Cell-Type Selective Markers Represented in Whole-Kidney RNA-Seq Data

Jevin Z. Clark, B.S.*, Lihe Chen, PhD*, Chung-Lin Chou, PhD*, Hyun Jun Jung, PhD*,

and Mark A. Knepper, MD, PhD*

*Epithelial Systems Biology Laboratory, Systems Biology Center, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland

Running title: RNA-Seq of Kidney

Corresponding Author:

Mark A. Knepper, MD, PhD

10 CENTER DR, MSC-1603

National Institutes of Health

Bethesda, MD 20892-1603

Phone: 301-496-3064 Fax: 301-402-1443 Email: knep@helix.nih.gov

Supplemental Datasets can be found at
https://hpcwebapps.cit.nih.gov/ESBL/Database/MouseWK/Data/

ABSTRACT

Bulk-tissue RNA-Seq is seeing increasing use in the study of physiological and pathophysiological processes in the kidney. However, the presence of multiple cell types in kidney complicates the interpretation of the data. Here we address the question, “What cell types are represented in whole-kidney RNA-Seq data?” to identify circumstances in which bulk-kidney RNA-Seq can successfully be interpreted. We carried out RNA-Seq in mouse whole kidneys and microdissected proximal S2 segments. To aid in the interpretation of the data, we compiled a database of cell-type selective protein markers for 43 cell types believed to be present in kidney tissue. The whole-kidney RNA-Seq analysis identified transcripts corresponding to 17742 genes, distributed over 5 orders of magnitude of expression level. Markers for all 43 curated cell types were detectable. Analysis of the cellular makeup of a mouse kidney, calculated from published literature, suggests that proximal tubule cells likely account for more than half of the mRNA in a kidney. Comparison of RNA-Seq data from microdissected proximal tubules with whole-kidney data supports this view. RNA-Seq data for cell-type selective markers in bulk-kidney samples provide a valid means to identify changes in minority-cell abundances in kidney tissue. Although proximal tubules make up a substantial fraction of whole-kidney samples, changes in proximal tubule gene expression could be obscured by the presence of mRNA from other cell types.

Keywords: transcriptome; bulk-tissue; proximal tubule

INTRODUCTION

RNA-Seq is a method for identifying and quantifying all mRNA species in a sample. Like RT-PCR, the first step of RNA-Seq is reverse transcription of all mRNAs to give corresponding cDNAs.[1,2] However, unlike RT-PCR, which amplifies only one cDNA target, RNA-Seq amplifies all cDNAs in the sample through use of adaptors that are ligated to the ends of each cDNA.[3] The read-out for RNA-Seq employs next-generation DNA sequencers to identify specific sequences that map to each mRNA transcript coded by the genome of a particular species (the 'transcriptome'). This allows counting of the number of sequence 'reads' for each transcript as a measure of the total amount of each transcript in the original sample. So, RNA-Seq can be viewed like quantitative RT-PCR, but more expansive and unbiased.[1] The abundance of a given transcript is assumed to be proportional to the number of independent sequence 'reads' normalized to the annotated exon length of each individual gene and to the total reads obtained for a sample. This calculation yields transcripts per million or 'TPM' as termed in this paper.[4]

RNA-Seq has seen increased use in recent years, in part because of the ease of execution and the availability of next-generation DNA sequencers.[5] Because of the existence of private-sector biotechnology companies, even small laboratories can successfully carry out RNA-Seq studies in lieu of quantitative RT-PCR. Many recent reports using RNA-Seq employ "bulk tissue RNA-Seq" in which complex tissues containing multiple cell types are analyzed. The limitation of this approach is that it is usually impossible to determine which cell types in the mixture are responsible for observed changes in mRNA abundances. Furthermore, strong responses in minority cell types may be masked by a lack of response in more abundant cell types.[6]

A solution to this problem in kidney is to isolate specific cell types using renal tubule micro-dissection prior to small sample RNA-Seq as described by Lee et al.[7,8] All 14 renal tubule segments plus glomeruli have been profiled in this way. In structures that contain more than one cell type, transcriptomes of each cell type can be determined using single-cell RNA seq.[9]

However, RNA-Seq in single tubules or single cells is not always feasible, e.g. in pathophysiological models. In this context, we ask the question, “Despite the existence of multiple cell types in bulk kidney tissue samples, what information about specific cell types can be gleaned from whole-kidney RNA-Seq?”

