RESEARCH ARTICLE

# The RNA encoding the microtubule-associated protein tau has extensive structure that affects its biology

**Jonathan L. Chen**[1], **Walter N. Moss**[2], **Adam Spencer**[1], **Peiyuan Zhang**[1], **Jessica L. Childs-Disney**[1], **Matthew D. Disney**[1]*

**1** Department of Chemistry, The Scripps Research Institute, Jupiter, Florida, United States of America, **2** Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa, United States of America

* Disney@scripps.edu

## Abstract

Tauopathies are neurodegenerative diseases that affect millions of people worldwide including those with Alzheimer's disease. While many efforts have focused on understanding the role of tau protein in neurodegeneration, there has been little done to systematically analyze and study the structures within tau's encoding RNA and their connection to disease pathology. Knowledge of RNA structure can provide insights into disease mechanisms and how to affect protein production for therapeutic benefit. Using computational methods based on thermodynamic stability and evolutionary conservation, we identified structures throughout the tau pre-mRNA, especially at exon-intron junctions and within the 5′ and 3′ untranslated regions (UTRs). In particular, structures were identified at twenty exon-intron junctions. The 5′ UTR contains one structured region, which lies within a known internal ribosome entry site. The 3′ UTR contains eight structured regions, including one that contains a polyadenylation signal. A series of functional experiments were carried out to assess the effects of mutations associated with mis-regulation of alternative splicing of exon 10 and to identify regions of the 3′ UTR that contain *cis*-regulatory elements. These studies defined novel structural regions within the mRNA that affect stability and pre-mRNA splicing and may lead to new therapeutic targets for treating tau-associated diseases.

## Introduction

RNA structures function in normal cellular processes, such as splicing, protein synthesis, and regulation of gene expression.[1] At the same time, mutations that disrupt RNA structure or formation of ribonucleoproteins (RNPs) can be deleterious and cause disease.[2] This has generated interest in targeting RNA with therapeutics. Evolutionarily conserved RNA structures across species may have common functions.[3] However, the superficial lack of sequence conservation in noncoding RNA (ncRNA) may complicate the search for conserved structures. [4,5] Thus, specialized techniques are needed for discovery of homologous structured regions

and the Huntington's Disease Society of America (J.L.C.).

in RNA.[4,6] One method to identify stable, conserved structures combines sequence alignment with thermodynamic-based folding algorithms.[6]

Tauopathies are a class of neurodegenerative diseases characterized by the presence of tau inclusion bodies.[7] Tauopathies such as Alzheimer's and Parkinson's diseases (AD and PD, respectively) are burdensome socioeconomically and affect more than 35 million and 6.3 million people, respectively, worldwide.[8] Currently available treatments are largely focused on symptoms and do not target underlying disease mechanisms.[7] The tau protein, which binds to microtubules and promotes microtubule assembly and stability, is encoded by the microtubule associated protein tau (*MAPT*) gene on chromosome 17.[8–11] The *MAPT* gene is well conserved, with 97 to 100% homology among primates.[12] This 134 kb gene is comprised of 16 exons, among which exons 2, 3, and 10 are known to be alternatively spliced, generating six isoforms ranging from 352 to 441 amino acids in length.[8,10–12] Exons 2 and 3 encode for N-terminal domains while exons 9 to 12 encode microtubule binding domains (MBD).[9] With such complex processing, the *MAPT* mRNA is likely rich in conserved regulatory structures that may have important functions and may be implicated in tau-associated diseases.

Tau proteins bind to and stabilize microtubules via their MBD repeat sequences that interact with negatively charged tubulin residues via their net positive charge.[9] Alterations in the protein coding content of the mRNA, including the number of MBDs, are due to alternative splicing. For example, exon 10 encodes an MBD and is alternatively spliced, resulting in protein isoforms with four (4R) or three (3R) microtubule-binding domains. In normal tissues, the 3R-to-4R (repeat) ratio is approximately 1. Deregulation of exon 10 alternative splicing due to mutation manifests itself in various diseases,[8–11] including frontotemporal dementia with parkinsonism-17 (FTDP-17) and Pick's disease, which express 4R and 3R tau at aberrantly high levels, respectively.[8] In FTDP-17, various mutations at the exon 10-intron 10 junction destabilize a stem-loop structure, resulting in increased interaction of the 5′ splice site with U1 snRNA and hence increased exon 10 inclusion and 4R tau levels.[13,14] These studies suggest that structures in *MAPT* pre-mRNA intron-exon junctions affect MAPT biology and that structures elsewhere in the mRNA could as well. Indeed, previous studies have shown that regions of the 3′ untranslated region (UTR) affect translation of the tau mRNA.[15]

Previous studies have identified conserved sequences in *MAPT* mRNA including the exon 6-intron 7 junction,[16] an intronic sequence upstream of exon 6 (180 nucleotides),[16] the intron downstream of exon 2,[17] and the intron downstream of exon 10, which contains several islands of conserved sequence.[18] Herein, computational studies were carried out to search for potentially conserved structures in *MAPT* pre-mRNA. Specifically, structured regions were identified at exon-intron junctions and within the long 3′ UTR isoform that may control tau expression. This method may be applied to diseases, such as tauopathies, in which RNA dysfunction contributes to disease development. Knowledge of RNA structure in the *MAPT* gene may lead to new therapeutics against RNA, such as small molecules for treatment of tauopathies.[19]

## Methods

### Structure analysis

RNA sequences were obtained from the National Center for Biotechnology Information (NCBI). Sequences were folded using RNAfold from the ViennaRNA 2.4.0 package.[20] Sequences 500 nt in length centered around each exon-intron junction were folded in 30 nt windows every 10 nt from the 5′ end. 500 nt sequences were used to identify potential long-range interactions near the splice site that may otherwise not be predicted with shorter sequences. Separately, a 4163 nt 3′ UTR sequence of the *MAPT* gene was folded in 150 nt

windows every 10 nt from the 5′ end.[15] For each sequence, the free energy of each window, the average free energy of a set of 50 randomized sequences for the same window, and a z-score for the difference between the native and average random free energies were calculated. Additionally, the native sequence ensemble diversity (ED) was calculated from a partition function.[6,21,22] The z-score is a measure of the number of standard deviations (SDs) by which the free energy of the minimum free energy (MFE) structure deviates from the average free energy of structures formed by randomized sequences.[23,24] A low z-score indicates that a structure is more stable than random and may be ordered to form structure (e.g. by evolution). The *p*-value represents the fraction of sequences in the randomized sequences whose free energy is lower than the MFE of a given sequence.[21] The ED is a measure of the difference (in base pairs) among Boltzmann-weighted structures in a predicted ensemble; here, a lower than average ED for a sequence indicates less conformationally diversity (e.g. a single dominant conformation).[24,25] Folded sequences whose z-scores were more than one SD below the average z-score of all windows were compiled and refolded to generate a new set of structures. Next, 250 nt sequences centered on each exon-intron junction were folded using 70 nt windows and 1 nt base step size. These parameters were chosen to identify relatively small motifs at exon-intron junctions that may be not be predicted by larger window or base step sizes.

NCBI BLASTn was used to search for sequence fragments in non-human primates from stable regions. Resulting sequence fragments were filtered to exclude duplicate fragments and fragments with less than 80% length of the query sequence. Sequences were aligned with the Multiple Alignment using Fast Fourier Transform (MAFFT) program[26] and folded with RNAalifold[20] to generate a consensus structure. The number of base pairs from each aligned sequence in the consensus structure was counted. The percent of canonical base pairs for each nucleotide pair and the total average percent of canonical base pairs were calculated.

Probabilities of loops in predicted and SHAPE-constrained structures of the 5′ UTR were calculated using ProbScan.[27] ProbScan estimates the probabilities of loops in secondary structure according to the frequency that a particular loop appears in an ensemble of structures, which corresponds to its approximate probability in the ensemble.[27] The APSI is a measure of similarity between aligned sequences.[6,28] The effects of mutations on structures at exon-intron junctions on structure were analyzed with RNA2DMut.[29] Using RNAfold, RNA2DMut calculates for a given input sequence, an MFE structure and its Gibbs folding energy ($\Delta G°_{37}$) in kcal/mol.[29] Additionally, it calculates a partition function, a centroid structure, defined as the structure that is most similar to other structures in the ensemble, and the ED for the input sequence.[29] Alternative algorithms such as CMFinder predict 2D structure from a set of unaligned sequences, but they are not suitable for identifying structures in single sequences, as reported in this work.[30]

To account for structural homology after motifs were identified, covariation among multiple MAFFT-aligned homologous motif sequences was measured using R-scape.[31] Covariation between two sites of an RNA molecule is observed when nucleotide changes occur at those sites in two or more species but base pairing is retained, and provides evidence for evolutionary conserved structure.[32,33] For each alignment column pair, an E-value is calculated to estimate the number of false positives relative to a covariation score.[31] A significant E-value is defined as 0.05 or lower.

