nature portfolio

https://doi.org/10.1038/s41586-026-10271-7

Supplementary information

Lifespan single-cell transcriptomic atlas of the human prefrontal cortex

In the format provided by the authors and unedited

Supplementary information

Lifespan single-cell transcriptomic atlas of the human prefrontal cortex

Hui Yang[1,2,3,4*], Tereza Clarence[1,2,3,4*], Madeline R. Scott[8*], Xinyi Wang[1,2,3,4], Prashant N.M.[1,2,3,4], Milos Pjanic[1,2,3,4], Sanan Venkatesh[1,2,3,4], Aram Hong[1,2,3,4], Clara Casey[1,2,3,4], Sarah R. Murphy[1,2,3,4], Alexander Kawah Yu[1,2,3,4], Zhiping Shao[1,2,3,4], Marcela Alvia[1,2,3,4], Stathis Argyriou[1,2,3,4], PsychAD Consortium[12], Nadejda M. Tsankova[2,5,7], Pavan K. Auluck[11], Stefano Marenco[11], Vahram Haroutunian[2,3,5,10], Georgios Voloudakis[1,2,3,4,6,9,10], Jaroslav Bendl[1,2,3,4], Colleen A. McClung[8], Donghoon Lee[1,2,3,4], John F. Fullard[1,2,3,4], Gabriel E. Hoffman[1,2,3,4,9,10], Kiran Girdhar[1,2,3,4 *#], Panos Roussos[1,2,3,4,9,10*#]

Affiliations

[1]Center for Disease Neurogenomics, [2]Friedman Brain Institute, [3]Department of Psychiatry, [4]Department of Genetics and Genomic Science, [5]Department of Neuroscience, [6]Department of Artificial Intelligence and Human Health, [7]Department of Pathology, Molecular, and Cell-Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA,

[8]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[9]Center for Precision Medicine and Translational Therapeutics, [10]Mental Illness Research, Education and Clinical Center VISN2, James J. Peters VA Medical Center, Bronx, NY, USA.

[11]Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, MD, USA.

[12]PsychAD Consortium. psychAD_list_authors
*These authors contributed equally to this work
Correspondence to: K.G. (kiran.girdhar@mssm.edu) and
P.R. (panagiotis.roussos@mssm.edu)

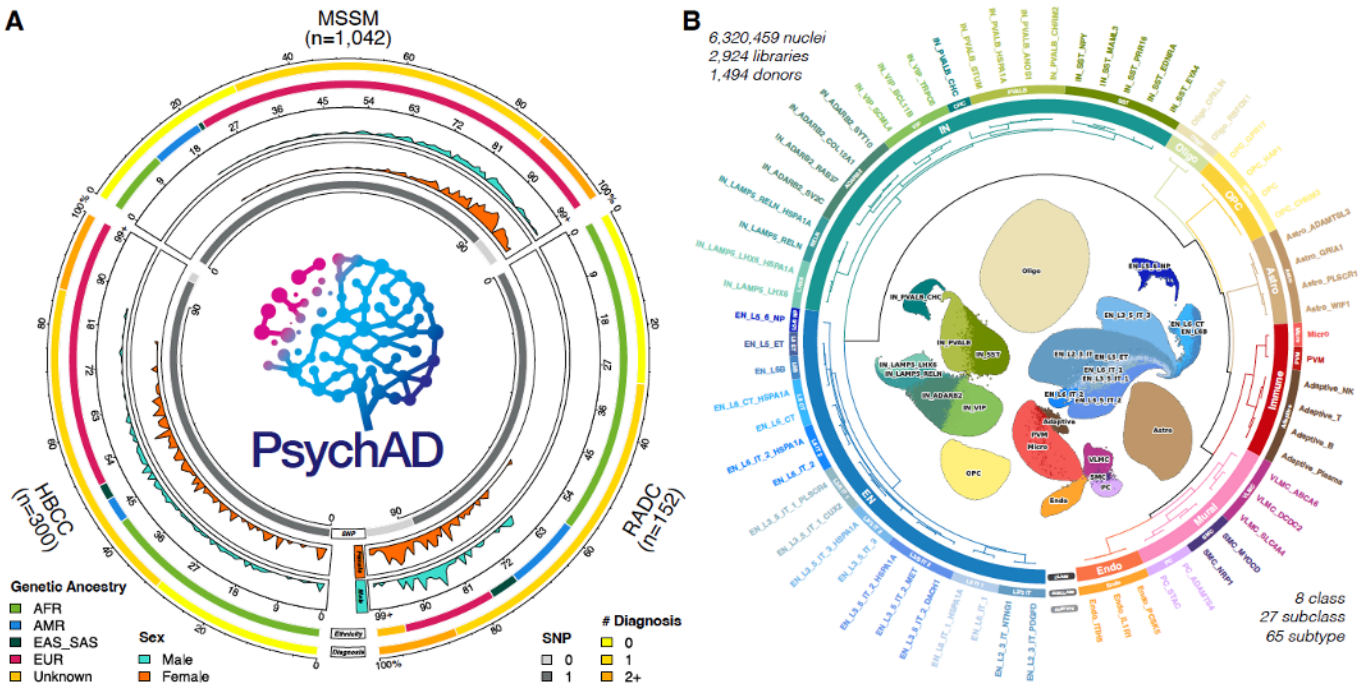
Supplementary Notes

The PsychAD dataset : To enhance our ability to identify shared and distinct molecular pathways, as well as the causal variants and genes involved in various diseases of the brain, we generated a population-scale single-cell gene expression dataset in the postmortem human prefrontal cortex. The PsychAD cohort comprises 1,494 donors, each of which was subjected to single nucleus RNA-seq (snRNA-seq), generating more than 6.3 million individual nuclei (**Supplementary Fig. N1A**). Human brain specimens were obtained from three sources, Mount Sinai NIH Brain Bank and Tissue Repository (MSSM; 1,042 samples), NIMH-IRP Human Brain Collection Core (HBCC; 300 samples), and Rush Alzheimer's Disease Center (RADC; 152 samples). For 1,381 (92%) donors, we also provide harmonized genotype data (either or both SNP array and whole genome sequencing). Importantly, less than 5% of the samples from either the MSSM[1] or RADC[2,3] cohort had previously been subjected to snRNA-seq. The PsychAD cohort consists of 1,074 donors affected by over 30 different disorders, including three represented by more than 100 cases (AD (n = 519), SCZ (n = 177), and DLBD (n = 112)) and three by at least 40 (vascular dementia (n = 85), BD (n = 72) and PD (n = 48)). In addition, the sample set also includes 420 neurotypical controls, as well as a number of cases with relatively understudied conditions, such as obsessive-compulsive disorder (n = 6), amyotrophic lateral sclerosis (n = 5), and progressive supranuclear palsy (n = 5). An important component of our data is the availability of phenotypic information on the nature and prevalence of neuropsychiatric symptoms (NPS). NPS frequently accompany AD and related dementias, and it has been estimated that, throughout the course of the disease, more than 80% of individuals will exhibit at least one NPS that significantly impacts their clinical outcomes[4]. So far, various studies have examined population data to characterize NPS along the AD continuum[5–7]. For example, depression and apathy are often the most observed symptoms in the early stages of AD, while delusions, hallucinations, and aggression become more prevalent as the disease advances[5]. However, research into the molecular basis of these NPS remains scarce, and we believe that our dataset provides a unique opportunity to explain their role in AD at a more granular level, leading to a better understanding of the broader disease and the potential identification of novel therapeutic targets.

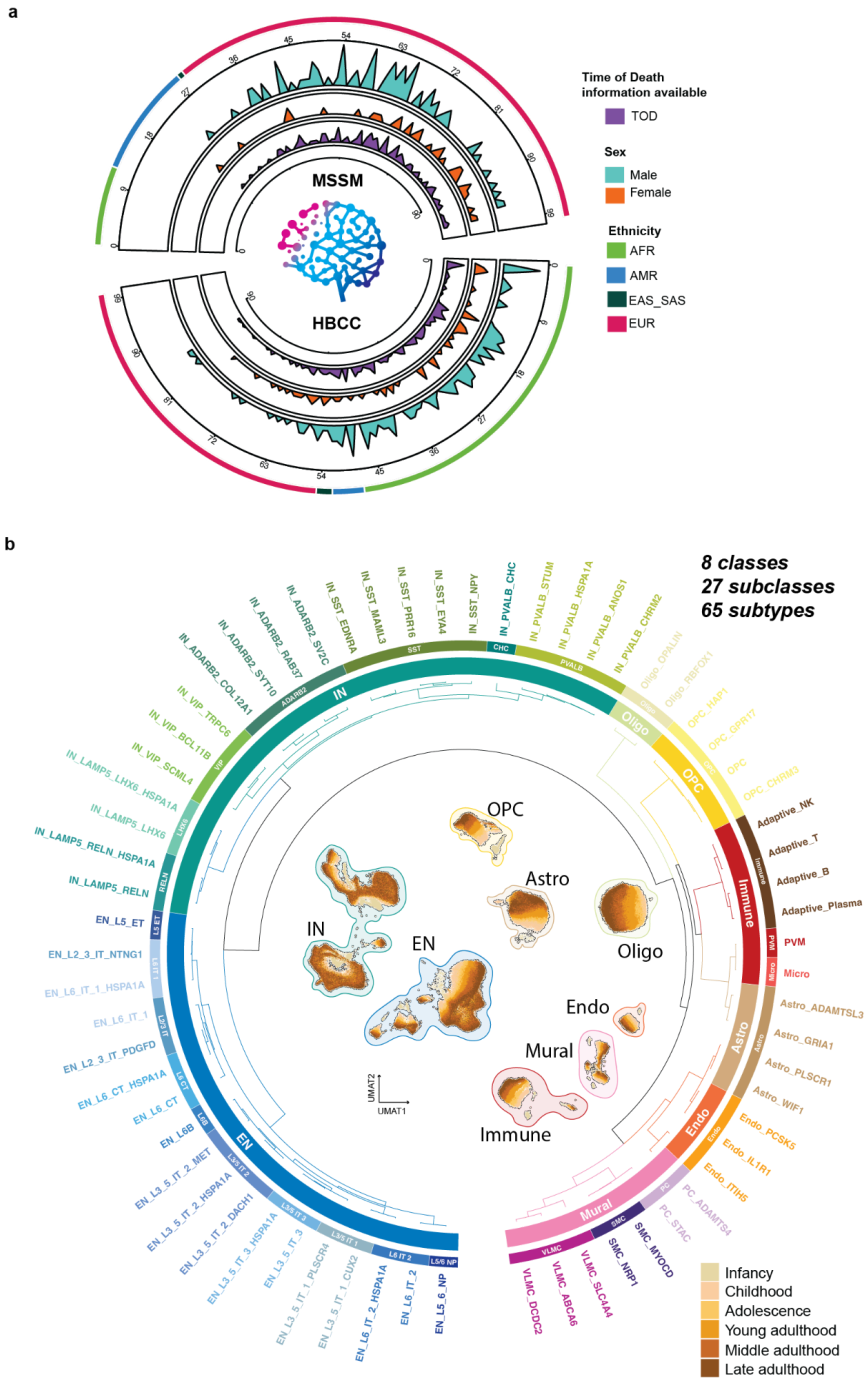
To understand complex and heterogeneous human cortical tissues in a disease context, we first generated a robust cell type taxonomy that is invariant to the aging lifespan, disease phenotypes, and various sampling and technical biases. Following unified computational processing, quality controls, and batch normalization of more than 6.3 million individual nuclei, we used the cellular taxonomy of the primate DLPFC[8] and human primary motor cortex[9] as a baseline reference to annotate the cell types of the human DLPFC. The resulting cellular taxonomy was organized using three levels of taxonomic hierarchy, consisting of 8 broad cell classes, 27 subclasses, and 65 functional subtypes (**Supplementary Fig. N1B**). Each level of the hierarchy represents different slices of the clustering dendrogram. The top-level (class) defines 8 major cell types, including two broad neuronal cell types: glutamatergic excitatory (EN) and GABAergic inhibitory neurons (IN), three glial (Astrocyte, Oligodendrocyte, OPC), and three non-neuronal cell types (Immune, Mural, Endothelial). Subsequent levels (subclass and subtype) were derived by iteratively re-clustering the subset of cells by gene matrix using a new set of variable genes relevant to the particular cell type. For example, the various EN subclasses were distinguished by their laminar organization and axon projection characteristics (IT: intra-telencephalic, ET: extra-telencephalic, NP: near projecting, CT: corticothalamic, and L6B), while IN subclasses were denoted using characteristic marker genes. This approach allowed us to identify 10 subclasses of EN and 7 subclasses of IN. The 10 EN subclasses were further subdivided into 18 functionally distinct subtypes, while the 7 IN subclasses comprised 21 subtypes.

The release of this dataset by the PsychAD consortium is accompanied by a series of manuscripts that are submitted as a package to Nature Portfolio journals describing the cross-disorder analysis of transcriptomic vulnerability (Lee *et al*, Single-cell atlas of transcriptomic vulnerability across multiple neurodegenerative and neuropsychiatric diseases), genetic regulation of gene expression (Zeng *et al*, Single nucleus, multi-ancestry

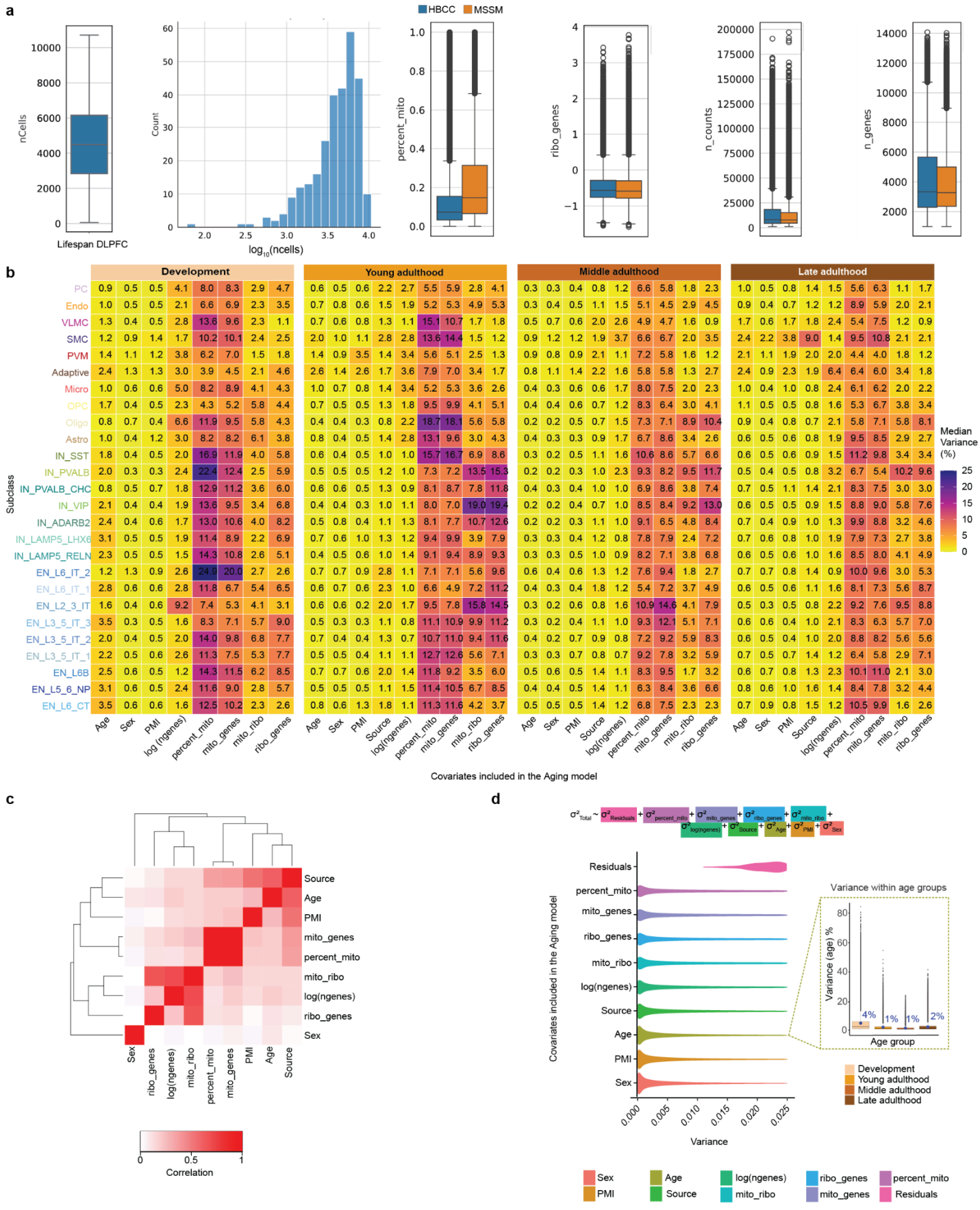
atlas of genetic regulation of gene expression in the human brain), and transcriptome-wide association studies (Venkatesh *et al*, Single-nucleus transcriptome-wide association study of human brain disorders). The consortium has also leveraged neurotypical controls to assemble a map of transcriptomic changes across the lifespan (Yang *et al*, A lifespan transcriptomic atlas of the human prefrontal cortex at single-cell resolution). Also, we performed two advanced analyses using emerging machine learning and AI approaches to detect phenotype-associated cells revealing potential novel cell subpopulations and expressed genes (PASCode: He *et al*, Phenotype Scoring of Population Scale Single-Cell Data Dissects Alzheimer's Disease Complexity) and quantify personalized importance scores of genes, cell types and regulatory networks for various PsychAD phenotypes (iBrainMAP: Chandrashekar *et al*, Single-cell transcriptomic dissection of 1,494 human brains reveals personalized functional genomics for Alzheimer's disease phenotypes). Furthermore, the computational scale and diversity of the generated data led to the development of several analytical tools, including *Dreamlet* for differential gene expression[10] and *Crumblr* for the identification of cell types with significant shifts in cell type proportions between disease cases and controls (Hoffman *et al*, Fast, flexible analysis of compositional data with crumblr). Lastly, we have prepared a separate manuscript (Fullard *et al*, Population-scale cross-disorder atlas of the human prefrontal cortex at single-cell resolution) that offers a comprehensive overview of the clinical and demographic donor information, as well as detailed descriptions of the techniques used in snRNA-seq and SNP array sample preparation, including bioinformatics preprocessing and quality control methods applied to the resulting data. The PsychAD dataset is publicly available at the AMP-AD Knowledge Portal in the study-specific folder https://doi.org/10.7303/syn60084804. This repository includes sample metadata, as well as raw and processed sequencing data for snRNA-seq and genotyping. snRNA-seq data can be inspected using the CELLxGENE visualization tool at https://cellxgene.cziscience.com/collections/84ce6837-548d-4a1f-919f-0bc0d9a3952f.



