Dysregulated coordination of *MAPT* exon 2 and exon 10 splicing underlies different tau pathologies in PSP and AD

Kathryn R. Bowles[1,2], Derian A. Pugh[1,2], Laura-Maria Oja[1,2], Benjamin M. Jadow[1,2], Kurt Farrell[2,3], Kristen Whitney[2,3], Abhijeet Sharma[2,8], Jonathan D. Cherry[4,5,6,7], Towfique Raj[1,2,8,9], Ana C. Pereira[2,8,9], John F. Crary[2,3,#] & Alison M. Goate[1,2,8*]

[1]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[2]Ronald M. Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[3]Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[4]Boston University Alzheimer's Disease and CTE Center, Boston University School of Medicine, Boston, MA, United States of America
[5]Department of Neurology, Boston University School of Medicine, Boston, MA, United States of America
[6]VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA, United States of America
[7]Department of pathology and laboratory medicine, Boston University School of Medicine, Boston, MA, United States of America
[8]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[9]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

[#]Neuropathology Brain Bank & Research Core, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

*Correspondence: alison.goate@mssm.edu

ABSTRACT

Understanding regulation of *MAPT* splicing is important to the etiology of many nerurodegenerative diseases, including Alzheimer disease (AD) and progressive supranuclear palsy (PSP), in which different tau isoforms accumulate in pathologic inclusions. *MAPT*, the gene encoding the tau protein, undergoes complex alternative pre-mRNA splicing to generate six isoforms. Tauopathies can be categorized by the presence of tau aggregates containing either 3 (3R) or 4 (4R) microtubule binding domain repeats (determined by inclusion/exclusion of exon 10), but the role of the N terminal domain of the protein, determined by inclusion/exclusion of exons 2 and 3 has been less well studied. Using an unbiased correlational screen in human brain tissue, we observed coordination of *MAPT* exons 2 and 10 splicing. Expression of exon 2 splicing regulators and subsequently exon 2 inclusion are differentially disrupted in PSP and AD brain, resulting in the accumulation of 1N4R isoforms in PSP and 0N isoforms in AD temporal cortex. Furthermore, we identified different N-terminal isoforms of tau present in neurofibrillary tangles, dystrophic neurites and tufted astrocytes, indicating a role for differential N-terminal splicing in the development of disparate tau neuropathologies. We conclude that N-terminal splicing and combinatorial regulation with exon 10 inclusion/exclusion is likely to be important to our understanding of tauopathies.

INTRODUCTION

Ninety-five percent of all human multi-exonic genes are subject to alternative pre-mRNA splicing[1,2]. Correct regulation of this mechanism is essential for proteomic diversity by the production of multiple distinct isoforms from a single gene[3]. The microtubule-associated protein tau (*MAPT*) is a neuronally expressed gene consisting of 16 exons, many of which are differentially spliced within the central nervous system and peripheral tissues. Tau proteins are involved in axonal transport, synaptic plasticity, and stabilization of the microtubule network[4–6]. In the human brain, the splicing of *MAPT* exons 2, 3 and 10 results in the expression of six different isoforms, which can be parsed into two groups depending on their inclusion or exclusion of exon 10 (Figure 1A).

The ratio of tau isoform expression changes during human brain development[7,8], with only the shortest 0N3R isoform of tau being expressed in fetal brain[7–10]. Following birth, there is a sudden shift in the expression of both exons 2 and 10; exon 10 inclusion increases dramatically during the perinatal period[8] where it reaches a stable 3R:4R ratio of roughly 1:1[11–13], whereas exon 2 expression increases gradually throughout the first decade of life[8]. The reason for this shift and the function of different *MAPT* isoforms is not fully understood. However, 4R tau has an increased affinity for binding microtubules that results in their increased stabilization[14], therefore shorter 3R isoforms may allow for greater neuroplasticity during brain development. It has been proposed that different N-terminal isoforms may also contribute to microtubule stabilization[15], and that the inclusion of exons 2 and 3 results in the extension of the acidic region of tau, lengthening its projection domain[16], which in turn may increase the distance between microtubules and increase bundling[17].

Understanding *MAPT* splicing is of critical importance to the etiology of tauopathies, which are characterized by the presence of neuronal and/or astroglial tau aggregates. Primary tauopathies include frontotemporal dementia (FTD), progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), primary age-related tauopathy (PART) and Pick's disease (PiD), while other tauopathies, such as AD, are secondary to amyloid-beta (Aß) deposition. In several primary tauopathies the regulation of *MAPT* splicing is altered and mis-spliced isoforms are differentially incorporated into neurofibrillary tangles (NFTs) and pathogenic inclusions[18,19]. The most striking evidence in support of the importance of regulated splicing is the numerous synonymous and intronic *MAPT* mutations, such as S305S[20,21], IVS10+16[22] and N296N[23], which result in increased exon 10 inclusion, and the subsequent development of autosomal dominant FTD. The ability for these mutations to induce tau pathology in the absence of an altered amino acid sequence is indicative of the relevance of *MAPT* splicing to disease pathogenesis, and the importance of maintaining the correct tau isoform ratio.

Alternative splicing of *MAPT* exon 10 in both healthy and diseased brain has been well characterized, although studies examining exon 10 expression in AD have yielded inconsistent results[11,13,24–27]. In contrast, less is known about the regulation of exons 2 and 3, and there have been no studies directly assessing the contribution of N-terminal tau isoforms to primary tauopathies. To date, splicing of exon 10 has been associated with the function of several candidate splicing factors (SFs) and RNA binding proteins (RBPs), the most comprehensively investigated of which are the serine and arginine-rich family of splicing factors (SRSFs). Multiple SRSFs have been associated with both exon 10 inclusion and exclusion[28–31], as have Tra2β[19,29,32,33], FUS[33], RBM4[34], NOVA1[35] and hnRNPs E2 and E3[35,36]. However, there is limited evidence for many of these associations, and their impact on *MAPT* splicing lacks robust replication.

Here, we report a correlational screen for SFs/RBPs in human brain tissues that revealed novel genes associated with *MAPT* splicing, and uncovered coordinated regulation of *MAPT* exons 2 and 10. We found that the splicing factor *RSRC1* bound directly to *MAPT* pre-mRNA and was associated with both exon 10 exclusion and exon 2 inclusion. Furthermore, *RSRC1* was also differentially expressed in PSP and AD brain, suggesting a regulatory role for this SF in disease pathogenesis. Consistent with discordant *RSRC1* expression, we observed increased expression of exon 2 and exon 10-containing transcripts in PSP brain, and increased expression of 0N transcripts in AD brain, which correlate with the accumulation of different N-terminal tau isoforms in different neuropathological features of AD and PSP. We therefore conclude that differential expression of *MAPT* splicing regulators in AD and PSP brain results in the loss of coordination between *MAPT* exon 10 and exon 2 splicing. In turn, this results in the expression and accumulation of different N-terminal isoforms in each disease, which may underlie the development of different neuropathological features characteristic of AD and PSP.

RESULTS

MAPT alternative splicing differs by brain region and 17q21.31 haplotype

MAPT is alternatively spliced at exons 2, 3 and 10 in the human brain (Figure 1A), and expression of these exons across brain regions has previously been measured by microarray analyses[12]. However, accurate measurement of gene or exon expression by microarray may be impacted by the specificity of probe design and a narrow dynamic range for signal detection. We therefore chose to characterize the relative expression of *MAPT* exons in multiple human postmortem RNA-seq datasets from the AMP-AD consortium (Religious Orders Study Memory and Aging Project [ROSMAP; Synapse syn3219045] N = 450, Mount Sinai Brain Bank [MSBB;

Synapse syn3159438] N = 230, and the Mayo Clinic [MAYO; Synapse syn3157268, syn5550404] N = 276) (Figure 1B-F, Figure S1A-H). We observed a pattern of exon expression consistent with that previously described[12,37] and known isoforms expressed in brain (i.e., constitutive expression of exons 1, 4, 5, 7, 9, 11 and 12, little to no expression of exons 4a, 6 and 8, and variable levels of exon 2, 3 and 10), which was consistent across AMP-AD datasets and brain regions (Figure 1B, Figure S1A-C).

To assess proportional exon expression across different brain regions included in MSBB and MAYO data, we calculated percent spliced in (PSI) values for the alternatively spliced exons 2, 3 and 10 using MISO (Mixture of Isoforms) (Figure 1C-D). While there were no differences in PSI values across MSBB Brodmann regions BM10 (frontal pole), BM22 (superior temporal gyrus), BM36 (fusiform gyrus) and BM44 (inferior frontal gyrus; Figure 1C), there was significantly increased inclusion of all three exons in cerebellum compared to the temporal cortex in the MAYO cohort (Exon 2 $p = 5.46\text{x}10^{-16}$; Exon 3 $p = 2.74\text{x}10^{-8}$; Exon 10 $p = 1.54\text{x}10^{-9}$; Figure 1D). This is consistent with previous reports that suggest *MAPT* splicing in the cerebellum differs from the forebrain[11,12].

We then compared PSI values between the major 17q21.31 *MAPT* H1 and H2 haplotypes, and observed increased exon 3 inclusion in H2 haplotype carriers compared to H1 across most AMP-AD datasets and brain regions (MSBB H1H2 $p=1.14\text{x}10^{-09}$, MSBB H2H2 $p=1.09\text{x}10^{-13}$; MAYO H1H2 $p< 2\text{x}10^{-16}$, MAYO H2H2 $p=1.9\text{x}10^{-04}$, Figure 1E-F, Figure S1D-E), which has been previously reported in other RNA-seq and microarray datasets[12,38,39]. In comparison, we did not observe any difference in exon 2 or 10 inclusion between haplotypes (Figure 1E-F, Figure S1D-E). In contrast to previous reports[40], we did not find altered total *MAPT* expression between these haplotypes in any dataset or brain region (Figure S1F-H).

MAPT exon 2 and 10 splicing is coordinated by SF/RBP expression

In order to identify novel *MAPT* splicing regulators in human prefrontal cortex, we carried out a correlational analysis between the expression (RPKM) of 294 known splicing factors (SFs) and RNA binding proteins (RBP) described in Gerstberger et al 2014[41] (Table S1) with *MAPT* exon 2, 3 and 10 PSI values, using ROSMAP and MSBB datasets (Figure 1G, Figure S2A-G). We did not observe any significant correlations that withstood Bonferroni multiple test correction between *MAPT* exon 3 and any SF/RBP (Figure S2A), likely due to the very low expression of exon 3 in human brain. However, when we separated the data by H1/H2 haplotype, we observed stronger associations between exon 3 inclusion and SF/RBP expression in H2 homozygotes compared to H1 homozygotes in both datasets (Figure S2B-C), although these still did not pass multiple test correction due to low frequency of the H2 haplotype in these datasets. While we did not have the statistical power to pursue analysis of exon 3 splicing regulators, these data indicate there may be additional regulation of exon 3 in the context of H2, consistent with its increased expression on this background.

In contrast, there were no significant differences in SF/RBP and *MAPT* exon 2/10 PSI value associations between H1/H2 haplotypes, which were highly correlated across both haplotypes (ROSMAP exon 2 $R^2 = 0.81$, $p < 0.001$, exon 10 $R^2 = 0.89$, $p < 0.001$; MSBB exon 2 $R^2 = 0.67$ $p < 0.001$, exon 10 $R^2 = 0.66$ $p < 0.001$) (Figure S2D-E). Associations between PSI values and SF/RBP expression were also significantly correlated between AD cases and controls (ROSMAP exon 2 $R^2 = 0.97$ $p < 2.2\text{x}10^{-16}$, exon 10 $R^2 = 0.96$ $p < 2.2\text{x}10^{-16}$; MSBB exon 2 $R^2 = 0.89$ $p < 2.2\text{x}10^{-16}$, exon 10 $R^2 = 0.93$ $p < 2.2\text{x}10^{-16}$) (Figure S2F-G), indicating no differences in the

regulatory effects of SF/RBP expression on *MAPT* splicing in AD. We therefore focused our analyses on exons 2 and 10 using pooled AD/control and H1/H2 data.

Unsupervised hierarchical clustering of the resulting Pearson's correlation coefficients between exon PSI values and SF/RBP expression revealed that exons 2 and 10 clustered separately from each other and had distinct patterns of association with SF/RBP expression (Figure 1G). In order to identify robust *MAPT* splicing regulator candidates, we selected SF/RBPs with significant (Bonferroni-corrected p-value < 0.05) associations with *MAPT* exons 2 or 10 in the same direction across the three most anatomically similar datasets; ROSMAP (prefrontal cortex), MSBB BM10 (frontal pole) and MSBB BM44 (inferior frontal gyrus). Many more SFs/RBPs were associated with exon 10 exclusion (94 genes, 69.1% of all significant and replicated correlations, defined by a negative correlation; Table S2) compared with its inclusion (5 genes replicated in 2/3 datasets, [3.7%], defined by a positive correlation; Table S2), suggesting that more complex regulation may be required to promote removal of exon 10 from pre-mRNA transcripts in brain.

Fewer SFs/RBPs were associated with exon 2 splicing (7 [5.1%] excluders, 16 [11.8%] includers). However, a proportion of SFs/RBPs were significantly associated with both exon 2 and exon 10 PSI values in opposing directions (Table 1). Specifically, 14 SF/RBPs (10.3%) were significantly correlated with both exon 10 exclusion and exon 2 inclusion, suggesting a coordinated regulation of *MAPT* N- and C-terminal splicing that has not been previously characterized. We therefore chose to focus on this subset of genes (Table 1).

RSRC1 and RBM11 directly bind *MAPT* pre-mRNA and alter *MAPT* splicing *in vitro*

In order to prioritize SFs/RBPs for validation, we identified genes known to be expressed in brain (queried through the GTex portal; www.gtexportal.org) and neurons (Barres RNA-seq browser; www.brainrnaseq.org), as we predicted these would be most relevant to the regulation of *MAPT*, which is primarily neuronally expressed. This resulted in a panel of seven exon 2 includers/exon 10 excluders (Table 1, bold text).

To validate the association of candidate SF/RBPs with *MAPT* splicing, we first assessed whether they were able to directly interact with *MAPT* pre-mRNA in the context of human brain tissue. We carried out RNA pull-downs using desthiobiotinylated *in vitro* reverse transcribed RNA generated from a mini-gene containing *MAPT* exons 9-11, with full intervening intronic sequences (LI9LI10)[42] and probed protein lysates derived from postmortem human prefrontal cortical tissue. The resulting eluates were examined by western blot for SFs/RBPs of interest. LI9LI10-derived *MAPT* pre-mRNA pulled down significant proportions of RSRC1 (41.62%, $p < 0.001$) and RBM11 (11.11%, $p < 0.01$) proteins from human brain lysates (Figure 2A-B), as well as a minimal, but not significant proportion of THOC3 (5.77%) and SNRNP25 (4.77%), suggesting that these factors may regulate *MAPT* exon 10 splicing via a direct interaction with its pre-mRNA.

To functionally validate whether *RSRC1* and *RBM11* could alter *MAPT* splicing *in vitro*, we overexpressed these SF/RBPs in human neuroblastoma SH-SY5Y cells as an immortalized cell line proxy to neuronal cells (Figure S3A-B). As expression of *MAPT* exon 10 was very low in this cell line, we co-expressed the LI9L10 minigene with each SF/RBP to facilitate measurement of exon 10 exclusion. Consistent with their putative roles as exon 10 excluders, overexpression of either *RBM11* or *RSRC1* significantly reduced the 4R:3R ratio, as measured by qRTPCR (*RBM11* $p < 0.001$, *RSRC1* $p < 0.01$) (Figure 2C). While *PPIH* and *SNRPB* did not directly bind *MAPT* pre-mRNA, their overexpression resulted in a significantly reduced 4R:3R ratio in SH-SY5Y cells

(*SNRPB* $p < 0.01$, *PPIH* $p < 0.001$) (Figure S3C). There was no significant effect of *THOC7*, *THOC3* or *SNRNP25* (Figure S3C), possibly due to poor overexpression efficiency. We then measured the ability for candidate SF/RBPs to alter N-terminal splicing of endogenous *MAPT*, and found that both *RBM11* and *RSRC1* overexpression significantly increased the Exon 2/0N ratio in SH-SY5Y cells (*RBM11* $p < 0.001$, *RSRC1* $p < 0.01$) (Figure 2C), consistent with their hypothesized role as exon 2 includers. In contrast, we did not see any effect of other candidate SF/RBPs on exon 2 inclusion (Figure S3D). We therefore conclude that *RBM11* and *RSRC1* may be important regulators of combinatorial N- and C-terminal *MAPT* splicing.

Regulators of *MAPT* N-terminal splicing and *MAPT* exon 2 are differentially expressed in PSP and AD brain

Altered *MAPT* C-terminal splicing is a characteristic of several tauopathies[18], including PSP[19] and AD, for which there is inconsistent data[11,13,24–27]. We therefore queried the AMP-AD MAYO temporal cortex RNA-seq dataset, which includes both AD and PSP cases, in order to investigate whether *MAPT* splicing regulation may be altered in tauopathy brain. We calculated the fold change (FC) expression of every significant SF/RBP from our initial correlational analysis (Table S2) in PSP and AD compared to controls (Figure 3A). While PSP and AD shared largely similar patterns of SF/RBP expression dysregulation compared to controls, there were many genes that exhibited differential patterns of expression in either disease (Figure 3A). The group of SFs/RBPs that exhibited increased expression in PSP and reduced expression in AD comprised 4/7 of our candidate SF/RBPs, including both *RBM11* and *RSRC1,* which were associated with both exon 2 inclusion and exon 10 exclusion. In contrast, SFs/RBPs that were increased in AD and reduced in expression in PSP were significantly enriched for exon 2 excluders (Fisher's exact test

$p = 0.0005$) (Figure 3A). Furthermore, the net fold change (FC) expression of all exon 2 includers was significantly higher in PSP temporal cortex compared to AD (PSP average FC = 0.024, AD average FC = -0.036, $p < 0.01$) (Figure 3B), suggesting differential regulation of *MAPT* exon 2 splicing between diseases. To confirm this hypothesis, we compared *MAPT* exon 2, 3 and 10 PSI values in control, PSP and AD brain from the same dataset (Figure 3D-F). Consistent with the observed patterns of exon 2 splicing regulator expression in PSP and AD, we observed significantly different exon 2 inclusion across cases compared to controls ($F_{(2,134)} = 4.01$, $p = 0.02$), with significantly increased exon 2 PSI in PSP brain compared to controls (Tukey HSD $p = 0.02$), and a trend towards reduced exon 2 PSI in AD brain compared to controls (Figure 3D). In contrast, there was no significant difference in exon 3 or exon 10 PSI values between either disease and controls (Figure 3E-F). Taken together, this suggests there is differential dysregulation of *MAPT* N-terminal splicing regulators between AD and PSP brain, which results in altered expression of *MAPT* exon 2.

RSRC1 **is differentially expressed in AD and PSP neurons in disease-relevant brain regions**

While MAYO temporal cortex bulk RNA-seq data indicated altered expression of exon 2 splicing regulators in PSP and AD brain, these data may be influenced by altered cell type proportions in the disease context. Therefore, in order to validate neuronal and disease-specific patterns of expression of exon 2 splicing regulators, we assessed single nuclei sequencing (snuc-seq) from AD entorhinal cortex[43] and PSP subthalamic nucleus[44], as well as single-soma sequencing of hyperphosphorylated tau (AT8) positive neurons from AD prefrontal cortex[45]. We found that while *RBM11* was not detectable in AD snuc-seq data, and was expressed at very low levels in the other datasets, *RSRC1* was consistently detected and more highly expressed across

data sets. While *RSRC1* expression was highest in microglia in both entorhinal cortex and subthalamic nucleus, we confirmed that it was also expressed in neurons in these regions (Figure S4A-B). Consistent with data from bulk temporal cortex tissue, we observed significantly lower *RSRC1* expression in AD tissue compared to controls (FC = 0.07, $p < 0.01$), as well as a lower proportion of *RSRC1*-expressing cells in AD compared to controls (8.9% vs 12.5%, respectively) (Figure 3C). We observed the same fold change expression and reduction in *RSRC1*-expressing cells when assessing neuronal populations specifically (Figure 3C), although due to the small proportion of neurons present in the data (Figure S4C), this was not significant. Interestingly, *RSRC1* expression was also higher in neurons derived from AD prefrontal cortex that were negative for hyperphosphorylated tau (AT8-), compared to AT8+ neurons (FC = 0.13, $p = 3.33 \times 10^{-33}$) (Figure 3C, Figure S4D), indicating that there may be an interaction between *RSRC1* expression, *MAPT* splicing regulation and the formation of tau pathology.

