Chapter 23

Methodologies for Following EMT In Vivo at Single Cell Resolution

Abdull J. Massri, Geoffrey R. Schiebinger, Alejandro Berrio, Lingyu Wang, Gregory A. Wray, and David R. McClay

Abstract

An epithelial-mesenchymal transition (EMT) occurs in almost every metazoan embryo at the time mesoderm begins to differentiate. Several embryos have a long record as models for studying an EMT given that a known population of cells enters the EMT at a known time thereby enabling a detailed study of the process. Often, however, it is difficult to learn the molecular details of these model EMT systems because the transitioning cells are a minority of the population of cells in the embryo and in most cases there is an inability to isolate that population. Here we provide a method that enables an examination of genes expressed before, during, and after the EMT with a focus on just the cells that undergo the transition. Single cell RNA-seq (scRNA-seq) has advanced as a technology making it feasible to study the trajectory of gene expression specifically in the cells of interest, in vivo, and without the background noise of other cell populations. The sea urchin skeletogenic cells constitute only 5% of the total number of cells in the embryo yet with scRNA-seq it is possible to study the genes expressed by these cells without background noise. This approach, though not perfect, adds a new tool for uncovering the mechanism of EMT in this cell type.

Key words Epithelial-mesenchymal transition, Single cell RNA-sequencing, Sea urchin, Tissue morphogenesis

1 Introduction

Most embryonic mesoderm cells are initially specified when they reside in an epithelium. An epithelial-to-mesenchymal transition (EMT) then removes them from the epithelial layer and they adapt a mesenchymal phenotype. In some cases, these cells again become epithelial and go through additional EMTs. This process of leaving the epithelium also occurs with carcinoma cells. Whether the two EMTs share mechanistic components of the process is a question that has often been asked. Literature reports indicate that they do indeed share multiple properties: they tend to use the same controlling transcription factors (*twist*, *snail*, and *zeb1*), though not always. They appear similar in behavior (the cells become motile,

Kyra Campbell and Eric Theveneau (eds.), *The Epithelial-to Mesenchymal Transition: Methods and Protocols*, Methods in Molecular Biology, vol. 2179, https://doi.org/10.1007/978-1-0716-0779-4_23,

change polarity, invade through the basement membrane, de-adhere from the adherens junction, and the plasma membrane is remodeled), though differences are observed in different model systems. What is clear is that in both the carcinoma and embryo systems, the molecular basis of this complex cellular event called EMT is incompletely understood. Indeed, of the several thousand papers a year on EMT, most focus on the epiphenomenon, that is, does the phenotypic change occur to an epithelial culture, or layer, under applied experimental conditions? Far fewer papers focus on the functional mechanics of that EMT in molecular detail.

A major reason for not understanding the EMT process in greater detail is that most systems are asynchronous, that is, the cells undergoing an EMT are at different states at any given time making it difficult to deduce the precise sequence of molecular events. A few cases of EMT in embryonic systems do offer synchrony, but each of these also has shortcomings. For example, ventral furrow formation in *Drosophila melanogaster (Drosophila)* provides a near synchronous EMT of mesoderm cells, and some genes necessary for the process have been identified. However, the difficulty in that system is that the number of mesoderm cells is small relative to the remaining cells of the embryo, and the EMT occurs relatively early in development, at a time when many maternally expressed genes are still expressed. This makes it difficult to exploit the power of *Drosophila* genetics to discover the genes mechanistically involved specifically in the EMT process [1]. Anchor cell invasion in *Caenorhabditis elegans* is another embryonic EMT in which one cell invades through the basement membrane as part of vulval assembly [2]. In this case the system is genetically tractable and a number of genes involved in the process have been identified. There is no question of synchrony, since only the one cell participates. However, a shortcoming of this system for EMT analysis is that the anchor cell does not complete an EMT. It breaches the basement membrane in a manner similar to that utilized by cells undergoing EMT in other systems, but it does not de-adhere from the epithelium. The sea urchin embryo also has a population of cells that undergo EMT at a precise time in early development and a gene regulatory network of specification is well established for those cells, making this a useful model system for understanding control of the process [3]. Nevertheless, this system also has shortcomings in that the skeletogenic cells that go through the EMT are only 5% of the population of cells in the embryo, making it a challenge to determine the sequence of molecular events in that small population.

