A cell atlas of the fly kidney

Jun Xu[1, #], Yifang Liu[1, #], Hongjie Li[2,9], Alexander J. Tarashansky[3,4], Colin H. Kalicki[3], Ruei-Jiun Hung[1], Yanhui Hu[1], Aram Comjean[1], Sai Saroja Kolluru[3,4], Bo Wang[3,5], Stephen R Quake[3,4], Liqun Luo[2], Andrew P. McMahon[6], Julian A.T. Dow[7], Norbert Perrimon[1,8, *]

[1] Department of Genetics, Blavatnik Institute, Harvard Medical School, Harvard University, Boston, MA 02115, USA

[2] Department of Biology, Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA

[3] Department of Bioengineering, Stanford University, Stanford, CA, USA

[4] Chan Zuckerberg Biohub, San Francisco, CA, USA

[5] Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA

[6] Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90089, USA

[7] Institute of Molecular, Cell & Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, UK

[8] Howard Hughes Medical Institute, Boston, MA, USA

[9] Present Address: Huffington Center on Aging, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[#] co-first authors
*Correspondence: perrimon@genetics.med.harvard.edu

SUMMARY

Like humans, insects rely on precise regulation of their internal environments to survive. The insect renal system consists of Malpighian tubules and nephrocytes that share similarities to the mammalian kidney. Studies of the *Drosophila* Malpighian tubules and nephrocytes have provided many insights into our understanding of the excretion of waste products, stem cell regeneration, protein reabsorption, and as human kidney disease models. Here, we analyzed single-nucleus RNA sequencing (snRNA-seq) data sets to characterize the cell types of the adult fly kidney. We identified 11 distinct clusters representing renal stem cells (RSCs), stellate cells (SCs), regionally specific principal cells (PCs), garland nephrocyte cells (GCs) and pericardial nephrocytes (PNs). Analyses of these clusters revealed many new interesting features. For example, we found a new, previously unrecognized cell cluster: lower segment PCs

that express *Esyt2*. In addition, we find that the SC marker genes *RhoGEF64c*, *Frq2*, *Prip* and *CG10939* regulate their unusual cell shape. Further, we identified transcription factors specific to each cluster and built a network of signaling pathways that are potentially involved in mediating cell-cell communication between Malpighian tubule cell types. Finally, cross-species analysis allowed us to match the fly kidney cell types to mouse kidney cell types and planarian protonephridia - knowledge that will help the generation of kidney disease models. To visualize this dataset, we provide a web-based resource for gene expression in single cells (https://www.flyrnai.org/scRNA/kidney/). Altogether, our study provides a comprehensive resource for addressing gene function in the fly kidney and future disease studies.

Key words: snRNA-seq; *Drosophila*; Malpighian tubules; nephrocytes; cross-species; kidney disease

INTRODUCTION

The functions of excretory systems are to remove toxins from the body and maintain homeostatic balance. For example, mammalian kidneys play important roles in several physiological processes, including maintaining water fluid homeostasis, removing metabolic waste products, controlling blood pressure, regulating blood cell composition, and regulating bone mineralization (Nielsen et al., 2012). Although the excretory systems of various animals have differences, they typically have in common two activities: filtration and tubular secretion/reabsorption (Denholm and Skaer, 2009). In mammals, the mature kidney consists of two connected parts: a nephron, derived from the metanephric mesoderm, and a collecting tubule derived from the ureteric bud (Nielsen et al., 2012; McMahon, 2016). The *Drosophila* renal system is composed of separated filtration nephrocytes and Malpighian (renal) tubules (Miller et al., 2013). In *Drosophila*, about 25 garland cell nephrocytes (GCs) and 120 pericardial nephrocytes (PNs) are found at the end of embryogenesis and are maintained during development into the adult stage. GCs form a ring around the junction between the proventriculus and esophagus, whereas PNs are located on both side of the heart tube (Helmstädter et al., 2017). These two types of nephrocytes, although derived from different cell lineages, share morphological, functional, and molecular features with podocytes, which form the glomerular filter in vertebrates, and possess a protein sequestration activity reminiscent of the proximal tubule (Helmstädter et al., 2017). The adult Malpighian tubules, considered to be analogous to the renal tubular system, develop from the ectodermal hindgut primordium and visceral mesoderm and consist of two pairs of epithelial tubes that empty into the hindgut at its junction with the posterior midgut (Jung et al., 2005).

 Drosophila Malpighian tubules and nephrocytes have been used to model human kidney diseases. Previously, a screen for genes involved in renal function identified over 70 genes required for nephrocyte function (Zhang et al., 2013). In addition, 30 human causative genes involved in Steroid resistant nephrotic syndrome (SRNS) have been analyzed in fly nephrocytes. Among them, *Cubilin* (*Cubn*) was found to be

required for nephrocytes endocytosis (Hermle et al., 2017). Further, the coenzyme Q10 (CoQ10) biosynthesis gene *Coq2*, involved in regulating the morphology of slit diaphragm, and ROS formation, contribute to a pathomechanism of COQ2-nephropathy (Hermle et al., 2017). In addition to model numerous human renal conditions such as chronic kidney disease and kidney stones, the Malpighian tubule is also an excellent model in which to study the neuroendocrine control of renal function and rapid fluid transport (Cohen et al., 2020).

Single-nucleus (snRNA-seq) and single-cell (scRNA-seq) RNA sequencing give us an opportunity to understand the cellular make-up of many organ systems, including the kidney. The mammalian kidney is composed of cell types with unique functions. Podocytes regulate the passage of proteins, and the function of principal cells and intercalated cells in the collecting duct balance systemic water, pH, and salt levels (Garg, 2018; Pearce et al., 2015; Roy et al., 2015). A detailed scRNA-seq study defined the whole landscape of the mouse kidney, with 32 distinct clusters of ontology (Ransick et al., 2019). Finally, scRNA-seq data can be used for the analysis of pseudotemporal ordering of cells, which can provide information about developmental trajectories of cellular lineages.

Here, we performed snRNA-seq of the adult *Drosophila* male and female kidney system to characterize the organization, origins and diversity of the various cell types. Specifically, we identified 11 distinct clusters representing renal stem cells (RSCs), stellate cells (SCs), principal cells (PCs), garland nephrocytes cells (GCs) and pericardial nephrocytes (PNs), and provide gene expression level data at single cell resolution. In addition, based on the snRNA-seq data, we identified a set of genes involved in regulating cell shape of SCs. We also analyzed cell-to-cell communication between clusters, cluster-specific transcription factors, and metabolic differences between clusters. Importantly, performing a cross-species analysis between the fly kidney, planarian protonephridia and mouse kidney allowed us to map kidney cell types across species. We also analyzed human kidney disease genes at the cluster level in the fly kidney. Finally, we built a web-based visualization resource (https://www.flyrnai.org/scRNA/kidney/) that allows users to browse snRNA-seq data and query the expression of genes of interest in different cell types.

RESULTS
snRNA-seq identifies 11 distinct clusters in the adult *Drosophila* kidney
The fly kidney consists of Malpighian tubules and nephrocytes that are located in different regions of the body. Malpighian tubules branch from two common ureters that drain into the gut at the midgut/hindgut junction. Nephrocytes represent garland nephrocytes cells (GCs) located near the esophagus and proventriculus, and pericardial nephrocytes (PNs) located in the abdominal tissue (Fig. 1A). As part of the Fly Cell Atlas (FCA) project, we dissected male and female Malpighian tubules (Li et al., 2021) and annotated the cell types. In addition, as nephrocytes were not included in the FCA, we performed snRNA-seq of both GCs and PNs. To visualize GCs and PNs during dissection, we expressed *UAS-GFP.nls* under the control of *Dot-Gal4,* which is expressed in both cell types. In total, 150 male and 150 female tissues were

1 dissected. Subsequently, four independent samples were processed for single nucleus
2 isolation and the mRNAs were barcoded and sequenced (Fig. 1A). We successfully
3 recovered 12,166 cells in the tubules. We also identified nephrocyte cell clusters that
4 include a GC cluster with 41 nuclei and a PN cluster with 93 nuclei. Details on the
5 number of cells and statistics are summarized in Supplementary Table 1.

6 We identified 11 distinct clusters representing renal stem cells (RSCs), stellate cells
7 (SCs), regionally specific principal cells (PCs), and nephrocyte cells (GCs and PNs)
8 (marker genes listed in Supplementary Table 2). Note that the tubule snRNA-seq data
9 were independently annotated at Harvard and FCA with highly concordant results (Fig.
10 S1). We annotated the clusters based on previous knowledge and validation of novel
11 marker genes (Supplementary Table 3). Previous studies have characterized enhancer
12 trap lines that identified six distinct regions and genetically separable cell types in the
13 tubules (Sözen et al., 1997). In addition, we validated new markers, identified as
14 cluster-specific, by driving fluorescent reporters with the appropriate GAL4 lines
15 (Supplementary Table 3).

16 The Malpighian tubule stem cell cluster is defined by the expression of *escargot*
17 (*esg*), *Notch* (*N*) and *Delta* (*Dl*) genes (Wang and Spradling, 2020). The two stellate
18 cell clusters (main segment SCs and bar-shaped SCs) both express *teashirt* (*tsh*), *kinin*
19 *receptor (lkr)* and *Secretory chloride channel* (*SecCl*) (Denholm et al., 2013; Radford
20 et al., 2002; Feingold et al., 2019). As there are no previously reported specific markers
21 of bar-shaped SCs, we characterized I_h *channel* (*ih*) (Fig. S2A). In addition, we
22 validated *CG30377* for main segment SCs and *α2-adrenergic-like octopamine receptor*
23 (*Octα2R*) for all SCs (Fig. S2B and C). We identified six PC clusters (initial and
24 transitional PCs, main segment PCs, lower tubule PCs, upper ureter PCs, lower ureter
25 PCs, and lower segment PCs). Initial and transitional PCs express *bifid* (*bi*) and *Death*
26 *executioner Bcl-2* (*Debcl*) (Fig. 1D). The main segment PCs express *urate oxidase*
27 (*Uro*) (Terhzaz et al., 2010, Fig. 1D). The lower segment PC cluster contains three sub-
28 clusters: lower tubule, upper ureter and lower ureter PCs. *Alkaline phosphatase 4* (*Alp4*)
29 is a known marker of lower segment PCs (Yang et al., 2000, Fig. 1D and 1E), upper
30 ureter PCs express *Sex peptide receptor* (*SPR*) (Fig. 1D and 1E), and lower ureter
31 PCs express *Wnt oncogene analog 4* (*Wnt4*) (Fig. 1D and 1E). Thus, marker genes
32 for lower ureter PC is *Wnt4*; upper ureter is *SPR*; lower tubule is *SPR*, *Alp4*^{hi}; and the
33 main segment PC is *Uro* (Fig. 1F). We also identified a small cell cluster expressing
34 *Extended synaptotagmin-like protein 2* (*Esyt2*) that we named lower segment PC. Note
35 that we find that *Alp4* is expressed in three clusters (the upper ureter PC, lower tubule
36 PC, and lower segment PC clusters) (Fig. 1D and 1E). *Wnt4*, *SPR*, *Esyt2* and *Debcl*
37 are new marker genes that had not been previously reported.

38 Nephrocyte clusters are defined by the expression of *sticks and stones* (*sns*), *Cubn*
39 and *prospero* (*pros*). *sns* encodes a core component of slit diaphragm (Zhuang et al.,
40 2009) and *Cubn* encodes a receptor for protein reabsorption (Zhang et al., 2013); both
41 are critical for NC function. These two genes have lower expression in GCs compared
42 to PNs. Two NC-specific marker genes, *Kruppel-like factor 15* (*klf15*) and *UDP-*
43 *glycosyltransferase family 36 member A1* (*Dot*) (Ivy et al., 2015; Zhang et al., 2013),
44 were not present in our data set, most likely due to technical limitations with the 10X

approach as these genes have very short 3'UTRs. *Pros* and *Hand* are known makers for GCs and PNs (Weavers et al., 2009, Fig. S2D and 2E). Details on the marker genes are listed in Supplementary Table 2. Finally, in order to make the dataset accessible to users, we developed a visualization web portal (https://www.flyrnai.org/scRNA/kidney/) that allows users to query the expression of any genes of interest in different cell types.

Reconciling physiology with clusters

The Malpighian tubule generates a primary urine not by paracellular filtration but by potent active cation transport. This is coupled to channel mediated anion flux to balance charge and water channels to allow rapid flux of osmotically obliged water (Cohen et al., 2020). The tubules can transport their own volume of water every six seconds, making them the fastest-secreting epithelium known (Dow et al., 1994). By contrast with the vertebrate nephron, the paracellular route in Malpighian tubules is tightly guarded by septate junctions, and solutes are excreted by a wide range of massively expressed transporters (Wang et al., 2004; Chintapalli et al., 2007). Many of the genes underlying these processes have been identified and some localized to cell types. However, the single cell dataset allows us to address various questions at a larger scale. In particular, do genes ascribed to particular processes co-locate to the same cell types or regions, what new insights can be gained as to regional specialization, and can we predict functions of previously uncharacterized genes based on their expression patterns?

The PC transcriptome broadly follows expectation (Fig. 2); the plasma membrane V-ATPase subunits all show elevated expression in the PCs along the whole length of the tubule, not just in the main segment (Fig. 2 and Fig. S3A). A candidate apical exchanger, *Na+/H+ hydrogen exchanger 3* (*Nhe3*), shows a similar expression pattern, as does the Na^+, K^+ ATPase that stabilizes cellular cation levels (Torrie et al., 2004) (Fig. 2). This implies that the basic transport machinery is an inherent property of the whole length of the tubule, not just the secretory region. By contrast, the inward rectifier K^+ channel family genes, all of which are strongly expressed in the tubule, show distinct patterns. *Inwardly rectifying potassium channel 1* (*Irk1*) marks PCs of only the secretory main segment of the tubule (Fig. 2), *irk2* is expressed in the main segment and lower tubule, and *irk3* is generally expressed in PCs (Fig. S3B). Control of transport is clearly critical, and receptors for the three major neuropeptides believed to act on PCs to stimulate secretion are all found in PCs of the main segment; however, their expression patterns are slightly different (Fig. 2). *Capa receptor* (*CapaR*) is confined to the initial, transitional and main segments, as is its effector, the cyclic GMP kinase *foraging* (*for*). *DH31-R* is similarly expressed but *DH44-R2* is present in the main segment and lower tubule. Surprisingly, it is also strongly expressed in SCs; this dual control of two cell types had not been predicted experimentally.

SCs are thought to provide a transcellular shunt for anions and water, and accordingly, the two chloride channels *Clc-a* (Cabrero et al., 2014) and *SecCl* (Feingold et al., 2019), as well as the two true aquaporins *Drip* and *Prip* (Cabrero et al., 2020), show strong SC-enriched expression. Tyramine signals identically to the kinin neuropeptide (Cabrero et al., 2013) and both their receptors show strong localization

1   to SCs, together with their downstream effector, protein kinase C, and the master
2   transcription factor, *teashirt* (*tsh*), which specifies SC fate (Denholm et al., 2013) (Fig.
3   2). However, there are surprises in this dataset; none of these genes show strong
4   expression in bar-shaped cells, characteristic of the initial segment of anterior tubules,
5   suggesting that although they are developmentally linked to stellate cells, bar-shaped
6   cells are not able to either receive diuretic signals or respond to them.

7   Junctional permeability is critical in epithelia. As the PC and SC lineages have
8   distinct origins (ectodermal and mesodermal, respectively) (Denholm et al., 2003), they
9   might not necessarily form heterotypic junctions. In fact, both cell types express the
10 septate (occluding) junctional genes *discs large 1* (*dlg1*), *snakeskin* (*ssk*) (Dornan et
11 al., 2020) and *Mesh* (Jonusaite et al., 2020) throughout the length of the tubule,
12 suggesting that both PC-PC and PC-SC junctions are equally tight. Interestingly,
13 although both cell types also contribute adherens junction components, the emphasis
14 is different, with *polychaetoid* (*ZO-1/pyd*) emphasized by SCs and *Armadillo* (*arm*) by
15 PCs. Of the gap junction (innexin) genes, three are strongly expressed in tubule
16 (Chintapalli et al., 2013); *inx2* and *inx7* are expressed in PCs but not SCs, and *inx3* is
17 enriched in stem cells. Thus, PCs have the abilities to communicate (for example,
18 sharing second messengers) and synchronize activities, but SCs are likely to be
19 functionally independent. There is experimental evidence to support this idea;
20 stimulation of PCs with *Capa* elevates intracellular calcium in PCs but not SCs (Rosay
21 et al., 1997), whereas the opposite holds for *Kinin* signaling (Radford et al., 2002). The
22 *Capa* and *Kinin* pathways thus act independently on two cell types without detectable
23 crosstalk, making functional interaction unlikely (MacMillan et al., 2018).

24   The tubules show strongly enriched expression of most organic solute transporter
25 families (Wang et al., 2004), including the ABC-transporters that underly eye color
26 (*white* (*w*), *scarlet* (*st*), *brown* (*bw*)), and these are all confined to main segment PCs
27 (Fig. S3E). Tubules are also excellent models for renal diseases (Cohen et al., 2020;
28 Dow and Romero, 2010) and readily develop oxalate kidney stones. Knockdown of the
29 oxalate transporter *prestin* increases these stones, presumably by preventing reuptake
30 of secreted oxalate (Hirata et al., 2012; Landry et al., 2016); *prestin* is expressed in
31 PCs of the reabsorptive (O'Donnell and Maddrell, 1995) lower tubule (Fig. S3E).
32 Similarly, transporters that have been implicated in the excretion of xenobiotics (Torrie
33 et al., 2004) are expressed only in PCs (Fig. S3F), confirming the role of these cells in
34 general-purpose solute transport.

35   As well as transport, tubules play a liver-like role in detoxification, and show
36 conspicuous expression of genes shown to detoxify insecticides (e.g. *Cyp6g1* and
37 *Cyp12d1*, Catania et al., 2004; Yang et al., 2007; Le Goff et al., 2003), and the master
38 transcriptional regulator *Hr96* (King-Jones et al., 2006); all of these genes show close
39 co-expression in PCs (Fig. S3G). Several transcription factors allow the clusters
40 imputed here to be resolved. For example, *tsh* and *tiptop* (*tio*) are SC-specific, *N* marks
41 stem cells, and *dachshund* (*dac*), *Dorsocross1* (*Doc1*), Homothorax (*Hth*), and *cut* (*ct*)
42 provide graded resolution of PC domains (Fig. S3K-M).

43

44 **Control of stellate cell shape**

SCs, which control channel-mediated Cl⁻ and water flux, have a cuboidal shape in third instar larvae. Subsequently, during the pharate adult stage, they adopt a star shape in the main segment and a bar shape in the initial segment (Beyenbach et al. 2010; Cabrero et al., 2020; Dow, 2012; Choubey et al., 2020). Previous studies have shown that disruption of SCs affects fly survival. For example, conditional downregulation of *Rab11* in SCs results in lethality at the pharate adult stage, and knockdown of *Snakeskin* (*Ssk*) in SCs result in loss of fluid integrity and a significant reduction in viability (Choubey et al., 2020; Dornan et al., 2020). Furthermore, ablation of SCs causes lethality, confirming the essential role of this cell type (Denholm et al., 2003). We identified two sub-clusters of SCs, bar-shaped SCs and main segment SCs (Fig. 3A). GO analysis revealed that bar-shaped SCs play important roles in cell-cell adhesion and potassium transport, while main segment SCs are mainly involved in water, hormone and neuropeptide flux (Fig. S4, Supplementary Table 4).

Next, we knocked down the top SCs marker genes using *tsh-Gal4* to study their functions, focusing on SCs shape and viability (Fig. S5A). Among the 18 genes analyzed, 13 were associated with reduced viability, four affected main segment SC cell shape, and two reduced main segment SC cell number (Fig. S5B-F). Among these genes, the top-ranking marker gene, *Rho guanine nucleotide exchange factor at 64C* (*RhoGEF64C*) (Fig. 3B), is an exchange factor for Rho GTPases. Knocking down *RhoGEF64c* affects cell shape of the main segment SCs (Fig. 3C and 3D) and viability (Fig. S7B). In humans, Rho-GTPases regulate the formation and maintenance of long cellular extensions/foot processes and their dysfunctions are associated with nephrotic syndrome (NS) (Matsuda et al., 2021). Further, following podocyte injury, Rho-GTPases orchestrate the rearrangement of the actin cytoskeleton (Matsuda et al., 2021). Interestingly, knockdown of *RhoGEF64c* results in loss of cytoarchitectural organization in main segment SCs (Fig. 3E and 3F) but did not affect septate junctions (Fig. S6A and S6B). This contrasts with knockdown of *Ssk,* which caused loss of both cytoarchitectural organization and septate junctions (Dornan et al., 2020). Finally, knockdown of other top marker genes, namely *Prip*, *Frequenin 2* (*Frq2*) and *CG10939,* did not affect the septate junctions (Fig. S6C and S6D).

Developmental trajectory analysis of principal cells

PCs are mitochondria-rich and transport protons through an apical, plasma membrane vacuolar H⁺-ATPase (V-ATPase) (Davies et al., 1996). The main functions of PCs are to set up a potassium gradient (Day et al., 2008; O'Donnell and Maddrell, 1995), which enters the cell basolaterally through the combined activity of Na⁺, K⁺-ATPase (Torrie et al., 2004), inward rectifier potassium channels (Evans et al., 2005; Wu et al., 2015; Swale et al., 2016), and potassium cotransports (Sciortino et al., 2001; Linton and O'Donnell 1999; Rodan et al., 2012). We identified six PC clusters from the scRNA-seq dataset and identified Gal4 lines that allowed us to precisely map their anatomical locations (Fig. 1D). To understand the functional differences of each PC cell cluster, we performed a GO analysis based on marker genes (Fig. S7A). Most of the GO terms refer to transport and responses to toxic substances, reflecting the main functions of the tubule. Interestingly, the top 10 terms in lower tubule PCs refer to transport,

7

suggesting that the function of lower tubule PC is to transport substances between Malpighian tubules and the hemolymph (GO information is in Supplementary Table 5).

scRNA-seq enables the exploration of the continuous differentiation trajectory of a developmental process. Thus, to analyze the developmental trajectory of PCs, we conducted a pseudotime analysis by ordering cells along a reconstructed trajectory using Monocle3 (Fig. S7B and S7C). Consistent with the distribution distance on the UMAP, inferred trajectories demonstrated gradual transitions from cells in lower ureter PCs, upper ureter PCs, lower tubule PCs, and lower segment PCs to main segment PCs, initial and transitional PCs (Fig. S7D). On the UMAP, lower segment PCs are close to lower ureter PCs, upper ureter PCs, and lower tubule PCs. The pseudotime analysis also showed that the state of lower segment PCs is a co-mixture of lower ureter PCs, upper ureter PCs, and lower tubule PCs (Fig. S7D). These results are consistent with our observation *in vivo* using *Esyt2-Gal4* flies (Fig. 1D and 1E), which suggested that lower segment PCs represent a new cell cluster that is different from previously reported *Alp4* expressed cells.