RESULTS

What mRNA species are detectable in whole kidney RNA-Seq analysis? We carried out RNA-Seq analysis in three whole-kidney samples from untreated 2-month-old male C57BL/6 mice. Figure 1 shows the distribution of all reads obtained in the three samples between 'uniquely mapped reads', 'multiple mapped reads', and 'unmapped' reads. The percentage of uniquely mapped reads exceeded 85% of the total reads indicating high data quality for all three samples. Total reads for each of the three samples exceeded 66 million reads. Figure 2 shows the reads that mapped to selected genes expressed over a broad range of TPM levels. It can be seen that faithful, selective mapping to exons was obtained down to a TPM value of about 0.15 in this study, or an expression rank of 17742. For example, the reads for Oxtr, coding for the oxytocin receptor (TPM=0.15), thought to be expressed selectively in macula densa cells, [10] are clearly mapped only to exons of the Oxtr gene indicating the specificity of the measurement for spliced Oxtr mRNA (see Supplemental Dataset 2 for mapping of reads for other transcripts with TPM around 0.15). In contrast, exon-specific mapping is ambiguous for Epo, the transcript that codes for erythropoietin (TPM=0.09). Overall, we conclude that 17742 transcripts out of approximately 21000 protein-coding genes in the genome can be detected and quantified in whole kidney samples with the technical approach used here. The whole-kidney TPM values for all transcripts down to rank 17742 are presented at a publicly accessible webpage (https://hpcwebapps.cit.nih.gov/ESBL/Database/MouseWK/index.html, temporary login is ***clp***; temporary password is ***Esbl!@#$***) and as Supplemental Dataset 3. Mapping of whole kidney RNA-Seq reads on a genome browser can be viewed by clicking on "UCSC Genome Browser" at this site. Since the data in this paper were obtained exclusively from 2-month-old male C57BL/6 mice, the reader is cautioned about possible differences that may occur on the basis of gender, age, mouse strain, animal species, food intake, etc. Further studies will be needed to identify the effects of these variables.

What cell types are represented in whole kidney RNA-Seq data? Based on a variety of data types (Methods), we curated a list of 43 cell types that are thought to exist in the kidney and representative protein markers that have been claimed to be specific to or selective for these cell types. The cell types, the markers and whole kidney TPM values for mRNAs corresponding to the markers are presented in Supplemental Dataset 1 and at a permanent, publicly available webpage (https://hpcwebapps.cit.nih.gov/ESBL/Database/MouseWK/WKMarkers.html, temporary login is ***clp***; temporary password is ***Esbl!@#$***). Selected values are presented in Tables 1 and 2. Table 1 shows TPM values for selected markers of epithelial cell types and Table 2 shows TPM values for selected markers of non-epithelial cell types. As seen in Table 1, markers for each epithelial cell type are highly expressed with the exception of macula densa cells. The TPM values for many non-epithelial cell type markers are above the TPM=0.15 threshold defined above (Table 2 and Supplemental Dataset 1). Overall, based on the markers that we have curated, we conclude that mRNAs from all 43 cell types are detectable in whole kidney RNA-Seq samples from mouse. This includes various blood-borne cells, stromal cells and endothelial cells.

How much do various kidney tubule cell types contribute to TPM values? Table 3 shows an accounting of the relative contributions of various renal epithelial cell types to the total makeup of the mouse renal tubule in terms of cell number and protein mass. These estimates were established by integrating several data sources relevant to quantitative renal anatomy in the mouse.[11,12,13,14] Proximal tubule cells account for roughly 45% of the estimated 51 million tubule epithelial cells. However, they account for approximately 66% of total tubule epithelial protein mass, by virtue of their large size compared to other renal tubule cells (Table 3). The second largest contribution is from the thick ascending limb of Henle, contributing 15% of cells and 11% of total protein (Table 3). If mRNA levels parallel protein levels, the contribution of proximal tubules to total mRNA in the renal tubule is also likely to be greater than 50%.

Wiggins et al. have quantified the cell types that make up the glomerulus in rats,[15] yielding a median value of 134 podocytes per glomerulus. If the value is the same in the mouse, each kidney would have 20220 glomeruli per kidney X 134 podocytes per glomerulus = 2.71 X 10^6 podocytes per kidney. This value is about 5% of the total number of tubule epithelial cells (5.1 X 10^7 [Table 4]). Thus, changes in glomerular transcripts are unlikely to be readily detectable in whole-kidney samples, unless they are specific to the glomerulus.

How do proximal tubule S2 segment TPM values correlate with whole kidney TPM values? Because the proximal tubule makes such a large contribution to total epithelial cell number and protein mass, it seems possible that whole kidney RNA-Seq measurements could be used as a surrogate for measurements of transcript levels in the proximal tubule. In order to compare the whole kidney transcriptome with that of the proximal tubule, we carried out RNA-Seq in microdissected S2 proximal tubules, manually dissected from the opposite (left) kidney from the one used for whole kidney RNA-Seq analysis. The S2 proximal data mapped to a total of 18767 genes with mean TPM values greater than 0.1 among the three animals. All of the 12 S2 proximal samples (4 replicates per kidney) had a percent of mapped reads greater than 85, consistent with high data quality (Figure 3). The mean TPM values are provided as a publicly accessible web page at https://hpcwebapps.cit.nih.gov/ESBL/Database/MusRNA-Seq/index.html (temporary login ***clp****;* temporary password ***Esbl!@#$***). This web page also gives TPM values of microdissected mouse cortical thick ascending limbs and cortical collecting ducts, mined from a prior study[9] and compares values from all three segments with the whole kidney RNA-Seq data from this paper. Figure 4A and 4B show plots of the base 2 logarithms of the whole kidney (WK) versus proximal S2 TPM values for housekeeping and nonhousekeeping genes, respectively. The list of housekeeping genes was taken from Lee et al. [7] The ratios for all genes were normalized such that the average WK/S2 TPM ratio is 1 for housekeeping genes that have TPM greater than 1. A tight correlation was seen for housekeeping transcripts (Figure 4A). As expected, WK/S2 ratios varied over a broad range for non-housekeeping transcripts.