## Cell-based assays

Mini-gene reporters of exon 10 splicing, whether wild type tau, tau with intron 9 `G(-10)T`, intron 10 `G(+3)A`, or intron 10 `C(+14)T` mutations[34], or fragments of a long 3′ UTR
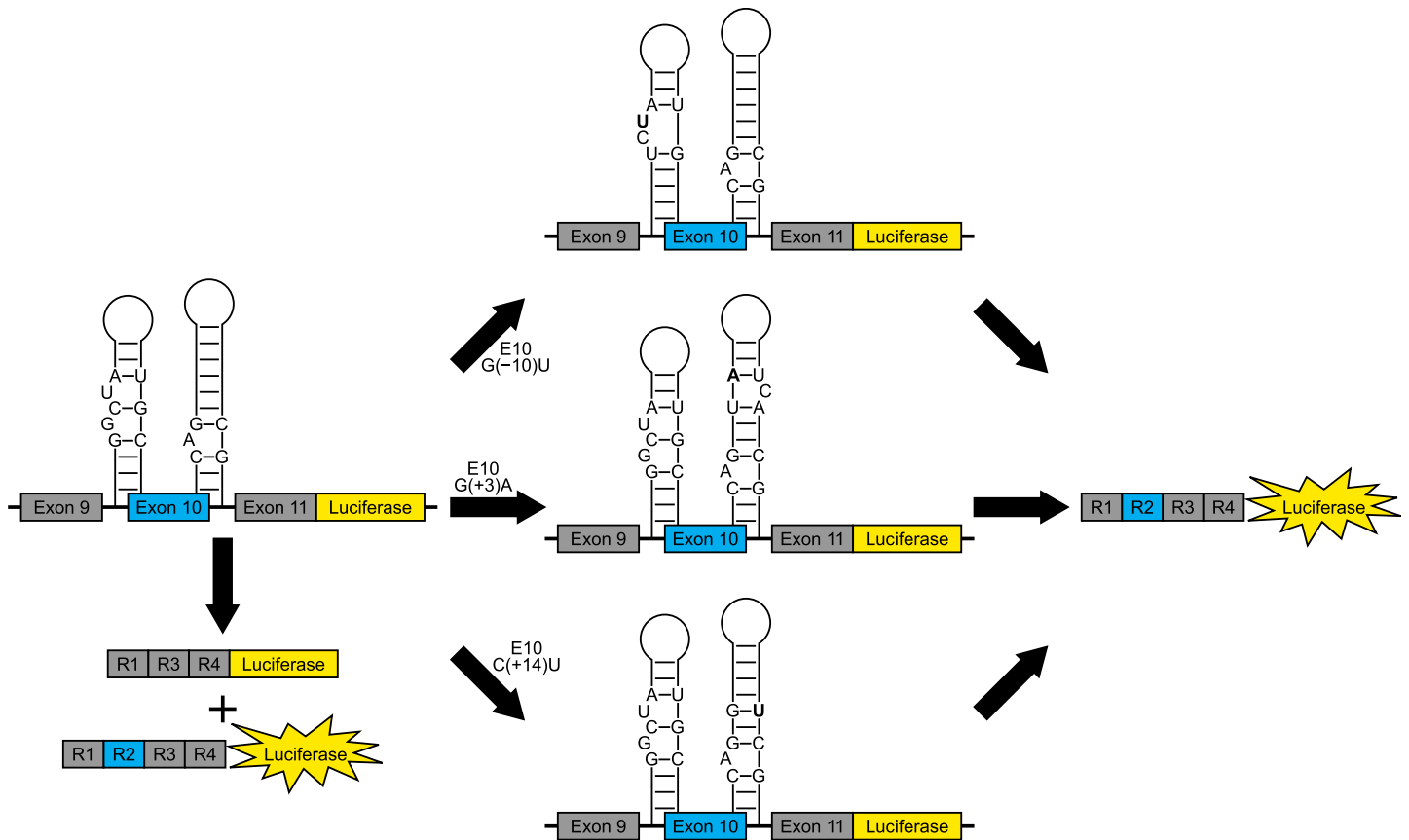
**Fig 1. Schematic representation of cell-based assays with mini-genes expressing wild type tau or tau with mutations in intron 9 or 10.** [34,36] Mini-genes contain exon 10 in frame with a firefly luciferase reporter gene. Wild type mini-genes express tau in a 4R-to-3R ratio of ~1. Mutations in this study (indicated in bold) that affect structures at splice sites adjacent to exon 10 lead to increased exon 10 inclusion and consequently, increased luciferase signal.

https://doi.org/10.1371/journal.pone.0219210.g001

isoform[15] were used in cell-based experiments. Mini-genes expressing wild type tau, tau with intron 10 C(+14)T mutations, and a long 3′ UTR isoform (4163 nt) were provided by Prof. Michael S. Wolfe (The University of Kansas).[34–36] Mini-genes expressing tau with intron 9 G(-10)T, intron 10 G(+3)A, and fragments of the long 3′ UTR isoform were generated using a QuikChange Site-Directed Mutagenesis Kit (Agilent) and the manufacturer's recommended protocol. Mini-genes containing mutations within intron 9 or 10 express a firefly luciferase reporter gene in-frame with exon 10, which leads to a decrease in signal when exon 10 is excluded (Fig 1).

HeLa cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) and 1% GlutaMax at 37 ˚C. For experiments with mini-genes containing intron 9-exon 10 or exon 10-intron 10 junctions, cells were transfected at ~90% confluency in a 100 mm dish using jetPRIME (PolyPlus-transfection SA) with 10 µg of mini-gene. A 60 mm dish was mock transfected using jetPRIME as a control. Cells were resuspended in growth medium, seeded in a 384-well plate at 10,000 cells per well, and incubated at 37 ˚C for 24 h. After, 5 µL of CellTiter Fluor reagent (Promega) was added to each well, the cells were incubated at 37 ˚C for 30 min, and fluorescence was measured (380 nm/505 nm excitation/ emission wavelegnths with a 495 nm cutoff filter) using a SpectraMax M5 plate reader (Molecular Devices). Next, 25 µL of ONE-Glo reagent (Promega) was added to each well, and cells

were incubated at room temperature for 3 min. Luminescence was measured on a SpectraMax M5 plate reader with a 500 ms integration time.

Mini-genes expressing fragments of a long 3′ UTR isoform were fused to firefly luciferase in the context of pmiRGLO vector, which also express Renilla luciferase used for normalization (see S1 File for cloning methods). Fragments can either increase or decrease firefly luciferase signal, corresponding to increased or decreased tau expression, respectively. HeLa cells were seeded at 250,000 cells per well in 6-well plates and 2 mL of growth medium. At ~60% confluency, cells were transfected with the mini-gene of interest, resuspended in growth medium, seeded in a 96-well plate at 50,000 cells per well density, and incubated at 37 ˚C. After 24 h, growth medium was removed from each well, and the cells were washed with 50 μL of 1× DPBS and lysed with 20 μL of lysis buffer. A 50 μL aliquot of firefly luciferin buffer was added to each well, and luminescence was measured as described above. Next, 50 μL of Renilla luciferase buffer was added to each well, the cells were incubated for 10 min at room temperature, and luminescence was measured as described above.

## Results and discussion

### Overview of computational analysis

The prediction of RNA secondary (2D) structure is an active area of research, and a number of algorithms attempt to find the native 2D conformation from sequence.[37] Despite varying approaches, almost all folding algorithms use the thermodynamic energy of folding in prediction. The free energy change (ΔG˚) going from the unfolded to the folded state of RNA 2D structure measures its energetic favorability. The ΔG˚ of RNA 2D structure can be calculated from the "Turner Rules," a compilation of free energy parameters for RNA 2D structures that is based on experimental measurements of small motifs.[38,39] In the most widely applied folding algorithms (e.g. mFold[40], RNAfold[41] and RNAstructure[42]) the Turner Rules are used to find the minimum free energy (MFE) 2D structure of a sequence in silico;[38,43] here the MFE (most stable) fold is assumed to best reflect the native structure.

The MFE ΔG˚ indicates the stability of a 2D structure, *but not its potential for function*. It was observed, however, that compared to random sequences, the MFE ΔG˚ of functional RNAs is lower.[44] This is attributed to the *evolved order and composition* of functional RNAs, which require structure to function; disrupting this ordered sequence also disrupts stabilizing base pairs leading to less favorable ΔG˚ values. This propensity for folding is quantified by the ΔG˚ z-score ($z - \text{score} = [\varDelta G°_{\text{native}} - \overline{\varDelta G°}_{\text{random}}]/\sigma$). That is, ΔG˚ z-score measures the stability of the native sequence vs. random sequence and suggests an evolved property of the RNA. [23,44] The z-score metric is indeed at the heart of functional 2D structure prediction approaches[45,46] and was used to identify conserved functional regions in the genomes of influenza viruses[6,47] and EBV,[21] as well as the mammalian Xist long noncoding (lnc) RNA.[48] These analyses led to the discovery of functional *cis*-regulatory RNA structures[49–51] and novel structured ncRNAs[21,52] that play important roles in both viruses, as well as multiple highly-conserved (across mammals) structures in Xist that are likely important to the mechanism of X chromosome inactivation.[48] This same approach was used to discover conserved regions of structure with *MAPT* pre-mRNA that may dictate biological function.

### Global structure of tau pre-mRNA

The tau pre-mRNA has 21 predicted structured exon-intron junctions, one predicted structured region in the 5′ UTR, and nine predicted structured regions in the 3′ UTR. Altogether, approximately 1000 bps at exon-intron junctions, 60 bps in the 5′ UTR, and 2,500 bps in the 3′

**Table 1. Exon-intron junctions containing predicted wild type structures.**

| Junction | Nucleotides | $\Delta G_{37}°$ (kcal/mol) | Average base pair conservation (%) | Number of species[a] |
|---|---|---|---|---|
| E1-I1 | 137–150 of E1 and E1+1 to E1+7 of I1 | −9.5 | 94.3 | 5 |
| I1-E2 | E2−29 to E2−1 of I1 and 1–3 of E2 | −9.0 | 97.4 | 6 |
| E2-I2 | 82–87 of E2 and E2+1 to E2+27 of I2 | −11.6 | 99.2 | 6 |
| E2-I2 | 68–87 of E2 and E2+1 to E2+46 of I2 | −19.2 | 99.0 | 6 |
| E3-I3 | 18–87 of E3 | −25.9 | 92.9 | 18 |
| I3-E4 | E4−43 to E4−1 of I3 and 1–15 of E4 | −14.0 | 100.0 | 2 |
| I3-E4 | E4−25 to E4−1 of I3 and 1–6 of E4 | −6.4 | 100.0 | 1 |
| I4-E4A | E4A−38 to E4A−1 of I4 and 1–31 of E4A | −24.1 | 97.5 | 3 |
| I4A-E5 | E5−63 to E5−1 of I4A and 1–5 of E5 | −15.4 | 100.0 | 5 |
| E5-I5 | 4–56 of E5 and E5+1 to E5+15 of I5 | −13.5 | 100.0 | 2 |
| E5-I5 | 50–56 of E5 and E5+1 to E5+53 of I5 | −16.5 | 100.0 | 2 |
| E6-I6 | 133–198 of E6 and E6+1 of I6 | −18.2 | 98.8 | 8 |
| E6-I6 | 164–198 of E6 and E6+1 to E6+29 of I6 | −11.7 | 100.0 | 1 |
| I6-E7 | E7−35 to E7−1 of I6 and 1–28 of E7 | −14.5 | 99.0 | 4 |
| I6-E7 | E7−1 of I6 and 1–48 of E7 | −17.0 | 96.3 | 9 |
| E7-I7 | E7+1 to E7+21 of I7 | −2.7 | 100.0 | 2 |
| I7-E8 | E8−19 to E8−1 of I7 and 1–27 of E8 | −13.0 | 100.0 | 1 |
| I7-E8 | E8−26 to E8−1 of I7 and 1–40 of E8 | −17.9 | 98.7 | 4 |
| I8-E9 | E9−41 to E9−1 of I8 and 1–21 of E9 | −11.0 | 100.0 | 4 |
| I8-E9 | E9−15 to E9−1 of I8 and 1–5 of E9 | −0.4 | 100.0 | 1 |
| I9-E10 | E10−14 to E10−1 of I9 and 1–7 of E10 | −4.5 | 94.1 | 7 |
| E10-I10 | 73–93 of E10 and E10+1 to E10+31 of I10 | −19.3 | 93.3 | 4 |
| I10-E11 | E11−3 to E11−1 of I10 and 1–30 of E11 | −6.4 | 94.4 | 5 |
| I10-E11 | E11−10 to E11−1 of I10 and 1–40 of E11 | −10.9 | 90.0 | 4 |
| E11-I11 | 78–82 of E11 and E11+1 to E11+60 of I11 | −17.7 | 92.1 | 5 |
| I11-E12 | E12−61 to E12−1 of I11 and 1–8 of E12 | −20.1 | 94.6 | 6 |
| I11-E12 | E12−11 to E12−1 of I11 and 1–51 of E12 | −16.0 | 95.4 | 5 |
| E12-I12 | 77–113 of E12 and E12+1 to E12+22 of I12 | −17.9 | 96.1 | 5 |
| E12-I12 | 100–113 of E12 and E12+1 to E12+53 of I12 | −11.9 | 94.9 | 5 |
| I12-E13 | E13−9 to E13−1 of I12 and 1–32 of E13 | −6.8 | 94.0 | 7 |

[a] Non-human primates.

