Supplementary information for:
Single cell RNA sequencing reveals mRNA splice isoform switching during kidney development

Yishay Wineberg[1,*], Tali Hana Bar-Lev[1,*], Anna Futorian[2,*], Nissim Ben-Haim[1,*], Leah Armon[2], Debby Ickowicz[2], Sarit Oriel[1], Efrat Bucris[1], Yishai Yehuda[1], Naomi Pode-Shakked[4,5,6], Shlomit Gilad[3], Sima Benjamin[3], Peter Hohenstein[7], Benjamin Dekel[4,5,6]**, Achia Urbach[2,**], and Tomer Kalisky[1,8,**]

[1]*Department of Bioengineering and Bar-Ilan Institute of Nanotechnology and Advanced Materials (BINA), Bar-Ilan University, Ramat Gan, Israel 52900*
[2]*The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel 52900*
[3]*The Nancy and Stephen Grand Israel National Center for Personalized Medicine (G-INCPM), Weizmann Institute of Science, Rehovot, Israel 7610001*
[4]*Pediatric Stem Cell Research Institute, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Hashomer, Israel 52621*
[5]*Division of Pediatric Nephrology, Sheba Medical Center, Tel-Hashomer, Israel 52621*
[6]*Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel*
[7]*Dept. Human Genetics, Leiden University Medical Center, Leiden, The Netherlands*
[*]*Co-first authors*
[**]*Co-senior authors*

[8]**Correspondence to:**
Tomer Kalisky
Department of Bioengineering, Bar-Ilan University, Ramat Gan, Israel 52900
Office: +972-3-738-4656
Email: tomer.kalisky@gmail.com

TABLE OF CONTENTS

SUPPLEMENTARY TEXT

Manual calculation inclusion levels:

A few alternatively spliced cassette/skipped exons (SE) and mutually exclusive exons (MXE) that we found in the literature were not included in our gene model. For these exons we calculated the inclusion levels manually as follows. The inclusion level ψ of an alternatively spliced exon is defined as the fraction of transcripts that include the exon out of the total number of transcripts that either include the exon or skip over it [1,2]. This can be estimated from the density of reads that align to the upstream splice junction, the alternative exon itself, and the downstream splice junction, vs. the density of reads that align to the skipping splice junction that directly connects the upstream exon to the downstream exon. It is also possible to use only reads that span the splice junctions.

We define I to be the number of reads that map to the exon inclusion isoform and S to be the number of reads that map to the exon skipping isoform. Since we are interested in the density of reads, we need to normalize the read counts I and S by the "effective lengths" – the lengths of the isoform-specific segments that they align to. We therefore define l_I as the effective length of the exon-inclusion isoform, that is, the number of the unique positions to which reads can be aligned in the upstream splice junction, the alternative exon itself, and the downstream splice junction. Likewise we define l_S as the effective length of the exon-skipping isoform, that is, the number of unique positions to which reads can be aligned in the skipping splice junction that directly connects the upstream exon to the downstream exon (see Fig. S1 of [1] and supplementary note of [2]).

Given the effective lengths l_I and l_S for each isoform, the inclusion level ψ can now be estimated from read counts I and S as:

$$\hat{\psi} = \frac{(I/l_I)}{(I/l_I) + (S/l_S)}$$

For our manual calculation we counted only reads that span the splice junctions, and approximated the effective lengths as the number of junctions from which each isoform is composed (see Fig. S1 of [1]), that is:

For a **skipped exon**: $l_I = 2$ and $l_S = 1$ - two splice junctions for the inclusion isoform and one splice junction for the skipping isoform.

For **mutually exclusive exons**: $l_I = 2$ and $l_S = 2$ - two splice junctions for the inclusion isoform of the first exon and two splice junctions for the inclusion isoform of the second exon.

Quantifying alternative splicing using DEXSeq:

We used DEXSeq [3] to count the number of reads that align to each exon (or segment of an exon) within each "bulk" *in-silico* transcriptome representing each cell population (Table S11), and then plotted the number of reads that align to selected exons. In order to take into account the different depths of each transcriptome due to the highly varying number of cells in each population, we normalized the exon-specific counts by dividing by the total number of reads that align to the specific gene.

DEXSeq can be used to complement rMATS in validating specific differential expression in specific exons for which the junction coverage is low.

The gene Wt1 undergoes splice isoform switching during kidney development:

We also inspected the Wt1 gene, which is essential for normal kidney development [4–7]. Mutations and alternative splicing in Wt1 were found to play an important role in developmental defects such as Denys-Drash syndrome and Frasier syndrome, as well as in Wilms' tumors. Wt1 is known to encode for multiple splice isoforms [4,8]: for example, three amino acids (K-T-S) at the 3' end of exon 9 may be included, creating a KTS+ isoform, or skipped, resulting in a KTS- isoform (Fig. S20). It was previously found that these isoforms differ in their affinity to DNA [9] and in their localization to different compartments within the nucleus [10]. Moreover, it was found that normal tissues have a KTS+:KTS- ratio of approximately 0.6, while in tissue from patients with Frasier syndrome the amount of KTS+ transcripts decreases, resulting in a lower KTS+:KTS- ratio of approximately 0.4 [5,11,12].

In our dataset we observed a decrease in the KTS+:KTS- isoform ratio, starting from from ~0.75 in the un-induced mesenchyme (UM) and the cap mesenchyme (CM), and converging to ~0.6 in the podocytes (PODO), early epithelial structures (PROX_1), and proximal tubules (PROX_2) (Fig. S20). It was interesting to observe the relative similarity of the KTS+:KTS- ratio in the podocytes and epithelial populations (PODO, PROX_1, and PROX_2), which was lower than the ratio for the mesenchymal populations (UM, CM). This might indicate the existence of a mechanism for stabilizing and tightly regulating this ratio in maturing renal cell populations. Likewise, we also observed a gradual increase in the inclusion levels of cassette exon 5 (Fig. S20).

We note that when inspecting the expression behavior of Wt1 we found multi-level expression differences between the different cell populations (Fig. S20). Wt1 is most highly expressed in the podocytes [13] (PODO), in which it was previously shown to be a key transcriptional regulator [7], moderately expressed in the un-induced mesenchyme (UM, partially) and cap-mesenchyme (CM), and under-expressed in the loop of Henle (LOH) and distal tubules/collecting duct (DIST/CD). In the early epithelial structures (PROX_1) we observed a wide distribution of Wt1 expression, probably due to the fact that some cells (e.g. those in the cleft of the S-shaped body [14]) are in the process of differentiating to podocytes while others are destined to become constituents of the proximal tubule, loop of Henle, or distal tubule.

SUPPLEMENTARY FIGURES

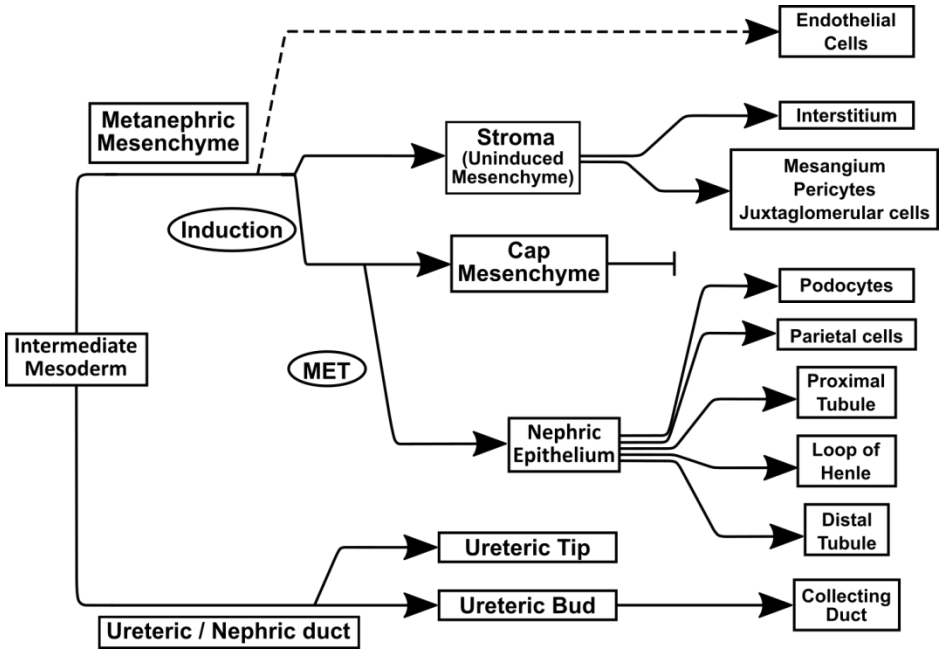


Figure S1: A sketch of the various cell lineages that co-exist in the nephrogenic zone of the developing fetal kidney [15].

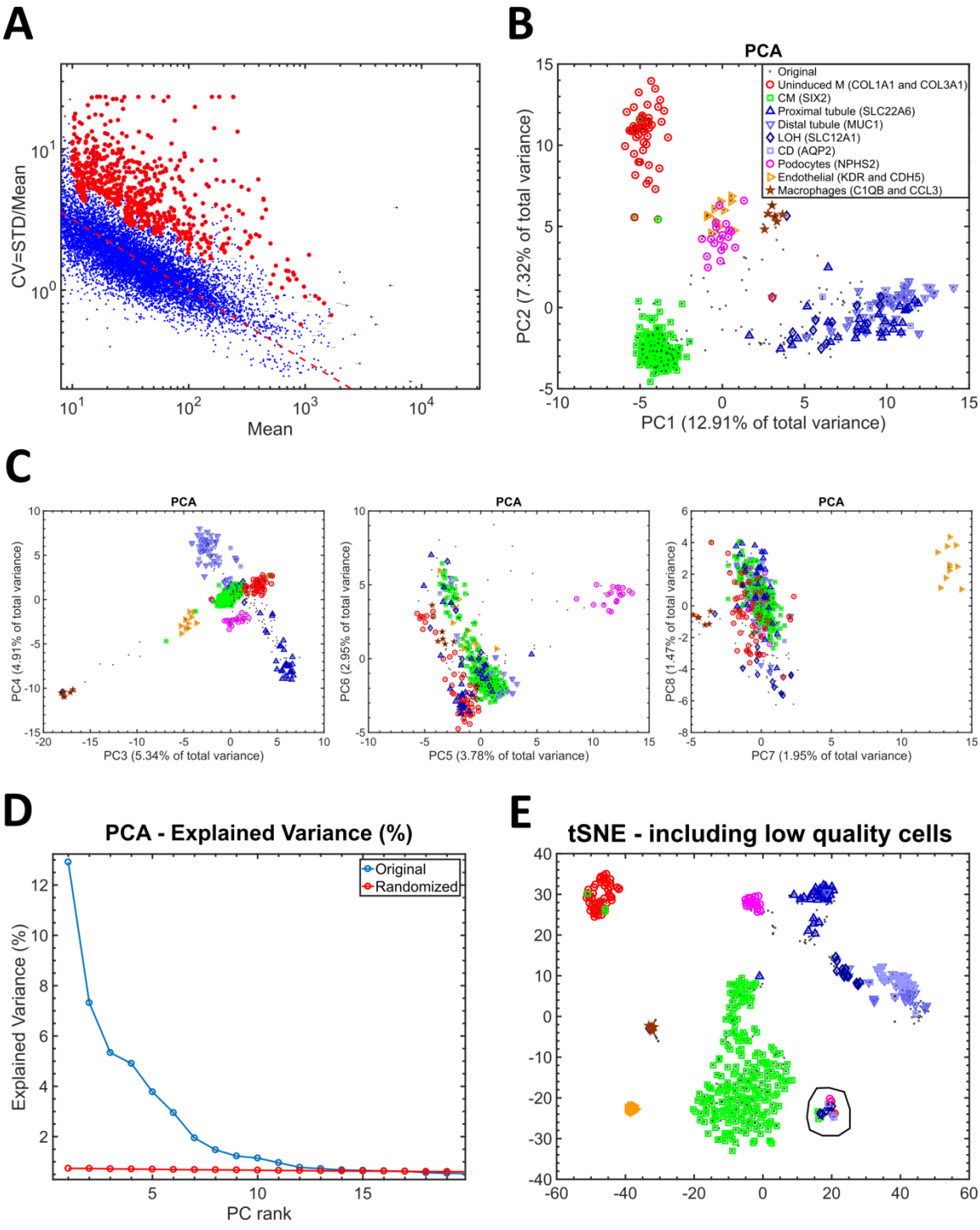


Figure S2: Preprocessing of single-cell data from the Smartseq2 protocol and removing low quality cells. Raw reads from 576 cells (6 x 96-well plates) were aligned to the mouse mm10 genome and the numbers of reads that align to each gene were counted and normalized. We filtered out 11 cells that expressed zero levels of the "housekeeping genes" Gapdh or Actb, resulting in 565 cells. (A) We chose highly variable genes. We followed the method by Macosko et al. [16] to select for genes whose cell-to cell

variance exceeds that which would have resulted from a Poisson distribution. These are actually genes whose cell to cell heterogeneity cannot be attributed to random distribution of transcripts between wells and are therefore likely to be actively over-expressed or under-expressed in the different cell populations. We first removed all non-expressed genes and plotted the mean vs. coefficient of variance (CV=STD/Mean) for each gene (blue dots). Then, we filtered out genes whose normalized counts are less than 10. We divided the mean expression (horizontal axis) into 100 equally sized bins in log10 space, and for each bin we calculated the mean and dispersion (=standard deviation) of the CV (vertical axis). We chose only genes whose CV exceeded the mean CV within their respective bin by at least one standard deviation (red large dots). This actually chooses genes whose variance exceeds those of other genes having a similar mean expression value. This step resulted in 647 highly variable genes, to which we added a list of genes from the literature that were previously shown to be involved in kidney development (Table S2). We also added an additional list of 48 genes from a previous single-cell qPCR study that we previously conducted on human fetal kidney cells [17] (Table S2), which, in retrospect, were not crucial to the identification of the different cell populations. These steps resulted in a gene expression matrix of 677 genes x 565 cells. Each gene was then modified-log-transformed [log2(1+expression)] and standardized by subtracting the mean, dividing by the standard deviation, and truncating to the range [-1,1] (B) Principal Components Analysis (PCA) plot of the resulting matrix consisting of 677 genes vs. 565 cells. The first two principal components (PC's 1-2) show a triangular shaped structure whose vertices correspond to the un-induced mesenchyme (UM), cap mesenchyme (CM), and epithelial populations. (C) Higher principal components (PC's 3-4, 5-6, etc.) capture additional heterogeneity arising from other cell populations such as macrophages, podocytes, and endothelial cells. (D) The explained variance of each principal component. PC's 14 and onwards do not explain more variance than is explained by a randomized matrix, that is, a matrix in which each gene (=row) was randomly permuted. (E) A tSNE plot of 565 single cell profiles, each consisting of 677 highly variable genes. Cells over-expressing genes that were previously shown to mark different populations are marked by additional symbols (as in Fig. 2A in the main text). We identified a population of 21 low quality cells that appears as "mixture" of many cell types (circled). These 21 low quality cells were filtered out, resulting in a 544 cells. The process of selecting for highly variable genes (and adding known genes related to kidney development as described above) was then repeated for the remaining high quality cells, resulting in a matrix of 728 genes x 544 cells, whose analysis is shown in the main text.

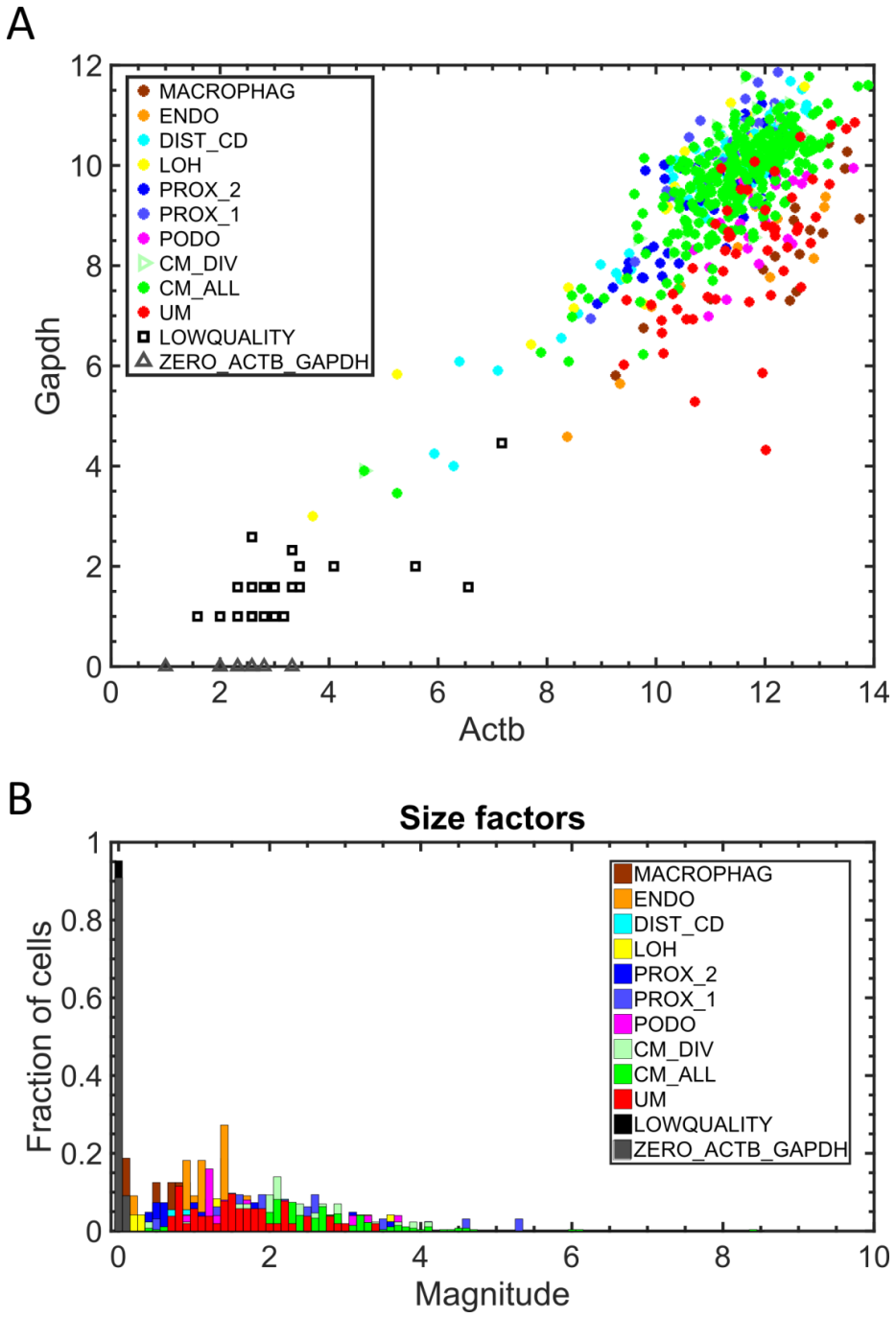


Figure S3: Removing low quality cells. (A) The population of 21 low quality cells (black squares) have low expression levels of the "housekeeping genes" Actb and Gapdh. (B) Likewise, they have low DESeq size factors [18,19]. The annotations "CM_ALL" and "CM" are used interchangeably to represent all cells that were classified as belonging to the cap mesenchyme. "CM_DIV" represents a subset of cells from the cap mesenchyme that are presumably dividing since they over-express the genes Mki67 and Top2a that are known to be over-expressed during the S-G2-M phase of the cell cycle.

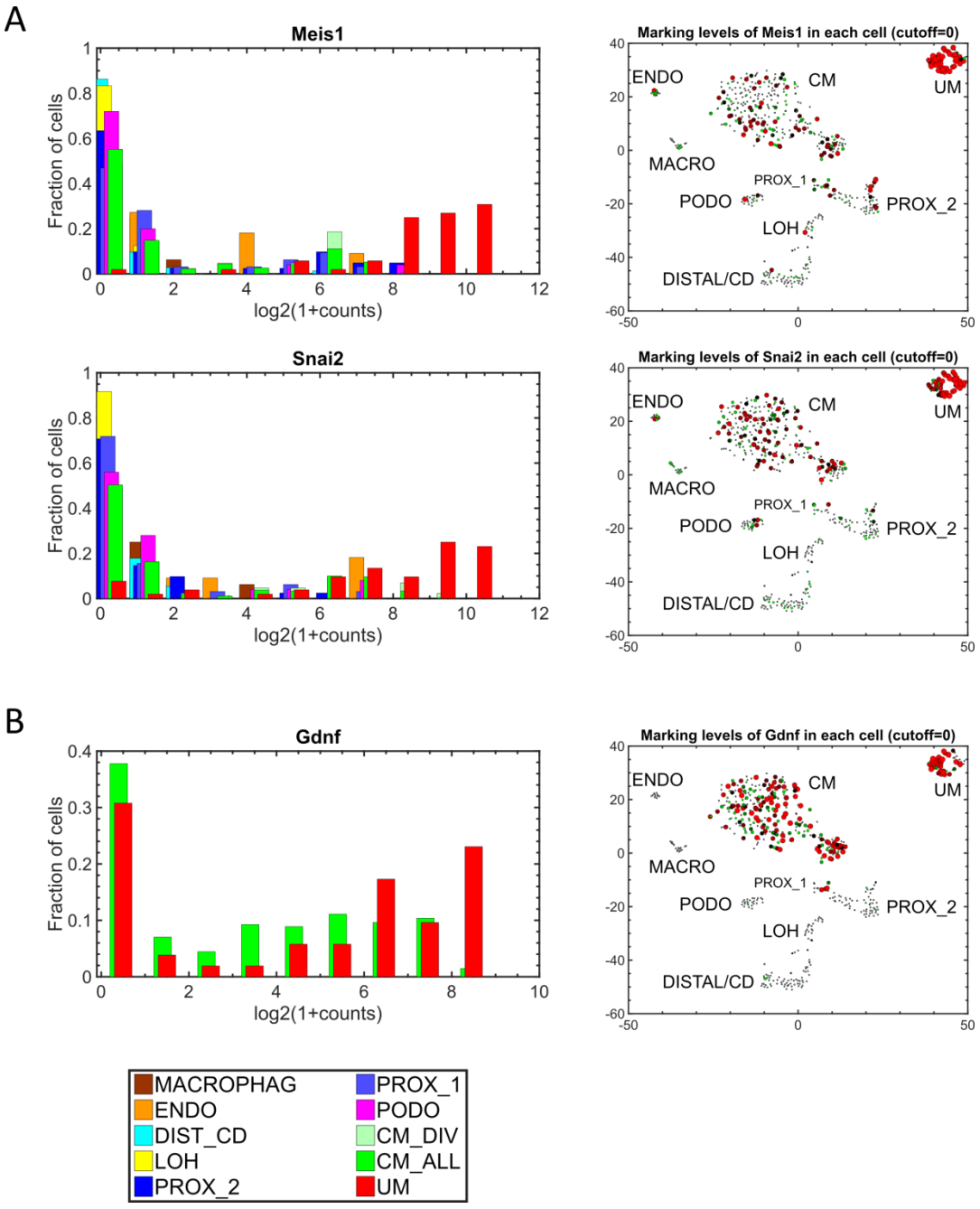


Figure S4: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the un-induced mesenchyme (UM) and stroma. Meis1 [20,21] and Snai2 [22] are more specific to the UM while Gdnf [21,23,24] is expressed in the un-induced mesenchyme (UM) and cap mesenchyme (CM) as previously shown [21].

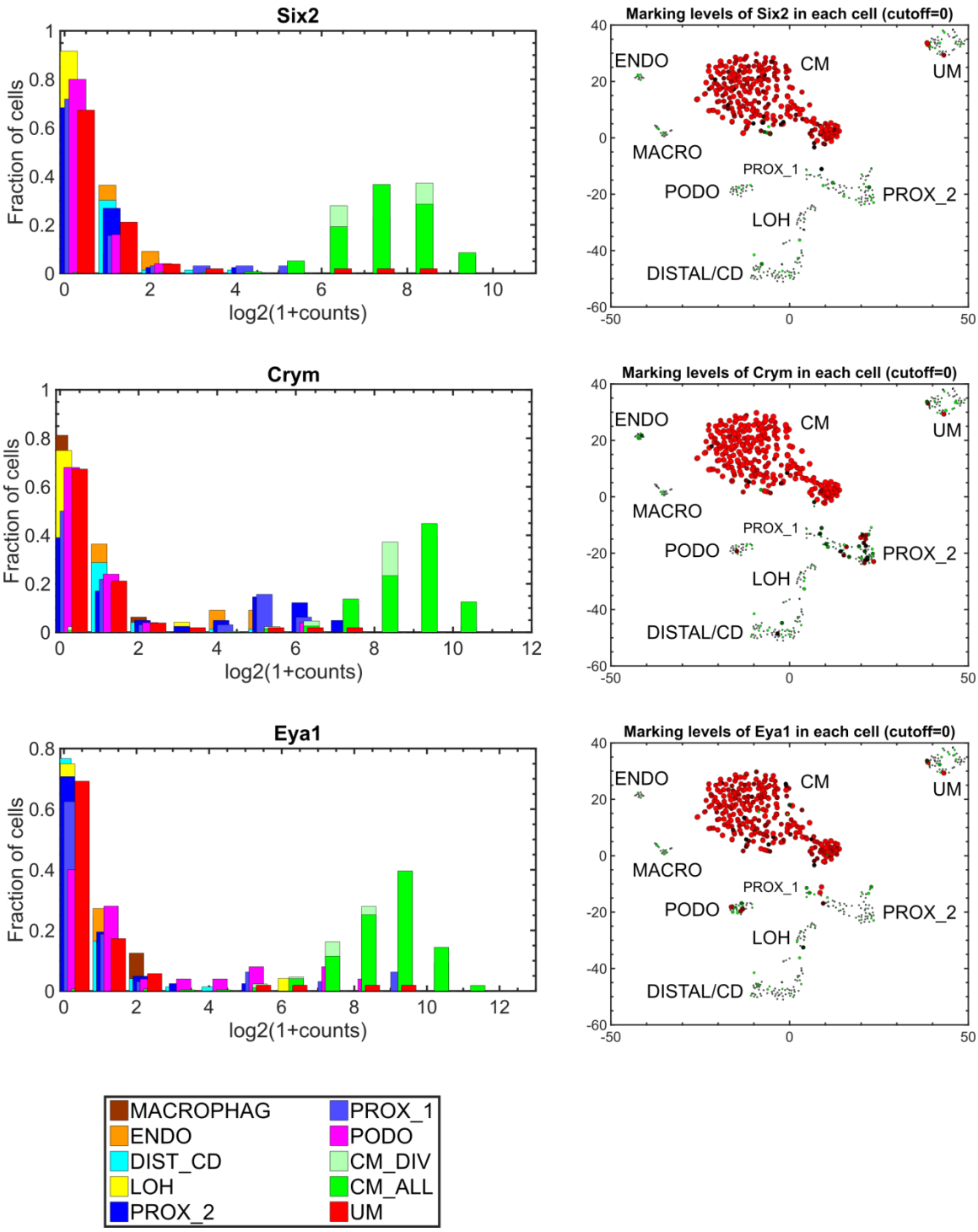


Figure S5: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the cap mesenchyme (CM) [24,25]. Note that Crym is also moderately expressed in the early epithelial structures (PROX_1) and proximal tubules (PROX_2) [21,23].

