Single-cell atlas of transcriptomic vulnerability across multiple neurodegenerative and neuropsychiatric diseases

Authors

Donghoon Lee[†#1,2,3,4], Mikaela Koutrouli[#1,2,3,4,5], Nicolas Y. Masse[#1,2,3,4], Gabriel E. Hoffman[#1,2,3,4,6,7], Seon Kinrot[1,2,3,4], Xinyi Wang[1,2,3,4], Prashant N.M.[1,2,3,4], Milos Pjanic[1,2,3,4], Tereza Clarence[1,2,3,4], Fotios Tsetsos[1,2,3,4,6], Deepika Mathur[1,2,3,4,6,7], David Burstein[1,2,3,4,6,7], Karen Therrien[1,2,3,4], 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], Jennifer Monteiro Fortes[1,2,3,4], Pavel Katsel[3], Pavan K. Auluck[8], Lisa L. Barnes[9,10], Stefano Marenco[8], David A. Bennett[9,10], PsychAD Consortium[11], Lars Juhl Jensen[5], Kiran Girdhar[1,2,3,4], Georgios Voloudakis[1,2,3,4,6,7,12], Vahram Haroutunian[2,3,7,13], Jaroslav Bendl[1,2,3,4], John F. Fullard[#1,2,3,4], Panos Roussos[†#1,2,3,4,6,7]

Affiliations

1 Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
2 Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
3 Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
4 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
5 Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
6 Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, USA.
7 Mental Illness Research, Education and Clinical Center VISN2, James J. Peters VA Medical Center, Bronx, NY, USA.
8 Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, MD, USA.
9 Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA.
10 Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA.
11 PsychAD Consortium
12 Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
13 Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

# These authors contributed equally to this work.
† Corresponding author: donghoon.lee@mssm.edu; panagiotis.roussos@mssm.edu

Abstract

Neurodegenerative and neuropsychiatric diseases impose a significant societal and public health burden. However, our understanding of the molecular mechanisms underlying these highly complex conditions remains limited. To gain deeper insights into the etiology of different brain diseases, we used specimens from 1,494 unique donors to generate a population-scale single-cell transcriptomic atlas of the human dorsolateral prefrontal cortex (DLPFC), comprising over 6.3 million individual nuclei. The cohort includes neurotypical controls as well as donors affected by eight common and complex brain disorders: Alzheimer's disease (AD), diffuse Lewy body disease (DLBD), vascular dementia (Vas), Parkinson's disease (PD), tauopathy, frontotemporal dementia, schizophrenia, and bipolar disorder. We show that inter-individual variation accounts for a substantial portion of gene expression variation in the DLPFC. By comparing transcriptomic variation across diseases, we reveal universal signatures enriched in basic cellular functions such as mRNA splicing and protein localization. After discounting these cross-disease signatures, we show strong genetic and transcriptomic concordance among AD, DLBD, Vas, and PD, largely driven by alteration of synaptic signaling functions in neurons. Furthermore, we characterize transcriptomic variation among different AD phenotypes that were distinct from healthy aging. We uncover mitigating effects of interneurons and aggravating effects of immune and vascular cells in AD dementia. Further exploring the effect of the neuropsychiatric symptoms frequently accompanying AD, we identify a link to deep layer excitatory neurons. By constructing transcriptome trajectories that capture AD progression, we show cell-type specific responses implicated in early and late stages of AD. Our atlas provides an unprecedented perspective of the transcriptomic landscape in neurodegenerative and neuropsychiatric diseases, shedding light on shared and distinct processes involving the neuro-immune-vascular systems, and identifying potential targets for therapeutic intervention.

Main

The human brain is a highly complex organ composed of billions of functionally diverse cells. Under pathogenic stress, cellular and molecular responses are often convoluted and contextual, so understanding their dysfunction in disease is challenging. Recent work has begun to unravel the molecular changes that occur at the single-cell level, which has been particularly helpful in understanding the vulnerability of specific cell types in various disease contexts, as well as the complex interplay between different cell types. In Alzheimer's disease (AD), it has been shown that a thorough exposition of cellular heterogeneity in the brain, the coordinated interactions between neurons and glia, and the selective depletion of vulnerable inhibitory neuronal subtypes are critical for understanding AD pathology[1,2].

Building a large-scale disease atlas at single-cell resolution creates an unprecedented opportunity to understand molecular responses at the cellular level and estimate population-level variation in the brain transcriptome. A large sample size provides the resolution needed to establish robust basal-level conditions and to sufficiently capture the full spectrum of disease pathology. By leveraging cross-disease atlases, studies have revealed shared and distinct patterns of gene-expression perturbations in major psychiatric diseases as well as shared genetic factors leading to molecular convergence[3,4]. As such, characterizing shared transcriptomic vulnerabilities and pathophysiology together has significant implications for early treatment and the development of effective therapeutics.

Here, we introduce the PsychAD cohort, which consists of 1,494 unique brain donors affected by various neurodegenerative and neuropsychiatric diseases in addition to neurotypical controls. The resulting single-nucleus RNA sequencing (snRNA-seq) dataset in the dorsolateral prefrontal cortex (DLPFC) of those donors, comprising over 6.3 million nuclei representing 27 distinct subclasses of cells, is sufficiently-powered to identify molecular signatures from multiple traits while accounting for individual variation. We utilize this dataset to uncover shared transcriptomic vulnerability across these diseases and, in so doing, to better understand specific transcriptional patterns and regulatory drivers underpinning each trait. By characterizing putative disease-driving molecular changes across multiple traits, we differentiate shared and novel disease- and cell type-specific associations. Concordance between heritability estimates and transcriptomic similarity identifies shared genetic factors underlying cross-disorder traits. Deep phenotyping of AD trajectories using Tau pathology and clinical dementia status suggests a potential link between immune and brain vasculature dysfunctions. In summary, our study provides a rich and comprehensive resource for exploring the cellular and molecular mechanisms of brain function and dysfunction across multiple neurodegenerative and neuropsychiatric diseases.

The PsychAD cohort represents diverse neurodegenerative and neuropsychiatric diseases across the lifespan

The PsychAD cohort comprises 1,494 unique brain donors (**Fig. 1a, Supplementary Table 1**). Brain tissue specimens were obtained from three sources: 1,042 donors from Mount Sinai NIH Neurobiobank (MSSM), 300 from Human Brain Collection Core (HBCC), and 152 from Rush Alzheimer's Disease Center (RADC). The cohort covers the whole lifespan of postnatal ages between 0 and 108, roughly equal numbers of males and females, and represents a diverse range of disease phenotypes, including Alzheimer's disease (AD), diffuse Lewy body disease (DLBD), vascular dementia (Vas), tauopathy (Tau), frontotemporal dementia (FTD), Parkinson's disease (PD), schizophrenia (SCZ), and bipolar disorder (BD). The cohort covers a diverse genetic background and over 30% of the donors were of non-European (EUR) ancestry.

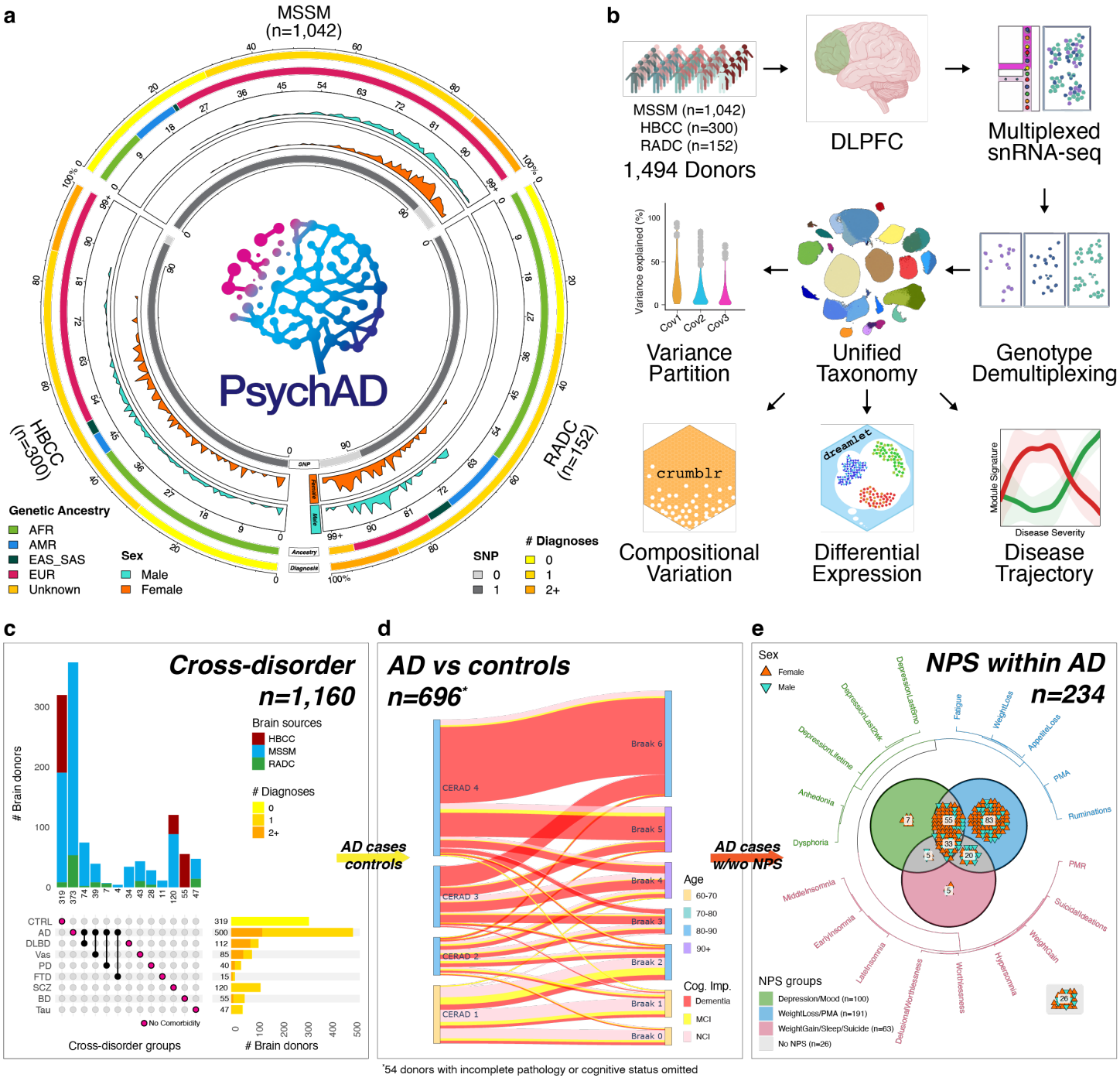
We streamlined the unified processing of the data as well as the harmonization of clinical, and technical metadata (**Fig. 1b, Supplementary Fig. 1a, Methods**). Frozen brain specimens were randomized and processed in batches of 6. Equal numbers of nuclei from each sample were pooled together and each pool

was subjected to snRNA-seq twice to generate a technical replicate. Following quality control (**Supplementary Figs. 1c-i**), the final data sets consisted of 6,320,459 nuclei. To characterize transcriptomic vulnerability across multiple neuropsychiatric and neurodegenerative diseases, we organized the analysis into three tiers.

First, we focused on the cross-disorder analyses (**Fig. 1c**). We targeted six neurodegenerative diseases (NDDs, including AD, DLBD, Vas, Tau, FTD, and PD) and two neuropsychiatric diseases (NPDs, including SCZ and BD), using a subset of 1,160 donors with minimal comorbidity aged ≥ 17. To estimate the sharing of transcriptomic vulnerability, we compared donors affected with NDDs and NPDs against the baseline of 319 neurotypical controls.

The second-tier analysis focused on the stage of AD progression based on pathological and cognitive impairment measures using a subset of 696 individuals (**Fig. 1d**). To uncover cell-type-specific roles in disease onset and disease trajectory, we compared two characteristic neuropathological abnormalities of the AD brain: the accumulation of amyloid-β (Aβ) plaques measured using CERAD plaque density score and tau-based neurofibrillary tangle (NFT) pathology measured using Braak staging, along with cognitive impairment.

In the third tier, we surveyed neuropsychiatric symptoms (NPS) within 234 individuals affected by AD pathology (**Fig. 1e, Supplementary Fig. 1b**). NPS are core features of AD and are common in patients with dementia[5]. We broadly categorized NPS into three groups, (1) depression or mood-related, (2) weight loss or psychomotor agitation (PMA), and (3) weight gain, insomnia, or suicidal ideation, based on co-occurrence estimates, sharing molecular mechanisms that lead to increased prevalence with disease severity.



*54 donors with incomplete pathology or cognitive status omitted

Fig. 1. **Overview of the PsychAD cohort and study design. (a)** Breakdown of donors by tissue source, number of diagnoses, genetic ancestry, distribution of age at death, sex, and % availability of genotype data. **(b)** Data generation and analysis workflow. **(c)** A subset of the PsychAD cohort (n=1,160 donors) focused on cross-disorder contrasts. This subset includes a group of neurotypical controls that were used to compare against 8 neurodegenerative and neuropsychiatric disorders. UpSet plot describing available disease phenotypes and number of unique donors by brain bank. **(d)** A subset of the PsychAD cohort (n=696 donors) focused on AD phenotype contrasts. Comparison of different measures of AD severity, including neuritic plaque density (CERAD), neurofibrillary tangle (NFT) pathology, and cognitive impairments. *n=54 donors with incomplete pathology or cognitive status were omitted from the figure. **(e)** A subset of the PsychAD cohort (n=234 donors) was used to characterize single-cell transcriptomic changes underlying NPS in AD. AFR: African, AMR: Ad Mixed American, EAS: East Asian, SAS: South Asian, EUR: European, NPS: Neuropsychiatric Symptom, PMA: Psychomotor Agitation, PMR: Psychomotor Retardation.

Unified processing and hierarchical cellular taxonomy of the human prefrontal cortex uncovers 27 distinct subclasses of cells

To understand heterogeneous human cortical tissues in disease contexts, we require a cell type taxonomy that is robust to aging, disease phenotypes, and various sampling and technical biases. Following unified computational processing, quality controls, and batch normalization of snRNA-seq libraries representing 1,494 dissections, processed in duplicate (**Methods**), we annotated cell types of the human DLPFC using the cell taxonomy of the primate DLPFC[6] and human primary motor cortex[7] as a baseline reference. The resulting human DLPFC cellular taxonomy was organized using three levels of hierarchy, identifying 8 broad cell classes, 27 subclasses, and 65 functionally distinct subtypes (**Fig. 2a, Supplementary Table 2**). Each level of the annotation hierarchy represents a slice in the clustering dendrogram. At the top, the “class level” of the annotation hierarchy defines 8 major cell types, including two broad neuronal cell types: glutamatergic excitatory (EN) and GABAergic inhibitory neurons (IN), three glial: astrocytes (Astro), oligodendrocytes (Oligo), and oligodendrocyte progenitor cells (OPC), and three non-neuronal cell types: immune cells (Immune), mural and vascular cells (Mural), and endothelial cells (Endo). Subsequent levels of the annotation hierarchy, subclasses and subtypes, 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 (see *iterative clustering* in **Methods**). The subclass level distinguished the EN class into 10 subclasses and the IN class into 7 subclasses (**Supplementary Figs. 2c,d**). Different types of neurons, especially ENs, are organized into six horizontal layers (L1-L6) that are distinct in both cytoarchitecture and function[8]. *In situ* spatial transcriptomics data was used to confirm that the EN subclasses were spatially distinct and found in their respective neocortical layers (**Fig. 2b**). The EN subclasses were denoted by their laminar organization (L2-6) and axon projection characteristics (IT: intra-telencephalic, ET: extra-telencephalic, NP: near projecting, CT: corticothalamic, and L6B). IN subclasses were determined using their characteristic marker genes (**Figs. 2c,d**); Ivy cells (IN_LAMP5_LHX6), neurogliaform cells (IN_LAMP5_RELN), basket cells (IN_PVALB), chandelier cells (IN_PVALB_CHC), Martinotti and non-Martinotti cells (IN_SST), VIP (IN_VIP), and homologs of mouse Sncg inhibitory neurons[6] (IN_ADARB2). Unlike laminar organization of EN, IN subclasses were distributed randomly throughout the gray matter, except for IN_ADARB2, which was predominantly found in the superficial layer of the neocortex (**Supplementary Fig. 2e**). Some previously annotated rare inhibitory neuron types, like SST NPY or SST HGF, were not distinguished at the subclass level, but were identified at the subtype level. The cellular taxonomy was relatively consistent, and the subtypes were well represented across all three brain sources (**Supplementary Figs. 2a,b**). Neuronal cells made up 38.4% (EN 23.0% and IN 15.4%), with oligodendrocytes being the next most abundant, at 36.1%. Major cell types matched well when compared to the previous cellular taxonomy of DLPFC[1,6], and subtypes were relatively concordant (**Supplementary Figs. 2f-h**). Neuronal subclasses exhibit

distinct functional characteristics from non-neuronal subclasses (**Fig. 2e**). ENs are enriched in pathways associated with synaptic vesicle priming and neurotransmitter secretion, whereas INs are enriched with functions related to receptor signaling and ion transport. Immune cell types show unique enrichments in functions related to cytotoxic immune responses, while mural cells are involved in vessel morphogenesis. It has been shown cell types are differentially implicated in mediating disease risks[9,10]. Therefore, we further annotated cell types with disease GWASs using the single-cell disease-relevance score (scDRS) and found that neurological diseases largely involve immune and glial cell types, whereas psychiatric diseases are predominantly associated with neuronal cell types (**Fig. 2f, Supplementary Fig. 2i**).

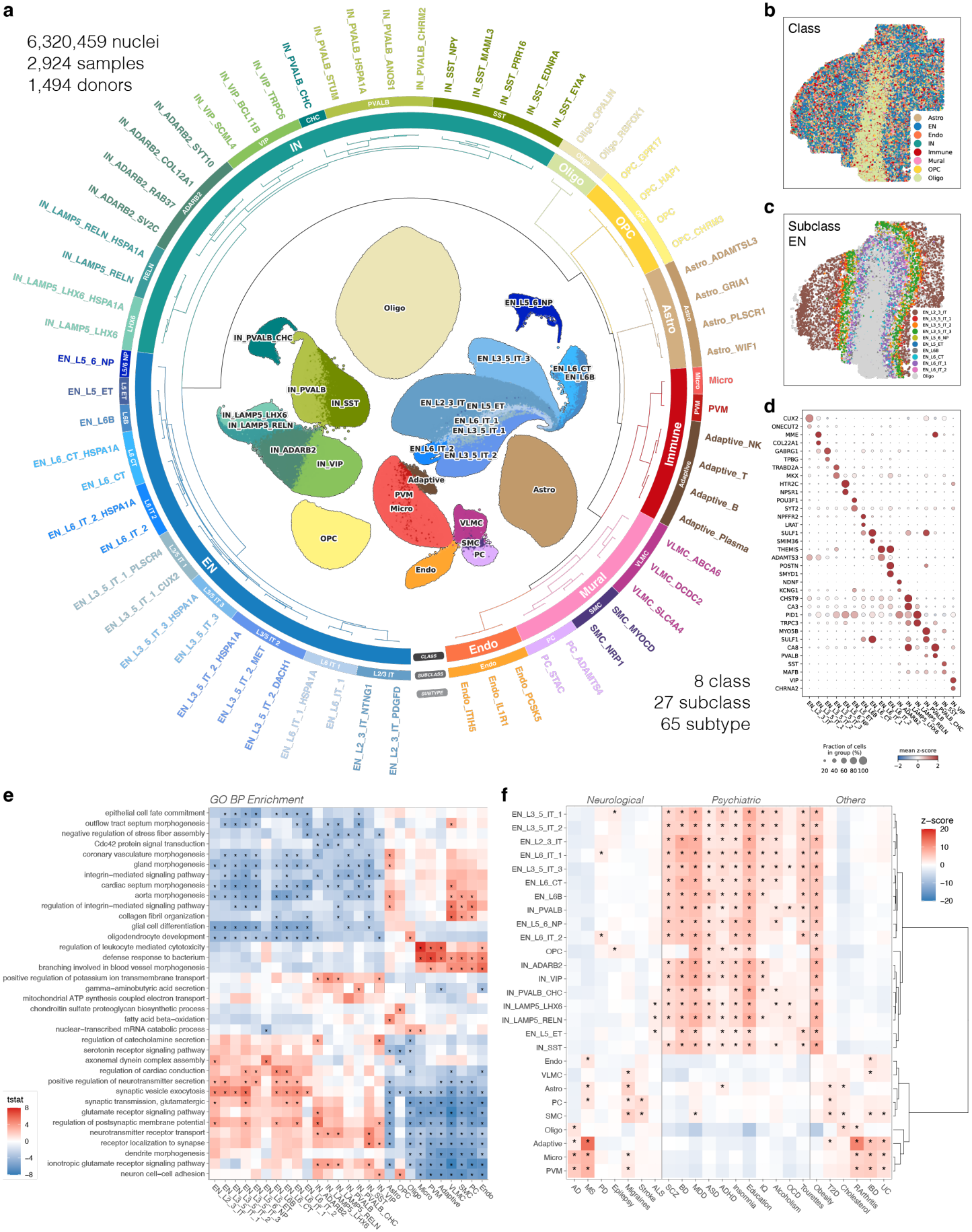


Fig. 2. **Unified processing of the single-cell transcriptomics atlas and hierarchical cellular taxonomy.** **(a)** Hierarchical structure of transcriptome-based cellular taxonomy. Taxonomic annotation at three levels of granularity; class (n=8), subclass (n=27), and subtype (n=65). **(b)** Spatial distribution of major cell classes. **(c)** Spatial distribution of EN subclasses. **(d)** Markers defining neuronal subclasses. **(e)** Functional enrichment of cellular subclasses using Gene Ontology Biological Process (GO BP). **(f)** Enrichment of heritable traits for cellular subclasses using previous GWAS studies (scDRS). AD[11], MS: multiple sclerosis[12], PD[13], Epilepsy: epilepsy focal[14], Migraines[15], Stroke[16], ALS: amyotrophic lateral sclerosis[17], SCZ[18], BD[19], MDD: major depressive disorder[20], ASD: autism spectrum disorder[21], ADHD: attention deficit hyperactivity disorder[22], Insomnia[23], Education: educational attainment[24], IQ: intelligence[25], Alcoholism[26], OCD: obsessive compulsive disorder[27], Tourettes: Tourette Syndrome[28], Obesity[29], T2D: type 2 diabetes mellitus[30], Cholesterol: cholesterol total[31], RArthritis: rheumatoid arthritis[32], IBD: inflammatory bowel disease[33], UC: ulcerative colitis[34].

Inter-individual variation of the human cortical transcriptome

Population-level transcriptomic variation is influenced by genetic differences among individuals, phenotypes such as age, sex, and disease status, as well as technical factors such as tissue source, dissection bias, and single-cell library preparation. We first set out to explore the determinants of the overall transcriptome variation in our sample cohort. Population-scale snRNA-seq data allowed partitioning of the gene expression variance by cellular variables (including cell type and fraction of mitochondrial and ribosomal genes), donor-level variables (including subject ID, age, sex, genetic ancestry, and diagnosis), and technical variables (including source of the tissue specimen, post-mortem interval, technical replicates, and sequencing depth) (**Fig. 3a**). Across all genes, a mean of 49.8% of the total expression variance can be attributed to variation across cell type, while inter-individual variation explains 9.8%, and unexplained residual variation accounts for 38.9%. The remaining variables explained less than 1% of the total variance on average. We prioritized several drivers of expression variation (**Fig. 3b**). Several genes with high cell type variation (including ATP8A2, GABRB3, and DOCK3) are dominated by genes differentially expressed between neuronal and non-neuronal cell types (**Fig. 3c**). As expected, genes varying across sexes were located in sex chromosomes (**Supplementary Fig. 3c**). Interestingly, genes with high variation across tissue sources were mostly mitochondrial genes (**Supplementary Fig. 3d**), possibly due to technical differences in physiological and environmental factors from dissection and handling at the respective tissue sources[35,36]. Top 3 genes with the highest variation across diagnosis (CIRBP, FGF1, and WSB1) were implicated in stress response and hypoxia. CIRBP, which also had a high variation across PMI (ranked 4th), was a gene induced in response to low temperature and hypoxia, and its expression was inversely associated with patient survival in cancer[37]. FGF1 was another gene linked to temperature stress and hypoxia[38], and WSB1 was a neuroprotective protein and regulator of many genes associated with the cellular response to hypoxia[39]. Top variable gene in PMI, HP1BP3, was chromatin organizing protein induced in hypoxic conditions[40]. This analysis examines expression variation shared across cell types, so the low variance fraction explained by diagnosis indicates the need for cell type specific analysis.

We observed that inter-individual differences explained 80.9% of the variation in ARL17B gene expression (**Fig. 3b,d**). Several adjacent genes, such as ARL17A and KANSL1, also have high inter-individual variation and are localized on the disease-associated MAPT locus (17q21.31) (**Fig. 3b**). ARL17B and KANSL1 are often found as a fusion transcript (KANSL1::ARL17B) and frequently undergo polymorphic translocation[41,42], and they have been implicated in neurological disorders such as ALS, PD, and MS[43–45]. To interrogate possible genetic causes for variation in expression, we examined normalized gene expression at the donor level, stratified by MAPT haplotypes (**Fig. 3e**, **Methods**). We observed two distinct patterns of ARL17B expression that could be potentially linked to H1 and H2 MAPT haplotypes, with lower expression linked to the H1H1 genotype. We observed stratification of haplotypes by genetic ancestry, where the H2H2 genotype was almost

222 exclusively found within EUR ancestry, consistent with previous reports[46] (**Supplementary Fig. 3f**). We
223 replicated a previous finding that the H1 haplotype is associated with PD susceptibility, and observed that the
224 H1H1 genotype increases PD risk with an odds ratio (OR) of 4.125 (**Supplementary Fig. 3g**; $P \leq 0.0273$),
225 much higher than previously reported (1.42[47] or 1.46[48]). In addition, we tested the contribution of the H1
226 haplotype to AD among non-ApoE4 carriers[49] but did not find a significant association ($P \leq 0.302$). The
227 variation in inter-individual expression is inversely correlated with genetic constraints, as measured by the
228 gnomAD Loss-of-function Observed/Expected Upper-bound Fraction (LOEUF) score (**Fig. 3f**, **Supplementary**
229 **Fig. 3h**). Genes with high inter-individual variability tend to be less constrained and are often associated with
230 the maintenance of basal cellular functions (i.e., housekeeping genes), including translation, RNA processing,
231 metabolism, signal transduction, and structural maintenance, consistent with previous findings[50,51] (**Fig. 3g**).
232

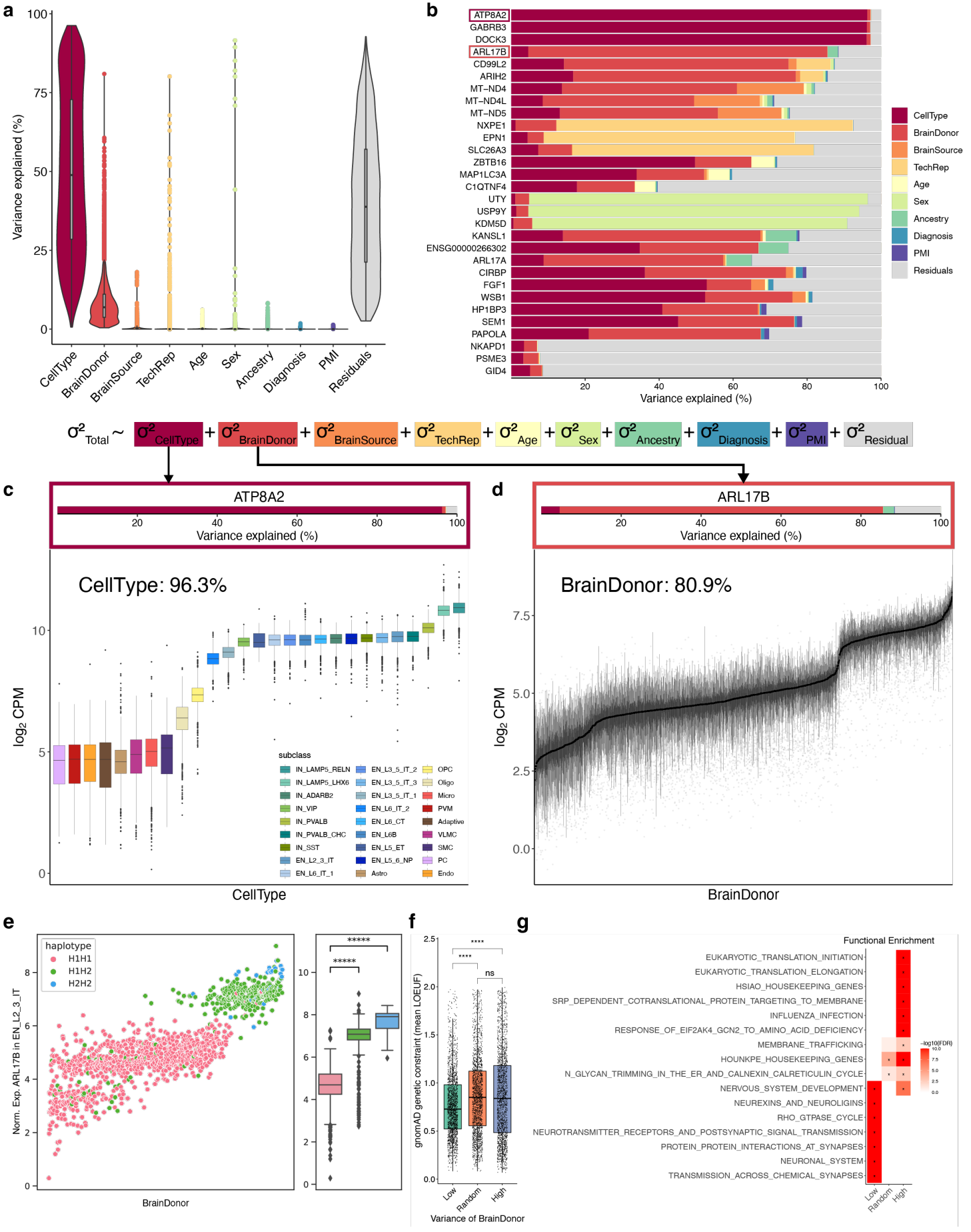


Fig. 3. **Sources of transcriptomic variation. (a)** Variance partition of the transcriptome. **(b)** Top genes for each technical and clinical covariate category. **(c)** ATP8A2 gene expression across cell types. **(d)** ARL17B gene expression across donors. **(e)** Stratification of ARL17B expression by MAPT haplotypes. **(f)** Genetic constraints of gene groups measured by average LOEUF score (upper bound of 90% confidence interval for o/e ratio for high confidence pLoF variants; lower values indicate more constrained). “Low”, indicates the bottom 2,000 genes with the lowest brain donor variance. “High”, indicates the top 2,000 genes with the highest brain donor variance. “Random”, indicates 2,000 random genes that are in neither the “Low” nor “High” category. **(g)** Functional enrichment for genes with high or low inter-individual variation.

Cross-disorder variation of cell type composition across 8 different neurodegenerative and neuropsychiatric diseases

Utilizing 318 neurotypical donors as a baseline, we systematically evaluated variation in the cellular composition of the DLPFC across eight different neurodegenerative (NDDs, including AD, DLBD, Vas, Tau, PD, and FTD) and neuropsychiatric diseases (NPDs, including SCZ and BD). Using all subclass-level cell types, we found the overall cell type composition changes were broadly stratified by NDDs and NPDs, and that they form distinct clusters (**Fig. 4a**). Notably, we observed a higher degree of similarity among AD, DLBD, and Vas. Focusing solely on neurons, we saw equal or greater correlations among the same class of diseases, underscoring the critical role of neurons in the etiology of neurological diseases. Additionally, similarities between FTD-AD (all cells) and FTD-Tau (neurons) were observed. Exploring each subclass further, we identified a notable overlap in the prevalence of neuronal and glial cell types within the same class of diseases (**Fig. 4b, Supplementary Table 3**). Specifically, we observed that NDDs were characterized by a higher abundance of non-neuronal cells, particularly vascular cell types (**Supplementary Fig. 4c**), as well as elevation of a specific IN, namely IN_LAMP5_RELN, IN_LAMP5_LHX6, and IN_ADARB2 subclasses. In contrast, NPDs were predominantly associated with an increase in neuronal cells, particularly deep layer ENs in L5-6. To further identify specific subtypes responsible for driving the compositional changes in vascular cell types, we used subtype-level annotation to analyze compositional variation in 8 NDDs and NPDs (**Fig. 4c, Supplementary Fig. 4d**). From this, we identified dominant subtypes of each subclass that further differentiated NDDs and NPDs. For example, vascular leptomeningeal cells (VLMCs) are barrier-forming fibroblasts of the brain[52], and they are transcriptionally segregated into three subtypes; two meningeal VLMCs (VLMC_DCDC2 and VLMC_SLC4A4) and one perivascular VLMC (VLMC_ABCA6). Our subtype-level analysis indicates a polarized response of meningeal VLMC_DCDC2 where their increased proportions are specifically associated with most NDDs.

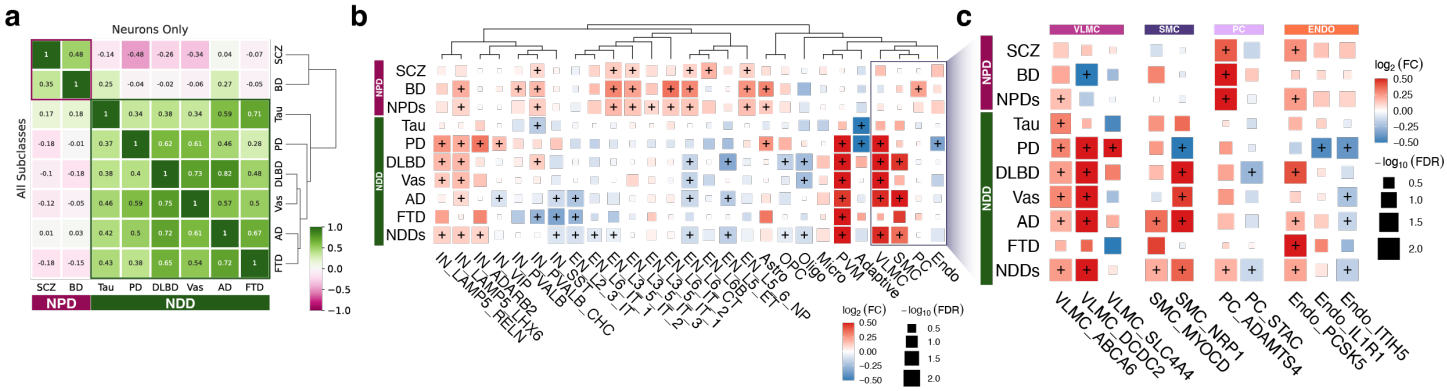


Fig. 4. **Cross-disorder variation of cell type composition comparing 8 different NDDs (AD, DLBD, Vas, Tau, PD, and FTD) and NPDs (SCZ and BD) against common neurotypical controls. (a)** Correlation of cell type composition using all cell types (bottom-left triangle) or limited to neuronal cell types (upper-right triangle). **(b)** Variation in cell type composition for each subclass in 8 different diseases. NDDs and NPDs indicate meta-analysis using broad disease categories. **(c)** Variation of subtype-level composition in the vascular cell class. Color intensity indicates effect size and dot size reflects the statistical significance of correlations.

Cross-disorder variation of gene expression and genetic concordance

Cross-disorder gene expression analysis has the potential to identify shared biological pathways and mechanisms, improve diagnostic accuracy, and help develop targeted treatments. To assess the extent of sharing between NDDs and NPDs, we performed a comprehensive analysis of differentially expressed genes (DEGs) by decomposing the total disease signatures into shared and distinct components. Using Dreamlet followed by Mashr[53], we performed composite tests to evaluate the specificity of disease effect (**Methods**), leading to the identification of shared disease signatures that are invariant across cell subclasses (**Fig. 5a, Supplementary Figs. 5a,b**). Genes sharing cross-disorder signatures encompassed crucial transcriptional processes, such as mRNA splicing and processing, and protein localization to mitochondria. The observed cross-disorder signatures, affecting genes, which are critical for the proper functioning of cellular processes, align with the omnigenic model[54], and further support the pleiotropy of those genes influencing multiple disorders, both genetically and transcriptionally. After discounting the shared cross-disorder signatures from the overall DEG expression profiles, we used the residual effects to quantify the pair-wise transcriptomic similarity between traits (**Fig. 5b, Supplementary Table 4**). Similarities between pairs of NDDs or NPDs were greater compared to the NDD-NPD contrast, with AD, DLBD, Vas, and PD being the most similar. Meta-analysis using disease-specific effects of these four transcriptomically similar traits (AD, DLBD, Vas, and PD) implicated neuronal development and synaptic signaling pathways involving interneurons (IN_LAMP5_RELN, IN_ADARB2, and IN_PVALB) as well as vasculature development from the VLMC subclass (**Fig. 5c, Supplementary Fig. 5c**).

Comparison of pairwise trait co-heritability against pairwise transcriptome similarity can identify shared genetic influences and underlying biological mechanisms, leading to better insights into disease etiology and potential therapeutic targets. We first used GWAS summary statistics to estimate the shared heritability (r_g) among NDDs and NPDs, revealing disease pairs (i.e., SCZ-BD, DLBD-PD) with a high degree of heritability overlap (**Fig. 5d**). This analysis differentiates between NDDs and NPDs, revealing lower shared heritability between NDDs and NPDs. We then quantified the pairwise transcriptome similarity (ρ_s) across NDDs and NPDs (**Fig. 5e**). Our findings revealed varying degrees of transcriptomic overlap, with some exhibiting high similarity, particularly in disease pairs such as AD and DLBD. We calculated the pair-wise trait heritability-transcriptome concordance (r_g-ρ_s) by comparing the shared heritability against the average transcriptome similarity across cell types. We observed a positive correlation between genetic and transcriptomic similarities (Spearman's ρ=0.56), indicating that diseases with higher shared genetic risk also tend to have more similar gene expression profiles (**Fig. 5f**). We extended the r_g-ρ_s comparison by considering the transcriptional concordance for each cell type (**Fig. 5g, Supplementary Fig. 5d**). Among all cell types, the transcriptional concordance of Chandelier cells (IN_PVALB_CHC) had the highest similarity with the pairwise trait heritability. In summary, our approach allowed us to dissect disease signatures into shared and distinct components, revealing significant overlap in gene expression and genetic risk across NDDs and NPDs. These findings bolster our understanding of common and unique disease mechanisms, paving the way for novel therapeutic strategies targeting shared or distinct pathways.

