MECHANISTIC UNDERSTANDING OF TAU ALTERNATIVE SPLICING IN NEURONS USING PROTEOMICS

MECHANISTIC UNDERSTANDING OF TAU ALTERNATIVE SPLICING IN NEURONS USING PROTEOMICS

BY

SANSI XING, M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of Requirement for the Degree of Doctor of Philosophy

Department of Biochemistry and Biomedical Sciences McMaster University

Descriptive note

Bachelor of Science, Biological Sciences (2013), Anhui Normal University, Wuhu, China.

Master of Science, Biological Engineering (2016), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.

Doctor of Philosophy, Biochemistry (2021), McMaster University, Hamilton, Ontario, Canada.

TITLE: Mechanistic Understanding of Tau Alternative Splicing in Neurons Using Proteomics

AUTHOR: Sansi Xing

SUPERVISOR: Dr. Yu Lu

NUMBER OF PAGES: xv, 151

Lay abstract

Neurodegenerative disorder is one of the most common diseases among the elderly population and is predicted to become the second leading cause of death with the ever-growing aging society. Abnormal aggregates of a protein named Tau in nerve cells are lethal in a subset of neurodegenerative diseases that are collectively named as tauopathies. In human adult brain, there are six Tau variants that are generated. It is believed that imbalanced expression of these Tau variants can trigger toxic Tau aggregate formation, yet the factors contributing to this process remain unclear. By using a mass spectrometry-based biochemical approach, we identified many upstream factors that are involved in regulating Tau variant balancing. Through in-depth analysis of one of these factors, hnRNPC, we revealed its novel role in controlling Tau variant distribution. HnRNPC may serve as a potential therapeutic target for the treatment of Tau-related neurodegenerative disorders. Separately, by using another mass spectrometry-based biochemical approach, we were able to understand the protein-protein interaction networks for individual Tau variants, which lays the foundation to understand how abnormal imbalance of Tau variants could lead to the pathological Tau aggregation and brain dysfunction. Overall, this work may help to develop new therapeutic strategies for the treatment of Tau-related neurodegenerative disorders.

Abstract

Tauopathies refer to a group of neurodegenerative diseases that are characterized by pathological aggregations of the microtubule-associated protein Tau (MAPT). Aberrant alternative splicing of Tau exon 10 leads to the imbalanced expression of Tau isoforms that contain either 3 or 4 microtubule binding repeats (3R-Tau or 4R-Tau) and this is sufficient to cause the formation of Tau inclusions. Nonetheless, the exact molecular mechanisms that regulate aberrant Tau exon 10 splicing regulation and subsequent Tau aggregation in tauopathies remain elusive. In my thesis research, I used RNA Antisense Purification by Mass Spectrometry (RAP-MS) to identify upstream regulators of Tau splicing events. Among the 15 identified novel protein candidates, I validated that hnRNPA2B1 and hnRNPC are required to promote 4R-Tau expression, whereas hnRNPH1 supports 3R-Tau expression. Separately, to elucidate the functional difference between 3R- and 4R-Tau isoforms, I performed proximity-dependent biotin identification (BioID2) for all six human central nervous system Tau isoforms in mouse primary neurons. Followed by tandem mass tag (TMT)-labeling proteomics and data analysis, I observed that 4R-Tau proximal proteins are highly enriched in endocytosis, whereas 3R-Tau proximal proteins show top enrichment in fatty acid metabolism. Through further biochemical validations, I found that MAT2A, a S-adenosylmethionine synthase, has higher binding affinity with 3R-Tau versus 4R-Tau. Overall, using novel proteomics methods, I discovered novel Tau splicing regulators and characterized the neuronal Tau isoform-specific proximity proteome networks. These proteins, once validated through future functional studies in cellular and animal models, can represent therapeutic and diagnostic targets for neurodegenerative tauopathies.

Acknowledgements

First, I would like to thank my supervisor Dr. Yu Lu, thank you for providing me with a chance to study in Canada. Thank you for being so supportive of my projects. Yu is an excellent professor who is always passionate about science. He is open-minded. Every time when I come up with a new idea, he is always happy to discuss with me and support me. Yu is also very patient. Whenever I reach out to him with a question, he would spend the time to explain to me until I fully understand the concept. To me, Yu is more like a friend than a professor. Thank you for all your help during my graduate study. I am so proud of being one of your graduate students and thank you for training me so well.

I would like to thank all of my committee members, Brad, Xu-dong, Karun. Thank you very much for guiding my projects in the right directions. Brad, I have learned a lot from you both within and outside of the academic setting. You are so kind and knowledgeable. I am inspired by your rigorous attitude towards science. Thank you for your positive comments on my committee meeting reports, which have brought me with a lot of encouragement. Brad, I am so lucky to meet you and have you as my committee member. Xu-dong, thank you for your support of my projects. You are the one who always provide me with practical suggestions. Without your kind help, my projects wouldn't have been progressing so smoothly. I am so appreciated to have you as my committee member. Karun, thank you for being my committee member and for helping me with my projects. Every time when I meet with you to discuss about my projects, you always provide me with detailed and reasonable suggestions. I would also want to thank our collaborators Marco

and John, thank you for all your support with my projects. Special thanks to Nadeem, thank you for providing me with the cultured primary neurons for my project.

I want to thank all of the Lu lab members and for all of the support, collaborations and company during my graduate study.

I want to thank my parents, thank you for supporting me to study in Canada. Thank you for your endless encouragement and love, all of which has become a big motivation for me to overcome the obstacles in scientific research.

Table of contents

Descriptive note ii
Lay abstract iii
Abstract iv
Acknowledgements v
List of figures and tables ix
Abbreviations xii
Declaration of Academic Achievement xv
Chapter 1: Introduction 1
1.1 Tau and Tauopathies 1
1.2 Non-standard functions of Tau 3
1.2.1 Functions of Tau in synapses 3
1.2.2 Functions of Tau in plasma membrane 4
1.2.3 Functions of Tau in the nucleus 6
1.3 Introduction to alternative splicing 7
1.4 Tau splicing regulation in physiology and pathology 10
1.4.1 Tau splicing regulation in the nervous system 10
1.4.2 Aberrant Tau splicing in tauopathy 11
1.4.3 Known factors contributing to Tau exon 10 splicing regulation 12
1.5 Earlier studies about Tau isoform specific functions 15
1.6 Mouse models for tauopathies 17
1.7 Tau-targeted therapeutic approaches for treatment of tauopathies 19
1.8 Efforts to pull down RNA and its associated RBPs 24
1.9 Enzyme-mediated proximity labeling techniques for *in vivo* interactome profiling 27
1.10 Mass Spectrometry Based Quantitative Proteomics 30
1.10.1 Label-free 31
1.10.2 Metabolic labeling-SILAC 32
1.10.3 Isobaric labeling-TMT 33
1.11 Summary of intent 35

Chapter 2: Identification of Novel Tau Exon 10 Splicing Regulators Using RNA Antisense Purification Proteomics 38
2.1 Abstract 38
2.2 Introduction 38
2.3 Methods 42
2.4 Results 50
2.5 Discussion 57
2.6 Main figures 62
2.7 Supplementary figures and tables 74
Chapter 3: Investigation of Tau Isoform-Specific Interactome in Mouse Primary Cortical Neurons 90
3.1 Abstract 90
3.2 Introduction 91
3.3 Methods 94
3.4 Results 101
3.5 Discussion 107
3.6 Main figures 111
3.7 Supplementary figures and tables 121
Chapter 4 Concluding remarks 131
REFERENCES 134

List of figures and tables

Chapter 1:

Figure 1: Two reaction steps for pre-mRNA splicing……..……8

Figure 2: Workflow of ChIRP–MS and RAP-MS…….....26

Figure 3: Structure of the Tandem Mass Tag……..34

Chapter 2:

Figure 1: Identification of proteins binding to the Tau minigene pre-mRNA using the RAP-MS method……....62

Figure 2: Loss-of-function assays to identify RBPs that control Tau exon 10 splicing….64

Figure 3: hnRNPC interacts with Tau pre-mRNA and promotes Tau exon 10 inclusion……..…66

Figure 4: hnRNPA2B1 interacts with Tau pre-mRNA and promotes Tau exon 10 inclusion…….....68

Figure 5: hnRNPH1 interacts with Tau pre-mRNA and promotes Tau exon 10 exclusion…….…....69

Figure 6: hnRNPC promotes Tau exon 10 inclusion through direct binding to U-rich tracts near Tau exons 10 and 11……..…71

Figure 7: Working model…….…..73

Figure S1: RNA antisense purification (RAP) proteomic method to identify RNA-binding proteins specifically interact with RNA targets……..…74

Figure S2: Detection of 3R- and 4R-Tau isoforms in SH-SY5Y and SK-N-SH neuroblastoma cells……75

Figure S3: UV-crosslink optimization in SH-SY5Y……..75

Figure S4: Expression of the Tau minigene in SH-SY5Y……...…76

Figure S5: Expression of the Tau minigene pre-mRNA in SH-SY5Y stably expressing Tau minigene……..76

Figure S6: Overexpression of hnRNPC promotes Tau exon 10 inclusion in HEK 293FT cells……77

Figure S7: Overexpression of hnRNPA2B1 promotes Tau exon 10 inclusion in HEK 293FT cells……77

Figure S8: The 4R-Tau splicing factors hnRNPC and hnRNPA2B1 are both upregulated in PSP patient brain samples……..78

Figure S9: The 3R-Tau splicing factors HNRNPH1 is downregulated during human brain development……...…78

Figure S10: Binding affinity of hnRNPC with E11-1 was disrupted after mutation of the overlapped U-tracts shown in E11-2……..79

Figure S11: Characterization of interaction between hnRNPA2B1 and Tau pre-mRNA ……80

Figure S12: Enrichment of endogenous Tau pre-mRNA in SH-SY5Y cells…….81

Table S1: ShRNA oligos and Primers used in this study……...……81

Table S2: DNA probes used in RAP-MS……84

Table S3: RNA binding proteins identified by proteomic analysis in Tau probes-based RAP-MS pulldown…….87

Chapter 3：

Figure 1: Expression of Tau BioID2……...….111

Figure 2: Schematic diagram of *in vivo* Tau BioID2 labeling and quantitative proteomics workflow……..……112

Figure 3: Expression of Tau BioID2 in primary mouse neuron at DIV 18…….….113

Figure 4: Proteome-wide quantification of identified proteins in Tau BioID2 ……..…..114

Figure 5: Human Tau isoform-specific interactome network…….…….115

Figure 6: Gene Ontology analysis for total Tau proximal proteins……..116

Figure 7: Functional module analysis for Tau proximal proteins……117

Figure 8: Functional distribution of proteins biotinylated by 3R- and 4R-Tau BioID2 ……..117

Figure 9: Overlap of 4R-Tau proximal proteins……...…118

Figure 10: Overlap of 3R-Tau proximal proteins…….....118

Figure 11: Validation of interaction between MAT2A and Tau isoforms by co-immunoprecipitation……119

Figure 12: Functional distribution of proteins biotinylated by 0N-, 1N- and 2N-Tau BioID2…….….120

Figure S1: TMT total intensity of each channels in three batches of MS experiments ……..….121

Figure S2: Comparison of the Tau proximal proteins identified in this Tau BioID2 system with one previously reported Tau interactome that identified by AP-MS in mouse brain

……..121

Figure S3: Validation of interaction between GLUL and Tau isoforms by co-immunoprecipitation……...….122

Table S1: Tau BioID2 samples and assigned TMT channels……..122

Table S2: Significant protein hits identified from Tau isoform-BioID2 experiments….123

Abbreviations

5'ss	5' splice site
ACE	A/C-rich enhancer
AD	Alzheimer's disease
AGD	Argyrophilic grain disease
ALS	Amyotrophic lateral sclerosis
AMPA	α-amino-3-hydroxy-5-methylisoxazole-4-propionic acid
AP	Affinity purification
APEX	Engineered ascorbate peroxidase
AP-MS	Affinity precipitation combined with mass spectrometry
AS	Alternative splicing
ASO	Antisense oligonucleotides
ATP	Adenosine triphosphate
BCCP	Biotin carboxyl carrier protein
bioAMP	Biotinoyl-5-AMP
BioID	Proximity dependent biotin identification
CBD	Corticobasal degeneration
CDK5	Cyclin Dependent Kinase 5
ChIRP-MS	Comprehensive identification of RNA binding proteins by mass spectrometry
CNS	Central nervous system
Co-IP	Co-immunoprecipitation
CRISPR	Clustered-regularly interspaced short palindromic repeats
CSF	Cerebrospinal fluid
DAB	Diaminobenzidine
DNA	Deoxyribonucleic acid
DSB	Double strand DNA breaks
Dyrk1A	Dual-specificity tyrosine-phosphorylated and regulated kinase 1A
ESE	Exon splicing enhancers
ESS	Exon splicing silencers
FBS	Fetal bovine serum
FTD	Frontotemporal dementia
FTDP-17	Frontotemporal dementia with parkinsonism linked to chromosome 17
FTLD	Frontotemporal lobar degeneration
FUS	Fused in sarcoma
GluR 2/3	Glutamate receptor 2/3
GSK-3β	Glycogen Synthase Kinase 3 Beta
HCD	Higher energy collision induced dissociation
hnRNPs	Heterogeneous nuclear RNPs

IgG	Immunoglobulin G
IR	Inter-repeat
ISE	Intron splicing enhancers
ISM	Intron splicing modulator
ISS	Intron splicing silencers
Keap1	Kelch-like ECH-associated protein-1
LC-MS	Liquid chromatography–mass spectrometry
lncRNA	Long non-coding RNA
MAP	Microtubule associated protein
MCI	Mild cognitive impairment
mRNA	messenger RNA
NAD	Davunetide
NFT	Neurofibrillary tangles
NLS	Nuclear localization sequence
NMR	Nuclear magnetic resonance
nt	Nucleotide
PAC	P1-derived artificial chromosome
PART	Primary age-related tauopathy
PCR	Polymerase chain reaction
PiD	Pick’s disease
PL	Proximity labeling
PNS	Peripheral nervous system
PP2A	PP2A
PPE	Polypurine enhancer
PPI	Protein-protein interaction
pre-mRNA	Precursor mRNA
PROTAC	Proteolysis-targeting chimaeras
PSP	Progressive supranuclear palsy
PTM	Post transcriptional modification
PUP-IT	Pupylation-based interaction tagging
RAP-MS	RNA Antisense Purification by Mass Spectrometry
RBPs	RNA binding proteins
RNA	Ribonucleic acid
RNase H1	Ribonuclease H1
SAM	S-adenosylmethionine
SF1	Splicing factor1
SF2	Serine/arginine-rich splicing factor 2
shRNA	Short hairpin ribonucleic acid
SILAC	Stable Isotope Labeling with Amino acids in Cell culture

snRNA	Small nuclear RNA
snRNPs	Small nuclear ribonucleoprotein particles
SR	Serine and arginine rich
SSOs	Splice-switching oligonucleotides
Syngr3	Synaptogyrin-3
TDP-43	Transactive response DNA-binding protein of 43 kDa
TFA	Trifluoroacetic acid
TMT	Tandem mass tag
U2AF1	U2 auxiliary factor1
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultraviolet
XIST	X-inactive specific transcript

Declaration of Academic Achievement

The experimental design, execution of all research studies described in this thesis, data analysis, data interpretation, and thesis writing were mainly completed by Sansi Xing. Dr. Yu Lu conceived and supervised all research described in this thesis. Dr. John Crary and Dr. Marco Hefti provided guidance and human postmortem brain samples for the research described in Chapter 2 and edited the manuscript included in Chapter 2. Dr. Karun Singh provided guidance for the research described in Chapter 3. Dr. Marco Hefti contributed to the RNA sequencing data of human fetal and adult postmortem brain samples. Jane Wang from Dr. Yu Lu's lab contributed to the UV cross-linking optimization in the research described in Chapter 2. Nadeem Murtaza from Dr. Karun Singh's lab contributed to the culturing of mouse primary neurons that were used in the research described in Chapter 3.

Chapter 1: Introduction

1.1 Tau and Tauopathies

Microtubule-associated protein Tau (MAPT) was first isolated from the microtubule fraction of porcine brain as a heat-stable protein by Weingarten et al. in 1975[1]. As a member of the microtubule associated protein (MAP) family, Tau is preferentially expressed in neurons and has low expression levels in astrocytes and glial cells[2]. Tau protein is considered as an intrinsically disordered protein with three characterized domains. The acidic N-terminal region, or projection domain, projects away from microtubules to interact with the plasma membrane and is involved in regulating microtubule dynamics and spacing[3]; the proline-rich middle domain functions to bind with both actin and SH-3 containing proteins[4, 5]. The C-terminal half contains four imperfect microtubule binding repeats that are separated by flanking regions, which can bind with the microtubule and regulate microtubule polymerization to maintain its stabilization[6].

In late 1980s, hyperphosphorylated Tau was discovered by several groups as the main component of neurofibrillary tangles (NFT), one of the major histopathological hallmarks in Alzheimer's disease (AD)[7-10]. Since then, numerous studies focused on unveiling the mechanisms that contributed to Tau hyperphosphorylation and aggregation. Tau inclusions were also identified in other neurodegenerative diseases, including progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Pick's disease (PiD), frontotemporal dementia (FTD), argyrophilic grain disease (AGD), primary age-related Tauopathy (PART), and other central and peripheral nervous system diseases[11-15]. These diseases are collectively classified as tauopathies.

The exact molecular cause of Tau aggregates is rather complex, as multiple factors contribute to the pathological process of tauopathies. Among these factors, Tau hyperphosphorylation is significantly associated with Tau aggregate formation[16-21]. Tau is a phosphorylated protein that contains 85 potential phosphorylation sites[22]. In human normal adult brain, each Tau molecule carries 2-3 phosphate groups, whereas in AD brain, there is, minimally, a 3-fold increase in Tau phosphorylation[23]. To date, over 45 phosphorylated sites have been revealed in the insoluble Tau filament that was isolated from AD brain[24]. Many kinases have been discovered to be involved in Tau phosphorylation, among which Glycogen Synthase Kinase 3 Beta (GSK-3β) and Cyclin Dependent Kinase 5 (CDK5) were reported to phosphorylate AD - associated sites[25, 26]. Conversely, protein phosphatase-2A (PP2A) is the major phosphatase for Tau dephosphorylation, and it has been reported that the catalytical activity of PP2A was reduced in AD brain[27, 28]. Tau can be normally hyperphosphorylated during normal brain development and hibernation[29, 30], but hyperphosphorylated Tau causes Tau aggregation and formation of neurofibrillary tangles that exhibit a detrimental role in neuronal activity. It has been reported that hyperphosphorylated Tau interacts with normal Tau, which disrupts microtubule assembly and stabilization[31]. The dysregulation of microtubules could further inhibit the intracellular trafficking[32]. Moreover, several studies have demonstrated that Tau hyperphosphorylation levels are correlated with cognitive impairment and memory deficits in transgenic mouse models that carry Tau genetic mutations[33, 34]. Another study has reported that hyperphosphorylated Tau inhibited proteasomal activity, which further prevented the clearance of the Tau aggregates[35]. However, further studies are

needed to confirm this hypothesis. Besides phosphorylation, Tau protein can also be regulated by other modifications, such as acetylation, truncation, nitration, ubiquitination and SUMOylation. These post translational modifications have also reported to be associated with Tau aggregation[36].

In addition, genetic mutation-mediated Tau expression dysregulation has also been linked to neurofibrillary degeneration, although it has received relatively less attention[37, 38]. The MAPT gene locus is exceptionally complex, being situated within an extended genomic locus characterized by a large inversion, gene duplications, numerous repeat sequences and rearrangements that have emerged during human evolution[38-41]. At the RNA level, reports in the literature, alongside public sequence databases, confirm that MAPT has an intricate splicing pattern. Alternative polyadenylation gives rise to additional structural complexity[38]. Furthermore, there are alternative promoters that may remove the first exon[38]. To date, the molecular mechanisms underpin Tau pathology remain not fully understood. Continuous efforts are still needed to fill this gap so that more and more therapeutic targets can be uncovered for tauopathy treatment.

1.2 Non-standard functions of Tau

Even though Tau, as a microtubule binding protein, is predominantly localized in axons of mature neurons[42, 43], Tau has also been observed in other cellular compartments such as synapse; plasma membrane and nuclei over the last decades, which indicates that Tau plays intricate roles in neuronal activity regulation[44-49].

1.2.1 Functions of Tau in synapses

Synaptic loss is considered as an early event for neuron dysfunction, and it has been reported that it is correlated with cognitive decline in AD and other tauopathies[50-52]. Several studies have reported synaptic localization of Tau in both normal and AD brains [53, 54]. A study from Chen et al. has shown that knock down of Tau significantly reduces the dendritic spine numbers in cultured hippocampal neurons [55], which implies that Tau plays a role in regulating synaptic integrity in physiological conditions. Under pathological conditions, mis-localization of disease-related Tau into dendrites is an early pathological feature prior to Tau accumulation in tauopathies. In 2010, a study from Hoover et al. described the preferential dendritic localization of Tau in rat neurons transfected with tauopathy related Tau mutant P301L compared to WT Tau[56]. Moreover, another study also observed that Tau was enriched in dendritic spines in P301L transgenic mouse models[57]. However, the molecular mechanism underlying the mis-localization of pathological Tau into dendrites and synaptic loss remains elusive. In 2018, a study from Verstreken's lab identified Synaptogyrin-3 (Syngr3), a transmembrane vesicle protein, as a novel Tau interactor[54]. Most importantly, they found that knock down of Syngr3 reduced pathological Tau-vesicles associations, which partially rescued synaptic functions, including increased vesicle mobility and neurotransmitter release[54]. Further research is still needed to fully understand synaptic Tau functions under physiological and pathological conditions in order to identify therapeutic targets and prevent disease progression in the early stage of tauopathies.

1.2.2 Functions of Tau in plasma membrane

Association of Tau with the cellular membrane has been studied for decades. Overexpression of a full length Tau and the N-terminal projection domain of Tau in rat PC12 pheochromocytoma cells demonstrated plasma membrane localization of Tau and the N-terminal domain was necessary for its membrane association [45]. Another study found that tubulin wasn't identified in plasma membrane fractions, hence they hypothesized that the interactions between Tau and cellular membrane might not rely on its microtubule binding function[58]. However, this hypothesis remains controversial. A study from Eliezer's lab showed that Tau interacted with the lipid membrane through the short amphipathic helices localized within the microtubule binding domain[59]. Several other studies reported that membrane localization of Tau was dependent on its phosphorylated states as a large amount of dephosphorylated Tau had been observed in membrane, which indicated that in Tauopathies, hyperphosphorylated Tau might lead to neuron dysfunction partially through disrupting the normal functions of Tau in the membrane[60, 61]. Even though numerous studies have indicated the existence of interaction between Tau and plasma membrane and Tau may play a role in cell signaling, the pathological relations between them remain limited. An early study has observed that paired helical filaments (PHF), composed of the truncated form of Tau, were enriched in cytomembrane[62]. Another study, by using a cellular model, found that the plasma membrane localized PHF core fragment could recruit full length native Tau into deposits[63]. Additionally, Elbaum-Garfinkle et al., used K18 (contains the microtubule binding domain of Tau), a Tau fragment that exhibited high aggregation property[64], to evaluate the impact of the lipid bilayer on Tau aggregation. They demonstrated that Tau aggregation happened when the amount of K18 that bounded to the

lipid bilayer reached to a critical surface density[65]. Recently, an *in vitro* study further confirmed the association between full length Tau and the lipid membrane by using X-ray and neutron scattering[66]. In this study, they also demonstrated that this strong interaction could induce Tau structure compaction and membrane disruption[66], which provides additional evidence supporting the role of lipid membrane plays in pathogenesis of Tauopathies.

1.2.3 Functions of Tau in the nucleus

In 1988, Metuzals et al. first discovered the presence of Tau paired helical filaments in the nuclei of brain neurons in AD patients through transmission electron microscopy[67]. Later, Binder and his colleagues observed nuclear Tau in both normal and AD brain[68]. Since Tau lacks the nuclear localization sequence (NLS), research efforts have been made to explain its nuclear localization. It has been considered that post transcriptional modification (PTM) may be involved in this translocation and several studies have confirmed the presence of both unphosphorylated and phosphorylated Tau in the nucleus[69-72]. In addition, nuclear Tau localization may also happen at the transcript level. Wang et al. found that a 2kb Tau transcript, which encodes for the full length Tau protein with differential polyadenylation sites usage compared with the classic 6 kb Tau transcripts in the central nervous system (CNS), was enriched in nucleus[73].

Since then, scientists started to explore the function of nuclear Tau in both physiological and pathological conditions. In 2011, Sultan et al. reported that the amount of nuclear Tau increased during acute oxidative and heat stress, which indicated that Tau might have a protective role in maintaining nuclear integrity under cellular stress[74]. Further

studies showed that Tau was capable of binding to both single and double stranded DNA[75]. It has also been reported that the proline rich domain and the second microtubule binding repeats of Tau were responsible for its interaction with DNA, based on nuclear magnetic resonance (NMR) data[76]. This interaction was facilitated by the minor groove structure of DNA with electrostatic and hydrophobic interactions[77, 78]. Most interestingly, the binding of Tau with DNA could cause DNA conformational changes [78], which indicated that Tau might function to protect DNA from damage. Indeed, loss of Tau proteins in mouse primary neurons induced a high rate of DNA damage upon celluar stress[79]. Besides the functional role that nuclear Tau plays in genome integrity and DNA repair, Tau is also reported to promote chromosome stability[44, 80] and to participate in ribosomal assembly, ribosome gene transcription and to control gene expression with the identification of Tau in the nucleolus [81-84]. Overall, the pathological role of nuclear Tau is still unclear, but it has been speculated that hyperphosphorylated Tau may disrupt normal Tau functions in the nucleus, which may further induce neuronal toxicity and neuronal death during the progression of tauopathy.

1.3 Introduction to alternative splicing

Genetic information is initially stored in the DNA sequence, and is then transcribed to messenger RNAs (mRNA) , which can be used as a template for translating functional proteins to support celluar activity[85]. In eukaryotes, after transcription, the nascent RNA transcripts not only contains the protein coding sequences (exons) but also the intervening noncoding sequences (introns)[86]. During the maturation of mRNA, introns are spliced out and exons are joined together. This process is termed as RNA splicing[87].

Splicing reaction is a complicated process that is catalyzed by the spliceosome, a dynamic megadalton machine, which is composed of five small nuclear ribonucleoprotein particles (snRNPs) (U1, U2, U4, U5 and U6 snRNPs) and hundreds of auxiliary proteins[88]. Spliceosome assembly on nascent RNA or precursor mRNA (pre-mRNA) requires the recognition of a 5'splice site GU sequence by the U1 snRNP, the binding of splicing factor 1 (SF1) to the branch site, U2 auxiliary factor 1 (U2AF1) to the polypyrimidine tract and U2AF2 to a 3' AG sequence[89]. The replacement of SF1 with U2 snRNP and the recruitment of U4/U5-U6 snRNPs result in conformational changes and remodeling of the spliceosome to initiate the following chemical reactions[89]. First, the 2' hydroxyl group of an adenosine located at the branch site attacks the 5' splice site, forming a lariat structure, which is accompanied with the release of the 5' exon[86]. Next, the 3' hydroxyl group of the released exon immediately reacts with the 3' splice site to form the joined exons, whereas the corresponding lariat-shaped intron is cut out and degraded as a byproduct (Figure 1.1)[86].

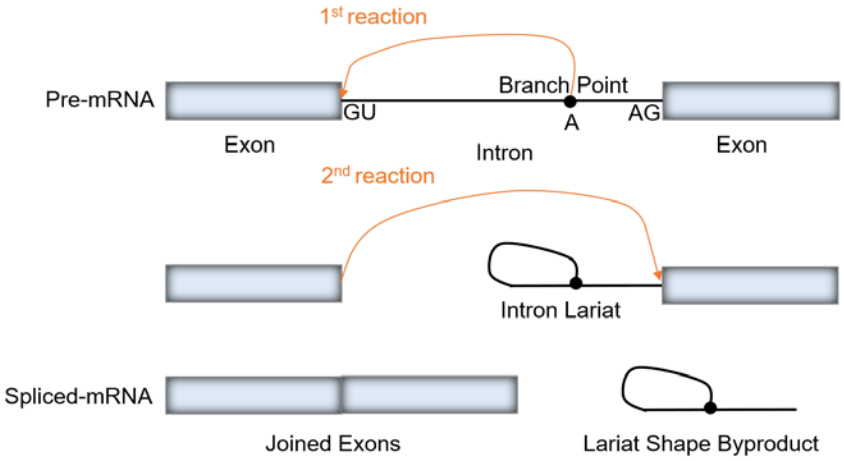


Figure 1: Two reaction steps for pre-mRNA splicing

Intriguingly, besides constitutive splicing, it has been observed that exons can be selectively ligated together in a process called alternative splicing[90]. During this process, pre-mRNA can be processed into one or more transcripts with different combinations of exons and introns, therefore, contributing to the proteomic diversity in eukaryotes[91-93]. The decision of which introns or exons should be included or skipped is determined by two groups of factors – cis-element and trans-acting factors[92, 94]. Cis-elements refer to short RNA sequences located in pre-mRNA, which can be broadly classified into four categories including exon splicing silencers (ESS), exon splicing enhancers (ESE), intron splicing silencers (ISS) and intron splicing enhancers (ISE), depending on their location and function[94]. Trans-acting factors mostly consist of two types of superfamilies – Serine and arginine rich (SR) /SR-like proteins and heterogeneous nuclear RNPs (hnRNPs)[94]. SR proteins functions as members of spliceosome, whereas hnRNPs play a major role in RNA metabolism, such as RNA transport, RNA stability and translational modulation[95-97]. During the process of alternative splicing, trans-acting proteins or RNA binding proteins can act on the regulatory elements to determine the inclusion or exclusion of specific exons/introns through modulating the activity of spliceosomes[98]. Alternative splicing can occur in many ways, such as alternative exon cassette usage, alternative 5’ and 3’ splice sites, mutually exclusive exons and retained introns[99]. It has been reported that more than 90% of human genes will undergo alternative splicing. The resulting proteome diversity is necessary for the precise regulation of cellular funcation.[100]. In addition, previous studies have shown that the diversity of protein variants generated by alternative splicing differs in

different cell types and tissues and alternative splicing can be dynamically regulated during development[101].

It has been reported that alternative splicing is actively regulated in the nervous system to support its functional complexity, such as in axonal transportation, synaptic plasticity, signal transduction and neurotransmitter release[102], whereas abnormal alternative splicing events have been described to correlate with the pathogenesis of neurodegenerative diseases[102, 103]. Aberrant alternative splicing occurs in two ways. One is via genetic mutations located in cis regulatory elements, while the other one is via dysregulation of trans-acting factors[104]. It has been reported that around 50% of the genetic mutations identified in human diseases are involved in altered RNA splicing by either disrupting, enhancing or creating the regulatory elements and modulating the activity of trans-acting factors [105].

Given the important role of aberrant splicing events in human diseases, increasing effort has been made to decipher the molecular mechanisms that contribute to altered splicing events in disease. To date, therapeutics have been designed to target aberrant alternative splicing by either blocking the splicing events with splice-switching oligonucleotides (SSOs) or modulating the expression and activity of upstream splicing factors[106].

1.4 Tau splicing regulation in physiology and pathology

1.4.1 Tau splicing regulation in the nervous system

Tau is encoded by the MAPT gene, which is located on chromosome 17[38]. The MAPT gene has 16 exons with exon 0 and exon 14 being transcribed but not translated[38]. Exons 2 and 3 encode for the two N-terminal inserts (29 amino acids each) at the N terminal half[107]. Exons 9, 10, 11 and 12 encode for four microtubule binding repeats[107]. There are six Tau isoforms in the CNS, which are driven by alternative splicing of exons 2 ,3 and 10[107]. With the inclusion or exclusion of exon 2 and exon 3, 0N,1N and 2N inserts are generated. Additionally, alternative splicing of exon 10 can produce either 3 microtubule repeats (3R) or 4 microtubule repeats (4R)[108]. Exon 2 cannot be transcribed in the absence of exon 3[107]. Hence, through alternative splicing of these exons, six isoforms (0N3R, 0N4R, 1N3R, 1N4R, 2N3R, 2N4R) with molecular masses ranging from 37 kDa to 46 kDa can be produced, which are corresponding to the Tau isoforms present in human brain[109]. Tau can also encode for a high molecular weight Tau isoform (110-120 kDa), but this long isoform is only present in the peripheral nervous system (PNS)[110].

In the CNS, Tau splicing is developmentally regulated with only the shortest 3R isoform detected in fetal brains, whereas all six isoforms are present in adulthood[111]. In human adult brain, the 2N-Tau species account for a small proportion (9%) of total Tau amount compared to 0N- and 1N-Tau isoforms, which make up 37% and 54% of the total, respectively[112]. However, the amount of 3R- and 4R-Tau species are relatively equal in healthy adult brain, which is critical to maintain the normal neuronal activity[113].

1.4.2 Aberrant Tau splicing in tauopathy

In 1998, MAPT mutations were first discovered in frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17)[114-116]. To date, more than 50 MAPT

mutation sites have been sequentially identified in FTDP-17. Some of these mutations lead to an imbalanced expression of 4R-/3R-Tau through modulating alternative splicing of exon 10[117], the resulting accumulated 3R-Tau or 4R-Tau in neurofibrillary tangles is a pathological hallmark for some FTDP-17 cases. To date, many transgenic Tau mouse models have been generated carrying FTDP-17 mutations, especially those with Tau exon 10 splicing related mutations such as N279K, N296H and IVS10+16 C>T[118-120]. All these transgenic mouse models can partially recapture the FTDP-17 like phenotype. For example, in 2013, Umeda et al. generated transgenic mice carrying Tau IVS10+16 C>T mutation. They demonstrated that these mice exhibited overproduction of 4R-Tau and FTDP like phenotypes, such as hyperphosphorylation of Tau, impaired behavior, and neuronal loss in latter stages[119]. Even though the molecular mechanisms regarding how excess production of 3R- or 4R-Tau leads to FTDP-17 remain elusive, the unbalanced expression of 4R-/3R-Tau has also been described in other sporadic tauopathies. For example, in PSP and CBD, overproduction of 4R-Tau contributes to Tau deposits, whereas 3R-Tau is dominant in Tau filaments in Pick's Disease[121]. All these findings imply that abnormal expression of Tau species due to aberrant Tau exon 10 splicing is sufficient to trigger the pathological accumulation of Tau proteins in tauopathies.

1.4.3 Known factors contributing to Tau exon 10 splicing regulation

Since distorted 4R/3R Tau ratio is implicated in tauopathy pathogenesis, efforts have been made over the last decades to investigate Tau exon 10 splicing regulation through both cis elements and transacting factors. Studies from the Schellenberg's lab demonstrated that Tau exon 10 expression was regulated by weak 5' and 3' splice sites. In their study, they

characterized the cis regulatory elements in both exon 10 and intron 10 where most of the FTDP-17 mutations were located based on known sequences or deletion coupled splicing tests [122-124]. Three ESEs, including SC35-like enhancer (targeted by splicing factor SC35), polypurine enhancer (PPE), and A/C-rich enhancer (ACE), were characterized in the 5' site of exon 10. Another ESS, followed by an ESE, was identified in the 3' site. An ISS element and an intron splicing modulator (ISM) were also uncovered in the 5' splice site of exon 10[123, 124]. Furthermore, the stem loop structure predicted by the complementary sequence located at the 5' splice site of exon 10 might also be involved in exon 10 splicing regulation through competing with U1 snRNPs[116]. The FDTP-17 mutations identified in the 5' splice site may disrupt or destabilize the stem loop structure, which further allows for the binding of U1snRNPs to the 5' splice site to promote exon 10 inclusion. Precise splicing regulatory functions of cis-elements rely on the cooperative binding of trans-acting factors. During the last decades, scientists have also been dedicated to identifying the splicing factors or RNA binding proteins (RPBs) that are involved in modulating exon 10 splicing. SRSFs and hnRNPs are the two major groups of proteins that participate in the alternative splicing process through marking or hiding a splice site to determine the splicing fate of an exon or intron in a cell type or tissue dependent mode[125-128]. SRSF1 or serine/arginine-rich splicing factor 2 (SF2) has been reported to be dysregulated in tumors and acts as an oncoprotein through regulating alternative splicing events of its downstream target genes during tumorigenesis[129]. As an extensively studied splicing factor in cancer, SRSF1 has a conserved binding motif, which is (A/G) GAAGAAC. In 2006, D'Souza et al. identified SRSF1 as an essential splicing factor promoting exon 10 inclusion by acting on the poly-

purine enhancer located at the 5' site of exon 10 and demonstrated that the SR domain is critical for SRSF1 to exert its function in Tau exon 10 splcing regulation [130]. Meanwhile, their study also showed that FTDP-17 mutations N279K and Δ280K located at the same motif had opposite effects for exon 10 splicing. While N279K showed an increase in exon 10 inclusion, Δ280K reduced exon 10 splicing, which could have resulted from an addtional AAG motif that was created by N279K[130]. Based on the known binding motif of SC35(SRSF2), SC35 was predicted to act on a SC35-like enhancer located at the 5' site of exon 10[131]. In 2011, Qian et al. confirmed the interaction between SC35 and this predicted motif, and futher demonstrated that SC35 indeed utilized this binding site to promote exon 10 inclusion.[132]

Besides SRSF1 and SRSF2, other SRSF family proteins including SRSF 3, SRSF 4, SRSF 6, SRSF 7, SRSF 9 and SRSF 11 were identified sequentially to regulate exon 10 splicing, among which SRSF 6 and SRSF 9 were validated to promote exon inclusion, whereas the others had shown the opposite effect[133]. The activity of SRSF family protein is known to be regulated by phosphorylation[134]. Hence, many research groups were trying to investigate upstream kinases with regulatory effects on Tau exon 10 splicing, which could potentially be used as therapeutic targets for tauopathy treatment. It had been reported that SC35 could be phosphorylated by Dyrk1A (dual-specificity tyrosine-phosphorylated and regulated kinase 1A), which suppressed SC35's function to promote exon 10 inclusion[132].

Other RBPs including RBM4, hnRNPG, hnRNPE2, hnRNPE3, DDX5, SWAP, CELF1, CELF3, TNRC4, PTBP2 and PSF were also reported to play a role in regulating

Tau exon 10 splicing[107, 133]. Recently, the identification of Tau splicing factors, Fused in Sarcoma (FUS) and Transactive Response DNA-binding Protein of 43 kDa (TDP-43), might help to explain the pathological Tau splicing patterns in neurodegenerative diseases[135, 136]. Both FUS and TDP-43 are RNA/DNA binding proteins with predominant expression in the nucleus, where they are involved in RNA processing. Mis-translocation of these two RBPs into the cytoplasm can lead to aggregate formation, which is one of the pathological hallmarks of amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD)[137, 138]. Loss of nuclear activity of these two RBPs may have contributed to the pathological changes in ALS and FTLD[139-141]. In 2012, Orozco et al. reported that knock down of FUS or mis-localization of FUS induced by ALS/FTLD related genetic mutations led to exon 10 inclusion[135]. In 2017, Liu's lab demonstrated that TDP-43 interacted with Tau pre-mRNA in intron 9 and functioned to promote exon 10 inclusion[136]. They also found that ALS/FTDP-associated TDP-43 mutations could increase the production of 4R-Tau compared to WT TDP-43[136]. Taken together, these findings suggest pathological connections between ALS/FTDP risk genes and aberrant Tau splicing during the pathogenesis of neurodegenerative diseases.

1.5 Earlier studies about Tau isoform specific functions

As mentioned earlier, Tau splicing is developmentally regulated in human brain[111]. Moreover, the identification of Tau mutations affecting the 4R-/3R-Tau ratio in FTDP-17 as well as the differential composition of Tau isoforms in Tau tangles among different types of tauopathies, indicates that there are functional differences between 3R- and 4R-Tau isoforms[117, 142]. Although the molecular mechanism underlying the dysregulation of 3R- and 4R-Tau directed

Tau aggregation and the subsequent neuronal loss and cognitive deficits remain unclear, early studies have revealed functional differences between 3R- and 4R-Tau in microtubule behavior regulation.

The "linear view" hypothesis is that 4R-Tau has higher microtubule binding affinity compared to 3R-Tau due to an additional microtubule binding repeat. However, further studies have shown that the inter-repeat region between the first and second microtubule binding repeats was involved in tubulin binding and assembly[143, 144]. In 1994, Gasket et al. systematically investigated the interaction between Tau and microtubules. Their study revealed 3-fold lower microtubule binding affinity of 3R-Tau compared to 4R-Tau based on the disassociation constant Kd values[145]. Meanwhile, in the same year, another study observed around 40-fold differences[143]. The binding differences between these two studies may come from the distinct translation systems that were used in their studies. The former study used a bacteria expression system, whereas the latter study used an *in vitro* translation system for Tau expression. Moreover, it was reported that 3R- and 4R-Tau have their own core microtubule binding regions. The first two repeats (R1-R3) and the intervening inter-repeat (IR) were the main regions involved in microtubule binding of 3R-Tau[144], whereas the R1-R2 IR was unique for 4R-Tau binding to microtubules[143]. Additionally, the C-terminal end of Tau protein was reported to modulate and increase microtubule association by several studies[143, 144, 146, 147]. Recently, Butner and the others further elaborated on the functional differences between 3R- and 4R-Tau in microtubule dynamics[147, 148]. They both found that compared to 4R-Tau, 3R-Tau lacked the ability to suppress microtubule shortening[147, 148]. A recent animal study reported that overexpression of the human 3R-Tau isoforms in mouse brains exhibited an increase in phosphorylated Tau, accumulation of Tau aggregates, and synapse and memory deficits[149]. In

the same study, through *in vivo* and *in vitro* assays, they found that overexpression of 3R-Tau could induce oxidation mediated double strand DNA breaks (DSB), while treatments with antioxidants such as vitamin C/E could alleviate the toxicity caused by 3R-Tau[149].

Although most studies have focused on the differences between 3R- and 4R- Tau, there were several studies that focused on Tau isoforms differing in the N-terminal region. In 2013, Liu et al. investigated the distribution of Tau isoforms in mouse adult brain. They found that 0N-Tau was enriched in the "classical region", cell body and axons, whereas the majority of the 2N-Tau was localized in the cell body and dendrites[150]. Intriguingly, 1N-Tau was enriched in the nucleus[150]. The differential distribution of Tau isoforms differing in their N-termini might correlate with their specific functions in different cellular compartments[150]. To test this hypothesis, they generated Tau isoform-specific antibodies targeting 0N-, 1N- and 2N- Tau isoforms and performed immunoprecipitations with mouse brain lysate. Through TMT-labeled quantitative proteomics and data analysis, they found proteins that have preferential binding with 2N-Tau were enriched in pathways that associated with neurodegenerative diseases[151].

1.6 Mouse models for tauopathies

Since hyperphosphorylated Tau is a major component of neurofibrillary tangles in different types of tauopathies, there have been many research groups studying the mechanisms underlying Tau accumulation and phosphorylation. As a promising therapeutic target, aggregation of Tau has been reported to be strongly correlated with the clinical symptoms and neurodegeneration in AD and primary tauopathies. To date, many transgenic mice have been generated to model AD or tauopathies. In 1999 and 2000, Tau transgenic mice were generated to overexpress the longest 2N4R and shortest 0N3R Tau isoforms,

respectively[152-154]. Increased Tau levels indeed produced Tau pathologies, such as pre-tangles, motor disturbances and behavioral deficits, but the neurofibrillary pathology was lacking with only rare tangles present in 0N3R transgenic mice up to 24 months of age[153]. Additionally, Duff et al. created P1-derived artificial chromosome (PAC) transgenic Tau mice containing the full sequence of the human Tau gene[155]. These mice produced all six Tau isoforms, but transgenic Tau splicing was altered with predominant expression of the 3R-Tau isoforms[155]. Even though there was an approximately 3-fold increase of the total Tau levels (mouse Tau and human Tau) compared to non-transgenic mice, these mice lacked Tau pathology up to 8 months of age[155]. Taken together, WT transgenic Tau mice display limited Tau pathology even when total Tau levels are increased.

