

Following the fate of cells *in vivo*

Researchers are building better barcodes and combining spatial and temporal tools to trace cell lineages from the earliest stages of animal development—and beyond.

Jim Kling

Researchers have long been fascinated by the developing embryo. 19th century scientists wondered whether the fertilized egg contains tiny structures that guide structural development down the line, or if the molecules contained in the egg preferentially adopt certain structures, like in a crystal. Others wondered if the parent passes instructions to the offspring epigenetically.

The solution seemed to be to track the fate of cells individually. In the early 1980s, researchers led by John Sulston at the University of Manchester used light and electron microscopy in the developing nematode *C. elegans* to track 671 cells generated during embryogenesis all the way to terminal differentiation or into formation of undifferentiated blast cells¹. Every cell division had to be recorded each nuclei sketched by hand and color-coded to indicate the cell's depth. Later divisions became impossible to track directly, especially in the interior, requiring researchers to infer a cell's ancestry by relying on repeating developmental patterns that allowed them to pinpoint previously identified cells. This Herculean effort gave insight to developmental patterns and informed years of future research in embryology.

The worm, however, has relatively few cells to follow compared to other animals, and it was unrealistic to painstakingly track cells in other, more complex organisms using microscopy alone. In the past decade, the field of cell lineage tracing has accelerated, aided by the surging capacity of single cell, high throughput sequencing to determine a cell's identity and the advent of gene editing tools such as CRISPR/Cas9, which can be used to introduce 'barcodes' for keeping track of cells as they divide. An algorithm then sorts the barcodes, making its best guesses as to where the cells fall on a sort of family tree.

Such information can greatly improve understanding of developmental processes, both normal and abnormal, according to Jan Philipp Junker, group leader at the Max Delbrück Center for Molecular Medicine. "It's unclear how many ways there are to build an embryo. In *C. elegans*, they found



With editing technologies such as CRISPR-Cas9, researchers can introduce 'barcodes' to keep track of the fate of different cells in an organism. Pairing these with additional technologies, such as transcriptomics and *in situ* hybridizations, can reveal additional details such as cell identity and spatial locations. Credit: Neil Leslei, Stockbyte, Getty

that each cell has a precisely defined role, but in vertebrates we know this is not the case. We just don't know how variable it is," said Junker, who studies developmental variation and stability in zebrafish embryos.

Lineage tracing can also reveal surprising relationships between cells. Two cells that look almost identical based on RNA transcripts, which researchers use to infer its function and identity, sometimes turn out to be only distantly related. "You (have discovered) new cell types," said Anna Alemany, a postdoc at the Hubrecht Institute in The Netherlands.

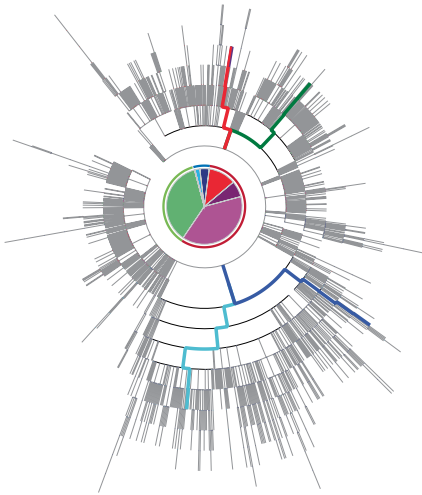
New cell types, or novel subpopulations, can improve understanding of biological systems and even lead to therapeutic leads. "Let's say in a disease we find that new cell types or states arise that we haven't seen before. Lineage tracing on a systematic high throughput level is very useful to understand

where these cell types come from. If a detrimental cell type forms, if we know where it comes from, maybe we can try to interfere with it in a more targeted way," said Junker.

There's been much recent progress, and the field of cell lineage tracing is still evolving. Researchers are working to make the technology itself better, and pairing it with emerging methods to trace cells in both time and in place.

Tracing gets CRISPR'ed

Cell lineage tracing operates by analyzing genetic changes that occur within cells. One of the limitations of current systems is that the timing of those changes is unpredictable, and generally much slower than the rate of cell division. That means that unknown generations of cell divisions may pass before another change occurs, leading to fuzzier



Reconstructed cell lineage tree for a mouse embryo. Reprinted with permission from Chan (2019)², Springer Nature.

lineages resembling family trees that are missing numerous aunts, grandparents, and cousins. “The resolution of these techniques is far lower than you would hope for,” said Michelle Chan, a postdoc in Jonathan Weissman’s lab at the University of California at San Francisco.

