A single-cell transcriptomic atlas of the prefrontal cortex across the human lifespan

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

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, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA,

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

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

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

[11]PsychAD Consortium.

*These authors contributed equally to this work

Correspondence to: K.G. (kiran.girdhar@mssm.edu) and P.R. (panagiotis.roussos@mssm.edu)

Abstract

The dorsolateral prefrontal cortex is central to higher cognitive functions and is particularly vulnerable to age-related decline. To advance our understanding of the molecular mechanisms underlying brain development, maturation, and aging, we constructed a detailed single-cell transcriptomic atlas of the human dorsolateral prefrontal cortex, encompassing over 1.3 million nuclei from 284 postmortem samples spanning the full human lifespan (0-97 years). This atlas reveals distinct phases of transcriptomic activity: a dynamic developmental period, stabilization during midlife, and subtle yet coordinated changes in late adulthood. Modeling non-linear age trends across the lifespan shows ten distinct trajectories of the entire transcriptome from all cell types, with notable findings in neurons and microglia, linked to neurodevelopmental disorders and Alzheimer's disease risk, respectively. Moreover, excitatory neurons exhibit a convergence of gene expression patterns across the lifespan, suggesting the emergence of a common molecular signature of aging. Pseudotime analysis tracing the progression of cellular lineages throughout life reveals key gene clusters with dynamic expression changes that reflect development, maturation, and aging, as well as their connection to brain-related diseases. We uncover significant circadian rhythm reprogramming in late adulthood, characterized by disruption of core clock gene rhythmicity and the emergence of new rhythmic patterns, particularly within microglia and oligodendrocytes. This comprehensive single-cell atlas provides a baseline for understanding the molecular transitions from development through successful aging in the human dorsolateral prefrontal cortex.

Main

The dorsolateral prefrontal cortex (DLPFC) plays a pivotal role in higher cognitive functions, including executive processes, decision-making, and working memory[1–3]. These functions are critical for maintaining mental health and cognitive integrity throughout life. However, the DLPFC is also one of the brain regions most vulnerable to age-related decline, contributing to the pathogenesis of various neuropsychiatric and neurodegenerative disorders, such as schizophrenia (SCZ)[4,5], major depressive disorder (MDD)[6], and Alzheimer's disease (AD)[7]. Despite its importance, our understanding of how the DLPFC transcriptome changes at single cell level across the human lifespan—particularly in response to aging—remains incomplete.

Single-nucleus RNA sequencing (snRNA-seq) has revolutionized our ability to study the human brain, enabling the exploration of molecular changes that occur within specific cell types across different life stages[8–11]. While previous studies[9,12,13] have provided valuable insights into DLPFC development and the aging process in adulthood, a comprehensive reference atlas that spans the entire human lifespan is lacking. Such an atlas is crucial for unraveling the molecular mechanisms that underlie neurodevelopment, the maintenance of cognitive function during midlife, and the adaptations that characterize successful aging.

The PsychAD Consortium (**Supplementary Notes "PsychAD dataset"**)[14,15] generated a population-level snRNA-seq dataset comprising over 6.3 million nuclei isolated from the DLPFC of 1,494 individual donors. In this study, we utilized a subset of the PsychAD cohort, encompassing over 1.3 million nuclei from 284 neurotypical samples, with donor ages ranging from 0 to 97 years, to provide a detailed transcriptomic atlas of the human DLPFC across the human lifespan (**Fig. 1a**). Our analysis provides a comprehensive overview of the transcriptomic changes that occur from development through to late adulthood, with a particular focus on the processes that contribute to successful development and aging. Moreover, the data provides a valuable resource for future studies aimed at elucidating the molecular basis of cognitive function and its decline in aging and disease.

Results

A single nuclei RNA-seq atlas of the DLPFC across the human lifespan

To understand dynamics of gene expression in the human DLPFC throughout life, we generated snRNA-seq libraries from the DLPFC of 284 neurotypical controls (87 female and 197 male donors), spanning the following time-points: neonatal (0-1 year, n=9), childhood (2-11 years, n=11), adolescence (12-19 years, n=33), and young (20-39 years, n=54), middle (40-59 years, n=95) and late adulthood (≥ 60 years, n=82) (**Fig. 1a**, **Fig. S1a**, **Extended Data Fig. 1a, b, Supplementary Data 1** and **Supplementary Notes "PsychAD dataset"**)[14,15]. Following quality filtering[14,15], we obtained 1,307,674 single nucleus transcriptome profiles, representing 8 major cell classes (excitatory (EN) and inhibitory (IN) neurons, astrocytes (Astro), immune cells, mural cells, endothelial cells (Endo), oligodendrocytes (Oligo) and oligodendrocyte progenitor cells (OPC), which were further categorized into 27 distinct subclasses and 65 subtypes[14,15] (**Fig. 1b**, **Extended Data Fig. 1c** and

Methods). Visual inspection of nuclei counts across subclasses revealed notable changes during early age groups, such as neurogenesis and OPC enrichment (**Fig. 1c**), which are further quantified in subsequent sections.

To explore the link between cellular subclasses and traits relevant to brain health, we investigated enrichment with polygenic disease risk (**Fig. 1d** and **Methods**) across three primary categories of disorders: neurological, psychiatric and others, which mainly consist of metabolic and immunological traits (**Supplementary Table 2** and **Supplementary Data 2**). Psychiatric traits were enriched in both neuronal classes, while MDD, autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), and educational attainment also demonstrated significant scores for OPCs. For neurological and other traits, we observed enrichment across immune cells, mural cells and astrocytes, with a particularly prominent effect in immune cells (including the adaptive, microglia (Micro), and perivascular macrophage (PVM) subclasses)[16,17]. Notably, the obesity trait displayed significant enrichment in both EN and IN classes, which is consistent with partial co-heritability between obesity and psychiatric traits[18,19–21].

We then proceeded to quantify the age-related variance in gene expression for each subclass using the *variancePartition* tool (**Methods**)[22]. Overall, the largest effect on gene expression was observed in the developmental group (neonatal, childhood and adolescence) (**Fig. 1e**) with a mean contribution of 4% in total variance (**Extended Data Fig. 1c)**, likely reflecting transcriptomic changes due to postnatal neurogenesis and gliogenesis[12,23]. The second largest effect was in the late adulthood group, explaining 2% of total variance (**Extended Data Fig. 1c** and **Supplementary Data 3**), indicating that successful aging coincides with alterations in gene expression[24–26]. Examining age-specific contributions to variance within broad taxonomy classifications [EN, IN, glia (consisting of Astro, Micro, Oligo, and OPC) and Other (consisting of Adaptive, Endo and Mural cells)] (**Fig. 1e** inset plot) revealed that the developmental group accounts for the largest contribution to variance in EN (mean 5.1%), followed by IN (mean 4.2%) and glia (mean 2.8%). In contrast, the largest contribution to variance in “Other” cells can be attributed to the late adulthood group (mean 3.1%).

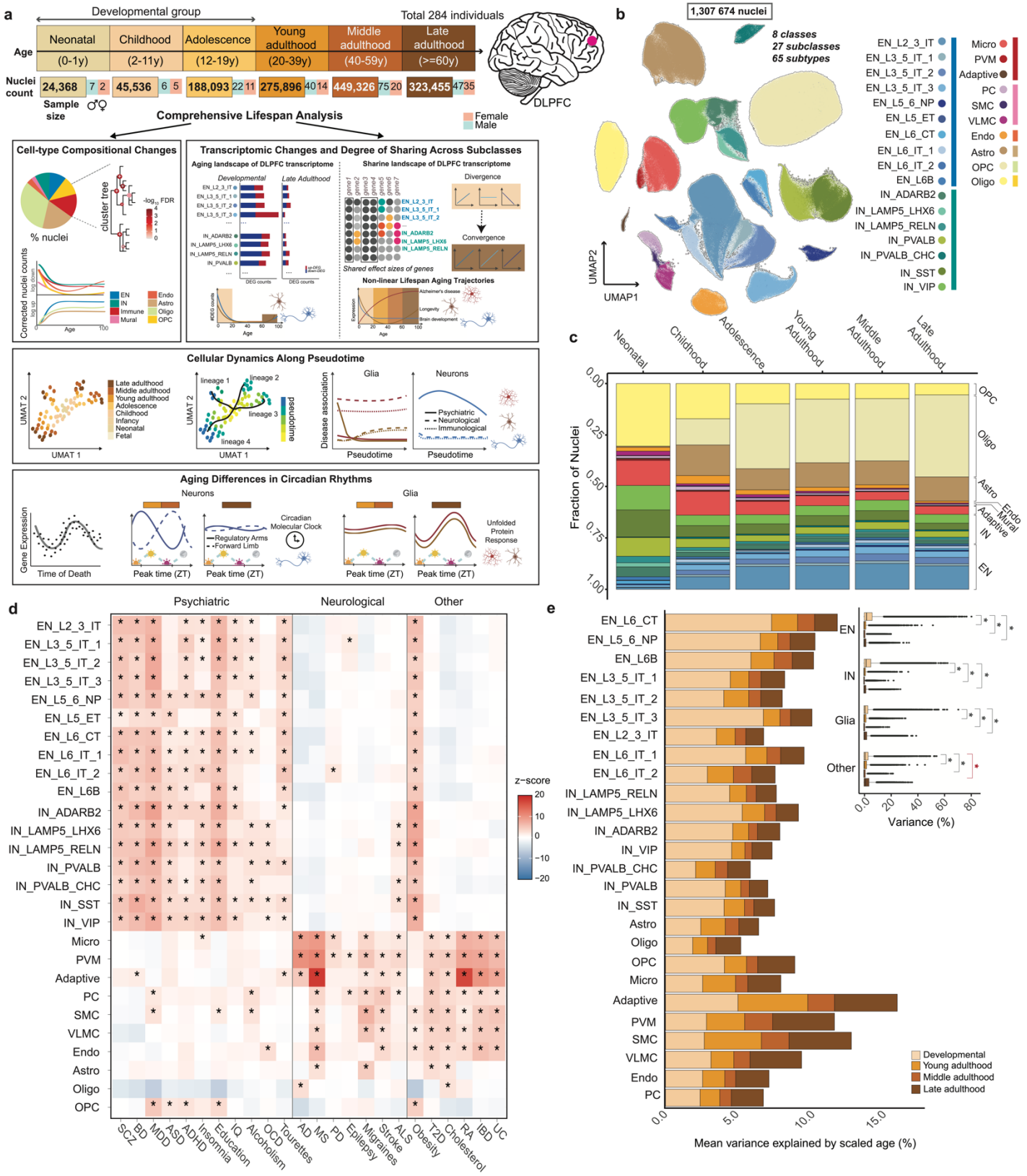
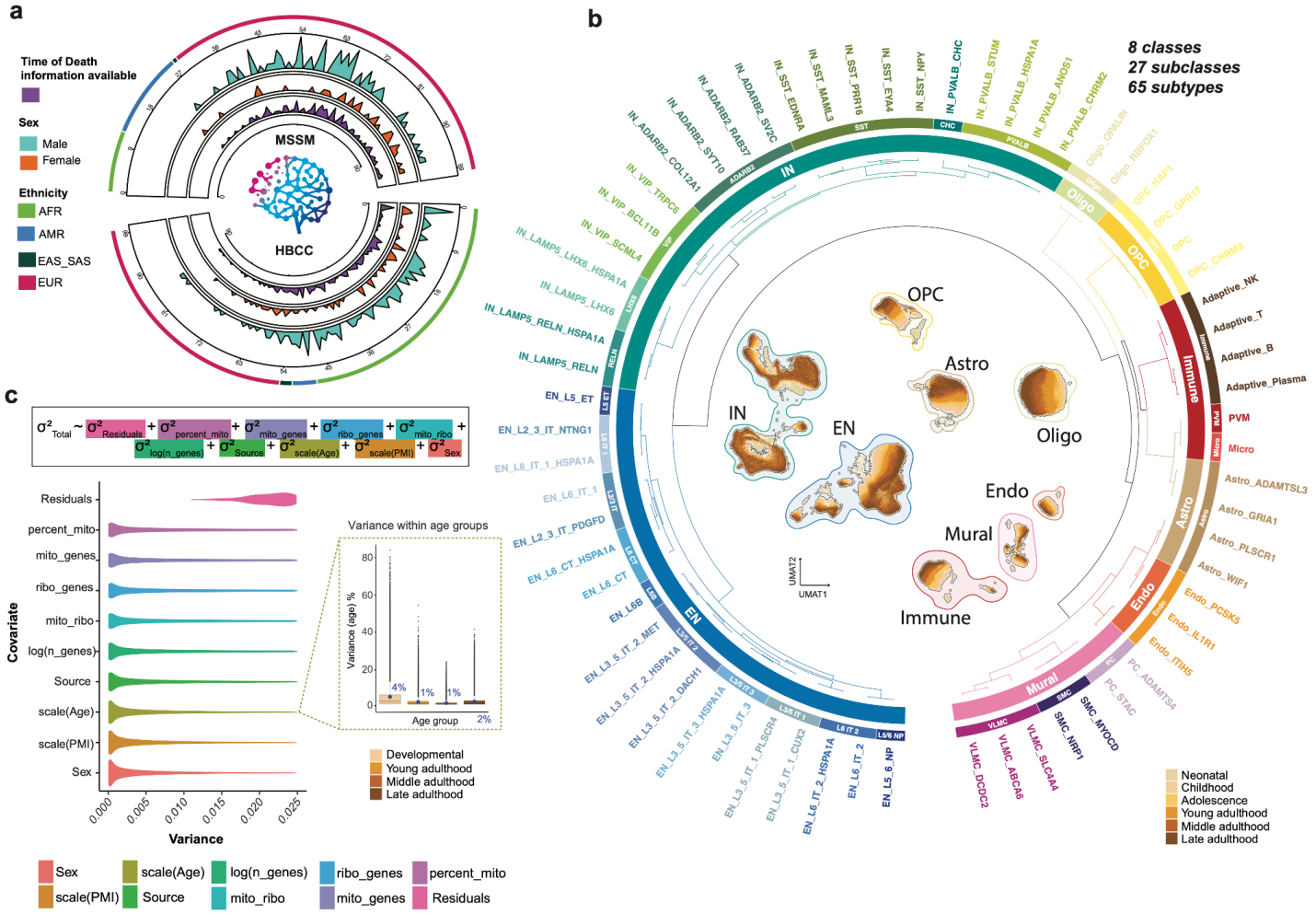


Figure 1. Lifespan transcriptomic atlas of the human prefrontal cortex. a, A schematic of the study design; the age groups used in this study are defined and include: age ranges, number of male and female donors, and total number of nuclei collected for each age group. Schematics (below) highlight the major analysis themes of the study. **b,** UMAP representation of the dataset colored by subclass. **c,** Fraction of nuclei across available age groups, colored by subclass. **d,** Polygenic disease risk scores, computed using the scDRS package[27], are depicted at the subclass level across various neurological, psychiatric and other selected GWAS traits (**Supplemental**

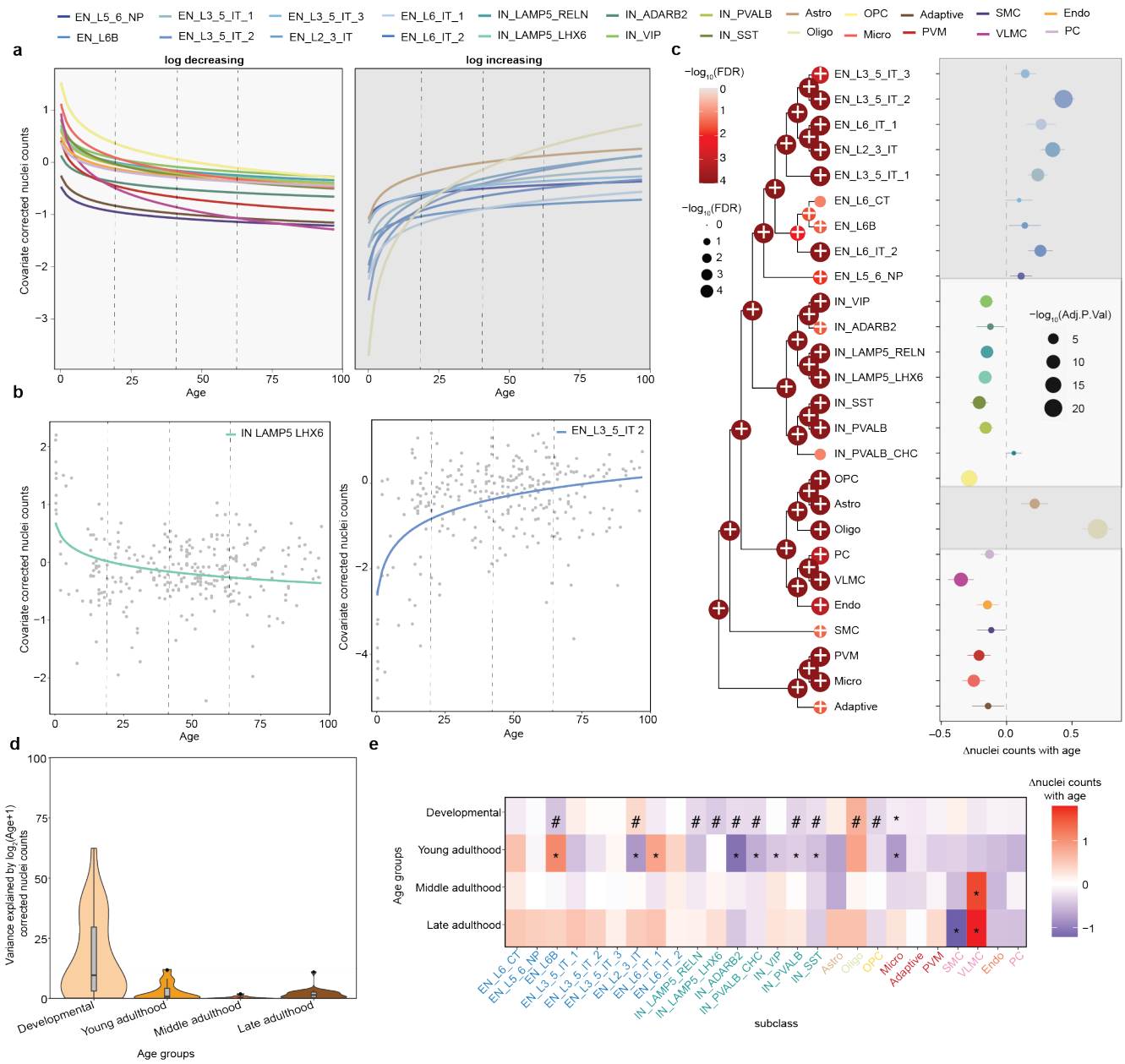
Table 2, Supplemental Data 2), presented as the mean association z-score. "*": indicates significance for enrichment in FDR-corrected P-value **e,** Age-group contribution to variance of gene expression, calculated from the variancePartition package[22]. Boxplot (inset) represents age-group variance within cell types. Statistical significance annotation between selected pairs of age groups in black, for pairs where mean variance of developmental group is larger (EN, IN, Glia), and red for pairs with the opposite effect (Other cells)."*": signifies P-value $< 2.22\times10^{-16}$. Box plots in inset show lower and upper hinges at the 25th and 75th percentiles, with whiskers extending to, at most, 1.5 times the interquartile range (IQR).



Extended Data Figure 1. Summary metadata of 284 donors. a, 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: ancestry (green: African, dark green: East and South Asian, blue: Ad Mixed American, magenta: European), sex (orange Female and cyan Male) and availability of metadata for time of death (purple) are represented as a distribution across donor age. **b,** Transcriptomic similarities across class, subclass and subtype. UMAP of maturation (UMAT)[12] 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. **c,** Combined variance partition[22] results (from developmental, young, middle and late adulthood) depicting respective contributions to variance in gene expression explained by listed covariates (scale(age), source (brain bank), scale(PMI) (post mortem interval), mitochondrial and ribosomal properties[28] (percent_mito, mito_genes, mito_ribo, ribo_genes), log(n_genes) and residuals), ordered from highest to lowest contribution. Boxplot (inset) highlights distribution of age-contributed variance within the respective age groups and shows lower and upper hinges at the 25th and 75th percentiles, with whiskers extending to at most 1.5 times the interquartile range (IQR).

Age-related variation in cell subclasses composition

Following the observation that prominent alterations in the fraction of nuclei counts occur across different age groups (**Fig. 1c**), we sought to understand the underlying changes in nuclei composition associated with age. We quantitatively assessed changes in composition for subclasses comprising at least 500 nuclei (26 out of 27 subclasses) using the *crumblr* (Count Ratio Uncertainty Modeling Based Linear Regression)[29] tool (**Methods**). *Crumblr* applies robust statistical evaluation of the shifts in cellular populations, which revealed that, compared to linear, a logarithmic (log) age trend provided a more accurate fit (**Fig. S2a**). We observed two different age trends: log-increasing, including ENs, Astro, Oligos, and log-decreasing, including INs, OPC, Micro and remaining subclasses (**Fig. S2b** and **Extended Data Fig. 2a-b**). After considering the hierarchical organization of the cell taxonomy using *crumblr*, we identified 24 out of 26 subclasses with significant compositional changes with age (**Extended Data Fig. 2c** and **Supplementary Table 3A**). We hypothesized that the majority of these changes occur during the developmental period, characterized by rapid differentiation and apoptosis of neuronal and non-neuronal subclasses[30]. To test this, we quantified age group and cell subclass specific contributions to changes in nuclei abundance. As expected, the developmental group showed the highest median variance (9.7%) of nuclei composition explained by age compared to adulthood groups (0.22-1.50%) (**Extended Data Fig. 2d**). Statistical analysis of compositional changes in nuclei count with age per subclass (**Supplementary Table 3B-E**) highlighted the developmental and young adulthood age groups, which is consistent with the rapid differentiation and elimination processes that occur during brain development (< 25 yo)[31] (**Extended Data Fig. 2e**), including an increased ratio of Oligo-to-OPC which aligns with the peak of myelination[12,32] (**Extended Data Fig. 2e** and see cellular dynamics of glial lineages across the lifespan section). Taken together, nuclei counts in the DLPFC exhibit a logarithmic trend with age and reach maturity after undergoing either proliferation/differentiation or apoptosis in young adulthood.



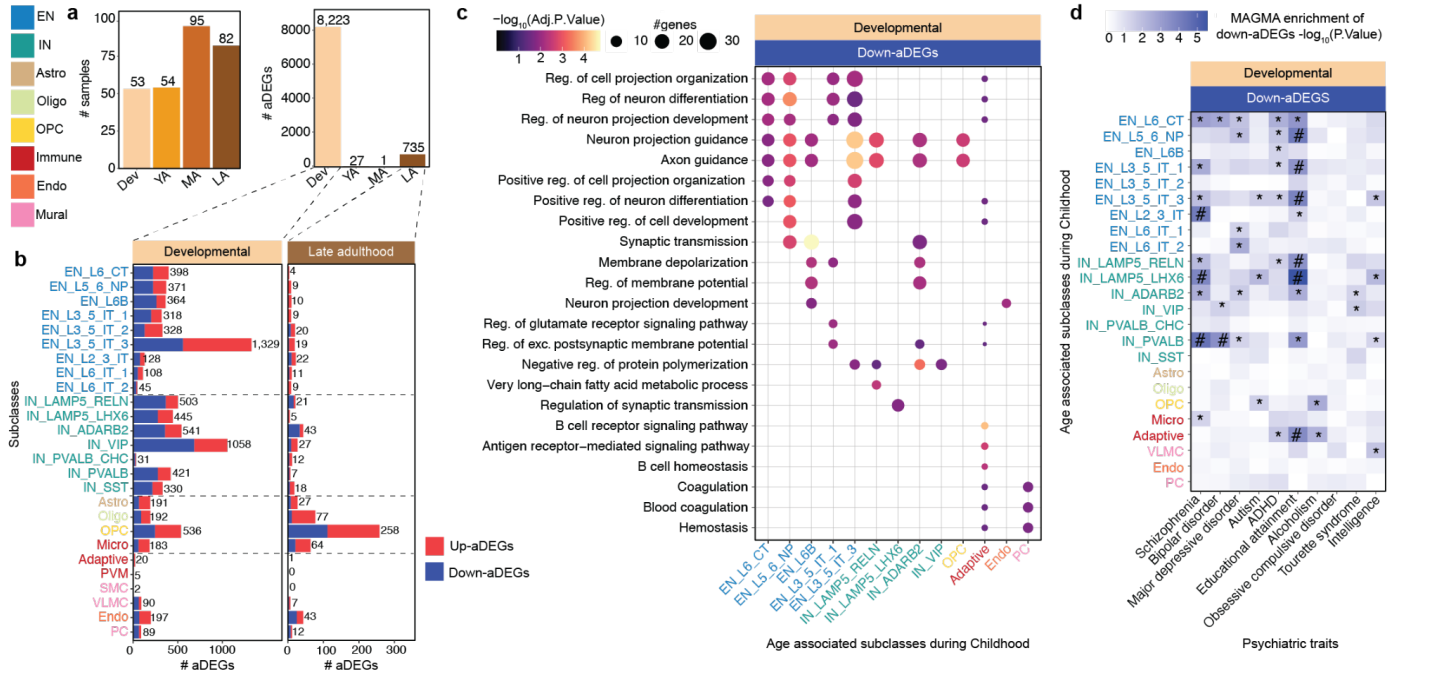
Extended Data Figure 2. Compositional changes of nuclei count across the lifespan. The top row displays the color scheme for each subclass. **a,** Stratification of lifespan trajectories of nuclei counts from all subclasses into logarithmic (log) increasing and decreasing groups. **b,** Examples of log-decreasing and log-increasing trends in IN LAMP5 LHX6 (left) and EN_L3_5_IT_3 (right) subclasses, respectively. **c,** *crumblr* results of univariate hypothesis testing on the leaves and multivariate hypothesis testing on the internal nodes shown on the hierarchical clustering, based on gene expression (left panel). Color and size of each node shows the FDR value from multivariate hypothesis testing. Right panel shows estimated effect size of age-associated changes in nuclei counts across the lifespan for all subclasses. Size of each dot depicts -$\log_{10}$ adjusted p-value from univariate testing. **d,** Fraction of variance in nuclei counts explained by age across early developmental, young, middle, and late adulthood age groups. Beige-brown color gradient depicts age groups ordered by age range. **e,** Estimated effect size of age-associated changes in nuclei counts across four age groups. * and # indicate nominal p-value < 0.05 and FDR < 0.05, respectively.

The DLPFC transcriptome undergoes three major phases between development and late adulthood

Prompted by the significant age-related variance in each subclass (**Fig. 1e)**, we employed *dreamlet*[33] to investigate age-associated changes in gene expression across cell subclasses (**Methods**). This analysis revealed three distinct phases of transcriptional alterations in the DLPFC across the lifespan (**Extended Data Fig. 3a-b** and **Supplemental Data 4**). The first phase, spanning the developmental group, showed significant transcriptional changes across all cell subclasses, with 8,223 **a**ge-associated **D**ifferentially **E**xpressed **G**enes (aDEGs). Major transcriptomic changes in subclasses coincided with significant changes in nuclei composition during this time point (**Extended Data Fig. 2d**). The majority of transcriptional changes were within neuronal subclasses, with 3,389 (41.2%) and 3,329 (40.5%) out of 8,223 aDEGs in EN and INs, respectively (**Extended Data Fig. 3b**). The second phase shows relative transcriptomic stability, with only 27 and 1 aDEGs in young and middle adulthood, respectively. However, the third phase in late adulthood undergoes widespread transcriptional changes, which primarily affect glia (426/735 aDEGs) compared to neurons (246/735 aDEGs) (**Extended Data Fig. 3b**). Counts of aDEGs were corroborated by the observation of pronounced age-associated effect sizes during the developmental phase compared to late adulthood (**Fig. S3**).

We performed gene set pathway analysis to identify biological processes that are enriched for aDEGs during the developmental stage (**Supplemental Data 5)**. Genes whose expression decreased with age (Down-aDEGs) were primarily associated with developmental processes, such as axon guidance and neuronal differentiation (**Extended Data Fig. 3c**). Conversely, aDEGs with elevated expression during developmental stage (Up-aDEGs) were linked to immune and metabolic pathways in EN_L6b, Oligo, Adaptive, Micro and Endo subclasses (**Fig. S4a**). The link between Down-aDEGs and neuronal developmental processes was further corroborated by significant enrichment for risk genes associated with psychiatric disorders, including SCZ and MDD, which have been linked with brain development previously[34,35] (**Extended Data Fig. 3d** and **Supplemental Data 6**). However, Up-aDEGs showed slight or no enrichment for psychiatric disorders across subclasses (**Fig. S4b**). Next, we tested if down-aDEGs are essential developmental genes that are under strong genetic selection, by comparing the tolerance to functional mutations of genes (pLI scores)[36] of aDEGs. Interestingly, down-aDEGs in subclasses such as EN_L3_5_IT_1/3, that were significantly associated with brain related traits (**Extended Data Fig. 3c**), had significantly higher tolerance for loss-of-function mutations compared to up-aDEGs indicating indeed stronger selection of these genes (**Fig. S4c**).

