A meta-analysis of mouse models of tauopathy reveals the cellular underpinnings of selective regional vulnerability

Justin Torok[1], Pedro D. Maia[2], Chaitali Anand[3], and Ashish Raj[1*]

[1]University of California, San Francisco, Department of Radiology, San Francisco, CA, 94143, United States
[2]University of Texas at Arlington, Department of Mathematics, Arlington, TX, 76019, United States
[3]University of California, San Francisco, Institute for Neurodegenerative Diseases, San Francisco, CA, 94143, United States
[*]Address correspondence to ashish.raj@ucsf.edu

ABSTRACT

Neurodegenerative diseases such as Alzheimer's disease (AD) exhibit pathological changes in the brain that proceed in a stereotyped and regionally specific fashion, but the cellular and molecular underpinnings of regional vulnerability are currently poorly understood. Recent work has identified certain subpopulations of neurons in a few focal regions of interest, such as the entorhinal cortex, that are selectively vulnerable to tau pathology in AD. However, the cellular underpinnings of regional susceptibility to tau pathology have not been studied at a whole-brain level, primarily because whole-brain maps of a comprehensive collection of cell types have been inaccessible. Here, we deployed a recent cell-type mapping pipeline, Matrix Inversion and Subset Selection (MISS), to determine the brain-wide distributions of pan-hippocampal and neocortical neuronal and non-neuronal cells in the mouse using recently available single-cell RNA sequencing (scRNAseq) data. We then performed a meta-analysis of 5 previously published transgenic mouse studies that quantified regional tau in 12 distinct PS19 mouse models to identify general principles of cell-type-based selective vulnerability using these cell-type distributions. Using our methodology, we were able to identify cell types and cell-type classes that conferred vulnerability and resilience to tau pathology. Hippocampal glutamatergic neurons as a whole were strongly positively associated with regional tau deposition, suggesting vulnerability, while cortical glutamatergic and GABAergic neurons were negatively associated. Among glia, we identified oligodendrocytes as the single-most strongly negatively associated cell type, whereas microglia were consistently positively correlated. Using parsimonious linear models of cell-types predictors, we showed that these results generalize across studies and across time points within studies. To test whether AD-associated genes are better predictors of tau vulnerability, we compared our cell-type results to associations between the regional expression of GWAS-identified genes and experimental tau burdens. Remarkably, cell-types present a far stronger effect size of vulnerability compared to AD risk genes. To understand the functional enrichment patterns of the genes that were markers of the identified vulnerable or resilient cell types, we performed gene ontology analysis. We found that the genes that are directly correlated to tau pathology are functionally distinct from those that constitutively embody the vulnerable cells. In short, we have demonstrated that regional cell-type composition is a compelling explanation for the selective vulnerability observed in tauopathic diseases at a whole-brain level, which has important implications for better understanding AD pathophysiology.

1 Introduction

A hallmark of all neurodegenerative diseases, including Alzheimer's disease (AD), frontotemporal lobar dementia (FTLD), and other tauopathies, is the display of spatially heterogeneous and stereotyped distributions of brain pathology. In AD, accumulations of misfolded microtuble-associated protein tau (τ), which are strongly associated with and precede neurodegeneration, are first observed in the locus coeruleus (LC) and then entorhinal cortex (EC), followed by orderly spread into limbic and temporal areas, the basal forebrain, and finally to neocortical association areas[1]. Despite this stereotyped progression of brain pathology, the questions of how and why pathology selectively targets specific regions while sparing others remains an open question.

A key observation partly explaining the pathological staging of pathology is that certain neuronal subtypes in affected regions harbor significantly more τ inclusions and degenerate at a faster rate than others. Great progress in exploring this phenomenon, known as *selective neuronal vulnerability*[2], has been facilitated by rapid advancements in single-cell sequencing technologies, allowing for the identification of key subpopulations of vulnerable neurons[3–7]. For instance, emerging evidence suggests that neurons in the noradrenergic nuclei might be instrumental in initiating tau pathology[8]. Within the EC, τ-inclusions are predominantly observed in *RORB*-expressing excitatory neurons, while other subpopulations appear relatively unaffected[7]. In the hippocampus, γ-aminobutyric-acid-secreting (GABAergic) inhibitory neurons, especially the *Pvalb*+ and

Sst+ interneurons, tend to colocalize with τ[9]. More generally, myriad of factors, from cytoarchitecture and potential exposure to external pathogens[10], to morphological attributes[11–16], are postulated to influence a cell's vulnerability or resilience to AD's pathological processes.

The role of non-neuronal cells in the disease process is also gaining wider attention. Oligodendrocytes, for instance, have been observed to sequester extracellular τ, potentially leading to its internal amplification and subsequent dissemination[17,18]. Microglia and the broader process of neuroinflammation are widely acknowledged as pivotal mediators of both τ and amyloid-β (Aβ) pathophysiology[19–27]. Astrogliosis, characterized by the proliferation of astrocytes in response to cellular damage, is also a significant focus in AD research[28–30].

This study seeks to gain insights into the foundations of *cell-type-based selective vulnerability* (SV-C) and *resilience* (SR-C) to τ at a whole-brain level from a statistical perspective. Most current evidence on SV-C and SR-C has understandably come from descriptive, hypothesis-driven mechanistic experimental bench or animal studies, and limited to selected brain structures or selected cell types. These studies, based on a chosen transgenic animal model or specific tau strain, are difficult to generalize. For instance, the roles of support cells such as microglia and astrocytes are complex, and whether they are helpful or harmful in tauopathy remains inconclusive and may be context-dependent[31,32]. We propose that a fuller assessment of broad phenomena like selective vulnerability would benefit from a principled integration of a diversity of data sources. We therefore employed a meta-analytical approach that integrated various histological τ data sourced from 5 distinct studies and encompassing a total of 12 experimental conditions, all of which used PS19 mouse models of tauopathy[33–37].

To derive the quantitative cell-type maps required to perform our analyses, we utilized the recently-described Matrix Inversion and Subset Selection (MISS) algorithm[38] to infer the distributions of 42 neuronal and non-neuronal cell types, utilizing the recent single-cell atlas developed by the Allen Institute for Brain Science (AIBS), which profiled approximately 1.3 million cells sampled across the hippocampal formation and neocortex[39]. This computational technique enabled us to significantly expand the number of covered cell types and their corresponding spatial coverage beyond those for cell types obtained by current *in situ* sequencing methods[40–44].

Our analysis revealed that, among the types of neurons present, hippocampal glutamatergic neurons were the only class that was consistently positively associated with τ pathology. Conversely, oligodendrocytes emerged as the cell type most negatively associated with pathology, suggesting a potential role in conferring resilience to τ at a whole-brain level. Furthermore, we found that combinations of cell types generally yielded more predictive multivariate models of τ pathology when compared with equivalent models utilizing the expression of 24 key AD risk genes[45,46]. A further exploration of the transcriptomic underpinnings of SV-C and SR-C revealed that there were surprising differences between the gene sets most strongly correlated with τ and those that were differentially expressed in vulnerable and resilient cell types. These differences were evident both in terms of the identified genes and their functional enrichment at the biological process level. Overall, our integrative modeling approach offers a complementary framework to interpret the extensive pathology and -omics data generated by bench scientists, yielding generalized insights into selective vulnerability in tauopathies.

2 Results

2.1 Inferring new brain-wide maps for neuronal and non-neuronal cell-types

We used the recently published Matrix Inversion and Subset Selection (MISS) algorithm to infer whole-brain cell-type distributions[38]. In brief, MISS determines distributions of cell types, as characterized by their single-cell RNA sequencing (scRNAseq) gene expression signatures, by performing a deconvolution on spatial gene expression data. We have previously demonstrated that MISS is an accurate, efficient, and robust tool for determining the whole-brain distributions of mouse neural cell types from two independently sampled scRNAseq datasets[47–49], where we used the coronal Allen Gene Expression Atlas (AGEA) as our source of spatially resolved gene expression data[38,50].

In the present study, we again leveraged the AGEA, but instead mapped the cell types recently characterized by Yao, *et al.*[39] (**Figure 1A**). We chose to produce new maps using this scRNAseq dataset (herein referred to as the "Yao, *et al.* dataset") rather than rely on previously published cell-type distributions for two reasons: 1) Yao, *et al.* comprehensively sampled the neocortex and hippocampal formation, sequencing approximately 1.3 million individual cells with excellent read depth[39]; and 2) capturing cell-type-specific vulnerability to tauopathy requires robust and specific delineation of cortical and hippocampal cell types. Following our established procedure, we first obtained an informative 1300-gene subset optimal for the spatial deconvolution problem, and then solved the nonnegative matrix inversion problem voxel-by-voxel (**Figure 1B**; see also **Methods**). The residual error at both the regional and voxel-wise levels was not greater outside of the neocortex and hippocampus, demonstrating that these whole-brain maps did not contain systematic errors outside of the regions where the mapped cell types were sampled (**Figure 1B**). We also validated our results using published distributions of *Pvalb*+, *Sst*+, and *Vip*+ interneurons in the neocortex and show excellent agreement with our inferred maps (**Figure 1B**).

In all, we mapped 36 neuronal and 6 non-neuronal cell types at the 200-μm resolution of the AGEA (**Figure 1C**). We further classified these cell types into four major classes: cortical glutamatergic neurons, hippocampal glutamatergic neurons,

GABAergic neurons, and non-neuronal cells (**Tables S1** and **S2** for complete descriptions of these cell types). The spatial structure of the resulting cell-type distributions was rich and diverse (**Figure 1D**), and in fact cell types and cell-type classes (e.g., non-neuronal cells) have more varied regional densities than gene expression profiles (**Figures S1-S5**). We emphasize that both the scRNAseq and spatial gene expression data were taken from healthy, adult, wild-type mice, and thus the MISS-inferred cell-type distributions represent *baseline* cell-type densities.

2.2 Tauopathy dataset curation

We assembled a set of 12 regional tauopathy datasets, each of which represents a unique experimental condition. Here, we define an experimental condition in terms of the genetic background of the mice used, the type of injectate used, and the injection site (see **Table S3** for brief descriptions of these datasets)[33–37]. We then coregistered the inferred MISS cell-type maps, which were parcellated using the Common Coordinate Framework version 2 (CCFv2)[51], into the region spaces of the individual datasets and performed a combination of univariate and multivariate analyses. This meta-analysis of mouse tauopathy allows us to answer broad questions about cell-type selective vulnerability at a whole-brain level.

2.3 Identifying vulnerable and resilient cell types

Figures 2A and **S6** depict the Pearson's correlations (R values) between the spatial distributions of cell types and the patterns of end-timepoint τ deposition across the 12 experimental conditions. Here, a positive R value indicates that a cell type may be more vulnerable to τ pathology, while a negative value suggests that it may confer resilience. In general, τ pathology showed a positive correlation with hippocampal glutamatergic neurons ($p = 4.3 \times 10^{-10}$) and negative correlations with both cortical glutamatergic and GABAergic neurons ($p = 3.3 \times 10^{-6}$ and $p = 4.3 \times 10^{-3}$, respectively). The hippocampal glutamatergic cell type CA1-ProS stood out as the single-most vulnerable cell type (Mean $\mathrm{R} = 0.32$), which is entirely confined to the hippocampal formation and in particular CA1 (**Figure 2C**). Non-neuronal cells, by contrast, did not exhibit a significant correlation with τ pathology overall (**Figure 2B**; **Table S4**). However, while non-neuronal cell types generally lacked a significant correlation to τ pathology (**Figure 2B**), certain types displayed either vulnerability or resilience. We identified oligodendrocytes (Oligo) to be the most resilient cell type, with a mean R of -0.36 (**Figure 2C**). Astrocytes (Astro) typically presented negative correlations with τ pathology, albeit less strongly than oligodendrocytes. By contrast, the immune cell subtype comprising microglia and perivascular macrophages (Micro-PVM) exhibited a strong positive correlation. The significance of these correlations, especially considering the known roles of these glial cells in disease progression, is further elaborated in the **Discussion**.

