2021

RNA SPLICING IN NEURON PHYSIOLOGY AND NEURODEGENERATIVE DISEASES

MD Faruk Hossain
Saint John's University, Jamaica New York

RNA SPLICING IN NEURON PHYSIOLOGY AND NEURODEGENERATIVE DISEASES

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

to the faculty of the

DEPARTMENT OF BIOLOGICAL SCIENCES

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY
New York
by
Md Faruk Hossain

Date Submitted 3/24/2021 Date Approved 4/29/2021

Md Faruk Hossain	Dr. Matteo Ruggiu

ABSTRACT

RNA SPLICING IN NEURON PHYSIOLOGY AND NEURODEGENERATIVE DISEASES

Md Faruk Hossain

Gene expression is regulated at multiple levels, including transcription, RNA editing, pre-mRNA splicing, mRNA export, translation, and posttranslational modifications. Alternative splicing is a process by which exons can be included or excluded, giving rise to multiple mRNA isoforms from the same transcript. Alternative splicing is an important mechanism in developmental, tissue- and cell-specific control of gene expression, and it is key for expanding proteomic diversity and complexity from a limited number of genes. Moreover, more than 95% of multiexon genes undergo alternative splicing in humans, and about half of all disease-causing point mutations in humans affect pre-mRNA splicing, including neurological disorders and cancer. The central nervous system comprises the tissues and cells with the highest rate of alternative splicing in the body, and RNA-binding proteins play a major functional role in neurons. However, the regulatory mechanisms of splicing are still poorly understood. This dissertation specifically aims to advance the understanding of regulatory mechanism of pre-mRNA splicing. To this end, we collaboratively performed two projects.

In the first project, we investigated how *NOVA*, a neuron-specific splicing factor, regulates nerve cell-specific alternative splicing of Z^+ Agrin — a molecule that is the master architect of nerve-muscle synapses at the neuromuscular junction (NMJ). We cloned the *Ciona* ortholog of *NOVA*, which is present as a single copy gene in tunicates, and that of Agrin, and *dissected* the regulatory mechanism of alternative splicing of Z^+

Agrin by Nova. Moreover, we characterized their function and expression pattern during larval development, which we will discuss in detail in Chapter 2 of this dissertation.

The second project was a case study where we investigated how mutations in the *SLC25A10* gene cause epileptic encephalopathy by disrupting pre-mRNA splicing. *SLC25A10* codes for a solute carrier protein and is a part of complex I in mitochondria. The patient inherited 3 mutations: 1 from the mother and 2 from the father. The maternal-derived mutation introduces a stop codon in exon 3. Mutations from the paternal allele are located in exon 9 and intron 10. Although the exonic mutation is a synonymous mutation, the patient had very low levels of SLC25A10 mRNA and lacked protein at detectable levels. Using minigene splicing assay we investigated the molecular mechanism underlying disease pathology in the patient. In Chapter 3 of this dissertation, we will discuss how paternal-derived mutations lead to aberrant splicing.

ACKNOWLEDGEMENTS

All praise is due to Allah (God). I express my utmost gratitude to Allah who has blessed me with opportunities and knowledge. I am grateful to my wonderful family members: my parents, my siblings, my lovely wife Sadia Afrin and my in-laws for their continuous support.

A special thanks to my mentor, Dr. Matteo Ruggiu, who has been patiently guiding me since the day I began working in his lab as a PhD student. I am lucky that I got direct supervision and guidance from Dr. Ruggiu on everything. Without his support it was not possible to complete this dissertation as well. I thank him for giving me the opportunity to learn science with the state-of-the art equipment in his lab. In every single tough time, Dr. Ruggiu supported me with his expertise and gave me the necessary freedom that I needed to move on.

I would like to sincerely thank our research collaborators Dr. Alberto Stolfi, Dr. Lionel Christiaen, Burcu Vitrinel who greatly contributed to the *Ciona* project. I would like to thank Dr. De Grassi for collaborating *SLC25A10* project with us. My sincere appreciation is to my committee members Dr. Ales Vancura, Dr. Ivana Vancurova, Dr. Javier Juarez and Dr. Yong Yu for their support.

I praise the help I got from the laboratory members throughout the years. I thank Ruizhi Wang, Jovan Mirkovic, and Samuel Sabzanov for helping me with many of the experiments, and Willy Munyao with the writing of this thesis. It would never have been so enjoyable without you guys. I also sincerely thank other faculty members, office secretaries, and friends in the biology department.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The focus towards the understanding of gene expression regulation largely shifted after the completion of human genome sequencing projects. Gene regulation is a set of complex biological processes that help genetic information to flow from DNA to RNA to proteins, which is known as the central dogma of biology. Gene regulation involves several processes involving chromatin remodeling, transcription, post-transcriptional regulation (pre-mRNA splicing/alternative splicing), mRNA export and editing, translation, and posttranslational modifications. Alternative splicing (AS) is one of the most critical steps in gene regulation, which results in multiple proteins with often distinct functions from a single gene. Compared with other tissue types, AS is highly abundant in the brain. Emerging evidence has shown that the disruption of AS plays a significant role in many human diseases, including cancer, diabetes, and neurodegenerative diseases. In this dissertation, we show the importance of pre-mRNA splicing in neuron physiology and neurodegenerative disorders. This chapter gives a brief introduction and background information related to splicing and its role in neurological disorders.

1.1 Gene splicing and alternative splicing

Gene expression is regulated at multiple levels, including transcription, RNA editing, pre-mRNA splicing, mRNA export, translation, and posttranslational modifications. After a gene is transcribed into pre-mRNAs, introns are spliced out and

exons are joined together to make mature mRNA transcript(s) in a process called pre-mRNA splicing. Pre-mRNA splicing was first detected in adenovirus in the 1970s (Chow *el al.,* 1977; Berget *el al.,* 1977) and later splicing was discovered in all eukaryotes, including yeast, plants, and animals. While constitutive pre-mRNA splicing produces a single transcript from a gene, AS on the other hand produces multiple mRNA from the same pre-mRNA transcript by specifically including or excluding alternative exon(s). Approximately 95% multiexonic human genes undergo AS (Gilbert, 1978; Pan *et al.,* 2008). In fact, AS mechanism explains the tremendous number of protein variants from a limited repertoire of ~25,000 genes in human and is a major source of proteomic diversity and complexity (Pan *et al.,* 2008). One of the noticeable examples of proteomic diversity by AS is the *Drosophila Dscam* gene. Dscam is a member of the immunoglobulin superfamily required for axon guidance (Schmucker *et al.,* 2000) and through 95 alternative exons has the potential to generate 38,016 distinct axon guidance receptors. That is a staggering number of different isoforms originating from a single gene that exceeds the total number of genes present in the *Drosophila* genome (Misra *et al.,* 2002).

1.2 The classification of alternative splicing

In eukaryotes, several different functional protein isoforms can be produced by AS. The regulation of AS depends on a combination of tissue-specific and universally expressed trans-acting RNA-binding factors that interact with cis-acting elements on the pre-mRNA molecule. Precise AS regulation is biologically and physiologically important for normal development. The completion of genome sequencing has provided important insights into the various modalities of pre-mRNA splicing. There are seven major types

of AS described so far. About 40% of the AS type is the skipped exon (SE), also known as cassette exon (Sugnet *et al.,* 2004; Kim *et al.,* 2007). Cassette exons can be included or excluded in the mature mRNA. Multiple cassette exons give rise to another splicing pattern known as mutually exclusive cassette exons. In this case the final processed mRNAs always include only one of the available alternative exon cassettes. Another two types of splicing modalities are the lengthening or shortening of an exon with the differential use of alternative 5' and alternative 3' splice sites. The regulatory potential of AS can be expanded by two other modalities where alternative polyadenylation sites and alternative promoter sites can be utilized. In another mode of AS, the intron can be included in the mature mRNA and the process is called intron retention. Moreover, AS can be combinatorial – meaning that multiple modes of the above-mentioned AS events can be observed in the mature mRNA (Black *et al.,* 2003).

1.3 The splicing machinery and regulatory network of alternative splicing

Constitutive splicing and AS are strictly controlled by complex interactions of many cis-regulatory elements and trans-acting RNA-binding proteins (RBPs). Splicing is performed by a large and dynamic complex called the spliceosome that is composed of small ribonucleoprotein particles (snRNPs). The splicesome core is composed of four small ribonucleoprotein particles (snRNP U1, U2, U4/U6 and U5) and numerous auxiliary proteins that are assembled together to execute a series of steps involved in splicing, including looping the intron, excising the intron, and joining the exons. Mass spectrometric analysis revealed that the spliceosome is one of the largest protein complexes in mammals and it is assembled by more than 300 proteins (Burge *et al.,*

1999; Zhou *et al.,* 2002; Barbosa-Morais *et al.,* 2006;). The fundamental splicing signals are universal; however, the level of conservation varies in different exons/introns and organisms. Each intron is almost invariantly marked by a GU dinucleotide at the 5' end (5' splice site, or 5'ss), and an AG dinucleotide at the 3'end (3' splice site or 3'ss). Moreover, a branch point sequence (BPS) upstream of the 3'ss and a polypyrimidine tract between the BPS and the 3'ss are other important signals. Normally, constitutive exons tend to have strong 5' and 3' ss, while alternative exons have relatively weaker 5' and 3' ss, therefore other RNA-binding proteins play a critical role on splice sites recognition (Stamm *et al.,* 1994).

RBPs/splicing factors act together with the spliceosome to from extensive protein-protein and protein-RNA complexes to control the expression of functionally distinct isoforms in specific tissues and developmental stages by promoting or blocking the inclusion of alternative exon(s) in the final processed mRNAs. While the spliceosome acts on the primary splicing signals (exon-intron junction, branchpoint A site in the intron, and intron-exon junction) to perform splicing, RBPs act on cis-acting specific sequences either in exons or in introns to execute AS. Cis-acting sequences on pre-mRNA are generally divided into four categories: exonic splicing enhancers and silencers (ESEs and ESSs), and intronic splicing enhancers and silencers (ISEs and ISSs) (Cartegni *et al.,* 2002). Interactions between the cis-acting regulatory elements and trans-acting splicing factors determine splicing outcome. RBPs are key regulators of AS, and their expression is often developmentally regulated and also tissue-specific. RBPS are of two types: AS activators and AS repressors. For instance, SR proteins promote inclusion of alternative exon(s) by binding to ESEs (Shen *et al.,* 2004) and heterogeneous

ribonucleoproteins (hnRNPs) block inclusion of alternative exon(s) when bound to ESSs (Del Gatto-Konczak et al, 1999). There are about 1,000 RBPs in the human genome with approximately 40 different types of RNA-binding motifs, including RNA-recognition motifs, K-homology (KH) domains etc (Lunde *et al.,* 2007). Expression of RBPs is often cell- or tissue-type specific. They can be expressed in different combinations to increase proteomic diversity, and they often have multiple target genes. Subsequently, aberrant expression or disrupting the function of a single RBP often affects posttranscriptional regulation of numerous pre-mRNA transcripts, a phenomenon frequently reported in human diseases associated with RBPs.

