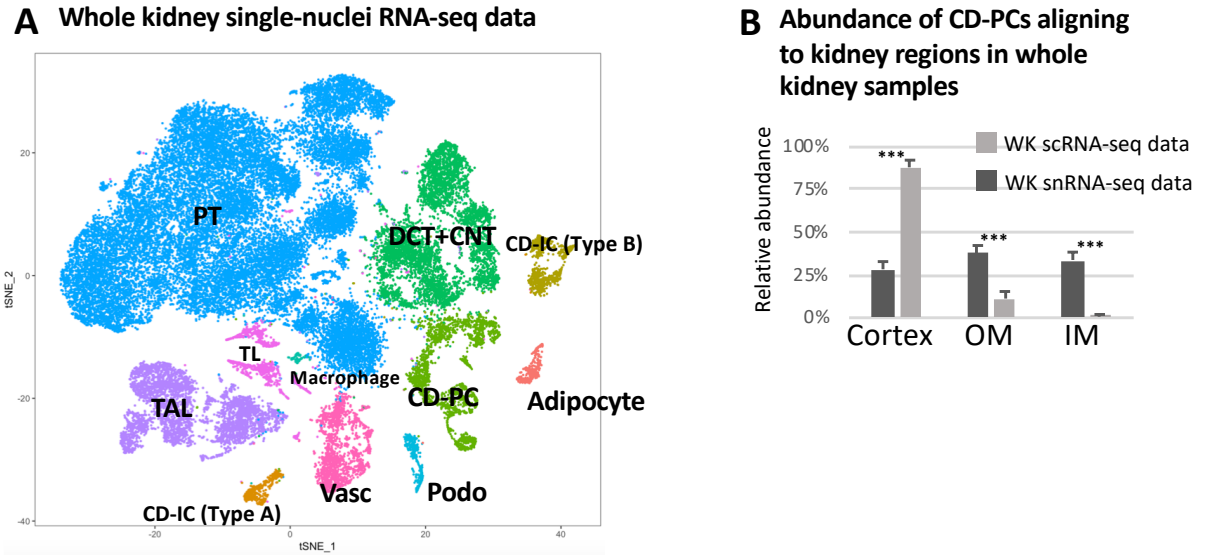
Siroky et al., 2017: "Primary cilia regulate the osmotic stress response of renal epithelial cells through TRPM3"						
Experimental strategy: Exposure of mouse IMCD3 and 176-5 cells to different surrounding osmolarities.						
Akr1b3	Bgt1	Slc6a12				
Colleen et al., 2010: "Osmoregulation of ceroid neuronal lipofuscinosis type 3 in the renal medulla"						
Experimental strategy: Primary mouse renal IM cell culture.						
Cln3						
Cohen 1999: "Clinical and Experimental Pharmacology and Physiology"						
Experimental strategy: Review presenting results from various sources.						
Egr1						
Aboudehen et al., 2017: "Hepatocyte Nuclear Factor-1β Regulates Urinary Concentration and Response to Hypertonicity"						
Experimental strategy: Mouse primary IMCD cells exposed to different surrounding osmolarities.						
Nr1h4						
Cai et al., 2005: "Pax2 expression occurs in renal medullary epithelial cells in vivo and in cell culture, is osmoregulated, and promotes osmotic tolerance"						
Experimental strategy: IMCD3 and primary IMCD cells from mice exposed to different surrounding osmolarities.						
Pax2						
Schulze Blasum et al., 2016: "The kidney-specific expression of genes can be modulated by the extracellular osmolality"						
Experimental strategy: Gene expression analysis of rat primary IMCD cells exposed to different surrounding osmolarities.						
Akr1b1	Akr7a3	Ampd1	Aqp2	Aqp3	Aqp4	Bsnd
Ca12	Capn13	Ccl20	Csrnp1	Ctse	Cxcl10	Egr1
Elf5	Fa2h	Fosb	Fxyd4	Guca2a	Guca2b	Hspa4l

