



Search

Advanced

User Guide

Save

Email

Send to

Display options 

> [Cell Genom.](#) 2024 Jul 10;4(7):100592. doi: 10.1016/j.xgen.2024.100592. Epub 2024 Jun 25.

Synthetic DNA barcodes identify singlets in scRNA-seq datasets and evaluate doublet algorithms

Ziyang Zhang¹, Madeline E Melzer², Keerthana M Arun¹, Hanxiao Sun¹, Carl-Johan Eriksson³, Itai Fabian⁴, Sagi Shaashua⁴, Karun Kiani⁵, Yaara Oren⁴, Yogesh Goyal⁶

Affiliations + expand
PMID: 38925122 PMCID: [PMC11293576](#) DOI: [10.1016/j.xgen.2024.100592](#)

Abstract

Single-cell RNA sequencing (scRNA-seq) datasets contain true single cells, or singlets, in addition to cells that coalesce during the protocol, or doublets. Identifying singlets with high fidelity in scRNA-seq is necessary to avoid false negative and false positive discoveries. Although several methodologies have been proposed, they are typically tested on highly heterogeneous datasets and lack a priori knowledge of true singlets. Here, we leveraged datasets with synthetically introduced DNA barcodes for a hitherto unexplored application: to extract ground-truth singlets. We demonstrated the feasibility of our framework, "singletCode," to evaluate existing doublet detection methods across a range of contexts. We also leveraged our ground-truth singlets to train a proof-of-concept machine learning classifier, which outperformed other doublet detection algorithms. Our integrative framework can identify ground-truth singlets and enable robust doublet detection in non-barcoded datasets.

Keywords: barcoding; benchmarking; doublet detection; lineage tracing; machine learning; scRNA-seq; single-cell genomics; singletCode; singlets.

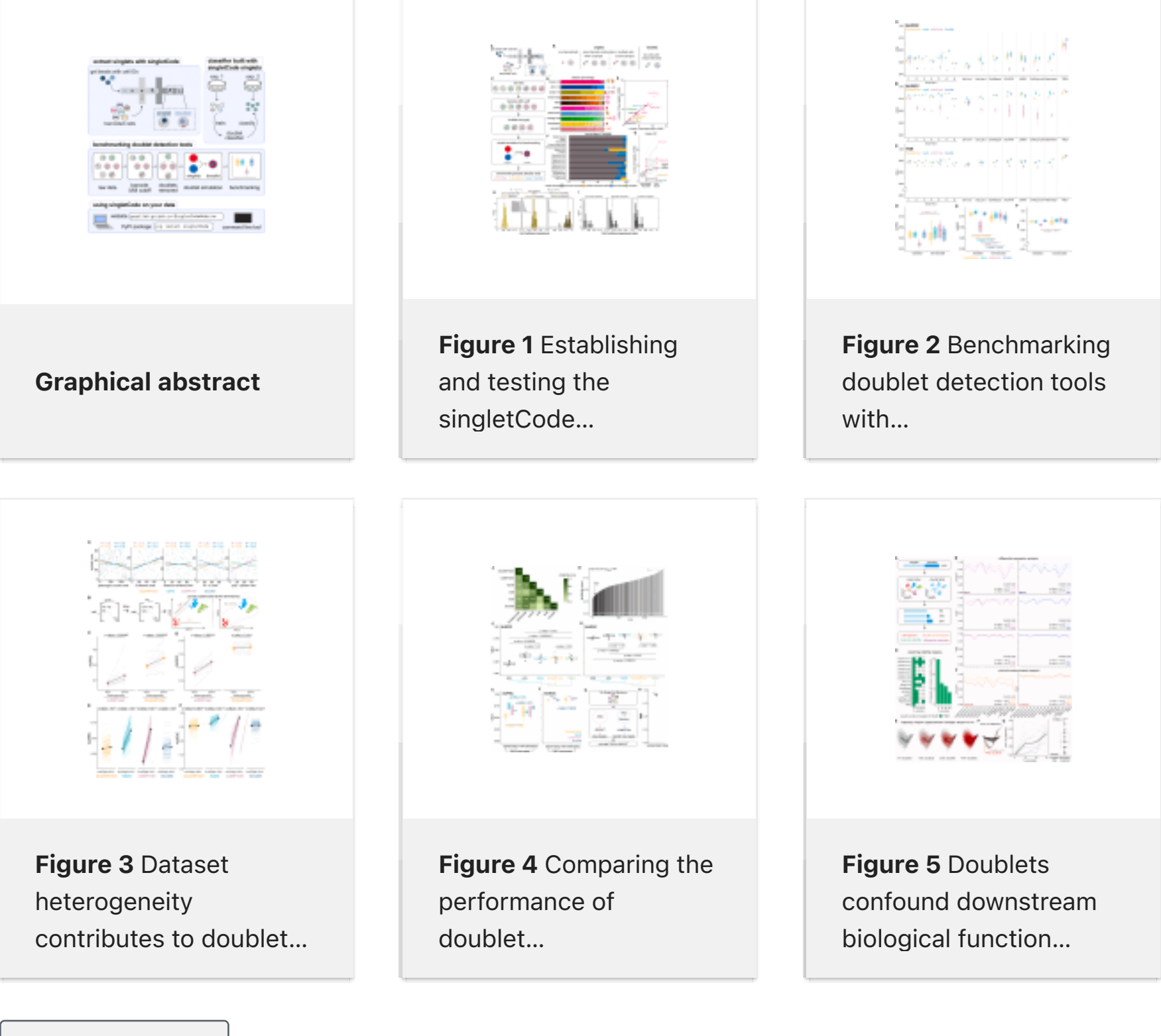
Copyright © 2024 The Author(s). Published by Elsevier Inc. All rights reserved.

[PubMed Disclaimer](#)

Conflict of interest statement

Declaration of interests Y.G. received consultancy fees from the Schmidt Science Fellows.

Figures



All figures (8)

Similar articles

[Chord: an ensemble machine learning algorithm to identify doublets in single-cell RNA sequencing data.](#)

Xiong KX, Zhou HL, Lin C, Yin JH, Kristiansen K, Yang HM, Li GB.
Commun Biol. 2022 May 30;5(1):510. doi: 10.1038/s42003-022-03476-9.
PMID: 35637301 [Free PMC article.](#)

[Double-jeopardy: scRNA-seq doublet/multiplet detection using multi-omic profiling.](#)

Sun B, Bugarin-Estrada E, Overend LE, Walker CE, Tucci FA, Bashford-Rogers RJM.
Cell Rep Methods. 2021 May 24;1(1):None. doi: 10.1016/j.crmeth.2021.100008.
PMID: 34278374 [Free PMC article.](#)

[Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data.](#)

Xi NM, Li JJ.
Cell Syst. 2021 Feb 17;12(2):176-194.e6. doi: 10.1016/j.cels.2020.11.008. Epub 2020 Dec 17.
PMID: 33338399 [Free PMC article.](#)

[Supervised application of internal validation measures to benchmark dimensionality reduction methods in scRNA-seq data.](#)

Koch FC, Sutton GJ, Voineagu I, Vafaee F.
Brief Bioinform. 2021 Nov 5;22(6):bbab304. doi: 10.1093/bib/bbab304.
PMID: 34374742 Review.

[Machine learning and statistical methods for clustering single-cell RNA-sequencing data.](#)

Petegrosso R, Li Z, Kuang R.
Brief Bioinform. 2020 Jul 15;21(4):1209-1223. doi: 10.1093/bib/bbz063.
PMID: 31243426 Review.

[See all similar articles](#)

Cited by

[Prevalence of and gene regulatory constraints on transcriptional adaptation in single cells.](#)

Mellis IA, Melzer ME, Bodkin N, Goyal Y.
Genome Biol. 2024 Aug 12;25(1):217. doi: 10.1186/s13059-024-03351-2.
PMID: 39135102 [Free PMC article.](#)

[Improving doublet cell removal efficiency through multiple algorithm runs.](#)

She Y, Wang C, Zhao Q.
Comput Struct Biotechnol J. 2025 Jan 15;27:451-460. doi: 10.1016/j.csbj.2025.01.009. eCollection 2025.
PMID: 39911841 [Free PMC article.](#)

