

# Coordinated cellular responses track cognitive outcomes in Alzheimer's disease

**Panos Roussos**

[panagiotis.roussos@mssm.edu](mailto:panagiotis.roussos@mssm.edu)

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0002-4640-6239>

**Pengfei Dong**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0003-3735-5158>

**Xuan Cao**

Icahn School of Medicine at Mount Sinai

**Seon Kinrot**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0001-5614-0983>

**Clara Casey**

Icahn School of Medicine at Mount Sinai

**Aram Hong**

Department of Psychiatry, Icahn School of Medicine at Mount Sinai

**Marcela Alvia**

Department of Psychiatry, Icahn School of Medicine at Mount Sinai

**Zhiping Shao**

Mount Sinai School of Medicine

**Joana Krause-Massaguer**

Icahn School of Medicine at Mount Sinai

**Vahram Haroutunian**

Mount Sinai and JJ Peters VA Medical Center <https://orcid.org/0000-0001-5860-2512>

**Gabriel Hoffman**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0002-0957-0224>

**Jaroslav Bendl**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0001-9989-2720>

**Donghoon Lee**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0003-0453-6059>

**John Fullard**

Icahn School of Medicine at Mount Sinai <https://orcid.org/0000-0001-9874-2907>

**Keywords:**

**Posted Date:** October 20th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-11296978/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

# Coordinated cellular responses track cognitive outcomes in Alzheimer's disease

## Authors:

Pengfei Dong<sup>1,2,3,4</sup>†, Xuan Cao<sup>1,2,3,4</sup>, Seon Kinrot<sup>1,2,3,4</sup>, Clara Casey<sup>1,2,3,4</sup>, Aram Hong<sup>1,2,3,4</sup>, Marcela Alvia<sup>1,2,3,4</sup>, Zhiping Shao<sup>1,2,3,4</sup>, Joana Krause-Massaguer<sup>1,2,3,4</sup>, Vahram Haroutunian<sup>2,3,6,7</sup>, Gabriel E. Hoffman<sup>1,3,4,5,6</sup>, Jaroslav Bendl<sup>1,2,3,4</sup>, Donghoon Lee<sup>1,2,3,4</sup>, John F Fullard<sup>1,2,3,4</sup>, Panos Roussos<sup>1,2,3,4,5,6</sup>†

## Affiliations:

<sup>1</sup>Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>2</sup>Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>3</sup>Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>4</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

<sup>5</sup>Center for Precision Medicine and Translational Therapeutics, James J. Peters VA Medical Center, Bronx, NY, USA.

<sup>6</sup>Mental Illness Research Education and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA.

<sup>7</sup>Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, USA

† Corresponding authors: pengfei.dong@mssm.edu; panagiotis.roussos@mssm.edu

## Abstract

Individuals with comparable Alzheimer's disease (AD) pathology can differ markedly in cognition, but the molecular nature of this heterogeneity remains unclear. Here we profiled paired gene expression and chromatin accessibility in 1.42 million nuclei from three cortical regions of 247 donors. A pathology-supervised transcriptomic score, supported by paired chromatin measurements, resolved cellular disease state differences within samples and between individuals. After adjustment for measured pathology, cellular responses were concordant across major cell classes and cortical regions. Higher pathology-adjusted responses were associated with lower relative abundance of vulnerable excitatory and somatostatin-expressing inhibitory neurons and poorer cognitive outcomes. Pathology-associated, pathology-adjusted, and within-sample cellular-state differences converged on overlapping molecular programs. Unsupervised analysis identified reproducible donor-associated transcriptional programs, including a proteostasis and cellular-stress program associated with higher pathology-adjusted responses, contrasting a lipid-metabolic and biosynthetic program showing the opposite relationship across regions and cohorts. Together, these findings identify a coordinated molecular phenotype associated with differences in neuronal composition and cognition among individuals with comparable measured neuropathology.

## Main

Alzheimer's disease (AD) is characterized by amyloid- $\beta$  (A $\beta$ ) plaques and phosphorylated tau (pTau) tangles, yet neuropathological burden does not fully account for its clinical expression<sup>1–3</sup>. Individuals with comparable measured pathology can differ substantially in cognitive impairment, neurodegeneration, and disease course<sup>2–6</sup>. Understanding the molecular differences accompanying this discordance could help clarify why similar pathological burdens are associated with different cellular and cognitive outcomes.

Large-scale single-nucleus studies have identified selectively vulnerable neuronal populations and cell-type-specific molecular programs associated with AD pathology and cognition<sup>7–15</sup>. At the same time, AD lesions are spatially heterogeneous<sup>1,16</sup>, and disease-associated and relatively preserved cellular states can coexist within the same tissue<sup>17–19</sup>. Donor-level neuropathological measurements therefore provide an important disease reference but do not directly characterize the molecular state of each cell. It remains unclear whether cellular responses among individuals with comparable measured pathology are coordinated across cell populations and brain regions, and whether this coordination relates to neuronal composition and cognition. Addressing this question requires distinguishing variation associated with pathological burden from within-sample cellular heterogeneity and differences in molecular response between individuals.

Here, as part of the PsychAD project<sup>8,20</sup>, we generated paired gene-expression and chromatin-accessibility profiles from 1.42 million nuclei across three cortical regions of 247 donors spanning AD neuropathology. We developed a pathology-supervised transcriptomic score, Dx\_cell, to characterize cellular responses within samples and between individuals, with support from paired chromatin measurements and external transcriptomic datasets. After adjustment for measured pathology, these responses were coordinated across major cell classes and cortical regions, and were associated with vulnerable-neuron composition and cognitive outcomes across cohorts. Unsupervised analysis identified proteostasis and lipid-biosynthetic programs with opposing associations with pathology-adjusted cellular response. These findings identify a coordinated molecular phenotype associated with differences in neuronal composition and cognition among individuals with comparable measured neuropathology.

## Results

### A population-scale paired single-nucleus multiome atlas defines cellular and regulatory identity across the AD cortex

We generated a population-scale paired single-nucleus (sn) multiome (snMultiome) atlas from 576 postmortem specimens from 247 unique donors spanning the full spectrum of AD pathology from the Mount Sinai NIH NeuroBioBank, including 68 late-stage AD donors, 111 early-stage AD donors, and 68 age-matched controls (Methods). We profiled three cortical regions, prefrontal cortex (PFC), superior temporal gyrus (STG), and parahippocampal gyrus (PHG) (**Fig. 1a**), which differ in their timing and extent of involvement by A $\beta$ <sup>21,22</sup> and neurofibrillary tangles<sup>23,24</sup>. Because AD-associated pathological measurements were strongly correlated across donors (**Fig. 1b** and **Extended Data Fig. 1**), we integrated these measurements using principal component analysis to derive a continuous donor-level pathological-burden score, Dx\_donor, with higher values reflecting greater pathological burden.

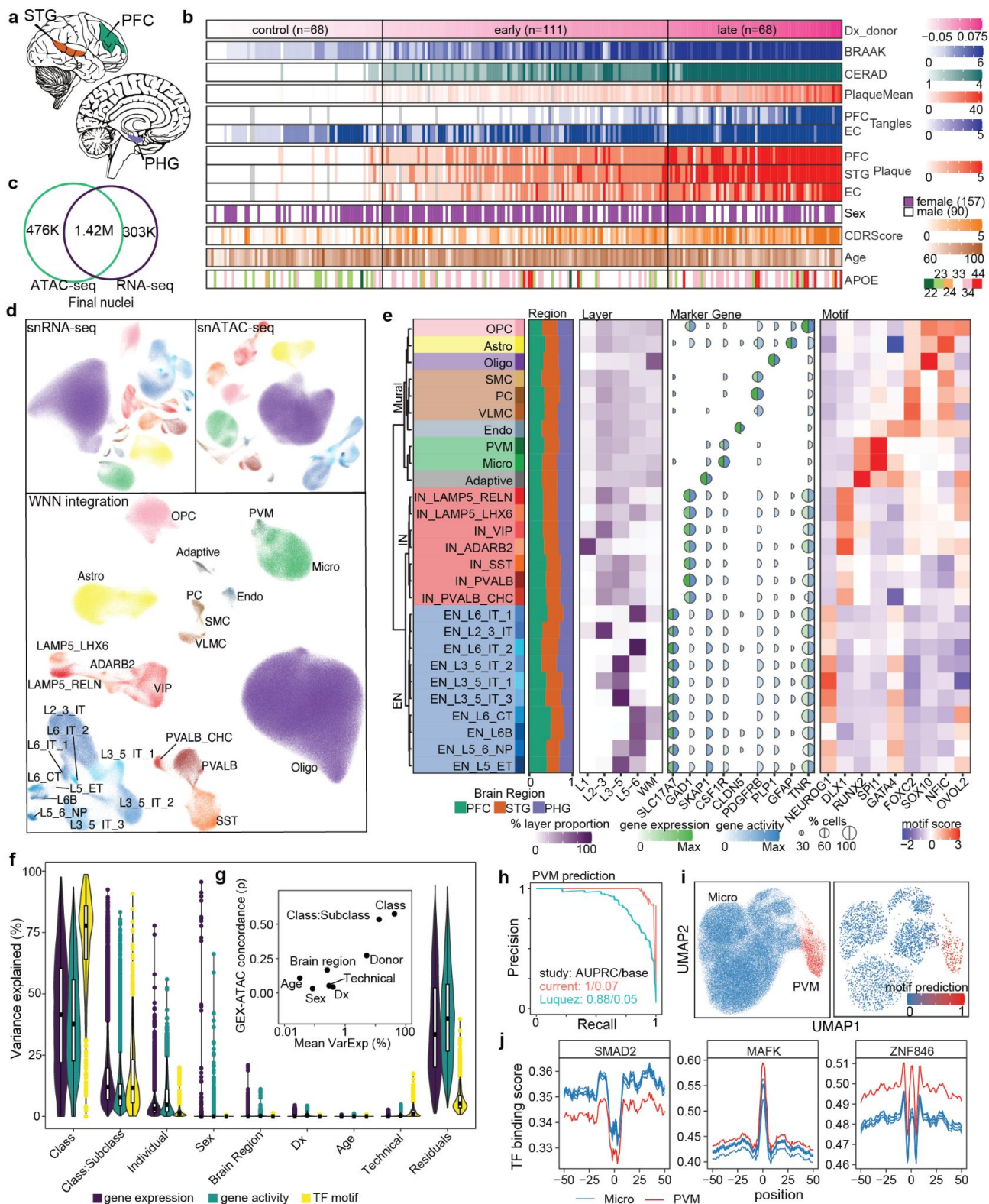


**Fig. 1. A population-scale paired snMultiome atlas across three brain regions and AD pathology.** **a**, Schematic of the three profiled brain regions. Adapted from cerebroViz<sup>25</sup>. **b**, Heatmap showing donor-level clinical and neuropathological measurements ordered by the continuous pathology score, Dx\_donor. **c**, Number of final nuclei retained after quality control. **d**, UMAP embeddings derived from RNA, ATAC, and WNN integration, colored by the 27-subclass taxonomy. **e**, Brain-region distribution, cortical-depth/layer annotation, class-level marker gene expression and gene activity scores, and class-level chromVAR motif activity for 27 subclasses. In the marker gene plot, the left half of each dot represents gene expression, and the right half represents the ATAC derived gene activity score; color indicates the mean normalized

expression or activity, and dot size indicates the percentage of nuclei with detectable expression or nonzero gene activity. **f**, Variance partitioning across subclass-by-sample pseudobulk profiles for gene expression, ATAC-derived gene activity, and TF motif activity. **g**, Scatter plot showing the mean variance explained (average across both modalities) and concordance between matched gene expression and gene activity ( $\rho$ ). Components explaining larger fractions of variance also showed high concordance across the two modalities. **h**, Precision–recall curves for three-motif prediction of PVM identity within the Micro\_PVM in the discovery cohort (current study) and validation cohort (Luquez et al 2026<sup>13</sup>). Area under the precision–recall curve (AUPRC) and prevalence-based baseline AUPRC are shown. **i**, Left, Micro\_PVM sn-UMAP colored by Micro and PVM annotation. Right, metacell UMAP colored by the three-motif PVM prediction score. **j**, Multiscale TF-binding profiles<sup>26</sup> for selected Micro- and PVM-associated TF motifs. Micro subtype-specific profiles are overlaid in the same color.

Using 10x Genomics snMultiome profiling, we obtained paired gene expression and chromatin accessibility profiles from 1.42 million high-quality nuclei after stringent quality control, with a median of 2,085 detected genes and 7,645 chromatin fragments per nucleus. An additional 778,243 nuclei passed quality control in a single modality and were retained through modality-specific label projection (**Fig. 1c** and **Methods**). To construct a unified cellular taxonomy, we performed iterative subclustering using weighted nearest-neighbor<sup>27</sup> (WNN) integration of RNA and ATAC modalities, followed by nearest-neighbor projection of single-modality nuclei (**Fig. 1d** and **Extended Data Fig. 2**). This procedure identified a hierarchical taxonomy comprising 9 major cell classes, 27 subclasses, and 90 subtypes. The nine major cell classes included excitatory neurons (EN), inhibitory neurons (IN), astrocytes (Astro), oligodendrocytes (Oligo), oligodendrocyte precursor cells (OPC), adaptive immune cells (Adaptive), microglia/perivascular macrophages (Micro\_PVM), mural cells (Mural), and endothelial cells (Endo). Subclass annotations were supported by concordant marker gene expression, ATAC-derived gene activity<sup>28</sup>, chromVAR TF motif activity<sup>29</sup>, cortical layer annotation<sup>8,30</sup>, and alignment to external reference datasets<sup>8,9,11,14,31</sup> (**Fig. 1e** and **Extended Data Figs. 2–4**).

We next quantified the major sources of transcriptomic and chromatin-derived variation across subclass-by-sample pseudobulk profiles (**Methods**). Across gene expression, ATAC-derived gene activity, and chromVAR TF motif activity<sup>29</sup>, major cell classes explained the largest fraction of variance, followed by subclass identity within the major cell class (**Fig. 1f**). Donor identity and brain region explained smaller fractions, whereas donor-level pathology, age, sex, and technical covariates explained lower median fractions. Variance components for matched gene expression and gene-activity measurements were concordant across biological sources, particularly cell class and subclass identity (**Fig. 1g**; **Extended Data Fig. 5**).

Sparse motif models based on major-class-specific candidate cis-regulatory elements (cCREs) and depth-matched metacells distinguished subclasses within major cell classes (**Methods** and **Extended Data Fig. 6**). A three-motif signature distinguished PVMs from closely related Micro and transferred to an independent multiome cohort<sup>13</sup> (**Fig. 1h, i**). Multiscale TF-binding scores<sup>26</sup> provided additional support for selected Micro- and PVM-associated motifs (**Fig. 1j**). Together, these analyses establish a paired transcriptomic and regulatory atlas that resolves broad cellular identity, subclass structure, and compact regulatory signatures across the AD cortex.

## Cellular disease-state score resolves within-sample molecular heterogeneity

Dx\_donor summarizes the overall pathological burden of each donor, but a donor-level pathology score does not represent molecular disease-state variation among cells from the same donor or brain region. This limitation is particularly relevant in AD, where pathology is spatially heterogeneous and cellular responses vary across cells and local tissue contexts<sup>16–19</sup> (**Fig. 2a**). The paired snRNA-seq and snATAC-seq measurements in our population-scale snMultiome atlas provide an opportunity to ask whether cellular disease states vary within the same tissue sample. We therefore developed Dx\_cell, an inferred pathology-associated cellular response trained on snRNA-seq gene expression, and evaluated its correspondence with paired ATAC-derived gene activity.

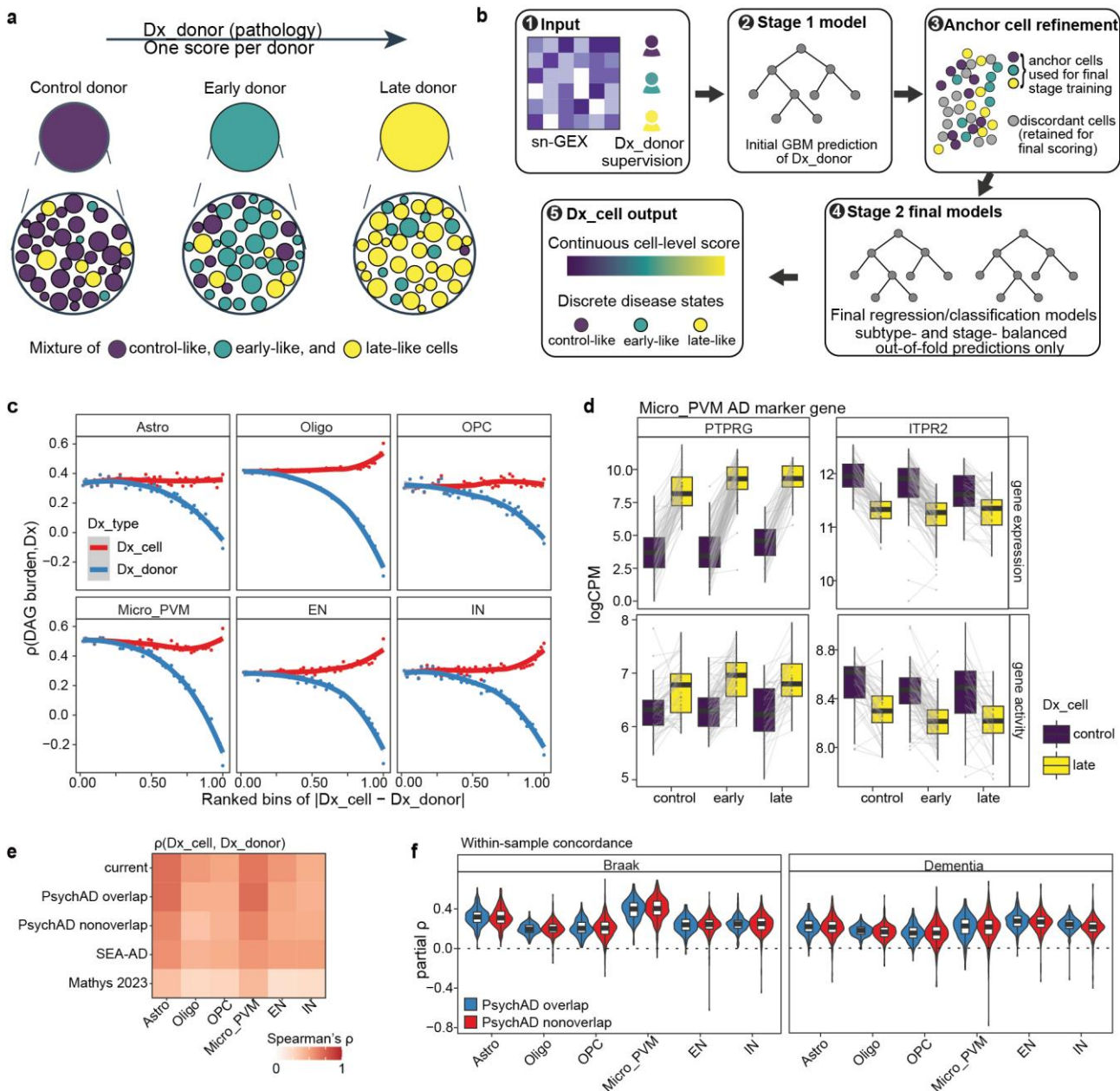


**Fig. 2. Cellular disease-state scoring resolves within-sample molecular heterogeneity.** **a**, Conceptual schematic showing donor-level pathology versus cellular disease-state heterogeneity. Donor-level pathology assigns one  $Dx\_donor$  value to each donor, whereas  $Dx\_cell$  represents a pathology-supervised cellular disease-state score that resolves mixtures of model-defined control-like, early-like, and late-like disease states within the same tissue sample. **b**, Two-step  $Dx\_cell$  training and prediction workflow. An initial regression model was trained using snRNA-seq gene expression to predict  $Dx\_donor$ . Donor-concordant anchor cells, identified using out-of-fold initial predictions, were then used to train final regression and classification models. Final  $Dx\_cell$  scores and discrete  $Dx\_cell$ -defined states were assigned to all cells using 20-fold cell-level cross-fitting, with predictions retained only for cells not included in the corresponding model-fitting data. **c**, Cross-modality evaluation of  $Dx\_cell$  using ATAC-seq-derived disease burden. Cells were divided into 50 ordered non-overlapping bins ranked by increasing  $|Dx\_cell - Dx\_donor|$ . Within each bin, Spearman's correlation coefficient,  $\rho$ , was calculated between ATAC-derived disease burden calculated from differential gene-activity genes (DAGs) and either  $Dx\_cell$  or  $Dx\_donor$ . **d**, Same-sample validation of  $Dx\_cell$ -defined disease states in Micro\_PVM cells. Gene expression and ATAC-derived gene activity levels of representative AD-associated marker genes, *PTPRG* and *ITPR2*, are shown for  $Dx\_cell$ -defined control-like and late-like cells within each sample. Lines connect paired values from the same sample. Paired two-sided Wilcoxon signed-rank tests were used to compare  $Dx\_cell$ -defined control-like and late-like pseudobulk values within each donor-level pathology stage, followed by FDR correction (all FDR < 0.05). **e**, Spearman correlations between sample-level mean  $Dx\_cell$  and cohort-specific measures of AD pathology across major cell classes. Correlations

are shown for the discovery multiome cohort; PsychAD donors overlapping and non-overlapping with the multiome cohort; the Mathys et al. 2023 ROSMAP cohort<sup>14</sup>; and SEA-AD<sup>9</sup>.  $Dx\_donor$  was used as the pathology measure for the multiome and PsychAD cohorts, global AD pathology was used for Mathys et al. 2023, and the CPS was used for SEA-AD. **f**, Within-sample concordance between  $Dx\_cell$  and independently trained PsychAD transcriptomic prediction scores<sup>8</sup>. Violin and box plots show the distribution of partial Spearman correlation coefficients between  $Dx\_cell$  and cellular Braak- or dementia-prediction scores across samples within each major cell class, conditioning on technical covariates. Results are shown separately for PsychAD donors overlapping and non-overlapping with the multiome cohort. Boxes indicate the median and interquartile range; whiskers extend to  $1.5 \times$  the interquartile range.

We trained major-cell-class-specific  $Dx\_cell$  models using donor-level pathology as supervision and single-nucleus gene expression as input (**Fig. 2b** and **Methods**). To reduce confounding from AD-associated differences in cellular composition, regional sampling, or region-specific cell loss, training was balanced across subtype, brain region, and donor pathology stage, such that cells from control, early-pathology and late-pathology donor groups contributed comparably within each subtype–region combination. We used a two-stage modeling strategy. An initial regression model identified donor-concordant anchor cells, which were used to train final regression and classification models yielding continuous scores and control-like, early-like, and late-like  $Dx\_cell$  defined states. Predictions were generated using 20-fold cell-level cross-fitting, retaining only out-of-fold predictions. Continuous and classification-derived estimates were concordant within samples (**Extended Data Fig. 7a-c**).

As an internal consistency test, we next asked whether  $Dx\_cell$  continued to reflect cellular transcriptional disease state when individual cells differed from the pathology level assigned to their donor. Because  $Dx\_donor$  is fixed for all cells from a given donor whereas  $Dx\_cell$  varies across cells, we quantified the extent to which each cell departed from its donor-level pathology. After rescaling  $Dx\_cell$  to the same mean and standard deviation as  $Dx\_donor$ , we defined cell–donor discordance as  $|Dx\_cell - Dx\_donor|$  (**Methods**). Cells were ranked by increasing discordance and divided into 50 ordered, non-overlapping bins. Within each bin, we calculated the Spearman correlation between cell-class-specific RNA-derived disease-associated transcriptional burden and either  $Dx\_cell$  or  $Dx\_donor$ . Among cells with low discordance, both measures tracked transcriptional disease burden. However, as cells increasingly departed from their donor-level pathology, the association with  $Dx\_donor$  progressively weakened, whereas the association with  $Dx\_cell$  remained consistently strong (**Extended Data Fig. 7d**). Thus,  $Dx\_cell$  captured cell-to-cell variation in transcriptional disease state that was not represented by the donor-level pathology measure.