Kcnj13	Klrk1	Krt23	Loc153328	Mal	Mmp13	Muc1
Mx2	Nupr1	Pcsk1	Ppp1r10	Ppyr1	Prss35	Ptgs2
Pygm	Ranbp3l	Rrad	S100a5	Serpina3	Serpinb2	Sla
Slc14a2	Slc4a11	Slc5a3	Slc6a12	Slco4a1	Sprr1a	

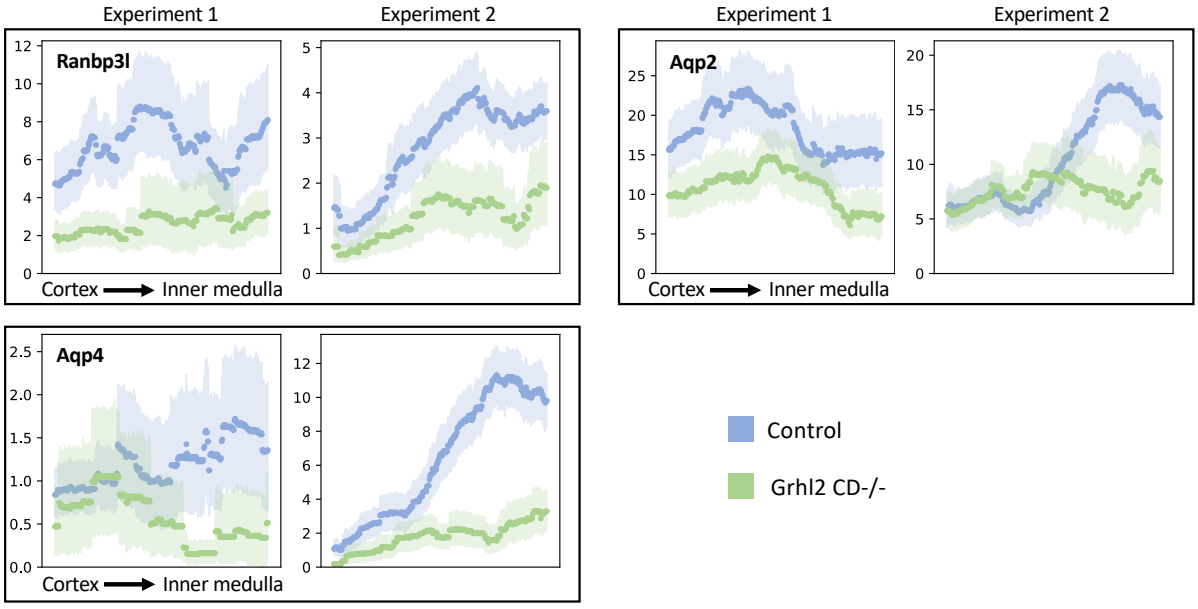
Supplemental Table S5: Osmogenes identified by literature research as indicated.



Supplemental Figure S6: Osmogene and hypoxia gene filtering. Filtered osmogenes (top) and hypoxia genes (bottom). Depicted are *in silico* gene expression gradient predictions as well as bar graphs from RNA-seq from dissected cortex, OM and IM (n=5). Since CD-PC abundances vary heavily between cortex and inner medulla, TPM values from bulk RNA-seq data were normalized to the abundance of CD-PC (mean ± SD, Suppl. Meth.) and adjusted to the mean in IM. An osmogene was considered to show an increasing expression gradient from cortex to inner medulla in RNA-seq data if its mean expression of CD-PC-normalized TPM values was strictly increasing from cortex to inner medulla and if it was at least significantly higher expressed in outer medulla than cortex or inner medulla than outer medulla (p-value<0.05). P-value: *<0.05, **<0.01, ***<0.001, n.s. – not significant.



Supplemental Figure S7: Whole kidney snRNA-seq data and spatial alignment of whole kidney snRNA-seq and scRNA-seq data. A. Joint TSNE plot for snRNA-seq data from two Grhl2 $^{CD\ -/-}$ kidneys and two control littermates (PT – proximal tubule, TL – thin limb, TAL – thick ascending limb, DCT – distal convoluted tubule, CNT – connecting tubule, CD-PC/IC – collecting duct principal/intercalated cells, Vasc – vasculature, Podo – podocytes). **B.** Applying the strategy presented in Fig. 6A to whole kidney single-cell and single-nuclei RNA-seq data reveals a significantly distinct regional distribution of CD-PCs in single-cell and single-nuclei RNA-seq data. To statistically compare results from both sortings, 10% of CD-PCs were randomly sampled without returning for each data type and spatially assigned (10 iterations in total, mean ± SD). P-value ***<0.001.