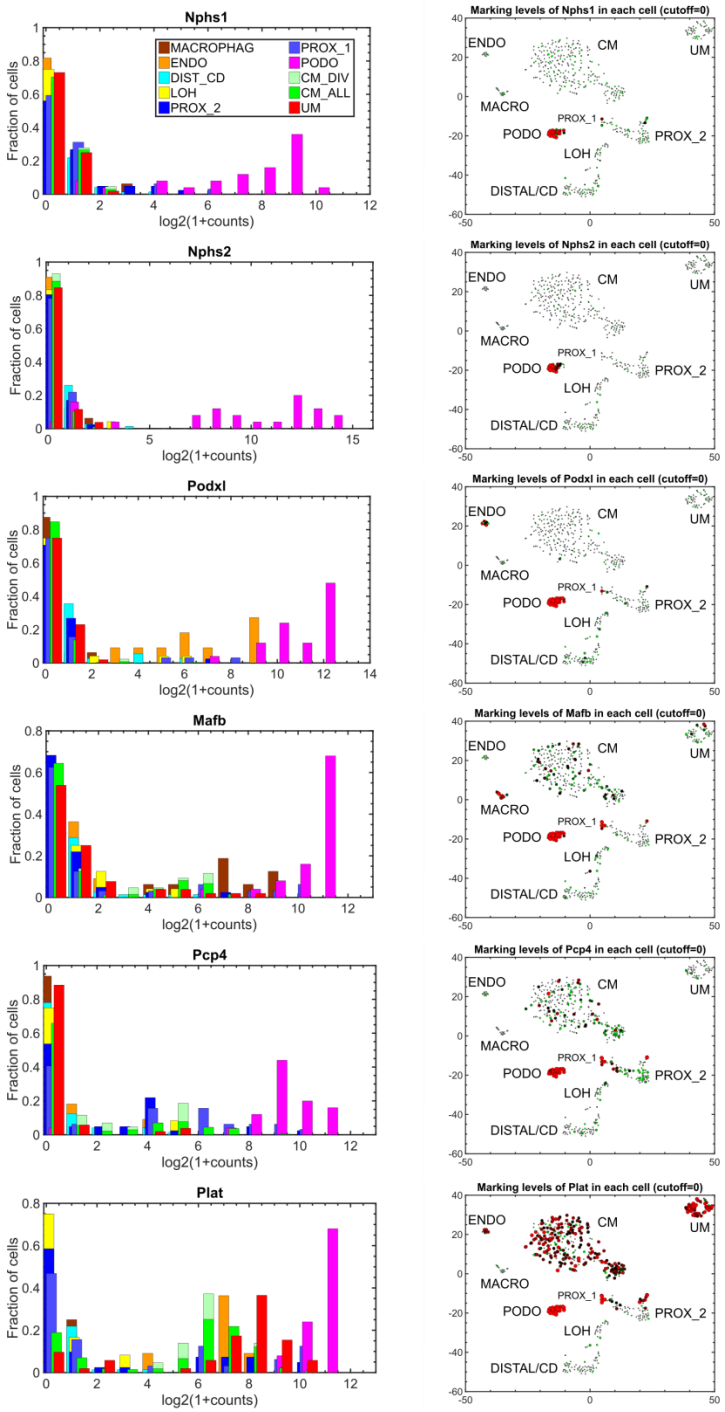


Figure S6: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the podocytes (PODO) [24]. Note that Plat is also moderately expressed in other early developmental lineages (CM, UM, PROX_1) [21,23].

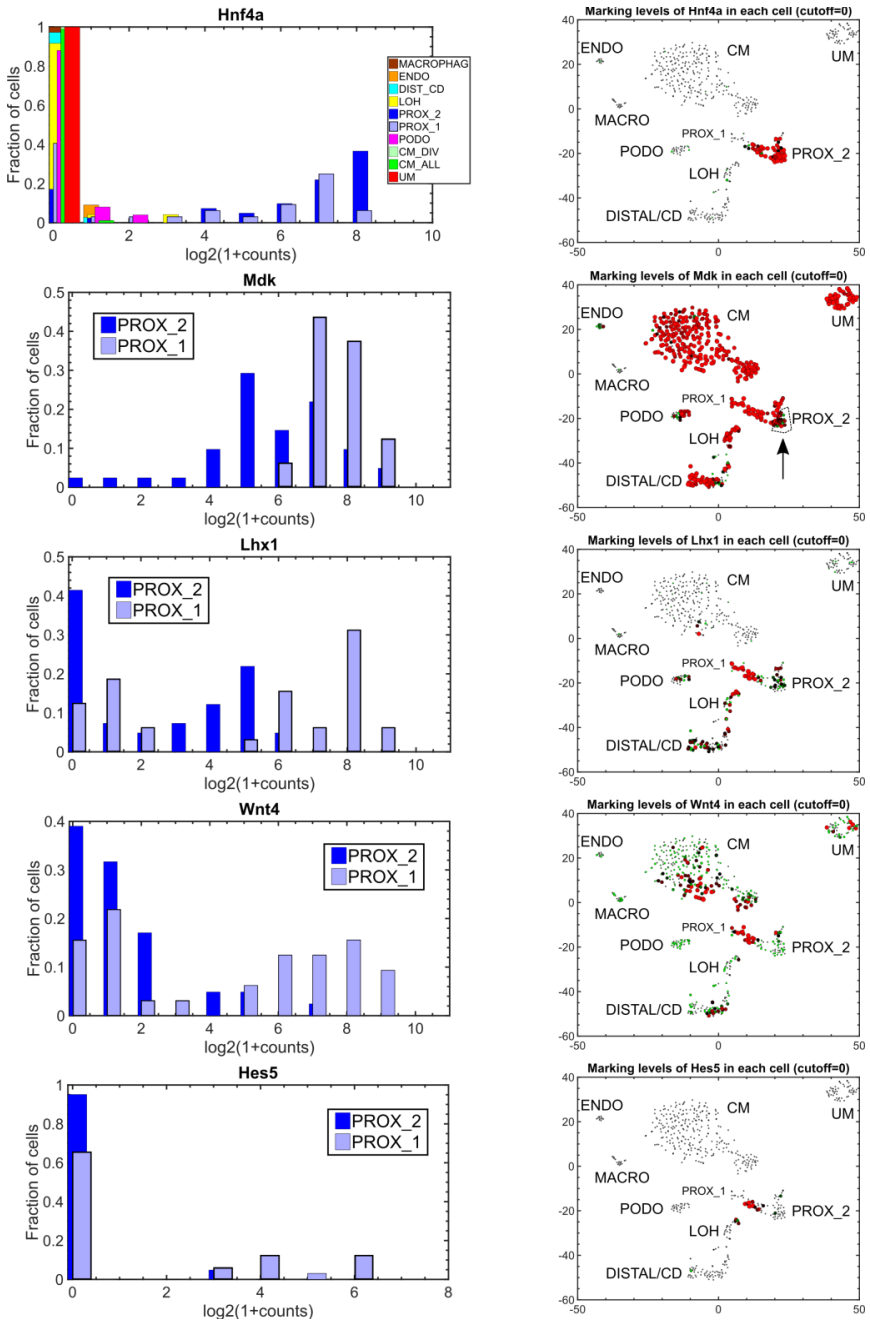


Figure S7: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the early epithelial structures (PROX_1) and proximal tubules (PROX_2) or genes that distinguish between them. Both populations express high levels of the proximal tubular marker Cdh6 [24,26] (Fig. 2D in main text). Hnf4a [20] is expressed in both PROX_1 and PROX_2, while Mdk [20,21], Lhx1 [20,24,27], Wnt4 [24,28,29], and Hes5 [20] are higher in the early epithelial structures (PROX_1) – which are presumably the pre-tubular aggregates, renal vesicles, and C/S-shaped bodies [21,23].

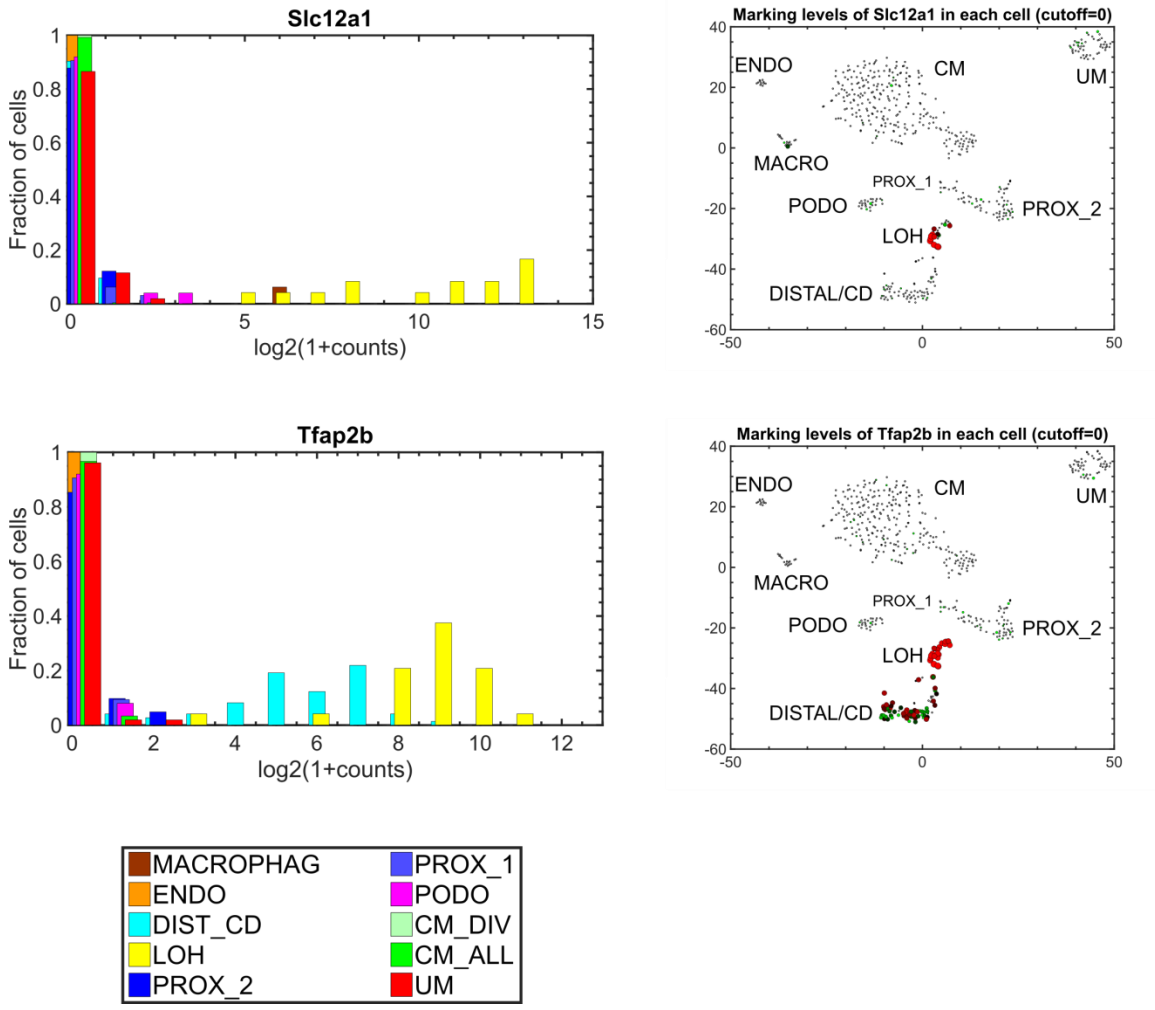


Figure S8: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the loop of Henle (LOH). Note that Tfap2b is also moderately expressed in the distal tubule and collecting duct (DIST_CD) [21,23].

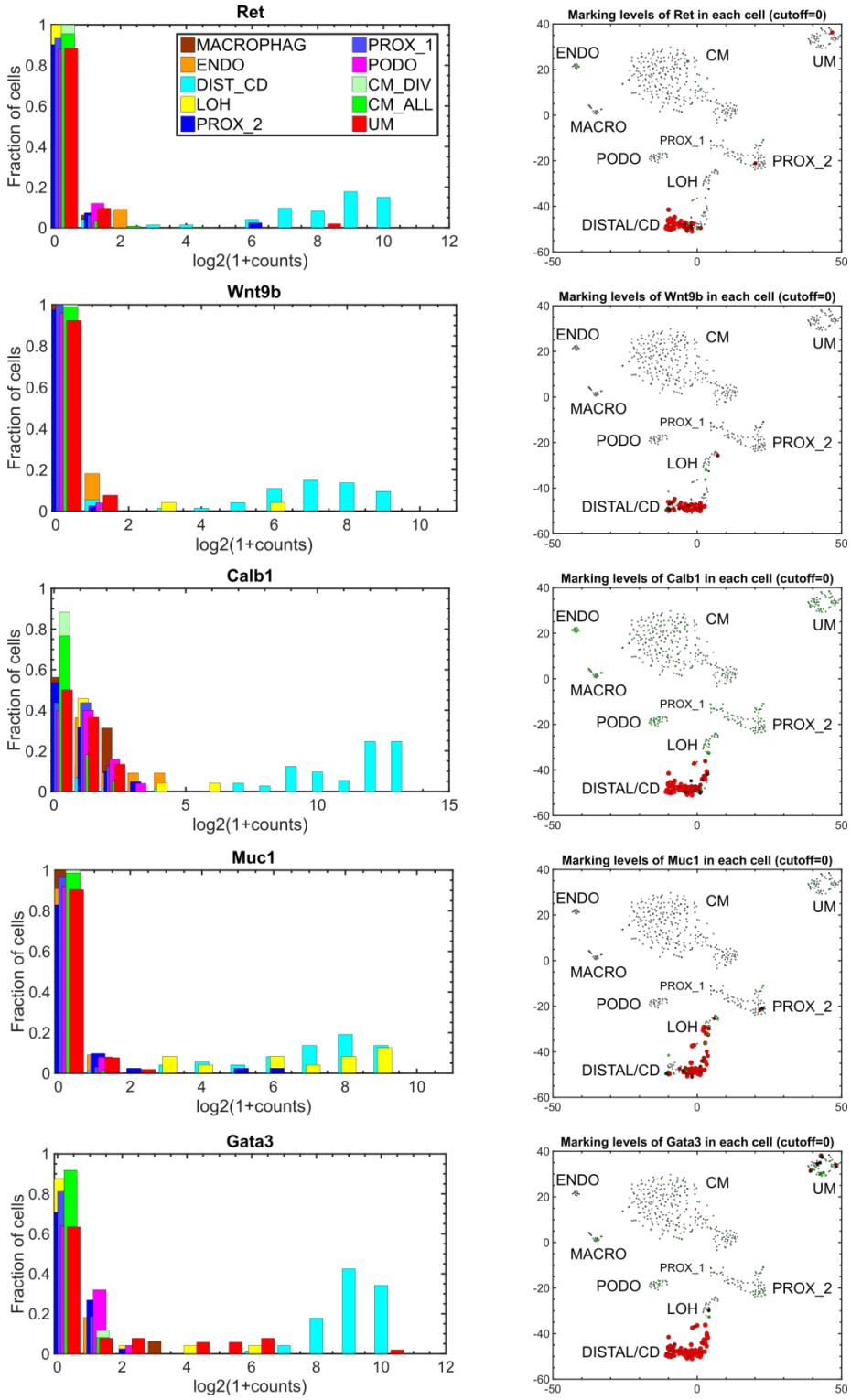


Figure S9: Confirmation of cell population identity. Shown are histograms and tSNE plots marking genes that are known to be over-expressed in the distal tubule and/or collecting duct (DIST_CD) [21,24,29–31]. We found it difficult to distinguish between the two populations, probably due to the small number of cells in our analysis. Note that the

genes Ret and Wnt9b, which are known to mark the collecting duct, are slightly more restricted that the other genes which are known to mark both the collecting duct and the distal tubules.

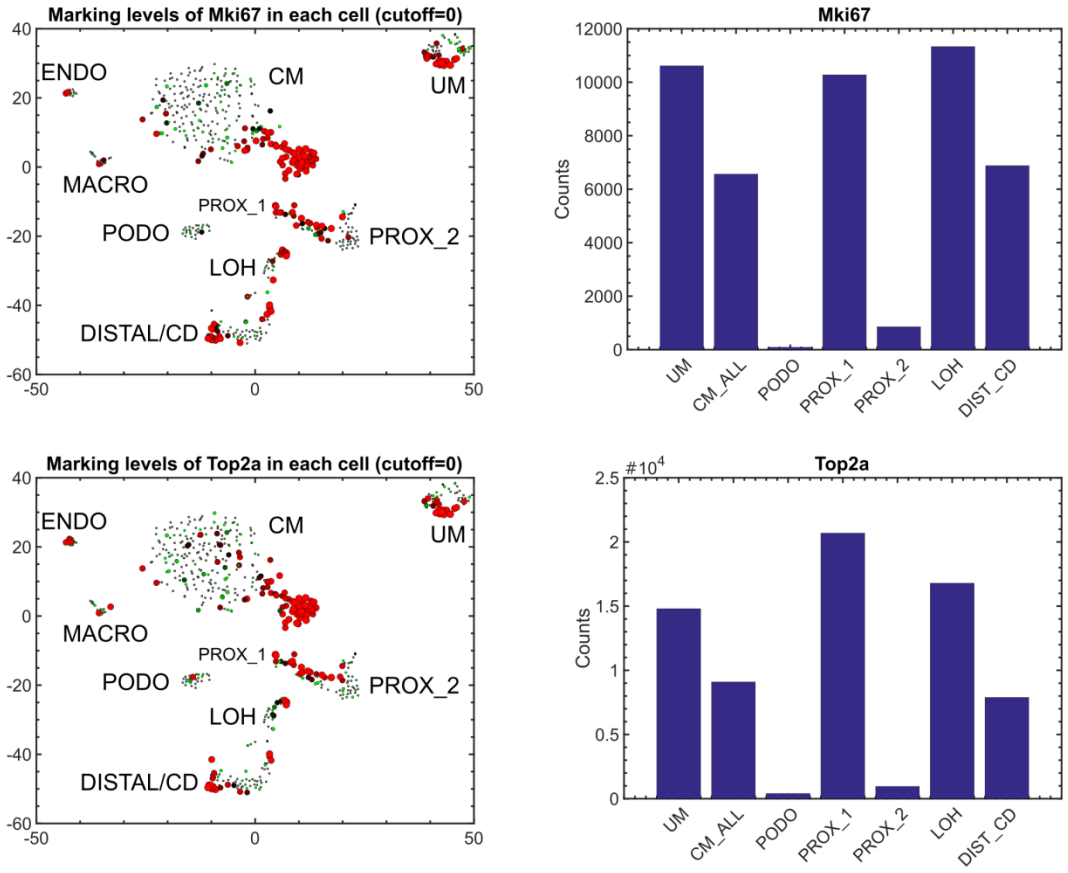
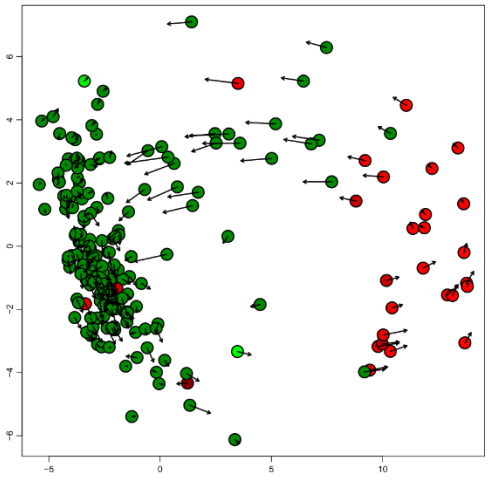
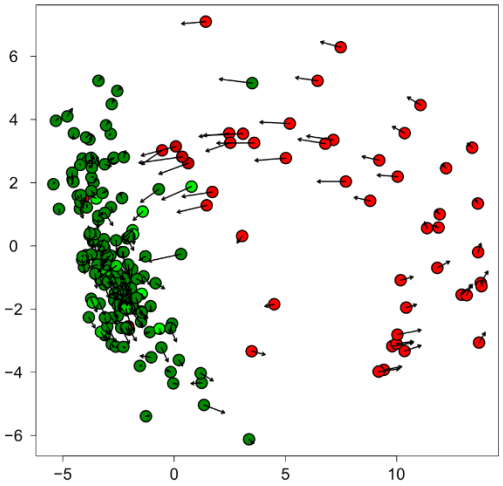


Figure S10: Cell division occurs in most cell populations of the developing kidney. Shown are tSNE plots marking the genes Mki67 and Top2a - genes that are typically expressed during the S-G2-M phases of the cell cycle - as well as barplots of their expression levels within the *in-silico* "bulk" transcriptomes representing the different populations. Note the higher expression in the early epithelial structures (PROX_1) with respect to the proximal tubules (PROX_2).

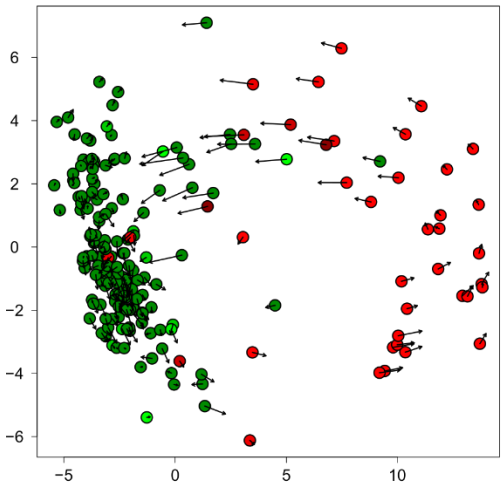
Anln



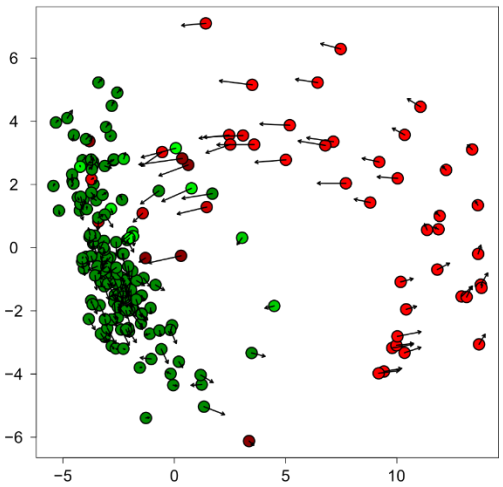
Birc5



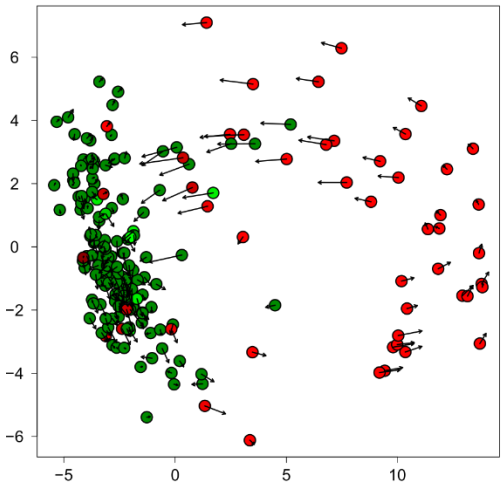
Ccna2



Mki67



Top2a



Arrows only

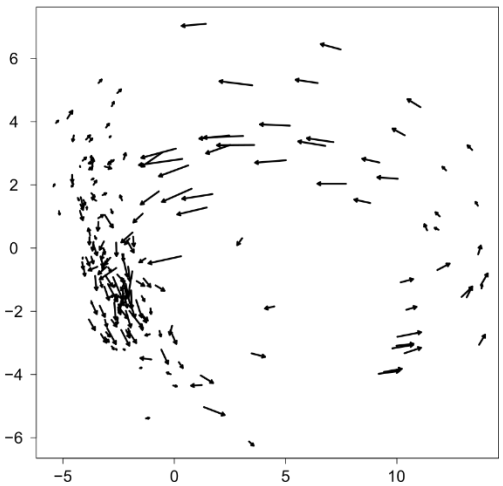


Figure S11: RNA velocity [32] of all Six2-high cells shows a consistent directional "flow" along the circular manifold created by the cells of the cap mesenchyme (CM) in gene expression space (PC 1 vs. 2). Each cell is represented by a dot. The arrows represent the directionality of each cell in gene expression space that is inferred from the difference between the spliced transcriptome (=present state) vs. yet-unspliced transcriptome (=near future state) [32]. Circle fill colors represent expression levels of selected genes (Red – high expression, green – low expression). It can be seen that cells over-expressing genes such as Top2a and Mki67 - genes that are known to be over-expressed in the S-G2-M phases of the cell cycle - are located in a specific segment of the circular manifold representing the S-G2-M segment of the cell cycle. At this segment the arrows are longest, indicating a rapidly changing transcriptional state.

