Epigenetics

https://doi.org/10.1038/s41576-026-00958-y

Reading cell division histories from the methylome

Every cell in the human body carries a record of its own history in the sequence of divisions that gave rise to it. This record contains information on how tissues are maintained, how cancers arise, and how ageing unfolds at the cellular level. Yet, in living humans, reading this record is extraordinarily difficult. The experimental strategies used to delineate cell lineages in model organisms, including chimeric mice, genetic barcodes and transgenic reporters are simply unavailable in humans. For decades, researchers could only identify cell histories indirectly from the distribution of rare somatic mutations. A more accessible molecular clock was needed.

DNA methylation offered an unexpected answer. The methylation state at CpG dinucleotides is accurately inherited across cell divisions, but also subject to stochastic errors, so-called epimutations, at a rate of several orders of magnitude higher than somatic DNA mutations. These errors accumulate gradually over a lifetime and are passed to daughter cells, causing methylation patterns to diverge progressively between lineages. In effect, the methylome keeps a probabilistic ledger of cell division history.

> "the methylome keeps a probabilistic ledger of cell division history"

The first systematic effort to harness this principle in human tissue was published by Yatabe et al. in 2001. Working with fresh human colectomy samples from ten patients, the authors isolated individual colonic crypts, each of which is maintained by a small pool of stem cells at its base. Using bisulfite sequencing, they profiled the methylation state of neutral CpG-containing loci within individually cloned DNA molecules from each crypt. The design was deliberately minimal: just three genomic loci (*MYOD1*, *CSX* and *BGN*), each bearing fewer than ten CpG sites. Yet the signals proved surprisingly informative.

Within a single crypt, methylation patterns across molecules were diverse but clearly related, consistent with descent from a common ancestor. By contrast, between crypts from the same patient, patterns were substantially more distinct, as though each crypt had accumulated its own idiosyncratic epigenetic signature independently over decades. This contrast — low within-crypt diversity, but high between-crypt diversity — identified the stem cell compartment as the site of methylation drift, rather than the rapidly cycling transit-amplifying cells that populate most of the crypt. Notably, upper and lower portions of bisected crypts bore similar methylation tags, confirming a stem cell origin for the observed patterns.

To interpret these observations, the authors developed stochastic simulations of crypt stem cell dynamics to contrast two competing models. The first model, assuming immortal stem cells and asymmetric division, predicts a steady accumulation of unique methylation tags. The second model involves a shared niche in which stem cells occasionally replace one another through symmetric division — a process of neutral drift — leading to periodic bottlenecks as lineages trace back to a single most recent common ancestor. Observed variance in tag diversity across crypts was inconsistent with the immortal model but aligned with niche-model predictions. Best-fit simulations suggested approximately 64 stem cells per crypt, with bottlenecks recurring every 8 years. This provided quantitative evidence in human tissue for neutral drift and clonal succession, dynamics later confirmed in mice by genetic lineage tracing.

The elegance of this 2001 study lay as much in its conceptual logic as in its technique. But its scope was limited: three loci could capture only a fraction of the available epigenetic signal, and the approach was tailored to a histologically uniform tissue, in which individual crypts could be isolated and examined in parallel. Extending methylation-based lineage tracing to heterogeneous tissues remained a formidable challenge, as cells of different types carry distinct methylation patterns that could dominate and obscure the underlying stochastic lineage signal.

This is precisely the barrier that my colleagues and I recently addressed with MethylTree. By coupling single-cell whole-genome bisulfite sequencing with a computational framework that corrects for measurement noise in sparse data and removes cell-type-specific methylation signals, we extracted lineage information from millions of CpG sites simultaneously, across diverse cell types and developmental stages. The key insight was that raw cell–cell methylation similarity reflects the superposition of cell-type identity and lineage history; subtracting the former reveals the latter. Benchmarked against ground-truth genetic lineage barcodes in mouse and human blood, MethylTree achieved near-perfect lineage reconstruction accuracy. Applied to human embryos, it identified stochastic early fate commitment towards the inner cell mass or trophectoderm at the 4-cell stage. In native mouse blood, it estimated approximately 250 distinct haematopoietic stem cell clones.

What began as a handful of methylation tags in isolated colon crypts has grown, two decades later, into a general approach for non-invasive, single-cell-resolution lineage tracing across the body —compatible with simultaneous transcriptome and chromatin profiling.

Shou-Wen Wang
Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China.
e-mail: wangshouwen@westlake.edu.cn

Competing interests
S.-W.W. is named inventor on a patent application for MethylTree (PCT/CN2024/095497).

Original articles: Yatabe, Y., Tavaré, S. & Shibata, D. Investigating stem cells in human colon by using methylation patterns. *Proc. Natl Acad. Sci. USA* **98**, 10839–10844 (2001); Chen, M. et al. High-resolution, noninvasive single-cell lineage tracing in mice and humans based on DNA methylation epimutations. *Nat. Methods* **22**, 488–498 (2025)