We chose *Best2*, *bifid (bi), Sarcoplasmic calcium-binding protein 2* (*Scp2*), *PDGF- and VEGF-receptor related* (*Pvr*), *Uro*, *salty dog* (*salt)*, *Alp4*, *SPR*, and *Transient receptor potential cation channel A1* (*TrpA1*) as representative genes for each cluster (Fig. S7E). A survey of our scRNA-seq dataset revealed that the expression of *TrpA1* gradually decreased along the pseudotime, followed by increased transcription of *Alp4* and *SPR*. The expression of *Pvr*, *Uro* and *salt* was elevated at the more geographical distant region of main segment PCs, while the progressive increase of *Best2*, *bi*, and *Scp2* expression was only observed in initial and transitional PCs (Fig. S7E). These results indicate that the patterns of expression of marker genes in each cluster are in concordance with the pseudotime analysis of the clusters.

Cell-type-specific expression of transcription factors and regulatory landscape
To investigate transcription factors (TFs) that may contribute to kidney differentiation and function, we identified 44 cell-type-specific transcription factors by setting up the parameter cutoff based on gene expression levels (fold change > 3) and adjusting the p-value (<0.05) (Fig. 4A-E). We also applied SCENIC, which is designed to reveal TF-centered gene co-expression networks (Aibar et al., 2017), for the simultaneous reconstruction of gene regulatory networks and identification of cell states (Fig. 4F, detail genes name in Fig. S8). By inferring a gene correlation network followed by motif-based filtration, SCENIC keeps only potential direct targets of each TF as modules (regulons).

Among the TFs, *esg*, *klu*, and *Sox100B* are specifically expressed in RCSs, which are essential for RSC proliferation and maintenance (Hung et al., 2020). SCENIC could also successfully infer multiple downstream target genes. For example, among the *esg* target genes are *fruitless* (*fru), N*, *Dl*, *klu* (Fig. 4G, Supplementary Table 6). *fru* is expressed male-specifically in the gonad stem cell (GSC) niche and plays important roles in the development and maintenance of GSCs (Zhou et al., 2021). We found that *fru* is also expressed in RCSs of both sexes (Fig. 4B), suggesting that it plays a critical role in RSC proliferation and/or maintenance.

tsh, tio, and Lim3 are expressed in both main segment and bar-shaped SCs (Fig. 4A). tsh and tio are paralogous genes that control SCs shape and the expression of genes required for terminal physiological differentiation (Laugier et al., 2005; Denholm et al., 2013). Interestingly, human TSHZ (homolog of tsh) genes are causal kidney disease loci, including ureteral smooth muscle differentiation and congenital pelvi-ureteric junction obstruction (Caubit et al., 2008; Jenkins et al., 2010). In addition, Dac, Doc1 and Doc2, which have been associated with tissue morphogenesis (Fig. 4D) (Brás-Pereira et al., 2016; Paul et al., 2018; Fan et al., 2021), are not only expressed in bar-shaped SCs but also in initial and transitional PCs. Finally, SCENIC also reveals that Lim3 is highly enriched in SCs (Fig. 4E and Fig. S8), which is consistent with the role of RhoGEF64C in controlling SC morphology (Fig. 3, Fig. 4G), as RhoGEF64C is a predicted target of Lim3 (Fig. 4G). Altogether, our analysis provides a list of possible TFs that control SCs morphology.

The two pairs of tubules are asymmetric both morphologically and transcriptionally. The anterior (right-side) tubules have an extended initial and transitional segment that typically contains calcareous concretions (stones) (Wessing and Eichelberg, 1978), and show selective expression of Doc1, Doc2 and dac (Chintapalli et al., 2012). These genes reflect the initial dorsal specification of the anterior tubules, before an embryonic rotation of the gut places them on the right side (Chintapalli et al., 2012). The single-cell data maps expression of these genes specifically to both bar-shaped cells and PCs of just the initial and transitional segments, suggesting a continuing role in maintaining the unique identities of these cell types. By contrast, PCs of the rest of the tubule, and SCs, show no such expression, implying that they are functionally equivalent in both sets of tubules. One interesting TF expressed in several PCs is Hepatocyte nuclear factor 4 (Hnf4). Its human orthologs are Hnf4γ and Hnf4α, a major regulator of renal proximal tubule development in mouse (Marable et al., 2020). Purine metabolites, including inosine, adenine, xanthine, hypoxanthine, and uric acid, are associated with increased diabetes risk and diabetic nephropathy, and are increased in Hnf4 mutant flies (Barry and Thummel, 2016). Interestingly, potential direct targets of Hnf4 include Arginine kinase (Argk) and midway (mdy) (Fig. 4E), with mdy acting as a repressor of Hnf4 and HNF4 controlling lipid metabolism in Drosophila nephrocytes (Marchesin et al., 2019). Additional information on these TFs can be found in Supplementary Table 7.

Metabolic pathway analysis

The basic functions of mammalian kidneys include metabolism of carbohydrates, proteins, lipids and other nutrients. As in mammalian kidneys, insect Malpighian tubules and nephrocytes play an essential role in the maintenance of ionic, acid–base and water balance, and elimination of metabolic and foreign toxins and homeostasis. To further understand metabolism in the fly kidney, we analyzed the KEGG metabolic pathways in UMAP of fly kidney snRNA-seq using AUCell software (Aibar et al., 2017). Among 86 KEGG metabolic pathways, purine metabolism, glycerophospholipid metabolism, nicotinate and nicotinamide metabolism, starch and sucrose metabolism were enriched (Fig. S9). Regarding purine metabolism, Xanthine oxidation is a

necessary step in the catabolic pathway for purines toward urate, allantoin and urea. Dysfunction of xanthine oxidase/dehydrogenase (XO/XDH) causes build-up of high levels of xanthine and hypoxanthine forming stones in humans and flies (Dent and Philpot, 1954; Ichida et al., 1997; Miller et al., 2013). Human ancestors lost the ability to synthesize a functional urate oxidase due to multiple point mutations in the *Uro* gene, resulting in increased serum and urinary uric acid (UA) levels (Mandal and Mount, 2015). In the UA pathway, humans and flies share some of the same steps. The product of the fly *Uro* gene which catalyzes formation of allantoin from UA (Fig. 5A). Most UA pathway genes are enriched in fly kidney cells, specifically in PCs (Fig. 5B and C). The enzymes that control the last three steps, which are encoded by *rosy* (*ry*), *Uro* and *CG30016,* are highly enriched in main segment PCs, suggesting that the last step occurs in this region (Fig. 5D). *ry* is the homolog of human XDH and loss-function of *ry* is associated with bloating in the lower tubules and formation of stones (Mitchell and Glassman, 1959). Metabolomic analysis of *ry* mutants showed significant changes up to five metabolites away from the metabolic lesion, with large increases in levels of hypoxanthine and xanthine, and undetectable levels of the downstream metabolite UA (Hobani et al. 2012). The product of *CG30016* is predicted to have hydroxyisourate hydrolase activity and to be involved in purine nucleobase metabolism. It will be interesting to see whether this gene also plays a role in maintaining fly urate levels.

Another important function of the insect kidney is detoxification. Many cytochromes P450 (CYPs) genes are involved in this process (Lu et al., 2021) (some are listed in Fig. S10A). For example, *Cyp6g1*, *Cyp6g2*, and *Cyp6A2* are involved in DDT insecticide resistance in flies (Seong et al., 2020; Bergé et al., 1998). *Cyp4e3* has been associated with permethrin insecticide resistance (Terhzaz et al., 2015) and Cyp12a5 in Nitenpyram resistance (Harrop et al. 2018). *Cyp307a2*, *Cyp18a1*, and *Cyp312a1* are involved in degradation of polychlorinated biphenyls (Idda et al., 2020), and *Cyp12d1* impacts caffeine resistance (Najarro et al., 2015). Among these genes, *Cyp6g1*, *Cyp6A2*, *Cyp4e3*, *Cyp12a5*, *Cyp307a2*, and *Cyp12d1* are mainly expressed in PCs (Fig. S10B), which is consistent with PCs playing a key role in detoxification.

Cross-species and human kidney disease analysis

Considering that the function of all animal excretory systems is to remove toxins from the body and maintain homeostatic balance, we next asked whether we could match fly kidney cell types to higher animals kidney cell types (mouse) and lower animal protonephridia cell types (planarian), and whether the single cell level data can help implicate new genes and cell types in human kidney diseases. We used the Self-Assembling Manifold mapping (SAMap) algorithm (Tarashansky et al., 2021) to map our fly single-cell transcriptomes with scRNA-seq data from mouse (Ransick et al., 2019) and planaria (Fincher et al., 2018). This method depends on two modules. First, a gene-gene bipartite graph with cross-species edges connects homologous gene pairs weighted by protein sequence similarity (all gene pairs are listed in Supplementary Table 8 and Supplementary Table 9). Second, a gene-gene graph projects two single-cell transcriptomic datasets into a joint, lower-dimensional manifold representation, from which each cell mutual cross-species neighbors are linked to

1   stitch the cell atlases together for fly and mouse kidney (Fig. S11A and B). With this
2   method, SAMap produced a combined manifold with a high degree of cross-species
3   alignment (Fig. S11C). After measuring the mapping strength between cell types by
4   calculating an alignment score (as edge width showed in Fig. 6A), which was defined
5   as the average number of mutual nearest cross-species neighbors of each cell relative
6   to the maximum possible number of neighbors, we generated a Sankey plot with 10 fly
7   kidney cell clusters matched to 26 mouse kidney cell clusters (Fig. 6A). A similar
8   analysis was performed for flies and planarians, with 9 fly kidney cell clusters matched
9   to 6 planarian protonephridia cell clusters (Fig. S12A-C and Fig. 6B).

10   The results of the fly/mouse analysis suggest that fly main segment PC, lower tubule
11   PC, upper ureter PC are similar to mouse proximal tubule (segment 1-3); fly bar-
12   shaped and main segment SCs map to mouse lower LOH (loop of Henle) thin limb of
13   inner medulla of juxtamedullary nephron; fly adult pericardial nephrocytes are similar
14   to mouse podocytes (visceral epithelium) and parietal epithelium; and fly adult GCs
15   map to mouse parietal epithelium (Fig. 6A). Thus, although pericardial and GC
16   nephrocytes are frequently considered to be interchangeable, they represent different
17   facets of the mammalian nephron. Interestingly, we found that the fly lower segment
18   PCs represents a discontinuous cell population located in the lower segment region
19   and match to mouse PCs of inner medullary collecting duct type 1/2, suggesting that
20   the fly lower segment PC cluster is a newly identified MT cell type.

21   The fly/planaria comparison suggests that fly stem cells are similar to planarian
22   Transition State 1 and Transition State 2 cluster cells, indicating that kidney stem cells
23   are present in both lower animal species but not mammals. Fly main segment PCs
24   map to planarian Proximal Tubule; fly lower tubule PCs map to planarian Collecting
25   Duct; and fly upper ureter PCs and initial and transitional PCs map to planarian Distal
26   Tubule. Interestingly, fly pericardial nephrocytes, GCs, bar-shaped and main segment
27   SCs map to planarian Flame Cells, suggesting that these cell types have conserved
28   function for removing waste materials (Fig. 6B).

29   Next, we chose several genes from homologous gene pairs (see Supplementary
30   Table 8 and Supplementary Table 9) to test the robustness of the comparative
31   analyses. Based on the SAMap, fly *pvf1* and *tsh* are highly expressed in SCs. Strikingly,
32   the homologous mouse genes *pdgfa* and *Tshz2* are highly expressed in lower LOH
33   thin limb of inner medulla of juxtamedullary nephron (Fig. S11D). Further, fly *Cyp6g1*
34   and *Na$^+$-dependent inorganic phosphate cotransporter* (*NaPi-T*) genes are highly
35   expressed in main segment PC, and their corresponding genes in the mouse, *Cyp4b1*
36   and *Slc22a6*, are highly expressed in mouse proximal tubules (Fig. S11D). *Esyt2* is a
37   marker gene for fly the lower segment PCs, and its homologous gene *Esyt1* is highly
38   expressed in PC of the inner medullary collecting duct type 1/2 (Fig. S11D). In the fly,
39   *sns* is highly expressed in nephrocytes, and the homologous gene in the mouse,
40   *Nphs1,* is highly expressed in mouse podocytes (Fig. S11D). With regards to planaria,
41   despite the lower extent of genome annotation, we identified some informative gene
42   pairs (see Supplementary Table 9) that include the fly nephrocyte marker gene *sns*,
43   the SC marker gene *Nep2*, the stem cell marker gene *esg*, and the main segment PC
44   marker gene *salt*, which could be mapped to the planaria cell clusters (Fig. S12D).

Altogether, these results indicate that the SAMap mapping results are well supported by conserved gene expression programs.

Finally, we examined whether the single cell data can help implicate cell clusters and gene targets in human kidney diseases, especially as a previous study in the mouse has shown that hereditary human kidney diseases characterized by the same phenotypic manifestations originate from the same cell types (Park et al., 2018). Strikingly, single cell distribution of human kidney diseases in the fly kidney showed that most of these genes were enriched in the orthologous cell types (Fig. S13). In particular, the fly homologs of 13 of 33 genes associated with monogenic inheritance of nephrotic syndrome in humans were expressed in fly nephrocytes. In the mouse, homologs of genes associated with the syndrome were expressed in podocytes (Park et al., 2018). Among the fly homologs, *sns*, *kin of irre* (*kirre*), and *cubn* have been shown to have key functions in fly nephrocytes (Helmstädter et al., 2017). The fly orthologs of human *Nphs1* and *Kirrel1, sns* and *kirre*, direct adhesion, fusion and formation of a slit diaphragm (SD) structure in insect nephrocytes (Zhuang et al., 2009). Knockdown of *sns* or *kirre* leads to a dramatic decrease in uptake of large proteins, consistent with the role of the SD in mammalian podocytes (Zhuang et al., 2009). Finally, the fly homologs of two genes associated with nephrolithiasis, *ATP6V1B1* and *ATP6V0A4* (*Vha55* and *Vha100-2* in flies), are highly expressed in Malpighian tubule PCs. Mutations in *ATP6V1B1* and *ATP6V0A4* have been identified in calcium oxalate kidney stone patients, suggesting that they are essential for calcium oxalate kidney stone formation (Dhayat et al., 2016). In the mouse, the orthologs of these two genes are hallmarks of intercalated cells, and one type of intercalated cell (intercalated type non-A non-B cell of nephron connecting tubule) matched with fly initial and transitional PCs (Fig. 6A). Interestingly, flies with knockdown of *Vha55* or *Vha100-2* in the Malpighian tubule also develop calcium oxalate kidney stones (Fan et al., 2020). Altogether, the analysis of the expression of fly homologs of human kidney disease-associated genes at the single cell level will help develop more accurate fly models of human kidney diseases.

DISCUSSION

Here, we surveyed the cell types of the adult fly kidney using snRNA-seq and identified all known cell types and their sub-types. Our dataset provides insights in SC shape, identifying RhoGEF64c as a key cell shape regulator. Interestingly, six clusters of PCs mapped to different regions of the tubule and we could associate them with different physiological functions. The dataset also provides information about potential gene expression networks of transcription factors. Of particular interest, we find that RSCs contain two clusters distinguishable by expression of *Dl*⁺ *klu*⁻ and *Dl*⁻ *klu*⁺ (detailed information in Supplementary text, Fig. S14 and Fig. S15), reminiscent of ISC/EB cells in the midgut (Hung et al., 2020). In addition, we used FlyPhoneDB (Liu et al, 2021) to predict ligand–receptor interactions between different cell clusters, a resource that will help analyze communication among kidney cells (detailed information in Supplementary text and Fig. S16). Altogether, this study will facilitate future work on the fly kidney and serve as a resource to understand cell-type identity and physiology.

1   Our study extends a previous report that performed scRNA-seq of the Malpighian
2   tubules with a focus on the ureter and lower tubule (Wang and Spradling, 2020). The
3   previous study captured 710 cells that did not include many types of PCs due to the
4   dissection and few SCs, as these oddly shaped cells were likely not captured efficiently
5   using scRNA-seq. Our study using single nuclei rather than single cells overcame this
6   difficulty, and altogether, we successfully recovered 12,166 cells representing 1,730
7   main segment SCs and 336 bar-shaped SCs. Nevertheless, snRNA-seq did not
8   capture well the GCs, which have double nuclei (Marchesin et al., 2019), or the
9   pericardial nephrocytes, which have large nuclei, as they were underrepresented in
10   our dataset - an issue also reported in several studies of the mouse kidney (Wu et al.,
11   2019; Ransick et al., 2019). We identified six sub-clusters of PCs that map to different
12   anatomical locations. Importantly, GO analysis showed that PCs in different
13   geographic locations have functional differences. For example, main segment PCs
14   respond to toxic substances, and consistent with this, genes related to insecticide
15   metabolism and genes encoding the last three steps enzymes of the uric acid pathway
16   were enriched in these cells (Fig. 5D and S10). The function of lower tubule PCs relates
17   to transport of different substances between Malpighian tubules and hemolymph (Fig.
18   S7). Altogether, PCs in different locations have distinct physiological functions,
19   highlighting that coordination of PC function is required for Malpighian tubules to
20   perform their normal function.

21   The cross-species analysis not only provided information about the potential
22   functions of unknow cell types, but also gave us a better comparative understanding
23   of kidney cells from lower species (planaria) to higher species (mouse). For example,
24   fly lower segment PCs map to mouse PCs of inner medullary collecting duct type 1/2
25   (Fig. 6A), but there is no corresponding cell type in planaria. Results of the cross-
26   species analysis will facilitate study of the functions of specific cell types found in higher
27   animals using lower species as models.

28   *Drosophila* Malpighian tubules and nephrocytes have been used successfully to
29   model human kidney diseases. For example, mutations in the vacuolar-type H^+-
30   ATPase (v-ATPase) subunit genes ATP6V1B1 and ATP6V0A4 in humans have been
31   identified in recurrent calcium oxalate kidney stones (Dhayat et al., 2016) and
32   knockdown of the fly homologs, *Vha55* and *Vha100-2,* using Uro-Gal4 led to increased
33   formation of calcium oxalate stones in Malpighian tubules (Fan et al., 2020). Our fly
34   kidney cell atlas will facilitate disease modeling and analysis. First, it will help narrow
35   down the number of genes to be tested in specific cell types, as our snRNA-seq has
36   identified cell type-specific transcriptomes. Second, as we were able to match cell
37   types between the fly and mouse, we are now able to associate human kidney disease-
38   associated genes with specific fly kidney cell types. This critical information should
39   facilitate the development of more accurate fly models of human kidney diseases.

40   In conclusion, our dataset provides detailed insights into: fly kidney cell type-specific
41   gene expression patterns, the specific transcription factors in each cell cluster,
42   potential cell-cell communication network, cross-species mapping of fly kidney cell

1 types to mouse and planarian kidney cell types; association of human kidney disease-
2 associated genes with specific cell clusters. We also provide a web-based resource
3 for visualization of gene expression in single cells of the fly kidney.

4

METHODS

Single nucleus isolation and sequencing

7 *Drosophila* Malpighian tubules and nephrocytes were dissociated to single nuclei as
8 previously described (Li et al. 2021) with a few modifications. Malpighian tubules were
9 dissected under a microscope from 5-7-day old *Drip-Gal4>GFP.nls* male and female
10 adult flies. Nephrocytes were dissected under a fluorescence microscope from 5-7-
11 day old *Dot-Gal4>GFP.nls* male and female adult flies. Ten flies at a time were
12 dissected and samples immediately transferred into 1.5 ml EP tube with Schneider's
13 medium on ice to avoid exposing the tissues to room temperature for a long period of
14 time. Once 50 flies were dissected, EP tubes were sealed with parafilm and put on dry
15 ice. Next steps involved spraying 100% ethanol to the dry ice near the tube to quickly
16 freeze the sample and storing samples at -80°C for long-term. After dissection,
17 samples were spined down (thaw samples from - 80°C) in 100 ul Schneider's medium
18 using a bench top spinner, medium was discarded, and 100 ul homogenization butter
19 (Li et al. 2021) was added. Subsequently, 900 ul homogenization buffer was added,
20 and 1000ul homogenized sample transferred into the 1 ml dounce. Nuclei were
21 released by 15-20 strokes with loose pestle and 15-20 tight pestle on ice. 1000 ul
22 sample was filtered through 5ml cell strainer (35 um), and then filter sample using 40
23 um Flowmi into 1.5ml EP tube. Centrifuge for 10 min at 1000 g at 4°C. Resuspend in
24 1000 ul PBS/0.5% BSA with RNase inhibitor. And filter sample using 40 um Flowmi
25 into a new EP tube. Hoechst 33342 was used to stain nuclei for more than 5 min. Then
26 FACS and collect single nuclei into a tube for 10x Genomics.

27 Ten thousand nuclei were targeted for each sample when loaded into the Chromium
28 Controller (10X Genomics, PN-120223) on a Chromium Single Cell B Chip (10X
29 Genomics, PN-120262), and processed to generate single cell gel beads in the
30 emulsion (GEM) according to the manufacturer's protocol (10X Genomics, CG000183).
31 The library was generated using the Chromium Single Cell 3′ Reagent Kits v3.1 (10X
32 Genomics, PN-1000121) and Chromium i7 Multiplex Kit (10X Genomics, PN-120262)
33 according to the manufacturer's manual. Quality control for constructed library was
34 performed by Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies,
35 5067-4626) for qualitative analysis. Quantification analysis was performed by Illumina
36 Library Quantification Kit (KAPA Biosystems, KK4824). The library was sequenced on
37 an Illumina NovaSeq system or Nextseq 500 instrument.

Dataset processing

The quality of the raw sequencing data was checked by FastQC software. The raw sequencing data were processed by cellranger count pipeline to generate the single cell matrix for each sample. The single cell matrix was analyzed by the Seurat package and Harmony was used for batch correction. The Malpighian tubules and nephrocytes samples were processed and clustered separately, and then were merged by filtering out unrelated cell clusters. The processed result of the gene expression matrix was used for the downstream data analysis.

To facilitate mining of the datasets, we developed a visualization web portal (https://www.flyrnai.org/scRNA/kidney/) that allows users to query the expression of any genes of interest in different cell types and to compare the expression of any 2 genes in individual cells. This MT dataset can also be mined at Fly Cell Atlas (https://flycellatlas.org/) along with other datasets generated by the FCA consortium (Li et al. 2021).