Here we describe a method that can be used on any system to at least partially overcome some of the shortcomings possessed by many systems. Single cell RNA-sequencing (scRNA-seq) has advanced to the point where one can obtain a profile of expressed RNA in each cell. Computational approaches along with a temporal

trajectory of single cells offers an approach to profile the molecular changes that occur in each cell undergoing the EMT over time. This approach therefore, has the potential of eliminating much of the noise introduced either by asynchrony of the EMT and/or inclusion of noninvolved cells, and the reward is provision of a temporal profile of molecular change.

It should be noted, however, that scRNA-seq is not the perfect solution. Because of the small amount of RNA obtained from each cell, amplification is necessary before sequencing. This and other limitations means that some rare RNA species are less likely to be included in the database than in bulk RNA-seq approaches. Nevertheless, the advances in scRNA-seq approaches provide the investigator with a valuable tool to penetrate EMT mechanisms to a level that heretofore has been unreachable.

2 The Single Cell RNA-Sequencing Approach, a Justification

Next generation sequencing (NGS) platforms increasingly allow in-depth analyses of gene expression and genetic interactions in many biological systems. The approaches allow the investigator unprecedented access to biological questions. The methodology begins with sample preparation, includes library production, sequencing, and data analysis. The latter is most important as software continues to be developed to enable the investigator to gain ever more detail about the biological process in question. As part of the description, the caveats and limitations of these technologies will be discussed. The focus will be on approaches that advance RNA-sequencing technologies and their application to understanding EMTs.

Two methods of RNA-sequencing are currently utilized, single cell RNA-sequencing (scRNA-seq) and bulk RNA-sequencing (RNA-seq). They each have their own individual advantages and disadvantages and are useful for addressing different biological questions. Bulk (whole-tissue) RNA-sequencing has many applications for research including comparative gene expression analyses between samples of various conditions, differential gene expression, identification of mRNA splice variants and small or long noncoding RNAs. RNA material collected from whole-tissue samples requires less or no amplification relative to scRNA-seq and the sample can be more deeply sequenced than that obtained from a single cell. Bulk RNA-seq is also easier: obtaining single cell suspensions from fixed or frozen tissue is nontrivial and may be very difficult for some samples. Thus, bulk RNA-sequencing is a good option in many applications. However, bulk RNA-seq is not as informative for identifying transcriptional differences within heterogeneous cell populations such as in developing and complex tissues because bulk RNA-seq measures the expression level of transcripts across a

population of various types of cells, therefore creating an average transcriptomic profile of the tissue. This can become an issue when rare cell types are of interest, because their signal is essentially lost in the noise and more abundant transcripts. One way to get around this issue is by enriching for the population of interest, using a cell surface marker, fluorescence or antibody, however, this will still provide an averaged transcriptome across cells and does not capture heterogeneity at the single cell level. Another way to improve the analysis is to perform a temporal trajectory of the material in question. For embryonic material this can be highly informative because it adds the element of time, although still, within each sample the heterogeneity produces noise.

Single cell RNA-sequencing has the potential to eliminate much of the noise within a mixed population of cells. With a temporal profile it enables the investigator to probe the transcriptional dynamics of heterogeneous cell populations because it measures the distribution of mRNA expression from individual cells. Single cell transcriptomes can be profiled for a number of purposes such as creating cell atlases, mapping cell lineage trajectories [4–10], modeling virtual in situ hybridization [11] and more [12]. Using scRNA, one can capture cell trajectories and developmental processes such as an EMT by applying a scRNA-seq time-course to construct a cell trajectory map [13]. Generating an EMT time-course to capture transient cell states at single cell resolution informs the investigator with information on how this dynamic process occurs over time, providing a resource that is not available in any other known way.

