The Company of
Biologists

REVIEW

Recording development with single cell dynamic lineage tracing

Aaron McKenna[1],* and James A. Gagnon[2,3],*

ABSTRACT

Every animal grows from a single fertilized egg into an intricate network of cell types and organ systems. This process is captured in a lineage tree: a diagram of every cell's ancestry back to the founding zygote. Biologists have long sought to trace this cell lineage tree in individual organisms and have developed a variety of technologies to map the progeny of specific cells. However, there are billions to trillions of cells in complex organisms, and conventional approaches can only map a limited number of clonal populations per experiment. A new generation of tools that use molecular recording methods integrated with single cell profiling technologies may provide a solution. Here, we summarize recent breakthroughs in these technologies, outline experimental and computational challenges, and discuss biological questions that can be addressed using single cell dynamic lineage tracing.

KEY WORDS: Lineage tracing, Lineage trees, Single cell

Introduction

The development of plants and animals has long been a source of fascination to those interested in the construction of complex organisms. In particular, researchers have sought to understand the collective pattern of cell divisions – termed a 'cell lineage tree' (Fig. 1A) – that transforms a single-celled zygote into a fully-formed adult organism. Two strategies to map cell lineage trees have emerged: imaging-based approaches and genetic approaches. Advances in imaging-based lineage tracing have been driven by technological progress in microscopy throughout the 20th century. Such advances include the work of Nicole Le Douarin, who generated quail/chicken chimeras, and the dye labeling and transplantation experiments carried out in frogs and fish by Spemann and others (Kretzschmar and Watt, 2012). The culmination was the generation of the complete *C. elegans* lineage tree by Sir John Sulston and colleagues (Sulston et al., 1983). More recently, advances in fluorescent protein engineering, and in confocal and light sheet microscopy, have enabled the lineage tracing of single cells within intact, living animals. This branch of lineage tracing has been recently reviewed (Kretzschmar and Watt, 2012; Liu and Keller, 2016). Suffice to say, these tools have advanced significantly in recent years, recording development through time with incredible cellular, and even subcellular, resolution (Liu et al., 2018; McDole et al., 2018). These methods all record lineages in living animals or tissues, and retain the spatial context that is so essential for understanding the growth and

maintenance of animals. However, imaging-based approaches are limited with regard to their temporal and molecular resolution. For example, animal development often occurs over timescales that make imaging impractical, and physiological growth conditions may be impossible to recreate under a microscope. It also remains difficult to capture detailed molecular information, such as the transcriptional state of individual cells, in conjunction with imaging-based lineage measurements, although recent advances in the scale of *in situ* imaging techniques are rapidly upending this assumption (Codeluppi et al., 2018 preprint; Shah et al., 2018; Wang et al., 2018; Rodriques et al., 2019; Eng et al., 2019).

Genetic or epigenetic changes can be used as an alternative to imaging-based techniques to infer lineage. A number of research groups have leveraged these naturally occurring changes to trace lineages, both in development and in cancer. Pioneering work in the 1960s, for example, used the silencing of specific alleles by X-chromosome inactivation to show the clonal nature of cancer (Linder and Gartler, 1965). Recent approaches leverage more abundant and variable genomic marks. Rare mutations, such as microsatellites and single nucleotide somatic variants, arise during development, and ensure that each cell within an individual has a slightly different genome (Fig. 1B). Relatedness between cells from the same individual can be determined by sequencing and comparing these mutations (Behjati et al., 2014; Brody et al., 2018; Carlson et al., 2012; Frumkin et al., 2005; Lodato et al., 2015; Salipante and Horwitz, 2006). However, these mutations are rare and scattered across the genome, necessitating genome-wide recovery approaches, and may not be ideal for many biological questions (Woodworth et al., 2017). Two recent studies used such an approach to quantify the number and dynamics of hematopoietic stem cells (HSCs) within single patients, taking advantage of the ability of HSCs to expand in culture (Lee-Six et al., 2018; Osorio et al., 2018). Recent cost reductions in genome sequencing and advances in variant calling allow these approaches to be scaled to the whole genomes of individual cells (Lodato et al., 2015). However, the unbiased organism-wide profiling of single-cell whole genomes is currently impractical. In the future, the signal from these somatic mutations could be combined with orthogonal lineage information from mitochondrial ATAC-seq reads or DNA methylation marks (Ludwig et al., 2019; Walther and Alison, 2016; Xu et al., 2019; Salas et al., 2018).

An alternative approach is to engineer 'lineage recorders' into the genome. Reporter systems, for example, have long been used to express a fluorescent marker in specific cell types (Fig. 1C). The fluorescent marker can be permanently activated by a recombinase or triggered by an external stimulus, such as heat shock or a drug. The marker is then passed on to descendant cells, and tissues or individual cells can later be examined for fluorescence, thereby allowing lineages to be traced. Moreover, the combinatorial expression of fluorescent proteins in systems such as Brainbow can label, and therefore trace, a larger number of clonal populations (Hadjieconomou et al., 2011; Livet et al., 2007; Pan et al., 2013). Cellular barcoding approaches, by contrast, employ a diverse pool of DNA tags ('barcodes'; see Glossary, Box 1) to mark individual cell lineages (Fig. 1D) (Kebschull and Zador, 2018). These approaches

[1]Department of Molecular and Systems Biology, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA. [2]Center for Cell and Genome Science, University of Utah, Salt Lake City, UT 84112, USA. [3]School of Biological Sciences, University of Utah, Salt Lake City, UT 84112, USA.

*Authors for correspondence (aaron.mckenna@dartmouth.edu; james.gagnon@utah.edu)

A.M., 0000-0001-8277-6512; J.A.G., 0000-0003-3978-6058

DEVELOPMENT

1

> **Box 1. Glossary**
> **Barcode.** A compact DNA sequence used to record lineage or other information, often composed of multiple target sites.
> **Dynamic lineage tracing.** Recording lineage information using barcodes integrated in the genome, often via genome editing or recombinase activity.
> **Edits/editing.** The error-prone repair of a CRISPR-Cas9 double-stranded break at a target site, often resulting in insertions or deletions of DNA.
> **GESTALT.** Genome editing of synthetic target arrays for lineage tracing (McKenna et al., 2016).
> **LINNAEUS.** Lineage tracing by nuclease-activated editing of ubiquitous sequences (Spanjaard et al., 2018).
> **Polylox.** An artificial DNA recombination locus based on the Cre-*loxP* system (Pei et al., 2017).
> **scGESTALT.** Single cell RNA sequencing with GESTALT (Raj et al., 2018).
> **ScarTrace.** Genome editing of engineered GFP targets coupled to recovery of single cell transcriptomes (Alemany et al., 2018; Junker et al., 2017).

overcome some of the limitations imposed by somatic mutation-tracing methods by vastly reducing the amount of material that must be recovered from sequencing of single cells. In addition, these static marks are retained by progeny and can be recovered from terminal cells. However, most cellular barcoding techniques are limited by the requirement that the DNA barcodes must be efficiently inserted into the genome of a population of cells. Owing to these integration constraints, cellular barcoding is currently limited to *ex vivo* studies, transplanted cells, viral integration or zygote microinjection. Additionally, it can be challenging to infer a branching lineage using either fluorescent reporters or cellular barcoding.

A more recently developed approach, which we term 'dynamic lineage tracing' (see Glossary, Box 1), aims to diversify barcode sequences in cells during development, allowing the generation of branching lineage trees (Fig. 1E). These approaches can generate thousands of evolving sequences that can be related to each other by shared mutations (Alemany et al., 2018; Frieda et al., 2017; Junker et al., 2017 preprint; McKenna et al., 2016; Pei et al., 2017; Peikon et al., 2014; Spanjaard et al., 2018). For example, CRISPR-Cas9 can

be directed to make double-stranded breaks in the genome. These breaks are then repaired by cellular machinery, sometimes in an error-prone way, resulting in insertions or deletions at the target locus. The resulting random 'edits' (see Glossary, Box 1) can differentiate one cell's copy of the CRISPR target sequence from another. Although the length and composition of these repairs is generally stochastic, some sequences generate a greater variation of repair sequences than others (Allen et al., 2018; Chen et al., 2018 preprint; Shen et al., 2018). A number of these CRISPR targets can then be grouped together as a barcode in a small locus in the genome, or scattered across many loci across the genome. These dynamic barcodes are also compact, and can be coupled with emerging single-cell techniques to simultaneously capture cell transcriptional state. A related strategy uses recombinases to generate barcode diversity (Pei et al., 2017; Peikon et al., 2014). In this approach, recombinase enzymes invert or excise a fragment of DNA flanked by two recognition sites. A huge diversity of DNA sequence combinations can thus be generated using this simple shuffling procedure and a large number of recognition sites closely packed into a barcode. Generally, only one locus is used in these recombinase strategies to avoid the excision of large stretches of DNA through engineered chromosome loss (Lewandoski and Martin, 1997). Notably, both CRISPR and recombinase-based dynamic lineage tracing approaches are prospective in nature; the recording construct has to be integrated into the genome of the organism, precluding their use in humans for ethical reasons.