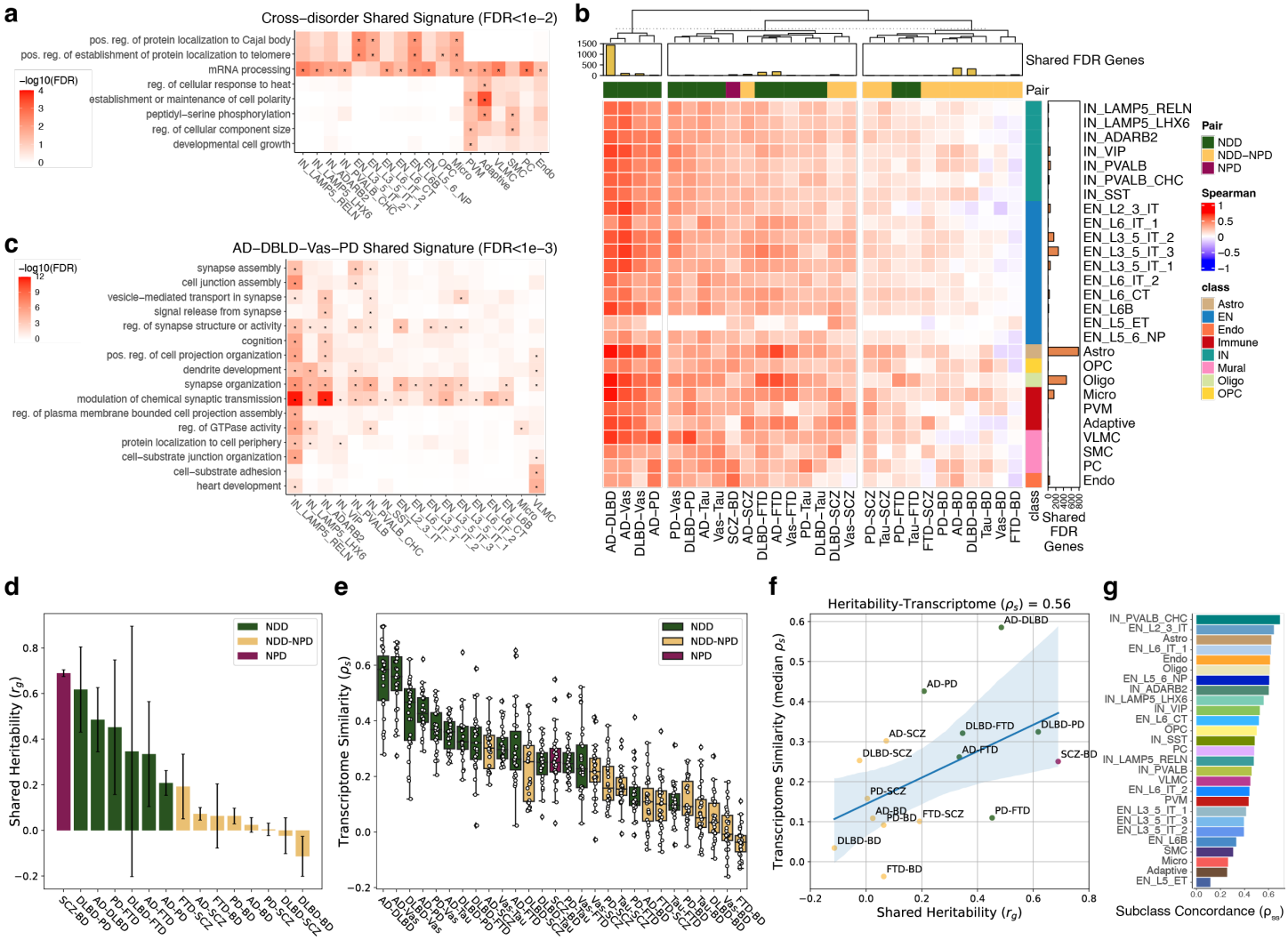


Fig. 5. **Cross-disorder variation of gene expression. (a)** Pathways implicated by shared gene expression changes across 8 disorders. Hypergeometric test with FDR ≤ 0.01 shown. **(b)** Transcriptome similarities between disease pairs. Correlation is measured using Spearman correlation. Cross-disease shared genes are discounted from the comparison. **(c)** Pathways implicated by shared signatures from AD, DLBD, Vas, and PD. Hypergeometric test with FDR ≤ 0.001 shown. **(d)** Shared heritability estimates for disease pairs. LD Score Regression (LDSC). **(e)** Transcriptome similarity measured by Spearman correlation. **(f)** Correlation between shared heritability and median of transcriptome similarity. **(g)** Correlation between shared heritability and transcriptome similarity per cell subclass.

Transcriptomic variation with AD pathology

Next, we focused on characterizing the transcriptomic variation in AD using case-control comparison, and analysis of different phenotypes that capture disease severity, including plaque density using CERAD scores, neurofibrillary tangle progression using Braak stage, and level of cognitive impairment. The analysis of variation in cell type composition reveals distinct patterns associated with AD pathology compared to normal aging (**Fig. 6a, Supplementary Fig. 6a, Supplementary Table 5**). Notably, changes unique to AD pathology, such as the observed increase in Micro and IN_LAMP5_LHX6 with higher CERAD scores, indicate a potential association between this neuronal subtype and Aβ pathology. Additionally, the loss of EN_L2_3_IT in AD (based on case-control and CERAD comparisons), which is not evident in normal aging, suggests a specific vulnerability for this neuronal subtype to AD. Placing subclasses based on compositional shifts of normal aging

and AD (**Fig. 6b**), we identified one of the vascular cell types, SMC, was having the opposite effect, suggesting an AD-specific vulnerability in SMC. Likewise, changes in most ENs were discordant suggesting the mechanisms that lead to neuronal loss are AD-specific. Taken together, our findings highlight the cell type-specificity of vulnerability associated with AD.

To better understand transcriptomic variation under AD pathology and to identify subclass most affected, we also analyzed the changes in AD cases due to co-occurring conditions. It is estimated that more than 80% of AD patients will exhibit at least one NPS over the course of their illness that significantly impacts their clinical outcomes[5,55] which suggests at least some shared molecular mechanisms between serious mental illness and AD. To evaluate cell type associations with the prevalence of NPS, we applied compositional variation analysis to three categories of NPS based on co-occurrence estimates. Notably, using age-matched groups of AD patients with or without NPS (**Supplementary Fig. 6b**), we found AD patients experiencing weight loss and PMA have an increased ratio of EN cell types, especially in deep layer neurons in L6 (**Fig. 6a, Supplementary Fig. 6a**). Our results support previous findings[56] that the deeper PFC layers (L5-6) are involved with certain types of NPS.

Upon identification of vulnerable cell subclasses in AD, we questioned their roles in AD pathology whether their changes are damaging, protective, causal, or derived. As demonstrated in previous studies[57], we employed mediation analysis to decipher the causal relationships between various cascades of events leading to disease onset and progression of AD pathology. We performed the analysis on the base hypothesis that polygenic risk (AD PRS; **Methods**) contributes to plaque accumulation (mean density of neuritic plaques), which in turn affects tau progression (Braak), leading to varying degrees of dementia. We tested 14 subclasses that were significantly altered in any of the contrasts and observed significant average causal mediation effects (ACME, $P < 0.05$) involving Microglia, VLMC, IN_SST, and IN_LAMP5_LHX6 (**Fig. 6c**). Initially, we showed PRS affects both plaque and tau pathology, but tau is also significantly affected by plaque accumulation (ACME = 0.0903, $P < 2e\text{-}16$). Then, tau progression results in more VLMC cells mediated by an increase in Microglia (ACME = 0.00477, $P = 0.005$). Conversely, an increase in microglia leads to a decrease IN_LAMP5_LHX6. Such changes in the levels of both VLMC (ACME = 0.00515, $P = 0.0002$) and IN_LAMP5_LHX6 (ACME = -0.00269, $P = 0.031$) appear to exacerbate dementia. Furthermore, plaque accumulation is mitigated by IN_SST cells (ACME = 0.1475, $P < 2e\text{-}16$). Increased microglia lowers IN_SST (coef = -0.18, $P = 5.04e\text{-}07$), as does increased tau (coef = -0.06, $P = 1.82e\text{-}04$), contributing to more plaque accumulation (coef = -2.01, $P = 7.47e\text{-}06$).

To determine how gene programs change in response to increasing severity of AD pathology, we characterized the DEGs in 27 subclasses (**Fig. 6d, Supplementary Fig. 6d, Supplementary Table 6**). DEGs in general have high concordance across different AD pathology variables, consistent with previous reports[1]. Overall, DEGs in AD (FDR < 0.05) can be summarized as up-regulation of genes in vascular cell classes (Mural and Endo) while down-regulation of genes in neurons (**Supplementary Fig. 6c**). We discovered genes differentially expressed in microglia were not observed in other subclasses, and these genes generally exhibited higher effect sizes. These include previously characterized up-regulated genes in Microglia, including DPYD, IL15, and PTPRG[53,58,59]. Further characterizing microglia-associated gene signatures, we found they are specifically enriched with pathways involved in negative regulation of cell motility and migration as well as response to lipoprotein particle (**Fig. 6e, Supplementary Fig. 6e**). Gene expression changes in neurons were largely affecting synaptic functions including synapse assembly, development, signaling, and membrane transport. Lastly, genes affecting VLMCs were implicated in muscle tissue development.

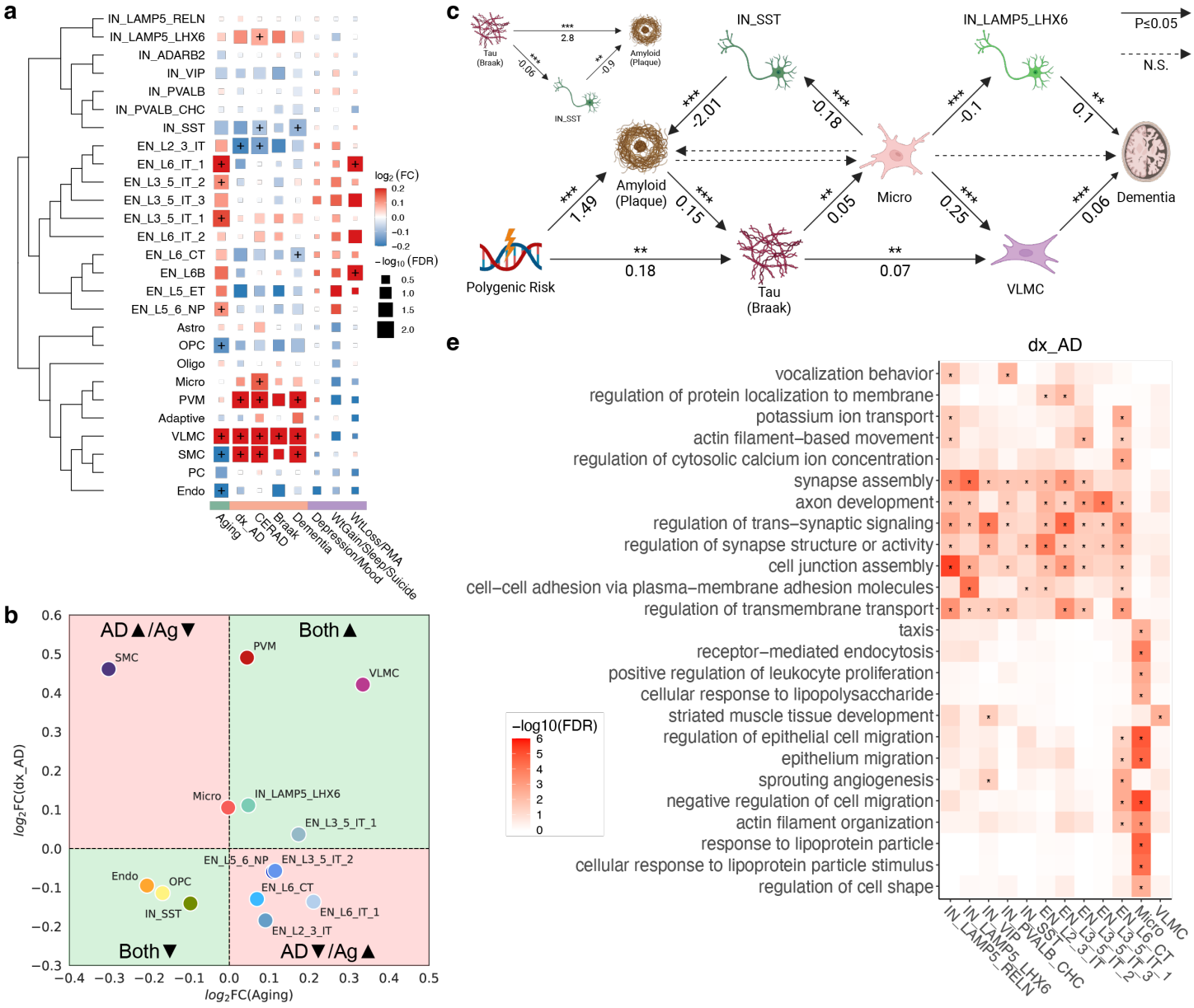


Fig. 6. **Transcriptome changes across AD neuropathology. (a)** Compositional variation analysis using normal aging, different measures of AD pathology (binary AD diagnosis (dx_AD), CERAD score, Braak staging, and ordinal dementia scale), and 3 categories of NPS within AD. **(b)** Comparison of compositional changes between aging and AD. Green indicates changes are concordant. Red indicates changes are discordant. Only subclasses with at least one FDR significant contrast shown. **(c)** Causal mediation analysis using PRS, mean plaque, Braak staging, and dementia scale. CLR-transformed subclass fractions were used for modeling. Statistical significances are shown above the arrow (P < 0.001 ***, < 0.01 **, and < 0.05 *), and numbers below indicate coefficients. Mediation effects of SST interneurons on Aβ plaque accumulation shown separately. **(d)** DEGs in AD phenotypes. Meta-analysis between brain banks. Top genes with FDR < 0.01 and effect size ≥ 0.35. **(e)** Functional enrichment analysis of DEGs by subclass using Gene Ontology Biological Process. Hypergeometric test with FDR ≤ 0.01 shown. GO terms were reduced using rrvgo.

Nonlinear dynamics of the AD pathological trajectory

To understand AD dementia mediated by tau proteinopathy, we modeled the transcriptome using variational autoencoder (VAE)-based latent manifold mapping (**Fig. 7a, Methods**). We inferred two independent cell-type-specific disease trajectories from semi-quantitative measures of AD progression (tau proteinopathy and severity of cognitive decline) and explored the dynamics of two paths leading to the onset and early stage of AD pathogenesis. To decorrelate Braak and dementia model predictions, we equally sampled all combinations of Braak stage and dementia status during model training (**Supplementary Fig. 7**). The accuracy of the Braak and dementia model predictions were significantly above chance for all eight cell classes (P < 1e-4, bootstrap) (**Fig. 7b, top**). We built a disease trajectory using the predicted Braak staging as a pseudo-temporal axis (**Supplementary Figs. 8a,b**) and calculated a dementia resilience score that measured how gene expression is correlated with predicted dementia, conditioned on the predicted Braak staging (**Methods**). Specifically, a gene is considered protective (i.e., resilient against dementia), if increased expression is associated with a decrease in predicted dementia, and vice versa (damaging if increased expression is associated with an increase in predicted dementia). Many past studies have suggested that gene expression can evolve nonlinearly with disease progression[60]. Thus, we measured the degree of nonlinearity for each gene trajectory (**Supplementary Fig. 8c**) and revealed that those within the immune and neuronal cell class were the most nonlinear and linear, respectively (both Ps < 1e-4, bootstrap) (**Fig. 7b, bottom**).

To further characterize the nonlinear evolution of gene expression, we defined an "early" and "late" disease stage (**Supplementary Fig. 8c**) and determined the top 32 GO BP pathways that best summarize the biological changes associated with early and late stages of tau proteinopathy and dementia resilience (**Fig. 7c, Supplementary Figs. 9-11, Supplementary Table 7)**. A semantic clustering of the pathways identified five functional clusters. The first cluster was primarily related to synaptic function, which were characterized by down-regulation in OPCs in the early stages of AD, and then up-regulation in neurons in later stages. For both cell types, increased expression was associated with diminished resilience (i.e., damaging). The late increase of genes implicated in these pathways could be linked to compensatory mechanisms following synaptic loss[61]. The second cluster was related to cell metabolism pathways implicated in protein translation[2,62], mitochondrial function[2,63], and acidification[64], which showed strong downregulation in neurons with increasing Braak, consistent with past studies[65]. Decreased expression in these metabolic pathways implied damaging association with decreased resilience (increase in metabolism was protective). The third cluster was related to cell stress, including pathways related to chaperone-mediated protein assembly, folding, and tau-kinase activity. In non-neuronal cells, increased cell stress was strongly associated with cognitive decline in both the early and late stages. The fourth cluster was related to immune response and inflammation. In immune cells, early increase in pathways such as microglial activation, phagocytosis, and B cell proliferation was observed, while late increase was involved with pathways implicated in adaptive immune responses such as antigen

presentation and T-cell response, all of which were associated with dementia resilience. Lastly, vascular cells were implicated in damaging changes such as reduced lipid transport and vascular permeability, response to interleukin-17 (IL-17), glycolysis, endothelial cell differentiation, and nitric oxide synthesis.

As previous studies have highlighted the importance of the immune response in AD[58,66–72], we examined this cell class in more detail (**Figs. 7d-g**); We also extended the analysis to EN, IN and OPC cells classes (**Supplementary Figs. 12-15**). To help visualize the non-linear dynamics of the immune response to AD, we categorized gene expression trajectories into four modules based on their response to increasing tau proteinopathy: early increasing, early decreasing, late increasing, and late decreasing (**Fig. 7d**). Gene enrichment of these four trajectory modules showed that pathways involved in macrophage colony stimulation and metal transport were all upregulated in the early stage (**Fig. 7e**), whereas migration, purinergic signaling, and negative regulation of inclusion body assembly were downregulated. In the later stage, lipid-related pathways such as adipose tissue development, and triglyceride metabolism increased, while synapse-related pathways were downregulated. Gene-set enrichment analysis using GWAS summary statistics for the top 250 genes in each trajectory module revealed that late increasing genes were the most strongly associated with AD (FDR ≤ 4.1e-6), although the other three modules were also significantly associated (FDR early increase ≤ 0.005, early decrease ≤ 0.048, late decrease ≤ 0.040 (**Fig. 7f, Supplementary Table 8, Supplementary Fig. 12b**). Interestingly, the late increasing trajectory module was also significantly correlated with other NDDs, including MS and PD. We further examined the genes associated with trajectory modules (**Fig. 7g,** list of top genes for all cell classes are found in **Supplementary Table 9**). Markers for homeostatic microglia[73], such as CX3CR1, NAV2 and P2RY12 were among the top 10 early decreasing genes with increasing proteinopathy (FRMD4A ranked 26th out 17,265 coding genes, **Supplementary Table 9**). Conversely, some of the early increasing genes, such as ACSL1, DPYD, and CD163 were recently implicated in a pathogenic lipid-droplet accumulation phenotype in individuals with AD who carry the APOE4/4 genotype (another implicated gene, NAMPT, ranked 54th. **Supplementary Table 9**)[70]. These genes, along with the later upregulation in lipid-related pathways such as adipose tissue development and triglyceride metabolism, lend support to the hypothesis that microglia develop a lipid-droplet accumulating state that potentially exacerbates disease progression[70–72].

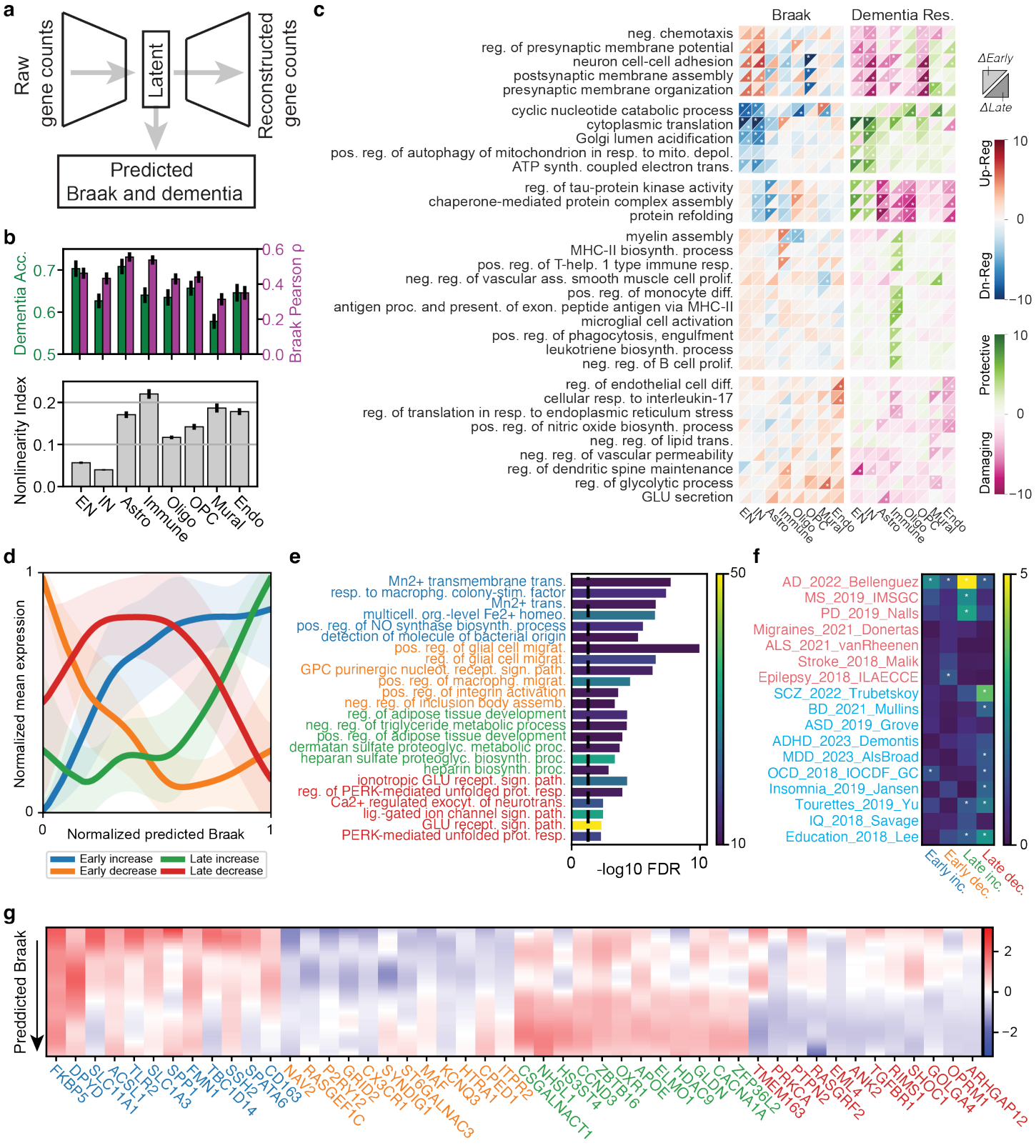


Fig. 7. **Modeling of AD using disease trajectory analysis. (a)** Overview of the trajectory inference model. **(b)** Accuracy of the model (above) and nonlinearity index by class (below). **(c)** Pathway enrichment by early (upper left triangles) and late (lower right triangles) phases of the AD trajectory, as predicted by Braak and dementia resilience. A gene is considered protective if increased expression is associated with a decrease in predicted dementia, and vice versa. Hue indicates the z-score (clipped between -10 and 10), and stars indicate FDR < 0.05. **(d)** Four representative disease trajectory modules for immune cell class. The mean normalized expression of the 250 genes with greatest early increasing (blue curve), early decreasing (orange), late increasing (green) and late decreasing (red) slopes based on Braak trajectories. **(e)** Pathway enrichment of disease trajectory modules. Text colored by the four trajectory modules. Hue indicates the number of genes in the pathway. **(f)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module. Text color indicates neurological traits (red) and psychiatric traits (blue). Hue indicates negative log10(FDR) (clipped at 5), asterisk indicates FDR < 0.05. **(g)** Top 12 genes for each trajectory module. Hue indicates the rate of normalized expression change (slope) based on the predicted Braak staging. For visualization, we only show each gene once even if it appears in more than one module.

Discussion

We report a comprehensive disease atlas of the human DLPFC using 1,494 donors affected with various complex neurological and psychiatric conditions. Our single-nucleus transcriptomic analyses provide novel insights about cellular heterogeneity and variability in the human brain, including the observation that about 10% of total transcriptomic variation can be attributed to inter-individual differences. Intriguingly, we find genes with higher inter-individual variability and lower genetic constraints that were implicated in certain housekeeping roles[51]. This is in contrast to the idea that, due to their essential nature, housekeeping genes are more likely to be conserved[74,75]. They are among the most universal genes in the cell, and cells adapt their protein synthesis machinery in response to physiological needs[76]. This flexibility is crucial for processes such as differentiation, proliferation, and response to stress. A higher degree of regulatory flexibility suggests a tolerance for variation within certain constraints.

We used the disease atlas to characterize cellular and disease-specific responses to pathologic conditions. Our findings reveal shared and distinct cellular composition profiles among NDDs and NPDs, furthering our understanding of their underlying pathophysiological mechanisms. Disease signatures shared across NDDs and NPDs are enriched with genes critical for the proper functioning of cellular processes, such as RNA splicing. This observation aligns with the omnigenic model[54], where most heritability in complex traits can be explained by effects on peripheral genes that often play indirect, subtle, and cumulative roles in disease. Thus, discounting cross-disease effects could facilitate identification of core disease-relevant functional genes. Using this approach, we found that disease pairs with higher genetic risk overlap tend to have greater transcriptomic concordance, in a cell-type-dependent manner, suggesting that genetic factors contributing to disease susceptibility can also influence transcriptomic alterations in similar ways[3].

In addition, we revealed that the brain vascular system is intricately linked to immune dysfunction in NDDs. In general, we saw a relative increase in vascular cell types in most NDDs and demonstrate that, in a seemingly protective role, levels of VLMCs rise in individuals who experience exacerbated cognitive impairment in AD. It has previously been shown that the meningeal lymphatic system plays multiple roles in the brain, including waste removal[77–79] and the adaptive immune response[52,80]. Given the role of VLMCs in these processes[81–84], they warrant further investigation in the context of NDDs.

Our analysis of the pathological trajectory of AD aligns with various proposed hypotheses[65,85–89], particularly for pathways affected in the earliest stages of the disease. Neuronal, immune, and vascular cells exhibit distinct vulnerabilities, with shared alterations across correlated pathways. In neurons, we observed a decrease in metabolic functions (cyclic nucleotide catabolic process, cytoplasmic translation, and ATP

synthesis) that have been closely linked to synaptic dysfunction and cognitive decline[65]. The immune response was generally protective against cognitive decline: an early innate immune activation[85] followed by an adaptive immune response[67–69] was associated with dementia resilience. The one exception was the response to IL-17, which was damaging in the early stages of AD for the Immune, Mural, and Endo classes. IL-17 has been associated with cognitive decline[90,91] and disruption of the blood-brain barrier[92], with anti-IL-17 treatment restoring cognitive function in mice[90,91]. In addition, chaperones have been closely associated with a number of disease related processes, including tau misfolding and aggregation, and neurotoxicity[86–89,93,94], however, their precise roles in AD pathogenesis remains unclear. Our results suggest that chaperones can play opposing roles; protective for neurons but damaging for glia and immune cells. Lastly, vascular cells are associated with negative regulation of vascular permeability and lipid transport, which are detrimental to the infiltration of perivascular immune cells[95] and the clearance of protein aggregates. Taken together, these insights deepen our understanding of AD pathogenesis and implicate cellular responses that warrant further investigation. Overall, the PsychAD single-cell disease atlas serves as a unique and foundational resource to further our understanding of population-level disease-associated transcriptomic variation in the human brain.

Methods

Collection and harmonization of clinical, pathological, and demographic metadata

Brain tissue specimens were sourced from two brain banks: the Mount Sinai NIH Neurobiobank (MSSM; 1,042 samples) and the NIMH-IRP Human Brain Collection Core (HBCC; 300 samples). Additionally, samples were obtained from five prospective cohort studies conducted at the Rush Alzheimer's Disease Center (RADC; 152 samples)[96,97]. As such, the available clinical data varied as a function of source (**Supplementary Fig. 4a**). We used the following scheme to harmonize available clinical, pathological, and demographic metadata: the CERAD scoring scheme for neuritic plaque density[98] was harmonized for consistency across multiple brain banks, where the scores range from 1 to 4, with increasing CERAD number corresponding to an increase in AD burden; 1=no neuritic plaque (normal brain), 2=sparse (possible AD), 3=moderate (probable AD), 4=frequent (definite AD). Samples from RADC used consensus summary diagnosis of no cognitive impairment (NCI), mild cognitive impairment (MCI), and dementia and its principal cause, Alzheimer's dementia[99–101]. MSSM samples used clinical dementia rating (CDR), which was based on a scale of 0-5; 0=no dementia, 0.5=questionable dementia (very mild), 1=mild dementia, 2=moderate dementia, 3=severe dementia, 4=profound dementia, 5=terminal dementia. After consulting with clinicians, we created a harmonized ordinal variable where dementia is categorized into three levels of cognitive decline, independent of AD diagnosis; 0=no cognitive impairment, 0.5=MCI (mild cognitive impairment), and 1-5=dementia. In addition to AD phenotype, we collected comprehensive demographic (age, sex, and genetic ancestry) and technical variables (tissue source, technician, sample batch, postmortem interval (PMI; measured in minutes), ApoE genotype) to describe each cohort (**Supplementary Table 1**). We briefly describe the process for assigning genetic ancestry[102]. In particular, we leveraged quadratic discriminant analysis (QDA) to infer genetic ancestry by training our model using data from the 1000 Genomes Project. We utilized 10-fold stratified cross validation to optimize the regularization parameter within QDA[103] as well as forward selection to identify the optimal number of principal components for genetic ancestry assignments. For samples without genotypic data we utilized race/ethnicity as a proxy for inferring genetic ancestry. We emphasize that while genetic ancestry is a distinct concept from the social constructs of race and ethnicity[104], we leveraged the correlated race/ethnicity variables as proxies to retain those samples in the analyses. Values for superpopulations included: EAS, SAS, AFR, AMR, EUR, and EAS_SAS, where the category "EAS_SAS" was assigned for samples with unavailable genotypes with an "Asian" value for race/ethnicity, which can potentially correspond to both EAS and SAS.

Clinical diagnosis of AD

For analysis comparing donors with AD cases and neurotypical controls, a binary clinical diagnosis variable for AD, dx_AD, was defined, as follows: Individuals with CERAD 2, 3, or 4, Braak ≥ 3, and CDR ≥ 1 for MSSM or Alzheimer's dementia for RADC were classified as AD cases. Controls were defined as individuals in *controls_neuropathological_clinical* category where CERAD={1}, Braak={0,1,2,3}, and secondary diagnosis (including dementia) is not allowed except for MCI.

Measuring AD neuropathology

For analysis comparing donors with pathologic AD, the following variables were used to measure the severity of AD neuropathology: **CERAD score**[98]. A quantitative measure of Aβ plaque density where 1 is normal, 2 is possible AD, 3 is probable AD, and 4 is definite AD[99]. **Braak AD-staging score** measuring progression of neurofibrillary tangle neuropathology (Braak & Braak-score, or BBScore). A quantitative measure of the regional patterns of neurofibrillary tangle (NFT) density across the brain, where 0 is normal and asymptomatic, 1-2 indicate initial stages where NFT begins to appear in the locus coeruleus and the transentorhinal region, 3-4 indicate progression to limbic regions, such as the hippocampus and amygdala, and 5-6 indicate NFT are widespread, affecting multiple cortical regions[105–107].

Measuring cognitive impairment

For analysis comparing donors with AD-related dementia, the following variable was used to measure the severity of cognitive impairment: **Clinical assessment of dementia.** A harmonized variable of cognitive status based on the CDR scale for MSSM or NCI, MCI, and Alzheimer's dementia for RADC. We used the three-level ordinal categories of clinical dementia to measure the severity of dementia, in which 0 indicates no dementia, 0.5 indicates minor cognitive impairment, and 1.0 indicates definite clinical dementia.

Definition of cross-disorder contrasts

For cross-disorder contrasts, we limited the analysis to any individual with Age ≥ 17. Neurotypical controls are defined as any individual CERAD={1}, Braak={0,1,2,3}, and secondary diagnosis is not allowed. AD is any individual with CERAD={2,3,4}, Braak={3,4,5,6}, clinically diagnosed as dementia, and secondary diagnosis not allowed. SCZ is any individual with SCZ diagnosis (*SCZ* | *Schizoaffective_bipolar* | *Schizoaffective_depressive*) and secondary diagnosis not allowed, except for metabolic and eating disorders. DLBD is any individual with DLBD diagnosis (*DLBD*) and secondary diagnosis can be only AD. Vascular is any individual with Vascular diagnosis (*Vascular*) and secondary diagnosis can be only AD. BD is any individual with BD diagnosis (*BD_unspecific* | *BD_I* | *BD_II* | *Schizoaffective_bipolar*) and secondary diagnosis not allowed except for metabolic and eating disorders. Tauopathy is any individual with CERAD={1}, Braak={4,5,6} and secondary diagnosis allowed. PD is any individual with PD diagnosis (*PD* | *PD_uncertain_plus_encephalitic*) and secondary diagnosis can be only AD. FTD is any individual with FTD diagnosis (*FTD*) and secondary diagnosis can be only AD. All disease contrasts used in the study can be found in **Supplementary Table 1**.

Isolation and fluorescence-activated nuclear sorting (FANS) of nuclei from frozen brain specimens with hashing

All buffers were supplemented with RNAse inhibitors (Takara). 25mg of frozen postmortem human brain tissue was homogenized in cold lysis buffer (0.32M Sucrose, 5 mM $CaCl_2$, 3 mM Magnesium acetate, 0.1 mM, EDTA, 10 mM Tris-HCl, pH8, 1 mM DTT, 0.1% Triton X-100) and filtered through a 40 µm cell strainer. The flow-through was underlaid with sucrose solution (1.8 M Sucrose, 3 mM Magnesium acetate, 1 mM DTT, 10 mM Tris-HCl, pH8) and centrifuged at 107,000 xg for 1 hour at 4˚C. Pellets were resuspended in PBS

supplemented with 0.5% bovine serum albumin (BSA). 6 samples were processed in parallel. Up to 2M nuclei from each sample were pelleted at 500 xg for 5 minutes at 4˚C. Nuclei were re-suspended in 100 µl staining buffer (2% BSA, 0.02% Tween-20 in PBS) and incubated with 1 µg of a uniuqe TotalSeq-A nuclear hashing antibody (Biolegend) for 30 min at 4˚C. Prior to FANS, volumes were brought up to 250 µl with PBS and 7-Aminoactinomycin D (7-AAD) (Invitrogen) added according to the manufacturer's instructions. 7-AAD positive nuclei were sorted into tubes pre-coated with 5% BSA using a FACSAria flow cytometer (BD Biosciences).

snRNA-seq and hashing library preparation

Following FANS, nuclei were subjected to 2 washes in 200 µl staining buffer, after which they were re-suspended in 15 µl PBS and quantified (Countess II, Life Technologies). Concentrations were normalized and equal amounts of differentially hash-tagged nuclei were pooled. Using 10x Genomics single cell 3' v3.1 reagents (10x Genomics), 60,000 (10,000 per donor) nuclei were run in each of x2 10x Genomics lanes to create a technical replicate. At the cDNA amplification step (step 2.2) during library preparation, 1 µl of 2 µm HTO cDNA PCR "additive" primer was added[108]. After cDNA amplification, supernatant from 0.6x SPRI selection was retained for HTO library generation. cDNA libraries were prepared according to the 10x Genomics protocol. HTO libraries were prepared as previously described[108]. cDNA and HTO libraries were sequenced at NYGC using the Novaseq platform (Illumina).

Processing of snRNA-seq data

Alignment. Paired-end snRNA-seq library reads were aligned to the hg38 reference genome using STAR solo[109,110] and sample pools were demultiplexed using genotype matching via vireoSNP[111]. After per-library count matrices were generated, the downstream processing was performed using Pegasus v1.7.0[112] and scanpy v1.9.1[113]. **QC**. We applied rigorous three-step QC to remove ambient RNA and retain high quality nuclei for subsequent downstream analysis. First, QC was applied at the cell level. Poor-quality nuclei were detected by thresholding based on UMI counts, gene counts, and mitochondrial content. We also checked for possible contamination from ambient RNA, fraction of reads mapped to non-mRNAs, like rRNA, sRNA, and pseudogenes, as well as known confounding features such as lncRNA MALAT1. Second, QC was applied at the feature level by removing features that were not robustly expressed by at least 0.05% of the nuclei. Lastly, QC was applied at the donor level by removing donors with very low nuclei counts, which can introduce noise to downstream analyses. We also removed donors with low genotype concordance. Further filtering was carried out by removing doublets using the Scrublet method[114]. **Batch correction**. We assessed the correlation between all pairs of technical variables using Canonical Correlation Analysis and used the Harmony method[115] to regress out unwanted variables such as the effect of brain tissue source.

Defining cellular taxonomy using iterative clustering

Cellular taxonomy was defined using a divide-and-conquer strategy. From the full dataset containing over 6 million nuclei, 8 major cell classes were defined using the following steps. We selected 6,000 highly variable genes (HVGs) from mean and dispersions trends[116] using the default parameters (min_mean=0.0125, max_mean=3, min_disp=0.5) and brain sources as batch variable after manually excluding sex and mitochondrial chromosomes and MT. We used the k-nearest-neighbor (kNN) graph calculated on the basis of harmony-corrected PCA embedding space to cluster nuclei of the same cell type using Leiden[117] clustering algorithms. We used UMAP[118] to visualize the resulting clusters. From the class-level clusters, we subsetted the data by each class. Re-calculating HVGs among cells in the same class allowed us to re-focus on a feature space that is more relevant for the same class of cells. We then calculated kNN graph on the basis of the harmony-corrected PCA of the selected HVGs. Leiden-clustering was used to annotate subclass-level annotations. We iterated to the second level of taxonomy yielding 67 subtypes of human brain cells.

Spatial validation of cellular taxonomy

Xenium *in situ* panel selection and custom panel design. Xenium Human Brain Gene Expression Panel (1000599, 10x Genomics) and a custom panel of 100 genes (**Supplementary Table 10**) were selected for the Xenium experiment. The 100 gene custom panel consisted mainly of subclass markers selected based on specificity and gene expression level. The custom gene list was sent to 10X genomics and the probe design was performed using their in-house pipeline.

Tissue preparation. Fresh frozen tissue specimens of DLPFC were dissected into small blocks on ice. Tissue blocks were snap frozen by submerging in an isopentane (320404-1L, Sigma-Aldrich) bath chilled with dry ice and stored in -80 °C. Before cryosectioning, tissue blocks were allowed to equilibrate to the cryostat (HM505, Microm) chamber temperature, and were mounted with OCT (Tissue-Tek® O.C.T. Compound, 4583, Sakura Finetek USA). After trimming, good quality 10 µm sections were flattened on the cryostat stage and placed on pre-equilibrated Xenium slides (Xenium Slides & Sample Prep Reagents, 1000460, 10x Genomics). 2-3 sections were placed on each slide. Sections were further adhered by placing a finger on the backside of the slide for a few seconds and were then refrozen in the cryostat chamber. Slides were sealed in 50 ml tubes and stored at -80 °C until Xenium sample preparation.

