Methods for lineage tracing on the organism-wide level

Spanjaard, B., Junker, J.P.

This is the final version of the accepted manuscript. The original article has been published in final edited form in:

Current Opinion in Cell Biology
2017 DEC DD ; 49: 16-21
2017 NOV 22 (first published online: final publication)
doi: 10.1016/j.ceb.2017.11.004

Publisher: Elsevier

Methods for lineage tracing on the organism-wide level

Bastiaan Spanjaard[1], Jan Philipp Junker[1,2]

[1] Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, 13092 Berlin-Buch, Germany

[2] correspondence: janphilipp.junker@mdc-berlin.de

Abstract

Determining the lineage origin of cell types is a major goal in developmental biology. Furthermore, lineage tracing is a powerful approach for understanding the origin of developmental defects as well as the origin of diseases such as cancer. There is now a variety of complementary approaches for identifying lineage relationships, ranging from direct observation of cell divisions by light microscopy to genetic labeling of cells using inducible recombinases and fluorescent reporters. A recent development, and the main topic of this review article, is the use of high-throughput sequencing data for lineage analysis. This emerging approach holds the promise of increased multiplexing capacity, allowing lineage analysis of large cell numbers up to the organism-wide level combined with simultaneous transcription profiling by single cell RNA sequencing.

Main text

Introduction

The rapid emergence of spatial patterns, and of tissues and organs, out of a single fertilized egg has fascinated biologists for centuries. It is therefore not surprising that experiments that aim to identify the precursor cells of anatomical structures have been central to developmental biology ever since it became clear in the 19th century that all cells arise from pre-existing cells. Early lineage tracing studies based on continuous observation of cell divisions under the microscope have yielded remarkable fate maps, but direct observation of cell cleavages was mostly limited to invertebrate species and to early developmental stages [1]. To overcome this limitation and extend the accessible time window, researchers turned to experimental manipulation of embryos such as dye injections and grafting of cells with different pigmentation [2]. Since then, lineage tracing has become a diverse field that includes a number of very different approaches, such as live imaging by light sheet microscopy, genetic switches that provide optical readouts of specific lineage decisions, and sequencing of somatic mutations or DNA barcodes.

Lineage tracing is now a powerful method in different fields of biology. This includes developmental biology, where it is used to understand the molecular mechanisms underlying cell fate decisions and their misregulation in congenital disease, as well as studies of regeneration and tissue homeostasis, where the origin of newly created cells is often a central question. Furthermore, lineage tracing is important in various aspects of medicine. The most well-known example is cancer biology, where lineage tracing is used to find the cell type of origin for specific cancer types and to identify which cells contribute to growth and spread of tumors. Another very interesting example is wound healing after injury in cutaneous tissues, where the extent of scar formation has been linked to the developmental origin of fibroblasts [3].

A shared characteristic between all of these questions is that they focus on time as an experimental parameter by aiming to identify the hidden source of events such as formation of an organ in the embryo or generation of a tumor in the adult body. Since most experimental

techniques used in molecular biology yield snapshot data in which temporal information is lost, and since direct and continuous observation of cell divisions is neither feasible nor practical in most cases, approaches for recording and reading information about individual cell fate decisions have become important tools in biology. In this review article, we discuss and compare current lineage tracing approaches, with a particular emphasis on emerging techniques for massively parallel lineage tracing on an organism-wide level.

Optical lineage tracing

Since the earliest fate mapping experiments, the two main strategies for lineage tracing have been approaches using either direct continuous observation of cell divisions, or labeling of cells for later analysis (Figure 1). The pinnacle of non-invasive lineage tracing probably consists of the famous study of the *C. elegans* lineage tree by John Sulston and coworkers [4]. Using differential interference contrast microscopy, they observed and videotaped all cell cleavages from the one-cell stage to the adult worm. This study established that *C. elegans* has an almost invariant lineage. Sulston's experiments have had a profound influence on how we think about lineage and its connection to cell fate decisions, even though lineage trees in higher organisms are more variable than in *C. elegans*. In a way, all method development in lineage tracing over the last 30 years was aimed at achieving lineage trees of the same quality, i.e. a complete history of all cell fate decisions, in other organisms. Technological advances in optical microscopy, in particular light sheet microscopy, have in recent years enabled live fluorescence microscopy of *Drosophila* and zebrafish development [5,6] (Figure 2A). In combination with computational approaches for analyzing the movements and divisions of single cells, such data allows generation of a digital embryo with information about positions, sizes and fluorescence intensities of nuclei over time. However, while such experiments are extremely powerful for studying general properties of development such as patterns of cell movement, they are mostly limited to relatively early developmental stages in selected species.

