nature portfolio

Corresponding author(s): Michael Lodato, Zhiping Weng

Last updated by author(s): 2025/07/01

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	No software was used in data collection.
Data analysis	The snRNA-seq reads were aligned to the human genome and assigned to genes (GENCODE v32) via Cell Ranger (v6.0.2 and v7.0.1). The barcode and UMI solved counts were further processed with Seurat (v4.3.0 and v5.1.0). scWGS and bulk WGS were first processed accordingly to GATK best practices. Reads were aligned to the human genome via bwa mem (v0.7.12) with default parameters. PCR duplicates were then filtered using picard, and the remaining reads were recalibrated via GATK (v4.1.8.1) "BaseRecalibrator" and "ApplyBQSR". Genotypes were then identified via GATK (v4.1.8.1) "HaplotypeCaller" and "GenotypeGVCFs". Finally, somatic SNVs were identified by comparing the scWGS to corresponding WGS from bulk tissues via SCAN2 (v1). Mutational signature analysis was performed using SignatureAnalyzer (version:GPU) and MutationalPatterns (v3.16.0). MERFISH data was decoded using Vizgen's analysis pipeline integrated within the MERSCOPE system. The Vizgen Post-processing tool (VPT) was used for cell segmentation and pre-filtering with a Gaussian filter and the CellPose algorithm. Gene Ontology analysis was performed in R using "gprofiler2" (v0.2.3) which sourced biological process terms from geneontology.org.

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

Raw RNA-seq and scWGS sequencing data not previously published, and MERFISH data are available at dbGaP (phs003445.v1.p1). Processed data are available at https://publications.wenglab.org/SomaMut/. An interactive genome browser of the snRNA-seq pseudo bulk expression data can be found at: https://genome.ucsc.edu/s/yutianxiong/Weng_Lodato_Aging. Previously published single cells analyzed in this study can be found at dbGaP (phs001485.v3.p1) and NIAGADS (NG00121). Previously published bulk DNA sequencing can be found at dbGap (phs001485.v1.p1), NCBI SRA (SRP041470 and SRP061939), and NIAGADS (NG00121). The previously published Velmeshev et al. data used for cell type annotation was downloaded from SRA accession number PRJNA434002. Previously published Herring et al. data was downloaded from GEO under accession number GEO: GSE168408. Previously published Ling et al. data was downloaded from the Neuroscience Multi-omic Data Archive (NeMo) (https://assets.nemoarchive.org/dat-bmx7s1t). Previously published Mathys et al. data was downloaded from the AD and aging brain atlas data repository (https://compbio.mit.edu/ad_aging_brain). GTEx data was downloaded from the GTEx portal (https://www.gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression). Gene Ontology Biological Process terms were sourced from https://geneontology.org/. DepMap data came from https://depmap.org/portal/ using version Public 22Q4.

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	Our previous work identified no impact of sex on somatic mutation dynamics so sex was not included as a covariate in this analysis. To avoid sex bias in RNA-seq data interpretation we included males and females in our adult and elderly samples. Our infant samples were both male, but we confirmed our findings in infants using a previously published dataset (Herring et al.) comprising males and females. Additionally, we used a linear model as an alternative method to identify genes that change during aging. We performed this analysis with and without sex included as a covariate and found no difference in the results.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were not considered in our analysis.
Population characteristics	Donor tissue was selected to encompass as wide of a window of human lifespan as possible and a balance of male and female donors. The presence of neuropathology meeting the criteria for a neurodegenerative disease diagnosis or a neurological diagnosis prior death was exclusionary.
Recruitment	n/a
Ethics oversight	All tissue was provided by the National Institutes of Health NeuroBioBank and Banner Sun Health Research Institute Brain and Body Donation Program, which obtained written authorization and informed consent for all donors. Tissue collection and distribution for research purposes was conducted in accordance with protocols approved by the NIH NeuroBioBank (IRB Protocol Number: HM-HP-00042077 or the The Human Brain and Spinal Fluid Resource Center (managed by Sepulveda Research Corporation), IRB Protocol Number: PCC#: 2015-060672, VA Project #: 0002) and Banner Sun Health Research Institute Brain and Body Donation Program (WCG IRB Protocol #20120821). Tissue was collected from post-mortem, deidentified donors and is thus this work is not considered human subjects research by our Institutional Review Board.

