Supplemental Information

RNA-seq in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes

Jae Wook Lee[a], Chung-Lin Chou[a], and Mark A. Knepper[a]

[a]Epithelial Systems Biology Laboratory, Systems Biology Center,

National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland

Full Methods

Microdissection of Renal Tubule Segments. Male Sprague-Dawley rats weighing 200 - 250 g were euthanized and the left kidney was perfused with bicarbonate-free, ice-cold dissection solution (135 mM NaCl, 1 mM Na_2HPO_4, 1.2 mM Na_2SO_4, 1.2 mM $MgSO_4$, 5 mM KCl, 2 mM $CaCl_2$, 5.5 mM glucose, 5 mM HEPES, pH 7.4) followed by digestion solution containing collagenase B from *Clostridium histolyticum* (Roche Applied Science, 1 mg/mL for cortex and outer medulla, 3 mg/mL for inner medulla) and bovine serum albumin (MP Biomedical, 1 mg/mL for cortex and outer medulla, 3 mg/mL for inner medulla). For outer and inner medulla, hyaluronidase was added (Worthington Biochemical Corporation, 1 mg/mL for cortex and outer medulla, 3 mg/mL for inner medulla). The left kidney was removed and small chunks of the tissue were incubated in the same digestion solution at 37°C for 25 minutes (cortex), 50 minutes (outer medulla), or 90 minutes (inner medulla). Tubule microdissection was carried out using a Wild M8 dissection stereomicroscope (Wild Heerbrugg) that transmitted light from below a specially designed microdissection table. This technique allows each renal tubule segment to be discriminated based on its appearance and topology.[1] Nomenclature for individual segments is given in Supplemental Figure 1A. For the proximal tubule segments, S1 was identified as the proximal tubule directly attached to the glomerulus, S2 was the straight part in the medullary ray, and S3 was the proximal tubule in the outer medulla. The short descending thin limb (SDL) and long descending thin limb of the outer medulla (LDLOM) were identified by their direct attachment to the S3 segment, and distinguished from each other based on their diameters and surface appearance. The long descending thin limb of the inner medulla (LDLIM) and the thin

ascending limb (tAL) were dissected from the inner medulla. The thin ascending limb was identified by its attachment to the medullary thick ascending limb. The medullary thick ascending limbs (mTALs) were dissected from the inner stripe of the outer medulla. The cortical thick ascending limbs (cTALs) were dissected from the medullary rays of the cortex. The distal convoluted tubule (DCT) was defined as a V-shaped or convoluted segment beyond the macula densa. The DCT was typically uniform in appearance for about 1 mm and underwent a gradual transition to the connecting tubule. We took only the uniform portion of the DCT. The transition region, sometimes referred to as DCT2, was discarded. The connecting tubule was identified by its cobblestone appearance and branching structure. The cortical collecting duct (CCD) was dissected from the medullary rays of the cortex. The outer medullary collecting duct (OMCD) was dissected chiefly from the inner stripe of the outer medulla. The inner medullary collecting duct (IMCD) was dissected from the middle portion of the inner medulla. Typically, several dissected tubules were pooled for one sample, yielding 4 - 5 mm of total tubule length. This would correspond to 1,000 - 2,000 cells per sample.

Construction of cDNA libraries for RNA-seq. Dissected tubule segments were transferred using long pipette tips, washed in 1× phosphate-buffered saline (PBS) under a second dissection microscope (Wild M8 stereomicroscope, Wild Heersbrugg), and transferred to a 0.5-mL PCR tube (Sorenson Bioscience) in ~2 μL of PBS. Based on prior experience, this wash step was crucial for elimination of contaminating cells. Reverse transcription with an oligo-dT primer and cDNA amplification were done following the single-cell RNA-seq protocol.[2] For this protocol, reverse transcription requires oligo-dT primers rather than random hexamers since there is no mRNA