An alternative strategy to deal with the complexity of vertebrate lineages is to apply optical labels to specific cells (reviewed in [7]). This approach is exemplified by Charles Kimmel's dye injection experiments in the zebrafish [8]. These studies showed that the position of cells at the beginning of gastrulation is the main determinant of their future cell type. Dye injection experiments allow for clonal analysis in wildtype organisms, but are restricted to relatively short time spans and provide only limited control about the number and identity of the cells that are labeled. These challenges were solved with the development of reporter transgenes such as β-galactosidase or green fluorescent protein (GFP), which enabled stable long-term labeling of cells and their progeny by e.g. retroviral infection [9,10]. Currently the most popular approach for lineage tracing by cell labeling is to use genetic recombination. Genetic recombination by means of the Cre/*lox* or Flp/*FRT* system provides exquisite control by enabling cell-type and stage specific labeling of cells [11]. The general strategy of this approach is to express Cre-recombinase under the control of a cell-type specific or inducible promoter. This leads to excision of a transcriptional roadblock (*loxP-STOP-loxP*) enabling specific expression of a reporter gene in the desired cells and their progeny (Figure 2B). The fact that cell labeling can be performed in a cell type dependent manner has turned Cre/*lox* lineage tracing into the preferred approach for adult tissue stem cells in e.g. the intestine or skin [12,13]. An important consideration when using optical markers for specific cells is that labeling has to be sufficiently sparse to distinguish individual clones. Furthermore, cross-sectional approaches combining Flp and Cre (or Dre and split-Cre) can greatly enhance the specificity of targeting. In many biological applications, for instance for identification of cancer cells of origin, it is important to combine lineage analysis with functional manipulation of cells. Mosaic analysis with double markers (MADM) achieves simultaneous cell labeling and gene knockout by Cre-mediated interchromosomal recombination resulting in expression of two fluorescent markers (e.g. GFP and RFP) that mark wildtype and mutant cells, respectively [14].

Using fluorophores with different spectral properties, as exemplified in *brainbow* [15,16], where cells randomly choose one out of several colors, increases the multiplexing capacity. In a landmark paper by the Clevers lab, the 4-color *Confetti* mouse was used to study patterns of stem cell self-renewal in tissue homeostasis in the intestine [17]. Short-term lineage tracing of sparsely labeled intestinal stem cells revealed their capacity for symmetric divisions expanding the stem cell population. Complementary long-term lineage tracing experiments in densely labeled animals showed that intestinal crypts undergo neutral drift towards clonality, with one clone eventually taking over the entire crypt. Other studies have used multicolor transgenes to study organ development [18] and tissue regeneration [19].

Sequence-based lineage tracing

A promising strategy for overcoming the limitations imposed by the small number of spectrally distinct genetically encoded fluorophores is to make use of the enormous multiplexing capacity of DNA for lineage tracing. Similar to optical lineage tracing, the different sequence-based approaches can be categorized into invasive and non-invasive methods (Figure 1).

In theory, naturally occurring somatic mutations (such as single nucleotide variants or copy number variations) are powerful lineage markers that can be read by sequencing (reviewed in [20]). In contrast to genetic recombination techniques, lineage tracing by somatic mutations is not limited to specific labeling times but allows reconstruction of full lineage trees (Figure 2C). Since lineage tracing by somatic mutations is non-invasive and does not require continuous observation, it is ideally suited for studying human samples. In an early study, the clonal origin of colonic adenomas was studied in an X0/XY mosaic individual with sequence probes for the Y chromosome [21]. Since then, many other studies have focused on analyzing clonality and lineage relationships in human tumors, where genomic aberrations are abundant [22-24]. In the last few years, pioneering studies have started to apply this strategy to early embryonic lineage decisions. In organoids derived from single mouse cells [25] and in human blood samples analyzed in bulk [26], analysis of somatic mutations allowed reconstruction of early embryonic lineage trees. In a recent landmark paper published by the Walsh lab, the authors placed neurons from post-mortem human brains in a developmental lineage tree after whole genome amplification and sequencing of single cells [27]. However, general applicability of this approach is currently hampered by the high cost of sequencing the whole genome of large numbers of single cells, and by the more fundamental challenge that the rate of somatic mutations may be too slow for lineage tracing in healthy tissues, in particular in those that have a high turnover such as skin or intestine. To some degree this obstacle can be overcome by targeted sequencing of mutation hotspots. Specifically, microsatellite repeats have been used successfully for lineage tracing [28,29]. As an alternative, a recent study used DNA hydroxymethylation for lineage tracing at the four-cell stage of mouse embryogenesis [30]. In theory, DNA methylation could be employed for lineage tracing beyond two cell divisions [31].

While the approaches discussed above are ideally suited for human samples, for model organisms, lineage tracing techniques that are based on experimental manipulation are typically the better choice due to the higher degree of control. Soon after the emergence of virus-induced optical cell labeling with β-galactosidase or GFP, researchers realized that introduction of DNA barcodes would allow a massive gain in multiplexing capacity (Figure 2D) [32,33]. The power of this approach increased with the emergence of high-throughput sequencing and has gained particular prominence for studying the hematopoietic system [34-36], where cells can be transfected *ex vivo* and then transplanted back into the animal. In the hematopoietic system, it is likely that clonal dynamics change upon transplantation of stem and progenitor cells. To study native unperturbed hematopoiesis, Fernando Camargo and co-workers developed a method that uses the integration sites of a DNA transposon as lineage markers [37] and observed that steady-state hematopoiesis is driven by a large number of multi-potent precursors, whereas earlier transplantation studies found dominant contributions by a smaller number of hematopoietic stem cells. Another recent method, termed *Polylox*

barcoding [38], avoids transplantation by using a synthetic recombination cassette consisting of ten *loxP* sites in alternating orientations, separated by unique spacer sequences. Upon activation of Cre recombinase, fragments get excised or inverted, which leads to generation of a high diversity of potential sequences. An intriguing future possibility is to modify the *Polylox* cassette in such a way that the full information about the sequence of editing events is encoded in the final pattern of excisions and inversions encodes, enabling reconstruction of complete lineage trees. Such synthetic recording systems have recently been successfully demonstrated in *E. coli* [39].

Novel experimental and computational approaches

The experimental lineage tracing approaches discussed in the previous paragraph have allowed a massive increase in labeling complexity, enabling analysis of a much larger number of clones per animal than would be possible with optical readouts. However, all of these methods are currently limited to labeling at a single time point, which has precluded reconstruction of full lineage trees. Current efforts that repurpose the CRISPR/Cas9 system for lineage tracing [40-43] aim to address this issue. The basic idea of CRISPR/Cas9 lineage tracing is to use Cas9 to create deletions or insertions in transgenic target sites in the genome and to use the resulting sequence modifications as lineage barcodes. The cellular DNA repair machinery has the capacity to create a high diversity of lineage barcodes by non-homologous end-joining, and the sequence complexity can be increased even further by using e.g. self-targeting sgRNAs [44,45].

