CoSpar identifies early cell fate biases from single cell transcriptomic and lineage information

Shou-Wen Wang*[,1], Michael J Herriges[2,3], Kilian Hurley[4,5], Darrell N. Kotton[2,3], Allon M. Klein*[,1]

1 Department of Systems Biology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA
2 Center for Regenerative Medicine of Boston University and Boston Medical Center, Boston, MA 02118, USA
3 The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA
4 Department of Medicine, Royal College of Surgeons in Ireland, Education and Research Centre, Beaumont Hospital, Dublin, Ireland
5 Tissue Engineering Research Group, Royal College of Surgeons in Ireland, Dublin, Ireland

*Email: shouwen_wang@hms.harvard.edu (S.W.W.); allon_klein@hms.harvard.edu (A.M.K.)

Abstract:

In tissue development, regeneration, and disease, cells differentiate into distinct, reproducible phenotypes. A ubiquitous challenge in studying these processes is to order events occurring during differentiation[1–3], and to identify events that drive cells towards one phenotype or another. This challenge is common to understanding mechanisms in embryo development, stem cell self-renewal, cancer cell drug resistance, and tissue metaplasia[1–3].

At least two observational strategies help to order cellular events. Single-cell genome-wide profiling – such as by single-cell RNA sequencing (scRNA-seq) – offers a universal and scalable approach to observing dynamic states by densely sampling cells at different stages[3–10]. However, scRNA-seq alone does not identify which early differences between cells drive or correlate with fate[2,11–13]. Conversely, lineage tracing offers a complementary family of methods that can clarify long-term dynamic relationships across multiple cell cycles. To carry out lineage tracing, individual cells are labeled at an early time point[1–3]. The state of their clonal progeny is analyzed at one or more later time points (Fig. 1**a**).

Recently, a number of efforts from us and others have integrated lineage-tracing with single-cell RNA sequencing (hereafter LT-scSeq) using unique, heritable, and expressed DNA barcodes[2,12,14–19]. These technologies identify cells that share a common ancestor and define their genomic state in an unbiased manner. LT-scSeq experiments have been used to successfully identify when fate decisions occur[12,15], novel markers for stem cells[18], and pathways which control cell fate choice[15,18]. The simplest of these methods labels cells at one time point[12] (Fig. 1**b**); more complex methods allow the accumulation of barcodes over successive cell divisions to reveal the substructure of clones[2,12,14–20] (Fig. 1**c**).

Emerging LT-scSeq methods have been successful at revealing regulators of cell fate[15,18] and the fate potential of early progenitors[12,15], but they also present challenges that may limit their utility in practice. At least five technical and biological challenges affect experimental design and interpretation (Fig. 1**f**): stochastic differentiation and variable expansion of clones[21] (Fig. 1**f-i**); cell loss during analysis (Fig. 1**f-ii**); barcode homoplasy wherein cells acquire the same barcode despite not having a lineage relationship[2] (Fig. 1**f-iii**); access to clones only at a single time point[22,23] (Fig. 1**f-iv**); and errors in determining the state of clonal progenitors due to a lag time between labeling cells and the first sampling ('clonal dispersion', Fig. 1**f-v**). Addressing these problems should greatly simplify the design and interpretation of LT-scSeq assays and put them in the hands of a wider research community.