The lower bound defines a ratio of about 0.25 and coincides with the location of S2 specific transcripts, e.g. Slc22a7 and Slc22a13, which mediate organic anion and organic cation secretion, respectively, key functions of the S2 segment.[16] This suggests that the S2 segment accounts for at least one quarter of whole kidney mRNA. Kap, a proximal tubule marker expressed in all three subsegments (S1, S2, and S3) is found near the 0.5 ratio line, suggesting that proximal may account for roughly 50% of whole kidney mRNA. However, general proximal tubule markers listed in Supplemental Dataset 4 (Gsta2, Agxt2, Cyp2e1, Cryl1, Glyat, Sord, Pdzk1, Upb1, Sod3, Hnf4a) have a median WK/S2 TPM ratio of 0.85, suggesting that that proximal tubule may account even more than 50% of whole kidney mRNA, assuming that these markers are distributed equally along the proximal tubule (see Supplemental Dataset 4 for distribution among S1, S2, and S3 of rat). 29% of transcripts had WK/S2 ratios greater than 2, signifying transcripts that are relatively excluded from proximal tubule samples, thus more strongly detectable in whole kidney. Some of these are labeled in Figure 4B.

DISCUSSION

In this paper, we asked the question, "What information about specific cell types can be gleaned from whole-kidney RNA-Seq?". To address this, we carried out RNA-Seq analysis of whole mouse kidney samples, yielding a data base of 17742 transcripts with TPM values above a threshold of 0.15, determined from examination of mapped reads for a variety of transcripts spanning TPM values from 0.10 to 621 (see Figure 2 and Supplemental Dataset 3). A full report of TPM values for all 17742 transcripts is given at a publicly accessible website. To identify cell types represented in these data, we compiled a list from literature of selective markers for 43 cell types likely present in kidney tissue. These are listed in Supplemental Dataset 1. (Note that we made no attempt to make the marker list totally comprehensive. Readers are encouraged to look up other transcripts of interest at the website of RNA-Seq data: https://hpcwebapps.cit.nih.gov/ESBL/Database/MouseWK/index.html). We detected markers for all 43 cell types, many of them presumably rare in the overall cell count for the kidney. Thus, even for rare cell types, bulk RNA-Seq data can be used to draw inferences about the abundance of a particular cell type or regulation of its marker. For example, an inflammatory process in the kidney is likely to be associated with increases in markers for macrophages (e.g. Adgre1 [F4/80] or Cd68) in whole kidney RNA-Seq data. Similarly, an increase in mRNA for renin in the kidney may be seen if either the number of afferent arteriolar granular cells increases or when the transcription of the renin gene is increased, both of which have been observed.[17]

Our analysis of the abundances of individual epithelial cell types confirms that proximal tubule cells account for a very large fraction of the total kidney substance, most likely at least 50%. The S2 segment alone appears to account for approximately 25% of whole kidney mRNA (Figure 4B). This raises the question of whether whole kidney measurements suffice to assess changes in the proximal tubule. Clearly, changes in proximal tubule mRNA abundance for a particular gene should be detectable in whole kidney samples, although the magnitude of

changes will be attenuated by dilution by other cell types. The main problem with interpreting whole kidney changes as tantamount to changes in the proximal tubule is that large changes that are specific to other segments would also be manifest in whole kidney samples. Furthermore, changes in the proximal tubule could be masked by opposite changes in other cell types. Consequently, we do not recommend using whole kidney or bulk tissue RNA-Seq as the sole methodology to address hypotheses about the proximal tubule. One approach that may be better in this setting is single-tubule RNA-Seq,[7] in which proximal tubules are first microdissected from the kidney and then subjected to small sample RNA-Seq analysis. In this paper, we present new single-tubule RNA-Seq data on the transcriptome of microdissected S2 proximal straight tubules and present a comparison with the whole-kidney RNA-Seq data. An additional problem with bulk tissue RNA-Seq occurs when small pieces of renal cortex are shaved from the surface of the kidney. There is considerable heterogeneity in cortical structure between the superficial cortex and deeper regions of the cortex,[18] and a lack of precision in sampling from the kidney surface will likely increase sample-to-sample variability.

The compendium of cell-type selective protein markers provided in this paper provides a resource that may be useful to investigators. We caution that the list is not necessarily comprehensive. The list includes multiple markers that have been claimed for certain cell types, many of which were chosen because the protein is present on the cell surface allowing cell sorting. The imprecise definition of the term “cell marker” may lead to uncertainty when interpreting different types of data, thus cell surface markers could be suboptimal for interpretation of RNA-Seq data. Furthermore, many markers have been claimed to be cell-type specific in several cell types, contradicting the specificity claim. In general, we believe that there is a need for a kidney-community oriented effort to define the best cell markers for various uses.

Summary. RNA-Sequencing (RNA-Seq) is seeing increasing use to assess gene expression in kidney. To discover pathophysiological mechanisms in animal models of kidney disease, RNA-Seq is often carried out in bulk kidney tissue, consisting of multiple cell types.

This study analyzes RNA-Seq data from whole mouse kidneys to identify the cell types represented in the data. Markers for 43 different cell types were clearly detectible including all epithelial cell types plus multiple types of vascular cells, stromal cells and bone-marrow derived cells. However, proximal tubule cells appear to account for half or more of total renal mRNA. Despite limitations created by the presence of multiple cell types, bulk-kidney RNA-Seq can be interpretable; particularly when changes in cell-type specific markers are observed.

METHODS

Animals. 2-month-old male C57BL/6 mice (Taconic, Hudson, NY) were maintained in standard conditions with free access to food and water. All animal experiments were conducted in accordance with NIH animal protocol H-0047R4.