We next asked whether these RNA-derived cellular disease states were also reflected in chromatin-associated molecular signals. Using the same 50-bin framework, we quantified the relationship between disease-associated burden derived from ATAC gene activity and either  $Dx\_cell$  or  $Dx\_donor$ . Similar to the transcriptomic analysis, across bins of increasing discordance, ATAC-derived disease burden remained associated with  $Dx\_cell$  while its correlation with  $Dx\_donor$  decreased (**Fig. 2c**). This cross-modality analysis supports the interpretation that  $Dx\_cell$  captures coordinated transcriptional and chromatin-associated disease-state variation, rather than reflecting RNA variation alone.

Finally, we examined whether  $Dx\_cell$  stratified cells into coherent disease states within the same sample. In Micro\_PVM cells, representative AD-associated genes, including *PTPRG* and *ITPR2*, showed clearer separation between  $Dx\_cell$ -defined control-like and late-like states than between donor pathology groups, both at the level of gene expression and ATAC-derived gene activity (**Fig. 2d**). Broader same-sample comparisons across additional major cell classes and marker genes showed consistent separation between  $Dx\_cell$ -defined control-like and late-like states (**Extended Data Fig. 7e**). As these comparisons were performed within the same samples, they support the interpretation that  $Dx\_cell$  captures within-sample molecular heterogeneity that is masked by assigning a uniform donor-level disease stage to every nucleus.

We next projected the frozen models into additional transcriptomic datasets. Sample-level mean  $Dx\_cell$  remained associated with cohort-specific measures of AD pathology in PsychAD<sup>8</sup>, including donors not represented in the discovery cohort, in the Mathys et al. 2023 ROSMAP cohort<sup>14</sup>, and in SEA-AD<sup>9</sup> (**Fig. 2e**).

These analyses used Dx\_donor in PsychAD, global AD pathology in Mathys et al. 2023, and the Continuous Pseudo-progression Score (CPS) in SEA-AD. Within individual PsychAD samples, Dx\_cell was positively associated with independently trained cellular Braak- and dementia-prediction scores after conditioning on technical covariates, with similar distributions in overlapping and non-overlapping donor sets (**Fig. 2f**). Together, the paired chromatin analyses and cross-dataset projections support Dx\_cell as a transferable cellular disease-state score that remains related to donor pathology while resolving heterogeneity within samples.

## Pathology-adjusted cellular response deviation is associated with AD-vulnerable neuronal composition

Having established that Dx\_cell resolves molecular heterogeneity among cells within the same tissue sample, we next asked a distinct question: do samples with comparable measured pathology differ systematically in their average cellular response? Although sample-level mean Dx\_cell generally increased with donor pathology, substantial variation remained among samples with similar Dx\_donor. We first aggregated Dx\_cell within each sample and major cell class and partitioned its variance across donor pathology, donor identity, and other biological and technical covariates (**Methods**). Donor pathology and donor identity together explained approximately 80% of the modeled variation in sample-level mean Dx\_cell across major cell classes (**Fig. 3a**). Mean Dx\_cell differed substantially between donors even after accounting for measured pathology. We defined Dx\_dev (**Fig. 3b**) as the deviation in mean Dx\_cell for each sample and major cell class after adjustment for Dx\_donor and demographic and technical covariates (**Methods; Supplementary Table 5**). Positive Dx\_dev indicates a stronger cellular response than expected for the measured pathological burden, whereas negative values indicate a weaker response.

Despite being derived from Dx\_cell models trained independently within each major cell class, Dx\_dev was concordant across cell classes (**Fig. 3c**). Dx\_dev was also reproducible across anatomical regions; consensus Dx\_dev values from PFC correlated with those from STG (Spearman's  $\rho = 0.86$ ) and PHG ( $\rho = 0.91$ ), and with PsychAD estimates from overlapping donors ( $\rho = 0.72$ ; **Fig. 3d**). Analogous pathology-adjusted deviations from independently trained Braak pathology- and dementia-prediction models showed concordance in both overlapping and non-overlapping PsychAD donors (**Extended Data Fig. 8a**). These analyses support Dx\_dev as a reproducible donor-associated, pathology-adjusted cellular disease-state phenotype that is shared across cell classes, brain regions, and transcriptomic platforms.

We next asked whether Dx\_dev was associated with the cellular-composition remodeling observed with increasing AD pathology. As a reference, we first quantified composition changes associated with Dx\_donor. Consistent with previous single-cell studies identifying selective vulnerability of excitatory and inhibitory neuronal populations in AD<sup>7-9,11,32</sup>, increasing pathology was associated with lower relative abundance of EN\_L2\_3\_IT and IN\_SST, with particularly strong effects among superficial-layer SST populations (**Extended Data Fig. 8b,c; Supplementary Table 6**). Changes in neuronal relative abundance were also regionally structured, with more pronounced changes in STG and PHG than in PFC. To test whether pathology-adjusted cellular response recapitulated this broader AD-associated remodeling, we derived Dx\_dev using glial cell classes and compared its composition effects with those associated with Dx\_donor. Across cell populations, Dx\_dev-associated composition effects were strongly concordant with the Dx\_donor-associated pattern in the discovery cohort (**Fig. 3e; Supplementary Table 6**). Thus, although Dx\_dev is defined after accounting for measured pathology, samples with higher Dx\_dev exhibited a cellular-composition profile resembling that observed with greater pathological burden.

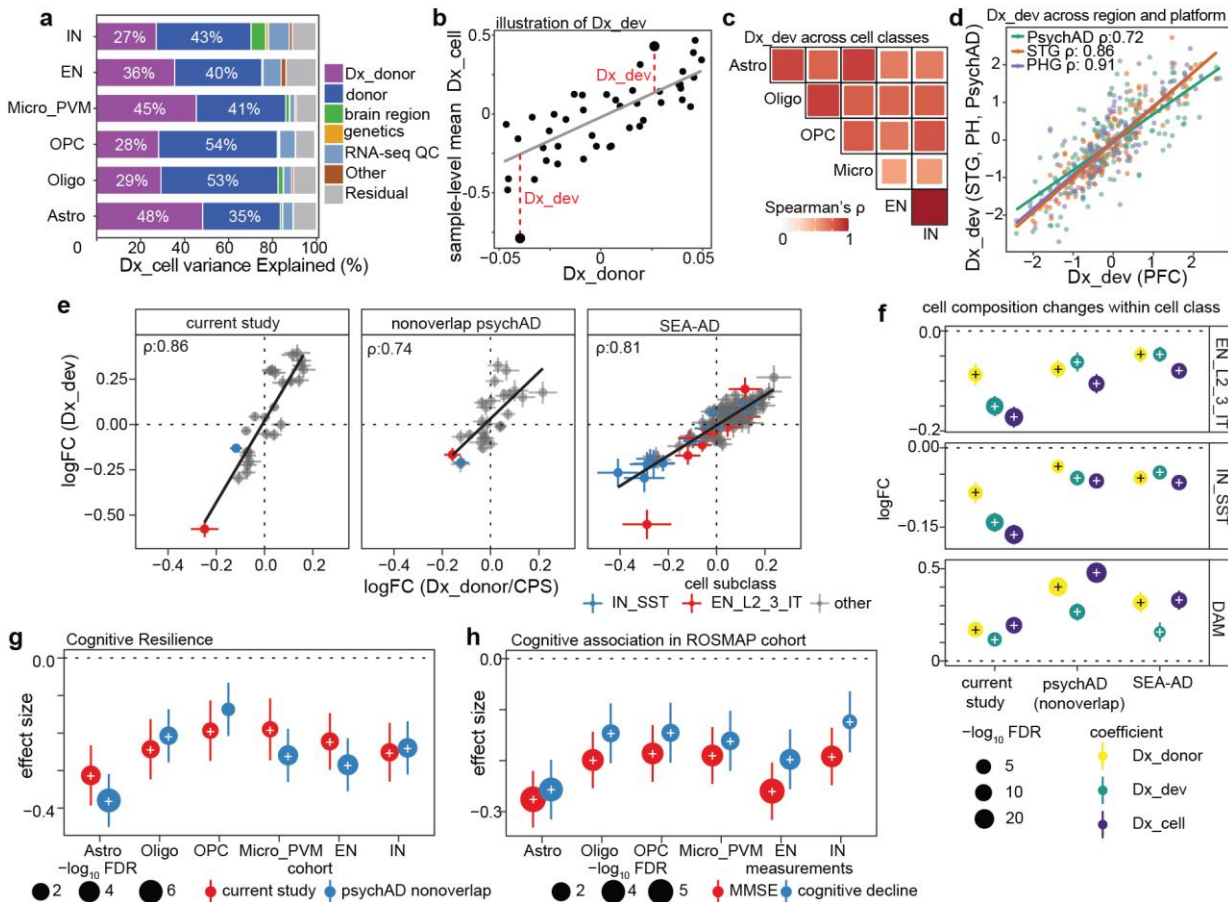


**Fig. 3. Pathology-adjusted cellular response deviation is associated with AD-related cellular composition changes and cognitive outcomes.** **a**, Variance partitioning of sample-level mean  $Dx_{cell}$  across major cell classes. Bars show the proportion of variance attributed to  $Dx_{donor}$ , donor identity, and other modeled biological and technical covariates. Genetics include the combined effects of APOE alleles and polygenic risk scores<sup>8,35,36</sup>. **b**, Conceptual illustration of  $Dx_{dev}$ . The relationship between sample-level mean  $Dx_{cell}$  and donor-level neuropathological burden ( $Dx_{donor}$ ) is shown schematically. For a given sample,  $Dx_{dev}$  is the vertical deviation between the observed sample-level mean  $Dx_{cell}$  and its expected value from the fitted model after accounting for  $Dx_{donor}$  and relevant covariates. Positive and negative  $Dx_{dev}$  indicate cellular disease states that are higher or lower, than expected given measured pathology. **c**, Pairwise Spearman correlations of sample-level  $Dx_{dev}$  across major cell classes. **d**, Cross-region and cross-dataset reproducibility of  $Dx_{dev}$ . Scatterplots compare consensus  $Dx_{dev}$ , calculated as the mean across major cell classes, between PFC and STG/PHG, and between PFC and PsychAD datasets for overlapping donors. Spearman correlation coefficients ( $\rho$ ) are shown. **e**, Concordance between AD-associated and  $Dx_{dev}$ -associated cell-composition changes across cohorts. Scatterplots compare cell-composition effect sizes associated with donor-level disease burden ( $Dx_{donor}$  for the multiome and PsychAD cohorts; CPS, for SEA-AD) with effects associated with glia-derived  $Dx_{dev}$ . Points represent cell subclasses; horizontal and vertical error bars indicate the standard errors (se) of the corresponding effect estimates. EN\_L2\_3\_IT and IN\_SST, two selectively vulnerable neuronal populations, are highlighted. For the multiome and PsychAD cohorts, all major cellular populations were used; because SEA-AD was FACS-enriched for neurons, the analysis was restricted to neuronal populations. **f**, Forest plots of prespecified AD-associated cellular composition changes across cohorts. Effect estimates, log fold-change ( $\log FC \pm se$ ), are shown for EN\_L2\_3\_IT as a fraction of EN, IN\_SST as a fraction of IN, and DAM as a fraction of Micro\_PVM. Associations with cellular disease state ( $Dx_{cell}$ ), donor-level disease burden ( $Dx_{donor}$  or CPS), and  $Dx_{dev}$  are shown for the multiome, PsychAD (non-overlapping donors), and SEA-AD cohorts. For neuronal outcomes,  $Dx_{dev}$  was derived from glial cell classes; for DAM,  $Dx_{dev}$  was derived excluding Micro\_PVM. **g**, Association between  $Dx_{dev}$  and cognitive resilience. Forest plots show the standardized regression coefficients ( $\beta \pm se$ ) for the association between  $Dx_{dev}$  and cognitive-resilience score in the multiome and non-overlapping PsychAD cohorts. **h**, Independent association of  $Dx_{dev}$  with cognitive outcomes in the ROSMAP cohort<sup>14</sup>. Forest plots show standardized regression coefficients ( $\beta \pm se$ ) for the associations between  $Dx_{dev}$  and last-visit MMSE or longitudinal cognitive slope, with

adjustment for donor pathology and relevant demographic and technical covariates. More negative slopes indicate faster decline in global cognition. (+ FDR < 0.05).

This relationship generalized across independent datasets. After projection of the frozen Dx\_cell models to non-overlapping PsychAD donors, glia-derived Dx\_dev-associated cell-composition changes again paralleled those associated with donor pathology (**Fig. 3e**). We next tested SEA-AD, which provided a complementary comparison because its CPS was derived from quantitative neuropathology measured in the same tissue sample used for molecular profiling<sup>9</sup>. Despite this more detailed, region-matched characterization of pathological burden, neuronal composition changes associated with glia-derived Dx\_dev remained broadly concordant with those associated with CPS (**Fig. 3e**).

We next focused on prespecified AD-associated populations while controlling for changes in their broader parent classes. EN\_L2\_3\_IT abundance was evaluated within EN, IN\_SST within IN, and disease-associated microglia (DAM) within Micro\_PVM<sup>9,13,33,34</sup>. To reduce same-lineage coupling between predictor and outcome, neuronal associations were tested using glia-derived Dx\_dev, whereas DAM associations used Dx\_dev derived from glial populations excluding Micro\_PVM. Higher Dx\_dev was associated with lower relative abundance of EN\_L2\_3\_IT and IN\_SST and higher relative abundance of DAM, with directions broadly paralleling those associated with Dx\_donor or CPS across the multiome, PsychAD and SEA-AD cohorts (**Fig. 3f**). These results indicate that pathology-adjusted cellular response is associated with both broad pathology-aligned composition changes and altered relative abundance of vulnerable and disease-associated populations, including when pathological burden is quantified directly in the profiled sample tissue.

## Pathology-adjusted cellular response deviation is associated with cognitive outcomes

Finally, we asked whether Dx\_dev was related to cognitive outcome beyond neuropathological burden. Using a residual-based operationalization of cognitive resilience<sup>37</sup>, we derived a continuous score representing cognitive function relative to that expected from pathological burden and demographic covariates (Methods), with higher values indicating greater resilience. In both the discovery cohort and non-overlapping PsychAD donors, Dx\_dev was associated with lower residual-based cognitive-resilience scores (**Fig. 3g**). As a complementary analysis, PsychAD donors classified as cognitively resilient based on substantial AD neuropathology but limited clinical dementia showed lower mean Dx\_cell across comparable levels of pathology (**Extended Data Fig. 8d**).

We further tested whether this relationship generalized to directly measured cognitive outcomes in the independent ROSMAP cohort<sup>14</sup>. After projection of the frozen Dx\_cell models, higher Dx\_dev was consistently associated across major cell classes with poorer cognitive performance at the last clinical assessment, measured by Mini-Mental State Examination (MMSE)<sup>38</sup>, and with greater longitudinal decline in global cognition<sup>39</sup>, after adjustment for donor pathology and relevant covariates (**Fig. 3h**). Both associations remained significant after correction for multiple testing across cell classes. Collectively, higher Dx\_dev was associated with lower relative abundance of vulnerable neuronal populations and poorer cognitive outcomes across cohorts.

## Cellular disease-state decomposition reveals shared molecular programs

The preceding analyses identified cellular disease-state variation at two distinct levels: heterogeneity among cells within the same sample and systematic differences between samples after accounting for measured pathology. We next asked whether these sources of heterogeneity reflected distinct molecular processes or different manifestations of shared AD-associated programs.

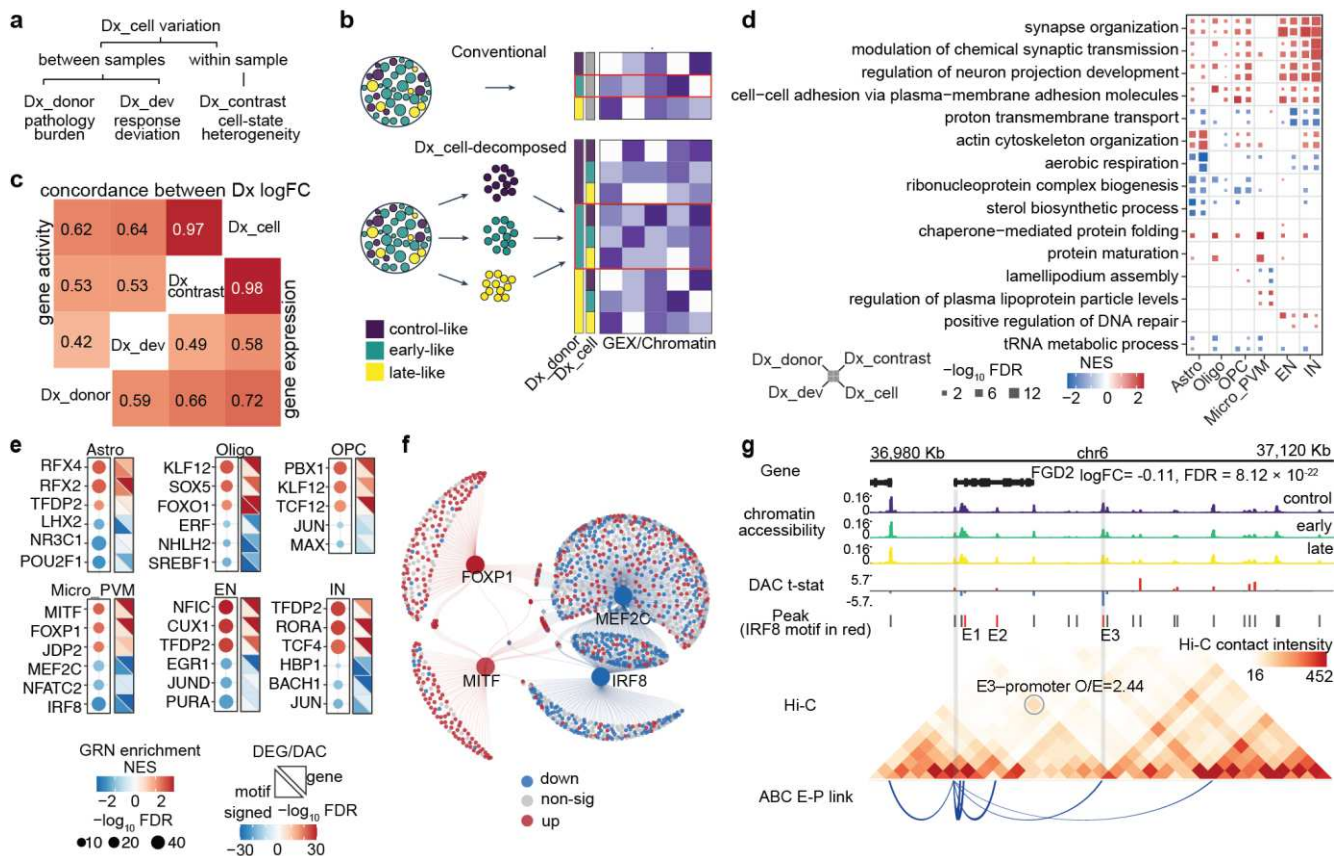


**Fig. 4. Cellular disease-state decomposition reveals shared molecular programs and regulatory architecture.**

**a**, Hierarchical representation of molecular variation captured by *Dx\_cell*. *Dx\_donor* represents measured donor-level pathological burden, *Dx\_dev* represents the pathology-adjusted cellular response deviation between samples, and *Dx\_contrast* represents the within-sample molecular contrast among *Dx\_cell*-defined states. *Dx\_contrast* was centered within each sample and major cell class. **b**, Schematic of conventional and *Dx\_cell*-decomposed pseudobulk analyses. In conventional pseudobulk analysis, cells from each sample and major cell class are aggregated into a single molecular profile. In the *Dx\_cell*-decomposed analysis, cells from the same sample and major cell class are further separated into control-like, early-like, and late-like states before aggregation, generating state-resolved pseudobulk profiles while retaining sample and donor structure in downstream models. **c**, Concordance of marginal molecular effects associated with *Dx\_cell*, *Dx\_donor*, *Dx\_dev* and *Dx\_contrast*. Heatmaps show pairwise Spearman correlations between genome-wide effect estimates across major cell classes for gene expression and ATAC-derived gene activity. *Dx\_donor*, *Dx\_dev*, *Dx\_contrast*, and *Dx\_cell* showed broad positive concordance consistent with shared AD-associated molecular programs. **d**, Pathway enrichment of molecular effects associated with *Dx\_cell*, *Dx\_donor*, *Dx\_dev* and *Dx\_contrast* across major cell classes. Gene-set enrichment analysis was performed using marginal differential-association statistics. Redundant pathways were reduced using semantic similarity, and representative pathways are shown. Representative pathways were selected using  $FDR < 0.05$  in at least one major cell class. For selected pathways, enrichment results with nominal  $P < 0.05$  are displayed, including results that do not individually meet the  $FDR$  threshold. Symbol shape denotes the corresponding disease-state measurement. **e**, GRN prioritization among concordant TF-motif candidates. For each major cell class, the three strongest up-associated and three strongest down-associated eRegulons were selected based on enrichment of predicted targets among *Dx\_cell*-associated genes (dot). Split squares show the signed significance of TF expression and cognate motif-group changes. **f**, Regulatory architecture of selected Micro\_PVM eRegulons. Networks show *MITF*-, *FOXP1*-, *MEF2C*-, and *IRF8*-centered regulatory modules and their predicted target genes. Target nodes are colored according to the direction of *Dx\_cell*-associated differential expression, and edges denote predicted TF-target relationships. **g**, Locus-level regulatory support for the *IRF8*-*FGD2* relationship in Micro\_PVM. The chr6 *FGD2* locus is shown with the *FGD2* gene model and its *Dx\_cell*-associated expression change ( $\log FC = -0.11$ ,  $FDR = 8.12 \times 10^{-22}$ ); chromatin accessibility profiles across control-, early-, and late-like *Dx\_cell* states; signed DAC t-statistics; accessible peaks with *IRF8*-family motif highlighted in red; Hi-C contact intensity; and inferred enhancer-promoter (E-P) links. Three *FGD2*-linked cCREs containing *IRF8*-family motifs were identified (E1-E3) and were linked to the *FGD2* promoter by activity-by-contact (ABC) E-P link analysis. E3 also showed

elevated three-dimensional contact with the *FGD2* promoter (O/E = 2.44), providing convergent locus-level support for the inferred *IRF8–FGD2* regulatory relationship.

To address this question, we decomposed cellular disease-state variation across three nested scales (**Fig. 4a**): donor-level pathological burden (Dx\_donor), pathology-adjusted differences in mean cellular response between samples (Dx\_dev), and contrasts among Dx\_cell-defined states coexisting within the same sample (Dx\_contrast). Dx\_contrast was calculated by centering state-specific mean Dx\_cell within each sample and major cell class and was therefore largely orthogonal to both Dx\_donor and Dx\_dev (**Extended Data Fig. 9a**). We further generated state-resolved pseudobulk profiles for each sample, major cell class, and Dx\_cell category, allowing minority cellular disease states to be analyzed without being diluted in conventional sample-level pseudobulk profiles (**Fig. 4b**).