Single cell RNA-sequencing is rapidly becoming a viable alternative to bulk RNA-sequencing, however, there are still some inherent issues with the platform. One challenge is due to the fact that RNA is harvested from only a single cell, and generally needs to be amplified with reverse transcription or PCR. This process of amplification can introduce bias that can lead to an incorrect interpretation. However, this can be overcome during the normalization and computational analysis by using Unique Molecular Identifiers (UMI), to uniquely label individual RNA molecules, greatly reducing amplification bias. Additionally, due to the sparsity of some RNA transcripts present in the cell and the inefficient cell capture process, sometimes a gene may have moderate expression in some cells, but cannot be detected in another cell. These occurrences, known as gene dropouts can be misleading because it is difficult to differentiate between inefficiency of transcript capture and a cell lacking that particular gene expression, or a gene that is expressed intermittently, therefore dimensionality reduction and normalization should to be performed computationally [14, 15].

3 Preparation of Single Cell Suspensions for scRNA-Seq

The key to any scRNA-seq experiment is generating a healthy representative single cell suspension from dissociated tissues or embryos. Therefore, it is imperative to develop a tissue dissociation protocol that properly captures single cells with minimal loss of integrity of the cells and minimal degradation of RNA. To achieve these goals, it is of utmost importance to minimize the time away from a cell's native environment while generating and handling single cell suspensions to accurately capture a cell's RNA identity, before alterations can occur. The transient nature of RNA expression can potentially be fixed in time following dissociation with a proper fixative, such as methanol, and the cells washed and rehydrated in 3× SSC rather than PBS, because rehydration in PBS can cause RNA degradation [16, 17]. Tissue types from various organisms and embryos are highly variable in their composition, therefore to generate a single cell suspension, different tissues require different enzymes, temperatures, salinity, and pH. Many groups have utilized enzymes that degrade extracellular matrix components to facilitate their dissociations. To establish a protocol, single cell preparations should be kept consistent, because altering the method of preparation can introduce a sampling bias. To establish the optimal conditions our single cell dissociation protocol was developed using a pilot study to establish the most reliable approach and as part of that, establish that a fixative such as methanol can be used to stabilize the RNA. The pilot study helped establish optimal scRNA-seq conditions for our system. The details of dissociation and stabilization of RNA are too varied to be covered in this chapter, but in each case the goals outlined above should be sought.

4 Considerations of Approach and Instrumentation Available for Library Preparation from Single Cells

To a research group beginning a scRNA-seq project, the next big question to ask is what platform should be used? Single cell RNA-sequencing has rapidly evolved since it was first used in 2009 [18]. When scRNA-seq was first introduced, it involved manually pipetting single cells into microwells and was relatively low throughput with a considerable amount of work required per cell. Since then, many groups have contributed to making scRNA-seq cost efficient and high throughput, and today many variations of these technologies exist. The introduction of multiplexing in 2011 [19, 20] was a major milestone where they showed many single cells could be sequenced together when UMIs were used. Additionally, in 2013 [21] integrated fluidic circuits, to allow for

higher throughput, and more reproducibility. In 2015, [22, 23] introduced droplet-based methods where single cells are placed in droplets using microfluidics and beads with UMIs to uniquely label RNA molecules in each cell.

Currently a number of platforms are available to choose between, each with its advantages and disadvantages. Platforms differ from each other by either method of RNA quantification, or by method of cell capture. RNA expression is quantified by measure of either full length cDNAs or by tag-based UMIs. There are three methods of cell capture, microwell-based, microfluidic-based, and droplet-based. With the various options, it may seem difficult to determine which method is best, and the answer is it depends on the question being asked. Ziegenhain et al. [24] and Svensson et al. [25] realized this and so to assist you in making an informed decision they compare and contrast the common scRNA-seq techniques' accuracy, sensitivity, precision, power, and cost efficiency. Based on their findings, Smart-seq2, had the best sensitivity, accuracy, precision, and the lowest gene dropout rate, however this approach provides relatively low throughput compared to droplet-based methods that are not as sensitive but significantly less costly. Smart-seq2 currently is the best option for increased sequencing depth but for a smaller number of cells, as cost can be quite considerable. If willing to sacrifice some sequencing depth, drop-seq is the most cost efficient of the methods, but requires a tedious multi day protocol to be performed. Labs and sequencing centers also are adapting commercial platforms that include Fluidigm's C1 microfluidic chip, Wafergen ICELL8, BioRad's ddSEQ, and perhaps the most popular, 10× Genomics Chromium. Other alternatives utilize combinatorial indexing such as sciRNA-seq, while SPLiT-seq utilizes a split and pool method of barcoding cells within wells [4, 26]. These allow for higher throughput and cost efficiency than 10× and Drop-seq, however, the sample preparation takes longer, and there is a potential for introduction of sampling bias. In addition, the cell quality reportedly is a bit lower than 10× and Drop-seq. With all these options, it can be difficult to identify which method is best, for your research question. For a process such as EMT which has a temporal component, and for a process that occurs within an in vivo model (in our case, the sea urchin), we sought a method that could process many single cells with the best depth possible. To satisfy such a requirement, 10× Genomics was our choice of platform. Following library construction of single cells via 10× Genomics protocol, cells are sequenced at ~50 k reads per cell and using a 150 bp paired end Illumina run. Similarly, other single cell library preparation protocols utilize Illumina's paired end sequencing but may have different run length of 75, 125, 250 bp and more. Depending on the number of cells and the run length, a variety of options will be available using Illumina. For example, using a total of 1 billion PE reads on the NovaSeq

