Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

DNA barcodes evolve for high-resolution cell lineage tracing

Nanami Masuyama[1,2,3,6], Hideto Mori[1,2,3,6] and Nozomu Yachie[1,2,3,4,5]

Mammalian development involves continuous dynamic processes in which cells propagate, differentiate, orchestrate, and decease to produce high-order functions. Although accurate cell lineage information can provide a strong foundation to understand such complex processes, the cell lineages involved in development of the whole mammalian body remain largely unclear, except for in early embryogenesis, which is observable under a microscope. With CRISPR genome editing, the concept of 'evolving DNA barcodes' has rapidly emerged for large-scale, high-resolution cell lineage tracing, where cell-embedded DNA barcodes continuously accumulate random mutations that are inherited from mother to daughter cells. Similar to evolutionary tree reconstruction using species' DNA sequences, cell lineages can be reconstructed using shared mutation patterns in the DNA barcodes identified using massively parallel sequencing. The dramatic developments of single-cell and imaging technologies have enabled analyses of the molecular and spatial architecture of heterogeneous cells. The evolving DNA barcodes can also consolidate this information on a reconstructed cell lineage tree and accelerate our understanding of multicellular organisms.

Addresses
[1] Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 153-8904, Japan
[2] Institute for Advanced Biosciences, Keio University, Tsuruoka 997-0035, Japan
[3] Graduate School of Media and Governance, Keio University, Fujisawa 252-0882, Japan
[4] Department of Biological Sciences, School of Science, The University of Tokyo, Tokyo 113-0033, Japan
[5] PRESTO, Japan Science and Technology Agency (JST), Tokyo 153-8904, Japan

Corresponding author:
Yachie, Nozomu (yachie@synbiol.rcast.u-tokyo.ac.jp)
[6] These authors contributed equally to this work.

Current Opinion in Chemical Biology 2019, **52**:63–71

This review comes from a themed issue on **Synthetic biology**

Edited by **Hirohide Saito** and **Yohei Yokobayashi**

For a complete overview see the Issue and the Editorial

Available online 15th June 2019

https://doi.org/10.1016/j.cbpa.2019.05.014

Introduction

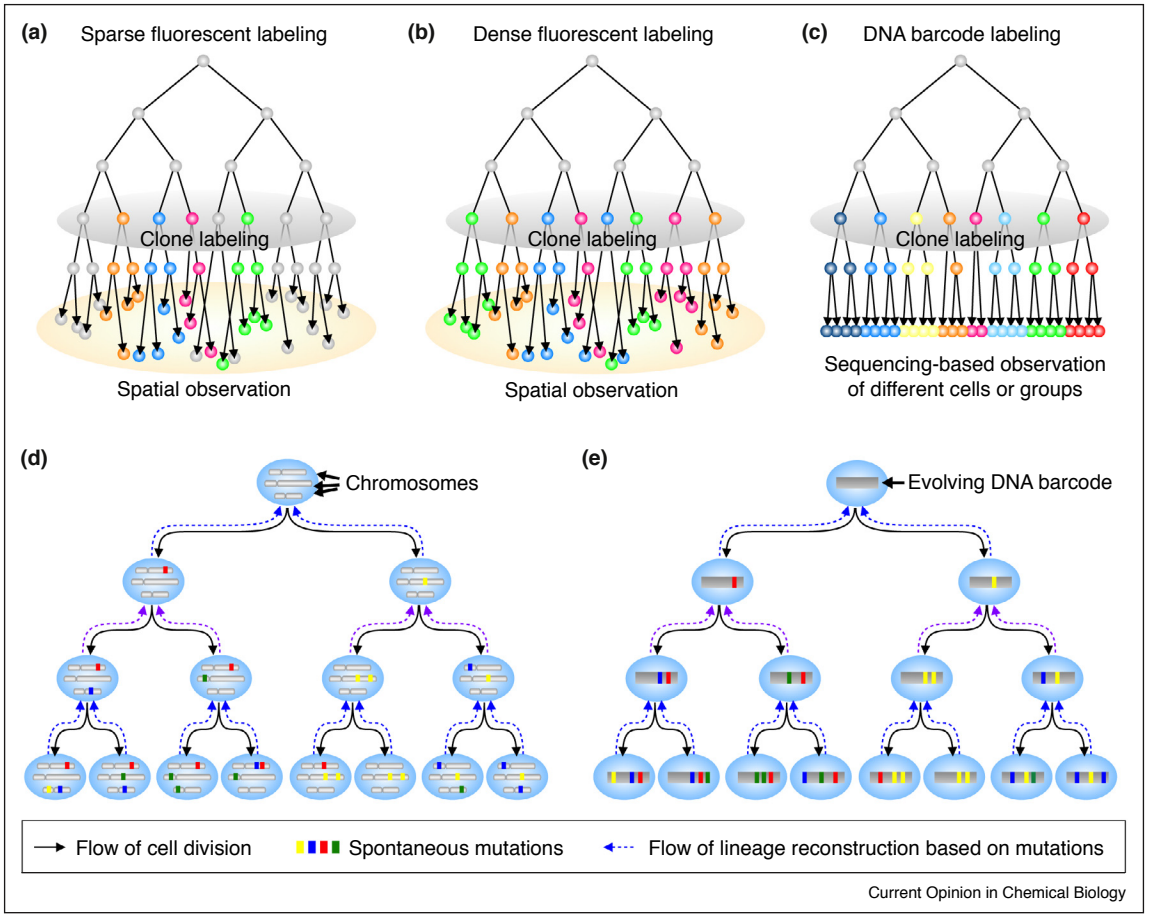
The development of animals and plants starts from a single fertilized cell and involves the formation of different organs and the whole body through cell division. In this process, cells with distinct molecular phenotypes and functions dynamically emerge, interact, and cooperate to organize functional units that also crosstalk at higher levels. Complete information on the whole cell lineage of a given organism would provide a strong basis for surveying its development. In the early 1980s, mapping of the entire cell lineage of the nematode *Caenorhabditis elegans* from the fertilized egg through to the hermaphroditic adult, for exactly 959 cells, was completed based on microscopic observations [1–3]. While this has long provided the foundation for study of the developmental biology of *C. elegans*, the developmental cell lineages of mammalian and other organisms have remained largely elusive.