Microdissection. Mice were euthanized by cervical dislocation. The right kidney was removed and immediately transferred to Trizol reagent for RNA extraction. The left kidney was placed in ice-cold dissection solution (135 mM NaCl, 1 mM Na_2HPO_4, 1.2 mM $MgSO_4$, 5 mM KCl, 2 mM $CaCl_2$, 5.5 mM glucose, 5 mM HEPES, 5mM Na acetate, 6mM alanine, 1mM trisodium citrate, 4mM glycine, 1mM heptanoate, pH 7.4) for microdissection. Proximal tubules were manually dissected in ice-cold dissection solution without protease treatment under a Wild M8 dissection stereomicroscope equipped with on-stage cooling. After a thorough wash in ice-cold PBS (2 times), the microdissected tubules were transferred to Trizol reagent for RNA extraction. 1 to 4 tubules were collected for each sample.

Whole kidney RNA-Seq and single-tubule RNA-Seq. These steps were conducted as previously reported.[9] Briefly, total RNA from whole kidney and microdissected proximal tubules were extracted using Direct-zol RNA MicroPrep kit (Zymo Research, Irvine, CA) and cDNA was generated by SMARTer V4 Ultra Low RNA kit (Clontech, Mountain View, CA) according to the manufacturer's protocols. 1 ng cDNA was fragmented and barcoded using Nextera XT DNA Sample Preparation Kit (Illumina, San Diego, CA). Libraries were generated by PCR amplification, purified by AmPure XP magnetic beads, and quantified using a Qubit 2.0 Fluorometer. Library size distribution was determined using an Agilent 2100 bioanalyzer with a High-Sensitive DNA Kit (Agilent, Wilmington, DE). Libraries were pooled and sequenced (paired-end 50bp) on Illumina Hiseq 3000 platform to an average depth of 60 million reads per sample.

Data processing and transcript abundance quantification. Data processing was performed as previously reported.[9] Briefly, raw sequencing reads were processed by FASTQC

(https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and aligned by STAR[19] to the mouse Ensembl genome (Ensembl, GRCm38.p5) with Ensembl annotation (Mus_musculus.GRCm38.83.gtf). Unique genomic alignment was processed for alignment visualization on the UCSC Genome Browser. Transcript abundances were quantified using RSEM[4] in the units of transcripts per million (TPM). Unless otherwise specified, the calculations were done on the NIH Biowulf High-Performance Computing platform.

Whole kidney and proximal tubule transcriptomes. The mean TPM values were calculated across all samples: 3 mice, (whole kidney, n=3) and (S2 proximal tubule, n=12). These filtered data are reported on specialized publicly accessible, permanent web pages to provide a community resource: https://hpcwebapps.cit.nih.gov/ESBL/Database/MusRNA-Seq/index.html. (Temporary login is “clp” and temporary password is “Esbl!@#$”).

Data deposition. The FASTQ sequences and metadata reported in this paper have been deposited in NCBI’s Gene Expression Omnibus (GEO) database, (accession number: GSE111837; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111837, secure token: crqzssqurzkbzsp).

Curation of list of cell-type selective genes. To identify a list of cell-type selective genes from renal tubule segments, we used data from microdissected rat renal tubules published by Lee et al.[7] as well as data from mouse microdissected tubules and single cells described by Chen et al.[9] For other cell types, markers were determined using a combination of the following sources: general PubMed searches for publicly accessible research articles, commercial information sources for recommended marker antibodies, and general reference textbooks. Specific sources are given in Supplementary Dataset 1. The curated list was designed to be representative but not exhaustive.

Disclosure

There are no conflicts of interest to disclose.

References

1. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621-628.
2. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009;10(1):57-63.
3. Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol.* 2000;18(6):630-634.
4. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
5. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46.
6. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: from vision to reality. *Nature.* 2017;550(7677):451-453.
7. Lee JW, Chou CL, Knepper MA. Deep Sequencing in Microdissected Renal Tubules Identifies Nephron Segment-Specific Transcriptomes. *J Am Soc Nephrol.* 2015;26(11):2669-2677.
8. Lee JW, Alsady M, Chou CL, et al. Single-tubule RNA-Seq uncovers signaling mechanisms that defend against hyponatremia in SIADH. *Kidney Int.* 2018;93(1):128-146.
9. Chen L, Lee JW, Chou CL, et al. Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. *Proc Natl Acad Sci U S A.* 2017;114(46):E9989-E9998.
10. Stoeckel ME, Freund-Mercier MJ. Autoradiographic demonstration of oxytocin-binding sites in the macula densa. *Am J Physiol.* 1989;257(2 Pt 2):F310-314.

11. Murawski IJ, Maina RW, Gupta IR. The relationship between nephron number, kidney size and body weight in two inbred mouse strains. *Organogenesis.* 2010;6(3):189-194.

12. Garg LC, Knepper MA, Burg MB. Mineralocorticoid effects on Na-K-ATPase in individual nephron segments. *Am J Physiol.* 1981;240(6):F536-544.

13. Vandewalle A, Wirthensohn G, Heidrich HG, Guder WG. Distribution of hexokinase and phosphoenolpyruvate carboxykinase along the rabbit nephron. *Am J Physiol.* 1981;240(6):F492-500.