UTR have structure. Five alternatively spliced exon-intron junctions have structures, which may regulate splicing. Base pair conservation at these junctions is generally high, but in relatively few non-human primates (typically less than 10). Within the 3′ UTR, eight of the nine structured regions have base pair conservation of ~90% or greater.

## Structures at exon-intron junctions

Structures were predicted at 19 exon-intron junctions, including five of the six that are adjacent to alternatively spliced exons (Table 1 and Figs 2 and 3 and Figure A in S1 File). Base pair conservation for all of these structures was above 90%. Thus, they are highly conserved. Regulation of splicing at these exon-intron junctions may be achieved by recognition of structures or splicing elements by splicing factors. Thus, mutations within sequences near the junctions that affect structure may disrupt normal splicing and cause disease.

Below, we focus on the stable predicted structures at exon-intron junctions that are adjacent to alternatively spliced exons (exons 2, 3, and 10). Each contains a U1 snRNA binding
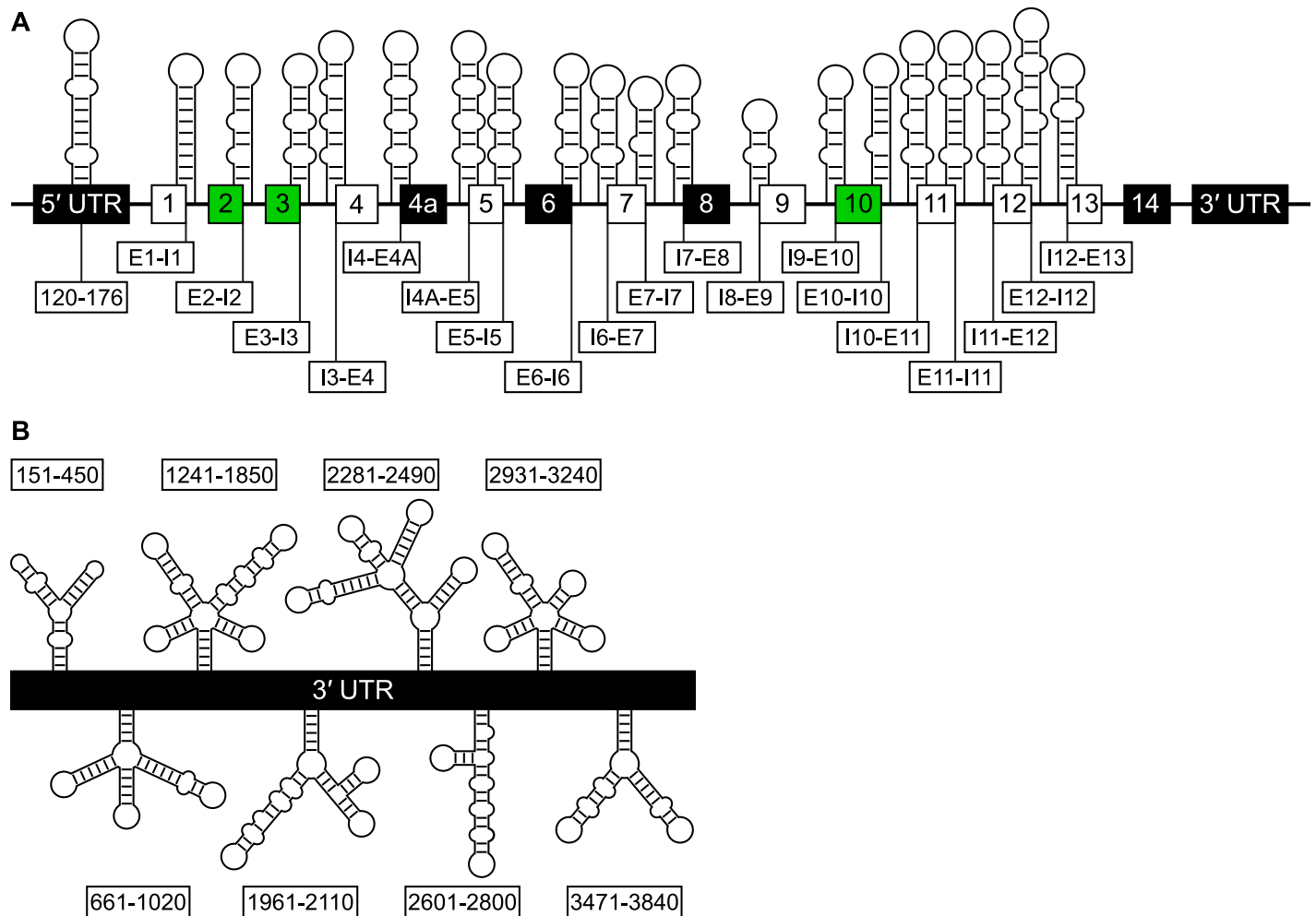
**Fig 2. Overview of structured regions in the tau pre-mRNA.** (a) Conserved structures in the 5′ UTR and exon-intron junctions. White boxes indicate constitutive exons. Black boxes indicate exons that are excluded in the human brain. Green boxes indicate exons that are alternatively spliced. (b) Conserved structures identified in the 3′ UTR.

site, where U1 5′ splice site recognition occurs by forming a duplex beginning at +1G of the 5′ splice site and C8 of U1 snRNA.[53] A maximum of 11 bps may form between the 5′ splice site and U1 snRNA, spanning the last 3 nt of the exon and first 8 nt of the intron.[53] RNA structure at the 5′ splice site may regulate splicing or compete with U1 snRNA for base pairing.[53]

The remainder of the structures and their associated metadata can be found in Supplementary Results in S1 File and Figures A and B in S1 File. There is interesting biology at these other intron-exon junctions. For example, a structured region was predicted at the exon 6-intron 6 junction. Deletion mutagenesis on exon 6 revealed that the 5′ and 3′ ends, the latter of which includes a 25–30 nt sequence within the predicted hairpin structure, decreased exon 6 inclusion, indicating that these regions act as strong splicing enhancers.[54] Exon 6 has two cryptic 3′ splice sites, which, when used, result in frameshift mutations and yield variants of tau protein lacking in microtubule binding domains.[54,55] These splice sites, termed 6p and 6d, are 101 and 170 nt long, respectively, are downstream of the canonical splice site. A stable structure was predicted at the 6p splice site (Fig 3A).[56] This same structure was predicted to
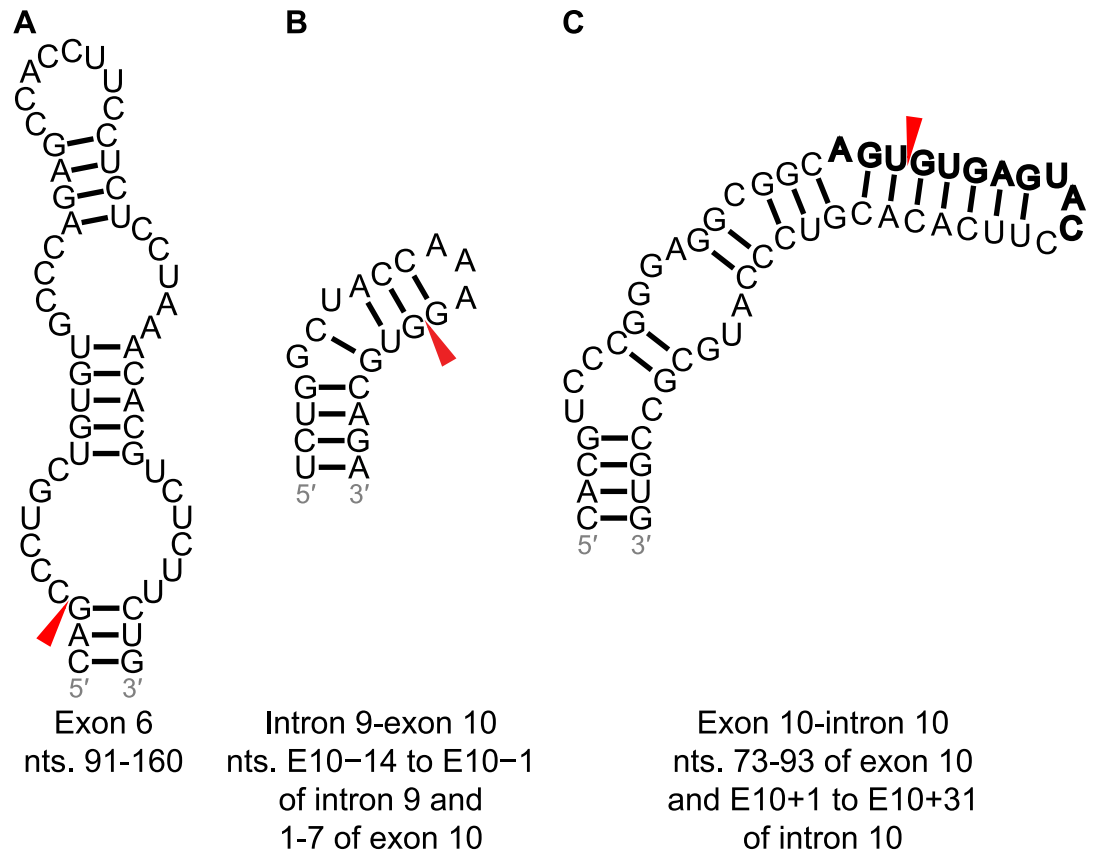
**Fig 3. Predicted structures at alternatively-spliced exon-intron junctions.** Splice sites are denoted with red arrowheads. Potential U1 snRNA binding sites are denoted by bold nucleotides. The enhancer in the hairpin structure at the exon 3-intron 3 junction is colored orange. Usage note: at a 3′ splice site, intronic nucleotides are denoted by a minus (−) sign and their position upstream of the adjacent exon, while exonic nucleotides are denoted by their position relative to the first nucleotide of the exon. At a 5′ splice site, intronic nucleotides are denoted by a (+) sign and their position downstream of the adjacent exon.

https://doi.org/10.1371/journal.pone.0219210.g003

form from the 250 nt sequence centered on the exon 6-intron 6 junction, but its z-score and ED were within 1 SD of average for all windows.