Tracing of clonal cell distributions using static markers

Lineage tracing by live imaging enables the analysis of developmental dynamics of only systems that can be observed under a microscope. Therefore, reverse (or retrospective) approaches have become the major means of studying cell lineages of more complex systems, in which heritable markers are introduced into cells at an early stage and their distributions at a later stage are analyzed after system progression. For example, Brainbow enabled the tracing of cell clones in the course of development of higher organisms, such as mice [4] and *Drosophila* [5], by labeling cells in an initial embryonic population with the expression of a random combination of fluorescent proteins induced via stochastic DNA recombination. This fluorescent labeling method enables the 'painting' of cells with different colors and analysis of how they are distributed at a later stage with histological information obtained by fluorescent microscope observations. However, definitive conclusions about the clonal relationship of two cells having the same fluorescent expression pattern cannot always be drawn when the number of different fluorescent patterns exceeds or is close to the number of initially labeled cells (Figure 1a and b).

The concept of a complex DNA barcode library has been adopted to expand the resolution of this approach (Figure 1c) [6,7•,8–11]. For example, by lentivirus

Figure 1

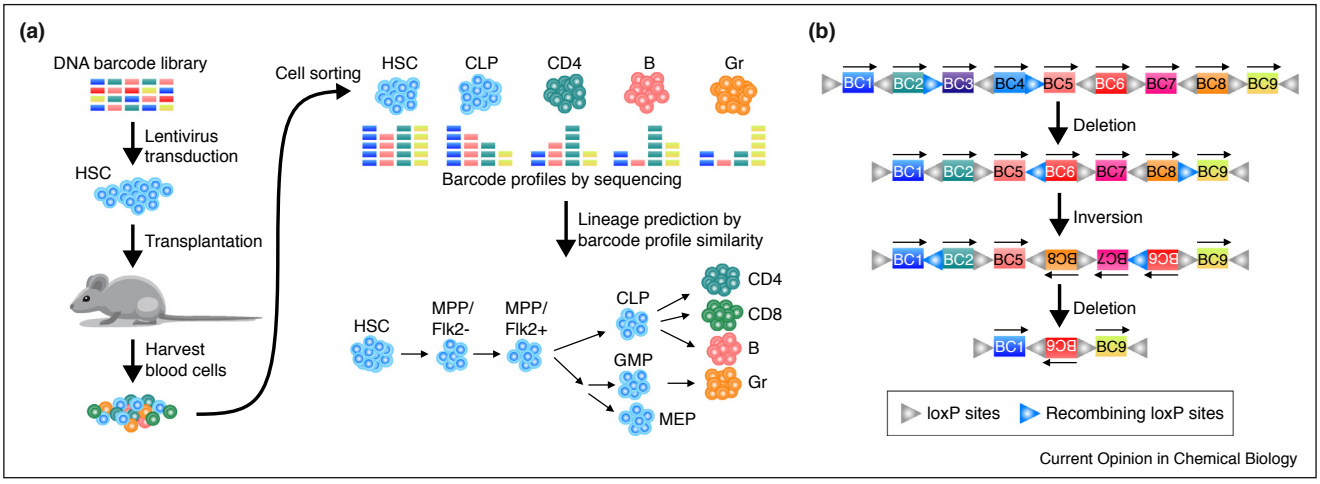


Various cell labeling methods for developmental lineage analyses. **(a, b)** Sparse fluorescent labeling (a) and dense fluorescent labeling (b). Cells in an initial population are labeled with a limited set of fluorescent markers. After cell progression, the spatial distribution of the labeled clones can be observed by imaging analysis. (a) Sparse fluorescent labeling traces distinct clones with low clone coverage. (b) Dense fluorescent labeling traces clones with high coverage but with ambiguous distinction of different clones. **(c)** DNA barcode-based labeling. This method enables high-coverage and distinct labeling of clones but the spatial information is lost in the pyrosequencing-based readout. **(d)** Lineage tree reconstruction using spontaneous somatic mutations. **(e)** Lineage tree reconstruction using an evolving DNA barcode.

transduction of random DNA barcodes, different clones in a cancer cell population can be uniquely labeled with different DNA barcodes [6]. Massively parallel sequencing of DNA barcodes obtained from time-course samples of the population enables *en masse* quantification of clone size dynamics in various conditions, such as in the presence of anticancer compounds. Other studies have also used a complex DNA barcode library for hematopoietic lineage tracing, where hematopoietic stem cells (HSCs) are harvested from mice, uniquely barcoded using lentivirus, and transplanted back into another mouse (Figure 2a) [7•,8]. After the *in vivo* differentiation of the barcoded HSCs, blood cells are harvested from the bone marrow and fractionated into different cell types according to their immunophenotypes using flow cytometry cell sorting. DNA barcodes of each fractionated population are then analyzed by massively parallel sequencing, and finally the genealogical relationships of different cell type groups are reconstructed according to the similarity between the DNA barcode distribution profiles.

