© 2020. Published by The Company of Biologists Ltd | Development (2020) 147, dev179788. doi:10.1242/dev.179788

SPOTLIGHT

Cellular diversity and lineage trajectory: insights from mouse single cell transcriptomes

Patrick P. L. Tam[1,2,*] and Joshua W. K. Ho[3]

ABSTRACT

Single cell RNA-sequencing (scRNA-seq) technology has matured to the point that it is possible to generate large single cell atlases of developing mouse embryos. These atlases allow the dissection of developmental cell lineages and molecular changes during embryogenesis. When coupled with single cell technologies for profiling the chromatin landscape, epigenome, proteome and metabolome, and spatial tissue organisation, these scRNA-seq approaches can now collect a large volume of multi-omic data about mouse embryogenesis. In addition, advances in computational techniques have enabled the inference of developmental lineages of differentiating cells, even without explicitly introduced genetic markers. This Spotlight discusses recent advent of single cell experimental and computational methods, and key insights from applying these methods to the study of mouse embryonic development. We highlight challenges in analysing and interpreting these data to complement and expand our knowledge from traditional developmental biology studies in relation to cell identity, diversity and lineage differentiation.

KEY WORDS: Single cell analytics, Embryo cell atlas, Developmental trajectory, Cell lineages, Bioinformatics

Introduction

During mouse development (Fig. 1), inner cell mass (ICM) cells of the blastocyst differentiate and contribute to the epiblast and the extra-embryonic endoderm lineages. At gastrulation, the multipotent epiblast cells differentiate into the three germ layers: the ectoderm, mesoderm (including embryonic and extra-embryonic cell lineages) and endoderm. A blueprint of the mouse body plan is then established during post-implantation development when diverse lineages are specified within each germ layer. The derivatives of these lineages are regionalized in domains defined by their position along the embryonic body axes, i.e. anterior-posterior, medial-lateral and dorsal-ventral patterning. The population of embryonic cells expands rapidly from ∼15 ICM cells in the blastocyst, to about 400,000 cells in the early organogenesis stage embryo (Kojima et al., 2014b), and around 13 million cells in the organogenesis stage embryo (Cao et al., 2019) (Fig. 1). In addition, there is extensive cell and tissue movement, morphogenesis, and growth at every step of embryogenesis (Tam and Gad, 2004; Tam and Loebel, 2007; Rivera-Pérez and Hadjantonakis, 2014). It is against this backdrop that cell specification and tissue patterning are accomplished, primarily through changes in the functional output of genomic activity (such as changes to the transcriptome, proteome and metabolome) together with the input of signalling activity, epigenetic modulation and mechanobiological function.

A significant body of knowledge about the cellular and molecular mechanisms of embryogenesis has been assembled from targeted studies of candidate genes and molecules (Arnold and Robertson, 2009; Solnica-Krezel and Sepich, 2012; Tam and Behringer, 1997; Tam and Loebel, 2007). These studies have highlighted the crucial role of transcription factors, epigenetic modifiers, signalling pathway factors, and mediators of inductive and physical tissue interactions in embryogenesis. However, owing to constrained sources of embryonic material, genome-wide studies of developmental drivers have previously been limited to the analysis of the bulk transcriptome and epigenome, based on the sum of millions of cells in whole embryos, organs, tissues or cell populations. These studies have shown global changes in gene expression and chromatin state profiles, revealing molecular changes that underlie the transition from pluripotency, the acquisition of lineage characteristics, and the modulation of signalling and morphogenetic activity during early post-implantation development (Kojima et al., 2014a; Mitiku and Baker, 2007). It is evident that bulk transcriptome analyses only provide an 'average' of genome activity, because they do not take the heterogeneity of cell populations into account. The averaged modality gives little insight into the specific and regionalized molecular activities that govern the specification and differentiation of individual lineages, nor the morphogenetic and patterning activity in the multitude of cell types. It is, therefore, imperative to refine the analysis to either sub-population or cellular resolutions.

Advances in high-resolution and enhanced-throughput analytics of transcriptome, epigenome and live imaging now offer the ability to perform single cell analysis of the mouse embryo. Here, we provide a snapshot of the methodology and knowledge gained from recent studies (Fig. 1; Table 1).

Molecular cell atlases of mouse development

Conventional cell atlases of whole organisms, organs and tissues have been constructed from histological attributes and knowledge of cell type-specific markers (i.e. protein and gene expression). While these atlases have informed the composition and architecture of cells and tissues, the delineation of cell types by morphological characteristics and known markers might not reveal the full suite of cellular diversity. In addition, morphological characteristics and markers may bear no relationship to the functionality and lineage relationships of specific cell types. In this context, previously uncharacterised cell types or functional cell states may be identified by 'clusters' of cells with similar transcriptomes that are distinct from other groups of cells based on scRNA-seq data (Fig. 2A).

[1]Embryology Unit, Children's Medical Research Institute, The University of Sydney, Sydney, NSW 2145, Australia. [2]The University of Sydney, School of Medical Sciences, Faculty of Medicine and Health, Sydney, NSW 2006, Australia. [3]School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

*Author for correspondence (devtam@cmri.usyd.edu.au)

