PLXNB1 and other signaling drives a pathologic astrocyte state contributing to cognitive decline in Alzheimer's Disease

Authors

Natacha Comandante-Lou[1], Tsering Lama[1], Kevin Chen[1], Jinglong Zhang[2,3], Bin Hu[2,3], Shuhui Liu[2,3], Sarah Heuer[4], Kristen J. Brennand[5], Julie Schneider[6], Lisa Barnes[6], Bin Zhang[2,3,7], Minghui Wang[2,3], Hongyan Zou[8,9], Roland H. Friedel[8,9], Yiyi Ma[1], Tracy Young-Pearse[4], Aiqun Li[2,3], Masashi Fujita[1], David Bennett[6], Ya Zhang[1], Vilas Menon[1], Hans-Ulrich Klein[1], Mariko Taga[1], Philip L. De Jager[1*]

Affiliations

[1]Center for Translational and Computational Neuroimmunology, Columbia University Irving Medical Center, Department of Neurology and the Taub Institute for the Study of Alzheimer's Disease and the Aging Brain, New York, NY, United States of America

[2]Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[3]Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[4]Brigham and Women's Hospital, Harvard Medical School, Ann Romney Center for Neurologic Diseases, Boston, MA, United States of America

[5]Departments of Psychiatry and Genetics, Wu Tsai Institute, Yale University School of Medicine, New Haven, CT, United States of America

[6]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, United States of America

[7]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[8]Nash Family Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[9]Department of Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

*Correspondence: Philip L. De Jager, MD PhD

Abstract

Alzheimer's disease (AD) is marked by the coordinated emergence of disease-associated cell states across multiple cell types. Here, we first performed a meta-analysis of single-nucleus transcriptomic (snRNAseq) data from 869 brains of diverse decedents, confirming the critical role of an *SLC38A2*^{high}*SMTN*^{high}*CACNA1D*^{high} astrocyte subset, Astrocyte 10 (Ast10), in AD and aging-related cognitive decline. We then investigated the signaling drivers of Ast10's emergence in the aging brain, focusing on interactions among microglial and astrocytic subsets. Analysis of the snRNAseq data prioritized a set of ligands and receptors that are robustly predictive of Ast10 proportions across participants, and we confirm our predictions in multiple studies. Independent validation with spatial transcriptomics reveals striking colocalization of these prioritized ligands with the Ast10 signature in AD brain tissue, but not with other astrocytic states. Genetic ablation of a top receptor *PLXNB1* in murine and human iPSC-derived astrocytes decreased the Ast10 signature, confirming its regulatory role. Finally, we find that Ast10 may contribute to cognitive decline through synaptic loss and is associated with cognitive decline independent of AD. Thus, Ast10 and its regulators are potential points of convergence for multiple neurodegenerative mechanisms and may be promising targets for therapeutic development to preserve cognitive function.

Introduction

Alzheimer's Disease (AD) changes the cellular composition of the brain[1,2]. The frequency of many different cell states of various cell types, including neuronal, glial and vascular subpopulations has been associated with AD[1,3–11];yet, major questions remain regarding the origins of these various AD-associated cell states and how they interact with one another to create cellular communities that may drive the trajectory towards AD. In particular, characterizing these cellular changes to understand which cell subpopulations to target to preserve cognition in the face of AD pathology remains a major challenge[12,13].

To elucidate the sequence of changes in cell subpopulations leading to AD, we previously built a cell atlas of 1.65 million single-nucleus transcriptomes from the dorsolateral prefrontal cortex (DLPFC) of 437 participants in the Religious Orders Study or the Rush Memory and Aging Project (ROSMAP)[14,15]. These participants are without dementia at study entry and agree to brain donation; as a result, they represent a wide spectrum of aging brain states found in older individuals, allowing us to reconstruct the trajectory of brain aging and of the accumulation of common neuropathologies, such as amyloid and tau that define AD[1]. Based solely on the cell-state composition observed in each individual, we resolved a trajectory of brain aging that leads to accumulating more AD pathology and dementia (Progression to AD trajectory, prAD) (**Figure 1A**). Importantly, this trajectory's endpoint is characterized by an increased frequency of an astrocytic state that we called Ast10 (**Figure 1B**)[1].

We found evidence that Ast10 may statistically mediate the effect of tau proteinopathy on cognitive decline[1]; thus, we hypothesized that the increased frequency of Ast10 contributes to the neuronal dysfunction that ultimately manifests as cognitive decline. This astrocytic state was defined transcriptionally, and its transcriptional signature includes genes such as *SLC38A2*, *CACNA1D*, *SMTN*, and *SGCD*, which are associated with ion trafficking and extracellular matrix organization. Interestingly, Ast10 frequency is also associated with certain lipid-laden microglial states, such as Microglia State 12 (Mic12) and 13 (Mic13), which highly express AD risk genes *APOE* and *GPNMB* (**Extended Data Figure 1A**)[16]. Our previous causal modeling placed Ast10 downstream of Mic13 and included a direct effect of Mic13 on Ast10, separate from its role as a microglial subset contributing to tau proteinopathy[1]. These glial subtypes are part of the same cellular community that drives the Progression to AD (prAD) trajectory[1] and suggests that cell-cell interactions among these glial subpopulations may be important in driving cellular changes leading to AD. A systematic understanding of the molecular basis for such cell-cell coordination, and its consequences on individual cell subsets in the cellular environment of the aging brain, may reveal novel therapeutic approaches to restore brain homeostasis. In principle, blocking specific cell-cell signals among the causal cell subsets may enable us to prevent the emergence of this pathologic cellular community.

In our prior study[1], we used cell-state frequencies inferred from bulk RNA sequencing data to replicate our discovery analysis of the associations between each cell subset and AD traits. Here, we completed a staged meta-analysis of four single-nucleus RNA sequencing (snRNAseq) datasets from a total of 869 brains to first replicate and then extend our earlier results. This effort also interrogated ancestrally diverse individuals to explore whether the prioritized cell subpopulations have a similar role among different human populations. We then elected to pursue one of these cell subpopulations, Ast10, more deeply by investigating the cell-cell signaling molecules that drive its emergence. We prioritized Ast10 because our modeling suggested that it was directly upstream of cognitive decline, which is the ultimate clinically meaningful target for therapeutic development. Further, understanding the origins of Ast10 enables us to explore the key transition from asymptomatic accumulation of proteinopathies to cognitive impairment.

To identify the ligand-receptor pairs regulating the Ast10 state, we use the microglial and astrocytic transcriptomes from the 437 human aging cortices in our Discovery study[1] and computationally inferred the upstream regulatory molecular interactions that lead to the emergence of the Ast10 state. We then validated these results with (1) two additional snRNAseq datasets from the same brain region and (2) spatial transcriptomic data that assesses the colocalization of the predicted ligands with the Ast10 signature. Further, we extend the validation for one of the top predicted receptors, *PLXNB1*, in tissue sections from an AD brain, confirming its upregulation in Ast10 cells by immunofluorescence. Additionally, we characterized the responses of *in vitro* human and *in vivo* murine astrocytes to genetic ablation of *PLXNB1* at the single-cell level. Finally, we provide evidence that synaptic loss may be the mechanism by which Ast10 contributes to the emergence of cognitive decline and that it may play a role in non-AD contributors to dementia. Altogether, this study provides a framework with which to systematically identify the cell-cell communication signals that lead to the expansion of a pathogenic glial state. It sets the stage for identifying interventions that could prevent or potentially reverse this process and restore homeostasis of the brain cellular communities.

Results

Meta-analysis of snRNAseq data to replicate known and uncover new AD-associated subtypes

To understand the cellular basis of AD, we previously performed a Discovery analysis where we statistically associated the frequencies of the 91 cortical cell subsets with AD-related quantitative traits[1]. We refer to this as the *Columbia University Irving Medical Center 1 (CUIMC1)* dataset. Since we had previously replicated the results of this Discovery study[1] with cellular frequencies inferred from bulk cortex RNAseq data using the CelMod method[2], we undertook a snRNAseq-based replication effort using three other datasets. **Supplementary Table 1** presents the demographic and clinicopathologic characteristics of the four sets of participants. All participants have data produced from the same brain region, the dorsolateral prefrontal cortex, and they are all participants in one of five studies based at Rush University: the Religious Order Study (ROS)[14], the Memory and Aging Project (MAP)[15], the Minority Aging Research Study (MARS)[17], the African American Clinical Core[18], or the Latino Core Study[19] (See Methods for detailed description of cohorts). All participants undergo the same ante- and post-mortem characterization[17,19–21], allowing us to combine results across datasets for our three primary outcome measures: the burden of amyloid and tau proteinopathies[22] and the person-specific slope of cognitive decline derived from longitudinal assessments using 19 neuropsychological tests[23–25].

One of the three replication datasets has not been reported previously, and we refer to it as the *CUIMC2* dataset. It consists of 240 ROSMAP participants (see Data Availability); 219 of these participants are not found in the *Discovery* dataset (*CUIMC1*) and were retained for the Replication analysis (**Figure 1C**). The second replication dataset was generated at the *Massachusetts Institute of Technology (MIT)* (n=427 ROSMAP participants[3]); 132 of these participants are not represented in *CUIMC1* or *CUIMC2* datasets and are used in our analysis. To begin to evaluate to what extent our findings are replicated in diverse populations, we also included snRNAseq data generated by the Accelerating Medicines Partnership for AD Diversity project (Synapse ID: syn52339332) as our third replication set (*Diversity* dataset). We focused the analysis on the 81 participants who are African American (69%), Hispanic White (25%), or American Indian or Alaska Native (5%) based on their self-reported race and were not already sampled in one of the other three datasets. **Figure 1D** illustrates the extent of overlap among the four datasets. As described above, only one instance of each participant is used in our meta-analyses (See Methods). Notable differences among the four datasets include a significantly higher percentage of female participants (74%) in the *Diversity* dataset and a significantly lower percentage of participants with a diagnosis of AD per the National Institute of Aging's Reagan Criteria[26] (45%) in the *MIT* dataset. There was no significant difference in the slope of cognitive decline (P = 0.43) across the four datasets (**Supplementary Table 1, Extended Data Figure S1B - F**).

Armed with these datasets, we focused on linking cell-state frequencies with three quantitative measures of AD traits available for these cohorts: (1) an immunofluorescence-based measure of neocortical amyloid-β (Aβ) (one of the three monoclonal antibodies against 1–40 and 1–42 Aβ: 4G8, 6F/3D and 10D5[22,27,28]), (2) an immunofluorescence-based measure of neocortical tau (phosphorylated tau antibody AT8[22]), and (3) the slope of cognitive decline prior to death (**Figure 1C**; See Methods for full variable descriptions). Using the *CUIMC1 Discovery* dataset as the reference, we computationally annotated individual cells in the replication datasets based on our previous cell-state taxonomy of each cell type[1] (**Extended Data Figure S2-S7**, See Methods). We then performed a meta-analysis to test for evidence of association between the frequency of each cell subset and the AD traits, using a linear regression model while controlling for sex, age, post-mortem interval, education, and library quality. We note that this snRNAseq-based replication effort assessed all 91 cell states as previously defined, which is more than our prior replication effort where the frequency of only 62 cell states could be inferred from bulk RNAseq data.

The first meta-analysis used the three strata of Replication data (n = 432 non-overlapping participants from the three datasets), which has no overlap with the 437 participants of the Discovery (*CUIMC1*) analysis. This new replication effort recapitulates the top results previously reported in the Discovery analysis, revealing the same top four cell subpopulations that are significantly associated with two or more AD traits (based on False Discovery Rate (*FDR*) <0.05): lipid-associated microglial states (Mic.13, Mic.12), stress-response oligodendrocyte state (Oli.7) and astrocyte state (Ast10) (**Figure 1E**). To maximize the statistical power and assemble all of our data, we also performed a second meta-analysis using all participants (Discovery + Replication, n = 869). This analysis summarizes all available data and highlights some previously unreported associations, including the role of a microglial state Mic16 (marked by upregulation of the senescence-related gene *SERPINE1*)[1] and an endothelial state End3 (upregulated genes associated with extracellular matrix organization e.g., *ADAMTS9*)[1] with faster cognitive decline (**Figure 1E**). We also calculated a measure of heterogeneity (I^2)[29] among the four datasets and found that Mic16 exhibits the most heterogeneity in effect among the studies (I^2 = 86%) (**Supplementary Table 1**). One example of a cell subtype with a consistent effect (I^2 = 0) across datasets is Ast10 with the rate of cognitive decline (**Figure 1F, Supplementary Table 1**). Its effect size is consistent in all studies; only the Diversity dataset is not significant, possibly due to its smaller sample size.

Overall, the meta-analysis that includes a total of 869 brains from the Discovery and three replication datasets highlights the robustness of our previously reported results[1] relating AD traits and specific glial subsets. Our analyses also begin to provide some evidence that their results are also relevant to participants who are not of European descent (**Figure 1F**). In particular, consistently in both meta-analyses, Mic.13 shows the most significant associations with increased amyloid (*FDR* $< 1.7 \times 10^{-5}$) and tau burden (*FDR* $< 3.0 \times 10^{-4}$), whereas Ast10 is most significantly associated with faster cognitive decline (*FDR* $< 4.8 \times 10^{-6}$) (**Extended Data Figure S1G**, **Supplementary Table 1**). We therefore elected to further explore the interactions among astrocyte and microglial subpopulations that may drive astrocytes towards the Ast10 state.

Inference of glial intercellular communications that drive Astrocyte 10 transcriptional programs

Our previous statistical modeling places Ast10 downstream of Mic13 (an $APOE^{high}GPNMB^{high}$ microglial state) within the same tightly coordinated cellular community driving the trajectory of AD[1]. Therefore, we focused on identifying intercellular signals from microglial or astrocytic subsets that drive the polarization of astrocytes towards the Ast10 state. To this end, we used the snRNA-seq profiles of 81,255 microglia and 223,063 astrocytes isolated from the 437 donors of the *CUIMC1* dataset; these nuclei are partitioned into 16 transcriptionally distinct microglial and 10 astrocytic states. **First**, we performed a system-wide cell-cell communication inference to prioritize the upstream ligand-receptor pairs that may regulate the previously defined Ast10 signature (**Figure 2A**). **Second**, we used an orthogonal data-driven approach to further prioritize these inferred ligands by using multivariate modeling to predict each donor's Ast10 frequency based on the combinatorial expression of these ligands. **Third,** to validate these ligands, we used a series of approaches: *in silico* validation using two of the replication datasets and spatially registered profiling of human tissue at the transcript and protein levels as well as genetic perturbation of induced-pluripotent stem cell-derived astrocytes (iAst) and of murine astrocytes.

To systematically identify relevant upstream interactions between Ast10 receptors and ligands expressed by microglial and/or astrocytic subpopulations, we used NicheNet[30], which computationally infers ligand-receptor pairs that are not only differentially expressed but also *functionally interacting*. This analysis integrates prior knowledge of the ligand's signaling and gene-regulatory networks[30]. Specifically, we prioritized ligand-receptor pairs based on three key criteria: (1) the degree to which their downstream signaling activity recapitulates the gene expression signature of Astrocyte 10 cells (**Figure 2B**), (2) differential expression of receptors in Ast10 compared to all microglial and astrocytic subsets (**Figure 2D**) and (3) differential expression of ligands in a given sender cell subset versus all glial subsets (**Figure 2C**). Based on these criteria, we prioritized 46 ligand-receptor pairs (consisting of 39 ligands and 26 receptors). Among the top ligand-receptor interactions, many are associated with signaling processes such as axon guidance (e.g., *SEMA4B/D – PLXNB1*), cell adhesion (e.g., *JAM2 – JAM2*, *ROBO1 – NCAM1*), and remodeling of the extracellular matrix (e.g., *COL5A3 – DDR*1). Most of the top ligand-receptor interactions involve astrocyte subsets (including Ast10 itself) as senders (**Figure**

2C), suggesting that astrocytes as a cell type may self-modulate the frequency of different cell-states through paracrine and/or autocrine signaling[31]. There are also several microglial-specific signals that are associated with cell-adhesion (e.g., *DSCAM, SEMA4D*), G-protein signaling (e.g., *GNAS*) and a gene associated with AD and zinc disequilibrium *A2M*[32,33]. Together, this systemic inference of ligand-receptor pairs and their activities prioritize a set of cell-cell interactions that may drive astrocytes towards the Ast10 state.

Glial state specific ligand and receptor gene expression predict variable Astrocyte 10 frequency across aging and AD human brains

Having prioritized a subset of ligand-receptor pairs that may play a role in driving an astrocyte towards the Ast10 fate, we next questioned to what extent these NicheNet-inferred ligands and receptors may explain the donor-to-donor variability in Ast10 frequency among ROSMAP participants. In most individuals, Ast10 represents <1% of all astrocytes, but the frequency can go up to 80% in rare cases (**Figure 3A**). Notably, the enrichment of Ast10 is especially pronounced in participants with a clinical diagnosis of AD (**Figure 3A**). To understand the molecular basis of such variability, we used a multivariate approach, partial least-squares regression (PLSR)[34]. PLSR simultaneously links the expression of the NicheNet-prioritized ligands from specific microglial or astrocytic subpopulations to Ast10 frequency across 171 donors from the *CUIMC1* Discovery dataset who have at least one Ast10 cell (**Figure 3B**). In this model, each donor is treated as an observation and is given a set of 28 ligand summary scores to predict the donor's Ast10 frequency. To calculate the summary score for a given ligand, we averaged its pseudobulk expression only from glial subpopulations whose expression of that ligand is significantly correlated with Ast10 frequency across participants (**Extended Data Figure S8A, B**, See Methods for details). PLSR then reduces the dimensionality of the data (171 donors by 28 ligand scores) by transforming it into a principal component space (a set of orthogonal coordinates), in a manner that maximally captures the covariance between the *input signal* variable (ligand summary score profile) and the *response* variable (Ast10 frequency) (**Figure 3B**). To train the model, we used 90% of the 171 donors (*CUIMC1* training set) and withheld the remaining 10% for validation (*CUIMC1* test set). The PLSR model performance was evaluated by calculating the fraction of variance in Ast10 frequency explained (R^2) or predicted (Q^2) by the ligand summary scores (**Figure 3C**). The cell-state specific ligand model shows remarkably high accuracy in model fit and prediction of the *CUIMC1* training data, with $R^2 = 0.62$ and $Q^2 = 0.57 \pm 0.09$ (using repeated 5-fold cross-validation) for the first PLSR component (**Figure 3C**). To independently validate the model, we then used the trained model to predict Ast10 frequency in the remaining 10% of the donors. The model demonstrates consistent accuracy, capturing 0.64 of the variance (Q^2) in Ast10 frequency in the *CUIMC1* test set (**Figure 3D**).

To further assess the robustness of the ligand model, we used the model trained on the *CUIMC1* dataset to independently predict Ast10 relative proportions across non-overlapping individuals from the *CUIMC2* and *MIT* datasets. The *Diversity* dataset did not have a sufficient number of individuals with Ast10 and was therefore not considered in this replication effort. In the *CUIMC2* data (**Figure 3E**), the ligand model retained most of its performance, predicting 0.42 of the variance (Q^2) in Ast10 frequency. In the *MIT* dataset (**Figure 3F**), the ligand model predicts 0.36 of the variance (Q^2). Despite the reduced quantitative accuracy, the ligand model captures the overall pattern of the Ast10 frequency in the MIT data, given the strong and significant correlation (Pearson's $R = 0.64$, $P = 9.6\times10^{-10}$) between predictions and the observed frequencies. Thus, we have demonstrated that our PLSR ligand model is robust, yielding reproducible results in two different datasets that do not overlap with the set of participants from which the model was derived.

The high predictability of the PLSR model validates the NicheNet-inferred ligands that we have prioritized as being important in driving Ast10 transcriptional programs. Importantly, it also indicates a significant link between signaling cues from specific glial states and astrocyte state compositions, explaining the substantial interindividual variability in Ast10 relative proportions (**Figure 3A**). Accordingly, donors with low to high Ast10 relative proportions can be ordered based on their PLSR scores along the first PLSR component (PLS 1) (**Figure 3G**). The interpretability of PLSR enables us to then dissect the PLSR components in molecular terms based on their loadings, which describe the contribution of each *signal* or *response* variable to a given PLSR component. Consistent with the scores plot, the *response* variable (i.e., Ast10 frequency) is highly and positively weighted along PLS 1 (**Figure 3H**). Loadings of the *signal* variables reveal two distinct groups of ligands that contribute to PLS 1 in opposite directions (**Figure 3H**). To further assess the significance of each contributing ligand to Ast10 frequency prediction, we calculated the variable importance in projection (VIP) scores using the first PLSR component, with which the model achieves maximal accuracy (**Figure 3I**). Among the significant ligands (determined by a magnitude of VIP ≥ 1), axon guidance and cell-cell adhesion genes (e.g., *JAM2, NCAN, SEMA4D, LRRC4B*) have positive PLS 1 loadings, thus correlating with Ast10 enrichment (**Figure 3H, I**). In contrast, another group of significant ligands correlates with Ast10 state depletion, given their negative weights along PLS 1, including genes associated with extracellular matrix organization (e.g., *FN1, EFEMP1*), epidermal growth factor signaling (*NRG3*), synaptogenesis (*NRXN1*) and astrocyte reactivity after injury (*LGALS3*[35]) (**Figure 3H, I**). Together, analysis of the ligand PLSR model generates hypotheses on the signal-response relationships between combinatorial patterns of specific ligands and the enrichment of Ast10 cell subpopulation.

Next, we asked whether the astrocyte expression of the 26 NicheNet-prioritized receptors also predict Ast10 frequency across donors (**Extended Data Figure S8C**). Applying the PLSR modeling approach, we found that the receptor model is moderately predictive of Ast10 relative proportion in the *CUIMC1 Discovery* data, predicting 0.47 of the variance (Q^2) in the training data and even higher in the validation set (Q^2 = 0.62) (**Extended Data Figure S8D, E**). The performance of the receptor model is also robust when tested on additional samples from the *CUIMC2* and *MIT* datasets, predicting 0.39 and 0.35 of the variance (Q^2) in Ast.10 frequency, respectively **(Extended Data Figure S8F, G)**. Individuals with higher Ast.10 frequency show elevated PLS1 scores (**Extended Data Figure S8H**), driven by upregulated receptors (**Extended Data Figure S8I**). Consistently, among the most significant receptors (based on their VIP scores), we find *ADORA1* (*GNAS* receptor), *PLXNB1* (*SEMA4D/B* receptor), and *PTPRS* (*LRRC4B* receptor) whose corresponding ligands are also significant contributors to the ligand model (**Extended Data Figure S8J, Figure 3I**).

Together, these results reveal that the Ast10 subpopulation is associated with cell-cell interactions involving specific microglial states or other astrocytic states. The gene expression of these ligands and receptors in specific microglial or astrocytic states are highly predictive of the relative abundance of the Ast10 subpopulation across hundreds of individuals from different datasets.

Prioritized ligand pathway activities are associated with Astrocyte 10 signature at the single-cell level

To evaluate the activities of the significant ligands from the PLSR model (*EFEMP1, JAM2, NRXN1 NCAN, LGALS3, SEMA4D, FN1, GNAS, SEMA4B, LRRC4B, NRG3, RGMA, LRIG1*), we quantified the association between genes downstream of these ligands with Ast10 signature across individual astrocytes from the snRNAseq data (*CUIMC1*). We reasoned that for the prioritized ligands to be active, genes from their downstream pathways should be informative of individual astrocytes' Ast10 signature (regardless of it being up- or down-regulated by the ligand) (**Figure 4A**). To quantify such flow of information at the single-cell level, we used a metric called conditional-Density Resampled Estimate of Mutual Information (DREMI) that is designed to capture the non-linear associations among genes in single-cell data[36,37]. To alleviate the challenge of dropouts inherent to snRNAseq data, we use a data-diffusion approach called MAGIC[38] to impute gene expression across astrocytes, recovering gene-gene relationships that were otherwise obscured by dropouts (See Methods). For each gene (excluding Ast10 signature genes themselves) that was used by NicheNet to link the prioritized ligands to downstream Ast10 signature genes, we calculated a DREMI score between its expression and the Ast10 signature score (See Methods for signature score calculation) across individual astrocytes.