[Genetically-Encoded Fluorescence Barcodes Allow for Single-Cell Analysis via Spectral Flow Cytometry.](#)

Lu X, Pritko DJ, Abravanel ME, Huggins JR, Ogunleye O, Biswas T, Ashy KC, Woods SK, Livingston MWT, Blenner MA, Birtwistle MR.
bioRxiv [Preprint]. 2025 Mar 28:2024.10.23.619855. doi: 10.1101/2024.10.23.619855.
Update in: [ACS Synth Biol.](#) 2025 May 16;14(5):1533-1548. doi: 10.1021/acssynbio.4c00807.
PMID: 39484616 [Free PMC article.](#) Preprint.

[Mapping cell-cell fusion at single-cell resolution.](#)

Gardner AL, Zheng L, Howland K, Saunders A, Ramirez A, Parker P, Iloegbunam C, Morgan D, Jost TA, Brock A.
bioRxiv [Preprint]. 2025 Jan 24:2024.12.11.627873. doi: 10.1101/2024.12.11.627873.
PMID: 39896473 [Free PMC article.](#) Preprint.

[Genetically Encoded Fluorescence Barcodes Allow for Single-Cell Analysis via Spectral Flow Cytometry.](#)

Lu X, Pritko DJ, Abravanel ME, Huggins JR, Ogunleye O, Biswas T, Ashy KC, Woods SK, Livingston MWT, Blenner MA, Birtwistle MR.
ACS Synth Biol. 2025 May 16;14(5):1533-1548. doi: 10.1021/acssynbio.4c00807. Epub 2025 May 6.
PMID: 40326708

References

- Cui L.-L., Kinnunen T., Boltze J., Nystedt J., Jolkkonen J. Clumping and Viability of Bone Marrow Derived Mesenchymal Stromal Cells under Different Preparation Procedures: A Flow Cytometry-Based In Vitro Study. Stem Cells Int. 2016;2016 - [PMC](#) - [PubMed](#)
- Kuonen F., Touvrey C., Laurent J., Ruegg C. Fc block treatment, dead cells exclusion, and cell aggregates discrimination concur to prevent phenotypical artifacts in the analysis of subpopulations of tumor-infiltrating CD11b(+) myelomonocytic cells. Cytometry A. 2010;77:1082-1090. - [PubMed](#)
- Bernstein N.J., Fong N.L., Lam I., Roy M.A., Hendrickson D.G., Kelley D.R. Solo: Doublet Identification in Single-Cell RNA-Seq via Semi-Supervised Deep Learning. Cell Syst. 2020;11:95-101.e5. - [PubMed](#)
- Xi N.M., Li J.J. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. Cell Syst. 2021;12:176-194.e6. - [PMC](#) - [PubMed](#)
- Luecken M.D., Theis F.J. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. 2019;15:e8746. - [PMC](#) - [PubMed](#)

Show all 81 references

MeSH terms

- > [Algorithms*](#)
- > [Animals](#)
- > [DNA Barcoding, Taxonomic* / methods](#)
- > [Humans](#)
- > [Machine Learning](#)
- > [Sequence Analysis, RNA / methods](#)
- > [Single-Cell Analysis* / methods](#)
- > [Single-Cell Gene Expression Analysis](#)

Related information

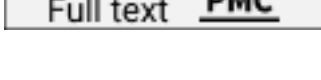
[GEO DataSets](#)
[MedGen](#)
[Related Project](#)
[SRA](#)

LinkOut - more resources

Full Text Sources
[ClinicalKey](#)
[Elsevier Science](#)
[PubMed Central](#)

Molecular Biology Databases
[NIAID Data Ecosystem - Find datasets on Infectious and Immune-mediated Diseases](#)

FULL TEXT LINKS



ACTIONS

 [Cite](#)

 [Collections](#)

SHARE



PAGE NAVIGATION

< [Title & authors](#)

Abstract
Conflict of interest statement
Figures
Similar articles
Cited by
References
MeSH terms
Related information
LinkOut - more resources

FOLLOW NCBI



Connect with NLM
  

National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20894

Web Policies
FOIA
HHS Vulnerability Disclosure

Help
Accessibility
Careers