2021

Cloning and Functional Characterizations of Circular RNAs from the Human *MAPT* Locus

Justin R. Welden
University of Kentucky, jrwe234@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-5525-3724
Digital Object Identifier: https://doi.org/10.13023/etd.2021.018

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Justin R. Welden, Student

Dr. Stefan Stamm, Major Professor

Dr. Trevor Creamer, Director of Graduate Studies

CLONING AND FUNCTIONAL CHARACTERIZATIONS OF CIRCULAR RNAS FROM THE HUMAN *MAPT* LOCUS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Medicine
at the University of Kentucky

By
Justin Ralph Welden
Lexington, Kentucky
Director: Dr. Stefan Stamm, Professor of Molecular and Cellular Biochemistry
Lexington, Kentucky
2021

https://orcid.org/0000-0001-5525-3724

ABSTRACT OF DISSERTATION

CLONING AND FUNCTIONAL CHARACTERIZATIONS OF CIRCULAR RNAS FROM THE HUMAN *MAPT* LOCUS

Under pathophysiological conditions, the microtubule protein tau (*MAPT*) forms neurofibrillary tangles that are the hallmark of sporadic Alzheimer's disease as well as familial frontotemporal dementias linked to chromosome 17 (FTDP-17). In this work, I report that *MAPT* forms circular RNAs through backsplicing of exon 12 to either exon 10 or exon 7 (12→10; 12→7), and that these circular RNAs are translated into proteins.

Using stable cell lines overexpressing the circular tau RNAs 12→7 and 12→10, we have discovered that the tau circular RNA 12→7 is translated in a rolling circle, giving rise to multiple proteins. This circular RNA has an endogenous start codon in exon 9 and no in frame stop codon. We purified the protein by immunoprecipitation using a 3X flag tag that is upstream of the start codon and that will only be translated in a circular RNA. The circular RNA 12→10 does not have a start codon and is translated when mutations that cause FTDP-17 create a start codon (K317M and V337M) or ADAR (Adenosine Deaminase acting on RNAs) enzymes are present. In addition, we show that the 12→10 wild type circular RNA can be translated due to RNA editing by ADAR1 and ADAR2. ADAR enzymes edit RNA sequences changing adenosines to inosines and are referred to as A-to-I editing (I = inosine). Inosines can be read as a guanosine, therefore editing an AUA sequence to AUI could create a new start codon in the RNA sequence.

The circular RNAs' translated region is similar to a seed sequence, known as K18, that promotes tau aggregation and neurodegeneration. We hypothesize that the tau proteins generated from the circular RNAs contribute to tau aggregation, which cause Alzheimer's Disease. We have designed siRNAs, specific for circular RNAs, as possible therapeutics. We show that the siRNAs target the circular tau RNAs backsplice junction, reducing their expression and preventing translation.

In summary, our data indicate that tau makes so far unknown circular RNAs that are unexpectedly translated into proteins that cause tau aggregations, a hallmark of tauopathies (AD and FTDP-17). Almost all FTDP-17 mutations are in the pre-mRNA regions contributing to circular RNA formation, suggesting they act through circular

RNAs. Finally, we identified oligonucleotides that abolish circular RNA formation, pointing to a novel therapeutic avenue for tauopathies.

I am thankful for the partial funding of this project by the College of Medicine Excellence in Graduate Research Fellowship and the Max Steckler Award.

KEYWORDS: Circular RNA, Tauopathies, Alu elements, *MAPT,* Alzheimer's Disease, FTDP-17

Justin Ralph Welden

(Name of Student)

01/14/2021

Date

CLONING AND FUNCTIONAL CHARACTERIZATIONS OF CIRCULAR RNAS FROM THE HUMAN *MAPT* LOCUS

By
Justin Ralph Welden

Dr. Stefan Stamm
Director of Dissertation

Dr. Trevor Creamer
Director of Graduate Studies

01/14/2021
Date

ACKNOWLEDGMENTS

I want to thank my mentor, Dr. Stefan Stamm, who has challenged me to be an excellent scientist. Stefan has always encouraged me not to give up and keep pushing myself in many aspects from the beginning as a lab technician through Grad school. I want to thank you, Stefan, for taking me to the RNA and Prader-Willi conferences and allowing me to give the talks and network with other scientists. Also, Stefan provided timely and instructive comments and evaluation at every stage of the dissertation process, allowing me to complete this project on schedule. Next, I wish to thank the complete Dissertation Committee and outside examiner, respectively: Drs. Matthew Gentry, Isabel Mellon, Jessica Blackburn, and Clark Kebodeaux. Each individual provided insights that guided and challenged my thinking, substantially improving the finished product.

I want to thank the former and present members of the Stamm Lab: Jay van Doorn, Samantha Danyi, Anna Pawluchin, Thomas Zerbes, LiYing Wu, Patrick Cordero, Jonah Dixon, Chi Jing Leow, Giorgi Margvelani, and Bhavani Gudlavalleti. With their help, this project was able to be completed on schedule.

In addition to the support I received above, nothing can compare to the love and support I have had from my parents Robert and Pamela Welden, and my brother Jason Welden and sister Stacy Jones. I would also like to thank my friends that have been by my side and supported me through this journey. Finally, I want to thank the Biochemistry Department Administrators and Personnel for helping the students with tasks outside of the lab.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SUPPLEMENTAL FIGURES

CHAPTER 1. INTRODUCTION

1.1 Overview

This dissertation investigates the human microtubule associated protein tau (*MAPT*) gene. The main finding is the discovery of novel circular RNAs that are generated from the tau locus (1). Not a lot is known about circular RNAs and very few circular RNAs have a known function. The tau protein is known to be associated with many different neurodegenerative diseases called tauopathies and is a major contributor to Alzheimer's disease (2). The tau protein can be hyperphosphorylated causing it to disassemble from tubulin forming tau aggregates, which leads to neurodegeneration (3). However, little is known about the mechanism that leads to this dysfunction of the protein.

In **chapter 2**, I give a review on how circular RNAs are formed in pre-mRNA processing and summarize their identified function (4).

In **chapter 3**, I describe the discovery of the tau circular RNAs and look into the correlation of circular RNA abundance with AD. The tau locus generates two circular RNAs that consist of exons 7,9-12 (12→7) and 10-12 (12→10) (1).

Chapter 4 describes an optimized protocol of cloning reporter genes that form circular RNAs in order to study in detail their function and processing. This protocol can be used for any gene that generates circular RNAs showing key components to consider when troubleshooting large complex sequences and that have repeat sequences (5). This optimized protocol allowed us to study the function of the tau circular RNAs and their relation to tauopathies.

In **chapter 5**, I look into the function of the circular tau RNAs and found that they are translated. Both the tau circular RNAs 12→7 and 12→10 contain an open reading frame, however, the 12→7 circular RNA contains one in-frame methionine start codon in exon 9 but no stop codon. I show that the 12→7 circular RNA is translated in a rolling circle giving rise to proteins of different sizes that are predicted to contain the regions of the microtubule binding domains. A peptide that consists of the four microtubule binding domains called K18 has been shown to promote tau aggregation and paired helical filament formation (6, 7). The four microtubule binding domains reside in exons 9-12 and when translated from the 12→7 circular RNA they can form a multimer that could initiate tau aggregation. The 12→10 circular RNA does not contain a methionine start codon and is not predicted to be translated. However, two particular mutations in tau exon 11 (K317M) and exon 12 (V337M) that cause Frontotemporal Dementia, create a start codon allowing translation to occur. We show that the wild type 12→10 circular RNA can be translated when there is an overexpression of the adenosine deaminase acting on RNA (ADAR) enzymes. The active ADAR enzymes, ADAR1 and ADAR2, may edit the RNA changing an AUA sequence to an AUI that now is read as an AUG start codon (I = Inosine). ADAR enzymes have also been shown to be deregulated in Alzheimer's disease which may contribute to circular tau RNA translation and tau pathology. Another function of the tau proteins discovered in

chapter 5 was the interaction with the eukaryotic translation initiation factor 4B (EIF4B). The translated proteins from the tau circular RNAs could be regulating gene expression or translation contributing to tauopathies.

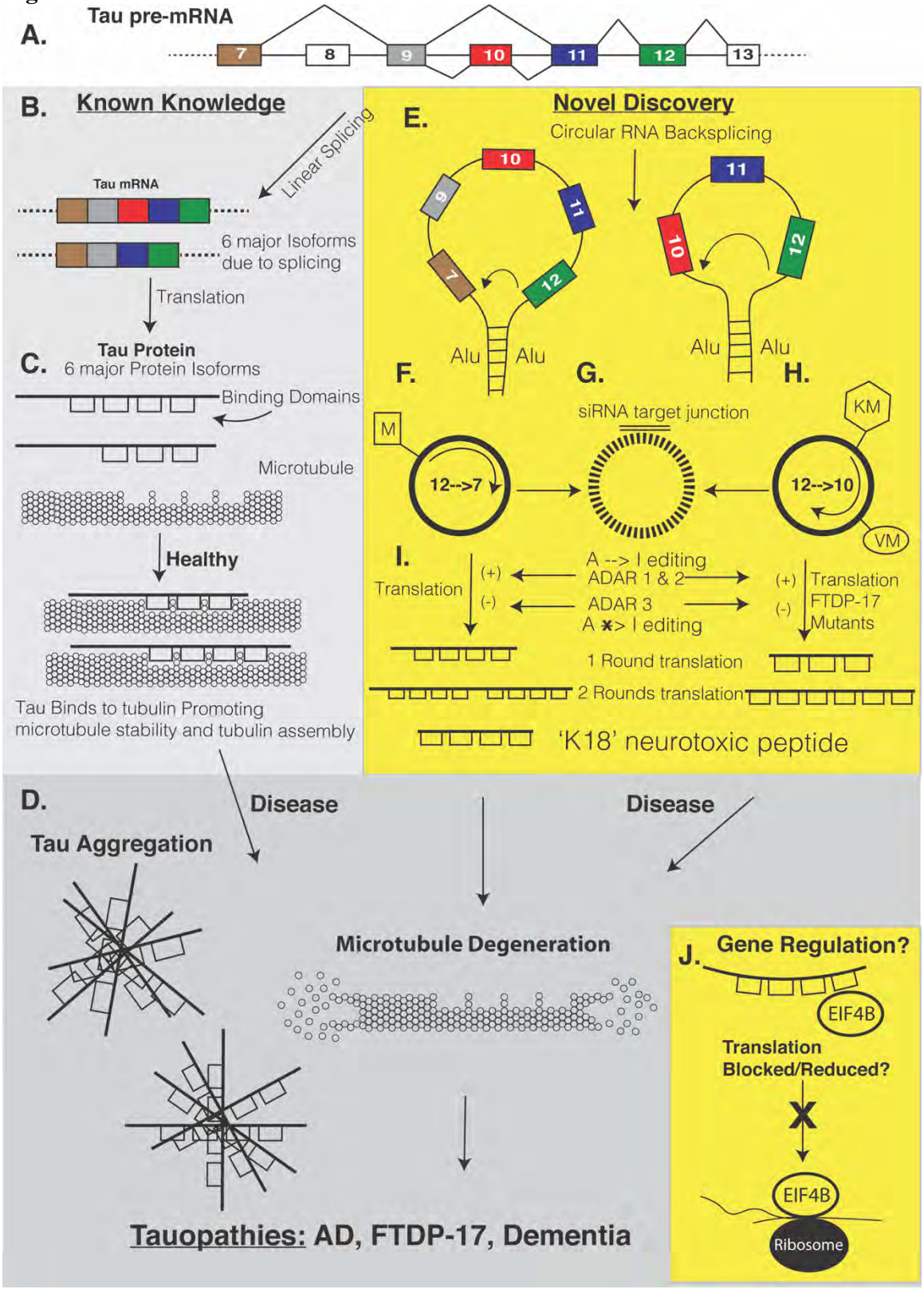
Lastly, I show that the circular tau RNAs can be reduced by siRNAs targeting against the backsplice junction preventing translation. The translated protein could be neurotoxic and with the siRNAs specifically targeting the circular and not the linear RNA the siRNAs could be used as a therapeutic drug for Alzheimer's disease and other tauopathies.

In summary, I propose a model on what is known in the processing of the tau protein shaded in gray in figure 1.1 A-D. The tau gene forms the linear pre-mRNA, which can be alternatively spliced giving rise to the six major isoforms of the tau protein (Figure 1.1 A-C). Tau binds to tubulin stabilizing microtubules promoting tubulin growth preventing neurodegeneration (Figure 1.1 C). However, due to aging or brain trauma, tau can be hyperphosphorylated promoting tau aggregation forming neurofibrillary tangles and paired helical filaments, leading to the destabilization of microtubules and neurodegeneration (Figure 1.1 D). I propose a model that includes the discovery of the circular tau RNAs in yellow in figure 1.1 E-I, we hypothesize that in order for the backsplicing to occur, Alu repeat elements are needed allowing the pre-mRNA to adopt a structure promoting backsplicing (Figure 1.1 E). Alu repeats are intronic regions that exhibit sequence complementarity, causing exon 12 to be close to either exon 7 or exon 10, allowing the backsplicing to occur.

The circular tau RNA 12→7 is translated, and its proteins may have a physiological role, but when there is an increase in translation due to overexpression of ADAR1 and ADAR2, this may lead to a gain of function promoting tau aggregation and neurodegeneration (Figure 1.1 D, F, I). The tau protein also binds to EIF4B and may play a role in gene regulation or translation where an overexpression of the protein may act as a gain of function decreasing gene expression (Figure 1.1 J). The circular RNA 12→10 is only translated when mutations create new start codons or when there is an overexpression of ADAR enzymes that can edit the RNA changing an adenosine to an inosine where the inosine and can be read as a guanosine (Figure 1.1 H, I). Editing the sequence AUA to AUI can be read as an AUG forming a new start codon. The translated protein from the 12→10 circular RNA may act as a seed sequence that can lead to tau aggregation and neurodegeneration causing Alzheimer's or other tauopathies (Figure 1.1 D, H, I).

The circular tau RNAs can be reduced by siRNAs, which could stop translation preventing tau aggregation and neurodegeneration (Figure 1.1 G). So far, no one has looked into this pathway from the circular RNAs, research has only focused on the linear tau RNAs. Thus, this work unveiled a new mechanism to target in the formation of the possibly neurotoxic tau proteins and can potentially prevent the progression of the tau pathology.

Figure 1-1 Model of Microtubule-Associated Protein Tau Linear vs Circular



Known knowledge is in gray and my finding is in yellow.

A. Schematic of the shortened pre-mRNA of the tau locus and how it is alternatively spliced.
B. After alternative splicing six major RNA isoforms are formed.
C. The six major protein isoforms are translated and interacts with tubulin stabilizing microtubules.
D. When tau no longer binds to tubulin and forms tau aggregates, disease pathology forms, causing tubulin disassembly leading to Tauopathies.
E. The novel discovery in yellow, two circular RNAs form due to the backsplicing of exon 12 onto either exon 7 or exon 10 (12→7, 12→10).
F. 12→7 circular tau RNA is translated in a rolling circle.
G. The circular tau RNAs can be destroyed or reduced with siRNAs targeting the backsplice junction inhibiting protein translation.
H. The 12→10 circular RNA is translated with FTDP-17 mutations K317M and V337M. Editing enzymes (ADAR1 and ADAR2) increase protein production due to RNA editing. The translated regions of the circular tau RNA 12→10 is predicted to be similar to the neurotoxic peptide called K18.
I. Both circular RNAs 12→ and 12→10 are translated forming multiple proteins of different sizes, a possible indication of multiple rounds of translation. Editing enzymes (ADAR1 and ADAR2) increase protein production due to RNA editing. The translated region of the circular tau RNAs is predicted to be similar to the neurotoxic peptide called K18.
J. EIF4B interacts with the protein from the tau circular RNAs possibly influencing translation.

CHAPTER 2. PRE-MRNAS FORMING CIRCULAR RNAS ARE A NEWLY DISCOVERED ELEMENT OF GENE EXPRESSION

This chapter was adapted from:

Welden, J. R. and Stamm, S. (2019). "Pre-mRNA structures forming circular RNAs." Biochim Biophys Acta Gene Regul Mech. 2019:194410 (4)

2.1 Introduction

Circular RNAs (circRNAs) are covalently closed RNAs that are expressed in all branches of life (8). Most circRNAs are generated through pre-mRNA back-splicing (9) where a downstream 5' splice site is connected to an upstream 3' splice site (Figure 2.1A). Most circRNAs contain exons of mRNAs, but circRNAs composed of pre-mRNA introns have also been detected (10). Other mechanisms to create circRNAs like tRNA splicing (11), self-splicing of tetrahymena rRNA (12), and the formation of circular viroid RNAs also exist (13, 14), but are far less common.

In general, circRNAs have a much lower abundance than linear mRNAs and early reports considered them splicing artifacts (15). As circRNAs lack a poly adenosine tail, they are diminished in libraries made by oligo dT priming. Their widespread expression became only apparent when next generation sequencing techniques were developed (16). Currently it is estimated that, in general, circRNAs comprise 0.8-1% of mRNAs (16, 17).

The overall function of circRNAs remains enigmatic. CircRNAs accumulate in the cytosol using an export pathway that depends on the RNA helicases UAP56 (DDX39B) and URH49 (DDX39A) (18), pointing towards functions in the cytosol. Some circRNAs have been shown to sequester miRNAs, acting as 'sponges' (19, 20), which indirectly influences mRNA abundance. Numerous circRNAs are associated with ribosomes (21) and proof of principle experiments indicated that they can be translated (22-27). In drosophila and humans, the highest diversity and expression of circRNAs has been found in the brain and increases during aging (28-30) suggesting tissue-specific roles for circRNAs.

The vast majority of exons in circRNAs are also present in their linear counterparts, suggesting a competition between linear splicing and back-splicing. In some cases, this competition can reduce linear mRNA expression (9, 31, 32), indicating a role of some circRNAs in mRNA expression.

2.2 Types of circular RNAs

The highest expressed individual circRNAs are generated by a single large exon through a back-splicing mechanism (33). Overall most circRNAs contain multiple exons, mostly two exons. The number of circRNA transcripts decreases with the number of their forming exons (33). Similar to linear mRNAs, circRNAs can be alternatively spliced (33, 34). Numerous circRNAs contain intronic sequences, but the majority of circRNAs lack intronic sequences, suggesting that multi-exon circRNAs are generated after parts of the

pre-mRNA have been already spliced together, indicating that frequently linear splicing likely occurs prior to back-splicing (28).

Reflecting the low abundance and cell-type specific expression, there is a large discrepancy between circRNAs identified in different experiments (17) and new circRNAs are being discovered (34). CircRNA sequences are available from various databases: (CIRCpedia (34), circBase (35), CircFunBase (36), CircNet (37), circRNADb (38), and PlantcircBase (39).

2.3 Mechanism of back-splicing

Back-splicing that creates circRNAs uses the same splicing machinery and most of the splicing enhancer and silencers as linear premRNA splicing. Thus, exons present in circRNAs are surrounded by canonical splice sites in the pre-mRNA and the formation of circRNAs is sensitive to splicing inhibition (40). siRNA mediated depletion of splicing factors increases formation of some circRNAs while the corresponding mRNAs were reduced, suggesting a coupling between circRNA and mRNA formation (41). Minigene analyses showed that, similar to linear splicing, the recognition of exons in circRNAs is based on combinatorial control that takes into account splice site strength (40), the concentration of hnRNPs, SR-proteins, SR-protein kinases (42, 43), and RNA-helicases (44). For circRNAs made from single exons, the yield of circRNA formation increases with exon length (45) and in drosophila long flanking introns favor circRNA formation (28).

2.4 Recognition of back-splicing sites depends on the pre-mRNA secondary structure

For back-splicing to occur, the splice sites involved have to be brought into close contact, either through a secondary structure in the RNA (Figure 2.1A) or through interactions (Figure 2.1B). In most cases studied, these structures are generated through base-pairing of complementary RNA sequences within the same pre-mRNA molecule. Protein interactions have been shown for quaking I (QKI) that binds to short intronic recognition sites flanking the back-splicing sites resulting in circRNA formation (46). Similarly, the muscle blind protein (MBL) binds to repeat sequences flanking back-splicing sites and also promotes circRNA formation (47). Bridging of exons through intronic hnRNP binding sites have been shown for other proteins, such as hnRNPA1 (48) and PTB/hnRNPI (49), suggesting a more general role for this mechanism in circRNA formation. In lower eukaryotes that have less inverted repeats, circRNAs can be formed through a lariat precursor, where back-splicing occurs within the circular lariat (50) (Figure 2.1C).

Due to its high expression, one of the first circRNAs identified was from the mouse SRY gene (sex determining region Y) (51). This single exon circRNA is located in a 2.7 kb long region flanked by a large inverted repeats (> 15.5 kb) that is necessary for the circularization (51, 52), providing the first evidence that often genomic repeat elements

flanking the back-splice sites generate the secondary structures necessary for proper alignment of the splice sites. In humans, these repeat elements are often provided by *Alu* elements (9) and their contribution to circRNA formation has been extensively studied experimentally (33, 45). Similarly, reverse complementary sequences not emanating from repeat elements have been detected in C. elegans (53). However, a genome-wide screen in drosophila failed to detect direct repeats for the majority of circRNAs (28), although detailed experiments using model RNAs revealed that some highly expressed drosophila circRNAs, like laccase 2 (Figure 2.3C) rely on repeats for their formation (42).

Genome wide cloning of human RNase resistant double stranded RNA revealed thousands of double stranded RNAs outside known repeats, further supporting the idea of RNA double stranded structures that occur independent of repeat elements (54). Thus, despite the current research focus on inverted repeat elements, especially Alu elements, other RNA structures likely also facilitate back-splicing.

2.5 *Alu* elements form secondary structures promoting human circular RNA formation

Sequencing of human circRNAs revealed that they are often flanked by Alu elements in their genomic location (9, 33). Alu elements are about 300 nucleotides (nt) long short interspersed nuclear elements (SINE) (55-57) that were derived from the 7SL-RNA in the early primate lineage (58, 59) (Figure 2.2A, B). There are more than one million *Alu* element copies in the human genome that comprise about 11% of the human genome (60). *Alu* elements continue to amplify through a polymerase III-derived RNA intermediate and it is estimated that there is about one new *Alu* insertion in 21 human births (61). They are predominantly located in gene-rich regions (62), possibly because their small size does not interfere strongly with gene regulation (55). *Alu* elements are subdivided into the J (Jurka, Jerzy), S (Smith, Temple) and Y (young) subfamilies, which reflect an alphabetical progression from the oldest (J) to the youngest (Y) *Alu* subtypes (63). Since different subfamilies of *Alu* elements are highly similar in sequence, they can form secondary structures when inserted in opposite orientation in a pre-mRNA transcript, referred to as inverted repeated *Alu* structure (IRAlus) (55) (Figure 2.2C, example in Figure 2.3A, B)

The ability of *Alu*-elements to form double-stranded secondary structures and thus their propensity to promote circRNA formation is reduced by adenosine to inosine (A->I) editing, which is the most common post-transcriptional RNA modification in primates (64, 65). A->I editing is performed by the adenosine deaminases acting on RNA (ADAR) family of enzymes that require RNA double strands as a template (66). Thus, the editing of Alu elements can be used to measure whether an Alu element is in a double stranded conformation. The double strandedness, (i.e. the formation of a secondary structure) of an Alu-element mostly depends on the distance to the closest reversely oriented Alu element, which is around 800 nt. The Alu-element editability decreases exponentially with longer distances (64). Other factors affecting double strandedness include the presence of additional competing Alu elements, which decrease double strandedness of a single Alu

element; and the subfamily of the Alu element, where related subfamilies are edited stronger (Figure 2.2C). ADAR knockdown increases circRNA expression in human cells (53), further suggesting a role of Alu-element modification in human circRNA formation. ADARs are not specific for double stranded RNAs formed by Alu elements, and thus ADAR knockdown in mouse promotes formation of some circRNAs as well (29).

Double stranded Alu-element structures can form within an intron as well as across an exon. Usually, only Alu elements localized across an exon in opposite directions can promote circRNA formation (33). Often, an Alu-element can base-pair with other distinct Alu-elements, leading to alternative circRNA formation (Figure 2.2D). Since linear splicing reduces the distance between Alu-elements that flank exons it promotes circRNA formation when it occurs faster than back-splicing (Figure 2.2E, F). Deletion analyses in several model systems showed that the full Alu element is not necessary for circRNA formation to occur (33). In fact complementary sequences as small a 30-40 nt are sufficient and give stronger circRNA formation than longer elements (45) (Figure 2.3A).

DHX9 is an RNA helicase that selectively binds to Alu-elements located in proximity (458 nt). DHX9 binds to an interferon-inducible form of ADAR, and knock down of DHX9 increases circRNA production (44, 67), suggesting that DHX9 destabilizes pre-mRNA structures necessary for circRNA formation.

There is emerging evidence that Alu-elements can also act in trans, i.e. between different pre-mRNA molecules, leading to trans-splicing (68), which could conceptually compete with circRNA splicing.

In summary, multiple factors control the occurrence of back-splicing and thus the formation of circRNAs. Most importantly, a double stranded pre-mRNA structure brings back-splicing sites together. Its formation is influenced by the availability of complementary sequences, which in turn are negatively influenced by RNA helicases, ADAR enzymes and single-stranded RNA binding proteins. Factors that influence intron splicing, such as splice site strength, the concentration of SR-proteins and hnRNPs as well as the availability of exon enhancers and silencers also impact on circRNA formation (Figure 2.2G).

2.6 Experimental Studies

The correlations between double stranded pre-mRNA structures and circRNA formation have been tested experimentally. Current reporter gene constructs and experimentally validated double stranded RNA structures are summarized in Figure 2.3. The double stranded structures can be surprisingly small. For example the experimentally determined structure needed to form circRNAs forming from ZKSCAN1 exon 2/3 is only 36 nt long and contains 7 mismatches, although the natural repeat is much longer (45), (Figure 2.3A). It is thus not surprising that circRNAs can form without known or identifiable repeats (40).

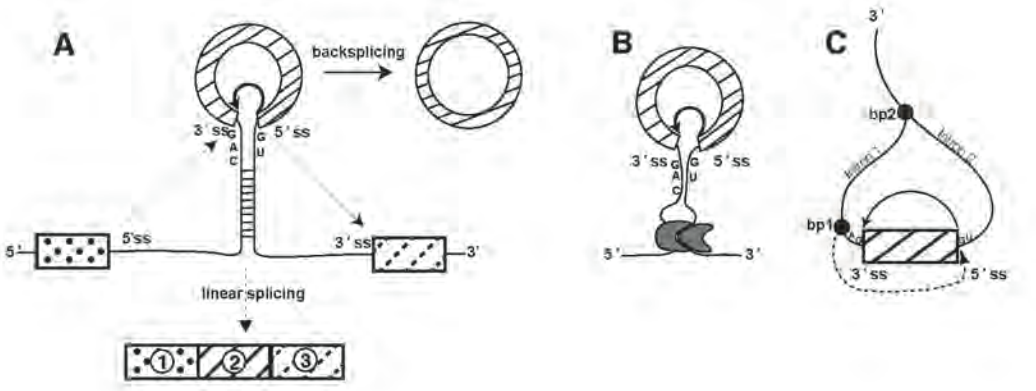
2.7 Outlook

Despite their low expression when compared to their linear counterparts, circRNAs could turn out to be biologically highly important. An increasing number of studies are showing now that circRNAs can form proteins (23, 69). CircRNAs are highly expressed in the brain, where they could encode new peptide hormones that are small, reflecting the short reading frames of circRNAs and work at small concentrations. For example, α-melanocyte-stimulating hormone (α-MSH), generated from POMC is a peptide composed of 13 amino acids that binds to the melanocortin 1 receptors with an affinity of 200 pM to 2 nM (70). Thus, the generation of novel brain-specific signaling peptides could be an important function of circRNAs. It is notable that the development of the brain cortex in primates correlates with the expansion of Alu elements (71), which could result in so far unknown peptide hormones.

The formation of circRNAs depends on intronic elements that form double-stranded RNA structures. CircRNAs and possibly their encoded peptides could be an evolutionary force reflecting differences in intron sequences caused by the presence of species-specific repeat elements, such as primate-specific Alu elements and SINEC_Cf elements characteristic for various dog species (72). Similarly, deep intronic mutations characteristic for cancer (73) and present in numerous hereditary diseases (74) could act by influencing circRNA expression.

So far, studies of pre-mRNA structures do not take RNA modifications other than A->I editing into account. N6-Adenosine methylation (m6A) is another common mRNA modification that could influence RNA structures and thus circRNA formation. M6A RNA modification creates a strong binding site for YTH-domain proteins (75) that could bridge back-splicing sites. In addition m6A modifications create hnRNPG or hnRNPC binding sites (76) which abolishes RNA double stranded structures and was shown to change linear alternative splicing (76). It is thus conceivable that circRNA formation is similarly influenced by m6A RNA modifications and could be thus under epigenetic control.

Figure 2-1 Generation of circRNAs through back-splicing

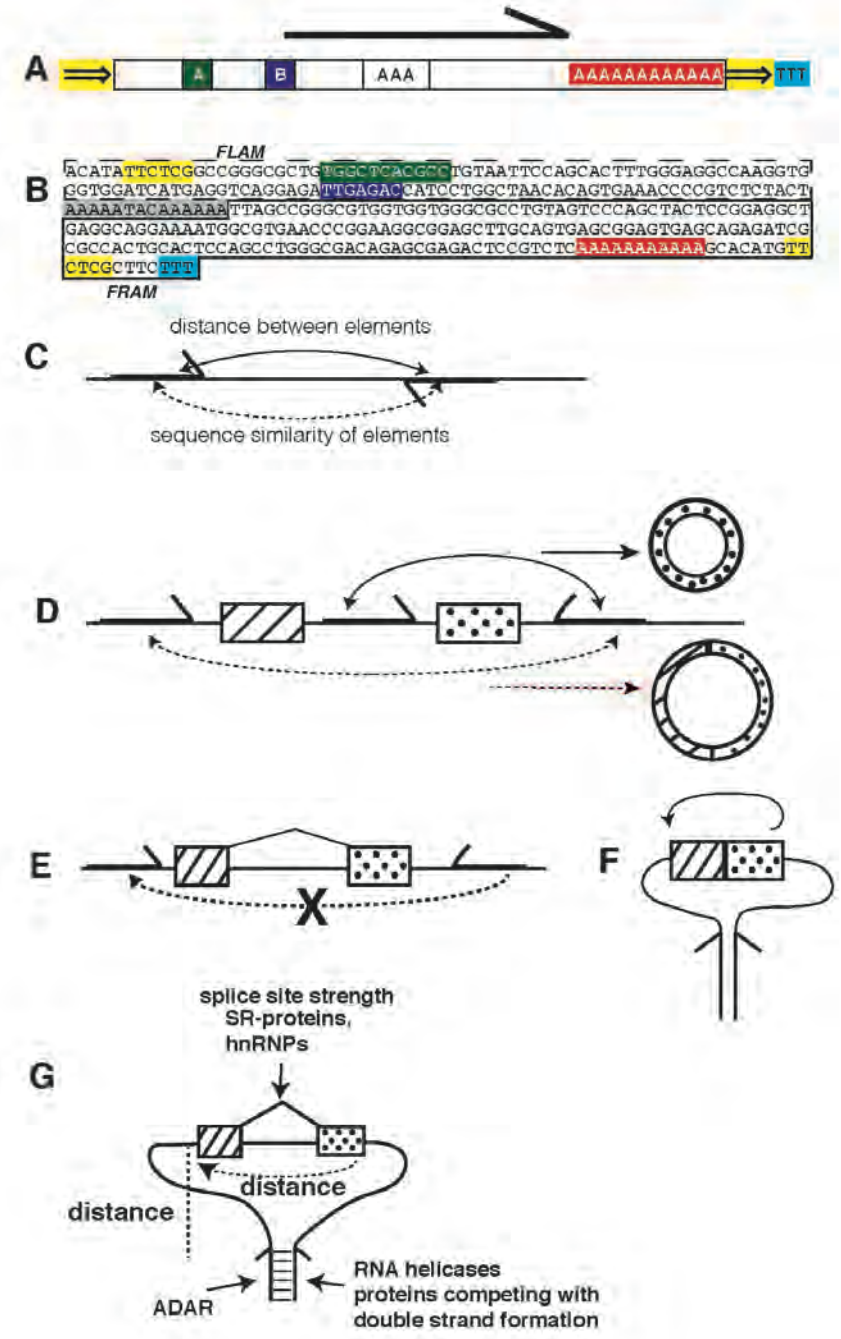


A. Back-splicing due to a pre-mRNA structure. A pre-mRNA containing exon 1, 2, and 3 is shown. The double stranded RNA structure formed by complementary regions in the pre-mRNA brings the 5' and 3' splice site of exon 2 into close proximity, allowing back-splicing (solid arrow) leading to the formation of a circRNA (striped circle). This process competes with the formation of linear mRNA, joining exons 1, 2 and 3, indicated by dashed lines with arrows.

B. Back-splicing can also occur when proteins (gray shapes) multimerize after binding to short recognition sequences that flank an exon (46), which brings the back-splicing sites into close proximity.

C. The presence of an exon in a lariat created during the splicing reaction can also lead to back-splicing of this exon. The exon is first present in a lariat formed by branch point 2 (bp2) in intron 2 of a three-exon substrate. Next a branch-point (bp1) in intron 1 forms a lariat with the guanosine of the 5' splice site (dashed line with arrow), allowing back-splicing to occur (solid line with arrow) (50).

Figure 2-2 Alu elements promote back-splicing through formation of structures in the pre-mRNA



A. General sequence of an Alu element, "A, B": A and B boxes as recognition sites for DNA polymerase III. AAA: short mid A stretch; AAAA (red): terminal A stretch, TTT: downstream T-stretch, acting as a pol III termination site, arrows, yellow direct repeats (genomic Alu insertion site). The common depiction of an Alu element is an arrow that shows the direction from the A box to the terminal A stretch (arrowhead).

B. Sequence of an AluJB element with the elements highlighted from the human MAPT gene (hg38 chr17:46,014,583-46,014,913). FLAM: free left Alu monomer (dashed box), FRAM: free right Alu monomer (solid box). The coloring of the elements is similar to panel A.

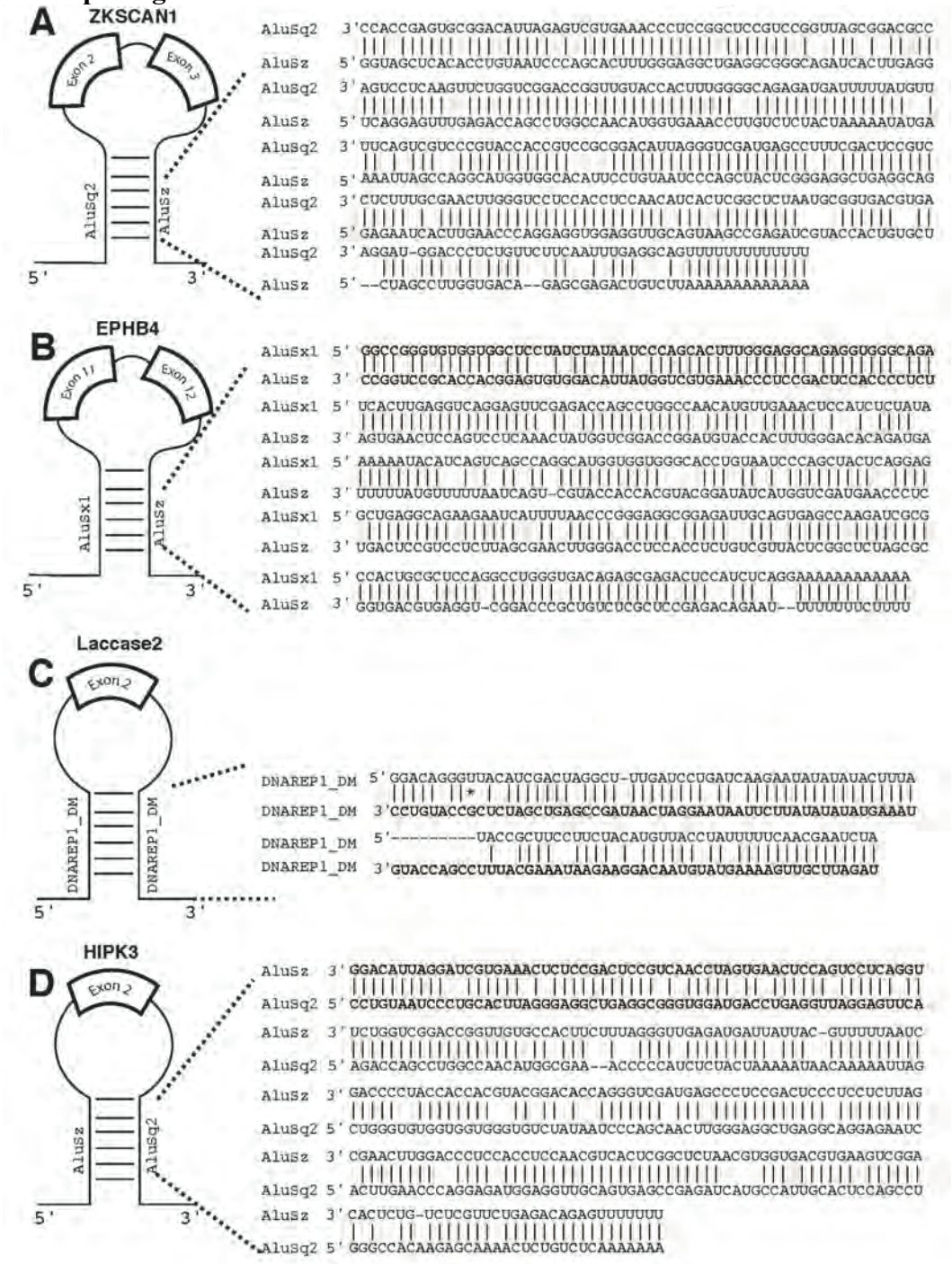
C. Determinants of Alu-element mediated secondary structures. Alu elements oriented in opposite direction can form double stranded regions, measured through editing of the Alu-elements. The probability to form a double stranded structure increases exponentially when the Alu-elements are getting closer.

D. Arrangement of Alu elements in a pre-mRNA can promote alternative circRNA splicing. Depending on the interaction (dotted line) of the Alu elements a circle between exon 1 and 2 or a circle composed of exon 2 (dotted) can be formed.

E, **F**. Removal of an intron through splicing can decrease the distance between dsRNA structure forming Alu-elements. **E**. shows the situation before splicing, where the Alu-elements are too far apart to allow back-splicing. **F**: Intron removal through splicing allows back-splicing and circRNA formation.

G. Elements controlling circRNA formation: hnRNPs and SR-proteins control general splice site recognition and influence competition between linear and circular splicing. Helicases influence the stability of the RNA structures promoting circular splicing, a larger distance between the back-splicing sites, i.e. the length of the circularized exon promotes back-splicing.

Figure 2-3 Experimentally characterized examples of RNA structures leading to back-splicing.



Schematic structures and double stranded region of **A.** ZKSCAN1 (45), **B.** EPHB4 (45), **C.** LACCASE2 (42) and **D.** HIPK3 (45) genes are shown. Experimentally validated double stranded pre-mRNA regions are shown on the right. A "*" indicated G-U base pairing.

CHAPTER 3. PROCESSING OF THE HUMAN *MAPT* LOCUS GENERATES CIRCULAR RNAs

With permission this chapter is adapted from:

Welden, J. R., van Doorn, J., Nelson, P. T. and Stamm S. (2018). "The human *MAPT* locus generates circular RNAs." Biochim Biophys Acta 1864(9 Pt B):2753-60 (1)

3.1 Introduction

The human microtubule-associated protein Tau is highly expressed in brain and promotes the assembly and stabilization of microtubules (77). Tau protein can mis-fold into paired helical filaments and neurofibrillary tangles, which characterize a group of neurodegenerative diseases known as tauopathies, that include Alzheimer's disease (AD), frontotemporal lobar degeneration (FTLD-TAU), progressive supranuclear palsy (PSP), chronic traumatic encephalopathy (CTE), and primary age-related tauopathy (PART) (78). The clearest connection between the *MAPT* gene and neurodegeneration is found in FTLD-MAPT, as the disease is caused by many different known mutations in the *MAPT* locus on chromosome 17 (79). Studies in mice indicate that tau protein is necessary for amyloid-beta induced neuronal cell death (80), and thus plays a central role in AD.

The human MAPT gene contains 16 exons, with exons 2, 3, 4a, 6, 8 and 10 being alternatively spliced cassette exons (Figure 3.1A). Alternative splicing of these exons in the normal adult human brain generates six major protein isoforms. These isoforms differ at the N-terminus due to alternative exons 2 and 3 in the tau projection domain and in the number of microtubule binding repeats due to alternative splicing of exon 10. Exon 10 encodes one of the four microtubule binding sites, and its alternative usage generates tau isoforms with either 3 or 4 binding sites (3R, 4R) that differ in their affinity to microtubules (81) and could thus 'fine tune' the interaction between the Tau protein and microtubules.

At least 19 mutations causing FTLD-MAPT have been identified in exon 10 and its 5' splice site (79). Pathological changes in exon 10 usage without mutations in exon 10 result in tauopathies, for example PSP and corticobasal degeneration (CBD) are characterized by 4R Tau, whereas Pick's disease is characterized by 3R Tau. Post mortem studies of AD brains indicated a slight increase of exon 10 usage (82-84), as well as deregulation of protein factors that regulate exon 10 splicing, namely the SR-like protein tra2-beta1 that promotes exon 10 usage and its kinase, clk2, that inhibits exon 10 usage (82, 85).

In addition to the well known linear RNAs, pre-mRNAs generate circular RNAs through a backsplicing mechanism, where a downstream 5' splice site is joined with an upstream 3' splice site (86). In most cases, circRNAs are generated when the pre-mRNA forms a loop containing the exons undergoing backsplicing. This loop can be formed by either a large lariat or more commonly through intramolecular RNA base pairing, leading

to double stranded RNA regions as short as 30-40 nt in length. Frequently, repeat elements that have regions of base complementarity provide the basis of loop formation. In humans, these elements are often Alu elements (9, 19), comprising around 11% of the human genome (57). Due to their self complementarity Alu elements form extensive double stranded RNA structures in pre-mRNA, which can influence alternative splicing (87) and promote the formation of circRNAs (9). With about 50 highly expressed exceptions, circRNAs comprise only 1-5% of their linear counterparts. Since they lack a 3' or 5' end, they escape the exonucleic degradation of linear RNAs and are thus more stable. The analysis of a few circRNAs showed that they mainly reside in the cytosol, where they can function as microRNA sponges (19), and can undergo translation in the presence of an internal ribosomal entry site (88). In addition, intron-containing circRNAs have been implicated in transcriptional control in the nucleus (89).

Here, we used a PCR approach to identify circular RNAs from the human tau locus that contain the alternatively spliced exon 10. These circular RNAs are low abundant, comprising less than 1% of the linear tau RNA. The data shows that not all RNAs generated from the human tau locus have been identified.

3.2 Materials and Methods

3.3 Primers

9 → 11 Forward: TGT CAA GTC CAA GAT CGG CT

9 → 11 Reverse: CTG GCC ACC TCC TGG TTT

11 → 10 Forward: GAC CTC CAA GTG TGG CTC AT

11 → 10 Reverse: TGG ACT GGA CGT TGC TAA GA

RPL13 Forward: GCC ATC GTG GCT AAA CAG GTA

RPL13 Reverse: GTT GGT GTT CAT CCG CTT GC

Tau 12→10 reverse: cag ctt ctt att aat tat ctg cac ctt tt

Tau 10→11 forward: gag gcg gca gtg tgc aa

HIPK3f: tcg gcc agt cat gta tca aa

HIPK3r: tgc ttg gct cta ctt tga gtt tc

Tau exon12 Rev: ccc aat ctt cga ctg gac tc

Tau exon 9 Forw: tgt caa gtc caa gat cgg ct

3.4 Minigene Generation

The tau exon 9-12 minigene was generated using Gibson cloning by assembling exons 9, 10, 11 and 12 in pcDNA3.1 using these primers:

Vector1 AAGCTTAAGTTTAAACGCTAGCCAGCTTG

Vector 2 CTCGAGTCTAGAGGGCCCGTTTAAACC

Exon 9f ctggctagcgtttaaacttaagcttACGCTGGCCGCAGGGATT

Exon 9r cctgtggatttttGGCCCACGAGTGGAGATGC

Exon 10f ccactcgtgggccAAAAATCCACAGGTGATTCTGATGCC

Exon 10r gagtggggtatctGCGGCAGCCCAGTCTCAG

Exon 11f actgggctgccgcAGATACCCCACTCCTGCCTTTCCA

Exon 11r aacatcctgtaaaccatgaccccacAGTAGCTGGGACTACAGGCG

Exon 12f GTGGGGTCATGGTTTACAG

Exon 12r tttaaacgggccctctagactcgagAAACTGCAGTGACTTAGGCC

3.5 RNA Isolation

Samples were derived from short-postmortem interval (PMI) autopsies. All methods conformed with a University of Kentucky IRB protocol. Premortem clinical evaluations and pathological assessments were as described previously (90). The inclusion criteria that were applied: PMI <4hrs; no evidence of frontotemporal dementia; no cancer in the brain parenchyma; and no large infarctions in the brain, or microinfarcts found within 3cm of the brain tissue samples. We also obtained information on agonal events for each subject, and additional criteria for exclusion from the study were an extended interval of premortem hypoxia, any medical ventilator use, brain edema, or large infarct. The characterization of the samples is shown in Supplemental Figure 3.1. The RNA isolation was performed using Trizol and the PureLink RNA mini kit from Ambion (Life Technologies).

3.6 Cell Fractionation

Cells were harvested through trypsination, washed once with PBS and then resuspendend in Buffer A (10mM HEPES, pH 7.9; 1.5mM MgCl2, 10mM KCl, 0.5mM DTT) and incubated on ice for 10 min. After swelling in buffer A, cells were ruptured in a 0.5 ml Dounce homogenizer using 10 strokes of the tight pestle “B”. Cytosol was recovered by centrifugation at 2000g for 10 min. Nuclei were resuspended in Buffer S1 (0.25M sucrose, 10mM MgCl2) and layered over 4 ml of Buffer S2 (0.35M sucrose, 10mM MgCl2). Nuclei were recovered through 20 min of centrifugation at 3500g.