P.P.L.T., 0000-0001-6950-8388

Development (2020) 147, dev179788. doi:10.1242/dev.179788

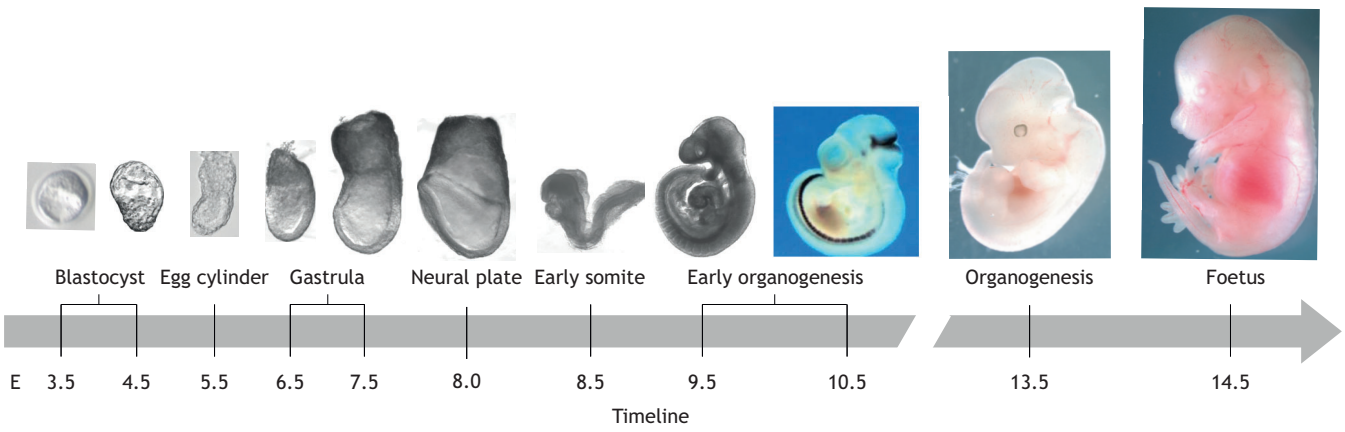

Fig. 1. A timeline of mouse development. E, embryonic day.

The utility of scRNA-seq has been demonstrated by the construction of two mouse cell atlases. First, the 'Tabula Muris' that delineates a multitude of cell clusters in organs and tissues of postnatal, adult and aging mice (6 weeks to 24 months) (The Tabula Muris Consortium, 2018, 2019 preprint; tabula-muris-senis.ds.czbiohub.org/). Second, the 'micro-well-seq' atlas that shows as many as 800 cell clusters in 51 tissues of postnatal (6-8 weeks) mice and E14.5 fetus (Han et al., 2018) (Table 1). These transcriptomic data have further enriched the molecular phenotype of cell types that were previously inadequately characterized.

Single cell atlases of the developing mouse embryo

In contrast to the adult cell atlases, which largely depict a snapshot of cellular composition at a homeostatic state, an embryonic cell atlas requires the incorporation of a developmental timeline. This is because the spectrum of cell types is expected to change over time, as multipotent embryonic cells transition to progenitors and intermediate cell types, and finally differentiate into mature cell types. Therefore, in order to capture the progressive changes in cells of every lineage, samples were often taken from embryos at a series of developmental stages.

Recently, several single cell transcriptomic analyses have been performed on mouse embryos between peri-implantation and organogenesis stages (Fig. 1; Table 1). The outcomes of these studies on the annotation of cell types have not only corroborated the findings of previous fate-mapping and lineage-tracing experiments but have also provided new insights into the molecular mechanism of lineage differentiation (Box 1). For example, in the pre-gastrulation embryo, the extra-embryonic visceral endoderm comprises the anterior visceral endoderm (AVE) and non-AVE cells that are spatially separated in the tissue layer (Cheng et al., 2019; Wen et al., 2017; Mohammed et al., 2017). In contrast, the epiblast contains three types of cells that display anterior and posterior epiblast-like properties and intermediate cell states, but they are not regionally segregated in the embryo (Cheng et al., 2019). These epiblast cells are comparable at the molecular level to the epiblast-like cells derived from the embryonic stem cells, which may represent the transition of cell state from naïve to formative pluripotency (Cheng et al., 2019; Mohammed et al., 2017). These findings thus define the molecular benchmark of formative pluripotency and the characteristics of cells in the transition state.

Based on an atlas of more than 116,000 single cells profiled across gastrulation and early organogenesis stages in mouse embryos, Pijuan and colleagues identified cell clusters representing the ectoderm, mesoderm and endoderm lineages (Pijuan-Sala et al., 2019). The number of cells in each cell cluster progressively increases and diverges into sub-clusters as the embryo develops. The study showed that the transient emergence of the epiblast cells that constitute the germ layer progenitors (e.g. EPI-1 Epi-2 and primitive streak cells) is accompanied by the decline of the population of pluripotent epiblast cells (Pijuan-Sala et al., 2019). Starting with only a few cell clusters at E6.5, around 21 cell clusters can be observed at the early organogenesis stage. Some clusters represent progenitors such as neuromesoderm progenitors (NMPs), spinal cord progenitors, primordial germ cells, cells of the haemogenic-endothelial lineages (encompassing the progenitors of macrophages and megakaryocytes), microglial macrophages, tissue-specific endothelial precursors and potentially other specialized cell types (Pijuan-Sala et al., 2019; Ibarra-Soria, et al., 2018; Chan et al., 2019).

In other studies, 38-98 cell clusters are identified in the mouse cell atlas at the organogenesis stage (Fig. 1; Table 1) (Cao et al., 2019; Han et al., 2018). Although in a developing embryo most cell types expand in numbers exponentially, the abundance of cells of different clusters may appear in proportionality relative to one another. For example, the average number of cells in a single cluster is 50,000, with connective tissue precursors and neutrophils having the highest (145,000) and lowest (100) abundance, respectively. There are about 570 subdivided clusters that could be delineated by the expression of known marker genes, but over 90 sub-divisions of clusters are no longer represented by the later fetal stage (Han et al., 2018), which suggests that they may be early lineage progenitors, transitory/intermediate cell types or replaced by definitive cell types, e.g. primitive erythroid progenitors are replaced by liver-resident hematopoietic stem cell-derived counterparts (Scialdone et al., 2016; Ibarra-Soria et al., 2018; Han et al., 2018). Indeed, mining the transcriptional activity among closely related cell clusters has revealed some of the molecular activities driving lineage differentiation.