However, this approach can potentially interfere with the native hematopoietic process, via lentiviral infection and HSC transplantation. Several *in vivo* cell barcoding methods have been developed to avoid this problem in hematopoietic lineage analysis [12•,13]. Using a concept similar to that applied in Brainbow, a recent study established a mouse line harboring a Polylox barcode, a chromosome-encoded array of DNA barcodes each sandwiched by *lox* sites orientated in the opposite directions (Figure 2b) [12•]. Inducing the expression of site-specific Cre recombinase in an early embryo of its offspring, cells including HSC progenitors were uniquely labeled by

Figure 2



Tracing hematopoiesis by a complex library of static DNA barcodes. **(a)** Lentiviral barcoding of cells. HSCs are harvested from mice and uniquely labeled by a highly complex lentiviral library of DNA barcodes. The barcoded cells are transplanted into a mouse. After differentiation, cells are harvested from the bone marrow and sorted into different cell types. The differentiation lineage can be inferred according to the similarities between the DNA barcode profiles of the fractionated cell groups. **(b)** An example of an *in vivo* Polylox barcoding pathway to randomly assign a unique barcode to a clone.

random drop-offs and inversions in the DNA barcode cassettes. Similar to the lentiviral barcoding approach, the pattern profiles of Polylox barcodes in different groups of differentiated cells successfully reconstructed the mosaic structure of HSC developmental patterns, where some major HSC clones gave rise to multilineage fates, such as the canonical tree-like view of HSC differentiation, but some other major clones committed to oligolineage fates.

Cell lineage tree reconstruction using somatic mutations

While the above-mentioned approaches capture how different markers that are introduced into cells of an initial population are distributed and expanded at the time of observation, these methods do not directly capture the structures of branching cell lineage trees. Chromosomes of dividing cells are replicated and segregated from mother cells to daughter cells with spontaneous mutations, each of which can be treated as a new label of cells discriminating their downstream branches of cell division from the others. Therefore, similar to phylogenetic tree reconstruction using the genomic sequences of diverged species, the reconstruction of cell lineages of daughter cells using their somatic mutation patterns was proposed (Figure 1d). Examples of this include inferences about abstract cell lineages using indels and single-nucleotide polymorphisms (SNPs) [14,15], mobile retrotransposon elements [16], microsatellites [17–20], and copy number variations (CNVs) [21,22].

In such studies, the resolutions of the reconstructed cell lineages are highly limited, mainly due to the fact that spontaneous mutations of individual cells rarely occur across the genome. Some studies have used single-cell genome sequencing to reconstruct cell lineages directly from daughter cells [15,17,18,22]. However, the sensitivity and coverage of single-cell genome sequencing technologies have been limited to thoroughly surveying *de novo* mutations and differences in chromosomal structures. Sequencing errors by these approaches could also lead to erroneous lineage reconstruction. Other studies have employed whole-genome sequencing of local tissue samples or distinct cell type groups of different immunophenotypes to reconstruct their lineages [16,19–21], assuming that such cell type groups are derived from similar cell lineage clades. This approach allows robust reconstruction of cell lineages from high-coverage genomic sequences for the upper lineage hierarchy above the population averages of the defined cell groups, but an invalid assumption about the lineage-restricted cell type groups might distort the results. Notably, another type of study successfully reconstructed an accurate cell lineage of mouse somatic cells using organoid technology, where organoids were established from the stomach, intestines, and prostate of mice to obtain abundant cells for high-coverage genome sequencing of their parental single cells [14]; however, in terms of its practical application, this method was limited to specific types of cell from which organoids can be established. In any of the genome sequencing-based lineage tracing approaches, the scalability in number of single-cell or local cell type group samples has been highly limited by the cost and throughput of current sequencing methods.

Evolving DNA barcodes

The DNA replication error rate of $\sim 10^{-8}$ and the DNA error correction efficiency of ~99% suggest that one new somatic mutation occurs in a mammalian genome of billions of base pairs in approximately every dozen cell divisions. Thus, a somatic 'mutatome' in the trillions of single cells comprising the mammalian body could nearly completely reconstruct the whole cell lineage of the body development. However, this concept is ahead of its time, as there is currently no technology that can efficiently and economically identify combinatorial mutation information spread across the genome of every cell. In contrast, CRISPR–Cas9 has emerged as a versatile genome editing tool that can induce mutations in target DNA regions of various organisms through sequence complementarity of guide RNA (gRNA); since its establishment, there have been rapid, turbulent implementations of the concept of 'evolving DNA barcodes' for lineage tracing (Figure 1e) [23••,24•,25,26••,27•,28–30,31••]. In this concept, a cell-embedded synthetic DNA barcode(s) continuously acquires new mutations and replicates along with cell divisions. Unlike somatic mutations in the genome, all of the lineage branching markers in the end-point daughter cells can be efficiently identified by sequencing of the local DNA barcode regions.

