TECHNOLOGIES AND TECHNIQUES

Unravelling cellular relationships during development and regeneration using genetic lineage tracing

Chloé S. Baron[1,2,3,4] *and Alexander van Oudenaarden*[1,2,3,4]*

Abstract | Tracking the progeny of single cells is necessary for building lineage trees that recapitulate processes such as embryonic development and stem cell differentiation. In classical lineage tracing experiments, cells are fluorescently labelled to allow identification by microscopy of a limited number of cell clones. To track a larger number of clones in complex tissues, fluorescent proteins are now replaced by heritable DNA barcodes that are read using next-generation sequencing. In prospective lineage tracing, unique DNA barcodes are introduced into single cells through genetic manipulation (using, for example, Cre-mediated recombination or CRISPR–Cas9-mediated editing) and tracked over time. Alternatively, in retrospective lineage tracing, naturally occurring somatic mutations can be used as endogenous DNA barcodes. Finally, single-cell mRNA-sequencing datasets that capture different cell states within a developmental or differentiation trajectory can be used to recapitulate lineages. In this Review, we discuss methods for prospective or retrospective lineage tracing and demonstrate how trajectory reconstruction algorithms can be applied to single-cell mRNA-sequencing datasets to infer developmental or differentiation tracks. We discuss how these approaches are used to understand cell fate during embryogenesis, cell differentiation and tissue regeneration.

Understanding cell-division dynamics, cell-fate decisions and spatial organization of cells is crucial for understanding development, homeostasis and disease. In particular, characterizing the descendants of individual cells (that is, cell lineages such as those formed during haematopoietic or neuronal differentiation) remains a long-standing aim of biomedical research. In a pioneer lineage tracing experiment, the fate of every cell of a *Caenorhabditis elegans* embryo was determined by time-lapse microscopy[1]. Such lineage tree reconstruction, based on direct observation, remains limited to small and transparent organisms with invariant lineages between individuals. To overcome these limitations and study lineage decisions in more complex organisms and tissues, scientists have developed methods to label and track cells of interest using dye injections, transplantations, viral transduction or genetic recombination of fluorescent proteins[2]. Although powerful, these approaches suffer from the small number of generations that remain labelled after dye injection, the non-physiological setting of cell transplantation, the low frequency of viral-barcode insertion or the limited number of fluorescent proteins available to label complex tissues[2].

The next-generation sequencing revolution has allowed the development of genetic lineage tracing methods, where dyes and fluorescent proteins are replaced by nucleotide sequences that serve as lineage barcodes[3]. Genetic lineage tracing methods rely on the introduction of unique and heritable DNA barcodes in single cells (FIG. 1). Barcodes are identified by sequencing, and cells sharing the same DNA barcode are identified as part of the same lineage, originating from the same founder cell. Examples of DNA barcodes include insertions and/or deletions generated by the CRISPR–Cas9 system or naturally occurring single-nucleotide variants (SNVs)[4–6]. Overall, the power of this approach lies in the high diversity of usable unique barcodes and in the high-throughput sequencing readout of these barcodes.

In this Review, we discuss recent advances in genetic lineage tracing, focusing on the biological findings and insights provided by these novel technologies. First, we discuss prospective genetic lineage tracing approaches such as methods using the hyperactive *Sleeping Beauty* transposon, Polylox and CRISPR–Cas9 systems (FIGS 2,3; TABLE 1). We explore how these approaches can be used to deepen our knowledge of the clonal dynamics of mouse haematopoiesis and of zebrafish embryonic development and organogenesis (FIGS 2–4). Next, we examine the retrospective genetic lineage tracing approaches, which rely on analysis of copy number variations (CNVs),

[1]*Oncode Institute, Utrecht, Netherlands.*

[2]*Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences, Utrecht, Netherlands.*

[3]*University Medical Center Utrecht, Utrecht, Netherlands.*

[4]*University Medical Center Utrecht, Cancer Genomics Netherlands, Utrecht, Netherlands.*

**e-mail: a.vanoudenaarden@hubrecht.eu*

https://doi.org/10.1038/s41580-019-0186-3

Long-term haematopoietic stem cells
(LT-HSCs). Blood stem cells able to self-renew and differentiate into all types of mature blood cells.

SNVs, retroelements and microsatellites (FIG. 5). We illustrate how these methods can be used to study human tissues in health or disease. Finally, we discuss how single-cell (m)RNA-sequencing (scRNA-seq) datasets can be used to reconstitute lineage trajectories using the Monocle and RNA velocity algorithms.

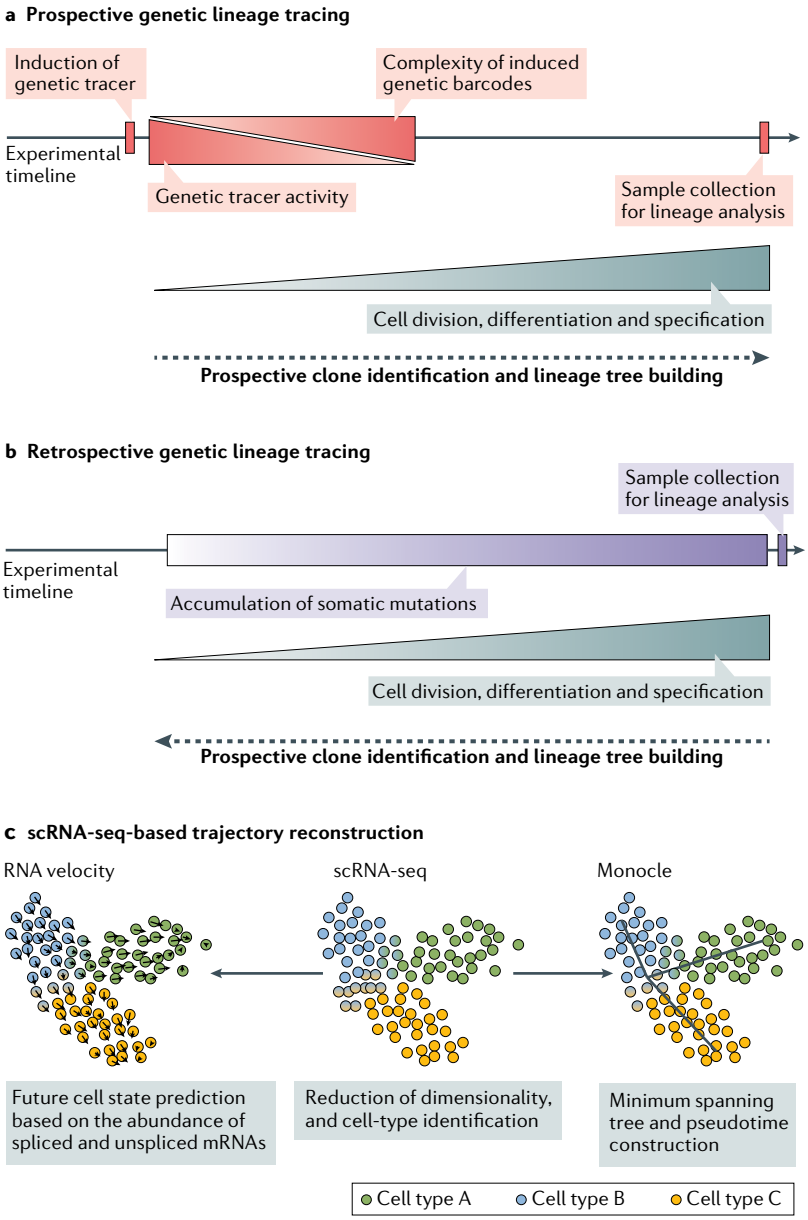


Fig. 1 | **Genetic lineage tracing approaches. a** | Set-up of prospective genetic lineage tracing. A genetic tracer is introduced, is activated and remains active for a given period. During tracer activity, genetic barcodes are dynamically generated. Tissue or cells of interest are collected at the end of the experiment and the genetic barcodes can be sequenced for clonal analysis. **b** | Set-up of retrospective genetic lineage tracing. Mutations naturally occur in cells of interest over time. The mutations are identified at the end of the experiment to reconstruct lineage relationships in the tissue of interest. **c** | Overview of single-cell (m)RNA sequencing (scRNA-seq)-based trajectory reconstruction using the algorithms Monocle and RNA velocity. Monocle is built upon the assumption that single cells with a similar transcriptional profile are likely to be close within a developmental or differentiation trajectory. Monocle projects single cells along a minimum spanning tree and creates a pseudotime (black line) representing the inferred lineage trajectory. RNA velocity uses the abundance of spliced and unspliced mRNAs in single cells to predict their future state (arrow).

Altogether, we demonstrate how advances in molecular and computational biology allow genetic lineage tracing to refine our fundamental knowledge of developmental and stem cell biology.

Prospective genetic lineage tracing

Prospective genetic lineage tracing requires the introduction of a genetic tracer that will uniquely and permanently label cells on its activation (FIG. 1a). The *Sleeping Beauty* transposase, Cre–*loxP* and CRISP–Cas9 systems fulfil the criteria required for prospective genetic lineage tracing methods. Once activated, these systems will remain active for a given period and allow the accumulation of genetic barcodes. At the end of the experiment, the progeny cells are collected and their genetic barcode is sequenced for clone identification, followed by lineage tree reconstruction.