Gene ontology (GO) analysis

Gene ontology (GO) analysis was performed by clusterProfiler. The marker genes identified using Seurat were used for GO analysis. The strength of enrichment was calculated as negative of log10(p-value), which was used to plot the barplot.

Cross-tissue analysis

Cross-tissue analysis data are from the Malpighian tubules processed dataset and midgut dataset (accession code: GSE120537). Before merging the two datasets, the number of Malpighian tubules cells was downsampled to the same size as the midgut dataset. The top markers of renal stem cells (RSC) were calculated by comparing the RSC with the rest of the merged dataset without intestinal stem cells (ISC). The top markers of ISC were calculated by comparing the ISC with the rest of the merged dataset without RSC. Results of the comparison are visualized on a Venn diagram.

Pseudotemporal ordering of cells using Monocle3

Processed Malpighian tubules dataset was analyzed using Monocle3 for pseudotemporal ordering. The state representing lower ureter PC was chosen as the starting time point. The Ridge plot was generated by extracting the cell clustering and pseudotime information and then visualized by Seurat RidgePlot function. The gene expression heatmap was generated by merging the cells into bins by the order of pseudotime and visualized by the pheatmap R package.

Transcription factors enrichment and SCENIC analysis

The top markers for each cluster were used as the candidates for transcription factors enrichment analysis. The markers were filtered by fold change > 3 and adjusted p-

15

1 value < 0.05. The transcription factors from the filtered markers were visualized by the
2 Seurat DoHeatmap

3   The analysis of regulon activity was conducted using the SCENIC pipeline (Aibar et
4 al., 2017). Cells from the previously processed dataset were selected as the input cells.
5 The TF and co-expressed genes were constructed by GRNBoost2. The TF co-
6 expression gene sets were filtered by the RcisTarget fly database. The regulon activity
7 score AUC (Area Under the Curve) was calculated by AUCell, and the active regulons
8 were determined by the AUCell default parameters. The regulon activity was visualized
9 by the average AUC score for each cluster.

10 **Cell-cell communication analysis**

11 The cell-cell communication analysis was performed using FlyPhoneDB (Liu et al.,
12 2021). The previously processed gene expression matrix and cell clustering
13 information were used as the input for the analysis. Ligand-receptor interaction scores
14 and specificity were then calculated. Cell communication at the signaling pathway level
15 was visualized by a circle plot. The interaction of ligand-receptor pairs between two
16 cell types was visualized by dot plot. The network was generated based on the MIST
17 database and TF2TG literature using Cytoscape (Hu et al., 2018; Otasek et al., 2019).

18 **Cross-species analysis**

19 Datasets included in the cross-species analysis were the processed dataset from this
20 study for the fly, the mouse kidney dataset GSE129798 and the planarian dataset
21 GSE111764. The analysis was conducted using the SAMap software (Tarashansky et
22 al., 2021). The input h5ad file for SAMap was processed by the Self-Assembling-
23 Manifold (SAM) algorithm. Alignments for each cell type in fly and mouse were
24 calculated by the get_mapping_scores function. Enriched gene pairs from the aligned
25 cell types were retrieved by find_all function with a default alignment score threshold
26 of 0.1. The SAMap results were visualized by the sankey_plot function.

27 **Fly genetics**

28 Fly husbandry and crosses were performed under a 12:12 hour light:dark photoperiod
29 at 25°C. *Hand-GFP; 4xHand-Gal4/CyO* and *Dot-Gal4* stocks are gifts from Dr. Han
30 Zhe. *esg-Gal4* and *Pros-Gal4* are from the Perrimon lab stock collection.
31   The following strains were obtained from the Bloomington *Drosophila* Stock
32 Center: *Drip-Gal4* (BL66782), *UAS-GFP.nls* (BL4776), *UAS-mCD8::RFP* (BL32219),
33 *UAS-mCD8::GFP* (BL32185), *c42-Gal4* (BL30835), *Uro-Gal4; tubGal80ts* (BL91415),
34 *Alp4-Gal4* (BL30840), *SPR-Gal4* (BL84692), *wnt4-Gal4* (BL67449), *Esyt2-Gal4*
35 (BL77712), *Debcl-Gal4* (BL81163), *ih-Gal4* (BL76162), *CG30377-Gal4* (BL67426),
36 *Octα2R-Gal4* (BL67637), *tsh-Gal4* (BL3040), *fru-Gal4* (BL30027), *Sba-Gal4*
37 (BL67640), *Doc2-Gal4* (BL26436), *Lim3-Gal4* (BL67450), *Pvf1-Gal4* (BL23032), *y v;*
38 *UAS-LucRNAi, attp2* (BL31603), *y v; UAS-wat-RNAi, attp40* (BL67801), *y v; UAS-tutl-*

16

1 *RNAi, attp40* (BL54850), *y v; UAS-axed-RNAi, attp40* (BL62928), *y v; UAS-prip-RNAi,*
2 *attp40* (BL50695), *y v; UAS-prip-RNAi, attp2* (BL44464), *y v; UAS-nep2-RNAi, attp40*
3 (BL61902), *y v; UAS-CG13323-RNAi, attp40* (BL53969), *y v; UAS-Lgr1-RNAi, attp2*
4 (BL51465), *y v; UAS-Octα2R-RNAi, attp2* (BL50678), *y v; UAS-CG30377-RNAi, attp40*
5 (BL51386), *y v; UAS-notum-RNAi, attp40* (BL55379), *y v; UAS-TkR99D-RNAi, attp2*
6 (BL27513), *y v; UAS-ih-RNAi, attp40* (BL58089), *y v; UAS-ncc69-RNAi, attp2*
7 (BL28682), *y v; UAS-frq2-RNAi, attp2* (BL28711), *y v; UAS-qvr-RNAi, attp40*
8 (BL58061), *y v; UAS-CG10939-RNAi, attp40* (BL65156), *y v; UAS-CG42594-RNAi,*
9 *attp2* (BL35006), *y v; UAS-RhoGEF64c-RNAi, attp2* (BL31130).

10 The following strains were obtained from the Vienna *Drosophila* Resource Center:
11 *y w (1118)*; *attp landing site* (v60100), *UAS-RhoGEF64c-RNAi* (v47121), *UAS-*
12 *RhoGEF64c-RNAi* (v105252).

13 For the screen shown in Fig. S5B, 8 *tsh-Gal4/CyO; UAS-CD::GFP* virgin females
14 were crossed with 4 RNAi males. Flies were raised at 22°C and the ratio of Cy+/Cy
15 was determined.

16 **Immunostaining and confocal microscopy**

17 *Drosophila* Malpighian tubules (and guts), PNs (included in the whole abdomen) and
18 GCs (and foreguts and crops) from adult females were fixed in 4% paraformaldehyde
19 in Phosphate-buffered saline (PBS) at room temperature for 1 hour, incubated for 1
20 hour in Blocking Buffer (5% normal donkey serum, 0.3% Triton X-100, 0.1% bovine
21 serum albumin (BSA) in PBS), and stained with primary antibodies overnight at 4°C in
22 PBST (0.3% Triton X-100, 0.1% BSA in PBS). Primary antibodies and their dilutions
23 used were: mouse anti-GFP (Invitrogen, A11120; 1:300) and mouse anti-discs-large
24 (DSHB, 4F3,1:50). After primary antibody incubation, the tissues were washed 3 times
25 with PBST, stained with 4′,6-diamidino-2-phenylindole (DAPI) (1:2000 dilution),
26 Phalloidin TRITC (Sigma-Aldrich, 1:2000) and Alexa Fluor-conjugated donkey-anti-
27 mouse (Molecular Probes, 1:1000), in PBST at 22°C for 2 hours, washed 3 times with
28 PBST, and mounted in Vectashield medium.

29 All images presented in this study are confocal images captured with a Nikon Ti2
30 Spinning Disk confocal microscope. Z-stacks of 15-20 images covering one layer of
31 the epithelium from the apical to the basal side were obtained, adjusted, and
32 assembled using NIH Fiji (ImageJ), and shown as a maximum projection. Details of
33 the imaging method are as follows: Samples were imaged with a Yokogawa CSU-W1
34 single disk (50 μm pinhole size) spinning disk confocal unit attached to a fully motorized
35 Nikon Ti2 inverted microscope equipped with a Nikon linear-encoded motorized stage
36 with a Mad City Labs 500 μm range Nano-Drive Z piezo insert, an Andor Zyla 4.2 plus
37 (6.5 μm photodiode size) sCMOS camera using a Nikon Plan Apo 60x/1.4 NA DIC oil
38 immersion objective lens with Cargille Type 37 immersion oil (cultured cells) or a Nikon
39 Plan Apo 20x/0.75 DIC air objective lens (tissue samples). The final digital resolution
40 of the image was 0.109 and 0.325 μm/pixel, respectively. Fluorescence from DAPI,
41 Alexa Fluor (AF)-488, and AF555 was collected by illuminating the sample with directly
42 modulated solid-state lasers 405 nm diode 100 mW (at the fiber tip) laser line, 488 nm

diode 100 mW laser line, and 561 nm DPSS 100 mW laser line in a Toptica iChrome MLE laser combiner, respectively. A hard-coated Semrock Di01-T405/488/568/647 multi-bandpass dichroic mirror was used for all channels. Signal from each channel was acquired sequentially with hard-coated Chroma ET455/50, Chroma ET525/36 nm, and Chroma ET605/52 nm emission filters in a filter wheel placed within the scan unit, for blue, green, and red channels, respectively. Nikon Elements AR 5.02 acquisition software was used to acquire the data. 2 µm range Z-stacks, set by indicating the middle focal plane and a z-step interval of 50 µm, were acquired using piezo Z-device, with the shutter closed during axial movement. Images were acquired by collecting the entire Z-stack in each color or by acquiring each channel in each focal plane within the Z stack. Data were saved as ND2 files.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request. Raw snRNA-seq reads have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession codes: (XXX, will provide soon). Processed datasets can be mined through a web tool (https://www.flyrnai.org/scRNA/kidney/) that allows users to explore genes and cell types of interest.

AUTHOR CONTRIBUTIONS

N.P. and J.X. conceptualized and designed the experiments. J.X. performed most experiments. Y.F.L. performed the bioinformatics analysis. H.J.L. and S.S.K. performed single nucleus isolation and RNA library construction. R.J.H. helped with dissection experiments and annotation. Y.H. and A.C. helped with bioinformatics analysis and the website. T.A., C.K. and B. W. contributed to cross-species analysis. S.R.Q. and L.L. contributed reagents and supervision of single nucleus isolation and RNA library construction. J.X., Y.F. L., J. A.T.D., A.P.M. and N.P. analyzed the data. J.X. wrote the first draft of the paper. N.P. and J.A.T.D. edited the paper. All authors discussed the results and commented on the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al., (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086.

Barry, W.E., and Thummel, C.S. (2016). The *Drosophila* HNF4 nuclear receptor promotes glucose-stimulated insulin secretion and mitochondrial function in adults. Elife 5, e11183.

Bergé, J.B., Feyereisen, R., and Amichot, M. (1998). Cytochrome P450 monooxygenases and insecticide resistance in insects. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *353*, 1701–1705.

Beyenbach, K.W., Skaer, H., and Dow, J.A.T. (2010). The developmental, molecular, and transport biology of Malpighian tubules. Annu. Rev. Entomol. *55*, 351–374.

Brás-Pereira, C., Potier, D., Jacobs, J., Aerts, S., Casares, F., and Janody, F. (2016). dachshund Potentiates Hedgehog Signaling during *Drosophila* Retinogenesis. PLoS Genet. *12*, e1006204.

Cabrero, P., Richmond, L., Nitabach, M., Davies, S.A., and Dow, J.A.T. (2013). A biogenic amine and a neuropeptide act identically: tyramine signals through calcium in *Drosophila* tubule stellate cells. Proc. Biol. Sci. *280*, 20122943.

Cabrero, P., Terhzaz, S., Dornan, A.J., Ghimire, S., Holmes, H.L., Turin, D.R., Romero, M.F., Davies, S.A., and Dow, J.A.T. (2020). Specialized stellate cells offer a privileged route for rapid water flux in *Drosophila* renal tubule. Proc. Natl. Acad. Sci. USA *117*, 1779–1787.

Cabrero, P., Terhzaz, S., Romero, M.F., Davies, S.A., Blumenthal, E.M., and Dow, J.A.T. (2014). Chloride channels in stellate cells are essential for uniquely high secretion rates in neuropeptide-stimulated *Drosophila* diuresis. Proc. Natl. Acad. Sci. USA *111*, 14301–14306.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene expression. Science *370*, eaba7721.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502.

Catania, F., Kauer, M.O., Daborn, PJ., Yen, J.L., Ffrench-Constant, R.H., and Schlotterer, C. (2004). World-wide survey of an Accord insertion and its association with DDT resistance in *Drosophila melanogaster*. Mol. Ecol. *13*, 2491–2504.

Caubit, X., Lye, C.M., Martin, E., Coré, N., Long, D.A., Vola, C., Jenkins, D., Garratt, A.N., Skaer, H., Woolf, A.S., et al. (2008). Teashirt 3 is necessary for ureteral smooth muscle differentiation downstream of SHH and BMP4. Development *135*, 3301–3310.

Chintapalli, V.R., Terhzaz, S., Wang, J., Al Bratty, M., Watson, D.G., Herzyk, P., Davies, S.A., and Dow, J.A.T. (2012). Functional correlates of positional and gender-specific renal asymmetry in *Drosophila*. PLoS One *7*, e32577.

Chintapalli, V.R., Wang, J., and Dow, J.A.T. (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. *39*, 715–720.

Chintapalli, V.R., Wang, J., Herzyk, P., Davies, S.A., and Dow, J.A.T. (2013). Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia. BMC Genomics *14*, 518.

Choubey, P.K., Nandy, N., Pandey, A., and Roy, J.K. (2020). Rab11 plays a key role in stellate cell differentiation via non-canonical Notch pathway in Malpighian tubules of *Drosophila melanogaster*. Dev.

Biol. *461*, 19–30.

Cohen, E., Sawyer, J.K., Peterson, N.G., Dow, J.A.T., and Fox, D.T. (2020). Physiology, Development, and Disease Modeling in the Drosophila Excretory System. Genetics *214*, 235–264.

Davies, S.A., Goodwin, S.F., Kelly, D.C., Wang, Z., Sözen, M.A., Kaiser, K., and Dow, J.A.T. (1996). Analysis and inactivation of vha55, the gene encoding the vacuolar ATPase B-subunit in *Drosophila melanogaster* reveals a larval lethal phenotype. J. Biol. Chem. *271*, 30677–30684.

Day, J.P., Wan, S., Allan, A.K., Kean, L., Davies, S.A., Gray, J.V., and Dow, J.A.T. (2008). Identification of two partners from the bacterial Kef exchanger family for the apical plasma membrane V-ATPase of Metazoa. J. Cell Sci. 121, 2612–2619.

de Boer, I.H., and Utzschneider, K.M. (2017). The kidney's role in systemic metabolism-still much to learn. Nephrol. Dial. Transplant. *32*, 588–590.

Denholm, B., and Skaer, H. (2009). Bringing together components of the fly renal system. Curr. Opin. Genet. Dev. *19*, 526–532.

Denholm, B., Sudarsan, V., Pasalodos-Sanchez, S., Artero, R., Lawrence, P., Maddrell, S., Baylies, M., and Skaer, H. (2003). Dual origin of the renal tubules in *Drosophila*: mesodermal cells integrate and polarize to establish secretory function. Curr. Biol. *13*,1052–1057.

Denholm, B., Hu, N., Fauquier, T., Caubit, X., Fasano, L., and Skaer, H. (2013). The *tiptop/teashirt* genes regulate cell differentiation and renal physiology in *Drosophila*. Development *140*, 1100–1110.

Dent, C.E., and Philpot, G.R. (1954). Xanthinuria, an inborn error (or deviation) of metabolism. Lancet *266*, 182–185.

Dhayat, N.A., Schaller, A., Albano, G., Poindexter, J., Griffith, C., Pasch, A., Gallati, S., Vogt, B., Moe, O.W., and Fuster, D.G. (2016). The Vacuolar H+-ATPase B1 Subunit Polymorphism p.E161K Associates with Impaired Urinary Acidification in Recurrent Stone Formers. J. Am. Soc. Nephrol. *27*, 1544–1554.

Dornan, A.J., Halberg, K.A., Beuter, L.-K., Davies, S.-A., and Dow, J.A.T. (2020). The septate junction protein Snakeskin is critical for epithelial barrier function and tissue homeostasis in the Malpighian tubules of adult *Drosophila*. bioRxiv. doi: https://doi.org/10.1101/2020.12.14.422678

Dow, J.A.T. (2012). The versatile stellate cell - more than just a space-filler. J. Insect Physiol. *58*, 467–472.

Dow, J.A., and Romero, M.F. (2010). Drosophila provides rapid modeling of renal development, function, and disease. Am. J. Physiol. Renal Physiol. *299*, F1237–F1244.

Dow, J.A.T., Krause, S.A., and Herzyk, P. (2021). Updates on ion and water transport by the Malpighian tubule. Curr. Opin. Insect Sci. *47*, 31–37.

Dow, J.A., Maddrell, S.H., Görtz, A., Skaer, N.J., Brogan, S., and Kaiser, K. (1994). The malpighian tubules of *Drosophila melanogaster*: a novel phenotype for studies of fluid secretion and its control. J. Exp. Biol. *197*, 421–428.

Evans, J.M., Allan, A.K., Davies, S.A., and Dow, J.A.T. (2005). Sulphonylurea sensitivity and enriched expression implicate inward rectifier K+ channels in Drosophila melanogaster renal function. J Exp Biol. *208*, 3771–3783.

Fan, Q.X., Gong, S.Q., Hong, X.Z., Feng, X.M., and Zhang, F.J. (2020). Clinical-grade Garcinia cambogia extract dissolves calcium oxalate crystals in Drosophila kidney stone models. Eur. Rev. Med. Pharmacol. Sci. *24*, 6434–6445.

Fan, Z., Zhang, J., Wang, D., and Shen, J. (2021). T-box transcription factors Dorsocross and optomotor-blind control *Drosophila* leg patterning in a functionally redundant manner. Insect Biochem. Mol. Biol. *129*, 103516.

Feingold, D., Knogler, L., Starc, T., Drapeau, P., O'Donnell, M.J., Nilson, L.A., and Dent, J.A.T. (2019). *secCl* is a cys-loop ion channel necessary for the chloride conductance that mediates hormone-induced fluid secretion in *Drosophila*. Sci. Rep. *9*, 7464.

Fincher, C.T., Wurtzel, O., de Hoog, T., Kravarik, K.M., and Reddien, P.W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. Science *360*, eaaq1736.

Garg, P. (2018). A Review of Podocyte Biology. Am. J. Nephrol. *1*, 3–13.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature *581*, 303–309.

Harrop, T.W., Denecke, S., Yang, Y.T., Chan, J., Daborn, P.J., Perry, T., and Batterham, P. (2018). Evidence for activation of nitenpyram by a mitochondrial cytochrome P450 in *Drosophila melanogaster*. Pest Manag. Sci. *74*, 1616–1622.

Helmstädter, M., Huber, T.B., and Hermle, T. (2017). Using the *Drosophila* Nephrocyte to Model Podocyte Function and Disease. Front. Pediatr. *5*, 262.

Hermle, T., Braun, D.A., Helmstädter, M., Huber, T.B., and Hildebrandt, F. (2017). Modeling Monogenic Human Nephrotic Syndrome in the *Drosophila* Garland Cell Nephrocyte. J. Am. Soc. Nephrol. *28*,1521– 1533.

Hirata T, Cabrero P, Berkholz DS, Bondeson DP, Ritman EL, Thompson JR, Dow JA, Romero MF. (2012). In vivo *Drosophilia* genetic model for calcium oxalate nephrolithiasis. Am. J. Physiol. Renal Physiol. *303*, F1555–F1562.

Hobani Y.H. (2012). Metabolomic analyses of *Drosophila* models for human renal disease. PhD theses, http://theses.gla.ac.uk/id/eprint/3222

Hung, R.J., Hu, Y., Kirchner, R., Liu, Y., Xu, C., Comjean, A., Tattikota, S.G., Li, F., Song, W., Ho Sui, S., et al. (2020). A cell atlas of the adult *Drosophila* midgut. Proc. Natl. Acad. Sci. USA *117*, 1514–1523.

Hu, Y., Vinayagam, A., Nand, A., Comjean, A., Chung, V., Hao, T., Mohr, S.E., and Perrimon, N. (2018). Molecular Interaction Search Tool (MIST): an integrated resource for mining gene and protein interaction data. Nucleic. Acids Res. *46*, D567–D574.

Ichida, K., Amaya, Y., Kamatani, N., Nishino, T., Hosoya, T., and Sakai, O. (1997). Identification of two mutations in human xanthine dehydrogenase gene responsible for classical type I xanthinuria. J. Clin. Invest. *99*, 2391–2397.

Idda, T., Bonas, C., Hoffmann, J., Bertram, J., Quinete, N., Schettgen, T., Fietkau, K., Esser, A., Stope, M.B., Leijs, M.M., et al. (2020). Metabolic activation and toxicological evaluation of polychlorinated biphenyls in *Drosophila melanogaster*. Sci. Rep. *10*, 21587.

Ivy, J.R., Drechsler, M., Catterson, J.H., Bodmer, R., Ocorr, K., Paululat, A., and Hartley, P.S. (2015). Klf15 Is Critical for the Development and Differentiation of Drosophila Nephrocytes. PLoS One *10*, e0134620.

Jenkins, D., Caubit, X., Dimovski, A., Matevska, N., Lye, C.M., Cabuk, F., Gucev, Z., Tasic, V., Fasano, L., and Woolf, A.S. (2010). Analysis of TSHZ2 and TSHZ3 genes in congenital pelvi-ureteric junction obstruction. Nephrol. Dial. Transplant. *25*, 54–60.

Jonusaite, S., Beyenbach, K.W., Meyer, H., Paululat, A., Izumi, Y., Furuse, M., and Rodan, A.R. (2020). The septate junction protein Mesh is required for epithelial morphogenesis, ion transport, and paracellular permeability in the *Drosophila* Malpighian tubule. Am. J. Physiol. Cell Physiol. *318*, C675–C694.

Jung, A.C., Denholm, B., Skaer, H., and Affolter, M. (2005). Renal tubule development in *Drosophila*: a closer look at the cellular level. J Am Soc Nephrol. *16*, 322–328.

