

Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development

Mike Adam, Andrew S. Potter and S. Steven Potter*

ABSTRACT

Single-cell RNA-seq is a powerful technique. Nevertheless, there are important limitations, including the technical challenges of breaking down an organ or tissue into a single-cell suspension. Invariably, this has required enzymatic incubation at 37°C, which can be expected to result in artifactual changes in gene expression patterns. Here, we describe a dissociation method that uses a protease with high activity in the cold, purified from a psychrophilic microorganism. The entire procedure is carried out at 6°C or colder, at which temperature mammalian transcriptional machinery is largely inactive, thereby effectively ‘freezing in’ the *in vivo* gene expression patterns. To test this method, we carried out RNA-seq on 20,424 single cells from postnatal day 1 mouse kidneys, comparing the results of the psychrophilic protease method with procedures using 37°C incubation. We show that the cold protease method provides a great reduction in gene expression artifacts. In addition, the results produce a single-cell resolution gene expression atlas of the newborn mouse kidney, an interesting time in development when mature nephrons are present yet nephrogenesis remains extremely active.

KEY WORDS: Single cell, Kidney development, Cell dissociation, RNA-seq, Artifacts

INTRODUCTION

Single-cell RNA sequencing (RNA-seq) is revolutionizing many areas of biology. There often exist heterogeneities at the cellular level that can only be dissected out with high-resolution approaches that define the characters of individual cells. For example, tumors consist of complex mixes of cells, including vascular cells, fibroblasts, invading immune cells, rapidly dividing cancer cells and quiescent cancer stem cells. Single-cell RNA-seq can be used to fully define the many types and subtypes of cells present (Tirosh et al., 2016). Similarly, during development there are often populations of cells with identical histology, where flanking cells might follow very distinct differentiation lineage pathways. Single-cell resolution studies are required to understand at the molecular level the nature of these early lineage decision processes (Chiang and Melton, 2003; Olsson et al., 2016; Brunskill et al., 2014). Studies that examine ensemble averages, or population pools of cells, can often provide only limited insight.

Recent technological advances allow the RNA-seq analysis of large numbers of single cells (Macosko et al., 2015; Klein et al., 2015). Nevertheless, important limitations remain. First, there is the

biological noise that results from the pulsatile, bursting nature of gene expression (Ross et al., 1994; Ozbudak et al., 2002; Raj et al., 2008). This means that the gene expression pattern of a single cell is in a constant state of flux. This expression state chatter is a consequence of the fundamental nature of gene expression and is unavoidable. A second limitation resides in the methodological measure of transcript levels when starting with only the 5-50 pg of total RNA typically found in a single cell. It is estimated that most current technologies only detect 5-50% of the mRNA molecules actually present (Svensson et al., 2017). These inefficiencies in reverse transcription/amplification add a considerable layer of technical noise to single-cell studies. Continual refinements of technology can be expected to improve future RNA detection rates (Sheng et al., 2017).

A third limitation is the element of artifactual gene expression introduced by the process of single-cell dissociation. The breakdown of an organ or tissue into a single-cell suspension often requires a lengthy and complex protocol. Almost invariably these procedures include 37°C incubation with a protease to help break attachments between cells. Among the enzymes and enzyme cocktails typically used are trypsin, TrypLE, pronase, collagenase, liberase and dispase, all of which are active at 37°C. At this temperature the enzymes within mammalian cells are also maximally active, including the transcriptional machinery. Because the cells are surrounded by an extremely foreign environment during this enzymatic digestion period one can predict that artifactual changes in gene expression will result.

We reasoned that this situation is analogous to that encountered during the advent of PCR. At first, it was necessary to add a fresh aliquot of *Escherichia coli* DNA polymerase after each denaturation cycle, as the heat inactivated the enzyme. The switch to DNA polymerases from thermophilic organisms, with enzymes adapted for high activity and stability at elevated temperatures, allowed the elimination of this cumbersome step.

Similarly, single-cell RNA-seq could incorporate the use of cold-adapted proteases that show high activity at low temperatures. This would allow the entire single-cell dissociation process to be carried out ‘on ice’, or at greatly reduced temperatures, at which mammalian enzymes show little activity, thereby minimizing gene expression artifacts. Just as there are thermophilic organisms that live in extremely hot environments, there are psychrophilic, or cryophilic, organisms that are adapted to thrive in the cold. These can be found on mountaintops, on glaciers, in the deep ocean and at extreme northern and southern latitudes, where cold temperatures prevail. These organisms can produce proteases that are adapted for high activity in the cold.

In this article, we develop and test a cold protease method of single-cell dissociation, showing that it can dramatically reduce the gene expression artifacts resulting from 37°C protease incubations. We apply the cold protease single-cell RNA-seq procedure to the newborn postnatal day 1 (P1) mouse kidney. This is a very

Cincinnati Children’s Hospital Medical Center, Division of Developmental Biology, Cincinnati, OH 45229, USA.

*Author for correspondence (steve.potter@cchmc.org)

 S.S.P., 0000-0003-0726-0350

Received 2 March 2017; Accepted 18 August 2017

interesting stage of mouse kidney development, when mature nephrons are present, as required to support postpartum life, and yet nephrogenesis is still quite active. All stages of nephron formation are represented, from cap mesenchyme progenitor cells to fully differentiated nephrons. The results provide a single-cell resolution gene expression atlas of the P1 developing kidney.

RESULTS

Psychrophilic protease dramatically reduces artifacts

To test the concept, we developed a method for single-cell dissociation using a cold active protease from *Bacillus licheniformis*, which is a soil bacterium that can grow in a wide range of temperatures and has been isolated from Himalayan glaciers (Joshi and Satyanarayana, 2013). The protocol, in essence, includes mincing the organ with a razor blade, digestion with *B. Licheniformis* protease for 15 min at 6°C, augmented with Miltenyi gentleMACS and trituration physical disruption, followed by passage through a 30 µm filter. The entire procedure is carried out at 6°C or colder.

