CHAPTER SEVEN

Uncovering cellular networks in branching morphogenesis using single-cell transcriptomics

Katharine Goodwin[a] **and Celeste M. Nelson**[b,c,*]

[a]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, United States
[b]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, United States
[c]Department of Molecular Biology, Princeton University, Princeton, NJ, United States
*Corresponding author: e-mail address: celesten@princeton.edu

Contents

Abstract

Single-cell RNA-sequencing (scRNA-seq) and related technologies to identify cell types and measure gene expression in space, in time, and within lineages have multiplied rapidly in recent years. As these techniques proliferate, we are seeing an increase in their application to the study of developing tissues. Here, we focus on single-cell investigations of branching morphogenesis. Branched organs are highly complex but typically develop recursively, such that a given developmental stage theoretically contains the entire spectrum of cell identities from progenitor to terminally differentiated. Therefore, branched organs are a highly attractive system for study by scRNA-seq. First, we provide an update on advances in the field of scRNA-seq analysis, focusing on spatial transcriptomics, computational reconstruction of differentiation trajectories, and integration of scRNA-seq with lineage tracing. In addition, we discuss the possibilities and limitations for applying these techniques to studying branched organs. We then discuss exciting advances made using scRNA-seq in the study of branching

Current Topics in Developmental Biology, Volume 143
ISSN 0070-2153
https://doi.org/10.1016/bs.ctdb.2020.09.004

morphogenesis and differentiation in mammalian organs, with emphasis on the lung, kidney, and mammary gland. We propose ways that scRNA-seq could be used to address outstanding questions in each organ. Finally, we highlight the importance of physical and mechanical signals in branching morphogenesis and speculate about how scRNA-seq and related techniques could be applied to study tissue morphogenesis beyond just differentiation.

Abbreviations

AT1	alveolar type 1
AT2	alveolar type 2
ECM	extracellular matrix
Fgf10	fibroblast growth factor 10
Fgfr2	fibroblast growth factor receptor 2
FACS	fluorescence activated cell sorting
FRET	Förster resonance energy transfer
GDNF	glial cell line-derived neurotrophic factor
PGK	3-phosphoglycerate kinase
pMLC	phosphorylated myosin light chain
Shh	Sonic hedgehog
TEB	terminal end bud
Wnt	Wingless-related integration site

1. Introduction

Branching morphogenesis requires reciprocal interactions between branching epithelia and their surrounding supporting cells within the mesenchyme, including endothelial cells and adipocytes. Patterning of branched organs is a complex process that integrates molecular and physical signals and involves crosstalk between cell types. The cellular networks that are generated during branching morphogenesis are critical to organ development and function. Therefore, it is essential to understand how these cellular networks arise and evolve throughout development.

With the advent of single-cell RNA sequencing (scRNA-seq) approaches, researchers can now simultaneously profile the gene expression patterns of diverse cell types. These techniques are currently widely used in all areas of biology to investigate many different species. Whole organisms can be sequenced at multiple developmental stages, as has been reported for *C. elegans*, *Ciona intestinalis*, and *Drosophila melanogaster* (Cao et al., 2019; Karaiskos et al., 2017; Packer et al., 2019). In larger organisms, including

mammals, transcriptomes have been constructed for the earliest stages of development (Nakamura et al., 2016; Sasaki et al., 2016). During later stages of development or in mature animals, maps of whole organs have been created (Schaum, Karkanias, Neff, et al., 2018).

The development of branched organs occurs recursively, with each stage of the branching process repeating over and over in different parts of the organ. Consequently, single snapshots in time could potentially include cells at each stage of the process of morphogenesis and differentiation. This property makes the study of branching morphogenesis by single-cell transcriptomics especially attractive. In this review, we ask: what can we uncover about the cellular networks that drive branching morphogenesis using single-cell transcriptomics? First, we discuss how technological and computational advances in scRNA-seq can be used to study the morphogenesis of branched organs. In Section 2, we provide an overview of scRNA-seq-related tools that are particularly applicable to studying development and differentiation. In Section 3, we describe examples of cellular networks and physical factors that direct cell differentiation and morphogenesis in branched organs. Specifically, we highlight recent advances in the study of lung, kidney, and mammary gland organogenesis that have been enabled by scRNA-seq approaches. We also discuss examples of how single-cell transcriptomics might be used to deepen our understanding of cellular networks in branching morphogenesis. Finally, in Section 4, we speculate about how single-cell transcriptomics and related technologies could be used to uncover the morphological and physical changes that influence organ development and cell differentiation.

2. scRNA-seq tools for studying morphogenesis

Cells evaluated by scRNA-seq are characterized based on the complement of genes that they express (their transcriptomes), allowing us to cluster them into specific cell types. Cells with similar transcriptional profiles will appear as neighbors in low-dimensional representations of scRNA-seq data, such as t-stochastic neighbor embedding (tSNE) or uniform manifold approximation and projection (UMAP) plots (McInnes, Healy, & Melville, 2018; van der Maaten & Hinton, 2008). For example, epithelial cell types of the lung, including ciliated and clara cells of the airways as well as alveolar type 1 (AT1) and type 2 (AT2) cells, would all be clustered by type and appear as neighbors in low-dimensional representations based on their transcriptomes (Fig. 1A–B). Commonly used clustering packages for scRNA-seq

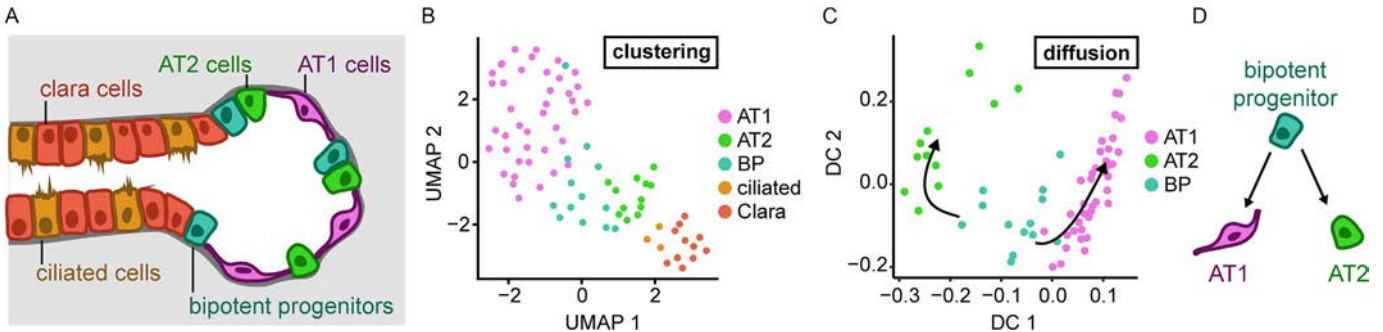

Fig. 1 Primer for scRNA-seq analyses. (A) Schematic illustrating some of the different epithelial cell types in a developing mouse lung. (B) UMAP representation of scRNA-seq data of mouse lung epithelial cells isolated at *E*18.5 (Treutlein et al., 2014) generated using the Seurat package (Butler, Hoffman, Smibert, Papalexi, & Satija, 2018) showing clusters of each epithelial cell type in (A). Note that neighbors in the UMAP are not physical neighbors within the tissue, but cells with similar transcriptomes. (C) Diffusion analysis of cell types indicated in (A) showing transitions between different cell identities implemented using the Destiny algorithm (Haghverdi, Buettner, & Theis, 2015). (D) Lineage diagrams inferred from diffusion analysis in (C).

data include Seurat (Butler et al., 2018) and Monocle (Trapnell et al., 2014). Importantly, the typical experimental pipeline of dissociating tissues into a single-cell suspension destroys all spatial information (such as physical neighbors, as opposed to neighbors in the high-dimensional space of transcriptomic data). In Section 2.1, we discuss several approaches that have been developed to address this caveat by computationally inferring or experimentally preserving spatial information in scRNA-seq experiments.

While clustering can be used to identify cell types, other algorithms have been developed to study transitions between them. These tools are especially useful for the study of morphogenesis, since developing tissues should contain a continuum of differentiating cells. For example, diffusion analysis implemented using the Destiny algorithm identifies transition probabilities between cells in high-dimensional space. Diffusion analysis works by allowing each cell to move around its position according to a defined wave function and computing the interference between the wave functions of each pair of cells (Haghverdi et al., 2015). The resulting diffusion components (similar to principal components in principal components analysis) capture variability in the dataset and can be used to infer differentiation trajectories. In our example above of epithelial cell types in the lung, we would expect that diffusion analysis could order cells along the trajectory from bipotent progenitor to differentiated cell, with separate paths for AT1 and AT2 cells (Fig. 1C–D). Diffusion analyses provide a simple way to infer differentiation trajectories from a single dataset, but of course biological transitions such as those occurring during morphogenesis are quite complex. In Section 2.2, we provide an overview of computational tools designed to reconstruct complex differentiation trajectories. Finally, simply observing a single time point by scRNA-seq may not provide sufficient temporal information, necessitating the use of lineage tracing in combination with transcriptomics (Wagner & Klein, 2020). In Section 2.3, we describe some experimental extensions to traditional scRNA-seq analyses that enable more detailed reconstructions of lineage decisions.

2.1 Integrating spatial information with scRNA-seq

The typical scRNA-seq pipeline discards all spatial information. In this section, we discuss different strategies for integrating spatial information with scRNA-seq. Novel cell types and marker genes can be mapped back onto the tissue of interest using experimental validation or computational inference. To visualize gene expression in situ at subcellular resolution, target

RNAs can be visualized using multiplexed fluorescence in situ hybridization (FISH). Alternatively, in situ RNA sequencing can be used to profile all RNAs within a tissue at subcellular resolution. Finally, slide-based methods can be used to capture scRNA-seq data in such a way that positional information is encoded.

The simplest approach to map clusters identified by scRNA-seq back onto the tissue is by immunostaining or in situ hybridization. These techniques have been used to confirm the presence of novel cell clusters or to validate the expression patterns of new marker genes. In one example, marker genes for computationally identified clusters of bipolar cell types of the retina were validated by combining sparse labeling of cell boundaries (to enhance spatial resolution in such a tightly-packed tissue) with FISH for genes expressed even at low levels (Shekhar et al., 2016).

Depending on the complexity of the architecture of the target tissue and the variety of cell types, it is sometimes possible to infer spatial coordinates within a tissue from single-cell data based on the analysis of published in situ hybridization datasets. In some cases, trends in the data happen to recapitulate spatial patterns in the embryo, as in a recent scRNA-seq study of the mouse gut endoderm in which the main diffusion component of the single-cell data reflected variability along the anterior–posterior axis of the gut tube (Nowotschin et al., 2019). Otherwise, more complex mappings need to be generated. The initial publication of the Seurat algorithm included such an approach to identify spatial expression patterns in the zebrafish embryo at the blastula stage (Satija, Farrell, Gennert, Schier, & Regev, 2015). Taking advantage of published in situ data and the simple geometry of the zebrafish embryo, a spatial gene-expression map was constructed that could be compared to single-cell transcriptomes. Based on similarity between expression patterns, the probable original locations of each cell in the scRNA-seq dataset within the embryo could be inferred. DistMap used a similar idea to generate three-dimensional maps of predicted gene expression in the *Drosophila* embryo at the onset of gastrulation (Karaiskos et al., 2017). Similar approaches have also been applied to spatial mapping of scRNA-seq data in the *Drosophila* wing imaginal disc (Deng et al., 2019).

Computationally mapping scRNA-seq data onto a target tissue is most easily accomplished when the tissue has a simple geometry that is reproducible between individuals. As a result, it may be challenging to generate faithful mappings of tissues with complex architecture like those of branched organs. It is therefore also important to consider less probabilistic,

experimental ways of mapping gene expression of novel clusters onto the tissue of interest or for discovering new spatial expression patterns. To address this need, a plethora of technologies have been developed to either spatially map the expression of more and more genes (multiplexing) or to obtain complete transcriptomes in situ (in situ sequencing).

Multiplexed FISH can be used to identify several individual mRNAs within cells while preserving spatial information. These approaches can be especially useful for mapping many different cell types identified by scRNA-seq within highly heterogeneous tissues. Barcodes made up of multiple probes that are conjugated to different fluorophores permit the simultaneous visualization of multiple mRNAs (more than the number of fluorophores), with the number increasing as barcode complexity increases (Chen, Boettiger, Moffitt, Wang, & Zhuang, 2015; Levsky, Shenoy, Pezo, & Singer, 2002; Lubeck & Cai, 2012). In one technique, different sets of labeled probes are hybridized to RNAs within a fixed cell over the course of multiple rounds of washing or digesting with DNase followed by imaging in order to spatially map RNA sequences of interest (Fig. 2Ai–ii). Other strategies include localizing mRNAs within fixed cells based on the spatial ordering or spectral overlap of fluorophores bound to probes that hybridize adjacent to each other on an RNA molecule of interest (Fig. 2Aii–iii) (Lubeck, Coskun, Zhiyentayev, Ahmad, & Cai, 2014). The single-molecule imaging required for multiplexed FISH is technically challenging, especially in tissues that scatter light or that are autofluorescent. To overcome these issues in order to study cell types in the lung, Nagendran et al. developed a multiplexed approach called proximity ligation in situ hybridization (PLISH). Briefly, rolling-circle replication is initiated at hybridized probes to generate long amplicons with many repeated barcodes to which several "imager" oligonucleotides can bind, thus increasing the number of fluorophores at a given RNA and drastically increasing the signal-to-noise ratio (Nagendran, Riordan, Harbury, & Desai, 2018).

