MOLECULAR BIOLOGY

Molecular recorders to track cellular events

DNA tapes could be used to record dynamic molecular and cellular events in animals

By **Nanami Masuyama[1,2], Naoki Konno[3], Nozomu Yachie[1,4]**

Starting from a fertilized egg, cells proliferate, pass their genomic information to their progeny, and dynamically change their functions to form diverse tissue structures. Throughout development, intracellular and environmental cues trigger patterns of gene expression that govern cell state transitions and produce further cellular and environmental cues, leading cells to self-organize into functional clusters in spatially distinct areas. How can these processes be investigated? High-resolution molecular snapshots of cells can be obtained by various omics technologies, but these require sample destruction, which precludes analyses over time. Live-cell imaging with fluorescent probes can analyze time-course dynamics but is limited to a small number of molecules or cells. These challenges could be overcome by DNA event recording, where molecular and cellular information is progressively stored in synthetic "DNA tapes" in cells within a multicellular organism, and the stored information is accessed by high-throughput DNA sequencing. However, there are technical limitations that need to be overcome.

One of the DNA event recording modalities, dynamic cell lineage recording using CRISPR-Cas9 genome editing, has accelerated developmental cell lineage tracing. In this system, chromosome-embedded synthetic DNA barcode arrays are continuously and randomly mutated by Cas9, which is recruited by the expression of many barcode-targeting guide RNAs (gRNAs); these mutated arrays are progressively inherited by daughter cells (*1–3*) (see the figure). Similar to an evolutionary tree reconstruction from genomic sequences of diverse species, the cell lineage of an animal initiated from a fertilized egg can be reconstructed according to the mutation patterns of the barcode sequences in descendant cells at the time of observation using high-throughput DNA sequencing. Such lineage tracing systems have also been linked to single-cell RNA sequencing (scRNA-seq), where individual cell types identified by scRNA-seq can be linked to their historical cell lineage information (*3*).

Today, new technologies are rapidly revising perspectives on developmental biology, such as the origins of the hypoblast and embryo-restricted epiblast (*4*), blood cell differentiation, and the deciphering of cell proliferation dynamics of metastatic cancers (*5*, *6*). Furthermore, ideas to develop other cell-embedded circuits that can convert environmental and cellular signals to genome-editing activities have emerged (*2*). Combined with a cell lineage recorder, such systems would enable the computational reconstruction and mapping of historical cellular events and cell state changes onto a cell lineage. These ideas could lead to the characterization of high-resolution cell lineages, cell state changes, and cell communications during whole mammalian body development, homeostasis, and malfunction. This would deepen knowledge about developmental and cancer biology that could be translated into therapeutic strategies. However, DNA event recording has yet to trigger the next gold rush in biology because there are some missing components. The current molecular event recording methods need to overcome four substantial limitations in information recording capacity, available molecular sensors to capture diverse biological events, the number of cells that can be analyzed, and computational capacity to reconstruct high-content biological history.

Information recording capacity could be increased by extending the total length of the DNA barcodes. For example, several lineage tracing methods have used multiple short DNA barcode units per cell for the random accumulation of CRISPR-induced mutations. To read out the combinatorial mutation patterns in these barcodes in distal genomic loci, constitutively active promoters are introduced to allow transcription, and the mutation information is captured by scRNA-seq. The current methods use either a few copies of long gRNA-targeting barcode arrays (up to 10 barcodes per array, whose size is restricted by short-read DNA sequencing methods) or many copies of short barcode arrays with single gRNA-targeting units. There seems to be no barrier to combining these two approaches to achieve a high-content event recording system (i.e., encoding many long barcode arrays per cell).

Altering the type of information-writing enzyme would also increase the information recording capacity. Most of the current cell lineage tracing systems use wild-type Cas9, which induces double-stranded DNA breaks that are toxic to cells and can also confer large deletions, especially when multiple breaks are introduced simultaneously. Additionally, these large deletions can overwrite acquired mutations and progressively collapse the barcodes. These limitations could be alleviated by using a CRISPR base editor (BE), fusing a catalytically inactive Cas9 to a nucleoside deaminase that introduces single-nucleotide substitutions at the gRNA target sequence with minimal double-stranded DNA breaks. Most BEs confer unidirectional substitutions (e.g., C→T or A→G) (*7*). This could be a downside, given that the lineage recording ability crumbles when all substrate C or A bases in the DNA barcodes are converted and thus depleted. However, the recent development of dual-function BEs has enabled the simultaneous introduction of C→T and A→G substitutions. Using such a dual-function BE that can confer reversible C:G⟷T:A mutations will enable more-accurate lineage tracing than wild-type Cas9 or single-function BEs. In evolutionary biology, phylogenetic trees have been efficiently reconstructed from revertant mutations. BEs have also been demonstrated to record cellular signals but only for a limited number of events in mammalian cells, such as Wnt signaling (*8*).