1.4 The role of alternative splicing in neuron physiology

Previous studies clearly showed that there is extensive AS regulation across brain regions and across developmental stages (Dillman *et al.,* 2013). Moreover, compared to other cell types neurons have the highest number of alternative exons, thus producing a greatest diversity of protein isoforms compared to other tissues (Porter *et al.,* 2018). For example, neurons generate neurotransmitter receptors with different specificities and coordinates the activity of protein networks at the synapse (Ule *et al.,* 2005). The voltage-gated Ca^{2+} channels (VGCCs) in the presynaptic terminal are key determinants of cell- and synapse-specific neurotransmitter release properties, and the genes encoding VGCC subunits have the potential to generate thousands of splice variants (Lipscombe *et al.,* 2013). For instance, AS of two mutually exclusive alternative exons (37a and 37b) in the Cav2.1 (the so-called P/Q-type Calcium channel) α1 subunit demonstrate acute shift of neuronal network activity, while one isoform promotes synaptic depression, the other drives synaptic facilitation (Thalhammer *et al.,* 2017).

The NMJ, possibly the best-understood mammalian synapse, is developed, formed, and maintained by a large number of molecules that enable precise signal transmission from spinal motor neurons to skeletal muscle. According to RefSeq gene annotation database which adopts only experimentally-proven splicing isoforms, 8 out of 16 genes (*ACHE, AGRN, CHAT, CHRNA1, CHRNB1, CHRND, CHRNE, CHRNG, COLQ, DOK7, LAMB2, LRP4, MUSK, RAPSN, SCN4A*, and SYT2) expressed at the NMJ are alternatively spliced (O'Leary *et al.,* 2016). According to the most extensive AceView gene annotation database (Thierry-Mieg and Thierry-Mieg, 2006), 13 of the 16 NMJ genes are alternatively spliced, whereas *CHRNE*, *CHRNG*, and *SCN4A* are not. In this dissertation, we will discuss the role of *AGRN* in NMJ development, formation, and maintenance and our investigation of regulatory mechanism of AS of *AGRN* by *NOVA* (see chapter 2 for detail).

1.5 Aberrant splicing in neurodegenerative disorders

The central nervous system comprises the tissues and cells with the highest rate of AS in the body (Xu *et al.,* 2002), and RBPs play a major functional role in neurons (Lenzken *et al.,* 2014), highlighting the critical importance of RBPs and splicing in human biology and disease pathogenesis. Moreover, up to 50% of all disease-causing mutations affect pre-mRNA splicing, including many cancers and neurodegenerative disorders (Teraoka *et al.,* 1999; Ars *et al.,* 2003) including, Alzheimer's disease (AD), Parkinson's disease (PD), Huntington's disease (HD), Spinal Muscular Atrophy (SMA), and Congenital Myasthenic Syndrome (CMS). Pre-mRNA splicing is disrupted in a large number of genes in AD patients (*PSEN1, PSEN2, GRN, APP, APOE, BACE, PIN1*, and *MAPT*), in PD patients (*PARK2, SNCA, SRRM2*), in HD patients (*BDNF*) (reviewed in

Mills and Janitz, 2012). To date, 30 genes (reviewed in McMacken *et al.,* 2017) have been reported to be CMS-related genes, many of those genes display abnormal AS including *CHRNE, DOK7, CHRNA1, RAPSAN, and COLQ* genes (reviewed in Rahman *et al.,* 2015). AS dysregulation has been recently reported in autism spectrum disorders (Irimia *et al.,* 2014; Parikshak *et al.,* 2016; Quesnel-Vallieres *et al.,* 2016; Xiong *et al.,* 2015) and schizophrenia (Cai *et al.,* 2020).

Pre-mRNA splicing patterns can be affected by mutations in splice sites, cis-regulatory elements, or trans-acting RBPs. An interesting example is the skipping of exon 7 of survival of motor neuron 2 (*SMN2*) due to a single nucleotide (nt) polymorphism (C to T) at position 6 that disrupts an ESE element recognized by SF2/ASF (Pellizzoni *et al.,* 1998; Cartegni and Krainer, 2002), and creates an ESS element recognized by hnRNP A1 (Kashima *et al.,* 2003). SMA is a devastating autosomal-recessive disorder associated with low expression levels of SMN1 protein due to homologous deletion or disruption of *SMN1*. *SMN2*, which is an almost identical copy of *SMN1* produces a truncated isoform lacking exon 7 which is unstable and nonfunctional (Khoo *et al.,* 2009; Lorson *et al.,* 1998; Lorson *et al.,* 1999; Burnett *et al.,* 2009). In Mattioli *et al.* (2020), 6 individuals have been reported with a *de novo* frameshift mutation in the RNA-binding protein *NOVA2* that results in a common C-terminal extension. All 6 of the individuals are affected by a severe form of neurodevelopment disorders. Using zebrafish ortholog of *NOVA2*, they have shown that downregulation of *NOVA2* affects neurite outgrowth. Moreover, downregulation of *NOVA2* alters the splicing of 41 genes in human neural cells. In chapter 3 of this dissertation, we will discuss a case study where we investigated aberrant splicing of the

SLC25A10 gene and explained disease pathology in a patient with epileptic encephalopathy.

CHAPTER 2

A CONSERVED NOVA-DEPENDENT ALTERNATIVE SPLICING PROGRAM CONTROLS NEUROMUSCULAR JUNCTION FUNCTION IN THE TUNICATE *CIONA ROBUSTA*

2.1 ABSTRACT

Tunicates are marine invertebrates and are the closest living relatives to vertebrates. The swimming larva of the tunicate *Ciona robusta* is an emerging animal model to study developmental and evolutionary biology. Its nervous system is comprised of a mere 177 neurons distributed rostro-caudally in a brain vesicle, a motor ganglion, and a nerve cord and its larval connectome has been completely mapped. Due to its small size, cellular simplicity, rapid development, and compact genome that has not undergone the duplications seen in vertebrates, *Ciona* is particularly amenable to molecular perturbation and imaging. Here we show that *Ciona* can be a powerful model organism to study NMJ biology and neurodegenerative and neuromuscular disorders, including congenital myasthenic syndrome (CMS).

NOVA1 and NOVA2 are neuron-specific AS factors and are target antigens in patients with an autoimmune neurodegenerative disorder. One of the targets of NOVA is a neuron-specific splice form of the ubiquitously expressed gene *AGRIN*. Termed Z^+ AGRIN, this splice variant activates the MuSK signaling pathway by interacting with the transmembrane receptor LRP4, thus promoting clustering of acetylcholine receptors

(AChRs) at the postsynaptic terminal., Interestingly, *AGRIN* mutations that mimic Z^- AGRIN cause CMS.

We cloned the *Ciona* ortholog of *NOVA*, which is present as a single copy gene in tunicates, and that of Agrin, and characterized their function and expression pattern during larval development. We discovered that, as in vertebrates, *Ciona* Agrin (CiAgrin) also undergoes AS to generate the Z^+ isoforms in *Ciona*, indicating that the Nova-Agrin-Lrp4 pathway for AChR clustering at the NMJ is shared between tunicates and mammals. Nova harbors 3 KH-type RNA-binding domains and specifically recognizes YCAY clusters on pre-mRNA. *Ciona* Nova (CiNova) requires the first two KH domains to mediate Z exon inclusion, and it does so via a bipartite intronic splicing enhancer downstream of the Z exons.

We also determined that at least two consecutive YCAY repeats from any of the two clusters are needed to promote exon inclusion at Z site. Moreover, we discovered unique function of the N/C-terminus and the KH3 domain of CiNova. It appears that, in *Ciona*, the KH3 domain is a negative regulator of AS and both N- and C- terminus act together to inhibit the negative action of KH3. We discovered that CRISPR KO of CiNova and CiAgrin phenocopies the findings in Nova and Agrin KO mice by drastically reducing the number of AChRs clusters at the NMJ. We provide evidence that the neural-specific AS program of Z^+ Agrin is conserved between tunicates and mammals, showing for the first time that *Ciona* can be a powerful organism to study the biology of neurodegenerative and neuromuscular disorders.

2.2 BACKGROUND

2.2.1 The *NOVA* protein family and its role in brain physiology

In mammals, the NOVA family is comprised of two highly homologous proteins encoded by two genes (*NOVA1* and *NOVA2*), which were first discovered as antigens in Paraplastic Opsoclonus Myoclonus Ataxia (POMA)- a manifestation of abnormal motor control at the level of brainstem and spinal cord (Darnell *et al.,* 1996; Buckanovich *et al.,* 1996; Buckanovich *et al.,* 1997). NOVA are neuron-specific AS factors that regulate the AS of neuronal pre-mRNA transcripts (such as transcripts of the *AGRN* gene) in the central nervous system (Buckanovich *et al.,* 1996; Buckanovich *et al.,* 1997). Both NOVA1 and NOVA2 possess three K-homology (KH) RNA-binding domains (KH1, KH2 and KH3) and it appears that it is the third KH domain (KH3) that specifically binds to YCAY motifs on pre-mRNA transcripts to regulate inclusion or skipping of specific alternative exons (Buckanovich *et al.,* 1997; Jensen *et al.,* 2000). Consistently, *de novo* frameshift mutations that ablate this RNA-binding domain abolish the pre-mRNA-binding activity of NOVA (Mattioli *et al.,* 2020).

In the central nervous system, Nova1 and Nova2 are expressed reciprocally, with Nova2 being highly expressed in the cortex and hippocampus, while Nova1 is highly expressed in midbrain and spinal cord (Yang *et al.,* 1998; Saito *et al.,* 2016). Moreover, Nova1 is highly expressed in the ventral spinal cord while Nova2 is highly expressed in the dorsal spinal cord (Buckanovich *et al.,* 1997). Both Nova1 and Nova2 are also expressed in white adipocytes (Vernia *et al.,* 2016). The finding that there are about 700 identified Nova1/Nova2 alternate exon targets on neuron-specific pre-mRNA (Zhang *et al.,* 2010), is a strong pointer to the role of both Nova1 and Nova2 in the regulation of AS of neuron-specific RNA transcripts (Saito *et al.,* 2016). Nova1 and Nova2 possess nearly

identical RNA-binding domains, and both recognize the same YCAY sequence on pre-mRNA (Buckanovich *et al.,* 1997; Jensen *et al.,* 2000; Sugimoto *et al.,* 2012). These two splicing regulators can therefore interact with and regulate AS on the same pre-mRNA transcript (Saito *et al.,* 2016). About 80% of the binding sites of Nova2 on pre-mRNA are found on introns, implying a preferential binding of this splicing regulator to introns, while Nova1 targets are almost equally distributed in introns and exons (Saito *et al.,* 2016). This spatial distribution of the RNA-binding targets suggests a possible mechanistic difference in the AS events driven by these two Nova homologues (Saito *et al.,* 2016).