Recent work has extended these kinds of techniques to live cells. For example, Atmanli et al. have designed a technique called multiplex analysis of gene expression in individual living cells (MAGIC), in which live cells are cotransfected with an RNA probe and an engineered double-stranded RNA-binding protein that are each labeled with an acceptor or donor fluorophore, respectively. When the probe and the binding protein are both bound to a target RNA, a FRET signal can be detected (Atmanli et al., 2019). Combining MAGIC probe and protein pairs with different fluorophores and/or FRET pairs scales this technique up to enable the

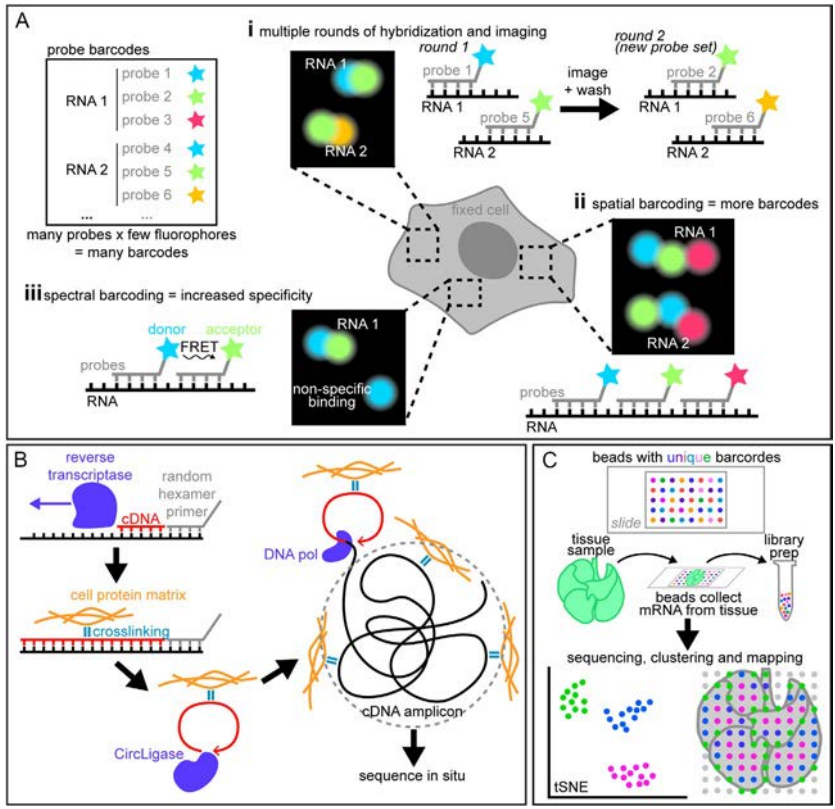

Fig. 2 Combining spatial information with scRNA-seq. (A) Multiplexed FISH, which combines many probes with few fluorophores, enables visualization of several different RNA sequences within a single sample. Increasing numbers of RNA sequences can be visualized by using successive rounds of hybridization, imaging, and washing (i) or by using spatial barcoding (ii), and higher-resolution FISH for denser mRNAs can be achieved using spectral barcoding (iii). In spatial barcoding (ii), probes for adjacent regions of RNA can be labelled with different fluorophores so that spatial ordering of fluorophores can identify different mRNAs. In spectral barcoding (iii), probes for adjacent regions of RNA can carry activator/emitter fluorophore pairs so that emitter signal is only detected when both probes have bound the same RNA molecule. (B) In situ RNA sequencing allows for detection of transcripts within a cell without the need for specific probes. Random hexamer primers initiate reverse transcriptase activity and generate cDNA within the sample, and the resultant cDNA is then cross-linked to the cell's protein matrix to preserve its location. After circularization, DNA polymerase generates cDNA amplicons that are also cross-linked in place and that can be sequenced in situ. (C) Slide-seq uses barcoded droplets (similar to Drop-seq) that have known spatial coordinates on a slide. Tissue sections placed on these slides will deposit mRNA onto each droplet, and a map of transcriptomes can be reconstructed based on droplet barcodes and the slide layout.

visualization of multiple RNAs within living cells, and could be used to track changes in gene expression during morphogenesis. However, this technique is subject to imaging constraints due to spectral overlap between FRET pairs, and therefore cannot be scaled up to the same extent as the approaches discussed above that have been applied to fixed cells.

Multiplex FISH and related techniques can be very informative but require prior knowledge of RNA sequences of interest and careful probe design. Unlike scRNA-seq, these approaches are therefore limited to known marker genes and cannot be used to discover new ones. To find new marker genes while also maintaining spatial context, alternative techniques such as RNA sequencing in situ have been developed. For example, fluorescent in situ sequencing (FISSEQ) involves reverse transcribing RNA within fixed cells to generate cDNA amplicons that are then cross-linked to the cellular protein matrix to preserve their position within the sample (Fig. 2B) (Lee et al., 2014). The amplicons can then be directly sequenced within the tissue using standard sequencing approaches based on imaging the sequential ligation of labeled oligonucleotides. This approach is most suitable for cells or thin sections and has been applied to whole fruit fly embryos. Extending this approach to thicker samples or adult tissues may require combining in situ sequencing approaches with sample clearing. For example, transcriptome mapping in thick sections of the mouse brain was accomplished by embedding samples in hydrogels to which amplicons could be crosslinked, and then removing proteins and lipids to render the samples optically transparent (Wang et al., 2018). This approach maintains physical context and enhances imaging capabilities to generate high-resolution spatial transcriptomic data. This adaptation could also be applied to carry out in situ RNA-sequencing of branching organs, given that a sufficient number of morphological features could be captured in thick sections. Of the techniques discussed here, multiplex FISH and in situ sequencing have the highest level of resolution since they can map transcripts to their subcellular locations.

Finally, slide-based spatial transcriptomics is an alternative approach to spatial RNA-sequencing that enables untargeted readouts of RNA content with positional information, albeit with lower spatial resolution. For example, in Slide-seq, tissue sections are transferred onto a slide that is coated with beads carrying known DNA barcodes, which map them to their positions on the slide (Fig. 2C) (Rodriques et al., 2019). Each bead collects RNA from the tissue sample around it, and then beads are recovered from the slide and can be processed by standard scRNA-seq protocols. Based on the barcodes,

a spatial map of RNA expression can be constructed (Fig. 2C). The main caveat to this approach is the spatial resolution: depending on inter-bead distance, a given bead could potentially collect RNA from adjacent cells (of different cell types). However, each bead from the slide can in principle be mapped to one or two cell types (Rodriques et al., 2019), demonstrating that close to single-cell resolution can achieved with this approach.

Naturally, the ideal spatial transcriptomic approach for studying branched tissues depends on the biological question being asked. The three-dimensional architecture of branched organs poses problems for many of the spatial gene-expression profiling approaches discussed above, which are most successful for (or exclusive to) tissue sections. However, since branched organs typically contain many repeated units of tissue structures, it may be possible to apply certain computational mappings usually reserved for simple geometries, or to make more confident conclusions based on tissue sections.

2.2 Reconstructing differentiation trajectories with scRNA-seq

Developing tissues, especially those that undergo recursive morphogenesis such as branching, contain cells at each stage of differentiation. Therefore, single cells isolated from a given tissue can theoretically be ordered along "pseudotime" from the most undifferentiated to the most terminally differentiated. Several pseudotemporal ordering algorithms have been developed for single-cell transcriptomics and other single-cell modalities, and most are available as open-source software packages. Importantly, these methods typically do not require sequential experimental time points, as would be required for bulk RNA-seq studies of differentiation, as they assume that temporal information can be inferred from the spectrum of cell states within a single sample.

In addition to its clustering utilities, Monocle was originally developed as a method for ordering cells along developmental trajectories (Trapnell et al., 2014). In this approach, dimensional reduction algorithms are first applied to transform data from a high-dimensional Euclidean space (in which each dimension corresponds to a gene) to a low-dimensional space that preserves the essential relationships between each cell's transcriptome. Next, a tree is constructed that connects all cells in low-dimensional space. In this context, the tree represents a graph connecting all cells to each other in low-dimensional space, with weights assigned to each branch between cells. The tree is then simplified to generate a minimum spanning tree (MST), which only retains branches such that the total weight of all branches is minimized (i.e., selecting

for branches that connect cells to their closest neighbor) and such that all cells are still attached to the tree. The Monocle algorithm then identifies the longest path through the MST to produce a trajectory along which cells can be ordered and that ideally recapitulates the incremental changes in gene-expression profiles that occur during differentiation. An alternative approach, called Wanderlust, starts by converting data to a nearest-neighbor graph, and then moves through the high-dimensional graph following the shortest path from a user-defined "initiator" cell, which gives the trajectory an orientation (e.g., from a progenitor "initiator" through differentiation). Longer trajectories are associated with more noise, so Wanderlust uses "waypoint" cells to more accurately map cells along the entire length of the trajectory (Bendall et al., 2014).

Whereas diffusion analysis and the original Monocle algorithm can be used to identify simple differentiation trajectories, the biological reality is often more complex. Progenitor cells can give rise to several differentiated cell types with distinct gene-expression programs, and transdifferentiation or convergent differentiation can complicate these trajectories even further. Branch points in datasets sampled from different time points can be identified using the SCUBA algorithm (Marco et al., 2014), for example, but one of the most attractive features of single-cell data is that an individual sample can contain all developmental time points. Methods for analysis of datasets without external temporal information are therefore highly desirable. To meet these needs, algorithms have been designed to identify branch points along pseudotime. The successor to Wanderlust, called Wishbone, specifically orders cells along bifurcating developmental trajectories (Setty et al., 2016). Wishbone infers trajectories from a low-dimensional representation of the data obtained using diffusion maps. Projecting the data onto the top diffusion components effectively removes noise and spurious connections between cells, which are especially problematic near branch points. It then employs the waypoints innovation from Wanderlust to identify branch points, since the perspectives of waypoints on different branches relative to each other and to the initiator cell will be different. Importantly, Wishbone can only recapitulate differentiation trajectories with one or two final fates.

Several approaches have also been developed to map trajectories with more branches. In the MPath algorithm, clusters from single-cell data are used to define landmarks (i.e., representative cell states or specific gene-expression profiles) that can be connected by graphs (Chen, Schlitzer, Chakarov, Ginhoux, & Poidinger, 2016). Cells are then compared to each landmark and placed on edges between those that they most closely

resemble. Finally, an MST is created that connects all landmarks and that takes into account how many cells occupy each edge in order to build the final trajectory. A similar approach, Slingshot, builds MSTs on low-dimensional data that have already been clustered, and then employs a curve-fitting approach to estimate the transitions between clusters (Street et al., 2018). To order cells in pseudotime, they are then assigned positions along the principal curves between clusters based on their Euclidean distance from the curve. Since these approaches rely on accurate landmark selection, they are limited by how well the data can be clustered initially. Developmental trajectories with stable intermediate states (i.e., where large numbers of cells at similar stages of differentiation can be sequenced) will likely be easier to cluster faithfully and thus provide more accurate landmarks.

Other approaches have been developed that map transitions between all cells (and that are therefore not as dependent upon initial clustering), including Monocle2 (Qiu et al., 2017) and Palantir (Setty et al., 2019), both evolutions of algorithms discussed above (Monocle and Wanderlust/Wishbone). Monocle2 employs a machine-learning approach called reverse graph embedding (RGE) that can infer differentiation trajectories without any prior knowledge of progenitor cell identity, the number or characteristics of terminally differentiated states, or the number of branch points (Qiu et al., 2017). The algorithm uses a subset of automatically identified genes based on their differential expression between cell types identified by tSNE and clustering (effectively reducing dimensionality by focusing on a subset of genes), and then applies RGE to find a faithful mapping between high- and low-dimensional space while simultaneously learning the shape of the principal graph (i.e., the branched curve that best passes through the data) in low-dimensional space. In contrast, Palantir builds upon the earlier tools to enable reconstruction of trajectories with multiple branches and estimates cell-fate probabilities by modeling differentiation as a Markov process: the algorithm assumes that progression along a differentiation trajectory occurs in stochastic, incremental steps (Setty et al., 2019). This approach allows Palantir to estimate probabilities between distant cells and to define "differentiation potentials" for each cell. Cells from early steps in pseudotime will have equal differentiation potential for each terminal state, whereas cells from further along pseudotime that have passed branch points will have a high differentiation potential for the terminal state at the end of the branch in which they reside and a low differentiation potential for the other terminal states.

