

Stem cells: lineage tracing lets single cells talk about their past

Multipotent stem cells can become a variety of cell types. A flurry of new approaches enable lineage tracing at single-cell resolution.

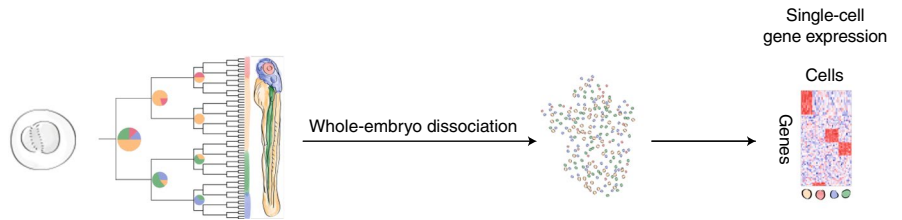
Vivien Marx

“So little by little I started looking by eye and drawing as I had for the larvae. At first it was hard, but I had the time to persist, and soon the structures became clearer in my mind. . . . Over the course of a year and a half it was finally done. We had the entire story of the worm’s cells from fertilised egg to adult.”

These are the words of developmental biologist John Sulston in his 2002 Nobel Prize lecture. Sulston, who passed away this year, was the founding director of the Wellcome Trust Sanger Institute¹. For 18 months and in two daily four-hour shifts, he peered through a microscope at developing nematode embryos and drew what happened to each cell: each cell division, migration, cell death.

Fate-mapping approaches before and after Sulston are varied. They include enzymes, genetically encoded fluorescent proteins and viral barcodes for labeling cells and their descendants². Now, a renaissance of lineage-tracing techniques is under way. With sequencing, labs can read out a large diversity of cellular barcodes and trace many cellular lineages in one experiment.

Some teams make heritable genomic changes with CRISPR–Cas9; others use Cre-mediated recombination. Researchers remind one another of Sulston’s feat, says Hans-Reimer Rodewald, an immunologist at the German Cancer Research Center (DKFZ) in Heidelberg who codeveloped a Cre-recombination-based approach to fate



New lineage-tracing techniques read out a diversity of cellular barcodes. In one experiment labs can trace many cellular lineages. (P. Olivares-Chauvet, Junker lab, MDC Berlin)

mapping. Sulston “was a huge inspiration to me developing this project, obviously,” says James Gagnon, developmental biologist at the University of Utah, who codeveloped single-cell GESTALT (scGESTALT).

There would be much to gain from lineage tracing at Sulston’s single-cell resolution. It can teach about cellular and organismal plasticity in vertebrates such as zebrafish and mice, says Jan Philipp Junker, a systems biologist at the Max Delbrück Center for Molecular Medicine in Berlin. Unlike nematodes, vertebrates assign cells to lineages variably, which allows their cells to react to change or injury so flexibly. Sulston’s manual approach to lineage tracing doesn’t scale well, but labs are exploring other routes.

CRISPR–Cas9 marks

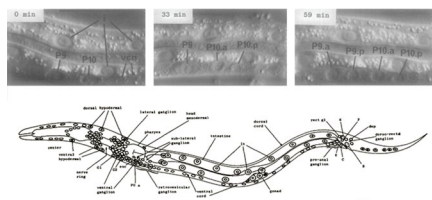
As a postdoctoral fellow in Alexander van Oudenaarden’s lab at the Hubrecht Institute in the Netherlands, Junker had been working on a lineage-tracing technique. Along came the CRISPR–Cas9-based lineage-tracing method GESTALT (genome editing of synthetic target arrays for lineage tracing) from Alexander Schier’s lab at Harvard Medical School. The paper “was a complete surprise to me,” says van Oudenaarden. It taught him to stay in touch with labs with similar interests.

Van Oudenaarden and Junker had been inducing CRISPR–Cas9-based ‘scars’ to mark the genome of zebrafish embryos with heritable, trackable labels.

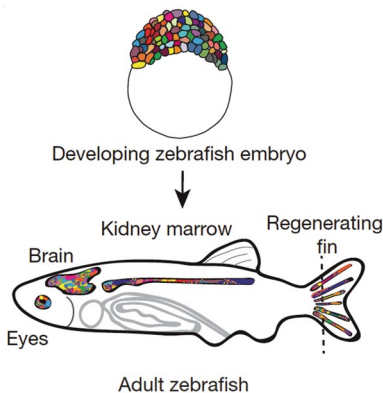
They uploaded their joint study to the preprint server bioRxiv, and then Junker left to start his own lab. The three teams stayed in touch, developing three different CRISPR–Cas9-based methods: ScarTrace, LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous sequences) and scGESTALT. These are but a few of the emerging lineage-tracing methods. Friendship has been an important factor, says Gagnon. He and Aaron McKenna, a co-author on the scGESTALT paper, pushed for cross-lab collaboration. Gagnon and McKenna have long been friends and are from nearby towns in Vermont. “If we all coordinate this and publish together, we all win, that’s my take on it,” says Gagnon.