2.4 Neuronal vulnerability is modulated by transcriptomic and spatial proximity to CA1-ProS neurons

In light of the wide disparities between cell types and cell-type classes in terms of their correlations to τ pathology shown in **Figures 2A** and **2B**, and in partricular among neuronal subtypes, we explored the more granular transcriptomic and spatial differences between cell types. We utilized a correlation dissimilarity measure, defined as $1 - \mathrm{R}$, to gauge each neuronal type's transcriptomic and spatial divergence from the most vulnerable subtype, CA1-ProS (**Figure 3A** and **3B**, respectively; see also **Figure S1**). Interestingly, we found a statistically significant negative association for both gene expression and spatial distribution distance with respect to end-timepoint τ pathology ($\mathrm{R}^2 = 0.25$ and 0.37, respectively).

The three neuronal types most similar in transcriptomic space to CA1-ProS (CA3, SUB-ProS, and L2/3 IT RHP, all hippocampal glutamatergic neurons) also had the next highest average correlations to τ pathology. Conversely, the three most distantly related (CR, Meis2, and Sst-Chodl, all GABAergic neurons) were all strongly anti-correlated with τ pathology (**Figure 3A**). Similarity to CA1-ProS among GABAergic neurons also appeared to partly explain the stark differences in τ pathology correlation between vulnerable (Lamp5, Sst, and Vip) and resilient (CR, Meis2, Pvalb, Sncg, and Sst-Chodl) subtypes (**Figure 2A**). More broadly, however, much of the variance in regional vulnerability remains unexplained by transcriptomic differences between neurons, and the regional distribution distance to CA1-ProS is more tightly associated with mean τ pathology (**Figure 3B**). Therefore, it is more likely that neuronal vulnerability is driven by spatial proximity to CA1-ProS-dense regions than transcriptomic similarity.

2.5 Spectral embedding of gene expression and spatial similarity matrices

To further understand the structure of these cell-type data, we created two similarity matrices: S_{gene} for gene expression and $S_{spatial}$ for spatial distribution. Each entry (i, j) of each matrix denotes the transcriptomic or spatial correlation-based similarity between cell types i and j, respectively. We then performed spectral embedding, which involves the eigencomposition of the Laplacian matrices derived from S_{gene} and $S_{spatial}$, and projected each cell type onto the first two nontrivial eigenvectors, v_2 and v_3 (**Figure 3C** and **3D**; see also **Methods**). In gene expression space, the cell types generally fell into three clusters: non-neuronal cells, GABAergic neurons, and glutmatergic neurons (**Figures 3C** and **S1A**). The evident overlap between hippocampal and neocortical glutamatergic neurons suggests that the delineation between these cell types does not strictly follow regional boundaries, aligning with the findings from Yao, *et al.*[39]. Clustering was much less distinct when looking at spatial similarity, with one small group of primarily neocortical glutamatergic neurons separating from the rest of the cell types

(**Figures 3D** and **S1B**). Separation with respect to τ vulnerability was also weak, although the cell types with higher v_3 loadings tended to be selectively resilient.

2.6 Correlating regional distributions of AD risk genes with tau pathology

We next examined the regional vulnerability or resilience to τ in terms of the expression of 24 known AD risk genes (**Table S5**)[45,46]. **Figure 4A** shows the variability in correlations between regional gene expression and τ pathology across tauopathy datasets and individual genes. The clusterin gene (*Clu*) stands out as the most correlated with τ (mean R = 0.36) and exhibits broad expression throughout the mouse brain, which is especially strong in the hippocampal regions. By contrast, the most anti-correlated gene, osteopontin (*Spp1*; mean R = -0.25) is primarily expressed in the hindbrain and olfactory areas, regions largely unaffected by τ pathology. Other genes also warrant special attention for their previously noted contributions to selective vulnerability. The nuclear receptor ROR-beta (*Rorb*) gene, a marker for layer-4 neocortical neurons that also distinguishes a subset of especially vulnerable EC neurons[7]. *Rorb* is broadly expressed throughout layer 4 of the neocortex (**Figure 4B**). However, with the notable exception of the unseeded Hurtado study[34], correlations to τ pathology were low in magnitude and trended negative, similar to the neocortical glutamatergic neurons (**Figure 2A**). The neuronal sorting receptor gene *Sorl1*, by contrast, exhibited strong and consistent correlations to τ pathology and is broadly expressed throughout the forebrain (**Figure 4B**). *Trem2*, a marker for disease-associated microglia (DAM), also generally aligned closely with overall τ distributions.

2.7 Cell types better predict regional vulnerability than AD risk genes

We next regressed brain-wide τ pathology distributions using cell-type maps as linear predictors. We used a Bayesian Information Criterion (BIC)-based approach for model selection to control for overfitting, ensuring that each linear model was constructed with only those cell types that were maximally informative (**Figure S9**; see also **Methods**). We found that all models exhibited statistical significance of at least $p < 0.001$ (**Figure 5A**; **Table 1**). Additionally, we found a remarkable diversity among the informative cell types (**Figure 5B**). Oligodendrocytes were selected the most frequently at a rate of 50%, followed by the microglia/perivascular macrophage cell type and three hippocampal glutamatergic neurons (CA1-ProS, CA3, and SUB-ProS). More generally, the four cell-type classes were not equally represented among the informative cell types ($\chi^2_{\mathrm{GOF}} = 15.0$, $p = 5.5 \times 10^{-4}$). Notably, the most underrepresented by overall frequency was cortical glutamatergic neurons, with only three cell types selected once each (L2/3 IT CTX, L5/6 NP CTX, and L6b CTX). To ensure that the BIC-based model selection procedure did not bias our results, we repeated these analyses using the five most correlated cell types per dataset and obtained qualitatively and quantitatively similar results (**Figure S10**; **Table S6**).

When we attempted to model τ pathology as a function of AD risk genes using a similar variable selection procedure, we found that the cell-type-based models had lower BIC values than their gene-based counterparts for 11 of the 12 tauopathy datasets, indicating that the former were generally superior (**Table 1**, **Figures S11**, and **S12**). For two of the datasets (DS7 and DS9 110), we were unable to find a collection of genes that could predict τ pathology at a 0.01 significance level (R^2 = 0.21 and 0.23, respectively), while both datasets were predicted well by the cell-type models (R^2 = 0.84 and 0.66, respectively). We also found that cell-type-based linear models universally outperformed AD-risk-gene models when using the five best correlates as predictors (**Figure S13** and **Table S6**), demonstrating that these results are robust to variable selection method.

2.8 Regional and cell-type vulnerability may change over the progression of the disease

The Hurtado *et al.* mouse model[34] is unique among the 12 tauopathy datasets because it exhibited τ pathology entirely endogenously. This model therefore provided an opportunity to study early tangle accumulation without the influence of external seed injections. Regressions on pathology data from different time points yielded significant associations with our model predictions (**Figure 6A**, **Figure S14**, and **Table S7**). Both observed and predicted τ-pathology also agree visually, capturing early regional involvement and its subsequent spread (**Figure 6B** and **6C**). Interestingly, the dominant feature in the linear models shifted across time points. At 2 months, where only caudal neocoritcal and entorhinal regions exhibited mild τ pathology, the EC-localized L3 IT ENT neuronal subtype was the most informative (**Figure 6D**, left panel). By contrast, Sst and CT SUB neurons, both mainly found in the forebrain, were the most informative at 4 and 6 months, respectively (**Figure 6D**, middle panels). At the final time point, oligodendrocytes stood out as the most significant cell type (**Figure 6D**, right panel); we note that the oligodendrocyte distribution is anti-correlated with τ for this mouse model (**Figure 2A**).

2.9 Distinct mechanisms underlie gene and cell-type-based selective vulnerability and resilience

To more deeply explore similarities and differences between cell-type-based and gene-based selective vulnerability and resilience, we identified four distinct sets of genes: 1) *SV-G*: Top 10% of genes positively correlated with τ across datasets; 2) *SV-C*: Top 10% of genes differentially expressed in vulnerable cell types; 3) *SR-G*: Top 10% of genes negatively correlated with τ across datasets; and 4) *SR-C*: Top 10% of genes differentially expressed in resilient cell types.

Details on gene set selection can be found in **Methods**. There was minimal overlap between the SV-G and SV-C gene sets or the SR-G and SR-C gene sets (**Figure 7A** and **7B**). We also note that, among the 24 AD risk genes explored above (**Table S5**),

only *Spp1* and *Trem2* appeared in the SR-G and SV-C sets, respectively. A subsequent gene ontology (GO) analysis revealed that these gene sets were also associated with distinct biological processes, albeit with some overlap. For instance, SV-G genes were predominantly associated with neuronal development, whereas SV-C genes were more associated with synaptic processes and cell signaling (**Figure 7C**). Similarly, the SR-G gene set was enriched in genes involved with axon maintenance and cognition, while SR-C gene set was linked to the macroscopic organization of the central nervous system during development (**Figure 7D**). These findings underscore that the genetic and mechanistic bases of selective vulnerability and resilience differ substantially when mediated by cell types versus genes.

3 Discussion

Differential susceptibility between regions to pathological protein species like τ is a hallmark of AD and many other neurodegenerative diseases. The discovery of neuronal subtypes in regions that are disproportionately afflicted by τ pathology has given credence to the idea that regional susceptibility may depend on its cell-type composition[2,5,7,9]. Here, we took a computational, meta-analytical approach towards exploring this cell-type-based selective vulnerability (SV-C) and resilience (SR-C) across the whole brain, leveraging available regional tauopathy data in twelve PS19 mouse models[33–37]. After inferring the baseline whole-brain distributions of 42 subclasses of cell types using the recently developed Matrix Inversion and Subset Selection (MISS) algorithm[38,39] (**Figures S2-5**), we performed various statistical analyses to answer the following questions: 1) Which individual cell types and cell-type classes contribute most to SV-C and SR-C? 2) How well are the distributions of τ pathology explained by baseline cell-type distributions? 3) How do SV-C and SR-C compare to the selective vulnerability and resilience conferred by AD risk genes? 4) Are SV-C and SR-C mediated by the same genes and functional gene networks as gene-expression-based selective vulnerability (SV-G) and resilience (SR-G)?

Among the four classes of cell types present in the Yao, *et al.* dataset, only hippocampal glutamatergic neurons showed a significantly positive association with τ pathology (**Figure 2B**). Conversely, we found net negative associations across cortical glutamatergic neurons and GABAergic neurons and no association with non-neuronal cell types. The discrepancy between the two classes of glutamatergic neurons is especially noteworthy given the fact that these types do not cleanly separate in gene expression space (**Figure S1**)[39]. Among all neurons, we also found a striking inverse relationship between gene expression distance from the most vulnerable neuronal type (CA1-ProS) and mean correlation to τ pathology across datasets (**Figure 3A**), which was even more pronounced when considering spatial distance from CA1-ProS instead (**Figure 3B**). Although we were able to draw broad conclusions about which classes of cells may contribute to SV-C and SR-C, we observed a high degree of variability between individual cell types within each class and between datasets (**Figure 2A** and **2B**). This suggests that a more nuanced understanding of the cellular underpinnings of τ vulnerability requires a more detailed investigation of cell types[7,9]. Below we discuss several notable individual cell types in light of their contributions to SV-C or SR-C.