All of the lineage tracing methods mentioned above have advantages, drawbacks and applications in modern developmental biology. In this Review, we focus on dynamic lineage tracing, which can be used to record deep branching lineage relationships. As with most new technologies, dynamic lineage tracing must overcome many hurdles to move past the demonstration phase to widespread acceptance and utility. Here, we discuss the experimental design, implementation and analytical challenges associated with dynamic lineage tracing, as well as its application to open questions in biology.

Barcode selection and integration
Dynamic lineage tracing approaches record information at a predefined location in the genome, commonly referred to as a
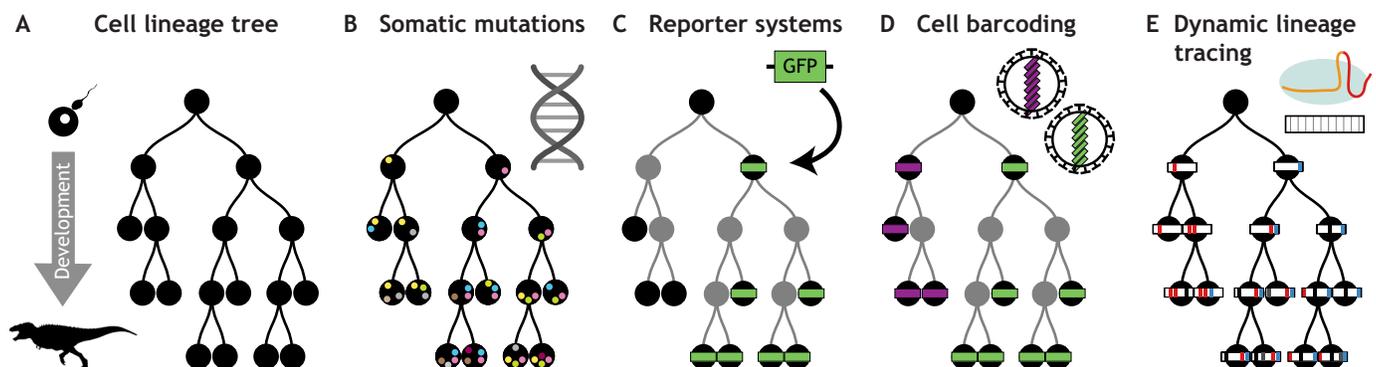

Fig. 1. Lineage tracing in development. (A) The cell lineage tree describes the successive cell divisions that generate all of the organs and tissues within an organism. Lineage tracing aims to capture these cellular relationships. (B) Rare mutations (marked by colored dots) occur sporadically across the genome during development. These mutations can be used to trace complete lineage trees, but require whole- or partial-genome sequencing. (C) A classic labeling approach is to tag a subset of cells with a dye or fluorescent marker such as GFP (green); this approach is good for tagging all descendants of the marked cell but does not reveal relationships within the marked population. (D) Cellular barcoding, using viruses or transposons, can label a population of cells with unique identifiers (purple and green). Descendant cells can then be assigned to common progenitors. These approaches can capture clonal relationships but cannot infer relationships within each subpopulation. (E) Dynamic lineage-tracing approaches add increasing information over developmental time. These approaches can assign cells to progenitors and determine branches within the lineage tree.

barcode. Barcodes contain one or more DNA elements that can be manipulated to record information. For example, CRISPR lineage-tracing barcodes can have one or more target sites within each barcode, while recombinase barcodes can have multiple pairs of recognition sites, all in a relatively short barcode segment. The number of barcodes and their mode of insertion into the genome have significant implications for both the downstream analysis and the ability to generate and maintain the respective genetically engineered animals. Single integration approaches can take advantage of identified 'safe harbor' locations in the genome, where barcodes can be integrated without perturbing cellular function, such as the ROSA26 locus in mice and phiC31 in flies (Bischof et al., 2007; Sadelain et al., 2011). Other strategies use multiple genomic barcodes integrated semi-randomly throughout the genome, generally with transposon or viral integration strategies. These multiplexed techniques have, in aggregate, a greater storage capacity but also require the recovery of multiple genomic targets (discussed below). In addition, as the different chromosomes carrying these integrations will segregate in progeny, multiple barcode approaches require breeding strategies that retain sufficient integrations in an animal line (Kalhor et al., 2018; Chan et al., 2018). Random barcode integration can also disrupt existing genes, leading to unintended phenotypes; indeed, it is known that lentiviral integration is biased towards open chromatin, i.e. active stretches of DNA (Demeulemeester et al., 2015). An alternative approach is to engineer all aspects of the recording construct into an artificial chromosome system (Oshimura et al., 2015). Such a system would allow all recording components to be integrated in a single step, and would be compatible with a number of mouse models. However, it is unclear how well these synthetic chromosomes would be retained by cells when they are repeatedly targeted for double-stranded breaks by enzymes such as CRISPR, or whether barcode diversity rates would be similar to those seen with endogenous chromosomes.

Once the barcode or barcodes are encoded in the genome, the next step is to activate the recording of lineage information. This usually entails delivering the recombinase or CRISPR system into cells to trigger barcode modification. To begin recording from the single cell zygote, CRISPR components, such as Cas9 mRNA or protein, and/or guide RNAs, can be injected directly into the founding zygote or egg (Alemany et al., 2018; Junker et al., 2017 preprint; McKenna et al., 2016; Spanjaard et al., 2018). Recording can begin almost immediately in all embryonic cells, but eventually the injected CRISPR components are diluted or degraded, constraining lineage recording to the early stages of development. A parallel approach has used mismatched guide RNAs to slow the rate of editing to better mirror the slower cell cycles observed in the early mouse embryo (Chan et al., 2018). Alternatively one can drive expression of a critical component such as the recombinase or the Cas9 enzyme using ubiquitous promoters integrated into the genome (Chan et al., 2018; Kalhor et al., 2018; Pei et al., 2017). However, these approaches often require additional breeding, line generation and assessment of the accuracy and recording efficiency. A more-targeted lineage recording approach can be taken when the biological question is more focused. For example, promoters that are active in specific cell types can be used to drive recording, restricting lineage recording to a particular subpopulation. Alternatively, environmental stimuli can be used to drive expression by coupling recorder components to drug- or heat-inducible promoters (Das et al., 2016; Raj et al., 2018). Such recorder constructs can be activated at various time points, providing a real-time dimension to lineage recordings (Raj et al., 2018). The two approaches can also be combined, with subsets of recording constructs being active during early development while others record later cell divisions.

Barcode recovery

In all dynamic lineage-tracing strategies, the lineage barcodes must be recovered from the cells of interest, i.e. the information collectively encoded in the barcode(s) must be captured and assigned to each individual cell (Fig. 2). There are two major recovery strategies in dynamic lineage-tracing systems: DNA sequencing and fluorescent probe hybridization. The choice of which recovery strategy to use is linked to the initial design of the recording construct.