With ScarTrace³, developed in the van Oudenaarden lab, a ‘scar’ sheds light on a cell’s history and the transcriptome helps with cell type identification. The scar is created during the repair of CRISPR–Cas9-based breaks. Cuts are dramatic events a cell quickly repairs. The repairs will differ: an extra T/T might be added or a C removed, says van Oudenaarden. “You really want sloppiness in the system,” he says, because this delivers unique, diverse barcodes. To obtain a set of scars in the same cell, the team uses eight copies of a histone–green fluorescent protein transgene.

To detect scars and measure gene expression in the same single cell, the team brought together two quite different protocols, says van Oudenaarden. Some lab members worked on reagents, others on



John Sulston mapped nematode cell lineage by hand and eye. (Adapted with permission from ref. ¹, Elsevier.)



With ScarTrace, researchers discovered that some immune cells do not arise in the marrow. (Adapted with permission from ref. ³.)

injections, buffer conditions, temperature, PCR timing or math. “That took us some time to really get that to work,” he says. “Now the protocol is there and now it works every day.” Using ScarTrace, the researchers found that progenitor cells in the zebrafish embryo slated to be specific brain or eye cells commit early, for example, to the right or left eye. That early commitment was also true for mesenchymal, epidermal and immune cells when the caudal fin was regenerated after a small injury. But immune-cell lineage tracing led to a surprise: a subpopulation of blood cells “are not born in the marrow, so that’s pretty interesting,” he says.

The possibilities with single-cell sequencing and Cas9 make for an “explosion of interest” in new ways to trace lineage, says van Oudenaarden. While he was at MIT, his lab was all about imaging. In his new lab in the Netherlands, single-cell sequencing was the focus. Microscopy gives labs spatial information “for free” that sequencing loses, he says. But the limited number of fluorophore colors limits the number of traceable lineages. The drawback with sequencing is “you don’t know what happened to the cell yesterday or the day before.” ScarTrace gives labs information they cannot get with imaging, he says, and the barcode diversity lets them infer thousands of lineages.

LINNAEUS, developed in the Junker lab, is a ScarTrace cousin that also applies heritable ‘scars’⁴. The method deploys a guide RNA to target 16–32 copies of a red fluorescent protein in the zebrafish line ZebraBOW-M. The readout is based on single-cell RNA-seq (scRNA-seq) and computational analysis of the barcodes, says Junker. The team reconstructed graph-based lineage trees for single cells from whole larvae as well as heart, brain and pancreas in the adult zebrafish.

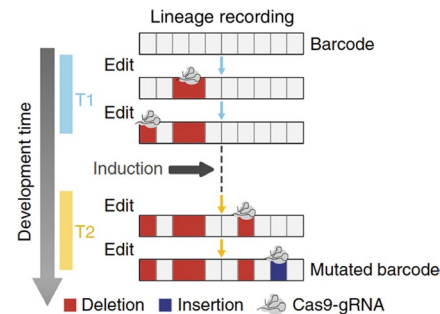
When RNA is collected from a single cell, says Junker, information always goes missing. He also noticed that some sequences are more commonly scarred than others. Identical barcodes might, for example, suggest falsely that a heart cell and a brain cell have a common ancestor, he says, “so you need to deal with this very carefully.”

The team established which scars are more probable and excluded them from the analysis. They built a filtering pipeline to address sequencing errors that occur with long barcodes even with deep sequencing. To add a focus on regeneration to lineage tracing, the lab collaborated with Nikolay Ninov’s team at the Technical University Dresden. Replenishing pancreatic beta cells is one area of research that can benefit from explorations of lineage and mechanisms of metabolic disease, says Junker. “I think it could work in mice,” he says of LINNAEUS. He and his team want to optimize aspects such as how to induce scarring not just early but also later in development, and to consider variants of Cas9.

Whole organisms, single cells

Lineage-tracing methods do not yet deliver complete lineage trees of whole organisms but they are moving in that direction, says Gagnon, codeveloper of scGESTALT. Now, labs can trace thousands, and potentially millions, of lineage branches in a single animal. One day, he says, one might build trees from many individuals to see the variable and fixed choices cells make during development.