The CA1-ProS neuron, among the 42 cell types evaluated, exhibited the strongest mean positive correlation with τ pathology across the 12 tauopathy datasets (**Figure 2A**) and was one of the most frequently selected cell types in the multivariate models (**Figure 5B**). We note that this association persisted after removing the seeded regions from consideration, indicating that it is not being directly driven by seeding site. Although the present work is only correlative and therefore limits the conclusions we can draw on a mechanistic level, this finding is notable because CA1 neurons have been previously identified as being especially susceptible to energy deprivation in APP/PS1 mouse models, especially when glucose and oxygen supply is compromised[52,53]. CA1 neurons have also been identified to be particularly vulnerable to hyperphosphorylated τ in mouse models using *in situ* cell-type mapping[54]. Neurons isolated from entorhinal regions, however, largely did not confer vulnerability to τ across the 12 datasets (**Figure 2A**) nor did they feature in the multivariate models (**Figure 5B** and **S10B**), despite the fact that the EC is one of the earliest regions to exhibit τ pathology in AD[1]. By contrast, scRNAseq performed on postmortem AD patients identified glutamatergic neurons in layer II of the entorhinal cortex[5], and in particular those expressing the gene *Rorb*[7], to be especially susceptible to τ. Intriguingly, the unseeded Hurtado *et al.* mouse model revealed pronounced early EC pathology (**Figure 6B**), which did exhibit strong correspondence to EC-isolated excitatory neurons such as L3 IT ENT (**Figure 2A** and **6D**)[34]. Therefore, it may be necessary to use mouse models that develop τ endogenously, more closely mimicking human disease conditions, in order to better study primary selective vulnerability[5] to τ pathology.

Recent studies have also identified a role for GABAergic neurons in AD pathophysiology. For instance, τ-dependent GABAergic synaptic dysfunction has been associated with and AD-specific pathological changes[55–57]. More recently, the first multimodal cell atlas of AD, SEA-AD, which profiled the middle temporal gyrus (MTG) of 84 human donors with AD at a single-cell level, identified subtypes of *Pvalb*+ and *Sst*+ interneurons as being prominently affected[9]. Here, we found wide variation in SV-C and SR-C with respect to GABAergic neurons, with *Sst*+ interneurons exhibiting consistently positive correlations with τ while *Pvalb*+ interneurons showed the opposite effect (**Figure 2A**). However, we note that the unseeded Aβ/τ mouse model[34] demonstrated positive associations with both *Pvalb*+ and *Sst*+ interneurons (**Figure 2A**). Because the this study was the only to use a hybrid Aβ/τ mouse model, it may be possible that the patterns of τ deposition are highly influenced by Aβ comorbidity. Recent work in human subjects has found that indeed there are distinct strains of τ that are

specifically observed in AD and not other tauopathic diseases[58–60]. Therefore, interneuron vulnerability to τ may also in fact be strain-specific, a hypothesis that warrants further investigation.

One of our most striking results is the pronounced *resilience* oligodendrocyte-rich regions exhibited to τ; it had the single-strongest correlation with τ among all cell types (**Figure 2A**) and was the most frequently selected cell type in the multivariate models (**Figure 5B**). These cells, which are primarily responsible for myelin production and maintenance[61], have been documented to play a role in tauopathic diseases, but whether their role is protective or harmful remains a subject of debate. Pathway analysis of gene expression changes in human AD found that oligodendrocyte-specific modules were among the most strongly disrupted[27]. AD patients exhibit detectable white matter lesions, suggesting a direct role for oligodendrocyte dysfunction in AD-related pathological changes[62]. Furthermore, although the mechanisms underpinning τ uptake in oligodendrocytes remain poorly understood[63], oligodendrocytes strongly co-localize with τ inclusions in mouse tauopathy and may facilitate τ seeding and propagation[17]. In the context of regional vulnerability, this suggests that oligodendrocytes may help to sequester τ in earlier stages of disease, they can also also serve as τ reservoirs that mediate inter-regional spread. Here, given the strongly negative association between oligodendrocyte density and τ deposition at a whole-brain level, we propose that the crucial homeostatic functions carried out by oligodendrocytes may be more easily disrupted in those regions that naturally have smaller pools of oligodendrocytes at baseline.

By contrast, we found a consistently positive association between τ pathology and Micro-PVM, the immune cell subclass of the Yao, *et al.* dataset (**Figure 2A**). Microglial activation in response to Aβ and τ is a key feature of AD pathophysiology. For instance, rare genetic variants of the microglial activation gene, *TREM2*, are associated with significantly increased risk of AD[64]. However, as with oligodendrocytes, the roles of microglia in the context of AD are complex and incompletely understood. Early in the disease process, microglia effectively clear Aβ pathology in mouse models[65,66], but over time, their capacity to remove plaques is attenuated[67]. Additionally, the neuroinflammatory response mounted by microglia in response to Aβ, which involves the release of inflammatory cytokines and the production of reactive oxygen species, can exacerbate protein pathology and induce neurodegeneration. Suppression of microglia in 5xFAD mice prevented hippocampal neuronal loss[68], and activated microglia with internalized τ were discovered in a postmortem examination of the brains of AD patients[69]. In the context of our results, regions with high baseline levels of microglia appear to have a greater vulnerability to τ pathology at later timepoints, suggesting that these cells may play a mediatory rather than protective role at more advanced stages of disease. Further bench work is needed to quantify microglial levels *in situ* to gain a more nuanced understanding of how these cells mediate pathology at a whole-brain level.

In addition to the role played by specific cell types in the CNS, non-cell-specific molecular factors may also mediate SV. Similar to the Yao, *et al.* cell types, the baseline gene expression profiles of 24 AD risk genes[45,46] were variably associated with τ pathology (**Figure 4**). However, we also found that cell-type distributions consistently explained end-timepoint τ pathology better than this set of risk genes (**Figure 5**; **Table 1**). Furthermore, the genes most directly contributing to SV-G and SR-G (that is, those directly and most strongly correlated with τ) were strikingly different from those differentiating vulnerable and resilient cell types (**Figure 7**). Gene ontology (GO) analysis revealed that these gene sets were enriched in distinct biological processes: SV-G genes were predominantly associated with neuronal development, whereas SV-C genes were more associated with synaptic processes and cell signaling. SR-G genes were involved with axon maintenance and cognition, while SR-C genes with the macroscopic organization of the CNS during development.

This suggests that SV-C/SR-C and SV-G/SR-G may be mediated independently, and may involve different processes. For instance, electrophysiological or morphological features of vulnerable neurons that are unrelated to baseline gene expression in adult mice may contribute to their propensity to accumulate τ[7,16]. Another extrinsic factor that contributes to the pathophysiology of tauopathies, which we did not explore here, is the trans-neuronal spread of τ along white matter tracts. Seminal work by Clavaguera, *et al.* demonstrated that the injection of pathological τ was sufficient to induce misfolding and aggregation of endogenous τ in distal regions[70], and *in vitro* experiments have directly shown that pathological τ can travel between neurons sharing a synapse[71–75]. In this context it is especially intriguing that genes responding to vulnerable cells are functionally enriched in synaptic and cell-cell processes that may be presumed to influence trans-neuronal spread of τ. Future experimental and computational work is required to reconcile cell-type- and gene-mediated effects with trans-neuronal spread to provide a more complete mechanistic understanding of selective vulnerability.

In summary, our study complements prior experimental approaches, offering a comprehensive exploration of the underpinnings of cell-type-specific regional vulnerability in *in vivo* tauopathy models. By deriving regional cell-type distributions using spatial deconvolution and then performing a meta-analysis of PS19 mouse τ pathology in the context of SV-C, we identified key cell types mediating vulnerability and resilience. We also demonstrated that SV-C is a robust mechanism for explaining diverse patterns of τ pathology across different mouse models. These results illustrate that integrative, computational approaches can reveal important insights into tauopathic disease as well as suggest avenues of future investigation.

4 Methods

4.1 Datasets

4.1.1 Gene expression

The scRNAseq data used to generate the cell-type maps come from Yao, *et al.* for the Allen Institute for Brain Science (AIBS), which sequenced approximately 1.3 million individual cells sampled comprehensively throughout the neocortex and hippocampal formation at 10x sequencing depth[39]. We used the provided trimmed means by cluster dataset (https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x), as the Matrix Inversion and Subset Selection (MISS) algorithm only requires the consensus profiles of cell types per cluster. Utilizing the hierarchical taxonomy provided by the authors, we grouped the 387 individual clusters into subclasses as we have done previously[38], resulting in 42 unique neuronal and non-neuronal cell types spanning four major classes: cortical glutamatergic, hippocampal glutamatergic, GABAergic, and non-neuronal (**Tables S1** and **S2**).

The spatial gene expression data come from the coronal series of the *in situ* hybridization (ISH)-based Allen Gene Expression Atlas (AGEA)[50]. While the sagittal atlas has better gene coverage, we chose to use the coronal atlas because of its superior spatial coverage, which provides an isotropic resolution of 200 μM per voxel. Furthermore, MISS uses a feature selection algorithm to remove uninformative and noisy genes, partly mitigating the effect of the reduced gene coverage. We performed unweighted averaging on genes for which multiple probes were available, resulting in a dataset of 4083 unique genes. Lastly, we removed the 320 genes that were not present in both the scRNAseq and ISH datasets, resulting in a final set of 3763 genes.

4.1.2 Tauopathy experiments

We queried five studies to obtain twelve individual mouse tauopathy datasets (which we refer to interchangeably as "experiments"): BoludaCBD[33], BoludaDSAD[33], DS4[37], DS6[37], DS6 110[37], DS7[37], DS9[37], DS9 110[37], Hurtado[34], IbaHippInj[35], IbaStrInj[35], and IbaP301S[36]. We summarize the key elements of each experiment in **Table S3**. We selected these studies for their spatial coverage ($>$ 40 regions quantified across both hemispheres) and the fact that they all utilized the same mouse tauopathy model (PS19), which contains a P301S τ transgene on a C57BL/6 background. The only exception is the Hurtado experiment, which contained an additional mutation in the amyloid precursor protein (APP) gene. This model is particularly insightful because the endogenous Aβ production induces τ pathology without the requirement of an injected seed.

4.2 Matrix Inversion and Subset Selection (MISS)

We applied the MISS algorithm to the Yao, *et al.* scRNAseq dataset[39] and the AGEA ISH dataset[50] as was described previously[38]. Briefly, MISS involves two steps: 1) subset selection, which utilizes a feature selection algorithm to remove low-information genes that add noise to the final prediction of cell-type density; and 2) matrix inversion, where the gene-subset spatial ISH-based gene expression matrix is regressed on the gene-subset scRNAseq-based gene expression matrix voxel-by-voxel to obtain cell-type densities. We outline each step below.

4.2.1 MRx3-based subset selection

As described previously[38], we first employed the Minimum-Redundancy-Maximum-Relevance-Minimum-Residual (MRx3) feature selection algorithm, which builds on the popular Minimum-Redundancy-Maximum-Relevance (mRMR) algorithm[76]. Let us first designate the $N_g \times N_v$ spatial gene expression matrix as E (normalized by gene), the $N_g \times N_t$ scRNAseq matrix as C (normalized by cell type), and the $N_t \times N_v$ target matrix of cell-type densities as D, where N_g, N_v, and N_t are the numbers of total genes, voxels, and cell types, respectively. At this stage, $N_g = 3763$, $N_v = 50246$, and $N_t = 42$. We seek to find an informative gene subset, S, which survives the following procedure.

The first step of MRx3 is to remove genes from the full gene set, G, which contribute noise to the prediction of the spatial gene expression matrix. We use a rank-1 update rule[77] to estimate the incremental noise added per gene, as assessed by the mean-squared error between the given E and the predicted $\hat{E}$, $||E - C \cdot \hat{D}||_2^2$), and then remove the upper 10% of genes in terms of noise added. This leaves the reduced gene set G^* and the reduced matrices E^* and C^*. We then apply mRMR on the genes of G^* that remain as candidates to be added to the optimal gene set, S. mRMR is a greedy algorithm which, at every iteration, adds the gene to S that maximizes the following criterion V_i:

$$V_i = \frac{F_i}{Redund(i|S)} \tag{1}$$

$$F_i = \sum_{j=1}^{N_t} \frac{(C^*(i,j) - \overline{C^*(i,:)})^2}{N_t - 1} \tag{2}$$

$$Redund(i|S) = \frac{1}{|S|} \sum j \in S|R(i,j|, \tag{3}$$

where the index i indicates the ith gene of C^*, F_i is the F-statistic of gene i, $Redund(i|S)$ is the *redundancy* of gene i with respect to set S, $C^*(i,j)$ is the column-normalized expression of gene i of cell type j, $\overline{C^*(i,:)}$ is the mean expression of gene i across all cell types, $|S|$ is the cardinality of S, and $|R(i,j)|$ is the absolute value of the Pearson's correlation between the expression of target gene i and included gene j of S across all cell types.