The distinct expression pattern of Nova1 and Nova2 denotes that each of these splicing regulators appears to have a unique role in different brain areas and different neuronal cell types (Saito *et al.,* 2016). In the brain cortex, where Nova2 is highly expressed, a total of 60 transcripts were significantly changed by Nova2 loss of function, while only 2 transcripts were altered in the brain cortex by Nova1 loss of function (Saito *et al.,* 2016). This scenario could be partially explained by the low expression levels of Nova1 in brain cortex. In the midbrain and hindbrain region of E18.5 mice, where Nova1 is highly expressed, 119 Nova1-dependent AS events were changed by loss of gene function (Saito *et al.,* 2016). Moreover, in these brain regions, only one AS event (Robo2 exon 6b), common to both Nova1 and Nova2 was slightly altered (Saito *et al.,* 2016).

Nova proteins-mediated splicing regulation plays a critical role in the development of the central nervous system (Leggere *et al.,* 2016; Saito *et al.,* 2016). The splicing regulatory role of Nova proteins has been implicated in neuronal migration in both cortical neurons and Purkinje neurons (Yano *et al.,* 2010). Nova2 is also important

for the induction of long-term potentiation of slow inhibitory post synaptic current in hippocampus neurons (Huang *et al.,* 2005). In addition to the above-mentioned Nova-mediated roles, four research articles (Ruggiu *et al.,* 2009; Leggere *et al.,* 2016; Saito *et al.,* 2016; and Saito *et al.,* 2019) expand the knowledgebase on the functional role of Nova family proteins in the development of the central nervous system. The AS outcome of Nova proteins is based on their binding position on the target pre-mRNA (Dredge *et al.,* 2005; Allen *et al.,* 2010). Nova proteins binding to their binding sites upstream or within an alternative exon represses exon inclusion (Dredge *et al.,* 2005). In contrast, Nova-mediated effect on alternative exon inclusion enhancement can be attributed to binding of the Nova protein downstream of an alternative exon (Dredge *et al.,* 2005). In addition, the Nova-mediated splicing effect on alternate exon inclusion may be due to its blocking of the binding of both essential splicing factors like U1 snRNP, and non-essential splicing factors like SR proteins, onto the pre-mRNA transcript (Dredge *et al.,* 2005). Nova2 expression is down regulated by the RE-1-Silencing Transcription factor (REST), a transcription repressor that is highly expressed in non-neuronal cells and early embryonic neuronal cells but is silenced in mature differentiated neuronal cells (Mikulak *et al.,* 2012). On the contrary, Nova1 expression does not appear to be regulated by REST in both neuronal and non-neuronal cells (Mikulak *et al.,* 2012).

2.2.2 Functions of Nova1

Nova1 is highly expressed in midbrain and ventral spinal cord (Yang *et al.,* 1998; Saito *et al.,* 2016) but is also expressed in non-neuronal cells, with high expression in pancreatic beta cells (at levels comparable to brain) and white adipose tissue, and low expression in the cervix, colon, muscle cells, liver, spleen, and lungs (Eizirik *et al.,* 2012;

Villate *et al.,* 2014; Meldolesi *et al.,* 2020; Vernia *et al.,* 2016). Although Nova1 null ($Nova1^{-/-}$) mice in Jensen *et al*. (2000) were indistinguishable from their littermates, they died postnatally due to a progressive motor defect caused by increased apoptotic death of spinal and brainstem neurons. In Ruggiu *et al.* (2009), $Nova1^{-/-}$ mice lacked defects in Agrin Z^+ AS in brain and spinal cord, and in AChR clustering or neuromuscular innervation at the NMJ. $Nova1^{-/-}$ mice in Saito *et al.,* 2016 did not portray any defects in neuronal development. They had normal parameters in dorsal interneuron development and differentiation, axon outgrowth and corpus callosum (CC) formation. In addition, these mutant mice lacked the defects displayed by $Nova2^{-/-}$ mice in both ventral diaphragmatic and auditory efferent innervation (Saito *et al.,* 2016; Leggere *et al.,* 2016).

Nova1 does not appear to have a regulatory role in the axon guidance process in the brain cortex (Saito *et al.,* 2016). Specifically, Nova1 does not appear to regulate the AS of key exons in axon guidance-related genes, including: *Dcc* exon 17, *Slit2* exon 28b, *Robo2* exons 6b and 21, *Epha5* exon 7, *Arhgef12* exon 4, *Ppp3cb* exon 10b, *Neo1* exon 26, and *Rock1* exon 27b (Saito *et al.,* 2016). The missing regulatory role of Nova1 in the brain cortex could however be due to the low expression levels of Nova1 in the brain cortex. It would be interesting to observe the effects of Nova1 overexpression in the brain cortex.

Nova1 is an enhancer of exon inclusion during AS. For example, Nova1 enhances the inclusion of exon 9 during the AS of the gamma-aminobutyric acid A receptor, gamma 2 subunit (Gabrg2) pre-mRNA in the brain and human pancreatic islets (Eizirik *et al.,* 2012; Dredge and Darnell, 2003; Jensen *et al.,* 2000a). Nova1 also enhances the

inclusion of exon 3A (E3A) during AS of the of glycine receptor a2 (GlyRα2) pre-mRNA (Jensen *et al.,* 2000a; Polydorides *et al.,* 2000).

In pancreatic cells, Nova1 has a functional role in both transcription and AS regulation, but its role is more pronounced in splicing regulation (Villate *et al.,* 2014). In these cells, Nova1 is a master regulator of the AS of 4961 isoforms, involved in a wide array of cellular functions, including: apoptosis, insulin receptor signaling, exocytosis, transcription regulation, and cell signaling (Villate *et al.,* 2014). Examples of Nova1-regulated genes in pancreatic cells include apoptosis genes (*Casp3*, *Apaf1*), insulin receptor signaling genes (*INSR*, *FoxO1*), calcium signaling genes (*Cacna1b*, *Cacna1c*, *Cacna1d*), exocytosis genes (*Apba1*, *Cadps*, *Cdc42*, *Gnai3*, *Snap25*), transcription regulation genes (*Pax6*, *FoxO1*, *FoxO3*) and some other cell signaling genes (Villate *et al.,* 2014).

The high expression of Nova1 in both the brain and pancreatic islets denotes a common splicing regulatory role and a common mechanism of action in these two organs. For example, the exocytosis process of neurotransmitter release in the neurons is similar to that of insulin release from pancreatic beta cells (Juan-Mateu *et al.,* 2017). In fact, 80% of Nova1-regulated genes are highly expressed in both the brain and pancreatic beta cells (Villate *et al.,* 2014). Examples of these genes are: *Gabrg2*, *Neuroligin* and *Neurexin* family members, inhibitory synapse-associated neuroligin and neurexin binding partners (Villate *et al.,* 2014; Eizirik *et al.,* 2012). The mechanism of Nova1 splicing regulation appears to be similar in both the brain and pancreatic beta cells (Villate *et al.,* 2014).

In pancreatic beta cells, Nova1 also regulates the *in vitro* splicing of several calcium channels proteins, including the alpha-1b, alpha-1c, alpha-1d subunits of the voltage-dependent N-type calcium channel (Cacna1b, Cacna1c, Cacna1d) mRNA (Villate *et al.,* 2014). On the contrary, the splicing regulation of calcium channel proteins in the brain is regulated by Nova2 (Allen *et al.,* 2010). In the brain, Nova2 enhances the inclusion of exon 24a in Cacna1a (CaV2.1) and Cacna1b (CaV2.2), but also represses the inclusion of exon 31a in both CaV2.1 and CaV2.2 mRNA (Allen *et al.,* 2010).

Nova1 loss of function in pancreatic cells decreases voltage-dependent calcium current due to splicing defects in the calcium channels transcripts and thus downregulates calcium signaling-mediated insulin exocytosis (Villate *et al.,* 2014). In addition, Nova1 splicing action on insulin secretion genes (phospholipase *PLCβ1* and the vesicle fusion protein *SNAP25*) also regulates insulin secretion (Villate *et al.,* 2014). Specifically, Nova1-mediated effect on insulin secretion entails the AS of the exon b-containing isoform of PLCβ1 (PLCβ1b) and enhances the inclusion of alternative exon 5b of SNAP25 pre-mRNA during AS (Villate *et al.,* 2014).

2.2.3 Functions of Nova2

Nova2 is highly expressed in the cortex and hippocampus regions of the brain (Saito *et al.,* 2016) and is sequentially expressed in a dorsal-ventral manner in the spinal cord, with the greatest expression in the dorsal spinal cord, although the large motor neurons of the ventral spinal cord express significant levels of both Nova1 and Nova2 (Yang *et al.,* 1998). In non-neuronal cells, Nova2 is expressed in the lungs (Meldolesi 2020; Yang *et al.,* 1998), endothelial cells (Giampietro *et al.,* 2015), and in white adipocytes (Vernia *et al.,* 2016).

Reduced expression of Nova2 (haploinsufficiency) leads to spontaneous epilepsy (Eom *et al.,* 2013). Five *de novo NOVA2* mutations (2 deletions, 1 insertion, and 2 duplications) and 1 deletion mutation not present in the mother, in the last and largest exon of *NOVA2* are implicated in a serious neuro-developmental disorder characterized by motor delay, speech delay, brain malfunction, seizures, CC thinning, hypotonia and feeding difficulties (Mattioli *et al.,* 2020). These mutations cause a frameshift at the mutation site of the fourth coding exon, that results in a long C-terminal (134 amino acids long) tail and the loss of the third KH (KH3) domain, in the stably expressed *NOVA2* protein (Mattioli *et al.,* 2018).

Nova2 knock out ($Nova2^{-/-}$) mice in Ruggiu *et al.,* 2019 had minimal defects in Z^{+} Agrin splicing in both the brain and spinal cord, AChR clustering or neuromuscular innervation in the motor neurons of the spinal cord. On the contrary, Nova2 in Saito *et al.,* 2016, had a profound splicing regulatory role in the brain cortex, and although $Nova2^{-/-}$ mice were similar in phenotype to their littermates at birth in this study, they died in less than three weeks due to progressive motor dysfunction.