GESTALT uses Cas9-induced mutations to generate genomic barcodes that are read out by sequencing. To get single-cell resolution, the team revamped the method so it was compatible with scRNA-seq, says Gagnon⁵. In both scGESTALT and LINNAEUS the method is a readout of scars and barcodes from the transcriptome, whereas ScarTrace reads scars from DNA and collects RNA from the same cell. Unlike both the van Oudenaarden and Junker labs, Gagnon and his colleagues generate some mutations early in development, and they



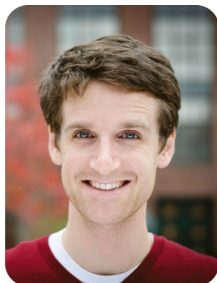
With scGESTALT, lineage tracing can take place at multiple points in development. (Adapted with permission from ref. ⁵.)

also deliver Cas9 as an inducible transgene. “In that way we can have a second round of editing happen later in development when many more cells are present,” says Gagnon. It enables lineage tracing of events that occur later in development.

The scientists used scGESTALT to trace lineages in the zebrafish brain on a single-cell level and to identify over 100 cell types. They injected the reagents—Cas9 protein and guide RNAs—into embryos at the one-cell stage and also added heat-shock-inducible Cas9 and guide RNAs for the edits later in development. They encapsulated cells from the zebrafish brain and sequenced the transcriptomes of around 66,000 brain cells and built a catalog of both progenitor and mature cell types.

Gagnon says that ScarTrace and LINNAEUS are easier systems than scGESTALT, which may be more flexible than the others, he says, because of the possibilities of multiple time points of editing. The lab will keep developing the method to try to enable continuous editing, barcoding and lineage tracing over the course of an organism’s lifetime. The team is optimizing the droplet-based recovery of cells, which captures the edited barcode in less than 30% of transcriptomes. Better cell and barcode recovery will deliver improved, more comprehensive lineage trees, he says. Perhaps more cellular history information can be embedded in scRNA transcriptomes so that they comprise all “that cell’s little experience,” he says. The scRNA-seq readout can then be about the cell’s present and its past.

The scGESTALT team chose to build on the data of Bushra Raj, a neuroscientist and postdoctoral fellow in the Schier lab who had generated an scRNAseq-based atlas of cell types in the zebrafish brain. It also built on a relationship with the lab of Jay Shendure at the University of Washington. They could have analyzed the heart, pancreas or blood because “the barcodes



Friendship has been an important factor in methods development, says James Gagnon. (T. Montague)

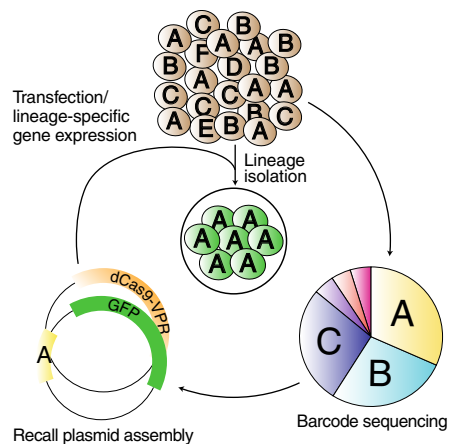
are present everywhere,” says Gagnon, a former postdoc in the Schier lab. In his view, scGESTALT can be used in a fly, a mouse or a human organoid.

It was challenging to advance GESTALT to scGESTALT, says Gagnon, to express the barcode as mRNA and then recover it in sequencing libraries. “I think there’s a lot of opportunity to improve how well that worked,” he says. They captured barcodes from a small percentage of the cells they sequenced and “I think that is true for the other approaches, too.” It would be hard, for example, to sequence every cell’s ‘full scars’. As groups advance techniques, he gets the sense that “we just cracked the door open.” Also on the to-do list: delivery, recovery, control and integration with other methods, such as imaging.

More barcoding

Separately, Amy Brock at the University of Texas at Austin has developed a Cas9-based lineage-tracing technology called COLBERT, for ‘control of lineages by barcode-enabled recombinant transcription’. It involves tagging a population of cells with a barcode gRNA that is regulated by a promoter. From a mixture of cells that contains populations with different barcoded gRNAs, the researchers can achieve lineage-specific gene expression and cell retrieval. For this, cells are transfected with a plasmid carrying a transcriptional activator variant of Cas9 and a ‘Recall’ plasmid that encodes the lineage barcode of interest that is upstream from the gene that is to be activated.