*Exon 10.* Exon 10 alternative splicing is the most well studied splicing event in tau pre-mRNA, owing to identified mutations that cause disease (Fig 4A to 4C and 4F to 4I). [13,57,58] Among 70 nt windows spanning the intron 9-exon 10 junction, the window with the lowest z-score (−1.05) contains a hairpin structure modeled using SHAPE mapping (Fig 3B).[59] A G-to-U mutation at position E10−10 near this junction was observed to increase the 4R-to-3R tau ratio (Fig 4D and 4E).[60] The previously validated hairpin and splice site regulator at the exon 10-intron 10 junction was predicted as part of a larger hairpin spanning nt 73–93 of exon 10 and E10+1 –E10+31 of intron 10, which contains a U1 snRNA binding site (Fig 3C). The 70 nt window containing this structure has a z-score and ED of −1.15 and 3.39, respectively, which were more than 1 SD below their respective averages.

This hairpin structure was also supported by SHAPE mapping.[59] Splicing of exon 10 is regulated by *cis*-elements, such as exonic splicing enhancers, silencers (ESEs and ESSs, respectively) and structured RNAs, and *trans*-factors, such as serine-arginine (SR) proteins that bind to ESEs adjacent to the 5′ and 3′ splice sites and promote association with U1 and U2 snRNP, respectively.[55,59,61,62] These splice sites are weak, and their activities (and exon 10 inclusion) are dependent on the ESEs.[62] Consistent with ESE-dependent splice sites is the
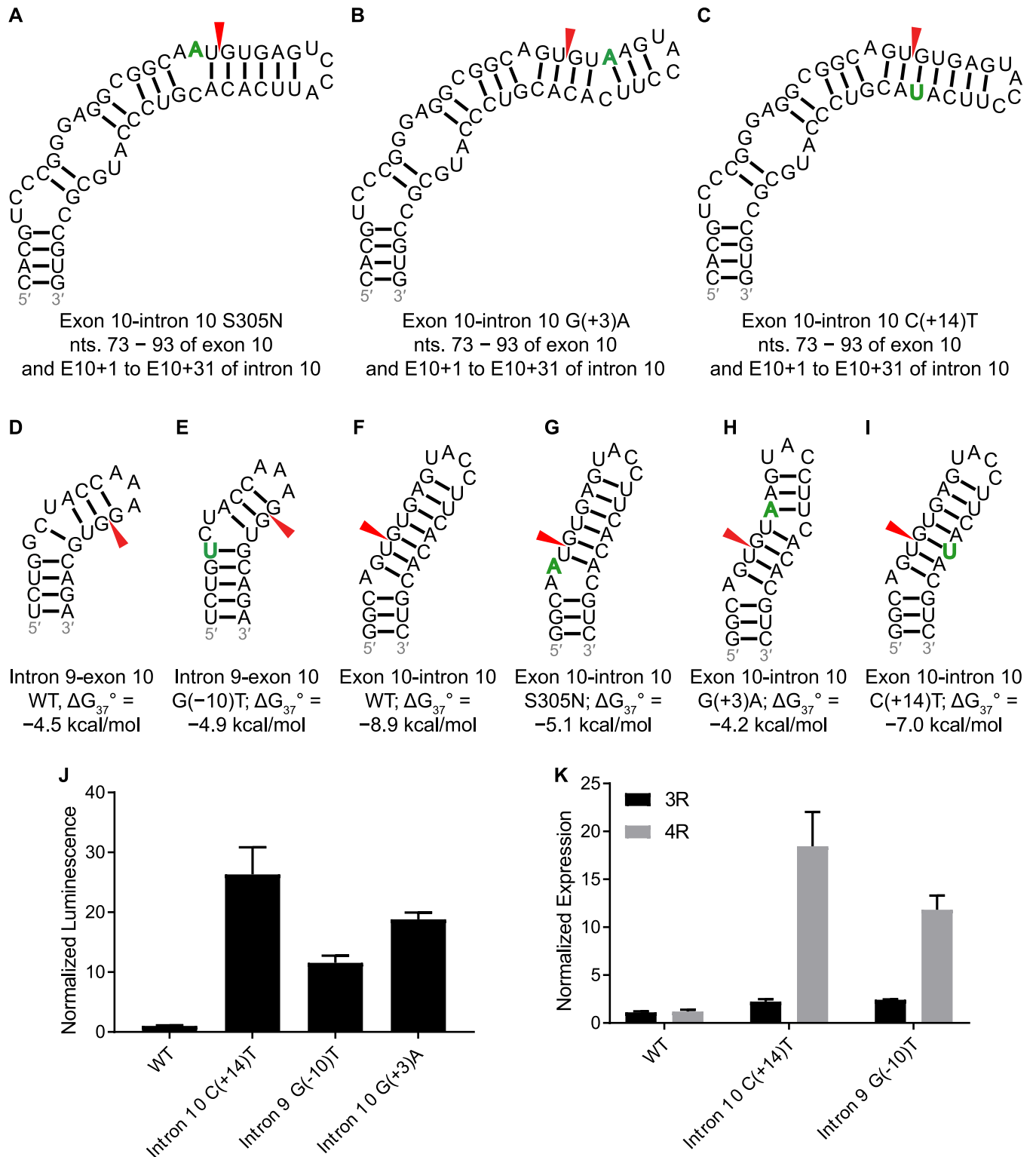
**Fig 4. Predicted structures at exon-intron junctions that contain mutations and experimental results with minigenes at exon-intron junctions.** Shown are structures containing mutations that affect the wild type structure. Mutations are denoted in green. Splice sites are denoted with red arrowheads. (a) to (c) Extended structures of hairpins at the exon 10-intron 10 junction containing S305N, G(+3)A, and C(+14)T mutations, respectively. (d) and (e) Wild type and G(−10)T mutant hairpin structures, respectively, at the intron 9-exon 10 junction. (f) to (i) Minimal wild type and mutant S305N, G(+3)A, and C(+14)T hairpin structures, respectively, at the exon 10-intron 10 junction. (j) Effect of mutations at the 5′ and 3′ splice sites of exon 10 on luciferase activity in transfected HeLa cells. (k) Effect of mutations at the 5′ and 3′ splice sites of exon 10 on 3R and 4R tau mRNA levels in HeLa cells as measured by RT-qPCR.

**Table 2. Exon-intron junctions containing predicted mutant structures.**

| Junction | Nucleotides | Mutation | $\Delta G_{37}°$ (kcal/mol) |
|---|---|---|---|
| I9-E10 | E10−14 to E10−1 of I9 and 1−7 of E10 | G(−10)T | −4.9 |
| E10-I10 | 73−93 of E10 and E10+1 to E10+31 of I10 | S305N | −16.6 |
| E10-I10 | 73−93 of E10 and E10+1 to E10+31 of I10 | G(+3)A | −14.6 |
| E10-I10 | 73−93 of E10 and E10+1 to E10+31 of I10 | C(+14)T | −17.4 |
| E11-I11 | 78−82 of E11 and E11+1 to E11+60 of I11 | P332S | −14.0 |
| I11-E12 | E12−61 to E12−1 of I11 and 1−8 of E12 | G335S | −20.5 |
| I11-E12 | E12−61 to E12−1 of I11 and 1−8 of E12 | G335V | −17.4 |
| I11-E12 | E12−11 to E12−1 of I11 and 1−51 of E12 | G335S | −16.0 |
| I11-E12 | E12−11 to E12−1 of I11 and 1−51 of E12 | G335V | −14.8 |
| I11-E12 | E12−11 to E12−1 of I11 and 1−51 of E12 | Q336H | −13.7 |
| I11-E12 | E12−11 to E12−1 of I11 and 1−51 of E12 | Q336R | −15.3 |
| I11-E12 | E12−11 to E12−1 of I11 and 1−51 of E12 | V337M | −16.6 |
| E12-I12 | 77−113 of E12 and E12+1 to E12+22 of I12 | V363I | −17.9 |
| E12-I12 | 77−113 of E12 and E12+1 to E12+22 of I12 | P364S | −14.9 |
| E12-I12 | 77−113 of E12 and E12+1 to E12+22 of I12 | G366R | −16.9 |
| E12-I12 | 77−113 of E12 and E12+1 to E12+22 of I12 | K369I | −17.9 |
| E12-I12 | 100−113 of E12 and E12+1 to E12+53 of I12 | K369I | −11.9 |

https://doi.org/10.1371/journal.pone.0219210.t002

presence of a weak polypyrimidine tract at the 3′ splice site.[62] Deletion of nt 7–15 of exon 10 inhibited binding and inhibition of splicing by SRp30c and SRp55.[63] This region is upstream of an enhancer, defined by a 5′-AAG motif spanning nt 16–18 of exon 10 that binds splicing activator htra2β1.[63–65] Deletion of this region, also known as the Δ280K mutation, decreased exon 10 inclusion and formation of 4R tau.[63–65] Co-transfection of HEK293 cells with SRp54 (signal recognition particle 54) and a mutant mini-gene without nt 16–21 of exon 10, which consists of a purine-rich 5′-AAGAAG enhancer element, eliminated the binding of SRp54.[66] These results indicate that htra2β1 and SRp54, which antagonizes its activity, bind to these regions.[63–66]

On the other hand, the N279K mutation, which changes a U to a G, increases exon 10 inclusion by strengthening an AG-rich region that acts as a splicing enhancer.[63–65] Similar to exon 2, the default splicing pattern is inclusion and regulation primarily occurs through inhibition.[55,63] The S305N mutation, caused by a NG_007398.1:c.914G>A substitution (cDNA numbering starting from A in the AUG initiation codon in the noted GenBank cDNA reference sequence),[67] converts a GC pair to a noncanonical AC pair (Table 2).[68] The result is destabilization of the exon 10-intron 10 SRE and increased exon 10 inclusion, similar to other intronic mutations at this junction (Table 2).[68] RNA2DMut predicted consistent MFE and centroid structures for this mutation, which increased ED from 3.39 to 11.16 and destabilized the structure by 3.8 kcal/mol. A G(+3)A mutation at the exon 10-intron 10 junction decreases the stability of the stem-loop by 4.7 kcal/mol by disrupting a GC pair and stabilizes formation of the U1 snRNP complex by changing a GU pair to an AU pair.[69] For this mutation, RNA2DMut predicted an increase in ED of 8.02 relative to WT for the centroid structure. A C(+14)T mutation converts a GC pair in the stem-loop to a GU pair, destabilizing the helix by 1.9 kcal/mol. RNA2DMut predicted a small 0.24 increase in ED.