As expected, most of the ligands, their receptors as well as their downstream signaling mediators and transcription regulators have significantly higher DREMI scores compared to random permutations of the data, indicating that they share certain levels of mutual information with the Ast10 signature (**Figure 4B**). In particular, *JAM2*, predicted by NicheNet to mediate signaling among astrocytes themselves (**Figure 2C**), has a high DREMI score of 1.17 (among top 0.7% of all tested ligands; See Methods) when associating its expression across astrocytes at the single-cell level (**Figure 4C**). Similarly, the top receptor predicted by NicheNet, *PLXNB1*, has a high DREMI score of 1.09 (among top 1% of all receptors) in association with Ast10 signature. As predicted by the PLSR model, ligand *NRXN1* is associated with decreasing Ast10 frequency, its downstream signaling mediator, *PRKCA*, also decreases with Ast10 signature with a high DREMI score of 1.18 (among top 0.2% of signaling mediators). In summary, the pathway analysis provides further evidence of the activities of the prioritized ligands in regulating Ast10 state at the single-cell level.

Expression of prioritized ligands predicts in the vicinity of Ast10 in spatially registered transcriptomic data.

To validate the group of ligands prioritized by NicheNet and PLSR, we analyzed spatial transcriptomic data (Visium platform) from 32 sections of frozen dorsolateral prefrontal cortex obtained from 17 individuals selected from our Discovery cohort[39]. Among them, 15 donors had a clinical diagnosis of AD dementia and advanced AD pathologies, while 2 donors were cognitively unimpaired and showed no pathological signs of AD. (see **Supplementary Table 7** for demographic and clinicopathologic characteristics). This version of Visium captures RNA in an array of 55 um spots, and the RNA is subsequently reverse transcribed to cDNA and sequenced in a spatially registered manner[40]. We focused on ligands that are predicted to upregulate Ast10 signature (based on VIP scores ≥ 1) and asked whether the expression of these ligands in the vicinity of a given spot can predict the expression level of the Ast10 signature of that spot. To operationalize this analysis, we assessed data found within a radius of 3 spots (468 um) from the center of the target spot. In most tissue sections (26 out of 32 sections), we observed significant and positive correlations between the top ligand signature within the neighborhood and the Ast10 signature (representative sample shown in **Figure 5A,** statistics of all samples shown in **Figure 5C**). In contrast, no significant correlations were observed when associating each spot's Ast10 signature with the ligand signature in randomly located spots (**Figure 5 B**). Importantly, for most samples, we did not observe significant and positive correlations when associating the prioritized ligands with the upregulated signatures of other astrocyte subtypes (**Figure 5C**), suggesting that the prioritized ligands are specifically predictive for the Ast10 signature. Interestingly, Ast3 appears to be significantly depleted in the target spot, suggesting that Ast10 and Ast3 may be somewhat mutually exclusive in regions with high Ast10 ligand expression.

To further explore the spatial relationship between individual prioritized ligands and Ast10 signature genes, we used MISTy[41], an explainable machine learning framework to analyze the importance of each ligand in predicting each Ast10 signature gene. Specifically, we used MISTy to estimate the contribution of ligands within the local neighborhood (radius of 3 spots) on a given Ast10 signature gene, while adjusting for the intrinsic contribution from other Ast10 markers within the same spot. We found that many of the ligands in the PLSR model are significantly predictive (based on MISTy importance scores > 0; see Methods for detailed explanation of the metric) of most of the Ast10 signature genes (**Figure 5D**). To assess the significance of the prioritized ligand set (13 ligands with |VIP| ≥ 1 in the PLSR model), we performed a permutation analysis. Specifically, we randomly selected 13 ligands from the full set of possible ligands 100,000 times, calculating the mean total importance score for each set (defined as the sum of importance z-scores across all Ast10 signature genes). The prioritized ligands in the PLSR model exhibited significantly higher average total importance scores compared to most randomly sampled ligand sets (empirical *P*-value = 1.4×10^{-3}, **Figure 5E**). Taken together, these results derived from a different datatype (spatially registered RNAseq data) suggest that we have identified a set of ligands expressed in microglia or astrocytes that putatively regulate the Ast10 state. These ligands are prioritized and validated *in silico* based on several orthogonal approaches: 1) their ability to recapitulate the Ast10 signature given by prior evidence of the signaling pathways, 2) their predictability of Ast10 frequencies across individuals in several independent single-nucleus datasets, and 3) their spatial colocalization with Ast10 signature in brain tissue. Across these several approaches, we found that the *SEMA4D – PLXNB1* ligand-receptor pair consistently showed up as a significant contributor to the Ast10 state. Given that we had previously investigated *PLXNB1*'s role as a key node in a molecular network leading to cognitive decline[42], we elected to focus on the *SEMA4D – PLXNB1* pair in further experimental validation.

Ast10 marker colocalizes with *PLXNB1* in human tissue at the protein level

We first tested, *in situ*, whether *PLXNB1* is upregulated at the protein level in Ast10-like astrocytes in human tissue at the single-cell level. To this end, we used immunofluorescence (IF) to measure the expression of PLXNB1, Ast10 marker SLC39A11 and astrocyte cell-type marker GFAP, capturing 40 images at 20x magnification of a formalin-fixed paraffin-embedded (FFPE) DLPFC sample derived from a donor with AD. SLC39A11 was selected as a marker for the Ast10 signature due to the lack of specific antibodies against more strongly upregulated transcriptional markers (e.g., *SLC38A2, SMTN, CACNA1D*) for IF on human brain sections. Nevertheless, SLC39A11 is significantly upregulated in Ast10 compared to all the other astrocyte states (**Supplementary Table S2**).

For image analysis, the images were segmented to identify individual GFAP+ astrocytes using a custom pipeline implemented in CellProfiler[43]; the resulting cell-by-mean fluorescence intensity feature matrix was used in downstream analyses (See Methods**, Figure 6A**). To determine whether PLXNB1 protein expression level in a given astrocyte is predictive of its expression of SLC39A11 (used as a marker protein for the Ast10 signature), we fit a mixed effects model controlling for the fixed effect of GFAP levels and the random effect across images (See Methods). We observed a strongly positive and significant effect size ($P < 2\times10^{-16}$) of PLXNB1 in predicting SLC39A11 expression level across individual astrocytes, confirming the association between astrocytic PLXNB1 receptor and the Ast10 signature in human tissue (**Figure 6 B, C**). Interestingly, GFAP expression is moderately but inversely correlated with SLC39A11, suggesting that the Ast10 signature and the classic reactive astrocyte state, characterized by elevated GFAP levels, may be mutually exclusive[8,44,45]. (**Extended Data Figure S9A**).

Altered expression of *PLXNB1* reduces Ast10 signature in a mouse model and in human iPSC-derived astrocytes

To validate the importance of *PLXNB1* in regulating Ast10 signature in a human cellular model, we accessed a human induced pluripotent stem cell (iPSC) line in which *PLXNB1* had been knocked out[46] and its isogenic control line. These lines are derived from a donor who does not have a history of AD. Both hiPSC lines were then differentiated into astrocytes (iAst) by viral induction of *NIFB* and *SOX9* expression. We then performed scRNA-seq to profile transcriptome-wide changes due to *PLXNB1* perturbation. To minimize the effect of possible cell-to-cell variability in differentiation, we used the reference-mapping approach described above to annotate the iAst at the cell-type level. We then computationally selected the most astrocyte-like cells for downstream analysis (See Methods).

We first evaluated the similarity between the Ast10 transcriptional program in the *baseline* iAst model and that observed in postmortem human astrocytes. To achieve this, we performed a gene-gene correlation analysis, comparing the coexpression patterns of Ast10 markers in the baseline iAst model (iAst Control) with those in the snRNAseq dataset (**Extended Data Figure S9B**). In the baseline iAst model, Ast10 markers were grouped into two distinct coexpression gene modules. However, in the snRNAseq data, these two gene modules exhibited stronger positive correlations with each other (**Extended Data Figure S9B**). This suggests that while many Ast10 markers are coordinately expressed in the iAst model, patterns of transcriptional program expression are not identical to postmortem human astrocytes in the model system.

Next, we assessed the impact of genetic depletion of *PLXNB1* on the Ast10 transcriptional program in the iAst model system. In this model system, the gene editing targeting *PLXNB1* in iAst has significantly reduced *PLXNB1* expression ($P = 1.9\times10^{-76}$) as well as the top Ast10 markers, such as *SMTN* ($P = 5.1\times10^{-77}$), *SLC38A2* ($P = 6.5\times10^{--39}$), *SLC39A11* ($P = 1.7\times10^{-5}$), and *BCAN* ($P = 5.2\times10^{-5}$) (**Figure 6D**). However, not all Ast10 markers were significantly downregulated upon *PLXNB1* deletion, consistent with the hypothesis that *PLXNB1* alone is unlikely to be sufficient to drive the entire Ast10 signature. Instead, a combination of receptors may be required to recapitulate the entire signature. Further, we observed distinct responses among the two gene sets defined in iAst Control cells: Gene Set 2 of Ast10 markers showed a significant reduction in expression upon *PLXNB1* deletion as hypothesized ($P = 9.5\times10^{-31}$, **Extended Data Figure S9D**). Notably, the proportion of cells expressing high levels of *PLXNB1* RNA and Gene Set 2 markers decreased from 39.4% to 10% following *PLXNB1* knockout (**Figure 6E**). In contrast, Gene Set 1 (primarily metallothionein genes) exhibited a significant increase in expression upon *PLXNB1* knockout ($P = 4.2\times10^{-48}$, **Extended Data Figure S9C**).

Since the human cellular model does not fully capture the cellular context in the brain, we also tested the regulatory role of *PLXNB1* for Ast10 signature *in vivo*, by repurposing a published single-cell RNAseq dataset collected from mouse cortical tissues to compare reference (wild-type) and *Plxnb1* knock-out (*Plxnb1*$^{-/-}$) animals[46]. To computationally identify the astrocytes from this dataset, we annotated individual cells by their major cell types, using our snRNAseq Discovery data as reference (See Methods). As expected, astrocytes from *Plxnb1*$^{-/-}$ mice have significantly lower *Plxnb1* expression compared to the control ($P = 0$); there appears to be residual *Plxnb1* truncated RNA expression, which may be derived from 5’ exons that are spliced to the gene trap cassette inserted into the *Plxnb1* locus[47]. As anticipated, reduced *Plxnb1* expression *in vivo* significantly decreases the same top Ast10 marker genes that were reduced in the human *PLXNB1*-KO iAst line above: *Smtn* ($P = 4.8\times10^{-15}$), *Slc38a2* ($P = 1.5\times10^{-112}$), *Slc39a11* ($P = 2.6\times10^{-9}$), and *Bcan* ($P = 1.2\times10^{-6}$) (**Figure 6F**, **Extended Data Figure S9E**) when we look at all astrocytes. To more comprehensively identify Ast10-like cells in the mouse data, we calculated the Ast10 signature score for individual astrocytes using all upregulated Ast10 marker genes that were expressed in the mouse data (25 out of 39 of the Ast10 marker genes from human snRNAseq data have an orthologue in the mouse data; See Methods). In wild-type animals, 30.8% of astrocytes were high in both *Plxnb1* and Ast10 signature. Reduced *Plxnb1* expression in the *Plxnb1*$^{-/-}$ animals dramatically shifts the distribution of astrocytes: only 0.5% of cells remained high in both while 69% of cells were observed to be low in both *Plxnb1* and the Ast10 signature (**Figure 6G**). These results validate the regulatory role of astrocytic *PLXNB1* in inducing the Ast10 state *in vivo*.

Altogether, these findings demonstrate that *PLXNB1* drives at least a part of the main transcriptomic signature of Ast10 consistently across two model systems; it is likely that the complete signature only emerges with the activation of several receptors and the convergence of multiple different signals. Future studies are needed to validate additional putative ligand-receptor pairs that synergize with *PLXNB1* in inducing a more complete Ast10 signature.

Ast10 has a pathology-independent effect on cognitive decline and may exert its effect through synaptic loss

The significant, robust association between Ast10 frequency and cognitive decline across multiple independent studies suggest an important, potentially causal role of Ast10 in AD-related cognitive decline[1]. However, while AD pathology is a major contributor to aging-related cognitive decline, many other factors contribute to this process, and more than half of the variance in cognitive decline remains unexplained after all known risk factors are evaluated[48]. To understand the extent to which other factors may explain the association between Ast10 frequency and cognitive decline, we deployed a linear regression model that includes all known aging-related proteinopathies (amyloid, tau, Lewy bodies and TAR DNA-binding protein 43), multiple measures of neurovascular pathology, and other factors (such as the level of education and hippocampal sclerosis) that are known to be associated with cognitive decline (**Extended Data Figure S10**). We then added an additional term for the frequency of Ast10 (**Figure 6H**) which remains significantly associated ($P = 3.1 \times 10^{-6}$) with the slope of cognitive decline after all listed variables are accounted for. Thus, we can say that, beyond its role in AD, Ast10 captures the consequences of other, as yet undiscovered, processes that contribute to aging-related cognitive decline.

To begin probing the mechanism by which Ast10 affects cognition, we hypothesized that Ast10 may reduce synapse integrity, ultimately leading to cognitive decline. Synaptosomal-associated protein 25 (*SNAP25*) levels are strongly associated by cognitive decline in ROSMAP and has previously been used as a proxy for synaptic integrity[49,50]. Thus, we tested whether *SNAP25* protein expression mediates the effect of Ast10 on cognitive decline, leveraging matching bulk DLPFC proteomic data from a subset of the *CUIMC1* ROSMAP participants (n = 273 have snRNAseq and proteomic data). Ast10 frequency is significantly associated with both cognitive decline ($P = 0.001$) and *SNAP25* expression ($P = 2.8\times10^{-5}$) in these ROSMAP participants (**Figure 6I**). Higher levels of *SNAP25* protein, as previously reported[51], is associated with slower cognitive decline ($P = 0.002$). We can then test for mediation, and *SNAP25* mediates 25% ($P = 0.002$) of the association between Ast10 frequency and cognitive decline (**Figure 6I**). This result suggests a new hypothesis: that Ast10 cells may contribute to cognitive decline by causing synaptic loss (and hence lower SNAP25 protein levels) and subsequent neuronal dysfunction. This hypothesis connects with several of the characteristics of Ast10, such as genes involved in extracellular matrix remodeling (e.g., *JAM2*, *EFEMP1, NCAN*, **Figure 3I**), which have been implicated in perisynaptic nets that contribute to the maintenance of synapses.

Discussion

To date, an increasing number of disease-associated cellular states of every major cell type has been reported to be involved in AD[1,3–11]. However, there remains a major gap in our knowledge when it comes to the origins of these cell subpopulations and their inter-cellular coordination over the course of brain aging and disease progression. Our work and other reports have proposed that these disease-associated cellular states often appear in concert[1,2,39,45,46,52–56], suggesting that their origins are likely orchestrated by cell-cell interactions or through responses to shared microenvironmental signals.

Here, we first completed an essential step in linking specific cellular states to AD-related traits: presenting evidence that the Discovery study's results are replicable in independent snRNAseq datasets. This is difficult to achieve for practical reasons: primarily, the high cost of producing a second single-nucleus dataset of equal or greater size which would have proper statistical power to replicate the results of a discovery study. Our own previous effort had used cell-state frequencies inferred from bulk RNAseq data from an independent set of participants[1,2]. This replication effort was successful, but uncertainty remained. Also, certain cell subtypes could not be inferred robustly from bulk data and were not considered in the published replication analysis[1]. It was therefore essential to complete a replication of the original results at the single-cell level with the same datatype. Our meta-analysis of three replication datasets (n=432) provides definitive evidence that our results are robust, and the most up-to-date results are those emerging from the meta-analysis assembling all four datasets (n = 869 brains) (**Figure 1E, F**). These analyses also present an initial assessment suggesting that the findings may be relevant among non-European individuals, although much more work is needed to fully address this issue. Further, as expected, the increased sample size uncovers the role of additional cell subpopulations, such as microglial state Mic16 and endothelial state End3 which seem to be influencing the appearance of cognitive symptoms. These cell subpopulations now need further characterization, but they are beyond the scope of this manuscript; here, we elected to pursue the characterization of Ast10 as an exemplar, developing an approach to identify the intercellular signals that lead to the differentiation of a prioritized cell state.

After analyzing 4.2 million nuclei collected from 869 brains, an *SLC38A2*^{high}*SMTN*^{high}*CACNA1D*^{high} astrocyte state, Ast10, is the cell subpopulation (among 91 tested) that is most strongly associated with cognitive decline, the ultimate target for therapeutic development as many individuals with evidence of AD proteinopathy are cognitively unimpaired[3,57]. In our earlier work, Ast10 defined the terminus of a trajectory of brain aging that leads to AD[1]. Ast10 is characterized by a distinct transcriptional signature associated with intracellular ion transport, extracellular matrix organization and oxidative stress. Key marker genes in this signature include ion transporters such as

SLC38A2, *SLC39A11* and *CACNA1D*, cytoskeleton components such as *SMTN* and *SGCD*, as well as metallothionein genes like *MT1F* and *MT1G*. This signature stands apart from the more commonly described Disease-Associated Astrocyte (DAA) state, which typically exhibits an acute inflammatory response profile, including *GFAP*, *SERPINA3*, and *OSMR*[8,44,45]. Although the origins and specific functions of the Ast10 state have not been studied, a similar astrocyte signature (e.g. *SLC38A2*, *MT1G*, *CACNA1D*) was upregulated in AD samples from a smaller, independent cohort[5]. In another independent snRNAseq study that sampled the spatiotemporal progression of AD across 32 participants, Ast10 markers such as *SLC38A2, SLC39A11, PLXNB1* were part of an astrocytic gene set that were elevated at later stages of AD[58]. Together, these studies further support an important role of the lesser known Ast10 state in AD.

Focusing on Ast10 as an important but understudied cell subset, we performed an analysis to identify the microglial or astrocytic signals that may drive the emergence of the Ast10 state. **Our approach is distinct from common cell-cell communication inference approaches** which only prioritize ligand-receptor interactions based on the cell-type/subtype specificity of expression alone[59–61]; here, we computationally prioritized a set of ligands expressed in specific microglial or astrocytic states based on how well their downstream signaling recapitulates the Ast10 signature. These predictions are robust, as we replicate them in two other independent snRNAseq datasets. Further, we have illustrated how our multivariate modeling with PLSR can yield specific hypotheses that can be validated experimentally (i.e. *PLXNB1*). We thus have outlined an approach to understanding the paracrine regulation of important cellular states; such an approach can be repurposed to explore our other cell subpopulations that are associated with AD traits (**Figure 1E**).

A key validation effort in this study is the use of spatial transcriptomics to provide direct, independent evidence for the role of prioritized ligands in regulating the Ast10 signature within human tissue while preserving its spatial context. We analyzed spatially registered transcriptomic data from 32 DLPFC tissue sections obtained from 17 donors, the majority of whom exhibited advanced AD pathology. The prioritized ligands in the local neighborhood were significantly and specifically predictive of the Ast10 signature at the center of the neighborhood, compared to other astrocytic states, providing strong evidence for their role in driving the Ast10 state in AD brains (**Figure 5A-D**). Furthermore, when evaluating the importance of these prioritized ligands relative to all possible ligands in the spatial context, we found that they outperformed most others, underscoring their significance (**Figure 5E**). Interestingly, this analysis also identified additional ligands with high importance, potentially expressed in other cell types not included in our initial NicheNet analysis, warranting future investigation.

Methodologically, this study highlights the power of integrating multi-modal data: first identifying disease-relevant cell subpopulations and their signaling regulators through well-powered single-cell transcriptomics and then situating them within the tissue context using spatial approaches. Notably, over half of the significant predictors from the ligand PLSR model and spatial analysis were associated with axon guidance, synapse assembly, or synaptic function, including *SEMA4D, LRRC4B, RGMA, NRXN1, SEMA4B, LRIG1*, and *NRG3*. This enrichment of synapse-related molecules aligns with the potential role of Ast10 in modulating synaptic integrity, further supported by our mediation analysis with *SNAP25* (**Figure 6H**).

For one of these prioritized receptors, *PLXNB1*, we extended the validation to include immunofluorescence measurement in human tissue, genetic perturbation of model systems *in vitro* and *in vivo*. First, although *PLXNB1* is not part of the defining gene set for Ast10, its expression at the protein level in astrocytes is correlated to the protein expression of *SLC39A11*, a marker gene selected to represent the Ast10 signature. Further, genetic ablation of *PLXNB1* reduces Ast10 signature as expected in both a mouse model and a human iAst model (**Figure 6**): we found that depleting *PLXNB1* expression significantly reduces the Ast10 signature in both murine astrocytes and hiPSC-derived astrocytes, consistent with the hypothesis that *PLXNB1* contributes to the regulation of the Ast10 state. *PLXNB1* has been associated with various central nervous systems disease contexts[46,62,63]. In AD, we previously identified *PLXNB1* as a key node of a gene module from bulk RNAseq data that is significantly associated with cognitive decline in the ROSMAP cohort[42]. Additionally, *PLXNB1* may be important in maintaining cell-cell distancing of microglia and astrocytes around plaques[46]. Our finding that *PLXNB1* regulates the Ast10 state therefore unify these separate observations by connecting them to a specific subset of astrocytes implicated in synaptic loss that is a feature of AD[64].

Our results also revealed a specific extracellular matrix (ECM) profile associated with the Ast10 state, adding to existing evidence that suggest an important role of astrocyte-ECM interactions in neurodegeneration[65,66]. In particular, our PLSR model shows that the enrichment of Ast10 in the cortex is associated with increased levels of the lectican family of chondroitin sulfate proteoglycans, such as neurocan (*NCAN*) and, to a lesser extent, with brevican (*BCAN*) (**Figure 3I**), which are key components of the perineuronal ECM[67]. In contrast, fibronectin (*FN1*) that makes up the basement membrane ECM surrounding the vasculature[68] are associated with diminished Ast10 frequency (**Figure 3I**). These results suggest that Ast10 may preferentially reside around neurons, possibly interacting with the perisynaptic ECM to alter synaptic functions. Profiling human cortical tissues at the single-cell level using spatial transcriptomic technologies (e.g. MERFISH[69]) will be needed to reveal the spatial organization of Ast10 and interacting neuronal subsets to test this hypothesis.

The exact role of such lectican-rich ECM microenvironment in AD remains unclear. While lectican may serve as a protective barrier for neurons[70], it may also mediate neuroinflammatory responses[68,71] or inhibit synaptic plasticity[72,73]. These seemingly opposite roles of lecticans likely depend on their versatile post-translational modifications. Since astrocytes are a primary source of lecticans, future studies should investigate the potential mutual influence between Ast10 and surrounding ECM, and how such interactions may impact synaptic function in AD.

Altogether, these results suggest that the increased frequency of Ast10 may lead to disruption of the established peri-synaptic nets and subsequent loss of synapses by neurons in AD. However, we also show that Ast10 explains a significant amount of the variance in aging-related cognitive decline that is not explained by known risk factors such as AD proteinopathies. Thus, Ast10 may also play a key role in other, as yet unknown, processes that contribute to loss of cognition in older age (**Figure 6H**). The Ast10 state may be a key nexus for multiple different pathophysiological processes that lead to synaptic loss and cognitive dysfunction, making it and its regulators promising targets for therapeutic development.

Materials and Methods

Participants

Human molecular profiles and clinicopathological traits used in this study are measured from participants from the following five community-based cohort studies of aging and dementia. All five studies are conducted by the Rush Alzheimer’s Disease Center (RADC) and share a large common core of data collected by the same team[19]. All five were approved by an Institutional Review Board of Rush University Medical Center. All participants enrolled without known dementia and agreed to annual clinical evaluation. All participants signed informed and repository consents and brain donors signed an Anatomic Gift Act. Below are summaries of the study designs and demographics:

Religious Orders Study (ROS)[14]

The Religious Orders Study (ROS) is a longitudinal epidemiologic clinical-pathologic cohort study focused on aging and Alzheimer's disease. Participants include older Catholic nuns, priests, and brothers from diverse groups across the US. Enrollment began in 1994, with participants agreeing to annual clinical evaluation, cognitive testing, and a subset have annual blood draw. All participants agree to brain donation and sign an Anatomic Gift Act.

Rush Memory and Aging Project (MAP)[15]

The Rush Memory and Aging Project (MAP) is a longitudinal epidemiologic clinical-pathologic cohort study, emphasizing common chronic conditions of aging, cognitive decline, and Alzheimer's risk. It includes older men and women recruited from retirement communities and senior housing facilities in Chicagoland and northeastern Illinois. Enrollment began in 1997, and all participants undergo annual clinical evaluation, cognitive testing, and annual blood draw. All participants agree to organ donation (e.g., brain, spinal cord, nerve and muscle) and sign an Anatomic Gift Act.

Minority Aging Research Study (MARS)[17]

The Minority Aging Research Study (MARS) is a longitudinal epidemiologic cohort study focusing on cognitive decline and Alzheimer's risk in older African Americans. Recruitment began in 2004, exclusively enrolling older African American men and women who undergo annual clinical evaluation, blood draw, and brain donation is optional.

