



HAL
open science

A computational tool suite to facilitate single cell lineage tracing analyses

Joshua Waterfall, Adil Midoun, Leïla Perié

► **To cite this version:**

Joshua Waterfall, Adil Midoun, Leïla Perié. A computational tool suite to facilitate single cell lineage tracing analyses. Cell Reports Methods, In press, 10.1016/j.crmeth.2024.100780 . hal-04579519

HAL Id: hal-04579519

<https://hal.science/hal-04579519>

Submitted on 17 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A computational tool suite to facilitate single cell lineage tracing analyses

Joshua J Waterfall^{1,2,†}, Adil Midoun³, Leïla Perié^{3,†}

1. Institut Curie, Université PSL, INSERM U830, Paris, France
2. Institut Curie, Université PSL, Department of Translational Research, Paris, France
3. Institut Curie, Université PSL, Sorbonne Université, CNRS UMR168, Laboratoire Physique de la Cellule et du Cancer, Paris, France.

† co-corresponding authors joshua.waterfall@curie.fr, leila.perie@curie.fr

Abstract

Tracking the lineage relationships of cell populations is of increasing interest in diverse biological contexts. In this issue of *Cell Reports Methods*, Holze et al present a suite of computational tools to facilitate such analyses and encourage their broader application.

Main text

Distinguishing how shared inheritance shapes cellular phenotypes is a central question in developmental biology, cancer heterogeneity, immunity, and many other biological fields. A powerful example of the utility of such lineage tracing experiments is the mapping of the developmental fate of every cell in the *Caenorhabditis elegans* embryo by direct observation via light microscopy¹. While classic techniques for eg lineage tracing via Cre-mediated reporter expression or T cell receptor or B cell receptor diversification in lymphocytes have been used for decades, recent years have seen an explosion in new approaches^{2,3}. These newer approaches include the use of “endogenous” barcodes generated by somatic mutations in nuclear and mitochondrial genomes as well as a wide variety of exogenous barcoding methods leveraging different genetic engineering approaches such as CRISPR/Cas9, lentiviruses, terminal deoxynucleotidyl transferase (TdT), RAG 1 and 2 enzymes, and more. Importantly, these exogenous barcoding technologies work both in human cell culture/ex vivo systems as well as in situ for a variety of model organisms including mouse and zebrafish.

Coupled with simultaneous high throughput molecular phenotyping of the cells, such as through single cell RNA sequencing, these techniques are powerful molecular “microscopes” that can link the current state of a cell with its ancestral history³. However, the introduction of PCR and sequencing errors and the unique aspects of each system pose substantial technical challenges for their analysis⁴. Currently, researchers interested in adding lineage information to their favorite model system have no shortage of experimental protocols to choose from, however there is a comparative lack of broadly applicable, flexible analysis tools capable of analyzing data from diverse protocols⁴. Publishing in *Cell Reports Methods*, Holze et al.⁵ have developed a tool suite that complements existing tools⁶⁻⁹ and could greatly enhance the widespread adoption of such techniques by non-specialist teams. It also helps to address the crucial need for clearer guidelines and standardized metrics to evaluate new datasets, improving the overall reproducibility of single cell lineage tracing results⁴.

The BARTab and bartools analysis suite reported by Holze et al.⁵ is a significant step forward by simplifying single cell lineage tracing results analysis for a broad community. Implemented both as a Nextflow pipeline and R package, these tools allow flexible analyses of diverse barcoding approaches

with a particular focus on quality control, barcode identification and quantification, multiple plotting options as well as statistical evaluation of barcode composition and diversity. The first tool, BARTab, is dedicated to early processing steps of fastq or bam files to identify and quantify barcodes. This can be applied both to bulk sequencing protocols as well as to single cell protocols. Together with the CellBarcode R tool recently released⁶, these are the first tools incorporating single cell input from a diversity of wet lab protocols with high flexibility in both the design and lengths of the barcodes. Therefore they could be broadly taken up by the community for standard analyses of diverse experimental approaches.