## Experimentally-determined structures at exon 10-intron junctions

The splicing regulator at the exon 10-intron 10 junction was predicted to form, with 93.3% base pair conservation (Figs 3 and 4). The C(+14)U mutation causes the disinhibition-

dementia-parkinsonism-amyotrophy (DDPAC) disease, one type of FTDP-17.[13,57,70] A 4R-to-3R ratio of 30:1 relative to wild type was measured by RT-PCR in HeLa cells transfected with mini-genes expressing the stem-loops.[34] In transfected luciferase assays, the 4R-to-3R ratio of DDPAC tau to wild type tau was similar.

A `C`-bulge was predicted to form in the structure containing the `G(+3)A` mutation. This mutation results in the multiple system tauopathy with presenile dementia (MSTD) form of FTDP-17.[58,69] A 4R-to-3R ratio of ~20:1 relative to wild type was measured in transfected luciferase assays with this mutation. No occurrence of this mutation was found in non-human primates.

A hairpin at the intron 9-exon 10 junction was predicted to form by SHAPE-constrained RNA folding algorithms (Fig 4D).[59] A `G-to-U` mutation at position E10−10 of this splice junction (Fig 4E) weakens a polypyridine tract and a 2 nt bulge loop was predicted to form as a result of the `G-to-U` mutation, which stabilizes the wild type hairpin by 0.4 kcal/mol. This mutation leads to increased selection of the splice site and exon 10 inclusion.[60,62] In cell-based assays completed herein, the `G-to-U` mutation led to an 11.5-fold increase in luciferase expression over wild type (Fig 4J).

## Structure of the 5′ UTR

The tau 5′ UTR contains a conserved oligopyrimidine sequence, `5′-CCTCCCC-3′` that promotes cap-dependent translation by activation of the mTOR pathway.[71,72] Separately, eukaryotic internal ribosome entry sites (IRESs) allow initiation of cap-independent translation of mRNA.[73,74] IRES activity may be mediated by RNA structure, short sequence motifs in unstructured or structured regions that bind to IRES trans-acting factors (ITAFs), or base pairing with 18S ribosomal RNA (rRNA).[73,74] Additionally, sequence motifs within loop regions of IRESs may allow for RNA-RNA interactions and formation of secondary or tertiary structure essential for IRES function.[73,74] The tau 5′ UTR contains a cis-regulatory internal ribosomal entry site (IRES) that contributed to 30% of the total translation of tau mRNA in luciferase assays.[71–73] Truncations to the IRES reduced or abolished IRES activity, suggesting that the entire sequence is required for IRES activity.[73] A secondary structure model of the 240 nt IRES generated by SHAPE analysis yielded two domains. Among them, Domain I is a hairpin and Domain II is a multibranch loop.[73] Mutations that disrupted the stem regions of Domains I and II abolished IRES activity.[73] Therefore, sequences and/or structures in these regions are required for IRES function.[73]

Our analysis predicts a structured region corresponding to nt 120–176 of the 304 nt 5′ UTR, which lies within Domain I of the IRES and has a favorable z-score and low ED (−1.03 and 5.82, respectively) (Fig 5). The structure is similar to Domain I, although the base pairs around the internal loop below the apical loop were shifted, resulting in a 3 × 1 instead of 3 × 4 internal loop. This may result from the lone cytidine on the 3′ side of the loop having SHAPE reactivity in experiments with transfected SK-N-SH cells, whereas other neighboring nucleotides were unreactive.[73] The free energies of 42 nt segments of the unconstrained and SHAPE-constrained Domain I hairpin structures (−21.0 and −20.6 kcal/mol, respectively) are relatively close. Therefore, these conformations may exist in equilibrium. The probabilities of formation of loops in the predicted and SHAPE-constrained 5′ UTR was analyzed with ProbScan as a measure of the accuracy of structure prediction compared to experiment.[27] The probabilities of the 3 × 4 internal loop and 9 nt hairpin loop calculated by ProbScan for the predicted structure are 0.584 and 0.601, respectively. Folding of the same fragment of sequence of the predicted structure into the SHAPE-constrained 3 × 1 internal loop and 10 nt hairpin loop resulted in loop probabilities of 0.301 and 0.323, respectively. These results suggest that
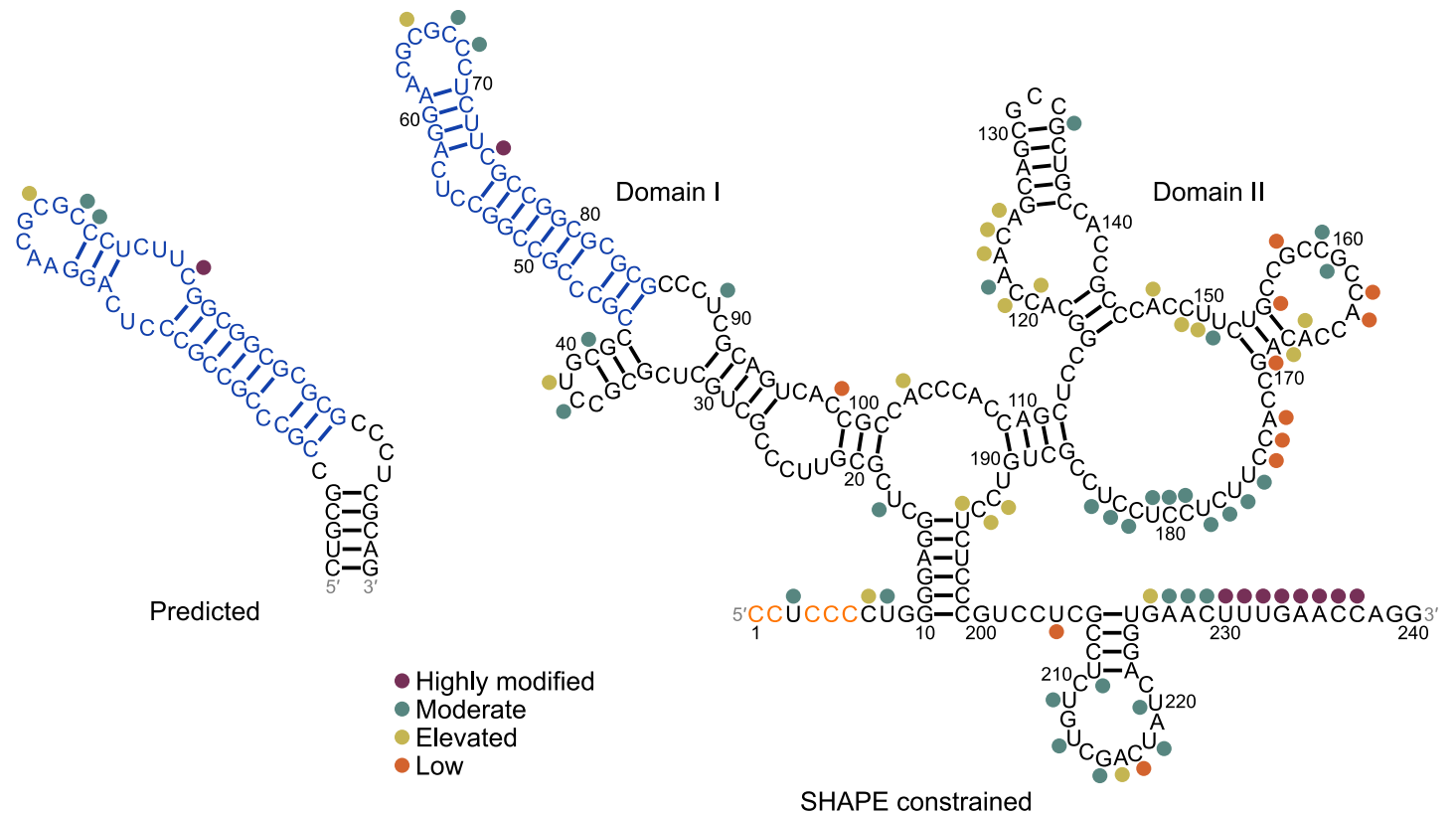
**Fig 5. Predicted and SHAPE constrained structures of the human 5′ UTR.** Predicted structure corresponds to nt 120–176. Structure determined by SHAPE analysis corresponds to the 240 nt tau IRES.[73] The conserved 5′-CCUCCCC-3′ sequence that promotes cap-dependent translation is colored orange. Blue nucleotides in the SHAPE constrained structure correspond to those in the predicted structure. Nucleotides in the SHAPE constrained structure that were modified in SHAPE analysis are indicated by circles colored according to SHAPE reactivity (see figure key). For reference, nucleotides in the predicted structure that correspond to those in the SHAPE-constrained structure are labeled with the respective colored circles. SHAPE reactivities were obtained from ref. [73].

https://doi.org/10.1371/journal.pone.0219210.g005

the predicted 3 × 4 loop and associated hairpin loop may be more likely to form than those in the SHAPE-constrained structure and are consistent with the slightly more favorable folding free energy of the conformation of the predicted structure relative to the SHAPE-constrained structure. When the full 5′ UTR was queried, a region within the human structure corresponding to base pairs C133:G160 to G142:C150 and containing a 3 × 4 internal loop was predicted from alignment and folding of the resulting sequences. Base pair conservation for this consensus structure is 94.6% in up to 20 non-human primates. R-scape predicted up to 39 potentially covarying base pairs in this region (Table D in S1 File). Altogether, a hairpin structure consistent with a SHAPE-constrained IRES structure was predicted to form in the 5′ UTR and contains structural and sequence elements required for IRES activity.

## Structure of the 3′ UTR

The 3′ UTR of mRNAs regulate various cellular events such as translation, and transcript stability, polyadenylation, and localization.[75] One mechanism by which regulation of gene expression occurs is by alternative polyadenylation (APA).[22,28,76,77] Polyadenylation at alternative sites in the 3′ UTR results in transcripts with 3′ UTRs of different lengths.[76] Longer 3′ UTRs may contain more miRNA binding sites or regulatory sequences that cause the mRNA to be more prone to negative regulation.[76] The human tau 3′ UTR contains two polyadenylation signals, which results in 3′ UTR isoforms 256 nt and 4163 nt in length.[15]

Dickson *et al.* found that the short isoform was expressed significantly higher than the long isoform.[15] The differential regulation of the 3′ UTR isoforms may be mediated by *cis*-elements and *trans*-factors.[15] Dickson *et al.* identified a binding site for miR-34a, which inhibits expression of endogenous tau, and reported that a luciferase reporter vector containing a mutant miR-34a binding site was insensitive to pre-miR-34a.[15] Additionally, cloning of fragments of the long 3′ UTR isoform into a luciferase reporter vector increased or decreased luciferase activity in transfected M17D and SH-SY5Y cells, suggesting that these fragments may differentially affect tau expression.[15,78] Secondary structures within the 3′ UTR are vital determinants of gene levels, and mutations affecting secondary structure or a polyadenylation signal can cause disease.[75]

Indeed, the tau 3′ UTR has been implicated in stabilization of tau mRNA in neuronal cells. [79] The ratio of 3′ UTR to coding DNA sequence is balanced in normal brain tissue but elevated in the AD brain, suggesting that the 3′ UTR may have regulatory roles in the disease.[78] The 3′ UTR contains several microRNA binding sites such as miR-34a, miR-132, miR-181c, miR-219, miR-485-5p, and miR-642 (Table B in S1 File).[15,78,80,81] In particular, expression of miR-219-5p and miR-485-5p are decreased in AD and may be attributed to overexpression of tau.[78,80–82] Thus, targeting structured regions containing microRNA binding sites may be an approach for treating tauopathies.