extraction step. For cell lysis, 20 μL of cell lysis buffer as prepared according to (ref. 2) was added to the collected tubule segments. The tube was then centrifuged for 30 s at 7,500 g at 4 °C, incubated at 70 °C for 90 s to release mRNA, and then centrifuged again for 30 s at 7,500 g at 4 °C. To start first-strand synthesis, we added 0.55 μL of cell lysate and 0.45 μL of reverse transcriptase mix [SuperScript III reverse transcriptase (13.2 U/μL, Invitrogen); RNase inhibitor (0.4 U/μL, Ambion); and T4 gene 32 protein (0.07 U/μL, Roche Applied Science)] to 4 μL of the cell lysis buffer. After first-strand synthesis, free primers were removed using exonuclease I (0.5 U/μL, New England Biolabs), poly(A)' tails were added to the 5'-ends of the DNA-RNA hybrid using dATP (3 mM, Invitrogen) and terminal transferase (0.75 U/μL, Invitrogen), and RNA template was removed using RNase H (0.1 U/ μL, Invitrogen). Second-strand synthesis was carried out using a pair of universal primers (1μM UP1, 5'-ATATGGATCCGGCGCGCCGTCGACTTTTTTTTTTTTTTTTTTTTTTT-3'; 1 μM UP2, 5'-ATATCTCGAGGGCGCGCCGGATCCTTTTTTTTTTTTTTTTTTTTTTT-3', Eurofins Genomics), dNTPs (0.25 mM each) and ExTaq HS DNA polymerase (0.05 U/μL, Clontech). The first-round PCR amplification (20 cycles) was performed using the same primers and DNA polymerase. For the second-round PCR (9 cycles), the primers were switched to 5'-NH_2-modified primers with the same nucleotide sequences as above to minimize the amount of the primers in the final RNA-seq libraries. Specifications of thermal cycles used in the above procedures are provided in [2]. After reverse transcription and amplification, cDNAs ranging from 100 to 3000 bp in length were selected on 2% agarose E-gels (Life Technologies) and extracted using Zymo Gel DNA Recovery kit (Zymo Research). Concentrations of cDNAs were measured using a Qubit

fluorometer (Life Technologies). Three hundred nanograms of cDNAs were sheared to generate ~200 bp fragments using a Covaris S2 sonication system (Covaris) according to the manufacturer's protocol. Sheared cDNAs were made into adapter-ligated cDNA libraries using an Ovation Ultralow library system (NuGen) and a Mondrian SP+ microfluidic sample preparation system (NuGen) following manufacturer's protocols. cDNAs ranging from 200 to 400 bp were selected on 2% agarose E-gel and recovered using Zymo Gel DNA Recovery kit. Paired-end sequencing was carried out on a HiSeq2000 platform (Illumina) to generate 50-bp FASTQ sequences.

Mapping to the rat reference genome. The raw FASTQ sequence files were inspected and nucleotides with sequencing quality score less than 30 (phred33) were trimmed using Trimmomatic 0.3.2[3], available at http://www.usadellab.org/cms/?page=trimmomatic. FASTQ sequences that passed this test were mapped to the rat reference genome (rn5) using Spliced Transcripts Alignment to a Reference (STAR) version 2.3.0.[4] A genomic index for rat was built from FASTA sequences of rat chromosomes (including 'random' assemblies) downloaded from the UCSC table browser website, and the FASTQ sequences were mapped to the rat reference genome using the following command: STAR --genomeDir <*genomedirectory*> --readFilesIn <*file1.fastq*> <*file2.fastq*> --runThreadN 8 --outFilterMismatchNmax 3 --genomeLoad LoadAndKeep --outSAMstrandField intronMotif --alignIntronMax 10000 --alignMatesGapMax 10000 --outFilterIntronMotifs RemoveNoncanonicalUnannotated. RNA-seq samples with <70% of mappability (i.e. fraction of reads mapped to the reference genome) were discarded. Only uniquely mapped reads were included in the downstream analysis. The numbers of mapped

reads and RPKMs for RefSeq transcripts were calculated using the HOMER software package.[5] To revise gene expression for a RefSeq transcript, RNA-seq reads in the interval between the RefSeq-annotated 3'-end and the PA-seq peak were included in calculating RPKMs. Mapped reads were converted to bedgraph format and visualized on the UCSC Genome Browser.

Polyadenylated mRNA-seq (PA-seq). To identify polyadenylation sites of renal tubule transcripts, we used a PA-seq protocol developed by the DNA Sequencing and Computational Biology Core of the National Heart, Lung, and Blood Institute [6]. This protocol captures nucleotide sequences at polyadenylation sites and sequences them in a strand-specific, paired-end manner. Ten micrograms of DNA-free total RNA prepared from a rat kidney using RNEasy Mini kit (QIAGEN) were made into a strand-specific cDNA library. This PA-seq library was sequenced using the Illumina HiSeq2000 platform and processed as described in [6] to call polyadenylation peaks. To revise the RPKM value for an underannotated RefSeq transcript, we included RNA-seq reads that mapped to the interval between the RefSeq-annotated 3'-end of the transcript and the tallest PA-seq peak within 5,000 bp downstream from the RefSeq-annotated 3'-end.