Tracing using transposon integration

In 2014, a prospective genetic lineage tracing method was developed in which a known DNA transposon is mobilized to random positions in the genome after activation of a transposase[7]. The toolkit for this method involves a doxycycline-dependent transcription activator under the control of the mouse *Rosa26* locus, and a hyperactive *Sleeping Beauty* (HSB) transposase and a known HSB-responsive transposon element, both encoded in the *Col1a1* locus (FIG. 2a; TABLE 1). A mouse line carrying all three alleles was generated, and on doxycycline administration, HSB is expressed, leading to transposon mobilization to a random genomic location. Transposon mobilization induces excision of a translation stop codon and the expression of a red fluorescent protein monitoring the mobilization. With use of a nested PCR strategy relying on a known ligated linker and the known transposon sequence, the genomic location of the mobilized transposon is sequenced and serves as a lineage barcode.

Biological application. Haematopoietic stem cells (HSCs) have been studied for decades using in vivo transplantation — an assay recognized as the gold standard for identifying HSCs and their lineage output[8,9]. With use of this approach, long-term haematopoietic stem cells (LT-HSCs) have been identified and placed at the top of the haematopoietic differentiation hierarchy, described as the exclusive multipotent and self-renewing cell population of the human blood system. Within the haematopoietic hierarchy, strict branching occurs between the myeloid and lymphoid lineages, and further differentiation occurs towards mature cell types through defined progenitor states[10,11] (FIG. 2c). With use of the *Sleeping Beauty* transposase system, the clonal dynamics of native and post-transplantation haematopoiesis was studied. First, analysis of the clonal output of native granulopoiesis revealed that the barcodes present in granulocytes are transient, indicating that in the steady state (unperturbed state of a tissue) granulopoiesis is polyclonal (~900 clones are made of a few single cells). A very minor barcode overlap was observed between granulocytes and lymphocytes, revealing that granulocyte-producing clones are restricted to the myeloid

lineage (FIG. 2d). However, a considerable barcode overlap was found between granulocytes and monocytes, indicating that clones that produce myeloid cells are at least bipotent. In contrast to the native state, different granulocyte clonal dynamics were observed after transplantation, as larger granulocyte clones were found to be stable over time. Finally, in a pioneer experiment, the clonal output of LT-HSCs at steady state and after transplantation were compared. To do so, the steady-state granulocyte barcodes in one animal were compared with the granulocyte barcodes present after transplantation of bone marrow cells of that same animal into an irradiated recipient. Considering that repopulation on transplantation is driven by LT-HSCs, the same barcodes should be found in both settings. Surprisingly, granulocyte clones were found to be different in the two experimental settings. This revealed that steady-state granulopoiesis is driven not by LT-HSCs but instead by progenitors that are not able to reconstitute the haematopoietic system of an irradiated recipient on transplantation. This important result showed for the first time that LT-HSCs have limited lineage output in the steady state compared with a transplantation setting.

In a follow-up study from the same laboratory, the *Sleeping Beauty* transposase system was combined with scRNA-seq to deepen our understanding of steady-state haematopoiesis, focusing on multipotent progenitors (MPPs) and megakaryocyte progenitors (MkPs)[12]. First, no MkP barcodes were found to be shared with erythroblasts, suggesting no shared clonality and arguing against the existence of a megakaryocyte–erythrocyte progenitor (FIG. 2d). Instead, a subset of MPPs was found that was responsible for the production of MkP clones. Furthermore, analysis of LT-HSC clonality showed an overall low but strongly MkP-biased output, accounting for half of MkP production. The combination of these two findings revealed that LT-HSCs and MPPs are two different, independent sources of MkP production at steady state. Second, transcriptome profiling of LT-HSCs and four phenotypically defined MPP populations (MPP1 to MPP4) was performed. The analysis revealed that MPPs are heterogeneous and contain both lineage primed cells and non-primed cells, in different ratios. For example, ~20% of MPP3 cells are granulocyte and monocyte primed cells and ~19% of MPP2 cells are erythroid primed cells. In line with the clonal analysis described above, a subset of LT-HSCs was found to be primed towards the megakaryocyte lineage. Finally, the output of LT-HSC clones at steady state was compared with the output in a transplantation setting. Strikingly, LT-HSC clones producing MkP at steady state were found to be capable of multipotency after transplantation, revealing once more the importance of studying native haematopoiesis separately from transplantation-induced haematopoiesis. Overall, these two studies emphasize the power of genetic lineage tracing compared with transplantation-based methods to study the haematopoietic system in physiological conditions. However, some limitations to this approach remain. First, the barcoding efficiency in vivo is reported to be around 30%, leaving a majority of HSCs unlabelled and therefore not taken into account for clonal analysis. Next, small clones potentially of high importance for in vivo dynamics of slowly dividing LT-HSCs can be missed owing to technical capture limitations. Finally, merging clonal information and cell type identification relies on the expression of a limited number of cell surface markers. Owing to the potential heterogeneity of phenotypically defined haematopoietic populations, identifying the lineage decision of single cells within a clone remains imperfect[10]. The *Sleeping Beauty* transposase method will hopefully continue to benefit from technological advances, mainly the simultaneous profiling of the transcriptome and of location of the mobilized transposon in the same single cell.

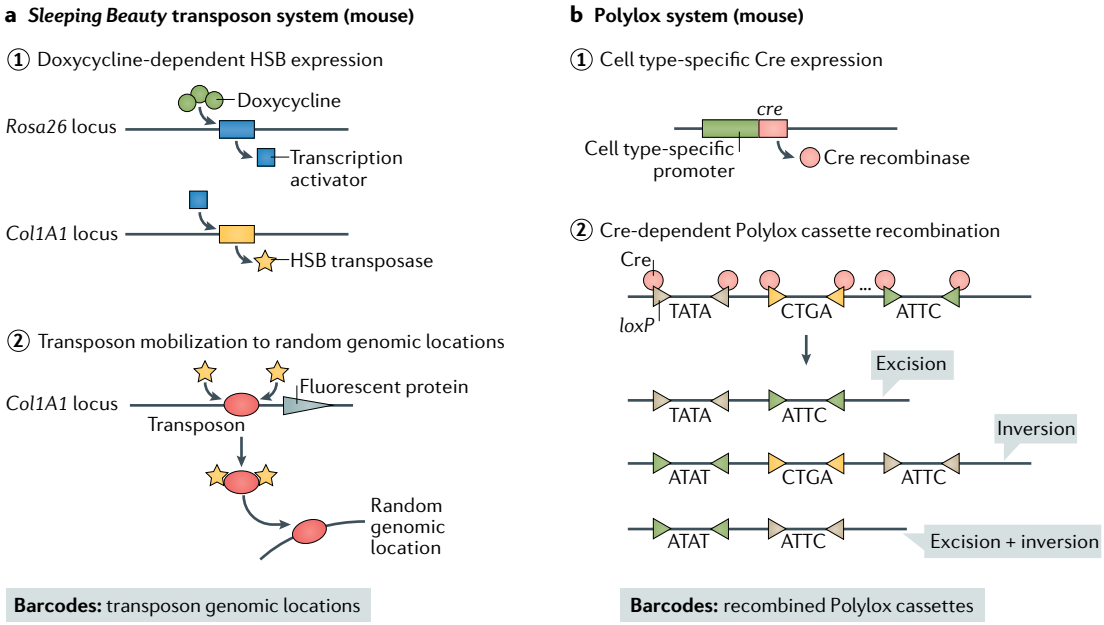
Primed cells
Differentiating cells with a determined cell fate.

Lineage tracing using Cre–loxP

The Cre–*loxP* system allows the tracking of genetic recombination by triggering the expression of a fluorescent protein in a cell type of choice[13–16]. The recombined fluorescent protein is permanently expressed and heritable, and its expression can be read out by microscopy. The Cre–*loxP* system requires two tools: an inducible Cre recombinase expressed under the control of a tissue-specific or cell-specific promotor, and a fluorescent reporter preceded by a *loxP*–stop codon–*loxP* sequence. On Cre induction at a desired time point, the two *loxP* sequences are recombined and the translation stop codon is excised, leading to the expression of the fluorescent protein in the cell population of choice. To increase the number of cell populations tracked, multiple fluorescent protein coding sequences flanked by *loxP* sites can be introduced[17–19]. In this case, on Cre induction, the *loxP* sites are randomly recombined, leading to multicolour, mosaic expression of different fluorescent proteins in the targeted cell populations. However, the limited number of clones (colours) that can be visualized by microscopy remains largely insufficient for the study of highly complex and heterogeneous tissues. To overcome this limitation, the fluorescent proteins were exchanged for DNA sequences to create the Polylox cassette, which is composed of 10 *loxP* sites interspaced with unique DNA sequences[20] (FIG. 2b; TABLE 1). On Cre induction, random excisions and inversions occur, generating a uniquely recombined Polylox cassette that is used as a barcode. If all 10 *loxP* sites are recombined, approximately one million unique barcodes are created, which can be used to study the clonal output of as many single cells. Additionally, the Polylox mouse line can be crossed with any cell type-specific, Cre-inducible mouse line to study a wide variety of cell populations and tissues during embryonic development, cell differentiation and/or tissue regeneration.