In 2016, GESTALT (genome editing of synthetic target array for lineage tracing) was the first reported approach among many developed for cell lineage tracing using the evolving DNA barcode [23••]. In this study, a transgenic zebrafish harboring an array of ten different DNA barcodes was first established. For lineage tracing, a one-cell embryo of this barcoded zebrafish was obtained and injected with Cas9 ribonucleoproteins (RNPs) containing gRNAs targeting the ten corresponding barcodes. After development, different organs were dissected, from which cells were fractionated into different cell type groups; their DNA barcodes were recovered by PCR amplification and identified by massively parallel sequencing. In the clades and subclades of the cell lineage reconstructed from the unified barcode sequencing results, there were increasing levels of lineage-specific restriction of specific cell type groups from the upper hierarchical levels of the lineage to downstream, indicating the validity of this approach.

Although GESTALT requires three different components—Cas9, gRNAs, and a DNA barcode array—a more compact system has been proposed. The original CRISPR–Cas9 genome editing requires the protospacer adjacent motif (PAM) sequence (5′-NGG-3′ for SpCas9 commonly used in these studies) next to the 3′ end of the target sequence, which discriminates the gRNA-encoding DNA region and its target having the same sequence. In contrast, the concept of engineered homing gRNA (hgRNA) [24•] or self-targeting gRNA (stgRNA) [25] has been proposed, in which the mutant gRNA encodes a PAM sequence adjacent to the targeting sequence region and iteratively mutates its own coding sequence. This gRNA mutant has also been used to trace cell lineages, but its resolution is limited because the iterative editing quickly terminates when its functional coding region including the intrinsic PAM is disrupted by mutations after a few rounds.

Similar to gRNA, the editing speed of hgRNA/stgRNA depends on its targeting sequence [24•] and is controllable by adjusting the Cas9 dose [25]. Furthermore, the number of self-editing cycles can be increased to a certain degree by adding an extra filler sequence upstream of the hgRNA/stgRNA-coding region, so that it can serve as a new targeting sequence when the original targeting sequence is deleted by the self-targeting [24•,25,26••]. Using combinatorial information of many different hgRNAs having various editing speeds, a proof-of-concept of mammalian developmental lineage tracing was recently demonstrated [26••]. In this study, a founder mouse MARC1 (mouse for actively recording cells) having 60 chromosome-encoded hgRNAs was derived from a population of engineered embryonic stem (ES) cells each having high copy number hgRNA transgenes introduced by PiggyBac transposon technology. Upon crossing a MARC1 mouse with a Cas9-expressing mouse, the development of their E12.5 conceptus was examined by analyzing mutation patterns of hgRNAs in different sections of the placenta and yolk sac, as well as the heart and a limb bud of the embryo. The developmental lineage for different brain sections was also reconstructed by this approach, showing that the anterior–posterior axis is established before the commitment to the lateral axis. Although this was the first demonstration of whole-body-level mammalian lineage tracing with the evolving DNA barcodes, this method had a similar issue to the somatic mutation-based lineage tracing, where mutations accumulated in hgRNAs of distant chromosomal loci in which combinatorial information is difficult to obtain from single end-point daughter cells.