This step of the MRx3 algorithm can be summarized as follows:

Result: S, the set of n MRx3-selected genes
Initialize $S_0 = \emptyset$ and $k = 1$;
while $|S| \leq n$ **do**
  $g_k = \underset{i \in G - S_{k-1}}{\text{argmax}} \frac{F_i}{Redund(i|S_{k-1})}$;
  $S_k = \{S_{k-1} \cup g_k\}$;
  $k = k + 1$;
end

After finding all gene sets S_n of cardinality $|S| = n$ using the above procedure, we then choose the optimal value of n, n_G, that balances minimizing the reconstruction residual error and number of included genes. For the Yao, *et al.* dataset used in the present study, $n_G = 1300$, using the same procedure previously described[38]. This produces matrices E_{red} and C_{red}, which only contain the 1300 rows corresponding to the genes in S_{n_G}. All hyperparameters of the method were consistent except that we considered sizes of S from 400 to 3100 genes: $|S| < 400$ produced maps with high residuals that excluded them from consideration, and singularities in the gene expression matrix produced null predictions for $|S| > 3100$.

4.2.2 Matrix inversion

We find the densities of the 42 cell types by solving the following equation voxel-by-voxel:

$$e^k_{red} = C_{red} d^k \tag{4}$$

where e^k_{red} is the 1300×1 gene expression vector for the kth voxel in E_{red} and d^k is the 42×1 cell-type density vector for the kth voxel in D. We used the `lsqnonneg` function in MATLAB to solve **Equation 4**.

4.2.3 MISS validation

To demonstrate the accuracy of our maps, we compared our inferred distributions of key interneurons to published regional densities in the neocortex[78]. In order to do a valid comparison between *Sst+* interneuron distributions, we averaged the densities of the *Sst+* subtypes Sst and Sst Chodl in each region (**Table S2**).

4.3 Coregistration and seed removal

In order to compare the cell-type distributions to τ pathology, we required a common regional space for both. Given the whole-brain, voxel-level resolution of the cell-type maps, we were limited in this study by the spatial sampling of the τ pathology data (**Table S3**). As mentioned above, the regional cell-type densities were calculated by averaging across the voxel-wise densities for the 424 regions of the AIBS CCFv2[51]. Then, for each quantified region per mouse experiment, we matched the corresponding regions between these two atlases, which in most cases had a clear 1:1 correspondence. The very few regions in the tauopathy datasets that did not cleanly correspond to the CCF parcellation were removed from consideration. In those cases where the CCF atlas contained multiple subregions of a larger region sampled in a tauopathy experiment, we averaged cell-type densities across subregions, weighted by subregion volume. Finally, each tauopathy dataset with the exception of Hurtado[34] and IbaP301S[36] were injected with a seed in a location that was uniquely defined within the CCF parcellation. For each of the other 10 experiments, we removed all seeded regions from both the tauopathy data and the cell-type maps prior to performing all downstream analyses. See **Table S3** for more complete details on each tauopathy experiment.

4.4 Modeling and statistical analysis

All analyses were performed using MATLAB v.2022b. For the linear modeling, we first performed feature selection to minimize the risk of overfitting. Because feature selection is a nontrivial combinatorial problem and the best criterion for choosing an "optimal" model is not unambiguous, we used two different feature selection methods to construct our linear models:

1. *BIC*: We ordered the 42 cell-type features by Pearson's correlation to regional τ pathology and removed the lower 75% from consideration. For the 24 genes, we removed the lower 50% to make sure the sizes of these reduced sets were roughly the same. We then calculated the Bayesian Information Criterion (BIC) for linear models of dimensionality 2 to $n_{\text{types}} + 1$ (including the intercept term), where the features were added in descending order of Pearson's correlation. The "optimal" model for each tauopathy experiment was that which had the lowest BIC value.

2. *Top 5*: We took the five most correlated features per tauopathy experiment and constructed linear models using all of them. By definition, the "optimal" models using the second method were identical to those of dimensionality 6 using the first method.

As mentioned above, an intercept term was included for all models. Both procedures yielded qualitatively similar results. All p-values provided were adjusted for false positives using the Bonferroni correction, except where indicated:

$$p_{\text{corr}} = \frac{p_{\text{nominal}}}{N_{\text{tests}}} \tag{5}$$

4.5 Spectral embedding

In order to perform spectral embedding to probe the structure of the Yao, *et al.* cell type data we first constructed correlation distance matrices D_{gene} and $D_{spatial}$, where $D = 1 - R$ and R is the Pearson's cross-correlation matrix (see **Figure S1**). D_{gene} used the correlations between cell types in terms of their expression of the 1300 MRx3 genes (**Figure S1A**) while $D_{spatial}$ used the correlations between cell types in terms of their MISS-derived densities across the 424 CCFv2 regions (**Figure S1B**). We then constructed similarity matrices S_{gene} and $S_{spatial}$ from D_{gene} and $D_{spatial}$, respectively, in the following way:

$$s_{ij} = \begin{cases} \frac{1}{d_{ij}} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{6}$$

After min-max normalizing these S matrices, we constructed the corresponding unnormalized graph Laplacians L_{gene} and $L_{spatial}$, where $L = \Delta - S$ and Δ is the degree matrix of S. To project the cell types into the gene expression and spatial eigenspaces, we performed the eigendecomposition of L_{gene} and $L_{spatial}$, respectively, and plotted each cell type's loadings on eigenvectors v_2 and v_3 (the first eigenvector, v_1, is trivial with an eigenvalue of 0 for Laplacian matrices). This method is very similar to the popular spectral clustering approach, which uses k-means clustering on these eigenvectors v of L to cluster data.

4.6 Gene ontology analysis

In order to perform gene ontology (GO) analysis, we first created four gene subsets: SV-C, SV-G, SR-C, and SR-G. The SV-G/SR-G gene subsets were constructed by taking the top 10% correlated and anti-correlated genes among the 3763 in the AGEA, respectively; this resulted in a 277-gene SV-G and 100-gene SR-G subsets (**Table S8** and **S9**, respectively). We required that the corresponding SV-C and SR-C subsets to contain equal numbers of differentially expressed (DE) genes per cell type as well as roughly equivalent sizes to these SV-G and SR-G subsets, respectively. This ensured that each cell type was roughly equally represented and that a meaningful comparison between subsets could be conducted. To this end, we first constructed the column-normalized 3763×42 matrix C_T, where each entry $c_T(i, j)$ represents the scRNAseq-based expression of gene i in cell type j. We then calculated the z-scores of each gene across cell types (in other words, for each row of C_T). This yielded a straightforward measure of DE that could be compared across genes to identify those that best discriminated each cell type from the others.

The vulnerable cell types that were incorporated into the SV-C gene set were those that had positive mean correlations across tauopathy datasets (**Figure 2A**) *and* were selected at least once in multivariate models (there were 16 such cell types; see also **Figure 5B**). We took the union set of the top n_{DEvuln} genes between these 16 cell types, $n_{DEvuln} = round(|\text{SV-G}|/n_{CTvuln}) = round(277/16) = 18$. After removing duplicate genes between vulnerable cell types, this resulted in a SV-C subset of 264 genes (**Table S10**). We used a similar procedure to construct the SR-C subset, resulting in a set of 104 genes (**Table S11**).

We conducted separate gene ontology (GO) analyses on each of the four gene subsets using the ShinyGO toolbox[79] (v0.77, http://bioinformatics.sdstate.edu/go/) and the biological process GO database, which allowed us to identify the biological processes most functionally enriched in each gene subset. Dot plots were constructed using the built-in ShinyGO visualization tool, where we show only those biological processes that were among the top 10 by fold enrichment and also survived an FDR cutoff of 0.05.

4.7 Code availability

All code for running the MISS algorithm to generate cell-type maps from gene expression data can be downloaded here: https://github.com/Raj-Lab-UCSF/MISS-Pipeline; see also the original publication[38]. All code for running the analyses and generating the plots can be downloaded here: https://github.com/Raj-Lab-UCSF/CellTypeVulnerability. Generating the glass brain images of 3-dimensional density maps requires the external package Brainframe, hosted here: https://github.com/Raj-Lab-UCSF/Brainframe.

References

1. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259, DOI: 10.1007/BF00308809 (1991).
2. Fu, H. *et al.* A tau homeostasis signature is linked with the cellular and regional vulnerability of excitatory neurons to tau pathology. *Nat. Neurosci.* **22**, 47–56, DOI: 10.1038/s41593-018-0298-7 (2019).
3. Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097, DOI: 10.1038/s41593-019-0539-4 (2019).
4. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337, DOI: 10.1038/s41586-019-1195-2 (2019).
5. Fu, H., Hardy, J. & Duff, K. E. Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1350–1358, DOI: 10.1038/s41593-018-0221-2 (2018).
6. Muratore, C. R. *et al.* Cell-type Dependent Alzheimer's Disease Phenotypes: Probing the Biology of Selective Neuronal Vulnerability. *Stem Cell Reports* **9**, 1868–1884, DOI: 10.1016/j.stemcr.2017.10.015 (2017).
7. Leng, K. *et al.* Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **24**, 276–287, DOI: 10.1038/s41593-020-00764-7 (2021).
8. Mather, M. & Harley, C. W. The Locus Coeruleus: Essential for Maintaining Cognitive Function and the Aging Brain. *Trends Cogn. Sci.* **20**, 214–226, DOI: 10.1016/j.tics.2016.01.001 (2016).
9. Gabitto, M. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. *Res. square* DOI: 10.21203/rs.3.rs-2921860/v1 (2023).
10. Braak, H. & Del Tredici, K. Where, when, and in what form does sporadic Alzheimer's disease begin? *Curr. Opin. Neurol.* **25**, 708–714, DOI: 10.1097/WCO.0b013e32835a3432 (2012).
11. Braak, H. *et al.* Pathology associated with sporadic Parkinson's disease — where does it end? *Park. Dis. Relat. Disord.* 89–97, DOI: 10.1007/978-3-211-45295-0_15 (2006).
12. Braak, H. *et al.* Stanley Fahn Lecture 2005: The staging procedure for the inclusion body pathology associated with sporadic Parkinson's disease reconsidered. *Mov. Disord.* **21**, 2042–2051, DOI: 10.1002/mds.21065 (2006).
13. Braak, H., Del Tredici, K., Schultz, C. & Braak, E. Vulnerability of Select Neuronal Types to Alzheimer's Disease. *Annals New York Acad. Sci.* **924**, 53–61, DOI: 10.1111/j.1749-6632.2000.tb05560.x (2006).
14. Braak, H. & Del Tredici, K. Alzheimer's pathogenesis: is there neuron-to-neuron propagation? *Acta Neuropathol.* **121**, 589–595, DOI: 10.1007/s00401-011-0825-z (2011).
15. Braak, H., Thal, D. R., Ghebremedhin, E. & Del Tredici, K. Stages of the Pathologic Process in Alzheimer Disease: Age Categories From 1 to 100 Years. *J. Neuropathol. Exp. Neurol.* **70**, 960–969, DOI: 10.1097/NEN.0b013e318232a379 (2011).
16. Stranahan, A. M. & Mattson, M. P. Selective Vulnerability of Neurons in Layer II of the Entorhinal Cortex during Aging and Alzheimer's Disease. *Neural Plast.* **2010**, 1–8, DOI: 10.1155/2010/108190 (2010).
17. Ferrer, I. *et al.* Involvement of Oligodendrocytes in Tau Seeding and Spreading in Tauopathies. *Front. Aging Neurosci.* **11**, DOI: 10.3389/fnagi.2019.00112 (2019).
18. Narasimhan, S. *et al.* Pathological Tau Strains from Human Brains Recapitulate the Diversity of Tauopathies in Nontransgenic Mouse Brain. *The J. Neurosci.* **37**, 11406–11423, DOI: 10.1523/JNEUROSCI.1230-17.2017 (2017).
19. Anand, C., Maia, P. D., Torok, J., Mezias, C. & Raj, A. The effects of microglia on tauopathy progression can be quantified using Nexopathy in silico (Nexis) models. *Sci. Reports* **12**, 21170, DOI: 10.1038/s41598-022-25131-3 (2022).
20. Dani, M. *et al.* Microglial activation correlates in vivo with both tau and amyloid in Alzheimer's disease. *Brain* **141**, 2740–2754, DOI: 10.1093/brain/awy188 (2018).
21. Griffin, W. S. *et al.* Glial-neuronal interactions in Alzheimer's disease: the potential role of a 'cytokine cycle' in disease progression. *Brain Pathol.* **8**, 65–72, DOI: 10.1111/j.1750-3639.1998.tb00136.x (1998).
22. Hamelin, L. *et al.* Early and protective microglial activation in Alzheimer's disease: a prospective study using 18 F-DPA-714 PET imaging. *Brain* **139**, 1252–1264, DOI: 10.1093/brain/aww017 (2016).
23. Hopp, S. C. *et al.* The role of microglia in processing and spreading of bioactive tau seeds in Alzheimer's disease. *J. Neuroinflammation* **15**, 269, DOI: 10.1186/s12974-018-1309-z (2018).