In contrast, we found significantly higher *RSRC1* expression in PSP subthalamic nucleus cells compared to controls (FC = 0.04, $p < 0.01$), as well as a higher proportion of *RSRC1*-expressing cells in PSP (10% compared to 7% controls) (Figure 3C). These differences were exacerbated when assessing neurons alone (FC = 0.08, 17% PSP neurons vs 10% control neurons) (Figure 3C), but similar to the AD data, the small proportion of neurons in this dataset (Figure S4E) precluded these data from reaching statistical significance. Interestingly, we were able to detect the opposite pattern of effect for the exon 2 excluders *QKI* and *PRPF38B* in AD and PSP neurons by snuc-seq: consistent with the temporal cortex bulk data, expression of *QKI* and *PRPF38B* were higher in AD neurons (*QKI* FC = 0.239, *PRPF38B* FC = 0.04) and lower in PSP neurons compared to controls (*QKI* FC = -0.03, 47.8% PSP vs 44.4% control neurons, *PRPF38B* FC = -0.125, 15.7% PSP vs 12.5% control) (Figure S4F). These data therefore support our

assertion of differential expression of *MAPT* exon 2 regulators between AD and PSP brain. Furthermore, this demonstrates that SF/RBP expression is altered in neurons in disease-relevant brain regions.

MAPT N-terminal isoforms are expressed at different ratios in AD and PSP brain

As we found evidence of coordinated splicing between *MAPT* exons 2 and 10, but observed differential expression of only exon 2 and exon 2 regulators in AD and PSP brain, we hypothesized that there may be loss of coordinated combinatorial splicing regulation between N- and C-terminal *MAPT* in tauopathy brain. We therefore carried out targeted *MAPT* isoform (iso)-seq on temporal cortex tissue from control, AD and PSP cases to assess the expression of full length transcripts (Figure 4A-C). We observed similar ratios of expression for each isoform as previously described by western blot analyses[12]: 1N3R and 0N3R were the two most highly expressed isoforms (~30% and 25%, respectively), followed by 0N4R (~21%) 1N4R (~15%) and finally very low expression of both 2N isoforms 2N3R (~1.4%) and 2N4R (~0.8%) (Figure 4A). This pattern of expression was largely similar in AD and PSP cases, however there was a trend towards increased expression of both 0N isoforms in AD brain compared to controls (AD 0N3R 42.5% ± 12.7% vs 25% ± 5% controls, AD 0N4R 31.6% ± 3% vs 21% ± 4.6% controls), although due to high variability, these differences did not reach statistical significance (Figure 4A). We then compared the 4R:3R ratios for each N-terminal isoform, and found that despite increased expression of 0N isoforms in AD, there was no difference for either AD or PSP in the 0N4R:0N3R ratio (Figure 4B). In contrast, we observed a trend towards an increase in exon 2-containing 4R:3R ratios in PSP cases compared to controls (1N4R:1N3R PSP = 1.1 ± 0.38 vs 0.53 ± 0.07 controls, 2N4R:2N3R PSP = 0.8 ± 0.13 vs 0.55 ± 4.9 controls) (Figure 4B), consistent with both our observation of increased exon 2 inclusion

in temporal cortex (Figure 3D) and the known pathological accumulation of 4R tau isoforms in PSP pathology. Lastly, when comparing N-terminal isoform expression, we found no differences in PSP brain, but a trend towards reduced 1N/0N and 2N/0N ratios in AD brain specifically (1N/0N AD = 0.43 ± 0.13 vs 1.2 ± 0.33 controls, 2N/0N AD = 0.02 ± 0.01 vs 0.05 ± 0.01 controls) (Figure 4C), likely due to the increase in expression of 0N isoforms (Figure 4A). Interestingly, we also observe an increase in the 2N/1N ratio in AD brain (0.1 ± 0.03 vs 0.05 ± 0.005 controls) (Figure 4C), which may be due to slightly decreased 1N expression or increased 2N3R expression we observe in AD compared to controls (Figure 4A).

To examine whether these transcriptional changes were apparent at the protein level, we carried out western blot analyses in the same tissues from the same individuals (Figure 4D-G). Consistent with the iso-seq data, there was a significantly increased 1N4R:1N3R ratio in PSP brain compared to controls (control 1N4R:1N3R ratio = 0.52 (SEM= 0.1), PSP 1N4R:1N3R ratio = 1.7 (SEM = 0.4), $p = 0.009$), which was largely driven by a reduction in 1N3R tau (Figure 4D-E). In contrast, there was no difference in either 0N or 2N 4R:3R ratios in PSP (Figure 4D-E). In AD brain, there was an accumulation of soluble 0N and 2N isoforms by western blot, consistent with the iso-seq data (Figure 4C, F-G). When examining 4R:3R ratios for each N-terminal isoform, we observed significantly increased 0N4R:0N3R (control ratio = 0.83 (SEM = 0.05), AD ratio = 1.49 (SEM = 0.15), $p = 0.002$) and 2N4R:2N3R ratios (control ratio = 0.83 (SEM = 0.05), AD ratio = 1.96 (SEM = 0.46), $p = 0.03$) in AD brain compared to controls (Figure 4F-G), indicating that while there were no transcriptional differences in the 4R:3R ratio for these isoforms, there may be impaired degradation of 0N4R and 2N4R tau in AD resulting from their increased expression.

Pathologically aggregated tau in PSP and AD brain are associated with different N-terminal isoforms

While western blot analysis of AD and PSP temporal cortex revealed differences in the accumulation of soluble N-terminal tau isoforms, we wanted to determine whether the formation of different neuropathological features between tauopathies may be due to the insoluble aggregation of different N-terminal tau isoforms. After validating the specificity of each N-terminal antibody by overexpressing different tau isoforms in N2a cells (Figure S5A-B), we carried out immunohistochemistry (IHC) on control, AD and PSP brain sections from the temporal cortex for hyperphosphorylated pathogenic tau (AT8; Figure 4H)) and each N-terminal isoform (Figure 4I). We observed the anticipated hyperphosphorylated tau neuropathology in both PSP and AD brain (Figure 4H), specifically the widespread presence of neurofibrillary tangles and neuropil threads throughout the AD cortex, and sparse glial plaques and neurofibrillary tangles in PSP tissue. In comparison, there was little to no signal in control brain (Figure 4H).

There was little signal for 0N tau in PSP brain (Figure 4I), although in one case we observed sparse labeling of neuropil threads or possible glial involvement (Figure 4I). In AD brain the 0N antibody did not label neurofibrillary tangles, but we did observe labeling of neuropil threads and dystrophic neurites surrounding amyloid plaques (Figure 4I). Interestingly, in one case the 0N antibody labelled thorny astrocytes consistent with an age-related tau astrogliopathy (ARTAG) pathology that was not visible with the AT8 antibody (Figure 4H-I). While it is surprising that there was little immunolabeling of 0N Tau in human brain, given the high levels of 0N transcripts, this pattern was replicated with a second antibody against 0N Tau (Figure S5C).

1N Tau immunostaining was primarily present in neurofibrillary tangles and pre-tangles (Figure 4I) throughout the cortex in all three PSP cases, as well as neuropil granules and threads.

In AD, this isoform was also the most prevalent in neurofibrillary tangles (Figure 4I), but was primarily present in dystrophic neurites surrounding amyloid plaques (Figure 4I). Finally, while 2N Tau was observed in some neurofibrillary tangles in PSP and AD brain (Figure 4I), these were less common than the 1N labelled neurons. 2N Tau was also observed in dystrophic neurites surrounding amyloid plaques in AD brain, but this labelling was far less dense than the 1N Tau and less tightly co-localized with plaques (Figure 4I). A second 2N antibody revealed a similar labeling of neurofibrillary tangles in AD and PSP brain, but less involvement in dystrophic neurites (Figure S5C).

In order to directly compare the accumulation of different tau N-terminal isoforms in pathogenic inclusions, we carried out Opal multiplex labelling of adjacent brain sections from the same individuals using all three N-terminal antibodies in conjunction with AT8 and β-amyloid staining (Figure 5A-B, Figure S6A-B). Consistent with the IHC staining, we observed primarily 1N and 2N tau in AT8-positive neurofibrillary tangles in PSP temporal cortex, but all three N-terminal isoforms were present in AD-associated tangles (Figure 5A-B, Figure 6A-B). Dystrophic neurites surrounding amyloid plaques were primarily associated with 0N and 1N accumulation, with less 2N involvement (Figure 5A, Figure S6A). In contrast, amyloid plaques present in either controls or PSP cases were not associated with any tau staining (Figure S6B). Interestingly, 2N tau was absent in glial pathology observed in both AD and PSP cases, suggesting that 2N tau is unable to accumulate in glia, while astrocytic tufts in PSP brain were labelled primarily by 1N tau (Figure 4I), consistent with the increased accumulation of 1N4R tau isoforms we observed by western blot and transcriptomic analyses.

DISCUSSION

To date, the contribution of N-terminal *MAPT* splicing to disease pathogenesis has largely been overlooked compared to evaluation of exon 10 splicing and 4R tau expression. However, N-terminal tau is relevant to disease pathogenesis: for example, N-terminal fragments are prominent in AD cerebrospinal fluid (CSF), and their secretion from cells can be inhibited by the presence of exon 2[15,46]. In contrast, 0N Tau is more readily cleaved and released from human neuroblastoma cells[46]. This is consistent with our IHC data, where we observe little 0N tau accumulation in neurons, but prominent 1N and 2N tau in neurofibrillary tangles. The regulation of tau release by N-terminal inserts also has implications for our understanding of tau spread and seeding, as different isoforms may be available extracellularly, and exhibit different seeding competencies. Furthermore, aberrant folding of N-terminal tau is one of the earliest pathological changes identified in tauopathies[47], thus supporting the assertion that N-terminal splicing is a relevant consideration when modeling and investigating tauopathy.

N-terminal tau splicing may play a role in modifying tau subcellular localization and aggregation propensity. The N-terminal contains a plasma-interacting domain[48] that interacts with synaptic proteins and Annexin A6[49,50], which results in retention of tau in the axonal compartment. While exons 2 and 3 are not within this domain, it is possible that they modify these interactions and impact the subcellular localization of tau, resulting in the increased propensity for certain isoforms to accumulate in these regions. Indeed, murine 0N tau has been found to localize in axons, whereas 1N was enriched in dendrites and 2N tau was depleted from cytoskeletal structures[51], indicating that N-terminal splicing likely modifies tau function. 1N tau has also been found to more readily accumulate and aggregate: the presence of exon 2 promotes the fibrillary extension

of tau filaments *in vitro*[52] and exon 2-containing proteins more readily polymerize than 0N tau[53], therefore increased exon 2 inclusion would be likely to worsen tau aggregation. This is consistent with our observation of strong 1N tau immunostaining in PSP astroglia and tangles in both PSP and AD neurons. Despite this, we observe increased 0N expression in AD brain. However, the resulting protein accumulation was soluble, and associated with dystrophic neurites surrounding amyloid plaques rather than in neurofibrillary tangles. N-terminal tau splicing has also been found to be required for specific interactions with proteins associated with synaptic signaling and the plasma membrane[54], supporting the assertion that different N-terminal isoforms likely facilitate distinct cellular functions. Characterizing the disruption of N-terminal splicing is therefore important for understanding tau biology and the mechanisms underlying disease pathogenesis.

We have identified coordinated regulation of *MAPT* exons 2 and 10, indicating that N- and C-terminal splicing of either region likely does not occur fully independently of the other. This phenomenon has been reported once previously using polony-based exon profiling[13]. Disrupting this coordination will therefore lead to imbalanced expression of different tau isoforms. Interestingly, while the expression of numerous exon 10 splicing regulators were altered between AD, PSP and control brains, we observed enrichment of exon 2 regulators with opposing patterns of differential expression between PSP and AD. As anticipated from this pattern of expression, we observed increased exon 2 inclusion in PSP and reduced exon 2 in AD. However, assessment of full length isoforms by iso-seq and western blot revealed that shifts in N-terminal splicing were coupled to alterations in the 4R:3R ratio, supporting our hypothesis that loss of coordinated regulation between N- and C-terminal splicing contributes to tauopathy pathogenesis. The change in the 4R:3R ratio for specific N-terminal isoforms may explain why we did not observe an increase in exon 10 inclusion in PSP by short-read bulk RNA-seq; the most highly expressed 0N

isoforms did not exhibit a change in the 4R:3R ratio in PSP brain, therefore the lack of change in 0N4R may have diluted out increases in 1N or 2N4R. Curiously, the increased 1N4R:1N3R ratio observed in PSP appeared to be largely due to a loss of 1N3R rather than an increase in 1N4R tau, thus raising the possibility that in combination to 1N4R tau accumulation, losing 1N3R tau expression and function could also be detrimental to neuronal health.

Aberrant regulation of splicing is a known phenotype of some forms of inherited tauopathy[55], which is hypothesized to be due to the sequestration and mislocalization of SF/RBPs to the cytosol and into stress granules[55]. SF/RBP dysfunction is therefore a likely mechanism underlying aberrant *MAPT* splicing in non-familial AD and PSP. While we do not characterize the mechanism of SF/RBP disruption in AD and PSP brain here, we do identify differential expression of numerous SF/RBPs between AD and PSP compared to controls, which is suggestive of separate downstream effects of splicing dysregulation that ultimately contribute to the pathogenesis of either disease.

We found that *MAPT* splicing is likely regulated by numerous splicing factors, of which *RSRC1* and *RBM11* may be of particular interest. However, it is likely that there are many other modifiers that were not identified in our screen. Several of our SF/RBPs of interest, including *SNRPB*, *SNRNP25*, *THOC7* and *THOC3*, are known to be components of protein complexes that regulate splicing, and therefore would be unlikely to be isolated by the RNA pull-down assay. Indeed, overexpression of some of these SF/RBPs was able to significantly shift the 4R:3R ratio in human neuroblastoma cell lines. While expressed in brain and in neurons, many of these SF/RBPs are ubiquitously expressed in many different tissues and cell types. Indeed, we observed the highest expression of *RSRC1* in microglia in human brain snuc-seq data. Therefore, while their dysregulation may be impacting *MAPT* splicing in neurons, there will likely be wider effects of

altered expression and splicing in other neural cell types that may also be relevant for disease pathogenesis.

RSRC1 is a member of the serine and arginine rich-related protein family, which are highly conserved regulators of alternative splicing, but how *RSRC1* regulates this process is unknown. It is hypothesized that RSRC1 plays a role in 3' splice site selection by interaction with splicing regulators U2AF and SRSF2[56], although it also possible that it may interact directly with pre-mRNA via its RS (arginine-serine) domain. However, there is no specific RNA motif that has been characterized as binding RS domains, thus the region in which RSRC1 may be binding *MAPT* pre-mRNA is currently unknown. Mutations within this gene have been associated with intellectual disability[57,58], whereas the silencing of *RSRC1* in SH-SY5Y cells has been associated with downregulation of genes associated with schizophrenia, Alzheimer's disease and dementia[58], thus indicating the relevance of this gene for brain function.

RBM11 is a brain specific splicing factor that exhibits fluctuating expression with brain development, with high expression throughout embryogenesis, peaking at perinatal days 0-3[59], after which *MAPT* exon 10 expression increases[8], consistent with our observations that *RBM11* expression may promote exon 10 exclusion. *RBM11* is associated with the choice of 5' splice sites and may antagonize the activity of other SF/RBPs such as *SRSF1*[59], potentially regulating *MAPT* splicing by direct interaction with pre-mRNA and by inhibiting binding of competing SFs/RBPs. *RBM11* is downregulated in the PS19 mouse model of tauopathy[55], consistent with the direction of effect we observe in AD brain. However, this mouse model expresses only 1N4R tau and as such the effect of *RBM11* disruption on *MAPT* splicing was not measurable in this model. It should be noted that we were unable to validate *RBM11* expression in snuc-seq data as its expression was

very low, therefore the relevance it may have to tauopathy and *MAPT* splicing requires additional investigation.

We were unable to identify regulators of *MAPT* exon 3 splicing from these data, or to compare its regulation between 17q21.31 H1 and H2 haplotypes due to the low frequency of H2 carriers. Increased Exon 3 inclusion is consistently identified on the H2 background[12,39], an effect we also observe in these data. However, the mechanism underlying increased exon 3 inclusion, and whether its expression is protective against tauopathy is currently unknown, although it may reduce the fibrillization of tau[52]. It should be noted that exon 3, and subsequently 2N tau expression, is very low in adult human brain (accounting for less than 3% of all transcripts in our targeted iso-seq data), therefore the extent to which it may contribute to disease is unclear. Nevertheless, we observe some alterations in 2N expression and accumulation in AD and PSP, the most striking of which was the absence of 2N tau in glial pathology, indicating that 2N tau is either not released from neurons or is not internalized or aggregated by astrocytes. These data have implications for the design of experimental models of tauopathy where a single tau isoform is expressed in order to ensure the most disease and pathology appropriate isoform is utilized.

In conclusion, we propose that changes in SF/RBP expression result in differential splicing of the *MAPT* N-terminus between AD and PSP, resulting in the expression of isoforms with different aggregation properties and subcellular localizations, thus explaining the distinct neuropathological phenotypes of each disease. It would therefore be of great interest to investigate the role of N-terminal splicing in other primary tauopathies associated with different pathologies, such as Pick's disease (PiD), primary age-related tauopathy (PART), age-related tau astrogliopathy (ARTAG) and chronic traumatic encephalopathy (CTE) to determine whether these diverse disorders also exhibit loss of *MAPT* exon 2 and 10 splicing coordination. These data indicate that

it is unlikely that exon 10 splicing is alone in underlying and regulating disease pathogenesis and tau neuropathology in either AD or PSP, but rather the combinatorial expression of specific and N- and C-terminal *MAPT* isoforms is relevant for understanding the development of tauopathy. While alterations in the 4R:3R ratio are undoubtedly important and relevant to our understanding of tau pathogenesis, N-terminal splicing is likely to be an important modifier of disease and pathology, and should be carefully assessed.

METHODS

RNA-seq data analysis

Aligned BAM files and gene expression count data were downloaded from the AMP-AD consortium through Synapse (https://adknowledgeportal.synapse.org/). *MAPT* exon level counts were calculated using the FeatureCounts feature within the Subread package[60,61], and PSI values were determined using the Mixture of Isoforms (MISO) package[62]. For statistical analysis, associations between *MAPT* PSI values and the expression of SFs/RBPs were carried out using a linear model in R with RNA integrity number (RIN), postmortem interval (PMI), sex, and age at death included in the model as covariates. The resulting *p*-values were Bonferroni-corrected for the number of comparisons. For heatmap plotting, correlation coefficients were generated using Pearson's correlation as part of the cor function in R. For the comparison of SF/RBP expression in PSP and AD brain, the fold change expression of each SF/RBP in disease brain was calculated in comparison to controls and plotted in R using ComplexHeatmap[63], and statistically significant differences were determined by linear regression of expression values including the previously described covariates.

At the time of analysis, genotype data were unavailable for the MSBB cohort, so 17q21.31 haplotype was determined by Taqman genotyping. The relevant DNA was obtained from the NIH Neurobiobank. Taqman genotyping was carried out for H2 tag SNPs rs8070723 and rs1052553 using commercially available assays. Haplotypes were determined for the other cohorts using the same tag SNPs from genotype data downloaded from the AMP-AD knowledge portal (https://adknowledgeportal.synapse.org/).

Single-nuclei and single-soma sequencing analysis

Snuc-seq processed gene counts and covariates derived from AD and control entorhinal cortex[43] were downloaded from the Gene Expression Omnibus (GEO) (GSE138852). Data were further normalized and analyzed in Seurat 3.0 dev[64,65]. Data from different individuals were integrated and scaled using SCTransform[66] while regressing out the percentage of mitochondrial genes, the number of genes per cell and the number of reads per cell. Principal components analysis (PCA) was carried out in Seurat using the top 3000 most variable genes, and data was reduced using UMAP[67]. Cell types present within each cluster were already annotated in the downloaded metadata. Differential gene expression analysis was carried out between AD cases and controls across the whole data set, or within the neuronal cluster only, using the MAST model applied to log normalized raw count data, including the percent of mitochondrial genes as a covariate.