Individual single cell transcriptome studies cover only a fraction of cells of the embryo: 16-62% coverage between gastrulation and early organogenesis stages, and 3-80% at the advanced organogenesis stage. In addition, the ensemble of RNA molecules in each cell (up to about 3500 in some cases) are not recovered. Collectively, however, these studies have collated the single cell transcriptomes with a good coverage of the cell population and a substantial depth of transcript reads, and from embryos across a

Development (2020) 147, dev179788. doi:10.1242/dev.179788

Table 1. Datasets available from single cell transcriptome analysis of mouse development

		Embryonic day (E)															Reference
		3.5	4.5	5.25	5.5	6.5	7.5	8.25	8.5	8.75	9.5	10.5	11.5	12.5	13.5	14.5	
Molecular cell atlas	Whole embryo			■	■	■											Cheng et al. (2019)
						■	■	■	■								Pijuan-Sala et al. (2019)
								■									Ibarra-Soria et al. (2018)
												■	■	■	■	■	Cao et al. (2019)
	Germ layers					■	■										Wen et al. (2017)
		■	■	■	■	■	■	■	■	■							Notwotschin et al. (2019)
	Tissues and organs											■	■	■	■	■	Cardoso-Moreira et al. (2019)
																■	Han et al. (2018)
Cell lineage relationship									■		■						Chan et al. (2019)
											■	■	■	■	■		Cao et al. (2019)
Transcriptional state/epigenetic landscape		■	■	■	■	■	■										Mohammed et al. (2017)
			■		■	■	■										Argelaguet et al. (2019)
				■	■	■											Cheng et al. (2019)
Molecular pathway/cellular network				■	■	■											Cheng et al. (2019)
						■	■	■	■								Pijuan-Sala et al. (2019)
						■	■	■	■								Scialdone et al. (2016)

series of developmental stages *in vivo* (Fig. 1; Table 1). It would be a worthwhile exercise to integrate the datasets from these studies (Lin et al., 2019) to construct a consensus, and perhaps more in-depth, molecular cell atlas, of the peri-implantation to organogenesis stage mouse embryo.

Towards a multi-omic atlas of mouse embryonic development

Beyond the transcriptome, the cell-type-specific molecular information in the cell atlas can be enriched by integration with data from parallel -omics analyses of the same cell (Adey, 2019; Welch et al., 2019). For example, protein-DNA interactions (scDamID&T-seq, Rooijers et al., 2019), chromatin status (scATAC-seq, Stuart et al., 2019; sc-ChIP-seq, Grosselin et al., 2019; histone ChIP-seq, Argelaguet et al., 2019); DNA methylation (methylome by bisulphite-seq, Argelaguet et al., 2019; sc-iMET, Mulqueen et al., 2017) and proteomes (Graf, 2019) can all be analysed at the single cell level. This additional information allows a refined categorization of cell types and provides insights into the molecular activity associated with the acquisition of cell fate.

In the context of epigenetic impact on transcriptional activity, the single cell transcriptome of pre-gastrulation embryos has revealed that the acquisition of cell states is influenced by the asymmetric activity of lineage determinants with bivalent chromatin status (the presence of both active and repressive chromatin histone modifications in the gene locus, Bernstein et al., 2006). In addition, the differentiation of the epiblast and primitive endoderm, and lineage allocation of primitive streak cells, is related to the repression of epigenetic modifiers, such as the genes targeted by the Polycomb complex (Mohammed et al., 2017).

Multi-omic single cell analyses of the methylome, histone chromatin immunoprecipitation sequencing (ChIP-seq) data and the transcriptome (Argelaguet et al., 2019) have revealed that epigenetic-transcriptional interaction underpins the exit of pluripotency. Specifically, the *de novo* methylation of CpG-depleted loci and reduced chromatin accessibility leads to repression of pluripotency regulatory networks and germline determinants. In another example, the acquisition of embryonic versus extra-embryonic lineage state has been shown in the context of epigenetic priming/remodelling, such as the loss of methylation in enhancers of lineage-driving regulatory genes, which accompanies cell fate decision of germ layer precursors and the increased diversity of transcriptional activity in cells. Then, during gastrulation, the expression of transcription of factors that determine lineage, tissue patterning and cell cycle progression is associated with the loss of methylation in distal enhancers (distally located regulatory elements) and an increase in chromatin accessibility. Finally, priming of the key transcription factors of ectoderm lineage has been shown to take place in the early epiblast, which enhances the propensity of ectoderm differentiation, as in embryonic stem cells. These ectoderm loci are repressed concurrently with the demethylation and chromatin opening of the loci related to specification of mesoderm and endoderm lineages (Argelaguet et al., 2019; Mohammed et al., 2017).