We performed similar gene set pathway analysis to identify biological processes that are enriched for aDEGs during the late adulthood stage (**Supplemental Data 5)**. Glial subclass (Oligo, OPC and Micro) Up-DEGs were enriched for processes such as leucine metabolic process, regulation of DNA binding and protein localization, while OPC and IN Down-DEGs were enriched for chemical synaptic transmission and modulation of postsynaptic potential (**Fig. S4d**). Glial subclasses and IN_ADRAB2 aDEGs were significantly enriched for risk genes related to AD and other neurological and immune-related traits (**Fig. S4e** and **Supplemental Data 6**). Comparison of pLI scores of ~15K coding genes[37] and late adulthood-aDEGs showed no significant association in contrast to developmental aDEGs (**Fig. S4f**).



Extended Data Figure 3. Age-associated changes in gene expression. **a,** Number of samples and the number of age-associated Differentially Expressed Genes (aDEGs) per age group, with a gradient from beige-brown representing the age groups from developmental (Dev), young adulthood (YA), middle adulthood (MA), to late adulthood (LA). **b,** Directionality of aDEGs, where red indicates genes increasing with age (Up-aDEGs) and blue indicates genes decreasing with age (Down-aDEGs), shown for childhood (left) and late adulthood (right). **c,** Functional pathway analysis of Down-aDEGs during development, highlighting subclasses significantly enriched for GO (gene ontology) biological processes with an adjusted p-value < 0.05. **d,** Association of Down-aDEGs with risk genes for a subset of brain related traits using MAGMA. * and # indicate nominal p-value < 0.05 and FDR < 0.05.

Non-linear gene expression trajectories across the lifespan of the DLPFC

After observing significant transcriptomic changes during development and late adulthood, we investigated the lifespan trends of aDEGs from these time points. Our goal for performing this analysis was to map the temporal expression patterns of aDEGs within specific subclasses, to reveal critical periods of transcriptome maturation and late adulthood alterations, and to identify age-related molecular mechanisms through similarities in transcriptional dynamics for each subclass. We first benchmarked lifespan gene expression trajectories for each subclass, by optimizing model selection of expression data from 334,689 genes and 26 subclasses to identify non-linear age-related trends explained by second-degree polynomials (**Fig. 2a, Fig. S5a** and **Methods**). We found that 40.3% (lifespan-aDEGs=135,120/334,689) of genes across all subclasses showed significant age-related changes, clustered into 10 trajectories (**Fig. S5b, Fig. 2b, Fig. S6a-b** and **Supplemental Data 7**). Functional pathway analysis revealed enrichment in processes from basic cellular functions to brain-related synaptic activities (**Supplemental Data 8**). For example, trajectory 1 (16,188 genes from 23 subclasses) highlighted basic cellular processes, while trajectory 10 (5,352 genes from 26 subclasses) indicated subclass-specific transcription programs like interferon signaling and immune activation (**Fig. S7**).

Stratification across the 10 trajectories showed that most of the developmental aDEGs (**Extended Figure 3a**) implicated EN and IN within trajectory 2 and trajectory 7, (**Fig. 2c**). Trajectory 2, which includes the majority of

down-aDEGs, showed a significantly higher association with brain-related traits (**Fig. 2e**) compared to other trajectories (**Fig. S6c** and **Supplemental Data 9**). The abundance of most of the developmental aDEGs (67.6% of genes that are included in trajectories 1, 2, 6 and 7) remained stable after peaking around 12-13 years. Among late adulthood aDEGs, 86.6% were found in trajectories 3-5 and 8-10 (**Fig. 2d**), indicating three distinct gene expression patterns: (a) significant changes during development followed by slower aging (trajectories 3, 8), (b) consistently higher changes across all ages (trajectories 5, 10), and (c) reversal after peaking around 10-12 years (trajectories 4, 9) (**Fig. S8** and **Supplemental Data 9**). Upward trends (trajectories 4, 8, and 10) showed higher prevalence for risk genes of neurological and immune-related disorders (**Fig. 2f**).

Micro late adulthood aDEGs in trajectory 10 showed a significant increase in the magnitude of gene expression with age (**Fig. S9a**) and were enriched for protein localization to the plasma membrane and calcium homeostasis (**Fig. S9b**). These aDEGs were also significantly enriched for AD risk genes compared to all Micro genes in the same trajectory, indicating the role of age associated-Micro in AD. (**Fig. S9c**). Additionally, Micro genes that were not late adulthood aDEGs in other trajectories, particularly 3 and 6, also showed enrichment for AD risk genes, suggesting that these trajectory-specific trends in Micro play a crucial role in AD (**Fig. S9c-d**). Functional pathway analysis revealed that trajectory 3 Micro genes were associated with chromatin modification, while trajectory 6 genes were linked to autophagy (**Supplemental Data 10)**—both molecular mechanisms previously associated with AD[38,39]. Among all neuronal late adulthood aDEGs in the trajectory clusters, only the IN_ADARB2 subclass aDEGs from Trajectory 2 exhibited significant GO molecular mechanisms. (**Fig. S9f** and **Supplemental Data 10**). The IN_ADARB2 genes in Trajectory 2 are downregulated during early brain development, and show a similar trend in late adulthood (**Fig. S9e**). This decline is associated with reduced synaptic plasticity—a molecular mechanism previously linked to aging[40,41].

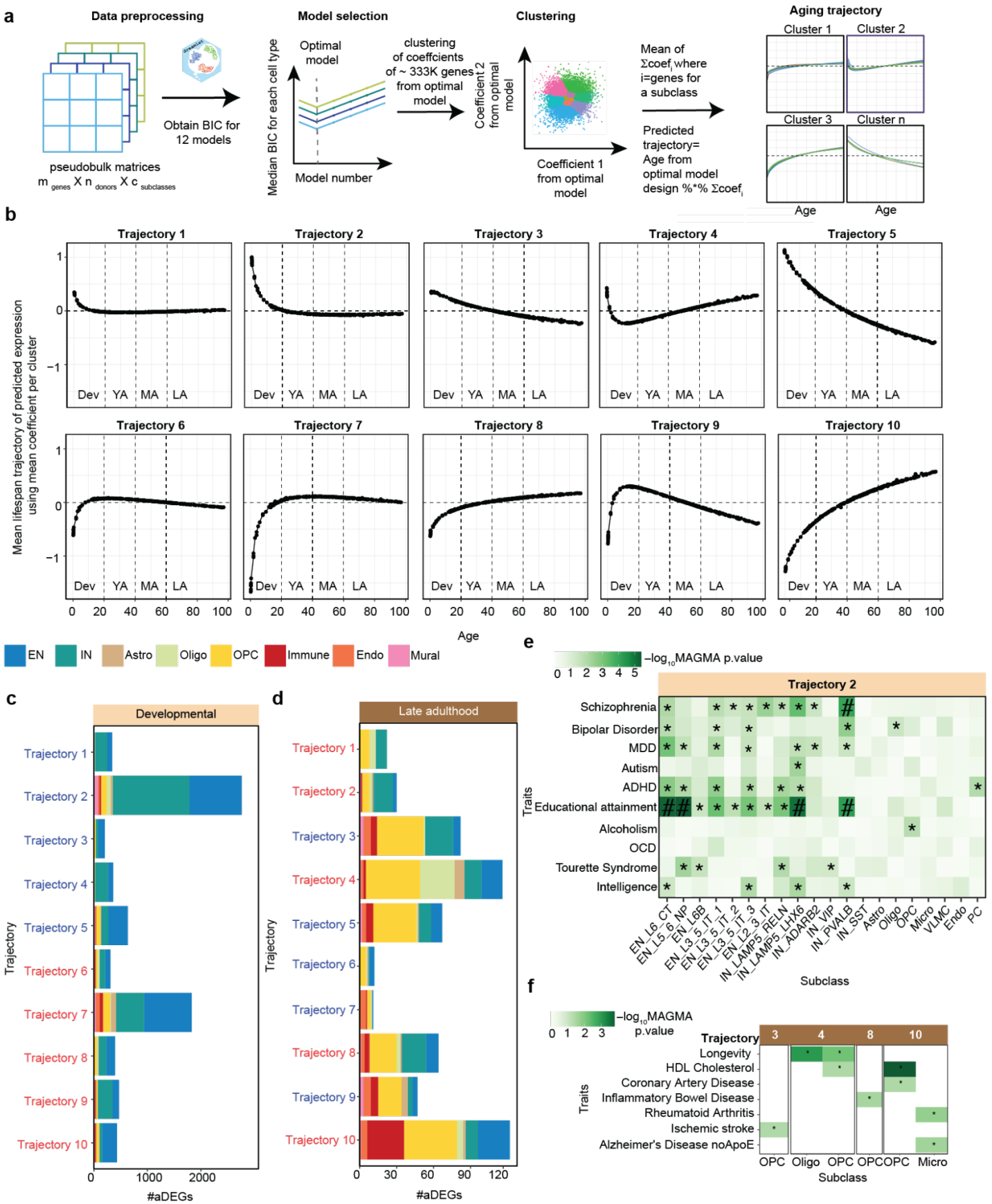


Figure 2. Lifespan gene expression trajectories. a, Schematic of the workflow illustrating the steps: data preprocessing, model selection for final trajectory analysis, clustering, and visualization of trajectories as a function of age. **b,** Ten characteristic trajectories

were identified, derived from the average gene expression of cellular subclasses within each cluster. The trajectories are organized by upward and downward trends. Stratification of **c** developmental aDEGs and **d** late adulthood aDEGs across the trajectories. Colors indicate the number of genes within each class. Trajectory names colored in red and blue indicate upregulation and downregulation, respectively, during development, in **c,** and late adulthood, in **d**. **e-f,** MAGMA enrichment of genes within developmental and late adulthood aDEGs highlighting their relevance to neuropsychiatric and late age diseases. * and # indicate nominal p-value < 0.05 and FDR < 0.05 from MAGMA enrichment.

Degree of sharing in age effect sizes across EN subclasses increases as the DLPFC ages.

We next sought to determine whether the degree of shared age-related changes in gene expression differs between cell subclasses and across different age groups, particularly comparing developmental stages with adulthood. To do this, we quantified the degree of sharing in age-related effect sizes across cell subclasses using *mashr*[42], an empirical Bayes approach that identifies and leverages patterns of similarity among conditions or cell subclasses to improve the age effect size estimates (**Fig. 3a**). We used composite probabilities (P_E, P_I, P_G) as a metric to assess the degree of sharing, based on *mashr* estimated effect sizes across 9 EN, 7 IN, and 4 Glial subclasses, respectively, within each age group (**Methods** and **Supplemental Data 11**). This analysis focused on genes with nominally significant age-associated effect sizes, identified by the *dreamlet* tool, in at least two cell subclasses (**Fig. S10a)**. Our analysis revealed a significant increase in degree of sharing (P_E, P_I, $P_G > 0.9$ in bin = 10) in adulthood compared to development, particularly in EN (**Fig. 3b)**, IN and glia (**Fig. S11a**).

Genes with low sharing ($P_E < 0.01$), such as *KCNH4* ($P_E = 0$), exhibited variable age-associated effect sizes (**Fig. S11b)**. This contrasts with concordant effect sizes across 9 EN subclasses for genes with high sharing ($P_E > 0.9$), such as *PARP2* ($P_E = 0.99$) (**Fig. 3c)**. Analyzing the mean P_E, P_I, P_G of each gene across all age groups revealed an increase in degree of sharing from development to adulthood groups in EN (**Fig. 3d)**, a trend less pronounced in IN and glial subclasses (**Fig. S11c)**. This suggests that developmental transcriptomic changes in EN are divergent (low sharing) but become more convergent (high sharing) with age, particularly in late adulthood. A similar observation was made for developmental and late adulthood aDEGs from the previous section, which are subsets of the genes analyzed here, further corroborating the shift in aging effect sizes across neuronal subclasses from low sharing to high sharing, particularly in EN (**Fig. S12a-b)**. To explore whether genes with high sharing are less likely to be cell type specific, we measured cell specificity using tau scores[43], which range from 0 (low specificity) to 1 (high specificity). Genes with higher sharing ($P_E, P_I, P_G > 0.9$) had significantly lower tau scores, indicating reduced cell specificity compared to genes with lower sharing ($P_E, P_I, P_G < 0.9$) across EN, IN, and glia (**Fig. S13, Fig. S12c** and **Supplementary Table 4**).

We next used functional pathway analysis to identify biological processes associated with shared genes across each age group. During development and young adulthood, shared genes in EN and IN were significantly associated with neuronal cell-cell adhesion and neurotransmitter uptake (**Supplemental Data 12**). In late adulthood, shared genes in EN were linked to DNA repair, a well-known aging mechanism[44,45] (**Fig. 3e**), and, in IN, to the degradation of modified macromolecules and proteins via ubiquitination (**Fig. 3e**). In glia, shared genes were involved in immune surveillance, including antigen processing and presentation, crucial for eliminating

damaged cells and maintaining tissue health during late adulthood[46,47] (**Fig. 3e**). No significant associations were observed in middle adulthood (**Supplemental Data 12**).

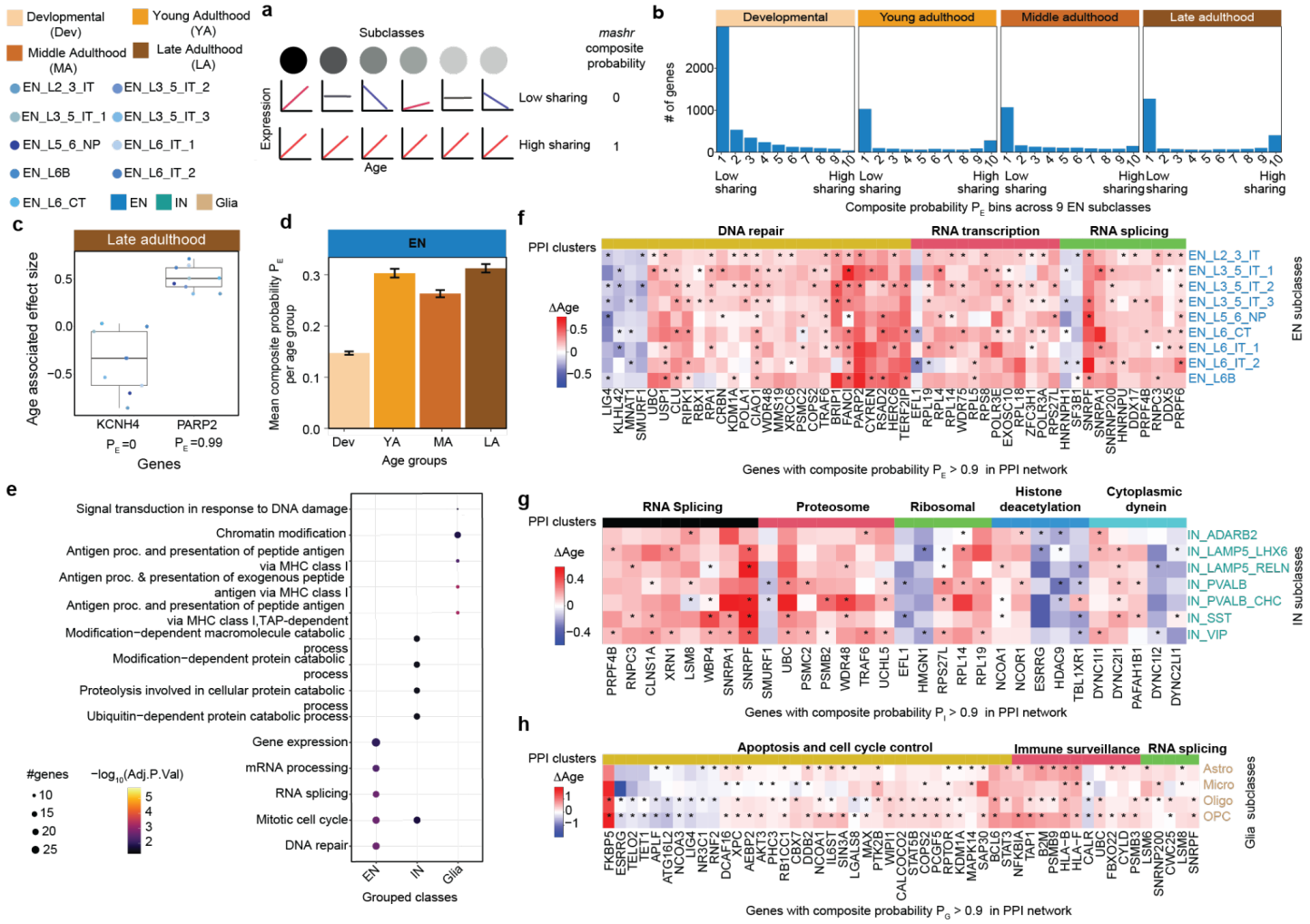


Figure 3. Degree of sharing in age effect sizes and associated biological mechanisms across cell subclasses. The top-left panel displays the color gradient and abbreviations for age groups, subclasses and classes: developmental (Dev), young adulthood (YA), middle adulthood (MA), late adulthood (LA), 9 EN subclasses and EN, IN and glia. **a,** A schematic illustrating low degree sharing with heterogeneous age effect sizes (composite probability = 0) and high degree sharing with concordant effect sizes across subclasses (composite probability = 1) **b,** Number of genes stratified into 10 equally sized bins based on composite probability (P_E) values for EN, ranging from 0 to 1, from development to late adulthood. **c,** Distribution of late adulthood age-associated effect sizes for KCNH4 and PARP2 genes with $P_E = 0$ and $P_E = 0.99$, respectively, across EN subclasses. **d,** Mean P_E of genes in each age group (4,767 Dev, 1,926 YA, 2,092 MA, 2,271 LA). This plot is the mean of probabilities of genes shown in Fig. 3a. **e,** Biological mechanisms of shared genes during late adulthood across 9 EN, 7 IN and 4 glia that show significant enrichment for GO terms after an adjusted p-value < 0.05. **f-h,** Heatmap of age-associated effect sizes of 51, 30 and 51 shared genes across 9 EN, 7 IN and 4 glia classes which showed significant PPI interactions (score > 0.9) and had at least 5 genes within the PPI network. Color bar on the top of the heatmap shows associated mechanisms obtained from k-means clustering of genes within the PPI network. "*" denotes genes that are nominally significant with p-value < 0.05 from age groups analysis using *dreamlet*.

To highlight specific cellular mechanisms associated with shared genes, we used STRING-db[48] to examine protein-protein interactions (PPI) per age group for each class (**Methods**). Clustering of identified PPIs showed

significant associations with basic cellular mechanisms in EN and IN during young and late adulthood, and in glia only during late adulthood. In young adulthood, PPI networks in EN revealed genes encoding proteins for mitochondrial and neuronal projection-specific complexes, while IN showed neuron-to-neuron synapse proteins (**Supplemental Data 13**). In late adulthood, PPIs revealed three clusters in EN and glia, and five clusters in IN. EN and IN clusters included genes crucial for DNA repair, RNA transcription, and RNA processing (**Fig. 3f-g, Fig. S14** and **Supplemental Data 13**). In glia, clusters included proteins involved in apoptosis, MHC class I, and RNA processing and splicing complexes (**Fig. 3h** and **Supplemental Data 13**). Most genes within PPI clusters were upregulated with age, reflecting activation of DNA repair in ENs, protein clearance in INs, and immune surveillance in glia, potentially to maintain cellular homeostasis.

In summary, the age-related landscape of EN shows a transition from divergent to convergent effect sizes from development to young, middle and late adulthood, resulting in a linear increase in the number of shared genes. These genes encode proteins involved in DNA repair, RNA transcription and splicing. In IN and glia, the transition is less clear, but shared genes peak during adulthood, indicating convergent effect sizes. These shared genes are linked to RNA transcription, splicing and immune surveillance in glia.

Cellular dynamics of glial lineages across the lifespan

Recognizing the variance observed during development and late adulthood periods (**Fig. 1e** and **3b**), we employed pseudotime trajectory analysis to delineate cellular dynamics of lineages of the DLPFC across the lifespan. We expanded our dataset to encompass fetal stages by integrating our data with published snRNA-seq from human DLPFC spanning gestation to adulthood[12]. The final dataset comprised 1,454,617 nuclei from 311 individuals (**Fig. S15a-d,** see **Methods**). We utilized UMAP of MATuration (UMAT)[12], which restricts UMAP neighbor selection to cells from adjacent stages, thereby enhancing the precision of representing transitional processes across the lifespan.

We initially investigated the pseudotime trajectories of glial lineages, including Astro, Micro, and the OPC to Oligo transition. For Astros, we identified two distinct age-related patterns (**Fig. 4a** and **Fig. S15e-f**), corresponding to fibrous Astros expressing high levels of glial fibrillary acidic protein (*GFAP*), and protoplasmic Astros with low *GFAP* and high glutamate transporter 2 (*SLC1A2*) (**Fig. 4a** and **Extended Data Fig. 4a**). Spatial gene expression data from DLPFC showed high *GFAP* expression in white matter and high *SLC1A2* expression in gray matter (**Fig. 4b** and **Extended Data Fig. 4b-c**), confirming the established distribution of fibrous and protoplasmic astrocytes in white and gray matter, respectively[49,50].

The two Astro pseudotime trajectories displayed distinct temporal patterns; protoplasmic Astros maturing later than fibrous Astros (**Fig. 4c**). This difference is likely due to their unique cellular contexts: fibrous Astros in white matter interact with myelinated axons and oligodendrocytes, while protoplasmic Astros in gray matter engage with neurons and synapses[51,52]. Furthermore, protoplasmic Astros exhibit consistently higher single cell disease relevance scores (scDRS) for migraines across pseudotime, especially during maturation and aging processes (**Fig. 4d** and **Extended Data Fig. 4d**). These results highlight the significance of cellular interactions and temporal dynamics in understanding Astro functions and their role in disease susceptibility.

To further investigate the lifespan gene expression dynamics of the two Astro pseudotime trajectories, we identified 827 differentially expressed genes along trajectories (traDEGs, FDR < 0.05, Moran's I ≥ 0.05) and grouped them into six patterns (**Methods**), broadly categorized into developmental (dev), mature (mat), and aging processes (**Fig. 4e**, **Extended Data Fig. 4e**, and **Supplemental Data 14**). GO functional enrichment analysis[53] (**Fig. 4f** and **Supplemental Data 15**) revealed that developmental genes are significantly involved in nervous system development, axonogenesis, and neuron projection guidance, consistent with the established roles of Astros in neural circuit formation[54,55]. Developmental and mature (dev-mat) genes are enriched in functions related to the blood-brain barrier and vascular transport, aligning with the roles of Astros in maintaining neuronal homeostasis and nutrient delivery[56,57]. Aging genes, on the other hand, show enrichment in regulation of macromolecule biosynthetic process and response to cytokine, consistent with the involvement of Astros in immune response biosynthesis and regulation[49]. Protoplasmic Astro-specific (PA-spec) genes are enriched in neuron projection development and cell projection morphogenesis, reflecting their contributions to shape neural architecture and facilitate synaptic connectivity. Conversely, fibrous Astro-specific aging (FA-spec aging) genes are linked to memory, negative regulation of amyloid fibril formation, and cellular response to organic cyclic compounds, underscoring their involvement in cognitive support, neuroprotection, and cellular defense mechanisms[49]. These findings emphasize the dynamic, process-specific roles of Astro subtypes throughout the lifespan, highlighting their distinct functional specializations.

Genetic risk gene enrichment analysis showed that developmental genes for both Astro categories are associated with several psychiatric traits (**Fig. 4g, Methods** and **Supplemental Data 16**), highlighting their critical role in early brain development and mental health[58–60]. In contrast, aging related genes are enriched in several neurological and immunological traits, consistent with their involvement in immune regulation (**Fig. 4f**)[61–63]. Notably, PA-spec genes are significantly enriched in SCZ and migraines, suggesting a cooperation between neurons and protoplasmic Astros in the pathophysiology of these two traits[64]. These findings underscore the significance of Astro gene expression in brain function and disease across the lifespan.

We also examined other glial lineages, including Micro and Oligos. For Micro, we identified a single pseudotime trajectory (**Fig. 4h** and **Extended Data Fig. 4f**) and observed significant and stable enrichment in several neurological traits, including AD and multiple sclerosis, as well as immunological traits, such as rheumatoid arthritis, IBD, ulcerative colitis (**Extended Data Fig. 4g**). Similar to astrocytes, developmental genes in the Micro lineage are associated with neuron development and psychiatric disorders, while aging genes are mainly associated with immunological functions (**Fig. 4i-j, g, Extended Data Fig. 4h** and **Supplemental Data 14-16**). These findings highlight Micro's roles in neurodegenerative and immune traits across the lifespan[65,66]. Furthermore, enrichment patterns from developmental to aging genes provide additional insights into AD progression beyond those identified in previous GWAS studies focused solely on aged donors, uncovering genetic influences that emerge later in life (**Fig. 4g, i** and **j**). For the Oligo lineage, we identified a trajectory from OPC to mature Oligos (**Fig. 4k** and **Extended Data Fig. 4i-j**) and observed a significant transition in trait enrichment along this trajectory (**Extended Data Fig. 4k**). Specifically, OPC shows notable enrichment in obesity as well as a number of psychiatric traits, whereas mature Oligos exhibit specific enrichment for AD. Additionally, developmental genes in the Oligo lineage are primarily associated with chemical synaptic transmission and multiple psychiatric disorders, while aging genes are linked to myelination, protein localization to plasma membrane and obsessive compulsive disorder (**Fig.4l-m**, **Extended Data Fig. 4l** and **Supplemental Data 14-16**).

Taken together, these findings highlight the importance of distinct temporal dynamics across the lifespan of glial cell lineages and susceptibility for psychiatric, neurological, and immune disorders.