The team has recently combined the method with scRNA-seq to study shifts in clonal dynamics and cell states. “It’s a very powerful new level of information about cell



With COLBERT, scientists obtain lineage-specific gene expression and can retrieve cells for analysis. (Brock lab, U. Texas Austin)

populations,” says Brock. With COLBERT, codeveloped by graduate student Aziz Al’Khafaji, the scientists were motivated by the limitations of DNA barcoding, she says. Because quantification by sequencing is destructive, it precludes the ability to isolate cells and get lineage information, she says.

Samantha Morris at Washington University and her team have developed ‘CellTagging’, a single-cell lentiviral-based approach for lineage-tracing analysis of clonal dynamics. It involves scRNA-seq and combinatorial indexing of cells in order to read out lineage information. The team applied ‘CellTags’ to study lineage reprogramming from fibroblasts to induced endoderm progenitor cells. Their method revealed the distinct phases of this process as they studied originating cells, their resulting clones and the heterogeneity arising from individual cells in the population.

Reza Kalhor in the Harvard Medical School lab of George Church and colleagues have developed the ‘homing CRISPR’ system⁶ and adapted it for in vivo lineage tracing. Their mouse line, MARC1, has heritable homing guide RNAs (hgRNAs)⁷. The Cas9:hgRNA complex targets and retargets the locus that encodes the hgRNA itself to generate new barcodes.



CellTagging can be used for lineage-tracing analysis of clonal dynamics, says Samantha Morris. (T. Druley)

Cre-loxP-based barcoding

Recombinases can be used for lineage tracing. Rodewald and Thomas Höfer, also at the DKFZ, and their teams engineered an artificial DNA recombination site, *Polylox*, to generate endogenous barcodes at a defined genomic location, and devised ways to analyze them for lineage tracing⁸. They tag cells with unique barcodes and track these barcodes across the mouse’s lifetime. *Polylox* is made up of ten *loxP* sites in various arrangements that are separated by nine unique 178-base sequences from an *Arabidopsis* gene, and this ‘cassette’ is targeted to the *Rosa26* locus in mouse embryonic stem cells. For testing, the stem cells were transfected with a plasmid containing tamoxifen-inducible Cre.



Labs want to trace cell lineages in mice with CRISPR-Cas9 and Cre-based methods. (Redmond Durrell/Alamy Stock Photo)

Applying *Polylox*, they studied hematopoietic stem cell fates in vivo in the mouse. Their data revealed, among other findings, a tree-like lineage in blood. They also induced recombination in the brain, an ectodermally derived tissue; in muscle and spleen, which are mesodermal in origin; and in liver and lung, which are endodermal tissues.

The lab has shipped its mice to labs doing a range of experiments with their own Cre lines, so it’s too early for results, but the scientists eagerly await them. The *Polylox* team studied blood because of their interest in the immune system’s numerous lineages and cellular variety. *Polylox* is “clearly not limited to that,” says Rodewald. They detected recombination in every tested organ. The *Rosa* locus is readily targeted and has long been used for reporter expression. “The gene targeting frequencies at that locus are very high,” he says. “There’s a huge zoo of Cre lines in the world” that represent decades of research, with many validated drivers and high-fidelity constructs with low background. Essentially, he says, *Polylox* is a new reporter for this existing zoo of Cre lines.

“We had this urge to get this into mice as quickly as possible,” says Rodewald, who believes that CRISPR-Cas9 lineage tracing will be feasible in mice soon. One advantage to the inducible Cre system is how it can be applied to adult animals, such as for studying processes associated with aging. Such options, he says, will emerge with CRISPR-Cas.

The *Polylox* system labels specific cells for which Cre drivers exist, which is not yet true in CRISPR-Cas9 systems: they lack cell-type specificity, says Höfer. It’s attractive to be able to assess lineage in vivo over the course of an animal’s lifetime without perturbing the cells in their physiological environment, he says. Unlike the CRISPR-Cas community, the Cre field has plenty of “little on-switches” to make these barcodes

work in a controlled way, in a chosen organ and targeted location and at a chosen time.

The inserted DNA, says Rodewald, is “dead DNA” chosen to avoid expression. With expressed barcodes, detection may be easier but labs can end up with hundreds or thousands of molecules per cell. The Heidelberg team is working on linking transcriptomes with barcodes, something already possible with CRISPR–Cas9-based lineage-tracing methods.