King-Jones, K., Horner, M.A., Lam, G., and Thummel, C.S. (2006). The DHR96 nuclear receptor regulates xenobiotic responses in *Drosophila*. Cell Metab. *4*, 37–48.

21

1 Landry, G.M., Hirata, T., Anderson, J.B., Cabrero, P., Gallo, C.J., Dow, J.A., and Romero, M.F. (2016). Sulfate and thiosulfate inhibit oxalate transport via a dPrestin (Slc26a6)-dependent mechanism in an insect model of calcium oxalate nephrolithiasis. Am. J. Physiol. Renal Physiol. *310*, F152–159.

4 Laugier, E., Yang, Z., Fasano, L., Kerridge, S., and Vola, C. (2005). A critical role of teashirt for patterning the ventral epidermis is masked by ectopic expression of *tiptop*, a paralog of *teashirt* in *Drosophila*. Dev. Biol. *283*, 446–458.

7 Le Goff, G., Boundy, S., Daborn, P.J., Yen, J.L., Sofer, L., Lind, R., Sabourault, C., Madi-Ravazzi, L., and ffrench-Constant, R.H. (2003). Microarray analysis of cytochrome P450 mediated insecticide resistance in *Drosophila*. Insect Biochem. Mol. Biol. *33*, 701–708.

10 Linton, S.M., and O'Donnell, M.J. (1999). Contributions of K+: Cl- cotransport and Na+/K+-ATPase to basolateral ion transport in malpighian tubules of *Drosophila melanogaster*. J. Exp. Biol. *202*, 1561–1570.

12 Li, H., Janssens, J., Waegeneer, M., Kolluru, S., Davie, K., Gardeux, V., Saelens, W., David, F., Brbić, M., Leskovec, J., et al. (2021) Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly. bioRxiv. https://doi.org/10.1101/2021.07.04.451050.

15 Liu, Y.F., Hu, Y.H., Li, J.S.S., Rodiger, J., Comjean, A., Attrill, H., Antonazzo, G., Brown, N.H., and Perrimon, N. (2020). FlyPhoneDB: An integrated web-based resource for cell-cell communication prediction in *Drosophila*. bioRxiv. doi: https://doi.org/10.1101/2021.06.14.448430

18 Lu, K., Song, Y., and Zeng, R. (2021). The role of cytochrome P450-mediated detoxification in insect adaptation to xenobiotics. Curr. Opin. Insect Sci. *43*, 103–107.

20 MacMillan, H.A., Nazal, B., Wali, S., Yerushalmi, G.Y., Misyura, L., Donini, A., and Paluzzi, J.P. (2018). Anti-diuretic activity of a CAPA neuropeptide can compromise *Drosophila* chill tolerance. J. Exp. Biol. *221*, jeb185884.

23 Mandal, A.K., and Mount, D.B. (2015). The molecular physiology of uric acid homeostasis. Annu. Rev. Physiol. *77*, 323–345.

25 Marable, S.S., Chung, E., and Park, J.S. (2020). Hnf4a Is Required for the Development of Cdh6-Expressing Progenitors into Proximal Tubules in the Mouse Kidney. J. Am. Soc. Nephrol. *31*, 2543–2558.

27 Marchesin, V., Pérez-Martí, A., Le Meur, G., Pichler, R., Grand, K., Klootwijk, E.D., Kesselheim, A., Kleta, R., Lienkamp, S., and Simons, M. (2019). Molecular Basis for Autosomal-Dominant Renal Fanconi Syndrome Caused by HNF4A. Cell Rep. *29*, 4407–4421.

30 Matsuda, J., Asano-Matsuda, K., Kitzler, T.M., and Takano, T. (2021). Rho GTPase regulatory proteins in podocytes. Kidney Int. *99*, 336–345.

32 McMahon, A.P. (2016). Development of the Mammalian Kidney. Curr. Top. Dev. Biol. *117*, 31–64.

33 Miao, Z., Balzer, M.S., Ma, Z., Liu, H., Wu, J., Shrestha, R., Aranyi, T., Kwan, A., Kondo, A., Pontoglio, M., et al. (2021). Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. Nat. Commun. 12, 2277.

36 Miller, J., Chi, T., Kapahi, P., Kahn, A.J., Kim, M.S., Hirata, T., Romero, M.F., Dow, J.A., and Stoller, M.L. (2013). *Drosophila melanogaster* as an emerging translational model of human nephrolithiasis. J. Urol. *190*,1648–1656.

39 Mitchell, H.K., and Glassman, E. (1959). Hypoxanthine in rosy and maroon-like mutants of *Drosophila melanogaster*. Science *129*, 268.

41 Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., King, E.G., Long, A.D., and Macdonald, S.J. (2015). Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in *Drosophila*. PLoS Genet. *11*, e1005663.

44 Nielsen, S., Kwon, T.H., Fenton, R.A., and Prætorious, J. (2012). Anatomy ofthe kidney. In Brenner &

rector's the kidney, 9th ed. (Elsevier/Saunders), pp. 31–93.

O'Donnell, M.J., and Maddrell, S.H. (1995). Fluid reabsorption and ion transport by the lower Malpighian tubules of adult female *Drosophila*. J Exp Biol. *198*, 1647–1653.

Otasek, D., Morris, J.H., Bouças, J., Pico, A.R., and Demchak, B. (2019). Cytoscape Automation: empowering workflow-based network analysis. Genome Biol. *20*, 185.

Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., and Suszták, K. (2018). Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science *360*, 758–763.

Paul, M.S., Dutta, D., Singh, A., Mutsuddi, M., and Mukherjee, A. (2018). Regulation of Notch signaling in the developing *Drosophila* eye by a T-box containing transcription factor, Dorsocross. Genesis *56*, e23251.

Pearce, D., Soundararajan, R., Trimpert, C., Kashlan, O.B., Deen, P.M., and Kohan, D.E. (2015). Collecting duct principal cell transport processes and their regulation. Clin. J. Am. Soc. Nephrol. *10*, 135–146.

Ransick, A., Lindström, N.O., Liu, J., Zhu, Q., Guo, J.J., Alvarado, G.F., Kim, A.D., Black, H.G., Kim, J., and McMahon, A.P. (2019). Single-Cell Profiling Reveals Sex, Lineage, and Regional Diversity in the Mouse Kidney. Dev. Cell. *51*, 399–413.

Radford, J.C., Davies, S.A., and Dow, J.A.T. (2002). Systematic G-protein-coupled receptor analysis in *Drosophila melanogaster* identifies a leucokinin receptor with novel roles. J. Biol. Chem. *277*, 38810–38817.

Rodan, A.R., Baum, M., and Huang, C.L. (2012). The *Drosophila* NKCC Ncc69 is required for normal renal tubule function. Am. J. Physiol. Cell Physiol. *303*, C883–894.

Rosay P, Davies SA, Yu Y, Sözen MA, Kaiser K, and Dow JA. (1997). Cell-type specific calcium signalling in a *Drosophila* epithelium. J. Cell Sci. *110*, 1683–1692.

Roy, A., Al-bataineh, M.M., and Pastor-Soler, N.M. (2015). Collecting duct intercalated cell function and regulation. Clin. J. Am. Soc. Nephrol. *10*, 305–324.

Sciortino, C.M., Shrode, L.D., Fletcher, B.R., Harte, P.J., and Romero, M.F. (2001). Localization of endogenous and recombinant Na (+)-driven anion exchanger protein NDAE1 from *Drosophila melanogaster*. Am. J. Physiol. Cell Physiol. *281*, C449–463.

Seong, K.M., Coates, B.S., and Pittendrigh, B.R. (2020). Post-transcriptional modulation of cytochrome P450s, *Cyp6g1* and *Cyp6g2*, by *miR-310s* cluster is associated with DDT-resistant *Drosophila melanogaster* strain *91-R*. Sci. Rep. *10*, 14394.

Sözen, M.A., Armstrong, J.D., Yang, M., Kaiser, K., and Dow, J.A.T. (1997). Functional domains are specified to single-cell resolution in a *Drosophila* epithelium. Proc. Natl. Acad. Sci. USA *194*, 5207–5012.

Swale, D.R., Engers, D.W., Bollinger, S.R., Gross, A., Inocente, E.A., Days, E., Kanga, F., Johnson, R.M., Yang, L., Bloomquist, J.R., et al. (2016). An insecticide resistance-breaking mosquitocide targeting inward rectifier potassium channels in vectors of Zika virus and malaria. Sci. Rep. *6*, 36954.

Tarashansky, A.J., Musser, J.M., Khariton, M., Li, P., Arendt, D., Quake, S.R., and Wang, B. (2021). Mapping single-cell atlases throughout Metazoa unravels cell type evolution. Elife *10*, e66747.

Tattikota, S.G., Cho, B., Liu, Y., Hu, Y., Barrera, V., Steinbaugh, M.J., Yoon, S.H., Comjean, A., Li, F., Dervis, F., et al. (2020). A single-cell survey of Drosophila blood. Elife *9*, e54818.

Terhzaz, S., Cabrero, P., Brinzer, R.A., Halberg, K.A., Dow, J.A., and Davies, S.A. (2015). A novel role of Drosophila cytochrome P450-4e3 in permethrin insecticide tolerance. Insect Biochem. Mol. Biol. *67*, 38–46.

Terhzaz, S., Finlayson, A.J., Stirrat, L., Yang, J., Tricoire, H., Woods, D.J., Dow, J.A., and Davies, S.A.

1   (2010). Cell-specific inositol 1,4,5 trisphosphate 3-kinase mediates epithelial cell apoptosis in response to
2   oxidative stress in *Drosophila*. Cell Signal. *22*, 737–748.

3   Torrie, L.S., Radford, J.C., Southall, T.D., Kean, L., Dinsmore. A.J., Davies, S.A., and Dow, J.A.T. (2004).
4   Resolution of the insect ouabain paradox. Proc. Natl. Acad. Sci. USA *101*, 13689–13693.

5   Wang, C., and Spradling, A.C. (2020) An abundant quiescent stem cell population in *Drosophila*
6   Malpighian tubules protects principal cells from kidney stones. Elife *9*, e54096.

7   Wang, J., Kean, L., Yang, J., Allan, A.K., Davies, S.A., Herzyk, P., and Dow, J.A.T. (2004). Function-
8   informed transcriptome analysis of *Drosophila* renal tubule. Genome Biol. *5*, R69.

9   Weavers, H., Prieto-Sánchez, S., Grawe, F., Garcia-López, A., Artero, R., Wilsch-Bräuninger, M., Ruiz-
10   Gómez, M., Skaer, H., and Denholm, B. (2009). The insect nephrocyte is a podocyte-like cell with a
11   filtration slit diaphragm. Nature *457*, 322–326.

12   Wessing, A., and Eichelberg, D. (1978). The genetics and biology of Drosophila Vol. 2c (eds A. Ashburner
13   & T.R.F. Wright) 1–42 (Academic Press).

14   Wu, H., Kirita, Y., Donnelly, E.L., and Humphreys, B.D. (2019). Advantages of Single-Nucleus over Single-
15   Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. J. Am.
16   Soc. Nephrol. *30*, 23–32.

17   Wu, Y., Baum, M., Huang, C.L., and Rodan, A.R. (2015). Two inwardly rectifying potassium channels, Irk1
18   and Irk2, play redundant roles in Drosophila renal tubule function. Am. J. Physiol. Regul. Integr. Comp.
19   Physiol. *309*, R747–756.

20   Yang, M.Y., Wang, Z., MacPherson, M., Dow, J.A., and Kaiser, K. (2000). A novel Drosophila alkaline
21   phosphatase specific to the ellipsoid body of the adult brain and the lower Malpighian (renal) tubule.
22   Genetics *154*, 285–297.

23   Yang, J., McCart, C., Woods, D.J., Terhzaz, S., Greenwood, K.G., ffrench-Constant, R.H., and Dow, J.A.T.
24   (2007). A *Drosophila* systems approach to xenobiotic metabolism. Physiol Genomics. 30, 223–231.

25   Zhang, F., Zhao, Y., and Han, Z. (2013). An in vivo functional analysis system for renal gene discovery in
26   Drosophila pericardial nephrocytes. J. Am. Soc. Nephrol. *24*,191–197.

27   Zhang, F., Zhao, Y., Chao, Y., Muir, K., and Han, Z. (2013). Cubilin and amnionless mediate protein
28   reabsorption in *Drosophila* nephrocytes. J. Am. Soc. Nephrol. 24, 209–216.

29   Zhou, H., Whitworth, C., Pozmanter, C., Neville, M.C., and Van Doren, M. (2021). Doublesex regulates
30   *fruitless* expression to promote sexual dimorphism of the gonad stem cell niche. PLoS Genet. *17*,
31   e1009468.

32   Zhuang, S., Shao, H., Guo, F., Trimble, R., Pearce, E., and Abmayr, S.M. (2009). Sns and Kirre, the
33   Drosophila orthologs of Nephrin and Neph1, direct adhesion, fusion and formation of a slit diaphragm-like
34   structure in insect nephrocytes. Development *136*, 2335–2344.

35

36   **Figure legends**

37   **Figure 1. snRNA-seq analysis of the fly adult kidney.** (A) Experimental design. The
38   location of the Malpighian tubules and two types of nephrocytes are shown. Nuclei
39   from Malpighian tubules and nephrocytes were processed separately and
40   encapsulated using 10x Genomics. Data analysis was conducted independently and
41   then combined to generate a single UMAP of the "fly kidney". Note, however, that
42   nephrocytes and tubules are not physically associated in vivo. (B) 11 distinct cell
43   clusters were annotated on the UMAP. (C) Expression levels and percentage of cells

24

expressing the marker genes in each cluster are shown as a dot plot. (D) GFP expression under the control of Gal4 lines specific for each of the six PC clusters. C42-Gal4 is expressed in all PCs. Scale bars = 500 µm. (E) Zoom-in of D panels to show the local features. Scale bars = 100 µm. (F) Malpighian tubule cell types are identified based on differentially expressed marker genes.

Figure 2. Mapping function to cell types and regions in the tubule. (A) Overview of the two-cell model of insect tubule fluid secretion and its control. Adapted from Dow et al. (2021). (B-D) UMAP distribution of genes involved in principal cell, stellate cell and junctions.

Figure 3. *RhoGEF64c* maintains stellate cell shape. (A) UMAP of two sub-clusters of SCs in the Malpighian tubules. (B) Gene expression level in the population of all SCs. (C) Cell shape visualized using tsh-Gal4 driving mCD8-GFP. DAPI (blue) is used to stain nuclei. White box indicates the zoom-in region. (D) Knockdown using VDRC line 47121^v of *RhoGEF64c* affects SC cell shapes. (E and F) Knockdown of *RhoGEF64c* results in loss of cytoarchitectural organization. Cell cytoarchitecture is visualized by Phalloidin (Phal; F-actin) staining. Arrows indicate SCs. Scale bars = 50 µm.

Figure 4. Cell type specific gene regulatory landscape of the fly kidney. (A) Heat map profile of the transcription factors (TFs) in all clusters. Genes were ranked based on expression levels (fold change > 3) and adjusted p-value (<0.05) in each condition in the heat map. (B-E) Expression of *fru*, *Sba*, *Doc2* and *Lim3* visualized using fru-Gal4, Sba-Gal4, Doc2-Gal4 and Lim3-Gal4 driven UAS-CD8::GFP expression, respectively. Scale bars = 100 µm in B and C. Scale bars = 500 µm in D and E. (F) SCENIC results of the fly kidney. The heatmap shows the gene expression level in each cluster. Low regulon activity is shown with blue color and high regulon activity is shown in red. See Sup. FigS8 for an enlarged version of the heat map with gene names. (G) UMAP depiction of regulon activity ("on-blue", "off-gray") and TF gene expression (blue scale) of RSCs (*esg*), SCs (*Lim3*), and PCs (*Hnf4*). Examples of target gene expression of the *esg* regulon (*Notch* (*N*) and *fruitless* (*fru*)), *Lim3* regulon (*RhoGEF64C* and *u-shape*d (*ush*)) and *Hnf4* regulon (*Arginine kinase* (*Argk*) and *midway (mdy)*) are shown in blue.

Figure 5. Gene distribution of the Uric Acid pathway in the fly kidney. (A) Uric acid pathway in human and fly. The end product is uric acid in human and allantoin in the fly. Right panel, the enzymes involved at each step. (B) Expression levels of each gene of the uric acid pathway visualized by UMAP. (C) The UMAP plot shows the gene set activity of all genes in the uric acid pathway. (D) Gene set activity of the last three steps, *rosy* (*ry*), *Urate oxidase* (*Uro*) and *CG30016,* visualized by UMAP plots.

1   **Figure 6. Cross-species analysis of fly, planarian and mouse kidneys using**
2   **SAMap.** (A and B) Sankey plot summarizing the cell type mappings. Edges with
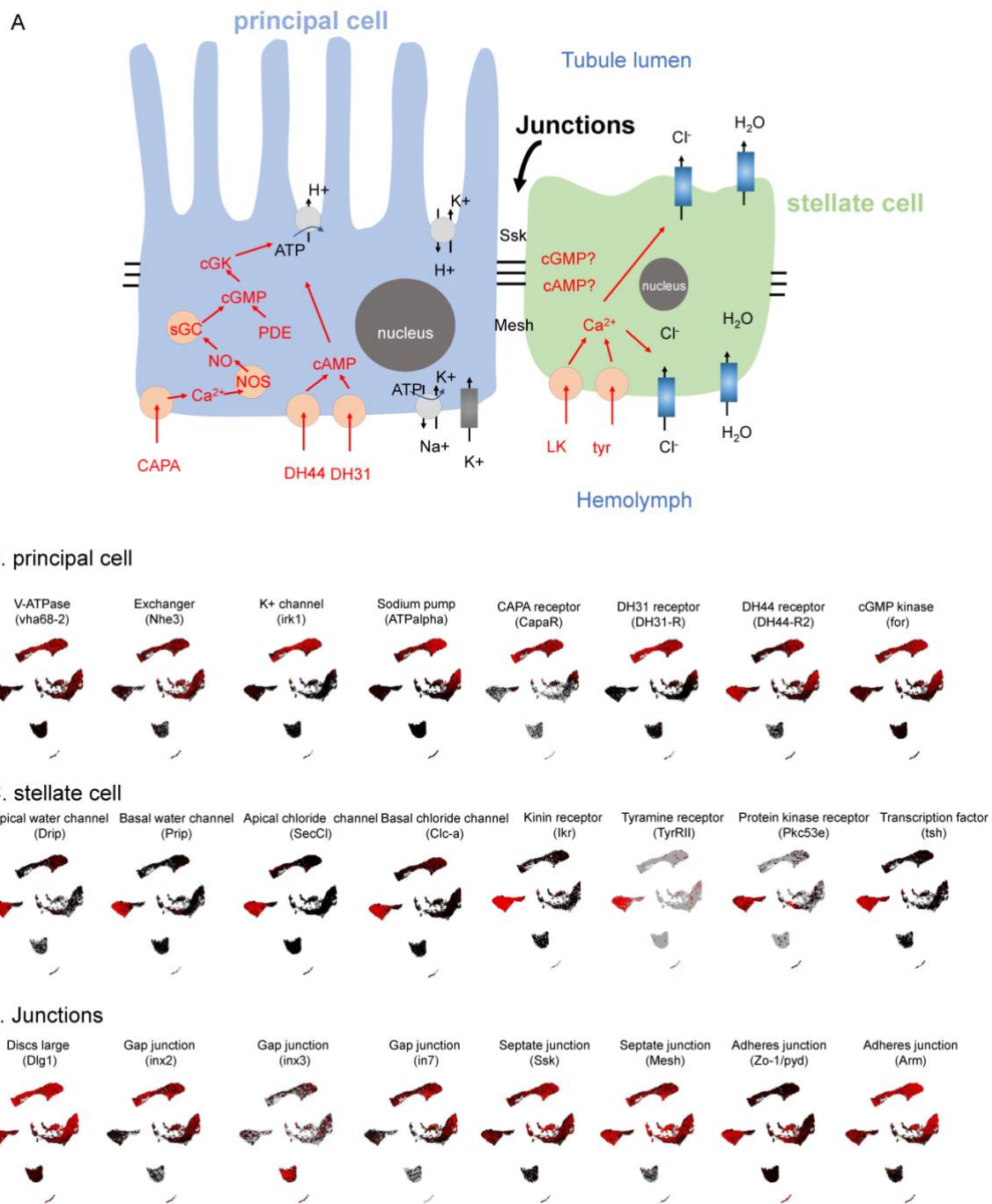3   alignment scores < 0.1 were omitted.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
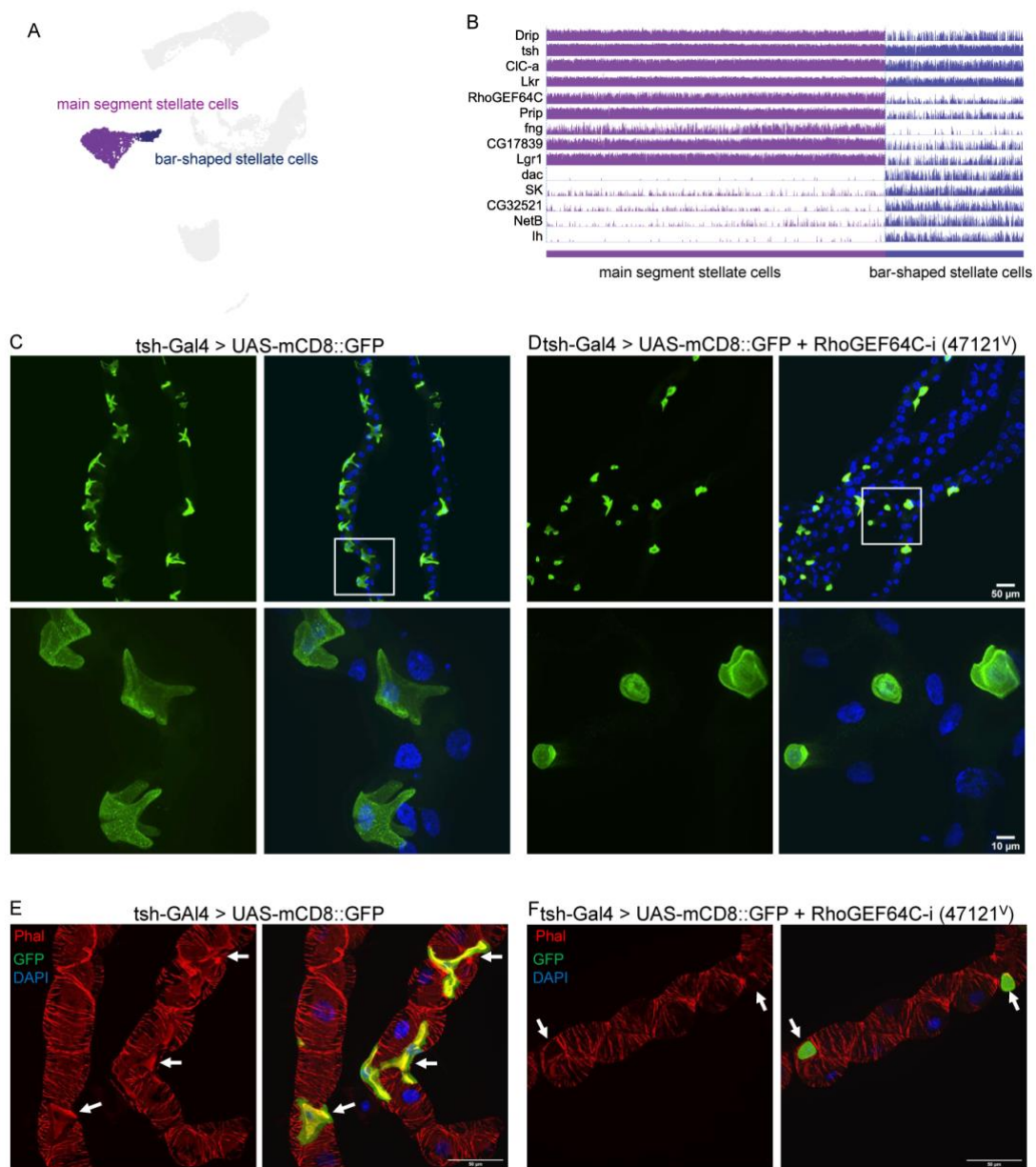33
34
35
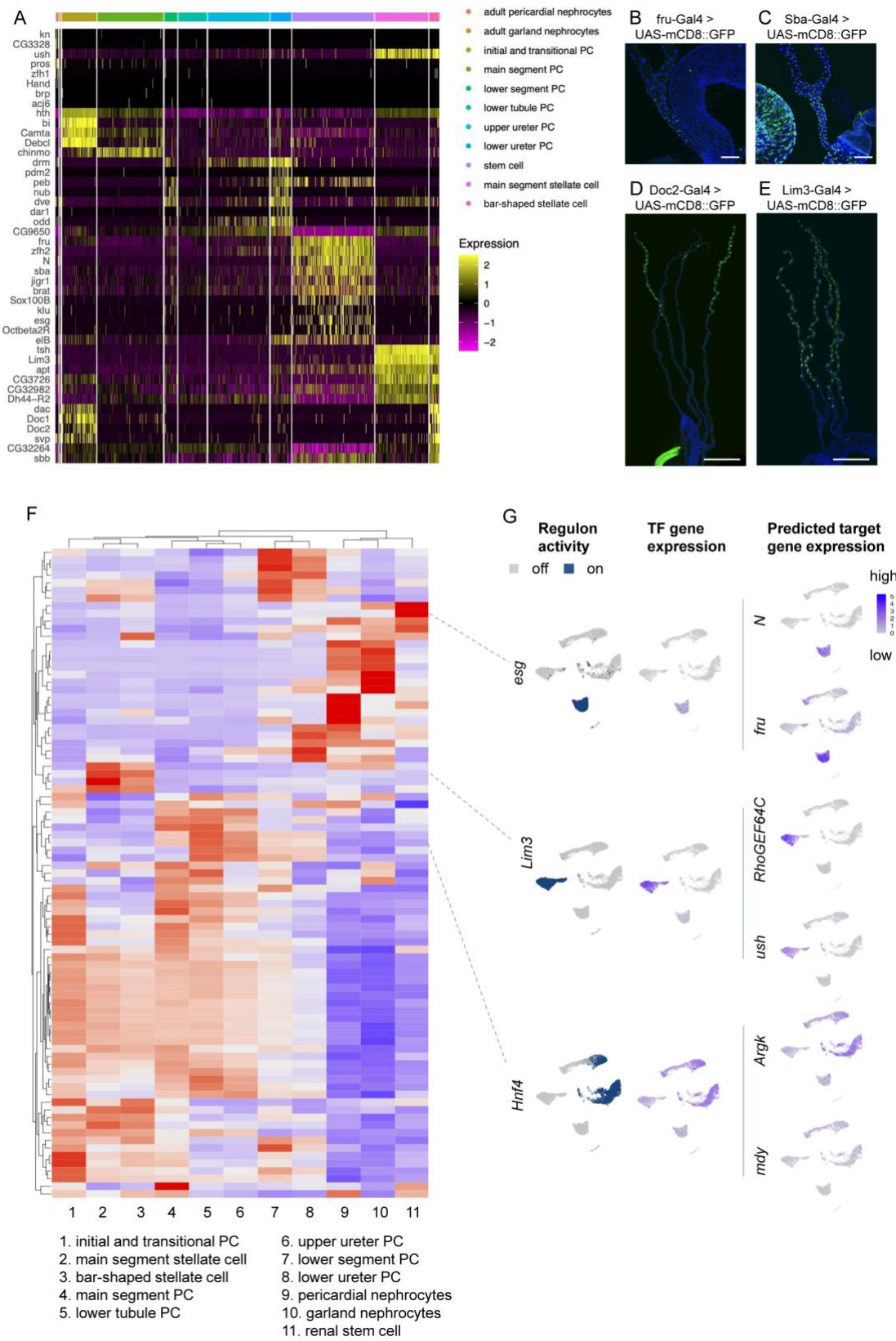36
37
38
39
40
41
42
43
44