Sample preparation. Xenium sample preparation was performed according to the manufacturer's protocol; "*Xenium In Situ for Fresh Frozen Tissues – Fixation & Permeabilization, CG000581, Rev C*" and *"Xenium In Situ Gene Expression - Probe Hybridization, Ligation & Amplification, User Guide, CG000582, Rev C"*. Briefly, fresh frozen sections mounted on Xenium slides from the previous step were removed from -80 °C storage on dry ice prior to incubation at 37 °C for 1 min. Samples were then fixed in 4% paraformaldehyde (Formaldehyde 16% in aqueous solution, 100503-917, VWR) in PBS for 30 min. After rinsing in PBS, the samples were permeabilized in 1% SDS (sodium dodecyl sulfate solution) for 2 min, and then rinsed in PBS before being immersed in the pre-chilled 70% methanol and incubated for 60 min on ice. After rinsing the samples in PBS, the Xenium Cassettes were assembled on the slides. Samples were incubated with a probe hybridization mix containing both the Xenium Human Brain Gene Expression Panel (1000599, 10x Genomics) and a 100 custom gene panel at 50°C overnight to allow the probes to hybridize to targeted mRNAs. After probe hybridization, samples were rinsed with PBST, and incubated with Xenium Post Hybridization Wash Buffer at 37°C for 30 min. Samples were then rinsed with PBST and ligation mix was added. Ligation was performed at 37°C for 2 hrs to circularize the hybridized probes. After rinsing the samples with PBST, Amplification Master Mix was added to enzymatically amplify the circularized probes at 30 °C for 2 hrs. After washing with TE buffer, auto-fluorescence was quenched according to the manufacturer's protocol and nuclei stained with DAPI prior to Xenium *in situ* analysis.

Data processing. The prepared samples were loaded into the Xenium analyzer and run according to manufacturer's instructions "*Xenium Analyzer User Guide CG000584 Rev B*". After the Xenium analyzer was initiated, the correct gene panel was chosen, and decoding consumables (Xenium Decoding Consumables, PN-1000487, 10x Genomics) and reagents (Xenium Decoding Reagents, PN-1000461, 10x Genomics) were loaded. The bottom of the slides were carefully cleaned with ethanol prior to loading. Once the samples were loaded and the run was initiated, the instrument scanned the whole sample area of the slides using the DAPI channel, and regions of interest were selected to maximize the capture area. Results were generated by the instrument using default settings. By default, the Xenium analyzer uses 15 µm nuclei expansion distance for segmentation of cells. To test the idea of nuclei only segmentation, we resegment the results with 0 µm nuclei expansion, by using the Xenium ranger and the following scripts:

```
xeniumranger resegment --id=demo --xenium-bundle=/path/to/xenium/files --
expansion-distance=0 --resegment-nuclei=True
```

Major cell type identification. After nuclei were segmented, cell x gene matrices were generated from the overlap of each segmented nuclear boundary with detected transcripts in the Xenium experiment. Nuclei were subsequently filtered by the number of detected transcripts, and only those containing at least 40 nuclear transcripts were retained for downstream analysis. Gene expression data from each sample was then log-normalized and normalized data were used for PCA, kNN graph calculation and Leiden clustering. Clusters were then assigned to one of 8 major cell types based on marker gene expression.

Label prediction from snRNA-seq data using scANVI. As an alternative to marker-based annotation, we used scANVI[119] to perform reference-based label transfer from the RADC dataset. Briefly, we followed the following steps. First, snRNA-seq gene expression data was subset to the genes shared with the Xenium gene panel. Next, we used the scvi-tools package[120,121] to train machine learning models for dimensionality reduction based on the reference dataset and its assigned labels (e.g., class and subclass). Models were run with 5 layers and 30 latent variables, and the scANVI model was trained for 20 epochs with a minimal sample of 100 cells per cluster per epoch. Lastly, a transfer model was trained for 100 epochs and applied to query data to assign labels based on those the model was trained on from the reference. To assess the performance of each transfer model, we asked the model to predict labels in the reference data (using the subset gene pool) and evaluated the rate of correct prediction and biases in label misassignment for each predicted label.

Subclass label transfer for EN (IN) cells. To assign subclass labels for EN and IN nuclei, nuclei from all samples were filtered to EN (IN) based on two alternative methods for major cell type prediction - that described above, and label transfer using scANVI (with the RADC dataset as a reference). After subsetting to nuclei labeled as EN (IN) by both methods, these were then used as a query for a second scANVI label transfer - this one trained on subclass labels. Accuracy for both major cell type and subclass models assessed by predicted labels in the RADC dataset based only on the Xenium gene panel was estimated at > 98%.

Processing of genotypes

DNA extraction and genotyping was performed as described previously[122]. In brief, genomic DNA was extracted from frozen brain tissue using the QIAamp DNA Mini Kit (Qiagen), according to the manufacturer's instructions. Samples were genotyped using the Infinium Psych Chip Array (Illumina) at the Mount Sinai Sequencing Core. Pre-imputation processing consisted of running the quality control script HRC-1000G-check-bim.pl from the McCarthy Lab Group (https://www.well.ox.ac.uk/~wrayner/tools/), using the Trans-Omics for Precision Medicine (TOPMed)[123]. Genotypes were then phased and imputed on the TOPMed Imputation Server (https://imputation.biodatacatalyst.nhlbi.nih.gov). Samples with a mismatch between one's self-reported and genetically inferred sex, suspected sex chromosome aneuploidies, high relatedness as defined by the KING kinship coefficient[124] (KING > 0.177), and outlier heterozygosity (+/- 3SD from mean) were removed. Additionally, samples with a sample-level missingness > 0.05 were removed and calculated within a subset of high-quality variants (variant-level missingness ≤ 0.02).

For ancestry assignment, genotypes were first merged with GRCh38 v2a 1000 Genomes Project data (https://wellcomeopenresearch.org/articles/4-50)[125] using BCFtools version 1.9[126]. PLINK 2.0[127] was then used to calculate the merged genotypes' principal components (PCs), following filtering (minor allele frequency (MAF) ≥ 0.01, Hardy-Weinberg equilibrium (HWE) $P \geq 1 \times 10^{-10}$, variant-level missingness ≤ 0.01, regions with high linkage disequilibrium (LD) removed) and LD pruning (window size = 1000 kb, step size = 10, r2 = 0.2) steps. For the samples of EUR ancestry assigned using the QDA method, autosomal bi-allelic variants with an imputation $R^2 > 0.8$, HWE $P \geq 1 \times 10^{-6}$, and variant-level missingness ≤ 0.02 were retained. Genotypes were

718 then annotated with ancestry-specific MAF values from the National Center for Biotechnology Information's
719 Allele Frequency Aggregator (ALFA) (https://ftp.ncbi.nih.gov/snp/population_frequency/latest_release/). Only
720 variants with an ancestry-specific ALFA MAF ≥ 0.01 were retained.
721
722 **Polygenic risk score calculation**
723 Polygenic risk scores (PRS) were computed for the PsychAD cohort using summary statistics from AD
724 GWAS[11]. The PRS-CS-auto method[128] was employed, which incorporates continuous shrinkage priors to
725 adjust the effect sizes from these summary statistics. An LD reference panel from the developers of PRS-CS,
726 based on data from the 1000 Genomes Project[125], was used (https://github.com/getian107/PRScs). The
727 default settings for PRS-CS were applied, including parameters a = 1 and b = 0.5 for the γ-γ prior, 1000
728 Markov Chain Monte Carlo (MCMC) iterations, 500 burn-in iterations, and a thinning factor of 5. The global
729 shrinkage parameter phi was determined using a fully Bayesian method. PLINK 2.0[127] was utilized to calculate
730 the individual-level PRS.
731
732 **Genetic heritability analysis of polygenic risk**
733 We established a standardized pipeline for Multi-marker Analysis of GenoMic Annotation (MAGMA)
734 followed by single-cell Disease-Relevance Scoring (scDRS). MAGMA incorporates the association P-values of
735 genetic variants from the latest genome-wide association study (GWAS). We used the following GWAS
736 summary stats in scDRS/MAGMA pipeline: AD[11], MS[12], PD[13], Epilepsy[14], Migraines[15], Stroke[16], ALS[17], SCZ[18],
737 BD[19], MDD[20], ASD[21], ADHD[22], Insomnia[23], Education[24], IQ[25], Alcoholism[26], OCD[27], Tourettes[28], Obesity[29],
738 T2D[30], Cholesterol[31], RArthritis[32], IBD[33], UC[34]. We applied MAGMA using a standard window of 35 kbp
739 upstream and 10 kbp downstream around the gene body. We executed scDRS using the top 1000 gene
740 weights, sorted by Z score. The MAGMA and scDRS pipeline was run using the following parameters. MAGMA
741 was run using `--snp-loc g1000_eur.bim` (SNP location file corresponding to the Phase 3 1000 Genome
742 Project) and `--gene-loc NCBI38.gene.loc` (gene location file from NCBI build 38). Both files were
743 obtained from https://ctg.cncr.nl/software/magma. For `scDRS`, default setting was applied.
744
745 **Variance partition analysis of gene expression**
746 After aggregating pseudobulk by library, assays (cell types) were stacked using the *StackedAssay* function
747 of Dreamlet. The resulting pseudobulk allowed us to perform analysis across cell types. Variance partition
748 analysis was performed on the resulting stacked pseudobulk. We used the following regression formula:
749
750 `Gene expression ~ (1|stackedAssay) + (1|Channel) + (1|SubID) + (1|Source) +`
751 `(1|Ethnicity) + dx_bit + scale(Age) + Sex + scale(PMI) + log(n_genes) +`
752 `percent_mito + mito_genes + ribo_genes + mito_ribo`
753
754 where *dx_bit* indicates binary disease status excluding metabolic and eating disorders. Technical covariates
755 *log(n_genes)*, *percent_mito*, *mito_genes*, *ribo_genes*, and *mito_ribo* were removed from the plotting and
756 subsequent analysis because they explained less than 1e-4 percent of overall gene expression variation.
757
758 **MAPT locus haplotyping**
759 From our harmonized genotype calls, we selected common variants in 17q21.31 locus (chr17:45307631-
760 46836264), performed PCA analysis of genotypes using 10 PCs, and used K-means clustering with k=3 to call
761 three genotype clusters, H1H1, H1H2, and H2H2. We additionally confirmed the H1 haplotype using two
762 published SNPs, rs17763050 and rs8070723, known to associate[129]. Haplotypes were estimated using Beagle
763 v5.4[130] on the selected genotypes of the 17q21.31 region. The estimation of the initial haplotype frequency

model converged after one burn-in iteration, and the estimate of the genotype phase converged after 23 phasing iterations. For testing association with PD diagnosis, we used logistic regression with age, sex, 10 genotype PCs, and H1H1 status:

```
PD ~ Age + Sex + Source + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 +
PC10 + H1H1
```

In addition, we tested the contribution of the H1 haplotype to AD among non-ApoE4 carriers[49] but did not find a significant association ($P \leq 0.302$). For testing association with AD diagnosis, we first subsetted for individuals who are not carriers of the ApoE4 allele and tested for AD association using logistic regression with the formula:

```
AD ~ Age + Sex + Source + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 +
PC10 + H1H1
```

Compositional variation analysis using Crumblr

We applied the Crumblr method (https://diseaseneurogenomics.github.io/crumblr) for testing the variation of cell type composition[131]. In summary, Crumblr scales the cell count ratio (i.e., fractions) data using centered log-ratio (CLR) transformation and applies linear models. Since CLR-transformed data is still highly heteroskedastic, the precision of measurements varies widely. Crumblr uses a fast asymptotic normal approximation of CLR-transformed counts from a Dirichlet-multinomial distribution to model the sampling variance of the transformed counts. Crumblr enables incorporating the sampling variance as precision weights to linear (mixed) models in order to increase power and control the false positive rate. Crumblr also uses a variance stabilizing transform based on the precision weights to improve the performance of PCA and clustering. Hypothesis testing was computed using the following formula:

```
Cell composition ~ scale(Age) + Sex + (phenotype of interest)
```

By including these variables, we account for potential confounders and improve the accuracy and reliability of our hypothesis testing (**Supplementary Fig. 4b**).

Differential gene expression analysis using Dreamlet

Due to the increased variable complexity in a large-scale disease atlas, scaling single-cell based approaches to millions of cells across a wide range of phenotypes presents computational challenges[132] and can be suboptimal[133–136]. To account for the scale of these data, complex study designs with repeated measures, and low read count per cell, we applied Dreamlet for differential expression analysis, which applies a pseudobulk approach. Building from the previously developed statistical tool Dream[137], it applies linear mixed models to the differential expression problem in single-cell omics data. It starts by aggregating cells by the donor using a pseudobulk approach[133,134] and fits a regression model and cell. For each feature and cell cluster, the following mixed model was applied:

```
Gene expression ~ scale(Age) + Sex + scale(PMI) + log(n_genes) + percent_mito +
mito_genes + mito_ribo + ribo_genes + (phenotype of interest)
```

where categorical and numerical variables were modeled as random and fixed effects, respectively. If the phenotype of interest was a categorical variable, we set the intercept as 0 and used pre-defined contrasts between two factors. We ran a gene set analysis using the full spectrum of gene-level t-statistics[138].

Meta-Analysis (between brain sources):

We conducted a meta-analysis to integrate results from different brain banks for the same disorder. Data tables from multiple brain banks were combined into a single list for each disorder and annotated with their respective sources. The meta_analysis function was used to perform the meta-analysis, which involved combining data tables into a single data frame, grouping the data by assay, and calculating standard errors using the formula `abs(logFC / t)`. The meta-analysis was performed using the `rma` function from the `metafor` package with a fixed effects model. P-values were adjusted using the False Discovery Rate (FDR) method, and the negative log10 of the FDR values were calculated. This method was applied to datasets of AD, DLBD, Vas, PD, Tau, FTD, SCZ, and BD.

Meta-Analysis (across the same disease category)

To further synthesize findings across multiple disorders, a `meta-of-meta` analysis was conducted, grouping the disorders into neurodegenerative and neuropsychiatric categories. Results from the initial meta-analyses for each disorder were combined into lists based on their categories. The `meta-of-meta` function was used to perform this higher-level analysis. This function combined the meta-analysis results into a single data frame, grouped the data by assay, and calculated standard errors using the formula `abs(estimate / statistic)`. This approach was applied to create `meta-of-meta` analyses for all disorders, neurodegenerative disorders, and neuropsychiatric disorders.

Evaluation of shared disease signatures using Mashr

Suppose the total disease signature can be written as the sum of shared and distinct components of DEGs for that disease:

$$\Delta\tau_k = \sum_{i \in G_{\text{shared}}} \Delta\tau_{i,k} + \sum_{j \in G_{\text{distinct}}} \Delta\tau_{j,k}$$

where $\Delta\tau_k$ denotes the total disease signature for a disease k, G_{shared} denotes the set of DEGs that are shared among the diseases, G_{distinct} denotes the set of DEGs that are distinct for diseases. This notation highlights that the total differential gene expression for each disease is composed of contributions from both shared genes (common to all diseases) and distinct genes (specific to each disease). To define shared disease signatures, we performed composite tests on Dreamlet results using Mashr to evaluate the specificity of an effect. We evaluated the posterior probability of a non-zero effect present in all 8 cross-disorder contrasts. We categorized a gene as part of the shared component if it had Mashr posterior probability greater than 0.01.

Construction of Correlation Matrix

To calculate the correlation matrix, we used a systematic approach to quantify the relationships between genetic estimates across different neuropsychiatric and neurodegenerative disorders. Spearman correlation coefficients were calculated to assess the strength and direction of association between genetic estimates across different disorders. The calculation was performed for the common to each pair of disorders, grouped by assay, after the exclusion of the shared disease signatures.

The wide-format correlation matrix was converted to a matrix suitable for heatmap visualization. Missing values were replaced with zeros. Annotations indicating the number of significant genes were added to the

rows and columns using the `rowAnnotation` and `HeatmapAnnotation` functions from the `ComplexHeatmap` package. The heatmap was generated with hierarchical clustering of both rows and columns, and colored based on the Spearman correlation values using a gradient from blue (negative correlation) to red (positive correlation).

Co-heritability analysis

We employed cross-trait LD score regression using the LDSC tool[139] to estimate the genetic correlation between a pair of traits. We used summary stats for the following GWAS traits (AD[11], DLBD[140], PD[13], FTD[141], SCZ[18], BD[19]) and calculated heritability for each of the traits and the genetic covariance and correlation between each of the pair of traits (in total 15 pairs of traits). The size of the cohort was provided to the function munge_sumstats.py for heritability estimates. Precomputed LD scores for 1000 Genomes EUR data were downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2. The SNP list for `munge_sumstats.py` was downloaded from https://data.broadinstitute.org/alkesgroup/LDSCORE/w_hm3.snplist.bz2. Standard error was obtained from the LDSC output. Script `munge_sumstats.py` was modified to include the parameter: `--chunksize 5e5`.

For each of the 15 possible combinations of traits, we calculated the level of correlation of gene expression using the Spearman rank correlation test. Genes were selected by applying the following criteria: logFC ≥ 0.5, FDR ≥ 0.05. Co-expression coefficient was calculated for the overall dataset and for each of the cell types. Next, to correlate co-expression and co-heritability, we calculated Spearman rank correlation coefficient between LDSC genetic correlation score and co-expression coefficient using 15 possible combinations of traits as data points for the Spearman rank correlation. Spearman was calculated for the overall dataset providing the genetic estimates of the expression similarities in the PsychAD cohort, and also per cell type to obtain the ranking of the cell types that contribute the most to the genetic to transcriptomic similarity in PsychAD.

Mediation analysis

Causal Mediation Analysis was performed using two R packages; mediation and psych. Results were cross-validated between the two methods (identical within a threshold) to ensure the estimated coefficients, and the mediation effects are statistically robust. From 696 individuals in AD contrast, we subsetted 645 individuals with European ancestry who have PRS calculations from the latest AD GWAS[11]. For each regression, we used the following covariates:

```
Age + Sex + PMI + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + PC7 + PC8 + PC9 + PC10
```

where PC1-10 indicate genotype PCs. For subclass proportion, we used CLR-transformed cell count fractions from crumblr analysis. For bootstrapping, we used 10,000 simulations with 50th percentile of the treatment variable used as the control condition and 90th percentile of the treatment variable used as the treatment condition.

Trajectory analysis using VAE model

Rationale. Traditionally, changes in gene expression as a function of disease state are measured using linear-based models. This approach has proven highly valuable and has enhanced our understanding of the biological mechanisms underlying many diseases. However, it is increasingly recognized that changes in gene expression can be highly nonlinear[60]; the interaction among numerous signaling pathways, many involving multiple feedback loops, can lead to complex dynamics that linear models may fail to capture. One approach to capturing potentially nonlinear changes in gene expression is pseudotime analysis (i.e., trajectory inference)[142,143]. In this approach, cells are assigned a relative pseudotime based on a metric that measures

the distance between gene expression vectors, coupled with specific assumptions about how trajectories can evolve. However, this approach requires prior assumptions that might obscure informative aspects of disease progression. For example, most methods require a dimensionality reduction operation (e.g., PCA or UMAP) before calculating each cell's nearest neighbors. The variance of lowly expressed genes, regardless of their importance to the disease, can be overshadowed by the variance of genes with greater expression but less relevance, or by the variance inherent in a sample from a highly diverse demographic. Furthermore, in the case of AD, it is known that the spread of neurofibrillary tangles and Aβ plaques is strongly, but not perfectly, correlated with dementia. Of special interest is understanding cases where individuals are resilient to dementia despite a high neurofibrillary tangle or Aβ plaque burden. Thus, we aimed to disentangle changes in gene expression associated with disease burden from those associated with the onset of dementia. Standard pseudotime analysis does not allow us to separate these two covariates effectively. To capture the potentially nonlinear gene expression dynamics during the course of AD progression and to disentangle the effects of disease burden and dementia, we used an alternative approach. As detailed below, we employed a VAE to predict the Braak stage and dementia status from the raw transcript counts of single cells. The model's predictions of the Braak stage and dementia status were then used as two independent pseudotime axes. Importantly, we trained the model by equally sampling all combinations of Braak stage and dementia status, thereby discouraging the model from learning spurious correlations between the two target variables.

Model architecture. We used a VAE, based on the scVI model[144], to predict the Braak stage and dementia status from single-cell gene counts. Both the gene encoder and decoder contained two 512-dimensional hidden layers. All hidden layers applied the ReLU activation function before a LayerNorm operation. Model predictions for Braak and dementia were generated using linear functions from the 32-dimensional latent layer.
Model training. Single-cell gene counts were log1p transformed in the input layer, and top 25,000 genes for each cell class, based on the percentage of cells the gene was expressed, were used for training. Genes found in the X or Y chromosomes were excluded to discourage the model from learning sex specific differences**.** The network was trained to minimize 1) the gene reconstruction loss, 2) the Kullback-Leiber divergence, and 3) the disease target prediction (Braak and dementia) loss:
$\mathcal{L} = \alpha(\mathcal{L}_{gene} + \mathcal{L}_{KL}) + \mathcal{L}_{target}$.
The scalar α was calibrated to properly weigh the contribution of the standard VAE loss terms (i.e., the gene reconstructions and KL losses) and the target and covariate prediction loss (see below). The gene reconstruction loss, L_{gene}, was the zero-inflated negative binomial[144]. For the binary target dementia, the softmax function was applied to the output, and the loss was the cross-entropy. For Braak, target values were first normalized to zero mean and unit standard deviation, and the mean-squared error loss was used. All loss terms were trained simultaneously. To prevent the model from overfitting the data, we applied a dropout with 0.25 probability in the input layer (after the log1p transform) and with probability of 0.5 to all hidden layers (after the ReLU and LayerNorm operations). For each cell class, we divided the cells into 10 splits. In each split, 90% of the cells were used for training, and the remaining 10% were used for inference. Within each split, cells from a single donor exclusively belonged to either the training set or the inference set, but never both. Thus, model predictions were always based on cross-validated data from different donors. We trained one model for each of the ten splits to generate predictions for all donors for that cell class. Models were trained using all donors from the MSSM and RADC brain sources, and then ran inference on only the 696 donors that focused on the AD phenotype contrast (**Fig. 1d**). There existed different numbers of cells for each cell class, thus the amount of training differed between classes. For neurons, astrocytes, oligodendrocytes and the immune cell class, we trained for 5 epochs, and then calculated the Braak and dementia model predictions by averaging the inference outputs generated from the last two epochs. For mural and endothelial cells, which contained less data, we trained for 15 and 20 epochs, respectively, and calculated the Braak and dementia model predictions by averaging the outputs generated from the last five epochs. Model accuracy for the OPC cell class evolved

more slowly over model training; thus, we also trained this cell class for 15 epochs, and calculated the Braak and dementia model predictions by averaging the outputs generated from the last five epochs. We used empirical testing to determine which values of α generated accurate Braak and dementia model predictions. Immune cells typically have reduced mean gene counts compared to the other cell classes, and we found that setting $\alpha = 0.005$. For all other cell classes, in which mean gene counts were on average greater, we set $\alpha = 0.002$. Network parameters were trained using the AdamW optimizer[145] with parameters$\beta_1 = 0.9$, $\beta_2 = 0.998$, $\epsilon = 1e-7$) and weight decay of 0.05 applied to all layers before the latent layer. We used a batch size of 512, and a learning rate schedule with a warmup and decay period using the formula $lrate = d^{-0.5} \cdot \min(n^{-0.5}, n \cdot warmup^{-1.5})$, where we set $d = 5e-4$, $warmup = 2000$, while n is the current training step. Finally, we clipped the gradient norm to 0.25 to stabilize training.

Model accuracy. We averaged the cell-level Braak and dementia model predictions to obtain donor-averaged scores. For dementia, we calculated the balanced classification score by determining the percentage of donors without dementia with a prediction score ≤ 0.5, the percentage of donors with dementia with a prediction score > 0.5, and then averaging these two values. For Braak, we calculated the Pearson R value between the actual Braak stage and the Braak predictions. Error bars were generated using a bootstrap procedure, in which we randomly sampled donors (with replacement), calculated the Braak and dementia prediction accuracy, and repeated this process 20,000 times. For each cell class, we included all donors with at least 5 cells.

Disentangling Braak and dementia. Braak stage and dementia status are significantly correlated (Pearson R = 0.582, P < 1e-60). This strong correlation between target variables implies that input features (i.e., changes in gene expression) associated with the two target variables are also likely to be correlated, which makes it challenging for the model to learn which input feature is predictive of which target variable. The result is when the model is trained in a standard fashion, the correlation between the Braak and dementia model predictions are close to 1 (**Supplementary Fig. 7a**, Immune class shown), suggesting that the model has learned spurious correlations between the input features and target variables. Fully removing spurious correlation in machine learning models is still an unresolved question. However, balancing the training data, such that each of the 14 combinations of Braak and Dementia (7 Braak values X 2 Dementia values) are equally sampled, can effectively reduce spurious correlations learned during training[146]. In practice, we equally sampled from 15 groups, in which the extra group consisted of donors whose Braak stage or dementia status had not been determined. Training with group balancing reduced the correlation between the predicted Braak and dementia values (**Supplementary Fig. 7b)**, did not adversely affect the model accuracy at the donor-level (**Supplementary Fig. 7c**).

Calculating gene trajectories. We wished to measure how gene expression varied as a function of the predicted Braak stage. First, gene counts were normalized so that each cell's total count was 10,000, followed by the log1p transformation. Second, for each cell class and each donor, we calculated the mean predicted Braak stage, and the mean normalized expression for each gene. Averaging within each donor reduced variability and ensured that donors with greater cell counts did not contribute disproportionately in downstream analysis. Third, we smoothed both the predicted donor-averaged Braak scores, and the donor-averaged gene expression with a Gaussian kernel (**Supplementary Fig. 8**). Specifically, for each donor i, we weighted all other donors j as

$$w_{i,j} = Z_i \exp\left(\frac{(\hat{b}_j - \hat{b}_i)^2}{2\sigma^2}\right)$$

where $\hat{b}_i$ is the predicted Braak score of donor i, σ^2 was set to half the variance of the predicted Braak distribution, and the normalization term Z_i was set such that $\sum_j w_{i,j} = 1$. This allowed us to calculate smoothed Braak, $\tilde{b}_i$ and gene expression, $\tilde{\mathbf{g}}_\mathbf{i}$, values

$$\tilde{b}_i = \sum_j w_{i,j}\hat{b}_j$$

,

$$\tilde{\mathbf{g}}_i = \sum_j w_{i,j}\mathbf{g}_j$$

where $\mathbf{g_i}$ is the gene expression vector of log1p normalized counts for donor i. After ordering the smoothed Braak scores, Braak gene trajectories are now represented as the tuple $\left(\tilde{b}_i, \tilde{\mathbf{g}}_i\right)$.

We only included donors with at least 5 cells for the cell class. We also removed the 10 donors with the least and greatest smoothed predicted Braak scores after smoothing to minimize the edge effects from smoothing.

Resilience against dementia. Since tau proteinopathy and dementia status are highly correlated, gene expression as a function of the two variables is also correlated, and thus partially redundant. Therefore, we aimed to measure how gene expression covaried with predicted dementia given the predicted Braak staging. To do so, we first calculated the expected predicted dementia and expected gene expression for donors with similar Braak staging. Specifically, we defined the expected dementia given Braak, d_i^b, and the expected gene expression given Braak, $\mathbf{g}_i^b$,

$$d_i^b = \sum_j w_{i,j}\hat{b}_j$$

,

$$\mathbf{g}_i^b = \sum_j w_{i,j}\mathbf{g}_j$$

.

The only modification to the kernel was that we set $w_{i,i} = 0$ so that each donor doesn't contribute to its own expected value. We then calculated the residuals between the donor's predicted dementia status and gene expression with its expected values:

$d_i^R = d_i - d_i^n$,

$\mathbf{g}_i^R = \mathbf{g}_i - \mathbf{g}_i^n$.

The dementia resilience score for each donor was then the product of these two terms. When calculating early and late resilience, donors were separated into early and late groups based on their predicted Braak staging before averaging within each group. Using this metric, we define genes as protective if gene expression increases as predicted dementia decreases, given the predicted Braak staging (i.e., the product of the terms defined above is negative). Conversely, we define genes as damaging if gene expression increases as predicted dementia increases, given the predicted Braak staging.

Trajectory nonlinearity. We were interested in measuring how gene expression potentially changes in a nonlinear manner as a function of Braak or dementia. As an initial step, we first wished to quantify the degree of nonlinearity in gene trajectories across the different cell classes. As outlined in **Supplementary Fig. 8**, this calculation was performed across a series of steps. Starting with the gene trajectories for each gene and each cell class as described above, in which smoothed gene expression varied as a function of predicted Braak, we fit each one with a linear fit, and then calculated its explained variance. We calculated the mean explained variance using all genes within each cell class with mean normalized expression above 0.01 and used this value as a baseline. Next, we fit each trajectory with a piecewise linear fit consisting of two consecutive linear segments, separated at some predicted Braak score $\tilde{b} = t$. For each cell class, we determined the optimal t that maximized the explained variance of the piecewise fits for all genes within that cell class. The nonlinearity

index (**Fig. 7b, bottom**) was then defined as the difference in explained variance between the piecewise linear fit and the linear fit. Error bars were determined by bootstrapping the model fits, in which we randomly sampled genes from donors (with replacement), calculating the nonlinearity index, and repeating the process 100 times. Significance between different cell classes was calculated by comparing the 100 x 100 bootstrapped nonlinear index values. We used the optimal time *t* was then used to define an "early" and "late" period. For the Immune cell class, which had the greatest nonlinearity index, the early period was defined as the first 140 donors with the lowest Braak model predictions (after the first 10 donors were removed to account for edge effects, see above). For the other cell classes, we adjusted 140 proportionally based on the number of donors for that cell class and defined this number as N; the early period for each cell class was then defined as the N donors with the lowest Braak model predictions. Since the rank ordering of donor-averaged Braak scores varies between cell classes, donors may be classified as early stage for some cell classes and late stage for others.

Trajectory gene enrichment. We wished to determine which genetic pathways were most significantly up or down-regulated during the progression of AD. To do so, we first extracted the slopes of the "early" and "late" linear fits for the Braak trajectories, and the mean early and late resilience scores (defined above). We used these slopes input to Zenith (https://bioconductor.org/packages/release/bioc/html/zenith.html) to calculate the changes across all GO BP pathways across the eight cell classes. For each GO BP pathway, we calculated the maximum - log10(FDR) score across the eight cell classes and across early and late Braak and Dementia stages. The score of each pathway was assigned this maximum value.

To obtain a condensed list of the most significantly genetic pathways, we first selected all pathways with FDR ≤ 0.05. Next, since we were only interested in pathways that could be informative of the mechanisms underlying AD progression, we excluded pathways containing words referring to overly broad behaviors or cognitive functions ("learning", "memory", "vocalization", "social", "auditory", "startle response", "behavior", "locomotor", "startle", "prepulse inhibition"), terms referring to anatomical structures other than the cortex ("substantia nigra development", "cardiac", "coronary", "aortic", "ventricular", "kidney", "metanephric", "retina", "optic", "bone", "respiratory", "pulmonary", "olfactory", "sperm", "placenta", "egg", "embryonic", "ovulation", "estrous", "placenta", "sperm", "mamary", "germ layer", "outflow tract septum", "adrenal", "epithelial", "skeletal", "otic", "head") or overly broad neural terms ("nervous system process", "cerebral cortex", "recognition", "host", "organ", "developmental growth"). To condense the remaining pathways into a more manageable size, we used rrvgo (https://www.bioconductor.org/packages/release/bioc/html/rrvgo.html). We selected the Wang semantic similarity metric[147], and set the threshold at 0.9 to obtain 56 GO BP pathways (**Supplementary Fig. 9**). For easier visualization, we selected 32 representative pathways from this set for **Fig 7c**. We also performed similar steps (pathways with minimum FDR < 0.05 included, Wang metric with threshold at 0.9) to obtain the top GO Molecular Function and Cellular Component (**Supplementary Figs. 10, 11**) pathways. In all cases, we only included pathways with at least 10 genes to ensure that results were statistically robust, and no more than 250 genes to ensure that pathways were not overly broad.

Mean trajectories and MAGMA enrichment. For both the mean normalized expression (**Fig. 7d**) and Magma enrichment analysis (**Fig. 7f**), we used the top 250 coding genes based upon the early and late slopes of the Braak trajectories. Late decreasing genes tend to also appear to contain an early increase (**Fig. 7d**). However, we cannot say whether this is biologically meaningful or the result of selection bias, as a strong late decrease must be preceded by a high baseline. For both the mean expression and Magma enrichment calculations, results were qualitatively similar if we used the top 500 or 1000 genes instead (data not shown).

Data availability

All data 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[102] detailing the data processing and data collection.

Code availability

All the source codes used in this study are available via GitHub https://github.com/DiseaseNeuroGenomics/PsychADxD.

Acknowledgments

We would like to express our deep gratitude to the patients and their families who generously donated the invaluable biological material essential for the success of this study. We are profoundly indebted to their participation and commitment to advancing scientific knowledge and improving human health. We acknowledge the National Institute on Aging for their generous support in funding this research with 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), the Rush Alzheimer's Disease Center (RADC; funding: P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG22018, U01AG46152, and U01AG61356), and NIMH-IRP Human Brain Collection Core (HBCC, project # ZIC MH002903). This work was also supported by the Novo Nordisk Foundation NNF14CC0001 and NNF20SA0035590. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.

Author contributions

Conceptualization: PR, VH, JFF, DL
Methodology: PR, JFF, DL, GEH
Software: GEH, DL, NYM
Validation: DL, XW, JMF, JFF, PK
Formal analysis: DL, MK, NYM, GEH, SK, MP, TC, FT
Investigation: JFF, AH, CC, ZS, MA, SA, XW, JMF
Resources: VH, DAB, SM, LLB, PKA
Data Curation: JB, PNM, DM, DB, KT, DL
Writing: DL, PR, JFF, NYM with supports from all co-authors.
Visualization: DL, MK, NYM
Supervision: PR, DL, JFF, JB, VH, GV, KG, LJJ
Project administration: DL, PR
Funding acquisition: PR, VH

All authors read and approved the final draft of the paper.

Competing interests declaration

The authors declare no competing interests.

Materials & Correspondence

Correspondence to Donghoon Lee or Panos Roussos.