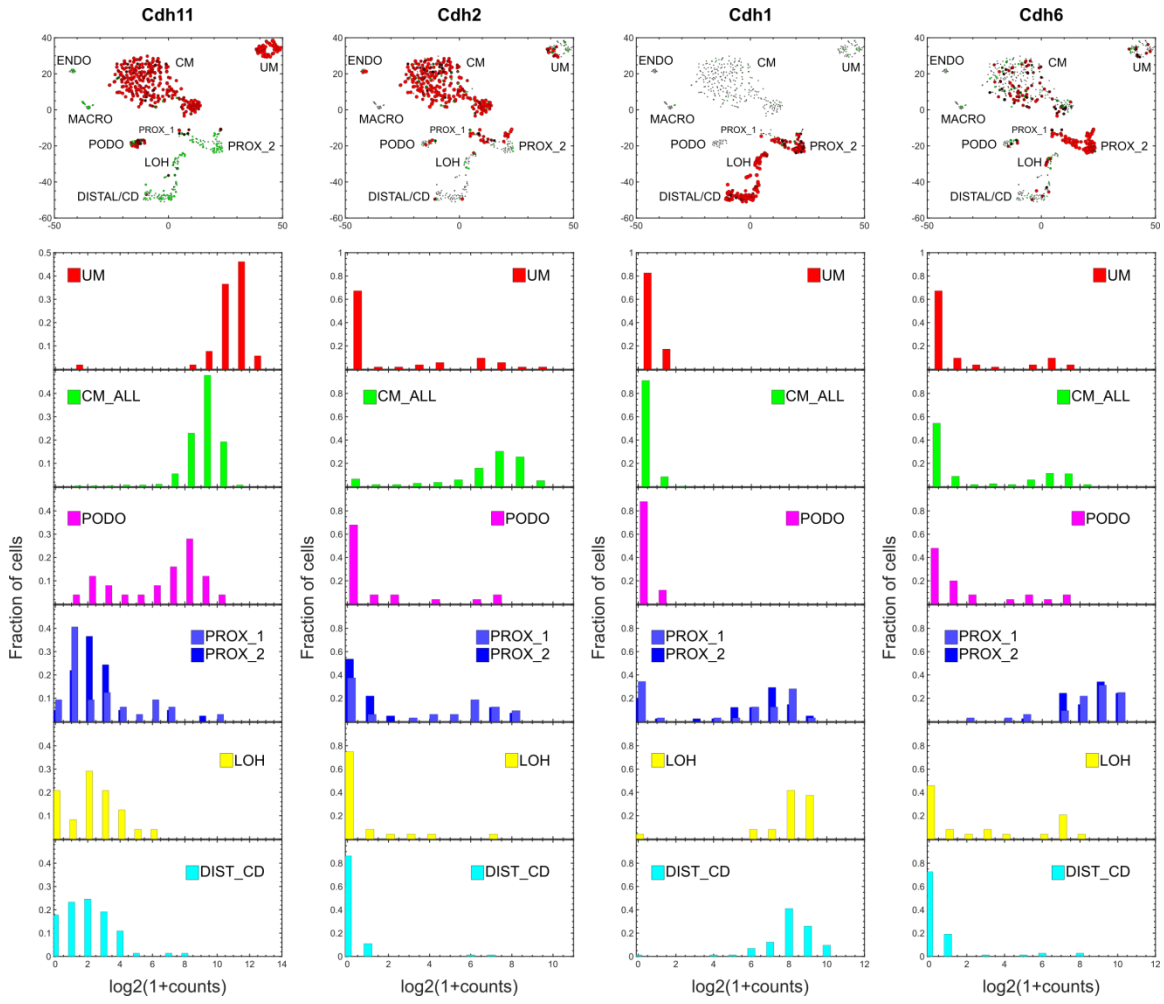


Figure S12: Mesenchymal to epithelial transition (MET) during kidney development. Shown are histograms and tSNE plots marking genes that are over-expressed in mesenchymal cells (Cdh11 [26] and Cdh2 [33]) and epithelial cells (Cdh1 and Cdh6 [26]). Note that Cdh11 is expressed in three levels: high expression in the un-induced mesenchyme (UM), moderate expression in the cap mesenchyme (CM), and low expression in the remaining epithelial populations (PROX_1, PROX_2, LOH, DIST_CD). In the podocytes we observed Cdh11 to have a bimodal distribution consisting of medium and low expression levels, indicating two distinct levels of differentiation.

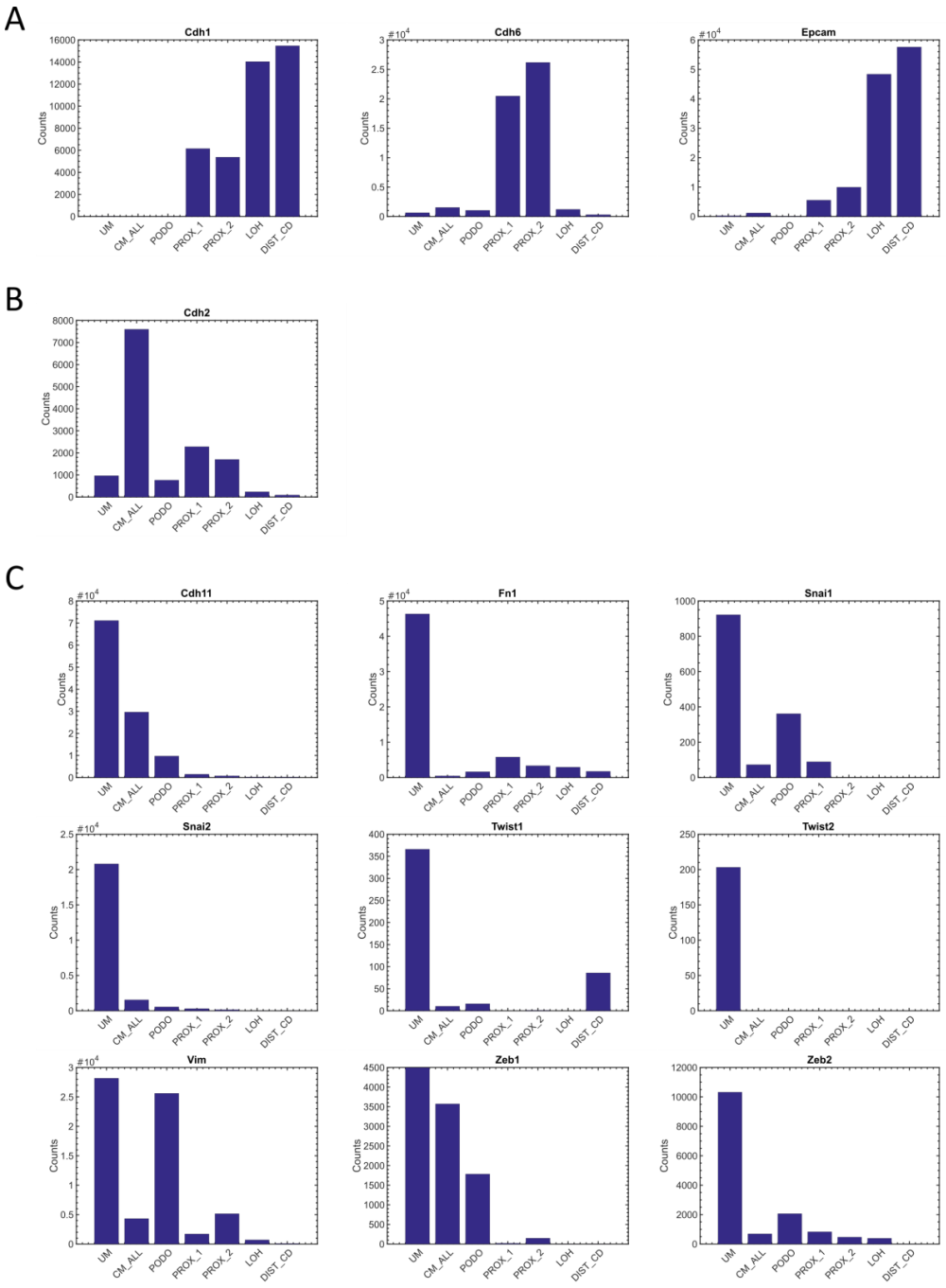


Figure S13: Mesenchymal to epithelial transition during kidney development. Shown are barplots of expression levels of selected genes within the *in-silico* "bulk" transcriptomes representing the different populations. (A) Epithelial markers (B) The mesenchymal marker Cdh2 [33] is over-expressed in the cap mesenchyme (CM) (C) Other mesenchymal markers. Note that Vimentin (Vim) is also highly expressed in the podocytes.

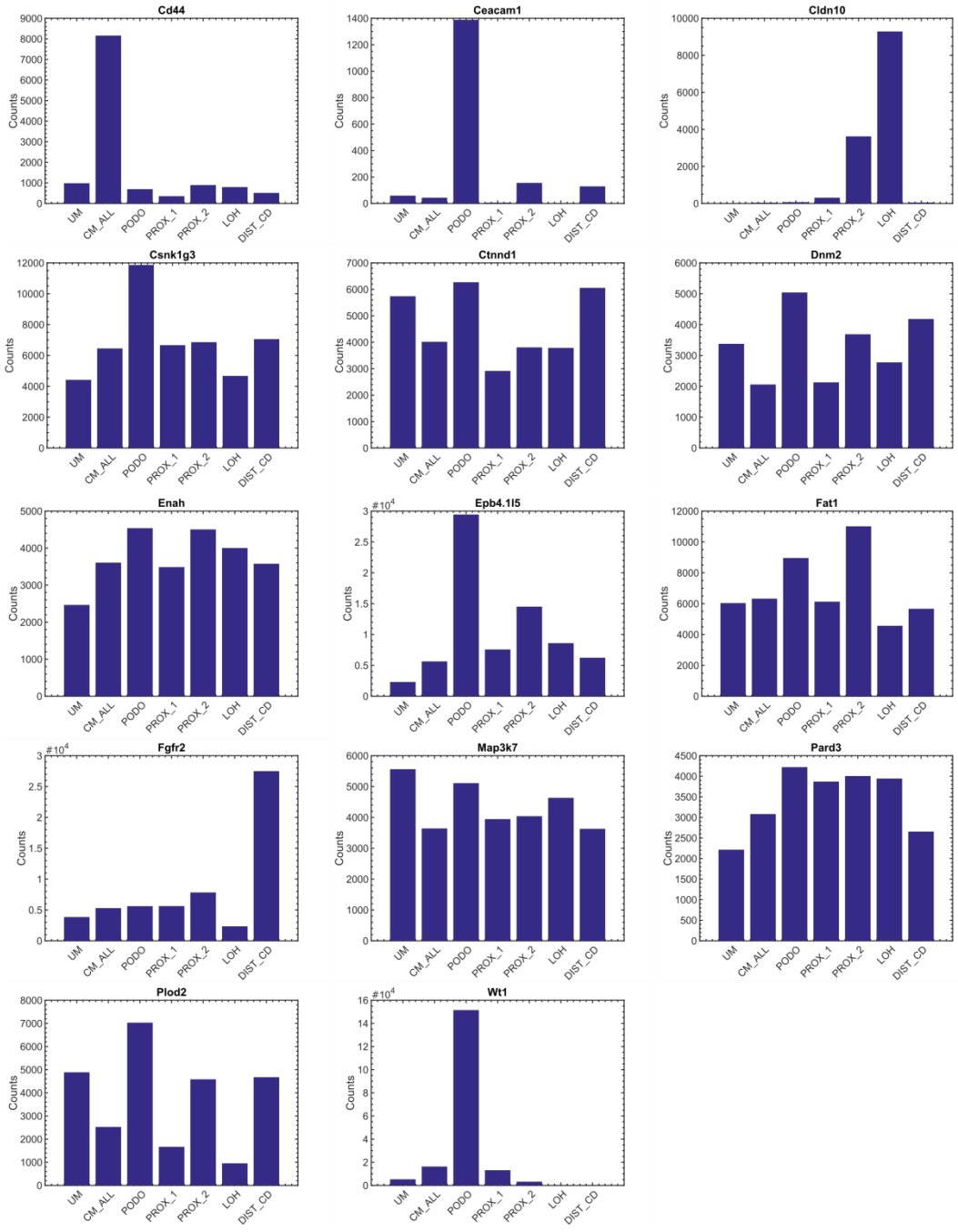


Figure S14: No obvious correlation between gene expression levels and inclusion levels of associated cassette exons. Shown are expression levels of selected genes in the *in-silico* "bulk" transcriptomes representing the different cell populations. For example, exons in the genes Dnm2 and Map3k7 have low inclusion levels in the mesenchymal populations and high inclusion levels in the epithelial populations (Fig. 3) but almost constant levels of gene expression.

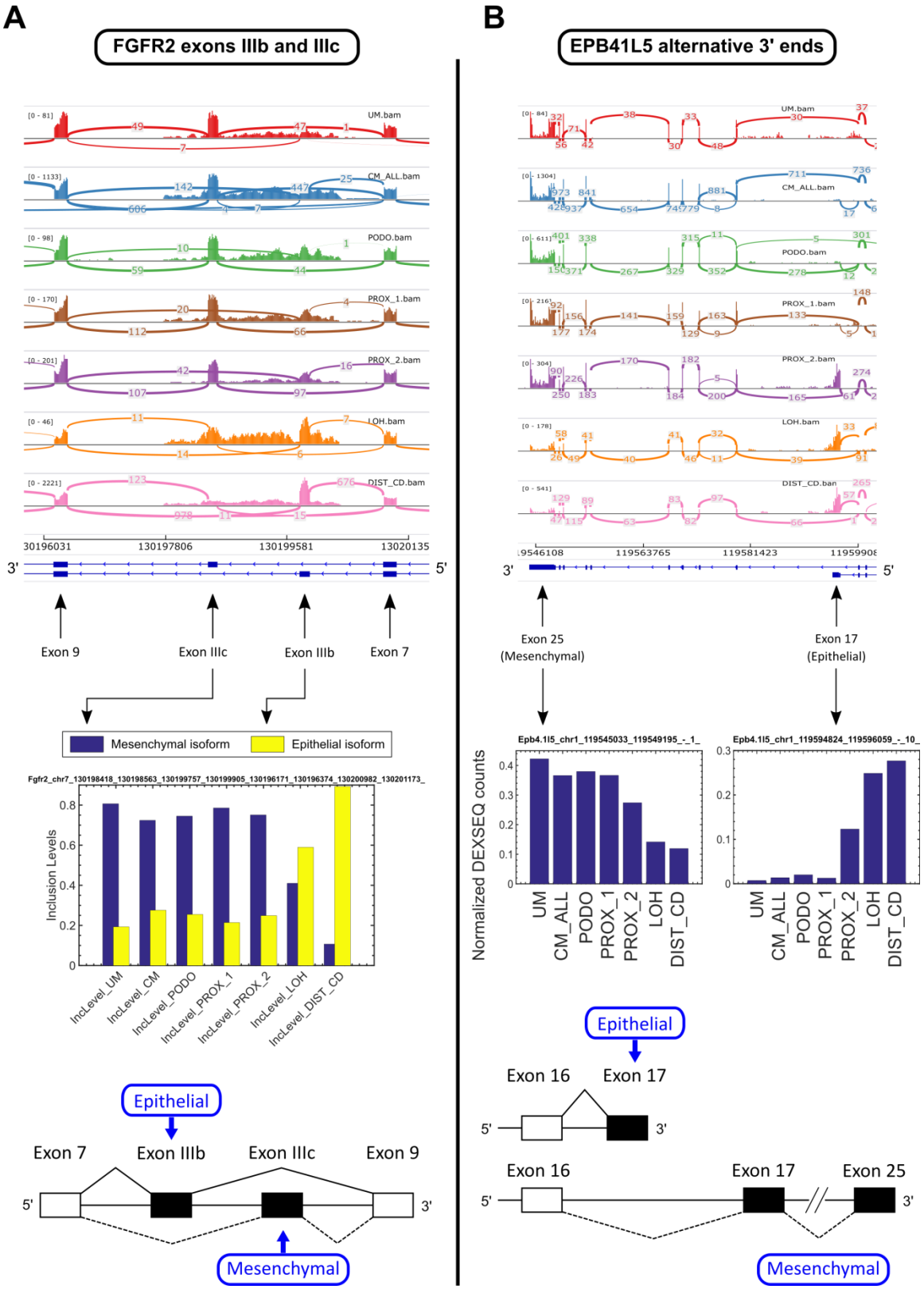


Figure S15: The genes Fgfr2 and Epb41l5 switch between mesenchymal and epithelial isoforms during kidney development. (A) The mesenchymal isoform of Fgfr2 (Fgfr2-IIIc)

[34,35] is predominantly expressed in the mesenchymal and early developmental cell populations (UM, CM, PODO, and to some extent PROX_1 and PROX_2) while the epithelial isoform (Fgfr2-IIIb) is expressed mostly in the mature epithelial cell populations (LOH and DIST_CD). Shown is a Sashimi plot and a bar plot of inclusion levels of each exon. Note that the tubular epithelial cell populations PROX_1, PROX_2, and LOH contain varying mixtures of the two isoforms. The dominance of the mesenchymal isoform of Fgfr2 (Fgfr2-IIIc) in the un-induced mesenchyme (UM) and cap mesenchyme (CM) is in agreement with previous observations that deletion of Fgfr2-IIIc (along with conditional deletion of Fgfr1 in the metanephric mesenchyme) results in poorly formed metanephric mesenchyme and unbranched ureteric buds [36,37]. (B) The longer mesenchymal isoform of Epb41l5 [34] is predominantly expressed in the mesenchymal and early developmental cell populations (UM, CM, PODO, and PROX_1) while the shorter epithelial isoform is expressed mostly in the mature epithelial cell populations (LOH and DIST_CD). Shown is a Sashimi plot and bar plots of normalized DEXSeq counts of reads that align to each exon. Note that the proximal tubular cell population (PROX_2) contains a mixture of the two isoforms.

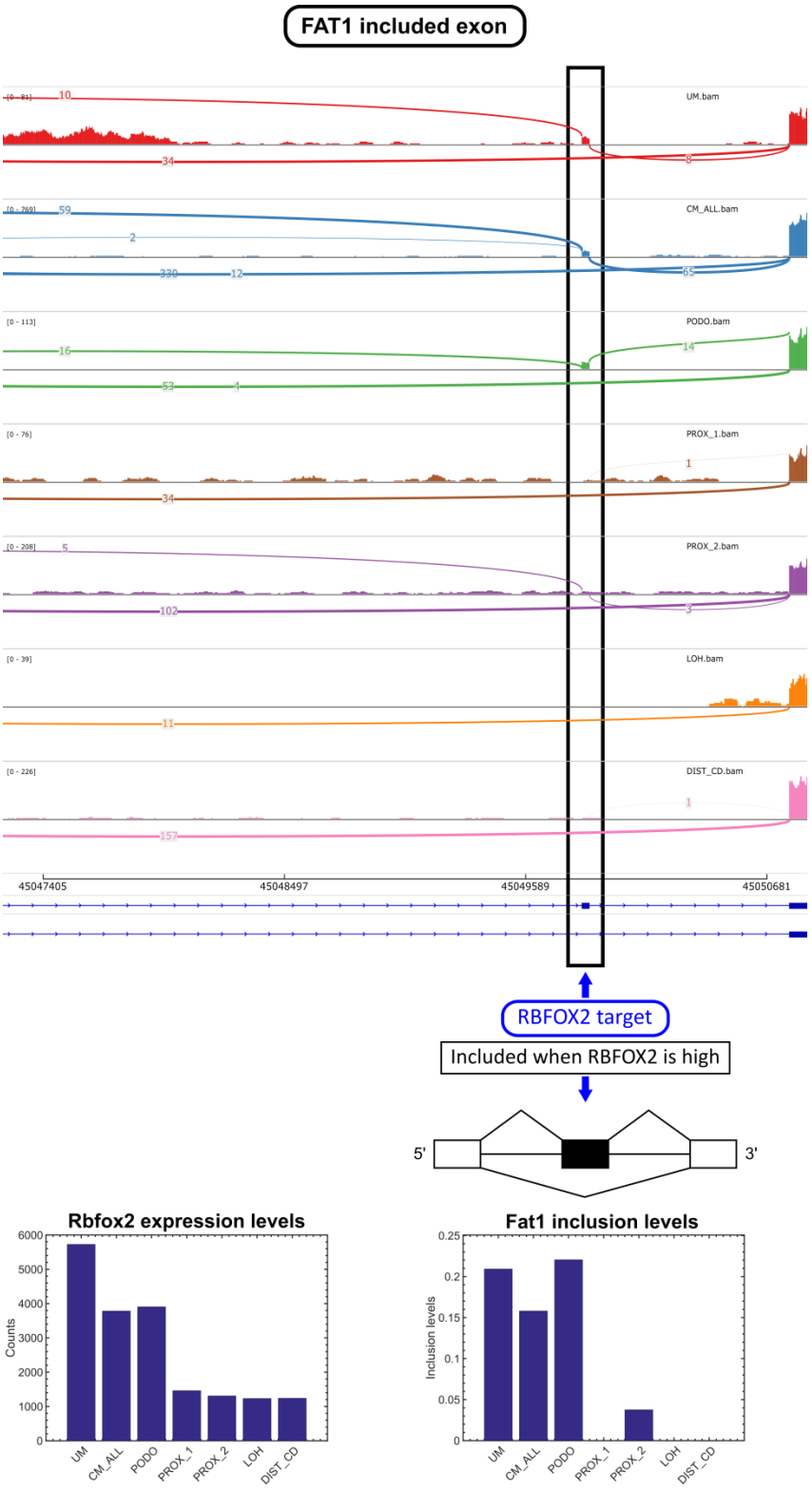


Figure S16: The gene Fat1, a known target of the splicing regulator Rbfox2 [38], undergoes splice isoform switching during kidney development. A cassette exon is expressed in the mesenchymal cell populations (UM, CM) and podocytes (PODO), and

repressed in the epithelial cell populations (PROX_1, PROX_2, LOH, and DIST_CD), in accordance with the expression levels of Rbfox2. This indicates that Rbfox2 acts as a splicing regulator during kidney development. Shown is a Sashimi plot and a bar plot of inclusion levels (manually calculated) for the alternatively spliced exon in Fat1, as well as a bar plot of expression levels for Rbfox2.

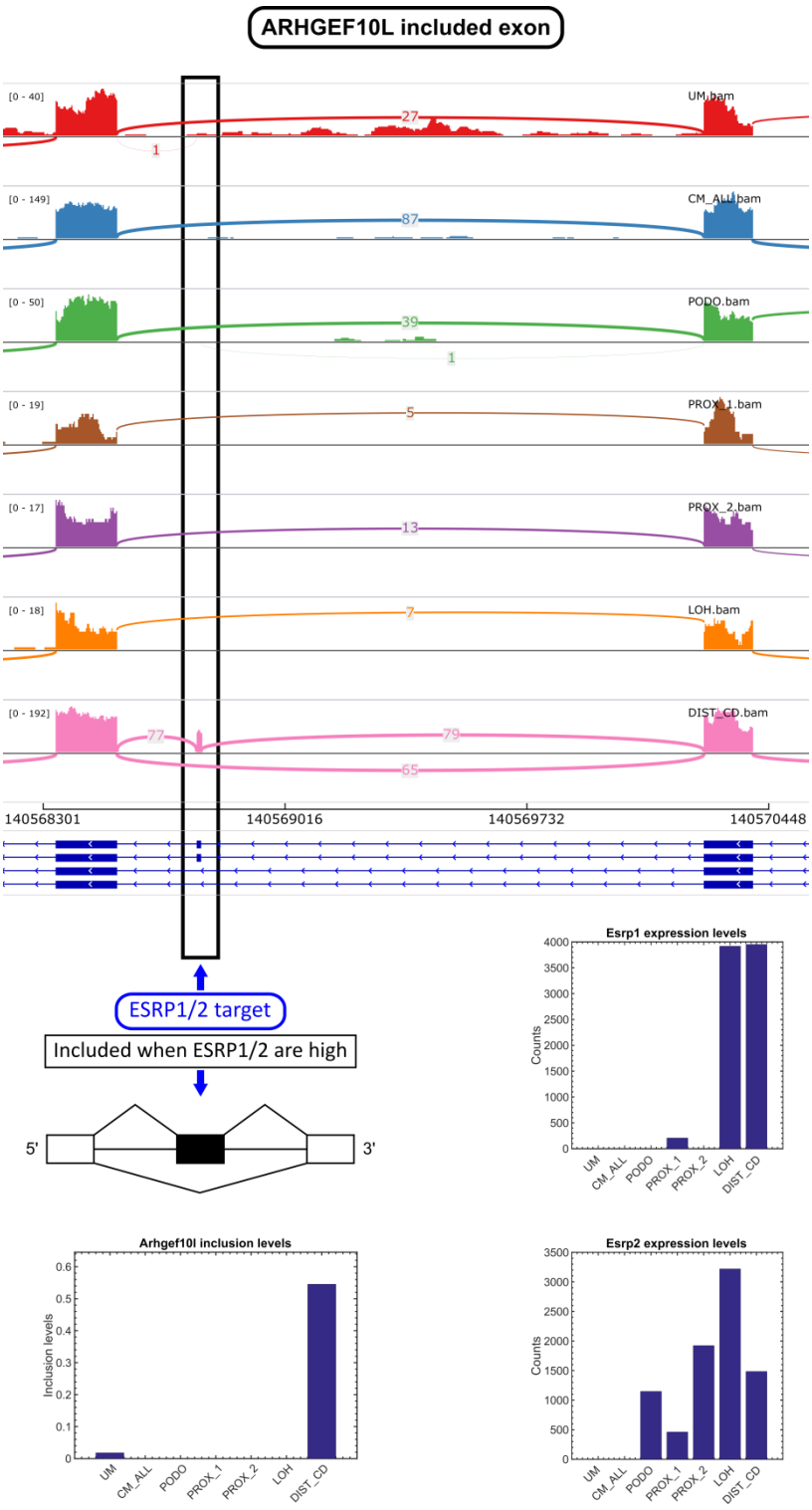


Figure S17: The gene Arhgef10l, a known target of the splicing regulators Esrp1/2 [39], undergoes splice isoform switching during kidney development. A cassette exon is repressed in the mesenchymal populations (UM,CM) and podocytes (PODO) and over-expressed in the distal tubules/collecting duct (DIST_CD). This is in accordance with the

expression levels of Esrp1 and Esrp2, which are jointly under-expressed in the mesenchymal populations (UM,CM) and over-expressed in the epithelial populations (mainly LOH and DIST_CD). This indicates that Esrp1 and Esrp2 act as splicing regulators during kidney development. Shown is a Sashimi plot and a bar plot of inclusion levels (manually calculated) for the alternatively spliced exon in Arhgef10l, as well as bar plots of expression levels for Esrp1/2. Note that Arhgef10l is sufficiently expressed only in the mesenchymal populations (UM, CM), podocytes (PODO), and distal tubules/collecting duct (DIST_CD) but not so much in the other epithelial populations (PROX_1, PROX_2, and LOH), thus limiting our ability to reliably measure the inclusion levels of the alternatively spliced exon in these populations.

