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. More specifically in AD, misfolded τ aggregates first accumulate in limbic and temporal areas before spreading to other parts of the brain, while leaving some areas relatively unaffected. One explanation for this apparent selective vulnerability is that there are specific cell types that are particularly susceptible or resistant to τ tangle formation. For instance, prior work has found that τ colocalizes with only a few select subpopulations of excitatory neurons in layer II of the entorhinal cortex as opposed to affecting all excitatory neurons equally. However, while detailed work has been performed in a few focal regions of interest, the cellular underpinnings of regional vulnerability to τ has 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 used an established cell-type mapping pipeline, Matrix Inversion and Subset Selection (MISS), to determine the distributions of pan-hippocampal and neocortical neuronal and non-neuronal cells in the mouse brain using recently available single-cell RNA sequencing (scRNAseq) data from the Allen Institute for Brain Science (AIBS). We then performed a meta-analysis of five studies that quantified regional τ in PS19 mice comprising a total of twelve experimental conditions to identify general principles of cell-type-based selective vulnerability using these cell-type distributions. We found significant variance between cell-type classes in terms of their vulnerability to τ, with hippocampal glutamatergic neurons as a whole being strongly positive associated with patterns of end-timepoint τ deposition while cortical glutamatergic neurons and inhibitory neurons exhibited negative associations. Interestingly, the single-strongest correlate across experiments was oligodendrocytes, whose distributions were anti-correlated with τ pathology, indicating resilience. We also constructed minimal linear models of collections of associated cell types for each experiment and show that not only can cell-type vulnerabilityexplain patterns of τ deposition in all twelve experiments, but that models using cell-type distributions outperform those using the expression patterns of AD risk genes. In short, we have demonstrated that regional cell-type composition is a compelling explanation for the selective vulnerability observed in tauopathic disease at a whole-brain level, which has important implications for better understanding AD pathophysiology and identifying mechanisms that are potential treatment targets.

1 Introduction

All neurodegenerative diseases, including Alzheimer's disease (AD) and other tauopathic diseases such as frontotemporal lobar dementia (FTLD), display *selective vulnerability* (SV), the display of spatially heterogeneous and stereotyped distributions of brain pathology. 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]. One of the most important outstanding questions in studying AD is how to identify both where the disease first emerges and how it subsequently affects specific structures in a patient-specific manner.

With the advent of technologies such as single-cell RNA sequencing (scRNAseq) and single-nucleus RNA sequencing (snRNAseq), it has been established that the transcriptomic programming of the brain is altered in AD in a cell-type-specific manner[2,3]. Accordingly, a key factor thought to contribute to the regional specificity in AD is the selective vulnerability of specific neuronal and non-neuronal cell types in the CNS; we refer to this as *selective cellular vulnerability* (SCV). In this view, the cell type composition of strongly affected regions, such as the hippocampus and EC, confers to them a higher innate susceptibility to AD pathology[4,5]. For instance, *RORB*-expressing excitatory neurons in layer II of the EC specifically exhibit

τ inclusions while other neuronal subpopulations do not[6]. Inhibitory (GABAergic) neurons have also displayed differential vulnerability to AD pathology, with hippocampal *Pvalb+* and *Sst+* interneurons specifically colocalizing with τ. Conversely, there may be cell types that confer not regional vulnerability but resilience to τ pathology. For instance, non-neuronal cells such as microglia, oligodendrocytes, and astrocytes at least in part participate in the remediation of AD-related pathology, although whether these cell types are ultimately helpful or harmful at various stages of disease is an open question.

Most prior literature on studying SCV has understandably focused on descriptive, mechanistic experimental work. Recent studies showing comprehensive scRNAseq profiling of specific regions of interest in AD patient brains, such as the entorhinal cortex[6] and middle temporal gyrus[7], have greatly advanced our understanding of the cellular underpinnings of AD-related pathological changes. However, a comparable exposition of SCV in a quantitative, whole-brain level without pre-selecting regions of interest is currently missing. This gap is entirely understandable: there do not currently exist spatially-resolved quantitative data on cell-type composition of the entire mammalian brain at the resolution and cell type specificity needed to fully address this question statistically. A broader investigation of the influence of regional cell-type composition on vulnerability to pathology, therefore, requires the mapping of a comprehensive set of cell types across the whole brain. While there has been direct quantification of specific types of cells at a regional level[8,9] and single-cell-level mapping of cell types in select areas of the mouse brain[10–12], a definitive, experimentally derived spatial atlas of cell types across the mouse brain has not yet been achieved. More recently, a computational technique called Matrix Inversion and Subset Selection (MISS)[13] produced highly accurate maps of cell types by using their transcriptomic signatures, as determined by single-cell RNA sequencing (scRNAseq)[14–16], to deconvolve the voxelwise gene expression data contained in the Allen Gene Expression Atlas (AGEA)[17]. The breadth of mapped neuronal and non-neuronal cell types coupled with their whole-brain spatial coverage makes the MISS-inferred distributions in particular highly useful for an investigation into selective vulnerability. However, while these MISS-derived maps do have whole-brain spatial coverage, they do not contain all of the relevant cell types to address questions relating to SCV because they lack comprehensive sampling of hippocampal and neocortical regions.

Here, we take advantage of the recent scRNAseq atlas of the mouse hippocampus and neocortex produced by Yao, *et al.* in association with the Allen Institute for Brain Science (AIBS)[18] and use MISS to produce quantitative maps of 42 individual cell types that can be broadly classified into four major types: cortical glutamatergic neurons, hippocampal glutamatergic neurons, GABAergic neurons, and non-neuronal cells. We then performed a meta-analysis of mouse tauopathy using histological τ data from five individual studies all using PS19-background mice[19–23]. We show that, although there is significant heterogeneity both between cell types of a given class and between the twelve experimental conditions explored, only hippocampal glutamatergic neurons as a class were strongly positively associated with τ pathology at the last quantified timepoint, with the CA1-ProS hippocampal glutamatergic neuron having the single-highest positive correlation of the 42 in the Yao, *et al.* dataset. However, the cell type with the strongest overall association with pathology, oligodendrocytes, was in fact negatively correlated, suggesting that high densities of this glial cell confer regional resilience to τ. Among neurons, similarity both in terms of gene expression and spatial distributions partly explains trends in vulnerability, although the association is generally weak. We also show that regional τ deposition can be modeled using combinations of cell types in a statistically significant way, which we could not fully replicate when using the expression of AD risk genes as predictors. Furthermore, we demonstrate that we can model the dynamic SCV of the mouse brain in an unseeded mouse tauopathy model[20]. Together, these results show differential roles of cell types in the etiology of τ-related pathological changes not just in individual regions but across the entire brain, thereby helping to elucidate the general principles of SCV in tauopathic disease.

2 Results

2.1 Cell-type selective vulnerability analysis pipeline

An overview of our data processing and analysis pipeline is shown in **Figure 1**. First, we used the recently published Matrix Inversion and Subset Selection (MISS) algorithm to infer whole-brain cell-type distributions[13]. 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 (**Figures 1A** and **1B**). 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[14–16], where we used the coronal Allen Gene Expression Atlas (AGEA), as our source of spatially resolved gene expression data[13,17]. In the present study, we again leveraged the AGEA, but instead mapped the cell types recently characterized by Yao, *et al.*[18]. 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 1.1 million individual cells with excellent read depth[18]; and 2) capturing cell-type-specific vulnerability to tauopathy requires robust and specific delineation of cortical and hippocampal cell types. In all, we mapped 36 neuronal and 6 non-neuronal cell types at the 200-μm resolution of the AGEA (**Figure 1**; see also **Tables S1** and **S2** for complete descriptions of these cell types). We emphasize that both the scRNAseq and ISH data were taken from healthy, adult, wild-type mice, and thus the MISS-inferred cell-type

distributions represent *baseline* cell-type densities.