Supplementary Information

PsychAD Consortium Authors

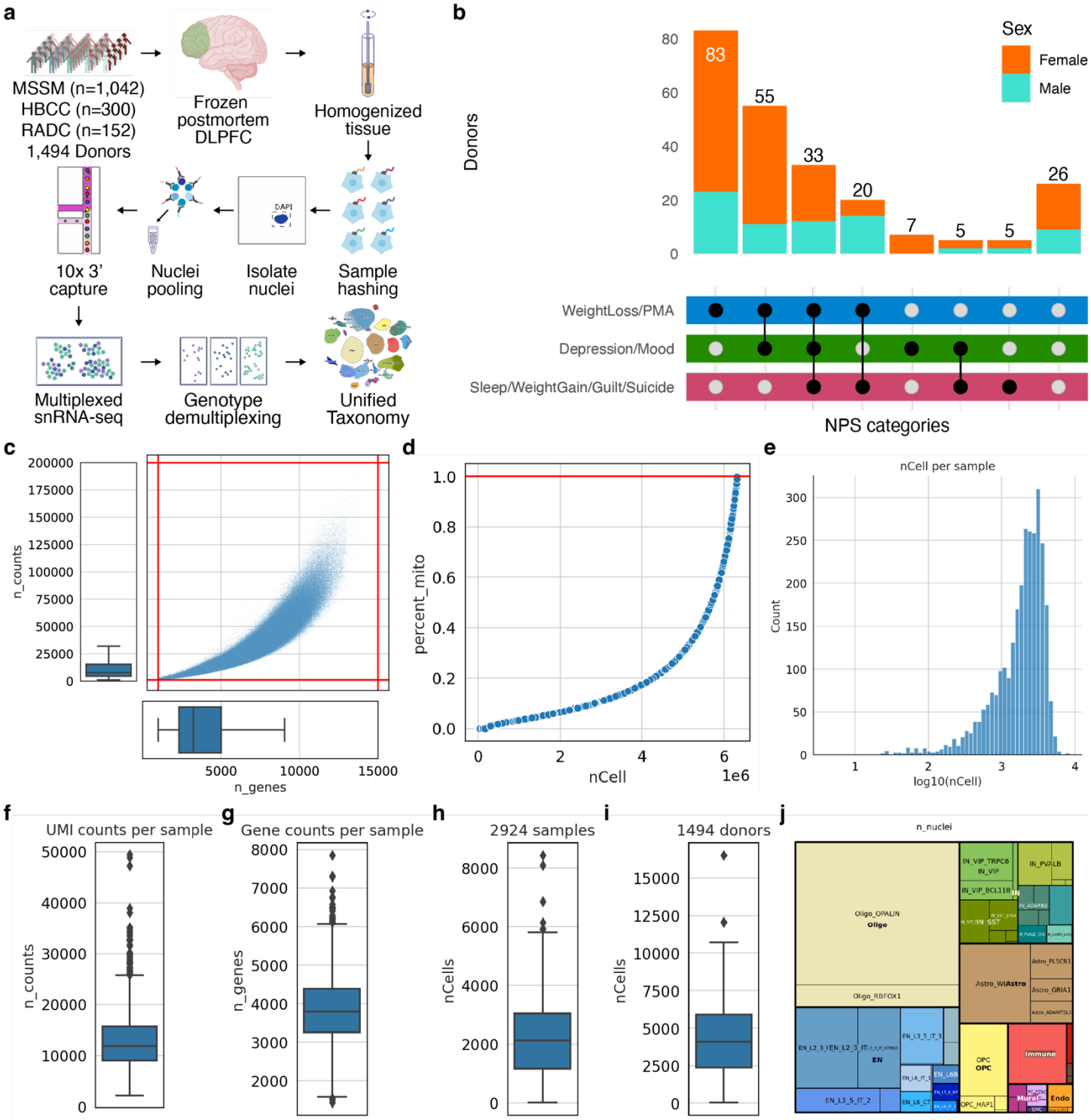
Aram Hong (1, 4, 6, 7); Athan Z. Li (10, 12); Biao Zeng (1, 4, 6, 7); Chenfeng He (9, 12); Chirag Gupta (9, 12); Christian Porras (1, 4, 6, 7); Clara Casey (1, 4, 6, 7); Colleen A. McClung (18); Collin Spencer (1, 4, 6, 7); Daifeng Wang (9, 10, 12); David A. Bennett (19); David Burstein (1, 2, 4, 6, 7, 8); Deepika Mathur (1, 4, 6, 7); Donghoon Lee (1, 4, 6, 7); Fotios Tsetsos (1, 2, 4, 6, 7); Gabriel E. Hoffman (1, 2, 4, 6, 7, 8); Genadi Ryan (13, 17); Georgios Voloudakis (1, 2, 3, 4, 6, 7, 8); Hui Yang (1, 4, 6, 7); Jaroslav Bendl (1, 4, 6, 7); Jerome J. Choi (11, 12); John F. Fullard (1, 4, 6, 7); Kalpana H. Arachchilage (9, 12); Karen Therrien (1, 4, 6, 7); Kiran Girdhar (1, 4, 6, 7); Lars J. Jensen (21); Lisa L. Barnes (19); Logan C. Dumitrescu (22, 23); Lyra Sheu (1, 4, 6, 7); Madeline R. Scott (18); Marcela Alvia (1, 4, 6, 7); Marios Anyfantakis (1, 4, 6, 7); Maxim Signaevsky (6, 7); Mikaela Koutrouli (1, 4, 6, 7, 21); Milos Pjanic (1, 4, 6, 7); Monika Ahirwar (13, 17); Nicolas Y. Masse (1, 4, 6, 7); Noah Cohen Kalafut (10, 12); Panos Roussos (1, 2, 4, 6, 7, 8); Pavan K. Auluck (20); Pavel Katsel (6); Pengfei Dong (1, 4, 6, 7); Pramod B. Chandrashekar (9, 12); Prashant N.M. (1, 4, 6, 7); Rachel Bercovitch (1, 4, 6, 7); Roman Kosoy (1, 4, 6, 7); Sanan Venkatesh (1, 2, 4, 6, 7); Saniya Khullar (9, 12); Sayali A. Alatkar (10, 12); Seon Kinrot (1, 4, 6, 7); Stathis Argyriou (1, 4, 6, 7); Stefano Marenco (20); Steven Finkbeiner (13, 14, 15, 16, 17); Steven P. Kleopoulos (1, 4, 6, 7); Tereza Clarence (1, 4, 6, 7); Timothy J. Hohman (22, 23); Ting Jin (9, 12); Vahram Haroutunian (5, 6, 7, 8); Vivek G. Ramaswamy (13, 17); Xiang Huang (12); Xinyi Wang (1, 4, 6, 7); Zhenyi Wu (1, 4, 6, 7); Zhiping Shao (1, 4, 6, 7)

PsychAD Consortium Affiliations

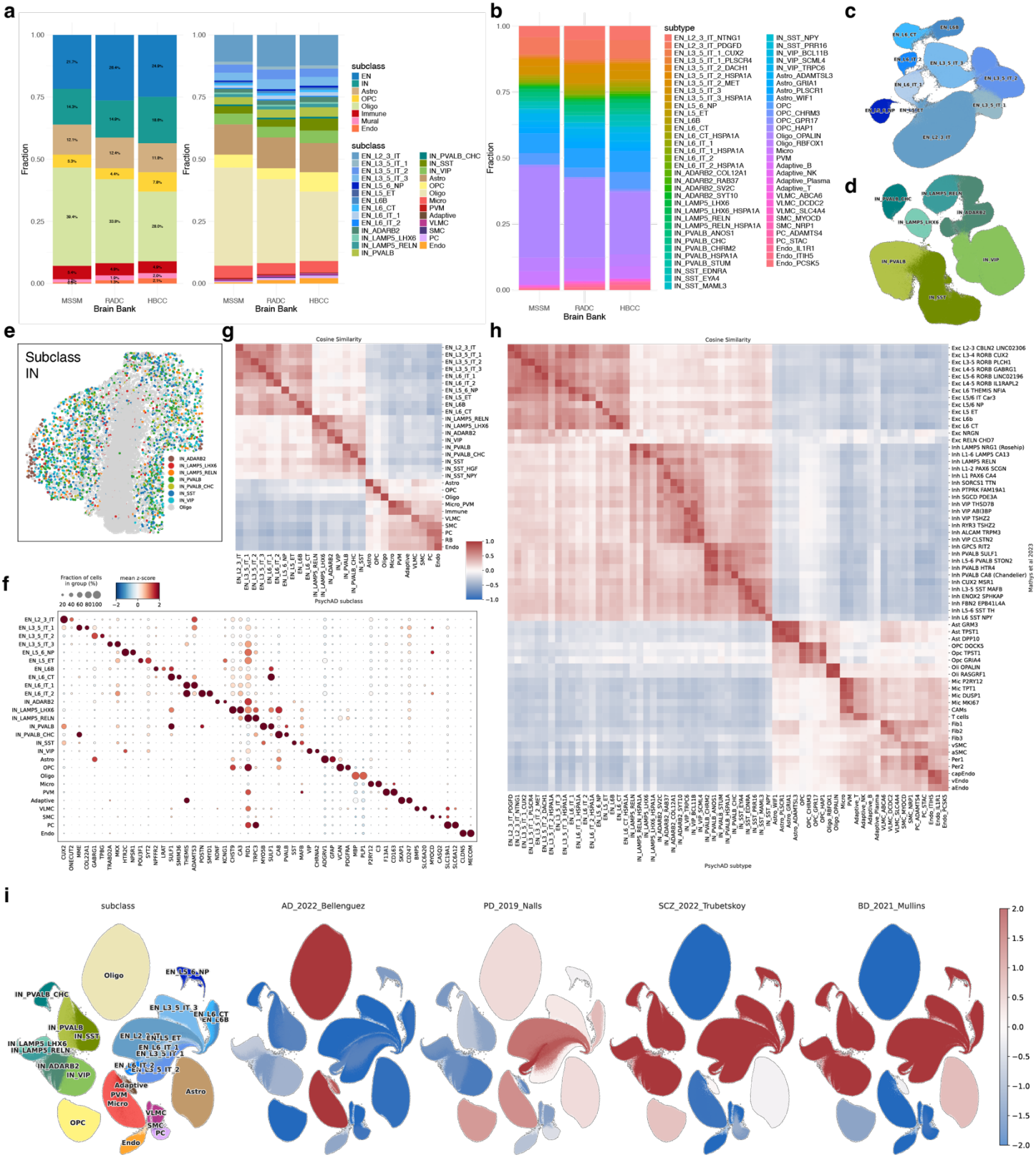
1: Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA
2: Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, USA
3: Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA
4: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
5: Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA
6: Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA
7: Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
8: Mental Illness Research, Education and Clinical Center VISN2, James J. Peters VA Medical Center, Bronx, NY, USA
9: Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA
10: Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA
11: Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI, USA
12: Waisman Center, University of Wisconsin-Madison, Madison, WI, USA
13: Center for Systems and Therapeutics, Gladstone Institutes, San Francisco, CA, USA
14: Department of Neurology, University of California San Francisco, San Francisco, CA, USA
15: Department of Physiology, University of California San Francisco, San Francisco, CA, USA
16: Neuroscience and Biomedical Sciences Graduate Programs, University of California San Francisco, San Francisco, CA, USA
17: Taube/Koret Center for Neurodegenerative Disease Research, Gladstone Institutes, San Francisco, CA, USA
18: Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
19: Rush Alzheimer's Disease Center and Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

20: Human Brain Collection Core, National Institute of Mental Health-Intramural Research Program, Bethesda, MD, USA
21: Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
22: Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA
23: Vanderbilt Memory & Alzheimer's Center, Vanderbilt University Medical Center, Nashville, TN, USA

Supplementary Figures

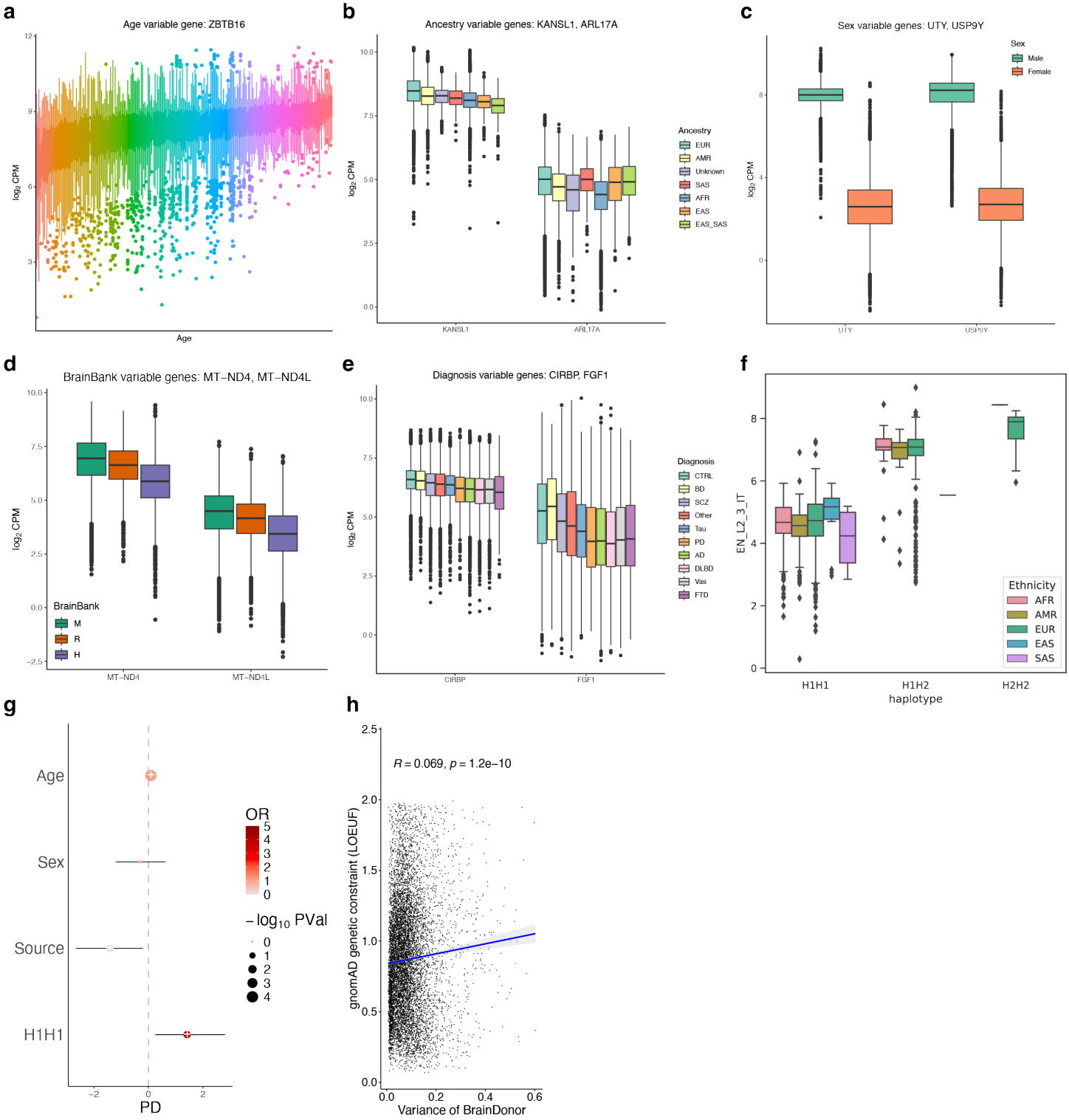


Supplementary Fig. 1. **(a)** Extended schematic overview of snRNA-seq experiments leading to unified taxonomy. **(b)** Donor breakdown by NPS categories and sex. **(c)** Distribution of gene counts by UMI counts after QC of 6M nuclei. **(d)** Cumulative nuclei count by percent mitochondrial genes. **(e)** Histogram of nuclei counts per sample. **(f)** Distribution of UMI counts per sample. **(g)** Distribution of gene counts per sample. **(h)** Distribution of nuclei count per sample. **(i)** Distribution of nuclei count per donor. **(j)** Treemap showing the color scheme of the unified taxonomy.



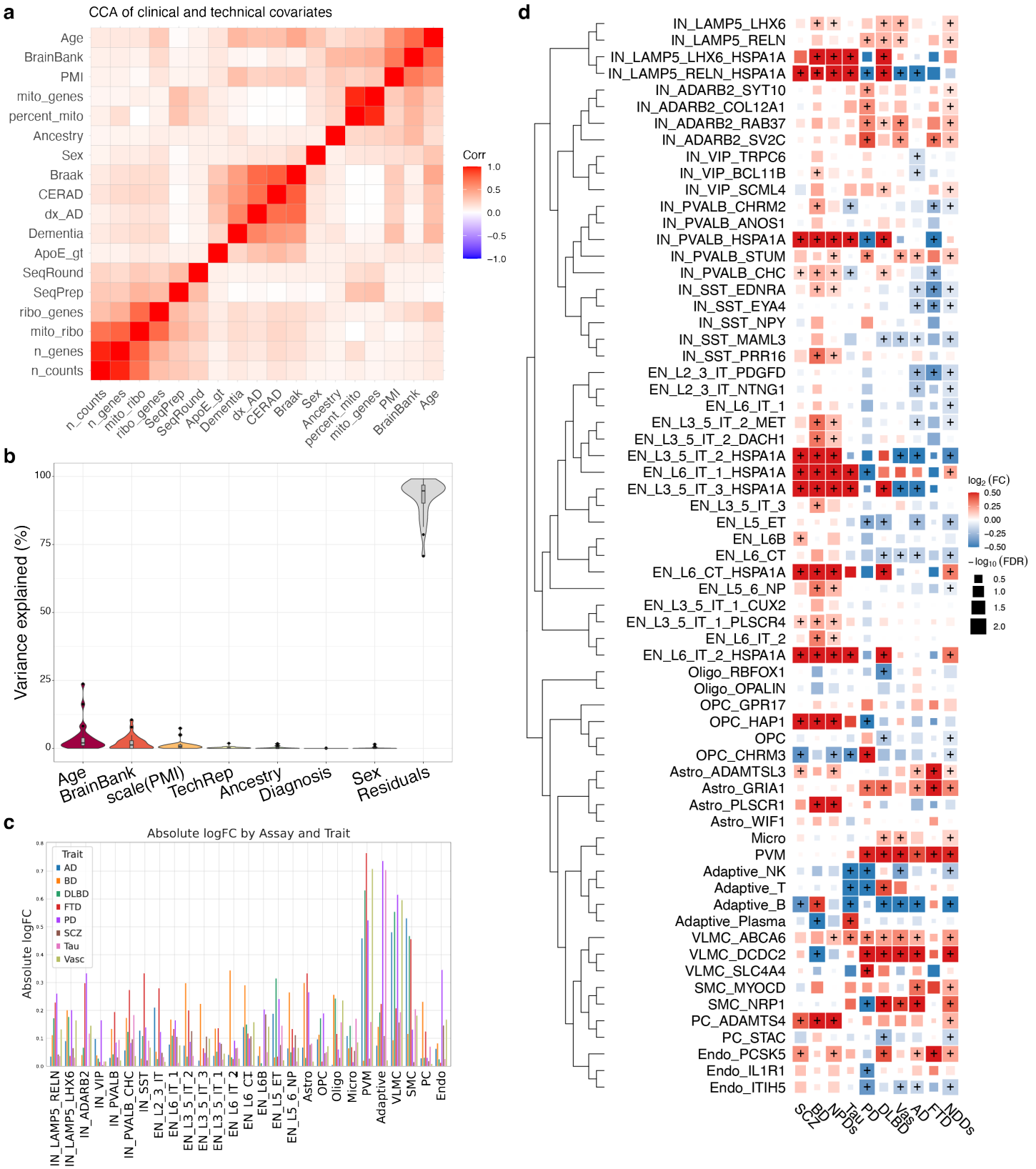
Supplementary Fig. 2. **(a)** Composition of cell types in each brain bank using class and subclass-level cellular taxonomy. **(b)** Composition of subtype in each brain bank. **(c)** Clustering of EN lineage. A UMAP focusing on the diversity of 10 subclasses of excitatory neurons. **(d)** Clustering of IN lineage. A UMAP focusing on the diversity of 7 subclasses of inhibitory neurons (IN). **(e)** Spatial distribution of IN subclass. **(f)** Expression and fraction of marker gene expression in 27 subclasses. Expression z-scaled and averaged across nuclei. **(g)** Comparison of subclass-level cellular taxonomy with Ma et al 2022. **(h)** Comparison of subtype-level cellular

taxonomy with Mathys et al 2023. **(i)** Disease enrichment scores based on GWAS (scDRS) for AD, PD, SCZ, and BD.



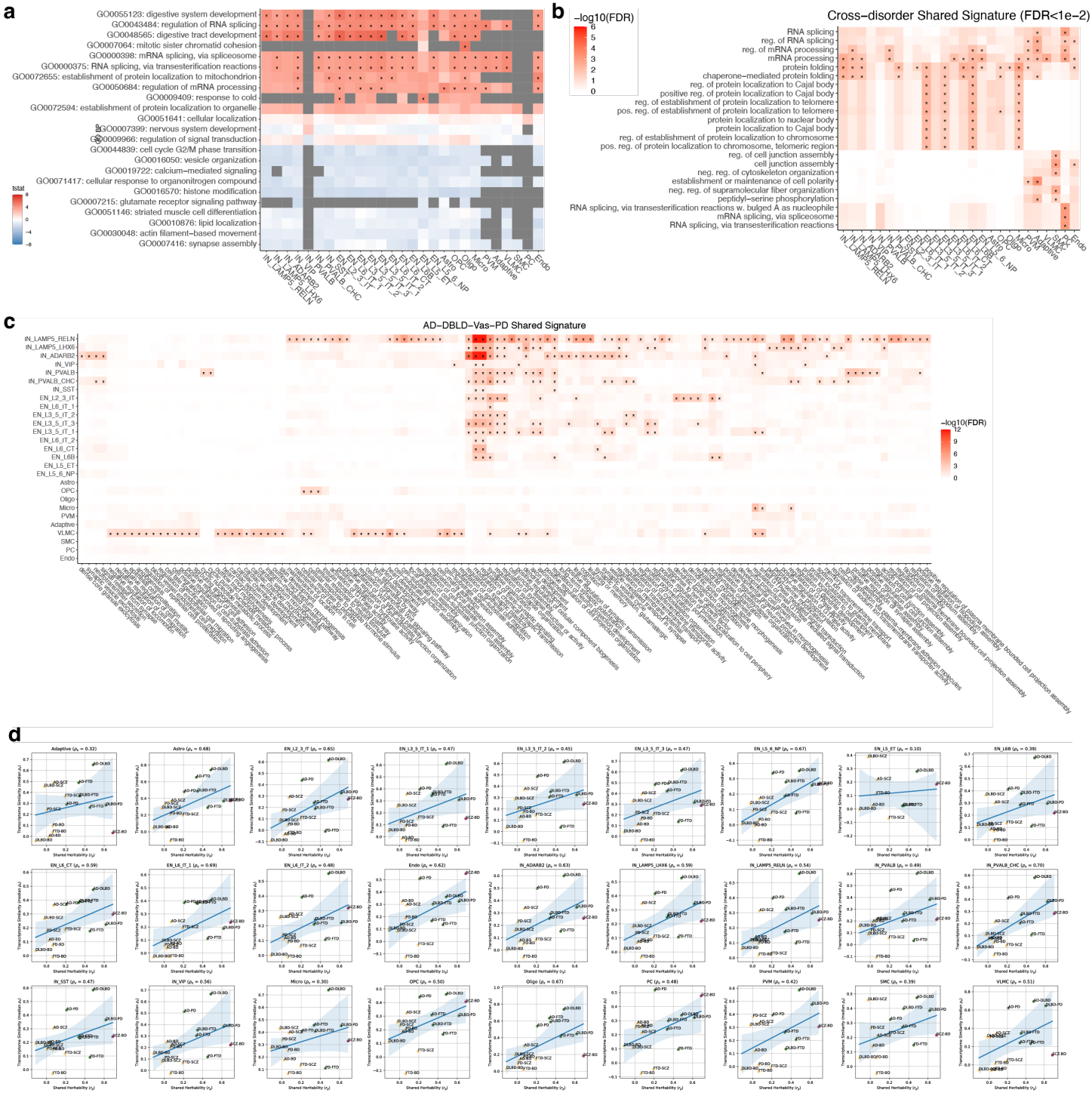
Supplementary Fig. 3. **(a)** Example of a gene with the highest variation across age. **(b)** Example of a gene with the highest variation across genetic ancestry. **(c)** Example of a gene with the highest variation across sex. **(d)** Example of a gene with the highest variation across brain banks. **(e)** Example of a gene with the highest variation across diagnosis. **(f)** Normalized expression of ARL17B gene within EN_L2_3_IT subclass stratified by genetic ancestry. **(g)** The logistic regression coefficient for PD risk using the haplotype of MAPT locus. **(h)**

Spearman correlation test between variance explained by inter-individual variation and average LOEUF conservation score.



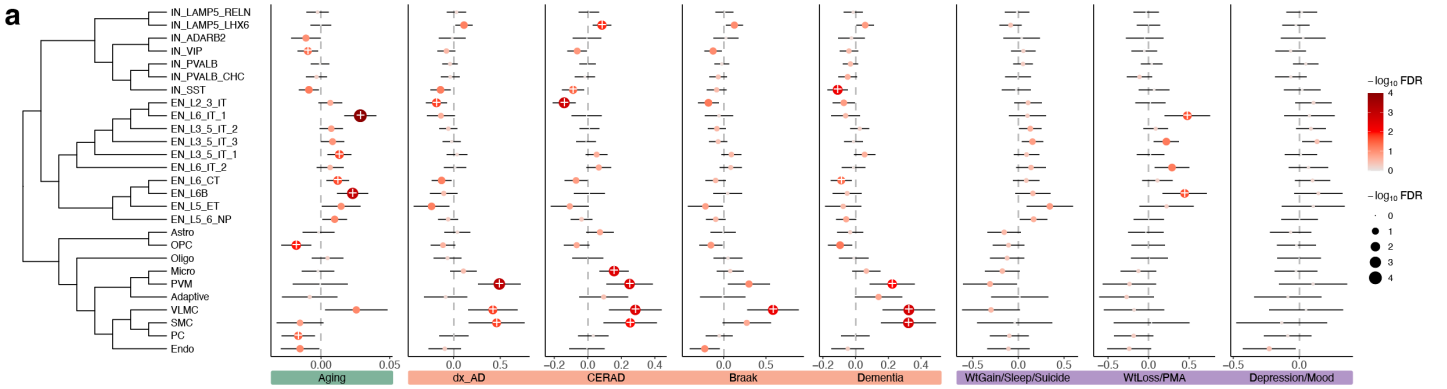
Supplementary Fig. 4. Cross-disorder variation of cell type composition comparing 8 different NDDs and NPDs against common neurotypical controls. **(a)** Correlation matrix showing the relationships between various

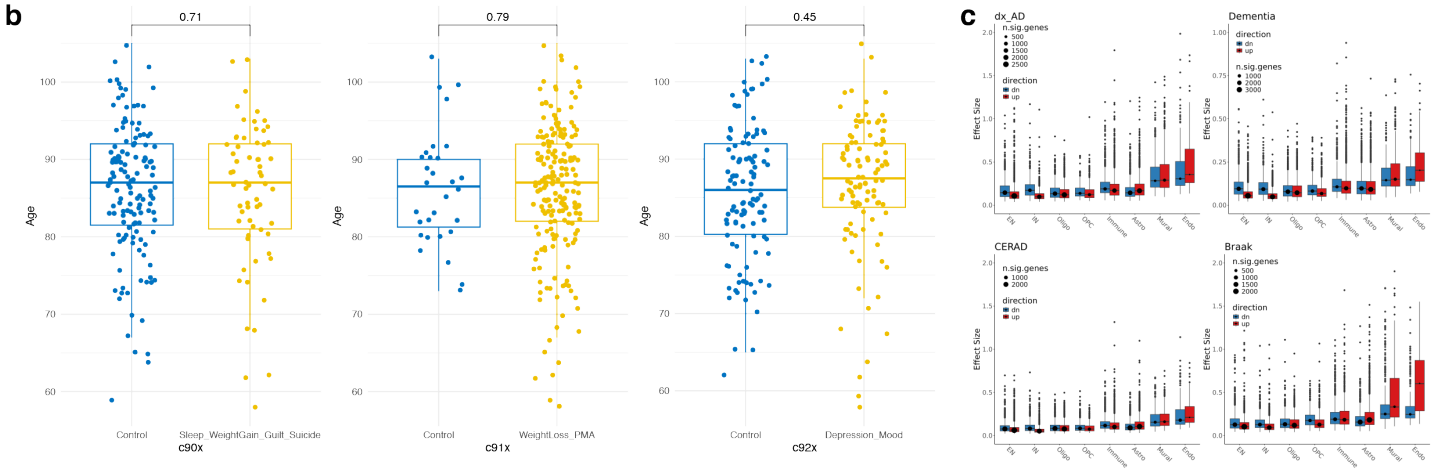
clinical and demographic variables for the selection of the covariates. The color intensity indicates the strength of the correlation, with red representing positive correlations and blue negative correlations. **(b)** Variance partition of cell type composition, displaying the variance captured by different covariates in the formula. **(c)** Bar plot depicting the absolute logFC values of the 8 different NDDs and NPDs across the cell types. **(d)** Variation in cell type composition for each subtype in 8 different diseases. NDDs and NPDs indicate meta-analysis using broad disease categories. Color intensity indicates effect size and dot size reflects the statistical significance of correlations.

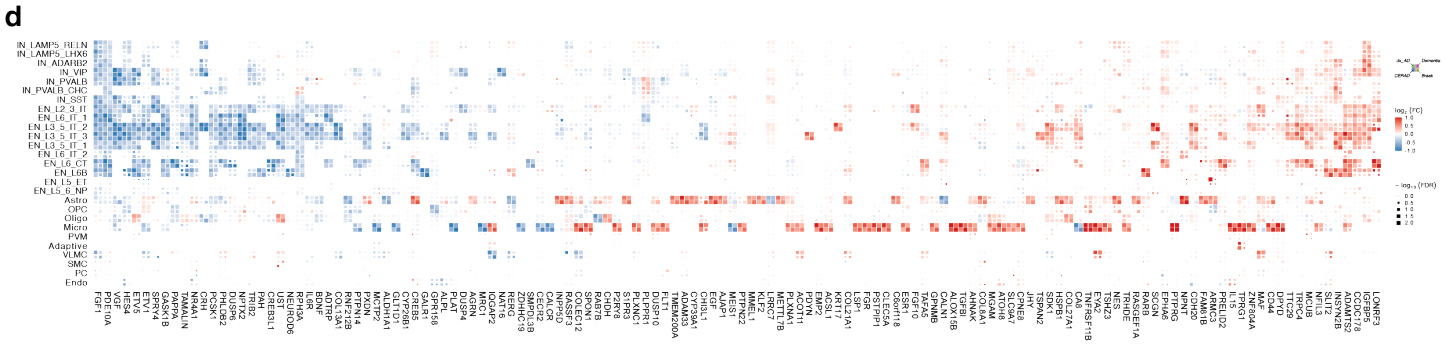


Supplementary Fig. 5. Cross-disorder gene expression signature. **(a)** GO BP pathways implicated by shared gene expression changes across 8 disorders using Zenith. T-statistic values indicate the direction and

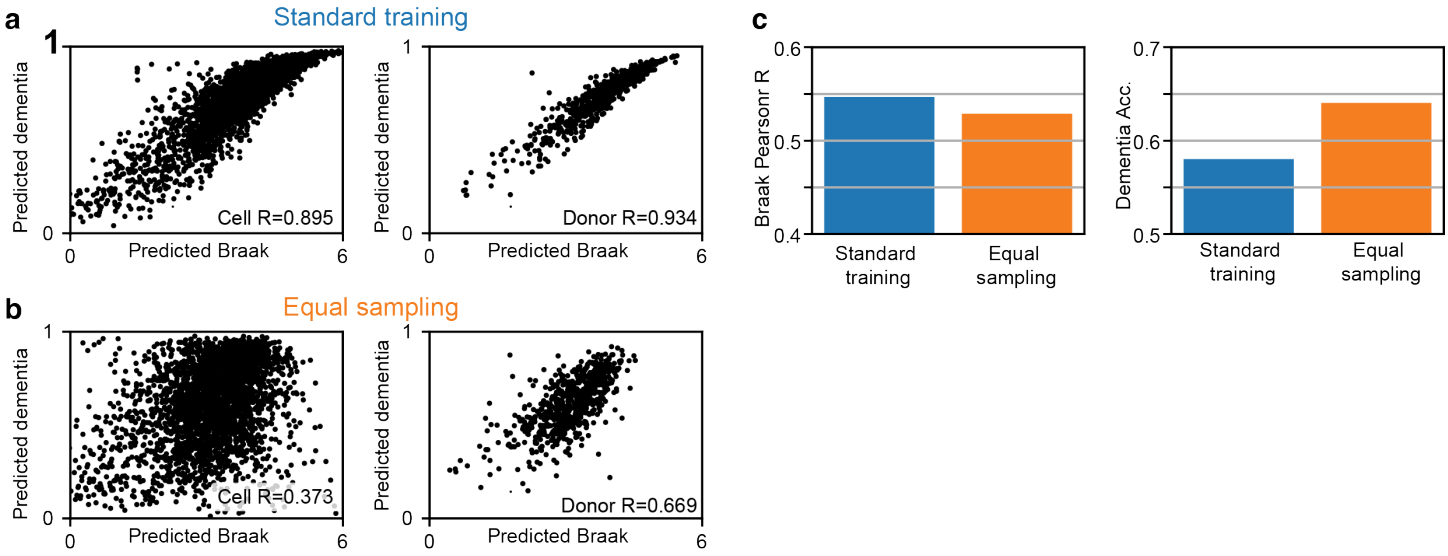
magnitude of the gene set enrichment. **(b)** Extended data for **Fig. 5a** with all 27 subclasses and without reduction in similar GO terms. **(c)** Pathways implicated by shared signatures from AD, DLBD, Vas, and PD. **(d)** Correlation between shared heritability and transcriptome similarity per cell subclass.



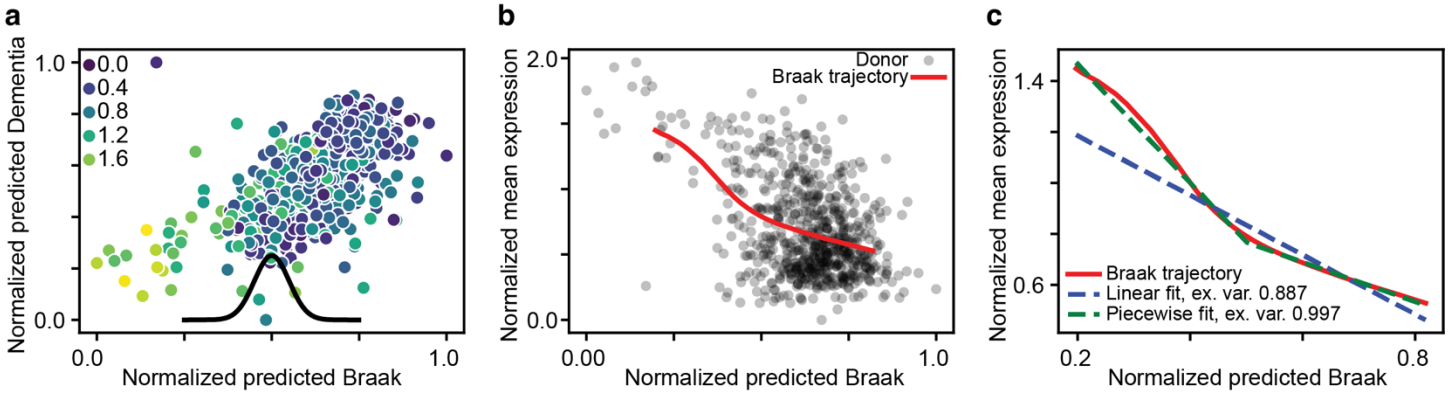




Supplementary Fig. 6. Transcriptomic changes across AD neuropathology. **(a)** Extended data for **Fig. 6a**. Compositional variation analysis using normal aging, different measures of AD pathology (binary AD diagnosis (dx_AD), CERAD score, Braak staging, and ordinal dementia scale), and 3 categories of NPS within AD. **(b)** Age distribution of AD cases with or without three NPS categories. **(c)** Mean effect sizes aggregated by direction of effect per cell class. **(d)** Extended data for **Fig. 6d**. DEGs in AD phenotypes. Meta-analysis between brain banks. Top genes with FDR < 0.01 and effect size ≥ 0.35. **(e)** Functional enrichment analysis of DEGs by subclass using Gene Ontology Biological Process. Hypergeometric test with FDR ≤ 0.01 shown.



Supplementary Fig. 7. The effect of different training methods on model prediction correlations and accuracy for the Immune cell class. **(a)** Correlation between Braak (x-axis) and dementia (y-axis) model predictions at the cell-level (left panel) and donor-averaged level (right panel) when all cells are equally sampled. To reduce the figure size, only 1 out of 20 cells shown in the left panel. **(b)** Same as (a), except that all combinations of Braak stage and dementia status are equally sampled during training. This has the effect of reducing the correlation between the predicted Braak and dementia values. **(c)** The correlation between the actual and predicted Braak values (left panel) and dementia classification accuracy (right panel), both measured at the donor-averaged level, are shown for the two training methods.

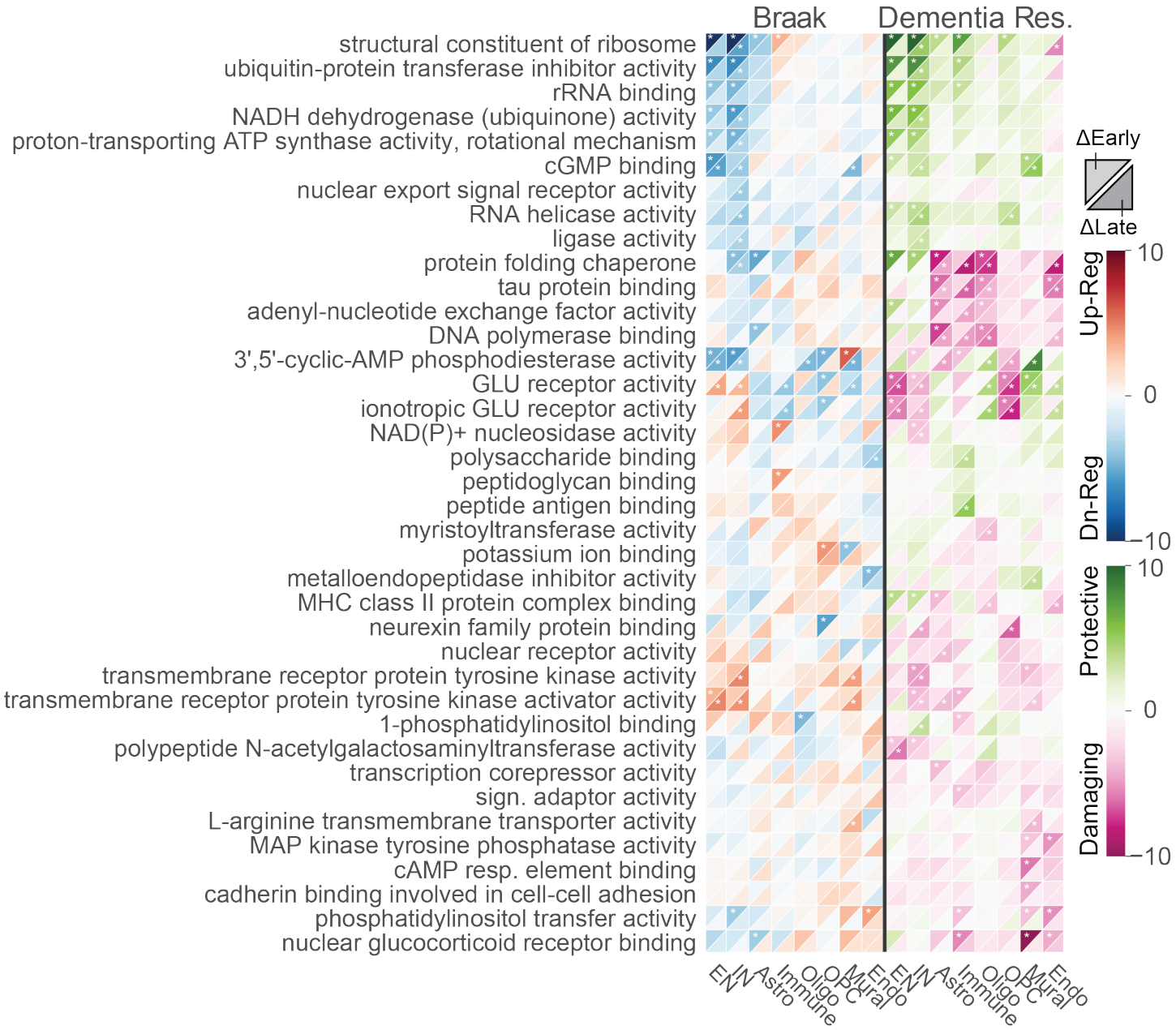


Supplementary Fig. 8. The steps involved in calculating the Braak trajectory for an example gene - NAV2 from the Immune cell class. **(a)** Given cell-based model predictions along with normalized gene counts, we can calculate the donor-averaged predicted Braak (x-axis), predicted dementia (y-axis), and normalized gene counts (hue). Predicted Braak and dementia values are normalized between 0 and 1. We then smooth the predicted Braak scores along with the gene counts with a Gaussian kernel (black curve, width is shown at scale). **(b)** After smoothing, we obtain the Braak trajectory for this gene, relating gene expression to the

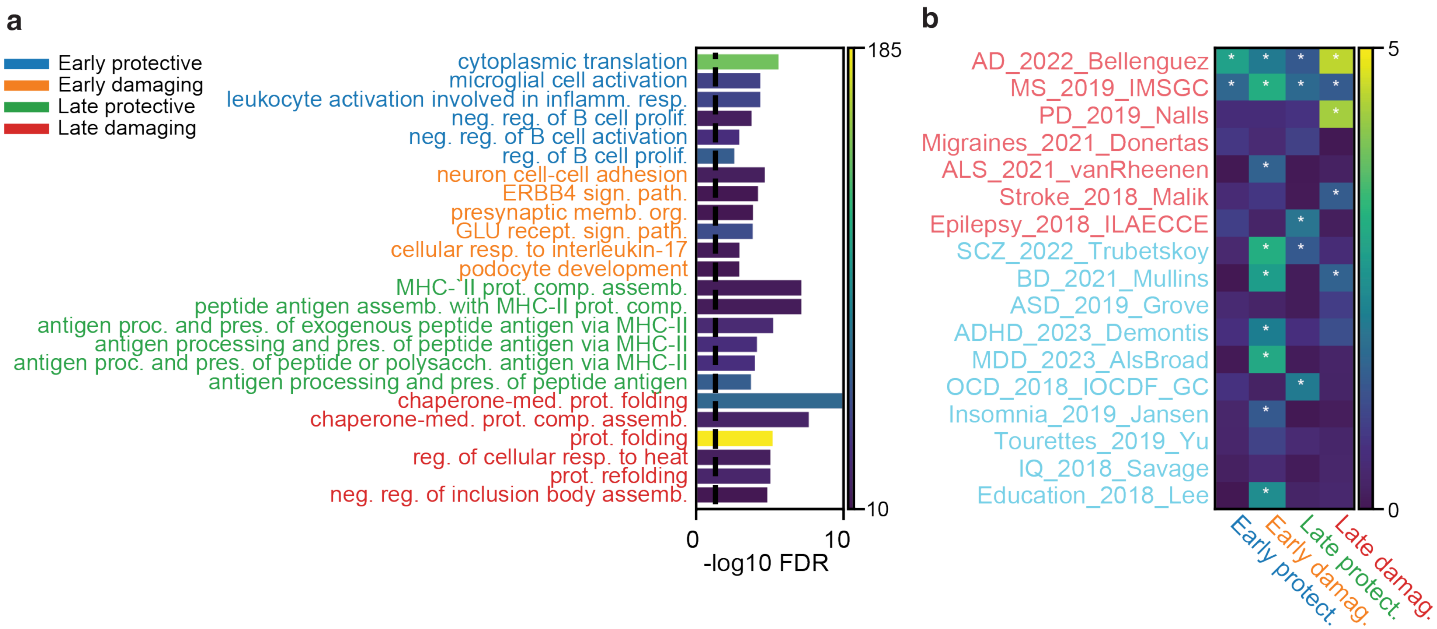
predicted Braak value. **(c)** We then compare how well we can fit this trajectory with a single linear model (blue dashed line) and a piecewise linear model (green dashed line). The piecewise linear model consists of two consecutive linear fits, in which the transition point has been optimized to maximize the explained variance. This transition point is then used to define an "early" and "late" disease stages (**Methods**).

Supplementary Fig. 9. Extended data for **Fig. 7c**. Expanded list of all 56 GO BP pathways. Upper left triangles indicate the early phases, and lower right triangles indicate the late phases of the AD trajectory predicted by Braak and dementia resilience. Hue indicates the z-score, and stars indicate FDR < 0.05.