Biological application. By this approach, the clonal output of mouse HSCs was studied by inducing recombination of the Polylox cassette in HSCs expressing the endothelial-specific receptor tyrosine kinase TIE2. Polylox recombination was first induced in embryonic day 9.5 (E9.5) embryos, the time at which the first HSCs are generated from TIE2$^+$ endothelial cells in the embryonic aorta. Analysis of the output of embryonic clones in adult mice revealed that single TIE2$^+$ HSC precursors give rise to large multipotent HSC clones. A considerable

fraction of barcodes detected in erythroid progenitors and granulocytes are absent from common myeloid progenitors, suggesting that erythroid–myeloid differentiation can occur independently of common myeloid progenitors (FIG. 2d). Next, Polylox recombination was induced in young adult mice and the clonal output was analysed at a later stage of adulthood. In this setting, the clonal output of HSCs was smaller than during



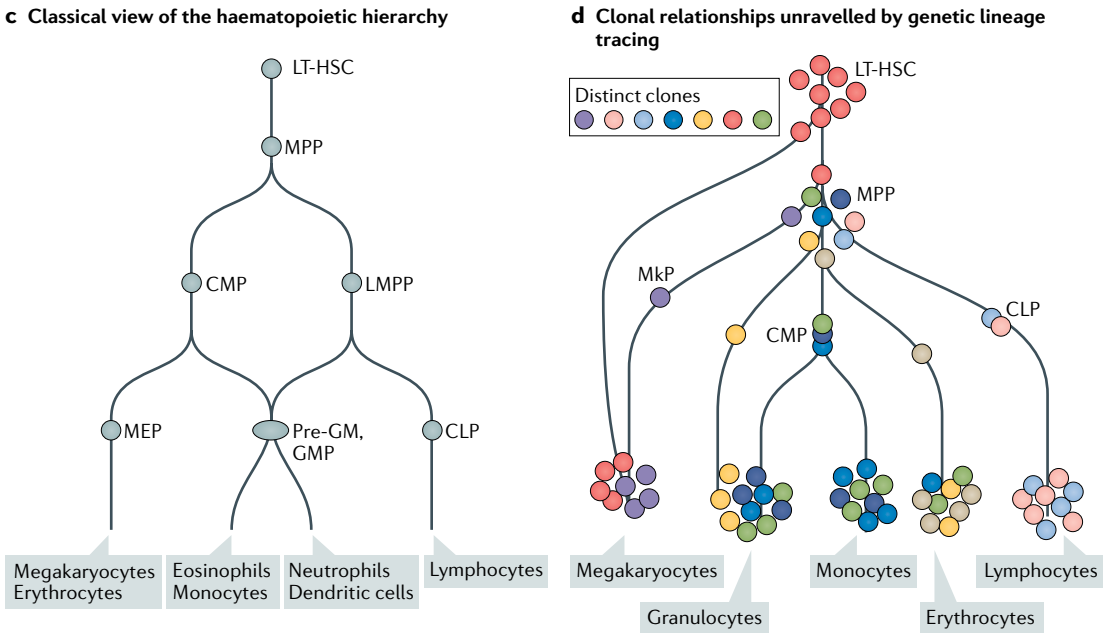


Fig. 2 | **Prospective genetic lineage tracing to study mouse haematopoiesis. a** | The *Sleeping Beauty* transposon system allows doxycycline-dependent mobilization of the hyperactive *Sleeping Beauty* (HSB) transposon to random genomic locations. Sequencing of the transposon integration loci, which serve as barcodes, allows clone identification. **b** | The Polylox system relies on the action of a Cre recombinase to rearrange a Polylox cassette of 10 *loxP* sites interspaced with unique DNA sequences. The randomly and stably recombined Polylox cassettes serve as barcodes. **c** | Classic tree-like organization of the haematopoietic system cell hierarchy. Haematopoietic stem cells are positioned at the top, giving rise to the different types of progenitor cell that will differentiate into all major blood cell types. **d** | Clonal relationships within the haematopoietic hierarchy as identified by genetic lineage tracing studies using the *Sleeping Beauty* transposon and Polylox systems. Alternative routes to the classic tree-like organization can be observed. CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte–monocyte progenitor; LMPP, lymphoid-primed multipotent progenitor; LT-HSC, long-term haematopoietic stem cell; MEP, megakaryocyte–erythrocyte progenitor; MkP, megakaryocyte progenitor; MPP, multipotent progenitor; pre-GM, pre-granulocyte–monocyte lineage cell.

embryogenesis, suggesting that fewer TIE2$^+$ HSC clones are active during adulthood. Further clonal analysis revealed clear clonal segregation between erythroid–myeloid and lymphoid cells, supporting a tree-like structure of haematopoiesis with a primary branching of the myeloid and lymphoid lineages. Although powerful, the clonal readout in this study lacks single-cell information and is restricted to phenotypically defined haematopoietic populations. This limitation does not allow one to conclude that single cells within a clone can give rise to several branches. To overcome this limitation, both the recombined Polylox cassette and the transcriptome need to be sequenced from single haematopoietic cells. Finally, to achieve an unbiased readout of the clonal output of all HSCs, TIE2 will need to be replaced by a unique, pan-HSC marker, the finding of which is a holy grail in the haematopoiesis field.

Lineage tracing using CRISPR–Cas9

Five years ago, the CRISPR–Cas9 system was described as a tool for genome editing in vitro[4,5,21,22]. Since this breakthrough discovery, scientists have greatly expanded the applicability, which now includes also genetic lineage tracing. Using a guide RNA (gRNA), the Cas9 nuclease targets a specific genomic region and generates a double-strand break, the repair of which introduces unique small insertions or deletions (indels) of variable length and position. Indels become permanent genomic barcodes (or scars) and the descendants of each barcoded cell will inherit the barcode, thereby allowing clonal tracking and lineage tree reconstruction.

CRISPR–Cas9 tracing tools in zebrafish. The GESTALT (genome editing of synthetic target arrays for lineage tracing) and ScarTrace methods were the first proof of principle for using Cas9-generated barcodes for lineage tracing in zebrafish[23,24]. By injection of Cas9 and gRNAs into zygotes, barcodes were introduced into pregastrulation embryos. In GESTALT, one gRNA targets 10 locations (a GESTALT cassette, which is integrated into the DNA of zebrafish embryos) at various efficiencies (FIG. 3a; TABLE 1). In ScarTrace, gRNA targets several copies of GFP inserted in tandem downstream of the ubiquitously expressed gene encoding histone H2A. Although both studies provided strong proof of principle, they lacked single-cell resolution and transcriptional readout. To overcome these limitations, scScarTrace, scGESTALT and LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous sequences) were developed[25–27]. In scGESTALT, a heat shock-inducible Cas9 provides a second wave of barcode generation at later time points during development. Thus, scGESTALT provides early barcode generation beginning at the one-cell stage, and late barcode generation following a heat shock at 30 h after fertilization. scScarTrace relies on injection of either Cas9 RNA or Cas9 protein at the one-cell stage. Following injection of Cas9 protein, barcode generation is observed between 0 and 3 h after fertilization, but when Cas9 mRNA is injected, barcode generation occurs until up to 10 h after fertilization. Use of CRISPR–Cas9 to edit identical sequences positioned in tandem inevitably leads to larger excisions and loss of barcode information from one or more target sites. To overcome this limitation, LINNAEUS targets red fluorescent protein transgenes that are scattered across the genome. Another difference between the methods is the detection of the scars from mRNA or DNA. Whereas scGESTALT and LINNAEUS rely on transcription of the scars to then capture edited mRNA, scScarTrace detects the indel directly from DNA, to remove the risk of differential expression of the scar in different tissues. Overall, the power of these CRISPR–Cas9-based methods lies in the dynamic barcoding process: Cas9 remains active for hours, which allows barcoding to occur sequentially, thereby generating intricate scar patterns that allow the reconstruction of multibranching lineages.

Once established, CRISPR–Cas9-based genetic lineage tracing was used to study the clonal dynamics of various biological systems. scGESTALT was applied to zebrafish brains and used to unravel lineage relationships between as many as 100 cell types. Clonal analysis revealed that most of the descendants of single progenitors are spatially restricted in the different brain regions (forebrain, midbrain or hindbrain)[28,29]. Only a small proportion of barcodes were found spread over several brain regions, suggesting that either very early barcoding events or only few progenitor cells give rise to descendants capable of migrating across brain regions, a finding confirmed using scScarTrace (FIG. 4a). Furthermore, barcodes were found to be not cell type specific, suggesting that the labelled progenitors were multipotent. By careful analysis of a small number of clones, and aiming to unravel divergent lineage trajectories, cell type-specific lineage trajectories were uncovered in the preoptic area of the hypothalamus. Precisely, Sst3$^+$ neurons and Penkb$^+$ neurons are found to be related to each other but not to Fezf1$^+$ or Hmx3a$^+$ neurons (FIG. 4a). However, these four cell populations possess a common ancestor from a population of GABAergic and glutamatergic neurons from the ventral forebrain, originating from the early edit. This reveals the existence of a common ancestor cell type, which splits after gastrulation to give rise to these specific lineages in the preoptic area. These findings were described in two zebrafish, where six and eight clones were found, respectively, each from ~100 single cells[26], thereby demonstrating the power of scGESTALT to trace rare lineage trajectories in an organ as complex as the brain, which is made of more than 100 cell types. Furthermore, lineage trees were hypothesized not only to reconstitute lineage trajectories of cells but also to reconstruct gene expression trajectories during development. The trajectory reconstruction algorithm Monocle2 was used to dissect the gene expression dynamics occurring during progenitor maturation towards an oligodendrocyte fate[30]. Finally, scScarTrace showed that brain-resident macrophages or microglia share clones with HSCs of the whole kidney marrow (WKM; a major site of haematopoiesis in adult zebrafish), thereby reflecting a common origin[31].