Second, we curated twelve regional tauopathy datasets, each of which represents a unique experimental condition, where an experimental condition is defined by mouse genetic background, type of injectate used, and injection site (**Figure 1C**; see also **Table S3** for brief descriptions of these datasets)[19–23]. We then coregistered the inferred MISS cell-type maps, which were parcellated using the Common Coordinate Framework version 2 (CCFv2)[24], into the region spaces of the individual datasets and performed a combination of univariate and multivariate analyses (**Figure 1D**). This meta-analysis of mouse tauopathy allows us to answer broad questions about cell-type selective vulnerability at a whole-brain level.

2.2 Correlations between regional cell-type distributions and end-timepoint tau pathology reveal differences across cell-type classes

Figures 2A and **S1** display the Pearson's R values and their associated p values, respectively, between the regional distributions of each of the 42 cell types in the Yao, *et al.* dataset and the deposition patterns of τ in each of the twelve datasets at their last quantified timepoint (**Figure S2**), with cell types grouped by class; from left to right: cortical glutamatergic neurons, hippocampal glutamatergic neurons, GABAergic neurons, and non-neuronal cells (**Tables S1** and **S2**). While there are generally wide variations in R-values both between τ datasets within a given cell type and between cell types, several macroscopic trends emerge. Generally, we found that τ pathology was positively correlated with the hippocampal glutamatergic neurons (one-sample t-test, $p = 4.3 \times 10^{-10}$), negatively correlated with cortical glutamatergic neurons and GABAergic neurons (($p = 3.3 \times 10^{-6}$ and 4.3×10^{-3}, respectively), and globally uncorrelated with non-neuronal cells (**Figure 2B** and **Table S4**). Further, hippocampal glutamatergic neurons were significantly different from each of the other cell-type classes in terms of their correlations with τ (**Figure 2B** and **Table S4**).

The cell type with the highest mean correlation across datasets (i.e., exhibited the highest vulnerability to τ) was the hippocampal glutamatergic cell type CA1-ProS ($\overline{\mathrm{R}} = 0.32$), whose distribution is shown in **Figure 2C**. Its distribution is almost exclusively confined to the hippocampus, with CA1 containing the highest density of this neuron. Although many of these mouse experiments were seeded with τ at CA1 (**Table S3**)[19,23], we specifically excluded the seed region from the analysis, so the high degree of similarity between this neuron and τ patterns is due to its distribution lying outside of CA1. The correlations of the τ pathology between timepoints in each study are also generally much less than 1 and go down over time (**Figure S3**), further emphasizing that the pathology at the seeded region alone is not driving these relationships. Although the non-neuronal cell types as a class exhibited no correlation to τ pathology across tauopathy datasets (**Figure 2B** and **Table S4**), several types did exhibit vulnerability or resilience to τ. Oligodendrocytes (Oligo), whose distribution is shown in **Figure 2C**, were the single-most resilient cell type in the Yao, *et al.* dataset ($\overline{\mathrm{R}} = -0.36$). Though less strong in magnitude, astrocytes (Astro) also tended to be negatively correlated with τ pathology across datasets. By contrast, the immune cell subtype consisting of microglia and perivascular macrophages (Micro-PVM) tended to be strongly postively correlated with end-timepoint τ pathology. As these various glial cells are well known to participate in the disease process, the directions of these relationships are noteworthy (see **Discussion**).

2.3 Regional vulnerability is weakly determined by cell type gene expression and spatial similarity

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 particular the stark differences between hippocampal and neocortical glutamatergic neurons, we explored the more granular transcriptomic and spatial differences between cell types. First, we investigated how each neurons's mean correlation to τ pathology across the twelve datasets varied with respect to distance from the most positively correlated cell type, CA1-ProS, both in terms of gene expression (**Figure 3A**) and regional distribution (**Figure 3B**), where "distance" is defined as 1 minus the pairwise Pearson's correlation between cell types (see also **Figure S4**). 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 highest average correlations to τ pathology aside from CA1-ProS; conversely, the three most distantly related (CR, Meis2, and Sst-Chodl, all GABAergic neurons) were all strongly anti-correlated with τ pathology (**Figure 3A**). Further, similarity to CA1-ProS among GABAergic neurons appears 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**). From a holistic perspective, 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 than transcriptomic similarity.

We also probed the global structure of the relationships between cell types by examining their projections in the eigenspaces of the Laplacians of the gene expression and spatial similarity matrices (**Figures 3C** and **3D**, respectively). In gene expression space, the cell types generally fall into three clusters: non-neuronal cells, GABAergic neurons, and glutmatergic neurons (**Figures 3C** and **S4A**). Notably, hippocampal and neocortical glutamatergic neurons are not easily separable, indicating that

these neurons' regions of origin weakly determine their transcriptomic similarity; this agrees with the taxonomy originally determined by Yao, *et al.*[18]. Vulnerability to τ, as represented by the color and size of each point in **Figure 3C**, also does not cleanly separate out in gene expression eigenspace. Clustering is much less distinct when looking at their similarity in spatial distribution, with one small group of primarily neocortical glutamatergic neurons separating from the rest of the cell types (**Figures 3D** and **S4B**). Separation with respect to τ vulnerability is also weak, although the cell types with higher v_3 loadings tended to be selectively resilient. We conclude that τ deposition patterns cannot be captured by a straightforward relationship between cell types as a whole, either in terms of their gene expression profiles or their distributions in the brain.

2.4 Regional vulnerability is better explained by cell types than AD risk genes

We next investigated whether we could reconstruct end-timepoint τ pathology using multilinear regression with cell-type distributions as independent predictors. To control for overfitting, we used a Bayesian Information Criterion (BIC)-based approach for model selection, ensuring that each linear model was constructed with only those cell types that were maximally informative (**Figure S5**; see also **Methods**). We found that all models exhibited statistical significance of at least $p < 0.001$ (**Figure 4A** and **Table 1**). Additionally, after tabulating the number of times a given cell type was selected by BIC across all tauopathy datasets, we found a remarkable diversity of informative cell types between datasets (**Figure 4B**), consistent with the wide variations in individual cell-type correlations (**Figure 2A**). 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 S6** and **Table S6**). Therefore, although there is no single cell type whose baseline distribution cleanly explains regional τ deposition patterns across experimental conditions in mouse models of disease, we find that cell types are a strong determinant of selective regional vulnerability, if we expand its definition to include multiple cell types simultaneously.

We also examined the regional vulnerability or resistance to τ in terms of the expression of 24 known AD risk genes (**Table S5**)[25,26]. Like the Yao, *et al.* cell types, the correlations between the regional expression of these genes and τ pathology varied widely between tauopathy datasets and between genes (**Figure 5A**). The single-most correlated gene, clusterin (*Clu*), is broadly expressed throughout the mouse brain, with notable expression in hippocampal areas (**Figure 5B**, left panel). Conversely, the most anti-correlated gene, osteopontin (*Spp1*), is predominantly expressed in hindbrain and olfactory areas, which are largely spared of τ pathology (**Figure 5B**, right panel). We make special mention of the nuclear receptor ROR-beta (*Rorb*) gene, which is a layer-4 neocortical neuron marker and also distinguishes a subpopulation of selectively vulnerable entorhinal cortex neurons[6]. As shown in the middle panel of **Figure 5B**, *Rorb* is broadly expressed throughout layer 4 of the neocortex; however, with the notable exception of the unseeded Hurtado study[20], correlations to τ pathology were low in magnitude and trended negative, similar to the neocortical glutamatergic neurons in the Yao, *et al.* dataset (**Figure 2A**).

When we attempted to model τ pathology as a function of AD risk genes using a similar variable selection procedure as above, 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 S7** and **S8**). Moreover, for two of the datasets (DS7 and DS9 110), we were unable to find a collection of genes that could predict τ pathology in a statistically significant sense ($R^2 = 0.21$ 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 S9** and **Table S6**), demonstrating that these results are robust to variable selection method. We conclude that the cell-type feature space contains richer spatial information for predicting regional susceptibility to τ than that of the AD risk genes.