1   **Fig.1**
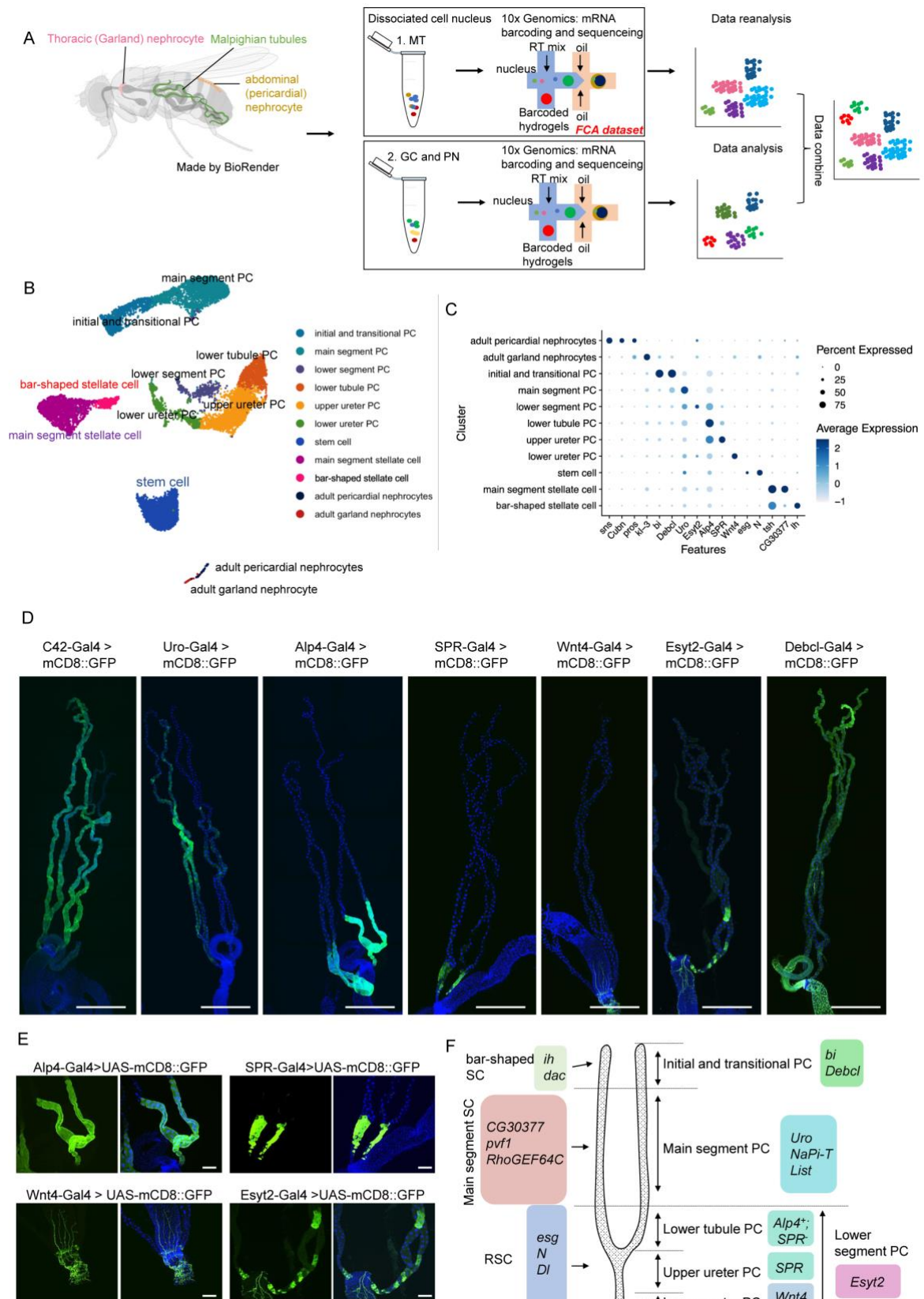
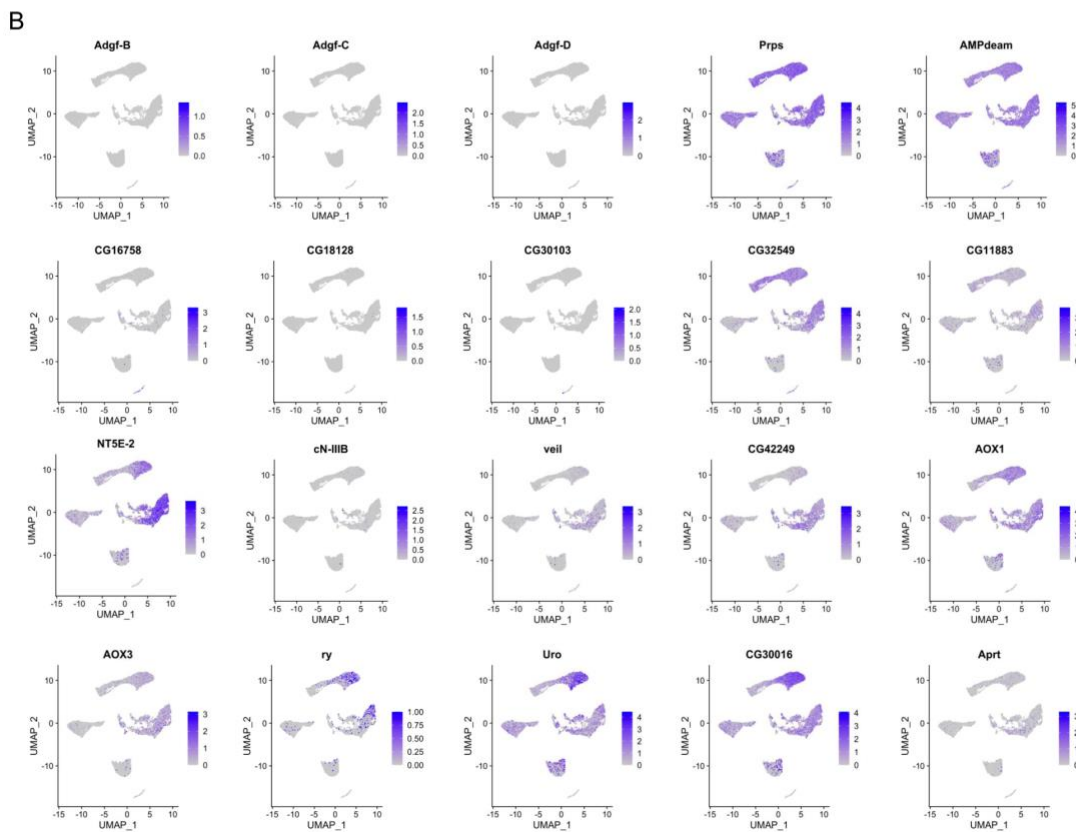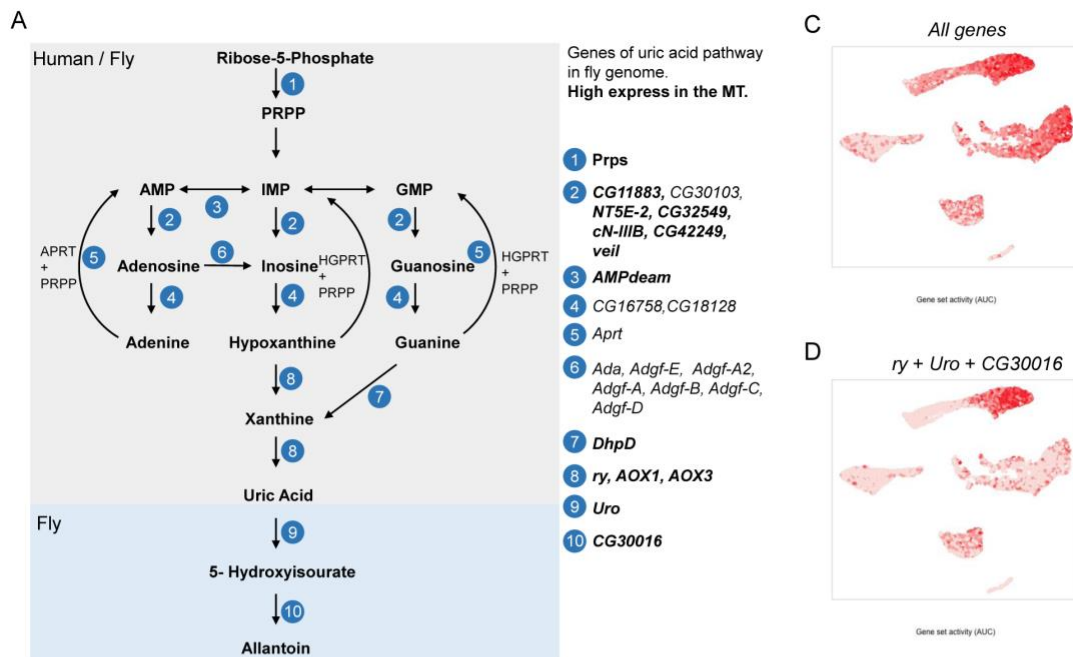2

1 **Fig.2**

2
3
4
5
6
7
8
9
10
11

1 **Fig.3**

2
3
4
5
6
7
8
9
10
11
12
13
14

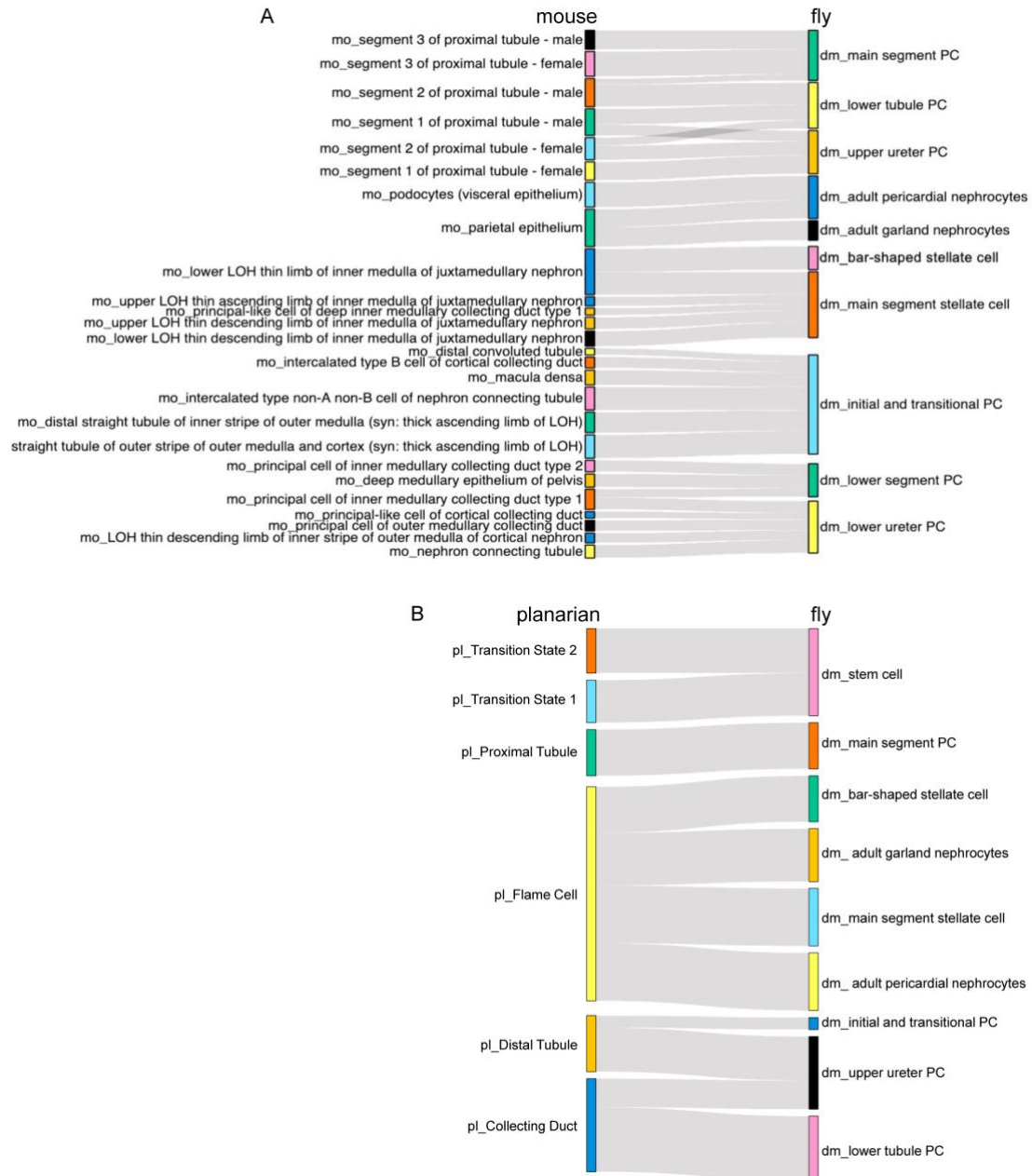29

1 **Fig.4**

2
3
4
5

1   **Fig.5**

2
3
4
5
6
7
8

1

2 **Fig.6**

3
4

5

6

7

8

9

1 **Supplemental Information**

2 **A cell atlas of the fly kidney**

3 Jun Xu, Yifang Liu, Hongjie Li, Alexander J. Tarashansky, Colin H. Kalicki, Ruei-Jiun

4 Hung, Yanhui Hu, Aram Comjean, Sai Saroja Kolluru, Bo Wang, Stephen R Quake,

5 Liqun Luo, Andrew P. McMahon, Julian A.T. Dow, Norbert Perrimon

6

7 **This supplementary information contains:**

8

9 **Supplementary text**

10 **Supplementary Figure 1- Supplementary Figure 16.**

11 **Supplementary references**

12

13

14 **Supplementary Table legends:**

15 **Supplementary Table 1.** Basic statistics of snRNA-seq libraries.

16 **Supplementary Table 2.** Differentially expressed genes in each cluster. Only

17 positive marker genes are shown.

18 **Supplementary Table 3.** Table of validated markers, from previous studies

19 and this study, allowing assignment of clusters to cell types or regions.

20 **Supplementary Table 4.** GO terms comparison of bar-shaped SCs and main

21 segment SCs.

22 **Supplementary Table 5.** GO terms of the six PC cell clusters.

23 **Supplementary Table 6.** Full list of regulons and their respective predicted

24 target genes.

25 **Supplementary Table 7.** List of cell type-specific transcription factors.

26 **Supplementary Table 8.** Gene pairs for fly and mouse cell type mappings.

27 **Supplementary Table 9.** Gene pairs for fly and planarian cell type mappings.

28 **Supplementary Table 10.** Gene list comparison of RSCs and ISCs.

29 **Supplementary Table 11.** GO terms comparison of RSCs and ISCs.

30 **Supplementary Table 12.** List of gene pairs for cell-cell communication

31 predictions.

32

33

34

35

36

37

38

39

40

41

Supplementary text

Similarities between renal and intestinal stem cells

Compared to a previous report that performed scRNAseq of the Mapighian tubules focusing on the stem cell zone (Wang and Spradling, 2020), our study captured more cells and more cell types. Wang and Spradling focused on the response of RSCs to tissue injury. RSCs were previously identified as a distinct population that expresses *esg* (Singh et al., 2007). RSCs are located in the lower ureter, the upper ureter and lower segment of the MT with small nuclei (Takashima et al., 2013). They respond to tissue injury by upregulating the JNK, EGFR/MAPK, Hippo/Yki and JAK/STAT pathways that promote RSC daughter differentiation (Wang and Spradling, 2020). RSCs originate from the same pool of adult midgut progenitors that generate the posterior midgut intestinal stem cell (ISCs) (Takashima et al., 2013; Xu et al., 2018). To examine how similar RSCs are to ISCs, we compared the snRNA-seq RSC data with previously reported scRNAseq ISC data (Hung et al., 2020). Consistent with their common origin (Takashima et al., 2013), the two stem cell clusters have high similarity at the gene expression level compared to other cell clusters (Fig. S14A).

The *esg* gene, a stem cell marker for both RSCs and ISCs (Fig. S14B and 14C), encodes a transcription factor that contributes to stem cell maintenance through modulation of Notch activity (Loza-Coll et al., 2014). In the intestine, *esg* is not only expressed in ISCs but also in AstC-EEs (enteroendocrine cells that express *Allatostatin C, AstC*) and NPF-EEs (EEs that express *neuropeptide F*, *NPF*) (Hung et al., 2020). In the intestine, ISCs are highly mitotic, especially during regeneration, and give rise to a transient progenitor, the enteroblast (EB) (Ohlstein and Spradling, 2006; Micchelli and Perrimon, 2006), whereas in the Malpighian tubule, RSCs normally divide very slowly (Wang and Spradling, 2020). Interestingly, 56 genes overlapped between RSCs and ISCs (MT$^+$gut$^+$, genes highly expressed in RSCs and ISCs), including *esg*, *N*, *Dl*, *klumpfuss* (*klu*), and *Sox100B* (Fig. S15A, all genes are listed in Supplementary Table 10. Hung et al., 2020). Gene Ontology (GO) analysis reveals that MT$^+$gut$^+$ genes are enriched in cell differentiation, proliferation, and stem cell division. However, GO terms of MT$^+$gut$^-$ (genes highly expressed in RSCs, low or not expressed in ISCs) mainly contain genes annotated as involved in growth, tube morphogenesis, and epithelial cell differentiation. Finally, GO terms of MT$^-$gut$^+$ (genes highly expressed in ISCs, low or no expressed in RSCs) mainly represent genes involved in protein folding, translational initiation, and peptide biosynthesis (Fig. S15B, all GO terms are listed in Supplementary Table 11). Protein folding is relevant to the stress response, reflecting damage to the gut caused by the food. Altogether, these analyses suggest that RSCs and ISCs have a common origin but are in different cell states.