To determine the efficacy of this protocol in reducing artifacts, we compared the single-cell RNA-seq profiles generated with cold active protease (CAP) with those resulting from more traditional methods requiring 37°C incubation. We used Drop-seq (Macosko et al., 2015) to examine P1 mouse kidneys. The Drop-seq method creates thousands of microdroplets that include a cell and a bead coated with uniquely barcoded oligonucleotides. The barcodes are incorporated into the cDNA, allowing the eventual DNA sequence reads to be assigned to individual cells.

To better define the potential time-dependent artifact-generating effects of 37°C incubations we also carried out Drop-seq after 15 min, 30 min and 60 min at 37°C, using a more conventional enzymatic dissociation method.

The single-cell RNA-seq datasets generated with the cold CAP protocol and the multiple 37°C incubation times were analyzed with Seurat (Macosko et al., 2015). The data included 4853 cells for CAP, 5261 cells for 15 min 37°C, 5870 cells for 30 min 37°C, and 4440 cells for 60 min 37°C, for a total of 20,424 cells. We first used unsupervised clustering to generate t-distributed stochastic neighbor embedding (t-SNE) plots that separated the cells into distinct groups. The CAP method produced the same cell groupings identified with 37°C incubation protocols (Fig. 1). The podocytes are octopus-shaped cells that intimately embrace the capillaries of the glomerulus and are particularly difficult to dissociate. We observed excellent representation of single-cell podocytes in the cold protease t-SNE plots, showing the power of the CAP method in separating cells.

We next looked for possible artifactual gene expression changes associated with 37°C incubations. Although this topic has been much debated in the single-cell research community, to our knowledge it has not been previously carefully addressed. The immediate-early response genes are prime candidates for changing expression during single-cell dissociation at 37°C. These genes are rapidly induced, even in the absence of protein synthesis, by a variety of conditions, including mitogens, growth factors and stress (O'Donnell et al., 2012). Members of the Fos and Jun gene families are prototypical immediate-early genes. Single-cell RNA-seq gene expression levels were quantified as previously described (Macosko et al., 2015). It is not possible to generate reads per kilobase of transcript per million mapped reads (RPKM) or fragments per kilobase of transcript per million mapped reads (FPKM) values with single-cell data. Comparison of gene expression levels of combined cell populations, including all cell types, showed that the CAP

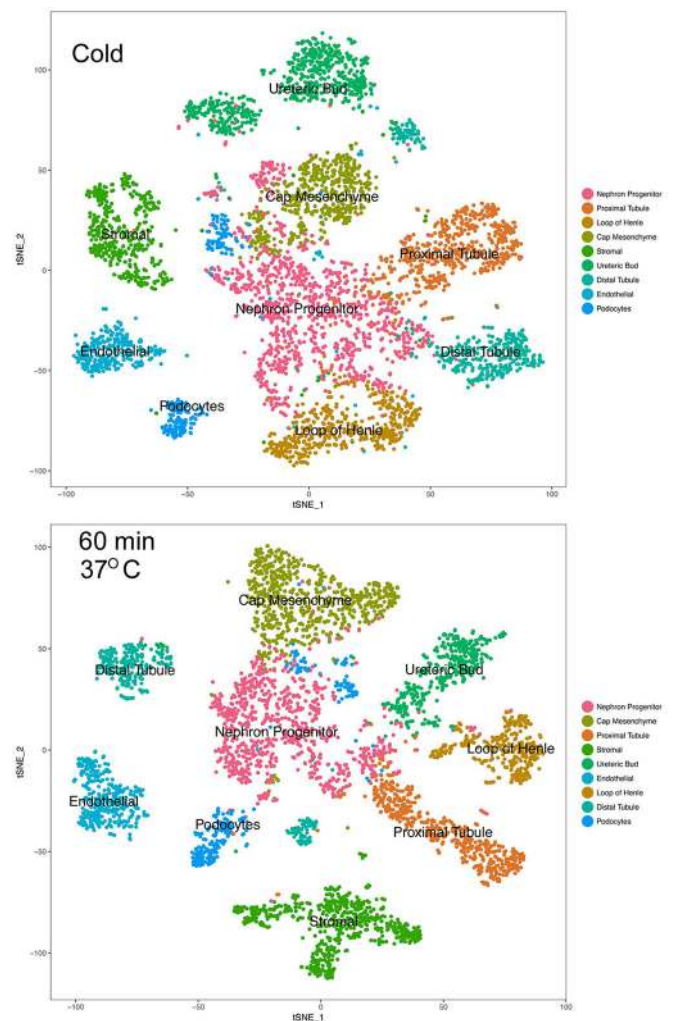


Fig. 1. t-SNE plots of cells from P1 mouse kidneys. Top panel shows t-SNE analysis of cells that were dissociated with the cold active protease method, allowing the entire procedure to be carried out at 6°C or colder. Single-cell RNA-seq data for 5261 cells were generated with Drop-seq and unsupervised clustering carried out with the Seurat program. Even cells that are difficult to separate, such as podocytes, were effectively dissociated. The labeled nephron progenitor cells are rather diverse, as they include progenitor renal aggregate, renal vesicle, comma- and S-shaped bodies and multiple differentiating lineages. Bottom panel, for comparison, shows a t-SNE plot of 4440 cells dissociated using a 60 min at 37°C dissociation protocol. Similar groups of cells are identified using both procedures, demonstrating the power of the cold method to dissociate all cell types. Of interest, several cell types are showing separation into subtypes. For example, in the top panel the ureteric bud cells localized to the tips are distinct from the other ureteric bud cells in their clustering. Similarly, the color-coded distal tubule cells separate into two clusters, and other cell types, such as the stromal and proximal tubule cells, show interesting subdivisions, even in this first-stage analysis.

method gave *Fos* expression of only 96, compared with the 37°C incubation method levels of 37,540 (15 min), 71,082 (30 min) and 34,786 (60 min). *Fosb* showed similar differences, with gene expression levels of only 5 for CAP, versus 8032 (15 min), 17,344 (30 min) and 15,944 (60 min) for 37°C incubations. The CAP method also gave the lowest *Jun* expression levels, with 3734 (CAP), 30,518 (15 min 37°C), 50,730 (30 min 37°C) and 38,517 (60 min 37°C). *Junb*, another member of the Jun gene family, again showed greatly reduced expression levels using the CAP method, with 483 (CAP), 12,819 (15 min 37°C), 23,848 (30 min 37°C) and

15,536 (60 min 37°C). In summary, these immediate-early response genes showed remarkably reduced expression using the low temperature CAP dissociation procedure compared with methods requiring incubation at 37°C.