Supplemental Discussion

Comparison of Results from Single-Tubule RNA-seq with RT-PCR Results for Specific Gene Targets

To provide further evidence that the single-tubule RNA-seq method gives reliable expression profiles along the renal tubule, we compare RNA-seq data with RT-PCR results from microdissected segments in what follows.

Epithelial Sodium Channel (ENaC) Subunits. The epithelial sodium channel is a heterotrimer made of three subunits: α, β, and γ (official gene symbols: *Scnn1a, Scnn1b* and *Scnn1g,* respectively). Within the kidney, it is generally accepted that the epithelial sodium channel has its major functional role in the principal cells of the collecting duct and connecting tubule where it mediates aldosterone dependent regulation of sodium ion reabsorption. Our RNA-seq analysis shows that all three ENaC subunits are expressed in the CNT, CCD, and OMCD, consistent with the generally held view (Supplemental Figure 4A). In addition, the expression of αENaC extends upward along the renal tubule to the DCT, cTAL and mTAL. Beyond this, there is expression of the beta subunit in the DCT and cTAL. Thus, according to the RNA-seq analysis, the only subunit that is uniquely expressed in principal cell-containing segments is the γ subunit. This pattern agrees with a prior study [7] that used single-tubule RT-PCR to localize ENaC subunit expression (Supplemental Figure S6A). Although not all segments were studied and a different animal species was used, the general pattern was seen to be the same as for our RNA-seq results in rats. Specifically, (a) significant levels of α-ENaC

mRNA are detectable upstream from the collecting duct, viz. in mTAL, cTAL, and DCT, and (b) γ-ENaC mRNA had the most restricted distribution with consistently detectable levels only in the CNT, CCD, and OMCD. Another study used in situ RT-PCR hybridization to localize mRNA for α subunit. This study also showed expression of the α subunit of ENaC in the mTAL and DCT in addition to CNT and collecting duct segments [8]. In conclusion, there appeared to be general agreement between RNA-seq results and RT-PCR results with regard to distribution of ENaC subunits along the renal tubule.

Basolateral Chloride Channel ClC-K2. ClC-K2 (gene symbol: *Clcnkb*) is a chloride channel expressed in the basolateral plasma membrane in distal nephron segments and in collecting ducts. It is believed to constitute the major pathway for basolateral chloride exit in transepithelial Na-Cl reabsorption in distal nephron and collecting duct. There was a very high degree of concordance between our RNA-seq data and the RT-PCR result [9] (Supplemental Figure S6B).

Urea Channels UT-A1, UT-A2 and UT-A3. Urea channel proteins play important roles in the urinary concentrating mechanism. All three urea channel mRNAs (UT-A1, UT-A2 and UT-A3) are coded by the *Slc14a2* gene, differing on the basis of differential promoter utilization and differential splicing. The distribution of RPKM values mapped to *Slc14a2* is shown in Supplemental Figure S6C. Mapping of RNA-seq data does not readily discriminate between UTA1 and UT-A2 because they have a common 3'-end. Previous RT-PCR studies in microdissected renal tubule segments found UT-A1 mRNA in the IMCD and UT-A2 mRNA in the SDL and LDLIM but did not identify transcripts in LDLOM, exactly matching the distribution shown in Supplemental Figure 1B [10] (The RT-

PCR study did not attempt to identify UT-A3 transcripts). Thus, there was a high degree of concordance between RNA-seq data and RT-PCR data for *Slc14a2* transcripts.

Parathyroid Hormone Receptor (*Pth1r*). The parathyroid hormone receptor is a G protein-coupled receptor that mediates effects of parathyroid hormone on transport processes in the kidney. Supplemental Figure S6D shows the distribution of *Pth1r* mRNA along the renal tubule based on RNA-seq experiments report in the present paper (left) and the distribution obtained by semiquantitative RT-PCR in microdissected segments from rats based on multiple replicates (right, [11]). There appears to be a high degree of concordance between the expression patterns obtained with the two methods.