We first asked whether these statistically distinct disease-state dimensions were associated with distinct or overlapping molecular responses. Joint models including Dx\_donor, leave-one-cell-class-out Dx\_dev and Dx\_contrast indicated that all three dimensions contributed to variation in established AD-associated expression and gene-activity features (**Extended Data Fig. 9b**). We then characterized the molecular program associated with each disease-state measure. Despite their near-orthogonality at the predictor level (**Extended Data Fig. 9a**), genome-wide gene expression and chromatin accessibility effects associated with Dx\_donor, Dx\_dev, and Dx\_contrast were substantially concordant (**Fig. 4c**; **Extended Data Fig. 9c-d**), indicating that higher Dx\_dev is associated with many of the same molecular programs observed with increasing pathological burden. Similar relationships were observed for ATAC-derived gene activity. The molecular effects were also concordant with published AD transcriptomic<sup>8,14</sup> and chromatin datasets<sup>13,40</sup> (**Extended Data Fig. 9e**).

Consistent with this molecular convergence, pathway analyses revealed substantial sharing of AD-associated biological processes across the different disease-state dimensions (**Fig. 4d**). Neuronal, metabolic, inflammatory, and protein-homeostasis programs previously implicated in AD<sup>41</sup>, were represented across multiple Dx measures, although their relative enrichment differed by cell class and disease-state dimension. Notably, Dx\_dev showed prominent associations with protein-folding and proteostasis-related processes across several major cell classes, suggesting that Dx\_dev varies along molecular programs that overlap those associated with AD pathology rather than representing an entirely distinct molecular signature. Together, these results indicate that Dx\_donor, Dx\_dev, and Dx\_contrast capture distinct disease-associated dimensions operating across donor, sample, and within-sample scales while converging on overlapping molecular programs.

## Paired multiome profiling resolves cell-type-specific regulatory programs associated with AD

We next leveraged the paired transcriptomic and chromatin profiles to characterize regulatory remodeling associated with Dx\_cell. Across major cell classes, we prioritized candidate TF regulators whose expression and corresponding chromVAR motif activity changed concordantly with Dx\_cell (**Extended Data Fig. 9f**; **Supplementary Table 7**). The resulting candidates included regulators shared across multiple major cell classes, including *FOXO1* and *FOXP1*, lineage-restricted signals such as *IRF8* in Micro\_PVM, and cell-type-dependent associations such as *CREB5*, which changed in opposite directions in Astro and Oligo. Because related TFs can recognize similar sequence motifs, concordant TF expression was used to prioritize candidate regulators among motif-sharing factors<sup>29,42</sup>, whereas chromatin effects were interpreted at the motif-group level rather than as evidence of TF-specific binding.

To further resolve candidate regulators, we integrated these concordant TF–motif signals with cell-type-specific eRegulons and tested whether predicted TF targets were enriched among Dx\_cell-associated up- or downregulated genes (**Fig. 4e**). This analysis identified distinct regulatory programs across cell classes. In Micro\_PVM, *MITF*, *FOXP1* and *JDP2* were among the strongest up-associated regulons, whereas *IRF8*, *MEF2C* and *NFATC2* were preferentially associated with downregulated genes (**Fig. 4f**; **Extended Data Fig. 9g**). Several of these prioritized regulators have previously been implicated in microglial state regulation; *FOXP1* has

been associated with an AD-enriched inflammatory human microglial state<sup>33</sup>, *MITF* was recently identified as an upstream regulator required to maintain a *GPNMB*+ AD-associated microglial state<sup>34</sup>, and *IRF8* has an established role in organizing microglial enhancer accessibility and transcriptional identity<sup>43</sup>. Functionally, down-associated Micro\_PVM regulons were broadly enriched for inflammatory, cytokine, and taxis-related processes, whereas *MITF* targets were preferentially associated with lipid transport (**Extended Data Fig. 9h**).

At the *FGD2* locus, expression decreased with Dx\_cell ( $\log FC = -0.11$ ,  $FDR = 8.12 \times 10^{-22}$ ), and three promoter-linked<sup>44,45</sup> cCREs contained IRF8-family motifs. The distal E3 element showed the strongest negative differential accessibility among linked cCREs and elevated Hi-C contact<sup>46</sup> with the *FGD2* promoter (observed/expected = 2.44; **Fig. 4g**). These complementary observations support a candidate *IRF8–FGD2* regulatory relationship and illustrate how paired transcriptomic and chromatin measurements inform the regulatory interpretation of cellular response states.

## Reproducible donor transcriptional axes capture pathology-adjusted cellular disease-state heterogeneity

Population-scale single-cell studies have identified substantial inter-individual transcriptional variation in the adult human brain<sup>47</sup>, but whether this variation is reproducible across cohorts and related to pathology-adjusted AD cellular response remains unclear. We first tested the reproducibility of donor-associated expression variation across datasets (**Fig. 5a**). For each major cell class, we compared the fraction of gene expression variance attributable to donor identity in the current cohort with two independent datasets, SEA-AD<sup>9</sup> and Luquez et al<sup>13</sup>. Gene-wise donor-associated variance was broadly concordant across all three cohorts, with particularly strong agreement in glial populations (**Fig. 5a**), indicating that a substantial component of inter-individual transcriptional variation is reproducible across datasets.

Given the concordance of Dx\_dev across major cell classes, we next sought to identify shared donor-level transcriptional programs associated with this variation. Within each cell class, we estimated gene-wise donor effects using best linear unbiased predictions (BLUPs) and performed factor analysis on autosomal protein-coding genes with high donor-attributable expression variance using MOFA<sup>48</sup> (**Extended Data Fig. 10a**). To reduce sensitivity to individual genomic regions or feature selection, we repeatedly estimated the axes after chromosome-block exclusion and additionally repeated the analysis using an expanded set of donor-variable genes (**Methods**). Axis reproducibility was further evaluated by projection into covariate-adjusted expression profiles across brain regions, into the larger PsychAD transcriptomic dataset, and into the independent SEA-AD and Luquez datasets. The three leading axes, F1–F3, showed high reproducibility across these analyses and were selected for downstream characterization (**Fig. 5b**; **Extended Data Fig. 10b**; **Supplementary Table 9**).

The three leading axes were mutually orthogonal in the discovery model (**Extended Data Fig. 10c**) and represented distinct biological programs (**Fig. 5b,c**). F1 was characterized by glial reactive and cytokine-associated signals, including prominent positive loadings for *SERPINA3* and *CHI3L1* in Astro and *KIAA0040* in OPC. F2 represented a broadly shared proteostasis and cellular-stress program, with strong loadings for heat-shock and chaperone genes and enrichment of protein-folding pathways across multiple cell classes. This unsupervised axis extended the protein-folding and cell-stress signal suggested by our earlier PFC analysis<sup>8</sup> by resolving it as a reproducible donor-associated program across cell classes and regions. In contrast, F3 captured a lipid-metabolic program, with cholesterol and other biosynthetic pathways enriched across neuronal and glial populations. These programs involved biological processes previously implicated in AD, including reactive glial and inflammatory states<sup>49,50</sup>, protein homeostasis and vulnerability to A $\beta$  and tau aggregation<sup>51</sup>, and lipid and cholesterol homeostasis<sup>52</sup>.

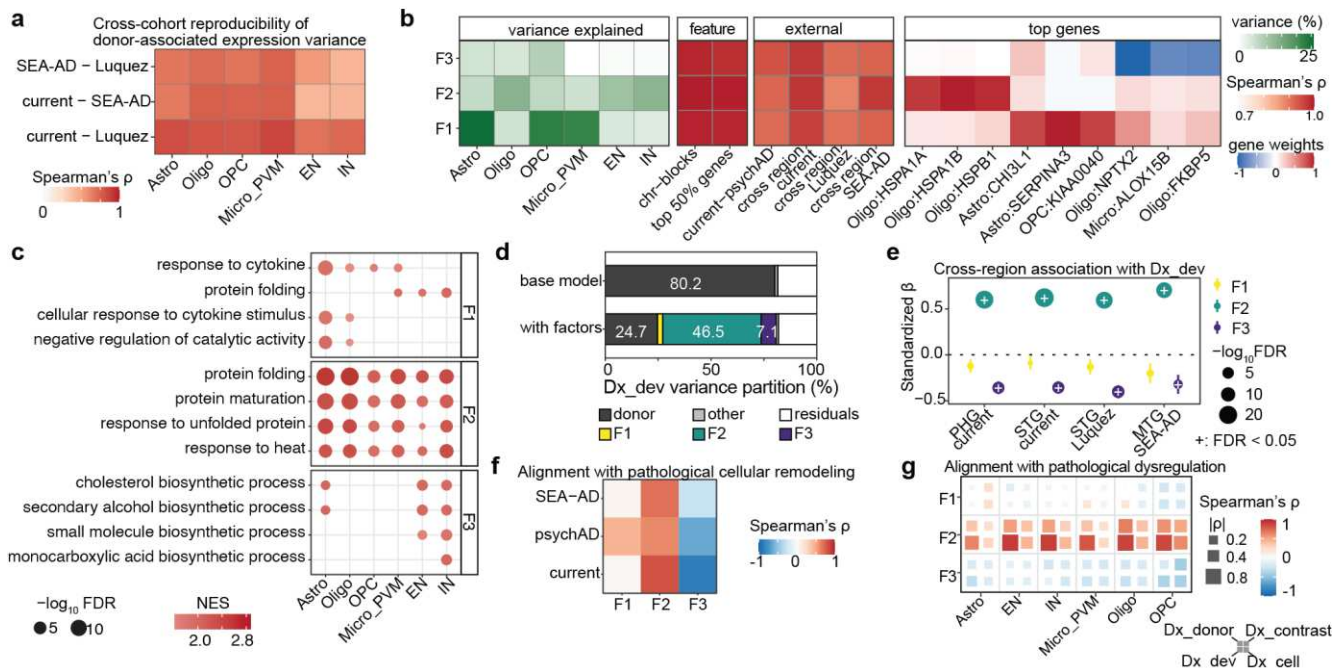


**Fig. 5. Reproducible donor-associated transcriptional axes capture pathology-adjusted disease-state heterogeneity.** **a**, Cross-cohort reproducibility of inter-individual transcriptional variability. Heatmap shows pairwise Spearman correlations of gene-wise expression variance attributable to donor identity across the current cohort, SEA-AD<sup>9</sup>, and Luquez et al.<sup>13</sup>, separately for major cell classes. Donor-associated expression variability was broadly concordant across cohorts, with stronger concordance in glial populations. **b**, Identification and robustness of the three leading donor-associated transcriptional axes (F1–F3). Left, fraction of donor-associated gene expression variance captured by each axis within each major cell class. Middle, reproducibility of factor scores across feature perturbations and independent projections, quantified by Spearman correlation after factor alignment. Feature robustness was evaluated using balanced leave-one-chromosome-block-out analyses and an expanded feature set comprising top 50% of donor-variable genes. External reproducibility was assessed by projection into PsychAD and by cross-region projection in the current cohort, Luquez et al.<sup>13</sup>, and SEA-AD. Right, representative high-loading genes for F1–F3. **c**, Biological processes associated with F1–F3. Dot plots show selected Gene Ontology pathways enriched among axis-associated genes across major cell classes. Only significant enrichments are shown. NES, normalized enrichment score. **d**, Donor-associated transcriptional axes capture a substantial component of Dx\_dev variation. Variance partitioning of Dx\_dev is shown for a base model containing donor identity and other modeled covariates and for a model additionally including F1–F3. Addition of the factors reduced the variance fraction attributed to donor identity from 80.2% to 24.7%, with F2 accounting for the largest axis-associated component. Bars show the percentage of total Dx\_dev variance attributed to each modeled component and residual variation. **e**, Forest plots show the standardized regression coefficient ( $\beta \pm se$ ) of cross-region association of donor-associated axes with Dx\_dev. Factor scores derived or projected from PFC were tested for association with Dx\_dev measured in independent brain regions in the current cohort and external datasets. Corresponding within-PFC associations are shown as a same-region reference in Extended Data Fig. 10f. **f**, Alignment of axis-associated cellular remodeling with pathology-associated cellular composition changes. For each cohort, Spearman correlation was calculated between the vector of cell-composition effects associated with each factor and the corresponding pathology-associated composition effects, using Dx\_donor in the current and PsychAD cohorts and CPS in SEA-AD. F2-associated composition changes consistently aligned with pathology-associated remodeling, whereas F3-associated changes were directionally opposed. **g**, Molecular alignment between donor-associated axes and disease-state dimensions. For each major cell class, MOFA gene loadings were compared with genome-wide marginal transcriptional effects associated with Dx\_donor, Dx\_dev, Dx\_contrast, and Dx\_cell. Color indicates the signed Spearman correlation, and square size indicates its absolute magnitude. F2 showed broad positive alignment with AD-associated transcriptional effects, including prominent concordance with Dx\_dev, whereas F3 showed predominantly opposing molecular associations; F1 showed comparatively weaker alignment.

We then tested the relationships of these donor programs with Dx\_dev. F2 showed a strong positive association with Dx\_dev, whereas F3 showed an opposing association; F1 showed comparatively little relationship with Dx\_dev. In contrast, all three axes showed relatively weak associations with Dx\_donor and age (**Extended Data**

**Fig. 10d**). In the base variance-partition model, donor identity accounted for 80.2% of modeled Dx\_dev variation. After F1–F3 were added jointly, the variance attributed to donor identity decreased to 24.7%, while F2 captured 46.5% and F3 7.1%, with a smaller contribution from F1 (**Fig. 5d**). Together, F1–F3 accounted for approximately 56% of modeled Dx\_dev variation, with the largest contribution from the proteostasis and cellular-stress axis.

We next asked whether the relationships between these axes and Dx\_dev extended across brain regions. If these axes capture donor-shared transcriptional variation rather than variation restricted to an individual sampled region, factor scores derived from one region should remain associated with Dx\_dev measured in another brain region from the same donor. PFC-derived factor scores were associated with Dx\_dev in both PHG and STG, with F2 consistently associated with higher Dx\_dev and F3 with lower Dx\_dev (**Fig. 5e**; **Extended Data Fig. 10e**). The corresponding cross-region effect estimates were similar in direction and magnitude to the within-PFC associations, which are shown as a same-region reference (**Extended Data Fig. 10f**). The opposing relationships were also observed across additional brain regions and after projection into the Luquez et al. and SEA-AD datasets. Together, these results indicate that the F2 and F3 relationships with Dx\_dev are not restricted to the local transcriptional state of the region from which the axes were estimated. We further examined whether the opposing F2 and F3 relationships with Dx\_dev were also reflected in AD-associated cellular composition changes. Using glia-projected factor scores, we compared factor-associated cell-composition effects with the corresponding pathology-associated effects. F2-associated composition changes were strongly aligned with pathology-associated remodeling in the current cohort (Spearman's  $\rho = 0.81$ ), PsychAD ( $\rho = 0.61$ ), and SEA-AD ( $\rho = 0.70$ ). In contrast, F3-associated effects were directionally opposed to pathology-associated remodeling ( $\rho = -0.88, -0.64$  and  $-0.29$ , respectively), whereas F1 showed little concordance in the current and SEA-AD cohorts and a more modest positive relationship in PsychAD (**Fig. 5f**; **Extended Data Fig. 10g**).

Finally, we examined how these donor-associated axes related to the molecular disease-state dimensions defined earlier (**Fig. 4a-c**). We compared continuous MOFA gene loadings with genome-wide marginal association statistics for Dx\_donor, Dx\_dev, Dx\_contrast, and Dx\_cell, avoiding dependence on differences in the number of significant genes identified for each measure. F2 loadings showed broad positive concordance with AD-associated transcriptional effects across cell classes, with particularly strong alignment with the Dx\_dev signature, whereas F3 showed predominantly opposing relationships (**Fig. 5g**). F1 showed substantially weaker genome-wide concordance. Consistent with these patterns, representative F2-loading genes, including the heat-shock and proteostasis genes *HSPA1A*, *HSPA1B* and *HSPB1*, also showed prominent Dx\_dev-associated expression changes (**Extended Data Fig. 10h**). Similar relationships with published AD transcriptional signatures were observed in external datasets<sup>8,14</sup> (**Extended Data Fig. 10i**).

Together, these analyses identify reproducible donor-associated proteostasis and lipid-metabolic programs with opposing associations with pathology-adjusted cellular response and AD-associated cellular composition.

## Discussion

Individuals with comparable measured AD pathology differed in a coordinated cellular response shared across major cell classes and cortical regions. Higher responses were associated with lower relative abundance of vulnerable neuronal populations and poorer cognitive outcomes across cohorts. These between-individual differences and within-sample molecular-state differences converged with pathology-associated changes on overlapping transcriptional and chromatin programs. Reproducible proteostasis and lipid-metabolic programs showed opposing associations with this response, identifying a molecular dimension of AD heterogeneity that complements measured neuropathological burden.

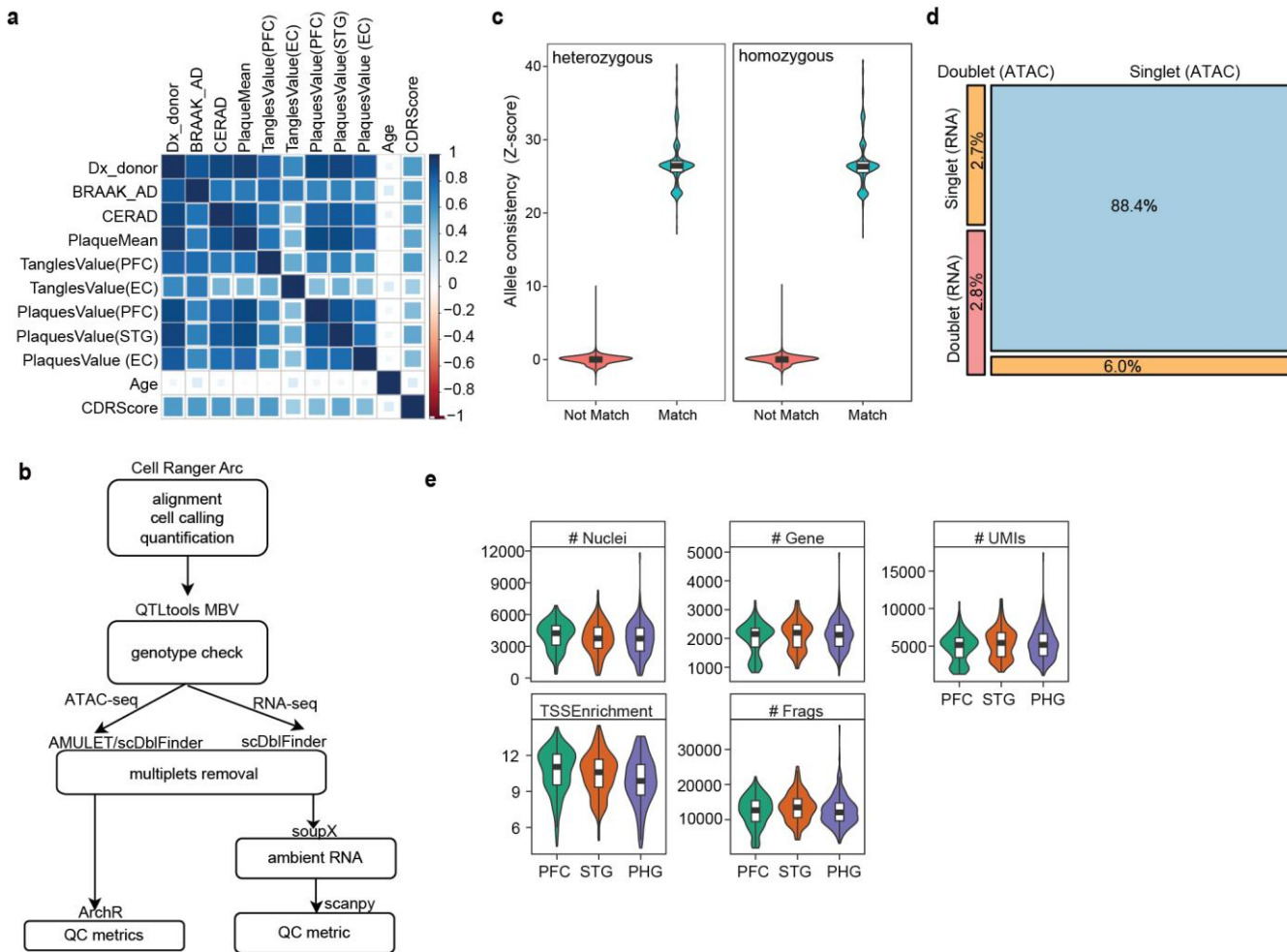
Within individual samples, Dx\_cell resolved substantial variation that cannot be represented by a single donor- or sample-level pathology value. Such heterogeneity is biologically plausible given the spatially restricted distribution of amyloid and tau lesions<sup>1,16</sup>, and emerging evidence that molecular responses vary across local tissue environments<sup>17–19</sup>. Dx\_cell models transferred to additional transcriptomic datasets, where the inferred scores retained expected associations. Importantly, RNA-defined Dx\_cell states were accompanied by

corresponding chromatin-associated changes in the same nuclei, supporting the interpretation that they represent coordinated molecular variation rather than RNA measurement noise alone. With donor pathology used as supervision, Dx\_cell should be interpreted as an inferred pathology-associated cellular response, rather than a direct measure of pathological burden within an individual cell. Because the primary Dx\_cell cross-fitting was performed at the cell rather than donor level, these scores should not be interpreted as donor-independent predictive performance; concordance with donor-held-out PsychAD models and transfer to external datasets provide complementary validation. Whether the observed within-sample variation primarily reflects local pathological exposure, intrinsic differences in cellular susceptibility, intercellular signaling, or combinations of these processes, will require spatially resolved measurements that directly integrate molecular states with amyloid, tau, and other pathological features.