The researchers plan to compare the transcriptomes of clones carrying the same barcodes, says Höfer, to then explore the differences shaping the fate of, for example, a given stem cell. With *Polylox*, around two million barcodes can be generated. In their experiment, the team believes they recovered around one-third of the generated barcodes. One issue is that if Cre activity is continuously stimulated, the barcodes are shuffled down. “If you let Cre go on for too long, you wind up with



The team brought two protocols together to detect ‘scars’ and measure gene expression in the same single cell, says Alexander van Oudenaarden. (Hubrecht Institute)

only 18,” says Höfer. “The recombinations cut part of the barcode out and you need to avoid that.” To do so, they let just four to six recombinations run, which leads to deletions or inversions that deliver diverse barcodes, he says. When they began the work, some in the lab worried that, for example, if there were many *loxP* sites Cre would recombine them all and they would disappear. That didn’t happen.

The two researchers met on a separate project that involved the first inducible Cre driver for use in hematopoietic stem cells.

The task needed mathematical analysis, so Rodewald reached out to Höfer. “We really have now two labs that are intertwined, and so they come up with predictions and we go back and test them,” says Rodewald.

“I think it’s an amazing idea,” says Junker, commenting on *Polylox*. He likes that the method is less dependent on endogenous cellular machinery than CRISPR–Cas9 and that the *Polylox* team can calculate barcode probabilities, which is not currently possible with CRISPR–Cas9. But ScarTrace, scGESTALT and LINNAEUS combine lineage tracing with scRNA-seq to provide cell-type identification, a resolution *Polylox* does not yet achieve, he says.

Van Oudenaarden also thinks highly of *Polylox* and says “what’s really important is that they did it in the mouse.” But the *Polylox* mouse is also quite complicated, he says. He wonders how long labs might need to optimize a mouse to, for example, study intestinal stem cells. Future experiments will show whether it is more straightforward to apply the Cas9 approach.

ScarTrace leverages the zebrafish’s quick development after delivering the reagents at the one-cell stage. A zebrafish embryo quickly develops into a ball of thousands of cells, which means thousands of cells are scarred to enable lineage tracing. In a mouse, several days can pass before an embryo reaches 100 cells. Delivering all reagents in a mouse’s one-cell stage will not lead to good options for lineage tracing. “I think the protocols will work for any cell type; of course I can’t prove this, but it would be my intuition,” says van Oudenaarden. To use ScarTrace in another system, a lab needs to ready the reagents and remember that Cas9 and guide RNAs remain active for around 8–10 hours before they are diluted out or fall apart.

Bright future

“I think we’re really at the beginning of this field,” says Junker. Labs might choose to combine new lineage-tracing techniques. One issue to address is targeting. “We need to find ways to increase the number of targets,” he says. The targeting sequence



Single-cell lineage tracing informs labs about cellular and organismal plasticity, says Jan Philipp Junker. (MDC Berlin)

determines how deep the lineage trees become. If a lab has four targets, at most four cell divisions can be scarred, possibly even fewer. Too little scarring hinders lineage tracing.

It’s too early to say which method will turn out to be a lab favorite, says Gagnon. “We all of course love our children.” He would like to find ways to barcode the genome at every cell division throughout the life of the entire animal to achieve whole-organism, whole-life lineage tracing. He is working on ways to layer information into barcodes to record cellular events of many kinds. Labs have developed techniques for recording lineage information with “wildly different” implementations, says Gagnon, “but we’re all trying to do that same thing, I think this just means this is the tip of the iceberg.” □

Vivien Marx

Technology editor for *Nature Methods*.
e-mail: v.marx@us.nature.com

Published online: 31 May 2018
<https://doi.org/10.1038/s41592-018-0016-0>

References

1. Sulston, J. E. & Horvitz, H. R. *Dev. Biol.* **56**, 110–156 (1977).
2. Woodworth, M. B., Girsakis, K. M. & Walsh, C. A. *Nat. Rev. Genet.* **4**, 230–244 (2017).
3. Alemany, A. et al. *Nature* **556**, 108–112 (2018).
4. Spanjaard, B. et al. *Nat. Biotechnol.* **36**, 469–473 (2018).
5. Raj, B. et al. *Nat. Biotechnol.* **36**, 442–450 (2018).
6. Kalhor, R., Mali, P. & Church, G. M. *Nat. Methods* **14**, 195–200 (2017).
7. Kalhor, R. et al. *bioRxiv* Preprint at <https://www.biorxiv.org/content/early/2018/03/12/280289> (2018).
8. Pei, W. et al. *Nature* **548**, 456–460 (2017).