14. Sperber I. Studios on the Mammalian kidney. *Zoologiska Bidrag fran Uppsala.* 1944;22:pp. 249-432.

15. Nishizono R, Kikuchi M, Wang SQ, et al. FSGS as an Adaptive Response to Growth-Induced Podocyte Stress. *J Am Soc Nephrol.* 2017;28(10):2931-2945.

16. Woodhall PB, Tisher CC, Simonton CA, Robinson RR. Relationship between para-aminohippurate secretion and cellular morphology in rabbit proximal tubules. *J Clin Invest.* 1978;61(5):1320-1329.

17. Taugner R, Hackenthal E, Nobiling R, Harlacher M, Reb G. The distribution of renin in the different segments of the renal arterial tree: immunocytochemical investigation in the mouse kidney. *Histochemistry.* 1981;73(1):75-88.

18. Seldin DW, Giebisch GH. *The kidney : physiology and pathophysiology.* 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2000.

19. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.

20. Knepper MA, Danielson RA, Saidel GM, Post RS. Quantitative analysis of renal medullary anatomy in rats and rabbits. *Kidney Int.* 1977;12(5):313-323.

Acknowledgments

The work was primarily funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute (project ZIA-HL001285 and ZIA-HL006129, M.A.K.). Next-generation sequencing was done in the National Heart, Lung and Blood Institute (NHLBI) DNA Sequencing Core Facility (Yuesheng Li, Director).

Author Contributions

L.C., C.-L.C, H.J.J and M.A.K designed research; L.C., C.-L.C, and H.J.J performed research; J.Z.C. and M.A.K analyzed data; L.C., J.Z.C. and M.A.K wrote the paper.

Table 1. Selected markers for renal epithelial cells in mouse whole kidney with corresponding TPM and Rank values. The full marker dataset values are listed in Supplementary Dataset 2.

Cell Type	Gene Symbol	Common Name	TPM	Rank
Podocyte	**Nphs2**	Podocin	53.3	2768
Proximal (S1)	**Slc5a2**	Type 2 Na-glucose cotransporter (SGLT2)	621.2	230
Thin Ascending Limb	**Clcnka**	Chloride channel, voltage sensitive, kidney type A	60.4	2511
Thick Ascending Limb	**Slc12a1**	Type 2 Na-K-2Cl cotransporter (NKCC2)	333.8	470
Macula Densa	**Ptgs2**	Prostaglandin-endoperoxide synthase 2 (COX2)	0.3	16278
Distal Convoluted Tubule	**Slc12a3**	Thiazide-sensitive Na-Cl cotransporter (NCC)	179.1	892
Connecting Tubule	**Calb1**	Calbindin 1	316.1	499
Principal Cell	**Aqp2**	Aquaporin-2	464.1	317
Intercalated Cell, Type A	**Slc4a1**	Chloride-bicarbonate transporter 1 (AE1)	17.4	5905
Intercalated Cell, Type B	**Slc26a4**	Pendrin	39.6	3484
Inner Medullary Collecting Duct Cell	**Slc14a2**	Urea channel, epithelial	20.1	5484
Transitional Epithelium	**Upk1a**	Uroplakin 1a	7.4	8472

Table 2. Selected markers for renal non-epithelial cells in mouse whole kidney with corresponding TPM and Rank values. The full marker dataset values are listed in Supplementary Dataset 2.

Cell Type	Gene Symbol	Common Name	TPM	Rank
Basophil	**Cd69**	Cd69 antigen	0.2	16835
B-Lymphocyte (follicular)	**Cd22**	B-cell receptor	0.2	17615
Dendritic Cell	**Adgre1**	Adhesion G protein-coupled receptor E1 (F4/80)	2.7	11123
Endothelial Cell	**Pecam1**	Platelet/endothelial cell adhesion molecule 1	16.8	6021
Fibroblast	**Pdgfrb**	Platelet derived growth factor receptor, beta	7.8	8347
Granular Cell of Afferent Arteriole	**Ren1**	Renin 1	111.3	1454
Macrophage	**Cd68**	Macrosialin	4.9	9629
Monocyte	**Cd14**	Monocyte differentiation antigen CD14	5.3	9436
Neuronal Cell (Axon Only)	**Stx1a**	Syntaxin 1A (brain)	0.5	14797
Smooth Muscle Cell	**Acta2**	Actin, alpha 2, smooth muscle	40.5	3418
Polymorphonuclear Leukocyte	**Csf3r**	colony stimulating factor 3 receptor (granulocyte)	0.2	16597
T-lymphocyte	**Cd4**	T-cell surface glycoprotein CD4	0.5	14893

Table 3. Contributions of epithelial cell types to whole kidney cell count and mass.

Segment /Cell Type[a]	Tubule Count[bcd]	Tubule Length (mm)[b]	Total Length per Kidney (mm)[e]	Density (cells/mm)[f]	Total Cells per Kidney[e]	Percent of Total Cells[e]	Protein Mass (ng/mm)[g]	Total Protein per Kidney (µg)[e]	Percent of Total Protein[e]
PT (Short looped nephrons)	15367	4.0	61469	280	17211264	33.9	225	13830	50.32
PT (Long looped nephrons)	4853	4.0	19411	280	5435136	10.7	225	4368	15.89
tDL - type 1	19465	0.6	11679	100	1167907	2.3	37	432	1.57
tDL - type 2	4853	0.6	2912	150	436752	0.9	37	108	0.39
tDL - type 3	4853	1.2	5823	125	727920	1.4	37	215	0.78
tAL	4853	1.2	5823	150	873504	1.7	37	215	0.78
mTAL	20220	1.2	24264	311	7546104	14.9	125	3033	11.03
cTAL	20220	0.7	13143	275	3614325	7.1	79	1038	3.78
DCT	20220	0.5	10110	374	3781140	7.5	159	1607	5.85
CNT Cell	20220	0.5	10110	263	2656908	5.2	76	764	2.78
CNT A-IC	20220	0.5	10110	44	442818	0.9	13	127	0.46
CNT B-IC	20220	0.5	10110	131	1328454	2.6	38	382	1.39
CCD PC	3370	1.3	4381	298	1306064	2.6	63	277	1.01
CCD B-IC	3370	1.3	4381	113	495403	1.0	24	105	0.38
CCD A-IC	3370	1.3	4381	103	450367	0.9	22	96	0.35
OMCD - PC	3370	1.2	4044	300	1213200	2.4	76	308	1.12
OMCD - A-IC	3370	1.2	4044	188	760272	1.5	48	193	0.70
OMCD - B-IC	3370	1.2	4044	12	48528	0.1	3	12	0.04
IMCD	variable	3.7	2494	500	1246900	2.5	150	374	1.36
Sum	**--**	**--**	**2.1×10^5**	**--**	**5.1×10^7**	**100**	**--**	**2.7×10^4**	**100**

a abbreviation, definition; A-IC, Alpha Intercalated Cells; B-IC, Beta Intercalated Cells; PC, Primary Cells; PT, Proximal Tubule; tDL, thin descending limb; tAL, thin ascending limb of the loop of Henle; mTAL, medullary thick ascending limb of the loop of Henle; cTAL, cortical thick ascending limb of the loop of Henle; DCT, distal convoluted tubule; CNT, connecting tubule; CCD, cortical collecting duct; OMCD, outer medullary collecting duct; IMC, inner medullary collecting duct

b Sperber [14]

c Murawski [11]

d Knepper [20]

e Calculated from data in this table

f Garg [12]

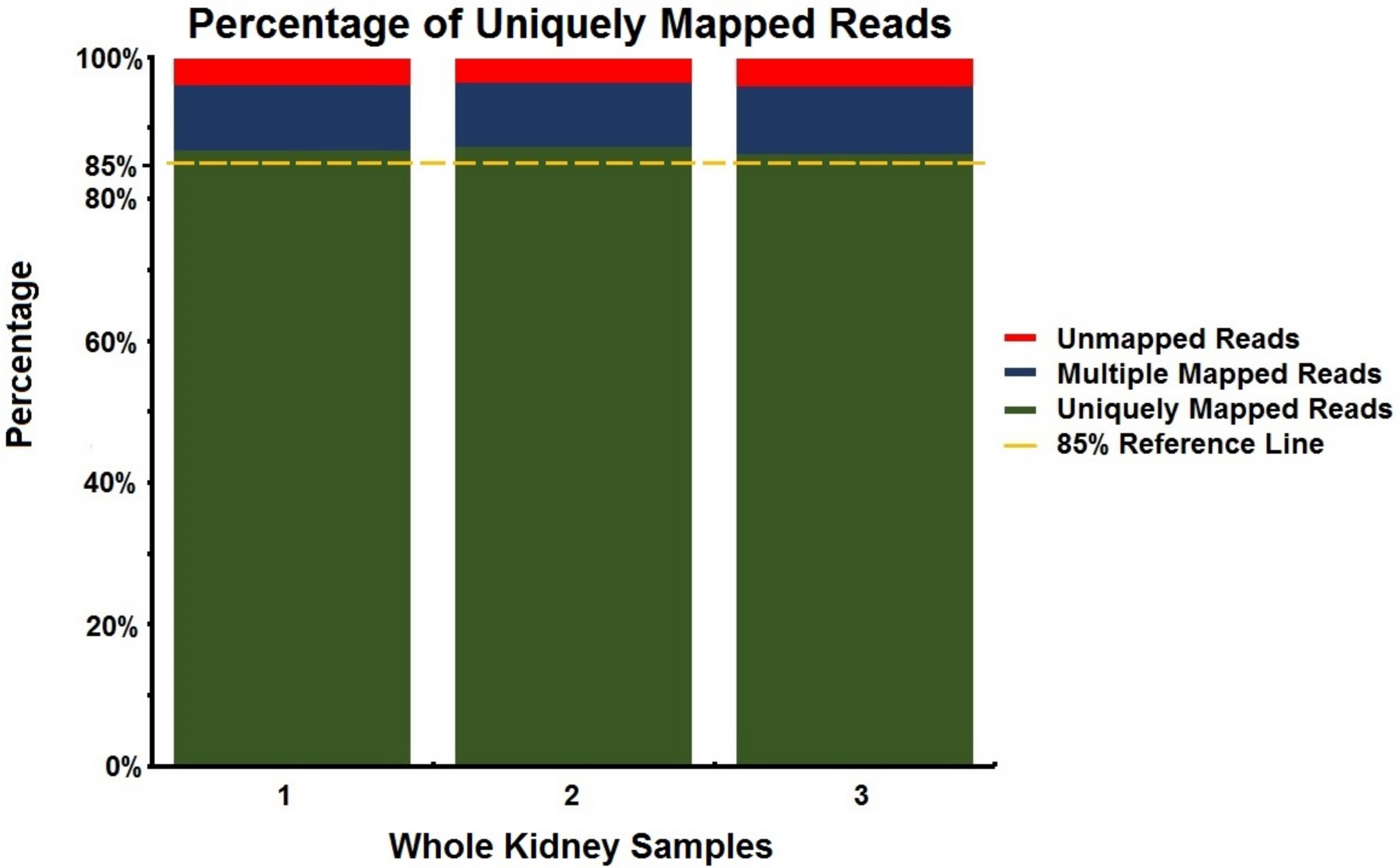
g Guder [13]