In 1998, several groups observed mutations in MAPT locus in FTDP-17[115, 116, 156]. Since then, researchers started to create Tau transgenic mice carrying these mutations to accelerate the process of NFT formation for therapeutic studies. In 2000 and 2001, Lewis et al. and Gotz et al. generated Tau P301L transgenic mice that expressed either the shortest 4R Tau isoform or the longest 4R Tau isoform with the P301L mutation in exon 10 [157, 158]. These transgenic mice both displayed motor and behavioral deficits at an early stage. NFTs that were mainly composed of straight Tau filaments were also observed in the spinal cord, cortex, brain stem and other regions, accompanied with a large portion of neuronal loss in 8-months old P301L transgenic mice. Similar phenotypes were also observed in other transgenic mouse models carrying missense Tau mutations, such as V337M[159]. In addition to Tau protein function related mutations, many Tau exon 10 splicing associated mutations have also been linked to FTDP-17. For example, MAPT N296H is a mutation site located

in exon 10 that gives rise to excess expression of 4R-Tau species. A study reported that a Tau transgenic mouse model carrying this mutation site successfully recapitulated the early major pathological features of FTDP-17 cases[120].

AD displays secondary tauopathy, which is characterized by two pathological hallmarks, beta amyloid (Aβ) plaques and NFTs. To study the pathological role of both Aβ and Tau, Oddo and his colleagues derived a triple transgenic mouse model (3xTg-AD) carrying Presenilin 1 (M146V), Amyloid precursor protein (Swe), and Tau (P301L) mutations[160]. In their study, transgenic constructs that encoded amyloid precursor protein (Swe) and Tau (P301L) were co-microinjected into the single-cell embryos of Presenilin 1 (M146V) knockin mice[160]. Presenilin 1(PS1) is one of the genetic risk factors that accounts for a subset of familial AD cases[161]. The inclusion of PS1 (M146V) was generally used to accelerate the aggregation rate of Aβ in mouse models. In the mouse model of 3xTg-AD, Aβ plaques where observed in the frontal cortex region in 6-month-old mice, whereas Tau pathology started to appear in the hippocampus at 12 months of age[160]. Tau paired helical filaments were not observed until 18 months[160]. Despite the similar expression of amyloid precursor protein (APP) and Tau mutants, Aβ deposits appeared much earlier compared to Tau pathology[160], which suggests that abnormal Aβ inclusions are the major causal factor in AD. Besides the pathological changes of these two proteins, 3xTg-AD also developed synaptic deficits[160], which suggest the intricate relationships among Aβ plaques, Tau tangles and synaptic dysfunction.

1.7 Tau-targeted therapeutic approaches for treatment of tauopathies

Identification of Tau dominant mutations in FTDP-17 has provided critical evidence to support that Tau dysregulation is the core factor to directly trigger neurodegeneration pathogensis[162-166]. In addition, many studies have demonstrated that Tau lesions are correlated with the clinical symptoms and severity of AD and other Tauopathies[167-171]. Therefore, targeting Tau or pathological Tau represents a promising approach for the development of therapeutics to alleviate disease progression. To date, significant efforts have been made to develop Tau targeted therapeutics, which are mainly divided into two categories depending on if the therapeutics are designed to target Tau loss of function or toxic gain of function[172].

The majority of current therapeutics for Tauopathies are designed to target the "gain of toxic Tau functions", including reducing total Tau or pathological Tau expression, inhibiting Tau aggregation, modifying Tau PTMs to reduce phosphorylated Tau levels, increasing Tau clearance with vaccines and monoclonal antibodies, and blocking Tau transfer from cell to cell[172]. In recent years, antisense oligonucleotide (ASO)-directed gene therapies are becoming more and more popular and have been used as precision medicine approaches in cancer and brain disorders such as Huntington's disease[173]. ASO is a short segment of DNA oligomer that is 18-30 nucleotides in length[174]. The sequence of ASO is designed to be complementary to the target RNA sequence[174]. The binding of the synthesized ASO with its RNA target provokes RNA degradation by ribonuclease H1 (RNase H1), leading to RNA translation blockade and the subsequent inhibition of production of disease related proteins[174]. In 2013, BIIB080, a Tau mRNA targeted ASO, was developed by IONIS in collaboration with a research group in the University of

Washington[175]. About 50% reduction of Tau mRNA expression was observed throughout the brain and spinal cord with the infusion of BIIB080 in non-transgenic mice[175]. Further study in Tau P301S transgenic mice showed that the BIIB080 treatment could reduce neuronal loss and rescued behavioral deficits[176]. Aside from modulating the total Tau levels, ASO have also been designed to alter Tau splicing pattern to restore the distorted 4R-/3R-Tau ratio in a subset of tauopathies such as PSP[177]. However, these are all preclinical studies, efforts are still needed to test the safety, tolerability and efficacy of Tau ASOs.

Hyperphosphorylated Tau is an early driver for neuronal dysfunction in AD and other tauopathies[178]. Numerous studies have been dedicated to decreasing phosphorylated Tau levels with kinase inhibitors. Potential inhibitors that target GSK3β, one of the key kinases that functions to promote Tau phosphorylation, have been developed as therapeutics for tauopathy treatment, including Lithium, valproic acid and Tideglusib[179-181]. Even though Lithium and valproic acid showed improvement in relieving Tau pathology and cognitive impairment in transgenic mice models[181, 182], they were both poorly tolerated and inefficacious in some of the tauopathy populations during clinical trials. The poor outcomes might result from off target toxic effects of Lithium and valproic acid, as both of them do not specifically target GSK-3β[172]. Unlike Lithium and valproic acid, Tideglusib was developed as a specific inhibitor of GSK-3β[180]. As expected, Tideglusib was well-tolerated in an early small scale clinical trial[183]. However, no significant efficacy was observed in the subsequent larger population of clinical trials. Besides inhibiting GSK-3β, there is also a group of studies focusing on other Tau kinases, such as Fyn and Abl, which

have shown some improvements in Tau pathology[172]. Despite these findings, larger clinical trials are needed to provide more reliable results.

To promote the clearance of Tau aggregates, the idea of immunotherapy has been integrated into the development of novel therapeutic drugs for dementia. Over the past few years, several vaccines using Tau fragments as epitopes have been developed to promote the clearance of pathological Tau. One of the most promising candidate Tau vaccines is AADvac 1, which showed a decrease of Tau aggregation and improvement of the clinical phenotype in transgenic rat models[184]. In a larger phase 2 clinical trial, 208 patients with mild AD received AADvac 1 treatment for 24 months[172]. Early data showed that the immunization was successful, the tolerance was good, and the clinical characteristics might be partially improved, although the data has not been fully released[172].

Besides vaccines, the development of humanized monoclonal antibodies targeting Tau, a passive immune approach, has become more and more popular. Given that a truncated form of Tau (eTau) containing the N-terminal and mid-domain regions has been identified in cerebrospinal fluid (CSF) of AD patients and other tauopathies[185, 186], the majority of Tau monoclonal antibodies were designed to target these two fragments. In transgenic mouse models, two Tau monoclonal antibodies, BIIB092(Gosuranemab) and C2N-8E12, have shown improved phenotypes including a reduction of secreted eTau, decreased Tau aggregation and rescue of cognitive impairment[187, 188]. BIIB092 also showed good safety profile and tolerability in an early clinical trial[189]. Larger clinical trials will start for these two antibodies. Another recently developed antibody named LY3303560 showed preferential binding with Tau aggregates compared to Tau monomer through *in vitro* testing,

suggesting that this antibody might serve as a drug candidate to selectively alleviate pathological Tau species[190]. However, the clinical results are still lacking.

In additional to immunotherapy-based Tau clearance, a new technology named proteolysis-targeting chimeras (PROTAC) has also been applied to promote Tau clearance through a protein degradation mechanism. PROTAC is a class of hetero-bifunctional molecules, each of which is composed of a short peptide that has binding affinity with E3 ubiquitin ligase, a small molecule/peptide that can bind to the target protein and a connecting linker sequence[191, 192]. With the above two functional domains, PROTAC acts as a mediator to recruit proteins of interest to E3 ubiquitin ligase for poly-ubiquitination and subsequent proteasome-mediated protein degradation[192]. The emergence of this technology has greatly attracted the attention of scientists to apply this method as a novel therapeutic approach for the treatment of cancer and neurodegenerative diseases. In 2018, Lu et al. developed a Kelch-like ECH-associated protein-1(Keap1) -dependent peptide PROTAC to achieve selective degradation of intracellular Tau[193]. Keap1 is an E3 ubiquitin ligase substrate adaptor[194]. In this study, they designed a BIOTAC, which contained two functional regions, a short peptide derived from beta-tubulin in the N-terminal region to specifically recognize Tau, and a Keap1 binding motif in the C-terminal region. Besides these two functional peptides, there was also a poly-D-arginine sequence at the C-terminus that functions to promote cell penetration [193]. By incubating these PROTAC molecules with neuroblastoma cells, they found that the expression level of Tau was decreased significantly, and this process employed in a proteasome-dependent degradation mechanism, which

greatly increased the possibility of using PROTAC-directed specific degradation of Tau in the treatment of tauopathy[193].

Another direction for Tau targeted therapy is restoring the microtubule functions that were disrupted by the loss of normal Tau activity in tauopathies. Several drugs that function as microtubule stabilizers, such as Davunetide (NAD) and Epothilond D, have been tested in clinical trials. Davunetide is an eight amino acid peptide derived from a neurotrophic protein[195], whereas Epothilond D is a chemical compound identified in the myxobacterium *Sorangium cellulosum*[196]. They both showed an increase in microtubule density and cognitive improvement in transgenic mouse models[197, 198]. In early clinical trials of Davunetide, improvement of the working memory was observed in patients diagnosed with mild cognitive impairment (MCI)[172]. However, in clinical trials of PSP patients, no curative effect was observed[199].

Although a certain number of Tau targeted drugs are in clinical trials, available therapeutic options to alleviate the disease progression of Tauopathies remain limited. Thus, more efforts are still needed to clarify the molecular mechanisms underlying the pathogenesis of tauopathy, which can pave the way for developing more promising therapeutic targets for tauopathy treatments.

1.8 Efforts to pull down RNA and its associated RBPs

Traditional approaches to identify splicing factors can be classified as in-solution RNA pull downs, where RNAs of interest can be transcribed *in vitro* and then be used to incubate with cell or tissue lysates[200-203]. The *in vitro* synthesized RNA targets can be labeled with

biotin, which allows for the identification of RNA-associated proteins through affinity pull down with streptavidin-coupled beads[201, 204, 205]. In-solution pull down is a powerful approach to validate RNA-protein interactions, but it suffers from prior cell lysis and re-association of RNA-protein complexes, which cannot accurately recapture the physiological interactions between RNAs and protein[206]. To overcome the limitation of in-solution RNA pull downs, *in vivo* UV crosslinking has been thought to preserve the physical interactions between RNAs and proteins. For example, Castello et al. successfully revealed global mRNA binding proteins under physiological conditions by using *in vivo* UV crosslinking and subsequent oligo dT pull down[207]. However, this technique cannot be applied to study RBPs that function in a specific splicing event.

Recently, two oligo-based pull-down methods, named comprehensive identification of RNA binding proteins by mass spectrometry (ChIRP-MS) and RNA antisense purification by mass spectrometry (RAP-MS), were developed to specifically target the long non-coding RNA (lncRNA), X-inactive specific transcript (XIST), to unravel its associated RBPs [208, 209]. Instead of using RNA to pull down its associated proteins, ChIRP-MS and RAP-MS fix the RNA-protein interactions in native state by *in vivo* crosslinking in living cells with either formaldehyde or UV irradiation and then targeting the RNA-protein complexes with DNA probes that are complementary to the RNA sequence *in vitro* (Figure 1.2). ChIRP-MS is optimized from ChIRP-seq, a high throughput method designed to discover the genomic binding sites of a lncRNA of interest[210]. Instead of looking for the DNA binding site of lncRNA, ChIRP-MS is developed to enrich for RNA and its associated proteins. Briefly, in the ChIRP-MS method, cultured cells are crosslinked with 3%

formaldehyde to preserve RNP complexes[208]. Following sonication, a pool of 20-nucleotide biotinylated DNA probes, paired with RNA of interests, are hybridized with cell lysate to target a specific RNA sequence [208]. By using streptavidin beads, the whole complex can be captured and purified. Next, free biotin was utilized to elute proteins from beads [208]. The protein eluate is then precipitated and purified by TCA and acetone [208]. After air drying, the protein pellets were resuspended in 1x laemmli sample buffer, followed by reverse cross linking, and the captured proteins are separated in a bis-tris SDS-PAGE gel and the size specific bands can be analyzed by mass spectrometry[208]. Similarly, *in vivo* cross linking and biotin labeled DNA probes are also used in RAP-MS. RAP-MS was demonstrated to identify direct RNA-protein interactions with high sensitivity[209]. In the RAP-MS method, ultraviolet (UV) at 254 nm is first used to create covalent bonds between RNAs and proteins, which are irreversible[209]. Secondly, longer DNA probes (90 nucleotides) are used in RAP-MS to ensure stable DNA/RNA hybrid formation allowing for capturing to occur in denaturing and reducing buffer conditions[209]. For example, 4M urea and 67°C hybridization conditions were introduced in the original protocol[209]. Urea can disrupt RNA secondary structure for efficient probe capturing and the high temperature can remove non-specific binding. These stringent buffer conditions can also help to reduce non-crosslinked background proteins. Compared to ChIRP-MS, RAP-MS is a more robust approach to capture direct interactions, but the low efficiency of UV crosslinking may limit its sensitivity to identify those proteins that are directly bound to the RNA of interest. However, this can be compensated for by increasing the cell number or tissue amount.

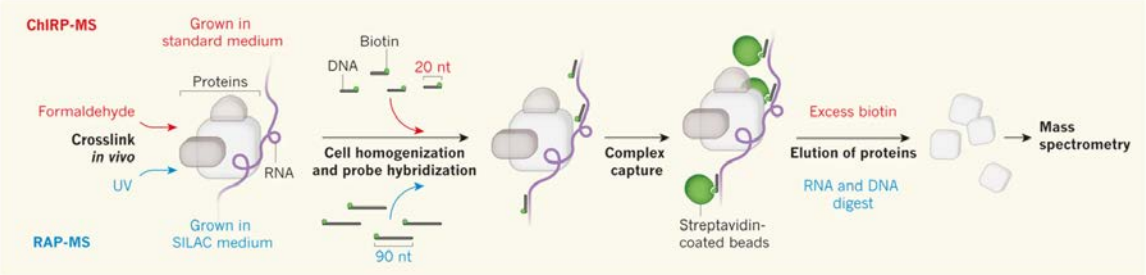


Figure 2: Workflow of ChIRP–MS and RAP-MS[211]

1.9 Enzyme-mediated proximity labeling techniques for *in vivo* interactome profiling

Precise regulation of cellular functions such as signal transduction and response to external stimuli, requires assembly of proteins to form functional complexes and the subsequent protein interaction networks. In addition, depiction of protein-protein interaction networks is also essential to study the molecular mechanisms underlying disease progression. Antibody-based affinity purification (AP) is the traditional approach that has been used for decades to study protein-protein interaction (PPI)[212]. It is a powerful way to validate PPI, but it has some drawbacks for discovering novel protein partners *in vivo*. First of all, antibody-based AP requires prior cell lysis, which is more likely to lose physical interactions, especially weak and transient interactions[213]. In addition, AP may produce high false positive interactions during destruction of subcellular structure[214]. Finally, the success of affinity purification is highly dependent on whether the bait protein has high quality antibodies, which greatly limits its wide application[213].

Over the last decade, enzymatic based proximity labeling (PL) has been developed and utilized to study compartmental protein composition and protein interaction networks in living cells[215, 216]. Proximity labeling relies on the transformation of the substrate into a

reactive radical to promiscuously tag the neighboring proteins. The resulting tagged proteins can be then isolated from the cells by affinity purification. To date, there are mainly three PL systems, including BioID (proximity dependent biotin identification), APEX (engineered ascorbate peroxidase), and PUP-IT (pupylation-based interaction tagging).

BioID relies on a mutant form of the biotin ligase BirA, which can non-specifically tag neighboring and interacting proteins[216]. Fusion of a bait protein with this modified biotin ligase (BirA*) allows for biotinylation of the proteins in proximity with the bait proteins. These biotin labeled proteins can be isolated by streptavidin pull down and analyzed by mass spectrometry[216]. Biotinylation is a rare post translational modification with high specificity for only a few proteins, such as carboxylases and decarboxylases[217]. BirA, a 35 kDa protein, is a biotin ligase identified in *Escherichia coli*[218]. BirA-directed protein biotinylation occurs in two steps. Firstly, with the help of adenosine triphosphate (ATP), biotin is activated to form biotinoyl-5-AMP (bioAMP)[219]. Next, this reactive radical will be held within the catalytic site of BirA until it tags to the lysine residues of substrate proteins with a biotin receptor sequence[219]. In 2004, Choi-Rhee et al. first reported a mutant form (R118G) of biotin protein ligase BirA or BirA* that can promiscuously biotinylate neighbor proteins[218]. This concept was then named as proximity dependent biotin identification or BioID by Roux et al. and applied to study protein interactomes in mammalian cells.[216]. Since then, BioID has been widely used to map subcellular protein compositions and protein interactome networks in cells and animals. In 2016, Kim et al. screened a smaller biotin ligase from *Aquifex aeolicus*, named BioID2[220]. Due to the lack of a DNA binding domain in the N-terminal half, the protein size of BioID2 (27 kDa) is

smaller than BirA*, which helps with the accurate localization of a tagged protein[220]. However, one limitation for BioID is its slow kinetics, usually taking around 18-24 hours for biotin labeling[216]. The long-term accumulation of biotin labeled proteins may be toxic to cellular processes[221]. To improve the catalytic efficiency, TurboID was generated by directed evolution based on BioID using yeast display[221]. TurboID biotin labeling was reported to be too much faster than that in BioID, which only takes 10 minutes for proximity labeling[221].

APEX is an engineered ascorbate peroxidase that was first reported by Ting's lab and it was initially described to act as an electron microscopy (EM) tag to achieve cellular structure imaging for all cell compartments[222]. With the presence of H_2O_2, APEX can catalyze oxidation of diaminobenzidine (DAB) to form localized deposits, followed by OSO4 staining, EM contrast will then be generated to achieve subcellular structure observation[222]. Additionally, genetically coded APEX is also a useful tool for proteomic mapping in living cells[223]. With the addition of H_2O_2, phenol-biotin can be catalyzed by APEX to produce highly reactive, short-lived biotin-phenoxyl radicals to tag the proximal proteins of a bait protein, the corresponding biotin labeled proteins can then be purified with streptavidin beads and detected by mass spectrometry[223]. To improve the sensitivity of APEX, through several rounds of direct evolution by using yeast display system, Ting's lab identified APEX2, an A134P mutant of APEX, with higher catalytic activity, which allowed for higher resolution of ultrastructural imaging and enabled the observation of dynamic proteome changes in living cells[224]. Compared to BioID, APEX directed intracellular proximity labeling has rapid kinetics with roughly one minute for labeling,

which can be utilized to study proteins with short half-lives, monitor the dynamic interactome changes upon drug treatment or activation of a specific signaling pathway[225]. However, the use of H_2O_2 in APEX based proximity labeling can cause cellular toxicity[221], which may influence normal biological processes and make the data interpretation difficult. Also, peroxidase-based approach can only tag the reactive radical to electron rich tyrosine residues[226], which may limit the total number of identified proteins.

Pup-IT is a newly developed PL system based on the bacterial PafA enzyme and has been used to study interactors of membrane proteins[227].PafA is a ligase that can tag proteins with pup, a 64 amino acids protein with Gly-Gly- Gln at the C-terminus. During pupylation, the C-terminal Gln is first deaminated to form Pup(E) (also known as PupGlu)[228]. Next, in the presence of ATP, PafA phosphorylates and conjugates Pup(E) to a lysine residue of the target protein[229, 230]. To detect pupylation, Pup(E) was fused with a bacterial-derived carboxylase domain (BCCP) so that it can be tagged with biotin by endogenous biotin ligase in mammalian cells. Then, Bio- Pup(E) and pupylated proteins can be detected and isolated[227]. Unlike BioID and APEX, PafA does not release the reactive tags, which ensures that only the proteins in close contact with PafA can be conjugated with Pup. On the flip side, its high specificity properties also limit its application in spatial proteomics studies[225].

1.10 Mass Spectrometry Based Quantitative Proteomics

Over the past decades, liquid chromatography–mass spectrometry (LC-MS) has been routinely used in quantitative proteomics to identify and quantify proteins across samples based on the mass-to-charge (m/z) ratio in systems biology and clinical research. To

quantify the relative protein amount, according to the “bottom-up” proteomic procedure, protein samples, such as cell lysate, will be digested with proteases[231], the resulting proteolytic peptides will then undergo fractionation and ionization before entering into a mass spectrometer[232, 233]. Given the limitation of the resolving power of mass analyzers, a selected peptide precursor ion will be further collided into small fragments ions by an inert gas for protein identification[234]. The spectrum of eluted peptide ions for a specific retention time are named MS1, whereas MS/MS or MS2 refer to the spectrum of eluted fragment ions[235]. The best peptide match can be found by comparing the m/z ratio generated from both MS1 and MS2 scans[236]. The corresponding ion intensity from MS1 or MS2 scans can be utilized for protein quantification. MS-based protein quantification can be divided into two categories, absolute and relative quantifications. Due to biased efficiencies in protein digestion and peptide ionization, MS signal is typically not a direct read-out of protein amount in a given sample[234]. However, absolute quantification can be achieved by spiking in a standard peptide with known concentrations. Nevertheless, the high cost of standard peptide has limited absolute quantification to small-scale proteomics assays[237]. Relative quantification has been developed with numerus advanced techniques being adapted, which enables relative protein quantification across two or more samples. Relative quantification methods include label-free and stable isotope labelled approaches, such as Stable Isotope Labeling with Amino acids in Cell culture (SILAC) and tandem mass tags (TMT).

1.10.1 Label-free

Label-free is the most traditional way to quantify proteins across different samples, which has no limitation for sample numbers. Recently, Johnson and co-workers have analyzed the

proteomic profile for hundreds of AD post-mortem brain samples based on this method[238]. Label-free methodology is based on the intrinsic quantification role of MS with the peak area of MS1 spectra as the read-out for the total ion amount of a given peptide. Even though the label-free method is cost-effective with no additional steps needed during sample preparation, missing value issues have always been considered as a big issue for accurate comparisons across samples[239]. Additionally, run-to-run variation may also contribute to inaccurate quantification in the label-free method, especially for large numbers of samples.

1.10.2 Metabolic labeling-SILAC

With the introduction of heavy isotope labels, multiplexed label-based quantification approaches have been developed and improved rapidly to overcome the limitations of label-free approaches. In 1996, Schnolzer et al. first introduced isotopic labels in proteomics by incorporating an ^{18}O atom at the C-terminal end of tryptic peptides[240]. Since then, many studies started to focus on chemical labeling at both the protein and peptide level to achieve relative protein quantification[241, 242]. In 2002, Ong et al. developed Stable Isotope Labeling with Amino acids in Cell culture (SILAC) with the incorporation of isotopic amino acids in cultured cells[243]. *In vitro* cultured cells cannot synthesize some amino acids, and these essential amino acids need to be supplied in the culturing medium, which creates the chance to incorporate isotope labeled amino acids in the metabolic cycles of cultured cells to achieve protein isotopic labeling in living cells. The mass difference between normal and heavy isotopic labeled peptide can be easily distinguished by a mass analyzer to achieve relative quantification across samples. Compared to label free approaches, SILAC enables co-analysis of different samples in a single MS run, which avoids run-to-run variations,

greatly enhances the accuracy of quantification, and saves on instrument usage time. Although SILAC is a powerful technique for multi-sample quantification, the high cost of the SILAC reagents greatly restricts its wide application in proteomics research. Additionally, SILAC is a MS1-based quantification method, multiplexing will increase sample complexity during MS1 quantification. Moreover, the inherent working principle of SILAC determines that SILAC cannot be applied to study clinical samples. In 2018, limitations of SILAC were partially alleviated by Overmyer and co-workers[244]. With the incorporation of neutron-encoded stable isotopes, minor mass differences were generated across samples. As such, they were able to achieve multiplexing labeling and quantification for up to 9 samples in their study[244]. Metabolism labeling has also been applied for *in vivo* studies. In 2004, Yates's lab achieved stable isotope labeling in mammalian organisms by feeding rats with a diet that was enriched with ^{15}N[245]. The labeling efficiency of this approach has then been further improved in slow turnover tissues (i.e., brain tissues) and has been widely used to evaluate the dynamic proteome changes in animal models of human diseases such as neurodegenerative diseases[246, 247].

1.10.3 Isobaric labeling-TMT

The development of isobaric labeling is a revolutionary achievement in quantitative proteomics. Unlike label-free or SILAC, which employ MS1-based quantification, isobaric labeling relies on MS2 based quantification, which allows for multiplex labeling without increasing the sample complexity during MS1 scans. The most widely used isobaric labeling approach is tandem mass tags (TMT), which are commercially available for research use. The concept of TMT was first introduced by Thompson et al. to achieve

isobaric labeling for two samples[248]. TMT refers to a chemical tag that is composed of a reporter group used for relative quantification, a balancer group to normalize the total mass of the tag and a primary amine reactive group to covalently tag the tryptic peptides[249]. In comparison to previous isotopic labeling methods, TMT ensures that peptide identification and quantification both occur at the MS2 level. Even though the reporter groups of TMT tags have different masses, pairs of peptides are still isobaric after labeling. As such, identical peptides from different samples will co-migrate through LC separation and exhibit a single peak in MS1 scans, which greatly improve the sensitivity of MS[234]. Additionally, only TMT labeled peptides will be considered for relative quantification rather than the unlabeled contamination to obtain high quality data[234]. To date, 2-plex TMT has been extended rapidly to achieve 6-plex, 10-plex, and 16-plex labeling with the incorporation of heavy isotopes in different chemical groups[249-251].

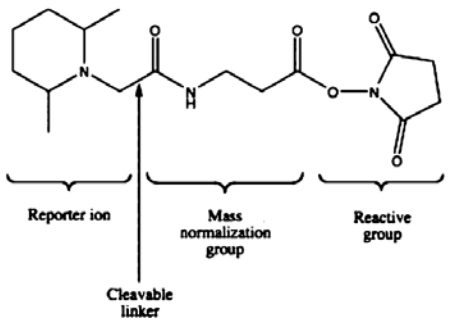


Figure 3: Structure of the Tandem Mass Tag[249]

The major limitation for isobaric-based TMT labeling is that the measured difference for a given protein across different samples is always smaller than the actual changes, as contaminating ions can also contribute to the relative quantification of a specific peptide[252]. Further studies showed that this “compression ratio” issue was

alleviated by a further MS scan or MS3 to isolate the peptide ions from M2 spectra followed by fragmentation[253]. However, additional MS scans might compromise total peptide identification.

1.11 Summary of intent

Tauopathies are a group of neurodegenerative diseases, including Alzheimer's Disease (AD) and other dementias, that share the common feature of neurofibrillary aggregates of the microtubule-associated protein Tau. It is believed that aberrant splicing of Tau at exon 10 is one of the causes for tauopathies. However, there has been limited efforts to study the upstream Tau splicing factors. Hence, comprehensive understanding of the molecular mechanisms underlying aberrant Tau splicing is crucial for the development of tauopathy therapeutics. The overall purpose of my research work described in this thesis is to use proteomic methods to unveil protein interactors that are involved in alternative Tau splicing events. My work can be divided into two sections: the first section is the identification of upstream splicing factors that contribute to Tau exon 10 splicing using RAP-MS; the second part is the investigation of isoform-specific human Tau interactome by BioID2 coupled quantitative proteomics, which are described in chapters 2 and 3, respectively.

Traditional methods to identify splicing factors were limited to *in vitro* RNA pull down, which is not helpful to unveil RNA-protein interactions in physiological conditions. Hence, it is crucial to develop a method to uncover RNA-protein interactions in a native state or in living cells. In chapter 2 of this thesis, the establishment and optimization of a RAP-MS method in our lab allows for capturing of RBPs that directly bind to Tau pre-mRNA in their native state. Compared to *in vitro* RNA pull down, RAP-MS enables

identification of RBPs that directly bind with the RNA target with higher specificity, which resuslts from two aspects. First, in the procedures of RAP-MS, UV crosslinking is used to covalently link the RBPs with RNA in cultured cells so that their native interaction is preserved. Meanwhile, crosslinked RBPs are compatible with denaturing conditions, which ensures the removal of non-direct and non-specific binding proteins; Secondly, biotinylated 90 nucleotides DNA probes spanning the RNA target sequence are used in our RAP-MS approach to capture RNA and its associated proteins. This allows for the formation of stable DNA/RNA hybrids and ensures that capturing can occur at a high temperatures that disrupt RNA secondary structures for efficient capturing. As described in chapter 2, RAP has been tested to successfully pull down several RNA molecules such as GAPDH and U1 small nuclear RNA (snRNA) with high specificity. Therefore, in combination with UV crosslinking, we propose that our RAP-MS method can be adapted to pull down Tau pre-mRNA and its conjugated RBPs. In chapter 2, through large scale RAP-MS assays with specific Tau pre-mRNA targeted probes, novel RBP candidates that associated with Tau pre-mRNA were identified. Followed by small scale short hairpin ribonucleic acid (shRNA) screening, their splicing functions on Tau exon 10 were then validated. With further validations, we confirmed that hnRNA2B1 and hnRNPC function to promote 4R-Tau production whereas hnRNPH1 was identified as a 3R-supporter. In particular, the functional regulatory cis-elements of hnRNPC, a protein shown to be upregulated in PSP, a sporadic tauopathy characterized with overproduction of 4R-Tau, were deeply studied as described in chapter 2. The research work described in chapter 2 uncovered a number of novel Tau splicing factors, which is crucial knowledge to comprehensively understand Tau

splicing regulation. Some of these identified RBPs, such as hnRNPC, might serve as therapeutic targets for sporadic tauopathy treatment.

The imbalanced distribution and expression of Tau isoforms during development indicate that functional differences might exist among Tau isoforms. It is believed that Tau plays multiple roles beyond microtubule regulation. However, previous functional studies of Tau isoforms, essentially 3R-and 4R-Tau isoforms, were limited to *in vitro* analysis of their regulatory roles for microtubule polymerization and stabilization. Even though some studies described the potential functions that Tau might play in the synapse and nucleus, the regulatory mechanisms remained unclear. Thus, it is important to utilize proteomic approaches to unravel the associated proteins for each Tau isoform, which is a crucial piece of information required to gain a global view of Tau isoform-specific interactome networks. This will set up a foundation for exploring the deeper molecular mechanisms underlying the pathogenesis of tauopathies, especially for those tauopathies that are characterized with Tau isoform-specific deposits in neurofibrillary tangles. To study Tau isoform specific interactomes, BioID2-based quantitative proteomics is described in chapter 3 to unveil the interactome network for each of the six human Tau isoforms in mouse primary neurons. Unique proximal proteins were identified for each of the Tau species. This part of work provides important resources to aid understanding the functional discrepancies among Tau isoforms. Moreover, some of these identified Tau isoform-associated proteins may serve as drug targets for tauopathy treatments.

Chapter 2: Identification of Novel Tau Exon 10 Splicing Regulators Using RNA Antisense Purification Proteomics

2.1 Abstract

Alternative splicing is a critical mechanism contributing to transcriptome and proteome diversity in eukaryotes. Aberrant alternative splicing is linked to a wide range of human diseases, including the neurodegenerative tauopathies, which are associated with imbalances in Tau splicing variants leading to abnormal accumulation of toxic neurofibrillary tangles. Identification of regulators upstream of pathogenic splicing events is key to understanding pathogenic mechanisms driving tauopathies and other diseases. Here, we used RNA Antisense Purification by Mass Spectrometry (RAP-MS) to identify candidate splicing regulators that interact with the Tau pre-mRNA around its exon 10 region. Among the 15 identified novel protein candidates, we validated that hnRNPA2B1 and hnRNPC are required to promote Tau exon 10 inclusion, whereas hnRNPH1 supports Tau exon 10 skipping. Intriguingly, hnRNPA2B1 and hnRNPC upregulation is associated with progressive supranuclear palsy (PSP), a tauopathy with accumulation of Tau proteoforms derived from transcripts containing exon 10. Separately, hnRNPH1 expression and Tau exon 10 skipping levels correlate in the course of human brain development. In summary, RAP-MS is a powerful tool to elucidate regulatory mechanisms upstream of key alternative splicing events in normal developmental and disease states.

2.2 Introduction

The tauopathies are a heterogeneous group of brain disorders with varying prevalence and symptomatology that share the common feature of accumulation of toxic forms of the

microtubule-associated protein Tau in cytoplasmic aggregates composed of different isoforms derived from alternative pre-mRNA splicing[254]. Human Tau is encoded by the microtubule-associated protein Tau (*MAPT*) gene located on chromosome 17. As a microtubule binding protein predominantly expressed in the nervous system, Tau plays a vital role in axonal transport and neuronal function[255]. Tau can form abnormal pathological aggregates in both neuron and glial cells in sporadic and familial tauopathies [11-15]. The molecular drivers of the accumulation of abnormal forms of Tau are complex, with multiple factors contributing to the pathological process, including abnormalities in alternative splicing [17-21, 256].

At the mRNA level, Tau exon 10 alternative splicing gives rise to Tau isoforms with either 3 or 4 microtubule-binding domain repeats, termed 3R- and 4R-Tau, respectively[108] (Figure 1A). Alternative splicing of exons 2 and 3, which encodes amino acids towards the N-terminus, gives rise to additional Tau isoforms mostly found in the central nervous system (CNS), namely 0N, 1N, 2N[108]. In the neurotypical fetal human brain, a single isoform predominates, without alternatively spliced N-terminal exons or exon 10 (0N3R)[111]. In the neurotypical adult human brain, however, a combination of six major Tau isoforms are present, where the ratio between the 3R and 4R isoforms is relatively equal[113]. Aberrant alternative splicing of Tau exon 10 leads to an imbalance of Tau splicing isoforms and is strongly linked to several inherited tauopathies such as the frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17)[257]. FTDP-17 is associated with at least 39 *cis*-element mutations around the *MAPT* exon 10 region that leads to 3R-Tau or 4R-Tau enrichment in Tau aggregates[116, 258]. In other tauopathies, such as progressive supranuclear

palsy (PSP) and Pick disease (PiD), post-mortem brain autopsies reveal that Tau aggregates are preferentially enriched with 4R-Tau and 3R-Tau isoforms, respectively[259]. Importantly, most PSP cases are sporadic with little connection to genetic causes including mutations in the *MAPT* locus[133]. These findings suggest that *trans*-acting factors may play important roles in aberrant Tau exon 10 splicing in sporadic PSP and other 3R- and 4R-tauopathies. While many RNA binding proteins (RBPs) were linked to Tau exon 10 splicing regulation[103, 107, 132, 136, 260-265], to date it remains uncertain how dysregulation of RBPs might contribute to the selective accumulation of 3R-Tau or 4R-Tau and disease pathogenesis. Therefore, we hypothesize that there are still unidentified splicing factors involved in mediating Tau exon 10 splicing.

Besides tauopathies, aberrant splicing caused by mutations in *cis*-elements or dysregulation of *trans*-acting RBPs has been implicated in a broad range of human diseases including cancer[266-268]. To study molecular mechanisms underlying alternative splicing control, genome-wide and exome-wide high-throughput sequencing technologies have been used to systematically identify *cis*-element mutations underlying pathogenic aberrant splicing events[269, 270]. When it comes to identification of *trans*-acting splicing factors directing disease-related splicing events, however, most studies to date have been based on known splicing factor binding motifs[271], which are not readily available for many potential splicing regulators. In the past two decades, several tools were developed to isolate RNA molecules of specific sequence and to identify their associated proteins. However, these methods all have their own limitations. RNA-binding protein enrichment from cell lysates using *in vitro* transcribed RNA bait could lead to non-physiologic protein-RNA localization

and association[272]. Tag-based RNA affinity purification, such as engineered chimeric RNA with aptamers or MS2 tag[203, 273, 274], requires complicated genetic engineering, while the physical interaction might also be disrupted during RNA purification. Furthermore, when these methods were used for RNA splicing studies, both direct and indirect binding proteins were detected, along with high background protein noise in the resulting proteomics data, making it difficult to identify functional splicing factors[201].

Recently, two oligo-based methods, Comprehensive identification of RBPs by mass spectrometry (ChIRP-MS)[208], and RNA antisense purification by mass spectrometry (RAP-MS)[209], were developed to identify RNA binding proteins in the context of long non-coding RNA functional regulation. Briefly, in both methods, DNA probes complimentary to the target RNA sequence are applied to capture the target RNA and its associated proteins. Physiological interactions between RNA and proteins are fixed in live cells by using UV irradiation or formaldehyde. These methods avoid post-lysis formation of RNA-protein complexes. Compared to the ChIRP-MS method, the RAP-MS method deploys longer probes (90 and above nt) to tile the entire sequence of RNA to ensure satisfactory capturing efficiency. Importantly, the use of UV irradiation fixation makes it possible for the subsequent probe-transcript hybridization step to occur at high temperature (> 67°C) to disrupt the RNA secondary structure for higher capture efficiency.

Given the promising capacity of the RAP-MS method to pull down specific RNA and its direct binding proteins in such a robust and sensitive way, in this study, we adapted and applied the RAP-MS method to detect RBPs that contribute to the Tau exon 10

alternative splicing event. Towards this end, we designed probes targeting the Tau pre-mRNA around its exon 10 region and performed RAP-MS proteomic analysis in the neuroblastoma cell line SH-SY5Y. For identified RBP candidates that interact with the Tau pre-mRNA, we validated their functional roles in Tau exon 10 splicing regulation and characterized their functional binding sites in the Tau pre-mRNA sequence.

2.3 Methods

Plasmids and antibodies

pLKO.1 puro (#8453), pCDH-EF1(#72266), pCDNA3.1-FLAG-HA (#52535) were obtained from Addgene. Mouse monoclonal anti-hnRNPC (# ab128049, 1:1000 dilution) and rabbit polyclonal anti-hnRNPH (#ab10374, 1:3000 dilution) were from Abcam. Rabbit polyclonal anti-hnRNPA2B1(#14813-1-AP, 1: 1000 dilution) was from Proteintech. Rabbit polyclonal anti-Tubulin (#2148S, 1:1000 dilution) and rabbit polyclonal anti-Actin (#4970S, 1:1000 dilution) were from Cell Signalling Technology. Mouse monoclonal anti-Flag (#F3165-2MG, 1:2000 dilution) was from Sigma.

Human tissue

Fresh frozen human post-mortem brain tissue was obtained from Neuropathology Brain Bank & Research Core (Mount Sinai, New York, NY). In all cases, brains with CNS malformation or neuropathologic evidence of hypoxic-ischemic injury were excluded. Tissue collection protocols were approved by the Mount Sinai Institutional Review Board (Protocol IRB-17-01313). All methods were carried out in accordance with the relevant guidelines, laws, and regulations.

Cell culture

HEK293FT and HeLa were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). The human neuroblastoma cell line SH-SY5Y was cultured in Ham's F12: Minimum Essential Medium Eagle(1:1), supplemented with 2mM GlutaMAX, 1% Non-Essential Amino Acids (NEAA) and 10% FBS.

Construction of Tau minigene plasmid

Tau minigene (4431 bp) was obtained by PCR amplification of three fragments from the genomic DNA of HEK-293FT cells between MAPT exons 9-11 (Figure 2C), which were then seamlessly cloned together to the pCDNA3.1-FLAG-HA vector (Addgene) in one step by using GeneArt® Seamless Cloning and Assembly Enzyme Mix (Thermo Scientific).

RT-PCR and Quantitative real-time PCR

Total RNA from cells was extracted using the RNeasy®mini kit (Qiagen) according to the manufacturer's protocol and were used for both semiquantitative and quantitative PCR. Endogenous 3R- and 4R-Tau mRNA expression were detected by RT-PCR (28 amplification cycles) with the following primers: forward5'- GCGAATTCGG TGAACCTCCAAAATCAGGGGATCG-3';reverse5'-CCCTGGTTTATGTTGCCTAAT GAG-3'. Primers for Tau minigene: forward5'-GGTGGAGGTGGAGGTTCTAGAAG TG-3'; reverse5'-ACCTGAGCAAGGTGACC TCCA-3'. RT-PCR products were detected on a 2 % agarose gel: endogenous 3R- and 4R-Tau RT-PCR products were 359 and 452 bp, respectively. Tau minigene 3R- and 4R-Tau RT-PCR products were 334 and 427 bp, respectively. Gel bands were quantified with ImageJ and Significance was tested using two-tailed Student's t test and $p < 0.05$ was considered significant. For quantitative real-time PCR (qRT-PCR), the RNA concentration was determined using a NanoDrop (Thermo

Scientific) and 2 μg RNA was reverse transcribed to cDNA using SuperScriptTM III First-Strand cDNA Synthesis. qRT-PCR was conducted with SsoAdvancedTM Universal SYBR Green Supermix (BIO-RAD). Primers used in qRT-PCR are shown in Table S1.

Western Blotting

Cells were collected and then lysed using 8M urea (Sigma)-100 mM ammonium bicarbonate (Sigma) on ice. Lysate protein concentrations were measured by micro BCA™ protein assay kit (Thermo Scientific). Same amounts of protein samples were then diluted in 1x LDS loading buffer containing 0.1 M DTT and were boiled for 10 minutes at 95°C . Boiled protein samples were then loaded to NuPAGE™ 4-12% Bis-Tris Protein Gels (Thermo Scientific). After gel electrophoresis, proteins were then transferred onto a nitrocellulose (NC) membrane (Thermo Scientific) using iBlot 2 Dry Blotting System (Thermo Scientific). Next, the above NC membrane was incubated with 5% skim milk that was prior dissolved in 1x TBST buffer for 1 hour with gentle shaking and was then incubated with primary antibodies (diluted in 5% skim milk-TBST) overnight at 4°C with gentle shaking. NC membrane was then washed with 1x TBST for 3 times (5 minutes per wash) . Then, secondary antibodies (Li-COR Biosciences) that diluted in 5% skim milk-TBST were incubated with the NC membrane for 1 hour with gentle shaking. Followed by washing the NC membrane for 3 times with 1x TBST (10 minutes per wash), the blot was imaged by using Odyssey Classic Imager (Li-COR Biosciences).

Generation of recombinant lentivirus, stable cell lines, cell transduction

PLKO.1 lentiviral vector that expresses the scramble shRNA (as the negative control) was purchased form Sigma. The sequence of shRNA oligos that target individual Tau splicing

factor candidates were selected based on Sigma's pre-validation assays with the highest target efficiencies for each target. Meanwhile the sequence specificity for each oligo was confirmed using Primer-BLAST[275]. All these checked shRNA oligos were cloned into pLKO.1 lentiviral vector under the U6 promoter. Recombinant vectors and packaging plasmids pSPAX2 and pMD2.G were transfected into HEK 293FT cells to produce recombinant lentivirus. To test the splicing profile of Tau exon 10, SH-SY5Y neuroblastoma cells were infected with lentivirus either expressing scramble-shRNA or candidates-shRNA followed by puromycin selection for 48 hrs. The sequence for shRNA oligos used in this study are shown in Table S1.

RNA Antisense Purification (RAP) Procedure

Probe design and generation: 90-mer DNA oligonucleotides spanning the target RNA with reverse complementary sequences were synthesized by Integrated DNA Technologies (IDT). Each DNA oligonucleotide probe carries a 5' biotin for streptavidin capturing. Probes that may hybridize to off-target sequences or have Tm value below 67°C were removed. RAP oligos targeting the Tau minigene and the GAPDH mature mRNA are listed in Tables S2.

UV cross linking: SH-SY5Y cells stably expressing the Tau-minigene were cultured in 150 mm dishes. Cells were UV irradiated with 600 mJ/cm^2 in ice cold PBS at 80% cell confluence (~10 million cells per dish). Cells were then scraped, washed twice with PBS, and pelleted at 800 g for 5 min at 4°C.