Supplemental Figure S8: **Spatial expression of filtered osmogenes.** Spatial expression from in silico prediction of snRNA-seq data of two independent experiments (experiment 1 and 2) for the remaining osmogenes from the filtering process depicted in Fig. 4C. Analogous to Fig. 6B, we show a sliding average over 150 spatial positions for each whole kidney snRNA-seq sample (dotted lines with 95% confidence intervals indicated by light shades). Expression values are library-normalized raw counts.

Supplemental Methods for

Kidney single-cell transcriptomes predict spatial corticomedullary gene expression and tissue osmolality gradients

Preparation of cell suspensions for SC-seq

Mouse kidneys were harvested and prepared as described in the methods section. The following protocol provides a step-by-step description of the digestion process.

Required Reagents:

Reagent	Supplier	Catalog no.	Storage Conditions
DPBS (no Ca, no Mg)	ThermoFisher	14190144	4° C
0.5 M EDTA	Ambion	AM9260G	room temp.
RBC Lysis Buffer	Sigma	R7757	4° C
Protease from *Bacillus Licheniformis*	Sigma	P5380	Store 100 µL aliquots (100 mg/mL) in DPBS at -80° C
DNAse	Applichem	A3778	Store 10 µL aliquots (250 U/10 µL) in DPBS at -80° C

Prepare *Bacillus Licheniformis* enzyme mix just prior to starting dissociation:

Volume (µL)	Reagent	Final concentration
894	DPBS	1X
1	0.5 M EDTA	0.5 mM
5	DNAse 1 (250 U/10 µL)	125 U / mL
100	*B. Lich* (100 mg/mL)	10 mg/mL

+25 mg tissue / 1 mL enzyme mix

After sacrificing the animal, perfuse mouse with 4°C cold PBS by puncturing the left cardiac ventricle and slitting the right atrium (perfusion speed 10ml/min, perfusion volume 30ml).

A. Isolate Kidney

1. Quickly dissect and isolate kidney in ice-cold PBS.
2. Remove fatty tissue and kidney capsule in ice-cold PBS.
3. Mince on petri dish, on ice (~2 min) until fine.
4. Weigh out 25 mg tissue per 1 mL enzyme mix (10 mg/mL).

B. Initial digestion

5. Place tissue in eppendorf tube containing 1 mL digest mix on ice.
6. After 2 min, triturate gently 20x using 1 mL pipet set to 700 µL.
7. During the first 8 minutes, remove tube and shake every minute to re-suspend tissue chunks in addition to triturating 10x every 2 minutes.
8. After 9 min, let tissue chunks settle for 1 min on ice.
9. Remove ~80% of supernatant and filter using 30 µM filter - rinse w/6 mL PBS/BSA into 50 mL conical. Save conical with filter for subsequent steps.

C. Second digestion

10. Add additional 1 mL enzyme mix to residual tissue chunks.
11. Triturate 10x w/ 1 mL pipet set to 700 µL.
12. Continue digesting while shaking in thermomixer at 1100 RPM.
13. Every 2 min remove tube to triturate 10X.
14. After 19 min total digest, let tissue chunks settle for 1 min on ice
15. Remove ~80% of supernatant and filter using 30 µM filter. Rinse w/6 mL PBS/BSA into the same 50 mL conical.

D. Third digestion

16. Add additional 1 mL enzyme mix to residual tissue chunks.
17. Continue digesting at 1200 RPM in thermomixer at 4 C.
18. Every 3 min passage 8X w/18 gauge needle w/1 mL syringe (2X total).
19. After 29 minutes total digest time, let tissue chunks settle on ice for 1 min.
20. Remove 80% of supernatant and filter using 30 µM filter. Rinse w/6 mL PBS/BSA into the same 50 mL conical.