We then examined the global gene expression changes associated with 37°C incubations. We first screened for differential gene expression in combined populations, including all cell types ($P \leq 0.001$, $FC \geq 2$). Compared with the CAP procedure we observed over 300 genes (339) changing in expression as a result of 15 min incubation at 37°C and over 500 (539) genes altered in expression following 60 min incubation at 37°C (Tables S1 and S2). These results suggest that there are indeed significant global gene expression artifacts associated with the 37°C incubations normally used for cell dissociation.

We then compared the gene expression profiles of specific cell types using the different dissociation methods. In this manner, it was possible to determine whether distinct types of cells show varied responses to 37°C incubations. How many of the observed gene expression changes were common to multiple cell types and how many were unique to a given class of cell? For this analysis, we focused on five diverse categories of cells: endothelial, proximal tubule, loop of Henle, podocytes and cap mesenchyme progenitor cells. For each cell type we compared the gene expression pattern of CAP to 15, 30 and 60 min 37°C incubations ($P \leq 0.001$, $FC \geq 2$). The resulting gene lists were used to define overlapping and cell type-specific genes with altered expression. For example, in comparing CAP with the 60 min 37°C condition we observed a common core of 38 genes that showed differential gene expression in all five types of cells (Fig. 2). A total of 742 genes showed altered expression in at least two cell types, providing a measure of cross validation. In addition there were many genes that changed in expression in only one type of cell (Fig. 2, Tables S3-S5).

As might be expected, the longer incubations at 37°C resulted in greater gene expression differences. The time course of changing expression was plotted for the 38 genes with greater than twofold altered expression in all five cell types following 60 min incubation at 37°C (Fig. 3). Basal gene expression levels were often very low, with dramatic upregulation following 37°C incubations. *Fosb*, for example, showed a 3000-fold increased expression resulting from 60 min incubation at 37°C. When comparing combined cell types 32 genes showed over tenfold change in expression (Table S2).

Gene ontology (GO) analysis of the genes with changed expression resulting from 60 min at 37°C (Table S2) was carried out with TopPgene (Chen et al., 2009). The strongest molecular function enrichment was for transcription factor activity ($P = 4.31 \times 10^{-23}$). Ninety transcription factor genes were altered in expression, indicating the onset of a massive break in the gene expression program. One of the greatest biological process enrichments was for regulation of cell death ($P = 7.34 \times 10^{-14}$). The programmed cell death gene expression profile could be related to the anoikis response that many cells show following physical separation.

There is possible concern that although the use of cold active proteases reduces the gene expression changes associated with 37°C dissociation protocols, it nevertheless introduces cold-shock stress artifacts associated with incubation at 6°C. Indeed, there is a well-documented cold-shock response following a period of exposure to mild hypothermia, typically 25-33°C, with several genes, including *Cirp*, *Rbm3*, *P53 (Trp53)* and *Il8 (Cxcl15)*, shown to be upregulated in various cell types (Sonna et al., 2002). Nevertheless, at much lower temperatures the mammalian transcriptional machinery is largely inactive and few if any RNA transcript level changes result. For example, no changes in gene transcript levels were found following incubation at temperatures below 5°C (Fujita, 1999). Although the cells surely experience a cold shock at these low temperatures, they are effectively unable to transcribe new RNA, or turnover old, and therefore their gene expression profiles remain unchanged.

Single-cell gene expression atlas of newborn kidney development

The CAP-generated single-cell RNA-seq data in this article provide a high-resolution atlas of the gene expression programs driving kidney development. This is a useful complement to the GUDMAP resource (<http://gudmap.org/>) (Harding et al., 2011), which includes many thousands of *in situ* hybridizations, many hundreds of microarrays for laser capture micro-dissected (LCM) kidney compartments, and considerable RNA-seq data for multiple fluorescence-activated cell sorting-isolated developing kidney cell types.

Molecular atlas projects, such as the Allen Brain Atlas (Lein et al., 2007), GUDMAP (Harding et al., 2011), FACEBASE

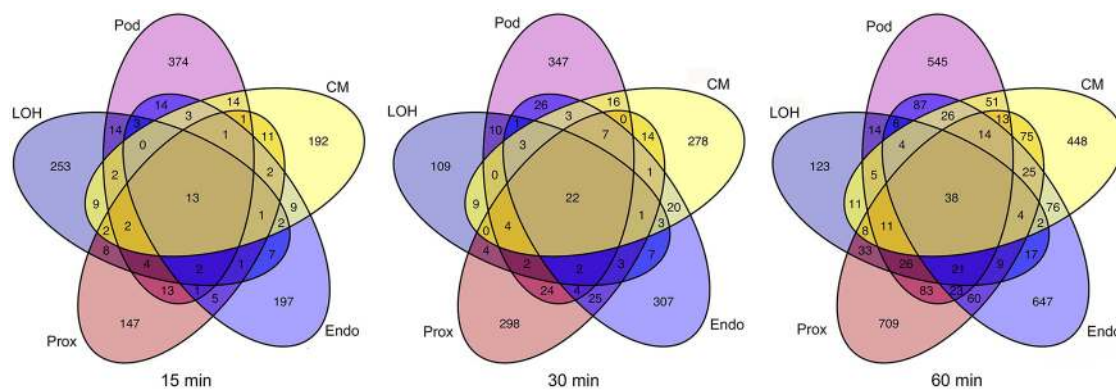


Fig. 2. Venn diagrams showing overlap of gene expression differences associated with incubation at 37°C for varying time periods. This analysis focuses on five cell types, endothelial (Endo), podocytes (Pod), proximal tubule (Prox), loop of Henle (LOH) and cap mesenchyme (CM) progenitors. For each cell type, the gene expression patterns of the cells incubated at 37°C were compared with the gene expression patterns of the same cell types generated using the cold active protease procedure ($P \leq 0.001$, $FC \geq 2$). Results are shown for 15, 30 and 60 min 37°C incubations. Cell type-specific as well as shared gene expression changes are observed. The shared genes, showing change in more than one cell type, are thereby cross-validated. There are increasing numbers of these genes as a function of time of incubation at 37°C.