Short-read sequencing is used for both static cell barcoding and CRISPR-based approaches that have a smaller DNA footprint (Fig. 2A,C,E) (Alemany et al., 2018; Junker et al., 2017 preprint; McKenna et al., 2016; Raj et al., 2018; Spanjaard et al., 2018). This approach was used in the first generation of dynamic lineage-tracing approaches, although they were limited in their resolution: lineage information was recovered from bulk DNA, precluding single cell approaches. These studies focused instead on entire embryos, dissected organs or sorted cell populations. The expression of lineage recordings as messenger RNA (mRNA) can allow capture alongside the transcriptome via single cell RNA sequencing. This technical advance was recently achieved using related approaches in zebrafish and mouse embryos (Alemany et al., 2018; Chan et al., 2018; Raj et al., 2018; Spanjaard et al., 2018). While each approach differed with regard to the details of the barcode and the mode of CRISPR delivery, they all captured the transcriptome and lineage recordings simultaneously from single cells using variants of single cell RNA sequencing. However, it is already clear that capturing the lineage barcode from all cells will be a major technical constraint; the methods published so far fail to capture barcodes from every cell. For example in the study by Raj et al., barcode transcripts were recovered from only 6-28% of zebrafish cells. This loss is potentially due to some combination of: (1) insufficient transcription, either in a cell type-specific manner or from transcriptional bursting; (2) poor RNA stability, polyadenylation and/or capture during library preparation; (3) the stochastic nature of amplification in single cell RNA sequencing (Shapiro et al., 2013); or (4) loss of the PCR primer handles from long deletions (Kosicki et al., 2018). Furthermore, even when a barcode is captured, intersite deletions may erase lineage information (McKenna et al., 2016; Salvador-Martínez et al., 2019). Simulations have shown that this can impact the accuracy of downstream tree reconstruction, having a larger effect as the number of target sequences increases within the barcode (Salvador-Martínez et al., 2019). Increasing the number of barcodes, each with fewer targets, lowers the chance of intersite deletions and increases the chance of partial barcode recovery, but also limits complete capture of all barcodes (Alemany et al., 2018; Junker et al., 2017 preprint; Spanjaard et al., 2018). It is possible that higher expression or stabilization of RNA species could increase capture (Horstick et al., 2015; Rabani et al., 2017), or perhaps new advances in single cell-based approaches will be required.

Recombinase approaches generally require longer DNA constructs, as pairs of recombinase sites and their intervening sequence are longer than CRISPR targets. As such, these barcodes must be recovered using long-read sequencing technologies such as PacBio sequencing (Fig. 2B) (Pei et al., 2017; Peikon et al., 2014). This currently limits their integration with single cell methods and limits the number of barcodes that can be recovered. However, the advantage of these recombinase-based recording constructs is that
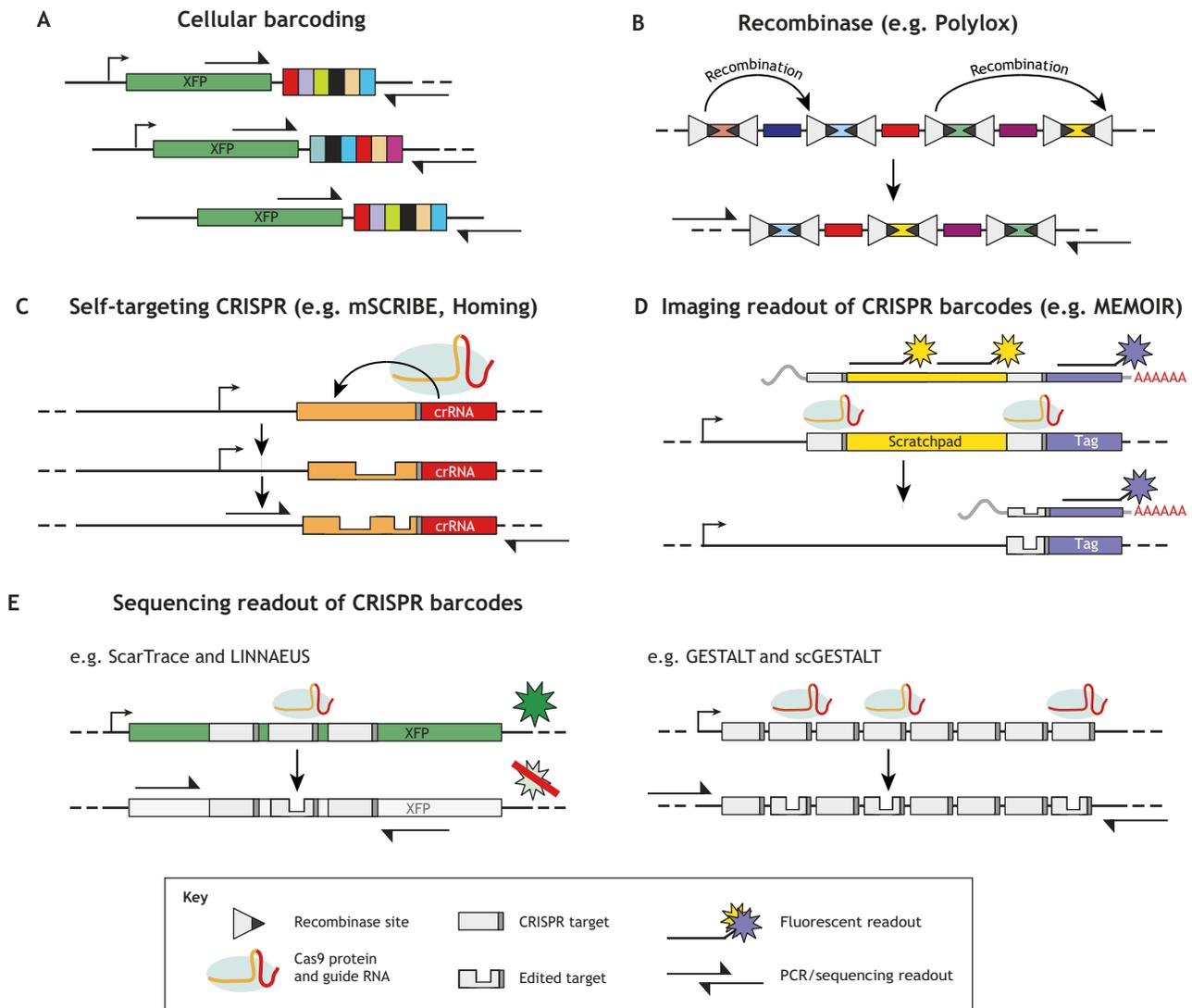
Fig. 2. DNA barcoding and recovery using prospective lineage recording. (A) Cellular barcoding introduces a diverse population of nucleotide tags into the genome, typically using viral integration or transposition. These static barcodes can be read by PCR and short-read sequencing. (B) Barcode diversity can also be generated with recombinases, whereby an array of DNA-binding sites can be scrambled when the recombinase is expressed. These barcodes are typically read with long-read sequencing due to their large size (see Peikon et al., 2014). (C) Self-targeting CRISPR guide RNAs target Cas9 editing to their own genomic sequence, resulting in repeated rounds of mutation that can generate barcode diversity. The resulting sequences can be read using short-read sequencing. (D) With MEMOIR, genetic regions termed 'scratchpads' are inserted into the genome and collapsed by CRISPR editing. The resulting scratchpads are expressed as mRNA and detected with fluorescent *in situ* hybridization. (E) Alternatively, multi-target barcodes – in one or more locations in the genome – can be edited with a single guide RNA or distinct guide RNAs to incrementally add information to the barcode during development. There barcodes are typically recovered using short-read sequencing.

they may provide predictable hybridization sequences for compatibility with existing *in situ* sequencing approaches (Frieda et al., 2017). Moreover, the inversion or deletions generated using recombinases resolve without creating double-stranded breaks, avoiding large unintended deletions, which can confound recovery in CRISPR systems (Kosicki et al., 2018).

Fluorescent probe hybridization has been adapted to recover barcodes *in situ* using a visual readout. For example, an approach termed MEMOIR (memory by engineered mutagenesis with optical *in situ* readout) leverages single molecule RNA fluorescence hybridization to recover multiple lineage barcodes from intact cells (Frieda et al., 2017). MEMOIR barcodes are expressed as mRNA and are composed of two parts – a unique tag that identifies each integration and a shared set of ten identical CRISPR targets called a scratchpad (Fig. 2D). When Cas9 and the guide RNA are expressed,

large deletions collapse these scratchpads stochastically over time. Barcode recovery involves hybridizing to the tag sequence to identify integrations, and hybridization to the common scratchpad to determine whether it is intact or deleted. Recovery via this approach is exceptional; 90% barcode signal recovery is observed in embryonic stem cells (Frieda et al., 2017). The challenges are now to intersect this method with transcriptome recovery and to automate barcode recovery for hundreds of thousands of cells. Regardless of the approach – sequencing or imaging – the challenge of large-scale lineage recovery will be an important consideration in future studies.