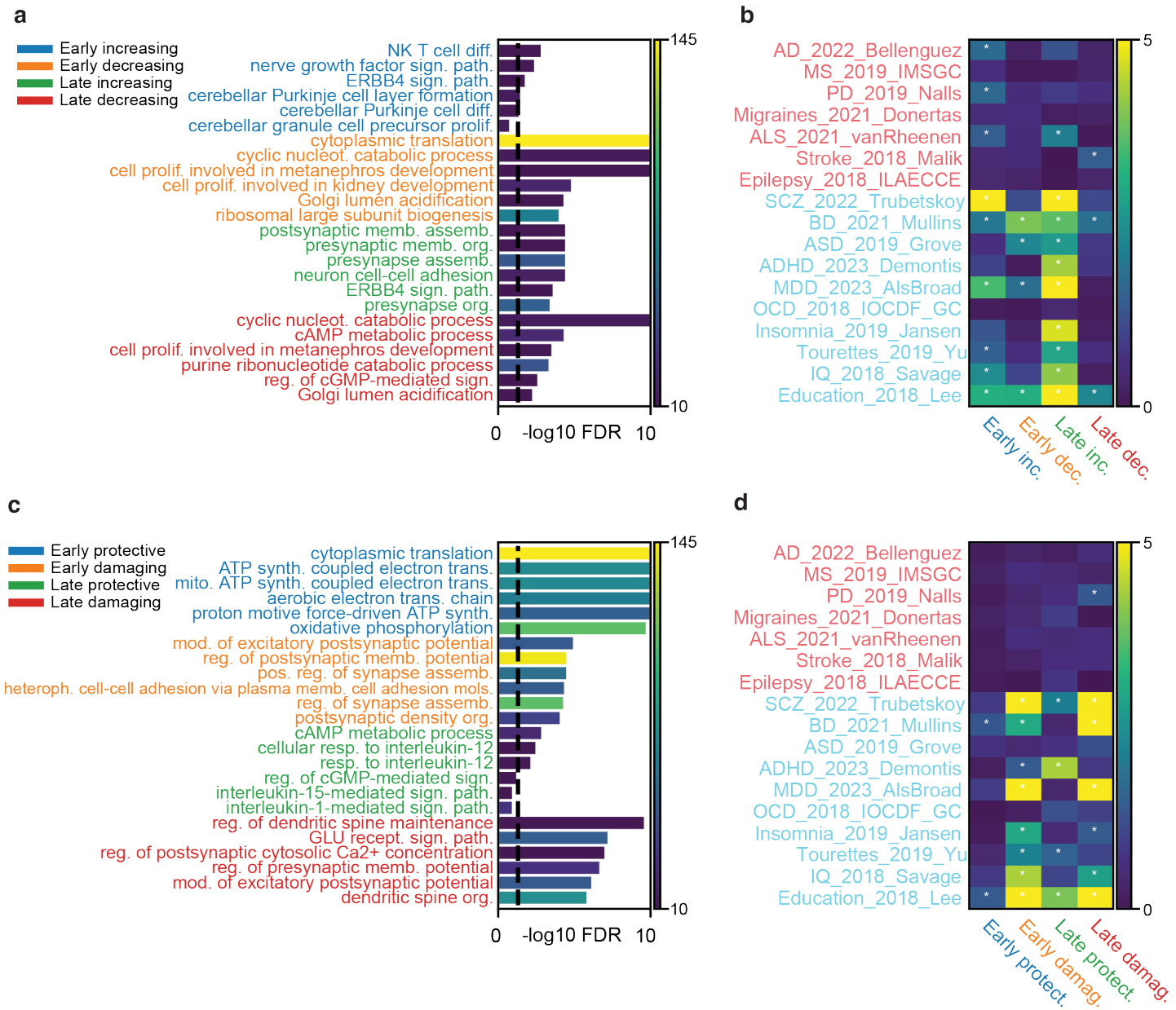
Supplementary Fig. 10. Pathway enrichment using the top GO Cellular Components pathways. Upper left triangles indicate the early phases, and lower right triangles indicate the late phases of the AD trajectory predicted by Braak and dementia resilience. Hue indicates the z-score, and stars indicate FDR < 0.05.



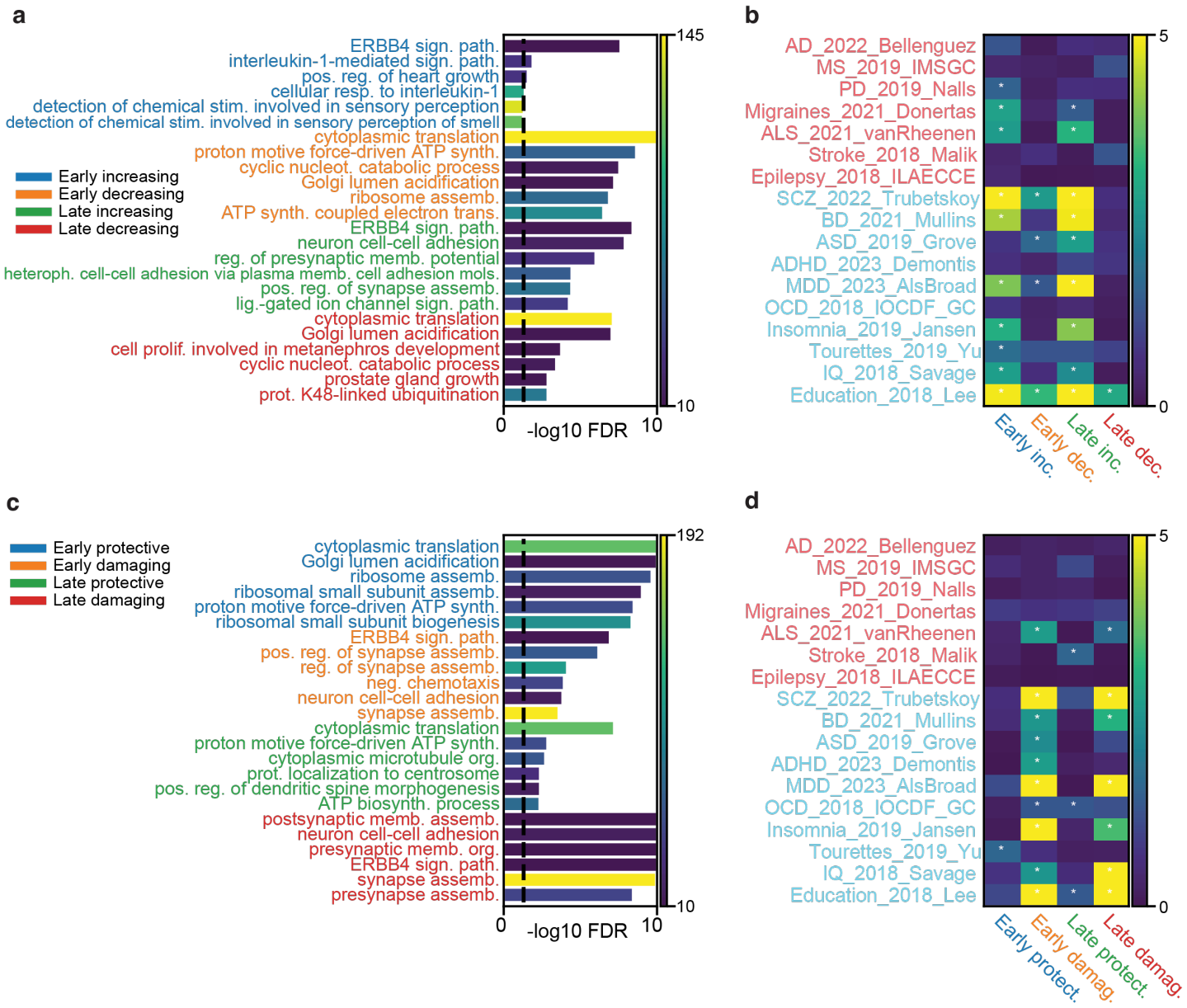
Supplementary Fig. 11. Pathway enrichment using the top GO Molecular Function pathways. Upper left triangles indicate the early phases, and lower right triangles indicate the late phases of the AD trajectory predicted by Braak and dementia resilience. Hue indicates the z-score, and stars indicate FDR < 0.05.



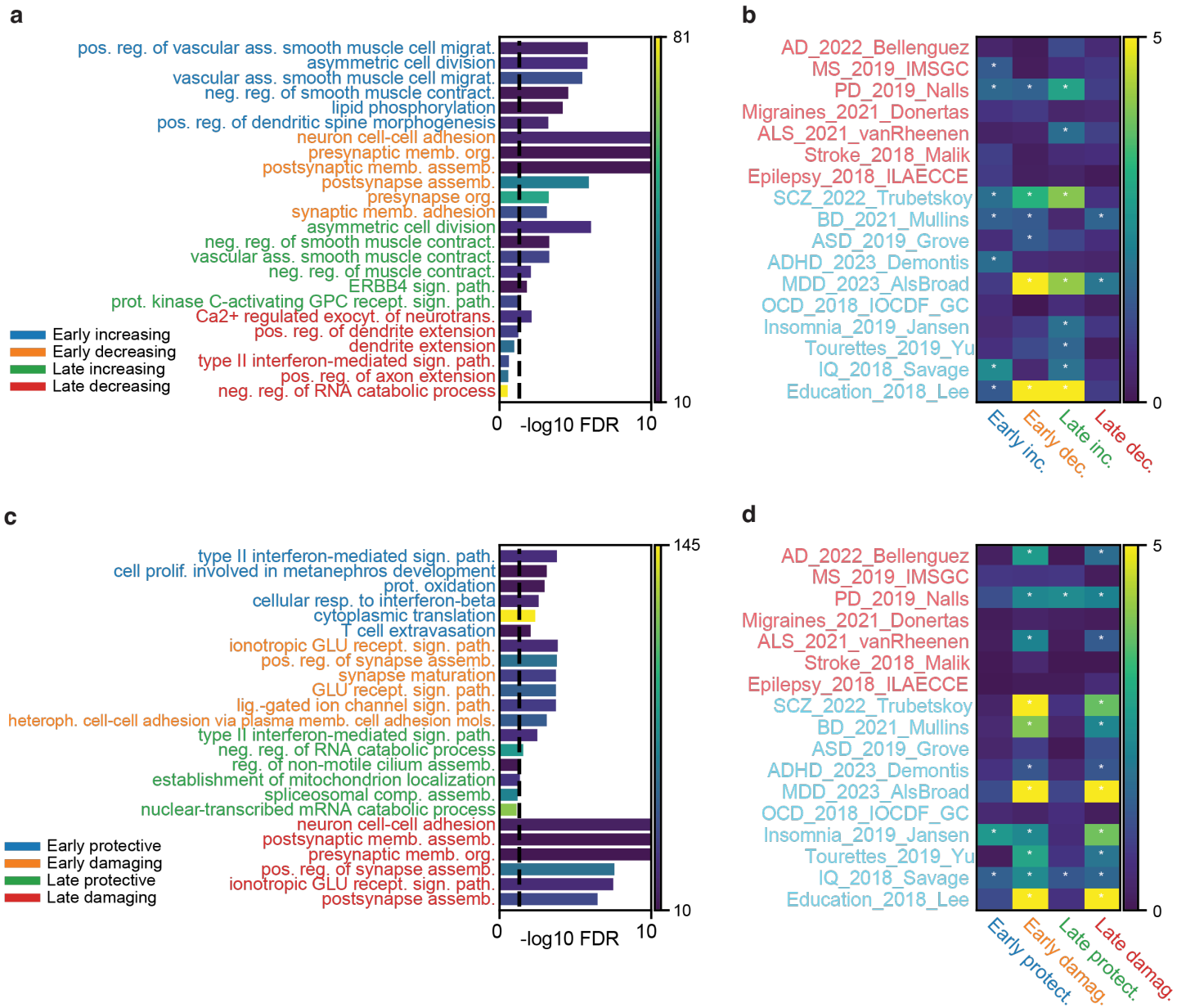
Supplementary Fig. 12. GO BP pathways and GWAS enrichment results for dementia resilience. **(a)** Top early protective pathways (i.e. resilience against dementia, blue), early damaging (i.e. associated with dementia, orange), late protective (green) and late damaging (red). Hue indicates the number of genes in the pathway. Negative log FDR clipped at 10. **(b)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module. Text color indicates neurological traits (light red) and psychiatric traits (cyan). Hue indicates negative log10(FDR), asterisk indicates FDR < 0.05. Negative log FDR clipped at 5.



Supplementary Fig. 13. GO BP pathways and GWAS enrichment results for the EN cell class based on their response to Braak (a,b) and dementia resilience (c,d). **(a)** Pathway enrichment of genes with early increasing response to increasing tau proteinopathy (blue), early decreasing (orange), late increasing (green) and late decreasing (red). Hue indicates the number of genes in the pathway. Negative log FDR clipped at 10. **(b)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module. Text color indicates neurological traits (light red) and psychiatric traits (cyan). Hue indicates negative log10(FDR), asterisk indicates FDR < 0.05. Negative log FDR clipped at 5. **(c)** Pathway enrichment of genes with early protective resilience against dementia (blue), early damaging associated with dementia (orange), late protective (green), and late damaging (red). **(d)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module.



Supplementary Fig. 14. GO BP pathways and GWAS enrichment results for the IN cell class based on their response to Braak (a,b) and dementia resilience (c,d). **(a)** Pathway enrichment of genes with early increasing response to increasing tau proteinopathy (blue), early decreasing (orange), late increasing (green) and late decreasing (red). Hue indicates the number of genes in the pathway. Negative log FDR clipped at 10. **(b)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module. Text color indicates neurological traits (light red) and psychiatric traits (cyan). Hue indicates negative log10(FDR), asterisk indicates FDR < 0.05. Negative log FDR clipped at 5. **(c)** Pathway enrichment of genes with early protective resilience against dementia (blue), early damaging associated with dementia (orange), late protective (green), and late damaging (red). **(d)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module.



Supplementary Fig. 15. GO BP pathways and GWAS enrichment results for the OPC cell class based on their response to Braak (a,b) and dementia resilience (c,d). **(a)** Pathway enrichment of genes with early increasing response to increasing tau proteinopathy (blue), early decreasing (orange), late increasing (green) and late decreasing (red). Hue indicates the number of genes in the pathway. Negative log FDR clipped at 10. **(b)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module. Text color indicates neurological traits (light red) and psychiatric traits (cyan). Hue indicates negative log10(FDR), asterisk indicates FDR < 0.05. Negative log FDR clipped at 5. **(c)** Pathway enrichment of genes with early protective resilience against dementia (blue), early damaging associated with dementia (orange), late protective (green), and late damaging (red). **(d)** Enrichment of heritability estimates (MAGMA) for each disease trajectory module.

Supplementary Tables

Supplementary Table 1. Clinical and technical metadata of 1,494 donors in the PsychAD cohort. Metadata includes binary contrasts and continuous variables used in the study.

Supplementary Table 2. Hierarchical cellular taxonomy of the PsychAD snRNA-seq data. Metadata including parent-child relationships, color hex codes, and number of nuclei per subtype.

Supplementary Table 3. Compositional variation analysis across 8 cross-disorder traits.

Supplementary Table 4. Differentially expressed genes across 8 cross-disorder traits.

Supplementary Table 5. Compositional variation analysis across AD phenotypes.

Supplementary Table 6. Differentially expressed genes across AD phenotypes.

Supplementary Table 7. The top 50 GO BP pathways for each cell class relative to early/late increases/decreases in predicted Braak staging, and relative to early/late dementia resilience

Supplementary Table 8. Gene-set enrichment analysis using GWAS summary data for each cell class. GWAS scores are calculated using the top 250 coding genes with the greatest slopes measured relative to early/late increases/decreases in predicted Braak staging, and the top 250 early/late protective/damaging coding genes.

Supplementary Table 9. The top 250 coding genes with the greatest slopes measured relative to early/late increases/decreases in predicted Braak staging, and the top 250 early/late protective/damaging coding genes.

Supplementary Table 10. List of gene markers used from Xenium human brain custom panel.

Supplementary Data

Supplementary Data 1. Metadata and model predictions for each cell. Each cell class contains a table that includes the cell barcode, donor information, the cell subclass, and both the actual Braak and dementia status along with the model predictions.

Reference

1. 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).
2. Gabitto, M. I. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. *bioRxiv* 2023.05.08.539485 (2024) doi:10.1101/2023.05.08.539485.
3. Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
4. Wingo, T. S. *et al.* Shared mechanisms across the major psychiatric and neurodegenerative diseases. *Nat. Commun.* **13**, 4314 (2022).
5. Argyriou, S. *et al.* Beyond memory impairment: the complex phenotypic landscape of Alzheimer's disease. *Trends Mol. Med.* (2024) doi:10.1016/j.molmed.2024.04.016.
6. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).
7. BRAIN Initiative Cell Census Network (BICCN). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* **598**, 86–102 (2021).
8. Kwan, K. Y., Sestan, N. & Anton, E. S. Transcriptional co-regulation of neuronal migration and laminar identity in the neocortex. *Development* **139**, 1535–1546 (2012).
9. Jagadeesh, K. A. *et al.* Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* **54**, 1479–1492 (2022).
10. 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).
11. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
12. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, (2019).
13. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

14. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address: epilepsy-austin@unimelb.edu.au. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **13**, 893–903 (2014).
15. Dönertaş, H. M., Fabian, D. K., Valenzuela, M. F., Partridge, L. & Thornton, J. M. Common genetic associations between age-related diseases. *Nat Aging* **1**, 400–412 (2021).
16. Malik, R. *et al.* Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann. Neurol.* **84**, 934–939 (2018).
17. van Rheenen, W. *et al.* Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat. Genet.* **53**, 1636–1648 (2021).
18. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
19. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
20. Als, T. D. *et al.* Depression pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders using genome-wide analyses. *Nat. Med.* **29**, 1832–1844 (2023).
21. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
22. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat. Genet.* **55**, 198–208 (2023).
23. Jansen, P. R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
24. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
25. Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
26. Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders

Identification Test (AUDIT) in Two Population-Based Cohorts. *Am. J. Psychiatry* **176**, 107–118 (2019).

27. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
28. Yu, D. *et al.* Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am. J. Psychiatry* **176**, 217–227 (2019).
29. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
30. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
31. Willer, C. J. *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
32. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
33. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
34. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* **43**, 246–252 (2011).
35. das Neves, R. P. *et al.* Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* **8**, e1000560 (2010).
36. Johnston, I. G. *et al.* Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput. Biol.* **8**, e1002416 (2012).
37. Zhu, X., Bührer, C. & Wellmann, S. Cold-inducible proteins CIRP and RBM3, a unique couple with activities far beyond the cold. *Cell. Mol. Life Sci.* **73**, 3839–3859 (2016).
38. Mouta Carreira, C., Landriscina, M., Bellum, S., Prudovsky, I. & Maciag, T. The comparative release of FGF1 by hypoxia and temperature stress. *Growth Factors* **18**, 277–285 (2001).
39. Haque, M., Kendal, J. K., MacIsaac, R. M. & Demetrick, D. J. WSB1: from homeostasis to hypoxia. *J.*

Biomed. Sci. **23**, 61 (2016).

40. Dutta, B., Yan, R., Lim, S. K., Tam, J. P. & Sze, S. K. Quantitative profiling of chromatome dynamics reveals a novel role for HP1BP3 in hypoxia-induced oncogenesis. *Mol. Cell. Proteomics* **13**, 3236–3249 (2014).
41. Boettger, L. M., Handsaker, R. E., Zody, M. C. & McCarroll, S. A. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat. Genet.* **44**, 881–885 (2012).
42. Wade-Martins, R. Genetics: The MAPT locus-a genetic paradigm in disease susceptibility. *Nat. Rev. Neurol.* **8**, 477–478 (2012).
43. Tian, Y. *et al.* Shared Genetics and Comorbid Genes of Amyotrophic Lateral Sclerosis and Parkinson's Disease. *Mov. Disord.* **38**, 1813–1821 (2023).
44. Català-Senent, J. F. *et al.* A deep transcriptome meta-analysis reveals sex differences in multiple sclerosis. *Neurobiol. Dis.* **181**, 106113 (2023).
45. Cooper, Y. A. *et al.* Functional regulatory variants implicate distinct transcriptional networks in dementia. *Science* **377**, eabi8654 (2022).
46. Steinberg, K. M. *et al.* Structural diversity and African origin of the 17q21.31 inversion polymorphism. *Nat. Genet.* **44**, 872–880 (2012).
47. Healy, D. G. *et al.* Tau gene and Parkinson's disease: a case-control study and meta-analysis. *J. Neurol. Neurosurg. Psychiatry* **75**, 962–965 (2004).
48. Zabetian, C. P. *et al.* Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann. Neurol.* **62**, 137–144 (2007).
49. Sánchez-Juan, P. *et al.* The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's Disease in APOE ε4 Non-carriers. *Front. Aging Neurosci.* **11**, 327 (2019).
50. Jovanovic, M. *et al.* Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038 (2015).
51. Emani, P. S. *et al.* Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
52. Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**,

885–892 (2022).

53. Hoffman, G. E. *et al.* Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. *bioRxiv* (2023) doi:10.1101/2023.03.17.533005.
54. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
55. 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).
56. Bhattacherjee, A. *et al.* Cell type-specific transcriptional programs in mouse prefrontal cortex during adolescence and addiction. *Nat. Commun.* **10**, 4169 (2019).
57. Green, G. S. *et al.* Cellular dynamics across aged human brains uncover a multicellular cascade leading to Alzheimer's disease. *bioRxiv* 2023.03.07.531493 (2023) doi:10.1101/2023.03.07.531493.
58. Lee, D. *et al.* Plasticity of Human Microglia and Brain Perivascular Macrophages in Aging and Alzheimer's Disease. *medRxiv* 2023.10.25.23297558 (2023) doi:10.1101/2023.10.25.23297558.
59. Roussos, P. *et al.* Alzheimer's disease transcriptional landscape in ex-vivo human microglia. *Res Sq* (2024) doi:10.21203/rs.3.rs-3851590/v1.
60. Kitano, H. Computational systems biology. *Nature* **420**, 206–210 (2002).
61. Bhembre, N., Bonthron, C. & Opazo, P. Synaptic Compensatory Plasticity in Alzheimer's Disease. *J. Neurosci.* **43**, 6833–6840 (2023).
62. Hernández-Ortega, K., Garcia-Esparcia, P., Gil, L., Lucas, J. J. & Ferrer, I. Altered Machinery of Protein Synthesis in Alzheimer's: From the Nucleolus to the Ribosome. *Brain Pathol.* **26**, 593–605 (2016).
63. Moreira, P. I., Carvalho, C., Zhu, X., Smith, M. A. & Perry, G. Mitochondrial dysfunction is a trigger of Alzheimer's disease pathophysiology. *Biochim. Biophys. Acta* **1802**, 2–10 (2010).
64. Lee, J.-H. *et al.* Faulty autolysosome acidification in Alzheimer's disease mouse models induces autophagic build-up of Aβ in neurons, yielding senile plaques. *Nat. Neurosci.* **25**, 688–701 (2022).
65. Butterfield, D. A. & Halliwell, B. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nat. Rev. Neurosci.* **20**, 148–160 (2019).
66. Mallach, A. *et al.* Microglia-astrocyte crosstalk in the amyloid plaque niche of an Alzheimer's disease

mouse model, as revealed by spatial transcriptomics. *Cell Rep.* **43**, 114216 (2024).

67. Marsh, S. E. *et al.* The adaptive immune system restrains Alzheimer's disease pathogenesis by modulating microglial function. *Proceedings of the National Academy of Sciences* **113**, E1316–E1325 (2016).
68. Grayson, J. M. *et al.* T cell exhaustion is associated with cognitive status and amyloid accumulation in Alzheimer's disease. *Sci. Rep.* **13**, 15779 (2023).
69. Su, W. *et al.* CXCR6 orchestrates brain CD8+ T cell residency and limits mouse Alzheimer's disease pathology. *Nat. Immunol.* **24**, 1735–1747 (2023).
70. Haney, M. S. *et al.* APOE4/4 is linked to damaging lipid droplets in Alzheimer's disease microglia. *Nature* **628**, 154–161 (2024).
71. Marschallinger, J. *et al.* Lipid-droplet-accumulating microglia represent a dysfunctional and proinflammatory state in the aging brain. *Nat. Neurosci.* **23**, 194–208 (2020).
72. Claes, C. *et al.* Plaque-associated human microglia accumulate lipid droplets in a chimeric model of Alzheimer's disease. *Mol. Neurodegener.* **16**, 50 (2021).
73. Lee, D. *et al.* Plasticity of human microglia and brain perivascular macrophages in aging and Alzheimer's disease. *bioRxiv* (2023) doi:10.1101/2023.10.25.23297558.
74. Zhu, J., He, F., Hu, S. & Yu, J. On the nature of human housekeeping genes. *Trends Genet.* **24**, 481–484 (2008).
75. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
76. Jorstad, N. L. *et al.* Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023).
77. Louveau, A. *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337–341 (2015).
78. Ahn, J. H. *et al.* Meningeal lymphatic vessels at the skull base drain cerebrospinal fluid. *Nature* **572**, 62–66 (2019).
79. Li, X. *et al.* Meningeal lymphatic vessels mediate neurotropic viral drainage from the central nervous

system. *Nat. Neurosci.* **25**, 577–587 (2022).

80. Fitzpatrick, Z. *et al.* Venous-plexus-associated lymphoid hubs support meningeal humoral immunity. *Nature* **628**, 612–619 (2024).
81. Da Mesquita, S. *et al.* Meningeal lymphatics affect microglia responses and anti-Aβ immunotherapy. *Nature* **593**, 255–260 (2021).
82. Louveau, A. *et al.* CNS lymphatic drainage and neuroinflammation are regulated by meningeal lymphatic vasculature. *Nat. Neurosci.* **21**, 1380–1391 (2018).
83. Da Mesquita, S., Fu, Z. & Kipnis, J. The Meningeal Lymphatic System: A New Player in Neurophysiology. *Neuron* **100**, 375–388 (2018).
84. Da Mesquita, S. *et al.* Functional aspects of meningeal lymphatics in ageing and Alzheimer's disease. *Nature* **560**, 185–191 (2018).
85. Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* **179**, 312–339 (2019).
86. Ryder, B. D., Wydorski, P. M., Hou, Z. & Joachimiak, L. A. Chaperoning shape-shifting tau in disease. *Trends Biochem. Sci.* **47**, 301–313 (2022).
87. Mok, S.-A. *et al.* Mapping interactions with the chaperone network reveals factors that protect against tau aggregation. *Nat. Struct. Mol. Biol.* **25**, 384–393 (2018).
88. Koren, J., 3rd *et al.* Chaperone signalling complexes in Alzheimer's disease. *J. Cell. Mol. Med.* **13**, 619–630 (2009).
89. Saha, I. *et al.* The AAA+ chaperone VCP disaggregates Tau fibrils and generates aggregate seeds in a cellular system. *Nat. Commun.* **14**, 560 (2023).
90. Brigas, H. C. *et al.* IL-17 triggers the onset of cognitive and synaptic deficits in early stages of Alzheimer's disease. *Cell Rep.* **36**, 109574 (2021).
91. Rosenzweig, N. *et al.* Sex-dependent APOE4 neutrophil-microglia interactions drive cognitive impairment in Alzheimer's disease. *Nat. Med.* (2024) doi:10.1038/s41591-024-03122-3.
92. Cipollini, V., Anrather, J., Orzi, F. & Iadecola, C. Th17 and Cognitive Impairment: Possible Mechanisms of Action. *Front. Neuroanat.* **13**, 95 (2019).

93. Criado-Marrero, M. *et al.* Hsp90 co-chaperones, FKBP52 and Aha1, promote tau pathogenesis in aged wild-type mice. *Acta Neuropathol Commun* **9**, 65 (2021).

94. Gorantla, N. V. & Chinnathambi, S. Tau Protein Squired by Molecular Chaperones During Alzheimer's Disease. *J. Mol. Neurosci.* **66**, 356–368 (2018).

95. Galea, I. The blood-brain barrier in systemic infection and inflammation. *Cell. Mol. Immunol.* **18**, 2489–2501 (2021).

96. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers. Dis.* **64**, S161–S189 (2018).

97. Marquez, D. X. *et al.* Representation of Older Latinxs in Cohort Studies at the Rush Alzheimer's Disease Center. *Neuroepidemiology* **54**, 404–418 (2020).

98. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).

99. Bennett, D. A. *et al.* Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844 (2006).

100. Bennett, D. A. *et al.* Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).

101. Bennett, D. A. *et al.* Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* **27**, 169–176 (2006).

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

103. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

104. Borrell, L. N. *et al.* Race and Genetic Ancestry in Medicine - A Time for Reckoning with Racism. *N. Engl. J. Med.* **384**, 474–480 (2021).

105. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).

106. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol.* **112**, 389–404 (2006).

107. Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the pathologic process in Alzheimer disease: age categories from 1 to 100 years. *J. Neuropathol. Exp. Neurol.* **70**, 960–969 (2011).

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

109. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

110. 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.

111. 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).

112. 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).

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

114. 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).

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

116. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

117. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

118. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018).

119. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
120. Gayoso, A. *et al.* A Python library for probabilistic analysis of single-cell omics data. *Nat. Biotechnol.* **40**, 163–166 (2022).
121. Virshup, I. *et al.* The scverse project provides a computational ecosystem for single-cell omics data analysis. *Nat. Biotechnol.* **41**, 604–606 (2023).
122. Kosoy, R. *et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat. Genet.* **54**, 1145–1154 (2022).
123. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
124. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
125. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
126. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
127. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
128. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
129. Bowles, K. R. *et al.* 17q21.31 sub-haplotypes underlying H1-associated risk for Parkinson's disease are associated with LRRC37A/2 expression in astrocytes. *Mol. Neurodegener.* **17**, 48 (2022).
130. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
131. Hoffman, G. E. & Roussos, P. *Count Ratio Uncertainty Modeling Based Linear Regression*. (2024). doi:10.5281/zenodo.12752107.
132. Lähnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biol.* **21**, 31 (2020).
133. Crowell, H. L. *et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-

condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).

134. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).

135. Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat. Commun.* **12**, 738 (2021).

136. Murphy, A. E., Fancy, N. N. & Skene, N. G. Avoiding false discoveries: Revisiting an Alzheimer's disease snRNA-Seq dataset. *bioRxiv* 2023.04.01.535040 (2023) doi:10.1101/2023.04.01.535040.

137. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).

138. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).

139. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

140. Chia, R. *et al.* Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat. Genet.* **53**, 294–303 (2021).

141. Ferrari, R. *et al.* Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.* **13**, 686–699 (2014).

142. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

143. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157**, 714–725 (2014).

144. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).

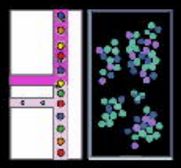
145. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. *arXiv [cs.LG]* (2017).

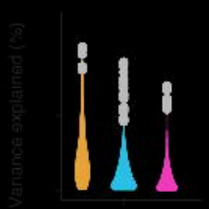
146. Idrissi, B. Y., Arjovsky, M., Pezeshki, M. & Lopez-Paz, D. Simple data balancing achieves competitive worst-group-accuracy. *arXiv [cs.LG]* (2021).

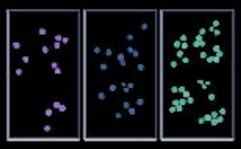
147. Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S. & Chen, C.-F. A new method to measure the semantic

1631 similarity of GO terms. *Bioinformatics* **23**, 1274–1281 (2007).


PsychAD

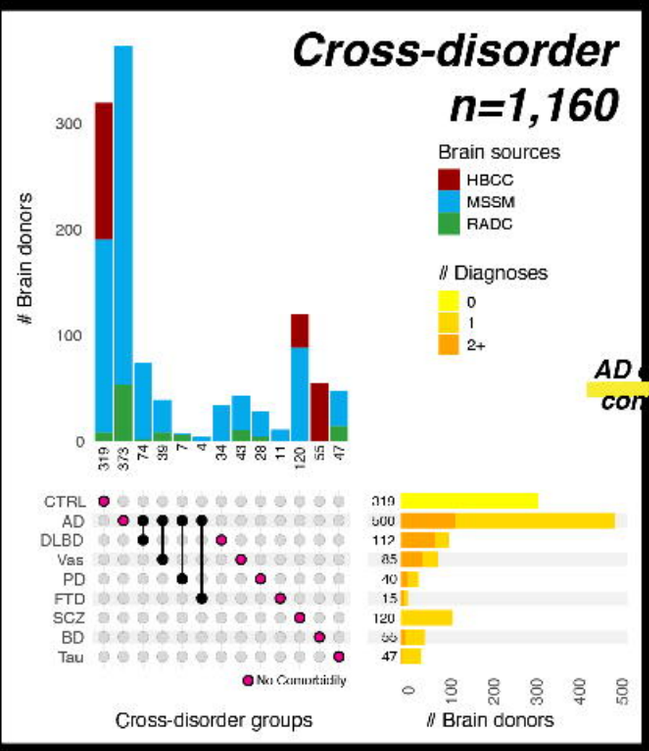


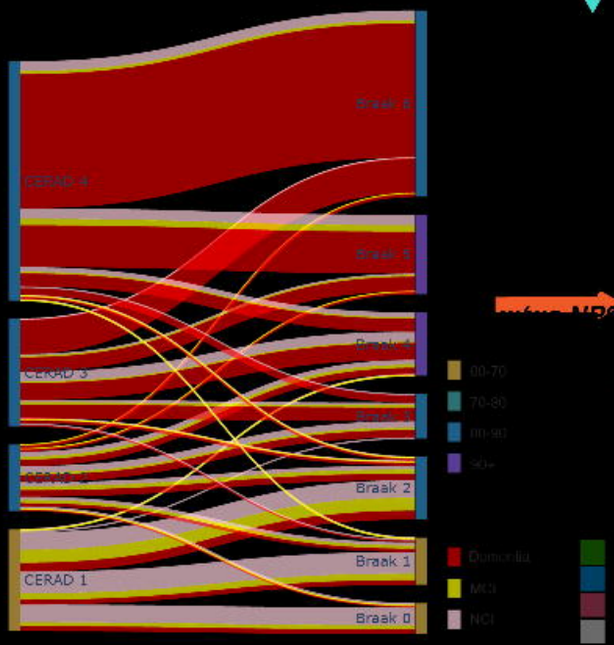

Variance explained (%)

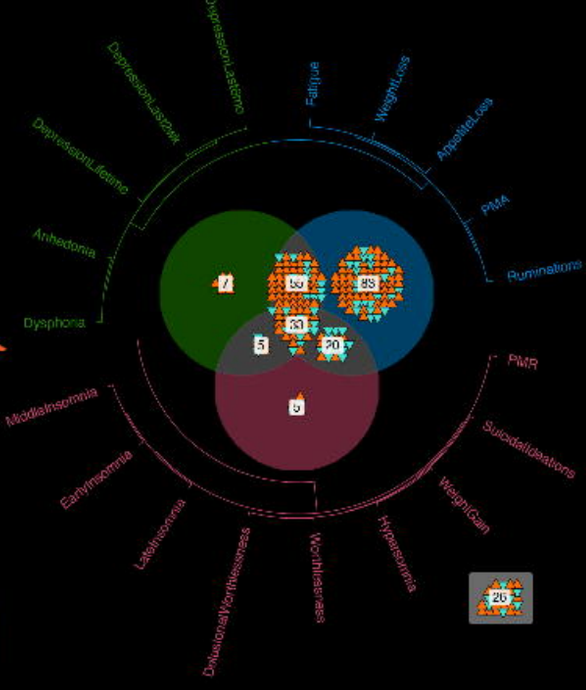


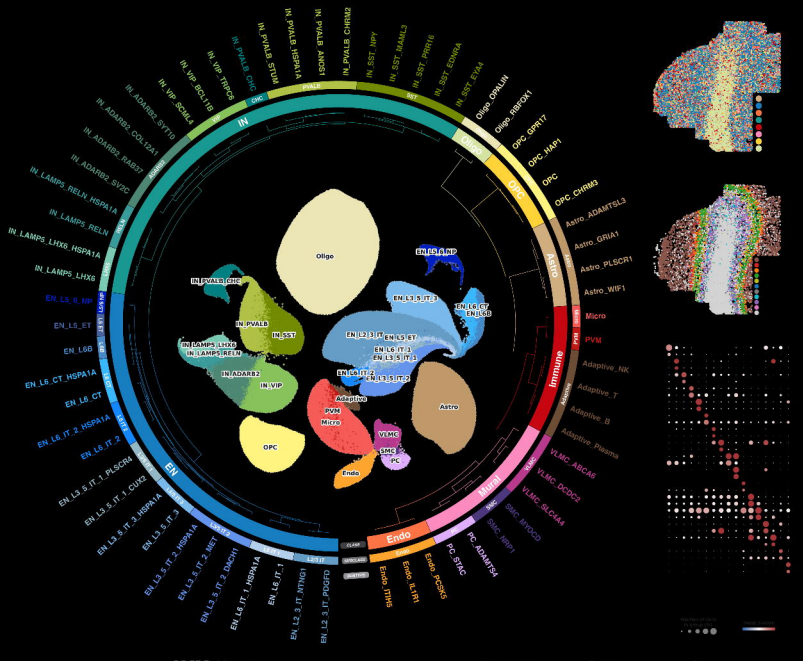

crumblr

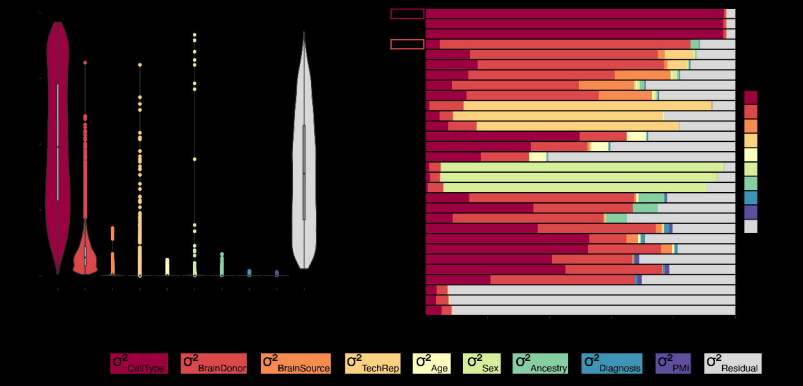

dreamlet

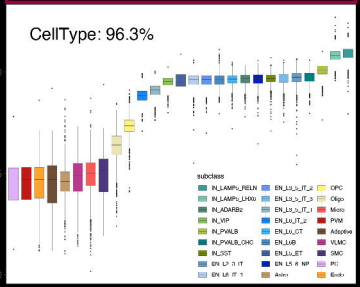

Cross-disorder
n=1,160
Brain sources
HBCC
MSSM
RADC
Diagnoses
0
1
2+
Brain donors
300
200
100
0
319
373
74
39
7
4
34
43
28
11
120
55
47
CTRL
AD
DLBD
Vas
PD
FTD
SCZ
BD
Tau
No Comorbidity
319
600
112
85
40
15
120
55
47
0
100
200
300
400
500
Cross-disorder groups
Brain donors

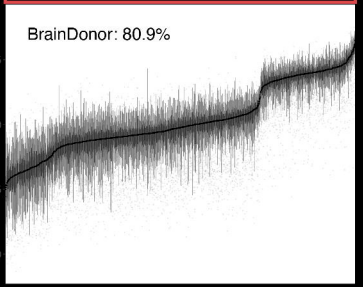

CERAD 4
CERAD 3
CERAD 2
CERAD 1
Braak 6
Braak 5
Braak 4
Braak 3
Braak 2
Braak 1
Braak 0
00-70
70-80
80-90
90+
Dementia
MCI
NCI