Pseudotime analyses could be quite useful for investigating branching morphogenesis, and choice of algorithm could depend on anticipated complexity of the differentiation tree in the tissue of interest. Importantly, these computational approaches (while usually informative) provide us only with probabilities that must be tested experimentally.

2.3 scRNA-seq combined with lineage tracing

Pseudotime analyses have provided invaluable insights into the mechanisms of cell differentiation in development and homeostasis, but they must often be validated with lineage-tracing experiments. To circumvent this requirement, some researchers have incorporated lineage tracing directly into scRNA-seq experiments. Pulse-seq is a simple and elegant way to incorporate true lineage tracing into scRNA-seq experiments, and was used to track airway epithelial cell differentiation during homeostasis (Montoro et al., 2018). In pulse-seq, mice expressing inducible fluorescent lineage reporters are harvested and sequenced at different stages after induction to label both progenitor cells as well as the differentiated cells they generate (Fig. 3A). If an airway epithelial basal-cell marker is used for lineage labeling, for example, then label-positive and -negative cells can be sorted and sequenced separately prior to re-integration for data analysis, or the lineage label could be read out in the scRNA-seq data. In data collected shortly after lineage-label induction, the label-positive cells will contribute only to the basal-cell cluster, while data collected weeks after label induction will include label-positive cells in all of the clusters that can be generated by the basal-cell progenitors. While informative, this approach requires prior knowledge of progenitor markers, preventing discovery of novel markers, and provides population-level lineage tracing, thus obscuring any variability among basal cells that could influence their differentiation.

An alternative strategy for lineage labeling is to incorporate barcodes into cells that can be used to reconstruct their histories. Several techniques have been developed to accomplish this using different approaches for barcode design and recovery (Kebschull & Zador, 2018; Wagner & Klein, 2020). One of the earliest techniques developed for cellular barcoding to infer lineage was called GESTALT, or genome editing of synthetic target arrays for lineage tracing (McKenna et al., 2016). Briefly, an array of CRISPR/Cas9 targets is engineered into the genome (downstream of a transcribed reporter gene) and cells harboring this barcode are transfected with a plasmid

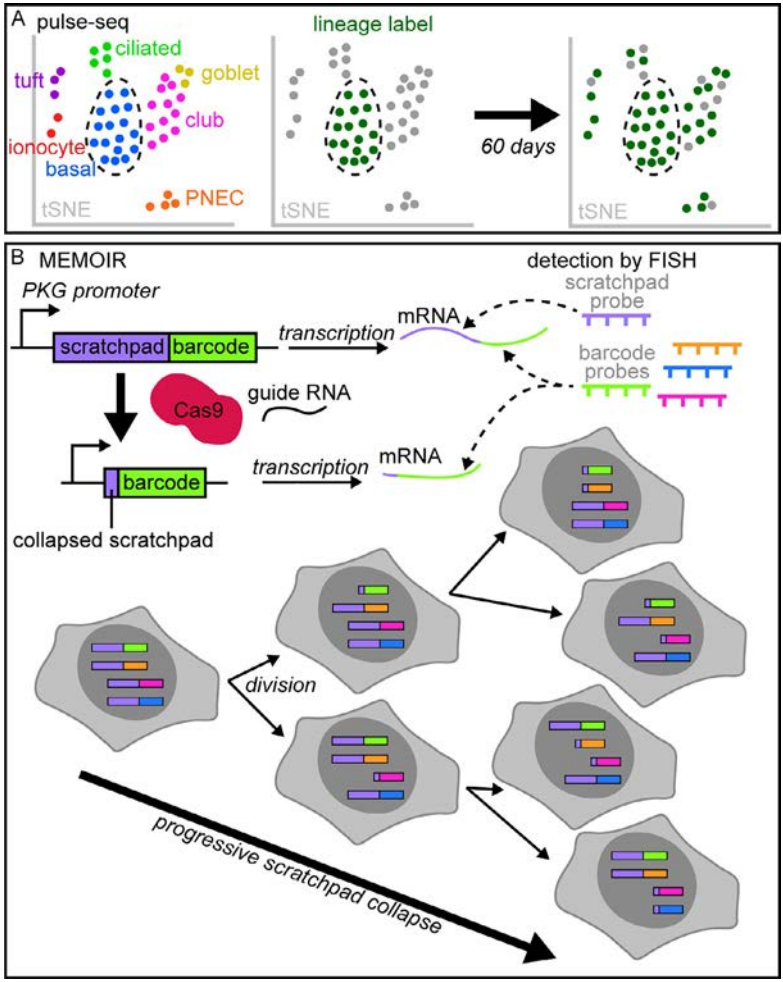

Fig. 3 Combining lineage information with scRNA-seq. (A) Classic lineage-tracing approaches like Cre-lox mouse models can be combined with scRNA-seq to determine how progenitor cells (that express Cre and a reporter) contribute to different cell types by collecting and analyzing lineage-labeled samples at different time points. In this example, the contribution of basal cells to other lineages in the airway epithelium is assessed using pulse-seq. (B) Dynamic barcoding can be used to trace cell lineages by multiplexed FISH or by scRNA-seq. In one such technique, a specific barcode and an adjacent scratchpad sequence in multiple places throughout the genome under the control of a PKG promoter. The scratchpad regions undergo stochastic Cas9-mediated mutagenesis as cells proliferate, and incremental changes in the sequences of a cell's scratchpads can be detected by FISH and used to infer lineage relationships.

encoding Cas9 and the guide RNA. The targets within the barcode are optimized and arranged such that the first is a perfect match for the guide RNA and the last is the least-favorable binding sequence. Over successive rounds of division, Cas9 edits the target sequences somewhat stochastically, causing random and heritable alterations to the barcode that can be read with DNA or RNA sequencing. By incorporating barcodes and the CRISPR/Cas9 components into zebrafish embryos at the one-cell stage, each organ could be harvested from adults and sequenced to determine how many different barcodes it contained. Most organs carried only a few barcodes, which tended to be distinct from those in other organs, suggesting that the majority of cells within each tissue in the adult are derived from just a few embryonic precursors.

A slightly different technology, memory by engineered mutagenesis with optical in situ readout (MEMOIR), also uses CRISPR/Cas9-mediated dynamic barcoding, but edits are made at several loci instead of within a single barcode region (Frieda et al., 2017). In MEMOIR, Cas9 generates double-stranded breaks within "scratchpad" regions adjacent to unique barcodes throughout the genome. As cells proliferate, their collection of scratchpads undergoes stochastic and incremental changes due to Cas9 activity that are heritable and can therefore be analyzed and traced to infer lineage relationships (Fig. 3B). The barcode and scratchpad regions are downstream of a 3-phosphoglycerate kinase (PGK) promoter, and their mRNA can be detected and identified by specific probes using sequential single-molecule FISH. At the same time, gene expression can be measured using single-molecule (and possibly multiplexed) FISH. This approach differs in functionality from GESTALT, as it can also be used to examine gene expression in each cell lineage. One interesting aspect of this technique is that barcodes and scratchpads can be identified in situ—conceivably, one could engineer an embryo or an organoid to express the MEMOIR components and then spatially map cellular histories.

To truly combine scRNA-seq with lineage tracing, barcodes need to be detected using the same experimental pipeline as that for a normal sequencing experiment and then faithfully assigned to cells along with their transcriptomes. Several groups have created techniques to accomplish this goal, including the developers of GESTALT, which could only be used to trace lineages at the tissue level and not in single cells. The evolution of this approach, termed scGESTALT, relies on transcription of the cell barcode into mRNA that can be detected using scRNA-seq (Raj et al., 2018). Other comparable techniques also applied to zebrafish development

include ScarTrace (Alemany, Florescu, Baron, Peterson-Maduro, & van Oudenaarden, 2018), which similarly uses a genomic barcode region at a single locus, and lineage tracing by nuclease-activated editing of ubiquitous sequences (LINNAEUS) and TracerSeq, which use multiple barcodes distributed throughout the genome (Spanjaard et al., 2018; Wagner et al., 2018). Similar progress has also been made in the study of mammalian development by introducing barcoding sequences (via a piggyBac transposon that efficiently inserts into the genome) along with sperm constitutively expressing Cas9 into the oocyte and then transplanting blastocysts into female mice (Chan et al., 2019), or by crossing females harboring the barcoding sequences to males constitutively expressing Cas9, allowing barcoded embryos to develop entirely in vivo (Kalhor et al., 2018). In both cases, continuous barcoding occurs throughout gestation after zygotic genome activation.

Importantly, all these approaches start barcoding from the beginning of embryogenesis, which may limit their use for the study of branching morphogenesis since most branched organs begin developing partway through embryogenesis or during puberty. Indeed, both scGESTALT and ScarTrace studies revealed that Cas9-mediated barcode editing activity ends by the time gastrulation is complete (Alemany et al., 2018; Raj et al., 2018). scGESTALT overcomes this issue by combining early and late barcoding: early barcode editing is carried out by Cas9 protein injected at the one-cell stage, while late barcode editing is carried out by inducible Cas9 expression under the control of a heatshock-sensitive promoter (Raj et al., 2018).

How these technologies might be adapted to study the developing mammalian lung or kidney, for example, is not yet clear. Inducible editing is a promising avenue, albeit by means other than heatshock, and could theoretically be carried out using tissue-specific Cre-mediated induction of Cas9. However, the computational challenges are enormous since barcoding would begin simultaneously in many cells. Barcoding of heterogeneous populations has been carried out in mammalian hematopoietic cells that are either allowed to proliferate in culture or are injected into mice and then harvested for sequencing (Weinreb, Rodriguez-Fraticelli, Camargo, & Klein, 2020). Perhaps the computational innovations from these investigations could also be applied to lineage tracing of heterogeneous populations of cells induced to express Cas9 that begin barcode editing in the early stages of lung or kidney development. Lineage tracing by pulse-seq-related methods is likely a useful tool for delving further into the contributions of cells expressing known lineage markers to the development of branched organs.

However, lineage tracing by barcoding in scRNA-seq remains challenging and highly labor intensive. Before these approaches are more accessible to the study of branching morphogenesis, certain technological advances in animal models that encode barcodes and computational approaches for reconstructing lineage trees are needed.

3. Cellular networks in branching morphogenesis

Building a branched organ and correctly patterning differentiation within it requires cooperation between many different cell types. As a result, diverse cellular networks exist between the cells of branched organs, and these typically encompass multiple tissue types including epithelia and mesenchyme. Single-cell transcriptomics of such organs provide us with an unprecedented level of detail into the composition of these cellular networks and shed light on how they are established and how they regulate organ patterning. With emphasis on branched epithelia within mammalian organs, we highlight important insights into branching morphogenesis that have been made using scRNA-seq and related techniques, and suggest areas where such approaches could help to answer outstanding questions specific to each organ.

3.1 Lung development

In the mouse embryo, the lung is initiated from the foregut endoderm around *E*9.5. By *E*10.5, the lung is a simple wishbone shape comprised of two epithelial tubes, and by *E*11.5, the four lobes of the right lung are established. During the pseudoglandular stage of development (*E*11.5-*E*15.5), a highly stereotyped program of branching morphogenesis constructs the tree-like architecture of the airways (Metzger, Klein, Martin, & Krasnow, 2008; Zepp & Morrisey, 2019). The airway epithelium branches into the surrounding pulmonary mesenchyme, while mesenchymal cells near branch tips differentiate into airway smooth muscle and wrap circumferentially around the epithelium (Zepp & Morrisey, 2019) (Fig. 4A). Elaboration of the tree is driven by bifurcations or domain (lateral) branches, the formation of which is regulated by molecular and physical stimuli (Metzger et al., 2008; Zepp & Morrisey, 2019).

