nature portfolio

Corresponding author(s): Kenta Nakai

Last updated by author(s): Nov 29, 2024

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	All data analyzed within this manuscript are publicly available. No additional software was used for the data collection process.
Data analysis	We performed all the data analysis using our developed STAIG software (v1.0.0) in this paper, which is available at [https://github.com/y-itao/STAIG] and [https://doi.org/10.5281/zenodo.14238885]. STAIG requires python 3.10.11. STAIG has many dependencies, including pytorch v1.13.1, pyg v2.3, byol-pytorch v0.6.0, opencv-python v4.7.0, pyyaml v6.0, rpy2 v3.5.11, pyyaml v6.0, pot v0.9.0, rpy2 v3.5.11, scanpy v1.9.3, numpy v1.24.3. The compared methods include GraphST v1.1.1 [https://github.com/JinmiaoChenLab/GraphST], DeepST v1.0.0 [https://github.com/JiangBioLab/DeepST], Seurat v4.1.1 [https://github.com/satijalab/seurat], SpaGCN v1.2.0 [https://github.com/jianhuupenn/SpaGCN], conST v1.0.0 [https://github.com/ys-zong/conST], STAGATE v1.0.1 [https://github.com/QIFEIDKN/STAGATE], stLearn v0.4.12 [https://github.com/BiomedicalMachineLearning/stLearn] and SEDR v1.0.0 [https://github.com/JinmiaoChenLab/SEDR], MuCoST v1.1.0, STAligner v1.0.0, Tesla v1.1.0 . mclust v6.0.0 is used for result visualization. Gene ontology analysis was conducted using the clusterProfiler v4.14.3 The datasets examined in our study are publicly available and can be accessed in their unprocessed form directly from the original sources, as detailed in Supplementary Tables S1 and S2. The data used in this study are available in the Zenodo under accession code [https://doi.org/10.5281/zenodo.10277127].

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

All data analyzed in this paper are available in raw form from their original authors.
1. The DLPFC dataset is accessible within the spatialL/BD package (http://spatial.libd.org/spatialLIBD).
2. The MouseBrain dataset and human breast cancer dataset is collected from the 10x Genomics website (https://support.10xgenomics.com/spatial-gene-expression/datasets).
3. The human placental bed dataset can be downloaded from (https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-12698)
4. The zebrafish melanoma dataset can be downloaded from (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159709)
5. The Stereo-seq mouse olfactory bulb dataset can be downloaded from (https://github.com/JinmiaoChenlab/SEDR_analyses)
6. The Slide-seqV2 mouse hippocampus can be downloaded from (https://portals.broadinstitute.org/single_cell/study/slide-seq-study)
7. The slide-seqV2 mouse olfactory bulb dataset can be downloaded from (https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatia-transcriptomics-at-near-cellular-resolution-with-slide-seqv2#study-summary)
8. The STARmap mouse visual cortex dataset can be downloaded from (https://www.dropbox.com/sh/f7ebheru1lbz91s/AADm6D54GSEFXB1feRy6OSASa/visual_1020/20180505_BY3_1kgenes?dl=0&subfolder_nav_tracking=1)

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	This study is based on analysis of publicly available spatial transcriptomic datasets. Direct data on sex and gender are not applicable as the datasets do not include individual-level human participant data.
Reporting on race, ethnicity, or other socially relevant groupings	The research utilizes spatial transcriptomic data that is publicly available and does not contain personal identifiers or demographic information related to race, ethnicity, or other socially relevant groupings.
Population characteristics	This study does not involve direct collection of population characteristics as it is an analysis of pre-existing datasets. The datasets used are publicly available and do not include individual-level data on human research participants.
Recruitment	Not applicable as this study did not involve the recruitment of human participants; it is an analysis of existing datasets.
Ethics oversight	Ethical oversight is not applicable to the analysis of the existing, publicly available datasets in this study. No new data from human participants were collected.

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	No sample size calculation was performed. All data used in this manuscript were sourced from public repositories to demonstrate the functionalities of STAIG. The DLPFC dataset was used to evaluate the accuracy of spatial domain recognition due to manual annotations. The mouse brain dataset was used for broader validation and to showcase the integrate feature. Datasets from platforms such as Stereo-seq, Slide-seqV2, and STARmap were utilized to assess performance across other platforms. The human placental bed dataset, due to partial overlaps between slices, was used to demonstrate the integrate functionality. The human breast cancer dataset and the zebrafish melanoma dataset were employed to evaluate STAIG's capability to identify spatial domains within the tumor microenvironment. These datasets validate the effectiveness of our work from various platforms.
Data exclusions	We performed quality control of spatial transcriptomic data based on the common used and pre-established criteria in this filed.
Replication	All Attempts at replication were successful and can be performed independently.
Randomization	The allocation was random

Blinding	All results are based on published data which have been studied in their original publications. Therefore, blinding from investigators is not applicable for computational algorithms.

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, as this study did not involve any plant material or seed stocks.
Novel plant genotypes	Not applicable, as this study focused on spatial transcriptomic data from animal and human subjects and did not involve any plant genotypes.
Authentication	Not applicable, as no plant seed stocks or genotypes were used in this study.