E. Fourth digestion

21. Add additional 1 mL enzyme mix to residual tissue chunks.
22. Continue digesting at 1400 RPM in thermomixer at 4 C.
23. Every 3 min passage 8X w/18 gauge needle w/1 mL syringe (2x total).
24. After 40 min total digest time, triturate 10x using 1 mL pipet and add to the same 30 µM filter. Rinse w/6 mL PBS/BSA.

F. Preparing cells for SC-seq

25. Spin 500 G for 6 min at 4°C.

26. Re-suspend pellet in 1 mL PBS/0.04% BSA. Analyze viability using hemocytometer with trypan blue.

Fix cells with methanol and proceed to Dropseq after rehydration of cells as previously published[2].

Preparation of suspensions for single-nuclei sequencing

Mouse kidneys were harvested and prepared as described in the methods section. The following protocol provides a step-by-step description of the generation of the single-nuclei suspension.

Add Inhibitors right before usage:
Prepare per Sample:
Nuclear Lysis Buffer 1 (NLB1)
4ml of Nuclear Lysis Buffer (Sigma) +
100µl of Ribolock (final concentration of 1U/µl) +
200µl VRC (200mM stock, leading to 10mM final concentration,
according to:
https://www.neb.com/-/media/catalog/datacards-or-manuals/s1402datasheet-lot0031207.pdf

Nuclear Lysis Buffer 2 (NLB2)
4ml of Nuclear Lysis Buffer (Sigma) +
100µl of Ribolock (final concentration of 1U/µl))

Nuclear suspension buffer (NSB)
Use 2ml PBS/0.04%BSA (=10x solution, see Doc. CG00053 from 10x) +
50µl Ribolock (final conc. 1U/µl)

Additional equipment:
60mm petri dish
razor blades

20µm, 35µm ,100µm strainer

Steps:

Pre-Cool centrifuges, falcon tubes, douncer and Nuclei Lysis Buffer (Sigma). Everything on ice unless stated otherwise.

1. Take one frozen kidney sample and put in petri dish on ice with 1ml of NLB1.
2. Mince thoroughly with a fresh razor blade, try to remove fatty tissue if possible.
3. Transfer into douncer.
4. Dounce slowly 25x with pastel A. Make sure to avoid air bubbles.
5. Pass the homogenate through a 100µm strainer and add another 1ml of NLB1.
6. Wash douncer with lysis buffer.
7. Transfer the homogenate again into the douncer and dounce slowly with pastel B 15x.
8. Transfer homogenate to 15ml falcon tube, add another 2 ml of NLB1 and incubate for 5 minutes on ice.
9. Pass the homogenate through a 35µm strainer.
10. Spin tube for 5 minutes at 500 g at 4°C.
11. Remove supernatant.
12. Carefully resuspend the pellet in 4ml of NLB2.
13. Underlay suspension with 1ml of EZ lysis buffer containing 10% sucrose and 25µl Ribolock.
14. Centrifuge at 4°C at 1400rpm for 5min.
15. Remove supernatant.
16. Carefully resuspend the pellet in 1ml of NSB.
17. Pass through 20µm strainer (maybe wet before). Note: Our nuclei range up to 8µm.

Nuclei suspensions were then stained with DAPI and FACS-sorted. Nuclei concentration (numbers of nuclei per volume) was determined using the DAPI stain. The nuclei suspension was then treated according to the protocol for v2 (experiment 1) or v3.1 (experiment 2) chemistry from 10x genomics targeting 10000 nuclei.

Computational methods

Multinomial model for regional assignment

Regional assignment was performed per cell type on microdissected samples. In general, input to the model is expression on selected genes, readout is categorical (cortex, OM, IM). In more detail, for each cell type, cells were randomly divided into training (2/3 of cells) and test dataset (remaining 1/3 of cells). Input to the multinomial model training was either expression of training cells on cell type-specific highly variable genes or random genes. We then used the glmnet package,
https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html#log),
with type multinomial for further analyses. Glmnet produces different models for different values of lambda which determines how hard overcomplexity of the respective model gets punished. Each so-generated model was tested on the test data and the model with the highest accuracy on the test data was determined (highest correct regional assignment to cortex, OM or IM on test data). This process was repeated 10 times for each cell type to get the presented statistics. For each iteration, the assignment to either test or training data was random.
Usually the numbers of cells of a certain cell type (e.g. CD-PC) in each region are not equal (more CD-PC in IM than in cortex). To avoid biases but still use 2/3 of all cells of a certain cell type from each region, cells from regions with lower cell number of the respective type are randomly oversampled in the training dataset.

