

Reporting Summary

Nature Research 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 Research 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 No software was used for data collection.

Data analysis

We compared the performance of 8 integration methods for predicting the spatial distribution of undetected transcripts: gimVI (Version 0.8.0b0), SpaGE(no version), Tangram(Version 1.0.0), Seurat (Version 3.6.3), SpaOTsc (Version 0.2), LIGER (Version 0.5.0), novoSpaRc (Version 0.4.3), stPlus(Version 0.0.6).

We compared the performance of 12 integration methods for predicting the cell type composition of spots: Cell2location(Version 0.6a0), DestVI((Version 0.14.4), SPOTlight((Version 0.1.7), SpatialDWLS(Version 1.0.4), Seurat(Version 4.0.5), Tangram((Version 1.0.0), RCTD(Version 1.2.0), Stereoscope(Version 0.14.4), STRIDE(Version 0.0.1b0), DSTG (Version 0.0.1), novoSpaRc (Version 0.4.3), SpaOTsc (Version 0.2).

We applied the Scuttle(Version 1.0.4) and Splatte(Version 1.19.3) package for the down-sampling of the spatial expression matrices of the datasets.

The code used in this paper is available at <https://github.com/QuKunLab/SpatialBenchmarking>.

The file provided at <https://github.com/QuKunLab/SpatialBenchmarking/Benchmarkingenvironment.yml> lists the software dependencies with version numbers.

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 Research [guidelines for submitting code & software](#) for further information.

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 list of figures that have associated raw data
- A description of any restrictions on data availability

Dataset 1 : seqFISH, <https://content.cruk.cam.ac.uk/jmlab/SpatialMouseAtlas2020/>; 10X Chromium, 'Sample 21' in MouseGastrulationData within the R/Bioconductor data package MouseGastrulationData.

Dataset 2 : seqFISH, <https://zenodo.org/record/3735329#.YY69HZMza3J>; Microwell-Seq, 'EmbryonicStemCells' in 'MCA_BatchRemoved_Merge_dge.h5ad' file in https://figshare.com/articles/dataset/MCA_DGE_Data/5435866.

Dataset 3 : seqFISH, <https://ars.els-cdn.com/content/image/1-s2.0-S0896627316307024-mmc6.xlsx>; 10X Chromium, "HIPP_sc_Rep1_10X sample" in GSE158450 in the GEO database.

Dataset 4 : seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>, and the spatial coordinate of each spot was generated using "stitchFieldCoordinates" function in Giotto; Smart-seq, mouse primary visual cortex (VISp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 5 : seqFISH+, <https://github.com/CaiGroup/seqFISH-PLUS>; Drop-seq, GSE148360 in the GEO database.

Dataset 6 : MERFISH, the 18th female parent mouse (Animal ID = 18) in <https://datadryad.org/stash/dataset/doi:10.5061/dryad.8t8s248>; 10X Chromium, GSE113576 in the GEO database.

Dataset 7 : MERFISH, the 'B1_cell' used in https://www.pnas.org/doi/suppl/10.1073/pnas.1912459116/suppl_file/pnas.1912459116.sd12.csv; 10X Chromium, BC22 in GSE152048 in the GEO database.

Dataset 8 : MERFISH, 'mouse1_slice162' in <https://caltech.box.com/shared/static/dzqt6ryytmjbgyai356s1z0phtnsbaol.gz>; 10X Chromium, https://data.nemoarchive.org/biccn/lab/zeng/transcriptome/scell/10x_v3/mouse/processed/analysis/10X_cells_v3_AIBS/.

Dataset 9 : MERFISH, <https://github.com/spacetx-spacejam/data/>; Smart-seq, mouse primary visual cortex (VISp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 10 : STARmap, "20180505_BY3_1kgenes" in <https://www.starmapresources.com/data>; Smart-seq, mouse primary visual cortex (VISp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 11 : STARmap, "20180419_BZ9_control" in <https://www.starmapresources.com/data>; 10X Chromium, "PFC_sc_Rep2_10X" in GSE158450 in the GEO database.

Dataset 12 : ISS, <https://github.com/spacetx-spacejam/data/>; Smart-seq, <https://portal.brain-map.org/atlas-and-data/rnaseq/human-mtg-smart-seq>.

Dataset 13 : ISS, <https://github.com/spacetx-spacejam/data/>; Smart-seq, mouse primary visual cortex (VISp) in the dataset in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 14 : FISH, <https://github.com/rajewsky-lab/distmap>; Drop-seq, GSE95025 in the Gene Expression Omnibus (GEO) database.

Dataset 15 : osmFISH, cortical regions in <http://linnarssonlab.org/osmFISH/>; Smart-seq, mouse somatosensory cortex (SSp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-smart-seq>.

Dataset 16 : BARISTASeq, <https://github.com/spacetx-spacejam/data/>; Smart-seq, mouse primary visual cortex (VISp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 17 : ExSeq, <https://github.com/spacetx-spacejam/data/>; Smart-seq, mouse primary visual cortex (VISp) in <https://portal.brain-map.org/atlas-and-data/rnaseq/mouse-v1-and-alm-smart-seq>.