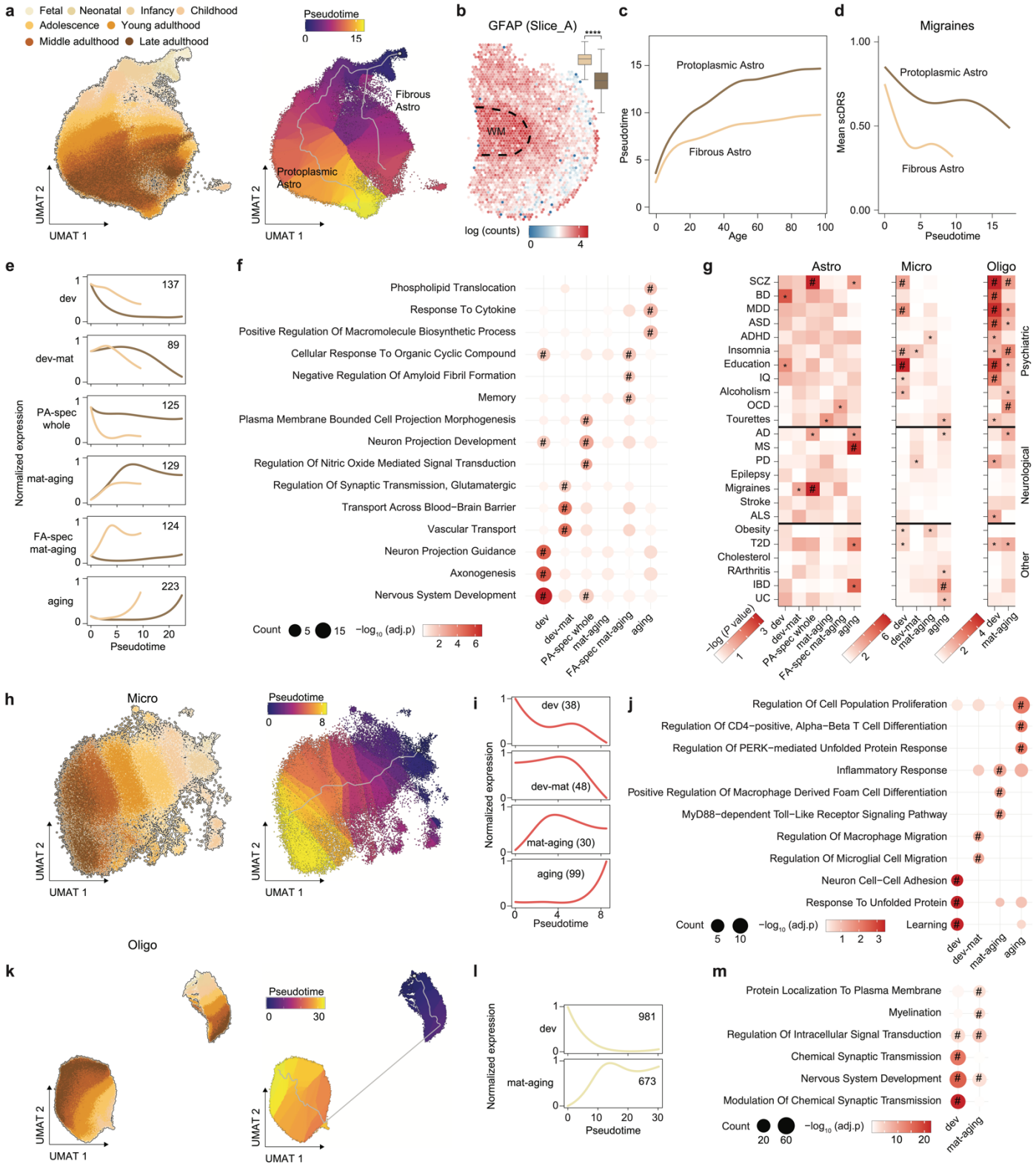
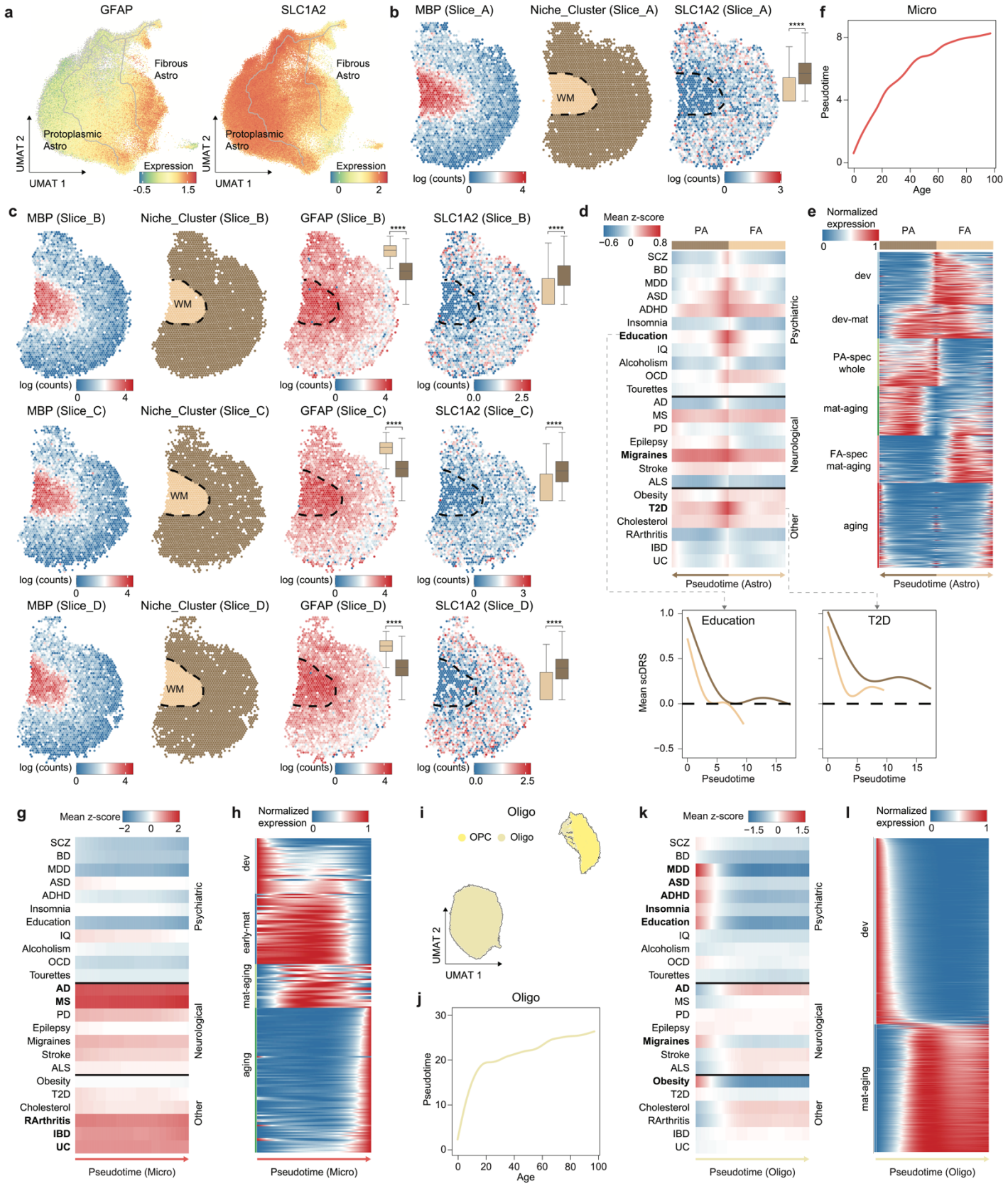


Figure 4. Cellular dynamics of glial lineages. a, UMAT representation of astrocyte lineage colored by stage (left) and pseudotime (right). **b**, Expression of *GFAP* in Visium spatial transcriptomics data of a section of cortex (Slice_A). The interface between white matter (WM) and gray matter is indicated by a dashed line. Two-sided p-value was computed using a Mann–Whitney *U*-test. ****: p-value < 0.0001. **c**, Maturation rates of two astrocyte categories. **d**, Association between scDRS disease score of migraines and pseudotime of the two astrocyte categories. **e**, Scaled (0-1) expression of differentially expressed genes along trajectories (traDEGs) in the astrocyte lineage clustered into six patterns. Fibrous astrocytes (FA) are indicated in light brown and protoplasmic astrocytes (PA) are indicated in dark brown. **f**, Enriched GO terms corresponding to traDEG clusters in the astrocyte lineage. #: adjusted p-value < 0.05. **g**, Enrichment adjusted *P* value of different classes of GWAS traits corresponding to traDEG clusters in Astro (left), Micro (middle) and Oligo (right) lineages. *: p-value < 0.05, #: adjusted p-value < 0.05. **h-m**, Cellular dynamics of the Micro and Oligo lineages. UMAT representation of the Micro (**h**) and Oligo (**k**) lineages, colored by stage (left) and pseudotime (right). Clusters of traDEGs in Micro (**i**) and Oligo (**l**) lineages, with corresponding enriched GO terms (**j**, **m**).


a
GFAP
SLC1A2
Fibrous Astro
Protoplasmic Astro
UMAT 2
UMAT 1
Expression
-0.5
1.5
0
2
b
MBP (Slice_A)
Niche_Cluster (Slice_A)
SLC1A2 (Slice_A)
WM
log (counts)
0
4
3
****
f
Micro
Pseudotime
Age
c
MBP (Slice_B)
Niche_Cluster (Slice_B)
GFAP (Slice_B)
SLC1A2 (Slice_B)
MBP (Slice_C)
Niche_Cluster (Slice_C)
GFAP (Slice_C)
SLC1A2 (Slice_C)
MBP (Slice_D)
Niche_Cluster (Slice_D)
GFAP (Slice_D)
SLC1A2 (Slice_D)
0.0
2.5
d
Mean z-score
−0.6
0.8
PA
FA
SCZ
BD
MDD
ASD
ADHD
Insomnia
Education
IQ
Alcoholism
OCD
Tourettes
AD
MS
PD
Epilepsy
Migraines
Stroke
ALS
Obesity
T2D
Cholesterol
RArthritis
IBD
UC
Psychiatric
Neurological
Other
Pseudotime (Astro)
e
Normalized expression
dev
dev-mat
PA-spec whole
mat-aging
FA-spec mat-aging
aging
Education
T2D
Mean scDRS
Pseudotime
g
Mean z-score
−2
2
Pseudotime (Micro)
h
Normalized expression
early-mat
i
Oligo
OPC
j
k
−1.5
1.5
Pseudotime (Oligo)
l

Extended Data Figure 4. Characteristics of glial lineages. a, Markers of fibrous and protoplasmic astrocytes. **b**,**c**, Expression of markers of fibrous and protoplasmic astrocytes in Visium spatial transcriptomics data of sections of cortex. *MBP* is a marker of white matter (WM). White and gray matter were annotated using ONTraC[68] and are separated by a dashed line. Two-sided p-values were computed using a Mann–Whitney *U*-test. ****: p-value < 0.0001. **d**, Fitted disease scores along the pseudotime trajectory for the two astrocyte categories (**Methods**). Traits that were significantly enriched are shown in bold and visualized in detail using the line plots below. **e**, Fitted expression of traDEG clusters along the pseudotime trajectory for the two astrocyte categories. **f-l**, Characteristics of the Micro and Oligo lineages. Maturation rates of the Micro (**f**) and Oligo (**j**) lineages. Fitted disease scores along the pseudotime trajectories for Micro (**g**) and Oligo (**k**) lineages. Fitted expression of traDEG clusters along the pseudotime trajectory for the Micro (**h**) and Oligo (**l**) lineages. UMAT representation of the OPC and Oligo lineages colored by subclass (**i**).

Cellular dynamics of neuronal lineages across the lifespan

We next focused on neuronal lineages, including ENs and INs. For ENs, we identified nine trajectories grouped into three categories based on their cortical distribution: upper-layer intratelencephalic projection neurons (Upper-IT), deep-layer intratelencephalic projection neurons (Deep-IT), and deep-layer non-intratelencephalic projection neurons (Deep-non-IT) (**Fig. 5a**, and **Extended Data Fig. 5a**). The pseudotime trajectories of these three categories of ENs displayed distinct temporal patterns, with Deep-non-IT ENs maturing earliest, followed by Deep-IT and Upper-IT ENs (**Fig. 5b** and **Extended Data Fig. 5b**), which is consistent with the well-known inside-out pattern of cortical development and migration[69,70].

scDRS analysis revealed that the lifespan processes of various ENs are significantly associated with most psychiatric diseases, with this enrichment observable from the earliest stages of development (**Fig. 5c** and **Extended Data Fig. 5c-d**). For example, SCZ shows significant associations during developmental and mature stages, suggesting a critical window for its emergence[35,71–73]. Notably, Tourette's syndrome and obesity become progressively associated during developmental processes of most ENs, indicating that neuronal maturation may play a crucial role in these traits. This extends previous findings, which primarily observed these associations with INs[74,75]. In ENs, we identified 2,686 traDEGs, forming eight distinct patterns (**Fig. 5d**, **Extended Data Fig. 5e** and **Supplementary Data 14**). Developmental genes are primarily associated with neuron generation and development, while mature and aging genes are broadly enriched in synapse transmission and ion transport (**Fig. 5e** and **Supplementary Data 15**), essential for synaptic function and structural integrity throughout life[76,77]. These genes show significant trajectory variability, especially between upper and deep layers. Specifically, upper-IT-specific developmental genes are enriched in neuron projection development and regulation of inclusion body assembly. Mature and aging genes in L5_6_NP ENs exhibit enrichment in non-neuronal processes such as extracellular matrix and external encapsulating structure organization, supporting a role in neural circuit integrity and adaptability[78]. Furthermore, the traDEGs show significant enrichment in psychiatric disorders (**Fig. 5f** and **Supplementary Data 16**) and, with the exception of bipolar disorder, this enrichment is more pronounced for developmental genes. Together, these findings underscore the critical roles for trajectory-dependent genes in neuronal development, synaptic maintenance, and disease associations, particularly in the context of psychiatric disorders.

For INs, we identified eleven trajectories that can be grouped into two categories: medial ganglionic eminence (MGE)-derived and caudal ganglionic eminence (CGE)-derived INs (**Fig. 5g** and **Extended Data Fig. 5f**). Overall, the maturation processes of CGE- and MGE-derived INs show no significant differences, but CGE-derived INs exhibit greater variability across different trajectories (**Fig. 5h** and **Extended Data Fig. 5g**),

suggesting distinct temporal dynamics in the lifespan of these IN categories. The scDRS analysis showed associations with most psychiatric diseases that are present from early development, similar to ENs (**Fig. 5i** and **Extended Data Fig. 5h-i**). In INs, we identified 2,036 traDEGs, forming eight distinct patterns (**Fig. 5j, Extended Data Fig. 5j** and **Supplementary Data 14**). IN development is enriched in genes associated with nervous system development, axonogenesis and axon guidance (**Fig. 5k** and **Supplementary Data 15)**. Additionally, CGE-specific developmental genes are involved in regulation of intracellular signal transduction. Mature and aging genes are broadly enriched in synaptic transmission and ion transport, contributing to the stability and adaptability of neural circuits throughout life[76,77]. Specifically, mature and aging genes in sporadic trajectories are enriched in processes related to maintaining extracellular matrix integrity and structural stability, similar to the observation for L5_6_NP ENs (**Fig. 5e**). Furthermore, enrichment in psychiatric disorders is also observed for inhibitory neuron traDEGs (**Fig. 5l** and **Supplementary Data 16**), with the enrichment being more pronounced in developmental genes. Notably, PVALB_CHC-specific mature and aging genes show significant enrichment in alcoholism and stroke, potentially due to their involvement in response to calcium ions (**Fig. 5k**). These findings highlight distinct temporal dynamics and disease associations among neuronal lineages, underscoring the importance of developmental, mature and aging genes in susceptibility to a range of psychiatric and neurodegenerative disorders.

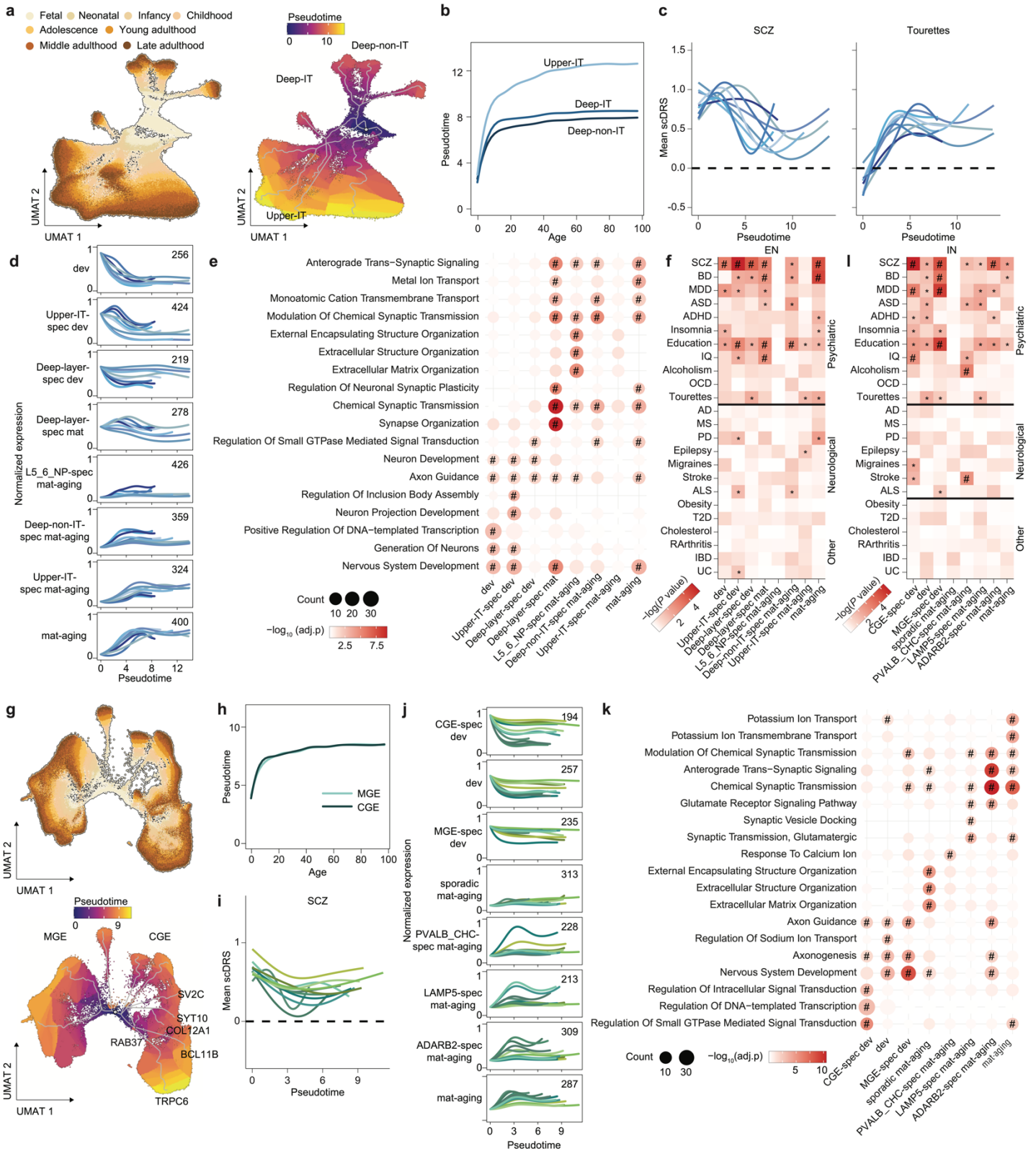
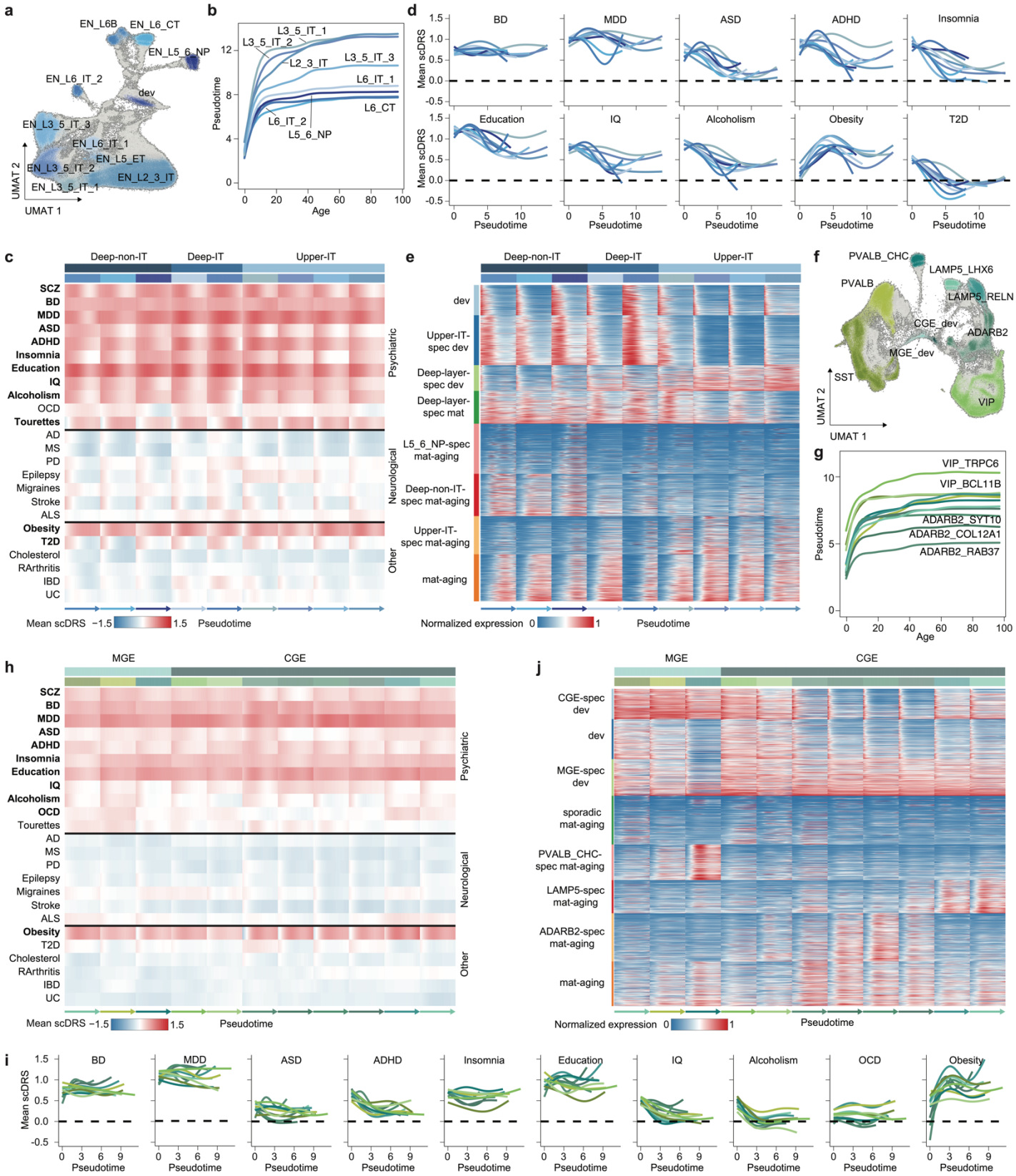


Figure 5. Cellular dynamics of neuronal lineages. a, UMAT representation of EN lineage colored by stage (left) and pseudotime (right). **b**, Maturation rates of three EN categories. **c**, Association between scDRS of SCZ (left) and Tourette's (right) and pseudotime of EN trajectories. **d**, Scaled (0-1) expression of the traDEGs in EN lineage clustered into eight patterns. **e**, Enriched GO terms corresponding to traDEG clusters in the EN lineage. #: adjusted p-value < 0.05. **f**, Enrichment adjusted p-value of different classes of

GWAS traits corresponding to traDEG clusters in EN lineage. *: p-value < 0.05, #: adjusted p-value < 0.05. **g**, UMAT representation of IN lineage, colored by stage (upper) and pseudotime (bottom). **h**, Maturation rates of two IN categories. **i**, Associations between scDRS of SCZ and pseudotime of IN trajectories. **j**, Clusters of traDEGs in IN lineage clustered into eight patterns. **k**, Enriched GO terms corresponding to traDEG clusters in the IN lineage. #: adjusted p-value < 0.05. **l,** Enrichment adjusted p-value of different classes of GWAS traits corresponding to traDEG clusters in IN lineage. *: p-value < 0.05, #: adjusted p-value < 0.05.


a
EN_L6B
EN_L6_CT
EN_L5_6_NP
EN_L6_IT_2
dev
EN_L3_5_IT_3
EN_L6_IT_1
EN_L5_ET
EN_L3_5_IT_2
EN_L2_3_IT
EN_L3_5_IT_1
UMAT 2
UMAT 1
b
L3_5_IT_1
L3_5_IT_2
L2_3_IT
L3_5_IT_3
L6_IT_1
L6_CT
L6_IT_2
L5_6_NP
Pseudotime
Age
d
BD
MDD
ASD
ADHD
Insomnia
Education
IQ
Alcoholism
Obesity
T2D
Mean scDRS
Pseudotime
c
Deep-non-IT
Deep-IT
Upper-IT
SCZ
BD
MDD
ASD
ADHD
Insomnia
Education
IQ
Alcoholism
OCD
Tourettes
AD
MS
PD
Epilepsy
Migraines
Stroke
ALS
Obesity
T2D
Cholesterol
RArthritis
IBD
UC
Psychiatric
Neurological
Other
Mean scDRS −1.5 1.5 Pseudotime
e
dev
Upper-IT-spec dev
Deep-layer-spec dev
Deep-layer-spec mat
L5_6_NP-spec mat-aging
Deep-non-IT-spec mat-aging
Upper-IT-spec mat-aging
mat-aging
Normalized expression 0 1 Pseudotime
f
PVALB_CHC
LAMP5_LHX6
PVALB
LAMP5_RELN
CGE_dev
ADARB2
MGE_dev
SST
VIP
UMAT 2
UMAT 1
g
VIP_TRPC6
VIP_BCL11B
ADARB2_SYT10
ADARB2_COL12A1
ADARB2_RAB37
Pseudotime
Age
h
MGE
CGE
SCZ
BD
MDD
ASD
ADHD
Insomnia
Education
IQ
Alcoholism
OCD
Tourettes
AD
MS
PD
Epilepsy
Migraines
Stroke
ALS
Obesity
T2D
Cholesterol
RArthritis
IBD
UC
Psychiatric
Neurological
Other
Mean scDRS −1.5 1.5 Pseudotime
j
MGE
CGE
CGE-spec dev
dev
MGE-spec dev
sporadic mat-aging
PVALB_CHC-spec mat-aging
LAMP5-spec mat-aging
ADARB2-spec mat-aging
mat-aging
Normalized expression 0 1 Pseudotime
i
BD
MDD
ASD
ADHD
Insomnia
Education
IQ
Alcoholism
OCD
Obesity
Mean scDRS
Pseudotime

Extended Data Figure 5. Cellular dynamics of EN and IN lineages. a, UMAT representation of the EN lineage colored by subclass. **b**, Maturation rates of nine EN trajectories. **c**, Fitted disease scores along the pseudotime trajectory for the EN lineage. Traits that were significantly enriched are shown in bold and visualized in detail using the line plot (**d**). **e**, Fitted expression of traDEG clusters along the pseudotime trajectory for the EN lineage. **f**, UMAT representation of the IN colored by subclass. **g**, Maturation rates of ten IN trajectories. **h**, Fitted disease scores along the pseudotime trajectory for the IN lineage. Traits that were significantly enriched are shown in bold and visualized in detail using the line plot at panel (**i**). **j**, Fitted expression of traDEG clusters along the pseudotime trajectory for the IN lineage.

Circadian reprogramming in the late adulthood DLPFC transcriptome

Having demonstrated that the DLPFC transcriptome exhibits the most dramatic changes during late adulthood compared to other adulthood groups, we next examined the vulnerability of the transcriptome to changes in circadian rhythms, a hallmark of aging[79,80]. To do this, we used a cosinor model to identify 24 h gene expression rhythms within groups of subjects with known times of death (TOD) (**Extended Data Fig. 6, Fig. 6a** and **Methods**). Our goal was to examine the rhythmicity of the transcriptome during young and middle adulthood compared to late adulthood.

We identified genes with a 24 h rhythm using covariate corrected expression matrices (**Methods**) of each cell subclass in young and middle adulthood (YA+MA, n = 116) and late adulthood (LA, n = 76) (**Fig. 6a** and **Supplementary Data 17**). 45 significant rhythmic genes (FDR < 0.05) were observed in 9 subclasses of YA+MA, with 84% (38/45) found in upper layer IT ENs (**Fig. 6b**). As expected, many of these genes are associated with the circadian molecular clock (**Extended Data Fig. 6b**). Versican (VCAN) was the only significant rhythmic gene (FDR < 0.05) observed in LA (**Fig. 6b**). For a deeper analysis, we loosened our rhythmicity significance threshold (p-value < 0.01) and found that, for most subclasses, more rhythmic genes were observed in YA+MA than in LA (**Fig. 6c** and **Extended Data Fig. 6c-d**). Some notable exceptions were EN_L6_CT, EN_L6B, IN_LAMP5_LHX6, Micro, and Oligo, which had more rhythmic genes in LA (**Fig. 6c** and **Extended Data Fig. 6c-d**). However, comparative rhythmicity analyses identified very few genes that were rhythmic in both groups within each subclass, indicating that the identity of transcripts with rhythmic gene expression differs between YA+MA and LA (**Fig. 6c** and **Supplementary Data 17**). Rhythmic genes peaked in expression either slightly after sunrise (~2 ZT), or in the evening (~12-14 ZT), regardless of age group (**Fig. 6d**). However, of the few genes that were identified as having rhythms in both age groups, 46% (28/61) had significantly different peak times (FDR < 0.05) of expression (**Supplementary Data 18**), further emphasizing the differences between these age groups. Overall, consistent with previous studies that find transcriptomic-wide age-dependent changes in rhythmic gene expression[80,81], these findings suggest that circadian reprogramming occurs in LA across subclasses.

We next determined which genes were rhythmic within each age group. Core circadian clock genes were the top rhythmic genes across neuronal subclasses in YA+MA (**Fig. 6e-h**, **Extended Data Fig. 6e** and **Supplementary Data 17**). The time of peak expression for each of these genes was consistent across subclasses, with the forward limb peaking at night, and regulatory arms peaking in the morning (**Fig. 6h** and **Extended Data Fig. 6b**). However, in LA these rhythms are lost (**Fig. 6e-h, Extended Data Fig 6d,** and **Supplementary Data 19**), and, as expected from non-rhythmic genes, the peak times calculated for these genes were not consistent across

subclasses (**Fig. 6h**). This is supported by enrichR[53] pathway analyses, which consistently identified circadian clock-related pathways as top pathways in subclasses of the YA+MA group, but not LA (**Fig. 6i, Fig. S16** and **Supplementary Data 20**). Alternatively, a notable pathway that was enriched in LA Micro and Oligo was Response to Unfolded Protein (**Fig 6i** and **Fig. S16**). This was particularly interesting as these two subclasses, uniquely, had more rhythmic genes in LA than in YA+MA. We also split genes by when their expression peaked ("Sunrise" vs "Sunset") and found that circadian pathways were enriched in genes that peaked around sunrise in YA+MA, while genes associated with response to unfolded protein peaked around sunset in LA in Micro, PC, VLMC, and Oligo (**Fig 6i** and **Supplementary Data 20**). Overall, while the expected circadian clock signature was observed in YA+MA neuronal subclasses, this signal was lost in LA. Genes that gained rhythmicity in LA were not consistent across subclasses, though Micro and Oligo did have a specific enrichment for the unfolded protein response, which may represent a response to increased cellular stress and/or activity of these subclasses as a component of aging[82].