Fibroblast growth factor 10 (Fgf10) expressed by the mesenchyme is required for lung development and signals to the epithelium via fibroblast growth factor receptor 2 (Fgfr2) to induce the expression of Sonic hedgehog (Shh). Epithelial Shh in turn signals to the mesenchyme to inhibit the

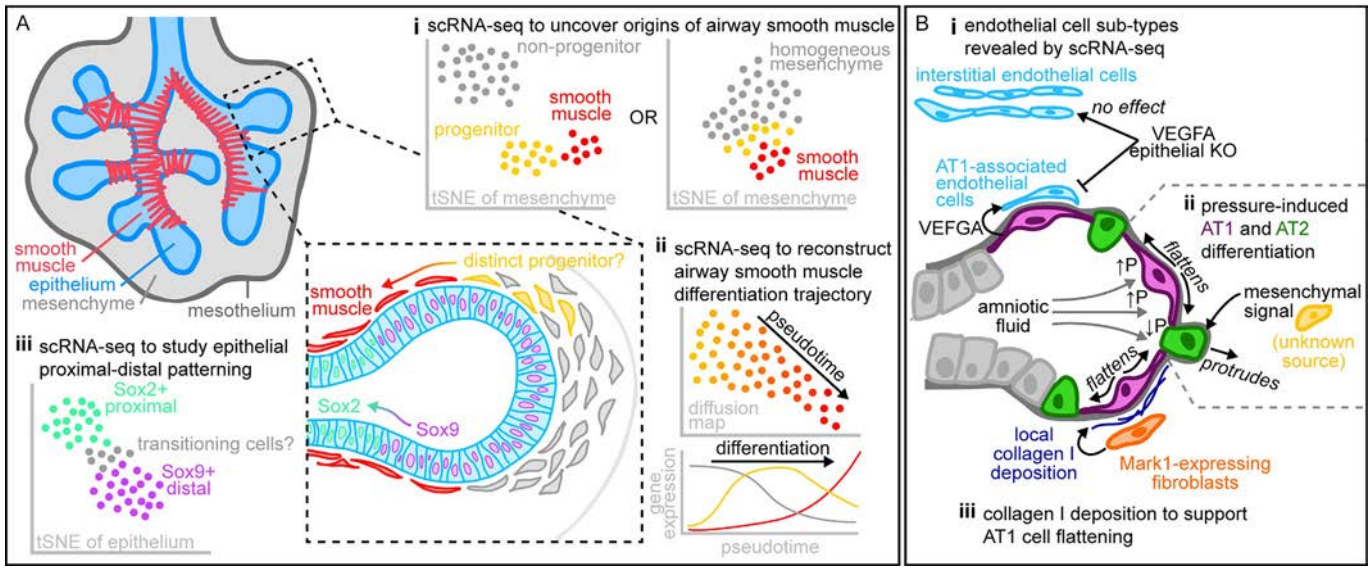

Fig. 4 scRNA-seq to study cellular networks and branching morphogenesis of the mammalian lung. (A) During the pseudoglandular stage of lung development, the branched architecture of the airway epithelium is constructed with help from the surrounding undifferentiated mesenchyme, smooth muscle, and mesothelium. The mesenchymal origins of airway smooth muscle could theoretically be identified using scRNA-seq (i), and pseudotime analyses could be used to reconstruct and examine gene-expression changes along the smooth muscle differentiation trajectory (ii). Lastly, as the epithelium branches, its constituent cells take on proximal or distal fates—scRNA-seq could be used to uncover the characteristics of transitioning cells and shed light on the mechanisms that regulate proximal-distal patterning (iii). (B) The alveolus is the functional unit of gas exchange, and its development involves differentiation of AT1 and AT2 cells and recruitment of endothelial cells. scRNA-seq analyses have revealed novel subpopulations of AT1-associated and interstitial endothelial cells, and shown that epithelial VEGFA is specifically required only for the former (i). Presumptive AT2 cells protrude into the mesenchyme to protect their apical surfaces from high pressure (P) due to inhalation of amniotic fluid; this pressure flattens unprotected cells and predisposes them to become AT1 cells (ii). In a separate mechanism, Mark1-expressing fibroblasts promote local deposition of collagen I to increase AT1 cell flattening (iii).

expression of Fgf10 and to promote differentiation of mesenchymal cells into airway smooth muscle. Smooth muscle differentiation is also regulated by epithelial expression of Wingless-related integration site 7b (Wnt7b) and mesenchymal expression of Wnt2 (Zepp & Morrisey, 2019). Patterned smooth muscle differentiation is required for bifurcation and domain branching of the airway epithelium; in cultured explants, inhibiting smooth muscle differentiation causes epithelial buds to dilate or buckle whereas promoting differentiation causes ectopic smooth muscle wrapping that impedes branching (Goodwin et al., 2019; Kim et al., 2015).

While some of the molecular requirements for airway smooth muscle differentiation have been identified, its origin and the signals that initiate its spatiotemporally patterned differentiation have yet to be elucidated. Lineage-tracing approaches have shown that Wnt- or Shh-responsive mesenchymal cells contribute (but not completely or exclusively) to the smooth muscle population, and that Ffg10-expressing mesenchymal cells can contribute to a lesser extent (Moiseenko et al., 2017). To search for the origin of airway smooth muscle without the need for specific lineage markers, mesenchymal cell identities in the early stages of lung development could be explored using scRNA-seq. Clustering of mesenchymal cells would reveal whether distinct subpopulations exist within the embryonic pulmonary mesenchyme and if any are more closely related to smooth muscle based on their gene-expression profiles (Fig. 4Ai). Further, at any given stage of branching, mesenchymal cells at each step of the smooth muscle differentiation process are presumably present. Therefore, at a single time point, one could computationally order cells along a differentiation trajectory using diffusion-based approaches. Tracing backwards along this trajectory would reveal the origin of smooth muscle cells. By examining gene-expression patterns along the trajectory, one could identify the signaling pathways that are activated at each stage of differentiation (Fig. 4Aii).

The broad inductive signals required for epithelial branching have been identified, and several studies have suggested that morphogen gradients generated in the mesenchyme drive collective migration or chemotaxis of the epithelium, in the way that the Fgf homolog *breathless* drives protrusive collective invasion of the thin tubules of the tracheae in *Drosophila* embryogenesis (Metzger & Krasnow, 1999). However, the tissue architecture and physical mechanisms of branching in *Drosophila* and in mammalian lungs are vastly different (Spurlin & Nelson, 2017), and genetic experiments have demonstrated that the focal sources of a morphogen, specifically Fgf10, are not required for branching of the lung epithelium (Volckaert et al., 2013).

scRNA-seq could be used to identify whether specific morphogen-producing cell types exist in the pulmonary mesenchyme, and provide us with additional markers of such populations that could be used to isolate or manipulate morphogen-producing cells.

It is unclear whether current tools in the scRNA-seq field can be used to study the physical mechanisms of branching morphogenesis. Specifically, we wonder whether cell transcriptomes can be used to distinguish between the proposed modes of epithelial branching in the lung: collective migration and tissue folding. Many of the molecular players broadly implicated in each process are similar, such as regulators of the actin cytoskeleton and of cell-matrix and cell-cell adhesions. To identify the physical processes occurring in airway epithelial cells based on their transcriptomes, we would need more specific classifications of these molecules (e.g., actin regulators that drive cell migration and those that may be activated in response to tissue folding and compression) and an understanding of which cellular transcriptomes could be hallmarks of these different morphogenetic movements. Perhaps a more fruitful endeavor for scRNA-seq studies of the airway epithelium could be to uncover how the proximal and distal cell types are patterned. Epithelial cell identities are specified as branching morphogenesis proceeds; cells in branch tips remain Sox9-positive while those that end up in branch stalks become Sox2-positive and eventually give rise to proximal cell types (Zepp & Morrisey, 2019). By comparing single-cell transcriptomes in Sox2- and Sox9-expressing cells, and in particular in those cells that seem to be transitioning from Sox9 to Sox2 expression as they end up in branch stalks, the signals that establish these cell identities could be uncovered (Fig. 4Aiii).

The pseudoglandular stage of lung development is followed by the canalicular and saccular stages, during which the terminal ends of branches are remodeled in preparation for alveologenesis (Zepp & Morrisey, 2019). Branching morphogenesis continues to some extent into these later stages to generate the future gas-exchange regions of the lung (Alanis, Chang, Akiyama, Krasnow, & Chen, 2014). During these stages, there is widespread differentiation of airway epithelial cells; for proximal cell types, evidence of differentiation is detectable during the pseudoglandular stage as well (Zepp & Morrisey, 2019). Proximal cells differentiate into basal, ciliated, and secretory cells, among others, and distal cells give rise to gas-exchanging AT1 and surfactant-producing AT2 epithelial cells. Initially, it was thought that a subset of airway epithelial cells that gave rise to AT2 cells would subsequently differentiate into AT1 cells based on lineage tracing of cells expressing the distal marker Id2 (Rawlins, Clark, Xue, & Hogan, 2009).

This was eventually disproven, and mapping the temporal changes in expression of several markers revealed that AT1 and AT2 cells arise separately from a bipotent progenitor that expresses markers of both cell types (Desai, Brownfield, & Krasnow, 2014).

More recently, these transitions in cell identities have been investigated using scRNA-seq, and some surprising and conflicting results have been reported. The first of these studies isolated airway epithelial cells from *E*18.5 embryos and identified AT1 cells, AT2 cells, and an intermediate cell type that the authors recognized could represent the bipotent progenitor identified experimentally (Treutlein et al., 2014). These data were then used to computationally reconstruct the changes in gene expression that occur as cells transition from a bipotent state to either alveolar cell type (Treutlein et al., 2014). A second study included lineage tracing of specific AT1 and AT2 markers from early stages in development and scRNA-seq analysis of airway epithelial cell clusters and transitions (Frank et al., 2019). The authors found that, as early as *E*15.5, lineage-labeled cells predominantly gave rise to clones of either AT1 or AT2 cells, but rarely of both cell types (Frank et al., 2019), suggesting that bipotent progenitors are rare even several days before AT1 and AT2 cells actually appear.

Lineage-labeling studies are limited to tracing cells that have expressed a single gene at some point in time and, traditionally, cell types have been named or defined by a specific marker gene that they express. However, scRNA-seq studies have demonstrated that cell identities may be better described by entire sets of marker genes, and the same study showed that this is also true for AT1 and AT2 cells and their progenitors. Using scRNA-seq at *E*17.5, alveolar cell populations were identified computationally, revealing cell-type specific sets of marker genes, and pseudotime analysis was used to reconstruct the diverging differentiation trajectories of AT1 and AT2 cells (Frank et al., 2019). However, the expression of these marker genes was never quite unique to a given cell type or its precursor, and precursor cell types overlapped significantly in low-dimensional representations of the data, suggesting that, transcriptionally, these cell types may not be as distinct as previously believed.

Alveologenesis requires the differentiation of many different cell types, including myofibroblasts and endothelial cells. Myofibroblasts differentiate to form a mesh around epithelial tips, depositing elastin and contracting to drive alveolar septation (Branchfield et al., 2016). Meanwhile, vasculature develops around alveoli to bring blood vessels into close apposition with gas-exchanging AT1 cells (Zepp & Morrisey, 2019). The cellular networks that

govern these fate decisions are being elucidated by combining genetic knockout models with scRNA-seq. For example, AT1 cells are the main cell type that expresses vascular endothelial growth factor A (VEGFA) in the lung, and knocking out VEGFA in AT1 cells specifically or in the entire airway epithelium leads to loss of a specific subset of endothelial cells and disrupts alveolar morphogenesis (Vila Ellis et al., 2020). scRNA-seq analysis of sorted endothelial cells revealed that they are a transcriptionally heterogeneous population comprised of two main cell types: one that associates very closely with AT1 cells and one that is further from AT1 cells in the interstitial space between alveoli (Fig. 4Bi). Importantly, the authors were able to show using scRNA-seq that these two populations are affected differently by epithelial-specific knockout of VEGFA: the former is completely lost, while the latter is preserved. Ablating this AT1 cell-associated endothelial population impaired alveolar septation, demonstrating that a local cellular network can have important consequences for the overall development of an organ.

Frequently, scRNA-seq studies are performed on cells isolated by fluorescence-activated cell sorting (FACS) based on the expression of a fluorescently tagged marker or of a fluorescent reporter under the control of a tissue-specific marker. This approach can enrich a sample for a cell type of interest, but also eliminates the possibility for studying interactions between different tissues and cell types. In the example of sacculation in the mouse lung, interactions between distal airway epithelial cells and their surrounding mesenchyme have been shown to influence AT1 or AT2 fate decisions and morphogenesis (Fumoto et al., 2019; Li et al., 2018). Inhalation of amniotic fluid during fetal breathing movements expands the distal airways and contributes to sacculation (Fig. 4Bii). At these stages, some of the undifferentiated airway epithelial cells protrude into the mesenchyme and accumulate apical myosin, thus fortifying themselves against the influx of fluid. These cells are thus protected from flattening and differentiate into AT2 cells (Li et al., 2018). The surrounding mesenchymal cells also influence AT1 differentiation via different mechanisms: Mark1-expressing fibroblasts promote epithelial cell flattening through Hedgehog-dependent deposition of type I collagen (Fig. 4Biii) (Fumoto et al., 2019).

scRNA-seq analyses that incorporate both epithelial and mesenchymal cell transcriptomes could shed light on the cellular networks that mediate these processes. Specifically, scRNA-seq could be used to identify the mesenchymal signaling source that induces only a subset of epithelial cells to migrate outwards, predisposing them to an AT2 fate (Fig. 4Bii), or to elucidate the characteristics of these particular epithelial cells that allow them to

respond to and migrate towards this mesenchymal signal. Further, using single-cell transcriptomes, the epithelial-mesenchymal signaling networks that mediate local deposition of type I collagen and cell flattening could be elucidated (Fig. 4Biii). Finally, the experimental tools developed to study the effects of fluid pressure on AT1 and AT2 cell differentiation (Li et al., 2018) could also be combined with scRNA-seq to understand how mechanical forces influence the cell transcriptome in vivo.