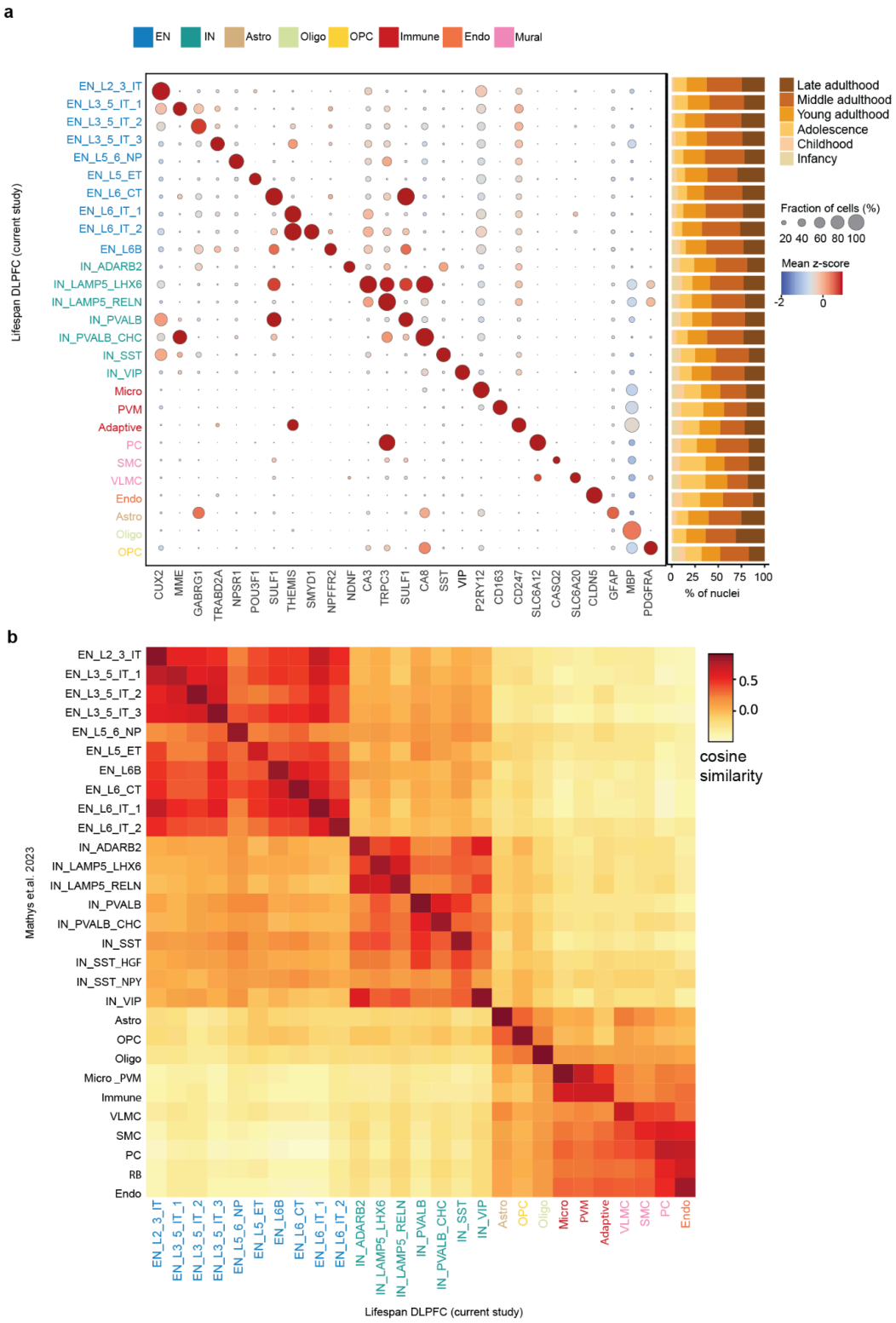
Supplementary Figure N1. The PsychAD cohort. A, Organization of the PsychAD study cohort. Breakdown of brain donors by tissue repository, number of neurological diagnoses, genetic ancestry, age, sex, and availability of genotype data. **B,** Unified processing of the single-cell transcriptomics atlas and hierarchical structure of cellular taxonomy of the overall transcriptome. Taxonomic annotation of cellular phenotype at three levels of granularity; class (n = 8), subclass (n = 27), and subtype (n = 65).



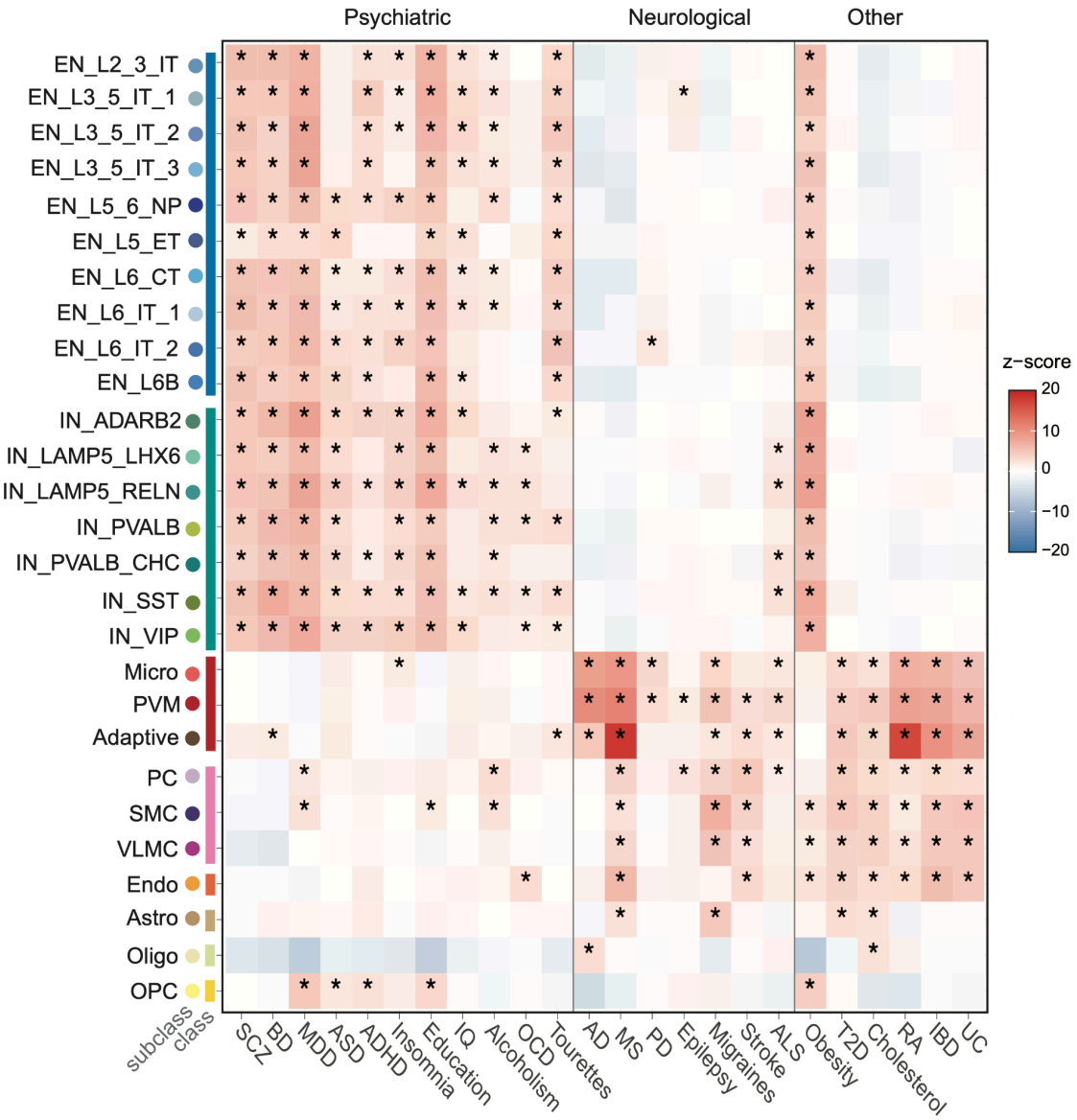
Supplementary Figure 1. Donor demographics and cell taxonomy in lifespan DLPFC dataset. a, Circos plot depicting demographic information of available donors. The plot is split by donors from the MSSM (top) and HBCC (bottom) brain banks, and, from outside-in, indicates: ethnicity (AFR: African, AMR: Ad Mixed American, EAS: East and South Asian, SAS: South Asian, EUR: European), sex (orange Female and cyan Male) and availability of metadata for time of death (purple) are represented as a distribution across age of donors. **b,** Transcriptomic similarities across class, subclass and subtype taxonomic levels. UMAP of maturation (UMAT) representation of the dataset is positioned within the taxonomy circos plot, colored by age group. The color of the halo encircling each cell class represents class category membership.


a
HBCC
MSSM
nCells
Lifespan DLPFC
Count
log10(ncells)
percent_mito
ribo_genes
n_counts
n_genes
b
Development
Young adulthood
Middle adulthood
Late adulthood
Subclass
PC
Endo
VLMC
SMC
PVM
Adaptive
Micro
OPC
Oligo
Astro
IN_SST
IN_PVALB
IN_PVALB_CHC
IN_VIP
IN_ADARB2
IN_LAMP5_LHX6
IN_LAMP5_RELN
EN_L6_IT_2
EN_L6_IT_1
EN_L2_3_IT
EN_L3_5_IT_3
EN_L3_5_IT_2
EN_L3_5_IT_1
EN_L6B
EN_L5_6_NP
EN_L6_CT
Age
Sex
PMI
Source
log(ngenes)
percent_mito
mito_genes
mito_ribo
ribo_genes
Median Variance (%)
Covariates included in the Aging model
c
Correlation
d
Covariates included in the Aging model
Variance
Variance within age groups
Variance (age) %
Age group
4%
1%
1%
2%
Development
Young adulthood
Middle adulthood
Late adulthood
Residuals
percent_mito
mito_genes
ribo_genes
mito_ribo
log(ngenes)
Source
Age
PMI
Sex

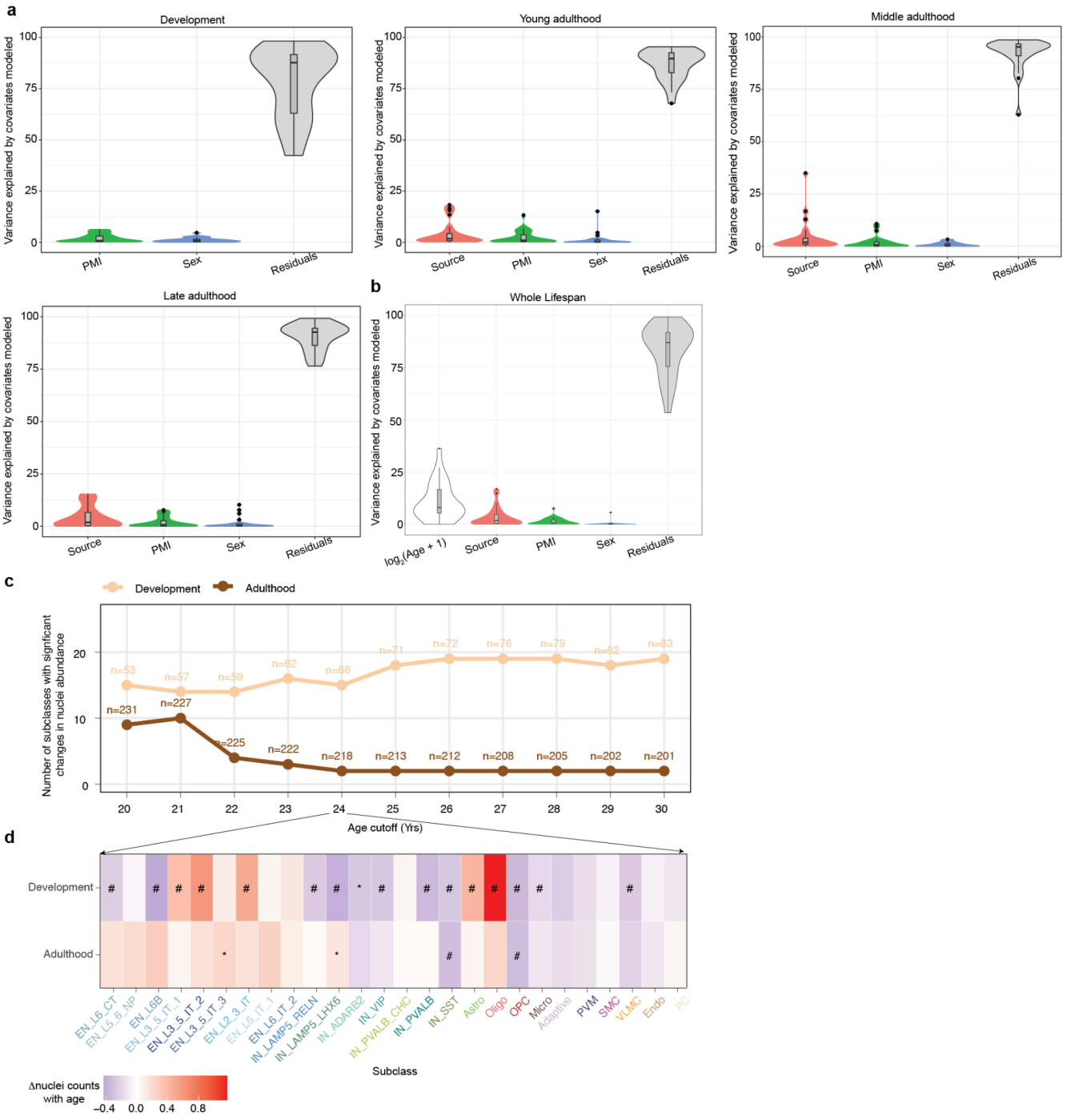
Supplementary Figure 2. Quality control of snRNA-seq and covariate modeling for pseudobulk gene expression analyses. a, Distribution of the number of nuclei per donor (first plot) and log-transformed nuclei per sample (second plot). The third to sixth plots show brain bank-specific distributions for percent mitochondrial reads, ribosomal gene expression, total UMI counts, and total detected genes (HBCC in blue, MSSM in orange). **b,** Heatmap showing the median percent variance explained by each covariate (x-axis) across subclasses (y-axis), stratified by age group. These covariates were included in all age associated and circadian genes analyses done at pseudobulk gene expression level. **c,** Heatmap of pairwise correlation of identified covariates and age. **d,** Combined variance partition results (from development, young, middle and late adulthood) depicting respective contributions to variance in gene expression explained by listed covariates (Age, source (brain bank), PMI (post mortem interval), mitochondrial and ribosomal properties (percent_mito, mito_genes, mito_ribo, ribo_genes), log(n_genes) and residuals), ordered from highest to lowest contribution. Highlighted inset boxplot shows distribution of age-contributed variance within the respective age groups. Boxplot in inset shows lower and upper hinges at the 25th and 75th percentiles, with whiskers extending to at most 1.5 times the interquartile range (IQR).



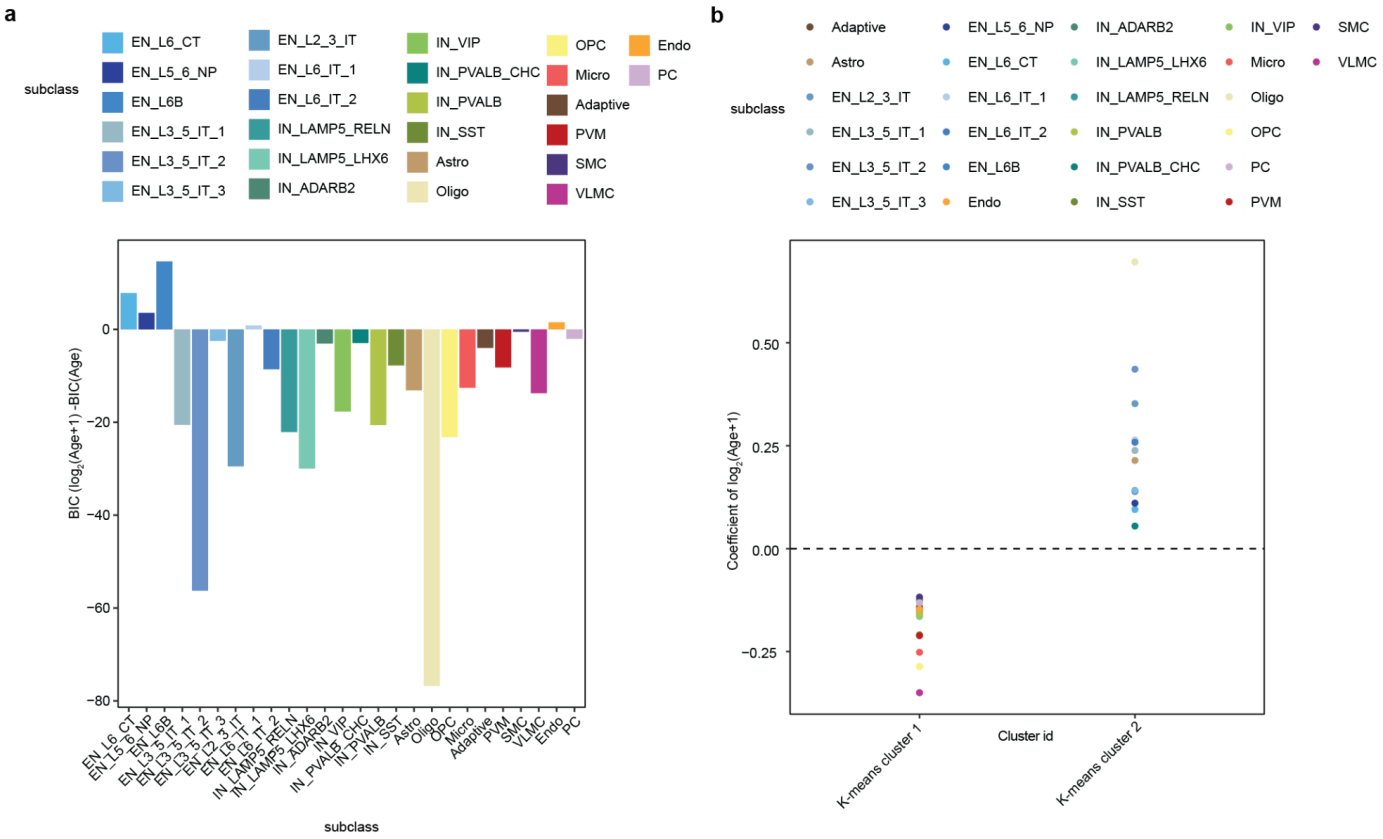
Supplementary Figure 3. Subclass specific marker validation and cross-dataset similarity analysis. a, Dot plot showcasing selected key markers for each subclass, along with the fraction of nuclei observed across different age groups. **b,** Cosine similarity analysis comparing subclass-level cellular taxonomy of the lifespan DLPFC atlas (current study) with the Mathys et al. 2023 dataset.