Aligned HDF5 feature barcode matrices for the single-soma sequencing data of AT8 positive and negative neurons from AD PFC[45] were downloaded from GEO (GSE129308) and processed in Seurat[64,65]. Data were filtered for cells expressing > 200 genes, < 2500 reads and < 10% mitochondrial genes. Data were integrated, transformed and reduced as described above. Differential gene expression analysis was carried out between AT8 positive and AT8 negative

cells across the whole data set, using the MAST model applied to log normalized raw count data, including the percent of mitochondrial genes, age, RNA integrity number (RIN) and postmortem interval (PMI) as covariates.

Aligned HDF5 feature barcode matrices for PSP snuc-seq data[44] were kindly shared by Drs. Pereira and Crary, and were filtered, integrated, transformed and reduced in the same manner as the AD snuc-seq and AT8 soma-seq data described above. Individual clusters were identified in Seurat using the default resolution factor 0.5. Cell types within each cluster were defined using visualization of specific markers utilized in the AD snuc-seq data[43]; *CD74* (microglia), *AQP4* (astrocytes), *MEGF11* (oligodendrocyte precursor cells (OPCs)), *MOBP* (oligodendrocytes), *SYT1* (neurons) and *FLT1* (endothelial cells). Differential gene expression analysis was carried out between PSP cases and controls across the whole data set, or within the neuronal cluster only, using the MAST model applied to log normalized raw count data, including the percent of mitochondrial genes and age as covariates.

Human brain tissue

Fresh frozen human control, AD and PSP temporal cortices were acquired from the Mount Sinai Neuropathology Core brain bank and the Harvard Brain Tissue Resource Center, University of Maryland Brain and Tissue Bank and Mount Sinai Brain Bank via the NIH Neurobiobank. Formalin fixed paraffin embedded sections from temporal cortex were acquired from the Mount Sinai Neuropathology Core brain bank, with neuropathological diagnosis being determined by Dr. John Crary. All post-mortem tissues were collected in accordance with the relevant guidelines and regulations of the respective institutions. A summary of tissues used in this project are described in Table S3.

Cell culture

All cell culture reagents were purchased from Thermo Fisher Scientific, unless otherwise stated. SH-SY5Y cells were grown in IMDM media supplemented with 1% penicillin/streptomycin and 10% FBS, and grown and maintained at 37°C with 5% CO_2 in a humid environment. Prior to transfection, cells were seeded into 6 well plates at a density of $1.6x10^5$ cells per well. The next day, cells were transfected with 1.25ug of LI9LI10 mini-gene and 1.25ug of plasmid DNA of the SF/RBP of interest (all Origene) using Lipofectamine 3000. Cells were collected for RNA extraction and analysis 48 hours later. For N-terminal antibody validation, N2a cells were transfected with 2.5ug of either a 0N3R, 1N3R or 2N3R *MAPT* cDNA vector (Origene), and after 48 hours were either fixed with 10% formalin (Sigma Aldrich) for 15 minutes at room temperature for immunofluorescence, or pelleted in PBS for protein extraction and western blotting.

RNA pull-down

The LI9LI10 mini-gene was digested by NotI (Cell Signaling Technologies) and SgfI (Promega) to excise the *MAPT* coding sequence and upstream T7 promoter from the PCI-Neo backbone. The resulting DNA was transcribed *in vitro* using the T7 Megascript kit (ThermoFisher Scientific), incubated at 37°C for 4 hrs, followed by 15 minutes treatment with DNase to degrade any remaining DNA template. The resulting RNA was isolated by Lithium Chloride precipitation, and examined on a 1% agarose gel for the anticipated product size, compared against Lambda DNA digested with HindIII and EcoRI (both Cell Signaling Technologies). RNA was labelled using the 3' end desthiobiotinylation RNA labelling kit (ThermoFisher), with 150ug/25nM RNA per reaction incubated at 16°C overnight. Labelled RNA was then isolated by chloroform and ethanol precipitation. A negative control consisting of a scrambled RNA sequence was labelled at

the same time. Labelling efficiency was measured by comparison of chemiluminescent signal from labelled sample RNA with a positive control using the ThermoFisher Scientific Chemiluminescent Nucleic Acid detection module. 25nM of labelled RNA was then bound to nucleic acid-compatible streptavidin magnetic beads using the Pierce Magnetic RNA-Protein Pull-down kit (ThermoFisher Scientific) and incubated with 100ug protein lysate from human brain overnight at 4°C. Bound protein was eluted from the beads and immediately subject to western blot analysis for SFs/RBPs of interest.

SDS-PAGE gel electrophoresis and western blot

Soluble protein was collected from cell pellets and human brain tissue by resuspension in Cell Lysis Buffer (Cell Signaling Technologies) supplemented with 10μM PMSF on ice. Cells or tissue were then sonicated briefly on ice and spun at 13,000xg for 10 minutes at 4°C to pellet debris. Protein concentration was determined by BCA assay (ThermoFisher Scientific). Prior to western blotting, human brain protein lysates were dephosphorylated in order to accurately determine Tau isoforms by size. Lysates were incubated with 100 units of Lambda protein phosphatase (LPP; Cell Signaling Technologies) per 10ug total protein, supplemented with 1x Protein MetalloPhosphatases buffer and 1mM $MnCl_2$, and incubated at 30°C for 3 hours before being analyzed by western blot.

For SDS-PAGE gel electrophoresis, 10-30ug of protein was incubated with 1x reducing agent and 1x LDS sample buffer at 70°C for 10 minutes before immediately being loaded onto a BOLT 4-16% Bis-Tris gel (ThermoFisher Scientific) in 1x MES buffer. For splicing factor analyses, electrophoresis was carried out for 20 minutes at 200V before blotting. For tau isoform analyses, electrophoresis was carried out for 60 minutes at 100V before blotting. Gels were blotted onto nitrocellulose membranes using the iBlot system (ThermoFisher Scientific), and blocked for

a minimum of 30 minutes in 5% milk in PBS-T. Primary antibodies were prepared at dilutions described in SI Table 4 in 5% milk in PBS-T and incubated with the membrane at 4°C overnight. Membranes were washed 3x in PBS-T, then incubated with either HRP Goat Anti-Rabbit or HRP Horse Anti-Mouse secondary antibodies (Vector laboratories) at a dilution of 1:20,000 in 5% milk in PBS-T for two hours at room temperature. Following three additional washes, membranes were then incubated with WesternBright ECL HRP substrate (Advansta) for 3 minutes before imaging on a UVP ChemiDoc. For re-staining, blots were incubated in Restore PLUS stripping buffer (ThermoFisher Scientific) for 15 minutes at room temperature, followed by one wash in PBS and re-blocking.

qRTPCR

Cell pellets were collected by washing and scraping into ice-cold PBS. RNA was extracted from cell pellets using the Qiagen RNeasy mini RNA extraction kit, and reverse transcribed using the high capacity RNA-to-cDNA kit (ThermoFisher Scientific). qRTPCR for specific *MAPT* exons and isoforms was carried out using SybrGreen mastermix with the following primers: *MAPT* 4R Forward 5'-CGGGAAGGTGCAGATAATTAA-3', Reverse 5'-GCCACCTCCTGGTTTATGATG-3'; *MAPT* 3R Forward 5'-AGGCGGGAAGGTGCAAATA-3', Reverse 5'-GCCACCTCCTGGTTTATGATG-3'; *MAPT* 0N Forward 5'-TTTGAACCAGGATGGCTGAG-3', Reverse 5'-ATGCCTGCTTCTTCAGCTTT-3'; *MAPT* Exon 2 Forward 5'-TTTGAACCAGGATGGCTGAG-3', Reverse 5'-CTGCAGGGGAGATTCTTTCA-3′. SF/RBP overexpression and knockdown was validated and quantified by qRTPCR using commercially available Taqman assays (ThermoFisher Scientific).

***MAPT* targeted iso-seq**

RNA was extracted from human temporal cortex brain tissue as described above and submitted to the Icahn School of Medicine at Mount Sinai Genomics CoRE for single molecule real time (SMRT) isoform sequencing (iso-seq) on the PacBio RS II platform using the following primers: Forward 5'-ATG GAA GAT CAC GCT GGG AC-3', Reverse 5'-GAG GCA GAC ACC TCG TCA G-3'. Raw sequencing reads were passed through the ISOseq3 pipeline to detect full-length transcripts expressed in each sample. Beginning with raw subreads, single consensus sequences were generated for each *MAPT* amplicon with a SMRT adapter on both ends of the molecule. SMRT adapter sequences were then removed and *MAPT*-specific primer sequences were identified to orient the isoforms. Isoforms were subsequently trimmed of poly(A) tails and concatemers were identified and removed. Isoform consensus sequences were then predicted using a hierarchical alignment and iterative cluster merging algorithm to align incomplete reads to longer sequences. Finally, clustered isoform sequences were polished using the arrow model and binned into groups of isoforms with predicted accuracy of either ≥ 0.99 (high quality) or < 0.99 (low quality). The resulting isoforms were aligned to hg38 using the GMAP aligner[68] and isoform calling, collapsing and measurements of abundance were carried out using the Cupcake/ToFU pipeline (https://github/Magdoll/cDNA_Cupcake).

Immunohistochemistry

Immunohistochemical staining was carried out on formalin fixed paraffin-embedded brain sections by the Neuropathology Brain Bank and Research CoRE at the Icahn School of Medicine at Mount Sinai using the Ventana BenchMark autostainer. Slides were scanned on a Leica SCN400 at 40x. A list of antibodies used and their relevant dilutions can be found in Table S4.

OPAL multiplexed immunofluorescence

Multiplexed immunofluorescent staining was carried out using the Opal Polaris 7 color IHC detection kit (Akoya biosciences) according to manufacturer's instructions. Briefly, slides were baked for 1 hour at 65°C, then deparaffinized with xylene and rehydrated with a graded series of ethanol concentrations. For epitope retrieval, slides were microwaved in AR buffer (provided with the OPAL IHC detection kit) for 45s at 100% power, followed by an additional 15 minutes at 20% power. After cooling, slides were blocked for 10 minutes in blocking buffer then incubated with the first primary antibody at room temperature for 30 minutes. Slides were rinsed three times in TBS-T, then incubated with the secondary polymer HRP for 1 hour at room temperature. After additional washes, the first Opal fluorophore was incubated with the slides for 10 minutes at room temperature, followed by further washes in TBS-T. This process was repeated from the microwave treatment step for each additional primary antibody, followed by one final repetition of the microwave treatment to strip the primary-secondary antibody complex from the tissue. Antibodies, concentrations and relevant Opal fluorophores can be found in Table S4. Once all primary antibodies had been introduced, slides were counterstained with DAPI for 5 minutes at room temperature, washed with TBS-T and coverslips were mounted using ProLong Diamond Antifade mounting reagent (ThermoFisher Scientific). Multispectral imaging was carried out using the Vectra Quantitative Pathology Imaging system, applying quantitative unmixing of fluorophores and removal of tissue autofluorescence. Images were visualized using the HALO image analysis platform (Indica Labs).

Statistical analysis

RNA-seq count and PSI data were analyzed as described above. Enrichment of SFs/RBPs in specific clusters was determined by Fisher's exact test in R. Western blot protein bands were quantified by densitometry analysis in ImageJ and normalized to GAPDH for each sample, and

the resulting values were subjected to unpaired student's t-test. For tau isoform analysis, ratios between each isoform were calculated per sample prior to statistical analysis. For assessment of RNA pull-downs, the amount of eluted SF/RBP protein was normalized to the total amount of SF/RBP protein (Flow-through + Eluate) to determine the percentage of SF/RBP protein bound to the labelled RNA, and this value was used for statistical analysis by student's t-test. For ISO-seq data, the expression of each isoform was calculated as a proportion of all detected isoforms, and average expression between control, AD and PSP cases was calculated. Statistical difference in isoform expression and expression ratios were calculated by one-way ANOVA with Tukey post-hoc testing. qRTPCR gene expression was analyzed using the ΔΔCt method, and expression was normalized to β-actin as endogenous controls. Statistical significance was determined by the appropriate one-way ANOVA and Bonferroni post-hoc testing. For cell culture experiments, all tests were conducted in triplicate in three independent experiments (total replicates = 9). For human brain analyses, tissue was acquired for 4-6 PSP cases, 4-6 AD cases and 4-6 healthy aged controls. Significant comparisons are labelled in figures as $*p < 0.05$, $**p < 0.01$ and $***p < 0.001$.

ACKNOWLEDGEMENTS

This work was supported by funding from the BrightFocus Foundation (KRB), Association for Frontotemporal Degeneration (KRB), CurePSP (KRB), the Rainwater Charitable Foundation (AMG, KRB), and Alzheimer's Association AARF-17-529888 (JDC). We thank the NIH Neurobiobanks at the University of Maryland, Harvard and Mount Sinai for supplying tissues for analyses. We also thank the Neuropathology brain bank and research CoRE and the Genetics CoRE at the Icahn School of Medicine at Mt. Sinai for supplying tissues and conducting

immunohistochemical staining, and for carrying out targeted iso-seq and data analysis. We are grateful to the study participants and their families for their contributions to research.

AUTHOR CONTRIBUTIONS

Conceptualization: KRB, AMG

Methodology: KRB, DAP, LMO, BMJ, KF, KW, AS, JDC

Validation: KRB, DAP, LMO

Formal analysis: KRB, KF

Investigation: KRB, DAP, LMO

Resources: JDC, TR, KF, KW, JFC, AP, AMG

Data curation: KRB

Writing – Original draft: KRB

Writing – Review and editing: KRB, DAP, LMO, KF, KW, JDC, JFC, AP, AMG

Visualization: KRB

Supervision: AMG

Funding Acquisition: KRB, JFC, AMG

DECLARATION OF INTERESTS

AMG: Scientific advisory board (SAB) for Denali Therapeutics (2015-2018), SAB for Pfizer (2019), SAB for Genentech, consultant for GSK, AbbVie, Biogen and Eisai. All other authors declare no competing interests.

FIGURE LEGENDS

Figure 1. Splicing factor and RNA binding protein expression is differentially correlated with the inclusion of *MAPT* exons 2 and 10.

A) *MAPT* exons 2, 3 and 10 are alternatively spliced, resulting in the expression of 6 different isoforms. At the N-terminus, exons 2 and 3 may be included or excluded, although exon 3 inclusion requires the inclusion of exon 2. The absence of either exon results in 0N isoforms, exon 2 alone results in 1N isoforms, and exon 3 inclusion defines 2N isoforms. At the C-terminus, the inclusion of exon 10 (encoding the second microtubule binding repeat domain) defines 4R isoforms, whereas its exclusion results in 3R isoforms.

B) *MAPT* exon expression (RPKM) in the AMP-AD ROSMAP cohort. Error bars ± SEM.

C) *MAPT* exons 2, 3 and 10 PSI across four Brodmann regions examined in the AMP-AD MSBB cohort. Error bars ± SEM.

D) PSI values for *MAPT* exons 2, 3 and 10 in cerebellum and temporal cortex in the AMP-AD MAYO cohort. Error bars ± SEM.

E-F) *MAPT* PSI values for exons 2, 3 and 10 between 17q21.31 H1 and H2 haplotypes in ***E.*** MSBB and ***F.*** MAYO AMP-AD datasets. Error bars ± SEM.

G) Pearson's correlation coefficients with unsupervised hierarchical clustering between SF/RBP expression (x-axis) and *MAPT* exon 2/exon 10 PSI values in ROSMAP and MSBB data. Blue indicates a positive correlation ("includers") and red indicates a negative correlation ("excluders"), while yellow denotes no association.

All comparisons carried out using linear regression model. ***$p < 0.001$

Figure 2. RBM11 and RSRC1 directly bind to *MAPT* pre-mRNA and regulate splicing

A) *MAPT* pre-mRNA pull-downs from human brain tissue protein lysates for each target SF/RBP using the LI9LI10 minigene sequence or scrambled (Scr) RNA sequence as bait. *FT* = flow through fraction not bound to *MAPT* pre-mRNA or non-specific control, *E* = eluate fraction bound to target RNA.

B) Quantification of RNA pull-down western blots in ***A***. Fraction of protein pulled down in *E* normalized to total protein present in *FT* and *E* fractions combined. N=3, each SF/RBP compared to scrambled control using t-test. $^{**}p < 0.01$, $^{***}p < 0.001$, ns = not significant.

C) Fold change (FC) expression of 4R:3R and Ex2:0N ratios in SH-SY5Y cells by qRTPCR following overexpression of either *RBM11* or *RSRC1*. Asterisks denote significantly different expression compared to empty vector control, represented by grey line intersecting plot at FC = 1. One-way ANOVA and post-hoc Bonferroni tests for multiple comparisons. N = 3 independent experiments with 3 replicates each. $^{**}p < 0.01$, $^{***}p < 0.001$.

Figure 3. *MAPT* exon 2 expression and regulation is differentially altered in PSP and AD brain

A) Fold change expression of each significant SF/RBP from the analysis in Figure 1G in PSP and AD brain compared to controls (AMP-AD MAYO temporal cortex), with unsupervised hierarchical clustering. Red indicates increased expression compared to controls, blue indicates reduced expression compared to controls. The direction of association of each SF/RBP with *MAPT* exon 2 and exon 10 splicing is indicated to the right of the figure. Clusters containing target SF/RBPs of interest from prior analyses and *MAPT* exon 2 excluders (purple) are indicated.

B) Sum of the fold change expression of all exon 2 includers in PSP and AD brain compared to controls (AMP-AD MAYO temporal cortex). Statistical comparison between sum expression in PSP compared to AD brain, t-test, Error bars ±SEM. $**p < 0.01$.

C) Single nuclei and single soma *RSRC1* expression in AD entorhinal cortex, AT8 positive/negative neurons and PSP subthalamic nucleus. FC = fold change expression disease/AT8 positive compared to controls/AT8 negative. Depth of color indicates scaled average expression, size of dot indicates proportion of cells expressing *RSRC1*, also denoted by percentage value above each dot. MAST linear model with Bonferroni correction, $**p < 0.01$, $***p < 0.001$

D-F) PSI values for exons 2 (***D***), 3 (***E***) and 10 (***F***) in control, AD and PSP brain (AMP-AD MAYO temporal cortex). Red dashed line indicates mean PSI value in controls for each exon. One-way ANOVA with post-hoc Tukey tests. Error bars ±SEM. $*p < 0.05$.

Figure 4. Coordination of N- and C-terminal splicing regulation is altered in AD and PSP brain, and is associated with tau pathology.

A-C) Proportion of each full length *MAPT* isoform (***A***), 4R:3R ratio for each N-terminal isoform (***B***) and N-terminal isoform ratios (***C***) as detected by targeted *MAPT* iso-seq in human control, AD and PSP temporal cortex. One-way ANOVA with post-hoc Tukey tests, all comparisons did not reach statistical significance. Error bars ±SEM.

D) Dephosphorylated tau isoform expression in PSP temporal cortex compared to tau ladder detected by specific N-terminal antibodies against 0N, 1N and 2N. Anti-1N tau showed poor detection for 1N4R isoforms, so anti-4R tau antibody RD4 was used for detection of this isoform. GAPDH used as loading control.

E) Quantification and 4R:3R ratio of N-terminal isoforms in ***D***. Each point denotes a different individual brain lysate, N = 6. Error bars ±SEM. Student's t-test $**p < 0.01$

F) Dephosphorylated tau isoform expression in AD temporal cortex compared to tau ladder with same antibody detection as in ***D***. GAPDH used as loading control.

G) Quantification and 4R:3R ratio of N-terminal isoforms in ***F***. Each point denotes a different individual brain lysate, N = 6. Error bars ±SEM. Student's t-test $*p < 0.05$, $**p < 0.01$

H) Representative images of labeling of control, AD and PSP temporal cortex with marker of hyperphosphorylated tau, AT8, indicating different tau pathologies present across cases. Sections counterstained with hematoxylin and eosin. N=3-4. Scale bar = 100µm.

I) Representative images of tau pathology labeling in control, AD and PSP temporal cortex with anti-0N, 1N and 2N tau antibodies from several individuals. Sections counterstained with hematoxylin and eosin. N=3-4. Scale bar = 100µm.