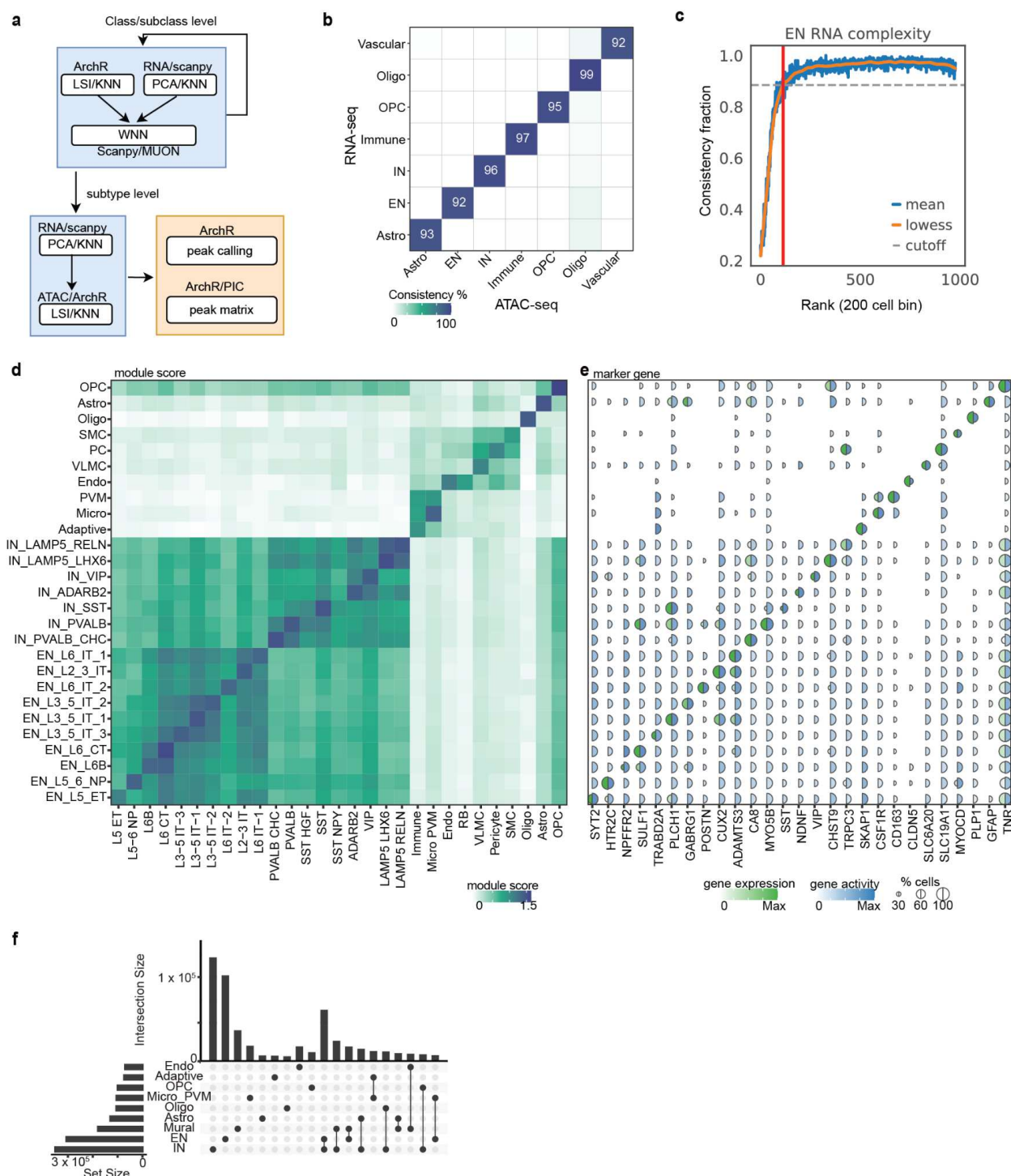
Cellular disease state also differed systematically between individuals after adjustment for measured pathology and relevant covariates. Dx\_dev was concordant across independently modeled cell classes, cortical regions, and datasets, indicating that this variation contains a reproducible donor-associated component rather than being restricted to a particular cell type or tissue specimen. Higher Dx\_dev was accompanied by AD-aligned cellular remodeling, including lower relative abundance of selectively vulnerable excitatory and inhibitory neuronal populations and greater representation of disease-associated microglia. These relationships generalized across independent cohorts and remained when neuronal outcomes were tested using glia-derived Dx\_dev, reducing direct coupling between the molecular predictor and neuronal composition. These associations persisted in SEA-AD, where quantitative neuropathology was measured in the profiled tissue, reducing concerns about anatomical mismatch between pathology and molecular measurements. Contributions from unmeasured pathology and measurement error nevertheless remain possible. Dx\_dev was also associated with lower cognitive resilience, poorer terminal cognitive performance, and greater longitudinal cognitive decline. Thus, a molecular phenotype defined independently of cognitive outcome nevertheless tracked clinically meaningful differences among individuals with similar measured pathology.

Unsupervised analysis identified reproducible donor-associated transcriptional programs related to Dx\_dev. A proteostasis and cellular-stress program was associated with higher Dx\_dev and AD-aligned molecular and cellular-composition differences, whereas a lipid-metabolic program showed opposing associations. A reactive glial and cytokine-associated program captured additional donor variation but was less strongly related to Dx\_dev. The factor analysis and disease-state models used the same underlying transcriptomic measurements; their concordance therefore supports a reproducible molecular relationship, with further support from associations across regions and cohorts. Higher stress-response expression may reflect cellular stress or compensatory responses, and the opposing lipid-metabolic association does not establish a protective function. Cross-sectional postmortem data cannot determine whether these programs modify cellular response or arise downstream of disease processes. Dx\_dev may also reflect unmeasured pathological features or other exposures and should not be interpreted as a pathology-independent resilience mechanism. Together, these findings provide a framework for investigating why similar measured neuropathological burdens are associated with different neuronal and cognitive outcomes.

## Extended Data Figures.

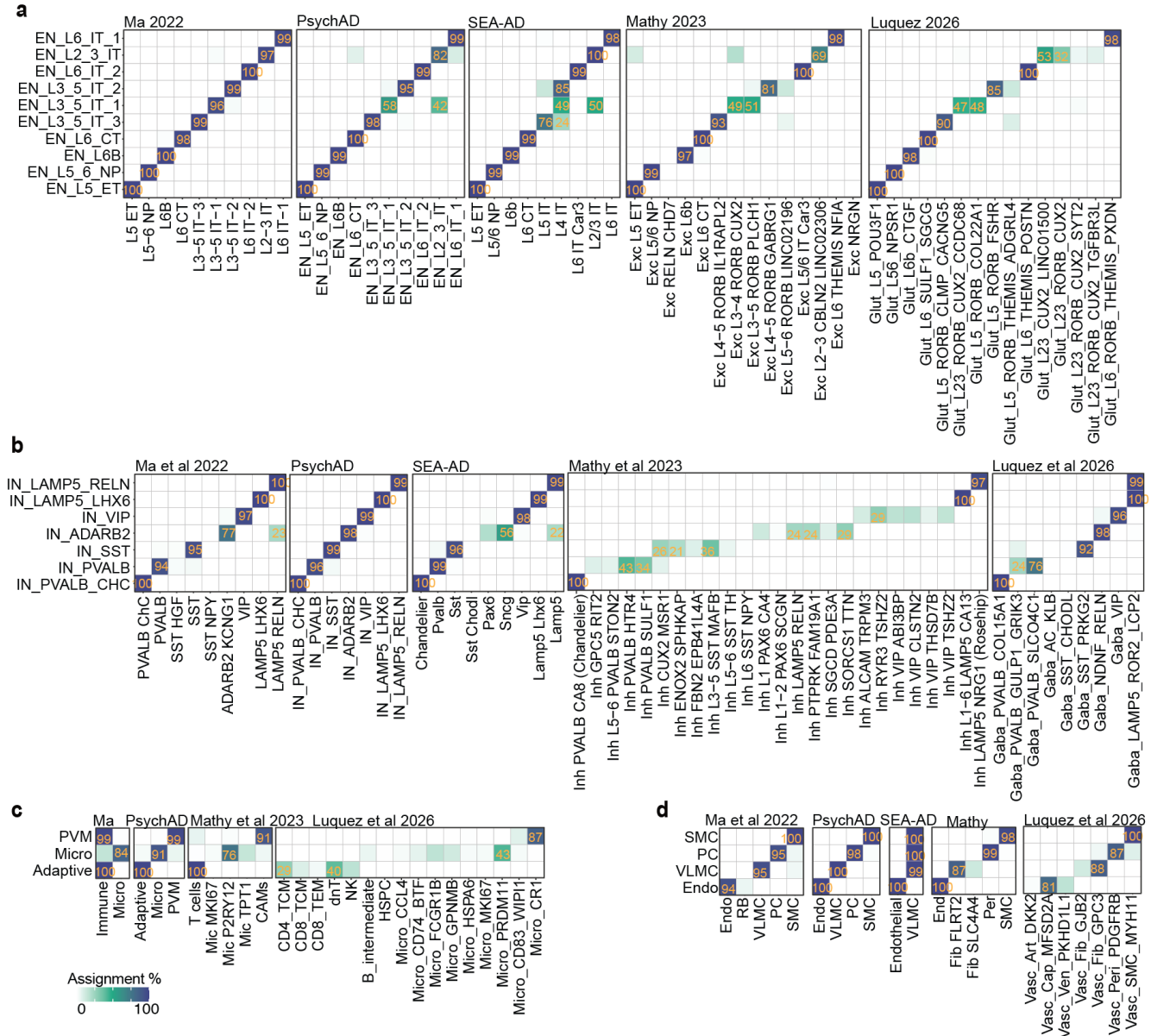


**Extended Data Fig. 1. Overview of cohort, sample processing, and QC.** **a**, Dot plot showing correlation between the donor-level pathology score, Dx\_donor, and AD-related neuropathological measurements. **b**, Overview of the snMultiome data processing workflow. **c**, Violin and box plots showing Z-score-normalized allele concordance between each snMultiome sample and matched or non-matched donor genotypes, shown separately for heterozygous and homozygous alleles. **d**, Mosaic plot showing the overlap between doublet calls derived from the RNA and ATAC modalities. RNA doublets were identified using scDbfFinder, and ATAC doublets were identified using the combined ATAC-based doublet-calling approach. **e**, Violin and box plots summarizing sample-level quality-control metrics for the final retained snMultiome data. Each point represents one sample. Metrics include the number of final nuclei (# Nuclei) and sample-level medians for the number of detected genes (# Genes), the number of unique molecular identifiers (# UMIs), the transcription start site enrichment score (TSSEnrichment), and the number of ATAC fragments (# Frags). The box plot indicates the median, interquartile range (IQR), and  $1.5 \times \text{IQR}$ .



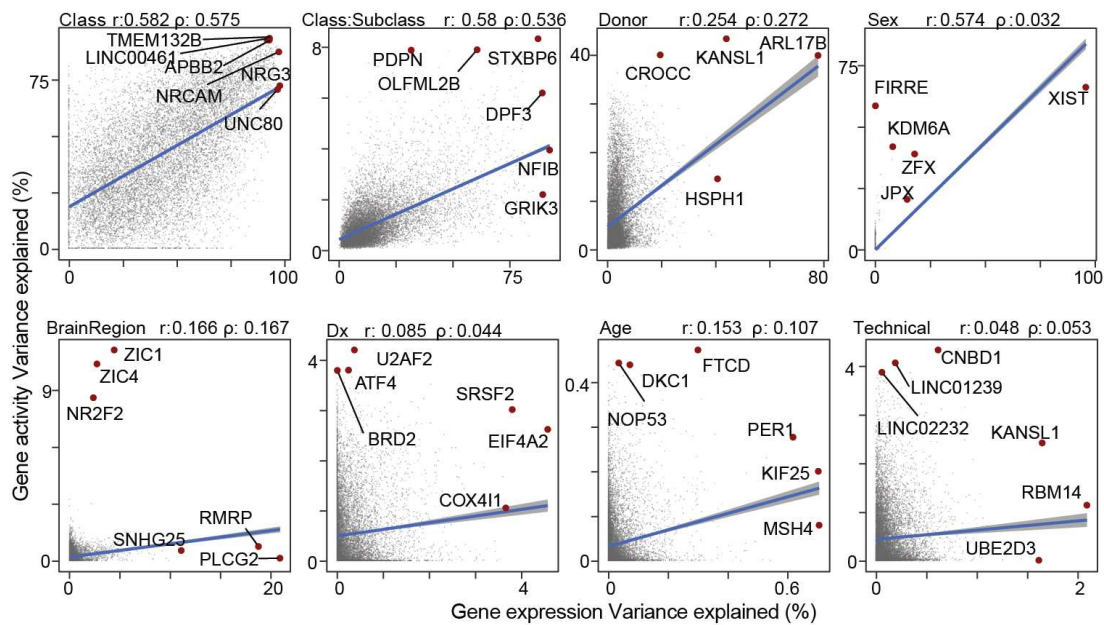
**Extended Data Fig. 2. Iterative snMultiome taxonomy construction.** **a**, Overview of the iterative snMultiome taxonomy workflow. **b**, Concordance between RNA-derived and ATAC-derived major-class annotations. Values are row-normalized by RNA-derived major cell class. **c**, Example of adaptive QC in excitatory neurons. RNA complexity was ranked and binned, and RNA-ATAC major-class concordance was modeled across bins. Low RNA-complexity regions showed reduced RNA-ATAC concordance and were removed during adaptive QC. **d**, Module scores across subclasses using marker genes from Ma et al. 2022<sup>31</sup>. **e**, Subclass marker gene expression and gene activity scores are shown as split dot plots. For each dot, the left half shows gene expression, and the right half shows gene activity score. Color represents the mean normalized expression or gene activity score, and dot size represents the percentage of cells with detectable gene expression or gene activity greater than zero. **f**, UpSet plot detailing shared and unique peaks across major cell classes. The plot

highlights the total peak count per major cell class alongside the top 10 largest intersections of peaks shared between multiple major cell classes.

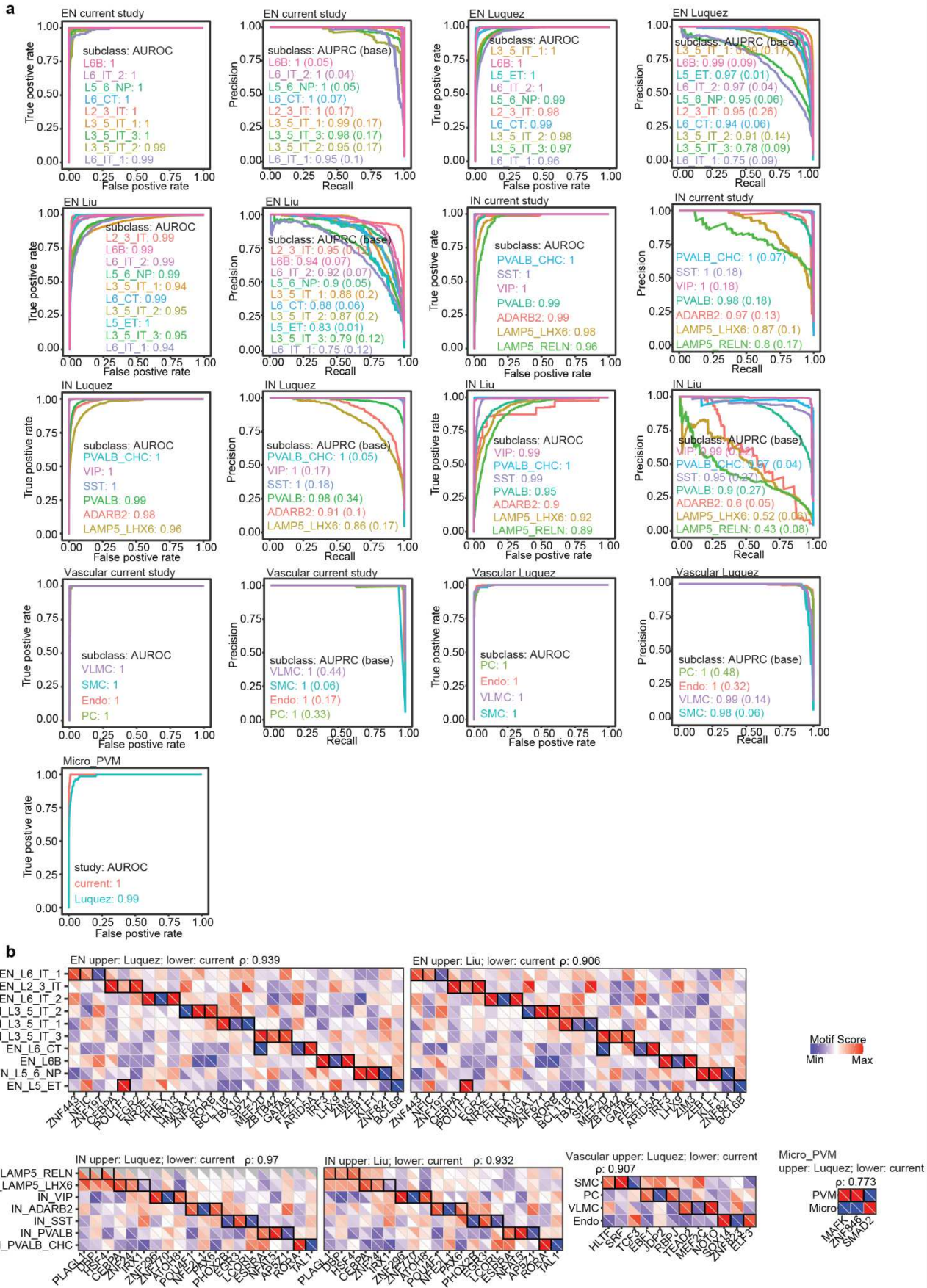


**Extended Data Fig. 3. Reference-based validation of subclass annotations by scANVI label transfer.** Heatmaps showing scANVI-based label transfer between our snMultiome subclasses and published reference taxonomies for **a**, EN; **b**, IN; **c**, immune cells, including Micro\_PVM and adaptive immune cells; and **d**, vascular cells, including Mural and Endo. Reference datasets included Ma et al. 2022<sup>31</sup>, PsychAD<sup>8</sup>, SEA-AD<sup>9</sup>, Mathys et al. 2023<sup>14</sup>, and Luquez et al. 2026<sup>13</sup>. For each panel, rows represent query subclasses from the current study, columns represent reference subclass labels, and colors indicate the percentage of cells from each query subclass assigned to each reference label. Percentages are normalized within each query subclass.

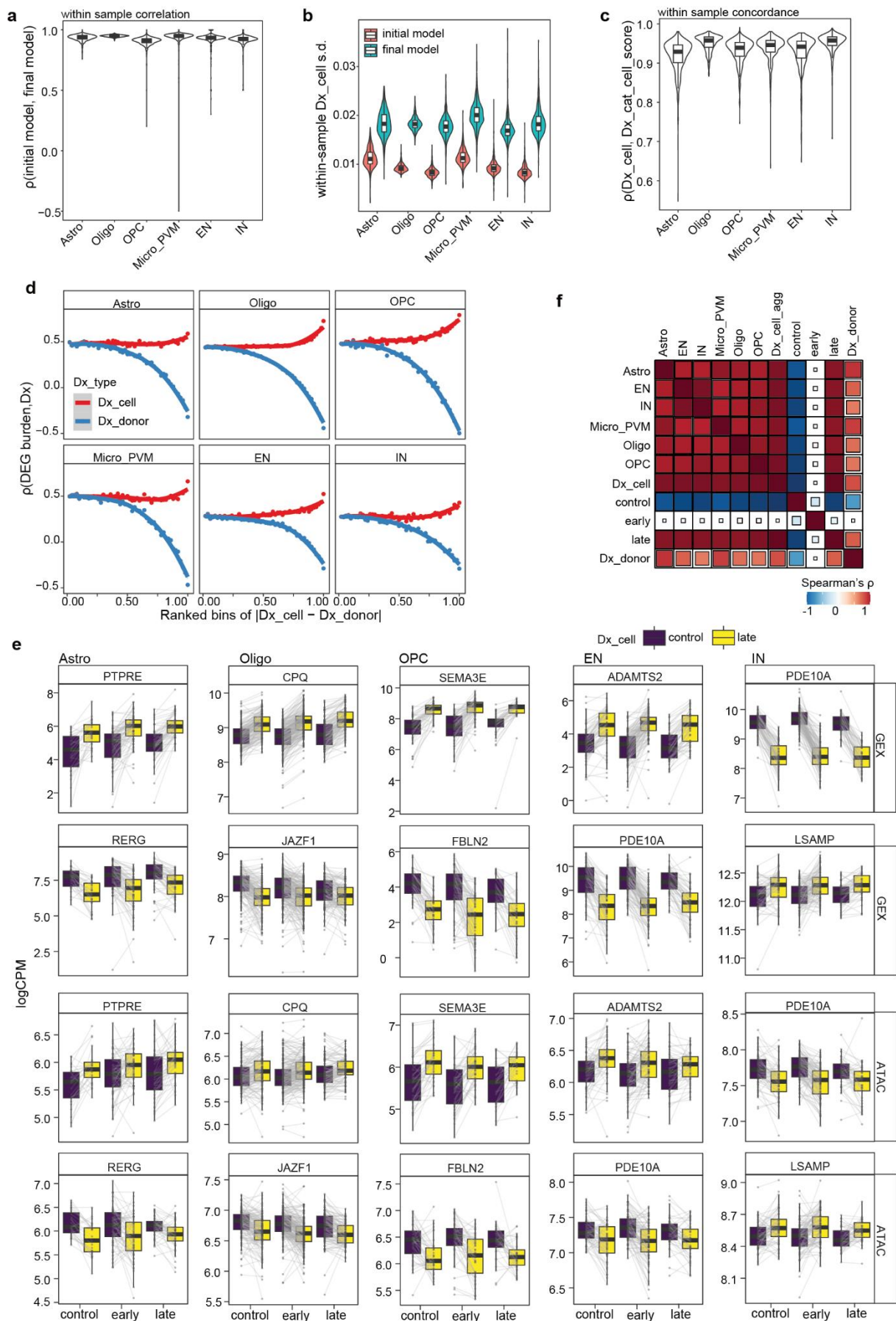




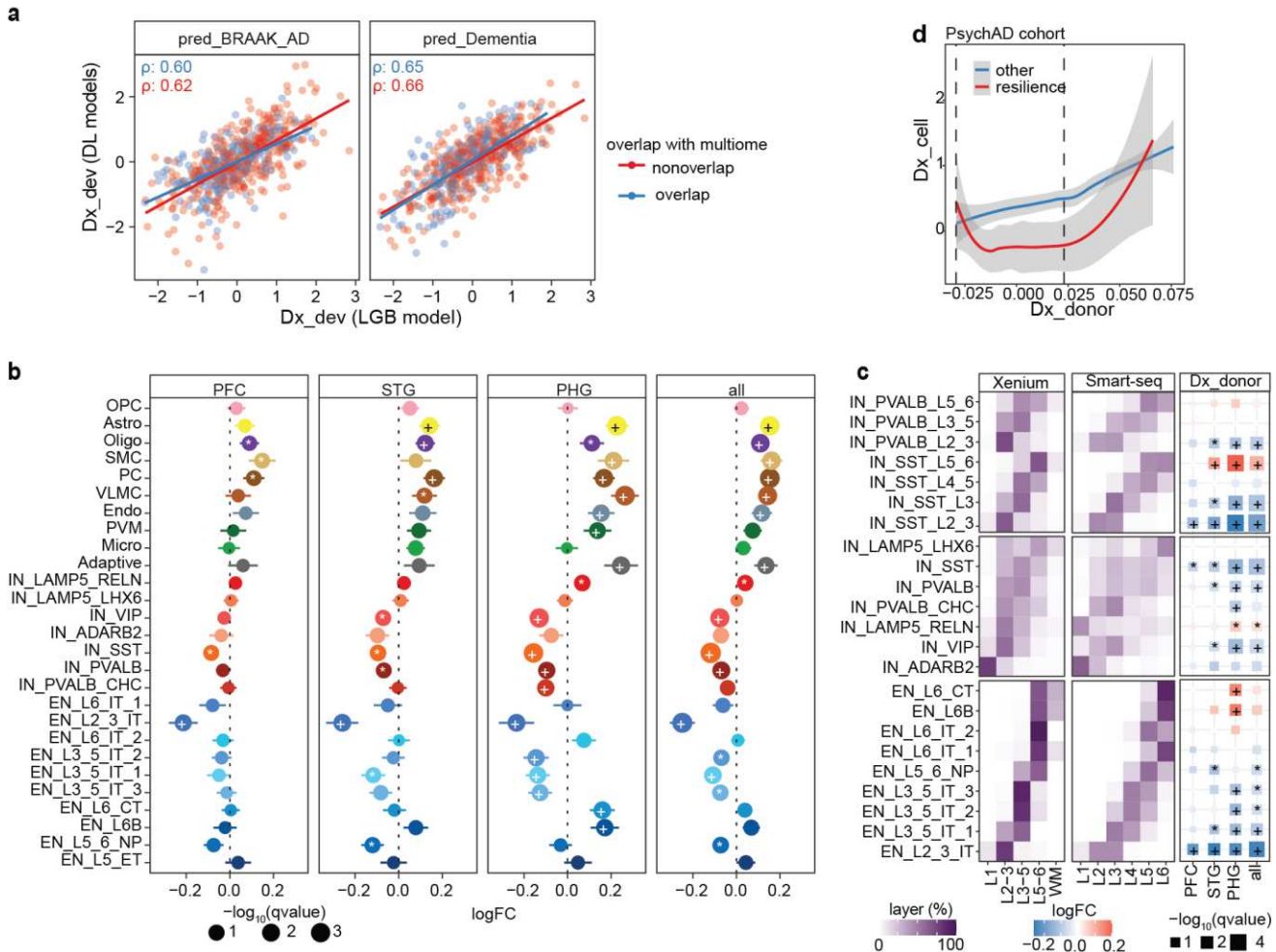
**Extended Data Fig. 5 Variance partitioning of gene expression and gene activity scores.** Matched gene expression and gene activity variance explained across different covariates. Pearson's correlation coefficient ( $r$ ) and Spearman's rank correlation coefficient ( $\rho$ ) are shown.