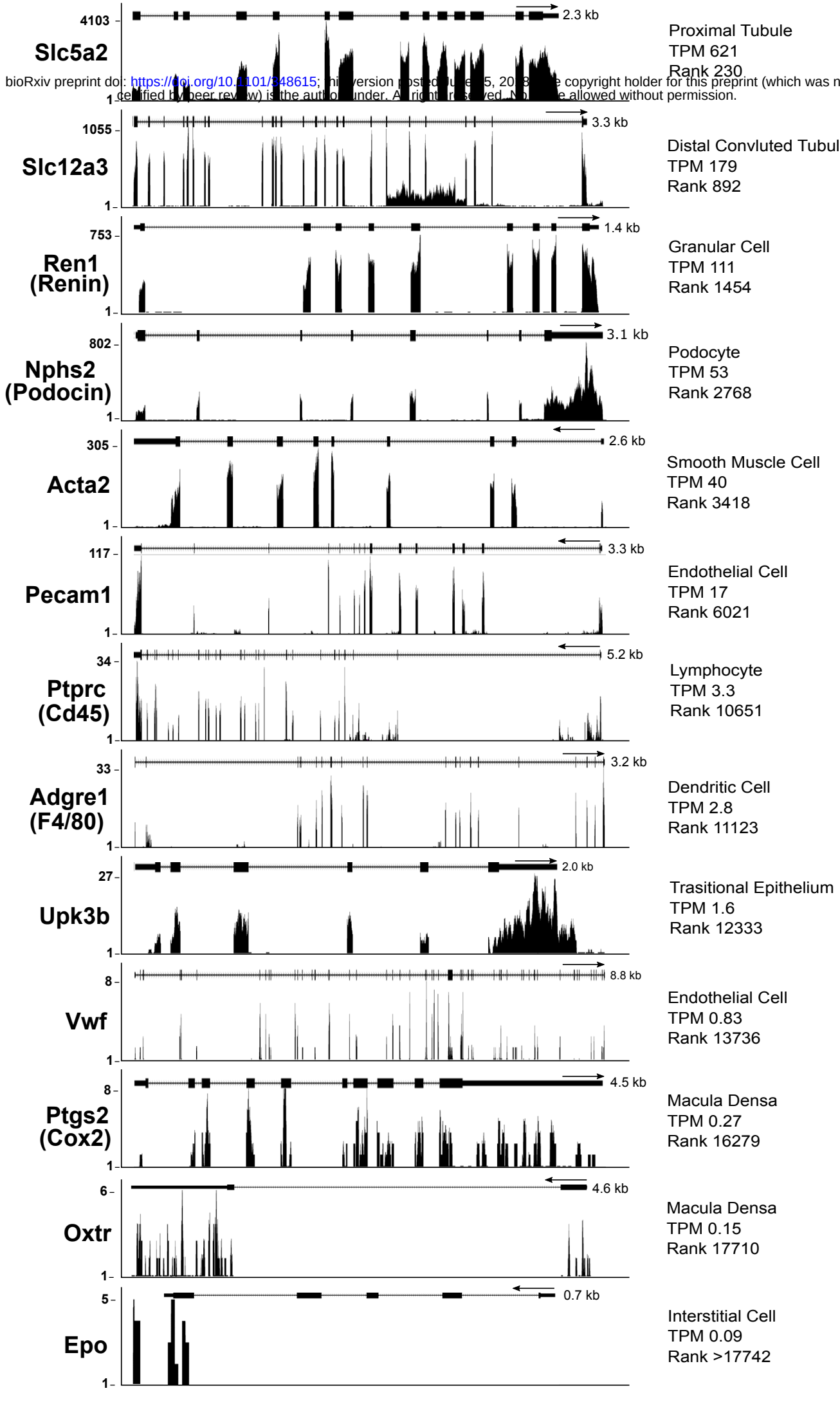
Figure 1. Mapping quality of the whole kidney RNA-Seq data. Distribution of reads shows that uniquely mapped reads exceeds 85% of total reads in all three whole kidney samples. Total reads were: sample 1, 66142467; sample 2, 68482027; sample 3, 69079531.