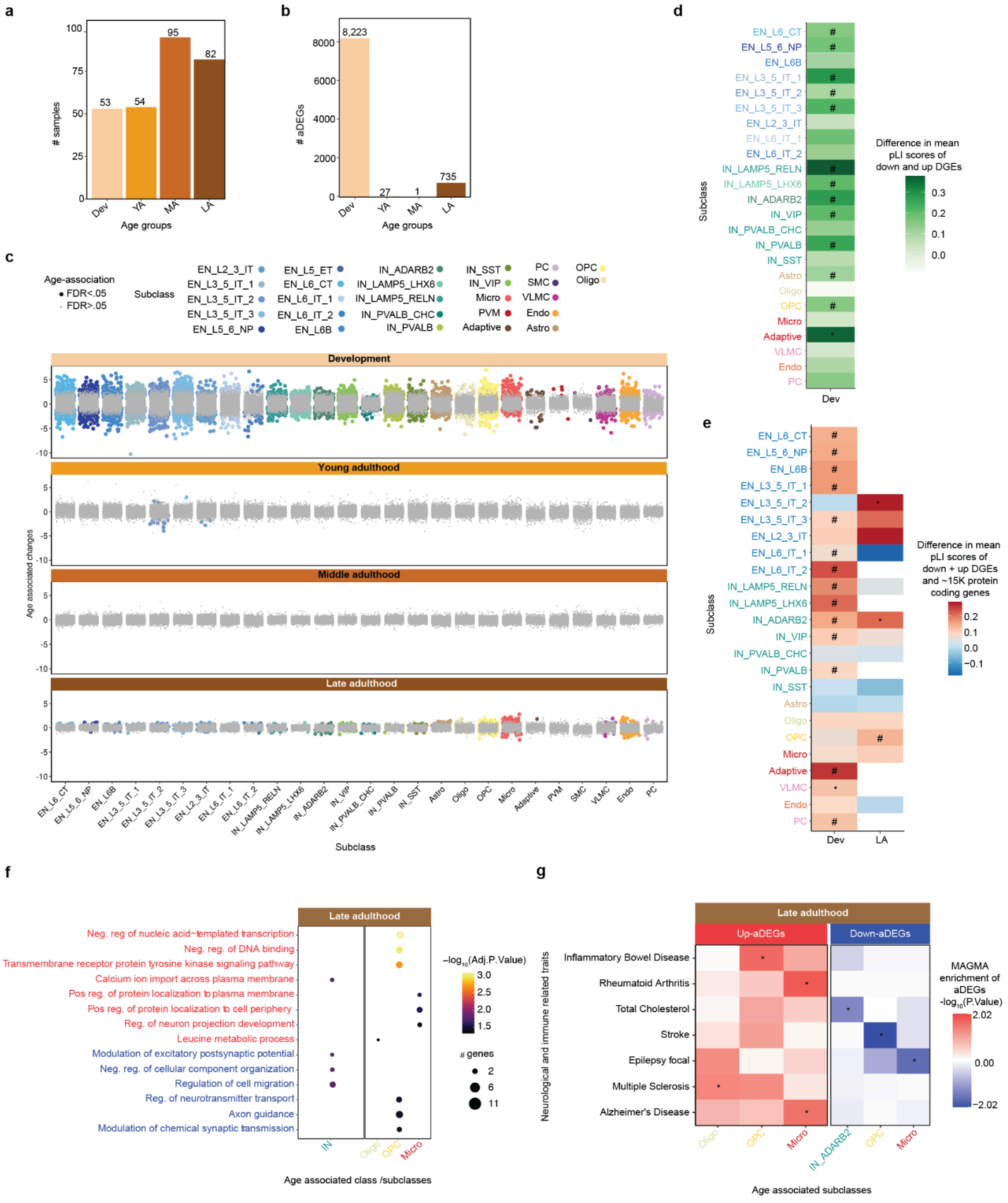
Supplementary Figure 4. Polygenic diseases risk scores. Disease scores computed using the scDRS tool, depicted at the subclass level across various neurological, psychiatric and other selected GWAS traits, presented as the mean association z-score. "*": indicates significance for enrichment in FDR-corrected P-value.



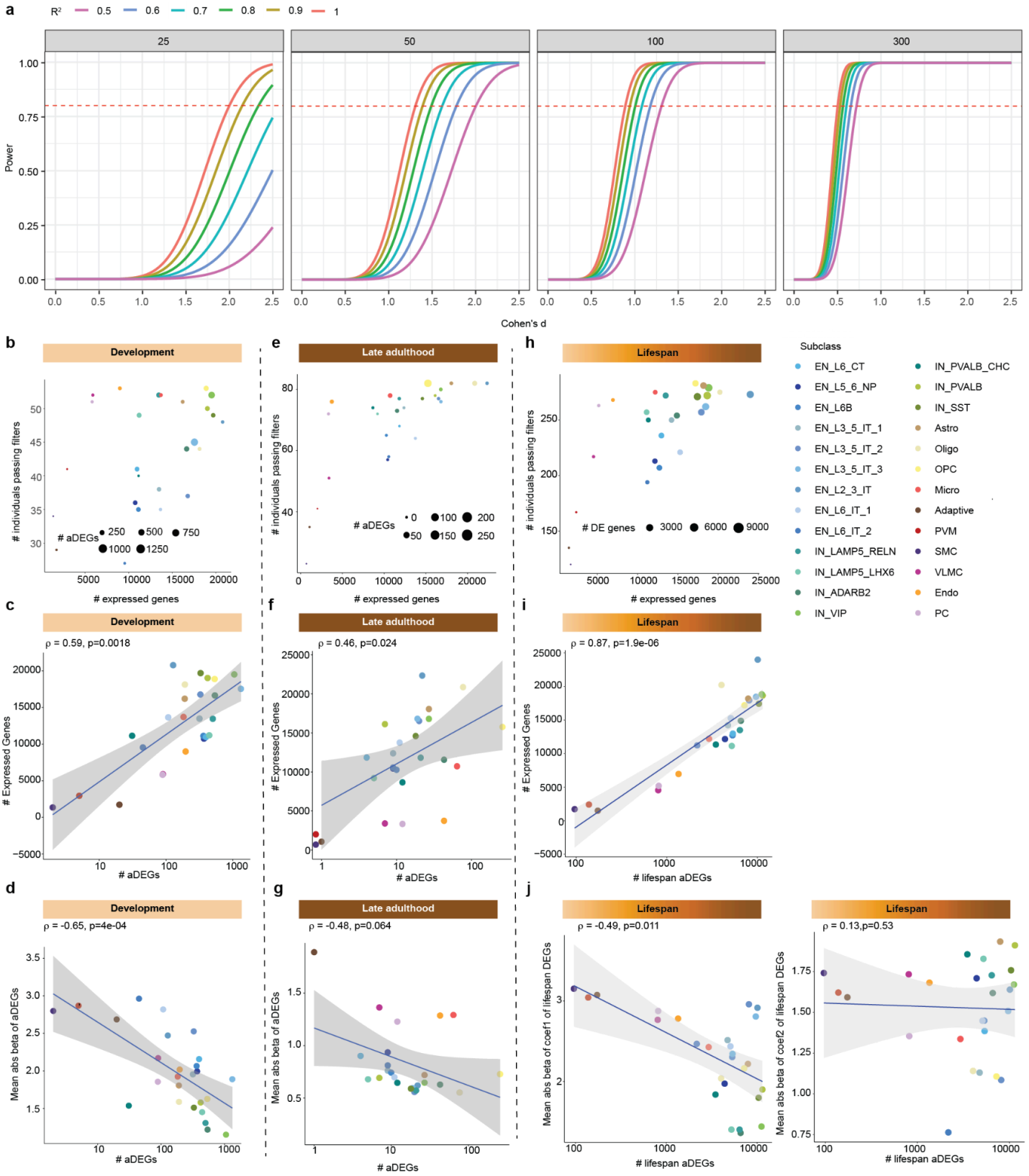
Supplementary Figure 5: Variance explained by modeled covariates in nuclei abundances across the lifespan and per age group. a-b, Violin plots show the distribution of variance explained by individual covariates across all subclasses, as estimated using the variancePartition. Plots are shown for four age groups in a panel and the entire lifespan in b panel. **c,** Number of subclasses showing significant age-related changes in nuclei abundance (FDR < 0.05) as a function of different age cutoffs used to define "development" (< cutoff) vs "adulthood" (≥ cutoff). Sample sizes (n) for each group at each cutoff are added as text. **d,** Heatmap showing estimated effect size of age-associated changes in nuclei abundance across two groups when data are split at age 24 years. * and # indicate nominal p-value < 0.05 and FDR < 0.05, respectively.



Supplementary Figure 6. Optimal model selection for nuclei counts. a, Bayesian Information Criterion (BIC) obtained from modeling nuclei counts using the formulas: nuclei counts ~ $\log_2$(Age+1) and nuclei counts ~ Age, utilizing crumblr for each subclass. **b,** K-means clustering coefficient from nuclei counts ~ $\log_2$(Age+1).

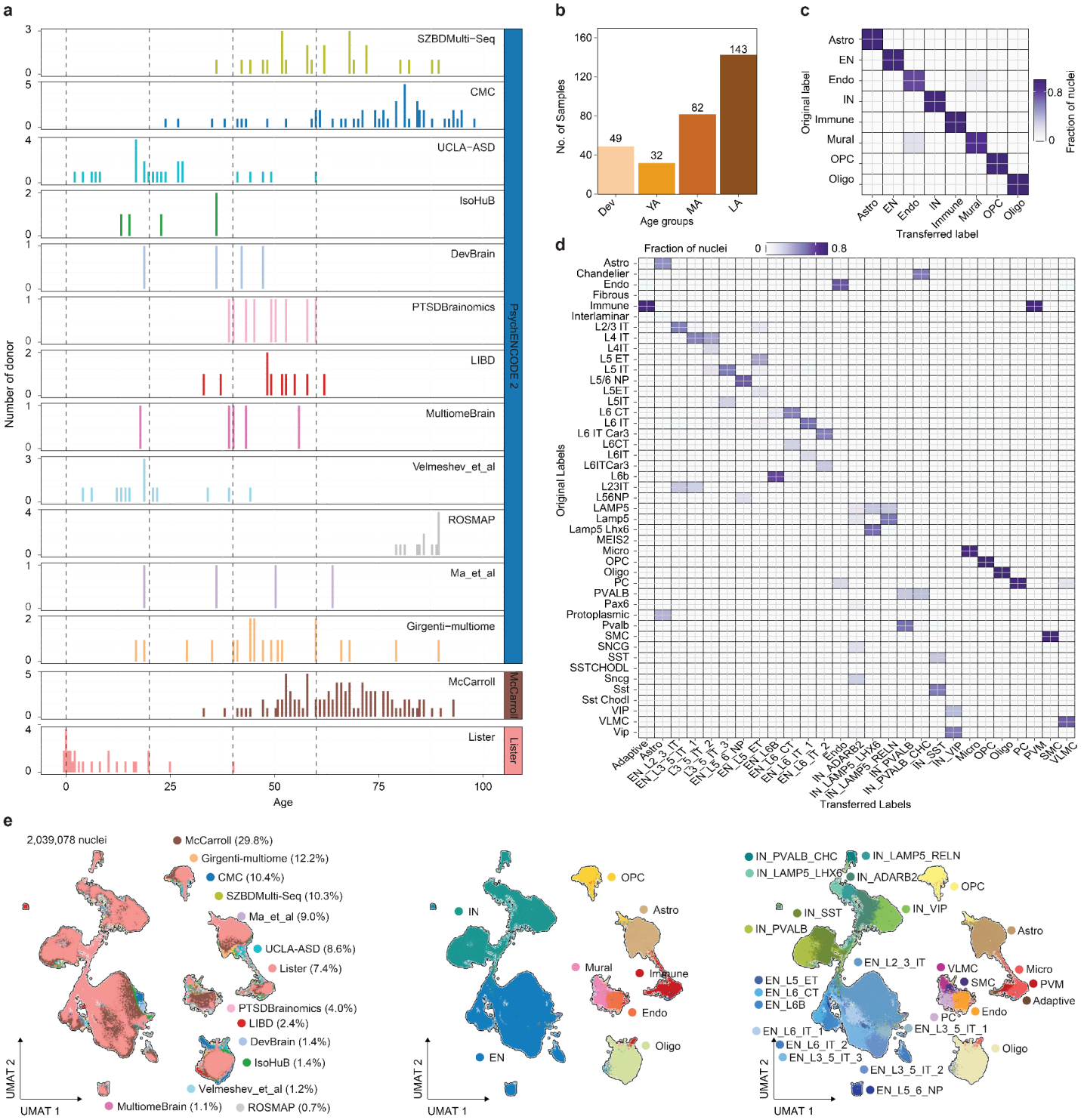

a
samples
Age groups
Dev
YA
MA
LA
53
54
95
82
b
aDEGs
8,223
27
1
735
c
Age-association
FDR<.05
FDR>.05
Subclass
Development
Young adulthood
Middle adulthood
Late adulthood
Age associated changes
Subclass
d
Subclass
Difference in mean pLI scores of down and up DGEs
e
Subclass
Difference in mean pLI scores of down + up DGEs and ~15K protein coding genes
f
Late adulthood
Neg. reg of nucleic acid−templated transcription
Neg. reg. of DNA binding
Transmembrane receptor protein tyrosine kinase signaling pathway
Calcium ion import across plasma membrane
Pos reg. of protein localization to plasma membrane
Pos reg. of protein localization to cell periphery
Reg. of neuron projection development
Leucine metabolic process
Modulation of excitatory postsynaptic potential
Neg. reg. of cellular component organization
Regulation of cell migration
Reg. of neurotransmitter transport
Axon guidance
Modulation of chemical synaptic transmission
−log10(Adj.P.Value)
genes
Age associated class /subclasses
g
Late adulthood
Up-aDEGs
Down-aDEGs
Inflammatory Bowel Disease
Rheumatoid Arthritis
Total Cholesterol
Stroke
Epilepsy focal
Multiple Sclerosis
Alzheimer's Disease
Neurological and immune related traits
MAGMA enrichment of aDEGs -log10(P.Value)
Age associated subclasses

Supplementary Figure 7. Age-associated gene expression across age groups. a, Number of samples and **b,** the number of DEGs across four age groups. **c,** Beeswarm plot of age-associated effect sizes for 26 subclasses across four age groups. Gray dots are genes with FDR > 0.05 and colored dots are genes with FDR < 0.05. **d,** Heatmap showing the mean difference in pLI scores between down-aDEGs and up-aDEGs during development. Colors represent the direction and magnitude of mean pLI differences for each subclass. * indicates nominal p-value < 0.05 (Wilcoxon rank-sum test) and # indicates FDR-adjusted p-values < 0.05 across all subclasses. **e,** Heatmap to show differences in loss-of-function mutation scores (pLI) of down+up-aDEGs and ~15K protein coding genes during development and late adulthood. * and # indicate nominal p-value < 0.05 and FDR < 0.05 from Wilcoxon test. **f,** Functional pathway analysis of up and down-aDEGs during late adulthood for subclasses that were significantly enriched for GO biological processes with an adjusted p-value < 0.05. **g,** Association of up and down-aDEGs during late adulthood with risk genes for neurological and immune-related traits using MAGMA.* and # indicate nominal p-value < 0.05 and FDR < 0.05 from MAGMA enrichment.

