Credit: Sean Prior/Alamy

⇥ DEVELOPMENT

Recording embryogenesis

" a CRISPR–Cas9-based 'molecular recorder' that can … report on both cellular state and cell lineage "

CRISPR–Cas9-based technologies have been used to give cells a 'genomic barcode', the mutation of which can be followed through successive cell divisions to trace cell lineage during development. However, these techniques have typically only allowed the genomic barcode to be mutated a few times, restricting the generation of traceable barcode diversity. Chan et al. now describe a CRISPR–Cas9-based 'molecular recorder' that can generate genomic barcode diversity in mice from fertilization through to adulthood and report on both cellular state and cell lineage. They use this recorder to assemble cell-fate maps from fertilization to gastrulation in mice.

The molecular recorder uses Cas9 to generate double-stranded DNA breaks, the repair of which results in heritable, recordable insertions or deletions (indels). Specifically, a DNA 'target-site' cassette, containing three Cas9 'cut sites' and an 'integration barcode', is embedded in the 3′ untranslated region of a constitutively transcribed fluorescent protein. Multiple copies of the target-site cassette are integrated into cells, which also express a cassette encoding three single-guide RNAs (sgRNAs). sgRNAs direct Cas9 to cut sites where it generates double-stranded breaks, the repair of which produces heritable indels; these indels can be identified and their lineage traced using single-cell RNA sequencing (scRNA-seq).

After confirming in cell culture that their recorder could identify different cell populations based on their indels (that is, based on their 'single-cell barcode'), the authors applied their system to mice. Mouse oocytes harbouring target-site and sgRNA cassettes were fertilized with Cas9–GFP-encoding sperm, and cells were isolated from embryos at around embryonic day (E)8.5 or E9.5. Whole transcriptome scRNA-seq data were collected for 7,364–22,264 cells from 7 embryos, from which 167–2,461 unique lineages were identified. By including single or dual mismatches in sgRNAs, the authors had generated guides with differing mutation rates. They observed that sgRNAs with the fastest mutation rates created the highest proportion of cells with identical indels, suggesting that mutations arise early in development and reduce indel diversity. Importantly, as most embryos contained cells with unmodified cut sites, the lineage tracer should record information post-E9.5.

The authors then determined the cellular phenotype of cells from six of the lineage-traced embryos by comparing their scRNA-seq profile to known gene annotations relating to wild-type mouse gastrulation (that is, E6.5–E8.5). The age of each embryo was also determined by comparing their tissue composition with wild-type references. To combine this information on cellular state with their lineage tracing data, the authors

reconstructed phylogenetic trees, in which each branch represented an indel, and overlaid these trees with their data on cellular phenotype. In general, they observed that fewer cell types are represented in the tree over time, effectively capturing functional restriction as it occurs over development. Furthermore, a smaller lineage distance between cells correlated with cells having a more closely related transcriptional profile, as has been previously suggested by scRNA-seq data alone.

As well as identifying known tissue relationships during development, analysis of the phylogenetic tree unveiled some surprises: for example, the extra-embryonic endoderm and the embryonic endoderm displayed an unexpectedly close ancestry, despite originating from the hypoblast and embryo-restricted epiblast, respectively. This observation may be explained by a subpopulation of cells that transcriptionally resemble embryonic endoderm but are placed within extra-embryonic branches by lineage analysis. Thus, the recorder can identify cases of convergent transcriptional regulation.

Finally, as each node within the tree represents a unique lineage that stems from a single reconstructed progenitor cell, the authors could estimate that the founding number of progenitors in early development is 1–6 totipotent cells, 10–20 early pluripotent progenitors and 18–51 late pluripotent progenitors.

Coupled with the fact that this system produced mice that maintained indel generation into adulthood, the data discussed suggest that this molecular recorder is a powerful tool for studying mammalian development.

Katharine H. Wrighton

ORIGINAL ARTICLE Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* https://doi.org/10.1038/s41586-019-1184-5 (2019)
FURTHER READING Woodworth, M. B., Girskis, K. M. & Walsh, C. A. Building a lineage from single cells: genetic techniques for cell lineage tracking. *Nat. Rev. Genet.* **18**, 230–244 (2017)