Figure 5. N-terminal isoforms accumulate differently in AD and PSP neuronal and glial pathologies

A) Representative images of multiplex immunofluorescent labeling of AD temporal cortex with AT8 (red), β-amyloid (green), 2N tau (blue), 1N tau (yellow) and 0N tau (orange), and overlay of all three N-terminal tau antibodies in 4 different individuals. Examples of tau accumulation in dystrophic neurites can be found in AD1 and AD2, thorny astrocytes present in AD3, and an example of a neurofibrillary tangle shown in AD4. N = 4, scale bar = 50µm.

B) Representative images of multiplex immunofluorescent labeling of PSP temporal cortex from 3 individuals as in ***A***. Examples of early and pre-tangles shown in PSP1 and PSP3, while examples of astrocytic tufts are shown in PSP2. N = 3, scale bar = 50µm.

Table 1. *MAPT* exon 10 excluders and exon 2 includers replicated across ROSMAP and MSBB datasets

Gene name	Exon 10 Exclusion						Exon 2 Inclusion					
	ROSMAP PFC		MSBB BM10		MSBB BM44		ROSMAP		MSBB BM10		MSBB BM44	
	Rank	Bonferroni	Rank	Bonferroni	Rank	Bonferroni	Rank	Bonferroni	Rank	Bonferroni	Rank	Bonferroni
CPSF3	18	7.04E-30	23	3.38E-09	38	5.62E-14	104	2.14E-12	51	0.00439307	31	1.77E-06
LSM4	16	2.63E-31	38	9.97E-09	20	9.70E-16	3	2.02E-32	47	0.00259308	47	0.000357
POLR2K	**11**	**3.12E-32**	**46**	**3.09E-08**	**29**	**5.11E-15**	**48**	**2.61E-21**	**30**	**3.36E-05**	**26**	**6.61E-07**
PPIE	32	1.44E-25	6	3.90E-12	62	5.60E-12	43	4.05E-22	56	0.01030525	32	6.43E-06
PPIH	6	3.72E-35	8	1.96E-11	4	3.60E-17	27	4.44E-25	37	0.00023555	11	3.92E-09
PUF60	19	8.33E-30	31	7.04E-09	109	2.03E-08	53	1.21E-20	26	2.43E-05	66	3.98E-07
RBM11	**24**	**1.12E-27**	**18**	**1.58E-09**	**36**	**3.28E-14**	**31**	**7.10E-24**	**13**	**1.35E-06**	**2**	**6.00E-11**
RSRC1	**9**	**1.10E-34**	**88**	**1.20E-05**	**42**	**2.61E-13**	**90**	**7.51E-15**	**52**	**0.00769628**	**42**	**0.000152**
RTCB	91	5.42E-15	51	7.76E-08	24	3.06E-15	39	1.57E-22	42	0.00057674	18	1.11E-07
SNRNP25	**34**	**2.92E-25**	**76**	**1.91E-06**	**22**	**1.87E-15**	**18**	**2.97E-27**	**44**	**0.00090501**	**35**	**1.04E-05**
SNRPB	**1**	**8.15E-45**	**1**	**1.43E-15**	**2**	**4.93E-20**	**37**	**3.59E-23**	**9**	**4.69E-07**	**6**	**1.08E-09**
THOC3	**53**	**1.68E-21**	**41**	**1.38E-08**	**41**	**1.80E-13**	**18**	**7.79E-27**	**23**	**1.44E-05**	**3**	**4.94E-10**
THOC7	**29**	**1.93E-26**	**72**	**8.51E-07**	**18**	**9.29E-16**	**5**	**1.93E-31**	**63**	**0.04167339**	**34**	**9.28E-06**
TXNL4A	41	9.62E-24	69	6.84E-07	10	8.71E-17	15	1.75E-27	55	0.00992996	33	8.57E-06

SUPPLEMENTARY INFORMATION

Supplementary Figure Legends

Figure S1. *MAPT* exon 3 inclusion varies between 17q21.31 haplotypes in multiple brain regions. Related to Figure 1.

A) *MAPT* exon expression (RPKM) for each AMP-AD MSBB brain region.

B-C) *MAPT* exon expression (RPKM) in temporal cortex (TCX, ***B.***) and cerebellum (CBM, ***C***.) in the AMP-AD MAYO cohort.

D) PSI values for *MAPT* exons 2, 3 and 10 for each MSBB brain region, split by 17q21.31 haplotype.

E) PSI values for *MAPT* exons 2, 3 and 10 for both MAYO brain regions and ROSMAP PFC, split by 17q21.31 haplotype.

F-H) Total *MAPT* expression (RPKM) for each 17q21.31 haplotype in ***F***. MSBB brain regions, ***G***. MAYO brain regions and ***H***. ROSMAP data. All error bars ± SEM

Figure S2. SF/RBP correlations with *MAPT* exons 2 and 10 inclusion are highly correlated between haplotypes and disease status. Related to Figure 1.

A) Pearson's correlation coefficients between SF/RBP expression and *MAPT* exon 2, 3 and 10 PSI values following unsupervised hierarchical clustering in ROSMAP and MSBB data.

B-C) Pearson's correlation coefficients between SF/RBP expression and *MAPT* exon 3 PSI split between *MAPT* 17q21.31 H1H1 an H2H2 haplotypes in ***B***. ROSMAP and ***C***. MSBB data.

D-E) Pearson's correlation coefficients between SF/RBP expression and *MAPT* exon 2 and exon 10 PSI values split between *MAPT* 17q21.31 H1H1 an H2H2 haplotypes in ***D***. ROSMAP and ***E.*** MSBB data.

F-G) Pearson's correlation coefficients between SF/RBP expression and *MAPT* exons 2, 3 and 10 PSI values, split between AD and control diagnosis in ***F***. ROSMAP and ***G.*** MSBB data.

Figure S3. SF/RBP overexpression influences *MAPT* splicing. Related to Figure 2.

A) Log_{10} fold change (FC) of each SF/RBP following overexpression in SH-SY5Y cells, normalized to *ACTB* endogenous control. Green line indicates average expression of empty vector control. N = 3 from 3 independent experiments. Student's t-test *$p < 0.05$, ***$p < 0.001$.

B) Representative images of western blot validation of SF/RBP overexpression in SH-SY5Y cells with densitometry quantification, normalized to GAPDH (SF/GAPDH). EV = Empty vector control, OE = SF/RBP overexpression. N = 3, Student's t-test *$p < 0.05$, ***$p < 0.001$, n.s = not significant

C-D) Expression fold change (FC) of the ***C***. 4R:3R ratio and ***D***. Exon2:0N ratio in SH-SY5Y cells following SF/RBP overexpression. Grey lines indicate average expression of empty

vector control. N = 3 from 3 independent experiments. Student's t-test *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, n.s = not significant.

Figure S4. *RSRC1* is expressed in multiple neural cell types, including neurons. Related to Figure 3.

A-B) Expression of *RSRC1* in neural cell types detected by snuc-seq in ***A.*** entorhinal cortex and ***B.*** subthalamic nucleus. Average expression is scaled across cell types, and deeper colors represent higher gene expression. Dot size indicates proportion of *RSRC1* expressing cells.

C) UMAP reduction of snuc-seq data from AD and control entorhinal cortex, with clusters colored by cell type, as defined in Grubman *et al*. 2019.

D) UMAP reduction of single-soma data from AD prefrontal cortex, colored by AT8 positive ("Tangle") or AT8 negative ("Non-Tangle") neurons.

E) UMAP reduction of snuc-seq data from PSP and control subthalamic nucleus, with clusters colored by cell type, as defined by positivity for markers described in Grubman *et al.* 2019.

F) Expression of *MAPT* exon 2 includer genes in AD (top) and PSP (bottom) neurons from snuc-seq data. Dot size represents the proportion of neurons expressing the gene, depth of color indicates normalized average gene expression.

Figure S5. N-terminal tau antibodies are specific for each tau isoform. Related to Figure 4.

A) N2a cells overexpressing either 0N3R (0N), 1N3R (1N), 2N3R (2N) or untransfected controls, labelled with Abcam 0N and 2N tau antibodies, and BioLegend 1N tau antibody.

B) Western blot of N2a cells overexpressing each tau isoform, detected by 0N, 1N and 2N tau N-terminal antibodies. Band size compared to tau ladder. GAPDH used as a loading control.

C) IHC detection of N-terminal tau in control, AD and PSP temporal cortex using alternative antibodies (Abcam 0N, 2N) to those in Figure 4 (BioLegend 0N, 2N).

Figure S6. N-terminal tau accumulates in AD and PSP brain. Related to Figure 5.

A) Representative images of multiplex immunofluorescent labeling of AD temporal cortex with AT8 (red), β-amyloid (green), 2N tau (blue), 1N tau (yellow) and 0N tau (orange), and overlay of all three N-terminal tau antibodies in 4 different individuals. Examples of tau accumulation in neurofibrillary tangles (AD1-4) and dystrophic neurites surrounding amyloid plaques (AD2-3). N=4, scale bar = 50µm.

B) Representative images of immunofluorescent labeling of PSP and control temporal cortex as in ***A.*** Examples of neurofibrillary tangles (PSP1, 3) and absence of tau-positive dystrophic neurites surrounding amyloid plaques in PSP (PSP3) and control brain (Control1). N=4, scale bar = 50µm

Supplementary tables

Table S1. SF/RBP genes included in *MAPT* exon PSI analyses

Gene Name	Gene Id	Gene Name	Gene Id	Gene Name	Gene Id	Gene Name	Gene Id	Gene Name	Gene Id
RBM5	ENSG00000003756	*CSTF2*	ENSG00000101811	*HNRNPR*	ENSG00000125944	*TSEN2*	ENSG00000154743	*RBM10*	ENSG00000182872
POLR2J	ENSG00000005075	*USB1*	ENSG00000103005	*RALY*	ENSG00000125970	*CCAR2*	ENSG00000158941	*DDX41*	ENSG00000183258
PTBP1	ENSG00000011304	*ESRP2*	ENSG00000103067	*SCAF1*	ENSG00000126461	*SON*	ENSG00000159140	*SF3A3*	ENSG00000183431
CLK1	ENSG00000013441	*ESRP1*	ENSG00000104413	*NSRP1*	ENSG00000126653	*CELF3*	ENSG00000159409	*ALYREF*	ENSG00000183684
AQR	ENSG00000021776	*HNRNPL*	ENSG00000104824	*HNRNPH2*	ENSG00000126945	*U2AF1*	ENSG00000160201	*SNRNP35*	ENSG00000184209
STRAP	ENSG00000023734	*SNRNP70*	ENSG00000104852	*NAA38*	ENSG00000128534	*U2AF1L4*	ENSG00000161265	*PRPF39*	ENSG00000185246
ZCCHC8	ENSG00000033030	*CLASRP*	ENSG00000104859	*SNRPN*	ENSG00000128739	*SRSF2*	ENSG00000161547	*RBM11*	ENSG00000185272
SKIV2L2	ENSG00000039123	*SF3A2*	ENSG00000104897	*LSM7*	ENSG00000130332	*SNRNP25*	ENSG00000161981	*RNPC3*	ENSG00000185946
POLR2B	ENSG00000047315	*POLR2I*	ENSG00000105258	*LSM4*	ENSG00000130520	*MAGOH*	ENSG00000162385	*SF3B3*	ENSG00000189091
THOC3	ENSG00000051596	*CACTIN*	ENSG00000105298	*PPAN*	ENSG00000130810	*ZNF326*	ENSG00000162664	*PRPF40A*	ENSG00000196504
THRAP3	ENSG00000054118	*HNRNPUL1*	ENSG00000105323	*RBM39*	ENSG00000131051	*CWC22*	ENSG00000163510	*PCBP2*	ENSG00000197111
CCAR1	ENSG00000060339	*PRPF31*	ENSG00000105618	*THOC6*	ENSG00000131652	*THOC7*	ENSG00000163634	*RPS26*	ENSG00000197728
SNRNP40	ENSG00000060688	*SUGP1*	ENSG00000105705	*RBM8A*	ENSG00000131795	*POLR2H*	ENSG00000163882	*AKAP17A*	ENSG00000197976
SFSWAP	ENSG00000061936	*RBM28*	ENSG00000106344	*SNRPA1*	ENSG00000131876	*LSM6*	ENSG00000164167	*TSEN15*	ENSG00000198860
U2AF2	ENSG00000063244	*LSM5*	ENSG00000106355	*ZRANB2*	ENSG00000132485	*TRA2A*	ENSG00000164548	*RBM20*	ENSG00000203867
TNPO3	ENSG00000064419	*CASC3*	ENSG00000108349	*RBM38*	ENSG00000132819	*SLU7*	ENSG00000164609	*RNPS1*	ENSG00000205937
SUGP2	ENSG00000064607	*C1QBP*	ENSG00000108561	*SRRM1*	ENSG00000133226	*RP9*	ENSG00000164610	*DHX16*	ENSG00000206486
YBX1	ENSG00000065978	*LUC7L3*	ENSG00000108848	*PRPF38B*	ENSG00000134186	*FASTK*	ENSG00000164896	*SCAF8*	ENSG00000213079
DHX8	ENSG00000067596	*EFTUD2*	ENSG00000108883	*RBM17*	ENSG00000134453	*HNRNPK*	ENSG00000165119	*RBMXL1*	ENSG00000213516
PABPC1	ENSG00000070756	*SUPT6H*	ENSG00000109111	*PRPF38A*	ENSG00000134748	*PCF11*	ENSG00000165494	*DDX47*	ENSG00000213782
CPSF1	ENSG00000071894	*DHX15*	ENSG00000109606	*SRPK2*	ENSG00000135250	*PRPF18*	ENSG00000165630	*SRSF10*	ENSG00000215699
ZNF638	ENSG00000075292	*PPARGC1A*	ENSG00000109819	*SYNCRIP*	ENSG00000135316	*CPSF2*	ENSG00000165934	*PAPOLB*	ENSG00000218823
GPATCH1	ENSG00000076650	*DCPS*	ENSG00000110063	*HNRNPA1*	ENSG00000135486	*NUDT21*	ENSG00000167005	*LSM2*	ENSG00000224979
MBNL3	ENSG00000076770	*PRPF19*	ENSG00000110107	*DHX9*	ENSG00000135829	*SNRPD1*	ENSG00000167088	*RBMY1J*	ENSG00000226941
XAB2	ENSG00000076924	*RPS13*	ENSG00000110700	*SRSF1*	ENSG00000136450	*SRRM2*	ENSG00000167978	*RBMY1A1*	ENSG00000234414
SNRPA	ENSG00000077312	*PRPF40B*	ENSG00000110844	*TRA2B*	ENSG00000136527	*POLR2G*	ENSG00000168002	*DDX39B*	ENSG00000237889
RBFOX1	ENSG00000078328	*MAGOHB*	ENSG00000111196	*PRPF4*	ENSG00000136875	*SF1*	ENSG00000168066	*PPIL3*	ENSG00000240344
THOC1	ENSG00000079134	*SRSF9*	ENSG00000111786	*NCBP1*	ENSG00000136937	*CDC40*	ENSG00000168438	*ISY1*	ENSG00000240682
DDX1	ENSG00000079785	*SRSF3*	ENSG00000112081	*BUD13*	ENSG00000137656	*SNRNP48*	ENSG00000168566	*RBMY1E*	ENSG00000242389
GEMIN5	ENSG00000082516	*QKI*	ENSG00000112531	*DBR1*	ENSG00000138231	*USP39*	ENSG00000168883	*RBMY1B*	ENSG00000242875
PPIE	ENSG00000084072	*PRPF4B*	ENSG00000112739	*HNRNPD*	ENSG00000138668	*HNRNPH1*	ENSG00000169045	*RBMY1D*	ENSG00000244395
RBM22	ENSG00000086589	*CLK4*	ENSG00000113240	*ZCRB1*	ENSG00000139168	*CD2BP2*	ENSG00000169217	*YTHDC1*	ENSG00000257413
SF3B2	ENSG00000087365	*PPWD1*	ENSG00000113593	*SCAF11*	ENSG00000139218	*ZRSR2*	ENSG00000169249	*FRG1*	ENSG00000260380
KHSRP	ENSG00000088247	*NCBP2*	ENSG00000114503	*SNRPF*	ENSG00000139343	*PCBP1*	ENSG00000169564	*CWC15*	ENSG00000261974
FUS	ENSG00000089280	*SF3B14*	ENSG00000115128	*HNRNPA1L2*	ENSG00000139675	*RBMY1F*	ENSG00000169800	*DDX5*	ENSG00000263077

PAPOLA	ENSG00000090060	*SF3B1*	ENSG00000115524	*SRRM4*	ENSG00000139767	*HNRNPF*	ENSG00000169813	*PRPF3*	ENSG00000265228
PDCD7	ENSG00000090470	*SRSF7*	ENSG00000115875	*MBNL2*	ENSG00000139793	*SF3B5*	ENSG00000169976	*RBFOX3*	ENSG00000267483
HNRNPC	ENSG00000092199	*TIA1*	ENSG00000116001	*NOVA1*	ENSG00000139910	*HNRNPA3*	ENSG00000170144	*AFF2*	ENSG00000269754
GEMIN2	ENSG00000092208	*SRSF4*	ENSG00000116350	*CELF6*	ENSG00000140488	*LSM3*	ENSG00000170860		
CDC5L	ENSG00000096401	*SFPQ*	ENSG00000116560	*DHX38*	ENSG00000140829	*TSEN34*	ENSG00000170892		
HNRNPH3	ENSG00000096746	*SRSF11*	ENSG00000116754	*TXNL4B*	ENSG00000140830	*NCBP2L*	ENSG00000170935		
HNRNPM	ENSG00000099783	*TTF2*	ENSG00000116830	*NOL3*	ENSG00000140939	*FAM98B*	ENSG00000171262		
POLR2E	ENSG00000099817	*PTBP2*	ENSG00000117569	*EIF4A3*	ENSG00000141543	*PLRG1*	ENSG00000171566		
SF3A1	ENSG00000099995	*SYF2*	ENSG00000117614	*TXNL4A*	ENSG00000141759	*PPIH*	ENSG00000171960		
SNRPD3	ENSG00000100028	*PPP1R8*	ENSG00000117751	*GEMIN7*	ENSG00000142252	*CLP1*	ENSG00000172409		
DGCR14	ENSG00000100056	*CPSF3*	ENSG00000119203	*SF3B4*	ENSG00000143368	*RBM4B*	ENSG00000173914		
TFIP11	ENSG00000100109	*PTBP3*	ENSG00000119314	*HNRNPLL*	ENSG00000143889	*RBM4*	ENSG00000173933		
NHP2L1	ENSG00000100138	*RBM25*	ENSG00000119707	*SNRPG*	ENSG00000143977	*PRPF8*	ENSG00000174231		
POLR2F	ENSG00000100142	*SMNDC1*	ENSG00000119953	*SNRNP200*	ENSG00000144028	*DDX23*	ENSG00000174243		
RTCB	ENSG00000100220	*WBP4*	ENSG00000120688	*POLR2D*	ENSG00000144231	*RSRC1*	ENSG00000174891		
THOC5	ENSG00000100296	*TARDBP*	ENSG00000120948	*DDX46*	ENSG00000145833	*LSM1*	ENSG00000175324		
ZMAT5	ENSG00000100319	*HNRNPA2B1*	ENSG00000122566	*NONO*	ENSG00000147140	*SART1*	ENSG00000175467		
RBFOX2	ENSG00000100320	*DDX39A*	ENSG00000123136	*RBMX*	ENSG00000147274	*CSTF3*	ENSG00000176102		
PHF5A	ENSG00000100410	*WDR83*	ENSG00000123154	*POLR2K*	ENSG00000147669	*CLK2*	ENSG00000176444		
SNW1	ENSG00000100603	*SRSF6*	ENSG00000124193	*CELF1*	ENSG00000149187	*POLR2L*	ENSG00000177700		
SRSF5	ENSG00000100650	*SNRNP27*	ENSG00000124380	*CPSF7*	ENSG00000149532	*HNRNPA0*	ENSG00000177733		
ACIN1	ENSG00000100813	*MPHOSPH10*	ENSG00000124383	*TRPT1*	ENSG00000149743	*CLK3*	ENSG00000179335		
PABPN1	ENSG00000100836	*SNRPC*	ENSG00000124562	*SAP18*	ENSG00000150459	*RBM15B*	ENSG00000179837		
PNN	ENSG00000100941	*UPF3B*	ENSG00000125351	*GEMIN6*	ENSG00000152147	*PUF60*	ENSG00000179950		
CSTF1	ENSG00000101138	*GTF2F1*	ENSG00000125651	*MBNL1*	ENSG00000152601	*POLR2A*	ENSG00000181222		
PRPF6	ENSG00000101161	*THOC2*	ENSG00000125676	*CWC27*	ENSG00000153015	*LSM10*	ENSG00000181817		
CRNKL1	ENSG00000101343	*SNRPD2*	ENSG00000125743	*HNRNPU*	ENSG00000153187	*SNRPE*	ENSG00000182004		
DHX35	ENSG00000101452	*SNRPB*	ENSG00000125835	*SREK1*	ENSG00000153914	*TSEN54*	ENSG00000182173		
CELF4	ENSG00000101489	*SNRPB2*	ENSG00000125870	*SRSF12*	ENSG00000154548	*ARL6IP4*	ENSG00000182196		

