

# Building a lineage from single cells: genetic techniques for cell lineage tracking

Mollie B. Woodworth<sup>1–3</sup>, Kelly M. Girsakis<sup>1–3</sup> and Christopher A. Walsh<sup>1–3</sup>

**Abstract** | Resolving lineage relationships between cells in an organism is a fundamental interest of developmental biology. Furthermore, investigating lineage can drive understanding of pathological states, including cancer, as well as understanding of developmental pathways that are amenable to manipulation by directed differentiation. Although lineage tracking through the injection of retroviral libraries has long been the state of the art, a recent explosion of methodological advances in exogenous labelling and single-cell sequencing have enabled lineage tracking at larger scales, in more detail, and in a wider range of species than was previously considered possible. In this Review, we discuss these techniques for cell lineage tracking, with attention both to those that trace lineage forwards from experimental labelling, and those that trace backwards across the life history of an organism.

## Fate-mapping methods

Approaches that apply a heritable mark to a given progenitor or class of progenitors, then use the inheritance of the mark to define the progeny of that cell or class.

Deriving lineage relationships between cells in a developing organism, and between an early dividing cell of unknown potential and its descendants, have been long-standing interests in developmental biology. Understanding these lineage relationships illuminates the fundamental mechanisms underlying normal development, and can provide insight into pathologies of development and cancer. Lineage relationships are experimentally revealed through fate-mapping methods, and when fate mapping is carried out at single-cell resolution it is known as lineage tracing (also known as lineage tracking).

Fundamental questions of lineage have been addressed since the earliest days of embryology, with technical sophistication increasing over time. Initially, embryologists were limited to visual observation of development in organisms that are small enough to be transparent, such as *Caenorhabditis elegans*, which enabled the discovery of genes that control cell proliferation, cell fate and cell death<sup>1,2</sup>. In species with larger numbers of cells, genetic mosaicism was leveraged to investigate cell fate, by creating chimeric embryos from mouse strains with differing coat colour genes<sup>3,4</sup> or by grafting quail cells into chicken embryos<sup>5</sup>. With the development of radioactive, enzymatic and fluorescent cellular labels, it became possible to selectively label one or more cells by direct injection and trace developmental potential directly<sup>6–9</sup>, although most available labels were subject to dilution with successive cell division<sup>10</sup>.

In recent years, many new methods have emerged to enable cell lineage tracking with increasing resolution, leading to substantial biological insights. In model organisms, novel cellular labels, such as barcoded retroviral libraries<sup>11</sup> and a rainbow of available fluorescent proteins<sup>12</sup>, have increased the number of founder cells that can be uniquely labelled and tracked. Labels can be delivered at different stages of development using various methods, including viral infection and *in utero* electroporation. Unlike most early cellular tracers, labels that are inserted into the genome can permanently mark lineages in a variety of experimental organisms without being diluted by cell division, and these modifications are facilitated by genome-editing technologies, such as the CRISPR–Cas9 system<sup>13</sup>. Furthermore, recent advances in sequencing enable naturally occurring somatic mosaic mutations to be used as lineage marks in cancerous tissue<sup>14,15</sup> and normal tissue<sup>16,17</sup>, illuminating a future in which lineage tracing moves from experimental organisms into humans.

In this Review, we present both historical and recently developed methods for lineage tracing. Following the common division of genetic approaches into ‘forward’ and ‘reverse’ genetics, we discuss methods according to whether they prospectively introduce lineage tracers and follow traced cells forwards in development (prospective lineage analysis), or whether they retrospectively identify lineage-specific tracers and use them to infer past developmental relationships

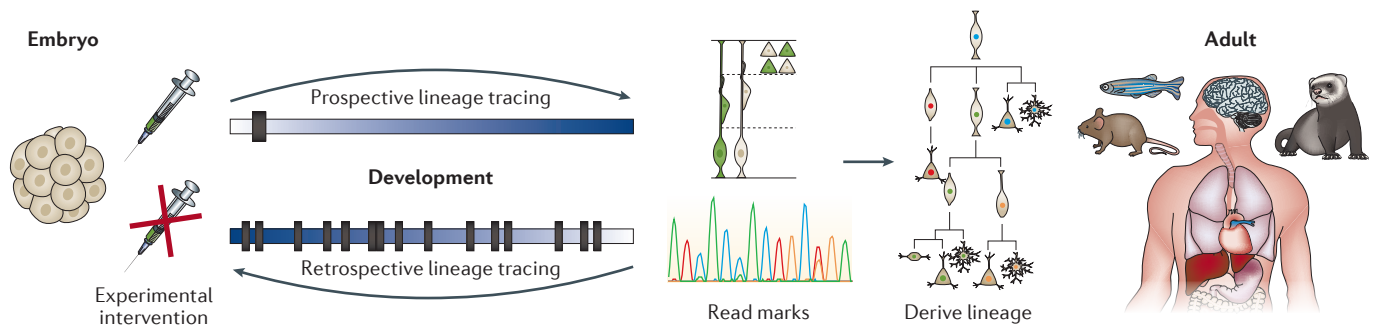
<sup>1</sup>Division of Genetics and Genomics, Manton Center for Orphan Disease, and Howard Hughes Medical Institute, Boston Children’s Hospital, Boston, Massachusetts 02115, USA.

<sup>2</sup>Departments of Neurology and Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA.

Correspondence to C.A.W. [christopher.walsh@childrens.harvard.edu](mailto:christopher.walsh@childrens.harvard.edu)

doi:10.1038/nrg.2016.159  
Published online 23 Jan 2017



**Figure 1 | Prospective and retrospective lineage tracing.** Prospective lineage tracing entails experimentally applying a lineage mark (grey rectangle on the blue timeline), then following cells forward to read its output at some later time. By contrast, retrospective lineage tracing follows cells backwards to read endogenous marks (multiple grey rectangles on the blue timeline) that have accumulated over the lifetime of an organism. Compared with retrospective lineage tracing, prospective lineage tracing generally requires greater experimental intervention at the onset of development (left), but less intervention to read the result of lineage tracing (right). In both experimental designs, cells are placed in a dendrogram according to their inferred relationships with each other.

(retrospective lineage analysis) (FIG. 1). We highlight technologies and methods that can make important contributions to the execution and the interpretation of lineage tracing experiments. We conclude with a discussion of systems and organs that present promising or challenging prospects for lineage tracing.

**Prospective methods of lineage tracing**

A classic approach to cell lineage analysis is to label a single founder cell and trace its progeny over time. This prospective method has been used since biological dyes mapped the fate of cells within chicken and mouse embryos in early observational studies, and continues to be used in current lineage tracking experiments<sup>18,19</sup>. Early developmental studies hoped to achieve clonal labelling by microinjecting small amounts of dye into an area of interest, whereas advances in genetic tools for prospective lineage tracing now allow for far greater cell and tissue specificity, recombinase-based intersectional analyses and single-cell resolution (FIG. 2; TABLE 1).

**Sparse retroviral labelling for lineage tracing.** Since the advent of recombinant DNA technology in the late 1980s, retroviral libraries that contain reporter transgenes such as  $\beta$ -galactosidase ( $\beta$ -gal) and green fluorescent protein (GFP) have been used for cell labelling and lineage tracing in vertebrate animal models<sup>20,21</sup>. Retroviral vector-mediated gene transfer allows viruses to introduce recombinant DNA into the genome of a host cell. Viruses are applied at limiting dilutions with the goal of labelling single founder cells. The integrated exogenous DNA is then inherited by all the descendants of the infected cell. The DNA encodes a histochemical or fluorescent protein that can be easily assayed to label cells of a ‘clone’ and to elucidate cell fate choices within that clone. Histological and morphological analyses of the progeny of virally infected cells allows for post hoc fate mapping within a clonally related cell population.

Sparse retroviral infection has also been used in live-cell imaging of progenitors and their progeny in organotypic slice culture. Mouse, ferret, chimpanzee and

human progenitors have all been analysed using time-lapse imaging. Individual progenitors that have been labelled with fluorescent reporter genes are visualized using confocal microscopy for multiple cellular divisions. At the end of the imaging experiment, immunohistochemistry and cellular morphology can then be used to analyse cell fate within the imaged clone<sup>22–26</sup>. Although *ex vivo* organotypic culturing conditions closely mimic the *in vivo* cellular environment, it is important to consider that progenitor divisions and behaviour may differ from that observed *in vivo*, and such experiments can usually be carried out for only a few days at most, and thus cannot typically relate clonal relationships to adult structure.

Initial studies using sparse retroviral labelling inferred clonality on the basis of the proximity of cells that express a reporter gene. Early studies in the cerebral cortex soon showed that sibling cells dispersed widely from one another in some clones<sup>27</sup>. To analyse such widespread clones, the first retroviral libraries were developed, encoding the lacZ gene as a reporter, but also using short DNA fragments to function as barcode tags<sup>28</sup>. Clonal relationships were then directly revealed through PCR amplification of the integrated barcode tags from cells dissected from tissue sections, rather than being inferred on the basis of proximity alone (FIG. 2a). Cells derived from a common progenitor share the same DNA tag at the vector integration site (IS) regardless of their patterns of migration, whereas clonally unrelated cells harbour different barcodes (FIG. 3). The first library of 100 tags soon expanded to 1,000 tags<sup>29,30</sup> and then to essentially unlimited complexity using random oligonucleotide barcodes of identical size but with distinct sequences<sup>31,32</sup>.

Advances in transgenic animal lines have also extended the applications of retroviral genetic tagging and fate mapping. Cell type specificity can now be achieved using transgenic mouse lines that express virus receptors under the control of a cell type-specific promoter<sup>33,34</sup>. Only cells that contain the virus receptor can be infected and express the reporter gene or barcode,

**Genetic mosaicism**

The state of containing more than one distinct genome within a single organism, whether achieved by experimental means (by combining early-stage embryos from different individuals or species) or by natural means (by considering differences in DNA from cell to cell).

**Prospective lineage analysis**

An approach that applies an experimental label to cells, which is then examined at some point in the future to construct a lineage tree looking forwards from development.

**Retrospective lineage analysis**

An approach that uses naturally occurring labels (for example, somatic mutations) to construct a lineage tree looking backwards at development.

**Intersectional analyses**

Using two attributes of a cell population (for example, the expression from two different promoters) to select only cells that display both attributes.

**Organotypic slice culture**

A culture system in which a slice of tissue is cultured, rather than a collection of dissociated cells, to more closely mimic the biological context of an organ.

allowing for more precise viral targeting *in vivo*. Barcode tags can then be recovered using fluorescence-activated cell sorting (FACS) with the fluorescent reporter transgene or using laser capture microdissection (LCM) techniques that can preserve cellular position within the infected tissue for future reconstruction and analysis.