24. Perea, J. R., Bolós, M. & Avila, J. Microglia in Alzheimer's Disease in the Context of Tau Pathology. *Biomolecules* **10**, 1439, DOI: 10.3390/biom10101439 (2020).

25. Španić, E., Langer Horvat, L., Hof, P. R. & Šimić, G. Role of Microglial Cells in Alzheimer's Disease Tau Propagation. *Front. Aging Neurosci.* **11**, DOI: 10.3389/fnagi.2019.00271 (2019).

26. Spangenberg, E. *et al.* Sustained microglial depletion with CSF1R inhibitor impairs parenchymal plaque development in an Alzheimer's disease model. *Nat. Commun.* **10**, 3758, DOI: 10.1038/s41467-019-11674-z (2019).

27. Wang, M. *et al.* Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Medicine* **8**, 104, DOI: 10.1186/s13073-016-0355-3 (2016).

28. DaRocha-Souto, B. *et al.* Brain Oligomeric β-Amyloid but Not Total Amyloid Plaque Burden Correlates With Neuronal Loss and Astrocyte Inflammatory Response in Amyloid Precursor Protein/Tau Transgenic Mice. *J. Neuropathol. Exp. Neurol.* **70**, 360–376, DOI: 10.1097/NEN.0b013e318217a118 (2011).

29. Fleeman, R. M. & Proctor, E. A. Astrocytic Propagation of Tau in the Context of Alzheimer's Disease. *Front. Cell. Neurosci.* **15**, DOI: 10.3389/fncel.2021.645233 (2021).

30. Obulesu, M. & Jhansilakshmi, M. Neuroinflammation in Alzheimer's disease: an understanding of physiology and pathology. *Int. J. Neurosci.* **124**, 227–235, DOI: 10.3109/00207454.2013.831852 (2014).

31. Hansen, D. V., Hanson, J. E. & Sheng, M. Microglia in Alzheimer's disease. *J. Cell Biol.* **217**, 459–472, DOI: 10.1083/jcb.201709069 (2018).

32. Monterey, M. D., Wei, H., Wu, X. & Wu, J. Q. The Many Faces of Astrocytes in Alzheimer's Disease. *Front. Neurol.* **12**, DOI: 10.3389/fneur.2021.619626 (2021).

33. Boluda, S. *et al.* Differential induction and spread of tau pathology in young PS19 tau transgenic mice following intracerebral injections of pathological tau from Alzheimer's disease or corticobasal degeneration brains. *Acta Neuropathol.* **129**, 221–237, DOI: 10.1007/s00401-014-1373-0 (2015).

34. Hurtado, D. E. *et al.* Aβ Accelerates the Spatiotemporal Progression of Tau Pathology and Augments Tau Amyloidosis in an Alzheimer Mouse Model. *The Am. J. Pathol.* **177**, 1977–1988, DOI: 10.2353/ajpath.2010.100346 (2010).

35. Iba, M. *et al.* Synthetic Tau Fibrils Mediate Transmission of Neurofibrillary Tangles in a Transgenic Mouse Model of Alzheimer's-Like Tauopathy. *J. Neurosci.* **33**, 1024–1037, DOI: 10.1523/JNEUROSCI.2642-12.2013 (2013).

36. Iba, M. *et al.* Tau pathology spread in PS19 tau transgenic mice following locus coeruleus (LC) injections of synthetic tau fibrils is determined by the LC's afferent and efferent connections. *Acta Neuropathol.* **130**, 349–362, DOI: 10.1007/s00401-015-1458-4 (2015).

37. Kaufman, S. K. *et al.* Tau Prion Strains Dictate Patterns of Cell Pathology, Progression Rate, and Regional Vulnerability In Vivo. *Neuron* **92**, 796–812, DOI: 10.1016/j.neuron.2016.09.055 (2016).

38. Mezias, C., Torok, J., Maia, P. D., Markley, E. & Raj, A. Matrix Inversion and Subset Selection (MISS): A pipeline for mapping of diverse cell types across the murine brain. *Proc. Natl. Acad. Sci.* **119**, e2111786119, DOI: 10.1073/pnas.2111786119 (2022).

39. Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241.e26, DOI: 10.1016/j.cell.2021.04.021 (2021).

40. Wang, F. *et al.* Enhancing Oligodendrocyte Myelination Rescues Synaptic Loss and Improves Functional Recovery after Chronic Hypoxia. *Neuron* **99**, 689–701.e5, DOI: 10.1016/j.neuron.2018.07.017 (2018).

41. Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324, DOI: 10.1126/science.aau5324 (2018).

42. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935, DOI: 10.1038/s41592-018-0175-z (2018).

43. Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature* **598**, 137–143, DOI: 10.1038/s41586-021-03705-x (2021).

44. Chen, X., Fischer, S., Zhang, A., Gillis, J. & Zador, A. Modular cell type organization of cortical areas revealed by in situ sequencing. *bioRxiv* **598**, DOI: 10.1101/2022.11.06.515380 (2022).

45. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436, DOI: 10.1038/s41588-022-01024-z (2022).

46. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430, DOI: 10.1038/s41588-019-0358-2 (2019).

47. AIBS. Allen Cell Types Database - Technical White Paper: Transcriptomics (2018).

48. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78, DOI: 10.1038/s41586-018-0654-5 (2018).

49. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22, DOI: 10.1016/j.cell.2018.06.021 (2018).

50. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176, DOI: 10.1038/nature05453 (2007).

51. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214, DOI: 10.1038/nature13186 (2014).

52. Casas, C. *et al.* Massive CA1/2 Neuronal Loss with Intraneuronal and N-Terminal Truncated Abeta42 Accumulation in a Novel Alzheimer Transgenic Model. *The Am. J. Pathol.* **165**, 1289–1300, DOI: 10.1016/S0002-9440(10)63388-3 (2004).

53. Hof, P. R. & Morrison, J. H. The aging brain: morphomolecular senescence of cortical circuits. *Trends Neurosci.* **27**, 607–613, DOI: 10.1016/j.tins.2004.07.013 (2004).

54. Zeng, H. *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446, DOI: 10.1038/s41593-022-01251-x (2023).

55. Hardy, J. *et al.* A disorder of cortical GABAergic innervation in Alzheimer's disease. *Neurosci. Lett.* **73**, 192–196, DOI: 10.1016/0304-3940(87)90016-4 (1987).

56. Koliatsos, V. E. *et al.* Early involvement of small inhibitory cortical interneurons in Alzheimer's disease. *Acta Neuropathol.* **112**, 147–162, DOI: 10.1007/s00401-006-0068-6 (2006).

57. Ramos-Miguel, A. *et al.* Loss of Munc18-1 long splice variant in GABAergic terminals is associated with cognitive decline and increased risk of dementia in a community sample. *Mol. Neurodegener.* **10**, 65, DOI: 10.1186/s13024-015-0061-4 (2015).

58. Janelidze, S. *et al.* Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia. *Nat. Medicine* **26**, 379–386, DOI: 10.1038/s41591-020-0755-1 (2020).

59. Palmqvist, S. *et al.* Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA* **324**, 772–781, DOI: 10.1001/jama.2020.12134 (2020).

60. Thijssen, E. H. *et al.* Diagnostic value of plasma phosphorylated tau181 in Alzheimer's disease and frontotemporal lobar degeneration. *Nat. Medicine* **26**, 387–397, DOI: 10.1038/s41591-020-0762-2 (2020).

61. Bradl, M. & Lassmann, H. Oligodendrocytes: biology and pathology. *Acta Neuropathol.* **119**, 37–53, DOI: 10.1007/s00401-009-0601-5 (2010).

62. Nasrabady, S. E., Rizvi, B., Goldman, J. E. & Brickman, A. M. White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. *Acta Neuropathol. Commun.* **6**, 22, DOI: 10.1186/s40478-018-0515-3 (2018).

63. Jiang, S. & Bhaskar, K. Degradation and Transmission of Tau by Autophagic-Endolysosomal Networks and Potential Therapeutic Targets for Tauopathy. *Front. Mol. Neurosci.* **13**, DOI: 10.3389/fnmol.2020.586731 (2020).

64. Gratuze, M. *et al.* Impact of TREM2R47H variant on tau pathology–induced gliosis and neurodegeneration. *J. Clin. Investig.* **130**, 4954–4968, DOI: 10.1172/JCI138179 (2020).

65. El Khoury, J. *et al.* Ccr2 deficiency impairs microglial accumulation and accelerates progression of Alzheimer-like disease. *Nat. Medicine* **13**, 432–438, DOI: 10.1038/nm1555 (2007).

66. Simard, A. R., Soulet, D., Gowing, G., Julien, J.-P. & Rivest, S. Bone Marrow-Derived Microglia Play a Critical Role in Restricting Senile Plaque Formation in Alzheimer's Disease. *Neuron* **49**, 489–502, DOI: 10.1016/j.neuron.2006.01.022 (2006).

67. Daria, A. *et al.* Young microglia restore amyloid plaque clearance of aged microglia. *The EMBO J.* **36**, 583–603, DOI: 10.15252/embj.201694591 (2017).

68. Spangenberg, E. E. *et al.* Eliminating microglia in Alzheimer's mice prevents neuronal loss without modulating amyloid-β pathology. *Brain* **139**, 1265–1281, DOI: 10.1093/brain/aww016 (2016).

69. Bolós, M. *et al.* Direct Evidence of Internalization of Tau by Microglia In Vitro and In Vivo. *J. Alzheimer's Dis.* **50**, 77–87, DOI: 10.3233/JAD-150704 (2015).

70. Clavaguera, F. *et al.* Transmission and spreading of tauopathy in transgenic mouse brain. *Nat. Cell Biol.* **11**, 909–913, DOI: 10.1038/ncb1901 (2009).

71. Freundt, E. C. *et al.* Neuron-to-neuron transmission of α-synuclein fibrils through axonal transport. *Annals Neurol.* **72**, 517–524, DOI: 10.1002/ana.23747 (2012).