African American Clinical Core[18]

The African American Clinical Core is a longitudinal epidemiologic clinical-pathologic cohort study examining the transition from normal aging to mild cognitive impairment (MCI) to early dementia. Recruitment of the current cohort began in 2008. Participants, primarily older African Americans from the Chicago area, undergo annual clinical evaluation, blood draw, and brain donation is optional.

Latino CORE Study (LATC)[19]

The Latino CORE Study (LATC) is a longitudinal epidemiologic cohort study investigating cognitive decline and Alzheimer's risk in older Latinos. Recruitment started in 2015, enrolling older Latino/Hispanic adults who undergo annual clinical evaluation, blood draw, and brain donation is optional.

AD phenotype measures

We focused on the following two continuous measures of AD pathological traits, which have higher statistical power than discrete classifications.

Rate of cognitive decline: estimated person-specific rate of change in the global cognition variable over time. Global cognition is assessed annually using 21 cognitive tests, 17 of which are shared across studies and used to derive summary measures for five cognitive domains: episodic memory, semantic memory, working memory, visuospatial ability, and perceptual speed. The global cognition score is computed by averaging the standardized scores of these 17 tests. The methods for assessing cognition are detailed in previous publications[23–25]. The person-specific slope of cognitive decline was then measured by modeling the annual global cognitive scores using a linear mixed effects model, controlling for age at baseline, sex and years of education[25].

Aβ and tau pathology burden: Pathological measures were collected (as described previously in detail[19,21,22]). Briefly, percent area of cortex occupied by Aβ or abnormally phosphorylated tau-positive neurofibrillary tangles, which were measured at death by immunohistochemistry. Aβ immunostaining was performed with one of three monoclonal antibodies: 4G8 (1:9000; Covance Labs, Madison, WI), 6F/3D (1:50; Dako North America Inc., Carpinteria, CA), or 10D5 (1:600; Elan Pharmaceuticals, San Francisco, CA). Paired helical filament (PHF) tau proteins were detected using the phosphorylated tau-specific antibody AT8 (Innogenetics, San Ramon, California, USA; 1:1000). Immunostaining was conducted as previously described on 20 μm paraffin-embedded tissue sections collected from 8 brain regions (the hippocampus, entorhinal cortex, anterior cingulate cortex, midfrontal cortex, superior frontal cortex, inferior temporal cortex, angular gyrus, and calcarine cortex), followed by image analysis for quantification[22,27,28]. Global measures for Aβ and tau burdens were summarized by averaging the measurements obtained from the eight brain regions. In the association analysis, we used the square-root transformed burden measures for better statistical properties.

Other neuropathological measures

Other neuropathologies were also measured using immunohistochemistry as previously described[74]. Presence of TAR DNA-binding protein 43 (TDP-43) cytoplasmic inclusions in neurons and glia were assessed in eight brain regions (amygdala, entorhinal cortex, hippocampus CA1, hippocampus dentate gyrus, anterior temporal pole cortex, midtemporal cortex, orbital frontal cortex, and midfrontal cortex) using a phosphorylated monoclonal TAR5P-1D3 antibody (pS409/410)[75]. TDP-43 pathology was staged into four categories: none, amygdala only, amygdala + limbic, and amygdala + limbic + neocortical. Lewy body disease was evaluated in four stages (none, nigral-predominant, limbic-type, neocortical-type) across seven regions (substantia nigra, anterior cingulate cortex, entorhinal cortex, amygdala, midfrontal cortex, superior or middle temporal cortex, inferior parietal cortex) using α-synuclein immunostaining with monoclonal antibody LB509[76]. Cerebral amyloid angiopathy (CAA) pathology was analyzed in four neocortical regions (midfrontal, midtemporal, parietal, calcarine cortices) using one of three monoclonal anti-amyloid-β antibodies: 4G8, 6F/3D, or 10D5[77]. Amyloid-β deposition in meningeal and parenchymal vessels was scored from 0 to 4, and the scores across regions were averaged to obtain a continuous composite score for CAA[77]. Additionally, large vessel cerebral atherosclerosis and arteriolosclerosis were graded based on visual and histological examination of arterial vessels after paraformaldehyde fixation[78,79]. Chronic cerebral infarctions and microinfarcts were assessed in a minimum of nine brain regions stained with hematoxylin and eosin (H&E)[80]. Hippocampal sclerosis was identified by neuronal loss and gliosis in CA1 or subiculum regions stained with H&E[81].

Generation of single-nucleus RNA & ATAC-seq dataset (CUIMC2 dataset)

Nuclei Isolation

Nuclei were isolated with EZ PREP buffer (Sigma, Cat #NUC-101). Tissue samples cut into pieces < 0.5 cm or cell pellets were homogenized using a glass dounce tissue grinder (Sigma, Cat #D8938) (25 times with pastel A, and 25 times with pastel B) in 2 ml of ice-cold EZ PREP and incubated on ice for 5 minutes, with additional 2 ml ice-cold EZ PREP. Nuclei were centrifuged at 500 x g for 5 minutes at 4°C, washed with 4 ml ice-cold EZ PREP and incubated on ice for 5 minutes. After centrifugation, the nuclei were washed in 4 ml Nuclei Suspension Buffer (NSB; consisting of 1x PBS, 0.01% BSA and 0.1% RNAse inhibitor (Clontech, Cat #2313A)). Isolated nuclei were resuspended in 100-500uL NSB depending on pellet size, filtered through a 35 μm cell strainer (Corning, Cat # 352235). Nuclei were counted using the Nexcelom Cellometer Vision 10x objective and an AO/PI stain. 20μl of the AO/PI was pipet mixed with 20ul of the nuclei suspension and 20μl was loaded onto a Cellometer cell counting chamber of standard thickness (Nexcelom catalog number: CHT4-SD100-002) and counted with the dilution factor set to 2. Determine volume needed for 250,000 nuclei from each donor to create a 3-plex donor pool. Pooled nuclei were centrifuged at 500 x g for 5 minutes at 4°C. Nuclei pellet was resuspended in 50-70uL NSB depending on pellet size, filtered through a 35 μm cell strainer. Nuclei were counted with AO/PI again. A final target of 21,500 nuclei was used for 10x Single Cell Multiome ATAC + Gene Expression experiments.

Single-nucleus RNA & ATAC-seq profiling

After nuclei isolation the 10x multiome protocol (Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle, PN-1000283) was followed according to manufacturer's instructions (10X Genomics, USA) and can be accessed at https://cdn.10xgenomics.com/image/upload/v1666737555/support-documents/CG000338_ChromiumNextGEM_Multiome_ATAC_GEX_User_Guide_RevF.pdf.

Briefly, following transposition GEMs were generated by combining barcoded Gel Beads, transposed nuclei, a Master Mix that includes reverse transcription (RT) reagents, and Partitioning Oil on a Chromium Next GEM Chip J (10× Genomics; PN-2000264). Incubation of the GEMs in a thermal cycler for 45 minutes at 37°C and for 30 minutes at 25°C generates full-length cDNA from poly-adenylated mRNA for gene expression (GEX) library and a Spacer sequence that enables barcode attachment to transposed DNA fragments for Assay for Transposase-Accessible Chromatin (ATAC) library. This was followed by a quenching step that stopped the reaction. Next, GEMs were broken, and pooled fractions were recovered. Silane magnetic beads were used to purify the first-strand cDNA from the post GEM-RT reaction mixture. Barcoded transposed DNA and barcoded full-length cDNA from poly-adenylated mRNA were pre-amplified by PCR and the products were used as input for both ATAC library construction and cDNA amplification for gene-expression library construction. Libraries were pooled and sequenced together on a NovaSeq 6000 with S4 flow cell (Illumina, San Diego, CA). The scATAC-seq libraries were sequences as follows: Read 1N, 50 cycles; i7 Index, 8 cycles; i5 Index, 24 Cycles; Read 2N, 49 cycles. The snRNA-seq libraries were sequenced as follows: Read 1, 28 cycles; i7 Index, 10 cycles; i5 Index, 10 Cycles; Read 2, 90 cycles.

Single-nucleus RNA & ATAC-seq data processing

Alignment and technical artifact adjustment

Through the application of CellRangerArc (v2.0.2), the raw sequence reads were aligned to human genome hg38. To adjust for technical artifacts from background RNA and random barcode swapping, the *remove-background* function from CellBender (v0.3.1) was conducted over the RNA matrices from CellRanger. In short, CellBender models from each library’s gene expression matrix levels of empty droplets (low UMI count/droplet with high droplet counts) and removes empty/contaminated droplets. The parameters set for CellBender *remove-background* include: epochs=150, fpr=0.01, expected-cells=5000, and total-droplets-included set to the number of cells in the library from the CellRanger gene expression matrix.

Demultiplexing

Demultiplexing was performed using the demuxlet software (v1.11). Given each library contains 3 or 4 individuals, and each individual has their own WGS VCF file from ROSMAP, we were able to identify which cell belonged to which study sample. Succinctly, BAM files with filtered VCF were taken from the output of CellRanger, popscle generates a pileup file required for demuxlet, and demuxlet assigned scores to identify corresponding cells to study samples.

Normalization and clustering

The normalization pipeline for the RNA gene expression matrices began with the *NormalizeData* method from Seurat (v4.4.0), followed by *FindVariableFeatures* (selection.method='vst', nfeatures=2000), *ScaleData*, dimensional reduction with *RunPCA* (npcs=30), creation of k-NN neighbor with *FindNeighbors* (dims=1:30), and cluster detection with *FindClusters* (resolution = 1.5, algorithm = 1).

Cell Type Classifications

Cell types were annotated with weighted ElasticNet linear regression trained off prior atlas of human ROSMAP dorsolateral prefrontal cortex tissue (n=24)[2] where 7 major cell types are present (excitatory neurons, inhibitory neurons, vascular cells, oligodendrocytes, astrocytes, microglia, and oligodendrocyte precursor cells). Weights were applied as the inverse of the number of nuclei per cell type to preserve classifications of smaller cell types like the vascular cells. Cross validation was performed to optimize regulation parameters using *glmnet* (*cv.glmnet*) (v4.1) with an alpha=0.25, nfolds=10, family='multinomial'.

Single-nucleus RNA & ATAC-seq data quality control

Low quality cell assignment and removal

Given biological differences in the numbers of genes expressed (nGene) and the number of UMIs (nUMI) between different cell types, we optimized seven different thresholds for each cell type where cells with nGene or nUMI under the thresholds would be labeled as low quality. To do so, we first picked 6 out of the 61 libraries (RA006, RA007, RA010, RA018, RA035, RA039) controlling for sex, age at death, and AD prevalence. We then manually identified clusters within each library that are low quality, judging by the overlap of cell type classifications, a high mitochondrial gene expression, and a low nGene/nUMI relative other clusters. Afterwards, we derived a harmonic mean using precision/recall to identify nUMI/nGene thresholds per cell type: excitatory neurons=2331/1068, inhibitory neurons=857/397, vascular cells=146/493, oligodendrocytes=776/490, astrocytes=528/442, microglia=291/150, and oligodendrocyte precursor cells=667/554.

Doublet removal

To account for doublets across samples, demuxlet doublet identification based off of WGS-VCF files of each sample was applied. Additionally, to address doublets across different cell types (e.g. Excitatory Neuron-Microglia doublet), we utilized Gilad Green's customized DoubletFinder's *DoubletFinder_v3* method (https://github.com/GreenGilad/DoubletFinder) where pN=0.5, pK=75/(1.5*(#nuclei/library), nExp=0, sct=T, labels=cell type information for each library. Thus any nuclei classified as either a demuxlet doublet or DoubletFinder doublet were subsequently removed.

Mitochondrial gene expression

Following guidelines from the 10X genomic multiome page, we filtered out cells possessing a mitochondrial RNA percent expression outside of 2 standard deviations within each library (https://www.10xgenomics.com/analysis-guides/common-considerations-for-quality-control-filters-for-single-cell-rna-seq-data).

Cell state annotations transfer across datasets

To replicate cell subset associations with AD pathologies using independent snRNAseq datasets, we first annotated the cellular states of individual cells from each cell type separately, based on the cell-state definitions from Green et al[1]. For each cell type, we mapped cells from each strata of the replication dataset as query onto our reference CUIMC1 Discovery dataset, using the Seurat query mapping pipeline[82]. Briefly, the reference and query data were first normalized, and the top variable genes were identified using the *SCTransform* function while regressing out percent of mitochondrial reads. A set of mapping anchors (pairs of cells that encode the relationships between the reference and query) was identified and scored using the *FindTransferAnchors* function with default parameters. These anchors were then used to transfer the cell-state annotations from the reference to the query data using the *TransferData* function. In this pipeline, we wevaried three key parameters in this pipeline, while keeping other parameters as default: the number of top variable genes (*variable.features.n*, ranged 1000 – 4000), the number of principal components for finding anchors (*npcs, ranged 10 –100)*, and the number of neighbors to consider when weighing anchors (*k.weight*, ranged 10 – 50). The parameters set for each cell type in each dataset are detailed in **Supplementary Table 1**.

Cell-state frequency calculation

We grouped the cell subpopulations into seven broad cell types as previously defined[1]: astrocytes, microglia, inhibitory neurons, excitatory neurons, oligodendrocytes, oligodendrocyte precursor cells

(OPCs) and vascular niche, which includes endothelial cells, arteriole cells, venule cells, pericytes, smooth muscle cells (SMC) and fibroblasts. We excluded rare subpopulations such as macrophages, monocytes, erythrocytes, CD8$^+$ T cells, NK cells and neutrophils due to the low abundances. Here, we define the frequency (or relative proportion) of a cell subpopulation as the number of cells of the *cell-state* of interest divided by the total number of cells of the same *cell-type*. For each donor, the frequency of subpopulation *s* is calculated as:

$$f_s = \frac{n_s}{\sum_{j=1}^{K} n_j}$$

where n_s is the number of cells for cell-state *s* and *K* is the total number of cell-states within the cell-type to which cell-state *s* belongs. Hence, the cell-state frequency here reflects the empirical probability of a cell of a given cell-type acquiring the transcriptional state of interest.

Statistical analysis associating AD traits and cell-state frequencies

For each dataset, statistical associations between AD traits and cell subset frequencies were tested using linear regression models as previously described[1]: $trait \sim \sqrt{f_s}$ + covariates, where f_s is the frequency of cell subpopulation *s* (defined in the previous section). The frequencies were square-root transformed to better capture the subtle changes of the less abundance cell subsets. To adjust for potential confounding variables, we included the following variables as covariates: 1) participants characteristics (sex, age, post-mortem interval, education), and 2) snRNA-seq library quality metrics (cell count and median of gene count for each participant). Results from the linear regression models were corrected for multiple hypothesis testing by calculating the false-discovery rate within each AD trait using the Benjamini-Hochberg correction method.

To integrate the statistical associations across datasets, we performed fixed-effect meta-analyses using the *metagen* function from the meta (v8.0-1) package[83] in R (v4.4.2). The meta-analyses were performed using an inverse-variance based approach by summing the effect size (β_i) weighted by the inverse of standard error (SE_i) of each study *i*. Mathematically, the summarized z-statistics for each association was calculated as:

$$Z = \beta / SE \text{ , where}$$

$$\beta = \frac{\sum_i \beta_i w_i}{\sum_i w_i} \;; \qquad SE = \sqrt{1/\sum_i w_i} \;; \qquad w_i = 1/SE_i^{\,2} \text{ for study } i$$

Meta-analysis results were then corrected for multiple-hypothesis testing using the Benjamini-Hochberg method within each AD trait. To assess the heterogeneity of statistical associations across studies, we calculated the I^2 statistics[29]

$$I^2 = max\left\{0, \frac{Q-(K-1)}{Q}\right\}$$

where K is the number of studies and Q is the Cochran's Q[84]:

$$Q = \sum_{i=1}^{K} w_i \, (\beta_i - \frac{\sum_i \beta_i w_i}{\sum_i w_i})^2$$

Identification of cell state signatures

To identify signature genes for a specific cell state, we performed pairwise differential expression (DE) analysis, comparing the gene expression of the cell state of interest against each of the other cell states within the same cell type. The input for the DE analysis was a pseudobulk expression matrix, generated by aggregating UMI counts for each gene across cells in each cell state for each individual. Lowly expressed genes were filtered using the *filterByExpr* function from the edgeR package (v3.40.2) with default parameters. To account for the mean-variance relationship across genes, we applied the voom function from the limma package (v3.54.2), which generated a set of precision weights. These weights were subsequently used to fit a linear model for each gene with the *lmFit* function. The statistics and fold changes of the linear model fit were computed using the *eBayes* function (with trend = FALSE and robust = FALSE).

To ensure specificity of the signature genes, we identified genes that were significantly differentially expressed (adjusted $P < 0.05$) in the same direction (either upregulated or downregulated) across *all* significant pairwise comparisons involving the cell state of interest. Genes exhibiting inconsistent regulation (upregulated in some comparisons and downregulated in others) were excluded. To further refine the selection of uniquely differentiating genes, we retained only those genes with at least one pairwise comparison showing a significant log-fold change (LFC) magnitude greater than 1, while ensuring that no comparisons had an LFC magnitude below 0.25. The final list of signature genes used for downstream analyses is provided in **Supplementary Table S2**.

Inference of cell-cell signaling drivers of glial states

To identify the cell-cell signaling that drives the transcriptional programs underlying a specific glial state (Ast10), we applied NicheNet[30] using its R implementation nichenetr (v2.0.0). NicheNet prioritizes ligand-receptor pairs according to their downstream signaling activities. Briefly, NicheNet models ligand activities by utilizing three types of prior knowledge from several complementary resources, including ligand-receptor, signaling and gene-regulatory interactions. NicheNet integrates such prior knowledge into an expansive intracellular signaling network where each ligand-to-target link is given a score, termed *regulatory potential*. Such prior information is then used to infer potentially active ligand-receptor interactions by searching for highly differentially expressed ligand-receptor pairs with high regulatory potential scores for the given set of target genes in receiver cells.

To uncover the cell-cell signaling mechanisms driving the transcriptional programs underlying the Astrocyte 10 (Ast10) astrocytic state, we utilized NicheNet[30], implemented via the R package nichenetr (v2.0.0). NicheNet prioritizes ligand-receptor pairs based on their downstream signaling activity. It leverages prior knowledge from complementary datasets on ligand-receptor interactions, signaling pathways, and gene-regulatory networks. This information is integrated into a comprehensive intracellular signaling network, where each ligand-to-target link is assigned a *regulatory potential* score. For a given set of target genes in receiver cells, NicheNet identifies active ligand-receptor pairs by prioritizing those with high regulatory potential scores and significant differential expression.

NicheNet requires three primary inputs: (1) the target gene set, (2) the background gene set, and (3) expression data for sender and receiver cell populations. For our analysis, we defined the Ast10 transcriptional signature (identified from differential expression analysis) as the target gene set and all genes available in the NicheNet ligand-target prior model as the background. Ast10 was designated as the receiver, while the 10 astrocyte subsets (including Ast10 itself) and 16 microglial subsets were set as sender populations.

Ligand activities were quantified using the NicheNet function *predict_ligand_activities* with default parameters, with activity measured as the area under the precision-recall curve (AUPR). This metric reflects the concordance between a ligand's predicted targets and the observed transcriptional signature relative to background genes. Ligands were then prioritized using weighted criteria, including scaled ligand activity (*activity_scaled* = 2), differential upregulation of sender ligands (*de_ligand* = 1) and receiver receptors (*de_receptor* = 1), and the average expression of sender ligands (*exprs_ligand* = 1) and receiver receptors (*exprs_receptor* = 1).

Based on this prioritization, we selected the top 100 sender-ligand-receptor combinations, resulting in 49 unique ligand-receptor interactions comprising 41 ligands and 27 receptors for further analysis.

Partial least squares regression (PLSR) modeling

To test whether NicheNet-prioritized ligands or receptors predict the frequency of the Ast10 subset across donors, we applied partial least squares regression (PLSR)[85]. PLSR is a supervised regression technique that models the relationship between a set of predictors (e.g., ligand summary scores or receptor expression) and an outcome variable (e.g., Ast10 frequency). Unlike principal component analysis (PCA), which identifies orthogonal components explaining the maximum variance in the data without regard to an outcome, PLSR identifies components that maximize covariance between the predictors and the outcome. This makes PLSR particularly suited for linking high-dimensional, covarying data to a particular outcome of interest.

Model construction

1. **Receptor Model:** The input matrix for the receptor model consisted of the pseudobulk expression (log-transformed mean expression) of NicheNet-prioritized receptors in all astrocytes across donors. The output vector consists of the log-transformed percentage of Ast10 in each donor with an offset of 1.
2. **Cell-state Specific Ligand Model:** For the ligand model, we considered the pseudobulk expression of NicheNet-prioritized ligands in sender cell subsets (microglial or astrocytic states). We first identified sender-ligand pairs where ligand expression significantly correlated with Ast10 frequency across donors ($|\text{Pearson's } R| \geq 0.1$, $P < 0.05$). For ligands with at least one significant correlation, we calculated a summary score by averaging the scaled expression (zero-centered) of the ligand across significant senders. In cases where a ligand (e.g., *NRG3*) exhibited opposing correlations (positive and negative) with Ast10 frequency across different sender subsets (**Extended Data Figure S8A**), we calculated separate summary scores for each direction. This approach reduces noise by excluding non-significant senders and addresses the challenge of sparsity in glial state representations across donors. The output vector for this model was identical to that used in the receptor model.

Model Training and Validation

All modeling steps, including training, cross-validation, and independent validation, were conducted in Python (v3.9.2) using the scikit-learn library (v0.24.1)[86].

Training and Testing: Donors were randomly split into 90% for training and 10% test set for independent validation. Input features (ligand summary scores or receptor expression) were normalized to zero mean and unit variance using the StandardScaler() function.

Modeling training and Cross-validation: PLSR models were trained using the *PLSRegression*() function. Model performance was evaluated by the fraction of variance explained (R^2) and predicted (Q^2), calculated using the *r2_score* function to assess the concordance between the empirical Ast10 frequency estimated from the data and model prediction. To evaluate model predictability on the training data, we performed repeated 5-fold cross-validation with the *RepeatedKFold*() function (*n_splits* = 5, *n_repeats* = 100) on PLSR models with varying number of PLS components (**Figure 3C**, **Extended Data Figure S8**)

Variable Importance in Projection (VIP)

To assess the relative importance of individual ligands or receptors in the PLSR models, we calculated the Variable Importance in Projection (VIP) scores for the first PLS component, where the models achieve their optimal performance. VIP scores summarize the contribution of each input variable (ligand or receptor) to the model, weighted by the variance explained by each PLS component. Mathematically, the VIP score for a given variable *r* is calculated as:

$$VIP_r = \sqrt{\frac{R\sum_{k=1}^{K}\omega_{kr}^2 \cdot SS_k}{\sum_{k=1}^{K} SS_k}}$$

where R is the total number of input variables (28 ligands or 26 receptors), w_{kr} is the weight of the r^{th} variable for the k^{th} PLS components, and SS_k is the sum of squares explained by the k^{th} PLS component. To help interpret the directionality of the associations, the VIP scores for each ligand or receptor were multiplied by the sign of the Pearson's correlation coefficient between the predictor (ligand or receptor expression) and Ast10 frequency.

Gene expression imputation

To address the challenge of data sparsity caused by stochastic mRNA capture, or drop-outs, which can obscure gene-gene relationships, we applied MAGIC[38] (Markov affinity-based graph imputation of cells) to impute gene expression. Such imputation is necessary when assessing the relationships between individual genes with Ast10 signature at the single-cell level. MAGIC learns the intrinsic structure of the data and uses data diffusion to share information across similar cells. Such a data diffusion approach can better recover the underlying gene-gene relationships that are typically obscured due to dropouts.

We implemented MAGIC in R using the Rmagic package (v2.0.3) to impute the gene expression of astrocytes in different datasets. For astrocytes in the Discovery snRNA-seq dataset, we first randomly sampled 2500 cells per astrocytic state to ensure equal representation, totaling 25,000 astrocytes. Raw count data were then normalized using the *library.size.normalize* function from the phateR package (v1.0.7), followed by a square-root transformation as input for MAGIC. Imputation was performed on all genes expressed in at least 10 cells, with parameters k-nearest neighbors (*knn*) = 10 and Markov matrix power (*t*) = 5.

For astrocytes in the mouse and iAst scRNA-seq datasets, we followed a similar workflow. Raw count data for astrocytes in each dataset were normalized, and MAGIC was applied to all genes expressed in at least 10 cells, but with dataset-specific parameters: *knn* = 5, *t* = 3 for mouse scRNA-seq data; *knn* = 5, *t* = 1 for iAst scRNA-seq data.

The resulting imputed gene expression data were used for downstream analyses, including the calculation of DREMI scores and Ast10 signature scores.

