

## SYSTEMS BIOLOGY

# Molecular memoirs of a cellular family

A system that introduces random modifications to barcode sequences embedded in cells' DNA allows lineage relationships between cells to be discerned, while preserving the cells' spatial relationships.

LAUREN E. BECK & ARJUN RAJ

Determining how large, multicellular organisms emerge from a single cell is a major goal for biologists, and a key to understanding many dynamic processes and diseases. A fundamental component of this goal is the mapping of cellular lineages, which is crucial for understanding not only embryonic development, but also processes such as the growth of cancers and stem-cell differentiation. In a paper online in *Nature*, Frieda *et al.*<sup>1</sup> describe a technique that uses genetic engineering to embed lineage information in a cell's DNA. The results can then be read out by direct visualization — an advance that stands to transform our ability to build spatio-temporal cell lineages.

One of the landmark achievements in the field of lineage tracing has been the ability to resolve the complete developmental cell-lineage tree for the roundworm *Caenorhabditis elegans*<sup>2</sup>. This tree revealed a great deal about how development progresses, and most notably uncovered an essential process called programmed cell death<sup>3</sup>. These successes were a result of, in equal parts, elegant manipulation and elbow grease. The worm's relatively small and invariant lineage tree, the optical transparency of *C. elegans* and the sheer determination of the researchers involved were crucial to mapping the embryonic lineages of this organism.

When it comes to more-complex organisms, how can we tell, just from looking, which cells are sisters, cousins, aunts or uncles? An ability to read out and compare the DNA-sequence variation between individuals and species has revolutionized studies of ancestry, and the same strategy can be applied to cells. However, whereas species have had time to accumulate many DNA mutations, individual cells accumulate natural mutations rarely and randomly, making them difficult to measure.

To get around this problem, Frieda *et al.* used a genome-editing system called CRISPR-Cas9, which can precisely delete specific DNA sequences. The authors constructed a series

of identical DNA sequences that they dubbed scratchpads. They then attached each one to a unique sequence called a barcode, and introduced all of these barcoded scratchpads into the genome of an individual cell. Over time, the Cas9 enzyme randomly removes scratchpads, leaving some barcodes without an associated scratchpad. The cell transcribes its barcoded scratchpads into RNA, which the researchers can detect using fluorescent probes — one set of probes that binds to the scratchpads and another set that binds to each barcode. In this way, the researchers can tell which barcodes had their corresponding scratchpads removed<sup>4,5</sup>.

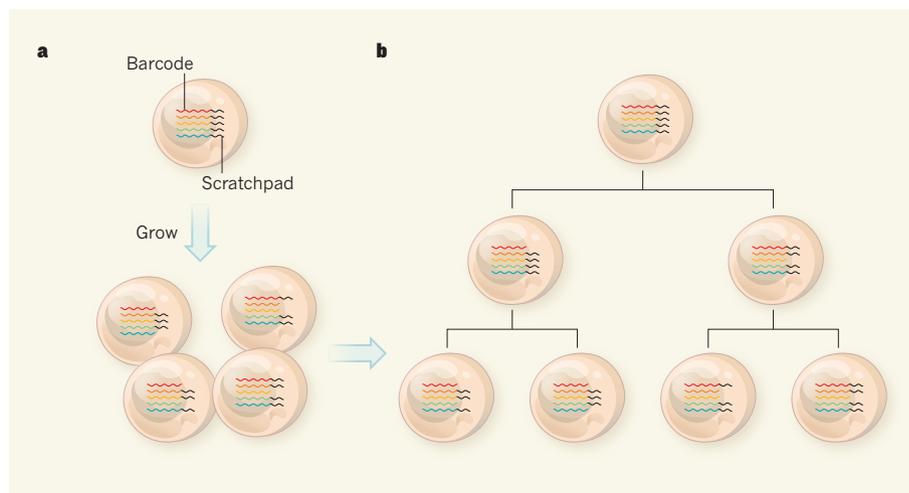
The cell's progeny inherit the deleted scratchpads and then go on to accumulate other deleted scratchpads, allowing the researchers to read out lineage information by determining (through fluorescent imaging) the subset of edited barcodes that differ between any two cells (Fig. 1). Frieda and

colleagues call this technique 'memory by engineered mutagenesis with optical *in situ* readout' (MEMOIR).