References

1. Wright PA, Burg MB, Knepper MA: Microdissection of kidney tubule segments. *Methods Enzymol* 191**:** 226-231, 1990.

2. Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, Lao K, Surani MA: RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nat Protoc* 5**:** 516-535, 2010.

3. Bolger AM, Lohse M, Usadel B: Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 2014.

4. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29**:** 15-21, 2013.

5. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK: Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38**:** 576-589, 2010.

6. Ni T, Yang Y, Hafez D, Yang W, Kiesewetter K, Wakabayashi Y, Ohler U, Peng W, Zhu J: Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy. *BMC Genomics* 14**:** 615, 2013.

7. Velazquez H, Silva T, Andujar E, Desir GV, Ellison DH, Greger R: The distal convoluted tubule of rabbit kidney does not express a functional sodium channel. *Am J Physiol Renal Physiol* 280**:** F530-539, 2001.

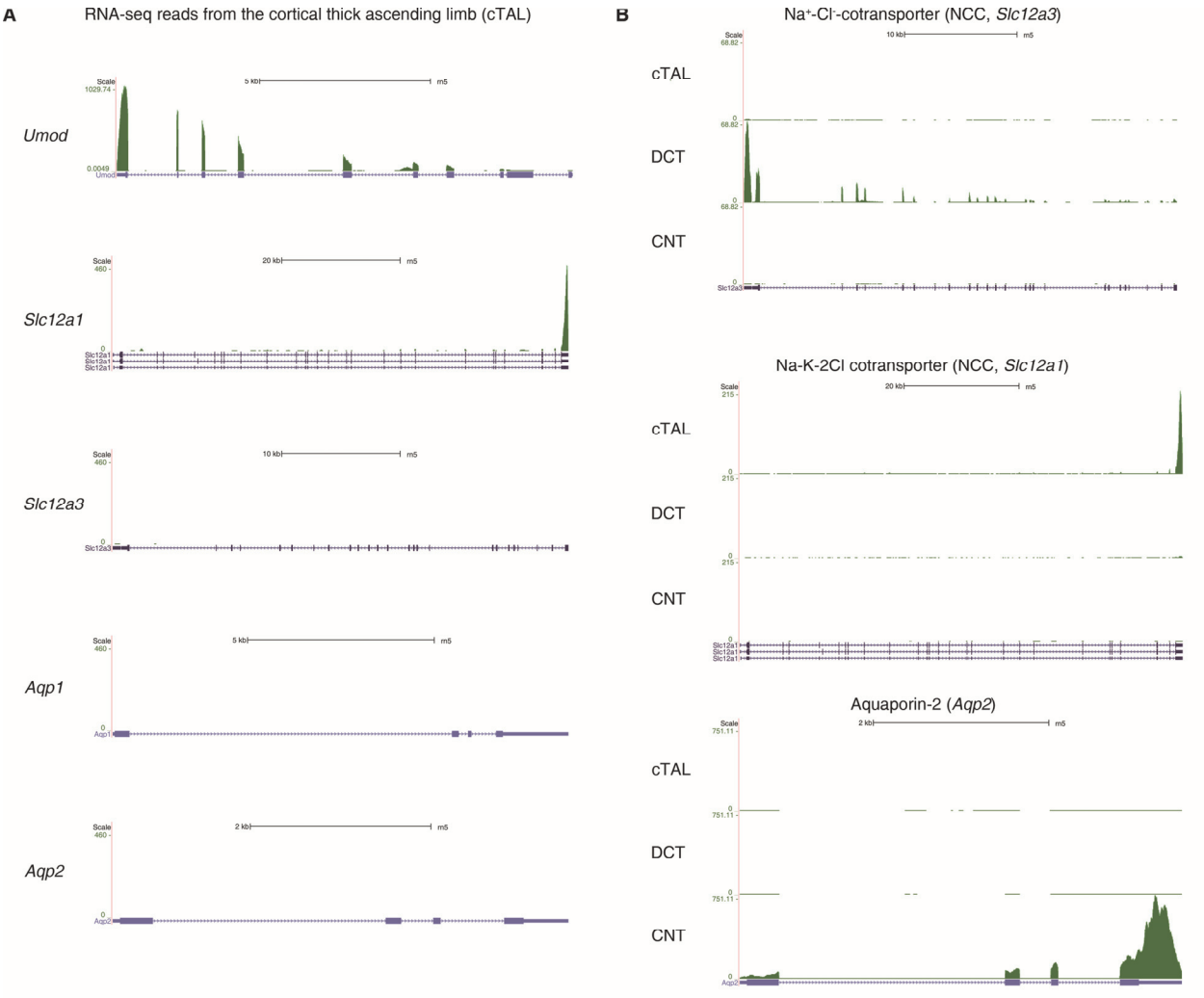
8. Ciampolillo F, McCoy DE, Green RB, Karlson KH, Dagenais A, Molday RS, Stanton BA: Cell-specific expression of amiloride-sensitive, Na(+)-conducting ion channels in the kidney. *Am J Physiol* 271**:** C1303-1315, 1996.