Lineage tree construction
The next step is to infer relationships between the recovered barcodes and build a lineage tree. This task has enormous computational challenges [reviewed by Tritschler et al., 2019

(also in this issue)]. For *n* sampled cells, there is a super-exponential number of ways to build a tree: over 34 million different rooted trees can be made from only 10 terminal cells (Felsenstein, 1978). The first step is to define a distance metric that scores relatedness between barcoded cells. This can be challenging, given the incomplete and noisy nature of the dynamic lineage-tracing barcodes, with issues such as incomplete capture, sequence alignment problems and intersite deletions in CRISPR approaches, or barcode excision in recombinase strategies. Some methods, such as the recombinase strategy Polylox (see Glossary, Box 1), enumerate all possible barcodes and assign a probability of the unaltered barcode recombining to each terminal state (Pei et al., 2017); such an approach could also be adapted to describe the distance between recombined barcodes. CRISPR approaches tend to compare the identity of editing events shared between two barcodes, treating each insertion or deletion as an independent event (Alemany et al., 2018; Chan et al., 2018; McKenna et al., 2016).

The next task is to convert these barcode relationships into a tree of cell divisions. Custom algorithms have been developed to achieve this based on maximum likelihood (Chan et al., 2018) or network graphs (Spanjaard et al., 2018). Others have adapted methods from the field of phylogenetics (see Box 2), which has a long history of addressing conceptually similar problems and has developed a number of techniques and approximations to tackle the complexity of tree construction (St. John, 2017). However, the data collected by these new lineage-tracing techniques upend several assumptions that make phylogenetics more tractable, and care should be taken when adapting these methods (see Box 2; Feng et al., 2019 preprint).

A common reconstruction technique borrowed from phylogenetics is maximum parsimony, which minimizes the number of changes (mutations or edits) over the tree as a whole. As the resulting branch lengths are equal to the number of mutations, this algorithm can simplify tree construction. However, it offers little information about the true developmental time between cell divisions. Care must also be taken to ensure that the underlying model assumptions are considered. For example, certain phylogenetic models, such as Camin-Sokal parsimony, assume an unaltered starting barcode and an irreversible transition from the unedited to the edited targets. When making comparisons with new emerging methods, it is important to consider these assumptions (Spanjaard et al., 2018). Other challenges include homoplasy, where identical marks are acquired independently on different branches of a tree. This is relevant for both CRISPR-Cas9 methods, where certain sequences produce stereotypical outcomes (Allen et al., 2018; Chen et al., 2018 preprint; Frieda et al., 2017; Shen et al., 2018; van Overbeek et al., 2016), and for recombinase methods, where there are a low number of potential outcomes (Pei et al., 2017; Peikon et al., 2014). If not dealt with, homoplasy can lead to incorrect relationships being assigned to unrelated cells. The field of tumor phylogenetics has already adapted many of these methods for the analysis of tumor evolution, and the reuse of strategies developed therein could benefit the lineage reconstruction field (Schwartz and Schäffer, 2017).

Last, the trees generated by the composite experimental and computational methods described above must be validated against known biology. Dynamic lineage tracing has recovered known cell/lineage relationships during germ layer patterning (McKenna et al., 2016), in regions of the brain (Raj et al., 2018) and within progenitors in the hematopoietic system (Alemany et al., 2018); but what resolution can we expect from these new lineage-tracing trees? Simulations offer insight (Frieda et al., 2017; McKenna et al., 2016; Salvador-Martínez et al., 2019), but experimental validation will also be important. Future benchmarking efforts that combine and

Box 2. From phylogenetics to lineage reconstruction

Many lineage-tracing tools borrow techniques and terminology from phylogenetics – the study of evolutionary history and relationships between biological entities. However, it is important to be aware of the assumptions made by phylogenetic models when adapting them for lineage tracing (Yang and Rannala, 2012). Phylogenetic studies based on nucleotide changes typically examine a limited number of species and appreciably more sites (i.e. nucleotides in a gene or genome). These data together represent a wide and short matrix (see Figure, top). Dynamic lineage tracing, by contrast, captures thousands of cells and their corresponding barcode sequences. These barcodes are made up of a limited number of mutable sites, leading to a tall and narrow matrix (see Figure, bottom). These differences make it difficult to directly apply phylogenetic methods designed for wide and short matrices to lineage-tracing data.

Dynamic lineage-tracing data also break several assumptions beyond these structural differences. First, CRISPR target sites can have highly variable mutability rates, with some sites highly mutated and others completely intact (1 in Figure). Mutations are also 'locked in' such that deletions or insertions prevent further recording. Phylogenetic models such as the general time reversible (GTR) model assume that nucleotide substitutions are reversible. Some phylogenetic models also assume that substitution rates are identical or very similar, and all models need at least a rate estimate of nucleotide substitutions. However, DNA repair can favor certain repair outcomes over others, and this can change across organisms or developmental timepoints (Haeussler et al., 2016). Second, large deletions can destroy PCR primer sites that flank the barcode, leading to biased dropout of data from subpopulations of cells (2 in Figure). Third, in multi-site barcodes, large deletions between two sites may erase previous mutations (striped deletion, 3 in Figure). In both single- and multiple-target barcodes, there is also a chance that a target is edited again, destroying a previous signal. Fourth, independence between mutation sites is a core assumption for many phylogenetic methods, which is violated in many CRISPR lineage-tracing systems where large deletions can ablate numerous targets within a single event (4 in Figure). These unique challenges present an opportunity for new tree building techniques, such as maximum likelihood models where branch lengths can indicate developmental time, and for computational strategies that integrate the resulting trees with other single cell data.
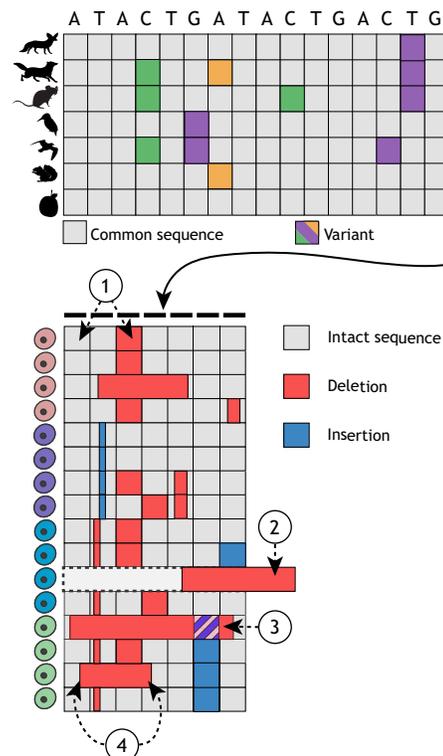
contrast multiple lineage-tracing methods in well-understood biological systems, or in model organisms such as *C. elegans* that have a known lineage, will provide the most comprehensive assessment (Packer et al., 2019 preprint).

Biological insights obtained from recent single cell dynamic lineage-tracing studies

The new tools discussed above have the potential to transform our understanding of tissue and organ development, both in terms of embryogenesis and homeostasis, and in the context of regeneration and disease. Below, we focus on two broad classes of biological questions that can be addressed with single cell lineage tracing: building the animal and maintaining the adult. First, we discuss how combining lineage tracing with single cell sequencing can allow historical annotation of embryonic cell fate specification and commitment. Second, we highlight how these approaches could resolve controversies over the differentiation potential of stem and progenitor cells within normal and regenerating tissues.

Building the animal: cell types and fate specification

When and where are cell types specified in the developing embryo? Decades of embryological observation and manipulation have given us a rough draft, but dynamic lineage tracing now offers single cell granularity of the molecular processes that govern differentiation. With the advent of massively parallel methods for single cell RNA sequencing of tissues, organs and whole animals, we will soon know the complete parts list of cells for constructing complex animals. A challenge for interpreting these atlases, however, is to understand how and where cells are made, and how they interact to form complex systems.

Three strategies have been used to infer paths of cellular differentiation from single cell atlases. First, when the sample contains both the progenitor and differentiated cell types, it is often possible to construct trajectories that describe cell-type differentiation (Bendall et al., 2014; Shin et al., 2015; Trapnell et al., 2014; see Tritschler et al., 2019 in this issue). Second, single-cell atlases of multiple time points can be stitched together into developmental trajectories that describe cell fate acquisition across developing embryos (Briggs et al., 2018; Cao et al., 2019; Farrell et al., 2018; Pijuan-Sala et al., 2019; Wagner et al., 2018). Third, the integration of atlases with molecular imaging or tissue sectioning [reviewed by Mayr et al., 2019 (in this issue)] allows inference of the spatial organization of cells within tissues and embryos (Achim et al., 2015; Junker et al., 2014; Karaiskos et al., 2017; Satija et al., 2015; Ståhl et al., 2016).