There are different implementations of this approach (Figure 3A-C). In GESTALT [40], the authors generated a new transgenic zebrafish line with synthetic concatemerized target sites in the 3' UTR of a GFP transgene, while alternative approaches (scartrace or LINNAEUS) use existing lines with multiple integrations of a transgene [41]. Injection of Cas9 and sgRNA into the zebrafish zygote led to successive accumulation of scar sequences until gastrulation stages. The Quake lab recently used a similar strategy in *C. elegans* [43]. These approaches have so far only been published for bulk material, but several manuscripts that report simultaneous lineage tracing and cell type identification in single cells from dissociated zebrafish have recently become available as preprints [46,47]. The MEMOIR technique [42] is based on a similar construct as in GESTALT, but with multiple integrations per cell. Collapse of the target sites is read by single-molecule FISH instead of sequencing. MEMOIR can read out lineage markers and expression of selected genes on the single cell level while retaining full spatial information, although this has so far only been demonstrated in cell culture.

All sequence-based lineage tracing approaches rely heavily on computational methods to handle the high complexity of the data and to correct limited diversity or biased frequency distributions of barcodes. Reconstruction of lineage trees from sequence barcodes is an emerging field. Methods such as maximum parsimony [48] and neighbor-joining [49] have been used for GESTALT and MEMOIR, respectively, but it will certainly be possible to use other tree building methods such as maximum likelihood methods [50] or Bayesian inference [51] as well. For proper adaptation of these techniques, it will be important to find ways to implement experiment-specific factors such as differential scar probabilities in CRISPR/Cas9 lineage tracing and loss of information due to dropout events in single-cell sequencing [46].

Outlook

Simultaneous lineage tracing and cell type identification in thousands of single cells is now becoming a reality. Beyond this important milestone, the next important experimental challenge lies in ensuring that full lineage information, i.e. every cell division, is recorded reliably, which will require a larger capacity of the molecular recorder than currently available. Furthermore, it will be informative to complement lineage tracing with molecular recording of

cellular exposure to specific signaling factors that influence cell fate decisions. It has become clear in recent years that transposons, Cre/*lox* and Cas9 can be successfully used for recording lineage information in the genome, and it will be interesting to see if other and potentially more powerful biological systems such as the CRISPR spacer acquisition system of Cas1 and Cas2 [52,53] can also be successfully repurposed for lineage tracing in multicellular organisms.

References

1. Conklin EG: **The organization and cell lineage of the ascidian egg**. *Journal of the Academy of Natural Sciences of Philadelphia* 1905, **13**:1–119.

2. Spemann H, Mangold H: **über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren**. *Dev Genes Evol* 1924, **100**:599–638.

3. Rinkevich Y, Walmsley GG, Hu MS, Maan ZN, Newman AM, Drukker M, Januszyk M, Krampitz GW, Gurtner GC, Lorenz HP, et al.: **Identification and isolation of a dermal lineage with intrinsic fibrogenic potential**. *Science* 2015, **348**:aaa2151–aaa2151.

4. Sulston JE, Schierenberg E, White JG, Thomson JN: **The embryonic cell lineage of the nematode Caenorhabditis elegans.** *Dev Biol* 1983, **100**:64–119.

5. Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EHK: **Reconstruction of Zebrafish Early Embryonic Development by Scanned Light Sheet Microscopy**. *Science* 2008, **322**:1065–1069.

6. Amat F, Lemon W, Mossing DP, McDole K, Wan Y, Branson K, Myers EW, Keller PJ: **Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data**. *Nat Methods* 2014, **11**:951–958.

7. Kretzschmar K, Watt FM: **Lineage Tracing**. *Cell* 2012, **148**:33–45.

8. Kimmel CB, Warga RM, Schilling TF: **Origin and organization of the zebrafish fate map.** *Development* 1990, **108**:581–594.

9. Turner DL, Cepko CL: **A common progenitor for neurons and glia persists in rat retina late in development.** *Nature* 1987, **328**:131–136.

10. Doetsch F, Caille I, Lim DA, Garcia-Verdugo JM, Alvarez-Buylla A: **Subventricular zone astrocytes are neural stem cells in the adult mammalian brain**. *Cell* 1999, **97**:703–716.

11. Harrison DA, Perrimon N: **Simple and efficient generation of marked clones in Drosophila**. *Current Biology* 1993, **3**:424–433.

12. Barker N, van Es J, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters P: **Identification of stem cells in small intestine and colon by marker gene Lgr5**. *Nature* 2007, **449**:1003–1007.