6000 and 150 bp PE run, roughly, 20 k cells can be sequenced at 50 k reads/cell to generate a single cell atlas capturing EMT. Indeed, the multitude of scRNA-seq techniques and methods are rapidly evolving, and as cost of scRNA-seq decreases, previous technologies will surely become obsolete. Research groups continue to push the limits and cost efficiency of scRNA-seq with methods like cell hashing that allow for "super loading" of cells, and it will only drive the cost down.

5 Bioinformatic Analysis: Overview of Procedure for Analysis of Results

Once single cell libraries are prepared and samples have been sequenced, the first step in analyzing the data is to create an expression matrix from the raw sequencing output. First, your bcl file should be demutiplexed using bcl2fastq to produce fastq files that can be checked for read quality control. A pipeline should be established early on, to identify what type of analyses will be performed (*see* Fig. 1 for a general ScRNA-seq pipeline that can be adapted). Following sequencing, Unique Molecular Identifiers (UMIs) should be extracted and reads demultiplexed using UMI-tools or zUMIs [28, 29]. To perform a quality check on reads, a common tool used is FastQC [30]. Once reads have been checked for quality control, they should be trimmed if a sample has poor base per sequence quality scores below 20, or if any exogenous nucleotides such as adapters were introduced. A few commonly used trimming tools are Trimmomatic, TrimGalore, and Cutadapt [31–33]. Trimmed reads can then be mapped back to a reference genome or transcriptome using a bulk RNA-seq aligner/pseudoaligner such as STAR/Kallisto or an aligner appropriate for your research question [34, 35]. Once reads have been mapped to genes, they are counted on a per gene and per cell basis to generate a single cell gene expression matrix [28, 36]. This matrix has a row for each cell and a column for each gene. The i, j entry encodes the number of molecules of mRNA for gene j in cell i. Therefore, each row encodes the expression profile of a cell as a point in a high-dimensional gene expression space, where there is a dimension for each gene.

With the expression matrix in hand, we are now ready to begin visualizing, exploring and analyzing the data. We begin by visualizing the high-dimensional single cell gene expression profiles in two or three dimensions. Some popular tools for visualizing single cell datasets include force layout embedding (FLE), UMAP, and tSNE [14, 15]. Instead of applying these tools directly to the single cell expression data, it can be helpful to first reduce the dimensionality from 20,000 down to ~1000 by selecting variable genes, and then down to ~100 using principal components analysis (PCA) or diffusion maps. This gradual decrease in dimensionality can help extract