Signature summary score

To evaluate the overall expression of a specific gene signature or gene set in individual cells, we utilized the *AddModuleScore* function in Seurat to derive a summary measure for that signature. The function computes the summary score for each cell as the average expression across the signature genes subtracted by the aggregated expression of a set of randomly selected control features[87]. Below details the summary score calculation for specific analyses:

- *DREMI analysis and analysis of the mouse scRNAseq data*: signature scores were calculated using the MAGIC-imputed expression of differentially *upregulated* genes in the Ast10 signature, as identified in the differential expression analysis.
- *Analysis of the iAst scRNAseq data*: In the analysis of the iAst dataset, we observed that the gene-gene correlation structure of Ast10 markers differed from that observed in the snRNA-seq human data (**Extended Data Figure S9B**). Consequently, the upregulated Ast10 markers were further divided into two sub-gene sets, and separate summary scores were calculated for each set using their MAGIC-imputed expression (**Extended Data Figure S9C, D**).
- *Spatial correlation analysis*: For the spatial analysis, we assessed correlations at the spot level in the Visium data between two summary measures, 1) the prioritized Ast10-predicting ligand signature and 2) astrocyte-state signatures. Here, we computed the ligand signature score using normalized expression values (See *Spatial Correlation Analysis* section below) of the set of Ast.10 ligands with significant and positive VIP scores (VIP ≥ 1). Similarly, signature scores for other astrocytic states were computed using the normalized expression of their respective differentially upregulated marker genes, except for Ast.1, which lacked uniquely upregulated markers.

To avoid confounding the correlations, any genes in question were excluded from the signature score calculations in all association analyses described above.

Identify ligand-associated pathway genes

For each ligand with a significant VIP score (|VIP| ≥ 1), we focused on reconstructing its pathway leading to the Ast10 differentially expressed genes that are among the top 300 targets of the ligand according to the NicheNet database. To determine the signaling pathways through which a ligand is predicted to regulate the Ast10 signature genes of interest, we used the *get_ligand_signaling_path_with_receptor* function from NicheNet to identify the intermediate genes. This function first identifies the top 5 transcription factors that best regulate the target genes and are closest to the upstream ligand, based on prior knowledge of the ligand-signaling and gene regulatory networks which are represented by a set of weights. These weights were then used to determine the shortest paths between these top transcription factors and the ligands. Genes along these shortest paths are considered as significant nodes within the pathway. To classify their roles within the pathway, these pathway nodes were annotated as one of the following categories: ligands, receptors, signaling mediators or transcriptional regulators, based on the OmniPath (R package OmnipathR, v3.9.6) *intercell* resources[88] (for ligand, receptor annotations) as well as the *TFcensus*[89] and *DoRothEA* resources[90] (for transcriptional regulator annotations).

DREMI to quantify pathway information flow

To quantify the relationships between a given ligand-associated pathway gene and Ast10 signature, we used a metric called DREMI (conditional-Density Resampled Estimate of Mutual Information)[36]. DREMI leverages the cell-to-cell variation in single-cell data to quantify the amount of information transmitted from gene X (i.e. pathway gene of interest) to gene Y (i.e. Ast10 signature score) in signaling networks. Notably, DREMI is a directional measure, as it quantifies the information flow using conditional density (i.e. probability of Y given X) rather than the joint density of X and Y. In estimating the conditional densities we adopted the k-nearest neighbor based-DREMI that was previously applied in scRNA-seq data following MAGIC imputation[38].

To set the baseline for the DREMI score, we randomly permutated the gene expression vector X, then calculated a DREMI score between the permuted X and the Ast10 signature score. We repeated this process for each pathway gene in question, then average their DREMI scores to set the baseline. To evaluate the significance of each pathway gene's DREMI score, we first derived DREMI scores for all genes available in each of these resources: OmniPath (R package OmnipathR, v3.9.3) *intercell* resources[88] (for the ligand, receptor gene categories), the *TFcensus*[89] and *DoRothEA* resources[90] (for the transcriptional regulator gene category), as well as 10779 randomly selected genes that do not belong to any of the above categories (for the signaling mediator gene category). We then calculated an empirical P-value (P_{emp}) *for* gene x, based on all genes within the same signaling node category:

$$P_{emp\ (x \subset K)} = S_K / N_K$$

where S_k is the number of genes with a higher DREMI score than gene x within the signaling node category K (i.e., ligand, receptor, signaling mediator or transcriptional regulator) and N_k is the number of genes in the signaling category K.

Visium data generation

The spatial transcriptomic data with Visium combined with immunofluorescence was generated as previously described[1,39]. Fresh-frozen dorsolateral prefrontal cortex (DLPFC) samples from ROSMAP participants were processed using the Visium Spatial Transcriptomics (ST) platform, coupled with immunofluorescence. Tissue sections encompassing all cortical layers and some white matter were selected based on RNA quality (RNA Integrity Number > 6). RNA was purified using the RNeasy Micro Kit (Qiagen, #74004) and RNA integrity was assessed using TapeStation 4150 and Bioanalyzer 2100 (Agilent Genomics).

Tissue blocks were prepared into ~6 × 6 mm², 10 µm thick sections embedded in Optimal Cutting Temperature (OCT; Sakura, 4583) and were mounted on Visium Tissue Optimization Slides and Gene Expression Slides (10x Genomics) in duplicates. Permeabilization time was optimized based on the Visium Spatial Tissue Optimization protocol (10x Genomics, CG000238 Rev D) with modifications. After fixation in cold 100% ethanol (–20°C for 30 minutes), sections were stained with Thioflavin S (ThioS; Sigma, T1892) to label neuritic plaques and an anti-GFAP Cy3-conjugated antibody (Millipore Sigma, MAB3402C3, dilution 1/50) to visualize astrocytes. Autofluorescence from lipofuscin was quenched using TrueBlack (Biotinum, 23007). Sections were then briefly immersed in 3X SSC buffer (Millipore Sigma, S6639L) and mounted with a solution containing 85% glycerol, 2 U/µL RNase inhibitor, and DAPI (Thermo Fisher, #62248).

Sections were imaged at 10X magnification using a Nikon Eclipse Ni-E immunofluorescence microscope at 10X magnification (Plan Apo λ, NA = 0.45). Following imaging, coverslips were removed, and sections were permeabilized to optimize mRNA release. A 3-minute permeabilization time was identified as optimal for the Visium workflow, balancing effective transcript release and minimal diffusion. Subsequent cDNA synthesis, library preparation, and sequencing were performed following manufacturer protocols (10x Genomics, Visium Spatial Gene Expression Reagent Kits, CG000239 Rev D), with sequencing conducted on a NovaSeq 6000 (Illumina). ST spots (diameter = 55µm) were aligned with brightfield images.

Visium data preprocessing and cortical layer annotations

Visium sequencing reads were aligned to the human reference genome (GRCh38) and were assigned to spots using Space Ranger (v2.0.0, 10x Genomics). The resulting h5 files for each sample were imported into R (v4.2.2) using the SpatialExperiment (v1.16.0). For quality control, mitochondrial and ribosomal genes, as well as genes expressed in fewer than 10 spots, were excluded. Additionally, spots located outside the tissue or those with ≤ 1,000 UMIs or with ≤ 500 expressed genes were removed.

Cortical layers and white matter were identified as previously described[39], by first selecting the top 10,000 most variable genes using the *modelGeneVar* function from the scran package (v1.26.2)[91]. To focus on layer-specific features, we extracted 3,003 marker genes by taking the top 429 marker genes for each of the six cortical layers and the white matter, as reported in a previous study[92]. We then intersected these marker genes with our highly variable genes, resulting in a final set of 1,357 highly variable marker genes. Principal component analysis (PCA) was then performed on this gene set, and the first 50 principal components (PCs) were used to mitigate batch effects using Harmony (v1.2.0), treating each tissue section as a separate batch and applying parameters *theta* = 3, *sigma* = 0.1, and *lambda* = 1. To identify spatial clusters, we applied BayesSpace (v1.8.2)[93] to the first 35 Harmony-corrected PCs, using a spatial smoothing parameter of *gamma* = 2. This approach partitioned the spots into seven clusters, corresponding to six cortical layers and white matter. White matter spots were excluded from all subsequent analyses.

Spatial correlation analysis of nearby Ast10-predicting ligands

For each tissue slide, we assessed Pearson's correlations between two summary measures at the spot level in the Visium data: 1) the averaged ligand signature predicting the Ast.10 state within the neighborhood, and 2) a specific astrocyte-state signature. Both signature scores were computed using normalized expression values obtained via Seurat's *SCTransform* function with default parameters, adjusting for the number of nuclei per spot and the percentage of mitochondrial reads.

To compute the ligand signature, we derived a summary score (See *Signature summary score* section above) focusing on the prioritized Ast.10 ligands with significant and positive VIP scores (VIP ≥ 1): *JAM2, NCAN, SEMA4D, GNAS, LRRC4B, RGMA* (**Figure 3G**). This ligand signature score was then averaged across spots within a 3-spot (468 µm) radius around the target spot, providing an estimate of the overall expression of Ast.10 ligands in the neighborhood. To establish a baseline for comparison, we also averaged ligand signature scores across the same number of spots randomly selected from within the tissue for each target spot.

The astrocyte-state signature score for the target spot was computed similarly, using each astrocytic state's respective marker genes (see *Identification of cell-state signatures* section). Since we focused on linking the ligands to enrichment of specific astrocyte states, we limited the analysis to states with uniquely *upregulated* markers for most astrocyte states.

Finally, all Pearson's correlation results were corrected for multiple hypothesis testing using the Benjamini-Hochberg method to calculate false discovery rates for each astrocytic state.

Estimation of the effects of nearby ligands on Astrocyte 10 signature genes using MISTy

To assess the importance of individual ligands to the expression of Astrocyte 10 (Ast10) signature genes in spatial contexts, we utilized the explainable machine learning framework MISTy[41], implemented via the R package mistyR (v1.10.0). This approach allowed us to decompose the contributions of intrinsic and spatial factors to Ast10 signature gene expression at the spot level using a multi-view predictive model. The model incorporated two distinct spatial contexts as predictors:

1. **Intrinsic view:** Captures the relationships among Ast10 signature genes within the same spot. The intrinsic view included Ast10 signature genes identified from the differential expression analysis (see *Identification of receiver glial state signature* section).
2. **Juxta view:** Reflects the cumulative expression of individual ligands from immediate neighboring spots, defined as spots within a three-spot radius. In the juxta view, all potential ligands, including those prioritized for their predictive relevance to Ast10 signature from the PLSR analysis, were included as candidate predictors.

The multi-view model was fitted independently for each of the 32 slides, generating a *standardized importance score* for each ligand. This score reflects the variance reduction in the ligand's prediction of the expression of a given Ast10 signature gene. For each ligand–signature gene pair, importance scores were averaged across all 32 slides to obtain an aggregated score (**Figure 5D**), representing the spatial association between a ligand and an Ast10 signature gene (though not implying causation). To evaluate each ligand's overall contribution to the collective Ast10 signature, we first z-score transformed the aggregated importance scores within each Ast10 signature gene and then summed these z-scores across all Ast10 signature genes to obtain the total importance score. To assess the significance of the PLSR-prioritized ligand set ($|\text{VIP}| \geq 1$) in predicting the overall Ast10 signature, we randomly sampled ligand sets of the same size 100,000 times, computing the mean total importance score for each set. The average total importance of the PLSR-prioritized ligand set was then compared against the distribution of randomly sampled sets to derive an empirical P-value (**Figure 5E**).

Immunofluorescence microscopy

A six µm formalin-fixed paraffin-embedded (FFPE) tissue section from the dorsolateral prefrontal cortex (Brodmann Area 9) was de-paraffinized with CitriSolv (xylene substitute, Decon Laboratories, 1601H) for 20 min. Heat-induced epitope retrieval was performed with citrate buffer (pH=6) using a microwave (800W, 30% power setting) for 25 minutes. The section was washed with phosphate-buffered saline (PBS) and blocked with a bovine serum albumin (BSA) medium (Sigma-Aldrich, A7906) for 30 minutes at room temperature (RT). The primary antibodies, anti-PLXNB1 (Bio-Techne, MAB37491, 1:100) and ZIP11 (Thermo Fisher, PA5-20679, 1:100), were incubated overnight at 4°C. The tissue section was washed three times with PBS and incubated with fluorochrome-conjugated secondary antibodies (Thermo Fisher, A21202, A31573, 1:500) for 1 hour at RT. After washing with PBS, GFAP-conjugated Cy3 (Millipore Sigma, MAB3402, 1:100) was incubated for 2 hours at RT. After washing, the section was treated with True Black Lipofuscin Autofluorescence Quencher (Invitrogen, P3693) to minimize autofluorescence. An anti-fading reagent with DAPI (Life technology, P36931) was used for tissue mounting.

Immunofluorescence image acquisition and analysis

Imaging was performed using the Nikon Eclipse Ni-E epifluorescence microscope at 20X magnification with the NIS-Elements Advanced Research software (v5.21.03). A Hamamatsu Orca-Fusion Digital Camera (C14440) was utilized to capture 40 images in a systematic zigzag pattern, ensuring coverage of all cortical layers for post-acquisition analysis. The Plan Apochromatic λ objective lens with a numerical aperture of 0.75 was used for high-performance resolution imaging across a range of wavelengths. The camera was set to a 1X1 binning configuration to enhance signal sensitivity and improve signal-to-noise ratio while maintaining a fast image acquisition. The epifluorescence microscope was equipped with a high color rendering LED light source from a Lumencor system, providing stable and efficient illumination for fluorescence imaging. For multi-channel fluorescence, a Multilaser Laser Induced Detection Area (LIDA) system was employed, integrating multiple lasers, each assigned to specific fluorophores, enabling simultaneous excitation and detection of their emissions. The excitation lasers used were 450 nanometers (nm), 550 nm, and 640 nm, with individual detection channels for the emitted light. The emission wavelengths were set as follows: DAPI (365 nm), FITC (488 nm), TRITC (561 nm), and Cy5 (640 nm). The exposure time for all images acquired at 20X magnification was 40 milliseconds.

To extract single-cell level protein measurements from the images, astrocytes were segmented using CellProfiler (v4.2.5)[94] based on DAPI+ and *GFAP*+ objects. First, nuclei (DAPI) were segmented using the “IdentifyPrimaryObjects” module with advanced settings. The object diameter was set between 18 and 80 pixels, and thresholding was performed with the “Robust Background” method, without adjustment. The lower and upper bounds of the threshold ranged from 0 to 1. For *GFAP* segmentation, the “EnhanceOrSuppressFeatures” module was used to augment the ramified features of astrocytes by applying the "Line Structures" method. *GFAP*+ objects were then segmented using the “IdentifyPrimaryObjects” module, with object diameters ranging from 10 to 300 pixels. Thresholding was performed again with the “Robust Background” method without adjustment. The lower and upper bounds of the threshold ranged between 0 to 1. The “SplitOrMergeObjects” module was used to group smaller objects, such as ramified structures, with larger nearby objects. To exclude *GFAP*+DAPI- objects, the “RelateObjects” module defined DAPI as the parent object and *GFAP* as the child object. The mean intensities of *SLC39A11*, *GFAP*, and *PLXNB1* were measured for each *GFAP*+DAPI+ object. All intensity measurements from CellProfiler were rescaled to a range between 0 to 16 bits by multiplying a factor of 65535 ($2^{16} – 1$), followed by log10 transformation.

To statistically test for the association between *PLXNB1* expression and Ast10 marker *SLC39A11* expression at the single-cell level, we fit a mixed effects linear regression model. The model predicts *SLC39A11* mean intensity based on *PLXNB1* mean intensity using all astrocytes (*GFAP*+DAPI+ cells). *GFAP* mean intensity was included as a covariate, while random effects accounted for variability in both intercepts and slopes for *GFAP* and *PLXNB1* mean intensities, as well as variability across images. All continuous variables were z-score normalized prior to model fitting. The model was implemented using the *lmer* function from the lme4 package (v1.1-35.5) in R. The structure of the model is as followed:

$$
\begin{aligned}
SLC39A11\ Mean\ Intensity &\sim PLXNB1\ Mean\ Intensity + GFAP\ Mean\ Intensity \\
&\quad + (1 + GFAP\ Mean\ Intensity + PLXNB1\ Mean\ Intensity \mid Image)
\end{aligned}
$$

Here, *PLXNB1* mean intensity and *GFAP* mean intensity were modeled as fixed effects; the terms in parentheses represent random effects, allowing intercepts and slopes for *GFAP* and *PLXNB1* to vary across images. This mixed-effects modeling approach accounts for both single-cell-level relationships and image-level variability, providing robust estimates of the association between *PLXNB1* and *SLC39A11* expression.

Generation of isogenic pairs of *PLXNB1* KO hiPSCs

The isogenic pair of *PLXNB1* KO hiPSCs from a control donor (CD1: 553-3) were generated using CRISPR/Cas9, as previously described[46]. Three guide RNAs (gRNAs) targeting *PLXNB1* Exon 3 were obtained through the Mount Sinai CRISPR Initiative Program. **Table 1** provides details on the gRNA sequences.

Table 1 | Details on the guide RNAs (gRNAs) used for generating *PLXNB1* KO hiPSCs.

Oligo ID	Location	gRNA	PAM	gRNA sequence
*PLXNB1*_1	48424318 - 48424340	#1	GGG	ATTCGGGTTGTTGGTAGGCT
*PLXNB1*_2	48424317 - 48424339	#2	TGG	GATTCGGGTTGTTGGTAGGC
*PLXNB1*_3	48424313 - 48424335	#3	AGG	AGCTGATTCGGGTTGTTGGT

Three gRNAs were individually packaged into lentiviruses using a third-generation system and transduced into Cas9-expressing hiPSCs (lentiCRISPR v2, Addgene #52961). After antibiotic selection, single-cell clones were plated, expanded, and screened. Successful *PLXNB1* knockout was confirmed by PCR amplification, TOPO cloning and Sanger sequencing.

Human iPSC-derived astrocyte induction

The isogenic pair of *PLXNB1* KO hiPSCs (from a single clone) were induced into astrocytes by FUW-M2rtTA (Addgene, 20342), TetO.Sox9.puro (Addgene, 117269), and TetO.NfiB.Hygro (Addgene, 117271), as previously described[46,95,96]. On day 0, add doxycycline (Dox) to induce *SOX9* and *NFIB* expression. On day 1, add dual antibiotics (puromycin and hygromycin) to enrich transduced hiPSCs, then expand cells in Expansion Medium (DMEM/F-12, HEPES, N2, 10% FBS, GlutaMAX). On day 3, switch to Astrocyte FGF Medium (Neurobasal, B27, GlutaMAX, CNTF, BMP4, FGF2) with Dox. On day 8, transition to Astrocyte Maturation Medium (DMEM/F-12, Neurobasal, N2, sodium pyruvate, EGF-like growth factor, CNTF, BMP4, cAMP, NAC) with Dox. On day 11, fully withdraw Dox. Collect hiPSC-derived astrocytes between days 21-24.

scRNAseq of hiPSC-derived astrocytes

The induced astrocytes were collected and filtered through a 70 µm cell strainer into a fresh falcon tube. The viability of single cells was assessed using Acridine Orange/Propidium Iodide viability staining reagent (Nexcelom), and debris-free suspensions of >80% viability were deemed suitable for the experiments. ScRNA-Seq was performed using the Chromium platform (10x Genomics) with the 5' gene expression (5' GEX) V2 kit, with a targeted recovery of 10,000 cells. Briefly, Gel-Bead in Emulsions (GEMs) were generated on the sample chip in the Chromium X system. Barcoded cDNA was extracted from the GEMs after Post-GEM RT-cleanup and amplified for 12 cycles. Full-length cDNA from poly-A mRNA transcripts underwent enzymatic fragmentation and size selection to optimize cDNA amplicon size (~400 bp) for library construction. The cDNA fragments were then subjected to end-repair, adapter ligation, and 10x-specific sample indexing according to the manufacturer's protocol (10x Genomics). The concentration of the single-cell library was accurately quantified using qPCR (Kapa Biosystems) to achieve appropriate cluster counts for paired-end sequencing on the NovaSeq 6000 (Illumina), aiming for a sequencing depth of 25,000 reads per cell.

Preprocessing scRNA-seq dataset of mice with genetic Plxnb1 perturbation

We used Seurat to preprocess the raw UMI count matrices generated by Huang et al using single-cell RNA sequencing of cortices from mice with or without *Plxnb1* deletion[46]. For quality control, we removed cells with ≤200 or ≥ 5000 genes, UMI count ≥ 50,000, as well as those with mitochondrial reads > 10%. We then removed genes that were expressed in less than 2 cells. After quality control, we retained 32,874 cells and 20,557 genes in total.

Preprocessing scRNA-seq data from hiPSC-induced astrocytes

Sequencing reads were aligned to the human reference genome (GRCh38) using Cell Ranger (v 5.0.1). Using similar preprocessing steps as in the mouse scRNA-seq analysis, we performed quality control on the hiPSC-induced astrocyte (iAst) scRNA-seq data, starting from the raw UMI count matrices generated by Cell Ranger. We retained cells that met the following quality control thresholds: UMI count > 1800 and < 50,000, number of genes > 800 and < 7000, and mitochondrial reads < 10%. Additionally, we removed genes expressed in fewer than 2 cells. Applying these quality control criteria resulted in 15,341 cells and 20,656 genes.

Annotating cell-types to identify astrocytes in scRNA-seq data

To annotate cell types and identify astrocytes in the scRNA-seq datasets, we used the Discovery dataset as reference and applied the same Seurat mapping approach[82] similar to the cell-state annotation process mentioned above. To map the mouse data onto the reference, which is based on human transcriptomes, we first converted all mouse genes to human genes using the *convert_mouse_to_human_symbols* function from the nichenetr package. To annotate the major cell types, we normalized each dataset using *SCTransform* (with *variable.features.n* = 4000 and 1000 for the mouse data and iAst data, respectively). Transfer anchors were identified using the *FindTransferAnchors* function with the 10 nearest neighbors (*k.anchor* = 10) and the first 30 principal components (*npcs* = 30), while keeping the other parameters as default. To transfer the cell-type labels, the anchors were then input to the *TransferData* function, using the 10 nearest neighbors for weighting anchors (*k.weight* = 10) and *npcs* = 30. To identify the most astrocyte-like cells, we selected cells with an astrocyte label prediction score of ≥ 0.9 for the mouse data and ≥ 0.5 for the iAst data. We chose a more lenient threshold for the iAst data due to the smaller number of annotated astrocytes available. Together, the preprocessing and annotation pipelines result in 7276 astrocytes annotated with high confidence in the mouse data and 1434 astrocytes in the iAst data. These annotated astrocytes were then selected for down-stream analysis.

Association between Ast10 frequency and cognitive decline adjusting for neuropathologies

To evaluate the extent to which the association between Ast10 frequency and cognitive decline can be explained by other factors, we conducted a linear regression analysis. Cognitive slope was modeled as a function of Ast10 frequency alongside neuropathological and demographic covariates available in either the Discovery or Replication cohorts (n = 869 participants). Included covariates were age, sex, education, study cohort, and neuropathological measures: amyloid, tau, Lewy bodies, TAR DNA-binding protein 43 (TDP-43), gross chronic infarcts, chronic microinfarcts, cerebral atherosclerosis, arteriolosclerosis, cerebral amyloid angiopathy, hippocampal sclerosis, and neuron proportion estimated from snRNA-seq (see Methods: *Other neuropathological measures* for details). Participants with complete data across all variables (n = 700) were included in the final model. Continuous variables were z-score normalized, and categorical variables were dichotomized for model parsimony. Linear regression was performed using the *lm* function from the stats package (v4.4.2) in R.

Mediation analysis of SNAP25 in the association between Ast10 frequency and cognitive decline

SNAP25 protein levels were measured using Tandem Mass Tag (TMT) multiplexed mass spectrometry in a subset of participants (n = 273) from the CUIMC1, CUIMC2 and MIT datasets. TMT data were preprocessed and normalized as previously described[97]. To assess whether SNAP25 mediates the relationship between Ast10 frequency and cognitive decline, we performed a mediation analysis using three linear models, each representing an "edge" of the mediation triad:

1. Total effect: Association between Ast10 frequency and cognitive decline without adjusting for SNAP25.
2. Mediator model: Association between Ast10 frequency and SNAP25 protein levels.
3. Outcome model: Association between SNAP25 protein levels and cognitive decline, adjusting for Ast10 frequency.

All models were adjusted for sex, postmortem interval, dataset, age at death, education, cell count, and median gene count per participant. The *mediate* function from the mediation package in R was used (with default parameters) to estimate the proportion of the effect mediated by SNAP25.

