nature portfolio

Corresponding author(s): Boyi Guo

Last updated by author(s): Mar 28, 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	All data used in this study were publicly available; no new data were collected or generated. No software was used for data collection
Data analysis	All code used for the current manuscript were conducted in R v4.4, Bioconductor v3.20. SpotSweeper (v.1.1.0) was developed for this manuscript and is publicly available as an R/Bioconductor package, and is also available from Github (https://github.com/MicTott/SpotSweeper).

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 DLPFC Visium datasets \cite{maynard_transcriptome-scale_2021, huuki-myers_integrated_2023} used for analyses in this manuscript can be obtained from \texttt{spatialLIBD} (\url{http://research.libd.org/spatialLIBD}) in \texttt{SpatialExperiment} format, which includes Manual Annotation labels from the original

sources. The DLPFC samples from Huuki-Myers et al. with dryspot and hangnail artifacts were samples "Br8325_ant" and "Br3942_mid", respectively. Other original datasets and any manual annotations are sourced from refs, including Visium ovarian cancer dataset \cite{Denisenko_2024} via the Gene Expression Omnibus series GSE211956, Slide-seqV2 mouse hippocampus dataset \cite{Stickels_2021} via an Rdata object made available on the \href{https://lulushang.org/SpatialPCA_Tutorial/slideseqV2.html}{SpatialPCA website}, Stereo-seq mouse embryonic dataset \cite{Chen_2022} via \href{https://db.cngb.org/stomics/datasets/STDS0000058/summary}{Spatial Transcriptic Omics DataBase}, MERFISH mouse colon dataset \cite{Cadinu_MERFISH_2024} via \texttt{MouseColonIbdCadinu2024} function from the \textit{MerfishData} R/Bioconductor package, STARmap mouse brain dataset via \cite{Shi_STARmap_2023} via the \texttt{STARmapPLUS_mousebrain} function from the \textit{STexampleData} R/Bioconductor package. All other data were sourced directly from 10x Genomics' publicly available datasets, including Visium Human Breast Cancer \url{(https://www.10xgenomics.com/datasets/human-breast-cancer-visium-fresh-frozen-whole-transcriptome-1-standard)}, Visium Coronal Mouse Brain \url{(https://www.10xgenomics.com/datasets/adult-mouse-brain-coronal-section-fresh-frozen-1-standard)}, Xenium 5K Coronal Mouse Brain \url{(https://www.10xgenomics.com/datasets/xenium-prime-fresh-frozen-mouse-brain)}, VisiumHD Human Breast Cancer \url{(https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-human-breast-cancer-fresh-frozen)}, and VisiumHD Coronal Mouse Brain \url{(https://www.10xgenomics.com/datasets/visium-hd-cytassist-gene-expression-mouse-brain-fresh-frozen)}. All other data supporting the findings of this study are available within the article and its supplementary files.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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	This study doesn't involve experiments to generate new data. All data used in the current manuscript were publicly available data from previous studies where appropriate sample sizes were determined. The number of datasets are decided to cover large majority of existing tissue types, spatial transcriptomics technologies.
Data exclusions	The only data points excluded in the current manuscript were Visium spots that did not lie underneath tissue samples. These spots do not contain useful biological information and it is standard practice to remove these spots.
Replication	The study is a computaional methods development, which doesn't require replication.
Randomization	The current study focuses on computational methodology development. It doesn't generate new data, and doesn't require randomization.
Blinding	The current study focuses on computational methodology development. It doesn't generate new data, and doesn't require blinding.

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

Methods

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