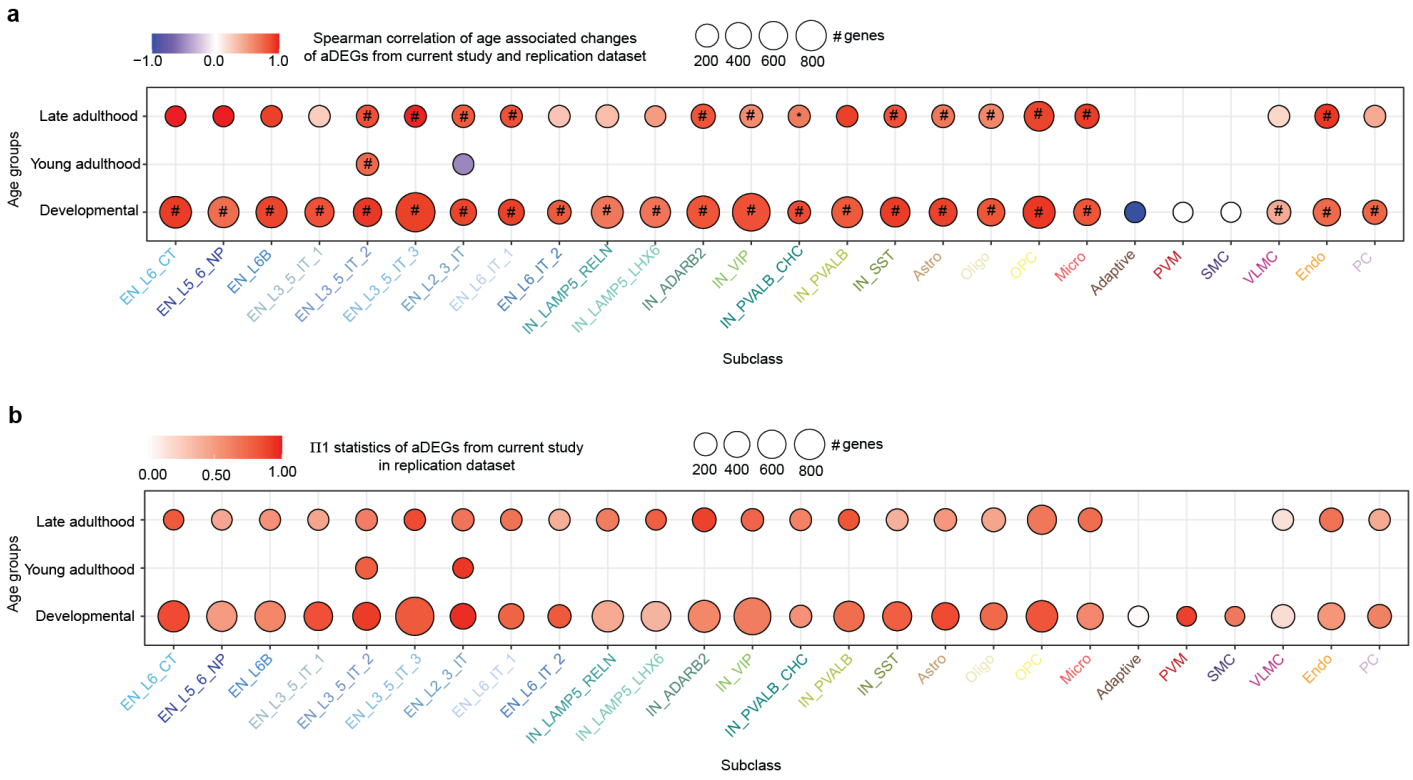


Supplementary Figure 8. Theoretical and empirical assessment of statistical power across subclasses. a, Theoretical power curves illustrating the relationship between effect size (Cohen's *d*) and statistical power across sample sizes (n = 25, 50, 100, 300) and moderated R^2 values (0.5–1.0). The horizontal dashed line indicates the 80% power threshold. **b,** Scatter plots showing the number of

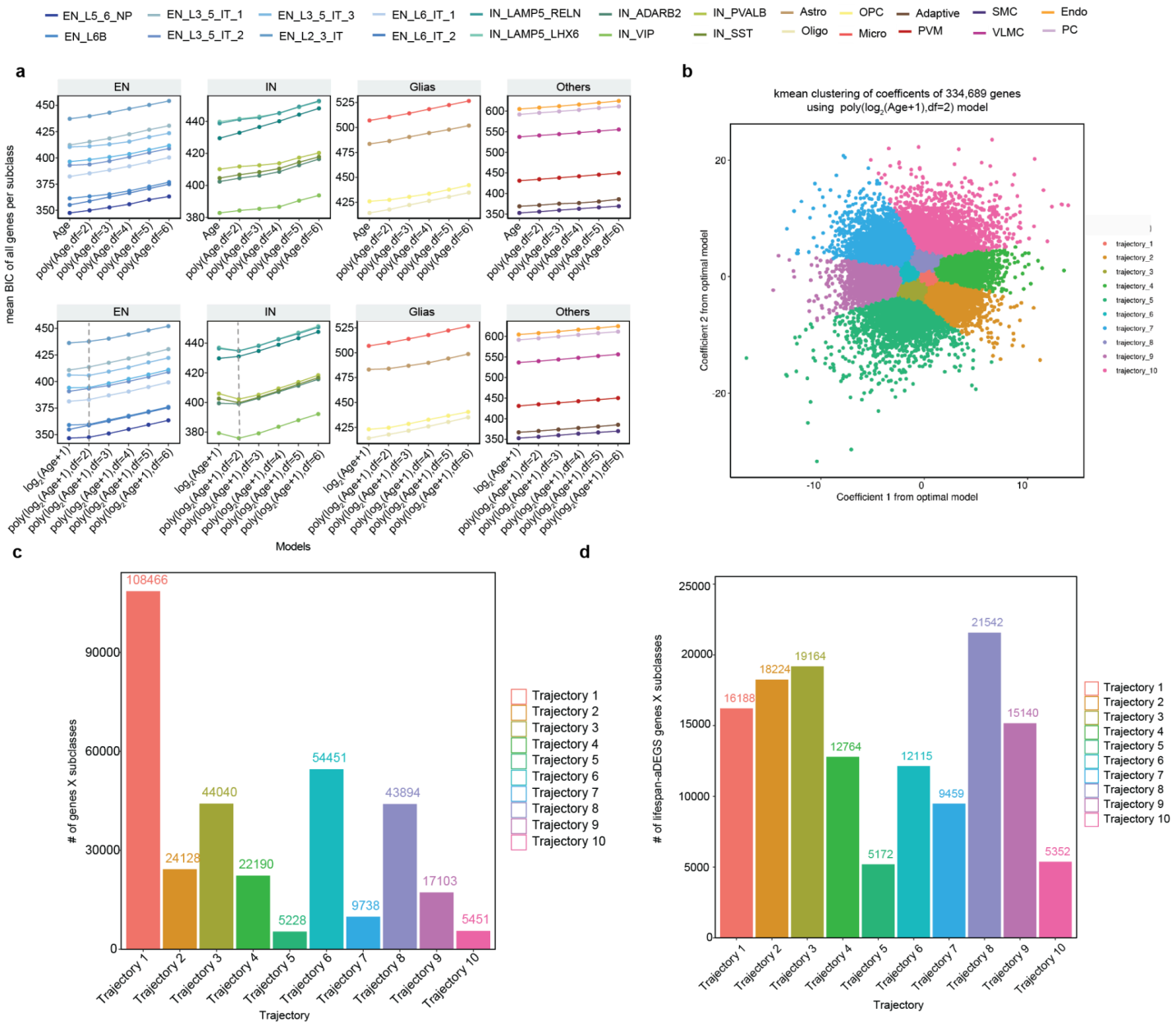
expressed genes and number of developmental donors passing quality control for each subclass (derived from dreamlet pseudobulk residuals). **c–d,** Spearman correlations between aDEG developmental counts and (c) number of expressed genes and (d) mean absolute age-associated expression changes during development per subclass. **e**, Same as panel b, but based on late adulthood donors. **f–g**, Spearman correlations between late adulthood aDEG counts and (f) number of expressed genes and (g) mean absolute age-associated expression changes during late adulthood per subclass. **h**, Same as panel b, but based on lifespan donors. **i–j**, Spearman correlations between lifespan aDEG counts and (i) number of expressed genes and (j) mean absolute age-associated expression changes during lifespan per subclass. In panel j, the left plot shows correlation with the linear age coefficient ($coef_1$), and the right plot shows correlation with the non-linear ($coef_2$) coefficient. Points in b, e and h are colored by subclass, and point size reflects the number of aDEGs during development, late adulthood and lifespan respectively.



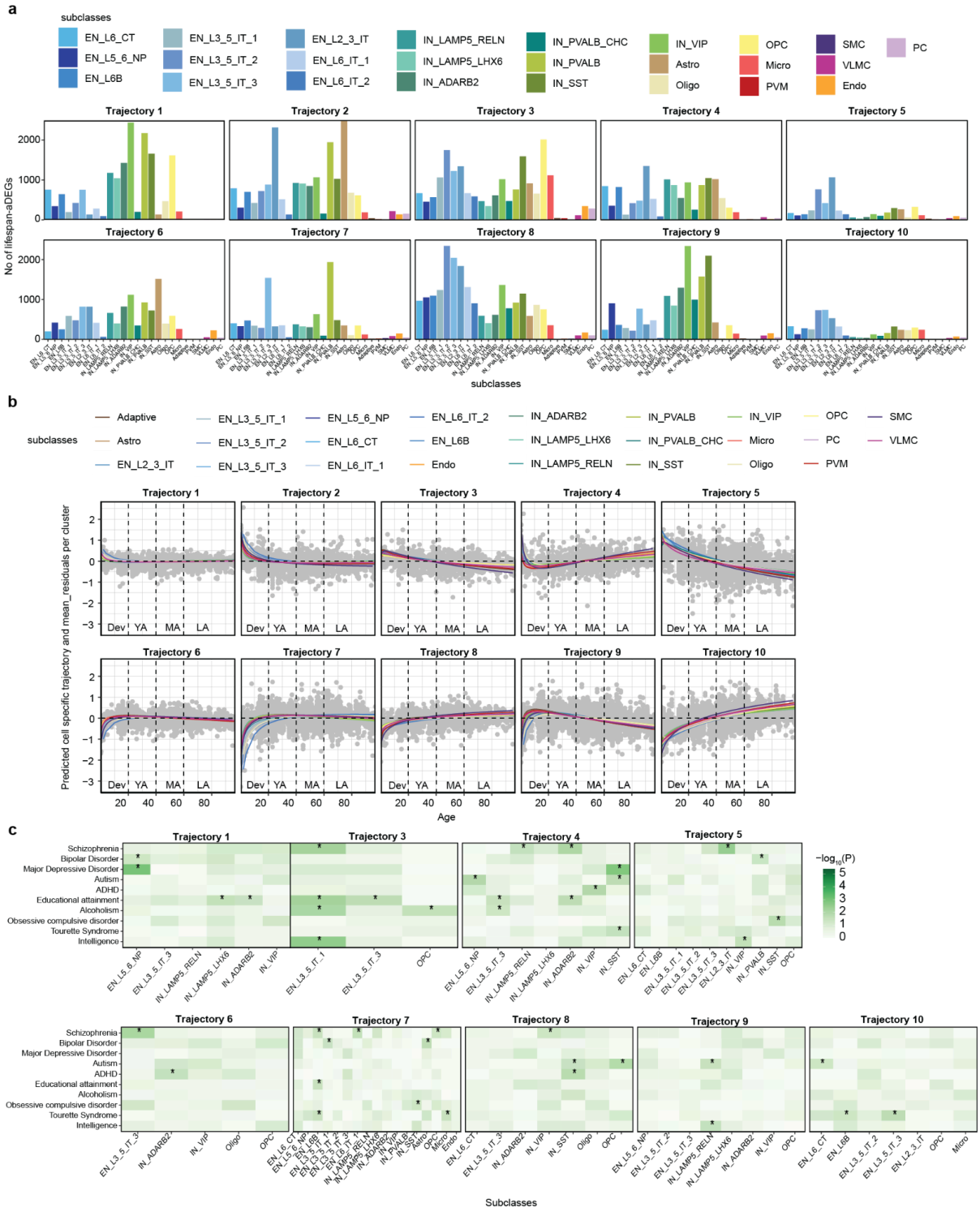
Supplementary Figure 9. Lifespan replication dataset. a, Age distribution of donors (x-axis) across 14 published cohorts (y-axis) included in the lifespan replication dataset across full human lifespan. **b**, Number of donor samples per age group included in the replication dataset. **c-d,** Concordance of class in c and subclass-level label transfer in d. Each cell shows the fraction of nuclei with a given original label assigned to each transferred class in c and subclass in d label confirming reproducibility of fine-grained annotations across datasets. **e**, UMAP plots of over 2 million nuclei from the replication dataset, colored by dataset of origin with the percentage of total nuclei contributed by each dataset (left), six broad cell types (middle), and subclass labels (right).



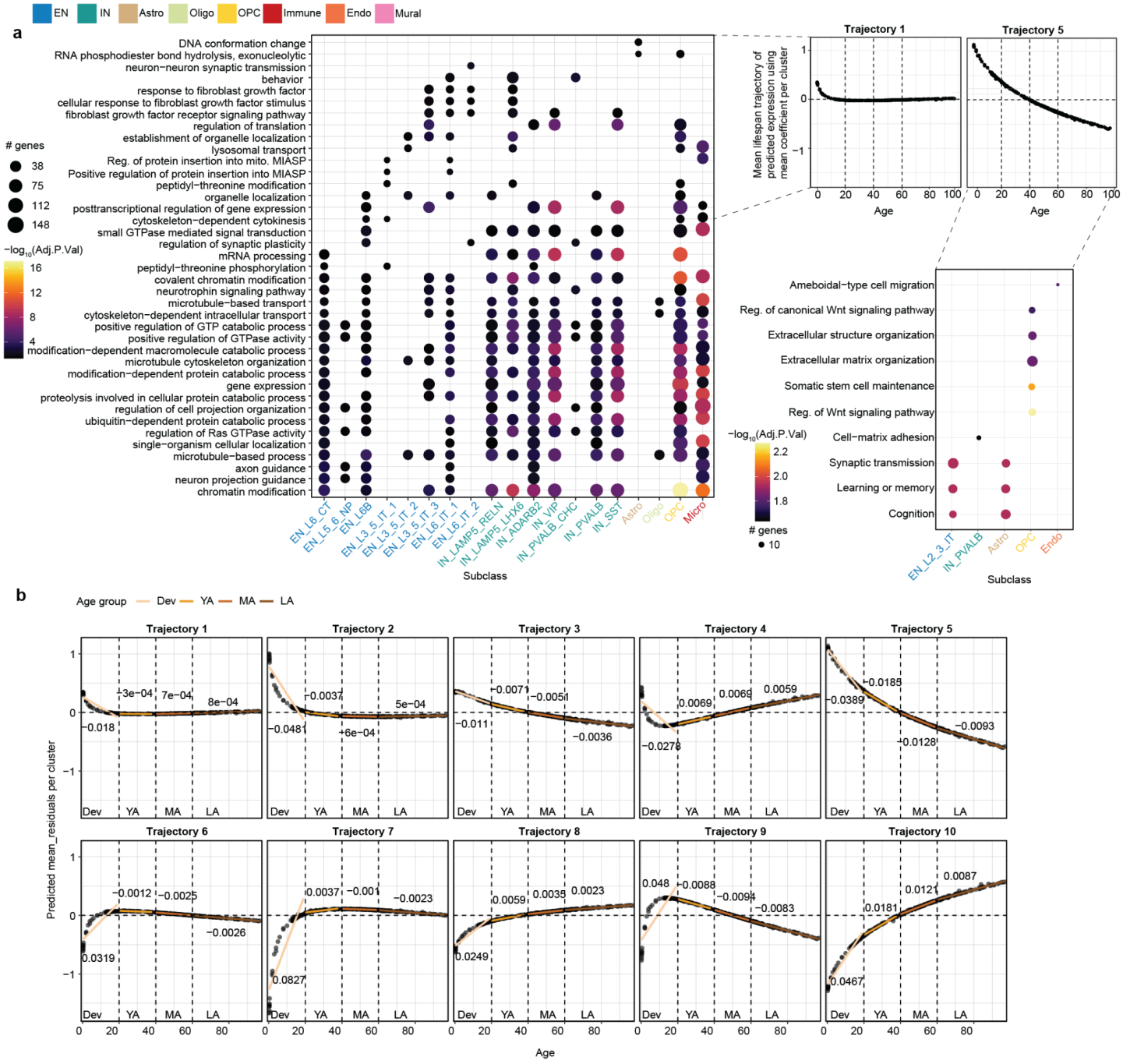
Supplementary Figure 10. Replication of age-associated gene expression changes across subclasses and age groups. a, Bubble plot displaying the Spearman correlation of age-associated differentially expressed genes (aDEGs; FDR < 0.05) between the lifespan DLPFC and the lifespan replication dataset. Dot size corresponds to the number of shared aDEGs, and color represents the spearman correlation. **b,** Bubble plot showing the π1 statistics (estimate of replication rate) of aDEGs from the DLPFC lifespan dataset in the replication dataset. Dot size reflects the number of aDEGs tested, and color intensity indicates the replication signal (π1 ranges from 0 to 1). In a panel: "*" indicates spearman correlation p-value < 0.05 and "#" indicates significance after BH correction across all subclasses.



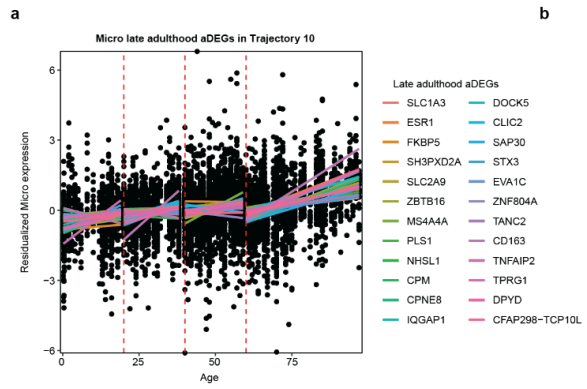
Supplementary Figure 11. Optimal model selection and clustering of gene expression trajectories. a, Mean BIC of linear and non-linear models fitting on all genes. X-axis displays all models implemented in the pipeline. The dashed line shows the optimal model used for final clustering of trajectories. **b,** k means clustering of non-linear coefficients from the optimal model of 334,689 genes across all subclasses. **c-d,** Bar plot to show all genes (n = 334,689) and lifespan aDEGs (n =135,120) after FDR < 0.05 from all subclasses within each trajectory cluster.