72. Pecho-Vrieseling, E. *et al.* Transneuronal propagation of mutant huntingtin contributes to non–cell autonomous pathology in neurons. *Nat. Neurosci.* **17**, 1064–1072, DOI: 10.1038/nn.3761 (2014).

73. Feiler, M. S. *et al.* TDP-43 is intercellularly transmitted across axon terminals. *J. Cell Biol.* **211**, 897–911, DOI: 10.1083/jcb.201504057 (2015).

74. Takeda, S. *et al.* Neuronal uptake and propagation of a rare phosphorylated high-molecular-weight tau derived from Alzheimer's disease brain. *Nat. Commun.* **6**, 8490, DOI: 10.1038/ncomms9490 (2015).

75. Katsikoudi, A. *et al.* Quantitative propagation of assembled human Tau from Alzheimer's disease brain in microfluidic neuronal cultures. *J. Biol. Chem.* **295**, 13079–13093, DOI: 10.1074/jbc.RA120.013325 (2020).

76. Peng, H., Long, F. & Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis Mach. Intell.* **27**, 1226–1238, DOI: 10.1109/TPAMI.2005.159 (2005).

77. Sherman, J. & Morrison, W. J. Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix. *The Annals Math. Stat.* **21**, 124–127, DOI: 10.1214/aoms/1177729893 (1950).

78. Kim, Y. *et al.* Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. *Cell* **171**, 456–469.e22, DOI: 10.1016/j.cell.2017.09.020 (2017).

79. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629, DOI: 10.1093/bioinformatics/btz931 (2020).

Author contributions statement

J.T. carried out analyses, figure generation, and drafting the manuscript. P.M. and C.A. assisted with the editing of the manuscript and study design. A.R. supervised this study and assisted with all aspects of manuscript preparation and editing.

Disclosures

The authors have no financial conflicts of interest to disclose.

Figures and Tables

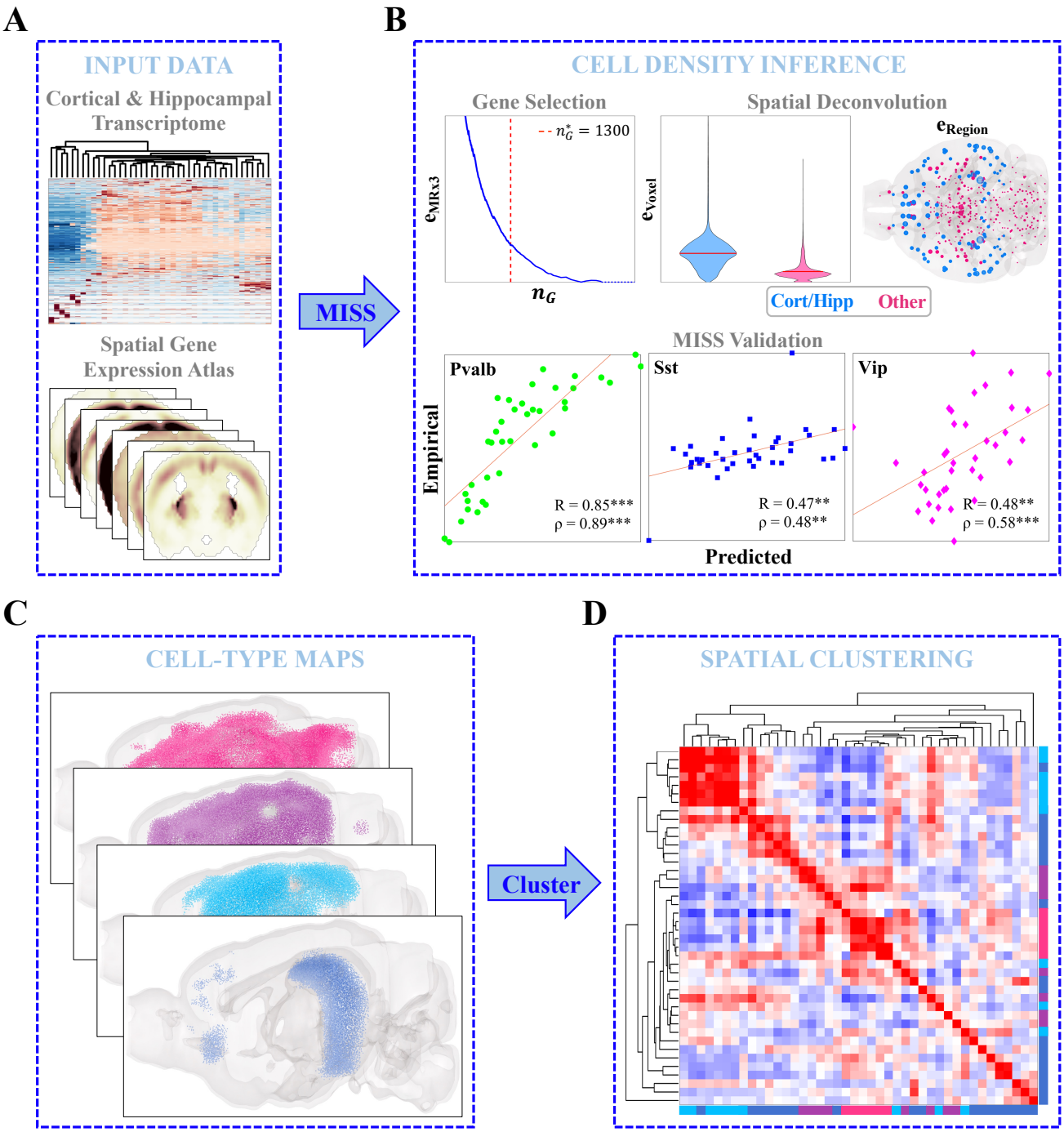
A
INPUT DATA
Cortical & Hippocampal Transcriptome
Spatial Gene Expression Atlas
MISS
B
CELL DENSITY INFERENCE
Gene Selection
$n_G^* = 1300$
e_{MRx3}
n_G
Spatial Deconvolution
e_{Voxel}
e_{Region}
Cort/Hipp
Other
MISS Validation
Pvalb
Sst
Vip
Empirical
Predicted
R = 0.85***
ρ = 0.89***
R = 0.47**
ρ = 0.48**
R = 0.48**
ρ = 0.58***
C
CELL-TYPE MAPS
Cluster
D
SPATIAL CLUSTERING

Figure 1. Matrix Inversion and Subset Selection (MISS) A. Neocortical and hippocampal scRNAseq data from Yao, *et al.*[39] and the Allen Gene Expression Atlas (AGEA)[50] were used to infer cell-type densities using the Matrix Inversion and Subset Selection (MISS) pipeline[38] **B.** MISS proceeds in two parts: informative gene selection and spatial deconvolution. The first step involves choosing an informative gene subset using the MRx3 subset selection algorithm (see **Methods**). Here we found an optimal 1300-gene subset using a similar procedure to that previously published (*top left*). The second step involves solving a nonnegative least-squares problem voxel-by-voxel to determine the whole-brain densities of the Yao, *et al.* cell types. The top center and top right panels show that the algorithm did not exhibit greater degrees of residual error inside of neocortical and hippocampal voxels and regions than elsewhere. We validated our results using published interneuron distributions in the neocortex[78], demonstrating that MISS is accurately inferring cell-type densities (*bottom panels*). **C.** MISS allowed us to create whole-brain maps of cell types at a resolution of 200 μm. **D.** Spatial clustering of these distributions at a regional level reflected a remarkable amount of diversity between cell types.

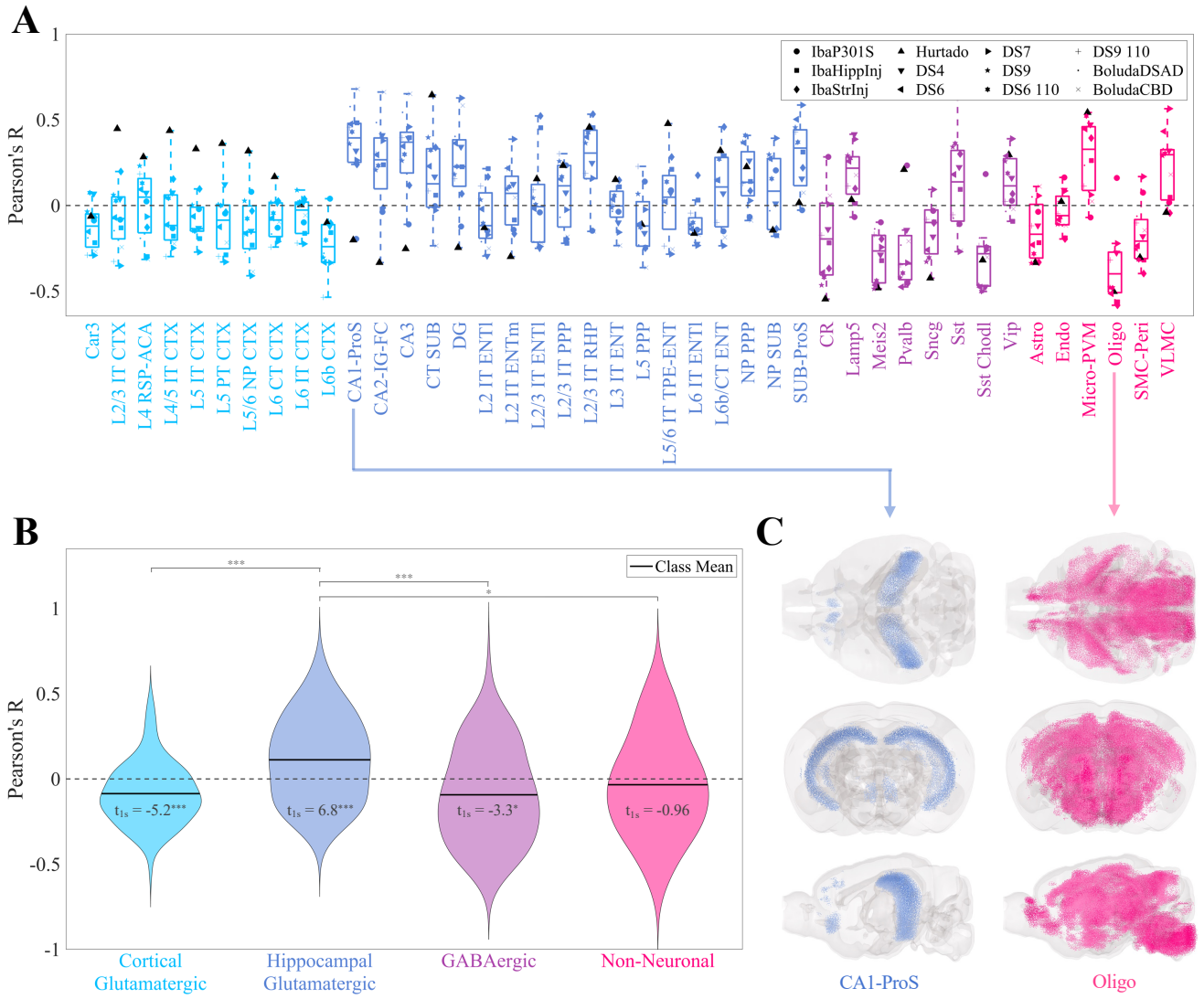

Figure 2. Univariate analysis of the relationship between cell-type density and end-timepoint pathology across mouse tauopathy models. A. Box plot of Pearson's correlation coefficients between the MISS-inferred distributions of the 42 cell types of the Yao, *et al.* dataset and the end tau pathology for 12 mouse tauopathy datasets. **B.** Violin plots of the data in **A**, where cell types have been grouped into four general classes. One-sample t-statistics are displayed within each violin plot and the statistically significant pairwise comparisons are bracketed. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$. **C.** Glass-brain representations of the cell type with the highest mean positive correlation (CA1-ProS, a glutamatergic hippocampal neuron) and the cell type with the highest mean negative correlation (Oligo, oligodendrocytes). Refer to the original publication for a full description of the cell-type annotations[39].