Incorporating spatial information

Unlike a conventional anatomical atlas, the molecular single cell atlases discussed so far do not map the spatial arrangement of various cell types in the mouse embryo. One method is to computationally infer spatial relationships among various cell types, through the so-called pseudo-space analysis. Pseudo-space analysis of position-specific marker genes has enabled the mapping of specific cell types to the gastrulating and neural plate/head fold stage embryo (Fig. 1; Table 1), thereby locating the progenitors of the embryonic and extra-embryonic mesoderm (Scialdone et al., 2016), and the haematopoietic-endothelial lineage (Ibarra-Soria et al., 2018). Furthermore, this approach has inferred the spatial proximity of mesodermal cells in the presomitic mesoderm on the basis of the dynamic expression patterns of the transcriptional regulator and its downstream effect on the somite maturation process (Ibarra-Soria et al., 2018). The spatio-temporal distribution of different endoderm cell types in the anterior-posterior axis of the

Development (2020) 147, dev179788. doi:10.1242/dev.179788

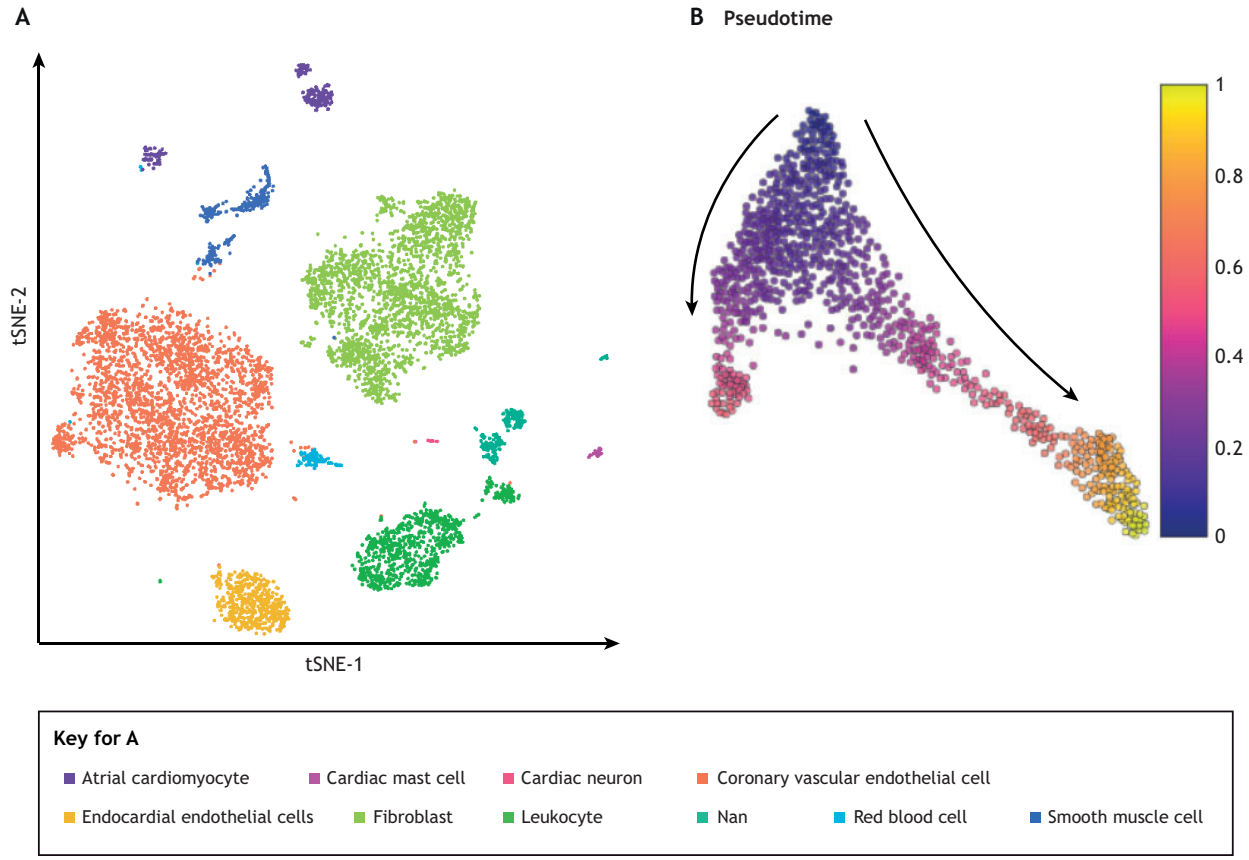

Fig. 2. Visualization of the single cell transcriptome data. (A) A representative tSNE plot illustrating the identification of cell clusters in a mouse single cell atlas (droplet based transcriptomic profiles of heart and aorta in Tabula Muris Senis). (B) A representative plot showing inferred pseudo-time (shown in a colour scale) trajectory of mouse embryonic E3.5-E4.5 cells from inner cell mass, epiblast and primitive endoderm (Nowotschin et al., 2019). This trajectory suggests the branching of one initial cell type into two differentiating cell types.

developing gut has been modelled by computing the distance of the transcriptome from known anterior-most cell types (Nowotschin et al., 2019) and by the expression of genes that are specific to each segment of the gut (Ibarra-Soria et al., 2018; Pijuan-Sala et al., 2019).

Beyond computational inference of spatial information based on standard scRNA-seq data, innovative methods are being developed to directly capture the spatial information of the cells. These methods use barcoded DNA sequences to mark the cellular origin of the transcripts that can be visualised spatially by multiple rounds of fluorescent *in situ* hybridization (Eng et al., 2019), *in situ* amplification of RNA sequences from cells at specific locations in the tissue (STARMAP, Wang et al., 2018) or capturing the RNAs from single cells that carry spatial information (SLIDE-seq, Rodriques et al., 2019). Despite the lower throughput with respect to the number of cells and transcripts, integrating *in situ* data with scRNA-seq data using these techniques has achieved the compatible clustering of the expected cell types (Stuart et al., 2019). Indeed, annotation of positional information of single cells would be a useful attribute of the next generation of embryo cell atlases.