Dataset 18 : 10X Visium, Vis5A in GSE161318 in the GEO database; 10X Chromium, D2_Ev3 in GSE159500 in the GEO database.

Dataset 19 : 10X Visium, Vis9A in GSE161318 in the GEO database; 10X Chromium, D7_Ev3 in GSE159500 in the GEO database.

Dataset 20 : 10X Visium, "CID3586" in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 21 : 10X Visium, "1160920F" in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 22 : 10X Visium, "CID4290" in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 23 : 10X Visium, "CID4465" in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 24 : 10X Visium, "CID44971" https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 25 : 10X Visium, "CID4535" in https://zenodo.org/record/4739739#.YY6N_pMzaWC; 10X Chromium, 'CID3586' in GSE176078 in the GEO database.

Dataset 26 : 10X Visium, "Visium-A" in GSE159709 in the GEO database; 10X Chromium, "SingleCell-E" in GSE159709 in the GEO database.

Dataset 27 : 10X Visium, "Visium-A1" in GSE160137 in the GEO database; 10X Chromium, "Pax2-GFP_SC-2" in GSE143806 in the GEO database.

Dataset 28 : 10X Visium, "D25" in GSE159697 in the GEO database; 10X Chromium, "V8" in GSE142489 in the GEO database.

Dataset 29 : 10X Visium, Sham Model in GSE171406 in the GEO database; 10X Chromium, wild-type sham mouse in GSE171639 in the GEO database.

Dataset 30 : 10X Visium, Ischemia Reperfusion Injury Model in GSE171406 in the GEO database; 10X Chromium, wild-type ischemic acute kidney injury mouse in GSE171639 in the GEO database.

Dataset 31 : 10X Visium, "section1" in GSE153424 in the GEO database; 10X Chromium, "brain1_cx" in GSE153424 in the GEO database.

Dataset 32 : 10X Visium, "Visium_10X" in GSE158450 in the GEO database; 10X Chromium, "PFC_sc_Rep1_10X" in GSE158450 in the GEO database.

Dataset 33 : 10X Visium, "Visium_10X" in GSE158450 in the GEO database; 10X Chromium, "HIPP_sc_Rep1_10X" in GSE158450 in the GEO database.

Dataset 34 : 10X Visium, GSE154107 in the GEO database; 10X Chromium, sample "(LPS36hr) scRNAseq" in GSE151658 in the GEO database.

Dataset 35 : 10X Visium, "ETOH" in GSE159697 in the GEO database; 10X Chromium, "V8" in GSE142489 in the GEO database.

Dataset 36 : 10X Visium, 'PBS' samples of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, 'PBS' samples in <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 37 : 10X Visium, Tumor A1 of Tissue 1 in <https://github.com/romain-lopez/DestVI-reproducibility>; 10X Chromium, <https://github.com/romain-lopez/DestVI-reproducibility>.

Dataset 38 : 10X Visium, <https://console.cloud.google.com/storage/browser/tommaso-brain-data>; 10X Chromium, 'batch 9' in 'tangram_demo_mop_sn_tutorial.h5ad' file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 39 : Slide-seq, <https://console.cloud.google.com/storage/browser/tommaso-brain-data>; 10X Chromium, 'batch 9' in 'tangram_demo_mop_sn_tutorial.h5ad' file from <https://console.cloud.google.com/storage/browser/tommaso-brain-data>.

Dataset 40 : Slide-seqV2, SCP948 in https://singlecell.broadinstitute.org/single_cell/; 10X Chromium, sample M003 of study SCP795 in https://singlecell.broadinstitute.org/single_cell/.

Dataset 41 : Slide-seqV2, "Puck_200115_08" in https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seqv2#study-download; Drop-seq, we randomly sampled 10,000 cells from 'GSE116470_F_GRCm38.81.P60Hippocampus.raw.dge.txt.gz' file in GSE116470 in the GEO database.

Dataset 42 : ST, GSM4284322 in the GEO database; 10X Chromium, "GSE144236_cSCC" in GSE144236 in the GEO database.

Dataset 43 : ST, wild-type replicate 1 in <https://data.mendeley.com/datasets/6s959w2zyr/1>; 10X Chromium, GSE116470 in the GEO database.

Dataset 44 : HDST, replicate1 in GSE130682 in the GEO database; 10X Chromium, WT1 samples used from GSE121891 in the GEO database.

Dataset 45 : Seq-scope, <https://deepblue.lib.umich.edu/data/downloads/gx41mj14n>; Smart-seq2, liver sample in GSE109774 in the GEO database.

The raw and processed data of these datasets are also available at : https://drive.google.com/drive/folders/1pHmE9cg_tMcouV1LFJFtbyBJNp7oQo9J?usp=sharing. A summary of these datasets is given in Supplementary Table 1.