Table S2. SF/RBPs significantly associated with *MAPT* exon 2 or 10 splicing in ROSMAP, MSBB BM10 and MSBB BM44 RNA-seq data

Exon 10 Excluders

Gene name	ROSMAP PFC		MSBB BM10		MSBB BM44	
	Bonferroni *p*	Rank	Bonferroni *p*	Rank	Bonferroni *p*	Rank
ALYREF	1.49E-16	82	1.27E-09	16	9.35E-16	19
ARL6IP4	0.00648	182	1.40E-08	42	2.24E-15	23
BUD13	3.51E-10	121	6.60E-07	68	1.18E-08	107
C1QBP	3.12E-24	40	1.30E-08	40	1.87E-14	32
CACTIN	1.69E-07	141	5.59E-05	100	0.00118829	166
CCAR2	4.46E-35	7	1.22E-05	89	1.32E-15	21
CD2BP2	8.26E-18	73	5.63E-05	101	3.40E-05	153
CDC40	1.73E-18	70	6.93E-06	85	2.73E-12	55
CELF1	3.56E-16	83	0.0329038	148	0.01450365	175
DBR1	8.46E-11	111	7.73E-10	12	3.61E-11	76
DCPS	5.57E-11	109	0.0002914	108	1.32E-11	70
DDX1	1.07E-31	15	6.76E-06	84	2.56E-16	14
DDX39A	1.44E-07	140	5.69E-05	102	0.11439467	193
DDX41	7.89E-39	3	1.54E-09	17	1.17E-18	3
DDX47	7.64E-15	92	0.0003432	111	5.61E-07	132
DHX38	1.03E-07	137	2.39E-08	44	3.51E-06	145
EFTUD2	3.23E-25	35	2.08E-05	93	6.25E-10	89
EIF4A3	7.04E-12	103	2.45E-09	20	1.14E-11	69
FAM98B	9.80E-14	97	0.0010702	120	8.50E-10	90
GTF2F1	1.50E-26	28	0.0004285	114	1.89E-09	95
HNRNPA0	7.07E-23	44	1.74E-07	57	3.95E-17	5
HNRNPA1L2	5.60E-16	85	4.76E-09	28	1.17E-16	13
HNRNPD	1.09E-24	38	8.12E-07	71	6.96E-08	119
HNRNPK	8.05E-07	149	1.47E-07	55	2.73E-11	74
HNRNPL	5.60E-31	17	2.48E-08	45	1.08E-08	105
HNRNPUL1	4.49E-06	156	0.0020035	124	0.00154667	168
ISY1	1.21E-07	138	0.0003112	109	4.03E-07	130
LSM10	2.47E-08	135	1.26E-07	54	1.12E-14	30
LSM3	3.04E-18	71	1.18E-09	15	4.76E-16	16
LSM6	0.000125	167	7.52E-09	34	1.63E-14	31
LSM7	2.11E-10	118	1.96E-07	58	4.37E-10	86
MAGOH	1.10E-10	114	5.43E-06	81	3.40E-12	58
MAGOHB	0.000187	169	9.87E-09	37	5.53E-12	61
MPHOSPH10	6.55E-19	67	7.47E-07	70	1.73E-12	52
NCBP2	7.40E-06	160	0.0320488	146	2.90E-05	152
NHP2L1	4.85E-22	50	1.07E-06	73	6.79E-17	8

NUDT21	5.03E-22	51	4.26E-07	65	2.17E-14	33
PHF5A	0.000281	171	7.63E-09	35	6.77E-12	65
POLR2A	9.49E-11	112	3.91E-07	64	9.88E-13	48
POLR2B	7.30E-18	72	2.32E-06	77	2.91E-12	57
POLR2E	1.71E-21	54	5.55E-06	82	1.23E-12	50
POLR2G	0.00014	168	9.69E-10	13	3.57E-13	45
POLR2I	5.37E-27	27	1.98E-09	19	1.50E-13	40
POLR2L	0.000231	170	1.43E-05	91	7.32E-07	135
PPAN	6.91E-32	13	8.06E-08	52	3.68E-07	129
PPIL3	3.61E-16	84	4.70E-09	27	4.19E-14	37
PPP1R8	0.000531	174	0.0003821	113	5.61E-09	102
PRPF19	1.45E-37	5	1.40E-10	10	3.37E-20	1
PRPF31	1.36E-16	80	1.63E-07	56	1.20E-10	78
PRPF4	1.71E-18	69	4.77E-09	29	4.20E-16	15
PRPF40A	0.0182	185	0.0012857	121	0.03739879	185
PRPF6	2.33E-27	26	1.12E-09	14	2.43E-12	54
RALY	2.01E-20	58	3.19E-07	62	6.17E-08	117
RBFOX1	4.80E-28	22	6.67E-06	83	1.02E-16	12
RBFOX2	3.48E-20	60	0.0128532	140	2.57E-10	83
RBM22	2.26E-07	142	0.0090851	136	1.31E-07	124
SAP18	0.00377	179	1.42E-05	90	1.85E-11	72
SART1	2.53E-11	107	1.59E-05	92	1.23E-07	123
SCAF1	2.44E-38	4	0.0001379	104	1.93E-06	139
SF3A1	2.71E-10	120	5.39E-05	99	3.13E-07	128
SF3A2	7.79E-21	56	3.05E-09	22	9.73E-07	136
SF3A3	7.07E-20	62	2.09E-07	59	3.65E-15	27
SF3B2	7.33E-11	110	0.0002469	107	3.55E-08	114
SF3B4	4.82E-09	128	7.13E-09	33	5.66E-10	88
SF3B5	1.59E-17	76	2.56E-09	21	8.74E-12	66
SKIV2L2	1.27E-14	94	0.0222023	144	2.60E-06	143
SNRNP200	3.41E-25	36	0.000319	110	2.41E-09	97
SNRNP40	3.48E-13	100	6.62E-09	30	9.71E-12	68
SNRPA	5.26E-06	158	4.64E-06	80	1.12E-08	106
SNRPA1	3.10E-22	47	6.53E-14	2	5.13E-17	6
SNRPC	2.79E-11	108	1.03E-08	39	2.68E-13	43
SNRPD1	4.28E-20	61	3.17E-07	61	8.48E-17	9
SNRPD2	7.79E-32	14	4.63E-09	26	3.50E-15	25
SNRPE	7.71E-12	104	7.06E-09	32	2.82E-14	35
SNRPF	1.33E-16	79	7.73E-10	11	4.48E-15	28
SNRPN	2.09E-24	39	0.0061774	132	6.11E-12	64
SNW1	5.24E-21	55	8.05E-06	86	2.10E-10	81
SRPK2	2.22E-10	119	0.0088929	135	1.49E-09	94
SRRM4	1.27E-19	63	0.004045	129	5.39E-10	87
SRSF2	1.88E-28	21	9.69E-13	4	3.37E-13	44
SRSF3	1.29E-05	162	4.51E-05	97	5.51E-08	115

SRSF5	3.72E-25	37	0.0001468	105	0.03313692	183
STRAP	4.08E-10	122	6.21E-05	103	6.01E-14	39
SUGP1	4.75E-22	49	2.47E-06	78	3.91E-06	146
SUPT6H	9.93E-17	78	0.0479101	149	0.01684213	177
THOC6	8.28E-05	165	4.20E-09	25	9.19E-10	92
THRAP3	1.10E-22	45	3.40E-09	24	2.82E-12	56
TRPT1	3.84E-26	31	2.38E-13	3	8.62E-10	91
TSEN34	3.61E-07	145	5.37E-11	9	1.52E-12	51
U2AF1L4	1.41E-21	52	6.33E-12	7	3.64E-15	26
USB1	1.09E-15	89	0.0064505	133	0.02563577	182
WDR83	4.08E-22	48	5.30E-05	98	2.10E-09	96
XAB2	3.28E-23	43	2.28E-07	60	0.01886529	178
ZRANB2	7.48E-16	88	0.0126366	139	0.00122737	167

Exon 10 Includers

Gene name	ROSMAP PFC		MSBB BM10		MSBB BM44	
	Bonferroni *p*	Rank	Bonferroni *p*	Rank	Bonferroni *p*	Rank
ESRP2	-	-	1.71E-06	2	1.54E-07	2
MBNL3	-	-	4.19E-05	16	5.46E-07	3
NAA38	-	-	2.99E-05	15	2.38E-06	6
PAPOLB	-	-	2.69E-07	6	2.26E-10	1
SRSF11	1.49E-09	6	0.0010663	22	-	-

Exon 2 Excluders

Gene name	ROSMAP PFC		MSBB BM10		MSBB BM44	
	Bonferroni *p*	Rank	Bonferroni *p*	Rank	Bonferroni *p*	Rank
CASC3	5.93E-05	9	0.000366	45	2.54E-05	85
NSRP1	0.0167442	12	0.0003789	46	9.43E-11	27
PRPF38B	2.20E-07	5	0.0007566	56	1.02E-05	79
QKI	2.55E-07	6	6.23E-08	13	2.49E-10	31
RBM38	0.00020408	10	3.70E-09	5	1.07E-08	52
YBX1	1.02E-06	7	2.92E-11	1	1.31E-12	9
ZNF326	3.64E-05	8	0.0001399	40	2.11E-09	47

Exon 2 Includers

Gene name	ROSMAP PFC		MSBB BM10		MSBB BM44	
	Bonferroni *p*	Rank	Bonferroni *p*	Rank	Bonferroni *p*	Rank
C1QBP	5.19E-28	11	0.0013978	19	8.97E-08	14
CD2BP2	4.18E-34	2	0.020757	28	0.04234665	53
CLP1	1.71E-16	51	1.20E-06	4	1.01E-09	4
DBR1	0.00019092	83	6.03E-05	13	2.11E-09	8
EIF4A3	1.39E-12	57	8.94E-05	14	1.39E-09	7
GEMIN2	2.33E-18	45	1.82E-06	6	8.53E-08	13
GEMIN6	3.07E-22	33	4.23E-05	12	5.93E-08	12

GEMIN7	1.30E-26	16	0.0008638	17	0.0002185	29
LSM1	3.25E-17	49	1.12E-06	3	1.25E-06	18
LSM6	1.43E-07	75	4.38E-08	1	7.56E-10	3
PLRG1	8.51E-11	69	0.007972	24	0.00038382	34
POLR2H	3.56E-06	78	0.0016605	20	0.02689017	52
PPP1R8	1.20E-08	72	3.51E-06	7	4.65E-07	16
PRPF4	8.18E-18	46	0.015142	27	3.86E-08	10
SF3B5	1.75E-25	19	3.52E-06	8	5.25E-08	11
TSEN15	9.93E-22	37	6.31E-06	9	1.26E-09	6

Table S3. Summary of human brain tissues and sources used in this study

ID	Dx	AoD	Sex	Source
Western blot & ISO-seq				
9	Control	55	Male	ISMMS Neuropathology Research Core and Brain Bank
10	Control	63	Male	ISMMS Neuropathology Research Core and Brain Bank
11	Control	73	Male	ISMMS Neuropathology Research Core and Brain Bank
13	Control	74	Female	ISMMS Neuropathology Research Core and Brain Bank
19	Control	18	Male	ISMMS Neuropathology Research Core and Brain Bank
22	Control	72	Female	ISMMS Neuropathology Research Core and Brain Bank
S08611	AD	81	Male	NIH Neurobiobank - Harvard
38917	AD	90	Female	NIH Neurobiobank - Mt.Sinai
S13471	AD	86	Female	NIH Neurobiobank - Harvard
S15985	AD	85	Female	NIH Neurobiobank - Harvard
S05093	AD	85	Male	NIH Neurobiobank - Harvard
3853	AD	85	Male	NIH Neurobiobank - Harvard
5737	PSP	73	Male	NIH Neurobiobank - Maryland
6191	PSP	62	Female	NIH Neurobiobank - Maryland
5126	PSP	72	Male	NIH Neurobiobank - Maryland
6043	PSP	67	Female	NIH Neurobiobank - Maryland
6225	PSP	87	Male	NIH Neurobiobank - Maryland
6085	PSP	88	Female	NIH Neurobiobank - Maryland
IHC & IF				
Control1	Control	73	Male	ISMMS Neuropathology Research Core and Brain Bank
Control2	Control	68	Female	ISMMS Neuropathology Research Core and Brain Bank
Control3	Control	66	Female	ISMMS Neuropathology Research Core and Brain Bank
AD1	AD	85	Female	ISMMS Neuropathology Research Core and Brain Bank
AD2	AD	95	Male	ISMMS Neuropathology Research Core and Brain Bank
AD3	AD	78	Female	ISMMS Neuropathology Research Core and Brain Bank
AD4	AD	66	Female	ISMMS Neuropathology Research Core and Brain Bank
PSP1	PSP	64	Male	ISMMS Neuropathology Research Core and Brain Bank
PSP2	PSP	71	Male	ISMMS Neuropathology Research Core and Brain Bank
PSP3	PSP	69	Male	ISMMS Neuropathology Research Core and Brain Bank

Dx = Diagnosis. AoD = Age of death.

Table S4. Summary of all antibodies and dilutions used in this study

Antibody	Source	Catalog number	WB	IF	IHC	Opal	Opal fluorophore
POLR2K	ThermoFisher Scientific	PA5-68172	1:200	-	-	-	-
RBM11	ThermoFisher Scientific	PA5-31249	1:1000	-	-	-	-
RSRC1	ThermoFisher Scientific	PA5-20899	1:1000	-	1:200	-	-
SNRPB	Abcam	ab155026	1:500	-	-	-	-
SNRNP25	Novus Biologicals	NBP2-32028	1:1000	-	-	-	-
THOC3	Bethyl Laboratories	A304-870A	1:2000	-	-	-	-
THOC7	ThermoFisher Scientific	PA5-31594	1:500	-	-	-	-
0N Tau	Abcam	ab218199	1:1000	1:500	1:400	1:500	620
0N Tau	BioLegend		-	-	1:500	-	-
1N Tau	BioLegend	823901	1:500	1:500	1:500	1:250	570
2N Tau	Abcam	ab218316	1:1000	1:500	1:400	1:500	480
2N Tau	BioLegend		-	-	1:500	-	-
AT8	ThermoFisher Scientific	MN1020	-	-	1:1000	1:2000	690
β-amyloid	BioLegend	800701	-	-	1:5000	1:1000	520
GAPDH	Abcam	ab181602	1:10,000	-	-	-	-
HRP Goat Anti-Rabbit	Vector Laboratories	PI-1000	1:20,000	-	-	-	-
HRP Horse Anti-Mouse	Vector Laboratories	PI-2000	1:20,000	-	-	-	-
Donkey anti-Rabbit AlexaFluor 488	ThermoFisher Scientific	A-21206	-	1:100	-	-	-
Donkey anti-Mouse AlexaFluor 568	ThermoFisher Scientific	A-10037	-	1:100	-	-	-
DISCOVERY Universal Secondary antibody	Roche	760-4205	-	-	Manufacturer's recommendation	-	-

REFERENCES

1. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**, 1413–1415 (2008).
2. Wang, F., Zhao, Y., Hao, Y. & Tan, Z. Identification of low-abundance alternatively spliced mRNA variants by exon exclusive reverse transcriptase polymerase chain reaction. *Anal. Biochem.* **383**, 307–310 (2008).
3. Nilsen, T. W. & Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–63 (2010).
4. Morfini, G. a *et al.* Minisymposium: Axonal Transport Defects in Neurodegenerative Diseases. *J. Neurosci.* **29**, 12776–12786 (2009).
5. Iqbal, K., Liu, F., Gong, C.-X. & Grundke-Iqbal, I. Tau in Alzheimer disease and related tauopathies. *Curr. Alzheimer Res.* **7**, 656–664 (2010).
6. Hernández, F., Merchán-Rubira, J., Vallés-Saiz, L., Rodríguez-Matellán, A. & Avila, J. Differences Between Human and Murine Tau at the N-terminal End. *Front. Aging Neurosci.* **12**, 1–6 (2020).
7. Hernández, F., Pérez, M., de Barreda, E. G., Goñi-Oliver, P. & Avila, J. Tau as a molecular marker of development, aging and neurodegenerative disorders. *Curr. Aging Sci.* **1**, 56–61 (2008).
8. Hefti, M. M. *et al.* High-resolution temporal and regional mapping of MAPT expression and splicing in human brain development. *PLoS One* **13**, 1–14 (2018).
9. Neve, R. L., Harris, P., Kosik, K. S., Kurnit, D. M. & Donlon, T. a. Identification of cDNA clones for the human microtubule-associated protein tau and chromosomal

localization of the genes for tau and microtubule-associated protein 2. *Mol. Brain Res.* **1**, 271–280 (1986).

10. Gao, Q. S. *et al.* Complex regulation of tau exon 10, whose missplicing causes frontotemporal dementia. *J. Neurochem.* **74**, 490–500 (2000).
11. Boutajangout, A., Boom, A., Leroy, K. & Brion, J. P. Expression of tau mRNA and soluble tau isoforms in affected and non-affected brain areas in Alzheimer's disease. *FEBS Lett.* **576**, 183–189 (2004).
12. Trabzuni, D. *et al.* MAPT expression and splicing is differentially regulated by brain region: Relation to genotype and implication for tauopathies. *Hum. Mol. Genet.* **21**, 4094–4103 (2012).
13. Conrad, C. *et al.* Single molecule profiling of tau gene expression in Alzheimer's disease. *J. Neurochem.* **103**, 1228–1236 (2007).
14. Goode, B. L. & Feinstein, S. C. Identification of a novel microtubule binding and assembly domain in the developmentally regulated inter-repeat region of tau. *J. Cell Biol.* **124**, 769–781 (1994).
15. Derisbourg, M. *et al.* Role of the Tau N-terminal region in microtubule stabilization revealed by new endogenous truncated forms. *Sci. Rep.* **5**, 1–10 (2015).
16. Caillet-Boudin, M. L. *et al.* Brain pathology in myotonic dystrophy: When tauopathy meets spliceopathy and RNAopathy. *Front. Mol. Neurosci.* **6**, 1–20 (2014).
17. Kanai, Y., Chen, J. & Hirokawa, N. Microtubule bundling by tau proteins in vivo: analysis of functional domains. *EMBO J.* **11**, 3953–3961 (1992).
18. Espinoza, M., de Silva, R., Dickson, D. W. & Davies, P. Differential incorporation of tau isoforms in Alzheimer's disease. *J. Alzheimers. Dis.* **14**, 1–16 (2008).

19. D'Souza, I. & Schellenberg, G. D. Regulation of tau isoform expression and dementia. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1739**, 104–115 (2005).

20. Stanford, P. M. *et al.* Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene Expansion of the disease phenotype caused by tau gene mutations. *Brain* 880–893 (2000).

21. Skoglund, L. *et al.* The tau S305S mutation causes frontotemporal dementia with parkinsonism. *Eur. J. Neurol.* **15**, 156–161 (2008).

22. Hutton, M. *et al.* Association of missense and 5' -splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, (1998).

23. Spillantini, M. G. *et al.* A Novel tau Mutation (N296N) in Familial Dementia with Swollen Achromatic Neurons and Corticobasal Inclusion. *Ann. Neurol.* **48**, 939–943 (2000).