Studies of lung development using single-cell transcriptomics have already led to the formation of databases, including LungGENS (Du et al., 2017; Du, Guo, Whitsett, & Xu, 2015), and the development of a variety of computational tools that could be applied to other systems (Guo, Bao, Wagner, Whitsett, & Xu, 2017; Guo, Wang, Potter, Whitsett, & Xu, 2015). Parallel single-cell analyses of lung development in vivo and in organoid culture have been used to identify airway cell types and the signals that control their differentiation and, importantly, to quantitatively compare cell identities that arise within organoids to those in vivo (Miller et al., 2020). Finally, recent work has highlighted the usefulness of scRNA-seq for the investigation of emerging model organisms. For example, Modepalli et al used scRNA-seq to study alveologenesis in a marsupial model, the grey short-tailed opossum *Monodelphis domestica* (Modepalli et al., 2018). scRNA-seq of less-studied animal models that may be costly or time-consuming to breed and isolate samples from provides us with a rich dataset that can be compared to the wealth of information about mouse development, for example, and that can be used to determine whether similar signaling pathways and morphogens are involved.

3.2 Kidney development

Kidney branching morphogenesis proceeds from *E*10.5-11 to *E*15.5 (Costantini & Kopan, 2010; McMahon, 2016; Short & Smyth, 2016). Branching is carried out by the ureteric bud epithelium, which evaginates from the nephric cord (of mesodermal origin) into the surrounding metanephric mesenchyme. As branching proceeds, the metanephric mesenchyme surrounding branch tips, called the cap mesenchyme, provides molecular signals important for ureteric bud growth and branching, including glial cell-derived neurotrophic factor (GDNF). Additionally, a subset of cap mesenchymal cells circles back to the "armpit" regions of bifurcated buds and differentiates into cells that give rise to nephrons, the functional units of the kidney (Fig. 5A). Nephron maturation occurs via intermediate structures called renal vesicles, comma-shaped bodies, and S-shaped bodies.

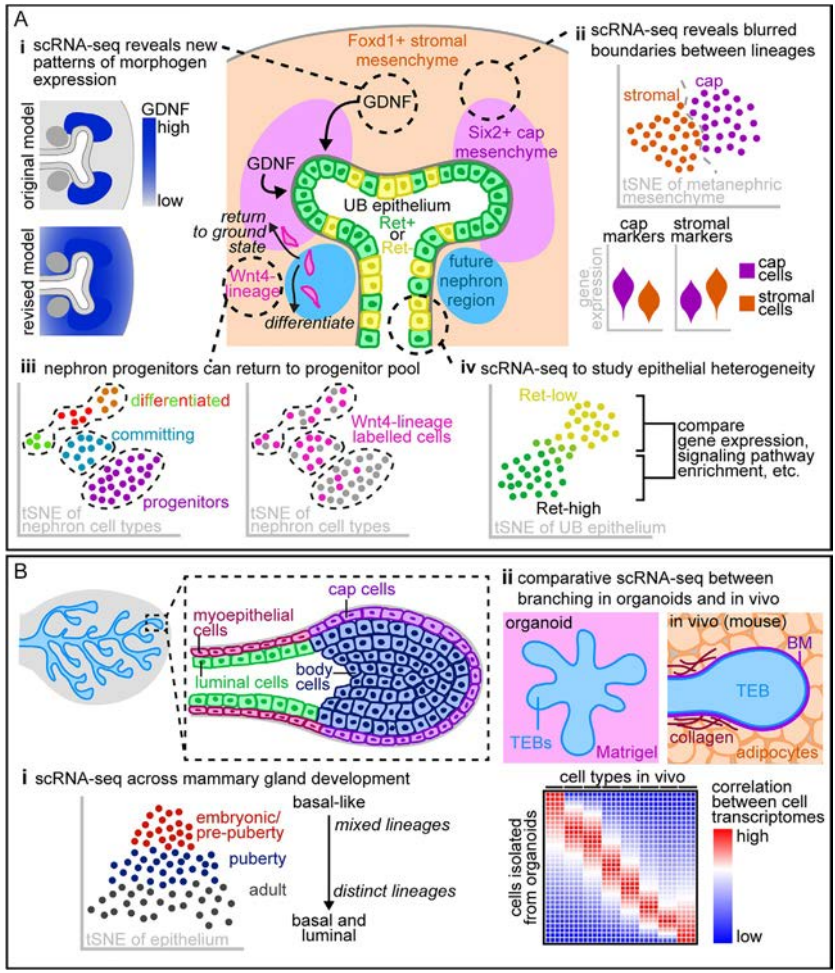

Fig. 5 scRNA-seq to study cellular networks and branching morphogenesis of the mammalian kidney and mammary gland. (A) During kidney branching morphogenesis, GDNF from the metanephric mesenchyme stimulates growth and bifurcation of the ureteric bud (UB) epithelium, and nephron progenitors from the cap mesenchyme initiate nephrogenesis in the armpits of epithelial bifurcations. scRNA-seq analyses have revealed new tissue-specific expression patterns of GDNF (i), and suggest that the lineage boundaries between mesenchymal compartments may not be as binary as previously thought (ii). Indeed, imaging and scRNA-seq have shown that Wnt4-lineage-labeled nephron progenitors can either differentiate or return to the progenitor "ground state" (iii). Cells without Ret are progressively excluded from branch tips—scRNA-seq analyses could be used to study epithelial heterogeneity in the UB to uncover mechanisms of kidney branching morphogenesis (iv). (B) Overview of mammary gland architecture and cell types in the terminal end bud (TEB). scRNA-seq analyses of mammary gland development have revealed a gradual diversification of epithelial cell identities, beginning with more homogeneous, basal-like, mixed-lineage populations and progressing towards distinct luminal and basal cell types (i). To gain insights into mammary gland morphogenesis, which is often studied using organoid models, scRNA-seq studies could be used to compare cell identities and gene expression within organoids to those in vivo (ii).

In contrast to the lung, the early stages of kidney development have already been interrogated using scRNA-seq (Brunskill et al., 2014). These types of analyses are particularly challenging because, at very early stages of organogenesis, tissues are in some ways more homogeneous (there are fewer fully differentiated cells and many uncommitted progenitors) while in other ways far more heterogeneous (undifferentiated cells tend to express a larger suite of genes (Gulati et al., 2020)). Indeed, these analyses revealed that at *E*11.5 uncommitted metanephric mesenchymal cells express markers for several differentiated cell types, a phenomenon termed "multi-lineage priming" (Brunskill et al., 2014). In line with this, induction of cap mesenchyme identity was found to involve not only activation of cap mesenchyme-specific genes but also repression of other genes. The primed transcriptional state of uncommitted progenitors could dictate the dynamics of their responses to external cues provided by cellular networks in the developing kidney.

Looking slightly later in kidney development, scRNA-seq analyses of *E*14.5 kidneys have been used to map cellular interactions by comparing ligands and receptors enriched in each cell population (Magella et al., 2018). This approach revealed expected interactions between Slit2 and Robo expressed in the ureteric duct epithelium and cap mesenchyme, respectively, but also uncovered some new interactions and surprising expression patterns. The analysis carried out by Magella et al predicted interactions between more mature nephron intermediates (comma and S-shaped bodies) and endothelial cells via R-spondin and Lgr signaling. They also found that the master regulator of kidney development, GDNF, was expressed not only in the cap mesenchyme but also in the surrounding stroma (Fig. 5Ai). GDNF was previously thought to be expressed solely in the cap mesenchyme (McMahon, 2016) and to act analogously to focal sources of Fgf10 in the lung (Metzger & Krasnow, 1999). These scRNA-seq results challenge this model and force us to re-examine the roles of morphogen-expression patterns in branching morphogenesis.

Similar to lineage-tracing of AT1 and AT2 cells in the mouse lung, lineage-tracing of the metanephric mesenchyme has long pointed to a binary model in which Six2-positive cells exclusively give rise to nephron progenitors and Foxd1-positive cells exclusively give rise to stromal cells. In the age of scRNA-seq analyses, we are no longer limited to single lineage markers and can compare whole transcriptomes for many cells. These analyses reveal that the boundaries between lineages may not be as binary as previously thought (Fig. 5Aii). Gene-expression profiles of distinct lineages overlap,

and low-dimensional representations of data suggest overlap between presumptively distinct progenitors in the metanephric mesenchyme (Brunskill et al., 2014; Magella et al., 2018) and in the airway epithelium (Frank et al., 2019). The degree and functional significance of these overlaps have yet to be clearly defined, and their interpretation thus far seems to be at the discretion of the investigators—in some cases the overlap is emphasized (Magella et al., 2018) and in others it is ignored (Frank et al., 2019).

Further support for the plasticity of cell differentiation in the developing kidney comes from combined lineage-tracing and scRNA-seq analysis (analogous to the pulse-seq experiments discussed in Section 2.3) of the *E*15.5 kidney (Lawlor et al., 2019). Motivated by the observation that lineage-traced Wnt4-expressing nephron progenitors (thought to be fully committed to the nephron fate) could re-enter the cap mesenchyme during time-lapse imaging of kidney development, Lawlor et al. carried out scRNA-seq and mapped lineage label expression onto the computationally identified cell clusters. They detected lineage-labeled cells throughout mesenchymal clusters, including in those corresponding to progenitors. Wnt4-lineage-traced cells in these clusters had transcriptomes indistinguishable from other progenitors, which the authors interpreted as a "return to ground" state (Fig. 5Aiii).

The majority of scRNA-seq analyses of embryonic kidney development seem to have focused on differentiation of the metanephric mesenchyme. We believe such analyses could also be used to study branching morphogenesis of the ureteric bud epithelium. While many molecular players have been identified, the physical mechanisms of branching in the kidney are still unknown (Costantini & Kopan, 2010; Goodwin & Nelson, 2020; Short & Smyth, 2016). However, mosaic genetic approaches have uncovered heterogeneity among epithelial cells of the ureteric bud that dictates whether these cells remain in branch tips or become excluded to stalks (Riccio, Cebrian, Zong, Hippenmeyer, & Costantini, 2016). Cells lacking Ret, the GDNF receptor, are less competitive than their wildtype counterparts and are progressively excluded from branch tips as branching morphogenesis proceeds. Whether there is a role for endogenous heterogeneity in GDNF-responsiveness among epithelial cells during kidney branching is still unknown, but one could begin to investigate this by comparing the expression of Ret or other receptors in scRNA-seq data. Perhaps by defining an endogenous population of Ret-low cells, one could examine the rest of the cells' transcriptomes to uncover clues as to how and why these cells behave differently and whether these behaviors might have consequences for overall tissue morphogenesis (Fig. 5Aiv).

Finally, scRNA-seq analyses are especially useful for the study of human development because samples are more difficult to procure and as much information as possible should be extracted from each one. Single-cell approaches are being used to identify the signaling pathways that have been extensively studied in animal models and that may also play a role in human kidney development, and to compare the various cell populations that make up the human embryonic kidney to those present in the mouse (Menon et al., 2018; Wang et al., 2018).

3.3 Mammary gland development

Morphogenesis of the mammary gland occurs in multiple phases and spans many different stages of the life of a mammal (Gjorevski & Nelson, 2011; Howard & Lu, 2014; Huebner & Ewald, 2014). At *E*12 in mouse embryos, the mammary epidermal placode invaginates into primary mammary mesenchyme. By *E*15-*E*16, this initial mass of epithelial cells forms a rudimentary branched network that extends into the secondary mammary fat pad mesenchyme. After birth, this network hollows out to form ducts and continues to undergo a limited amount of branching until puberty (Howard & Lu, 2014). During puberty, branching morphogenesis of the mammary epithelium constructs an extensive tree-like structure that permeates the entire adipose-rich stroma of the fat pad. Branch extension in the mouse mammary gland occurs as terminal end buds (TEBs), multi-layered, loosely-connected agglomerations of cells, propel themselves forward into the stroma, leaving behind ducts that resolve into bi-layered tubes (Fig. 5B). The final phase of mammary gland development occurs cyclically during pregnancy and lactation, when branch tips form lobulo-alveolar units in the mouse (terminal ductal lobular units in the human) and their constituent epithelial cells differentiate to produce milk.

scRNA-seq analyses have been carried out at various stages of mammary gland development. For example, some studies have focused on how mammary epithelial cell identities and gene expression evolve over the lifetime of a mouse from prepuberty to pregnancy (Pal et al., 2017) or from embryonic stages to adulthood (Giraddi et al., 2018). These analyses revealed that the embryonic and pre-pubertal epithelium are comprised of fairly homogeneous cell populations that are characterized by the expression of basal cell markers. Over the course of development, the epithelium diversifies into basal and luminal cell populations (Fig. 5Bi). Similar to the metanephric mesenchyme (Brunskill et al., 2014), a subset of epithelial cells isolated in

these scRNA-seq analyses have mixed expression of lineage markers (Giraddi et al., 2018; Pal et al., 2017). Additionally, epithelial cells at *E*18 have accessible chromatin near genes of both basal and luminal lineages (Giraddi et al., 2018). Overall, these scRNA-seq studies suggest that the epithelium of the mammary gland initially contains a continuum of cells without clear lineage markers, and provide further evidence that lineage boundaries may be blurrier than classically believed.