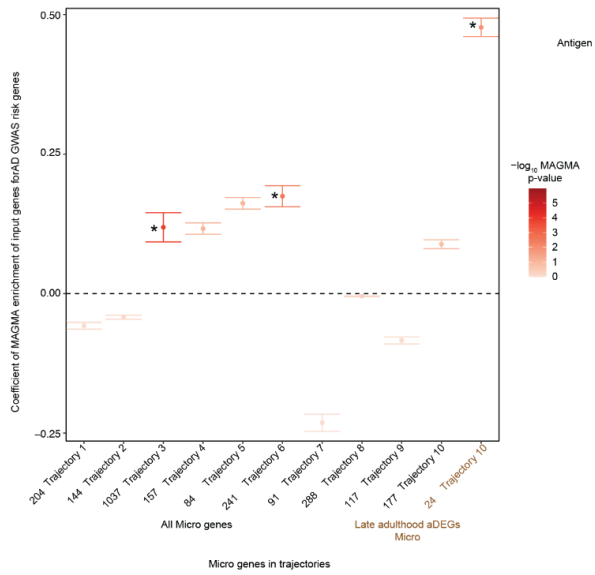
Supplementary Figure 12. Subclass specificity of trajectories. a, Distribution of lifespan-aDEGs across 27 subclasses stratified by ten clusters. **b,** Plot of covariates corrected gene expression as a function age shown as gray points and fitted polynomial using the mean of coefficients from each subclass. The polynomial fit is shown in subclass specific colours. **c,** Association of subclass specific development aDEGs stratified by clusters of trajectories with risk genes for brain-related traits using MAGMA. * and # indicate nominal p-value < 0.05 and FDR < 0.05 from MAGMA enrichment. * and # indicate nominal p-value < 0.05 and FDR < 0.05 from MAGMA enrichment.

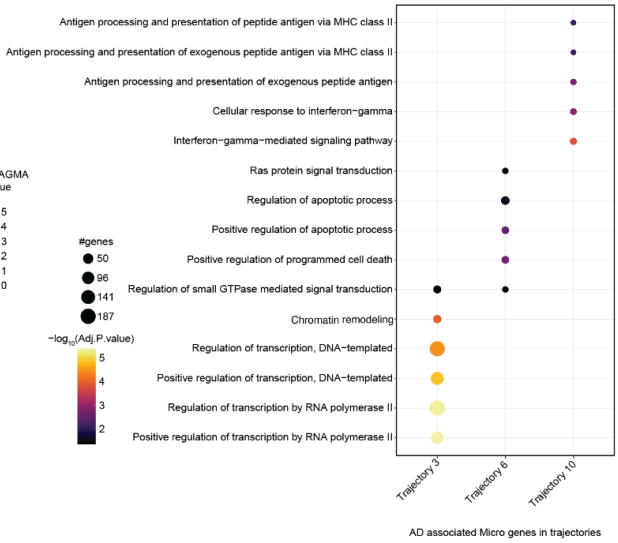


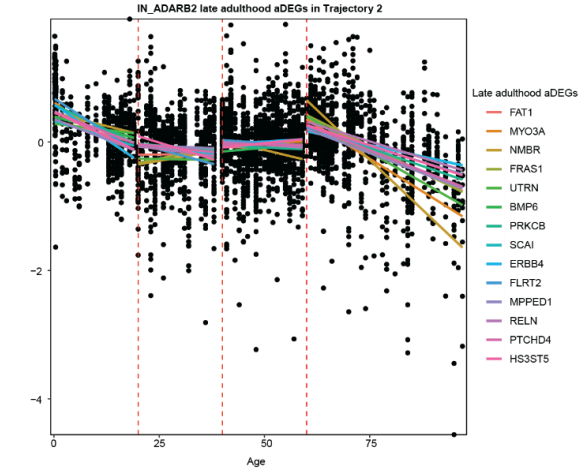
Supplementary Figure 13. Biological mechanism exhibited by trajectories. a, Functional pathway analysis of subclass-specific genes in trajectory 1 (left) and trajectory 5 (right), highlighting subclasses significantly enriched for enrichR GO biological processes with an adjusted p-value < 0.05. **b,** Plot of predicted trajectory using the mean of coefficients of all genes within each trajectory. The linear fit across age groups shows the regression line from average trajectory to age from each age group**.** The text above the age-groups specific fitted line figure shows the beta from model: lm(expression~Age+covariates). All beta values had p-value < 2e-16.


a
Micro late adulthood aDEGs in Trajectory 10
Residualized Micro expression
Age
Late adulthood aDEGs
SLC1A3
ESR1
FKBP5
SH3PXD2A
SLC2A9
ZBTB16
MS4A4A
PLS1
NHSL1
CPM
CPNE8
IQGAP1
DOCK5
CLIC2
SAP30
STX3
EVA1C
ZNF804A
TANC2
CD163
TNFAIP2
TPRG1
DPYD
CFAP298−TCP10L


b
acidic amino acid transport
L−glutamate transmembrane transport
dicarboxylic acid transport
L−amino acid transport
amide transport
L−aspartate import across plasma membrane
amino acid import across plasma membrane
L−alpha−amino acid transmembrane transport
steroid hormone mediated signaling pathway
intracellular estrogen receptor signaling pathway
response to estrogen
steroid hormone mediated signaling pathway
multinuclear osteoclast differentiation
auditory receptor cell stereocilium organization
auditory receptor cell morphogenesis
regulation of myeloid cell differentiation
neuron projection organization
neuron projection development
neuron projection extension
positive regulation of chondrocyte differentiation
chondrocyte differentiation
actin filament network formation
actin filament bundle assembly
organelle fusion
purine nucleobase metabolic process
neurotransmitter receptor transport to postsynaptic membrane
positive regulation of protein localization to plasma membrane
protein localization to chromatin
protein localization to chromosome
regulation of chemotaxis
positive regulation of vascular associated smooth muscle cell migration
fructose transmembrane transport
synaptic vesicle fusion to presynaptic active zone membrane
negative regulation of miRNA transcription
regulation of cardiac muscle contraction by regulation of the release of sequestered calcium ion
cellular response to calcium ion
cellular response to metal ion
response to salt
response to estradiol
exocytosis
import into cell
synaptic vesicle exocytosis
regulation of phospholipase C activity
negative regulation of stem cell population maintenance
myoblast differentiation
myoblast fusion
chloride transport
negative regulation of calcium channel activity
glomerular epithelial cell development
mesonephros development
dense core granule cytoskeletal transport
superoxide anion generation
regulation of ossification
transepithelial transport
ameboidal−type cell migration

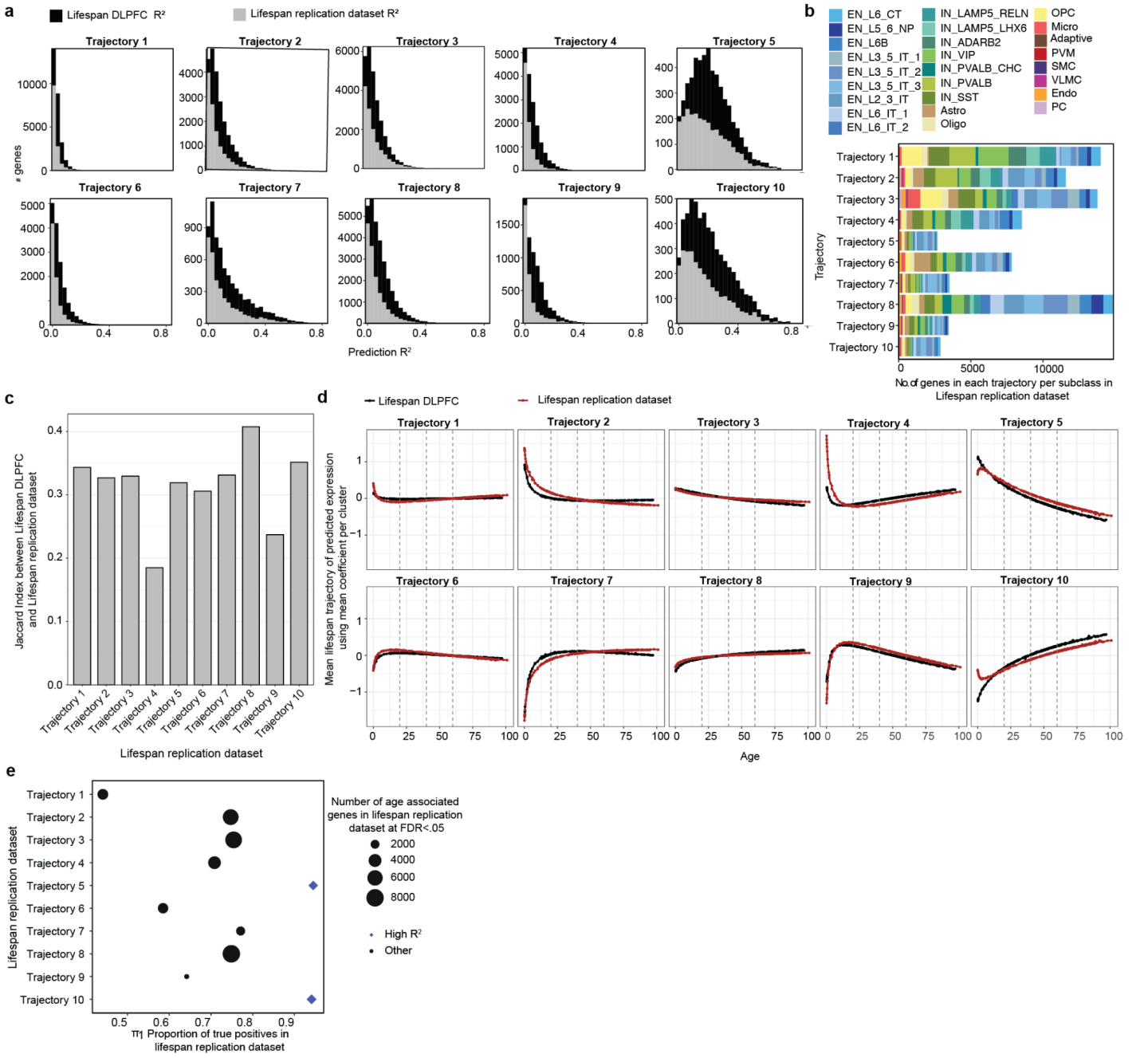

c
Coefficient of MAGMA enrichment of input genes forAD GWAS risk genes
−log10 MAGMA p-value
All Micro genes
Late adulthood aDEGs Micro
Micro genes in trajectories
204 Trajectory 1
144 Trajectory 2
1037 Trajectory 3
157 Trajectory 4
84 Trajectory 5
241 Trajectory 6
91 Trajectory 7
288 Trajectory 8
117 Trajectory 9
177 Trajectory 10
24 Trajectory 10


d
Antigen processing and presentation of peptide antigen via MHC class II
Antigen processing and presentation of exogenous peptide antigen via MHC class II
Antigen processing and presentation of exogenous peptide antigen
Cellular response to interferon−gamma
Interferon−gamma−mediated signaling pathway
Ras protein signal transduction
Regulation of apoptotic process
Positive regulation of apoptotic process
Positive regulation of programmed cell death
Regulation of small GTPase mediated signal transduction
Chromatin remodeling
Regulation of transcription, DNA−templated
Positive regulation of transcription, DNA−templated
Regulation of transcription by RNA polymerase II
Positive regulation of transcription by RNA polymerase II
#genes
−log10(Adj.P.value)
Trajectory 3
Trajectory 6
Trajectory 10
AD associated Micro genes in trajectories

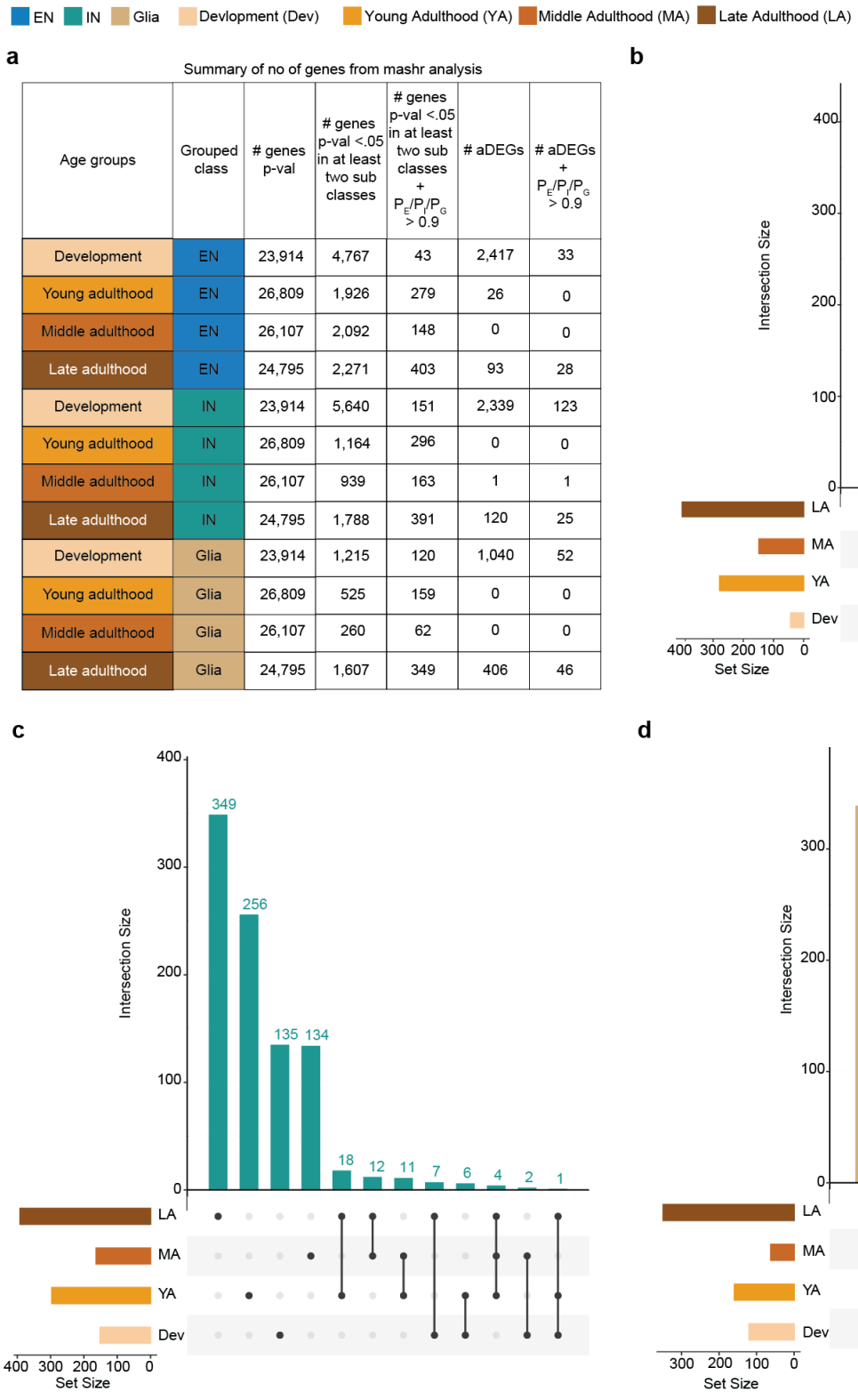

e
IN_ADARB2 late adulthood aDEGs in Trajectory 2
Age
Late adulthood aDEGs
FAT1
MYO3A
NMBR
FRAS1
UTRN
BMP6
PRKCB
SCAI
ERBB4
FLRT2
MPPED1
RELN
PTCHD4
HS3ST5


f
cerebral cortex cell migration
positive regulation of dendritic spine morphogenesis
positive regulation of cardiac muscle cell proliferation
spinal cord development
protein phosphorylation
positive regulation of kinase activity
positive regulation of protein phosphorylation
negative regulation of glucose transmembrane transport
negative regulation of transmembrane transport
positive regulation of cell−cell adhesion
regulation of hemopoiesis
forebrain development
cerebral cortex cell migration
regulation of gliogenesis
regulation of cardiac muscle cell proliferation
positive regulation of AMPA receptor activity
regulation of AMPA receptor activity
regulation of transmembrane transport
positive regulation of cell cell adhesion mediated by integrin
regulation of hemopoiesis
positive regulation of cell−cell adhesion mediated by integrin
positive regulation of antigen receptor−mediated signaling pathway
regulation of B cell receptor signaling pathway
heart development
positive regulation of cardiac muscle tissue growth
positive regulation of dendritic spine development
positive regulation of long−term synaptic potentiation
positive regulation of SMAD protein signal transduction
positive regulation of adherens junction organization
positive regulation of vascular permeability
heparan sulfate proteoglycan biosynthetic process, enzymatic modification
positive regulation of signal transduction
positive regulation of excitatory postsynaptic potential
regulation of SMAD protein signal transduction
modulation by symbiont of host cell−cell adhesion
positive regulation of fatty acid biosynthetic process
lipopolysaccharide−mediated signaling pathway
cellular response to iron ion
response to iron ion
mitotic nuclear membrane disassembly
regulation of coagulation
glandular epithelial cell development
mammary gland epithelial cell differentiation
positive regulation of SMAD signal transduction
peptidyl−tyrosine modification
peptidyl−tyrosine phosphorylation
regulation of peptidyl−tyrosine phosphorylation
lipoprotein localization
protein sulfation
positive regulation of CREB transcription factor activity
positive chemotaxis
type B pancreatic cell differentiation
positive regulation of alcohol biosynthetic process
negative regulation of cell−cell adhesion mediated by cadherin
mitochondrial fragmentation involved in apoptotic process
Notch−mediated signaling pathway
basement membrane organization

Supplementary Figure 14. Late adulthood IN_ADARB2 and Micro aDEGs. a, Plot of Micro specific aDEGs residualized gene expression from trajectory 10 shown as black markers. Age related changes of every gene is shown as a colored regression line within each group. **b,** Functional pathway analysis of micro-specific aDEGs from trajectories 10 was conducted using rrvgo. The input GO IDs were obtained from enrichR analysis of aDEGs with an adjusted p-value < 0.10 for microglia. **c,** The coefficient of enrichment of AD GWAS risk genes for all micro genes from the 10 trajectories and late-adulthood specific aDEGs in trajectory 10 is shown in brown. **d,** Functional pathway analysis of all micro genes from trajectories 3, 6, and 10, which were significantly enriched for GO biological processes with an adjusted p-value < 0.05, is also included. **e,** Plot of IN_ADARB2 specific aDEGs residualized gene expression from trajectory 2 shown as black markers. Age related changes of every gene is shown as a colored regression line within each group. **f,** Functional pathway analysis of IN_ADARB2 aDEGs from trajectories 2 was conducted using rrvgo. The input GO IDs were obtained from enrichR analysis of aDEGs with an adjusted p-value < 0.05.