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

We used 45 paired spatial transcriptomics and scRNA-seq datasets from published studies. The spatial transcriptomic datasets were produced by 13 different spatial transcriptomics approaches. A summary of the individual accession numbers is given in Supplementary Table 1. The details of these datasets are listed as follows:

Dataset 1 has 8425 spots in spatial transcriptomics data and 4651 cells in scRNA-seq data;
 Dataset 2 has 175 spots in spatial transcriptomics data and 9991 cells in scRNA-seq data;
 Dataset 3 has 3585 spots in spatial transcriptomics data and 8596 cells in scRNA-seq data;
 Dataset 4 has 524 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 5 has 2050 spots in spatial transcriptomics data and 31217 cells in scRNA-seq data;
 Dataset 6 has 4975 spots in spatial transcriptomics data and 31299 cells in scRNA-seq data;
 Dataset 7 has 645 spots in spatial transcriptomics data and 9234 cells in scRNA-seq data;
 Dataset 8 has 6963 spots in spatial transcriptomics data and 7240 cells in scRNA-seq data;
 Dataset 9 has 2399 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 10 has 1549 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 11 has 1380 spots in spatial transcriptomics data and 7737 cells in scRNA-seq data;
 Dataset 12 has 6000 spots in spatial transcriptomics data and 15928 cells in scRNA-seq data;
 Dataset 13 has 6000 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 14 has 3039 spots in spatial transcriptomics data and 1297 cells in scRNA-seq data;
 Dataset 15 has 3405 spots in spatial transcriptomics data and 5613 cells in scRNA-seq data;
 Dataset 16 has 11426 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 17 has 1154 spots in spatial transcriptomics data and 14249 cells in scRNA-seq data;
 Dataset 18 has 982 spots in spatial transcriptomics data and 4748 cells in scRNA-seq data;
 Dataset 19 has 995 spots in spatial transcriptomics data and 4816 cells in scRNA-seq data;
 Dataset 20 has 4784 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 21 has 4895 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 22 has 2432 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 23 has 1211 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 24 has 1162 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 25 has 1127 spots in spatial transcriptomics data and 6178 cells in scRNA-seq data;
 Dataset 26 has 2425 spots in spatial transcriptomics data and 1911 cells in scRNA-seq data;
 Dataset 27 has 198 spots in spatial transcriptomics data and 3415 cells in scRNA-seq data;
 Dataset 28 has 277 spots in spatial transcriptomics data and 4740 cells in scRNA-seq data;
 Dataset 29 has 1835 spots in spatial transcriptomics data and 10872 cells in scRNA-seq data;
 Dataset 30 has 2064 spots in spatial transcriptomics data and 13600 cells in scRNA-seq data;
 Dataset 31 has 3805 spots in spatial transcriptomics data and 8798 cells in scRNA-seq data;
 Dataset 32 has 3024 spots in spatial transcriptomics data and 3512 cells in scRNA-seq data;
 Dataset 33 has 3024 spots in spatial transcriptomics data and 8653 cells in scRNA-seq data;
 Dataset 34 has 1888 spots in spatial transcriptomics data and 8346 cells in scRNA-seq data;
 Dataset 35 has 744 spots in spatial transcriptomics data and 4561 cells in scRNA-seq data;
 Dataset 36 has 369 spots in spatial transcriptomics data and 7268 cells in scRNA-seq data;
 Dataset 37 has 2125 spots in spatial transcriptomics data and 7185 cells in scRNA-seq data;
 Dataset 38 has 2669 spots in spatial transcriptomics data and 3499 cells in scRNA-seq data;
 Dataset 39 has 9852 spots in spatial transcriptomics data and 3499 cells in scRNA-seq data;
 Dataset 40 has 41674 spots in spatial transcriptomics data and 26252 cells in scRNA-seq data;
 Dataset 41 has 6000 spots in spatial transcriptomics data and 10000 cells in scRNA-seq data;
 Dataset 42 has 1145 spots in spatial transcriptomics data and 48164 cells in scRNA-seq data;
 Dataset 43 has 604 spots in spatial transcriptomics data and 15095 cells in scRNA-seq data;
 Dataset 44 has 6000 spots in spatial transcriptomics data and 10259 cells in scRNA-seq data;
 Dataset 45 has 2177 spots in spatial transcriptomics data and 981 cells in scRNA-seq data;

Data exclusions

No data was excluded from the study. In the data preprocessing step of scRNA-seq data, we filtered cells that captured >200 RNA transcripts.

Replication

To make sure that the experimental findings are reproducible, we (1) verified the performance of the 8 integration methods for predicting the spatial distribution of undetected transcripts on 45 datasets and 32 simulated datasets, (2) verified the performance of the 8 integration

methods under a varied set of down-sampling settings to assess the robustness of each method. (3) verified the performance of the 10 integration methods for predicting the cell type composition of spots on 32 simulated datasets

Randomization

This is not relevant to our study because we reanalyzed publicly available data or generated our own synthetic data.

Blinding

In the comparison, all the methods were blinded to the ground truth of the spatial transcriptomics data. The outputs from the methods were then compared to the ground truth available in the respective datasets.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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