In many contexts, scRNA-seq has been used to identify distinct progenitor or stem cells (Kumar, Tan, & Cahan, 2017). This is not always possible in tissues undergoing morphogenesis, where a spectrum of cells is present instead of a distinct progenitor. To search for mammary stem cells, scRNA-seq analyses were carried out with ductal and TEB cells of pubertal mammary glands (Scheele et al., 2017). However, computationally clustering epithelial cells revealed no distinct subpopulation of mammary stem cells. Instead, the cells analyzed had highly heterogeneous gene-expression profiles and only loosely clustered into their respective lineages. The authors therefore suggested that most of the cells within TEBs are mammary stem cells. In line with this, other scRNA-seq studies have referred to the heterogeneous population present at early developmental stages (*E*18) as functional mammary stem cells (Giraddi et al., 2018). These findings shed light on the nature of developing tissues, which appear to be initially comprised almost entirely of "stem cells," evidenced by overlapping cell clusters and mixed-lineage gene-expression profiles and chromatin-accessibility landscapes. As development progresses, these cells give rise to various differentiated populations, leading to diversification of cell types and emergence of more specific gene-expression profiles.

Lineage-tracing studies had shown that by postnatal day 1, mammary epithelial cells are unipotent, suggesting that they were already lineage-committed by the end of embryogenesis (Wuidart et al., 2018). It has been argued that the lack of distinct luminal and basal progenitors in scRNA-seq datasets of early mammary gland development could be due to limitations of the scRNA-seq method itself. Transcripts of genes expressed at low levels are often missed entirely, and a cell's identity or state is not necessarily encoded only in its transcriptome. Additional information is encoded in the chromatin landscape, which can be investigated using assay for transposase-accessible chromatin (ATAC)-seq. Recent technological advances have enabled researchers to carry out single-nucleus (sn) ATAC-seq (Preissl et al., 2018). In snATAC-seq, the chromatin landscape can be mapped for individual cells and thus an alternative (or additional, if scRNA-seq and snATAC-seq are

carried out on the same cells (Chen, Lake, & Zhang, 2019)) readout of cell identity can be observed. snATAC-seq of embryonic mammary glands revealed that epithelial cells could be classified more distinctly into basal and luminal clusters based on their open chromatin at distal sites, as compared to classification based on their mRNA content or by their open chromatin at proximal sites (Chung et al., 2019; Giraddi et al., 2018). This heterogeneity in accessible chromatin sites may explain why analyses of bulk ATAC-seq data suggested that epithelial cells have both luminal and basal chromatin landscapes in the early stages of mammary gland development.

After branching morphogenesis, the adult mammary gland continues to elaborate and specialize during pregnancy and lactation (Inman, Robertson, Mott, & Bissell, 2015). Following lactation, the mammary gland is remodeled, and many epithelial cells undergo apoptosis in a process called involution. Each of these stages has been characterized by scRNA-seq analysis (Bach et al., 2017). Similar to prepubertal and pubertal stages, the cycling adult mammary gland contains a continuum of progenitor and differentiated epithelial cell types within the basal and luminal compartments. Focusing first on pre-pregnancy, Bach et al. used diffusion maps to reconstruct differentiation trajectories: in the resultant map, luminal progenitor cells clustered together at the root, and two branches led to the secretory alveolar lineage and the hormone-sensing lineage. To determine if cell identities revert to their pre-pregnancy state after involution, the authors mapped luminal progenitor cells from the post-pregnancy dataset onto this diffusion map. Post-pregnancy luminal progenitors were spread throughout the secretory alveolar branch, suggesting that luminal progenitor cells remain primed to differentiate. Luminal cells of the post-pregnancy mammary gland have a distinct DNA-methylation signature, particularly at genes expressed during pregnancy (Dos Santos, Dolzhenko, Hodges, Smith, & Hannon, 2015). Overall, these data suggest that post-pregnancy luminal progenitor cells have a unique epigenome and transcriptome that may enhance their response to a second pregnancy and drive robust alveologenesis (Bach et al., 2017; Dos Santos et al., 2015). Using snATAC-seq or related techniques in combination with scRNA-seq, future work could uncover exactly which cells retain an epigenetic memory of pregnancy and whether gene-expression profiles of luminal progenitors reflect epigenetic modifications.

scRNA-seq investigations into development and morphogenesis also have exciting implications for understanding disease mechanisms, as developmental pathways are often coopted in cancer (Aiello & Stanger, 2016). Indeed, comparing developmental trajectories obtained from samples

isolated during mammary epithelial branching morphogenesis to those detected in tumor samples revealed that mammary epithelial carcinoma cells reactivate some of the gene expression programs of mammary epithelial stem cells (Giraddi et al., 2018). Specifically, metabolic signaling pathways favoring glycolysis and disfavoring mitochondrial metabolism and oxidative phosphorylation were highly enriched in both mammary stem cells (*E*18 epithelial cells) and basal-like tumors, but not in luminal-like tumors. Future work could compare single-cell data from tumors to developing mammary glands to determine whether tumor cell heterogeneity recapitulates developmental heterogeneity.

Our current physical models of mammary branching morphogenesis are primarily informed by observations of organoids. Dissociated mammary epithelial cells embedded in Matrigel and treated with growth factors assemble into organoids and undergo robust branching morphogenesis (Huebner & Ewald, 2014). This experimental model is amenable to live-imaging, unlike the optically unfavorable, adipocyte-rich fat pad. In mammary epithelial organoids, branch extension occurs by collective, noninvasive migration of TEB-like structures (Ewald, Brenot, Duong, Chan, & Werb, 2008), and myoepithelial cells provide hoop stresses around ducts to allow for TEB propulsion (Neumann et al., 2018). Whether similar mechanisms drive branching in vivo or in human mammary glands, which are surrounded by a very different, fibroblast-rich stroma, is unknown (Goodwin & Nelson, 2020). Comparative scRNA-seq analyses of mammary glands in vivo and in organoids, similar to that carried out in studies of human embryonic lung organoids and tissue samples (Miller et al., 2020), could help determine how closely organoids recapitulate morphogenesis in vivo, and point the way towards identifying key components of the native microenvironment that regulate mammary epithelial cell gene expression (Fig. 5Bii).

Most scRNA-seq studies of the mammary gland have been carried out in sorted cell populations to enrich for epithelial cells based on specific cell-surface markers. Since scRNA-seq data can be readily sorted computationally, physically sorting may not always be necessary. Presorting could bias the cells analyzed, and excludes supporting cell types, some of which may have surprising roles. For example, recent work identified a population of macrophages nestled between the luminal and myoepithelial cells of the ducts that support mammary gland homeostasis and remodeling (Dawson et al., 2020). In scRNA-seq studies of perturbations to the mammary gland, sequencing all cells would uncover effects on the network of epithelial cells and macrophages instead of obfuscating potentially important effects on other cell types by sequencing only sorted epithelial cells.

3.4 Development of other branched organs

Finally, it is important to describe a few scRNA-seq studies of branching morphogenesis in organs not discussed above, including the salivary gland, pancreas, and prostate. The literature on these organs is currently less extensive, but scRNA-seq analyses have already produced insights into the development of and cell-type heterogeneity within these tissues.

scRNA-seq analyses have been carried out in the embryonic and postnatal salivary gland (Sekiguchi et al., 2019; Song et al., 2018). These studies have shown that at *E*12 the salivary gland mesenchyme is more heterogeneous than the epithelium, and that gland-specific mesenchymal clusters from the salivary and parotid glands emerge during clustering analyses (Sekiguchi et al., 2019). Postnatally, there is increased epithelial diversity and acinar, myoepithelial, and basal lineages can be clearly demarcated (Song et al., 2018). Multi-stage data or data from later stages in the developing organ could be used to computationally reconstruct differentiation trajectories into each of these lineages. The physical mechanisms of salivary gland branching morphogenesis have been extensively investigated and are better understood than the mechanisms of branching in other commonly-studied organs (Goodwin & Nelson, 2020). Given the recent publication of embryonic salivary gland scRNA-seq data (Sekiguchi et al., 2019), there is an exciting opportunity to compare our current physical models of branching morphogenesis with cell transcriptomes. For example, outer cells of the buds of developing salivary glands are more motile than inner cells and protrude through their basement membrane to contact the surrounding mesenchyme (Daley et al., 2017; Harunaga, Doyle, & Yamada, 2014). If these cells could be identified within scRNA-seq data, we could learn more about the genes that regulate these behaviors and, conversely, we could begin to define features of a cell transcriptome indicative of specific cellular activities important for morphogenesis.

Development of the pancreas between *E*9.5 and *E*17.5 was studied using scRNA-seq analyses of cells isolated and sorted from embryos expressing various reporters of pancreatic progenitors and differentiated cell types (Yu et al., 2019). This experimental design comes with rich metadata (developmental stage, lineage label) that can be overlaid on computationally-identified clusters and reconstructed differentiation trajectories to verify their biological relevance. Focusing on specific lineage-restriction decisions, the authors used Monocle2 analyses to study differentiation within tips, ducts, endocrine progenitors, and α- and β-cell lineages separately. By isolating each lineage decision, they were able to move step-wise throughout pancreatic development

and identify key signaling pathways at each stage in the morphogenesis of this complex organ. Ductal cell differentiation in the human pancreas has also been investigated using scRNA-seq (Qadir et al., 2020). Similar to developing embryonic tissues, ductal cells within the adult pancreas exist in a continuum of states from progenitor to fully differentiated.

To our knowledge, no scRNA-seq studies of embryonic prostate development have been carried out, although some have examined postnatal (Boufaied et al., 2017) and adult (Kwon et al., 2019) stages. When such datasets are eventually generated, it will be valuable to compare cell types and transitions in the developing prostate mesenchyme to those in the pulmonary mesenchyme. Prostate branching morphogenesis may be influenced by smooth muscle differentiation from the surrounding mesenchyme (Toivanen & Shen, 2017) in a manner similar to airway branching (Goodwin & Nelson, 2020). Comparative scRNA-seq analyses could determine whether similar pathways regulate mesenchymal patterning and smooth muscle cell recruitment in each organ.

4. Using scRNA-seq to determine how physical signals dictate cell identity

Thus far, scRNA-seq has revealed that branching tissues are comprised of many different cell types, and that they typically begin as more homogeneous populations from which distinct progenitors cannot always be discerned. As branching morphogenesis and differentiation proceed, scRNA-seq has allowed us to observe lineage choices and differentiation trajectories. However, branching morphogenesis is more than just differentiation into cell types; to build a branched organ, tissues also need to expand, elongate, fold, push, and squeeze to generate the final architecture necessary for organ function. In this final section, we speculate about whether we can use a cell's transcriptome to understand the physical aspects of branching morphogenesis, and whether scRNA-seq analyses could shed light on the physical signals that cells receive during morphogenesis.

One approach to address this question would be to begin with a simpler morphogenetic event, in which the forces and cell-shape changes are well-defined, and compare our understanding of the physical aspects of the system with single-cell transcriptomes. Recently, scRNA-seq analyses were carried out on the *Drosophila* wing imaginal disc, a model system that has been extensively used to study the coupling between growth, patterning, and tissue mechanics (Deng et al., 2019). First, the authors identified sets of

coregulated genes that were related to proliferation and growth; one set contained genes related to DNA replication, and the other contained genes related to protein translation. Then, the authors clustered cells based on their expression of these gene sets into four classes. In wild-type wing discs, cells of each class were found in each anatomically distinct region, as expected, but in mutants for the tumor suppressor *scribble*, the majority of cells were members of classes with enhanced expression of either protein translation genes or both DNA replication and protein translation genes. This analysis showed that the growth state of a cell could be effectively inferred from its transcriptome.

The mechanical state of a cell is the sum of many different factors, and can be characterized in part based on the expression of proteins within specific networks including the adhesome, contractome, or mechanobiome (Horton et al., 2016; Kothari, Johnson, Sandone, Iglesias, & Robinson, 2019; Zaidel-Bar, Zhenhuan, & Luxenburg, 2015). Mechanotransduction causes many immediate changes in protein conformation, binding, recruitment, etc., and, over longer time scales, in gene expression. Transcription factors that respond to mechanical signals regulate diverse aspects of cell behavior including proliferation, differentiation, metabolism, cytoskeletal contraction, cell-cell and cell-matrix binding, and chromatin remodeling. These responses can occur directly downstream of changes in gene expression, but some are independent of transcription and are instead regulated at the protein level. Therefore, scRNA-seq may not provide a sufficient readout of cell mechanical state. Other single-cell, big data-style techniques would need to be used to probe other aspects of the mechanobiome, such as single-cell proteomics (Zhu et al., 2019).