9. Vitzthum H, Castrop H, Meier-Meitinger M, Riegger GA, Kurtz A, Kramer BK, Wolf K: Nephron specific regulation of chloride channel CLC-K2 mRNA in the rat. *Kidney Int* 61**:** 547-554, 2002.

10. Shayakul C, Knepper MA, Smith CP, DiGiovanni SR, Hediger MA: Segmental localization of urea transporter mRNAs in rat kidney. *Am J Physiol* 272**:** F654-660, 1997.

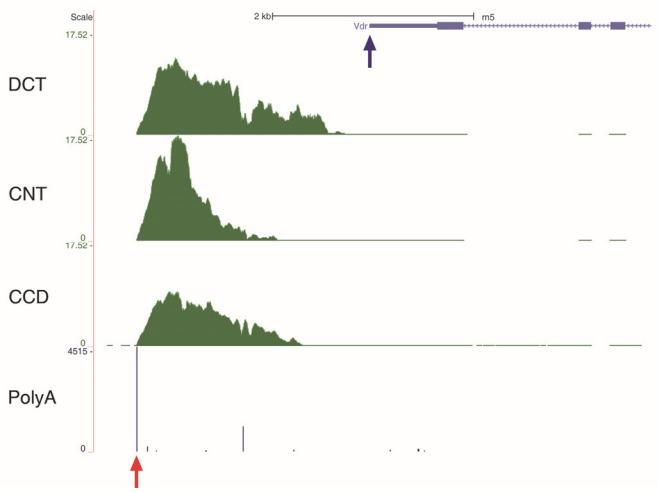
11. Yang T, Hassan S, Huang YG, Smart AM, Briggs JP, Schnermann JB: Expression of PTHrP, PTH/PTHrP receptor, and Ca(2+)-sensing receptor mRNAs along the rat nephron. *Am J Physiol* 272**:** F751-758, 1997.

Supplemental Figures.