Statistical information

All boxplots highlight the median, first and third quartiles, with whiskers extending to 1.5 times the interquartile range. Sample sizes are provided in the figure legend. The significance of DREMI scores for prioritized pathway genes was assessed empirically by comparison to other genes within the same category (i.e., ligand, receptor, signaling mediator, and transcriptional regulator). The MISTy importance score for the prioritized ligand set was evaluated using a permutation test. The effect of PLXNB1 knockout in astrocytes was tested using a two-sided Wilcoxon signed-rank test.

Author contributions

P.L.D. and N.C.-L. designed the study; Y.Z. prepared the single nucleus libraries and performed sequencing; M.F., K.C. performed the sequence alignment and demultiplexing analysis; N.C.-L. performed the computational- and statistical analyses, with the guidance from P.L.D., V.M., M.F., H.U.K.; J.Z., B.H., S.L. and M.W. generated isogenic pairs of *PLXNB1* KO hiPSCs-derived astrocytes and performed the single-cell RNA-seq, with the supervision from K.J.B., B.Z., H.Z., R.H.F., and A.L.; M.T, T.L. performed immunohistochemistry experiment and image analysis; M.T., H.U.K. generated the ST data; N.C.-L., H.U.K analyzed the ST data. N.C.-L. and P.L.D. wrote the manuscript, and all co-authors edited it for critical comments. D.A.B. is the PI of the parent ROS, and MAP studies and obtained funding and performed study supervision. J.S. is the PI of the parent LATC and AA-Core studies and obtained funding and performed study supervision. L.L.B is the PI of the parent study MARS and obtained funding and performed study supervision. All authors critically reviewed the manuscript. P.L.D. obtained funding for the project.

Data Availability

All data and analysis output are available via the AD Knowledge Portal (https://adknowledgeportal.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.org/DataAccess/Instructions). Data can be access from the Synapse database (https://www.synapse.org): for *CUIMC1* snRNA-seq data (syn31512863); for *MIT* snRNA-seq data (syn52293417); for AMP-AD 2.0 *Diversity* snRNA-seq (syn52339332); for *CUIMC2* snRNA-seq data (pending); for 10x Visium spatial transcriptomic data (syn62109935); for *PLXNB1*-KO mice single-cell RNA-seq data (syn26448163); for ROSMAP proteomic data (syn51150434); for *PLXNB1*-KO iAst single-cell RNA-seq data (pending); for immunofluorescence raw images (pending). RADC resources can be requested at https://www.radc.rush.edu.

Code Availability

All code used in this study are available on GitHub: https://github.com/cu-ctcn/Ast10_Communication

Declaration of conflict of interests

All authors have no relevant conflicts of interest.

Acknowledgements

We thank the individuals who have generously donated their brains to research through the Rush University Alzheimer's Disease Center. The five RADC cohorts are supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, U01AG61356, and R01AG22018. This work is supported by the National Institutes of Health's National Institute of Aging (NIH-NIA), through the following grants: Accelerating Medicines Partnership - Alzheimer's Disease Target Discovery and Preclinical Validation (AMP-AD, U01AG061356), the Alzheimer's Disease Sequencing Project (ADSP) Functional Genomics Consortium (FunGen-AD, U01AG072572), R21AG077168 (to M.W. and A.L.), RF1AG077828 (to R.H.F., H.Z., and M.W.), U01AG046170, RF1AG057440, R01AG068030, RF1AG074010 (to B.Z.). This work is also supported by the Alzheimer's Association, through the following grants: AARF-24-1313800 (to. N.C.-L.), AARG-22-928419 (to A.L. and M.W.), ADSF-21-816675 (to H.U.K. and M.K.) and by the New York State Department of Health DOH01-C38330GG and DOH01-C39068GG (to H.Z.). We also thank Drs. Zhihong Chen, Seunghee Kim-Schulze, and other members of the Human Immune Monitoring Center (HIMC) at Icahn School of Medicine at Mount Sinai for single-cell RNA sequencing.

Supplementary Tables

Supplementary Table S1 – Dataset demographics, Cell-state frequencies across studies, Meta-analysis results

Supplementary Table S2 – Astrocytic state signature genes, Pairwise differential gene expression results

Supplementary Table S3 – NicheNet results

Supplementary Table S4 – Ligand PLSR modeling input, performance and validation

Supplementary Table S5 – Receptor PLSR modeling input, performance and validation

Supplementary Table S6 – Prioritized Pathway and DREMI scores, MAGIC-imputed astrocyte expression

Supplementary Table S7 – Visium data demographics, Spatial correlation results per sample, MISTy results

Supplementary Table S8 – Raw CellProfiler analysis output

Supplementary Table S9 – MAGIC-imputed astrocyte expression of Ast10 signature in the mouse and iAst single-cell RNA-seq data.

Supplementary Table S10 – Data used in cognition model; Data used in SNAP25 mediation analysis

References

1. Green, G. S. *et al.* Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* 1–12 (2024) doi:10.1038/s41586-024-07871-6.
2. Cain, A. *et al.* Multicellular communities are perturbed in the aging human brain and Alzheimer's disease. *Nat Neurosci* **26**, 1267–1280 (2023).
3. 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).
4. Pandey, S. *et al.* Disease-associated oligodendrocyte responses across neurodegenerative diseases. *Cell Reports* **40**, (2022).
5. Sadick, J. S. *et al.* Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer's disease. *Neuron* **110**, 1788-1805.e10 (2022).
6. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).
7. Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).
8. Habib, N. *et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci* **23**, 701–706 (2020).
9. Tsartsalis, S. *et al.* A single nuclear transcriptomic characterisation of mechanisms responsible for impaired angiogenesis and blood-brain barrier function in Alzheimer's disease. *Nat Commun* **15**, 2243 (2024).
10. Luquez, T. *et al.* Cell type-specific changes identified by single-cell transcriptomics in Alzheimer's disease. *Genome Medicine* **14**, 136 (2022).
11. Tuddenham, J. F. *et al.* A cross-disease resource of living human microglia identifies disease-enriched subsets and tool compounds recapitulating microglial states. *Nat Neurosci* **27**, 2521–2537 (2024).
12. Self, W. K. & Holtzman, D. M. Emerging diagnostics and therapeutics for Alzheimer disease. *Nat Med* **29**, 2187–2199 (2023).
13. Kampmann, M. Molecular and cellular mechanisms of selective vulnerability in neurodegenerative diseases. *Nat. Rev. Neurosci.* **25**, 351–371 (2024).
14. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr Alzheimer Res* **9**, 628–645 (2012).
15. Bennett, D. A. *et al.* Overview and Findings from the Rush Memory and Aging Project. *Curr Alzheimer Res* **9**, 646–663 (2012).
16. Hüttenrauch, M. *et al.* Glycoprotein NMB: a novel Alzheimer's disease associated marker expressed in a subset of activated microglia. *acta neuropathol commun* **6**, 108 (2018).
17. L. Barnes, L., C. Shah, R., T. Aggarwal, N., A. Bennett, D. & A. Schneider, J. The Minority Aging Research Study: Ongoing Efforts to Obtain Brain Donation in African Americans without Dementia. *CAR* **9**, 734–745 (2012).

18. Schneider, J. A., Aggarwal, N. T., Barnes, L., Boyle, P. & Bennett, D. A. The Neuropathology of Older Persons with and Without Dementia from Community versus Clinic Cohorts. *Journal of Alzheimer's disease : JAD* **18**, 691 (2009).
19. Marquez, D. X. *et al.* Representation of Older Latinxs in Cohort Studies at the Rush Alzheimer's Disease Center. *Neuroepidemiology* **54**, 404–418 (2020).
20. Arvanitakis, Z., Bennett, D. A., Wilson, R. S. & Barnes, L. L. Diabetes and Cognitive Systems in Older Black and White Persons. *Alzheimer Disease & Associated Disorders* **24**, 37 (2010).
21. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J Alzheimers Dis* **64**, S161–S189 (2018).
22. Boyle, P. A. *et al.* To what degree is late life cognitive decline driven by age-related neuropathologies? *Brain* **144**, 2166–2175 (2021).
23. Wilson, R. S. *et al.* Education and cognitive reserve in old age. *Neurology* **92**, e1041–e1050 (2019).
24. Wilson, R. S., Wang, T., Yu, L., Bennett, D. A. & Boyle, P. A. Normative Cognitive Decline in Old Age. *Annals of Neurology* **87**, 816–829 (2020).
25. De Jager, P. L. *et al.* A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiology of Aging* **33**, 1017.e1-1017.e15 (2012).
26. Newell, K. L., Hyman, B. T., Growdon, J. H. & Hedley-Whyte, E. T. Application of the National Institute on Aging (NIA)-Reagan Institute Criteria for the Neuropathological Diagnosis of Alzheimer Disease. *Journal of Neuropathology & Experimental Neurology* **58**, 1147–1155 (1999).
27. Bennett, D. A., Schneider, J. A., Wilson, R. S., Bienias, J. L. & Arnold, S. E. Neurofibrillary Tangles Mediate the Association of Amyloid Load With Clinical Alzheimer Disease and Level of Cognitive Function. *Archives of Neurology* **61**, 378–384 (2004).
28. Wilson, R. S., Arnold, S. E., Schneider, J. A., Tang, Y. & Bennett, D. A. The relationship between cerebral Alzheimer's disease pathology and odour identification in old age. *Journal of Neurology, Neurosurgery & Psychiatry* **78**, 30–35 (2007).
29. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539–1558 (2002).
30. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**, 159–162 (2020).
31. Li, J. *et al.* Astrocyte-to-astrocyte contact and a positive feedback loop of growth factor signaling regulate astrocyte maturation. *Glia* **67**, 1571–1597 (2019).
32. Kovacs, D. M. α2-Macroglobulin in late-onset Alzheimer's disease. *Experimental Gerontology* **35**, 473–479 (2000).
33. Mocchegiani, E. & Malavolta, M. Zinc Dyshomeostasis, Ageing and Neurodegeneration: Implications of A2M and Inflammatory Gene Polymorphisms. *Journal of Alzheimer's Disease* **12**, 101–109 (2007).
34. Geladi, P. & Kowalski, B. R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **185**, 1–17 (1986).
35. Sirko, S. *et al.* Astrocyte reactivity after brain injury—: The role of galectins 1 and 3. *Glia* **63**, 2340–2361 (2015).
36. Krishnaswamy, S. *et al.* Conditional density-based analysis of T cell signaling in single-cell data. *Science* **346**, 1250689 (2014).
37. Kuchroo, M. *et al.* Single-cell analysis reveals inflammatory interactions driving macular degeneration. *Nat Commun* **14**, 2589 (2023).
38. Dijk, D. van *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e27 (2018).
39. Karaahmet, B. *et al.* Spatial Transcriptomic Analysis Identifies a *SERPINA3* -Expressing Astrocytic State Associated with the Human Neuritic Plaque Microenvironment. Preprint at https://doi.org/10.1101/2024.11.13.623438 (2024).

40. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
41. Tanevski, J., Flores, R. O. R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology* **23**, 97 (2022).
42. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* **21**, 811–819 (2018).
43. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* **7**, R100 (2006).
44. Hasel, P., Rose, I. V. L., Sadick, J. S., Kim, R. D. & Liddelow, S. A. Neuroinflammatory astrocyte subtypes in the mouse brain. *Nat Neurosci* **24**, 1475–1487 (2021).
45. Liddelow, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
46. Huang, Y. *et al.* Regulation of cell distancing in peri-plaque glial nets by Plexin-B1 affects glial activation and amyloid compaction in Alzheimer's disease. *Nat Neurosci* 1–16 (2024) doi:10.1038/s41593-024-01664-w.
47. Friedel, R. H. *et al.* Gene targeting using a promoterless gene trap vector ("targeted trapping") is an efficient method to mutate a large fraction of genes. *Proceedings of the National Academy of Sciences* **102**, 13188–13193 (2005).
48. Yu, L. *et al.* Residual decline in cognition after adjustment for common neuropathologic conditions. *Neuropsychology* **29**, 335–343 (2015).
49. Honer, W. G. *et al.* Cognitive reserve, presynaptic proteins and dementia in the elderly. *Transl Psychiatry* **2**, e114–e114 (2012).
50. Poirier, A. *et al.* PTPRS is a novel marker for early Tau pathology and synaptic integrity in Alzheimer's disease. *Sci Rep* **14**, 14718 (2024).
51. Bereczki, E. *et al.* Synaptic proteins predict cognitive decline in Alzheimer's disease and Lewy body dementia. *Alzheimer's & Dementia* **12**, 1149–1158 (2016).
52. Hasel, P., Aisenberg, W. H., Bennett, F. C. & Liddelow, S. A. Molecular and metabolic heterogeneity of astrocytes and microglia. *Cell Metabolism* **35**, 555–570 (2023).
53. Gazestani, V. *et al.* Early Alzheimer's disease pathology in human cortex involves transient cell states. *Cell* **186**, 4438-4453.e23 (2023).
54. İş, Ö. *et al.* Gliovascular transcriptional perturbations in Alzheimer's disease reveal molecular mechanisms of blood brain barrier dysfunction. *Nat Commun* **15**, 4758 (2024).
55. Mallach, A. *et al.* Microglia-astrocyte crosstalk in the amyloid plaque niche of an Alzheimer's disease mouse model, as revealed by spatial transcriptomics. *Cell Reports* **43**, 114216 (2024).
56. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991.e19 (2020).
57. Neuner, S. M. *et al.* Translational approaches to understanding resilience to Alzheimer's disease. *Trends in Neurosciences* **45**, 369–383 (2022).
58. Serrano-Pozo, A. *et al.* Astrocyte transcriptomic changes along the spatiotemporal progression of Alzheimer's disease. *Nat Neurosci* **27**, 2384–2400 (2024).
59. Armingol, E., Baghdassarian, H. M. & Lewis, N. E. The diversification of methods for studying cell–cell interactions and communication. *Nat Rev Genet* 1–20 (2024) doi:10.1038/s41576-023-00685-8.
60. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021).
61. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat Protoc* **15**, 1484–1506 (2020).
62. Clark, I. C. *et al.* Barcoded viral tracing of single-cell interactions in central nervous system inflammation. *Science* **372**, eabf1230 (2021).

63. Evans, E. E. *et al.* Semaphorin 4D is upregulated in neurons of diseased brains and triggers astrocyte reactivity. *Journal of Neuroinflammation* **19**, 200 (2022).
64. Tzioras, M., McGeachan, R. I., Durrant, C. S. & Spires-Jones, T. L. Synaptic degeneration in Alzheimer disease. *Nat Rev Neurol* **19**, 19–38 (2023).
65. Hall, C. M., Moeendarbary, E. & Sheridan, G. K. Mechanobiology of the brain in ageing and Alzheimer's disease. *European Journal of Neuroscience* **53**, 3851–3878 (2021).
66. Keough, M. B. *et al.* An inhibitor of chondroitin sulfate proteoglycan synthesis promotes central nervous system remyelination. *Nat Commun* **7**, 11312 (2016).
67. Huber, R. E., Babbitt, C. & Peyton, S. R. Heterogeneity of brain extracellular matrix and astrocyte activation. *Journal of Neuroscience Research* **102**, e25356 (2024).
68. Anwar, M. M., Özkan, E. & Gürsoy-Özdemir, Y. The role of extracellular matrix alterations in mediating astrocyte damage and pericyte dysfunction in Alzheimer's disease: A comprehensive review. *European Journal of Neuroscience* **56**, 5453–5475 (2022).
69. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
70. Miyata, S., Nishimura, Y. & Nakashima, T. Perineuronal nets protect against amyloid β-protein neurotoxicity in cultured cortical neurons. *Brain Research* **1150**, 200–206 (2007).
71. Schachtrup, C. *et al.* Fibrinogen Triggers Astrocyte Scar Formation by Promoting the Availability of Active TGF-β after Vascular Damage. *J. Neurosci.* **30**, 5843–5854 (2010).
72. Végh, M. J. *et al.* Hippocampal Extracellular Matrix Levels and Stochasticity in Synaptic Protein Expression Increase with Age and Are Associated with Age-dependent Cognitive Decline *. *Molecular & Cellular Proteomics* **13**, 2975–2985 (2014).
73. Végh, M. J. *et al.* Reducing hippocampal extracellular matrix reverses early memory deficits in a mouse model of Alzheimer's disease. *acta neuropathol commun* **2**, 76 (2014).
74. Klein, H.-U. *et al.* Characterization of mitochondrial DNA quantity and quality in the human aged and Alzheimer's disease brain. *Molecular Neurodegeneration* **16**, 75 (2021).
75. Nag, S. *et al.* TDP-43 pathology in anterior temporal pole cortex in aging and Alzheimer's disease. *Acta Neuropathologica Communications* **6**, 33 (2018).
76. Schneider, J. A. *et al.* Cognitive impairment, decline and fluctuations in older community-dwelling subjects with Lewy bodies. *Brain* **135**, 3005–3014 (2012).
77. Boyle, P. A. *et al.* Cerebral amyloid angiopathy and cognitive outcomes in community-based older persons. *Neurology* **85**, 1930–1936 (2015).
78. Arvanitakis, Z. *et al.* The Relationship of Cerebral Vessel Pathology to Brain Microinfarcts. *Brain Pathology* **27**, 77–85 (2017).
79. Buchman, A. S., Leurgans, S. E., Nag, S., Bennett, D. A. & Schneider, J. A. Cerebrovascular Disease Pathology and Parkinsonian Signs in Old Age. *Stroke* **42**, 3183–3189 (2011).
80. Arvanitakis, Z., Leurgans, S. E., Barnes, L. L., Bennett, D. A. & Schneider, J. A. Microinfarct Pathology, Dementia, and Cognitive Systems. *Stroke* **42**, 722–727 (2011).
81. Nag, S. *et al.* Hippocampal sclerosis and TDP-43 pathology in aging and Alzheimer disease. *Annals of Neurology* **77**, 942–952 (2015).
82. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
83. Schwarzer, G. meta: General Package for Meta-Analysis. 8.0–1 https://doi.org/10.32614/CRAN.package.meta (2006).
84. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10**, 101–129 (1954).
85. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130 (2001).
86. Pedregosa, F. *et al. Scikit-Learn: Machine Learning in Python*. http://arxiv.org/abs/1201.0490 (2018) doi:10.48550/arXiv.1201.0490.
87. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

88. Türei, D., Valdeolivas, A. & Saez-Rodriguez, J. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Molecular Systems Biology* **17**, e9923 (2021).
89. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252–263 (2009).
90. Holland, C. H., Szalai, B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194431 (2020).
91. L. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* **17**, 75 (2016).
92. Maynard, K. R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425–436 (2021).
93. Zhao, E. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat Biotechnol* **39**, 1375–1384 (2021).
94. Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).
95. Canals, I. *et al.* Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. *Nat Methods* **15**, 693–696 (2018).
96. Dobrindt, K. *et al.* Publicly Available hiPSC Lines with Extreme Polygenic Risk Scores for Modeling Schizophrenia. *Complex Psychiatry* **6**, 68–82 (2021).
97. Wingo, A. P. *et al.* Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. *Nat Neurosci* **23**, 696–700 (2020).
98. Schneider, J. A., Arvanitakis, Z., Bang, W. & Bennett, D. A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* **69**, 2197–2204 (2007).

Figures

Figure 1

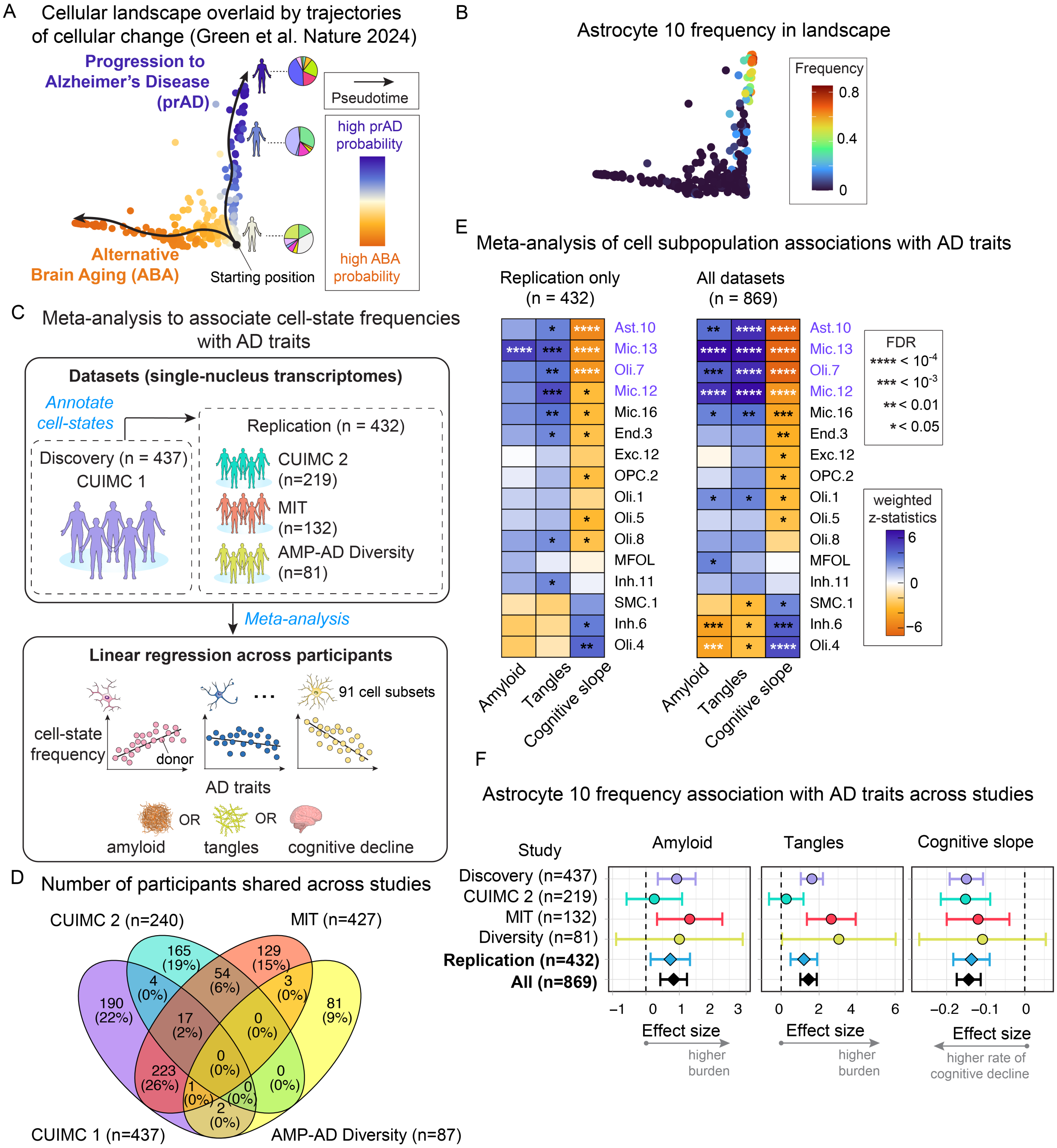


Figure 1 | Meta-analysis of DLPFC single-nucleus transcriptomic datasets reveals cell subpopulations associated with AD traits in 869 older individuals.

(A) Cellular landscape manifold of the aged neocortex, with each dot representing a Discovery dataset participant (n = 437) projected based on their 91 cell subpopulation composition[1]. Pseudotime analysis reveals two trajectories: Progression to AD (prAD) and Alternative Brain Aging (ABA)[1].

(B) Same manifold as in (A), except each participant is colored by their frequency of the Astrocyte 10 (Ast10) subpopulation, which is enriched along the prAD trajectory.

(C) Meta-analysis strategy for linking cell-state frequencies to three AD traits (Aβ, tau tangles, and cognitive decline). The Replication set includes three single-nucleus RNA sequencing (snRNAseq) datasets (*CUIMC2*, *MIT*[3], *Diversity*). Cells were annotated by their cell-state using the Discovery (*CUIMC1*) dataset[1] as a reference, and associations were tested via linear regression, followed by fixed-effect meta-analysis.

(D) Number of participants shared across studies. 869 non-overlapping samples were used for analyses (see Methods).

(E) Meta-analysis results for the Replication set (left) and all participants (Discovery + Replication, right). Heatmaps display effect sizes, ranging from negative (orange) to positive (purple). Positive effect sizes on amyloid and tangles correspond to higher pathology burdens, while negative effect size of cognitive slope corresponds to a higher rate of cognitive decline. Asterisks indicate significance: ****FDR $< 10^{-4}$, ***FDR $< 10^{-3}$, **FDR $< 10^{-2}$, *FDR < 0.05.