Scaling up lineage tracing using single-cell technologies

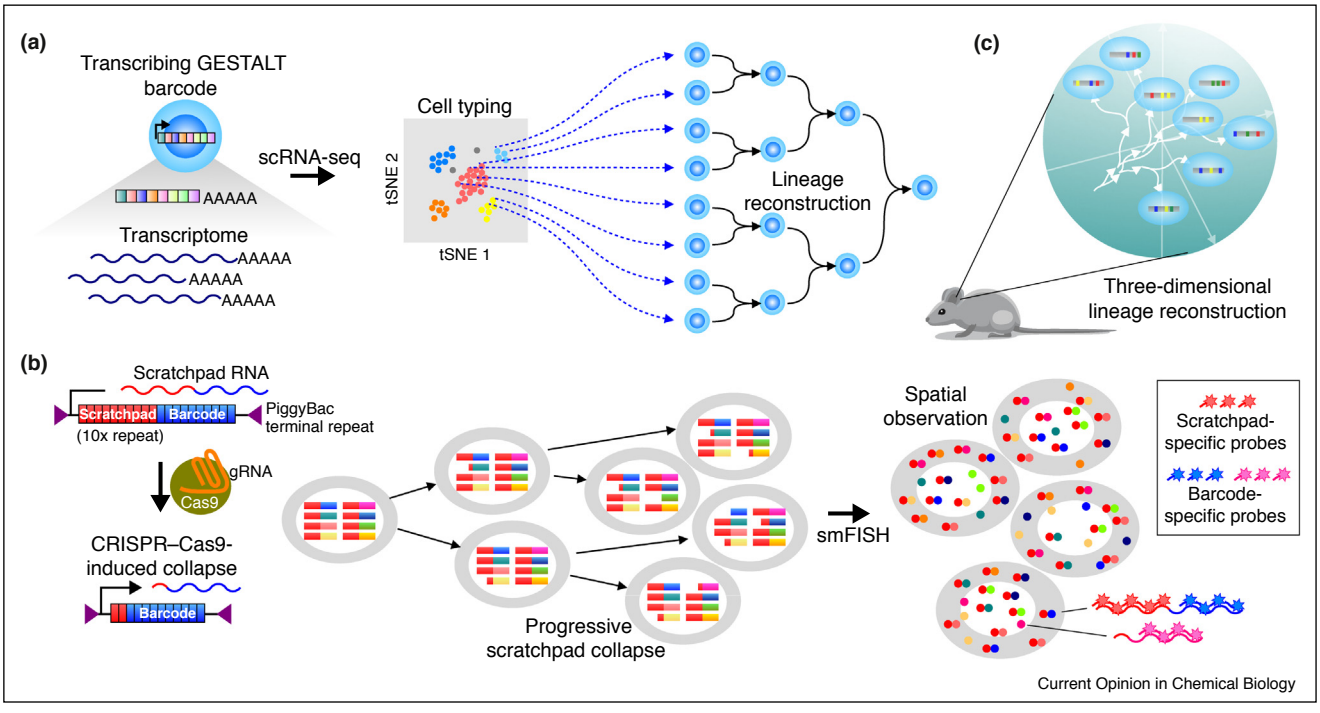
The evolving DNA barcodes only record information for topological structures of lineage trees. Therefore, mapping of functional or histological annotations of end-point daughter cells or population groups to the reconstructed lineage is necessary to address their biology. The rapid progress in single-cell technologies has enabled high-throughput molecular profiling of cells and revealed that heterogeneous biological systems are composed of many but distinct groups of unknown cell types, presenting a new challenge regarding how they emerge and orchestrate functions and structures within a system [32]. The throughputs of single-cell transcriptome (single-cell RNA sequencing or scRNA-seq) methods have also been dramatically enhanced by the molecular DNA barcode concept. For example, in the water-in-oil droplet-based

scRNA-seq technologies [33,34], single cells are encapsulated by droplets with a single unique bead particle containing oligo(dT) reverse-transcription primers concatenated to bead-specific DNA barcodes, which eventually serve as 'cell-specific' barcodes. The transcriptome of each cell is reverse-transcribed and labeled by the cell-specific barcodes in a droplet. After breaking up the droplets, the barcoded cDNAs are pooled for sequencing, and the sequencing reads are demultiplexed according to the cell-specific barcodes.

The evolving DNA barcode concept has been integrated into this scRNA-seq approach. In this concept, the evolving DNA barcodes are transcribed as polyadenylated RNA (evolving RNA barcodes), so their sequence or induced mutation patterns can be identified with transcriptome profiles at single-cell resolution. scGESTALT was one such example that achieved simultaneous single-cell profiling of cell types and their cell lineages for zebrafish brain development (Figure 3a) [27•]. Using scGESTALT, more than 100 distinct cell types with unique gene expression patterns were identified along with their cell lineages from ~60 000 single-cell transcriptomes of juvenile zebrafish brain, shedding light on how gene expression cascades commit different cell types to restricted lineages and how different brain regions develop.

In the MARC1 mouse-based lineage tracing using many different hgRNAs, lineage marker mutations arise in distant loci of the genome and their combinations cannot be identified efficiently at the single-cell level. In GESTALT (and in scGESTALT), random mutations are all induced in a single array of targeting barcodes, whereby the mutational combinations of a cell can all be identified in a single sequencing read of massively parallel pyrosequencing. However, owing to the limit of sequencing read length, the barcode array cannot be elongated so it can harbor a higher number of mutational variants, capable of recoding lineages of all of the trillions of cells in the adult mammalian body. A computational simulation study has also suggested that some of the current lineage tracing methods have a limitation in amendable information content in their evolving DNA barcodes to uniquely label all of the cell lineage branches in a developmental process [35]. These issues could also be resolved using the concept of the evolving RNA barcode with scRNA-seq. In LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous

Figure 3



Lineage tracing methods for simultaneous annotation of cells. **(a)** scGESTALT. An evolving DNA barcode is transcribed as polyadenylated RNA. Mutation patterns in the evolving RNA barcodes of different cells can be identified with their corresponding single-cell transcriptome data, providing simultaneous observation of cell type and lineage. **(b)** MEMOIR. The founder cell has many transgenes expressing different barcodes concatenated to a common 'scratchpad' encoding a tandem repeat of gRNA targeting sites. As the cell divides, each scratchpad is collapsed by CRISPR–Cas9. Scratchpad statuses of different barcodes are identified for each cell with spatial information using seqFISH. **(c)** Coupled with high-capacity tissue imaging and *in situ* RNA sequencing technologies, an evolving DNA barcode system would enable high-resolution lineage reconstruction of cells from the whole mammalian body with their spatial information.

sequences) [28] and ScarTrace [29], developmental lineages of zebrafish were reconstructed via the CRISPR–Cas9 editing of ubiquitous sequences in RFP or GFP transgenes expressed from multiple independent genomic loci, in which the whole of the induced mutations in each cell was read by scRNA-seq together with cell type information. A better approach for lineage tracing in mouse development has been proposed more recently [30]; in this method, unlike in LINNAEUS or ScarTrace, unique DNA identifiers are introduced into the different transgene barcode units, all of which have a common array of three gRNA targeting sites, so mutation products of the same barcode loci in different single cells can be related directly. However, although the capacity for recording cell lineage information in an evolving DNA barcode system can be elevated accordingly using scRNA-seq, a bottleneck inhibiting large-scale cell lineage reconstruction using this strategy has remained as an obstacle, due to limited scalability of scRNA-seq, which is currently capable of processing only up to dozens of thousands of single cells.