13. Nowak JA, Polak L, Pasolli HA, Fuchs E: **Hair Follicle Stem Cells Are Specified and Function in Early Skin Morphogenesis**. *Cell Stem Cell* 2008, **3**:33–43.

14. Liu C, Sage JC, Miller MR, Verhaak RGW, Hippenmeyer S, Vogel H, Foreman O,

Bronson RT, Nishiyama A, Luo L, et al.: **Mosaic Analysis with Double Markers Reveals Tumor Cell of Origin in Glioma**. *Cell* 2011, **146**:209–221.

15. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW: **Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system**. *Nature* 2007, **450**:56–62.

16. Roy E, Neufeld Z, Livet J, Khosrotehrani K: **Concise Review: Understanding Clonal Dynamics in Homeostasis and Injury Through Multicolor Lineage Tracing**. *Stem Cells* 2014, **32**:3046–3054.

17. Snippert HJ, Van Der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD, et al.: **Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells.** *Cell* 2010, **143**:134–144.

18. Gupta V, Poss KD: **Clonally dominant cardiomyocytes direct heart morphogenesis**. *Nature* 2012, **484**:479–484.

19. Gurevich DB, Nguyen PD, Siegel AL, Ehrlich OV, Sonntag C, Phan JMN, Berger S, Ratnayake D, Hersey L, Berger J, et al.: **Asymmetric division of clonal muscle stem cells coordinates muscle regeneration in vivo.** *Science* 2016, **353**:aad9969.

20. Woodworth MB, Girskis KM, Walsh CA: **Building a lineage from single cells: genetic techniques for cell lineage tracking**. *Nature Reviews Genetics* 2017, **18**:230–244.

21. Novelli MR, Williamson JA, Tomlinson I, Elia G, Hodgson SV, Talbot IC, Bodmer WF, Wright NA: **Polyclonal origin of colonic adenomas in an XO/XY patient with FAP**. *Science* 1996, **272**:1187–1190.

22. McDonald SAC, Greaves LC, Gutierrez Gonzalez L, Rodriguez Justo M, Deheragoda M, Leedham SJ, Taylor RW, Lee CY, Preston SL, Lovell M, et al.: **Mechanisms of Field Cancerization in the Human Stomach: The Expansion and Spread of Mutated Gastric Stem Cells**. *Gastroenterology* 2008, **134**:500–510.

23. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al.: **Tumour evolution inferred by single-cell sequencing**. *Nature* 2012, **472**:90–94.

24. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, Li F, Tsang S, Wu K, Wu H, et al.: **Single-Cell Exome Sequencing Reveals Single-Nucleotide Mutation Characteristics of a Kidney Tumor**. *Cell* 2012, **148**:886–895.

25. Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G, et al.: **Genome sequencing of normal cells reveals developmental lineages and mutational processes.** *Nature* 2014, **513**:422–425.

* 26. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, Wedge DC, Davies HR, Ramakrishna M, Fullam A, et al.: **Somatic mutations reveal asymmetric cellular dynamics in the early human embryo.** *Nature* 2017, **543**:714–718.

Reconstruction of early human developmental lineage tree based on bulk analysis of human blood samples taken from breast cancer patients.

* 27. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Lee S, Chittenden TW, D'Gama AM, Cai X, et al.: **Somatic mutation in single human neurons tracks developmental and transcriptional history**. *Science* 2015, **350**:94–98.

To reconstruct a developmental lineage tree, the authors isolated single neurons from post-mortem human brains and identified somatic mutations by whole genome amplification and sequencing of single cells.

28. Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E: **Genomic Variability within an Organism Exposes Its Cell Lineage Tree**. *PLoS Comput. Biol.* 2005, **1**:e50–13.

29. Shlush LI, Chapal-Ilani N, Adar R, Pery N, Maruvka Y, Spiro A, Shouval R, Rowe JM, Tzukerman M, Bercovich D, et al.: **Cell lineage analysis of acute leukemia relapse uncovers the role of replication-rate heterogeneity and microsatellite instability**. *Blood* 2012, **120**:603–612.