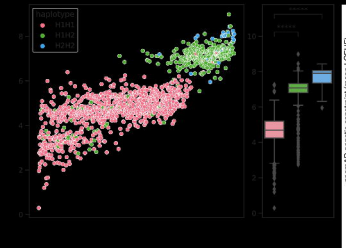

DepressionLastYear
DepressionLifetime
Anhedonia
Dysphoria
MiddleInsomnia
EarlyInsomnia
LateInsomnia
DiurnalMorbidness
Hypersomnia
WeightGain
SuicidalIdeations
PMR
Rumination
PMA
AppetiteLoss
WeightLoss
Fatigue
DepressionLifetime
7
52
83
33
5
20
5
26

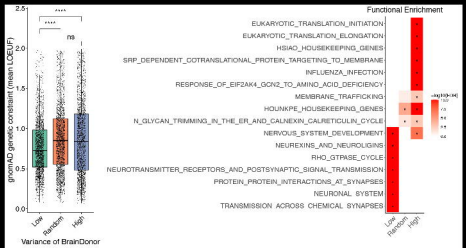

IN
EN
Non-Neuronal
Immune
Vascular
Mural
Endo
Astro
OPC
Oligo
Micro
PVM
Adaptive_NK
Adaptive_T
Adaptive_B
Adaptive_Plasma
VLMC_ABCA6
VLMC_DCDC2
VLMC_SLC4A4
SMC_MYH20
SMC_RBP1
PC_ADAMTSL4
PC_STAC
Endo_RUNX1
Endo_PCDH9
Endo_NTM
Astro_PLSCR1
Astro_WIF1
Astro_GRIA1
Astro_ADAMTSL3
OPC_CHRM3
OPC
OPC_NAP1
OPC_GPR17
Oligo_RBFOX1
OLIG_SPAUN1
IN_SST_EGR4
IN_SST_ERBB4
IN_SST_PRIM1
IN_SST_MAML3
IN_SST_NPY
IN_PVALB_CHRM2
IN_PVALB_ANO3
IN_PVALB_HSPA1A
IN_PVALB_DNM1
IN_PVALB_STUM
IN_PVALB_CHI3
IN_VIP_TRPC6
IN_VIP_SCML1
IN_VIP_BCL11B
IN_ADARB2_SYT6
IN_ADARB2_COL12A1
IN_ADARB2_RASIP1
IN_ADARB2_SYZP
IN_LAMP5_RELN_HSPA1A
IN_LAMP5_RELN
IN_LAMP5_LHX6_HSPA1A
IN_LAMP5_LHX6
EN_LS_6_NP
EN_L5_ET
EN_L6B
EN_L6_CT_HSPA1A
EN_L6_CT
EN_L6_IT_2_HSPA1A
EN_L6_IT_2
EN_L3_5_IT_1_PLSCR1
EN_L3_5_IT_1_CUX2
EN_L3_5_IT_3_HSPA1A
EN_L3_5_IT_3
EN_L3_5_IT_2_HSPA1A
EN_L3_5_IT_2_MEF2C
EN_L5_IT_1_DACH1
EN_L5_IT_1_HSPA1A
EN_L5_IT_1
EN_L2_3_IT_NTNG1
EN_L2_3_IT_PDGFD
EN_L2_3_IT
EN_L3_5_IT_2
EN_L5_IT_3
EN_L5_6_IT_Car3
EN_L5_6_NP
EN_L5_ET
EN_L6B
EN_L6_CT
EN_L6_IT_2
IN_PVALB_CHI3
IN_PVALB
IN_SST
IN_LAMP5_LHX6
IN_LAMP5_RELN
IN_ADARB2
IN_VIP
Adaptive
PVM
Micro
VLMC
SMC
PC
Endo
OPC
Oligo
Astro
Neurological
Psychiatric
Others
GO BP Enrichment
epithelial cell tube commitment
neuron tract neuron morphogenesis
negative regulation of stress fiber assembly
Golgi to plasma membrane protein transport
coronary vasculature morphogenesis
gland morphogenesis
integrin−mediated signaling pathway
cardiac septum morphogenesis
aorta morphogenesis
regulation of integrin−mediated signaling pathway
collagen fibril organization
glial cell differentiation
oligodendrocyte development
regulation of leukocyte mediated cytotoxicity
defense response to bacterium
branching involved in blood vessel morphogenesis
positive regulation of potassium ion transmembrane transport
gamma−aminobutyric acid secretion
mitochondrial ATP synthesis coupled electron transport
dolichol−linked oligosaccharide biosynthetic process
fatty acid beta−oxidation
nuclear−transcribed mRNA catabolic process
regulation of catecholamine secretion
serotonin receptor signaling pathway
anaphase−promoting complex−dependent catabolic process
regulation of cardiac conduction
positive regulation of neurotransmitter secretion
synaptic vesicle exocytosis
synaptic transmission, glutamatergic
glutamate receptor signaling pathway
regulation of postsynaptic membrane potential
neurotransmitter receptor transport, postsynaptic
receptor localization to synapse
dendrite morphogenesis
synchronic glutamate receptor−mediated EPSC pathway
neuron cell−cell adhesion
EN_L3_5_IT_1
EN_L3_5_IT_2
EN_L3_5_IT_3
EN_L5_IT_1
EN_L5_IT_2
EN_L5_IT_3
EN_L6_CT
EN_L6B
IN_PVALB_CHI3
IN_LAMP5_LHX6
IN_LAMP5_RELN
EN_L5_ET
IN_SST
IN_VIP
VLMC
Astro
PC
SMC
Endo
Oligo
Adaptive
Micro
PVM
total
8
4
0
−4
−8

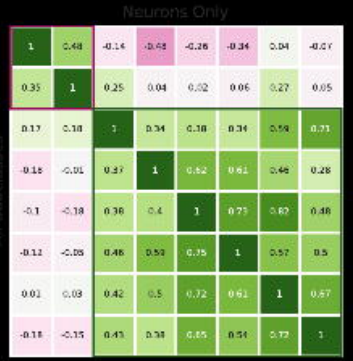

σ² CellType
σ² BrainDonor
σ² BrainSource
σ² TechRep
σ² Age
σ² Sex
σ² Ancestry
σ² Diagnosis
σ² PMI
σ² Residual

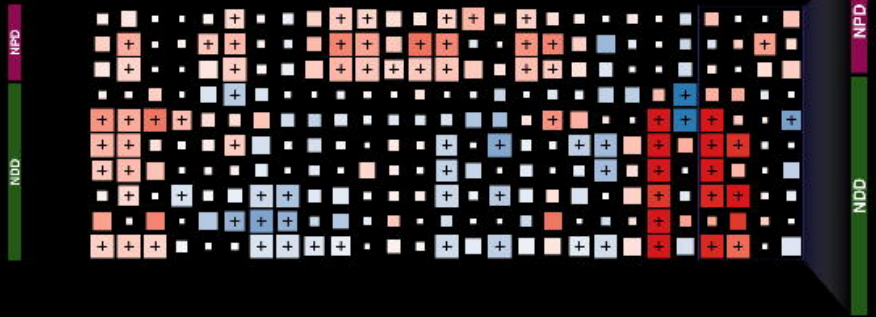

CellType: 96.3%
subclass
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_VIP
IN_PVALB
IN_PVALB_CHC
IN_SST
EN_L2_3_IT_1
EN_L2_3_IT_2
EN_L2_3_IT_3
EN_L3_5_IT_1
EN_L5_IT_2
EN_L5_ET
EN_L5_6_NP
Astro
OPC
Oligo
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo

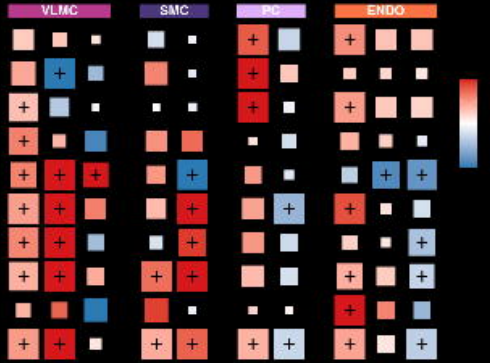

BrainDonor: 80.9%

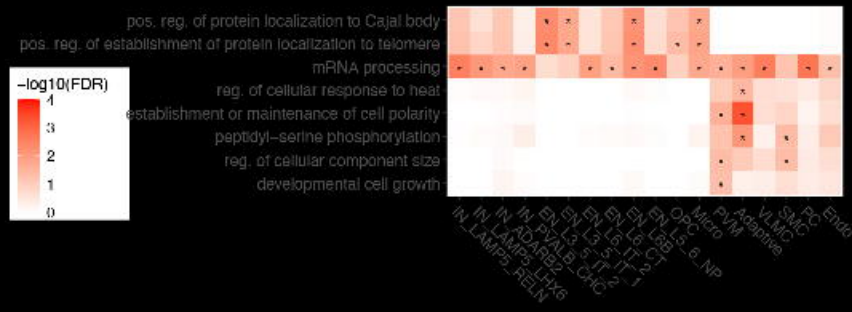

Haplotype
H1H0
H1H1
H2H2

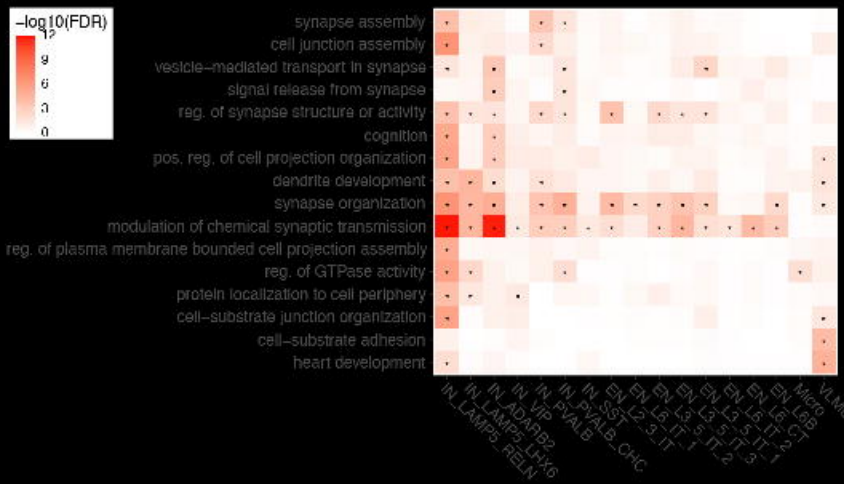

groomD genetic constraint (mean LOEUF)
Low
Medium
High
Variance of BrainDonor
Functional Enrichment
EUKARYOTIC_TRANSLATION_INITIATION
EUKARYOTIC_TRANSLATION_ELONGATION
HSIAO_HOUSEKEEPING_GENES
SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
INFLUENZA_INFECTION
RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
MEMBRANE_TRAFFICKING
HOUNKPE_HOUSEKEEPING_GENES
N_GLYCAN_TRIMMING_IN_THE_ER_AND_CALNEXIN_CALRETICULIN_CYCLE
NERVOUS_SYSTEM_DEVELOPMENT
NEUREXINS_AND_NEUROLIGINS
RHO_GTPASE_CYCLE
NEUROTRANSMITTER_RECEPTORS_AND_POSTSYNAPTIC_SIGNAL_TRANSMISSION
PROTEIN_PROTEIN_INTERACTIONS_AT_SYNAPSES
NEURONAL SYSTEM
TRANSMISSION ACROSS CHEMICAL SYNAPSES
Low
Medium
High

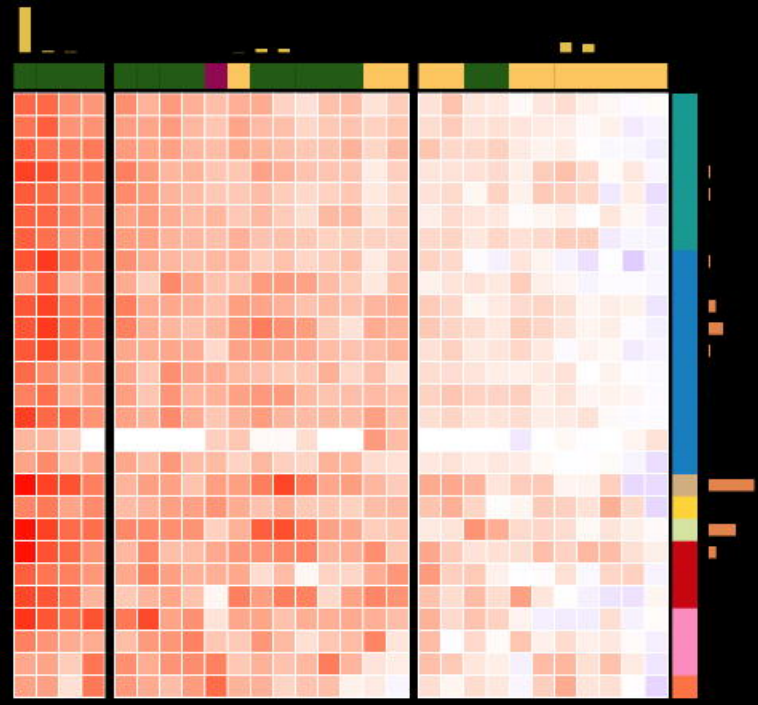

Neurons Only
All Subclasses
NPD
NDD

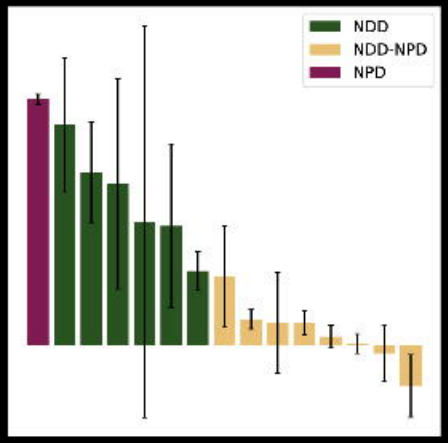

NPD
NDD
NPD
NDD

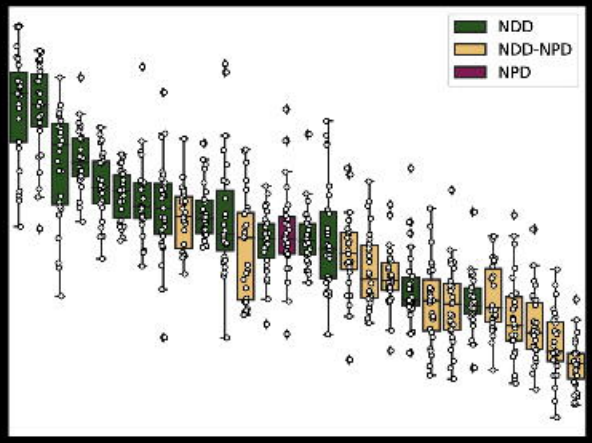

VLMC
SMC
PC
ENDO


pos. reg. of protein localization to Cajal body
pos. reg. of establishment of protein localization to telomere
mRNA processing
reg. of cellular response to heat
establishment or maintenance of cell polarity
peptidyl−serine phosphorylation
reg. of cellular component size
developmental cell growth
−log10(FDR)
4
3
2
1
0
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_PVALB_CHC
EN_L2_3_IT
EN_L5_5_IT_1
EN_L5_5_IT_2
EN_L6_IT_2
EN_L6_CT
EN_L6B
EN_L5_6_NP
OPC
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo

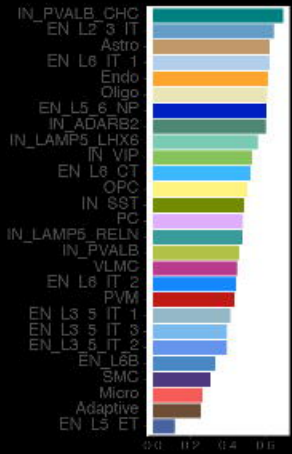

synapse assembly
cell junction assembly
vesicle−mediated transport in synapse
signal release from synapse
reg. of synapse structure or activity
cognition
pos. reg. of cell projection organization
dendrite development
synapse organization
modulation of chemical synaptic transmission
reg. of plasma membrane bounded cell projection assembly
reg. of GTPase activity
protein localization to cell periphery
cell−substrate junction organization
cell−substrate adhesion
heart development
−log10(FDR)
12
9
6
3
0
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_VIP
IN_PVALB
IN_SST
IN_PVALB_CHC
EN_L2_3_IT
EN_L6_IT_1
EN_L5_5_IT_3
EN_L5_5_IT_2
EN_L6_IT_2
EN_L6_CT
EN_L6B
Micro
VLMC

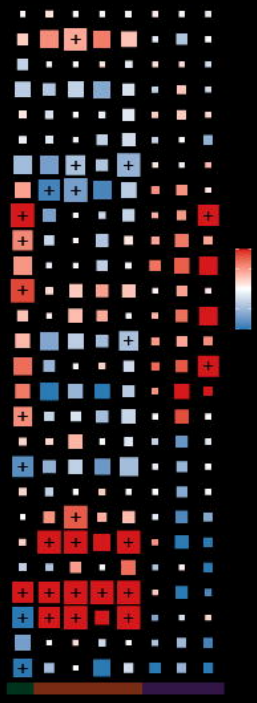


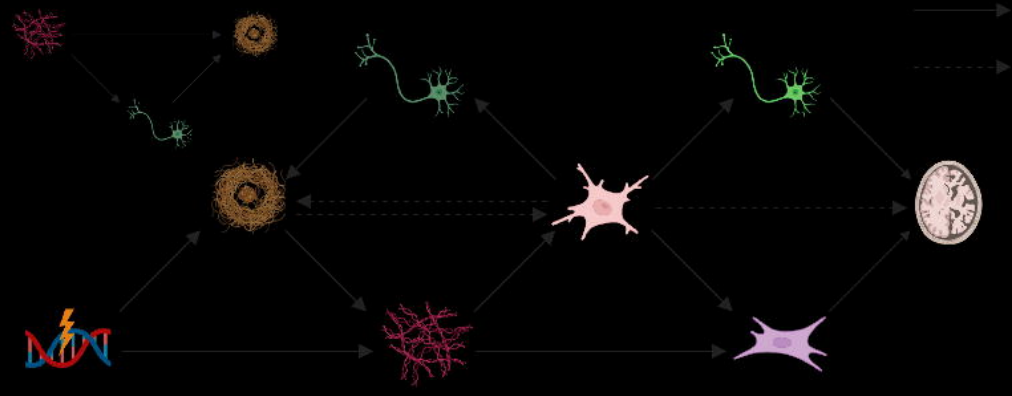

NDD
NDD-NPD
NPD

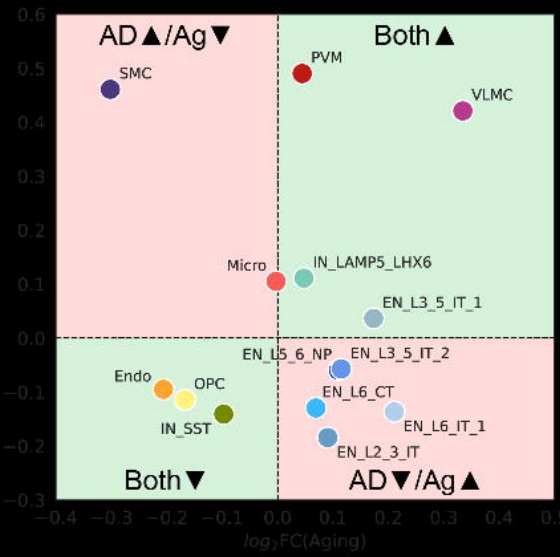

NDD
NDD-NPD
NPD


AD-DLBD
AD-PD
AD-SCZ
DLBD-FTD
DLBD-PD
DLBD-SCZ
AD-FTD
SCZ-B
PD-SCZ
AD-BD
PD-BD
FTD-SCZ
PD-FTD
DLBD-BD
FTD-BD

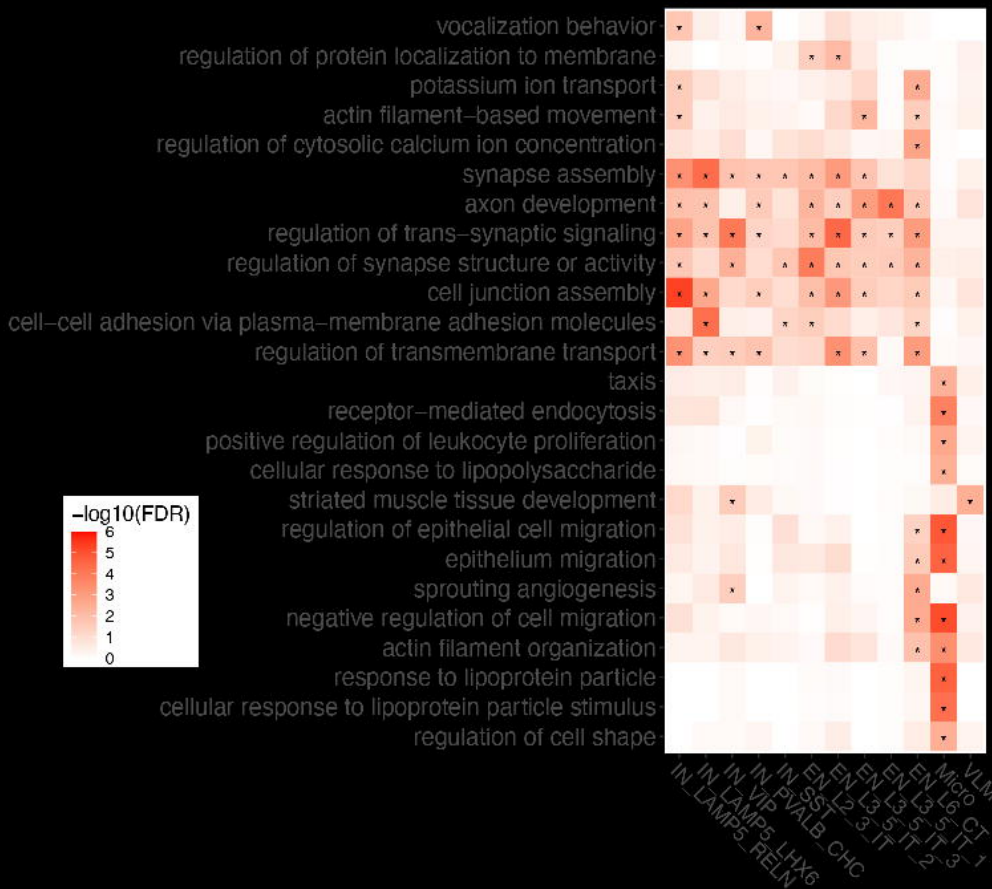

IN_PVALB_CHC
EN_L2_3_IT
Astro
EN_L6_IT_1
Endo
Oligo
EN_L5_6_NP
IN_ADARB2
IN_LAMP5_LHX6
IN_VIP
EN_L6_CT
OPC
IN_SST
PC
IN_LAMP5_RELN
IN_PVALB
VLMC
EN_L6_IT_2
PVM
EN_L3_5_IT_1
EN_L3_5_IT_3
EN_L3_5_IT_2
EN_L6B
SMC
Micro
Adaptive
EN_L5_ET
0.0
0.2
0.4
0.6

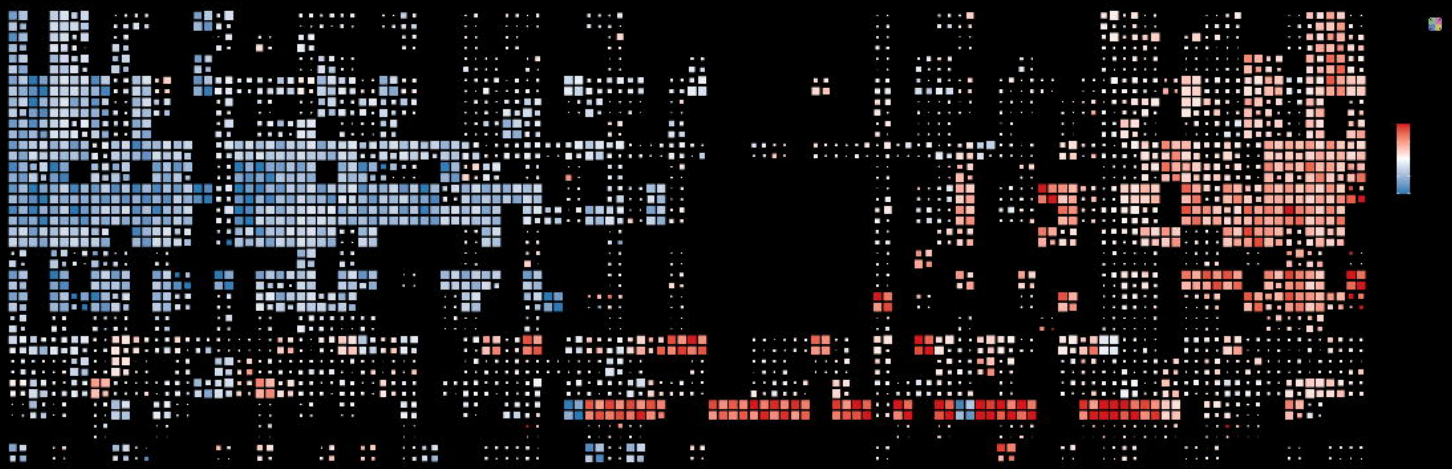


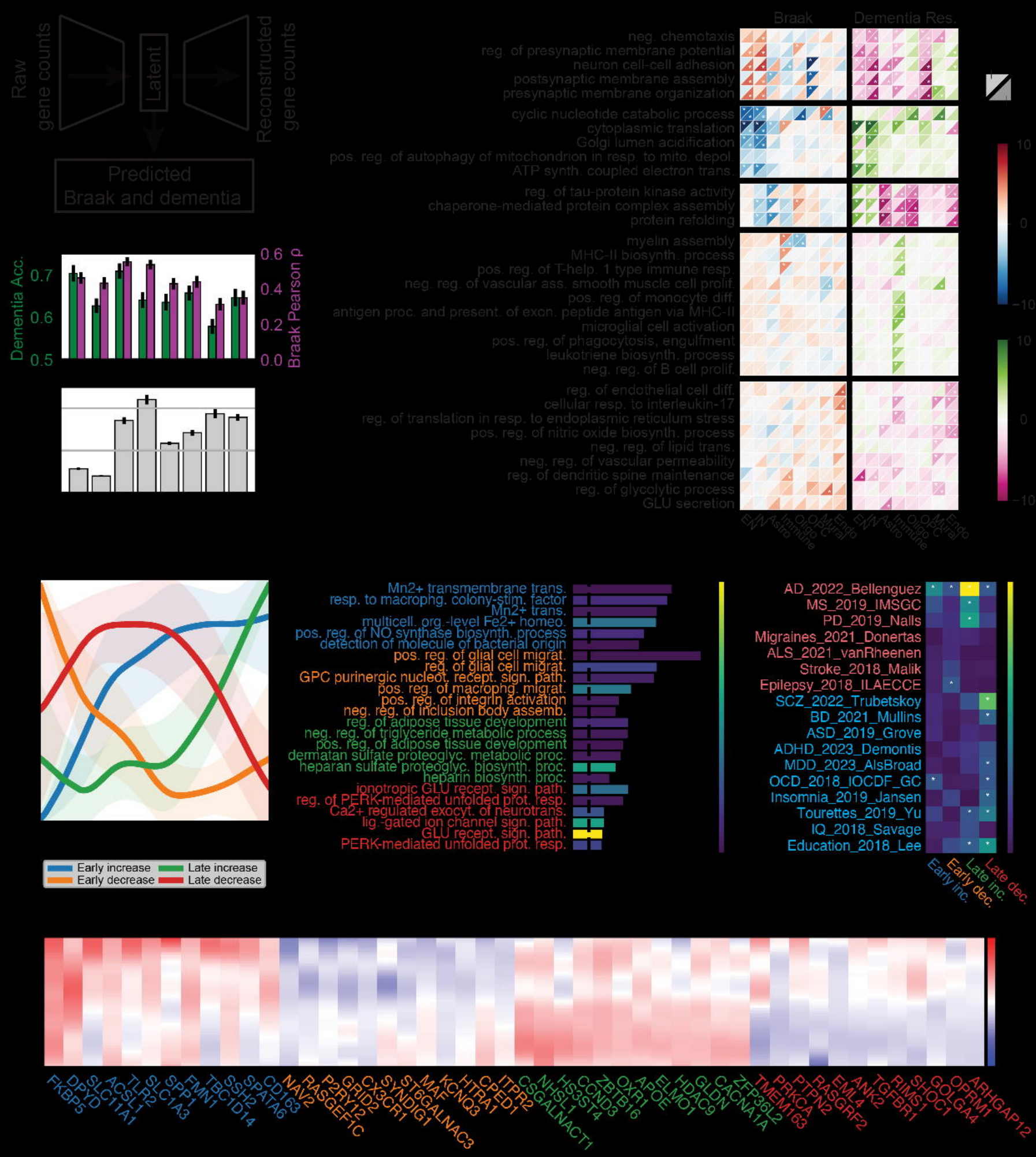


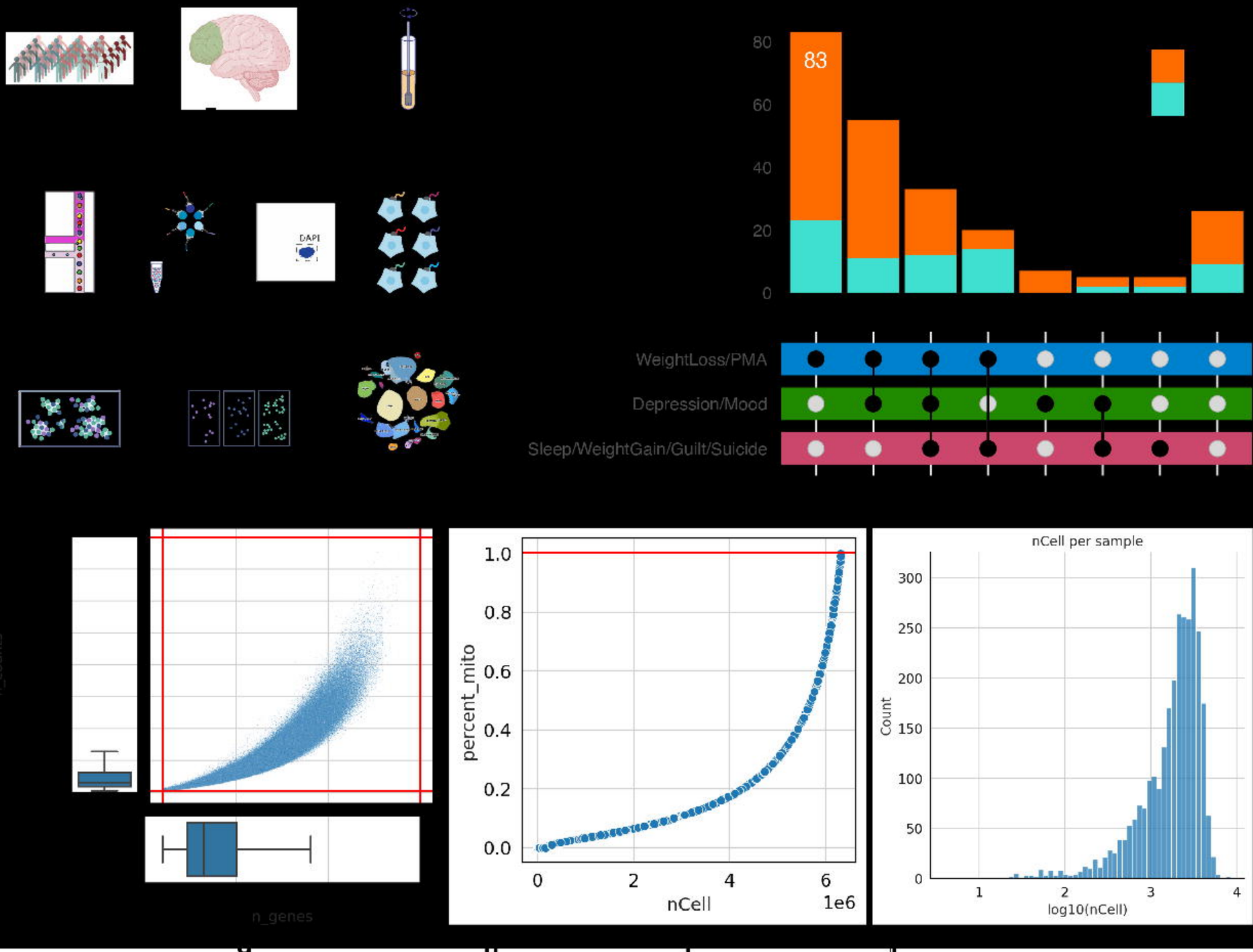

AD▲/Ag▼
Both▲
Both▼
AD▼/Ag▲
SMC
PVM
VLMC
Micro
IN_LAMP5_LHX6
EN_L3_5_IT_1
EN_L5_6_NP
EN_L3_5_IT_2
Endo
OPC
EN_L6_CT
IN_SST
EN_L6_IT_1
EN_L2_3_IT
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4 −0.3 −0.2 −0.1 0.0 0.1 0.2 0.3 0.4 0.5
log2FC(dx_AD)
log2FC(Aging)

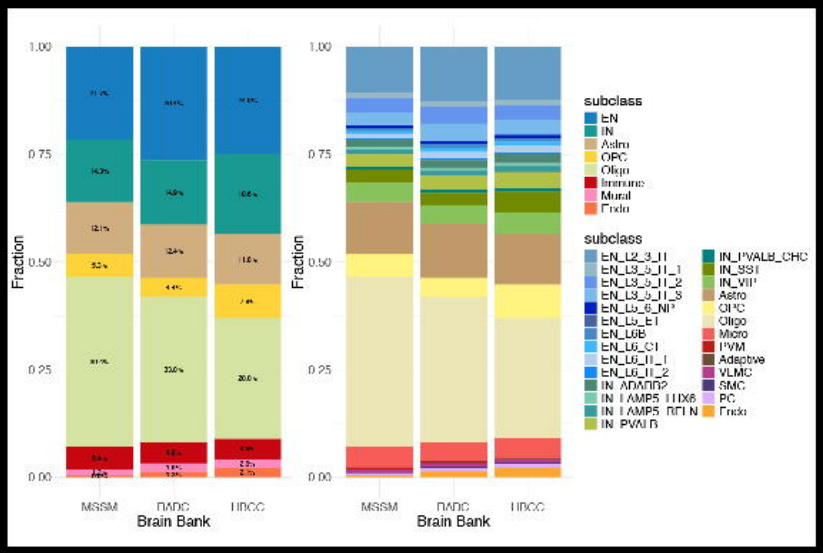
−log10(FDR)
6
5
4
3
2
1
0

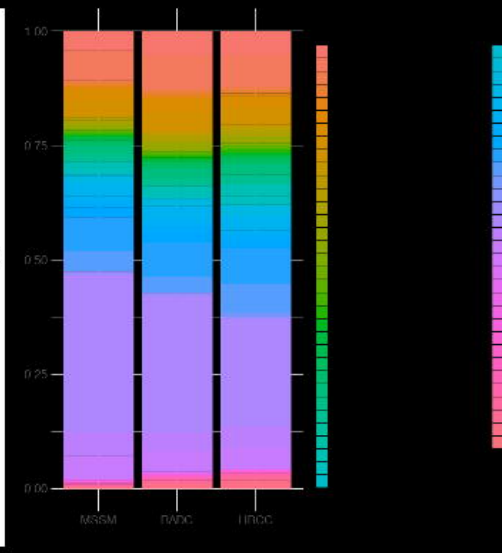

vocalization behavior
regulation of protein localization to membrane
potassium ion transport
actin filament−based movement
regulation of cytosolic calcium ion concentration
synapse assembly
axon development
regulation of trans−synaptic signaling
regulation of synapse structure or activity
cell junction assembly
cell−cell adhesion via plasma−membrane adhesion molecules
regulation of transmembrane transport
taxis
receptor−mediated endocytosis
positive regulation of leukocyte proliferation
cellular response to lipopolysaccharide
striated muscle tissue development
regulation of epithelial cell migration
epithelium migration
sprouting angiogenesis
negative regulation of cell migration
actin filament organization
response to lipoprotein particle
cellular response to lipoprotein particle stimulus
regulation of cell shape
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_PVALB_CHC
IN_SST
IN_VIP
EN_L2_3_IT
EN_L3_5_IT_2
EN_L3_5_IT_3
EN_L6_CT
Micro
VLMC

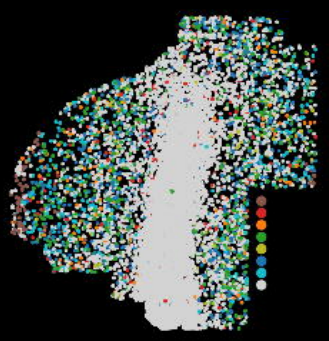


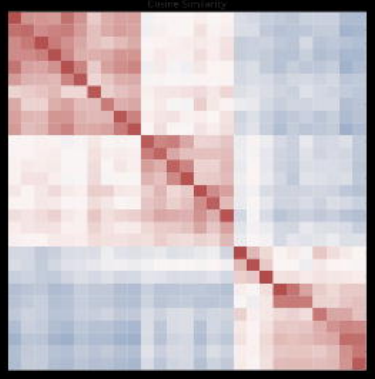

Raw gene counts
Latent
Reconstructed gene counts
Predicted Braak and dementia
Dementia Acc.
Braak Pearson ρ
Braak
Dementia Res.
neg. chemotaxis
reg. of presynaptic membrane potential
neuron cell-cell adhesion
postsynaptic membrane assembly
presynaptic membrane organization
cyclic nucleotide catabolic process
cytoplasmic translation
Golgi lumen acidification
pos. reg. of autophagy of mitochondrion in resp. to mito. depol.
ATP synth. coupled electron trans.
reg. of tau-protein kinase activity
chaperone-mediated protein complex assembly
protein refolding
myelin assembly
MHC-II biosynth. process
pos. reg. of T-help. 1 type immune resp.
neg. reg. of vascular ass. smooth muscle cell prolif.
pos. reg. of monocyte diff.
antigen proc. and present. of exon. peptide antigen via MHC-II
microglial cell activation
pos. reg. of phagocytosis, engulfment
leukotriene biosynth. process
neg. reg. of B cell prolif.
reg. of endothelial cell diff.
cellular resp. to interleukin-17
reg. of translation in resp. to endoplasmic reticulum stress
pos. reg. of nitric oxide biosynth. process
neg. reg. of lipid trans.
neg. reg. of vascular permeability
reg. of dendritic spine maintenance
reg. of glycolytic process
GLU secretion
EN
Astro
Immune
Oligo
OPC
Mural
Endo
Early increase
Late increase
Early decrease
Late decrease
Mn2+ transmembrane trans.
resp. to macrophg. colony-stim. factor
Mn2+ trans.
multicell. org.-level Fe2+ homeo.
pos. reg. of NO synthase biosynth. process
detection of molecule of bacterial origin
pos. reg. of glial cell migrat.
reg. of glial cell migrat.
GPC purinergic nucleot. recept. sign. path.
pos. reg. of macropng. migrat.
pos. reg. of integrin activation
neg. reg. of inclusion body assemb.
reg. of adipose tissue development
neg. reg. of triglyceride metabolic process
pos. reg. of adipose tissue development
dermatan sulfate proteoglyc. metabolic proc.
heparan sulfate proteoglyc. biosynth. proc.
heparin biosynth. proc.
ionotropic GLU recept. sign. path.
reg. of PERK-mediated unfolded prot. resp.
Ca2+ regulated exocyt. of neurotrans.
lig.-gated ion channel sign. path.
GLU recept. sign. path.
PERK-mediated unfolded prot. resp.
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early inc.
Early dec.
Late inc.
Late dec.
FKBP5
DPYD
SLC11A1
ACSL1
TLR2
SLC1A3
SPP1
FMN1
TBC1D14
SSH2
SPATA6
CD163
NAV2
RASGEF1C
P2RY12
GRID2
CX3CR1
SYNDIG1
ST6GALNAC3
MAF
KCNQ3
HTRA1
CPED1
ITPR2
CSGALNACT1
NHSL1
HS3ST4
CCND3
ZBTB16
OXR1
APOE
ELMO1
HDAC9
GLDN
CACNA1A
ZFP36L2
TMEM163
PRKCA
PTPN2
RASGRF2
EML4
ANK2
TGFBR1
RIMS1
SHOC1
GOLGA4
OPRM1
ARHGAP12