Dynamic lineage tracing can serve as a parallel and complementary approach to trajectory inference by providing a ground truth framework for interpreting atlases. For example, single cell lineage tracing can identify cells with the same transcriptional state but with divergent origins. This is particularly important when studying cells of similar identity in spatially discrete regions of the embryo (e.g. eye development or somitogenesis) or for dissecting the contributions of individual cells within a population of a given cell type (e.g. HSCs or other stem cells). As an example, an approach called Tracer-seq used cellular barcoding in zebrafish to trace the contribution of embryonic cells to various cell types (Wagner et al., 2018). When overlaid onto a developmental trajectory of zebrafish embryogenesis, these barcodes traced explicit lineage relationships between cells. A similar strategy has been used to trace hematopoiesis in mice (Weinreb et al., 2018 preprint).

Dynamic lineage tracing can also record cell divisions after the tags have been introduced, thereby delineating the relationships

between descendants (Fig. 1D). For example, CRISPR lineage tracing has been implemented in zebrafish and worms (Junker et al., 2017 preprint; McKenna et al., 2016; Schmidt et al., 2017), and more recently in mice (Chan et al., 2018; Kalhor et al., 2018), to capture early regional specification events during embryogenesis. This includes the initial segregation of multipotent cells into mesoderm, endoderm and ectoderm, and the diversification of these germ layers into more specialized precursor cells. These approaches generally sampled whole organs from adult animals. Barcodes found within organs and tissues clustered more closely with others from the same germ layer instead of across germ layers. These barcodes also allowed tracing of the entire zebrafish hematopoietic system back to a handful of pre-gastrulation embryonic cells (Junker et al., 2017 preprint; McKenna et al., 2016).

These approaches can also be used in more targeted studies. One such application labeled embryonic progenitors in the zebrafish tail bud, providing insights to the cell fate decision between neural and mesodermal fates (Attardi et al., 2018). Another study applied CRISPR lineage tracing to discover how the spatial coordinates of the mouse brain are established, determining that the anterior-posterior axis is formed earlier in development than that of the left-right axis (Kalhor et al., 2018). Finally, these approaches can be combined with classic genetic and cell transplantation techniques. For example, genetically mutant zebrafish cells labeled with CRISPR-edited barcodes have been transplanted into host embryos to rule out a cell-autonomous requirement for the transcription factor Nanog during germ layer specification (Gagnon et al., 2018).

The intersection of CRISPR lineage tracing and single cell RNA sequencing has further improved the resolution of these studies. In an approach termed scGESTALT (see Glossary, Box 1), the lineage relationships within a single cell atlas of the juvenile zebrafish brain were defined (Raj et al., 2018). Using editing at both embryonic and larval stages, it was shown that progenitors remain spatially restricted to distinct regions of the growing brain, with minimal cell movement between these regions. Furthermore, the generation of lineage trees that relate different cell types within in the brain uncovered a novel lineage branchpoint that generates distinct cell types in the hypothalamus. In a similar vein, the LINNAEUS technique (see Glossary, Box 1) was used to perform embryonic editing in zebrafish to study the relationships between cell types in various organs (Spanjaard et al., 2018). Specifically, this method was used to reveal lineage hierarchies that describe aspects of organogenesis in the heart, liver and telencephalon. Moreover, the combination of lineage and cell type information in this context allowed the authors to define distinct lineages that contribute to the endocardium and myocardium in the heart. In another recent study, embryonic CRISPR editing and single cell transcriptional state profiling was used to define the split of embryonic neuroectoderm into the left and right aspects of the brain and eyes, providing an example of how these tools can provide spatial information missing from developmental trajectories alone (Alemany et al., 2018).

Lineage tracing can also uncover convergent events in developmental lineages, where distinct lineage paths lead to a shared cell type. For example, delays and detours along developmental trajectories were recently identified within single-cell atlases of zebrafish and mouse embryogenesis (Cao et al., 2019; Farrell et al., 2018; Pijuan-Sala et al., 2019; Wagner et al., 2018). Cells in several developmental trajectories (axial mesoderm, neural plate/lateral plate, endoderm and others) appeared to differentiate based on their molecular identity but later converged into a common transcriptional state. Convergence of lineages to a common cell type is also seen in the heart: cardiomyocytes have embryonic origins in

both the mesoderm and neural crest (Keyte and Hutson, 2012), with consequences for adult heart function (Abdul-Wajid et al., 2018). The application of CRISPR lineage tracing in mouse embryos has also described the convergence of discrete embryonic and extra-embryonic lineages into a common endodermal cell type (Chan et al., 2018). Although these studies have uncovered new and exciting insights into developmental trajectories, they also raise a number of questions that pave the way for future studies. For example, how common are these convergences in differentiation? Are the resulting cells equipotent and, if so, what are the implications for the formation, maintenance and regeneration of healthy tissues?

Maintaining the adult: stem and progenitor cell potential

Lineage tracing can also help us understand the developmental potential of cells within adult organs, a topic of great interest in recent years. For practical reasons, hematopoiesis has served as a testbed for technological advances in clone tracing and cellular barcoding (Yamamoto et al., 2018). Indeed, studies of hematopoiesis established the paradigm of dedicated multipotent stem cells at the top of a differentiation hierarchy. The transplantation of barcoded HSCs back into host animals uncovered clonal dynamics between transplanted cells, highlighting variable contributions to different branches of the blood lineage tree and a change in the composition of the bone marrow over time (Carrelha et al., 2018; Lu et al., 2011; Naik et al., 2013). However, these transplanted cells may not recapitulate native stem cell activity. A creative solution to this problem is to use random transposon insertions to uniquely mark endogenous HSCs in an animal. Such an approach has been used to trace population dynamics and cell contributions during native hematopoiesis in the mouse, defining population clonal dynamics over time and discrete lineage contributions between HSCs (Rodriguez-Fraticelli et al., 2018; Sun et al., 2014). In a parallel approach, the Polylox system was applied to mouse hematopoiesis to delineate lineage hierarchies of blood development (Pei et al., 2017).

Dynamic lineage tracing provides a complementary approach for understanding lineage hierarchies in hematopoiesis and in other stem and progenitor populations. For example, CRISPR-based lineage tracing has been applied to understand the lineage connections between progenitor and differentiated cells in zebrafish: tracing clones of embryonic cells using GESTALT (see Glossary, Box 1) suggested that clonal dynamics might drive population drift within all adult tissues and organs, not just blood (McKenna et al., 2016). More recently, scGESTALT has been

applied to the developing brain to trace the continuous production of differentiated cells from pools of neural and oligodendrocyte precursors (Raj et al., 2018). Another method, termed ScarTrace (see Glossary, Box 1), has been used to identify immune cells in the zebrafish fin, showing that they have an embryonic origin distinct from the rest of the haematopoietic system (Alemany et al., 2018). These approaches can be applied in virtually any tissue or organ, and in the future may clarify controversies of whether there are 'professional' stem cells outside of the blood (Clevers and Watt, 2018). Is it possible that other cell types moonlight as stem or progenitor cells? Although examples of multipotent stem cell potential exist in the skin, intestine, liver, lung and other tissues, this activity often cannot be directly ascribed to a dedicated stem cell, and instead is found distributed across a population of progenitor cells or emerges after injury (Merrell and Stanger, 2016). New lineage-tracing tools will allow researchers to better label cells and track their contributions to tissue homeostasis in normal and diseased states.

Single-cell dynamic lineage tracing could also be applied to studies of regeneration (Mokalled and Poss, 2018; Rajagopal and Stanger, 2016). A fundamental question of regeneration is whether there are regeneration-specific lineage programs, or whether regeneration can be seen as the recapitulation of normal organ development. In the zebrafish caudal fin, a well-studied regeneration system, a consensus has emerged to suggest that the blastema (the region of regenerating tissue) contains few if any multipotent dedifferentiated cells. However, osteoblasts retain some potential to dedifferentiate and contribute to other cell types in a regeneration-specific manner, a finding that was initially demonstrated with classic lineage tracing and recently confirmed using dynamic lineage tracing (Alemany et al., 2018; Junker et al., 2017 preprint; Stewart and Stankunas, 2012; Tornini et al., 2017; Tu and Johnson, 2011). A blastema is also formed from adult cells during limb regeneration in axolotls (Haas and Whited, 2017; Tanaka, 2016). Recently, time course single cell RNA sequencing was applied to this process to reveal cells that dedifferentiate into embryonic-like cells that exhibit multipotency during limb regeneration (Gerber et al., 2018; Leigh et al., 2018). Together with a second study using CRISPR lineage tracing (Flowers et al., 2017), these data argue against professional regenerative stem cells and instead support a model in which adult cells can dedifferentiate and proliferate during axolotl regeneration.