3.7 Transfection Assays

Transfection assays were performed as described (91). Briefly, DNA was incubated with PEI (Polyethyleneimine, 1 mg/ml, (Sigma), final concentration 3 μg/μg PEI/DNA, per 300,000 cells) for 10 min and then added to HEK293 cells that were 60% confluent in six-well plates. A GFP construct was added to visualize transfection efficiency (> 80%). The

RNA was isolated after 24 h using the PureLink RNA Mini Kit (Ambion, Life Technologies).

3.8 Reverse Transcription Polymerase Chain Reaction

RT-PCR was performed as described (92) using an H– MMLV reverse transcriptase (100 U per reaction, Invitrogen) 500 μM, dNTPs, 0.5–1 μg RNA, 500 nM gene-specific primers. PCR was performed using Platinum Taq Polymerase (Invitrogen) (200 μM dNTPs, 500 nM primer). PCR was performed on a Mastercycler Nexus2 (Eppendorf). Denaturation was at 94° for 30 s, extension at 68° for 30 s. Annealing temperatures and cycles used were: 46°, 20Å~, (SNORD2), 55°, 20Å~ (HIPK3), 55°, 30Å~ (linear tau). Circular tau was amplified using touchdown, starting at 65° lowered per cycle by 0.5° to 58° (14 cycles), followed by 16 cycles at 58°.

3.9 RNase Protection

RNase protection was performed as described (93) using the RPAIII kit (Ambion/Invitrogen) and one million cpm of a 32P uniformly labeled probe. Hybridization was overnight, the digestion used RNaseA and T1 for one hour.

3.10 Results

3.11 Tau generates circular RNAs through exon 12 backsplicing

No human circular *MAPT* RNAs have been reported in databases. Given the extensive alternative pre-mRNA processing of the human tau gene (Figure 3.1A), we used a PCR approach to search for possible human circular MAPT RNAs. We concentrated on exon 10, since this exon is alternatively spliced and deregulated in both Alzheimer's disease and FTLD-MAPT (82, 83). To amplify circular RNAs, we used a reverse primer upstream of the forward primer, i.e. a reverse primer in exon 10 and a forward primer in exon 1, which amplifies circular, but not linear RNAs (Figure 3.1A, C). To further enrich for circular RNAs, we digested the RNA with RNase R, a bacterial RNA exonuclease that removes linear RNAs (84)

As expected, RNAse R treatment removed the signal for linear *MAPT*, detected by primers in exons 9 and 12 and RPL3 mRNA, but enriched the signal for circular RNAs (Figure 3.1B). The bands from RNAse R treated RNA generated by the circular RNA primers were subcloned and sequenced. Two of the bands corresponded to circular RNAs made from exon12 back splicing, either to exon 10 or 7 (Figure 3.1C). The exon sequences present in the linear *MAPT* mRNA were completely present in the circular RNAs, indicating usage of the linear splice sites. Both circRNAs were divisible by 3 (288, and 681 nt, respectively) and circ12→7 contained an in-frame AUG start codon. No AUG start codon was present in circ12→10 (Figure 3.1D, E). The other bands contained non-canonical splice sites and thus their mechanism of generation is unclear, and they could be PCR artifacts.

3.12 A minigene spanning exons 9 to 12 generates circ12→10

We concentrated on circ12→10, as it was the most abundant RNA. To determine the sequences necessary for its formation, we created a minigene consisting of exons 9, 10, 11 and 12, each flanked by 1-2 kb of intronic sequence (Figure 3.2A). This minigene was transfected into HEK293 cells and circular RNAs were detected using PCR exon junction primers selective for the 12→10 backsplice, consisting of 10→12 reverse primer and a forward primer in exon 10 (Figure 3.2B). In addition, I amplified linear RNA, using primers in exons 9 and 12 (Figure 3.1A). The minigene contained the repetitive elements surrounding exon 12 and an intronic repetitive element upstream of exon 9. After minigene transfection, I could detect both the exon 9 to 12 linear RNA containing the alternative exon 10, as well as the circ12→10 backsplice RNA (Figure 3.2C). Next, we tested the expression of circRNAs using a different method. We choose RNase protection analysis that has a linear readout but is less sensitive than RT-PCR. Using a uniformly radioactively labeled probe exhibiting sequence complementarity towards the exon12→10 backsplice junction (Figure 3.2B, D), we could detect a faint signal in RNA from transfected HEK293 cells, which was enriched by RNAse R treatment (Figure 3.2E). In addition, the probe detected parts of exon 10 and 12, which derived from linear RNA indicating that RNAse R treatment does not remove all linear RNAs. A signal of the same length could also be detected from 50 µg of total human cortex RNA. The circular RNA signal was less than 1% of the linear RNA signal in transfected cells, indicating that despite our ability to detect circ12→10 using RT-PCR, this RNA is very weakly expressed (Figure 3.2E). Since our minigene can produce circular RNA, we conclude that all cis-acting sequence elements to generate the tau circ12->10 RNA are present in exons 9 to 12 and their immediate intronic vicinity.

3.13 Tau circ12→10 is regulated by clk2 and sensitive to mutations in exon 10

Tau circ12->10 contains exon 10, which is alternatively spliced in linear MAPT. Several exonic mutations causing FTLD/FTD have been identified that change exon 10 usage (79). In addition, exon 10 is regulated by several trans-acting factors, most notable tra2-beta1/TRA2B(85), promoting its inclusion as well as the cdc2-like kinase CLK2 (82, 85), the SR-protein SRSF7/9G8, protein phosphatase 1 (PP1) and its nuclear inhibiter NIPP1 promoting exon 10 skipping (94-96).

To test a possible regulation, I transfected the tau 9-12 minigene with plasmids expressing EGFP, CLK2, the inactive CLK2 variant CLK2KR, tra2-beta1, DYRK, 9G8, SRPK1, NIPP1 and PP1. The SR-protein kinase SRPK1 (97), was tested as it phosphorylates tra2-beta1. To investigate the effect of an FTLD/FTD mutation on the formation of the circular RNA, we introduced the N279K mutation promoting exon 10 inclusion into the minigene (79).

First, I tested the effect of the trans-acting factors in MAPT linear RNA using PCR primers in exon 9 and 12. CLK2, 9G8, NIPP1 and PP1 caused exon 10 skipping, as expected from previous data using exon 9-12 reporter genes (Figure 3.3A). This

experiment was repeated with a minigene harboring the N279K mutation. This exon 10 mutation (ATAA to AGAA) creates a stronger tra2-beta1 binding site, as tra2-beta1 binds to NGAA sequences (98). The generation of this tra2-beta1 binding site promotes inclusion of exon 10, by promoting the formation of tra2-beta1 dependent splicing enhancer complexes. Importantly, the N279K mutation causes FTLD/FTD (79). Testing the trans-acting factors on this minigene revealed that the effect on exon 10 skipping was strongly reduced for CLK2, 9G8, NIPP1 and PP1.

RNA from these experiments was next tested for the expression of the tau12→10 circRNA. CLK2 and 9G8, NIPP1 and PP1 strongly reduced the expression of this circRNA. A similar effect was observed when I used the N279K minigene. The trans-acting factors had no detectable effect on the HIPK3 circular RNA (Figure 3.3C), and I thus quantified our data by calculating the ratio between circtau12→10 and circHIPK3. The analysis of three independent experiments showed a statistically significant reduction of circtau12->10 caused by 9G8, NIPP1 and PP1 in both minigenes and by CLK2 in the wild-type minigene context (Figure 3.3D).

Finally, I investigated possible synergistic effects between the SR-proteins and their kinases and cotransfected CLK2 and SRPK1 together with tra2-beta1and 9G8 expression clones. SRPK1 slightly increased the effect of tra2-beta1 and 9G8 on tau 12→10 circRNA formation in the wild-type exon 10 context and had no detectable effect on the N279K mutations. In contrast, clk2 reduced the effect of tra2-beta1 in both wild-type and N279K background, whereas it had no effect on 9G8 (Figure 3.3E).

The data indicate that the abundance of circtau12→10 RNA can be regulated by the cell through trans-acting factor expression, whose actions is in turn regulated through their kinases. In addition, an FTLD/FTD mutation that promotes exon 10 inclusion also influences tau circ12→10 RNA formation.

3.14 Tau circ12→10 RNA is predominantly localized in the cytosol

I detected tau12->10 circRNA in the human neuroblastoma cell line SH-SY5Y and identified the endogenous RNA's cellular localization. I employed cell fractionation using different lysis of plasma and nuclear membranes followed by RT-PCR. The tau circ12->10 RNA was exclusively localized in the cytosol (Figure 3.4), similar to other circRNAs analyzed (9). circHIPK3 RNA showed a similar cytosolic localization, whereas the C/D box snoRNA SNORD2 was exclusively nuclear.

3.15 No significant correlation between circ12→10 expression and Braak stages

Tau exon 10 and Clk2 splicing is deregulated in temporal cortex in sporadic Alzheimer's disease (82, 85). I therefore tested brain samples from 15 subjects for a correlation between tau circ10→12 circRNA and Braak stages (Supplemental Data 3.1, Figure 3.5A-C). RNA from the gray matter of Superior and middle temporal gyri was used. The amount of tau circ12→10 was normalized to circular HIPK3. Although I found expression of tau circ 12→10 in all subjects, there was no statistically significant

correlation with Braak stages, when the circ12→10 RNA was normalized to circHIPK3 RNA (Figure 3.5D). For the statistical analysis, only RNAs from samples with RIN numbers larger than seven were used. In addition, we tested these samples for expression of the kinase clk2 using primers spanning its alternative exon 4. Skipping of this exon generates an inactive kinase isoform (99). Again, there was no correlation between the amount of full-length clk2 and tau12→10 circRNA (Supplemental Figure 3.2).

3.16 Discussion

3.17 *MAPT* as a contributor to human disease

The MAPT gene and its polypeptide products are acutely relevant to human diseases. A deregulation of MAPT is involved in Alzheimer's Disease, Chronic Traumatic Encephalopathy, Frontotemporal Lobar Degeneration-tau, and many other 'tauopathies' (2).

3.18 Circular RNAs from the *MAPT* locus could be species specific

Here we present evidence that the human *MAPT* locus generates circular RNAs through backsplicing from exon 12 to either exon 10 or 7. Our data show that the regulation of *MAPT* gene expression and its molecular biology are still incompletely understood, despite numerous studies focusing on alternative splicing of *MAPT* (100).

We concentrated on circ12→10, as it is the most predominant circular RNA. Currently, human circular tau RNAs have not been reported in database. For mouse, a backsplice from exon six to four has been identified (101, 102). We could not amplify the human orthologue for this mouse circ6→4 RNA, nor could we identify mouse orthologues for our described 12→10 and 12→7 circRNAs. It is thus possible that *MAPT* generates species-specific circRNAs, similar to exon 10 usage that is alternatively spliced in adult humans, whereas it is constitutively used in adult mouse (100).

CircRNA formation through backsplicing is facilitated by repetitive elements that form regions of complementarity in the pre-mRNA, allowing to position exons for backsplicing (9). Human *MAPT* contains at least 83 Alu elements, 56 on the sense strand and 27 on the antisense strand (103). Since Alu elements are primate-specific, they could cause a difference in circular RNA formation between mouse and humans.

We generated a minigene consisting of exons 9 to 12 with about 2 kb flanking intron each. I found that this construct forms circtau12→10 RNA when transfected into HEK293 cells, indicating that these regions are sufficient to generate the circular RNA. Importantly, in this construct, exon 10 is not surrounded by repetitive elements, suggesting that far-distance interaction, for example between repeats near exon 12 and 9 present in our minigene could form the basis for circRNA formation.

3.19 Possible functions of tau circRNAs

To gain insight into the possible function of tau circ12→10, I first determined its cellular localization in human SH-SY5Y neuroblastoma cells. Similar to other circular RNAs (9), tau circ 12→10 is almost exclusively cytosolic. Inspection of the sequence of tau circ 12→10 and circtau12→7 showed that both RNAs contain a number of nucleotides divisible by three. Tau circ12→10 contains a reading frame without a stop or start codon. This reading frame is identical to a portion of the tau protein containing the microtubule binding site encoded by exon 10. Despite the absence of an AUG codon, it is conceptually possible that translation could occur, for example one of the two in frame AUA triplets could be edited by ADAR1 or 2 to AUI, where the inosine (I) is read as a G by the initiator tRNA, leading to protein synthesis. Further, repeat-associated non-ATG (RAN) translation (104) from CAGs that are also present in tau circ 12→10 are possible and finally a translation of the RNA could theoretically occur in mitochondria, that recognize AUA start codons (105). Initiation at any of these non-AUG start codons could generate a tau protein fragment containing a microtubule binding site. Since the circtau12→10 RNA is divisible by three, it could be translated through a rolling circle mechanism that has been shown for model circular RNAs (106, 107). A similar translational mechanism is possible for tau circ12→7 that contains an in frame start codon, but no stop codon. However, both circular Tau transcripts are very weakly expressed. Thus, their translation will be a rare event. If translation occurs, the expected product would be a high molecular weight tau multimer containing several microtubule binding sites.

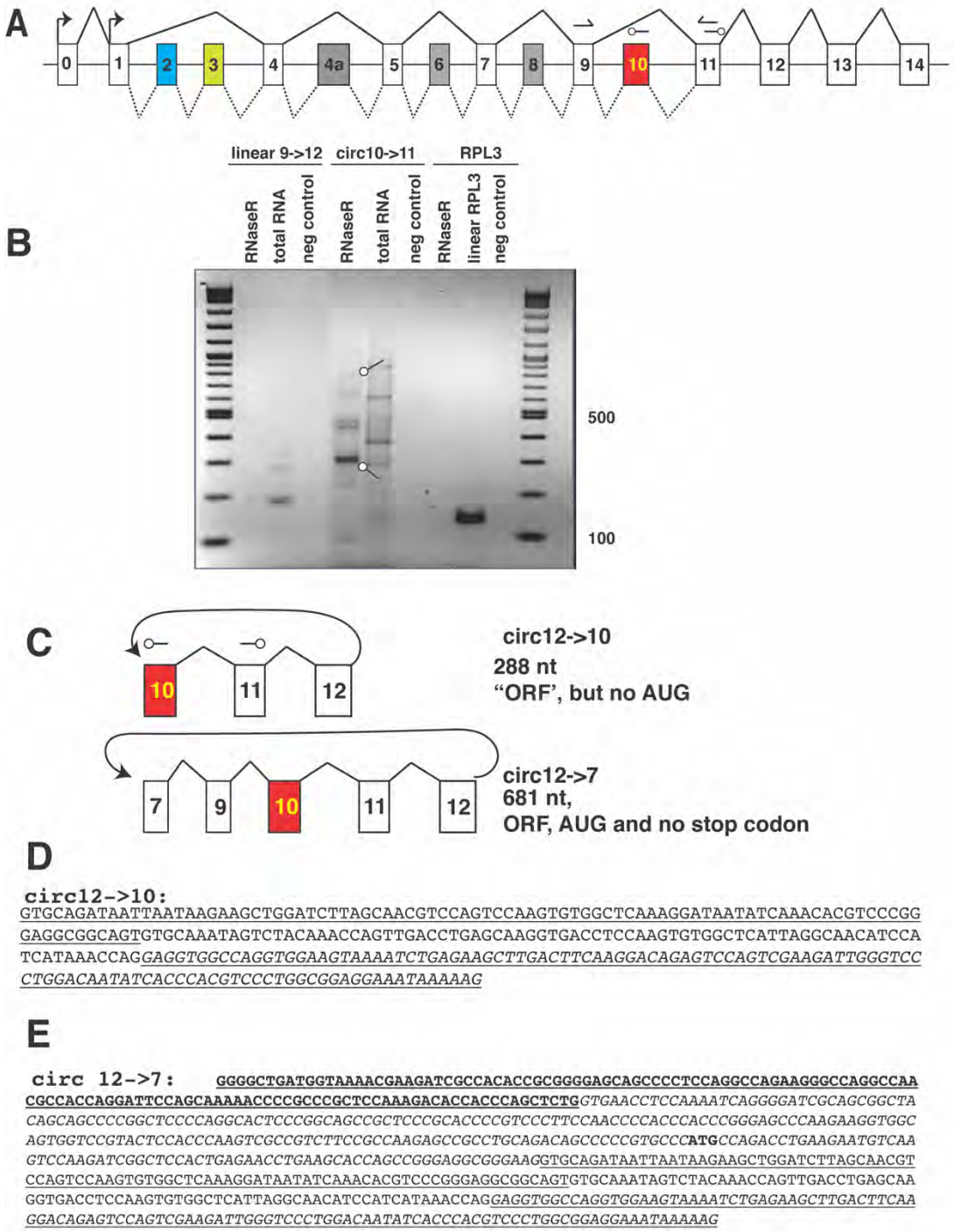
3.20 Relation to disease

The alternative splicing regulation of tau exon 10 has been extensively studied, due to exon 10's involvement in human disease. Mutations that interfere with the inclusion ratio of tau exon 10 cause frontotemporal dementia (79) and exon 10 usage as well as CLK2 splicing isoforms are changed in Alzheimer's disease (82). We thus tested the expression of circtau12->10 in AD brains of various Braak stages, using gray matter of the superior and middle temporal gyrus (SMTG), but did not find a statistically significant correlation between tau circ12->10 expression, normalized to circHIPK3. However, the translation of circular RNAs depends on the creation of an internal ribosomal entry site likely caused by RNA methylation (108) and thus the biological effect of the tau circRNAs might be correlated with modifications and not abundance. In addition, our sample number was low and we thus cannot rule out a connection with AD.

In a minigene, we re-created one FTDP-17 mutation that strengthens exon 10 usage and found that it generates an amount of tau circ12->10 RNA similar to the wild-type in transfection assays, indication that mutations causing frontotemporal dementia can also form circular tau RNAs.

In summary, our data show that the MAPT locus can generate circRNAs that could play a role in neurodegenerative diseases, which warrants further investigation.

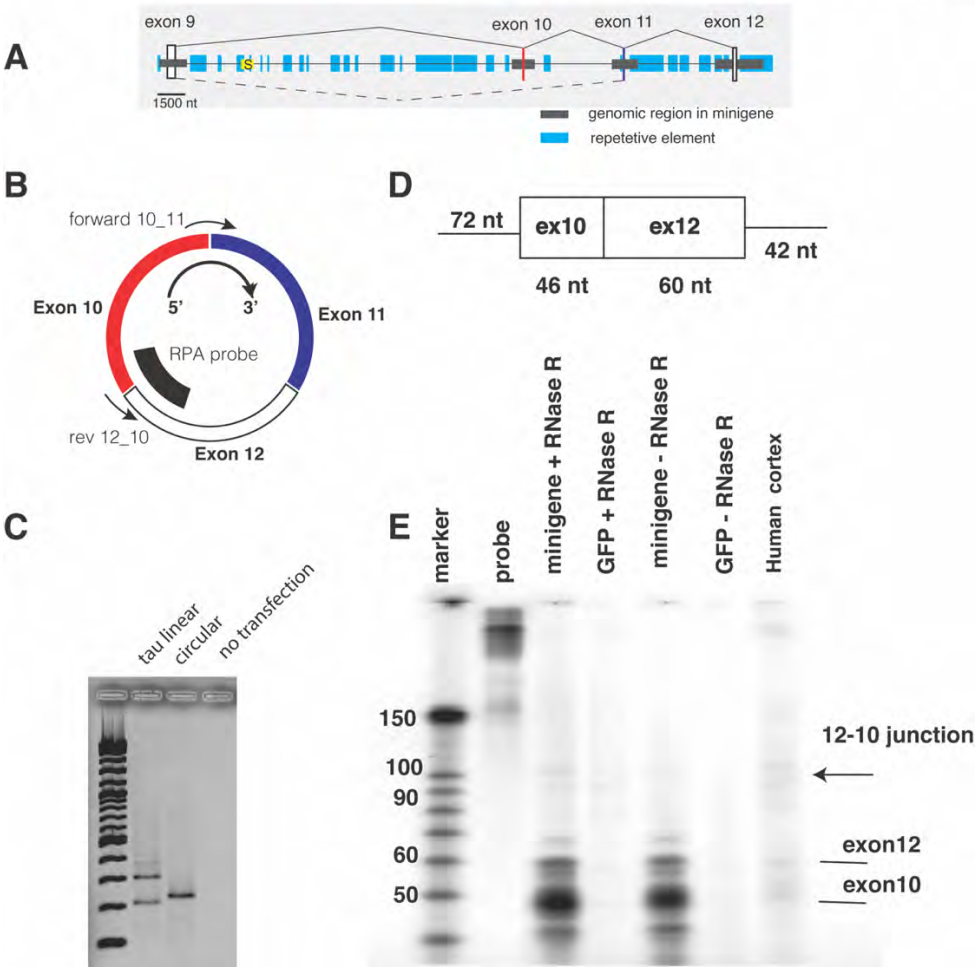
Figure 3-1The human MAPT gene generates circular RNAs through exon 12 backsplicing



A. Schematic structure of the human MAPT gene
Constitutive exons are in white, alternative exons are colored. Splicing patterns are indicated by lines. The two transcriptional start sites in exons 0 and 1 are indicated by arrows. The numbering of exons is from (100). The arrows in exons 9 and 11 indicate the

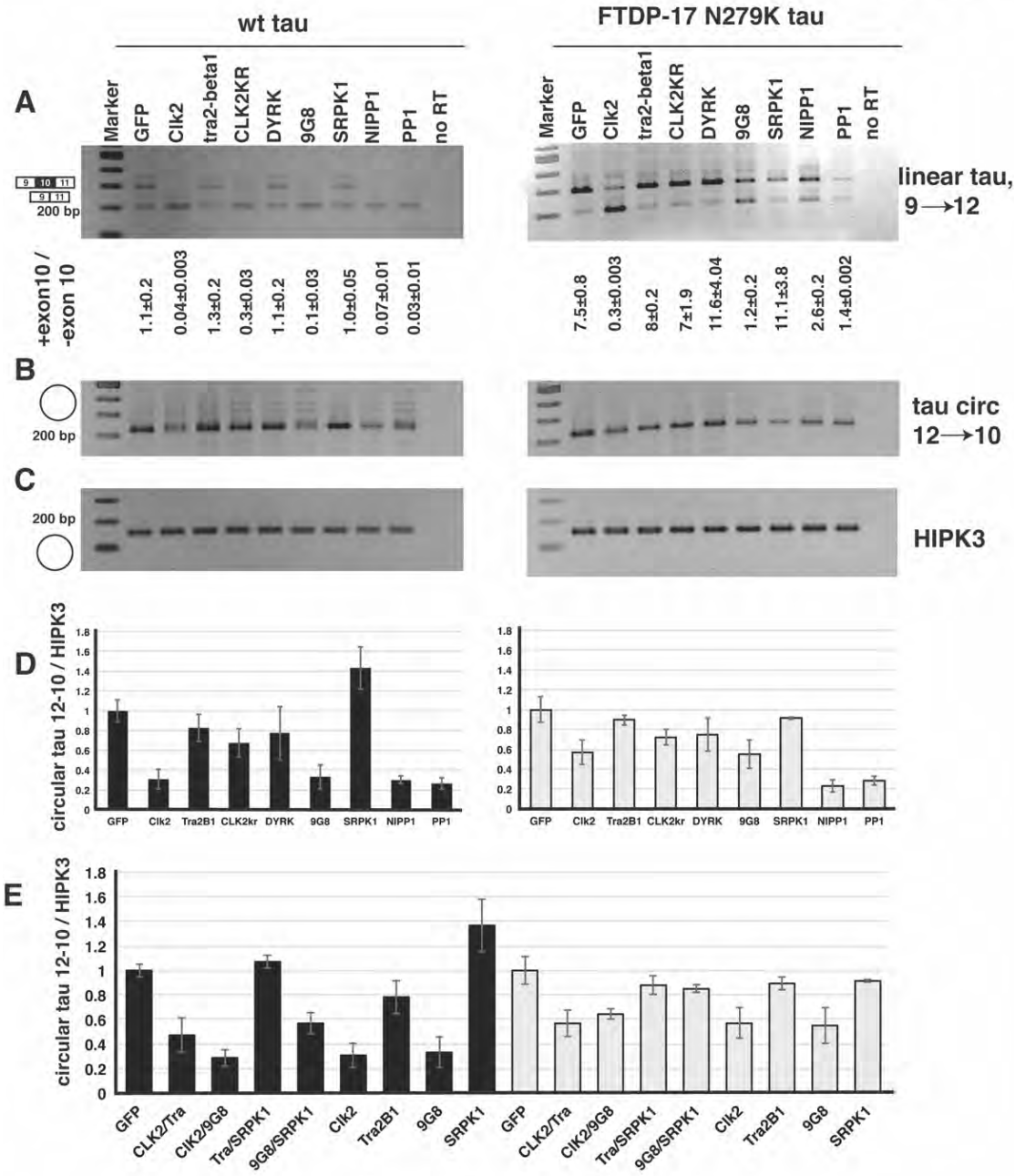
position of primers to amplify linear MAPT mRNA, arrows in exons 10 and 11 indicate primers to detect circular RNAs. The arrowheads indicate the 3' ends.
B. Circular RNAs generated through backsplicing of exon 12
10 µg of total human hippocampus RNA was amplified with primers pointing outwards in exon 10 and 11. Prior to RT-PCR, an equal part of the RNA was digested with RNase R. The negative control was performed without reverse transcription. The PCR-products were gel purified, subcloned and sequenced. Two bands indicated by arrows were circular RNAs generated through backsplicing of exon 12. The other bands were PCR artifacts, i.e. amplicons from different gene regions. *This figure was produced by Jay van Doorn.
C. Structure of the RNAs generated through exon 12 backsplicing.
D. Sequence and predicted ORFs for circ12->10 The RNA part corresponding to exon 10 is underlined, exon 11 is in regular letters and exon 12 is underlined and italic.
E. Sequence and predicted ORFs for circ12->7. Exon 7 is bold, exon 9 italic, exon 10 is underlined, exon 11 is in regular letters and exon 12 is underlined and italic. The in frame AUG is shown in bold.

Figure 3-2 A minigene containing exons 9-12 generates circ12→10



A. Structure of the minigene. Exons are indicated as vertical boxes, horizontal gray boxes indicate the genomic regions used for cloning the minigene. Blue boxes indicate repetitive elements defined by the UCSC genome browser repeat masker. The yellow box with an 'S' depicts the intronless Saitohin reading frame. The drawing is to scale, as indicated.
B. Detection of circ12->10. The orientation of the circular RNA is clockwise 5'->3'. The location of the detection primer for10_11 and rev12_10 is indicated. The probe used for RPA is shown as a bold line.
C. Detection of RNAs made from the exon 9-12 minigene. 1 µg of the minigene was transfected into HEK293 cells and after 24 hrs, RNAs were detected by RT-PCR. The negative control is untransfected HEK293 cells, using circRNA primers.
D. RNase protection probe to detect circ12->10 RNA. The T7 antisense RNA is shown with the nucleotide lengths indicated.
E. RNAse protection using RNA from transfected cells as well as human cortex. 50 µg total RNA was used, which was digested with RNaseR, as indicated. *Dr. Stefan Stamm produced this figure.

Figure 3-3 The abundance of exon 10 containing circ RNA is regulated by the kinase clk2

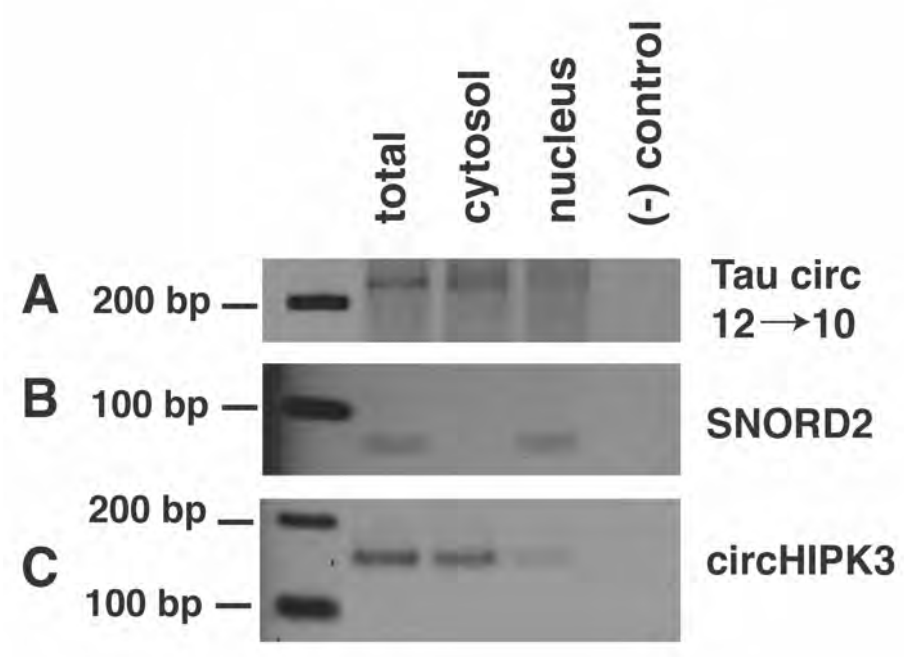


One µg of the tau exon 9-12 minigene was transfected into HEK293 cells together with 1 µg of the plasmids expressing the trans-acting factors indicated.

A. Linear RNA was amplified using primers in exon 9 and 12
B. Circular RNA was amplified using primers rev12-10 and forward_exon 10.
C. The circular HIPK3 RNA was amplified as a loading control
D. Quantification of the circtau to circHIPK3 ratio

E. Quantification of a cotransfection of the SR-proteins tra2-beta1 and 9G8 in the presence of their kinases CLK2 and SRPK1. 500 ng of each expressing vectors were cotransfected and three independent experiments were analyzed.

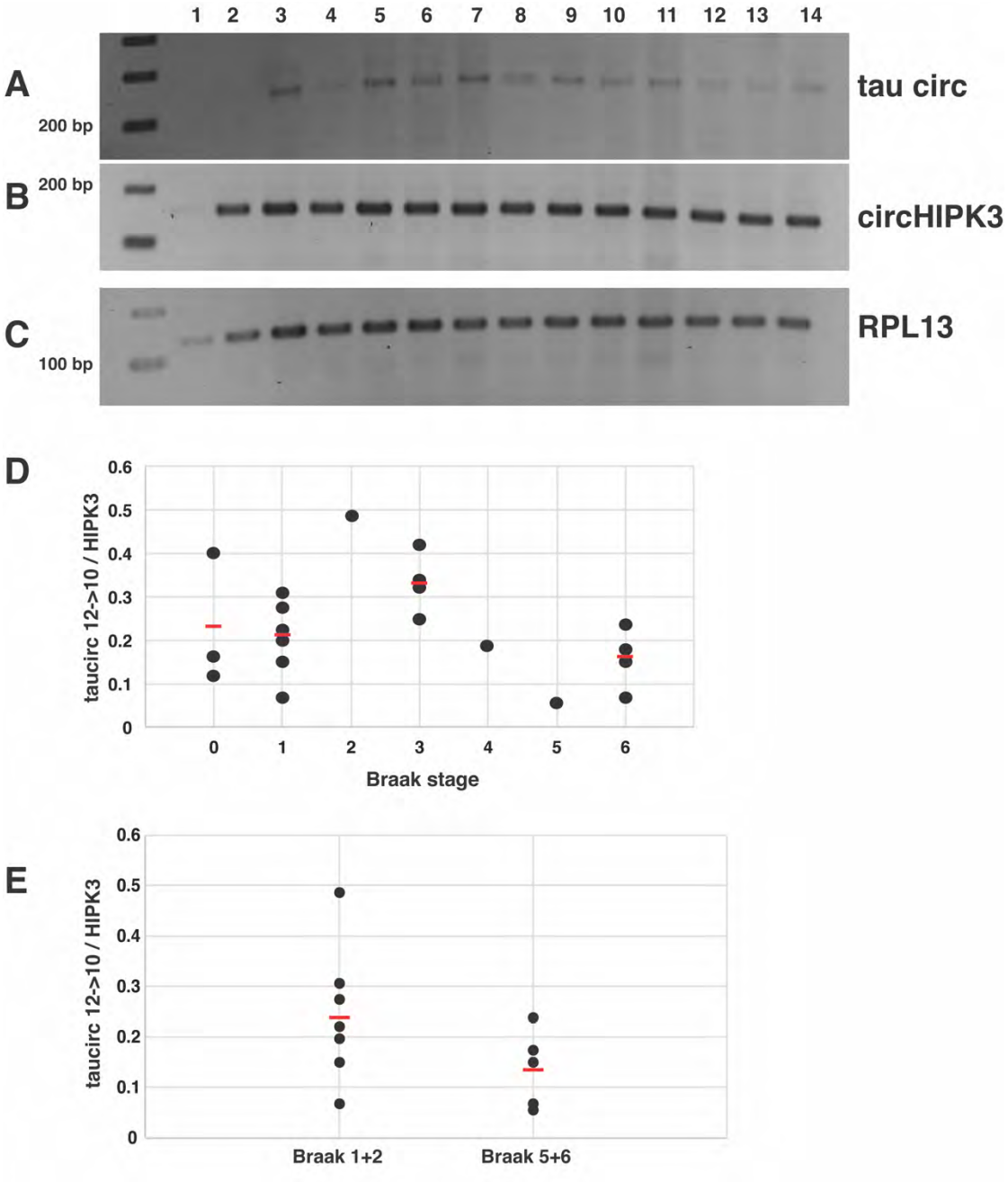
Figure 3-4 Tau circ12→10 is cytosolic.



SH-SY5Y cells were separated into cytosol and nucleus and RNA isolated from each fraction.

A. Detection of circ12→10
B. Detection of SNORD2, a C/D box snoRNA with nuclear localization
C. Detection of circHIPK3, a circular RNA known to be cytosolic.

Figure 3-5 The ratio between tau circ12→10 and circHIPK3 differs between individuals



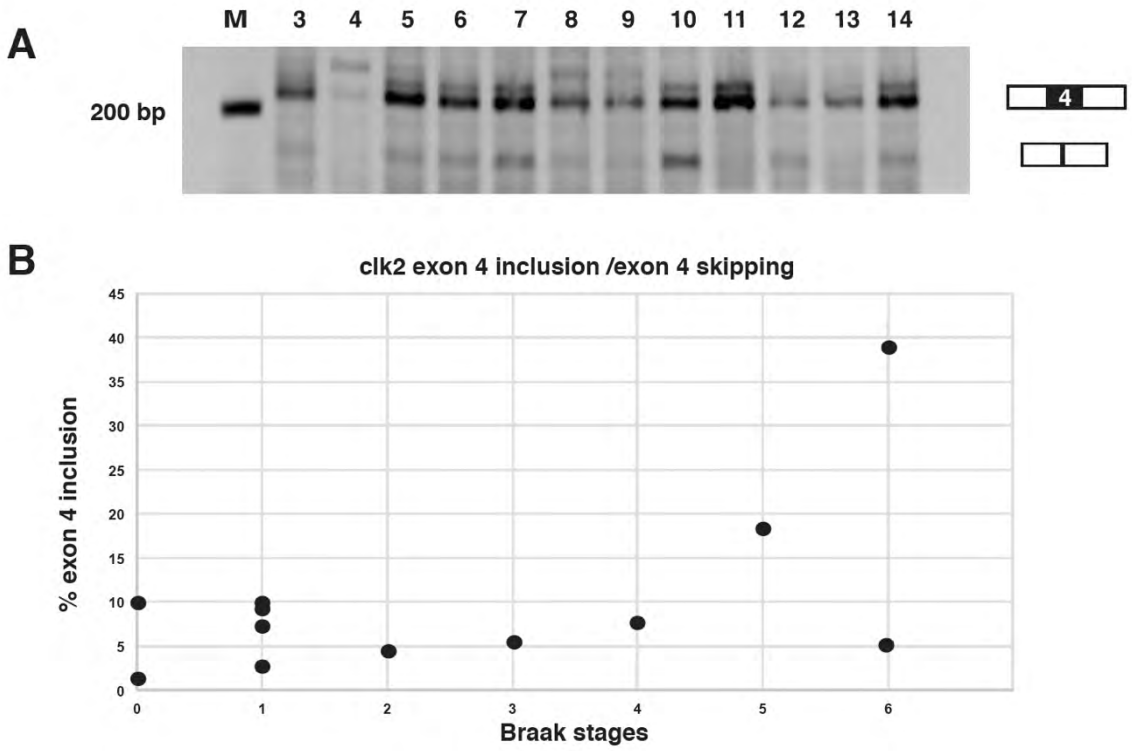
Temporal cortex from 14 patients with Braak stages 0 to 6 were analyzed by RT-PCR amplifying (**A**) cricTau12->10 and (**B**) the abundant circHIPK3 RNA, as well as (**C**) linear RPL3 mRNA. (**D**) shows the quantification of tau circ12->10 normalized to HIPK3. There are no statistically significant changes between the Braak stages. (**E**) Summary of tau circ12→10 abundance in Braak stages 1 and 2, compared with Braak stages 5 and 6. Red lines indicate the averages.

Supplemental Figure 3-1 Clinical Specimens used.

Number	Braak Stage	PMI (h)	age	gender	RIN
1	6	3	93	f	2.6
2	6	3.3	84	f	2.6
3	4	3.5	87	m	8.9
4	2	2.4	87	m	2.7
5	1	10	76	f	9
6	6	2.8	90	f	8
7	3	2.9	82	f	8.1
8	1	1.8	86	f	4
9	1	4.1	90	f	7.1
10	0	2.4	84	f	8
11	6	2.8	84	m	9.2
12	1	2.9	84	f	8
13	5	2	99	f	7.7
14	0	3.5	95	f	8.4

RIN: RNA integrity number
PMI: postmortem interval
Specimens were received by Dr. Peter Nelson from the Sanders Brown Center of Aging, University of Kentucky Brain Bank.

Supplemental Figure 3-2 CLK2 splicing abundance



Amplification of RNA using primers in exons 3 and 5 of clk2. The primers amplify two alternative splice variants, a full-length form containing exon 4 encoding an active kinase, and a truncated form, lacking exon 4. There was no statistically significant correlation between exon 4 usage or total full-length clk2 abundance with tau circ12 → 10. *This Figure was produced by Jay van Doorn.

CHAPTER 4. MOLECULAR AND BIOLOGICAL APPROACH TO ANALYZE CIRCULAR RNAS

With permission from JoVE and the Authors this chapter is adapted from:

Welden, J. R., Pawluchin, A., van Doorn, J., Stamm, S. Use of Alu Element Containing Minigenes to Analyze Circular RNAs. J. Vis. Exp. (157), e59760, doi:10.3791/59760 (2020)." (5)

4.1 Summary

This chapter describes a method to clone and analyze reporter genes generating circular RNAs. These reporter genes are larger than constructs to analyze linear splicing and contain Alu elements, which is challenging for cloning. To investigate the circular RNAs, the constructs are transfected into cells and resulting RNA is analyzed using RT-PCR after removal of linear RNA.

4.2 Introduction

4.3 Circular RNAs

Circular RNAs (circRNAs) are covalently closed single stranded RNAs that are expressed in most organisms. They are generated by joining a downstream 5' splice site to an upstream 3' splice site, which is called back-splicing (Figure 4A) (9). Sequences in the pre-mRNA that exhibit base complementary as short as 30-40 nt bring back-splice sites into proper alignment for circRNA formation (33). In humans, Alu elements (9), representing about 11 % of the genome (57), form extensive double stranded RNA structures in pre-mRNA due to their self-complementarity (64, 109) and thus promote the formation of circRNAs (9).

Currently, three major functions of circRNAs have been described. Some circRNAs bind microRNAs (miRNAs) and through sequestration act like miRNA sponges (19). CircRNAs have been implicated in transcriptional and post transcriptional regulation, through competition with linear splicing (110) or modulation of transcription factor activity (89). Finally, circRNAs contain short open reading frames and proof of principle studies show that they can be translated (22, 111). However, the function of most circRNAs remains enigmatic. The majority of circular RNAs have been detected using next-generation sequencing methods (112). Detailed analyses of individual genes using targeted RT-PCR approaches reveal that a large number of circular RNAs remains to be discovered (44).

4.4 Use of reporter genes to analyze pre-mRNA processing

The analysis of mRNA derived from DNA reporter constructs transfected into cells is a well-established method to study alternative pre-mRNA splicing, which can be applied to circular RNAs. In general, the alternative exon and its surrounding introns and

constitutive exons are amplified and cloned into a eukaryotic expression vector. Frequently, the introns are shortened. The constructs are transfected into eukaryotic cells and usually analyzed by RT-PCR (113, 114). This approach has been extensively used to map regulatory splicing sites and trans-acting factors in co-transfection experiments [reviewed in (113, 115-118)]. In addition, the generation of protein-expressing minigenes allowed for screening of substances that change alternative splicing (119, 120).

The method has been applied to circular RNAs. Currently, at least 12 minigenes backbones have been described in the literature and are summarized in Table 1. With the exception of tRNA based expression system (121, 122), they are all dependent on polymerase II promoters. Here, we describe a method to generate human reporter minigenes to determine cis and trans-acting factors involved in the generation of circular RNAs. An overview of the method using sequences of a published reporter gene (1) is shown in (Figure 4.1).

4.5 Protocol

4.6 Design of the constructs

Use the UCSC genome browser (123) to identify repetitive elements necessary for circular RNA formation and incorporate them in the constructs. Importantly, primers for amplification need to be outside the repetitive elements.

Paste your circular RNA sequence (Supplemental Figure 4.1 is a test sequence) into https://genome.ucsc.edu/cgi-bin/hgBlat?command=start and make sure you select the right organism. Submit the sequence and go to browser view, zoom out 1.5-fold or as appropriate (Figure 4.2A). The search sequence appears in the top line (Figure 4.2A, 1). Depending on the order of the exons in the circular RNA, BLAT will not connect all the exons. In this example, exon 12 (Figure 4.2A, 4) is not connected to 11 (Figure 4.2A, 2, 3), because exon 12 is upstream of exon 11 in the circular RNA sequence (Supplemental Figure 4.1). The repetitive elements are in the 'repeat masker' track, indicated by boxes, where black to gray color indicates the evolutionary conservation (Figure 4.2A, 5).

Mouse over the repetitive elements to identify their subtype in a floating window. Alu elements are in the SINE (short interspersed nuclear element) line. Use the 'default tracks' button under the window to reset your browser if you obtain a different picture than Figure 4.2.

Mousing over the exons in the gene display generates a window with exon numbers that are computer generated. These numbers do not correspond to the exon numbering established in the literature and also change between isoforms.

4.7 Select the sequence to be cloned in an expression vector

Download the DNA sequence shown in the window by going to view → DNA in the top line of the UCSC genome browser. In the 'sequencing formatting option' select 'extended case/color options'.

Select default case: lower and select toggle case for 'NCBI refseq'. Select underline, and bold, and italic for 'repeat masker'. Click submit, you will see the exons as capital letters and introns as small letters. Check the exon/intron borders.

In this example there is a 'ccctttacCTTTTT' sequence, indicating that the browser shows the reverse complement. If this is the case, go back and select the reverse complement box until you see the correct exon intron borders (agEXONgt), in this example AAAAAGgtaaaggg.

Copy the file with the correct orientation (internal exons are surrounded by intronic ag…gt) into a word processing document and highlight the exons (Supplemental Figure 4.2).

In this example, the genomic fragment encompassing exons 9-12 is around 24 kb and thus too large to be amplified from genomic DNA. Therefore, each exon surrounded by about 1 kb intronic region is individually amplified and these four fragments assembled in a cloning vector. Select fragments to be amplified (exon +/- 500 nt intron). Make sure that the intron does not begin or end in a repetitive region, as primers in these regions will not amplify specific sequences. The selected regions are shown in Figure 4.2B, their sequences are shown in Supplemental Figure 4.3.

Note: In general, the larger the constructs, the more difficult the cloning will be. The fragments can be assembled either step wise, i.e. exon 9 is combined with the vector and in the next step exon 10 is introduced into this construct via cloning until all exons are in place, or alternatively, all fragments are assembled simultaneously. A step-wise approach always works, but requires more time. A simultaneous assembly does not always work and depends how well the individual fragments can be amplified from genomic DNA and on the overall size of the construct. We therefore usually start with both approaches simultaneously.

4.8 Design primers for cloning

Use a web tool (https://nebuilder.neb.com/#!/) to design the primers for cloning. For example, enter fragments 9, 10, 11 and 12 and the vector sequence (Supplemental Figure 4.3) to this tool.

For the vector sequence, the insertion site is added as the last nucleotide and subsequently the fragments are added. Since the vector numbering does not start with a given insertion site, the site of insertion in the vector is located and the downstream part is put in in front of the upstream sequence. In this example the inserts start directly after the HindIII [AAGCTT] site and ends directly after the PmeI [GTTTAAAC] site of pcDNA3.1.

The sequence from cccgctgatcag….. until ccgtaaaaaggccgc is pasted before position 1 in the pcDNA3.1 sequence (gttgctggcgtttttcc….).

Adjust primers if their melting points are more than 4 ºC apart and do not work in amplification. The assembly of the fragments and primer sequences designed are shown in Supplemental Figure 4.4. Primers for cloning can also be designed manually ((124)).

4.9 PCR and amplicon detection

Standard PCR Reaction: Make a reaction mix for total volume of 50 µL per reaction, below is for one reaction using polymerase 1

10 µL 5X Reaction Buffer; 1 µL 10 mM dNTPs; 2.5 µL 10 µM Forward Primer; 2.5 µL 10 µM Reverse Primer; 0.5 µL Polymerase 1; *Optional 10 µL 5X GC Enhancer if product has high GC content, you can make a separate mix containing GC Enhancer to test PCR reaction; 32.5 µL Nuclease free H2O

Aliquot 49 µL of the mix into PCR tubes per reaction sample

Add 1 µL DNA to the PCR, the amount ranges from 10 pg to 1 ng.