Eight regions within the long 3′ UTR isoform were predicted to have z-scores less than 1 SD below the average of all windows (Table 3). Six of the regions had base pair conservation above 90%. Positions 151–300 have an average canonical base pair conservation of 97.6%. Within this region is a 5′–AAUAAApolyadenylation signal, which lies within a 5 × 5 internal loop and spans nt 230–235 of the 3′ UTR.[35] Cleavage and polyadenylation at this site generates the short 3′ UTR isoform.[35]

The structured region predicted to form between nt 1961–2110 in the 3′ UTR (average canonical base pair conservation of 89.2%) is similar to the multibranch loop formed in the tau coding region (nt 1989–2015). Positions 2281–2490 and 2931–3240 have an average canonical base pair conservation of 95.2% and 89.2%, respectively. R-scape predicted that two hairpin structures in the nt 2931–3240 region contain base pairs with evidence of covariation (Table G in S1 File). These results suggest that these regions are highly likely to form functional structure.

Other regions in the 3′ UTR are also highly conserved, which span (or are adjacent to) microRNA binding sites, nt 661–1020 (94.0%) and nt 1241–1850 (93.8%). Interestingly, deletion of nt 986–1479, which includes binding sites for miR-34a and miR-485-5p and partly overlaps with this region, from a luciferase reporter construct significantly increased luciferase activity in transfected HEK293 and differentiated SH-SY5Y cells compared to cells expressing

**Table 3. Regions of the 3′ UTR with predicted structure.** Secondary structures for these regions are in Fig 6 and Figures C–H in S1 File.

| Nucleotides | $\Delta G_{37}°$ (kcal/mol) | Average base pair conservation (%) |
|---|---|---|
| 151–450 | −91.1 | 97.7 |
| 661–1020 | −147.1 | 94.0 |
| 1241–1850 | −219.6 | 93.8 |
| 1961–2110 | −50.2 | 89.2 |
| 2281–2490 | −81.7 | 95.2 |
| 2601–2800 | −66.0 | 43.7 |
| 2931–3240 | −98.3 | 89.2 |
| 3471–3840 | −149.6 | 94.9 |

https://doi.org/10.1371/journal.pone.0219210.t003
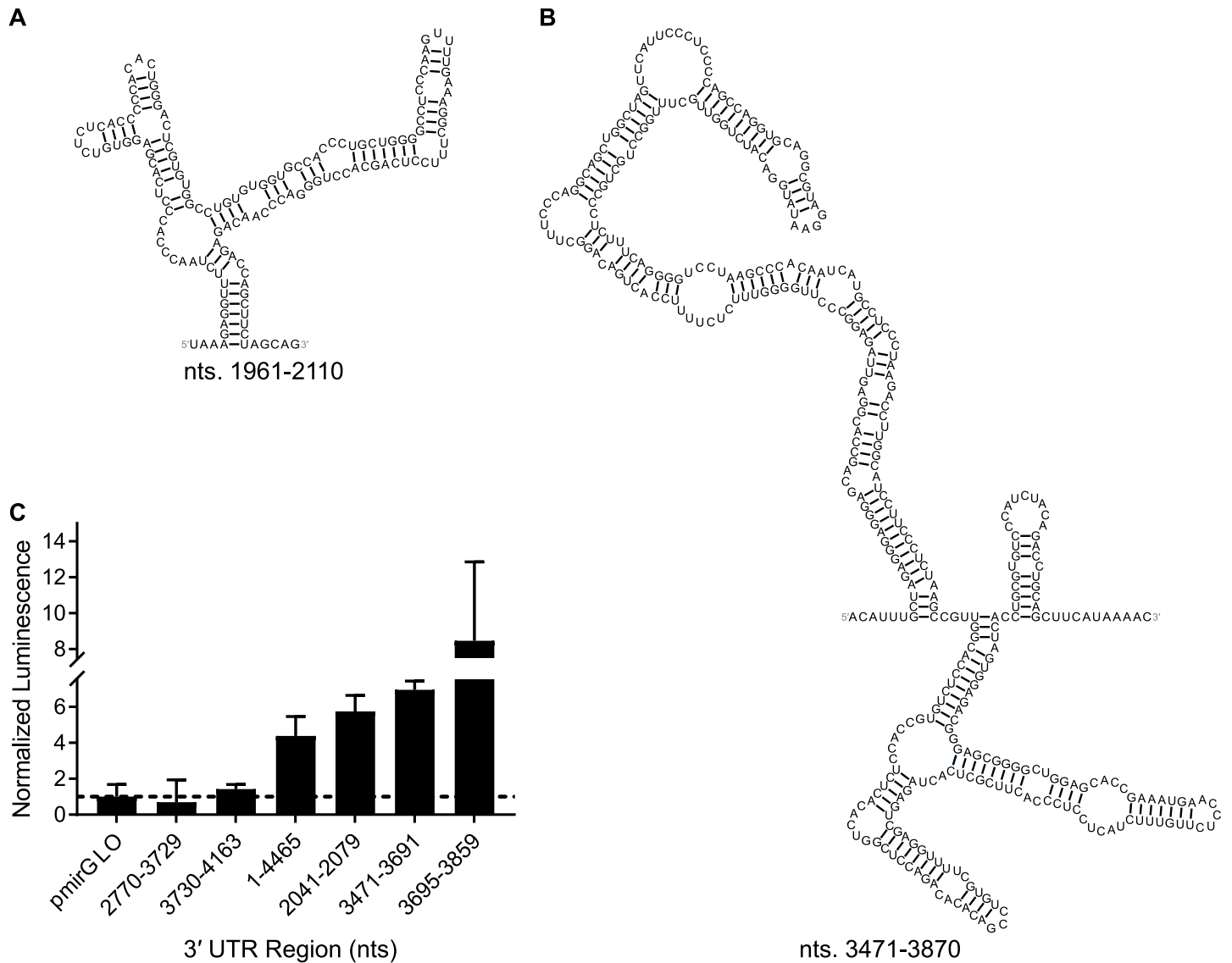
A



nts. 1961-2110

B



nts. 3471-3870

C



**Fig 6. Structured regions of the 3′ UTR that upregulated luciferase activity in Dickson *et al*[15] and results of cell-nased luciferase assays completed herein with fragments of the 3′ UTR.**

a wild type 3′ UTR construct.[35] This suggests that this region may contain a *cis*-element that downregulates expression of tau.[35] Indeed, a binding site for miR-485-5p is located upstream of the structured region comprised by nt 1241–1850, in addition to the RNA structure itself.[80] 43 potentially conserved base pairs in this region were predicted by R-scape in a consensus structure containing a multibranch loop that resembles one predicted in this region of the human sequence (Table E in S1 File). Together, these may control expression of tau.

Two structured regions, corresponding to nt 1961–2110 and 3471–3870, lie in segments that upregulated luciferase activity in transfected cells reported by Dickson et al (Fig 6A and 6B).[15] In cell-based assays carried out in this work, three fragments of the 3′ UTR were found to upregulate luciferase activity (Fig 6C). These three fragments, corresponding to nt 2041–2079, 3471–3691, and 3695–3859, contain regions that were predicted to be structured. In other studies, deletion of nt 3501–4000 increased luciferase activity in cells relative to those

expressing WT tau.[35] Therefore, the predicted structures may contain binding sites for *cis*-elements or trans factors that regulate tau expression.

## Conclusions

In summary, stable structures were predicted throughout the *MAPT* pre-mRNA. Structures were predicted at 19 exon-intron junctions, including five adjacent to alternatively spliced exons (Fig 1). Two folds were predicted at the exon 2-intron 2 junction, the larger of which is an extended form of a hairpin where the stem contains two single nucleotide bulges. The predicted structures at the 5′ and 3′ splice sites of exon 10 were previously identified by experiment. The `G(+3)A` and `C(+14)T` mutations that are associated with disease were verified by cell-based assays to upregulate exon 10 inclusion. These mutations destabilize the hairpin structure at the exon 10-intron 10 junction. A structure was predicted to form in exon 6 at a cryptic splice site, which results in frameshift. The IRES within the 5′ UTR contains a predicted structure, mutations within which had been demonstrated to decrease IRES activity. Eight structured regions were identified throughout the 3′ UTR, including regions that regulate expression of tau. RNA secondary structure plays important roles in regulating gene expression, and splice site junctions and UTRs are known to contain gene regulatory element. These regions of pre-mRNA may interact with *cis*-elements or *trans*-factors via sequence- or structure-specific binding. The results of this study may be applied to future studies to identify new therapeutic targets against tau-associated diseases.[36,83–86]

## Supporting information

**S1 File.** Table A. Primers Used to Clone Fragments of Tau 3′ UTR into a pmirGLO Vector. Table B. MicroRNAs Predicted to Bind to the Tau 3′ UTR. Table C. R-scape Calculations for nt 137–150 of Exon 1 and E1+1 to E1+7 of Intron 1 at the Exon 1-Intron 1 Junction. Table D. R-scape Calculations for the 5′ UTR. Table E. R-scape Calculations for Positions 1241–1850 of the 3′ UTR. Table F. R-scape Calculations for Positions 2601–2800 of the 3′ UTR. Table G. R-scape Calculations for Positions 2931–3240 of the 3′ UTR. Table H. R-scape Calculations for Positions 3471–3870 of the 3′ UTR. Figure A. Predicted Structures at Other Exon-Intron Junctions. Figure B. Predicted Structures at Exon-Intron Junctions that Contain Mutations. Figure C. Positions 151–450 of the 3′ UTR. Figure D. Positions 661–1020 of the 3′ UTR. Figure E. Positions 1241–1850 of the 3′ UTR. Figure F. Positions 2281–2490 of the 3′ UTR. Figure G. Positions 2601–2800 of the 3′ UTR. Figure H. Positions 2931–3240 of the 3′ UTR. (DOCX)

## Author Contributions

**Conceptualization:** Jonathan L. Chen, Walter N. Moss, Jessica L. Childs-Disney, Matthew D. Disney.

**Data curation:** Jonathan L. Chen, Walter N. Moss, Adam Spencer, Matthew D. Disney.

**Formal analysis:** Jonathan L. Chen, Walter N. Moss, Adam Spencer, Jessica L. Childs-Disney, Matthew D. Disney.

**Funding acquisition:** Walter N. Moss, Jessica L. Childs-Disney, Matthew D. Disney.

**Investigation:** Jonathan L. Chen, Walter N. Moss, Adam Spencer, Peiyuan Zhang, Matthew D. Disney.

**Methodology:** Jonathan L. Chen, Walter N. Moss, Adam Spencer, Matthew D. Disney.

**Project administration:** Jessica L. Childs-Disney, Matthew D. Disney.

**Resources:** Jonathan L. Chen, Walter N. Moss, Jessica L. Childs-Disney, Matthew D. Disney.

**Software:** Walter N. Moss.

**Supervision:** Walter N. Moss, Jessica L. Childs-Disney, Matthew D. Disney.

**Validation:** Jonathan L. Chen, Walter N. Moss, Adam Spencer.

**Visualization:** Jonathan L. Chen, Walter N. Moss.

**Writing – original draft:** Jonathan L. Chen, Walter N. Moss, Jessica L. Childs-Disney, Matthew D. Disney.

**Writing – review & editing:** Jonathan L. Chen, Walter N. Moss, Adam Spencer, Peiyuan Zhang, Jessica L. Childs-Disney, Matthew D. Disney.