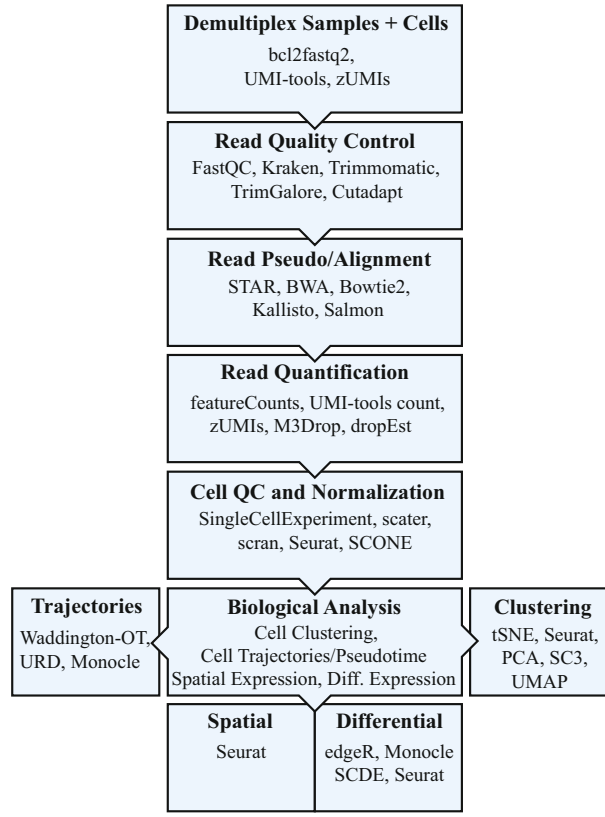

Fig. 1 General scRNA-seq pipeline. Figure adapted from and inspired by the single cell RNA-sequencing course [27]. Bioconductor is a repository that houses toolkits for sequencing and cell quality control, analysis, visualization, exploration, and more. Common packages used for each step in the pipeline are included. Using these methods, each gene's expression during EMT can be quantitatively measured in single cells, allowing for a deeper understanding of the underlying mechanistic structure of EMT

meaningful signals in the visualization. This visualization results in a set of x, y (and maybe z) coordinates that can be used to plot cells as points in 2 or 3 dimensions. Cells can be colored according to time of collection, batch, or expression of individual genes or gene signatures. The second component of exploratory data analysis involves searching for sets of cells with coherent gene expression programs. There are two main ways to do this. The first is to cluster cells (e.g., using Louvain clustering in diffusion component space). The second is to define cell sets according to expression of gene signatures. A gene signature is a list of genes (10 to 100 genes) related to a specific biological process or cell state (e.g., Epithelial Identity). To define an Epithelial cell state, we could select the top 10% of cells with highest expression of the Epithelial Identity gene signature.

In a time-course experiment, an expression matrix is obtained for each time point. The exploratory analysis described above can be applied to all time-points together in order to learn about general trends in expression over time. But, in order to learn about the different developmental trajectories and gene regulatory networks controlling differentiation, we must perform trajectory analysis.

The first goal of trajectory analysis is to infer ancestor–descendant relationships between pairs of time-points. This is crucial because scRNA-seq kills cells; therefore, we cannot use it to directly measure the change in gene expression of any individual cell over time. Live-cell imaging with fluorescent reporters can address this, but only for a handful of genes at a time. Many algorithms have been proposed to recover trajectories from scRNA-seq data. Waddington-OT is the only algorithm developed to date that is capable of modeling cell growth and development in a scRNA-seq time-course. All other algorithms either cannot incorporate known information about time of collection, or assume that all cells grow at the same rate (and therefore give rise to the same number of descendants). Waddington-OT infers ancestor-descendant relationships between pairs of time-points by leveraging a classical mathematical tool called optimal transport (OT). Intuitively, OT is based on the principle that cells can't change expression of all genes by large amounts in a short period of time. Therefore, cells are connected to "putative descendants" in a way that minimizes the total net change in expression over time. Each cell is allocated a certain amount of "descendant mass" according to an estimate of its proliferative ability and apoptosis rate (i.e., more proliferative cells are connected to more descendants). These growth rates are initially based on gene signatures of cell cycle and apoptosis, but are ultimately learned from data. The output of this first step of trajectory analysis is a "transport matrix" connecting each pair of time-points. The transport matrix has a row for each cell at time 1 and a column for each cell at time 2. The entries of the matrix indicate the amount of descendant mass each cell from time 1 gives rise to at time 2 (if we hadn't killed the cells).