Bulk RNA-seq has been used to investigate cellular responses to microenvironmental stiffness in culture models (Moreno-Vicente et al., 2018); in such systems, cell heterogeneity is arguably much lower, so bulk RNA-seq is likely to capture important changes in gene expression in response to stiffness. In studies of more complex organ systems, however, this may not be sufficient. Bulk RNA-seq of lungs under high and low transmural pressure revealed that pressure enhances branching morphogenesis and causes differential expression of many different genes (Nelson et al., 2017). Single-cell-level information in this system would shed light on how each cell population responds to transmural pressure, and could elucidate whether each tissue responds directly to physical pressure or indirectly via pressure-dependent signaling in a neighboring tissue (Fig. 6A).

scRNA-seq analyses have not been extensively used to study tissue mechanics, but some recent exciting studies have either begun to or have

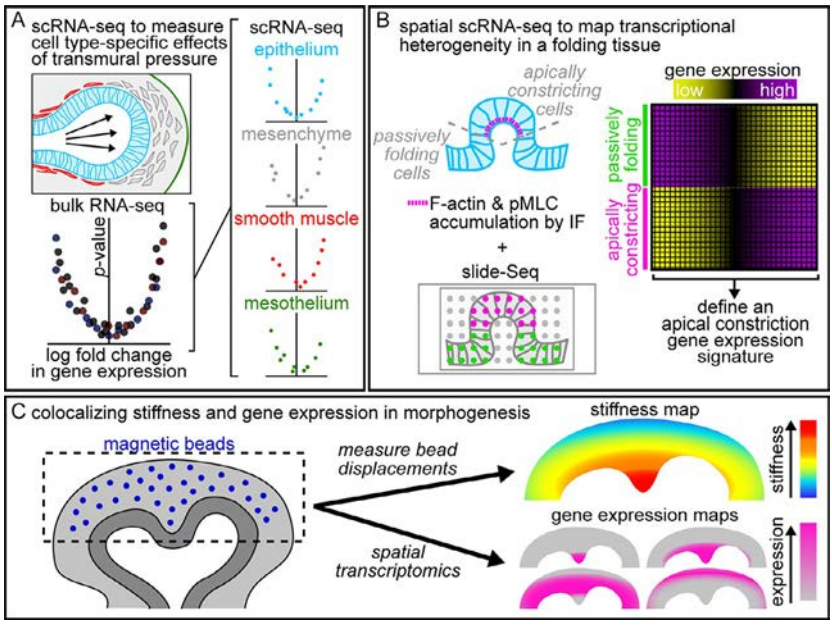

Fig. 6 Using scRNA-seq to study physical and mechanical aspects of morphogenesis. (A) Transmural pressure stimulates embryonic mouse lung development and leads to many changes in gene expression as determined by bulk RNA seq. With scRNA-seq, cell-type-specific effects of pressure could be measured, and the bulk effects could be disentangled to determine which tissues respond directly to pressure and how. (B) Serial sections of a folding tissue could be used for immunofluorescence analysis (for example, to visualize F-actin or pMLC and infer regions of active constriction) and for slide-seq to define a gene-expression profile unique to constricting cells compared to passively folding cells of the same cell type. (C) Biophysical techniques for measuring local tissue stiffness, for example embedding magnetic beads in the mesenchyme around a bifurcating airway, could be used in parallel with spatial transcriptomics to generate and compare stiffness maps with gene-expression maps.

already generated datasets that could be used to explore this area. For example, Abbas et al. carried out a comparative study of the decidua, placenta, and endometrium, all tissues at the maternal-fetal interface. For each tissue, they measured stiffness using atomic force microscopy and profiled gene expression using scRNA-seq (Abbas et al., 2019). This dataset was used to determine which extracellular matrix (ECM) components were expressed by each tissue type and could be used to relate cell transcriptomes to tissue stiffness. scRNA-seq datasets of lung epithelial cells in healthy and fibrotic tissue (Xu et al., 2016) or in lung tissue after pneumonectomy (Wu et al., 2020) could be used to study the response of lung epithelial cells to an altered

mechanical microenvironment. Indeed, the latter study identified a role for mechanical tension in regulating harmful signaling in AT2 cells during pulmonary fibrosis (Wu et al., 2020).

Finally, we propose that spatial scRNA-seq approaches such as those discussed in Section 2.3 could be used in combination with readouts of local tissue shape and mechanics to directly compare single-cell transcriptomics to tissue mechanics. For example, simple staining for F-actin or phosphorylated myosin light chain (pMLC) in serial sections combined with slide-seq would reveal local transcriptional signatures associated with cytoskeletal tension (Fig. 6B). More direct measures of cellular mechanics such as Förster resonance energy transfer (FRET)-based tension sensors (Tao et al., 2019), magnetic bead displacements (Zhu et al., 2020), and oil-drop deformation (Mongera et al., 2018) are now being applied to increasingly complex morphogenetic processes. Conceivably, one could map tissue mechanics with these approaches followed by spatial transcriptomics to understand how endogenous forces regulate cell identity (Fig. 6C).

5. Conclusions and outlook

Single-cell transcriptomics serve as a powerful approach to map the cell-intrinsic characteristics and changes that define cell types and differentiation trajectories. However, it is still challenging to uncover the external factors and cellular networks that influence the cell transcriptome. Preserving spatial information and inclusive sequencing (as opposed to sequencing only sorted cells) will be instrumental in identifying these networks, as will new techniques to preserve connections between cells, such as PIC-seq, an approach for sequencing pairs of physically interacting cells (Giladi et al., 2020).

Transcriptomic characterizations of progenitor cells within developing or adult organs, including several examples discussed here, have consistently revealed that lineage boundaries are blurrier than previously thought, and that distinct progenitors cannot always be discerned. While this may reflect the underlying biology, it might also be a result of insufficient resolution. Indeed, one of the studies discussed here showed that snATAC-seq could resolve lineages where scRNA-seq could not (Chung et al., 2019). Incorporating lineage-tracing and barcoding technologies into single-cell transcriptomics experiments is a promising avenue for testing the boundaries between lineages (Wagner & Klein, 2020). Additionally, improvements are being made to scRNA-seq itself to enhance sensitivity, detect more transcripts, and distinguish between isoforms (Hagemann-Jensen et al., 2020).

With these newer technologies, we will either resolve progenitors or confirm that progenitor identity is not always so distinct.

scRNA-seq analyses have already shed light on differentiation within lineages of branched organs, but they have not yet been used to understand the physical aspects of branching morphogenesis. We hope that, as single-cell technologies and applications continue to develop, we will learn more about how mechanical and physical signals influence the cell transcriptome, and eventually devise methods of extracting physical or mechanical information from single-cell transcriptomics.

References

Abbas, Y., Carnicer-Lombarte, A., Gardner, L., Thomas, J., Brosens, J. J., Moffett, A., et al. (2019). Tissue stiffness at the human maternal-fetal interface. *Human Reproduction*, *34*(10), 1999–2008.

Aiello, N. M., & Stanger, B. Z. (2016). Echoes of the embryo: Using the developmental biology toolkit to study cancer. *Disease Models & Mechanisms*, *9*, 105–114.

Alanis, D. M., Chang, D. R., Akiyama, H., Krasnow, M. A., & Chen, J. (2014). Two nested developmental waves demarcate a compartment boundary in the mouse lung. *Nature Communications*, *5*, 3923.

Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J., & van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. *Nature*, *556*, 108–112.

Atmanli, A., Hu, D., Deiman, F. E., van de Vrugt, A. M., Cherbonneau, F., Black, L. D., 3rd, et al. (2019). Multiplex live single-cell transcriptional analysis demarcates cellular functional heterogeneity. *eLife*, *8*.

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D. J., Marioni, J. C., et al. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nature Communications*, *8*, 2128.

Bendall, S. C., Davis, K. L., Amir, A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., et al. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*, *157*, 714–725.

Boufaied, N., Nash, C., Rochette, A., Smith, A., Orr, B., Grace, O. C., et al. (2017). Identification of genes expressed in a mesenchymal subset regulating prostate organogenesis using tissue and single cell transcriptomics. *Scientific Reports*, *7*, 16385.

Branchfield, K., Li, R., Lungova, V., Verheyden, J. M., McCulley, D., & Sun, X. (2016). A three-dimensional study of alveologenesis in mouse lung. *Developmental Biology*, *409*, 429–441.

Brunskill, E. W., Park, J. S., Chung, E., Chen, F., Magella, B., & Potter, S. S. (2014). Single cell dissection of early kidney development: Multilineage priming. *Development*, *141*, 3093–3101.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, *36*, 411–420.

Cao, C., Lemaire, L. A., Wang, W., Yoon, P. H., Choi, Y. A., Parsons, L. R., et al. (2019). Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature*, *571*, 349–354.

Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., et al. (2019). Molecular recording of mammalian embryogenesis. *Nature*, *570*, 77–82.

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*, aaa6090.

Chen, S., Lake, B. B., & Zhang, K. (2019). High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nature Biotechnology*, *37*, 1452–1457.

Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F., & Poidinger, M. (2016). Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nature Communications*, *7*, 11988.

Chung, C. Y., Ma, Z., Dravis, C., Preissl, S., Poirion, O., Luna, G., et al. (2019). Single-cell chromatin analysis of mammary gland development reveals cell-state transcriptional regulators and lineage relationships. *Cell Reports*, *29*. 495–510 e496.

Costantini, F., & Kopan, R. (2010). Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. *Developmental Cell*, *18*, 698–712.

Daley, W. P., Matsumoto, K., Doyle, A. D., Wang, S., DuChez, B. J., Holmbeck, K., et al. (2017). Btbd7 is essential for region-specific epithelial cell dynamics and branching morphogenesis in vivo. *Development*, *144*, 2200–2211.

Dawson, C. A., Pal, B., Vaillant, F., Gandolfo, L. C., Liu, Z., Bleriot, C., et al. (2020). Tissue-resident ductal macrophages survey the mammary epithelium and facilitate tissue remodelling. *Nature Cell Biology*, *22*, 546–558.

Deng, M., Wang, Y., Zhang, L., Yang, Y., Huang, S., Wang, J., et al. (2019). Single cell transcriptomic landscapes of pattern formation, proliferation and growth in Drosophila wing imaginal discs. *Development*, *146*.

Desai, T. J., Brownfield, D. G., & Krasnow, M. A. (2014). Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*, *507*, 190–194.

Dos Santos, C. O., Dolzhenko, E., Hodges, E., Smith, A. D., & Hannon, G. J. (2015). An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Reports*, *11*, 1102–1109.

Du, Y., Guo, M., Whitsett, J. A., & Xu, Y. (2015). 'LungGENS': A web-based tool for mapping single-cell gene expression in the developing lung. *Thorax*, *70*, 1092–1094.

Du, Y., Kitzmiller, J. A., Sridharan, A., Perl, A. K., Bridges, J. P., Misra, R. S., 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.

Ewald, A. J., Brenot, A., Duong, M., Chan, B. S., & Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Developmental Cell*, *14*, 570–581.

Frank, D. B., Penkala, I. J., Zepp, J. A., Sivakumar, A., Linares-Saldana, R., Zacharias, W. J., et al. (2019). Early lineage specification defines alveolar epithelial ontogeny in the murine lung. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(10), 4362–4371.

Frieda, K. L., Linton, J. M., Hormoz, S., Choi, J., Chow, K. K., Singer, Z. S., et al. (2017). Synthetic recording and in situ readout of lineage information in single cells. *Nature*, *541*, 107–111.

Fumoto, K., Takigawa-Imamura, H., Sumiyama, K., Yoshimura, S. H., Maehara, N., & Kikuchi, A. (2019). Mark1 regulates distal airspace expansion through type I pneumocyte flattening in lung development. *Journal of Cell Science*, *132*.

Giladi, A., Cohen, M., Medaglia, C., Baran, Y., Li, B., Zada, M., et al. (2020). Dissecting cellular crosstalk by sequencing physically interacting cells. *Nature Biotechnology*, *38*(5), 629–637.

Giraddi, R. R., Chung, C. Y., Heinz, R. E., Balcioglu, O., Novotny, M., Trejo, C. L., et al. (2018). Single-cell transcriptomes distinguish stem cell state changes and lineage specification programs in early mammary gland development. *Cell Reports*, *24*. 1653–1666 e1657.