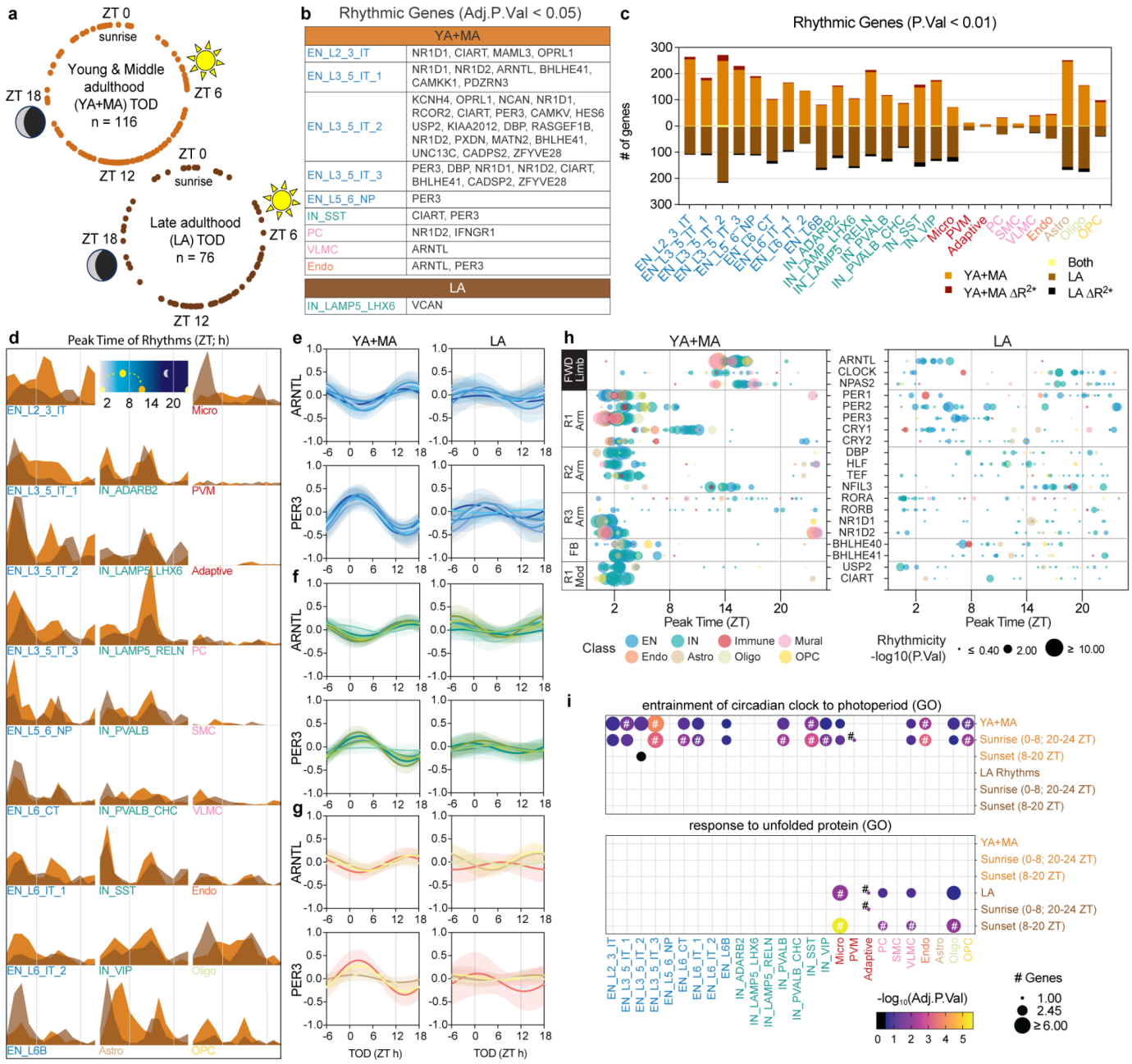
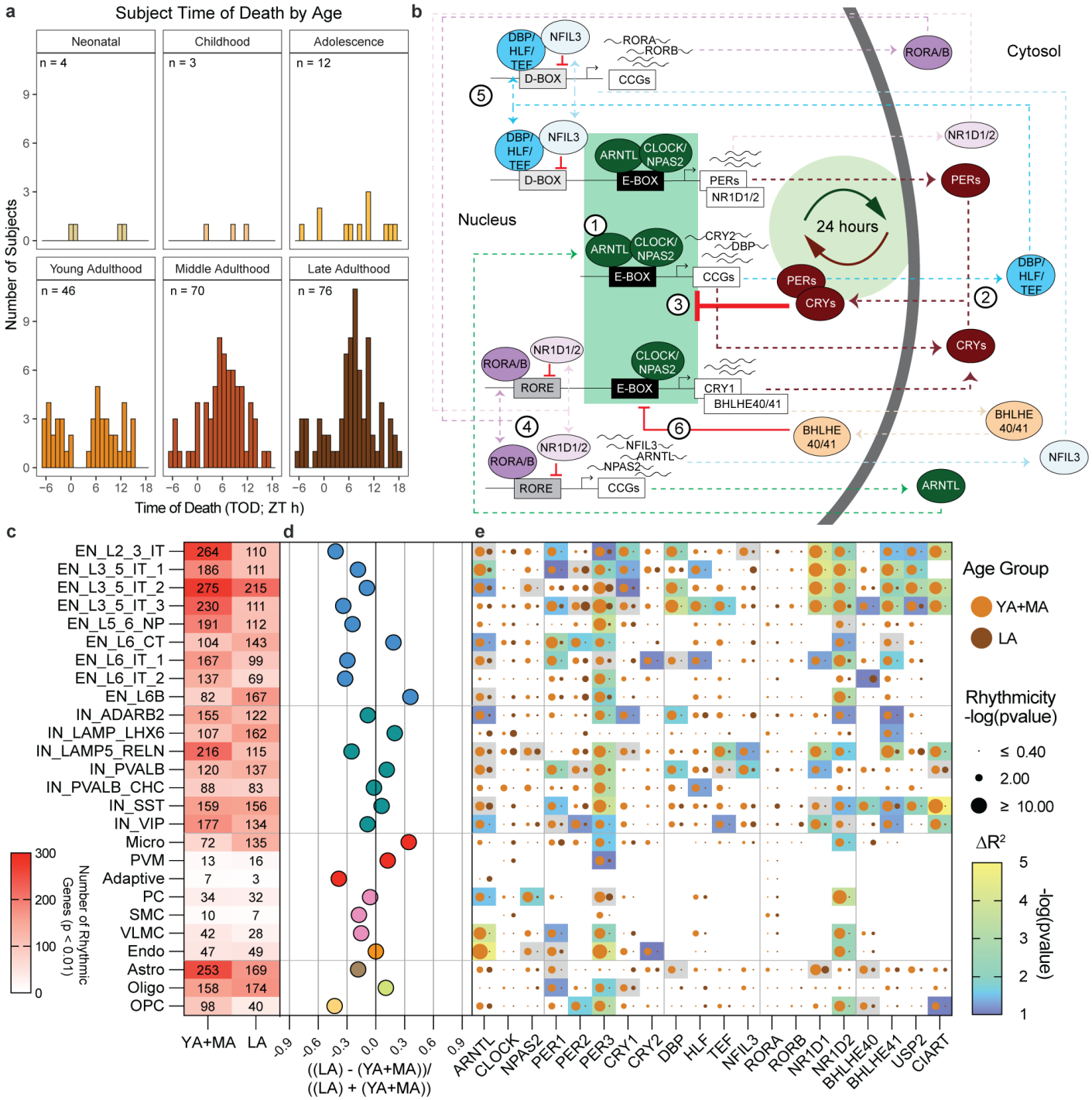


Figure 6. Circadian reprogramming in late adulthood. a, Time of death (TOD) information for subjects used in the rhythmicity analysis. Each dot is an individual's TOD in zeitgeiber time (ZT). For full 24 h coverage, Young and Middle adulthood groups were combined (YA+MA; 21 - 60 years), and then compared to the Late adulthood (LA; ≥ 61 years) group. **b,** Genes identified as rhythmic (FDR < 0.05). **c,** Number of genes identified as rhythmic ($p < 0.01$) for YA+MA, LA or Both groups. Rhythmic genes with significantly different R^2 values ($p < 0.01$) are denoted as dark red or black. **d**, Peak expression times for all rhythmic genes. **e - g,** Example rhythms for ARNTL and PER3, two canonical molecular clock genes, in **e,** Excitatory Neuron (EN) subclasses, **f**, Inhibitory Neuron (IN) subclasses, **g**, and glia, including microglia, astrocytes, oligodendrocytes and oligodendrocyte progenitor cells (OPC). Solid lines are the calculated 24 h oscillation, transparent area is the 95% confidence interval. Individual plots are available in **Fig. S17-18**. **h**, Rhythmicity and timing of genes associated with the circadian molecular clock in YA+MA and LA. Represented pathways include the forward limb (FWD limb), the primary PER/CRY regulatory arm (R1), the secondary D-Box regulatory arm (R2), the secondary ROR/NR1D regulatory arm (R3), the BHLHE40/41 feedback loop (FB), and modifiers of the R1 arm. (R1 mod). **i**, Important pathways identified by enrichR. # represents FDR < 0.05.



Extended Data Figure 6. Age associated changes in 24 h gene expression rhythms. a, Time of Death (TOD) distributions within each age group. **b,** The core of the molecular clock is a forward limb (1) that drives 3 major regulatory arms (2/3, 4, 5) in an interconnected series of transcription-translation feedback loops (reviewed in[83]). **c**, Number of rhythmic genes ($p < 0.01$) in each subclass in YA+MA and LA. **d**, Percent difference in the number of rhythmic genes within each subclass between YA+MA and LA. **e**, Comparing rhythmicity (size) of circadian clock genes between YA+MA and LA. Additionally, the results of testing for difference in rhythmicity (ΔR^2) is included for genes that were significantly rhythmic ($p < 0.01$) in at least one group.

Discussion

We present a single-cell transcriptomic atlas of the human DLPFC, generated from over 1.3 million nuclei across 284 postmortem samples, spanning birth to 97 years of age. Using this resource, we gained several key insights into the dynamic changes of the DLPFC transcriptome across the lifespan. During development, 2.9% and 1.7% of age-related gene expression changes occurred in neuronal and glial subclasses, respectively. These changes stabilized after age 20 but significantly resurged after age 60, predominantly affecting glial subclasses compared to neurons. While previous studies noted the vulnerability of neurons during development[84] and the increased susceptibility of glial cells with aging[85,86], our study quantitatively compared the aging landscape across cells and demonstrated neurons are transcriptionally more resilient to aging.

Quantification of the degree of sharing in transcriptomes across all age groups uncovered three patterns. First, during development, different cell types exhibit distinct transcriptome programs, which is consistent with previous studies showing that cell-specific gene expression patterns are tightly regulated during development, leading to divergent transcriptomic profiles among different cell types[87]. Second, during aging, cellular stressors such as DNA damage, oxidative stress, and inflammatory responses become more pronounced, which activate shared pathways across different cell types, leading to a convergence of transcriptome signatures. Neurons are particularly vulnerable to DNA damage due to their high metabolic activity and low regenerative capacity, activating shared gene expression signatures related to DNA repair and RNA splicing[88–90]. Third, compared to glia, there is a higher convergence in the neuronal transcriptome during late adulthood. Neurons, as long-lived post-mitotic cells, rely on common pathways for maintaining integrity and function under aging conditions, leading to activation of common stress response pathways across neuronal subtypes[90].

Mapping of distinct trends in the transcriptome across the lifespan organized them into ten distinct non-linear age trajectories, highlighting that 67.6% of developmental transcriptomes remain stable after peaking around ages 12-13. The stability of gene expression after early adolescence is crucial for maintaining the fundamental functions of the brain, including cognitive processes, memory, and neural circuitry, and shows similar age-related patterns with normative brain growth quantified based on MRI imaging[91]. Trajectory 2 showed a downward trend during development and was linked to multiple risk genes for neurodevelopmental disorders, suggesting that disruptions that occur before this critical stabilization may have lasting impacts on brain function[92]. Interestingly, the IN_ADARB2 subclass in trajectory 2 showed downregulation in both early development and late adulthood, suggesting a shared molecular mechanism affecting genes related to synaptic plasticity. Synaptic plasticity is crucial for learning, memory, and cognitive function, and its regulation is vital during periods of significant brain remodeling, such as in early development and aging[93].

We applied pseudotime approaches to identify dynamically expressed genes along lineage trajectories, and their disease associations, revealing critical windows of neurodevelopmental vulnerability and aging. Psychiatric disorder-associated genes maintain high expression in neuronal lineages across the lifespan, indicating sustained involvement in brain function and diverse roles in development and maintenance[84,91,94]. These genes are also highly expressed during the development of astrocyte and oligodendrocyte lineages, highlighting the crucial glial-neuronal interplay necessary for proper brain function[64]. Conversely, genes linked to neurodegenerative diseases are highly expressed in microglia throughout the lifespan and prevail in oligodendrocyte lineages during aging, suggesting distinct roles for those two cell types in age-related degeneration[66,95]. Further analysis of traDEGs reveals trajectory-specific disease associations, including the enrichment of SCZ risk in various EN cell types and

gene modules across development and aging. In glial lineages, there is a notable transition from enrichment of neurodevelopmental and psychiatric traits in developmental genes to neuroinflammation and neurodegenerative traits in aging genes, reflecting their evolving roles from supporting development to maintaining homeostasis and responding to age related damage[66,96].

We explored circadian rhythm disruption, a hallmark of aging, by using cosinor analysis to analyze gene expression rhythms in different age groups. Our findings suggest that an intact, rhythmic molecular clock, which is crucial for regulating sleep-wake patterns, cognitive function, and cellular metabolism[79–81], is a fundamental component during young and middle adulthood, particularly in neuronal subclasses. However, in late adulthood, rhythmicity of core circadian clock genes is lost, with different genes gaining a 24-hour rhythm, consistent with circadian reprogramming observed in aging across various species[79,80,81]. Notably, microglia and oligodendrocytes showed an increase in rhythmic genes in late adulthood, enriched for genes involved in the unfolded protein response, indicating a potential adaptive response to increased protein synthesis demands and cellular stress during aging[81]. This study, the first to examine molecular rhythms in the human DLPFC at the single-cell level, provides insights into the mechanisms underlying age-related functional changes in this brain region.

In conclusion, our comprehensive single-cell transcriptomic atlas of the human DLPFC provides unprecedented insights into the dynamic molecular landscape of the brain across the human lifespan. This resource reveals how distinct cellular processes are tightly regulated during critical periods of development and aging, with key findings highlighting the resilience of neuronal transcriptional programs and the vulnerability of glial cells in late adulthood. The identification of trajectory-specific disease associations, including the role of circadian rhythm disruption and the differential expression of psychiatric and neurodegenerative disorder-associated genes, underscores the complexity of brain aging and the potential for targeted interventions. Our findings not only enhance our understanding of the molecular underpinnings of brain function but also pave the way for future research aimed at mitigating age-related cognitive decline and neurodegenerative diseases.

Methods

DLPFC lifespan study design

Brain tissue specimens were obtained from NIMH-IRP Human Brain Collection Core (HBCC): 172 samples, ages 0.2-85 years and The Mount Sinai NIH Neurobiobank (MSSM): 112 samples, ages 20-97 years. In total, 284 neurotypical controls of age range 0-97 years from "PsychAD dataset" were included in this study (**Fig. S1a**). The majority of the samples are from European descent (n = 158) followed by African (n = 95), American (n = 26), East Asian and South Asian (n = 5). **Extended Data Figure 1a** shows the demographic information at donor level, including sex, age, time of death and ancestry, stratified by corresponding brain bank. The data was categorized into four groups: 1) developmental which constitutes neonatal (0-1 year), childhood (2-11 years), adolescence (12-19 years), 2) young (20-39 years), 3) middle (40-59 years) and 4) late adulthood (≥ 60 years) samples.

We utilized the available neuropathology details on MSSM samples for control samples selection. The selection criterion for MSSM included the following:

1) CERAD scores: For neuritic plaque density of MSSM samples, only those with a score of 1 (no AD) were included.
2) Braak stage: Samples with Braak stages 0, 1, or 2 were retained.
3) Secondary diagnoses: Donors with any additional brain-related diagnosis, including neurodegenerative (e.g. AD and PD) and neuropsychiatric diseases (e.g. SCZ and BD) as well as the presence of mild cognitive impairment, were not kept.

In principle, we applied equivalent selection criteria to the HBCC samples. Although the HBCC cohort did not provide specific Braak and CERAD values, we confirmed through review of neuropathological reports that the selected donors did not exhibit significant plaque or tangle pathology. Thus, all selected donors who lacked brain-related diagnoses were deemed reliable neurotypical controls, despite the absence of detailed neuropathological data.

The link to the complete demographic and clinical information of the present study population is provided in **Supplemental Table 1** and **Supplemental Data 1**.

Fluorescence-activated nuclear sorting (FANS) protocol and snRNA-seq hashing from frozen brain tissue

All libraries from the "PsychAD" dataset were prepared using a standardized protocol for nuclei isolation and hashing[35]. The lifespan dataset was generated after completing snRNA-seq preprocessing and the taxonomy step, as described in this and succeeding sections. The process involved isolating and sorting nuclei from frozen brain specimens using fluorescence-activated nuclear sorting (FANS). 25 mg of frozen postmortem human brain tissue was homogenized in a cold lysis buffer with RNAse inhibitors. The homogenate was filtered through a 40 µm cell strainer, and the flow-through was underlaid with sucrose solution before centrifugation at 107,000 xg for 1 hour at 4 ˚C. The resulting pellets were resuspended in PBS with 0.5% bovine serum albumin (BSA). Six samples were processed simultaneously, with up to 2 million nuclei per sample pelleted at 500 xg for 5 minutes at 4 ˚C. The nuclei were then resuspended in 100 µl staining buffer and incubated with 1 µg of a unique TotalSeq-A nuclear hashing antibody (BioLegend) for 30 minutes at 4 ˚C. Prior to FANS, the volumes were adjusted to 250 µl with PBS, and 7-Aminoactinomycin D (7-AAD) was added according to the manufacturer's instructions. The 7-AAD positive nuclei were sorted into tubes pre-coated with 5% BSA using a FACSAria flow cytometer (BD Biosciences).

After FANS, the nuclei were washed twice with 200 µl of staining buffer, resuspended in PBS and quantified using the Countess II (Life Technologies). The concentrations were adjusted, and equal volumes of differentially hash-tagged nuclei were combined. Using 10x Genomics single cell 3' v3.1 reagents, 60,000 nuclei (10,000 per donor) were processed in each of two lanes to create technical replicates. During cDNA amplification for library preparation, 1 µl of a 2 µm HTO cDNA PCR additive primer[97] was included. The supernatant from a 0.6x SPRI selection was reserved for HTO library generation. Both cDNA and HTO libraries were prepared according to manufacturer's instructions (10x Genomics and BioLegend, respectively). Sequencing was performed at the NYGC using the Novaseq platform (Illumina).

Preprocessing of the snRNA-seq dataset

Paired-end snRNA-seq library reads were mapped to the hg38 reference genome using STARsolo[98]. Sample pools were then demultiplexed by matching genotypes with vireoSNP[99]. After generating count matrices for each library, further downstream analysis was conducted with *Pegasus* (v1.7.0)[100] and *Scanpy* (v1.9.1)[28].

A thorough three-step quality control (QC) process was implemented to eliminate ambient RNA and ensure the retention of high-quality nuclei for subsequent analysis. Initially, QC was conducted at the cellular level, where low-quality nuclei were identified by setting thresholds based on UMI counts, gene counts, and mitochondrial content. We also assessed potential contamination from ambient RNA, as well as the proportion of reads mapped to non-mRNA categories like rRNA, sRNA, and pseudogenes, in addition to examining confounding factors such as the lncRNA MALAT1. The second step involved feature-level QC, where features not robustly expressed in at least 0.05% of nuclei were excluded. Finally, QC was performed at the donor level, removing donors with low nuclei counts, which could introduce noise into downstream analyses, and those with poor genotype concordance. Additionally, doublets were filtered out using the Scrublet method[101]. Next, we applied *Harmony* (v0.1)[102] to correct for unwanted factors, including the influence of the brain tissue source.

Joint cellular taxonomy and dimensionality reduction

A modular approach was employed to define the cellular taxonomy of the "PsychAD dataset" [14,15] from which we isolated the Aging cohort. Starting with a dataset of over 6 million nuclei ("PsychAD dataset"), 8 major cell classes were identified through several key steps. Initially, 6,000 highly variable genes (HVGs) were selected based on mean and dispersion trends, using default parameters (min_mean = 0.0125, max_mean = 3, min_disp=0.5) and excluding sex chromosomes, mitochondrial genes, and using "MT" as batch variables. A k-nearest-neighbor (kNN) graph was generated using the harmony-corrected PCA embedding space, which facilitated clustering of nuclei by cell type using the Leiden clustering algorithm. UMAP projection was then utilized to visualize these clusters. For each class-level cluster, data were subsetted, and HVGs were recalculated within each class to fine-tune the feature space relevant to that class. A new kNN graph was generated based on the harmony-corrected PCA of these HVGs, followed by Leiden clustering to annotate subclass-level identities. This iterative process resulted in the identification of 65 subtypes of human brain cells. Canonical markers of major cell types, reference datasets[95,103] as well as spatial validation were utilised for class/subclass/subtype-level labels of the joint "PsychAD dataset". The Aging cohort was then extracted and the UMAP re-calculated using the same approach; for the full Aging dataset, 6,000 HVGs were identified using the `hvg` function in *Pegasus* (v1.7.0)[100], while excluding mitochondria and sex-specific genes. Principal component analysis (PCA) was performed on the gene expression data, followed by batch correction using *Harmony* (v0.1)[102]. Subsequently, UMAP dimensionality reduction was generated based on the first 30 harmony-adjusted principal components[104] (**Fig. 1b**). Finally, a nearest-neighbour graph using the first 30 harmony-adjusted principal components was calculated using the `sc.pp.leiden` function. UMAT calculation (**Extended Data Fig.1b**) was run based on the approach introduced in[12]. The taxonomy labels were retained from the joint "PsychAD dataset".

Single-cell polygenic disease risk score

We utilized *scDRS* package (v1.0.1)[28] to evaluate the combined expression of potential disease-associated genes obtained from GWAS summary statistics through MAGMA analysis [105]. Each gene's contribution was weighted by its MAGMA z-score from GWAS and inversely weighted by its gene-specific technical noise level in the single-cell data. This process was performed across each cell of the Aging single nuclei dataset to generate raw disease scores specific to each. Additionally, we generated 200 sets of raw control scores, matched in gene set size, mean expression, and expression variance to the disease-associated genes. Subsequently, we normalized both the raw disease scores and raw control scores for each cell, resulting in normalized scores. These calculations were executed using the `scdrs.score_cell` function, with the following parameters:

scdrs.score_cell(--ctrl_match_key="mean_var",--n_ctrl=200,--weight_opt="vs", --return_ctrl_raw_score=False,--return_ctrl_norm_score=True,--verbose=False)

For further analysis, subclass-level examinations were conducted to link predefined subclasses to disease and evaluate the heterogeneity in disease association across cells within each predefined subclass level of taxonomy. This was achieved using the `scdrs.method.downstream_group_analysis` function with default settings. The output from this step is provided in **Supplemental Data 2.** All 24 GWAS traits utilized in the study are listed in **Supplemental Table S2**.

scDRS analysis for the pseudotime section was performed consistently across each cell of the integrated single nuclei dataset. Single-cell disease scores along each trajectory were compressed into 500 meta-cells per trajectory. Those scores were then modelled using a generalized linear model $disease\ score \sim splines::ns(pseudotime, df = 3)$.

Psuedobulk data aggregation and covariate selection

To quantify the variance explained by age, age-associated changes in the transcriptome and lifespan trends, we pseudobulked gene expression data by aggregating the counts for each donor from 1,307,674 nuclei, stratified by subclass and four age groups (developmental, young, middle and late adulthood) using the `aggregateToPseudoBulk` function from *dreamlet* (v1.1.17) R package. Note, all steps in this section are done using functions from the *dreamlet* (v1.1.17) R package. Subsequently, we applied voom normalization to the age group-subclass-specific gene-by-donor matrix using the `processAssays` function, which filters for samples with at least 5 nuclei per subclass and genes with a minimum of 5 reads per donor. **Fig. S1c-d** shows the counts of donors and genes obtained after this step. This process generated four lists containing genes x donor matrices for each subclass corresponding to each age group.

With $Source$, Sex, and PMI as the base model, technical and biological covariates were identified using a forward stepwise regression approach on subclass-specific pseudobulk expression data on the entire "PsychAD dataset". Since the lifespan nuclei in this study are a subset of the PsychAD dataset, we implemented the same model for i subclass and the j gene in our analysis as described in equation (eq) (1). From hereon we use these variables as $covariates$ variable in other models in succeeding sections.

$$expression_{i,j} = Sex + PMI + log(n_genes) + mito_genes + mito_ribo + percent_mito + ribo_genes \quad (1)$$

The details of these covariates are as follows: PMI (post-mortem interval in hours) is coded as a scaled numerical variable, while $Source$ and Sex are categorical variables. Source is excluded from the equation for any downstream analyses for the developmental group, since all samples are from a single brain bank (HBCC) (**Supplemental Data1**). The number of genes from each nuclei n_genes, and the proportion of mitochondrial genes per donor $percent_mito$ are summarized for each donor and were obtained using the `qc_metrics` function from the *pegasus* (v1.8.1)[100] package. Scores for reference mitochondrial genes from chrM $mito_genes$ and mitochondrial ribosomal genes not from chrM $mito$, as well as ribosomal genes from both large and small units $ribo_genes$, are estimated for each nucleus and summarized for each donor using the `calc_signature_score` function from *pegasus* (v1.8.1)[100]. **Fig. S1b** shows the pairwise correlation of these covariates with age.

Quantification of variance in age

We quantified the variance explained by age for age group-subclass-specific pseudobulk expression data for i subclass and the j gene as $expression_{i,j} = Age + covariates$ in the `fitVarPart` function from *dreamlet* (v1.1.17) R package. The covariates were obtained from eq(1). **Fig. 1e** highlights the mean contribution of numerical variable age in pseudobulk expression for each subclass across four age groups. **Extended Data Fig.1c** shows the combined contribution of covariates and age from four groups. The output from this section is provided in **Supplemental Data 3**.

Lifespan dynamics of nuclei counts

To quantify changes in nuclei counts as a function of age we utilized `crumblr` function from *crumblr*[106] (Count Ratio Uncertainty Modeling Based Linear Regression) (v0.99.6) R package. The package employs a three-step process to quantify the association of changes in nuclei counts with user-defined independent variables. 1) Normalization: Nuclei counts are normalized using a Dirichlet multinomial distribution. 2) Modeling: A standard dream precision-weighted linear mixed model with empirical Bayes estimation is implemented to obtain association statistics for each measurement. 3) Multivariate hypothesis testing: This step enables the joint analysis of internal nodes in a hierarchical clustering of subclasses, which is crucial for accounting for correlations among related subclasses, such as those among EN subclasses. The sample code of these steps section is provided in the repository (**code availability section**)

In this section, we performed two analyses to quantify changes in nuclei counts: (a) as a function of age across the entire lifespan, and (b) within specific age groups. These analyses were limited to subclasses with at least 500 nuclei counts resulting in 26 out of 27 subclasses. For both analyses, we explored the optimal model to represent nuclei counts as a function of age, aiming to determine whether a linear or logarithmic relationship would be optimal. To test these models, we first regressed out the subset of covariates: sex, PMI, and source from nuclei

counts for i subclass across 284 donors using the equations (2) and (3). The other covariates are more relevant to gene expression analysis so we kept them out for nuclei composition analysis.

$$\widehat{nuclei_counts} \sim Sex + PMI + Source \quad (2)$$

$$res_counts_i = nuclei_counts_i - \widehat{nuclei_counts} \quad (3)$$

After this, we tested two models $res_counts_i \sim Age$, $res_counts_i = log_2(Age+1)$ and obtained the difference of Bayesian Information Criterion (BIC) ΔBIC_i for i subclass. Using ΔBIC_i metric, we found $log_2(Age+1)$ showed optimal relationship between age and changes in nuclei counts for most subclasses as shown in **Fig. S2a**. Applying k-means clustering to subclass specific coefficients from the $log_2(Age+1)$ model **(Fig. S2b)**, we identified two distinct clusters: one with positive coefficients indicating a log-increasing trend, and another with negative coefficients indicating a log-decreasing trend, as illustrated in **Extended Data Fig. 2a**. To further quantify these lifespan associations, we performed *crumblr* enabled modeling and multivariate test analysis using the eq(4). The coefficient of $log_2(Age+1)$ is shown in **Extended Data Fig. 2c** and output is provided in **Supplemental Table 3**.

$$nuclei_counts_i = log_2(Age+1) + Sex + PMI + Source \quad (4)$$

Next, we performed age-group specific analysis by first quantifying the variance explained by age in normalized nuclei counts for each age group using `fitExtractVarPartModel` from *variancePartition*[22] (v1.33.11) R package including eq(4) as a model. **Extended Data Fig. 2d** shows the distribution of variance explained by $log_2(Age+1)$ in 26 subclasses from four age groups. Subsequently, we quantified the coefficients of $log_2(Age+1)$ and performed *crumblr* enabled modeling and multivariate hypothesis test using the eq(4) for each age group ("source" covariate was removed from eq(4) for developmental group). **Extended Data Fig. 2c** shows the heatmap of coefficient of $log_2(Age+1)$ for each subclass and age group**.** The results of this analysis are provided in **Supplemental Table 3**.