(F) Forest plots of Ast10 frequency associations with AD traits across individual studies and meta-analyses (bold), shown for Replication ($n = 432$) and all participants ($n = 869$). Dots represent effect sizes; lines indicate 95% confidence intervals.

Figure 2

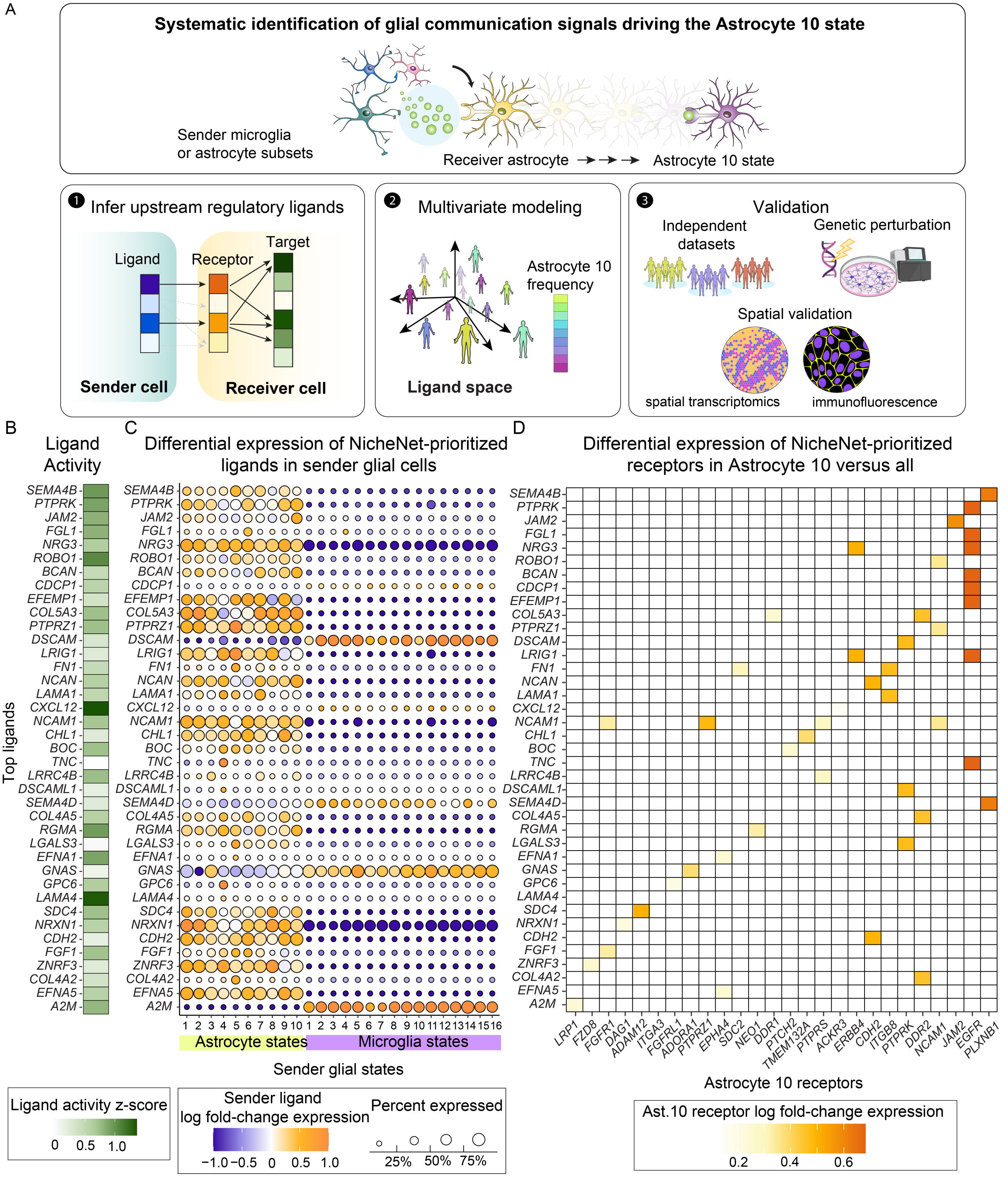


Figure 2 | Systematic identification of cell-cell communication signals regulating the $SLC38A2^{high}SMTN^{high}CACNA1D^{high}$ Astrocyte 10 state.

(A) Overview of our systematic approach to identify microglial and/or astrocytic cell-cell signals regulating Astrocyte 10 (Ast10). **(1)** NicheNet[30] prioritizes ligand-receptor pairs based on their expression and how well their downstream signaling activities recapitulating the Ast10 transcriptional signature. **(2)** Partial Least Squares Regression (PLSR) models predict Ast10 frequency per donor using expression patterns of prioritized ligands or receptors. **(3)** Validation includes replication in independent datasets, spatial transcriptomics to confirm ligand-Ast10 colocalization, immunohistochemistry for coexpression of an Ast10 marker with a top receptor, and genetic depletion of the top receptor in iPSC-derived and murine astrocytes, followed by scRNA-seq.

(B) Ligand activity z-scores from NicheNet for the top 100 sender-ligand-receptor interactions. A high z-score indicates that a ligand's predicted target genes are enriched for Ast10 signature genes. A positive z-score reflects above-average activity relative to all other ligands analyzed.

(C) Differential expression of the top ligands across all analyzed astrocytic and microglial sender states. Color indicates log fold-change (logFC) in ligand expression relative to other sender populations; circle size represents the percentage of cells expressing each ligand.

(D) Differential expression of the receptors for top-ranked ligands from (B). Color denotes logFC of receptor expression in Ast10 compared to other astrocytic and microglial subsets.

Figure 3

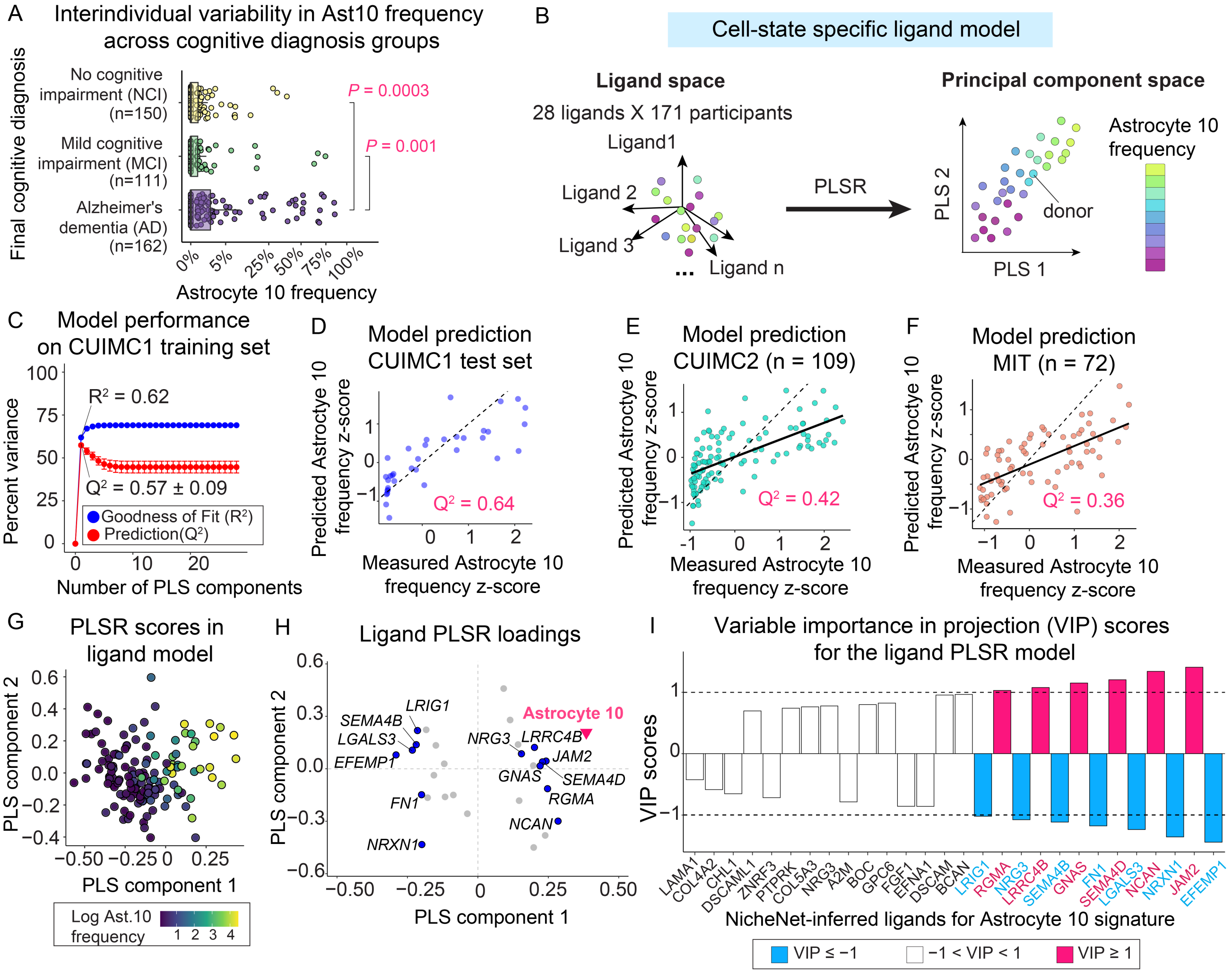


Figure 3 | Expression of prioritized ligands in specific glial states predict variation in Ast10 frequency across participants.

(A) Astrocyte 10 (Ast10) frequency varies across participants in the Discovery cohort, grouped by final consensus cognitive diagnosis: No cognitive impairment (NCI), mild cognitive impairment (MCI), and Alzheimer's dementia (AD). Clinical diagnoses were determined by neurology experts based on all available clinical data at the time of death[98]. *P*-values: two-sided Wilcoxon signed-rank test.

(B) Partial Least Squares Regression (PLSR) modeling predicts Ast10 frequency per donor based on prioritized ligand expression. The model input consists of 28 ligand summary scores, each summarizing ligand expression across specific glial states (see Methods). PLSR projects the high-dimensional ligand space into a principal component space, capturing covarying ligands that predict Ast10 variation. Representative principal components (PLS 1 and PLS 2) are shown.

(C) Ligand PLSR model performance, evaluated by variance explained (R^2) and variance predicted (Q^2) using five-fold cross-validation across 100 iterations, plotted against the number of PLS components.

(D) Predicted versus empirical (measured) Ast10 frequency z-scores in the *CUIMC1* test set, a 10% subset withheld from model training for independent validation.

(E- F) Predicted versus empirical Ast10 frequency z-scores in the *CUIMC2* **(E)** and *MIT* **(F)** snRNAseq datasets. Predictions were made using the ligand PLSR model trained on the *CUIMC1* Discovery dataset.

(G) PLSR scores (the first two PLS components) for each participant, colored by their log-transformed Ast10 frequency.

(H) PLSR loadings (the first two PLS components) for each ligand predictor and the response variable (Ast10 frequency). Top contributing ligands are highlighted.

(I) Variable Importance in Projection (VIP) scores from PLSR quantifying the importance of each ligand summary score in predicting Ast10 frequency. The sign of the VIP score indicates whether the ligand contributes positively or negatively to increasing Ast10 frequency. Ligands with |VIP| > 1 (indicating significant contribution) are highlighted.

Figure 4

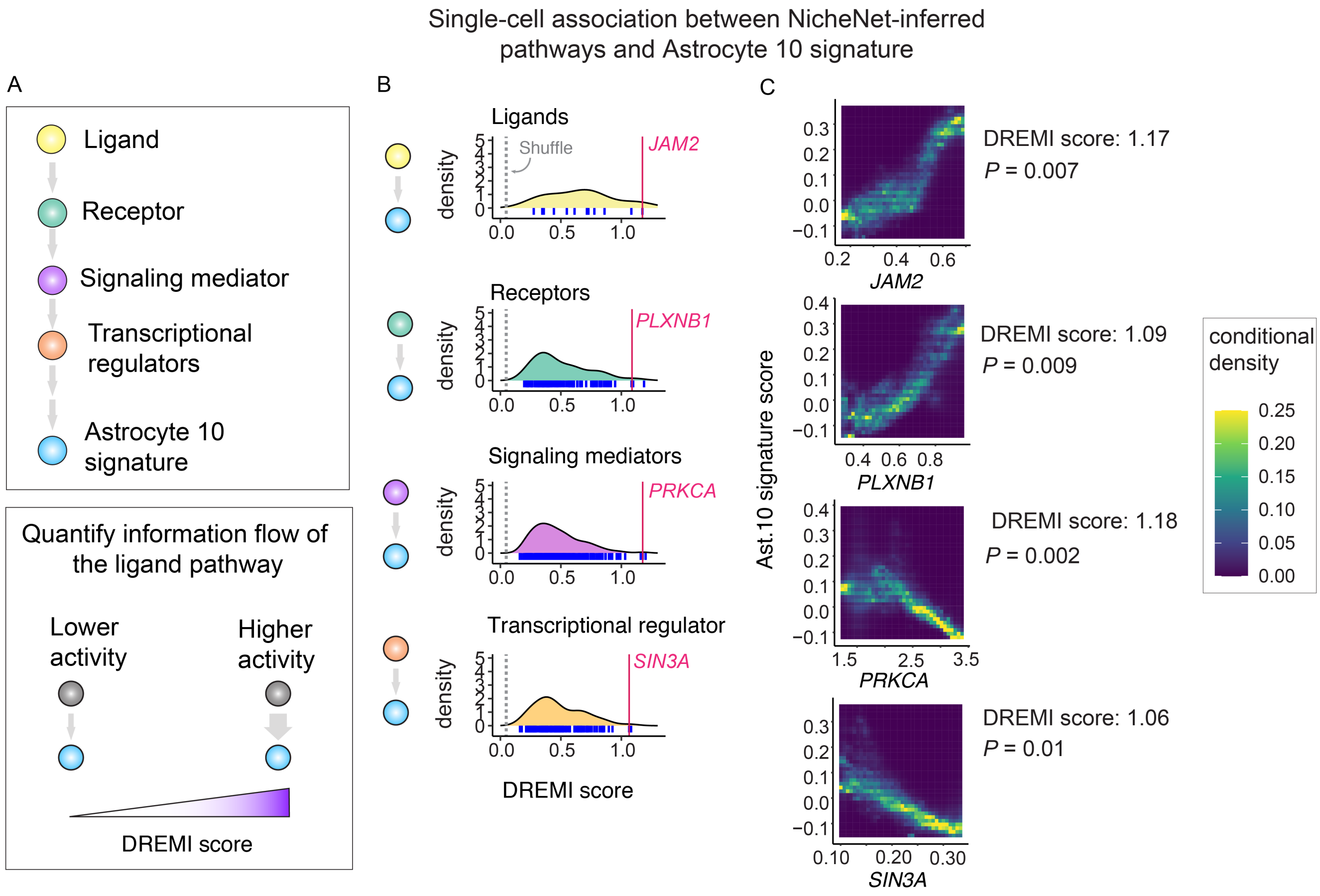


Figure 4 | Pathway genes of the prioritized ligands share significant mutual information with the Astrocyte 10 signature at the single-cell level.

(A) Schematic of the analysis quantifying the associations between prioritized ligand pathway genes and the Astrocyte 10 (Ast10) signature at the single-cell level. A simplified signaling cascade illustrates four signaling node categories: ligands, receptors, signaling mediators, and transcriptional regulators. To assess signal transfer activity, we applied Density Rescaled Mutual Information (DREMI)[36] to quantify the mutual information between each signaling node and the Ast10 signature score across individual astrocytes in the Discovery snRNAseq dataset. A higher DREMI score indicates stronger signal transfer activity from the pathway node to the Ast10 transcriptional signature.

(B) DREMI score distributions for each signaling node category: ligands, receptors, transcriptional regulators, and signaling mediators. Representative top-ranked pathway genes in each category are highlighted in red. Grey lines indicate the mean DREMI scores from shuffled data, serving as a baseline for comparison.

(C) Density Rescaled Visualization of Information (DREVI)[36] plots showing the relationship between representative pathway genes and the Ast10 signature score across individual astrocytes. Color intensity represents the conditional density distribution of astrocytes, illustrating how the Ast10 signature score (y-axis) varies with expression level of the pathway gene of interest (x-axis). DREMI scores and empirical *P*-values for each signaling node are displayed. *P*-values were derived from permutation tests across all possible genes in each category: ligands (n = 621), receptors (n = 657), signaling mediators (n = 11,124), and transcriptional regulators (n = 1,079) (see Methods for details).

Figure 5

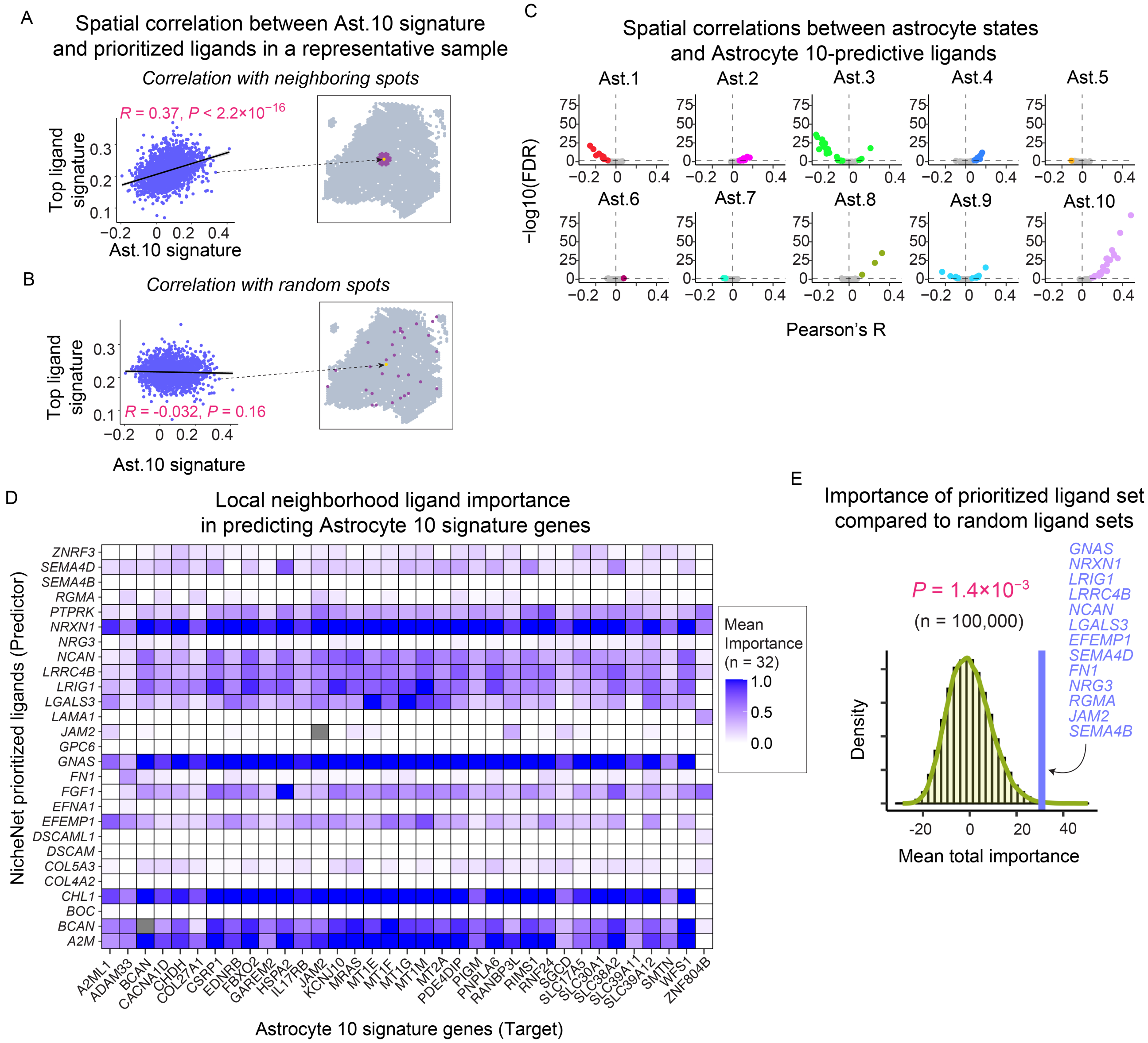


Figure 5 | Prioritized ligands predict nearby Ast10 signature but not other astrocytic states in spatial transcriptomic data

(A) Pearson's correlation between Ast10 signature scores in individual spatial transcriptomic spots and the mean expression of prioritized ligands in their local neighborhood. Each point represents a single spatial transcriptomic spot within a representative tissue slice. Inset: A spatially registered tissue slice highlighting an example spot of interest (yellow) and its neighboring spots (purple) used in the analysis.

(B) Control analysis for (A) using randomly selected spatial spots. Pearson's correlation between Ast10 signature scores in individual spots and the mean expression of prioritized ligands in *randomly selected* spots within the same tissue slice from (A). Inset: The same spatially registered tissue slice, highlighting the example spot of interest (yellow) and the randomly assigned spots (purple) used for comparison.

(C) Volcano plots showing Pearson's R and statistical significance for the spatial correlation between prioritized Ast10-predictive ligands and the astrocyte state signatures (Ast1–Ast10) across all samples. State signatures were derived from differentially upregulated marker genes, except for Ast1, where downregulated markers were used due to the absence of upregulated markers. To aid interpretation, Pearson's R values for Ast1 are inverted so that negative correlations indicate depletion of the Ast1 signature (see Methods). Each point represents a tissue sample; statistically significant correlations (FDR < 0.05) are highlighted.

(D) MISTy-derived importance scores quantifying the local influence of NicheNet-prioritized ligands (within a 3-spot radius) on Ast10 signature genes averaged across 32 spatial transcriptomic slides from 17 donors. Importance scores > 0 indicate significant influence. (see Methods for metric details).

(E) Total importance scores of PLSR-prioritized ligands (|VIP| ≥ 1) compared to randomly sampled ligand sets of the same size (see Methods). Empirical *P*-value calculated from 100,000 random samplings. Prioritized ligands are listed.

Figure 6

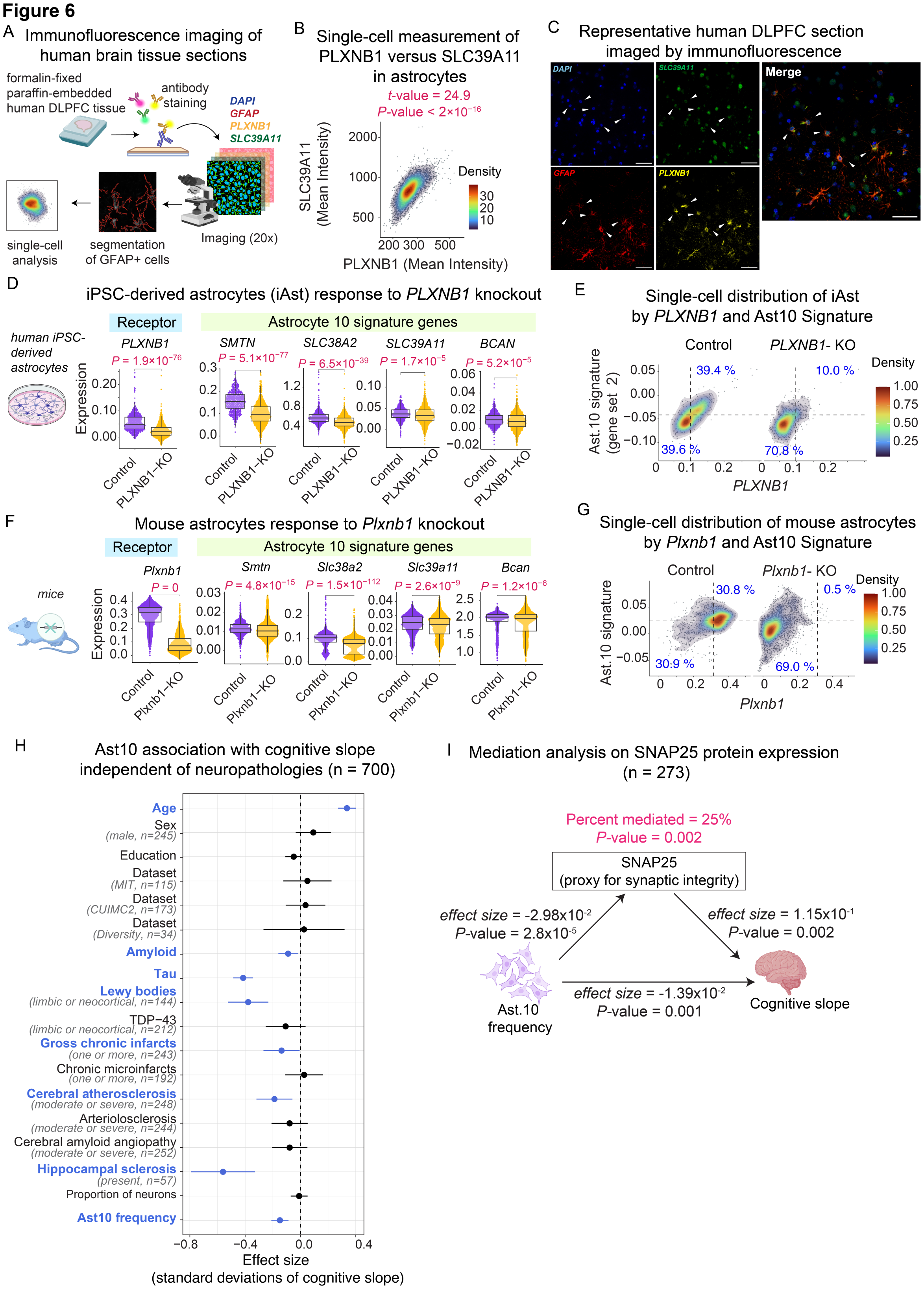


Figure 6 | Validation of the role of *PLXNB1* in Ast10 and mechanism of Ast10 action

Figure 6 | Validation of the role of *PLXNB1* in Ast10 and mechanism of Ast10 action

(A) Schematic of the immunofluorescence approach to validate PLXNB1 coexpression with the Ast10 marker SLC39A11 in human dorsal lateral prefrontal cortex (DLPFC) tissue from an AD donor. The FFPE section was stained for DAPI (nuclei), GFAP (astrocytes), SLC39A11 (Ast10 marker), and PLXNB1. Forty 20x images were systematically collected across cortical layers. Astrocytes were identified via DAPI and GFAP segmentation, and mixed-effect modeling assessed single-cell covariation between PLXNB1 and SLC39A11.