The CRISPR prime editor (PE), fusing a catalytically inactive Cas9 to reverse transcriptase, has recently been proposed for scalable, high-content event recording. In these genome-editing systems, a PE gRNA (pegRNA) encodes a synthetic short payload sequence and recruits the PE to the target sequence, in which—coupled with endogenous DNA repair machineries—the reverse transcription activity of PE enables the integration of the payload sequence to the target site (*7*). This enables more diverse data encoding per barcode than wild-type Cas9 and BEs. Preliminary data show that the gene expression activities of 300 enhancers can be recorded to a single target DNA locus of a cell population using a library of pegRNAs, each of which is encoded downstream of a specific enhancer element (*9*). PE has also been programmed for a cascade reaction, where the order of complex events can be recorded by arraying event-specific payload sequences

[1]School of Biomedical Engineering, Faculty of Applied Science and Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada. [2]Institute for Advanced Biosciences, Keio University, Yamagata, Japan. [3]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [4]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. Email: nozomu.yachie@ubc.ca

Downloaded from https://www.science.org at University of British Columbia on July 28, 2022

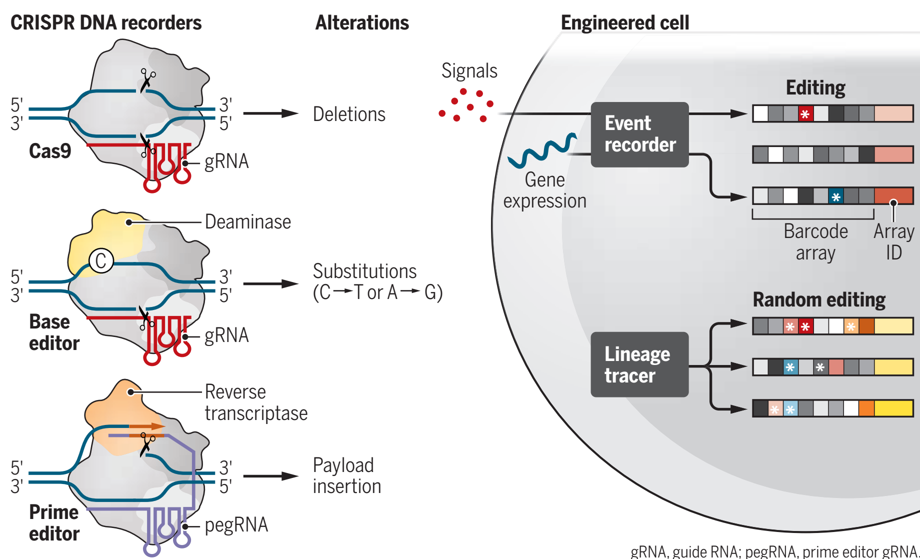
at a target DNA locus like a ticker tape (*10*). However, these approaches have not yet been implemented in animal models.

High-content event reconstruction necessitates high-capacity scRNA-seq. A cell lineage recording system that confers at least one unique mutation per cell division would yield the perfect reconstruction of cell lineage if all cells of a body can be analyzed. Analysis of a limited number of cells enables the historical reconstruction of only an upstream cell lineage whose descendant information can be covered by the analytes. The current state-of-the-art scRNA-seq can process multiple millions of cells, whereas a mouse body consists of multiple billions of nucleated cells. Given the rapid development of scRNA-seq since the advent of the droplet-based methods that enabled the analysis of multiple thousands of cells (*11*), it is not unreasonable to predict that technologies that can capture a greater magnitude of cells will appear in the next 5 years. Additionally, spatial RNA-seq has enabled spatial profiling of gene expression patterns in a tissue section or a three-dimensional block in high resolution (*12*). Coupled with DNA event recording systems whose readout information is converted into RNA, this modality also enables the reconstruction of cell lineage and cell state trajectories throughout a physical space. However, spatial RNA-seq allows only the reconstruction of later events that contributed to its formation. Therefore, more-scalable methods need to be developed to capture the whole developmental process of an animal body.