Overall, these studies highlight promising applications of dynamic lineage tracing to monitor the contributions of all cells during development or within a tissue during tissue homeostasis and
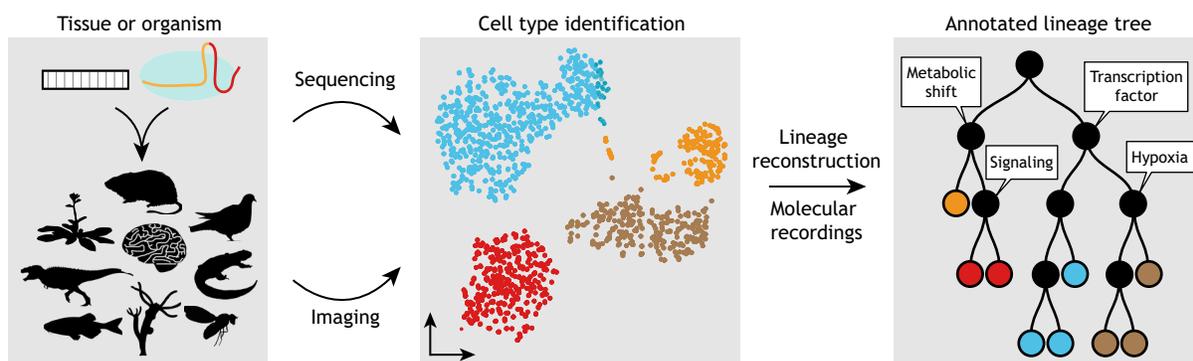

Fig. 3. The future of dynamic lineage tracing. Single cell transcriptomes and barcode recordings can be recovered from a variety of animals, tissues and organoids using sequencing and/or imaging approaches. These data can be organized into cell types using dimensionality reduction approaches. The cell lineage tree serves as an organizing scaffold to relate single cells and cell types with embedded recordings of spatial context, signaling landmarks and other measures of cellular state. The output will be richly annotated trees of development and disease.

response to injury. Initial studies are encouraging, but must be caveated by the fact that they are based on very small numbers of individuals. Remaining challenges include finding ways to scale these approaches and to develop methods to compare across individuals or species.

Conclusions and perspectives

The recent explosion of single cell data sets, including whole-organism atlases, provides an unprecedented view of development. The challenge is to link these static views of cell states together into a dynamic view of how cell types arise, and how they are regulated, maintained and ultimately dysregulated in disease (Fig. 3). It is becoming clear that single cell transcriptomes alone cannot solve this problem. Lineage tracing has shown promise as a scaffold to link cell state to cell fate, and this union with single cell transcriptomics provides the potential to explore the underlying molecular mechanisms that drive cell fate decisions (Kester and van Oudenaarden, 2018; Woodworth et al., 2017). There will be opportunities to enrich this scaffold with fate choices documented by a new generation of DNA recording constructs that respond to transient endogenous events, allow time-resolved recordings, use distributed barcodes for deeper recording and provide independence from potentially dangerous DNA breaks (Biddy et al., 2018; Farzadfard et al., 2018 preprint; Halperin et al., 2018; Perli et al., 2016; Schmidt et al., 2018; Sheth et al., 2017; Shipman et al., 2016; Tang and Liu, 2018). It is also an enticing challenge to intersect prospective lineage tracing with the powerful tools of microscopy. An ideal system would recover molecular cell state and historical recordings, all in the native spatial context of an intact organism. Although it is early days, there have already been promising steps towards this goal (Chen et al., 2015; Frieda et al., 2017; Lee et al., 2014; Shah et al., 2017; Wang et al., 2018). Interesting avenues for future work are the aggregation of data from multiple recording constructs within an organism, and the aggregation of stochastic lineage choices across many individuals. This future field of statistical development will incorporate developmental variation into models of fate choices, plasticity and function in the homeostasis of adult tissues.

With all of these evolving tools, the current challenge for biologists is to choose the best technology for their question. Dynamic lineage tracing tools are still emerging, and each approach comes with strengths and weaknesses. For computational biologists, there will also be exciting opportunities to develop new tools, especially as lineage information is coupled with single-cell measurements such as RNA sequencing, ATAC sequencing and others (Kelsey et al., 2017; Stuart and Satija, 2019). The merger of these efforts will no doubt shed light on previously intractable questions in biology, and allow the characterization and engineering of fate choices in animal models and human organoids. Ultimately, a multifaceted view of development will be provided by dynamic lineage tracing, the broad capture of single cell states and high-dimensional phenotyping. These tools will provide enormous datasets for developmental biologists looking to understand one of the most complex construction projects on earth – the development of the adult human.

References

Abdul-Wajid, S., Demarest, B. L. and Yost, H. J. (2018). Loss of embryonic neural crest derived cardiomyocytes causes adult onset hypertrophic cardiomyopathy in zebrafish. *Nat. Commun.* **9**, 4603. doi:10.1038/s41467-018-07054-8

Achim, K., Pettit, J.-B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D. and Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503-509. doi:10.1038/nbt.3209

Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. and van Oudenaarden, A. (2018). Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108-112. doi:10.1038/nature25969

Allen, F., Crepaldi, L., Alsinet, C., Strong, A. J., Kleshchevnikov, V., De Angeli, P., Páleníková, P., Khodak, A., Kiselev, V., Kosicki, M. et al. (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64-72. doi:10.1038/nbt.4317

Attardi, A., Fulton, T., Florescu, M., Shah, G., Muresan, L., Lenz, M. O., Lancaster, C., Huisken, J., van Oudenaarden, A. and Steventon, B. (2018). Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. *Development* **145**, dev166728. doi:10.1242/dev.166728

Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., Martincorena, I., Petljak, M., Alexandrov, L. B., Gundem, G. et al. (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422-425. doi:10.1038/nature13448

Bendall, S. C., Davis, K. L., Amir, E.-A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., Shenfeld, D. K., Nolan, G. P. and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714-725. doi:10.1016/j.cell.2014.04.005

Biddy, B. A., Kong, W., Kamimoto, K., Guo, C., Waye, S. E., Sun, T. and Morris, S. A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219-224. doi:10.1038/s41586-018-0744-4

Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific ϕC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312-3317. doi:10.1073/pnas.0611511104

Briggs, J. A., Weinreb, C., Wagner, D. E., Megason, S., Peshkin, L., Kirschner, M. W. and Klein, A. M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780. doi:10.1126/science.aar5780

Brody, Y., Kimmerling, R. J., Maruvka, Y. E., Benjamin, D., Elacqua, J. J., Haradhvala, N. J., Kim, J., Mouw, K. W., Frangaj, K., Koren, A. et al. (2018). Quantification of somatic mutation flow across individual cell division events by lineage sequencing. *Genome Res.* **28**, 1901-1918. doi:10.1101/gr.238543.118

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

Carlson, C. A., Kas, A., Kirkwood, R., Hays, L. E., Preston, B. D., Salipante, S. J. and Horwitz, M. S. (2012). Decoding cell lineage from acquired mutations using arbitrary deep sequencing. *Nat. Methods* **9**, 78-80. doi:10.1038/nmeth.1781

Carrelha, J., Meng, Y., Kettyle, L. M., Luis, T. C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A. et al. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106-111. doi:10.1038/nature25455

Chan, M. M., Smith, Z. D., Grosswendt, S., Kretzmer, H., Norman, T. M., Adamson, B., Jost, M., Quinn, J. J., Yang, D., Jones, M. G. et al. (2019). Molecular recording of mammalian embryogenesis. *Nature* 384925. doi:10.1038/s41586-019-1184-5

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090. doi:10.1126/science.aaa6090

Chen, W., McKenna, A., Schreiber, J., Yin, Y., Agarwal, V., Noble, W. S. and Shendure, J. (2018). Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *bioRxiv* 481069. doi:10.1101/481069

