MethylTree: exploring epimutations for accurate and non-invasive lineage tracing

Non-invasive lineage tracing in humans relies on rare somatic mutations, which have limited throughput and temporal resolution. We developed a computational method, 'MethylTree', which uses epimutations on DNA methylation to accurately infer lineages across cell types, developmental stages and species, providing a superior alternative for non-invasive lineage tracing in humans and other organisms.

This is a summary of:
Chen, M. et al. High-resolution, noninvasive single-cell lineage tracing in mice and humans based on DNA methylation epimutations. *Nat. Methods* https://doi.org/10.1038/s41592-024-02567-1 (2025).

Published online: 16 January 2025

The problem

Cell lineages (the history of cell divisions) within an individual encode the fundamental principles of development and disease. Many powerful approaches have been developed to enable scalable single-cell lineage tracing in vivo[1], including our DARLIN mouse with approximately 10^{18} distinct lineage barcodes and superior barcode editing and capture[2]. However, these approaches inevitably require invasive genetic manipulation that only works in certain model organisms. Non-invasive lineage tracing relies on endogenous lineage markers, such as somatic mutations[3]. Detecting these rare mutations is technically challenging, costs about US $200 per cell, and produces a lineage tree with limited temporal resolution and no cell state information. Mitochondrial DNA mutations are more frequent than somatic mutations, but their use in lineage tracing is controversial owing to the complex inheritance pattern over cell division.

Epimutations on DNA methylation are a promising alternative for non-invasive lineage tracing (Fig. 1a). DNA methylation in mammals usually occurs at the cytosine residue in CpG dinucleotides. The epimutation rate in humans is estimated to be ~0.001 per CpG site per division[4], leading to ~30,000 epimutations per division, which remain stable over ~1,000 cell divisions. However, single-cell DNA methylation measurements are highly sparse, with only ~5% of the genome covered. Moreover, methylation is thought to change dynamically during differentiation to produce cell-type-specific methylation patterns. Global modulation of the DNA methylation level during early embryo development might also erase shared epimutations between sister cells[5]. These are shared challenges across mammals.

The solution

We developed MethylTree, a computational method for non-invasive lineage tracing based on epimutations on DNA methylation. The 5% genomic coverage of DNA methylation per cell implies 0.25% CpG sites that are jointly observed in two cells. We used these shared CpG sites to compute a cell–cell similarity matrix, which defines their relationship. This approach was used because the 'stochastic' epimutations that are crucial for lineage inference would be averaged out when using conventional methylation data analysis approaches, such as averaging methylation values over large genomic bins (for example, 100 kbp) or imputing missing methylation values. We also developed a generic approach to correct the heterogeneous measurement noises on the cell–cell similarity matrix. To remove the cell-type-specific methylation signals, we developed a mathematical approach to regress out such signals on the similarity matrix. We adopted this strategy instead of the more intuitive, but technically challenging, approach that selects only the lineage-specific CpG sites to define the similarity matrix. The resulting cell–cell similarity matrix was used to infer lineage histories (Fig.1b).

We systematically benchmarked our approach using in silico simulation, serial clonal expansion of cell lines, in vitro differentiation of mouse and human blood, in vitro early human embryo development, and in vivo expansion of haematopoietic stem cells (HSCs), each with ground-truth lineage labels that are known or measured experimentally. MethylTree accurately reconstructed cell lineages across all these contexts with almost 100% accuracy, with a temporal resolution that could separate individual cell divisions. Applying MethylTree, we identified early fate bias at the four-cell stage in human embryos and estimated approximately 250 HSC clones in mouse blood.

The implications

We have demonstrated that epimutations on DNA methylation can be used for high-resolution, non-invasive lineage tracing. Joint single-cell DNA methylation and transcriptomic profiling enables single-cell multi-omic lineage tracing, with the lineage tree inferred from the DNA methylome. MethylTree, which is publicly available, should greatly benefit the broad research community interested in lineage tracing in humans or other organisms. Single-cell DNA methylome data are being actively generated in many laboratories across the world to study epigenetic regulation and its role in aging – our study opens a new avenue to exploit this dataset. Finally, MethylTree also represents a new dimension-reduction approach to deal with highly sparse data.

Further developments are needed to make epimutation-based lineage tracing more scalable and robust. A method for affordable, scalable profiling of the single-cell DNA methylome (about US $5 per cell) has recently been reported[6]; however, more targeted profiling could further reduce the cost and increase the throughput. How epimutations accumulate over time and across tissues remains to be investigated. Finally, further investigation is needed to identify genomic regions that will enable more robust lineage inference.

Shou-Wen Wang
Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China.