Inferring developmental trajectories and constructing lineage trees

Pseudo-time trajectories

More than 50 bioinformatics tools have been developed for inferring developmental trajectories from scRNA-seq data (Saelens et al., 2019). Most methods infer ‘pseudo-time’ ordering of cells, which orders cells based on their similarity in the transcriptome (or continuity in terms of expression gradient) in a linear or tree-like trajectory (Kester and van Oudenaarden, 2018) (Fig. 2B). Based on the assumption that cells that are of similar developmental stages along a differentiation lineage tend to have comparable gene expression profiles, and that the degree of similarity is proportional to how ‘far apart’ two cells are along the developmental trajectory, the pseudo-time ordering structure is a suitable proxy for visualizing connectivity in cell lineages.

Analyses of single cell transcriptome have shown that the proximity between cell clusters – and therefore cells with similar transcriptomes – correlates with the genealogy relationship between cell types. Examples include the connection of epiblast with neural ectoderm and the primitive streak, of the primitive streak with mesoderm and endoderm, and of NMPs with trunk neuroectoderm and mesoderm (Pijuan-Sala et al., 2019). The haematopoietic-endothelial lineage can be subclustered into descendants of progenitors that give rise to first wave of primitive erythrocytes, and descendants of endothelial progenitors in the second wave of haematopoiesis (Pijuan-Sala et al., 2019). Sub-populations of endothelial cells (the allantoic and non-allantoic types) that display different molecular identities and are allocated to different regions/tissues of the embryo are predicted to have a common origin (Ibarra-Soria et al., 2018). Recently, a large single cell atlas of gastrulating endodermal tissues revealed the convergence of trajectories of visceral endoderm-derived and epiblast-derived endoderm cells (Nowotschin et al., 2019) (Box 2). In the

Development (2020) 147, dev179788. doi:10.1242/dev.179788

Box 1. Examples of molecular activities underlying lineage differentiation that can be identified or supported by single cell transcriptomic analysis

Signalling activity

- Differential activity of FGFR1, FGFR2 and FGFR4 drives the segregation of epiblast and primitive endoderm, and of epiblast and visceral endoderm (Nowotschin et al., 2019).
- Changes in FGF and Gata6/Nanog/Fgf4 activity regulate differentiation of the epiblast (Nowotschin et al., 2019).
- The signalling activity of the visceral endoderm in patterning the pre-gastrulation epiblast (Cheng et al., 2019), and mesoderm and endoderm during early gastrulation (Wen et al., 2017).
- Mapping of signalling domains of FGF, WNT, Notch, retinoic acid, BMP, Hedgehog, JAK/STAT and HIPPO pathways in the embryonic gut (Nowotschin et al., 2019).
- The dynamic regulation of cell signalling and proliferation in the apical ectodermal ridge of the limb bud.

X-chromosome activity

- The dynamic reactivation and inactivation of the X-chromosome (in XX cells) (Cheng et al., 2019; Mohammed et al., 2017).

Molecular drivers

- The role of Tal1 in diverting mesoderm progenitors to the primitive erythroid, megakaryocyte and myeloid cells during the second wave of haematopoiesis (Pijuan-Sala et al., 2019).
- The role of the leukotriene pathway in specifying haematopoietic progenitor cells (Scialdone et al., 2016; Ibarra-Soria et al., 2018).
- The molecular determinants driving the differentiation of skeletal muscles of mesenchymal lineages (Cao et al., 2019).

organogenesis stage embryo, there are simple linear trajectories as well as branching (divergence of cell types in an inferred lineage; Fig. 2B) and converging trajectories (similar cell types being generated in parallel trajectories), and trajectories with multiple pseudo-time starting points and end points in a developmental timeline (Cao et al., 2019). These studies demonstrate that developmental trajectories can be inferred from a well-designed scRNA-seq study by computational trajectory analysis. Many computational methods are available, most performed similarly on simple trajectory structures and when the data were relatively clean (Saelens et al., 2019).

It is important to appreciate that pseudo-time analysis of scRNA-seq data is not a perfect substitution for the traditional lineage tracing of cells (Kester and van Oudenaarden, 2018). The pseudo-time analysis of lineage relationship may be fraught with confounding caveats, such as the uneven or insufficient sampling of time point and tissue, and the presence of non-linear intermediates, such as cell clusters that only appear transiently at a specific time point or the transdifferentiation of certain cell types. Finally, the data are not free of systematic variation caused by batch effect and variable sequencing depth or quality.

In fact, even normalization and processing methods applied to scRNA-seq can impact on the algorithms of trajectory inference (Saelens et al., 2019; Tian et al., 2019). A pedigree map can be generated from scRNA-seq data even when the cells have no inherent developmental relationship. Pseudo-time trajectory analysis of good-quality scRNA-seq data that profile a large proportion of cells in a well-defined time series cell differentiation system, such as cultured hematopoietic cells, may reflect lineage relationships. However, this may not be the case where samples consist of mostly terminally differentiated cells. It is crucial to leverage temporal information instead of assuming that all cell states exist simultaneously at a specific time point. It is in this context that computational trajectory inference is particularly relevant to studying mouse embryonic development as the current single cell technologies are able to capture a good proportion of cells in an embryo with a good mix of differentiating and differentiated cells, across several embryonic time points.

Three recent studies have pushed the concept of computational trajectory inference beyond discovering a pseudo-time and branching structure (Setty et al., 2019; La Manno et al., 2018; Schiebinger et al., 2019) (Box 3). New bioinformatics methods are attempting to infer cell fate potential of the future cell state of individual cells within a lineage in the context that the expression profile of single cell could predict the origin of the cell and the direction of differentiation.