Total cell lysate preparation: About 240 million cells (12x 20 million) were lysed by completely resuspending frozen cell pellets in ice-cold detergent-based cell lysis buffer (10

mM Tris pH 7.5, 500 mM LiCl, 0.5% dodecyl maltoside (DDM, Sigma), 0.2% sodium dodecyl sulfate (SDS, Ambion), 0.1% sodium deoxycholate (Sigma), 1× Protease Inhibitor Cocktail (Roche) and Murine RNase Inhibitor (New England Biolabs)). Samples were then incubated for 10 minutes on ice to allow lysis to proceed. During this incubation period, each cell sample was passed 5 times through a 26-gauge needle attached to a 1 mL syringe to disrupt the pellets and shear genomic DNA. Each sample was then sonicated at 20% pulse power for 3 times in intermittent pulses (1 second on, 2 seconds off). The samples were then treated for 10 minutes at 37 °C with 2.5 mM MgCl2, 0.5 mM CaCl2, and 20 U of TURBO DNase (Ambion) to digest DNA. Samples were returned to ice and the reaction was immediately terminated by the addition of 10 mM EDTA and 5 mM EGTA. Disulfide bonds were reduced by addition of 2.5 mM Tris-(2-carboxyethyl) phosphine (TCEP) and samples were then mixed with twice the lysate volume of 1.5× LiCl/Urea Buffer (the final1 × Buffer contains 10 mM Tris pH 7.5, 5 mM EDTA, 500 mM LiCl, 0.5% DDM, 0.2% SDS, 0.1% deoxycholate, 4M urea, 2.5 mM TCEP). Lysates were incubated on ice for 10 minutes then cleared by centrifugation for 10 minutes at maximum speed. Supernatants were saved. RNA Antisense Purification of crosslinked complexes: Lysate samples from 240 million cells (half lysate used for Tau minigene pre-mRNA pull-down, the other half lysate used for GAPDH mature transcript control pull-down) were pre-cleared with 500 µl Streptavidin coated Sepharose beads (GE) at 37 °C for 30 minutes. Biotinylated 90-mer DNA oligonucleotide probes specific for the RNA target of interest (12 µg Tau probes and 12 µg GAPDH probes) were heat-denatured at 85 °C for 3 minutes and then snap-cooled on ice. Probes and pre-cleared lysate were mixed and incubated at 67 °C using Thermomixer R

(Eppendorf) with intermittent shaking (1100 rpm, with 30 seconds on and 30 seconds off) for 2 hours. Hybrids of biotinylated DNA probes and target RNA were then bound to streptavidin beads at 67 °C for 30 minutes with intermittent shaking as above. Beads with captured hybrids were washed 5 times with LiCl/Urea Hybridization Buffer at 67 °C for 5 minutes to remove non-specifically associated proteins.

Elution and analysis of RNA samples: For qPCR examination of RNA captures, beads with hybrids were centrifuged at 500g for 2 minutes at 4 °C and then the supernatant was discarded. Beads were then resuspended by pipetting in 20 µL NLS RNA Elution Buffer (20 mM Tris pH 8.0, 10 mM EDTA, 2% NLS, 2.5 mM TCEP). To release the target RNA, beads were heated for 2 minutes at 95 °C. Then eluted target RNA molecules were incubated with 1 µg Proteinase K (Thermo Scientific) for 1 hour at 55 °C to remove all proteins. The remaining nucleic acids were then purified by ethanol precipitation onto SILANE beads (Thermo Scientific). DNA probes were removed by digestion with TURBO DNase (Ambion). Further, qRT-PCR was performed to quantify RNA enrichment.

Elution and analysis of protein samples: For mass spectrometry analysis of captured protein contents, on beads trypsin digestion was performed. In brief, after oligo capturing, beads were further washed 6 times with 50 mM ammonium bicarbonate (pH=8.0). 1.2 µg trypsin/ 120 million cells were added to the beads and incubated overnight at 37 °C. 0.8 µg trypsin/120 million cells were added the next day and incubated for another 2-4 hrs. Trypsin Supernatant was collected for mass spectrometry analysis.

LC-MS/MS analysis

Peptide samples were resuspended in 0.1% Trifluoroacetic acid (TFA), loaded onto a home-made trap-column (5cm x 200 µm inner diameter; POROS 10µm 10R2 C18 resin) and a home-made analytical column (50cm x 50µm inner diameter; Reprosil-Pur 120 C18-AQ 5µm resin), and eluted with a 120min reversed-phase gradient at 70nl/min on a Thermo Fisher Ultimate 3000 RSLCNano UPLC (Ultra-Performance Liquid Chromatography) system coupled to a Thermo QExactive HF quadrupole-Orbitrap mass spectrometer. A parent ion scan was performed using a resolving power of 120,000 and then up to the 20 most intense peaks were selected for MS/MS.

Data were analyzed using Thermo Proteome Discoverer 2.2. For protein identification, search was against the SwissProt human proteome database (42,173 protein isoform entries), while the search parameters specified a parent ion mass tolerance of 10ppm, and an MS/MS fragment ion tolerance of 0.02Da, with up to two missed cleavages allowed for trypsin.

Ribonucleoprotein immunoprecipitation (RIP) analysis

RIP assay was performed in SH-SY5Y cells with hnRNPC antibody (Abcam), hnRNPH1 antibody (Abcam) or IgG control (Millipore). 10 million SH-SY5Y cells were lysed with RIP lysis buffer (50mM Tris-HCl pH 7.5, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% Sodium Deoxycholate, 1x protease inhibitor cocktail). After sonication, cell lysate was precleared with protein G beads for half an hour in 4°C. Meanwhile, 5 µg antibody was incubated with protein G beads for 1hr at room temperature. Next, precleared cell lysate was incubated with antibody coupled protein G beads for overnight at 4°C. After 9 times of wash with NET-2 buffer (5mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1% TritonX-100),

RNA was isolated with Trizol-LS (Thermo Scientific) and was then reverse transcribed and measured by qRT-PCR, with 18S rRNA as the internal control. Primers are shown in Table S1.

***In vitro* transcription**

DNA fragments were *in vitro* transcribed into RNA using the MEGAscript T7 transcription kit (Thermo Scientific) incorporated with biotinylated-CTP (Thermo Scientific) at a molar ratio of 1:1 with non-biotinylated CTP. RNA products were visualized in 6% or 10% TBE-Urea gel (Thermo Scientific) using SYBR Gold (Thermo Scientific).

***In vitro* RNA Pull-down**

For each reaction, 10 pmol of each *in vitro* transcribed RNA and 1.5 mg of SH-SY5Y cell lysate (in RIP lysis buffer) were used. The transcribed RNA was first incubated at 65°C for 10 mins, then cooling down slowly to 4°C. The pre-cleared cell lysate was then incubated with the RNA at 4°C for 4 hrs. After streptavidin beads capture and 4 washes with NET-2 buffer, proteins were eluted with the 1x LDS loading buffer and were probed with specific antibody by western blot. Samples with beads-only and human GAPDH (Rat Gapdh for hnRNPH1[201]) mRNA 3'UTR fragment pull-down were used as negative controls.

RNA sequencing

RNA sequencing for post-mortem human fetal and adult brain samples was carried out as previously described[276]. Briefly, total RNA was isolated from flash frozen cortex using a QIAsymphony automated RNA extraction robot (Qiagen), reverse transcribed with a high-capacity cDNA reverse transcription kit with RNase inhibitor (Life Technologies) and rRNA depleted with a Ribozero rRNA Removal Kit (human/mouse/rat probe) (Illumina),

followed by library preparation with NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was performed in 2×150 paired end configuration with an Illumina HiSeq instrument. Alignment to the hg38 human reference genome was carried out with STAR with quantification by featureCounts[277, 278].

PSP RNA-seq dataset was obtained from Mayo clinic [279](Synapse:syn5049298). The tissues were collected from cerebellum region of 84 PSP and 77 controls without neurodegeneration diagnoses.

Raw counts from two RNA-seq cohorts described above were reprocessed and analysed by limma[280], an R/Bioconductor software package, to identify differentially expressed genes in R studio. P-value was adjusted with Benjamini–Hochberg method. A gene was considered significant after Benjamini–Hochberg false discovery rate of 0.05.

2.4 Results

RAP-MS screen identifies novel Tau exon 10 splicing regulators

We first designed biotinylated DNA probes targeting U1 snRNA and GAPDH mRNA transcripts in HEK 293FT cells to perform the RAP RNA pull-down workflow (Figure S1A). As shown in Figure S1B-D, we confirmed the effectiveness and specificity of targeted RNA enrichment using the RAP method. Next, we worked on adapting the RAP-MS method to identify RBPs upstream of Tau exon 10 splicing in the SH-SY5Y neuroblastoma cells. Towards this end, we first confirmed that the SH-SY5Y cells express both 3R-Tau (skipping the exon 10) and 4R-Tau (including the exon 10) isoforms (Figure S2). We optimized the UV irradiation crosslinking dose at 600 mJ/cm2 in SH-SY5Y cells (Figure S3). To identify the regulators that selectively bind to the Tau pre-mRNA transcript

around its exon 10 region, we generated a Tau minigene cassette that includes full-length exons 9, 10, 11, flanked by the partial sequences (first and last ~1kb each) of intron 9 and intron 10 (~ 4.5 kb fragment in total) (Figure 1B). The exon 10 alternative splicing profile of the Tau minigene transcript is similar to that of the endogenous Tau transcript in the SH-SY5Y cells (Figure S4). Once we generated an SH-SY5Y cell line that stably expressed the Tau minigene, we found the Tau minigene pre-mRNA level was ~80 fold higher than the endogenous Tau pre-mRNA level (Figure S5), which would benefit the sensitivity of our RAP-MS experiment.

In our initial experiment, to capture the Tau minigene pre-mRNA, we used thirty-one 90-mer biotinylated oligo probes (Figure 1C, Table S2) that tile the entire length of the Tau minigene sequence (non-specific and exon-targeting probes were removed). With these probes, Tau minigene pre-mRNA, but not the non-specific U1 snRNA, was highly enriched using the RAP procedure (Figure 1D-E). We then performed RAP-MS experiment (Figure 1F) with ~240 million SH-SY5Y cells expressing the Tau minigene using oligo sets that target either the Tau minigene pre-mRNA or GAPDH mature mRNA (negative control). From two RAP-MS runs with the Tau-targeting probes, we identified 66 RBPs, which included 12 Tau exon 10 splicing regulators that were previously reported[132, 200, 204, 262, 263, 281, 282](Table S3). When we compared these results to proteins that were also identified in the control RAP-MS runs using GAPDH-targeting probes, we narrowed our results down to 16 RBPs that specifically interact with the Tau minigene pre-mRNA (Figure 1G). Among these RBP candidates, only one (SFPQ) was previously reported to regulate Tau exon 10 splicing[263].

Loss of function assays validate hnRNPC, hnRNPA2B1 and hnRNPH1 as Tau exon 10 splicing regulators

To test the function of the 15 newly identified RBPs in Tau exon 10 splicing regulation, we knocked down each candidate gene using short-hairpin RNAs (shRNAs) in wild-type SH-SY5Y cells (Figure 2A). For 9 of these 15 proteins, namely LUC7L3, THRAP3, EMG1, DDX21, hnRNPD, hnRNPU, hnRNPH1, hnRNPC and hnRNPA2B1, their individual knockdown significantly ($p < 0.05$) impacted Tau exon 10 splicing (Figure 2B-2C). We then knocked down the top 4 RBP hits (each with $p < 0.01$) in SH-SY5Y cells expressing the Tau minigene (Figure 2D) and demonstrated that, among these 4 RBP candidates, hnRNPC and hnRNPA2B1 knockdown significantly promoted Tau exon 10 skipping, whereas hnRNPH1 knockdown significantly promoted Tau exon 10 inclusion (Figures 2E-2F).

HnRNPC and HnRNPA2B1 interact with Tau pre-mRNA and promote exon 10 inclusion

To confirm the functional role of hnRNPC in Tau exon 10 splicing regulation, we induced hnRNPC knockdown using two independent shRNAs in SH-SY5Y cells and found that 4R/3R-Tau ratio was significantly increased (Figure 3A-C). Further, we demonstrated that overexpression of hnRNPC can promote inclusion of Tau exon 10 in the minigene in both HeLa cells and HEK 293FT cells (Figure 3D-E, S6). Given earlier evidence generated by our RAP-MS experiments showing that hnRNPC associated with Tau pre-mRNA, we performed Ribonucleoprotein immunoprecipitation (RIP) assay to validate this protein-RNA interaction. After we immunoprecipitated hnRNPC in SH-SY5Y cell extracts, we

measured Tau pre-mRNA using three sets of primers (Figure 3F). The interaction was quantified by the enrichment of Tau pre-mRNA in hnRNPC immunoprecipitation (IP) samples compared to immunoglobulin G (IgG) IP controls (Figure 3F). To find out the binding regions of the hnRNPC protein around Tau exon 10, we performed *in vitro* RNA fragment pull-down in SH-SY5Y cell extracts with the seven *in vitro* synthesized RNA fragments (~700 nt with 150 nt overlap with each other) from the Tau minigene sequence (Figure 3G). Pull-down analysis showed that hnRNPC was associated with the E10-1, E11-1, and E11-2 fragments (Figure 3G).

Similar to our assays for hnRNPC, we induced hnRNPA2B1 knockdown using two independent shRNAs in wild-type SH-SY5Y cells and found a consistent reduction of 4R/3R-Tau ratio upon hnRNPA2B1 knockdown (Figure 4A-C). In addition, we found that hnRNPA2B1 overexpression can promote Tau exon 10 inclusion in the minigene and in both HeLa cells and HEK 293FT cells (Figure 4D-E, S7). We next performed RIP assay and demonstrated that hnRNPA2B1 binds to the Tau pre-mRNA around its exon 10 (Figure 4F). To further map the binding region of hnRNPA2B1 protein on Tau pre-mRNA, we performed *in vitro* RNA pull-down and found that hnRNPA2B1 was associated with the E10-1 fragment (Figure 4G).

Since we demonstrated that hnRNPC and hnRNPA2B1 both promote Tau exon 10 inclusion, we decided to look into their potential clinical association with sporadic 4R-tauopathies such as PSP, to which no known genetic cause has been linked. Towards this end, through analysing the RNA-seq data of post-mortem PSP patient brain samples released by Mayo Clinic[279] (Synapse:syn5049298), we found that both hnRNPA2B1 and

hnRNPC are significantly higher expressed (adjusted $p < 0.001$) in PSP patient post-morterm brain samples versus sex- and age-matched controls (Figure S8A-C), which suggests that elevated hnRNPA2B1 and hnRNPC levels associate with 4R-Tau aggregation in PSP.

HnRNPH1 interacts with Tau pre-mRNA and promotes exon 10 skipping

Next, we confirmed the functional role of hnRNPH1 in Tau exon 10 splicing. Using two independent shRNAs targeting hnRNPH1 in wild-type SH-SY5Y cells (Figure 5A), hnRNPC knockdown significantly increased 4R/3R-Tau ratio (Figure 5B-C). Further, we found that hnRNPH1 overexpression can promote Tau exon 10 exclusion in the minigene in HeLa cells (Figure 5D-E). Next, we performed RIP assay and found that hnRNPH1 can physically bind to Tau pre-mRNA near exon 10 and exon 11 region in SH-SY5Y cells (Figure 5F). To further investigate the binding region of hnRNPH1, we performed *in vitro* RNA fragment pull-down. Our results showed that hnRNPH1 has multiple binding sites within the Tau minigene pre-mRNA both before and after the Tau exon 10 (Figure 5G).

Intriguingly, as a 3R-Tau splicing factor, hnRNPH1 transcript level decreases between fetal and adult post-mortem brain samples in our RNA-seq studies (Figure S9A), which is in strong correlation with the increased 4R/3R-Tau ratio along the course of human brain development[38]. We further confirmed that hnRNPH1 protein level is also reduced between fetal and adult human post-mortem brain samples (Figure S9B).

HnRNPC promotes Tau exon 10 splicing through direct binding to U-tracts around exon 10

Given the binding specificity of hnRNPC and hnRNPA2B1 shown in our Tau RNA fragments pull-down experiments (Figure 3G and Figure 4G), we decided to further investigate the binding sites in Tau pre-mRNA of these two RBPs. As reported in previous studies, hnRNPC has a conserved binding motif with a poly-uridine sequence or U-tract[283]. Our pull-down analysis showed that hnRNPC was associated with the E10-1 and E11-2 fragments (Figure 3G), both of which have multiple U-tracts binding sites. We also noticed that hnRNPC weakly associated with the E11-1 fragment. However, when we mutated the two U-tracts that overlapped between E11-1 and E11-2 by replacing the middle uridine with cytosine, the binding between hnRNPC and E11-1 was abolished (Figure S10), suggesting that this weak interaction is resulted from the U-tracts overlapping between E11-1 and E11-2. To further study the functional binding sites for hnRNPC around the Tau exon 10 region, we mutated all U-tracts found in the E10-1 and E11-2 fragments separately (Figure 6A), through replacing the middle uridine with cytosine as described above. As a result, we found that the binding affinity of hnRNPC to either of these two fragments was totally abolished with the mutations (Figure 6B-C). We further made constructs of Tau minigene with mutated U-tracts in the E10-1, E11-2, or both regions. We then co-transfected these mutated Tau minigene constructs with hnRNPC in HEK 293FT cells. As shown (Figure 6D), overexpression of hnRNPC with wild-type Tau minigene in 293FT cells significantly promoted Tau minigene exon 10 inclusion. In comparison, the impact of hnRNPC overexpression on Tau minigene exon 10 inclusion was fully blocked in Tau minigene with U-tract mutations in the E10-1 and E11-2 regions. These results demonstrate that U-tracts within the E10-1 and E11-2 regions can be recognized and utilized by

hnRNPC to promote Tau exon 10 inclusion. To further narrow down the functional binding sites for hnRNPC in the E11-2 region, we divided E11-2 into four sub-fragments (150 nt with 20-bp overlap with each other) as shown (Figure 6E). Next, we performed *in vitro* RNA fragment pull-down and found that hnRNPC was predominantly associated with the first two out of four sub-fragments within the E11-2 region (Figure 6F). Through combined mutation for U-tracts within these two sub-fragments, we found that the first four U-tracts are necessary for interaction with hnRNPC (Figure 6G-H). As the binding of hnRNPC to E10-1 was weak, we were not able to further narrow down hnRNPC binding sites within this region through *in vitro* RNA fragment pull-down (data not shown). We then mutated the first 4 U-tracts in the E11-2 fragment and showed that the hnRNPC binding affinity was mostly abolished, which is similar to the phenotype in full U-tracts mutant pull-down of the E11-2 fragment (Figure 6I). Further, we showed that Tau minigene exon 10 inclusion promoted by hnRNPC overexpression was largely abolished with the 5 U-tracts in the E10-1 region and the first 4 U-tracts in the E11-2 region all mutated in the Tau minigene, mostly recapitulating the phenotype of the Tau minigene with all E10-1 and E11-2 U-tracts mutated (Figure 6J). Hence, the first 4 U-tracts in the E11-2 region together with U-tracts in E10-1 play a crucial role for hnRNPC to promote Tau exon 10 inclusion.

Separately, hnRNPA2B1 was shown to have a strong association with E10-1 (Figure 4G). Within the E10-1 sequence, we found three UAG[A/G] sequences (Figure S11A), which is the predicted binding motif for hnRNPA2B1[284]. When we performed further *in vitro* RNA fragment pull-down within the E10-1 sequence, we found that hnRNPA2B1 was only enriched in the first fragment that contains the first UAGG binding

motif for hnRNPA2B1 (Figure S11B). We mutated the UAGG sequence within the first fragment of E10-1, which abolished the association between hnRNPA2B1 to the E10-1 fragment (Figure S11C). However, Tau minigene exon 10 inclusion promoted by hnRNPA2B1 overexpression was not disrupted with this mutation incorporated in the E10-1 region (Figure S11D), suggesting that this binding site is not sufficient for hnRNPA2B1 to promote Tau exon 10 inclusion, or hnRNPA2B1 might function through alternative indirect mechanisms to modulate alternative splicing of Tau exon 10.

2.5 Discussion

Previous studies have identified several splicing regulators that affect Tau exon 10 inclusion or exclusion. Besides serine- and arginine-rich (SR) family proteins [107], other RNA binding proteins have also been shown to modulate exon 10 splicing, such as RBM4[261], hnRNPG[285], SWAP[285], Tra2β[130], hTRA2-beta1[285], CELF3[286] and SFPQ[263]. However, there is no evidence to functionally link these reported factors to sporadic 3R- and 4R-tauopathies, in which dysregulated splicing factors may contribute to aberrant Tau exon 10 splicing and subsequent Tau aggregation. Most of the reported Tau exon 10 splicing factors were identified based on their known binding motifs in the Tau pre-mRNA around its exon 10 region as in the cases of SRSF2[132] and Tra2β[130], or gain-of-function overexpression screening for known splicing factors along with a Tau-splicing GFP-reporter, as demonstrated for RBM4[261] and SRSF11[262]. Through our RAP-MS experiments, we were able to capture the RNA-protein interactions *in situ* and use unbiased proteomics to systematically discover a group of novel RBP candidates that directly interact with the Tau pre-mRNA around its exon 10. Among these newly identified hits, we further validated

hnRNPA2B1 and hnRNPC as Tau exon 10 splicing factors that promote 4R-Tau expression, while hnRNPH1 promotes 3R-Tau expression (Figure 7).

HnRNPC is well known to regulate RNA processing and pre-mRNA splicing[287]. Previous studies reported that hnRNPC might play a role in cancer progression, such as breast cancer and colon cancer[288-290]. In our study, we found that hnRNPC physically interacts with Tau pre-mRNA around its exon 10 through the predicted U-tract binding motif for hnRNPC[283]. Several regulatory elements including exonic and intronic enhancers and silencers have been characterized within exon 10 and 5' splice site (5'ss) present at the exon 10-intron 10 junction[123]. Previous work also predicted that the exon-intron interface at the 3'end of exon 10 forms a stem loop structure and plays an essential role in Tau exon 10 splicing[133]. Many FTDP-related mutations were identified in this stem loop region, which may destabilize the specific structure for splicing factor binding[258]. Conversely, in our study, we found that the hnRNPC-binding U-tracts around the 5'-end of both exons 10 and 11, rather than inside of exon 10 or the aforementioned stem loop region at the 3'-end of exon 10, are more important for hnRNPC to direct Tau exon 10 inclusion (Figure 7). Our findings provide insights for novel molecular mechanisms that direct Tau exon 10 splicing. Interestingly, we demonstrated that hnRNPC expression level is associated with the 4R-tauopathy PSP, which suggests that hnRNPC elevation may play a role in PSP pathogenesis.

HnRNPA2B1 is a splicing factor implicated in cancer development and progression, including breast cancer, lung cancer, pancreatic cancer, glioblastoma, and gastric adenocarcinoma[291-296]. It was also reported that hnRNPA2B1 is associated with

amyotrophic lateral sclerosis (ALS) through regulating neurodegeneration-related RNA network [297]. In our study, we demonstrated that hnRNPA2B1 is both required and sufficient to promote 4R-Tau expression. From previous work, the RBP TAR DNA-binding protein (TDP-43), which interacts with hnRNPA2B1[298], was also reported to promote Tau exon 10 inclusion and 4R-Tau expression[136]. Therefore, future work testing whether hnRNPA2B1 and TDP-43 coordinate with each other to functionally regulate Tau exon 10 splicing may help understand the molecular mechanisms underlying tauopathy pathogenesis. Interestingly, as in the case of hnRNPC, we discovered that hnRNPA2B1 is abnormally upregulated in brain samples of PSP patients, suggesting its link to PSP pathogenesis. Intriguingly, re-balancing the Tau isoform ratio can reduce phenotypic abnormalities in the mouse model of disease[263]. Therefore, further studies are needed to confirm the causal link between the upregulation of hnRNPC/hnRNPA2B1 and PSP pathogenesis, as targeting hnRNPC and hnRNPA2B1 might potentially serve as a novel way to treat 4R-tauopathies like PSP.

HnRNPH1 is developmentally regulated and characterized as an important splicing factor in the central nervous system. Through a genome-wide transcriptomic analysis and related research, hnRNPH1 was identified to regulate splicing events and gene expression in oligodendrocytes[299-301], and it is functional in neuronal differentiation[201]. In our study, we showed that hnRNPH1 is required to promote 3R-Tau expression. Importantly, we found that hnRNPH1 is remarkably decreased at both RNA and protein levels in adult human brain versus fetal brain, which tightly correlates with the expression of 3R-Tau

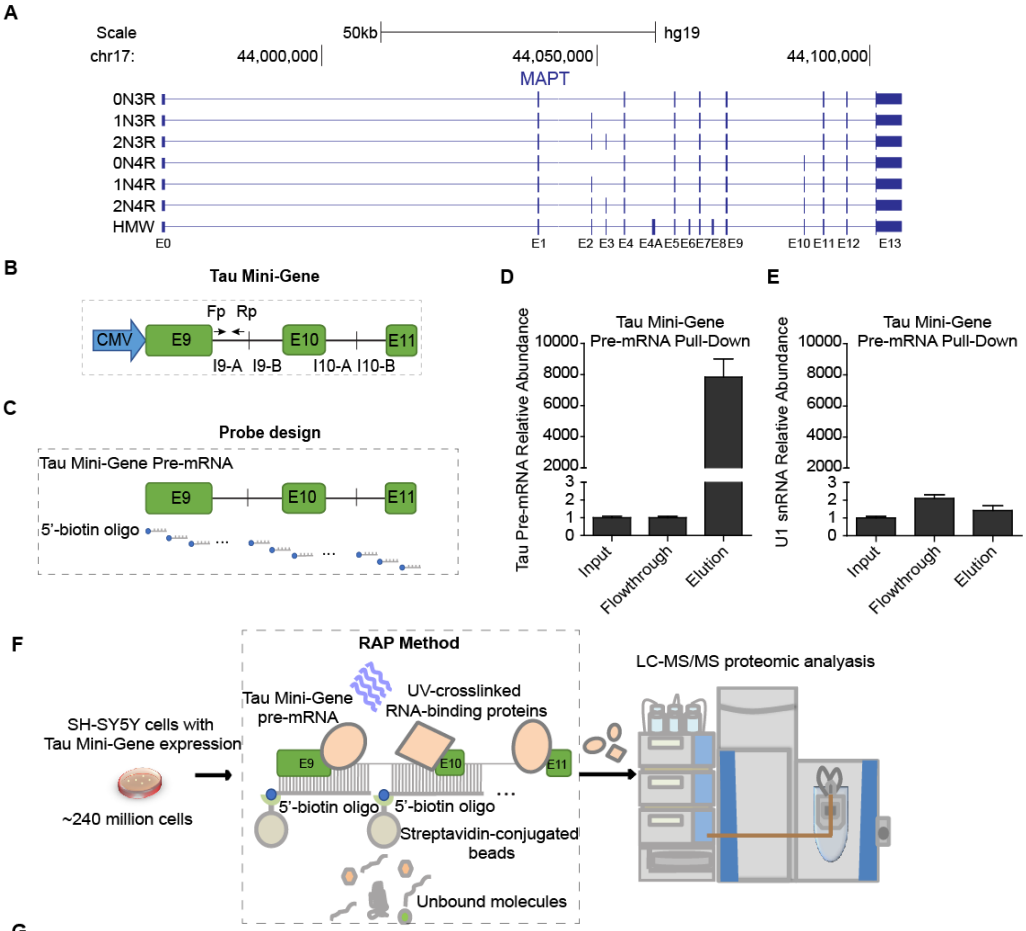
through the brain development course[111]. Our findings further highlight the multiple functional roles that hnRNPH1 plays in the central nervous system.

Using the RAP-MS method, we identified a total of 66 RBPs that bind to the Tau pre-mRNA around its exon 10, covering 12 previously reported Tau exon 10 splicing regulators. Among the 15 novel RBP candidates that we found to uniquely interact with Tau pre-mRNA, 9 RBPs significantly impact Tau exon 10 splicing regulation. All these results demonstrated the effectiveness of applying the RAP-MS method to discover RBPs mediating specific alternative splicing events. Compared to most RNA splicing regulator studies that depend on known binding motifs of these proteins[271], the RAP-MS method allows for rapid narrowing-down and identification of novel regulators upstream of specific splicing events. It is also a much more directed and efficient way to identify splicing regulators upstream of specific splicing events than large-scale RBP eCLIP-seq analysis and loss-of-function screens that are currently available to the splicing research community.

To further improve the detection sensitivity and reproducibility of the RAP-MS method, one option is to use tandem mass tags (TMT) to accommodate multiple RAP pull-down samples within the same LC-MS run. Separately, in this study we needed to deliver the Tau minigene to identify regulators that specifically mediate Tau exon 10 splicing, which is a limitation of our current method. Although we were able to efficiently capture endogenous Tau pre-mRNA in addition to the Tau minigene pre-mRNA (Figure S12), to directly identify RBPs that interact with the endogenous Tau pre-mRNA around its exon 10, we will need to develop a new procedure to selectively isolate and enrich for the Tau pre-mRNA sequence in that region.

In summary, with our successful identification and validation of novel Tau exon 10 splicing regulators, the RAP-MS method provides high potential as a useful tool to unveil unique splicing factors that mediate aberrant splicing events clinically linked with human diseases such as cancer and neurodegeneration.

2.6 Main figures



G

Protein name	Tau_1	Tau_2	GAPDH_1	GAPDH_2
	# of peptide			
LUC7 like 3 pre-mRNA splicing factor(LUC7L3)	2	3	0	2
thyroid hormone receptor associated protein 3(THRAP3)	4	3	0	1
EMG1, N1-specific pseudouridine methyltransferase(EMG1)	1	1	0	1
pre-mRNA processing factor 19(PRPF19)	2	2	0	2
ADP ribosylation factor like GTPase 6 interacting protein 4(ARL6IP4)	1	1	0	1
DEAH-box helicase 15(DHX15)	2	3	0	2
DExD-box helicase 21(DDX21)	4	3	0	3
heterogeneous nuclear ribonucleoprotein U(HNRNPU)	10	3	1	4
heterogeneous nuclear ribonucleoprotein D(HNRNPD)	2	1	0	1
heterogeneous nuclear ribonucleoprotein K(HNRNPK)	7	3	0	0
heterogeneous nuclear ribonucleoprotein A3(HNRNPA3)	1	1	0	0
heterogeneous nuclear ribonucleoprotein H1 (H)(HNRNPH1)	7	0	0	0
heterogeneous nuclear ribonucleoprotein C (C1/C2)(HNRNPC)	1	1	0	1
heterogeneous nuclear ribonucleoprotein R(HNRNPR)	1	3	0	0
heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	2	10	1	4

Figure 1: Identification of proteins binding to the Tau minigene pre-mRNA using the RAP-MS method. (A) Alternative splicing of exon10 (E10) gives rise to Tau with either

3 or 4 microtubule-binding domain repeats, termed 3R and 4R, respectively. Alternative splicing of exons 2 and 3 (E2 and E3) towards the N-terminus gives rise to additional Tau isoforms (HMW) in peripheral tissues but are found at low frequency in the central nervous system (CNS). Chromosome map adapted from UCSC Genome Brower. (B) Construct of Tau minigene expressed in the neuroblastoma cell lines. I9-A: +1 to +990 nt in intron 9; I9-B: +12645 to +13645 nt in intron 9; I10-A: +1 to +1000 nt in intron10; I10-B: +2840 to +3840 nt in intron 10. (C) 5' biotin labelled 90 nt DNA oligos that are complimentary to Tau minigene pre-mRNA sequence were designed as shown in the diagram. (D) The Tau minigene pre-mRNA abundance through RAP process was measured by qRT-PCR using the primers Fp and Rp as shown in Figure 1B. (E) The U1 snRNA abundance was measured by qRT-PCR with the same samples as (D). GAPDH was used as the internal standard in both (D) and (E). (F) Identification of Tau exon 10 splicing regulator candidates by performing the RAP enrichment procedure in SH-SY5Y cells that stably express the Tau minigene cassette, followed by proteomics. (G) Identified RBPs that are selectively enriched with Tau minigene probes (Tau_1 and Tau_2 are biological replicates), along with the number of peptides found by mass spec analysis, are listed. RAP-MS runs using biotinylated probes that target GAPDH mature transcript were included as controls.

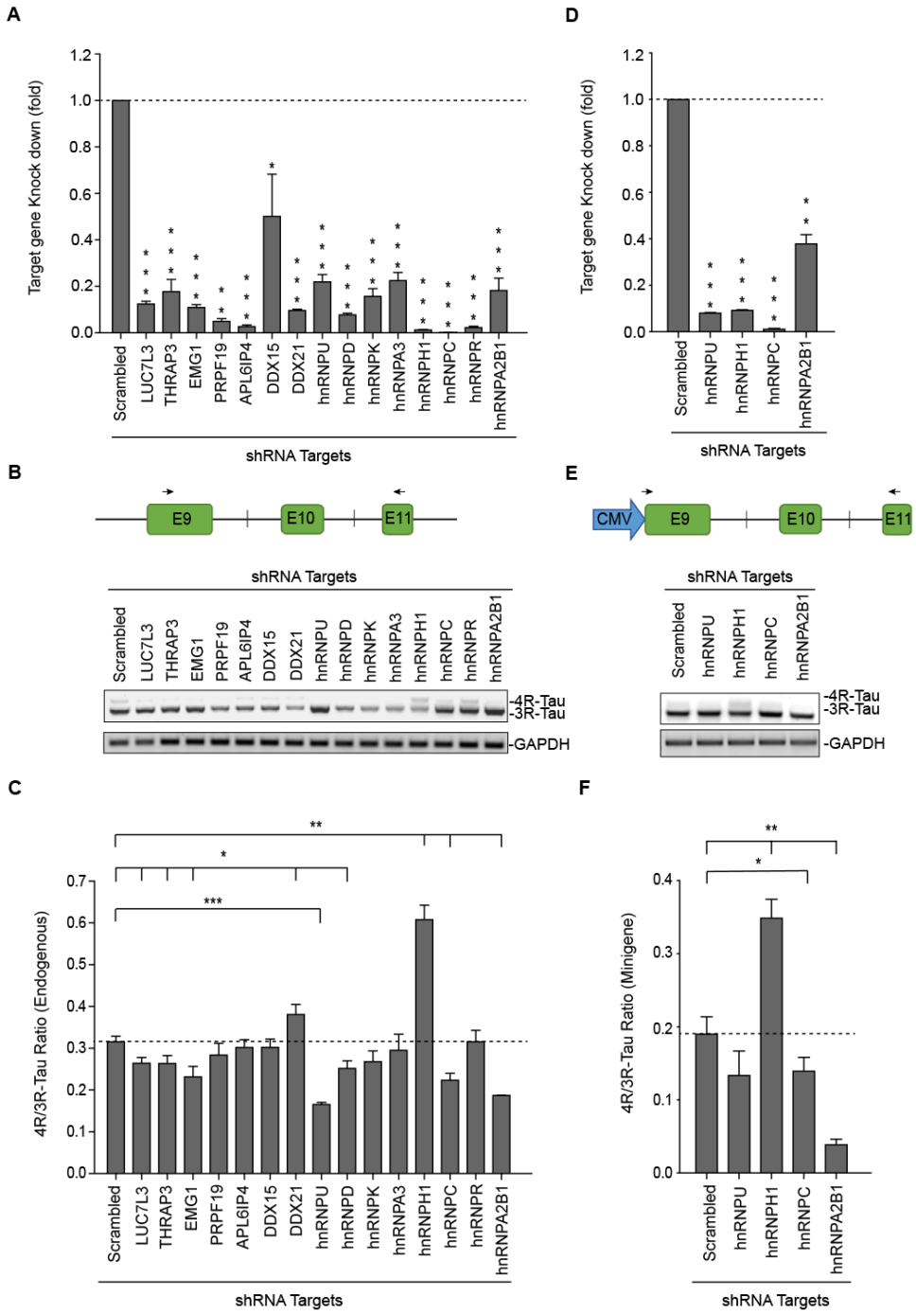


Figure 2: Loss-of-function assays to identify RBPs that control Tau exon 10 splicing. (A) RBP knockdown was confirmed by the remaining RBP mRNA levels in the RBP-targeting shRNA infection groups compared with the scrambled shRNA infection group in SH-SY5Y cells, as assessed by qRT-PCR. (B-C) Agarose gels showing RT-PCR products

from the endogenous Tau exon 10 splicing, GAPDH was used as internal standard; each lane represents a different cell population in which a given RBP identified by mass spectrometry was knocked down using shRNA; control cells transduced with scrambled shRNA were included. Arrows, primers used to amplify the spliced segment (B). The levels of ratio (4R-Tau relative to 3R-Tau bands) from the endogenous MAPT gene were calculated (C). (D-F) RBPs showing significant role ($p < 0.01$) in regulating 4R/3R-Tau ratio in (A-C) were knocked down in SH-SY5Y cells stably expressing the Tau minigene cassette. RBP knockdown efficiencies were confirmed by qRT-PCR (D). (E) Agarose gels showing RT-PCR products from the Tau minigene pre-mRNA; GAPDH was used as the internal standard. Arrows, primers used to amplify the spliced segment in the minigene pre-mRNA. The ratios (4R-Tau relative to 3R-Tau bands) from the Tau minigene splicing products were calculated (F). Data shown represent the means ± SEM of three independent experiments (*$p < 0.05$, **$p < 0.01$, **$p < 0.001$; two-tailed Student's t test).

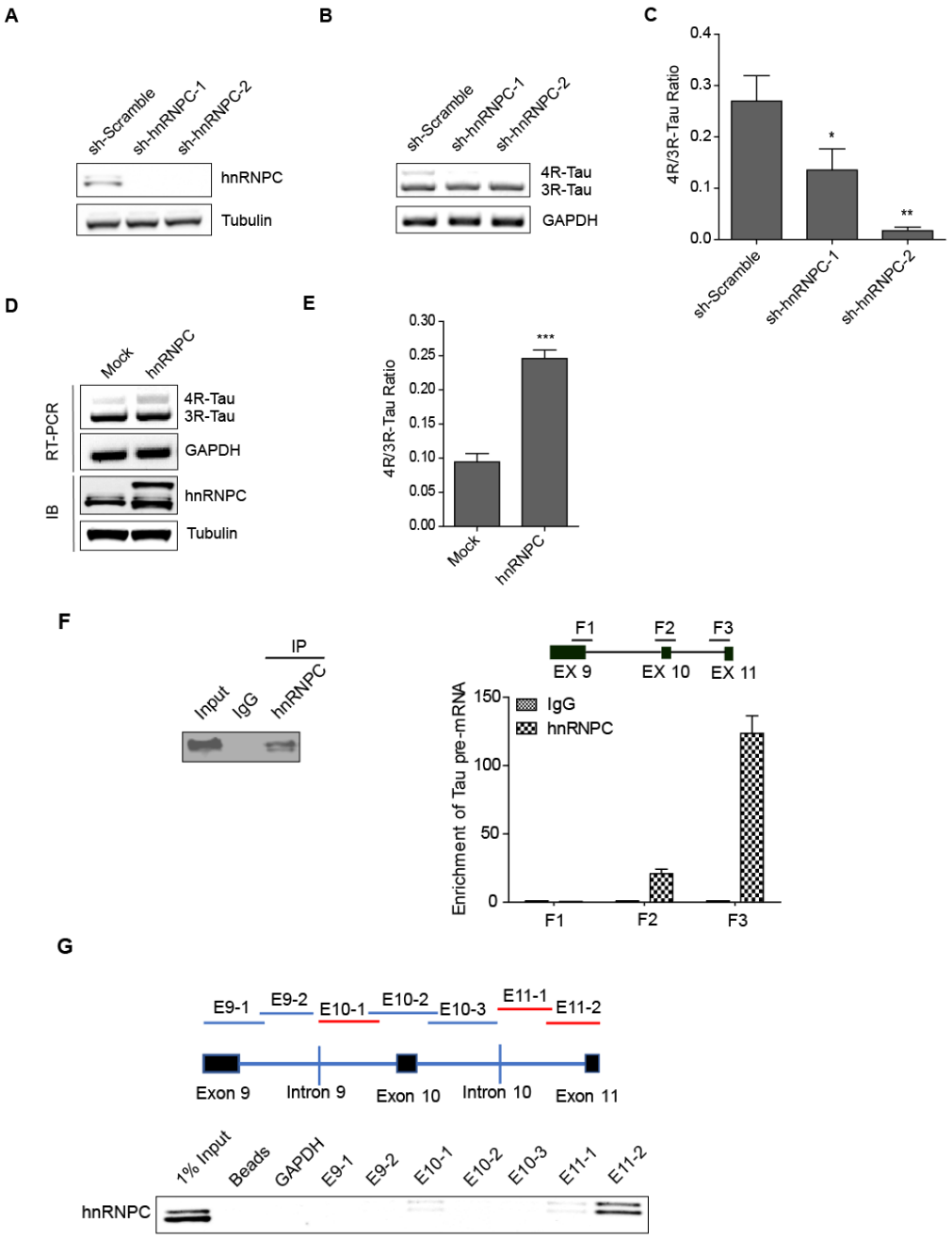


Figure 3: hnRNPC interacts with Tau pre-mRNA and promotes Tau exon 10 inclusion. (A) SH-SY5Y cells were transduced with lentivirus expressing two independent shRNAs targeting hnRNPC. Knockdown of hnRNPC was confirmed by western blot analysis, with the Tubulin level as the loading control. (B-C) RT-PCR products from the endogenous Tau exon 10 splicing after hnRNPC knockdown (B). The 4R/3R-Tau ratios were calculated (C).

(D-E) Overexpression of hnRNPC promotes Tau minigene exon 10 inclusion in HeLa cells (D). The 4R/3R-Tau ratios from the Tau minigene were calculated (E). Data shown represent the means ± SEM of three independent experiments (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; two-tailed Student's t test). (F) RIP analysis to measure the enrichment of Tau pre-mRNA in hnRNPC immunoprecipitation (IP) relative to IgG IP as measured by qRT-PCR using the primers target F1, F2 and F3 in SH-SY5Y cells, 18s rRNA was used as internal control. (G) hnRNPC interacts with Tau pre-mRNA at the boundary between Intron 10 and Exon 11 and Intron 9 flanking Exon 10, as detected by *in vitro* RNA/protein interaction assay.