2.5 Unseeded mouse tauopathy model exhibits differences in regional vulnerability over time

Of special significance among the tauopathy datasets within this study is the Hurtado, *et al.* mouse model[20], which developed τ pathology entirely endogenously. Therefore, unlike the other experiments, it is possible to examine early tangle accumulation without the confound of an injected seed. We fit each of the four quantified timepoints individually using the same BIC-based model selection procedure as above and found strongly statistically significant associations between model predictions and observed τ pathology (**Figure 6A** and **Table S7**). We also achieved excellent qualitative correspondence between observed (**Figure 6B**) and predicted (**Figure 6C**) τ deposition patterns, capturing early entorhinal and caudolateral neocortical involvement that is followed by the progressive spread to and accumulation within other limbic and neocortical structures. Surprisingly, we found that the most informative feature in each of the linear models differed between each timepoint. At two months, the regions exhibiting mild τ pathology were the entorhinal, ectorhinal, perirhinal, temporal association, and visual areas; correspondingly, the single-most important cell type, L3 IT ENT, is localized almost entirely within the entorhinal cortex

(**Figure 6D**, left panel). We note that the other BIC-selected cell type for the two-month linear model, L6b/CT ENT, is also an entorhinal-derived neuron. As pathology spreads profusely at four and six months, the most important cell types are Sst and CT SUB neurons, respectively (**Table S7**). Each is strongly localized to the forebrain, with CT SUB being more restricted to neocortical and hippocampal regions (**Figure 6D**, middle panels). Interestingly, oligodendrocytes become the most significant cell type at the last quantified time point, with a strikingly different distribution from observed pathology (**Figure 6D**, right panel). It also is the only cell type of the four to be a negative predictor of τ pathology (**Figure 2A** and **Table S7**). Overall, these results suggest that, insofar as cell-type-based mechanisms contribute to τ accumulation, the cell types that are involved are highly dynamic and span multiple taxonomic classes of neural cells.

3 Discussion

3.1 Summary of major findings

In the present study, we inferred the whole-brain distributions of 42 recently characterized neuronal and nonneuronal cells[13,18], and then used these data to evaluate selective regional vulnerability to τ pathology in twelve PS19-background mouse models (**Table S3**). We summarize our broad findings as follows:

1. Only hippocampal glutamatergic neurons tended to to be positively correlated on average with end-timepoint τ deposition, although remarkable diversity exists both between cell types of a given class and between tauopathy datasets (**Figure 2**; **Figures S2** and **S4**). We found similar results when comparing the regional expression of AD risk genes and τ pathology (**Figure 5**).

2. While correlations were pronounced for many individual cell types, no cell type exhibited a mean correlation across tauopathy datasets that exceeded 0.4 in magnitude, nor did any single correlation exceed 0.7. This suggests that the endemic susceptibility of any given cell type cannot solely explain the observed τ distributions in any given experiment or in PS19 mouse tauopathy as a whole. In other words, differences in τ burden across regions must be due to their cell-type compositions more holistically, cell-type-extrinsic factors, or both.

3. We were able to construct linear models of cell-type distributions that exhibited strongly statistically significant correspondences with regional τ pathology from all twelve experiments (**Figure 4** and **Table 1**). We found that cell-type-based models outperformed AD-risk-gene-based models using either a BIC feature selection criterion (**Figure S7** and **Table 1**) or using the top five most correlated features as predictors (**Figure S6**, **S9**, and **Table S6**), suggesting that the cell-type feature space was more robust than that of the AD risk genes

4. Cell-type linear models were statistically significantly associated with τ pathology in an unseeded mouse model at all quantified time points (**Figure 6**). Interestingly, we found that the most relevant cell types in the linear models differed between time points.

We elaborate on some of the implications of these results below.

3.2 How do cell types govern tau-related vulnerability or resilience?

The current study aimed to quantitatively uncover the cellular correlates of brain-wide τ-related vulnerability. As such, it is not a suitable framework to address mechanistic questions about *how* various cell types mediate regional susceptibility to τ pathology. Nevertheless, our findings generally agree with previous work investigating contributions of individual types of cells. Below we discuss several notable cell types identified by the present work in the context of experimental and clinical evidence.

3.2.1 Limbic glutamatergic projection neurons

Overall, we found that the most vulnerable class of neurons within the Yao, *et al.* dataset were glutamatergic projection neurons located in the hippocampal formation, with the CA1-ProS neuron having the single-highest positive correlation with regional τ pathology across the twelve experiments (**Figure 2A** and **B**). Noteworthy is that although many of these experiments contained hippocampal seeding sites (**Table S3**), all correlations presented within this study exclude these regions, indicating that seeding alone cannot explain why CA1-ProS neurons in particular are most strongly associated with τ. This finding is consistent with prior work identifying these neurons as particularly vulnerable to decreased glucose and oxygen delivery through the vasculature and thus to energy deprivation[27]. More recently, *in situ* mapping of single cells in the limbic areas of a mouse model of Alzheimer's disease demonstrated that excitatory CA1 neurons were particularly vulnerable to pathological τ[7]. Combined with our whole-brain cell-type-mapping through MISS and meta-analysis of twelve different experimental mouse models, this suggests that one consistent hallmark of tauopathy may be the disease of CA1 glutamatergic neurons.

It is well known that one of the earliest regions to exhibit τ pathology in AD is the entorhinal cortex (EC)[1]. More recently, scRNAseq performed on postmortem AD pateints identified glutamatergic neurons in layer II of the entorhinal cortex[4], and in particular those expressing the gene *Rorb*[6], are especially susceptible to τ. However, although Yao, *et al.* dataset contains several subclasses of neuron annotated as originating from layer II of the EC (**Table S1**), none of these exhibited strong correspondence to regional τ deposition across mouse experiments (**Figure 2A**), nor did they feature in the multivariate models (**Figure 4B** and **Figure S6B**). While we acknowledge that differences between the mouse disease models explored here and clinical phenotype may partly explain this discrepancy (see **Discussion: Limitations** below), we also note that neither our maps for the *Rorb* gene (**Figure 5B**) nor those of the three layer-II EC neurons in the Yao, *et al.* dataset are entorhinal-cortex-specific. This suggests homologs of these neurons exist outside of the sampled region, and furthermore, that neither gene expression nor cell-type identity is sufficient to explain why the entorhinal cortex is specifically afflicted in AD.

3.2.2 GABAergic neurons

While most work has explored the effects of τ pathology on excitatory neurons, the effects of AD-related pathology on GABAergic neurons has become a focus of study in recent years. In general, the levels of glutamic acid decarboxylase (GAD), an enzyme essential for producing GABA, is reduced in the brains of AD patients, albeit with significant study-to-study and region-to-region heterogeneity[28]. Other studies have pointed to an association between GABAergic synaptic dysfunction and AD-related pathological changes[29], including in response to τ[30,31]. GABAergic signaling and the expression of different GABA receptor subunits is known to be altered in AD, both in response to Aβ and τ pathology; however, it is unclear how these changes are related to pathogenesis[28]. More direct evidence of the vulnerabilty of GABAergic neurons comes from mouse models, which showed decreased numbers of *Pvalb*+ and *Sst*+ interneurons in various hippocampal regions as well as colocalization of pathological τ with these cell types[32,33]. These findings appear to be corroborated in humans by SEA-AD, the first multimodal cell atlas in AD, which profiled the middle temporal gyrus (MTG) of 84 human donors with AD and associated the absolute numbers of different types of cells assessed by single-nucleus RNAseq (snRNAseq) with disease severeity; of the GABAergic neurons, subtypes of *Pvalb*+ and *Sst*+ interneurons were among the most prominently affected[34].

Overall, we found that GABAergic neurons as a class were on the whole negatively associated with end-timepoint τ pathology at a whole-brain level across mouse experiments (**Figure 2B**), although several types exhibited consistently positive correlations (**Figure 2A**). Furthermore, while we consistently found that the distributions of *Sst*+ interneurons were positively correlated with τ pathology at a whole-brain level, *Pvalb*+ interneurons were strikingly *negatively* correlated (**Figure 2A**). The predominant proximate cause for the anti-correlation observed for *Pvalb*+ interneurons is that this cell type is present in high densities throughout the neocortex[8], while τ pathology in most experiments is largely absent from the neocortex (**Figure S2**). While we cannot fully discount the possibility of seed-based effects (see **Limitations** below), a likely contributor is the fact that patterns of τ deposition are strongly tied to the biochemical properties of the τ strain itself[23]. Recent studies demonstrating that specific phosphoepitopes of AD-τ can be used to distinguish AD from other tauopathic dementias in clinical patients suggests such a strain-specific hypothesis of AD pathogenesis[35–37]. Notably, within the present work, we found that the Hurtado, *et al.* unseeded mouse model, which exhibited comorbid Aβ pathology[20] and prominent neocortical τ deposition (**Figure 6B**), also demonstrated positive associations with both *Pvalb*+ and *Sst*+ interneurons (**Figure 2A**). Therefore, it is possible that interneuron vulnerability to τ may also in fact be strain-specific.