After inferring ancestor–descendant relationships, the second goal of trajectory analysis is to infer gene regulatory networks controlling development and differentiation. To do this, Waddington-OT looks for transcription factors that are most predictive of transitions to various cell sets. For example, in iPSC reprogramming which transcription factors are responsible for pushing cells toward the stem cell state? Waddington-OT also allows us to analyze the shared ancestry connecting pairs of cell sets. This allows us to answer—does this pair of cell sets share a common ancestor near the beginning of the time-course and when does the pair diverge? We can then look for transcription factors that explain the bifurcation.

One common drawback of all these techniques is that spatial information is lost when cells are dissociated into suspension, however, the robust characterization of spatial markers within a tissue and developing embryo make it possible to reconstruct spatial patterning in silico. To reconstruct spatial information from dissociated tissues or embryos, Seurat can be employed to estimate a cell's likely position within spatial domains of the original tissue or embryo. As software matures and techniques improve in resolution, spatial transcriptomic technologies like Spatial Transcriptomics, Slide-Seq, and Seurat can provide more accurate spatial transcriptomic distributions [37, 38].

An outcome sought from this long list of computational options is a list of genes to be used in follow-up mechanistic studies. The question of how to reduce the size of that list varies with the goals in the system. In the case of the EMT, one approach might be to eliminate RNAs that are constitutively expressed since the EMT is fundamentally a change. Then, the direction of change and its timing can be considered using trajectories of RNAs and clustering programs. To that, data on perturbations, either based on known transcription factor control or perhaps on known drug effects can provide differential expression data that helps narrow the candidate list. Ultimately the goal is to identify candidates that are essential to the EMT and will help the investigator understand how the process works. To that end scRNA-seq provides an excellent tool.

Acknowledgements

The authors thank members of the McClay and the Wray labs for their critical input. We also appreciate the help provided by the Duke Core facility and the Benfey lab in the Biology Department. Support for this project was provided by NIH to DRM (RO1 HD 14483 and PO1 HD037105), and by NSF to GAW (IOS-1457305) and AJM for his NSF predoctoral fellowship (DGE-1644868). GS is supported by a Career Award at the Scientific Interface from the Burroughs Welcome Fund, and by funds from the Klarman Cell Observatory.

References

1. Schafer G, Narasimha M, Vogelsang E, Leptin M (2014) Cadherin switching during the formation and differentiation of the Drosophila mesoderm - implications for epithelial-to-mesenchymal transitions. J Cell Sci 127 (Pt 7):1511–1522. https://doi.org/10.1242/jcs.139485
2. Schindler AJ, Sherwood DR (2013) Morphogenesis of the Caenorhabditis elegans vulva. Wiley Interdiscip Rev Dev Biol 2(1):75–95. https://doi.org/10.1002/wdev.87
3. Saunders LR, McClay DR (2014) Sub-circuits of a gene regulatory network control a developmental epithelial-mesenchymal transition. Development 141(7):1503–1513. https://doi.org/10.1242/dev.101436
4. Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN,

Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357 (6352):661–667. https://doi.org/10.1126/science.aam8940