A clonal analysis of the WKM was performed using scScarTrace, which revealed that all blood cells derive from a small number of early embryonic progenitor cells, which is in line with findings reported from use of fluorescence-based lineage tracing methods[32] (FIG. 4b). LINNAEUS was used to study the clonal make-up

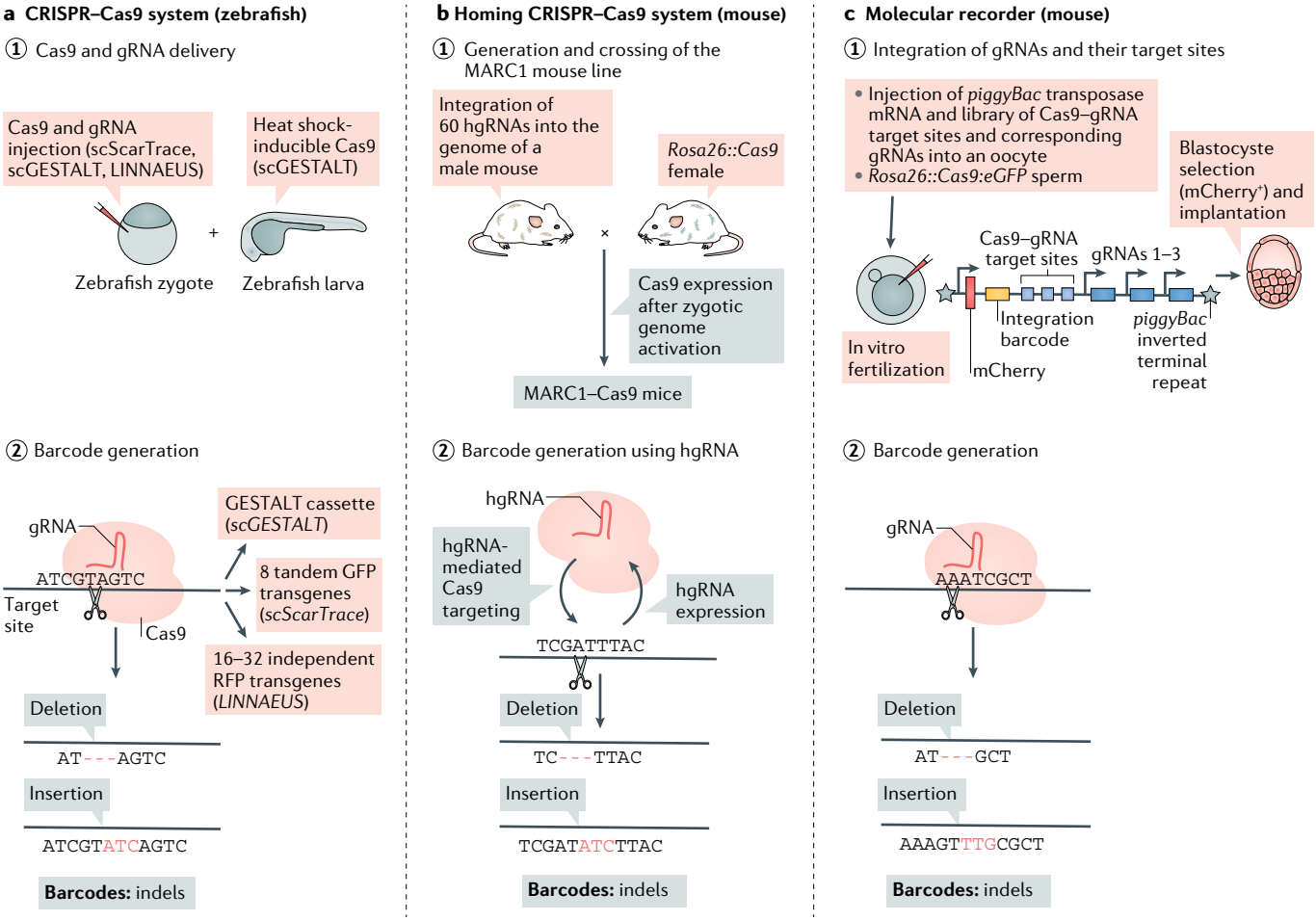


Fig. 3 | **Prospective genetic lineage tracing using the CRISPR–Cas9 system.** CRISPR–Cas9 genome editing can be used to generate insertion and deletion mutations (indels), which serve as lineage barcodes. **a** | In zebrafish, single-cell genome editing of synthetic target arrays for lineage tracing (scGESTALT), scScarTrace and lineage tracing by nuclease-activated editing of ubiquitous sequences (LINNAEUS) allow the generation of DNA barcodes during early embryogenesis. In the three methods, Cas9 and guide RNAs (gRNAs) targeting a sequence of choice are injected into the zygote to induce dynamic barcoding during early embryogenesis. Later barcoding can be induced in scGESTALT by heat shock-mediated induction of Cas9. **b** | In the MARC1 (mouse for actively recording cells 1) line, 60 sequences encoding homing gRNAs (hgRNAs) are integrated across the genome. By crossing MARC1 animals with Cas9-expressing animals, barcoding in all hgRNA-expressing sites is initiated at the onset of transcription in the zygote. **c** | With use of the *piggyBac* transposase, a library of DNA sequences containing Cas9 target sites and encoding the gRNAs targeting them are integrated into the genome of a mouse zygote created from the fusion of an oocyte into which the library and *piggyBac* mRNA have been injected with Cas9-expressing sperm. Barcoded blastocysts are then transferred into pseudopregnant mothers for further embryonic development. RFP, red fluorescent protein.

of zebrafish larvae at 5 days after fertilization and identified ~70 cell types. With focus on the lateral plate mesoderm, all blood cells were found to share the same barcodes with the exception of erythrocytes (FIG. 4b). This observation reflects the distinct origin of embryonic erythrocytes, which are produced during the primitive wave of haematopoiesis, whereas all adult blood cell types (including HSCs) are generated from endothelial cells during the definitive wave of haematopoiesis[33,34]. In accordance with this endothelial origin of adult HSCs, endothelial cells and blood cells of independent origin from erythrocytes were found to share clones.

scScarTrace was used to study right–left body axis specification by separately analysing the clonal output in right and left eyes and brain regions. Although the right and left midbrain shared clones, right and left eyes mostly shared none, suggesting an early (less than 10 h after fertilization) specification of the right–left identity of the eyes. To refine this time window, Cas9 was specifically injected into one of the cells of the two-cell-stage embryo to induce scars in only half of the embryo until the dome stage. From analysis of clones 3 weeks later, scars were found in the right and left eyes, indicating that both eyes receive cells by the time of the dome stage. By comparing lineage trees reconstructed after scaring with Cas9 protein and Cas9 mRNA, the study concluded that progenitor cells commit to the right eye or left eye shortly after the dome stage (FIG. 4c).

LINNAEUS was used to study the clonal relationship between the telencephalon, heart, liver and pancreatic

Lateral plate mesoderm
A subset of mesodermal tissue found in developing embryos that will form the body walls and circulatory system.

Primitive wave of haematopoiesis
A process by which the embryo generates transient haematopoietic cells, which are necessary for embryonic development.

Definitive wave of haematopoiesis
A process by which definitive adult haematopoietic stem cells are generated during embryogenesis.

Dome stage
A developmental stage of zebrafish embryos that is reached at 4.3 h of development.

islets of a single fish. All organs showed early clone separation, thereby revealing the existence of early differentiation into segregated lineages; in the pancreas, α-cells, β-cells and δ-cells mainly shared their origin, with the exception of an interesting clone, in which δ-cells were depleted, suggesting the existence of an α-cell and β-cell restricted progenitor[27] (FIG. 4d).

Finally, the clonal dynamics of zebrafish caudal fin regeneration was studied using scScarTrace. Consistent with the findings of previous work, most of the cell types in the caudal fin were found to be clonally restricted, in physiological conditions and following regeneration[35]. However, an increase of clones shared between osteoblasts and mesenchyme cells following one or two cycles of regeneration revealed a degree of plasticity between the two lineages (FIG. 4e). Most of the immune cells in the fin were found to share clonality with WKM HSCs. Strikingly, a subset of resident immune cells share clonality with epidermal and mesenchymal cells. Since this particular immune population was found in both healthy and regenerated fin, it suggests the existence of a novel developmental trajectory for tissue-resident immune cells. This observation was confirmed and refined by the identification of a new ectoderm-derived myeloid-like cell population in the zebrafish epidermis, termed 'metaphocytes'[36]. In summary, three distinct experimental strategies relying on CRISPR–Cas9 to introduce DNA barcodes in single cells during zebrafish embryogenesis were used to study the clonal history of a wide variety of larval and adult organs.