Molecular barcoding methods

The developmental trajectory of tissue lineages at a single cell resolution has been documented by simultaneous tracking of lineage progression with the collation of transcriptional phenotype. An emerging approach to harness single-cell -omic data for inferring lineage history is to use natural sequence variation among endogenous somatic cells in tissues to perform lineage reconstruction. For example, it is possible to identify nucleotide variations on RNA sequences from scRNA-seq data. These variations could come from somatic mutations or, perhaps more likely, RNA editing (Ding et al., 2019). Another potential source of endogenous sequence variation is from mutations in mitochondrial DNA, which is particularly abundant in ATAC-seq data (Xu et al., 2019). Nonetheless, as mitochondria have the potential to undergo exchange with other cells, these types of mitochondria-based methods need further evaluation for their applicability in mouse embryos. For these novel analyses that exploit endogenous sequence variations to infer genetic lineages, proper experimental design is still a crucial requisite. For example, it may be important to also obtain whole-genome or exome sequencing data to identify the full set of sequence variations, which will greatly facilitate the discovery and quantification of mutations at the single cell level. In identifying sequence variations in coding genes, it may be useful to profile the full-length transcriptome, rather than using a 5′- or 3′-biased protocol.

Recently, single cell analysis has used dynamic (time-reiterative) molecular barcoding for tracing cell lineages *in vivo* (McKenna and

Box 2. An insight from inferred developmental trajectory of cell lineages

The convergence of distinct lineage descendants to a cell population that shares a common molecular phenotype is exemplified by the developmental connection between the visceral endoderm (VE)-derived and epiblast-derived definitive endoderm (DE) cells in the gut endoderm. The descendants of different lineages contribute to intermixing subtypes of endoderm cells, with enrichment of DE in anterior and VE in posterior segments of the gut. The VE and DE cells show positional identity and comparable transcription factor signatures of 20 core genes in early endoderm primordia (Nowotschin et al., 2019). The origin of the endoderm in the gut was inferred by the analysis of extra-embryonic endoderm (ExE) signature that distinguishes the VE-derived endoderm from epiblast-derived DE in the hindgut (Pijuan-Sala, et al., 2019; Nowotschin et al., 2019). The inferred ontogeny reveals the developmental connectivity of distal (at E5.5) and anterior visceral endoderm (at E6.5), and a rare contribution of the epiblast of the pre-gastrulation embryo to the visceral endoderm lineage. The ExE-specific genes are expressed in visceral endoderm prior to intercalation of DE and expression of these genes declines after intercalation followed by the upregulation of a common set of DE-specific genes upon DE intercalation and epithelial remodelling (Nowotschin et al., 2019).

Development (2020) 147, dev179788. doi:10.1242/dev.179788

Box 3. Developmental trajectory analysis beyond pseudo-time inference

Palantir is a computational method that models cell fate decision, not as a series of discrete branching events at specific time points, but as continuous changes in the probability of which cell will differentiate into different future cell fates (Setty et al., 2019). Importantly, such a probability distribution of cell fates can be inferred from scRNA-seq data. Cell fate is inferred by evaluating two quantities that can be determined for every cell with respect to an early anchored cell: branching probabilities and differentiation potential. A drop in plasticity or transcriptional noise at each state indicates lineage restriction. Palantir identified the trajectory of embryonic gut endoderm that descends from visceral endoderm and epiblast-derived definitive endoderm (Box 2) (Nowotschin et al., 2019).

Another recent study generated a scRNA-seq time series during the reprogramming of fibroblasts to induced pluripotent stem cells (Schiebinger et al., 2019). In this example, each single cell is viewed as a point in a multi-dimensional gene expression space. Developmental trajectory is modelled by moving a collection of cells from the area in this space to another area. Assuming that a cell prefers to 'differentiate' using the shortest path (or smallest changes in gene expression), it is possible to infer where a group of cells is likely to be located in this gene expression space in the past and future, thereby enabling the reconstruction of a developmental trajectory of a specific cell lineage.

In addition to using the gene expression levels alone, it is possible to use the sequence information in scRNA-seq to provide additional information to predict future gene expression. An interesting observation is that the relative abundance of unspliced transcripts is correlated with its rate of active transcription – termed RNA velocity (La Manno et al., 2018). Using this information, it is possible to identify direction of change of all the expressed genes in a scRNA-seq data set, and therefore provide a means to predict the future cell states on a time scale of hours. This type of analysis is potentially powerful as it provides a biophysical basis for predicting future cell states.

Gagnon, 2019). In this system, CRISPR-controlled generation of mutations in an integrated barcode sequence, driven by promoters of tunable dynamics, provides a lineage record that can be read simultaneously with the cell-specific transcriptome of individual cells (Chan et al., 2019). Indeed, one study analyzed lineage-tagged single cells at the early organogenesis stage to produce a lineage tree demonstrating progressive cell fate restriction. In addition, emergent sub-lineages at various branching points of the tree indicate divergence of differentiated cell types from intermediate cell types, such as NMPs. Importantly, the proximity of cell state delineated by the transcriptome does not always reflect lineage relationships, as illustrated by the similar transcriptomes of endoderm cells, which are derived from separate populations of visceral endoderm and epiblast-derived definitive endoderm (Chan et al., 2019; Nowotschin, et al., 2019; Pijuan-Sala et al., 2019) (Box 2). The phenotypic convergence of the descendants of two endoderm lineages may be an example of cell transdifferentiation *in vivo*. Lineage trees enable the quantification of parental cells for the tissue lineages. For example, the developing mouse embryo comprises (on average) 1-6 totipotent cells, 10-20 pluripotent cells and 18-51 late multipotent cells at different stages; and, in decreasing order of abundance over time, ectoderm and mesoderm precursors, blood, mesendoderm and endoderm, and primordial germ cells (Chan et al., 2019). It may be noted that molecular barcoding presents caveats in the efficiency of recovering barcodes and uneven cell sampling that may skew the outcome of lineage reconstruction. It is imperative, however, to cross reference the lineage trajectories inferred computationally by pseudotime analysis and constructed by tracking molecular barcode, to evaluate any discrepancy between the findings of the two approaches of single cell analytics.

