nature portfolio

Corresponding author(s): Yiqiao Zhong

Last updated by author(s): May 15, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The datasets used in this study are publicly available and we collected the datasets from the following resources. CIFAR-10 dataset: obtained at https://www.cs.toronto.edu/~kriz/cifar.html. Describable Textures dataset: obtained at https://www.robots.ox.ac.uk/~vgg/data/dtd/. Mouse brain single-cell ATAC-seq data: obtained at https://doi.org/10.6084/m9.figshare.12420968. Mouse embryonic stem cell differentiation data: obtained under NCBI GEO accession code GSE98664, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98664. Single-cell RNA-seq dataset generated from PBMCs treated with interferon-beta is obtained from the R package Seurat (version 5.0.3) under the name ifnb. The single-cell RNA-seq data of human pancreatic tissues is obtained from the smartseq2 dataset in the R package Seurat (version 5.0.3) under the name panc8. Mouse mammary epithelial single-cell data is obtained from the R package scRNAseq (version 2.20.0) under the name BachMammaryData. The simulated data are generated using R (version 4.2.1) and R package MGMM (version 1.0.1.1). The computer code and the processed data are available on GitHub https://github.com/zhexuandliu/MapContinuity-NE-Reliability and on Zenodo https://doi.org/10.5281/zenodo.15384394.
Data analysis	The computer code is available on GitHub https://github.com/zhexuandliu/MapContinuity-NE-Reliability and on Zenodo https://doi.org/10.5281/zenodo.15384394. We used the following softwares/packages across all analysis: Python (version 3.11.0), scikit-learn (version 1.2.0), numpy (version 1.23.5), pandas (version 1.5,2), dynamicviz (version 0.0.3), EMBEDR (version 2.1.1), R (version 4.2.1), scDEED (version

0.1.0), dbscan (version 1.2-0), clusterSim (version 0.51-3), stats (version 4.2.1), MGMM (version 1.0.1.1), Rtsne (version 0.17), Seurat (version 5.0.3), scRNAseq (version 2.20.0), ggplot2 (version 3.5.1), egg (version 0.4.5), uwot (version 0.2.2), viridis (version 0.6.5), Rfast (version 2.1.4), parallel (version 4.2.1), FNN (version 1.1.4.1).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The datasets used in this study are publicly available and can be accessed through the following sources. The CIFAR-10 dataset used in this study is available at https://www.cs.toronto.edu/~kriz/cifar.html. The Describable Textures dataset used in this study is available at https://www.robots.ox.ac.uk/~vgg/data/dtd/. The pretrained ResNet-18 model is available at Hugging Face https://huggingface.co/edadaltocg/resnet18_cifar10. The mouse brain single-cell ATAC-seq data can be downloaded from Figshare https://doi.org/10.6084/m9.figshare.12420968. Mouse embryonic stem cell differentiation data is available with NCBI GEO accession code GSE98664, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98664. The single-cell RNA-seq dataset generated from PBMCs treated with interferon-beta is available from the R package Seurat (version 5.0.3) under the name ifnb, and is also available with NCBI GEO accession code GSE96583 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96583. The single-cell RNA-seq data of human pancreatic tissues is available from the smartseq2 dataset in the R package Seurat (version 5.0.3) under the name panc8, and is also available in EMBL-EBI with accession code E-MTAB-5061 https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-5061. Mouse mammary epithelial single-cell data is available from the R package scRNAseq (version 2.20.0) under the name \texttt{BachMammaryData}, and is also available with NCBI GEO accession code GSE106273 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106273. The computer code and the processed data are available on GitHub https://github.com/zhexuandliu/MapContinuity-NE-Reliability and on Zenodo https://doi.org/10.5281/zenodo.15384394.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not calculated beforehand. We used publicly available datasets. The sample sizes of individual datasets are provided in Supplementary File Table 1.
Data exclusions	Our analysis of mouse embryonic stem cell differentiation dataset, the mouse brain single-cell ATAC-seq dataset, and the mouse mammary epithelial cell dataset in Section "Use case 2: enhancing interpretation of single-cell data" included all samples from the original datasets without exclusions. In the analysis of CIFAR-10 and Describable Texture datasets, we applied random subsampling. We also randomly subsampled the datasets listed in Supplementary File Table S2 to create datasets of different sample sizes to empirically verify the LOO assumption.
Replication	We replicated the success of our methods by verifying its success on multiple datasets. Each experiment was performed independently.

Randomization	Not applicable. We did not have experimental groups. This study only uses publicly available datasets, and the randomization details for these datasets are documented in their respective original publications.
Blinding	Not applicable. This study only uses publicly available datasets, and the blinding procedures for individual datasets are documented in their respective original publications.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable