nature portfolio

Corresponding author(s): Cindy van Velthoven & Hongkui Zeng

Last updated by author(s): Jun 5, 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	scRNA-seq data was collected using the 10x Genomics Chromium Single Cell 3' v3, v3.1 and v4 kits. Sequencing was performed on the Illumina HiSeq2500 and NovaSeq6000. MERFISH data were acquired using the Vizgen MERSCOPE platform.

Data analysis	Most of the methods used to analyze the data have been described before (Yao et al., 2023) and the following methods are either newly introduced or modified version for this manuscript. We performed PCA (R package stats, v4.4.1, RRID:SCR_025968, https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html). Followed by creation of UMAPs (v0.5.6, RRID:SCR_018217, https://github.com/lmcinnes/umap) Constellation plots were generated using scrattch.bigcat (v0.0.5, https://github.com/AllenInstitute/scrattch.bigcat). We performed independent component analysis using fastICA (v1.2-5.1, RRID:SCR_013110, https://cran.r-project.org/web/packages/fastICA/). We applied UCell (v2.8.0, DOI: 10.18129/B9.bioc.UCell) to assign a "gene module score" to each cell. BANKSY was used to perform spatial domain clustering on MERFISH data (v0.99.12, https://github.com/prabhakarlab/Banksy) We performed mapping of cells from each adult external dataset to the 10xv3 whole-brain dataset using treeMap function from scrattch.mapping package (v0.55, https://github.com/AllenInstitute/scrattch-mapping). Raw scRNA-seq fastq files were processed with the CellRanger v6.1.1 pipeline using default parameters. Doublets were identified using a modified version of the DoubletFinder algorithm which is available in scrattch.hicat (v0.1.0, RRID:SCR_01809, https://github.com/AllenInstitute/scrattch.hicat). Seurat was used for initial clustering of developmental data (v5.1.0, RRID:SCR_016341) For dimensionality reduction and batch correction, the raw counts of the highly variable genes were used to train the scvi-tools (v0.17.1, RRID:SCR_026673) scVI variational autoencoder.

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

Sequencing reads were aligned to the mouse reference transcriptome (M21, GRCm38.p6)

The scRNA-seq and MERFISH datasets for this study are part of the Allen whole mouse brain (WMB) cell type atlas and are accessible through Neuroscience Multi-omic Data Archive (NeMO, RRID:SCR_016152) and Brain Image Library (BIL, RRID:SCR_017272). The adult 10x scRNA-seq datasets and the developmental scRNA-seq datasets are being made available through BRAIN Initiative Cell Atlas Network (BICAN, RRID:SCR_022794) (FASTQ files) and are available at NeMO under identifier https://assets.nemoarchive.org/dat-zqurqvh. The MERFISH dataset is available at BIL under DOI https://doi.org/10.35077/g.610.

The Subpallium-GABA cell type taxonomy is available from the Allen Brain Cell (ABC) Atlas (RRID:SCR_024440) to visualize both sc/snRNA-seq and MERFISH datasets. Instruction for access of the processed 10x scRNA-seq data is available at https://github.com/AllenInstitute/abc_atlas_access/blob/main/descriptions/WMB-10X.md, and instruction for access of the processed MERFISH data is available at https://github.com/AllenInstitute/abc_atlas_access/blob/main/descriptions/MERFISH-C57BL6J-638850.md.

The following published developmental datasets were used; Allaway et al. 2021, GSE164570; Kaplan et al. 2025, GSE280964; Lee et al. 2022, GSE167013 and GSE190593; Mayer et al. 2018, GSE103983 and GSE104156; La Manno et al. 2021, PRJNA637987; Turrero García et al. 2021, GSE184879.