Nova2 has a profound role in the development of the brain cortex and is integral for axon pathfinding and outgrowth in cortical CC axons (Leggere *et al.,* 2016; (Mattioli *et al.,* 2020). Nova2 is also required in the formation and extension of axonal tracts in the brain (Mattioli *et al.,* 2020). Nova2 expression increases gradually during mouse development from E12.5 to E18.5 (Saito *et al.,* 2016), with high expression in the cortical plate and subplate (a region where post-mitotic neurons are found at E18.5 in high numbers), implying a developmental regulation role in neural progenitor cells differentiation (Saito *et al.,* 2016).

The role of Nova2 in this developmental process is related to the AS of five axon guidance-related genes: *Dcc*, *Robo2*, *Epha5*, *Slit2*, and *Neo1*, in which the loss of function of Nova2 results in the aberrant inclusion of developmentally-regulated exons into these axon guidance gene transcripts (Saito *et al.,* 2016). Although Nova2 is highly expressed in the dorsal spinal cord (Yang *et al.,* 1998; Saito *et al.,* 2016), its splicing regulatory effect is only evident in motor neurons innervating the ventral diaphragm muscles and auditory efferent axons (Saito *et al.,* 2016). Nova2 regulates the AS of key genes involved in neurogenesis (Mattioli *et al.,* 2018). For example, it promotes inclusion of exon 26 of *Neogenin1* (*NEO1*) and represses the inclusion of exon 14 of Amyloid Beta A4 Precursor- like protein 2 (*APLP2*). Both *NOVA2* silencing and frameshift mutations affecting the last exon of *NOVA2* (Mattioli *et al.,* 2020; Mattioli *et al.,* 2018) perturb the Nova2-mediated splicing events of these two genes in *in vitro* models (Mattioli *et al.,* 2020; Mattioli *et al.,* 2018). The loss of *NOVA2* function greatly affects axonal formation and synapse function in both *in vivo* and *in vitro* settings (Mattioli *et al.,* 2020; Mattioli *et al.,* 2018; Ruggiu *et al.,* 2009). Important to note is that the mutant *NOVA2* variants (specifically *Mut1*) and the wild type *NOVA2* counterpart had a comparable effect in axonal outgrowth *in vivo* (Mattioli *et al.,* 2020). This might be attributed to the dominant effect of the non-mutated allele (Mattioli *et al.,* 2020), a notion that can be further supported by the finding that human *Mut1 NOVA2* (mutant variant 1) alone could not rescue the splicing defects caused by *NOVA2* orthologue silencing in zebrafish but could do so when injected together with wild type *NOVA2* mRNA (Mattioli *et al.,* 2020).

The axon guidance regulation of Nova2 in the brain cortex appears to be attributed to the developmentally-regulated AS of two netrin receptors: Dcc exon 17 (Dcc

long) and Neo1 exon 27 (Saito *et al.,* 2016). In the spinal cord however, the dual action of Nova1 and Nova2 is required for the regulation of Dcc exon 17 in the developing spinal commissural neuron (Leggere *et al.,* 2016). Netrin proteins are guidance protein that are produced in the ventral floor plate and that promote commissural neuron outgrowth by attracting it towards the ventral plate (Stoeckli, 2018; Duman-Scheel, 2009). Although Nova2 loss of function leads to agenesis of the corpus callosum (ACC), Dcc-long alone could not rescue the ACC defect in Nova2$^{-/-}$ mice (Saito *et al.,* 2016). Dcc-long, however, was able to restore the normal spinal commissural neuron development in Nova1/2 dKO mice (Leggere *et al.,* 2016).

The splicing regulatory role of *NOVA2* is also critical for brain development, neurite outgrowth, and neuronal cell differentiation (Saito *et al.,* 2019; Mattioli *et al.,* 2020). Loss of function of *NOVA2* significantly affects neuronal migration (Yano *et al.,* 2010) and neurite outgrowth (Mattioli *et al.,* 2020). The *de novo* frameshift mutations in Mattioli *et al.* (2020) also significantly altered neuronal cell differentiation and neurite outgrowth *in vivo*.

Nova2 mediates its splicing regulatory effect in different parts of the central nervous system by discriminately binding to specific neurons and neuronal cell types, hence regulating differently the same RNA transcripts in different neurons and neuronal cells (Saito *et al.,* 2019). By using the Nova2 cTag-CLIP method, Saito *et al.* (2019) discovered that the binding profiles of Nova2 are significantly different between inhibitory (GABAergic) and excitatory (glutamatergic) neurons in the developing brain cortex, and that the AS events in these two neuronal populations are different. This discriminatory splicing effect of Nova 2 appears to be critical to the proper development

of hippocampus and neocortex. The discriminatory binding of Nova2 is also evident in the cerebellum, where Nova2 binds selectively to transcripts in Purkinje cells inhibitory neurons, with high binding specificity to 3' UTR binding sites on RNA transcripts (Saito *et al.,* 2019). Loss of this Nova-mediated effect in the Purkinje cells leads to cell-specific progressive motor discoordination and cerebellar atrophy. Taken together, these findings suggest a Nova2-specific and discriminatory binding on the same transcript in different neurons and cell types (Saito *et al.,* 2019).

The mechanism of Nova2 AS regulation appears to be due to its negative interaction effect with another splicing regulator, PTBP2 (Saito *et al.,* 2019). To this effect, Saito *et al.* (2019) found out that Nova2 increases the removal of introns in the alternatively spliced transcripts, hence blocking the binding of PTBP2 to the intronic regions of the mRNA transcript.

Nova2 is also highly expressed in endothelial cells during angiogenesis and its AS role is critical for vascular lumen formation and endothelial cell polarity (Giampietro *et al.,* 2015). Loss of Nova2 disrupts these processes and leads to altered endothelial cells polarity and impaired vascular lumen formation (Giampietro *et al.,* 2015). Nova2 splicing regulatory role in these cells involves the AS of transcripts of key effectors of endothelial cell polarity, including: Par3 (exon 7 inclusion), Magi1 (exon 13a inclusion), Rap1GAP (exon 18a suppression), Dock6 (exon 24 suppression), Dock9 (exon 37a suppression), DBS (exon 37 inclusion) and Pix-α (exon 17 suppression) (Giampietro *et al.,* 2015).

2.2.4 Dual functional role of Nova1 and Nova2

Nova1 appears to play a cooperative function with Nova2 (Ruggiu *et al.,* 2009; Leggere *et al.,* 2016; Saito *et al.,* 2016). Nova1 and Nova2 proteins appeared to play a

dual role in motor nerve function through: (1) NMJ formation, through the AChR clustering function of the Nova-spliced, Z exon-containing (Z^+) Agrin proteins and (2) the motor nerve functioning, through the motor firing functional role of unidentified Nova targets at the proximal motor nerve region (Ruggiu *et al.,* 2009). Nova1/2 double knock out (dKO) mice in Ruggiu *et al.* (2009) were paralyzed due to defective motor function both at the motor nerve and NMJ. Nova1/2 loss of function also disrupts dorsal interneuron development and completely affects auditory efferent innervation (Saito *et al.,* 2016).

During the spinal commissural neuron development, Nova 1/2 dKO mice display defects in the netrin-dependent axon outgrowth process due to disruption of the AS of Dcc (deleted in colorectal carcinoma) long isoform (Leggere *et al.,* 2016). Dcc is a netrin-specific receptor on the axon growth cone (Stoeckli 2018; Duman-Scheel 2009). The mechanism of the cooperative action of both Nova1 and Nova2 in the brain and spinal cord is not clear. These two Nova homologues are expressed in a reciprocal manner in central nervous system and their synergistic effect on neuronal development is still unelucidated.

Both Nova1 and Nova2 are highly expressed in white adipocytes and they both regulate a wide array of AS events associated with diet-induced obesity and adipose tissue thermogenesis, including exon inclusion (768 exons), mutually exclusive exon repression (128 exons), intron activation (99 introns), 64 alternative 5' splice sites (64 sites), and 110 alternative 3' splice (Vernia *et al.,* 2016). The splicing regulation of both Nova1 and Nova2 is essential for metabolism and energy expenditure of high fiber diet and is thus a promoter of diet induced obesity (Vernia *et al.,* 2016).

2.2.5 Nova/Agrin function is critical for NMJ formation, development, and maintenance

The *AGRN* gene encodes a large protein named AGRIN comprised of around 2000 amino acids (~200 kDa) (Nitkin *et al.,* 1987), and has been cloned from several vertebrates including rat (Rupp *et al.,* 1991), chick (Tsim *et al.,* 1992; Denzer *et al.,* 1995) marine ray (*Torpedo californica*) (Smith *et al.,* 1992), man (Groffen *et al.,* 1998), and the invertebrate *C. elegans* (Hurs *et al.,* 2007). Agrin is one of the first proteins found to be involved in the formation and development of the NMJ (Burden *et al.,* 2018). NMJ formation is a multistep process requiring sophisticated interaction between presynaptically-secreted motoneuron-derived Agrin and LRP4 (low-density lipoprotein receptor-related protein 4) that activates the receptor tyrosine kinase MuSK (muscle-specific kinase) on postsynaptic membrane. Consequently, AChRs aggregate into clusters at the postsynaptic membrane, an essential prerequisite for fast, robust, and reliable synaptic transmission (Zhang *et al.,* 2008; Zong *et al.,* 2012; Zong and Rongsheng, 2013). A reduced number of AChRs at the NMJ leads to defective synaptic transmission and is responsible for variety of CMS.

Agrin undergoes AS at three different sites termed X, Y, and Z sites with inserts of 3/12, 4, and 8/11/19 amino acids (AAs), respectively (Gautam *et al.,* 1996), making Agrin a unique model gene to study regulatory mechanism of AS. In mammals, there are two alternative exons at the Z site, termed Z8 and Z11 as they encode for 8 and 11 amino acid (AA) peptides, respectively (Gautam *et al.,* 1996). Expression of Z exons is neuron-specific, thus generating Z^+ Agrin in neurons, while Z^- Agrin is expressed ubiquitously (Gautam *et al.,* 1996). It has been shown that Z8 Agrin is sufficient to induce AChRs clusters and is ~1,000-fold more active than Z^- Agrin (Gautam *et al.,* 1996). Agrin null

mice and motoneuron-specific Z exons KO (knock out) mice ($Agrin^{Z-/Z-}$) are unable to form NMJs and die at birth from diaphragmatic paralysis (Buckanovich *et al.,* 1993; Darnell *et al.,* 2003). AS at Z site of Agrin is regulated by Nova (Ruggiu *et al.,* 2009). Nova harbors 3 KH-type RNA-binding domains and each KH domain has a GXXG motif and a variable loop (Buckanovich *et al.,* 1993; Jensen *et al.,* 2000; Hollingworth *et al.,* 2012). *In vitro* RNA selection (Buckanovich *et al.,* 1993; Jensen *et al.,* 2000; Licatalosi *et al.,* 2008) along with X-ray crystallography (Lewis *et al.,* 2000) revealed that Nova specifically recognizes YCAYs clusters on pre-mRNA transcripts. Moreover, three distinct methodologies, including CLIP (Licatalosi *et al.,* 2008; Ule *et al.,* 2003) splicing microarrays (Ule *et al.,* 2005), and bioinformatics analysis (Ule *et al.,* 2006) have led to the identification of numerous targets of Nova at the neuronal synapses and also showed that Nova binds to YCAY clusters on pre-mRNA *in vivo*. An *in silico* study predicted that binding of Nova to intronic YCAY clusters promotes inclusion of alternative exons (Ule *et al.,* 2006). In mammals, Nova has two members in its gene family: Nova1 and Nova2. Nova dKO mice show dramatic reduction in the inclusion of Z exon of Agrin, and the animal dies immediately after birth from diaphragmatic paralysis (Ruggiu *et al.,* 2009).