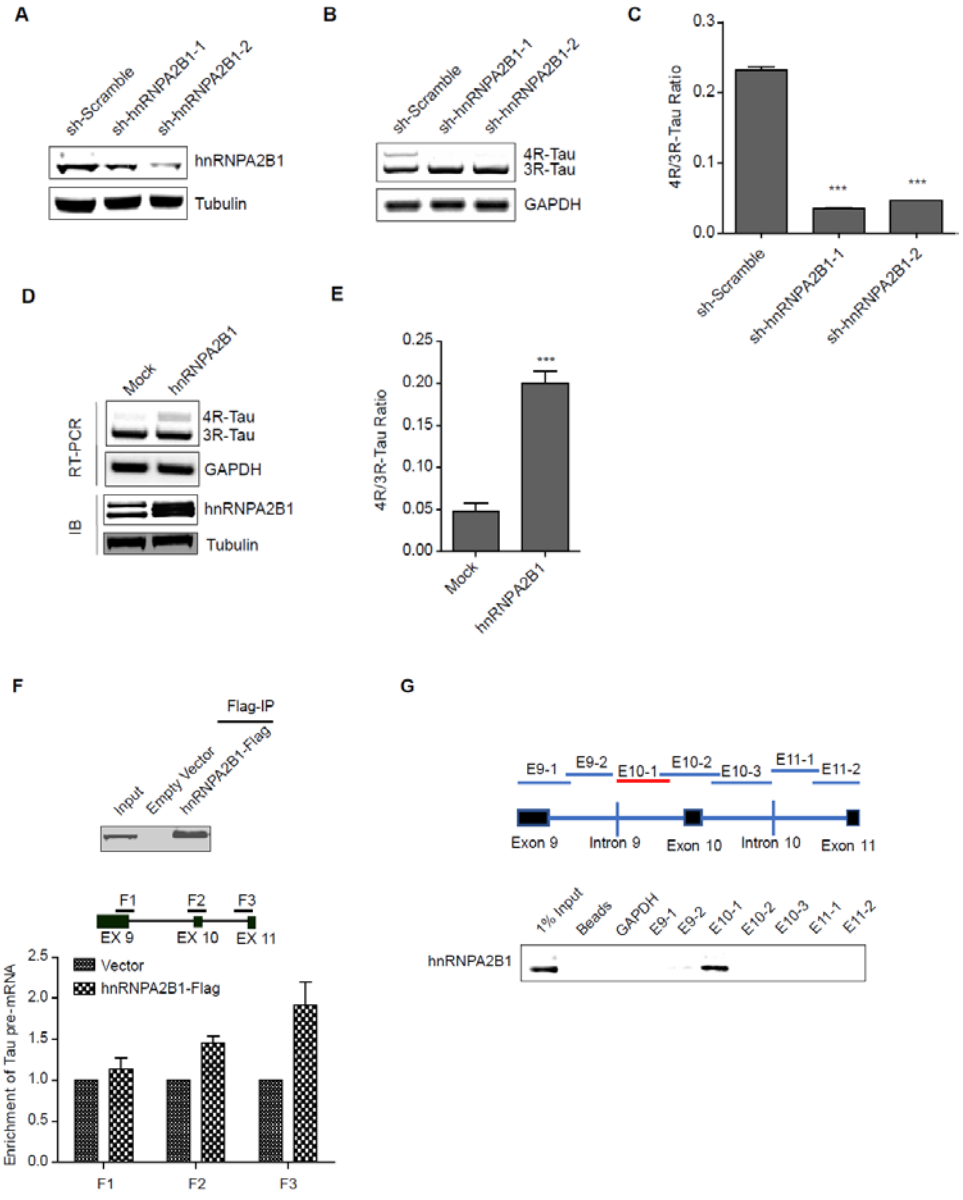


Figure 4: hnRNPA2B1 interacts with Tau pre-mRNA and promotes Tau exon 10 inclusion. (A) Confirmation of hnRNPA2B1 knockdown was done by western blot, with the Tubulin level as the loading control. (B-C) RT-PCR products from the endogenous Tau exon 10 splicing after hnRNPA2B1 knockdown (B). The 4R/3R-Tau ratios were calculated (C). (D-E) Overexpression of hnRNPA2B1 promotes Tau minigene exon 10 inclusion in HeLa cells (D). The 4R/3R-Tau ratios from the Tau minigene were calculated (E). Data shown represent the means ± SEM of three independent experiments (***$p < 0.001$; two-tailed Student's t test). (F) Exogenously expressed

hnRNPA2B1 fused with Flag tag in SH-SY5Y cells was pulled down through IP. The enrichment of Tau pre-mRNA was measured by qRT-PCR. (G) hnRNPA2B1interacts with Tau pre-mRNA at the region near exon 10, as detected by *in vitro* RNA/protein interaction assay.

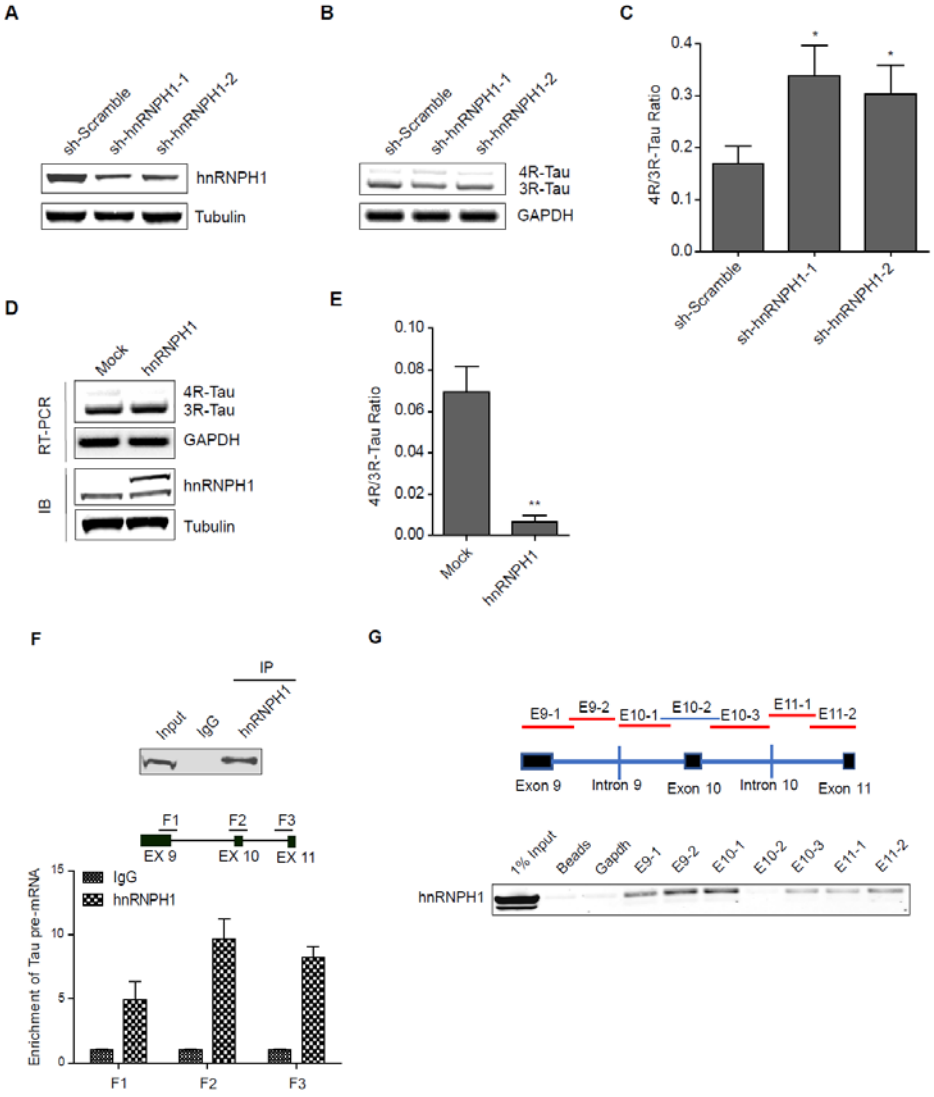


Figure 5: hnRNPH1 interacts with Tau pre-mRNA and promotes Tau exon 10 exclusion. (A) Confirmation of hnRNPH1 knockdown was done by western blot, with the Tubulin level as the loading control. (B-C) RT-PCR products from the endogenous Tau exon 10 splicing after hnRNPH1 knockdown (B). The 4R/3R-Tau ratios were calculated (C). (D-E) Overexpression of hnRNPH1 promotes Tau minigene exon 10 exclusion in HeLa cells (D). The 4R/3R-Tau ratios from the Tau minigene were calculated (E). Data

shown represent the means ± SEM of three independent experiments (*$p < 0.05$, **$p < 0.01$; two-tailed Student's t test). (F) RIP analysis to measure the enrichment of Tau pre-mRNA in hnRNPH1 IP relative to IgG IP as measured by qRT-PCR in SH-SY5Y cells. (G) hnRNPH1 interacts with Tau pre-mRNA at multiple regions near exons 9,10,11, as detected by *in vitro* RNA/protein interaction assay.

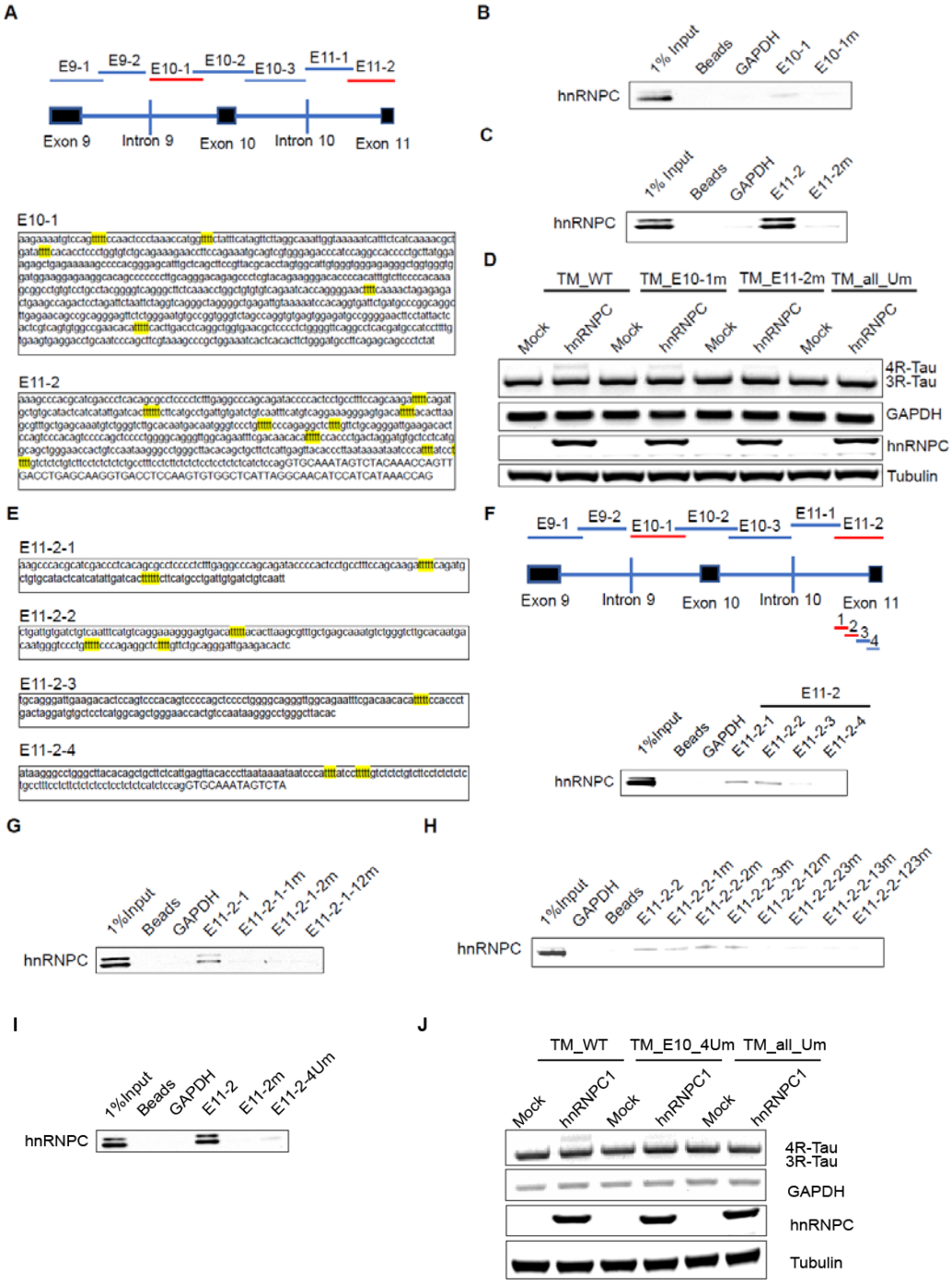


Figure 6: hnRNPC promotes Tau exon 10 inclusion through direct binding to U-rich tracts near Tau exons 10 and 11. (A) Sequences for E10-1 and E11-2. Red indicates

binding regions of hnRNPC in Tau minigene pre-mRNA. U-tract (at least 4 uridines in a row) positions are highlighted in yellow. Lowercase letters indicate intronic sequence, capital letters indicate exonic sequence. (B) Interaction between hnRNPC and Tau pre-mRNA E10-1 fragment is abolished after point mutations disrupting all U-tracts (by replacing the middle uridine with cytosine in each U-tract) in the fragment sequence, as measured by *in vitro* RNA/protein interaction assay. E10-1m represents the E10-1 fragment with all U-tracts mutated as described above. (C) Interaction between hnRNPC and Tau pre-mRNA E11-2 fragment is abolished after point mutations disrupting all U-tracts (shown as E11-2m). (D) The impact of hnRNPC on Tau exon 10 inclusion was partially abolished when co-expressed with Tau minigene containing U-tract mutations in either E10-1 or E11-2 (shown as TM_E10-1m and TM_E11-2m separately). The hnRNPC-induced Tau exon 10 inclusion was completely abolished when co-expressed with Tau minigene containing U-tract mutations in both of these fragments (shown as TM_all_Um). TM_WT represents the wild-type Tau minigene. (E) E11-2 is divided into four sub-fragments, 150 nt each with 20 nt overlap. U-tract positions are highlighted in yellow. (F) hnRNPC interacts with E11-2-1 and E11-2-2 as detected by *in vitro* RNA/protein interaction assay. The binding fragments are marked in red. (G) Interaction between hnRNPC and the E11-2-1 sub-fragment with mutated U-tract motifs (by replacing the middle uridine with cytosine) separately and altogether. E11-2-1-1m represents first U-tract was mutated; E11-2-1-2m represents second U-tract was mutated; E11-2-1-12m represents both U-tracts were mutated. (H) Interaction between hnRNPC and the E11-2-2 sub-fragment with mutated U-tract motifs following the same procedure described above. (I) Interaction between hnRNPC and E11-2 fragment is largely abolished with the mutations in top four U-tracts (shown as E11-2-4Um) compared to mutations in all U-tracts. (J) Top four U-tracts in E11-2 fragment together with the U-tracts in E10-1 fragments were mutated in Tau minigene (shown as TM_E10_4Um). The impact of hnRNPC on Tau exon 10 inclusion was mostly abolished when co-expressed with this mutated Tau minigene compared to full U-tracts mutated Tau minigene that described in (D).

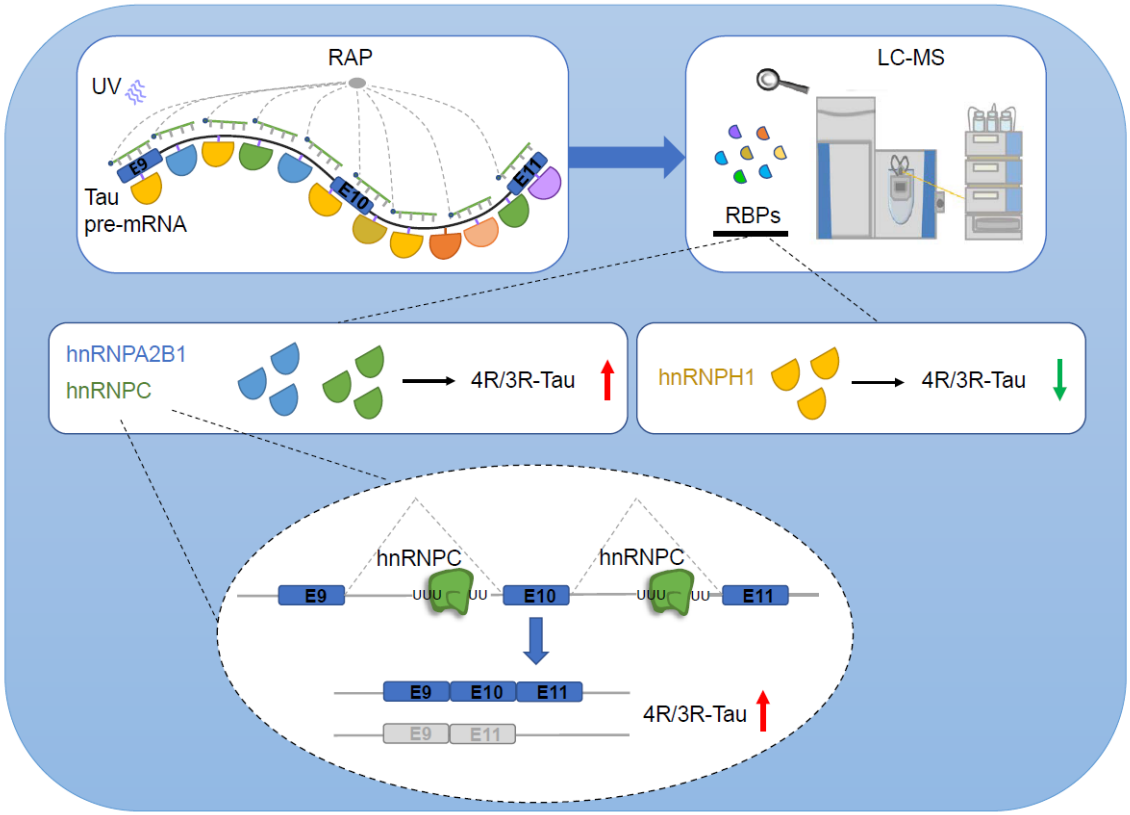


Figure 7: Working model. Novel RBPs that interact with Tau pre-mRNA around its exon 10 region were identified using the adapted and optimized RAP-MS proteomic method. Among these RBP candidates, HNRNPC (binding to the U-tracts at the 3'-ends of Tau introns 9 and 10) and HNRNPA2B1 are both required to promote Tau exon 10 inclusion, while HNRNPH1 promotes Tau exon 10 skipping.

2.7 Supplementary figures and tables

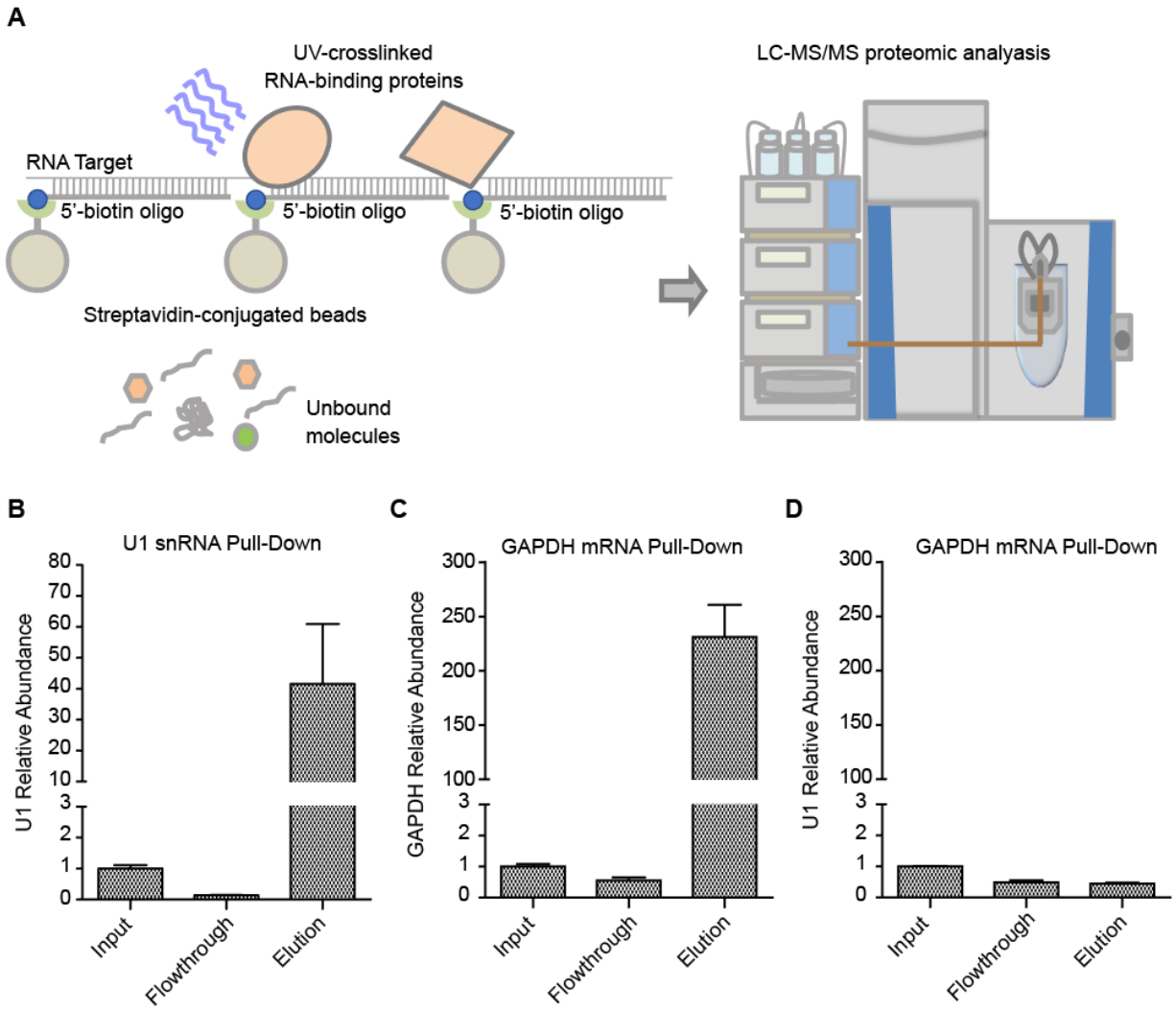


Figure S1: RNA antisense purification (RAP) proteomic method to identify RNA-binding proteins specifically interact with RNA targets. (A) RAP workflow. Cells are irradiated with UV light (254nm) to induce RNA-protein crosslinking. After cell lysis, target RNA molecules are hybridized with a pool of 90mer oligos with 5'-biotinylationand and are enriched together with the cross-linked RBP interaction partners by streptavidin beads. These RBPs are digested on beads and are identified using LC-MS/MS proteomics. (B) Enrichment of U1 snRNA in HEK 293FT cells using RAP method. U1 snRNA abundance was measured by qRT-PCR, with GAPDH as the internal standard. (C-D)

Specific enrichment of GAPDH mRNA in HEK 293FT cells using the RAP method. (C) GAPDH mRNA and (D) U1 snRNA abundances were measured by qRT-PCR, with ACTB as the internal standard.

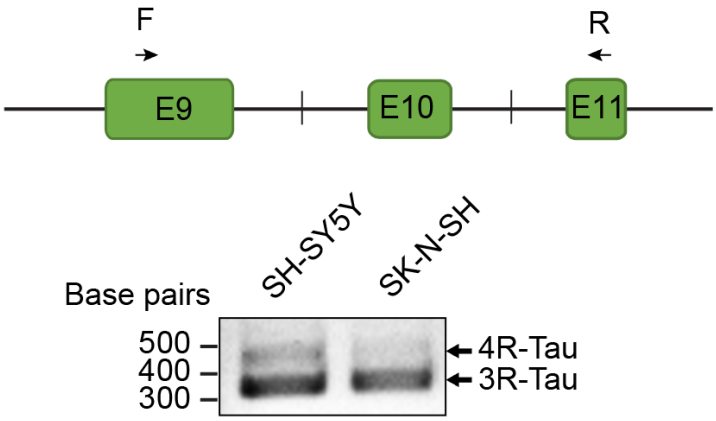


Figure S2: Detection of 3R- and 4R-Tau isoforms in SH-SY5Y and SK-N-SH neuroblastoma cells. Alternative splicing profile of Tau was measured by RT-PCR, using primers F and R.

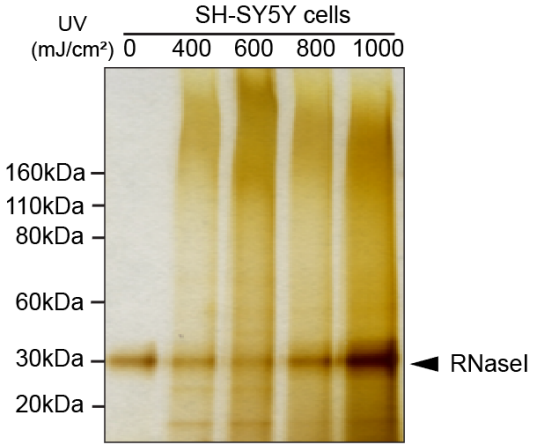


Figure S3: UV-crosslink optimization in SH-SY5Y. Proteins cross-linked to mRNAs with UV irradiation were captured by oligo-dT beads and released by RNase I digestion, before being loaded onto the SDS-PAGE and detected by silver staining.

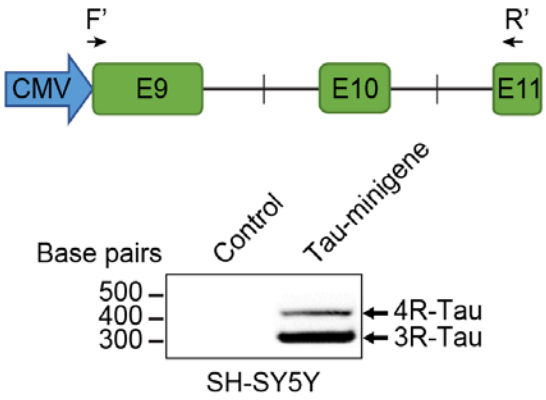


Figure S4: Expression of the Tau minigene in SH-SY5Y. Tau 3R and 4R isoforms expressed from the minigene measured using RT-PCR with the primers F' and R'.

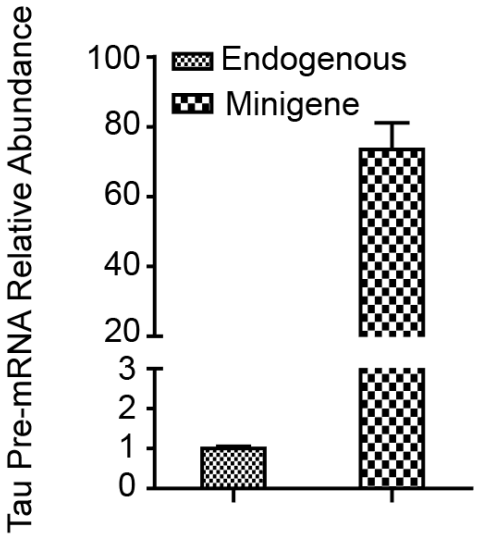


Figure S5: Expression of the Tau minigene pre-mRNA in SH-SY5Y stably expressing Tau minigene. Cells transduced and selected with Tau minigene or blank vector as control were harvest separately for qRT-PCR detection, with primers Fp and Rp as shown in Figure 1B.

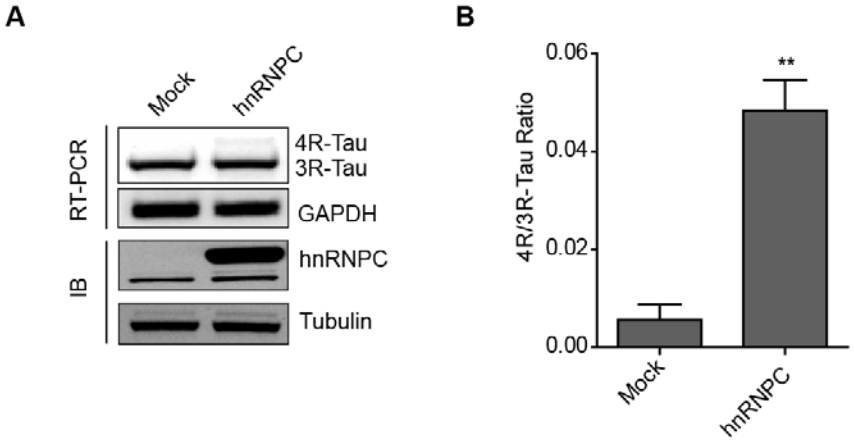


Figure S6: Overexpression of hnRNPC promotes Tau exon 10 inclusion in HEK 293FT cells. Forty-eight hours post-transfection, cells were harvested for Tau exon 10 splicing detection using RT-PCR. GAPDH was used as the internal control(A). The 4R/3R-Tau ratios from the Tau minigene were calculated (B). Data shown represent the means ± SEM of three independent experiments (**p < 0.01; two-tailed Student's t test)

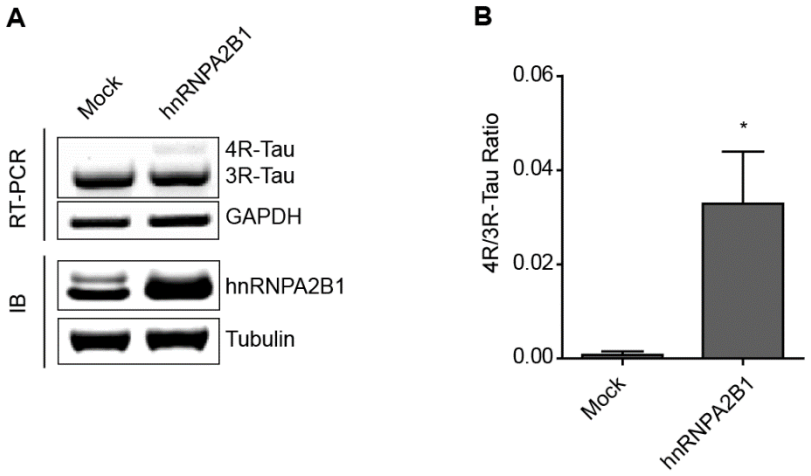


Figure S7: Overexpression of hnRNPA2B1 promotes Tau exon 10 inclusion in HEK 293FT cells. Forty-eight hours post-transfection, cells were harvested for Tau exon 10 splicing detection using RT-PCR. GAPDH was used as the internal control(A). The 4R/3R-Tau ratios from the Tau minigene were calculated (B). Data shown represent the means ± SEM of three independent experiments (*p < 0.05; two-tailed Student's t test).

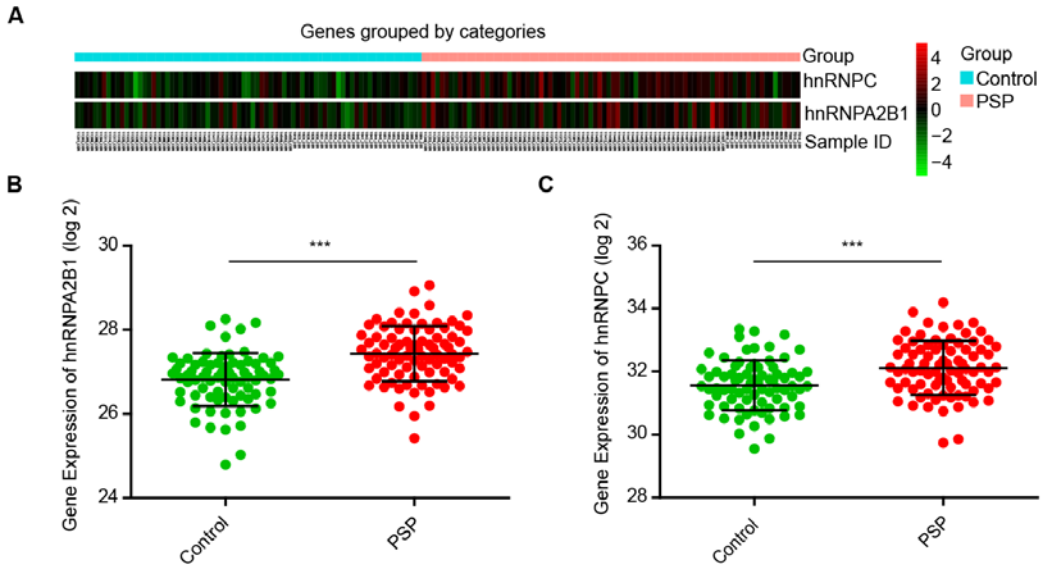


Figure S8: The 4R-Tau splicing factors hnRNPC and hnRNPA2B1 are both upregulated in PSP patient brain samples. (A) RNA-seq results for Control (n = 77) versus PSP (n = 84) cerebellum released from the Mayo clinic (see method). Heatmap shows gene expression of hnRNPA2B1 and hnRNPC in both groups. (B-C) hnRNPA2B1 (B) and hnRNPC (C) expression levels were significantly increased in PSP compared to Control (***adjusted p < 0.001; Benjamini-Hochberg correction).

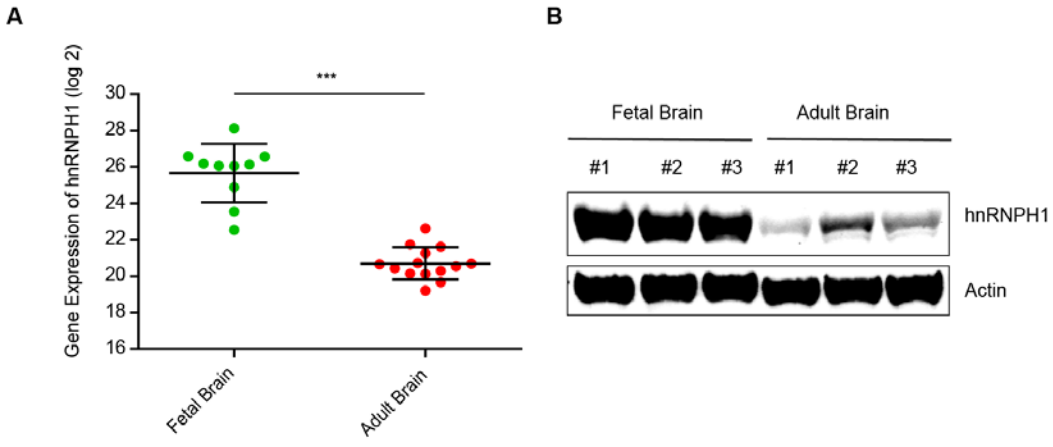


Figure S9: The 3R-Tau splicing factors hnRNPH1 is downregulated during human brain development. (A) RNA-seq analysis in ten post-mortem fetal brain versus fourteen adult brain (***adjusted p < 0.001; Benjamini-Hochberg correction). Gene expression of

hnRNPH1 was indicated. Green dots: fetal brain samples; Red dots: adult brain samples. (B) Western blot analysis of protein level of hnRNPH1 in three post-mortem fetal brain versus three adult brain samples, Actin was used as loading control.

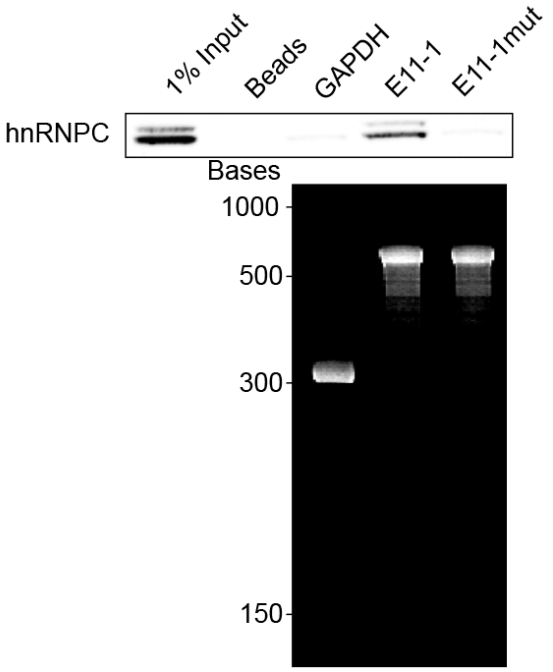


Figure S10: Binding affinity of hnRNPC with E11-1 was disrupted after mutation of the overlapped U-tracts shown in E11-2. Top: The binding affinity of hnRNPC to this region was detected by *in vitro* RNA pull down. E11-1mut: U-tracts mutant in overlapped region. E11-1 is a positive binding control. Bottom: 20% Input of RNA fragments were loaded in the 6% TBE-Urea gel and visualized with SYBR Gold.

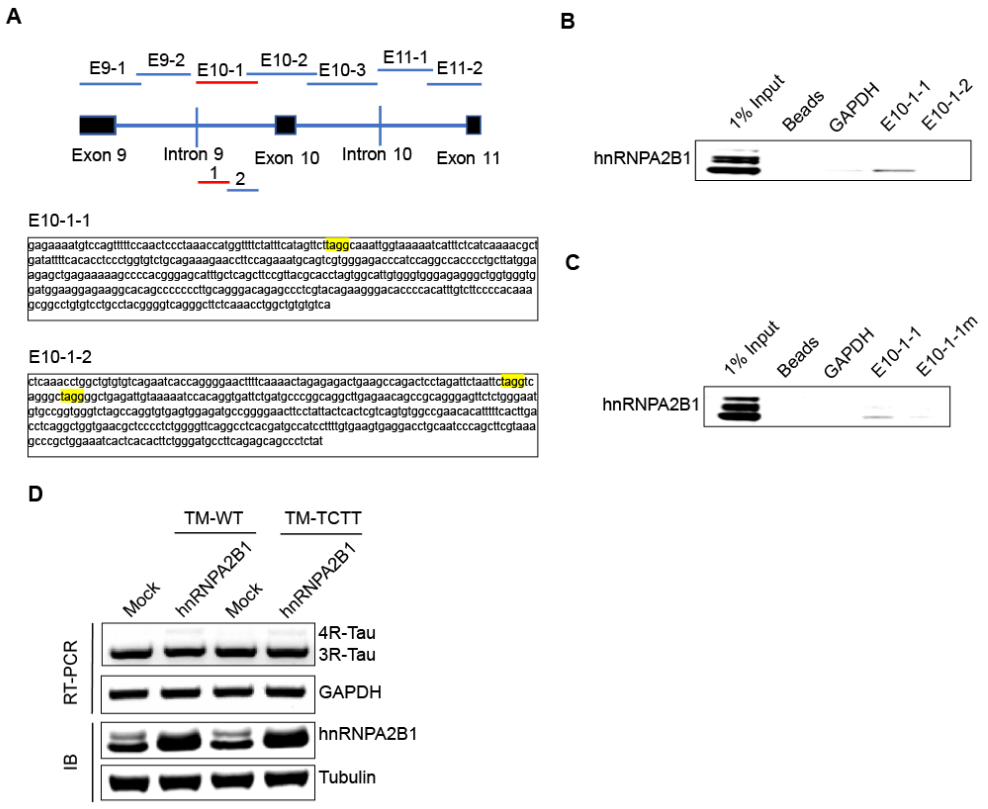


Figure S11: Characterization of interaction between hnRNPA2B1 and Tau pre-mRNA. (A) The E10-1 Tau minigene fragment was divided into fragment 1 and 2 with 150 bp overlap as shown in the top panel, the sequence for both sub-fragments are shown at the bottom; UAGR positions are highlighted in yellow. (B) hnRNPA2B1 interacts with fragment 1 as highlighted in (A), as detected by the *in vitro* RNA/protein interaction assay described above. (C) Interaction between hnRNPA2B1 and Tau pre-mRNA E10-1-1 fragment was abolished after point mutations replacing UAGG to UCTT in the E10-1-1 sequence(E10-1-1m), as measured by *in vitro* RNA/protein interaction assay. (D) hnRNPA2B1 was co-transfected with either wild type Tau minigene (TM-WT) or Tau minigene with TCTT mutation (TM-TCTT) in the E10-1-1 region, after forty-eight hours, cells were harvest for Tau splicing detection. GAPDH was used as the internal standard.

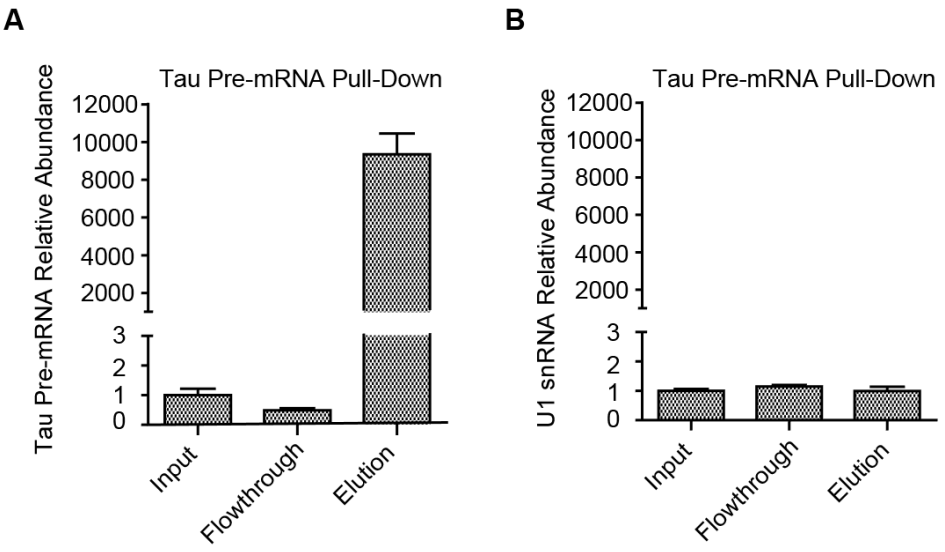


Figure S12: Enrichment of endogenous Tau pre-mRNA in SH-SY5Y cells. (A) Endogenous Tau pre-mRNA abundance through RAP process was measured by qRT-PCR using the primers Fp and Rp as shown in Figure 1B. (B) The U1 snRNA abundance was measured by qRT-PCR with the same samples as (A).

Table S1: ShRNA oligos and Primers used in this study.

ShRNA oligos:

Gene	Target Sequence
hnRNPD	TCGAAGGAACAATATCAGCAA
EMG1	GTAGTAGAACCTGTTCTGAAA
PRPF19	GAACGGATGTGGAAGGAAGAA
DDX21	CCTGAGGTTGATTTGGTTATA
hnRNPA3	GCTTTGAAACTACAGATGATA
hnRNPH1	GATCCACCACGAAAGCTTATG
hnRNPC	GCCTTCGTTCAGTATGTTAAT
hnRNPA2B1	CAGAAATACCATACCATCAAT
ARL6IP4	GAGAAAGAAGAAGAGCAGGAA
hnRNPR	GAGGGTTTGGTGGACGTTATT
LUC7L3	CCGGGATCGAAAGTCATATAA
DHX15	GTTGGTTCGATAATGGCCTTT
hnRNPU	CAGTGCTTCTTCCCTTACAAT

THRAP3	AGATCTCGTTCTCGTTCATTT
hnRNPK	CGTTATTGTTGGTGGTTTAAA
hnRNPA2B1-SH-2	GCTTCTTCCTATTTGCCATGG(UTR)
hnRNPC-SH-2	GCGCTTGTCTAAGATCAAATT(UTR)
HnRNPH1-SH-2	TTGCCCTTTGCCACGTTAAAT

qRT-PCR primers:

Gene	**Forward sequence 5' to 3'**	**Reverse sequence 5' to 3'**
GAPDH	ACCACAGTCCATGCCATCAC	ATGACCTTGCCCACAGCCTT
ACTB	TGTGACGTGGACATC	ACCTGAGCAAGGTGACCTCCA
U1 SnRNA	TGGCAGGGGAGATACCATGA	TGCAGTCGAGTTTCCCACAT
hnRNPH1	GGGATAACATTGCCGGTGGA	TGGCCATAAGCTTTCGTGGT
hnRNPR	ACAGTGTGTACTCTGGCGTG	GACCGGACAGTGGATCCATC
DDX15	GTTGGTGAGACTGGGTCTGG	TCCTGGCCCAACATCACATC
hnRNPC	TCGAAACGTCAGCGTGTATC	TCCAGGTTTTCCAGGAGAGA
hnRNPA2B1	CAAGAGAGGAGTCTGGGAA	ATGGTCATCAAAGGTAACAA
ARL6IP4	TGACCAAGGAGGAGTGGGAT	TTTCCTCTAGGACCTCGCCA
CLASRP	CGCAAGAGAGAAGCTGACCA	TTTCCTCTAGGACCTCGCCA
LUC7L3	GGACCAAGATCAGAACGTGTATTTG	CAGTTGTTGGATGAGTTAATGGGC
DHX9	CAGGAGAGAGAGTTACTGCCT	CTCTGCTGCTCGGTCATTCTG
HNRNPU	GAGATTGCTGCCCGAAAGAAGC	TTCGCTGGAAGCCTGCAAACAG
YBX1	TCGCCAAAGACAGCCTAGAGA	TCTGCGTCGGTAATTGAAGTTG
hnRNPK	AACACTCAGACAACAATCA	TCCTCCAATAAGAACAACTC
THRAP3	CAGAAGGAGGAGAGTGCTGC	ATCACCCAAGCCTGTGTCAG
EMG1	GAAGTGAATCCCCAGACCCG	CACTGACAACCGGGATGGAA
HNRNPA3	GGTAATTTTGGCCGTGGTGG	CTATAACCAGGACCACCGCC
HNRNPD	GCCATTCAAACTCCTCCCCA	GCACAAAGCCAAAACCCCTT
C1QBP	GCAAGAAGGCCCTTGTGTTG	GACTCGCCAGTGGACTGAAA
PRPF19	TGCCAAGTTCCCAACCAAGT	GCTGAGCTCTTCTGGCTTCA
DDX21	CAGGATGCTCAGTCCTTGCA	CGCCCGGATCGATGAATGTA
Tau_Intron 9	AGGCTCTCATAGTCAGGGCA	AACCTAAGCAGATGGCCCAC
Tau-MAP-F1	CCAAGATCGGCTCCACTGAG	CACGACTCCACGCTCAACC

Tau-MAP-F2	CTGGCTACCAAAGGTGCAGA	TACGCACTCACACCACTTCC
Tau-MAP-F3	CTGTCCAATAAGGGCCTGGG	CACTTGGAGGTCACCTTGCT
18S rRNA	GTAACCCGTTGAACCCCATT	CCATCCAATCGGTAGTAGCG

***In vitro* RNA pull-down primers:**

Fragment Name	**Forward sequence 5' to 3'**	**Reverse sequence 5' to 3'**
E9-1	TAATACGACTCACTATAGGGGTGAACCTCCAAAATCAGGGG	ACCTAAGCAGATGGCCCAC
E9-2	TAATACGACTCACTATAGGGTGTTAATCGGACAGAGATGGCAG	CTGAGTAGCTGGGATTATAGGCG
E10-1	TAATACGACTCACTATAGGGGAGAAAATGTCCAGTTTTTCCAACT	ATAGAGGGCTGCTCTGAAGG
E10-1m	TAATACGACTCACTATAGGGGAGAAAATGTCCAGTTCTTCCAACT	ATAGAGGGCTGCTCTGAAGG
E10-2	TAATACGACTCACTATAGGGTCACTTGACCTCAGGCTGGTG	CATGAGTGGAAGGAGACTTTGATGG
E10-3	TAATACGACTCACTATAGGGTTCCTCATATTCTAGGAGGGGGT	GTTTTTTGAGACAAAGTCTCACTCTGT
E11-1	TAATACGACTCACTATAGGGGCCACTGCTGGAGGCCCT	AATTGACAGATCACAATCAGGCATGA
E11-2	TAATACGACTCACTATAGGGAAAGCCCACGCATCGACCCT	CTGGTTTATGATGGATGTTGCCTAATG
E11-2-1	TAATACGACTCACTATAGGGAAAGCCCACGCATCGACCCT	AATTGACAGATCACAATCAGGCATG
E11-2-2	TAATACGACTCACTATAGGGCTGATTGTGATCTGTCAATTTCATG	GAGTGTCTTCAATCCCTGCA
E11-2-3	TAATACGACTCACTATAGGGTGCAGGGATTGAAGACACTCC	GTGTAAGCCCAGGCCCTTAT
E11-2-4	TAATACGACTCACTATAGGGATAAGGGCCTGGGCTTACAC	TAGACTATTTGCACCTGGAGATGAG

E10-1-1	TAATACGACTCACTATAGGG GAGAAAATGTCCAGTTTTTC CAACT	TGACACACAGCCAGGTTTGA
E10-1-2	TAATACGACTCACTATAGGG CTCAAACCTGGCTGTGTGTC	ATAGAGGGCTGCTCTGAAGG

Table S2: DNA probes used in RAP-MS.