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 does not involve human research participants.
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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

Life sciences study design

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

Sample size	No statistical methods were used to predetermine sample sizes For scRNA-seq data, sample size was determined based on extensive prior experience using these technologies in brain (Zeisel et al. 2018, Yao et al 2021). Post-hoc analysis of clustering shows that it is robust to downsampling, which indicates that the sample size is sufficient to distinguish cell types as described in the manuscript. For MERFISH data we evenly sampled one mouse brain to capture most brain regions and get an even distribution of cells types throughout the brain.
Data exclusions	Cells from scRNA-seq data underwent stringent QC and cell quality was assessed based on gene detection, qc score, and doublet score. The qc score was calculated by summing the log transformed expression of a set of genes whose expression level is decreased significantly in poor quality cells. We used this qc score to quantify the integrity of cytoplasmic mRNA content, which tended to show bimodal distribution. Cells at the low-end were very similar to single nuclei, which we removed for downstream analysis. Doublets were identified using a modified version of the DoubletFinder algorithm (available in scrattch.hicat, https://github.com/AllenInstitute/scrattch.hicat, v1.0.9) and removed when doublet score > 0.3. Post-clustering we excluded noise clusters which are clusters with significantly lower gene detection due to extensive drop out, and clusters due to doublets or contamination. We first identified doublet clusters based on the co-expression of any pair of broad class marker genes using find_doublet_by_marker function in scrattch.bigcat package. To identify other doublet clusters, we searched for triplets of clusters A, B and C, wherein A was the putative doublet cluster, such that up-regulated genes of A relative to B largely overlapped with up-regulated genes in C relative to B, and up-regulated genes in A relative to C largely overlapped with up-regulated genes of B relative to C. After removing all doublet clusters, we then identified clusters with lower gene detection. To do that, we identified pairs of clusters such that one cluster with at least 50% fewer UMIs or >100 lower QC score, smaller size, and no more than one up-regulated gene relative to another cluster was identified as the low-quality cluster. For MERFISH data the cell-by-gene table containing segmented cells was filtered to keep cells with a volume > 100 µm3 and < 3,000 µm3, that have at least 15 genes detected and contain a minimum of 40 but no more than 3,000 mRNA molecules to remove low quality cells and doublets that are outside of these ranges.
Replication	Most of the scRNA-seq experiments were carried out at least twice independently and at least 2 mice and multiple brain dissections per mouse were used. All attempts at replication were successful. For MERFISH, 59 sections from one mouse brain were generated.
Randomization	Randomization of animals to different groups is not relevant to our study design. There were no experimental vs. control groups.
Blinding	Prior to clustering, single cell transcriptomes were analyzed for previously known marker genes and were segregated into large groups: non-neuronal, glutamatergic and GABAergic. Clustering was then performed blind to the cell source or any other metadata that could reveal sample identity.

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	All animals used in this study were house mice (Mus musculus) maintained on the C57BL/6J background. Animals were euthanized at E11.5-E14.5 (n=6), P0 (n=14), P4 (n=12), P8 (n=16), P12 (n=12), P14 (n=14). For the developmental dataset, each animal's unique ID, sex, age, and genotype are listed in Supplementary Table 4. The donor information for the adult mice can be found in Yao et al., 2023 (https://doi.org/10.1038/s41586-023-06812-z) Supplemental Table 2. Mice had ad libitum access to food and water and were group-

	housed within a temperature- (21-22°C), humidity- (40-51%), and light- (14/10hr light/dark cycle) controlled room within the vivariums of the Allen Institute for Brain Science.
Wild animals	This study did not involve wild animals.
Reporting on sex	For ages P0 to P14, both male and female mice were used to collect scRNA-seq data. For the embryonic ages E11.5 to E14.5 we collected brains and performed post-hoc sex identification. Therefore, sexes are not fully balanced for the embryonic samples.
Field-collected samples	This study did not involve field-collected samples.
Ethics oversight	All experimental procedures related to the use of mice were approved by the Institutional Animal Care and Use Committee of the Allen Institute for Brain Science, in accordance with NIH guidelines.

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*