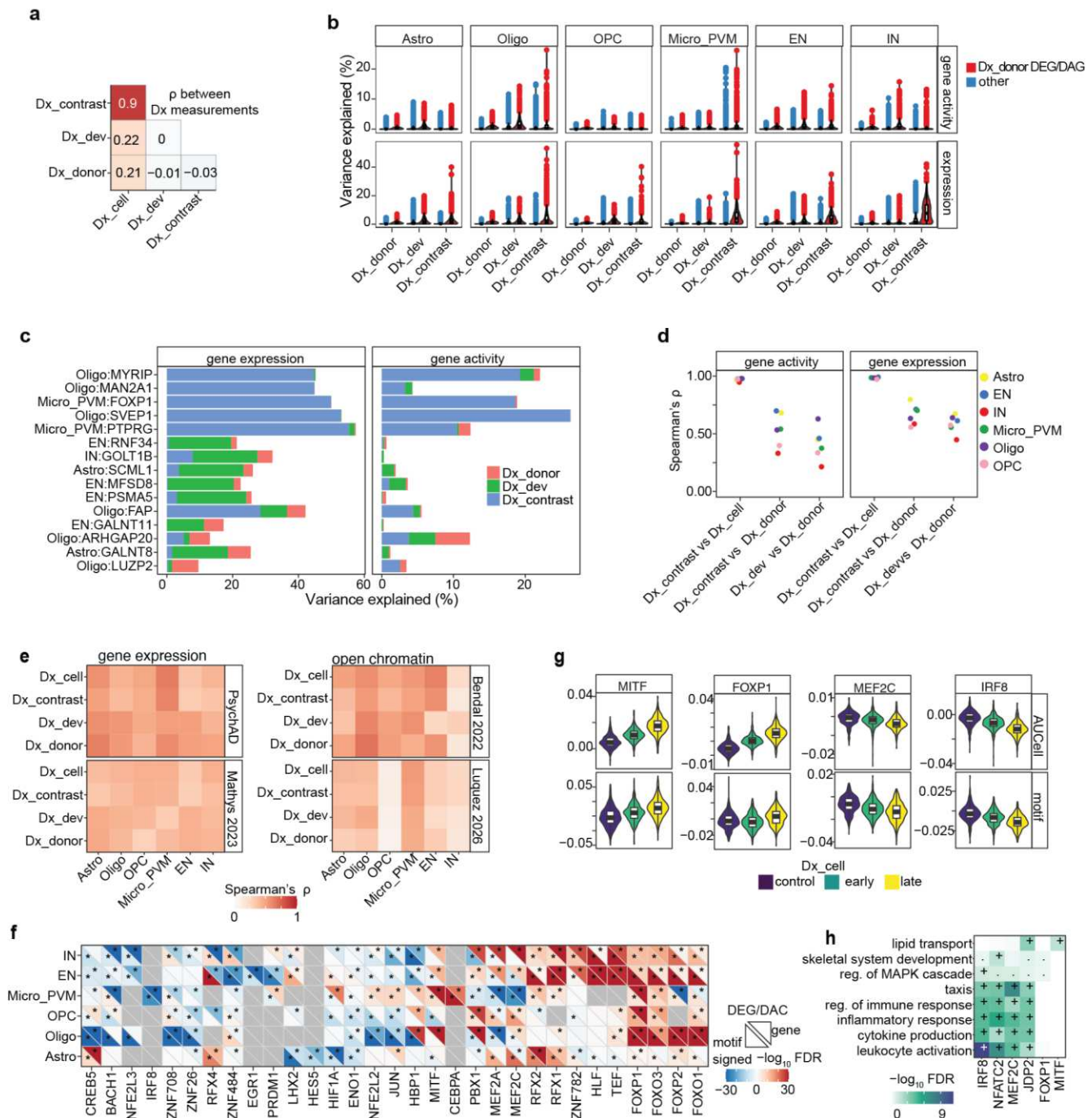
**Extended Data Fig. 6 Cross-cohort validation of subclass-associated chromatin motif signatures.** **a**, Performance of three-motif signatures for predicting subclass identity within major cell classes. Receiver operating characteristic (ROC) and precision–recall (PR) curves are shown for the current study and two independent validation cohorts (Luquez et al 2026<sup>13</sup> and Liu et al 2025<sup>12</sup>). AUROC is reported for ROC curves, and area under the precision–recall curve (AUPRC) together with the prevalence-based baseline AUPRC is reported for PR curves. **b**, Cross-cohort reproducibility of subclass-associated chromatin motif activity. chromVAR scores are shown for the selected signature motifs used to distinguish subclasses in the discovery cohort and the two validation cohorts. Signature motifs are highlighted in black. Spearman’s correlation coefficient ( $\rho$ ) summarizes concordance of motif activity between the discovery and validation datasets.



**Extended Data Fig. 7. Robustness and same-sample validation of Dx\_cell models.** **a**, Within-sample Spearman correlation between Dx\_cell scores from the initial model and the final anchor-refined model across six major cell classes. Each data point represents one sample within one major cell class. **b**, Distribution of within-sample standard deviation for raw, unscaled Dx\_cell scores from the initial and final anchor-refined models across six major cell classes. Statistical significance was assessed using paired two-sided Wilcoxon signed-rank tests followed by FDR correction (all FDR <  $1 \times 10^{-16}$ ). **c**, Within-sample Spearman correlation between continuous regression-derived Dx\_cell and the probability-weighted classification-derived disease-state score, Dx\_cat\_score, across six major cell classes. Dx\_cat\_score was calculated as  $0 \times P_{\text{control-like}} + 1 \times P_{\text{early-like}} + 2 \times P_{\text{late-like}}$ . **d**, Validation of Dx\_cell using RNA-seq-derived disease burden. Cells were divided into 50 ordered non-overlapping bins ranked by increasing  $|Dx_{\text{cell}} - Dx_{\text{donor}}|$ . Within each bin, Spearman's correlation coefficient,  $\rho$ , was calculated between transcriptomic disease burden derived from PsychAD DEGs and either Dx\_cell or Dx\_donor. **e**, Same-sample validation of Dx\_cell-defined disease states across major cell classes. AD-associated marker gene expression or ATAC-derived gene activity values are shown for Dx\_cell-defined control-like and late-like cells within the same sample. Lines connect paired values from the same sample. Analyses were restricted to samples with at least 20 control-like and 20 late-like cells in the corresponding major cell class. Statistical significance was assessed using paired two-sided Wilcoxon signed-rank tests comparing control-like and late-like pseudobulk values within each donor-level pathology category, followed by FDR correction across all same-sample marker tests (all FDR < 0.05). Box plots show the median, interquartile range, and  $1.5 \times$  interquartile range. **f**, Correlation matrix comparing sample-level mean Dx\_cell scores across Astro, Oligo, OPC, Micro\_PVM, EN, and IN, together with aggregate mean Dx\_cell across cell classes, disease-state cell composition, and Dx\_donor.



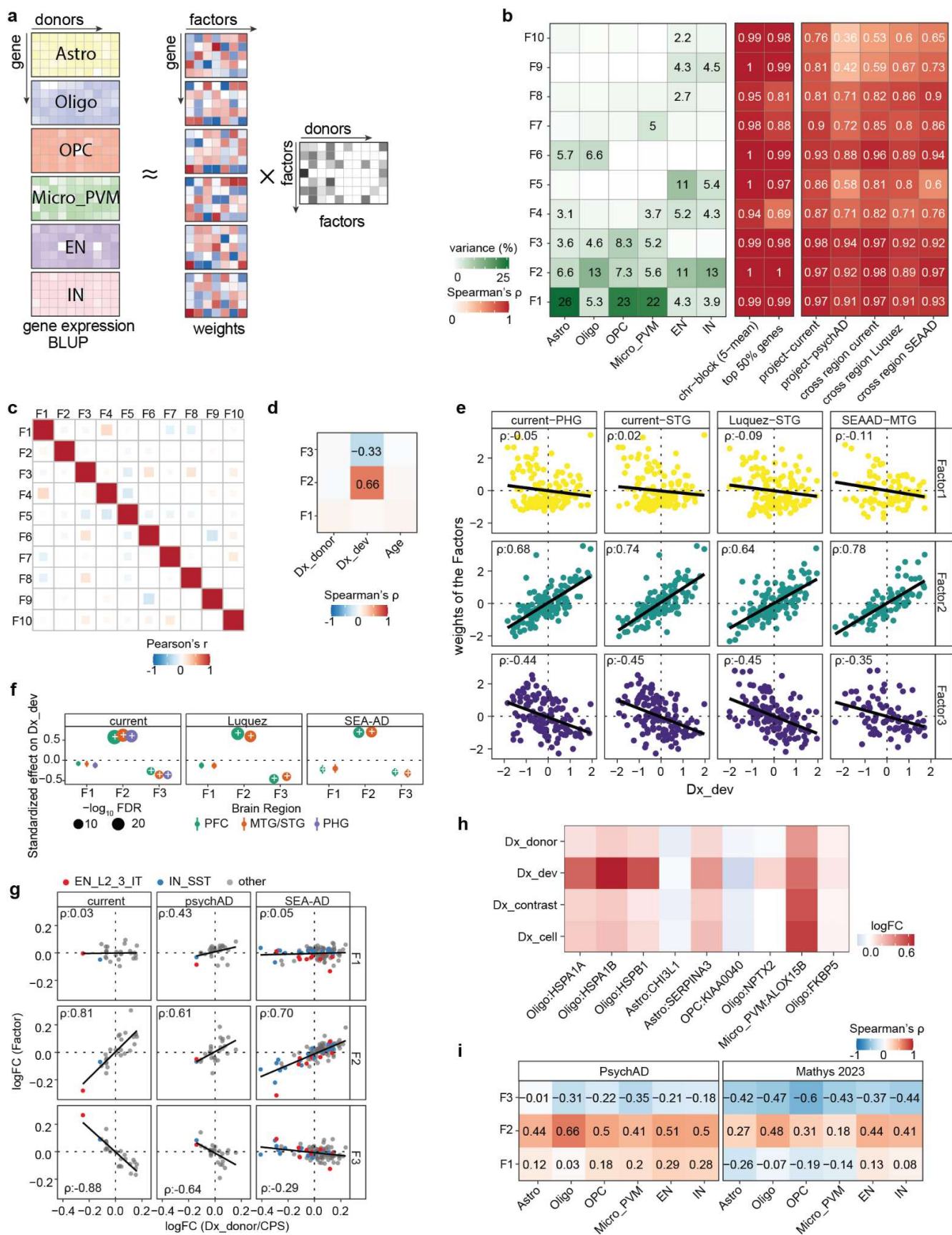
**Extended Data Fig. 8. Reproducibility and phenotypic characterization of Dx\_dev.** **a**, Cross-model concordance of Dx\_dev. Scatterplots compare Dx\_dev derived from the frozen LightGBM Dx\_cell model with analogous pathology-adjusted deviations derived from previously trained transcriptomic models predicting Braak pathology and dementia. Results are shown separately for donors overlapping and non-overlapping with the multiome cohort. Spearman correlation coefficients ( $\rho$ ) are shown. **b**, Brain-region-specific AD-associated cell-composition changes at subclass resolution. Composition effects are shown across PFC, STG, and PHG. EN\_L2\_3\_IT and IN\_SST showed lower relative abundance, with more pronounced changes in STG and PHG than in PFC. **c**, Relationship between neuronal vulnerability and cortical depth. AD-associated composition changes are shown across neuronal subclasses and, at finer resolution, across layer-associated IN\_SST and IN\_PVALB subtypes. **d**, Relationship between donor pathology, aggregate cellular disease state, and cognitive resilience. Smoothed relationships between Dx\_donor and mean Dx\_cell are shown separately for pathology-positive donors classified as cognitively resilient or non-resilient according to the PsychAD resilience definition. Shaded regions indicate confidence intervals around the fitted curves.



**Extended Data Fig. 9. Molecular characterization and external validation of cellular disease-state decomposition.**

**a**, The heatmap shows pairwise Spearman correlations among Dx\_cell, Dx\_donor, Dx\_dev, and Dx\_contrast. **b**, Joint variance partitioning of molecular variation across disease-state dimensions. Variance partitioning of gene expression and ATAC-derived gene activity using state-resolved pseudobulk profiles. Models included Dx\_donor, Dx\_dev, and Dx\_contrast simultaneously, together with the corresponding covariates and random effects, to estimate the molecular variance uniquely attributable to each disease-state dimension. Violin and box plots summarize the percentage of molecular variance attributed to Dx\_donor, Dx\_dev, and Dx\_contrast across tested genes. Genes identified as significant donor-level differentially expressed genes (DEGs) or differential gene-activity genes (DAGs) in the conventional Dx\_donor analysis are highlighted in red; other tested genes are shown in blue. **c**, Representative genes dominated by distinct disease-state dimensions. For each disease-state dimension, the five genes with the largest gene expression variance attributable to Dx\_donor, Dx\_dev or Dx\_contrast are shown. Stacked bars indicate the fraction of gene expression variance attributable to each of the three disease-state dimensions. Corresponding variance partitions for ATAC-derived gene-activity scores of the same genes are shown alongside the expression results. Examples illustrate genes whose molecular variation is

preferentially associated with donor-level pathology, pathology-adjusted cellular response, or within-sample cellular-state heterogeneity. **d**, Concordance of marginal molecular effects associated with Dx\_cell, Dx\_donor, Dx\_dev and Dx\_contrast across major cell classes. **e**, External concordance of disease-associated transcriptional and chromatin effects. Cell-type-specific molecular effects identified in the current study were compared with external AD datasets using Spearman correlation of signed genome-wide association statistics. For gene expression, effects were compared with published differential-expression results from PsychAD and Mathys et al. For chromatin accessibility, effects were compared with Luquez et al.<sup>13</sup> and previously generated sorted-cell chromatin data<sup>40</sup>. Luquez ATAC-seq data were reprocessed using the peak set defined in the current study, and differential accessibility was tested against CERAD using dreamlet. Correlations were calculated using matched molecular features and cell classes across datasets. **f**, Integrated TF expression and chromatin motif remodeling associated with Dx\_cell across major cell classes. Each square is divided into two triangles representing the Dx\_cell association of TF expression and the corresponding chromVAR motif-group activity. \*: FDR-significant associations. We prioritized the TFs shown based on concordant directions of TF expression and motif activity changes. **g**, eRegulon and motif activity across Dx\_cell-defined states. Violin plots with embedded boxplots show eRegulon AUCCell scores (top) and corresponding chromVAR motif-group scores (bottom) for *MITF*, *FOXP1*, *MEF2C* and *IRF8* across control-, early- and late-like Dx\_cell states. **h**, Functional enrichment of selected Micro\_PVM regulons. Heatmap shows representative Gene Ontology processes enriched among predicted targets of *IRF8*, *NFATC2*, *MEF2C*, *JDP2*, *FOXP1*, and *MITF*.



**Extended Data Fig. 10. Identification, robustness and molecular characterization of donor transcriptional axes. a,** Donor transcriptional axes analysis framework. Schematic illustrating estimation of gene-wise donor effects using best linear

unbiased predictions (BLUPs) within each major cell class and integration of the resulting donor-by-gene matrices using Multi-Omics Factor Analysis (MOFA)<sup>48</sup>. Factor discovery was restricted to the top 25% of protein-coding genes ranked by donor-attributable expression variance. MOFA decomposed the cell-class-specific donor-expression profiles into donor factor scores and cell-class-specific gene loadings. **b**, Variance contribution and robustness of donor transcriptional axes. Left, fraction of donor-associated gene expression variance captured by each latent factor across major cell classes. Middle and Right, factor reproducibility across feature perturbations and independent projections, quantified by Spearman correlation after factor matching and sign alignment. Robustness analyses include five chromosome-block analyses excluding approximately 20% of autosomes per run, an expanded top-50% feature-set sensitivity analysis, projection into held-out/current data, cross-region reproducibility within the current cohort, and projection into independent PsychAD, SEA-AD, and Luquez datasets. **c**, Orthogonality of donor factors. Heatmap showing pairwise Pearson correlations among donor factor scores in the discovery cohort. **d**, Relationships of donor factors with cellular disease state, pathology, and age. Heatmap showing Spearman correlations of F1–F3 with Dx\_dev, Dx\_donor, and donor age. F2 and F3 showed opposing associations with Dx\_dev, whereas associations with Dx\_donor and age were comparatively weak. **e**, Cross-region portability of factor–Dx\_dev associations. Scatterplots showing associations between F1–F3 estimated from PFC and Dx\_dev measured in independent brain regions from the same donors in the current cohort, Luquez et al, and SEA-AD. Spearman correlation coefficients  $\rho$  are shown. **f**, Within- and cross-region associations of donor-associated transcriptional axes with Dx\_dev. Factor scores derived or projected from PFC were tested for association with Dx\_dev measured in PFC and independent brain regions in the current cohort, Luquez et al., and SEA-AD. The PFC–PFC association provides a within-region reference; cross-region comparisons include PFC–PHG and PFC–STG in the current cohort, PFC–STG in Luquez et al., and PFC–MTG in SEA-AD. Forest plot showing standardized regression coefficient ( $\beta \pm se$ ). **g**, Alignment of factor-associated and pathology-associated cellular remodeling. Scatterplots comparing cell-population composition effects associated with F1–F3 (glia-projected) with the corresponding pathology-associated composition effects in the current cohort and independent datasets. Each point represents a cell population; Spearman correlations summarize concordance across populations. Pathology was represented by Dx\_donor where available and by CPS in SEA-AD. **h**, Disease-associated expression of representative factor-loading genes. Heatmap showing molecular effects of Dx\_donor, Dx\_dev, Dx\_contrast and Dx\_cell for representative high-loading genes from F1–F3. All Dx predictors were standardized before differential analysis. **i**, Alignment of donor-associated axes with published AD transcriptional signatures. Heatmaps showing Spearman correlations between cell-class-specific F1–F3 gene loadings and published AD-associated differential-expression statistics from PsychAD and Mathys et al. Effects are compared using the common set of genes measured in each dataset, with factor orientation fixed according to the discovery cohort.

## Methods

### Study cohort

Single-nucleus multiome data were generated from 576 tissue specimens collected from the prefrontal cortex (PFC; Brodmann areas 9/44/46), superior temporal gyrus (STG; BA22), and parahippocampal gyrus (PHG; BA36) of 247 postmortem human donors obtained from the Mount Sinai NIH NeuroBioBank (MSSM) (**Supplementary Table 1**). Neuropsychological, diagnostic, and autopsy protocols were approved by the Institutional Review Boards of Mount Sinai and the James J. Peters VA Medical Center. As part of the PsychAD project<sup>8,20</sup>, the current study included donors with either no discernible neuropathology or cognitive impairment (controls) or AD-associated neuropathology without substantial comorbid neuropathological lesions, such as significant cerebrovascular disease or Lewy body pathology. Clinical and pathological metadata, genotype data, and polygenic risk scores were obtained from the PsychAD project<sup>8,20</sup>.

### Donor-level AD pathology score and disease staging

Neuropathological assessments and evaluations of cognitive, medical, and neurological status were performed according to established procedures<sup>53</sup>, as described previously in detail<sup>8,40</sup>. Because multiple measures of AD neuropathology were highly correlated (**Extended Data Fig. 1a**), we derived an aggregated continuous measure of AD pathological burden using principal component analysis (PCA). Eight neuropathological measures were included: Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuritic plaque score<sup>54</sup>, Braak

neurofibrillary tangle stage<sup>55,56</sup>, mean neuritic plaque density (PlaqueMean), neuritic plaque burden in the middle frontal gyrus, entorhinal cortex, and superior temporal gyrus, and neurofibrillary tangle burden in the middle frontal gyrus and entorhinal cortex.

The pathology measures were standardized before PCA; the first principal component (PC1) was initially derived from 573 eligible PsychAD MSSM donors aged  $\geq 60$  years with complete measurements for all eight neuropathological variables. Eligible donors excluded those with schizophrenia, bipolar disorder, Parkinson's disease, dementia with Lewy bodies, frontotemporal dementia, vascular disease, or brain tumor. The direction of PC1 was oriented such that higher values represented greater AD pathological burden. PC1 explained 72.8% of the variance across the neuropathological measures and was used as the donor-level pathology score (Dx\_donor). For an additional 66 eligible donors lacking one or more of the eight neuropathological measurements, Dx\_donor was estimated by linear regression. For each donor, we identified the subset of neuropathological variables available for that donor and fitted a linear regression model predicting Dx\_donor using the corresponding variables among donors with observed Dx\_donor. The fitted model was then used to estimate Dx\_donor for the donor with incomplete measurements. A model using CERAD and Braak stage alone explained 91% of the variance in Dx\_donor (adjusted  $R^2 = 0.91$ ), indicating that these core neuropathological measures captured most of the variation represented by the composite pathology score. Together, Dx\_donor was available for 639 eligible PsychAD MSSM donors, including the 247 donors profiled by single-nucleus multiome in the current study.

To define anchor groups along the AD pathological continuum, donors were classified using CERAD and Braak criteria. Controls were defined as CERAD = 1 and Braak < 3. Early-stage donors were defined as CERAD > 1 and Braak > 2 with a combined CERAD and Braak score of 5–9, excluding donors who met the late-stage criteria. Late-stage donors were defined as CERAD > 3 and Braak > 4. Donors that did not meet these anchor definitions were initially left unassigned. We then fitted a cumulative link model with a probit link using the ordered anchor groups (control < early < late) as the outcome and standardized Dx\_donor as the sole predictor. The fitted model was applied to all donors to estimate probabilities of membership in each disease stage ( $P_{\text{control}}$ ,  $P_{\text{early}}$ , and  $P_{\text{late}}$ ), providing a continuous probabilistic representation of position along the AD pathological continuum.

## Cognitive resilience

We used a continuous cognitive-resilience phenotype defined in previous PsychAD work<sup>8,20</sup> using a residual-based approach<sup>37</sup>, in which better-than-expected cognitive status for a given level of neuropathological burden indicates greater resilience. The continuous Cognitive\_Resilience score was defined as the residual from a linear regression of Clinical Dementia Rating sum-of-boxes (CDR-SB) score on age at death, sex, postmortem interval (PMI), CERAD score, and Braak stage. Residuals were oriented such that higher values indicated better-than-expected cognitive status given the model covariates. Because CERAD and Braak stage were included as predictors, this measure captures variation in cognitive status not explained by these measures of plaque and tangle burden and the other model covariates. Given the limited number of ante-mortem cognitive measures available in the MSSM cohort, this phenotype primarily represents pathology-adjusted variation in CDR-SB.

In parallel, we used a previously defined categorical cognitive-resilience phenotype<sup>8</sup>. Donors were classified as AD\_resilient according to a graded combination of CERAD neuritic plaque burden and cognitive status: CERAD = 2 with CDR-SB = 0, CERAD = 3 with CDR-SB  $\leq 2$ , or CERAD = 4 with CDR-SB  $\leq 3$ . This categorical definition was used as a complementary resilience phenotype (**Extended Data. Fig. 8d**).