Deciphering cell lineages with spatial information

Immediately after the first CRISPR-based evolving DNA barcode GESTALT, another proof-of-concept method, MEMOIR (memory by engineered mutagenesis with optical *in situ* readout), was reported for the simultaneous identification of spatial information of cells and their dividing lineages at a small scale [31••]. In this study, multiple different transcribing barcodes concatenated to a common 'scratchpad', a tandem repeat of ten identical gRNA targeting sequences, were introduced into parental mouse ES cells (Figure 3b). The scratchpads of cells were then irreversibly modified by Cas9 and gRNA along with cell divisions. By a multiplexed single-molecule RNA fluorescence hybridization (smFISH) method, namely, seqFISH [36,37], the transcribed barcode molecules were detected using the probing intensity for remaining intact repeats in the scratchpad, allowing reconstruction of the lineage of daughter cells based on the combinatorial barcoded scratchpad statuses. Combined with the simultaneous smFISH imaging of the pluripotency regulator *Esrrb*, this MEMOIR experiment showed that the short-term state switching of ES cells is also lineage-restricted from parental cells.

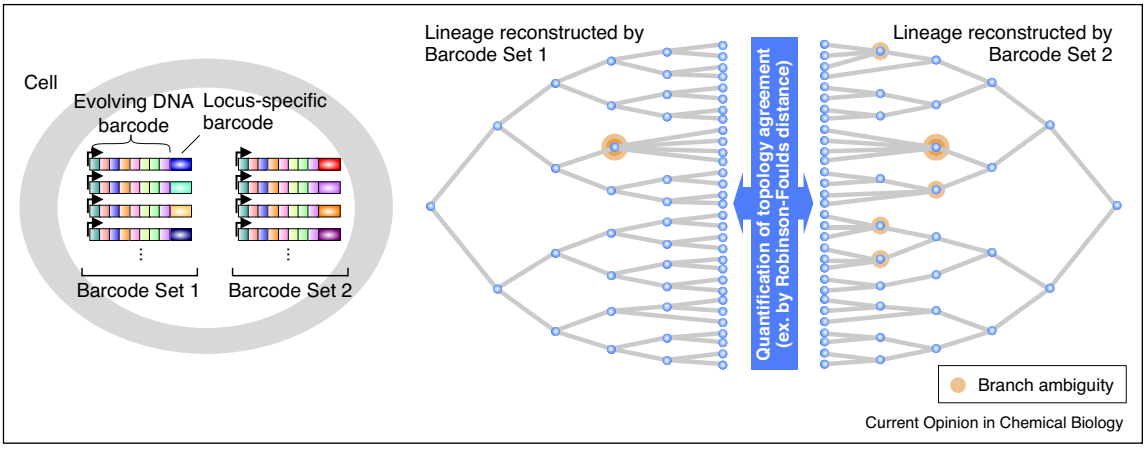
Although MEMOIR has been demonstrated only for small-scale lineage tracing from a single parental cell to up to eight daughter cells, this concept provided a foundational idea for current technological developments towards interpretation of the development and functional expression of multicellular organisms with both their underlying lineage and spatial structures. The rapid development of recent tissue clearing and imaging methods has enabled the exploration of spatial molecular distributions at the whole-organ and whole-body levels [38]. Whereas these methods have mainly focused on tissue localization of a given set of target proteins tagged by immunofluorescent probes or fused to fluorescent marker proteins, deep-tissue RNA imaging coupled with the MEMOIR-like approach could achieve spatial tracing of the lineages of large organs and bodies. Furthermore, unlike FISH-based imaging, which requires prior information about target RNA sequences, extensive efforts have been devoted to the *de novo* sequencing of RNAs in three-dimensional intact tissues, as represented by STARmap technology [39]. This prompted us to envision the spatial reconstruction of a high-resolution cell lineage tree by a parallel set of scGESTALT-like evolving RNA barcodes that continuously accumulate random mutations to label unique lineage branches in a high-content manner (Figure 3c).

Outlook and discussion

The process of cell division can be traced by live imaging under a microscope only for a small system, similar to the development of *C. elegans*. Fluorescent labeling of cells allows us to analyze the spatial distributions of labeled clones in a larger system with a certain resolution, which is limited by the number of fluorescent markers. DNA barcode-based unique labeling of a large number of cells coupled with pyrosequencing sacrifices spatial information of cells but enables high-capacity measurements of clone size dynamics and the distribution of clones into different types of cell, as seen in the examples of hematopoietic cell lineage analysis.