3.2.3 Oligodendrocytes

One of our most striking results is the disproportionate role played by oligodendrocytes in regional *resilience* to τ. In addition to being the cell type with the greatest magnitude of association with end-timepoint τ pathology across experiments (**Figure 2A** and **Figure S1**), but it was also the most prominently featured cell type in combined linear models of τ pathology (**Figure 4**) and the most important predictor of end-timepoint pathology in the Hurtado dataset (**Figure 6D**). Together, these results consistently indicate that regions with high baseline densities of oligodendrocytes are especially resistant to τ pathology.

The primary role of oligodendrocytes is the production and maintenance of the myelin sheath surrounding axonal projections, which acts as an insulator and facilitates fast signal conduction between neurons[38]. While there is strong neurobiological evidence for oligodendrocyte involvement in tauopathic disease, whether they are protective or harmful is the subject of some debate. Pathway analysis of gene expression changes in human AD found that oligodendrocyte-specific modules were among the most strongly disrupted[39]. AD patients exhibit detectable white matter lesions, suggesting a direct role for oligodendrocyte dysfunction in AD-related pathological changes[40]. Furthermore, although the mechanisms underpinning τ uptake in oligodendrocytes remain poorly understood[41], oligodendrocytes strongly co-localize with τ inclusions in mouse tauopathy and may facilitate τ seeding and propagation[42]. A mouse model in which neurons do not express τ still exhibited inter-regional spread along white matter tracts, a process that was mediated by oligodendrocytes[43]. 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. Put another way, regions with higher numbers of oligodendrocytes may be better able to tolerate pathological insults.

3.3 Trans-synaptic spread as an alternative mechanism of SV

Worth underscoring is that, while correlations were pronounced for many individual cell types, no cell type exhibited a mean correlation across tauopathy datasets that exceeded 0.4 in magnitude, nor did any single correlation exceed 0.7 (**Figure 2A**). This suggests that the endemic susceptibility of any given cell type cannot solely explain the observed τ distributions in any given experiment or in PS19 mouse tauopathy as a whole. In other words, differences in τ burden across regions must be due to their cell-type compositions more holistically, cell-type-extrinsic factors, or both.

Another hypothesis about τ pathophysiology, one that is not mutually exclusive with selective vulnerability, is that misfolded conformers of τ travel between regions trans-synaptically and intracellularly via white matter tracts in a prion-like manner. Seminal work by Clavaguera, *et al.* demonstrated that in mice expressing humanized τ, the injection of pathological τ was sufficient to induce misfolding and aggregation of endogenous τ both in the seeding location as well as distal regions[44]. More direct evidence comes from experiments using microfluidically isolated neurons that show τ uptake in one neuron, translocation down its axonal processes, and synaptic release and reuptake in a second neuron[45–49]. Although this has not been directly observed in humans, the fact that τ has been demonstrated to be transported trans-synaptically via exosomal release and reuptake[50] and the presence of τ-containing exosomes in the plasma of AD patients[51,52] is suggestive that this is a key process in τ pathogenesis. Furthermore, mathematical models of network-based spread of pathology both in humans and in mouse models of disease have been shown to be highly predictive[53–58].

It is challenging to disambiguate the effects of cell-type selective vulnerability and trans-synaptic spread in practice because some of the most vulnerable cell types are precisely the excitatory neurons along whose projections pathological τ propagates to neurons in distant gray-matter regions. However, given the strong evidence in support of both hypotheses, it is likely that endogenous factors mediate trans-synaptic spread in a cell-type-specific fashion. For instance, it is possible that certain vulnerable excitatory neurons have a lesser capacity to clear τ pathology as it develops, and as a result τ accumulates in these neurons and then spreads along their projections. One avenue of exploration is to combine these effects directly in a mathematical model; for instance, recent work demonstrated the efficacy of including microglial signature genes in a network spread model of mouse tauopathy in a single model, Nexis[59]. Further computational, experimental, and clinical work will be required to fully explore the mechanisms underpinning τ pathogenesis in the context of selective vulnerability and trans-synaptic spread.

3.4 Limitations

We must state several important caveats when interpreting our results. First, since both the scRNAseq dataset from Yao, *et al.*[18] and the AGEA[17] profile brain tissue in healthy mice, we cannot directly correlate the cell-type distributions in a disease condition to τ deposition, let alone in the same individual; we do note, however, that with the exception of the mutant P301S τ transgene in the PS19 mouse line, the mice used in the tauopathy and the gene expression studies all have the same genetic background (C57BL/6). We are also limited by the tauopathy dataset resolution, which was only quantified at the regional level (**Table S3**). We therefore can only claim that our analyses can only associate τ vulnerability with respect to *baseline* cell-type distributions. Furthermore, although more precise imaging technologies such as STARmap Plus are starting to allow for the simultaneous characterization of cell types and τ pathology *in situ*, they have not been scaled up to perform whole-brain imaging across different model organisms[7,11]. By contrast, MISS does not suffer from these limitations and therefore we can ask broader questions about the nature of regional vulnerability with respect to cell-type distributions in tauopathy more generally.

We also note that transgenic mouse models of neurodegenerative disease, while an invaluable tool for exploring pathological processes in ways that cannot be accessed clinically, carry significant limitations. First, the requirement of an exogenous τ seed for PS19 mice hinders the ability to study early disease processes, constraining us to explore end-timepoint associations between cell-type distributions and τ deposition. We have taken care to minimize the effect of the seed by removing it from the analyses presented herein and we note that τ pathology changes appreciably between time points (**Figure S3**), but we cannot entirely eliminate the effects of artificially inducing pathology. We also observe vast heterogeneity between tauopathy models as a function of seeding location and the nature of the seed itself (**Figure S2** and **Table S3**), which consequently produces wide variations in the correlations we observe for each cell type across datasets (**Figure 2**). However, we believe that this is also strength of the present work, because those cell types that have consistent effects across datasets *despite* the vast differences in experimental conditions suggest generalizable relationships between those types and their vulnerability to pathological τ. We have also included the doubly transgenic Hurtado, *et al.* mouse model (**Figure 6**)[20], which does not suffer the above limitations as it developed τ pathology entirely endogenously. Notably, we found that linear models of cell-type distributions could accurately capture the disease process at all timepoints both qualitatively and quantitatively. Given the significant differences in experimental conditions between this study and the others explored here, as well as the fact that the Hurtado model developed

endogenous Aβ pathology[20], the fact that there is consistency about the involvement of oligodendrocytes and microglia in the disease process between these datasets is notable (**Figure 2**).

3.5 Future directions

Although we have taken care to analyze a wide range of datasets, extending this work to other mouse tauopathy models will allow us to increase the robustness of our claims and potentially yield new insights. We would also like to apply the same analysis to Aβ datasets, for instance using the popular 5xFAD model as a baseline as we have used PS19 in the present study. Comparing and contrasting the cellular correlates of regional vulnerability to amyloidopathy with respect to tauopathy would have great relevance to understanding the various aspects of AD pathophysiology. As mentioned above, further step in deepening our understanding of AD and tauopathies in general would be to create joint models of network-based spread and cell-type-based regional vulnerability. Such a model, Nexis, has already been developed and implemented to jointly explore network spread of τ and its interactions with the disease-associated microglia marker *Trem2*, demonstrating that the best mathematical models of τ pathology in mice incorporate both mechanisms[59]. We would also like to generate MISS-inferred cell-type distributions using available scRNAseq data and the human gene expression atlas from the AIBS[60] and then repeat these analyses, where τ and Aβ pathology data come from positron emission tomography (PET) imaging of human AD patients. This more direct comparison will allow us for the first time answer questions about cell-based selective vulnerability at a whole-brain level, which are currently cannot be addressed with current imaging technologies.