As a proof of principle, the authors applied MEMOIR to mouse embryonic stem cells grown in culture. As a reference against which to compare MEMOIR's results, they used time-lapse microscopy to track lineage development over three or four cell divisions. Impressively, despite all the random events required for the method to work, they were still able to reconstruct even these relatively fine-grained family trees up to about 72% of the time. These results were squarely within the range obtained when they used computer simulations to predict the accuracy of an idealized MEMOIR experiment, showing that the method performs optimally from a technical perspective.

The real beauty of the system is that, because the ultimate measurement occurs *in situ* (in intact cells in their native positions), it both preserves the spatial relationships of cells and allows for measurements of other features, such as gene expression. To demonstrate the power of this approach, Frieda *et al.* studied *Esrrb*, a gene that has highly variable expression in embryonic stem cells, at the same time as measuring lineages using MEMOIR.

If *Esrrb* expression slowly switches between on and off, one would expect the progeny of a cell that has 'on' expression also to have 'on' expression (and the same would be true for 'off' expression). However, if *Esrrb* expression rapidly switches between on and off, one would expect the expression state of a daughter cell to be largely independent of that of its



**Figure 1 | Lineage tracing finds its place.** **a**, Frieda *et al.*<sup>1</sup> use genetic engineering to introduce a series of identical DNA sequences called scratchpads, each attached to a unique DNA sequence called a barcode, into the genome of a cell. As the cell proliferates, scratchpads are stochastically and permanently deleted. Microscopy, combined with RNA-detection techniques, can be used to identify the RNA transcripts produced from each barcode in each cell — in this way, the barcodes attached to deleted scratchpads can be identified. **b**, This information can be used to reconstruct a complete cell-lineage tree. In this example, the tree reveals that the two cells on the left in **a** are sisters, and that the two on the right are their cousins.

parent. In their test, the authors inferred expression dynamics for *Esrrb* — the probabilities that expression would flip from off-to-on or on-to-off were 0.09 and 0.04 per generation, respectively. Although the current iteration of MEMOIR serves mainly to prove its efficacy for lineage tracing, it is remarkable that these temporal quantities can be inferred without having to make dynamic measurements.

As the number of barcodes increases, the fidelity and complexity of the relationships that MEMOIR can read out stands to increase. Frieda and colleagues used tens of barcodes, but advances in *in situ* RNA-detection methods<sup>6–9</sup> should enable the use of hundreds or thousands, vastly increasing the power of MEMOIR to discern family trees with high fidelity. Moreover, the methods for RNA detection used by the authors can be applied in animal tissues, in theory allowing family trees to be directly visualized on top of an

image of, say, an organ or potentially an entire organism. Combining MEMOIR with techniques for imaging in tissues may facilitate such developments.

Notably, a study<sup>10</sup> this year outlined a complementary approach to lineage tracing, which also used CRISPR–Cas9 to induce genomic mutations, but that read out the results with DNA or RNA sequencing, rather than imaging. Sequencing comes at the expense of spatial information, but does enable hundreds of thousands of cells to be interrogated. Developments in spatial genomics might help to circumvent the spatial limitations of sequencing-based lineage strategies<sup>11</sup>. These techniques, complemented by methods such as MEMOIR, stand to reinvigorate the field of lineage tracing in the coming years. ■

**Lauren E. Beck and Arjun Raj** are in the Department of Bioengineering, University of

Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

e-mails: [labeck@seas.upenn.edu](mailto:labeck@seas.upenn.edu); [arjunraj@seas.upenn.edu](mailto:arjunraj@seas.upenn.edu)

1. Frieda, K. L. *et al.* *Nature* <http://dx.doi.org/10.1038/nature20777> (2016).
2. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. *Dev. Biol.* **100**, 64–119 (1983).
3. Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110–156 (1977).
4. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. *Nature Methods* **5**, 877–879 (2008).
5. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. *Nature Methods* **11**, 360–361 (2014).
6. Levesque, M. J. & Raj, A. *Nature Methods* **10**, 246–248 (2013).
7. Levisky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. *Science* **297**, 836–840 (2002).
8. Lubeck, E. & Cai, L. *Nature Methods* **9**, 743–748 (2012).
9. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. *Science* **348**, aaa6090 (2015).
10. McKenna, A. *et al.* *Science* **353**, aaf7907 (2016).
11. Crosetto, N., Bienko, M. & van Oudenaarden, A. *Nature Rev. Genet.* **16**, 57–66 (2015).