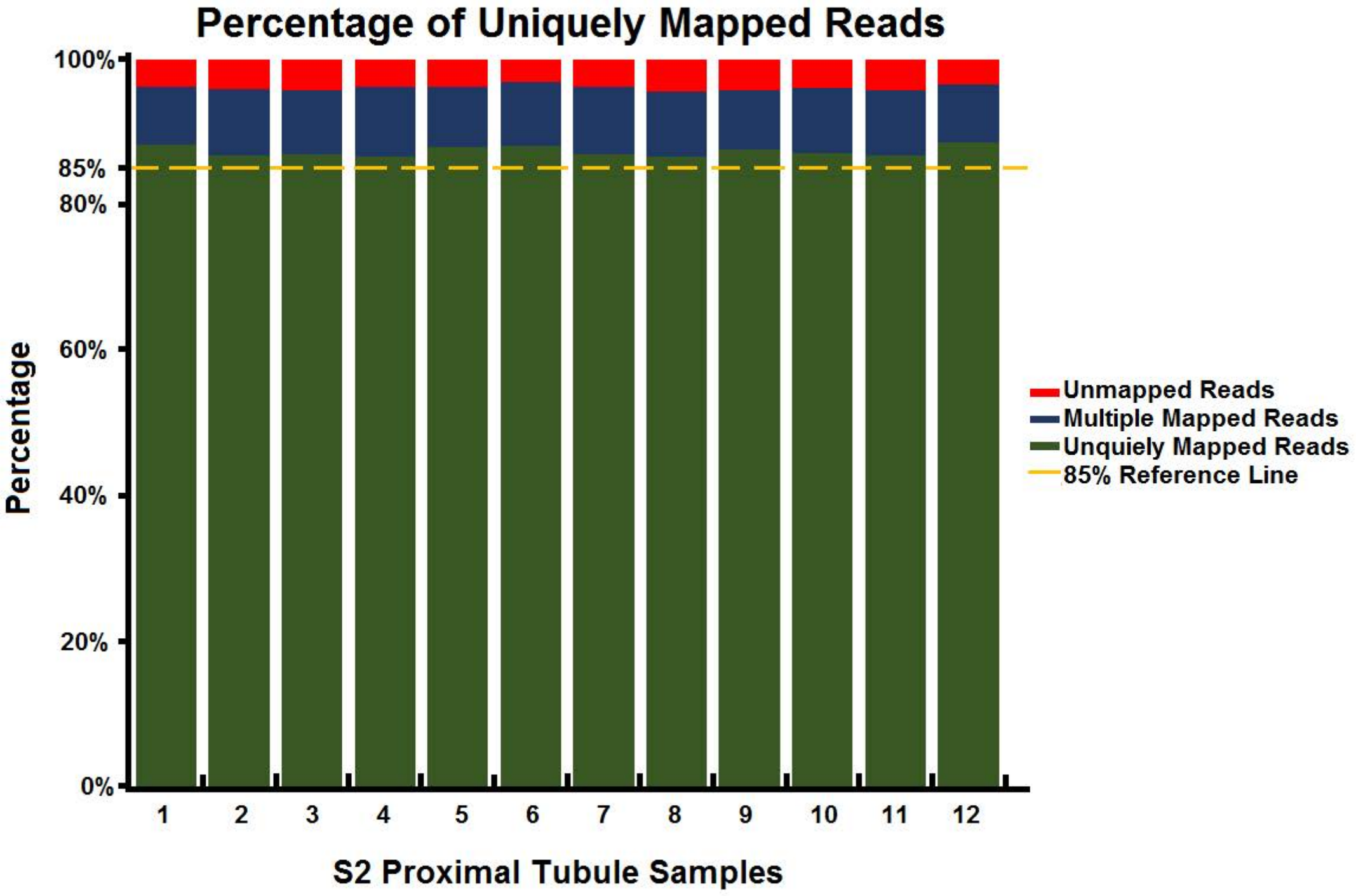
Figure 2. Visualization of the RNA-Seq reads for representative transcripts. Cell-type selective genes from indicated cell types with their mRNA length, TPM, and Rank values. Genes with TPM greater than 0.15 are within a confident detectable range. Data were visualized in the UCSC Genome Browser. Vertical axis shows read counts. Map of exon/intron organization of each gene is shown on top of individual panels.

Figure 3. Mapping quality of the microdissected proximal tubule (S2) RNA-Seq data. Distribution of reads shows that uniquely mapped reads exceeds 85% of total reads in all twelve S2 proximal tubule samples. Total reads were: sample 1, 69808466; sample 2, 84962667; sample 3, 75565121; sample 4, 74862689; sample 5, 76598350; sample 6, 78381995; sample 7, 70858077; sample 8, 77120838; sample 9, 64935558; sample 10, 69894298; sample 11, 70091668; sample 12, 67011247.

Figure 4. Correlation between whole kidney RNA-seq and microdissected proximal tubule (S2) RNA-seq. (A) Housekeeping Genes were plotted between whole kidney RNA-Seq and microdissected proximal tubule (S2) RNA-Seq. (B) Nonhousekeeping Genes were plotted between whole kidney RNA-Seq and microdissected proximal tubule (S2) RNA-Seq. The dashed lines represent the whole-kidney versus proximal tubule (S2) ratios. Each dot is an individual transcript with TPM greater than 0.15. Data are $\log_2$-transformed and normalized before plotting (Methods).


Percentage of Uniquely Mapped Reads
100%
85%
80%
60%
40%
20%
0%
Percentage
1
2
3
Whole Kidney Samples
Unmapped Reads
Multiple Mapped Reads
Uniquely Mapped Reads
85% Reference Line




Percentage of Uniquely Mapped Reads
100%
85%
80%
60%
40%
20%
0%
Percentage
1
2
3
4
5
6
7
8
9
10
11
12
S2 Proximal Tubule Samples
Unmapped Reads
Multiple Mapped Reads
Unquiely Mapped Reads
85% Reference Line