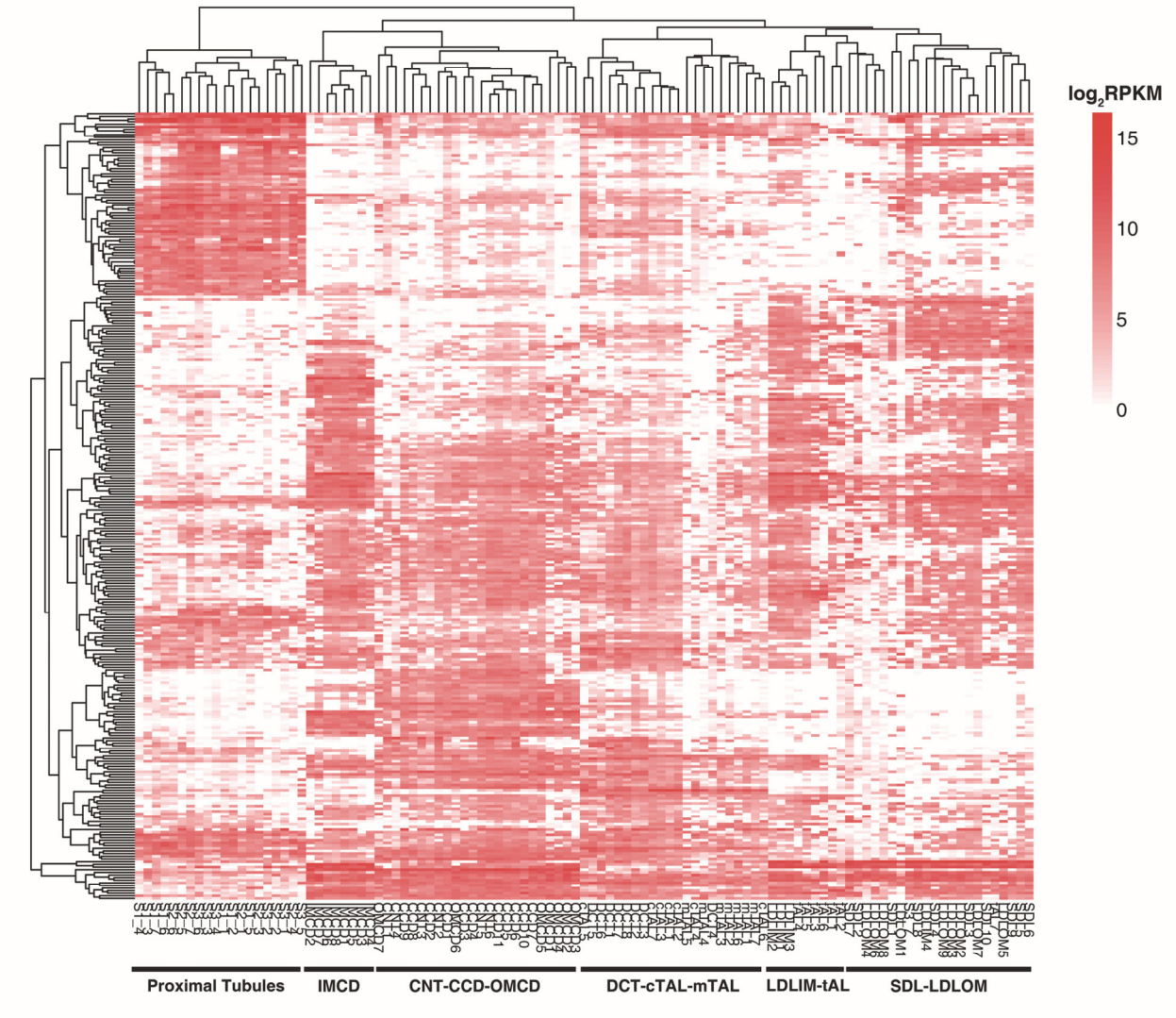


Supplemental Figure 1. RNA-seq of microdissected cortical thick ascending limbs and adjacent segments. A. RNA-seq reads from microdissected cortical thick ascending limbs (cTAL). The distribution of reads along the gene body is depicted as green histograms, aligned with diagram showing intron-exon structure. Direction of transcription is indicated by arrows. Classical markers for cTAL (*Umod* and *Slc12a1*), are highly expressed, whereas markers for neighboring segments (*Slc12a3,* distal convoluted tubule; *Aqp1,* proximal convoluted tubule; and *Aqp2,* cortical collecting duct) are not expressed. **B.** Expression of *Slc12a3*, *Slc12a1*, and *Aqp2* in the cTAL, distal

convoluted tubule (DCT), and connecting tubule (CNT). *Slc12a3* is highly expressed in the DCT, and not in the cTAL and CNT. *Aqp2* is exclusively expressed in the CNT.



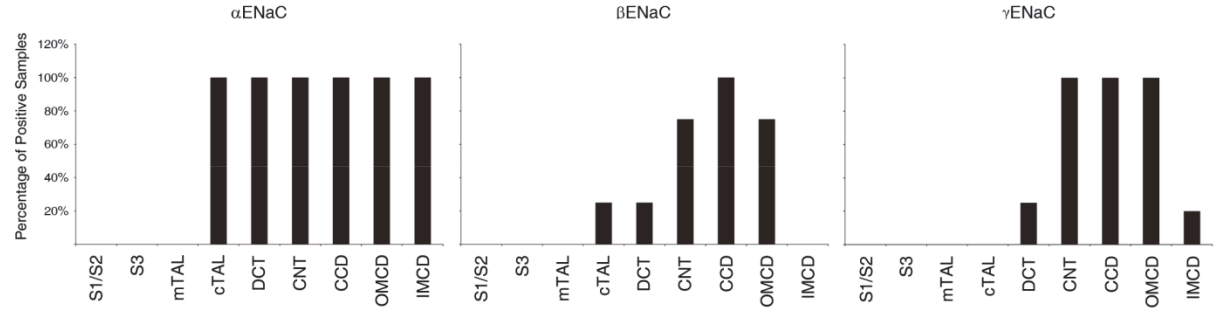
Supplemental Figure 2. Mapped RNA-seq reads for vitamin D receptor (*Vdr*). Note that most reads mapped to sites downstream from the RefSeq-annotated 3'-end (blue arrow). The distal end of this region matches the polyadenylation peak (red arrow) as called by PA-seq. These additional reads were included in the calculation of RPKMs (Datasets S3 and S4).