ID	Probe sequence 5' to 3'
U1-1	/5Biosg/CTG GGA AAA CCA CCT TCG TGA TCA TGG TAT CTC CCC TGC CAG GTA AGT AT
U1-2	/5Biosg/GGG AAA TCG CAG GGG TCA GCA CAT CCG GAG TGC AAT GGA TAA GCC TCG CC
U1-3	/5Biosg/AAC GCA GTC CCC CAC TAC CAC AAA TTA TGC AGT CGA GTT TCC CAC ATT TG
GAPDH-2	/5Biosg/CCCCATGGTGTCTGAGCGATGTGGCTCGGCTGCGCGGAGG GAGAGAACAG TGAGCGCCTAGTGGCCCCGGCCGACACACACGCCTCCCCT
GAPDH-3	/5Biosg/GGC AAC AAT ATC CAC TTT ACC AGA GTT AAA AGC AGC CCT GGT GAC CAG GCG CCC AAT ACG ACC AAA TCC GTT GAC TCC GAC CTT CAC CTT
GAPDH-7	/5Biosg/GCA GGA GGC ATT GCT GAT GAT CTT GAG GCT GTT GTC ATA CTT CTC ATG GTT CAC ACC CAT GAC GAA CAT GGG GGC ATC AGC AGA GGG GGC
GAPDH-8	/5Biosg/AGT GAT GGC ATG GAC TGT GGT CAT GAG TCC TTC CAC GAT ACC AAA GTT GTC ATG GAT GAC CTT GGC CAG GGG TGC TAA GCA GTT GGT GGT
GAPDH-9	/5Biosg/GCC AGT AGA GGC AGG GAT GAT GTT CTG GAG AGC CCC GCG GCC ATC ACG CCA CAG TTT CCC GGA GGG GCC ATC CAC AGT CTT CTG GGT GGC
GAPDH-13	/5Biosg/GGC CAT GTG GGC CAT GAG GTC CAC CAC CCT GTT GCT GTA GCC AAA TTC GTT GTC ATA CCA GGA AAT GAG CTT GAC AAA GTG GTC GTT GAG
Tau-6	/5Biosg/CAA GAG AAG AGC CAG AAC GTT CTC TCG CCT GAG GAG GGC ACT CAC AGT GTA GTG GAG AGC CCA ATA AAG AAC TGA ACT TCC TTG AAG AGG
Tau-7	/5Biosg/GGA ACC GCT GTG GGT GCC CTG ACT ATG AGA GCC TCC GGC CGT GGA GAC ACA GCC CTG CCA TCT CTG TCC GAT TAA CAG GGG CTG AAG GGG

Tau-8	/5Biosg/GAA TCC AGA GGA ACC CAA CCT AAG CAG ATG GCC CAC GAG TGG AGA TGC TGA CCT ATG GGT GGC AGT GTA TTC TGC CCA GAA GGC AGG TGG
Tau-9	/5Biosg/TTC TCC ACC TTT AGG AAG CTG GAG AGC TCG CGA GCA CCC ACT TGC TCC CAG AAG AAT CAG CTG ATC AAA ACA GAA CCC CCA ATC TCC CCA
Tau-10	/5Biosg/GCT TTG TAT GTC AGC CCT AAC GCA CCC AGA CGA GGG GCA CAG GGA CAG GAA TGG GGA AAG GGA TCC AGG CCA GGC CCC CGA AGT CTG TGC
Tau-11	/5Biosg/GCT ACC CAC GGC CAC AGA GAG GCC TTG GAA GAA ACC CAC CCG CCC ACA CCT GGC TAG TGA GGA GGC ATA CTG CTG TTC TTT CAC TGT GGT
Tau-12	/5Biosg/TGC CAT GGT GTT CTT ATA AGA AGA TGA GCA TGC GCA GAG CTA CCA GGA GCA CAC ACA GAG GGA AGA CTG CAG CGG TGC AGG ACA GAG GTG
Tau-16	/5Biosg/TCC ATA AGC AGG GGT GGC CTG GAT GGG TCT CCC ACG ACT GCA TTT CTG GAA GGT TCT TTC TGC AGA CAC CAG GGA GGT GTG AAA ATA TCA
Tau-17	/5Biosg/ATC CAC CCA CCA GCC CTC TCC CAC CCA CAA TGC CAC TAG GTG CGT AAC GGA AGC TGA GCA AAT GCT CCC GTG GGG CTT TTT CTC AGC TCT
Tau-18	/5Biosg/ACA GGC CGC TTT GTG GGG AAG ACA AAT GTG GGG TGT CCC TTC TGT ACG AGG GCT CTG TCC CTG CAA GGG GGG GCT GTG CCT TCT CCT TCC
Tau-21	/5Biosg/CCA CAC TGA CGA GTG AGT AAT AGG AAG TTC CCC GGC ATC TCC ACT CAC ACC TGG CTA GAC CCA CCG GCA CAT TCC CAG AGA ACT CCC TGC
Tau-22	/5Biosg/GTC CTC ACT TCA CAA AAG GAT GGC ATC GTG AGG CCT GAA CCC CAG AGG GGA GCG TTC ACC AGC CTG AGG TCA AGT GAA AAA TGT GTT CGG
Tau-23	/5Biosg/CCA GGG GAG CTG AAG GGA TAG AGG GCT GCT CTG AAG GCA TCC CAG AAG TGT GAG TGA TTT CCA GCG GGC TTT ACG AAG CTG GGA TTG CAG
Tau-24	/5Biosg/CCC TGG ACC CGC CTA CTT GCT CGC AAG GAT GCC TCC ACT TTC GGA CTT GGC AGA GGC AGT CTG GGG AGT GAC GGG AGG TTG AGT CAC ATC
Tau-25	/5Biosg/TAG CCA GAA AAA AGG ATG AGT GAC ACG CCA CCC TGG ACC CGC CTG CTT GCT CGC AAG GAC GCC TCC ACT TTC GAT GAG TGA CAT GCG CCA
Tau-28	/5Biosg/CCG TGT GGG GGA ATG TGG GGA GCC GGG CTA CAT TCA CCC AGA GGT CGC AGC CAG ATC CTG AGA GCC CAA GAA GGA TTT ATT CTA TGC AGT

Tau-29	/5Biosg/TGT CCA CAC AGC CTG CGC AGG CTG GGG GAC CCA GAA ATG CTG GGA CAC CCC CTC CTA GAA TAT GAG GAA GGG GCT TCT GGG AAC AGT GGA
Tau-30	/5Biosg/AGG GCT GGG CAT GAG TGG AAG GAG ACT TTG ATG GGT ATT CTC AGC TTC CCA GGC AGG GTC CGG AGA GAG GGT CCG TCA TCT GCC CTA TTC
Tau-32	/5Biosg/CAG CAG CTG AGC AGG ACA GTA AGA CCT TGG TGT CCG CTC ATC AGA ACA AAC GGG TCA CCT GTC CAC AGT GGA GCC ACT CAA CGT CAA CCT
Tau-33	/5Biosg/CTT TGC ATC TGG AGT CTC TGC CAG AGG GTG GCG CAG GCT AAG CAT AAG CTG GAG GCT TAG GAA TCG GCT CCT GCC TTG AAC GTG TAG GAG
Tau-34	/5Biosg/GAG GCA GCC AAA TCT AAA AGG GCC AAA AGC AGC ATT TCT CTG CCA CAC CTC AAC GCT GGG CAG AGG GAC CTG TCG CAC TTT GGT TTG GCT
Tau-39	/5Biosg/AAA GAA CCA GCT CCC CTG TGG GGG GCA CGC TGT AGG GAC TCC ACA GTG GCA GAC AGA CAG GTC CCT TGG AGG CAG GCC TCT ATG CAG CCA
Tau-40	/5Biosg/CGG ACG CAG ACA CAC ATG AGG AGG CCT GTG GCC TCC AGG GCA CCG CTC CCT CTG GGA ATG ATG ACC CTG AGC TGC CAG CTG ATC TCA GTC
Tau-41	/5Biosg/TGT CTT TAG CAA TAG GAA CAA AGC AAC ACC AAC AGG GTG GGT AGG TGG CAC AGC AGA AAC CAA CAG ATT TAG TGT CTC AGT AAG CTC GAG
Tau-42	/5Biosg/AAG ACC ATC AGT AAG CAG GAC CCA GCG GAG TTC AGC AGC TCG GAC GTG AGA ACC AGC AGG CAC CTG CAC ACT CAG TGT CCT GGA CAT TCC
Tau-43	/5Biosg/TGG GCC TCA AAG AGG GGA GGC GCT GTG AGG GTC GAT GCG TGG GCT TTA TTC CAG GAA AAG CTT CCA CGG ACC CTG GAA AGC ACT AGA GCA
Tau-44	/5Biosg/CAA TCA GGC ATG AAG AAA AAA GTG ATC AAT ATG ATG AGT ATG CAC AGC ATC TGA AAA ATC TTG CTG GAA AGG CAG GAG TGG GGT ATC TGC
Tau-45	/5Biosg/ATT GTC ATT GTG CAA GAC CCA GAC ATT TGC TCA GCA AAC GCT TAA GTG TAA AAA TGT CAC TCC CTT TCC TGA CAT GAA ATT GAC AGA TCA
Tau-46	/5Biosg/CAA CCC TGC CCC AGG GGA GCT GGG GAC TGT GGG ACT GGA GTG TCT TCA ATC CCT GCA GAA CAA AAG AGC CTC TGG GAA AAA CAG GGA CCC
Tau-47	/5Biosg/TAA GCC CAG GCC CTT ATT GGA CAG TGG TTC CCA GCT GCC ATG AGG AGC ACA TCC TAG TCA GGG TGG AAA AAT GTG TTG TCG AAA TTC TGC

Tau-48	/5Biosg/GGA AAG GCA GAG AGA GAG GAA GAC AGA GAG ACA AAA AGG ATA AAA TGG GAT TAT TTT ATT AAG GGT GTA ACT CAA TGA GAA GCA GCT GTG

Table S3: RNA binding proteins identified by proteomic analysis in Tau probes-based RAP-MS pulldown.

Protein name	**Tau-1**	**Tau-2**	**GAPDH-1**	**GAPDH-2**
ADP ribosylation factor like GTPase 6 interacting protein 4(ARL6IP4)	1	1	0	1
CLK4 associating serine/arginine rich protein (CLASRP)	1	0	0	0
DEAD-box helicase 17(DDX17)	0	1	0	0
DExD-box helicase 21(DDX21)	4	3	0	3
DEAD-box helicase 23(DDX23)	1	0	0	0
DExD-box helicase 39B(DDX39B)	2	4	2	3
DEAD-box helicase 46(DDX46)	0	2	0	1
DEAD-box helicase 5(DDX5)	4	0	0	0
DEAH-box helicase 15(DHX15)	2	3	0	2
DExH-box helicase 9(DHX9)	3	3	2	3
dyskerin pseudouridine synthase 1(DKC1)	1	0	0	0
EMG1, N1-specific pseudouridine methyltransferase (EMG1)	1	1	0	1
fibrillarin-like 1(FBLL1)	0	2	0	1
FUS RNA binding protein (FUS)	4	0	0	0
heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	4	12	3	0
heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	2	10	1	4
heterogeneous nuclear ribonucleoprotein A3(HNRNPA3)	1	1	0	0
heterogeneous nuclear ribonucleoprotein A/B(HNRNPAB)	0	1	0	0
heterogeneous nuclear ribonucleoprotein C (C1/C2) (HNRNPC)	1	1	0	1
heterogeneous nuclear ribonucleoprotein D(HNRNPD)	2	1	0	1
heterogeneous nuclear ribonucleoprotein F(HNRNPF)	0	1	0	1
heterogeneous nuclear ribonucleoprotein H1 (H)(HNRNPH1)	7	0	0	0

heterogeneous nuclear ribonucleoprotein H3(HNRNPH3)	1	0	0	0
heterogeneous nuclear ribonucleoprotein K(HNRNPK)	7	3	0	0
heterogeneous nuclear ribonucleoprotein M(HNRNPM)	1	0	0	0
heterogeneous nuclear ribonucleoprotein R(HNRNPR)	1	3	0	0
heterogeneous nuclear ribonucleoprotein U(HNRNPU)	10	3	1	4
heat shock protein family A (Hsp70) member 8(HSPA8)	10	13	6	11
KH RNA binding domain containing, signal transduction associated 1(KHDRBS1)	1	0	0	0
leucine rich pentatricopeptide repeat containing (LRPPRC)	27	38	29	29
LUC7 like 3 pre-mRNA splicing factor (LUC7L3)	2	3	0	2
nucleolar protein 6(NOL6)	1	0	0	0
non-POU domain containing, octamer-binding (NONO)	1	1	2	0
nucleophosmin(NPM1)	1	0	0	0
nudix hydrolase 21(NUDT21)	0	1	0	0
poly(rC) binding protein 3(PCBP3)	1	0	0	0
pleiotropic regulator 1(PLRG1)	0	1	0	0
pinin, desmosome associated protein (PNN)	6	3	4	5
pre-mRNA processing factor 19(PRPF19)	2	2	0	2
pre-mRNA processing factor 38A(PRPF38A)	2	0	1	0
pre-mRNA processing factor 40 homolog A(PRPF40A)	0	2	0	1
poly(U) binding splicing factor 60(PUF60)	0	2	0	1
RNA binding motif protein 39(RBM39)	6	7	9	8
RNA binding motif protein, X-linked like 2(RBMXL2)	0	1	0	0
RNA binding protein with serine rich domain 1(RNPS1)	23	16	13	16
SAP domain containing ribonucleoprotein (SARNP)	0	1	0	1
splicing factor 3b subunit 3(SF3B3)	10	8	3	7
splicing factor proline and glutamine rich (SFPQ)	2	1	0	0
SON DNA binding protein (SON)	2	1	1	1

splicing regulatory glutamic acid and lysine rich protein 1(SREK1)	0	3	0	2
serine and arginine repetitive matrix 1(SRRM1)	0	2	0	1
serine/arginine repetitive matrix 2(SRRM2)	7	12	8	8
serine and arginine rich splicing factor 1(SRSF1)	24	29	44	23
serine and arginine rich splicing factor 10(SRSF10)	2	4	4	4
serine and arginine rich splicing factor 11(SRSF11)	3	7	3	7
serine and arginine rich splicing factor 2(SRSF2)	12	1	16	5
serine and arginine rich splicing factor 3(SRSF3)	17	13	36	22
serine and arginine rich splicing factor 5(SRSF5)	5	0	3	0
serine and arginine rich splicing factor 6(SRSF6)	7	0	16	2
serine and arginine rich splicing factor 7(SRSF7)	8	14	27	19
serine and arginine rich splicing factor 8(SRSF8)	1	0	0	0
serine and arginine rich splicing factor 9(SRSF9)	0	7	3	2
thyroid hormone receptor associated protein 3(THRAP3)	4	3	0	1
transformer 2 alpha homolog (TRA2A)	4	4	3	2
transformer 2 beta homolog (Drosophila)(TRA2B)	3	6	1	6
Y-box binding protein 1(YBX1)	2	2	10	7

***** Spectra counts of each RBPs from two biological MS runs are listed. The corresponding identified spectra counts for GAPDH pulldown are also listed. Known Tau exon 10 splicing regulators identified in our RAP-MS experiments are highlighted in blue.

Chapter 3: Investigation of Tau Isoform-Specific Interactome in Mouse Primary Cortical Neurons

3.1 Abstract

Tau is a microtubule binding protein that is preferentially expressed in nervous system to support neuronal function. Alternative splicing of Tau pre-mRNA generates multiple isoforms and is developmentally regulated in human brain. Alternative splicing of the N-terminal domain produces 0N-, 1N-, 2N-Tau isoforms, whereas alternative splicing of exon 10 leads to isoforms with three or four microtubule binding repeats named 3R- and 4R-Tau. Tau filament is the major detrimental pathological feature in a large group of neurogenerative diseases collectively termed tauopathy. Interestingly, the isoform composition of Tau aggregates differs among different types of tauopathy, which indicates the functional discrepancy among Tau isoforms. To investigate the functional difference of human Tau isoforms, we fused six human central nervous system (CNS) Tau isoforms with a mutant form of biotin ligase, BioID2, that can promiscuously label proximal and interacting protein with biotin. Using subsequent tandem mass tag (TMT) multiplexed quantitative mass spectrometry (MS), we revealed the protein-protein interacting networks for each of the six Tau isoforms. In this study, we found that the total Tau interacting proteins were associated with endocytosis, neurodegeneration, and synaptic transmission pathways. Through gene ontology enrichment analysis, we demonstrate that 1N-Tau isoforms may have putative roles in fatty acid metabolism, PPAR and glucagon signaling pathways. Intriguingly, regarding the interactome network comparison between 3R- and 4R-Tau, we observed that 4R-Tau proximal proteins are involved in endocytosis, whereas

proteins in proximity with 3R-Tau are enriched in fatty acid metabolism. Through our BioID2 proteomic analysis, we further identified a novel Tau interactor, MAT2A, that exhibited preferential binding with 3R-Tau compared to 4R-Tau, which we also confirmed with follow-up validation experiments. This finding may imply the specific functional association between MAT2A and 3R-Tau isoforms. Altogether, the results described above demonstrate that Tau isoforms might exert different functions in nerves system and the interactome networks identified here may provide useful clues to further dissect the regulatory mechanisms underpin isoform-specific aggregation in a subset of tauopathies.

3.2 Introduction

Microtubule-associated protein Tau is a microtubule binding protein and characterized by accelerating microtubule assembling and stabilizing microtubule structure[255]. Tau aggregation is a key pathological feature in a group of neurodegenerative diseases known as tauopathies, including Alzheimer's Disease (AD), Frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17), progressive supranuclear palsy (PSP), and other neurodegenerative disorders[302]. Tau aggregation is central to the progressive neuronal dysfunction and damage in tauopathy patients.

Interestingly, as a member of the microtubule-associated protein (MAP) family, Tau is predominantly expressed in the brain, suggesting that Tau might have specific functions in the nervous system beyond its basic role as a microtubule organizer[303-305]. Over the past decades, Tau has been reported to modulate axongenesis and post-synapse activity[306, 307]. In addition, Tau protein has been proved to promote genome integrity and chromosome stability in nucleus[44, 80]. Furthermore, Tau is a highly modified protein with

multiple residues containing a variety of post translational modifications including phosphorylation, acylation, ubiquitination, and SUMOylation, of which also associated with the structure diversity of tauopathy strains[308-311]. These evidence collectively suggest that Tau plays a significant regulatory role in neurons, far beyond what we have understood by far.

The Tau protein found in human adult brain has six isoforms, resulting from combinatory alternative splicing of the cassette exons 2, 3 and 10. With the alternatively spliced exon 10, Tau can be classified into 3R-Tau isoforms (with three microtubule binding repeats) or 4R-Tau isoforms (with four microtubule binding repeats). The N-terminal exons 2 and 3 can be alternatively spliced, giving rise to 0N- ,1N- and 2N-Tau isoforms[255]. Intriguingly, these Tau isoforms are developmentally regulated, as only the shortest 0N3R isoform shows in fetal brain, where all six isoforms are present in adult brain with an almost equal levels of 3R-Tau and 4R-Tau[255]. The balance of 3R- and 4R-Tau is crucial for neurons to maintain their physiological functions. Mutations identified in FTDP-17 patients triggering imbalanced expression of 3R-Tau and 4R-Tau, which is intimately linked to tauopathy pathogenesis[257]. Distorted 4R/3R-Tau ratios have also been observed in sporadic tauopathies such as PSP, CBD (Tau filament composed of 4R-Tau), and Pick's disease (Tau filament composed of 3R-Tau)[259]. Taken together, the above compelling evidence demonstrates that there might be functional differences among isoforms in the context of neuronal developmental and Tau pathological process and the Tau interactome differences might contribute to isoform-specific Tau aggregates in different types of tauopathies.

Characterization of the protein-protein interaction network for Tau isoforms is essential to understand the molecular mechanisms that lead to Tau aggregation and neurodegeneration. In recent years, besides cytoskeleton associated proteins, Tau has been reported to interact with ribosome proteins and RNA binding proteins (RBPs) to regulate RNA metabolism and translation[312-314]. In particular, several proteins have been observed to show binding preference among mouse Tau isoforms with apoA1 displayed higher binding affinity with 2N-Tau, whereas β-synuclein showed preference for 0N-tau[151]. Despite these findings, to date there has been no systematic characterization of *in situ* neuronal Tau protein interactome, which is critical for understanding the functional role of Tau in physiological and pathological conditions in the brain.

Over the last decades, antibody-based affinity precipitation combined with mass spectrometry (AP-MS) has been widely used as a tool to reveal protein-protein interaction (PPI) in the context of cells and tissues[315-317]. However, AP-MS is not available to capture *in situ* protein-protein interactions. AP-MS requires prior cell lysis, making it difficult to preserve native protein-protein interaction and may cause loss of the transient and weak bindings. Moreover, false positive interactions can be introduced because of the disruption of subcellular structure. Recently, a novel *in situ* protein biotinylation labeling method named proximity-dependent biotin identification (BioID) was developed to identify proteins that are in proximity of a target protein. The BioID method is based on a promiscuous bacterial biotin ligase called BirA1* (R118G), which can label nearby proteins by proximity-dependent cellular biotinylation. Genetically fused with the bait protein, BioID can be applied to study protein interactome in living cells. Through

streptavidin pull-down, the biotinylated proteins in close proximity to the bait protein can be identified by mass spectrometry[318-320]. These candidates identified by the BioID method represent direct interactors, indirect interactors, and/or vicinal proteins that do not physically interact with the fusion protein. In 2016, Roux et al. improved the previous method by screening a smaller biotin ligase from *Aquifex aeolicus*, called BioID2, which is a more efficient way to label proximate proteins[220].

In this study, we applied the proximity protein labeling method using BioID2, followed by tandem mass tag (TMT) multiplexed quantitative proteomics, to unveil the interactome network for each of the six human Tau isoforms in mouse primary neurons. Through functional enrichment analysis, we found that 1) proteins in proximity with 1N-Tau are enriched in fatty acid metabolic pathways and signaling transduction; 2) 4R-Tau interacting proteins were specifically involved in endocytosis, whereas 3R-Tau has a putative role in fatty acid related metabolism. Furthermore, we identified MAT2A as a novel Tau interactor. Specifically, MAT2A exhibited 3R-Tau preferential binding affinity.

3.3 Methods

Plasmids and antibodies

The pLJM1 EGFP-Tau, PET28-his-3C_Tau352 and Tau/pET29b plasmids were obtained from Addgene, with which contains sequence for Tau 0N4R, 0N3R and 2N4R respectively. Six Tau isoforms were then cloned to pLV-hSyn-TurboGFP-P2A-BioID2 by using infusion cloning strategy. pLV-hSyn-TurboGFP-P2A-luciferase-P2A-BioID2 plasmid was kindly provided by Dr. Karun Singh's lab. Rabbit polyclonal anti-Tubulin (#2148S, 1:1000 dilution) was from Cell Signalling Technology. Mouse monoclonal anti-Flag (#F3165-

2MG, 1:6000 dilution) was from Sigma. Mouse monoclonal anti-Tau (#MN1000, 1:1000 dilution) was from Thermo Fisher. Rabbit polyclonal anti-HA (#ab9110, 1:4000 dilution) was from Abcam.

Cell culture

HEK293FT cells were cultured in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% fetal bovine serum (FBS). Mouse primary neuron (CD-1, E16) were cultured in neural basal medium supplemented with 50x B27 (Life Technology),100x Pen/Strep (Life Technology) and 100x Glutamine (Life Technology).

Immunofluorescence

Primary mouse neurons were fixed on DIV18 with 4% formaldehyde. After blocking and permeabilization with 10% Normal Donkey Serum (Sigma) /0.3% Triton X-100 in PBS (B/P buffer) for 45 minutes, primary neurons were incubated overnight at 4°C with primary antibody dilutions: mouse anti-Flag (1:1000 dilution, Sigma #F3165-2MG), chicken anti-Map2 (1:2000 dilution, Cedarlane #CLN182), rabbit anti-TurboGFP (1:2000 dilution, Thermo Fisher #PA5-22688), Alexa Fluor® 647 (1:500 dilution, Jackson ImmunoResearch # 016-600-084). Primary antibodies were diluted in B/P buffer. Neurons were then washed with PBS for 3 times, 10 minutes incubation for each washes. After wash steps, neurons were incubated with fluorescent- conjugated secondary antibodies for 1.5 hour. Secondary antibodies including Cy™3 AffiniPure Donkey Anti-Mouse IgG (H+L) (1:500 dilution, #715-165-151), DyLight™ 405 AffiniPure Donkey Anti-Chicken IgY (IgG) (H+L) (1:500 dilution, #703-475-155), Alexa Fluor® 488 AffiniPure Donkey Anti-Rabbit IgG (H+L)

(1:500 dilution, #711-545-152), secondary antibodies were diluted in 50% B/P buffer and 50% PBS. Followed by PBS washes for 3 times, neurons were mounted onto slides with ProLong Gold Antifade mounting medium. Stained samples were imaged by using Zeiss LSM700 confocal microscope on the next day. Images were scanned with 11 Z-stacks (0.67 μm step size) and were analyzed with the maximum intensity by ImageJ software.

Proximity-dependent biotin identification (BioID) in primary neuron

Lentivirus transduction and cell harvest

Primary mouse cortical neurons were cultrued from E16-E17 CD-1 mouse pups and plate at 600 000 cells/well in PDL and laminin-coated 12-well plates. After titration with primary neuron on DIV 3-4, primary neurons on DIV 14 were transduced with lentivirus at MOI of 0.7. 24 hours later, topping up media by adding 2mL fresh media to each well. 72 hours post-transduction, cells were treated with 50uM Biotin.18-24 hours later, cells were washes twice with 1mL ice cold PBS and harvested by 120uL RIPA with proteinase inhibitor and cell lysates were transferred to an 2ml Eppendorf LoBind tube. Samples were then flash frozen with liquid nitrogen and transferring to -80 freezer. BioID sample preparation will be performed until all samples are done.

BioID sample preparation

Stored frozen samples were thawed on ice. Then 1uL benzonase nuclease (250U, Sigma E1014) was added to each sample and incubated for 1hr on an end-over-end rotator at 4 °C. Followed by sonication at 3X 10s on high, samples were centrifuged at 15,000rpm for 30min, 4°C. Supernatant was taken for later pull-down assay. Streptavidin-Sepharose beads (70uL beads per sample, GE ref# 17-5113-01) was balanced in RIPA buffer before added

to the supernatant sample above. Next, these samples were incubated for 3hrs at 4°C on an end-over-end rotator. After wash with 1mL RIPA buffer, another 1 ml of RIRA buffer was added to suspend the beads and then the whole mixture was transferred into a 1.5mL Eppendorf tube. The beads were then washed 6 times with 1mL 100 mM Triethylammonium bicarbonate (TEAB) (Sigma). Samples were centrifuged at 2000 rpm for 2 min between each washes. Finally, the beads were resuspended in 200uL 100 mM TEAB.

Trypsin Digestion and Tandem Mass Tag (TMT) Labeling

Enriched proteins on beads in each sample were treated with 2 μg Sequencing grade Trypsin (Promega) for overnight at 37 °C on an end-over-end rotator. 1 μg Trypsin was then added to each sample the next morning and incubated for another 2 hours at 37 °C. Next, beads were pelleted at 2000 rpm for 2 minutes and supernatant was then transferred to new 1.5mL Eppendorf LoBind tubes. Then 150 μL 100 mM TEAB was added twice to elute most of the digested peptides. The eluted samples were then combined with the previous supernatant samples and spun at 14000 rpm for 10 minutes. The resulting supernatants were then transferred to new 1.5mL Eppendorf LoBind tubes, leaving 20-30 μL at bottom of each tube to avoid transferring of residual beads. Next, samples were dried with speed vacuum for 2-3 hours and then resuspended with 30 μl of 100 mM TEAB. Aliquots of 10 μl from each sample were taken for TMT labeling. Separately, 5 μl from each sample were pooled thoroughly. And 10 μl of this pooled sample was used for each TMT batch, serving as a normalization control channel across different batches. In our experiment, three sets of TMT10plex™ Isobaric Label Reagent were used: biological

triplicates of 0N3,0N4 plus the luciferase-expressing negative control, together the normalization control; 1N3,1N4 plus the luciferase control and the normalization control; 2N3,2N4 plus the luciferase control and the normalization control. For TMT labeling step, the 10 μl aliquot from each sample was labeled with 4 μl of TMT reagents (0.08 mg) for 1 hour at room temperature, then the reaction was quenched by adding 4 μl of 5% hydroxylamine for 15 minutes. Within each batch of TMT10plex labeling, 10 channels were combined before speed vacuum drying. Finally, three batches of TMT10plex samples were each resuspended in 20 μl 5% acetonitrile-0.1% formic acid for LC-MS/MS analysis.

Liquid Chromatography and Tandem Mass Spectrometry (LC-MS/MS)

From each of the three TMT10plex sample batches, 1 μl out of 20 μl was injected for LC-MS/MS analysis. Liquid chromatography was conducted using a home-made trap-column (5 cm x 200 μm inner diameter) and a home-made analytical column (110 cm x 50 μm inner diameter) packed with Reprosil-Pur 120 C18-AQ 5 μm particles (Dr. Maisch), running a 360-min reversed-phase gradient at 70nl/min on a Thermo Fisher Ultimate 3000 RSLCNano UPLC (Ultra-Performance Liquid Chromatography) system coupled to a Thermo QExactive HF quadrupole-Orbitrap mass spectrometer. A parent ion scan was performed using a resolving power of 120,000 and then up to the 30 most intense peaks were selected for MS/MS (minimum ion counts of 1000 for activation), using higher energy collision induced dissociation (HCD) fragmentation. Dynamic exclusion was activated such that MS/MS of the same m/z (within a range of 10ppm; exclusion list size=500) detected twice within 5s were excluded from analysis for 50s.

Proteomic Data Analysis

LC-MS/MS raw files were searched using Proteome Dis-coverer V2.2 (Thermo Fisher) against the SwissProt mouse database (released in June 2017), with the following parameters: full tryptic specificity with up to two missed cleavage sites, carbamidomethylation of cysteine residues, along with TMT labeling of peptide N-termini and lysine residues as fixed modification, and protein N-terminal acetylation plus methionine oxidation as variable modifi-cations. The false discovery rate was set to 0.01. In total, 2570 proteins were identified in these three LC-MS/MS runs. Separately, these files were searched again against the SwissProt human proteome database (released in June 2017) to identify human Tau specific peptides. The search parameters are the same as above.

Human Tau specific peptide, DGTGSDDKK, located at the common region of the Tau isoforms was identified and quantified in all three LC-MS/MS runs and was used to evaluate the Tau expression levels within biological triplicate of each Tau isoform. Within biological replicates for each Tau isoform BioID MS data, any channel(s) with the TMT reporter intensity of the specific Tau peptide over 1.5-fold of the median value within the group was disqualified for further analysis. Based on the above criteria, biological triplicate No.1 of 0N4R in TMT10plex batch 1 (~1.68 fold), biological triplicate No.2 of 1N4R in TMT10plex batch 2 (~1.63 fold) and biological triplicate No.2 of 2N3R in TMT10plex batch 3 (~2.62 fold) were removed. In the adjusted dataset, the total TMT reporter intensities across channels in each batch of the TMT10plex reaction were within a 20% variation (Figure S1), indicating that there was not much difference within the total amount of enriched proteins in the BioID procedure. Next, within each TMT10plex batch, TMT reporter values in each TMT channel was corrected based on the ratio of the total TMT

reporter intensity of this channel and the total TMT reporter intensity of the TMT-126 channel. Interactome difference between 0N3R and 0N4R, 1N3R and 1N4R, 2N3R and 2N4R were compared within batches. For proximity proteome data analysis of each Tau isoform, ratios of the mean TMT reporter intensity for each identified protein within the Tau isoform group to the corresponding mean TMT reporter intensity in the luciferase control group, along with and mean TMT reporter intensity values in the Tau isoform group, were submitted to the Perseus Software[321] to calculate the significance B values (< 0.05). Meanwhile one tailed Student's t-distribution p-values were calculated for each identified protein (< 0.05). To be noted that proteins didn't pass both sigB and p-value criteria in original dataset were considered as false positive hits and were removed in final protein hits identified in each Tau isoform BioID pull down.

Co-immunoprecipitation (Co-IP)

Transfected cells were lysed using ice cold co-IP buffer (50mM Tris-HCl, 150 mM NaCl, 1% NP40, 10 mM EDTA, pH=7.4) and incubated for 20 mins. Followed by sonication at 8X 5s on high, samples were centrifuged at 15,000 rpm for 30min, 4°C. 10% of Supernatant was taken as the input sample, the rest of supernatant was saved for later IP assay. Magnetic Flag beads (10uL beads per sample, Sigma) were balanced in co-IP buffer before added to the saved supernatant sample above. Next, these samples were incubated overnight at 4°C on an end-over-end rotator. After wash with 1mL co-IP buffer for 3 times, beads were boiled in 1x LDS loading buffer for 15 mins at 95°C. After pelleted the beads using magnetic rack, the supernatant was saved as IP samples.

Western Blotting

Same procedure as described in Chapter 2.

Bioinformatic Analysis

The functional enrichment analysis including gene ontology (GO) and KEGG pathway enrichment were performed using DAVID functional enrichment tool[322] with Benjamini-Hochberg FDR correction method applying significance threshold of 0.05. Pathway analysis for total Tau proximal proteins were performed using Reactome FI APP in Cytoscape (version 3.82). The genes that enriched in similar pathways were clustered together with a Benjamini-Hochberg FDR less than 0.05.

3.4 Results

Human Tau isoform-specific BioID2

The six Human CNS Tau isoforms were cloned into pLV-hSyn-TurboGFP-P2A-BioID2 plasmid as shown in Figure 1A, Human synapsin 1 gene promoter in the plasmid ensures neuron-specific gene expression. pLV-hSyn-TurboGFP-P2A-luciferase-P2A-BioID2 was used as a BioID2 background control. Within LTR sequence of the vector, the week promoter can ensure the low expression of insert genes. With that, HEK 293 FT cells were transfected with Tau-BioID2 constructs to examine their expression. As shown in Figure 1B, the expressed proteins were probed by Flag antibody (Sigma) and Tau Monoclonal antibody (HT7) and confirmed by their molecular size, ~ 65 to 75 kDa.

Primary cortical neurons were purified from embryonic brain of CD-1 (E16) mouse and seeded at 600,000 cells per well in a 12-well cell culture plate. As shown in Figure 2, one 12-well plate is used for one condition of BioID2 experiments. There are three groups of experiments and three condition of BioID2 experiments within each group, which

including: 1) 0N3R-BioID2, 0N4-BioID2 plus BioID2 control; 2) 1N3R-BioID2, 1N4-BioID2 plus BioID2 control; 3) 2N3R-BioID2, 2N4-BioID2 plus BioID2 control. BioID2 was performed with three biologically independent mice for each condition of BioID2 experiments. To perform BioID2 experiment, primary neurons were cultured in neurobasal medium plus B27 to 2 weeks (DIV 14), then lentivirally transduced with Tau BioID2. At DIV 17, cells were treated with 50 uM Biotin for 22 hours before harvesting. Meanwhile, expression and localization of Tau-BioID2 was confirmed by immunofluorescence, as shown in Figure 3. Streptavidin pull-down was performed at the same time once all conditions of samples were harvested. In the next step, the pulled down elutes from each samples within each group (nine samples total with three biological replicates) were labeled with the first nine channels (126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C) of TMT10plex (Figure 2).

Identification of Tau-interacting Proteins by Mass Spectrometry

After combining TMT labeled samples in each group as shown in Figure 2, we performed LC-MS/MS for these TMT labeled samples (Table S1). With the TMT-based LC-MS/MS experiments, we were able to identify Tau isoform specific interacting proteins and quantify their binding affinity. Through our analysis, we identified 2570 proteins in total through three batches of TMT-labeled MS runs. Given the compressed intensity ratio caused by TMT chemical background issues, Significance B values (SigB)[323] were generated to describe protein ratio to select the significant hits as interactors (as shown in method). Meanwhile one tailed Student's t-distribution p-values were calculated for each identified protein among three biological triplicate experiments. Based on the threshold of less than

0.05 for both SigB and p-values, we detected 38 proteins as 0N3R proximal proteins; 92 proteins as 0N4R proximal proteins; 41 proteins as 1N3R proximal proteins; 36 proteins as 1N4R proximal proteins; 33 proteins as 2N3R proximal proteins; 66 proteins as 2N4R proximal proteins (Table S2). Identified proteins for each Tau isoforms were shown in Figure 4, red dots indicated the protein that passed above criteria and top 10 significant interactors for each isoforms were labeled in the graph as well. As shown in Figure 4, microtubule associated proteins (such as MAP2 and MAP4) were always identified as the most significant interactors for each isoform, which is consistent with the fundamental role of Tau that acts as a microtubule binding protein.

With the significant protein hits identified in the above Tau isoform-specific BioID2 experiments, we generated proximal proteome networks for all human six Tau isoforms (Figure 5). Through combining the proximal proteins from each of the Tau isoforms, we in total identified 187 Tau proximity proteins that may represent as the major component of Tau interactome network. By overlapping with the previously reported Tau interactors that generated by Tau antibody-based affinity purification in mice brain[151], we found 8 proteins that were also identified in our BioID2 experiments (Figure S2). Moreover, of the 187 protein hits, we found several risk genes that linked to neurodegenerative diseases, which include AD associated genes: BIN1[324] and KNCD1[325]; PD associated genes: SNCA[326] and NOS1[327]; FTLD-ALS associated genes: VCP and SQSTM1[328]. Besides AD, which is a secondary tauopathy, pathological Tau was also identified as one of the clinical features in high population of ALS and PD. Therefore, the discovery of the potential association of

Tau with their ALS and PD risk genes might provide clues to explain the implicated role that Tau plays in the disease progression of neurodegenerative diseases.

Functional Enrichment Analysis for Tau Interacting Proteins

To further characterize these 187 proteins that were identified in close proximity to Tau by our BioID2 experiments, we performed gene ontology (GO) analysis using the DAVID functional enrichment tool[322]. Through GO analysis under the categories of cellular component, biological process and molecular function, we found Tau proximal proteins shows top enrichment in cytoskeleton and microtubule and regulates microtubule functions (Figure 6A-C), which further validates the sensitivity and specificity of our Tau BioID2 methods. To further characterize the functional network of Tau proximal proteins, we clustered these proteins using the Reactome FI app in Cystoscope, followed by pathway enrichment. Through these analyses, we found besides cytoskeleton regulation, these Tau potential interactors may also play a role in endocytosis, protein regulation, synaptic transmission and neurodegeneration pathways (Figure 7).

Functional Enrichment Analysis for 3R- and 4R-Tau proximal proteins

To investigate the interactome differences between Tau isoforms that differ in the C-terminal microtubule binding repeats, we generated the proximal protein lists for 3R- and 4R- Tau (significant protein hits identified in at least one isoform for either 3R- or 4R-Tau in our BioID2 experiments). These two protein lists were then uploaded to DAVID functional enrichment tool[322] for KEGG functional pathway analysis. We found that 3R-Tau proximal proteins are significantly enriched in fatty acid metabolic pathways whereas 4R-Tau proximal proteins are specifically involved in endocytosis (Figure 8). The

functional discrepancy between 3R- and 4R-Tau might provide additional evidence to explain the intricate role that distorted expression of 3R-/4R-Tau plays in the pathogenesis of neurodegenerative disease.

MAT2A shows preferential binding with 3R-Tau compared to 4R-Tau

To identify the pan-4R-Tau interactors, we overlapped the protein list identified from all three 4R-Tau isoforms. The majority of the significant protein hits shared by all 4R-Tau isoforms are microtubule binding proteins, which confirms previous reports suggesting that 4R-Tau has higher binding affinity with microtubules (Figure 9). As for pan-3R-Tau protein hits, we found MAT2A and GLUL are the only two proteins that shown higher TMT ratio in 3R-Tau BioID2, suggesting these two protein may have selective binding affinity with 3R-Tau (Figure 10). GLUL is a glutamine synthetase, which has been identified as a candidate interactor that may preferentially binding to 0N-Tau [151], however, its interaction with Tau has yet to be validated. Here, we performed Co-IP assay in HEK 293FT cells and demonstrated that GLUL can bind with Tau isoforms (Figure S3). MAT2A is a methionine adenosyltransferase that catalyzes the formation of S-adenosylmethionine (SAM), which is the key methyl donor that used for methylation of DNA, RNA and proteins [329]. Through both Co-IP and reverse IP validation, we found that MAT2A exhibited stronger binding affinity with 3R-Tau relative to 4R-Tau, which is also consistent with our MS data (Figure 11).