Gjorevski, N., & Nelson, C. M. (2011). Integrated morphodynamic signalling of the mammary gland. *Nature Reviews. Molecular Cell Biology*, *12*, 581–593.
Goodwin, K., Mao, S., Guyomar, T., Miller, E., Radisky, D. C., Kosmrlj, A., et al. (2019). Smooth muscle differentiation shapes domain branches during mouse lung development. *Development*, *146*.
Goodwin, K., & Nelson, C. M. (2020). Branching morphogenesis. *Development*, *147*.
Gulati, G. S., Sikandar, S. S., Wesche, D. J., Manjunath, A., Bharadwaj, A., Berger, M. J., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. *Science*, *367*, 405–411.
Guo, M., Bao, E. L., Wagner, M., Whitsett, J. A., & Xu, Y. (2017). SLICE: determining cell differentiation and lineage based on single cell entropy. *Nucleic Acids Research*, *45*, e54.
Guo, M., Wang, H., Potter, S. S., Whitsett, J. A., & Xu, Y. (2015). SINCERA: A pipeline for single-cell RNA-Seq profiling analysis. *PLoS Computational Biology*, *11*, e1004575.
Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks, G.-J., Larsson, A. J. M., et al. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology*, *38*(6), 705–714.
Haghverdi, L., Buettner, F., & Theis, F. J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, *31*, 2989–2998.
Harunaga, J. S., Doyle, A. D., & Yamada, K. M. (2014). Local and global dynamics of the basement membrane during branching morphogenesis require protease activity and actomyosin contractility. *Developmental Biology*, *394*, 197–205.
Horton, E. R., Humphries, J. D., James, J., Jones, M. C., Askari, J. A., & Humphries, M. J. (2016). The integrin adhesome network at a glance. *Journal of Cell Science*, *129*, 4159–4163.
Howard, B. A., & Lu, P. (2014). Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Seminars in Cell & Developmental Biology*, *25-26*, 43–51.
Huebner, R. J., & Ewald, A. J. (2014). Cellular foundations of mammary tubulogenesis. *Seminars in Cell & Developmental Biology*, *31*, 124–131.
Inman, J. L., Robertson, C., Mott, J. D., & Bissell, M. J. (2015). Mammary gland development: Cell fate specification, stem cells and the microenvironment. *Development*, *142*, 1028–1042.
Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., et al. (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science*, *361*.
Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., et al. (2017). The Drosophila embryo at single-cell transcriptome resolution. *Science*, *358*, 194–199.
Kebschull, J. M., & Zador, A. M. (2018). Cellular barcoding: Lineage tracing, screening and beyond. *Nature Methods*, *15*, 871–879.
Kim, H. Y., Pang, M. F., Varner, V. D., Kojima, L., Miller, E., Radisky, D. C., et al. (2015). Localized smooth muscle differentiation is essential for epithelial bifurcation during branching morphogenesis of the mammalian lung. *Developmental Cell*, *34*, 719–726.
Kothari, P., Johnson, C., Sandone, C., Iglesias, P. A., & Robinson, D. N. (2019). How the mechanobiome drives cell behavior, viewed through the lens of control theory. *Journal of Cell Science*, *132*.
Kumar, P., Tan, Y., & Cahan, P. (2017). Understanding development and stem cells using single cell-based analyses of gene expression. *Development*, *144*, 17–32.
Kwon, O. J., Zhang, Y., Li, Y., Wei, X., Zhang, L., Chen, R., et al. (2019). Functional heterogeneity of mouse prostate stromal cells revealed by single-cell RNA-Seq. *iScience*, *13*, 328–338.
Lawlor, K. T., Zappia, L., Lefevre, J., Park, J. S., Hamilton, N. A., Oshlack, A., et al. (2019). Nephron progenitor commitment is a stochastic process influenced by cell migration. *Elife*, 8.

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science*, *343*, 1360–1363.

Levsky, J. M., Shenoy, S. M., Pezo, R. C., & Singer, R. H. (2002). Single-cell gene expression profiling. *Science*, *297*, 836–840.

Li, J., Wang, Z., Chu, Q., Jiang, K., Li, J., & Tang, N. (2018). The strength of mechanical forces determines the differentiation of alveolar epithelial cells. *Developmental Cell*, *44*. 297–312 e295.

Lubeck, E., & Cai, L. (2012). Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods*, *9*, 743–748.

Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., & Cai, L. (2014). Single-cell in situ RNA profiling by sequential hybridization. *Nature Methods*, *11*, 360–361.

Magella, B., Adam, M., Potter, A. S., Venkatasubramanian, M., Chetal, K., Hay, S. B., et al. (2018). Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf. *Developmental Biology*, *434*, 36–47.

Marco, E., Karp, R. L., Guo, G., Robson, P., Hart, A. H., Trippa, L., et al. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, E5643–E5650.

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*. e-prints.

McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, *353*, aaf7907.

McMahon, A. P. (2016). Development of the mammalian kidney. *Current Topics in Developmental Biology*, *117*, 31–64.

Menon, R., Otto, E. A., Kokoruda, A., Zhou, J., Zhang, Z., Yoon, E., et al. (2018). Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Development*, *145*.

Metzger, R. J., Klein, O. D., Martin, G. R., & Krasnow, M. A. (2008). The branching programme of mouse lung development. *Nature*, *453*, 745–750.

Metzger, R. J., & Krasnow, M. A. (1999). Genetic control of branching morphogenesis. *Science*, *284*, 1635–1639.

Miller, A. J., Yu, Q., Czerwinski, M., Tsai, Y. H., Conway, R. F., Wu, A., et al. (2020). In vitro and in vivo development of the human airway at single-cell resolution. *Developmental Cell*, *53*. 117–128 e116.

Modepalli, V., Kumar, A., Sharp, J. A., Saunders, N. R., Nicholas, K. R., & Lefevre, C. (2018). Gene expression profiling of postnatal lung development in the marsupial gray short-tailed opossum (Monodelphis domestica) highlights conserved developmental pathways and specific characteristics during lung organogenesis. *BMC Genomics*, *19*, 732.

Moiseenko, A., Kheirollahi, V., Chao, C. M., Ahmadvand, N., Quantius, J., Wilhelm, J., et al. (2017). Origin and characterization of alpha smooth muscle actin-positive cells during murine lung development. *Stem Cells*, *35*, 1566–1578.

Mongera, A., Rowghanian, P., Gustafson, H. J., Shelton, E., Kealhofer, D. A., Carn, E. K., et al. (2018). A fluid-to-solid jamming transition underlies vertebrate body axis elongation. *Nature*, *561*, 401–405.

Montoro, D. T., Haber, A. L., Biton, M., Vinarsky, V., Lin, B., Birket, S. E., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature*, *560*, 319–324.

Moreno-Vicente, R., Pavon, D. M., Martin-Padura, I., Catala-Montoro, M., Diez-Sanchez, A., Quilez-Alvarez, A., et al. (2018). Caveolin-1 modulates mechanotransduction

responses to substrate stiffness through actin-dependent control of YAP. *Cell Reports*, *25*. 1622–1635 e1626.
Nagendran, M., Riordan, D. P., Harbury, P. B., & Desai, T. J. (2018). Automated cell-type classification in intact tissues by single-cell molecular profiling. *Elife*, 7.
Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., et al. (2016). A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, *537*, 57–62.
Nelson, C. M., Gleghorn, J. P., Pang, M. F., Jaslove, J. M., Goodwin, K., Varner, V. D., et al. (2017). Microfluidic chest cavities reveal that transmural pressure controls the rate of lung development. *Development*, *144*, 4328–4335.
Neumann, N. M., Perrone, M. C., Veldhuis, J. H., Huebner, R. J., Zhan, H., Devreotes, P. N., et al. (2018). Coordination of receptor tyrosine kinase signaling and interfacial tension dynamics drives radial intercalation and tube elongation. *Developmental Cell*, *45*. 67–82 e66.
Nowotschin, S., Setty, M., Kuo, Y. Y., Liu, V., Garg, V., Sharma, R., et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature*, *569*, 361–367.
Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., et al. (2019). A lineage-resolved molecular atlas of *C. elegans* embryogenesis at single-cell resolution. *Science*, *365*(6459), eaax1971.
Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A. C., et al. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nature Communications*, *8*, 1627.
Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D. U., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nature Neuroscience*, *21*, 432–439.
Qadir, M. M. F., Alvarez-Cubela, S., Klein, D., van Dijk, J., Muniz-Anquela, R., Moreno-Hernandez, Y. B., et al. (2020). Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 10876–10887.
Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*, *14*, 979–982.
Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., et al. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nature Biotechnology*, *36*, 442–450.
Rawlins, E. L., Clark, C. P., Xue, Y., & Hogan, B. L. (2009). The Id2+ distal tip lung epithelium contains individual multipotent embryonic progenitor cells. *Development*, *136*, 3741–3745.
Riccio, P., Cebrian, C., Zong, H., Hippenmeyer, S., & Costantini, F. (2016). Ret and Etv4 promote directed movements of progenitor cells during renal branching morphogenesis. *PLoS Biology*, *14*, e1002382.
Rodriques, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., et al. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, *363*, 1463–1467.
Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., et al. (2016). The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. *Developmental Cell*, *39*, 169–185.
Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, *33*, 495–502.
Schaum, N., Karkanias, J., Neff, N. F., et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*, 367–372.

Scheele, C. L., Hannezo, E., Muraro, M. J., Zomer, A., Langedijk, N. S., van Oudenaarden, A., et al. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature*, *542*, 313–317.

Sekiguchi, R., Martin, D., Genomics Computational Biology Core, & Yamada, K. M. (2019). Single-cell RNA-seq identifies cell diversity in embryonic salivary glands. *Journal of Dental Research*. 22034519883888.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., & Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology*, *37*, 451–460.

Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., et al. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnology*, *34*, 637–645.

Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., et al. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell*, *166*. 1308–1323 e1330.

Short, K. M., & Smyth, I. M. (2016). The contribution of branching morphogenesis to kidney development and disease. *Nature Reviews. Nephrology*, *12*, 754–767.

Song, E. C., Min, S., Oyelakin, A., Smalley, K., Bard, J. E., Liao, L., et al. (2018). Genetic and scRNA-seq analysis reveals distinct cell populations that contribute to salivary gland development and maintenance. *Scientific Reports*, *8*, 14043.

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N., et al. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nature Biotechnology*, *36*, 469–473.

Spurlin, J. W., 3rd, & Nelson, C. M. (2017). Building branched tissue structures: from single cell guidance to coordinated construction. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 372.

Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., et al. (2018). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*, 477.

Tao, H., Zhu, M., Lau, K., Whitley, O. K. W., Samani, M., Xiao, X., et al. (2019). Oscillatory cortical forces promote three dimensional cell intercalations that shape the murine mandibular arch. *Nature Communications*, *10*, 1703.

Toivanen, R., & Shen, M. M. (2017). Prostate organogenesis: Tissue induction, hormonal regulation and cell type specification. *Development*, *144*, 1382–1398.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*, 381–386.

Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., et al. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature*, *509*, 371–375.

van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Vila Ellis, L., Cain, M. P., Hutchison, V., Flodby, P., Crandall, E. D., Borok, Z., et al. (2020). Epithelial VEGFA specifies a distinct endothelial population in the mouse lung. *Developmental Cell*, *52*. 617–630 e616.

Volckaert, T., Campbell, A., Dill, E., Li, C., Minoo, P., & De Langhe, S. (2013). Localized Fgf10 expression is not required for lung branching morphogenesis but prevents differentiation of epithelial progenitors. *Development*, *140*, 3731–3742.

Wagner, D. E., & Klein, A. M. (2020). Lineage tracing meets single-cell omics: Opportunities and challenges. *Nature Reviews. Genetics*.

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G., & Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science*, *360*, 981–987.

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science*, *361*.

Wang, P., Chen, Y., Yong, J., Cui, Y., Wang, R., Wen, L., et al. (2018). Dissecting the global dynamic molecular profiles of human fetal kidney development by single-cell RNA sequencing. *Cell Reports*, *24*. 3554–3567 e3553.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., & Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, *367*.

Wu, H., Yu, Y., Huang, H., Hu, Y., Fu, S., Wang, Z., et al. (2020). Progressive pulmonary fibrosis is caused by elevated mechanical tension on alveolar stem cells. *Cell*, *180*. 107–121 e117.

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., et al. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nature Cell Biology*, *20*, 666–676.

Xu, Y., Mizuno, T., Sridharan, A., Du, Y., Guo, M., Tang, J., et al. (2016). Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight*, *1*, e90558.

Yu, X. X., Qiu, W. L., Yang, L., Zhang, Y., He, M. Y., Li, L. C., et al. (2019). Defining multistep cell fate decision pathways during pancreatic development at single-cell resolution. *The EMBO Journal*, *38*.

Zaidel-Bar, R., Zhenhuan, G., & Luxenburg, C. (2015). The contractome- -a systems view of actomyosin contractility in non-muscle cells. *Journal of Cell Science*, *128*, 2209–2217.

Zepp, J. A., & Morrisey, E. E. (2019). Cellular crosstalk in the development and regeneration of the respiratory system. *Nature Reviews. Molecular Cell Biology*, *20*, 551–566.

Zhu, Y., Scheibinger, M., Ellwanger, D. C., Krey, J. F., Choi, D., Kelly, R. T., et al. (2019). Single-cell proteomics reveals changes in expression during hair-cell development. *eLife*, *8*.

Zhu, M., Tao, H., Samani, M., Luo, M., Wang, X., Hopyan, S., et al. (2020). Spatial mapping of tissue properties in vivo reveals a 3D stiffness gradient in the mouse limb bud. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 4781–4791.