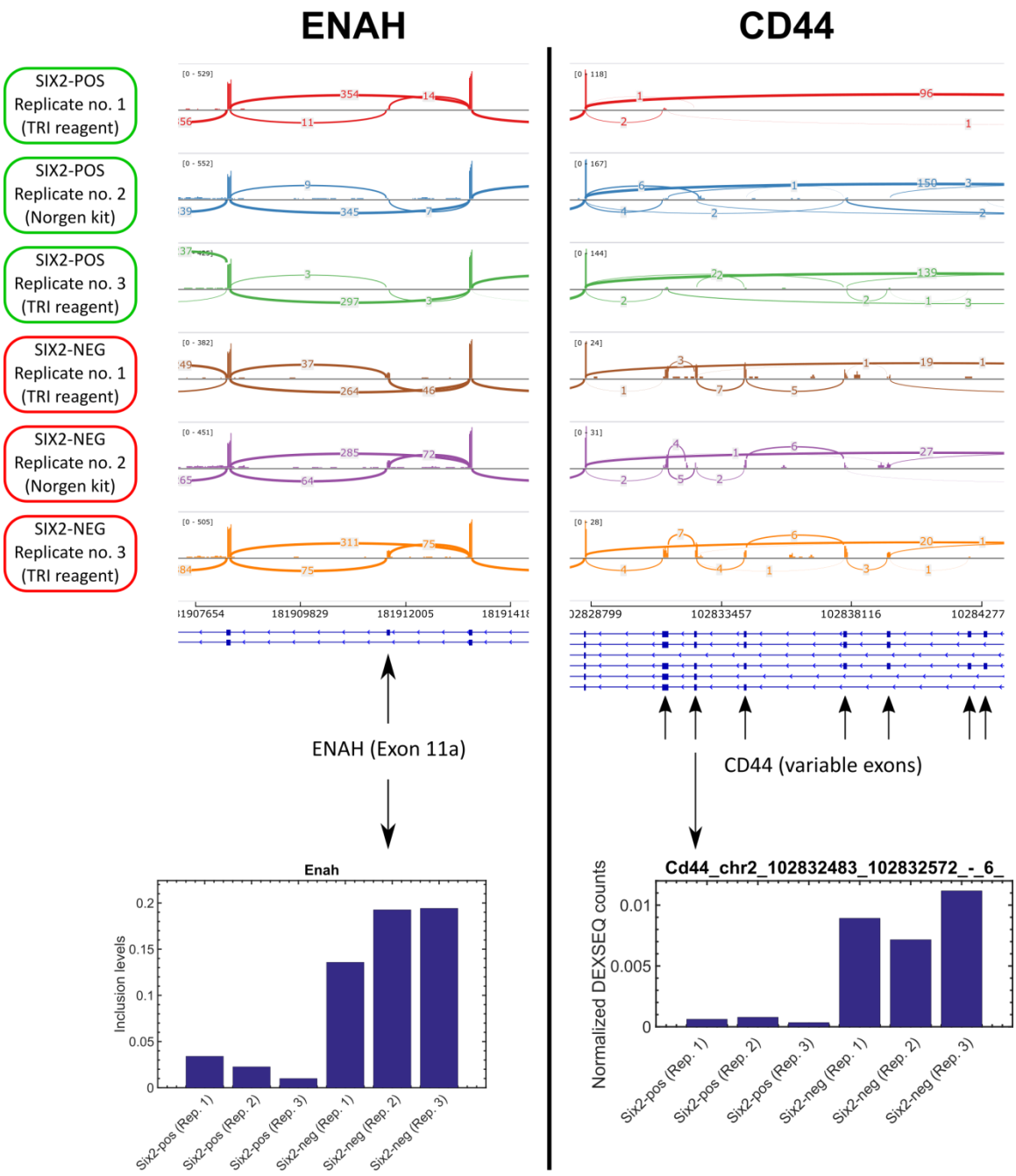


Figure S18: Alternative splicing in the genes Enah and CD44, two genes that are known to undergo splice isoform switching during EMT [34,35,38,40–42]. Shown are Sashimi plots from replicates of "bulk" RNA samples that were isolated from Six2-high and Six2-low mouse fetal kidney cells. We also show a bar plot of inclusion levels (manually calculated) of the alternatively spliced exon in Enah and a bar plot of normalized DEXSeq counts of reads that align to a selected alternatively spliced exon in Cd44. Since these cells contain a GFP under the control of a Six2 promoter, and since Six2 is highly expressed in the cap mesenchyme, the Six2-high fraction is predominantly

mesenchymal, as it contains mainly cells from the cap mesenchyme (CM), whereas the Six2-low cell fraction contains all other cell populations including all the epithelial cells (PODO, PROX_1, PROX_2, LOH, DIST_CD) as well as some mesenchymal cells from the un-induced mesenchyme (UM). Accordingly, the cassette exons are repressed in the Six2-high cell fraction (that is enriched for the cap mesenchyme, CM) and over-expressed in Six2-low cell fraction (that is enriched with the epithelial cell populations). We isolated total RNA from 3 replicates of cells that were sorted by FACS for high and low levels of the Six2-GFP reporter gene. RNA was isolated using two different kits (TRI reagent and Norgen) and sequenced. We note that the splice isoform switching in Enah and CD44 was difficult to observe in the single cell RNAseq dataset, probably due to the bias and relative sparsity of reads in the single cell data, but was more apparent in the "bulk" RNAseq dataset.

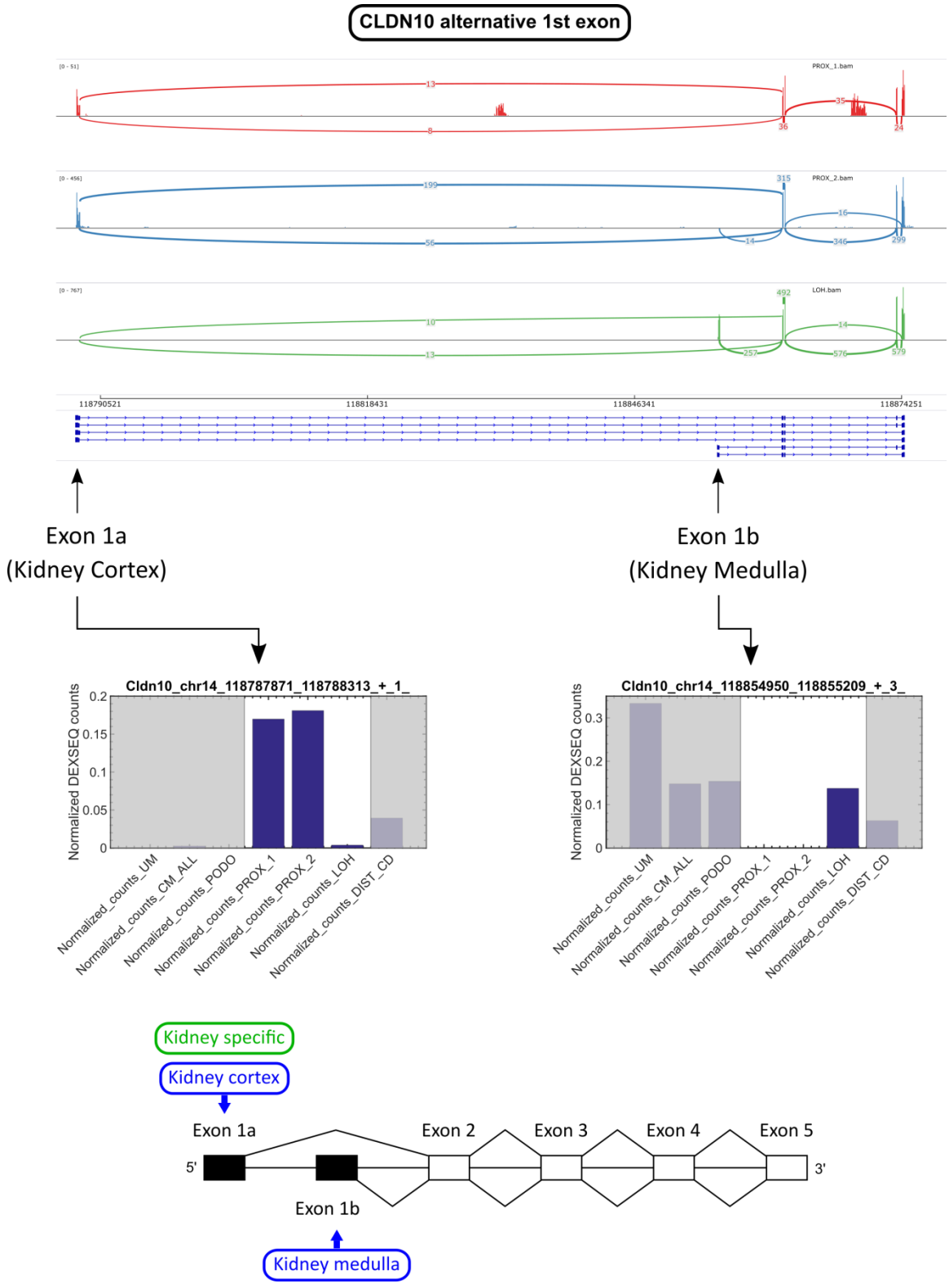


Figure S19: Alternative splicing of the gene Cldn10. The isoform Cldn10-1a, a kidney-specific isoform known to be expressed in the kidney cortex [43], was observed to be predominantly expressed in the early epithelial structures (PROX_1) and the proximal

tubular cells (PROX_2). The isoform Cldn10-1b, which is known to be expressed in the kidney medulla, was observed to be predominantly expressed in the loop of Henle (LOH). Shown is a Sashimi plot and bar plots of normalized DEXSeq counts of reads that align to exons that are alternatively spliced in these two isoforms. The different isoforms might be related to the variable paracellular permselectivity (i.e., the permeability of the epithelial tight junctions to specific ions) along the different segments of the nephron tubules [43]: Cldn10-1a creates “leaky” pores in the tubular epithelium, and is therefore over-expressed in the proximal tubule which is more “leaky” in order to allow for rapid reabsorption of essential ions, whereas Cldn10-1b is over-expressed in the loop of Henle in which ion transport is more tightly regulated.

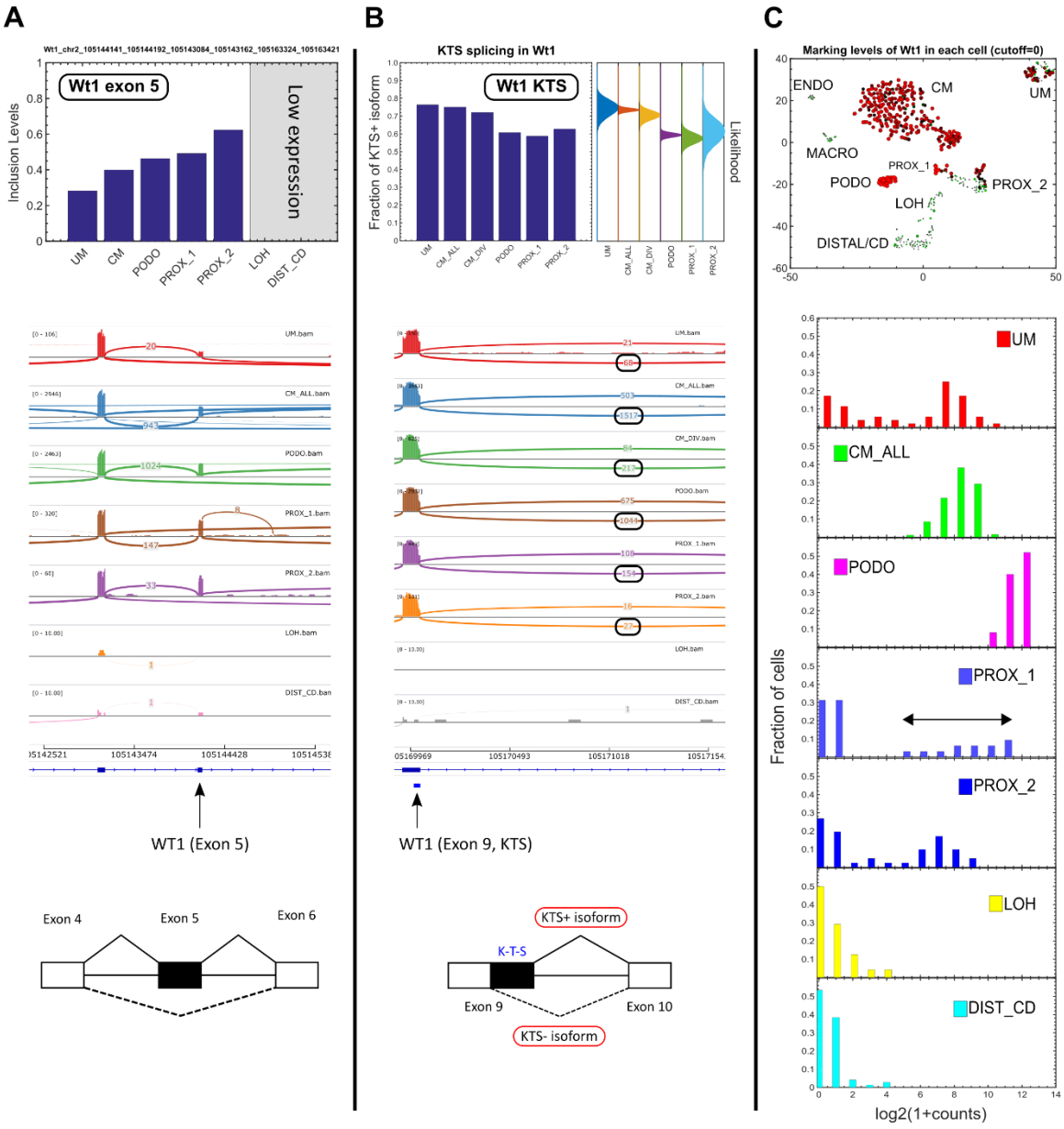


Figure S20: Splice isoform switching and gene expression analysis of the gene Wt1, a central gene in kidney development, whose deletion or mutation can lead to developmental defects and pediatric tumors of the kidney (e.g. Wilms' tumors) [4–8]. (A) Cassette exon 5 increases gradually during development. Shown are a barplot of inclusion levels and a Sashimi plot of the *in-silico* "bulk" transcriptomes representing the different populations. (B) KTS alternative splicing in exon 9. Shown are a barplot of inclusion levels, a likelihood plot for the inclusion levels (see below), and a Sashimi plot. An alternative splice donor site at exon 9 inserts or skips the three amino acids lysine (K), threonine (T), and serine (S). The ratio between the Wt1(+KTS) isoform - that includes the KTS segment - and the Wt1(−KTS) isoform – that skips the KTS segment - is typically 60:40 [5,11,12]. A disruption of this ratio to 30:70 is associated with Frasier syndrome, a kidney developmental defect. Here we see that the KTS+:KTS- ratio is

approximately 70:30 in the mesenchymal cell populations (the un-induced mesenchyme [UM], the cap mesenchyme [CM], and CM_DIV which consists of only the actively dividing cells within the cap mesenchyme) and converges to approximately 60:40 in the epithelial cell populations (podocytes [PODO], early epithelial structures [PROX_1], and proximal tubules [PROX2]). (C) Single-cell gene expression analysis of Wt1. Shown is a tSNE plot and histograms showing the expression levels of Wt1 in each of the different populations. It can be seen that Wt1 is most highly expressed in the podocytes [13] (PODO), moderately expressed in the un-induced mesenchyme (UM) and cap-mesenchyme (CM), and under-expressed in the loop of Henle (LOH) and distal tubules/collecting duct (DIST/CD). Note the wide distribution of expression in the early epithelial structures (PROX_1), which is probably due to the fact that some cells (e.g. those in the cleft of the S-shaped body [14]) are in the process of differentiating to podocytes while others are destined to become constituents of the proximal tubule, loop of Henle, or distal tubule. The area of each circle in the tSNE plots is proportional to log2(1+expression) of Wt1 in that particular cell.

Calculation of KTS inclusion levels: To calculate the likelihood function for the KTS inclusion level ψ, which is defined as the fraction of transcripts that include the KTS segment out of the total number of transcripts that either include the KTS segment or skip over it [1,2], we used both rMATS and manual read counting and obtained similar results.

For manual read counting, we counted the number of reads that span the junction between exons 9 and 10, while either including the KTS segment (I) or skipping it (S).

Since:

$$I|\psi \sim Binomial(N = I + S, p = \psi)$$

$$Prob(I|N = I + S, p = \psi) = \binom{I+S}{I} \psi^I (1-\psi)^S$$

Note that since we only counted reads that span the junction between exons 9 and 10, the effective lengths [1] (that is, the number of unique isoform-specific read positions) of the KTS-inclusion isoform and the KTS-skipping isoform are taken to be the same.

Therefore, for given values of I and S, we get the likelihood:

$$L(\psi|I,S) = \binom{I+S}{I} \psi^I (1-\psi)^S$$

Where the, maximum likelihood is $\psi_{ML} = I/(I+S)$.

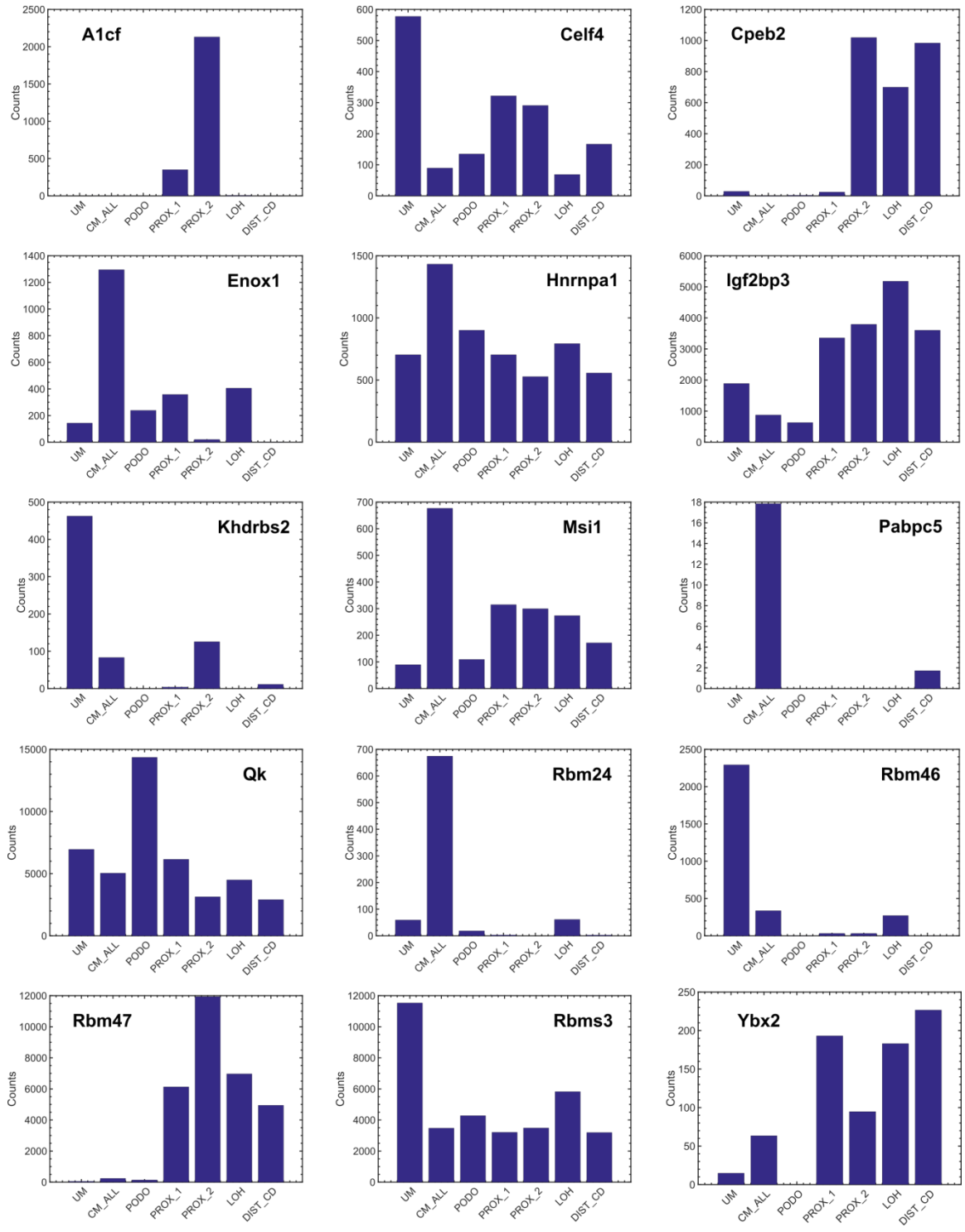


Figure S21: Additional putative splicing regulators. Shown are barplots for selected RNA binding proteins (RBP's) [44] that change their expression during the transition from mesenchymal to epithelial states.

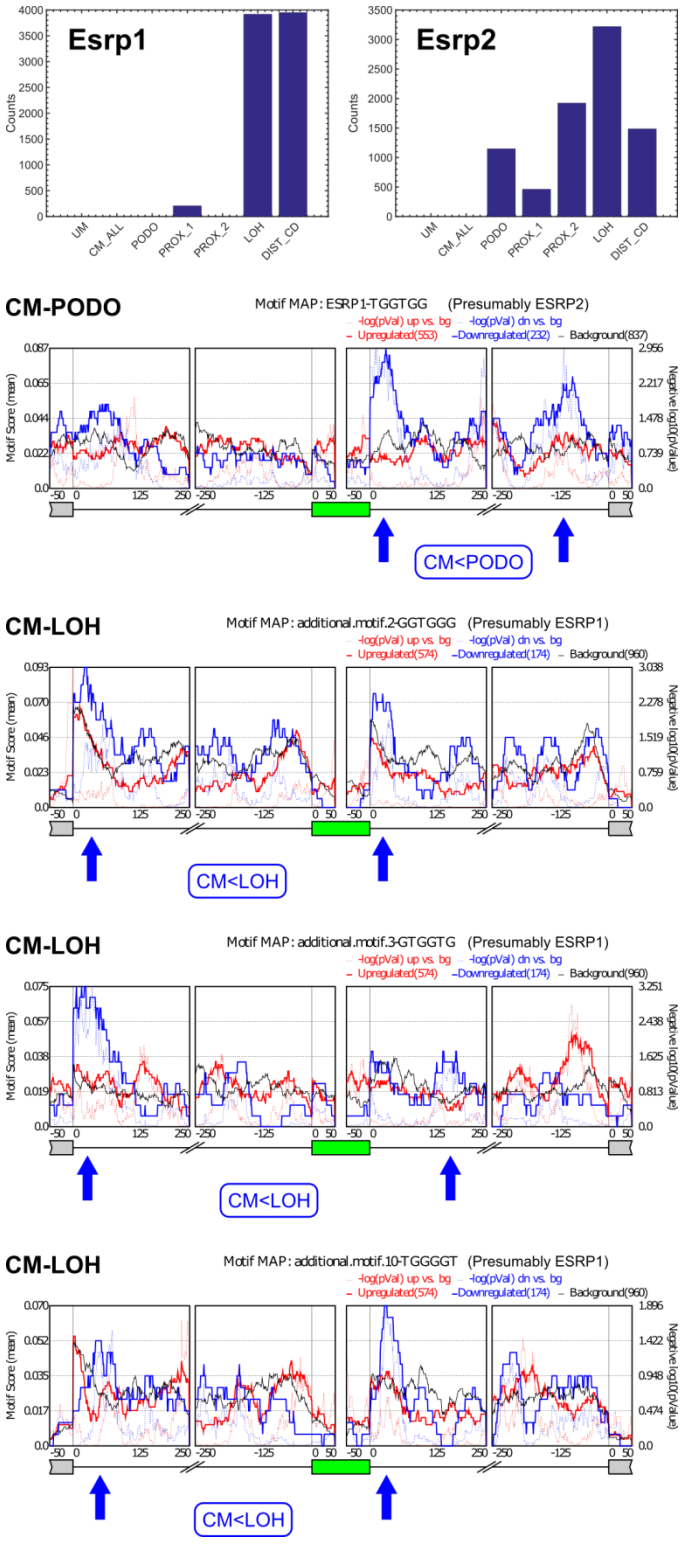


Figure S22: Motif enrichment analysis for additional UGG-enriched motifs that were previously found to be binding sites for the RNA binding proteins Esrp1 [39,45] and Esrp2 [46]. Cassette exons that are over-expressed in the epithelial populations contain a significant enrichment of ESRP1/2 binding motifs in their downstream 3'-flanking

introns, and in some cases (CM vs. LOH), also in the far 5' end of their upstream 5'-flanking introns as was previously observed using SELEX-Seq experiments in EMT [45]. This further indicates that ESRP1/2 are splicing regulators involved in Mesenchymal to Epithelial Transition (MET) during kidney development. Note that although the TGGTGG motif (2nd panel from top) is usually regarded as an RNA binding site for Esrp1 [44,45] (Table S10), we hypothesize that in this case it serves as a binding site for Esrp2 since Esrp1 is not expressed in the podocytes and since this motif is similar to the Esrp2 binding motif TGGTG [46,47].