In our previous gut scRNA-seq study, ISCs/EBs were annotated as one cluster based on the expression of *Dl* and *esg*. However, this cluster could be split into ISCs and EBs, as one subset of cells in the ISC/EB cluster is *Dl$^+$ klu$^-$* and another subset is *Dl$^-$ klu$^+$* (Hung et al., 2020). Interestingly, we could also identify two RSC sub-clusters based on the expression of *Dl$^+$ klu$^-$* and *Dl$^-$ klu$^+$* (Fig. S14D and S14E). The *Dl$^+$ klu$^-$* sub-cluster specifically expresses *Dl*, *N*, and *esg*, reminiscent to the *Dl$^+$ klu$^-$* sub-

34

cluster of ISC (Fig. S14F), whereas the *Dl⁻ klu⁺* sub-cluster specifically expresses *E(spl)m3-HLH*, *E(spl)malpha-BFM, E(spl)mbeta-HLH*, which are transcription factors executing Notch-mediated cellular differentiation (Couturier et al., 2019; Lu and Li, 2015, Fig. S14F).

Cell-cell communication networks in the fly kidney

Previous studies have indicated that the survival, renewal, and differentiation of PCs and SCs are largely regulated through their cross-talk with RSCs (Singh et al., 2007; Takashima et al., 2013; Wang and Spradling, 2020). We used FlyPhoneDB (Liu et al, 2021) to explore cell–cell communication between the different fly kidney cell clusters. FlyPhoneDB was established recently and provides predictions of ligand-receptor interactions based on fly scRNA-seq data. We analyzed 13 major pathways and indicated their cell–cell interaction pairs between the different cell clusters (Fig.S16A). Strikingly, the Notch ligand only has interaction within RSCs and does not pair with other cell clusters (Fig. S16A). This is consistent with previous studies showing that differential Notch activity is required for RSC homeostasis and that damage activates Notch signaling, which in turn regulates differentiation of RSCs to PCs (Li et al., 2014; Wang and Spradling, 2020). Further, we found that the EGFR signaling pathway connects RSCs and all SCs and PCs, with a preferentially strong interaction with main segment SCs and main segment PCs (Fig. S16A). These are consistent with previous studies showing that EGFR is dispensable for RSC maintenance but required for RSC proliferation (Li et al., 2015). In addition, FlyPhoneDB predicts a strong interaction from main segment SCs to main segment PCs with the Pvf1-Pvr ligand-receptor pair (Fig. S16B). This interaction is based on the gene expression pattern in cell clusters of main segment SCs and PCs via MIST database and TF2TG literatures (Fig. S16C). Consistent with this, *pvf1* was highly expressed in main segment SCs as detected using *pvf1-Gal4>mCD8::GFP* (Fig. S16D). Altogether, FlyPhoneDB predicts a number of specific signaling events between Malpighian tubule cell clusters. The full list of predicted gene pairs can be found in Supplementary Table 12.

1
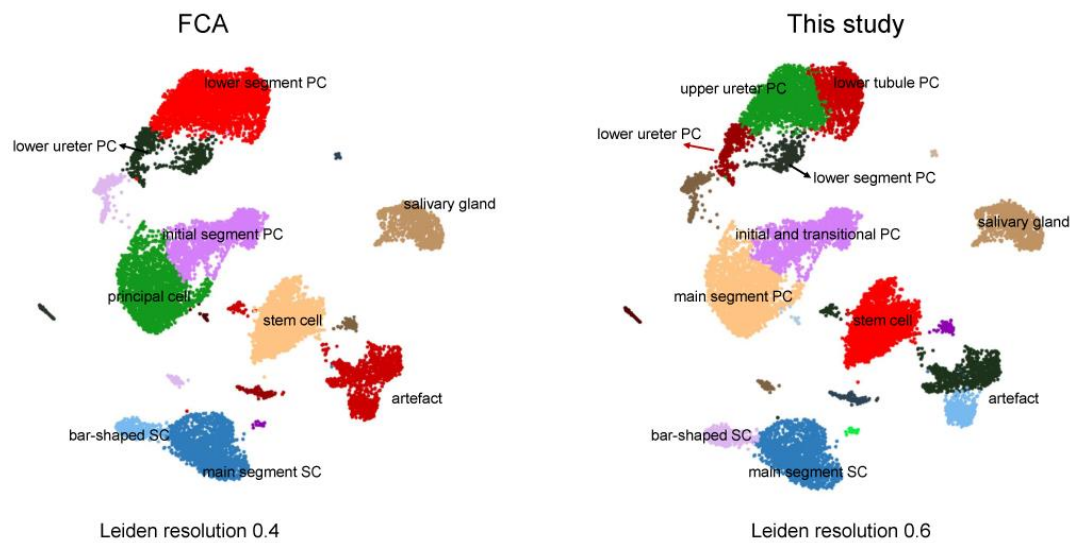
Figure S1. High resolution snRNA-seq analysis of the adult MT. Left, UMAP of the MTs FCA data set at Leiden resolution 0.4 (Li et al., 2021). Right, annotation of the same data set at Leiden resolution 0.6 (this study). The FCA analysis reports four clusters for principal cells: lower ureter PC, lower segment PC, principal cell, and initial segment PC. In this study, we defined six clusters for principal cells based on Gal4 reporter lines: lower ureter PC, lower ureter PC, lower tubule PC, lower segment PC, main segment PC, and initial segment PC. Note that this figure contains all the original clusters, including non-Malpighian tubule cell clusters (salivary glands, artefacts), which we did not include in Fig. 1B.
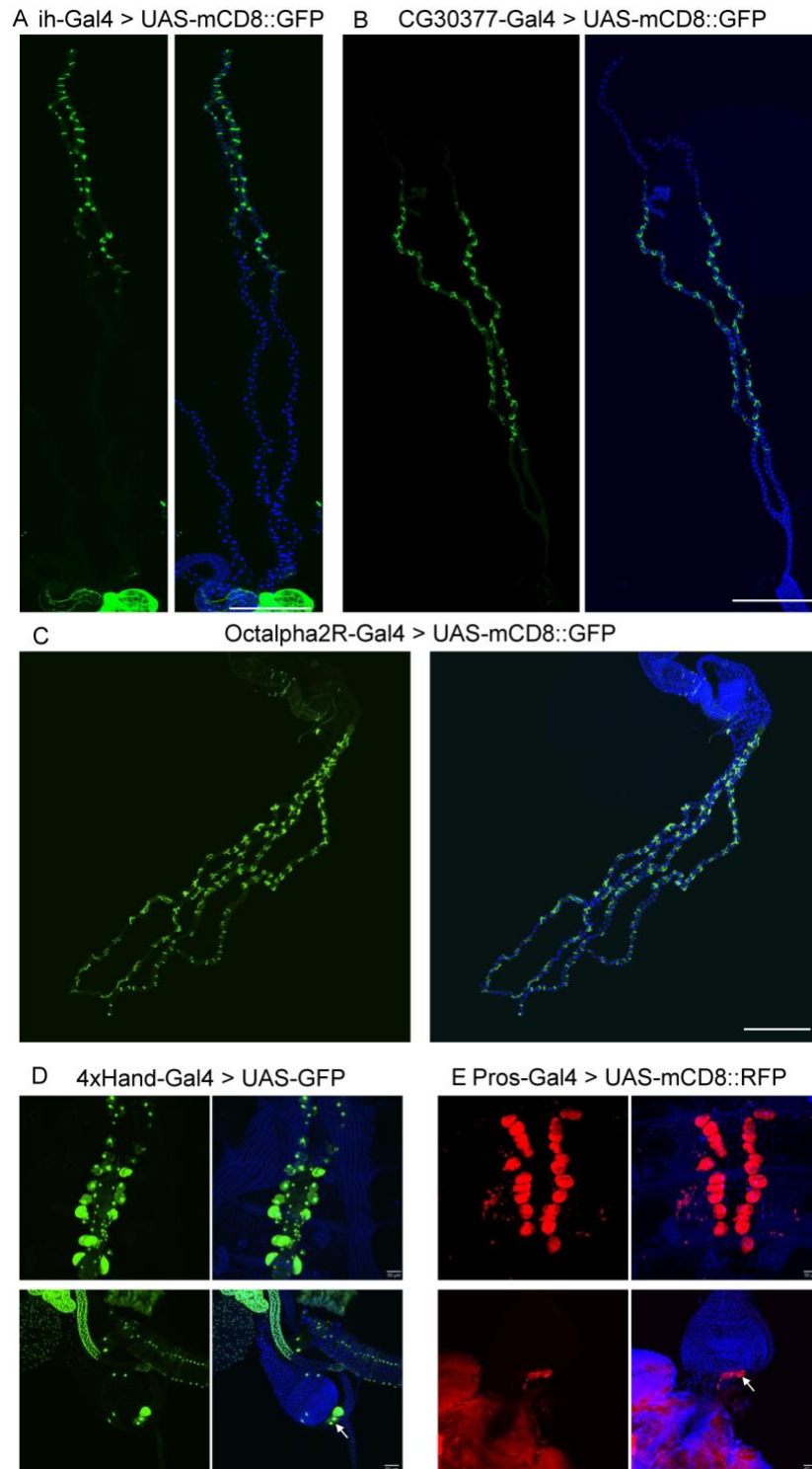
Figure S2. Gene expression in stellate cell clusters and nephrocytes. (A-C) The expression of three new marker genes for SCs is shown using Gal4 lines driving UAS-mCD8::GFP: I_h *channel* (*ih*) (bar-shaped SC), *CG30377* (main segment SCs) and *Octalpha2R* (all SCs). Scale bars = 500 µm. (D and E) *Hand* and *Prospero (Pros)* genes are previously known marker genes. *4xHand-Gal4* line is a four copy enhancer sequences of the *Hand* gene driving Gal4 (Zhu et al., 2017). Arrows indicate GCs. Scale bars = 50 µm.

37
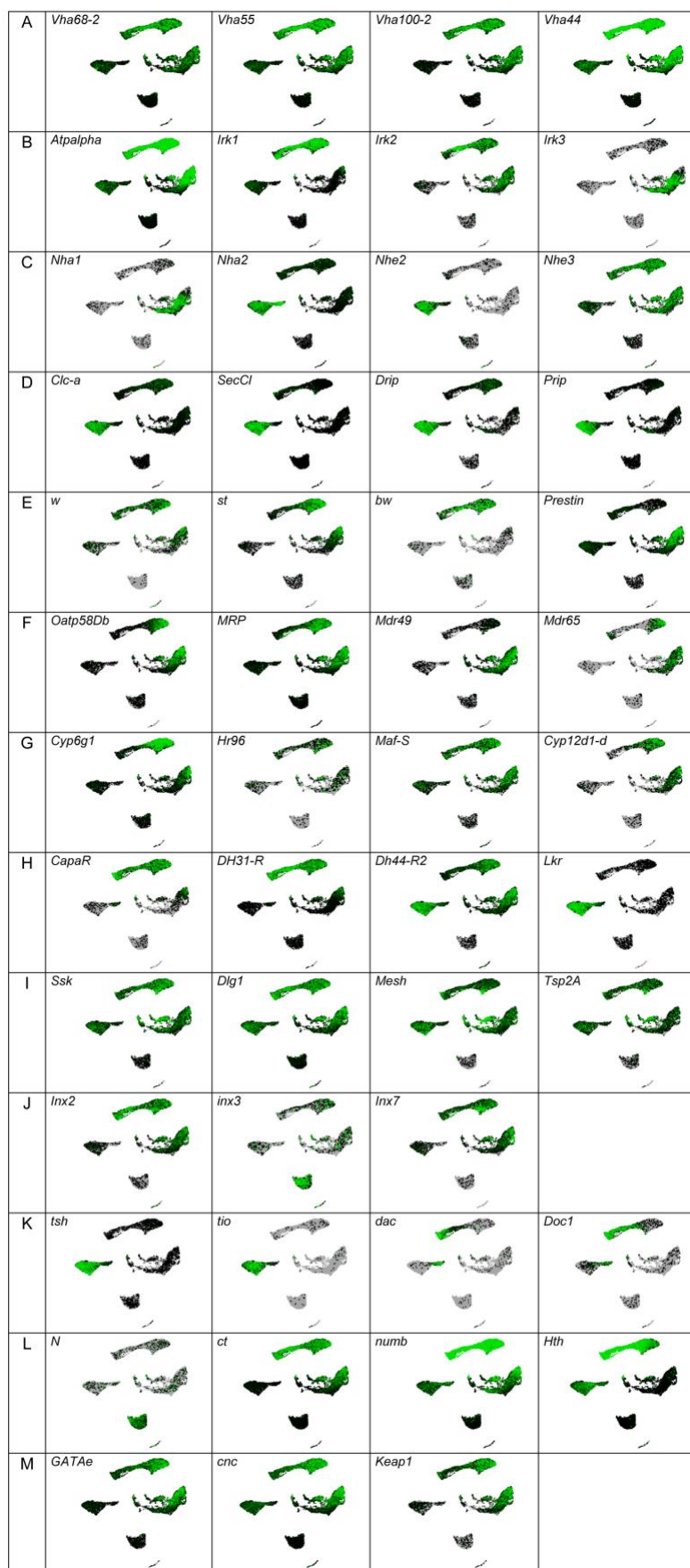
Figure S3. Expression encyclopedia of function-linked tubule genes.

1
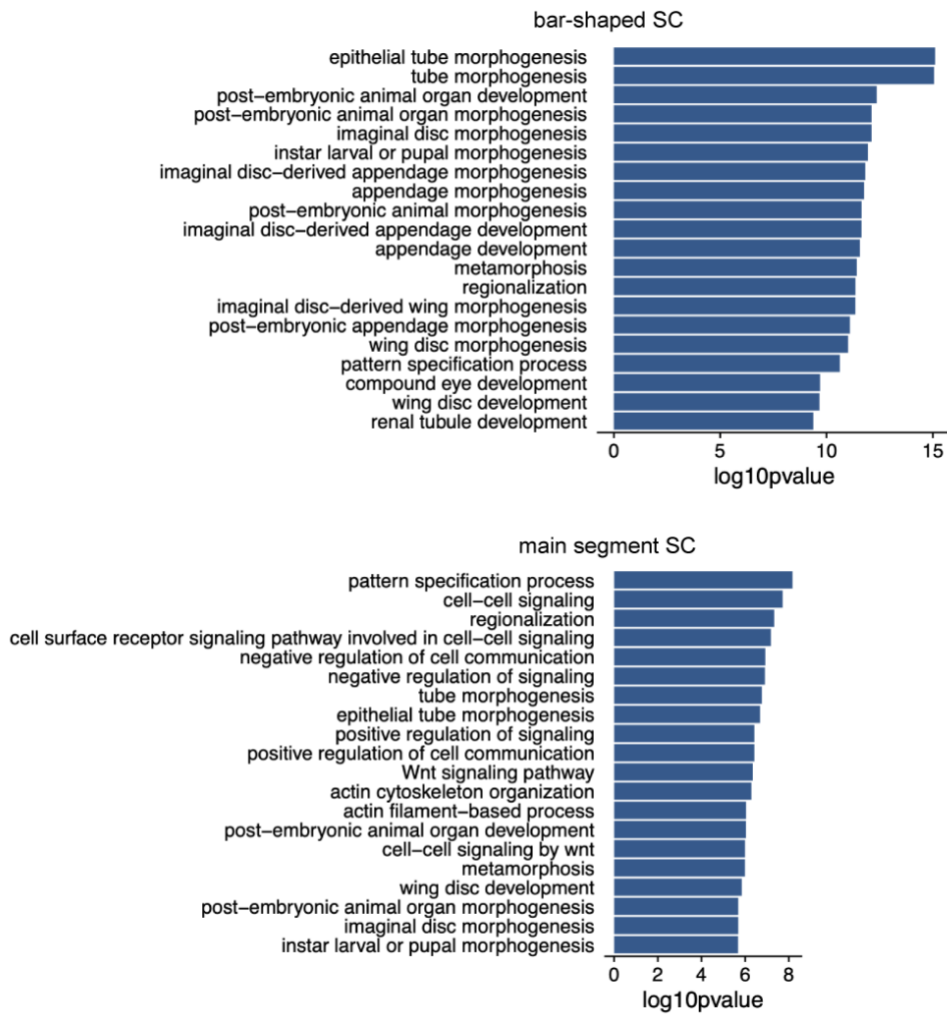
2

3 **Figure S4. Gene Ontology (GO) analysis of bar-shaped SC and main segment SC**
4 **clusters.** Bar-shaped SCs are enriched for epithelial tube morphogenesis and tube
5 morphogenesis GO terms. These cells specifically express the cell adhesion genes
6 *turtle* (*tutl*), *Tenascin accessory* (*Ten-a*) and *echinoid* (*ed*), the transcription factors
7 *dachshund* (*dac*), *Dorsocross2* (*Doc2*) and *Doc1*, and the potassium channels *tiwaz*
8 (*twz*), I_h *channel (ih)* and *small conductance calcium-activated potassium channel* (*SK*)
9 (Supplementary Table 4). Main segment SCs are enriched for pattern specification
10 process, cell-cell signaling, regionalization, and cell surface receptor signaling pathway
11 involved in cell-cell signaling GO terms. They express a number of hormones and
12 neuropeptide receptors (*Leucine-rich repeat-containing G protein-coupled receptor 1*
13 (*Lgr1*), *Octα2R*, *Tachykinin-like receptor at 99D* (*TkR99D*) and *Leucokinin receptor*,
14 (*Lkr*)), chloride channels (*SecCl* and *Chloride channel-a* (*Clc-a*)), and aquaporins (*Prip*
15 and *Drip*) (Supplementary Table 4). The top 20 terms are displayed.

39

A

♀ : tsh-Gal4/*Cy*; UAS-mCD8::GFP

x

♂ : TRiP-RNAi / VDRC-RNAi

↓

1. Count fly number of *Cy* or *Cy+* phenotype (confirmed by GFP)

2. Shape change of stellate cell

3. Number change of stellate cell

B

	Gene name	Identifier	Cy+	Cy	Cy+/Cy	Lethal stage	Shape of SC	Number of SC	Function
1	luciferase	BL31603	183	192	95.31%	-	-	-	-
2	tutl	BL54850	71	331	21.45%	early adult	-	-	Cell–cell adhesion mediator
3	prip	BL50695	63	261	24.14%	lava, embryo	Changed	-	aquaporin
4	wat	BL67801	55	200	27.50%	most die at pupa	-	-	Waterproof, liquid clearance
5	CG30377	BL51386	19	48	39.58%	lava, embryo	-	-	copper ion homeostasis
6	Lgr1	BL51465	30	74	40.54%	lava, embryo	-	-	hormone receptor
7	CG10939	BL65156	60	139	43.17%	lava, embryo	Changed	-	scaffold linking the plasma membrane
8	qvr	BL58061	65	134	48.51%	lava, embryo	-	-	GPI anchor binding
9	ih	BL58089	88	155	56.77%	lava, embryo	-	-	Potassium channel
10	axed	BL62928	123	196	62.76%	most die at pupa	-	-	neurogenesis
11	Octα2R	BL50678	83	125	66.40%	lava, embryo	-	-	octopamine receptor
12	notum	BL55379	92	135	68.15%	lava, embryo, pupa	-	-	pectinacetylesterase
13	TkR99D	BL27513	80	115	69.57%	-	-	-	tachykinin receptor
14	CG13323	BL53969	162	211	76.78%	-	-	-	unknown
15	prip	BL44464	63	75	84.00%	-	-	-	aquaporin
16	CG42594	BL35006	68	77	88.31%	-	-	-	ion leak channel
17	nep2	BL61902	144	163	88.34%	-	-	-	Peptidase
18	ncc69	BL28682	132	149	88.59%	-	-	-	sodium chloride cotransporter
19	frq2	BL28711	109	103	105.83%	-	Changed	Reduced	neuromuscular junction development
20	RhoGEF64C	BL31130	36	43	83.72%	-	-	-	regulation of Rho protein signal transduction
21	VDRC-Con	V60100	43	44	97.73%	-	-	-	-
22	RhoGEF64C-V1	V47121	4	111	3.60%	pupa, lava, embryo	Changed	Reduced	-
23	RhoGEF64C-V2	V105252	19	56	33.93%	pupa, lava, embryo	Changed	Reduced	-

C

D

E tsh-Gal4 > UAS-mCD8::GFP + Luc-i

F tsh-Gal4 > UAS-mCD8::GFP + Prip-i

G tsh-Gal4> UAS-mCD::-GFP + CG10939-i

H tsh-Gal4 > UAS-mCD8::GFP + frq2-i

Figure S5. Phenotypes associate with the top stellate cell marker genes. (A) Screening strategy. SCs were visualized by tsh-Gal4 driving mCD8::GFP. (B) List of the genes tested in the screen, RNAi line identifiers, ratio of Cy+ versus Cy progenies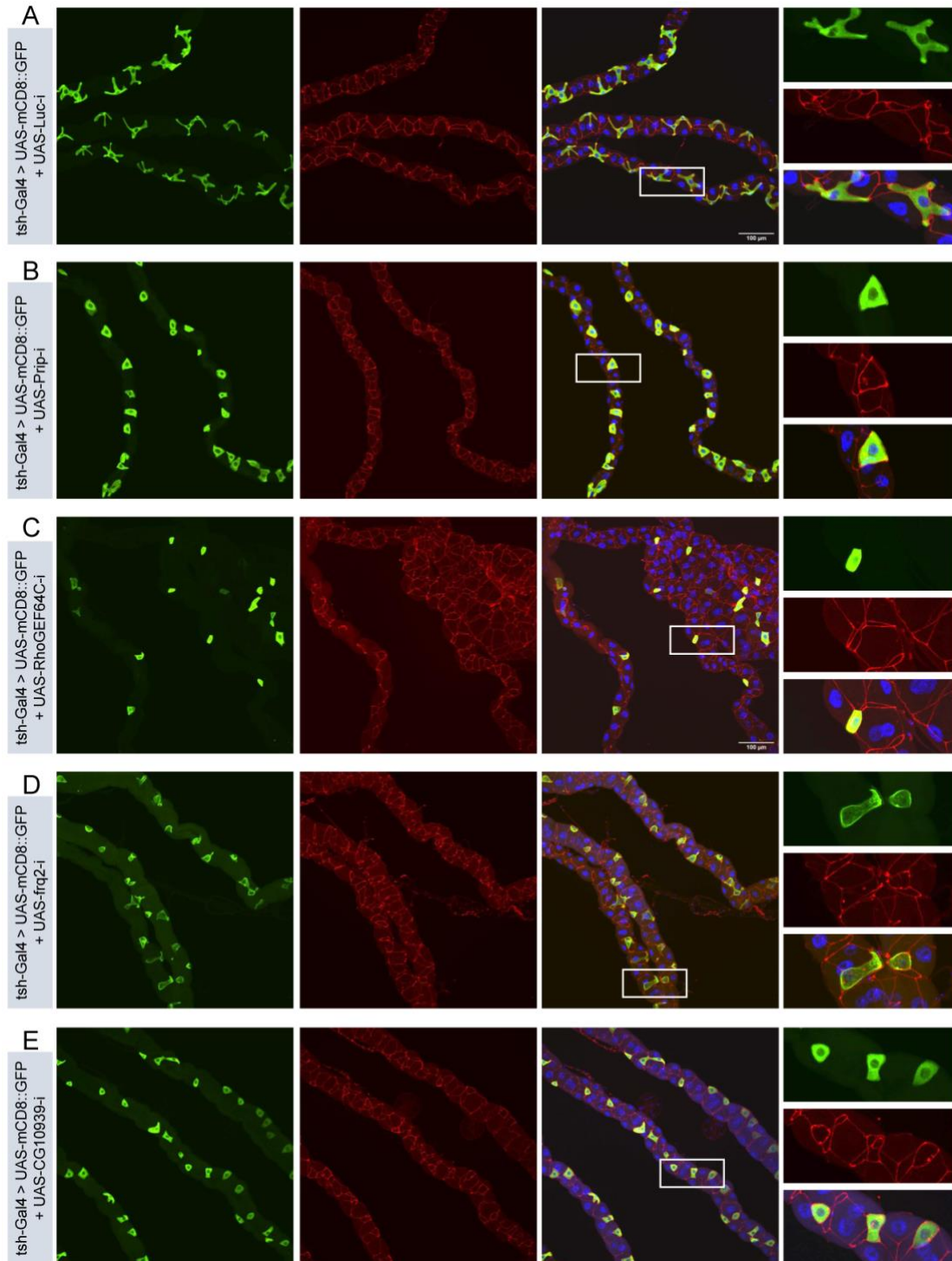, stage of lethality, effects on cell shape and number, and a short description of gene function. (C) Dot plots indicating the expression level of candidate genes in bar-shaped SCs and main segment SCs. (D) Histogram showing the survival rate. (E-H) SC cell shape phenotypes associated with RNAi knockdown of *Prip*, *CG10939* or *Frequenin 2* (*Frq2)*. DAPI (blue) staining for nuclei.

