nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

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 single cell multiome data for PBMC are available at 10X Genomics website. The human AD single cell multiome data are available at https://personal.broadinstitute.org/bjames/AD_snATAC/. The ChIA-PET chromatin interaction loop files are available from the ENCODE portal. Chromatin contact maps generated by sn-m3C-seq are available at https://salkinstitute.app.box.com/s/fp63a4j36m5k255dhje3zcj5kfuzkyj1. Hi-C data from an AD study profiling from post-mortem prefrontal cortex tissues from both dementia-free elderly individuals and patients with AD are available at http://menglab.pub/hic/. The CRISPR dataset generated in the K562 cell line from ENCODE is available at https://www.encodeproject.org/files/ENCFF968BZL/. The CRISPRi dataset generated in human primary astrocytes from AstroREG is available at https://github.com/Voineagulab/astrocyte_crispri.
Data analysis	Analyses were conducted using R v4.4 and Python 3.10.12 on the University of Florida HiPerGator HPC cluster (Linux, x86_64-pc-linux-gnu). Key Python packages included PyTorch v2.9.0+cu128, NumPy v1.26.0, pandas v2.1.1, SciPy v1.14.0, scikit-learn v1.3.1, and h5py v3.8.0. Custom inference software: DeepCNet R package (https://github.com/lichen-lab/DeepCNet), implementing the deep learning framework described in the manuscript. Single-cell preprocessing: Seurat (v5.4.0), Signac (v1.14.0). Genomic analysis: GenomicRanges (v1.56.2); EnsDb.Hsapiens.v86 (hg38, v2.99.0), EnsDb.Hsapiens.v75 (hg19, v2.99.0); rtracklayer (v1.64.0) for liftover. Statistical analysis: DescTools (v0.99.60), MatchIt (v4.7.1), pROC (v1.18.5), and PRROC (v1.3.1).

Motif and TF binding analysis: motifmatchr (v1.26.0), TFBSTools (v1.42.0), JASPAR2020 (v0.99.10); chromVAR (v1.26.2) for background peak generation.

Network visualization: Cytoscape (v3.10.3).

All software versions and dependencies are documented in the GitHub repository. Demo data and a README with installation and usage instructions are included in the R package.

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 single cell multiome data for PBMC are available at 10X Genomics website. The human AD single cell multiome data are available at https://personal.broadinstitute.org/bjames/AD_snATAC/. The ChIA-PET chromatin interaction loop files are available from the ENCODE portal. Chromatin contact maps generated by sn-m3C-seq are available at https://salkinstitute.app.box.com/s/fp63a4j36m5k255dhje3zcj5kfuzkyj1. Hi-C data from an AD study profiling from post-mortem prefrontal cortex tissues from both dementia-free elderly individuals and patients with AD are available at http://menglab.pub/hic/. The CRISPR dataset generated in the K562 cell line from ENCODE is available at https://www.encodeproject.org/files/ENCFF968BZL/. The CRISPRi dataset generated in human primary astrocytes from AstroREG is available at https://github.com/Voineagulab/astrocyte_crispri.

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 did not involve new data collection from human participants. All analyses were performed on publicly available datasets. Sex information was obtained from the original data sources where available. The PBMC single-cell multi-omics dataset was generated from a healthy 25-year-old female donor (10X Genomics). Sex information for the MFC cohort was reported in the original study by Xiong et al.
Reporting on race, ethnicity, or other socially relevant groupings	No new participant-level demographic data were collected in this study. The PBMC dataset from 10x Genomics did not provide individual-level race or ethnicity information. Race and ethnicity information for the MFC cohort was reported in the original study by Xiong et al.
Population characteristics	Not applicable. No new human subjects were recruited for this study. Data used were previously published and publicly available.
Recruitment	Not applicable. No recruitment of human participants was undertaken.
Ethics oversight	Not applicable to this study. All data analyzed were obtained from previously published sources that had received ethics approval from their respective institutions, as described in the original publications.

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 is a computational/method development study. We used publicly available single-cell multiome datasets of fixed size as provided by the original sources. No additional samples were collected.
Data exclusions	Low-quality cells and peaks were filtered following standard QC procedures implemented in Seurat/Signac, as described in the Methods. No arbitrary exclusions were made.
Replication	Key analyses were replicated across multiple cell types and independent datasets. Model performance was evaluated using chromosome-based cross-validation, and regulatory interaction predictions were further validated using independent chromatin-interaction and

perturbation datasets where available.

Randomization: Publicly available datasets were used and no randomization was performed by the authors.

Blinding: Not applicable. No experimental groups or manual annotations requiring blinding were involved. All data analyses were computational.

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: *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.*