Implementation of CRISPR–Cas-based tracing in mice. Following the successful development of CRISPR–Cas9 genetic lineage tracing methods in zebrafish, the technology was implemented in mice. As described earlier, zebrafish CRISPR–Cas9 barcoding relies mainly on injecting the gRNA and Cas9 nuclease into the one-cell embryo. Use of such an approach in the mouse embryo is not attractive because mouse embryonic development is slower, and thus after injection, Cas9 would be active only during the first cell division. By contrast, the barcodes need to be generated over a longer period and the induction of the system needs to be compatible with the in utero development of the zygote. To this end, a barcoding strategy was created that relies on homing gRNAs (hgRNAs), which allow targeting of Cas9 to a larger number of genomic sites than canonical gRNAs[37–39]. Mice of the MARC1 (mouse for actively recording cells 1) line expressing 60 hgRNAs, of which the genomic locations of the integration of 54 are known, were generated and crossed with a constitutively expressing Cas9 transgenic line (FIG. 3b; TABLE 1). On fertilization and as soon as zygotic genome activation begins, the hgRNAs are transcribed and barcodes are made by the activity of Cas9 at the different loci. Characterization of the barcoding dynamics revealed that certain hgRNAs were mutated with variable dynamics, between shortly after Cas9 activation and as late as birth. As a proof of principle, the lineage tree reflecting the branching of trophectoderm, primitive endoderm and epiblast by E4.5 was reconstructed by analysis of barcodes in E12.5 embryos[39]. Furthermore, axis establishment during neural tube development was studied by comparison of the embryonically generated barcodes found in various right–left and anterior–posterior regions of the adult brain. The analysis revealed that, following initial specification around the time of gastrulation, the nervous system axes are established around E8.5 in an order that remains unclear. Tree reconstruction resulted in the conclusion that during embryogenesis the anterior–posterior axis of the neural tube is established earlier than the right–left axis. The power of this study lies in the hgRNA-expressing mouse line, which requires for lineage tracing only crossing with any inducible or cell-specific Cas9-expressing line. In addition, and similarly to the approach of LINNAEUS, the 60 sequences encoding hgRNAs in the MARC1 line are scattered across the genome, which dramatically reduces the occurrence of the large deletions found when Cas9 target loci are located in tandem. Although only 41 of the 60 hgRNAs are active, the theoretical number of possible barcodes is still astonishing — 10^{74}. However, the readout of the barcodes generated in MARC1 offspring still does not allow transcriptome analysis of single cells.

In another study, a molecular recorder of mammalian embryogenesis was developed, using the CRISPR–Cas9 system on the basis of the capacity of a transposase to integrate the target sequences in various genomic locations[40]. With use of the *piggyBac* transposase, several synthetic DNA molecules, each containing three different Cas9 target sites together with the sequences encoding their corresponding gRNAs, are integrated into the genome; each gRNA is transcribed under the control of a distinct promoter and the target sites are transcribed under the control of a constitutive promoter and fused to a fluorescent protein (mCherry) (FIG. 3c; TABLE 1). To perform in vivo lineage tracing, in vitro fertilization was performed using mouse oocytes into which *piggyBac* transposase mRNA had been injected, multiple Cas9 target sites to be integrated into the genome and sperm expressing Cas9–eGFP. Healthy blastocysts screened for exhibiting high mCherry fluorescence were transferred into pseudopregnant mice for implantation, and embryos were collected at ~E9 for lineage analysis. As proof of principle, this approach successfully reconstituted the lineage segregation between placenta, yolk sac, embryonic body, head and tail. Since the Cas9-edited target sites are transcribed, they can be profiled together with the transcriptome of thousands of single cells. Many tissues and cell types were profiled, and despite some heterogeneity in the number of target sites recovered per cell, lineage relationships were determined. A shared ancestor between presomitic mesoderm and neural tissues was found, suggesting they have a common origin from neuromesodermal progenitor cells. Strikingly, a subpopulation of embryonic endoderm cells was found to derive from an extraembryonic origin: despite overall transcriptional similarity with embryonic endoderm, this subpopulation was found to express *Trap1a* and *Rhox5*, which are markers of extraembryonic tissues. This finding reveals the power of genetic lineage tracing combined with scRNA-seq to uncover subtle yet crucial lineage and transcriptional decisions during embryonic development. Finally, the number of totipotent

Table 1 | **Overview of the prospective genetic lineage tracing methods**

Method		Type of readout	Readout level	Genetic barcode	Barcode induction	Biological system studied	Ref.
Tagging *Sleeping Beauty* transposase		Lineage barcode	Single cell	Genomic location of randomly integrated transposons	Induction of hyperactive *Sleeping Beauty* transposase by doxycycline injection	Mouse haematopoiesis	7
Polylox barcoding		Lineage barcode	Cell population	Recombined Polylox cassette (10 *loxP* sites interspaced with unique DNA sequences)	Cre recombinase induction by tamoxifen	Mouse haematopoiesis	20
CRISPR–Cas9 in zebrafish	scGESTALT	Lineage barcode and transcriptome	Single cell	Cas9-generated indels in a GESTALT cassette (array of 10 CRISPR–Cas9 targets)	Injection of Cas9 and gRNAs 1–4 into zygotes. Heat shock induction of Cas9 and constitutive expression of gRNAs 5–9	Zebrafish brain	26
	scScarTrace	Lineage barcode and transcriptome	Single cell	Cas9-generated indels in 8 tandem histone–GFP transgenes	Injection of Cas9 (mRNA or protein) and gRNA into zygotes	Zebrafish haematopoiesis, brain, and caudal fin regeneration	25
	LINNAEUS	Lineage barcode and ranscriptome	Single cell	Cas9-generated indels in red fluorescent protein transgenes (16–32 independent integrations)	Injection of Cas9 and gRNA into zygotes	Zebrafish larvae (5 days after fertilization), adult heart, liver, pancreas and telencephalon	27
CRISPR–Cas9 in mouse	MARC1 mouse line	Lineage barcode	Cell population	60 hgRNAs integrated throughout the genome	Constitutive Cas9 expression starts after maternal-to-zygotic transition	Mouse early embryogenesis	39
	Molecular recorder	Lineage barcode and transcriptome	Single cell	Cas9-generated indels in synthetic DNA target sites randomly integrated in the genome	Constitutive Cas9 expression starts after maternal-to-zygotic transition	Mouse early embryogenesis	40

GESTALT, genome editing of synthetic target arrays for lineage tracing; gRNA, guide RNA; hgRNA, homing guide RNA; indels, insertion and deletion mutations; LINNAEUS, lineage tracing by nuclease-activated editing of ubiquitous sequences; MARC1, mouse for actively recording cells 1; sc, single cell.

cells (1–6) and early and late multipotent embryonic progenitor cells (5–17 and 15–52, respectively) present in the mouse embryo was estimated. One current limitation of this molecular recorder lies in the scarcity of the available barcode data, with a barcode recovery rate (percentage of cells with at least one target site recovered by sequencing) ranging from 15% to 73%. This technical limitation forces the averaging of single cells within a cluster to extract lineage identity from the average tissue[40].

The mouse models described above are limited by the difficulty to control the site and number of insertions of the desired construct. This limitation requires a necessary screening step to identify animals qualifying for downstream lineage tracing experiments. Overall, strong proofs of principle for mouse genetic lineage tracing using CRISPR–Cas9 are now available, paving the way for further technological improvements and novel discoveries in the field of embryology.

Retrospective lineage tracing

The genetic lineage tracing methods require the introduction of a DNA barcode into the cell of interest to mark the onset of lineage tracking. This is therefore a powerful approach to study lineage decisions in model organisms, where genetic manipulations can be performed. However, such manipulations are impossible in the context of human development and disease[41]. Luckily, the human genome contains naturally occurring somatic mutations, which can be used as lineage barcodes to track the origins of cells. Many types of mutations match the requirements for a lineage barcode, as they are permanent and transmitted to the progeny[42]. These natural barcodes, when occurring in the nuclear DNA, fall mainly into four categories: CNVs, SNVs, long interspersed nuclear element 1 (LINE-1) retroelements and microsatellite repeats. In addition, naturally occurring mutations can also be found in mitochondrial DNA (mtDNA)[43].

CNVs are stretches of DNA more than 1 kb long present in different copy numbers when compared with the reference genome (FIG. 5a). Many CNVs have been linked to human diseases (especially to cancer), and there is recent evidence of the presence of CNVs also in healthy tissues (for example, in the skin and brain)[44–48]. CNVs have been used to reconstruct the clonal dynamics of breast tumour initiation, invasion and metastasis in multiple studies[49–51]. CNVs were detected in single nuclei of ductal carcinoma following whole-genome amplification and sequencing, despite a reported low coverage of ~6% of the genome of a single cell[49]. After tumour cells had been distinguished from healthy cells (which have a perfectly diploid genome), CNVs between single tumour cells were used to build phylogenetic trees.

Nuc-seq
RNA-sequencing technology for nuclear RNA capture from frozen tissue samples.