The concept of evolving DNA barcodes has appeared to enable the reconstruction of cell lineage trees of complex systems using CRISPR–Cas9 genome editing. However, these methods require further improvement because CRISPR-based evolving DNA barcodes have been demonstrated in several studies to saturate or stop in earlier phases of the developmental process [23••,26••,27•]. An ideal system would instead continuously accumulate unique new mutations in evolving barcodes at every cell division to entirely label all of the branches of the target cell lineage. Although the validity of the current lineage tracing methods has often been shown only by specific cell types being restricted or localized in the reconstructed lineages, none of the previous studies quantitatively assessed the accuracy of their lineage reconstruction methods. We here propose that the performance of an evolving DNA barcode system could be quantified if the same target lineage can be reconstructed using orthogonal sets of independent evolving DNA barcodes (Figure 4). For example, if multiple evolving DNA barcodes with distinct identifier sequences are transcribed from different chromosomal loci, subgroups of them can be used for separate lineage reconstructions with scRNA-seq and the agreement between the reconstructed lineages can then be quantitatively compared, for example, using Robinson–Fould distance.

Figure 4



Quantitative measurement of the resolution and accuracy of a lineage tracing method. For a lineage tracing method that utilizes many evolving DNA barcodes per cell, multiple cell lineage trees can be predicted from orthogonal sets of evolving DNA barcodes and the accuracy of the method can be measured by their agreement. Since mother cells always divide into two daughter cells, the resolution of the reconstructed lineage can also be formulated by combining the lineage agreement and distribution in numbers of branching edges of middle-layer nodes.

DNA barcodes serve not only as indices of cell clones and lineage branches, but also as addresses to link single-cell profiles and spatial positions of molecules and cells when coupled with scRNA-seq and imaging technologies, respectively. Several studies have demonstrated simultaneous analyses of cell types and their lineages at certain scales using scRNA-seq. MEMOIR has successfully traced a small-scale lineage of eight mouse ES cells with their spatial positions using seqFISH. With further advancements in single-cell and imaging technologies, an evolving DNA barcode system could accurately decipher whole-body cell lineages of large multicellular organisms beyond *C. elegans* with high-content cell type and spatial information. This approach, however, only maps information of cells obtained at the time of observation to the marginal edges of a reconstructed cell lineage, which is only the one that constitutes information tracing back into the past, and cannot address how different cells dynamically change their statuses across the lineage. An interesting simulation study was reported in the MEMOIR paper, in which the developmental histories of cellular events could also be reconstructed with cell lineages by transmitting targeted event signals to scratchpad editing [31••]. As discussed in another review on this issue [40], the development of DNA event recording systems has become another large field of synthetic biology, where various environmental and intracellular events are recorded in DNA tapes of cells using CRISPR-related methods and DNA recombinations. While their current implementations remain at the proof-of-concept stage, many of these approaches should theoretically be compatible to lineage tracing using the evolving DNA barcodes. Accordingly, the evolving DNA barcodes should further accelerate our understanding of various biological systems, including the entire developmental process of mammals.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank members of the Yachie laboratory for discussions about different aspects of the current and past cell lineage tracing approaches. The research projects of the Yachie laboratory in this field are supported by grants from the Japan Science and Technology Agency (JST) PRESTO and CREST programs, the New Energy and Industrial Technology Development Organization (NEDO) Smart Cell Project, the Japan Agency for Medical Research and Development (AMED) PRIME program, the Shimadzu Science Foundation, the Takeda Science Foundation, the SECOM Science and Technology Foundation, the Naito Foundation, the Nakajima Foundation, and the Asahi Glass Foundation.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Sulston JE, Horvitz HR: **Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans***. *Dev Biol* 1977, **56**:110-156.

2. Deppe U, Schierenberg E, Cole T, Krieg C, Schmitt D, Yoder B, von Ehrenstein G: **Cell lineages of the embryo of the nematode *Caenorhabditis elegans***. *Proc Natl Acad Sci U S A* 1978, **75**:376-380.

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

4. 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.

5. Hampel S, Chung P, McKellar CE, Hall D, Looger LL, Simpson JH: **Drosophila Brainbow: a recombinase-based fluorescence**

labeling technique to subdivide neural expression patterns. *Nat Methods* 2011, **8**:253-259.

6. Bhang H-C, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, Singh AP, Kao I, Rakiec D, Shaw P *et al.*: **Studying clonal dynamics in response to cancer therapy using high-complexity barcoding**. *Nat Med* 2015, **21**:440-448.

7. • 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**. *Nat Biotechnol* 2011, **29**:928-933.
The hematopoietic process was reconstructed by transplantation of DNA barcoded HSCs to an adult mouse followed by an analysis of barcode distribution after differentiation into different cell types using high throughput sequencing.

8. 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.

9. Mayer C, Jaglin XH, Cobbs LV, Bandler RC, Streicher C, Cepko CL, Hippenmeyer S, Fishell G: **Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries**. *Neuron* 2015, **87**:989-998.

10. Harwell CC, Fuentealba LC, Gonzalez-Cerrillo A, Parker PR, Gertz CC, Mazzola E, Garcia MT, Alvarez-Buylla A, Cepko CL, Kriegstein AR: **Wide dispersion and diversity of clonally related inhibitory interneurons**. *Neuron* 2015, **87**:999-1007.