24. Wesseling, H. *et al.* Tau PTM Profiles Identify Patient Heterogeneity and Stages of Alzheimer's Disease. *Cell* **183**, 1699-1713.e13 (2020).

25. Glatz, D. C. *et al.* The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease. *J. Neurochem.* **96**, 635–644 (2006).

26. Ingelsson, M. *et al.* No alteration in tau exon 10 alternative splicing in tangle-bearing neurons of the Alzheimer's disease brain. *Acta Neuropathol.* **112**, 439–449 (2006).

27. Ginsberg, S. D., Che, S., Counts, S. E. & Mufson, E. J. Shift in the ratio of three-repeat tau and four-repeat tau mRNAs in individual cholinergic basal forebrain neurons in mild cognitive impairment and Alzheimer's disease. *J. Neurochem.* **96**, 1401–1408 (2006).

28. Qian, W. & Liu, F. Regulation of alternative splicing of tau exon 10. *Neurosci. Bull.* **30**,

367–377 (2014).

29. Bruch, J., Xu, H., De Andrade, A. & Höglinger, G. Mitochondrial Complex 1 Inhibition Increases 4-Repeat Isoform Tau by SRSF2 Upregulation. *PLoS One* **9**, e113070 (2014).

30. van Abel, D. *et al.* SFRS7-mediated splicing of tau exon 10 is directly regulated by STOX1A in glial cells. *PLoS One* **6**, (2011).

31. D'Souza, I. & Schellenberg, G. D. Arginine/serine-rich protein interaction domain-dependent modulation of a tau exon 10 splicing enhancer: Altered interactions and mechanisms for functionally antagonistic FTDP-17 mutations ??280K and N279K. *J. Biol. Chem.* **281**, 2460–2469 (2006).

32. Jiang, Z. *et al.* Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2β. *J. Biol. Chem.* **278**, 18997–19007 (2003).

33. Storbeck, M. *et al.* Neuronal-specific deficiency of the splicing factor Tra2b causes apoptosis in neurogenic areas of the developing mouse brain. *PLoS One* **9**, (2014).

34. Kar, A., Havlioglu, N., Tarn, W. Y. & Wu, J. Y. RBM4 interacts with an intronic element and stimulates tau exon 10 inclusion. *J. Biol. Chem.* **281**, 24479–24488 (2006).

35. Wang, Y., Gao, L., Tse, S. W. & Andreadis, A. Heterogeneous nuclear ribonucleoprotein E3 modestly activates splicing of tau exon 10 via its proximal downstream intron, a hotspot for frontotemporal dementia mutations. *Gene* **451**, 23–31 (2010).

36. Broderick, J., Wang, J. & Andreadis, A. Heterogeneous nuclear ribonucleoprotein E2 binds to tau exon 10 and moderately activates its splicing. *Gene* **331**, 107–114 (2004).

37. Caillet-Boudin, M.-L., Buée, L., Sergeant, N. & Lefebvre, B. Regulation of human MAPT gene expression. *Mol. Neurodegener.* **10**, 28 (2015).

38. Beevers, J. E. *et al.* MAPT Genetic Variation and Neuronal Maturity Alter Isoform Expression Affecting Axonal Transport in iPSC-Derived Dopamine Neurons. *Stem Cell Reports* **9**, 587–599 (2017).

39. Lai, M. C. *et al.* Haplotype-specific MAPT exon 3 expression regulated by common intronic polymorphisms associated with Parkinsonian disorders. *Mol. Neurodegener.* **12**, 1–16 (2017).

40. Jong, S. De *et al.* Common inversion polymorphism at 17q21 . 31 affects expression of multiple genes in tissue-specific manner. *BMC Genomics* **13**, 1 (2012).

41. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).

42. Yu, Q., Guo, J. & Zhou, J. A minimal length between tau exon 10 and 11 is required for correct splicing of exon 10. *J. Neurochem.* **90**, 164–172 (2004).

43. 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 (2019).

44. Sharma, A. *et al.* Single-cell atlas of progressive supranuclear palsy reveals a distinct hybrid glial cell population. *bioRxiv* 2021.04.11.439393 (2021).

45. Otero-Garcia, M. *et al.* Single-soma transcriptomics of tangle-bearing neurons in Alzheimer's disease reveals the signatures of tau-associated synaptic dysfunction Marcos. *bioRxiv* (2020).

46. Kim, W. H., Lee, S. & Hall, G. F. Secretion of human tau fragments resembling CSF-tau in Alzheimer's disease is modulated by the presence of the exon 2 insert. *FEBS Lett.* **584**, 3085–3088 (2010).

47. Combs, B., Hamel, C. & Kanaan, N. M. Pathological conformations involving the amino terminus of tau occur early in Alzheimer's disease and are differentially detected by monoclonal antibodies. *Neurobiol. Dis.* **94**, 18–31 (2016).

48. Brandt, R., Léger, J. & Lee, G. Interaction of tau with the neural plasma membrane mediated by tau's amino-terminal projection domain. *J. Cell Biol.* **131**, 1327–1340 (1995).

49. Gauthier-Kemper, A. *et al.* Annexins A2 and A6 interact with the extreme N terminus of tau and thereby contribute to tau's axonal localization. *J. Biol. Chem.* **293**, 8065–8076 (2018).

50. Stefanoska, K. *et al.* An N-terminal motif unique to primate tau enables differential protein–protein interactions. *J. Biol. Chem.* **293**, 3710–3719 (2018).

51. Liu, C. & Götz, J. Profiling murine tau with 0N, 1N and 2N isoform-specific antibodies in brain and peripheral organs reveals distinct subcellular localization, with the 1N isoform being enriched in the nucleus. *PLoS One* **8**, 1–18 (2013).

52. Zhong, Q., Congdon, E. E., Nagaraja, H. N. & Kuret, J. Tau isoform composition influences rate and extent of filament formation. *J. Biol. Chem.* **287**, 20711–20719 (2012).

53. King, M. E., Gamblin, T. C., Kuret, J. & Binder, L. I. Differential assembly of human tau isoforms in the presence of arachidonic acid. *J. Neurochem.* **74**, 1749–1757 (2000).

54. Liu, C., Song, X., Nisbet, R. & Götz, J. Co-immunoprecipitation with Tau isoform-specific antibodies reveals distinct protein interactions and highlights a putative role for 2N Tau in disease. *J. Biol. Chem.* **291**, 8173–8188 (2016).

55. Apicco, D. J. *et al.* Dysregulation of RNA Splicing in Tauopathies. *Cell Rep.* **29**, 4377-4388.e4 (2019).

56. Cazalla, D., Newton, K. & Cáceres, J. F. A Novel SR-Related Protein Is Required for the

Second Step of Pre-mRNA Splicing. *Mol. Cell. Biol.* **25**, 2969–2980 (2005).

57. Scala, M. *et al.* RSRC1 loss-of-function variants cause mild to moderate autosomal recessive intellectual disability. *Brain* **1**, (2020).

58. Perez, Y. *et al.* RSRC1 mutation affects intellect and behaviour through aberrant splicing and transcription, downregulating IGFBP3. *Brain* **141**, 961–970 (2018).

59. Pedrotti, S., Busà, R., Compagnucci, C. & Sette, C. The RNA recognition motif protein RBM11 is a novel tissue-specific splicing regulator. *Nucleic Acids Res.* **40**, 1021–1032 (2012).

60. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

61. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner : fast , accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, (2013).

62. Katz, Y., Wang, E. T., Airoldi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).

63. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

64. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions , technologies , and species. *Nat. Biotechnol.* **36**, (2018).

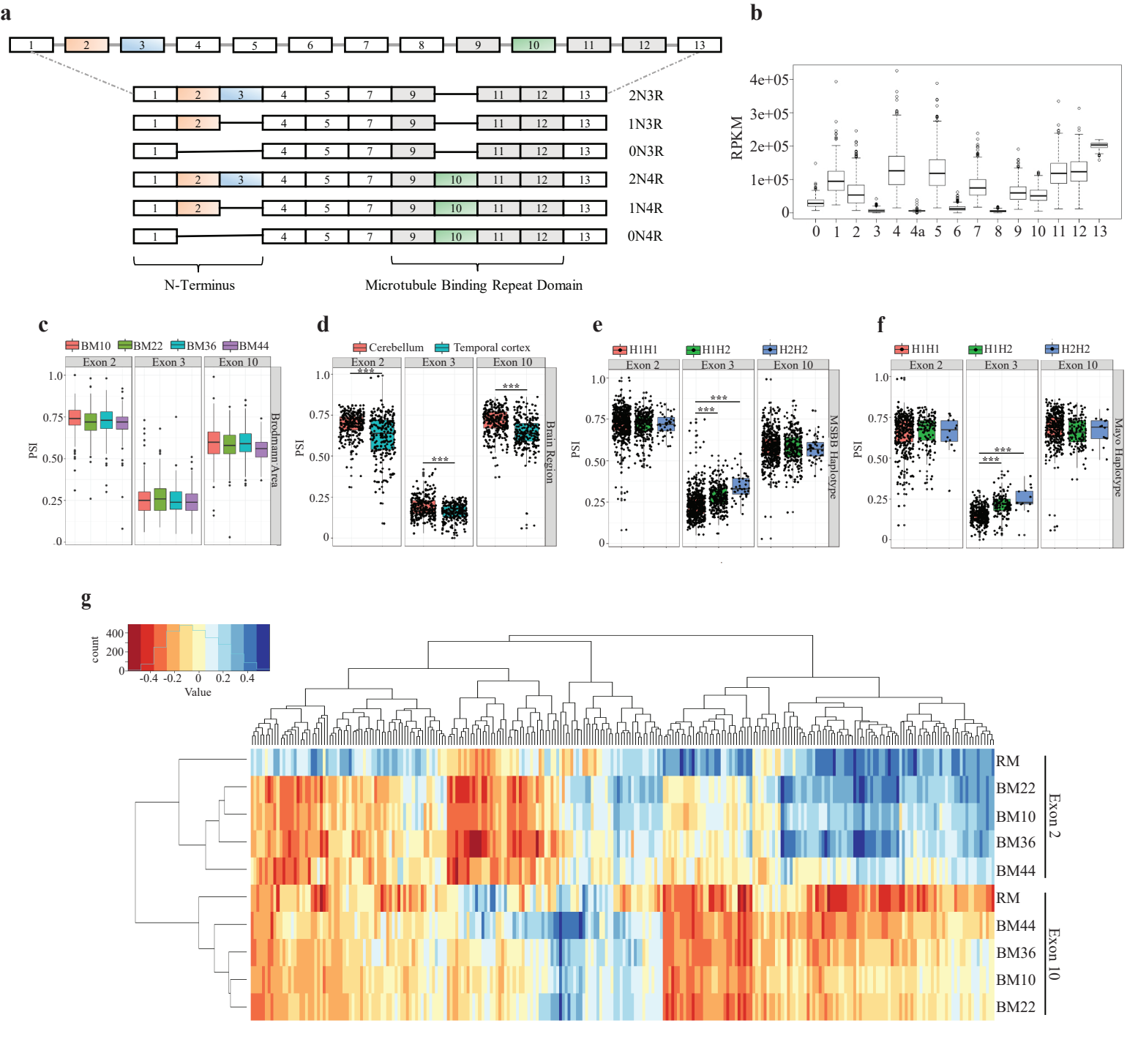
65. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

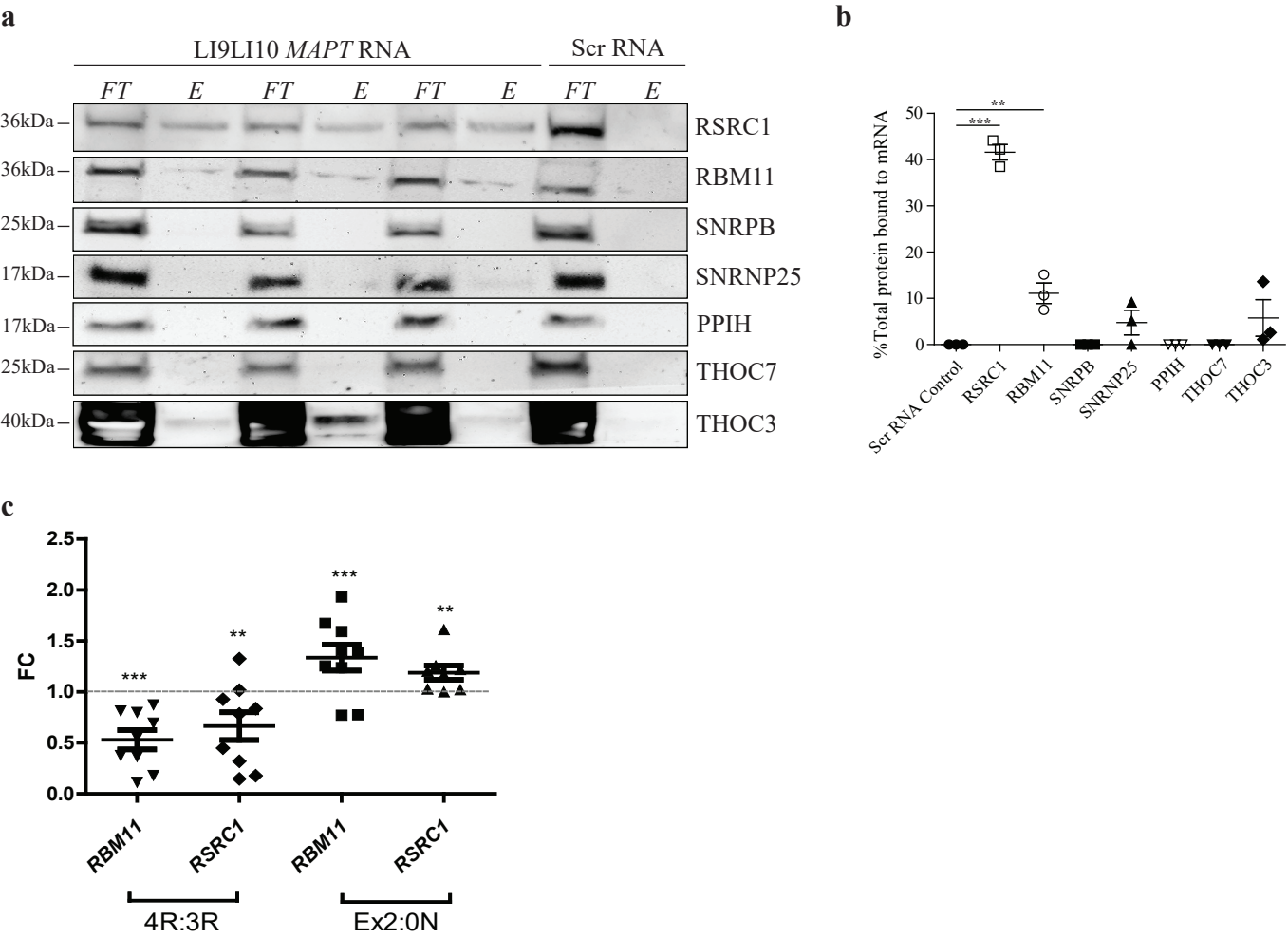
66. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-

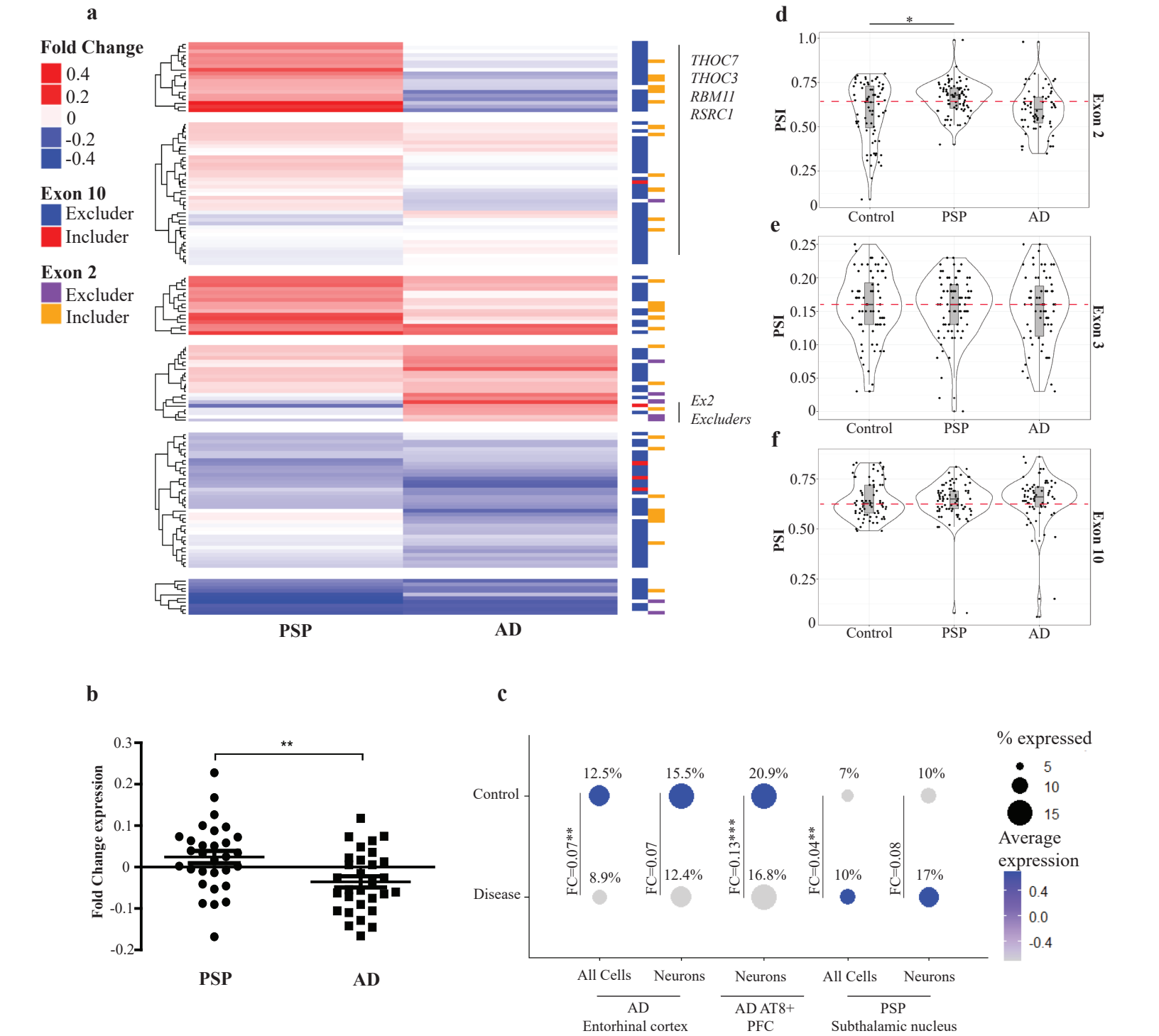
seq data using regularized negative binomial regression. *Genome Biol.* **20**, 1–15 (2019).