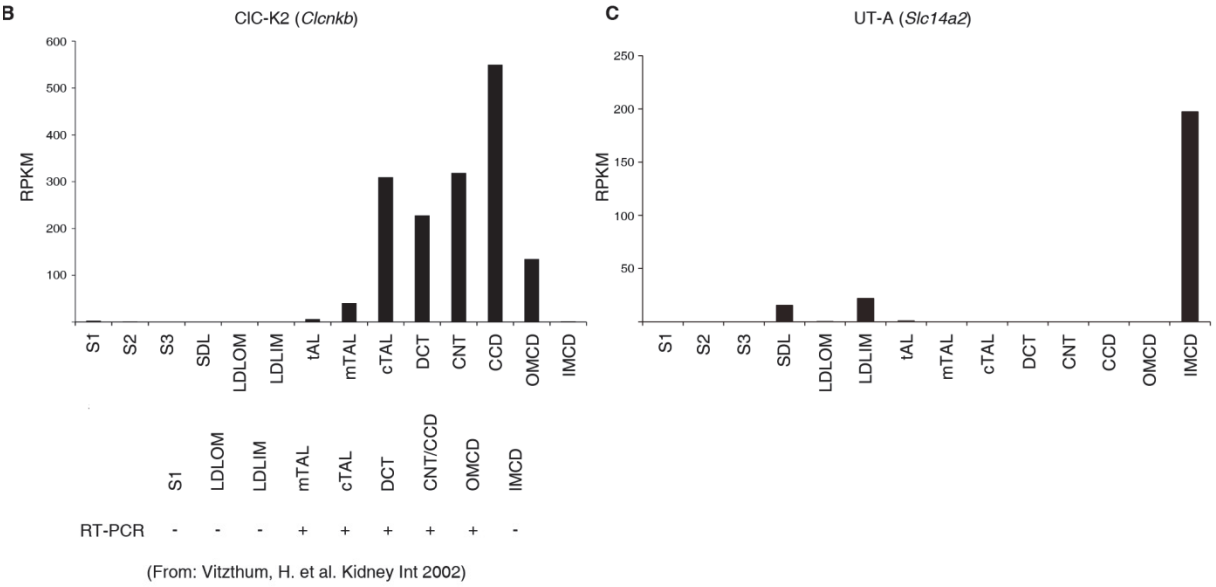


Supplemental Figure 3. A heatmap for all replicates of dissected renal tubule segments. All replicates are represented in columns, and the top 300 genes with highest variance in log_2RPKM were selected to draw this heatmap. Replicates from the same tubule segment clustered more closely with each other than with those from other segments. Replicates from 14 different renal tubule segments were found to form 6 distinct clusters according to their anatomical organization (shown by horizontal bars at the bottom of the heatmap).

A

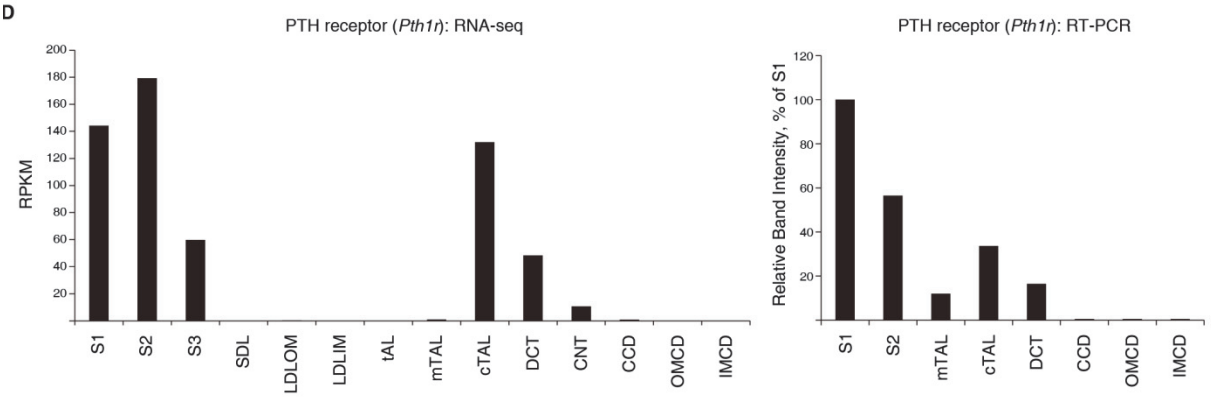
	Gene symbol	S1	S2	S3	SDL	LDLOM	LDLIM	tAL	mTAL	cTAL	DCT	CNT	CCD	OMCD	IMCD
αENaC	*Scnn1a*	0.5	0.2	0.1	0.1	0.1	0.1	0.1	10.8	39.2	68.9	88.2	104.2	47.2	0
βENaC	*Scnn1b*	0	0.1	0	0	0	0.1	0	0.1	2.8	32	160.1	190.9	110.5	0.1
γENaC	*Scnn1g*	0	0.1	0	0	0	0	0	0	0	0.2	52.9	88.7	61.1	0