(B) Single-cell density plot showing SLC39A11 versus PLXNB1 expression, quantified by mean fluorescence intensity. Each point represents a GFAP+ astrocyte across 40 images. Effect size and significance for PLXNB1 predicting SLC39A11 levels were determined using a mixed-effect model (see Methods).

(C) Representative tissue section showing DAPI, GFAP, PLXNB1 and SLC39A11 staining. Astrocytes coexpressing high levels of PLXNB1 and Ast10 marker SLC39A11 ($\text{GFAP}^{\text{high}}\text{SLC39A11}^{\text{high}}\text{PLXNB1}^{\text{high}}$ cells) are marked by white arrows. Scale bars: 50 μm.

(D, F) *PLXNB1* and Ast10 marker expression measured in astrocytes from human iPSCs **(D)** and mice **(F)** with (PLXNB1-KO) or without *PLXNB1* knockout (Control). Two-sided Wilcoxon signed-rank test. Sample size: one isogeneic iAst pair (with or without PLXNB1-KO) derived from a single iPSC clone; one mouse for each genotype.

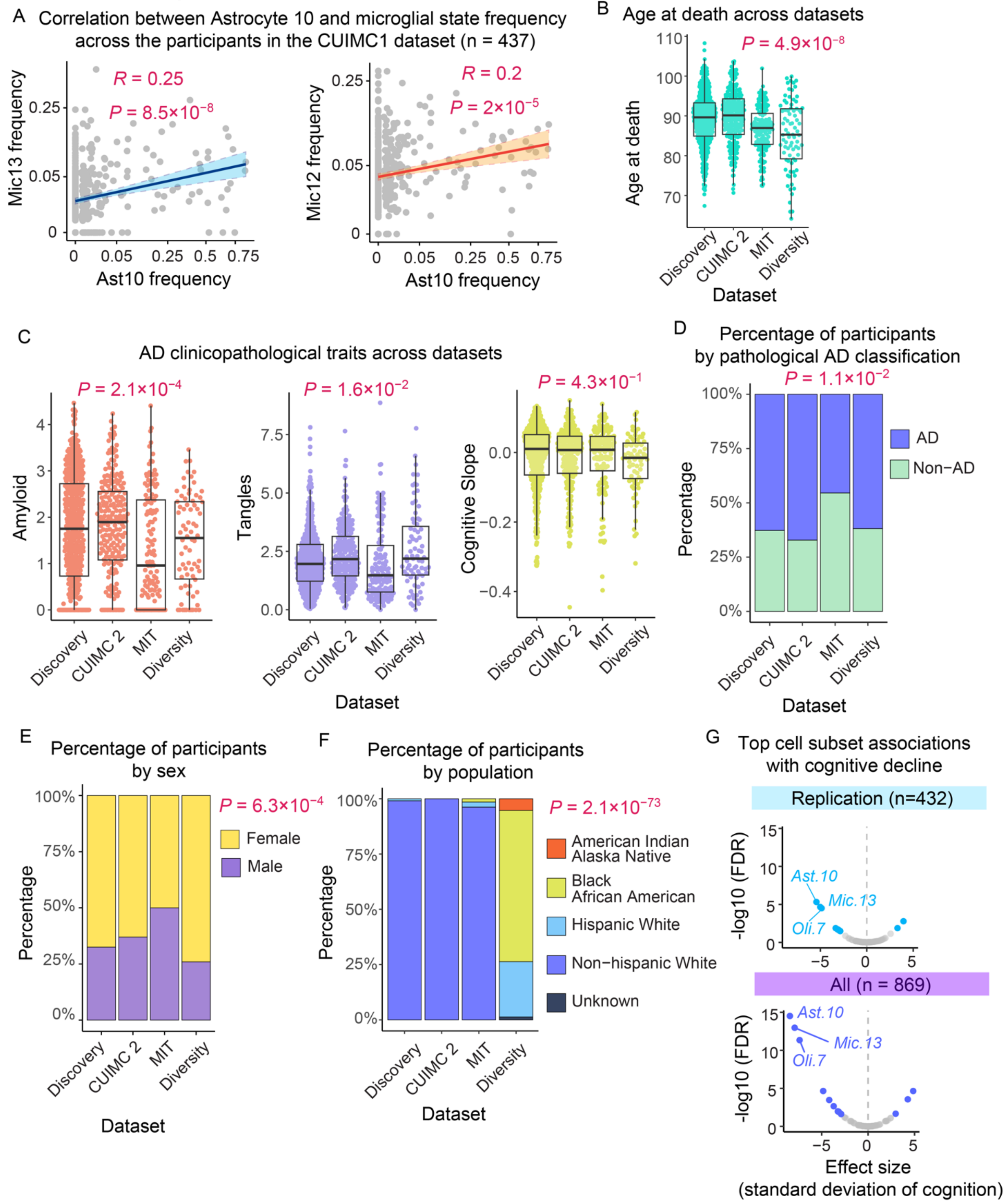
(E, G) Single-cell covariation between *PLXNB1* and the Ast10 signature score in Control and PLXNB1-KO astrocytes from iPSCs **(E)** and mice **(G)**. Dashed lines mark the median. Percentages of cells in each quadrant are indicated. Color scale represents cell density.

(H) Multivariate regression model predicting cognitive decline based on Ast10 frequency, neuropathologies, and demographics. Effect sizes (dots) and 95% confidence intervals (error bars) are shown; significant variables (blue) are highlighted. A total of n = 700 samples across all snRNAseq datasets (Discovery + Replication) with complete measurements of all variables were analyzed.

(I) Mediation analysis testing SNAP25 protein expression as a mediator of the Ast10–cognitive decline association in 273 ROSMAP DLPFC samples. SNAP25 significantly ($P = 0.002$) mediated 25% of the association.

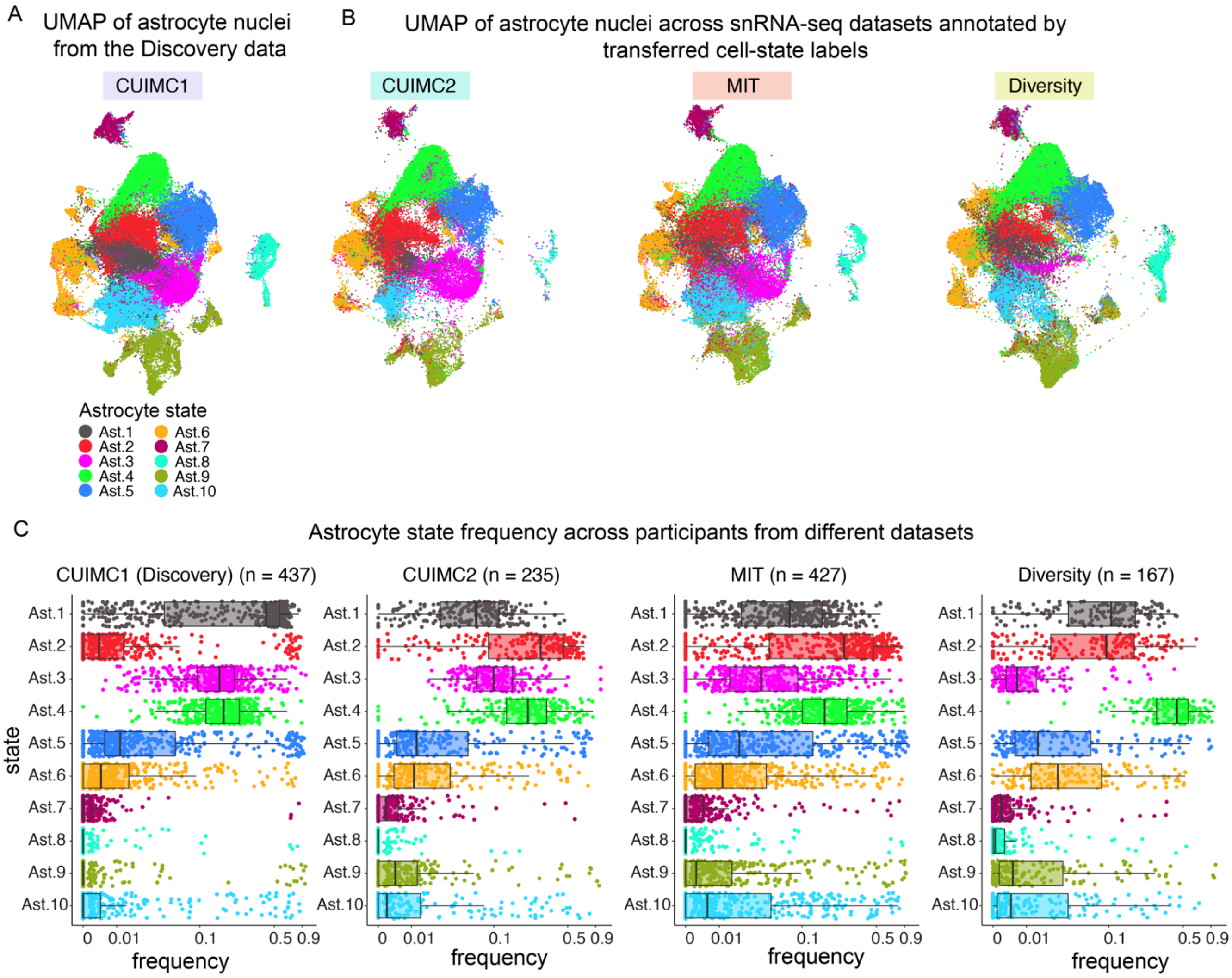
Extended Data

Extended Data Figure S1



Extended Data Figure 1 | Participant characteristics and meta-analysis of cell subset associations with cognitive decline.
(A) Pearson's correlation between the frequency of Astrocyte state 10 (Ast.10) and Microglia state 13 (Mic.13, left) or Microglia state 12 (Mic.12, right) across 437 participants in the CUIMC1 single-nucleus RNA sequencing (snRNA-seq) dataset[1].
(B–F) Clinicopathologic characteristics of the 869 participants across the four snRNA-seq datasets. Additional details are provided in Table 1.
(G) Effect sizes and statistical significance of associations between 91 cell subpopulations and cognitive decline, assessed via meta-analysis of the Replication set (Replication, n = 432) and all datasets combined (All, n = 869). Significant cell subpopulations (False Discovery Rate < 0.05) are highlighted, with the three most significant subpopulations indicated.

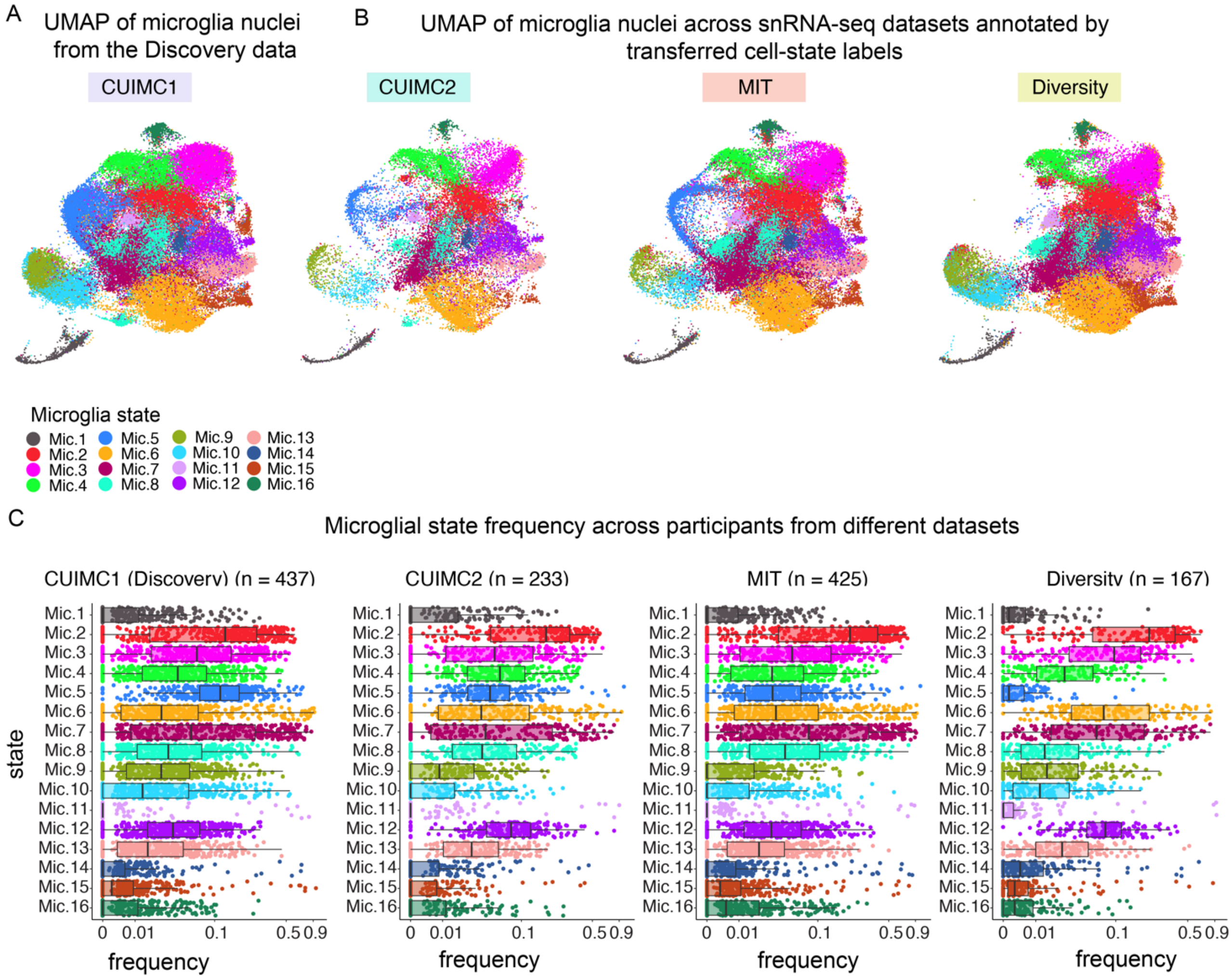
Extended Data Figure S2



Extended Data Figure 2 | Uniform astrocytic state taxonomy applied across four snRNA-seq datasets.

- **(A)** Uniform Manifold Approximation and Projection (UMAP) of individual astrocyte nuclei from the Discovery CUIMC1 dataset, colored by the astrocyte-state taxonomy defined in Green et al[1].
- (B) UMAP embeddings of individual astrocyte nuclei from each dataset in the Replication set, mapped onto the reference UMAP space of CUIMC1 (shown in A). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for details). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).
- (C) Frequency distributions of astrocytic states across participants in the four snRNA-seq datasets.

Extended Data Figure S3



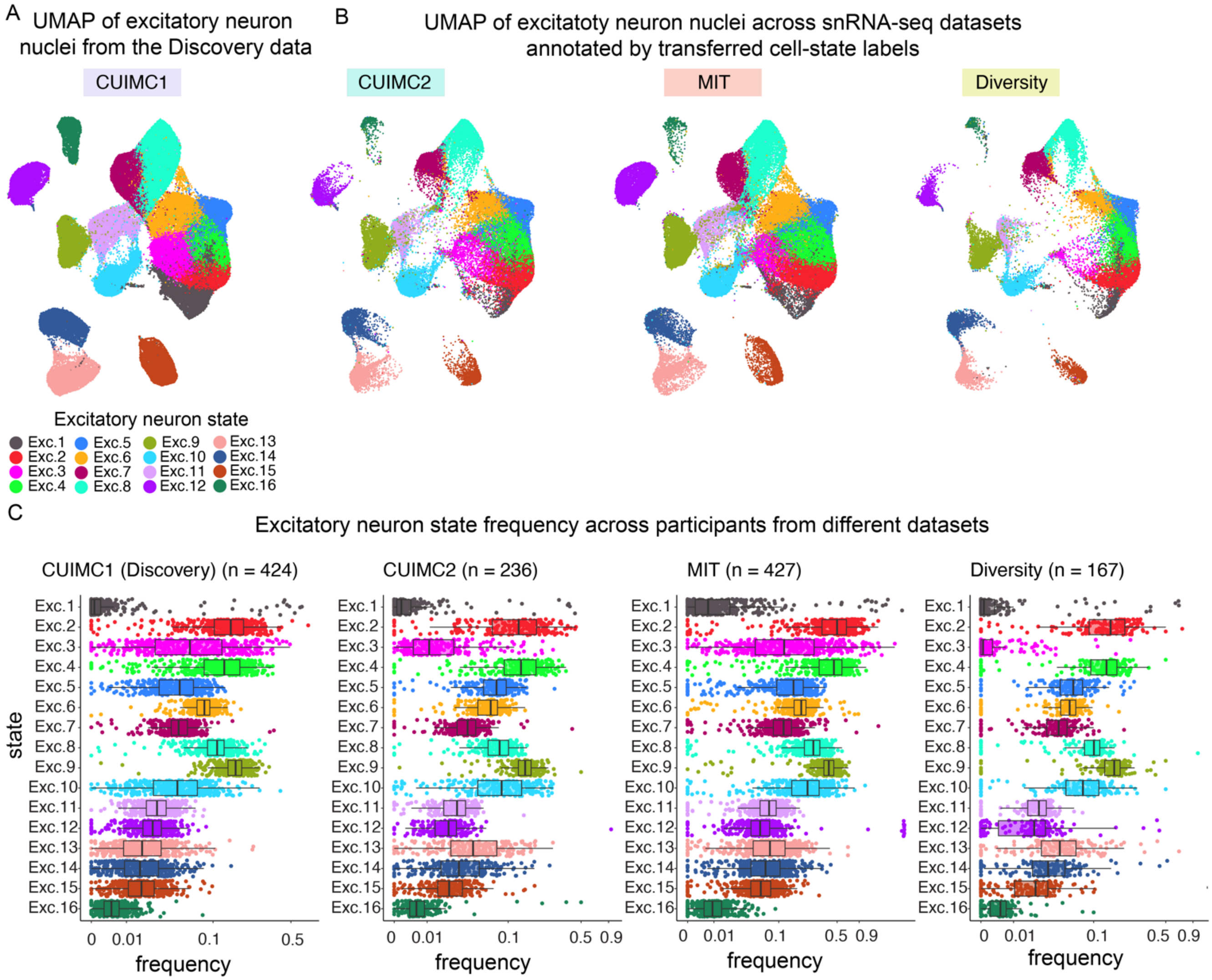
Extended Data Figure 3 | Uniform microglial state taxonomy applied across four snRNA-seq datasets.

(A) Uniform Manifold Approximation and Projection (UMAP) of individual microglia nuclei from the Discovery CUIMC1 dataset, colored by the microglial state taxonomy defined in Green et al[1].

(B) UMAP embeddings of individual microglia nuclei from each dataset in the Replication set, mapped onto the reference UMAP space of CUIMC1 (shown in A). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for detail). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).

(C) Frequency distributions of microglial states across participants in the four snRNA-seq datasets.

Extended Data Figure S4



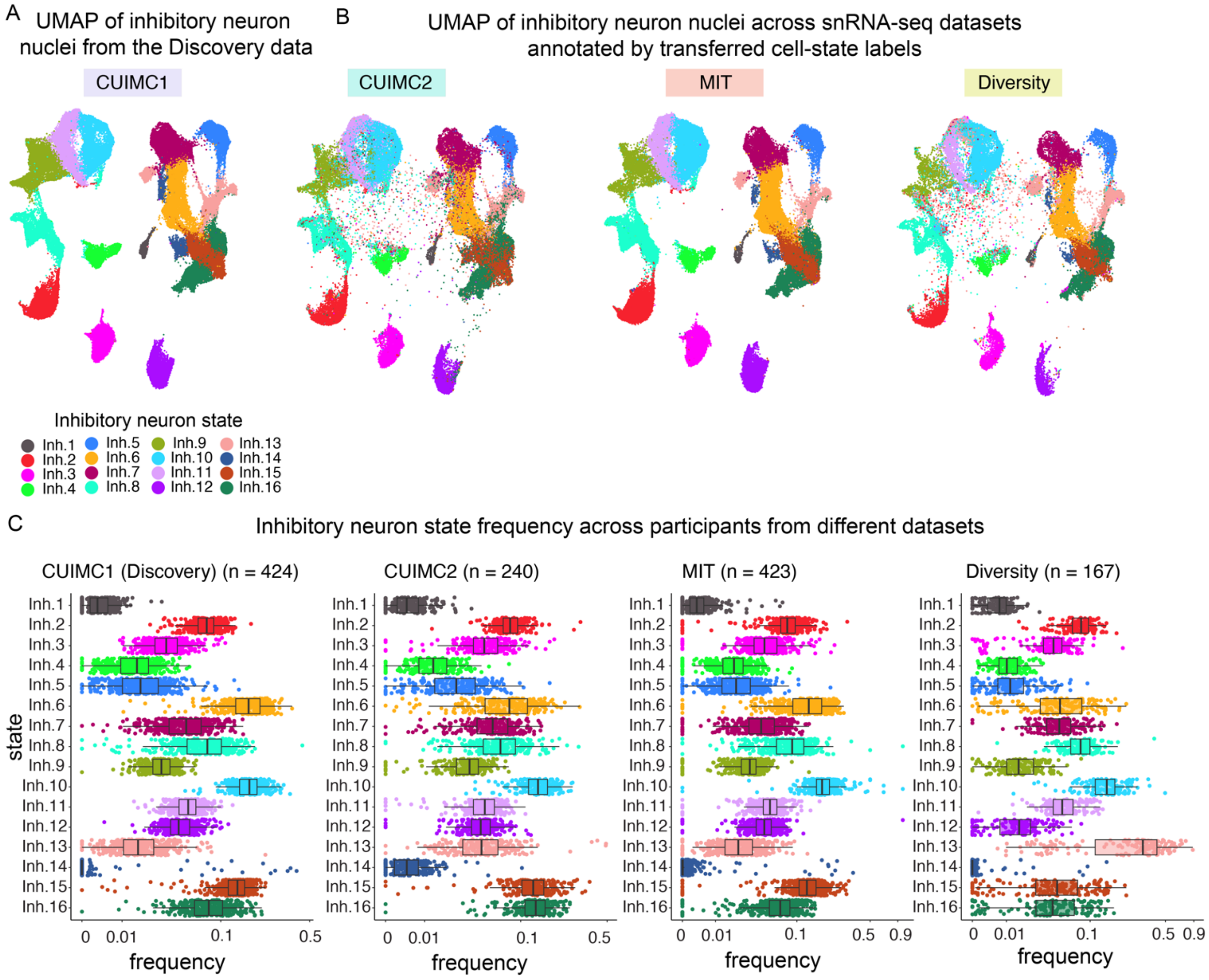
Extended Data Figure 4 | Uniform excitatory neuron cell-state taxonomy applied across four snRNA-seq datasets.