11. Fuentealba LC, Rompani SB, Parraguez JI, Obernier K, Romero R, Cepko CL, Alvarez-Buylla A: **Embryonic origin of postnatal neural stem cells**. *Cell* 2015, **161**:1644-1655.

12. • Pei W, Feyerabend TB, Rossler 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 Polylox barcoding method was developed for *in vivo* barcoding of cells, whereby a non-invasive clone distribution analysis revealed that HSC clones gave rise to both multilineage and oligolineage fates.

13. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, Panero R, Patel SH, Jankovic M, Sun J, Calogero RA, Klein AM, Camargo FD: **Clonal analysis of lineage fate in native haematopoiesis**. *Nature* 2018, **553**:212-216.

14. 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.

15. 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.

16. Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann HS, Park PJ *et al.*: **Cell lineage analysis in human brain using endogenous retroelements**. *Neuron* 2015, **85**:49-59.

17. Frumkin D, Wasserstrom A, Itzkovitz S, Stern T, Harmelin A, Eilam R, Rechavi G, Shapiro E: **Cell lineage analysis of a mouse tumor**. *Cancer Res* 2008, **68**:5924-5931.

18. 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.

19. Reizel Y, Chapal-Ilani N, Adar R, Itzkovitz S, Elbaz J, Maruvka YE, Segev E, Shlush LI, Dekel N, Shapiro E: **Colon stem cell and crypt dynamics exposed by cell lineage reconstruction**. *PLoS Genet* 2011, **7**:e1002192.

20. Salipante SJ, Kas A, McMonagle E, Horwitz MS: **Phylogenetic analysis of developmental and postnatal mouse cell lineages**. *Evol Dev* 2010, **12**:84-94.

21. Baslan T, Kendall J, Ward B, Cox H, Leotta A, Rodgers L, Riggs M, D'Italia S, Sun G, Yong M *et al.*: **Optimizing sparse sequencing of single cells for highly multiplex copy number profiling**. *Genome Res* 2015, **25**:714-724.

22. 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* 2011, **472**:90-94.

23. •• 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**:eaaf7907.
GESTALT was the first large-scale lineage tracing method that demonstrated the evolving DNA barcode concept by harnessing CRISPR–Cas9 genome editing. From this study a whole-body zebrafish developmental lineage was inferred.

24. • Kalhor R, Mali P, Church GM: **Rapidly evolving homing CRISPR barcodes**. *Nat Methods* 2016, **14**:195-200.
The two-component evolving DNA barcode was developed using a mutated self-targeting gRNA system. See also Ref. [25].

25. Perli SD, Cui CH, Lu TK: **Continuous genetic recording with self-targeting CRISPR-Cas in human cells**. *Science* 2016, **353**: eaag0511.

26. •• Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, Church GM: **Developmental barcoding of whole mouse via homing CRISPR**. *Science* 2018, **361**:eaat9804.
This paper reported the first mammalian whole body cell lineage tracing method using evolving DNA barcodes. A MARC1 founder mouse harboring 60 hgRNAs was established as a toolkit mouse line in this study.

27. • 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**. *Nat Biotechnol* 2018, **36**:442-450.
An evolving DNA barcode was transcribed and readout by scRNA seq thereby providing both transcriptome-based cell type information of end point daughter cells and cell lineage information. See also Refs. [28–30].

28. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP: **Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars**. *Nat Biotechnol* 2018, **36**:469-473.

29. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A: **Whole-organism clone tracing using single-cell sequencing**. *Nature* 2018, **556**:108-112.

30. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, Jost M, Quinn JJ, Yang D, Jones MG *et al.*: **Molecular recording of mammalian embryogenesis**. *Nature* 2019, **570**:77-82.

31. •• 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* 2016, **541**:107-111.
MEMOIR was the first method that successfully demonstrated the reconstruction of cell lineages with the spatial position of daughter cells by harnessing CRISPR–Cas9 genome editing and seqFISH RNA imaging.

32. Stuart T, Satija R: **Integrative single-cell analysis**. *Nat Rev Genet* 2019, **20**:257-272.

33. Macosko Evan Z, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas Allison R, Kamitaki N, Martersteck Emily M *et al.*: **Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets**. *Cell* 2015, **161**:1202-1214.

34. Klein Allon M, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz David A, Kirschner Marc W: **Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells**. *Cell* 2015, **161**:1187-1201.

35. Salvador-Martínez I, Grillo M, Averof M, Telford MJ: **Is it possible to reconstruct an accurate cell lineage using CRISPR recorders?** *eLife* 2019, **8**:e40292.

36. Shah S, Lubeck E, Zhou W, Cai L: ***In situ* transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus**. *Neuron* 2016, **92**:342-357.

37. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L: **Single-cell *in situ* RNA profiling by sequential hybridization**. *Nat Methods* 2014, **11**:360-361.

38. Richardson DS, Lichtman JW: **Clarifying tissue clearing**. *Cell* 2015, **162**:246-257.

39. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J *et al.*: **Three-dimensional intact-tissue sequencing of single-cell transcriptional states**. *Science* 2018, **361**:eaat5691.

40. Ishiguro S, Mori H, Yachie N: **DNA event recorders send past information of cells to the time of observation**. *Curr Opin Chem Biol* 2019, **52**:54-62.