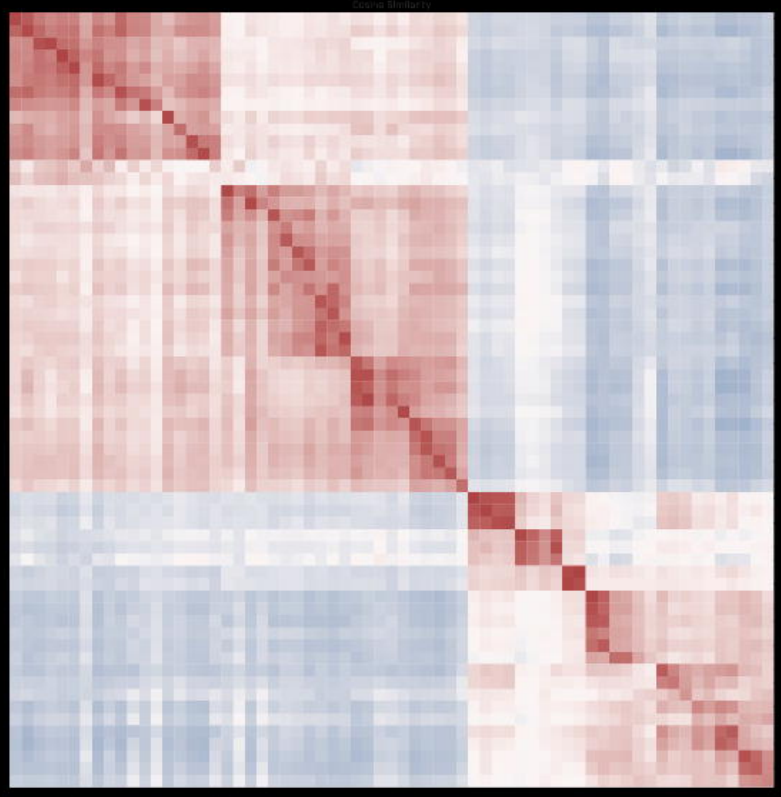

DAPI
83
WeightLoss/PMA
Depression/Mood
Sleep/WeightGain/Guilt/Suicide
n_counts
n_genes
percent_mito
nCell
nCell per sample
Count
log10(nCell)
UMI counts per sample
Gene counts per sample
2924 samples
1494 donors
n_counts
n_genes
nCells
nCells
n_nuclei
Oligo_OPALIN
Oligo
Oligo_RBFOX1
Astro_WIAstro
Astro_PLSCR1
Astro_GRIA1
IN_VIP_TRPC6
IN_VIP
IN_VIP_BCL11B
IN_PVALB
IN_SST
EN
EN_L2_3_IT_2
EN_L3_5_IT_2
OPC
OPC
OPC_HAP1
Immune
Endo

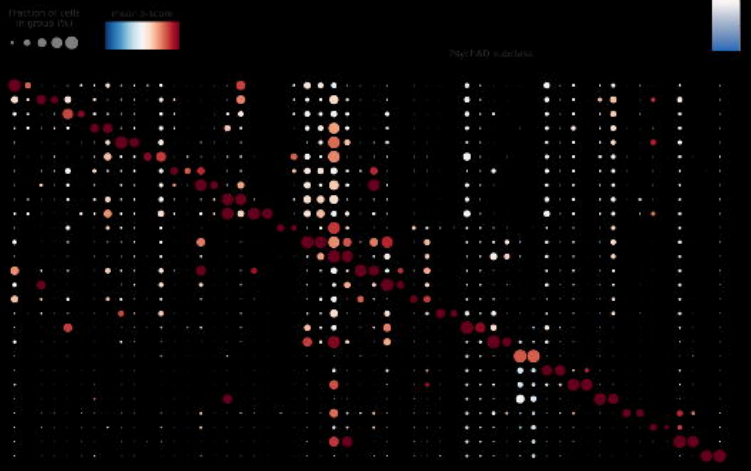

1.00
0.75
0.50
0.25
0.00
Fraction
MSSM
PADC
UPCC
Brain Bank
subclass
EN
IN
Astro
OPC
Oligo
Immune
Mural
Endo
subclass
EN_L2_3_IT
EN_L3_5_IT_1
EN_L3_5_IT_2
EN_L3_5_IT_3
EN_L5_6_NP
EN_L5_ET
EN_L6B
EN_L6_CT
EN_L6_IT_1
EN_L6_IT_2
IN_ADARB2
IN_LAMP5_LHX6
IN_LAMP5_RELN
IN_PVALB
IN_PVALB_CHC
IN_SST
IN_VIP
Astro
OPC
Oligo
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo

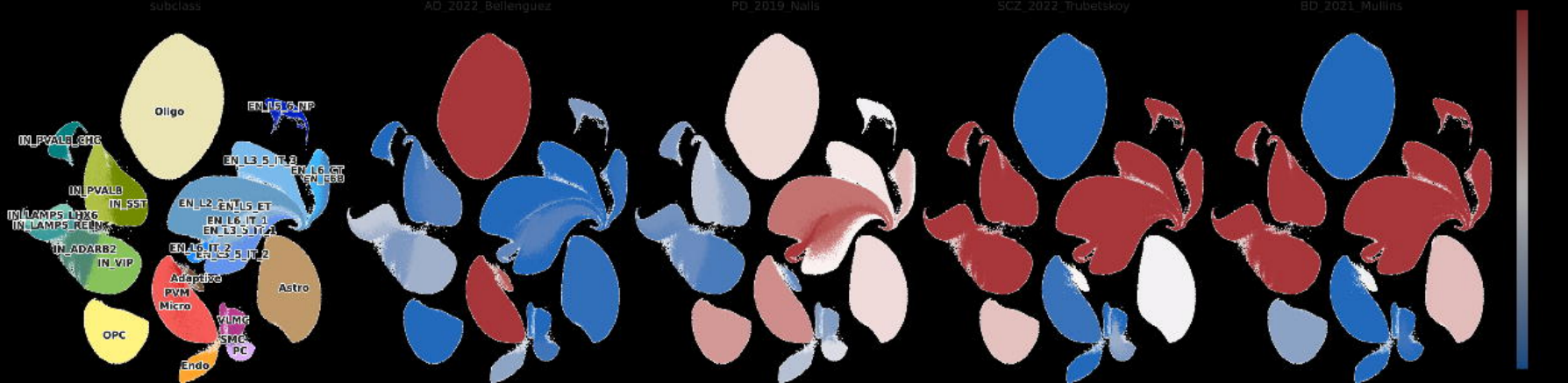

1.00
0.75
0.50
0.25
0.00
MSSM
PADC
UPCC

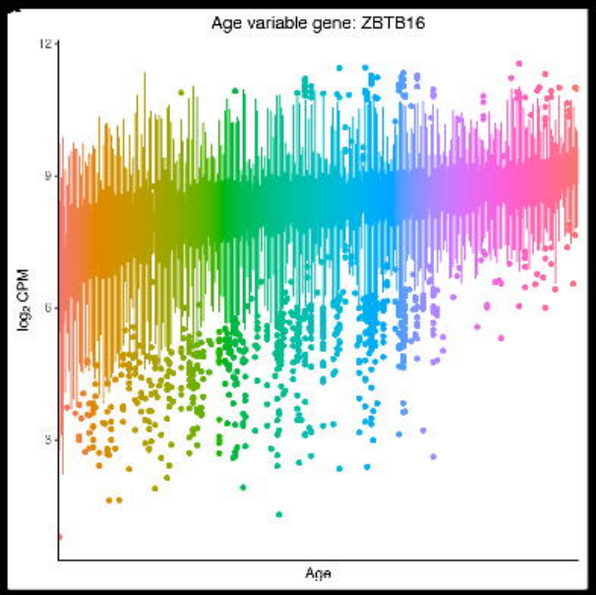


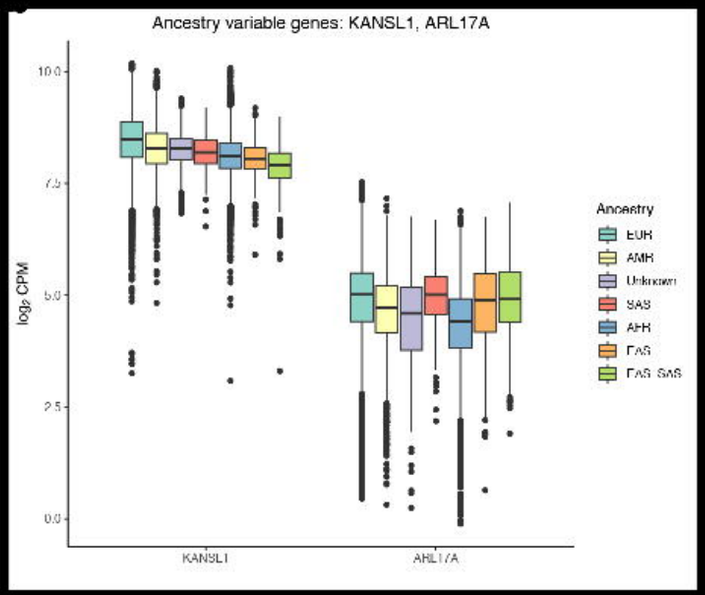


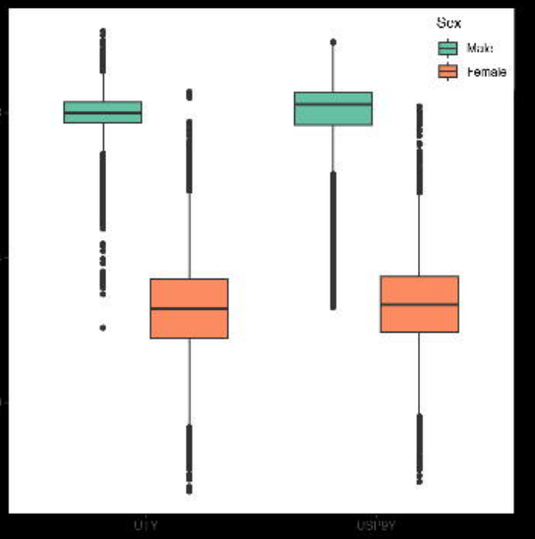


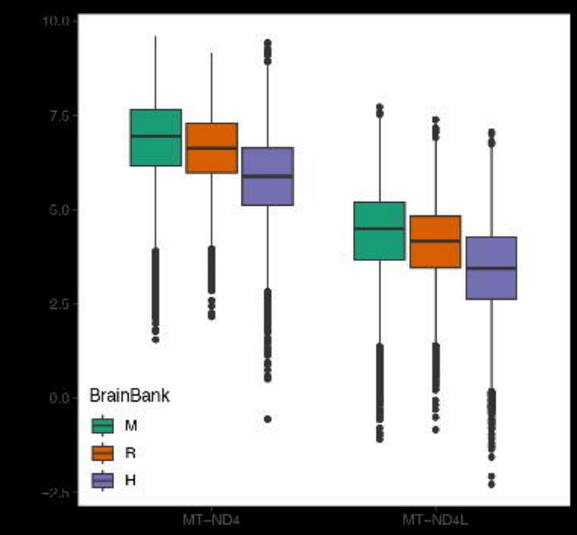


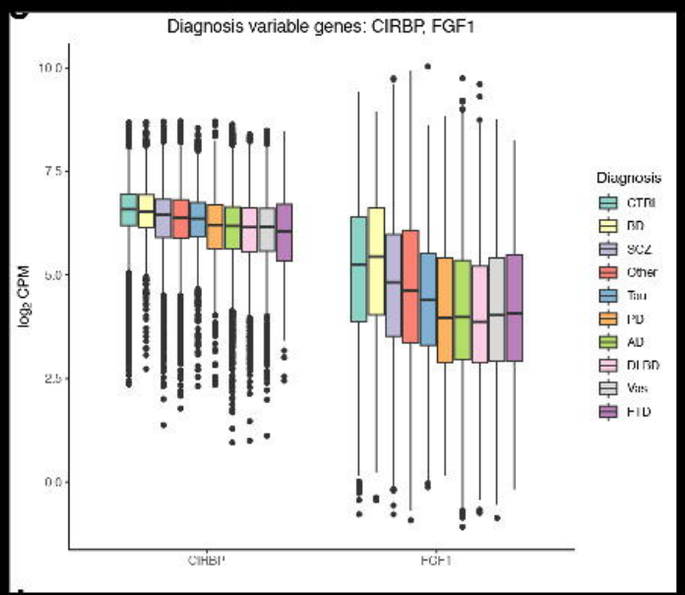

subclass
AD_2022_Bellenguez
PD_2019_Nalls
SCZ_2022_Trubetskoy
BD_2021_Mullins
Oligo
EN_L5_6_NP
IN_PVALB_CHC
EN_L3_5_IT_3
EN_L6_CT
EN_L6B
IN_PVALB
IN_SST
EN_L2_3_IT
EN_L5_ET
EN_L6_IT_1
EN_L3_5_IT_1
IN_LAMP5_LHX6
IN_LAMP5_RELN
EN_L6_IT_2
EN_L3_5_IT_2
IN_ADARB2
IN_VIP
Adaptive
PVM
Micro
Astro
VLMC
SMC
PC
OPC
Endo

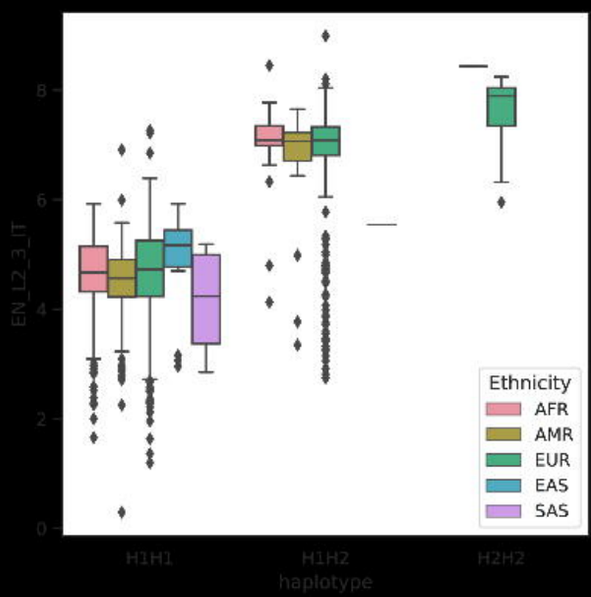

Age variable gene: ZBTB16
log2 CPM
Age

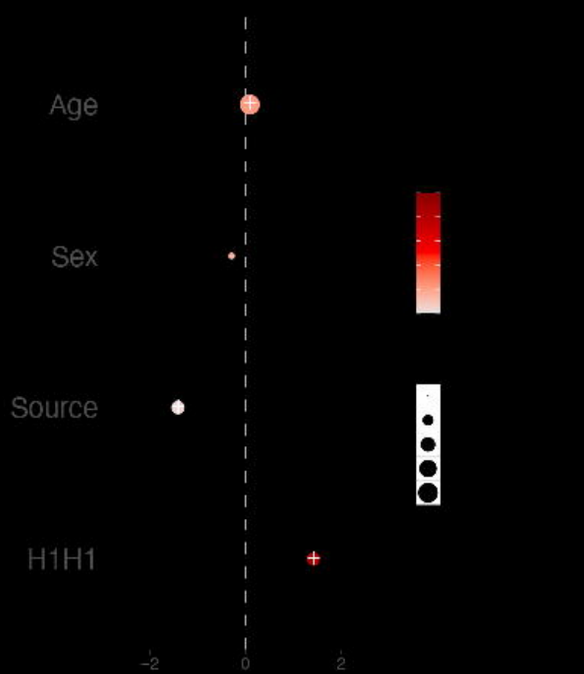

Ancestry variable genes: KANSL1, ARL17A
log2 CPM
KANSL1
ARL17A
Ancestry
EUR
AMR
Unknown
SAS
AFR
EAS
EAS_SAS

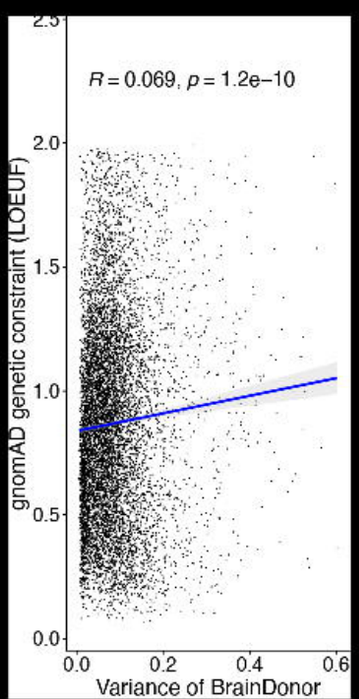

Sex
Male
Female

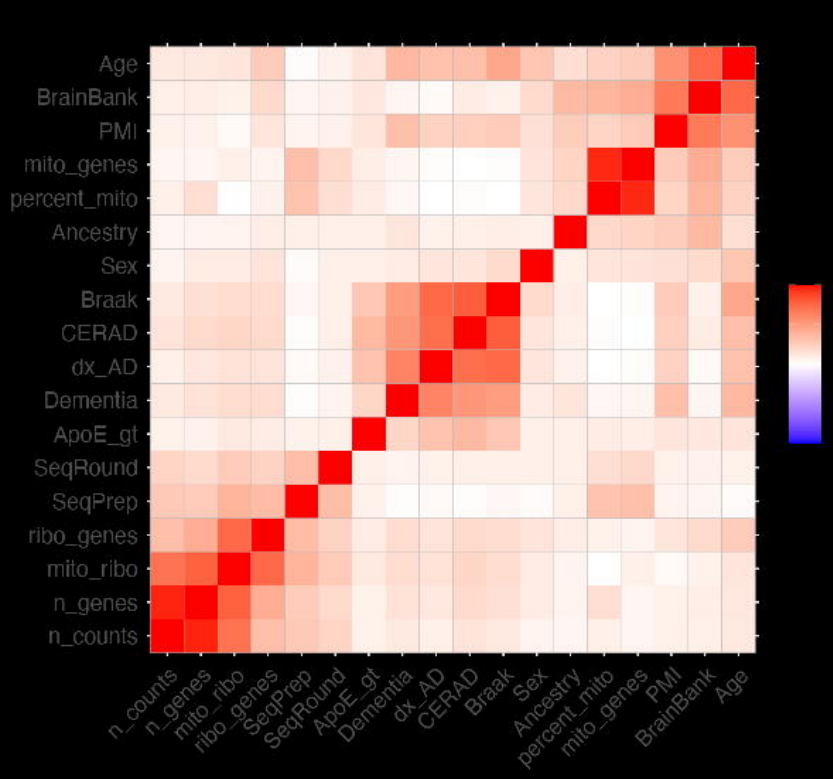

BrainBank
M
R
H
MT-ND4
MT-ND4L

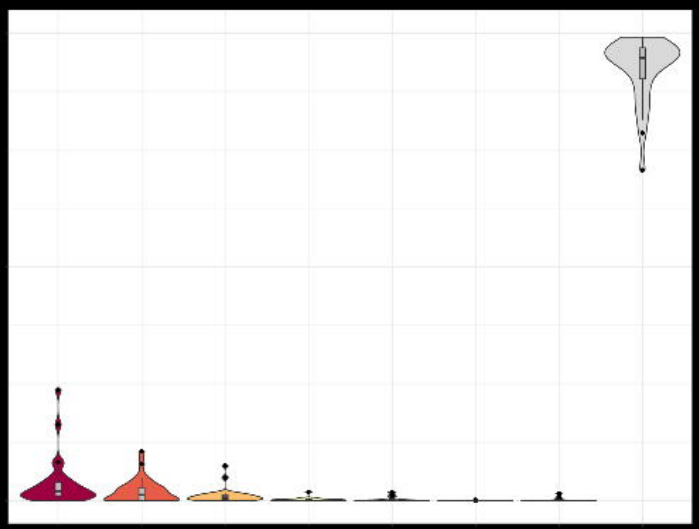

Diagnosis variable genes: CIRBP, FGF1
log2 CPM
Diagnosis
CTRL
BD
SCZ
Other
Tau
PD
AD
DLBD
Vas
FTD

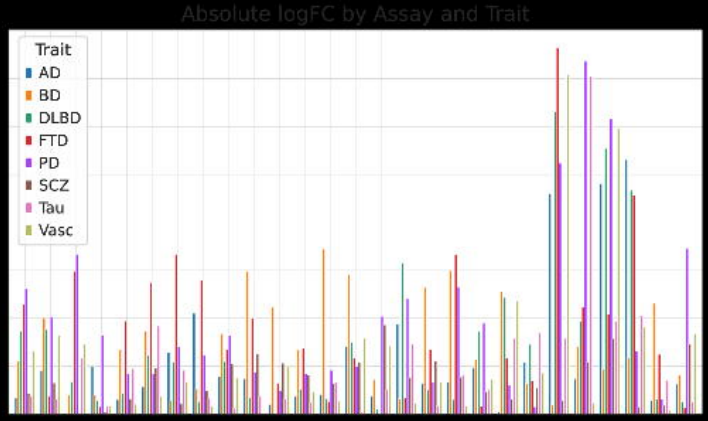

EN_L2_3_IT
Ethnicity
AFR
AMR
EUR
EAS
SAS
H1H1
H1H2
H2H2
haplotype

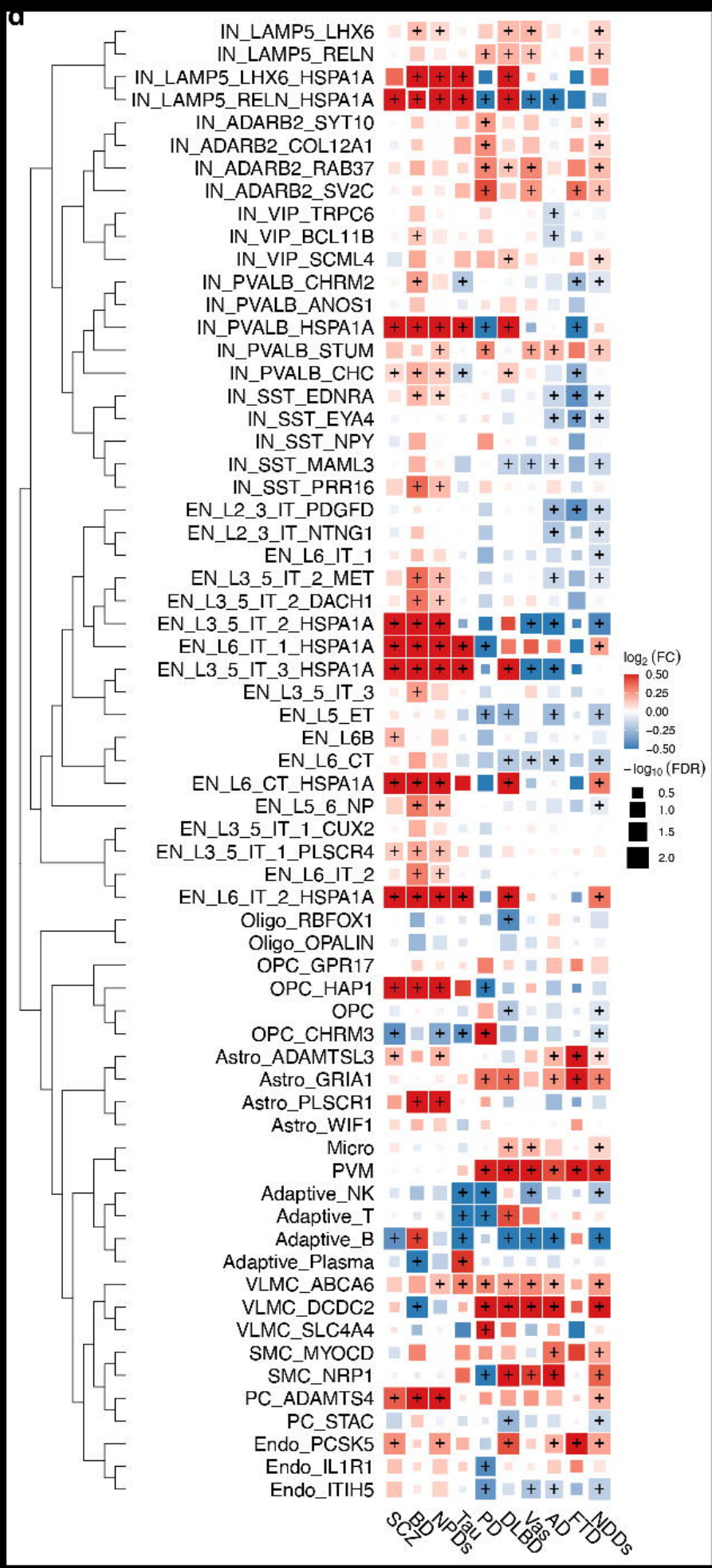

Age
Sex
Source
H1H1


R = 0.069, p = 1.2e−10
gnomAD genetic constraint (LOEUF)
Variance of BrainDonor

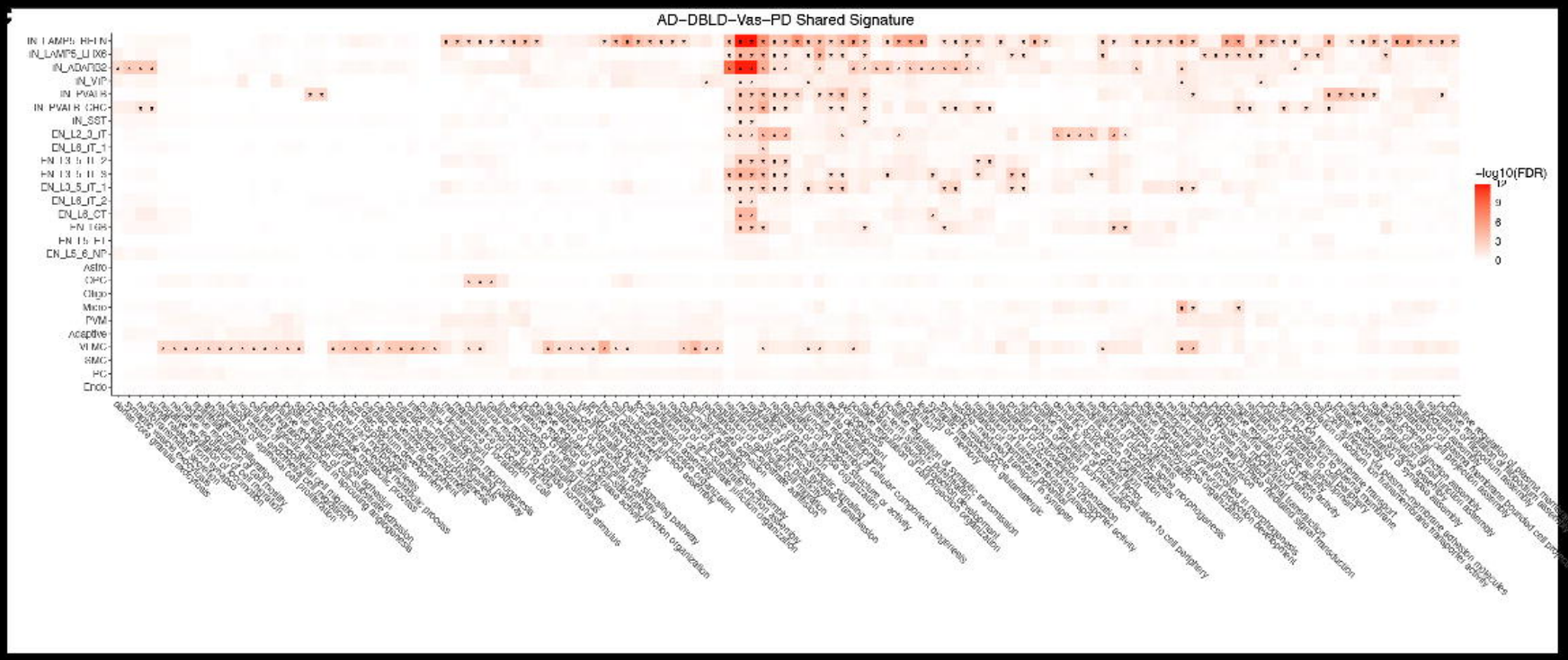

Age
BrainBank
PMI
mito_genes
percent_mito
Ancestry
Sex
Braak
CERAD
dx_AD
Dementia
ApoE_gt
SeqRound
SeqPrep
ribo_genes
mito_ribo
n_genes
n_counts

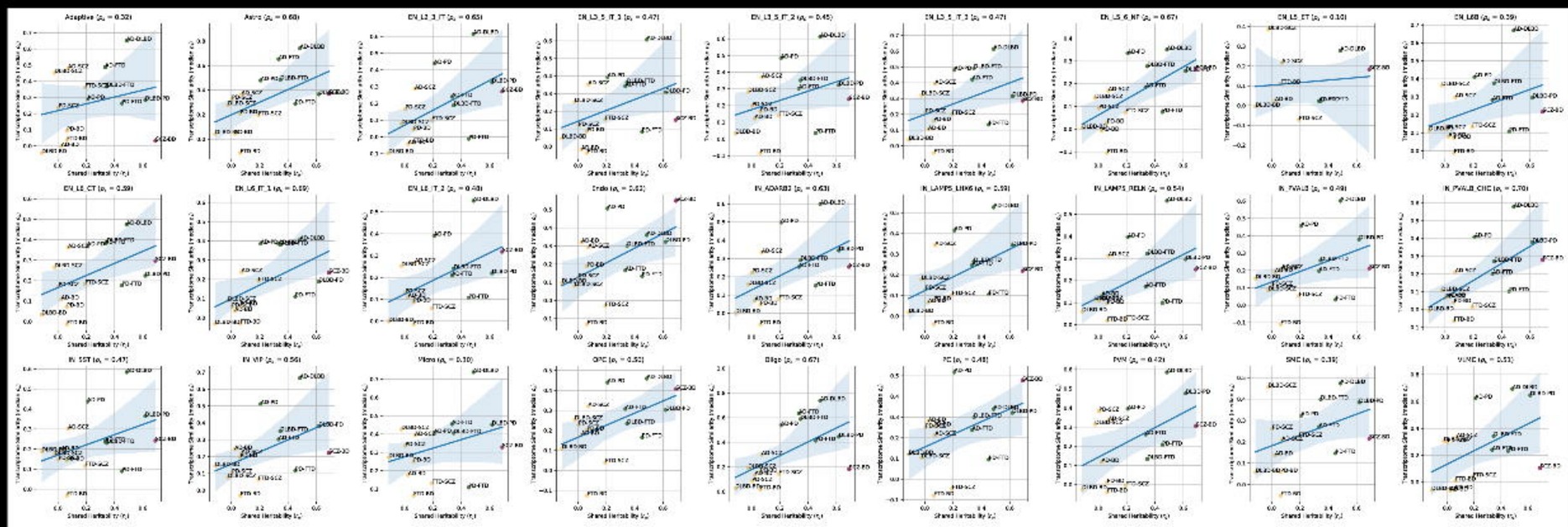


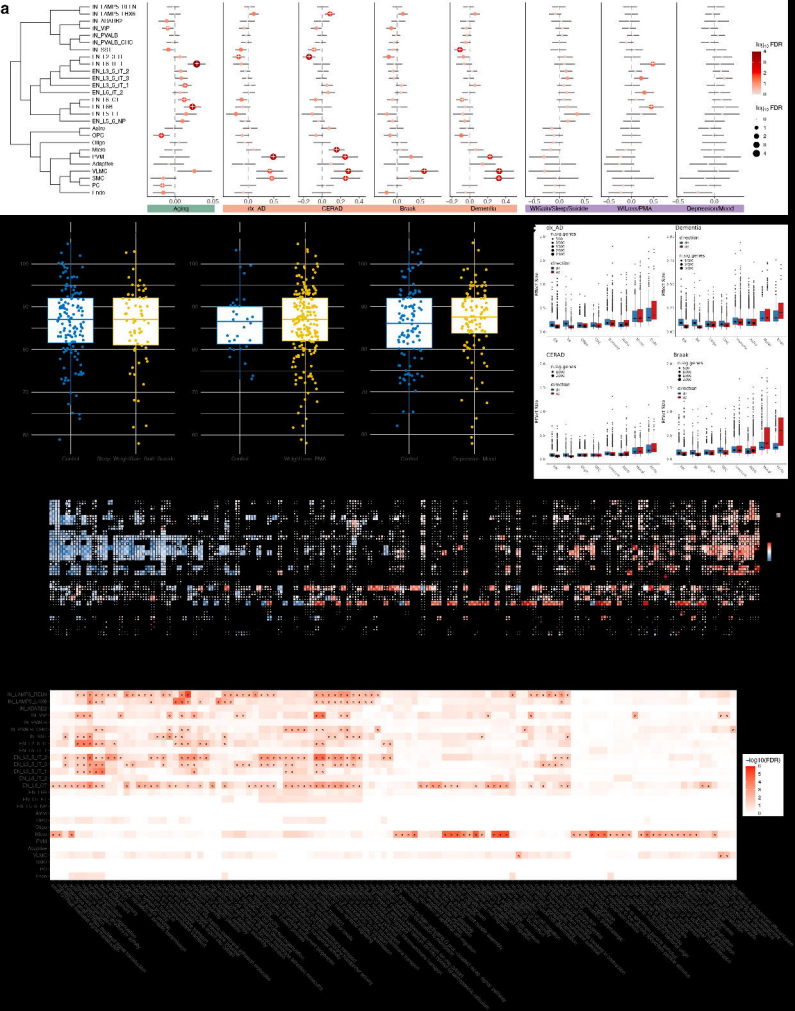

Absolute logFC by Assay and Trait
Trait
AD
BD
DLBD
FTD
PD
SCZ
Tau
Vasc
Absolute logFC


a
IN_LAMP5_LHX6
IN_LAMP5_RELN
IN_LAMP5_LHX6_HSPA1A
IN_LAMP5_RELN_HSPA1A
IN_ADARB2_SYT10
IN_ADARB2_COL12A1
IN_ADARB2_RAB37
IN_ADARB2_SV2C
IN_VIP_TRPC6
IN_VIP_BCL11B
IN_VIP_SCML4
IN_PVALB_CHRM2
IN_PVALB_ANOS1
IN_PVALB_HSPA1A
IN_PVALB_STUM
IN_PVALB_CHC
IN_SST_EDNRA
IN_SST_EYA4
IN_SST_NPY
IN_SST_MAML3
IN_SST_PRR16
EN_L2_3_IT_PDGFD
EN_L2_3_IT_NTNG1
EN_L6_IT_1
EN_L3_5_IT_2_MET
EN_L3_5_IT_2_DACH1
EN_L3_5_IT_2_HSPA1A
EN_L6_IT_1_HSPA1A
EN_L3_5_IT_3_HSPA1A
EN_L3_5_IT_3
EN_L5_ET
EN_L6B
EN_L6_CT
EN_L6_CT_HSPA1A
EN_L5_6_NP
EN_L3_5_IT_1_CUX2
EN_L3_5_IT_1_PLSCR4
EN_L6_IT_2
EN_L6_IT_2_HSPA1A
Oligo_RBFOX1
Oligo_OPALIN
OPC_GPR17
OPC_HAP1
OPC
OPC_CHRM3
Astro_ADAMTSL3
Astro_GRIA1
Astro_PLSCR1
Astro_WIF1
Micro
PVM
Adaptive_NK
Adaptive_T
Adaptive_B
Adaptive_Plasma
VLMC_ABCA6
VLMC_DCDC2
VLMC_SLC4A4
SMC_MYOCD
SMC_NRP1
PC_ADAMTS4
PC_STAC
Endo_PCSK5
Endo_IL1R1
Endo_ITIH5
SCZ
BD
NPDs
Tau
PD
DLBD
Vas
AD
FTD
NDDs
log2 (FC)
0.50
0.25
0.00
−0.25
−0.50
−log10 (FDR)
0.5
1.0
1.5
2.0

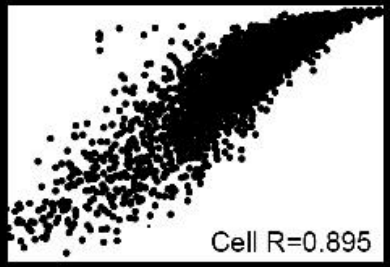

GO0055123: digestive system development
GO0043484: regulation of RNA splicing
GO0048565: digestive tract development
GO0007064: mitotic sister chromatid cohesion
GO0000398: mRNA splicing, via spliceosome
GO0000875: RNA splicing, via transesterification reactions
GO0072655: establishment of protein localization to mitochondrion
GO0050684: regulation of mRNA processing
GO0009409: response to cold
GO0072594: establishment of protein localization to organelle
GO0051641: cellular localization
GO0007399: nervous system development
GO0009966: regulation of signal transduction
GO0044839: cell cycle G2/M phase transition
GO0016050: vesicle organization
GO0019722: calcium-mediated signaling
GO0071417: cellular response to organonitrogen compound
GO0016570: histone modification
GO0007215: glutamate receptor signaling pathway
GO0051146: striated muscle cell differentiation
GO0010876: lipid localization
GO0030048: actin filament-based movement
GO0007416: synapse assembly
−log10(FDR)
6
5
4
3
2
1
0
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_VIP
IN_PVALB
IN_PVALB_CHC
EN_L2_3_IT
EN_L3_5_IT_1
EN_L3_5_IT_2
EN_L3_5_IT_3
EN_L5_ET
EN_L5_6_NP
EN_L6_CT
EN_L6_IT_1
EN_L6_IT_2
EN_L6B
Astro
OPC
Oligo
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo
RNA splicing
reg. of RNA splicing
reg. of mRNA processing
mRNA processing
protein folding
chaperone−mediated protein folding
reg. of protein localization to Cajal body
positive reg. of protein localization to Cajal body
reg. of establishment of protein localization to telomere
pos. reg. of establishment of protein localization to telomere
protein localization to nuclear body
protein localization to Cajal body
reg. of establishment of protein localization to chromosome
pos. reg. of protein localization to chromosome, telomeric region
reg. of cell junction assembly
cell junction assembly
neg. reg. of cytoskeleton organization
establishment or maintenance of cell polarity
neg. reg. of supramolecular fiber organization
peptidyl−serine phosphorylation
RNA splicing, via transesterification reactions w. bulged A as nucleophile
mRNA splicing, via spliceosome
RNA splicing, via transesterification reactions
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_VIP
IN_PVALB
IN_PVALB_CHC
EN_L2_3_IT
EN_L3_5_IT_1
EN_L3_5_IT_2
EN_L3_5_IT_3
EN_L5_ET
EN_L5_6_NP
EN_L6_CT
EN_L6_IT_1
EN_L6_IT_2
EN_L6B
Astro
OPC
Oligo
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo

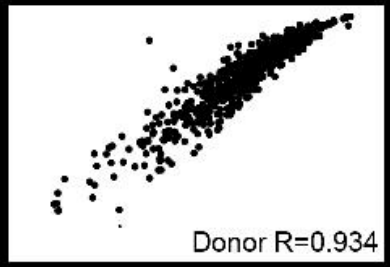

AD−DBLD−Vas−PD Shared Signature
IN_LAMP5_RELN
IN_LAMP5_LHX6
IN_ADARB2
IN_VIP
IN_PVALB
IN_PVALB_CHC
IN_SST
EN_L2_3_IT
EN_L6_IT_1
EN_L3_5_IT_2
EN_L3_5_IT_1
EN_L6_IT_2
EN_L6_CT
EN_L6B
EN_L5_ET
EN_L5_6_NP
Astro
OPC
Oligo
Micro
PVM
Adaptive
VLMC
SMC
PC
Endo
−log10(FDR)
12
9
6
3
0

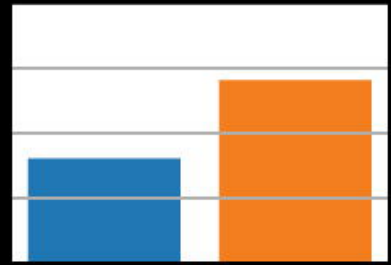

Adaptive (p1 = 0.32)
Astro (p1 = 0.60)
EN_L2_3_IT (p1 = 0.65)
EN_L3_5_IT_3 (p1 = 0.47)
EN_L3_5_IT_2 (p1 = 0.45)
EN_L3_5_IT_1 (p1 = 0.47)
EN_L5_6_NP (p1 = 0.67)
EN_L5_ET (p1 = 0.30)
EN_L6B (p1 = 0.39)
EN_L6_CT (p1 = 0.39)
EN_L6_IT_1 (p1 = 0.59)
EN_L6_IT_2 (p1 = 0.48)
Endo (p1 = 0.62)
IN_ADARB2 (p1 = 0.63)
IN_LAMP5_LHX6 (p1 = 0.59)
IN_LAMP5_RELN (p1 = 0.54)
IN_PVALB (p1 = 0.49)
IN_PVALB_CHC (p1 = 0.70)
IN_SST (p1 = 0.47)
IN_VIP (p1 = 0.56)
Micro (p1 = 0.30)
OPC (p1 = 0.52)
Oligo (p1 = 0.67)
PC (p1 = 0.48)
PVM (p1 = 0.42)
SMC (p1 = 0.39)
VLMC (p1 = 0.53)
Shared Heritability (r1)
Transcriptome−Wide Heritability (r2)


a
Aging
tlx_AD
CERAD
Braak
Dementia
WMcoxSteg/Scalice
WMcoxPMA
DepressionMixed
log10 FDR
Control
Resp Weight/Gum Graft Suicide
Control
Weight/Gum PMA
Control
Depression Mixed
tlx_AD
Dementia
CERAD
Braak
direction
up
down
up genes
down genes
Effect Size

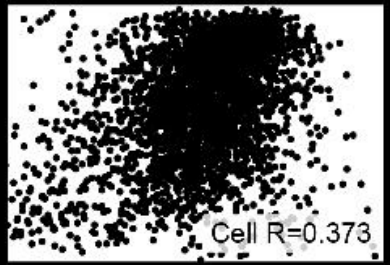
Standard training

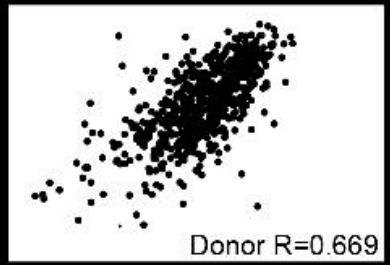


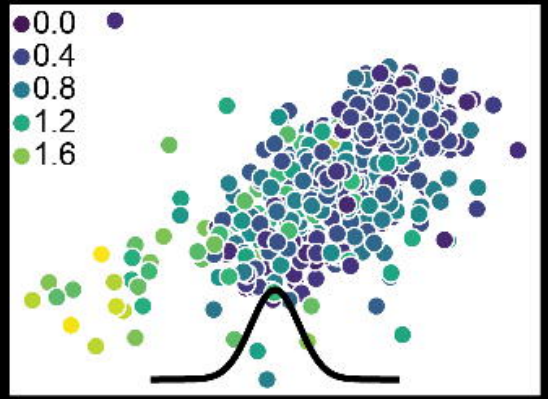


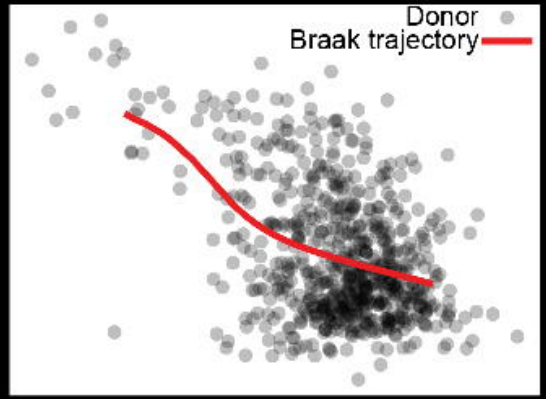


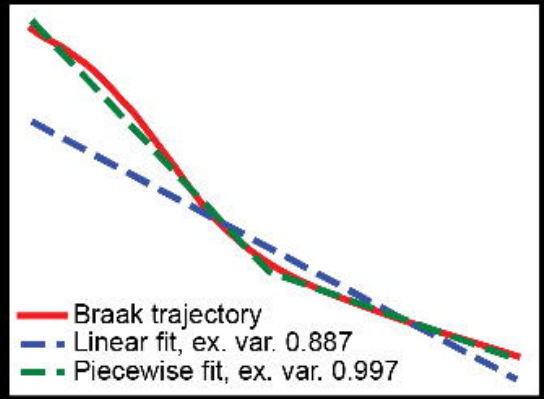
Equal sampling

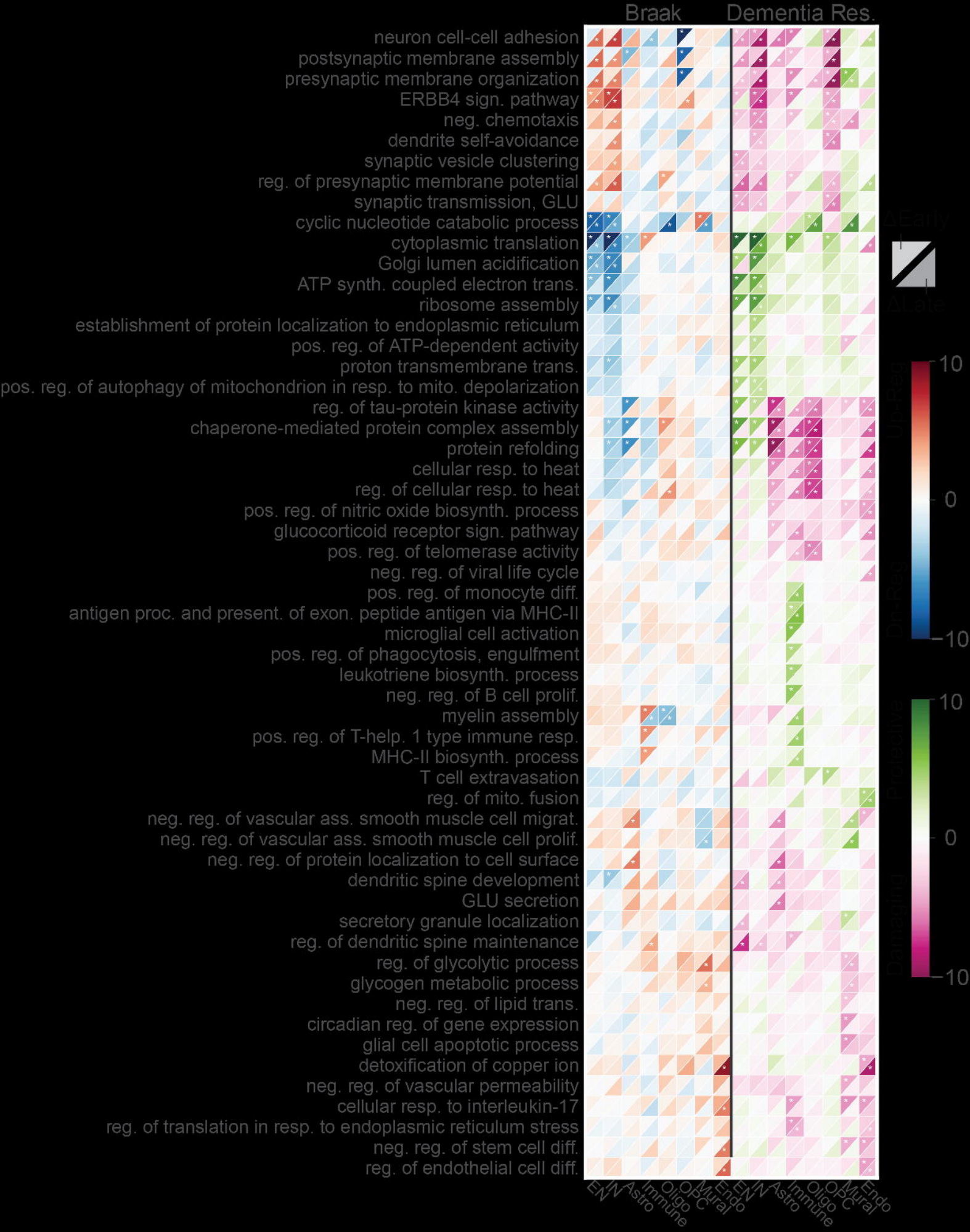


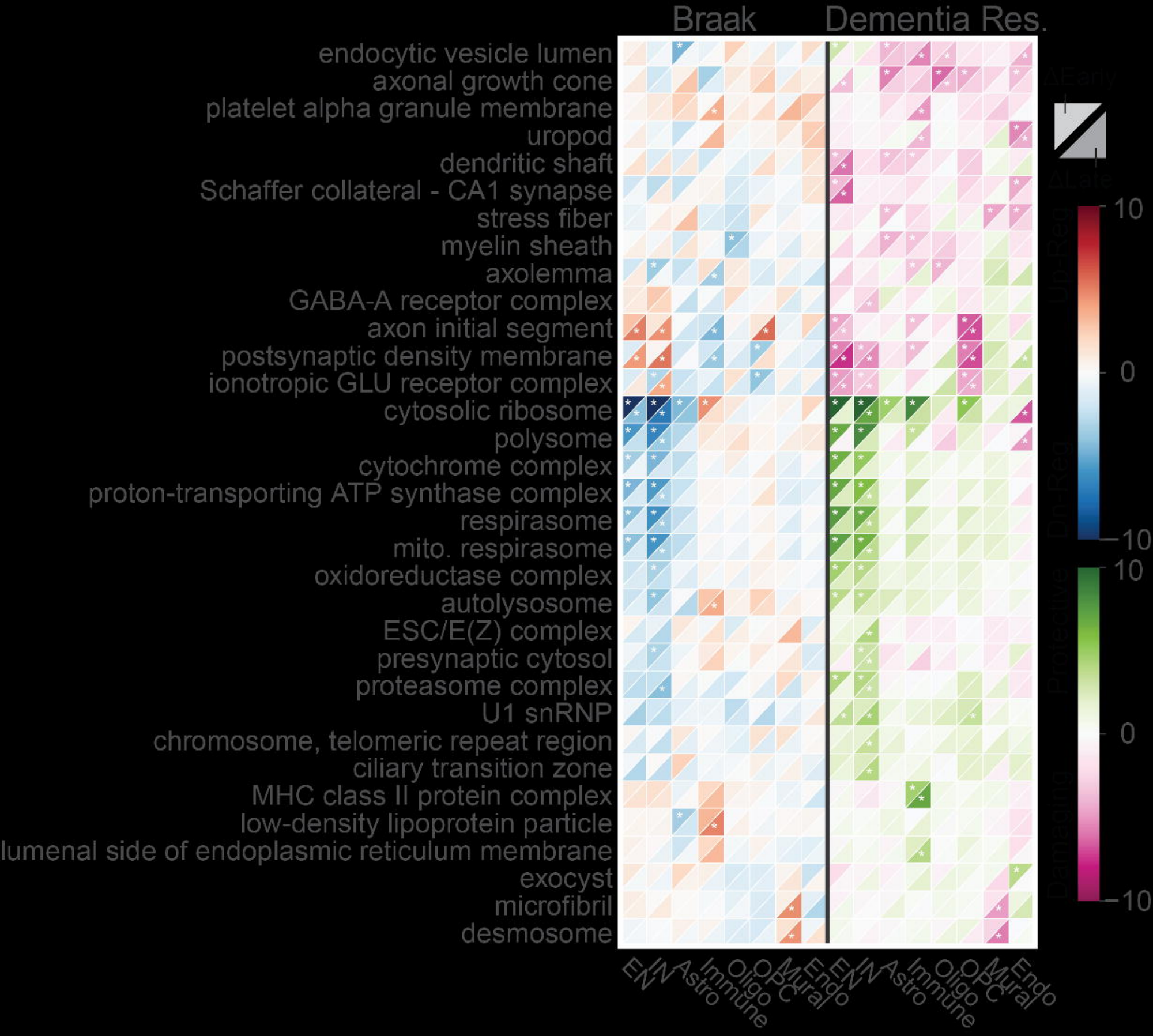



0.0
0.4
0.8
1.2
1.6

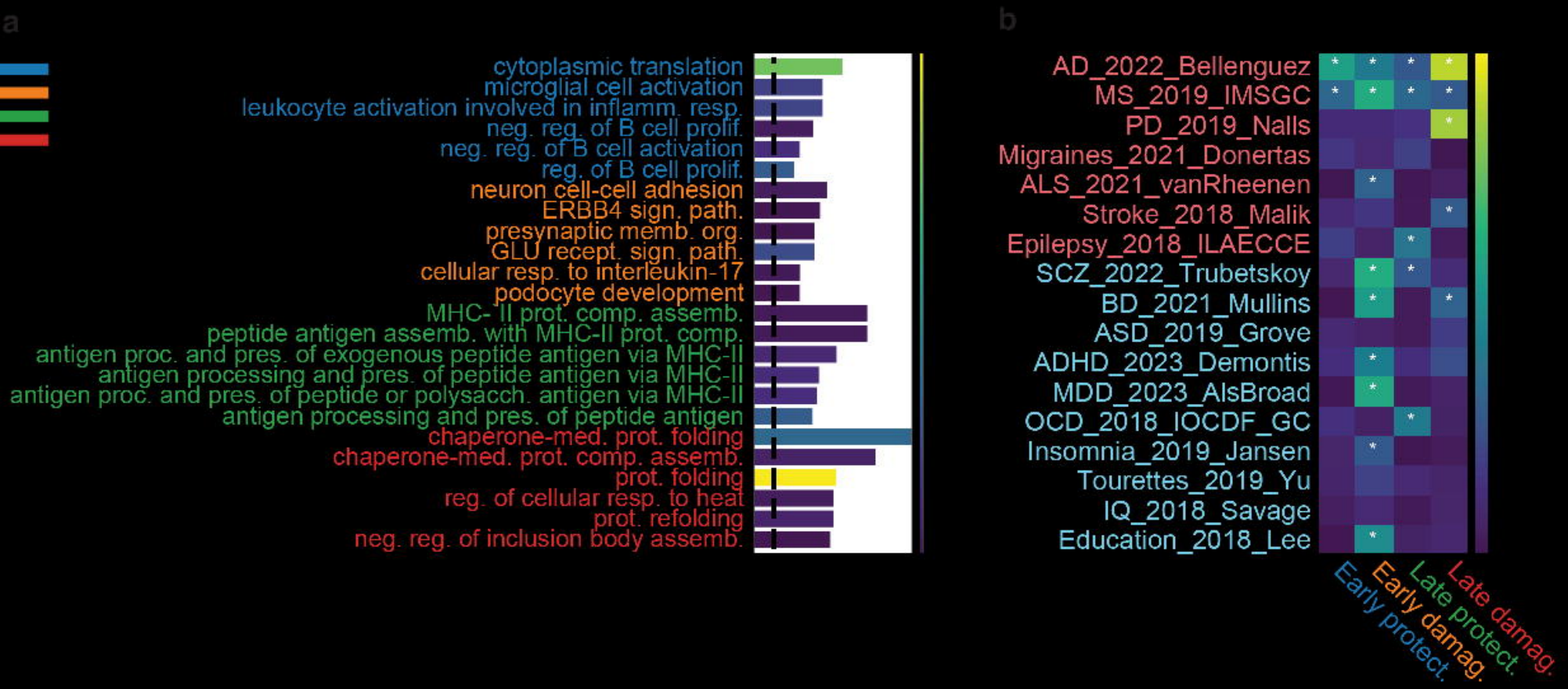

Donor
Braak trajectory

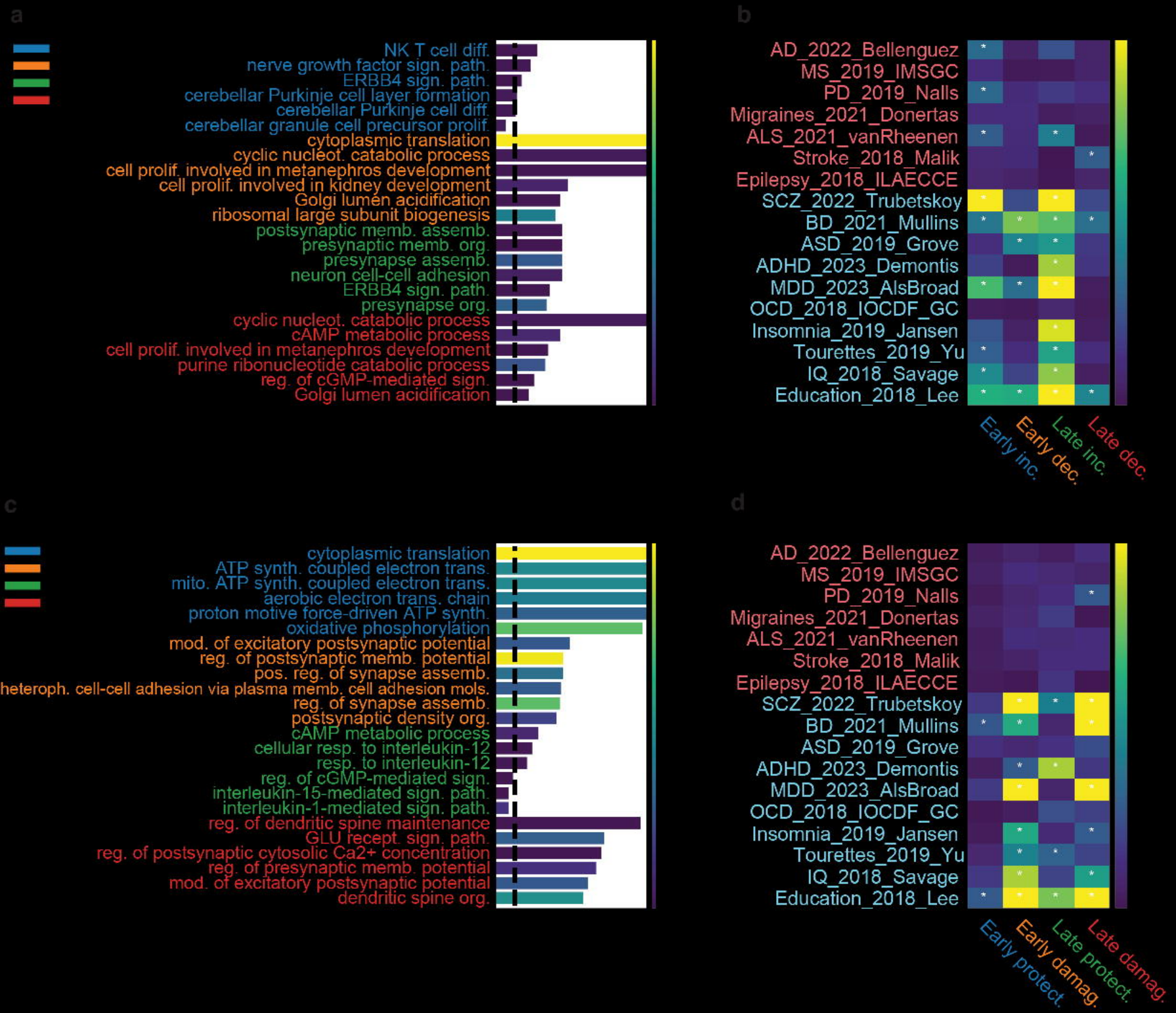

Braak trajectory
Linear fit, ex. var. 0.887
Piecewise fit, ex. var. 0.997

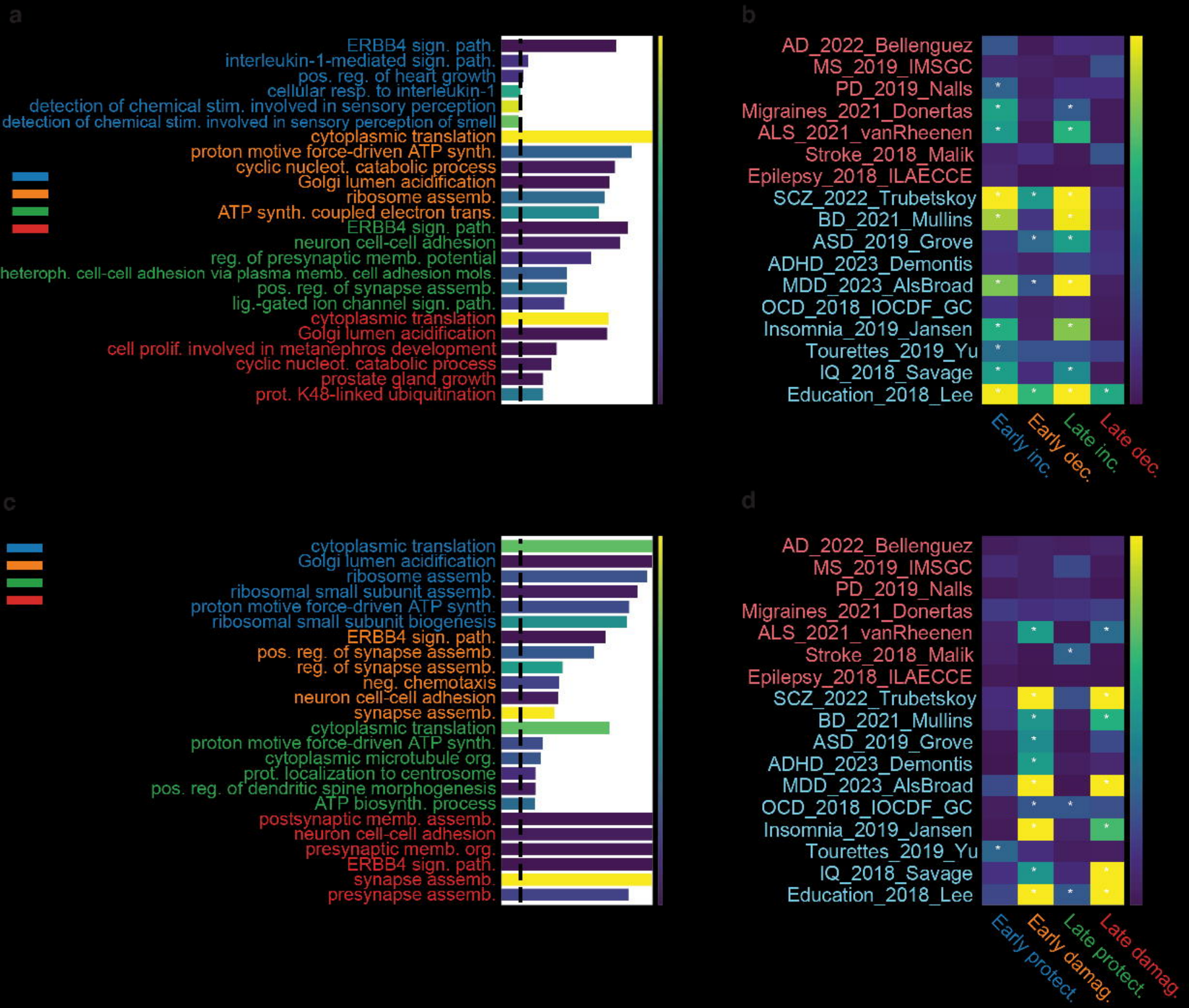

Braak
Dementia Res.
neuron cell-cell adhesion
postsynaptic membrane assembly
presynaptic membrane organization
ERBB4 sign. pathway
neg. chemotaxis
dendrite self-avoidance
synaptic vesicle clustering
reg. of presynaptic membrane potential
synaptic transmission, GLU
cyclic nucleotide catabolic process
cytoplasmic translation
Golgi lumen acidification
ATP synth. coupled electron trans.
ribosome assembly
establishment of protein localization to endoplasmic reticulum
pos. reg. of ATP-dependent activity
proton transmembrane trans.
pos. reg. of autophagy of mitochondrion in resp. to mito. depolarization
reg. of tau-protein kinase activity
chaperone-mediated protein complex assembly
protein refolding
cellular resp. to heat
reg. of cellular resp. to heat
pos. reg. of nitric oxide biosynth. process
glucocorticoid receptor sign. pathway
pos. reg. of telomerase activity
neg. reg. of viral life cycle
pos. reg. of monocyte diff.
antigen proc. and present. of exon. peptide antigen via MHC-II
microglial cell activation
pos. reg. of phagocytosis, engulfment
leukotriene biosynth. process
neg. reg. of B cell prolif.
myelin assembly
pos. reg. of T-help. 1 type immune resp.
MHC-II biosynth. process
T cell extravasation
reg. of mito. fusion
neg. reg. of vascular ass. smooth muscle cell migrat.
neg. reg. of vascular ass. smooth muscle cell prolif.
neg. reg. of protein localization to cell surface
dendritic spine development
GLU secretion
secretory granule localization
reg. of dendritic spine maintenance
reg. of glycolytic process
glycogen metabolic process
neg. reg. of lipid trans.
circadian reg. of gene expression
glial cell apoptotic process
detoxification of copper ion
neg. reg. of vascular permeability
cellular resp. to interleukin-17
reg. of translation in resp. to endoplasmic reticulum stress
neg. reg. of stem cell diff.
reg. of endothelial cell diff.
EN
IN
Astro
Immune
Oligo
OPC
Mural
Endo
ΔEarly
ΔLate
10
0
−10
10
0
−10

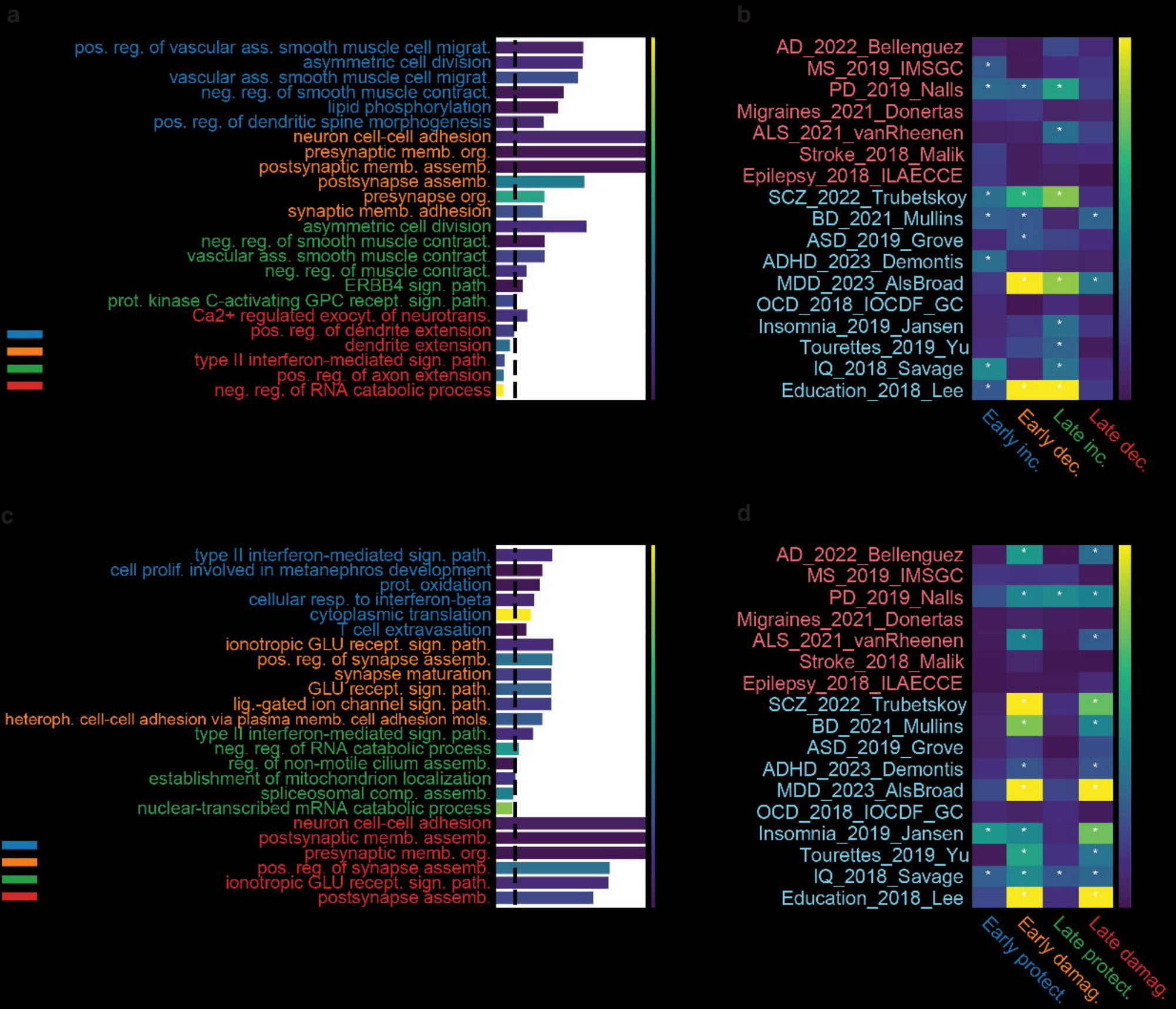

Braak
Dementia Res.
endocytic vesicle lumen
axonal growth cone
platelet alpha granule membrane
uropod
dendritic shaft
Schaffer collateral - CA1 synapse
stress fiber
myelin sheath
axolemma
GABA-A receptor complex
axon initial segment
postsynaptic density membrane
ionotropic GLU receptor complex
cytosolic ribosome
polysome
cytochrome complex
proton-transporting ATP synthase complex
respirasome
mito. respirasome
oxidoreductase complex
autolysosome
ESC/E(Z) complex
presynaptic cytosol
proteasome complex
U1 snRNP
chromosome, telomeric repeat region
ciliary transition zone
MHC class II protein complex
low-density lipoprotein particle
lumenal side of endoplasmic reticulum membrane
exocyst
microfibril
desmosome
EN
Astro
IN
Immune
Oligo
OPC
Mural
Endo
ΔEarly
ΔLate
10
0
−10
10
0
−10


Braak
Dementia Res.
structural constituent of ribosome
ubiquitin-protein transferase inhibitor activity
rRNA binding
NADH dehydrogenase (ubiquinone) activity
proton-transporting ATP synthase activity, rotational mechanism
cGMP binding
nuclear export signal receptor activity
RNA helicase activity
ligase activity
protein folding chaperone
tau protein binding
adenyl-nucleotide exchange factor activity
DNA polymerase binding
3',5'-cyclic-AMP phosphodiesterase activity
GLU receptor activity
ionotropic GLU receptor activity
NAD(P)+ nucleosidase activity
polysaccharide binding
peptidoglycan binding
peptide antigen binding
myristoyltransferase activity
potassium ion binding
metalloendopeptidase inhibitor activity
MHC class II protein complex binding
neurexin family protein binding
nuclear receptor activity
transmembrane receptor protein tyrosine kinase activity
transmembrane receptor protein tyrosine kinase activator activity
1-phosphatidylinositol binding
polypeptide N-acetylgalactosaminyltransferase activity
transcription corepressor activity
sign. adaptor activity
L-arginine transmembrane transporter activity
MAP kinase tyrosine phosphatase activity
cAMP resp. element binding
cadherin binding involved in cell-cell adhesion
phosphatidylinositol transfer activity
nuclear glucocorticoid receptor binding
EN
IN
Astro
Immune
Oligo
OPC
Mural
Endo
10
0
−10
10
0
−10


a
cytoplasmic translation
microglial cell activation
leukocyte activation involved in inflamm. resp.
neg. reg. of B cell prolif.
neg. reg. of B cell activation
reg. of B cell prolif.
neuron cell-cell adhesion
ERBB4 sign. path.
presynaptic memb. org.
GLU recept. sign. path.
cellular resp. to interleukin-17
podocyte development
MHC- II prot. comp. assemb.
peptide antigen assemb. with MHC-II prot. comp.
antigen proc. and pres. of exogenous peptide antigen via MHC-II
antigen processing and pres. of peptide antigen via MHC-II
antigen proc. and pres. of peptide or polysacch. antigen via MHC-II
antigen processing and pres. of peptide antigen
chaperone-med. prot. folding
chaperone-med. prot. comp. assemb.
prot. folding
reg. of cellular resp. to heat
prot. refolding
neg. reg. of inclusion body assemb.
b
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early protect.
Early damag.
Late protect.
Late damag.


a
NK T cell diff.
nerve growth factor sign. path.
ERBB4 sign. path.
cerebellar Purkinje cell layer formation
cerebellar Purkinje cell diff.
cerebellar granule cell precursor prolif.
cytoplasmic translation
cyclic nucleot. catabolic process
cell prolif. involved in metanephros development
cell prolif. involved in kidney development
Golgi lumen acidification
ribosomal large subunit biogenesis
postsynaptic memb. assemb.
presynaptic memb. org.
presynapse assemb.
neuron cell-cell adhesion
ERBB4 sign. path.
presynapse org.
cyclic nucleot. catabolic process
cAMP metabolic process
cell prolif. involved in metanephros development
purine ribonucleotide catabolic process
reg. of cGMP-mediated sign.
Golgi lumen acidification
b
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early inc.
Early dec.
Late inc.
Late dec.
c
cytoplasmic translation
ATP synth. coupled electron trans.
mito. ATP synth. coupled electron trans.
aerobic electron trans. chain
proton motive force-driven ATP synth.
oxidative phosphorylation
mod. of excitatory postsynaptic potential
reg. of postsynaptic memb. potential
pos. reg. of synapse assemb.
heteroph. cell-cell adhesion via plasma memb. cell adhesion mols.
reg. of synapse assemb.
postsynaptic density org.
cAMP metabolic process
cellular resp. to interleukin-12
resp. to interleukin-12
reg. of cGMP-mediated sign.
interleukin-15-mediated sign. path.
interleukin-1-mediated sign. path.
reg. of dendritic spine maintenance
GLU recept. sign. path.
reg. of postsynaptic cytosolic Ca2+ concentration
reg. of presynaptic memb. potential
mod. of excitatory postsynaptic potential
dendritic spine org.
d
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early protect.
Early damag.
Late protect.
Late damag.


a
ERBB4 sign. path.
interleukin-1-mediated sign. path.
pos. reg. of heart growth
cellular resp. to interleukin-1
detection of chemical stim. involved in sensory perception
detection of chemical stim. involved in sensory perception of smell
cytoplasmic translation
proton motive force-driven ATP synth.
cyclic nucleot. catabolic process
Golgi lumen acidification
ribosome assemb.
ATP synth. coupled electron trans.
ERBB4 sign. path.
neuron cell-cell adhesion
reg. of presynaptic memb. potential
heteroph. cell-cell adhesion via plasma memb. cell adhesion mols.
pos. reg. of synapse assemb.
lig.-gated ion channel sign. path.
cytoplasmic translation
Golgi lumen acidification
cell prolif. involved in metanephros development
cyclic nucleot. catabolic process
prostate gland growth
prot. K48-linked ubiquitination
b
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early inc.
Early dec.
Late inc.
Late dec.
c
cytoplasmic translation
Golgi lumen acidification
ribosome assemb.
ribosomal small subunit assemb.
proton motive force-driven ATP synth.
ribosomal small subunit biogenesis
ERBB4 sign. path.
pos. reg. of synapse assemb.
reg. of synapse assemb.
neg. chemotaxis
neuron cell-cell adhesion
synapse assemb.
cytoplasmic translation
proton motive force-driven ATP synth.
cytoplasmic microtubule org.
prot. localization to centrosome
pos. reg. of dendritic spine morphogenesis
ATP biosynth. process
postsynaptic memb. assemb.
neuron cell-cell adhesion
presynaptic memb. org.
ERBB4 sign. path.
synapse assemb.
presynapse assemb.
d
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early protect.
Early damag.
Late protect.
Late damag.


a
pos. reg. of vascular ass. smooth muscle cell migrat.
asymmetric cell division
vascular ass. smooth muscle cell migrat.
neg. reg. of smooth muscle contract.
lipid phosphorylation
pos. reg. of dendritic spine morphogenesis
neuron cell-cell adhesion
presynaptic memb. org.
postsynaptic memb. assemb.
postsynapse assemb.
presynapse org.
synaptic memb. adhesion
asymmetric cell division
neg. reg. of smooth muscle contract.
vascular ass. smooth muscle contract.
neg. reg. of muscle contract.
ERBB4 sign. path.
prot. kinase C-activating GPC recept. sign. path.
Ca2+ regulated exocyt. of neurotrans.
pos. reg. of dendrite extension
dendrite extension
type II interferon-mediated sign. path.
pos. reg. of axon extension
neg. reg. of RNA catabolic process
b
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early inc.
Early dec.
Late inc.
Late dec.
c
type II interferon-mediated sign. path.
cell prolif. involved in metanephros development
prot. oxidation
cellular resp. to interferon-beta
cytoplasmic translation
T cell extravasation
ionotropic GLU recept. sign. path.
pos. reg. of synapse assemb.
synapse maturation
GLU recept. sign. path.
lig.-gated ion channel sign. path.
heteroph. cell-cell adhesion via plasma memb. cell adhesion mols.
type II interferon-mediated sign. path.
neg. reg. of RNA catabolic process
reg. of non-motile cilium assemb.
establishment of mitochondrion localization
spliceosomal comp. assemb.
nuclear-transcribed mRNA catabolic process
neuron cell-cell adhesion
postsynaptic memb. assemb.
presynaptic memb. org.
pos. reg. of synapse assemb.
ionotropic GLU recept. sign. path.
postsynapse assemb.
d
AD_2022_Bellenguez
MS_2019_IMSGC
PD_2019_Nalls
Migraines_2021_Donertas
ALS_2021_vanRheenen
Stroke_2018_Malik
Epilepsy_2018_ILAECCE
SCZ_2022_Trubetskoy
BD_2021_Mullins
ASD_2019_Grove
ADHD_2023_Demontis
MDD_2023_AlsBroad
OCD_2018_IOCDF_GC
Insomnia_2019_Jansen
Tourettes_2019_Yu
IQ_2018_Savage
Education_2018_Lee
Early protect.
Early damag.
Late protect.
Late damag.