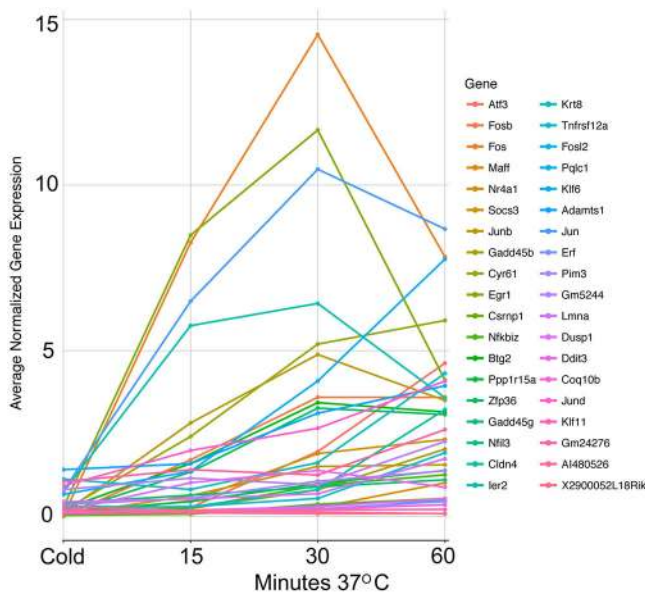


Fig. 3. Changing expression of genes as a function of dissociation carried out at increasing times at 37°C. Gene expression after dissociation under the four different conditions is shown for the 38 genes with changed expression in all five cell types (endothelial, podocytes, proximal tubule, loop of Henle and cap mesenchyme) when incubated at 37°C for 60 min ($P \leq 0.001$, $FC \geq 2$). Cold represents expression levels with the cold active protease method. These basal expression levels, defined by the cold method, are often very low. Some genes, such as *Fosl*, show over 1000-fold increase in expression following 37°C incubation. The average normalized gene expression values (y -axis) were determined by first dividing the read count for each gene by the total of number reads in that cell, and multiplying by 10,000. These expression values were then averaged by adding for all cells, and dividing by the number of cells.

(Hochheiser et al., 2011) and LungMap (Du et al., 2017), provide useful resources for the research community. They define the gene expression blueprints of organogenesis. They can aid in the screening of candidate single nucleotide polymorphisms in human exome sequencing projects. They can identify novel compartment-specific markers. They provide a wild-type baseline for subsequent mutant analysis. Further, when they include RNA-seq data, they can define enhancer transcription, noncoding RNAs, and all transcription factors, growth factors and receptors expressed during the differentiation of specific cell types.

The unsupervised clustering of the P1 kidney single-cell RNA-seq data generates the cell groupings shown in Fig. 1. Although spatial information is lost following single-cell dissociation, it is possible to assign cell types to the resulting clusters based on known markers. A heatmap showing the differential expression of genes in these groups is shown in Fig. S1 and the genes are listed in Table S6.

A particular strength of the single-cell approach is the ability to carry out the analysis in reiterative fashion and characterize subtypes of cells. The initial analysis separated out the most distinct types of cells, and it is then possible to focus on each one of these groups, and to identify subgroups. For example, consider the proximal tubules. The cells defined as proximal tubules (Fig. 1), by virtue of expression of *Hnf4a* and other proximal tubule markers, were singled out and analyzed further. They were first divided into two subgroups. One of these represented early proximal tubule progenitor cells, expressing midkine, which is associated with the early proximal tubule (Sato and Sato, 2014), as well as *Osr2*, *Lhx1* and *Jag1*, which are markers of early nephron developmental stage cells, preceding the proximal tubule (Fig. S2). This is consistent

with our previous single-cell study of early nephrogenesis, showing that progenitors can show robust yet stochastic mRNA expression of multiple markers of later potential differentiated states (Brunskill et al., 2014). The second group of proximal tubule cells, representing a more mature developmental stage, was then further divided into two subgroups. One of these expressed *Gpx3*, *Slc3a2*, *Rbp1*, *Nrep*, *Acaa2* and *Nox4*, all of which are known markers of the S1 segment of the proximal tubule (Lee et al., 2015), whereas the other group expressed *Napsa*, *Ttc36*, *Slc3a1*, *Slc23a1*, *Slc27a2* and *Kap*, which are known markers of the S2,3 segments of the proximal tubule (Lee et al., 2015; Thiagarajan et al., 2011) (Fig. 4, Table S7). The S3 segment is small, and shows significant overlapping gene expression with the S2 segment, making it difficult to distinguish with the number of cells in this study (4853 CAP cells). To our knowledge, these results provide the first single-cell analysis of the development of the proximal tubule, from progenitor, including intermediate stages, to differentiated segments. It is interesting to note that the early proximal tubule cells express *Aldh1a2*, which synthesizes retinoic acid (RA) from retinol. In zebrafish, RA is a key driver of proximal tubule development (Wingert and Davidson, 2011; Cheng and Wingert, 2015). Therefore, the observed RA synthesis in the murine proximal tubule progenitors could represent autocrine signaling.

The collecting duct system, which connects the nephrons to the renal pelvis, is at a very dynamic stage in the newborn mouse. Additional analysis of the collecting duct cells identified in Fig. 1 further divided these cells into three subgroups (Fig. 5, Table S8). One group corresponds to the collecting duct tip cells, which at P1 continue to drive nephrogenesis in the flanking cap mesenchyme. These cells express tip-associated genes, including *Wnt11*, *Ret*, *Krt23*, *Sox9*, *Etv4*, *Etv5* and *Gfra1*. Most of the remaining cells express markers of the principal cells that help control sodium and potassium balance as well as carry out water resorption. Principal cell marker genes include the aquaporins *Aqp2*, *Aqp3*, *Aqp4*, and the genes encoding the components of the epithelial sodium channel ENaC, *Scnn1a*, *Scnn1b* and *Scnn1g*. The single-cell RNA-seq data provides a global gene expression profile of these early differentiating principal cells.