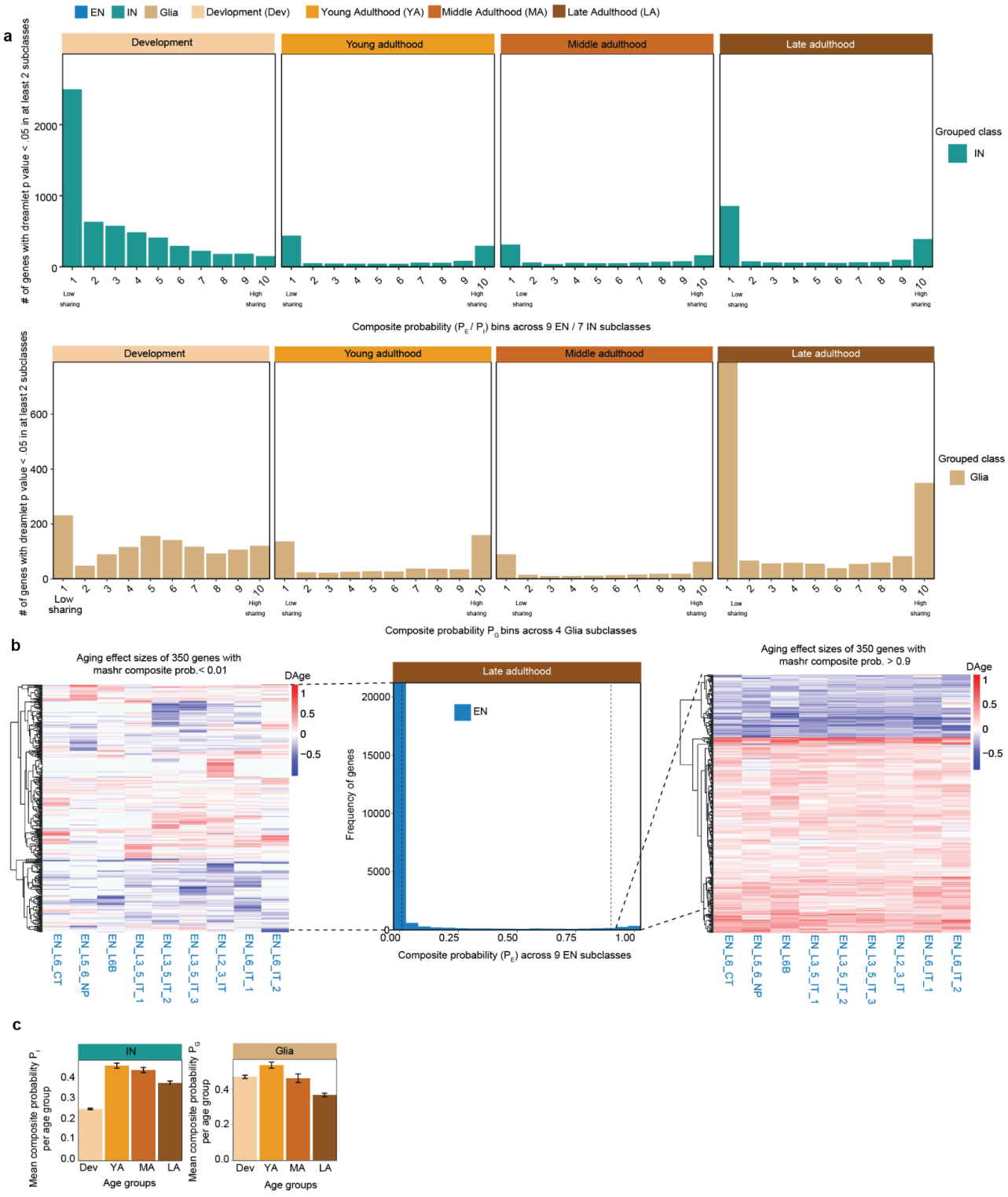


Supplementary Figure 15. Replication of lifespan trajectories. a, Distribution of gene-level R^2 values comparing predicted versus observed expression across all genes in each trajectory. R^2 values are shown for both the discovery lifespan DLPFC dataset (black) and the lifespan replication dataset (gray). **b,** Barplot showing the number of genes per trajectory stratified by subclass in the Lifespan replication dataset. **c,** Jaccard index between discovery and replication trajectory gene sets, summarizing the degree of gene-level overlap for each trajectory. **d,** Mean trajectory fits for each trajectory across donor age. Predicted expression trends from the discovery dataset (black) and the replication dataset (red). **e,** Scatterplot showing replication performance across trajectories. The x-axis shows the proportion of true positives (π_1) from the discovery dataset that were recovered in the replication dataset. Dot size corresponds to the number of replicated age-associated genes at FDR < 0.05. Blue diamonds indicate trajectories with high R^2 in both datasets (Trajectories 5 and 10). In panels **c–e** Trajectory patterns derived independently by re-running the full trajectory pipeline (Fig. 2a) on the Lifespan replication dataset whereas in panel.

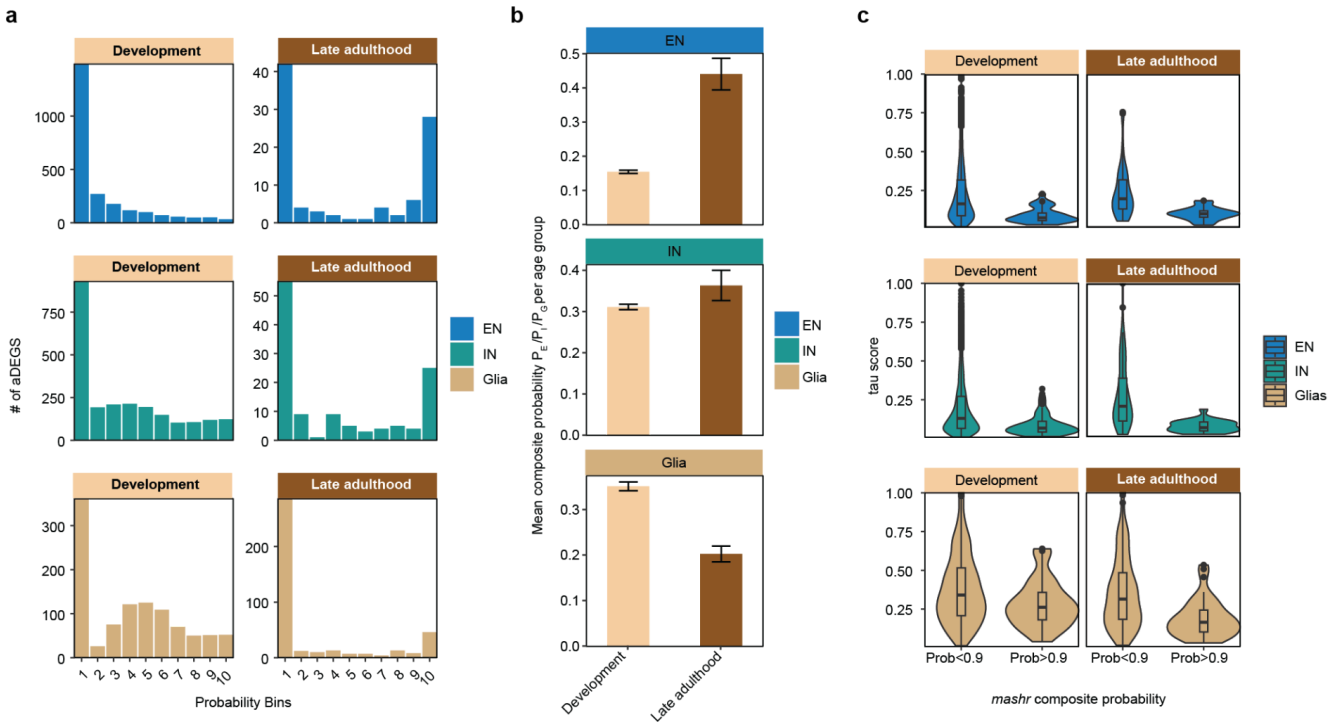


Age groups	Grouped class	# genes p-val	# genes p-val <.05 in at least two sub classes	# genes p-val <.05 in at least two sub classes + $P_E/P_I/P_G$ > 0.9	# aDEGs	# aDEGs + $P_E/P_I/P_G$ > 0.9
Development	EN	23,914	4,767	43	2,417	33
Young adulthood	EN	26,809	1,926	279	26	0
Middle adulthood	EN	26,107	2,092	148	0	0
Late adulthood	EN	24,795	2,271	403	93	28
Development	IN	23,914	5,640	151	2,339	123
Young adulthood	IN	26,809	1,164	296	0	0
Middle adulthood	IN	26,107	939	163	1	1
Late adulthood	IN	24,795	1,788	391	120	25
Development	Glia	23,914	1,215	120	1,040	52
Young adulthood	Glia	26,809	525	159	0	0
Middle adulthood	Glia	26,107	260	62	0	0
Late adulthood	Glia	24,795	1,607	349	406	46

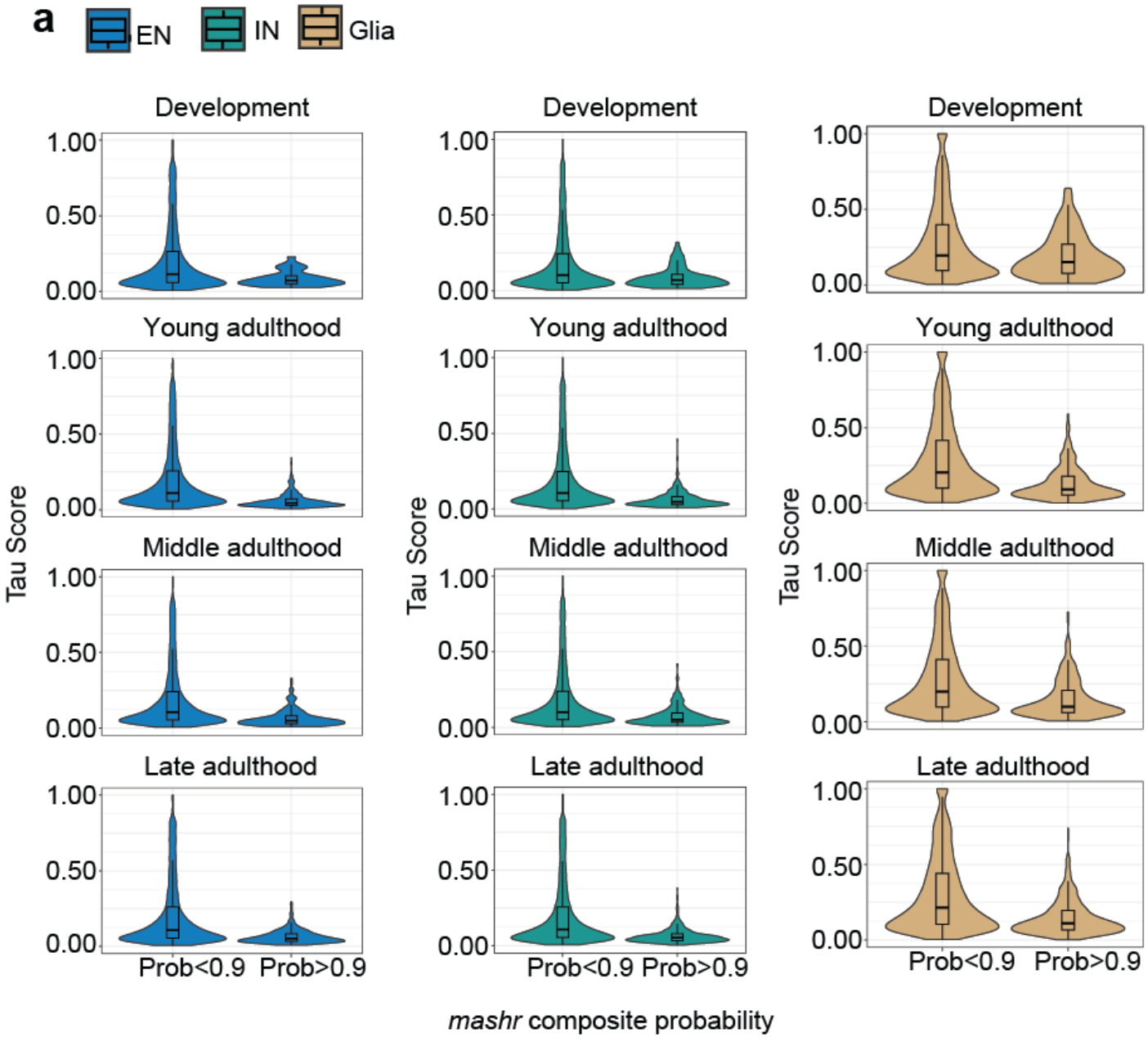
Supplementary Figure 16. Summary of number of shared genes from mashr analysis. a, Table displaying the total number of genes to run the mashr analysis and shared genes at composite probability > 0.9. Table also displays the counts of genes that had age effect size at significant p-value < 0.5 in at least two subclasses**. b-d,** Upset plot to show the overlap of shared genes at composite probability > 0.9 across age groups in EN, IN and Glia grouped classes.



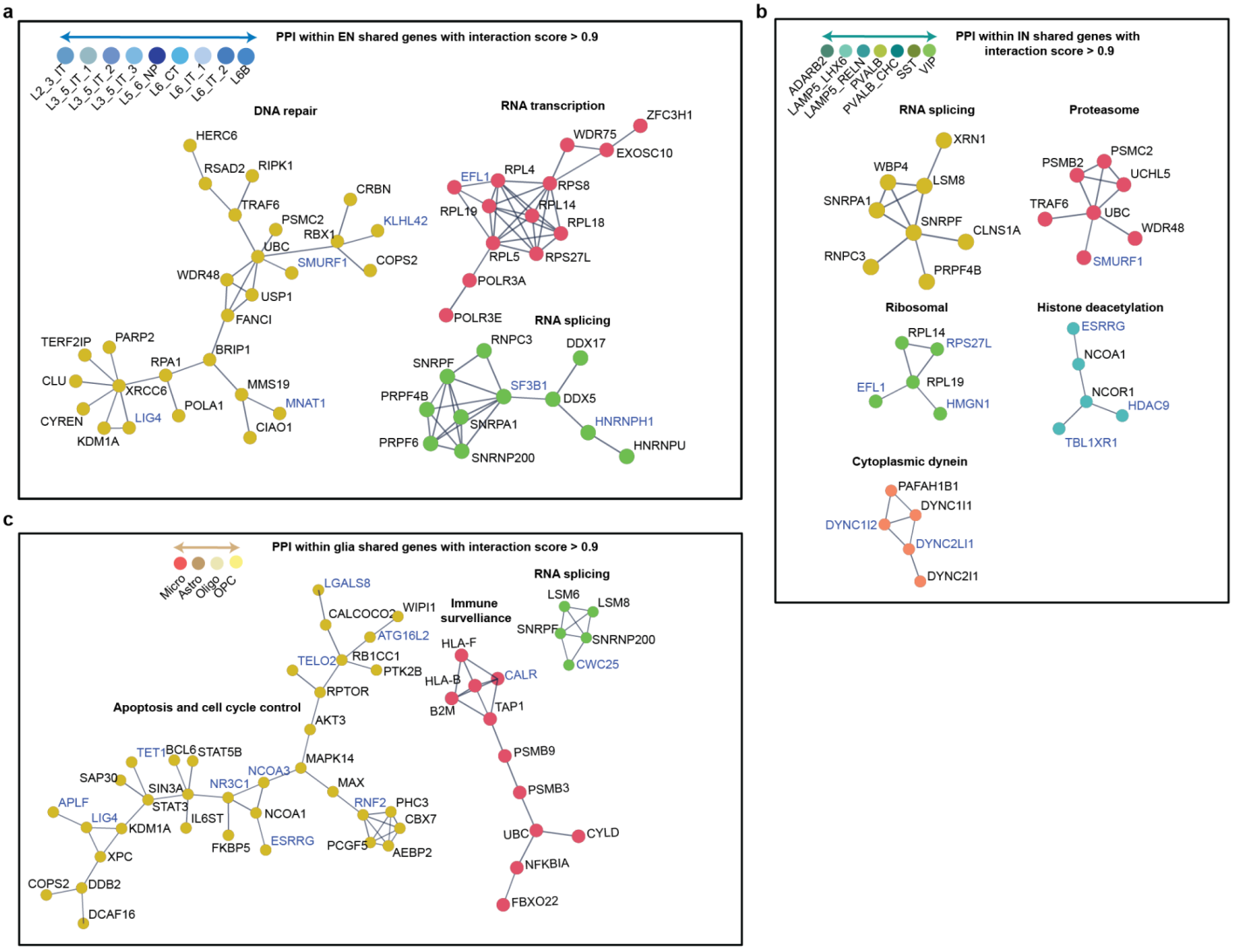
Supplementary Figure 17. Degree of sharing across subclasses. a, Bar plot displaying the number of genes stratified into 10 equally sized bins based on composite probability (P_I and P_G) values for IN and glia respectively, ranging from 0 to 1 from development to late adulthood group. The number of input genes to make these bar-plots is displayed in Fig. S16a. **b,** Heatmap of age associated effect sizes across 9 EN for genes with composite probability ($P_E < 0.01$ in left) and ($P_E > 0.9$ in right) across late adulthood. **c,** Mean P_I (left) and P_G (right) of genes in each age group. This plot is the mean of probabilities of genes shown in a panel.