Functional Enrichment Analysis for 0N-, 1N-, 2N-Tau proximal proteins

To compare the interactome difference among Tau isoforms that differ in N-terminus (0N-, 1N-, and 2N-Tau), the 1504 overlapping proteins (~75% of total proteins) that present in

all three MS runs (Figure 12A) was used to re-generate significant protein hits for each isoforms. Thanks to the use of TMT tags, we were able to avoid the “missing value” issue that usually occurs in typical label-free proteomics experiments across multiple LC-MS/MS runs. In this case, we observed 129 Tau interacting proteins including 29 0N3R interactors, 71 0N4R interactors, 30 1N3R interactors, 31 1N4R interactors, 25 2N3R interactors and 50 2N4R interactors. Next, we generated the sub-protein lists based on the N-terminus of Tau isoforms. Based on the similar idea to generate 3R- and 4R-Tau proximal proteins, here, we found 80 0N-Tau proximal proteins, 47 1N-Tau proximal proteins, 58 2N-Tau proximal proteins (Figure 12A). With the above protein lists, we performed GO analysis and found that 0N-Tau proximal proteins are widely distributed in neurons with the enriched localization in microtubule, axon, dendrites, synapse, synaptic vesicles, and membrane whereas proteins in proximity with 1N-Tau shown specific localization in mitochondrial compared to 0N- and 2N-Tau (Figure 12B). With biological process analysis, we observed that besides fundamental role of microtubule organization, the three groups of proteins all have specific enriched GO items. As shown in Figure 12C, 0N-Tau proximal proteins have specific enrichment in cell-cell adhesion and neuron projection development while 1N-Tau proximal proteins are specifically involved in fatty acid metabolic process and 2N-Tau proximal proteins are associated with drug response. Although we only identified 47 proteins in proximity with 1N-Tau, this protein list is highly enriched in several pathways, such as fatty acid metabolism and cell signaling pathways, through KEGG functional analysis (Figure 12D). Conversely, none of the pathways were

enriched for 0N- and 2N- Tau proximal proteins. Based on these findings, we propose that 1N-Tau might play an important role in fatty acid metabolism and signaling transduction.

3.5 Discussion

In this study, we combined biotin ligase-based proximity labeling (BioID2) and TMT based quantitative proteomics to profile the interactome of six human CNS Tau isoforms. In our experiments, a humanized mutant form of biotin ligase, named BioID2, was genetically fused with each of the six human Tau isoforms and lentivirally transduced to primary neurons to identify the interacting proteins. Compared to traditional antibody-based affinity purification (AP), enzyme-based proximity labeling (PL) greatly reduced the false positive interactions produced by prior cell lysis in AP. In addition, compared to antibody based-AP, enzyme-based PL is available to capture transient and week protein-protein interaction. Known studies to unveil Tau interactome were limited to antibody- based AP, whose intrinsic false positive issues could make the data interpretation less accurate to recapitulate the Tau interactome in the *in situ* physiological conditions. In addition, previous work was based on either the full-length human Tau or three adult mice Tau isoforms[151, 312]. Hence, the interactome study for all human Tau isoforms present in CNS is still lacking, which is the key knowledge to unveil the functional difference among Tau isoforms and the underlying mechanisms may help to discover novel therapeutics for Tauopathy treatment. In this study, we depicted the interactome network for all six human Tau isoforms in living neurons under native states. The novel Tau protein interactors that identified in this study might compensate previous findings. Also, the identification of genetic risk factors

annotated in other neurodegenerative diseases such as PD and ALS/FTLD might explain the implicated role that Tau plays in these related brain disorders.

Distorted 4R/3R-Tau ratio is believed as one of the causal factors for pathogenesis of a subset of Tauopathies. In this study, we were able to identify 89 and 146 proteins that are in proximity with at least one of 3R-Tau and 4R-Tau isoforms, respectively. Through GO analysis, we found that the localization of 3R- and 4R-Tau proximal proteins were not completely identical. Specifically, 3R-Tau proximal proteins are specifically enriched in mitochondrial, while proteins that biotinylated by 4R-Tau-BioID2 are highly enriched in cell junction and synaptic vesicles. The distinction for localization determined their functional differences. For example, in this study, we found 3R-Tau proximal proteins are enriched in fatty acid metabolism whereas proteins identified in 4R-Tau BioID2 systems were specifically enriched in endocytosis pathway. It has been reported that pathological Tau induces presynaptic dysfunctions in AD mouse model[330], which indicates that Tau might function to regulate endocytosis, a crucial biological process for synaptic transmission. A recent study reported that overexpression of human Tau in mouse cortical neurons impairs endocytosis through miR-132/MeCP2/dynamin 1 pathway[331]. The Tau isoform that used in their study is the full-length Tau (with 4 microtubule binding repeats), which is consistent with our findings that 4R-Tau is associated with endocytosis. It has been reported that Tau aggregation can be induced by fatty acid, however the underlying mechanism is still unclear[332]. Here we established the potential functional linkage between 3R-Tau (as well as 1N-Tau) and fatty acid metabolism, which lays the foundation for future studies to decipher the associated mechanisms.

Thanks to our TMT-based experimental design (Figure 2), we were able to quantitatively compare the number of individual biotinylated proteins identified by mass spectrometer in 3R- and 4R-Tau BioID2 system. Based on the quantitively data and further validation, we confirmed that MAT2A exhibited preferential binding with 3R-Tau compared to 4R-Tau. MAT2A functions to catalyze the biosynthesis of S-adenosylmethionine (SAM) from L-methionine and ATP[333]. SAM provides methyl group for methyl transfer reactions of DNA, RNA and proteins that catalyzed by methyltransferase[334]. Thus, MAT2A plays an important role in global epigenetic regulation and biological processes such as cell proliferation and cell cycle. It has been well documented that dysregulation of MAT2A has been observed in many types of cancer, such as liver cancer, colon cancer and leukemia[334-336]. For example, in hepatocellular carcinoma, some studies have been reported that the pathogenic switch of hepatic MAT isoform MAT1A to extrahepatic isoform MAT2A induced the decreased SAM level[329, 337]. MAT2A has been used as a therapeutic target in cancer and several MAT2A inhibitors such as PF-9366[338] and AG-270[339] has been subsequentially developed to alleviate cancer proliferation. Accumulating evidence have shown that epigenetic dysregulation is also implicated in neurodegenerative diseases. Moreover, overt reduction of SAM has been observed in brain tissues and cerebrospinal fluid (CSF) of AD patients versus age matched control[340]. However, the functional role of MAT2A plays in nervous system was poorly defined. It has been reported that the expression level of MAT2A is significantly reduced in AD postmortem brains samples versus age matched control[341]. In addition, a recent CRISPR (Clustered-regularly interspaced short palindromic repeats) screening study identified

MAT2A as an essential gene for iPSC-derived neurons instead of iPSC, which only few genes were changed upon MAT2A depletion[342]. Through RNA sequencing, they found they found the genes that modulated by MAT2A are enriched in synaptic functions[342], which indicates the neuron specific functions for MAT2A. Moreover, they found loss of MAT2A dramatically reduced the expression of microtubule associated proteins with Tau as the top downregulated genes[342]. In our study, we identified the preferential binding of MAT2A with 3R-Tau, it is possible that 3R-Tau and 4R-Tau function differently to modulate MAT2A activity to further influence the global epigenetic landscape and the subsequentially changes for gene expression profiling and cellular function. Another possibility is that 3R-Tau might function as a scaffold protein to connect MAT2A with its target proteins to modulate their methylation levels, which may further impact their cellular function.

There has been limited efforts to compare the functional differences among Tau isoforms differing in N-terminus. In 2016, Liu et al. performed Co-IP in adult mice brain tissues (with 0N4R, 1N4R and 2N4R Tau) and demonstrated a putative role of 2N-Tau in disease[151]. In our study, we didn't observe the enrichment of 2N-Tau proximal proteins in disease pathway. Conversely, we found 1N-Tau proximal proteins are enriched in many pathways such as fatty acid metabolism and signaling transduction. This inconsistent findings may result from the different systems that have been used to unveil the protein-protein interaction. Moreover, their experiment was designed to capture interactors of mouse Tau whereas our study was aiming to reveal the interactome of human Tau isoforms.

Taken together, our study unveiled the proximity proteome network for individual human Tau isoforms that present in CNS and demonstrated that 3R-Tau and 1N-Tau has a putative role in fatty acid metabolism whereas 4R-Tau has potential role in endocytosis pathway. Specifically, we identified MAT2A as a novel Tau interactor that shown preferential binding affinity with 3R-Tau versus 4R-Tau.

3.6 Main figures

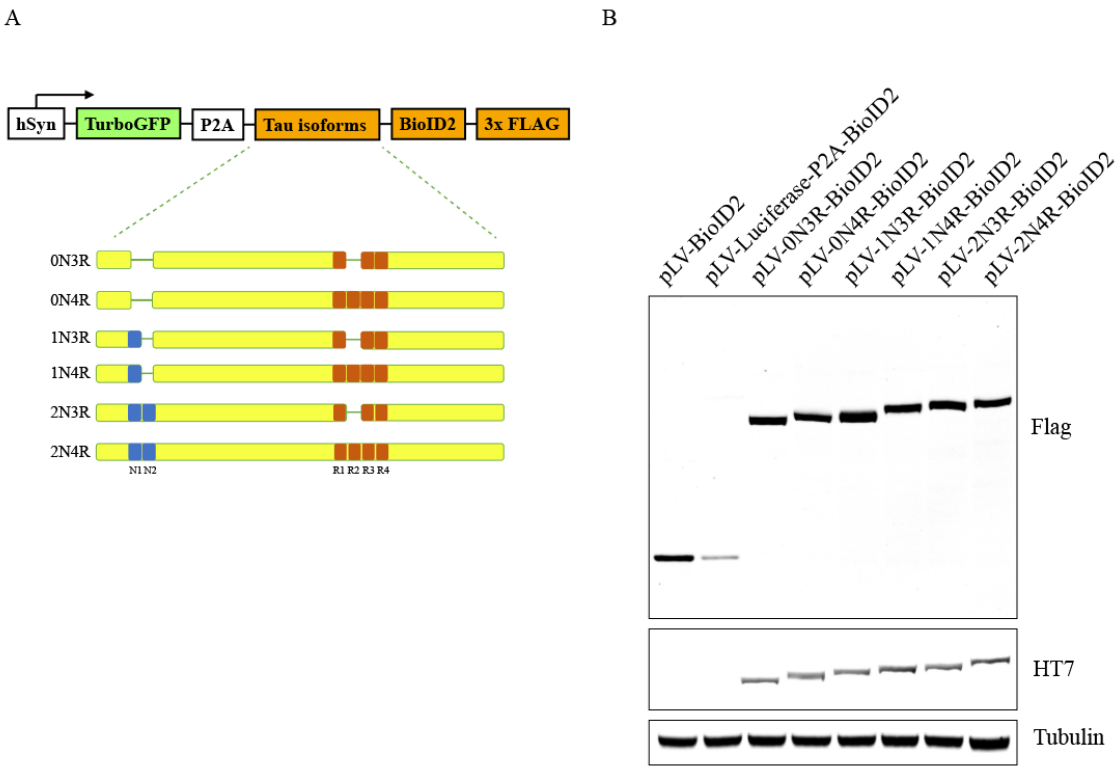


Figure 1: Expression of Tau BioID2. A) Schematic of Tau BioID2 constructs. B) 2ug of Tau BioID2 plasmids plus BioID2 control were transfected into HEK293 FT cells, cells were lysed after 48 hours, and cell lysate were then probed with Flag and HT7 (Tau) antibody. Tubulin is probed as a loading control.

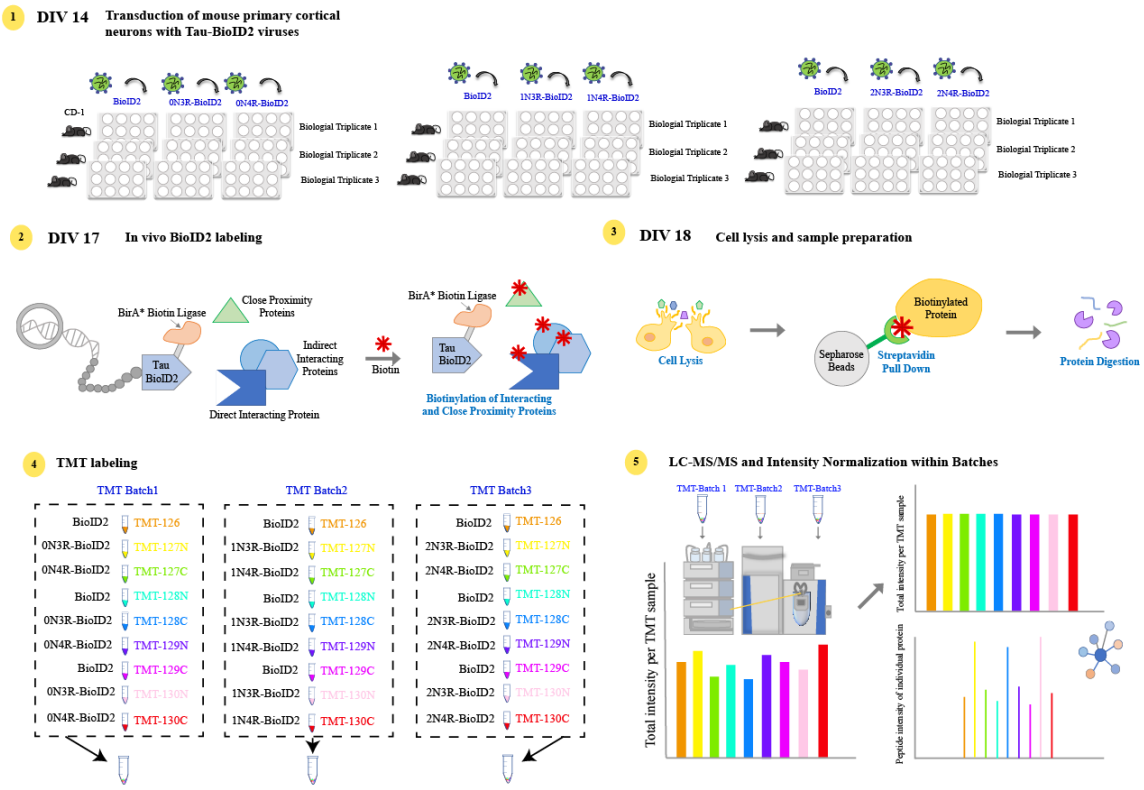


Figure 2: Schematic diagram of *in vivo* Tau BioID2 labeling and quantitative proteomics workflow. 1) Lentivirally transduction of Tau BioID2 into primary cortical neuron at DIV14. 2) Addition of Biotin for 22 hours to induce Tau BioID2 labeling at DIV17. 3) At DIV18, cells were lysed. And streptavidin beads pull down were performed in parallel once all samples were harvested. Elutes were then digested into peptides by trypsin. 4) Isobaric tandem mass tag (TMT) labeling of each pulled down samples as indicated, labeled samples from each batch were mixed for MS analysis. 5) Three batches of TMT labeled samples were separated by liquid chromatography (LC), followed by a measurement by mass spectrometry (MS) and quantification of relative protein intensities by tags.

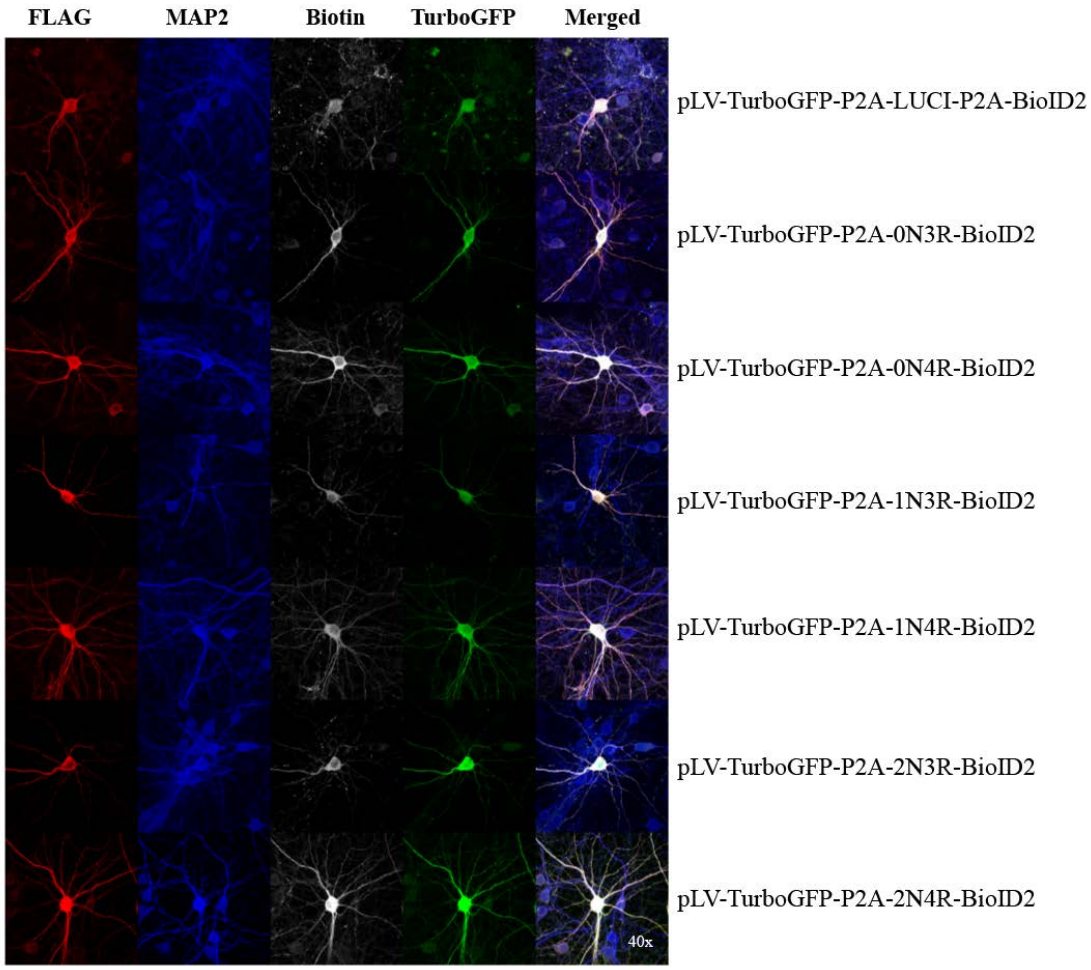


Figure 3: Expression of Tau BioID2 in primary mouse neuron at DIV 18. Primary neurons are transduced with Tau BioID2 lentivirus and incubated with Biotin at DIV 17 and fixed on DIV18. Immunofluorescence was performed with Flag, Map2, Biotin and TurboGFP antibody (This experiment was done by Nadeem Murtaza from Dr. Karun Singh's lab).

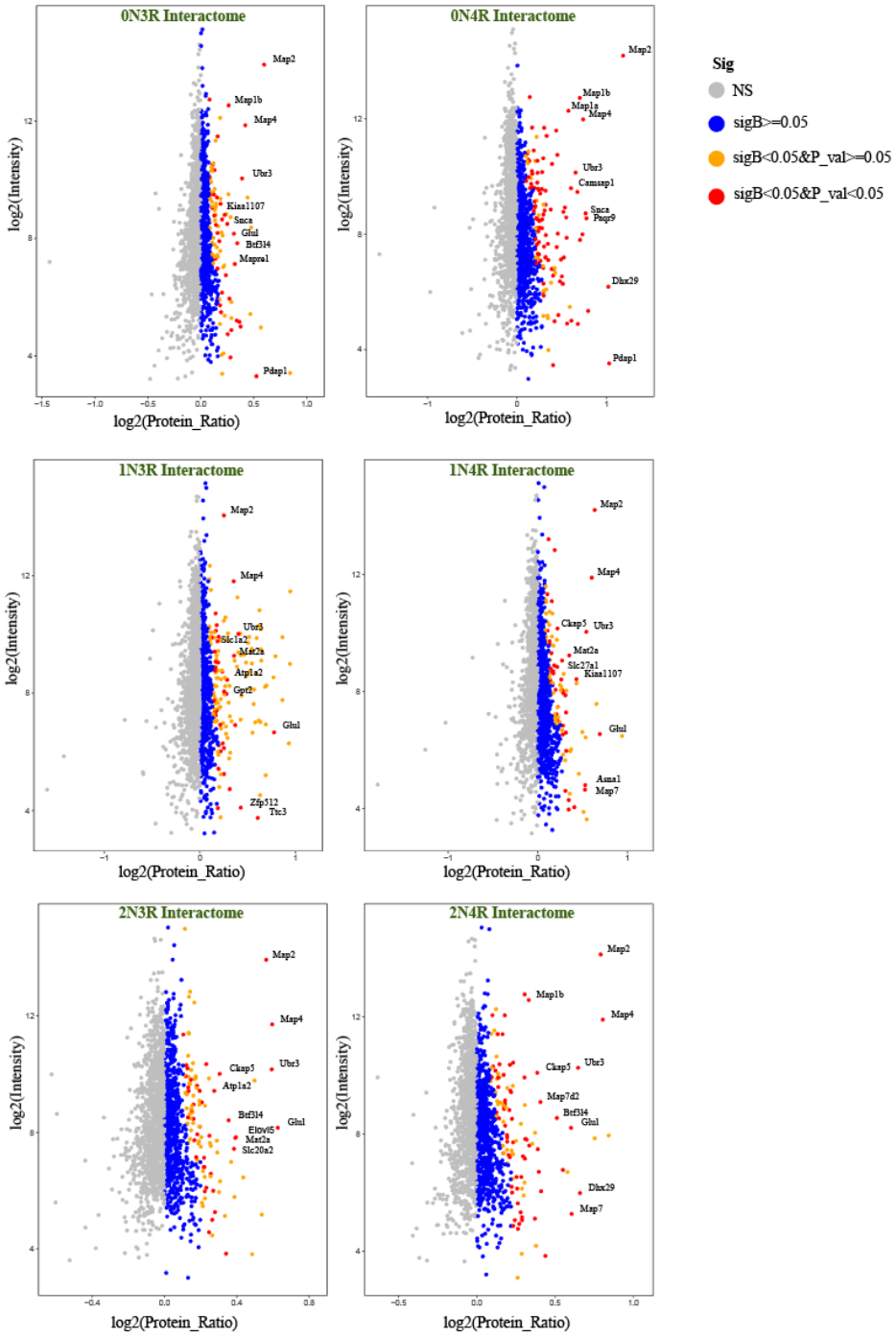


Figure 4: Proteome-wide quantification of identified proteins in Tau BioID2. Protein ratios (log2) are plotted against mean values of summed peptide intensities (log2). For the left side, gray dots (NS) marked the protein hits with ratio less or equal to 1 compared to

BioID2 control. For the right sites, the data points are colored by their sigB and Pval values, with blue dots having sigB > =0.05, yellow dots having sigB < 0.05 and Pval > = 0.05, and red dots having sigB < 0.05 and Pval < 0.05.

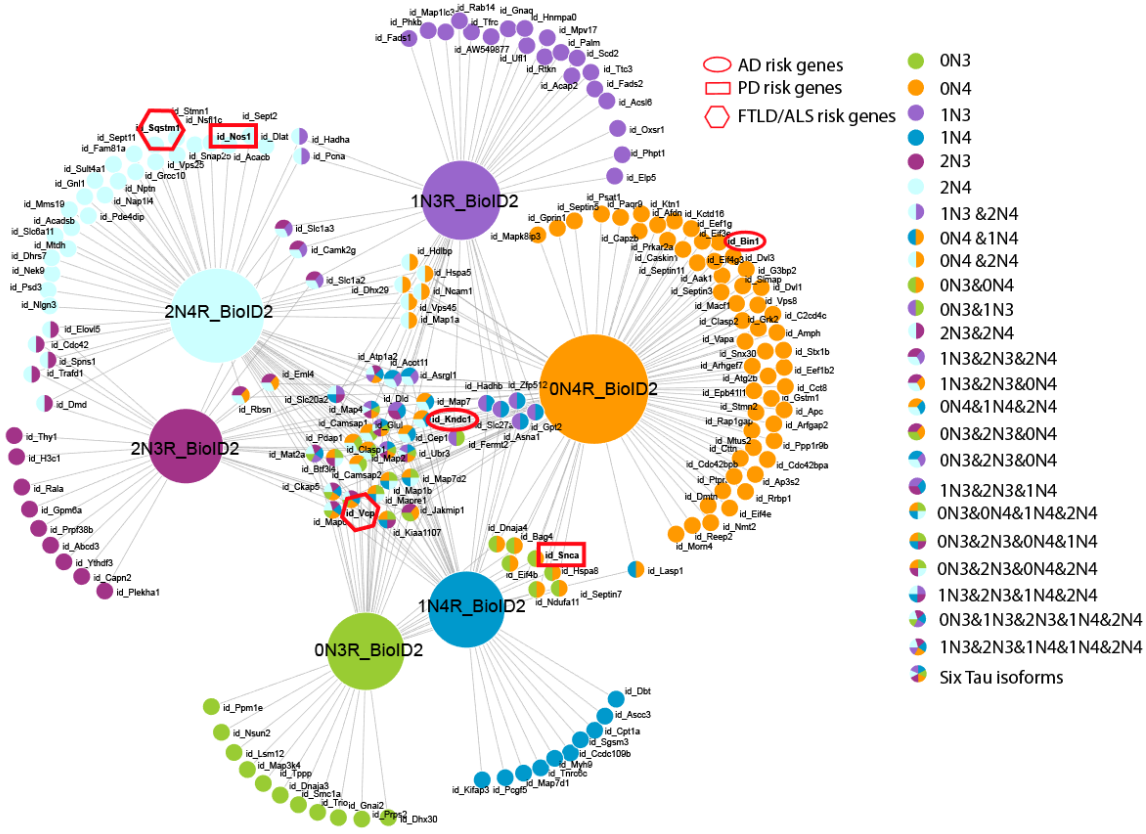


Figure 5: Human Tau isoform-specific interactome network. Biotinylated proteins identified from six Tau isoform BioID2 system. As shown in the legend, proximal proteins for each Tau isoforms were labeled with specific color. The overlapped proteins were labeled with the corresponding mixed color. Published risk genes from AD, PD, FTLD/ALS were highlighted with red box in the diagram.

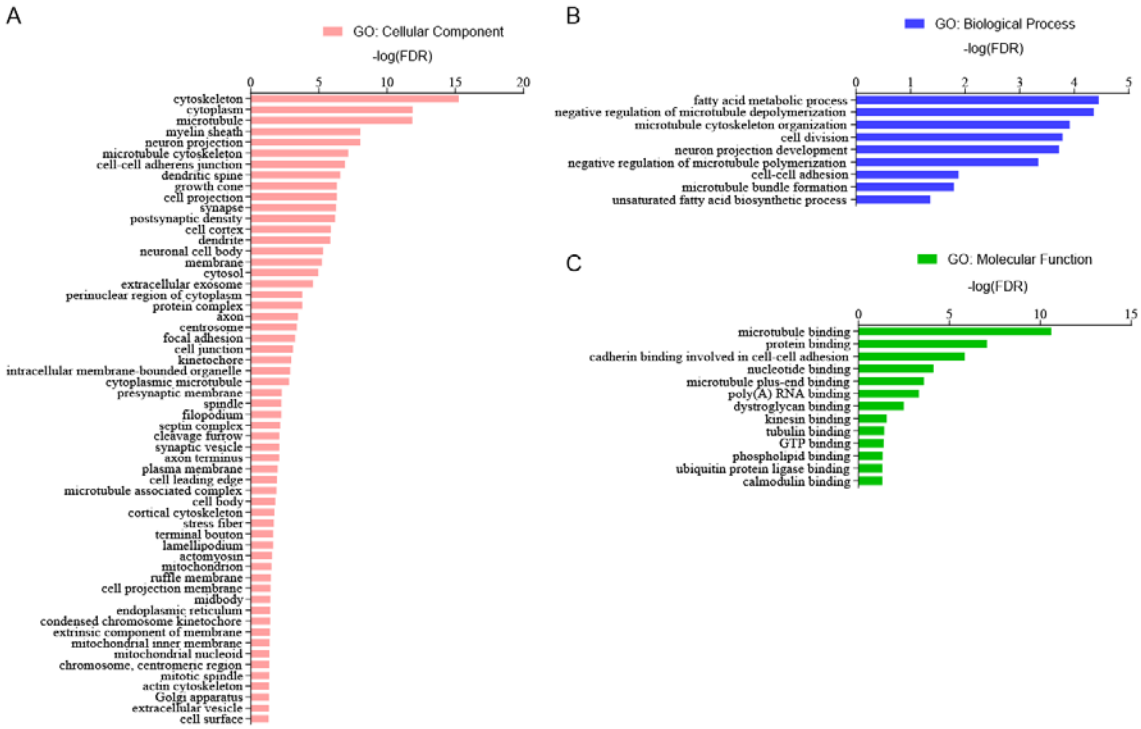


Figure 6: Gene Ontology analysis for total Tau proximal proteins. A-C) Proximal protein identified in all six Tau isoform BioID2 systems were uploaded to DAVID for gene ontology analysis under the categories of cellular component, biological process, and molecular function. The enrichment items with FDR less than 0.05.

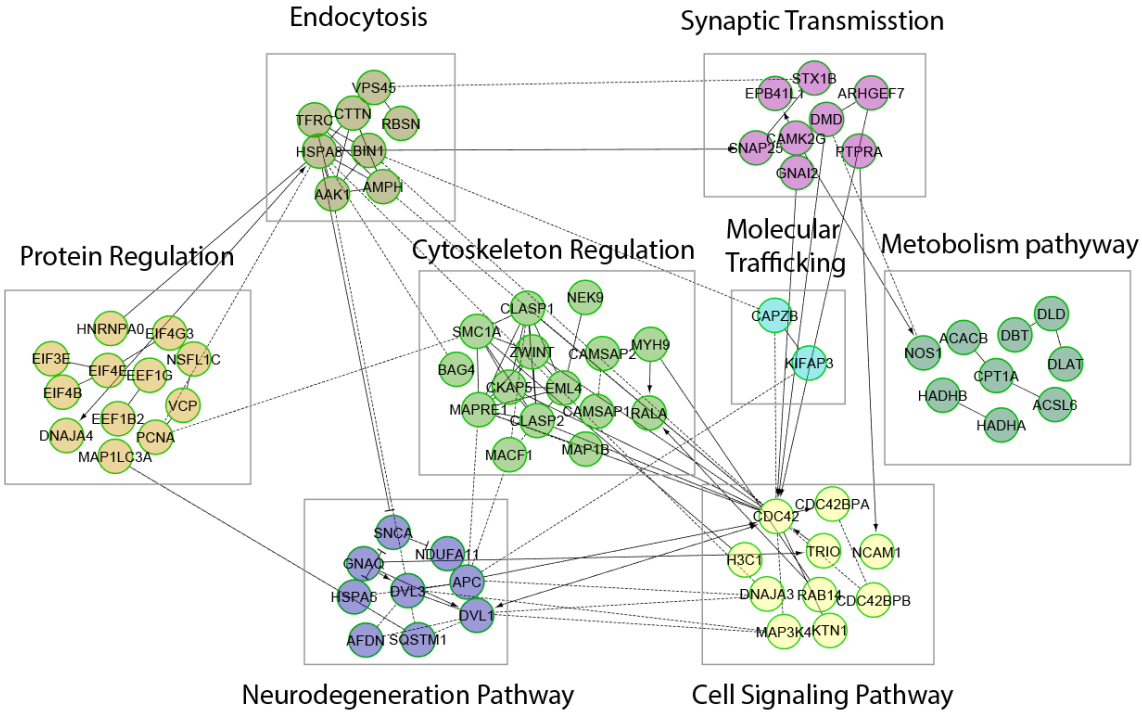


Figure 7: Functional module analysis for Tau proximal proteins. Total Tau proximal protein list was uploaded to cystoscope and analyzed with Reactome FI app. Genes enriched in similar pathways were manually clustered together.

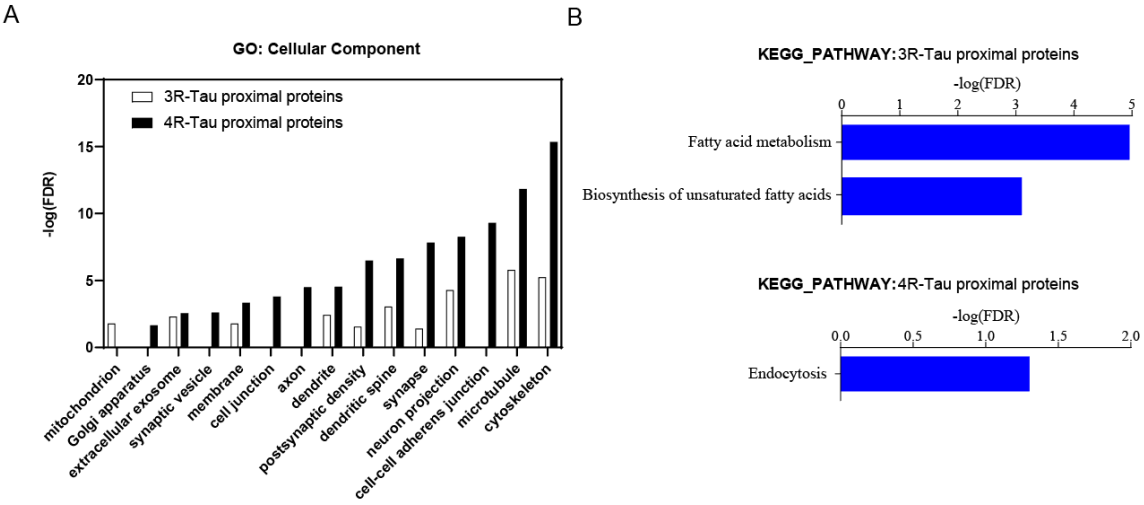


Figure 8: Functional distribution of proteins biotinylated by 3R- and 4R-Tau BioID2. 3R- and 4R-Tau proximal proteins refer to the biotinylated proteins that identified in at least one isoform for either 3R- or 4R-Tau with the BioID2 systems. A) cellular component analysis of the 3R- and 4R-Tau proximal proteins. B) KEGG pathway enrichment analysis of 3R- and 4R-Tau proximal proteins. Enriched items with FDR less than 0.05 were listed.

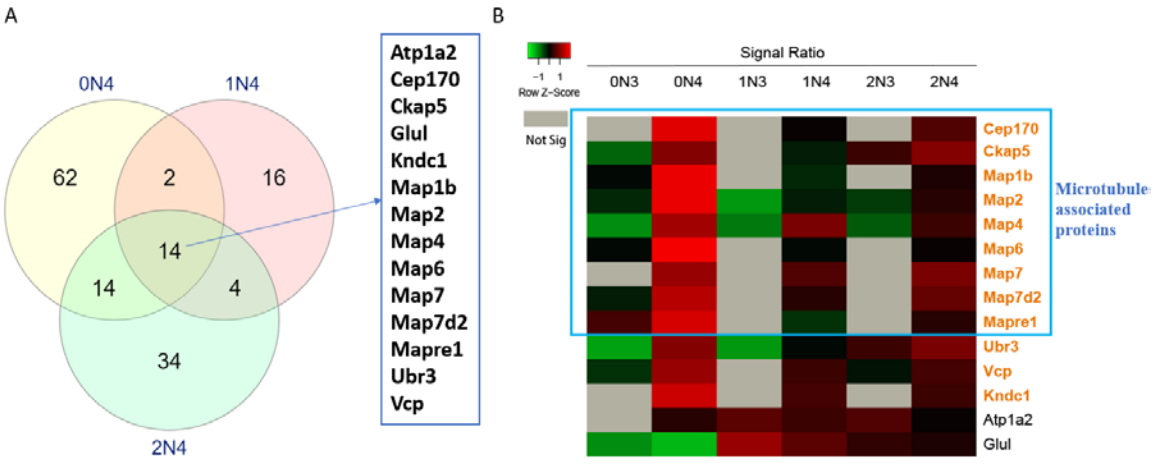


Figure 9: Overlap of 4R-Tau proximal proteins. A) Venn diagram showing overlapped proteins that identified in 0N4R-,1N4R- and 2N4R- BioID2 systems. B) heatmap showing TMT signal ratio of these overlapped proteins in individual Tau BioID2 systems. Microtubule associated proteins were highlighted with blue box. Proteins that shown higher signal ratio in 4R-Tau BioID2 labeling compared to 3R-Tau were highlighted in orange.

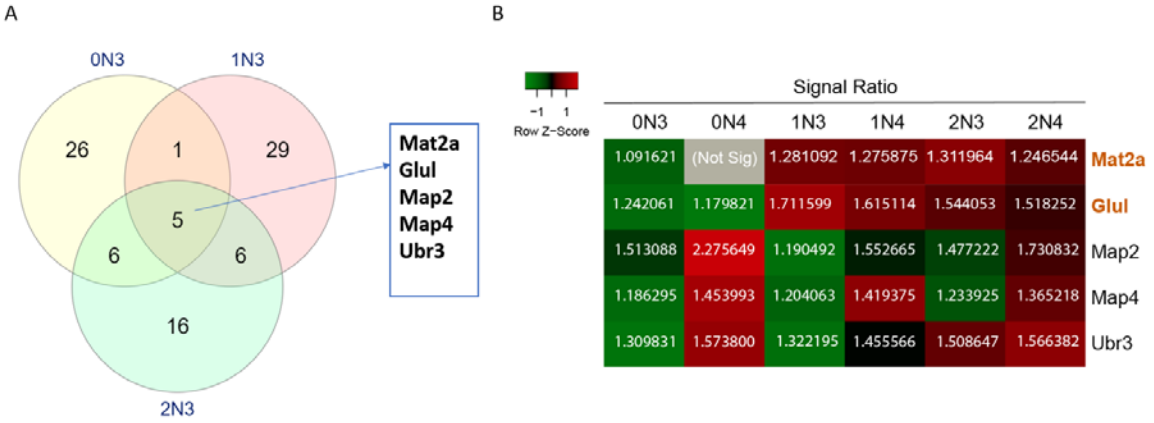


Figure 10: Overlap of 3R-Tau proximal proteins. A) Venn diagram showing overlapped proteins identified in 0N3R-,1N3R- and 2N3R- BioID2 systems. B) heatmap showing TMT signal ratio of these overlapped proteins in individual Tau BioID2 systems. Proteins that shown higher signal ratio in 3R-Tau BioID2 labeling compared to 4R-Tau were highlighted in orange.

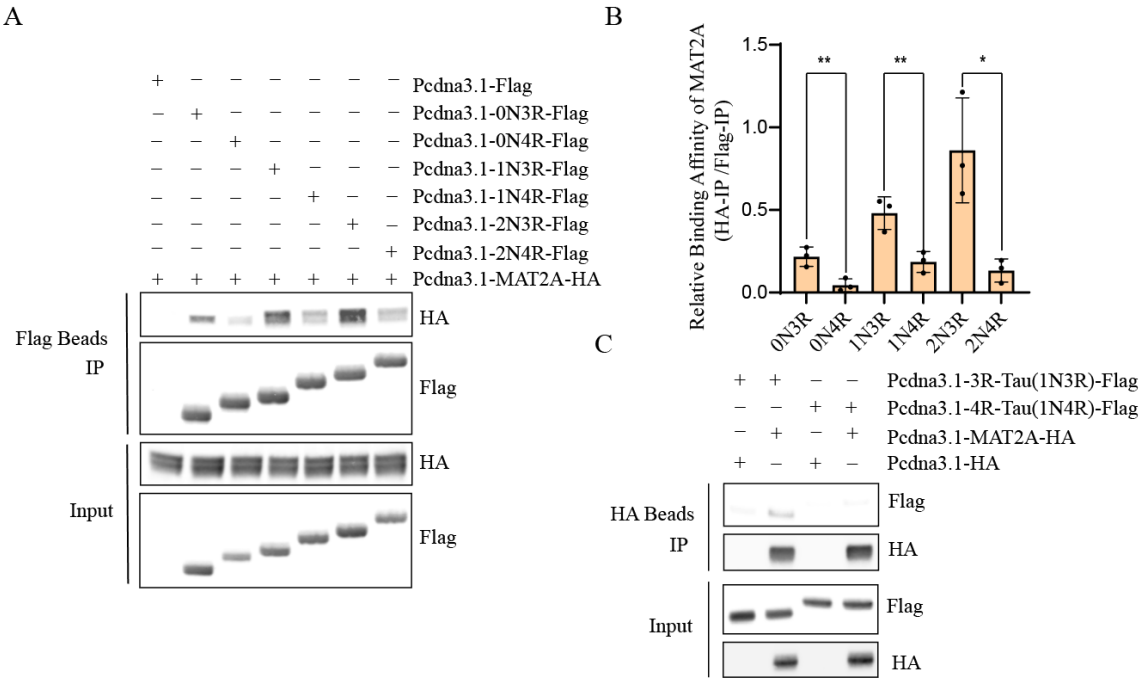


Figure 11: Validation of interaction between MAT2A and Tau isoforms by co-immunoprecipitation. A) MAT2A interacts with Tau isoforms in HEK 293FT cells through Flag beads IP. B) 3R-Tau shows significantly higher binding affinity with MAT2A compared to 4R-Tau through quantitative analysis of Co-IP assays. Data shown represent the means ± SD of three biological independent experiments (*$p < 0.05$, **$p < 0.01$; Student's t test). C) Reverse Co-IP confirmed the preferential binding of 3R-Tau for MAT2A.

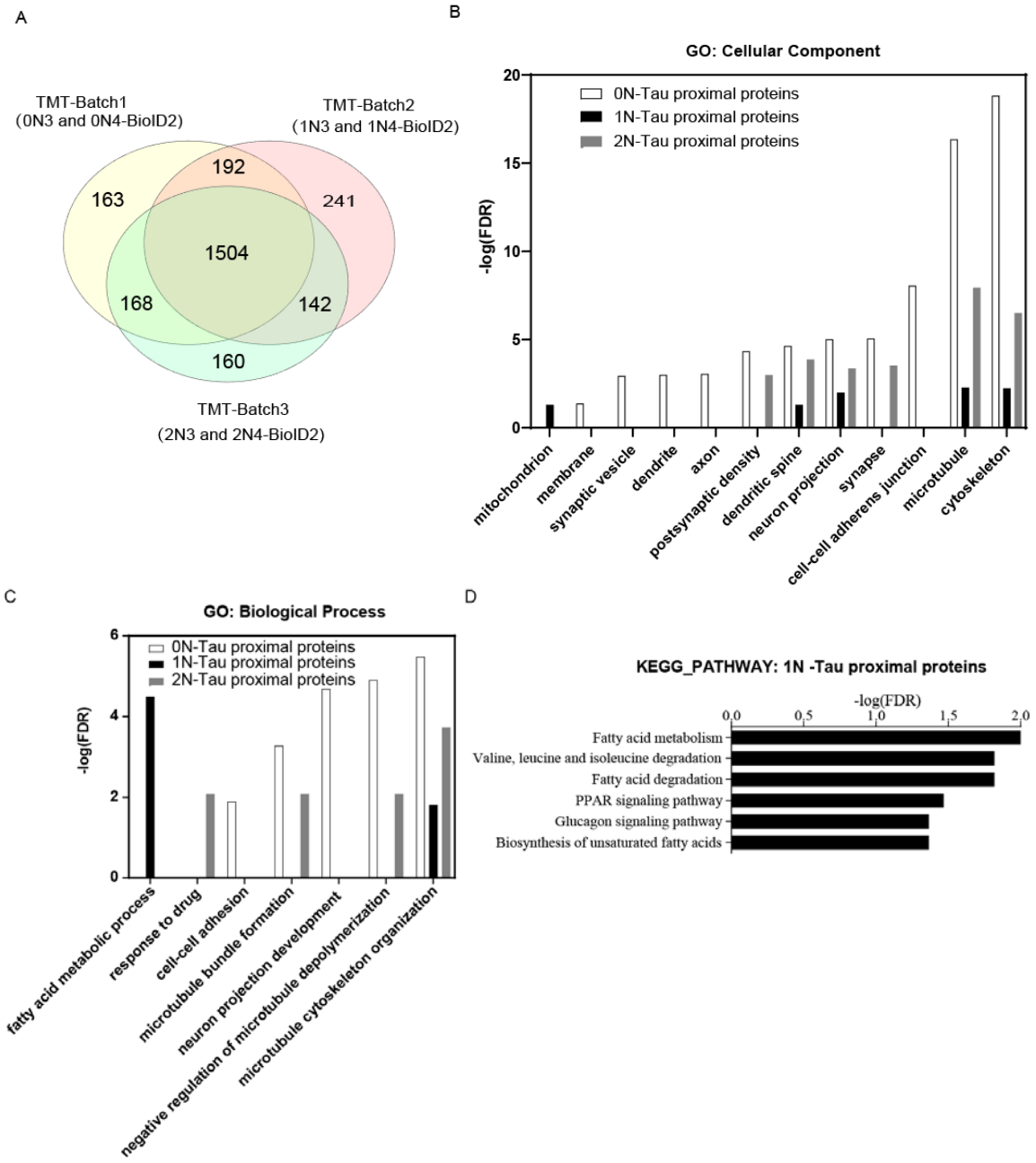


Figure 12: Functional distribution of proteins biotinylated by 0N-, 1N- and 2N-Tau BioID2. A) Venn diagram showing 1504 overlapped proteins identified from three batches of MS runs. These overlapped proteins were used to re-generate the BioID2 protein hits for each Tau isoforms. The resulting 0N-, 1N, 2N-Tau proximal proteins were used for functional analysis in (B-D). B-C) Gene ontology of 0N-, 1N- and 2R-Tau proximal proteins. D) KEGG pathway enrichment analysis of 1N-Tau proximal proteins. Enriched items with FDR< 0.05 were listed.