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 included 100 nuclei analyzed by scWGS from 19 different donors, 3-9 nuclei per donor. The number of donors and nuclei was informed by our previous works, Lodato et al. 2018 (PMID: 29217584) and Miller et al. 2022 (PMID: 35444284). The same 19 donors were used for the snRNA-seq studies to allow for gene expression and somatic mutation comparisons within each donor, determining the sample size. 8,710-36,921 nuclei were analyzed per donor for snRNA-seq, across 2-3 replicates per donor.

Data exclusions	The following filtering criteria were applied to each sample and cell: more than 100 cells in the sample, reads from mitochondrially-encoded genes less than 5%, and more than 500 expressed gene in the cell. We further filtered samples "5817 200102", "5288 200128", and "5887 PFC 210601" according to their poor consistency with other samples in the following clustering results. To minimize false discovery and focus on universal changes in aging, mitochondrially-encoded genes and genes in sex chromosomes were removed in the downstream analysis.
Replication	snRNA-seq library preparations were performed in duplicate or triplicate in multiple batches, with batches comprising donors from multiple age groups. After sequencing all cells across replicates were analyzed together in one group.
Randomization	Samples were processed in batches comprising young, adult, and elderly samples to control for batch effects.
Blinding	Investigators were not blinded. In order to appropriately control for batch effects and perform analyses to draw conclusions about changes across age, the investigators needed to know the age of each donor included in the study.

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

Antibodies

Antibodies used	MAB377X, mouse anti-NeuN antibody 488 conjugated from Millipore, Donkey anti-Rabbit IgG 647 conjugated (Fisher).
Validation	MAB377X was validated as neuron-specific using single-cell transcriptomics, published in our previous manuscript doi: 10.1038/s41586-022-04640-1. Secondary antibody is a commonly used reagent and was not validated.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Fresh frozen samples were processed using a 7ml glass dounce homogenizer with approximately 20µg tissue in 5 ml of filter-sterilized tissue lysis buffer (0.32M sucrose, 5mM CaCl2, 3mM MgAc2, 0.1mM EDTA, 10mM Tris-HCl (pH 8), 0.1% Triton X-100, 1mM fresh DTT). The homogenized solution was loaded on top of a filter-sterilized sucrose cushion (1.8M sucrose, 3mM MgAc2, 10mM Tris-HCl (pH 8), 1mM DTT) and spun in an ultracentrifuge in an SW28 rotor (13,300 RPM, 2hrs, 4□C) to separate nuclei. Supernatant was removed and nuclear pellets were resuspended in ice-cold resuspension buffer (8.5mL 1xPBS/3mM MgCl2 + 1mL 1xPBS/3mM MgCl2/1% BSA + 500µL Sucrose Cushion), filtered with a 40µm cell strainer, then stained with anti-NeuN antibody (directly conjugated to Alexa Fluor 488; Millipore cat. No. MAB377X, clone A60; 1:1,250) and an anti-rabbit IgG Alexa Fluor 647 antibody as a negative control for 30min. Using a BD biosciences FACSAria Fusion machine and BD FACSDiva Software, forward scatter A (FSC-A) was first used to isolate large non-replicating cells. This non-replicating high-FSC-A plus high-NeuN population was confirmed to comprise an excitatory neuron population, comprising 2–5% of the total population of nuclei in each sample.
Instrument	BD biosciences FACSAria Fusion
Software	BD FACSDiva™ Software
Cell population abundance	Our targeted large, excitatory neuron population comprised 2-5% of the total sample
Gating strategy	Nuclei were gated away from debris using side scatter (SSC) and forward scatter (FSC). Doublets were gated out using FSC-area vs. FSC-height and SSC-area vs. SSC_height. NeuN staining produced a bimodal signal distribution, distinguishing NeuN+ and NeuN– nuclei. Large neuronal nuclei, representing excitatory pyramidal neurons, were further identified by collecting the nuclei with highest NeuN signal among the NeuN+ neuronal fraction, and gating for the population with the highest FSC-A signal and excluding Alexa-Fluor-647-high events.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.