2.2.6 *Ciona robusta* as an animal model to study RNA-regulatory networks at the NMJ

Phylogenetically, tunicates (or sea squirts) are invertebrate chordates and are the closest living relatives to vertebrates (Delsuc *et al.,* 2006). The invertebrate *C. robusta* is a suitable model organism with many experimental advantages, i.e. small size, rapid development, easy maintenance, cellular simplicity, streamlined and compact genome that has not undergone the duplications seen in vertebrates. *Ciona* is particularly amenable to molecular perturbation and imaging, and its connectome (Bentley *et al.,*

2016) is only the second one to be completely mapped after that of the nematode *Caenorhabditis elegans* (Hrus *et al.,* 2007), which makes it a powerful tool for genetic analysis. Adult sea squirts are simple, sessile, filter-feeding animals and their free-swimming tadpole-like larvae are composed of only ~2,600 cells and display a simplified body plan that is chordate in a mode of development (Satoh, 1994; Satoh *et al.,* 1995). The tunicate *Ciona* has emerged as a powerful model for studying chordate-specific developmental mechanisms and evolutionary biology (Pennisi, 2002). The dorsally-located larval central nervous system of *Ciona robusta* comprises only 177 neurons distributed rostrocaudally in a brain vesicle, a motor ganglion, and a nerve cord (Ryan *et al.,* 2016). To investigate the regulatory mechanism of AS of neural Agrin by Nova, we have cloned cDNA of putative Agrin and Nova genes from the tunicate *Ciona robusta*. We used *Ciona* for the first time to study regulatory mechanism of AS of neural Agrin. The findings are discussed in the following sections.

2.3 MATERIALS AND METHODS

2.3.1 Animal model and mammalian cell

The tunicate, *Ciona robusta*, a marine invertebrate was used as an animal model in our study. The tissue sample of *Ciona* was collected to clone full length Agrin and Nova. Human HEK293T cells (ATCC) were used to transfect/co-transfect all the constructs generated in this study.

2.3.2 Cloning procedures

A total of 33 CiAgrin minigene constructs (WT and mutants) and 17 CiNova constructs (WT and mutants) were prepared to investigate the regulatory mechanism of Nova/Agrin splicing and to discover the specific RNA-binding domain(s). Moreover, 16 mouse Nova (WT and mutants) were prepared to investigate splicing of mouse *Agrin* minigene mAgrin_31-34-3x-Flag (Saito *et al.,* 2016) (a gift of Dr. Robert B. Darnell from Rockefeller University). A mouse *Dcc* minigene (Leggere *et al.,* 2016) (a gift of Dr. Zhe Chen from University of Colorado, Boulder) was also used for verification of our splicing assay in our lab. To investigate the disease pathology of epileptic encephalopathy in a proband, 9 *SLC25A10* minigene constructs, including WT and mutants, were constructed. *Ciona* minigenes and *SLC25A10* minigenes were cloned in pCi-neo vector (Promega). Mouse *Snap25* minigene was cloned in the exon trapping vector pSPL3 (Nisson *et al.,* 1994; Tompson *et al.,* 2017). Mouse *Dcc* minigene, is cloned in pDEST26 (Invitrogen). cDNA constructs of *Ciona* and mouse Nova were cloned in pEGFP-C1 vector (Clontech). All these vectors contain a CMV promoter, including *mAgrin*_31-34-3x-Flag (Saito *et al.,* 2016). Moreover, cDNA constructs of mouse Nova1, Nova2, Rbfox1, Ptbp2,

and Mbln2 were cloned in pCAGGS-3x-Flag vector (a gift from Dr. Chaolin Zhang from Columbia University).

2.3.3 Proof-reading PCR

A mixture of 5 X buffer (10 µL); 10 mM dNTPs (1 µlL; 10 µM forward primer (2.5 µL); 10 µM reverse primer (2.5 µL); template genomic DNA from *Ciona robusta* (10 ng); Q5 Hot Start High-Fidelity DNA Polymerase 2,000 U/ ml (New England BioLabs) (0.3 µlL was prepared in total volume of 50 µL (diluted in sterile nuclease-free H_2O) in a PCR tube. Some of the reactions were supported by adding 10 µl of 5 M Betaine Solution (Sigma), for a final concentration of 1 M Betaine. PCR reactions were performed in a T100 thermal cycler (Bio-Rad) with initial denaturation for 5 minutes at 98 °C; variable number of cycles, followed by 2 minutes of final extension at 72 °C and a hold at 4 °C. The annealing temperature for each primer pair was calculated using New England BioLabs Tm calculator version 1.13.0 (http://tmcalculator.neb.com/#!/ as of April 2021).

2.3.4 Site-Directed Mutagenesis

Point mutations (YCAY to YAAY) in Agrin, and deletion mutations in Nova and its KH domains (GXXG to GDDG) were introduced by Q5 Hot Start DNA Polymerase (Q5 Site-Directed Mutagenesis Kit, New England Biolabs) according to the manufacturer. The mutants were cloned into appropriate vectors and were confirmed by DNA sequencing.

2.3.5 Gel extraction

The PCR product was then purified with a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer. The DNA fragment was excised from the agarose gel (made

in TAE buffer) with a scalpel. The gel piece was weighed in an Eppendorf tube. 3 volumes of Buffer QG were added to 1 volume gel. This was followed by an incubation at 50° C for 10 minutes until the gel slice was completely dissolved (vortexing the tube every 2-3 minutes). 1 volume of Isopropanol (BDH) was added to the mixture and vortexed properly. The mixture was then loaded into the QIAquick spin column (DNA was bound to the membrane of the column) that was attached to a vacuum manifold. The column was washed two times with 750 µL of Buffer PE in the vacuum manifold. Then the column was placed in the provided 2 mL collection tube and centrifuged for 1 minute at 14,500 rpm in order to remove residual wash buffer. The QIAquick spin column was then placed into a sterile 1.5 mL Eppendorf tube and DNA was eluted with the addition of 50 µL Buffer EB to the center of the QIAquick membrane, letting the column stand for 2 minutes and centrifuging for 2 minutes at 14,500 rpm. Purified PCR products were stored at -20° C.

2.3.6 Vector and insert preparation

The mammalian expression vectors (pCi-neo and pEGFP-C1) and the purified Q5® Hot Start High-Fidelity DNA Polymerase products (i.e. *Ciona* Agrin and *Ciona* Nova) were digested with the appropriate restriction enzymes (New England BioLabs) according to the manufacturer. Both vector and PCR product digestion mixtures were prepared with: 10X CutSmart Buffer (10 µL), the enzyme pair used for cloning (3 µL each) in a final volume of 100 µL. 2 µg of vector and the whole purified PCR product were digested in a total reaction volume of 100 µL (diluted in sterile ddH_2O) in separate tubes. The mixtures were incubated O/N at 37° C. 1 µL of each enzyme was added and

the mixtures were incubated another hour at 37° C. This was followed by a phenol chloroform extraction.

2.3.7 Vector dephosphorylation

In order to prevent self-ligation, the 5' phosphate groups rewe removed from the digested vector prior to ligation. A mixture of the whole amount of digested plasmid (20 µL); CutSmart buffer (3 µL); Shrimp Alkaline Phosphatase (rSAP) (New England BioLabs) (1.5 µL); 5.5 µL sterile H_2O was prepared in an Eppendorf tube. The mix was then incubated for 30 minutes at 37° C, followed by an incubation for 5 minutes at 65° C to inactivate the enzyme according to the manufacturer. This was followed by a phenol chloroform extraction.

2.3.8 Phenol Chloroform extraction and DNA/RNA precipitation

The Phenol Chloroform extraction was performed by adding 180 µL Phenol:Chlorophorm (AMRESCO) to the DNA or RNA + enzymes mixture and H_2O to 360 µl. The resulting mix was vortexed and spun for 5 minutes at 12,000g. The supernatant was transferred to a new Eppendorf tube. This was followed by addition of 3 M Sodium Acetate (AMRESCO) pH 5.2, (20 µl); GlycoBlue 15 mg/ml (Ambion) (2 µL) and 2.5 volumes of 100% Ethanol (Fisher Schientific) (500 µL). The mixture was vortexed and incubated for 30 minutes at - 80° C. It was then spun for 30 minutes at 4° C at 1,200g. The supernatant was removed, and the pellet was washed with 70% Ethanol (500 µL). This was followed by a 10-minute centrifugation at 4° C and 12,000g. The supernatant was removed, and the pellet was dried at 37° C. H_2O was added according to expected concentration, and the nucleic acid concentration was measure using a

BioSpectrometer spectrophotometer (Eppendorf). Resulting pure DNA was stored at -20° C. Resulting pure RNA was stored at -80° C.

2.3.9 Ligation of insert and plasmid

Ciona Agrin inserts (and other minigenes) were cloned into the pCI-neo mammalian expression vector (Promega) (5472 bp) and *Ciona* Nova (and other splicing factors) inserts were cloned into pEGFP-C1 vector (Clontech) (4731 bp). The insert integration was performed using T4 DNA Ligase (New England BioLabs) and an insert-to-vector ratio of 3:1. A mix of 50 ng of vector (1 µL); 2 µL 10X Ligation Reaction Buffer; 1 µL T4 DNA Ligase was prepared in an Eppendorf tube. It was centrifuged briefly and incubated O/N at 16° C. The mix was then used to transform *Escherichia coli* DH5α competent cells.