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 1.1 million individual cells sampled comprehensively throughout the neocortex and hippocampal formation at 10x sequencing depth[18]. We used the provided trimmed means by cluster dataset (available here: https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x), as MISS 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[13], 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)[17]. 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 throughout as "experiments"): BoludaCBD[19], BoludaDSAD[19], DS4[23], DS6[23], DS6 110[23], DS7[23], DS9[23], DS9 110[23], Hurtado[20], IbaHippInj[21], IbaStrInj[21], and IbaP301S[22]. We summarize the key elements of each experiment in **Table S3**; refer to the original publications for further methodological details. 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 MISS

We applied the Matrix Inversion and Subset Selection (MISS) algorithm to the Yao, *et al.* scRNAseq dataset[18] and the AGEA ISH dataset[17] as was described previously[13]. 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-subsetted spatial ISH-based gene expression matrix is regressed on the gene-subsetted 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[13], 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[61]. 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 criteria.

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[62] to estimate the incremental noise added per gene, as assessed by the mean-squared error between the given E and the predicted $\hat{E}$ (i.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 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 gene i and the jth gene 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, $n_G = 1300$. This produces matrices E_{red} and C_{red}, which only contain the 1300 rows corresponding to the genes in S_{n_G}.

4.2.2 Matrix inversion

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

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

where e_{red}^k 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**.

4.3 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[24]. 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[20] and IbaP301S[22] 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 obvious, 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 $n_{\text{types}} + 1$, 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 is identical to the model of dimensionality 6 using the first method.

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 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[13]. 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. 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).
3. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337, DOI: 10.1038/s41586-019-1195-2 (2019).
4. 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).
5. 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).
6. 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).
7. 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).
8. 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).
9. Li, X., Xing, Y., Schwarz, S. T. & Auer, D. P. Limbic grey matter changes in early Parkinson's disease. *Hum. Brain Mapp.* **38**, 3566–3578, DOI: 10.1002/hbm.23610 (2017).
10. 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).

11. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691, DOI: 10.1126/science.aat5691 (2018).

12. 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).

13. 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).

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

15. 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).

16. 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).

17. 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).

18. 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).

19. 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).

20. 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).

21. 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).

22. 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).

23. 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).

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

25. 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).

26. 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).

27. 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).

28. Govindpani, K., Vinnakota, C., Waldvogel, H., Faull, R. & Kwakowsky, A. Towards a Better Understanding of GABAergic Remodeling in Alzheimer's Disease. *Int. J. Mol. Sci.* **18**, 1813, DOI: 10.3390/ijms18081813 (2017).

29. 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).

30. 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).

31. 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).

32. Levenga, J. *et al.* Tau pathology induces loss of GABAergic interneurons leading to altered synaptic plasticity and behavioral impairments. *Acta Neuropathol. Commun.* **1**, 34, DOI: 10.1186/2051-5960-1-34 (2013).

33. Loreth, D. *et al.* Selective degeneration of septal and hippocampal GABAergic neurons in a mouse model of amyloidosis and tauopathy. *Neurobiol. Dis.* **47**, 1–12, DOI: 10.1016/j.nbd.2012.03.011 (2012).

34. Gabitto, M. *et al.* Integrated multimodal cell atlas of Alzheimer's disease. *Res. square* DOI: 10.21203/rs.3.rs-2921860/v1 (2023).

35. 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).

36. 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).

37. 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).

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

39. 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).

40. 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).

41. 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).

42. 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).

43. Narasimhan, S. *et al.* Human tau pathology transmits glial tau aggregates in the absence of neuronal tau. *J. Exp. Medicine* **217**, DOI: 10.1084/jem.20190783 (2020).

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

45. 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).

46. 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).

47. 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).

48. 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).

49. 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).

50. Wang, Y. *et al.* The release and trans-synaptic transmission of Tau via exosomes. *Mol. Neurodegener.* **12**, 5, DOI: 10.1186/s13024-016-0143-y (2017).

51. Perrotte, M. *et al.* Profile of pathogenic proteins in total circulating extracellular vesicles in mild cognitive impairment and during the progression of Alzheimer's disease. *Neurobiol. Aging* **86**, 102–111, DOI: 10.1016/j.neurobiolaging.2019.10.010 (2020).

52. Zhao, Y., Gu, Y., Zhang, Q., Liu, H. & Liu, Y. The Potential Roles of Exosomes Carrying APP and Tau Cleavage Products in Alzheimer's Disease. *J. Clin. Medicine* **12**, 1883, DOI: 10.3390/jcm12051883 (2023).

53. Iturria-Medina, Y. Anatomical Brain Networks on the Prediction of Abnormal Brain States. *Brain Connect.* **3**, 1–21, DOI: 10.1089/brain.2012.0122 (2013).

54. Mezias, C., LoCastro, E., Xia, C. & Raj, A. Connectivity, not region-intrinsic properties, predicts regional vulnerability to progressive tau pathology in mouse models of disease. *Acta Neuropathol. Commun.* **5**, 61, DOI: 10.1186/s40478-017-0459-z (2017).

55. Pandya, S., Mezias, C. & Raj, A. Predictive Model of Spread of Progressive Supranuclear Palsy Using Directional Network Diffusion. *Front. Neurol.* **8**, DOI: 10.3389/fneur.2017.00692 (2017).

56. Raj, A., Kuceyeski, A. & Weiner, M. A Network Diffusion Model of Disease Progression in Dementia. *Neuron* **73**, 1204–1215, DOI: 10.1016/j.neuron.2011.12.040 (2012).

57. Raj, A. *et al.* Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer's Disease. *Cell Reports* **10**, 359–369, DOI: 10.1016/j.celrep.2014.12.034 (2015).

58. Weickenmeier, J., Kuhl, E. & Goriely, A. Multiphysics of Prionlike Diseases: Progression and Atrophy. *Phys. Rev. Lett.* **121**, 158101, DOI: 10.1103/PhysRevLett.121.158101 (2018). 1804.01958.

59. 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).

60. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399, DOI: 10.1038/nature11405 (2012).

61. 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).

62. 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).