## Isolation and fluorescence-activated nuclear sorting (FANS)

All buffers were supplemented with RNase inhibitors (Takara). 25mg of frozen postmortem human brain tissue was homogenized in cold lysis buffer (0.32 M Sucrose, 5 mM  $\text{CaCl}_2$ , 3 mM Magnesium acetate, 0.1 mM EDTA, 10mM Tris-HCl, pH 8, 1 mM DTT, 0.1% Triton X-100) and filtered through a 40  $\mu\text{m}$  cell strainer. The flow-through

was overlaid with sucrose solution (1.8 M Sucrose, 3 mM Magnesium acetate, 1 mM DTT, 10 mM Tris-HCl, pH 8) and centrifuged at 107,000 ×g for 1 hour at 4°C. Pellets were resuspended in 250 µl PBS supplemented with 0.5% bovine serum albumin (BSA). Before FANS, 7AAD (Invitrogen) was added according to the manufacturer's instructions. 7AAD-positive nuclei were sorted into tubes pre-coated with 5% BSA using a FACSAria flow cytometer (BD Biosciences).

### **snMultiome ATAC and Gene Expression library preparation**

Following FANS, nuclei were subjected to 2 washes in 200 µl wash buffer (10x Genomics), re-suspended in 30 µl nuclei buffer (10x Genomics), and quantified (Countess II, Life Technologies). 10,000 nuclei from each sample were subjected to the Chromium Next GEM Single Cell Multiome ATAC and Gene Expression protocol (10x Genomics), according to the manufacturer's instructions. ATAC-seq and gene expression libraries were quantified using the KAPA library quantification kit (KAPA Biosystems) and Qubit (Invitrogen), respectively, and fragment sizes were determined by Tapestation (Agilent). All libraries were sequenced at the New York Genome Center (NYGC) using the NovaSeq platform (Illumina).

### **snMultiome data preprocessing and quality control**

snMultiome data preprocessing was performed according to the workflow shown in **Extended Data Fig. 1b**. Raw sequencing data were processed using Cell Ranger ARC v2.0.0 (10x Genomics) with default parameters and the GRCh38 reference genome. Cell Ranger ARC generated aligned reads, filtered barcode-associated gene-expression count matrices, and chromatin-accessibility fragment files. Nucleus-associated barcodes were identified jointly using information from the RNA and ATAC modalities.

To confirm donor identity, genotype concordance was evaluated between the snRNA-seq data and the PsychAD genotype reference panel using QTLtools MBV v1.3<sup>57</sup>. Fifteen samples showed discordant genotype assignments and were reassigned to the genotype-matched donor before downstream analysis (**Extended Data Fig. 1c**).

For the gene-expression modality, quality control was performed using Scanpy v1.9.3<sup>58</sup>. Nuclei were retained if they contained <100,000 unique molecular identifiers (UMIs), >500 detected genes, and <1% mitochondrial reads. Ambient RNA contamination was corrected using SoupX v1.6.2<sup>59</sup>. For the chromatin-accessibility modality, quality control was performed using ArchR v1.0.2<sup>28</sup>. Nuclei identified by Cell Ranger ARC were retained if they had transcription start site (TSS) enrichment >3 and >1,000 unique fragments. Additional major-class-specific quality control was applied after initial cell-class annotation as described below.

Potential doublets were evaluated independently using the RNA and ATAC modalities. RNA doublets were identified using scDbtFinder v1.14.0<sup>60</sup>. For the ATAC modality, scDbtFinder was combined with AMULET v1.0<sup>61</sup> following the scDbtFinder scATAC-seq workflow. The scDbtFinder doublet score was transformed as 1 – scDbtFinder score and combined with the AMULET P value using Fisher's method<sup>62</sup>. Values below 0.001 were set to 0.001 before combination to limit the influence of extremely small values. FDR-adjusted Combined P values <0.05 were classified as ATAC doublets. Nuclei identified as doublets by either RNA-based scDbtFinder or the combined ATAC-based analysis were excluded from downstream analyses. Concordance between RNA- and ATAC-based doublet detection is shown in **Extended Data Fig. 1d**, and sample-level QC metrics for retained nuclei are shown in **Extended Data Fig. 1e**.

### **snMultiome dimensionality reduction and initial major-class annotation**

Cell taxonomy was constructed using an iterative multimodal workflow (**Extended Data Fig. 2a**). For the RNA modality, gene-expression count matrices were library-size normalized and log-transformed, and taxonomy construction was restricted to autosomal, non-mitochondrial, non-ribosomal protein-coding genes. For initial major-class annotation, 5,000 highly variable genes were selected based on mean–dispersion relationships

using default parameters<sup>63</sup>, whereas 3,000 highly variable genes were used for subsequent higher-resolution analyses performed separately within each major cell class. PCA was used for dimensionality reduction, followed by Harmony<sup>64</sup> correction for sample-associated variation, k-nearest-neighbor (KNN) graph construction, and Leiden clustering<sup>65</sup> using Scanpy. Each tissue sample was treated as an individual batch covariate for Harmony correction. Harmony-adjusted embeddings were used for taxonomy construction, whereas uncorrected molecular measurements were retained for downstream association analyses.

For the ATAC modality, open-chromatin regions were identified separately within each RNA-defined major cell class using ArchR with MACS2 v2.2.7.1<sup>66</sup>. Peak accessibility was quantified using paired-end insertion counting<sup>67</sup>, binarized, and normalized using term frequency–inverse document frequency (TF-IDF) transformation. Dimensionality reduction was performed using a modified iterative latent semantic indexing (LSI) workflow in ArchR. The first LSI component was excluded because it was primarily associated with technical and quality-control variation. An initial LSI embedding was generated using highly accessible peaks. Harmony correction was applied to the initial LSI space before clustering. Harmony-adjusted low-resolution clusters were used to calculate cluster-level average accessibility and identify the 150,000 most variable features for subsequent LSI iterations. After the final LSI iteration, Harmony correction was again applied to the LSI embedding to account for sample-associated variation. KNN graph construction and Leiden clustering were then performed on the Harmony-adjusted final LSI embedding using Scanpy.

We subsequently performed a joint multimodal embedding by integrating RNA PCs and ATAC-seq LSIs using the weighted nearest-neighbor<sup>27</sup> (WNN) implementation in Muon v0.1.3<sup>68</sup>. The first 30 RNA principal components and ATAC LSI components 2–30 were used. This cross-class WNN representation was used for multimodal visualization (**Fig. 1**) and assessment of the overall cell taxonomy.

Initial annotations identified major cell classes based on canonical marker-gene expression, including excitatory neurons (EN), inhibitory neurons (IN), astrocytes (Astro), oligodendrocytes (Oligo), oligodendrocyte precursor cells (OPC), immune cells (Micro\_PVM and adaptive immune cells), and vascular cells (endothelial and mural cells).

## Cell-type annotation and adaptive quality control

Major-class annotations were evaluated using both RNA and ATAC information. RNA major-class labels were assigned based on canonical marker-gene expression, whereas ATAC major-class assignments were supported by gene activity scores for canonical marker genes. Concordance between RNA- and ATAC-derived major-class assignments from the same nucleus was then used as an additional quality-control criterion. Nuclei with discordant RNA- and ATAC-derived major-class annotations were excluded from downstream analyses (**Extended Data Fig. 2b**).

To further remove low-quality nuclei while accounting for major-class-specific differences in QC distributions, adaptive QC was performed independently within each major cell class. For each QC metric, nuclei were ranked according to the metric and divided into bins of 100 nuclei. Within each bin, we calculated the fraction of nuclei with concordant RNA- and ATAC-derived major-class annotations. A concordance threshold was defined as the mean minus three standard deviations of concordance rates across the central 60% of bins. Concordance across the ranked QC distribution was then smoothed using LOWESS regression (fraction = 0.05), and the first point at which the smoothed concordance exceeded this threshold was used to define the corresponding QC cutoff (**Extended Data Fig. 2c**). Adaptive thresholds were evaluated for RNA library complexity and total UMI counts and for ATAC TSS enrichment, fragment depth, and blacklist ratio. Metrics were sign-inverted where necessary to evaluate technical outliers at the opposite end of the distribution. RNA library complexity was defined within each major cell class as the residual of log<sub>10</sub>-transformed total RNA UMI counts after linear regression against duplicate fraction, with the regression fitted among nuclei with concordant RNA and ATAC major-class annotations. Final major-class-specific adaptive QC thresholds are provided in **Supplementary Table 2**.

## Taxonomy refinement and multimodal label transfer

Following adaptive QC, nuclei from each major cell class were subsetted and reclustered independently. Within each class, 3,000 highly variable genes were recalculated for the RNA modality, and ATAC embeddings were recomputed using class-specific peak sets. RNA PCA and ATAC LSI embeddings were then integrated within each major cell class using the WNN implementation in Muon. These within-class WNN analyses were used to define subclass-level clusters, resulting in 27 subclasses used for the main atlas and downstream analyses.

For nuclei in which only one modality passed quality control, subclass labels were transferred from high-quality paired multiome nuclei using within-class nearest-neighbor mapping based on the available modality. To evaluate cross-modality label-transfer consistency, 20% of paired multiome nuclei within each major cell class were held out, and subclass labels were predicted using only the ATAC modality. Agreement with paired-modality annotations was quantified using adjusted Rand index (ARI) and classification accuracy. Subclass assignment showed high concordance, with an ARI of  $0.957 \pm 3.30 \times 10^{-3}$  and accuracy of  $0.981 \pm 5.91 \times 10^{-3}$  (mean  $\pm$  standard error, se; **Supplementary Table 3**).

Because chromatin accessibility provided lower resolution at the finest taxonomic level<sup>12</sup>, subtype annotation was performed primarily using the RNA modality, resulting in 91 subtypes (**Supplementary Table 4**). ATAC-based subtype labels were assigned by nearest-neighbor transfer from paired multiome nuclei. Subtype-level cross-modality concordance was evaluated using the same held-out procedure, with an ARI of  $0.5 \pm 3 \times 10^{-3}$  and accuracy of  $0.79 \pm 6 \times 10^{-3}$  (**Supplementary Table 3**). Astrocyte and Micro\_PVM subtypes were further refined and validated using marker genes, module scores derived from external reference datasets<sup>11,31,33,34</sup>, and spatial or cortical-layer mapping.

During astrocyte subclustering, a rare population of ependymal-like nuclei expressing cilia- and ependymal-associated genes was identified (**Extended Data Fig. 4**). These nuclei were retained in the final annotation metadata but excluded from main atlas visualizations and downstream disease-state analyses because they were rare, sample-restricted, enriched in PHG/BA36 dissections, and likely reflected inclusion of adjacent tissue during dissection (Supplementary Notes).

## Taxonomy validation

Taxonomy annotations were evaluated using multiple complementary approaches. Canonical marker-gene expression, gene activity scores, and chromVAR motif activities were examined at the major-class and subclass levels (**Fig. 1e**; **Extended Data Fig. 2d–e**). Module scores based on the top 50 marker genes for each subclass from a published brain reference<sup>31</sup> were used to assess subclass specificity. RNA marker genes were identified within each major cell class using Wilcoxon rank-sum tests implemented in Scanpy, with FDR-adjusted  $P < 0.05$  and  $\log_2$  fold change  $> 1$ .

The taxonomy was further compared with the parent PsychAD snRNA-seq atlas<sup>8</sup> and independent ROSMAP<sup>13,14</sup> and SEA-AD<sup>9</sup> reference datasets using scANVI-based label transfer<sup>69,70</sup> from scvi-tools v1.0.2 within major cell classes (**Extended Data Fig. 3**).

## Peak calling and gene activity scores

Final open-chromatin regions were identified using ArchR with MACS2. To increase sensitivity for regulatory elements associated with distinct cellular states while maintaining sufficient coverage for peak calling, peaks were called separately for each annotated cell subtype. For each subtype, 2–5 independent pseudobulk replicates were generated specifically for peak calling, with approximately 500–3,000 nuclei contributing to each pseudobulk depending on subtype abundance. Only peaks reproducibly detected across pseudobulk replicates were retained.

Reproducible subtype-level peak sets were merged to construct consensus open-chromatin region sets across the complete taxonomy and separately within each major cell class. Overlapping candidate peaks were resolved using the iterative peak-merging/reduction procedure implemented in ArchR, producing a non-overlapping consensus atlas of 644,407 open-chromatin regions (**Extended Data Fig. 2f**). Major-cell-class-specific consensus peak sets were used for downstream within-class ATAC analyses, whereas the full consensus peak atlas was used for analyses spanning cell classes.

Gene activity scores were calculated from chromatin accessibility using ArchR's distance-weighted gene-score model<sup>28</sup>, which integrates accessibility surrounding each gene with greater weight assigned to proximal regulatory regions.

## Reference-based Xenium cell-type annotation with scANVI

Cell-type annotations were transferred to the Xenium dataset<sup>8</sup> using scANVI, with the snMultiome gene-expression dataset generated in the current study as the reference. Xenium nuclei had previously been assigned major cell-class labels as described in our previous study<sup>8</sup>; these labels were used to restrict subsequent higher-resolution label transfer to the corresponding major cell class. The broad spatial and cortical laminar domains were identified from local cell-type composition using an extension of our previously described approach<sup>8</sup>, additionally including layer 1-specific IN\_ADARB2 to annotate the layer 1 spatial domain (Supplementary Notes).

For each major cell class, reference snMultiome nuclei and Xenium query nuclei were subsetting to genes represented in the Xenium panel. Models were trained using scvi-tools for dimensionality reduction and reference-based label transfer. Models used five hidden layers and 30 latent dimensions. The scANVI reference model was trained for 20 epochs with a minimum sampling of 100 nuclei per annotation label per epoch, followed by training of the query-transfer model for 100 epochs. The resulting model was applied to the Xenium query data to assign subclass and subtype labels based on the snMultiome reference taxonomy. Label transfer was performed hierarchically. Major cell-class labels were first used to partition the Xenium dataset, after which subclass and subtype prediction was performed independently within each corresponding major cell class.

To independently relate the snMultiome taxonomy to cortical laminar identity, we used a previously published SMART-seq dataset generated from microdissected cortical layers<sup>30</sup> as a reference for scANVI-based label transfer. Layer annotations from the SMART-seq reference were transferred to snMultiome nuclei, and the resulting predicted layer assignments were used to quantify the laminar composition of each neuronal subclass and subtype (**Extended Data Figs. 4g, 8c**). This analysis also supported the enrichment of IN\_ADARB2 in layer 1, consistent with its superficial-layer localization in our previous PsychAD spatial study.

## Motif signature selection from metacell chromatin accessibility profiles

To reduce sparsity in single-nucleus chromatin accessibility measurements, nuclei were aggregated into locally homogeneous metacells within each major cell class. Individual peak-accessibility profiles were first binarized. For each major cell class, a target metacell accessibility depth corresponding to approximately 20 average nuclei was defined as 20 times the mean number of accessible peaks per nucleus. Seed nuclei were sampled within each subclass, and neighboring nuclei were added according to proximity in the corresponding low-dimensional taxonomy embedding until the target accessibility depth was reached. Metacells were required to have  $\geq 95\%$  subclass purity. The number of metacells generated per subclass was proportional to its total accessibility depth and capped at 1,000. For the Micro\_PVM major class, subtype labels were used during metacell construction and motif-model training to distinguish PVM from microglial subtypes; only the PVM subclass-associated motif signature was used and reported in the present study.

Motif activity was quantified using chromVAR v1.28.0<sup>29</sup> deviation Z-scores and cisBP<sup>71</sup> motif annotations. Peaks with aggregate accessibility  $\leq 10$  across metacells were excluded before chromVAR analysis. GC-matched background peaks were generated using chromVAR, and the resulting motif deviation Z-scores were used

directly for motif selection and prediction models. To reduce redundancy among highly similar motifs, motifs were grouped according to motif-clustering results generated using TOBIAS<sup>72</sup>. Within each motif cluster, subclass-discriminatory activity was quantified as the mean absolute one-versus-rest Welch t-statistic across subclasses, and the motif with the largest value was retained as the representative motif.

Motif selection was performed independently within each major cell class. Metacells were divided into stratified training and testing sets while preserving subclass proportions. On the training set, multinomial LASSO models were fitted using glmnet v4.1-8<sup>73</sup>, with subclass identity as the response and chromVAR deviation Z-scores as predictors. The regularization parameter was selected by five-fold cross-validation using the one-standard-error rule. Model fitting was repeated across 100 stratified bootstrap resamples of the training metacells, and selection stability was evaluated separately for each subclass. Motifs with nonzero coefficients in at least 60% of bootstrap iterations were retained as stable candidates. Because stable motifs could represent related TF families or correlated motif activities, compact subclass-specific signatures were further selected using minimum redundancy maximum relevance (mRMR) analysis<sup>74</sup> on the training set. Up to three nonredundant motifs were retained for each subclass as the final motif signature.

To assess internal motif-signature performance, one-versus-rest logistic models were fitted using the selected motifs and evaluated on a stratified 20% internal test set of metacells. Prediction performance was summarized using area under the receiver operating characteristic curve (AUROC) and area under the precision–recall curve (AUPRC; **Extended Data Fig. 6**). Because subclass frequencies differed across major cell classes and validation datasets, AUPRC was interpreted relative to the prevalence-based baseline.

External validation was performed using independent multiome and snATAC-seq datasets, and was restricted to cortical regions. For the Luquez et al. 2026 multiome cohort<sup>13</sup>, external chromatin fragments were quantified over discovery-defined OCRs, and metacells were generated using harmonized subclass labels assigned by scANVI-based label transfer (**Extended Data Fig. 3**). For the Liu et al. 2025 snATAC-seq dataset<sup>12</sup>, validation was restricted to neuronal populations. External peaks were retained if they overlapped the corresponding major-class-specific OCRs or were accessible in >1.5% of nuclei within the corresponding major class. chromVAR deviation Z-scores were calculated using the same motif annotations as in the discovery analysis. Discovery-trained motif models were then directly applied to the validation cohorts, and performance was evaluated using harmonized external subclass labels (**Extended Data Fig. 6**). chromVAR activity profiles of the selected signature motifs were also compared across subclasses between discovery and validation cohorts (**Extended Data Fig. 6**).

## Cellular disease-state score training

Cellular disease-state models were trained independently for six major cell classes with sufficient representation: Astro, Oligo, OPC, Micro\_PVM, EN, and IN. Input features consisted of total-count-normalized and log-transformed snRNA-seq expression values for protein-coding genes, excluding mitochondrial and ribosomal genes. To reduce differences in training-cell composition across donor pathology groups, cells were stratified by subtype, brain region, and donor-level pathology category. Within the eligible subtype–region strata, equal numbers of cells were sampled from control-, early-, and late-stage donors. Donor-level pathology categories were defined as described above.

Dx\_cell was trained using a two-stage framework (**Fig. 2b**). In the first stage, a LightGBM v3.3.5<sup>75</sup> gradient-boosted decision-tree regression model was trained to predict the continuous donor-level pathology score, Dx\_donor, from single-nucleus gene expression. Initial cell-level predictions used for anchor-cell selection were generated by cell-level cross-fitting and linearly rescaled to the mean and standard deviation of Dx\_donor. A cell was defined as a donor-concordant anchor cell when the absolute difference between its rescaled initial prediction and Dx\_donor was less than one standard deviation of Dx\_donor.

Anchor cells were subsequently used to train final regression and classification models. The regression model generated the continuous cellular disease-state score,  $Dx\_cell$ . The classification model predicted probabilities of control-like, early-like, and late-like cellular states, using the donor-level pathology category of each anchor cell as the training label.

Final  $Dx\_cell$  values and disease-state classifications were generated using 20-fold cell-level cross-fitting. Within each major cell class, cells were partitioned into 20 folds, and predictions for each fold were generated from models trained without cells in that fold. Final continuous  $Dx\_cell$  values were linearly rescaled to the mean and standard deviation of  $Dx\_donor$  before downstream analyses involving  $Dx\_cell$ – $Dx\_donor$  deviation. Raw, unscaled  $Dx\_cell$  values were used only to compare within-sample score dispersion between the initial and anchor-refined models (**Extended Data Fig. 7b**). A probability-weighted classification score,  $Dx\_cat\_score$ , was calculated as  $0 \times P_{control-like} + 1 \times P_{early-like} + 2 \times P_{late-like}$ . Concordance between continuous regression-derived  $Dx\_cell$  and classification-derived  $Dx\_cat\_score$  was evaluated within each sample and major cell class using Spearman correlation (**Extended Data Fig. 7c**).

LightGBM models were trained for up to 1,000 boosting rounds, with early stopping after 50 rounds without improvement. Regression models used a Huber objective ( $\alpha = 0.9$ ) and mean absolute error as the evaluation metric. Classification models used a multiclass objective with multiclass log loss and three output classes. Shared hyperparameters were `num_leaves = 31`, `learning_rate = 0.05`, and `feature_fraction = 0.8`. Regression models additionally used `bagging_fraction = 0.8` and `bagging_freq = 5`. The same hyperparameter settings were used across major cell classes.

## Evaluation and validation of $Dx\_cell$

We first evaluated whether  $Dx\_cell$  was concordant with established cell-class-specific AD-associated transcriptional signatures. For each major cell class, the top 100 upregulated and top 100 downregulated genes from the corresponding PsychAD AD-versus-control differential-expression analysis<sup>8</sup> were used to construct an RNA disease-burden score. For each cell, the score was calculated as the difference between the mean standardized expression of the upregulated genes and the mean standardized expression of the downregulated genes.

Cells were ranked according to the absolute discordance between cellular and donor-level disease estimates,  $|Dx\_cell - Dx\_donor|$ , after rescaling  $Dx\_cell$  to the same scale as  $Dx\_donor$ , and divided into 50 ordered, non-overlapping equal bins. Within each bin, Spearman correlations were calculated between the RNA disease-burden score and either  $Dx\_cell$  or  $Dx\_donor$ . Analyses were performed independently for each major cell class (**Extended Data Fig. 7d**). Because the RNA disease-burden score was derived from an external PsychAD disease signature but evaluated in the same molecular modality as  $Dx\_cell$ , this analysis was used as a transcriptional concordance benchmark rather than an independent validation of the model.

We next evaluated cross-modality concordance using ATAC-derived gene activity. Cell-class-specific AD-associated gene-activity signatures were derived from the current multiome dataset, and ATAC disease-burden scores were calculated as the difference between the mean standardized gene activity of the top 100 AD-increased genes and the mean standardized gene activity of the top 100 AD-decreased genes. Cells were ranked and binned according to  $|Dx\_cell - Dx\_donor|$  as described above. Within each bin, Spearman correlations were calculated between the ATAC disease-burden score and either  $Dx\_cell$  or  $Dx\_donor$  (**Fig. 2c**).

To assess whether  $Dx\_cell$ -defined states captured molecular heterogeneity within individual tissue samples, we compared marker-gene expression and ATAC-derived gene activity between  $Dx\_cell$ -defined control-like and late-like cells from the same sample. Analyses were restricted to sample–cell-class combinations containing at least 20 control-like and 20 late-like cells. For each marker and molecular modality, values were averaged separately within each sample and  $Dx\_cell$ -defined state to generate paired pseudobulk estimates. Control-like

and late-like states were compared within each donor-level pathology category. Multiple-testing correction was performed across the 72 same-sample marker comparisons shown in **Fig. 2d** and **Extended Data Fig. 7e**.

For sample-level analyses, continuous Dx\_cell scores were averaged separately for each sample and major cell class. The frozen major-cell-class-specific models were applied without retraining to PsychAD<sup>8</sup>, SEA-AD<sup>9</sup>, and Mathys et al. 2023 ROSMAP snRNA-seq datasets<sup>14</sup>. External gene-expression matrices were processed using the same normalization procedure as the discovery dataset, consisting of total-count normalization followed by log transformation. Cells were assigned to the corresponding major cell classes using the final filtered annotations provided by each study. Expression matrices were harmonized to the discovery-model feature space and feature order. For model features absent from an external dataset, the mean normalized expression of that gene in the corresponding discovery reference was used. PsychAD donors were analyzed separately according to whether they overlapped with the multiome discovery cohort. Cohort-level counts for the indicated regions and six modeled major cell classes are summarized in **Supplementary Table 10**; the effective sample sizes for individual analyses may differ after phenotype and covariate filtering. For each dataset and major cell class, sample-level mean Dx\_cell was compared with the available donor-level disease measure using Spearman correlation (**Fig. 2e**). Dx\_donor was used for the discovery multiome and PsychAD datasets, Continuous Pseudo-progression Score (CPS) for SEA-AD, and global AD pathology for Mathys et al. 2023 ROSMAP.