The third group of collecting duct cells, however, is perhaps the most interesting. These cells express the principal cell marker genes *Aqp2,3,4* and *Scnn1a,b,g*, but they also express *Foxl1* and *Slc26a4*, indicating an intercalated cell character. These cells appear to be in transition, with dual principal/intercalated cell features. The intercalated cells can be further divided into two types: α and β . The β cells give rise to the α cells, which can be viewed as a more differentiated intercalated cell (Al-Awqati, 2013). At P1, we see cells expressing *Slc26a4* (also known as pendrin), a marker of the β cells, and fewer cells expressing *Slc4a1* (also known as AE1), a marker of α intercalated cells (Al-Awqati, 2013). To our knowledge, these single-cell RNA-seq results provide the first global definition of the gene expression patterns of these intermediate stages of intercalated cell differentiation.

The interstitial cells (stroma), located between the nephrons, were further analyzed in similar reiterative fashion, revealing four subgroups (Fig. 6, Table S9). A number of genes with elevated expression in the first group of cells showed greatest expression in the outer, cortical stroma, including *Gpc3*, *Aldh1a2*, *Mest* and *Alcam* (Fig. 7). It is important to note that these genes were identified on the basis of their differential expression in the various subgroups of stroma, but their expression is not necessarily stroma specific. The second group of stromal cells expressed genes with elevated expression in the medullary stroma (Fig. 7). Some of these

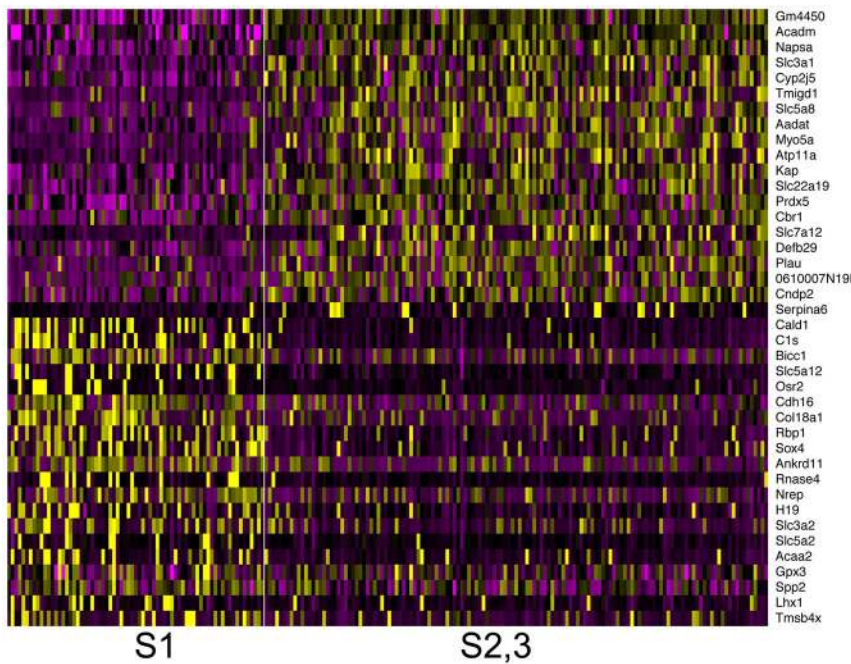


Fig. 4. Heatmap showing subgroups of proximal tubule cells. Unsupervised clustering of the more mature proximal tubule cells generated the two clusters shown, with one expressing multiple segment specific genes associated with the S1 segment, and the other expressing genes known to be expressed in the S2,3 segments. Yellow indicates elevated expression; magenta indicates lower expression; black indicates equal expression.

genes, such as *Zeb2* and *Dcn*, showed expression throughout the medullary stroma, whereas others, such as *Penk* and *Coll5a1*, showed expression in more restricted subdomains of the medullary stroma. It was surprising to find *Wnt11*, a classic marker of the collecting duct tips, in the list of genes separating medullary from cortical stroma. *In situ* hybridizations, however, show that *Wnt11* is, as expected, most strongly expressed in the collecting duct tips, but it also shows weaker, yet significant, expression in the medullary, but not cortical stroma (Fig. 7). The third group of stromal cells appeared to represent the mesangium of the glomerulus. The fourth

group showed elevated expression of a large number of genes associated with cell cycle and mitosis, and therefore represented actively dividing stromal cells.

DISCUSSION

The process of single-cell dissociation for RNA-seq analysis is challenging. The goal is to generate a suspension of single cells, with as few cell doublets or clumps as possible, with as little cell death and lysis as possible, and with optimal preservation of the *in vivo* gene expression patterns. Standard procedures use proteolytic