Although retroviral library labelling is an advantageous method for determining lineage relationships both *in vivo* and *ex vivo*, this technique does have some considerations and limitations: only cells with the capacity to divide will propagate the barcode to progeny; retroviral vectors can spontaneously silence, such that many retrovirally transfected cells are no longer histochemically labelled even though their DNA can be detected in the tissue; and barcode tag recovery from single cells can be challenging<sup>32,33</sup>. However, experiments can be designed to mitigate these drawbacks, and current technological advances may circumvent some of these limitations. Retroviral silencing is thought to be a stochastic event, thus overall clonal size or complexity may be underestimated but should not otherwise skew experimental results. To circumvent this challenge, new studies have been combining retroviral library labelling with high-throughput next-generation sequencing. This advance, which has been used to track mouse haematopoietic stem cells *in vivo*, allows for not only a more precise barcode identification and quantification compared with Sanger sequencing but also single-cell sensitivity<sup>35</sup>.

**Plasmid transfection labelling for lineage tracing.** In addition to viral infection, reporter transgenes for cell labelling and fate mapping can be introduced into cells by DNA plasmid transfection. Lipofection, which is a common lipid-based system, has been used to transfect the developing *Xenopus laevis* retina and to trace retinal cell fate *in vivo*<sup>36</sup>. Lipofection continues to be a popular method for both *in vivo* and *in vitro* lineage studies. Electroporation, which is an alternative non-viral delivery method, uses electrical fields to increase cell membrane permeability to recombinant DNA. Electroporation has also been used to deliver reporter transgenes that encode fluorescent proteins to track cells both *in vitro* and in various vertebrate animal models<sup>37,38</sup>. To introduce recombinant DNA plasmids into neural progenitors *in vivo*, *in utero* electroporation (IUE) has proved to be an efficient technique<sup>38</sup>. Reporter gene plasmids can be injected into the ventricles of the developing brain and then introduced into neural progenitors that line the ventricular wall by electrical pulses. A reporter transgene, such as GFP, is then carried episomally by the progenitor cell and passed on to subsequent daughter cells. Unlike retroviral labelling, however, plasmid DNA is not integrated into the genome of the progenitor and becomes diluted or inactivated in progeny after serial cellular divisions. Plasmid electroporation techniques, therefore, are transient and fail to label the entire lineage<sup>39</sup>.

A solution to plasmid loss and inactivation is a DNA transposon system, which stably integrates the reporter transgene into the genome of the progenitor (FIG. 2b). Transposon systems include Mos1, Tol2, Sleeping Beauty

(SB) and piggyBac (PB) which mobilize through a cut-and-paste mechanism<sup>40–42</sup>. The typical transposon system is used in a dual plasmid format, in which a donor plasmid contains the reporter transgene of interest and a helper plasmid expresses the transposase. The donor plasmid includes terminal repeats that flank the transgene, which allows for random genomic integration by the transposase. The transgene is then propagated to all progeny within the lineage but with limited further transpositional mobility because the transposase, like any episomal plasmid, will be diluted over cellular divisions. Expression from donor and helper plasmids can be driven by different promoters, allowing for cell type specificity and genetic intersectional analyses. Compared with other transposon systems, piggyBac has a more precise cut-and-paste mechanism, higher transposition efficiency and a larger cargo capacity<sup>43</sup>. These attributes have made the piggyBac transposon system particularly popular. In addition, piggyBac transposase can be co-electroporated with multiple fluorescent reporter constructs, each of which is driven by a cell type-specific promoter. In this experimental design, multiple lineages can be examined in a single animal<sup>44</sup>. PiggyBac has been successfully used in multiple mammalian cell lines and in combination with *in utero* electroporation to track and manipulate cell lineages in animal models<sup>44–47</sup>.

The piggyBac transposon plasmid system allows for remarkable flexibility and cell type specificity, but as with any random genomic insertion event, the precise location and number of transposition occurrences introduces a risk of confounded results owing to mutagenesis. Transposition of the reporter transgene may cause endogenous genes at or near the insertion site to become unintentionally dysregulated. One study, however, found no evidence of mutagenesis by transposon insertion in cells that were labelled using the piggyBac IUE method<sup>43</sup>. Transposase plasmid systems are a remarkable tool for transgenesis and cell lineage tracking in both classically genetically modifiable animal models, such as mice, and in otherwise non-genetically tractable animals, such as the ferret.

**Genetic recombination for lineage tracing.** Cell lineage tracing by genetic recombination is able to leverage the expression of recombinase enzymes in a cell-specific or tissue-specific manner to activate the expression of a conditional reporter gene. Two genetically encoded, site-specific recombination systems include *Cre-loxP* and *FLP-FRT*. In the *Cre-loxP* system, mice are engineered to express Cre recombinase under the control of a chosen promoter, limiting Cre expression to a specific tissue or cell type<sup>48</sup> (FIG. 2c). These lines are then crossed with a second line in which a reporter transgene, such as lacZ or GFP, is preceded by a *loxP*-flanked transcriptional stop (*loxP*-STOP-*loxP*) cassette. In cells that express Cre recombinase, the STOP sequence is excised and the reporter transgene is expressed. Temporal control of recombination can be gained by using an inducible Cre system, which selectively activates Cre under promoters that may also be active at undesired time points, such as embryogenesis. In an inducible system, Cre recombinase

#### Cut-and-paste mechanism

Method of mobilization by class II DNA transposable elements, in which the transposon excises itself from its genomic location using transposase protein and integrates into a new target site.

#### Cre-loxP

A genetic system derived from P1 bacteriophage and adapted for use in genetically modifiable organisms. The site-specific recombinase Cre inverts or recombines any sequence located between 34 bp *loxP* sites, depending on their orientation.

#### FLP-FRT

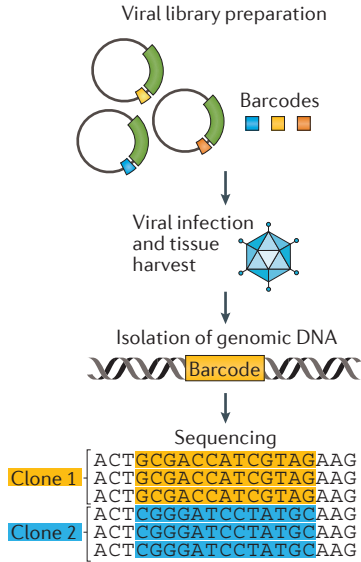
A genetic system derived from *Saccharomyces cerevisiae* and adapted for use in genetically modifiable organisms. The site-specific recombinase FLP inverts or recombines any sequence located between 34 bp *FRT* sites, depending on their orientation.

#### loxP-STOP-loxP

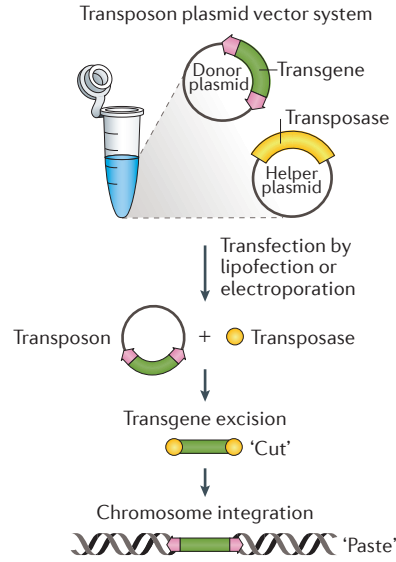
A DNA element containing a transcription termination sequence flanked by *loxP* sequences, allowing the transcription termination sequence to be removed by the activity of Cre recombinase.



**a Retroviral library infection**

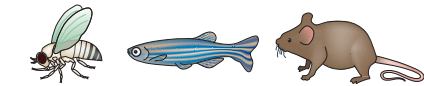
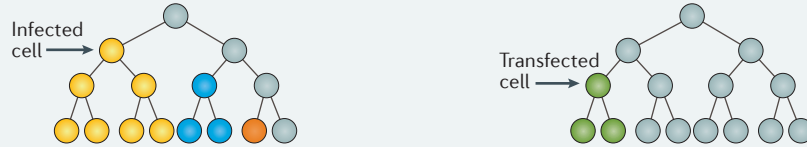


**b Plasmid transfection**

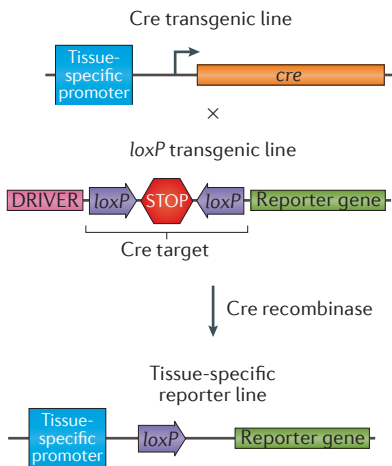


**Reconstruction of lineage:**

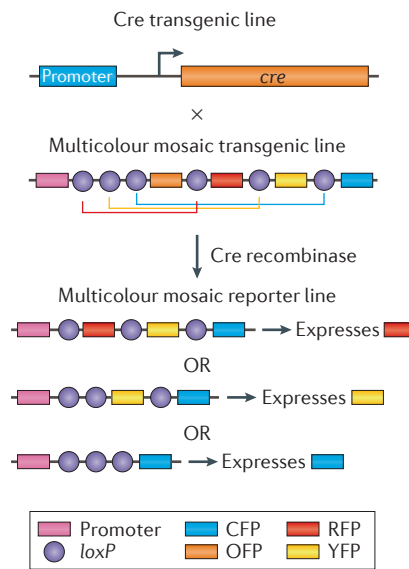
Developmental time



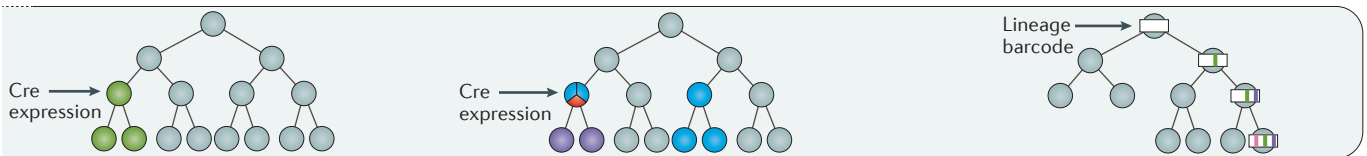
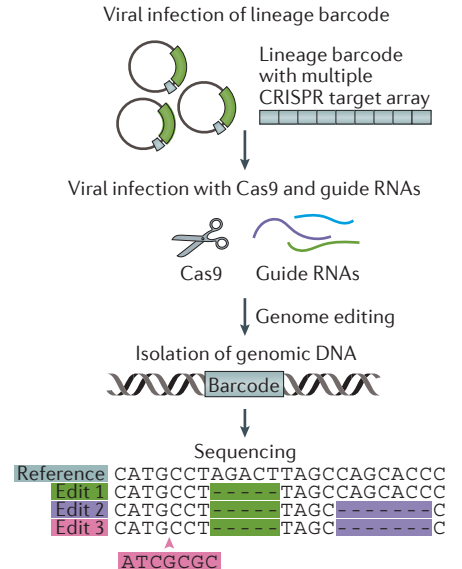
**c Tissue-specific genetic recombination**



**d Multicolour mosaics**



**e CRISPR—Cas9 genome-editing systems**





**Pulse**

An experiment in which a brief bolus of label is followed by a period with no label, allowing events that occurred within a specific time window to be marked.