Like many who use lineage tracing methods, the Weissman group employs CRISPR-Cas9 to modify ‘barcodes’ embedded in the model organism’s genome—in their case, the mouse². These introduced sequences, which have no effect on cell function themselves, are where the CRISPR-Cas9 machinery sets to work. A guide sequence leads the Cas9 enzyme to the barcodes, where it makes a cut. The cell’s machinery then repairs the damage, usually causing a deletion. With every cell division, this change is passed down to all daughter cells. When individual daughter cells undergo further CRISPR-Cas9-induced changes at other barcodes, the result is a series of nested mutation trees. When it’s time for an analysis, researchers sequence the messenger RNA of each cell individually, and an algorithm sorts through the varying combinations of barcode mutations, generating a best fit family tree for the cells. Other techniques rely on viruses or other methods to introduce changes to a barcode, but the general idea is the same.

There are some limitations of CRISPR methods. There’s the impact of CRISPR-based alterations—those double-stranded breaks are stressful to cells, and can sometimes kill them. Another key limitation is the limited amount of storage available in a given barcode. With each deletion introduced by the

Box 1 | Tracking beyond embryos

Lineage tracing isn’t just limited to studying cell fate in embryos. Junker has expanded his studies past development, to the mature zebrafish heart. Unlike humans, zebrafish can regenerate heart tissue, but little is understood about the cells involved in the regenerative niche that is formed after injury. “We want to identify the cells in the regenerative niche, and then understand their origins by lineage tracing, and try to understand their function based on the genes that they express and the pathways they’re involved in,” said Junker.

His team used a CRISPR-Cas9 system to identify and trace the lineage of fibroblast-like cells that appear after an injury. “We have a couple of interesting examples of the function that these cells then acquire,” said Junker.

Weissman’s group has also begun to apply its CRISPR approach to medical

cell’s repair machinery, significant portions of the barcode are removed, thus erasing some of the evidence of the cell’s lineage. After a certain number of alterations, any barcode will lose its capacity to store lineage information, which typically limits lineage tracing to a short period of the organism’s development.

But lineage tracing is valuable in other longer-term applications as well (Box 1). For example, cancer experiments, which may last six months or longer, use lineage tracing to examine the origin, evolution, and metastatic patterns of tumors in mouse models. “We knew we were going to do experiments in embryogenesis which were going to be about 10-day experiments, and cancer experiments which were far longer. So we wanted to create one technology that would provide useful information for both those time scales,” says Chan.

There are several approaches to lengthening the lifespan of a barcode. For experiments designed to run for longer periods of times, such as studies of tumor formation, the Cas9 system can be slowed down. For example, Weissman’s group can ‘tune’ their CRISPR-based recorder in mice by introducing mismatches within the guide sequences that recognize these barcodes, which slows the process of guiding Cas9 to the barcode. As a result, the rate of barcode changes slows and more time passes before barcodes get deleted. This lengthens the useful lifetime of that barcode’s memory. Cas9 can be added exogenously, or it can be engineered into the organism

problems. In a *bioRxiv* preprint⁶, the group described metastatic progression of lung cancer transplanted into a xenograft mouse model. The implanted cells were engineered to express Cas9, ten barcodes, guide RNAs, and luciferase to track the progeny using live imaging. The guide RNAs contained mismatches to slow the lineage recording rate and allow lineage tracing over the 54 days that lapsed before the animal was sacrificed.

Researchers then dissected tumor tissue and worked out lineage trees. The work revealed phylogenies for tens of thousands of descendant cells, and the team discovered cell populations with wide-ranging characteristics: some were aggressively metastatic, others incapable of spread. The live imaging revealed specific tissues where metastasis occurred, as well as the direction of movement.

with a promoter that responds to dietary doxycycline. That can offer the ability to create ‘pulses’ of lineage tracing activity by adding or removing doxycycline, providing more information about specific times during development, says Chan.

The team is also working to make it possible to restrict lineage tracing to specific tissues. The idea is to have Cas9 expression under the control of a promoter that can be limited to a tissue type, and then cross the mice with mice carrying that promoter in the tissue of interest. Selected offspring then produce Cas9 in specific tissue, such as the developing brain, allowing researchers to produce cell lineage profiles there.

Though CRISPR has greatly enabled cell lineage tracing, it still only captures one to two percent of cell divisions. “That’s still thousands of cell divisions per organism, which is an order or two of magnitude higher than we’ve seen before. So there’s a lot of interesting data you can get from only sampling one percent of trees,” said James Gagnon, assistant professor of biology at the University of Utah.