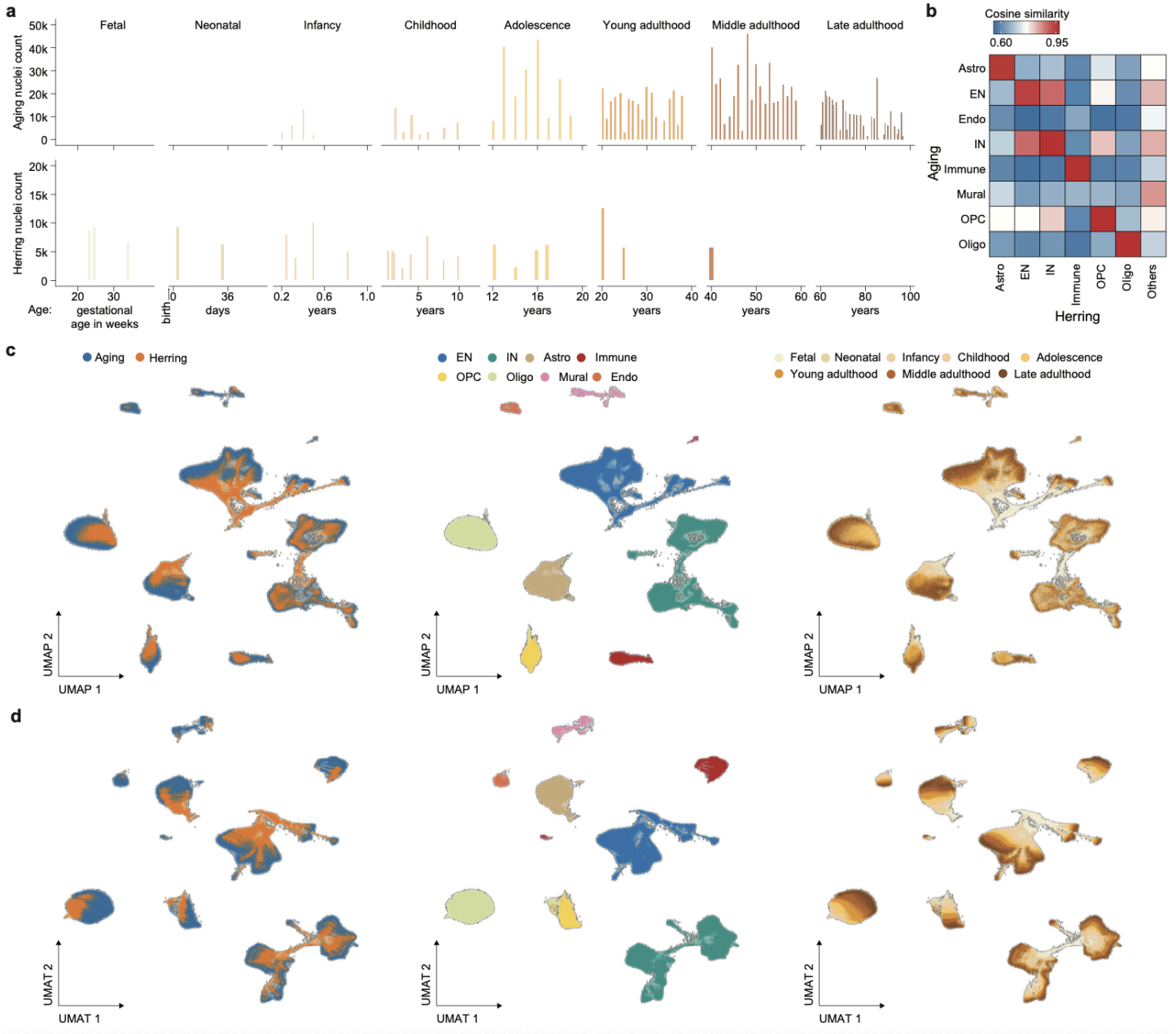
Supplementary Figure 18. Composite probability of aDEGs. a, Bar plot displaying the number of aDEGs stratified into 10 equally sized bins based on composite probability (P_E, P_I , P_G) values for EN, IN and glia ranging from 0 to 1 for development and late adulthood groups. **b,** Mean P_E, P_I , P_G of all development and late adulthood aDEGs. **c,** Violin plot of tau scores of genes that have age associated with nominal p-value from dreamlet tool in at least two subclasses. The x-axis is stratified by genes with P_E, P_I , $P_G < 0.9$ and > 0.9.



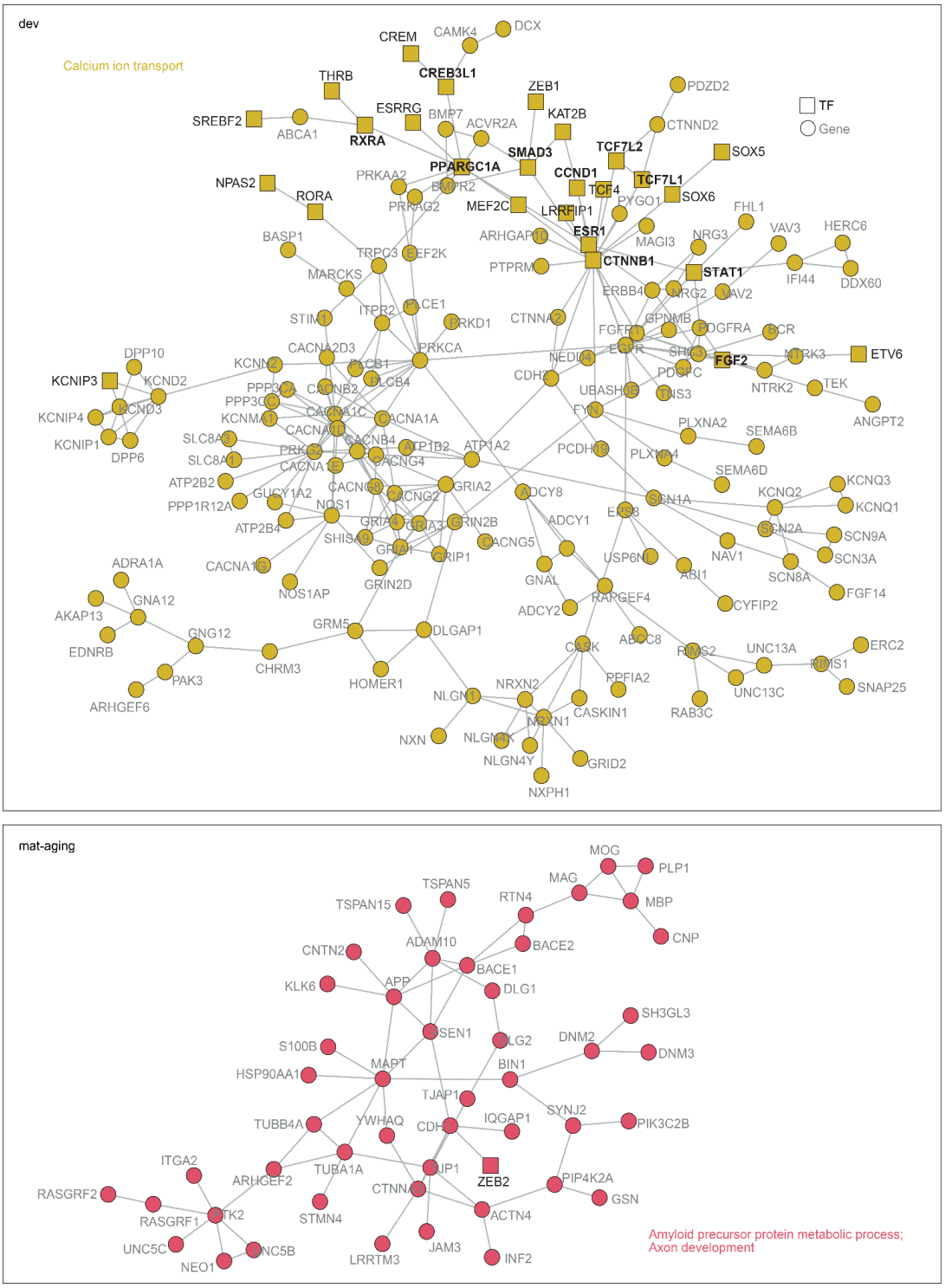
Supplementary Figure 19. Distribution of cell specificity tau score. a, Violin plot of tau scores of genes that have age associated with nominal p-value from dreamlet tool in at least two subclasses. The x-axis is stratified by genes with P_E, P_I, $P_G < 0.9$ and > 0.9. These tau scores were obtained from 9 EN and 7 IN and 4 Glia expressions separately for each age group.



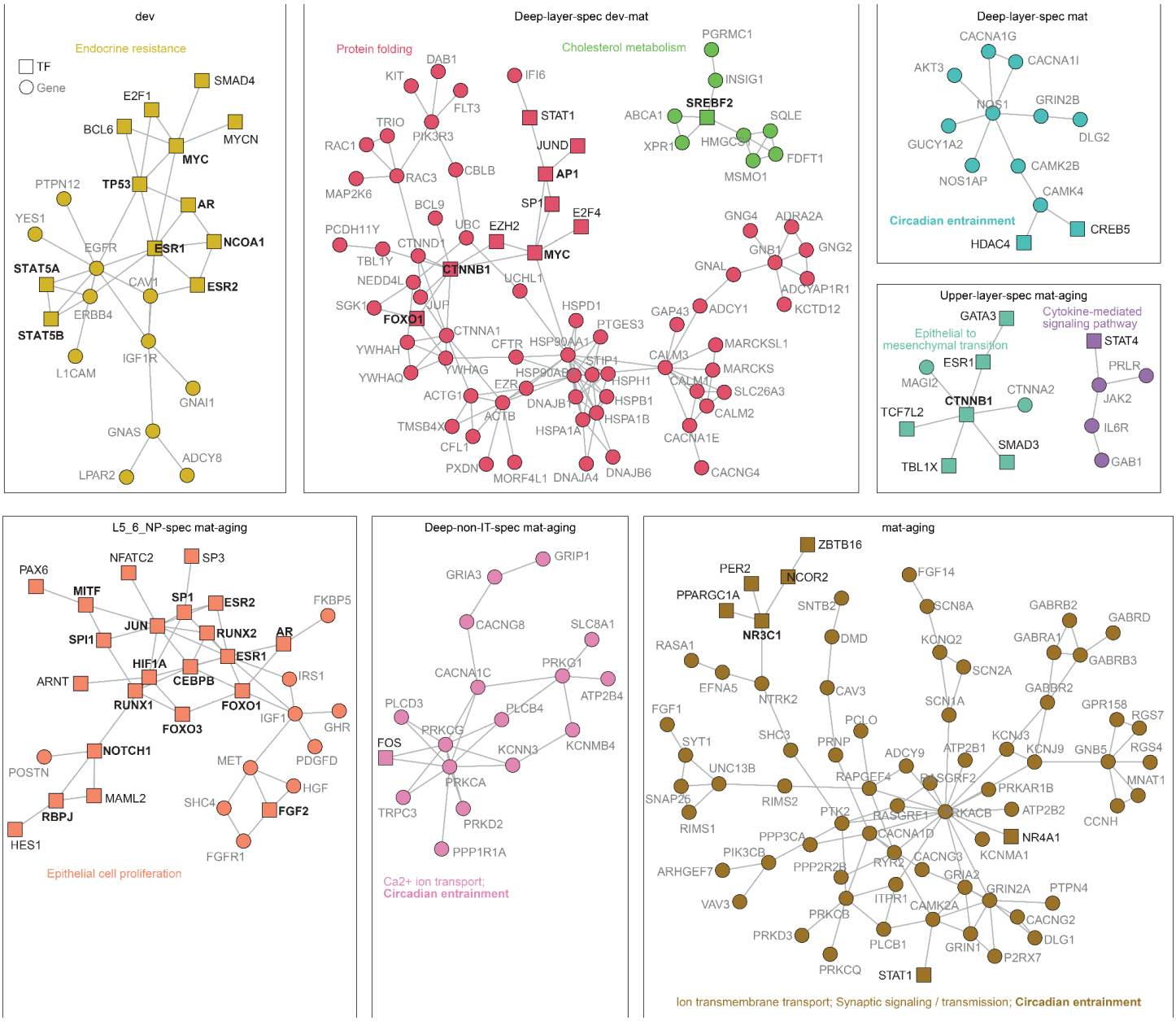
Supplementary Figure 20. Protein-protein interactions among shared genes. a-c Identified clusters of networks of genes that show significant PPI interactions with score > 0.9 obtained from the analysis of shared genes within 9 EN, 7 IN and 4 glia classes. Genes highlighted in blue indicate downregulation with age and remaining are upregulated. EN clusters in b include genes crucial for 1) DNA repair, 2) RNA transcription, and 3) RNA splicing-specific complexes colored as gold, pink and mantis respectively. IN clusters in c include 1) spliceosome assembly, 2) the proteasome complex, 3) the histone deacetylase complex, and 4) components for maintaining cellular function and intracellular transport colored as yellow, pink, mantis, sea blue and salmon respectively. Glia clusters in d include 1) proteins involved in apoptosis and cell cycle control, 2) the MHC class I protein complex and peptide loading complex, and 3) RNA processing and splicing complexes colored as yellow, pink and mantis respectively.



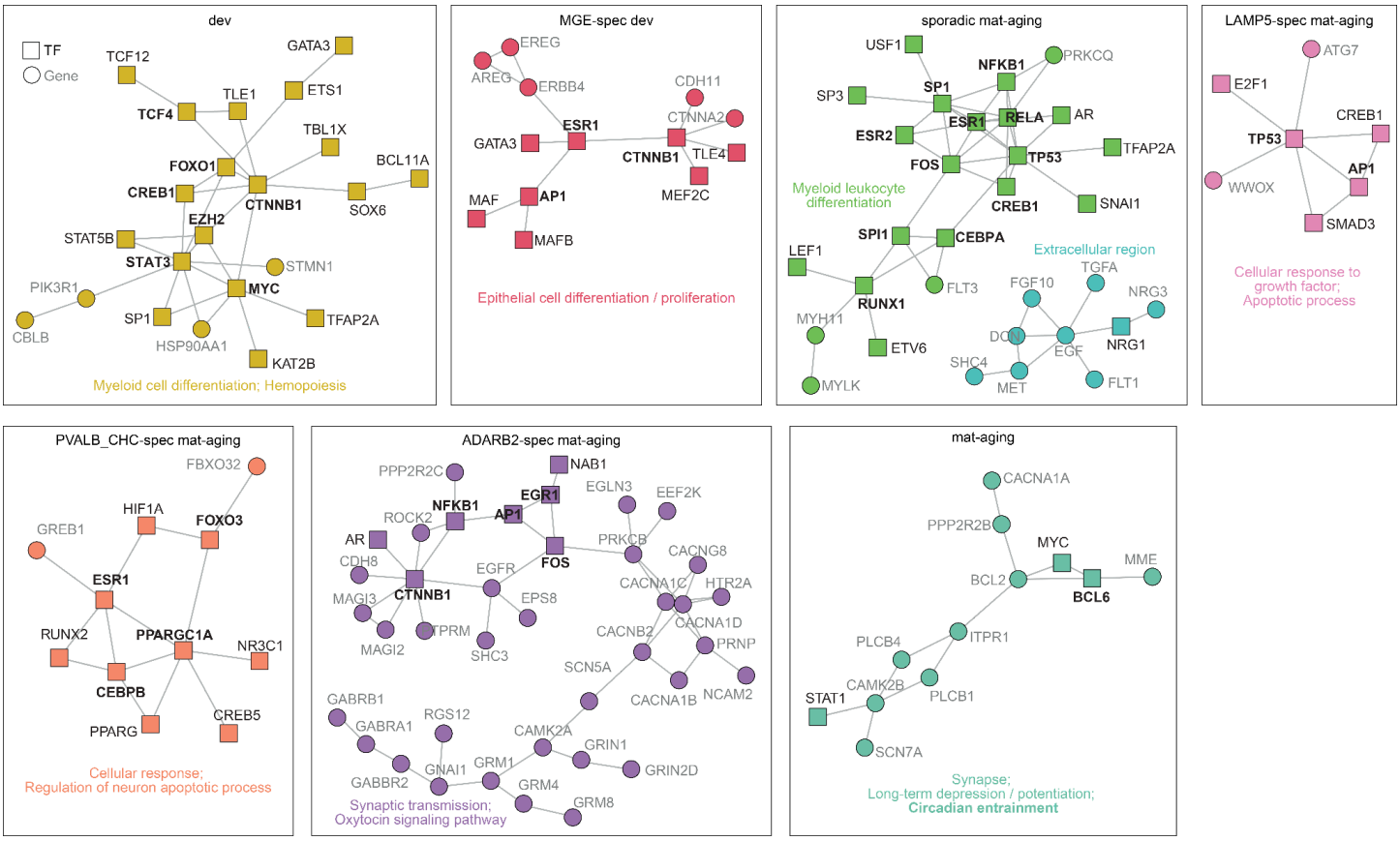
Supplementary Figure 21. Characteristics and similarity of the combined snRNA-seq dataset. a, Nuclei counts in current (top) and previous (bottom) snRNA-seq datasets of the human DLPFC, categorized by age group and color-coded. **b**, Cell type similarity between the two datasets at the class level. **c**,**d**, Integration of the current dataset with previously published data. UMAP (**c**) and UMAT (**d**) representations of the combined snRNA-seq dataset colored by dataset (left), class (middle), and stage (right).