**Leakiness**

Activity in the absence of inducing signal.

is fused to the human oestrogen or progesterone receptor and activated only in the presence of an anti-oestrogen, such as tamoxifen, or an anti-progestin, respectively. A pulse of tamoxifen administration with an inducible Cre system can be used to determine lineage relationships<sup>49</sup>. Leakiness is a common problem of inducible Cre systems<sup>50</sup> but, nonetheless, these inducible systems have been used for lineage tracing in many adult tissues.

To gain even more cell type specificity, an intersectional approach with the Split-Cre or a combination of the Cre-*loxP* and FLP-*FRT* site-specific recombination systems may be used. The Split-Cre system expresses two cleaved, inactive Cre fragments with each driven by a different promoter. Only when both promoters are concurrently expressed within the same cells of a population will the Cre enzyme be reconstituted, the STOP cassette be excised and the reporter transgene expressed<sup>48</sup>. Cre-*loxP* can also be combined with the FLP-*FRT* system for higher-resolution fate mapping and a reduction in background leakage<sup>51–53</sup>. In this intersectional method, both site-specific recombinases are required to excise two STOP cassettes and activate the reporter transgene. Once the STOP cassettes are removed, all progeny will also express the reporter transgenes.

In addition to single reporter transgene recombination mouse lines, dual or multicolour reporter lines have become increasingly popular for tracking cell lineage relationships. Mosaic analysis with double markers (MADM) uses a Cre-*loxP* system to express GFP and red fluorescent protein (RFP) in cell populations of interest<sup>54</sup>. Before recombination, no reporter transgene is expressed, but after Cre recombinase is activated, one or both of the transgenes are reconstituted. Green, red or double-labelled yellow cells are generated depending on the recombination and the chromosomal segregation type. MADM can be used with cell type-specific

and inducible Cre systems to provide single-cell resolution and to more precisely examine progenitor division patterns<sup>33,54–56</sup>. Multicolour lineage tracing is also possible with recent mouse reporter lines, including Brainbow and Confetti<sup>57,58</sup> (FIG. 2d). The Brainbow mouse lines harness stochastic Cre-mediated recombination using incompatible loxP sites to drive the combinatorial expression of fluorescent reporter transgenes. The Brainbow mouse can label individual cells with as many as 90 distinguishable colours through the stochastic expression of several fluorescent reporter transgenes. Cells that express a particular colour share a common lineage. A modified line, the Confetti mouse, ubiquitously expresses Cre from the *Rosa26* locus and has been used to track individual stem cell lineages in the mouse intestinal crypt<sup>58</sup>. With the expression of a multitude of unique colours, co-staining with antibodies to determine protein expression within Brainbow or Confetti mice is nearly impossible. Endogenous fluorescence of the reporter genes, however, can be used for imaging clones. Advances in microscopy, such as the two-photon microscope, continue to make these lines an attractive choice for *in vivo* cell lineage tracing.

It is important to note that these lineage-marking strategies are not mutually exclusive. Viral vectors can carry Cre recombinase to sparsely activate transgene recombination, and the Cre-*loxP* system can be used to drive the conditional expression of viral receptors, adding a greater level of cell type specificity<sup>59,60</sup>. Viral libraries may combine exogenous DNA barcodes with multicolour reporters to trace cell lineage. Using this type of marking strategy, clonal relationships of murine hepatocytes and leukaemic cells were recently investigated<sup>61</sup>. Other multicolour mosaic constructs such as Brainbow can also be expressed in the form of a viral vector library for random colour expression in infected cells. This method has been used successfully to visualize multiple clones in a single developing mouse embryo<sup>12</sup>. The multiaddressable genome-integrative colour (MAGIC) marker toolkit, which is a recent transposon-based Brainbow transgene method, has been used to track progenitors in both the embryonic mouse brain and spinal cord<sup>62</sup>. As new transgenic animal models and genetic labelling tools are developed, experiments that harness multiple technologies in concert will continue to remain a powerful approach for cell-specific and tissue-specific tracing in model organisms and culture.

**Recent methodological advances in prospective lineage tracing.** Innovations in both microfluidic platforms and genome-editing strategies have also recently been used to prospectively track cell lineage<sup>63,64</sup> (TABLE 1). Advances in microfluidic technologies now allow for the capture and culture of single progenitor cells and up to five generations of their progeny on a single chip. *In vitro* time-lapse imaging for both division kinetics and the identification of lineage relationships can be coupled with on-chip immunohistochemistry to assess cell fate within the captured clones. Clones can also be retrieved after culturing for single-cell transcriptomics with known lineage relationships. Kimmerling *et al.*<sup>63</sup>

◀ **Figure 2 | Highlighted genetic methods and strategies for prospective lineage tracing in vertebrate animal models and cell culture.** Early observational lineage studies used biological dyes for cell labelling and analysis, whereas advances in recombinant DNA technology, transgenesis and genome-editing platforms have revolutionized prospective lineage tracing. Although not mutually exclusive, these featured techniques are commonly used for the tracking of cell lineage and cell fate in animal models and cell culture. **a** | Sparse retroviral labelling integrates a reporter transgene and a short DNA barcode tag into the genome of the host cell. After propagation to progeny, cells derived from a common progenitor share the same barcode, whereas clonally unrelated cells harbour different barcodes. **b** | In a transposon plasmid vector system, such as piggyBac, a helper plasmid expressing a transposase excises ('cut') and integrates ('paste') a reporter transgene from a donor plasmid into the genome of a cell. Once the transgene is integrated, all daughter cells within that lineage will express the reporter. **c** | Genetic recombination systems, such as Cre-*loxP*, leverage the expression of recombinase enzymes to activate the expression of reporter genes in a cell-specific or tissue-specific manner. Once Cre is activated within a cell, all progeny will express the exogenous reporter gene. **d** | Much like single-colour reporters, multicolour mosaic systems harness recombination to label lineages with multiple unique colours. In the schematic, stochastic recombination at various loxP sites allows for the combinatorial expression of multiple fluorophore colour combinations. **e** | Genome-editing systems express a lineage barcode with a CRISPR target array that progressively and stably accumulates mutations over cellular divisions. Much like retrospective tracing, lineage relationships are reconstructed on the basis of the pattern of shared mutations among cells. CFP, cyan fluorescent protein; OFP, orange fluorescent protein; RFP, red fluorescent protein; YFP, yellow fluorescent protein.

used this microfluidic trap array technology, paired with single-cell RNA sequencing (RNA-seq), to look at both interclonal and intracolon variability in activated CD8<sup>+</sup> T cells; they demonstrated that lineage-dependent transcriptional profiles corresponded to functional cellular phenotypes. This study was the first to link single-cell transcriptomics with cell lineage history<sup>63</sup>. Combining prospective lineage tracing with RNA-seq allows for the overlay of phenotypic cell identity with genetic lineage information for a more comprehensive view of clonal relationships. Moving forwards, the layering of ‘omics’ technologies such as transcriptomics and proteomics with genetic-based tracking will allow for the deeper analysis of identity and lineage within a cell population.

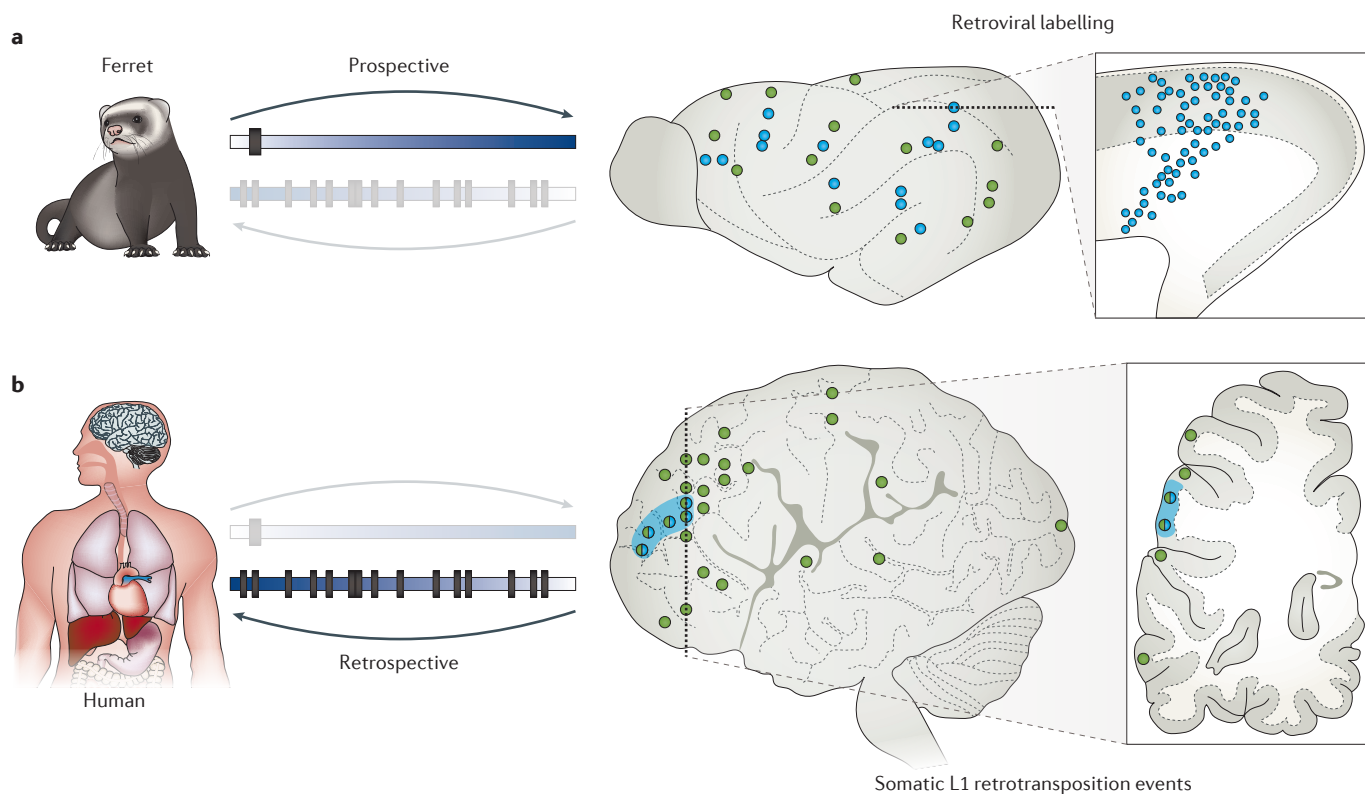
Recently, the CRISPR–Cas9 genome-editing technology has been used to track and to synthetically reconstruct cell lineage relationships in complex, multicellular organisms (FIG. 2e). McKenna *et al.*<sup>64</sup> developed genome editing of synthetic target arrays for lineage tracing (GESTALT), a highly multiplexed method that uses barcodes that consist of multiple CRISPR–Cas9 target sites<sup>64</sup>. These barcodes progressively and stably