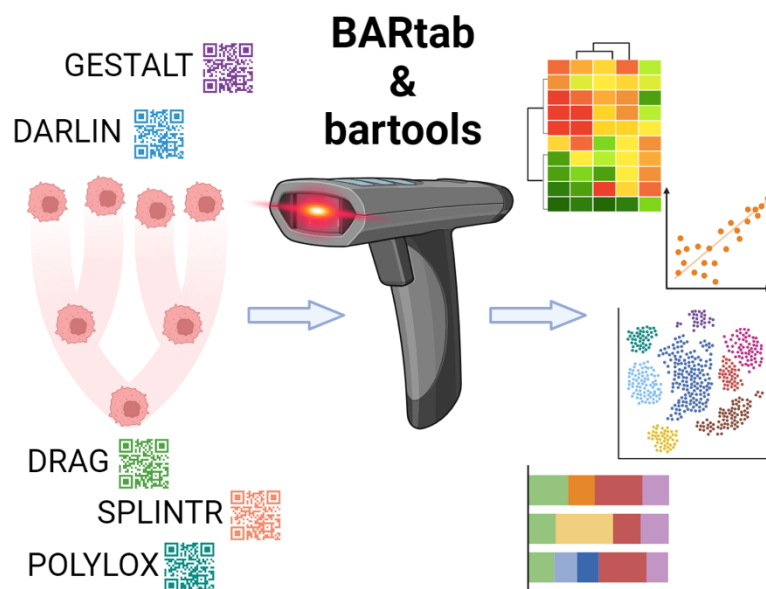
The second package, bartools, is dedicated to the further analysis of quantified barcode abundance estimates and can import both the BARTab output as well as results from other software packages in simple csv format. Bartools focuses in particular on barcode normalization, visualization, and evaluation of barcode composition and diversity alongside basic statistical testing. The introduction of intuitive visualization techniques and accompanying statistics greatly aids exploratory analysis of such large datasets and rapid QC evaluation of new experiments. For non-specialist teams, this will enable the incorporation of such approaches in their experimental toolbox. bartools does require the user to be familiar with the R environment, while other tools have provided Rshiny apps for even easier access^{8,9}. The authors describe extensive testing against previously published datasets, both from their own lab as well as others. This includes not only barcode assays performed by bulk sequencing as well as scRNAseq, but also spatial transcriptomics, demonstrating the most up to date range of utility.

The field faces upcoming challenges in providing guidelines on how to identify barcodes from noise and provide accurate barcode identification and quantification⁴. It is easy to generate “normal looking” data that is in fact polluted by technical artifacts such as PCR or sequencing errors. Simulation tools incorporated into analysis suites have already started to address such issues⁶ but more diverse barcode designs as well as biological and technical parameters need to be added. Additionally, such tools need to be extended to incorporate evolvable barcodes that introduce added complexity to the analysis⁷. While evolvable barcodes have their own diversity of designs, incorporating them would provide a truly flexible suite of tools for all single cell lineage analysis. Finally, maintaining and updating the existing tools over time is often a challenge in academia. The successful examples of Seurat and scanpy for scRNAseq analyses should guide the field.

References

1. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
2. Sankaran, V. G., Weissman, J. S. & Zon, L. I. Cellular barcoding to decipher clonal dynamics in disease. *Science* **378**, (2022).
3. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics* **21**, (2020).
4. Lyne, A. M. *et al.* A track of the clones: new developments in cellular barcoding. *Experimental Hematology* **68**, (2018).
5. Holze et al. Cell Reports Methods 2024. S2667-2375(24)00094-8.

6. Sun, W. *et al.* Extracting, filtering and simulating cellular barcodes using CellBarcode tools. *Nat. Comput. Sci.* **4**, (2024).
7. Jones, M. G. *et al.* Inference of single-cell phylogenies from lineage tracing data using Cassiopeia. *Genome Biol.* **21**, (2020).
8. Hadj Abed, L., Tak, T., Cosgrove, J. & Perié, L. CellDestiny: A RShiny application for the visualization and analysis of single-cell lineage tracing data. *Front. Med.* **9**, (2022).
9. Espinoza, D. A., Mortlock, R. D., Koelle, S. J., Wu, C. & Dunbar, C. E. Interrogation of clonal tracking data using barcodetrackR. *Nat. Comput. Sci.* **1**, (2021).



Cell lineage can be determined by multiple experimental barcoding protocols (left). BARTab and bartools, published by Holze et al in this issue of Cell Reports Methods, provide flexible, broadly applicable computational analysis tools that can analyze diverse barcoding experimental protocols, providing QC assessment, diverse visualization tools, and statistical testing. Figure created using biorender.com.