Merging of the clonal information with spatial information (obtained by microdissection before isolation of nuclei) showed that cells with similar copy number profiles were in close proximity, thereby revealing that tumour growth occurred through clonal expansion. The development of nuc-seq allowed the profiling of CNVs after whole-genome and whole-exome sequencing of single nuclei, and revealed a large number of de novo

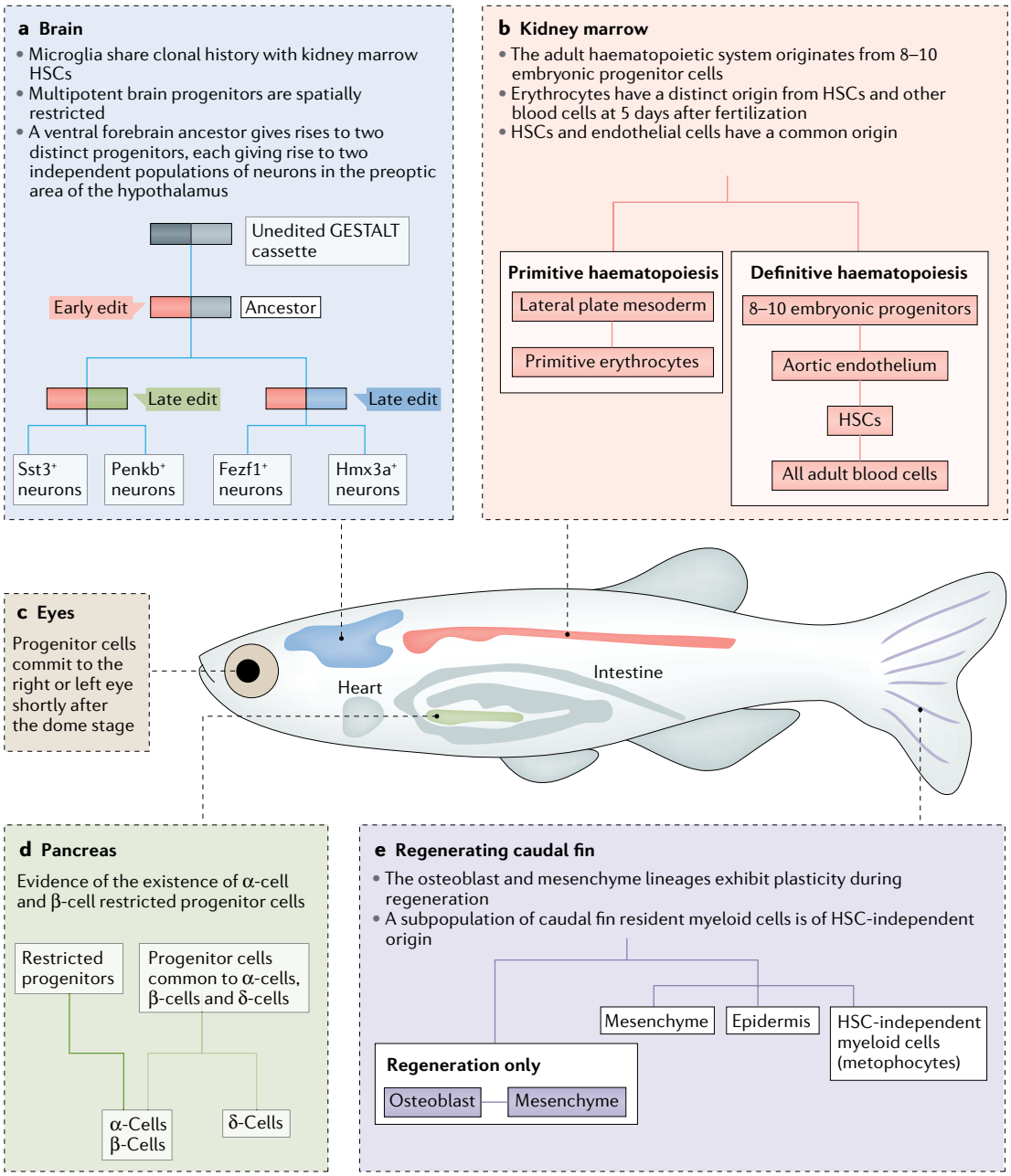


Fig. 4 | **Zebrafish development studies using CRISPR–Cas9 genetic lineage tracing.** Schematic representation of the biological findings made using CRISPR–Cas9 genetic lineage tracing in zebrafish. **a** | In the brain, single-cell genome editing of synthetic target arrays for lineage tracing (GESTALT) and single-cell ScarTrace (scScarTrace) unravelled the clonal origin of microglia and multipotent progenitor cells. Interesting lineage trajectories were identified in the preoptic area of the hypothalamus. **b** | In the kidney marrow, scScarTrace allowed the quantification of the number of embryonic progenitor cells responsible for the production of the haematopoietic system. Lineage tracing by nuclease-activated editing of ubiquitous sequences (LINNAEUS) revealed that erythrocytes circulating in the embryo at 5 days after fertilization are of independent origin compared with the rest of the blood system. **c** | With use of scScarTrace, the clonal composition of the right and left eyes was analysed to pinpoint the developmental stage in which progenitor cells commit to each of the eyes. **d** | LINNAEUS analysis of the pancreas revealed the existence of α-cell and β-cell restricted progenitor cells. **e** | In the caudal fin, scScarTrace revealed the existence of a resident immune cell population sharing its origin with the epidermis and mesenchyme cells. Additionally, clonal analysis of the regenerated fin revealed cell-type plasticity in the osteoblast precursor compartment. HSCs, haematopoietic stem cells.

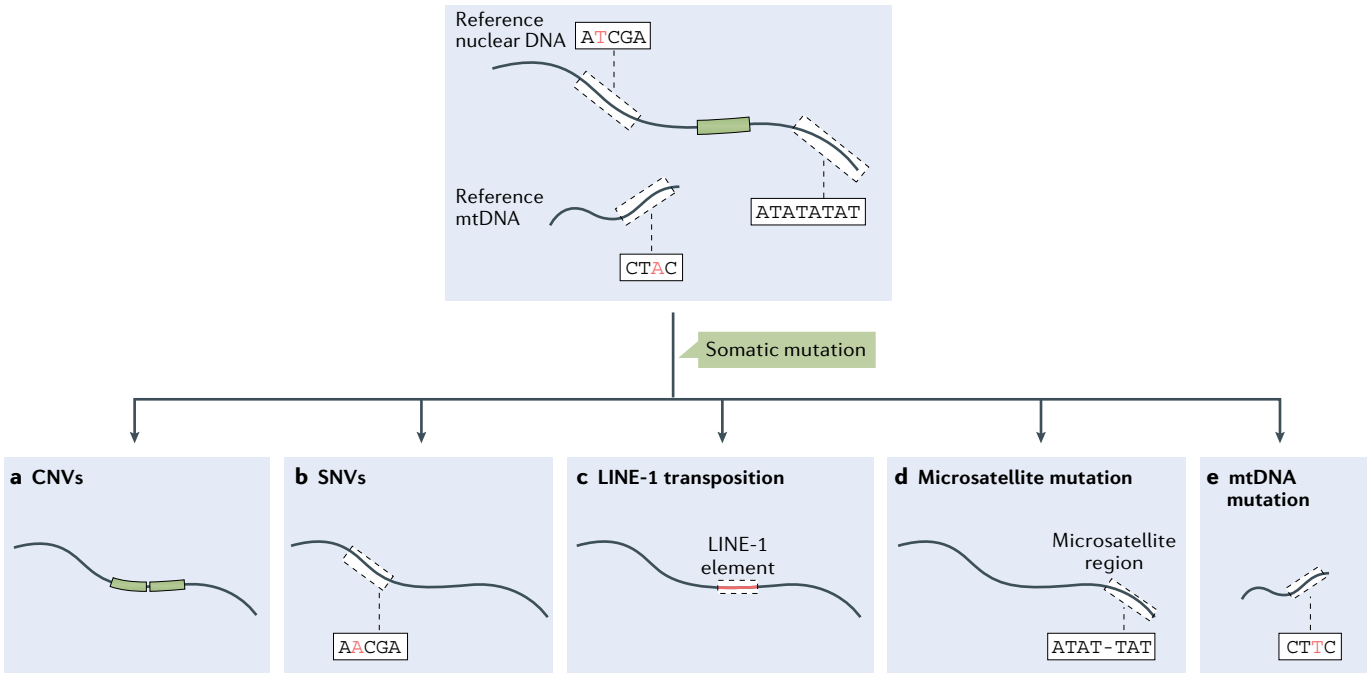


Fig. 5 | **Naturally occurring mutations are used for retrospective lineage tracing.** Several types of mutations can be used for retrospective lineage tracing. Mutation calling is always done by comparing the sequenced material with a reference genome. Several categories of naturally occurring mutations have been used for lineage tracing. **a** | Copy number variations (CNVs) represent variation in the number of a tandemly repeated long DNA sequence. **b** | Single-nucleotide variants (SNVs) are single-base changes. **c** | Long interspersed nuclear element 1 (LINE-1) is a retroelement that can undergo transposition across the genome. **d** | Microsatellite repeats are simple, highly repetitive sequences that are highly susceptible to mutation. **e** | Mutations in mitochondrial DNA (mtDNA) are frequent and can be captured by single-cell RNA sequencing or by single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq).

mutations occurring within clones and generating extensive subclonal diversity. Finally, the recent development of topographic single-cell sequencing allows CNV profiles to be obtained while the special context of single cells within the tumour tissue is preserved[51]. In topographic single-cell sequencing, the tissue of choice is cryosectioned, stained and imaged to generate a global map of the tissue. Then, laser-capture microdissection is performed to capture and transfer single cells into single tubes, where whole-genome amplification and sequencing is performed. Application of this method to 10 ductal carcinomas revealed that CNVs and other de novo mutations occur before tumour invasion and that multiple clones invade the adjacent tissue during tumour growth.