Spin down the samples to remove residue off the sides and place them in a PCR machine; It is important to use the same machine, as well as, same spots in the machine when optimizing the PCR conditions.

4.10 Optimization for longer DNA fragments for use of different polymerases

4.11 Temperature

Long range Polymerase 2, are optimized as follows:
Lower denaturation temperature (Polymerase 1: 98 ºC, Polymerase 2: 94 ºC)
Longer denaturation time (Polymerase 1: 10 s, Polymerase 2: 30 s)
Longer annealing time (Polymerase 1: 30 s, Polymerase 2: 60 s)
Longer extension time (Polymerase 1: 30 s/kb, Polymerase 2: 50 s/kb)
Lower extension temperature (Polymerase 1: 72 ºC, Polymerase 2: 65 ºC)
Annealing temperature 5 ºC below Tm of the primers
Optimize primer concentrations (five times less to five times more than the original primer concentration)
DNA concentrations were optimized from 10 pg to 50 pg, 100 pg, 1 ng, 5 ng, 10 ng
Example of a PCR program to amplify a 15 kb product size using Polymerase 2
Initial denaturation at 94 ºC for 30 s,
DNA is denatured at 94 ºC for 30 s followed by annealing at 58 ºC for 30 s (Ta specific for our primers determined by web-based temperature calculations that are specific for each polymerase) and extended at 65 ºC for 12 min 30 s *Extension times may vary and need to be optimized if fragments are larger than 10 kb.
Final Extension was at 65 ºC for 10 min with a final 4 ºC hold.

4.12 Extension times and DNA concentrations

Longer extension times for DNA fragments over 6 kb should be 1 min / 1 kb Longer fragments should use less DNA. Do dilutions of DNA to find optimum DNA concentration for amplification, usually 1 pg to 1 ng for plasmids or 1 ng to 1 µg for genomic DNA.

Note: For most cloning use polymerase 1, engineered by fusing the Sso7d DNA binding domain to a proprietary thermostable DNA polymerase(125). The polymerase has a low error rate and due to the Sso7d domain a high processivity, needed to amplify large (10-15 kb) genomic fragments. Due to this large fragment size, enzymatic assembly of DNA molecules (126) is used for insertion into vectors, as this method does not require restriction enzymes. The amplification of fragments longer than 15 kb gets increasingly difficult with Q5 polymerase. For large fragment amplification use polymerase 2 made from pyrococcus-like proofreading polymerase fused to Sso7d or the Long range PCR kit that gives, however, higher error rates.

4.13 Purification of PCR products and cloning

Run ½ of the PCR products on 1% agarose gels containing 1x Gel green. Gel-green intercalates into DNA, similar to ethidium bromide, but it is excited by light around 500 nm (cyan), a wavelength that does not damage DNA, which highly improves cloning. The visualization of gel green stained gels is on a Dark Reader Transilluminator.

In parallel run ½ of the PCR products on a 1% gel that is stained post-run with ethidium bromide as Gel green stained DNA does not run true to size. (Figure 4.3A, B).

Excise bands of the right size from the Gel-green stained gel and isolate DNA using a gel and PCR cleanup kit. In deviation from its standard protocol, the DNA is eluted in only 20 µL of double distilled water, as usually the DNA concentrations are low.

Prior to cloning, the isolated DNA is checked on an agarose gel for concentration and integrity (Figure 4B). Aim for a 2:1 molar insert to vector ratio. When using several inserts, the ratio is 2:2:2:1.

4.14 DNA assembly and clone detection

Cloning is done using a enzymatic DNA assembly kit, with minor modifications. The assembly is performed for 60 min at 50 ºC and generally the lower range of DNA is used for assembly (20-100 fmol/20 µL reaction). The whole reaction mix is added to chemical competent cells and the whole cells plated out on a 6 cm agar plate.

Combine the vector and insert in a 1:2 molar ratio (20-500 fmol each) in 10 µL of water.
Add 10 µL DNA Assembly Master Mix.
Samples are then incubated for 60 minutes at 50 ºC.
Next, transform competent cells with the total assembly reaction. Here, E.coli strain 1 cells are used for shorter constructs or E. coli strain 2 cells for longer or unstable constructs. Cells should be in 50 µL volume.

Thaw cells on ice and add 2 µL of the chilled assembled product to the competent cells. Mix by gently flicking tube 4-5 times. Do not Vortex.
Let mixture sit on ice for 30 min.
Heat shock at 42 ºC for 30 s. Do not mix.
Transfer tube back on ice for 2 min.
Add 950 µL of room-temperature SOC Media to the tube.
Incubate the reaction tube at 37 ºC for 60 min. Shake vigorously (300 rpm)
Warm selection plates with the appropriate antibiotic during incubation at 37 ºC.
Pellet the cells through centrifugation (10,000 g, 30 seconds) and plate out on ¼ and ¾ of cells two selection plates and incubate at 37 ºC overnight.

4.15 Validation of clones

Use colony PCR, employing PCR primers spanning the assembly sites (Figure 4.1C) for detection.

Take a sterile toothpick and touch a single bacterial colony.

Touch the bottom of a PCR tube with the toothpick that has the colony and then streak the toothpick on an antibiotic agar plate, incubate the streaked plate at 37 ºC or room temperature for sensitive constructs overnight.

Overlay the PCR tube touched with the colony with a PCR mix containing the detection primers. These are designed using primer 3 (http://bioinfo.ut.ee/primer3-0.4.0/). The program has the option to force primer design across a defined sequence using the "[ccc]" signs to select primers across the assembly junction. DNA from positive strains is isolated and verified by sequencing.

4.16 Analysis of circular RNA expressing reporter genes

For analysis, reporter genes are transfected into eukaryotic cells. Here, routinely HEK293 cells (ATTC #CRL-1573) are used as they give high transfection efficiency. To save costs, lipofectamine is substituted with PEI (Polyethyleneimine) solutions.

PEI solution: Dissolve Linear Polyethyleneimine Max (PEI) (Polysciences, 24765-1) at 1 mg/mL in water at low pH, pH 2, bringing pH up to 7 with NaOH. Sterile filter with 0.22 µm filters and store at 4 ºC.

Split cells into six wells (approximately 150,000 cells per well) and let them grow over night in 10% FBS in DMEM media (Sigma, D5796);

Aliquot 1 µg of reporter gene in a sterile tube and add 200 µL of sterile filtered 150 mM NaCl, mix with the DNA by vortexing;

Add PEI solution to this mix and vortex, briefly centrifuge to collect samples at bottom of tube. Use a ratio of 1 µg DNA/3 µL PEI;

Incubate at room temperature for 10 min and then add directly to HEK 293 cells;

Incubate HEK293 cells at 37 ºC, 5% CO2, overnight;

RNA is then isolated for RT-PCR via an RNA isolation kit.

4.17 RNase R treatment to remove linear RNAs

Use 10 µg of total RNA in an RNase-free tube

Add 10 µL 10x RNase R buffer to RNA (0.2 M Tris-HCl (pH 8.0), 1 M KCl, 1 mM MgCl2);

Add RNase R to RNA;

Add 1 µL glycol blue to the RNA and bring volume up to 100 µL with sterile water.

Incubate samples at 37 ºC for 30 min

Add 100 µL phenol/ chloroform, vortex for 1 min;

Centrifuge at 21,000 g for 1 min to separate phases;

Take the supernatant (aqueous phase) and add 1 volume (around 80 µL Chloroform);

Vortex for 1 min, centrifuge for 1 min to separate phases;

Take supernatant, add 1/10 vol KAc, and 2.5 vol ethanol, precipitate at -20 ºC for 1-4 h, centrifuge at 4 ºC for 30 min at full speed (21,000 g). There will be a small blue pellet at the bottom;

Remove supernatant, wash with 80% ethanol, let air dry for 5 min at room temperature, dissolve in 10 µL water.

4.18 RT-PCR analysis

Use 1 µg of RNA per RT reaction;

Make reaction mix for a final total volume of 20 µL per reaction, below is for one reaction;
1 µL 10 mM dNTPs;
1 µL 0.1 M Dithiothreitol (DTT);
4 µL 5X First-Strand Buffer;
0.5 µL Reverse Transcriptase

Aliquot 6.5 µL of reaction mix into new PCR tubes;

Mix up to 5 primers in one mix. However, it is important to note some primers may miss-pair with other primers in the PCR (Figure 4.5), Primer sequences are in Table 4.2. We routinely use gene-specific exon-junction primers, but priming with random hexamers is also possible.

Add 1 µL 10 µM Reverse primer to desired RT reaction tube;

Add 1 µg of RNA to PCR reaction tube;

Add RNase Free H2O up to a total volume of 20 µL;

Spin tubes down to remove residue on side of tubes and place in thermocycler;

Run the RT reaction in thermocycler at 50 ºC for 50 minutes;

Store RT cDNA at -20 °C or proceed to PCR reaction.

4.19 Results

Reporter genes allow determining regulatory factors that influence circular RNA formation. However, these reporter genes are large and contain repetitive elements that often make DNA constructs unstable. Due to their large size, it is often necessary to delete parts of the introns, which is achieved by amplifying genomic pieces containing the exons and smaller flanking intronic parts. These DNA pieces are enzymatically assembled, allowing construction without restriction enzymes.

The example of a circular RNA generated from the microtubule associated protein tau (MAPT) shows an application of the minigene approach to analyze circular RNAs. The tau 9→12 minigene used in this example was co-transfected with different splicing factors and the effect of these splicing factors was detected by RT-PCR (Figure 4.6). Different trans-acting factors influence both circular RNA and linear pre-mRNA formation. The experiment also shows that all the sequence elements necessary for circular RNA formation are localized in the cloned fragment.

4.20 Discussion

In general, circular RNAs are low abundant (9), which complicates the study of their function and formation. Similar to linear RNAs (113), the use of reporter minigenes allows the identification of cis and trans-acting factors that regulate the formation of circular RNAs. Thus, this approach generates hypotheses that can be further tested using the endogenous genes.

The most critical step is the design of the reporter gene. The enzymatic assembly of DNA fragments (“Gibson cloning (126)”) facilitates this design, as it allows construction of large reporter genes independent of restriction sites.

The back-splicing sites are brought together through flanking inverted repeats, which should be taken into account in reporter gene construction. The repeats are annotated in the genome browser ‘repeat track’ and selecting them shows their orientation. It should be kept in mind that proteins can also force the back-splicing sites into a secondary structure needed for circular RNA expression (46) and for an unbiased analysis 1-2 kb of flanking intronic regions should be investigated.

To ensure stability of the constructs, an important consideration is the type of bacterial strains and their growth conditions. For shorter, simple constructs standard cloning bacteria are used, which are almost identical to DH5-alpha (huA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17). For longer fragments,

containing more than 6 Alu elements, “stable” competent cells are used that lack a recombinase (recA) and endonuclease (endA1) (F' proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10 (TetR) Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC). If problems appear with recombination, indicated by low transformation counts, plate the transformed bacteria on two plates and let them grow at 30 and 37 ºC, respectively. Due to the presence of numerous repetitive elements in the minigenes, they need to be fully sequenced using next generation sequencing, which is commercially available for around $150 per plasmid at 2019 rates. The sequence of the example is shown in Supplemental Figure 4.5. In addition, restriction fragment length polymorphism analysis for new larger preparation of the constructs is routinely performed. For example, using sites that cut 1-4 times results in a characteristic band pattern that rules out recombination (Figure 4.4). Enzymes should be selected that give a characteristic band pattern of fragments that can be separated on an agarose gel.

Circular RNAs are analyzed with RT-PCR using exon junction primers that overlap with the backsplicing event (Figure 4.1F). Due to the circular nature of the RNA, the reverse (i.e. antisense) primer is upstream of the forward (i.e. sense) primer (Figure 4.1G). Primers detecting the abundantly expressed homeodomain-interacting protein kinase 3 (HIPK3) circular RNA (9) are used as a positive control. HIPK3 and minigene specific reverse primers are reverse transcribed in the same tube, which allows their comparison. PCR reactions are performed with primers amplifying the linear mRNAs to compare processing patterns of circular and linear pre-mRNAs. We frequently observed aberrant bands when primers for linear and circular RNAs were mixed (Figure 4.5), and thus keep the reverse transcription of these samples separate.

RT-PCR analysis of circular RNAs is challenging and need to be carefully controlled. While sensitive and convenient, it can produce artifacts unique for circular RNAs (17). The reverse transcriptase can move several times around the RNA circle, which generates concatemers. Most circular RNA reporter genes generate both circular and linear RNA, which can cross-hybridize, leading to more PCR artifacts (42, 69, 121). It is thus imperative to sequence the PCR products and validate findings using different techniques using Northern blots (40) or RNase protections (1).

Unexplained bands can also originate from aberrant amplification of linear RNA. Linear RNA can be removed using the exonuclease RNaseR, which enriches circular RNAs(84)(Figure 4.5C). RNaseR treatment helps in the initial optimization of detection primers and can often be omitted once primers are optimized.

Alternative back-splicing can also contribute to unexplained bands as multiple circular RNAs can be formed from a genomic locus (34). This alternative back-splicing is often the result of competing pre-mRNA structures formed by more than two inverted repeat elements. In addition cryptic back-splice sites can occur (40, 45). Depending on the experimental goal, Alu-elements can be repeated or added to the constructs. The complementary regions flanking back-splicing sites can be as short as 30-40 nt (45) and

replacement of Alu elements with shorter complementary regions can increase circular RNA formation (33), which can be tested to improve circular RNA formation. Once the pre-mRNA sequences that cause back-splicing have been identified, it is thus possible to shorten circular RNA expressing constructs, which can improve transfection efficiency in some cases.

Table 4-1 List of current minigenes expressing circular RNAs

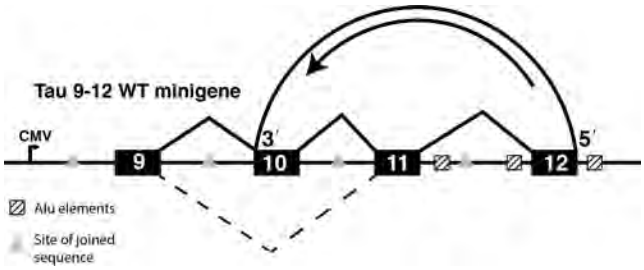
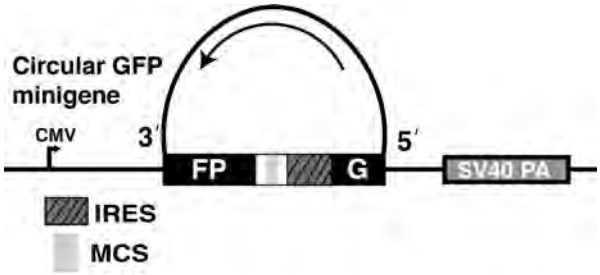
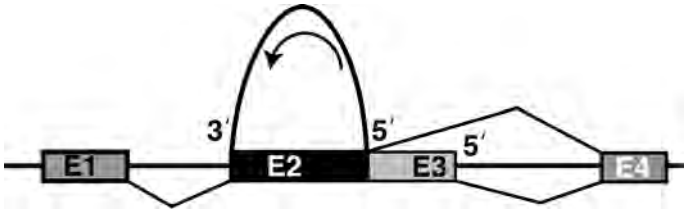
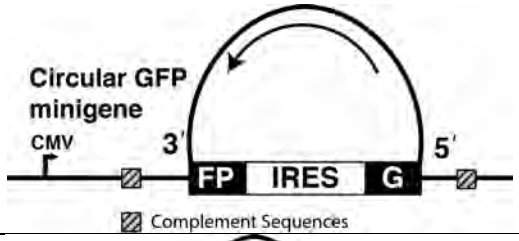
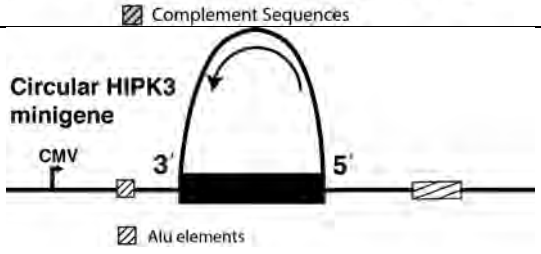
Microtubule-Associated Protein Tau		(1)
Circular RNA with 1 Exon (GFP)		(88)
circRNA from LPAR1		(40)
Circular GFP		(127)
Circular HIPK3		(128)

Table 4-1 Continued

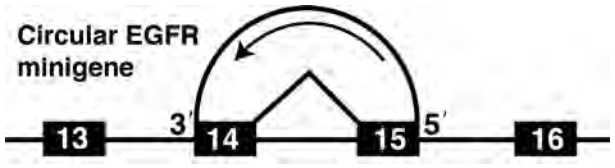
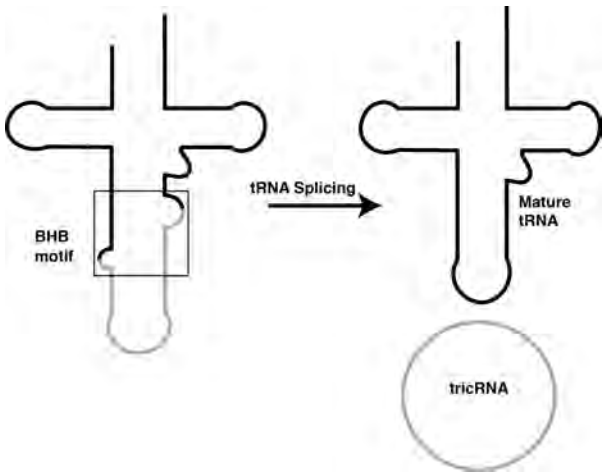
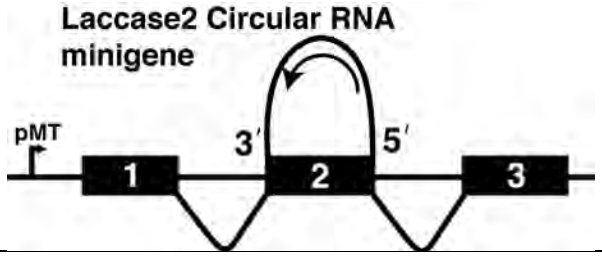
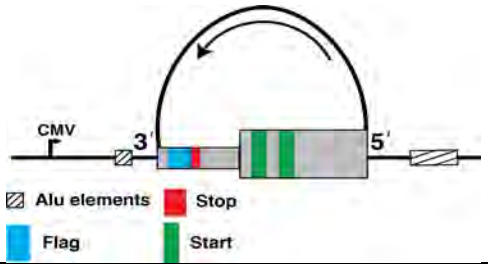
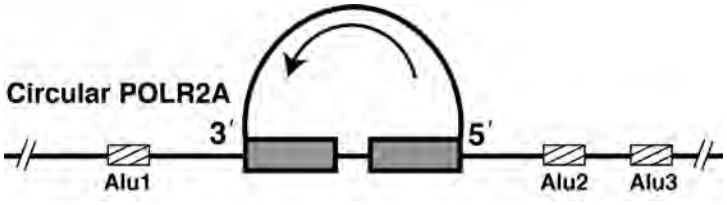
epidermal growth factor receptor (EGFR)		(129)
circular RNA from tRNAs		(121, 122)
Laccase2 CircRNA		(130)
CircZNF609		(23)
Circular RNA from POL2RA		(33, 34)

Table 4-1 Continued

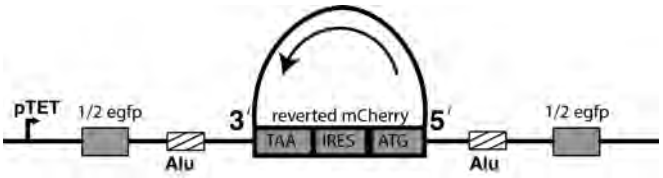
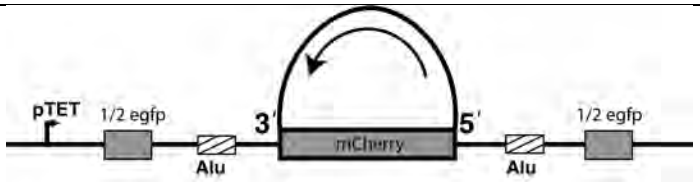
mCherry circular RNA with IRES		(131)
mCherry Circular RNA		(132)

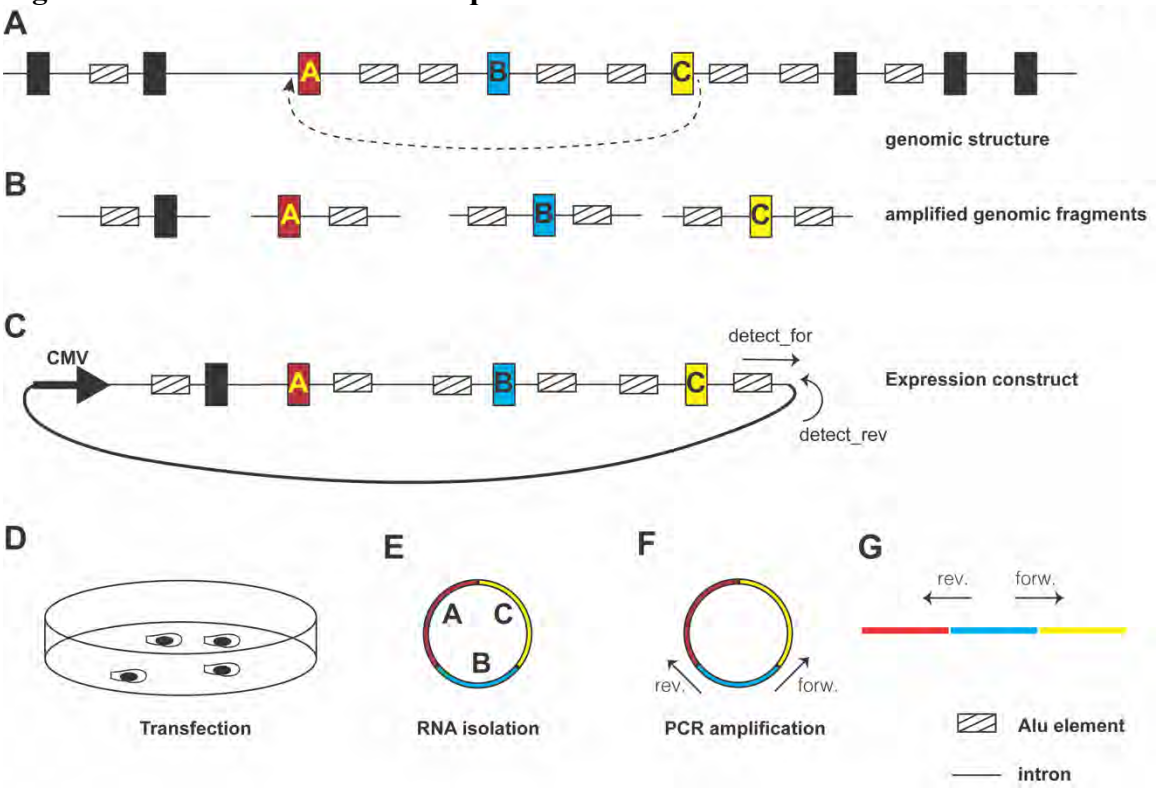
Table 4-2 List of Primers

Circular HIPK3 Control Primers	
HIPK3 Reverse	TGCTTGGCTCTACTTTGAGTTTC
HIPK3 Forward	TCGGCCAGTCATGTATCAAA
Linear Primers	
Tau Exon 12 Reverse	CCCAATCTTCGACTGGACTC
Tau Exon 9 Forward	TGTCAAGTCCAAGATCGGCT
Circular Primers	
circTau exon12_10 Reverse	CAGCTTCTTATTAATTATCTGCACCTTTT
circTau exon10-11 Forward	GAGGCGGCAGTGTGCAA

Table 4-3 Table of Materials

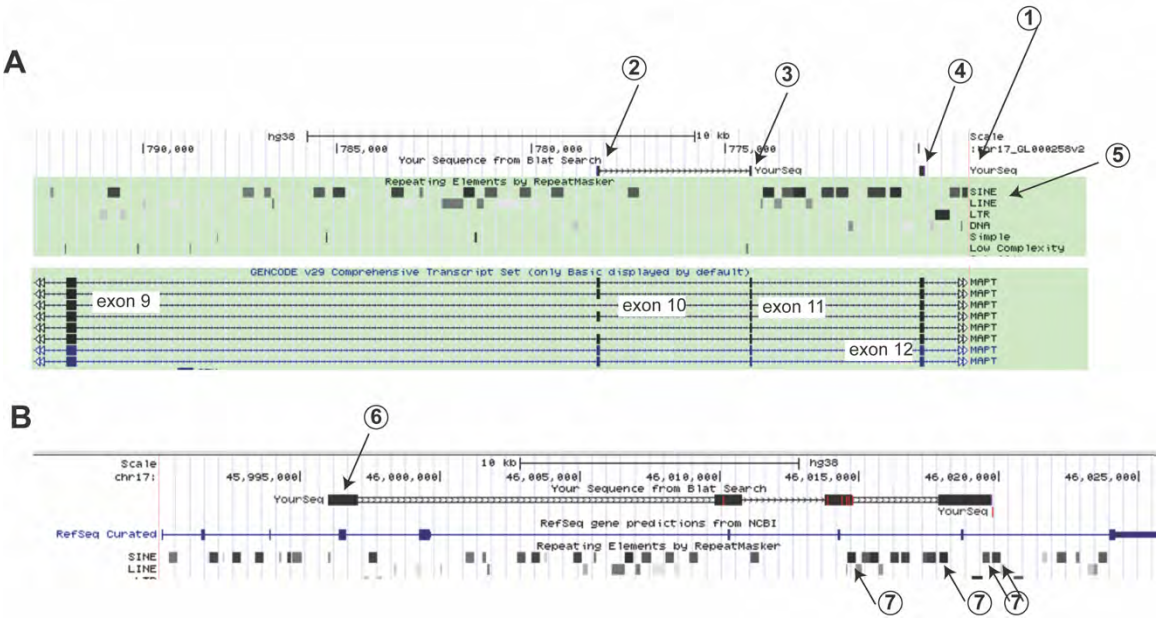
Item	Company / provider	item
Web tool to design primers	NEB builder	https://nebuilder.neb.com/#!/
pcDNA3.1 cloning site	Polycloning site	https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets/LSG/manuals/pcdna3_1_man.pdf
polymerase 1	NEB	Q5 DNA polymerase (NEB #M0491L) Ipswich, MA 01938-2732, USA
Polymerase 2	Biorad, Qiagen	Long range polymerase (NEB # E5200S) Ipswich, MA 01938-2732, USA, iproof (BioRad #1725310) Hercules, CA, USA, Qiagen long range polymerase kit (#206402) Hilden, Germany, EU
web-based temperature calculations	NEB	https://tmcalculator.neb.com/#!/main
Dark Reader Transilluminator	Clare Chemical Research	Clare Chemical Research, Dolores, CO 81323, USA
the gel and PCR cleanup kit	Promega	(Promega #A9282) Madison, WI 53711 USA
enzymatic DNA assembly kit	NEB	NEB # E2621S Ipswich, MA 01938-2732, USA
(PEI) Hydrochloride	Polysciences	(Polysciences #24765-1) Warrington, PA 18976, USA
RNA isolation kit	Life Tech	Ambion by Life Technologies #12183025 Carlsbad, CA, USA
RNAse R	Lucigen	Epicenter/Lucigen #RNR07250 Middleton, WI 53562, USA
Glyco Blue	Thermo Fisher	Invitrogen By Thermo Fisher Scientific #AM9516 Waltham, MA, USA
Reverse Transcriptase	Thermo Fisher	ThermoFisher #18080044 Waltham, MA, USA
Standard cloning bacteria	NEB	NEB5-alpha competent # C2988J Ipswich, MA 01938-2732, USA
"stable" competent cells	NEB	NEB stable cells # C3040H Ipswich, MA 01938-2732, USA
Builder tool	NEB	NEB builder tool https://nebuilder.neb.com/#!/

Figure 4-1 Overview of the technique



A. A hypothetical gene is shown. Introns are lines, exons are boxes, Alu elements are smaller striped boxes. Backsplicing from exon C to A creates a circular RNA. The structure of this circular RNA is shown in panel E.
B. To create a reporter gene, exons and surrounding introns (at least 500 nt on each side) are amplified. The constructs should contain repetitive elements, which are usually Alu elements in humans. An exon upstream of exon A was included to provide an additional Alu element. The genomic fragments will overlap with their flanking 25 nts.
C. Fragments are cloned into an expression vector, driven by a CMV promoter. The successful recombination is detected by detection primers and validated by sequencing.
D. Cells are transfected with this construct and
E. circular RNA is isolated and
F. amplified using circular RNA specific primers, preferable exon junction primers. During PCR amplification, linear RNA can also be amplified (G).
G. Orientation of the primers used to detect circular RNAs. The forward primer is in sense orientation (i.e. has the same sequence as the RNA) and the reverse primer is in antisense orientation (i.e. is the reverse complement of the RNA). Note that different from RT-PCR for linear mRNAs, the reverse primer is upstream of the forward primer.

Figure 4-2 Selection of the sequence for minigene construction



A. Browser display after the sequence shown in Supplemental Figure 1 is run against the human genomic database using BLAT. Literature exon numbers(77) are indicated in the gene display, they are different from the numbers given by the browser.

1. The aligned sequences are shown under 'YourSeq'

2-4. Note that due to the circularity of the RNA, BLAT does not connect all exons with lines as it does in linear RNA. Exons 10 and 11 (corresponding to 2 and 3) are connected, but exon 12 (corresponding to 4) is not connected to exon 11.

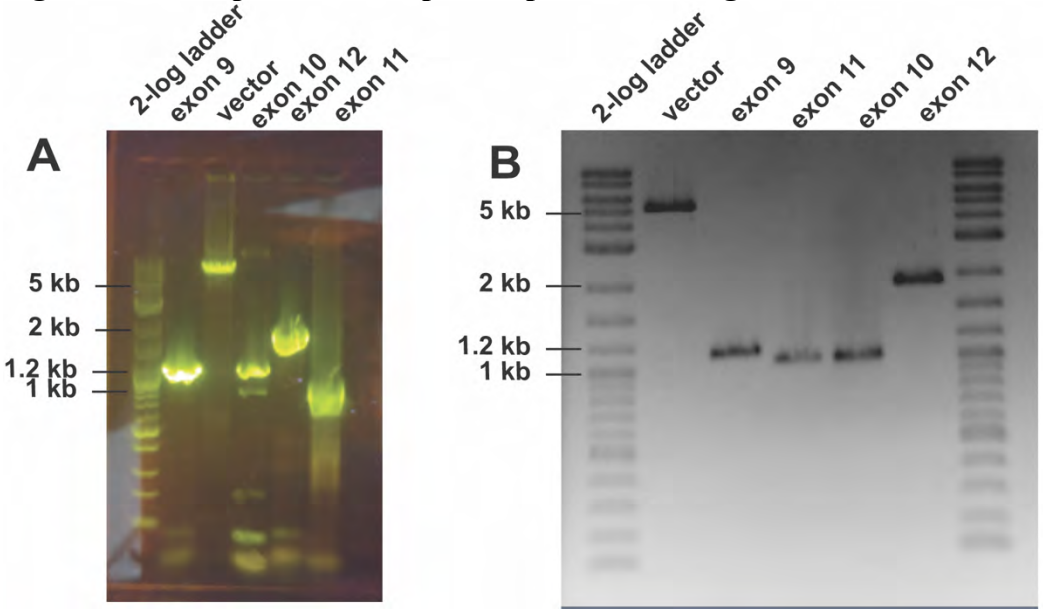
5. Alu elements are shown in the repetitive element track.

B. Sequence alignment between the planned construct and genomic DNA.

6. The planned construct was run against the database using BLAT.

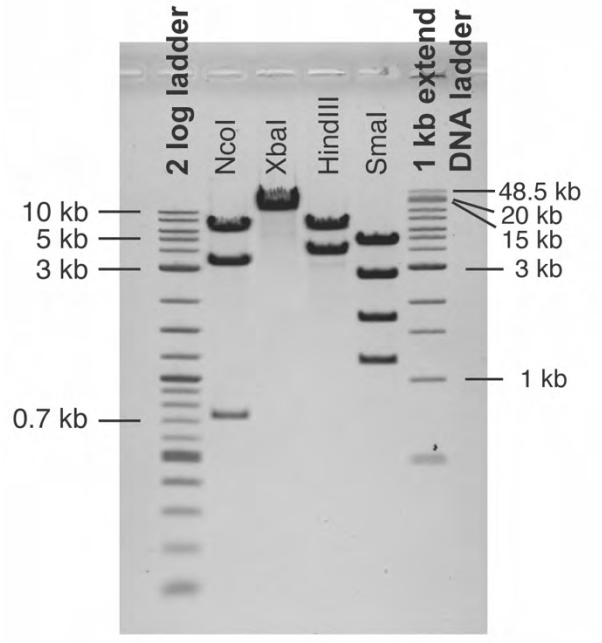
7. Note the inclusion of several Alu elements in the construct.

Figure 4-3 Examples of the amplicons prior to cloning



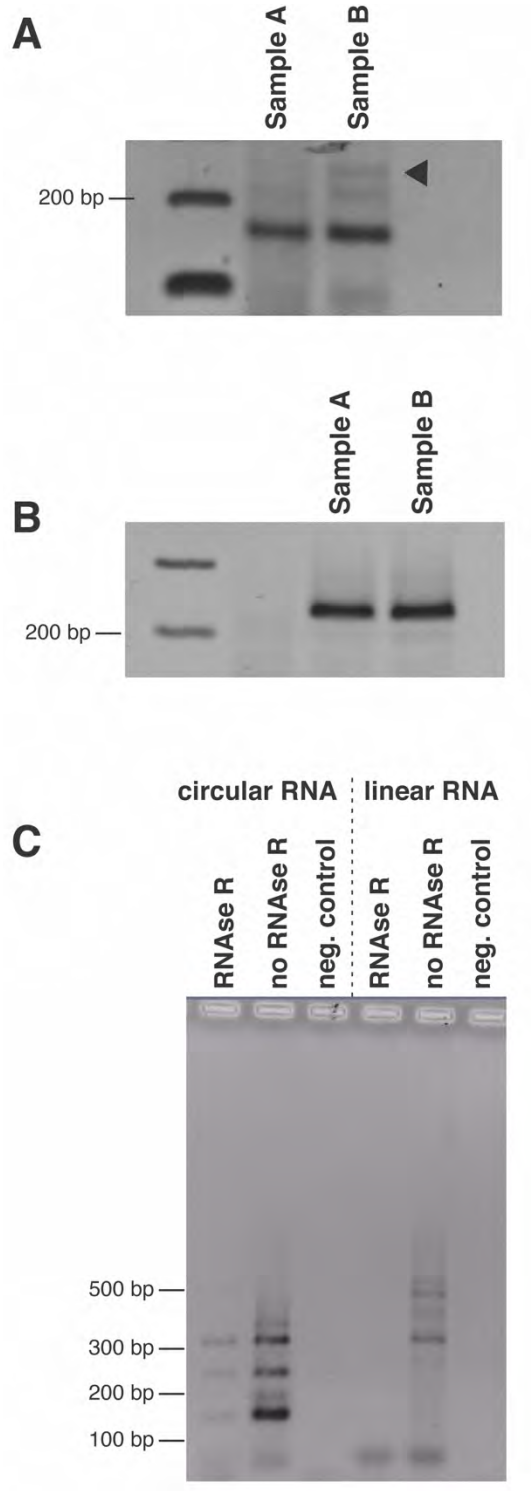
A. Optimized PCR products separated on a 1% agarose gel containing 1x gel green. The individual bands represent the PCR products that will be used enzymatic DNA assembly. **B.** The bands from A were cut out from the gel and purified. The purified PCR products were separated on a 1% agarose gel, which was subsequently stained with ethidium bromide. Both gels were run in order to determine the correct banding size, as the gel green does not run true to size.

Figure 4-4 Restriction analysis of reporter genes



The tau 9-12 minigene used as an example was cut with restriction enzymes indicated to rule out major recombinations. **Lane 1**: cut with NcoI expected sizes 735 bp, 3345 bp, 6266 bp, **lane 2** cut with XbaI expected size 10346 bp, **lane 3** cut with HindIII expected sizes 3951 bp, 6395 bp, **lane 4** cut with SmaI expected sizes 1168 bp, 1688 bp, 2708 bp, 4782 bp.

Figure 4-5 Effect of primer multiplexing and RNase R treatment on circular RNA detection



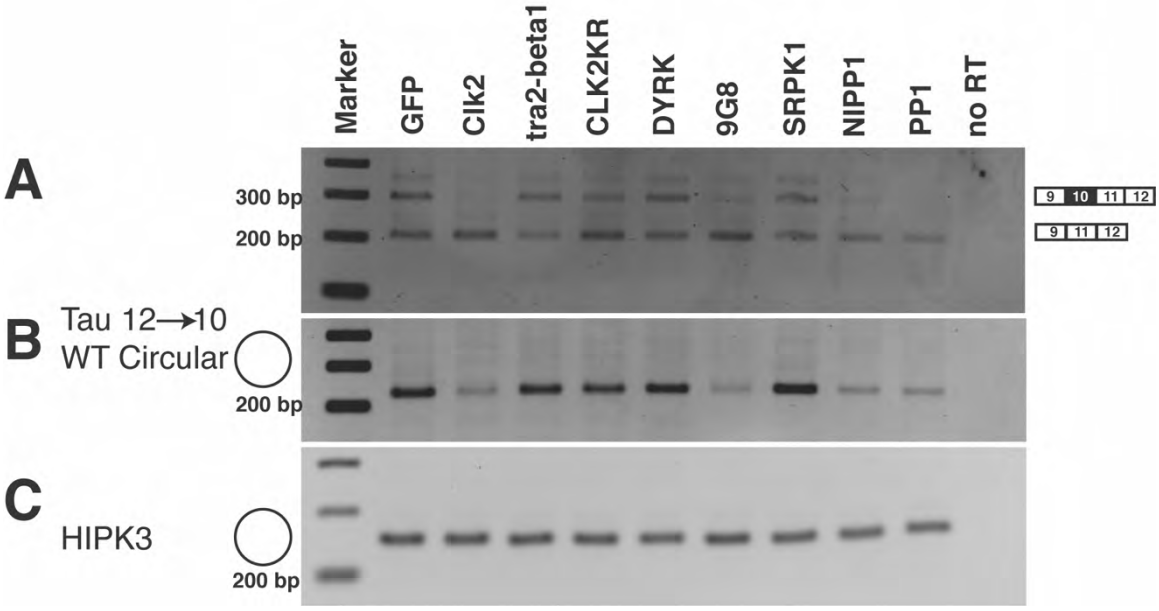
A. cDNA from samples A and B derived from human brain tissues was amplified with circular RNA primers circTau exon12_10 Reverse and circTau exon10_11 Forward. The reverse transcription for the cDNA was performed with the primers for linear and circular

tau RNA. The expected band corresponding to tau circular RNA is shown by a triangle. The other strong bands are artifacts that did not match the human genome.

B. The experiment was repeated with identical PCR conditions, but the reverse transcription was performed only with the circTau exon12_10 Reverse primer. Only the expected band was amplified and validated through sequencing.

C. The RNA was treated with RNase R that removes linear RNA. The circular RNA is detectable after the treatment (left), whereas linear RNA gives no longer a detectable signal (right).

Figure 4-6 Example of an analysis of a circRNA reporter gene



1 µg of the tau 9→12 (133) reporter gene was transfected with 1 µg of splicing factors indicated. RNA was isolated 24 h post transfection and analyzed by RT-PCR.
A. Amplification of the linear tau mRNA. Due to alternative splicing of exon 10 two bands are observed. Their ratio changes due to the overexpression of splicing factors (85, 134).
B. Amplification of the circular 12→10 tau RNA (1). Note the dependency of tau circRNA expression on expression of some splicing factors, especially the cdc2 like kinase clk2 and the SR protein 9G8.
C. The circular RNA of HIPK3 was used as a positive control indicating equal loading.

Supplemental Figure 4-1 Tau circular RNA test sequence

GTGAACCTCCAAAATCAGGGGATCGCAGCGGCTACAGCAGCCCCGGCTCCCC
AGGCACTCCCGGCAGCCGCTCCCGCACCCCGTCCCTTCCAACCCCACCCACCC
GGGAGCCCAAGAAGGTGGCAGTGGTCCGTACTCCACCCAAGTCGCCGTCTTC
CGCCAAGAGCCGCCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGAAGAAT
GTCAAGTCCAAGATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGC
GGGAAGGAGGTGGCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGG
ACAGAGTCCAGTCGAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGG
CGGAGGAAATAAAAAGgtgcagataattaataagaagctggatcttagcaacgtccagtccaagtgtggctcaa
aggataatatcaaacacgtcccgggaggcggcagtGTGCAAATAGTCTACAAACCAGTTGACCTG
AGCAAGGTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAG

Test sequence corresponding to a circular RNA from the MAPT locus. Different exons are indicated by underline, small caps and large caps.