Age-associated transcriptomic changes

To assess the age-related transcriptomic changes for each subclass and age group, we performed *dreamlet* enabled limma modeling on age group-subclass-specific pseudobulk expression data as described in the "Psuedobulk data aggregation and covariates selection". The model implemented for i subclass and the j gene is $expression_{i,j} = Age + covariates$. We conducted final multiple testing correction at the study-wide level (26 subclasses × number of genes per subclass) for each group. The total number of study-wide genes from all subclasses were: developmental (*n* = 328,991), young adulthood (*n* = 333,471), middle adulthood (*n* = 340,084), and late adulthood (*n* = 292,748). After applying a threshold of FDR < 0.05, we obtained age-associated differential expression genes (aDEGs) per age group, as shown in **Extended Data Fig. 3a-b** and **Fig. S3**. We provided the results of this analysis in **Supplemental Data 4**. For gene set pathway analysis to identify biological processes, we used the `enrichr` function from the *enrichR* (v3.2) R library. The database used was GO_Biological_Process_2021, and the results are shown in **Extended Data Fig. 3c** and **Fig. S4d.** The pathway enrichment analysis for all aDEGs per subclass is provided in **Supplemental Data 5.**

Lifespan trends of gene expression

To obtain the lifespan trend of each gene (n = 334,689 genes) across 26 subclasses, analysis was performed in three steps. 1) model selection, 2) clustering and 3) prediction of average trajectory as depicted in **Fig. 2a**. The details of these steps are explained below.

Model selection: Our objective was to identify a single optimal model that could capture age related non-linear trends in gene expression across all 26 subclasses. We achieved this by fitting 12 different models to expression of i subclass and the j gene for a total of 334,689 genes using the `dream` function as described in eq(5-8).

$$expression_{i,j} = Age + covariates \tag{5}$$
$$expression_{i,j} = poly(Age, df = n) + covariates \tag{6}$$
$$expression_{i,j} = log_2(Age + 1) + covariates \tag{7}$$
$$expression_{i,j} = poly(log_2(Age + 1), df = n) + covariates \tag{8}$$

n in eq(6,8) represent degrees of freedom which are $2, 3, 4, 5, 6$ and `poly` is R function from *stats* base R library, which produces orthogonal polynomials with a given degree of freedom in df argument. To identify the optimal model, we collected the BIC values from each of the 12 models and searched for the minimum values. Although BIC values across all models were not conclusive for glia and other subclasses, a notable decrease in BIC was observed for the $poly(log_2(Age + 1), df = 2) + covariates$ model specifically in neurons, particularly IN (**Fig. S5a**). Since approximately 60% of the subclasses are neuronal, and for simplicity, we adopted the $poly(log_2(Age + 1), df = 2) + covariates$ model as the optimal lifespan model for all 26 subclasses. All steps in this section were performed using the dream function from *dreamlet* (v1.1.17) R package. The output from optimal model consists of linear $coef^1_{ij}$ and non-linear coefficient $coef^2_{ij}$ from the model for 334,689 genes.

Clustering: Our goal was to identify the optimal number of unique characteristic curves for the lifespan trends of 334,689 genes. To achieve this, we applied k-means clustering to $coef^1_{ij}$ and $coef^2_{ij}$ from the dreamlet summary statistics of the optimal lifespan model for 334,689 genes. Clustering all coefficients revealed that k = 10 provided the optimal clusters with non-overlapping lifespan trends (**Fig. S5b**). **Fig. S5c** shows the stratification of all genes across these 10 clusters. For all downstream analyses, we focused on the 135,120 genes that remained after applying study-wide multiple testing correction (FDR < 0.05) from the optimal model on the 334,689 genes. **Fig. S5d** illustrates the stratification of lifespan-aDEGs (135,120 genes) across the 10 trajectories. The output table containing $coef^1_{ij}$ and $coef^2_{ij}$, along with other summary statistics from the *dreamlet* tool for each gene and subclass, and trajectory cluster numbers, is provided in **Supplemental Data 7**.

Prediction of average lifespan trend: Finally, to visualise an average trajectory per cluster, we calculated the mean of the coefficients for all genes within each cluster, as follows:
$\overline{coef^1_m} = \sum_{k=1}^{N} coef^1_{k,m}$, $\overline{coef^2_m} = \sum_{k=1}^{N} coef^2_{k,m}$ where N is total number of genes in m^{th} cluster and k is j gene in i subclass in m^{th} cluster. Next we obtained non-linear form of age using $\hat{Age} \sim 0 + poly(log_2(Age + 1), df = 2)$. Using the mean of the coefficients and the polynomial form of age $\hat{Age}$, we obtained lifespan trajectory for m cluster $lifespan_trajectory_m = \hat{Age} * (\overline{coef^1_m}, \overline{coef^2_m})$ as shown in **Fig. 2b. Figure S6b** shows predicted average trajectory for each subclass in a trajectory cluster using the mean of subclass coefficients

$lifespan_trajectory_{m,j} = \hat{Age} * (\overline{coef^1_{m,j}}, \overline{coef^2_{m,j}})$. To further obtain biological insight on genes within trajectory clusters, we performed gene set pathway analysis using the `enrichr` function from *enrichR* (v 3.2) R library. The database used was GO_Biological_Process_2021. **Figure S7** shows an example of pathway enrichment from trajectory 1 and 10. The full table of pathways is provided in **Supplemental Data 8**.

Degree of sharing across EN, IN and Glial subclasses

To identify genes with similarity in effect sizes across subclasses, the association statistics from *dreamlet* was not sufficient. For example, if the age-associated effect size for a gene was significant in one subclass but not in another, one might have concluded that the age effect was subclass-specific. However, the absence of a significant effect in another subclass did not necessarily mean that the effect size was zero. This scenario often occurs when statistical power is limited, or varies significantly between subclasses. To overcome this limitation, we utilized *mashr* [42] (v 0.2.79) R library, which used an empirical Bayes approach to learn patterns of similarity in effect sizes across subclasses and then leveraged these prior patterns to improve the accuracy of effect size estimates.

We assess gene sharing across 9 EN, 7 IN and 4 glial subclasses. The EN group included 9 subclasses (EN_L6_CT, EN_L5_6_NP, EN_L6B, EN_L3_5_IT_1, EN_L3_5_IT_2, EN_L3_5_IT_3, EN_L2_3_IT, EN_L6_IT_1, EN_L6_IT_2), the IN group included 7 subclasses (IN_LAMP5_RELN, IN_LAMP5_LHX6, IN_ADARB2, IN_VIP, IN_PVALB_CHC, IN_PVALB, IN_SST), and the glial group included 4 subclasses (Astro, Oligo, OPC, Micro). From hereafter, we refer to EN, IN and Glial subclasses sets as c_E, c_I *and* c_G respectively.

In this section, we conducted two analyses: 1) we quantified the degree of sharing for each gene based on the similarity of age-associated effect size patterns across EN, IN, and glial subclasses using `run_mash` function built in *dreamlet* adapted from *mashr* (v0.2.79)[42] R library, and 2) we calculated the tau score to quantify the cell specificity of each gene, allowing us to stratify shared and non-shared genes.

<u>*Mashr analysis*</u>*:* Using the age-associated summary statistics per age group from **Supplemental Data 4**, we built two matrices with genes as rows and subclasses as columns, containing $\log_2$FC ($\beta_{j,i}$) and standard errors ($SE_{j,i}$). The number of genes included are: developmental (*n* = 23,914 genes), young adulthood (*n* = 26,107 genes), middle adulthood (*n* = 26,809 genes), and late adulthood (*n* = 24,795 genes) time points. Any gene without summary statistics for a subclass was replaced with 0. Using the empirical bayesian framework, *mashr* computed posterior mean and variance by combining covariance matrix derived from prior distribution of effect sizes and matrix of observed variance to assess sharing of j gene across subclasses sets c_E, c_I *and* c_G. The posterior mean $\hat{\beta_{j,i}}$ provides a more accurate estimate of the true effect size $\beta_{j,i}$, while the posterior variance quantifies the uncertainty in $\hat{\beta_{j,i}}$. After this step, *mashr* reported the local false positive rate, which is defined as the probability that the true effect size $\beta_{j,i}$ had the opposite sign of the estimated effect size $\hat{\beta_{j,i}}$. For example, for j gene within c_E set, $lfsr_{j,i \in c_E} = min\left[p(\beta_{j,i} \geq 0|\hat{\beta_{j,i}}), p(\beta_{j,i} \leq 0|\hat{\beta_{j,ij}})\right]$ where $\beta_{j,i}$ is true age effect size and $\hat{\beta_{j,i}}$ is the estimated effect size. Then $p_{j,i} = 1 - \mathrm{lfsr}_{j,i}$ is the probability that the true age effect size and estimated effect size are concordant. Thus, for j gene, the degree of sharing or composite posterior probability for concordance in effect sizes of EN, IN, and glial classes for each age group can be quantified as the product of probabilities across their

respective subclasses, i.e., $P_{j,E} = \prod_{i \in c_E} p_{j,i},\ P_{j,I} = \prod_{i \in c_I} p_{j,i},\ P_{j,G} = \prod_{i \in c_G} p_{j,i}$. To ensure that each gene has true effect size $\beta_{j,i}$ with p-value < 0.05 from *dreamlet* regression as well, we applied an additional filter: a dreamlet p-value of the gene < 0.05 in at least two subclasses in a grouped class. **Figure S10a** shows the number of genes obtained after this step. Composite posterior probability P_E, P_I, P_G were stratified into ten equally sized bins to display the number of genes in each bin. Bin 1 represented genes with a composite prob. < 0.1, indicating less sharing, while bin 10 included genes with a composite prob. > 0.9, indicating high sharing (**Fig. S11a** and **Fig 3b). Fig. S11b** demonstrates an example of the distribution of composite posterior probability from EN P_E from 24,795 genes during late adulthood. The heatmap of true age effect sizes $\beta_{j,i}$ shows a highly concordant pattern across 9 subclasses of ENs with a composite probability greater than 0.9. In contrast, the effect sizes are heterogeneous across the nine EN subclasses for genes with a probability less than 0.01 (**Fig. S11b).** Using the threshold $P_{j,E}, P_{j,I}, P_{j,G} > 0.9$, we obtained the shared genes for each grouped class across four age groups (**Fig. S10a)**. Interestingly, no significant overlap was observed in bin 10 across the age groups for each subclass (**Fig. S10b-d).** The output from this step is provided as P_E, P_I, P_G for each gene per age group in **Supplemental Data 11.**

Using these shared genes per age group per grouped class we identified a network of genes which encode significant protein-protein interactions (PPI) using *STRING-db* [48] (v 12.0). **Fig. S14a-c** show the significant PPI with highest confidence interaction score > 0.9 and genes in networks of at least 5 genes. The clusters of each PPI per grouped class per age group are given in **Supplemental Data 13.**

Cells specificity score: Next we reasoned that shared genes with $P_{j,E}, P_{j,I}, P_{j,G} > 0.9$ have significantly lower cell specificity compared to genes < 0.9. To validate this we estimated tau score adapted from GTEX [43] studies of each gene for each grouped class per age group. Tau score indicates how specifically a gene is expressed across various subclasses. In other words, genes with a tau score close to 1 are more specifically expressed in one subclass, while those with a tau score closer to 0 are equally expressed across all subclasses within a grouped class. To do this, we integrated pseudobulk gene expression data from a grouped class per age group using the `stackedAssays` function, resulting in expression of j genes and i subclass * k donors specifically 17,205 X 2,393; 17,205 X 1,945 and 17,205 X 1,129 for EN; IN and glia respectively. The analysis was limited to protein coding genes. Next, we estimated the median expression of donors across each subclass using the equations below:

$$expression_{j,i} = median(expression_{i,j,1}, expression_{i,j,2} ... expression_{i,j,k})$$

$$expression_{j,i} = log_2(expression_{j,i})$$

$$max_expression_j = median(expression_{j,1}, expression_{j,2} ... expression_{j,g})$$

Finally tau score was estimated as $Tau\ Score_j = \frac{N - \sum_{i=1}^{g} expression_{j,i}/max_expression_j}{N-1}$, where N is the total number of subclasses. Tau score was kept for cell specificity measurement and was retained for only those genes with sum of median expression value across all subclasses > 10 cpm. **Fig. S12c** and **Fig. S13** shows the distribution of tau scores stratified by shared and non-shared genes for each grouped across and each age group for a list of genes from *mashr* analysis and aDEGs. The tau score of each age group across 3 grouped classes are provided in **Supplemental Table 4**. The gene set pathways of shared genes, shown in **Fig. 3e**, were obtained using the *enrichr* function from the *enrichR* (v 3.2) R library and the full table is provided in **Supplemental Data 12**. The analysis utilized the GO_Biological_Process_2021 database and was limited to protein-coding genes.

snRNA-seq dataset integration for dimension reduction and clustering

Filtered count matrices from our in-house lifespan dataset and published snRNA-seq data from human DLPFC spanning from gestation to adulthood[12] were combined into a *Scanpy* (v1.9.3) object for processing and filtering. Observation columns of concatenated snRNA-seq datasets were organized by classifying cell types, developmental stages, and other metadata, standardizing formats, and to prepare the data for further analysis and visualization. Genes not observed in ≥ 5 nuclei across all batches were removed, and potential nuclei doublets were identified and removed using *Scrublet* (v0.2.3) with 10 principal components for each batch individually.

To remove sampling bias from differing sequencing depths between batches, the integrated data were downsampled to 1,000 Unique Molecular Identifier counts (UMIs) per nucleus. Downsampling was performed by randomly sampling, without replacement, 1,000 UMIs from the total UMIs for a given nucleus, and nuclei with less than 1,000 total UMIs were excluded. The count matrix was then scaled to Counts Per Million (CPM) post-downsampling and transformed using natural log plus one. Six thousand HVGs were selected using the `highly_variable_genes` function in *Scanpy*, while excluding mitochondria and sex-specific genes. Data dimensions were reduced via principal component analysis (PCA) to components explaining 50% of the variance. A neighborhood graph was constructed using *Scanpy*'s preprocessing neighbors function with 100 neighbors on the reduced components, and a 2-dimensional Uniform Manifold Approximation and Projection (UMAP) embedding was generated. Umap of MATuration (UMAT) embedding was generated as previously described[12].

Pseudotime analysis and dynamically expressed gene identification and downstream analysis

Trajectory reconstruction and identification of differentially expressed genes along trajectories (traDEGs) were performed as previously described[9]. For each lineage, corresponding cells were selected and genes not observed in ≥ 5 nuclei were excluded. Six thousand HVGs were selected, data dimensions were reduced via PCA to components explaining 50% of the variance, and the UMAT embedding was recalculated. Pseudotime trajectory analysis was then conducted using *Monocle3* (v1.0.0) based on the UMAT embedding. The shortest path between the developmental node and the node in the mature subclass/subtype clusters was isolated as the corresponding trajectory graph. Cells along the trajectory were selected, and traDEGs were identified using *Monocle3*'s modified `graph_test` function with Moran's *I* test, including covariates (Sex, Batch, PMI, and log_n_genes) to ensure results were not affected by uneven contributions from different subjects and conditions. Genes with an adjusted p-value < 0.05 and Moran's $I \geq 0.05$ were considered statistically significant DEGs. To cluster the DEGs in each lineage, single-cell expression data along each trajectory were compressed using a sliding window along pseudotime, averaging the expression of neighboring cells to generate 500 meta-cells per trajectory. Each gene's expression was then modeled using a generalized linear model $(expression \sim splines::ns(pseudotime, df = 3))$, and k-means clustering was performed on the fitted expressions.

GO-term analysis of traDEG clusters was performed using *enrichR* (v3.2) and the GO Biological Process 2023 dataset. To investigate whether the traDEG clusters play a role in neurological, psychiatric, and other traits, we quantified their colocalization with common risk variants from 24 GWAS (**Supplementary Table 2**) using MAGMA analysis.

Spatial validation of astrocyte lineages

Visium experiments were performed according to the 10x Genomics user guide.

Tissue freezing and cryosectioning: DLPFC tissues were prepared according to 10x Genomics Visium Spatial Protocols (Tissue Preparation Guide, CG000240 Rev B). Tissue blocks were snap-frozen by submerging in an isopentane bath (320404-1L, Sigma-Aldrich) chilled with liquid nitrogen. Once frozen, tissue blocks were embedded with chilled OCT (Tissue-Tek O.C.T. Compound, 4583, Sakura Finetek USA), and stored at -80 ˚C until further use. Before cryosectioning, tissue blocks were equilibrated to the temperature of the cryostat chamber (HM505, Microm). After trimming, regions of interest (ROI) were identified. To fit in the 6 mm x 6 mm Visium capture area, the ROIs containing intact gray matter and white matter were scored by cutting the tissue surface with a razor blade. High quality 10 µm sections, encompassing intact ROIs, were flattened and placed on the capture areas of the pre-equilibrated Visium slides (Visium Spatial Gene Expression Slide, 2000233, 10x Genomics). To firmly adhere the sections to the Visium slides, the sections were melted by placing a finger on the backside of the slide for a few seconds, and refrozen in the cryostat chamber. Visium slides were sealed in 50 ml conical tubes and stored at -80 ˚C until further processing.

Sample preparation: Visium sample preparation was performed according to the 10x Genomics protocols "*Methanol Fixation, H&E Staining & Imaging for Visium Spatial Protocols CG000160 Rev C"* and *"Visium Spatial Gene Expression Reagent Kits User Guide CG000239 Rev E"*. Briefly, Visium slides prepared in the previous step were retrieved from -80 ˚C storage and placed on a 37 ˚C thermocycler adapter for 1 minute to dry. The slides were then fixed in pre-chilled methanol for 30 minutes at -20 ˚C. After fixation, isopropanol was applied to the tissue sections and incubated for 1 minute at room temperature and, following removal of isopropanol, the tissue sections were air dried for 5 minutes. For H&E staining, hematoxylin (Hematoxylin, Mayer's (Lillie's Modification), Agilent, S330930-2) was applied to cover the tissue sections and incubated for 7 minutes. After incubation, the hematoxylin was discarded and the slides were rinsed with water. To enhance staining, bluing buffers (Bluing Buffer, Dako, Agilent, CS70230-2) were added and incubated for 2 minutes, bluing buffer was then discarded and the Visium slides were rinsed. The sections were subsequently incubated with eosin (Eosin, Dako, Agilent, CS70130-2) mixed for 1 minute, rinsed, and dried for 5 minutes at 37 ˚C. The slides were then coverslipped and scanned.

Following imaging, coverslips were removed by immersing the slides in 3X SSC buffer. The Visium Slide Cassettes were then assembled on the Visium slides. Tissue sections were permeabilized with the permeabilization enzyme (Visium Spatial Gene Expression Reagent Kit, PN-1000189, 10x Genomics) at 37 ˚C for 24 minutes, based on tissue optimization results from adjacent brain sections, followed by incubation with 0.1X SSC. Reverse transcription was initiated by incubating samples with RT master mix (Visium Spatial Gene Expression Reagent Kit, PN-1000189, 10x Genomics) at 53 ˚C for 45 minutes. Samples were then treated with 0.08M KOH, and single strand cDNA converted to double strand DNA by incubation with second strand mix (Visium Spatial Gene Expression Reagent Kit, PN-1000189, 10x Genomics) at 65 ˚C for 15 minutes. Samples were then washed with buffer EB, and 0.08M KOH was added to denature the new synthesis double strand DNA. Samples were neutralized with 1M Tris pH7.0. To amplify the cDNA, cDNA amplification mix (Visium Spatial Gene Expression Reagent Kit, PN-1000189, 10x Genomics) was added, and PCR was performed with 16 cycles, as determined by qPCR. PCR products were purified using SPRIselect reagent (Beckman Coulter SPRIselect Reagent Kit B23318) and quantified with an Agilent TapeStation.

Library Construction: To construct the spatial gene expression library, purified cDNA samples were fragmented, end-repaired and A-tailed using the fragmentation mix (Library Construction Kit, PN-1000196, 10x Genomics). Following fragmentation, samples were purified and size-selected using SPRIselect reagent. Adapters for adding the sequencing indexes (Adaptor Ligation Mix, Library Construction Kit, PN-1000196,10x Genomics) were ligated to the cDNA, followed by post-ligation cleanup using SPRIselect Reagent. Sequencing indexes were added to the cDNA through 15 cycles of PCR using the amp mix (Library Construction Kit, PN-1000196,10x Genomics) and indexes (Dual Index Plate TT Set A, PN-1000215, 10x Genomics). Following the PCR, the indexed PCR products were size-selected using SPRIselect reagent, and post-library construction QC was performed using an Agilent Tapestation with 1:10 diluted samples.

Data processing: Filtered matrices for each sample were imported into *Scanpy*, excluding spots with > 30% mitochondrial reads and focusing exclusively on protein-coding genes located in euchromatin regions. Normalized log-transformed counts per sample were calculated, and highly variable genes were selected for clustering using *Scanpy*'s default parameters. To obtain spatial domains, a spatial nearest neighbor graph was created using *Squidpy* (v1.4.1)[107] and combined with the transcriptomic nearest neighbor graph. This combined graph was used for Leiden clustering, resulting in the primary annotated clusters. The samples were then merged, and the cluster information was input to *ONTraC* (v0.0.7)[68] for final cluster annotation.

Enrichment of brain and non-brain related risk genes in age and pseudotime associated genes.

All analyses to evaluate the enrichment of brain and non-brain related traits were conducted using MAGMA version 1.08b[105]. MAGMA calculates gene-level P-values for each gene and trait by assessing the joint association of all SNPs within the gene region, while accounting for linkage disequilibrium (LD) between SNPs. Gene regions were defined with a window of 35 kb upstream and 10 kb downstream, and LD estimates were derived from the European panel of the 1000 Genomes Project[108] (phase 3). MAGMA then applies a linear regression framework to determine whether differentially expressed genes show stronger associations with GWAS traits compared to the rest of the genome. Genes overlapping the MHC region (chr6:25-35 MB) were excluded from the analysis. Heatmaps of brain- and non-brain-related traits were generated using the MAGMA pipeline as described. The outputs are provided in the following Supplemental Data: Supplemental Data 6 includes results from Fig. S4b and Extended Data Fig. 3d, Supplemental Data 9 includes Fig. 2e-f and Fig. S6c, and Supplemental Data 16 includes Fig. 4g and Fig. 5f, 5l.

Rhythmicity Analysis

Prior to rhythmicity analysis, time of death (TOD) for each subject was normalized to a Zeitgeber Time (ZT) scale as described by Seney *et al.*[109]. Briefly, the TOD for each subject was collected at local time then converted to coordinated universal time by adjusting for time zone and daylight savings time. Coordinated universal time was further adjusted to account for longitude and latitude of death place. Each subject's TOD was then set as ZT = t hours (h) after previous (if $t < 18$) or before next (if $t \geq -6$) sunrise. The distribution of subject TODs across age groups is shown in (**Extended Data Fig. 6a**). The neonatal, childhood and adolescent groups had too few subjects with known TODs to run further rhythmicity analyses, and the young adulthood group had large gaps in their distribution. As such, we excluded the neonatal, childhood, and adolescent groups and combined the young

and middle adulthood groups (YA+MA; n = 116), which was then compared to the late adulthood group (LA; n = 76) (**Fig. 6a**).

To measure the rhythmicity in i subclass and j genes and for other downstream analyses in this section, we first regressed out the effects of identified technical and biological covariates in the previous section along with age using the eq(9) described below.

$$\hat{expression}_{i,j} = Age + covariates$$
$$residuals_{i,j} = expression_{i,j} - \hat{expression}_{i,j} \quad (9)$$

Then, using the DiffCircaPipeline workflow[110], 24 h rhythms in gene expression were detected and compared between YA+MA and LA groups. First, samples were ordered by TOD and residualized expression for each transcript was fit to a cosinor model separately in each group and categorized by type of rhythmicity (rhythmic in YA+MA, rhythmic in LA, both, or arrhythmic) (**Supplementary Data 17**). These categories were then used to determine which genes further analyses were performed in to reduce the likelihood of Type I error. As such, differences in goodness of fit (ΔR^2) between YA+MA and LA were determined through a permutation test (1000 permutations) in transcripts identified as rhythmic in YA+MA, LA, or Both (**Supplementary Data 19**). Additionally, a global differential parameter test was performed in transcripts identified as rhythmic in Both groups, followed by *post hoc* analyses to determine whether differences were due to amplitude, MESOR, or phase (**Supplementary Data 18**). enrichR was then used to identify pathways enriched in rhythmic genes (**Supplementary Data 20**). This was done for all rhythmic genes within a subclass in YA+MA and LA, and then separately for rhythmic genes that peaked between 0-8 h ZT and 20-24 h ZT (the "Sunrise" group) and 8 - 20 h ZT (the "Sunset" group).

Data availability

The DLPFC lifespan snRNA-seq profiling data can be accessed via Synapse, as part of the PsychAD Study. The dataset, analysis outputs are available via the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.synapse.org/Data%20Access). The data are available under controlled use conditions set by human privacy regulations. To access the data, a data use agreement is needed. The registration is in place solely to ensure the anonymity of the study participants. In addition, we have a data descriptor manuscript detailing the data processing and data collection.

Code availability

Code deposit at Zenodo is available for crumblr analysis: https://doi.org/10.5281/zenodo.12752107

Acknowledgments

We extend our deep gratitude to the patients and their families for their generous donation of invaluable biological material, which was essential for the success of this study. Their unwavering participation and dedication to advancing scientific knowledge and enhancing human health are deeply appreciated. We also acknowledge the generous support of the National Institute on Aging, who provided funding for this research through the following NIH grants: R01AG067025, R01AG082185, and R01AG065582. Human tissues were obtained from the NIH NeuroBioBank at the Mount Sinai Brain Bank (MSSM; supported by NIMH-75N95019C00049), and NIMH-IRP Human Brain Collection Core (HBCC, project # ZIC MH002903). The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.

Author contributions

Conceptualization: PR, KG
Methodology & Software: GEH, KG, HY, MRS
Validation: HY, XW
Formal analysis: KG, HY, MRS, TC
Investigation: JFF, XW, MP, SV, AH, CC, ZS, MA, SA
Resources: VH, PKA, SM
Data Curation: JB, PNM, SV, MP, DL, HY
Writing: HY, KG, MRS, TC, PR, JFF, XW, DL with support from all co-authors.
Visualization: KG, HY, MRS, TC
Supervision: PR, KG, JFF, CAM, GV, JB, VH, DL, GEH
Project administration: KG, PR
Funding acquisition: PR, VH
All authors read and approved the final draft of the paper.

Competing interests

The authors declare no competing interests.

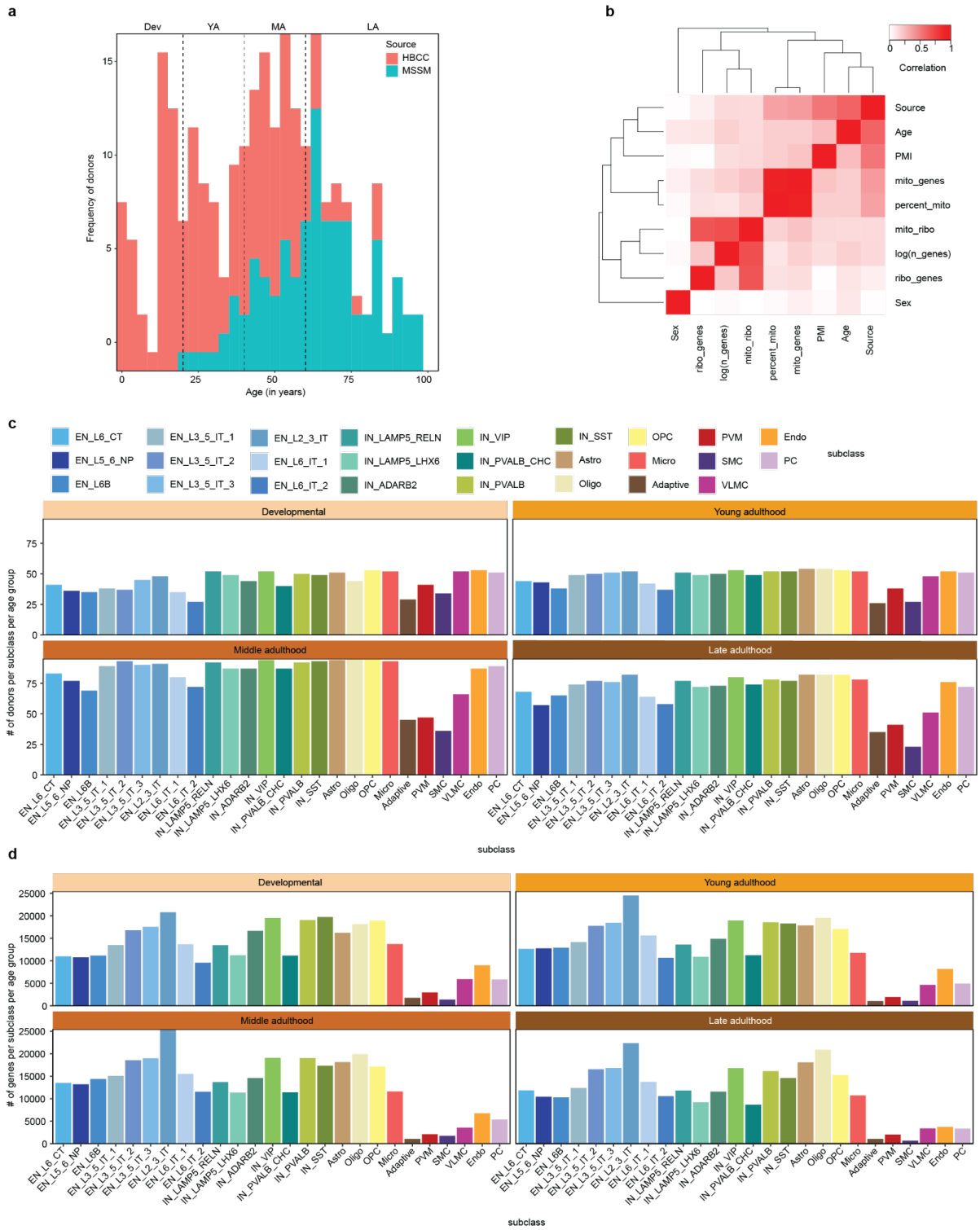
Materials & Correspondence

Correspondence to Kiran Girdhar or Panos Roussos.