accumulate unique mutations over cellular divisions and can be recovered using targeted sequencing. Cell lineage relationships are determined on the basis of the pattern of shared mutations among analysed cells. Although prospective in the sense that the barcode is introduced at the start of the experiment, the GESTALT method also parallels retrospective, somatic-mutation-based tracking (discussed below). The incrementally edited barcodes from thousands of cells were then used in large-scale reconstructions of multiple cell lineages within cell culture and zebrafish. Although the precise anatomical position and cell type of each assayed cell cannot be determined using this method, this study and other emerging studies demonstrate the potential for cumulative and combinatorial barcode editing in prospective lineage tracing of whole organisms<sup>64–68</sup>. Advances during the past 30 years, since the advent of genetic barcoding and recombinase-based transgenic animals, have allowed prospective cell lineage tracking experiments not only to uncover clonal relationships at the single-cell level but also to map cell fate choices in a wide variety of cells, tissues and model organisms.

Table 1 | Lineage reconstruction techniques for prospective tracing

Lineage marking method	Reconstruction strategy	Requirements	Refs
<b>Microscopy or live imaging</b>			
Retroviral infection	After sparse viral infection, all progeny of that lineage will carry the reporter gene. Lineages expressing the reporter are visualized using microscopy.	Virus and microscope	22–27
Plasmid transfection	After transfection, progeny of the lineage will carry the reporter gene until cellular divisions dilute the episomal plasmid. Transposon systems can be used to integrate the transgene into the genome of the host to mitigate plasmid loss. Lineages expressing the reporter are visualized using microscopy.	Transfection agent or electroporator and microscope	36–39, 43–47
Tissue-specific genetic recombination	After Cre-based recombination, all progeny of that lineage will carry the reporter gene. Lineages expressing the reporter are visualized using microscopy.	Genetically modifiable lines and microscope	51–53
Multicolour mosaics	After Cre recombination, all progeny of that lineage will carry a combination of fluorophores. Lineages expressing a certain hue are visualized using microscopy.	Genetically modifiable lines and microscopes capable of resolving multiple colours	12, 54–62
Microfluidic capture	A single founder cell is cultured on a microfluidic chip and the next five generations of progeny are captured downstream using a hydrodynamic trap. Time-lapse imaging is used to visualize cellular divisions in real time.	Microfluidic chip and microscope	63
<b>Sequencing of viral barcodes</b>			
Retroviral library infection	After viral infection, all progeny of that lineage will carry the reporter gene and a unique barcode. Once cells are isolated and sequenced, clones will harbour the same barcode. If LCM is used, the precise anatomical position of the cell within a clone may be recovered. Using sequenced barcodes, lineages can be clustered and plotted into dendrograms.	Constructed barcoded library, method for isolating cells for sequencing (FACS or LCM) and sequencing analysis	28–35
<b>Sequencing of edited barcodes</b>			
CRISPR–Cas9 genome-editing systems	After viral infection, the lineage barcode will incrementally accumulate mutations in progeny over cellular divisions. Once cells are isolated and sequenced, lineage barcode hierarchies can be determined using maximum parsimony methods and plotted into dendrograms.	Virus with target CRISPR array barcode, Cas9, guide RNAs and sequencing analysis	64–66

FACS, fluorescence-activated cell sorting; LCM, laser-capture microdissection.



**Figure 3 | Prospective and retrospective lineage tracing of brain development.** This figure illustrates studies for assessing questions of neuronal migration and lineage, including whether neurons that share a common origin are physically adjacent to each other in the brain, and whether closely related cells are more likely to be adjacent than more distantly related cells, by prospective and retrospective lineage tracing. **a** | Using a prospective method in a model organism (ferret), cortical cells are traced using the injection of a tagged retroviral library, revealing two clonal lineages that are widely distributed across the brain (blue and green). A sagittal section of parietal cortex is shown following analysis by microscopy, demonstrating that blue-lineage neurons migrate into the cortex and spread laterally in a cone-shaped structure<sup>136,137</sup>. **b** | A similar, but retrospective, analysis carried out in human brain identifies somatic long interspersed nuclear element 1 (L1) retrotransposition events by sequencing and digital droplet PCR, revealing a widespread clone resulting from an early retrotransposition event (green) and a smaller clone that is restricted to a small region of frontal cortex, resulting from a later retrotransposition event (blue)<sup>81</sup>. Results from both approaches are consistent, leading to the conclusion that lineages that are marked early in mammalian neuronal development are spread across the brain and intermingled with other lineages, but those that are marked late in neuronal development are more spatially restricted and physically coherent. Part **a** is adapted from REF. 136, Ware, M. L., Tavazoie, S. F., Reid, C. B. & Walsh, C. A. Coexistence of widespread clones and large radial clones in early embryonic ferret cortex, *Cereb. Cortex*, 1999, **9** (6), 636–645, by permission of Oxford University Press, and from REF. 137, republished with permission of The Company of Biologists Ltd, from Clonal dispersion and evidence for asymmetric cell division in ferret cortex, Reid, C. B., Tavazoie, S. F. & Walsh, C. A. **124** (12), 1997. Part **b** is adapted with permission from REFS 17,81, AAAS and Elsevier, respectively.

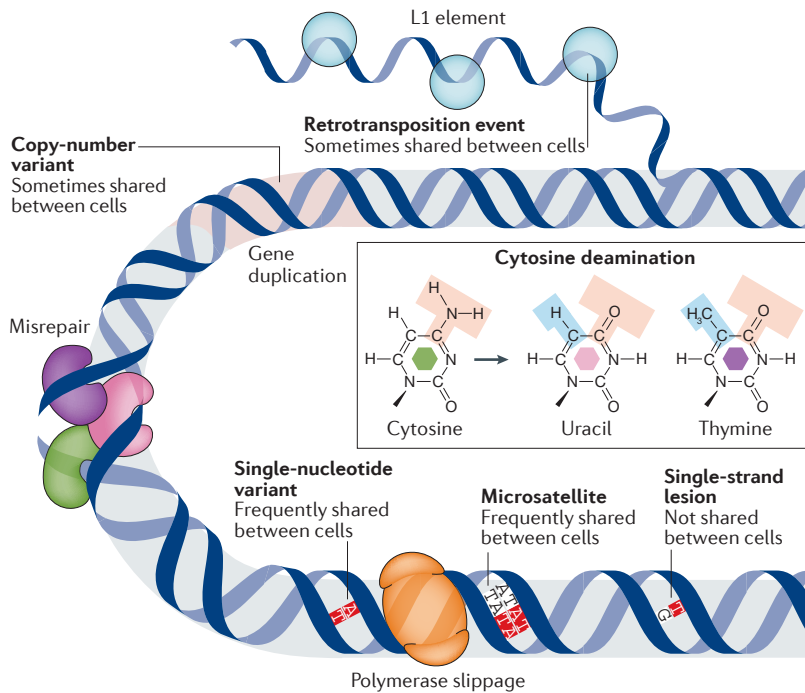
### Retrospective methods of lineage tracing

It has only recently become possible to harness naturally occurring mutations to infer cell lineage information retrospectively, mostly owing to advances in the genome sequencing of single cells. Similar to prospective lineage tracers in model organisms, somatic mutations indelibly mark the progeny of the dividing cell in which they occurred, and the cells bearing these naturally occurring lineage marks can be later analysed to reconstruct the genealogy of organs and cell types<sup>69</sup>. To use naturally occurring somatic mutations for lineage tracking, it is first necessary to discover the mutations that are shared between multiple cells from that individual, but because somatic mutations are, by their nature, low frequency, they are difficult to identify through the sequencing

of a mixed population of cells at conventional depths. Innovations in next-generation sequencing, such as the declining cost of deep next-generation genome sequencing and the advent of single-cell genome sequencing, have made it possible to discover rare mutations that mark minority lineages within a larger cellular population<sup>70</sup>. These variants, from the least frequently somatically mutated to the most frequently mutated, include retrotransposons, copy-number variants (CNVs), single-nucleotide variants (SNVs) and microsatellites (FIG. 4). The different rates at which these variants occur in somatic tissues allow lineage tracing experiments to be conducted at different levels of granularity according to the types of variants, tissue and disease state selected (TABLE 2), although precise frequency estimates for each

#### Microsatellites

(Also known as short tandem repeats). Short genomic repeats consisting of a set of tandem nucleotides, with repeat numbers varying between different alleles.



**Figure 4 | Somatic mutation in the genome.** Somatic mutations in the genome include (in order of increasing frequency): long interspersed nuclear element 1 (L1) retrotransposition events, copy-number variation, single-nucleotide variants, microsatellite (short tandem repeat) variants and single-strand lesions. Each class of mutations is caused by different environmental stressors, such as DNA polymerase slippage for microsatellites and cytosine deamination for single-nucleotide lesions. Furthermore, each class of mutation has different functional consequences for the genome of the cell in which it occurs, such as gene or enhancer disruption (L1 retrotransposition) and increased protein production (copy-number variation).

type of variant have not been measured in a consistent way, and are therefore not provided here. Single-cell genome sequencing promises to revolutionize lineage tracking in humans; however, whole-genome sequencing currently requires considerably more DNA than the 6 picograms that are present in a single cell, necessitating pre-sequencing genome amplification, which can introduce technical artefacts and complications to a lineage-tracing experiment<sup>71,72</sup>.