Abi1

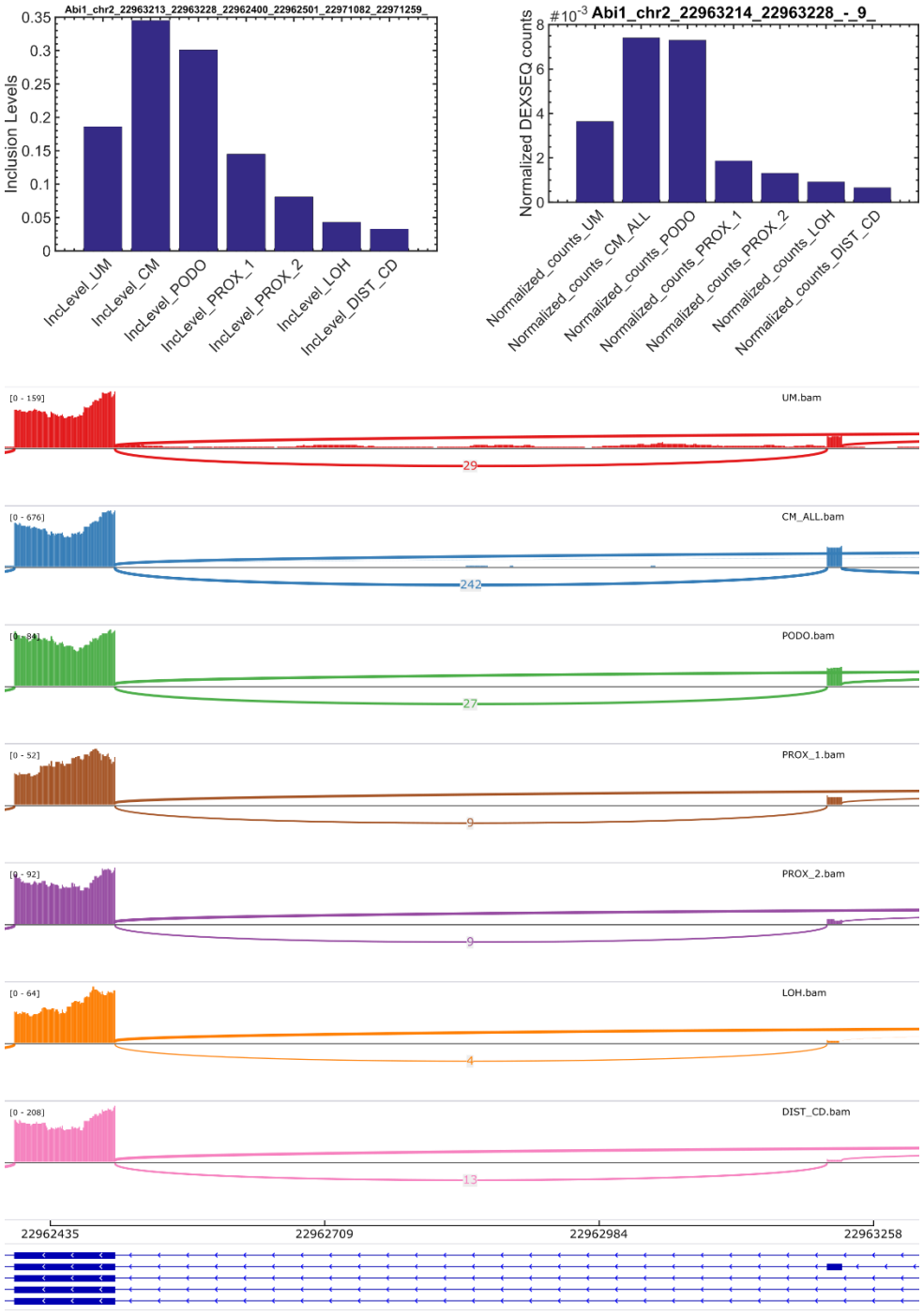


Figure S23: The gene Abi1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Add3

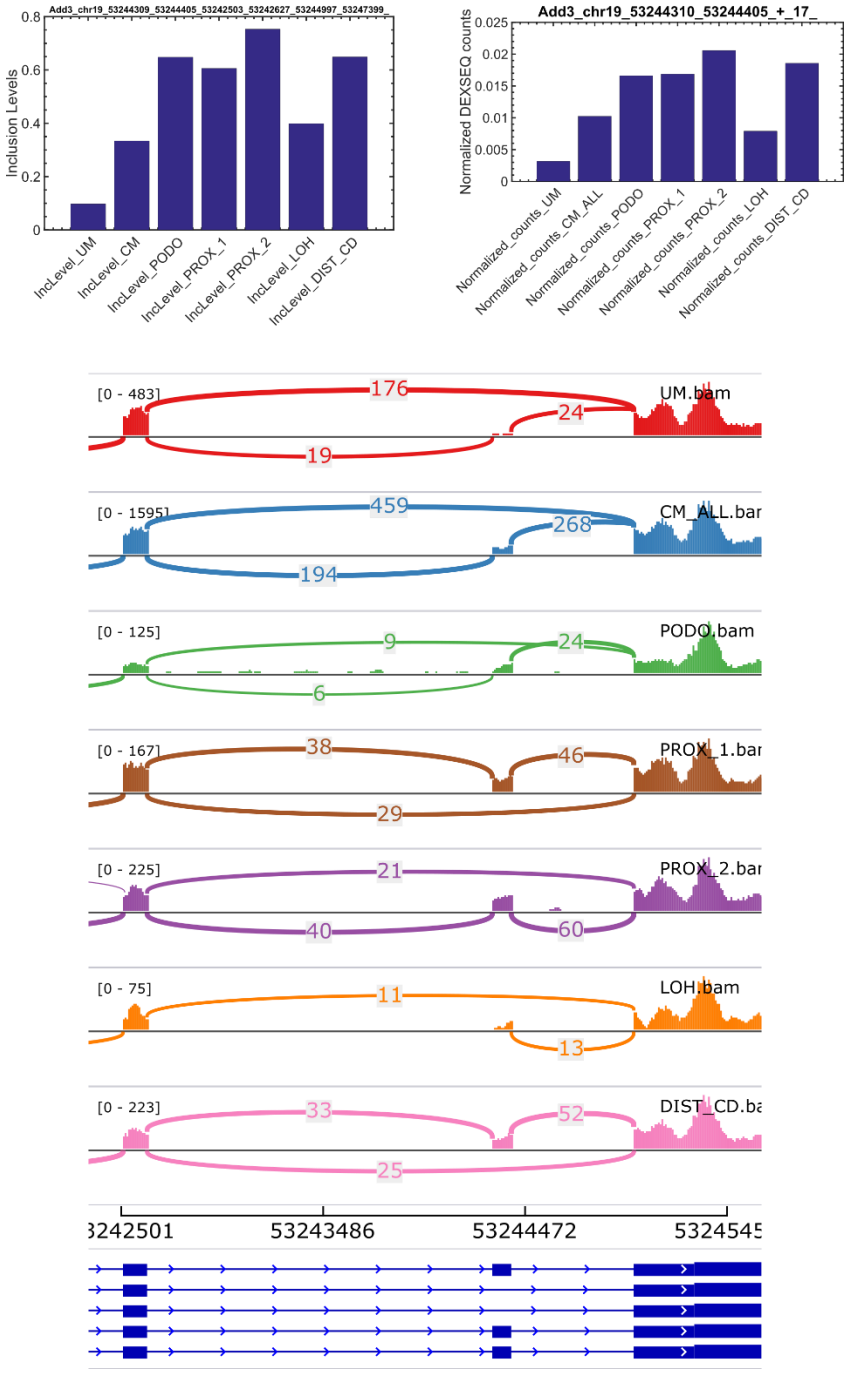


Figure S24: The gene Add3 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

App

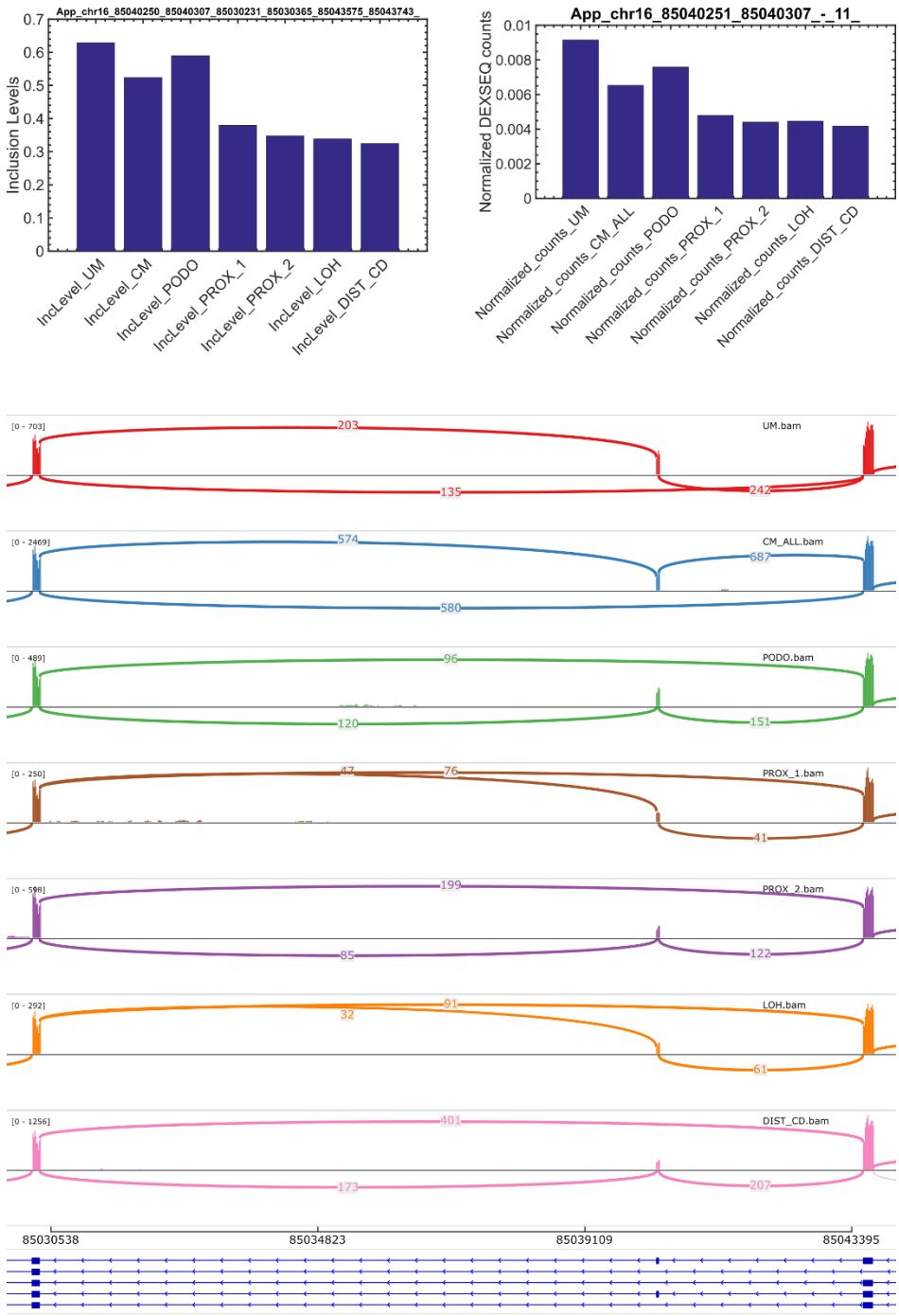


Figure S25: The gene App undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Csnk1g3

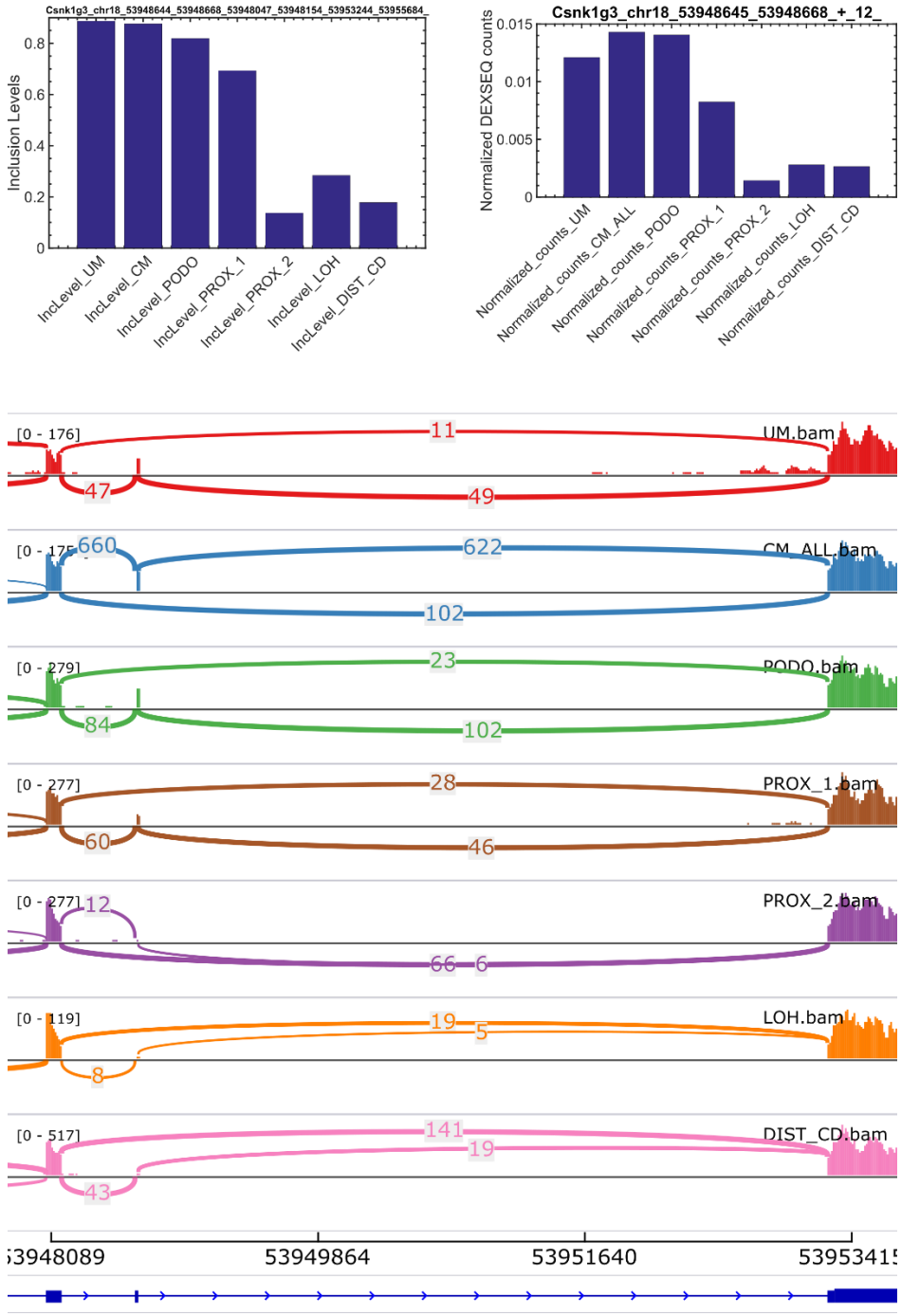


Figure S26: The gene Csnk1g3 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Ctage5

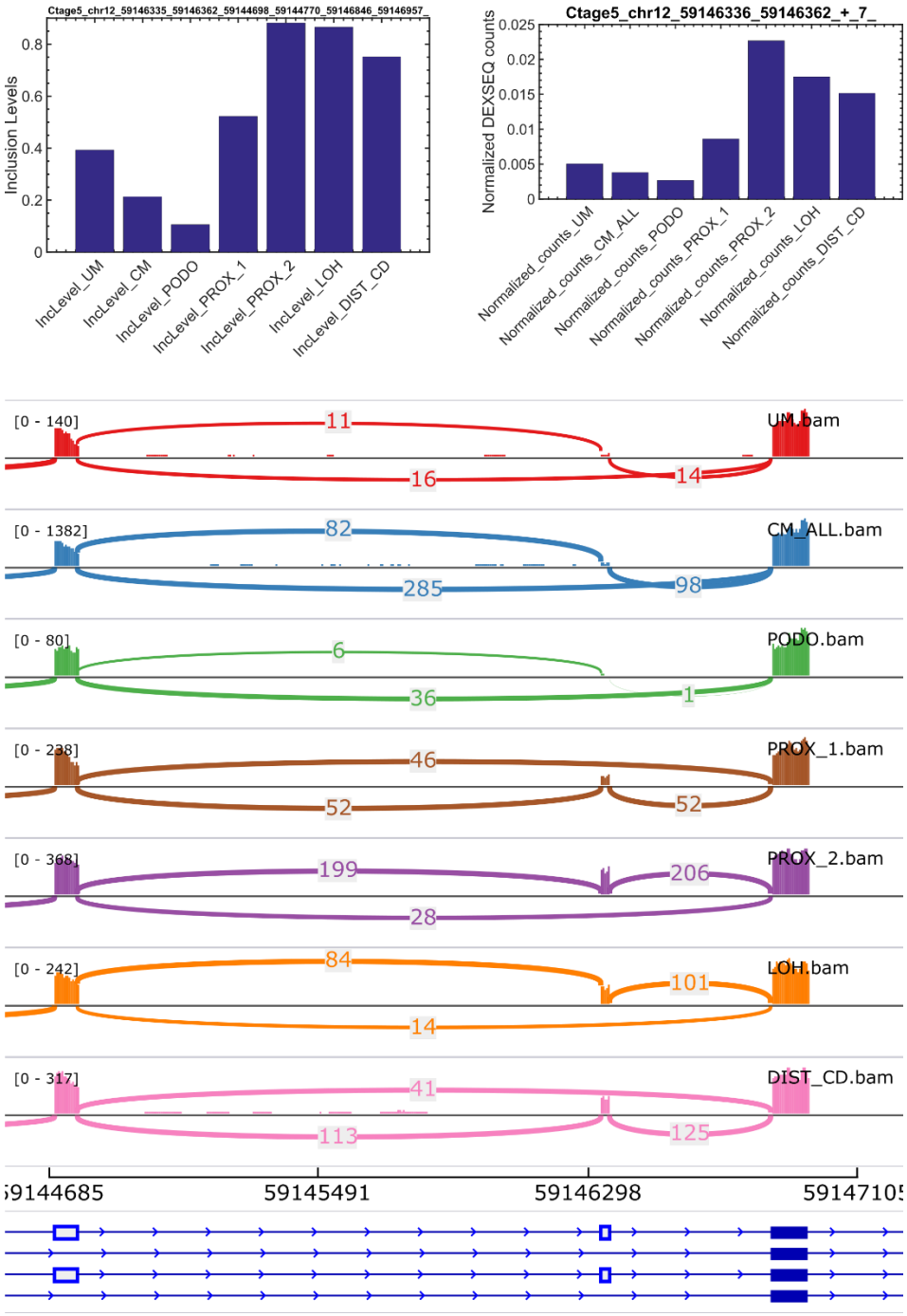


Figure S27: The gene Ctage5 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Ctnnd1

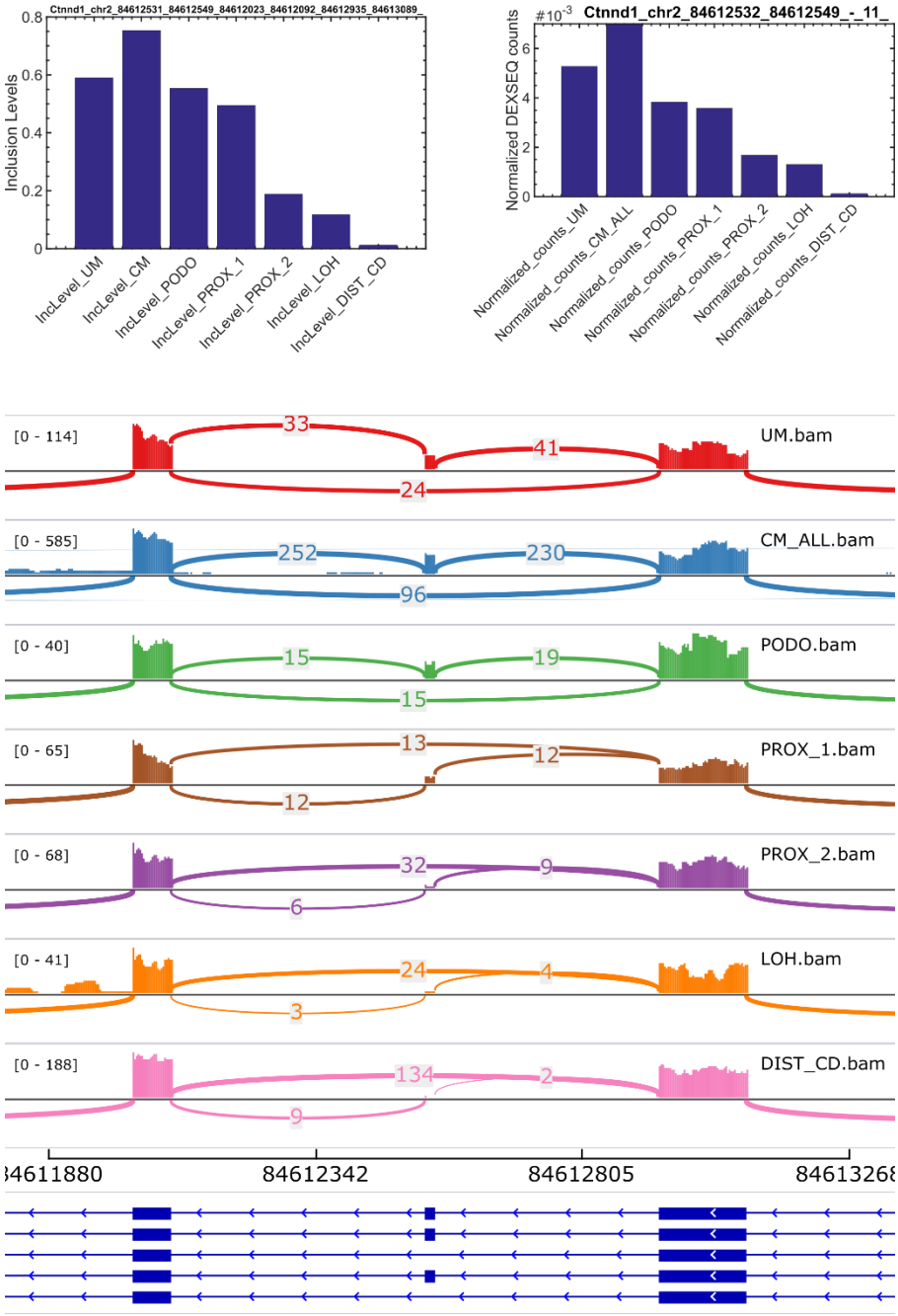


Figure S28: The gene Ctnnd1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Dnm2

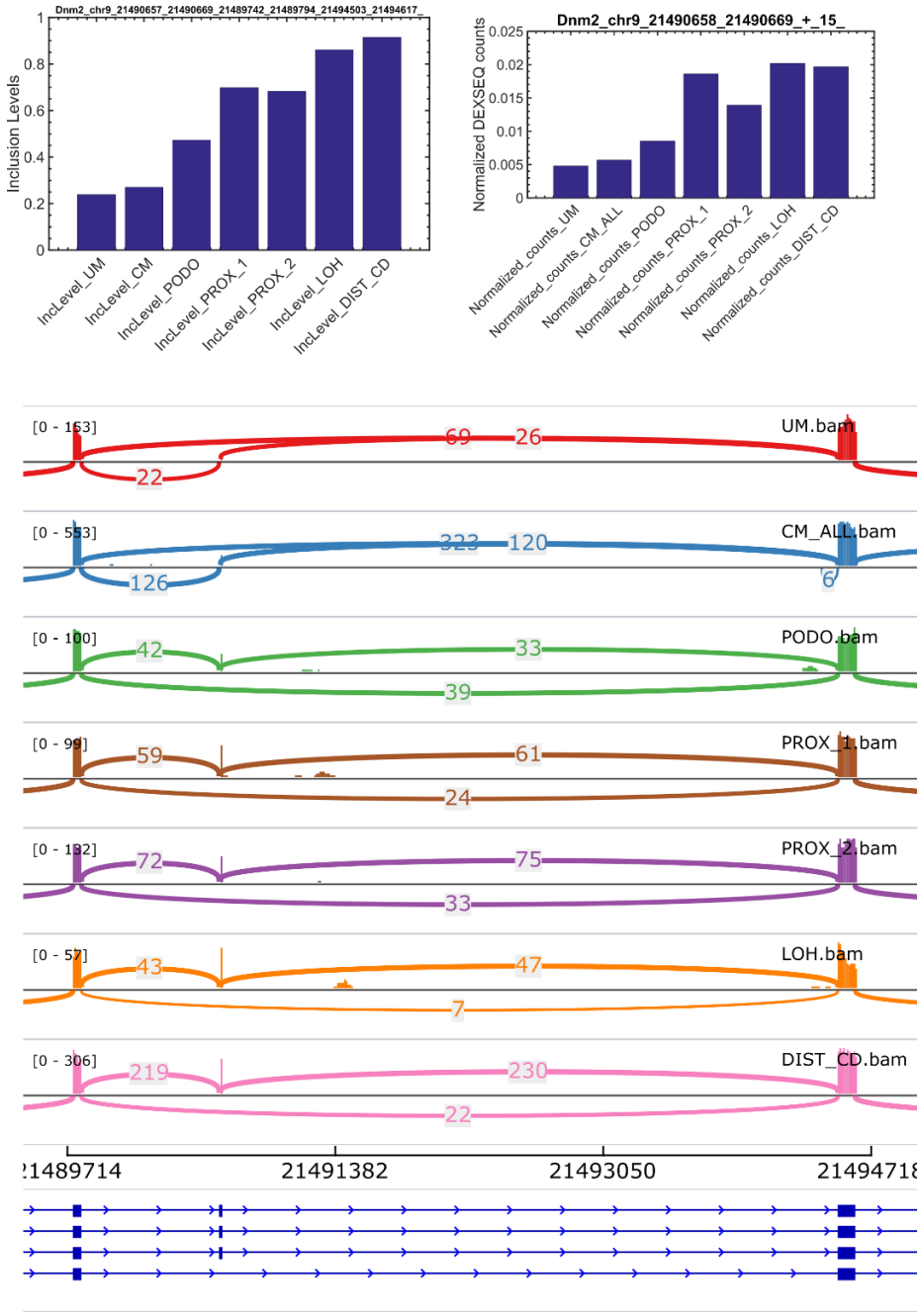


Figure S29: The gene Dnm2 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Dtnb