Still, he and others are working to expand the recording capacity of cell lineage tracing in order to capture more cell divisions. His group has taken advantage of the fact that a number of different CRISPR systems have been found in a wide range of bacteria, many of which can act completely independently of one another. Some can make specific changes to individual bases within a barcode, or cause insertion of defined sequences into the double-stranded

break, rather than a deletion, which would reduce the loss of lineage information. Others employ CRISPR inhibitors produced by the viruses that CRISPR targets. That can switch CRISPR off and on during key developmental time points. “We hope they’ll let us be really clear about when and where mutations are happening,” said Gagnon. His group has also combined two, and more recently three different CRISPR systems (unpublished work) to allow them to record cell lineages at various time points.

They are also experimenting with using longer barcode arrays, which expands the memory storage capacity, as well as more advanced single cell sequencing technologies to read out the mRNA from the barcodes

(scRNAseq). Existing systems don’t capture all of the mRNA in any given cell, so some lineage information never gets read. As a postdoc in at the University of Basel, and together with fellow postdoc Bushra Raj, Gagnon applied CRISPR mutation editing of barcodes with scRNAseq to development of the zebrafish brain. It’s an exciting model, he says, because researchers can track development from fertilization to a point at which a 1-month old fish is displaying multiple adult behaviors, including feeling and schooling. But the fish are still growing. The team found that almost 20% of the cells in these fish look like stem cells they rapidly divide, and they generate new neurons³. “These neurons are added to a functional brain, which I think is really cool,” said Gagnon.

They also examined the forebrain, midbrain, and hindbrain, and found that each had distinct populations of progenitor cells, all churning out neurons and seemingly specialized to their own region. “So there’s already spatial organization in the brain, and if we look specifically at regions like the hypothalamus, we can identify all of the cell types in that region, and see lineage relationships that led down to those structures. That tells us a lot about the development process by which the brain is generated,” said Gagnon.

Combinations in time and space

As labs are introducing emerging technologies to improve barcodes and identify cell types, others are returning to cell lineage tracing’s roots, combining it with newer, more advanced imaging to add a sense of place to lineage information. Michael Elowitz, and Long Cai, biological engineers at the California Institute of Technology, have combined CRISPR-Cas9 genome editing tools with spatial, imaging-based readouts of cells. Like other methods, the CRISPR system produces heritable changes in cells that can be grouped to determine lineage. But rather than using sequencing to read the barcodes, Elowitz and Cai developed the MEMOIR system⁴. MEMOIR leaves cells intact and

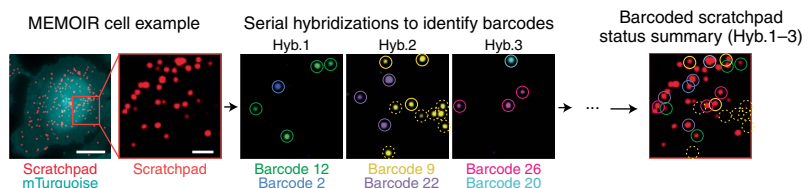
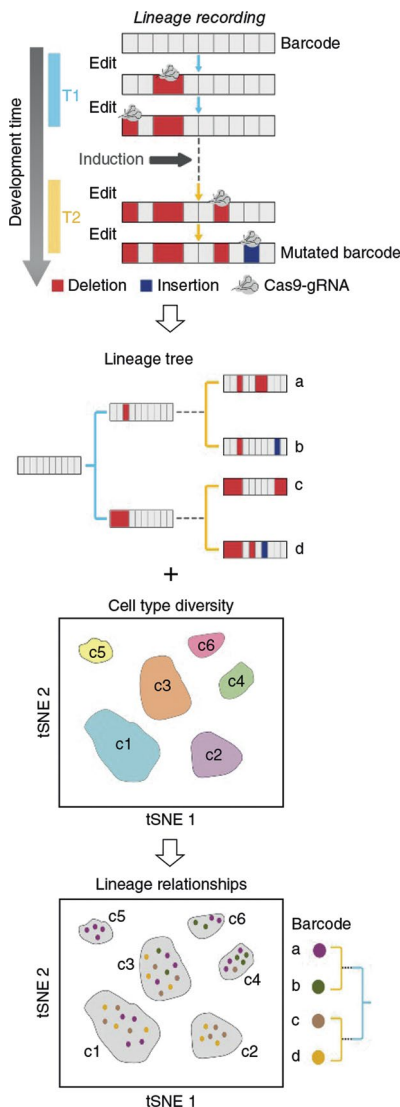
in place and then employs sequential single molecule fluorescence in situ hybridization (seqFISH), a technology for spatial visualization of cellular identity.