Clevers, H. and Watt, F. M. (2018). Defining adult stem cells by function, not by phenotype. *Annu. Rev. Biochem.* **87**, 1015-1027. doi:10.1146/annurev-biochem-062917-012341

Codeluppi, S., Borm, L. E., Zeisel, A., La Manno, G., van Lunteren, J. A., Svensson, C. I. and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932-935. doi:10.1038/s41592-018-0175-z

Das, A. T., Tenenbaum, L. and Berkhout, B. (2016). Tet-on systems for doxycycline-inducible gene expression. *Curr. Gene Ther.* **16**, 156-167. doi:10.2174/1566523216666160524144041

Demeulemeester, J., De Rijck, J., Gijsbers, R. and Debyser, Z. (2015). Retroviral integration: site matters: mechanisms and consequences of retroviral integration site selection. *BioEssays* **37**, 1202-1214. doi:10.1002/bies.201500051

Doench, J. G., Hartenian, E., Graham, D. B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B. L., Xavier, R. J. and Root, D. E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262-1267. doi:10.1038/nbt.3026

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R. et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184-191. doi:10.1038/nbt.3437

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

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A. and Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar313. doi:10.1126/science.aar3131

Farzadfard, F., Gharaei, N., Higashikuni, Y., Jung, G., Cao, J. and Lu, T. K. (2018). Single-nucleotide-resolution computing and memory in living cells. *bioRxiv* 263657. doi:10.1101/263657

Felsenstein, J. (1978). The number of evolutionary trees. *Syst. Zool.* **27**, 27-33. doi:10.2307/2412810

Feng, J., DeWitt, W. S., McKenna, A., Simon, N., Willis, A. and Matsen, F. A. (2019). Estimation of cell lineage trees by maximum-likelihood phylogenetics. *bioRxiv* 595215. doi:10.1101/595215

Flowers, G. P., Sanor, L. D. and Crews, C. M. (2017). Lineage tracing of genome-edited alleles reveals high fidelity axolotl limb regeneration. *Elife* **6**, e25726. doi:10.7554/eLife.25726

Frieda, K. L., Linton, J. M., Hormoz, S., Choi, J., Chow, K.-H. K., Singer, Z. S., Budde, M. W., Elowitz, M. B. and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107-111. doi:10.1038/nature20777

Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. and Shapiro, E. (2005). Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput. Biol.* **1**, e50. doi:10.1371/journal.pcbi.0010050

Gagnon, J. A., Valen, E., Thyme, S. B., Huang, P., Ahkmetova, L., Pauli, A., Montague, T. G., Zimmerman, S., Richter, C. and Schier, A. F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS ONE* **9**, e98186. doi:10.1371/journal.pone.0098186

Gagnon, J. A., Obbad, K. and Schier, A. F. (2018). The primary role of zebrafish nanog is in extra-embryonic tissue. *Development* **145**, dev147793. doi:10.1242/dev.147793

Gerber, T., Murawala, P., Knapp, D., Masselink, W., Schuez, M., Hermann, S., Gac-Santel, M., Nowoshilow, S., Kageyama, J., Khattak, S. et al. (2018). Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. *Science* **362**, eaaq0681. doi:10.1126/science.aaq0681

Haas, B. J. and Whited, J. L. (2017). Advances in decoding axolotl limb regeneration. *Trends Genet.* **33**, 553-565. doi:10.1016/j.tig.2017.05.006

Hadjieconomou, D., Rotkopf, S., Alexandre, C., Bell, D. M., Dickson, B. J. and Salecker, I. (2011). Flybow: genetic multicolor cell labeling for neural circuit analysis in Drosophila melanogaster. *Nat. Methods* **8**, 260-266. doi:10.1038/nmeth.1567

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J. et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148. doi:10.1186/s13059-016-1012-2

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V. and Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248-252. doi:10.1038/s41586-018-0384-8

Horstick, E. J., Jordan, D. C., Bergeron, S. A., Tabor, K. M., Serpe, M., Feldman, B. and Burgess, H. A. (2015). Increased functional protein expression using nucleotide sequence features enriched in highly expressed genes in zebrafish. *Nucleic Acids Res.* **43**, e48. doi:10.1093/nar/gkv035

Junker, J. P., Noël, E. S., Guryev, V., Peterson, K. A., Shah, G., Huisken, J., McMahon, A. P., Berezikov, E., Bakkers, J. and van Oudenaarden, A. (2014). Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159**, 662-675. doi:10.1016/j.cell.2014.09.038

Junker, J. P., Spanjaard, B., Peterson-Maduro, J., Alemany, A., Hu, B., Florescu, M. and van Oudenaarden, A. (2017). Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *bioRxiv* 056499. doi:10.1101/056499

Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P. and Church, G. M. (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804. doi:10.1126/science.aat9804

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N. and Zinzen, R. P. (2017). The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**, 194-199. doi:10.1126/science.aan3235

Kebschull, J. M. and Zador, A. M. (2018). Cellular barcoding: lineage tracing, screening and beyond. *Nat. Methods* **15**, 871-879. doi:10.1038/s41592-018-0185-x

Kelsey, G., Stegle, O. and Reik, W. (2017). Single-cell epigenomics: recording the past and predicting the future. *Science* **358**, 69-75. doi:10.1126/science.aan6826

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

Keyte, A. and Hutson, M. R. (2012). The neural crest in cardiac congenital anomalies. *Differentiation* **84**, 25-40. doi:10.1016/j.diff.2012.04.005

Kosicki, M., Tomberg, K. and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765-771. doi:10.1038/nbt0918-899c

Kretzschmar, K. and Watt, F. M. (2012). Lineage tracing. *Cell* **148**, 33-45. doi:10.1016/j.cell.2012.01.002

Lee, J. H., Daugharthy, E. R., Scheiman, J., Kalhor, R., Yang, J. L., Ferrante, T. C., Terry, R., Jeanty, S. S. F., Li, C., Amamoto, R. et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363. doi:10.1126/science.1250212

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E. et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473-478. doi:10.1038/s41586-018-0497-0

Leigh, N. D., Dunlap, G. S., Johnson, K., Mariano, R., Oshiro, R., Wong, A. Y., Bryant, D. M., Miller, B. M., Ratner, A., Chen, A. et al. (2018). Transcriptomic landscape of the blastema niche in regenerating adult axolotl limbs at single-cell resolution. *Nat. Commun.* **9**, 5153. doi:10.1038/s41467-018-07604-0

Lewandoski, M. and Martin, G. R. (1997). Cre-mediated chromosome loss in mice. *Nat. Genet.* **17**, 223-225. doi:10.1038/ng1097-223

Linder, D. and Gartler, S. M. (1965). Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science* **150**, 67-69. doi:10.1126/science.150.3692.67

Liu, Z. and Keller, P. J. (2016). Emerging imaging and genomic tools for developmental systems biology. *Dev. Cell* **36**, 597-610. doi:10.1016/j.devcel.2016.02.016

Liu, T.-L., Upadhyayula, S., Milkie, D. E., Singh, V., Wang, K., Swinburne, I. A., Mosaliganti, K. R., Collins, Z. M., Hiscock, T. W., Shea, J. et al. (2018). Observing the cell in its native state: imaging subcellular dynamics in multicellular organisms. *Science* **360**, eaaq1392. doi:10.1126/science.aaq1392

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R. and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56-62. doi:10.1038/nature06293

Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., Lee, S., Chittenden, T. W., D'Gama, A. M., Cai, X. et al. (2015). Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98. doi:10.1126/science.aab1785

Lu, R., Neff, N. F., Quake, S. R. and Weissman, I. L. (2011). Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928-933. doi:10.1038/nbt.1977

Ludwig, L. S., Lareau, C. A., Ulirsch, J. C., Christian, E., Muus, C., Li, L. H., Pelka, K., Ge, W., Oren, Y., Brack, A. et al. (2019). Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325-1339.e22. doi:10.1016/j.cell.2019.01.022

Mayr, U., Serra, D. and Liberali, P. (2019). Exploring single cells in space and time during tissue development, homeostasis and regeneration. *Development* **146**, dev176727. doi:10.1242/dev.176727

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., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F. and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907. doi:10.1126/science.aaf7907

Merrell, A. J. and Stanger, B. Z. (2016). Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. *Nat. Rev. Mol. Cell Biol.* **17**, 413-425. doi:10.1038/nrm.2016.24