	S1	LDLOM	LDLIM	mTAL	cTAL	DCT	CNT/CCD	OMCD	IMCD
RT-PCR	-	-	-	+	+	+	+	+	-

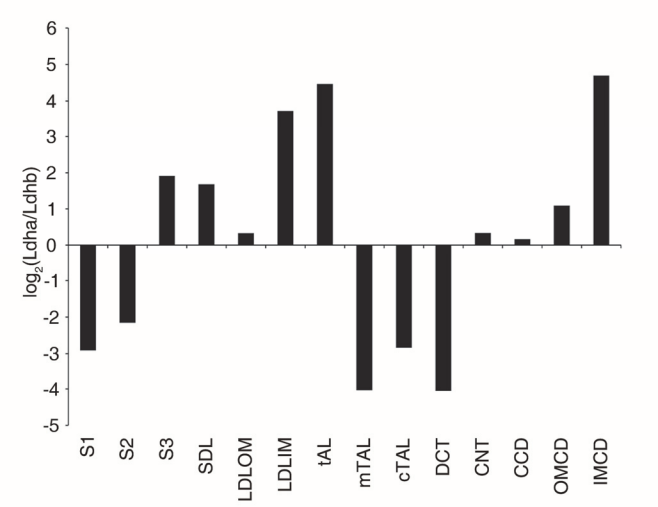
(From: Vitzthum, H. et al. Kidney Int 2002)



Supplemental Figure 4. Comparison of Results from Single-Tubule RNA-seq with RT-PCR Results for Specific Gene Targets. **A.** (top) Median RPKMs for subunits of the

epithelial sodium channel (ENaC). (Bottom) relative expression of ENaC subunits across renal tubule segments. **B.** RT-PCR expression level of chloride channel ClC-K2 (*Clcnkb*). **C.** RT-PCR expression level of urea transporter UT-A (*Slc14a2*). **D.** Comparison of RNA-seq and RT-PCR results for parathyroid receptor (*Pth1r*).

Supplemental Figure 5. A heatmap for expression of transcription factors along the renal tubule segments.



Supplemental Figure 6. Relative expression of two lactate dehydrogenase isoforms, *Ldh* and *Ldhb* along renal tubule.

Supplemental Data

Because of the large size of the datasets, we make the supplemental datasets available via a link to a publicly accessible web server. Please click on the links to access data: https://helixweb.nih.gov/ESBL/Database/NephronRNAseq/Supplemental.html (temporary login ID: clp password: Esbl!@#$)

Supplemental Data 1. Information on microdissected glomeruli and tubule segments and gene expression in each tubule segment. The tab "Microdissected tubule" contains the information on all samples of microdissected tubule segments. "RPKM_withoutPAseq" contains RPKMs for RefSeq transcripts calculated without considering PA-seq data. "RPKM_withPAseq" contains RPKMs for RefSeq transcripts supplemented with PA-seq data. The RNA-seq reads between the annotated 3'-end and the polyadenylation peak were included in the calculation of RPKMs.

Supplemental Data 2. This file contains information on chromosomal coordinates of polyadenylation-site peaks found by PA-seq. Since PA-seq is a strand-specific RNA-seq protocol, peaks on forward and reverse strands are presented on separate tabs.

Supplemental Data 3. This file contains genes that are enriched in the nephron- and collecting-duct segments. The result of DAVID analysis is stored in multiple tabs.

Supplemental Data 4. This file contains genes that are enriched in the renal medullary segments. Medulla-enriched genes and result of DAVID analysis are stored in multiple tabs.