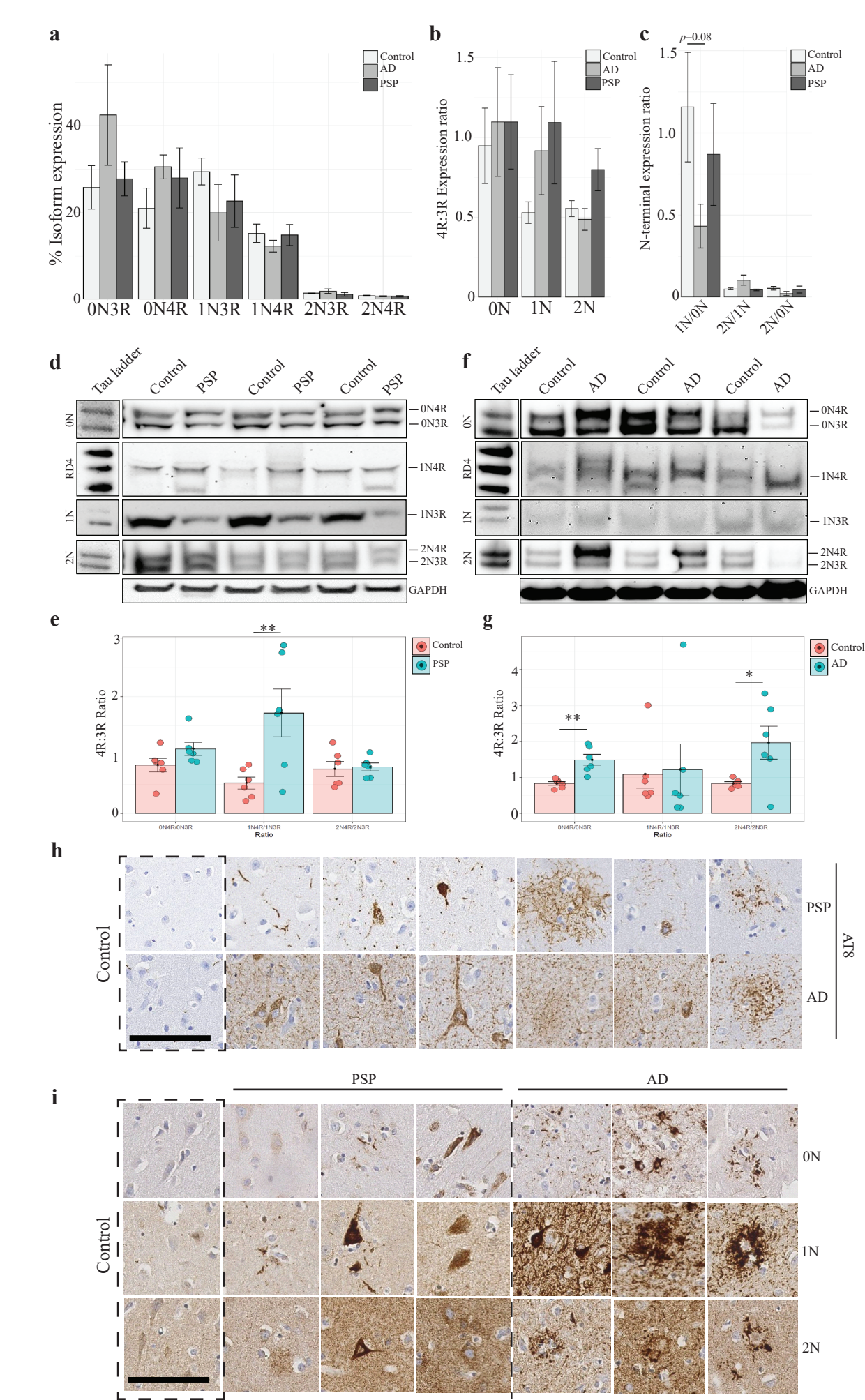
67. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* (2018).

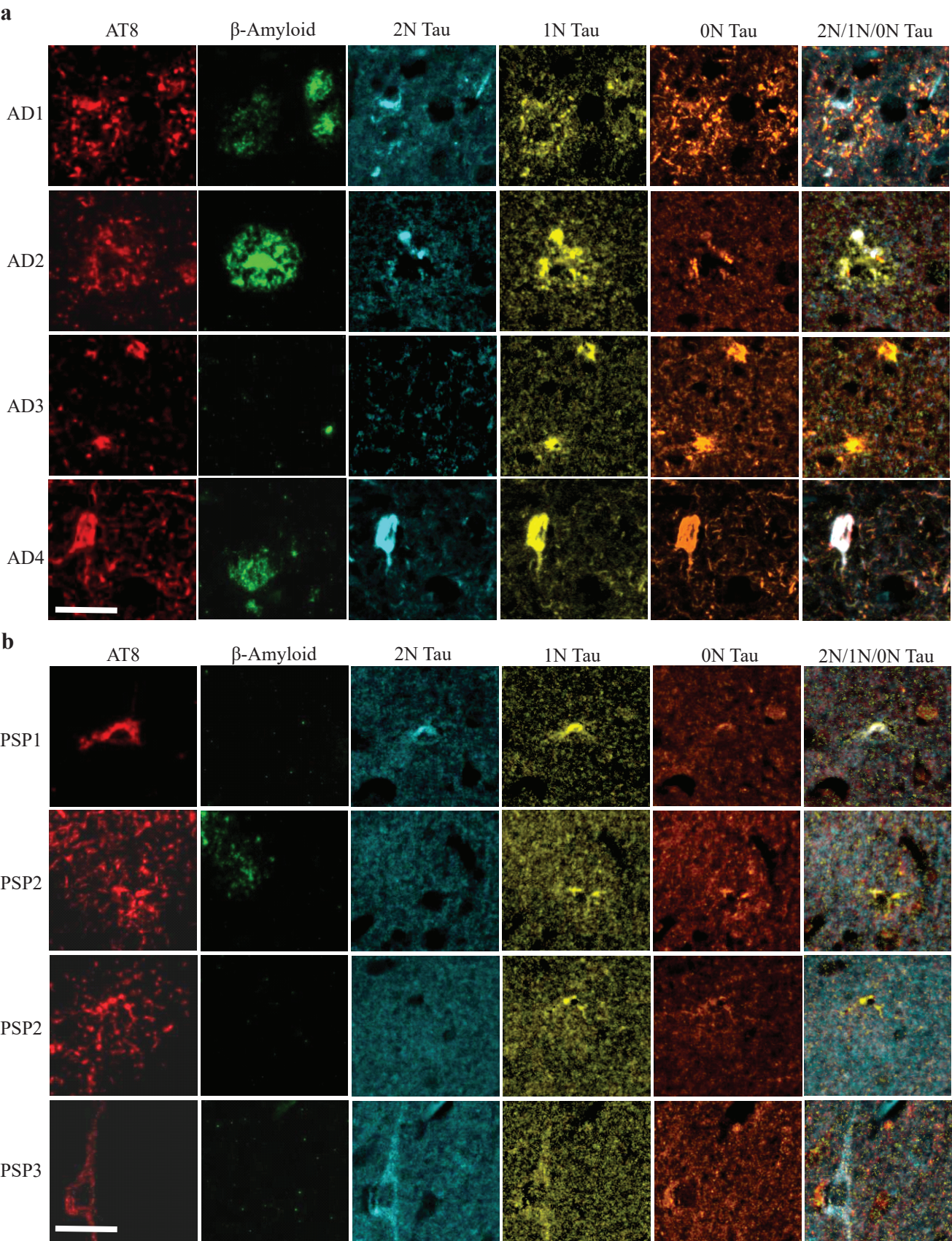
68. Wu, T. D. & Watanabe, C. K. Sequence analysis GMAP : a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).

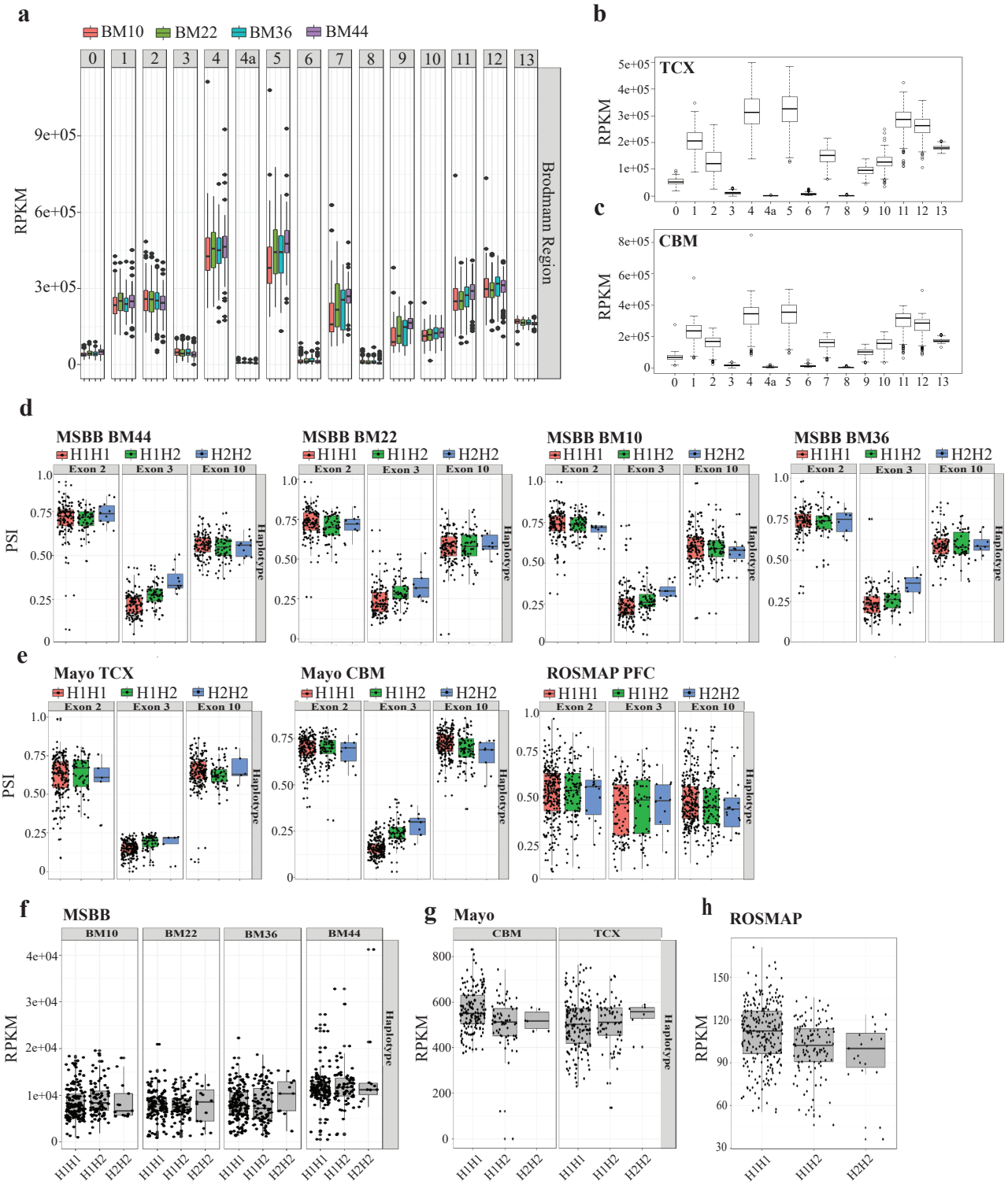

a
1 2 3 4 5 6 7 8 9 10 11 12 13
2N3R
1N3R
0N3R
2N4R
1N4R
0N4R
N-Terminus
Microtubule Binding Repeat Domain
b
RPKM
4e+05
3e+05
2e+05
1e+05
0
0 1 2 3 4 4a 5 6 7 8 9 10 11 12 13
c
BM10 BM22 BM36 BM44
Exon 2 Exon 3 Exon 10
PSI
Brodmann Area
d
Cerebellum Temporal cortex
Exon 2 Exon 3 Exon 10
PSI
Brain Region
***
e
H1H1 H1H2 H2H2
Exon 2 Exon 3 Exon 10
PSI
MSBB Haplotype
***
f
H1H1 H1H2 H2H2
Exon 2 Exon 3 Exon 10
PSI
Mayo Haplotype
***
g
count
400
200
0
-0.4 -0.2 0 0.2 0.4
Value
RM
BM22
BM10
BM36
BM44
Exon 2
RM
BM44
BM36
BM10
BM22
Exon 10

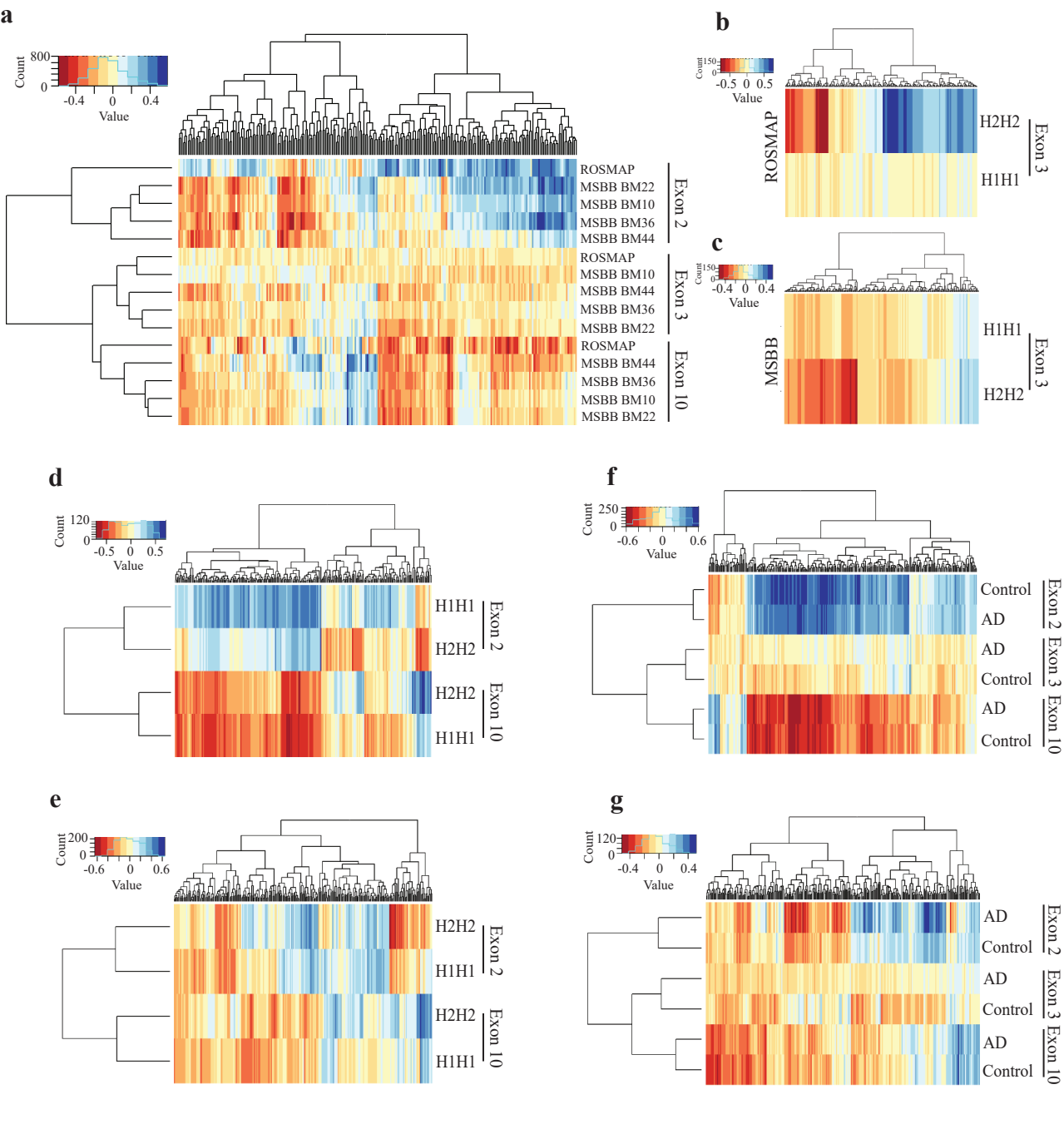

a
LI9LI10 MAPT RNA
Scr RNA
FT
E
FT
E
FT
E
FT
E
36kDa
36kDa
25kDa
17kDa
17kDa
25kDa
40kDa
RSRC1
RBM11
SNRPB
SNRNP25
PPIH
THOC7
THOC3
b
%Total protein bound to mRNA
50
40
30
20
10
0
**
***
Scr RNA Control
RSRC1
RBM11
SNRPB
SNRNP25
PPIH
THOC7
THOC3
c
FC
2.5
2.0
1.5
1.0
0.5
0.0
***
**
***
**
RBM11
RSRC1
RBM11
RSRC1
4R:3R
Ex2:0N

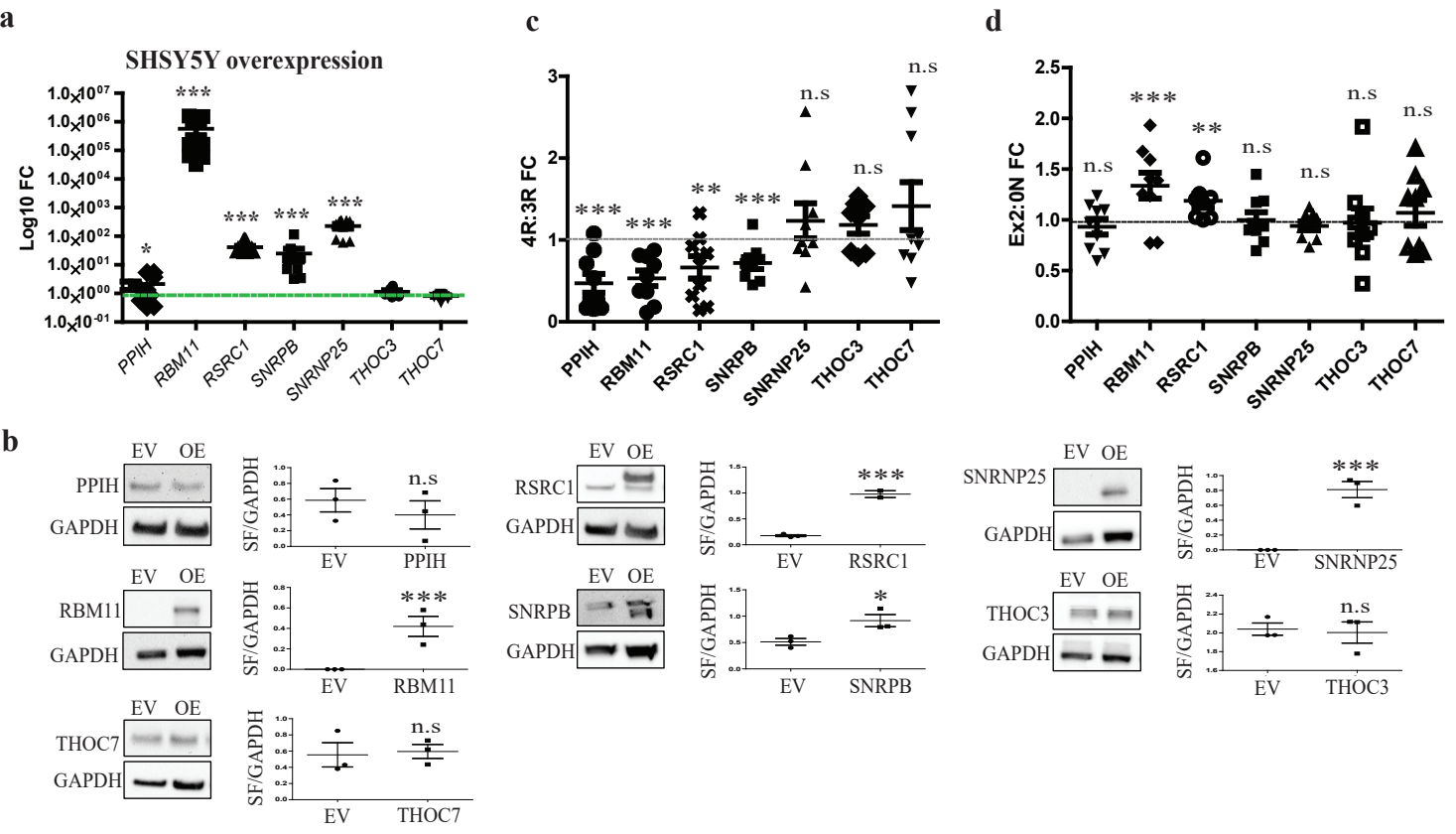

a
Fold Change
0.4
0.2
0
-0.2
-0.4
Exon 10
Excluder
Includer
Exon 2
Excluder
Includer
THOC7
THOC3
RBM11
RSRC1
Ex2
Excluders
PSP
AD
d
*
PSI
Exon 2
Control
PSP
AD
e
PSI
Exon 3
Control
PSP
AD
f
PSI
Exon 10
Control
PSP
AD
b
**
Fold Change expression
PSP
AD
c
Control
Disease
12.5%
15.5%
20.9%
7%
10%
8.9%
12.4%
16.8%
10%
17%
FC=0.07**
FC=0.07
FC=0.13***
FC=0.04**
FC=0.08
All Cells
Neurons
Neurons
All Cells
Neurons
AD
Entorhinal cortex
AD AT8+
PFC
PSP
Subthalamic nucleus
% expressed
5
10
15
Average expression
0.4
0.0
-0.4