High-performance computing methods also need to be developed with these engineering efforts. Although many of the current cell lineages have been generated by repurposing existing evolutionary tree reconstruction tools, their algorithms need to be optimized for information recording biases in specific lineage tracing methods (*13*). Efficient algorithms for other biological events have yet to be established. Furthermore, no current phylogeny estimation software tool can reconstruct a lineage of more than one million inputs, raising concerns about how to reconstruct a biological history forming a mammalian body of at least hundreds of millions of cells. A new computing framework called deep distributed computing has been proposed to overcome this limitation (*14*). This framework has enabled accurate cell lineage tracing of more than two hundred million mutated sequences. In brief, starting from a single computing node, it reconstructs an upstream lineage and assigns all the input sequences into its subclades, where an independent computing node recursively resolves the next upstream lineage of each subclade. Like a spreading virus, a single computing task autonomously generates many distributed computing tasks, each of which iterates the same procedure. There is no conceptual barrier to implementing other event reconstruction processes into this framework.

The DNA event recording concept

Cell-embedded synthetic CRISPR-Cas information-writing enzymes can progressively record cellular events and lineage transitions into DNA tapes (or barcode arrays) by introducing various types of DNA alteration. DNA tapes are read out by single-cell sequencing, and the biological event histories are computationally reconstructed. Cell lineage can be recorded by continuous accumulation of random mutations into DNA tapes. Other event recording systems necessitate editing event-specific DNA tape positions or inserting specific DNA payloads into the DNA tape.



gRNA, guide RNA; pegRNA, prime editor gRNA.

Technologies for high-resolution DNA event recording have been maturing, which could expand the horizons of biology in different animal models—perhaps starting from mice for mammalian species. It is expected that this approach will identify unknown cell differentiation behaviors and their underlying molecular and cellular regulatory mechanisms. Analysis of many individuals' developmental processes in a single species in high resolution would characterize deterministic and stochastic cell differentiation patterns and quantify the ability of cells to undergo state transitions, enabling more precise computational modeling of dynamic multicellular systems with biophysical perspectives. Direct observation of human developmental procedures with the same approach is impossible because of ethical considerations regarding the genetic manipulation of human embryos. However, high-resolution developmental maps of diverse nonhuman species and comparative genomics approaches could also reveal human developmental processes. Furthermore, as seen in the emerging CRISPR cell lineage tracing studies of human organoid development (*15*) and tumor progression to metastasis in a mouse model (*5*, *6*), embedding a high-content DNA event recorder system in human pluripotent stem cells for organoid studies and in cancer cells for animal xenograft studies can more directly contribute to understanding their behaviors. Combined with genome-editing technologies that can reconstitute disease mutations in the human genome, these new model systems will also serve as testbeds with substantially expanded data monitoring abilities in therapeutic development for genetic disorders and cancers. ■

REFERENCES AND NOTES

1. A. McKenna *et al.*, *Science* **353**, aaf7907 (2016).
2. K. L. Frieda *et al.*, *Nature* **541**, 107 (2017).
3. S. VanHorn, S. A. Morris, *Dev. Cell* **56**, 7 (2021).
4. M. M. Chan *et al.*, *Nature* **570**, 77 (2019).
5. K. P. Simeonov *et al.*, *Cancer Cell* **39**, 1150 (2021).
6. D. Yang *et al.*, *Cell* **185**, 1905 (2022).
7. A. V. Anzalone *et al.*, *Nat. Biotechnol.* **38**, 824 (2020).
8. F. Farzadfard *et al.*, *Mol. Cell* **75**, 769 (2019).
9. W. Chen *et al.*, bioRxiv 10.1101/2021.11.05.467434 (2021).
10. J. Choi *et al.*, *Nature* 10.1038/s41586-022-04922-8 (2022).
11. R. Stark *et al.*, *Nat. Rev. Genet.* **20**, 631 (2019).
12. S. K. Longo *et al.*, *Nat. Rev. Genet.* **22**, 627 (2021).
13. W. Gong *et al.*, *Cell Syst.* **12**, 810 (2021).
14. N. Konno *et al.*, *Nat. Biotechnol.* **40**, 566 (2022).
15. Z. He *et al.*, *Nat. Methods* **19**, 90 (2022).

ACKNOWLEDGMENTS

We thank members of the Yachie laboratory for useful insights, especially S. King and A. Adel. The authors are supported by the Canada Research Chair program, a project grant (PJT 178040) by the Canadian Institutes for Health Research, and a Pilot Innovation Fund (PIF003) by Genome British Columbia. N.M. and N.K. are supported by fellowships from the Japan Society for the Promotion of Science. N.M. and N.K. contributed equally to this work.

10.1126/science.abo3471

GRAPHIC: K. HOLOSKI/*SCIENCE*

Downloaded from https://www.science.org at University of British Columbia on July 28, 2022

Science

Molecular recorders to track cellular events

Nanami MasuyamaNaoki KonnoNozomu Yachie

Science, 377 (6605), • DOI: 10.1126/science.abo3471

View the article online
https://www.science.org/doi/10.1126/science.abo3471
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at University of British Columbia on July 28, 2022

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.
Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works