**Somatic mutations for lineage tracing in normal tissue.** Endogenous retroelements, which principally include long interspersed nuclear element 1 (L1; also known as LINE-1) elements, constitute much of the human genome; L1 elements alone constitute nearly one-fifth of the genome<sup>73</sup>. A very small number of these L1 elements retain the ability to mobilize in humans and can insert into a new genomic location during somatic cell division<sup>74</sup>, which has raised substantial interest in their potential contribution to somatic diversity, especially within complex tissues, such as the brain<sup>75</sup>. Large numbers of apparent somatic L1 mobilization events were suggested by initial experiments using quantitative PCR (qPCR)<sup>76</sup> or DNA sequencing<sup>77</sup> from bulk human brain, but more precise estimates of L1 mobilization frequency that have been derived by sorting single neurons, amplifying the whole genome and analysing L1

retrotransposition at a single-cell level<sup>78</sup>, suggest fewer than one somatic insertion per neuronal genome on average<sup>78</sup>. A second study suggests higher rates (10–15 somatic insertions per genome)<sup>79</sup>, but this study is subject to criticism for the inclusion of sequencing and other technical artefacts, the removal of which reduces the estimated rate to <1 somatic insertion per neuron<sup>80</sup>. A single-neuron whole-genome sequencing study<sup>81</sup> confirms the low rate of L1 retrotransposition events but also illustrates the striking spatial distribution patterns of clonal retrotransposition events, providing strong proof of principle for the use of spontaneous somatic L1 events for lineage tracing. Using a digital droplet PCR assay, one somatic L1 insertion was found across the cortex, and the other somatic L1 insertion was restricted to a small region of prefrontal cortex, indicating that the first L1 insertion occurred early in brain development, with the second occurring later<sup>81</sup> (FIG. 3).

Subchromosomal somatic copy-number variation is common in human tissues, and somatic CNVs are potentially useful lineage-tracing tools owing to the relative ease with which they can be detected from single-cell sequencing data. Large subchromosomal somatic CNVs can be detected in normal skin<sup>82,83</sup> and brain<sup>83–85</sup>, and these studies report large proportions of skin cells and neurons, approximately 30–70%, that contain at least one somatic CNV, including a small number of shared CNVs that arose during development<sup>85</sup>. Furthermore, the analysis of clonal CNVs can also illuminate genes and lineages that are responsible for disease; for example, brain tissue from patients with hemimegalencephaly contains neurons with somatic copy-number gains of chromosome 1q (containing the growth-promoting gene *AKT3*)<sup>85</sup>. CNVs are particularly promising as lineage-marking somatic variants; unlike other types of somatic mutations, they can be identified from low-coverage (<1×) sequencing, given sufficiently even genome amplification (see ‘Methodological considerations for retrospective lineage tracing’, below), making the sequencing of many single cells for variant discovery a cost-effective strategy<sup>86</sup>.

SNVs are a major source of evolutionary and disease-causing mutations, although they can also occur very frequently in non-coding portions of the genome without functional effects on somatic cells<sup>87</sup>. Thus, somatic SNVs represent a rich source of lineage-marking mutations, as they are both abundant and can be expected to be frequently functionally neutral. Indeed, pioneering work in mouse stomach, intestine and prostate<sup>16</sup>, and mouse brain<sup>88</sup> and human brain<sup>17</sup>, suggests that somatic SNVs can be identified from single cells or clones and used to reconstruct developmental lineages. These works disagree as to the precise rates of mutation, which is potentially attributable to differences in species and methodology, as two studies amplified mouse single-cell genomes *in vivo* by organoid cell culture<sup>16</sup> or somatic cell nuclear transfer<sup>88</sup>, estimating approximately 100–600 somatic SNVs per cell, and one study amplified human single-cell genomes from post-mortem tissue *in vitro* by multiple displacement amplification (MDA)<sup>17</sup>, estimating approximately 1,500 somatic SNVs per cell. These

**Digital droplet PCR**

A polymerase chain reaction in which the reaction is divided into thousands of small droplets, allowing absolute quantification of PCR products.

**Organoid**

A three-dimensional culture model of a whole or partial organ or tissue.



Table 2 | Examples of somatic mosaic mutations identified by sequencing

Tissue	Genome source	Sequencing method	Approximate depth of coverage	Genome amplification method	Refs
<b>L1 retrotransposition event</b>					
Normal human brain	Single cell	L1 insertion profiling (L1-IP)	Low (0.35x)	MDA and MALBAC	78
Normal human brain	Single cell	WGS	High (40x)	MDA	80
Normal human brain	Single cell	Retrotransposon capture sequencing (RC-seq)	Low (0.35x)	MALBAC	79
<b>Copy-number variant</b>					
Human breast cancer	Single cell	WGS	Low (0.05x)	<i>In vivo</i> amplification and DOP-PCR	14
Normal human skin	Single cell	WGS	High (20x)	<i>In vivo</i> amplification	82
Human cancer cell line	Single cell	WGS	High (25x)	MALBAC	119
Normal human brain	Single cell	WGS	Low (0.04x)	MDA and GenomePlex	84
Human breast cancer	Single cell	WGS	High (50x)	<i>In vivo</i> amplification and DOP-PCR	15
Normal human brain	Single cell	WGS	Low (0.2x)	GenomePlex	85
Human breast cancer line	Single cell	WGS	Low (0.05x)	DOP-PCR	86
Normal rat brain	Single cell	WGS	Low (1x)	MALBAC and GenomePlex	118
Normal human brain, keratinocytes	Single cell	WGS	Low (0.1x)	DOP-PCR	83
<b>Single-nucleotide variant</b>					
Human leukaemia	Tumour and relapse samples	WGS	High (25x)	None	105
Human myoproliferative neoplasm	Single cell	WES	High (15x)	MDA	108
Human kidney tumour	Single cell	WES	High (15x)	MDA	100
Human breast cancer	Single cell	WGS	High (50x)	<i>In vivo</i> amplification and DOP-PCR	15
Human bladder cancer	Single cell	WES	High (40x)	MDA	99
Normal mouse gut	Single cell	WGS	High (30x)	<i>In vivo</i> amplification	16
Human leukaemia	Single cell	Targeted sequencing	Not provided	MDA	109
Normal human skin	Biopsy	Targeted sequencing	Very high (500x)	None	106
Normal human brain	Single cell	WGS	High (40x)	MDA	17
Normal mouse brain	Single cell	WGS	High (40x)	<i>In vivo</i> amplification	88
<b>Microsatellite</b>					
Mutant mouse tumour	Single cell	Targeted genotyping	NA	MDA	103
Normal mouse (various tissues)	Single cell	Targeted genotyping	NA	<i>In vivo</i> amplification	69
Mutant mouse colon	Single cell	Targeted genotyping	NA	MDA	93
Mutant mouse oocytes	Single cell	Targeted genotyping	NA	MDA	94
Human leukaemia	Single cell	Targeted genotyping	NA	MDA	104
Normal human brain	Single cell	WGS	High (40x)	MDA	81

Recent work identifying somatic mutations is organized by type of variant interrogated and date of publication. Each study's source tissue and source genome are listed, together with the genome amplification and sequencing approach used and approximate depth of coverage (if given). DOP-PCR, degenerate oligonucleotide priming PCR; L1, long interspersed nuclear element 1; MALBAC, multiple annealing and looping-based amplification cycles; MDA, multiple displacement amplification; NA, not applicable; WES, whole-exome sequencing; WGS, whole-genome sequencing.

discrepancies could be resolved by the development of new algorithms that are specifically designed for the interpretation of single-cell genome-sequencing data<sup>89</sup>, and also by increasing the variety of cell types and tissues subjected to single-cell genome sequencing. Regardless of the precise rate of SNV mutation in somatic tissues, it is clear that somatic SNVs can be used as endogenous lineage tracers; in one study, 9 of 16 sequenced neurons, and 136 of 226 total neurons from the same area of cortex, could be placed in a lineage tree with four independent clades that diverged before gastrulation. One clade contained a nested set of 11 somatic mutations, which were progressively regionally restricted across the brain and were present in progressively decreasing frequency in bulk tissue<sup>17</sup>, suggesting that the analysis of such nested mutations might allow the examination of the progressively branching lineage trees that characterize the developing embryo.

The most frequently mutated somatic loci are likely to be microsatellites, as DNA polymerase slipping makes them highly variable both between and within individuals<sup>90</sup>. Owing to the instability of microsatellite repeats, the analysis of all microsatellite locations in the genome is predicted to be capable of reconstructing the entire cell lineage tree of an organism<sup>91</sup>, using methods adapted from organism-level phylogenetic analysis<sup>92</sup>. Microsatellites have been used to reconstruct the cell lineage decisions that lead to the development of colonic crypts<sup>93</sup> and the female germ line<sup>94</sup>; in the female germ line study, 81 microsatellite loci were analysed in mismatch repair-deficient mice (which have elevated rates of microsatellite instability), allowing the oocyte lineage to be reconstructed in comparison to cells from the bone marrow and ovarian cumulus cells. The oocytes formed a lineage that was distinct from both bone marrow and cumulus cells, but oocytes from the left and right ovaries did not form distinct sub-clusters, demonstrating that oocytes were generated at a time before the segregation of somatic cells on the left and right sides of the body, and thus were not generated during postnatal life<sup>94</sup>. Similar to microsatellites, the polyadenylated tracts following somatic L1 retrotransposition events are subject to frequent DNA polymerase slippage during replication and, therefore, lineages that are defined by a somatic L1 retrotransposition event can be further delineated by analysing poly(A) tail polymorphisms<sup>81</sup>.

**Somatic mutations for tracing cancer evolution.** Cell lineage tracing is useful for describing the natural history of a tumour, as lineage analysis can identify the source of a metastasis or the accumulation of mutations that lead to unchecked growth. Although frequently mutated microsatellite loci were identified in cancer cells two decades ago and used to mark tumour lineages<sup>95,96</sup>, tracing complex mutational paths in cancerous tissue required the advent of more rapid and comprehensive methods, especially single-cell sequencing and bioinformatic analysis, to identify minority clones with multiple progressive mutations<sup>97,98</sup>. Lineage tracing in cancer tissue has several advantages compared with lineage tracing in normal tissue, including the ability to compare a tumour

sample with a paired normal sample<sup>99,100</sup>, and the availability of tumour samples from living individuals owing to surgical resection<sup>15</sup>. Furthermore, the rapid mutation rate and genomic instability of cancer cells generates large numbers of clonal mutations and rearrangements, which themselves facilitate lineage-tracing analyses<sup>101</sup>, enabling cancer biologists to draw important biological insights even in the few years since single-cell sequencing has become possible.