Supplemental Figure 4-2 Genomic sequence containing the planned minigene

> chr17:45,995,463-46,019,282

gttccgtgggccacactttggaaaatacagacccatgagatagaataccagactgttgaa
gtgtaacgggggcctgggaagtgcagtaacagaagcaagtttgagggtaaaggacaccca
gaggagggagggacagcatctgcatggagaggagaagagacccccagcagcttccaggg
tgttggaagggtgcgctagtaactgctatgcatggcaggtggggaactgtacgtcagggc
acagcagcatgaagcggtatggctcgtgtggacagctagggacaggcaggcgtggagcag
gcatcctgttctgaaggccaaatcccacagaggagccagggtgctggcaggagccctgaa
ctagccgaacagctgaacagctgaacattcaccctgtggggaaagggtcagaagcgtcca
ggcttgagggcacagctgggtctcgtcactgcatcacccttatttaggataaaggccctg
aagaattgtattagaggttggcaaagcatatctaccacctcctggagccacgctggcc***gc***
agggattataattatttccattttcaaattaaggcctctgagctcagagaggggaagttta
cttgtctgaggccacacagcttgttggagcccatctcttgacccaaagactgtggagccg
agttggccacctctctgggagcgggtattggatggtggttgatggttttccattgctttc
ctgggaaaggggtgtctctgtccctaagcaaaaaggcagggaggaagagatgcttcccca
gggcagccgtctgctgtagctgcgcttccaacctggcttccacctgcctaacccagtggt
gagcctgggaatggacccacgggacaggcagcccccagggccttttctgaccccacccac
tcgag***tcctggcttcactcccttccttccttcc***cagGTGAACCTCCAAAATCAGGGGATC
GCAGCGGCTACAGCAGCCCCGGCTCCCCAGGCACTCCCGGCAGCCGCTCCCGCACCCCGT
CCCTTCCAACCCCACCCACCCGGGAGCCCAAGAAGGTGGCAGTGGTCCGTACTCCACCCA
AGTCGCCGTCTTCCGCCAAGAGCCGCCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGA
AGAATGTCAAGTCCAAGATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGCGGGAAG
gtgagagtggctggctgcgcgtggaggtgtgggggctgcgcctggaggggtagggct
gtgcctggaagggtagggctgcgcctggaggtgcgcggttgagcgtggagtcgtgggact
gtgcatggaggtgtggggctccccgcacctgagcacccccgcataacaccccagtcccct
ctggaccctcttcaaggaagttcagttctttattgggctctccactacactgtgagtgcc
ctcctcaggcgagagaacgttctggctcttctcttgccccttcagcccctgttaatcgga
cagagatggcagggctgtgtctccacggccggaggctctcatagtcagggcacccacagc
ggttccccacctgccttctgggcagaatacactgccacccataggtcagcatctccactc
gtgggccatctgcttaggttgggttcctctggattctggggagattgggggttctgtttt
gatcagctgattcttctgggagcaagtgggtgctcgcgagctctccagcttcctaaaggt
ggagaagcacagacttcgggggcctggcctggatccctttccccattcctgtccctgtgc
ccctcgtctgggtgc***gttagggctgacatacaaagcaccacagtgaaagaacagcagtat***
gcctcctcactagccaggtgtgggcgggtgggtttcttccaaggcctctctgtggccgtg
ggtagccacctctgtcctgcaccgctgcagtcttccctctgtgtgtgctcctggtagctc
tgcgcatgctcatcttcttataagaacaccatggcagctgggcgtagtggctcacgccta
taatcccagcactttgggaggctgaggcaggcagatcacgaggtcaggagttcgagacca
acctgaccaacagggtgaaacctcgtctctactaaaaatacaaaaatacctgggcgtggt
ggtggtgcgcgcctataatcccagctactcaggaggctgaggcaggagaatcgcttgaac
ccaggaggcagaggttgcagtgagccgagatagtgccactgcactccagtttgagcaaca
gagcgagactctgtctcaaaacaaaataaaacaaaccaaaaaaacccaccatggcttagg
gcccagcctgatgacctcatttttcacttagtcacctctctaaaggccctgtctccaaat
agagtcacattctaaggtacgggggtgttggggaggggggttagggcttcaacatgtgaa
tttgcggggaccacaattcagcccaggaccccgctcccgccacccagcactggggagctg
gggaagggtgaagaggaggctgggggtgagaaggaccacagctcactctgaggctgcaga
tgtgctgggccttctgggcactgggcctcggggagctaggggggctttctggaaccctggg
cctgcgtgtcagcttgcctcccccacgcaggcgctctccacaccattgaagttcttatca
cttgggtctgagcctggggcatttggacggagggtggccaccagtgcacatgggcacctt
gcctcaaaccctgccacctcccccccacccaggatcccccctgcccccgaacaagcttgtg
agtgcagtgtcacatcccatcgggatggaaatggacggtcgggttaaaagggacgcatgt
gtagaccctgcctctgtg***catcaggcctctttgagagtccctgcgtgccaggcggtgca***
cagaggtggagaagactcggctgtgccccagagcacctcctctcatcgaggaaaggacag
acagtggctcccctgtggctgtggggacaagggcagagctccctggaacacaggagggag
ggaaggaagagaacatctcagaatctccctcctgatggcaaacgatccgggttaaattaa
ggtccggccttttcctgctcaggcatgtggagcttgtagtggaagaggctctctggaccc
tcatccaccacagtggcctggttagagaccttggggaaataactcacaggtgacccaggg
cctctgtcctgtaccgcagctgagggaaactgtcctgcgcttccactggggacaatgcgc
tccctcgtctccagactttccagtcctcattcggttctcgaaagtcgcctccagaagccc

catcttgggaccaccgtgactttcattctccagggtgcctggccttggtgctgcccaaga
ccccagaggggccctcactggcctttcctgccttttctcccattgcccacccatgcaccc
ccatcctgctccagcacccagactgccatccaggatctcctcaagtcacataacaagcag
cacccacaaggtgctcccttcccccctagcctgaatctgctgctccccgtctggggttccc
cgcccatgcacctctgggggcccctgggttctgccataccctgccctgtgtcccatggtg
g***ggaatgtccttctctccttatctcttcccttcccttaaatccaagttcagttgccatct***
cctccaggaagtcttcctggattcccctctctcttcttaaagcccctgtaaactctgacc
***acactga*GCATGTGTCTGCTGCTCCCTAGTCTGGGCCATGAGTGAGGGTGGAGGCCAAGT**
CTCATGCATTTTTGCAGCCCCCACAAGACTGTGCAGGTGGCCGGCCCTCATTGAATGCGG
GGTTAATTTAACTCAGCCTCTGTGTGAGTGGATGATTCAGGTTGCCAGAGACAGAACCCT
CAGCTTAGCATGGGAAGTAGCTTCCCTGTTGACCCTGAGTTCATCTGAGGTTGGCTTGGA
AGGTGTGGGCACCATTTGGCCCAGTTCTTACAGCTCTGAAGAGAGCAGCAGGAATGGGGC
TGAGCAGGGAAGACAACTTTCCATTGAAGGCCCCTTTCAGGGCCAGAA***CTGTCCCTCCCA***
CCCTGCAGCTGCCCTGCCTCTGCCCATGAGGGGTGAGAGTCAGGCGACCTCATGCCAAGT
GTAGAAAGGGGCAGATGGGAGCCCCAGGTTATgacgtcaccatgctgggtggaggcagca
cgtccaaatctactaaagggttaaaggagaaagggtgacttgacttttcttgagatattt
tgggggacgaagtgtggaaaagtggcagaggacacagtcacagcctcccttaaatgccag
gaaagcctagaaaaattgtctgaaactaaacctcagccataacaaagaccaacacatgaa
tctccaggaaaaaagaaaaagaaaaatgtcatacagggtccatgcacaagagcctttaaa
atgacccgctgaagggtgtcaggcctcctcctcctggactggcctgaaggctccacgagc
ttttgctgag***accttgggtccctgtggcctcatgtagtacccagtatgcagtaagtgct***
caataaatgtttggctacaaaagaggcaaagctggcggagtctgaagaatccctcaaccg
tgccggaacagatgctaacaccaaagggaaaagagcaggagccaagtcacgtttgggaac
ctgcagaggctgaaaactgccgcagattgctgcaaatcattgggggaaaaacggaaaacg
tctgttttccc***ctttgtgcttttctctgttttcttctttgtgcttttctctgtttt***cagg
atttgctacagtgaacatagattgctttggggccccaaatggaattattttgaaaggaaa
atgcagataatcaggtggccgcactggagcaccagctgggtaggggtagagattgcaggc
aaggaggaggagctgggtggggtgccaggcaggaagagcccgtaggccccgccgatcttg
tgggagtcgtgggtggcagtgttccctccagactgtaaaagggagcacctggcgggaaga
gggaattcttttaaacatcattccagtgcccgagcctcctggacctgttgtcatcttgag
gtgggcctcccctgggtgactctagtgtgcagcctggctgagactcagtggccctgggtt
cttactgctgacacctaccctcaacctcaaccactgcggcctcctgtgcaccctgatcca
gtggctcattttccactttcagtcccagctctatcctatttgcagtttccaagtgcctg
gtcctcagtcagctcagacccagccaggccagcccctggttcccacatccccttgccaa
gctcatccccgccctgtttggcctgcgggagtgggagtgtgtccagacacagagacaaag
gaccagcttttaaaacatttgttgg***ggccaggtgtggtggctcacacctaatcccaaca***
cctggggaggccaaggcagaaggatcacttgagtccaggagttcaagaccagcctgggca
acatagggagaccctgtctctacaatttttttttaattagctgggcctgttggcactct
cctgtagttccagctactctagaggctgaggtgggaggactgcttgagcctgggaggtca
gggctgcaatgagccatgttcacaccactgaacgccagcctgggcgagaccctgtatcaa
aaaagtaaagtaaaatgaatcctgtacgttatattaaggtgccccaaattgtacttagaa
ggatttcatagttttaaatacttttgttatttaaaaaattaaatgactgcagcatataaa
ttaggttcttaatggaggggaaaaagagtacaagaaaagaaataagaatctagaaacaaa
gataagagcagaaataaaccagaaaacacaaccttgcactcctaacttaaaaaaaaaaat
gaagaaaacacaaccagtaaaacaacatataacagcattaagagctggctcctg***gctggg***
cgcggtggcgcatgcctgtaatcccaacactttgggaggccgatgctggaggatcacttg
agaccaggagttcaaggttgcagtgagctatgatcataccactacaccctagcctgggca
acacagtgagactgagactctattaaaaaaaaaatgctggttccttccttatttcattcc
ttt***attcattcattcagacaacatttatggggcacttctgagcaccaggctctgtgctaa***
gagcttttgcccccagggtcca***ggccaggggacaggggcaggtgagcagagaaacaggg***c
cagtcacagcagcaggaggaatgtaggatggagagcttggccaggcaaggacatgcaggg
ggagcagcctgcacaagtcagcaagccagagaagacaggcagacccttgtttgggacctg
ttcagtggcctttgaaaggacagcccccacccggagtgctgggtgcaggagctgaaggag
gatagtggaacactgcaacgtggagctcttcagagcaaaagcaaaataaacaactggagg
cagctggggcagcagagggtgtgtgttcagcactaaggggtgtgaagcttgagcgctagg
agagttcacactggcagaagagaggttggggcagctgcaagcctctggacatcgcccgac
aggacagagggtggtggacggtggccctgaagagaggctcagttcagctggcagtggccg
tgggagtgctgaagcaggcaggctgtcggcatctgctggggacggttaagcaggggtgag

ggcccagcctcagcagcccttcttgggggtcgctgggaaacatagaggagaactgaaga
agcagggagtcccagggtccatgcagggcgagagagaagttgctcatgtggggcccaggc
tgcaggatcaggagaactggggaccctgtgactgccagcggggagaaggggtgtgcagg
atcatgcccagggaagggcccagggcccaagcatgggggggcctggttggctctgagaa
gatggagctaaagtcactttctcggaggatgtccaggccaatagttgggatgtgaagacg
tgaagcagcacagagcctggaagcccaggatggacagaaacctacctgagcagtggggct
ttgaaagccttggggcgggggtgcaatattcaagatggccacaagatggcaatagaatg
ctgtaactttcttggttctgggccgcagcctgggtggctgcttccttccctgtgtgtatt
ga***tttgtttctctttttgagacagagtcttgctgggttgcccaggctggagtgcagtgg***
tgcgatcatagctcactgcagccttgaagtcctgagctcaagagatccttccacctcagc
ctcctgagtagttgggaccacaggcttgcaccacagtgcccaactaatttcttatatttt
ttgtagagatggggtttcactgtgtcgcccaggatggtcttgaactcctgggctcaagtg
atcctcctgcctcagcctcgcaaattgctgggattacaggtgtgagccaccatgcccgac
cttctctttttaagggc***gtgtgtgtgtgtgtgtgtgtgtg***ggcgcactctcgtcttcacc
ttcccccagccttgctctgtctctacccagtcacctctgcccatctctccgatctgtttc
tctctccttttaccctctttcctccctcctcatacaccactgaccattatagagaactg
agtattctaaaaataca***ttttatttatttattttgagacagagtctcactctgtcaccca***
ggctggagtgcagtggtgcaatctcggctcactgcaacctccgcctcccaggttgaagca
actctcctgcctcagcctccctagtagctgggattacaagcacacaccaccatgcctagc
aaatttttatattttagtagaggaggggtgtcaccatgtttgccaagctggtctcaaac
tcctggcctcaggtgatctgcctaccttggtctcccaaagtgctgggattacaggtgtga
gccaccacgcctgcccttaaaaatacattatatttaatagcaaagccccagttgtcactt
taaaaagcatctatgtagaacatttatgtggaataaatacagtgaatttgtacgtggaat
cgtttgcctctcctcaatcagggccagggatgcaggtgagcttgggctgagatgtc***agac***
cccacagtaagtgggggcagagccaggctgggaccctcctctaggacagctctgtaact
ctgagaccctccaggcatcttttcctgtacctcagtgcttctgaaaaatctgtgtgaatc
aaatcatttaaaggagcttgggttcatcactgtttaaaggacagtgtaaataattctga
aggtgactctaccctgttatttgatctcttctttggccagctgacttaacaggacataga
caggttttcctgtgtcagttcctaagctgatcaccttggacttgaagaggaggcttgtgt
gggcatccagtgcccaccccgggttaaactcccagcagagtattgcactgggcttgctga
gcctggtgaggcaaagcacagcacagcgagcaccaggcagtgctggagacaggccaagtc
tgggccagcctgggagccaactgtgaggcacggacggggctgtggggctgtggggctgca
ggcttggggccagggagggagggctgggctctttggaacagccttgagagaactgaaccc
aaacaaaaccagatcaaggtctagtgagagcttagggctgctttgggtgctccaggaaat
tgattaaaccaagtggacacacacccccagccccacctcaccacagcctctccttcaggg
tcaaactctgaccacagacatttctcccctgactaggagttccctggatcaaaattggga
gcttgcaacacatcgttctctcccttgatggtttttgtcagtgtctatccagagctgaag
tgtaatatatatgttactgtagctgagaaattaaatttcaggattctgatttcataatga
caaccattcctcttttctctcccttctgtaaatctaagattctataaacggtgttgactt
aatgtgacaattggcagtagttcaggtctgctttgtaaataccttgtgtctattgtaaa
atctcacaaaggcttgttgccttttttgtggggttagaacaagaaaaagccacatggaaa
aaaaatttc***ttttttgttttttttgtttgcttgttttttttgagacagagtttcactctgtc***
gcccaggctggagtgcagtggtgcgatctccgcccactgcaagctccacctcccgggttcc
atgctattctcctgtctcagcctcccaagtagctgggactgcaggtgcccgccaccacac
ctggctaatttttttgtattttagtagagacggggtttcaccgtgttagccaggatggtt
ctcaatctcctgacctcgtcatctgcctgcctcggcctcccaaagtgctgagattacagg
cgtgagccaccgtgcccggccagaaaaaaacatttctaagtatgtggcagatactgaat***t***
attgcttaatgtcctttgattcatttgtttaatttctttaatggattagtacagaaaaca
aagttctcttccttgaaaaactggtaagttttctttgtcagataaggagagttaaataac
ccatgacatttccctttttgcctcggcttccaggaagctcaaagttaaatgtaatgatca
ctcttgtaattatcagtgttgatgcccttcccttcttctaatgttactctttacattttc
ctgctttattattgtgtgtgttttctaattctaagctgttcccactcctttctgaaagca
ggcaaatcttctaagccttatccactgaaaagttatgaataaaaaatgatcgtcaagcct
acaggtgctgag***gctactccagaggctgaggccagaggaccacttgagcccaggaatttg***
agacctgggctgggcagcatagcaagactctatctccattaaaactatttttttttattt
aaaaaataatccgcaaagaaggagtttatgtgggattccttaaaatcggagggtggcatg
aattgattcaaagacttgtgcagagggcgacagtgactccttgagaagcagtgtgagaaa
gcctgtcccacctccttccgcagctccagcctgggctgaggcactgtcacagtgtctcct

tgctggcaggagagaatttcaacattcaccaaaaagtagtattgtttttattaggtttat
gaggctgtagccttgaggacagcccaggacaactttgttgtcacatagatagcctgtggc
tacaaactctgagatctagattcttctgtggctgcttctgacctgagaaagttgcggaac
ctcagcgagcctcacatggcctccttgtccttaacgtggggacggtgggcaagaaaggtg
atgtggcactagagatttatccatctctaaaggaggagtggattgtacattgaaacacca
gagaaggaattacaaaggaagaatttgagtatctaaaaatgtaggtcag***gcgctcctgtg***
ttgattgcagggctattcacaatagccaagatttggaagcaacccaagtgtccatcaaca
gacaaatggataaagaaaatgtggtgcatatacacaatggaatactattcagccatgaaa
aagaatgagaatctgtcatttgaaacaacatggatggaactggaggacattatgttaagt
gaaataagccagacagaaggacagacttcacatgttctcacacatttgtgggagctaaaa
attaaactcatggagatagagagtagaaggatggttaccagaggctgaggagggtggagg
ggagcagggagaaagtagggatggttaatgggtacaaaaacgtagttagcatgcatagat
ctagtattggatagcacagcagggtgacgacagccaacagtaatttatagtacatttaaa
aacaactaaaagagtgtaattggactggctaacatggtgaaaccccgtctctactaaaaa
tacaaaaattagctgggcatggtggctcacgcctgtaatcccagcactttgggaggccga
ggcgggccgatcacgaggtcaggagatcgagaccatcctagctaacatggtgaaaccccg
tctctactacaaatacaaaaaaaagaaaaaattagccgggcatggtggtgggcgcctgta
gtcccagctactcgggaggctgaggcaggagaatggcgtgaacccgggaggcggagcttg
cagtgagccgagatcgcgccactgcactccagcctgggcgacaaggcaagattctatctc
aaaaaaataaaaataaaataaaataaaataatataaaataaaataaaataaaataaaataaa
taaaataaaataaaatgtataattggaatgtttataacacaagaaatgataaatgcttga
ggtgatagataccccattcaccgtgatgtgattattgcacaatgtatgtctgtatctaaaa
tatctcatgtaccccacaagtatatacacctactatgtacccatataaatttaaaattaa
aaaattataaaacaaaaataaataagtaaattaaaatgtaggctggacaccgtggttcac
gcctgtaatcccagtgctttgtgaggctgaggtgagagaatcacttgagcccaggagttt
gagaccggcctgggtgacatagcgagaccccatcatcacaaagaatttttaaaaattagcc
tgggcgtggtagcacataccggtagttccagctacttgggagaccgaggcaggaggattgc
cttgagcccaggagtttaaggctgcagtgagctacgatggcgccactgcattccagcctg
ggtgacagagtgagagcttgtctctatttttaaaaataataaaaagaataaataaaaataa
attaaaatgtaaatatgtgcatgttagaaaaaatacacccatcagcaaaaaggggtaaa
ggag***cgatttcagtcataattggagagatgcagaataagccagcaatgcagtttctttta***
ttttggtcaaaaaaaataagcaaaacaatgttgtaaacacccagtgctggcagcaatgtg
gtgaggctggctctctcaccagggctcacagggaaaactcatgcaacccttttagaaagc
catgtggagagttgtaccgagaggttttagaatatttataactttgacccagaaattcta
ttctaggactctgtgttatgaaaataacccatcatatggaaaaagctcctttcagaaaga
ggttcatgggaggctgtttgtatttttttttctttgcatcaaatccagctcctgcagga
ctgtttgtattattgaagtacaaagtggaatcaatacaaatgttggatagcaggggaaca
atattcacaaaatggaatgggacatagtattaaacatagtgcttctgatgaccgtagacc
atagacaatgcttaggatatgatatcacttcttttgttgtttttgtattttgagacgaa
gtctcattctgtcacccaggctggagttcagtggcgccatctcagctcactgcaacctcc
atctcccgggttcaagctattctccttcctcaacctcccgagtagctgggttgcgcacca
ccatgcctggctaacttttgtatttttagtacagacggggtttcaccacgttggccaggc
tgctcttgaactcctgacgtcaggtgatccaccagccttgacctcccaaagtgctaggat
tacaggagccactgtacccagcctaggatatgatatcacttcttagagcaagatacaaaa
ttgcatgtgcacaataattctaccaagtataggtatacaggggtagttatatataaatga
gacttcaaggaaatacaacaaaatgcaatcgtgattgtgttagggtggtaagaaaacggt
ttttgctttgatgagctctgtttttttaaaatcgttatattttctaataaaaatacatagt
cttttgaaggaacataaaagattatgaagaaatgagttagatattgattcctattgaaga
ttcagacaagtaaaattaaggggaaaaaaaacgggatgaaccagaagtcaggctggagtt
ccaaccccagatccgacagcccaggctgatggggcctccagggcagtggtttccacccag
cattctcaaaagagccactgaggtctcagtgccattttcaagatttcggaagcggcctgg
gcacggctggtccttcactgggatcaccacttggcaattatttacacctgagacgaataa
aaaccagagtgctgagatta***caggcatggtggcttacgcttgtaatcggctttgggaagc***
cgaggtgggctgattgcttgagcccaggagtttcaaactatcctggacaacatagcatga
cctcgtctctacaaaaaatacaaaaaatttgccaggtgtggtggcatgtgcctgtggtcc
cagctacttgggaggctgaagtaggagaatccccctgagccctgggaagtcgaggctgcac
tgagccgtgatggtgtcactgcactccagcctgggtgacaaagtgagaccctatctcaca
aagaaaaaaaacaaaacaaaaaacccaaagcacactgtttccactgtttccagagttcct

gagaggaaaggtcaccgggtgaggaagacgttctcactgatctggcagagaaaatgtcca
gtttttccaactccctaaaccatggttttctatttcatagttcttaggcaaattggtaaa
aatcatttctcatcaaaacgctgatattttcacacctccctggtgtctgcagaaagaacc
ttccagaaatgcagtcgtgggagacccatccaggccacccctgcttatggaagagctgag
aaaaagccccacgggagcatttgctcagcttccgttacgcacctagtggcattgtgggtg
ggagagggctggtgggtggatggaaggagaaggcacagccccccttgcagggacagagc
cctcgtacagaagggacaccccacatttgtcttccccacaaagcggcctgtgtcctgcct
acggggt***cagggcttctcaaacctggctgtgtgtcagaatcaccaggggaacttttcaaa***
actagagagactgaagccagactcctagattctaattctaggtcagggctaggggctgag
attgtaaaaatccacaggtgattctgatgcccggcaggcttgagaacagccgcagggagt
tctctgggaatgtgccggtgggtctagccaggtgtgagtggagatgccggggaacttcct
attactcactcgtcagtgtggccgaacacattttcacttgacctcaggctggtgaacgc
tcccctctggggttcaggcctcacgatgccatccttttgtgaagtgaggacctgcaatcc
cagcttcgtaaagcccgctggaaatcactcacacttctgggatgccttcagagcagccct
ctatcccttcagctcccctgggatgtgactcaacctcccgtcactccccagactgcctct
gccaagtccgaaagtggaggcatccttgcgagcaagtaggcgggtccagggtggcgcatg
tcactcatcgaaagtggaggcgtccttgcgagcaagcaggcgggtccagggtggcgtgtc
actcatcctttttctggctaccaaagGTGCAGATAATTAATAAGAAGCTGGATCTTAGC
AACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAACACGTCCCGGGAGGCGGCAGT
gtgagtaccttcacacgtcccatgcgccgtgctgtggcttgaattattaggaagtggtgt
gagtgcgtacacttgcgagacactgcatagaataaatccttcttgggctctcaggatctg
gctgcgacctctgggtgaatgtagcccggctccccacattcccccacacggtccactgtt
cccagaagccccttcctcatattctaggaggggtgtcccagcatttctgggtcccccag
cctgcgcaggctgtgtggacagaatagggcagatgacggaccctctctccggaccctgcc
tgggaagctgagaatacccatcaaagtctccttccactcatgcccagccctgtccccagg
agccccatagcccattggaagttgggctgaaggtggtggcacctgagactgggctgccgc
ctcctcccccgacacctgggcaggttgacgttgagtggctccactgtggacaggtgaccc
gtttgttctgatgagcggacaccaaggtcttactgtcctgctcagctgctgctcctacac
gttcaaggcaggagccgattcctaagcctccagcttatgcttagcctgcgccaccctctg
gcagagactccagatgcaaagagccaaaccaaagtgcgacaggtccctctgcccagcgtt
gaggtgtggcagagaaatgctgcttttggccctttagatttggctgcctctt***gccagga***
gtggtggctcgtgcctgtaattccagcactttgggagactaaggcgggaggttcgcttga
gcccaggagttcaagaccagcctgggcaacaatgagacccctgtgtctacaaaaagaatt
aaaattagccaggtgtggtggcacgcacctgtagtcccagctacttgggaggctgaggtg
ggaggattgcctgagtccgggaggcggaagttgcaaggagccatgatcgcgccactgcac
ttcaacctaggcaacagagtgagactttgtctcaaaaaacaatcatataataatttaaa
ataaatagatttggcttcctctaaatgtccccggggactccgtgcatcttctgtggagtg
tctccgtgagattcgggactcagatcctcaagtgcaactgacccacccgataagctgagg
cttcatcatcccctggccggtctatgtcgactgggcacccgaggctcctctcccaccagc
tctcttggtcagctgaaagcaaactgttaacaccctggggagctggacgtatgagaccct
tggggtgggaggcgttgatttttgagagcaatcacctggccctggctggcagtaccggga
cactgctgtggctccggggtgggctgtctccagaaaatgcctggcctgaggcagccaccc
gcatccagcccagagggtttattcttgcaatgtgctgctgcttcctgccctgagcacctg
gatcccggcttctgccctgaggccccttgagtcccacaggtagcaagcgcttgccctgcg
gctgctgcatggggctaactaacgcttcctcaccagtgtctgctaagtgtctcctctgtc
tcccacgccctgctctcctgtccccccagtttgtctgctgtgaggggacagaagaggtgt
gtgccgcccccaccctgcccgggcccttgttcctgggattgctgttttcagctgtttga
gctttgatcctggttctctggcttcctcaaagtgagctcggccagaggaggaaggccatg
tgctttctggttgaagtcaagtctggtgccctggtggaggctgtgctgctgaggcggagc
tggggagagagtgcacacgggctgcgtggccaacccctctgggtagctgatgcccaaaga
cgctgcagtgcccaggacatctgggacctccctggggcccgcccgtgtgtcccgcgctgt
gttcatctgcgggctagcctgtgacccgcgctgtgctcgtctgcgggctagcctgtgtcc
cgcgctctgcttgtctgcggtctagcctgtgacctggcagagagccaccagatgtcccgg
gctgagcactgccctctgagcaccttcacaggaagcccttctcctggtgagaagagatgc
cagcccctggcatctgggggcactggatccctggcctgagccctagcctctccccagcct
gggggcccctccccagcaggctggccctgctccttctctacctgggacccttctgcctcc
tggctggaccctggaagctctgcagggcctgctgtccccctccctgccctccaggtatcc
tgaccaccggccctggctcccactgccatccactcctctcctttctggccgttccctggt

ccctgtcccagcccccctcccctctcacgagttacctcacccaggccagagggaagagg
gaaggaggccctggtcataccagcacgtcctcccacctccctcggccctggtccaccccc
tcagtgctggcctcagagcacagctctctccaagccaggccgcgcgccatccatcctccc
tgtcccccaacgtccttgccacagatcatgtccgccctgacacacatgggtctcagccat
ctctgccccagttaactccccatccataaagagcacatgccagctgacaccaaaataatt
cgggatggttccagtttagacctaagtggaaggagaaaccaccacctgccctgcaccttg
ttttttggtgaccttgataaaccatcttcagccatgaagccagctgtctcccaggaagct
ccagggcggtgcttcctcgggagctgactgataggtgggaggtggctgccccttgcacc
ctcaggtgaccccacacaaggccactgctggaggccctggggactccaggaatgtcaatc
agtgacctgcccccccaggccccacacagccatggctgcatagaggcctgcctccaaggga
cctgtctgtctgccactgtggagtccctacagcgtgcccccccacaggggagctggttctt
tgactgagatcagctggcagctcagggtcatcattcccagagggagcggtgccctggagg
ccacaggcctcctcatgtgtgtctgcgtccgctcgagcttactgagacactaaatctgtt
ggtttctgctgtgccacctacccaccctgttggtgttgctttgttcctattgctaaagac
aggaatgtccaggacactgagtgtgcaggtgcctgctggttctcacgtccgagctgctga
actccgctgggtcctgcttactgatggtctttgctctagtgctttccagggtccgtggaa
gcttttcctggaataaagcccacgcatcgaccctcacagcgcctcccctctttgaggccc
agcagataccccactcctgcctttccagcaagatttttcagatgctgtgcatactcatca
tattgatcacttttttcttcatgcctgattgtgatctgtcaatttcatgtcaggaaaggg
agtgacattttacacttaagcgtttgctgagcaaatgtctgggtcttgcacaatgacaa
tgggtccctgtttttcccagaggctcttttgttctgcagggattgaagacactccagtcc
cacagtccccagctcccctggggcagggttggcagaatttcgacaacacatttttccacc
ctgactaggatgtgctcctcatggcagctgggaaccactgtccaataagggcctgggctt
acacagctgcttctcattgagttacacccttaataaaataatcccatttttatccttttg
tctctctgtcttcctctctctctgcctttcctcttctctctcctcctctctcatctccag
GTGCAAATAGTCTACAAACCAGTTGACCTGAGCAAGGTGACCTCCAAGTGTGGCTCATTA
GGCAACATCCATCATAAACCAGgtagccctgtggaaggtgagggttgggacgggagggtg
caggggtggaggagtcctggtgaggctggaactgctccagacttcagaaggggctggaa
aggatattttaggtagacctacatcaaggaaagtgttgagtgtgaaacttgcgggagccc
aggaggcgtggtggctccagctcgctcctgcccaggccatgctgcccaagacaaggtgag
gcggga***gtgaagtgaaataaggcaggcacagaaagaaagcacatattctcggccgggcgc***
tgtggctcacgcctgtaattccagcactttgggaggccaaggtgggtggatcatgaggtc
aggagattgagaccatcctggctaacacagtgaaaccccgtctctactaaaaatacaaaa
aattagccgggcgtggtggtgggcgcctgtagtcccagctactccggaggctgaggcagg
aaaatggcgtgaacccggaaggcggagcttgcagtgagcggagtgagcagagatcgcgcc
actgcactccagcctgggcgacagagcgagactccgtctcaaaaaaaaaaaagcacatgtt
ctcgcttctttgtgggatccaggagatagagaatagaaggatggttaccagaggctggga
agggtagtgaggggatggtgggggatggtcaatgggtacaaaaaaaatagaataagacc
tagtatttgatagtgcaacagggtgactatagtcaataataatttaattgtacatttaaa
aataactaaaagatagccgggtgcagtggcttacgtctgtaatcccagtactttgggagg
ctgaggtgggcgtttgagaccagcctggccaacatggtgaaaccccatctctactaaaaa
tacaaaaattagccaggcatggtggcgggcgcctgtaatcccagctactcgggaggctga
ggcaggagaatcacttgaacctgggaggcagaggttgcagtgagccgagatcttgccact
gcactccagcctgggtgacagtgaaactccgtctcaaaaataaaaataaaaataca***gctg***
ggcacggtggctcacgcctgtaatcccagcactttgggaggccgaggcgagcggatcaca
aggtcaggagatatagaccatcctggctaacacggtgaaacccggtctctactaaaaata
caaaaaattagccaggcgtggtggcaggtgcctatagtcccagctactcacaaggctgag
gcaggagaatggcatgaacctgggaggcggagcttgcagtgagccgagattgtgccactg
cactccagcctgggcgagagagtgagactccgtctcaaaacaaaaacaaaaacaaaaaca
aaaacaaacacacaacaaaaacctaaaagaatataaatggattgtttgtaacacaaagga
caaatgtttgaggggatggataccccatttccatgatgtgattattatacattgtgtgt
ctgtatcaaaacatctcatgagccccataaatatatacacctaactatgtacccacaaaa
attaaaaaaatatatttttaaggtgaagagggaggcgagatgctggccttaaccctaa
cccgttgttctccctgcaagctgtccacagggcctctcagactcgaggttcagctatatg
gatgcatgagcttggtccccagccaacatgggagacacttcaccatcggcagcagctaca
gcacaggaaccctgggtcactgccatgtcccctctgtgactttgtttaaacagaaaatga
tgctctg***ggccggctgtggtggcccacacctataatcccagcaccttgggaggcgggggt***
gggcagattgcctgaggtcaggagttggagatcagcctggccgacatggcgaaaccccat

gtctactaaaaatacaaaaactagccaggcatggtggcacatgcctgtaatcccagctac ttgggaggctgaagcaggagaatcacttgaacccaggaggcagaggctgagtgagccaag atcgtgccaatgcactccagcttgggtgagggagtgagactccgtctcaaaaaaaaaaaa aaagaaagaaaaagaaaagaaagtgatcctactggaaccatgcttactcccctccccacc tcacactgtgtagaaattagtgctgtc***ggccaggcgcggtggctcatgcctgtaatcgca gcactttgggaggccaaggcaggcggatcacgaggtcaggagatcaagaccatcctggct aacacagtgaaaccctgtctctactaaaaatacaaaaaattagccgggcatggtggcagg cacctgtagtcccaactacttgggaggctgaggcaggagaatggcatgaacctgggaggc ggagcttgcagtgagccaagatcgcgccactgcataccagcctaggtgacagagtgagac tcagcaaaaaaagaaagaaagaaagaaagaaa***t***cagtgctgtctatacttctttctgcag tgatggaaatattctgtatctgtgctgtccagtatagtagccactagctacatgtggcac ttgaaacatggctggtacagttgaggaa***gagtggctgccatatcggacgacacagctata gattctgtcaccccaccccgagagtccagagcggggacttctgccttaggccctattcag ggctgatttttacttgaacccttactgtgggaagagaaggccatgagaagttcagtctag aatgtgactccttattttctggctcccttggacactttgtgggatttagtctccctgtgg aaagtattccacaagtggtgccaccaccccagctgtgagagcagctgggagctgctttg tcatctttccctggaaagtcctgtgggctgtctcttcctcatgccttgtcccatgcttgg gcatggtgtcaagcgtcaggagggagaaagggtcc***ttatttatttatttagagagggacc cttcttctgttcccaggctggagtgcagtggtgcgatctcggctcactgcaacctccgcc tcctgggttcaagtgattctcctgcctcagcctcctgagtagctgagattacaggcacat gccaacatgcctggctaatttttttttttttttttttttttttttgagatggagttg tactctcattgcccaggctggaatgtaatggcacaatctcggctcactgcaacctccacc tcctggattcaagcaattctcctgtctcagcttcccaagtagctgggattacaggtgccc gccaccatgctcaactaatttttgtatttttttttagtagagacgaggtttcaccatgt tggtcagactggtctcaaactcctgacctcaggtgatccacctgcctcggcctcccaaag tgctaggattacaggcatgagccaccacgcccggcc***tgaaagggttcttatttagtgtgc atttgacattcaatttaattccaaggtcttgtggggtcatggtttacaggatgttgata tagaaaagacttcacttaatg***ggccgggcgcagtggctcatgcctgtaatcccagcactt tgggaggccgaggcaggcagatcaggaggtcaggagattgagaccatcctggctaacaca gtgaaaccccatctctactgaaaatacaaaaaattagctgggcgtggtggcaggcacctg tagtcccagccactcggttggctgaggcaggagaatggcatgaacccgggaggcggagct tgcagtgagcagagaccatgccactgcactccagcctgggcgacagagcaagactctgtc tcaagaaaaaaaaaaaaaaaa***cagactttacttactggaagccaaccaatgtatatttag agtaatttttcct***gggctgagctgtcatttacttttgcagtatctcaagaagaagagttt acagtgtaaatatttgatgcacactttgattatatagatgaagcaaactattttcaagag ctttgcaaggacttacttgtatccaaacaccattctaaaggagtcttacctacttctaaa ggctggtctctacttggaaccacttgcttggccctggttcaagtcctgctgcaaacctgg aagt***cctgtcattgtcttcttccctccagagcagtggcacccaatctaatttttgctgtg ccccagcagcccctggcactttgccctgtagactgcagacctcatgtaatgtatgttaag tccacagaaccacagaagatgatggcaagatgctcttgtgtgtgttgtgttctag***GAGGT GGCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAGTCGAAGATT GGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAAAAG***gtaaaggggta gggtgggttggatgctgcccttgggtatatgggcattaatcaagttgagtggacaaaggc tggtccagttcccagaggaggaaaacagaggcttctgtgttgactggctggatgtgggcc ctcagcagcatc***cagtgggtctccactgcctgtctcaatcacctggagctttagcacgtt tcacacctgggccccaacctggagaggctgaccaatgggtctcaggggcagctcggttgc tggagttttt***gttttttatttatttttatgtatttaaggcagggtctc***tgtattagtccat tctcacactgctaataaagacatacccaagactgggtaatttataaaggaaagaggttta atggactcacagttccacatggctggggaggcctcaaaatcatggcggaaggcaaaggag aagcaaaggcatttcttacatggcgacaggcaagagagcgtgtgcaggggaactcccatt tataaaaccatcagacctcatgagatttattcactatcatgagaacagcatgggaaagac***

Exons are highlighted in color and repetitive elements are ***underlined, italic and bold***. Gray shading indicates flanking regions of low complexity that can be used to generate primers.

Supplemental Figure 4-3 Sequences of the planned reporter gene Vector

cccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttccttgaccctggaaggtg
ccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggggtggggtggg
gcaggacagcaaggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcgga
aagaaccagctggggctctagggggtatccccacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgc
agcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttcccc
gtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatg
gttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaa
actggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctg
atttaacaaaaatttaacgcgaattaattctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcaga
agtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaag
catgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgcccagttccgcccattctc
cgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgaggaggct
tttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcaagagacaggatgaggatcgtt
tcgcatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaaca
gacaatcggctgctctgatgccgccgtgttccggctgtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggt
gccctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacg
ttgtcactgaagcgggaagggactggctgctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccga
gaaagtatccatcatggctgatgcaatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaaca
tcgcatcgagcgagcacgtactcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgc
cagccgaactgttcgccaggctcaaggcgcgcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgcc
gaatatcatggtggaaaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcg
ttggctacccgtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattc
gcagcgcatcgccttctatcgccttcttgacgagttcttctgagcgggactctggggttcgaaatgaccgaccaagcgacgccca
acctgccatcacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatg
atcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaata
gcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtatac
cgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacata
cgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgcttt
ccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttc
cgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtta
tccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgc
gttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgac
aggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgt
ccgcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagct
gggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagaca
cgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtg
gtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggta
gctcttgatccggcaaacaaaccaccgctggtagcggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaa
gaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaa
ggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatg
cttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatac

gggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacc
agccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagcta
gagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggc
ttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctcc
gatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaa
gatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaa
tacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggat
cttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggt
gagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttttc
aatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgc
gcacatttccccgaaaagtgccacctgacgtcgacggatcgggagatctcccgatcccctatggtgcactctcagtacaatctgc
tctgatgccgcatagttaagccagtatctgctccctgcttgtgtgttggaggtcgctgagtagtgcgcgagcaaaatttaagctaca
acaaggcaaggcttgaccgacaattgcatgaagaatctgcttagggttaggcgttttgcgctgcttcgcgatgtacgggccagat
atacgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgtta
cataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagt
aacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatat
gccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttccta
cttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttga
ctcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcg
taacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaactag
agaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcgtttaaacttaagctt

Exon 9 fragment

acgctggccgcagggattataattatttccattttcaaattaaggcctctgagctcagagaggggaagttacttgtctgaggccaca
cagcttgttggagcccatctcttgacccaaagactgtggagccgagttggccacctctctgggagcgggtattggatggtggttg
atggttttccattgctttcctgggaaaggggtgtctctgtccctaagcaaaaaggcagggaggaagagatgcttccccagggcag
ccgtctgctgtagctgcgcttccaacctggcttccacctgcctaacccagtggtgagcctgggaatggacccacgggacaggca
gcccccagggccttttctgaccccacccactcgagtcctggcttcactcccttccttccttcccagGTGAACCTCCAAA
ATCAGGGGATCGCAGCGGCTACAGCAGCCCCGGCTCCCCAGGCACTCCCGGC
AGCCGCTCCCGCACCCCGTCCCTTCCAACCCCACCCACCCGGGAGCCCAAGA
AGGTGGCAGTGGTCCGTACTCCACCCAAGTCGCCGTCTTCCGCCAAGAGCCG
CCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGAAGAATGTCAAGTCCAAG
ATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGCGGGAAGgtgagagtgg
ctggctgcgcgtggaggtgtgggggctgcgcctggaggggtagggctgtgcctggaagggtagggctgcgcctggaggtg
cgcggttgagcgtggagtcgtgggactgtgcatggaggtgtggggctccccgcacctgagcacccccgcataacaccccagt
cccctctggaccctcttcaaggaagttcagttctttattgggctctccactacactgtgagtgccctcctcaggcgagagaacgttct
ggctcttctcttgccccttcagcccctgttaatcggacagagatggcagggctgtgtctccacggccggaggctctcatagtcag
ggcacccacagcggttccccacctgccttctgggcagaatacactgccacccataggtcagcatctccactcgtgggcc

Exon 10 fragment

aaaaatccacaggtgattctgatgcccggcaggcttgagaacagccgcagggagttctctgggaatgtgccggtgggtctagc
caggtgtgagtggagatgccggggaacttcctattactcactcgtcagtgtggccgaacacatttttcacttgacctcaggctggt
gaacgctcccctctggggttcaggcctcacgatgccatccttttgtgaagtgaggacctgcaatcccagcttcgtaaagcccgct

ggaaatcactcacacttctgggatgccttcagagcagccctctatcccttcagctcccctgggatgtgactcgacctcccgtcact
ccccagactgcctctgccaagtccgaaagtggaggcatccttgcgagcaagtaggcgggtccagggtggcgcatgtcactcat
cgaaagtggaggcgtccttgcgagcaagcaggcgggtccagggtggcgtgtcactcatccttttttctggctaccaaagGTG
CAGATAATTAATAAGAAGCTGGATCTTAGCAACGTCCAGTCCAAGTGTGGCT
CAAAGGATAATATCAAACACGTCCCGGGAGGCGGCAGTgtgagtaccttcacacgtcccat
gcgccgtgctgtggcttgaattattaggaagtggtgtgagtgcgtacacttgcgagacactgcatagaataaatccttcttgggctc
tcaggatctggctgcgacctctgggtgaatgtagcccggctccccacattcccccacacggtccactgttcccagaagccccttc
ctcatattctaggaggggtgtcccagcatttctgggtcccccagcctgcgcaggctgtgtggacagaatagggcagatgacgg
accctctctccggaccctgcctgggaagctgagaatacccatcaaagtctccttccactcatgcccagccctgtccccaggagcc
ccatagcccattggaagttgggctgaaggtggtggcacctgagactgggctgccgcagataccccactc

Exon 11 fragment

tgcctttccagcaagatttttcagatgctgtgcatactcatcatattgaccacttttttcttcatgcctgattgtgatctgtcaatttcatgt
caggaaagggagtgacatttttacacttaagcgtttgctgagcaaatgtctgggtcttgcacaatgacaatgggtccctgtttttccc
agaggctcttttgttctgcagggattgaagacactccagtcccacagtccccagctcccctggggcagggttggcagaatttcga
caacacatttttccaccctgactaggatgtgctcctcatggcagctgggaaccactgtccaataagggcctgggcttacacagctg
cttctcattgagttacacccttaataaaataatcccatttttatccttttgtctctctgtcttcctctctctctgcctttcctcttctctcctc
ctctctcatctccagGTGCAAATAGTCTACAAACCAGTTGACCTGAGCAAGGTGACCT
CCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGgtagccctgtggaaggtgagg
gttgggacgggaaggtgcaggggtggaggagtcctggtgaggctggaactgctccagacttcagaaggggctggaaagga
tattttaggtagacctacatcaaggaaagtgttgagtgtgaaacttgcgggagcccaggaggcgtggtggctccagctcgctcct
gcccaggctatgctgcccaagacaaggtgaggcgggagtgaagtgaaataaggcaggcacagaaagaaagcacatattctc
ggccgggcgctgtggctcacgcctgtaatcccagcactttgggaggccaaggtgggtggatcatgaggtcaggagattgagac
catcctggctaacacagtgaaaccccgtctctactaaaaatacaaaaaattagccgggcgtggtggcgggcgcctgtagtccca
gctact

exon 12 fragment

gtggggtcatggtttacaggatgttgatatagaaaagacttcacttaatgggccgggcgcagtggctcatgcctgtaatcccagca
ctttgggaggccgaggcaggcagatcaggaggtcaggagattgagaccatcctggctaacacagtgaaaccccatctctactg
aaaatacaaaaaattagctgggcgtggtggcaggcacctgtagtcccagccactcggttggctgaggcaggagaatggcatga
acccgggaggcggagcttgcagtgagcagagaccatgccactgcactccagcctgggcgacagagcaagactctgtctcaa
gaaaaaaaaaaaaaaaacagactttacttactggaagccaaccaatgtatatttagagtaatttttcctgggctgagctgtcatttact
tttgcagtatctcaagaagaagagtttacagtgtaaatatttgatgcacactttgattatatagatgaagcaaactattttcaagagcttt
gcaaggacttacttgtatccaaacaccattctaaaggagtcttacctacttctaaaggctggtctctacttggaaccacttgcttggc
cctggttcaagtcctgctgcaaacctggaagtcctgtcattgtcttcttccctccagagcagtggcacccaatctaatttttgctgtgc
cccagcagcccctggcactttgccctgtagactgcagacctcatgtaatgtatgttaagtccacagaaccacagaagatgatggc
aagatgctcttgtgtgtgttgtgttctagGAGGTGGCCAGGTGGAAGTAAAATCTGAGAAGCTT
GACTTCAAGGACAGAGTCCAGTCGAAGATTGGGTCCCTGGACAATATCACCC
ACGTCCCTGGCGGAGGAAATAAAAAGgtaaaggggtagggtgggttggatgctgcccttgggtatat
gggcattaatcaagttgagtggacaaaggctggtccagttcccagaggaggaaaacagaggcttctgtgttgactggctggatg
tgggccctcagcagcatccagtgggtctccactgcctgtctcaatcacctggagctttagcacgtttcacacctgggccccaacct
ggagaggctgaccaatgggtctcaggggcagctcggttgctggagttttgttttttatttatttttatgtatttaaggcagggtctctgt
attagtccattctcacactgctaataaagacatacccaagactgggtaatttataaaggaaagaggtttaatggactcacagttcca

catggctggggaggcctcaaaatcatggcggaaggcaaaggagaagcaaaggcatttcttacatggcgacaggcaagagag
cgtgtgcaggggaactcccatttataaaaccatcagacctcatgagatttattcactatcatgagaacagcatgggaaagacccg
cccccatgattcagttacctcccactgggtccctcccatgacacatggaattatgggagctacaattcaagatgagatttgggtgg
ggacacagccaaaccatatcagtctccctctgtcatccaggctggagtgcactggcatgatctcggctcactgcagcctctacct
ccctgggtcaggtgatcttcccacctcagcctcccaggtagctggaactacaggtacctgccactatgcctggctaaatattttgta
tttcctgtggagacgaggttttgccacgttgcccaggctggtcttgaactcctgaggtcaagcaatatgcccacctcggcctccca
aggtgctgggattacaggtgtgagccacagtgctcggcctaagtcactgcagtttctcgagtctagagggcccgtttaaa

The vector sequence and the planned genomic fragments are shown.

Supplemental Figure 4-4 Primer design for assembly

New Assembly

Created: 1/15/2019, 2:37:35 PM
Saved: not saved

Component Fragments

Name	Length	Produced by	5' End	3' End
pcDNA3.1	5360	PCR	Fwd Primer (auto)	Rev Primer (auto)
exon 9	1124	PCR	Fwd Primer (auto)	Rev Primer (auto)
exon 10	1053	PCR	Fwd Primer (auto)	Rev Primer (auto)
exon 11	1011	PCR	Fwd Primer (auto)	Rev Primer (auto)
exon 12	1920	PCR	Fwd Primer (auto)	Rev Primer (auto)

Notes

- A 60 minute reaction is recommended for the assembly of more than 3 fragments.
- Primer pcDNA3.1_fwd has %GC outside of desired range (35-65%) in the annealing segment.
- Primer exon 9_fwd has %GC outside of desired range (35-65%) in the annealing segment.
- Primer exon 9_rev contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primer exon 9_rev has %GC outside of desired range (35-65%) in the annealing segment.
- Primer exon 10_fwd contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primer exon 10_rev contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primer exon 10_rev has %GC outside of desired range (35-65%) in the annealing segment.
- Primer exon 11_fwd contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primer exon 11_rev contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primer exon 12_fwd contains a run of 4+ repeats of a mono/di/trinucleotide.
- Primers for fragment pcDNA3.1 have internal complementarity.

Required oligos

Name	Primer 5' (overlap/spacer/ANNEAL) 3'	Len	%GC	3' %GC	3' Tm	3' Ta
pcDNA3.1_fwd	gggcccgtttaaaCCCGCTGATCAGCCTCGA	31	61	67	70.3	68.4
pcDNA3.1_rev	tgcggccagcgtAAGCTTAAGTTTAAACGCTAGCCAGC	38	53	42	67.4	68.4
exon 9_fwd	taaacttaagcttACGCTGGCCGCAGGGATT	31	48	67	73.8	72.0
exon 9_rev	cctgtggatttttGGCCCACGAGTGGAGATGC	32	56	68	71.6	72.0
exon 10_fwd	cactcgtgggccAAAAATCCACAGGTGATTCTGATGC	37	51	40	65.4	66.4
exon 10_rev	tgctggaaaggcaGAGTGGGGTATCTGCGGC	31	61	67	68.6	66.4
exon 11_fwd	gataccccactcTGCCTTTCCAGCAAGATTTTTC	34	47	41	63.6	63.6
exon 11_rev	accatgaccccacAGTAGCTGGGACTACAGG	31	58	56	62.6	63.6
exon 12_fwd	gtcccagctactGTGGGGTCATGGTTTACAG	31	55	53	62.9	62.0

NEBuilder Assembly Tool v2.0.7 https://nebuilder.neb.com 1/15/2019, 2:50:06 PM

exon 12_rev	ctgatcagcgggTTTAAACGGGCCCTCTAG	30	57	50	61.0	62.0

Build Settings

Property	Value
Product/Kit	#E5520 NEBuilder HiFi DNA Assembly Cloning Kit
Minimum Overlap	25 nt
Minimum Overlap Tm	48 °C
Circularize	Yes
PCR Polymerase/Kit	Q5 High-Fidelity DNA Polymerase
PCR Primer Conc.	500 nM
Min. Primer Length	18 nt

The sequence from Supplemental Figure 4.3 was entered into the builder tool.