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

Figures

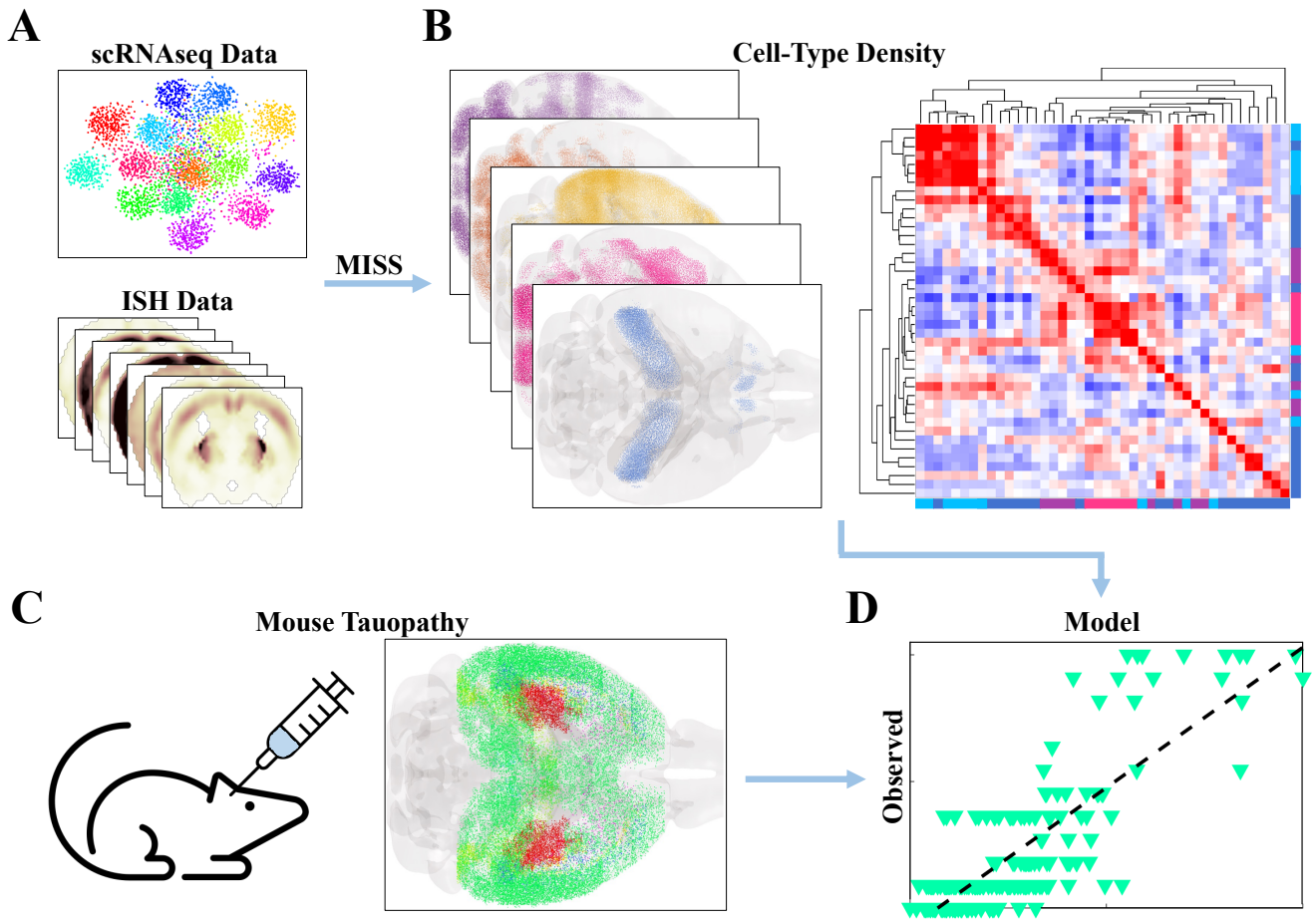

Figure 1. Data analysis pipeline. Neocortical and hippocampal scRNAseq data from Yao, *et al.* and the Allen Gene Expression Atlas (AGEA) (**A**) are used to infer cell-type densities (**B**) using the Matrix Inversion and Subset Selection (MISS) pipeline. Example cell-type densities and the spatial cross-correlation matrix between all cell types are shown in **B**. These cell-type distributions were then used to fit mouse tauopathy data (**C**) that were gathered from a collection of studies using BIC-based model selection (**D**)

.

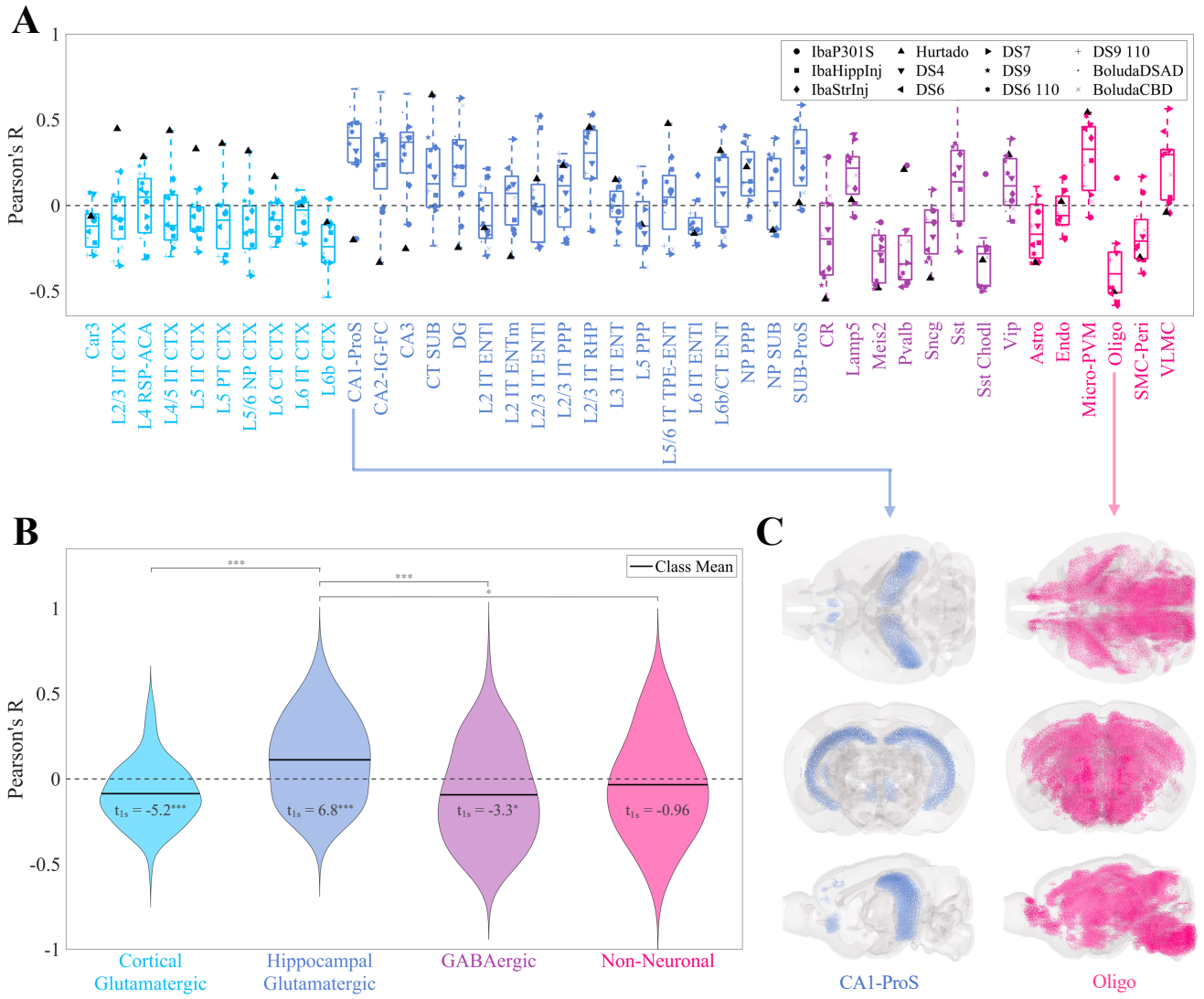

Figure 2. Univariate analysis of the relationship between cell-type density and end-timepoint pathology across mouse tauopathy models. A. Box plot of Pearson's correlation coefficients between the MISS-inferred distributions of the 42 cell types of the Yao, *et al.* dataset and the end tau pathology 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[18].