2.3.10 Transformation of *E. coli* DH5α cells

Ligation products were transformed into competent *E. coli* DH5α cells (New England BioLabs). The competent cells were thawed on ice. About 4 µL of the ligation mixture was added to 50 µL of competent cells and mixed by flicking the tube 2-3 times. The mix was then placed on ice for 30 minutes. This was followed by a heat shock at 42° C for 30 seconds. After this, 950 µL of room temperature SOC media was added, and the cells were shaken at 250 rpm for 60 minutes at 37° C. The cells were then spread on LB agar (AMRESCO) plates under antibiotic selection and incubated overnight at 37° C.

2.3.11 Miniprep

Colonies from transformation plates were picked with a toothpick and grown overnight at 37° C while shaking at 250 rpm in 5 mL LB medium (AMRESCO)

supplemented with the appropriate antibiotic (AMRESCO). Before the miniprep procedure, each cell culture was streaked on LB agar plates (AMRESCO) containing the appropriate antibiotic; plates were incubated overnight at 37° C, and then stored at 4° C. The bacterial culture was harvested by centrifugation at 8,000 rpm in a standard bench top microcentrifuge for 2 minutes at room temperature. The supernatant was removed. All purification steps were carried out in a table-top microcentrifuge at 12,000g with a GeneJET Plasmid Miniprep Kit (Thermo Scientific). The pellet was completely resuspended in 250 µL Resuspension Solution and transferred to a microcentrifuge tube. 250 µL Lysis Solution was then added and the tube was inverted 4-6 times until the solution became viscous and slightly clear. 350 µL Neutralization Solution was added and the tube was immediately inverted 4-6 times. The mix was centrifuged for 5 minutes to pellet cell debris and chromosomal DNA. The supernatant was transferred to a GeneJET spin column placed on a vacuum manifold and vacuum was applied to bind plasmid DNA to the column membrane. The column was washed with 500 µL of Wash Solution twice. Then the column was placed in the provided 2 mL collection tube and centrifuged for 1 minute in order to remove residual Wash Solution. The GeneJET spin column was then placed into a clean 1.5 mL Eppendorf tube and DNA was eluted with the addition of 50 µl Elution Buffer to the center of the membrane, incubating for 2 minutes at room temperature and centrifuging for 2 minutes. Purified plasmid DNA was stored at -20° C.

2.3.12 Test cut

The isolated constructs from the miniprep were digested with the restriction enzymes (New England BioLabs) used for cloning according to manufacturer. The

digestion mixture was prepared with: 10X CutSmart Buffer (1 µL); the enzyme pair used for cloning (0.5 µL each). 200 ng of construct in a total volume of the reaction of 10 µL (diluted in sterile ddH_2O). The mixtures were incubated 30 minutes at 37° C. A following incubation for 5 minutes at 65° C was done in order to inactivate the enzymes.

2.3.13 Gel electrophoresis

The digested constructs were then run on an agarose gel. The gels were made with variable percentages of Agarose (AMRESCO) in TBE or TAE buffers. 10X TBE was prepared by mixing Tris Base (AMRESCO) (60.55 g); Boric Acid (AMRESCO) (30.9 g); 0.5M EDTA pH 8.0 (AMRESCO) (20 mL) and bringing the volume to 1L with ddH_2O. 50X TAE was prepared by mixing Tris Base (AMRESCO) (242 g); Glacial Acetic Acid (AMRESCO) (57.1 mL); 0.5M EDTA pH 8.0 (AMRESCO) (100 mL) and bringing the volume to 1L with ddH_2O. The buffers were autoclave sterilized. TBE gels were used for gel extraction. For the visualization of DNA, we used 3x GelRed (Biotium) (33 mL in a 100 mL gel), with exception of the gels used for running digested vectors which we stained with Ethidium Bromide, 10 mg/ml (AMRESCO) (5 µL in a 100 mL gel).

2.3.14 Sequencing

The plasmids containing the inserts of expected length were Sanger sequenced and the clones with correct insert were used for maxiprep for future cell line transfection. All the constructs were confirmed by the sequencing. The sequencing was performed by Genewiz.

2.3.15 Maxiprep

The appropriate bacterial cells were picked from the streak plate prepared before the miniprep procedure. They were grown in 1 mL of LB broth supplemented with the proper antibiotic for 4 hours and then transferred to a flask containing 200 mL of LB broth for overnight growth at 37° C while shaking at 250 rpm. 50 % Glycerol stocks were prepared from the bacterial culture and stored at -80° C (937 µL of sterile 80 % Glycerol + 563 µL bacterial culture). The bacterial culture was harvested by centrifugation at 5,000g for 10 minutes at room temperature. The supernatant was removed. All purification steps were carried out with a maxiprep kit (GeneJET Plasmid maxiprep kit from Thermo Scientific and Macherey Nagel maxiprep kit from Macherey Nagel) at room temperature. The bacterial cell pellet was resuspended in 6 mL Resuspension Solution. 6 mL Lysis Solution was added and mixed gently by inverting the tube 4-6 times until the solution became viscous and slightly clear. This was followed by incubation for 3 minutes. 6 mL Neutralization Solution was then added and mixed immediately by inverting the tube 5-8 times. This was followed by addition of 0.8 mL Endotoxin Binding Reagent and mixing by inverting the tube 5-8 times, followed by incubation for 5 minutes. 6 mL of 96% ethanol was then added and mixed by inverting the tube 5-8 times. This was followed by centrifugation at 5,000g for 40 minutes to pellet cell debris and chromosomal DNA. The supernatant was transferred to a 50 mL tube. 6 mL of 96% ethanol was added and mixed by inverting the tube 5-8 times. The sample was transferred to the column placed on a vacuum manifold and vacuum was applied to bind plasmid DNA to the column membrane. The column was washed with 8 mL Wash Solution I and then twice with 8 mL Wash Solution II. The column was then placed in the collection tube and centrifuged for 5 minutes at 3,000g in a swinging bucket rotor to

remove residual Wash Solution. The column was transferred to a fresh collection tube. 1 mL Elution Buffer was added to the center of the purification column membrane. This was followed by incubation for 2 minutes and centrifugation for 5 minutes at 3,000g to elute plasmid DNA. The purified plasmid DNA was stored at -20° C.

2.3.16 Transfection procedure for splicing assay

The day before the transfection 0.6 x 10^6 HEK293T cells were seeded per well in a 6-well plate (USA Scientific) in DMEM culture medium. On the day of transfection, a total of 2.5 µg DNA of minigene, cDNA construct, and empty vector was used to transfect each of 6 well plate(s) and 7.5 µL of linear polyethylenimine (PEI; Polysciences), MW 25,000 (1mg/mL) was used in a ratio of 1:3 (DNA : PEI). 0.5 µg (= 1x) of minigene DNA was used in each well to test splicing with different amount of splicing factor (0µg = 0x, 0.5µg = 1x, and 2.0µg = 4x). Empty vector was used to bring the total amount of DNA to 2.5 µg (2.0µg = 4x, 1.5µg = 3x, and 0µg = 0x) per well. The total volume of the DNA mixture was 200 µL (Table 6 and 7). First, the exact amount of DNA in µL was pipetted in 1.5 µL Eppendorf tube (Eppendorf) and Opti-MEM media (Thermo Scientific) was used to bring the volume to 192.5 µL. Then the mixture was vortexed thoroughly. Finally, 7.5 µL of PEI was added, vortexed, and centrifuged briefly. The mixture was then incubated for 15 minutes at room temperature. In the meantime, medium in the cells was aspirated and 2 mL of fresh DMEM medium was added. After a 15 minutes incubation, 200 µL of reaction mixture was added to the cell and the plate was cross-shaked gently. The plate was then incubated for 48 hours at 37 ° C.

2.3.17 RNA extraction

RNA from transfected HEK293T cells was extracted 48 hours after transfection using RiboZol RNA Extraction Reagent (AMRESCO) or IBI Isolate (IBI Scientific) according to the manufacturer. HEK293T cells were homogenized using pipette tip in 500 µL RiboZol and transferred into a 1.5 mL Eppendorf tube and incubated for 10 minutes at room temperature to ensure complete dissociation of the nucleoprotein complexes. This was followed by addition of 100 µL chloroform and shaking the tube vigorously for 15 seconds. The sample was then incubated for 3 minutes at room temperature. This was followed by centrifugation at 12,000x rpm for 15 minutes at 4° C. After centrifugation 3 phases were visible: a) lower red, phenol-chloroform phase; b) white interphase; c) colorless, upper, aqueous phase. RNA locates exclusively in the upper phase and about 80% of this phase were transferred to a new tube without touching the interphase. RNA was then precipitated by adding 250 µL of Isopropanol (BDH) and 2 µL of RNA grade glycogen (Thermo Scientific). The samples were incubated for 10 minutes at room temperature and then centrifuged at 12,000x rpm for 10 minutes at 4° C. After centrifugation a white pellet of RNA was visible at the bottom of the tubes. The supernatant was removed without disturbing the pellet. The pellet was then washed with 500 µL of 70% ethanol prepared with RNase-free water – vortexed and centrifuged at 14,000x rpm for 5 minutes at 4° C. After this the pellet was air-dried for 5 minutes and re-dissolved in RNase-free water (Thermo Scientific) by passing it several times through a pipette tip and incubating for 10 minutes at 50° C to completely dissolve. After reading the concentration of RNA with a spectrophotometer the samples were stored at -80° C.

2.3.18 DNase I treatment

A total of 5 µg of RNA were digested with DNase I in order to remove any genomic DNA contamination from RNA samples in a 50 µL reaction containing 1.5 µL of TurboDNase (Thermo Scientific), 5 µL of 10X Buffer, and ddH_2O to 50 µL. After 30 minutes of incubation at 37 °C another 1.5 µL of TurboDNase was added to the mixture and incubated for another 30 minutes. After a total of one-hour incubation 10 µL of TurboDNAse Inactivation Reagent was added and samples were kept at room temperature for 5 minutes, and the tubes were flicked every 2 minutes to resuspend the inactivation reagent. Then the tubes were centrifuged at 10,000 rpm for 90 seconds. An adequate amount of supernatant was collected for further processing.

2.3.19 Reverse transcription

From total RNA we synthesized cDNA using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). A mix of 250 ng RNA and 1 µL of oligo $(dT)_{18}$ 500 ng/µL was prepared in a total volume of 12 µL (diluted in sterile ddH_2O). The mix was incubated for 5 minutes at 65 °C in a PCR machine. After this, an RT reaction mix was prepared combining the mixture above with the following ingredients in a total volume of 20 µL: 5X RT Buffer (4 µL); RiboLock RNase Inhibitor 20 U/µl (0.5 µL); 10 mM dNTPs (2 µL); RevertAid RT 200 U/µl (0.5 µl). This mixture was incubated for 1 hour at 42 °C followed by 5 minutes at 72 °C in a PCR machine. After the incubation, 5 µL of H_2O were added to each tube bringing the volume to a total of 25 µL. Each RT reaction mix had a concentration of 10 ng of starting RNA/µL. 5 µL from each RT reaction, equivalent to 50 ng of starting RNA, were used as template for each RT-PCR.