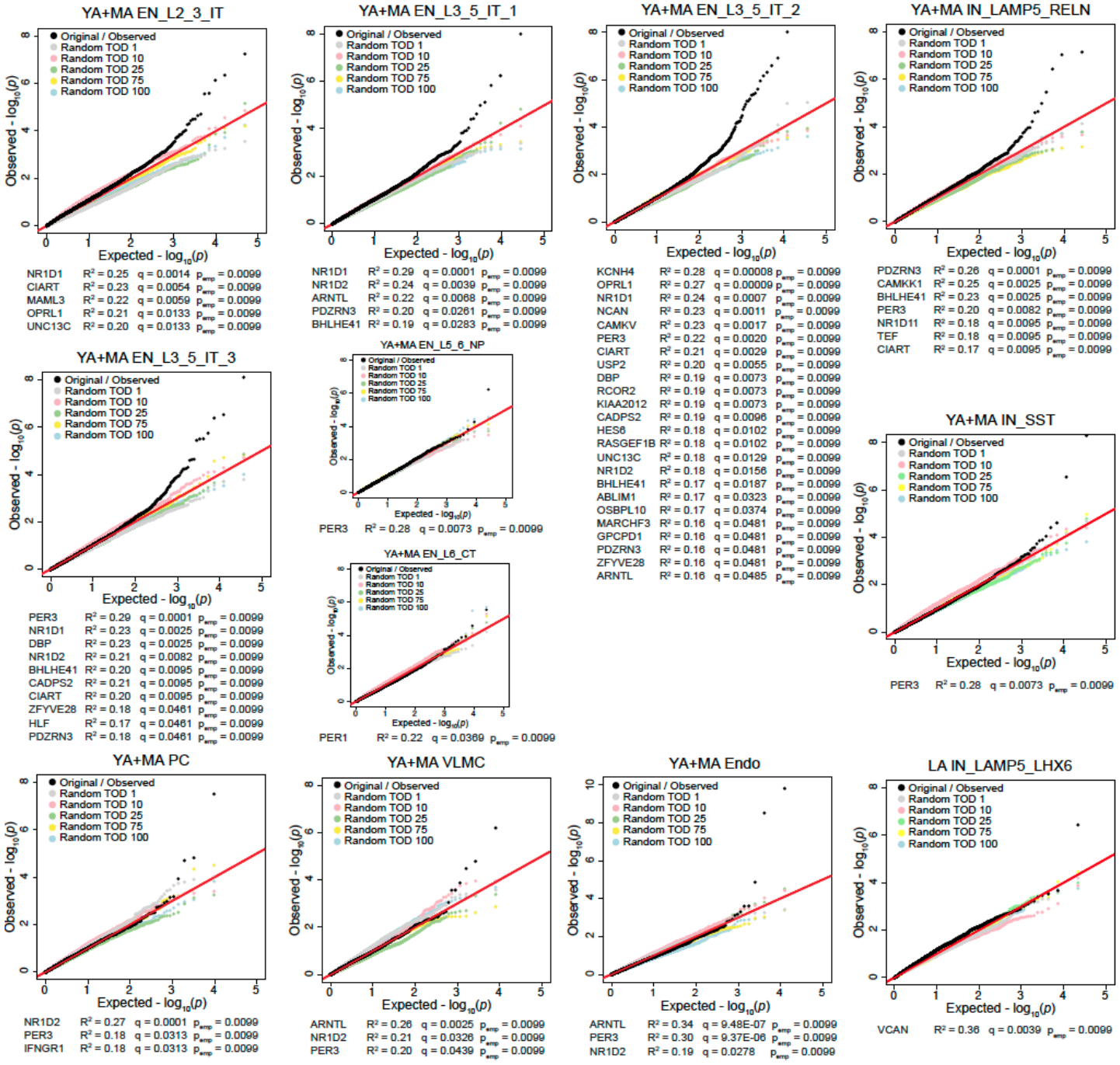
Supplementary Figure 22. Regulatory networks for traDEG modules in oligodendrocytes. TF-gene regulatory networks inferred for Oligo traDEG modules, with enriched biological processes annotated based on network composition. Squares indicate TFs; circles indicate target genes. TF labels are shown in black; gene labels are shown in gray. Bolded TF names denote key regulators with high connectivity (connected components ≥ 5 nodes and degree > 3).



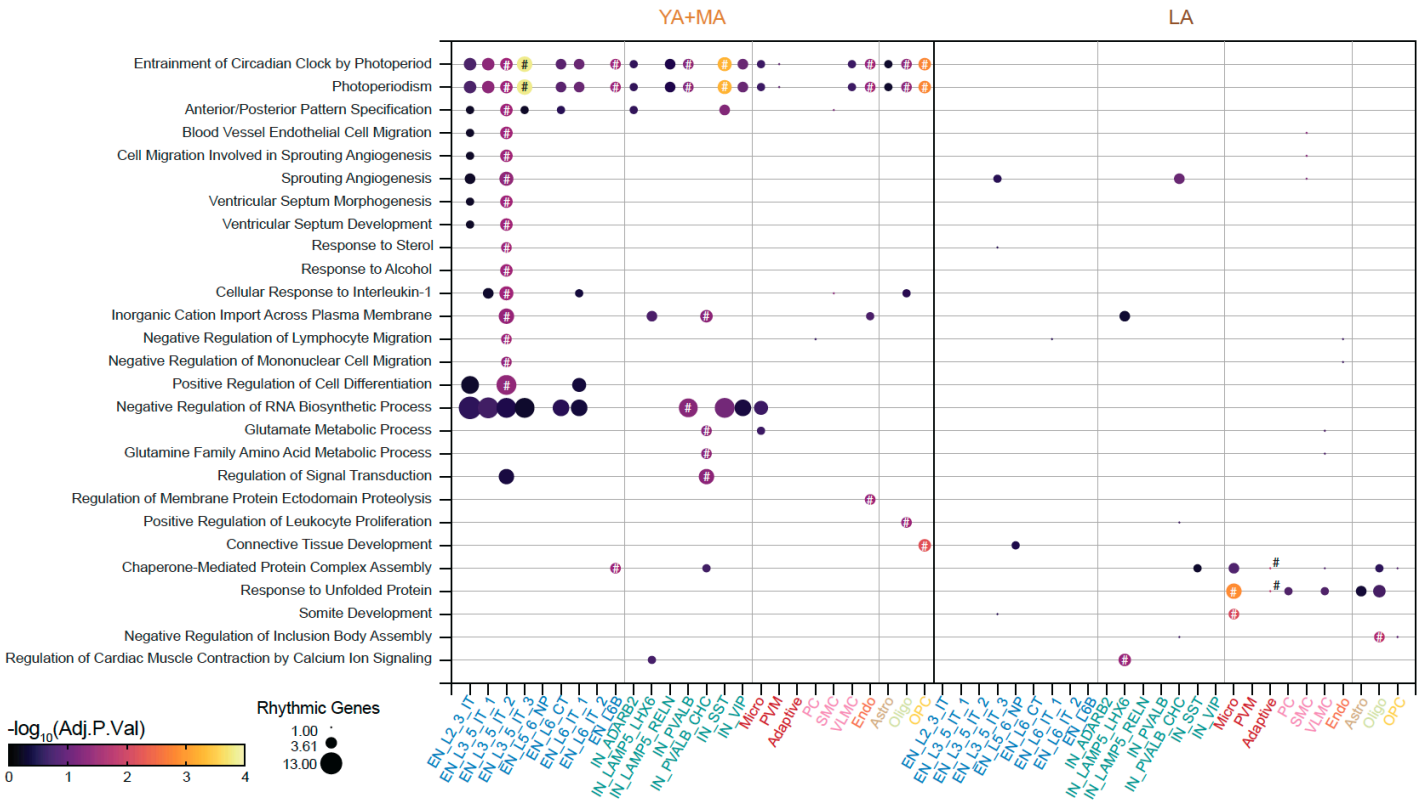
Supplementary Figure 23. Regulatory networks for traDEG modules in excitatory neurons. TF-gene regulatory networks inferred for EN traDEG modules, with enriched biological processes annotated based on network composition. Squares indicate TFs; circles indicate target genes. TF labels are shown in black; gene labels are shown in gray. Bolded TF names denote key regulators with high connectivity (connected components ≥ 5 nodes and degree > 3).



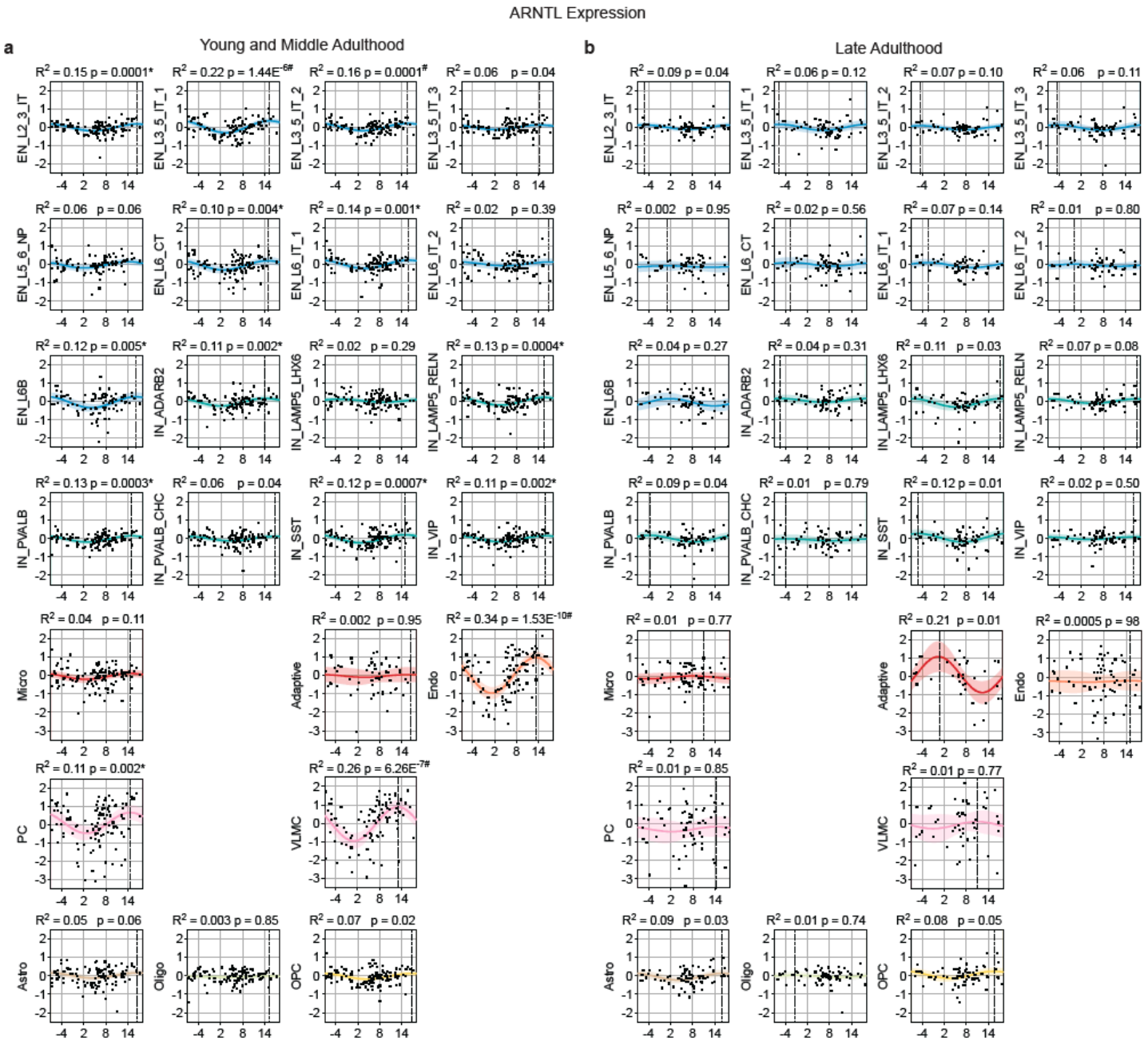
Supplementary Figure 24. Regulatory networks for traDEG modules in inhibitory neurons. TF-gene regulatory networks inferred for IN traDEG modules, with enriched biological processes annotated based on network composition. Squares indicate TFs; circles indicate target genes. TF labels are shown in black; gene labels are shown in gray. Bolded TF names denote key regulators with high connectivity (connected components ≥ 5 nodes and degree > 3).



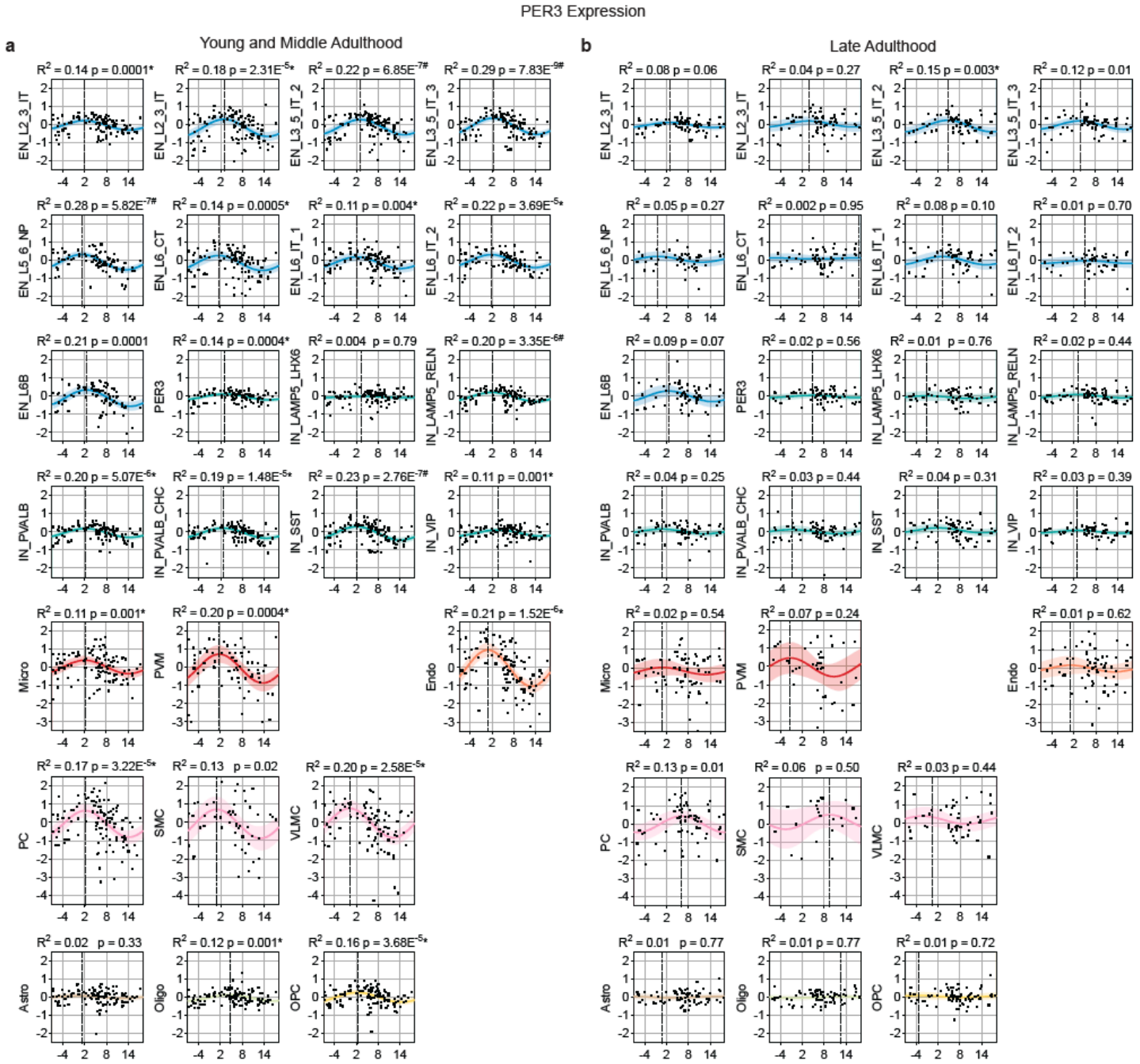
Supplementary Figure 25. Q-Q Plots of representative rhythmicity analyses with randomized TOD. Q-Q Plots comparing the expected and observed p-value distributions of rhythmicity analyses. For each subclass, the original/observed distribution (black) is compared to 5 representative samples in which TOD was randomized. The subclasses shown had at least one gene that was significantly rhythmic (F_{Rhy}, $FDR_{BH} < 0.05$, 1-tailed) in the original/observed analysis. Below each Q-Q plot, the genes that were originally identified as rhythmic are listed, along with their goodness-of-fit (R2), FDR_{BH} (q), and the final calculated empirical p-value.



Supplementary Figure 26. Pathways enriched in rhythmic genes. enrichR pathway analysis of rhythmic genes (F_{Rhy}, nominal p-value < 0.01, 1-tailed) in YA+MA and LA subclasses. Pathways shown were identified as significantly enriched (FDR_{BH} < 0.05, represented by #) in at least one subclass. Other points represent classes in which the pathway was only nominally significant (p < 0.01).



Supplementary Figure 27. ARNTL rhythms. Plots of residualized gene expression across time of death (TOD) for the circadian gene ARNTL in **a**, YA+MA and **b**, LA subclasses. A dotted vertical line indicates the peak expression time for each curve.



Supplementary Figure 28. PER3 rhythms. Plots of residualized gene expression across time of death (TOD) for the circadian gene PER3 in **a**, YA+MA and **b**, LA subclasses. A dotted vertical line indicates the peak expression time for each curve.

References

1. Ruzicka, W. B. *et al.* Single-cell multi-cohort dissection of the schizophrenia transcriptome. *Science* **384**, eadg5136 (2024).

2. Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385.e27 (2023).

3. Green, G. S. *et al.* Cellular dynamics across aged human brains uncover a multicellular cascade leading to Alzheimer's disease. *bioRxiv* (2023) doi:10.1101/2023.03.07.531493.

4. Lyketsos, C. G. *et al.* Prevalence of neuropsychiatric symptoms in dementia and mild cognitive impairment: results from the cardiovascular health study. *JAMA* **288**, 1475–1483 (2002).

5. Lyketsos, C. G. *et al.* Neuropsychiatric symptoms in Alzheimer's disease. *Alzheimers. Dement.* **7**, 532–539 (2011).

6. Zhao, Q.-F. *et al.* The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis. *J. Affect. Disord.* **190**, 264–271 (2016).

7. Pless, A. *et al.* Understanding neuropsychiatric symptoms in Alzheimer's disease: challenges and advances in diagnosis and treatment. *Front. Neurosci.* **17**, 1263771 (2023).

8. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).

9. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).

10. Hoffman, G. E. *et al.* Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. *bioRxiv* 2023.03.17.533005 (2023) doi:10.1101/2023.03.17.533005.