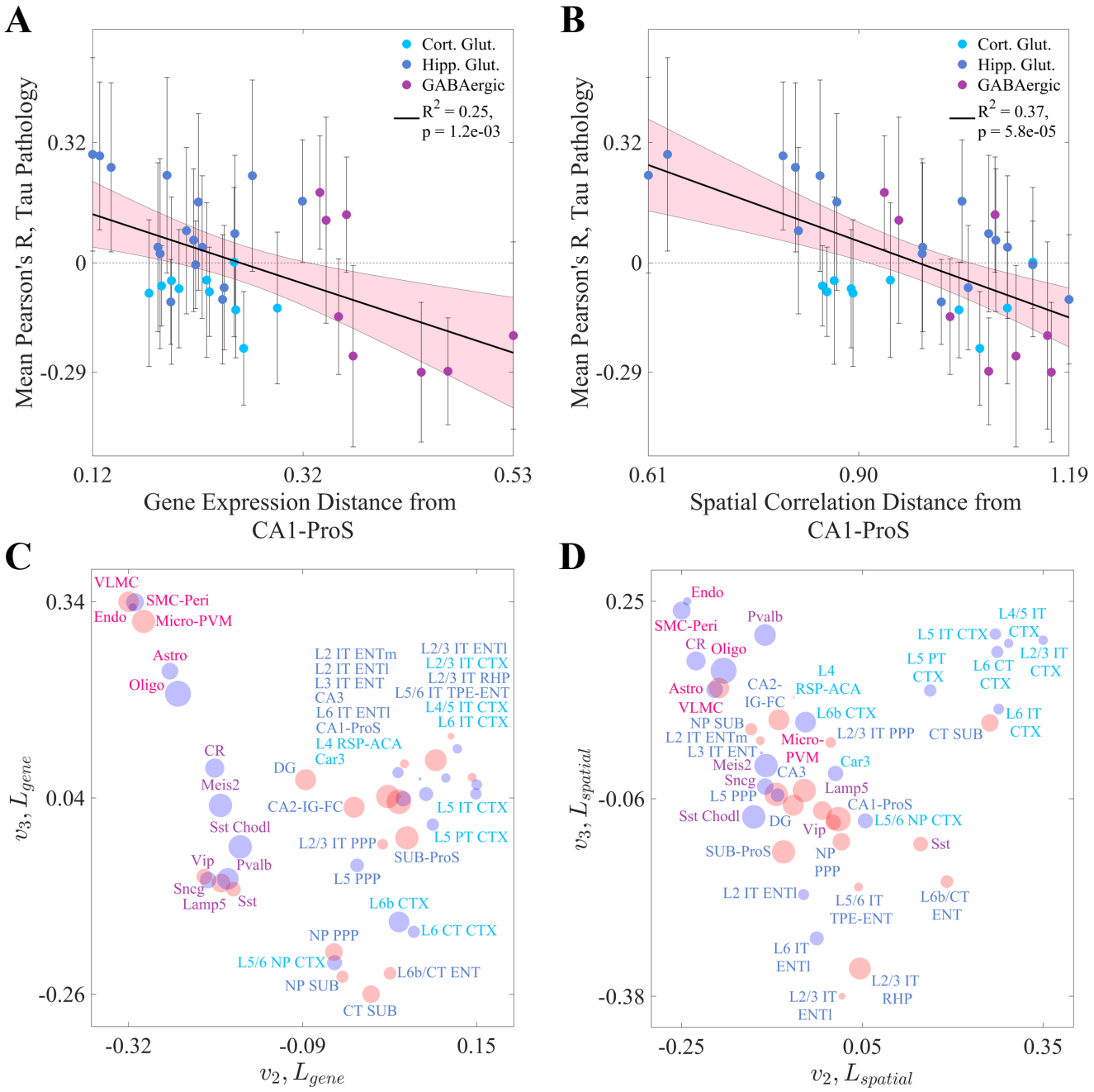

Figure 3. Analysis of inter-cell-type distance in gene expression and spatial distribution. A. Linear models showing the negative relationship between gene expression correlation distance (1 - R) and correlation with end time-point tau pathology for neuronal cell types in the Yao, *et al.* dataset with 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[18]. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$.

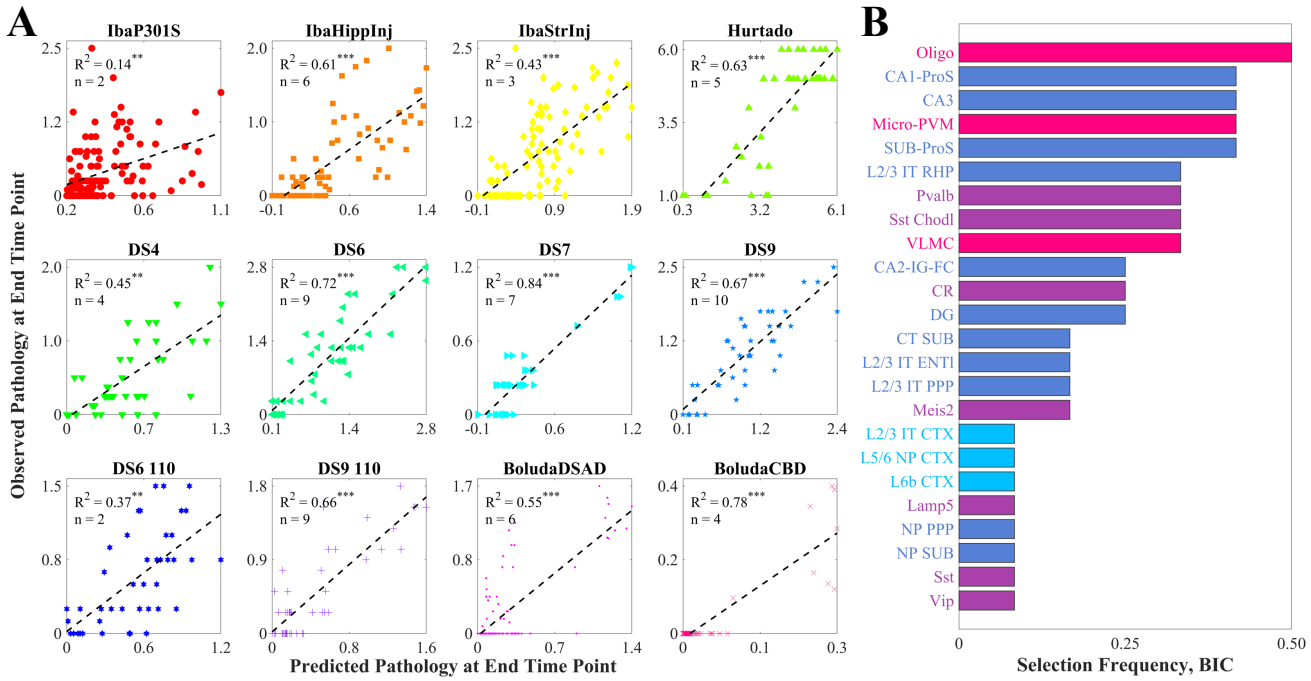

Figure 4. Multivariate analysis of end-timepoint pathology. A. Scatter plots of the optimal cell-type-based models of tau pathology at the end time points for each of the nine mouse tauopathy studies, along with their associated R^2 values and the number of cell types chosen using the BIC. *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$. **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.