To further evaluate whether Dx\_cell captured within-sample disease-associated transcriptional heterogeneity, cellular Dx\_cell values were compared with previously developed PsychAD transcriptomic prediction scores for Braak pathology and dementia<sup>8</sup>. In contrast to the current Dx\_cell framework, which uses cell-level cross-fitting and therefore permits cells from the same donor to occur in both training and prediction folds, the PsychAD Braak and dementia models were trained using donor-level cross-validation, such that all cells from a given donor were excluded from model training when generating predictions for that donor. Thus, concordance with these independently trained donor-held-out models provides a complementary model-based check that Dx\_cell captures disease-associated transcriptional variation beyond same-cell reuse or direct fitting to the evaluated cells. Analyses were performed independently within each sample, and major cell class (**Fig. 2f**). Partial Spearman correlations were calculated between Dx\_cell and each PsychAD prediction score while conditioning on the RNA-seq covariates used in the PsychAD analysis, including number of detected genes (n\_genes), mitochondrial-gene fraction (mito\_genes), ribosomal-gene fraction (ribo\_genes), and the mitochondrial ribosomal gene metric (mito\_ribo). Samples containing fewer than 10 cells for the corresponding major cell class were excluded. Correlation coefficients were summarized across samples within each major cell class and PsychAD donor-overlap group.

### Definition of pathology-adjusted cellular disease-state deviation (Dx\_dev)

To quantify sources of variation in cellular disease state across tissue samples, continuous Dx\_cell scores were averaged separately for each sample and major cell class. Variance components of sample-level mean Dx\_cell were estimated independently within each major cell class using the variancePartition v1.36.3 framework<sup>76</sup>. The model included donor-level pathology (Dx\_donor), donor identity, brain region, age at death, sex, postmortem interval (PMI), APOE genotype, AD polygenic risk score excluding the APOE locus, and RNA-seq quality covariates, including n\_genes, ribo\_genes, and mito\_genes. Categorical factors were modeled as random effects for variance decomposition, whereas continuous covariates were included as scaled continuous terms.

The variance-partitioning model was specified as:

$$\text{mean\_Dx\_cell} \sim \text{scale}(\text{Age}) + (1|\text{Sex}) + (1|\text{Donor}) + \text{scale}(\text{Dx\_donor}) + \text{scale}(\text{PMI}) + (1|\text{BrainRegion}) + (1|\text{APOE}) + \text{scale}(\text{PRS\_AD\_noAPOE}) + \text{scale}(\text{n\_genes}) + \text{scale}(\text{ribo\_genes}) + \text{scale}(\text{mito\_genes})$$

The proportion of variance attributable to each modeled variable was estimated separately for each major cell class.

We next defined `Dx_dev` as the pathology-adjusted cellular disease-state deviation. Within each major cell class, sample-level mean `Dx_cell` was regressed on `Dx_donor`, age at death, sex, PMI, and sample-level RNA-seq covariates, including `n_genes`, `ribo_genes`, and `mito_genes`. Sex and the other covariates were included as fixed effects in this residualization model. The resulting residual was retained as `Dx_dev`. Positive `Dx_dev` therefore indicates a higher mean `Dx_cell` than expected given measured pathology and modeled covariates, whereas negative `Dx_dev` indicates a lower mean `Dx_cell` than expected. Donor identity and brain region were intentionally not removed when defining `Dx_dev`, allowing donor- and region-associated biological variation to remain in the phenotype.

Cohort-specific `Dx_dev` values were derived using the same framework after projection of the frozen `Dx_cell` models. Within each external dataset and major cell class, `Dx_cell` was first averaged within each tissue sample. Sample-level mean `Dx_cell` was then regressed on the cohort-specific measure of AD pathology together with age, sex, PMI, and the same available RNA-seq quality covariates used in the discovery `Dx_dev` model. The resulting residual was defined as `Dx_dev`. PsychAD analyses used `Dx_donor` as the pathology covariate, SEA-AD analyses used CPS, Mathys et al. 2023 ROSMAP analyses used global AD pathology, and Luquez et al. 2026 analyses used CERAD and Braak measures. `Dx_dev` was estimated separately within each cohort because the available neuropathological and technical covariates differed among datasets.

Analogous pathology-adjusted deviation measures were derived from the previously developed donor-held-out PsychAD transcriptomic prediction scores for Braak pathology and dementia by regressing out `Dx_donor`, age, sex, PMI, and RNA-seq covariates.

For analyses integrating `Dx_dev` across multiple cell classes, `Dx_dev` was standardized separately within each major cell class and then averaged across available cell classes within each tissue sample. In addition to the overall cross-cell-class mean, glial-restricted and Astro/OPC/Oligo-restricted mean `Dx_dev` scores were calculated for prespecified downstream analyses. Samples from different brain regions of the same donor were retained as separate observations and were not averaged across regions.

## **Cell-composition analysis using `crumblr`**

Cell-composition associations were tested using `crumblr` v0.99.19<sup>77</sup>. `Crumblr` applies a centered log-ratio transformation to cell counts and estimates the sampling variance of transformed proportions using an asymptotic approximation to the Dirichlet-multinomial distribution. These sampling variances are incorporated as precision weights in downstream linear or mixed-effects models to account for heteroskedasticity associated with differences in cell recovery across samples.

`Dx_donor`, sample-level mean `Dx_cell`, and `Dx_dev` were standardized before association testing. Associations between cell composition and `Dx_donor` were evaluated across the subclass taxonomy. Models combining multiple brain regions adjusted for age, sex, PMI, and brain region as fixed effects and included donor identity as a random intercept to account for repeated tissue samples from the same donor. Region-specific analyses were performed separately without brain-region and donor identity covariates.

Associations of cell composition with sample-level mean `Dx_cell` and `Dx_dev` were evaluated using analogous models with additional adjustment for `Dx_donor`, thereby testing associations conditional on measured pathological burden. Composition changes were evaluated across the full taxonomy and, for prespecified AD-associated populations, within their corresponding parent classes, including `EN_L2_3_IT` within `EN`, `IN_SST` within `IN`, and `DAM` within `Micro_PVM`. For subclass-level composition analyses, we used glia-derived `Dx_cell` or `Dx_dev` scores. The glia-derived `Dx_cell` score was calculated by standardizing sample-level mean `Dx_cell` separately within Astro, Oligo, OPC, and `Micro_PVM`, followed by averaging the standardized values across these four classes within each sample. Glia-derived `Dx_dev` scores were calculated analogously. For analyses

of Micro\_PVM composition, the corresponding scores were calculated using only Astro, Oligo, and OPC, excluding Micro\_PVM to avoid direct overlap between the predictor and the composition outcome.

Replication analyses were performed in non-overlapping PsychAD donors and the SEA-AD cohort using analogous covariate models. Because the SEA-AD sampling strategy used fluorescence-activated cell sorting that imposed study-specific neuron/glia sampling proportions, broad subtype-level composition analyses in SEA-AD were restricted to neuronal populations to minimize effects of artificial variation in the neuron-to-glia ratio. Prespecified within-parent-class analyses, including DAM as a fraction of Micro\_PVM, were analyzed separately within their parent classes.

### **Association of Dx\_dev with cognitive phenotypes and cognitive resilience**

In the multiome discovery cohort, associations with cognitive resilience were tested separately within each major cell class using linear mixed-effects models. Standardized Dx\_dev was modeled as the outcome and standardized cognitive phenotype as the predictor. Thus, each regression coefficient estimates the change in standardized Dx\_dev per standard-deviation increase in the cognitive phenotype, conditional on pathology and the specified covariates. Models adjusted for Dx\_donor, age at death, sex, PMI, brain region, and APOE genotype as fixed effects and included donor identity as a random intercept to account for multiple regional tissue samples from the same donor. P values were adjusted using the Benjamini–Hochberg procedure across the six major cell classes separately as implemented in dreamlet for each cognitive phenotype and cohort. Analyses were restricted to donors classified within the early- or late-stage AD groups.

Replication analyses were performed in PsychAD donors not overlapping with the multiome discovery cohort. Scaled Dx\_dev was modeled as the outcome and the continuous Cognitive\_Resilience score as the predictor using linear models adjusted for Dx\_donor, age at death, sex, PMI, and APOE genotype.

As a complementary categorical analysis, previously defined cognitive-resilience status was tested for association with scaled Dx\_dev. The relationship between donor pathology and sample-level mean Dx\_cell was also visualized separately for resilient and non-resilient pathology-positive donors.

In the Mathys et al. 2023 ROSMAP cohort, analogous analyses were performed using available cognitive phenotypes including last-visit Mini-Mental State Examination score<sup>38</sup> (cts\_mmse30\_iv) and longitudinal cognitive slope<sup>39</sup> (cogng\_demog\_slope). Each cognitive phenotype was standardized and tested separately as the predictor of scaled Dx\_dev within each major cell class, using the corresponding cohort-specific pathology and covariates. Effect directions were interpreted according to the orientation of each cognitive measure.

### **Cross-cell-type variance partitioning of gene expression, gene activity, and TF motif activity**

To quantify major sources of molecular and regulatory variation across the atlas, we constructed sample-by-subclass pseudobulk profiles for gene expression, ATAC-derived gene activity, and chromVAR motif activity. This variance-partitioning analysis was the only molecular analysis performed jointly across major cell classes; subsequent differential analyses were performed separately within each major cell class.

Pseudobulk gene-expression and gene-activity profiles were generated using dreamlet v1.4.1<sup>78</sup>, with observations defined by sample and subclass. Pseudobulk matrices were processed using voom-based mean–variance modeling<sup>79</sup> before linear mixed-model analysis. For chromVAR analysis, ATAC peak counts were first aggregated at the sample-by-subclass level. Motif deviation scores were then recomputed directly from these pseudobulk peak-accessibility profiles using chromVAR<sup>29</sup> and cisBP motif annotations<sup>71</sup>, rather than averaging single-cell chromVAR scores.

Variance partitioning was performed separately for each molecular modality using dreamlet/variancePartition. For each feature, the model was:

$Y \sim \text{scale}(\text{Dx\_donor}) + (1|\text{Donor}) + (1|\text{Sex}) + (1|\text{BrainRegion}) + (1|\text{Class}) + (1|\text{Class:Subclass}) + \text{scale}(\text{Age}) + \text{scale}(\text{PMI}) + \text{scale}(\text{technical covariates})$

where Y represents the corresponding sample-by-subclass molecular feature. Donor identity, sex, brain region, major cell class, and subclass nested within major cell class were modeled as random effects for variance decomposition. Subclass labels were unique within each major cell class; for major classes represented by a single subclass, including Astro, Oligo, and OPC, the class and subclass terms were equivalent. Continuous covariates were modeled as scaled fixed effects.

For gene-expression models, technical covariates included n\_genes, ribo\_genes, and mito\_genes, following the PsychAD framework<sup>8</sup>. For ATAC-derived gene activity and chromVAR models, technical covariates included TSS enrichment and ATAC fragment depth (nFragments). Variance fractions were estimated independently for each feature and modality.

To assess concordance of variance structure across modalities, variance fractions for matched genes were compared between RNA gene expression and ATAC-derived gene activity. Pearson and Spearman correlations were calculated across matched genes separately for each modeled variance component (**Extended Data Fig. 5**).

### Conventional pseudobulk differential analysis

Conventional differential analyses were performed separately within each major cell class using dreamlet<sup>78</sup>, which extends the linear mixed-model framework implemented in dream<sup>80</sup> to pseudobulk single-cell data. Gene-expression and ATAC-derived gene-activity measurements were aggregated to the sample level within each major cell class using dreamlet pseudobulk aggregation and analyzed using voom-based precision-weighted linear mixed models.

Associations with donor-level AD pathology were tested using the model:

$Y \sim \text{scale}(\text{Dx\_donor}) + (1|\text{Donor}) + \text{Sex} + \text{BrainRegion} + \text{scale}(\text{Age}) + \text{scale}(\text{PMI}) + \text{scale}(\text{technical covariates})$

where Y represents the pseudobulk molecular feature. Sex and brain region were modeled as fixed effects, whereas donor identity was included as a random intercept to account for multiple tissue samples from the same donor. RNA-derived models included n\_genes, mito\_genes, and ribo\_genes as technical covariates. ATAC-derived models included TSS enrichment and fragment depth.

Differential gene-activity results from these models were subsequently used to construct the ATAC-derived disease-burden signatures used for Dx\_cell evaluation. Gene-expression models were also used to derive donor-level best linear unbiased predictors (BLUPs) for downstream donor-factor analyses.

### Differential analysis of decomposed cellular disease states

To resolve molecular differences among cellular disease states coexisting within the same tissue sample, cells from each sample and major cell class were further partitioned into Dx\_cell-defined control-like, early-like, and late-like states (**Fig. 4b**). Gene expression, ATAC-derived gene activity, and peak accessibility were aggregated separately for each sample × major-cell-class × Dx\_cell-state combination. chromVAR motif deviation scores were additionally calculated from decomposed pseudobulk ATAC peak-count profiles.

For ATAC analyses, only peaks defined within the corresponding major cell class were used. Because partitioning cells into three disease-state groups introduced additional variation in pseudobulk cell number, the log-transformed number of nuclei contributing to each pseudobulk was included as an additional covariate. RNA

models were additionally adjusted for `n_genes`, `mito_genes`, `ribo_genes`, and `mito_ribo`, whereas ATAC-derived models additionally adjusted for TSS enrichment, fragment depth, nucleosome ratio, and blacklist ratio.

Four complementary continuous disease-state measures were evaluated (**Fig. 4a**). `Dx_donor` represented donor-level pathological burden. `Dx_cell` represented the mean cellular disease-state score of each decomposed pseudobulk. `Dx_dev` represented pathology-adjusted sample-level cellular-state deviation. To avoid using the molecular state of the focal cell class to define its own `Dx_dev` predictor in these differential analyses, `Dx_dev` was standardized within the cell class and averaged across the other five major cell classes within the same tissue sample. Thus, the `Dx_dev` predictor represented a leave-one-cell-class-out measure of sample-level cellular disease-state deviation.

`Dx_contrast` was defined to capture within-sample differences among the three coexisting `Dx_cell`-defined states. For each sample and major cell class, the mean `Dx_cell` score was calculated separately for the control-like, early-like, and late-like groups. The unweighted mean of these three group-specific scores was then calculated, and `Dx_contrast` for each decomposed pseudobulk was defined as the difference between its group-specific mean `Dx_cell` and this three-state mean. By construction, `Dx_contrast` therefore represented a within-sample-centered cellular disease-state measure, whereas `Dx_donor` captures between-donor pathology and `Dx_dev` captures pathology-adjusted variation shared across other cell classes.

Differential associations with `Dx_donor`, `Dx_cell`, `Dx_dev`, and `Dx_contrast` were tested separately using marginal dreamlet models with a common covariate structure that included donor identity as a random intercept and age, sex, brain region, PMI, pseudobulk cell number, and modality-specific technical covariates as fixed effects.

To determine the relative contribution of donor pathology, sample-level deviation, and within-sample cellular-state contrast to disease-associated molecular variation, we fitted joint variance-partitioning models separately within each major cell class. Analyses focused on genes or gene-activity features associated with `Dx_donor` in the conventional pseudobulk analysis.

The joint model included:  $Y \sim \text{scale}(\text{Dx\_donor}) + \text{scale}(\text{Dx\_dev}) + \text{scale}(\text{Dx\_contrast}) + (1|\text{Donor}) + (1|\text{Sex}) + (1|\text{BrainRegion}) + \text{scale}(\text{Age}) + \text{scale}(\text{PMI}) + \text{scale}(\log\_nuclei) + \text{scale}(\text{technical covariates})$ , where `log_nuclei` is the log-transformed number of nuclei contributing to each pseudobulk, `Dx_dev` was the leave-one-cell-class-out composite described above, and `Dx_contrast` represented the within-sample-centered disease-state score. For variance decomposition, sex and brain region were modeled as random effects, whereas continuous covariates were included as scaled terms; modality-specific technical covariates were defined as above. Variance fractions attributable to `Dx_donor`, `Dx_dev`, and `Dx_contrast` were estimated jointly for each molecular feature.

These three terms represent distinct levels of disease-associated variation: donor-level pathological burden (`Dx_donor`), pathology-adjusted sample-level cellular deviation shared across cell classes (`Dx_dev`), and within-sample differences among coexisting cellular disease states (`Dx_contrast`).

## Gene-set enrichment analysis

Gene-set enrichment analysis (GSEA) was performed using `fgsea v1.32.2`<sup>81</sup>. Genes were ranked according to the t-statistic from the corresponding dreamlet differential model. Gene Ontology Biological Process (GO-BP) gene sets were obtained using `clusterProfiler` with `org.Hs.eg.db` annotations.

Enrichment analyses were performed separately for each disease-state measure and major cell class. Significant pathways were defined at  $\text{FDR} < 0.05$ . For visualization, the two significant GO-BP terms with the smallest nominal P values were selected for each disease-state measure  $\times$  major-cell-class combination, without separately selecting positive and negative enrichment directions. Redundant GO terms were reduced using

semantic similarity calculated with GOSemSim v2.32.0<sup>82</sup>. For pairs of terms with semantic similarity >0.5, the term with the smaller enrichment P value was retained.

## Comparison with published and external differential effects

Gene-expression differential effects were compared with published AD-associated transcriptional changes from the Mathys et al. 2023 ROSMAP dataset, using global AD pathology as the disease measure, and from the PsychAD using CERAD pathology (**Extended Data Fig. 9e**). Comparisons were performed between matched genes and cell classes using Spearman correlation of logFC.

ATAC differential-accessibility effects were evaluated using two complementary datasets. First, peak-level effects were compared with our previously published FACS-sorted ATAC-seq study of the MSSM cohort<sup>40</sup>. Overlapping accessible regions between the two assays were used for comparison, providing cross-assay concordance within the same donor resource rather than an independent cohort replication.

Second, raw ATAC fragments from the cortical region of Luquez et al. 2026 multiome cohort<sup>13</sup> were reprocessed using ArchR. Fragment counts were quantified directly over the major-cell-class-specific peak sets defined in the present study, thereby placing the external cohort in the same peak feature space. Pseudobulk accessibility profiles were generated, and differential accessibility with CERAD pathology was tested using the same dreamlet framework.

For cross-study peak-level comparisons, analyses were restricted to peaks showing nominal  $P < 0.05$  in at least one of the four discovery analyses (Dx\_donor, Dx\_cell, Dx\_dev, or Dx\_contrast) or in the published datasets. Spearman correlations were then calculated between matched peak-effect estimates across datasets.

## Enhancer-driven gene regulatory network inference

Enhancer-driven gene regulatory networks (eGRNs) were inferred using SCENIC+<sup>83</sup> (v0.1.dev454+g9320635) from the paired snRNA-seq and snATAC-seq dataset. Because direct network inference across the full atlas was computationally intensive, 100,000 nuclei were subsampled with balanced representation across major cell classes for eGRN construction. The inferred regulatory relationships were subsequently evaluated and scored across the full dataset.

Candidate regulatory regions were identified using pycisTopic<sup>84</sup> (v1.0.0). Topic modeling of the snATAC-seq profiles was performed using latent Dirichlet allocation with 16 topics, 100 iterations,  $\alpha = 50$ , and  $\eta = 0.1$ . Candidate regions were defined using both topic-associated regions and differentially accessible regions (DARs). Topic-region distributions were binarized using Otsu thresholding, and the top 3,000 regions associated with each topic were additionally retained. DARs were identified using Wilcoxon rank-sum tests and filtered at log2 fold change >0.5 and FDR <0.05.

Motif enrichment within candidate regulatory regions was performed using pycisTarget (v1.0.2) with the differential enrichment of motifs (DEM) algorithm to identify candidate TF-binding cisromes. SCENIC+ subsequently integrated raw gene-expression counts, imputed chromatin accessibility, and motif-enrichment results to infer TF–region–gene relationships. Region-to-gene relationships were inferred within a maximum search distance of 150 kb, excluding regions within 1 kb of the gene boundary, using gradient-boosting regression<sup>85</sup>. TF–target gene relationships were inferred independently using gradient-boosting regression based on gene-expression profiles.

eRegulons were constructed using the SCENIC+ GSEA-based build\_grn procedure. Region-to-gene relationships were binarized according to gradient-boosting importance using the 85th, 90th, and 95th percentiles and the top 5, 10, and 15 regions per gene, with min\_target\_genes = 10, adj\_pval\_thr = 1, and min\_regions\_per\_gene = 0. SCENIC+ classifies eRegulons according to the direction of the inferred TF–gene

and region–gene relationships. For downstream analyses, we retained only ++ eRegulons, in which both the TF–target gene association and the regulatory-region–target gene association were positive. This resulted in 494 ++ eRegulons with predicted TF–region–gene relationships.

For each retained eRegulon, cell-level enrichment scores for target genes and regulatory regions were calculated across the full paired multiome dataset using AUCell<sup>86</sup>.

### Cell-class-specific refinement of inferred eRegulons

Because the initial SCENIC+ networks were inferred jointly across major cell classes, we further refined TF–target relationships within each major cell class before downstream disease-state analyses. Gene-expression profiles were aggregated to sample-level pseudobulks within each major cell class using the conventional dreamlet framework described above. logCPM expression values were adjusted for the RNA-seq technical covariates used throughout the study, including number of detected genes, mitochondrial-gene fraction, and ribosomal-gene fraction.

For each predicted TF–target pair, we tested the correlation between covariate-adjusted TF and target-gene expression across sample-level pseudobulks within the corresponding major cell class. TF–target relationships were retained when the correlation was positive and statistically significant after multiple-testing correction (Spearman correlation >0 and FDR <0.05). eRegulons containing fewer than 30 retained target genes after this cell-class-specific filtering were excluded from subsequent analyses.

This refinement was used to focus the atlas-derived SCENIC+ networks on TF–target relationships showing coordinated expression within the cell class being analyzed (**Supplementary Table 8**).

### Prioritization of Dx\_cell-associated TFs and eRegulons

Candidate regulatory programs associated with cellular disease state were identified by integrating TF expression, chromatin motif activity, and the cell-class-refined eRegulons. Within each major cell class, TF expression and chromVAR motif deviation scores were tested for association with Dx\_cell using the differential-analysis framework described above. TFs were prioritized when TF expression and compatible motif activity changed concordantly with Dx\_cell (**Extended Data Fig. 9f**).

All individual motifs were retained in this analysis. Because closely related TFs can recognize highly similar sequence motifs, motifs were annotated by TF family and motif evidence was interpreted at the motif-family level rather than as direct evidence of TF-specific occupancy. Concordant TF-expression changes were used together with motif-family activity to prioritize candidate regulators among TFs with similar DNA-binding specificity.

To evaluate whether the refined eRegulons captured Dx\_cell-associated transcriptional changes, genes were ranked by the t-statistic from the corresponding Dx\_cell dreamlet differential-expression analysis. Predicted target genes for each eRegulon were then tested using fgsea for enrichment toward the positive or negative ends of the ranked gene list. eRegulons were prioritized according to the strength and direction of target-gene enrichment. For visualization in **Fig. 4e**, the three strongest positively associated and three strongest negatively associated eRegulons within each major cell class were selected.