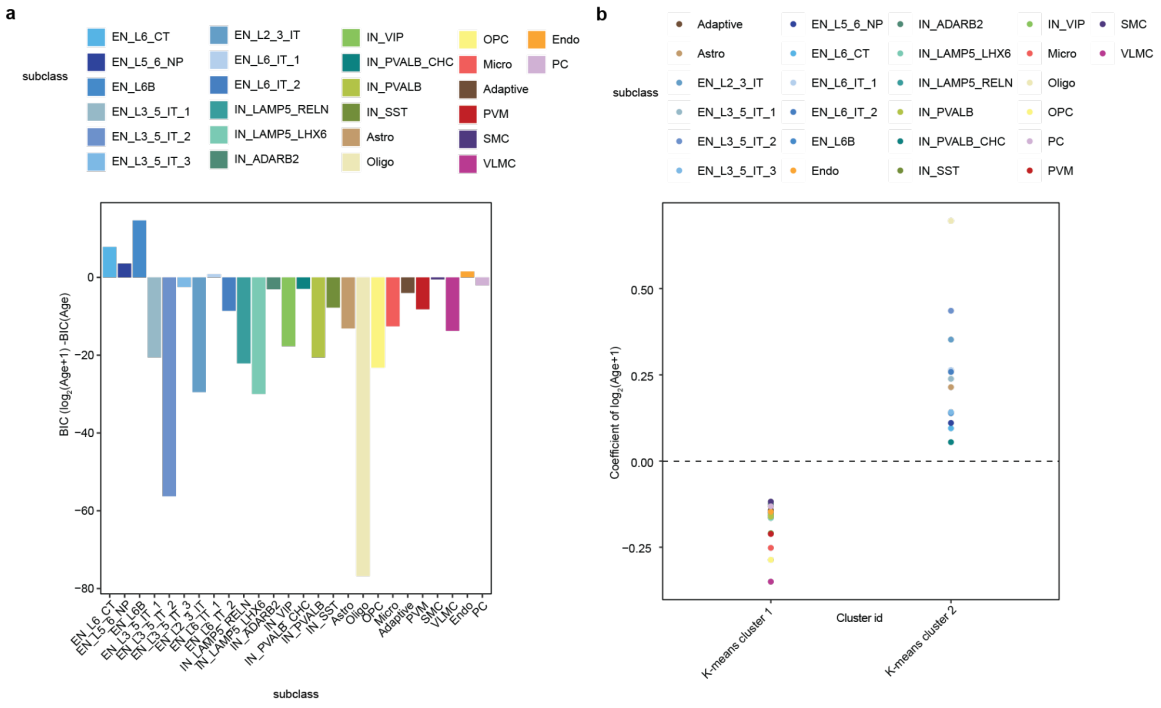
Supplementary Information

This supplementary information includes:

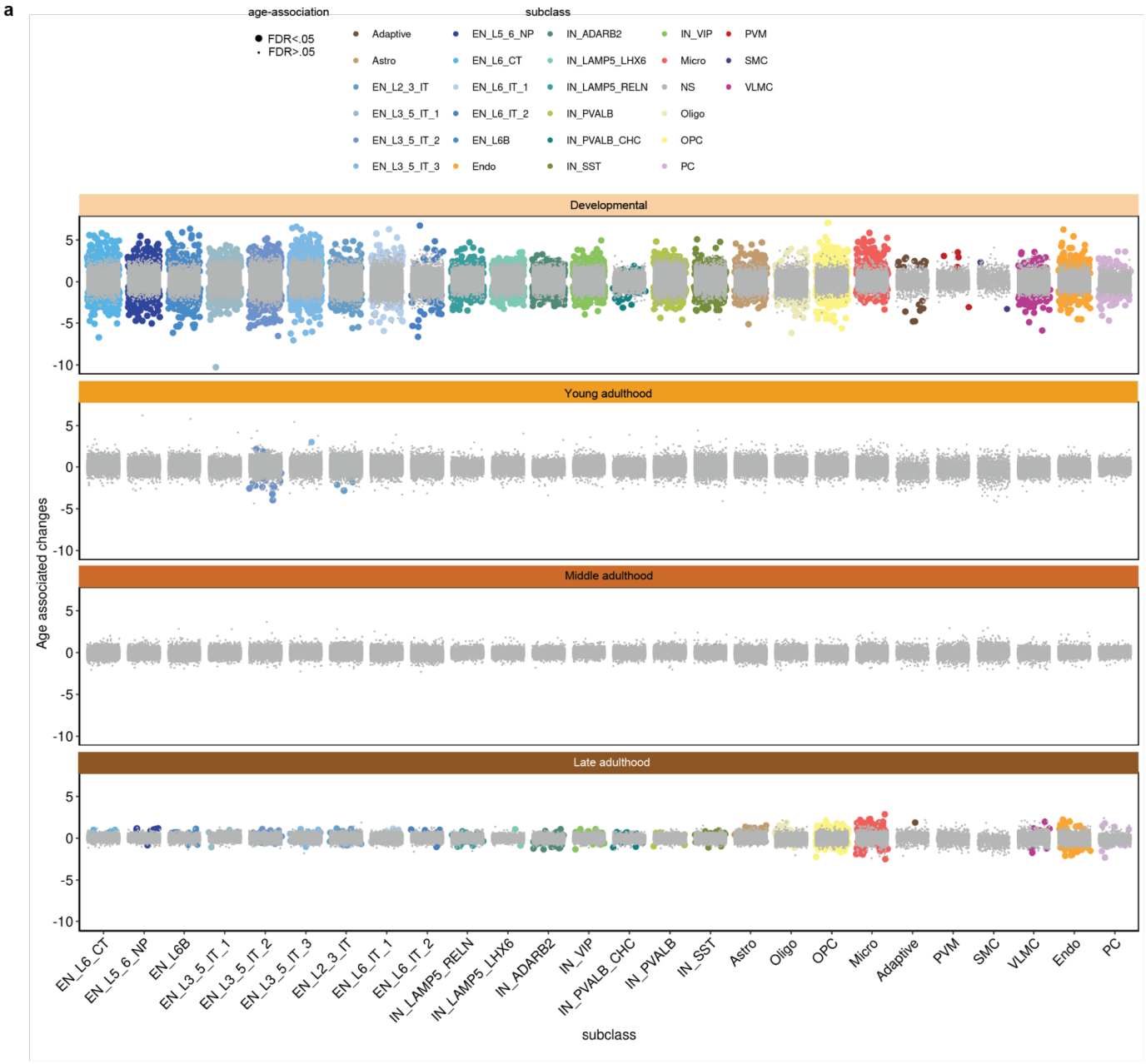
- Supplementary Figs. 1-18
- Captions for Supplementary Tables 1-4
- Captions for Supplementary Data 1-20
- Supplementary Notes



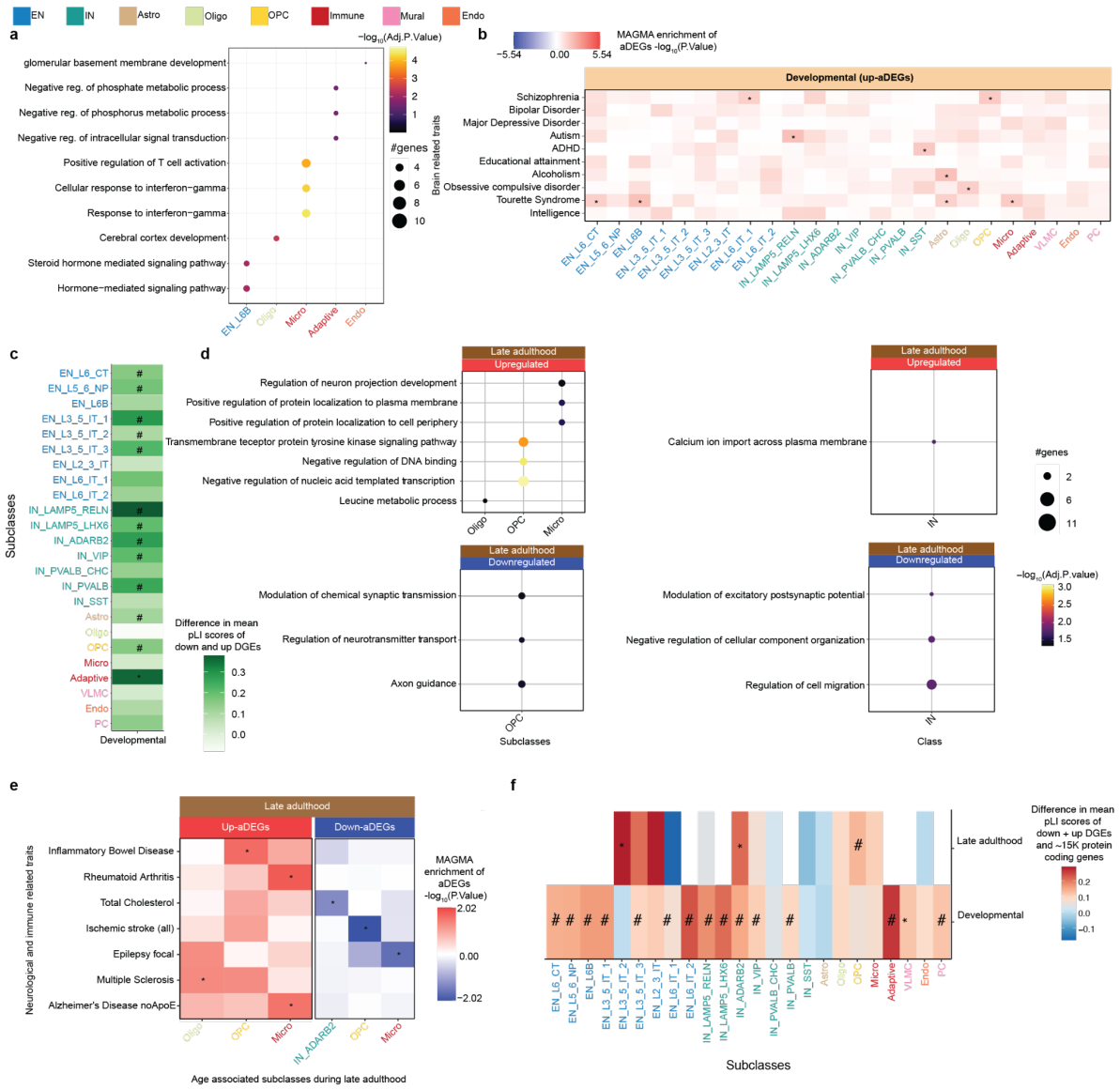
Supplementary Figure 1. Age distribution of lifespan data and global summary of pseudobulk expression. a, Histogram of age of 284 donors colored by brain bank. Dev, YA, MA and LA are abbreviations for age groups developmental, young adulthood, middle adulthood and late adulthood respectively. Dashed line demarcates the age range across four groups. **b,** Heatmap of pairwise correlation of identified covariates and age. **c-d,** Barplot to show no. of donors and genes across each subclass stratified by four age groups. These counts were obtained after applying the `processAssays` function to pseudobulk expression, which enforces the following criteria: each sample must have at least 5 nuclei, each gene must have at least 5 reads, and subclass must include at least 4 samples (see methods).



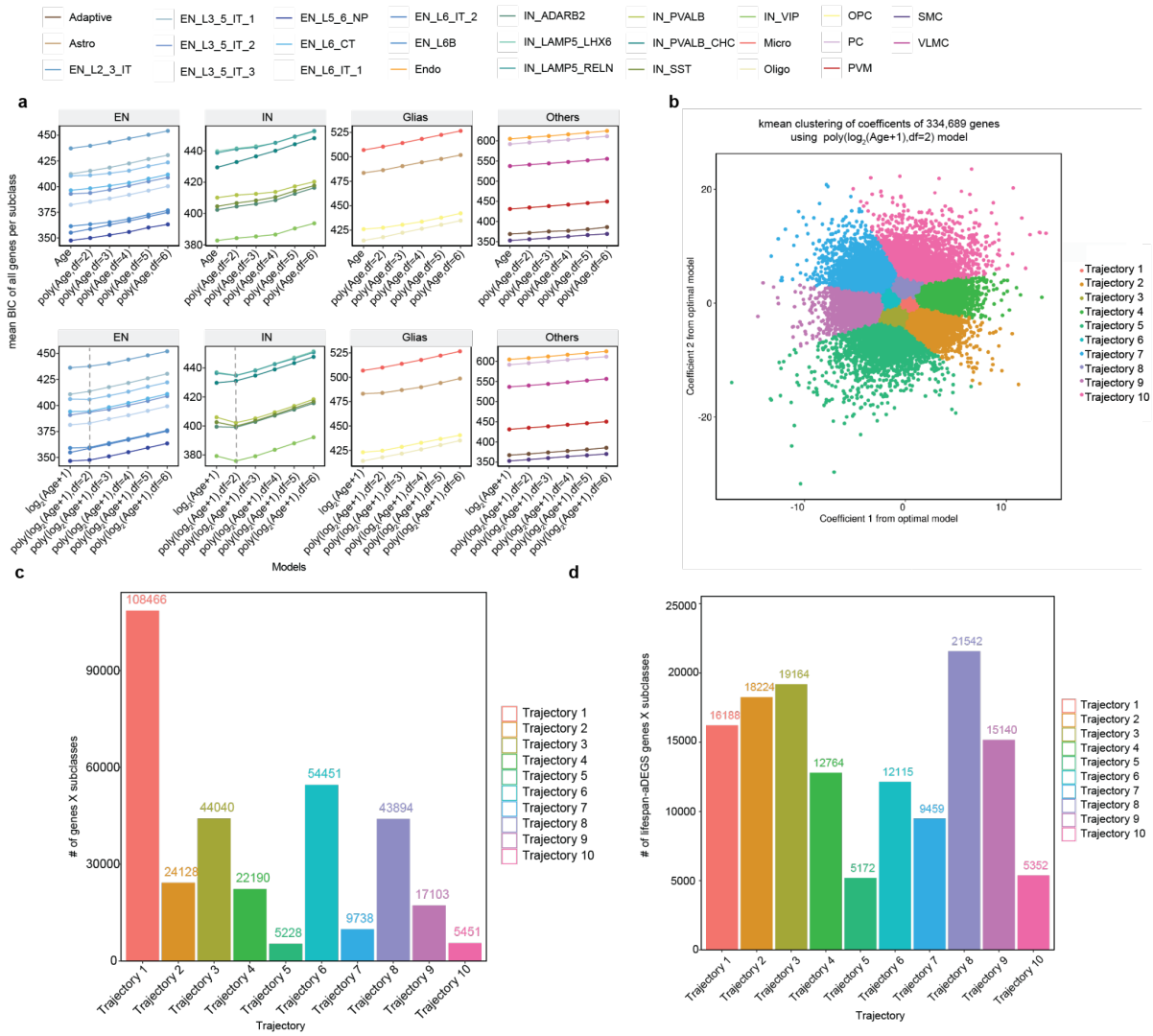
Supplementary Figure 2. Optimal model selection for nuclei counts. a, Bayesian Information Criterion (BIC) obtained from modeling nuclei counts using the formulas: nuclei counts ~ $\log_2$(Age+1) and nuclei counts ~ Age, utilizing *crumblr* for each subclass. **b,** K-means clustering coefficient from nuclei counts ~ $\log_2$(Age+1).



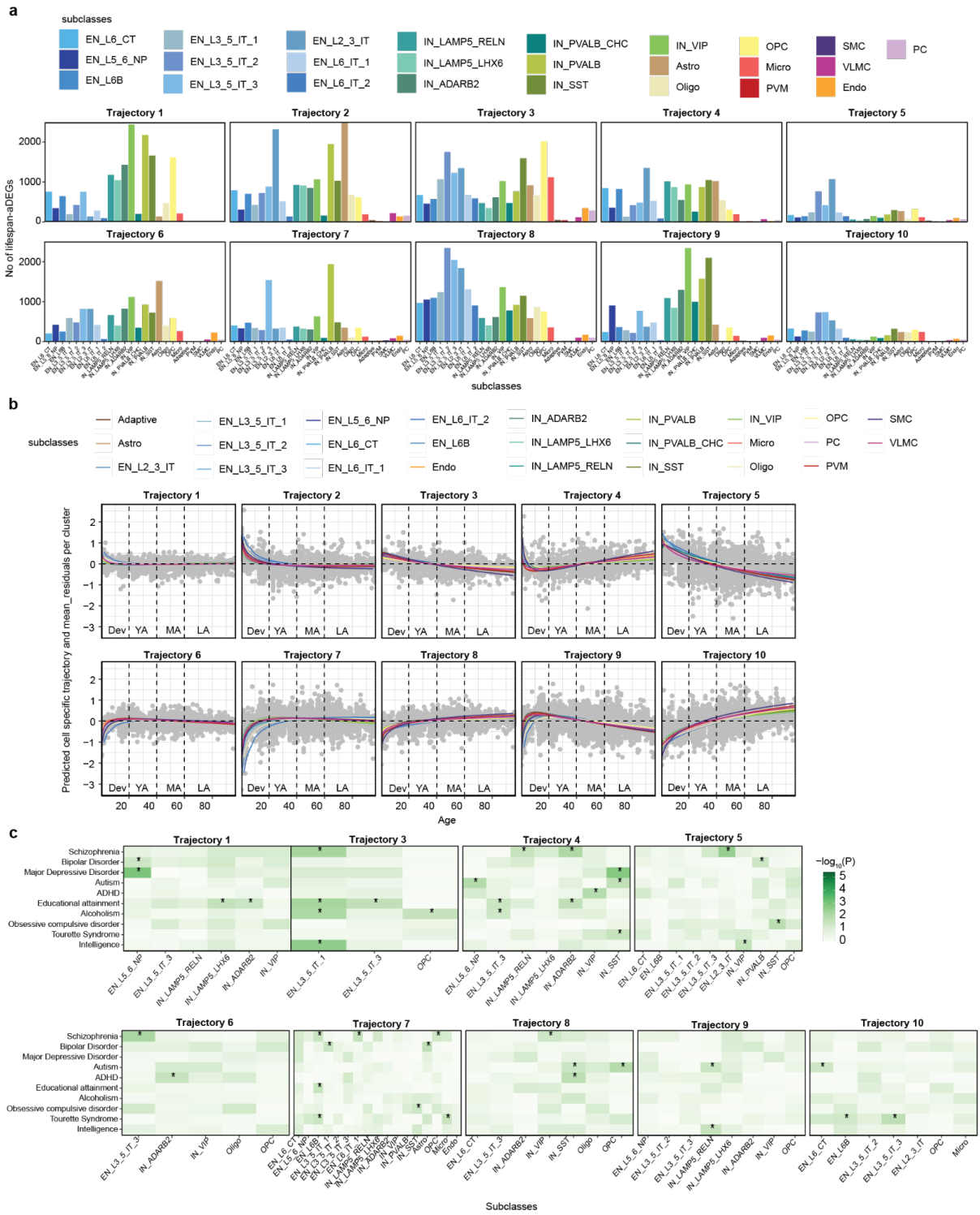
Supplementary Figure 3. Age associated effect sizes. a, Beeswarm plot of age-associated effect sizes for 26 subclasses across four age groups. Gray dots are genes with FDR < 0.05 and colored dots are significant genes with FDR < 0.05.



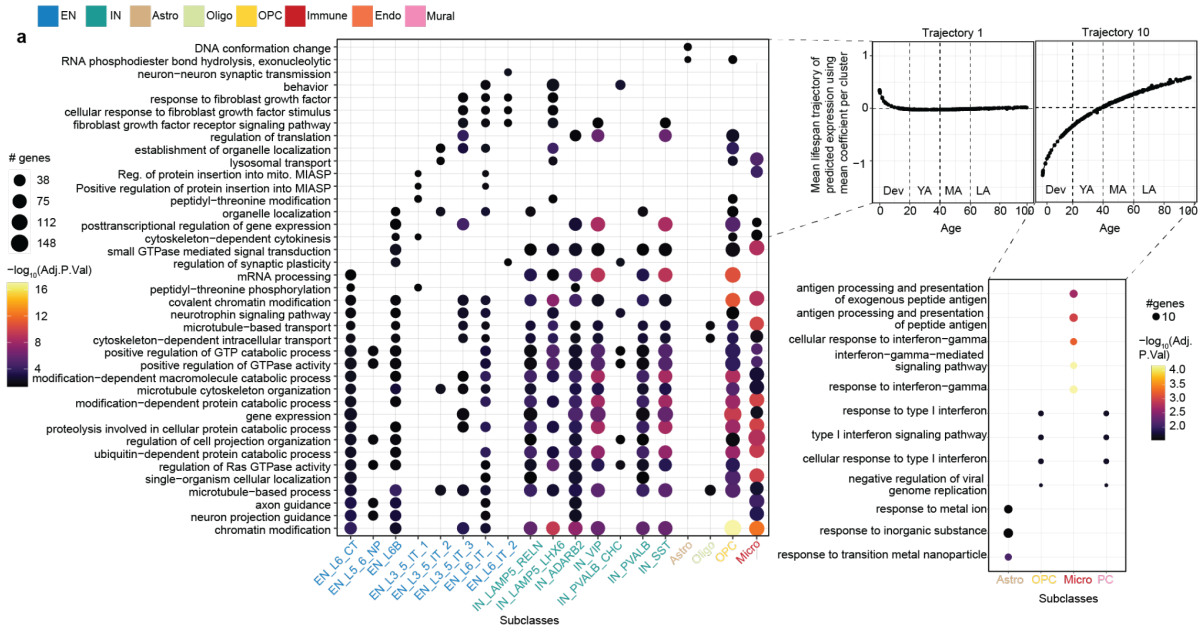
Supplementary Figure 4. Age-associated gene expression across age groups. a, Functional Pathway analysis of up-aDEGs during development for subclasses that were significantly enriched for GO biological processes with an adjusted p-value < 0.05. **b,** Association of up-aDEGs with risk genes for brain-related traits using MAGMA. * and # indicate nominal p-value < 0.05 and FDR < 0.05 from MAGMA enrichment. **c,** Heatmap to show differences in loss-of-function mutation scores (pLI) of down and up-aDEGs during development: * and # indicate nominal p-value < 0.05 and FDR < 0.05 from Wilcoxon test. **d,** 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. **e,** 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. **f,** 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.



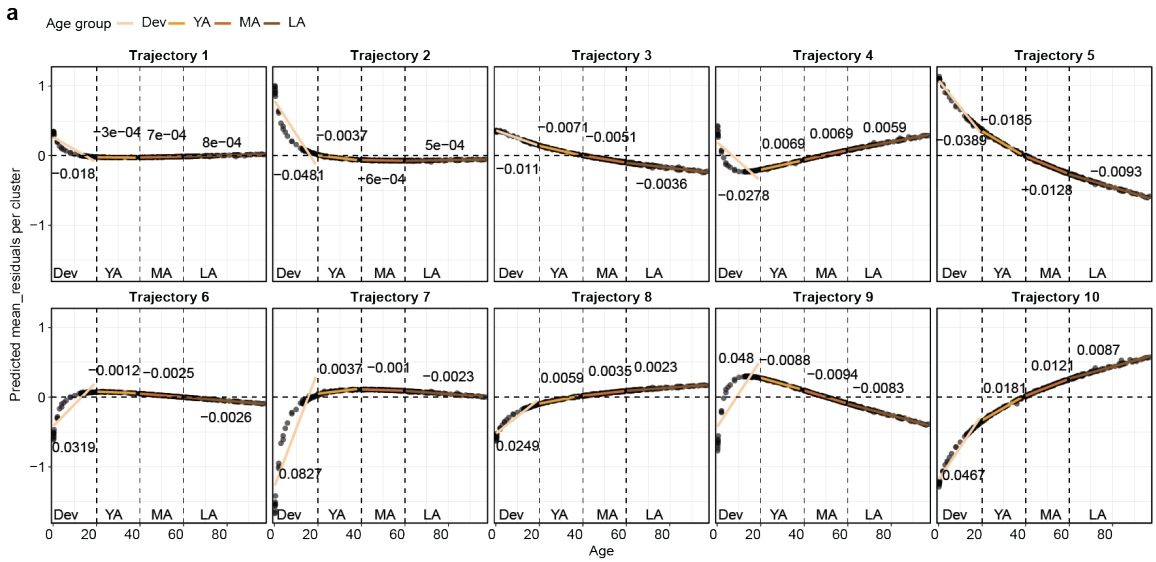
Supplementary Figure 5. Optimal model selection and clustering of gene expression trajectories. a, Mean BIC of linear and non-linear models fitting on all genes. X-axis displays all models implemented in the pipeline. The dashed line shows the optimal model used for final clustering of trajectories.**b,** 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.



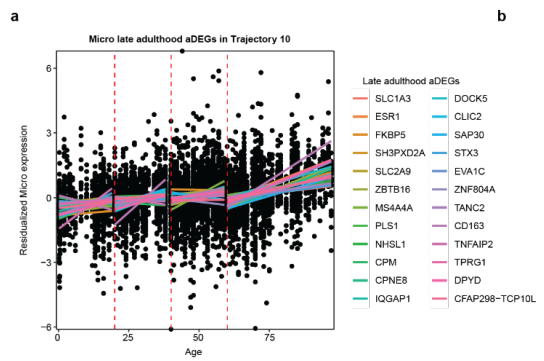
Supplementary Figure 6. Subclass specificity of trajectories. a, Distribution of lifespan-aDEGs across 27 subclasses stratified by ten clusters. **b,** Plot of covariates corrected gene expression as a function age shown as gray points and fitted polynomial using the mean of coefficients from each subclass. The polynomial fit is shown in subclass specific colours. **c,** Association of subclass specific developmental 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. **e,** Functional pathway analysis of late adulthood aDEGs for subclasses that were significantly enriched for GO biological processes with an adjusted p-value < 0.05.



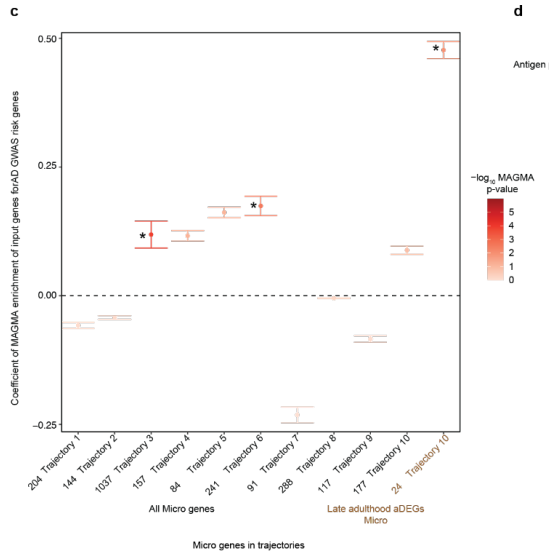
Supplementary Figure 7. Biological mechanism exhibited by trajectories. a, Functional pathway analysis of subclass-specific genes in trajectory 1 (left) and trajectory 10 (right), highlighting subclasses significantly enriched for GO biological processes with an adjusted p-value < 0.05. **b,** Functional pathway analysis of late-adulthood aDEGs for each subclass and trajectory that were significantly enriched for GO biological processes with an adjusted p-value < 0.05.