SNVs are frequent variations in a single nucleotide that have been shown to cause different diseases, such as cystic fibrosis and β-thalassaemia[52] (FIG. 5b). However, SNVs also occur in non-coding regions in somatic cells, with no phenotypic consequences. Since SNVs are faithfully transmitted to descendant cells, they allow clone identification and lineage tree reconstruction. Like CNVs, SNVs can be detected by whole-genome or whole-exome sequencing of single cells[53]. A limitation of analysing SNVs in single cells is the sparsity of single-cell data, which makes it difficult to capture the same SNV in a large number of single cells. However, this approach has been used successfully to study clonality and lineage decisions in several healthy and diseased tissues, such as the human brain, colorectal cancer and kidney tumours[54–56]. In the healthy human brain, single nuclei from neurons of three individuals were profiled, and ~1,500 SNVs were identified in each brain, which were generated by transcription-related DNA damage and allowed the reconstruction of the life history of post-mitotic neurons. Five neuronal clades were identified, which reflect the existence of five pluripotent progenitor cells in early embryonic development. The future use of deeper or targeted sequencing approaches is anticipated to improve our ability to identify more SNVs in larger numbers of cells to resolve further the lineage history of human tissues.

Retroelement transposons, especially LINE-1, are abundant in the genome and possess the ability to undergo transposition into new genomic sites on cell division[57]. Owing to their abundance, the unique genomic locations of these mobilized LINE-1 elements can be used as a lineage barcode[58] (FIG. 5c). Most of the work on LINE-1 elements for clone identification and lineage reconstruction was done in the human brain[59,60]. By single-cell whole-genome sequencing, a small number of LINE-1 insertions were identified. Spatial analysis of these insertions in the cortex revealed that one element insertion was spread throughout the cortex, whereas the other insertion was restricted to a smaller region, suggesting that the latter happened later in development. In one of these studies, LINE-1 elements were used in combination with SNVs to refine the lineage map, demonstrating the capacity to combine different naturally occurring mutations for retrospective lineage tracing[60].

Minimum spanning tree
The shortest way to connect all edges of a graph.

Pseudotime
A quantitative measure of biological progression through a process such as cell differentiation.

Microsatellites are simple and short tandem repeats, which are very abundant in the genome[61] (FIG. 5d). Somatic mutations occur frequently in microsatellites as DNA polymerases are likely to slip at such repetitive sequences during DNA replication; the microsatellite mutations can be used for clone tracking. The genomic locations of microsatellite repeats are known, allowing the use of targeted sequencing instead of whole-genome sequencing and thus increasing coverage and reducing cost. So far, microsatellites have been used in mice to reconstruct and analyse the lineage decisions of the female germline and of colonic crypts[62,63].

Finally, mutations in mtDNA occur 10–100 times more frequently than in nuclear DNA and can be used to reconstruct clonal relationships[64] (FIG. 5e). Importantly, the small size of the mitochondrial genome reduces the cost of sequencing for capturing mtDNA barcodes. Mitochondrial sequences can also be analysed by scRNA-seq and by single-cell chromatin accessibility assays such as single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq), thereby offering the possibility of dual readout of clonality and cell state. Such an approach was recently used to retrospectively analyse family relationships in healthy and leukaemic haematopoietic cells[43].

The main limitation of using naturally occurring mutations to build lineage trees is the rarity of these mutations, which therefore require costly deep sequencing of the genome or exome of single cells to capture sufficient data to allow tree reconstruction. Recently, combined measurement of cell type or cell state and mutation profiling in thousands of single cells was performed, representing an important technological improvement for lineage tree building using somatic mutations[65]. Further progress, such as targeted amplification and capture of mutated loci, will allow us to gain a better understanding of mutation occurrence and dynamics during development, differentiation and disease.

scRNA-seq trajectory reconstruction

In parallel with the emergence of prospective and retrospective genetic lineage tracing methods, a large number of computational strategies were developed to infer developmental trajectories from scRNA-seq datasets (FIG. 1c). scRNA-seq provides only a static snapshot of the transcriptional profile of individual cells, but scRNA-seq datasets are now made of thousands of cells spanning different states of a developmental or differentiation trajectory. By assuming that cells with similar transcriptomes are found in close proximity within these differentiation trajectories, it becomes possible to perform trajectory reconstruction. It is important to keep in mind that these methods do not fall under the umbrella of genetic lineage tracing since no genetic tracer is introduced. Nevertheless, by comparison of the output of the molecular recorder described above with the results of the analysis of cell-to-cell transcriptome similarity, it was concluded that cells that are closely related genetically have more similar transcriptomes[40]. A large variety of experimental validations of developmental trajectories inferred from scRNA-seq have been performed and supported this approach. A detailed review of all the trajectory reconstruction algorithms was recently published[66]; therefore, we will focus our discussion on Monocle and RNA velocity, which are two widely used and pioneer algorithms for trajectory reconstruction.

A pioneering method of trajectory reconstruction was Monocle, and its successor is Monocle2 (REFS[30,67]). Briefly, Monocle reduces highly dimensional datasets and projects a minimum spanning tree that connects cells that have similar transcriptional profiles. Then, a backbone is drawn, which connects the least and most differentiated cells studied using the shortest route possible. Finally, all single cells are projected onto this backbone, forming a pseudotime that serves as a lineage trajectory of differentiation of the system studied. Monocle2 improved this process and allows the creation of lineage trees with several branches. Monocle has been used to study a large variety of biological processes, including mouse myoblast differentiation[67], HSC differentiation in wild-type and mutant bone marrow[30], mouse kidney development and differentiation[68], HSC generation in the mouse embryonic aorta[69], the development of human prefrontal cortex[70] and the differentiation trajectory of T cells in breast cancer[71].

RNA velocity, which was recently introduced, relies on the abundance of spliced and unspliced transcripts to predict the future transcriptome of single cells[72]. RNA velocity allows prediction of the future of single cells in a short time window (a few hours). Several biological systems were studied using RNA velocity, including the differentiation of chromaffin cells (neuroendocrine cells from the adrenal medulla) from Schwann cell precursors in E12.5 mouse embryos, the transcription dynamics of neutrophil maturation in the mouse adult bone marrow and the complex lineage branching of the developing mouse hippocampus. RNA velocity is able to study human cell differentiation, such as the differentiation of glutamatergic neurons in the developing human forebrain at 10 weeks after conception.

Conclusions and future perspective

Introducing a DNA barcode or reading out a naturally occurring mutation during development or cell differentiation is now possible. Since introducing a barcode to perform prospective lineage tracing requires genetic manipulation, its use is restricted to in vitro studies or to model organisms. On the other hand, because mutations naturally occur in human development and disease, they can be used for retrospective lineage tracing in the context of human development and/or cancer.

A mouse embryo or a human tumour is made of millions of cells and is highly heterogeneous. So far, obtaining cell identity and lineage information for that many cells has not been achieved. However, only 10 years have passed since the first single-cell transcriptome was sequenced, and today, as many as two million cells can be profiled from developing mouse embryos[73–75]. It is therefore only natural to expect technological improvements in the coming years that will scale up the genetic lineage tracing methods discussed herein and potentially allow the reconstruction of the full lineage tree of highly complex biological systems.

A major limitation of genetic lineage tracing is the absence of spatial resolution, both at the time of barcode induction and at the time of analysis. Despite the development of MEMOIR (memory by engineered mutagenesis with optical in situ readout), which allowed the reconstruction of the clonal dynamics of embryonic stem cells in situ, it is not yet possible in vivo to link the initial position of a cell to its position when its DNA barcode is read out[76]. This limitation still restricts our ability to fully understand developmental processes such as cell migration differentiation. By contrast, DNA barcodes induced with the CRISPR–Cas9 system were shown to be detectable from their transcribed mRNA sequence[26,27,40]. Therefore, spatial transcriptomics could be used to read barcodes, thereby providing a spatial resolution of genetic lineage tracing datasets[77–80]. Recent advances revealed new scalable spatial transcriptomics approaches, potentially capable of analysing a complete zebrafish or mouse embryo[81]. Overall, spatially resolving lineage trees inferred from genetic lineage tracing experiments is a very exciting future development.

Naturally occurring mutations can be used to identify clones in complex human tissues, healthy or cancerous. This retrospective lineage tracing approach has two drawbacks: traceable mutations are quite rare and computational identification of mutations in single cells is difficult. Indeed, looking for an SNV in whole-genome sequencing data is like looking for a needle in a haystack. To confidently call SNVs in single cells, it is necessary to capture at least one of the two copies of the locus of interest and be certain that the sequencing library preparation method did not introduce any sequence error. The end goal is to obtain a robust readout from many cells in the system of interest. Currently, experimental and computational technologies are not robust enough to achieve these criteria. One way to overcome these limitations would be to combine measurements from several genetic elements in the system of interest (for example, CNVs, SNVs and LINE-1 elements) to increase the pool of information available for reconstructing cellular families. Along those lines, characterization of the transcriptome and the T cell receptor repertoire in single T cells allowed clonal tracking of the immune system in colorectal cancer[82]. Finally, whereas so far we have explored only DNA sequences as lineage barcodes, there is considerable potential for using other DNA features as lineage barcodes. For example, cytosine 5-hydroxymethylation patterns were used to track back sister cells in the mouse early embryo[83]. Additionally, DNA methylation has been used to resolve clonal patterns in human fetal haematopoietic progenitor cells and in human adult leukocytes[84–86]. It is highly likely that several other (epi)genetics marks could be used to link cells to their ancestors.