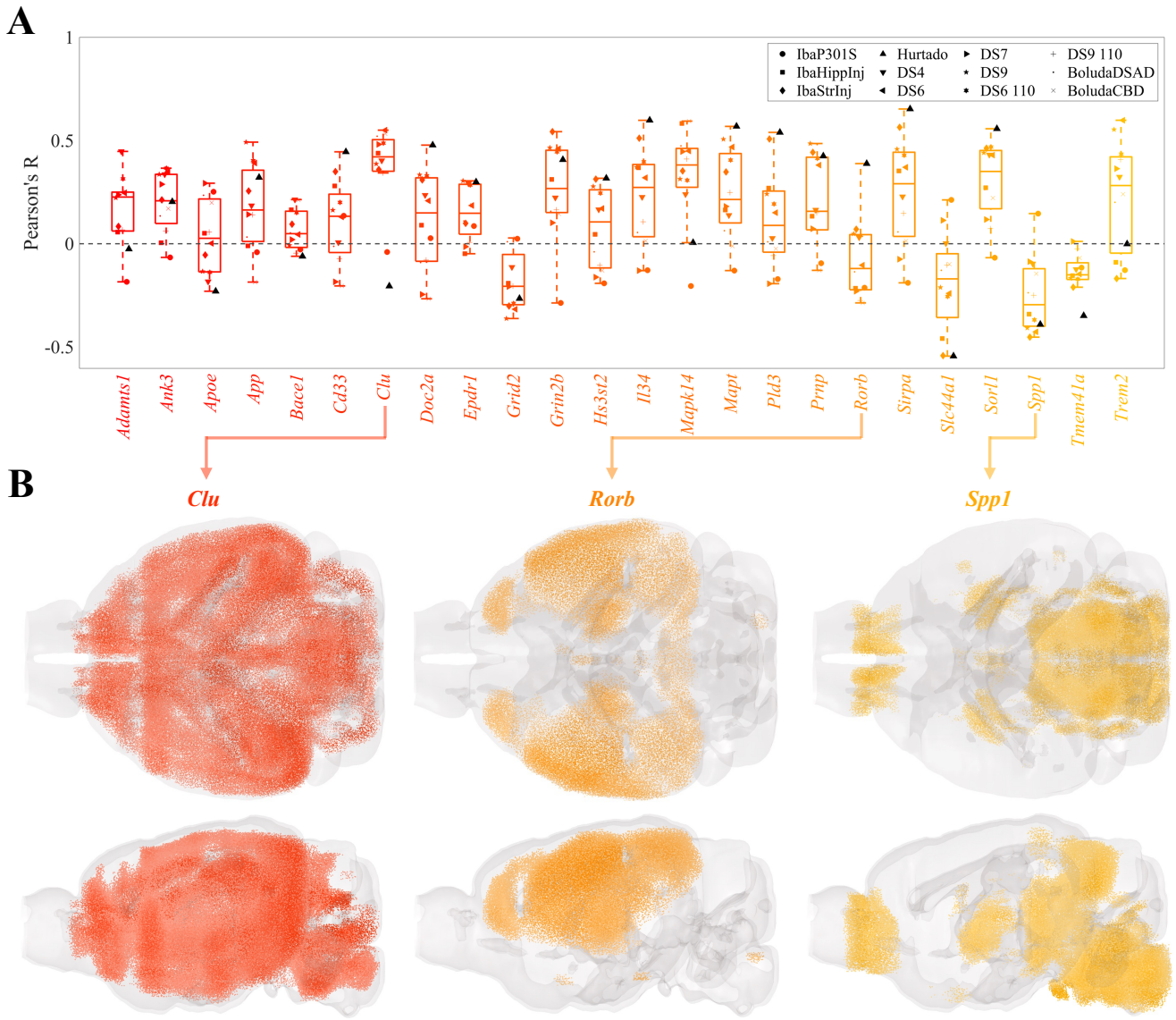

Figure 5. Univariate analysis of the relationship between AD risk genes and end-timepoint pathology across mouse tauopathy models. A. Box plot of Pearson's correlation coefficients between the MISS-inferred distributions of the 24 AD risk genes and the end tau pathology for 12 mouse tauopathy datasets. The list of AD risk genes is provided in **Table S5**. **B.** Glass brain renderings of genes of interest.

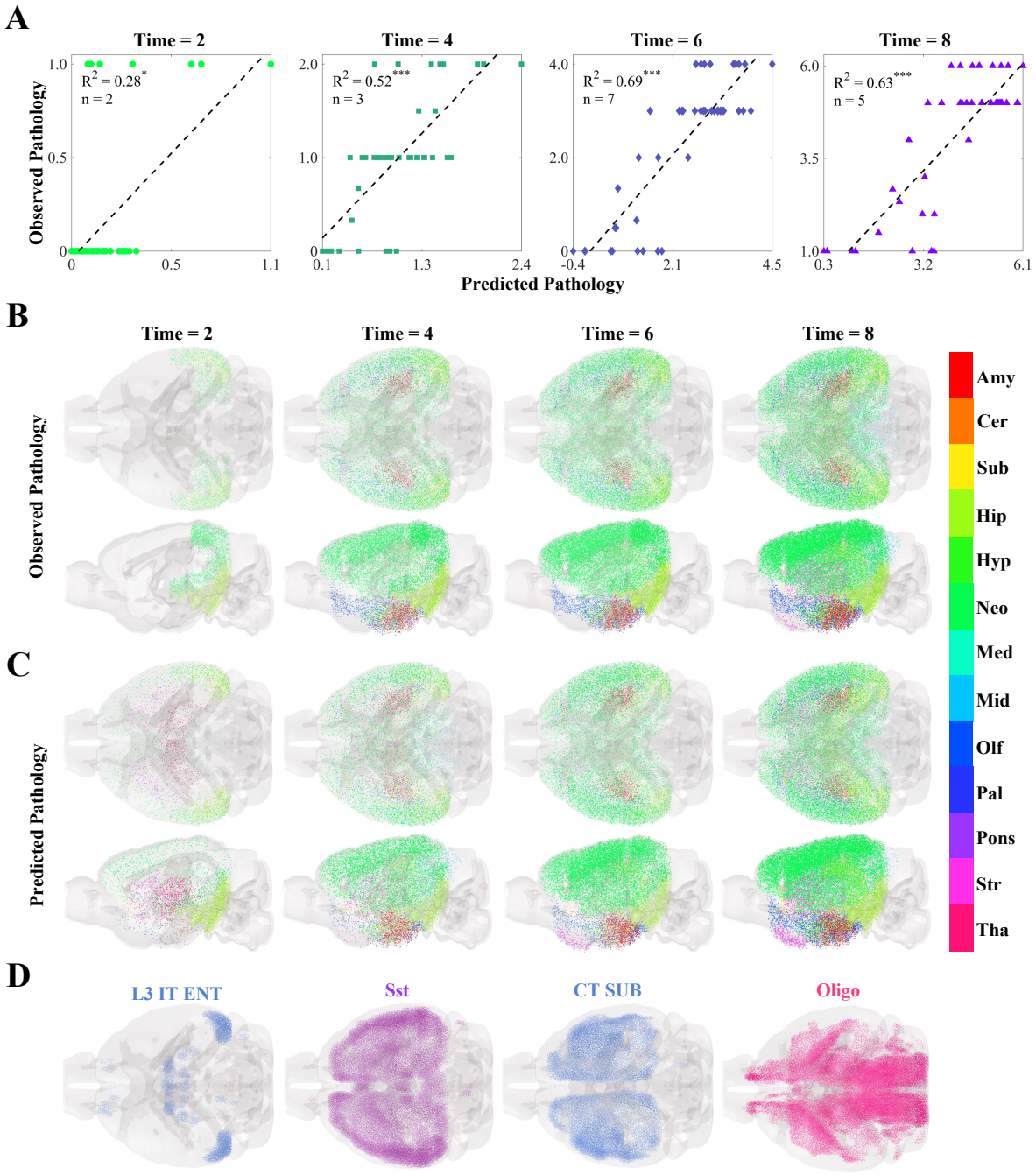

Figure 6. Multivariate analysis of the temporal progression of pathology in an unseeded mouse tauopathy model. A. Scatter plots of the optimal cell-type-based models of tau pathology for each time point in the unseeded Hurtado, *et al.* mouse tauopathy dataset[20], along with their associated R^2 values 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[18].

Tables

Dataset	Cell types (BIC)			AD risk genes (BIC)		
	R^2	n	BIC	R^2	n	BIC
IbaP301S[22]	0.14^{**}	2	180.2	$\mathbf{0.28}^{***}$	**8**	**176.4**
IbaHippInj[21]	$\mathbf{0.61}^{***}$	**6**	**81.0**	0.43^{***}	3	109.0
IbaStrInj[21]	$\mathbf{0.43}^{***}$	**3**	**163.5**	0.31^{***}	1	174.4
Hurtado[20]	$\mathbf{0.63}^{***}$	**5**	**153.3**	0.41^{***}	1	163.8
DS4[23]	$\mathbf{0.45}^{**}$	**4**	**50.8**	0.35^{**}	2	52.5
DS6[23]	$\mathbf{0.72}^{***}$	**9**	**79.3**	0.55^{***}	2	82.0
DS7[23]	$\mathbf{0.84}^{***}$	**7**	**-28.7**	0.21	1	23.2
DS9[23]	$\mathbf{0.67}^{***}$	**10**	**70.3**	0.39^{**}	2	75.4
DS6 110[23]	$\mathbf{0.37}^{**}$	**2**	**47.3**	0.32^{*}	2	50.4
DS9 110[23]	$\mathbf{0.66}^{***}$	**9**	**49.9**	0.23	2	67.1
BoludaDSAD[19]	$\mathbf{0.55}^{***}$	**6**	**68.0**	0.34^{***}	2	87.8
BoludaCBD[19]	$\mathbf{0.78}^{***}$	**4**	**-167.7**	0.54^{***}	4	-127.3

Table 1. BIC linear regression model statstics. Statistics corresponding to the linear models shown in **Figure 4** and **Figure S8**. Bold font indicates the best model by the Bayesian Information Criterion (BIC). *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$.