NovoSpaRc sorting and spatial assignment of whole kidney snRNA-seq data

NovoSpaRc sorting was performed on log-normalized data from the Seurat's object data slot (dge). Scripts and datasets for sorting of dissected samples as well as whole kidney single-nuclei data are online available at www.mdc-berlin.de/kidneyspatial.zip.

Identification of regionally restricted genes

CD-PCs in our dataset were identified as described above. For the dataset in Ransick et al.[1], CD-PCs were identified according to the instructions in the original publication. CD-PCs from both datasets were sorted using novoSpaRc and CD-PC-specific HVGs. In

case of the Ransick dataset, 843 of the 872 original CD-PC HVGs were expressed and hence used for sorting using 100 spatial positions.

We then used the Ransick dataset as discovery dataset. The reason for this was that we saw more genes with multiple peaks in the Ransick dataset. This might be due to more expressed genes using the 10x platform when compared to Dropseq.

Genes were filtered for having only one major peak along the corticomedullary axis in the Ransick dataset using scipy.signal.find_peaks with window=20 to avoid directly adjacent local maxima. Expression values on the interval borders (0 and 100) were evaluated manually and compared to the inner-interval maxima.

For expression level filtering, we calculated the distributions of maximum gene expression levels along the corticomedullary axis using a sliding window of 10 spatial positions. This resulted in a maximum gene expression distribution in every spatial position. We then used the 95^{th} percentile as expression cutoff for each spatial position, separately.

In a third step, we aimed to filter for sufficient maximum expression above background for the remaining genes. For this, we used again a sliding window of 10 spatial positions. For each gene with identified maximum expression the respective interval, we calculated the ratio between maximum expression and background expression (remaining interval outside the 20 positions window of the peak). Using all genes above the 75^{th} percentile at each position resulted in the final list of 104 genes.

Genes were considered validated by our dataset if their spatial maximum expression was comparable in our dataset. For this, we calculated maximum gene expression as in the other dataset and plotted maximum position of known marker gene expression in both datasets (markers as in Fig. 3C). We then interpolated the so-generated points using numpy.polyfit with degree 3. A gene of the filtered genes was considered to be regionally consistent if the position of its maximum expression was in a 25% interval around the cubic interpolation of the marker genes in our dataset when compared to Ransick et al.

CD-PC-normalization of RNA-seq data

To account for large differences between cortex, OM and IM in CD-PC abundance, we normalized RNA-seq TPM values from cortex, OM and IM for each gene using the

following formula (the formula was used for each gene and cortex, OM and IM, separately):

$E = (1-f_{CD})*E_{others} + f_{CD}*E_{CD}$

f_{CD} - fraction of CD-PC (this was calculated as mean abundance from our microdissected SC data and from a second microdissected SC dataset[1] for cortex, OM and IM, separately and was 2.55% for cortex, 7.3% for OM and 38.7% for IM), E – total expression value of gene from Ransick et al., E_{others} – mean expression of gene in all non-CD-PC from Ransick et al., E_{CD} - mean expression of gene in CD-PC in respective region from Ransick et al.

This leads to:

$E_{CD} = [1- (1- f_{CD})* E_{others}/E]/ f_{CD} * E = f * E.$

The factor f was calculated for each gene for each region (cortex, OM, IM). The TPM values from bulk RNA-sequencing of cortex, OM and IM were then multiplied by the factor for the respective gene and region.

References:

1. Ransick, A*., et al.* Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. *Dev Cell* **51**, 399-413 e397 (2019).
2. Alles, J*., et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol* **15**, 44 (2017).