Lineage tracking has been powerfully applied to investigate tumour evolution over time through the comparison of initial tumour samples with metastases or relapse samples<sup>102</sup>. In one pioneering study, 37 cells from a primary tumour, two secondary metastases and surrounding normal tissue were removed from a mismatch repair-deficient mouse using LCM. Cellular genomes were subjected to whole-genome amplification and genotyped at 100 microsatellite loci. Tumour cells formed a coherent clade and phylogenetically clustered away from normal surrounding cells, with physically adjacent tumour cells tending to be more closely related by lineage than non-adjacent tumour cells, indicating that the tumour mostly grew in place without substantial cellular migration<sup>103</sup>. Using similar methods, microsatellite-based lineage tracking was applied to paired original and relapse samples from patients with leukaemia, identifying some relapses that resulted from slowly dividing cells that were present in the original sample, and others that resulted from the enrichment of particular subclones that were present in the original sample or from a lineage that was almost entirely distinct from the original sample<sup>104</sup>. Although functionally neutral microsatellite mutations in single cells are likely to provide a more unbiased survey of lineage variation, mutational burden in leukaemia is high enough that sequencing candidate somatically mutated genes in original and relapse samples, even in bulk rather than in single cells, can also establish a detailed picture of clonal evolution in relapse. In one study, two major patterns of evolution were identified: in the first case, a fairly homogeneous initial clone remained the dominant clone in relapse; and in the second case, a minor subclone from the initial sample became the dominant clone in relapse, with the other original subclones lost following initial treatment. In both cases, the dominant subclone in relapse tended to acquire further mutations, possibly as a result of the treatment itself<sup>105</sup>.

As the preceding studies make clear, a more comprehensive understanding of cellular heterogeneity is crucial for understanding both the development of cancerous lesions and their resistance to treatment. Deep sequencing of small skin biopsy samples demonstrates that normal skin, which is an organ that is exposed to considerable environmental mutagens, carries a heavy burden of mutation, and this heterogeneity in normal tissue provides ample raw material for the development of malignancies, as oncogenic mutations are frequently found and positively selected even in healthy skin tissue<sup>106</sup>. Similarly, primary pancreatic tumours contain substantial numbers of deleterious mutations up to a decade before the origination of subclones that are

#### Clade

A group on a dendrogram (tree diagram) that is separate from another group.

#### Nested

To have a set fully contained within a broader set.

capable of giving rise to metastases<sup>107</sup>. The expansion of mutation-carrying subclones is not strictly necessary for the development of some malignancies, such as kidney clear cell renal cell carcinoma, which seem to be genetically heterogeneous when subjected to single-cell exome sequencing, with little evidence of dominant subpopulations by principal component analysis or phylogenetic clustering<sup>100</sup>. By contrast, as suggested by the relapse studies discussed above, leukaemia samples generally demonstrate a clear clonal structure when analysed using single-cell sequencing, with dominant oncogenic clones<sup>108</sup> and clonal structural variation occurring before the acquisition of oncogenic point mutations<sup>109</sup>.

As with species-level evolution by natural selection, single-cell lineage analyses strongly suggest that tumours evolve irregularly, with periods of mutational stasis followed by punctuated expansions. In breast cancer tissue, 100 single genomes were analysed from two tumours for ploidy and copy-number variation, and investigators identified a few distinct primary subclones in each case, rather than a large number of more closely related subclones. This indicates that large-scale genomic aberrations accumulated in punctuated bursts, with mutated cell populations rapidly emerging to dominate the cancer cell population<sup>14</sup>. By contrast, the analysis of point mutations in breast cancer suggests that these primary subclones are established early and remain stable through later tumour evolution, but are quite genetically heterogeneous at the single-nucleotide level, with each cell carrying a unique mutational burden<sup>15</sup>. Similarly, single-cell exome sequencing of bladder cancer cells demonstrated the presence of two late-occurring subclones that constituted approximately 70% of the tumour, indicating that continuing single-nucleotide mutation can generate highly proliferative clones that can be positively selected and that can come to dominate the tumour population in a short time period<sup>99</sup>.

**Methodological considerations for retrospective lineage tracing.** Retrospective lineage tracing based on the analysis of somatic mutation often entails analysing the genomes of single cells or small groups of cells, and so the DNA must be amplified to generate enough material for next-generation sequencing. The process of amplification, like cell division itself, is inherently error-prone, and can create amplicons that contain sequence or structural errors, which produce false-positive mosaic structural variants, microsatellite variability and SNVs. In addition, uneven amplification across the genome can produce false-positive CNV calls, as well as false-negative sequence calls, in the case of allelic dropout<sup>72,110</sup>. When designing single-cell sequencing experiments, it is therefore important to consider the frequencies and types of errors that are introduced and to select an approach that best balances signal and noise for the experiment at hand<sup>71</sup> (TABLE 2).

One broad class of whole-genome amplification strategies is based on amplifying the genome *in vitro* using highly processive DNA polymerases, and another is based on amplifying the genome *in vivo*, in cells or whole organisms, by cloning and cell culture (TABLE 2).

The earliest *in vitro* approach to be developed, MDA, takes advantage of the high processivity of  $\Phi$ 29 DNA polymerase to generate long linear amplicons. Secondary priming and extension occur from newly synthesized amplicons, increasing amplification efficiency<sup>111</sup>. MDA generates 15–20  $\mu$ g of DNA from a single nucleus<sup>78</sup>, and MDA-amplified single-cell DNA is sufficiently high quality for calling somatic retrotransposition events<sup>81</sup> and SNVs<sup>17,112</sup>. Several groups have recently described methods for partitioning MDA reactions into nanolitre-sized droplets<sup>113–115</sup> or by using microfluidic devices<sup>116</sup>, which increases the uniformity of amplification and reduces reagent costs. A second approach, degenerate oligonucleotide priming PCR (DOP-PCR), involves the fragmentation of the genome into small pieces, followed by amplification with random priming<sup>117</sup>. This method amplifies the genome more evenly than MDA, and is thus particularly well-suited to studying copy-number variation<sup>14,85,86,118</sup>. Hybrid methods, including multiple annealing and looping-based amplification cycles (MALBAC) and PicoPlex, include pre-amplification with a tagged primer. Full amplicons contain complementary sequences at each end, creating hairpins or loops, preventing them from overamplification<sup>119</sup>. MALBAC-based amplification is more even across the genome than MDA-based amplification, but error rates are higher<sup>72</sup>. With this more even amplification, hybrid methods are appropriate for investigating CNVs<sup>118</sup>, structural variants<sup>120</sup> and retrotransposition events<sup>79</sup>, although chimeric amplification products that occur in the MALBAC and MDA reactions are of particular concern for interpreting retrotransposition events and structural variants<sup>80</sup>.

As sequence errors introduced by DNA polymerase, as well as chimeric amplification products, can create difficulties for interpreting sequencing from DNA amplified *in vitro*, some groups have developed approaches that use cell division to amplify genomic DNA. Even selecting cells in G2/M phase for sequencing, after they have replicated their genomes, leads to remarkable improvement in dropout rates and false-positive calls<sup>15,121</sup>, although this approach is not applicable to non-dividing cell populations. Other groups have turned to selecting single cells (and reproducing their genomes by somatic cell nuclear transfer if they are terminally differentiated), and then growing clonal populations in induced pluripotent stem cell (iPSC) or organoid culture<sup>16,82</sup>, or in a cloned experimental organism<sup>88</sup>, and sequencing in bulk. These methods are an interesting solution to the problem of errors occurring due to DNA polymerase, although it is not clear to what degree the single-stranded lesions that exist in the genomes of terminally differentiated cells are stable after re-activating cell cycle-dependent repair processes<sup>122</sup>, which would tend to deflate the number of somatic mutations recovered.

A further methodological consideration in single-cell sequencing is selecting single cells for analysis. For many cell populations, fluorescence-activated sorting can be used to sort single cells or nuclei<sup>78,123,124</sup>, but this method partly depends on finding an antibody or

#### Allelic dropout

One of two alleles at a genomic locus fails to amplify and is therefore not recovered in sequencing data. Compare with locus dropout, in which both alleles at a given locus fail to amplify.

#### Chimeric amplification

In whole-genome amplification, when one amplicon misprimed another locus, leading to a hybrid DNA product with sequences from the original amplicon adjacent to those from the second locus.

#### Induced pluripotent stem cell

(iPSC). A cell that is capable of giving rise to daughter cells of many or all lineages, derived by reprogramming of an adult cell using pluripotency factors.

cellular characteristic that is specific to the cell population under study. Alternatively, cells or nuclei can be triturated and manually selected under microscope guidance<sup>100,108,125</sup>, or FACS-purified then subjected to manual selection for single cells<sup>79</sup>. Manual selection is less expensive than FACS, and more broadly applicable to cells that lack a specific antibody marker, but it requires finer motor control on the part of the operator. Laser-capture microdissection allows the selection of single cells in their native tissue context, although it is difficult to ensure the capture of precisely one nucleus without leaving chromosomes behind<sup>126,127</sup>.

As retrospective lineage tracing by single-cell sequencing matures, it will be crucial to develop methods that allow the visualization of mutations *in situ*, thus maintaining the tissue context of a mutation-carrying cell. CNVs and retrotransposition events are large enough for detection by traditional fluorescence *in situ* hybridization (FISH), but SNVs are more challenging to detect *in situ*<sup>128</sup>. Because coding SNVs are present in many copies of mRNA, they can be analysed using a modified FISH protocol, with mRNAs amplified by rolling circle amplification and detected with padlock probes<sup>129</sup>; SNVs in the mitochondrial genome can be detected using a similar method<sup>130</sup>, and SNVs in genes that are expressed at high levels in a given tissue are amenable to detection by fluorescent *in situ* sequencing (FISSEQ)<sup>131–133</sup>. Fortunately for investigators using SNVs that occur in the large proportion of the genome that is non-coding or poorly expressed, and for those using archival fixed tissue, a new FISH method that is sensitive to SNVs in genomic DNA has been developed recently using allele-specific PCR<sup>134</sup>.

### Perspectives

When designing a lineage tracing experiment, it is important to consider the strengths and weaknesses of either a prospective or a retrospective approach (FIG. 3). For success in prospective lineage tracing, there must be genetic access to the population in question, whether through a regionally directed method such as viral injection and electroporation, or by using population-specific marker lines and promoters. Because prospective lineage tracing depends on labelling and follow-up analysis, its use is restricted to experimental organisms and cell culture systems, whereas retrospective lineage tracing can investigate lineage directly in human tissue. This unprecedented access to human lineage information provides investigators with a wealth of data relevant to human development and disease. However, investigators must carefully select subjects and cells to identify pools of informative variants that differ between the experimental populations in question. Retrospective lineage tracing heavily relies on sequencing, often of single cells, and is therefore currently lower throughput and more expensive than most prospective methods. Although emerging prospective lineage systems engineer revolutionary ways to investigate lineage in model organisms, it will always be necessary to retrospectively map lineage in a naturally occurring tissue without engineered lineage marks.