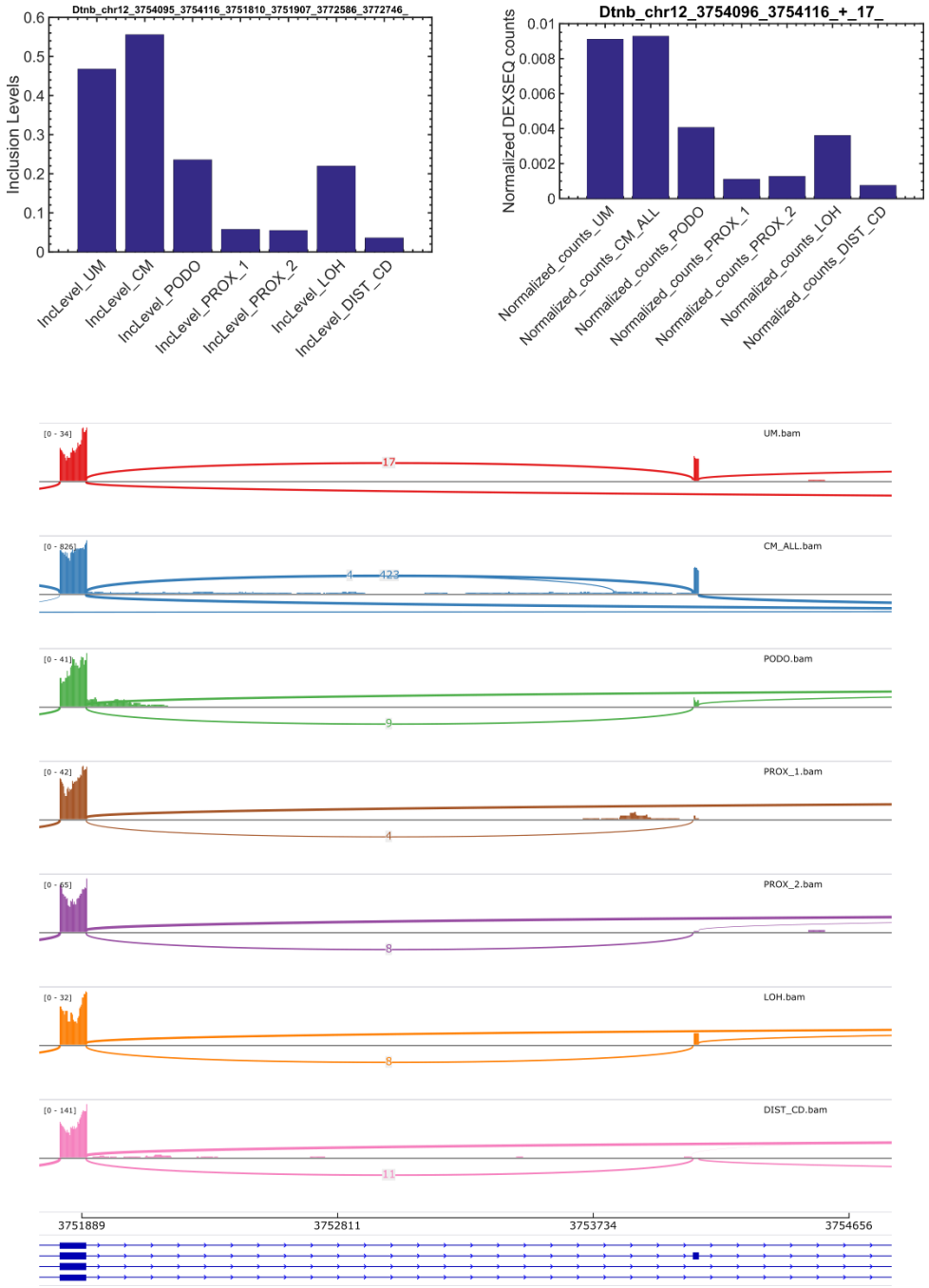


Figure S30: The gene Dtnb undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Gyk

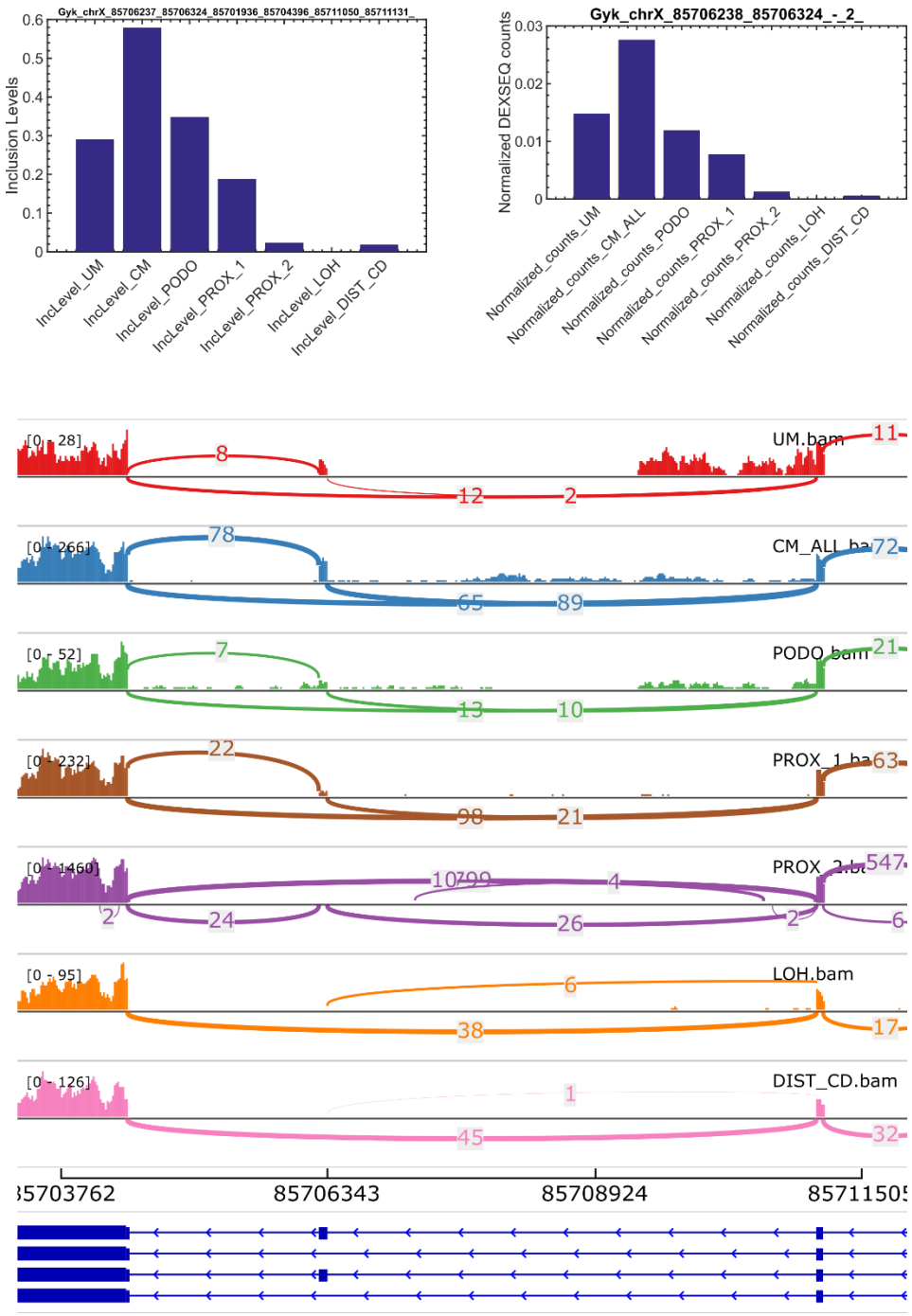


Figure S31: The gene Gk undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Lrrfip1

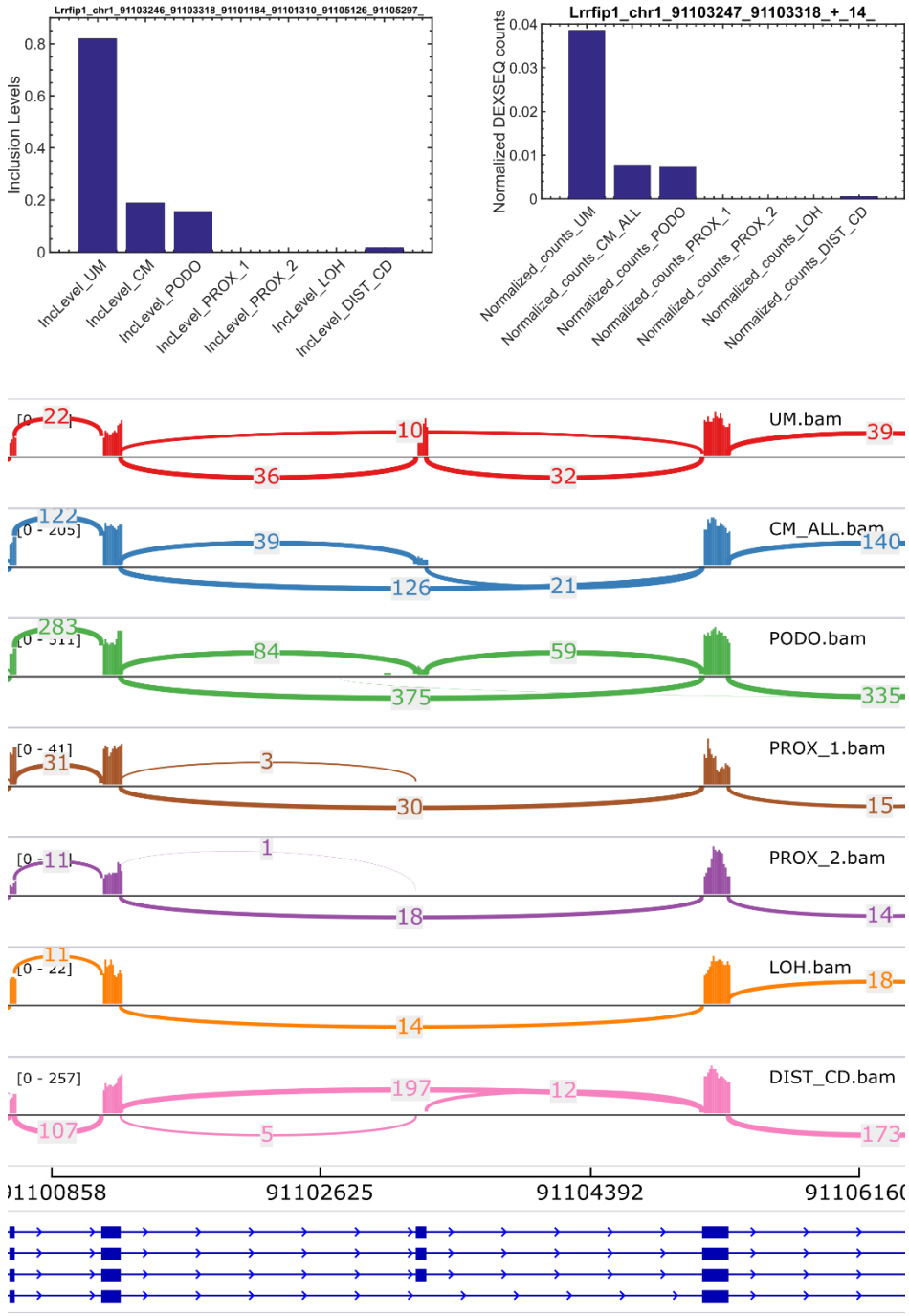


Figure S32: The gene Lrrfip1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Map3k7

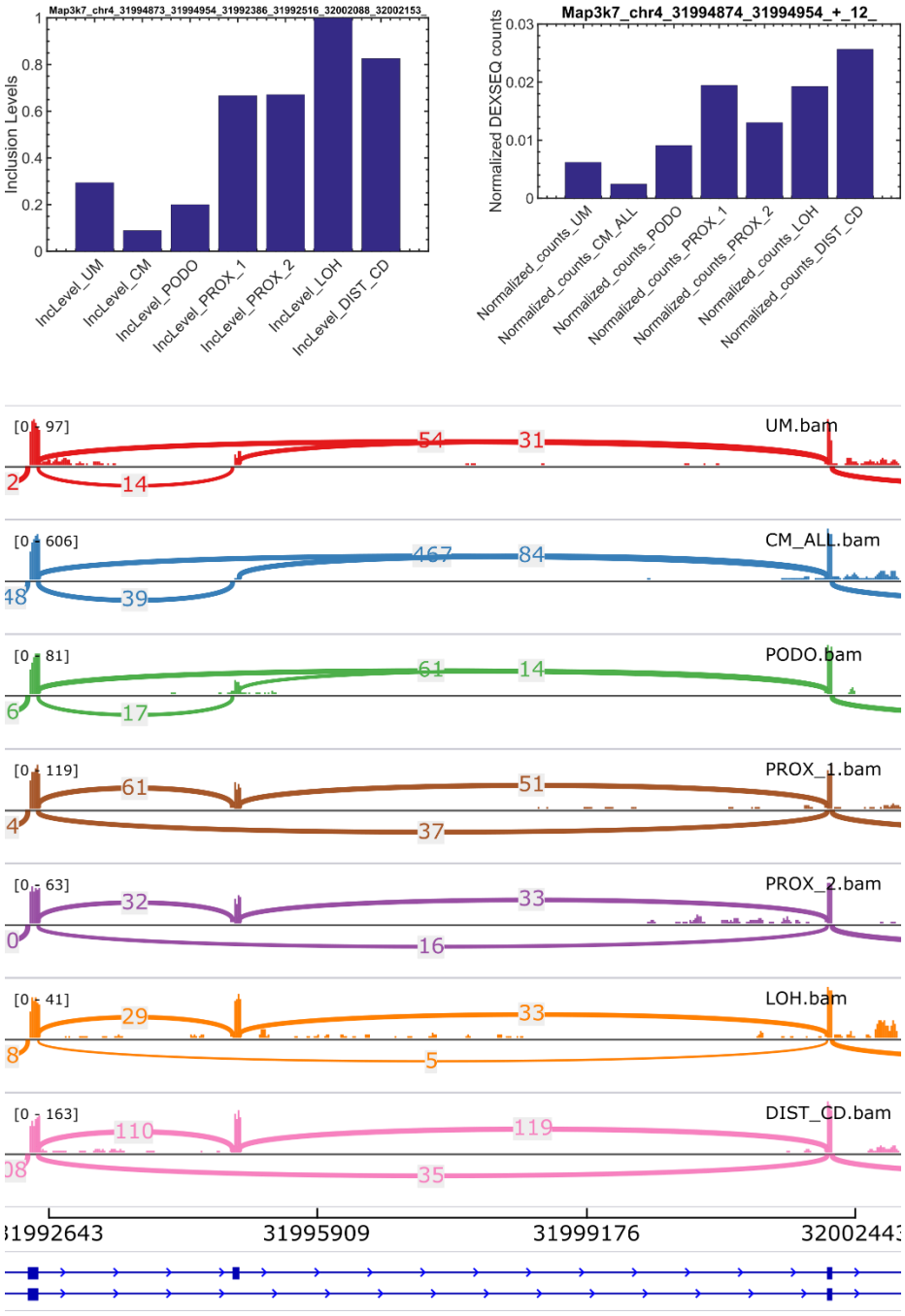


Figure S33: The gene Map3k7 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Mark3

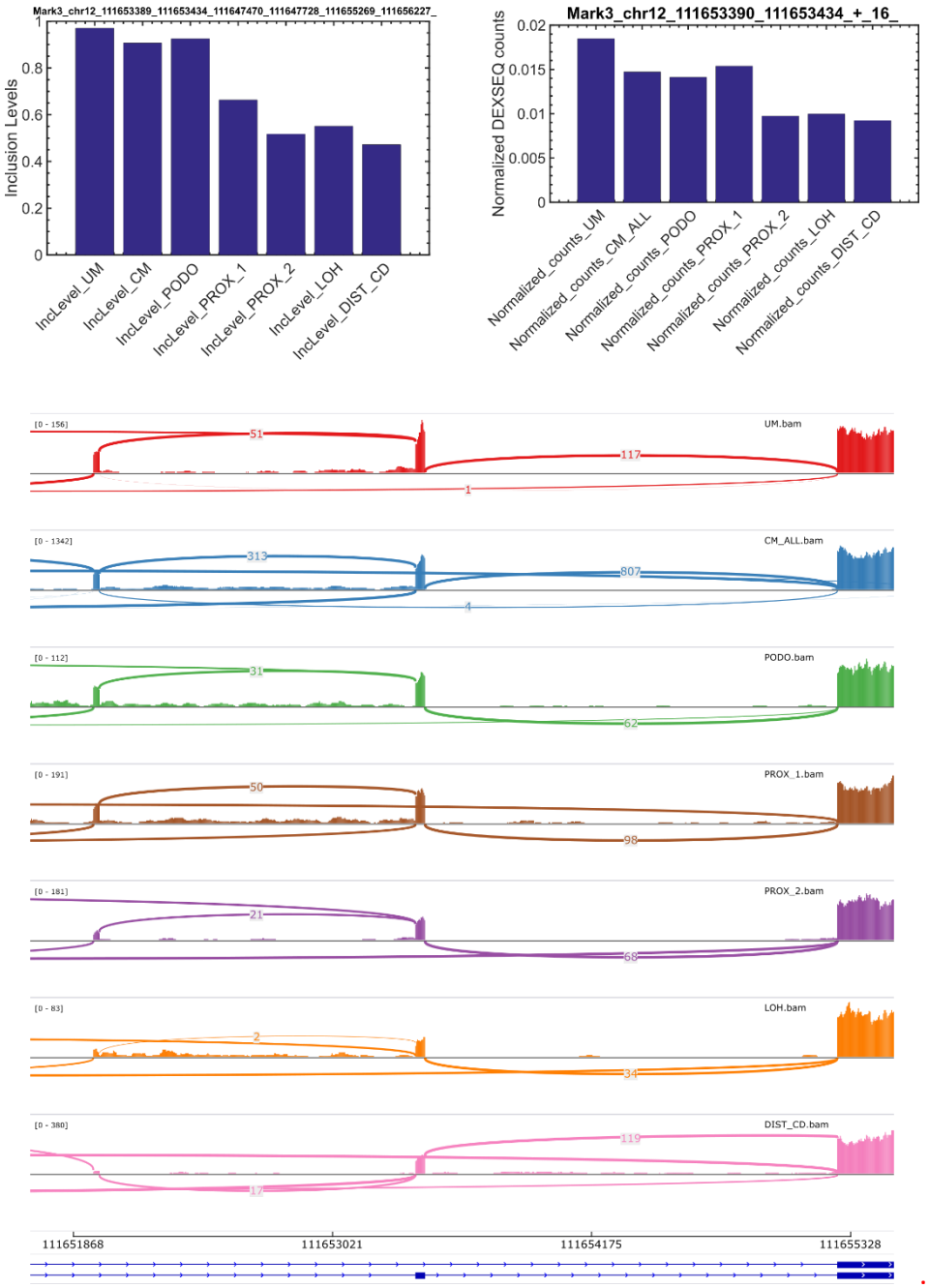


Figure S34: The gene Mark3 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Mprip

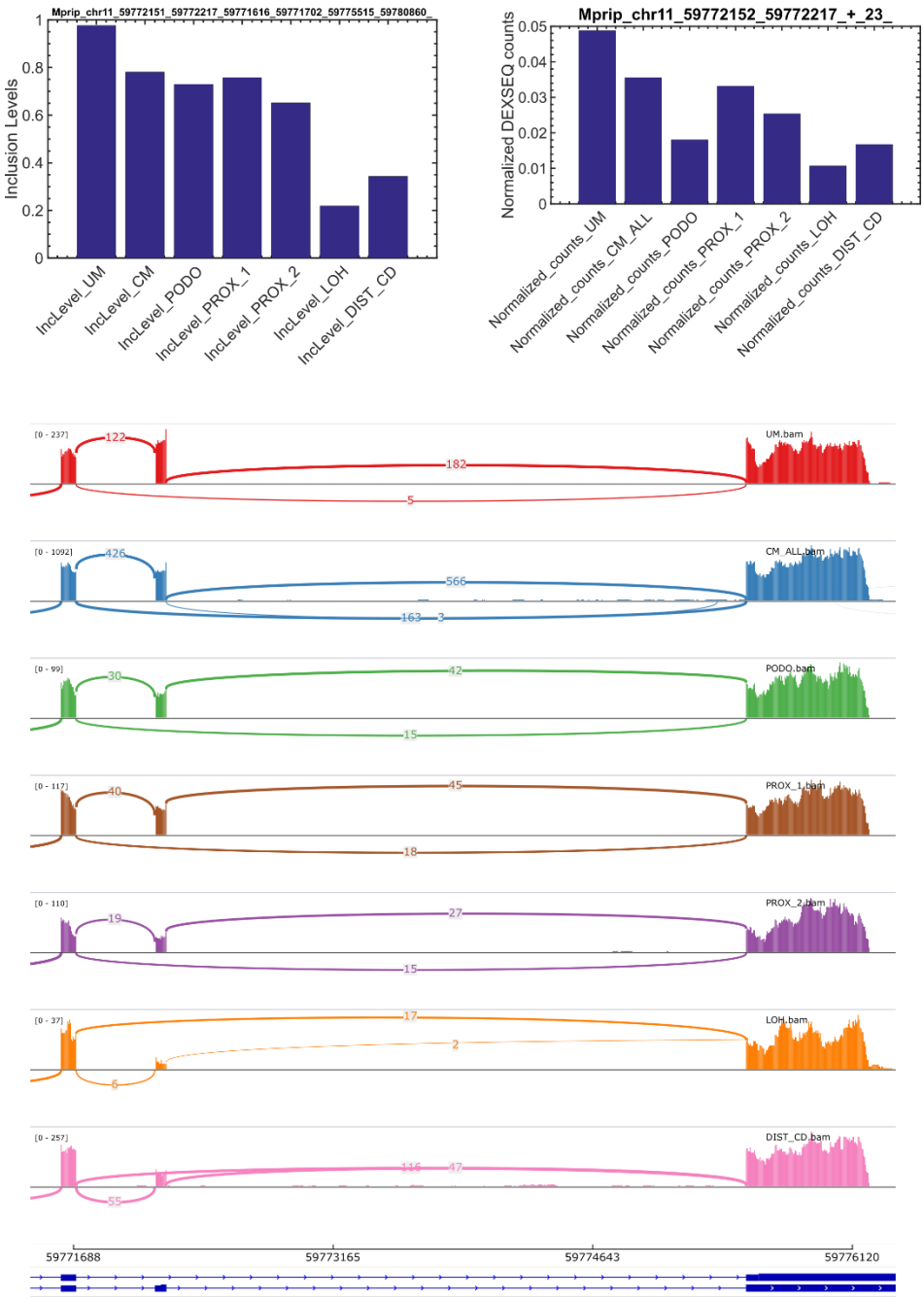


Figure S35: The gene Mprip undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Myo1b

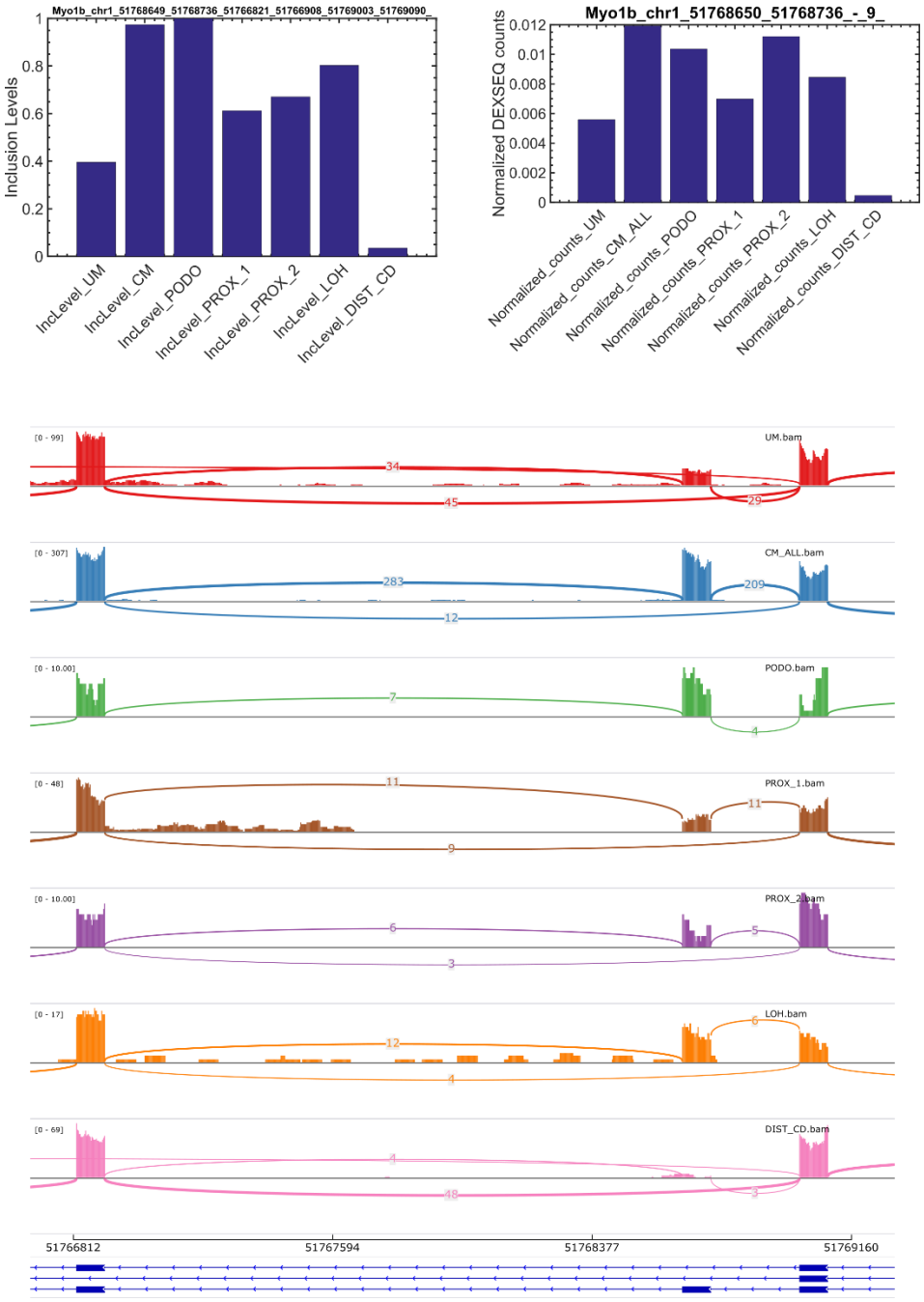


Figure S36: The gene Myo1b undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Nadkd1