# References

1. Strobel EJ, Watters KE, Loughrey D, Lucks JB (2016) RNA systems biology: Uniting functional discoveries and structural tools to understand global roles of RNAs. Curr Opin Biotechnol 39: 182–191. https://doi.org/10.1016/j.copbio.2016.03.019 PMID: 27132125

2. Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136: 777–793. https://doi.org/10.1016/j.cell.2009.02.011 PMID: 19239895

3. Schoch KM, DeVos SL, Miller RL, Chun SJ, Norrbom M, Wozniak DF, et al. (2016) Increased 4R-tau induces pathological changes in a human-tau mouse model. Neuron 90: 941–947. https://doi.org/10.1016/j.neuron.2016.04.042 PMID: 27210553

4. Freyhult EK, Bollback JP, Gardner PP (2007) Exploring genomic dark matter: A critical assessment of the performance of homology search methods on noncoding RNA. Genome Res 17: 117–125. https://doi.org/10.1101/gr.5890907 PMID: 17151342

5. Seemann SE, Mirza AH, Hansen C, Bang-Berthelsen CH, Garde C, Christensen-Dalsgaard M, et al. (2017) The identification and functional annotation of RNA structures conserved in vertebrates. Genome Res 27: 1371–1383. https://doi.org/10.1101/gr.208652.116 PMID: 28487280

6. Moss WN, Priore SF, Turner DH (2011) Identification of potential conserved RNA secondary structure throughout influenza A coding regions. RNA 17: 991–1011. https://doi.org/10.1261/rna.2619511 PMID: 21536710

7. Orr ME, Sullivan AC, Frost B (2017) A brief overview of tauopathy: Causes, consequences, and therapeutic strategies. Trends Pharmacol Sci 38: 637–648. https://doi.org/10.1016/j.tips.2017.03.011 PMID: 28455089

8. Bourdenx M, Koulakiotis NS, Sanoudou D, Bezard E, Dehay B, Tsarbopoulos A (2017) Protein aggregation and neurodegeneration in prototypical neurodegenerative diseases: Examples of amyloidopathies, tauopathies and synucleinopathies. Prog Neurobiol 155: 171–193. https://doi.org/10.1016/j.pneurobio.2015.07.003 PMID: 26209472

9. Morris M, Maeda S, Vossel K, Mucke L (2011) The many faces of tau. Neuron 70: 410–426. https://doi.org/10.1016/j.neuron.2011.04.009 PMID: 21555069

10. Brunden KR, Trojanowski JQ, Lee VMY (2009) Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies. Nat Rev Drug Discov 8: 783–793. https://doi.org/10.1038/nrd2959 PMID: 19794442

11. van Swieten J, Spillantini MG (2007) Hereditary frontotemporal dementia caused by Tau gene mutations. Brain Pathol 17: 63–73. https://doi.org/10.1111/j.1750-3639.2007.00052.x PMID: 17493040

12. Caillet-Boudin ML, Buee L, Sergeant N, Lefebvre B (2015) Regulation of human *MAPT* gene expression. Mol Neurodegener 10: 28. https://doi.org/10.1186/s13024-015-0025-8 PMID: 26170022

13. Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, et al. (1998) Association of missense and 5′-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393: 702–705. https://doi.org/10.1038/31508 PMID: 9641683

14. McCarthy A, Lonergan R, Olszewska DA, O'Dowd S, Cummins G, Magennis B, et al. (2015) Closing the tau loop: the missing tau mutation. Brain 138: 3100–3109. https://doi.org/10.1093/brain/awv234 PMID: 26297556

15. Dickson JR, Kruse C, Montagna DR, Finsen B, Wolfe MS (2013) Alternative polyadenylation and miR-34 family members regulate tau expression. J Neurochem 127: 739–749. https://doi.org/10.1111/jnc.12437 PMID: 24032460

16. Andreadis A, Brown WM, Kosik KS (1992) Structure and novel exons of the human tau gene. Biochemistry 31: 10626–10633. https://doi.org/10.1021/bi00158a027 PMID: 1420178

17. Li K, Arikan MC, Andreadis A (2003) Modulation of the membrane-binding domain of tau protein: splicing regulation of exon 2. Brain Res Mol Brain Res 116: 94–105. PMID: 12941465

18. Wang J, Gao QS, Wang Y, Lafyatis R, Stamm S, Andreadis A (2004) 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. https://doi.org/10.1046/j.1471-4159.2003.02232.x PMID: 15009664

19. Angelbello AJ, Chen JL, Childs-Disney JL, Zhang P, Wang ZF, Disney MD (2018) Using genome sequence to enable the design of medicines and chemical probes. Chem Rev 118: 1599–1663. https://doi.org/10.1021/acs.chemrev.7b00504 PMID: 29322778

20. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. (2011) ViennaRNA Package 2.0. Algorithm Mol Biol 6: 26–26.

21. Moss WN, Steitz JA (2013) Genome-wide analyses of Epstein-Barr virus reveal conserved RNA structures and a novel stable intronic sequence RNA. BMC Genomics 14: 543. https://doi.org/10.1186/1471-2164-14-543 PMID: 23937650

22. Dela-Moss LI, Moss WN, Turner DH (2014) Identification of conserved RNA secondary structures at influenza B and C splice sites reveals similarities and differences between influenza A, B, and C. BMC Res Notes 7: 22. https://doi.org/10.1186/1756-0500-7-22 PMID: 24405943

23. Freyhult E, Gardner PP, Moulton V (2005) A comparison of RNA folding measures. BMC Bioinformatics 6: 241. https://doi.org/10.1186/1471-2105-6-241 PMID: 16202126

24. Moss WN (2018) The ensemble diversity of non-coding RNA structure is lower than random sequence. Noncoding RNA Res 3: 100–107. https://doi.org/10.1016/j.ncrna.2018.04.005 PMID: 30175283

25. Andrews RJ, Baber L, Moss WN (2017) RNAStructuromeDB: A genome-wide database for RNA structural inference. Sci Rep 7: 17269. https://doi.org/10.1038/s41598-017-17510-y PMID: 29222504

26. Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059–3066. https://doi.org/10.1093/nar/gkf436 PMID: 12136088

27. Sloma MF, Mathews DH (2016) Exact calculation of loop formation probability identifies folding motifs in RNA secondary structures. RNA 22: 1808–1818. https://doi.org/10.1261/rna.053694.115 PMID: 27852924

28. Juan V, Wilson C (1999) RNA secondary structure prediction based on free energy and phylogenetic analysis. J Mol Biol 289: 935–947. https://doi.org/10.1006/jmbi.1999.2801 PMID: 10369773

29. Moss WN (2018) RNA2DMut: A web tool for the design and analysis of RNA structure mutations. RNA 24: 273–286. https://doi.org/10.1261/rna.063933.117 PMID: 29183923

30. Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder—A covariance model based RNA motif finding algorithm. Bioinformatics 22: 445–452. https://doi.org/10.1093/bioinformatics/btk008 PMID: 16357030

31. Rivas E, Clements J, Eddy SR (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods 14: 45–48. https://doi.org/10.1038/nmeth.4066 PMID: 27819659

32. Lindgreen S, Gardner PP, Krogh A (2006) Measuring covariation in RNA alignments: physical realism improves information measures. Bioinformatics 22: 2988–2995. https://doi.org/10.1093/bioinformatics/btl514 PMID: 17038338

33. Parsch J, Braverman JM, Stephan W (2000) Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics 154: 909–921. PMID: 10655240

34. Donahue CP, Muratore C, Wu JY, Kosik KS, Wolfe MS (2006) Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. J Biol Chem 281: 23302–23306. https://doi.org/10.1074/jbc.C600143200 PMID: 16782711

35. Dickson JR (2013) The role of the human tau 3′-untranslated region in regulating tau expression [Dissertation]. Boston, MA: Harvard University. 119 p.

36. Luo Y, Disney MD (2014) Bottom-up design of small molecules that stimulate exon 10 skipping in mutant MAPT pre-mRNA. ChemBioChem 15: 2041–2044. https://doi.org/10.1002/cbic.201402069 PMID: 25115866

37. Fallmann J, Will S, Engelhardt J, Gruning B, Backofen R, Stadler PF (2017) Recent advances in RNA folding. J Biotechnol 261: 97–104. https://doi.org/10.1016/j.jbiotec.2017.07.007 PMID: 28690134

38. Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A 101: 7287–7292. https://doi.org/10.1073/pnas.0401799101 PMID: 15123812

39. Turner DH, Mathews DH (2010) NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38: D280–282. https://doi.org/10.1093/nar/gkp892 PMID: 19880381

40. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415. https://doi.org/10.1093/nar/gkg595 PMID: 12824337

41. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429–3431. https://doi.org/10.1093/nar/gkg599 PMID: 12824340

42. Bellaousov S, Reuter JS, Seetin MG, Mathews DH (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res 41: W471–474. https://doi.org/10.1093/nar/gkt290 PMID: 23620284

43. Mathews DH, Moss WN, Turner DH (2010) Folding and finding RNA secondary structure. Cold Spring Harb Perspect Biol 2: a003665. https://doi.org/10.1101/cshperspect.a003665 PMID: 20685845

44. Clote P, Ferre F, Kranakis E, Krizanc D (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. RNA 11: 578–591. https://doi.org/10.1261/rna.7220505 PMID: 15840812

45. Fu Y, Xu ZZ, Lu ZJ, Zhao S, Mathews DH (2015) Discovery of Novel ncRNA Sequences in Multiple Genome Alignments on the Basis of Conserved and Stable Secondary Structures. PLoS One 10: e0130200. https://doi.org/10.1371/journal.pone.0130200 PMID: 26075601