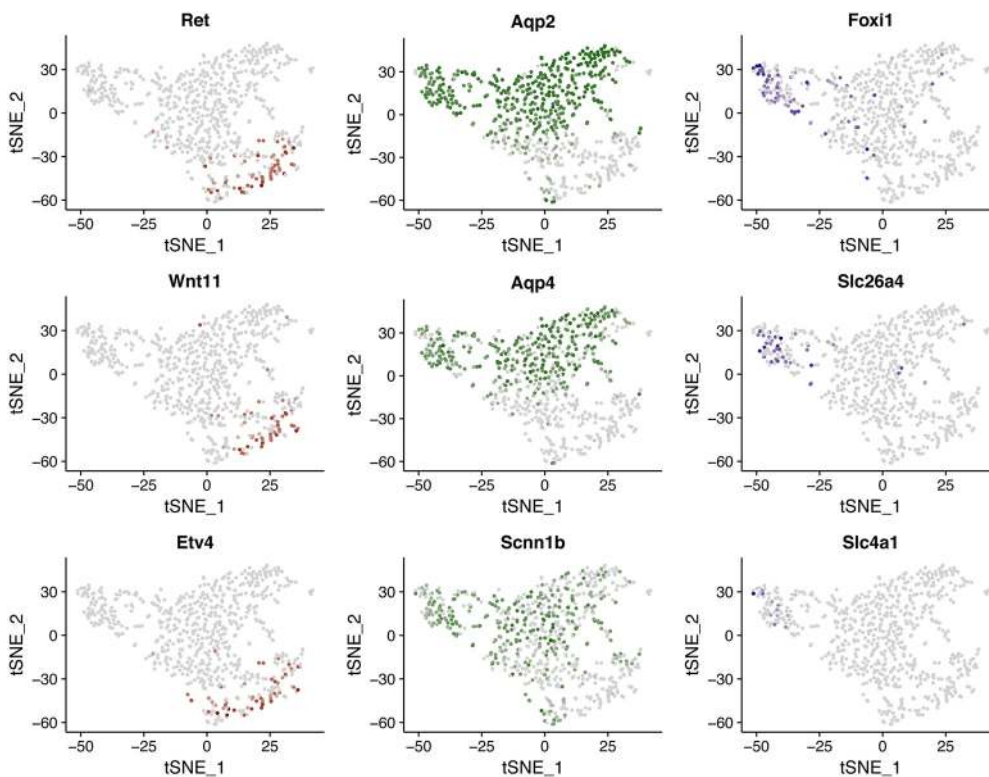


Fig. 5. Cells of the collecting duct system. Unsupervised clustering of the collecting duct cells identified a group expressing ureteric bud tip genes, including *Ret*, *Wnt11* and *Etv4*, as shown in the left column (red cells). Most of the remaining cells expressed principle cell-associated genes, including *Aqp2*, *Aqp4* and *Scnn1b*, as shown in the middle column (green cells). Of interest, a subset of the cells expressing principle cell genes also expressed genes that mark intercalated cells, as shown in the right column (blue cells), suggesting that they represent transitional cells. Many of these cells expressed the β intercalated cell-specific gene *Slc26a4*, whereas few expressed the more differentiated α intercalated cell gene *Slc4a1*.

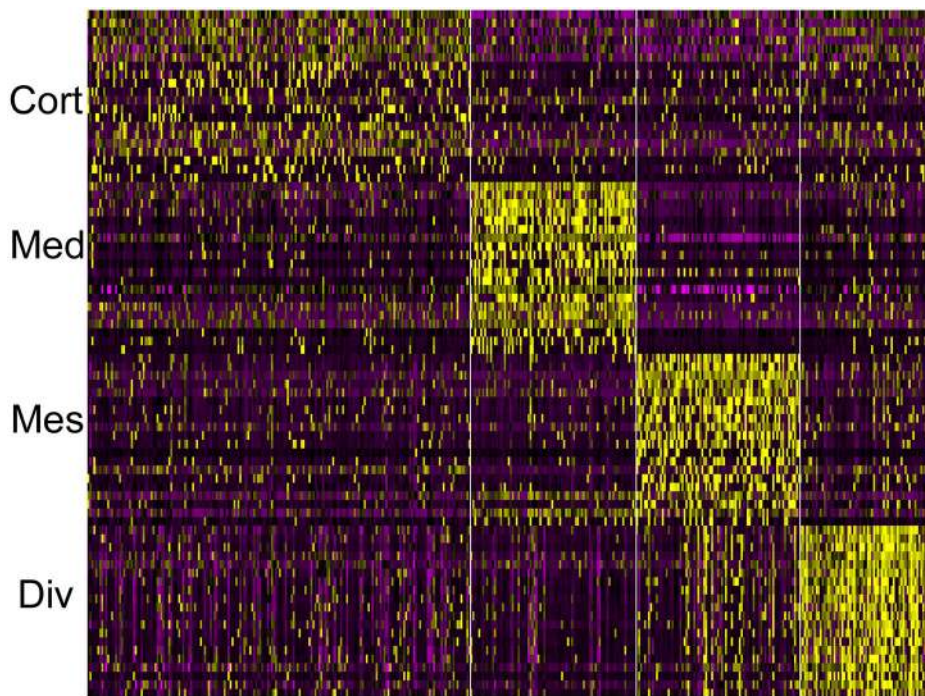


Fig. 6. Heatmap of stromal cell clusters.

Unsupervised clustering of the stromal cells identified four subgroups: cells expressing genes active in the cortical stroma (Cort); cells with genes expressed in the medullary stroma (Med); cells expressing genes associated with the mesangial stroma (Mes); and cells expressing genes associated with active cell division (Div). Yellow indicates elevated expression; magenta indicates lower expression; black indicates equal expression.

digestion at 37°C to help break attachments between cells. In this study, we show that there are large-scale gene expression changes associated with 37°C proteolytic incubations. Some immediately response genes increase in expression level over 1000-fold after only 15 min at 37°C. Cell death-related genes, perhaps related to an anikis response, also show dramatic increases in expression. Some of the gene expression changes associated with 37°C proteolytic incubation were common to all cell types examined,

whereas others were cell type specific. In total, many hundreds of genes showed artifactual changes in expression.

In this article, we show that the gene expression changes associated with single-cell dissociation can be minimized through the use of proteases isolated from psychrophilic organisms adapted to live in the cold. We describe a dissociation method using a protease from *Bacillus licheniformis* that allows the entire procedure to be carried out at 6°C or colder. Under these conditions,