30. Mooijman D, Dey SS, Boisset J-C, Crosetto N, van Oudenaarden A: **Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction**. *Nature Biotechnology* 2016, **34**:871–875.

31. Kelsey G, Stegle O, Reik W: **Single-cell epigenomics: Recording the past and predicting the future.** *Science* 2017, **358**:69–75.

32. Lemischka IR, Raulet DH, Mulligan RC: **Developmental Potential and Dynamic Behavior of Hematopoietic Stem-Cells**. *Cell* 1986, **45**:917–927.

33. Walsh C, Cepko CL: **Widespread Dispersion of Neuronal Clones Across Functional Regions of the Cerebral-Cortex**. *Science* 1992, **255**:434–440.

34. Lu R, Neff NF, Quake SR, Weissman IL: **Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding.** *Nature Biotechnology* 2011, **29**:928–933.

35. Naik SH, Perié L, Swart E, Gerlach C, van Rooij N, de Boer RJ, Schumacher TN: **Diverse and heritable lineage imprinting of early haematopoietic progenitors**. *Nature* 2013, **496**:229–232.

36. Lin D, Kan A, Gao J, Crampin E, Hodgkin PD, Naik SH: **Assessment of clonal kinetics reveals multiple trajectories of dendritic cell development**. *bioRxiv* 2017, doi:10.1101/167635.

37. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho Y-J, Klein A, Hofmann O, Camargo FD: **Clonal dynamics of native haematopoiesis**. *Nature* 2014, **514**:322–327.

* 38. Pei W, Feyerabend TB, Rössler J, Wang X, Postrach D, Busch K, Rode I, Klapproth K, Dietlein N, Quedenau C, et al.: **Polylox barcoding reveals haematopoietic stem cell fates realized in vivo.** *Nature* 2017, **548**:456–460.

The authors generated a transgenic mouse with a cassette of loxP sites. Upon activation of Cre, excision and inversion of the spacer sequences produces a large diversity of barcodes that can be used for clonal analysis.

39. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK: **Synthetic recombinase-based state machines in living cells.** *Science* 2016, **353**:aad8559–aad8559.

* 40. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J: **Whole-organism lineage tracing by combinatorial and cumulative genome editing.** *Science* 2016, **353**:aaf7907.

Introduces GESTALT method: Massively parallel lineage tracing in zebrafish and cell culture based on CRISPR/Cas9 editing of synthetic target sites. Lineage markers are read out by bulk sequencing.

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

* 42. Frieda KL, Linton JM, Hormoz S, Choi J, Chow K-HK, Singer ZS, Budde MW, Elowitz MB, Cai L: **Synthetic recording and in situ readout of lineage information in single cells.** *Nature* 2017, **541**:107–111.

Introduces MEMOIR: Massively parallel lineage tracing in cultured cells based on CRISPR/Cas9 editing of synthetic target sites. Optical readout in single cells by sequential single-molecule FISH.

43. Schmidt ST, Zimmerman SM, Wang J, Kim SK, Quake SR: **Quantitative Analysis of Synthetic Cell Lineage Tracing Using Nuclease Barcoding.** *ACS Synth Biol* 2017, **6**:936–942.

44. Kalhor R, Mali P, Church GM: **Rapidly evolving homing CRISPR barcodes.** *Nat Methods* 2017, **14**:195–200.

45. Perli SD, Cui CH, Lu TK: **Continuous genetic recording with self-targeting CRISPR-Cas in human cells.** *Science* 2016, **353**.

46. Spanjaard B, Hu B, Mitic N, Junker JP: **Massively parallel single cell lineage tracing using CRISPR/Cas9 induced genetic scars**. *bioRxiv* 2017, doi:10.1101/205971.

47. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier AF: **Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain by scGESTALT**. *bioRxiv* 2017, doi:10.1101/205534.