Whether one chooses a prospective or a retrospective lineage tracing method, the choice of organ or cell population to investigate has an effect on the ease of lineage tracing and the questions that can be asked. Organs that are relatively homogenous in terms of cellular composition, such as the liver, will require less information about specific genetic or protein population markers to investigate than will more diverse organs, such as the immune system, and a diverse organ is likely to have a more complex lineage structure, which could be more difficult to fully investigate. If the tissue is primarily composed of post-mitotic cells, such as the composition of the kidney, lineage information from development will be preserved for a retrospective lineage analysis, but the lack of proliferative capacity means that forward lineage tracing is restricted beyond a certain developmental point. In a tissue with continuously proliferating progenitors, such as the skin, prospective lineage tracing is possible throughout life, but the loss and replenishment of post-mitotic cells removes important sources of retrospective lineage data. Organs that are structured with proliferating populations located adjacent to a luminal space, such as the intestine or embryonic brain, are more accessible to injection and infection than those without a lumen. Finally, for human studies, the accessibility of the tissue in question from patients or donors is crucial. Some specimens can be obtained from routine biopsies, such as skin, or from minimally invasive procedures, such as blood draws, whereas others are not possible to obtain from living subjects. Blood and tissue banks have been established for certain tissues, particularly the brain, and several disease-specific post-mortem tissue banks allow researchers to study lineage in pathological conditions.

Several major recent funding initiatives aim to trace lineage in whole organs or organisms, and scalable methods supported by these initiatives are beginning to bear fruit. Notably, the Paul G. Allen Foundation issued a call for proposals in 2014 for strategies that tracked lineage by barcoding large numbers of cells, which funded the development of an innovative whole-organism approach to lineage tracing using genome-editing-based barcodes<sup>64</sup>. The US National Institutes of Health (NIH) BRAIN Initiative also solicited applications in 2014 for proposals that generated a census of cell types in the brain, and several awardees proposed the identification of lineage relationships between cell types in addition to enumerating the cell types themselves. One project funded by this initiative has produced a method for RNA sequencing of single cells in nanolitre-sized droplets, enabling the sequencing of many cells with reduced cost and preparation time compared with other single-cell RNA-seq methods. As a proof of principle, this sequencing was carried out on 44,000 retinal cells, and the data were used to derive classes of cells and the gene expression relationships between them, which may relate to their lineage relationships<sup>135</sup>. As funding initiatives promote the development of new large-scale lineage-tracing methods, it will increasingly become possible to trace the lineage of organs and organisms at a scale that the early developmental biologists could never have imagined.

#### Triturated

A homogeneous solution created by mixing or grinding, such as pipetting cells up and down to create a uniform suspension.



No longer limited to observing the development of transparent organisms or tracking a small number of cells with serially diluted dyes, biologists can now access a variety of methods for tracing lineage forwards from the application of a genetic label. In addition, recent advances in sequencing, particularly genome sequencing of single cells, allow lineage tracing to be carried out retrospectively, reconstructing lineage decisions that occurred months or years before sequencing. Retrospective lineage tracing can be carried out in normal tissue, examining developmental relationships between cells, and in pathological states such as cancer,

enabling the reconstruction of tumour evolution. In both prospective and retrospective lineage tracing experiments, biological differences between tissues and experimental organisms inform appropriate choices in experimental approach. Furthermore, within a broad experimental strategy, the choice of amplification, sequencing and visualization methods must be adapted to the biological question under study. One hundred years after the first investigations of cell lineage, developmental biologists have built a tremendously enriched genetic toolkit for examining the developmental fate decisions that construct a whole organism.

- Deppe, U. *et al.* Cell lineages of the embryo of the nematode *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **75**, 376–380 (1978).
- 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).
- Mintz, B. Gene control of mammalian pigmentary differentiation. I. Clonal origin of melanocytes. *Proc. Natl Acad. Sci. USA* **58**, 344–351 (1967).
- Kelly, S. J. Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. *J. Exp. Zool.* **200**, 365–376 (1977).
- Le Douarin, N. M. & Teillet, M. A. The migration of neural crest cells to the wall of the digestive tract in avian embryo. *J. Embryol. Exp. Morphol.* **30**, 31–48 (1973).
- Rosenquist, G. C. The location of the pregut endoderm in the chick embryo at the primitive streak stage as determined by radioautographic mapping. *Dev. Biol.* **26**, 323–335 (1971).
- Johnson, M. H. & Zimek, C. A. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71–80 (1981).
- Lawson, K. A., Meneses, J. J. & Pedersen, R. A. Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer. *Dev. Biol.* **115**, 325–339 (1986).
- Pedersen, R. A., Wu, K. & Bakakier, H. Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Dev. Biol.* **117**, 581–595 (1986).
- Clarke, J. D. & Tickle, C. Fate maps old and new. *Nat. Cell Biol.* **1**, E103–E109 (1999).
- Gerrits, A. *et al.* Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618 (2010).
- Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W. & Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013).
- Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
- Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011). **The authors pioneered the sequencing of dividing cancer cells by single-nucleus sequencing (SNS) to construct the lineage of a breast tumour.**
- Wang, Y. *et al.* Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160 (2014).
- Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014). **This study uses cell culture as an *in vivo* DNA-amplification method, sequencing clonal lines of mouse cells to identify lineages in adult tissues.**
- Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015). **The authors sequence normal human brain tissue and identify SNVs, which they use to place adult neurons in a lineage tree.**
- Beddington, R. An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *Development* **64**, 87–104 (1981).
- Serbedzija, G. N., Bronner-Fraser, M. & Fraser, S. E. A vital dye analysis of the timing and pathways of avian trunk neural crest cell migration. *Development* **106**, 809–816 (1989).
- Turner, D. L. & Cepko, C. L. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**, 131–136 (1987).
- Frank, E. & Sanes, J. R. Lineage of neurons and glia in chick dorsal root ganglia: analysis *in vivo* with a recombinant retrovirus. *Development* **111**, 895–908 (1991).
- Brown, K. N. *et al.* Clonal production and organization of inhibitory interneurons in the neocortex. *Science* **334**, 480–486 (2011).
- Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).
- Noctor, S. C., Martinez-Cerdeño, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat. Neurosci.* **7**, 136–144 (2004).
- Gertz, C. C., Lui, J. H., LaMonica, B. E., Wang, X. & Kriegstein, A. R. Diverse behaviors of outer radial glia in developing ferret and human cortex. *J. Neurosci.* **34**, 2559–2570 (2014).
- Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
- Walsh, C. A. & Cepko, C. L. Clonally related cortical cells show several migration patterns. *Science* **241**, 1342–1345 (1988).
- Walsh, C. A. & Cepko, C. L. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* **255**, 434–440 (1992).
- Walsh, C. A. & Cepko, C. L. Clonal dispersion in proliferative layers of developing cerebral cortex. *Nature* **362**, 632–635 (1993).
- Reid, C. B., Liang, I. & Walsh, C. Systematic widespread clonal organization in cerebral cortex. *Neuron* **15**, 299–310 (1995).
- Golden, J. A., Fields-Berry, S. C. & Cepko, C. L. Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc. Natl Acad. Sci. USA* **92**, 5704–5708 (1995).
- Fuentealba, L. C. *et al.* Embryonic origin of postnatal neural stem cells. *Cell* **161**, 1644–1655 (2015).
- Mayer, C. *et al.* Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* **87**, 989–998 (2015).
- Harwell, C. C. *et al.* Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* **87**, 999–1007 (2015).
- Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat. Biotechnol.* **29**, 928–933 (2011). **The authors demonstrate a method for combining viral genetic barcoding with high-throughput next-generation sequencing, then apply this method to follow the differentiation of single haematopoietic stem cells.**
- Holt, C. E., Garlick, N. & Cornel, E. Lipofection of cDNAs in the embryonic vertebrate central nervous system. *Neuron* **4**, 203–214 (1990).
- Emerson, M. M. & Cepko, C. L. Identification of a retina-specific Otx2 enhancer element active in immature developing photoreceptors. *Dev. Biol.* **360**, 241–255 (2011).
- Fukuchi-Shimogori, T. & Grove, E. A. Neocortex patterning by the secreted signaling molecule FGF8. *Science* **294**, 1071–1074 (2001).
- LoTurco, J., Manent, J.-B. & Sidiqi, F. New and improved tools for *in utero* electroporation studies of developing cerebral cortex. *Cereb. Cortex* **19** (Suppl. 1), i120–i125 (2009).
- Wu, S., Meir, Y. & Coates, C. J. piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc. Natl Acad. Sci. USA* **103**, 15008–15013 (2006).
- VandenDriessche, T., Ivics, Z., Izsvák, Z. & Chuah, M. K. L. Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood* **114**, 1461–1468 (2009).
- Yoshida, A. *et al.* Simultaneous expression of different transgenes in neurons and glia by combining *in utero* electroporation with the Tol2 transposon-mediated gene transfer system. *Genes Cells* **15**, 501–512 (2010).
- Chen, F. & LoTurco, J. A method for stable transgenesis of radial glia lineage in rat neocortex by piggyBac mediated transposition. *J. Neurosci. Methods* **207**, 172–180 (2012).
- Sidiqi, F. *et al.* Fate mapping by piggyBac transposase reveals that neocortical GLAST+ progenitors generate more astrocytes than nestin+ progenitors in rat neocortex. *Cereb. Cortex* **24**, 508–520 (2014).
- Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005). **This study demonstrated that the piggyBac transposon could be used for genetic manipulation in human and mouse cells.**
- Wilson, M. H., Coates, C. J. & George, A. L. PiggyBac transposon-mediated gene transfer in human cells. *Mol. Ther.* **15**, 139–145 (2007).
- Woltjen, K. *et al.* PiggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766–770 (2009).
- Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6861–6865 (1992).
- Branda, C. S. & Dymecki, S. M. Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell* **6**, 7–28 (2004).
- Greig, L. C., Woodworth, M. B., Greppi, C. & Macklis, J. D. Ctip1 controls acquisition of sensory area identity and establishment of sensory input fields in the developing neocortex. *Neuron* **90**, 261–277 (2016).
- Awatramani, R., Soriano, P., Rodriguez, C., Mai, J. J. & Dymecki, S. M. Cryptic boundaries in roof plate and choroid plexus identified by intersectional gene activation. *Nat. Genet.* **35**, 70–75 (2003).
- Farago, A. F., Awatramani, R. B. & Dymecki, S. M. Assembly of the brainstem cochlear nuclear complex is revealed by intersectional and subtractive genetic fate maps. *Neuron* **50**, 205–218 (2006).
- Yamamoto, M. *et al.* A multifunctional reporter mouse line for Cre- and FLP-dependent lineage analysis. *Genesis* **47**, 107–114 (2009).
- Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
- Bonaguidi, M. A. *et al.* *In vivo* clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* **145**, 1142–1155 (2011).