Towards a spatial lineage trajectory

With the knowledge of cell lineages, it would be feasible to retrospectively map the single cell populations collected from the time series onto the tree-like structure to reconstruct the *in vivo* developmental trajectory of specific cell types. Further understanding of the lineage relationship of cells in the embryo and the molecular activity that drives lineage specification during embryonic development requires additional information on the developmental journey and final destination of the lineage of interest (i.e. the whereabouts of the cells from the beginning to the end of lineage development in the embryo), which is missing from conventional single cell transcriptomic datasets. Embedded in the spatio-temporal transcriptome of cell populations in pre-gastrulation to gastrulation stage embryos are sets of transcripts that are uniquely expressed by cells at a defined position in the embryo, termed the zipcode. By mining the zipcode from the transcriptome of the single cells, it is now possible to infer the address of a single cell in the embryo by mapping the cell to a spatially delimited cell population (Peng et al., 2016, 2019). Recently, a live-imaging study of morphogenetic cell movements in the mouse embryo has tracked the journey of individual cells and their clonal descendants during gastrulation and early organogenesis (McDole et al., 2018). It is therefore feasible to map the journey and destination of the descendants of individual cells at any inferred position of the embryo to provide the requisite positional information for constructing a spatial rendition of the lineage trajectory.

Future perspectives

Future single cell atlases will likely be increasingly multi-omics and incorporate spatial information (HuBMAP Consortium, 2019). Combining the data on the spatial location, the inferred developmental trajectory and the cell type-specific molecular attributes may help to compile a compendium of the lineage trajectory and molecular activity that underpins the differentiation of cell lineage and tissue patterning in the mouse embryo. At this juncture, most analyses focused on trajectory of individual cell types. Nonetheless, it is clear that inter-cellular communications between cell types at any particular time point is important. New computational and experimental technology should be developed to further refine the cell-cell communication and changes in signalling pathways during embryonic development. In this context, incorporating the spatial arrangement of cells is particularly important. Leveraging and comparing single cell atlases across different studies, across developmental stages (embryonic vs adult) and even across species (e.g. human cell atlas versus Tabula muris), presents new opportunities for advanced bioinformatics methods for data integration. With advances in computational and experimental trajectory analysis, we are now developing a much better picture of embryonic development.

Competing interests

The authors declare no competing or financial interests.

Funding

P.P.L.T. is a Senior Principal Research Fellow of the National Health and Medical Research Council of Australia (1110751). This work is partially supported by the Hong Kong University-University of Sydney Strategic Partnership Fund (to J.W.K.H.).

References

Adey, A. C. (2019). Integration of single-cell genomics datasets. *Cell* **177**, 1677-1679. doi:10.1016/j.cell.2019.05.034

Argelaguet, R., Mohammed, H., Clark, S. J., Stapel, L. C., Krueger, C., Kapourani, C.-A., Xiang, Y., Hanna, C., Smallwood, S., Ibarra-Soria, X. et al. (2019). Single cell multi-omics profiling reveals a hierarchical epigenetic landscape during mammalian germ layer specification. *Nature* **576**, 487-491. doi:10.1038/s41586-019-1825-8

Arnold, S. J. and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91-103. doi:10.1038/nrm2618

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326. doi:10.1016/j.cell.2006.02.041

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J. et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502. doi:10.1038/s41586-019-0969-x

Cardoso-Moreira, M., Halbert, J., Volloton, D., Velten, B., Chen, C., Shao, Y., Leichti, A., Ascenção, K., Rummel, C., Ovchinnikova, S. et al. (2019). Gene expression across mammalian organ development. *Nature* **571**, 505-509. doi:10.1038/s41586-019-1338-5

Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., Jost, M., Quinn, J. J., Yang, D. and Jones, M. G. et al. (2019). Molecular recording of mammalian embryogenesis. *Nature* **570**, 77-82. doi:10.1038/s41586-019-1184-5

Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P. P. L., Jing, N. and Deng, Q. (2019). Single-cell RNA-seq reveals cellular heterogeneity of pluripotency transition and X chromosome dynamics during early mouse development. *Cell Rep.* **26**, 2593-2607.e3. doi:10.1016/j.celrep.2019.02.031

Ding, J., Lin, C. and Bar-Joseph, Z. (2019). Cell lineage inference from SNP and scRNA-Seq data. *Nucleic Acids Res.* **47**, e56. doi:10.1093/nar/gkz146

Eng, C.-H. L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C. et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235-239. doi:10.1038/s41586-019-1049-y

Graf, T. (2019). Transcription factor stoichiometry drives cell fate: single-cell proteomics to the rescue. *Cell Stem Cell* **24**, 673-674. doi:10.1016/j.stem.2019.03.002

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O. et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060-1066. doi:10.1038/s41588-019-0424-9

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F. et al. (2018). Mapping the mouse cell atlas by microwell-Seq. *Cell* **172**, 1091-1107.e17. doi:10.1016/j.cell.2018.02.001

HuBMAP Consortium. (2019). The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187-192. doi:10.1038/s41586-019-1629-x

Ibarra-Soria, X., Jawaid, W., Pijuan-Sala, B., Ladopoulos, V., Scialdone, A., Jörg, D. J., Tyser, R. C. V., Calero-Nieto, F. J. Mulas, C., Nichols, J. et al. (2018). Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20**, 127-134. doi:10.1038/s41556-017-0013-z

Kester, L. and van Oudenaarden, A. (2018). Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166-179. doi:10.1016/j.stem.2018.04.014