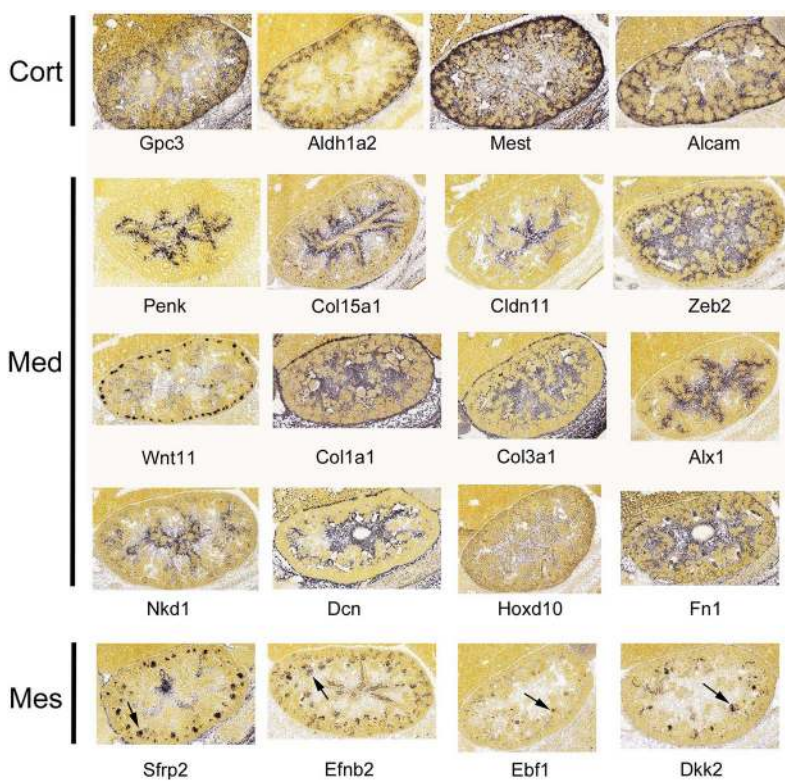


Fig. 7. Expression patterns of stromal-associated genes.

Compartment-enriched expression of the genes associated with the cortical (Cort), medullary (Med) and mesangial (Mes) stromal cells (arrows). The genes shown do not necessarily show stroma-specific expression, but rather serve to separate the different stromal compartments. Of interest, *Wnt11* showed moderate expression in the medullary stroma, and no detected expression in the cortical stroma, although the highest expression was, as expected, in the ureteric bud tips. Images are from *in situ* hybridizations generated as a part of the Allen Brain Atlas project (Jones et al., 2009). (The Allen Institute site contents are a free open resource, provided as a public courtesy.)

mammalian enzymes are effectively inactivated, thereby ‘freezing in’ the *in vivo* gene expression patterns. Using this CAP method, the immediate-early response genes, for example, show dramatically reduced expression compared with protocols requiring 37°C incubation. The use of the CAP method will allow single-cell RNA-seq studies to generate more precise, artifact-free, gene expression datasets.

MATERIALS AND METHODS

Single-cell dissociation of P1 mouse kidneys by cold active protease (CAP)

Twelve P1 CD1 mice were euthanized by decapitation and their bodies immersed in ice-cold PBS. The kidneys were rapidly dissected in ice-cold PBS and then finely minced in a Petri dish on ice using a razor blade. The same kidney mince pool was used for all methods – CAP and 37°C. Fifty milligrams of tissue was added to 2 ml of protease solution [5 mM CaCl₂, 10 mg/ml *Bacillus Licheniformis* protease (Creative Enzymes NATE0633) and 125 U/ml DNase (Applichem, A3778) in Dulbecco’s phosphate-buffered saline (DPBS)] and incubated in a 6°C water bath with trituration using a 1 ml pipet (15 s every 2 min). After 7 min, the solution was transferred to a Miltenyi C-tube (on ice) and the Miltenyi gentleMACS brain_03 program was run twice in a cold room (4°C). The incubation was repeated at 6°C with trituration for an additional 8 min. Single-cell dissociation was confirmed by microscopic examination. Cells were transferred to a 15 ml conical tube, and 3 ml ice-cold PBS with 10% fetal bovine serum (PBS/FBS) was added. Cells were pelleted by centrifugation (1200 g, 5 min), the supernatant was discarded and cells were re-suspended in 2 ml PBS/FBS. Cells were passed through a 30 µm filter (Miltenyi MACS smart strainer) and the filter rinsed with 2 ml PBS/0.01% bovine serum albumin (PBS/BSA). Cells were pelleted (1200 g, 5 min), and re-suspended in 5 ml PBS/BSA, and the PBS/BSA wash was repeated once more. The cells were re-suspended in PBS/BSA and cell concentration determined with a hemocytometer and adjusted to 100,000 cells/ml for Drop-seq.

P1 kidney cell dissociation by 37°C incubation

The same kidney mince pool of cells, prepared as described above, was used for both CAP and 37°C methods. Fifty milligrams of minced kidneys was added to 2 ml digest buffer [3 mg/ml type 2 collagenase (Worthington), 2 mg/ml ProNase E (Sigma, P6911), 125 U/ml DNase (Applichem, A3778), 5 mM CaCl₂ in DPBS]. The tissue was digested at 37°C for 15, 30 or 60 min with trituration every 3 min for 15 s using a 1 ml pipette. After 15, 30 or 60 min, the digestion was stopped by adding 2 ml 10% PBS/FBS. Note that for the 30 min digestion the digest buffer contained 1.5 mg/ml collagenase and 1 mg/ml ProNase E, and for the 60 min digestion it contained 0.75 mg/ml collagenase and 0.5 mg/ml ProNase E. The cells were pelleted by centrifugation at 1200 g for 5 min, the supernatant was discarded and the cells were re-suspended in 6 ml PBS/FBS then filtered with a 30 µm filter, which was then rinsed with 3 ml PBS/BSA. The cells were transferred to a 15 ml conical tube and the volume adjusted to 10 ml with PBS/BSA. Cells were pelleted again, the supernatant discarded and cells re-suspended in 10 ml PBS/BSA and the pelleting procedure was repeated with a final re-suspension in 2 ml PBS/BSA. The cell concentration was determined with a hemocytometer and adjusted to 100,000 cells/ml.

Animals

Outbred CD1 mice were used. Humane procedures were used, following Cincinnati Children’s Hospital Medical Center Institutional Animal Care and Use Committee guidelines (protocol 2D12115).