3.7 Supplementary figures and tables

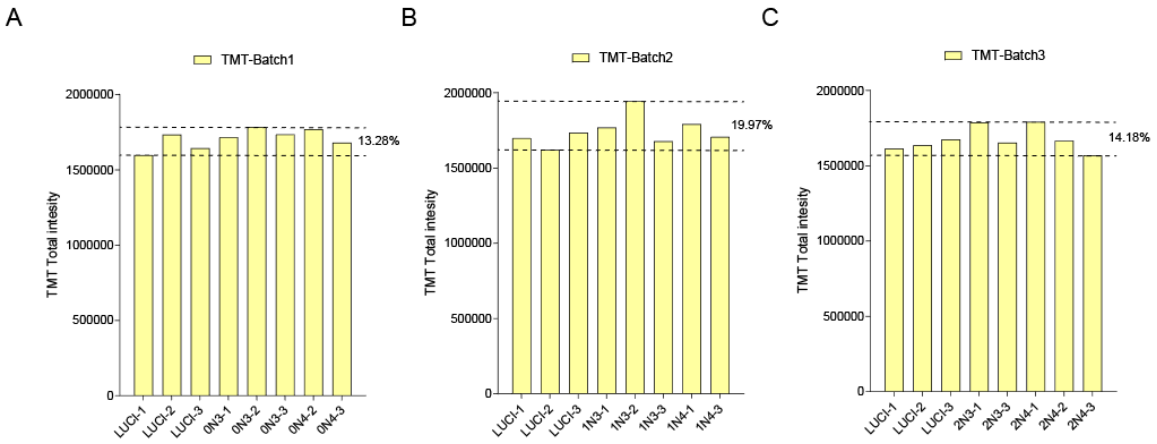


Figure S1: TMT total intensity of each channels in three batches of MS experiments. A-C) TMT total intensity in each TMT labeled sample. The variation of TMT intensity among each channel within one batch of TMT experiment was indicated.

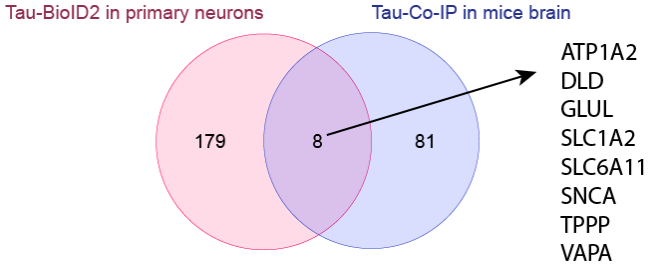


Figure S2: Comparison of the Tau proximal proteins identified in this Tau BioID2 system with one previously reported Tau interactome that identified by AP-MS in mouse brain [151]. Overlapped proteins from these two datasets are listed.

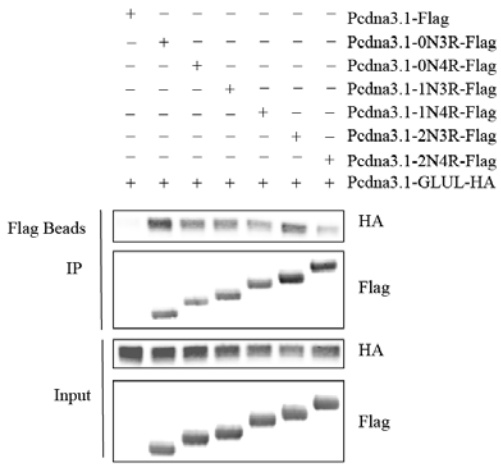


Figure S3: Validation of interaction between GLUL and Tau isoforms by co-immunoprecipitation. Six Tau isoforms (flag tagged) as well as the blank vector (negative control) were individually co-transfected with HA tagged GLUL in HEK 293FT cells, after 48 hrs., cells were harvest for immunoprecipitation with Flag beads.

Table S1: Tau BioID2 samples and assigned TMT channels

TMT-10plex-set1			**TMT-10plex-set2**			**TMT-10plex-set3**		
No.	sample name	TMT channel	No.	sample name	TMT channel	No.	sample name	TMT channel
1	LUCI-P2A-BioID2	126	1	LUCI-P2A-BioID2	126	1	LUCI-P2A-BioID2	126
2	0N3-BioID2	127N	2	1N3-BioID2	127N	2	2N3-BioID2	127N
3	0N4-BioID2	127C	3	1N4-BioID2	127C	3	2N4-BioID2	127C
4	LUCI-P2A-BioID2	128N	4	LUCI-P2A-BioID2	128N	4	LUCI-P2A-BioID2	128N
5	0N3-BioID2	128C	5	1N3-BioID2	128C	5	2N3-BioID2	128C
6	0N4-BioID2	129N	6	1N4-BioID2	129N	6	2N4-BioID2	129N
7	LUCI-P2A-BioID2	129C	7	LUCI-P2A-BioID2	129C	7	LUCI-P2A-BioID2	129C
8	0N3-BioID2	130N	8	1N3-BioID2	130N	8	2N3-BioID2	130N
9	0N4-BioID2	130C	9	1N4-BioID2	130C	9	2N4-BioID2	130C

Table S2: Significant protein hits identified from Tau isoform-BioID2 experiments.

0N3R-Tau proximal proteins

Accession	Gene	Ratio	SigB	p_value
P20357	Map2	1.513088	4.91E-63	0.005652
Q3UHX2	Pdap1	1.439424	7.12E-11	0.012696
P27546	Map4	1.337963	9.56E-29	0.001375
Q5U430	Ubr3	1.309831	1.63E-24	0.003191
Q8CI61	Bag4	1.297556	8.27E-06	0.016097
Q9CQU5	Zwint	1.287601	1.61E-05	0.04464
Q9CQH7	Btf3l4	1.269812	4.19E-09	0.003782
Q1HFZ0	Nsun2	1.265181	6.67E-05	0.039271
Q61166	Mapre1	1.249758	4.54E-08	0.011788
P15105	Glul	1.242061	1.09E-07	0.0452
Q9D8B4	Ndufa11	1.222523	0.000756	0.015744
O08648	Map3k4	1.214473	0.001148	0.01807
Q99PU8	Dhx30	1.206319	0.00173	0.00378
P14873	Map1b	1.199951	2.32E-11	0.00269
Q80TL0	Ppm1e	1.192341	0.003388	0.01523
O55042	Snca	1.189309	2.03E-08	0.040001
Q9JMC3	Dnaja4	1.178805	0.006265	0.005815
Q80TK0	Kiaa1107	1.168694	4.94E-07	0.011922
Q8BGD9	Eif4b	1.149454	0.021032	0.033347
A2AG50	Map7d2	1.149398	7.3E-06	0.000676
P08752	Gnai2	1.141797	0.028077	0.012231
Q0KL02	Trio	1.138222	3.05E-05	0.001905
Q9CU62	Smc1a	1.134253	0.036919	0.01028
Q8BVL9	Jakmip1	1.128658	0.002675	0.00405
Q7TSJ2	Map6	1.116501	6.35E-05	5.08E-05
Q9CS42	Prps2	1.115745	0.006145	0.041003
Q7TQD2	Tppp	1.115584	0.006206	0.026
A2AHC3	Camsap1	1.111503	0.000632	0.045379
Q01853	Vcp	1.104313	0.001305	0.045479
Q8CIB5	Fermt2	1.099152	0.0163	0.031434
A2AGT5	Ckap5	1.093105	0.001168	0.000178
Q3THS6	Mat2a	1.091621	0.004272	0.027534
O55131	Septin7	1.090036	0.004913	0.006807
Q8C1B1	Camsap2	1.078909	0.005273	0.02991
Q9D0R8	Lsm12	1.076943	0.014531	0.031959
Q80TV8	Clasp1	1.073675	0.018682	0.014987
P63017	Hspa8	1.05856	0.032743	0.020691

Q99M87	Dnaja3	1.056715	0.037906	0.001227

0N4R-Tau proximal proteins

Accession	**Gene**	**Ratio**	**SigB**	**p_value**
P20357	Map2	2.275649	8.6E-156	0.001206
Q3UHX2	Pdap1	2.042697	1.18E-21	0.013632
Q6PGC1	Dhx29	2.029413	3.83E-21	0.011272
Q9CQU5	Zwint	1.736085	1.78E-11	0.007671
Q6TCG2	Paqr9	1.713416	4.17E-23	0.010763
O55042	Snca	1.701472	2.13E-22	0.003043
P27546	Map4	1.6686	9.66E-46	0.001846
Q9CQH7	Btf3l4	1.66072	3.03E-18	0.00087
Q6R891	Ppp1r9b	1.629864	1.03E-16	0.006647
P14873	Map1b	1.627269	1.15E-40	0.000917
O88735	Map7	1.60257	4.06E-08	0.009504
P97390	Vps45	1.596773	1.12E-16	0.005521
Q5U430	Ubr3	1.5738	1.47E-34	0.000498
Q9D8B4	Ndufa11	1.524493	1.88E-06	0.001254
A2AHC3	Camsap1	1.519321	7.31E-29	0.001204
Q9QYR6	Map1a	1.489449	5.75E-26	0.007945
Q3UHD3	Mtus2	1.474558	1.68E-05	0.000388
P18052	Ptpra	1.468259	7.28E-10	0.020511
Q80Y56	Rbsn	1.464315	1.01E-10	0.002944
Q80TK0	Kiaa1107	1.449561	3.82E-10	0.000657
Q61166	Mapre1	1.437624	8.72E-09	0.007555
A2AG50	Map7d2	1.43238	1.7E-09	0.002753
Q8BGD9	Eif4b	1.426467	0.000114	0.005596
P12367	Prkar2a	1.415404	0.000172	0.020348
Q8BSZ2	Ap3s2	1.405795	9.7E-08	0.000999
Q8C1B7	Septin11	1.395847	0.00035	0.009129
O70311	Nmt2	1.377407	0.000662	0.008905
Q8BVL9	Jakmip1	1.370895	1.12E-06	0.000782
Q80XI3	Eif4g3	1.367033	9.08E-16	0.006646
Q61315	Apc	1.360197	2.27E-06	0.011896
Q7TSJ2	Map6	1.354889	6.66E-15	0.001633
Q8CI61	Bag4	1.337296	0.002416	0.023384
Q3UMY5	Eml4	1.332106	0.00283	0.000197
Q99K28	Arfgap2	1.32274	0.003744	0.045418
A2AGT5	Ckap5	1.311707	4.88E-12	0.000169
Q9JMC3	Dnaja4	1.308772	0.00561	0.008219
P55821	Stmn2	1.29904	2.85E-05	0.002555

Q0P5W1	Vps8	1.274417	0.000329	0.011419
Q6PGF2	Morn4	1.270509	0.000401	0.010747
P51141	Dvl1	1.258032	6.13E-09	0.01625
Q9WV69	Dmtn	1.250381	0.001065	0.01749
Q3UU96	Cdc42bpa	1.248014	0.000498	0.014448
Q9QZQ1	Afdn	1.247844	0.000502	0.009728
P10649	Gstm1	1.247536	0.001216	0.014527
P97379	G3bp2	1.246919	0.001251	0.011018
Q9ESN9	Mapk8ip3	1.238772	0.000795	0.010359
Q6A065	Cep170	1.233633	1.07E-07	0.001547
P63073	Eif4e	1.233053	0.00234	0.002635
Q99MK8	Grk2	1.229779	0.041461	0.008164
Q9QXZ0	Macf1	1.227955	0.043175	0.000832
Q7TQF7	Amph	1.227046	0.04405	0.02462
Q61062	Dvl3	1.225694	2.57E-07	0.017497
Q3URD3	Slmap	1.225606	0.045469	0.036144
Q8CE50	Snx30	1.224461	0.046625	0.000401
Q01853	Vcp	1.216388	0.002321	0.017492
O70251	Eef1b2	1.215227	7.85E-07	0.020089
Q61595	Ktn1	1.213666	0.00533	0.022144
Q8VDJ3	Hdlbp	1.212	0.002834	0.013633
P61264	Stx1b	1.211213	0.002936	0.042591
P47757	Capzb	1.197417	0.005368	0.007054
Q0KK55	Kndc1	1.188628	0.007747	0.004721
Q8VCD6	Reep2	1.186983	0.014982	0.010153
Q6PIE5	Atp1a2	1.18587	0.015603	0.020503
O55131	Septin7	1.183619	0.009492	0.005271
Q5HZI2	C2cd4c	1.180204	0.010874	0.005513
P15105	Glul	1.179821	0.019392	0.003066
Q80TV8	Clasp1	1.178568	2.78E-05	3.56E-06
P60229	Eif3e	1.175408	0.013117	0.012664
Q60598	Cttn	1.172578	4.74E-05	0.019655
Q8C1B1	Camsap2	1.171781	5.08E-05	0.002415
Q9Z2H5	Epb41l1	1.170061	5.9E-05	0.025165
Q6P9K8	Caskin1	1.168171	0.028997	0.010994
Q5DTY9	Kctd16	1.165003	0.019435	0.017739
A2ALS5	Rap1gap	1.163927	0.033395	0.029863
O08539	Bin1	1.160108	0.023233	0.038162
Q61792	Lasp1	1.157318	0.041378	0.038047
Q99K85	Psat1	1.154581	0.045127	0.026687
Q80XK6	Atg2b	1.153048	0.029839	0.007448
Q99PL5	Rrbp1	1.145488	0.038643	0.039525

Q9ES28	Arhgef7	1.145218	0.000449	0.043937
P13595	Ncam1	1.138501	0.000745	0.024589
Q7TT50	Cdc42bpb	1.126116	0.001811	0.000133
Q3UNH4	Gprin1	1.125943	0.001833	0.047633
P20029	Hspa5	1.123995	0.002096	0.006041
Q9D8N0	Eef1g	1.122632	0.002301	0.006558
Q9Z1S5	Septin3	1.116067	0.003564	0.00784
Q9Z2Q6	Septin5	1.106016	0.006745	0.017875
P63017	Hspa8	1.103729	0.007757	0.01616
Q3UHJ0	Aak1	1.10041	0.009467	0.014851
Q9WV55	Vapa	1.094733	0.013184	0.042024
Q8BRT1	Clasp2	1.092114	0.015296	0.012809
P42932	Cct8	1.073085	0.041674	0.028849

1N3R-Tau proximal proteins

Accession	**Gene**	**Ratio**	**SigB**	**p_value**
P15105	Glul	1.711599	8.23E-20	0.004286
O88196	Ttc3	1.521099	1.66E-11	0.005867
Q69Z99	Zfp512	1.344654	5.86E-06	0.033455
Q5U430	Ubr3	1.322195	1.38E-16	0.001078
Q80UP8	Slc20a2	1.291945	0.000107	0.000338
Q3THS6	Mat2a	1.281092	1.31E-07	0.000576
P27546	Map4	1.277711	7.66E-13	0.000661
O54984	Asna1	1.241356	0.001186	0.03062
Q6PIE5	Atp1a2	1.219148	3.28E-05	0.005022
Q8BGT5	Gpt2	1.212813	5.96E-05	0.012043
Q8VHQ9	Acot11	1.192382	0.000266	0.009797
Q91VR7	Map1lc3a	1.19103	0.008944	0.007687
Q8C6B2	Rtkn	1.19079	0.009023	0.01134
P20357	Map2	1.190492	5.42E-07	0.012171
Q6P9R2	Oxsr1	1.179802	0.013358	0.003834
Q8C0M9	Asrgl1	1.174775	0.000814	0.00497
Q6ZQK5	Acap2	1.165675	0.021574	0.031941
Q7TSH2	Phkb	1.163374	0.023263	0.014044
Q9DAK9	Phpt1	1.1587	0.002375	0.022239
P43006	Slc1a2	1.14665	8.60E-05	0.023823
Q62351	Tfrc	1.145832	0.040347	0.038424
Q9Z0P4	Palm	1.143913	0.042741	0.01842
Q60714	Slc27a1	1.141751	0.00598	0.033357
Q8BR90	AW549877	1.139082	0.049304	0.045191
Q9Z0R9	Fads2	1.136577	0.000235	0.027132

Q920L1	Fads1	1.129457	0.000462	0.036713
Q8CCJ3	Ufl1	1.127597	0.013209	0.02727
P17918	Pcna	1.12245	0.017065	0.044084
Q99JY0	Hadhb	1.122276	0.000885	0.005967
P13011	Scd2	1.121422	0.017945	0.018697
Q8BMS1	Hadha	1.120744	0.001014	0.005066
P56564	Slc1a3	1.12057	0.018056	0.003619
Q91WC3	Acsl6	1.11882	0.019664	0.007688
Q91V41	Rab14	1.115997	0.02252	0.027663
P19258	Mpv17	1.115848	0.023435	0.042207
Q99L85	Elp5	1.11298	0.025965	0.035062
P21279	Gnaq	1.112125	0.027867	0.021043
Q8CIB5	Fermt2	1.106646	0.035696	0.029436
O08749	Dld	1.105457	0.036598	0.006083
Q923T9	Camk2g	1.08703	0.014155	0.009475
Q9CX86	Hnrnpa0	1.083971	0.017421	0.011596

1N4R-Tau proximal proteins

Accession	**Gene**	**Ratio**	**SigB**	**p_value**
P15105	Glul	1.615114	5.75E-10	0.013206
P20357	Map2	1.552665	3.40E-56	0.002864
P27546	Map4	1.518259	9.69E-50	3.72E-05
Q5U430	Ubr3	1.455566	5.12E-39	0.00477
O54984	Asna1	1.442954	1.01E-05	0.001207
O88735	Map7	1.440266	1.15E-05	0.003488
Q80TK0	Kiaa1107	1.348411	1.73E-14	0.007185
Q69Z99	Zfp512	1.326839	0.001309	0.020931
Q3THS6	Mat2a	1.275875	1.29E-09	0.005473
Q3UK78	Pcgf5	1.270319	0.008564	0.008357
Q8VCZ6	Sgsm3	1.270219	0.00859	0.004601
Q61792	Lasp1	1.24616	0.017361	0.037993
E9PZJ8	Ascc3	1.24444	5.74E-05	0.00296
A2AG50	Map7d2	1.238416	8.84E-05	0.008317
Q8BGT5	Gpt2	1.218139	0.00035	0.033139
Q80UP8	Slc20a2	1.215063	0.039671	0.005535
Q810S1	Ccdc109b	1.212716	0.042066	0.002694
Q8VHQ9	Acot11	1.209628	0.000603	0.028033
Q60714	Slc27a1	1.20695	5.45E-06	0.024867
Q6PIE5	Atp1a2	1.199077	1.23E-05	0.012647
A2AGT5	Ckap5	1.164171	1.06E-06	0.000108
O08749	Dld	1.162231	0.000372	0.008924

Q01853	Vcp	1.161894	0.008874	0.009591
P70188	Kifap3	1.147699	0.017546	0.038624
P14873	Map1b	1.139594	2.80E-05	0.031857
Q3UHC0	Tnrc6c	1.13556	0.002972	0.028878
P97742	Cpt1a	1.126921	0.005437	0.046795
Q8C0M9	Asrgl1	1.118072	0.00975	0.040211
Q7TSJ2	Map6	1.114291	0.000505	0.029985
Q61166	Mapre1	1.111411	0.014788	0.003584
Q0KK55	Kndc1	1.104988	0.021692	0.044658
Q8VDD5	Myh9	1.100582	0.027916	0.035611
A2AJI0	Map7d1	1.095057	0.003304	0.003265
P53395	Dbt	1.090986	0.046965	0.010491
Q99JY0	Hadhb	1.07771	0.014262	0.01111
Q6A065	Cep170	1.076379	0.015814	0.007005

2N3R-Tau proximal proteins

Accession	**Gene**	**Ratio**	**SigB**	**p_value**
P15105	Glul	1.544053	2.47E-23	0.023652
P27546	Map4	1.511513	1.17E-33	0.000909
Q5U430	Ubr3	1.508647	2.61E-33	0.009645
P20357	Map2	1.477222	1.35E-29	0.011377
Q8BHI7	Elovl5	1.315167	3.96E-09	0.000634
Q3THS6	Mat2a	1.311964	5.57E-09	0.00295
Q80UP8	Slc20a2	1.305673	1.08E-08	0.016313
Q9CQH7	Btf3l4	1.278924	1.35E-09	0.018976
Q80SY5	Prpf38b	1.266448	0.000724	0.045004
A2AGT5	Ckap5	1.235935	1.12E-08	0.005882
Q6PIE5	Atp1a2	1.209892	3.82E-06	0.001219
Q3UDK1	Trafd1	1.205393	0.008782	0.039777
Q8BUL6	Plekha1	1.200911	0.010328	0.00751
P11531	Dmd	1.194158	0.013116	0.039269
P60766	Cdc42	1.188109	0.000285	0.005926
Q8BYK6	Ythdf3	1.186171	0.017258	0.023019
Q3UMY5	Eml4	1.17383	0.025909	0.009129
P43006	Slc1a2	1.173586	1.98E-05	0.025492
P63321	Rala	1.165533	0.033645	0.039236
P01831	Thy1	1.162007	0.037488	0.028394
P56564	Slc1a3	1.153435	0.000576	0.013896
O08749	Dld	1.144547	0.001125	0.03017
P35802	Gpm6a	1.144458	0.004323	0.019207
Q923T9	Camk2g	1.132401	0.002666	0.004368

Q8R0G7	Spns1	1.129177	0.009818	0.037773
Q8BVL9	Jakmip1	1.121061	0.014769	0.007525
Q80TK0	Kiaa1107	1.117939	0.006891	0.030613
Q01853	Vcp	1.117533	0.007069	0.038597
O08529	Capn2	1.106577	0.006609	0.021521
Q80Y56	Zfyve20	1.096925	0.023607	0.03174
P68433	Hist1h3g	1.095091	0.026068	0.004486
P55096	Abcd3	1.092123	0.017255	0.028411
Q7TSJ2	Map6	1.074902	0.047423	0.035402

2N4R-Tau proximal proteins

Accession	Gene	Ratio	SigB	p_value
P27546	Map4	1.747453	3.2E-77	6.02E-05
P20357	Map2	1.730832	5.49E-74	0.006892
Q6PGC1	Dhx29	1.57961	1.59E-12	0.034255
Q5U430	Ubr3	1.566382	8.74E-46	0.004565
O88735	Map7	1.522009	2.08E-10	0.018321
P15105	Glul	1.518252	5.82E-17	0.008429
P11531	Dmd	1.464445	1.66E-08	0.002053
Q9CQH7	Btf3l4	1.426197	1.89E-20	0.003456
Q3UHX2	Pdap1	1.354898	1.77E-05	0.044748
Q9CZ44	Nsfl1c	1.329336	7.01E-05	0.014731
A2AG50	Map7d2	1.325864	1.14E-12	0.001188
Q9CQ80	Vps25	1.316273	0.000136	0.043373
Q80UP8	Slc20a2	1.310259	5.32E-07	0.013059
A2AGT5	Ckap5	1.307338	2.51E-15	4.02E-05
Q8VHQ9	Acot11	1.293476	0.000411	0.026856
P31650	Slc6a11	1.261369	2.38E-05	0.008217
P14873	Map1b	1.258797	1.65E-11	0.00574
Q3THS6	Mat2a	1.246544	6.7E-05	2.21E-05
Q8BHI7	Elovl5	1.239607	0.000107	0.000269
P97300	Nptn	1.237914	0.000119	0.005864
Q9QYR6	Map1a	1.235864	6.52E-10	0.032755
P97390	Vps45	1.235137	7.29E-10	0.006486
Q8R0G7	Spns1	1.226992	0.000241	0.003897
P63046	Sult4a1	1.221891	0.008067	0.026466
Q8K1R7	Nek9	1.221762	0.008105	0.000539
Q61166	Mapre1	1.218053	0.009271	0.046362
Q8BYM5	Nlgn3	1.21427	0.010614	0.041486
Q64337	Sqstm1	1.206927	0.01372	0.001618
P36916	Gnl1	1.200326	0.017173	0.025357

Q3UXZ6	Fam81a	1.19931	0.017766	0.038713
Q3UMY5	Eml4	1.195978	0.019843	0.002522
Q9CXR1	Dhrs7	1.192262	0.022406	0.028009
O35127	Grcc10	1.183434	0.029674	0.042971
Q80WJ7	Mtdh	1.17501	0.038415	0.038534
P17918	Pcna	1.174781	0.004666	0.02643
A2AHC3	Camsap1	1.174462	2.78E-06	0.019717
Q3UDK1	Trafd1	1.169819	0.044823	0.028916
Q01853	Vcp	1.168369	0.000192	0.010281
P42208	Septin2	1.166535	0.049327	0.009683
Q6PIE5	Atp1a2	1.164077	0.000275	0.00443
Q80Y56	Zfyve20	1.160587	0.000367	0.032518
P43006	Slc1a2	1.15806	1.8E-05	0.030958
Q78ZA7	Nap1l4	1.157286	0.010864	0.030422
P60879	Snap25	1.154897	0.000581	0.018329
Q2PFD7	Psd3	1.151422	0.014188	0.004812
Q80YT7	Pde4dip	1.144332	0.019381	0.002698
P60766	Cdc42	1.143729	0.019892	0.009469
Q9D071	Mms19	1.142928	0.020587	0.02344
Q8C1B7	Septin11	1.13358	0.002861	0.040355
Q80TV8	Clasp1	1.13324	0.00293	0.001728
P56564	Slc1a3	1.13309	0.002961	0.027126
Q6A065	Cep170	1.132144	0.000253	0.005624
E9Q4Z2	Acacb	1.127448	0.004365	0.03506
P54227	Stmn1	1.126049	0.000447	0.008672
Q8BMF4	Dlat	1.125136	0.042509	0.022863
Q7TSJ2	Map6	1.120734	0.000721	0.004236
Q8C1B1	Camsap2	1.105684	0.00256	0.002736
Q0KK55	Kndc1	1.1035	0.019487	0.016613
Q923T9	Camk2g	1.100066	0.023678	0.005852
P13595	Ncam1	1.098996	0.00432	0.031962
Q8BMS1	Hadha	1.096683	0.005148	0.015313
Q9Z0J4	Nos1	1.090945	0.038794	0.046018
Q8VDJ3	Hdlbp	1.085935	0.011193	0.015932
Q8C0M9	Asrgl1	1.074862	0.023346	0.045522
P20029	Hspa5	1.071575	0.028676	0.038705
Q9DBL1	Acadsb	1.068413	0.034762	0.002286

Chapter 4 Concluding remarks

Tauopathy is a group of neurodegenerative diseases that are characterized by the accumulation of Tau aggregates in affected brain regions. Mounting evidence indicates that the formation of Tau filament tangles is strongly linked to the severity of cognitive deficits in dementia, which highlights the central role that pathological Tau plays in the disease progression of AD and other neurodegenerative disorders. It is believed that the aberrant alternative splicing of Tau is one of the pathogenic factors in the pathogenesis of a subset of Tauopathies. However, there has been limited success to systematically unravel the upstream splicing regulators that contribute to abnormal Tau splicing regulation. Meanwhile, the regulatory mechanisms regarding how excess expression of 3R-Tau or 4R-Tau leads to pathological aggregates remain unclear as well. To address this scientific question, it is crucial to understand the interactome network for each Tau isoform, especially 3R- and 4R-Tau isoforms, and to establish functional basis to dissect the mechanisms underpinning the development of Tau pathology in tauopathies.

The lack of techniques to identify the locus-specific splicing factors has become a major problem in splicing regulation field over the last decades. Previous studies to identify Tau splicing regulators were limited to *in vitro* RNA pull down, which does not accurately recapitulate the protein-RNA interaction under physiological conditions. Moreover, the high background noises also make the data interpretation difficult. As described in Chapter 2, by applying a robust and sensitive oligo-based proteomic approach, RAP-MS, Tau pre-mRNA (transcribed from Minigene) along with its covalently coupled RBPs were captured and analyzed by mass spectrometry. Following validations, nine novel Tau exon 10 splicing

regulators, including hnRNPA2B1, hnRNPC, hnRNPU, hnRNPH1, hnRNPD, THRAP3, DDX21, LUC7L3 and EMG1 were identified. Specifically, hnRNPC was validated to bind to and utilize the U-Tracts motifs located in the 3' ends of intron 9 and intron 10 to promote exon 10 inclusion. It has been found that Tau mutations are frequently identified in the stem loop structure located in the 5' splice site of intron 10. Hence, our results are complementary to previous findings. Modulating the expression of hnRNPC may be able to restore the distorted Tau splicing in familial tauopathies that are bearing genetic mutations in this stem loop region. Moreover, a large publicly available RNA-seq dataset has shown that hnRNPC is upregulated in PSP, a sporadic tauopathy that is characterized by the accumulation of 4R-Tau in neuron and glial cells. Hence, it is possible that hnRNPC can be developed as a therapeutic target for PSP.

Separately, through applying biotin ligase-based PL (BioID2) coupled with TMT-quantitative proteomics, the proximity interactome networks for each of the six human CNS Tau isoforms were uncovered in mouse primary neurons. The identified interactome networks suggested that functional differences beyond microtubule association existed among Tau isoforms. As described in Chapter 3, the proteins in proximity with 3R-Tau were enriched in the fatty acid metabolic pathway whereas 4R-Tau proximal proteins were identified to have enrichment in endocytosis, which is related with synaptic transmission. When it comes to the difference between Tau isoforms differing in the N-terminus, 1N-Tau was identified to be associated with many pathways, such as fatty acid metabolism, PPAR signaling pathway and glucagon signaling pathway. In contrast, no pathways were significantly enriched for 0N- and 2N-Tau proximal proteins. The functional differences

among Tau isoforms indicate that the maintenance of intrinsic expression for each Tau isoform may be important to achieve functional integration to support the complex biological processes in neurons. Once this balance is disrupted due to the elevation or reduction of certain isoforms, neuronal activity regulation might be interfered, which may further lead to Tau aggregation and the subsequent neuronal loss. However, additional functional assays in cellular and animal models are needed to validate this hypothesis.

Specifically, MAT2A was validated as a Tau interactor that showed preferential binding with the 3R-Tau compared to the 4R-Tau. Further mechanistic study, such as the mapping interaction sites and loss of function assays, are needed to decipher the functional regulatory mechanisms that 3R-Tau may play in modulating the activity of MAP2A for global epigenetic regulation and methylation directed protein functional modulation.

Altogether, my thesis elaborated on the importance of applying RAP-proteomics to systematically uncover splicing factors that contribute to Tau exon 10 splicing, which may provide potential therapeutic targets for Tauopathy treatment. Moreover, the successful application of RAP-MS in Tau splicing regulation indicates that this method can be utilized for other splicing regulation studies, specifically to unveil the dysregulated splicing factors/RBPs that are involved in the production of disease-specific splicing isoforms for developing therapeutics. Separately, the discovery of Tau isoform-specific interactome networks implies that the functional differences exist among Tau isoforms. In the future, substantial efforts are still required to dissect how the interactome network differences of among the Tau isoforms contribute to the isoform-specific aggregation in a subset of tauopathy patients.

REFERENCES

1. Weingarten, M.D., Lockwood, A.H., Hwo, S.Y. & Kirschner, M.W. A protein factor essential for microtubule assembly. *Proc Natl Acad Sci U S A* **72**, 1858-1862 (1975).
2. Tashiro, K., Hasegawa, M., Ihara, Y. & Iwatsubo, T. Somatodendritic localization of phosphorylated tau in neonatal and adult rat cerebral cortex. *Neuroreport* **8**, 2797-2801 (1997).
3. Derisbourg, M. *et al.* Role of the Tau N-terminal region in microtubule stabilization revealed by new endogenous truncated forms. *Sci Rep* **5**, 9659 (2015).
4. He, H.J. *et al.* The proline-rich domain of tau plays a role in interactions with actin. *BMC cell biology* **10**, 81 (2009).
5. Sottejeau, Y. *et al.* Tau phosphorylation regulates the interaction between BIN1's SH3 domain and Tau's proline-rich domain. *Acta Neuropathol Commun* **3**, 58 (2015).
6. Croft, C.L. *et al.* Novel monoclonal antibodies targeting the microtubule-binding domain of human tau. *PLoS One* **13**, e0195211 (2018).
7. Wood, J.G., Mirra, S.S., Pollock, N.J. & Binder, L.I. Neurofibrillary tangles of Alzheimer disease share antigenic determinants with the axonal microtubule-associated protein tau (tau). *Proc Natl Acad Sci U S A* **83**, 4040-4043 (1986).
8. Wischik, C.M. *et al.* Isolation of a fragment of tau derived from the core of the paired helical filament of Alzheimer disease. *Proc Natl Acad Sci U S A* **85**, 4506-4510 (1988).
9. Nukina, N. & Ihara, Y. One of the antigenic determinants of paired helical filaments is related to tau protein. *J Biochem* **99**, 1541-1544 (1986).
10. Grundke-Iqbal, I. *et al.* Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *J Biol Chem* **261**, 6084-6089 (1986).
11. Spillantini, M.G. & Goedert, M. Tau pathology and neurodegeneration. *Lancet Neurol* **12**, 609-622 (2013).
12. Wang, Y. & Mandelkow, E. Tau in physiology and pathology. *Nat Rev Neurosci* **17**, 5-21 (2016).
13. Caillet-Boudin, M.L. *et al.* Brain pathology in myotonic dystrophy: when tauopathy meets spliceopathy and RNAopathy. *Front Mol Neurosci* **6**, 57 (2014).
14. Crary, J.F. *et al.* Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol* **128**, 755-766 (2014).
15. Sergeant, N., Delacourte, A. & Buee, L. Tau protein as a differential biomarker of tauopathies. *Biochim Biophys Acta* **1739**, 179-197 (2005).
16. Avila, J., Santa-Maria, I., Perez, M., Hernandez, F. & Moreno, F. Tau phosphorylation, aggregation, and cell toxicity. *J Biomed Biotechnol* **2006**, 74539 (2006).
17. Buee, L. *et al.* From tau phosphorylation to tau aggregation: what about neuronal death? *Biochem Soc Trans* **38**, 967-972 (2010).
18. Hong, Y. *et al.* SRPK2 phosphorylates tau and mediates the cognitive defects in Alzheimer's disease. *J Neurosci* **32**, 17262-17272 (2012).
19. Stoothoff, W.H. & Johnson, G.V. Tau phosphorylation: physiological and pathological consequences. *Biochim Biophys Acta* **1739**, 280-297 (2005).
20. Morris, M. *et al.* Tau post-translational modifications in wild-type and human amyloid precursor protein transgenic mice. *Nat Neurosci* **18**, 1183-1189 (2015).

21. Iqbal, K., Gong, C.X. & Liu, F. Hyperphosphorylation-induced tau oligomers. *Front Neurol* **4**, 112 (2013).
22. Haj-Yahya, M. *et al.* Site-Specific Hyperphosphorylation Inhibits, Rather than Promotes, Tau Fibrillization, Seeding Capacity, and Its Microtubule Binding. *Angewandte Chemie (International ed. in English)* **59**, 4059-4067 (2020).
23. Köpke, E. *et al.* Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease. *J Biol Chem* **268**, 24374-24384 (1993).
24. Hanger, D.P., Anderton, B.H. & Noble, W. Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. *Trends in molecular medicine* **15**, 112-119 (2009).
25. Liu, F., Iqbal, K., Grundke-Iqbal, I. & Gong, C.X. Involvement of aberrant glycosylation in phosphorylation of tau by cdk5 and GSK-3beta. *FEBS Lett* **530**, 209-214 (2002).
26. Wang, J.Z., Wu, Q., Smith, A., Grundke-Iqbal, I. & Iqbal, K. Tau is phosphorylated by GSK-3 at several sites found in Alzheimer disease and its biological activity markedly inhibited only after it is prephosphorylated by A-kinase. *FEBS Lett* **436**, 28-34 (1998).
27. Liu, F., Grundke-Iqbal, I., Iqbal, K. & Gong, C.X. Contributions of protein phosphatases PP1, PP2A, PP2B and PP5 to the regulation of tau phosphorylation. *The European journal of neuroscience* **22**, 1942-1950 (2005).
28. Gong, C.X. *et al.* Phosphatase activity toward abnormally phosphorylated tau: decrease in Alzheimer disease brain. *J Neurochem* **65**, 732-738 (1995).
29. Yu, Y. *et al.* Developmental regulation of tau phosphorylation, tau kinases, and tau phosphatases. *J Neurochem* **108**, 1480-1494 (2009).
30. Su, B. *et al.* Physiological regulation of tau phosphorylation during hibernation. *J Neurochem* **105**, 2098-2108 (2008).
31. Alonso, A.C., Zaidi, T., Grundke-Iqbal, I. & Iqbal, K. Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. *Proc Natl Acad Sci U S A* **91**, 5562-5566 (1994).
32. Salehi, A., Delcroix, J.D. & Mobley, W.C. Traffic at the intersection of neurotrophic factor signaling and neurodegeneration. *Trends in neurosciences* **26**, 73-80 (2003).
33. Boekhoorn, K. *et al.* Improved long-term potentiation and memory in young tau-P301L transgenic mice before onset of hyperphosphorylation and tauopathy. *J Neurosci* **26**, 3514-3523 (2006).
34. Schindowski, K. *et al.* Alzheimer's disease-like tau neuropathology leads to memory deficits and loss of functional synapses in a novel mutated tau transgenic mouse without any motor deficits. *The American journal of pathology* **169**, 599-616 (2006).
35. Ren, Q.G., Liao, X.M., Chen, X.Q., Liu, G.P. & Wang, J.Z. Effects of tau phosphorylation on proteasome activity. *FEBS Lett* **581**, 1521-1528 (2007).
36. Wang, J.Z., Xia, Y.Y., Grundke-Iqbal, I. & Iqbal, K. Abnormal hyperphosphorylation of tau: sites, regulation, and molecular mechanism of neurofibrillary degeneration. *Journal of Alzheimer's disease : JAD* **33 Suppl 1**, S123-139 (2013).
37. Komuro, Y., Xu, G., Bhaskar, K. & Lamb, B.T. Human tau expression reduces adult neurogenesis in a mouse model of tauopathy. *Neurobiology of aging* **36**, 2034-2042 (2015).
38. Caillet-Boudin, M.L., Buee, L., Sergeant, N. & Lefebvre, B. Regulation of human MAPT gene expression. *Mol Neurodegener* **10**, 28 (2015).

39. Pittman, A.M. *et al.* The structure of the tau haplotype in controls and in progressive supranuclear palsy. *Hum Mol Genet* **13**, 1267-1274 (2004).
40. Boettger, L.M., Handsaker, R.E., Zody, M.C. & McCarroll, S.A. Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat Genet* **44**, 881-885 (2012).
41. Caffrey, T.M. & Wade-Martins, R. Functional MAPT haplotypes: bridging the gap between genotype and neuropathology. *Neurobiol Dis* **27**, 1-10 (2007).
42. Scholz, T. & Mandelkow, E. Transport and diffusion of Tau protein in neurons. *Cellular and molecular life sciences : CMLS* **71**, 3139-3150 (2014).
43. Binder, L.I., Frankfurter, A. & Rebhun, L.I. The distribution of tau in the mammalian central nervous system. *J Cell Biol* **101**, 1371-1378 (1985).
44. Sotiropoulos, I. *et al.* Atypical, non-standard functions of the microtubule associated Tau protein. *Acta Neuropathol Commun* **5**, 91 (2017).
45. Brandt, R., Leger, 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).
46. Pooler, A.M. *et al.* Dynamic association of tau with neuronal membranes is regulated by phosphorylation. *Neurobiology of aging* **33**, 431 e427-438 (2012).
47. Gil, L. *et al.* Aging dependent effect of nuclear tau. *Brain research* **1677**, 129-137 (2017).
48. Bukar Maina, M., Al-Hilaly, Y.K. & Serpell, L.C. Nuclear Tau and Its Potential Role in Alzheimer's Disease. *Biomolecules* **6**, 9 (2016).
49. Baquero, J. *et al.* Nuclear Tau, p53 and Pin1 Regulate PARN-Mediated Deadenylation and Gene Expression. *Front Mol Neurosci* **12**, 242 (2019).
50. Terry, R.D. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann Neurol* **30**, 572-580 (1991).
51. Scheff, S.W., Price, D.A., Schmitt, F.A. & Mufson, E.J. Hippocampal synaptic loss in early Alzheimer's disease and mild cognitive impairment. *Neurobiology of aging* **27**, 1372-1384 (2006).
52. Hilton, K.J., Cunningham, C., Reynolds, R.A. & Perry, V.H. Early Hippocampal Synaptic Loss Precedes Neuronal Loss and Associates with Early Behavioural Deficits in Three Distinct Strains of Prion Disease. *PLoS One* **8**, e68062 (2013).
53. Tai, H.C. *et al.* The synaptic accumulation of hyperphosphorylated tau oligomers in Alzheimer disease is associated with dysfunction of the ubiquitin-proteasome system. *The American journal of pathology* **181**, 1426-1435 (2012).
54. McInnes, J. *et al.* Synaptogyrin-3 Mediates Presynaptic Dysfunction Induced by Tau. *Neuron* **97**, 823-835.e828 (2018).
55. Chen, Q. *et al.* Tau protein is involved in morphological plasticity in hippocampal neurons in response to BDNF. *Neurochemistry international* **60**, 233-242 (2012).
56. Hoover, B.R. *et al.* Tau mislocalization to dendritic spines mediates synaptic dysfunction independently of neurodegeneration. *Neuron* **68**, 1067-1081 (2010).
57. Kopeikina, K.J. *et al.* Synaptic alterations in the rTg4510 mouse model of tauopathy. *The Journal of comparative neurology* **521**, 1334-1353 (2013).
58. Maas, T., Eidenmuller, J. & Brandt, R. Interaction of tau with the neural membrane cortex is regulated by phosphorylation at sites that are modified in paired helical filaments. *J Biol Chem* **275**, 15733-15740 (2000).