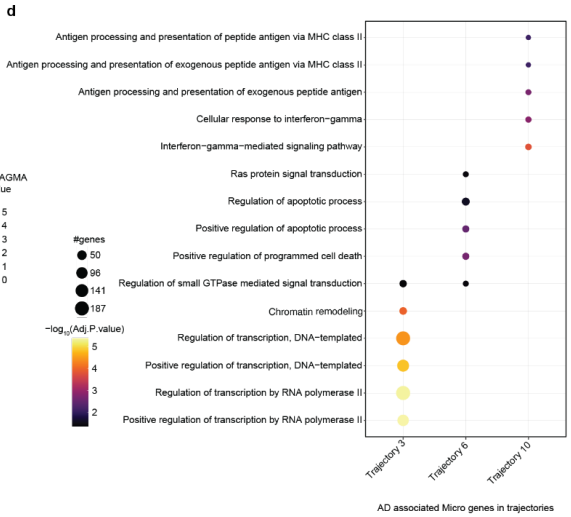


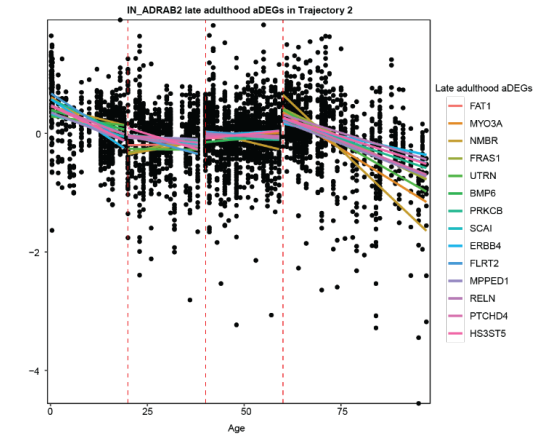
Supplementary Figure 8. Magnitude of aging of average lifespan trajectories for each age group. a, Plot of predicted trajectory using the mean of coefficients of all genes within each trajectory. The linear fit across age groups shows the regression line from average trajectory to age from each age group**.** 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
steroid hormone mediated signaling pathway
multinuclear osteoclast differentiation
neuron projection organization
positive regulation of chondrocyte differentiation
actin filament network formation
purine nucleobase metabolic process
neurotransmitter receptor transport to postsynaptic membrane
fructose transmembrane transport
synaptic vesicle fusion to presynaptic active zone membrane
negative regulation of miRNA transcription
cellular response to metal ion
negative regulation of stem cell population maintenance
myoblast fusion
chloride transport
regulation of phospholipase C activity
superoxide anion generation
mesonephros development
dense core granule cytoskeletal transport
regulation of ossification

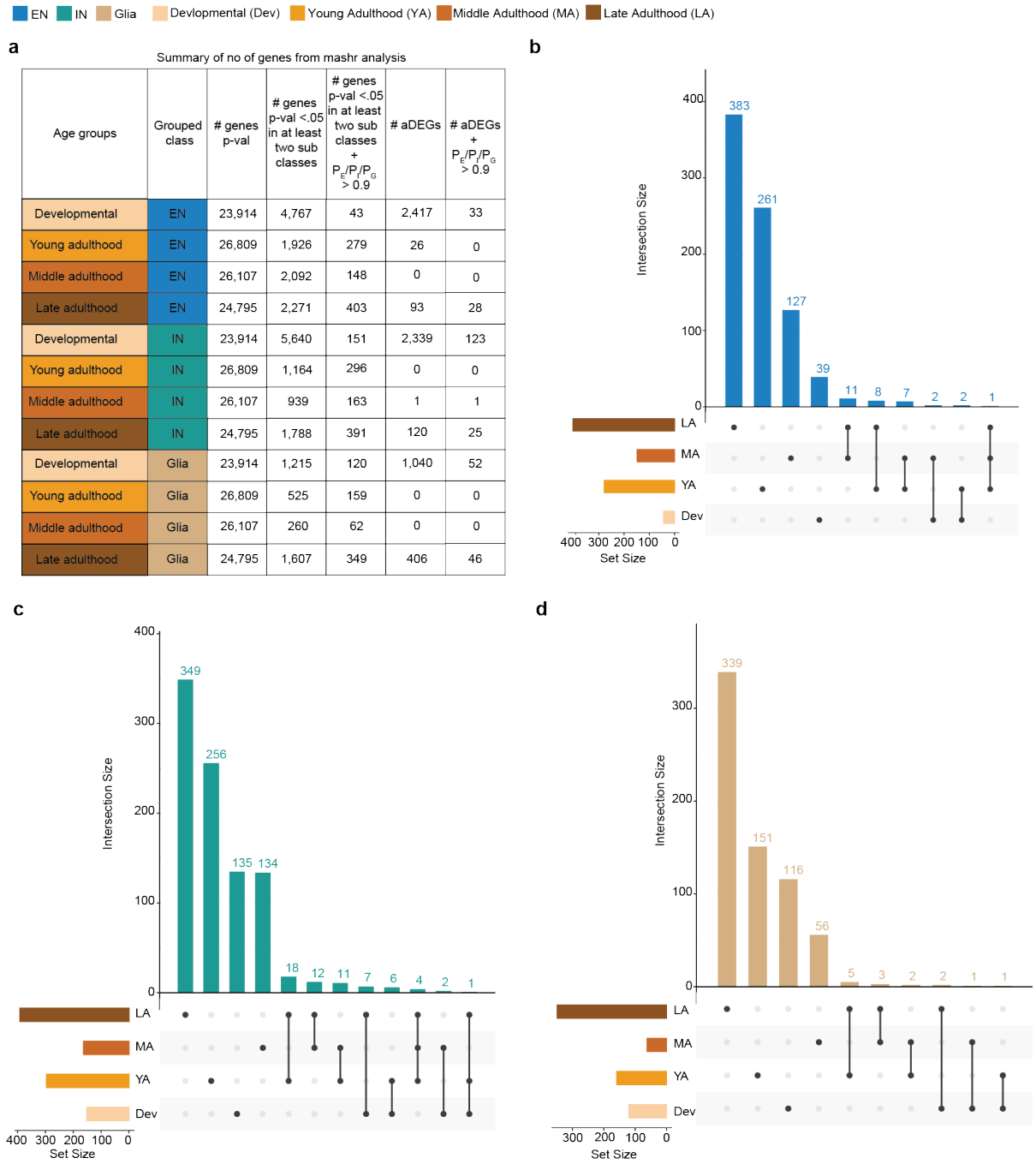

c
Coefficient of MAGMA enrichment of input genes forAD GWAS risk genes
−log10 MAGMA p-value
204 Trajectory 1
144 Trajectory 2
1037 Trajectory 3
157 Trajectory 4
84 Trajectory 5
241 Trajectory 6
91 Trajectory 7
288 Trajectory 8
117 Trajectory 9
177 Trajectory 10
24 Trajectory 10
All Micro genes
Late adulthood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


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
50
96
141
187
−log10(Adj.P.value)
Trajectory 3
Trajectory 6
Trajectory 10
AD associated Micro genes in trajectories


e
IN_ADRAB2 late adulthood aDEGs in Trajectory 2
Age
Late adulthood aDEGs
FAT1
MYO3A
NMBR
FRAS1
UTRN
BMP6
PRKCB
SCAI
ERBB4
FLRT2
MPPED1
RELN
PTCHD4
HS3ST5

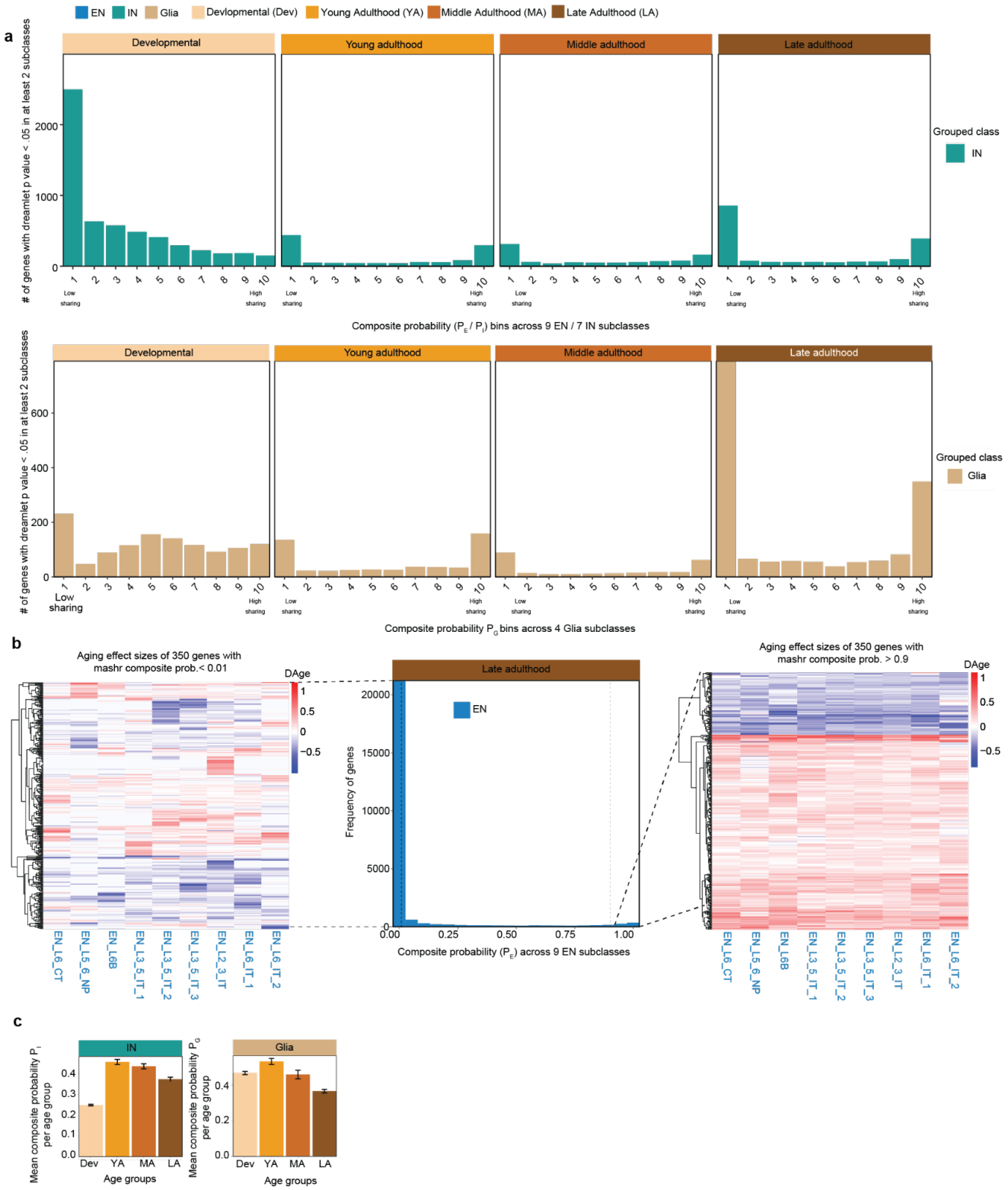
f
cerebral cortex cell migration
positive regulation of kinase activity
positive regulation of AMPA receptor activity
regulation of hemopoiesis
phosphorylation
heparan sulfate proteoglycan biosynthetic process, enzymatic modification
positive regulation of signal transduction
positive regulation of vascular permeability
mitotic nuclear membrane disassembly
regulation of coagulation
lipoprotein localization
protein sulfation
positive regulation of CREB transcription factor activity
mammary gland epithelial cell differentiation
positive regulation of peptidyl-tyrosine phosphorylation
type B pancreatic cell differentiation

Supplementary Figure 9. 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.

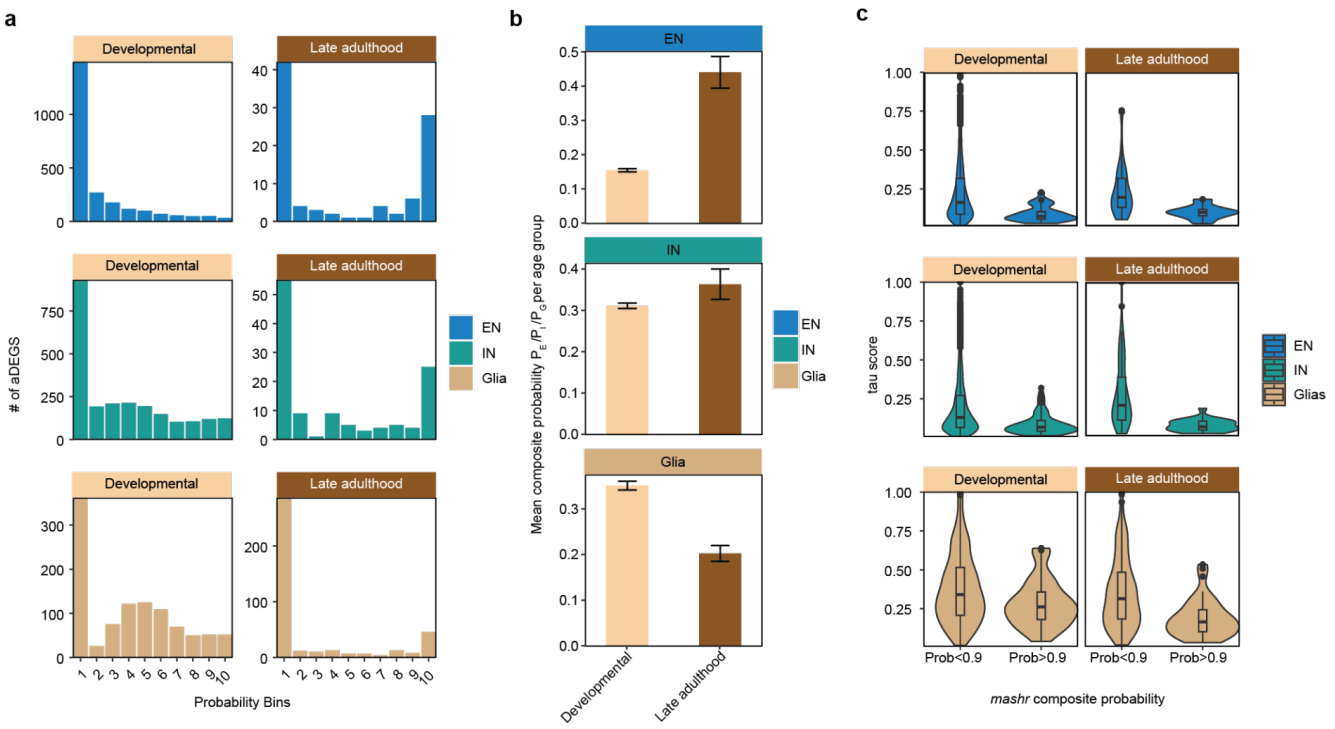


Age groups	Grouped class	# genes p-val	# genes p-val <.05 in at least two sub classes	# genes p-val <.05 in at least two sub classes + $P_E/P_I/P_G$ > 0.9	# aDEGs	# aDEGs + $P_E/P_I/P_G$ > 0.9
Developmental	EN	23,914	4,767	43	2,417	33
Young adulthood	EN	26,809	1,926	279	26	0
Middle adulthood	EN	26,107	2,092	148	0	0
Late adulthood	EN	24,795	2,271	403	93	28
Developmental	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
Developmental	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

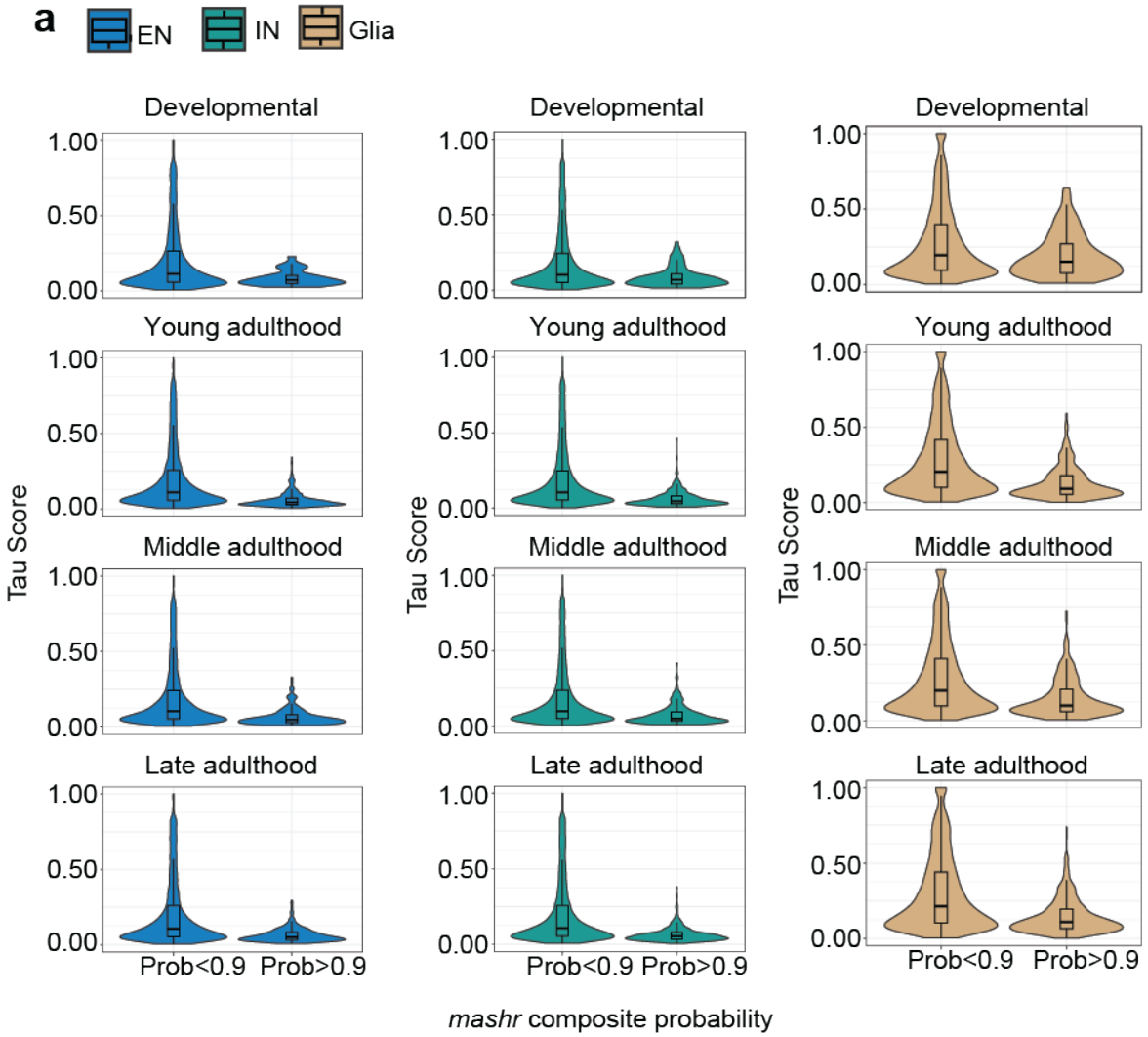
Supplementary Figure 10. Summary of number of shared genes from *mashr* analysis. a, Table displaying the total number of genes to run the *mashr* analysis and shared genes at composite probability > 0.9. Table also displays the counts of genes that had age effect size at significant p-value < 0.5 in at least two subclasses. These genes are used to plot Fig. S11a and Fig 3a**. 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.



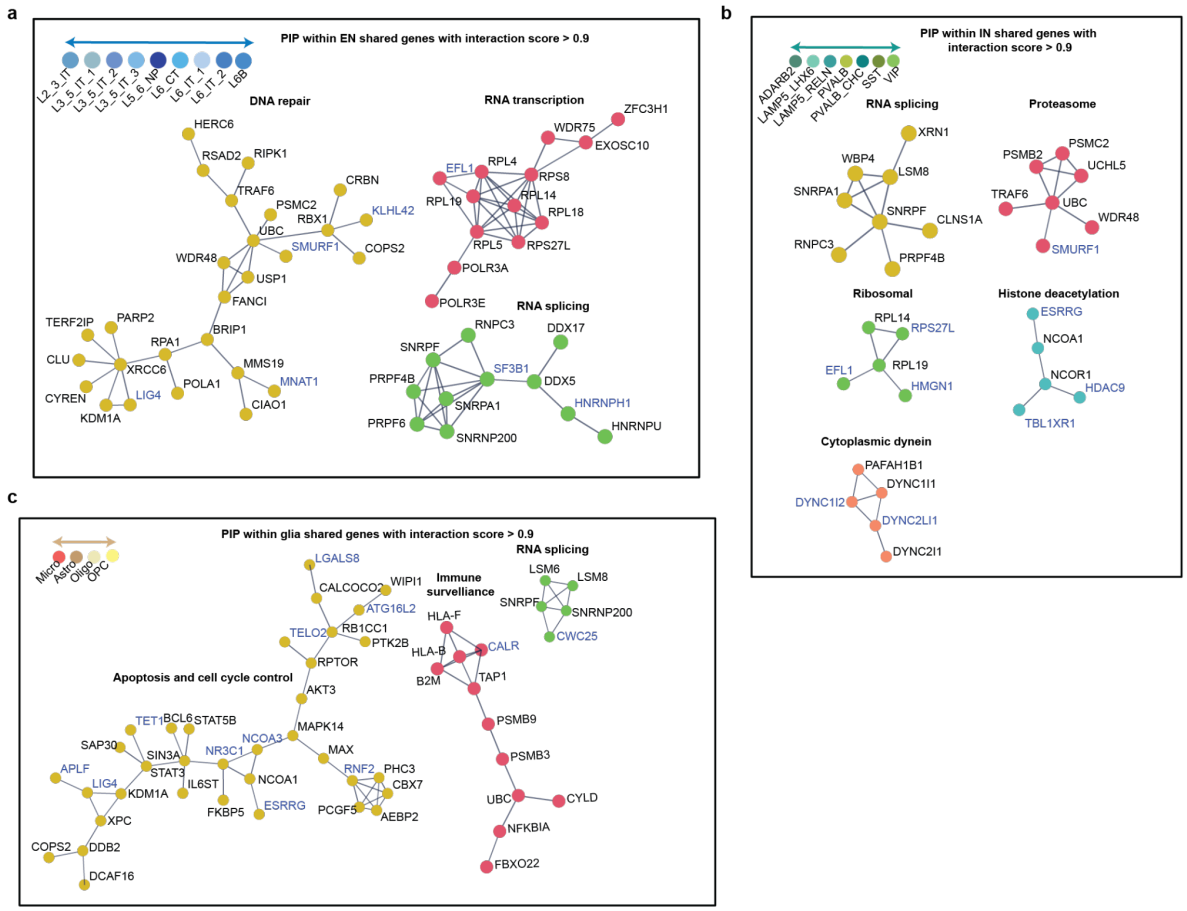
Supplementary Figure 11. Degree of sharing across subclasses a, Bar plot displaying the number of genes stratified into 10 equally sized bins based on composite probability (P_I and P_G) values for IN and glia respectively, ranging from 0 to 1 from development to late adulthood group. The number of input genes to make these bar-plots is displayed in Fig. S11a. **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 Fig. S12a.



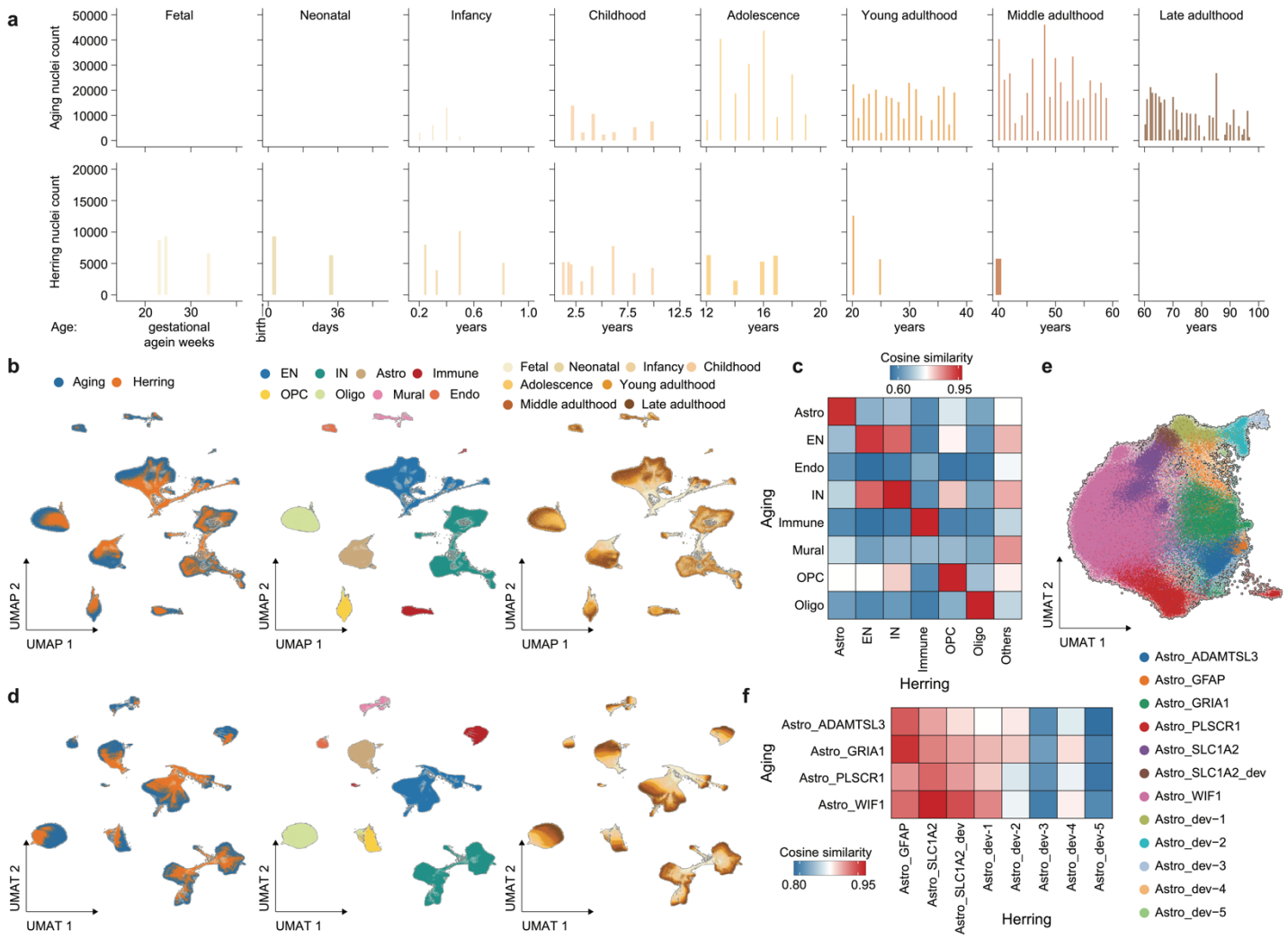
Supplementary Figure 12. Composite probability of aDEGs. a, Bar plot displaying the number of aDEGs stratified into 10 equally sized bins based on composite probability (P_E, P_I, P_G) values for EN, IN and glia ranging from 0 to 1 for development and late adulthood groups. **b,** Mean P_E, P_I, P_G of all developmental and late adulthood aDEGs. This plot is the mean of probabilities shown in Fig. S12a. **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.



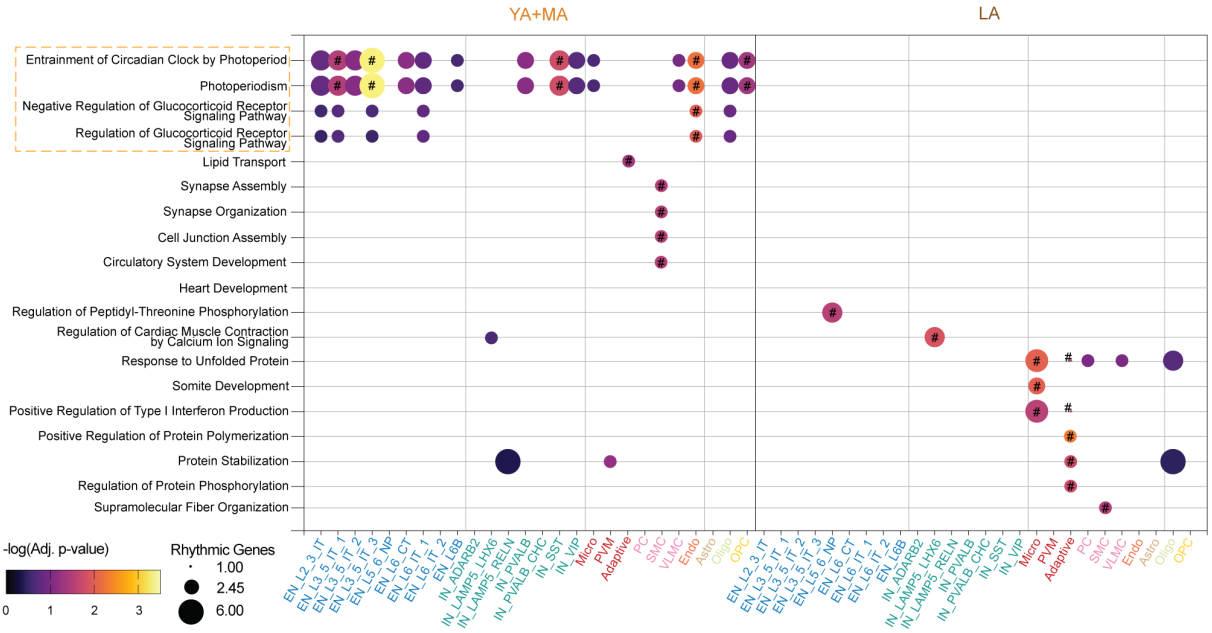
Supplementary Figure 13. Distribution of cell specificity tau score. a, Violin plot of tau scores of genes that have age associated with nominal p-value from dreamlet tool in at least two subclasses. The x-axis is stratified by genes with P_E, P_I, P_G < 0.9 and > 0.9. These tau scores were obtained from 9 EN and 7 IN and 4 Glia expressions separately for each age group.