That spatial information is key to understanding developmental processes, since the fate of a cell is believed to be tied both to its lineage and the influence of nearby cells. “You want to see what the tissue looks like where the cell is, and then recover the lineage within that context. These are classic questions in developmental biology: How much of the developmental program is internal to the cell, and how much is cells responding to cues coming from other cells,” said Elowitz.

The MEMOIR method relies on the repair machinery’s tendency to delete or ‘collapse’ the barcode, which Elowitz and his team call a scratchpad. Each scratchpad has a known sequence adjacent to it, and represents a bit, much like the ‘0’ or ‘1’ value of a computer bit. The scratchpad exists in one of two stages depending on whether it has been untouched by genome editing or has experienced a deletion. The researchers then use fluorescent tags designed to recognize either the deleted or untouched stage, hybridize the tags to cells to reveal expressed mRNA, and then use microscopy to distinguish various colors. The color combinations can be distinguished on each barcoded scratchpad and within every cell in the intact tissue sample, thus revealing both the lineage history and the spatial context of the cell.

Recently, Elowitz and Cai teamed up with Carlos Lois’ lab at CalTech to use MEMOIR to examine cell lineage and fates in the fly brain⁵. They discovered that related cells were more similar in cell fate than unrelated cells at the same spatial distance, says Elowitz.

That isn’t too surprising, but it helps resolve the relative contributions of shared ancestry and spatial environment on the control of cell fate. The next step pushes the technique further, and could reveal mechanisms of development. Elowitz plans to examine the role of developmental signaling pathways like Hedgehog and



Barcodes plus transcriptomes in the scGESTALT approach yield information about cell lineage and identity. Reprinted with permission from Raj (2018)³, Springer Nature.

A MEMOIR scratchpad example. Through serial hybridizations, different barcodes receive fluorescent labels that add spatial information to temporal lineage. Reprinted with permission from Frieda (2017)⁴, Springer Nature.

bone morphogenetic protein (BMP) during development. These proteins appear during development in concentration gradients. Elowitz reasoned that cell lineage tracing and spatial information could lead to even more powerful insights into development if they could also record a cell's history of exposure to these signaling molecules and relate those signaling histories to the cell's individual fate decisions.

With that in mind, MEMOIR also has the capacity to place the genome editing system under control of one of those signaling molecules. In that case, the rate of memory editing will be directly proportional to the activity of a signaling pathway within the cell. Furthermore, multiple, independent editing systems could run simultaneously in an animal. "One could be running at a constant speed and giving you the lineage, and the other one could be conditional, increasing

or decreasing the rate at which edits accumulate, depending on the activity of the signaling pathway," said Elowitz.

As Elowitz and others seek to combine cell lineage tracing with imaging, other groups continue to push for greater granularity. Is the field mature enough to handle these combined approaches, or should exiting cell lineage tracing approaches be further optimized? "Do we want to build a better version of the technology, or apply it now to a biological question? That's a tension that the field is sort of struggling with," says Gagnon.

One way or another, the work initially pioneered by John Sulston will continue until scientists have a thorough understanding of development throughout the animal kingdom. "I'm really excited to see these new technologies that split in the past converge again into a common tool, where we're using molecular

recording and collecting molecular information but somehow retain the spatial context of where cells are relative to each other. It's so important for understanding how embryos actually work," says Gagnon. □

Jim Kling ✉

Freelance science writer, Bellingham, WA, USA.

✉e-mail: jkling@gmail.com

Published online: 23 July 2020

<https://doi.org/10.1038/s41684-020-0602-5>

References

1. Sulston, J. E., Schierenberg, E., White, J. G. & Thomas, J. N. *Dev. Biol.* **100**, 64–119 (1983).
2. Chan, M. M. et al. *Nature* **570**, 77–82 (2019).
3. Raj, B. et al. *Nat. Biotechnol.* **36**, 442–450 (2018).
4. Frieda, K. L. et al. *Nature* **541**, 107–111 (2017).
5. Chow, K. H. K. et al. *bioRxiv* <https://doi.org/10.1101/2020.02.21.958678> (2020).
6. Quinn, J. J. et al. *bioRxiv* <https://doi.org/10.1101/2020.04.16.045245> (2020).