59. Georgieva, E.R., Xiao, S., Borbat, P.P., Freed, J.H. & Eliezer, D. Tau binds to lipid membrane surfaces via short amphipathic helices located in its microtubule-binding repeats. *Biophysical journal* **107**, 1441-1452 (2014).
60. Maas, T., Eidenmüller, J. & Brandt, R. Interaction of tau with the neural membrane cortex is regulated by phosphorylation at sites that are modified in paired helical filaments. *J Biol Chem* **275**, 15733-15740 (2000).
61. Arrasate, M., Pérez, M. & Avila, J. Tau dephosphorylation at tau-1 site correlates with its association to cell membrane. *Neurochemical research* **25**, 43-50 (2000).
62. Gray, E.G., Paula-Barbosa, M. & Roher, A. Alzheimer's disease: paired helical filaments and cytomembranes. *Neuropathol Appl Neurobiol* **13**, 91-110 (1987).
63. Campos-Peña, V., Tapia-Ramírez, J., Sánchez-Torres, C. & Meraz-Rios, M.A. Pathological-like assembly of tau induced by a paired helical filament core expressed at the plasma membrane. *Journal of Alzheimer's disease : JAD* **18**, 919-933 (2009).
64. Wille, H., Drewes, G., Biernat, J., Mandelkow, E.M. & Mandelkow, E. Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau in vitro. *J Cell Biol* **118**, 573-584 (1992).
65. Elbaum-Garfinkle, S., Ramlall, T. & Rhoades, E. The role of the lipid bilayer in tau aggregation. *Biophysical journal* **98**, 2722-2730 (2010).
66. Jones, E.M. *et al.* Interaction of tau protein with model lipid membranes induces tau structural compaction and membrane disruption. *Biochemistry* **51**, 2539-2550 (2012).
67. Metuzals, J., Robitaille, Y., Houghton, S., Gauthier, S. & Leblanc, R. Paired helical filaments and the cytoplasmic-nuclear interface in Alzheimer's disease. *J Neurocytol* **17**, 827-833 (1988).
68. Brady, R.M., Zinkowski, R.P. & Binder, L.I. Presence of tau in isolated nuclei from human brain. *Neurobiology of aging* **16**, 479-486 (1995).
69. Greenwood, J.A. & Johnson, G.V. Localization and in situ phosphorylation state of nuclear tau. *Experimental cell research* **220**, 332-337 (1995).
70. Cross, D.C., Munoz, J.P., Hernandez, P. & Maccioni, R.B. Nuclear and cytoplasmic tau proteins from human nonneuronal cells share common structural and functional features with brain tau. *J Cell Biochem* **78**, 305-317 (2000).
71. Loomis, P.A., Howard, T.H., Castleberry, R.P. & Binder, L.I. Identification of nuclear tau isoforms in human neuroblastoma cells. *Proc Natl Acad Sci U S A* **87**, 8422-8426 (1990).
72. Papasozomenos, S.C. Nuclear tau immunoreactivity in presenile dementia with motor neuron disease: a case report. *Clin Neuropathol* **14**, 100-104 (1995).
73. Wang, Y., Loomis, P.A., Zinkowski, R.P. & Binder, L.I. A novel tau transcript in cultured human neuroblastoma cells expressing nuclear tau. *J Cell Biol* **121**, 257-267 (1993).
74. Sultan, A. *et al.* Nuclear tau, a key player in neuronal DNA protection. *J Biol Chem* **286**, 4566-4575 (2011).
75. Krylova, S.M. *et al.* Tau protein binds single-stranded DNA sequence specifically--the proof obtained in vitro with non-equilibrium capillary electrophoresis of equilibrium mixtures. *FEBS Lett* **579**, 1371-1375 (2005).
76. Qi, H. *et al.* Nuclear magnetic resonance spectroscopy characterization of interaction of Tau with DNA and its regulation by phosphorylation. *Biochemistry* **54**, 1525-1533 (2015).
77. Hua, Q. *et al.* Microtubule associated protein tau binds to double-stranded but not single-stranded DNA. *Cellular and molecular life sciences : CMLS* **60**, 413-421 (2003).

78. Multhaup, G., Huber, O., Buée, L. & Galas, M.C. Amyloid Precursor Protein (APP) Metabolites APP Intracellular Fragment (AICD), Aβ42, and Tau in Nuclear Roles. *J Biol Chem* **290**, 23515-23522 (2015).
79. Olive, P.L., Banath, J.P. & Durand, R.E. Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells measured using the "comet" assay. *Radiat Res* **122**, 86-94 (1990).
80. Rossi, G. *et al.* A new function of microtubule-associated protein tau: involvement in chromosome stability. *Cell cycle (Georgetown, Tex.)* **7**, 1788-1794 (2008).
81. Mansuroglu, Z. *et al.* Loss of Tau protein affects the structure, transcription and repair of neuronal pericentromeric heterochromatin. *Sci Rep* **6**, 33047 (2016).
82. Thurston, V.C., Zinkowski, R.P. & Binder, L.I. Tau as a nucleolar protein in human nonneural cells in vitro and in vivo. *Chromosoma* **105**, 20-30 (1996).
83. Maina, M.B. *et al.* The involvement of tau in nucleolar transcription and the stress response. *Acta Neuropathol Commun* **6**, 70 (2018).
84. Maina, M.B., Bailey, L.J., Doherty, A.J. & Serpell, L.C. The Involvement of Aβ42 and Tau in Nucleolar and Protein Synthesis Machinery Dysfunction. *Frontiers in cellular neuroscience* **12**, 220 (2018).
85. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561-563 (1970).
86. Crick, F. Split genes and RNA splicing. *Science* **204**, 264-271 (1979).
87. Newman, A. RNA splicing. *Current biology : CB* **8**, R903-905 (1998).
88. Hoskins, A.A., Gelles, J. & Moore, M.J. New insights into the spliceosome by single molecule fluorescence microscopy. *Current opinion in chemical biology* **15**, 864-870 (2011).
89. Chen, M. & Manley, J.L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**, 741-754 (2009).
90. Zhang, G.W., Song, H.D. & Chen, Z. [Molecular mechanism of mRNA alternative splicing]. *Yi chuan xue bao = Acta genetica Sinica* **31**, 102-107 (2004).
91. Stamm, S. *et al.* Function of alternative splicing. *Gene* **344**, 1-20 (2005).
92. Cáceres, J.F. & Kornblihtt, A.R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends in genetics : TIG* **18**, 186-193 (2002).
93. Yura, K. *et al.* Alternative splicing in human transcriptome: functional and structural influence on proteins. *Gene* **380**, 63-71 (2006).
94. Lee, Y. & Rio, D.C. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annual review of biochemistry* **84**, 291-323 (2015).
95. Shepard, P.J. & Hertel, K.J. The SR protein family. *Genome Biol* **10**, 242 (2009).
96. Martinez-Contreras, R. *et al.* hnRNP proteins and splicing control. *Adv Exp Med Biol* **623**, 123-147 (2007).
97. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum Genet* **135**, 851-867 (2016).
98. Black, D.L. & Grabowski, P.J. Alternative pre-mRNA splicing and neuronal function. *Progress in molecular and subcellular biology* **31**, 187-216 (2003).
99. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomedical reports* **3**, 152-158 (2015).
100. Wongpalee, S.P. & Sharma, S. The pre-mRNA splicing reaction. *Methods Mol Biol* **1126**, 3-12 (2014).

101. Ellis, J.D. *et al.* Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol Cell* **46**, 884-892 (2012).
102. Fu, R.H. *et al.* Aberrant alternative splicing events in Parkinson's disease. *Cell transplantation* **22**, 653-661 (2013).
103. Andreadis, A. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. *Biochim Biophys Acta* **1739**, 91-103 (2005).
104. Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. *Signal transduction and targeted therapy* **6**, 78 (2021).
105. Cartegni, L., Chew, S.L. & Krainer, A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298 (2002).
106. Lee, S.C. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat Med* **22**, 976-986 (2016).
107. Qian, W. & Liu, F. Regulation of alternative splicing of tau exon 10. *Neurosci Bull* **30**, 367-377 (2014).
108. Goedert, M., Spillantini, M.G., Jakes, R., Rutherford, D. & Crowther, R.A. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3**, 519-526 (1989).
109. Guo, T., Noble, W. & Hanger, D.P. Roles of tau protein in health and disease. *Acta Neuropathol* **133**, 665-704 (2017).
110. Fischer, I. & Baas, P.W. Resurrecting the Mysteries of Big Tau. *Trends in neurosciences* **43**, 493-504 (2020).
111. Hefti, M.M. *et al.* High-resolution temporal and regional mapping of MAPT expression and splicing in human brain development. *PLoS One* **13**, e0195771 (2018).
112. Goedert, M. & Jakes, R. Expression of separate isoforms of human tau protein: correlation with the tau pattern in brain and effects on tubulin polymerization. *EMBO J* **9**, 4225-4230 (1990).
113. Goedert, M., Spillantini, M.G., Potier, M.C., Ulrich, J. & Crowther, R.A. Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats: differential expression of tau protein mRNAs in human brain. *EMBO J* **8**, 393-399 (1989).
114. Spillantini, M.G. *et al.* Mutation in the tau gene in familial multiple system tauopathy with presenile dementia. *Proc Natl Acad Sci U S A* **95**, 7737-7741 (1998).
115. Poorkaj, P. *et al.* Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Ann Neurol* **43**, 815-825 (1998).
116. Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature* **393**, 702-705 (1998).
117. Ghetti, B. *et al.* Invited review: Frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging. *Neuropathol Appl Neurobiol* **41**, 24-46 (2015).
118. Dawson, H.N., Cantillana, V., Chen, L. & Vitek, M.P. The tau N279K exon 10 splicing mutation recapitulates frontotemporal dementia and parkinsonism linked to chromosome 17 tauopathy in a mouse model. *J Neurosci* **27**, 9155-9168 (2007).

119. Umeda, T. *et al.* Neurodegenerative disorder FTDP-17-related tau intron 10 +16C → T mutation increases tau exon 10 splicing and causes tauopathy in transgenic mice. *The American journal of pathology* **183**, 211-225 (2013).
120. Wobst, H.J. *et al.* Increased 4R tau expression and behavioural changes in a novel MAPT-N296H genomic mouse model of tauopathy. *Sci Rep* **7**, 43198 (2017).
121. Nakano, M. *et al.* Unclassified four-repeat tauopathy associated with familial parkinsonism and progressive respiratory failure. *Acta Neuropathol Commun* **8**, 148 (2020).
122. D'Souza, I. & Schellenberg, G.D. Determinants of 4-repeat tau expression. Coordination between enhancing and inhibitory splicing sequences for exon 10 inclusion. *J Biol Chem* **275**, 17700-17709 (2000).
123. D'Souza, I. & Schellenberg, G.D. tau Exon 10 expression involves a bipartite intron 10 regulatory sequence and weak 5' and 3' splice sites. *J Biol Chem* **277**, 26587-26599 (2002).
124. D'Souza, I. & Schellenberg, G.D. Regulation of tau isoform expression and dementia. *Biochim Biophys Acta* **1739**, 104-115 (2005).
125. Fu, X.D. The superfamily of arginine/serine-rich splicing factors. *RNA (New York, N.Y.)* **1**, 663-680 (1995).
126. Dreyfuss, G., Kim, V.N. & Kataoka, N. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**, 195-205 (2002).
127. Zahler, A.M., Neugebauer, K.M., Lane, W.S. & Roth, M.B. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* **260**, 219-222 (1993).
128. Kamma, H., Portman, D.S. & Dreyfuss, G. Cell type-specific expression of hnRNP proteins. *Experimental cell research* **221**, 187-196 (1995).
129. Karni, R. *et al.* The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14**, 185-193 (2007).
130. 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 Delta280K AND N279K. *J Biol Chem* **281**, 2460-2469 (2006).
131. Mabon, S.A. & Misteli, T. Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo. *PLoS biology* **3**, e374 (2005).
132. Qian, W. *et al.* Regulation of the alternative splicing of tau exon 10 by SC35 and Dyrk1A. *Nucleic Acids Res* **39**, 6161-6171 (2011).
133. Liu, F. & Gong, C.X. Tau exon 10 alternative splicing and tauopathies. *Mol Neurodegener* **3**, 8 (2008).
134. Zhou, Z. & Fu, X.D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191-207 (2013).
135. Orozco, D. *et al.* Loss of fused in sarcoma (FUS) promotes pathological Tau splicing. *EMBO reports* **13**, 759-764 (2012).
136. Gu, J. *et al.* Transactive response DNA-binding protein 43 (TDP-43) regulates alternative splicing of tau exon 10: Implications for the pathogenesis of tauopathies. *J Biol Chem* **292**, 10600-10612 (2017).

137. Jovičić, A., Paul, J.W., 3rd & Gitler, A.D. Nuclear transport dysfunction: a common theme in amyotrophic lateral sclerosis and frontotemporal dementia. *J Neurochem* **138 Suppl 1**, 134-144 (2016).
138. Mackenzie, I.R., Rademakers, R. & Neumann, M. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol* **9**, 995-1007 (2010).
139. Vance, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208-1211 (2009).
140. Dormann, D. *et al.* ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *Embo j* **29**, 2841-2857 (2010).
141. Feneberg, E., Gray, E., Ansorge, O., Talbot, K. & Turner, M.R. Towards a TDP-43-Based Biomarker for ALS and FTLD. *Molecular neurobiology* **55**, 7789-7801 (2018).
142. Buee, L., Bussiere, T., Buee-Scherrer, V., Delacourte, A. & Hof, P.R. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. *Brain Res Brain Res Rev* **33**, 95-130 (2000).
143. 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**, 769-782 (1994).
144. Goode, B.L., Chau, M., Denis, P.E. & Feinstein, S.C. Structural and functional differences between 3-repeat and 4-repeat tau isoforms. Implications for normal tau function and the onset of neurodegenetative disease. *J Biol Chem* **275**, 38182-38189 (2000).
145. Gustke, N., Trinczek, B., Biernat, J., Mandelkow, E.M. & Mandelkow, E. Domains of tau protein and interactions with microtubules. *Biochemistry* **33**, 9511-9522 (1994).
146. Butner, K.A. & Kirschner, M.W. Tau protein binds to microtubules through a flexible array of distributed weak sites. *J Cell Biol* **115**, 717-730 (1991).
147. Bunker, J.M., Wilson, L., Jordan, M.A. & Feinstein, S.C. Modulation of microtubule dynamics by tau in living cells: implications for development and neurodegeneration. *Mol Biol Cell* **15**, 2720-2728 (2004).
148. Panda, D., Samuel, J.C., Massie, M., Feinstein, S.C. & Wilson, L. Differential regulation of microtubule dynamics by three- and four-repeat tau: implications for the onset of neurodegenerative disease. *Proc Natl Acad Sci U S A* **100**, 9548-9553 (2003).
149. Xu, C. *et al.* Co-Expression of Three Wild-Type 3R-Tau Isoforms Induces Memory Deficit via Oxidation-Related DNA Damage and Cell Death: A Promising Model for Tauopathies. *Journal of Alzheimer's disease : JAD* **73**, 1105-1123 (2020).
150. 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**, e84849 (2013).
151. Liu, C., Song, X., Nisbet, R. & Gotz, 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).
152. Spittaels, K. *et al.* Prominent axonopathy in the brain and spinal cord of transgenic mice overexpressing four-repeat human tau protein. *The American journal of pathology* **155**, 2153-2165 (1999).
153. Ishihara, T. *et al.* Age-dependent emergence and progression of a tauopathy in transgenic mice overexpressing the shortest human tau isoform. *Neuron* **24**, 751-762 (1999).

154. Probst, A. *et al.* Axonopathy and amyotrophy in mice transgenic for human four-repeat tau protein. *Acta Neuropathol* **99**, 469-481 (2000).
155. Duff, K. *et al.* Characterization of pathology in transgenic mice over-expressing human genomic and cDNA tau transgenes. *Neurobiol Dis* **7**, 87-98 (2000).
156. Dumanchin, C. *et al.* Segregation of a missense mutation in the microtubule-associated protein tau gene with familial frontotemporal dementia and parkinsonism. *Hum Mol Genet* **7**, 1825-1829 (1998).
157. Lewis, J. *et al.* Neurofibrillary tangles, amyotrophy and progressive motor disturbance in mice expressing mutant (P301L) tau protein. *Nat Genet* **25**, 402-405 (2000).
158. Gotz, J., Chen, F., Barmettler, R. & Nitsch, R.M. Tau filament formation in transgenic mice expressing P301L tau. *J Biol Chem* **276**, 529-534 (2001).
159. Tanemura, K. *et al.* Formation of filamentous tau aggregations in transgenic mice expressing V337M human tau. *Neurobiol Dis* **8**, 1036-1045 (2001).
160. Oddo, S. *et al.* Triple-transgenic model of Alzheimer's disease with plaques and tangles: intracellular Abeta and synaptic dysfunction. *Neuron* **39**, 409-421 (2003).
161. Sorbi, S. *et al.* Genetic risk factors in familial Alzheimer's disease. *Mechanisms of ageing and development* **122**, 1951-1960 (2001).
162. Yen, S.H., Hutton, M., DeTure, M., Ko, L.W. & Nacharaju, P. Fibrillogenesis of tau: insights from tau missense mutations in FTDP-17. *Brain pathology (Zurich, Switzerland)* **9**, 695-705 (1999).
163. Hasegawa, M., Smith, M.J., Iijima, M., Tabira, T. & Goedert, M. FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. *FEBS Lett* **443**, 93-96 (1999).
164. Spillantini, M.G., Van Swieten, J.C. & Goedert, M. Tau gene mutations in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). *Neurogenetics* **2**, 193-205 (2000).
165. Spillantini, M.G. & Goedert, M. Tau mutations in familial frontotemporal dementia. *Brain* **123 (Pt 5)**, 857-859 (2000).
166. Miyasaka, T. *et al.* Molecular analysis of mutant and wild-type tau deposited in the brain affected by the FTDP-17 R406W mutation. *The American journal of pathology* **158**, 373-379 (2001).
167. Bejanin, A. *et al.* Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain* **140**, 3286-3300 (2017).
168. Jeong, S. Molecular and Cellular Basis of Neurodegeneration in Alzheimer's Disease. *Molecules and cells* **40**, 613-620 (2017).
169. Saha, P. & Sen, N. Tauopathy: A common mechanism for neurodegeneration and brain aging. *Mechanisms of ageing and development* **178**, 72-79 (2019).
170. Malpetti, M. *et al.* Microglial activation and tau burden predict cognitive decline in Alzheimer's disease. *Brain* **143**, 1588-1602 (2020).
171. Giannakopoulos, P. *et al.* Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology* **60**, 1495-1500 (2003).
172. VandeVrede, L., Boxer, A.L. & Polydoro, M. Targeting tau: Clinical trials and novel therapeutic approaches. *Neuroscience letters* **731**, 134919 (2020).
173. Fischbeck, K.H. & Wexler, N.S. Oligonucleotide Treatment for Huntington's Disease. *The New England journal of medicine* **380**, 2373-2374 (2019).

174. Scoles, D.R., Minikel, E.V. & Pulst, S.M. Antisense oligonucleotides: A primer. *Neurology. Genetics* **5**, e323 (2019).
175. DeVos, S.L. *et al.* Antisense reduction of tau in adult mice protects against seizures. *J Neurosci* **33**, 12887-12897 (2013).
176. DeVos, S.L. *et al.* Tau reduction prevents neuronal loss and reverses pathological tau deposition and seeding in mice with tauopathy. *Science translational medicine* **9** (2017).
177. Schoch, K.M. *et al.* Increased 4R-Tau Induces Pathological Changes in a Human-Tau Mouse Model. *Neuron* **90**, 941-947 (2016).
178. Murray, P.S. *et al.* Hyperphosphorylated tau is elevated in Alzheimer's disease with psychosis. *Journal of Alzheimer's disease : JAD* **39**, 759-773 (2014).
179. Chen, G., Huang, L.D., Jiang, Y.M. & Manji, H.K. The mood-stabilizing agent valproate inhibits the activity of glycogen synthase kinase-3. *J Neurochem* **72**, 1327-1330 (1999).
180. Martinez, A., Alonso, M., Castro, A., Pérez, C. & Moreno, F.J. First non-ATP competitive glycogen synthase kinase 3 beta (GSK-3beta) inhibitors: thiadiazolidinones (TDZD) as potential drugs for the treatment of Alzheimer's disease. *Journal of medicinal chemistry* **45**, 1292-1299 (2002).
181. Noble, W. *et al.* Inhibition of glycogen synthase kinase-3 by lithium correlates with reduced tauopathy and degeneration in vivo. *Proc Natl Acad Sci U S A* **102**, 6990-6995 (2005).
182. Qing, H. *et al.* Valproic acid inhibits Abeta production, neuritic plaque formation, and behavioral deficits in Alzheimer's disease mouse models. *The Journal of experimental medicine* **205**, 2781-2789 (2008).
183. del Ser, T. *et al.* Treatment of Alzheimer's disease with the GSK-3 inhibitor tideglusib: a pilot study. *Journal of Alzheimer's disease : JAD* **33**, 205-215 (2013).
184. Kontsekova, E., Zilka, N., Kovacech, B., Novak, P. & Novak, M. First-in-man tau vaccine targeting structural determinants essential for pathological tau-tau interaction reduces tau oligomerisation and neurofibrillary degeneration in an Alzheimer's disease model. *Alzheimer's research & therapy* **6**, 44 (2014).
185. Meredith, J.E., Jr. *et al.* Characterization of novel CSF Tau and ptau biomarkers for Alzheimer's disease. *PLoS One* **8**, e76523 (2013).
186. Borroni, B. *et al.* Pattern of Tau forms in CSF is altered in progressive supranuclear palsy. *Neurobiology of aging* **30**, 34-40 (2009).
187. Sopko, R. *et al.* Characterization of tau binding by gosuranemab. *Neurobiol Dis* **146**, 105120 (2020).
188. Yanamandra, K. *et al.* Anti-tau antibodies that block tau aggregate seeding in vitro markedly decrease pathology and improve cognition in vivo. *Neuron* **80**, 402-414 (2013).
189. Boxer, A.L. *et al.* Safety of the tau-directed monoclonal antibody BIIB092 in progressive supranuclear palsy: a randomised, placebo-controlled, multiple ascending dose phase 1b trial. *Lancet Neurol* **18**, 549-558 (2019).
190. Alam, R. *et al.* Preclinical Characterization of an Antibody [Ly3303560] Targeting Aggregated Tau. *Alzheimer's & Dementia* **13**, P592-P593 (2017).
191. Konstantinidou, M. *et al.* PROTACs- a game-changing technology. *Expert opinion on drug discovery* **14**, 1255-1268 (2019).
192. Zou, Y., Ma, D. & Wang, Y. The PROTAC technology in drug development. *Cell biochemistry and function* **37**, 21-30 (2019).

193. Lu, M. *et al.* Discovery of a Keap1-dependent peptide PROTAC to knockdown Tau by ubiquitination-proteasome degradation pathway. *European journal of medicinal chemistry* **146**, 251-259 (2018).
194. Cuadrado, A. *et al.* Therapeutic targeting of the NRF2 and KEAP1 partnership in chronic diseases. *Nature reviews. Drug discovery* **18**, 295-317 (2019).
195. Gozes, I. *et al.* A novel signaling molecule for neuropeptide action: activity-dependent neuroprotective protein. *Annals of the New York Academy of Sciences* **897**, 125-135 (1999).
196. Clark, J.A., Chuckowree, J.A., Dyer, M.S., Dickson, T.C. & Blizzard, C.A. Epothilone D alters normal growth, viability and microtubule dependent intracellular functions of cortical neurons in vitro. *Sci Rep* **10**, 918 (2020).
197. Shiryaev, N. *et al.* NAP protects memory, increases soluble tau and reduces tau hyperphosphorylation in a tauopathy model. *Neurobiol Dis* **34**, 381-388 (2009).
198. Brunden, K.R. *et al.* Epothilone D improves microtubule density, axonal integrity, and cognition in a transgenic mouse model of tauopathy. *J Neurosci* **30**, 13861-13866 (2010).
199. Boxer, A.L. *et al.* Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. *Lancet Neurol* **13**, 676-685 (2014).
200. Kondo, S. *et al.* Tra2 beta, SF2/ASF and SRp30c modulate the function of an exonic splicing enhancer in exon 10 of tau pre-mRNA. *Genes Cells* **9**, 121-130 (2004).
201. Grammatikakis, I. *et al.* Alternative Splicing of Neuronal Differentiation Factor TRF2 Regulated by HNRNPH1/H2. *Cell reports* **15**, 926-934 (2016).
202. Pastor, T. & Pagani, F. Interaction of hnRNPA1/A2 and DAZAP1 with an Alu-derived intronic splicing enhancer regulates ATM aberrant splicing. *PLoS One* **6**, e23349 (2011).
203. Yoon, J.H. & Gorospe, M. Identification of mRNA-Interacting Factors by MS2-TRAP (MS2-Tagged RNA Affinity Purification). *Methods Mol Biol* **1421**, 15-22 (2016).
204. Kar, A. *et al.* RNA helicase p68 (DDX5) regulates tau exon 10 splicing by modulating a stem-loop structure at the 5' splice site. *Mol Cell Biol* **31**, 1812-1821 (2011).
205. Miro, J., Bourgeois, C.F., Claustres, M., Koenig, M. & Tuffery-Giraud, S. Identification of Splicing Factors Involved in DMD Exon Skipping Events Using an In Vitro RNA Binding Assay. *Methods Mol Biol* **1687**, 157-169 (2018).
206. Zheng, X. *et al.* Detecting RNA-Protein Interaction Using End-Labeled Biotinylated RNA Oligonucleotides and Immunoblotting. *Methods Mol Biol* **1421**, 35-44 (2016).
207. Castello, A. *et al.* System-wide identification of RNA-binding proteins by interactome capture. *Nat Protoc* **8**, 491-500 (2013).
208. Chu, C. *et al.* Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).
209. McHugh, C.A. *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).
210. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E. & Chang, H.Y. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* **44**, 667-678 (2011).
211. Roth, A. & Diederichs, S. Molecular biology: Rap and chirp about X inactivation. *Nature* **521**, 170-171 (2015).
212. Dunham, W.H., Mullin, M. & Gingras, A.C. Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* **12**, 1576-1590 (2012).

213. Zhou, Y. & Zou, P. The evolving capabilities of enzyme-mediated proximity labeling. *Current opinion in chemical biology* **60**, 30-38 (2020).
214. Snider, J. *et al.* Fundamentals of protein interaction network mapping. *Mol Syst Biol* **11**, 848 (2015).
215. Gingras, A.C., Abe, K.T. & Raught, B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. *Current opinion in chemical biology* **48**, 44-54 (2019).
216. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801-810 (2012).
217. Chapman-Smith, A. & Cronan, J.E., Jr. The enzymatic biotinylation of proteins: a post-translational modification of exceptional specificity. *Trends in biochemical sciences* **24**, 359-363 (1999).
218. Choi-Rhee, E., Schulman, H. & Cronan, J.E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Sci* **13**, 3043-3050 (2004).
219. Lane, M.D., Rominger, K.L., Young, D.L. & Lynen, F. THE ENZYMATIC SYNTHESIS OF HOLOTRANSCARBOXYLASE FROM APOTRANSCARBOXYLASE AND (+)-BIOTIN. II. INVESTIGATION OF THE REACTION MECHANISM. *J Biol Chem* **239**, 2865-2871 (1964).
220. Kim, D.I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Mol Biol Cell* **27**, 1188-1196 (2016).
221. Branon, T.C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **36**, 880-887 (2018).
222. Martell, J.D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat Biotechnol* **30**, 1143-1148 (2012).
223. Xu, Y., Fan, X. & Hu, Y. In vivo interactome profiling by enzyme-catalyzed proximity labeling. *Cell & bioscience* **11**, 27 (2021).
224. Lam, S.S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* **12**, 51-54 (2015).
225. Bosch, J.A., Chen, C.L. & Perrimon, N. Proximity-dependent labeling methods for proteomic profiling in living cells: An update. *Wiley interdisciplinary reviews. Developmental biology* **10**, e392 (2021).
226. Rhee, H.W. *et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328-1331 (2013).
227. Liu, Q. *et al.* A proximity-tagging system to identify membrane protein-protein interactions. *Nat Methods* **15**, 715-722 (2018).
228. Striebel, F. *et al.* Bacterial ubiquitin-like modifier Pup is deamidated and conjugated to substrates by distinct but homologous enzymes. *Nat Struct Mol Biol* **16**, 647-651 (2009).
229. Pearce, M.J., Mintseris, J., Ferreyra, J., Gygi, S.P. & Darwin, K.H. Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science* **322**, 1104-1107 (2008).
230. Guth, E., Thommen, M. & Weber-Ban, E. Mycobacterial ubiquitin-like protein ligase PafA follows a two-step reaction pathway with a phosphorylated pup intermediate. *J Biol Chem* **286**, 4412-4419 (2011).

231. McCormack, A.L. *et al.* Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem* **69**, 767-776 (1997).
232. Dowell, J.A., Frost, D.C., Zhang, J. & Li, L. Comparison of two-dimensional fractionation techniques for shotgun proteomics. *Anal Chem* **80**, 6715-6723 (2008).
233. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
234. Pappireddi, N., Martin, L. & Wühr, M. A Review on Quantitative Multiplexed Proteomics. *Chembiochem : a European journal of chemical biology* **20**, 1210-1224 (2019).
235. Schubert, O.T., Röst, H.L., Collins, B.C., Rosenberger, G. & Aebersold, R. Quantitative proteomics: challenges and opportunities in basic and applied research. *Nat Protoc* **12**, 1289-1294 (2017).
236. Eng, J.K., McCormack, A.L. & Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* **5**, 976-989 (1994).
237. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci U S A* **100**, 6940-6945 (2003).
238. Johnson, E.C.B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med* (2020).
239. Lazar, C., Gatto, L., Ferro, M., Bruley, C. & Burger, T. Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *Journal of proteome research* **15**, 1116-1125 (2016).
240. Schnolzer, M., Jedrzejewski, P. & Lehmann, W.D. Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17**, 945-953 (1996).
241. Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999 (1999).
242. Stewart, II, Thomson, T. & Figeys, D. 18O labeling: a tool for proteomics. *Rapid Commun Mass Spectrom* **15**, 2456-2465 (2001).
243. Ong, S.E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).
244. Overmyer, K.A. *et al.* Multiplexed proteome analysis with neutron-encoded stable isotope labeling in cells and mice. *Nat Protoc* **13**, 293-306 (2018).
245. Wu, C.C., MacCoss, M.J., Howell, K.E., Matthews, D.E. & Yates, J.R., 3rd Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* **76**, 4951-4959 (2004).
246. Savas, J.N. *et al.* Amyloid Accumulation Drives Proteome-wide Alterations in Mouse Models of Alzheimer's Disease-like Pathology. *Cell reports* **21**, 2614-2627 (2017).
247. McClatchy, D.B., Dong, M.Q., Wu, C.C., Venable, J.D. & Yates, J.R., 3rd 15N metabolic labeling of mammalian tissue with slow protein turnover. *Journal of proteome research* **6**, 2005-2010 (2007).

248. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **75**, 1895-1904 (2003).
249. Dayon, L. *et al.* Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. *Anal Chem* **80**, 2921-2931 (2008).
250. Werner, T. *et al.* Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal Chem* **86**, 3594-3601 (2014).
251. Thompson, A. *et al.* TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal Chem* **91**, 15941-15950 (2019).
252. Savitski, M.M. *et al.* Measuring and managing ratio compression for accurate iTRAQ/TMT quantification. *Journal of proteome research* **12**, 3586-3598 (2013).
253. Ting, L., Rad, R., Gygi, S.P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods* **8**, 937-940 (2011).
254. Kovacs, G.G. Invited review: Neuropathology of tauopathies: principles and practice. *Neuropathol Appl Neurobiol* **41**, 3-23 (2015).
255. Barbier, P. *et al.* Role of Tau as a Microtubule-Associated Protein: Structural and Functional Aspects. *Front Aging Neurosci* **11**, 204 (2019).
256. Avila, J. Tau phosphorylation and aggregation in Alzheimer's disease pathology. *FEBS Lett* **580**, 2922-2927 (2006).
257. Kar, A., Kuo, D., He, R., Zhou, J. & Wu, J.Y. Tau alternative splicing and frontotemporal dementia. *Alzheimer Dis Assoc Disord* **19 Suppl 1**, S29-36 (2005).
258. Hong, M. *et al.* Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17. *Science* **282**, 1914-1917 (1998).
259. Lamb, R., Rohrer, J.D., Lees, A.J. & Morris, H.R. Progressive Supranuclear Palsy and Corticobasal Degeneration: Pathophysiology and Treatment Options. *Curr Treat Options Neurol* **18**, 42 (2016).
260. 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 Tra2 beta. *J Biol Chem* **278**, 18997-19007 (2003).
261. 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).
262. Wu, J.Y., Kar, A., Kuo, D., Yu, B. & Havlioglu, N. SRp54 (SFRS11), a regulator for tau exon 10 alternative splicing identified by an expression cloning strategy. *Mol Cell Biol* **26**, 6739-6747 (2006).
263. Ishigaki, S. *et al.* Altered Tau Isoform Ratio Caused by Loss of FUS and SFPQ Function Leads to FTLD-like Phenotypes. *Cell reports* **18**, 1118-1131 (2017).
264. 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).
265. Wang, Y. *et al.* Tau exons 2 and 10, which are misregulated in neurodegenerative diseases, are partly regulated by silencers which bind a SRp30c.SRp55 complex that either recruits or antagonizes htra2beta1. *J Biol Chem* **280**, 14230-14239 (2005).
266. Luo, C. *et al.* SRSF2 Regulates Alternative Splicing to Drive Hepatocellular Carcinoma Development. *Cancer Res* **77**, 1168-1178 (2017).
267. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat Commun* **7**, 10615 (2016).

268. David, C.J., Chen, M., Assanah, M., Canoll, P. & Manley, J.L. HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**, 364-368 (2010).
269. Dufner-Almeida, L.G., do Carmo, R.T., Masotti, C. & Haddad, L.A. Understanding human DNA variants affecting pre-mRNA splicing in the NGS era. *Adv Genet* **103**, 39-90 (2019).
270. Group, P.T.C. *et al.* Genomic basis for RNA alterations in cancer. *Nature* **578**, 129-136 (2020).
271. Ule, J. & Blencowe, B.J. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* **76**, 329-345 (2019).
272. Barnes, C. & Kanhere, A. Identification of RNA-Protein Interactions Through In Vitro RNA Pull-Down Assays. *Methods Mol Biol* **1480**, 99-113 (2016).
273. Leppek, K. & Stoecklin, G. An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins. *Nucleic Acids Res* **42**, e13 (2014).
274. Yoon, J.H., Srikantan, S. & Gorospe, M. MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. *Methods* **58**, 81-87 (2012).
275. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
276. Marquez, A.M. *et al.* Oxygen Exposure During Cardiopulmonary Resuscitation Is Associated With Cerebral Oxidative Injury in a Randomized, Blinded, Controlled, Preclinical Trial. *J Am Heart Assoc* **9**, e015032 (2020).
277. 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).
278. Baruzzo, G. *et al.* Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods* **14**, 135-139 (2017).
279. Allen, M. *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data* **3**, 160089 (2016).
280. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
281. 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).
282. Gao, L., Wang, J., Wang, Y. & Andreadis, A. SR protein 9G8 modulates splicing of tau exon 10 via its proximal downstream intron, a clustering region for frontotemporal dementia mutations. *Mol Cell Neurosci* **34**, 48-58 (2007).
283. Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152**, 453-466 (2013).
284. Hutten, S. & Dormann, D. hnRNPA2/B1 Function in Neurodegeneration: It's a Gain, Not a Loss. *Neuron* **92**, 672-674 (2016).
285. Wang, J. *et al.* Tau exon 10, whose missplicing causes frontotemporal dementia, is regulated by an intricate interplay of cis elements and trans factors. *J Neurochem* **88**, 1078-1090 (2004).
286. Chapple, J.P. *et al.* Expression, localization and tau exon 10 splicing activity of the brain RNA-binding protein TNRC4. *Hum Mol Genet* **16**, 2760-2769 (2007).
287. Swanson, M.S., Nakagawa, T.Y., LeVan, K. & Dreyfuss, G. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain

structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol Cell Biol* **7**, 1731-1739 (1987).
288. Huang, H. *et al.* HNRNPC as a candidate biomarker for chemoresistance in gastric cancer. *Tumour Biol* **37**, 3527-3534 (2016).
289. Wu, Y. *et al.* Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. *EMBO J* **37** (2018).
290. Fischl, H. *et al.* hnRNPC regulates cancer-specific alternative cleavage and polyadenylation profiles. *Nucleic Acids Research* **47**, 7580-7591 (2019).
291. Dowling, P. *et al.* Abnormal levels of heterogeneous nuclear ribonucleoprotein A2B1 (hnRNPA2B1) in tumour tissue and blood samples from patients diagnosed with lung cancer. *Mol Biosyst* **11**, 743-752 (2015).
292. Hu, Y. *et al.* Splicing factor hnRNPA2B1 contributes to tumorigenic potential of breast cancer cells through STAT3 and ERK1/2 signaling pathway. *Tumour Biol* **39**, 1010428317694318 (2017).
293. Chen, Z.Y. *et al.* Fyn requires HnRNPA2B1 and Sam68 to synergistically regulate apoptosis in pancreatic cancer. *Carcinogenesis* **32**, 1419-1426 (2011).
294. Golan-Gerstl, R. *et al.* Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res* **71**, 4464-4472 (2011).
295. Jing, G.J. *et al.* Aberrant expression and localization of hnRNP-A2/B1 is a common event in human gastric adenocarcinoma. *J Gastroenterol Hepatol* **26**, 108-115 (2011).
296. Yan-Sanders, Y., Hammons, G.J. & Lyn-Cook, B.D. Increased expression of heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP) in pancreatic tissue from smokers and pancreatic tumor cells. *Cancer Lett* **183**, 215-220 (2002).
297. Martinez, F.J. *et al.* Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron* **92**, 780-795 (2016).
298. He, F., Krans, A., Freibaum, B.D., Taylor, J.P. & Todd, P.K. TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Hum Mol Genet* **23**, 5036-5051 (2014).
299. Wang, E., Dimova, N. & Cambi, F. PLP/DM20 ratio is regulated by hnRNPH and F and a novel G-rich enhancer in oligodendrocytes. *Nucleic Acids Res* **35**, 4164-4178 (2007).
300. Wang, E. & Cambi, F. Heterogeneous nuclear ribonucleoproteins H and F regulate the proteolipid protein/DM20 ratio by recruiting U1 small nuclear ribonucleoprotein through a complex array of G runs. *J Biol Chem* **284**, 11194-11204 (2009).
301. Wang, E. *et al.* Global profiling of alternative splicing events and gene expression regulated by hnRNPH/F. *PLoS One* **7**, e51266 (2012).
302. Orr, M.E., Sullivan, A.C. & Frost, B. A Brief Overview of Tauopathy: Causes, Consequences, and Therapeutic Strategies. *Trends in pharmacological sciences* **38**, 637-648 (2017).
303. Matus, A. Microtubule-associated proteins. *Curr Opin Cell Biol* **2**, 10-14 (1990).
304. Dehmelt, L. & Halpain, S. The MAP2/Tau family of microtubule-associated proteins. *Genome Biol* **6**, 204 (2005).
305. Halpain, S. & Dehmelt, L. The MAP1 family of microtubule-associated proteins. *Genome Biol* **7**, 224 (2006).
306. Mandell, J.W. & Banker, G.A. A spatial gradient of tau protein phosphorylation in nascent axons. *J Neurosci* **16**, 5727-5740 (1996).

307. Ittner, A. & Ittner, L.M. Dendritic Tau in Alzheimer's Disease. *Neuron* **99**, 13-27 (2018).
308. Johnson, G.V. & Stoothoff, W.H. Tau phosphorylation in neuronal cell function and dysfunction. *J Cell Sci* **117**, 5721-5729 (2004).
309. Min, S.W. *et al.* Acetylation of tau inhibits its degradation and contributes to tauopathy. *Neuron* **67**, 953-966 (2010).
310. Luo, H.B. *et al.* SUMOylation at K340 inhibits tau degradation through deregulating its phosphorylation and ubiquitination. *Proc Natl Acad Sci U S A* **111**, 16586-16591 (2014).
311. Arakhamia, T. *et al.* Posttranslational Modifications Mediate the Structural Diversity of Tauopathy Strains. *Cell* **180**, 633-644 e612 (2020).
312. Gunawardana, C.G. *et al.* The Human Tau Interactome: Binding to the Ribonucleoproteome, and Impaired Binding of the Proline-to-Leucine Mutant at Position 301 (P301L) to Chaperones and the Proteasome. *Mol Cell Proteomics* **14**, 3000-3014 (2015).
313. Vanderweyde, T. *et al.* Interaction of tau with the RNA-Binding Protein TIA1 Regulates tau Pathophysiology and Toxicity. *Cell reports* **15**, 1455-1466 (2016).
314. Koren, S.A. *et al.* Tau drives translational selectivity by interacting with ribosomal proteins. *Acta Neuropathol* **137**, 571-583 (2019).
315. Ewing, R.M. *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* **3**, 89 (2007).
316. Trinkle-Mulcahy, L. *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J Cell Biol* **183**, 223-239 (2008).
317. Gavin, A.C., Maeda, K. & Kuhner, S. Recent advances in charting protein-protein interaction: mass spectrometry-based approaches. *Curr Opin Biotechnol* **22**, 42-49 (2011).
318. Choi-Rhee, E., Schulman, H. & Cronan, J.E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Sci* **13**, 3043-3050 (2004).
319. Cronan, J.E. Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. *J Nutr Biochem* **16**, 416-418 (2005).
320. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801-810 (2012).
321. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
322. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
323. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
324. Karch, C.M. & Goate, A.M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biological psychiatry* **77**, 43-51 (2015).
325. Zhao, L. *et al.* A quantitative trait rare variant nonparametric linkage method with application to age-at-onset of Alzheimer's disease. *European journal of human genetics : EJHG* **28**, 1734-1742 (2020).
326. Delamarre, A. & Meissner, W.G. Epidemiology, environmental risk factors and genetics of Parkinson's disease. *Presse medicale (Paris, France : 1983)* **46**, 175-181 (2017).

327. Huang, H. *et al.* Genetic association of NOS1 exon18, NOS1 exon29, ABCB1 1236C/T, and ABCB1 3435C/T polymorphisms with the risk of Parkinson's disease: A meta-analysis. *Medicine* **95**, e4982 (2016).
328. Pottier, C., Ravenscroft, T.A., Sanchez-Contreras, M. & Rademakers, R. Genetics of FTLD: overview and what else we can expect from genetic studies. *J Neurochem* **138 Suppl 1**, 32-53 (2016).
329. Frau, M., Feo, F. & Pascale, R.M. Pleiotropic effects of methionine adenosyltransferases deregulation as determinants of liver cancer progression and prognosis. *Journal of hepatology* **59**, 830-841 (2013).
330. Polydoro, M. *et al.* Soluble pathological tau in the entorhinal cortex leads to presynaptic deficits in an early Alzheimer's disease model. *Acta Neuropathol* **127**, 257-270 (2014).
331. Xie, A.J. *et al.* Tau overexpression impairs neuronal endocytosis by decreasing the GTPase dynamin 1 through the miR-132/MeCP2 pathway. *Aging cell* **18**, e12929 (2019).
332. Gamblin, T.C., King, M.E., Kuret, J., Berry, R.W. & Binder, L.I. Oxidative regulation of fatty acid-induced tau polymerization. *Biochemistry* **39**, 14203-14210 (2000).
333. Zhao, C. *et al.* MAT2A promotes porcine adipogenesis by mediating H3K27me3 at Wnt10b locus and repressing Wnt/β-catenin signaling. *Biochimica et biophysica acta. Molecular and cell biology of lipids* **1863**, 132-142 (2018).
334. Secker, K.A. *et al.* MAT2A as Key Regulator and Therapeutic Target in MLLr Leukemogenesis. *Cancers* **12** (2020).
335. Murray, B., Barbier-Torres, L., Fan, W., Mato, J.M. & Lu, S.C. Methionine adenosyltransferases in liver cancer. *World journal of gastroenterology* **25**, 4300-4319 (2019).
336. Yang, H., Li, T.W., Peng, J., Mato, J.M. & Lu, S.C. Insulin-like growth factor 1 activates methionine adenosyltransferase 2A transcription by multiple pathways in human colon cancer cells. *The Biochemical journal* **436**, 507-516 (2011).
337. Cai, J., Sun, W.M., Hwang, J.J., Stain, S.C. & Lu, S.C. Changes in S-adenosylmethionine synthetase in human liver cancer: molecular characterization and significance. *Hepatology (Baltimore, Md.)* **24**, 1090-1097 (1996).
338. Quinlan, C.L. *et al.* Targeting S-adenosylmethionine biosynthesis with a novel allosteric inhibitor of Mat2A. *Nature chemical biology* **13**, 785-792 (2017).
339. Kalev, P. *et al.* MAT2A Inhibition Blocks the Growth of MTAP-Deleted Cancer Cells by Reducing PRMT5-Dependent mRNA Splicing and Inducing DNA Damage. *Cancer cell* **39**, 209-224.e211 (2021).
340. Linnebank, M. *et al.* S-adenosylmethionine is decreased in the cerebrospinal fluid of patients with Alzheimer's disease. *Neuro-degenerative diseases* **7**, 373-378 (2010).
341. Schrötter, A. *et al.* The amyloid precursor protein (APP) family members are key players in S-adenosylmethionine formation by MAT2A and modify BACE1 and PSEN1 gene expression-relevance for Alzheimer's disease. *Mol Cell Proteomics* **11**, 1274-1288 (2012).
342. Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron* **104**, 239-255.e212 (2019).