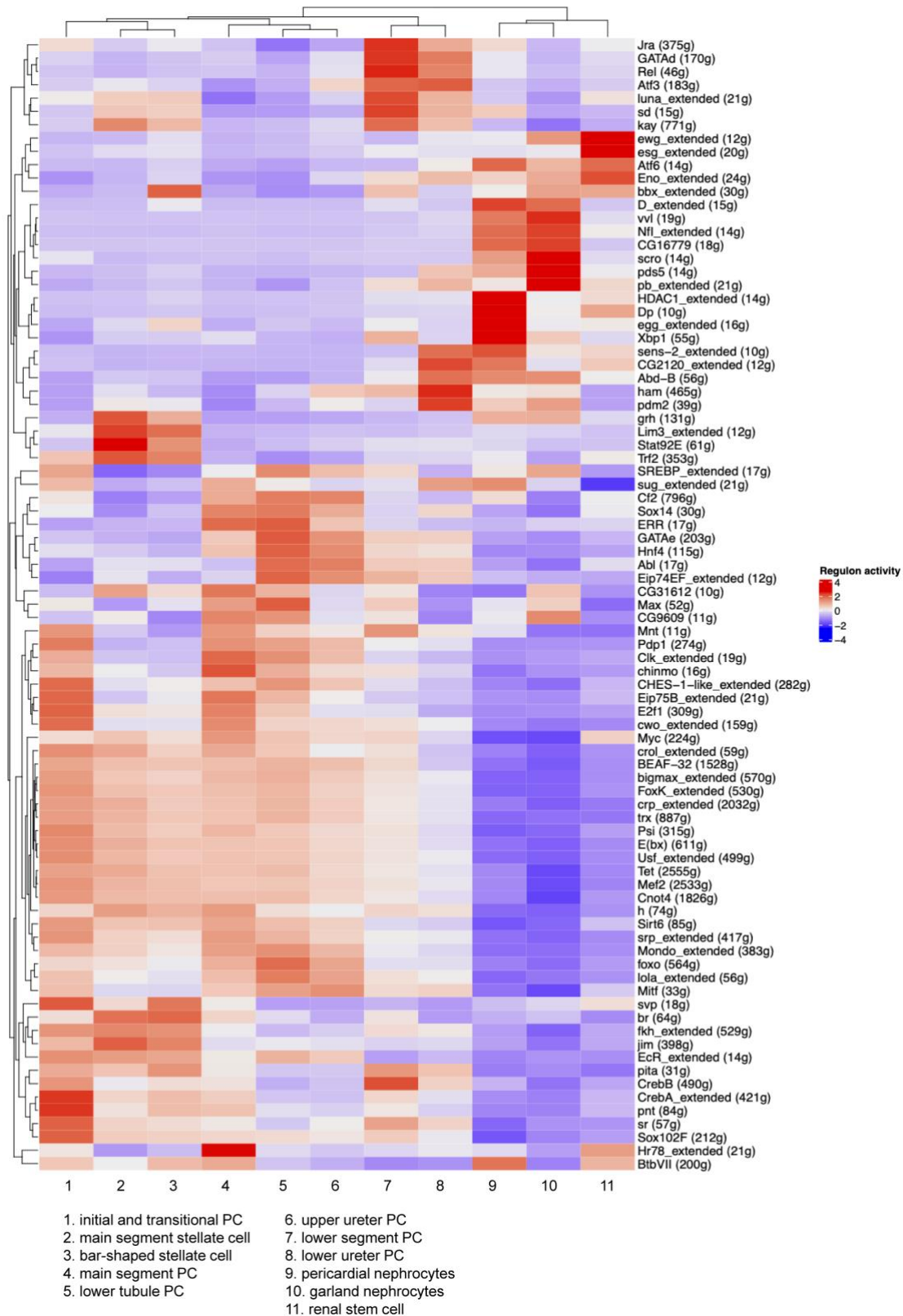**Figure S6. Morphology of septate junctions in Malpighian tubules.** tsh-Gal4 drives mCD8::GFP expression in the adult Malpighian tubules in evenly spaced SCs. Septate junctions are labelled using anti-Dlg (red). White boxes indicate the zoomed-in regions. DAPI (blue) staining for nuclei. (A) *Luciferase* RNAi control. (B-E) Phenotypes associated with RNAi knockdown of *Prip*, *RhoGEF64c*, *Frq2* or *CG10939*. Scale bars = 100 μm.
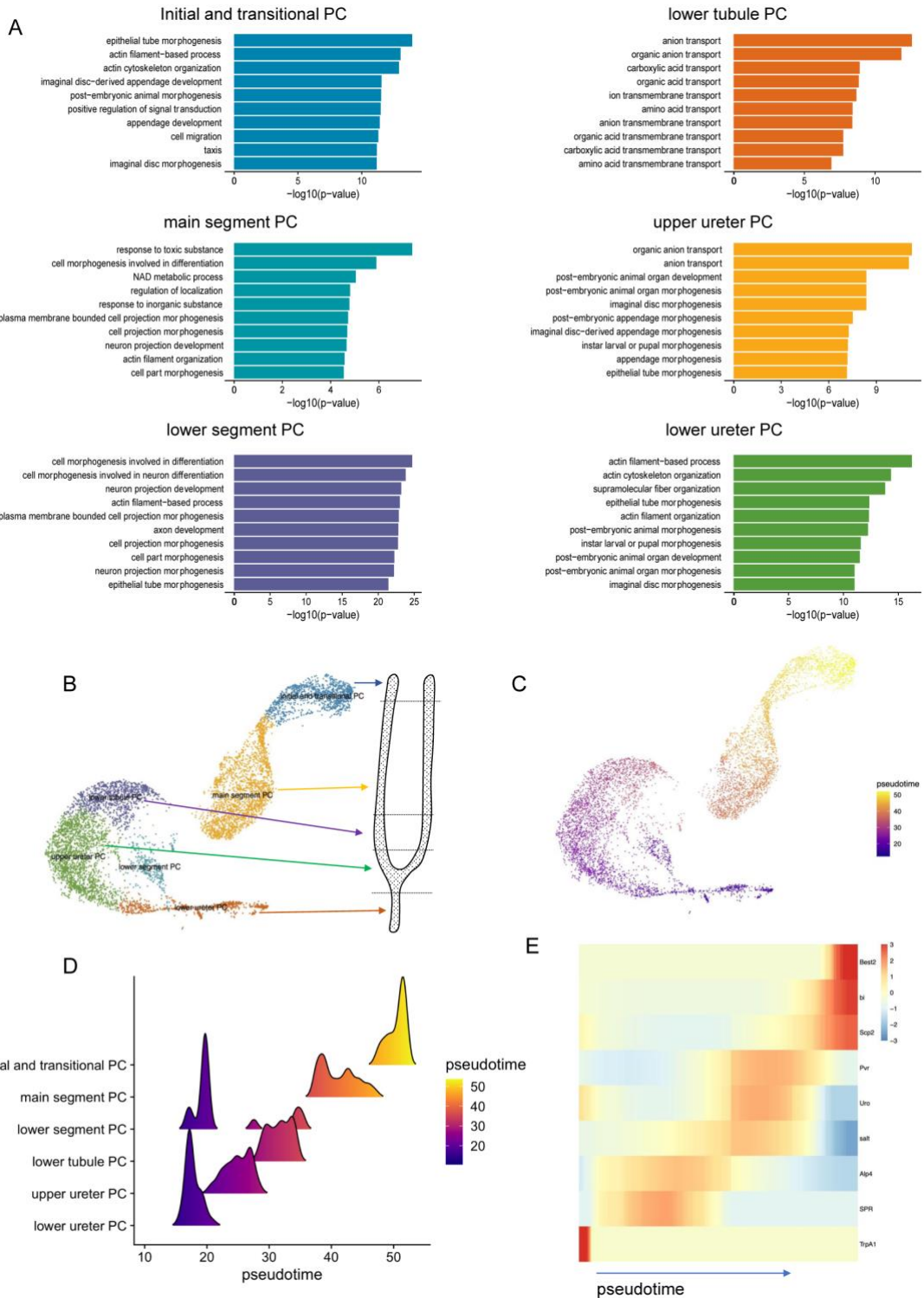
Figure S7. Pseudotime and GO analysis of the six PC sub-clusters. (A) UMAP of PCs showing a geographical map of the tubule. (B) Cell pseudotime was inferred using Monocle3. Purple at the beginning becomes yellow over pseudotime. (C) Sub-cell type populations for each inferred cellular trajectory. The x-axis indicates the inferred pseudotime and the y-axis indicates the height of density estimated and visualized by the RidgePlot function of Seurat R package. (D) Heatmap showing gene expression patterns during differentiation along pseudotime. (E) GO analysis of each PC cluster.

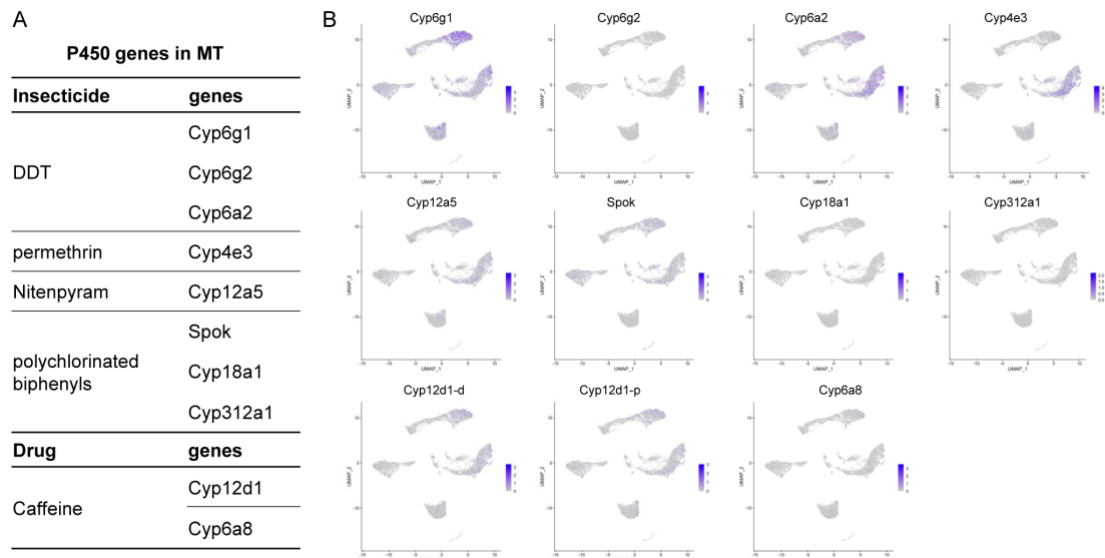Initial and transitional PCs include epithelial tube morphogenesis, actin-filament based process, and actin cytoskeleton organization. Main segment PCs include response to toxic substance, cell morphogenesis involved in differentiation, and NAD metabolic process. Lower segment PCs include cell morphogenesis involved in differentiation, cell morphogenesis involved in neuronal differentiation, and actin-filament based process GO terms. Lower tubule PCs include terms such as anion transport, organic anion transport, and carboxylic acid transport. Upper ureter PC terms include organic anion transport, and anion transport. Lower ureter PC GO terms include actin filament-based process, actin cytoskeleton organization, and supramolecular fiber organization. All the top 10 terms in lower tubule PCs were related to transport. The top 10 terms are displayed.

1. initial and transitional PC
2. main segment stellate cell
3. bar-shaped stellate cell
4. main segment PC
5. lower tubule PC
6. upper ureter PC
7. lower segment PC
8. lower ureter PC
9. pericardial nephrocytes
10. garland nephrocytes
11. renal stem cell

Figure S8. Detailed information for Fig. 4. SCENIC results for the fly kidney. Heatmap, gene expression levels in each cluster. Low regulon activity is shown in blue and high regulon activity is shown in red.

44

Figure S9. Gene set activity of 86 KEGG pathways in the UMAP fly kidney. For each pathway the color represents the gene set activity level. Red intensity reflects high gene set activity.

A

P450 genes in MT

Insecticide	genes
DDT	Cyp6g1
	Cyp6g2
	Cyp6a2
permethrin	Cyp4e3
Nitenpyram	Cyp12a5
polychlorinated biphenyls	Spok
	Cyp18a1
	Cyp312a1
Drug	**genes**
Caffeine	Cyp12d1
	Cyp6a8