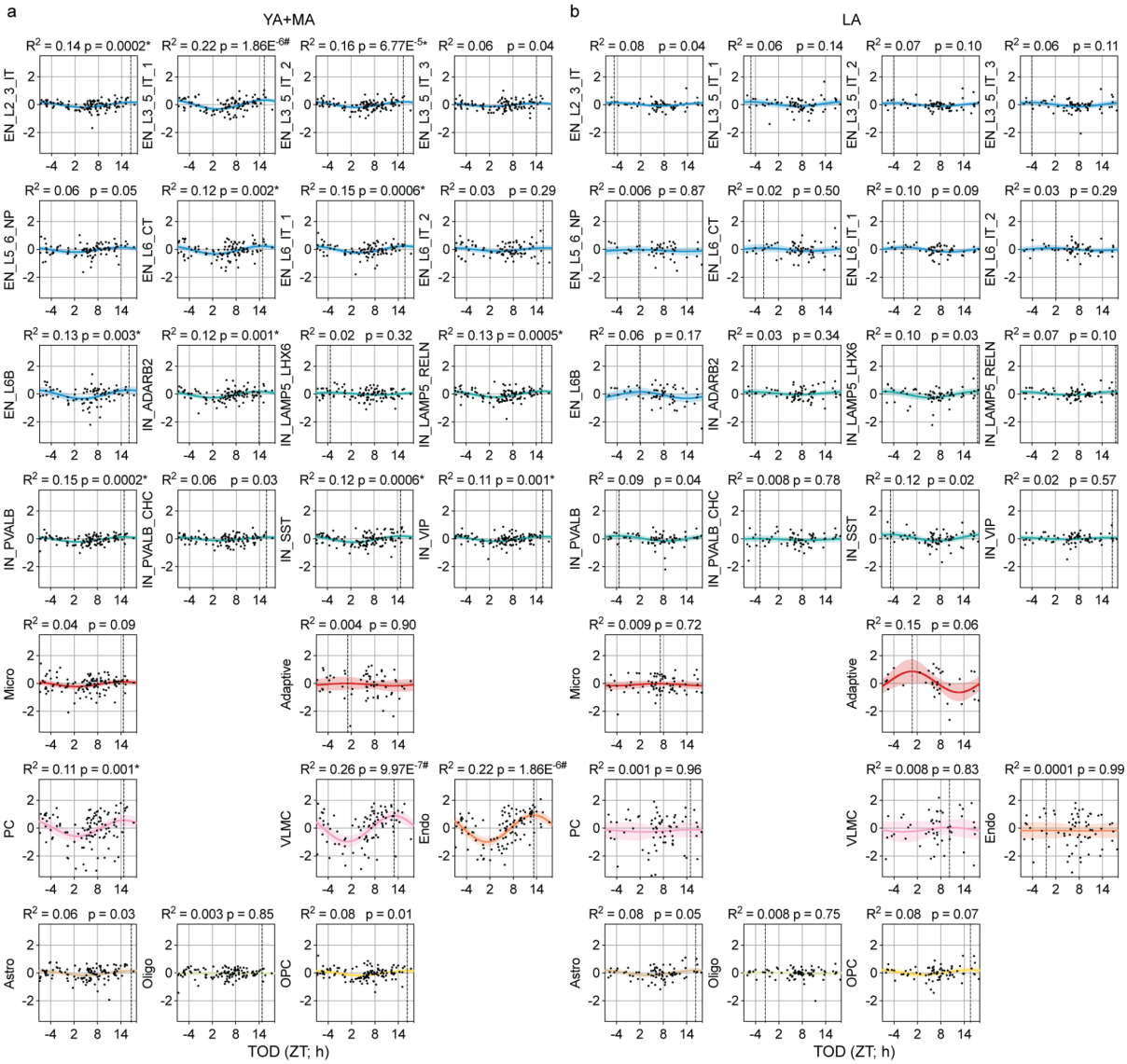
Supplementary Figure 14. Protein-protein interactions among shared genes. a-c Identified clusters of networks of genes that show significant PPI interactions with score > 0.9 obtained from the analysis of shared genes within 9 EN, 7 IN and 4 glia classes. Genes highlighted in blue indicate downregulation 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.



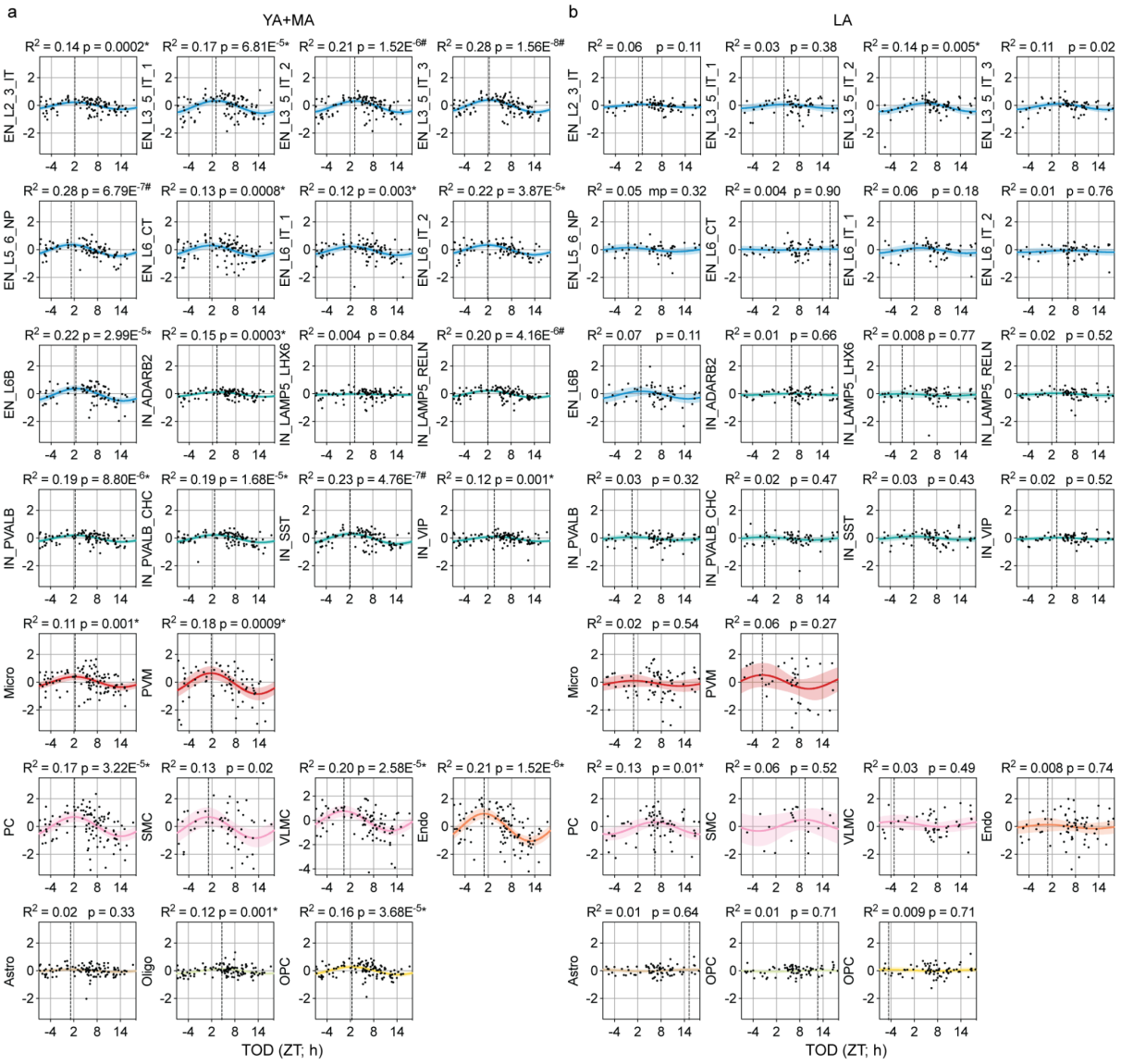
Supplementary Figure 15. Characteristics and similarity of the combined snRNA-seq dataset. a, Nuclei counts in current (top) and previous (bottom) snRNA-seq datasets of the human DLPFC, categorized by age group and color-coded. **b**, Integration of the current dataset with previously published data. UMAP representation of the combined snRNA-seq dataset colored by dataset (left), class (middle), and stage (right). **c**, Cell type similarity between the two datasets at the class level. **d**, UMAT representation of the combined snRNA-seq dataset colored as in panel **b**. **e**, UMAT representation of the astrocyte lineage colored by subtype. **f**, Cell type similarity of the astrocyte lineage between the two datasets at the subtype level.



Supplementary Figure 16. Pathways enriched in rhythmic genes. enrichR pathway analysis of rhythmic genes (p-value < 0.01) in YA+MA and LA subclasses. Pathways shown were identified as significantly enriched (adjusted p-value < 0.05, represented by #) in at least one subclass. Other points represent classes in which the pathway was only nominally significant (p < 0.01). The dashed box surrounds pathways that were driven by circadian clock genes.



Supplementary Figure 17. 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 18. 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.

Captions for Supplementary table 1-4

Supplementary Table 1 | Summary metadata of 284 donors ranges from 0 to 97 years.
Supplementary Table 2 | List of GWAS traits and references
Supplementary Table 3 | Summary statistics of changes in nuclei composition using *crumblr* tool.
Supplementary Table 4 | Tau scores of each gene across EN, IN and glia in four age groups.

Captions for Supplementary Data 1-20

Supplementary Data 1 | Metadata of 284 donors with neuropathology details.
Supplementary Data 2 | Overall scDRS statistics from 1,307,674 nuclei and 284 donors for a subset of neurological, psychiatric, metabolic and immunological traits.
Supplementary Data 3 | Variance explained by technical and biological covariates for each gene across 26 subclasses and four age groups.
Supplemental Data 4 | Dreamlet age association summary statistics of genes for four age groups across 26 subclasses.
Supplemental Data 5 | Functional pathway analysis of four age groups specific aDEGs across 26 subclasses.
Supplemental Data 6 | Enrichment of brain and non-brain related risk genes from GWAS in four age groups specific aDEGs across 26 subclasses.
Supplemental Data 7 | Lifespan age association summary statistics of each gene from a polynomial log model across 26 subclasses.
Supplemental Data 8 | Functional pathway analysis of genes within each subclass across ten trajectories.
Supplemental Data 9 | Enrichment of brain and non-brain related risk genes from GWAS in developmental and late adulthood aDEGs across ten trajectories.
Supplemental Data 10 | Functional pathway analysis of developmental and late adulthood aDEGs across ten trajectories.
Supplemental Data 11 | Composite *mashr* probabilities representing degree of sharing of a gene across 9 EN, 7 IN and 4 glial subclasses across four age groups.
Supplemental Data 12 | Functional pathway analysis of shared genes across all cell classes and PPI cluster index in EN and IN from young and late adulthood, and in glia from late adulthood.
Supplemental Data 13 | PPI clusters in EN and IN from young and late adulthood, and in glia from late adulthood from STRINGDB.
Supplemental Data 14 | Clusters of traDEGs within Astro, Micro, OPC+Oligo, EN and IN lineages.
Supplemental Data 15 | Pathway enrichment analyses of traDEG clusters within Astro, Micro, OPC+Oligo, EN and IN lineages.
Supplemental Data 16 | Enrichment of brain and non-brain related risk genes from GWAS in traDEG clusters within Astro, Micro, OPC+Oligo, EN and IN lineages.
Supplementary Data 17 | Rhythmic gene expression analyses.
Supplementary Data 18 | Gene expression rhythm parameter changes between YA+MA and LA within each subclass.
Supplementary Data 19 | Gene expression rhythms - ΔR^2 between YA+MA and LA within each subclass.
Supplementary Data 20 | enrichR pathway analysis of rhythmic genes.

Supplementary Notes

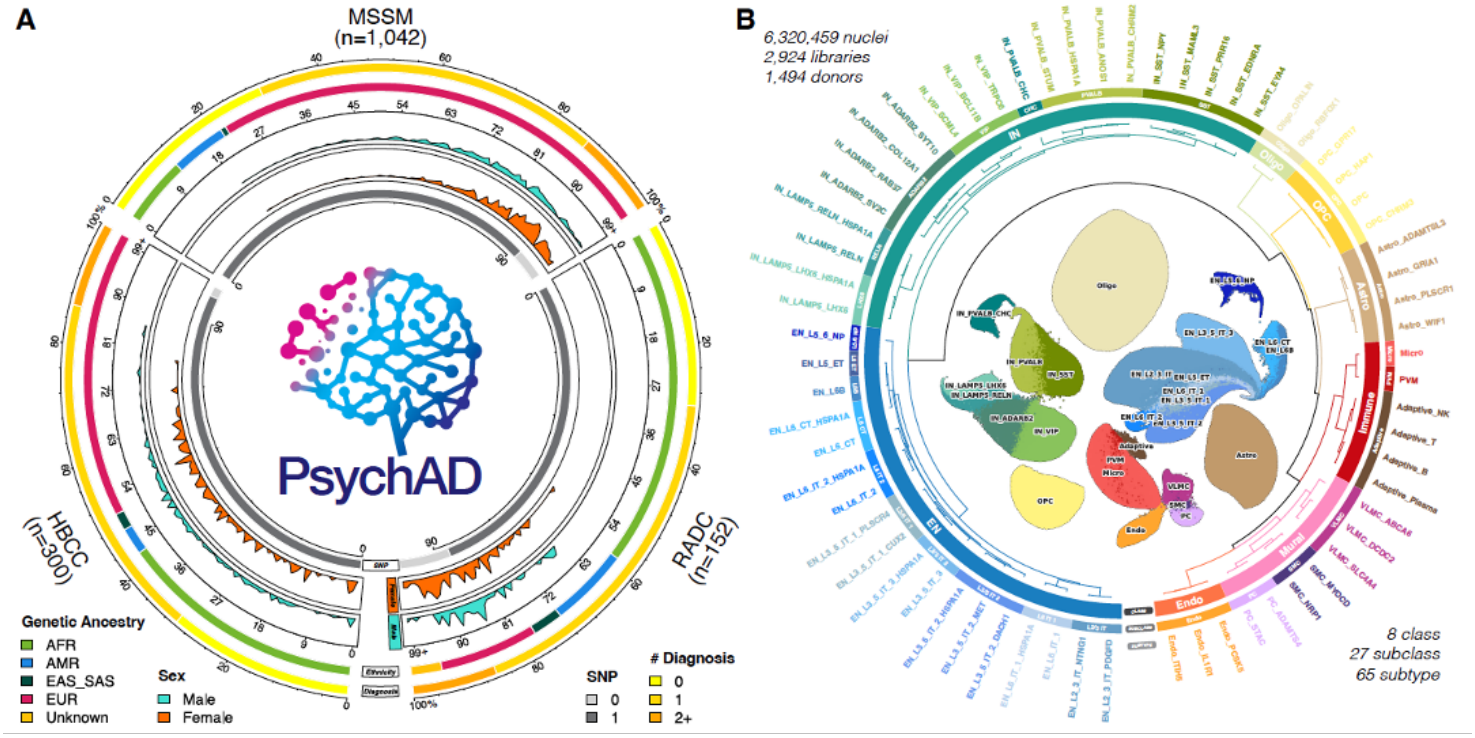
The PsychAD dataset

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[35] or RADC[85,95] 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[111]. So far, various studies have examined population data to characterize NPS along the AD continuum[112–114]. 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[112]. 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[115] and human primary motor cortex[116] 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[33] 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. This repository includes sample metadata, as well as raw and processed sequencing data for snRNA-seq and genotyping.



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).

References

1. Nejati, V., Majdi, R., Salehinejad, M. A. & Nitsche, M. A. The role of dorsolateral and ventromedial prefrontal cortex in the processing of emotional dimensions. *Sci. Rep.* **11**, 1971 (2021).
2. Barbey, A. K., Koenigs, M. & Grafman, J. Dorsolateral prefrontal contributions to human working memory. *Cortex* **49**, 1195–1205 (2013).
3. Hertrich, I., Dietrich, S., Blum, C. & Ackermann, H. The Role of the Dorsolateral Prefrontal Cortex for Speech and Language Processing. *Front. Hum. Neurosci.* **15**, 645209 (2021).
4. Smucny, J., Dienel, S. J., Lewis, D. A. & Carter, C. S. Mechanisms underlying dorsolateral prefrontal cortex contributions to cognitive dysfunction in schizophrenia. *Neuropsychopharmacology* **47**, 292–308 (2022).
5. Girdhar, K. *et al.* Chromatin domain alterations linked to 3D genome organization in a large cohort of schizophrenia and bipolar disorder brains. *Nat. Neurosci.* **25**, 474–483 (2022).
6. Pizzagalli, D. A. & Roberts, A. C. Prefrontal cortex and depression. *Neuropsychopharmacology* **47**, 225–246 (2022).
7. Cain, A. *et al.* Multicellular communities are perturbed in the aging human brain and Alzheimer's disease. *Nat. Neurosci.* **26**, 1267–1280 (2023).
8. Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428–4447.e28 (2022).

9. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *Science* **382**, eadf0834 (2023).

10. Zeng, B. *et al.* The single-cell and spatial transcriptional landscape of human gastrulation and early brain development. *Cell Stem Cell* **30**, 851–866.e7 (2023).

11. Eze, U. C., Bhaduri, A., Haeussler, M., Nowakowski, T. J. & Kriegstein, A. R. Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nat. Neurosci.* **24**, 584–594 (2021).

12. Herring, C. A. *et al.* Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution. *Cell* **185**, 4428–4447.e28 (2022).

13. Duffy, M. F. *et al.* Divergent patterns of healthy aging across human brain regions at single-cell resolution reveal links to neurodegenerative disease. *bioRxivorg* (2023) doi:10.1101/2023.07.31.551097.

14. Fullard, J. F. *et al.* Population-scale cross-disorder atlas of the human prefrontal cortex at single-cell resolution.

15. Lee, D. *et al.* Single-cell atlas of transcriptomic vulnerability across multiple neurodegenerative and neuropsychiatric diseases.

16. Muraoka, S. *et al.* Enrichment of Neurodegenerative Microglia Signature in Brain-Derived Extracellular Vesicles Isolated from Alzheimer's Disease Mouse Models. *J. Proteome Res.* **20**, 1733–1743 (2021).

17. Felsky, D. *et al.* Polygenic analysis of inflammatory disease variants and effects on microglia in the aging brain. *Mol. Neurodegener.* **13**, 38 (2018).

18. Bahrami, S. *et al.* Shared Genetic Loci Between Body Mass Index and Major Psychiatric Disorders: A Genome-wide Association Study. *JAMA Psychiatry* **77**, 503–512 (2020).

19. Marcus, M. D. & Wildes, J. E. Obesity: is it a mental disorder? *Int. J. Eat. Disord.* **42**, 739–753 (2009).

20. Rajan, T. M. & Menon, V. Psychiatric disorders and obesity: A review of association studies. *J. Postgrad. Med.* **63**, 182–190 (2017).

21. Leutner, M. *et al.* Obesity as pleiotropic risk state for metabolic and mental health throughout life. *Transl. Psychiatry* **13**, 1–12 (2023).

22. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).

23. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. doi:10.1101/2022.10.24.513555.

24. Ham, S. & Lee, S.-J. V. Advances in transcriptome analysis of human brain aging. *Exp. Mol. Med.* **52**, 1787–1797 (2020).

25. Chaudhari, P. R., Singla, A. & Vaidya, V. A. Early Adversity and Accelerated Brain Aging: A Mini-Review. *Front. Mol. Neurosci.* **15**, 822917 (2022).

26. Chiou, K. L. *et al.* Multiregion transcriptomic profiling of the primate brain reveals signatures of aging and the social environment. *Nat. Neurosci.* **25**, 1714–1723 (2022).

27. Zhang, M. J. *et al.* Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat. Genet.* **54**, 1572–1580 (2022).

28. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

29. Count ratio uncertainty modeling base linear regression. https://diseaseneurogenomics.github.io/crumblr/.

30. Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol. Rev.* **20**, 327–348 (2010).

31. Arain, M. *et al.* Maturation of the adolescent brain. *Neuropsychiatr. Dis. Treat.* **9**, 449–461 (2013).

32. Perlman, K. *et al.* Developmental trajectory of oligodendrocyte progenitor cells in the human brain revealed by single cell RNA sequencing. *Glia* **68**, 1291–1303 (2020).

33. 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.

34. El Marroun, H. *et al.* Prenatal exposure to maternal and paternal depressive symptoms and brain morphology: A population-based prospective neuroimaging study in young children. *Depress. Anxiety* **33**, 658–666 (2016).

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

36. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

37. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

38. Wang, Z. *et al.* Microglial autophagy in Alzheimer's disease and Parkinson's disease. *Front. Aging Neurosci.* **14**,

1065183 (2022).

39. Santana, D. A., Smith, M. de A. C. & Chen, E. S. Histone Modifications in Alzheimer's Disease. *Genes* **14**, (2023).

40. Lynch, G., Rex, C. S. & Gall, C. M. Synaptic plasticity in early aging. *Ageing Res. Rev.* **5**, 255–280 (2006).

41. Arcos-Burgos, M., Lopera, F., Sepulveda-Falla, D. & Mastronardi, C. Neural Plasticity during Aging. *Neural Plast.* **2019**, 6042132 (2019).

42. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat. Genet.* **51**, 187–195 (2019).

43. Lüleci, H. B. & Yılmaz, A. Robust and rigorous identification of tissue-specific genes by statistically extending tau score. *BioData Min.* **15**, 31 (2022).

44. Schumacher, B., Pothof, J., Vijg, J. & Hoeijmakers, J. H. J. The central role of DNA damage in the ageing process. *Nature* **592**, 695–703 (2021).

45. Maynard, S., Fang, E. F., Scheibye-Knudsen, M., Croteau, D. L. & Bohr, V. A. DNA Damage, DNA Repair, Aging, and Neurodegeneration. *Cold Spring Harb. Perspect. Med.* **5**, (2015).

46. Bernier, L.-P. *et al.* Microglial metabolic flexibility supports immune surveillance of the brain parenchyma. *Nat. Commun.* **11**, 1559 (2020).

47. Kellogg, C. M. *et al.* Microglial MHC-I induction with aging and Alzheimer's is conserved in mouse models and humans. *bioRxiv* (2023) doi:10.1101/2023.03.07.531435.

48. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2022).

49. Sofroniew, M. V. & Vinters, H. V. Astrocytes: biology and pathology. *Acta Neuropathol.* **119**, 7–35 (2010).

50. Siletti, K. *et al.* Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).

51. Molofsky, A. V. *et al.* Astrocytes and disease: a neurodevelopmental perspective. *Genes Dev.* **26**, 891–907 (2012).

52. Köhler, S., Winkler, U. & Hirrlinger, J. Heterogeneity of astrocytes in grey and white matter. *Neurochem. Res.* **46**, 3–14 (2021).

53. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–7 (2016).

54. Perez-Catalan, N. A., Doe, C. Q. & Ackerman, S. D. The role of astrocyte-mediated plasticity in neural circuit

development and function. *Neural Dev.* **16**, 1 (2021).

55. Chung, W.-S., Allen, N. J. & Eroglu, C. Astrocytes control synapse formation, function, and elimination. *Cold Spring Harb. Perspect. Biol.* **7**, a020370 (2015).
56. Abbott, N. J., Rönnbäck, L. & Hansson, E. Astrocyte-endothelial interactions at the blood-brain barrier. *Nat. Rev. Neurosci.* **7**, 41–53 (2006).
57. Attwell, D. *et al.* Glial and neuronal control of brain blood flow. *Nature* **468**, 232–243 (2010).
58. Lee, H.-G., Wheeler, M. A. & Quintana, F. J. Function and therapeutic value of astrocytes in neurological diseases. *Nat. Rev. Drug Discov.* **21**, 339–358 (2022).
59. Kruyer, A., Kalivas, P. W. & Scofield, M. D. Astrocyte regulation of synaptic signaling in psychiatric disorders. *Neuropsychopharmacology* **48**, 21–36 (2023).
60. Allen, N. J. & Lyons, D. A. Glia as architects of central nervous system formation and function. *Science* **362**, 181–185 (2018).
61. Sanmarco, L. M. *et al.* Gut-licensed IFNγ+ NK cells drive LAMP1+TRAIL+ anti-inflammatory astrocytes. *Nature* **590**, 473–479 (2021).
62. Linnerbauer, M., Wheeler, M. A. & Quintana, F. J. Astrocyte crosstalk in CNS inflammation. *Neuron* **108**, 608–622 (2020).
63. Wheeler, M. A. *et al.* MAFG-driven astrocytes promote CNS inflammation. *Nature* **578**, 593–599 (2020).
64. Ling, E. *et al.* A concerted neuron-astrocyte program declines in ageing and schizophrenia. *Nature* **627**, 604–611 (2024).
65. Chen, X. *et al.* Microglia-mediated T cell infiltration drives neurodegeneration in tauopathy. *Nature* **615**, 668–677 (2023).
66. Bartels, T., De Schepper, S. & Hong, S. Microglia modulate neurodegeneration in Alzheimer's and Parkinson's diseases. *Science* **370**, 66–69 (2020).
67. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
68. Wang, W., Zheng, S., Shin, S. C. & Yuan, G.-C. Characterizing spatially continuous variations in tissue microenvironment through niche trajectory analysis. *bioRxivorg* 2024.04.23.590827 (2024).

69. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).

70. Silbereis, J. C., Pochareddy, S., Zhu, Y., Li, M. & Sestan, N. The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016).

71. Zhu, K. *et al.* Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Sci. Adv.* **9**, eadg3754 (2023).

72. Forsyth, J. K. & Lewis, D. A. Mapping the consequences of impaired synaptic plasticity in schizophrenia through development: An integrative model for diverse clinical features. *Trends Cogn. Sci.* **21**, 760–778 (2017).

73. Ma, Y. *et al.* Activity-Dependent Transcriptional Program in NGN2+ Neurons Enriched for Genetic Risk for Brain-Related Disorders. *Biol. Psychiatry* **95**, 187–198 (2024).

74. Johnson, K. A. *et al.* Tourette syndrome: clinical features, pathophysiology, and treatment. *Lancet Neurol.* **22**, 147–158 (2023).

75. Lowe, C. J., Reichelt, A. C. & Hall, P. A. The prefrontal cortex and obesity: A health neuroscience perspective. *Trends Cogn. Sci.* **23**, 349–361 (2019).

76. Südhof, T. C. The cell biology of synapse formation. *J. Cell Biol.* **220**, (2021).

77. Chen, M. B., Jiang, X., Quake, S. R. & Südhof, T. C. Persistent transcriptional programmes are associated with remote memory. *Nature* **587**, 437–442 (2020).

78. Fawcett, J. W., Oohashi, T. & Pizzorusso, T. The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat. Rev. Neurosci.* **20**, 451–465 (2019).

79. Stowe, T. A. & McClung, C. A. How Does Chronobiology Contribute to the Development of Diseases in Later Life. *Clin. Interv. Aging* **18**, 655–666 (2023).

80. Wolff, C. A. *et al.* Defining the age-dependent and tissue-specific circadian transcriptome in male mice. *Cell Rep.* **42**, 111982 (2023).

81. Chen, C.-Y. *et al.* Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 206–211 (2016).

82. Uddin, M. S., Yu, W. S. & Lim, L. W. Exploring ER stress response in cellular aging and neuroinflammation in Alzheimer's disease. *Ageing Res. Rev.* **70**, 101417 (2021).

83. Kim, Y. H. & Lazar, M. A. Transcriptional Control of Circadian Rhythms and Metabolism: A Matter of Time and Space. *Endocr. Rev.* **41**, 707–732 (2020).

84. Li, M. *et al.* Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

85. 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.

86. Su, Y. *et al.* A single-cell transcriptome atlas of glial diversity in the human hippocampus across the postnatal lifespan. *Cell Stem Cell* **29**, 1594–1610.e8 (2022).

87. Nowakowski, T. J. *et al.* Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).

88. Reid, D. A. *et al.* Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* **372**, 91–94 (2021).

89. Li, H., Wang, Z., Ma, T., Wei, G. & Ni, T. Alternative splicing in aging and age-related diseases. *Translational Medicine of Aging* **1**, 32–40 (2017).

90. Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).

91. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).

92. Khodosevich, K. & Sellgren, C. M. Neurodevelopmental disorders-high-resolution rethinking of disease modeling. *Mol. Psychiatry* **28**, 34–43 (2023).

93. Rozycka, A. & Liguz-Lecznar, M. The space where aging acts: focus on the GABAergic synapse. *Aging Cell* **16**, 634–643 (2017).

94. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).

95. 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).

96. Thion, M. S., Ginhoux, F. & Garel, S. Microglia and early brain development: An intimate journey. *Science* **362**, 185–189 (2018).

97. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).

98. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* 2021.05.05.442755 (2021) doi:10.1101/2021.05.05.442755.

99. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).

100. Li, B. *et al.* Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).

101. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).

102. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

103. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).

104. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2018).

105. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

106. Count ratio uncertainty modeling base linear regression. https://diseaseneurogenomics.github.io/crumblr/.

107. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).

108. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

109. Seney, M. L. *et al.* Diurnal rhythms in gene expression in the prefrontal cortex in schizophrenia. *Nat. Commun.* **10**, 3355 (2019).

110. Xue, X. *et al.* DiffCircaPipeline: a framework for multifaceted characterization of differential rhythmicity. *Bioinformatics* **39**, (2023).

111. 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).

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

113. 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).

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

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

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