Published online 5 November 2019

1. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode caenorhabditis elegans. *Dev. Biol.* **100**, 64–119 (1983).
2. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148**, 33–45 (2012).
3. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
4. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
5. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
6. Navin, N. et al. Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
7. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014). **This work introduces the *Sleeping Beauty* transposase system for prospective genetic lineage tracing in the mouse and reveals a limited contribution of HSCs to steady-state blood production.**
8. Till, J. E. & Mc, C. E. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* **14**, 213–222 (1961).
9. Becker, A. J., Mc, C. E. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).
10. Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. *Nat. Cell Biol.* **21**, 2–8 (2019).
11. Laurenti, E. & Gottgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).
12. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
13. Metzger, D., Clifford, J., Chiba, H. & Chambon, P. Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc. Natl Acad. Sci. USA* **92**, 6991–6995 (1995).
14. Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
15. Feil, R. et al. Ligand-activated site-specific recombination in mice. *Proc. Natl Acad. Sci. USA* **93**, 10887–10890 (1996).
16. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237**, 752–757 (1997).
17. Rinkevich, Y., Lindau, P., Ueno, H., Longaker, M. T. & Weissman, I. L. Germ-layer and lineage-restricted stem/progenitors regenerate the mouse digit tip. *Nature* **476**, 409–413 (2011).
18. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
19. Snippert, H. J. et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143**, 134–144 (2010).
20. Pei, W. et al. Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. *Nature* **548**, 456–460 (2017).
21. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). **This work introduces the Cre–*loxP*-based Polylox system for inducible recombination of DNA barcodes in mice, which is used to study haematopoiesis.**
22. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
23. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
24. Junker, J. P. *et al*. Massively parallel clonal analysis using CRISPR/Cas9 induced genetic scars. *bioRXiv* Preprint at https://doi.org/10.1101/056499 (2017).
25. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018). **This study presents scScarTrace, a method for prospective genetic lineage tracing in zebrafish using CRISPR–Cas9 to study the clonality of blood, brain, eyes and caudal fin.**
26. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018). **This article describes the developmenr of scGESTALT, a method for prospective genetic lineage tracing in zebrafish using CRISPR–Cas9, and the characterization of lineage relationships between cell types in the adult brain.**
27. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018). **This study introduces LINNAEUS, a method for prospective genetic lineage tracing in zebrafish using CRISPR–Cas9 and uses it to study multiple organs of zebrafish larvae.**
28. Woo, K. & Fraser, S. E. Order and coherence in the fate map of the zebrafish nervous system. *Development* **121**, 2595–2609 (1995).
29. Solek, C. M., Feng, S., Perin, S., Weinschutz Mendes, H. & Ekker, M. Lineage tracing of dlx1a/2a and dlx5a/6a expressing cells in the developing zebrafish brain. *Dev. Biol.* **427**, 131–147 (2017).
30. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). **This study presents Monocle2, an scRNA-seq-based trajectory reconstruction algorithm that connects single cells on the basis of transcriptional similarity to infer a developmental or differentiation pseudotime.**
31. Xu, J. et al. Temporal-spatial resolution fate mapping reveals distinct origins for embryonic and adult microglia in zebrafish. *Dev. Cell* **34**, 632–641 (2015).
32. Henninger, J. et al. Clonal fate mapping quantifies the number of haematopoietic stem cells that arise during development. *Nat. Cell Biol.* **19**, 17–27 (2017).
33. Kissa, K. & Herbomel, P. Blood stem cells emerge from aortic endothelium by a novel type of cell transition. *Nature* **464**, 112–115 (2010).
34. Bertrand, J. Y. et al. Haematopoietic stem cells derive directly from aortic endothelium during development. *Nature* **464**, 108–111 (2010).
35. Tu, S. & Johnson, S. L. Fate restriction in the growing and regenerating zebrafish fin. *Dev. Cell* **20**, 725–732 (2011).
36. Lin, X. et al. An ectoderm-derived myeloid-like cell population functions as antigen transporters for

Langerhans cells in zebrafish epidermis. *Dev. Cell* **49**, 605–617 (2019).
37. Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR–Cas in human cells. *Science* **353**, aag0511 (2016).
38. Kalhor, R., Mali, P. & Church, G. M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* **14**, 195–200 (2017).
39. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, aat9804 (2018).
This study reports the generation of a mouse line (MARC1) for prospective genetic lineage tracing using CRISPR–Cas9, which is used to build lineage trees of early embryogenesis.
40. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
The authors develop a method for prospective genetic lineage tracing using CRISPR–Cas9 in a mouse model to study lineage relationships during early embryogenesis.
41. Biasco, L. et al. In vivo tracking of human haematopoiesis reveals patterns of clonal dynamics during early and steady-state reconstitution phases. *Cell Stem Cell* **19**, 107–119 (2016).
42. Woodworth, M. B., Girskis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017).
43. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339 (2019).
44. Zarrei, M., MacDonald, J. R., Merico, D. & Scherer, S. W. A copy number variation map of the human genome. *Nat. Rev. Genet.* **16**, 172–183 (2015).
45. Abyzov, A. et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–442 (2012).
46. Knouse, K. A., Wu, J. & Amon, A. Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res.* **26**, 376–384 (2016).
47. McConnell, M. J. et al. Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).
48. Cai, X. et al. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **8**, 1280–1289 (2014).
49. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
50. Wang, Y. et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
51. Casasent, A. K. et al. Multiclonal invasion in breast tumors identified by topographic single cell sequencing. *Cell* **172**, 205–217 e212 (2018).
This work demonstrates the capture of CNV by whole-genome amplification in single cells while preserving spatial information using laser-capture microdissection.
52. Katsonis, P. et al. Single nucleotide variations: biological impact and theoretical interpretation. *Protein Sci.* **23**, 1650–1666 (2014).
53. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
54. Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
55. Leung, M. L. et al. Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res.* **27**, 1287–1299 (2017).
56. Xu, X. et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumour. *Cell* **148**, 886–895 (2012).
57. Ostertag, E. M. & Kazazian, H. H. Jr. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
58. Hwang, B. et al. Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements. *Nat. Commun.* **10**, 1234 (2019).
59. Evrony, G. D. et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483–496 (2012).
60. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
61. Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**, 435–445 (2004).
62. Reizel, Y. et al. Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. *PLOS Genet.* **7**, e1002192 (2011).
63. Reizel, Y. et al. Cell lineage analysis of the mammalian female germline. *PLOS Genet.* **8**, e1002477 (2012).
64. Stewart, J. B. & Chinnery, P. F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat. Rev. Genet.* **16**, 530–542 (2015).
In this article, the authors demonstrate that mitochondrial mutations can be detected by scRNA-seq and scATAC-seq and can be used to reconstruct clonal relationships of haematopoietic and leukaemic cells.
65. Nam, A. S. et al. Somatic mutations and cell identity linked by genotyping of transcriptomes. *Nature* **571**, 355–360 (2019).
66. Kester, L. & van Oudenaarden, A. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166–179 (2018).
67. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
68. Park, J. et al. Single-cell transcriptomics of the mouse kidney reveals potential cellular targets of kidney disease. *Science* **360**, 758–763 (2018).
69. Baron, C. S. et al. Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta. *Nat. Commun.* **9**, 2517 (2018).
70. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
71. Savas, P. et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat. Med.* **24**, 986–993 (2018).
72. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
The authors present RNA velocity, a trajectory reconstruction algorithm that uses the abundance of spliced and unspliced mRNA detected by scRNA-seq to predict the future state of individual cells.
73. Tang, F. et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382 (2009).
74. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
75. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
76. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
77. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
78. Stahl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
79. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376 (2018).
80. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
81. Vickovic, S. *et al*. High-density spatial transcriptomics arrays for in situ tissue profiling. *bioRXiv* Preprint at https://doi.org/10.1101/563338 (2018).
82. Zhang, L. et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
83. Mooijman, D., Dey, S. S., Boisset, J. C., Crosetto, N. & van Oudenaarden, A. Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* **34**, 852–856 (2016).
84. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
85. Accomando, W. P., Wiencke, J. K., Houseman, E. A., Nelson, H. H. & Kelsey, K. T. Quantitative reconstruction of leukocyte subsets using DNA methylation. *Genome Biol.* **15**, R50 (2014).
86. Salas, L. A. et al. Tracing human stem cell lineage during development using DNA methylation. *Genome Res.* **28**, 1285–1295 (2018).

Acknowledgements
The authors thank A. Alemany and J. Yeung for valuable feedback on the manuscript.

Author contributions
C.S.B. researched data and wrote the manuscript. A.v.O. reviewed and edited the manuscript.

Competing interest
The authors declare no competing financial interests.

Peer review information
Nature Reviews Molecular Cell Biology thanks Bin Zhou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.