2.3.20 RT-PCR

A mixture of 10X PCR Buffer (5 μL); dNTPs 10 mM (1 μL); forward and reverse primers each 10 μM (1 μL); 5 U/μL HotStarTaq Plus DNA polymerase (Qiagen) (0.4 μL) or 5 U/μL Dream Taq Hot Start DNA polymerase (Thermo Scientific) (0.4 μL) and RT reaction (5 μL) in a total volume of 50 μL (diluted in sterile ddH_2O) was prepared in a PCR tube. The PCR reaction was performed with initial denaturation for 5 minutes at 95 °C; variable number of cycles of: denaturation for 30 seconds at 94° C, annealing for 30 seconds at 60°C and elongation for 30 seconds at 72 °C. This was followed by a final extension of 7 minutes at 72 °C and a hold at 12 °C.

2.3.21 Western Blotting

The cDNA constructs all the pEGFP-Nova (WT and mutants) were co-transfected with Agrin minigenes into HEK293 cells. After 48 hours cells were collected and resuspended in lysis buffer: 0.5% deoxycholic acid sodium salt (Fisher), 0.1% SDS (AMRESCO), 0.5% NP-40 (Calbiochem), 1x PBS and 50% glycerol (AMRESCO)) with protease inhibitor cocktail (AMRESCO). The lysate was left on ice for 20'. After sonication at 50% amplitude (Fisher Scientific Sonic Dismembrator Ultrasonic Processor FB120, 120 W 20 kHz), the lysate was subjected to centrifugation at 14,000 rpm at 4° C for 15 minutes and the supernatant collected in a fresh tube. Total protein amounts were calculated using a standard Bradford assay and 10 μg of protein extract from each sample were loaded on a 8% SDS-polyacrylamide gel. Proteins were then transferred from gel onto PVDF membrane (Millipore, Immobilon-FL) and blocked in Odyssey blocking buffer (TBS, Li-cor). Membranes were blotted with mouse monoclonal anti-GFP antibody (Santa Cruz biotech, sc-9996, 1:1000 dilution) and then donkey anti-mouse

IRDye 800CW secondary antibody (Li-cor, 926-32212, 1:5000 dilution). The signal was detected using the Odyssey CLx imaging system (Li-cor).

2.4 RESULTS

2.4.1 Agrin's Z exons and Nova are conserved in *Ciona robusta*

To dissect the regulatory mechanism of Agrin splicing by Nova, we cloned full-length Agrin and Nova cDNAs from hatched larvae at 22.5 hours post fertilization (hpf) at 20° C and brain from adult *Ciona robusta* (Fig. 1A, B). *Ciona* Agrin (CiAgrin) and *Ciona* Nova (CiNova) cDNAs are ~8 kb and ~2 kb long, respectively. The Z exons at Z site of CiAgrin are 18 and 15 bases long and code for 6 and 5 amino acids, respectively. Therefore, we termed the Z exons from *Ciona* Agrin Z6 and Z5, respectively. By direct cloning and sequencing of RT-PCR product, we detected all the different isoforms of Z^+ Agrin from larvae at 22.5 hpf and from adult brain (Fig. 1C). Interestingly, robust expression of Z5 and Z6 splice isoforms was detected in brain tissues while Z11 appears to be the predominant form in larvae. To investigate the molecular mechanism regulating AS of the Z exons of CiAgrin, we generated a minigene construct spanning the genomic region of *Ciona* Agrin encompassing the Z exons and flanking introns and including the upstream and downstream constitutive exons (exon 40 and 41; Fig. 2A). This genomic region was amplified by PCR and cloned into the mammalian expression vector pCi-neo. We cloned 3 isoforms of CiNova form *Ciona* and we named them CiNova_MMM, CiNova_MLN, and CiNova_MEY respectively based on the first 3 AA at the N-terminus (Fig. 1D). These 3 isoforms are the product of AS where two different first exons can be used: Inclusion of a distal first exon, termed exon 1a, gives rise to CiNova_MMM, while inclusion of a proximal first exon, termed exon 1b, gives rise to CiNova_MLN and CiNova_MEY. Of these 3 isoforms, only CiNova_MMM has a nuclear localization signal (NLS) present in first exon 1a. CiNova_MLN and CiNova_MEY, on the other hand, include an alternative first exon 1b that does not include an NLS. Therefore, in

Ciona, Nova undergoes AS at the first exon to generate two isoforms: one that contains an NLS, and a second one that does not. Compared to CiNova_MLN, CiNova_MEY uses a downstream first codon and therefore lacks the first 4 AA at its N-terminus. The 3 full-length *Ciona* Nova cDNAs were cloned into the pEGFP-C1 mammalian expression vector in order to generate Nova constructs tagged with EGFP in-frame at their N-termini.

2.4.2 Nova promotes inclusion of Agrin's Z exons in a dose-dependent manner and Z exons splicing is species specific

To investigate the mechanism regulating AS of the Z exons of CiAgrin, we generated a minigene construct (Fig. 2A) spanning the genomic region of *Ciona* Agrin encompassing the Z exons and flanking introns and including the upstream and downstream constitutive exons (exon 40 and 41). This genomic region was amplified by PCR and cloned into the mammalian expression vector pCi-neo. To perform *in vitro* minigene splicing assays, we co-transfected constant amounts of our *Ciona* Agrin minigene construct with increasing amounts of EGFP-CiNova in HEK293T cells. Total RNA was extracted, quantified, normalized and subjected to AS analysis by semi-quantitative RT-PCR. Using this splicing assay, we discovered that all 3 CiNova constructs promote inclusion of CiAgrin's Z exons in a dose-dependent manner. This means that all *Ciona* isoforms of Nova are able to splice independently of their N-terminus; that is, independently of whether the N-terminus includes the canonical NLS or not. In details, in this *in vitro* splicing assay we were able to detect two Z isoforms of Z^+ Agrin: Z5 and Z11 (Z6+Z5), but not Z6 (Fig. 2D). To confirm the expression of the CiNova proteins tagged with EGFP at the N-terminus, we co-transfected all 3 EGFP-CiNova constructs with our CiAgrin minigene in HEK293T cells and analyzed EGFP-

Nova protein expression by western blotting. All Nova proteins were robustly expressed in a dose-dependent manner (Fig. 2E). For simplicity, all further splicing assays of CiAgrin (WT and mutants) were performed using CiNova_MLN.

In mammals, there are two different *NOVA* genes which encode for two highly homologous proteins, NOVA1 and NOVA2. Ruggiu *et al.* (2009) reported that mouse Nova (mNova1 and mNova2) regulates AS of mouse Agrin and mNova dKO mice significantly abolish inclusion of Z exons and the dKO mice die immediately after birth. We generated two cDNA constructs of mouse Nova into the pEGFP-C1 mammalian expression vector to further investigate the splicing of Z exons of mouse Agrin in our *in vitro* splicing assay system. We used a mouse Agrin minigene spanning the genomic region of Agrin encompassing the Z exons (32 and 33) and flanking introns and including the upstream and downstream constitutive exons (exon 31 and 34) used in previous studies (Fig. 2B). Co-transfection was carried out in HEK293T cells with constant amounts of mouse Agrin minigene and increasing amounts of EGFP-mNova1 or mNova2 and we found that both mNova1 and mNova2 promote inclusion of Z exons in a dose-dependent manner. In our minigene splicing assay we were able to detect only the Z8 isoform of Z^+ Agrin (Fig. 2F).

RBFOX1 (also known as A2BP1 or FOX1) is a neuron-specific RNA-binding protein that exerts both positive and negative regulatory effects on AS (Underwood *et al.,* 2005). RBFOX1 has been implicated in numerous neurodevelopmental and neuropsychiatric disorders including autism spectrum disorder, mental retardation and epilepsy (Ballah *et al.,* 2004; Martin *et al.,* 2007; Sebta *et al.,* 2007; Wang *et al.,* 2009; Voineagu *et al.,* 2011; Wintle et al. 2011), attention deficit hyperactivity disorder (Elia *et*

al., 2010), bipolar disorder, schizoaffective disorder and schizophrenia (Xu *et al.,* 2008; Le-Niculescu *et al.,* 2009; Hamshere *et al.,* 2009). Two other important alternative splicing regulators in brain development Ptbp2 (Licatalosi *et al.,* 2012) and Mbln2 (Charizanis *et al.,* 2012) were also used to investigate any potential alternative splicing regulation of mouse Agrin and Snap25 minigenes. To explore if mouse Rbfox1 regulates AS of the Z exons of mouse Agrin, we co-transfected Rbfox1 with the mouse Agrin minigene in HEK293T cells and discovered that Rbfox1 regulates the splicing of Agrin by promoting inclusion of the Z exons in a dose-dependent manner as well (Fig. 2G). To our knowledge there is no known cis-acting element for Rbfox1 in mouse Agrin and its splicing regulation has never been reported by Rbfox1. Moreover, we have not observed any splicing regulation when mouse Agrin minigene was tested against splicing factors Ptbp2 and Mbnl2 (data not shown), suggesting that the regulation of Agrin's AS by Rbfox is specific. Leggere *et al.,* 2016 reported that Nova regulates the AS of Dcc pre-mRNA. We carried out splicing assays in HEK293T cells by co-transfecting mouse Nova with a Dcc minigene containing the genomic DNA between exons 16 and 17 and verified that we could replicate the results by Leggere *et al.,* 2016 (Fig. 2H, I).

2.4.3 Expression of Agrin's Z isoforms is developmentally regulated

We collected tissue samples from different developmental stages of *Ciona*, extracted total RNA, and investigated the expression of Agrin and Nova by semi-quantitative RT-PCR. In this analysis we also included unfertilized eggs, and adult brain and heart tissue. We found that, while total Agrin and Nova are robustly expressed throughout development (Fig. 3, first panel and third panel), Agrin's Z exons start to be included at 10 hpf (Fig. 3, first panel). Interestingly, our RT-PCR data indicate that Z11

is the most abundant Z^+ Agrin isoform, peaking in hatched larvae at 22.5 hpf during development, while the smaller Z5 isoform appears to be the predominant one in adult brain (Fig. 3, first panel). We also observed weak expression of Nova and Z^+ Agrin in heart muscle of *Ciona*. Interestingly, total Agrin is also robustly expressed in unfertilized eggs, suggesting that Agrin is maternally inherited in *Ciona*, and it may play an important role during early stages of larval development. The expression of Actin was monitored as experimental (housekeeping) control (Fig. 3, last panel).