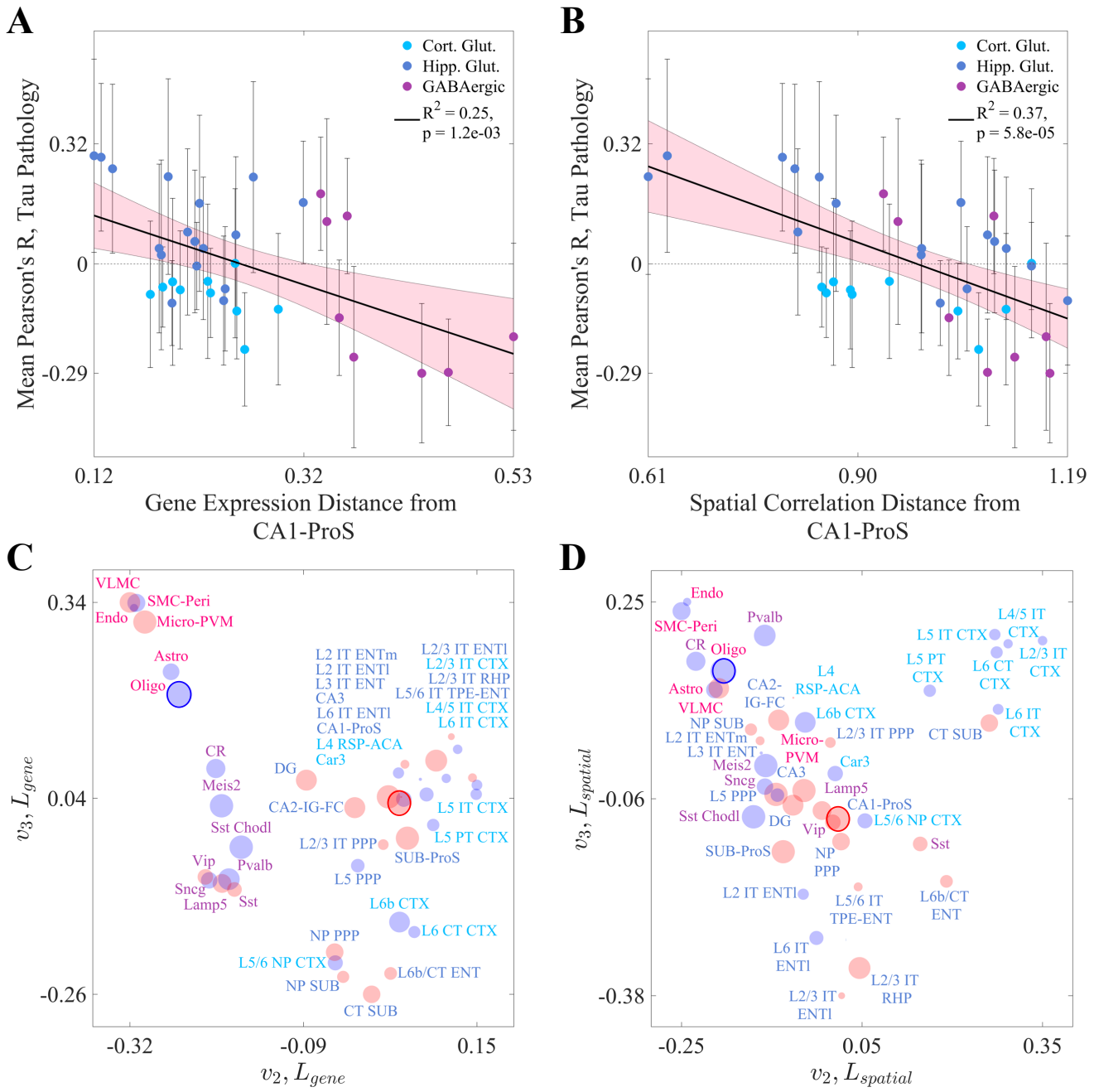

Figure 3. Analysis of inter-cell-type distance in gene expression and spatial distribution. A. Linear models showing the negative relationship between gene expression correlation distance (1 - R) and correlation with end time-point tau pathology for neuronal cell types in the Yao, *et al.* dataset with respect to the neuron most correlated to pathology, CA1-ProS (see also **Figure S3A**). The pink shaded area represents the 95% confidence interval for the least squares regression. **B.** Linear model showing the negative relationship between end-timepoint pathology and inter-cell-type distance with respect to CA1-ProS as in **A**, but using spatial correlation distance instead of gene expression correlation distance (see also **Figure S3B**). **C.** All cell types in the Yao, *et al.* dataset plotted in terms of their projections on v_2 and v_3, the second and third eigenvectors of Laplacian of the gene expression correlation similarity matrix (L_{gene}). The radius of each point corresponds to the magnitude of that cell type's mean correlation to τ pathology and the color indicates the sign (red = positive, blue = negative). **D.** All cell types in the Yao, *et al.* dataset plotted in terms of their projections as in **C**, except v_2 and v_3 are the second and third eigenvectors of the Laplacian of the spatial correlation similarity matrix ($L_{spatial}$). Refer to the original publication for a full description of the cell-type annotations[39]. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$.

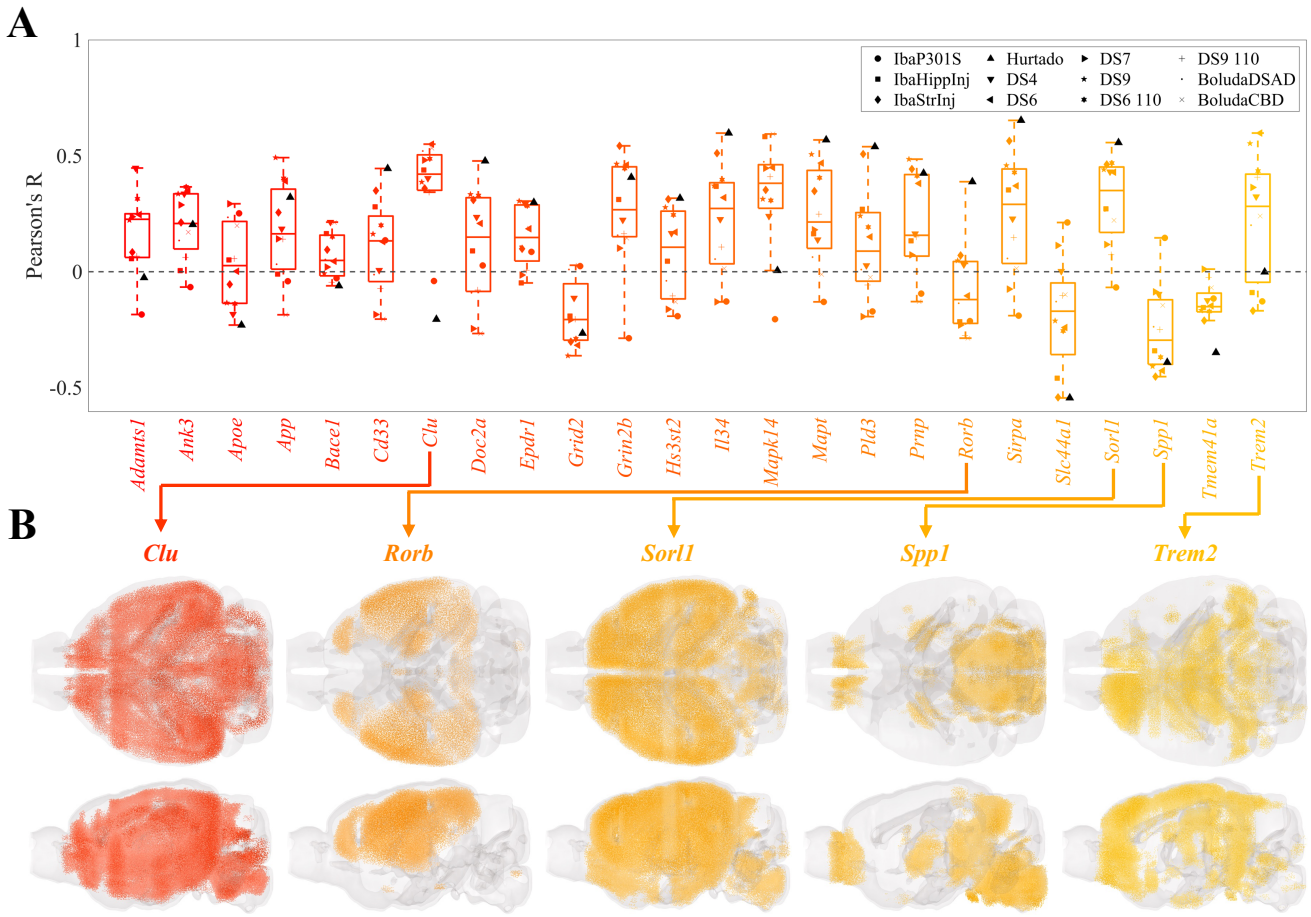

Figure 4. Univariate analysis of the relationship between AD risk genes and end-timepoint pathology across mouse tauopathy models. A. Box plot of Pearson's correlation coefficients between the MISS-inferred distributions of the 24 AD risk genes and the end tau pathology for 12 mouse tauopathy datasets. The list of AD risk genes is provided in **Table S5**. **B.** Glass brain renderings of genes of interest.

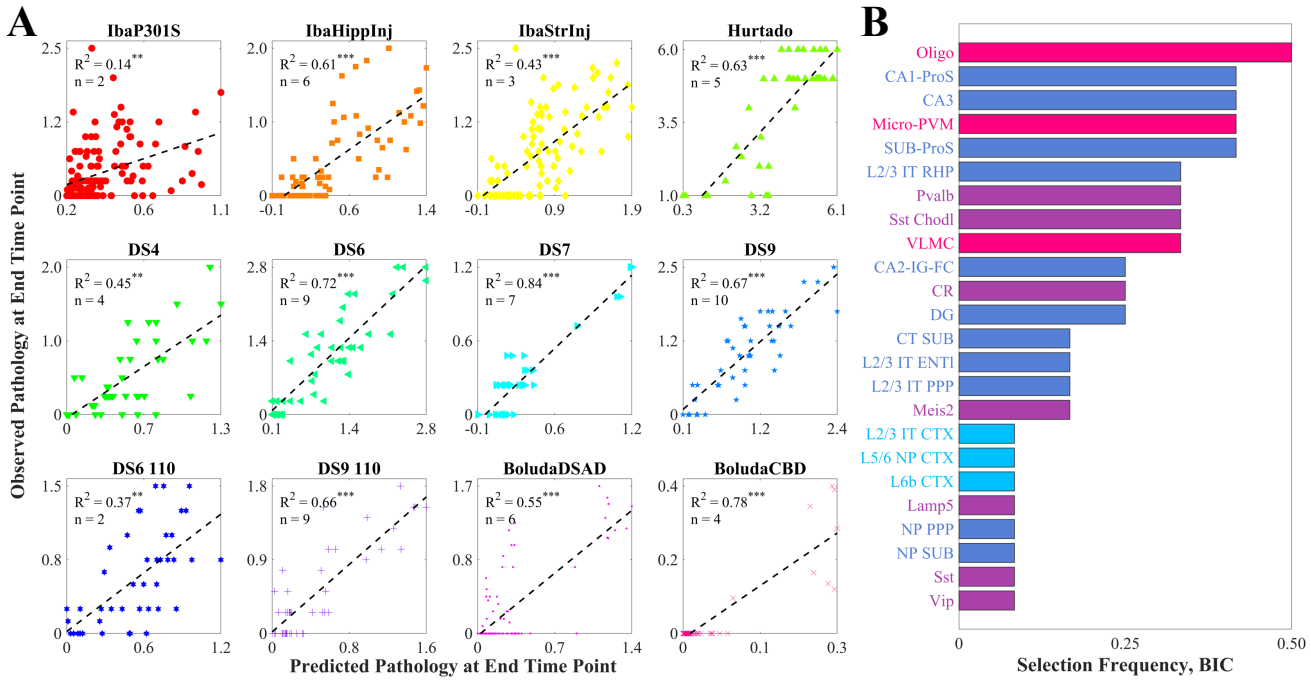

Figure 5. Multivariate analysis of end-timepoint pathology. A. Scatter plots of the optimal cell-type-based models of tau pathology at the end time points for each of the nine mouse tauopathy studies, along with their associated R^2 values and the number of cell types chosen using the BIC. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$. **B.** Bar plot of the frequency with which cell types were included in the BIC-based linear models in **A**. Of the 42 cell types in the Yao, *et al.* dataset, 24 were selected at least once.

