

GENOMICS

Cas9, the cellular genealogist

Genome editing marks cells in developing fish for lineage studies.

What began as informal conversations about science at meet-ups and weddings eventually led long-time friends Aaron McKenna and James Gagnon into a productive scientific collaboration. With co-first author Greg Findlay, the scientists ultimately brought together their lab heads, Jay Shendure at the University of Washington and Alex Schier at Harvard University, for regular video conference sessions. The groups focused their backgrounds in genomics and developmental biology on using Cas9, a genome-targeting nuclease, to study the lineage relationships between cells of the zebrafish, a trusted model of vertebrate biology.

Methods for lineage marking have a long history. Over a century ago, plant biologists traced the descendants of enlarged polyploid cells through development. Scientists then used irradiation, dye injection and transplantation to mark individual cells, followed by molecular approaches such as reporter gene expression and viral barcoding. Shendure's first project as a graduate student in the lab of George Church was to generate a binary code to mark cells. "I spent the first six months ... trying to build a massive array of recombinase targets, a series of flipping switches," he recalls.

Many of these methods are laborious or limited to small sublineages, whereas complex organisms require permanent and highly diverse markers. Somatic mutations in simple sequence repeats and even whole genomes fit the bill, but they "have the downside of being very diffuse and expensive to collect," says Shendure. "In many ways, it's a problem of information content."

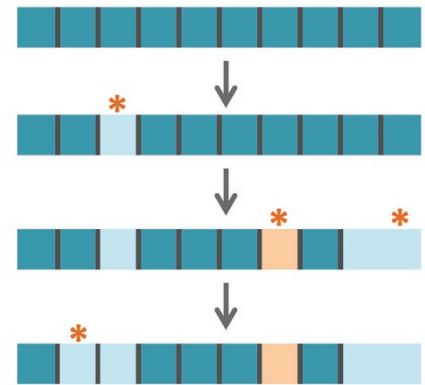
The researchers solved the problem by inserting a series of synthetic Cas9 target sites arranged as a compact barcode in the zebrafish genome. When injected into a single-cell embryo along with single guide RNAs (sgRNAs), Cas9 can cut target sites, generating insertions or deletions during repair. These unique scars accumulate dur-

ing development and are passed to a cell's offspring. Sequencing barcodes from an animal or organ makes it possible in many cases to infer which cells originated from a common progenitor on the basis of shared scars. The method was aptly named GESTALT, for 'genome editing of synthetic target arrays for lineage tracing'.

Using the system, the team detected up to a few thousand unique barcodes at various embryonic stages or from adult organs. Some barcodes were shared by approximately half of all sampled cells, implying that they were generated in the two-cell embryo, and the researchers estimated that the injected sgRNAs were exhausted just before or during gastrulation. Their results revealed that most organs are populated by descendants from a small number of progenitors; one blood sample appeared to mostly originate from just five cells. Very few studies have been able to connect the early embryo to adult organs—"I think this is a step forward for the fish," says Schier.

Inferring histories from a snapshot is not entirely straightforward. Cells migrate or die and their barcodes are lost. It could be that a few progenitors founded stable lineages, neutral drift caused a small subset to take over the stem cell niche, or a large number of clones competed to eventually produce a few winners. "That's a super exciting question that we don't yet have the answer for," says Schier. By regulating the timing of Cas9 or sgRNA expression, GESTALT could be used to capture later lineage relationships and address these questions, something that the collaborators are pursuing.

Most gene editing research focuses on making edits efficient and homogeneous, but GESTALT relies on the stochastic nature of cutting and repair. "It could have been a disaster, where you just lose the whole barcode at the one-cell stage and you don't generate any diversity," says Schier. There are many ways to tune barcode editing. The researchers successfully tested two classes of barcode, one with a perfect target site



Cas9-mediated editing (asterisks) of a lineage barcode generates a permanent and cumulative record of insertions (orange bars) and deletions (light blue bars) over time, enabling cellular lineage reconstruction. Adapted from McKenna *et al.* (2016) with permission from AAAS.

and multiple lower-efficiency off-target sites, which requires only a single sgRNA, and another with multiple targets for different sgRNAs. "There's a huge opportunity here just around engineering of the system, of guides, targets and levels of Cas9," says Shendure.

Ideally, the system would evolve to generate unique barcodes as frequently as once per division. The ultimate goal is to combine lineage with spatial and sequencing-based information about cell type and state, which can be achieved with single-cell sequencing. "In the long run what we want is X, Y, Z, lineage and molecular profiles, whether that be RNA, epigenetic or both," says Shendure.

A number of related strategies are being reported in preprints, pointing to building interest in Cas9 as a lineage tool. "There was something almost romantic about the way this collaboration worked out," says Schier. "It was great fun."

Tal Nawy

RESEARCH PAPERS

McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* <http://dx.doi.org/10.1126/science.aaf7907> (2016).