(A) Uniform Manifold Approximation and Projection (UMAP) of individual excitatory neuron nuclei from the Discovery CUIMC1 dataset, colored by the excitatory neuron cell-state taxonomy defined in Green et al[1].

(B) UMAP embeddings of individual excitatory neuron nuclei from each dataset in the Replication set, mapped onto the reference UMAP space of CUIMC1 (shown in A). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for details). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).

(C) Frequency distributions of excitatory neuron states across participants in the four snRNA-seq datasets.

Extended Data Figure S5



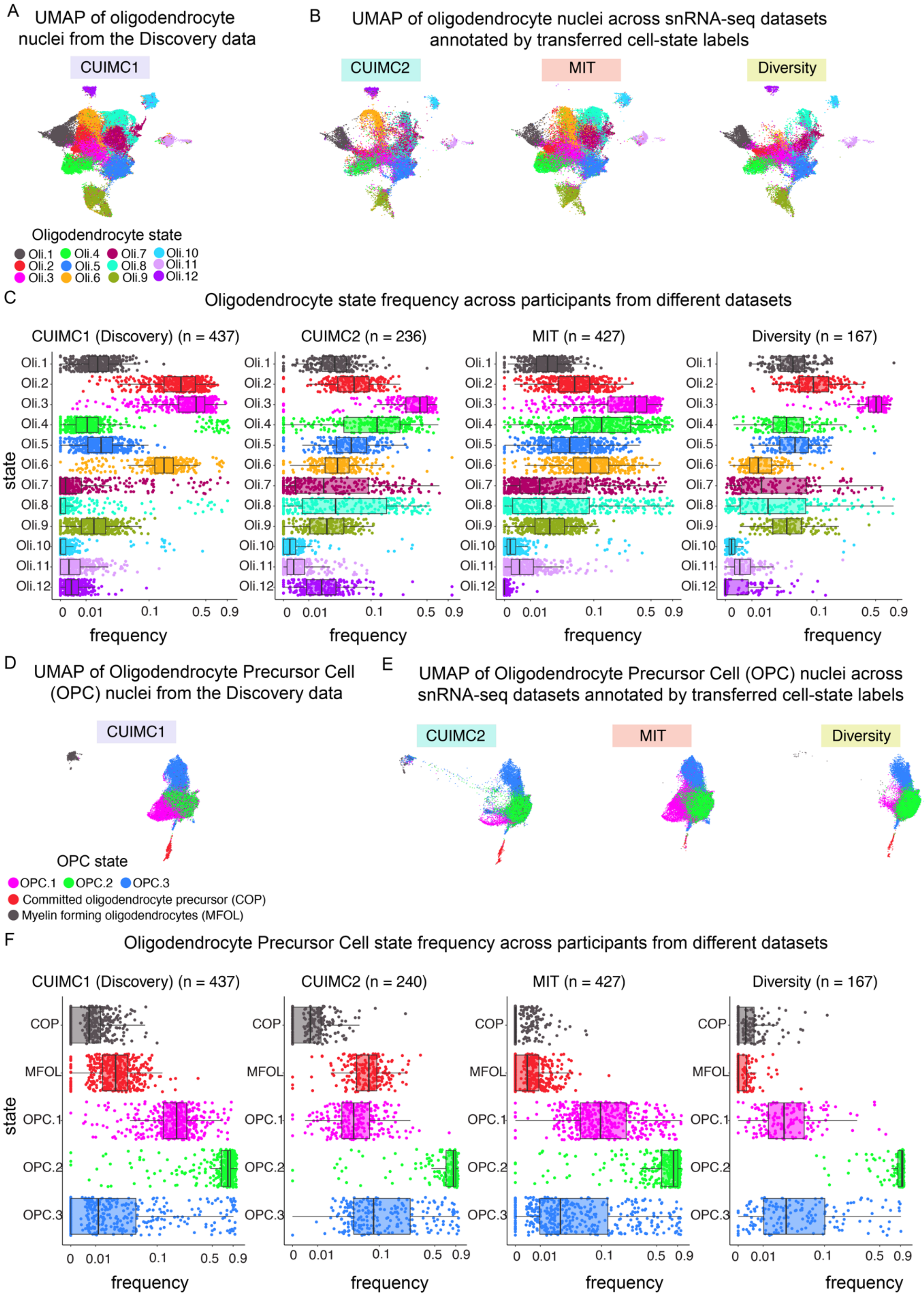
Extended Data Figure 5 | Uniform inhibitory neuron cell-state taxonomy applied across four snRNA-seq datasets.

(A) Uniform Manifold Approximation and Projection (UMAP) of individual inhibitory neuron nuclei from the Discovery CUIMC1 dataset, colored by the inhibitory neuron cell-state taxonomy defined in Green et al[1].

(B) UMAP embeddings of individual inhibitory neuron nuclei from each dataset in the Replication set, mapped onto the reference UMAP space of CUIMC1 (shown in A). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for details). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).

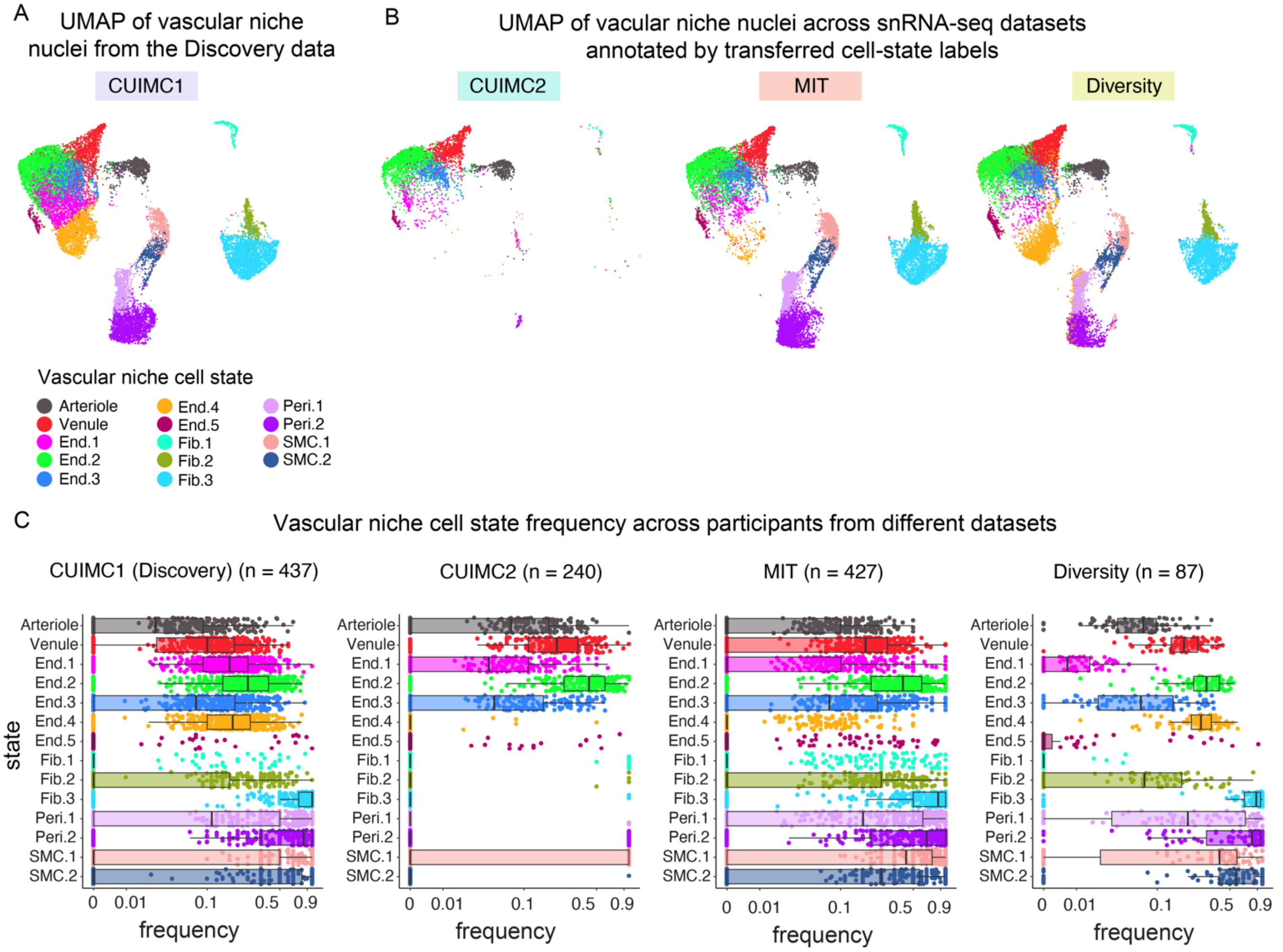
(C) Frequency distributions of inhibitory neuron states across participants in the four snRNA-seq datasets.

Extended Data Figure S6



Extended Data Figure 6 | Uniform oligodendrocyte and oligodendrocyte precursor cell-state taxonomy applied across four snRNA-seq datasets.
(A,D) Uniform Manifold Approximation and Projection (UMAP) of individual oligodendrocyte nuclei **(A)** or oligodendrocyte precursor cell (OPC) nuclei **(D)** from the Discovery CUIMC1 dataset, colored by the cell-state taxonomy defined in Green et al[1].
(B,E) UMAP embeddings of individual oligodendrocyte nuclei **(B)** or OPC nuclei **(E)** (from each dataset in the Replication set, mapped onto the reference UMAP space of the CUIMC1 (shown in A, D). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for details). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).
(C,F) Frequency distributions of oligodendrocyte states **(C)** or OPC states **(F)** across participants in the four snRNA-seq datasets.

Extended Data Figure S7

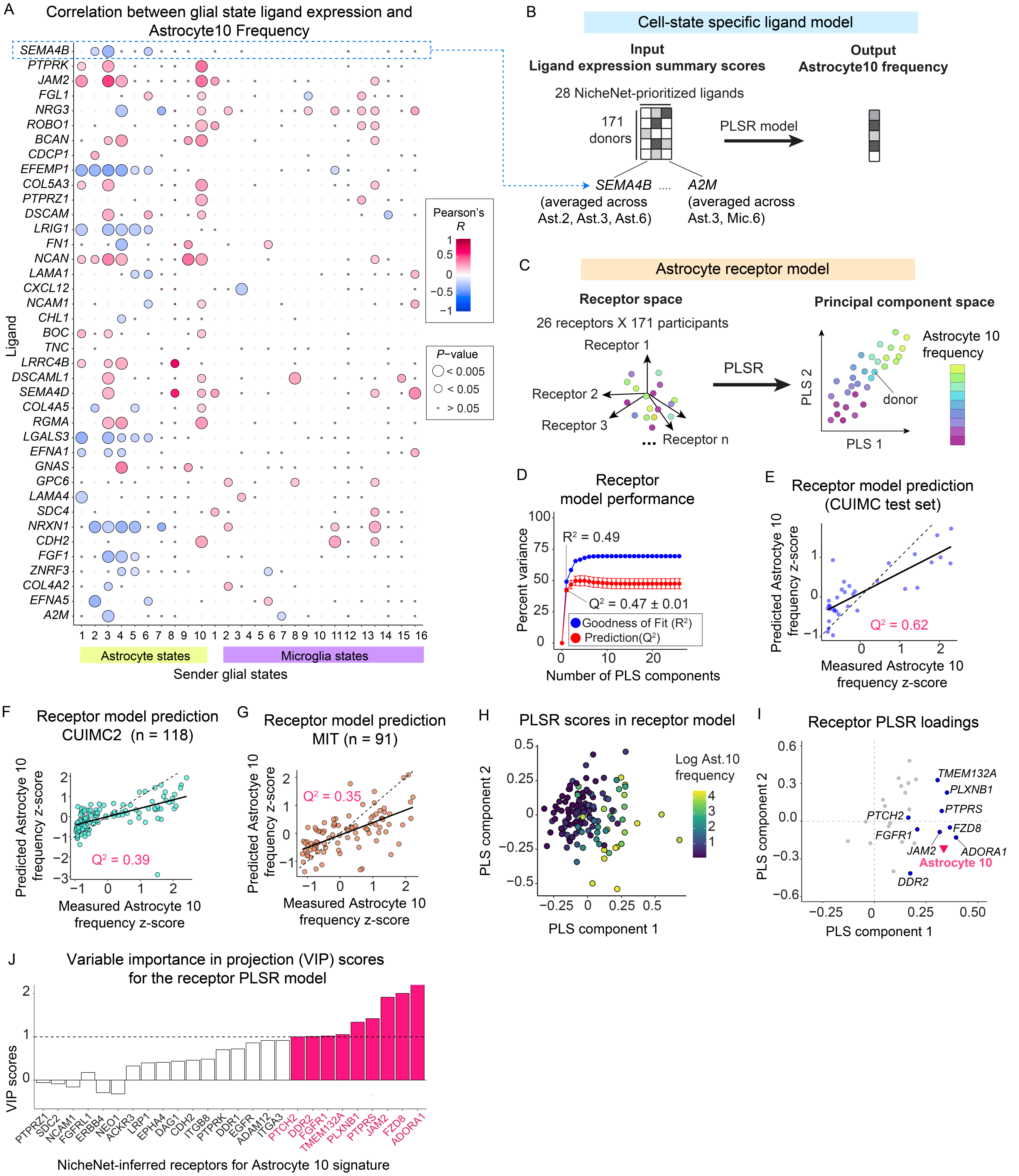


Extended Data Figure 7 | Uniform cell-state taxonomy applied to vascular cell subpopulations across four snRNA-seq datasets
(A) Uniform Manifold Approximation and Projection (UMAP) of individual nuclei from the vascular niche of the Discovery CUIMC1 dataset, colored by the cell-state taxonomy defined in Green et al[1].

(B) UMAP embeddings of individual vascular niche nuclei from each dataset in the Replication set, mapped onto the reference UMAP space of CUIMC1 (shown in A). Each cell is colored by the predicted cell-state label, assigned using CUIMC1 as the reference (See Methods for details). The Replication set includes three snRNA-seq datasets: Columbia University Irving Medical Center dataset 2 (CUIMC2), Massachusetts Institute of Technology dataset (MIT)[3] and the Accelerating Medicines Partnership for AD Diversity dataset (Diversity).

(C) Frequency distributions of vascular cell subsets across participants in the four snRNA-seq datasets.

Extended Data Figure S8



Extended Data Figure 8 | Partial Least Squares Regression modeling links ligands and receptors to Astrocyte state 10 frequency across participants.

(A) Pearson's correlations between Astrocyte state 10 (Ast10) frequency and NicheNet-prioritized ligands expressed across different sender astrocytic or microglial states. For each ligand, expression levels across significantly associated sender subpopulations (|Pearson's R| ≥ 0.1, P < 0.05) are averaged to generate a summary score, which serves as model input (see Methods for details). An example ligand, SEMA4B, is highlighted.

(B) Partial Least Squares Regression (PLSR) modeling approach based on cell-state-specific ligands. The model input consists of a matrix of ligand summary scores across donors in the Discovery dataset (n = 171 with Ast10), trained to predict each donor's Ast10 frequency (output). The calculation of the summary score for the example ligand SEMA4B (highlighted in A) is illustrated: expression of SEMA4B in significantly associated sender astrocyte states (Ast2, Ast3, Ast6) is averaged to derive the summary score.

(C) PLSR modeling predicts Ast10 frequency based on the prioritized receptors. For each donor, the input vector includes the pseudobulk astrocyte expression of 26 receptors. PLSR reduces the high-dimensional input receptor space to a principal component space, which captures covarying receptor patterns that predict Ast10 variation. Representative principal components (PLS 1 and PLS 2) are shown.

(D) Performance of the receptor-based PLSR model, evaluated by computing percentage of variance in Ast10 frequency explained (R^2) or predicted using repeated five-fold cross validation across 100 iterations (Q^2) with increasing number of PLS components.

(E) Comparison between the predicted Ast10 frequency z-scores from the receptor PLSR model versus the empirical (measured) Ast10 frequency z-scores in the CUIMC1 test set, which is 10% of the CUIMC1 data withheld from model training for independent validation.

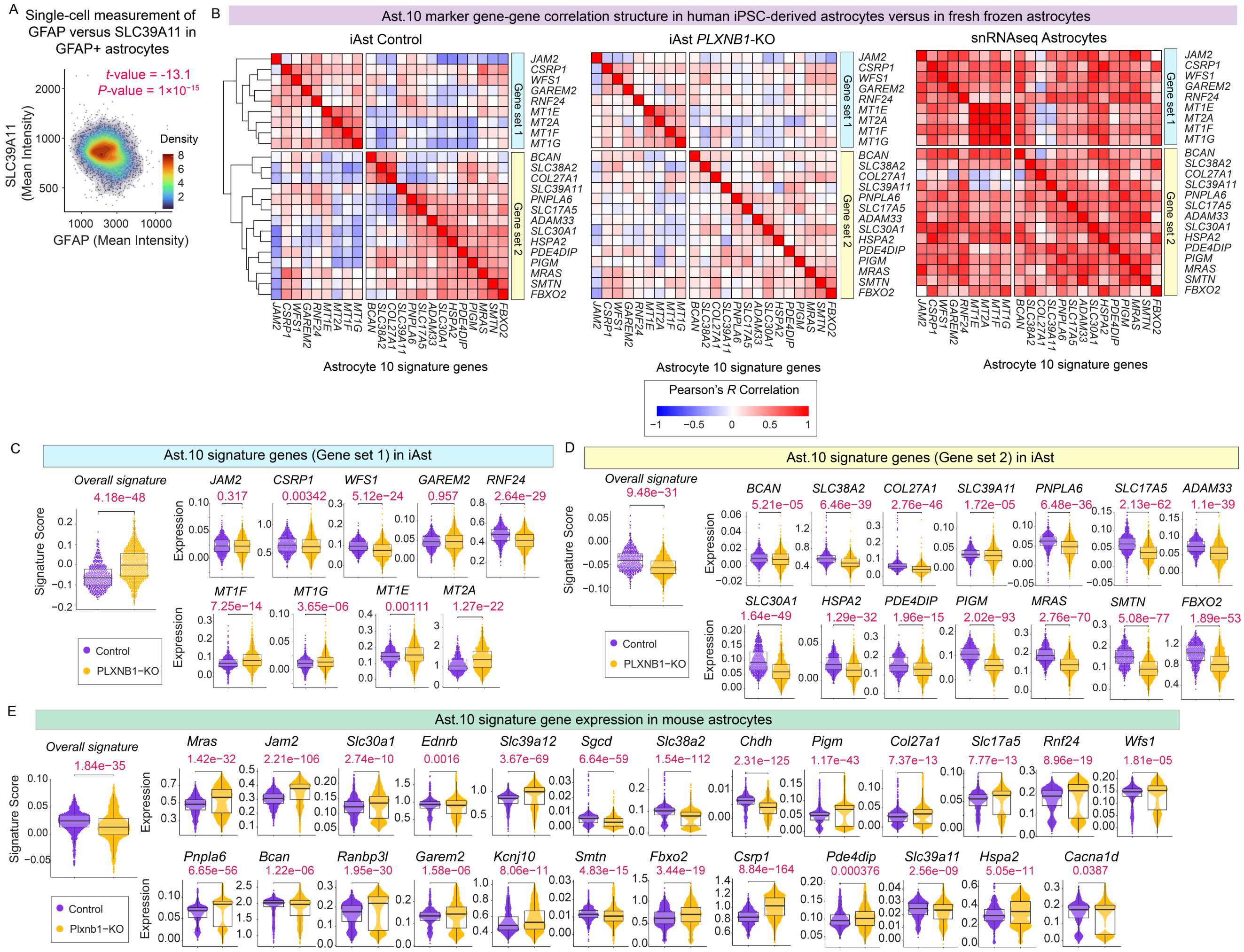
(F-G) Comparison between the predicted Ast10 frequency z-scores from the receptor PLSR model versus the empirical (measured) Ast10 frequency z-scores in the CUIMC2 dataset (F) and in the MIT snRNAseq dataset (G). Predictions were made using the PLSR trained on the CUIMC1 Discovery dataset.

(H) Receptor PLSR scores (of the first two PLS components) for each participant colored by their log-transformed Ast10 frequency.

(I) Receptor PLSR loadings (of the first two PLS components) for each ligand predictor and the response variable (Ast10 frequency). Top contributing receptors are highlighted.

(J) Variable Importance in Projection (VIP) scores from PLSR, quantifying each astrocytic receptor's contribution to predicting Ast10 frequency. The sign of the VIP score indicates whether the receptor positively or negatively influences Ast10 frequency prediction. Receptors with |VIP| ≥1, indicating significant contributions, are highlighted.

Extended Data Figure S9



Extended Data Figure 9 | Validation of the role of PLXNB1 on Ast10

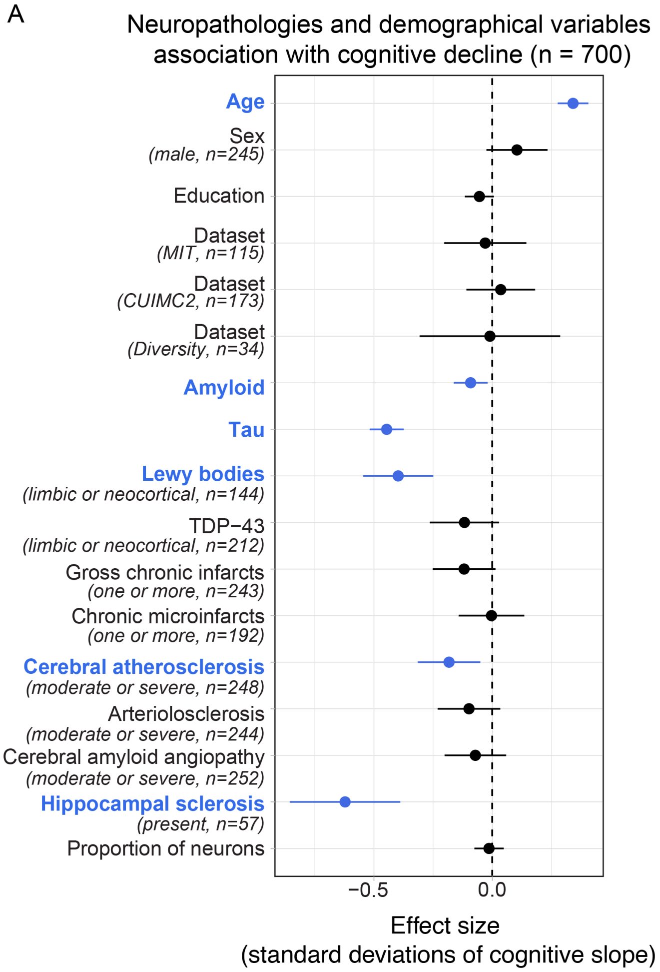
(A) Single-cell density plot showing the covariation between *SLC39A11* and *GFAP* expression, quantified based on mean fluorescence intensity measured from immunofluorescence imaging of an AD donor. Each data point represents an astrocyte identified by *GFAP*-based segmentation across 40 images. Statistics indicate the effect size and significance of *GFAP* expression for predicting *SLC39A11* levels, adjusting for *SLC39A11* expression and image variability, derived from fitting a mixed effect model (See Methods for details).

(B) Gene-gene correlation structure of Ast10 marker genes across three astrocyte systems: human induced pluripotent stem cell-derived (hiPSC) astrocytes at the baseline (iAst Control) or with *PLXNB1* knockout (iAst *PLXNB1*-KO), and fresh frozen astrocytes from the Discovery CUIMC1 dataset. For each gene pair, Pearson's correlation was computed across all astrocytes based on their MAGIC-imputed expression (see Methods for details). Genes were clustered using hierarchical clustering (with the "complete" agglomeration method) of the gene-gene correlation matrix from the iAst Control system.

(C-D) Comparison of single-cell distributions of Ast10 signature gene expression in iAst with and without *PLXNB1* knockout. Expression of Ast10 signature genes from Gene set 1 **(C)** and Gene set 2 **(D)** are shown, along with overall signature scores calculated for each gene sets (see Methods for details).

(E) Comparison of single-cell distributions of Ast10 signature gene expression in murine astrocytes with and without *PLXNB1* knockout. Expression levels of individual Ast10 signature genes and their summarized signature scores are shown.

Extended Data Figure S10



Extended Data Figure 10 | Variables associated with cognitive decline in older participants

(A) Forest plot from a multivariate regression model predicting cognitive decline based on pathologic variables and demographic variables (y-axis). Effect sizes (dots) and the respective 95% confidence intervals (error bars) are shown. Significant variables are highlighted in blue. All continuous variables were z-scored. Categorical variables were dichotomized; their respective factor levels and case numbers are indicated in brackets. A total of n = 700 samples across all available snRNAseq datasets (Discovery + Replication) with complete measurements for all variables were analyzed.