For selected Micro\_PVM eRegulons, cell-level AUCell scores and corresponding chromVAR motif-family activities were examined across Dx\_cell-defined control-like, early-like, and late-like cellular states (**Extended Data Fig. 9g**).

Functional enrichment of selected Micro\_PVM eRegulons was performed using GO-BP terms. The retained target genes of each eRegulon were tested for enrichment using genes expressed in Micro\_PVM and retained after dreamlet expression filtering as the background gene universe. Multiple-testing correction was performed across tested ontology terms (**Extended Data Fig. 9h**).

## Enhancer–promoter linking

Enhancer–promoter relationships were inferred using the activity-by-contact (ABC) framework<sup>44,45</sup>, following our previously described promoter-isoform-resolved implementation. To account for alternative promoter usage in the human brain, promoter isoforms rather than genes alone were treated as candidate regulatory targets. Promoters were defined as regions extending 500 bp upstream and downstream of annotated transcription start sites (TSSs), and promoter isoforms whose promoter regions overlapped accessible chromatin were included as candidate targets.

ABC scores were calculated separately within each subclass by integrating subclass-specific chromatin accessibility with cell-type-appropriate chromatin-contact maps. Microglia Hi-C maps<sup>46</sup> were used for Micro\_PVM, whereas neuronal or glial Hi-C maps<sup>87</sup> were used for the corresponding neuronal and non-microglial cell classes.

Candidate enhancer–promoter links with ABC score >0.02 were retained. Subclass-level links were subsequently aggregated to the corresponding major cell class, and a link was retained at the major-class level when it exceeded the ABC threshold in at least one constituent subclass.

## Locus-level regulatory integration

For selected loci, Dx\_cell-associated gene expression, differential chromatin accessibility, motif occurrence, ABC enhancer–promoter links, and three-dimensional chromatin contacts were jointly visualized. Hi-C contact maps were analyzed at 5-kb resolution using cooler-normalized matrices<sup>88</sup>, and locus-level tracks were generated using plotgardener v1.12.0<sup>89</sup>.

TF motif occurrences within candidate regulatory elements were identified using motifmatchr v1.28.0 and the same motif-resource framework used for the chromVAR analyses. Because related TFs can recognize similar motifs, motif occurrences were interpreted at the TF-family level.

At the *FGD2* locus in Micro\_PVM, accessible candidate regulatory elements linked to the *FGD2* promoter by ABC analysis were examined for IRF-family motif occurrences. Differential accessibility across Dx\_cell-defined cellular states, ABC enhancer–promoter links, and local Hi-C contact strength were integrated with the Dx\_cell-associated expression change of *FGD2* to evaluate locus-level support for a candidate IRF8–FGD2 regulatory relationship (**Fig. 4g**).

## Donor-associated transcriptional variation across cohorts

To quantify reproducible inter-individual transcriptional variation, we estimated the fraction of gene-expression variance attributable to donor identity using the conventional pseudobulk dreamlet/variancePartition framework described above. Analyses were performed separately within each major cell class. Equivalent analyses were performed in external datasets containing multiple brain-region samples per donor, including Luquez et al.<sup>13</sup> and SEA-AD<sup>9</sup>. For Luquez et al., PFC and STG were included, whereas anterior caudate was excluded because of its subcortical origin and distinct cellular composition. For SEA-AD, PFC and MTG were included. For each major cell class, donor-attributable variance fractions were compared across datasets for genes measured in both cohorts using Spearman correlation.

## Identification of donor-associated transcriptional axes

To identify coordinated transcriptional programs underlying donor-associated expression variation, we extracted gene-wise donor best linear unbiased predictions (BLUPs) from the conventional pseudobulk mixed models<sup>90,91</sup>. BLUPs were estimated separately within each major cell class after adjustment for the biological and technical covariates included in the corresponding differential-expression models.

For the primary discovery analysis, donor-by-gene BLUP matrices were integrated using Multi-Omics Factor Analysis (MOFA2) v0.6.6 Python package<sup>48</sup>, treating each major cell class as a separate data view (**Extended Data Fig. 10a**). Within each cell class, analysis was restricted to protein-coding, non-mitochondrial, and non-ribosomal genes in the top 25% ranked by donor-attributable expression variance. MOFA models were fitted with 10 latent factors.

To reduce sensitivity to individual genomic regions, factor estimation was repeated using five chromosome-block exclusion runs. Autosomes were divided into five approximately balanced groups, and one chromosome group, corresponding to approximately 20% of autosomal genes, was excluded from each run. Thus, each gene contributed to four of the five runs. Latent factors were aligned across runs according to the correlation of their donor scores. Factors were matched one-to-one based on maximum absolute correlation, signs were oriented consistently, and donor scores were standardized before constructing consensus axes. Consensus donor scores were obtained by averaging the aligned standardized scores across the five runs. Gene loadings were aligned using the same factor assignments and sign orientations and averaged across runs in which the corresponding gene was present. Loadings were rescaled to account for standardization of donor scores, preserving the relative contribution of each axis to the original MOFA decomposition. The variance explained by each aligned axis was summarized across chromosome-block runs.

The resulting consensus donor scores and cell-class-specific gene loadings defined the donor-associated transcriptional axes used in subsequent analyses.

### **Robustness and projection of donor-associated transcriptional axes**

Robustness of the transcriptional axes was evaluated using complementary feature-perturbation and projection analyses. First, reproducibility of donor scores was assessed across the five chromosome-block exclusion runs after factor matching and sign alignment. Second, as a sensitivity analysis to feature selection, MOFA was repeated using the top 50% rather than the top 25% of genes ranked by donor-attributable expression variance, and the resulting axis scores were compared with the primary consensus axes.

We additionally projected the consensus axes into sample-level expression profiles. Gene-expression values were first residualized for the specified biological and technical covariates, including disease/pathology, age, sex, brain region, and RNA-seq quality covariates, while retaining donor-associated expression variation. For genes shared with the consensus loading matrices, projected axis scores were calculated by matrix multiplication of the residualized expression profiles and the corresponding consensus gene loadings. Projected scores were subsequently standardized separately for each axis.

Within the current cohort, projection was used to assess concordance of axis scores across brain regions from the same donor. The axes were also projected into the larger PsychAD snRNA-seq cohort and the independent Luquez et al.<sup>13</sup> and SEA-AD<sup>9</sup> datasets using the common set of genes available in each dataset. Cross-region and cross-dataset concordance of projected axis scores was evaluated using Spearman correlation.

The three leading donor-associated transcriptional axes, F1–F3, were selected for downstream analyses based on their contribution to donor-associated expression variation and their reproducibility across chromosome-block exclusion, the expanded top-50% feature sensitivity analysis, brain-region projections, and independent datasets.

### **Functional annotation of donor-associated transcriptional axes**

Biological processes associated with F1–F3 were characterized using their consensus cell-class-specific gene loadings. Within each major cell class, genes were ranked by signed loading for each axis, and GO-BP enrichment was performed using fgsea. Significant enriched pathways were summarized across major cell classes to characterize the biological programs represented by each axis.

## **Association of donor-associated transcriptional axes with Dx\_dev**

Associations between F1–F3 and cellular disease-state measures were first evaluated using Spearman correlations of standardized axis scores with Dx\_dev, Dx\_donor, and donor age.

We next used variancePartition to quantify the contribution of the transcriptional axes to Dx\_dev variation. A base model partitioned Dx\_dev variance among donor identity and the biological and technical covariates defined for the Dx\_dev analysis. An extended model additionally included F1, F2, and F3 simultaneously. Comparison of the two models was used to quantify the variance associated with the transcriptional axes and the corresponding change in variance attributed to donor identity. Because this analysis was performed for variance decomposition, categorical variables were modeled as random effects and continuous variables as scaled continuous terms, consistent with the variancePartition analyses described above.

## **Cross-region association of donor-associated transcriptional axes with Dx\_dev**

To determine whether relationships between the transcriptional axes and Dx\_dev reflected donor-shared programs rather than transcriptional states restricted to the same sampled region, PFC-derived axis scores were tested for association with Dx\_dev measured in other brain regions from the same donors.

F1–F3 scores derived from PFC were tested separately against Dx\_dev measured in PHG and STG, with the corresponding PFC–PFC associations included as same-region references. Both axis scores and Dx\_dev were standardized before analysis. Models were adjusted for the corresponding cohort-specific pathology measure, age, sex, and PMI. Equivalent analyses were performed after projection of the transcriptional axes into the Luquez et al. and SEA-AD datasets. In Luquez et al., PFC-projected factor scores were tested against STG Dx\_dev; in SEA-AD, PFC-projected factor scores were tested against MTG Dx\_dev.

## **Association of donor-associated transcriptional axes with cellular composition**

Associations between donor-associated transcriptional axes and cellular composition were evaluated using crumblr. To focus these analyses on transcriptional programs represented within glial populations, axis scores were projected using glial expression profiles only. Models adjusted for pathology, age, sex, PMI, and brain region as fixed effects and included donor identity as a random effect when multiple regional samples from the same donor were analyzed.

To assess whether axis-associated cellular remodeling resembled pathology-associated composition changes, the vector of cell-population effect estimates associated with each axis was compared with the corresponding pathology-associated effect estimates using Spearman correlation. Pathology was represented by Dx\_donor in the current and PsychAD non-overlapping cohorts and by CPS in SEA-AD.

## **Molecular concordance between donor-associated transcriptional axes and disease-state dimensions**

To relate the donor-associated transcriptional axes to the molecular disease-state dimensions defined above, cell-class-specific consensus gene loadings for F1–F3 were compared with genome-wide marginal differential-expression effect estimates for Dx\_donor, Dx\_cell, Dx\_dev, and Dx\_contrast. Analyses were performed separately within each major cell class using genes shared between the corresponding loading and differential-expression results, and concordance was quantified using Spearman correlation.

Analogous analyses were performed using published AD-associated differential-expression statistics from PsychAD (CERAD) and Mathys et al. 2023 (global pathology). Comparisons were restricted to genes shared between each external dataset and the corresponding cell-class-specific consensus loading matrix, with the orientation of each axis fixed according to the discovery analysis.

## Data availability

Single-nucleus multiomic data, including raw sequencing data, processed data, cell- and sample-level metadata, and complete differential expression and chromatin accessibility results, will be made available through the AD Knowledge Portal prior to publication. Interactive visualization of the snRNA-seq and snATAC-seq data, including cell annotations and Dx\_cell scores, is available through the UCSC Cell Browser at <https://cells.ucsc.edu/?ds=psychad-snmultiome>.

External single-nucleus transcriptomic and chromatin-accessibility datasets were used for replication, label transfer, regulatory analyses, and projection of disease-state and donor-associated transcriptional axes. PsychAD snRNA-seq. PsychAD single-nucleus RNA-seq data and associated annotations from the Mount Sinai Brain Bank were obtained from the AD Knowledge Portal (Synapse ID: syn60084804). Donors aged  $\geq 60$  years without major neurological or psychiatric comorbidities were selected for the analyses described here. SEA-AD snRNA-seq. SEA-AD single-nucleus RNA-seq data and annotations were obtained from the Allen Institute SEA-AD data release (<https://registry.opendata.aws/allen-sea-ad-atlas/>). Middle temporal gyrus (MTG) and PFC samples were included in the relevant replication and donor-associated transcriptional axes analyses, while cell composition changes were restricted to the MTG region. Luquez et al. snMultiome. Single-nucleus multiome data from Luquez et al., including ATAC-seq fragment-level data, gene-by-cell expression matrices, and cell annotations, were obtained from the AD Knowledge Portal (Synapse ID: syn52339332). PFC and STG samples were included. Mathys et al. 2023 snRNA-seq. Single-nucleus RNA-seq gene-by-cell expression matrices and associated annotations from Mathys et al. 2023 were obtained from the AD Knowledge Portal (Synapse ID: syn52293442). Liu et al. 2025 snATAC-seq. Single-nucleus ATAC-seq data from Liu et al. 2025 were obtained from [https://compbio.mit.edu/AD\\_Multiomic\\_MultiRegion/](https://compbio.mit.edu/AD_Multiomic_MultiRegion/). Analyses were restricted to cortical samples, and the provided cell-by-cCRE accessibility matrix and annotations were used. Ma et al. 2022. Data from Ma et al. 2022 were obtained from the Neuroscience Multi-Omic Archive (NeMO Archive; <https://assets.nemoarchive.org/dataset/zkgd0it>). BICCN MTG. Human middle temporal gyrus reference data from BICCN were obtained from the NeMO Archive ([https://data.nemoarchive.org/publication\\_release/Great\\_Ape\\_MTG\\_Analysis/human\\_mat.RDS](https://data.nemoarchive.org/publication_release/Great_Ape_MTG_Analysis/human_mat.RDS)). SMART-seq cortical-layer reference. The SMART-seq reference dataset used for cortical-layer annotation was obtained from CELLxGENE (<https://datasets.cellxgene.cziscience.com/7ac99f87-4c02-45cc-aa3d-57d3540ee0f2.h5ad>).

## Code availability

All source code used in this study is available on GitHub at <https://github.com/DiseaseNeuroGenomics/PsychAD-snMultiome>. The frozen Dx\_cell models are available through Zenodo at <https://zenodo.org/records/23190790><sup>92</sup>.

## Acknowledgments

We are deeply grateful to the patients and their families who generously donated the biological materials that made this study possible. We sincerely appreciate their participation and commitment to advancing scientific knowledge and improving human health. We thank Dr. Masse and our colleagues at the Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, for helpful discussions and valuable feedback.

This research was supported by the National Institute on Aging of the National Institutes of Health under award numbers R01AG067025, R01AG082185, R01AG065582, U24AG087563, R01AG050986 and R01AG095776. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Author contributions

P.R. and P.D. conceptualized the study. J.F.F., C.C., A.H., M.A., Z.S., and J.K.M. performed experiments and generated the snMultiome data, with supervision from J.F.F. and P.R. V.H. provided and coordinated resources. P.D. and P.R. designed the analytical approach. P.D. led the computational analyses, with analytical contributions and supervision from G.E.H., J.B., D.L., and P.R. X.C. performed the gene regulatory network analyses. P.D. and S.K. performed the spatial transcriptomic analyses. P.D. and P.R. wrote the original manuscript, with contributions from all authors. All authors read and approved the final manuscript.

## Competing interests declaration

The authors declare no competing interests.

## References

1. DeTure, M. A. & Dickson, D. W. The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* **14**, 32 (2019).
2. Frisoni, G. B. *et al.* New landscape of the diagnosis of Alzheimer's disease. *Lancet* **406**, 1389–1407 (2025).
3. Wenk, G. L. Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiatry* **64 Suppl 9**, 7–10 (2003).
4. Gómez-Isla, T. & Frosch, M. P. Lesions without symptoms: understanding resilience to Alzheimer disease neuropathological changes. *Nat. Rev. Neurol.* **18**, 323–332 (2022).
5. Scheltens, P. *et al.* Alzheimer's disease. *Lancet* **397**, 1577–1590 (2021).
6. Arenaza-Urquijo, E. M. & Vemuri, P. Resistance vs resilience to Alzheimer disease: Clarifying terminology for preclinical studies. *Neurology* **90**, 695–703 (2018).
7. Dharshini, S. A. P. *et al.* Molecular signatures of resilience to Alzheimer's disease in neocortical layer 4 neurons. *Nat. Commun.* **17**, (2026).
8. Lee, D. *et al.* Single-cell atlas of transcriptomic vulnerability across brain disorders. *Nature* **657**, 988–1002 (2026).
9. Gabitto, M. I. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* **27**, 2366–2383 (2024).
10. Gazestani, V. *et al.* Early Alzheimer's disease pathology in human cortex involves transient cell states. *Cell* **186**, 4438–4453.e23 (2023).
11. Mathys, H. *et al.* Single-cell multiregion dissection of Alzheimer's disease. *Nature* **632**, 858–868 (2024).

12. Liu, Z. *et al.* Single-cell multiregion epigenomic rewiring in Alzheimer's disease progression and cognitive resilience. *Cell* **188**, 4980-5002.e29 (2025).
13. Luquez, T. *et al.* Cell-type signatures of Alzheimer's disease shared across population groups. *Nature* <https://doi.org/10.1038/s41586-026-10793-0> (2026) doi:10.1038/s41586-026-10793-0.
14. 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).
15. Green, G. S. *et al.* Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* **633**, 634–645 (2024).
16. Serrano-Pozo, A. *et al.* Reactive glia not only associates with plaques but also parallels tangles in Alzheimer's disease. *Am. J. Pathol.* **179**, 1373–1384 (2011).
17. Lu, A. *et al.* Human microglial transitions at the A $\beta$ -tau inflection point associate with divergent pathways to dementia and resilience. *Nat. Med.* **32**, 2047–2059 (2026).
18. Otero-Garcia, M. *et al.* Molecular signatures underlying neurofibrillary tangle susceptibility in Alzheimer's disease. *Neuron* **110**, 2929-2948.e8 (2022).
19. Zeng, H. *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446 (2023).
20. Fullard, J. F. *et al.* Population-scale cross-disorder atlas of the human prefrontal cortex at single-cell resolution. *Sci. Data* **12**, 954 (2025).
21. Grothe, M. J. *et al.* In vivo staging of regional amyloid deposition. *Neurology* **89**, 2031–2038 (2017).
22. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of A $\beta$ -deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791–1800 (2002).
23. Schöll, M. *et al.* PET imaging of tau deposition in the aging human brain. *Neuron* **89**, 971–982 (2016).
24. Lowe, V. J. *et al.* Widespread brain tau and its association with ageing, Braak stage and Alzheimer's dementia. *Brain* **141**, 271–287 (2018).

25. Bahl, E., Koomar, T. & Michaelson, J. J. cerebroViz: an R package for anatomical visualization of spatiotemporal brain data. *Bioinformatics* **33**, 762–763 (2017).
26. Hu, Y. *et al.* Multiscale footprints reveal the organization of cis-regulatory elements. *Nature* **638**, 779–786 (2025).
27. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
28. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
29. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
30. Jorstad, N. L. *et al.* Comparative transcriptomics reveals human-specific cortical features. *Science* **382**, eade9516 (2023).
31. Ma, S. *et al.* Molecular and cellular evolution of the primate dorsolateral prefrontal cortex. *Science* **377**, eabo7257 (2022).
32. Cain, A. *et al.* Multicellular communities are perturbed in the aging human brain and Alzheimer’s disease. *Nat. Neurosci.* **26**, 1267–1280 (2023).
33. Sun, N. *et al.* Human microglial state dynamics in Alzheimer’s disease progression. *Cell* **186**, 4386–4403.e29 (2023).
34. Lee, D. *et al.* Plasticity of human microglia and brain perivascular macrophages in aging and Alzheimer’s disease. *Nat. Genet.* **58**, 2152–2163 (2026).
35. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
36. 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).
37. Hohman, T. J. *et al.* Asymptomatic Alzheimer disease: Defining resilience. *Neurology* **87**, 2443–2450 (2016).

38. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
39. Oveisgharan, S. *et al.* Estrogen receptor genes, cognitive decline, and Alzheimer disease. *Neurology* **100**, e1474–e1487 (2023).
40. Bendl, J. *et al.* The three-dimensional landscape of cortical chromatin accessibility in Alzheimer's disease. *Nat. Neurosci.* **25**, 1366–1378 (2022).
41. Murdock, M. H. & Tsai, L.-H. Insights into Alzheimer's disease from single-cell genomic approaches. *Nat. Neurosci.* **26**, 181–195 (2023).
42. Lambert, S. A. *et al.* The human transcription factors. *Cell* **172**, 650–665 (2018).
43. Saeki, K., Pan, R., Lee, E., Kurotaki, D. & Ozato, K. IRF8 defines the epigenetic landscape in postnatal microglia, thereby directing their transcriptome programs. *Nat. Immunol.* **25**, 1928–1942 (2024).
44. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
45. Dong, P. *et al.* A multi-regional human brain atlas of chromatin accessibility and gene expression facilitates promoter-isoform resolution genetic fine-mapping. *Nat. Commun.* **15**, 10113 (2024).
46. Kosoy, R. *et al.* Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat. Genet.* **54**, 1145–1154 (2022).
47. Emani, P. S. *et al.* Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
48. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).
49. Kim, H. *et al.* Reactive astrocytes transduce inflammation in a blood-brain barrier model through a TNF-STAT3 signaling axis and secretion of alpha 1-antichymotrypsin. *Nat. Commun.* **13**, 6581 (2022).
50. Lananna, B. V. *et al.* Chi3l1/YKL-40 is controlled by the astrocyte circadian clock and regulates neuroinflammation and Alzheimer's disease pathogenesis. *Sci. Transl. Med.* **12**, eaax3519 (2020).

51. Freer, R. *et al.* A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease. *Sci. Adv.* **2**, e1600947 (2016).
52. Blanchard, J. W. *et al.* APOE4 impairs myelination via cholesterol dysregulation in oligodendrocytes. *Nature* **611**, 769–779 (2022).
53. Haroutunian, V., Katsel, P. & Schmeidler, J. Transcriptional vulnerability of brain regions in Alzheimer's disease and dementia. *Neurobiol. Aging* **30**, 561–573 (2009).
54. Fillenbaum, G. G. *et al.* Consortium to Establish a Registry for Alzheimer's Disease (CERAD): the first twenty years. *Alzheimers Dement* **4**, 96–109 (2008).
55. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239–259 (1991).
56. Braak, H., Alafuzoff, I., Arzberger, T., Kretschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol* **112**, 389–404 (2006).
57. Fort, A. *et al.* MBV: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* **33**, 1895–1897 (2017).
58. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018).
59. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, (2020).
60. Germain, P.-L., Lun, A., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDbtFinder. *F1000Res.* **10**, 979 (2021).
61. Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* **22**, 252 (2021).
62. Fisher, S. R. A. *Statistical Methods for Research Workers.* (1963).

63. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
64. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
65. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
66. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
67. Miao, Z. & Kim, J. Uniform quantification of single-nucleus ATAC-seq data with Paired-Insertion Counting (PIC) and a model-based insertion rate estimator. *Nat. Methods* **21**, 32–36 (2024).
68. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol.* **23**, 42 (2022).
69. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
70. 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).
71. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
72. Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
73. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
74. De Jay, N. *et al.* mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* **29**, 2365–2368 (2013).

75. Ke, G. *et al.* LightGBM: a highly efficient gradient boosting decision tree. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 3149–3157 (Curran Associates Inc., Red Hook, NY, USA, 2017).
76. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).
77. Hoffman, G. E. & Roussos, P. Fast, flexible analysis of differences in cellular composition with crumblr. *Nat. Commun.* **17**, (2026).
78. Hoffman, G. E. *et al.* Efficient differential expression analysis of large-scale single-cell transcriptomics data using Dreamlet. *Nat. Commun.* **17**, (2026).
79. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
80. Hoffman, G. E. & Roussos, P. Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics* **37**, 192–201 (2021).
81. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* (2016) doi:10.1101/060012.
82. Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978 (2010).
83. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
84. Bravo González-Blas, C. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).
85. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
86. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).

87. Rahman, S. *et al.* Lineage specific 3D genome structure in the adult human brain and neurodevelopmental changes in the chromatin interactome. *Nucleic Acids Res.* **51**, 11142–11161 (2023).
88. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
89. Kramer, N. E. *et al.* Plotgardener: cultivating precise multi-panel figures in R. *Bioinformatics* **38**, 2042–2045 (2022).
90. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models Using lme4. *J. Stat. Softw.* **67**, (2015).
91. Robinson, G. K. That BLUP is a good thing: The estimation of random effects. *Stat. Sci.* **6**, 15–32 (1991).
92. Dong, P. Trained cellular disease-state models for the PsychAD single-nucleus multiome study. Preprint at <https://doi.org/10.5281/ZENODO.23190790> (2026).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ADmultiomeSup.docx](#)
- [SupplementaryTables.xlsx](#)