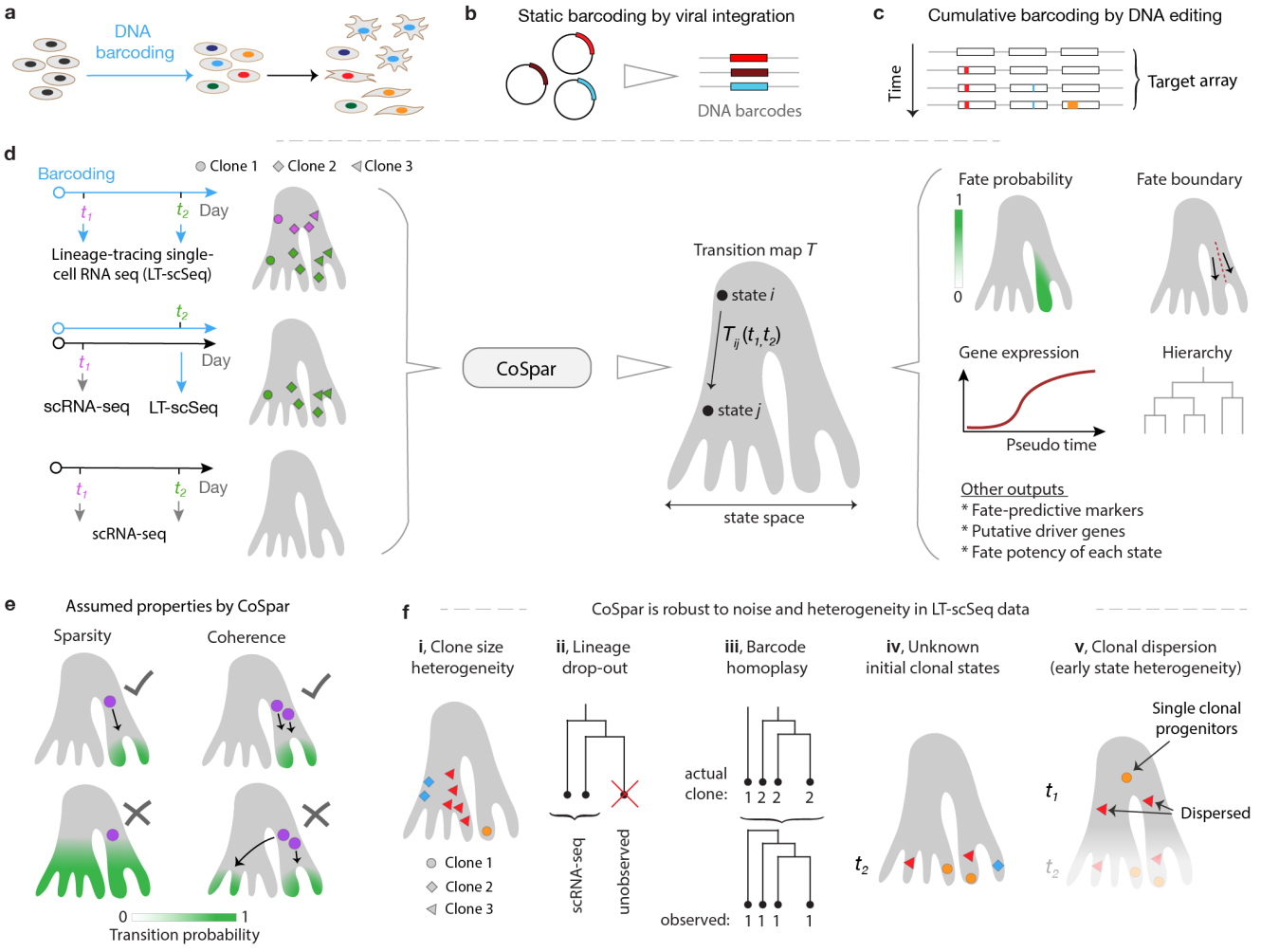

Fig. 1. Integrative analysis of lineage tracing and transcriptome data.

Here, we advance on recent efforts[24,25] to develop robust, computationally-efficient and generalizable approaches to analyze LT-scSeq experiments. We begin with a model of clonal dynamics in which cells divide, differentiate, or are lost from the sampled tissue in a stochastic manner, with rates that are state-dependent. We use this model to learn from the

data the fraction of progeny of cells, initially in one state, which are found to occupy a second state after some time interval (Fig. 1**d**). Our approach captures differentiation bias and fate hierarchies, and can reveal genes whose early expression is predictive of future fate choice.

In this computational approach, we develop coherent, sparse optimization (CoSpar) to infer cell dynamics from single-cell transcriptomics integrated with lineage tracing. Built on assumptions of coherence and sparsity of transition maps, CoSpar is robust to severe down-sampling and dispersion of lineage data, which enables simpler experimental designs and requires less calibration. In datasets representing hematopoiesis, reprogramming, and directed differentiation, CoSpar identifies early fate biases not previously detected, predicting transcription factors and receptors implicated in fate choice. Documentation and detailed examples for common experimental designs are available at https://cospar.readthedocs.io/. This work is recently published at Nature Biotechnology.

References

1. 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).
2. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* (2020) doi:10.1038/s41576-020-0223-2.
3. Kester, L. & van Oudenaarden, A. Single-Cell Transcriptomics Meets Lineage Tracing. *Cell Stem Cell* **23**, 166–179 (2018).
4. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).
5. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0591-3.
6. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
7. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
8. Qiu, X. *et al.* Mapping Vector Field of Single Cells. 696724 (2019) doi:10.1101/696724.
9. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928–943.e22 (2019).
10. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
11. Tritschler, S. *et al.* Concepts and limitations for learning developmental trajectories from single cell genomics. *Development* **146**, (2019).
12. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science* **367**, (2020).
13. Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2467–E2476 (2018).
14. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).

15. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
16. Bowling, S. *et al.* An engineered CRISPR/Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. *Cell* 797597 (2019) doi:10.1101/797597.
17. Chan, M. M. *et al.* Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
18. Rodriguez-Fraticelli, A. E. *et al.* Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. *Nature* (2020) doi:10.1038/s41586-020-2503-6.
19. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
20. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
21. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822–825 (2010).
22. Hurley, K. *et al.* Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell* (2020) doi:10.1016/j.stem.2019.12.009.
23. Yao, Z. *et al.* A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. *Cell Stem Cell* **20**, 120–134 (2017).
24. Prasad, N., Yang, K. & Uhler, C. Optimal Transport using GANs for Lineage Tracing. *arXiv [cs.LG]* (2020).
25. Forrow, A. & Schiebinger, G. LineageOT is a unified framework for lineage tracing and trajectory inference. *Nat. Commun.* **12**, 4940 (2021).