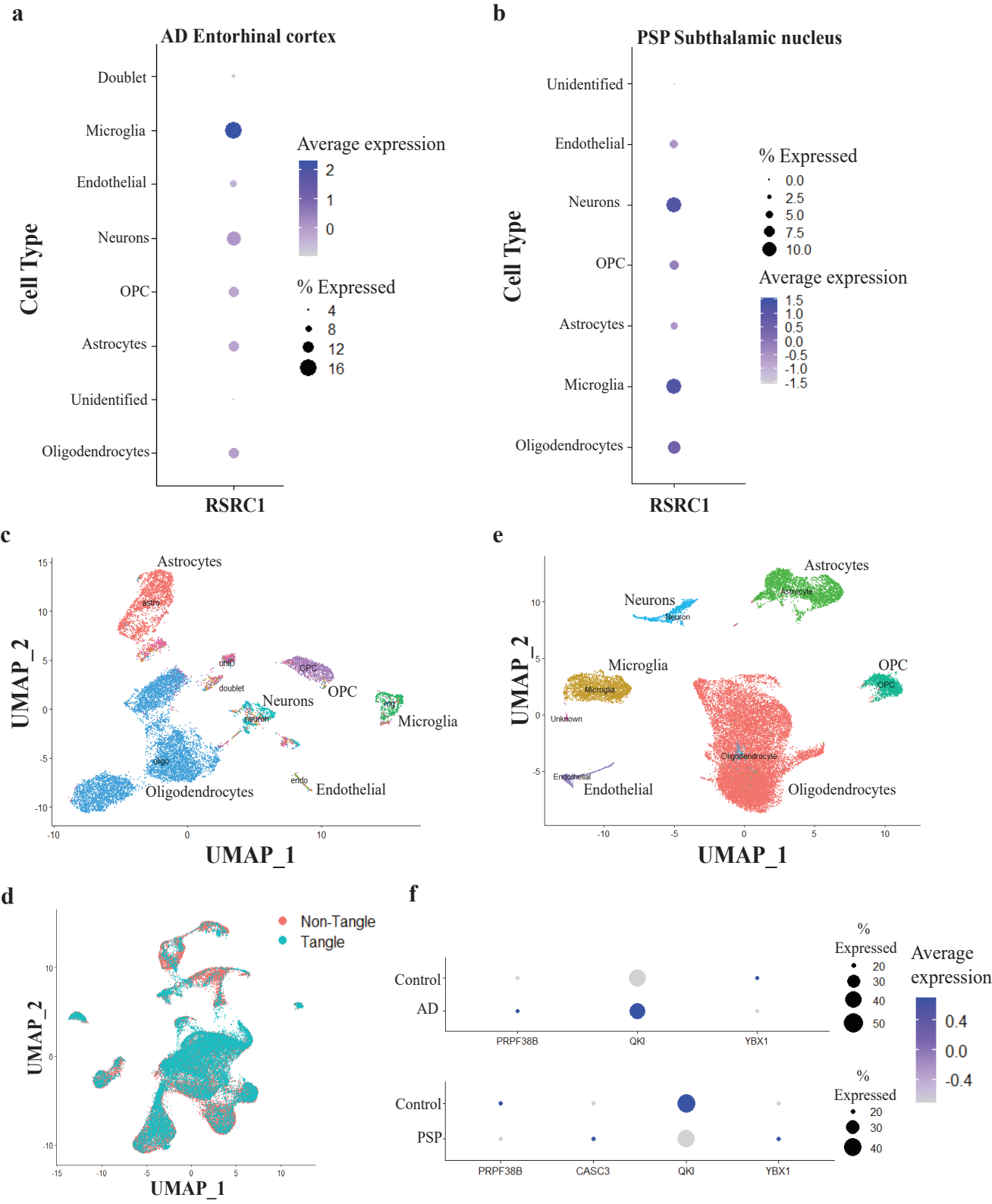

a
% Isoform expression
Control
AD
PSP
0N3R
0N4R
1N3R
1N4R
2N3R
2N4R
b
4R:3R Expression ratio
0N
1N
2N
c
p=0.08
N-terminal expression ratio
1N/0N
2N/1N
2N/0N
d
Tau ladder
Control
PSP
0N
RD4
1N
2N
0N4R
0N3R
1N4R
1N3R
2N4R
2N3R
GAPDH
f
Tau ladder
Control
AD
e
4R:3R Ratio
0N4R/0N3R
1N4R/1N3R
2N4R/2N3R
Ratio
g
h
Control
PSP
AD
AT8
i
PSP
AD
Control
0N
1N
2N

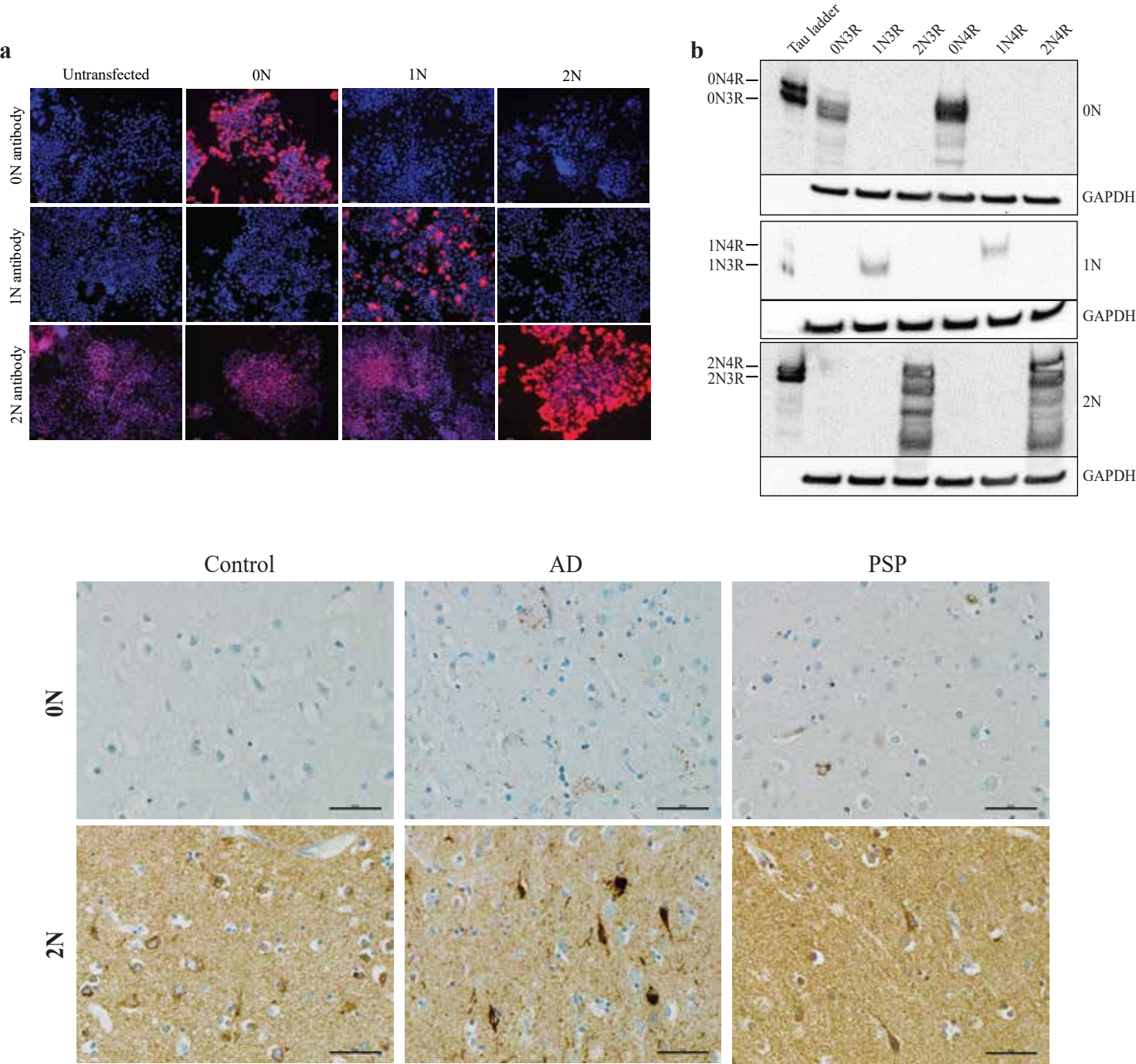

a
AT8
β-Amyloid
2N Tau
1N Tau
0N Tau
2N/1N/0N Tau
AD1
AD2
AD3
AD4
b
AT8
β-Amyloid
2N Tau
1N Tau
0N Tau
2N/1N/0N Tau
PSP1
PSP2
PSP2
PSP3

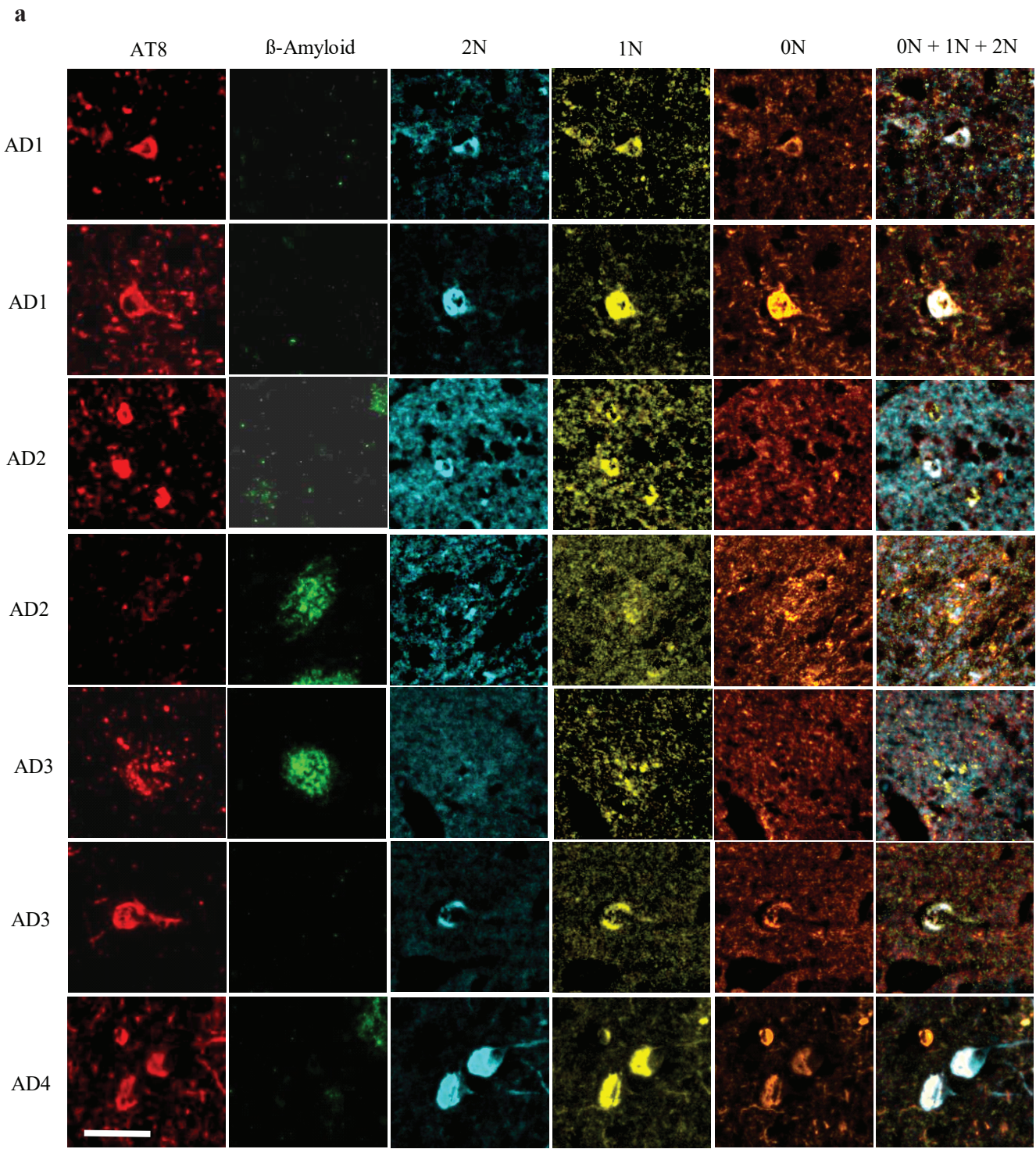

a
BM10
BM22
BM36
BM44
0
1
2
3
4
4a
5
6
7
8
9
10
11
12
13
Brodmann Region
RPKM
9e+05
6e+05
3e+05
0
b
TCX
RPKM
5e+05
4e+05
3e+05
2e+05
1e+05
0
c
CBM
RPKM
8e+05
6e+05
4e+05
2e+05
0
d
MSBB BM44
MSBB BM22
MSBB BM10
MSBB BM36
H1H1
H1H2
H2H2
Exon 2
Exon 3
Exon 10
PSI
Haplotype
e
Mayo TCX
Mayo CBM
ROSMAP PFC
f
MSBB
BM10
BM22
BM36
BM44
RPKM
4e+04
3e+04
2e+04
1e+04
0
g
Mayo
CBM
TCX
800
600
400
200
h
ROSMAP
150
120
90
60
30

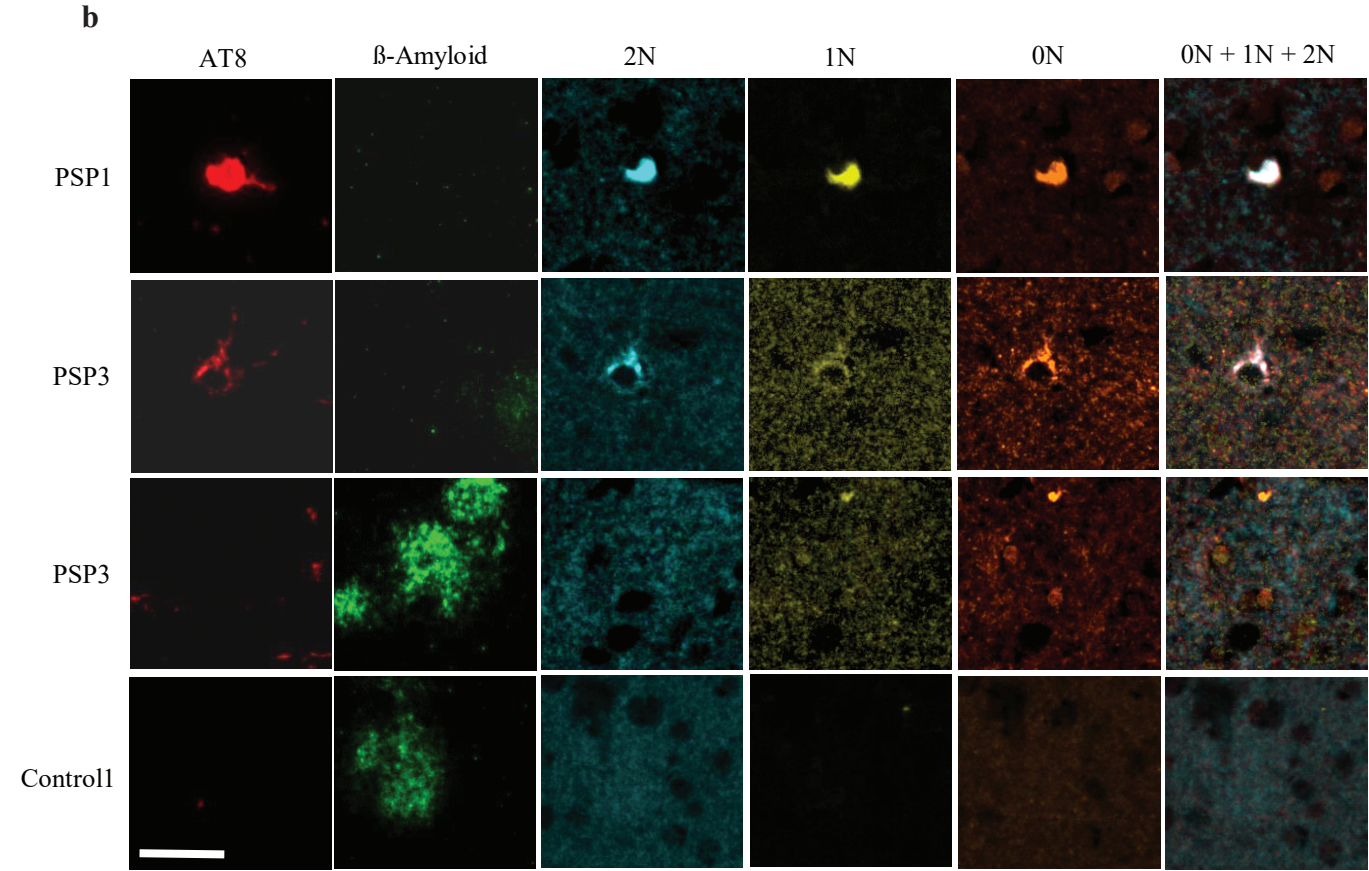

a
Count
800
0
-0.4
0
0.4
Value
ROSMAP
MSBB BM22
MSBB BM10
MSBB BM36
MSBB BM44
Exon 2
ROSMAP
MSBB BM10
MSBB BM44
MSBB BM36
MSBB BM22
Exon 3
ROSMAP
MSBB BM44
MSBB BM36
MSBB BM10
MSBB BM22
Exon 10
b
Count
150
0
-0.5
0
0.5
Value
ROSMAP
H2H2
H1H1
Exon 3
c
Count
150
0
-0.4
0
0.4
Value
MSBB
H1H1
H2H2
Exon 3
d
Count
120
0
-0.5
0
0.5
Value
H1H1
H2H2
Exon 2
H2H2
H1H1
Exon 10
e
Count
200
0
-0.6
0
0.6
Value
H2H2
H1H1
Exon 2
H2H2
H1H1
Exon 10
f
Count
250
0
-0.6
0
0.6
Value
Control
AD
AD
Control
AD
Control
Exon 2
Exon 3
Exon 10
g
Count
120
0
-0.4
0
0.4
Value
AD
Control
AD
Control
AD
Control
Exon 2
Exon 3
Exon 10


a
SHSY5Y overexpression
Log10 FC
PPIH
RBM11
RSRC1
SNRPB
SNRNP25
THOC3
THOC7
c
4R:3R FC
n.s
d
Ex2:0N FC
b
EV
OE
PPIH
GAPDH
SF/GAPDH
RBM11
RSRC1
SNRPB
THOC7
SNRNP25
THOC3


a
AD Entorhinal cortex
Doublet
Microglia
Endothelial
Neurons
OPC
Astrocytes
Unidentified
Oligodendrocytes
Cell Type
RSRC1
Average expression
2
1
0
% Expressed
4
8
12
16
b
PSP Subthalamic nucleus
Unidentified
Endothelial
Neurons
OPC
Astrocytes
Microglia
Oligodendrocytes
Cell Type
RSRC1
% Expressed
0.0
2.5
5.0
7.5
10.0
Average expression
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
c
Astrocytes
astro
unID
doublet
OPC
OPC
Neurons
neuron
Microglia
oligo
Oligodendrocytes
endo
Endothelial
UMAP_2
UMAP_1
e
Astrocytes
Neurons
Microglia
OPC
Unknown
Oligodendrocyte
Endothelial
Endothelial
Oligodendrocytes
UMAP_2
UMAP_1
d
Non-Tangle
Tangle
UMAP_2
UMAP_1
f
Control
AD
PRPF38B
QKI
YBX1
Control
PSP
PRPF38B
CASC3
QKI
YBX1
% Expressed
20
30
40
50
Average expression
0.4
0.0
-0.4
% Expressed
20
30
40


a
Untransfected
0N
1N
2N
0N antibody
1N antibody
2N antibody
b
Tau ladder
0N3R
1N3R
2N3R
0N4R
1N4R
2N4R
0N4R
0N3R
0N
GAPDH
1N4R
1N3R
1N
GAPDH
2N4R
2N3R
2N
GAPDH
Control
AD
PSP
0N
2N


a
AT8
ß-Amyloid
2N
1N
0N
0N + 1N + 2N
AD1
AD1
AD2
AD2
AD3
AD3
AD4


b
AT8
ß-Amyloid
2N
1N
0N
0N + 1N + 2N
PSP1
PSP3
PSP3
Control1