Mokalled, M. H. and Poss, K. D. (2018). A regeneration toolkit. *Dev. Cell* **47**, 267-280. doi:10.1016/j.devcel.2018.10.015

Naik, S. H., Perié, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R. J. and Schumacher, T. N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229-232. doi:10.1038/nature12013

Oshimura, M., Uno, N., Kazuki, Y., Katoh, M. and Inoue, T. (2015). A pathway from chromosome transfer to engineering resulting in human and mouse artificial

chromosomes for a variety of applications to bio-medical challenges. *Chromosome Res.* **23**, 111-133. doi:10.1007/s10577-014-9459-z

Osorio, F. G., Rosendahl Huber, A., Oka, R., Verheul, M., Patel, S. H., Hasaart, K., de la Fonteijne, L., Varela, I., Camargo, F. D. and van Boxtel, R. (2018). Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. *Cell Rep.* **25**, 2308-2316.e4. doi:10.1016/j.celrep.2018.11.014

Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J. et al. (2019). A lineage-resolved molecular atlas of C. elegans embryogenesis at single cell resolution. *bioRxiv* 565549. doi:10.1101/565549

Pan, Y. A., Freundlich, T., Weissman, T. A., Schoppik, D., Wang, X. C., Zimmerman, S., Ciruna, B., Sanes, J. R., Lichtman, J. W. and Schier, A. F. (2013). Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development* **140**, 2835-2846. doi:10.1242/dev.094631

Pei, W., Feyerabend, T. B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C. et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456-460. doi:10.1038/nature23653

Peikon, I. D., Gizatullina, D. I. and Zador, A. M. (2014). In vivo generation of DNA sequence diversity for cellular barcoding. *Nucleic Acids Res.* **42**, e127. doi:10.1093/nar/gku604

Perli, S. D., Cui, C. H. and Lu, T. K. (2016). Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511. doi:10.1126/science.aag0511

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

Rabani, M., Pieper, L., Chew, G.-L. and Schier, A. F. (2017). A massively parallel reporter assay of 3′ UTR sequences identifies in vivo rules for mRNA degradation. *Mol. Cell* **68**, 1083-1094.e5. doi:10.1016/j.molcel.2017.11.014

Raj, B., Wagner, D. E., McKenna, A., Pandey, S., Klein, A. M., Shendure, J., Gagnon, J. A. and Schier, A. F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442-450. doi:10.1038/nbt.4103

Rajagopal, J. and Stanger, B. Z. (2016). Plasticity in the adult: how should the waddington diagram be applied to regenerating tissues? *Dev. Cell* **36**, 133-137. doi:10.1016/j.devcel.2015.12.021

Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M. and Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212-216. doi:10.1038/nature25168

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). Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **263**, 1463-1467. doi:10.1126/science.aaw1219

Sadelain, M., Papapetrou, E. P. and Bushman, F. D. (2011). Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51-58. doi:10.1038/nrc3179

Salipante, S. J. and Horwitz, M. S. (2006). Phylogenetic fate mapping. *Proc. Natl. Acad. Sci. USA* **103**, 5448-5453. doi:10.1073/pnas.0601265103

Salas, L. A., Wiencke, J. K., Koestler, D. C., Zhang, Z., Christensen, B. C. and Kelsey, K. T. (2018). Tracing human stem cell lineage during development using DNA methylation. *Genome Res.* **28**, 1285-1295. doi:10.1101/gr.233213.117

Salvador-Martínez, I., Grillo, M., Averof, M. and Telford, M. J. (2019). Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? *eLife* **8**, e40292. doi:10.7554/eLife.40292

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495-502. doi:10.1038/nbt.3192

Schmidt, S. T., Zimmerman, S. M., Wang, J., Kim, S. K. and Quake, S. R. (2017). Quantitative analysis of synthetic cell lineage tracing using nuclease barcoding. *ACS Synth. Biol.* **6**, 936-942. doi:10.1021/acssynbio.6b00309

Schmidt, F., Cherepkova, M. Y. and Platt, R. J. (2018). Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380-385. doi:10.1038/s41586-018-0569-1

Schwartz, R. and Schäffer, A. A. (2017). The evolution of tumour phylogenetics: principles and practice. *Nat. Rev. Genet.* **18**, 213-229. doi:10.1038/nrg.2016.170

Shah, S., Lubeck, E., Zhou, W. and Cai, L. (2017). seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. *Neuron* **94**, 752-758.e1. doi:10.1016/j.neuron.2017.05.008

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H. L., Koulena, N., Cronin, C., Karp, C., Liaw, E. J. et al. (2018). Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363-376.e16. doi:10.1016/j.cell.2018.05.035

Shapiro, E., Biezuner, T. and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618-630. doi:10.1038/nrg3542

Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., Cassa, C. A., Liu, D. R., Gifford, D. K. and Sherwood, R. I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646-651. doi:10.1038/s41586-018-0686-x

Sheth, R. U., Yim, S. S., Wu, F. L. and Wang, H. H. (2017). Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457-1461. doi:10.1126/science.aao0958

Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., Enikolopov, G., Nauen, D. W., Christian, K. M., Ming, G.-L. et al. (2015). Single-Cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* **17**, 360-372. doi:10.1016/j.stem.2015.07.013

Shipman, S. L., Nivala, J., Macklis, J. D. and Church, G. M. (2016). Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175. doi:10.1126/science.aaf1175

Spanjaard, B., Hu, B., Mitic, N., Olivares-Chauvet, P., Janjuha, S., Ninov, N. and Junker, J. P. (2018). Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469-473. doi:10.1038/nbt.4124

Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M. et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 57-63. doi:10.1126/science.aaf2403

Stewart, S. and Stankunas, K. (2012). Limited dedifferentiation provides replacement tissue during zebrafish fin regeneration. *Dev. Biol.* **365**, 339-349. doi:10.1016/j.ydbio.2012.02.031

St John, K. (2017). Review paper: the shape of phylogenetic treespace. *Syst. Biol.* **66**, e83-e94. doi:10.1093/sysbio/syw025

Stuart, T. and Satija, R. (2019). Integrative single-cell analysis. *Nat. Rev. Genet.* **20**, 257-272. doi:10.1038/s41576-019-0093-7

Sulston, J. E., Schierenberg, E., White, J. G. and Thomson, J. N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64-119. doi:10.1016/0012-1606(83)90201-4

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O. and Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. *Nature* **514**, 322-327. doi:10.1038/nature13824

Tanaka, E. M. (2016). The molecular and cellular choreography of appendage regeneration. *Cell* **165**, 1598-1608. doi:10.1016/j.cell.2016.05.038

Tang, W. and Liu, D. R. (2018). Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992. doi:10.1126/science.aap8992

Tornini, V. A., Thompson, J. D., Allen, R. L. and Poss, K. D. (2017). Live fate-mapping of joint-associated fibroblasts visualizes expansion of cell contributions during zebrafish fin regeneration. *Development* **144**, 2889-2895. doi:10.1242/dev.155655

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381-386. doi:10.1038/nbt.2859

Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H. and Theis, F. J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, dev170506.

Tu, S. and Johnson, S. L. (2011). Fate restriction in the growing and regenerating zebrafish fin. *Dev. Cell* **20**, 725-732. doi:10.1016/j.devcel.2011.04.013

van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., Reece-Hoyes, J. S., Nye, C., Gradia, S., Vidal, B. et al. (2016). DNA Repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633-646. doi:10.1016/j.molcel.2016.06.037

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G. and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987. doi:10.1126/science.aar4362

Walther, V. and Alison, M. R. (2016). Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers. *Wiley Interdiscip. Rev. Dev. Biol.* **5**, 103-117. doi:10.1002/wdev.203

Wang, X., Allen, W. E., Wright, M. A., Sylwestrak, E. L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J. et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691. doi:10.1126/science.aat5691

Weinreb, C., Rodriguez-Fraticelli, A. E., Camargo, F. D. and Klein, A. M. (2018). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *bioRxiv* 467886. doi:10.1101/467886

Woodworth, M. B., Girskis, K. M. and Walsh, C. A. (2017). Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230-244. doi:10.1038/nrg.2016.159

Xu, J., Nuno, K., Litzenburger, U. M., Qi, Y., Ryan Corces, M., 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

Yamamoto, R., Wilkinson, A. C. and Nakauchi, H. (2018). Changing concepts in hematopoietic stem cells. *Science* **362**, 895-896. doi:10.1126/science.aat7873

Yang, Z. and Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303-314. doi:10.1038/nrg3186

DEVELOPMENT