Figure S10. Gene expression of insecticide- and drug-related P450 genes in the fly kidney. (A) Insecticide- and drug-related P450 genes in the fly (based on Seong et al., 2020; Bergé et al., 1998; Terhzaz et al., 2015; Harrop et al. 2018; Idda et al., 2020; Najarro et al., 2015). (B) Expression levels of each P450 gene visualized by UMAP plots.
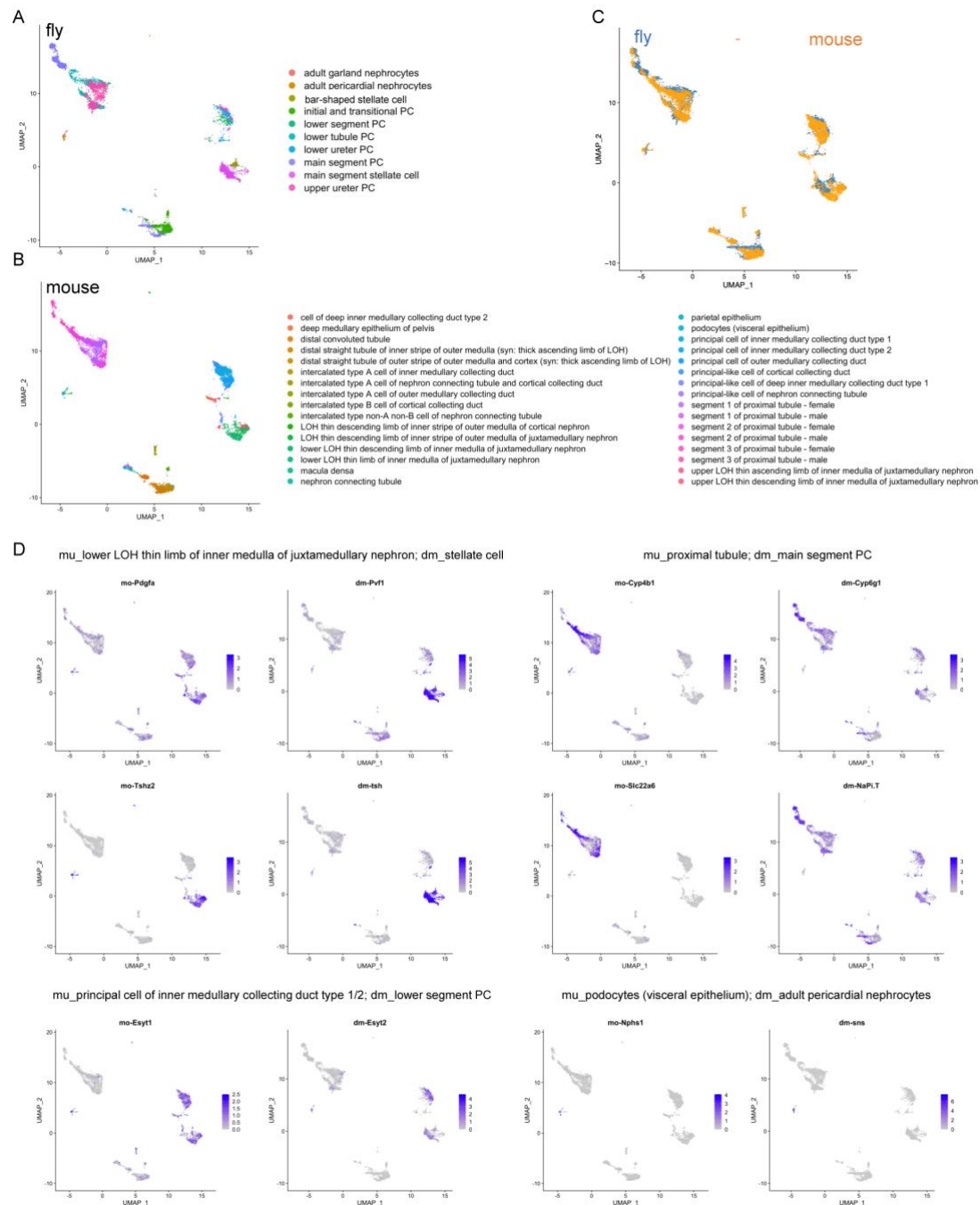
Figure S11. Cross-species analysis of fly kidney and mouse using SAMap. (A and B) Low dimensional representations of the cell atlases through homologous gene pairs in the mouse and fly using SAMap. (C) UMAP projection of the combined mouse (yellow) and fly (blue) manifolds. (D) Expression of orthologous gene pairs on the UMAP projection. Expressing cells are in blue and cells with no expression are shown in gray.
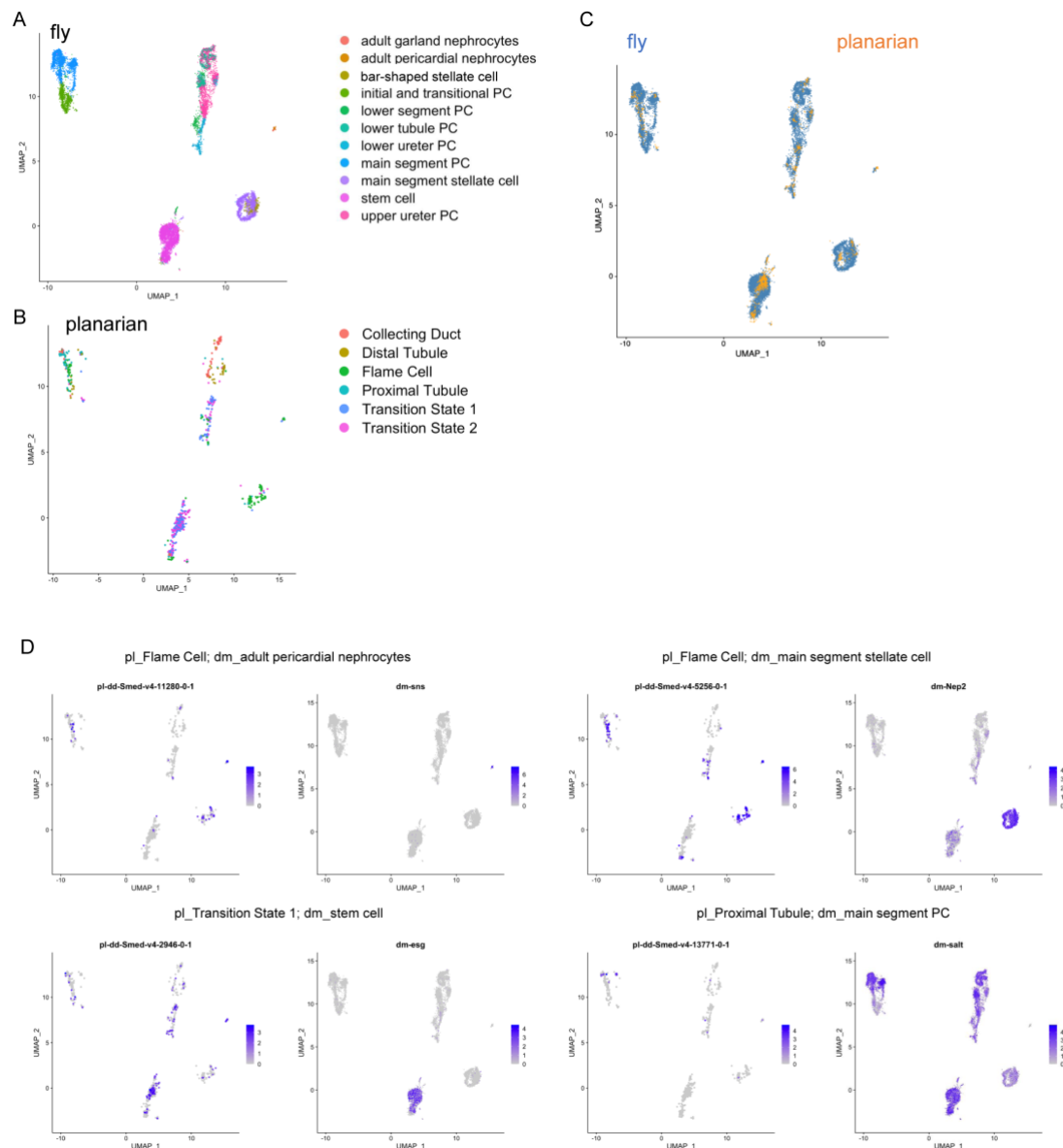
Figure S12. Cross-species analysis of fly kidney and planaria protonephridia using SAMap. (A and B) Low dimensional representations of the cell atlases through homologous gene pairs in the planaria and fly using SAMap. (C) UMAP projection of the combined planaria (yellow) and fly (blue) manifolds. (E) Expression of orthologous gene pairs on the UMAP projection. Expressing cells are in blue and cells with no expression are shown in gray.
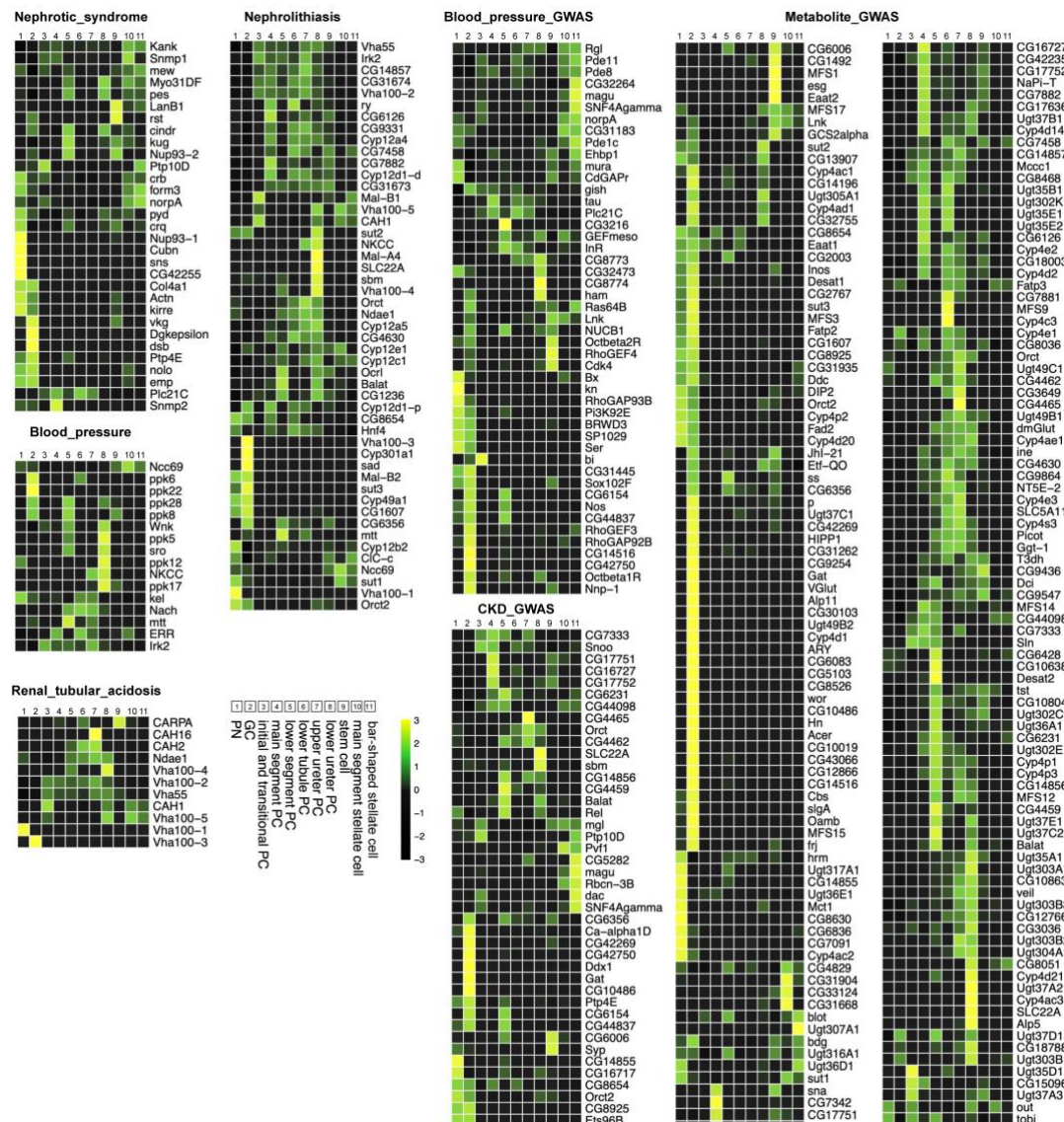
Figure S13. Expression of fly orthologs of human kidney disease-associated genes in specific fly kidney cell types. Average expression in single cell clusters of fly orthologs of human monogenic disease genes and complex-trait genes identified from genome-wide association studies (GWAS) (Park et al., 2018). Mean expression values of the genes were calculated for each cluster. The color scheme is based on z-score distribution (-3 < z-scores < 3). Each row in the heat map represents one gene and each column a single cell type.
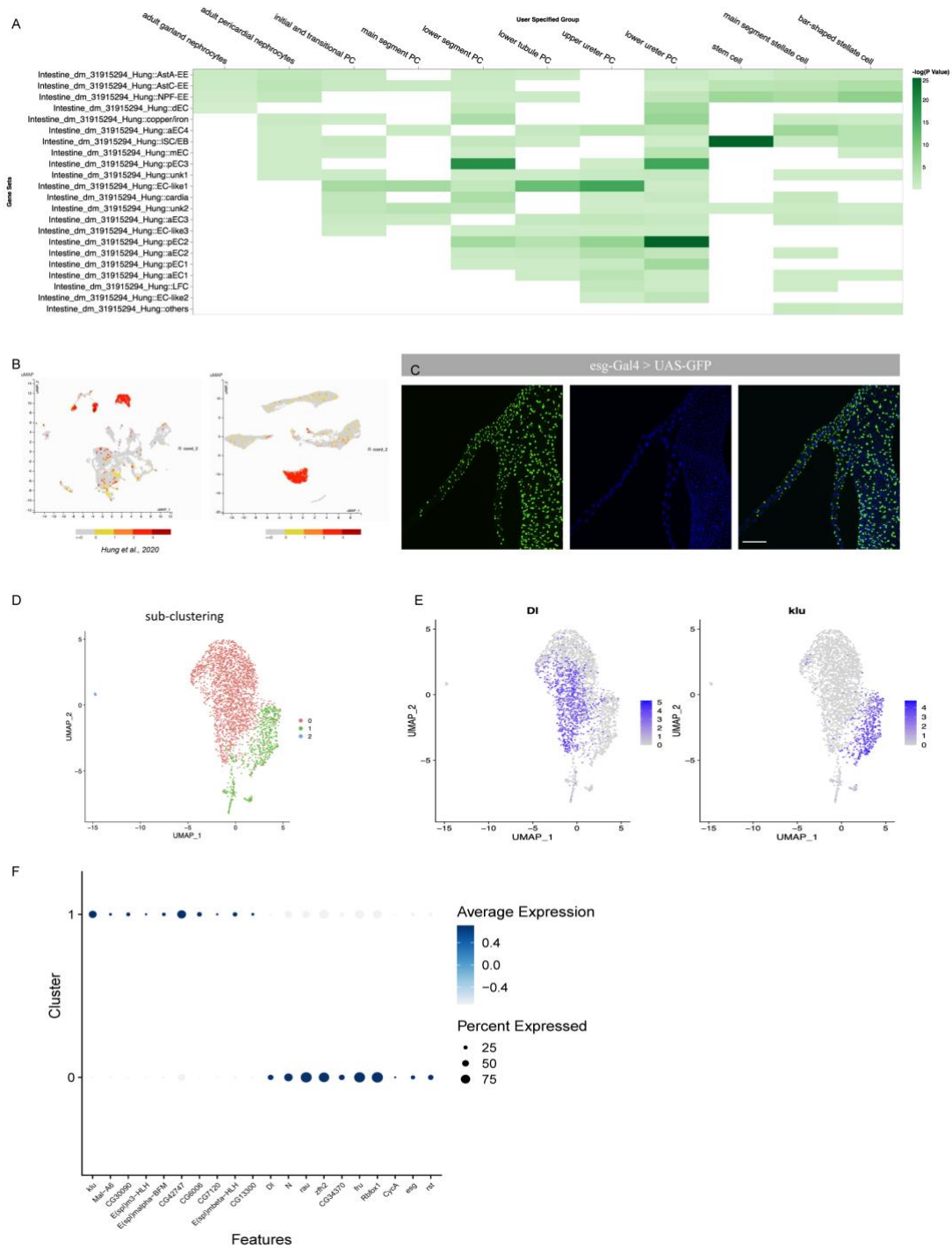
Figure S14. Comparison of renal stem cell and intestinal stem cell clusters. (A) snRNA-seq midgut clusters are from Hung et al. 2020. The X axis shows the 11 integrated clusters from Fig. 1B. The Y axis shows the midgut clusters from Hung et al. (2020). Colors represent gene expression similarities. Note that RSCs are highly similar to ISCs. In addition, lower ureter PCs share high similarity with pEC2 (posterior enterocytes). (B) *escargot (esg)* expression in the gut and Malpighian tubule UMAPs. (C) *esg* expression in the Malpighian tubules visualized using esg-Gal4 driving UAS-GFP expression. Scale bars = 100 μm. (D) UMAP distribution of different sub-clusters

1   of RSCs. (E) *Delta (Dl)* and *klumpfuss (klu)* expression in the RSC UMAP subclusters.

2   (F) Dot plot showing the expression levels and percentage of cells expressing the

3   various markers in the *Dl* and *klu* sub-clusters.

4

5

6
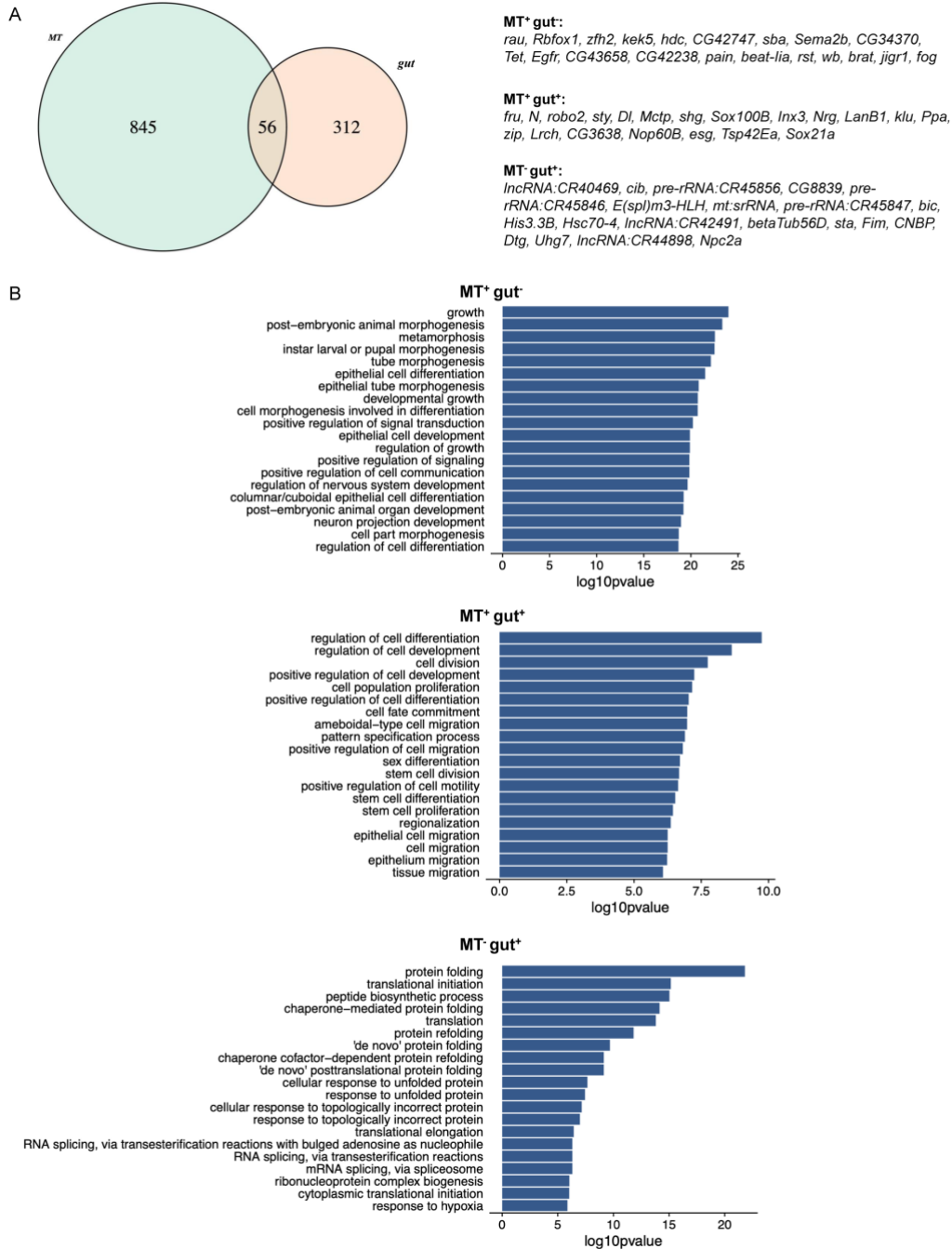
Figure S15. Comparison of renal stem cell and intestinal stem cell clusters. (A) Venn diagram of the overlap between Malpighian tubule (MT) and gut stem cell top marker genes (log2FC > 0.25 and adjust p-value < 0.05). On the right are identities of the top genes in the three regions of the diagram. (B) Gene Ontology (GO) analysis of MT⁺gut⁻, MT⁺gut⁺ and MT⁻gut⁺. The top 20 terms are displayed.
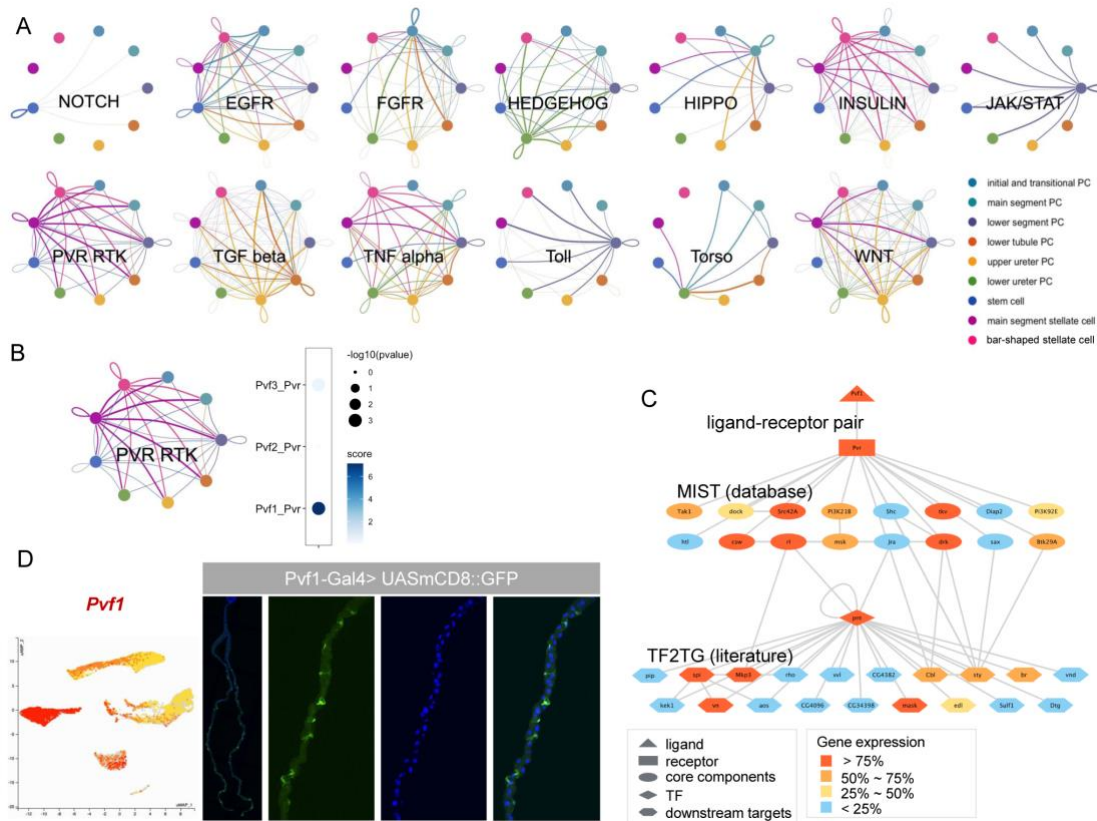
12

Figure S16. Cell–cell communication analysis in the adult fly kidney. (A) Network of 13 signaling pathways in fly kidney cell clusters. Each of the 11 cell clusters is displayed in a different color. The predicted interaction between two clusters is indicated by a color curve. The thickness of the curve indicates the strength of the interaction. The full list of predicted ligand/receptor pair genes can be found in Supplementary Table 12. (B) Ligand-receptor interaction between *Pvf1* and its receptor *Pvr* in main segment SCs and main segment PCs. The panel on the right shows a dot plot of the interaction score and specificity of ligand-receptor pairs between main segment principal cell cluster and main segment stellate cell cluster. (C) *Pvr-pvf1* interaction network based on MIST and TF2TG. (D) *pvf1* expression in SCs visualized using pvf1-Gal4 driving UAS-mCD8::GFP expression.

52

Supplementary references

Bergé, J.B., Feyereisen, R., and Amichot, M. (1998). Cytochrome P450 monooxygenases and insecticide resistance in insects. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *353*, 1701–1705.

Couturier, L., Mazouni, K., Corson, F., and Schweisguth, F. (2019). Regulation of Notch output dynamics via specific E(spl)-HLH factors during bristle patterning in *Drosophila*. Nat. Commun. *10*, 3486.

Harrop, T.W., Denecke, S., Yang, Y.T., Chan, J., Daborn, P.J., Perry, T., and Batterham, P. (2018). Evidence for activation of nitenpyram by a mitochondrial cytochrome P450 in *Drosophila melanogaster*. Pest Manag. Sci. *74*, 1616–1622.

Hung, R.J., Hu, Y., Kirchner, R., Liu, Y., Xu, C., Comjean, A., Tattikota, S.G., Li, F., Song, W., Ho Sui, S., et al. (2020). A cell atlas of the adult *Drosophila* midgut. Proc. Natl. Acad. Sci. USA *117*, 1514–1523.

Idda, T., Bonas, C., Hoffmann, J., Bertram, J., Quinete, N., Schettgen, T., Fietkau, K., Esser, A., Stope, M.B., Leijs, M.M., et al. (2020). Metabolic activation and toxicological evaluation of polychlorinated biphenyls in *Drosophila melanogaster*. Sci. Rep. *10*, 21587.

Li, H., Janssens, J., Waegeneer, M., Kolluru, S., Davie, K., Gardeux, V., Saelens, W., David, F., Brbić, M., Leskovec, J., et al. (2021) Fly Cell Atlas: a single-cell transcriptomic atlas of the adult fruit fly. bioRxiv. https://doi.org/10.1101/2021.07.04.451050.

Li, Z, Liu S, and Cai Y. (2014). Differential Notch activity is required for homeostasis of malpighian tubules in adult *Drosophila*. J. Genet. Genomics *41*, 649–652.

Li, Z, Liu S, and Cai Y. (2015). EGFR/MAPK signaling regulates the proliferation of Drosophila renal and nephric stem cells. J. Genet. Genomics *42*, 9–20.

Liu, Y.F., Hu, Y.H., Li, J.S.S., Rodiger, J., Comjean, A., Attrill, H., Antonazzo, G., Brown, N.H., and Perrimon, N. (2020). FlyPhoneDB: An integrated web-based resource for cell-cell communication prediction in *Drosophila*. bioRxiv. doi: https://doi.org/10.1101/2021.06.14.448430

Loza-Coll, M.A., Southall, T.D., Sandall, S.L., Brand, A.H., and Jones, D.L. (2014). Regulation of *Drosophila* intestinal stem cell maintenance and differentiation by the transcription factor Escargot. EMBO J. *33*, 2983–2996.

Lu, Y., and Li, Z. (2015). Notch signaling downstream target E(spl)mbeta is dispensable for adult midgut homeostasis in *Drosophila*. Gene *560*, 89–95.

Micchelli, C.A., and Perrimon, N. (2006). Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. Nature *439*, 475–479.

Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., King, E.G., Long, A.D., and Macdonald, S.J. (2015). Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in *Drosophila*. PLoS Genet. *11*, e1005663.

Ohlstein, B., and Spradling, A. (2006). The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. Nature *439*, 470–474.

Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., and Suszták, K. (2018). Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. Science *360*, 758–763.

Seong, K.M., Coates, B.S., and Pittendrigh, B.R. (2020). Post-transcriptional modulation of cytochrome P450s, *Cyp6g1* and *Cyp6g2*, by *miR-310s* cluster is associated with DDT-resistant *Drosophila melanogaster* strain *91-R*. Sci. Rep. *10*, 14394.

Singh, S.R., Liu, W., and Hou, S.X. (2007). The adult *Drosophila* malpighian tubules are maintained by multipotent stem cells. Cell Stem Cell *1*, 191–203.

Takashima, S., Paul, M., Aghajanian, P., Younossi-Hartenstein, A., and Hartenstein, V. (2013). Migration

of *Drosophila* intestinal stem cells across organ boundaries. Development *140*, 1903–1911.

Terhzaz, S., Cabrero, P., Brinzer, R.A., Halberg, K.A., Dow, J.A.T., and Davies, S.A. (2015). A novel role of Drosophila cytochrome P450-4e3 in permethrin insecticide tolerance. Insect Biochem. Mol. Biol. *67*, 38–46.

Wang, C., and Spradling, A.C. (2020) An abundant quiescent stem cell population in *Drosophila* Malpighian tubules protects principal cells from kidney stones. Elife *9*, e54096.

Xu, K., Liu, X., Wang, Y., Wong, C., and Song, Y. (2018). Temporospatial induction of homeodomain gene cut dictates natural lineage reprogramming. Elife *7*, e33934.

Zhu, J.Y., Fu, Y., Nettleton, M., Richman, A., and Han, Z. (2017). High throughput in vivo functional validation of candidate congenital heart disease genes in *Drosophila*. Elife *6*, e22617.