Supplemental Figure 4-5 Sequence of the tau 9→12 reporter gene used as an example.

GTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATC
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGG
CGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTT
ACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAG
CTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCT
GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTAT
CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCA
CTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTT
GAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGC
GCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
GCAAACAAACCACCGCTGGTAGCGGTTTTTTTGTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCT
GACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTAT
CAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAG
TGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGAC
TCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAG
TGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCA
ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTAT
CCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCG
CCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTC
ACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGC
GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGC
AGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGA
CTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAG
TTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACT
TTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGA
TCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGA
TCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
GCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACT
CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCG
CGCACATTTCCCCGAAAAGTGCCACCTGACGTCGACGGATCGGGAGATCTCC
CGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAA
GCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGC
AAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATC
TGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACG
CGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATT
AGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCC
CGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTA

TGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT
ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGT
ACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC
ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATA
GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA
GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCC
GCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA
GCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATT
AATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCT
TACGCTGGCCGCAGGGATTATAATTATTTCCATTTTCAAATTAAGGCCTCTGA
GCTCAGAGAGGGGAAGTTACTTGTCTGAGGCCACACAGCTTGTTGGAGCCCA
TCTCTTGACCCAAAGACTGTGGAGCCGAGTTGGCCACCTCTCTGGGAGCGGG
TATTGGATGGTGGTTGATGGTTTTCCATTGCTTTCCTGGGAAAGGGGTGTCTC
TGTCCCTAAGCAAAAAGGCAGGGAGGAAGAGATGCTTCCCCAGGGCAGCCG
TCTGCTGTAGCTGCGCTTCCAACCTGGCTTCCACCTGCCTAACCCAGTGGTGA
GCCTGGGAATGGACCCACGGGACAGGCAGCCCCCAGGGCCTTTTCTGACCCC
ACCCACTCGAGTCCTGGCTTCACTCCCTTCCTTCCTTCCCAGGTGAACCTCCA
AAATCAGGGGATCGCAGCGGCTACAGCAGCCCCGGCTCCCCAGGCACTCCCG
GCAGCCGCTCCCGCACCCCGTCCCTTCCAACCCCACCCACCCGGGAGCCCAA
GAAGGTGGCAGTGGTCCGTACTCCACCCAAGTCGCCGTCTTCCGCCAAGAGC
CGCCTGCAGACAGCCCCCGTGCCCATGCCAGACCTGAAGAATGTCAAGTCCA
AGATCGGCTCCACTGAGAACCTGAAGCACCAGCCGGGAGGCGGGAAGGTGA
GAGTGGCTGGCTGCGCGTGGAGGTGTGGGGGGCTGCGCCTGGAGGGGTAGG
GCTGTGCCTGGAAGGGTAGGGCTGCGCCTGGAGGTGCGCGGTTGAGCGTGGA
GTCGTGGGACTGTGCATGGAGGTGTGGGGCTCCCCGCACCTGAGCACCCCCG
CATAACACCCCAGTCCCCTCTGGACCCTCTTCAAGGAAGTTCAGTTCTTTATT
GGGCTCTCCACTACACTGTGAGTGCCCTCCTCAGGCGAGAGAACGTTCTGGC
TCTTCTCTTGCCCCTTCAGCCCCTGTTAATCGGACAGAGATGGCAGGGCTGTG
TCTCCACGGCCGGAGGCTCTCATAGTCAGGGCACCCACAGCGGTTCCCCACC
TGCCTTCTGGGCAGAATACACTGCCACCCATAGGTCAGCATCTCCACTCGTGG
GCCAAAAATCCACAGGTGATTCTGATGCCCGGCAGGCTTGAGAACAGCCGCA
GGGAGTTCTCTGGGAATGTGCCGGTGGGTCTAGCCAGGTGTGAGTGGAGATG
CCGGGGAACTTCCTATTACTCACTCGTCAGTGTGGCCGAACACATTTTTCACT
TGACCTCAGGCTGGTGAACGCTCCCCTCTGGGGTTCAGGCCTCACGATGCCAT
CCTTTTGTGAAGTGAGGACCTGCAATCCCAGCTTCGTAAAGCCCGCTGGAAA
TCACTCACACTTCTGGGATGCCTTCAGAGCAGCCCTCTATCCCTTCAGCTCCC
CTGGGATGTGACTCGACCTCCCGTCACTCCCCAGACTGCCTCTGCCAAGTCCG
AAAGTGGAGGCATCCTTGCGAGCAAGTAGGCGGGTCCAGGGTGGCGCATGTC
ACTCATCGAAAGTGGAGGCGTCCTTGCGAGCAAGCAGGCGGGTCCAGGGTGG
CGTGTCACTCATCCTTTTTTCTGGCTACCAAAGGTGCAGATAATTAATAAGAA
GCTGGATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAA
CACGTCCCGGGAGGCGGCAGTGTGAGTACCTTCACACGTCCCATGCGCCGTG

CTGTGGCTTGAATTATTAGGAAGTGGTGTGAGTGCGTACACTTGCGAGACAC
TGCATAGAATAAATCCTTCTTGGGCTCTCAGGATCTGGCTGCGACCTCTGGGT
GAATGTAGCCCGGCTCCCCACATTCCCCCACACGGTCCACTGTTCCCAGAAG
CCCCTTCCTCATATTCTAGGAGGGGGTGTCCCAGCATTTCTGGGTCCCCCAGC
CTGCGCAGGCTGTGTGGACAGAATAGGGCAGATGACGGACCCTCTCTCCGGA
CCCTGCCTGGGAAGCTGAGAATACCCATCAAAGTCTCCTTCCACTCATGCCCA
GCCCTGTCCCCAGGAGCCCCATAGCCCATTGGAAGTTGGGCTGAAGGTGGTG
GCACCTGAGACTGGGCTGCCGCAGATACCCCACTCCTGCCTTTCCAGCAAGA
TTTTTCAGATGCTGTGCATACTCATCATATTGACCACTTTTTTCTTCATGCCTG
ATTGTGATCTGTCAATTTCATGTCAGGAAAGGGAGTGACATTTTTACACTTAA
GCGTTTGCTGAGCAAATGTCTGGGTCTTGCACAATGACAATGGGTCCCTGTTT
TTCCCAGAGGCTCTTTTGTTCTGCAGGGATTGAAGACACTCCAGTCCCACAGT
CCCCAGCTCCCCTGGGGCAGGGTTGGCAGAATTTCGACAACACATTTTTCCAC
CCTGACTAGGATGTGCTCCTCATGGCAGCTGGGAACCACTGTCCAATAAGGG
CCTGGGCTTACACAGCTGCTTCTCATTGAGTTACACCCTTAATAAAATAATCC
CATTTTATCCTTTTTGTCTCTCTGTCTTCCTCTCTCTCTGCCTTTCCTCTTCTCTC
TCCTCCTCTCTCATCTCCAGGTGCAAATAGTCTACAAACCAGTTGACCTGAGC
AAGGTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGGTA
GCCCTGTGGAAGGTGAGGGTTGGGACGGGAAGGTGCAGGGGGTGGAGGAGT
CCTGGTGAGGCTGGAACTGCTCCAGACTTCAGAAGGGGCTGGAAAGGATATT
TTAGGTAGACCTACATCAAGGAAAGTGTTGAGTGTGAAACTTGCGGGAGCCC
AGGAGGCGTGGTGGCTCCAGCTCGCTCCTGCCCAGGCTATGCTGCCCAAGAC
AAGGTGAGGCGGGAGTGAAGTGAAATAAGGCAGGCACAGAAAGAAAGCAC
ATATTCTCGGCCGGGCGCTGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
GCCAAGGTGGGTGGATCATGAGGTCAGGAGATTGAGACCATCCTGGCTAACA
CAGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGG
CGGGCGCCTGTAGTCCCAGCTACTGTGGGGTCATGGTTTACAGGATGTTGATA
TAGAAAAGACTTCACTTAATGGGCCGGGCGCAGTGGCTCATGCCTGTAATCC
CAGCACTTTGGGAGGCCGAGGCAGGCAGATCAGGAGGTCAGGAGATTGAGA
CCATCCTGGCTAACACAGTGAAACCCCATCTCTACTGAAAATACAAAAAATT
AGCTGGGCGTGGTGGCAGGCACCTGTAGTCCCAGCCACTCGGTTGGCTGAGG
CAGGAGAATGGCATGAACCCGGGAGGCGGAGCTTGCAGTGAGCAGAGACCA
TGCCACTGCACTCCAGCCTGGGCGACAGAGCAAGACTCTGTCTCAAGAAAAA
AAAAAAAAAAACAGACTTTACTTACTGGAAGCCAACCAATGTATATTTAGAG
TAATTTTTCCTGGGCTGAGCTGTCATTTACTTTTGCAGTATCTCAAGAAGAAG
AGTTTACAGTGTAAATATTTGATGCACACTTTGATTATATAGATGAAGCAAAC
TATTTTCAAGAGCTTTGCAAGGACTTACTTGTATCCAAACACCATTCTAAAGG
AGTCTTACCTACTTCTAAAGGCTGGTCTCTACTTGGAACCACTTGCTTGGCCC
TGGTTCAAGTCCTGCTGCAAACCTGGAAGTCCTGTCATTGTCTTCTTCCCTCC
AGAGCAGTGGCACCCAATCTAATTTTTGCTGTGCCCCAGCAGCCCCTGGCACT
TTGCCCTGTAGACTGCAGACCTCATGTAATGTATGTTAAGTCCACAGAACCAC
AGAAGATGATGGCAAGATGCTCTTGTGTGTGTTGTGTTCTAGGAGGTGGCCA
GGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAGTCGAAG

ATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAAAAGG
TAAAGGGGGTAGGGTGGGTTGGATGCTGCCCTTGGGTATATGGGCATTAATC
AAGTTGAGTGGACAAAGGCTGGTCCAGTTCCCAGAGGAGGAAAACAGAGGC
TTCTGTGTTGACTGGCTGGATGTGGGCCCTCAGCAGCATCCAGTGGGTCTCCA
CTGCCTGTCTCAATCACCTGGAGCTTTAGCACGTTTCACACCTGGGCCCCAAC
CTGGAGAGGCTGACCAATGGGTCTCAGGGGCAGCTCGGTTGCTGGAGTTTTT
GTTTTTATTTATTTTTATGTATTTAAGGCAGGGTCTCTGTATTAGTCCATTCTC
ACACTGCTAATAAAGACATACCCAAGACTGGGTAATTTATAAAGGAAAGAGG
TTTAATGGACTCACAGTTCCACATGGCTGGGGAGGCCTCAAAATCATGGCGG
AAGGCAAAGGAGAAGCAAAGGCATTTCTTACATGGCGACAGGCAAGAGAGC
GTGTGCAGGGGAACTCCCATTTATAAAACCATCAGACCTCATGAGATTTATTC
ACTATCATGAGAACAGCATGGGAAAGACCCGCCCCCATGATTCAGTTACCTC
CCACTGGGTCCCTCCCATGACACATGGAATTATGGGAGCTACAATTCAAGAT
GAGATTTGGGTGGGGACACAGCCAAACCATATCAGTCTCCCTCTGTCATCCA
GGCTGGAGTGCACTGGCATGATCTCGGCTCACTGCAGCCTCTACCTCCCTGGG
TCAGGTGATCTTCCCACCTCAGCCTCCCAGGTAGCTGGAACTACAGGTACCTG
CCACTATGCCTGGCTAAATATTTTGTATTTCCTGTGGAGACGAGGTTTTGCCA
CGTTGCCCAGGCTGGTCTTGAACTCCTGAGGTCAAGCAATATGCCCACCTCG
GCCTCCCAAGGTGCTGGGATTACAGGTGTGAGCCACAGTGCTCGGCCTAAGT
CACTGCAGTTTCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCG
ACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCC
TTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGC
AGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATG
CGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGG
GTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTT
ACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC
TTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAA
TCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA
AAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGAC
GGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTT
CCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAG
GGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAA
TTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCC
CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAG
CAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAA
GCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATC
CCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAAT
TTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGA
AGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGG
AGCTTGTATATCCATTTTCGGATCTGATCAAGAGACAGGATGAGGATCGTTTC
GCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGA
GAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCC

GCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGA
CCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGG
CTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAG
CGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTC
ATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGC
GGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACA
TCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGAT
GATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGC
TCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGC
CTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACT
GTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCG
TGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTT
ACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGAC
GAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCC
CAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGG
GCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGA
TCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATG
GTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCA
CTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
ATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGC
ATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTG
CGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCAT
TAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATA
ACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTA
AAAAGGCCGC

CHAPTER 5. FUNCTION OF *MAPT* CIRCULAR RNAS

5.1 Introduction

After finding that the human *MAPT* gene generates circular RNAs due to the processing of the pre-mRNA (43) (Figure 5.1 A-C), I started to investigate their functional characteristics. The circular tau RNA 12→7 contains one in-frame start codon in exon 9 and no stop codon, as mentioned in chapter 3 (Figure 3.1 C). We found that the circular tau RNA 12→7 is translated in a rolling circle giving rise to a multimeric protein. The 12→7 protein could then potentially promote tau aggregation and neurofilament tau tangles, causing neurodegeneration and tauopathies. Since tau is one of the major contributors to Alzheimer's disease and other neurodegenerative diseases, this mechanistic approach would be a novel pathway for the disease pathology. The protein generated from the 12→7 circular RNA may play a role under normal physiological conditions in the brain.

The circular RNA 12→10 does not contain a start codon or a stop codon. However, it has been found that 53 mutations in the *MAPT* gene give rise to hereditary frontotemporal dementia (79). The majority of these mutations are located in exons 9-13, regions that control circular RNA formation, where two specific mutations create methionine start codons in exon 11 (K317M) and exon 12 (V337M). These mutations would create a start codon in the circular RNA 12→10 causing translation to occur. We show that these mutations allow this circular RNA to be translated, and the encoded protein could potentially lead to tau aggregation or neurofilament tangles.

It is not clear how neurofibrillary tau tangles (NFT) develop. It was shown that the 'K18' peptide encompassing tau repeats R1-R4, acts as a 'seed peptide', promoting tau aggregation into paired helical filaments (6) and neurofibrillary tau tangles (NFT) (7). Strikingly, the proteins encoded by tau circular RNA encode R1-R4 (tau circ 12→7) or R2-R4 (tau circ12→10). Thus, tau multimers encoded by circular RNAs could act as a seed sequence that promotes tau aggregation and NFT formation, similar to the structurally related 'K18' peptide (Figure 5.1 D, E).

RNA editing enzymes, known as adenosine deaminase acting on RNA (ADAR), can also contribute to circular tau RNA translation. The catalytically active enzymes, ADAR1 and ADAR2, edit adenosines to inosines (A-to-I), where the inosine will be read as a guanosine. The sequence AUA could be edited to AUI and be read as an AUG start codon. We show that the 12→10 circular RNA that does not have a start codon is translated when the catalytically active ADAR enzymes are overexpressed. It has been shown that a deregulation of RNA editing can contribute to Alzheimer's disease (135). The deregulation of the ADAR enzymes may act on the tau circular RNAs and the circular RNAs translated proteins could potentially contribute to tauopathies.

5.2 Experimental Approach

Two systems were utilized to test for circular RNA expression and protein translation. First, we tested endogenous tau circular RNA expression, in different areas of human brain tissue. The protein could not be tested, as there is no antiserum specific for the protein generated by the circular RNA versus the linear. Antiserum targeting the peptide junction specific for the circular RNA protein and not the linear RNA protein was made but did not work. The second system used was Human Embryonic Kidney 293T cells (HEK 293T). Human embryonic kidney 293T cells were stably or transiently transfected to express the tau circular RNAs 12→7 and 12→10.

In order to test protein expression, I used an expression vector that contains Alu components from the Zinc finger protein with KRAB and SCAN domains 1 (ZKSCAN1) gene. ZKSCAN1 encodes a strongly expressed circular RNA. The processing of this circular RNA is due to the short intronic repeat elements, known as Alu elements, that have complementary sequences that bind and bring the exons into close proximity (45). The ZKSCAN1 Alu elements were cloned into our tau cDNA constructs that contained exons 7, 9-12 and exons 10-12 (Figure 5.2 A-D, Supplemental Figure 5.1-5.4, 5.8). In order to fully recapitulate the human system and make the tau expression constructs more authentic, I removed the ZKSCAN1 intronic regions and inserted a shortened sequence of the tau intron. The tau intronic regions contain the tau Alu element and repeat sequences in both the left and right intron (Figure 5.3 A-C, Supplemental Figure 5.5-5.8).

The ZKSCAN1 model has a more robust expression than the model with authentic tau Alu. Thus, I used the ZKSCAN1 model for mass spectrometry purposes to have enough protein for detection.

5.3 Materials and Methods

5.4 Tau Primers for cloning

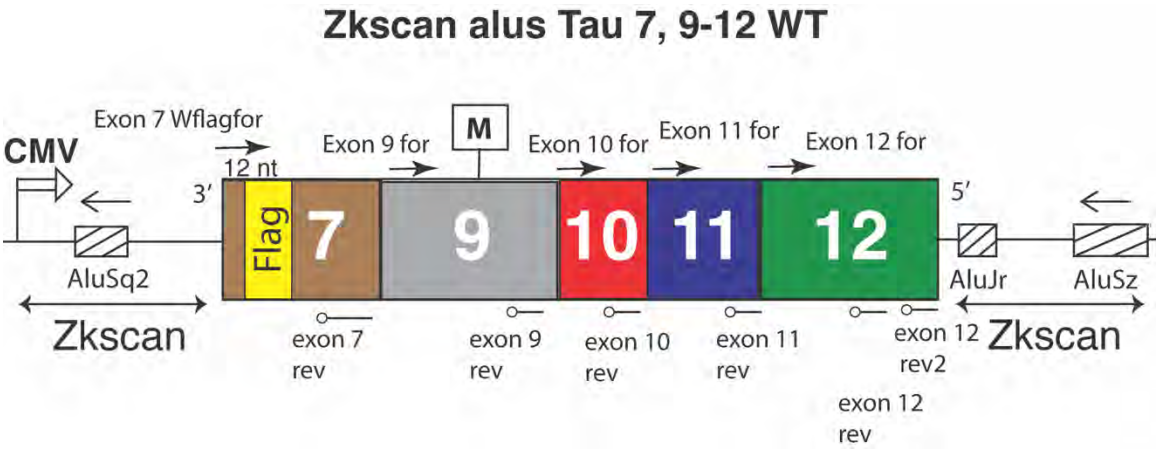
Primer Name	Primer Sequence
pcDNA 9-12_fwd	AGGGAGTTCTCTGGGAATGTG
pcDNA 9-12_rev	GGCCCACGAGTGGAGATG
JB and MER5A_fwd	agcatctccactcgtgggccCATAGTCTTTTGAAGGAA CATAAAAGATTATGAAGAAATG
JB and MER5A_rev	acattcccagagaactccctGCGGCTGTTCTCAAGCCTG
V337_to_M_F	AGGTGGCCAGaTGGAAGTAAA
V337_to_M_R	CCTGGTTTATGATGGATGTTG
K_TO_M Mut_F	GACCTGAGCAtGGTGACCTCC
K_TO_M Mut_R	AACTGGTTTGTAGACTATTTGC

7_12_both_auth_alu_vec_fwd	GGGGGCTGATGGTGACTACAAAGACCATG
7_12_both_auth_alu_vec_rev	GTTTCCTTACTTCTAAATACATTCAAATATGT ATCC
7_12_both_aut_alu_insert_fwd	GTATTTAGAAGTAAGGAAACCACCTTTG
7_12_both_aut_alu_insert_rev	TGTAGTCACCATCAGCCCCCTGTAAATG

5.5 Tau Sequencing primers

Primer Name	**Primer Sequence**
JB_MER5Adetect_for	ccacccataggtcagcatct
JB_MER5Adetect_rev	gcctgacttctggttcatcc
Zkscan forward	GAATTCAGTGACAGTGGAG
ZKscan rev	TAATGCCTACCAGGTTCATTAC
KM_mut_seq_for	ACA CGT CCC GGG AGG C
Tau exon 7 rev	CTGCGATCCCCTGATTTTGG
Tau exon 11 for	GTGTGCAAATAGTCTACAAACCA
Tau exon 12 rev	AGGGACCCAATCTTCGACTG
Exon 12 forward	GAG GTG GCC AGG TGG AA
Exon 10 forward	GTG CAG ATA ATT AAT AAG AAG CTG
Exon 10 rev	CTG CCG CCT CCC GGG
Exon 11 reverse	TGG TTT ATG ATG GAT GTT GCC T
Exon 7Wflag_for	GGG GCT GAT GGT GAC TAC AA
Exon 9 forward	GTG AAC CTC CAA AAT CAG GG
Exon 9 reverse	CTT CCC GCC TCC CGG C

5.6 Tau Primer location



5.7 Cell Culture

Human Embryonic Kidney Cells (HEK) 293T (ATCC, CRL-3216) were cultured in Dulbecco's Modified Eagles Medium – High Glucose (DMEM)(Sigma, D5796) with 10% fetal bovine serum (Azer Scientific, ES56817). Cells were grown in a humidified incubator at 37˚C with 5% CO_2. Stable cell lines were selected and maintained by adding 1µg/ml of Hygromycin B (Sigma-Aldrich, 10843555001) to the cell culture media.

5.8 Generation of Stable Cells

HEK 293T cells were co-transfected with 5 µg of the pcDNA3.1 expression vector that contained the tau minigenes, with 0.33µgs of linear hygromycin (Takara Bio, 631625). The DNA was combined and mixed in 200µl sterile-filtered 150 mM NaCl (Fisher Scientific, BP-358-212), then Linear Polyethyleneimine Hydrochloride (PEI)(Polysciences, 24765-1) was added to the DNA mix at 3µg/µg (PEI/DNA). The DNA was incubated for 10 mins then added directly to cells in 100mm cell culture dish (Azer Scientific, ES56269). The media was changed after 24 hrs. The cells were then split at 25% confluency with Trypsin-EDTA (0.25%), phenol red (Fisher Scientific, 25200114), into a multiwell cell culture plate. The cells were selected by adding 1µg/ml of Hygromycin B (Sigma-Aldrich, 10843555001) to the cell culture media and then expanded.

5.9 Transient Transfection

Transient transfection was performed in HEK 293T cells. The controls and tau expression constructs for this cell line were in a pcDNA 3.1 vector. 2µg of DNA was mixed with 200 µl of sterile 150 mM NaCl (Fisher Scientific, BP-358-212) and 6 µl of Linear Polyethyleneimine Hydrochloride (PEI) (Polysciences, 24765-1). The DNA was incubated at room temperature for 10 mins and then added to the cells. The cells were cultured in a 150 mm dish (Azer Scientific, ES56268), and transfected at 60% confluency. The cells were lysed and analyzed 96 hrs post transfection.

5.10 siRNA Transfection

Transient transfection was performed in HEK 293T cells. Cells were plated in a 6-well plate (VWR, 10062-892) at a density of 300,000 cells per well. All of the samples had 200 ng of eGFP for transfection efficiency and loading control. One microgram of the tau expression construct, zkscan 7-12 WT, was mixed with 150 µl of sterile Opti-MEM medium (ThermoFisher Scientific, 31985062). The 12to7 junction siRNA was added to the mix at different concentrations, 0.40 µM, 2 µM, and 4 µM, and then vortexed and incubated at room temperature for 5 minutes. In another tube, 150 µl of sterile Opti-MEM medium and 10 µl of Lipofectamine 2000 (Fisher Scientific, 11668030) were mixed. The two mixtures were combined, vortexed and incubated for 10 minutes at room temperature. After 10 minutes, the media was removed, and the DNA mixture was added to the cells. The cells were incubated for 20 minutes at 37˚C 5% CO_2. After 20 minutes, 2 ml of DMEM 10% FBS media were added to the cells and incubated for 24 hrs. After 24hrs, the media was replaced with fresh media for 48 hrs. After 48 hrs, the cells are lysed and analyzed for protein. The same was done for the zkscan 10-12 VM expression construct and 12to10 junction siRNAs. The siRNAs sequences are shown in Table 5.1.

5.11 Immunoprecipitation

The cells were spun at 14,000g at 4˚C for 10 mins to pellet. The cell pellet was washed 3 times with ice-cold phosphate buffered saline (PBS) (Sigma-Aldrich, 59331C-1000ML) to remove culture medium contaminants. The cell pellets were then resuspended in ice-cold lysis buffer: 50 mM Tris-HCL, pH 7.4 at 4˚C (Sigma-Aldrich, 648315-100ML), 150 mM NaCl (Fisher Scientific, BP-358-212), 1 mM EDTA (Bioland Scientific, EDTA01); supplemented with 0.5% Nonidet P-40 (Sigma-Aldrich, I8896-50ML), and complete min EDTA-free protease inhibitor cocktail (Sigma-Aldrich, 11836170001), at a 1:4 pellet weight:volume ratio. The protease inhibitor cocktail tablets were added fresh before each use. The lysed cell pellet was then freeze thawed two times by incubating on dry ice for ~10 mins, then transferred to 37˚C bath until only a small amount of ice remains. The lysed cell pellet was spun again at 14,000g at 4˚C for 10 mins to remove cell insoluble products.

The supernatant was then transferred to a new clean tube and boiled at 95˚C for 5 mins. The supernatant was spun again at 14,000g at 4˚C for 10 mins. The supernatant was transferred to a new clean tube and analyzed by Qubit 4 Fluorometer (Fisher Scientific, Q33238, Q33211) for the total protein concentration.

The supernatant was then pre-cleared with mouse IgG Agarose beads (Sigma-Aldrich, A0919-2ML). The volume of beads used was determined via Qubit at 10 µl of beads to each 1 mg of total protein. Benzonase (Sigma-Aldrich, E1014-25KU) with 1 mM Magnesium Chloride ($MgCl_2$) (Sigma-Aldrich, M8266-100G) were added to the supernatant mixture with the mouse IgG beads for 30 mins shaking at 37˚C and then rotating at 4˚C for 30 mins. The supernatant was spun at 14,000g at 4˚C for 5 mins to pellet the mouse IgG agarose beads. The supernatant was then transferred to a new tube and

mixed with Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich, M8823-1ML). The anti-FLAG M2 magnetic beads and lysate were incubated at 4˚C on a rotating wheel overnight.

The beads were then pelleted using a magnetic rack, and the supernatant was removed and disposed. The magnetic flag M2 beads were washed three times in lysis buffer, then another three times in fresh 50 mM Ammonium Bicarbonate (Sigma-Aldrich, 09830-500G). The beads were resuspended in 5X Sodium Dodecyl Sulfate (SDS) loading buffer: 0.25% Bromophenol Blue (Fisher Scientific, B-392-5G), 0.5M Dithiothreitol (DTT) (IBI Scientific, IB21040), 50% Glycerol (Fisher Scientific, BP229-1), 10% SDS (Sigma-Aldrich, L3771-500G), 0.25M pH 6.8 Tris-HCl (Sigma-Aldrich, 10812846001). The flag magnetic beads were then boiled at 95˚C for 2 mins. The beads were pelleted again using the magnetic rack and the supernatant "sample" was transferred to a new clean tube.

5.12 Western Blot

The IP samples were run on precast any kD Mini-PROTEAN acrylamide gels (Bio-Rad, 4569035) for 45 mins at 160 Volts and 2 amps in 1X SDS running buffer. The protein samples were then transferred from the gel to a nitrocellulose membrane (Bio-Rad, 1704270) in a Trans-Blot Turbo Transfer System (Bio-Rad, 1704150) 25V 2.5 Amps for 3 mins. The stacking sponge was presoaked in Transfer buffer (Bio-Rad, 1704270).

The nitrocellulose membranes were blocked with 10 ml of blocking buffer (Phosphate Buffered Saline, 10% Tween 20 (PBST), 5% Milk) three times for 20 mins each. After blocking the nitrocellulose membrane, the blocking buffer is discarded, and the primary antibody solution (6 ml PBST, 6 µl primary antibody (1:1000 dilution), 1X Bovine Serum Albumin (BSA), 0.1% Sodium Azide) is added. The membrane was placed on a rocking platform overnight at 4˚C. The primary antibodies used were: Anti 4R-tau (Cosmo Bio, CAC-TIP-4RT-P01), Anti-FLAG M2-Peroxidase (Sigma-Aldrich, A8592-1MG), Anti-EIF4B (Cell Signaling Technology, 3592S). The next day, the primary antibodies are then discarded, and the membrane is washed for 20 mins two times with PBST. The secondary antibody solution (10 ml PBST, 5% Milk, 1:10000 dilution secondary antibody) is added to the nitrocellulose membrane at room temperature rocking on a platform for 1 hr. The secondary antibody used for Anti 4R-Tau and Anti-EIF4B was Anti-Rabbit IgG H&L (Horse Radish Peroxidase) (Abcam, ab6802).

After the secondary antibody incubation, the solution is discarded. The membrane is then washed three times for 20 mins with PBST. After the last wash discard the solution and add 1:1 ratio, 500 µl each, of the SuperSignal West Pico PLUS Chemiluminescent Substrate (Fisher Scientific, 34577). Let the membrane sit for 2 mins then Image using ChemiDoc MP Imaging system (Bio-Rad, 12003154).

5.13 RNase Protection

RNase protection was performed as described (93) using the RPAIII kit (Ambion/Invitrogen) and one million cpm of a 32P uniformly labeled probe. Hybridization was overnight, the digestion used RNaseA and T1 for one hour (43).

5.14 Mass Spectrometry

The immunoprecipitation was prepared as stated above (Figure 5.7) and the samples on the Flag M2 magnetic beads were sent to Dr. Alex Campos from the Medical Discovery Institute in La Jolla, California for analysis.

5.15 Results

5.16 Expression of tau circular RNA 12→7

To detect the endogenous circular RNA 12→7 from human brain tissue (Table5.2) and transfected cells, RNase protection probes were designed to hybridize to the backsplice junction with 60 nucleotides (nt) hybridizing to exon 12 and 40 nt hybridizing to exon 7 (Figure5.4 A, C. Figure 5.5 A-C). The protected band of 100 nt indicates the circular RNA, and two bands at 60 and 40 nt indicate the linear RNA. We show that the endogenous expression of the tau circular RNA 12→7 is expressed in different regions of human brain tissue (Figure 5.6).

In order to find sequences necessary and sufficient for circular RNA expression, I used the cloning procedure as described previously in chapter 2 (5) to overexpress the tau circular RNA 12→7. In this model, we first used a heterologous system, ZKSCAN1. ZKSCAN1 encodes a zinc-protein and expresses one of the most abundant human circular RNAs. We used the intronic regions that contained the Alu elements, which promotes circular RNA formation due to the short intronic repeats that have base complementarity, bringing the downstream exon 3 into close proximity to the upstream exon 2 (45). The tau exons 7,9-12 were cloned in a pcDNA3.1 expression vector between the two ZKSCAN1 short intronic repeat segments (Figure 5.2 D). The RNase protection probes were designed to detect the endogenous circular RNAs (Figure 5.4. A), and the circular RNA that contains the artificial 3X flag tag in the ZKSCAN1 model (Figure 5.4 C). Stably transfected human embryonic kidney (HEK) 293T cells, that express the ZKSCAN1 tau 12→7 circular RNA, were validated for circular RNA expression (Figure 5.6).

The circular RNA 12→7 is expressed in both human brain tissue and transfected cells. However, there are no antibodies available to distinguish between the protein generated from the linear RNA and the circular RNA. Therefore, we designed the expression constructs incorporating a 3X flag tag upstream of the start codon of the circular RNA. Translation of the 3X flag tag will only occur from the circular RNA. We purify the protein by immunoprecipitating with an anti-flag antibody and detect the protein with an anti-tau antibody by western blot.

To test for translation the immunoprecipitation protocol was optimized (Figure 5.7), where we lysed the cells and performed two freeze thaw cycles to break open the cells. The cell debris was recovered by centrifugation, and the lysate supernatant was boiled at 95˚C for five minutes, to break apart tubulin and other enzymes and protein complexes from the tau protein, as tau is heat stable (136). In pilot experiments, we found that the protein encoded by the circular RNA is also heat stable. The lysate is spun again and mixed with Mouse IgG agarose beads to pre-clear any non-specific binding proteins, along with Benzonase that degrades DNA and RNA that is present in the lysate. The pre-clearing step is incubated at 4˚C for one hour on a rotating wheel. This step allows some denatured proteins to renature and bind again to interacting proteins. The lysate is now immunoprecipitated with anti-flag M2 magnetic beads to purify the protein generated from the circular RNA. The flag tag is located in exon 7 with the start codon in exon 9. Therefore, the only way to generate the flag sequence is by translation of a circular RNA. The immunoprecipitation is incubated at 4˚C on a rotating wheel overnight and then washed and boiled in SDS loading buffer. The boiled eluent is separated on an acrylamide gel, then transferred to a nitrocellulose membrane, and analyzed by western blot with anti-flag and anti-tau antibodies.

The circular tau 12→7 RNA is translated due to the naturally occurring methionine (start codon) in exon 9. The circular tau 12→7 RNA gives rise to multiple proteins at different sizes, potentially indicating the rolling circle translation and multimeric protein (Figure 5.8, A-B). As there is no stop codon, an unusual sequence element or RNA structure may implement termination of the translated protein. Translation could potentially be stalled due to RNA recoding or editing. RNA editing, by ADAR enzymes, edit adenosines to inosines and have been shown to promote translational stalling by inducing context-dependent recoding (137). The immunoprecipitation purifies the protein with the flag tag and then the protein is detected using two antibodies, anti-flag and anti-tau.

5.17 Frontotemporal Dementia mutations create new start codons

Human brain tissue was analyzed by RNase protection for the circular 12→10 RNA expression described in chapter 3 (Figure 3.2 E). The cellular expression constructs expressing the ZKSCAN1 and tau authentic Alu element 12→10 circular RNAs were validated by RNase protection using probes that hybridize to the backsplice junction (Figure 5.4 B, D. Figure 5.5 A-C). The probes contain 60 nucleotides of exon 12 and 40 nucleotides of exon 10, and the ZKSCAN1 constructs contain a partial 3X flag tag sequence in exon 10. The ZKSCAN1 probes were designed for the stable and transient transfected cells expressing the circular RNA 12→10 with the 3X flag tag, the endogenous probes do not have the 3X flag tag sequence. The circular 12→10 RNA was detected by RNase protection for both ZKSCAN1 and tau authentic Alu element constructs (Figure 5.9 A-B).

Two mutations in tau exons 11 (K317M) and 12 (V337M) lead to Frontotemporal Dementia (138), creating start codons in the circular RNA 12→10. These mutations are

predicted to initiate translation forming a protein that contains the microtubule binding domains R2-R4. Immunoprecipitation was performed for the 12→10 circular RNA, as indicated previously for the 12→7 translated circular RNA (Figure 5.7). The circular tau RNA 12→10 with ZKSCAN1 and authentic tau Alu element have a 3X flag tag incorporated in exon 10, similar to the 12→7 circular RNA. Translation of the 3X flag tag will only occur from the circular RNA 12→10, but not the linear since the start codons are located downstream in exon 11 and 12. The ZKSCAN1 10-12 constructs were tested for translation together with the constructs containing the tau authentic Alu element (Figure 5.10 A-C). The wild type circular RNA is not translated and is only translated with the FTDP-17 mutations that incorporate a start codon (Figure 5.10 B-C). Protein translation is reduced when the ZKSCAN1 Alu element is replaced with the authentic tau Alu element (Figure 5.10 C).

To analyze the protein generated from the circular RNA 12→10 with the V337M mutation, I immunoprecipitated the protein described before using the anti-flag M2 magnetic beads and added Benzonase to degrade DNA and RNA in the lysate to avoid nucleic acid mediated interactions. The beads were incubated overnight at 4˚C on a rotating wheel and washed the next day in lysis buffer and fresh 50 mM ammonium bicarbonate. The magnetic beads with the purified protein bound were then frozen and shipped to Dr. Alex Campos at the Medical Discovery Institute in La Jolla, California for Mass Spectrometry analysis. Dr. Campos digested the protein on the beads with trypsin and then performed mass spectrometry. In the analysis we detected the sequence containing the V337M mutation from the circular RNA with amino acids upstream of the methionine, indicating the protein was translated from the circular RNA (Figure 5.11).

5.18 ADAR enzymes promote circular tau RNA translation

Adenosine deaminase acting on RNA (ADAR) is a protein coding gene that encodes enzymes responsible for RNA editing. The enzyme recodes the RNA sequence deaminating adenosines into inosines and is referred to as A-to-I editing. This modification changes the RNA sequence, where adenosines are inosines and are now read as guanosines by the ribosome. For instance, the nucleotide sequence AUA can be converted to AUI, which would be read as an AUG, forming a new start codon. It has been shown that a deregulation of RNA editing can contribute to Alzheimer's disease (135). Two editing enzymes, ADAR also known as ADAR1 and ADARB1 also known as ADAR2, are catalytically active enzymes that catalyze the hydrolytic deamination of adenosines to inosines of double-stranded RNA structures. ADARB2 also known as ADAR3 is catalytically inactive and lacks editing activity. ADAR3 prevents the binding of other ADAR enzymes to their targets in vitro and decreases the efficiency of these enzymes.

The effect of ADARs on protein translation was investigated with the catalytically active ADARs, ADAR1 and ADAR2, and the inactive ADAR3. Transient transfected HEK 293T cells were co-transfected with the ADAR enzymes and the ZKSCAN1 tau 10-12 V337M and 7-12 WT constructs, and protein analysis was performed as previously described above. Protein expression was increased when overexpressing the catalytically

active ADAR1 and ADAR2 enzymes (Figure 5.12 A-B) for both the 10-12 VM and 7-12 WT expression constructs. Protein expression was not affected with the inactive ADAR3 (Figure 5.12 A-C). Translation of the circular RNA 12→10 wild type was not observed. However, when there is an overexpression of the editing enzymes, ADAR1 and ADAR2, the circular RNA 12→10 wild type is translated (Figure 5.12 C). A deregulation of ADAR enzymes may be acting on the circular RNAs causing more editing to occur and allowing more protein to be translated from the circular tau RNAs, which could contribute to tauopathies and Alzheimer's disease.

5.19 Interaction with EIF4B

Eukaryotic translation initiation factor 4B (EIF4B) is a helicase ribosomal protein required for mRNA to bind with ribosomes and to be translated. In our mass-spectrometry analysis EIF4B was detected as one of the abundant peptides from the immunoprecipitation with the circular RNA 12→10 V337M. We validated this interaction by immunoprecipitating with anti-flag beads and detected endogenous EIF4B with anti-serum by western blot. The experiment was done in the presence of Benzonase, an endonuclease that attacks and degrades all forms of DNA and RNA (Figure 5.13 A). Binding of endogenous EIF4B was detected for both the tau circular RNAs 12→7 and 12→10, after being boiled in the lysate and pulled down with an anti-flag immunoprecipitation (Figure 5.13 A-C).

5.20 siRNAs can be used as potential therapy for Tauopathies

I showed that the circular tau RNA 12→7 can be translated, and in instances where mutations and editing introduce start codons, the circular tau RNA 12→10 is translated. These translated proteins, when over expressed, may lead to a neurotoxic protein that could mimic the 'K18' peptide promoting neurodegeneration (6). Therefore, we designed siRNAs targeting the backsplice junction of the circular tau RNAs (Table 5.1). The siRNA would be specific for the circular tau RNA and not the linear due to this design and could therefore be used as a therapeutic drug, preventing tau pathologies.

The siRNAs were transfected into stable HEK 293T cells at different concentrations for both the ZKSCAN1 tau 7-12 wild type and 10-12 V337M. The protein was purified by immunoprecipitating with a 3X flag tag as previously described and detected by western blot with an anti-flag antibody. Protein levels were reduced with increasing concentrations indicating translation inhibition (Figure 5.14 A-C).

5.21 Discussion

5.22 MAPT as a contributor to tauopathies

The microtubule-associated protein tau, under pathophysiological conditions, forms neurofibrillary tangles that are the hallmark of sporadic Alzheimer's disease and familial frontotemporal dementia linked to chromosome 17 (FTDP-17) (3).

5.23 Function of tau circular RNAs

In this chapter we have discovered a function for the circular tau RNAs showing that the endogenous tau 12→7 circular RNA with the original start codon in exon 9 is translated. The translated protein is much larger than most circular RNAs (681 nt = 227 amino acids) and may have a physiological role. For smaller circular RNAs, the translated peptides may be acting as a neuropeptide due to the smaller size (150 nt = 50 amino acids), which is a characteristic of most circular RNAs. The circular tau RNA 12→10 is translated when mutations or possible editing introduce start codons. The mutations that cause Frontotemporal Dementia (K317M and V337M) may be acting on the circular RNA causing the translated protein to induce the pathology. The proteins generated from the 12→10 circular RNA could be causing tau aggregation and neurofilament tangles, since it contains the region of the microtubule binding domains similar to the K18 seed segment.

Adenosine deaminase acting on RNA (ADAR) is a protein coding gene that encodes an enzyme responsible for RNA editing by site-specific deamination of adenosines. This enzyme destabilizes double-stranded RNA through conversion of adenosines to inosines. The catalytically active enzymes, ADAR1 and ADAR2, catalyze the hydrolytic deamination of adenosine to inosine in double-stranded RNA referred to as A-to-I RNA editing (139-141). ADAR3 is catalytically inactive lacking the editing activity. ADAR3 prevents the binding of other ADAR enzymes to their targets and decreases the efficiency of these enzymes. ADAR3 is capable of binding to both double stranded and single stranded RNA. Deregulation of ADAR enzymes have been linked to a variety of neurodegenerative and neurological disorders (142, 143). The ADAR enzymes that are deregulated in Alzheimer's disease may be acting on the circular RNAs, which contribute to the pathology. We have shown that an overexpression of the active enzymes, ADAR1 and ADAR2, increase circular tau RNA translation and protein production. We also discovered that the 12→10 wild type circular RNA, that has no start codon, is translated when ADAR1 and ADAR2 are present. The tau circular RNAs' editing may introduce new start codons, and a deregulation of RNA editing may cause an overabundance of protein production, causing the tau pathology.

We have found that the circular tau RNAs protein interacts with the eukaryotic translation initiation factor 4B (EILF4B). EIF4B is a protein encoding gene and is required for the initiation of mRNA to ribosomes. EIF4B functions in close association with other initiation factors, EIF4-F and EIF4-A. EIF4B binds near the 5′-terminal cap of mRNA in the presence of EIF4-F and ATP. EIF4B promotes the ATPase activity and the ATP-dependent RNA unwinding activity of both EIF4-A and EIF4-F. The interaction of the circular tau RNAs protein with EIF4B may explain a potential function for the circular RNAs. The circular tau RNAs may be regulating translation of other genes by reducing the amount of EIF4B needed to bind competing mRNAs to ribosomes.

5.24 Tau circular RNAs disease relevance

The *MAPT* gene generates so far unknown circular RNAs that are unexpectedly translated producing proteins that are predicted to be similar to a neurotoxic peptide known as K18. The K18 peptide is a segment of the tau microtubule binding domains and promotes tau aggregation and neurofibrillary tangles (6). Mutations in the tau locus that cause FTDP-17 reside mainly in exon 9-13, regions that control circular RNA formation and may be acting on the circular RNAs instead of the linear RNAs. This novel pathway has never been looked into before with the circular tau RNAs producing proteins that can contribute to neurodegeneration. The circular tau RNAs can be a new target for drug therapies in combating tau pathology. We have shown that protein production can be reduced with siRNAs that target the backsplice junction of both the 12→7 and 12→10 circular RNA. Currently there is no cure for tauopathies, and treatment is limited in slowing down the pathology. The siRNAs could be used as a potential therapy by preventing the circular tau RNAs from being translated, preventing the tau pathology from occurring or accumulating.