Data generation and analysis

Drop-seq was carried out as previously described (Macosko et al., 2015). Sequencing was carried out on an Illumina Hi-Seq2500, using one flow cell (about 300 million reads) per sample. Cell-type clusters and marker genes were identified using the R3.3.2 library Seurat1.3.3 (Macosko et al., 2015). All clustering was unsupervised, without driver genes. The influence of the number of unique molecular identifiers was minimized using Seurat’s

RegressOut function. Initial cell filtering selected cells that expressed >1000 genes. Genes included in the analysis were expressed in a minimum of three cells. Only one read per cell was needed for a gene to be counted as expressed per cell. The resulting gene expression matrix was normalized to 10,000 molecules per cell and log transformed (Macosko et al., 2015). Cells containing high percentages of mitochondrial, histone or hemoglobin genes were filtered out. Genes with the highest variability among cells were used for principal components analysis. Clustering was performed with Seurat’s t-SNE implementation using significant principal components determined by JackStraw plot. Marker genes were determined for each cluster using Seurat’s FindAllMarkers function using genes expressed in a minimum of 15% of cells and fold change threshold of 1.3. Over/under clustering was verified via gene expression heatmaps. Subclustering of select clusters was performed as above. Differential expression comparing CAP and 37°C was determined using DEGseq (Wang et al., 2010), and Venn diagrams were prepared with the R package VennDiagram. Data are available in Gene Expression Omnibus under accession number GSE94333.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: S.S.P.; Methodology: A.S.P.; Formal analysis: M.A.; Writing - original draft: S.S.P.; Supervision: S.S.P.; Funding acquisition: S.S.P.

Funding

This work was supported by a grant from the National Institutes of Health (UO1DK107350 to S.S.P.). Deposited in PMC for release after 12 months.

Data availability

RNA-seq data have been deposited in Gene Expression Omnibus under accession number GSE94333.

Supplementary information

Supplementary information available online at <http://dev.biologists.org/lookup/doi/10.1242/dev.151142.supplemental>

References

- Al-Awqati, Q. (2013). Cell biology of the intercalated cell in the kidney. *FEBS Lett.* **587**, 1911–1914.
- Brunskill, E. W., Park, J.-S., Chung, E., Chen, F., Magella, B. and Potter, S. S. (2014). Single cell dissection of early kidney development: multi-lineage priming. *Development* **141**, 3093–3101.
- Chen, J., Bardes, E. E., Aronow, B. J. and Jegga, A. G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311.
- Cheng, C. N. and Wingert, R. A. (2015). Nephron proximal tubule patterning and corpuscles of Stannius formation are regulated by the sim1a transcription factor and retinoic acid in zebrafish. *Dev. Biol.* **399**, 100–116.
- Chiang, M.-K. and Melton, D. A. (2003). Single-cell transcript analysis of pancreas development. *Dev. Cell* **4**, 383–393.
- Du, Y., Kitzmiller, J. A., Sridharan, A., Perl, A. K., Bridges, J. P., Misra, R. S., Pryhuber, G. S., Mariani, T. J., Bhattacharya, S., Guo, M. et al. (2017). Lung Gene Expression Analysis (LGEA): an integrative web portal for comprehensive gene expression data analysis in lung development. *Thorax* **72**, 481–484.
- Fujita, J. (1999). Cold shock response in mammalian cells. *J. Mol. Microbiol. Biotechnol.* **1**, 243–255.
- Harding, S. D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-Macgilp, S., Pi, X., Roochun, Y. et al. (2011). The GUDMAP database – an online resource for genitourinary research. *Development* **138**, 2845–2853.
- Hochheiser, H., Aronow, B. J., Artinger, K., Beaty, T. H., Brinkley, J. F., Chai, Y., Clouthier, D., Cunningham, M. L., Dixon, M., Donahue, L. R. et al. (2011). The FaceBase Consortium: a comprehensive program to facilitate craniofacial research. *Dev. Biol.* **355**, 175–182.
- Jones, A. R., Overly, C. C. and Sunkin, S. M. (2009). The Allen Brain Atlas: 5 years and beyond. *Nat. Rev. Neurosci.* **10**, 821–828.
- Joshi, S. and Satyanarayana, T. (2013). Biotechnology of cold-active proteases. *Biology (Basel)* **2**, 755–783.
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A. and Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201.

- Lee, J. W., Chou, C.-L. and Knepper, M. A. (2015). Deep sequencing in microdissected renal tubules identifies nephron segment-specific transcriptomes. *J. Am. Soc. Nephrol.* **26**, 2669-2677.
- Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168-176.
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M. et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214.
- O'donnell, A., Odrowaz, Z. and Sharrocks, A. D. (2012). Immediate-early gene activation by the MAPK pathways: what do and don't we know? *Biochem. Soc. Trans.* **40**, 58-66.
- Olsson, A., Venkatasubramanian, M., Chaudhri, V. K., Aronow, B. J., Salomonis, N., Singh, H. and Grimes, H. L. (2016). Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature* **537**, 698-702.
- Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. and van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69-73.
- Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877-879.
- Ross, I. L., Browne, C. M. and Hume, D. A. (1994). Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* **72**, 177-185.
- Sato, W. and Sato, Y. (2014). Midkine in nephrogenesis, hypertension and kidney diseases. *Br. J. Pharmacol.* **171**, 879-887.
- Sheng, K., Cao, W., Niu, Y., Deng, Q. and Zong, C. (2017). Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods*.
- Sonna, L. A., Fujita, J., Gaffin, S. L. and Lilly, C. M. (2002). Invited review: effects of heat and cold stress on mammalian gene expression. *J. Appl. Physiol.* **92**, 1725-1742.
- Svensson, V., Natarajan, K. N., Ly, L.-H., Miragaia, R. J., Labalette, C., Macaulay, I. C., Cvejic, A. and Teichmann, S. A. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381-387.
- Thiagarajan, R. D., Georgas, K. M., Rumballe, B. A., Lesieur, E., Chiu, H. S., Taylor, D., Tang, D. T. P., Grimmond, S. M. and Little, M. H. (2011). Identification of anchor genes during kidney development defines ontological relationships, molecular subcompartments and regulatory pathways. *PLoS ONE* **6**, e17286.
- Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., Fisher, J. M., Rodman, C., Mount, C., Filbin, M. G. et al. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglia. *Nature* **539**, 309-313.
- Wang, L., Feng, Z., Wang, X., Wang, X. and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136-138.
- Wingert, R. A. and Davidson, A. J. (2011). Zebrafish nephrogenesis involves dynamic spatiotemporal expression changes in renal progenitors and essential signals from retinoic acid and *irx3b*. *Dev. Dyn.* **240**, 2011-2027.