48. Camin JH, Sokal RR: **A Method for Deducing Branching Sequences in Phylogeny**. *Evolution* 1965, **19**:311–326.

49. Saitou N, Nei M: **The Neighbor-Joining Method - a New Method for Reconstructing Phylogenetic Trees**. *Mol. Biol. Evol.* 1987, **4**:406–425.

50. Felsenstein J: **Evolutionary Trees From Dna-Sequences - a Maximum-Likelihood Approach**. *J. Mol. Evol.* 1981, **17**:368–376.

51. Rannala B, Yang ZH: **Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference**. *J. Mol. Evol.* 1996, **43**:304–311.

52. Shipman SL, Nivala J, Macklis JD, Church GM: **Molecular recordings by directed CRISPR spacer acquisition.** *Science* 2016, **353**:aaf1175.

53. Shipman SL, Nivala J, Macklis JD, Church GM: **CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria.** *Nature* 2017, **547**:345–349.

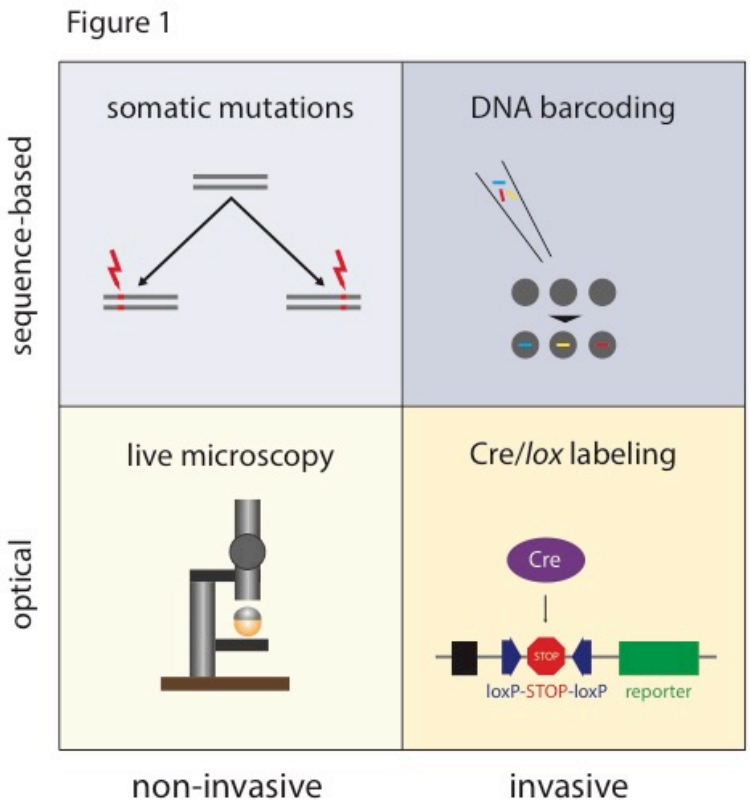

Figure 1. Classes of lineage tracing techniques. Methods for lineage tracing can be categorized along two axes: Intrinsic versus extrinsic lineage markers (x axis), and optical versus sequence-based lineage readout (y axis). Cartoons show methods that are representative for the four classes of techniques.