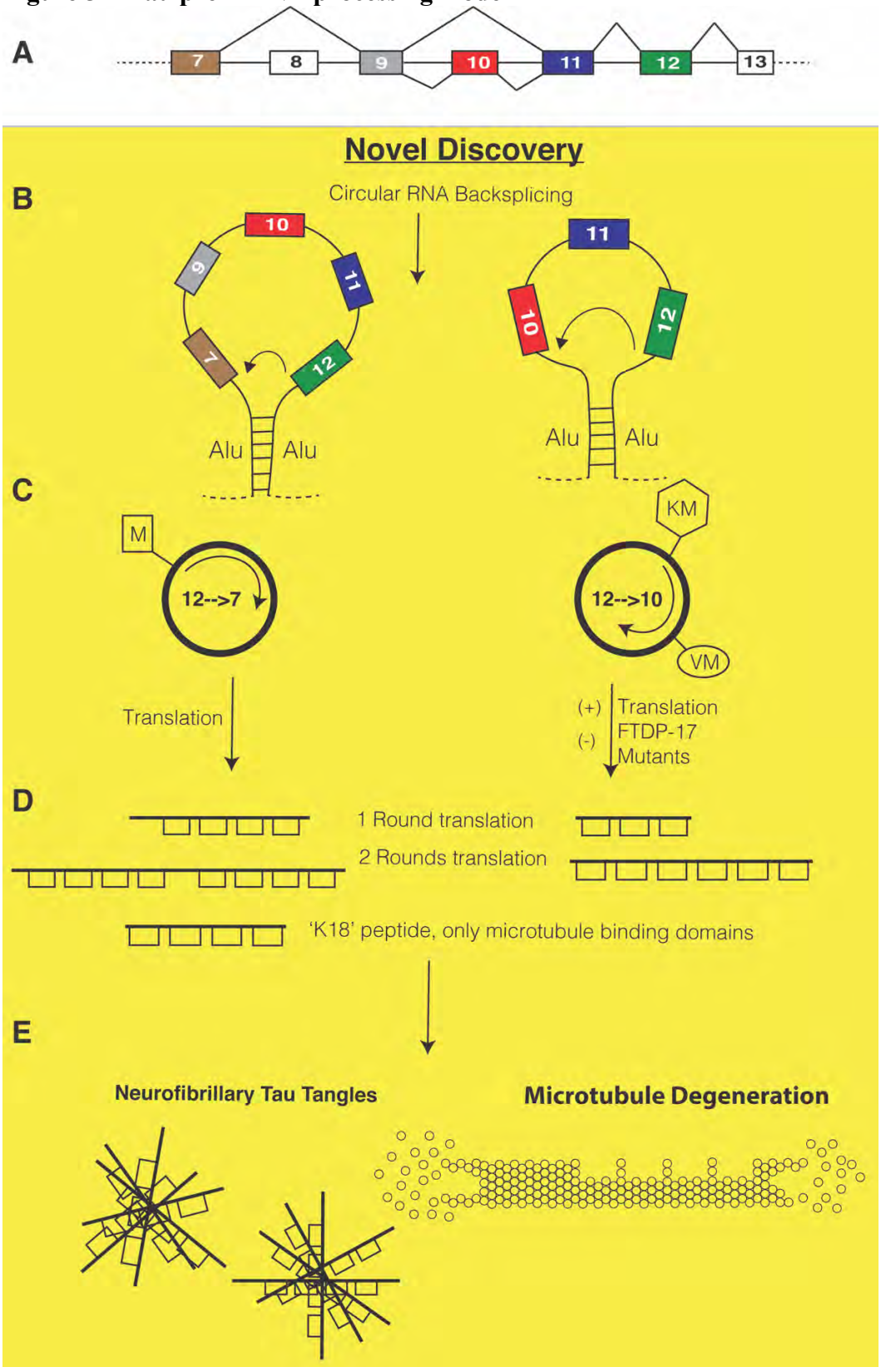
Table 5-1 siRNA sequences targeting circular tau RNA backsplice junctions. Provisional Patent Filed UK: #2539

12to10 junction	Silencer® Select by Life Technologies Sense Strand: 5′AUAAAAAGGUGCAGAUAAUtt 3′ Antisense Strand: 3′AUUAUCUGCACCUUUUUAUtt 5′
12→10 RNA Sequence	**Exon 12/Exon 10:** AUAAAAAG/GUGCAGAUAAU
12to7 junction V3	Silencer® Select by Life Technologies Sense Strand: 5′AGGAAAUAAAAAGGGGGCUtt 3′ Antisense Strand: 3′AGCCCCCUUUUUAUUUCCUcc 5′
12→7 RNA Sequence	**Exon 12/Exon 7:** AGGAAAUAAAAAG/GGGGCUtt

Table 5-2 Human Brain Samples

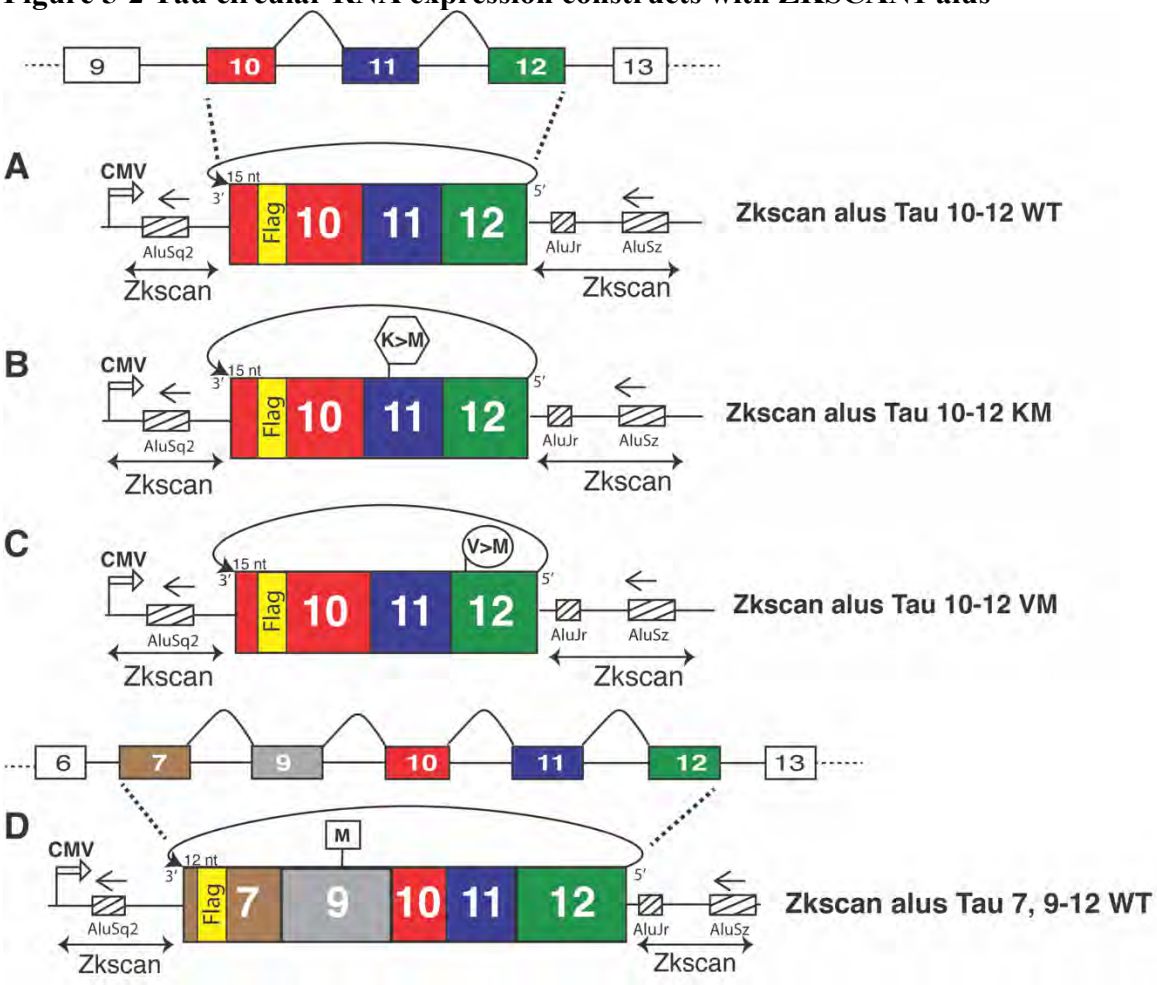
mRef#	Age	Gender	Race	Braak Stage	PMI (Hours)	Brain Weight (Grams)	Brain Bank
5464	31	Female	Asian	N/A	16	1210	NIH NeuroBioBank
5357	51	Female	White	N/A	8	1320	NIH NeuroBioBank
5298	31	Male	White	N/A	10	N/A	NIH NeuroBioBank
1161	84	Female	White	0	2.50	1230	UK Sanders Brown-Center on aging

Figure 5-1 Tau pre-mRNA processing model



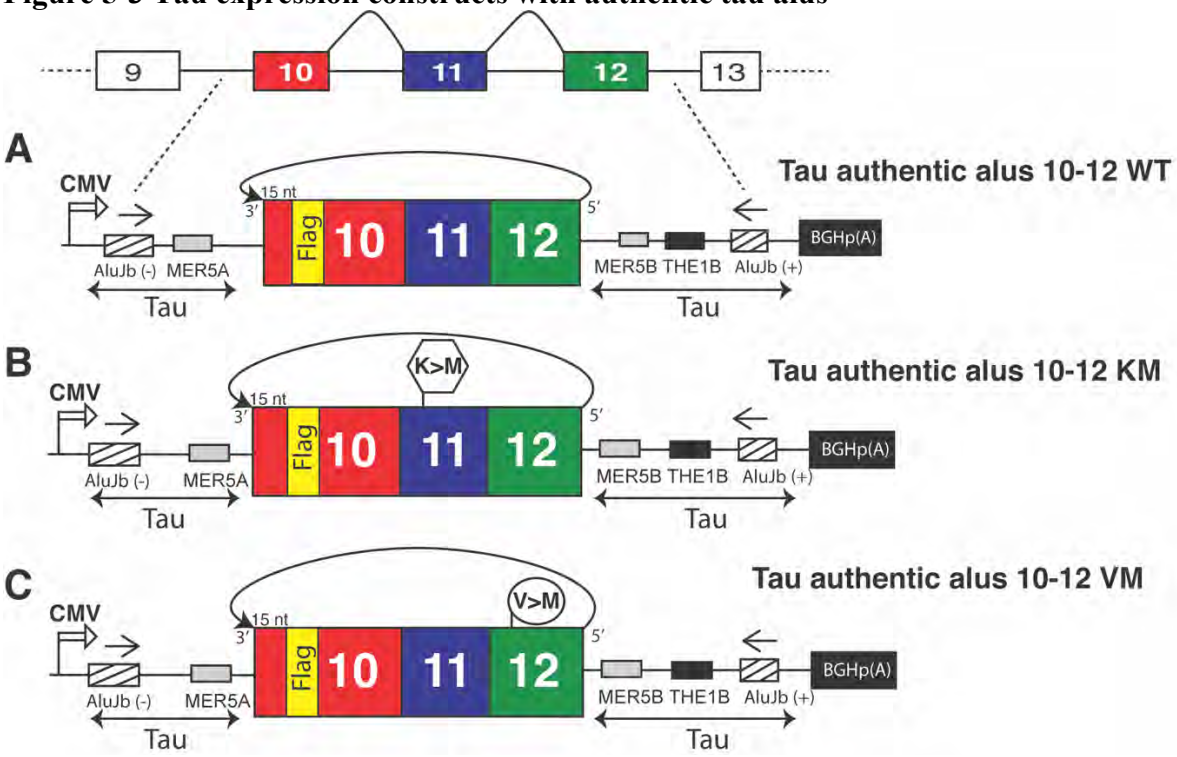
A. Shortened schematic of the tau locus and alternative splicing of the exons.
B. Backsplicing of exon 12 onto exon 7 or exon 10 forming the two tau circular RNAs. The Alu elements have base complementarity forming the secondary structure needed for backsplicing to occur.
C. Tau circular RNAs 12→7 is translated forming proteins of different sizes due to possible multiple rounds of translation. The 12→10 circular RNA is translated when FTDP-17 mutations are introduced that create start codons, also giving rise to multiple proteins of different sizes.
D. Both circular RNAs proteins are predicted to be similar to the neurotoxic 'K18' peptide that contains only the microtubule binding domains.
E. The proteins translated from the circular tau RNAs may act like the 'K18' peptide and promote tau aggregation and neurofibrillary tangles creating the tau pathology

Figure 5-2 Tau circular RNA expression constructs with ZKSCAN1 alus



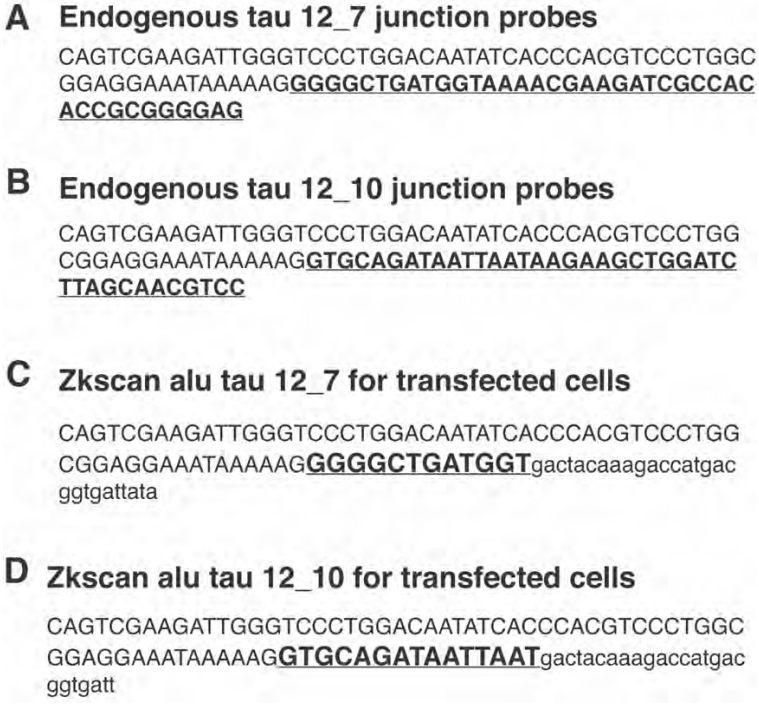
ZKSCAN Alu constructs with cDNA from tau exons 10-12 WT (**A**), 10-12 K317M (**B**), 10-12 V337M (**C**), and 7,9-12 WT (**D**). The yellow box is 3X flag tag and striped boxes are ZKSCAN Alu sequences.

Figure 5-3 Tau expression constructs with authentic tau alus



Both authentic Tau Alu constructs with cDNA from tau exons 10-12 WT (**A**), 10-12 K317M (**B**), 10-12 V337M (**C**). The yellow box is 3X flag tag and striped boxes are Alu sequence. Grey and black boxes are repeat sequences.

Figure 5-4 RNase protection probe sequences



RNase protection probe sequences used to detect endogenous and zkscan tau linear and circular RNA 12→7 and 12→10.

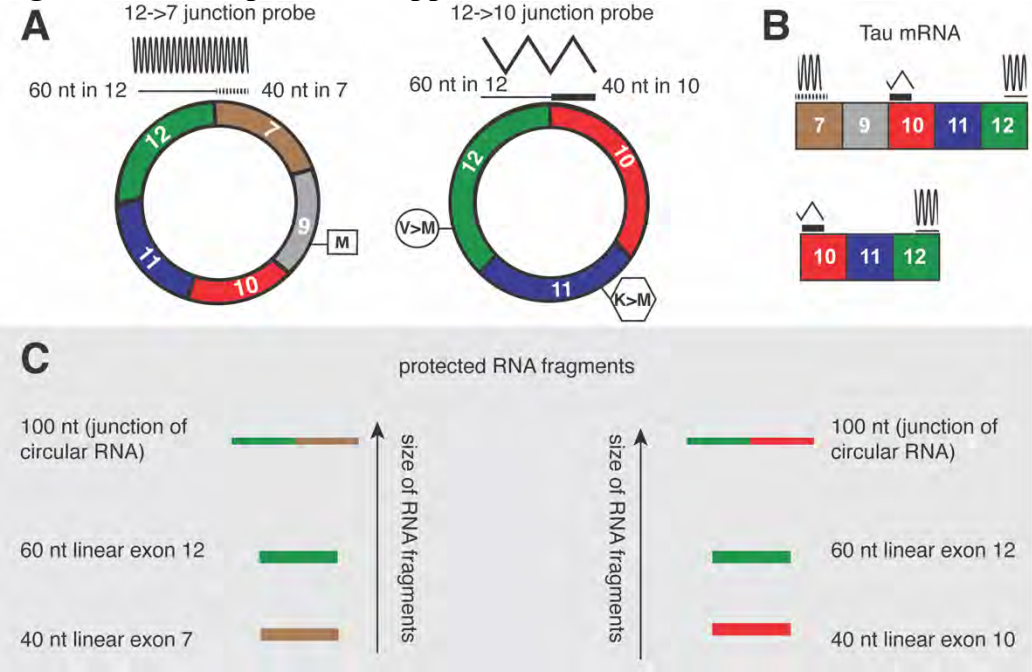
A. Tau exon 12 all caps regular text on the left and exon 7 all caps, bold, and underlined on the right

B. Tau exon 12 all caps regular text on the left and exon 10 all caps, bold, and underlined on the right

C. Tau exon 12 all caps regular text on the left and exon 7 all caps, bold, and underlined in the middle and lower case regular text is part of flag sequence.

D. Tau exon 12 all caps regular text on the left and exon 10 all caps, bold, and underlined in the middle and lower case regular text is part of flag sequence.

Figure 5-5 RNase protection approach to detect circular RNAs

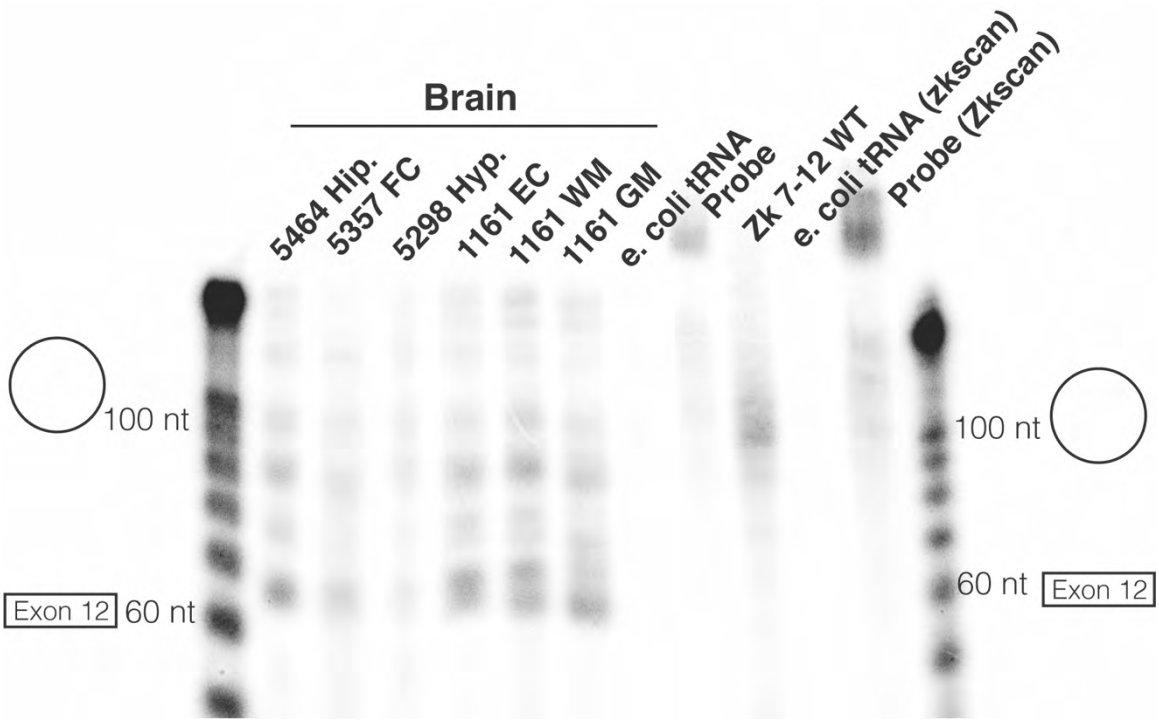


A. Schematic overview: circular RNAs are detected with a radioactive antisense RNA. This hybridizes to the circular RNA junction (12→7 or 12→10). 60 nt in exon 12 are protected and 40 nt in exon 7 or 10 respectively.

B. Hybridization to linear RNA. A tau mRNA isoform from exon 7 to 12 is shown. The hybridization of the probe to this mRNA is indicated.

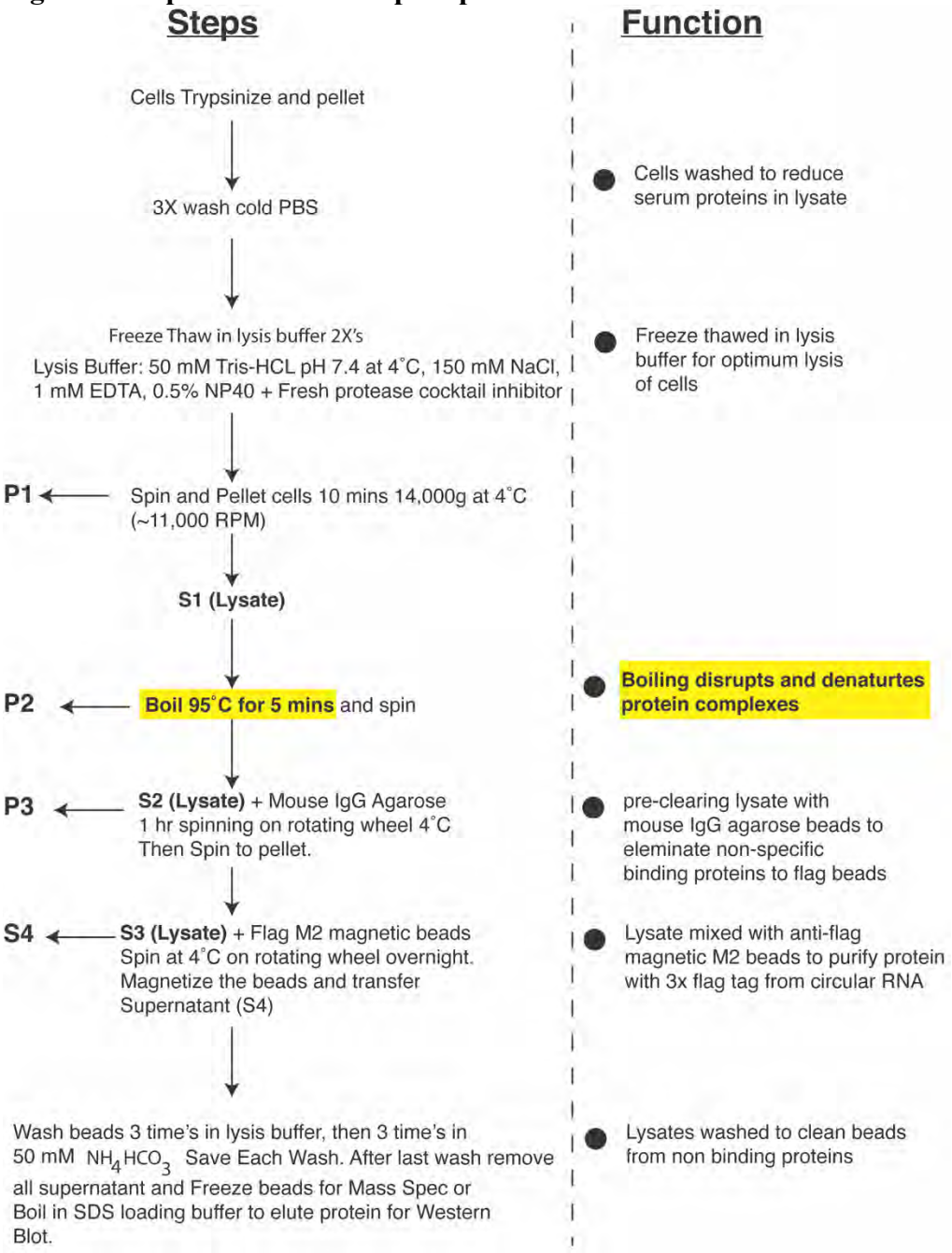
C. Anticipated results: The full-length, 100 nt long protected fragment corresponds to the circular RNA and is very weakly expressed. The probe also hybridizes to Linear forms of tau, i.e. the linear form that contains exon 12, 7 and 10, resulting in much stronger bands. Exon 10 is alternatively spliced, and thus the intensity of this band varies.

Figure 5-6 RNase protection Brain regions and stable cells expressing tau circular 12→7



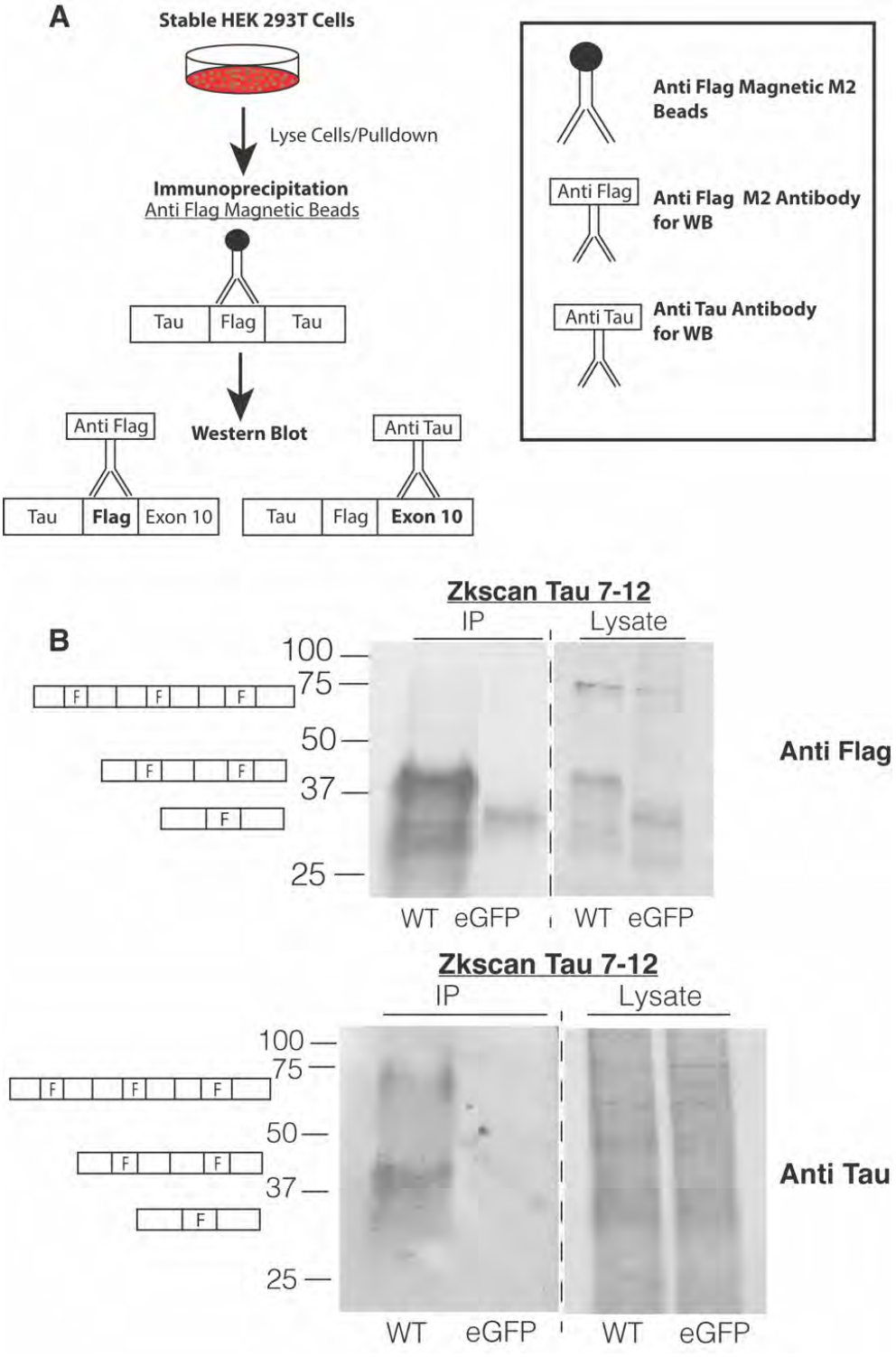
20 µg total brain RNA from 3 individuals from various regions were analyzed using RNAse protection with probe A (endogenous tau 12_7_junction probe) (Figure 5.3A). Bands at 100 nt are protected circular 12→7 junction. Numbers indicate length in nucleotides. Patient 5464 Hippocampus. Patient 5357 Frontal cortex. Patient 5298 Hypothalamus. Patient 1161 Entorhinal Cortex. Patient 1161 White Matter. Patient 1161 Grey Matter. e. coli tRNA (junction). Probe junction. Stable HEK 293T ZKSCAN1 tau 7-12 WT. e. coli tRNA (ZKSCAN1). Probe (ZKSCAN1). *This figure was produced by Dr. Giorgi Margvelani.

Figure 5-7 Optimized Immunoprecipitation Protocol



Immunoprecipitation protocol to purify the protein generated from circular tau RNA using a 3X flag tag. The IP used anti-flag M2 magnetic beads to purify protein generated from the circular RNA and not the linear RNA.

Figure 5-8 Protein expression of stable cells expressing ZKSCAN1 tau 7-12 WT circular RNA

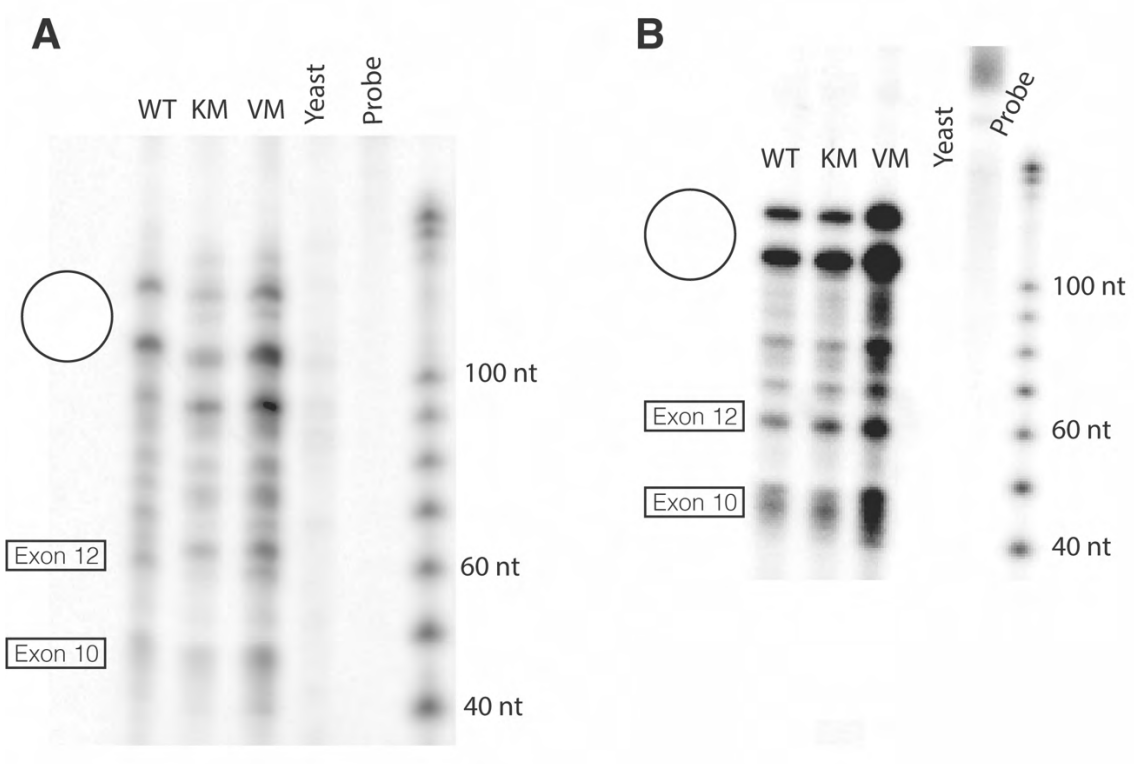


Stable cell lines expressing circular 12→7 RNA generated from ZKSCAN1 tau 7-12 WT construct were lysed and Immunoprecipitated using a 3X flag pulldown. All tau cellular expression constructs contain 3X flag tag. eGFP contained 3X flag tag.

A. Immunoprecipitation schematic pulling down protein generated from circular tau RNA with 3X flag tag. The Western Blot was detected with anti-flag antibody and an anti-tau antibody that has an epitope in exon 10.

B. Protein expression generated from ZKSCAN1 tau 7-12 WT circular RNA. Detection by anti-flag and tau antibodies. The peptide diagrams on the left are depicted as 1, 2 and 3 rounds of translation. The F is for 3X flag tag.

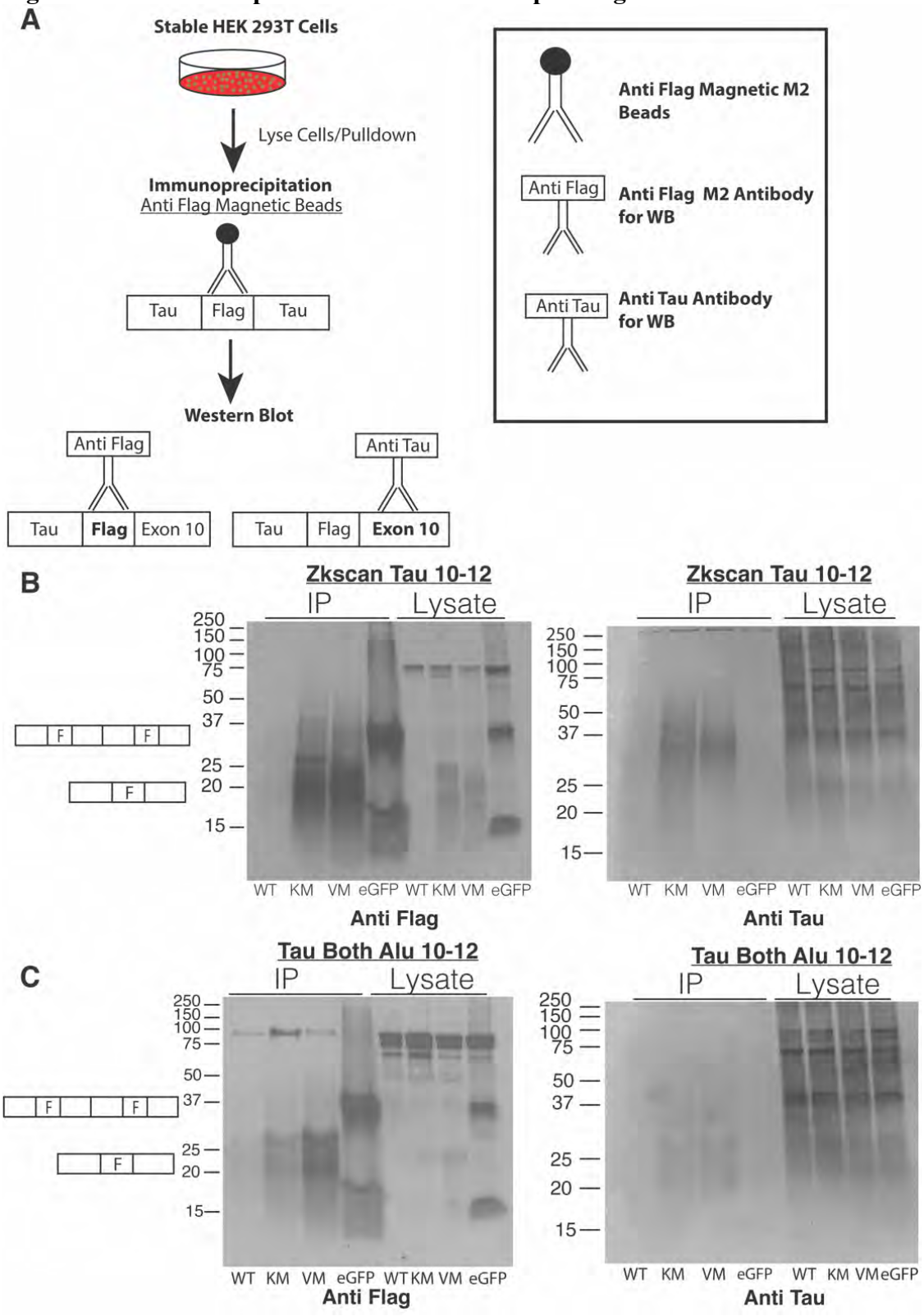
Figure 5-9 RNase protection HEK 293T cells expressing circular tau 12→10



RNase protection from stable HEK 293T cell lines expressing tau 12→10 circular RNAs. *This figure was produced by Dr. Giorgi Margvelani.

- **A.** Stable HEK 293T cells expressing Tau authentic alus 10-12 WT, KM, VM. Band at 100 nt indicates protected splice junction. The circle depicts the protected junction from the circular RNA and the linear fragments exons 12 and 10.
- **B.** Stable HEK 293T cells expressing ZKSCAN1 tau 10-12 WT, KM, VM. Band at 100 nt indicates protected splice junction. The circle depicts the protected junction from the circular RNA and the linear fragments exons 12 and 10.

Figure 5-10 Protein expression of stable cells expressing 10-12 circular RNAs



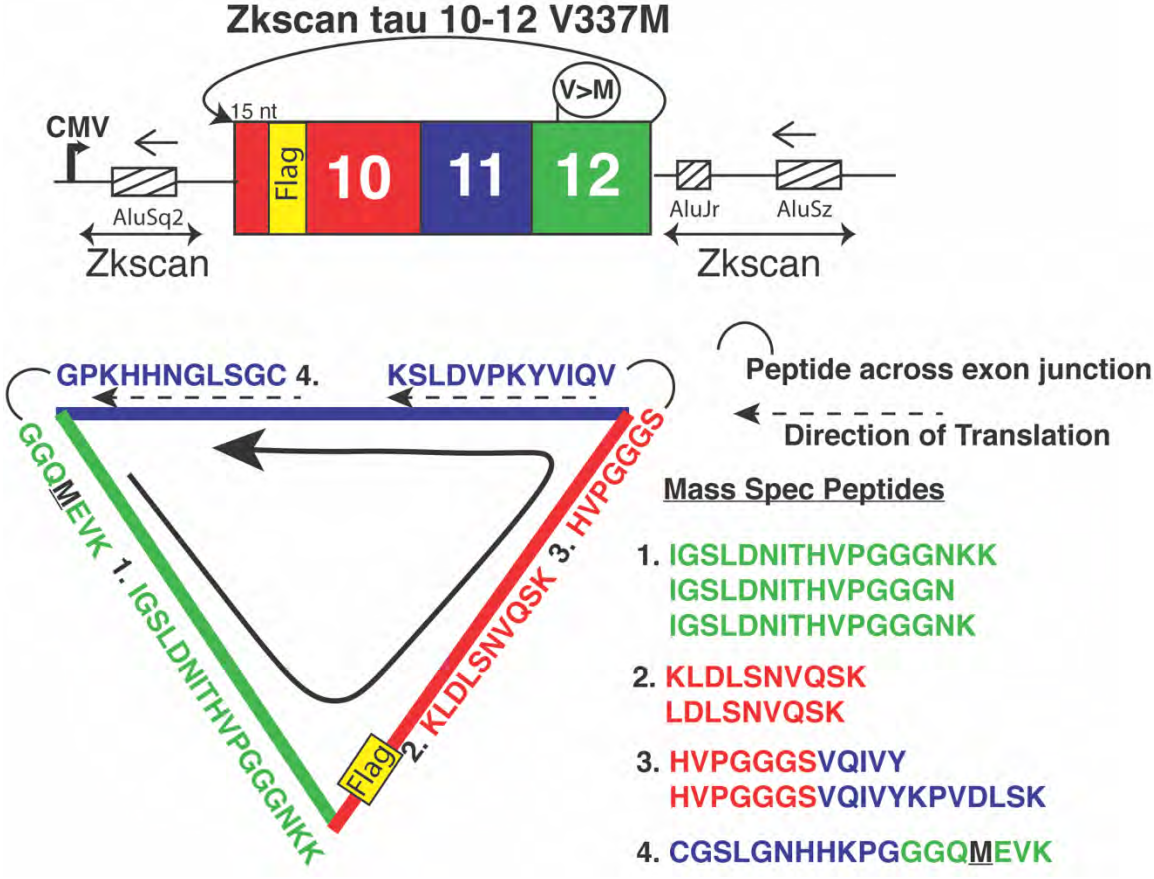
Stable cell lines expressing circular 12→10 RNA generated from ZKSCAN1 and the authentic tau alus 10-12 WT, K317M, and V337M constructs were lysed and Immunoprecipitated using a 3X flag pulldown. All tau cellular expression constructs contain 3X flag tag. eGFP contained 3X flag tag.

A. Immunoprecipitation schematic pulling down protein generated from circular tau RNA with 3X flag tag. The Western Blot was detected with anti-flag and anti-tau antibodies. The anti-tau antibody has an epitope in exon 10.

B. Protein expression with ZKSCAN1 alus in circular tau 12→10 constructs. Detected by anti-flag and tau antibodies. Only the mutants K317M and V337M are translated. The protein diagrams on the left are depicted as 1 round and 2 round translation. The F is for 3X flag tag.

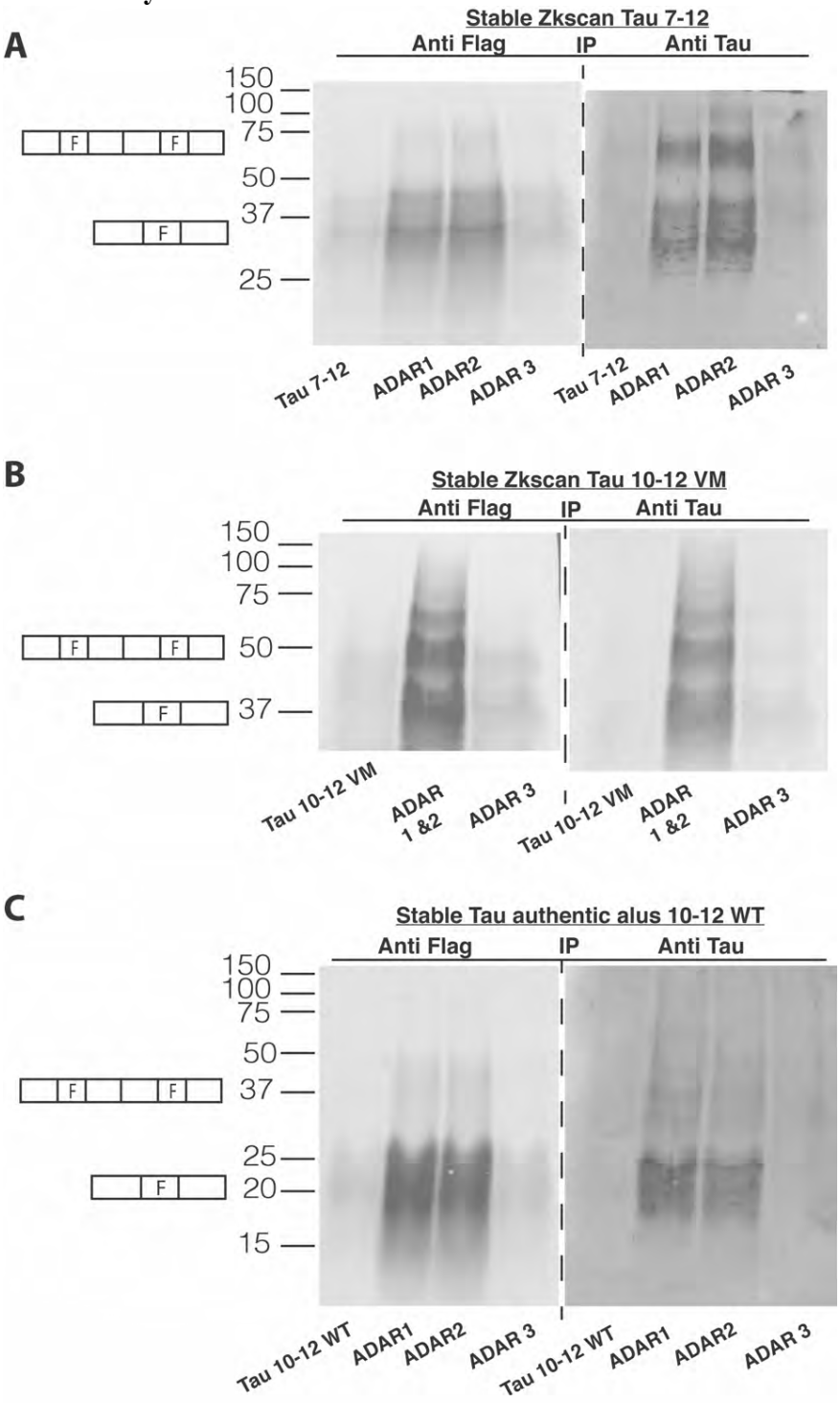
C. Protein expression with authentic tau alus in circular tau 12→10 constructs. Detected by anti-flag and tau antibodies. Only the mutants K317M and V337M are translated. The protein diagrams on the left are depicted as 1 round and 2 round translation. The F is for 3X flag tag.

Figure 5-11 Mass Spectrometry Analysis Zkscan 10-12 VM



Mass Spectrometry analysis from stable HEK 293T cells expressing ZKSCAN1 tau 10-12 VM circular RNA. Peptide number 4 depicts the V337M mutation and start codon, indicating circular translation. Protein was pulled down using flag M2 magnetic beads and digested with Trypsin.