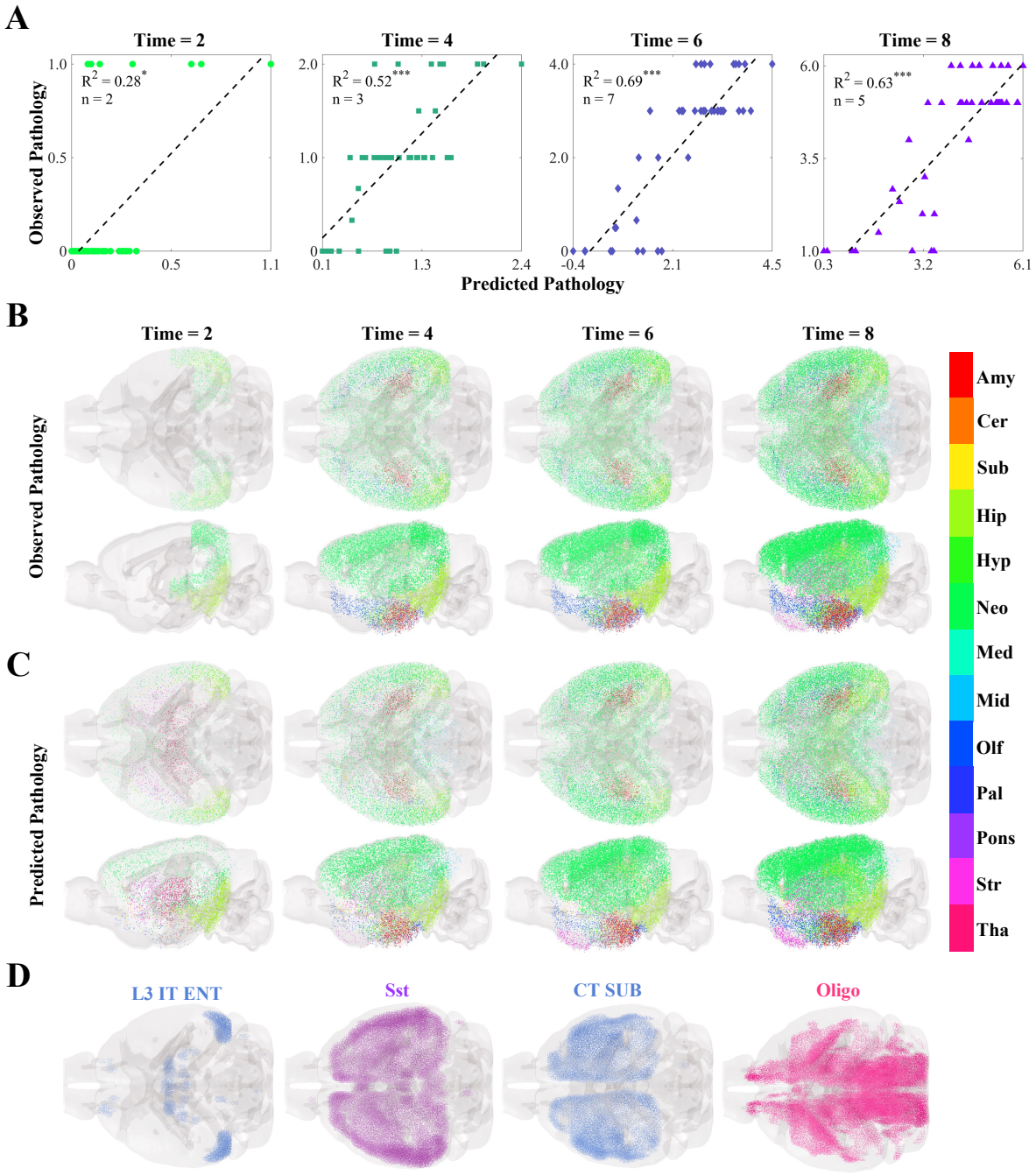

Figure 6. Multivariate analysis of the temporal progression of pathology in an unseeded mouse tauopathy model. A. Scatter plots of the optimal cell-type-based models of tau pathology for each time point in the unseeded Hurtado, *et al.* mouse tauopathy dataset[34], along with their associated R^2 values and the number of cell types chosen using the BIC. Time is given in months. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$. **B,C.** Glass-brain representations of the observed **B** and predicted **C** pathology plotted along the y-axis in **A** over time. The color indicates major region-group: Amy - amygdala; Cer - cerebellum; Sub - cortical subplate; Hip - hippocampus; Hyp - hypothalamus; Neo - neocortex; Med - medulla; Mid - midbrain; Olf - olfactory; Pal - pallidum; Pons - pons; Str - striatum; Tha - thalamus. **D.** Glass-brain representations of the voxel-wise distributions of the four cell types that appear in the BIC-optimal linear models in **A** for all four time points. L3 IT ENT - layer-3 intratelencephalic entorhinal neuron; Sst - telencephalic somatostatin-expressing GABAergic neuron; CT SUB - Corticothalamic neuron of the subiculum; Oligo - oligodendrocytes. Please refer to the original publication for a full description of the cell-type annotations[39].

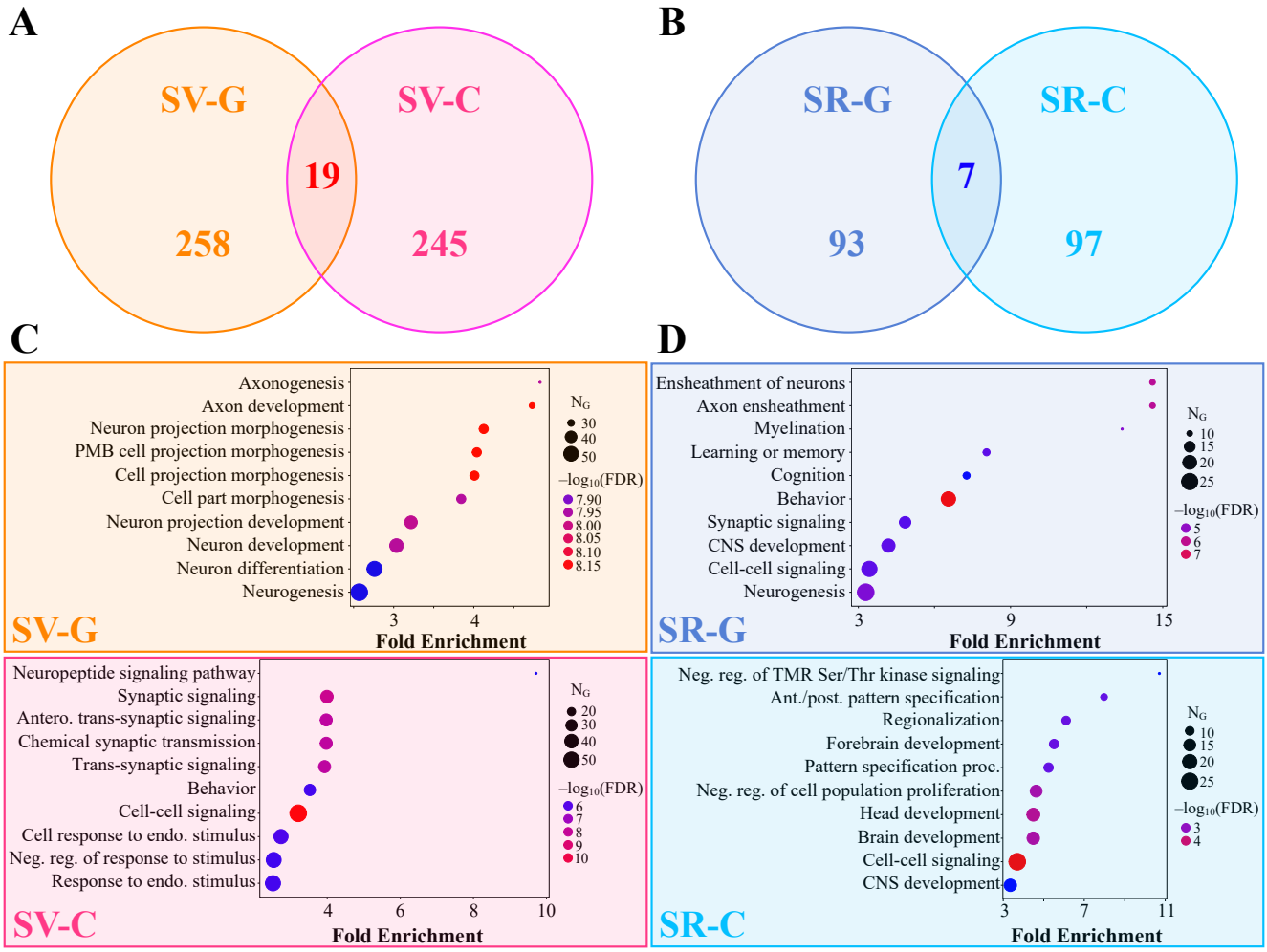

Figure 7. Gene ontology analysis of vulnerable and resilient gene sets. A. Venn diagram of the gene-based selective vulnerability (SV-G) and cell-type-based selective vulnerability (SV-C) genes. **B.** Venn diagram of the gene-based selective resilience (SR-G) and cell-type-based selective resilience (SR-C) genes. **C.** Top 10 biological processes by fold enrichment represented in the SV-G (*top*) and SV-C (*bottom*) gene sets. **D.** Top 10 biological processes by fold enrichment represented in the SR-G (*top*) and SR-C (*bottom*) gene sets. PMB – plasma membrane bounded; TMR – transmembrane receptor; CNS – central nervous system.

Dataset	Cell types (BIC)			AD risk genes (BIC)		
	R^2	n	BIC	R^2	n	BIC
IbaP301S[36]	0.14**	2	180.2	**0.28*** **	**8**	**176.4**
IbaHippInj[35]	**0.61*** **	**6**	**81.0**	0.43***	3	109.0
IbaStrInj[35]	**0.43*** **	**3**	**163.5**	0.31***	1	174.4
Hurtado[34]	**0.63*** **	**5**	**153.3**	0.41***	1	163.8
DS4[37]	**0.45** **	**4**	**50.8**	0.35**	2	52.5
DS6[37]	**0.72*** **	**9**	**79.3**	0.55***	2	82.0
DS7[37]	**0.84*** **	**7**	**-28.7**	0.21	1	23.2
DS9[37]	**0.67*** **	**10**	**70.3**	0.39**	2	75.4
DS6 110[37]	**0.37** **	**2**	**47.3**	0.32*	2	50.4
DS9 110[37]	**0.66*** **	**9**	**49.9**	0.23	2	67.1
BoludaDSAD[33]	**0.55*** **	**6**	**68.0**	0.34***	2	87.8
BoludaCBD[33]	**0.78*** **	**4**	**-167.7**	0.54***	4	-127.3

Table 1. BIC linear regression model statistics. Statistics corresponding to the linear models shown in **Figure 5** and **Figure S8**. Bold font indicates the best model by the Bayesian Information Criterion (BIC). *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$.