46. Gruber AR, Findeiss S, Washietl S, Hofacker IL, Stadler PF (2010) RNAz 2.0: improved noncoding RNA detection. Pac Symp Biocomput: 69–79. PMID: 19908359

47. Dela-Moss LI, Moss WN, Turner DH (2014) Identification of conserved RNA secondary structures at influenza B and C splice sites reveals similarities and differences between influenza A, B, and C. BMC Res Notes 7: 22. https://doi.org/10.1186/1756-0500-7-22 PMID: 24405943

48. Fang R, Moss WN, Rutenberg-Schoenberg M, Simon MD (2015) Probing Xist RNA Structure in Cells Using Targeted Structure-Seq. PLoS Genet 11: e1005668. https://doi.org/10.1371/journal.pgen.1005668 PMID: 26646615

49. Jiang T, Nogales A, Baker SF, Martinez-Sobrido L, Turner DH (2016) Mutations Designed by Ensemble Defect to Misfold Conserved RNA Structures of Influenza A Segments 7 and 8 Affect Splicing and Attenuate Viral Replication in Cell Culture. PLoS One 11: e0156906. https://doi.org/10.1371/journal.pone.0156906 PMID: 27272307

50. Moss WN, Dela-Moss LI, Kierzek E, Kierzek R, Priore SF, Turner DH (2012) The 3' splice site of influenza A segment 7 mRNA can exist in two conformations: a pseudoknot and a hairpin. PLoS One 7: e38323. https://doi.org/10.1371/journal.pone.0038323 PMID: 22685560

51. Moss WN, Dela-Moss LI, Priore SF, Turner DH (2012) The influenza A segment 7 mRNA 3' splice site pseudoknot/hairpin family. RNA Biol 9: 1305–1310. https://doi.org/10.4161/rna.22343 PMID: 23064116

52. Moss WN, Lee N, Pimienta G, Steitz JA (2014) RNA families in Epstein-Barr virus. RNA Biol 11: 10–17. https://doi.org/10.4161/rna.27488 PMID: 24441309

53. Roca X, Akerman M, Gaus H, Berdeja A, Bennett CF, Krainer AR (2012) Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides. Genes Dev 26: 1098–1109. https://doi.org/10.1101/gad.190173.112 PMID: 22588721

54. Wang J, Tse SW, Andreadis A (2007) Tau exon 6 is regulated by an intricate interplay of trans factors and cis elements, including multiple branch points. J Neurochem 100: 437–445. https://doi.org/10.1111/j.1471-4159.2006.04252.x PMID: 17144905

55. Andreadis A (2005) Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. Biochim Biophys Acta 1739: 91–103. https://doi.org/10.1016/j.bbadis.2004.08.010 PMID: 15615629

56. Wang Z, Burge CB (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. RNA 14: 802–813. https://doi.org/10.1261/rna.876308 PMID: 18369186

57. Hong M, Zhukareva V, Vogelsberg-Ragaglia V, Wszolek Z, Reed L, Miller BI, et al. (1998) Mutation-specific functional impairments in distinct tau isoforms of hereditary FTDP-17. Science 282: 1914–1917. https://doi.org/10.1126/science.282.5395.1914 PMID: 9836646

58. Spina S, Farlow MR, Unverzagt FW, Kareken DA, Murrell JR, Fraser G, et al. (2008) The tauopathy associated with mutation +3 in intron 10 of Tau: Characterization of the MSTD family. Brain 131: 72–89. https://doi.org/10.1093/brain/awm280 PMID: 18065436

**59.** Lisowiec J, Magner D, Kierzek E, Lenartowicz E, Kierzek R (2015) Structural determinants for alternative splicing regulation of the MAPT pre-mRNA. RNA Biol 12: 330–342. https://doi.org/10.1080/15476286.2015.1017214 PMID: 25826665

**60.** Malkani R, D'Souza I, Gwinn-Hardy K, Schellenberg GD, Hardy J, Momeni P (2006) A MAPT mutation in a regulatory element upstream of exon 10 causes frontotemporal dementia. Neurobiol Dis 22: 401–403. https://doi.org/10.1016/j.nbd.2005.12.001 PMID: 16503405

**61.** Wang Y, Wang Z (2014) Systematical identification of splicing regulatory cis-elements and cognate trans-factors. Methods 65: 350–358. https://doi.org/10.1016/j.ymeth.2013.08.019 PMID: 23974071

**62.** D'Souza I, Schellenberg GD (2002) tau exon 10 expression involves a bipartite intron 10 regulatory sequence and weak 5′ and 3′ splice sites. J Biol Chem 277: 26587–26599. https://doi.org/10.1074/jbc.M203794200 PMID: 12000767

**63.** Wang Y, Wang J, Gao L, Lafyatis R, Stamm S, Andreadis A (2005) Tau exons 2 and 10, which are mis-regulated 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. https://doi.org/10.1074/jbc.M413846200 PMID: 15695522

**64.** Jiang Z, Tang H, Havlioglu N, Zhang X, Stamm S, Yan R, et al. (2003) Mutations in tau gene exon 10 associated with FTDP-17 alter the activity of an exonic splicing enhancer to interact with Tra2β. J Biol Chem 278: 18997–19007. https://doi.org/10.1074/jbc.M301800200 PMID: 12649279

**65.** Wang J, Gao QS, Wang Y, Lafyatis R, Stamm S, Andreadis A (2004) 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. https://doi.org/10.1046/j.1471-4159.2003.02232.x PMID: 15009664

**66.** Wu JY, Kar A, Kuo D, Yu B, Havlioglu N (2006) SRp54 (SFRS11), a regulator for tau exon 10 alternative splicing identified by an expression cloning strategy. Mol Cell Biol 26: 6739–6747. https://doi.org/10.1128/MCB.00739-06 PMID: 16943417

**67.** Ogino S, Gulley ML, den Dunnen JT, Wilson RB, Committtee AfMPTaE (2007) Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. J Mol Diagn 9: 1–6. https://doi.org/10.2353/jmoldx.2007.060081 PMID: 17251329

**68.** Hasegawa M, Smith MJ, Iijima M, Tabira T, Goedert M (1999) FTDP-17 mutations N279K and S305N in tau produce increased splicing of exon 10. FEBS Lett 443: 93–96. https://doi.org/10.1016/s0014-5793(98)01696-2 PMID: 9989582

**69.** Jiang Z, Cote J, Kwon JM, Goate AM, Wu JY (2000) Aberrant splicing of tau pre-mRNA caused by intronic mutations associated with the inherited dementia frontotemporal dementia with parkinsonism linked to chromosome 17. Mol Cell Biol 20: 4036–4048. https://doi.org/10.1128/mcb.20.11.4036-4048.2000 PMID: 10805746

**70.** Wilhelmsen KC, Lynch T, Pavlou E, Higgins M, Nygaard TG (1994) Localization of disinhibition-dementia-parkinsonism-amyotrophy complex to 17q21-22. Am J Hum Genet 55: 1159–1165. PMID: 7977375

**71.** Morita T, Sobue K (2009) Specification of neuronal polarity regulated by local translation of CRMP2 and Tau via the mTOR-p70S6K pathway. J Biol Chem 284: 27734–27745. https://doi.org/10.1074/jbc.M109.008177 PMID: 19648118

**72.** Nemkul N (2016) Mechanisms of regulation of tau IRES mediated translation [Dissertation]. Houston, Texas: The University of Texas Health Science Center at Houston. 152 p.

**73.** Veo BL, Krushel LA (2012) Secondary RNA structure and nucleotide specificity contribute to internal initiation mediated by the human tau 5′ leader. RNA Biol 9: 1344–1360. https://doi.org/10.4161/rna.22181 PMID: 22995835

**74.** Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N,Yakhini Z, et al. (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. Science 351: aad4939. https://doi.org/10.1126/science.aad4939 PMID: 26816383

**75.** Barrett LW, Fletcher S, Wilton SD (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell Mol Life Sci 69: 3613–3634. https://doi.org/10.1007/s00018-012-0990-9 PMID: 22538991

**76.** Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Mol Cell 43: 853–866. https://doi.org/10.1016/j.molcel.2011.08.017 PMID: 21925375

**77.** Gruber AR, Findeiss S, Washietl S, Hofacker IL, Stadler PF (2010) RNAz 2.0: Improved noncoding RNA detection. Pac Symp Biocomput: 69–79. PMID: 19908359

**78.** Garcia-Escudero V, Gargini R, Martin-Maestro P, Garcia E, Garcia-Escudero R, Avila J (2017) Tau mRNA 3′ UTR-to-CDS ratio is increased in Alzheimer disease. Neurosci Lett 655: 101–108. https://doi.org/10.1016/j.neulet.2017.07.007 PMID: 28689927

**79.** Aronov S, Marx R, Ginzburg I (1999) Identification of 3′ UTR region implicated in tau mRNA stabilization in neuronal cells. J Mol Neurosci 12: 131–145. https://doi.org/10.1007/BF02736927 PMID: 10527457

80. Crary J, Perez IS-M (2013) Micrornas that silence tau expression.  United States of America:  Columbia University of New York

81. Santa-Maria I, Alaniz ME, Renwick N, Cela C, Fulga TA, Van Vactor D, et al. (2015) Dysregulation of microRNA-219 promotes neurodegeneration through post-transcriptional regulation of tau. J Clin Invest 125: 681–686. https://doi.org/10.1172/JCI78421 PMID: 25574843

82. Lau P, Bossers K, Janky R, Salta E, Frigerio CS, Barbash S, et al. (2013) Alteration of the microRNA network during the progression of Alzheimer's disease. EMBO Mol Med 5: 1613–1634. https://doi.org/10.1002/emmm.201201974 PMID: 24014289

83. Velagapudi SP, Cameron MD, Haga CL, Rosenberg LH, Lafitte M, Duckett DR, et al. (2016) Design of a small molecule against an oncogenic noncoding RNA. Proc Natl Acad Sci U S A  113: 5898–5903. https://doi.org/10.1073/pnas.1523975113 PMID: 27170187

84. Velagapudi SP, Gallo SM, Disney MD (2014) Sequence-based design of bioactive small molecules that target precursor microRNAs. Nat Chem Biol  10: 291–297. https://doi.org/10.1038/nchembio.1452 PMID: 24509821

85. Costales MG, Haga CL, Velagapudi SP, Childs-Disney JL, Phinney DG, Disney MD (2017) Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. J Am Chem Soc 139: 3446–3455. https://doi.org/10.1021/jacs.6b11273 PMID: 28240549

86. Rzuczek SG, Colgan LA, Nakai Y, Cameron MD, Furling D, Yasida R, et al. (2017) Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat Chem Biol 13: 188–193. https://doi.org/10.1038/nchembio.2251 PMID: 27941760