5. Chen J, Renia L, Ginhoux F (2018) Constructing cell lineages from single-cell transcriptomes. Mol Asp Med 59:95–113. https://doi.org/10.1016/j.mam.2017.10.004
6. Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW (2018) Cell type transcriptome atlas for the planarian Schmidtea mediterranea. Science 360(6391). https://doi.org/10.1126/science.aaq1736
7. Plass M, Solana J, Wolf FA, Ayoub S, Misios A, Glazar P, Obermayer B, Theis FJ, Kocks C, Rajewsky N (2018) Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science 360(6391). https://doi.org/10.1126/science.aaq1723
8. Tintori SC, Osborne Nishimura E, Golden P, Lieb JD, Goldstein B (2016) A transcriptional lineage of the early C. elegans embryo. Dev Cell 38(4):430–444. https://doi.org/10.1016/j.devcel.2016.07.025
9. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, Huang D, Xu Y, Huang W, Jiang M, Jiang X, Mao J, Chen Y, Lu C, Xie J, Fang Q, Wang Y, Yue R, Li T, Huang H, Orkin SH, Yuan GC, Chen M, Guo G (2018) Mapping the mouse cell atlas by microwell-seq. Cell 172 (5):1091–1107. e1017. https://doi.org/10.1016/j.cell.2018.02.001
10. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM (2018) Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science 360 (6392):981–987. https://doi.org/10.1126/science.aar4362
11. Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, Kipar C, Kocks C, Rajewsky N, Zinzen RP (2017) The Drosophila embryo at single-cell transcriptome resolution. Science 358(6360):194–199. https://doi.org/10.1126/science.aan3235
12. Haque A, Engel J, Teichmann SA, Lonnberg T (2017) A practical guide to single-cell RNA--sequencing for biomedical research and clinical applications. Genome Med 9(1):75. https://doi.org/10.1186/s13073-017-0467-4
13. Griffiths JA, Scialdone A, Marioni JC (2018) Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol 14(4):e8046. https://doi.org/10.15252/msb.20178046
14. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW (2018) Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. https://www.ncbi.nlm.nih.gov/pubmed/30531897
15. Van der Maaten L, Hinton G (2008) Visualizing data using T-SNE. J Mach Learn Res 9:2579–2605
16. Juliano C, Swartz SZ, Wessel G (2014) Isolating specific embryonic cells of the sea urchin by FACS. Methods Mol Biol 1128:187–196. https://doi.org/10.1007/978-1-62703-974-1_12
17. Chen J, Cheung F, Shi R, Zhou H, Lu W, Consortium CHI (2018) PBMC fixation and processing for chromium single-cell RNA sequencing. J Transl Med 16(1):198. https://doi.org/10.1186/s12967-018-1578-4
18. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6(5):377–382. https://doi.org/10.1038/nmeth.1315
19. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, Linnarsson S (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 21(7):1160–1167. https://doi.org/10.1101/gr.110882.110
20. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lonnerberg P, Linnarsson S (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11(2):163–166. https://doi.org/10.1038/nmeth.2772
21. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V, Baying B, Benes V, Teichmann SA, Marioni JC, Heisler MG (2013) Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 10(11):1093–1095. https://doi.org/10.1038/nmeth.2645
22. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161 (5):1202–1214. https://doi.org/10.1016/j.cell.2015.05.002
23. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW (2015) Droplet barcoding for single-cell transcriptomics applied

to embryonic stem cells. Cell 161 (5):1187–1201. https://doi.org/10.1016/j.cell.2015.04.044

24. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65(4):631–643. e634. https://doi.org/10.1016/j.molcel.2017.01.023
25. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA (2017) Power analysis of single-cell RNA-sequencing experiments. Nat Methods 14(4):381–387. https://doi.org/10.1038/nmeth.4220
26. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, Graybuck LT, Peeler DJ, Mukherjee S, Chen W, Pun SH, Sellers DL, Tasic B, Seelig G (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360(6385):176–182. https://doi.org/10.1126/science.aam8999
27. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, Natarajan KN, Reik W, Barahona M, Green AR, Hemberg M (2017) SC3: consensus clustering of single-cell RNA-seq data. Nat Methods 14(5):483–486. https://doi.org/10.1038/nmeth.4236
28. Smith TS, Heger A, Sudbery I (2017) UMI-tools: modelling sequencing errors in unique molecular identifiers to improve quantification accuracy. Genome Res 27 (3):491–499. https://doi.org/10.1101/gr.209601.116
29. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I (2018) zUMIs - a fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience 7(6). https://doi.org/10.1093/gigascience/giy059
30. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data
31. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170
32. Krueger F (2012) http://www.bioinformatics.babraham.ac.uk/projects/trim_galore
33. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):10–12
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635
35. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34 (5):525–527. https://doi.org/10.1038/nbt.3519
36. Andrews TS, Hemberg M (2019) M3Drop: dropout-based feature selection for scRNASeq. Bioinformatics 35(16):2865–2867. https://doi.org/10.1093/bioinformatics/bty1044
37. Eng CL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, Yun J, Cronin C, Karp C, Yuan GC, Cai L (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature 568 (7751):235–239. https://doi.org/10.1038/s41586-019-1049-y
38. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ (2019) Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science 363(6434):1463–1467. https://doi.org/10.1126/science.aaw1219