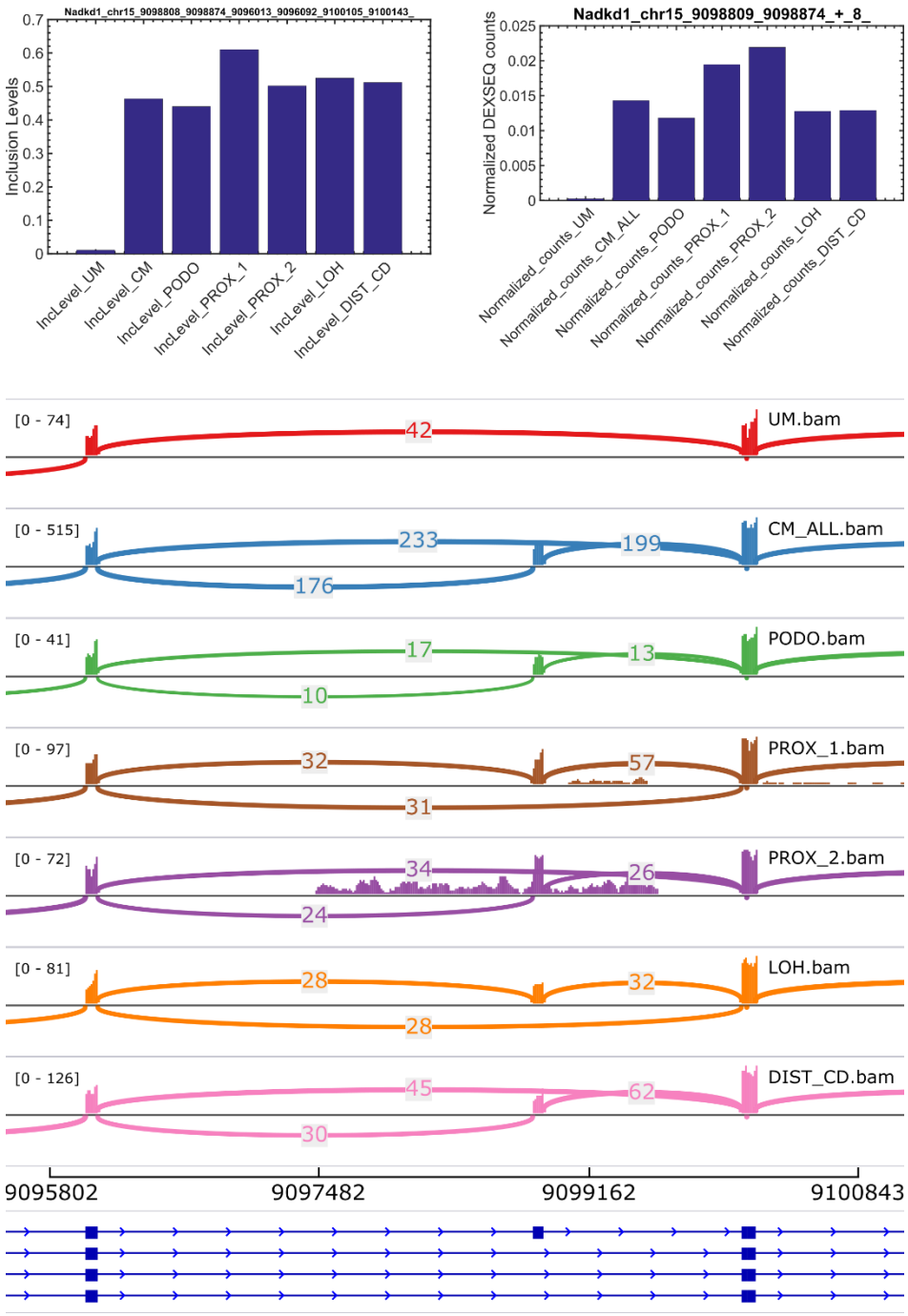


Figure S37: The gene Nadkd1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Osbpl9

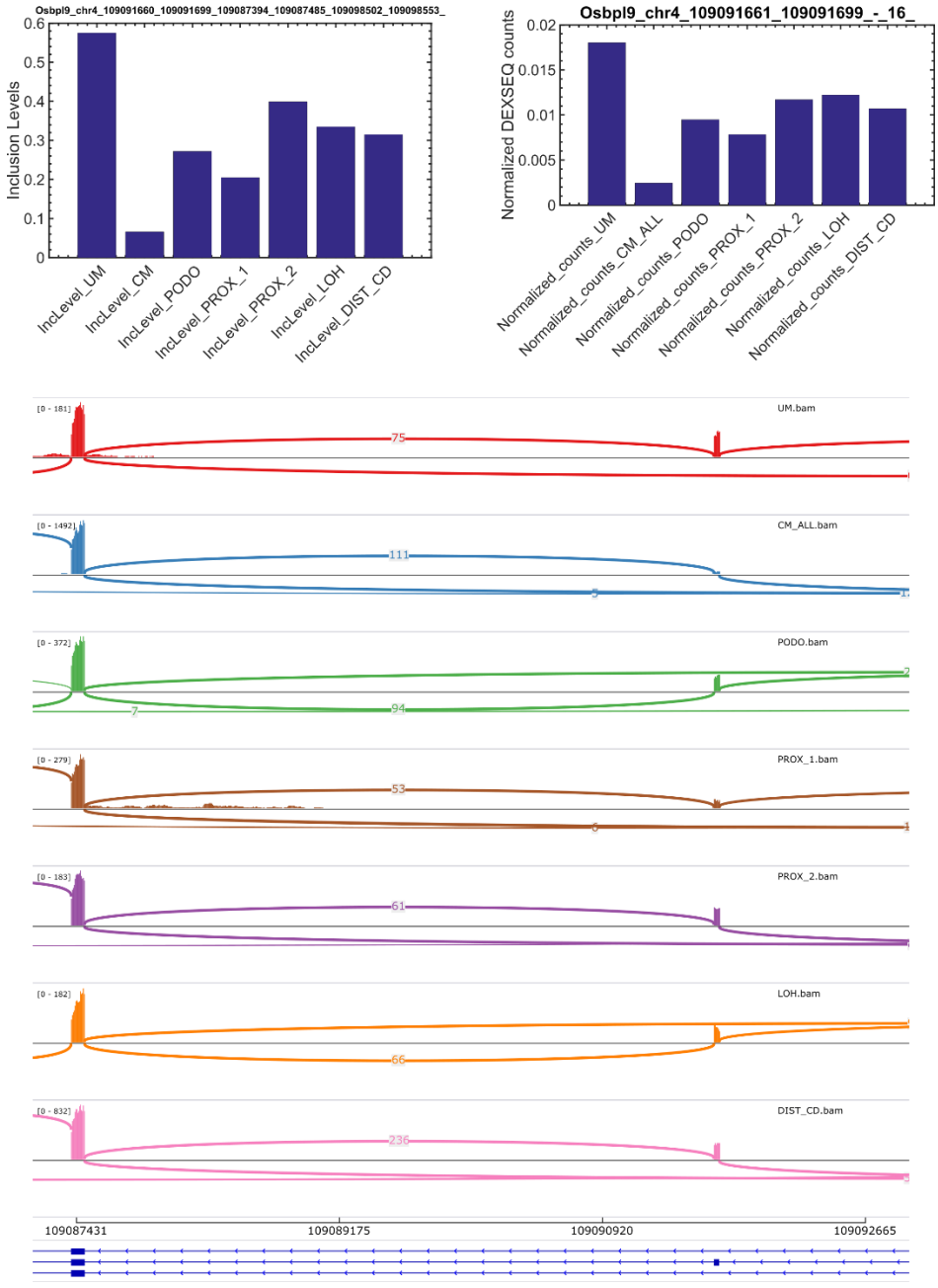


Figure S38: The gene Osbpl9 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Pard3

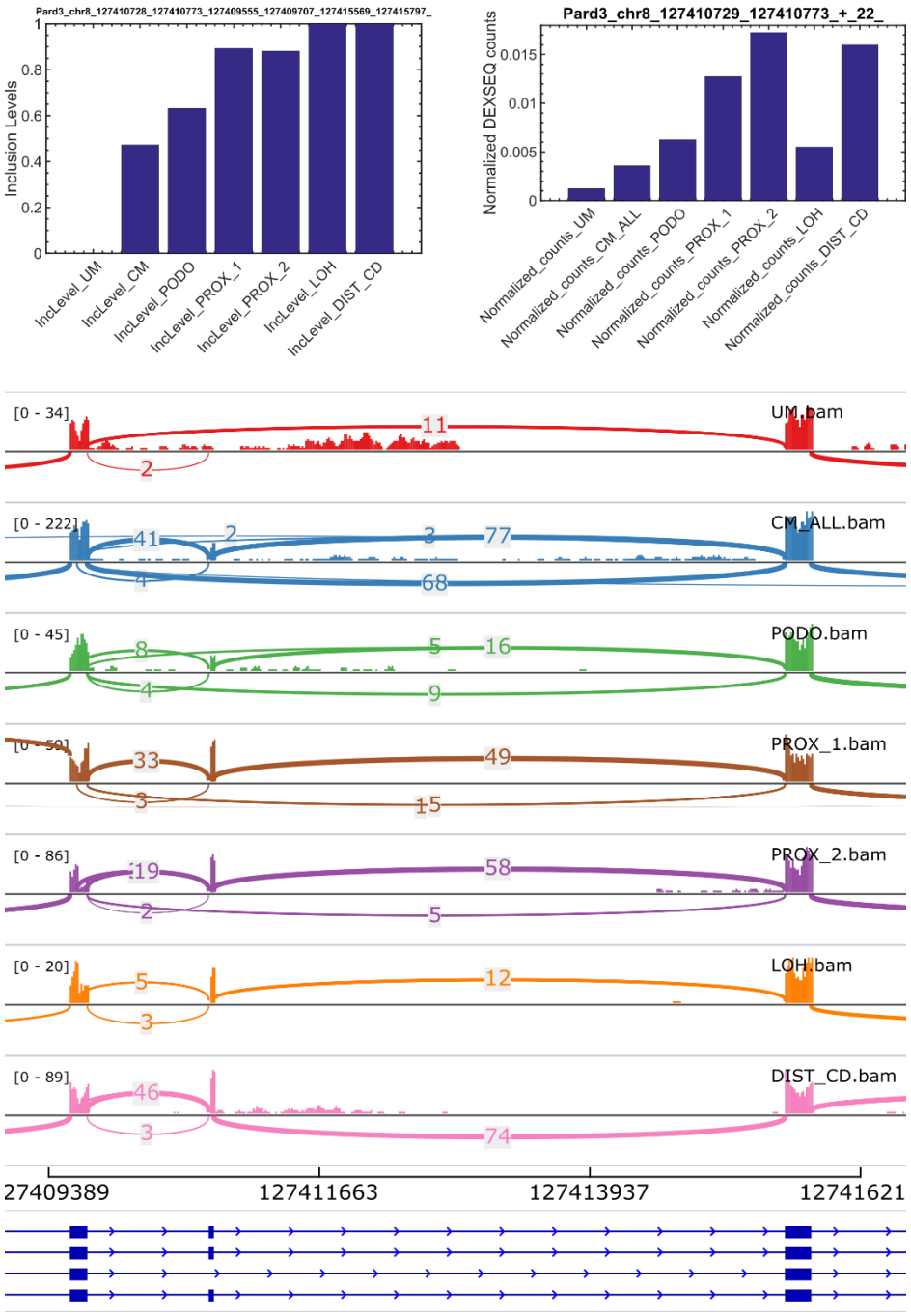


Figure S39: The gene Pard3 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Plod2

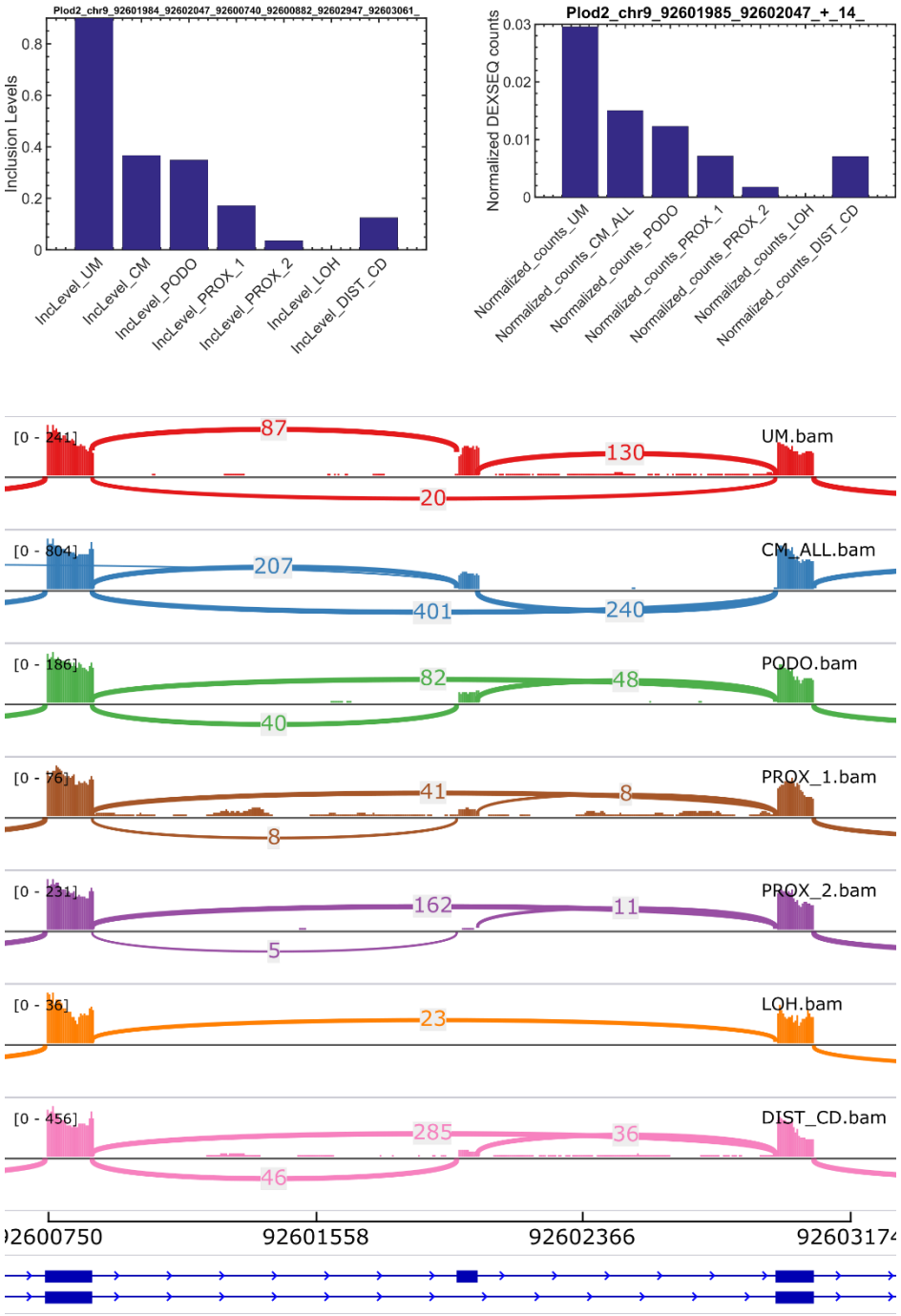


Figure S40: The gene Plod2 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Rps24

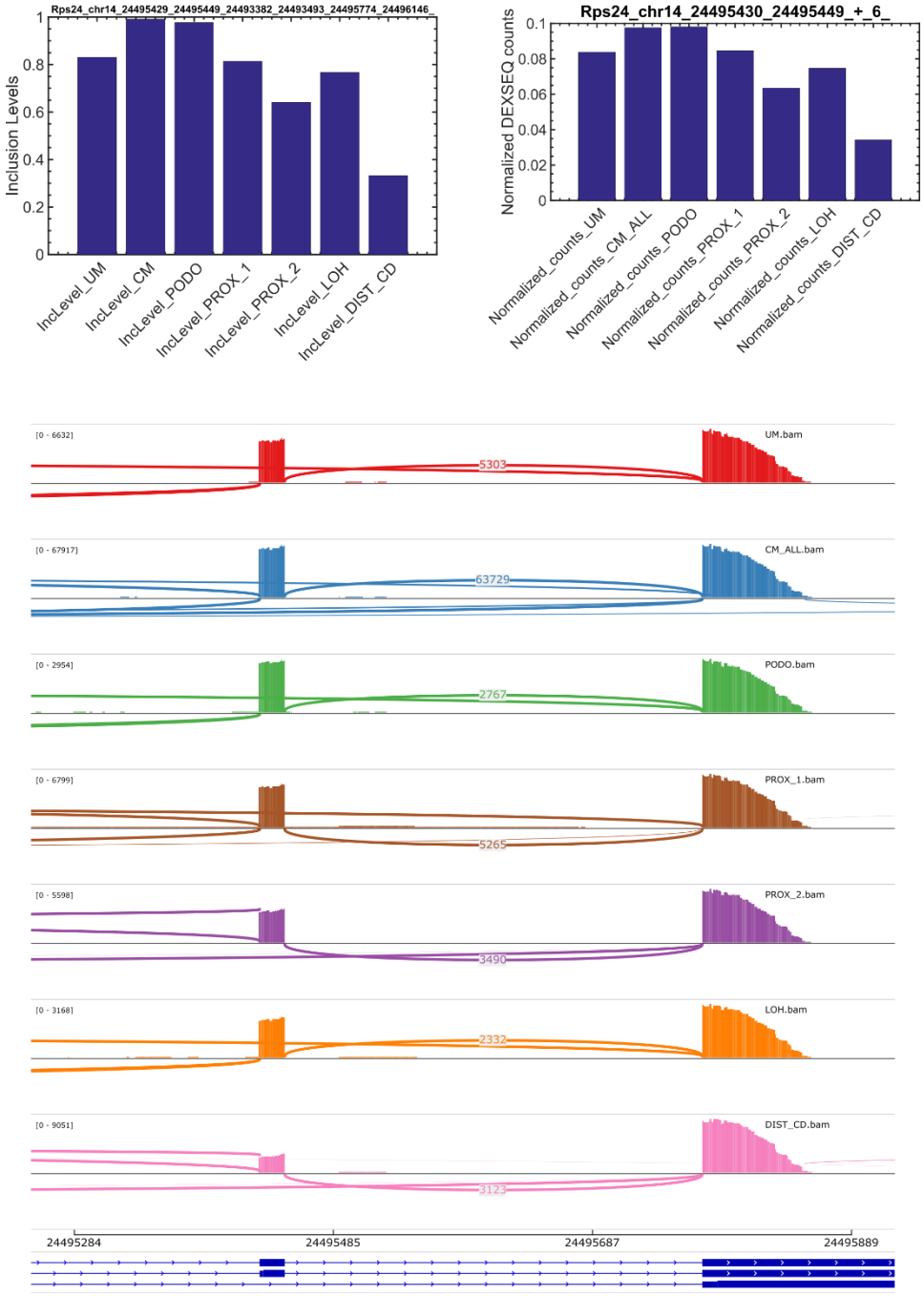


Figure S41: The gene Rps24 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively spliced exon (skipped exon, SE), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exon.

Actn1

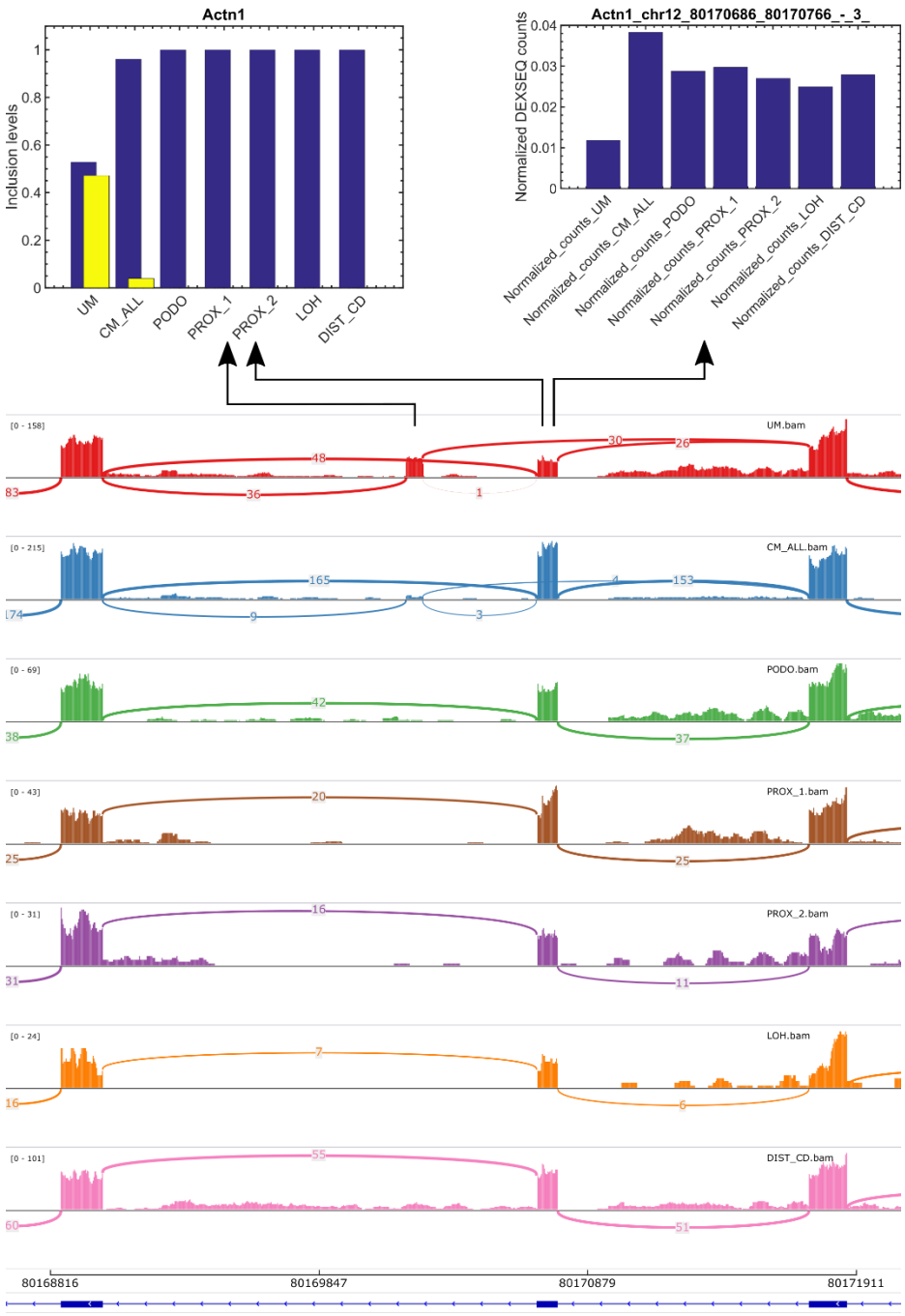


Figure S42: The gene Actn1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, bar plots of inclusion levels of the alternatively spliced mutually exclusive exons (MXE, calculated manually), and bar plot of normalized DEXSeq counts of reads that align to one of the alternatively spliced exons.

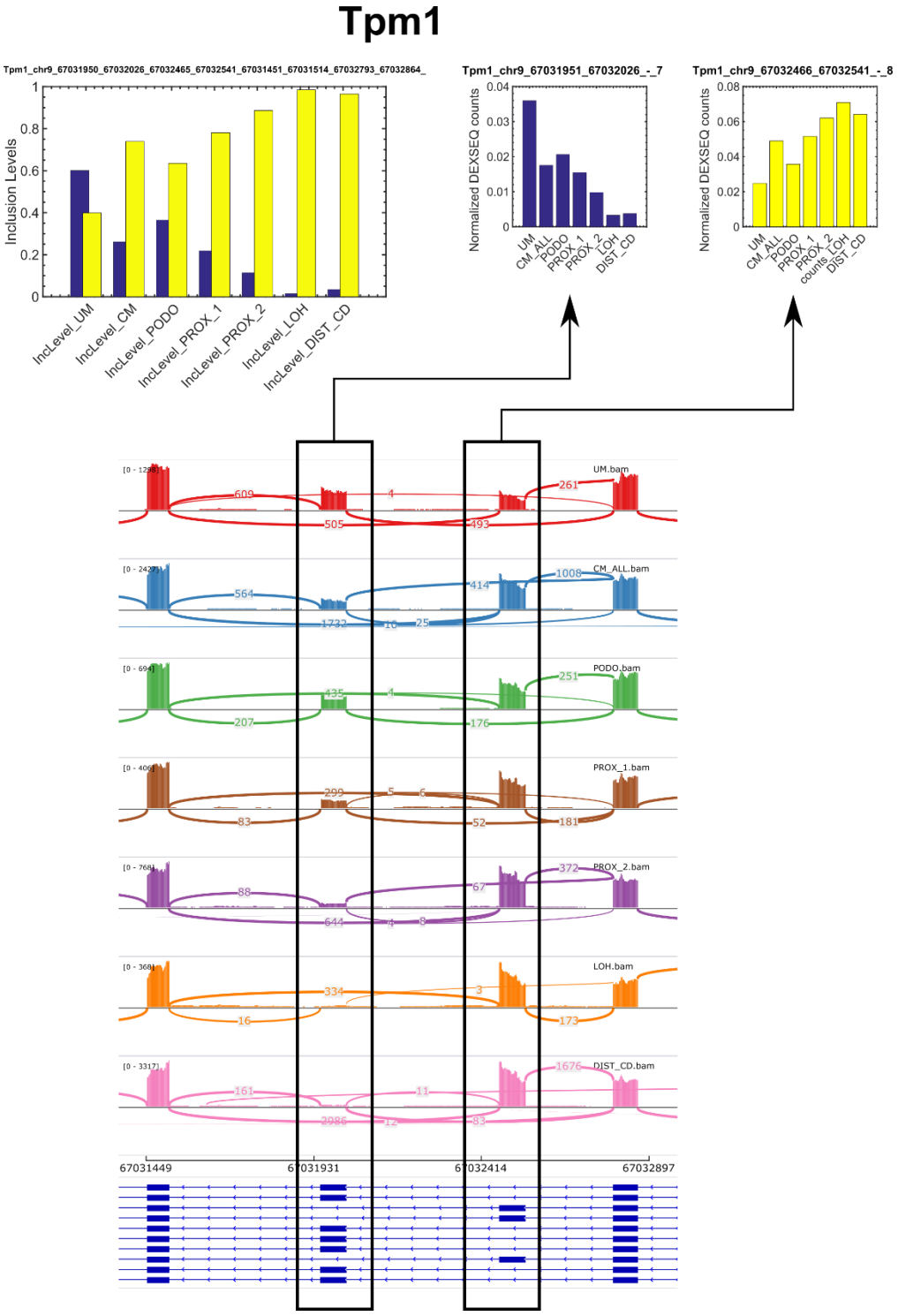


Figure S43: The gene Tpm1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, bar plots of inclusion levels of the alternatively spliced mutually exclusive exons (MXE), and bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced exons.