Figure 2

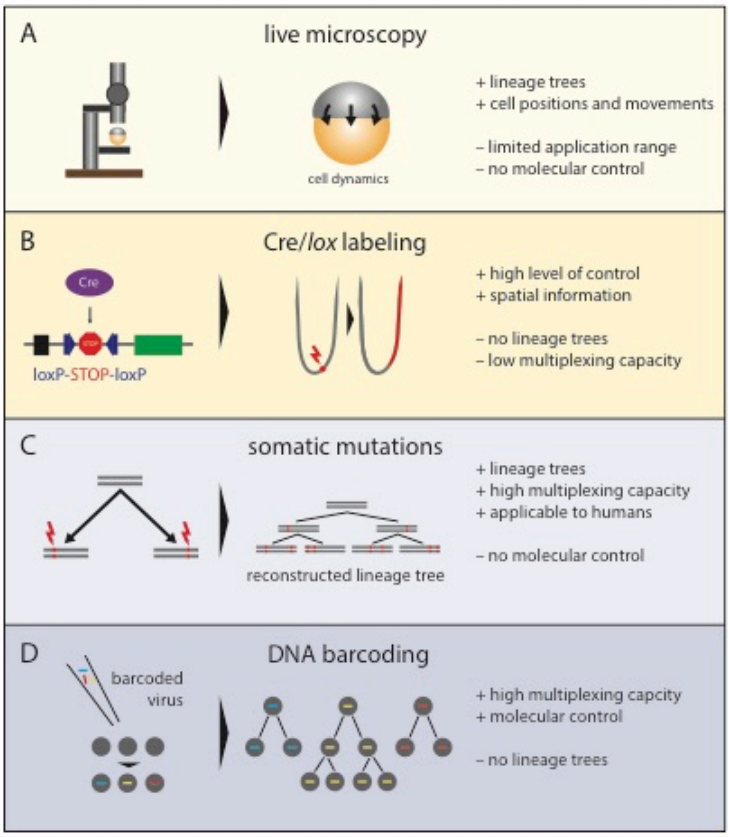

Figure 2. Possible results of lineage tracing experiments. Representative methods for non-invasive/invasive labeling and microscopy-/sequencing-based detection **A.** Live microscopy yields information about positions and movements of cells in addition to lineage trees. **B.** Genetic labeling of cells with the Cre/*lox* system provides a high level of control and is ideally suited for targeted questions in adult stem cell systems. **C.** Somatic mutations allow reconstruction of lineage trees in humans. **D.** Integration of viral barcodes allows unique labeling of large numbers of cells in a controlled fashion.

Figure 3

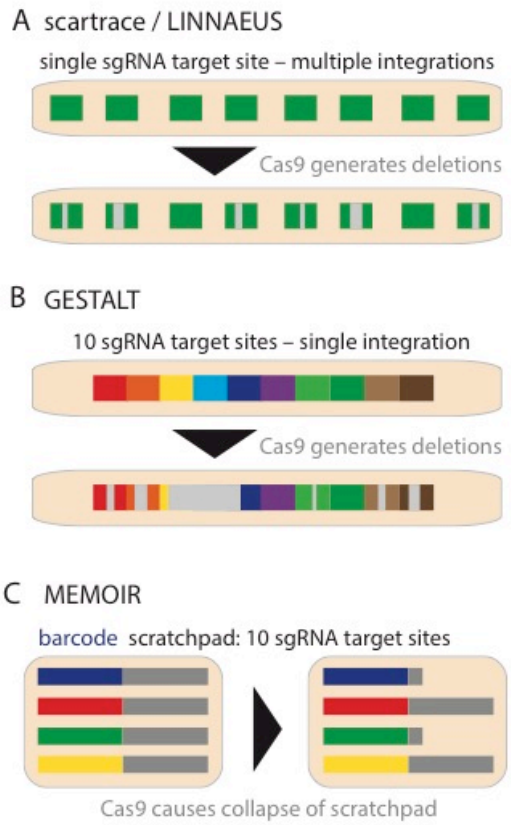

Figure 3. Massively parallel lineage tracing using the CRISPR/Cas9 system. A. In scartrace/LINNAEUS, an existing fish line with multiple integrations of a transgene is targeted by Cas9. The sequences of the resulting "genetic scars" (light gray) are used as lineage markers. **B.** GESTALT uses the same principle, but a new line with concatemerized target sites is used. Compared to scartrace/LINNAEUS, this has the advantage that the individual sites (different colors) can be distinguished. Furthermore, even in bulk data the information which scars were in the same cells is preserved. A downside is that some target sites get lost due to excision events (here, the light blue target site), reducing the available lineage information. **C.** MEMOIR uses multiple integrations of concatemerized target sites ("scratchpad", gray) that can be distinguished via barcodes (colored bars). Collapse of the scratchpad is visualized by smFISH, preserving spatial information. However, lineage information capacity is lower than in scartrace/LINNAEUS or GESTALT since the exact sequence of the collapsed scratchpad is not detectable by microscopy.