56. Hippenmeyer, S. *et al.* Genetic mosaic dissection of *Lis1* and *Ndel1* in neuronal migration. *Neuron* **68**, 695–709 (2010).
57. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).  
**In this work, the authors present the Brainbow method for stochastic multicolour labelling in mice *in vivo*.**
58. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing *Lgr5* stem cells. *Cell* **143**, 134–144 (2010).
59. Sutherland, K. D. *et al.* Cell of origin of small cell lung cancer: inactivation of *Trp53* and *Rb1* in distinct cell types of adult mouse lung. *Cancer Cell* **19**, 754–764 (2011).
60. Beier, K. T., Samson, M. E. S., Matsuda, T. & Cepko, C. L. Conditional expression of the TVA receptor allows clonal analysis of descendants from Cre-expressing progenitor cells. *Dev. Biol.* **353**, 309–320 (2011).
61. Cornils, K. *et al.* Multiplexing clonality: combining RGB marking and genetic barcoding. *Nucleic Acids Res.* **42**, e56 (2014).
62. Loulier, K. *et al.* Multiplex cell and lineage tracking with combinatorial labels. *Neuron* **81**, 505–520 (2014).
63. Kimmerling, R. J. *et al.* A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat. Commun.* **7**, 10220 (2015).
64. McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).  
**This study represents a proof of principle for the use of genome editing to simultaneously mark large numbers of lineages in a developing organism, opening the door for the next generation of high-throughput lineage tracing methods.**
65. Junker, J. P. *et al.* Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars. Preprint at *bioRxiv* <http://dx.doi.org/10.1101/056499> (2016).
66. Kalhor, R., Mali, P. & Church, G. M. Rapidly evolving homing CRISPR barcodes. *Nat. Methods* <http://dx.doi.org/10.1038/nmeth.4108> (2016).
67. Frieda, K. L. *et al.* Synthetic recording and *in situ* readout of lineage information in single cells. *Nature* <http://dx.doi.org/10.1038/nature20777> (2016).
68. Schmidt, S. T., Zimmerman, S. M., Wang, J., Kim, S. K. & Quake, S. R. Cell lineage tracing using nuclease barcoding. Preprint at *arXiv* <https://arxiv.org/abs/1606.00786> (2016).
69. Salipante, S. J., Kas, A., McMonagle, E. & Horwitz, M. S. Phylogenetic analysis of developmental and postnatal mouse cell lineages. *Evol. Dev.* **12**, 84–94 (2010).
70. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
71. Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810 (2015).
72. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).
73. Ostertag, E. M. & Kazanietz, H. H. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
74. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903–910 (2005).
75. Erwin, J. A., Marchetto, M. C. & Gage, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat. Rev. Neurosci.* **15**, 497–506 (2014).
76. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131 (2009).
77. Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537 (2011).
78. Evrony, G. D. *et al.* Single-neuron sequencing analysis of 11 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483–496 (2012).
79. Upton, K. R. *et al.* Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**, 228–239 (2015).
80. Evrony, G. D., Lee, E., Park, P. J. & Walsh, C. A. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5**, e12966 (2016).
81. Evrony, G. D. *et al.* Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).  
**This study demonstrated that somatic L1 retrotransposition events could be used to identify distinct lineages in normal human brain.**
82. Abyzov, A. *et al.* Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438–442 (2012).
83. 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).
84. McConnell, M. J. *et al.* Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).
85. 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).
86. Baslan, T. *et al.* Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. *Genome Res.* **25**, 714–724 (2015).
87. Frank, S. A. Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer & neurodegeneration. *Proc. Natl Acad. Sci. USA* **107** (Suppl. 1), 1725–1730 (2010).
88. Hazen, J. L. *et al.* The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron* **89**, 1225–1236 (2016).
89. Zafar, H., Wang, Y., Nakhleh, L., Navin, N. & Chen, K. Monovar: single-nucleotide variant detection in single cells. *Nat. Methods* **13**, 505–507 (2016).
90. Ellegren, H. Microsatellites: simple sequences with complex evolution. *Nat. Rev. Genet.* **5**, 435–445 (2004).
91. Frumkin, D., Wasserstrom, A., Kaplan, S., Feige, U. & Shapiro, E. Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput. Biol.* **1**, e50 (2005).
92. Salipante, S. J. & Horwitz, M. S. Phylogenetic fate mapping. *Proc. Natl Acad. Sci. USA* **103**, 5448–5453 (2006).
93. Reizel, Y. *et al.* Colon stem cell and crypt dynamics exposed by cell lineage reconstruction. *PLoS Genet.* **7**, e1002192 (2011).
94. Reizel, Y. *et al.* Cell lineage analysis of the mammalian female germline. *PLoS Genet.* **8**, e1002477 (2012).
95. Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816–819 (1993).
96. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**, 558–561 (1993).
97. Turajlic, S., McGranahan, N. & Swanton, C. Inferring mutational timing and reconstructing tumour evolutionary histories. *Biochim. Biophys. Acta* **1855**, 264–275 (2015).
98. Ross, E. M. & Markowitz, F. OncoNEM: inferring tumor evolution from single-cell sequencing data. *Genome Biol.* **17**, 69 (2016).
99. Li, Y. *et al.* Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *Gigascience* **1**, 12 (2012).
100. Xu, X. *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148**, 886–895 (2012).
101. Shibata, D., Navidi, W., Salovaara, R., Li, Z. H. & Aaltonen, L. A. Somatic microsatellite mutations as molecular tumor clocks. *Nat. Med.* **2**, 676–681 (1996).
102. Naxerova, K. & Jain, R. K. Using tumour phylogenetics to identify the roots of metastasis in humans. *Nat. Rev. Clin. Oncol.* **12**, 258–272 (2015).
103. Frumkin, D. *et al.* Cell lineage analysis of a mouse tumor. *Cancer Res.* **68**, 5924–5931 (2008).  
**After previous work by the authors demonstrated the theoretical feasibility of using somatic microsatellite variability to reconstruct lineage, this study carried out the experiments, demonstrating that cells in a mouse tumour could be traced back to their somatic origin.**
104. Shlush, L. I. *et al.* Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability. *Blood* **120**, 603–612 (2012).
105. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
106. Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
107. Yachida, S. *et al.* Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
108. Hou, Y. *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* **148**, 873–885 (2012).
109. Gawad, C., Koh, W. & Quake, S. R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl Acad. Sci. USA* **111**, 17947–17952 (2014).
110. Wang, Y. & Navin, N. E. Advances and applications of single-cell sequencing technologies. *Mol. Cell* **58**, 598–609 (2015).
111. Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci. USA* **99**, 5261–5266 (2002).
112. Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* **150**, 402–412 (2012).
113. Fu, Y. *et al.* Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc. Natl Acad. Sci. USA* **112**, 11923–11928 (2015).
114. Sidore, A. M., Lan, F., Lim, S. W. & Abate, A. R. Enhanced sequencing coverage with digital droplet multiple displacement amplification. *Nucleic Acids Res.* **44**, e66 (2016).
115. Rhee, M., Light, Y. K., Meagher, R. J. & Singh, A. K. Digital Droplet Multiple Displacement Amplification (ddMDA) for whole genome sequencing of limited dna samples. *PLoS ONE* **11**, e0153699 (2016).
116. Szulwach, K. E. *et al.* Single-cell genetic analysis using automated microfluidics to resolve somatic mosaicism. *PLoS ONE* **10**, e0135007 (2015).
117. Barker, D. L. *et al.* Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Res.* **14**, 901–907 (2004).
118. Ning, L. *et al.* Quantitative assessment of single-cell whole genome amplification methods for detecting copy number variation using hippocampal neurons. *Sci. Rep.* **5**, 11415 (2015).
119. 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).
120. Voet, T. *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res.* **41**, 6119–6138 (2013).
121. Leung, M. L. *et al.* Highly multiplexed targeted DNA sequencing from single nuclei. *Nat. Protoc.* **11**, 214–235 (2016).
122. Chow, H.-M. & Herrup, K. Genomic integrity and the ageing brain. *Nat. Rev. Neurosci.* **16**, 672–684 (2015).
123. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisén, J. Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005).
124. Leung, M. L., Wang, Y., Waters, J. & Navin, N. E. SNES: single nucleus exome sequencing. *Genome Biol.* **16**, 55 (2015).
125. Knouse, K. A., Wu, J., Whittaker, C. A. & Amon, A. Single cell sequencing reveals low levels of aneuploidy across mammalian tissues. *Proc. Natl Acad. Sci. USA* **111**, 13409–13414 (2014).
126. Espina, V. *et al.* Laser-capture microdissection. *Nat. Protoc.* **1**, 586–603 (2006).
127. Gutierrez-Gonzalez, L. *et al.* Analysis of the clonal architecture of the human small intestinal epithelium establishes a common stem cell for all lineages and reveals a mechanism for the fixation and spread of mutations. *J. Pathol.* **217**, 489–496 (2009).
128. Crossetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
129. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. *In situ* detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010).

130. Larsson, C. *et al.* In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. *Nat. Methods* **1**, 227–232 (2004).
131. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
132. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
133. Lee, J. H. *et al.* Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
134. Janiszewska, M. *et al.* In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat. Genet.* **47**, 1212–1219 (2015).
135. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
136. Ware, M. L., Tavazoie, S. F., Reid, C. B. & Walsh, C. A. Coexistence of widespread clones and large radial clones in early embryonic ferret cortex. *Cereb. Cortex* **9**, 636–645 (1999).
137. Reid, C. B., Tavazoie, S. F. & Walsh, C. A. Clonal dispersion and evidence for asymmetric cell division in ferret cortex. *Development* **124**, 2441–2450 (1997).

#### Acknowledgements

The authors thank members of the Walsh laboratory, especially M. Lodato, for helpful comments. This work was supported by the Manton Center for Orphan Disease Research and grants from the National Institute of Neurological Disorders and Stroke (NINDS) (R01 NS032457, R01 NS079277 and U01 MH106883) to C.A.W. M.B.W. is supported by the Leonard and Isabelle Goldenson Research Fellowship. K.M.G. is supported by US National Institutes of Health (NIH) grant T32 MH20017. C.A.W. is a Distinguished Investigator of the Paul G. Allen Family Foundation and an Investigator of the Howard Hughes Medical Institute.

#### Competing interests statement

The authors declare no competing interests.