EXPERT OPINION

"This paper develops an approach for lineage tracing in single cells using CpG methylation patterns and their similarity. The core idea is based on epimutations that provide lineage information but are also confounded by cell-type differences. The authors utilize Pearson correlation between cell-by-region methylation matrices of two cells to compute cell-by-cell similarity. They then remove noise and cell-type effects from these similarities and then infer a phylogenetic tree. No doubt this work is of significance, and I appreciate the many different real datasets the tool has been demonstrated on." **Ferhat Ay, La Jolla Institute for Immunology, La Jolla, CA, USA.**

FIGURE

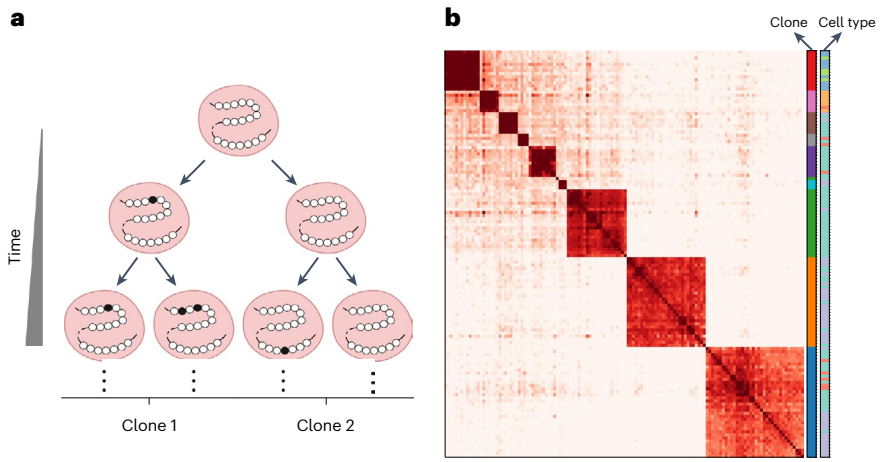

Fig. 1 | Epimutation-based lineage inference. a, Schematic of epimutation accumulation over cell division. These stable and rich epimutations lay the foundation for lineage inference from the DNA methylome. **b**, Demonstration of accurate lineage inference with MethylTree. Here, $CD34^+$ cells were sorted from human umbilical cord blood, labeled with a lineage and RNA recovery (LARRY) barcode via lentiviral infection, differentiated in vitro for 13 days, and profiled to generate the DNA methylome, transcriptome and actual clonal barcode for each cell. Methylation-based cell–cell similarity correctly reconstructed all the clones defined by the lineage barcode, despite the presence of seven cell types with distinct transcriptomic profiles. Panel a was created in BioRender, http://BioRender.com/m32z330.

BEHIND THE PAPER

During our previous work on the DARLIN mouse model, my collaborator Li Li developed Camellia-sequencing to jointly profile the DNA methylome, transcriptome, chromatin accessibility and cell lineages in single-cells[2]. Initially, we wanted to reveal epigenetic fate regulators for HSCs. However, after analyzing this dataset for several months, I realized that the most pressing challenge was to devise a computational approach that can extract biologically meaningful signals from this sparse dataset. This approach resulted in the discovery of strong clonal memory on the DNA methylome, but not on the transcriptome or chromatin accessibility, which led me to speculate that the DNA methylome might contain enough information for lineage tracing. Realizing its great importance, I raced to develop MethylTree and supervised my students to perform the validation experiments within just a year, with help from Li Li. **S.-W. Wang.**

REFERENCES

1. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
   This review article highlights advances in lineage tracing as well as caveats for data interpretation.
2. Li, L. et al. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183–5199.e22 (2023).
   This article presents DARLIN, the advanced lineage tracing mouse model, and reports strong clonal memory on the DNA methylome.
3. Abyzov, A. & Vaccarino, F. M. Cell lineage tracing and cellular diversity in humans.*Annu. Rev. Genomics Hum. Genet.* **21**, 101–116 (2020).
   This review article summarizes existing approaches for non-invasive lineage tracing in humans.
4. Ushijima, T. et al. Fidelity of the methylation pattern and its variation in the genome. *Genome Res.* **13**, 868–874 (2003).
   This article presents an estimation of epimutation rate.
5. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
   This review article discusses functional methylation changes in cells.
6. Cao, Y. et al. Single-cell bisulfite-free 5mC and 5hmC sequencing with high sensitivity and scalability. *Proc. Natl Acad. Sci. U.S.A.* **120**, e2310367120 (2023).
   This article reports a scalable and affordable method for profiling the single-cell DNA methylome.

FROM THE EDITOR

"Lineage tracing has broad applications in developmental biology and disease research, yet most experimental methods rely on genetic manipulation. Non-invasive lineage tracing in humans, however, remains limited. This work presents MethylTree, a method that leverages somatic DNA methylation changes to trace cell lineages non-invasively across a range of biological contexts." **Lei Tang, Senior Editor, *Nature Methods*.**