Acsl4

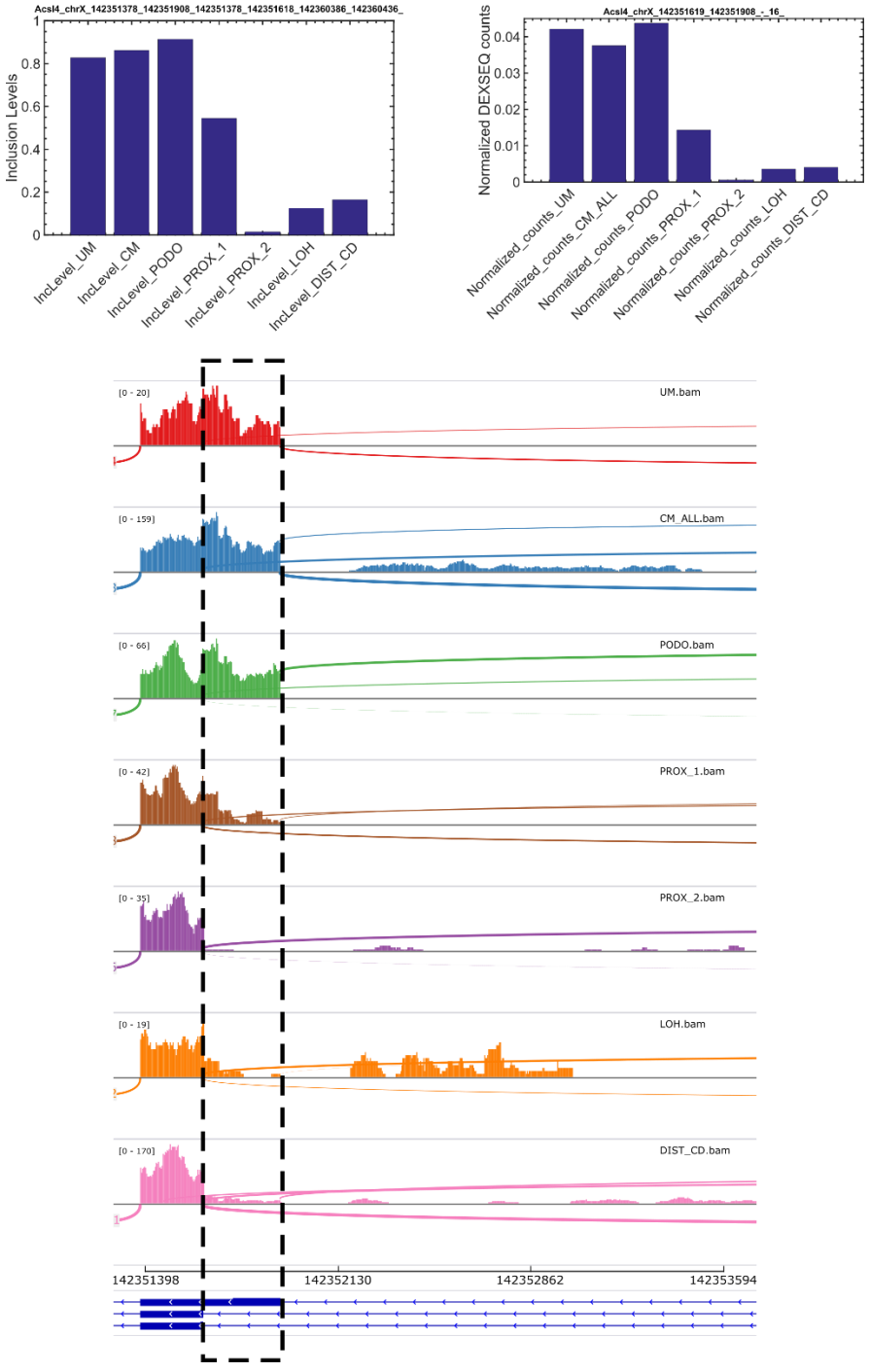


Figure S44: The gene Acsl4 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternative 3' splice site (A3SS), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced segment of the exon.

Bmp1

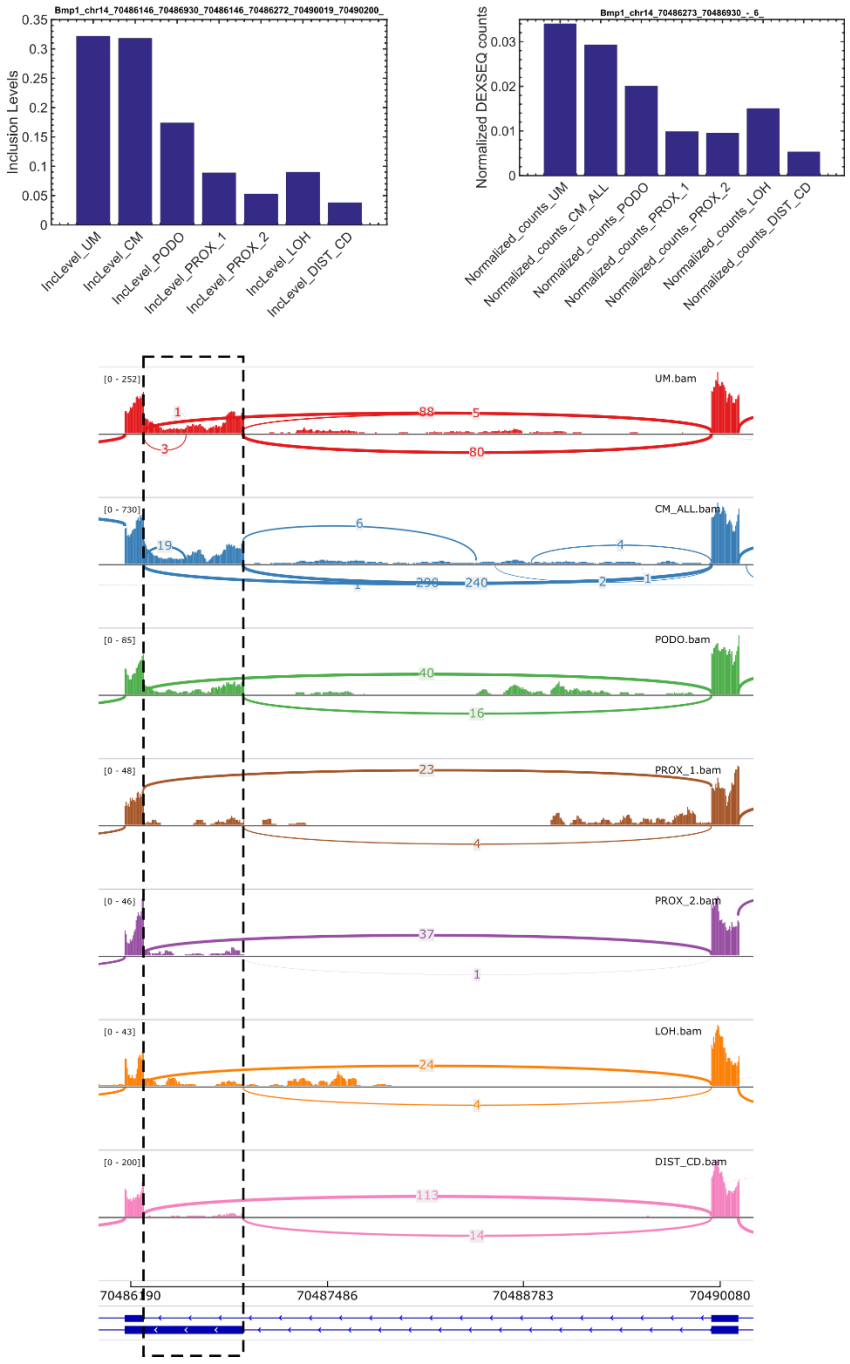


Figure S45: The gene Bmp1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternative 3' splice site (A3SS), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced segment of the exon.

Polr2k

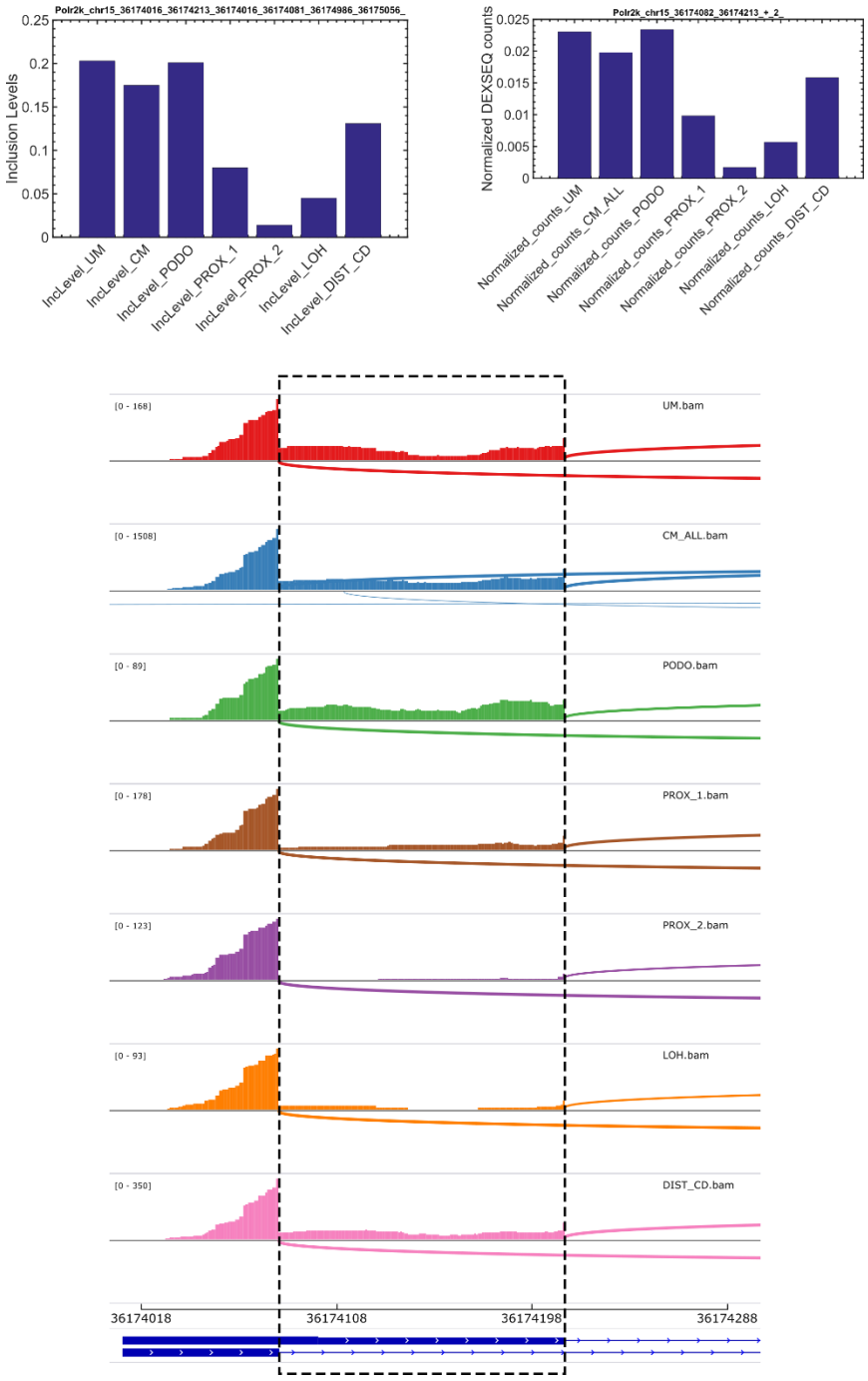


Figure S46: The gene Polr2k undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternative 5' splice site (A5SS), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively spliced segment of the exon.

Srsf1

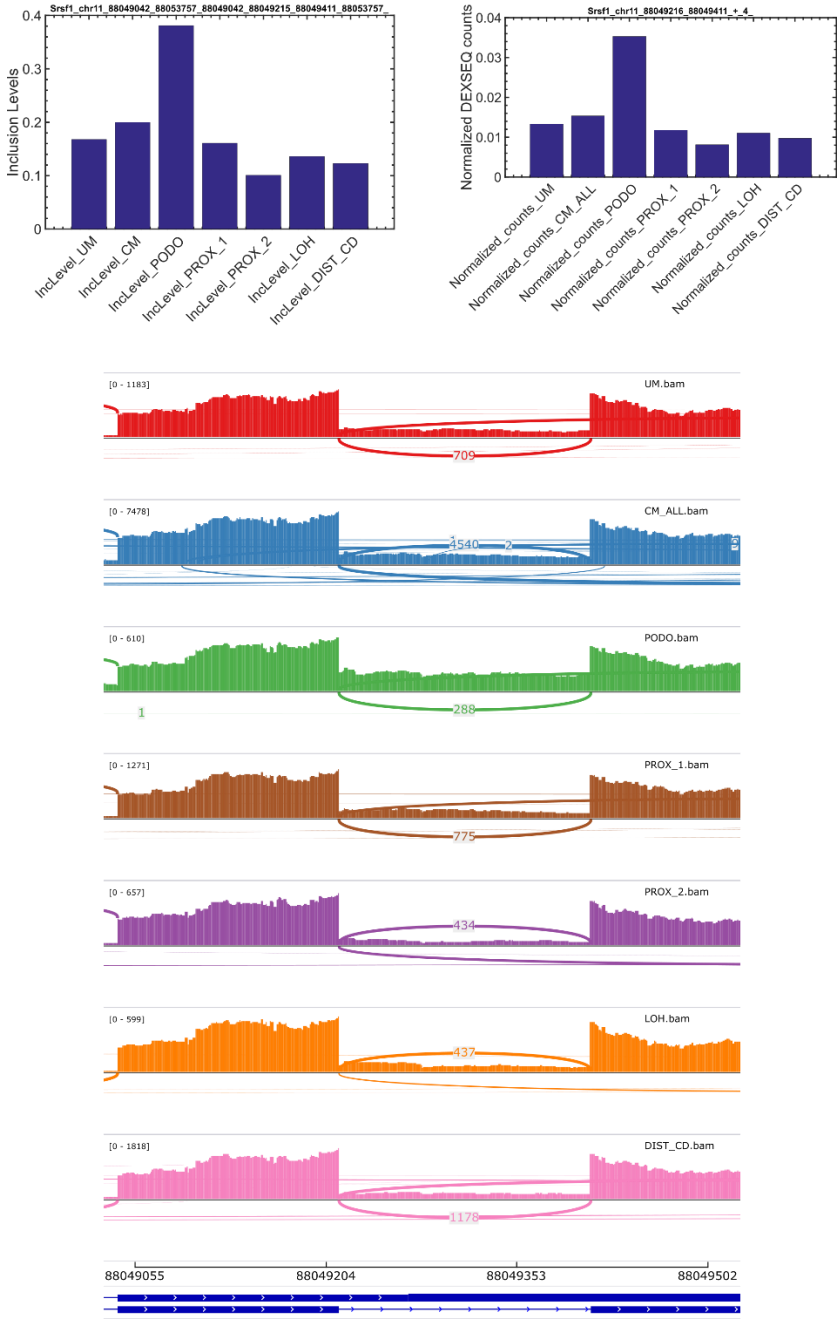


Figure S47: The gene Srsf1 undergoes splice isoform switching during kidney development. Shown is a Sashimi plot, a bar plot of inclusion levels of the alternatively retained intron (RI), and a bar plot of normalized DEXSeq counts of reads that align to the alternatively retained intron. The intron is expressed at high levels in the podocytes, but does not differ much between the mesenchymal and epithelial cell populations.

REFERENCES

1. Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014;111: E5593-601. doi:10.1073/pnas.1419161111
2. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7: 1009–1015. doi:10.1038/nmeth.1528
3. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012;22: 2008–2017. doi:10.1101/gr.133744.111.Freely
4. Hohenstein P, Hastie ND. The many facets of the Wilms' tumour gene, WT1. Hum Mol Genet. 2006;15: 196–201. doi:10.1093/hmg/ddl196
5. Ozdemir DD, Hohenstein P. Wt1 in the kidney - A tale in mouse models. Pediatr Nephrol. 2014;29: 687–693. doi:10.1007/s00467-013-2673-7
6. Hastie ND. The genetics of Wilms' tumor--a case of disrupted development. Annu Rev Genet. 1994;28: 523–58. doi:10.1146/annurev.ge.28.120194.002515
7. Lefebvre J, Clarkson M, Massa F, Bradford ST, Charlet A, Buske F, et al. Alternatively spliced isoforms of WT1 control podocyte-specific gene expression. Kidney Int. Elsevier Masson SAS; 2015;88: 321–331. doi:10.1038/ki.2015.140
8. Haber DA, Sohn RL, Buckler AJ, Pelletier J, Call KM, Housman DE. Alternative splicing and genomic structure of the Wilms tumor gene WT1. Proc Natl Acad Sci U S A. 1991;88: 9618–9622. doi:10.1073/pnas.88.21.9618
9. Laity JH, Dyson HJ, Wright PE. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. Proc Natl Acad Sci. 2002;97: 11932–11935. doi:10.1073/pnas.97.22.11932
10. Larsson SH, Charlieu JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, Ross A, et al. Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. Cell. 1995;81: 391–401. doi:10.1016/0092-8674(95)90392-5
11. Barbaux S, Niaudet P, Gubler MC, Grünfeld JP, Jaubert F, Kuttenn FF, et al. Donor splice-site mutations in WT1 are responsible for Frasier syndrome. Nat Genet. 1997;17: 467–470. doi:10.1038/ng1297-467
12. Hammes A, Guo JK, Lutsch G, Leheste JR, Landrock D, Ziegler U, et al. Two splice variants of the wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. Cell. 2001; doi:10.1016/S0092-8674(01)00453-6
13. Menke AL, Schedl A. WT1 and glomerular function. Semin Cell Dev Biol. 2003;14: 233–240. doi:10.1016/S1084-9521(03)00026-0
14. Schell C, Wanner N, Huber TB. Glomerular development - Shaping the multi-cellular filtration unit. Semin Cell Dev Biol. Elsevier Ltd; 2014;36: 39–49. doi:10.1016/j.semcdb.2014.07.016
15. Little MH, McMahon AP. Mammalian kidney development: Principles, progress, and projections. Cold Spring Harb Perspect Biol. 2012;4: 3. doi:10.1101/cshperspect.a008300

16. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. Elsevier; 2015;161: 1202–1214. doi:10.1016/j.cell.2015.05.002
17. Pode-Shakked N, Gershon R, Tam G, Omer D, Gnatek Y, Kanter I, et al. Evidence of In Vitro Preservation of Human Nephrogenesis at the Single-Cell Level. Stem Cell Reports. ElsevierCompany.; 2017;9. doi:10.1016/j.stemcr.2017.04.026
18. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8
19. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. BioMed Central Ltd; 2010;11: R106. doi:10.1186/gb-2010-11-10-r106
20. Brunskill EW, Aronow BJ, Georgas K, Rumballe B, Valerius MT, Aronow J, et al. Atlas of Gene Expression in the Developing Kidney at Microanatomic Resolution. Dev Cell. Elsevier Ltd; 2008;15: 781–791. doi:10.1016/j.devcel.2008.09.007
21. Magella B, Adam M, Potter AS, Venkatasubramanian M, Chetal K, Hay SB, et al. Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. Dev Biol. Elsevier Inc.; 2017;434: 1–12. doi:10.1016/j.ydbio.2017.11.006
22. Brunskill EW, Park J-S, Chung E, Chen F, Magella B, Potter SS. kh. Development. 2014;141: 3093–3101. doi:10.1242/dev.110601
23. Adam M, Potter AS, Potter SS. Psychrophilic proteases dramatically reduce single cell RNA-seq artifacts: A molecular atlas of kidney development. Development. 2017;1: dev.151142. doi:10.1242/dev.151142
24. Little M, Georgas K, Pennisi D, Wilkinson L. Kidney Development: Two Tales of Tubulogenesis. In: Thornhill BA, Chevalier RL, editors. Current Topics in Developmental Biology. 2010. pp. 193–229. doi:10.1016/S0070-2153(10)90005-7
25. Kobayashi A, Valerius MT, Mugford JW, Carroll TJ, Self M, Oliver G, et al. Six2 Defines and Regulates a Multipotent Self-Renewing Nephron Progenitor Population throughout Mammalian Kidney Development. Cell Stem Cell. 2008;3: 169–181. doi:10.1016/j.stem.2008.05.020
26. Cho E a, Patterson LT, Brookhiser WT, Mah S, Kintner C, Dressler GR. Differential expression and function of cadherin-6 during renal epithelium development. Development. 1998;125: 803–812. Available: http://www.ncbi.nlm.nih.gov/pubmed/9449663
27. Lindström NO, Lawrence ML, Burn SF, Johansson J a, Bakker ER, Ridgway R a, et al. Integrated β-catenin, BMP, PTEN, and Notch signalling patterns the nephron. Elife. 2015;4: 1–29. doi:10.7554/eLife.04000
28. Brown AC, Deepthi S, Oxburgh L, Brown AC, Muthukrishnan SD, Oxburgh L. A Synthetic Niche for Nephron Progenitor Cells Technology A Synthetic Niche for Nephron Progenitor Cells. Dev Cell. 2015;34: 229–241. doi:10.1016/j.devcel.2015.06.021
29. Iglesias DM, Hueber P-A, Chu L, Campbell R, Patenaude A-M, Dziarmaga AJ, et al. Canonical WNT signaling during kidney development. Am J Physiol Renal Physiol.

2007;293: F494-500. doi:10.1152/ajprenal.00416.2006
30. Little MH, Hale LJ, Howden SE, Kumar S V. Generating Kidney from Stem Cells. Annu Rev Physiol. 2019;81: 335–357. doi:10.1146/annurev-physiol-020518-114331
31. Leroy X, Copin MC, Devisme L, Buisine MP, Aubert JP, Gosselin B, et al. Expression of human mucin genes in normal kidney and renal cell carcinoma. Histopathology. 2002;40: 450–457. doi:10.1046/j.1365-2559.2002.01408.x
32. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. Springer US; 2018;560: 494–498. doi:10.1038/s41586-018-0414-6
33. Yang Y, Park JW, Bebee TW, Warzecha CC, Guo Y, Shang X, et al. Determination of a Comprehensive Alternative Splicing Regulatory Network and Combinatorial Regulation by Key Factors during the Epithelial-to-Mesenchymal Transition. Mol Cell Biol. 2016;36: 1704–1719. doi:10.1128/MCB.00019-16
34. Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA Biol. 2009;6: 546–562. doi:10.4161/rna.6.5.9606
35. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 Are Epithelial Cell-Type-Specific Regulators of FGFR2 Splicing. Mol Cell. Elsevier Ltd; 2009;33: 591–601. doi:10.1016/j.molcel.2009.01.025
36. Bates CM. Role of fibroblast growth factor receptor signaling in kidney development. Pediatr Nephrol. 2011;26: 1373–1379. doi:10.1007/s00467-010-1747-z
37. Sims-Lucas S, Cusack B, Baust J, Eswarakumar VP, Masatoshi H, Takeuchi A, et al. Fgfr1 and the IIIc isoform of Fgfr2 play critical roles in the metanephric mesenchyme mediating early inductive events in kidney development. Dev Dyn. 2011;240: 240–249. doi:10.1002/dvdy.22501
38. Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, et al. An emt-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS Genet. 2011;7. doi:10.1371/journal.pgen.1002218
39. Bebee TW, Park JW, Sheridan KI, Warzecha CC, Cieply BW, Rohacek AM, et al. The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. Elife. 2015;4: 1–27. doi:10.7554/eLife.08954
40. Di Modugno F, Iapicca P, Boudreau A, Mottolese M, Terrenato I, Perracchio L, et al. Splicing program of human MENA produces a previously undescribed isoform associated with invasive, mesenchymal-like breast tumors. Proc Natl Acad Sci U S A. 2012;109: 19280–5. doi:10.1073/pnas.1214394109
41. Brown RL, Reinke LM, Damerow MS, Perez D, Chodosh L a., Yang J, et al. CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. J Clin Invest. 2011;121: 1064–1074. doi:10.1172/JCI44540
42. Sneath RJ, Mangham DC. The normal structure and function of CD44 and its role in neoplasia. Mol Pathol. 1998;51: 191–200. doi:10.1136/mp.51.4.191

43. Van Itallie CM, Rogan S, Yu A, Vidal LS, Holmes J, Anderson JM. Two splice variants of claudin-10 in the kidney create paracellular pores with different ion selectivities. Am J Physiol Renal Physiol. 2006;291: F1288–F1299. doi:10.1152/ajprenal.00138.2006
44. Park JW, Jung S, Rouchka EC, Tseng YT, Xing Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. Nucleic Acids Res. 2016;44: W333–W338. doi:10.1093/nar/gkw410
45. Dittmar KA, Jiang P, Park JW, Amirikian K, Wan J, Shen S, et al. Genome-Wide Determination of a Broad ESRP-Regulated Posttranscriptional Network by High-Throughput Sequencing. Mol Cell Biol. 2012;32: 1468–1482. doi:10.1128/MCB.06536-11
46. Bhate A, Parker DJ, Bebee TW, Ahn J, Arif W, Rashan EH, et al. ESRP2 controls an adult splicing programme in hepatocytes to support postnatal liver maturation. Nat Commun. Nature Publishing Group; 2015;6: 8768. doi:10.1038/ncomms9768
47. Bangru S, Arif W, Seimetz J, Bhate A, Chen J, Rashan EH, et al. Alternative splicing rewires Hippo signaling pathway in hepatocytes to promote liver regeneration. Nat Struct Mol Biol. 2018;25: 928–939. doi:10.1038/s41594-018-0129-2