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., de Alencastro, G., Logan, G. J. et al. (2014a). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107-120. doi:10.1016/j.stem.2013.09.014

Kojima, Y., Tam, O. H. and Tam, P. P. L. (2014b). Timing of developmental events in the early mouse embryo. *Semin. Cell Dev. Biol.* **34**, 65-75. doi:10.1016/j.semcdb.2014.06.010

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A. et al. (2018). RNA velocity of single cells. *Nature* **560**, 494-498. doi:10.1038/s41586-018-0414-6

Lin, Y., Ghazanfar, S., Wang, K. Y. X., Gagnon-Bartsch, J. A., Lo, K. K., Su, X., Han, Z.-G., Ormerod, J. T., Speed, T. P., Yang, P. et al. (2019). scMerge leverages factor analysis, stable expression, and pseudoreplication to merge multiple single-cell RNA-seq datasets. *Proc. Natl. Acad. Sci. USA* **116**, 9775-9784. doi:10.1073/pnas.1820006116

McDole, K., Guignard, L., Amat, F., Berger, A., Malandain, G., Royer, L. A., Turaga, S. C., Branson, K. and Keller, P. J. (2018). In toto imaging and reconstruction of post-implantation mouse development at the single-cell level. *Cell* **175**, 859-876.e33. doi:10.1016/j.cell.2018.09.031

McKenna, A. and Gagnon, J. A. (2019). Recording development with single cell dynamic lineage tracing. *Development* **146**, dev169730. doi:10.1242/dev.169730

Mitiku, N. and Baker, J. C. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. *Dev. Cell* **13**, 897-907. doi:10.1016/j.devcel.2007.10.004

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J. et al. (2017). Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **20**, 1215-1228. doi:10.1016/j.celrep.2017.07.009

Mulqueen, R. M., Pokholok, D., Norberg, S. J., Torkenczy, K. A., Fields, A. J., Sun, D., Sinnamon, J. R., Shendure, J., Trapnell, C., O'Roak, B. J. et al. (2017). Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428-431. doi:10.1038/nbt.4112

Nowotschin, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C. et al. (2019). The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367. doi:10.1038/s41586-019-1127-1

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G. et al. (2016). Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* **36**, 681-697. doi:10.1016/j.devcel.2016.02.020

Peng, G., Suo, S., Cui, G., Yu, F., Wang, R., Chen, J., Chen, S., Liu, Z., Chen, G., Qian, Y. et al. (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528-532. doi:10.1038/s41586-019-1469-8

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L. et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495. doi:10.1038/s41586-019-0933-9

Rivera-Pérez, J. A. and Hadjantonakis, A.-K. (2014). The dynamics of morphogenesis in the early mouse embryo. *Cold Spring Harb. Perspect. Biol.* **7**, a015867. doi:10.1101/cshperspect.a015867

Rodriques, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., Welch, J., Chen, L. M., Chen, F. and Macosko, E. Z. (2019). Slideseq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467. doi:10.1126/science.aaw1219

Rooijers, K., Markodimitraki, C. M., Rang, F. J., de Vries, S. S., Chialastri, A., de Luca, K. L., Mooijman, D., Dey, S. S. and Kind, J. (2019). Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37**, 766-772. doi:10.1038/s41587-019-0150-y

Saelens, W., Cannoodt, R., Todorov, H. and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547-554. doi:10.1038/s41587-019-0071-9

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P. et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928-943.e22. doi:10.1016/j.cell.2019.01.006

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N. K., Macaulay, I. C., Marioni, J. C. and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. *Nature* **535**, 289-293. doi:10.1038/nature18633

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L. and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451-460. doi:10.1038/s41587-019-0068-4

Solnica-Krezel, L. and Sepich, D. S. (2012). Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* **28**, 687-717. doi:10.1146/annurev-cellbio-092910-154043

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., III, Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Tam, P. P. L. and Behringer, R. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3-25. doi:10.1016/S0925-4773(97)00123-8

Tam, P. P. L. and Gad, J. M. (2004). Chapter 16: Gastrulation of the mouse embryo. In *Gastrulation* (ed. C. D. Stern), pp. 223-262. New York, CSHL Press.

Tam, P. P. L. and Loebel, D. A. F. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368-381. doi:10.1038/nrg2084

The Tabula Muris Consortium. (2018). Single-cell transcriptomics of 20 mouse organs creates a *Tabula Muris*. *Nature* **562**, 367-372. doi:10.1038/s41586-018-0590-4

The Tabula Muris Consortium. (2019). A single-cell transcriptomic atlas characterizes aging tissues in the mouse. *BioRxiv*. doi:10.1101/661728

Tian, L., Dong, X., Freytag, S., Lê Cao, K.-A., Su, S., JalalAbadi, A., Amann-Zalcenstein, D., Weber, T. S., Seidi, A., Jabbari, J. S. et al. (2019). Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nat. Methods* **16**, 479-487. doi:10.1038/s41592-019-0425-8

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnam, C., Liu, J. et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Sciences* **361**, 380-388. doi:10.1126/science.aat5691

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C. and Macosko, E. Z. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873-1887.e17. doi:10.1016/j.cell.2019.05.006

Wen, J., Zeng, Y., Fang, Z., Gu, J., Ge, L., Tang, F., Qu, Z., Hu, J., Cui, Y., Zhang, K. et al. (2017). Single-cell analysis reveals lineage segregation in early post-implantation mouse embryos. *J. Biol. Chem.* **292**, 9840-9854. doi:10.1074/jbc.M117.780585

Xu, J., Nuno, K., Litzenburger, U. M., Qi, Y., Corces, M. R., Majeti, R. and Chang, H. Y. (2019). Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**, e45105. doi:10.7554/eLife.45105