Figure 5-12 Protein translation of circular tau RNAs 12→7 and 12→10 affected by ADAR enzymes



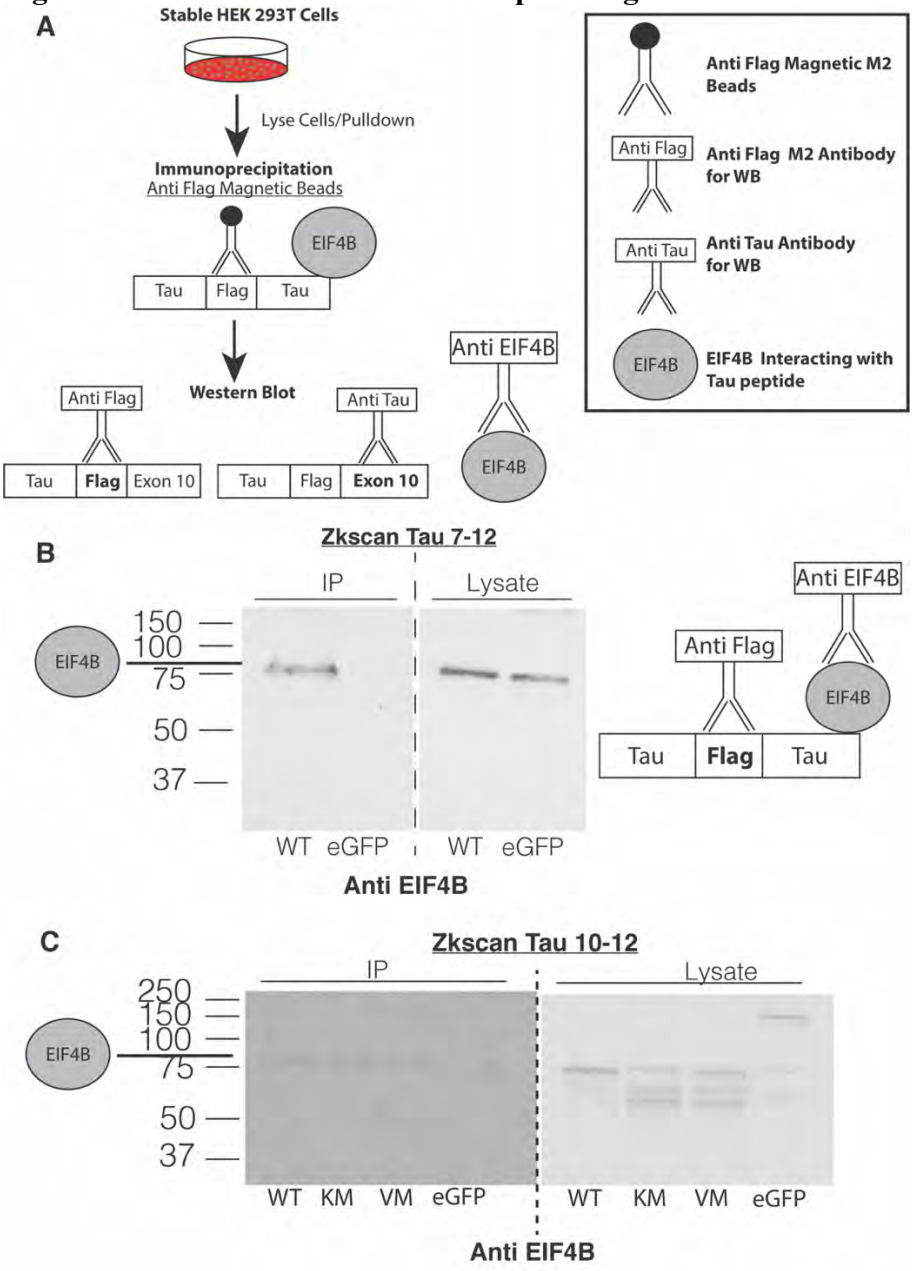
Stable cell lines expressing circular 12→7 WT, 12→10 V337M, and 12→10 WT transfected with ADARs 1-3. 96 hrs post transfection cells were immunoprecipitated and analyzed by western blot with anti-flag and tau antibodies.

A. Circular RNA 12→7 increased protein expression with ADAR1 and ADAR2. There was no effect on protein translation with ADAR3. Western blot was detected with both anti-flag and anti-tau antibodies. 4 µgs each of ADAR1, ADAR2 and ADAR3 were transfected. The protein diagrams on the left are depicted as 1 round and 2 round translation. The F is for 3X flag tag.

B. Circular RNA 12→10 V337M increased protein expression with ADAR1 and ADAR2. There was no effect on protein translation with ADAR3. Western blot was detected with both anti-flag and anti-tau antibodies. 2 µgs of ADAR1 and ADAR2 and 4 µgs of ADAR3 were transfected. The protein diagrams on the left are depicted as 1 round and 2 round translation. The F is for 3X flag tag.

C. Circular RNA 12→10 WT is translated with ADAR1 and ADAR2 active enzymes. Translation does not occur without the ADAR enzymes 1 and 2. WT and the inactive ADAR3 does not generate protein. Western blot was detected with both anti-flag and anti-tau antibodies. 4 µgs each of ADAR1, ADAR2 and ADAR3 were transfected in the cells. The peptide diagrams on the left are depicted as 1 round and 2 round translation. The F is for 3x Flag tag

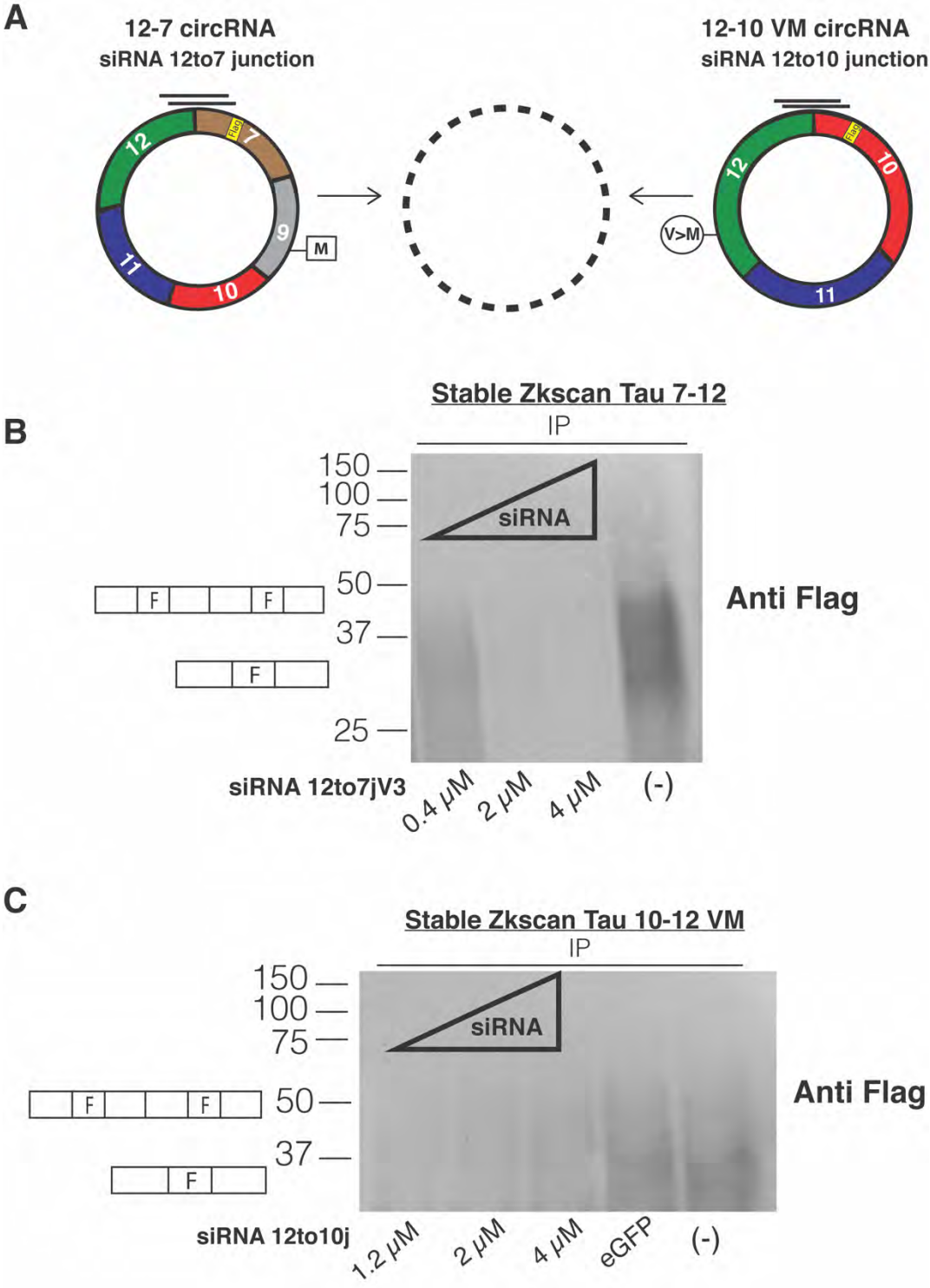
Figure 5-13 EIF4B interacts with tau protein generated from the circular RNA.



EIF4B detected from Mass Spectrometry after tau IP pulldown with anti-flag antibody was blotted with Anti-EIF4B and detected the endogenous enzyme.

- **A.** Immunoprecipitation schematic pulling down peptide generated from circular tau RNA with 3X flag tag. The Western Blot was detected with anti EIF4B antibody.
- **B.** Endogenous expression of EIF4B after 3X flag IP pulldown from stable cells expressing ZKSCAN1 tau circular RNA 12→7.
- **C.** Endogenous expression of EIF4B after 3X flag IP pulldown from stable cells expressing ZKSCAN1 tau circular RNA 12→10.

Figure 5-14 siRNAs destroy circular tau RNAs and prohibit protein translation



Stable cells expressing ZKSCAN1 tau 7-12 WT and ZKSCAN1 10-12 VM were transfected in a 6-well plate with siRNAs at different concentrations. Cells were lysed 72 hrs post transfection, Immunoprecipitated, and analyzed by western blot.

A. Schematic depicting siRNA binding to backsplice junction of 12→7 and 12→10 circular RNA. Yellow box in exon 7 and exon 10 is 3X flag.

B. Increasing concentrations of siRNA targeting 12-7 backsplice junction prohibits protein translation from the 12→7 circular RNA. The negative control was stable cells not transfected with siRNA. The peptide diagrams on the left are depicted as 1 and 2 rounds of translation. The F is for 3X flag tag.

C. Increasing concentrations of siRNA targeting 12-10 backsplice junction prohibits protein translation from the 12→10 circular RNA. 3X flag eGFP was included as negative control along with stable cells without siRNAs. The peptide diagrams on the left are depicted as 1 and 2 rounds of translation. The F is for 3X flag tag.

Supplemental Figure 5-1 ZKSCAN1 tau 7-12 WT Sequence

AGTGACAGTGGAGATTGTACAGTTTTTTCCTCGATTTGTCAGGAttttttttttttgacgga
gtttaacttcttgtctcccaggtaggaagtgcagtggcgtaatctcggctcactacaacctccacctcctgggttcaagcgtttctcc
tgcctcagctttccgagtagctgggattacaggcgcctgccaccatgccctgctgacttttgtatttttagtagagacggggtttcac
catgttggccaggctggtcttgaactcctgaccgcaggcgattggcctgcctcggcctcccaaagtgctgagattacaggcgtg
agccaccaccccggccTCAGGAGCGTTCTGATAGTGCCTCGATGTGCTGCCTCCTATA
AAGTGTTAGCAGCACAGATCACTTTTTGTAAAGGTACGTACTAATGACTTTTT
TTTTATACTTCAGGGGGCTGATGGTGACTACAAAGACCATGACGGTGATTAT
AAAGATCATGACATCGATTACAAGGATGACGATGACAAGAAAACGAAGATC
GCCACACCGCGGGGAGCAGCCCCTCCAGGCCAGAAGGGCCAGGCCAACG
CCACCAGGATTCCAGCAAAAACCCCGCCCGCTCCAAAGACACCACCCAGCT
CTGGTGAACCTCCAAAATCAGGGGATCGCAGCGGCTACAGCAGCCCCGGC
TCCCCAGGCACTCCCGGCAGCCGCTCCCGCACCCCGTCCCTTCCAACCCC
ACCCACCCGGGAGCCCAAGAAGGTGGCAGTGGTCCGTACTCCACCCAAGT
CGCCGTCTTCCGCCAAGAGCCGCCTGCAGACAGCCCCCGTGCCC**ATG**CCA
GACCTGAAGAATGTCAAGTCCAAGATCGGCTCCACTGAGAACCTGAAGCAC
CAGCCGGGAGGCGGGAAGGTGCAGATAATTAATAAGAAGCTGGATCTTAGC
AACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAACACGTCCCGGGAG
GCGGCAGTGTGCAAATAGTCTACAAACCAGTTGACCTGAGC**AAG**GTGACCTC
CAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGGAGGTGGCCAG**GTG**
GAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAGTCGAAGATTG
GGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAAAAGGTAAG
AAGCAAGGTTTCATTTAGGGGAAGGGAAATGATTCAGGACGAGAGTCTTTGT
GCTGCTGAGTGCCTGTGATGAAGAAGCATGTTAGTcctgggcaacgtagcgagaccccatct
ctacaaaaaatagaaaaattagccaggtatagtggcgcacacctgtgattccagctacgcaggaggctgaggtgggaggattg
cttgagcccaggaggttgaggctgcagtgagctgtaatcatgccactactccaacctgggcaacacagcaaggaccctgtctca
aaaGCTACTTACAGAAAAGAATTAggctcggcacggtagctcacacctgtaatcccagcactttgggaggc
tgaggcgggcagatcacttgaggtcaggagtttgagaccagcctggccaacatggtgaaaccttgtctctactaaaaatatgaaa
attagccaggcatggtggcacattcctgtaatcccagctactcgggaggctgaggcaggagaatcacttgaacccaggaggtg
gaggttgcagtaagccgagatcgtaccactgtgctctagccttggtgacagagcgagactgtcttaaaaaaaaaaaaaaaaaaa
aaagaattaattaaaaatttaaaaaaaaatgaaaaaaaGCTGCATGCTTGTTTTTTGTTTTTAGTTATT
CTACATTGTTGTCATTATTACCAAATATTGGGGAAAATACAACTTACAGACCA
ATCTCAGGAGTTAAATGTTACTACGAAGGCAAATGAACTATGCGTAATGAAC
CTGGTAGGCATTA

Tau Exon 7 is highlighted in Brown, 3X flag tag in yellow, exon 9 highlighted in Grey, exon 10 is highlighted in Red, Exon 11 in Blue, Exon 12 in Green. ZKSCAN1 Alu elements indicated in lowercase. Highlighted in light green are sequencing primers.

Supplemental Figure 5-2 ZKSCAN1 tau 10-12 WT sequence

AGTGACAGTGGAGATTGTACAGTTTTTTCCTCGATTTGTCAGGAttttttttttttgacgga
gtttaacttcttgtctcccaggtaggaagtgcagtggcgtaatctcggctcactacaacctccacctcctgggttcaagcgtttctcc
tgcctcagctttccgagtagctgggattacaggcgcctgccaccatgccctgctgacttttgtatttttagtagagacggggtttcac
catgttggccaggctggtcttgaactcctgaccgcaggcgattggcctgcctcggcctcccaaagtgctgagattacaggcgtg
agccaccaccccggccTCAGGAGCGTTCTGATAGTGCCTCGATGTGCTGCCTCCTATA
AAGTGTTAGCAGCACAGATCACTTTTTGTAAAGGTACGTACTAATGACTTTTT
TTTTATACTTCAGGTGCAGATAATTAATGACTACAAAGACCATGACGGTGATT
ATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAAGAAGCTGG
ATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAACACGT
CCCGGGAGGCGGCAGTGTGCAAATAGTCTACAAACCAGTTGACCTGAGCAA
GGTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGGAGGT
GGCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAG
TCGAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATA
AAAAGGTAAGAAGCAAGGTTTCATTTAGGGGAAGGGAAATGATTCAGGACG
AGAGTCTTTGTGCTGCTGAGTGCCTGTGATGAAGAAGCATGTTAGTcctgggcaac
gtagcgagaccccatctctacaaaaaatagaaaaattagccaggtatagtggcgcacacctgtgattccagctacgcaggaggc
tgaggtgggaggattgcttgagcccaggaggttgaggctgcagtgagctgtaatcatgccactactccaacctgggcaacaca
gcaaggaccctgtctcaaaaGCTACTTACAGAAAAGAATTAggctcggcacggtagctcacacctgtaatc
ccagcactttgggaggctgaggcgggcagatcacttgaggtcaggagtttgagaccagcctggccaacatggtgaaaccttgt
ctctactaaaaatatgaaaattagccaggcatggtggcacattcctgtaatcccagctactcgggaggctgaggcaggagaatca
cttgaacccaggaggtggaggttgcagtaagccgagatcgtaccactgtgctctagccttggtgacagagcgagactgtcttaa
aaaaaaaaaaaaaaaaaaaagaattaattaaaaatttaaaaaaaaatgaaaaaaaGCTGCATGCTTGTTTTTTG
TTTTTAGTTATTCTACATTGTTGTCATTATTACCAAATATTGGGGAAAATACA
ACTTACAGACCAATCTCAGGAGTTAAATGTTACTACGAAGGCAAATGAACTA
TGCGTAATGAACCTGGTAGGCATTA

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. ZKSCAN1 Alu elements indicated in lowercase. Highlighted in light green are sequencing primers.

Supplemental Figure 5-3 ZKSCAN1 tau 10-12 K317M Sequence

AGTGACAGTGGAGATTGTACAGTTTTTTCCTCGATTTGTCAGGAttttttttttttgacgga
gtttaacttcttgtctcccaggtaggaagtgcagtggcgtaatctcggctcactacaacctccacctcctgggttcaagcgtttctcc
tgcctcagctttccgagtagctgggattacaggcgcctgccaccatgccctgctgacttttgtatttttagtagagacggggtttcac
catgttggccaggctggtcttgaactcctgaccgcaggcgattggcctgcctcggcctcccaaagtgctgagattacaggcgtg
agccaccaccccggccTCAGGAGCGTTCTGATAGTGCCTCGATGTGCTGCCTCCTATA
AAGTGTTAGCAGCACAGATCACTTTTTGTAAAGGTACGTACTAATGACTTTTT
TTTTATACTTCAGGTGCAGATAATTAATGACTACAAAGACCATGACGGTGATT
ATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAAGAAGCTGG
ATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAACACGT
CCCGGGAGGCGGCAGTGTGCAAATAGTCTACAAACCAGTTGACCTGAGC**ATG**
GTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGGAGGTG
GCCAGGTGGAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAGTC
GAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAA
AAGGTAAGAAGCAAGGTTTCATTTAGGGGAAGGGAAATGATTCAGGACGAG
AGTCTTTGTGCTGCTGAGTGCCTGTGATGAAGAAGCATGTTAGTcctgggcaacgtag
cgagaccccatctctacaaaaaatagaaaaattagccaggtatagtggcgcacacctgtgattccagctacgcaggaggctgag
gtgggaggattgcttgagcccaggaggttgaggctgcagtgagctgtaatcatgccactactccaacctgggcaacacagcaa
ggaccctgtctcaaaaGCTACTTACAGAAAAGAATTAggctcggcacggtagctcacacctgtaatcccag
cactttgggaggctgaggcgggcagatcacttgaggtcaggagtttgagaccagcctggccaacatggtgaaaccttgtctcta
ctaaaaatatgaaaattagccaggcatggtggcacattcctgtaatcccagctactcgggaggctgaggcaggagaatcacttga
acccaggaggtggaggttgcagtaagccgagatcgtaccactgtgctctagccttggtgacagagcgagactgtcttaaaaaaa
aaaaaaaaaaaaaaagaattaattaaaaatttaaaaaaaaatgaaaaaaaGCTGCATGCTTGTTTTTTGTTTT
TAGTTATTCTACATTGTTGTCATTATTACCAAATATTGGGGAAAATACAACTT
ACAGACCAATCTCAGGAGTTAAATGTTACTACGAAGGCAAATGAACTATGCG
TAATGAACCTGGTAGGCATTA

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. ZKSCAN1 Alu elements indicated in lowercase. Highlighted in light green are sequencing primers. K317M mutation AAG to ATG is Bold and underlined in exon 11.

Supplemental Figure 5-4 ZKSCAN1 tau 10-12 V337M Sequence

AGTGACAGTGGAGATTGTACAGTTTTTTCCTCGATTTGTCAGGAttttttttttttgacgga
gtttaacttcttgtctcccaggtaggaagtgcagtggcgtaatctcggctcactacaacctccacctcctgggttcaagcgtttctcc
tgcctcagctttccgagtagctgggattacaggcgcctgccaccatgccctgctgacttttgtatttttagtagagacggggtttcac
catgttggccaggctggtcttgaactcctgaccgcaggcgattggcctgcctcggcctcccaaagtgctgagattacaggcgtg
agccaccaccccggccTCAGGAGCGTTCTGATAGTGCCTCGATGTGCTGCCTCCTATA
AAGTGTTAGCAGCACAGATCACTTTTTGTAAAGGTACGTACTAATGACTTTTT
TTTTATACTTCAGGTGCAGATAATTAATGACTACAAAGACCATGACGGTGATT
ATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGAAGAAGCTGG
ATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATAATATCAAACACGT
CCCGGGAGGCGGCAGTGTGCAAATAGTCTACAAACCAGTTGACCTGAGCAAG
GTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAAACCAGGAGGTG
GCCAG**ATG**GAAGTAAAATCTGAGAAGCTTGACTTCAAGGACAGAGTCCAGTC
GAAGATTGGGTCCCTGGACAATATCACCCACGTCCCTGGCGGAGGAAATAAA
AAGGTAAGAAGCAAGGTTTCATTTAGGGGAAGGGAAATGATTCAGGACGAG
AGTCTTTGTGCTGCTGAGTGCCTGTGATGAAGAAGCATGTTAGTcctgggcaacgtag
cgagaccccatctctacaaaaaatagaaaaattagccaggtatagtggcgcacacctgtgattccagctacgcaggaggctgag
gtgggaggattgcttgagcccaggaggttgaggctgcagtgagctgtaatcatgccactactccaacctgggcaacacagcaa
ggaccctgtctcaaaaGCTACTTACAGAAAAGAATTAggctcggcacggtagctcacacctgtaatcccag
cactttgggaggctgaggcgggcagatcacttgaggtcaggagtttgagaccagcctggccaacatggtgaaaccttgtctcta
ctaaaaatatgaaaattagccaggcatggtggcacattcctgtaatcccagctactcgggaggctgaggcaggagaatcacttga
acccaggaggtggaggttgcagtaagccgagatcgtaccactgtgctctagccttggtgacagagcgagactgtcttaaaaaaa
aaaaaaaaaaaaaaagaattaattaaaaatttaaaaaaaaatgaaaaaaaGCTGCATGCTTGTTTTTTGTTTT
TAGTTATTCTACATTGTTGTCATTATTACCAAATATTGGGGAAAATACAACTT
ACAGACCAATCTCAGGAGTTAAATGTTACTACGAAGGCAAATGAACTATGCG
TAATGAACCTGGTAGGCATTA

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. ZKSCAN1 Alu elements indicated in lowercase. Highlighted in light green are sequencing primers. V337M mutation GTG to ATG is Bold and underlined in exon 12.

Supplemental Figure 5-5 Tau authentic alus 10-12 WT sequence

gtaaggaaaccacctttgaaaagaaccaggctgctctgctgtggtttgcaaatgtggggtttgttta***tttg ttttttagcctcaaagacctttcttcaaatgagttctggcatagaagcaccgtgtaaaatagttagaattc tgggcaaaggggaaaagagagctgggggccatccctctcagcaccccacaggctctcatagcagcagctcc taagacacctggtgggacc***ttggtttcgaaatcgctactctaa***ggctgggcacggtggctcacacctgtaa tcccagctctttaggaggccgaggagggtggatcacctgagatcaggagttcgagaccagcctggctaaca tggcaaaaccctgtctctactaaaaatacaaaaattagccgggcgtggtgttatgcgtggtggtaatcgca gctactcgggaggctgaggcacaaggattgcttgaaccccagaggcagaggttgtagttagctccagcttg ggcgacagagcaagaccctgtcgcaaaaattgtttaaaaaacaaa***cccaaaattgctactctcattgggtt ccttgcccattcctgattttggcaagagaaatgcttccagattgccctgatctgggtaggacagcatcac gccatagcaacactgccccgtgagctcactgccccctcaactagcttgtggtccttggttaatgtcagttt ctttttgagtttgtgttatgtctaagggtcatctgctgggtaacggaacccagggactgccctagtccct agactgtgccatgcccgactctgccagctttgtcagtgatgctggtgctcgcctcctcgggtgctcgcctg gtctgagcacacccaaggagttcttgaggccttagggttgtttgcgagagaatgaaagaacacgacctagc tctctttagcatccttggtcaggttcaacactgcccccaggggcctctggtggagccaaccaccatcagcc aaataaatccataattagagtcagaaaatggatgtctgcatatgtgtagtgcactaatgtcctgccgatga ttgacatggagtggagagtgacctgatcattgctgtgagctctgctggccttggcacaactcatgctgata actaatgcacacagttcctctgggaggaaatgtcctcagggaacttggagtttgggtggggatgtgggttt gtgtgcccagcaagcccttgtggttgtagcagacactagtggcatctaggaggcaaagggtcaccccagtc ttagccacgttttgagtcaaggtggcggagtggggctggtgttgactcttggtggcagtaacttttcccaa***tggtgaaaaaccccctctatcatgtttcatttacag**GTGCAGATAATTAAT**GACTACAAAGA CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC AAG**AAGAAGCTGGATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATA ATATCAAACACGTCCCGGGAGGCGGCA***GT*GTGCAAATAGTCTACAAACCA**GTTG ACCTGAGC**AAG**GTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAA ACCAG**GAGGTGGCCAG**GTG**GAAGTAAAATCTGAGAAGCTTGACTTCAAGGA CAGAGTC*CAGTCGAAGATTGGGTCCCT*GGACAATATCACCCACGTCCCTGGCG GAGGAAATAAAAAG***gtaaaggggtagggtgggttggatgctgcccttgggtatatgggcattaa tcaagttgagtggacaaaggctggtccagttcccagaggaggaaaacagaggcttctgtgttgactggctg gatgtgggccctcagcagcatc***cagtgggtctccactgcctgtctcaatcacctggagctttagcacgttt cacacctgggccccaacctggagaggctgaccaatgggtctcaggggcagctcggttgctggagttttt***gt ttttatttatttttatgtatttaaggcagggtctc***tgtattagtccattctcacactgctaataaagacat acccaagactgggtaatttataaaggaaagaggtttaatggactcacagttccatatggctggggaggcct caaaatcatggcggaaggcaaaggagaagcaaaggcatgtcttacatggcaacaggcaagagagcgtgtgc aggggaactcccatttataaaaccatcagacctcatgagatttattcactatcatgagaacagcatgggaa agacccgccccatgattcagttacctcccactgggtccctcccatgacacatggaattatgggagctaca attcaagatgagatttgggtggggacacagccaaaccatatcagtctccctctgtcatccaggctggagtg cactggcatgatctcggctcactgcagcctctacctccctgggtcaggtgatcttcccacctcagcctccc aggtagctggaactacaggtacctgccactatgcctggctaaatattttgtatttcctgtggagacgaggt ttgccacgttgcccaggctggtcttgaactcctgaggtcaagcaatatgcccacctcggcctcccaaggt gctgggattacaggtgtgagccacagtgctcggcc***taagtcactgcag***ttttta***

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. Tau Alu elements indicated in lowercase bold and underlined. Sequencing primers Exon 11 forward: *GTGTGCAAATAGTCTACAAACCA* Exon 12 reverse: AGGGACCCAATCTTCGACTG.

Supplemental Figure 5-6 Tau both authentic alus 10-12 K317M sequence

gtaaggaaaccacctttgaaaagaaccaggctgctctgctgtggtttgcaaatgtggggtttgttta***tttg ttttttagcctcaaagacctttcttcaaatgagttctggcatagaagcaccgtgtaaaatagttagaattc tgggcaaaggggaaaagagagctgggggccatccctctcagcaccccacaggctctcatagcagcagctcc taagacacctggtgggacc***ttggtttcgaaatcgctactctaa***ggctgggcacggtggctcacacctgtaa tcccagctctttaggaggccgaggagggtggatcacctgagatcaggagttcgagaccagcctggctaaca tggcaaaaccctgtctctactaaaaatacaaaaattagccgggcgtggtgttatgcgtggtggtaatcgca gctactcgggaggctgaggcacaaggattgcttgaaccccagaggcagaggttgtagttagctccagcttg ggcgacagagcaagaccctgtcgcaaaaattgtttaaaaaacaaa***cccaaaattgctactctcattgggtt cctttgcccattcctgattttggcaagagaaatgcttccagattgccctgatctgggtaggacagcatcac gccatagcaacactgccccgtgagctcactgccccctcaactagcttgtggtccttggttaatgtcagttt ctttttgagtttgtgttatgtctaagggtcatctgctgggtaacggaacccagggactgccctagtccct agactgtgccatgcccgactctgccagctttgtcagtgatgctggtgctcgcctcctcgggtgctcgcctg gtctgagcacacccaaggagttcttgaggccttagggttgtttgcgagagaatgaaagaacacgacctagc tctctttagcatccttggtcaggttcaacactgcccccaggggcctctggtggagccaaccaccatcagcc aaataaatccataattagagtcagaaaatggatgtctgcatatgtgtagtgcactaatgtcctgccgatga ttgacatggagtggagagtgacctgatcattgctgtgagctctgctggccttggcacaactcatgctgata actaatgcacacagttcctctgggaggaaatgtcctcagggaacttggagtttgggtggggatgtgggttt gtgtgcccagcaagcccttgtggttgtagcagacactagtggcatctaggaggcaaagggtcaccccagtc ttagccacgttttgagtcaaggtggcggagtggggctggtgttgactcttggtggcagtaacttttcccaa***tggtgaaaaaccccctctatcatgtttcatttacag**GTGCAGATAATTAAT**GACTACAAAGA CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC AAG**AAGAAGCTGGATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATA ATATCAAACACGTCCCGGGAGGCGGCA***GT*GTGCAAATAGTCTACAAACCA**GTTG ACCTGAGC**ATG**GTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAA ACCAG**GAGGTGGCCAG**GTG**GAAGTAAAATCTGAGAAGCTTGACTTCAAGGA CAGAGTC*CAGTCGAAGATTGGGTCCCT*GGACAATATCACCCACGTCCCTGGCG GAGGAAATAAAAAG**gtaaaggggtagggtgggttggatgctgcccttgggtatatgggcattaa tcaagttgagtggacaaaggctggtccagttcccagaggaggaaaacagaggcttctgtgttgactggctg gatgtgggccctcagcagcatc***cagtgggtctccactgcctgtctcaatcacctggagctttagcacgttt cacacctgggccccaacctggagaggctgaccaatgggtctcaggggcagctcggttgctggagttttt***gt ttttatttatttttatgtatttaaggcagggtctc***tgtattagtccattctcacactgctaataaagacat acccaagactgggtaatttataaaggaaagaggtttaatggactcacagttccatatggctggggaggcct caaaatcatggcggaaggcaaaggagaagcaaaggcatgtcttacatggcaacaggcaagagagcgtgtgc aggggaactcccatttataaaaccatcagacctcatgagatttattcactatcatgagaacagcatgggaa agacccgccccatgattcagttacctcccactgggtccctcccatgacacatggaattatgggagctaca attcaagatgagatttgggtggggacacagccaaaccatatcagtctccctctgtcatccaggctggagtg cactggcatgatctcggctcactgcagcctctacctccctgggtcaggtgatcttcccacctcagcctccc aggtagctggaactacaggtacctgccactatgcctggctaaatattttgtatttcctgtggagacgaggt ttgccacgttgcccaggctggtcttgaactcctgaggtcaagcaatatgcccacctcggcctcccaaggt gctgggattacaggtgtgagccacagtgctcggcc***taagtcactgcag***ttttta***

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. Tau Alu elements indicated in lowercase bold and underlined. Sequencing primers Exon 11 forward: *GTGTGCAAATAGTCTACAAACCA* Exon 12 reverse: AGGGACCCAATCTTCGACTG. K317M mutation AAG to ATG is Bold and underlined in exon 11.

Supplemental Figure 5-7 Tau both authentic alus 10-12 V337M sequence

gtaaggaaaccacctttgaaaagaaccaggctgctctgctgtggtttgcaaatgtggggtttgttta***tttg ttttttagcctcaaagacctttcttcaaatgagttctggcatagaagcaccgtgtaaaatagttagaattc tgggcaaaggggaaaagagagctgggggccatccctctcagcaccccacaggctctcatagcagcagctcc taagacacctggtgggacc***ttggtttcgaaatcgctactctaa***ggctgggcacggtggctcacacctgtaa tcccagctctttaggaggccgaggagggtggatcacctgagatcaggagttcgagaccagcctggctaaca tggcaaaaccctgtctctactaaaaatacaaaaattagccgggcgtggtgttatgcgtggtggtaatcgca gctactcgggaggctgaggcacaaggattgcttgaaccccagaggcagaggttgtagttagctccagcttg ggcgacagagcaagaccctgtcgcaaaaattgtttaaaaaacaaa***cccaaaattgctactctcattgggtt cctttgcccattcctgattttggcaagagaaatgcttccagattgccctgatctgggtaggacagcatcac gccatagcaacactgccccgtgagctcactgccccctcaactagcttgtggtccttggttaatgtcagttt ctttttgagtttgtgttatgtctaagggtcatctgctgggtaacggaacccagggactgccctagtccct agactgtgccatgcccgactctgccagctttgtcagtgatgctggtgctcgcctcctcgggtgctcgcctg gtctgagcacacccaaggagttcttgaggccttagggttgtttgcgagagaatgaaagaacacgacctagc tctctttagcatccttggtcaggttcaacactgcccccaggggcctctggtggagccaaccaccatcagcc aaataaatccataattagagtcagaaaatggatgtctgcatatgtgtagtgcactaatgtcctgccgatga ttgacatggagtggagagtgacctgatcattgctgtgagctctgctggccttggcacaactcatgctgata actaatgcacacagttcctctgggaggaaatgtcctcagggaacttggagtttgggtggggatgtgggttt gtgtgcccagcaagcccttgtggttgtagcagacactagtggcatctaggaggcaaagggtcaccccagtc ttagccacgttttgagtcaaggtggcggagtggggctggtgttgactcttggtggcagtaacttttcccaa***tggtgaaaaaccccctctatcatgtttcatttacag**GTGCAGATAATTAAT**GACTACAAAGA CCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGAC AAG**AAGAAGCTGGATCTTAGCAACGTCCAGTCCAAGTGTGGCTCAAAGGATA ATATCAAACACGTCCCGGGAGGCGGCA***GT*GTGCAAATAGTCTACAAACCA**GTTG ACCTGAGC**AAG**GTGACCTCCAAGTGTGGCTCATTAGGCAACATCCATCATAA ACCAG**GAGGTGGCCAG**ATG**GAAGTAAAATCTGAGAAGCTTGACTTCAAGGA CAGAGTC*CAGTCGAAGATTGGGTCCCT*GGACAATATCACCCACGTCCCTGGCG GAGGAAATAAAAAG***gtaaaggggtagggtgggttggatgctgcccttgggtatatgggcattaa tcaagttgagtggacaaaggctggtccagttcccagaggaggaaaacagaggcttctgtgttgactggctg gatgtgggccctcagcagcatc***cagtgggtctccactgcctgtctcaatcacctggagctttagcacgttt cacacctgggccccaacctggagaggctgaccaatgggtctcaggggcagctcggttgctggagttttt***gt ttttatttatttttatgtatttaaggcagggtctc***tgtattagtccattctcacactgctaataaagacat acccaagactgggtaatttataaaggaaagaggtttaatggactcacagttccatatggctggggaggcct caaaatcatggcggaaggcaaaggagaagcaaaggcatgtcttacatggcaacaggcaagagagcgtgtgc aggggaactcccatttataaaaccatcagacctcatgagatttattcactatcatgagaacagcatgggaa agacccgccccatgattcagttacctcccactgggtccctcccatgacacatggaattatgggagctaca attcaagatgagatttgggtggggacacagccaaaccatatcagtctccctctgtcatccaggctggagtg cactggcatgatctcggctcactgcagcctctacctccctgggtcaggtgatcttcccacctcagcctccc aggtagctggaactacaggtacctgccactatgcctggctaaatattttgtatttcctgtggagacgaggt ttgccacgttgcccaggctggtcttgaactcctgaggtcaagcaatatgcccacctcggcctcccaaggt gctgggattacaggtgtgagccacagtgctcggcc***taagtcactgcag***ttttta***

Tau Exon 10 is highlighted in Red, 3X flag tag in Yellow, Exon 11 in Blue, Exon 12 in Green. Tau Alu elements indicated in lowercase bold and underlined. Sequencing primers Exon 11 forward: *GTGTGCAAATAGTCTACAAACCA* Exon 12 reverse: AGGGACCCAATCTTCGACTG. V337M mutation GTG to ATG is Bold and underlined in exon 12.

Supplemental Figure 5-8 Description of Expression Constructs

Expression Construct Name	Description
ZKSCAN Alu Tau 10-12 WT	pcDNA3.1 expression vector with CMV promoter. Contains ZKSCAN1 Alu-elements, short intronic regions, on the left and right side of tau cDNA. Contains Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10.
ZKSCAN Alu Tau 10-12 KM	pcDNA3.1 expression vector with CMV promoter. Contains ZKSCAN1 Alu-elements, short intronic regions, on the left and right side of tau cDNA. Contains Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10. FTDP-17 mutation K317M in exon 11.
ZKSCAN Alu Tau 10-12 VM	pcDNA3.1 expression vector with CMV promoter. Contains ZKSCAN1 Alu-elements, short intronic regions, on the left and right side of tau cDNA. Contains Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10. FTDP-17 mutation V337M in exon 12.
ZKSCAN Alu Tau 7-12 WT	pcDNA3.1 expression vector with CMV promoter. Contains ZKSCAN1 Alu-elements, short intronic regions, on the left and right side of tau cDNA. Contains Tau cDNA exons 7,9-12 with a 3X flag tag 12 nt downstream in exon 7.
Tau authentic Alu 10-12 WT	pcDNA3.1 expression vector with CMV promoter. Contains authentic tau Alu-elements, short intronic regions in the left intron of exon 10 along with the MER5A repeat. The right intron of exon 12 contains the MER5B and THE1B repeat with the tau Alu. Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10.
Tau authentic Alu 10-12 KM	pcDNA3.1 expression vector with CMV promoter. Contains authentic tau Alu-elements, short intronic regions in the left intron of exon 10 along with the MER5A repeat. The right intron of exon 12 contains the MER5B and THE1B repeat with the tau Alu. Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10. Contains the FTDP-17 mutation K317M in exon 11.

Tau authentic Alu 10-12 VM	pcDNA3.1 expression vector with CMV promoter. Contains authentic tau Alu-elements, short intronic regions in the left intron of exon 10 along with the MER5A repeat. The right intron of exon 12 contains the MER5B and THE1B repeat with the tau Alu. Tau cDNA exons 10-12 with a 3X flag tag 15 nt downstream in exon 10. Contains the FTDP-17 mutation V337M in exon 12.

CHAPTER 6. CONCLUDING REMARKS

6.1 Summary

6.2 Major Findings

In this dissertation, we have **discovered** that the *MAPT* gene generates two circular RNAs consisting of exons 10-12 (12→10) and exons 7, 9-12 (12→7) (43). The tau locus contains repeat elements that have regions of base complementarity and provide the basis of loop formation. Some of the repeat elements are known as Alu elements, and the tau locus has at least 83 Alu elements, where 56 are on the sense strand and 27 are on the antisense strand (144). In humans, Alu elements comprise around 11% of the human genome (57) and are primate specific. Mice do not have Alu elements or the human specific tau circular RNAs 12→7 and 12→10 and therefore may explain why they do not get Alzheimer's like humans.

We then looked for a correlation of circular RNA abundance and Alzheimer's disease severity and found no statistically significant correlation. The tau circular RNAs have unique properties where both circular RNAs are divisible by three, giving them a reading frame. Neither the 12→7 nor 12→10 circular RNA has a stop codon, and when translation occurs, we predicted it would be translated in a rolling circle giving rise to multiple proteins at different sizes. The 12→7 circular RNA has one methionine start codon in exon 9, where the 12→10 circular RNA does not have a methionine. Therefore, we looked into whether or not the circular RNAs are translated into protein and **identified** that the 12→7 circular RNA is translated forming proteins at approximately 25 kDa, 30 kDa, 40 kDa. The 12→10 circular RNA is only translated when mutations introduce methionine start codons or catalytically active RNA editing enzymes (ADARs) are present. The protein generated from the tau circular RNA is predicted to consist of only the microtubule binding domains of the linear protein, where the 12→7 circular RNA contains repeats R1-R4 and the 12→10 circular RNA contains R2-R4. The K18 neurotoxic peptide consist only of the tau microtubule binding domains, repeat regions R1-R4 (6, 7). The tau circular RNAs protein may also contribute to tau pathology and form neurofibrillary tangles and paired helical filaments.

We identified an interaction and potential **function** of the protein generated from both tau circular RNAs with the eukaryotic initiation factor 4B (EIF4B). The protein generated from the tau circular RNAs may have a physiological role, but when the system gets stressed, there may be a gain of function that deregulates gene regulation and translation.

The novel pathway discovered with the circular tau RNAs can now be a new target for drug therapy. Research has mainly been focused on the linear RNA and its respected protein which could explain why there has not been much advancement in tauopathies. We designed siRNAs to target the backsplice junction of the tau circular RNAs 12→7 and

12→10 so that they will not affect the linear RNA. The siRNAs can be used as a **therapeutic drug** since they prohibit the production of the protein generated from the circular RNAs.

The pathway involving the human tau circular RNAs significantly increases our knowledge of the *MAPT* locus. Now that we have identified a new pathway, further research can look into this mechanism for other tauopathies.

6.3 Limitations and Future Directions

In order to validate circular RNA expression, it must be noted that reverse transcription polymerase chain reaction (RT-PCR) for circular RNAs can give false positives and PCR artifacts. One way we eliminated this, was by validating the circular RNA detection by RNase protection. RNase protection is less sensitive, and by designing probes that bind to the junction, one will see the hybridized probe protected from RNase digestion.

Also, when transfecting cells with pcDNA3.1 expression vector that contains a CMV promoter, multiple copies of the transfected DNA are incorporated into the genome. Due to multiple copies being incorporated the results of the circular RNA from cells could also give a false positive and actually be a concatemer. The concatemer would be the linear repeat giving rise to the exon junction that would appear to only be from the circular RNA. The tau circular RNAs are expressed in human brain tissue. Therefore, we use the cell culture expressing the tau circular RNAs to detect the protein.

We are currently making stable cell lines that will incorporate only one copy into the genome using HEK 293 Flp-In T-REx cells. HEK 293 Flp-In T-REx cells contain a single stably integrated FRT site at a transcriptionally active genomic locus. Co-transfecting an expression vector, pcDNA 5/FRT that contains your gene of interest with the Flp recombinase vector, pOG44, results in target integration of the expression vector to the same locus in every cell, ensuring homogenous levels of gene expression. The tau cDNA constructs are cloned into the pcDNA 5/FRT expression vector. The polyA tail was removed so that this expression construct would only give rise to the circular RNAs and have one copy incorporated into the genome. The HEK 293 Flp-In T-REx cell line will validate the authenticity of the circular RNAs being made are circular and not linear and that they are not concatemers.

We are analyzing human brain tissue by Mass Spectrometry to detect the protein encoded by the backsplice junction generated from the circular 12→7 RNA. The expected protein sequence is rather difficult to analyze because the junction sequence contains two lysine residues that will be cleaved by trypsin. Trypsin is a common digestive enzyme used in mass spectrometry because it is specific and active. The enzymes we need to digest the protein to obtain the backsplice junction are less specific and are being optimized under different conditions.

We are moving our model *in vivo* using zebrafish to test the pathology generated from the wild type and mutant circular RNAs. So far, the zebrafish are sterile. Different human brain tissue regions will also be analyzed and compared for circular RNA abundance and severity of Alzheimer's pathology.

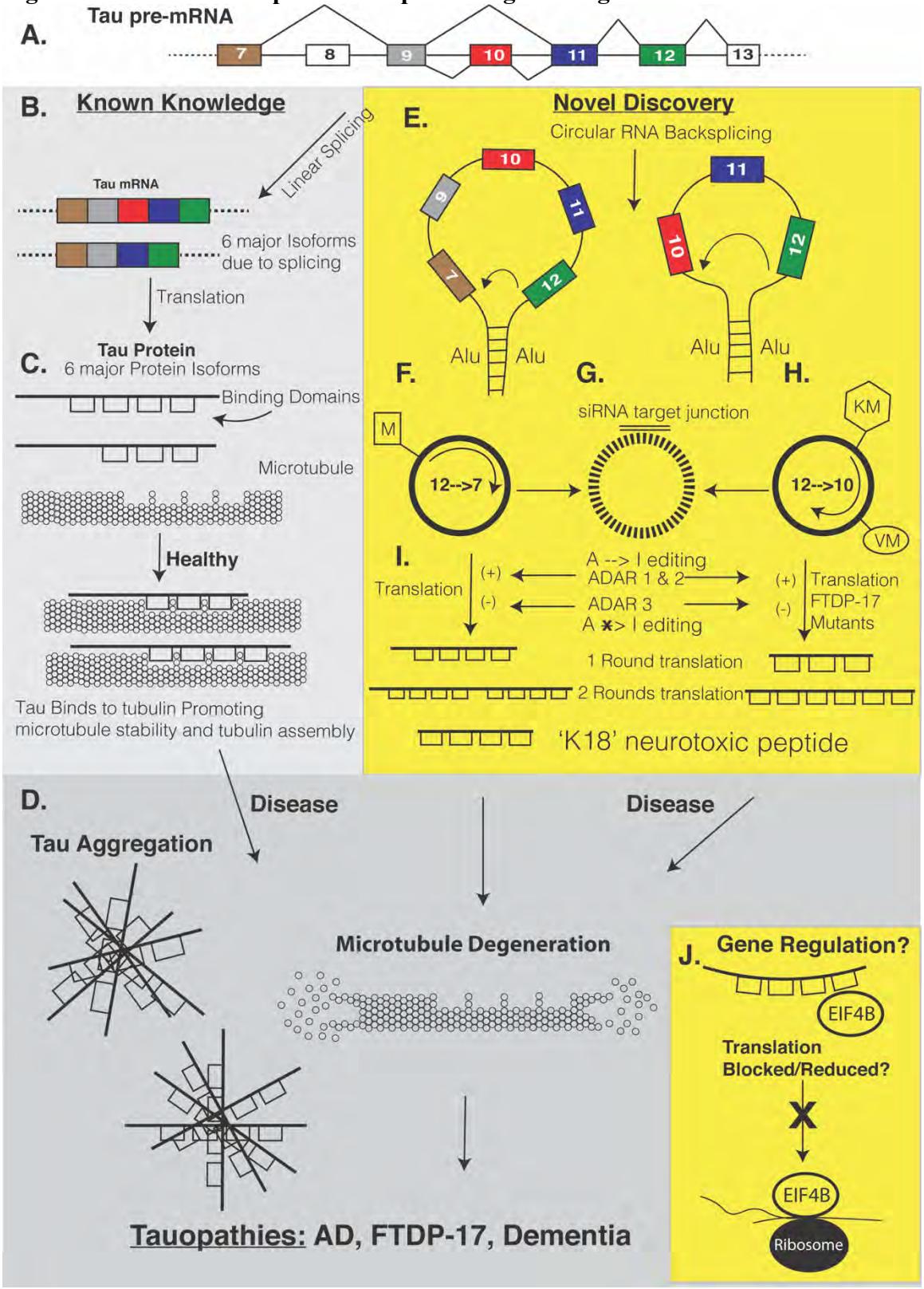
6.4 Conclusion

We propose a model that adds a new feature (Figure 6.1 in Yellow) to the human tau biology. The *MAPT* gene generates circular RNAs that contain exons 7,9-12 and 10-12 due to exon 12 backsplicing. In Alzheimer's disease it has been shown that there is more exon 10 inclusion in these patients. However, an increase in exon 10 inclusion does not affect circular RNA abundance, nor is there an effect on circular RNA abundance and Alzheimer's severity indicated by Braak stages. The circular RNAs are cytosolic and in synaptosomes indicating they can play a role in translation or gene regulation.

These circular RNAs are naturally occurring in humans where the 12→7 is translated due to the methionine start codon in exon 9. The 12→10 circular RNA is only translated when mutations introduce start codons or RNA editing enzymes (ADARs) are present. Since EIF4B is required for translational initiation of highly structured mRNAs, the circular tau RNAs function may be to regulate gene expression due to the interaction with EIF4B. ADAR enzymes have been shown to be deregulated in Alzheimer's disease (145) and have an effect on tau circular RNA translation. The circular RNAs may be edited creating new start codons enabling more protein to be produced. Increasing protein production may lead to tau aggregation and neurodegeneration.

Thus, this work allows a novel therapeutic approach with siRNAs that target the tau circular RNAs backsplice junction reducing their expression and prohibiting translation, as an intervention for Alzheimer's disease.

Figure 6-1 Tau model of pre-mRNA processing forming linear and circular RNAs



Known knowledge is in gray and my finding is in yellow.

A. Schematic of the shortened pre-mRNA of the tau locus and how it is alternatively spliced.
B. After alternative splicing six major RNA isoforms are formed.
C. The six major protein isoforms are translated and interacts with tubulin stabilizing microtubules.
D. When tau no longer binds to tubulin and forms tau aggregates, disease pathology forms, causing tubulin disassembly leading to Tauopathies.
E. The novel discovery in yellow, two circular RNAs form due to the backsplicing of exon 12 onto either exon 7 or exon 10 (12→7, 12→10).
F. 12→7 circular tau RNA is translated in a rolling circle.
G. The circular tau RNAs can be destroyed or reduced with siRNAs targeting the backsplice junction inhibiting protein translation.
H. The 12→10 circular RNA is translated with FTDP-17 mutations K317M and V337M. Editing enzymes (ADAR1 and ADAR2) increase protein production due to RNA editing. The translated regions of the circular tau RNA 12→10 is predicted to be similar to the neurotoxic peptide called K18.
I. Both circular RNAs 12→ and 12→10 are translated forming multiple proteins of different sizes, a possible indication of multiple rounds of translation. Editing enzymes (ADAR1 and ADAR2) increase protein production due to RNA editing. The translated region of the circular tau RNAs is predicted to be similar to the neurotoxic peptide called K18.
J. EIF4B interacts with the protein from the tau circular RNAs possibly influencing translation

APPENDIX: LIST OF ABBREVIATIONS

ADAR	Adenosine Deaminase Acting on RNA
ADAR1	ADAR
ADAR2	ADARB1
ADAR3	ADARB2
AD	Alzheimer's Disease
DNA	Deoxyribonucleic Acid
EIF4B	Eukaryotic Initiation Factor 4B
FTDP-17	Frontotemporal Dementia with Parkinsonism linked to Chromosome 17
HEK 293T	Human Embryonic Kidney Cells
IP	Immunoprecipitation
KM	Lysine to Methionine Mutation
MAPT	Microtubule Associated Protein Tau
mRNA	Messenger RNA
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEI	Polyethyleneimine
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RNA	Ribonucleic Acid
RPA	RNase Protection Assay
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulphate
siRNA	Small Interfering RNA
VM	Valine to Methionine Mutation
WB	Western Blot
WT	Wild Type

BIBLIOGRAPHY

1. Welden JR, van Doorn J, Nelson PT, Stamm S. The human MAPT locus generates circular RNAs. Biochim Biophys Acta Mol Basis Dis. 2018:2753-60. doi: 10.1016/j.bbadis.2018.04.023. PubMed PMID: 29729314.
2. Iqbal K, Liu F, Gong CX. Tau and neurodegenerative disease: the story so far. Nat Rev Neurol. 2016;12(1):15-27. doi: 10.1038/nrneurol.2015.225. PubMed PMID: 26635213.
3. Spillantini MG, Goedert M. Tau protein pathology in neurodegenerative diseases. Trends Neurosci. 1998;21(10):428-33. doi: 10.1016/s0166-2236(98)01337-x. PubMed PMID: 9786340.
4. Welden JR, Stamm S. Pre-mRNA structures forming circular RNAs. Biochim Biophys Acta Gene Regul Mech. 2019:194410. Epub 2019/08/20. doi: 10.1016/j.bbagrm.2019.194410. PubMed PMID: 31421281.
5. Welden JR, Pawluchin A, van Doorn J, Stamm S. Use of Alu Element Containing Minigenes to Analyze Circular RNAs. J Vis Exp. 2020(157). Epub 2020/03/10. doi: 10.3791/59760. PubMed PMID: 32225139.
6. Shammas SL, Garcia GA, Kumar S, Kjaergaard M, Horrocks MH, Shivji N, Mandelkow E, Knowles TP, Klenerman D. A mechanistic model of tau amyloid aggregation based on direct observation of oligomers. Nat Commun. 2015;6:7025. Epub 2015/04/30. doi: 10.1038/ncomms8025. PubMed PMID: 25926130; PMCID: PMC4421837.
7. Guo JL, Lee VM. Neurofibrillary tangle-like tau pathology induced by synthetic tau fibrils in primary neurons over-expressing mutant tau. FEBS Lett. 2013;587(6):717-23. Epub 2013/02/05. doi: 10.1016/j.febslet.2013.01.051. PubMed PMID: 23395797; PMCID: PMC3678381.
8. Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, Dinneny JR, Brown PO, Salzman J. Circular RNA is expressed across the eukaryotic tree of life. PLoS One. 2014;9(6):e90859. doi: 10.1371/journal.pone.0090859. PubMed PMID: 24609083; PMCID: PMC3946582.
9. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. Rna. 2013;19(2):141-57. Epub 2012/12/20. doi: 10.1261/rna.035667.112. PubMed PMID: 23249747; PMCID: 3543092.
10. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, Zhu S, Yang L, Chen LL. Circular intronic long noncoding RNAs. Molecular cell. 2013;51(6):792-806. Epub 2013/09/17. doi: 10.1016/j.molcel.2013.08.017. PubMed PMID: 24035497.
11. Lu Z, Filonov GS, Noto JJ, Schmidt CA, Hatkevich TL, Wen Y, Jaffrey SR, Matera AG. Metazoan tRNA introns generate stable circular RNAs in vivo. RNA. 2015;21(9):1554-65. doi: 10.1261/rna.052944.115. PubMed PMID: 26194134; PMCID: PMC4536317.
12. Zaug AJ, Grabowski PJ, Cech TR. Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. Nature. 1983;301(5901):578-83. PubMed PMID: 6186917.
13. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like

structures. Proc Natl Acad Sci U S A. 1976;73(11):3852-6. PubMed PMID: 1069269; PMCID: PMC431239.
14. Toptan T, Abere B, Nalesnik MA, Swerdlow SH, Ranganathan S, Lee N, Shair KH, Moore PS, Chang Y. Circular DNA tumor viruses make circular RNAs. Proc Natl Acad Sci U S A. 2018;115(37):E8737-E45. doi: 10.1073/pnas.1811728115. PubMed PMID: 30150410; PMCID: PMC6140489.
15. Cocquerelle C, Mascrez B, Hetuin D, Bailleul B. Mis-splicing yields circular RNA molecules. FASEB J. 1993;7(1):155-60. PubMed PMID: 7678559.
16. Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. PLoS One. 2012;7(2):e30733. doi: 10.1371/journal.pone.0030733. PubMed PMID: 22319583; PMCID: PMC3270023.
17. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. Nat Biotechnol. 2014;32(5):453-61. doi: 10.1038/nbt.2890. PubMed PMID: 24811520; PMCID: PMC4121655.
18. Huang C, Liang D, Tatomer DC, Wilusz JE. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. Genes Dev. 2018;32(9-10):639-44. doi: 10.1101/gad.314856.118. PubMed PMID: 29773557; PMCID: PMC6004072.
19. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. Natural RNA circles function as efficient microRNA sponges. Nature. 2013;495(7441):384-8. doi: 10.1038/nature11993. PubMed PMID: 23446346.
20. Panda AC. Circular RNAs Act as miRNA Sponges. Adv Exp Med Biol. 2018;1087:67-79. doi: 10.1007/978-981-13-1426-1_6. PubMed PMID: 30259358.
21. Ragan C, Goodall GJ, Shirokikh NE, Preiss T. Insights into the biogenesis and potential functions of exonic circular RNA. Sci Rep. 2019;9(1):2048. doi: 10.1038/s41598-018-37037-0. PubMed PMID: 30765711; PMCID: PMC6376117.
22. Abe N, Matsumoto K, Nishihara M, Nakano Y, Shibata A, Maruyama H, Shuto S, Matsuda A, Yoshida M, Ito Y, Abe H. Rolling Circle Translation of Circular RNA in Living Human Cells. Sci Rep. 2015;5:16435. doi: 10.1038/srep16435. PubMed PMID: 26553571; PMCID: PMC4639774.
23. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatica A, Santini T, Andronache A, Wade M, Laneve P, Rajewsky N, Bozzoni I. Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. Mol Cell. 2017;66(1):22-37 e9. doi: 10.1016/j.molcel.2017.02.017. PubMed PMID: 28344082; PMCID: PMC5387670.
24. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, Xu B, Wu C, Zhou Q, Hu W, Wu C, Jiang J. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via Hippo-YAP signaling. Mol Cancer. 2019;18(1):47. doi: 10.1186/s12943-019-1010-6. PubMed PMID: 30925892; PMCID: PMC6440158.
25. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, Liu H, Xu J, Xiao F, Zhou H, Yang X, Huang N, Liu J, He K, Xie K, Zhang G, Huang S, Zhang N. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. Nat Commun. 2018;9(1):4475. doi: 10.1038/s41467-018-06862-2. PubMed PMID: 30367041; PMCID: PMC6203777.

26. Li XF, Lytton J. A circularized sodium-calcium exchanger exon 2 transcript. J Biol Chem. 1999;274(12):8153-60. PubMed PMID: 10075718.
27. Costello A, Lao NT, Barron N, Clynes M. Continuous translation of circularized mRNA improves recombinant protein titer. Metab Eng. 2019;52:284-92. doi: 10.1016/j.ymben.2019.01.002. PubMed PMID: 30615942.
28. Westholm JO, Miura P, Olson S, Shenker S, Joseph B, Sanfilippo P, Celniker SE, Graveley BR, Lai EC. Genome-wide analysis of drosophila circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation. Cell Rep. 2014;9(5):1966-80. Epub 2014/11/26. doi: 10.1016/j.celrep.2014.10.062. PubMed PMID: 25544350; PMCID: PMC4279448.
29. Rybak-Wolf A, Stottmeister C, Glazar P, Jens M, Pino N, Giusti S, Hanan M, Behm M, Bartok O, Ashwal-Fluss R, Herzog M, Schreyer L, Papavasileiou P, Ivanov A, Ohman M, Refojo D, Kadener S, Rajewsky N. Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. Mol Cell. 2015;58(5):870-85. doi: 10.1016/j.molcel.2015.03.027. PubMed PMID: 25921068.
30. Veno MT, Hansen TB, Veno ST, Clausen BH, Grebing M, Finsen B, Holm IE, Kjems J. Spatio-temporal regulation of circular RNA expression during porcine embryonic brain development. Genome Biol. 2015;16:245. doi: 10.1186/s13059-015-0801-3. PubMed PMID: 26541409; PMCID: PMC4635978.
31. Chao CW, Chan DC, Kuo A, Leder P. The mouse formin (Fmn) gene: abundant circular RNA transcripts and gene-targeted deletion analysis. Mol Med. 1998;4(9):614-28. PubMed PMID: 9848078; PMCID: PMC2230310.
32. Gualandi F, Trabanelli C, Rimessi P, Calzolari E, Toffolatti L, Patarnello T, Kunz G, Muntoni F, Ferlini A. Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. J Med Genet. 2003;40(8):e100. PubMed PMID: 12920092; PMCID: PMC1735543.
33. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. Cell. 2014;159(1):134-47. Epub 2014/09/18. doi: 10.1016/j.cell.2014.09.001. PubMed PMID: 25242744.
34. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, Chen LL, Yang L. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. Genome Res. 2016;26(9):1277-87. Epub 2016/06/30. doi: 10.1101/gr.202895.115. PubMed PMID: 27365365; PMCID: PMC5052039.
35. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA. 2014;20(11):1666-70. Epub 2014/09/18. doi: 10.1261/rna.043687.113. PubMed PMID: 25234927; PMCID: PMC4201819.
36. Meng X, Hu D, Zhang P, Chen Q, Chen M. CircFunBase: a database for functional circular RNAs. Database (Oxford). 2019;2019. Epub 2019/01/01. doi: 10.1093/database/baz003. PubMed PMID: 30715276; PMCID: PMC6360206.
37. Liu YC, Li JR, Sun CH, Andrews E, Chao RF, Lin FM, Weng SL, Hsu SD, Huang CC, Cheng C, Liu CC, Huang HD. CircNet: a database of circular RNAs derived from transcriptome sequencing data. Nucleic Acids Res. 2016;44(D1):D209-15. Epub 2015/10/07. doi: 10.1093/nar/gkv940. PubMed PMID: 26450965; PMCID: PMC4702939.
38. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. Sci Rep.

2016;6:34985. Epub 2016/10/11. doi: 10.1038/srep34985. PubMed PMID: 27725737; PMCID: PMC5057092.
39. Chu Q, Zhang X, Zhu X, Liu C, Mao L, Ye C, Zhu QH, Fan L. PlantcircBase: A Database for Plant Circular RNAs. Mol Plant. 2017;10(8):1126-8. Epub 2017/03/16. doi: 10.1016/j.molp.2017.03.003. PubMed PMID: 28315753.
40. Starke S, Jost I, Rossbach O, Schneider T, Schreiner S, Hung LH, Bindereif A. Exon circularization requires canonical splice signals. Cell Rep. 2015;10(1):103-11. Epub 2014/12/24. doi: 10.1016/j.celrep.2014.12.002. PubMed PMID: 25543144.
41. Liang D, Tatomer DC, Luo Z, Wu H, Yang L, Chen LL, Cherry S, Wilusz JE. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. Mol Cell. 2017;68(5):940-54.e3. Epub 2017/11/22. doi: 10.1016/j.molcel.2017.10.034. PubMed PMID: 29174924; PMCID: PMC5728686.
42. Kramer MC, Liang D, Tatomer DC, Gold B, March ZM, Cherry S, Wilusz JE. Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. Genes Dev. 2015;29(20):2168-82. Epub 2015/10/08. doi: 10.1101/gad.270421.115. PubMed PMID: 26450910; PMCID: PMC4617980.
43. Welden JR, van Doorn J, Nelson PT, Stamm S. The human MAPT locus generates circular RNAs. Biochim Biophys Acta. 2018;1864(9 Pt B):2753-60. Epub 2018/05/03. doi: 10.1016/j.bbadis.2018.04.023. PubMed PMID: 29729314.
44. Ottesen EW, Luo D, Seo J, Singh NN, Singh RN. Human Survival Motor Neuron genes generate a vast repertoire of circular RNAs. Nucleic Acids Res. 2019;47(6):2884-905. doi: 10.1093/nar/gkz034. PubMed PMID: 30698797; PMCID: PMC6451121.
45. Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. Genes Dev. 2014;28(20):2233-47. doi: 10.1101/gad.251926.114. PubMed PMID: 25281217; PMCID: PMC4201285.
46. Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, Roslan S, Schreiber AW, Gregory PA, Goodall GJ. The RNA binding protein quaking regulates formation of circRNAs. Cell. 2015;160(6):1125-34. doi: 10.1016/j.cell.2015.02.014. PubMed PMID: 25768908.
47. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, Evantal N, Memczak S, Rajewsky N, Kadener S. circRNA biogenesis competes with pre-mRNA splicing. Mol Cell. 2014;56(1):55-66. Epub 2014/09/18. doi: 10.1016/j.molcel.2014.08.019. PubMed PMID: 25242144.
48. Nasim FU, Hutchison S, Cordeau M, Chabot B. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. RNA. 2002;8(8):1078-89. doi: 10.1017/s1355838202024056. PubMed PMID: 12212851; PMCID: PMC1370318.
49. Lamichhane R, Daubner GM, Thomas-Crusells J, Auweter SD, Manatschal C, Austin KS, Valniuk O, Allain FH, Rueda D. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proc Natl Acad Sci U S A. 2010;107(9):4105-10. Epub 2010/02/16. doi: 10.1073/pnas.0907072107. PubMed PMID: 20160105; PMCID: PMC2840148.
50. Barrett SP, Wang PL, Salzman J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor. Elife. 2015;4:e07540. Epub 2015/06/09. doi: 10.7554/eLife.07540. PubMed PMID: 26057830; PMCID: PMC4479058.

51. Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, Goodfellow P, Lovell-Badge R. Circular transcripts of the testis-determining gene Sry in adult mouse testis. Cell. 1993;73(5):1019-30. doi: 10.1016/0092-8674(93)90279-y. PubMed PMID: 7684656.
52. Dubin RA, Kazmi MA, Ostrer H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. Gene. 1995;167(1-2):245-8. doi: 10.1016/0378-1119(95)00639-7. PubMed PMID: 8566785.
53. Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, Piechotta M, Levanon EY, Landthaler M, Dieterich C, Rajewsky N. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. Cell Rep. 2015;10(2):170-7. Epub 2014/12/31. doi: 10.1016/j.celrep.2014.12.019. PubMed PMID: 25558066.
54. Shen M, Eyras E, Wu J, Khanna A, Josiah S, Rederstorff M, Zhang MQ, Stamm S. Direct cloning of double-stranded RNAs from RNase protection analysis reveals processing patterns of C/D box snoRNAs and provides evidence for widespread antisense transcript expression. Nucleic Acids Res. 2011;39(22):9720-30. Epub 2011/08/31. doi: 10.1093/nar/gkr684. PubMed PMID: 21880592; PMCID: PMC3239178.
55. Chen LL, Yang L. ALUternative Regulation for Gene Expression. Trends Cell Biol. 2017;27(7):480-90. Epub 2017/02/10. doi: 10.1016/j.tcb.2017.01.002. PubMed PMID: 28209295.
56. Häsler J, Samuelsson T, Strub K. Useful 'junk': Alu RNAs in the human transcriptome. Cell Mol Life Sci. 2007;64(14):1793-800. doi: 10.1007/s00018-007-7084-0. PubMed PMID: 17514354.
57. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011;12(12):236. doi: 10.1186/gb-2011-12-12-236. PubMed PMID: 22204421; PMCID: PMC3334610.
58. Quentin Y. Fusion of a free left Alu monomer and a free right Alu monomer at the origin of the Alu family in the primate genomes. Nucleic Acids Res. 1992;20(3):487-93. doi: 10.1093/nar/20.3.487. PubMed PMID: 1741283; PMCID: PMC310412.
59. Quentin Y. Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. Nucleic Acids Res. 1992;20(13):3397-401. doi: 10.1093/nar/20.13.3397. PubMed PMID: 1378589; PMCID: PMC312495.
60. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe

C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Consortium IHGS. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921. doi: 10.1038/35057062. PubMed PMID: 11237011.

61. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB. Mobile elements create structural variation: analysis of a complete human genome. Genome Res. 2009;19(9):1516-26. Epub 2009/05/13. doi: 10.1101/gr.091827.109. PubMed PMID: 19439515; PMCID: PMC2752133.

62. Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res. 2003;13(9):1998-2004. Epub 2003/08/12. doi: 10.1101/gr.1649303. PubMed PMID: 12915492; PMCID: PMC403669.

63. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. Standardized nomenclature for Alu repeats. J Mol Evol. 1996;42(1):3-6. doi: 10.1007/BF00163204. PubMed PMID: 8576960.

64. Bazak L, Levanon EY, Eisenberg E. Genome-wide analysis of Alu editability. Nucleic Acids Res. 2014;42(11):6876-84. doi: 10.1093/nar/gku414. PubMed PMID: 24829451; PMCID: PMC4066801.

65. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2004;2(12):e391. Epub 2004/11/09. doi: 10.1371/journal.pbio.0020391. PubMed PMID: 15534692; PMCID: PMC526178.

66. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. Nat Rev Mol Cell Biol. 2016;17(2):83-96. doi: 10.1038/nrm.2015.4. PubMed PMID: 26648264; PMCID: PMC4824625.

67. Aktas T, Avsar Ilik I, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, Manke T, Backofen R, Akhtar A. DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. Nature. 2017;544(7648):115-9. doi: 10.1038/nature21715. PubMed PMID: 28355180.

68. Chuang TJ, Chen YJ, Chen CY, Mai TL, Wang YD, Yeh CS, Yang MY, Hsiao YT, Chang TH, Kuo TC, Cho HH, Shen CN, Kuo HC, Lu MY, Chen YH, Hsieh SC, Chiang TW. Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-backsplicing in human cells. Nucleic Acids Res. 2018;46(7):3671-91. doi: 10.1093/nar/gky032. PubMed PMID: 29385530; PMCID: PMC6283421.
69. Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E, Shenzis S, Samson M, Dittmar G, Landthaler M, Chekulaeva M, Rajewsky N, Kadener S. Translation of CircRNAs. Mol Cell. 2017;66(1):9-21 e7. doi: 10.1016/j.molcel.2017.02.021. PubMed PMID: 28344080; PMCID: PMC5387669.
70. Siegrist W, Solca F, Stutz S, Giuffre L, Carrel S, Girard J, Eberle AN. Characterization of receptors for alpha-melanocyte-stimulating hormone on human melanoma cells. Cancer Res. 1989;49(22):6352-8. PubMed PMID: 2804981.
71. Mattick JS, Mehler MF. RNA editing, DNA recoding and the evolution of human cognition. Trends Neurosci. 2008;31(5):227-33. doi: 10.1016/j.tins.2008.02.003. PubMed PMID: 18395806.
72. Wang W, Kirkness EF. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res. 2005;15(12):1798-808. doi: 10.1101/gr.3765505. PubMed PMID: 16339378; PMCID: PMC1356118.
73. Diederichs S, Bartsch L, Berkmann JC, Frose K, Heitmann J, Hoppe C, Iggena D, Jazmati D, Karschnia P, Linsenmeier M, Maulhardt T, Mohrmann L, Morstein J, Paffenholz SV, Ropenack P, Ruckert T, Sandig L, Schell M, Steinmann A, Voss G, Wasmuth J, Weinberger ME, Wullenkord R. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. EMBO Mol Med. 2016;8(5):442-57. Epub 2016/03/20. doi: 10.15252/emmm.201506055. PubMed PMID: 26992833.
74. Vaz-Drago R, Custodio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017;136(9):1093-111. doi: 10.1007/s00439-017-1809-4. PubMed PMID: 28497172.
75. Meyer KD, Jaffrey SR. Rethinking m(6)A Readers, Writers, and Erasers. Annu Rev Cell Dev Biol. 2017;33:319-42. doi: 10.1146/annurev-cellbio-100616-060758. PubMed PMID: 28759256; PMCID: PMC5963928.
76. Liu N, Zhou KI, Parisien M, Dai Q, Diatchenko L, Pan T. N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. Nucleic Acids Res. 2017;45(10):6051-63. doi: 10.1093/nar/gkx141. PubMed PMID: 28334903; PMCID: PMC5449601.
77. Wang Y, Mandelkow E. Tau in physiology and pathology. Nat Rev Neurosci. 2016;17(1):5-21. doi: 10.1038/nrn.2015.1. PubMed PMID: 26631930.
78. Lee VM, Goedert M, Trojanowski JQ. Neurodegenerative tauopathies. Annu Rev Neurosci. 2001;24:1121-59. doi: 10.1146/annurev.neuro.24.1.1121. PubMed PMID: 11520930.
79. van Swieten J, Spillantini MG. Hereditary frontotemporal dementia caused by Tau gene mutations. Brain Pathol. 2007;17(1):63-73. doi: 10.1111/j.1750-3639.2007.00052.x. PubMed PMID: 17493040.
80. Roberson ED, Scearce-Levie K, Palop JJ, Yan F, Cheng IH, Wu T, Gerstein H, Yu GQ, Mucke L. Reducing endogenous tau ameliorates amyloid beta-induced deficits in an

Alzheimer's disease mouse model. Science. 2007;316(5825):750-4. doi: 10.1126/science.1141736. PubMed PMID: 17478722.
81. 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. 1990;9(13):4225-30. PubMed PMID: 2124967; PMCID: PMC552204.
82. Glatz DC, Rujescu D, Tang Y, Berendt FJ, Hartmann AM, Faltraco F, Rosenberg C, Hulette C, Jellinger K, Hampel H, Riederer P, Moller HJ, Andreadis A, Henkel K, Stamm S. The alternative splicing of tau exon 10 and its regulatory proteins CLK2 and TRA2-BETA1 changes in sporadic Alzheimer's disease. J Neurochem. 2006;96:635-44. PubMed PMID: 16371011.
83. Conrad C, Zhu J, Schoenfeld D, Fang Z, Ingelsson M, Stamm S, Church G, Hyman BT. Single molecule profiling of tau gene expression in Alzheimer's disease. J Neurochem. 2007;103(3):1228-36. PubMed PMID: 17727636.
84. Suzuki H, Tsukahara T. A view of pre-mRNA splicing from RNase R resistant RNAs. Int J Mol Sci. 2014;15(6):9331-42. Epub 2014/05/29. doi: 10.3390/ijms15069331. PubMed PMID: 24865493; PMCID: 4100097.
85. Hartmann AM, Rujescu D, Giannakouros T, Nikolakaki E, Goedert M, Mandelkow EM, Gao QS, Andreadis A, Stamm S. Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors. Mol Cell Neurosci. 2001;18(1):80-90.
86. Chen LL. The biogenesis and emerging roles of circular RNAs. Nat Rev Mol Cell Biol. 2016;17(4):205-11. doi: 10.1038/nrm.2015.32. PubMed PMID: 26908011.
87. Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. Intronic Alus influence alternative splicing. PLoS Genet. 2008;4(9):e1000204. doi: 10.1371/journal.pgen.1000204. PubMed PMID: 18818740; PMCID: PMC2533698.
88. Wang Y, Wang Z. Efficient backsplicing produces translatable circular mRNAs. RNA. 2015;21(2):172-9. doi: 10.1261/rna.048272.114. PubMed PMID: 25449546; PMCID: PMC4338345.
89. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, Zhong G, Yu B, Hu W, Dai L, Zhu P, Chang Z, Wu Q, Zhao Y, Jia Y, Xu P, Liu H, Shan G. Exon-intron circular RNAs regulate transcription in the nucleus. Nat Struct Mol Biol. 2015;22(3):256-64. doi: 10.1038/nsmb.2959. PubMed PMID: 25664725.
90. Nelson PT, Jicha GA, Schmitt FA, Liu H, Davis DG, Mendiondo MS, Abner EL, Markesbery WR. Clinicopathologic correlations in a large Alzheimer disease center autopsy cohort: neuritic plaques and neurofibrillary tangles "do count" when staging disease severity. J Neuropathol Exp Neurol. 2007;66(12):1136-46. doi: 10.1097/nen.0b013e31815c5efb. PubMed PMID: 18090922; PMCID: PMC3034246.
91. Stoss O, Stoilov P, Hartmann AM, Nayler O, Stamm S. The in vivo minigene approach to analyze tissue-specific splicing. Brain Res Brain Res Protoc. 1999;4(3):383-94. doi: 10.1016/s1385-299x(99)00043-4. PubMed PMID: 10592349.
92. Falaleeva M, Pages A, Matuszek Z, Hidmi S, Agranat-Tamir L, Korotkov K, Nevo Y, Eyras E, Sperling R, Stamm S. Dual function of C/D box snoRNAs in rRNA modification and alternative pre-mRNA splicing Proc Natl Acad Sci U S A. 2016;113(12):E1625-34. doi: 10.1073/pnas.1519292113. PubMed PMID: 26957605.
93. Shen M, Eyras E, Wu J, Khanna A, Josiah S, Rederstorff M, Zhang MQ, Stamm S. Direct cloning of double-stranded RNAs from RNase protection analysis reveals processing patterns of C/D box snoRNAs and provides evidence for widespread antisense

transcript expression. Nucleic acids research. 2011;39(22):9720-30. Epub 2011/09/02. doi: 10.1093/nar/gkr684. PubMed PMID: 21880592; PMCID: 3239178.
94. Ding S, Shi J, Qian W, Iqbal K, Grundke-Iqbal I, Gong CX, Liu F. Regulation of alternative splicing of tau exon 10 by 9G8 and Dyrk1A. Neurobiol Aging. 2012;33(7):1389-99. doi: 10.1016/j.neurobiolaging.2010.11.021. PubMed PMID: 21215488; PMCID: PMC3085640.
95. 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. 2007;34(1):48-58. doi: 10.1016/j.mcn.2006.10.004. PubMed PMID: 17137791; PMCID: PMC1866282.
96. Novoyatleva T, Heinrich B, Tang Y, Benderska N, Butchbach ME, Lorson CL, Lorson MA, Ben-Dov C, Fehlbaum P, Bracco L, Burghes AH, Bollen M, Stamm S. Protein phosphatase 1 binds to the RNA recognition motif of several splicing factors and regulates alternative pre-mRNA processing. Hum Mol Genet. 2008:52-70.
97. Jamros MA, Aubol BE, Keshwani MM, Zhang Z, Stamm S, Adams JA. Intra-domain Cross-talk Regulates Serine-arginine Protein Kinase 1-dependent Phosphorylation and Splicing Function of Transformer 2β1. J Biol Chem. 2015;290(28):17269-81. Epub 2015/05/26. doi: 10.1074/jbc.M115.656579. PubMed PMID: 26013829; PMCID: PMC4498066.
98. Clery A, Jayne S, Benderska N, Dominguez C, Stamm S, Allain FH. Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-beta1. Nature structural & molecular biology. 2011;18(4):443-50. Epub 2011/03/15. doi: 10.1038/nsmb.2001. PubMed PMID: 21399644.
99. Duncan PI, Stojdl DF, Marius RM, Bell JC. In vivo regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. Mol Cell Biol. 1997;17(10):5996-6001. doi: 10.1128/mcb.17.10.5996. PubMed PMID: 9315658; PMCID: PMC232448.
100. Andreadis A. Tau gene alternative splicing: expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. Biochem Biophys Acta. 2005;1739:91-103.
101. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, Loewer A, Ziebold U, Landthaler M, Kocks C, le Noble F, Rajewsky N. Circular RNAs are a large class of animal RNAs with regulatory potency. Nature. 2013;495(7441):333-8. doi: 10.1038/nature11928. PubMed PMID: 23446348.
102. Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA. 2014;20(11):1666-70. doi: 10.1261/rna.043687.113. PubMed PMID: 25234927; PMCID: PMC4201819.
103. Caillet-Boudin ML, Buee L, Sergeant N, Lefebvre B. Regulation of human MAPT gene expression. Mol Neurodegener. 2015;10:28. doi: 10.1186/s13024-015-0025-8. PubMed PMID: 26170022; PMCID: PMC4499907.
104. Cleary JD, Ranum LP. Repeat-associated non-ATG (RAN) translation in neurological disease. Hum Mol Genet. 2013;22(R1):R45-51. doi: 10.1093/hmg/ddt371. PubMed PMID: 23918658; PMCID: PMC3782068.
105. Suzuki T, Nagao A, Suzuki T. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 2011;45:299-329. doi: 10.1146/annurev-genet-110410-132531. PubMed PMID: 21910628.

106. Abe N, Hiroshima M, Maruyama H, Nakashima Y, Nakano Y, Matsuda A, Sako Y, Ito Y, Abe H. Rolling circle amplification in a prokaryotic translation system using small circular RNA. Angew Chem Int Ed Engl. 2013;52(27):7004-8. doi: 10.1002/anie.201302044. PubMed PMID: 23716491.
107. Abe N, Matsumoto K, Nishihara M, Nakano Y, Shibata A, Maruyama H, Shuto S, Matsuda A, Yoshida M, Ito Y, Abe H. Rolling Circle Translation of Circular RNA in Living Human Cells. Sci Rep. 2015;5:16435. Epub 2015/11/11. doi: 10.1038/srep16435. PubMed PMID: 26553571; PMCID: 4639774.
108. Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, Jin Y, Chen LL, Wang Y, Wong CC, Xiao X, Wang Z. Extensive translation of circular RNAs driven by N. Cell Res. 2017;27(5):626-41. Epub 2017/03/10. doi: 10.1038/cr.2017.31. PubMed PMID: 28281539; PMCID: PMC5520850.
109. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, Olshansky M, Rechavi G, Jantsch MF. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol. 2004;22(8):1001-5. doi: 10.1038/nbt996. PubMed PMID: 15258596.
110. Kelly S, Greenman C, Cook PR, Papantonis A. Exon Skipping Is Correlated with Exon Circularization. J Mol Biol. 2015;427(15):2414-7. doi: 10.1016/j.jmb.2015.02.018. PubMed PMID: 25728652.
111. Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, Jin Y, Yang Y, Chen LL, Wang Y, Wong CC, Xiao X, Wang Z. Extensive translation of circular RNAs driven by N(6)-methyladenosine. Cell Res. 2017;27(5):626-41. doi: 10.1038/cr.2017.31. PubMed PMID: 28281539; PMCID: PMC5520850.
112. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. PLOS Genetics. 2013;9(9):e1003777. Epub 2013/09/17. doi: 10.1371/journal.pgen.1003777. PubMed PMID: 24039610; PMCID: 3764148.
113. Stoss O, Stoilov P, Hartmann AM, Nayler O, Stamm S. The in vivo minigene approach to analyze tissue-specific splicing. Brain Research Protocols. 1999;4:383-94.
114. Mardon HJ, Sebastio G, Baralle FE. A role for exon sequences in alternative splicing of the human fibronectin gene. Nucl Acids Res. 1987;15:7725-33.
115. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol. 2010;653:249-57. doi: 10.1007/978-1-60761-759-4_15. PubMed PMID: 20721748.
116. Cooper TA. Use of minigene systems to dissect alternative splicing elements. Methods. 2005;37(4):331-40. PubMed PMID: 16314262.
117. Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet. 2005;42(10):737-48. doi: 10.1136/jmg.2004.029538. PubMed PMID: 16199547; PMCID: PMC1735933.
118. Percifield R, Murphy D, Stoilov P. Medium throughput analysis of alternative splicing by fluorescently labeled RT-PCR. Methods Mol Biol. 2014;1126:299-313. doi: 10.1007/978-1-62703-980-2_22. PubMed PMID: 24549673.
119. Stoilov P, Lin CH, Damoiseaux R, Nikolic J, Black DL. A high-throughput screening strategy identifies cardiotonic steroids as alternative splicing modulators. Proc Natl Acad Sci U S A. 2008;105(32):11218-23. PubMed PMID: 18678901.

120. Shen M, Bellaousov S, Hiller M, de La Grange P, Creamer TP, Malina O, Sperling R, Mathews DH, Stoilov P, Stamm S. Pyrvinium pamoate changes alternative splicing of the serotonin receptor 2C by influencing its RNA structure. Nucleic acids research. 2013;41(6):3819-32. Epub 2013/02/09. doi: 10.1093/nar/gkt063. PubMed PMID: 23393189; PMCID: 3616728.
121. Noto JJ, Schmidt CA, Matera AG. Engineering and expressing circular RNAs via tRNA splicing. RNA Biol. 2017:1-7. doi: 10.1080/15476286.2017.1317911. PubMed PMID: 28402213.
122. Schmidt CA, Noto JJ, Filonov GS, Matera AG. A Method for Expressing and Imaging Abundant, Stable, Circular RNAs In Vivo Using tRNA Splicing. Methods Enzymol. 2016;572:215-36. doi: 10.1016/bs.mie.2016.02.018. PubMed PMID: 27241756.
123. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Karolchik D, Hinrichs AS, Haeussler M, Guruvadoo L, Navarro Gonzalez J, Gibson D, Fiddes IT, Eisenhart C, Diekhans M, Clawson H, Barber GP, Armstrong J, Haussler D, Kuhn RM, Kent WJ. The UCSC Genome Browser database: 2018 update. Nucleic Acids Res. 2018;46(D1):D762-D9. doi: 10.1093/nar/gkx1020. PubMed PMID: 29106570; PMCID: PMC5753355.
124. Grozdanov PN, MacDonald CC. Generation of plasmid vectors expressing FLAG-tagged proteins under the regulation of human elongation factor-1alpha promoter using Gibson assembly. J Vis Exp. 2015(96). doi: 10.3791/52235. PubMed PMID: 25742071; PMCID: PMC4354628.
125. Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. Nucleic Acids Res. 2004;32(3):1197-207. doi: 10.1093/nar/gkh271. PubMed PMID: 14973201; PMCID: PMC373405.
126. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343-5. doi: 10.1038/nmeth.1318. PubMed PMID: 19363495.
127. Yang Y, Wang Z. Constructing GFP-Based Reporter to Study Back Splicing and Translation of Circular RNA. Methods Mol Biol. 2018;1724:107-18. doi: 10.1007/978-1-4939-7562-4_9. PubMed PMID: 29322444.
128. Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, Luo Y, Lyu D, Li Y, Shi G, Liang L, Gu J, He X, Huang S. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. Nat Commun. 2016;7:11215. doi: 10.1038/ncomms11215. PubMed PMID: 27050392; PMCID: PMC4823868.
129. Jia W, Xu B, Wu J. Circular RNA expression profiles of mouse ovaries during postnatal development and the function of circular RNA epidermal growth factor receptor in granulosa cells. Metabolism. 2018;85:192-204. doi: 10.1016/j.metabol.2018.04.002. PubMed PMID: 29634953.
130. Liang D, Tatomer DC, Luo Z, Wu H, Yang L, Chen LL, Cherry S, Wilusz JE. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. Mol Cell. 2017;68(5):940-54 e3. doi: 10.1016/j.molcel.2017.10.034. PubMed PMID: 29174924; PMCID: PMC5728686.
131. Li X, Liu CX, Xue W, Zhang Y, Jiang S, Yin QF, Wei J, Yao RW, Yang L, Chen LL. Coordinated circRNA Biogenesis and Function with NF90/NF110 in Viral Infection.

Mol Cell. 2017;67(2):214-27 e7. doi: 10.1016/j.molcel.2017.05.023. PubMed PMID: 28625552.
132. Zhang Y, Xue W, Li X, Zhang J, Chen S, Zhang JL, Yang L, Chen LL. The Biogenesis of Nascent Circular RNAs. Cell Rep. 2016;15(3):611-24. doi: 10.1016/j.celrep.2016.03.058. PubMed PMID: 27068474.
133. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature. 2009;457(7232):1028-32. Epub 2009/01/27. doi: 10.1038/nature07759. PubMed PMID: 19169241; PMCID: 2719882.
134. Gao QS, Memmott J, Lafyatis R, Stamm S, Screaton G, Andreadis A. Complex regulation of tau exon 10, whose missplicing causes frontotemporal dementia. J Neurochem. 2000;74(2):490-500. PubMed PMID: 10646499.
135. Gaisler-Salomon I, Kravitz E, Feiler Y, Safran M, Biegon A, Amariglio N, Rechavi G. Hippocampus-specific deficiency in RNA editing of GluA2 in Alzheimer's disease. Neurobiol Aging. 2014;35(8):1785-91. Epub 2014/03/01. doi: 10.1016/j.neurobiolaging.2014.02.018. PubMed PMID: 24679603.
136. Cleveland DW, Hwo SY, Kirschner MW. Purification of tau, a microtubule-associated protein that induces assembly of microtubules from purified tubulin. J Mol Biol. 1977;116(2):207-25. doi: 10.1016/0022-2836(77)90213-3. PubMed PMID: 599557.
137. Licht K, Hartl M, Amman F, Anrather D, Janisiw MP, Jantsch MF. Inosine induces context-dependent recoding and translational stalling. Nucleic Acids Res. 2019;47(1):3-14. doi: 10.1093/nar/gky1163. PubMed PMID: 30462291; PMCID: PMC6326813.
138. Ghetti B, Oblak AL, Boeve BF, Johnson KA, Dickerson BC, Goedert M. Invited review: Frontotemporal dementia caused by microtubule-associated protein tau gene (MAPT) mutations: a chameleon for neuropathology and neuroimaging. Neuropathol Appl Neurobiol. 2015;41(1):24-46. doi: 10.1111/nan.12213. PubMed PMID: 25556536; PMCID: PMC4329416.
139. Kim U, Wang Y, Sanford T, Zeng Y, Nishikura K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. Proc Natl Acad Sci U S A. 1994;91(24):11457-61. doi: 10.1073/pnas.91.24.11457. PubMed PMID: 7972084; PMCID: PMC45250.
140. Patterson JB, Samuel CE. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. Mol Cell Biol. 1995;15(10):5376-88. doi: 10.1128/mcb.15.10.5376. PubMed PMID: 7565688; PMCID: PMC230787.
141. Cho DS, Yang W, Lee JT, Shiekhattar R, Murray JM, Nishikura K. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. J Biol Chem. 2003;278(19):17093-102. Epub 2003/03/04. doi: 10.1074/jbc.M213127200. PubMed PMID: 12618436.
142. Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. Genome Med. 2013;5(11):105. Epub 2013/11/29. doi: 10.1186/gm508. PubMed PMID: 24289319; PMCID: PMC3979043.
143. Singh M. Dysregulated A to I RNA editing and non-coding RNAs in neurodegeneration. Front Genet. 2012;3:326. Epub 2013/01/22. doi: 10.3389/fgene.2012.00326. PubMed PMID: 23346095; PMCID: PMC3551214.

144. Caillet-Boudin ML, Buée L, Sergeant N, Lefebvre B. Regulation of human MAPT gene expression. Mol Neurodegener. 2015;10:28. Epub 2015/07/14. doi: 10.1186/s13024-015-0025-8. PubMed PMID: 26170022; PMCID: PMC4499907.
145. Khermesh K, D'Erchia AM, Barak M, Annese A, Wachtel C, Levanon EY, Picardi E, Eisenberg E. Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. RNA. 2016;22(2):290-302. Epub 2015/12/11. doi: 10.1261/rna.054627.115. PubMed PMID: 26655226; PMCID: PMC4712678.

VITA

Justin Ralph Welden

Education

University of Kentucky College of Medicine, Lexington, Kentucky 8/2017-01/2021

PhD in Molecular and Cellular Biochemistry

University of Kentucky College of Arts and Science, Lexington, KY 8/2012-5/2016

Bachelor of Arts in Biology, Minor in Psychology

Henderson Community College, Henderson, Kentucky 8/2009-12/2011

Associate of Arts

Research Experience

University of Kentucky, Biochemistry Department Lexington, KY 5/2016-current

- Mentor: Stefan Stamm, PhD
- Project: Investigating the role of *MAPT* circular RNAs in Tauopathies.

University of Kentucky, Microbiology, Immunology & Molecular Genetics Lexington, KY 1/2016-5/2016

- Mentor: Beth Garvy, PhD
- Project: Investigating the mechanism behind the trophic forms of a fungal pathogen, Pneumocystsis Carinii, how they are recognized and cleared from the lungs

Awards and Honors

Max Steckler Award, $8,350 2019-2020

ASBMB Travel Award, $1,000 2019

College of Medicine Excellence in Graduate Research Fellowship, $12,500 2018-2019

Publications

1. **Welden JR**, Pawluchin A, van Doorn J, Stamm S. Use of Alu Element Containing Minigenes to Analyze Circular RNAs. J Vis Exp. 2020(157). Epub 2020/03/10. doi: 10.3791/59760. PubMed PMID: 32225139.
2. **Welden JR**, Stamm S. Pre-mRNA structures forming circular RNAs. Biochim Biophys Acta Gene Regul Mech. 2019:194410. Epub 2019/08/20. doi: 10.1016/j.bbagrm.2019.194410. PubMed PMID: 31421281.

3. **Welden JR**, van Doorn J, Nelson PT, Stamm S. The human MAPT locus generates circular RNAs. Biochim Biophys Acta. 2018;1864(9 Pt B):2753-60. Epub 2018/05/03. doi: 10.1016/j.bbadis.2018.04.023. PubMed PMID: 29729314.
4. **Welden JR**, Zhang Z, Duncan MJ, Falaleeva M, Wells T, Stamm S. The posterior pituitary expresses the serotonin receptor 2C. Neurosci Lett. 2018;684:132-9. Epub 2018/06/30. doi: 10.1016/j.neulet.2018.06.051. PubMed PMID: 29969651.
5. Falaleeva M, **Welden JR**, Duncan MJ, Stamm S. C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. Bioessays. 2017;39(6). Epub 2017/05/15. doi: 10.1002/bies.201600264. PubMed PMID: 28505386; PMCID: PMC5586538.

Provisional Patent

siRNAs targeting circular Tau RNA backsplice junction. "Combating Alzheimer's Disease using oligonucleotides against circular RNAs". (63/137, 405)

Meetings Attended

- RNA Conference, Kraków, Poland 2019 (Talk & Poster)
- Foundation for Prader-Willi Syndrome, New Orleans, Louisiana 2019 (Poster)
- ASBMB Annual Conference, Orlando, Florida 2019 (Poster)
- Foundation for Prader-Willi Syndrome, Las Vegas, Nevada 2018 (Poster)
- RNA Conference, UC Berkeley, California 2018 (Poster)
- Foundation for Prader-Willi Syndrome, Indianapolis, Indiana 2017 (Poster)
- RNA Conference, Czech Republic, Prague 2017 (Poster)