2.4.4 Nova requires its first two KH domains to splice Agrin's Z exons while KH3 is dispensable

KH (hnRNP K-homology, ~70 AA long) domains are evolutionarily conserved RNA-binding domains that are present in both eukaryotes and prokaryotes (Musco *et al.,* 1996; Siomi *et al.,* 1993; Grishin *et al.,* 2001) and are found in a wide range of nucleic acid binding proteins, including Nova that harbors 3 KH domains and is exclusively expressed in the central nervous system and in certain tumors (Zhang *et al.,* 2016; Buckanovich *et al.,* 1993; Darnell *et al.,* 2003). KH domain-harboring proteins play key roles in many cellular processes, such as translation, AS of pre-mRNA, and mRNA localization (Buckanovich *et al.,* 1993; Darnell *et al.,* 2003; Siomi *et al.,* 1993) and loss of function of KH domains is associated with several diseases, including paraneoplastic syndromes and some cancers (Valverde *et al.,* 2008). Previous work showed that a missense mutation in the KH domain of *FMRP1*, the fragile X mental retardation gene, impairs RNA binding and causes severe mental retardation in humans (Siomi *et al.,* 1994). KH domains have a conserved GXXG motif and a variable loop (Buckanovich *et al.,* 1997; Jensen *et al.,* 2000; Hollingworth *et al.,* 2012) and it has been shown that

mutating the conserved GXXG motif to GDDG abolishes RNA-protein interactions while keeping the protein structure intact (Hollingworth *et al.,* 2012).

To determine which KH domain or domains of CiNova is/are required for splicing of Z exons, we generated KH GDDG mutants in all possible combinations including single (KH1, KH2, KH3), double (KH1/KH2, KH2/KH3, KH1/KH3), and triple (KH1/KH2/KH3). To this end, we co-transfected HEK293T cells with a constant amount (1x) of CiAgrin minigene with varying amounts (0x, 1x, 4x) of WT and all the GDDG mutants of CiNova. Total RNA was extracted, quantified, and subjected to RT-PCR analysis to determine the inclusion of Agrin's Z exons by WT and GDDG CiNova mutants. We observed that WT CiNova promotes inclusion of Z exons in a dose-dependent manner, while CiNova KH1 and KH2 GDDG mutations abolish splicing (Fig. 4B). However, mutation of the third KH domain has no negative effect on splicing. RT-PCR analysis of co-transfection experiments including KH GDDG double and triple mutants further suggests that CiNova requires both KH1 and KH2 domains for promoting Z exons inclusion, while KH3 is dispensable (Fig. 4B). Since CiNova_MMM and CiNova_MLN have different N-terminal, we therefore generated all the combinations of KH GDDG mutants for CiNova_MMM and co-transfected them with our CiAgrin minigene in HEK293T cells. Despite their different N-termini, in our splicing assays we were unable to detect any differences in splicing activity between CiNova_MLN and CiNova_MMM GDDG mutants (Fig. 6A, B). To investigate KH domain requirement by mouse Nova, we also generated all the combinations of KH GDDG mutants of both mouse Nova1 and Nova2 and co-transfected them with our mouse Agrin minigene in HEK293T cells. We found that only KH3 GDDG mutant of Nova2 abolishes splicing of

Z exons of mouse Agrin (Fig. 7C), while none of the individual KH mutant of Nova1 abolishes splicing (Fig. 7A). However, double and triple KH GDDG mutants of both Nova1 and Nova2 completely abolish splicing of Z exons of mouse Agrin (Fig. 7B, D).

2.4.5 Novel function of N/C-terminals and KH3 domain of Nova

To investigate possible novel function of N and C-terminals and KH3 domain, we generated 7 deletion mutants of CiNova by deleting N and C-terminals, KH3 domain and various combinations of them. We discovered that deleting either N- or C-terminals or both of them together completely abolishes splicing of Z exons of CiAgrin even though both KH1 and KH2 domains are intact (Fig. 5A,6C). Consistent with our data indicating that KH3 is dispensable for splicing, KH3 deletion mutant has no effect on splicing, as expected. Interestingly, splicing of Z exons was rescued when deletion of either N- and C-terminus or both together was coupled with the deletion of KH3 domain, a finding that suggests a complex regulatory mechanism of splicing between different intramolecular CiNova domains (Fig. 5B, 6D).

2.4.6 A bipartite intronic splicing enhancer mediates Nova-dependent inclusion of Agrin's Z exons

Nova targets are enriched in YCAY clusters as Nova specifically recognizes YCAY sequences on pre-mRNA (Jensen *et al.,* 2000; Ule *et al.,* 2003; Ule *et al.,* 2005; Ule *et al.,* 2006). YCAY-rich sequences on Nova targets can be exonic or intronic and are conserved between mouse and human (Ule *et al.,* 2006). The position of the YCAY cluster and Nova's binding to specific cluster(s) determines splicing outcome (Ule *et al.,* 2006). Previous work has shown that Nova promotes inclusion of alternative exons when it binds to intronic YCAY clusters downstream of alternative exons, but binding to a

cluster located on the alternative exon or upstream of it, on the other hand, doesn't promote inclusion but skipping (Ule *et al.,* 2006). Mutating the targets of Nova from YCAY to YAAY has been shown to abolish both Nova binding (Buckanovich *et al.,* 1997) and splicing (Leggre *et al.,* 2016). In CiAgrin intronic YCAY sequences are exclusively concentrated in the intron downstream of exon Z5 (18 YCAY sequences in intron 40) and only 5 YCAY sequences are present on flanking constitutive exons: one in exon 40 and 4 in exon 41 (Fig. 8A).

To dissect the cis-regulatory YCAY elements of CiAgrin that may mediate Nova-dependent inclusion of the Z exons, we generated 32 mutant minigenes of CiAgrin by mutating its 23 YCAY sequences to YAAY in different combinations (Fig. 8E). To this end, co-transfection of constant amounts (1x) of CiAgrin minigene (WT and YAAY mutants) with varying amounts (0x, 1x, 4x) of CiNova in HEK293T cells was performed. Total RNA was extracted, quantified, and subjected to semi-quantitative RT-PCR analysis to monitor the inclusion of Agrin's Z exons by CiNova. Our results suggest that we have identified two Nova-dependent intronic splicing enhancers (NISE) elements, termed NISE1 and NISE2, in intron 40 (Fig. 8E). NISE1 contains the first 6 YCAY sequences while NISE2 covers YCAY11 to YCAY14 (Fig. 8A, E). Exonic YCAY sequences, on the other hand, are not required for splicing at the Z site of CiAgrin (Fig. 8B). We noticed that in CiAgrin the average nt distance between the intronic YCAY repeats is 32. Interestingly, the average nt distance in the NISE elements is reduced to 10 nt and 22 nt in NISE1 and NISE2, respectively. However, our findings suggest that the spacing between single YCAY sequences does not appear to be critical for splicing (Fig. 8F). For example, YCAY 6 and 7, the closest YCAYs, are only 2 nt apart and double

mutant YAAY 6-7 cannot abolish splicing. Conversely, YCAY 12 and 13 are also 2 nt apart but double mutant YAAY 12-13 is unable to splice. At the other end of the spectrum, YCAY 13 and 14 are 56 nt apart, and double mutant YAAY 13-14 is unable to splice, while ICAY 17 and 18 are 111 nt apart but double mutant YAAY 17-18 is still able to splice.

2.4.7 Agrin's Z exons splicing is species specific

To determine species specificity of Agrin's Z exons splicing, we co-transfected CiAgrin minigene with individual cDNA constructs of Nova from either mouse or *Ciona* in HEK293T cells. Total RNA was extracted and subjected to semi-quantitative RT-PCR analysis. The contemporary experiment was also performed where HEK293T cells were co-transfected with mouse Agrin minigene and Nova from either mouse or *Ciona* and subjected to semi-quantitative RT-PCR analysis after total RNA extraction. We observed that the Z exons of Agrin are spliced only by its species-specific Nova: Nova from mouse cannot splice CiAgrin minigene and mouse Agrin minigene is not spliced by mouse Nova (Fig. 9A, B).

2.4.8 Exons 5a and 5b of *Snap25* are regulated by mouse Nova1/2, Rbfox1, and Ptbp2 but not Mbnl2

SNAP25 plays a crucial role in neuroexocytosis by linking synaptic vesicles to the plasma membrane during regulated neurotransmitter release. Exon 5 of *SNAP25* is a perfect example of exon duplication, a process of proteomic diversification. Exons 5a and 5b are subjected to mutually exclusive AS that results in two splice variants, SNAP25a and SNAP25b (Letunic *et al.,* 2002). Several mutations affecting SNAP25b isoform have been reported in patients with neurological illnesses including myasthenia, intellectual

disability, and cerebral ataxia (Engel *et al.,* 2018; Shen *et al.,* 2014; Fukuda *et al.,* 2018). Snap25 knockout mice revealed that the Snap25b isoform shows a higher capability in stabilizing primed vesicles than Snap25a (Sørensen *et al.,* 2003). Transgenic mouse models have revealed that the splicing ratio of exons 5a and 5b of Snap25 is developmentally regulated and decreased expression of 5b resulted in developmental defects, spontaneous seizures, and impaired short-term synaptic plasticity (Johansson *et al.,* 2008). From iCLIP study Gehman *et al.,* 2011 reported that Rbfox1 has many (U)GCAUG motifs adjacent to exon 5b suggesting a possible splicing regulation of this exon by Rbfox1.

To specifically explore the splicing of exon 5a and 5b, a minigene pSP-Snap25-5a/b containing 2625 nt of genomic region encompassing exon 5a and 5b from the Snap25 gene (Fig. 11A) was generated and tested in minigene splicing assays with various splicing factors including Nova1, Nova2, Rbfox1, Ptbp2, and Mbnl2. All 5 of these splicing factors were cloned in pCAGGS-3x-Flag vector. Since exon 5a and 5b are identical in size (118 nt), it is not possible to distinguish them by gel electrophoresis after semi-quantitative RT-PCR. Interestingly, all constructs promote inclusion of either 5a or 5b or maybe both (Fig. 11B). To determine which exon is spliced in in the RNA transcript, we digested the RT-PCR products with restriction enzymes that are specific for each alternative exon: an NdeI restriction site present in exon 5a, and an AvrII restriction site present in exon 5b. Digestion of the RT-PCR products with NdeI generates two fragments of 194 and 186 bp if isoform 5a is present, while digestion of the RT-PCR products with AvrII generates two fragments of 211 and 171 bp if isoform 5b is present. By observing the size of bands on a gel, we confirmed which exon was regulated by

which splicing factor. Running the digested RT-PCR transcripts on gel, we discovered that Nova2 and Rbfox1 strongly promote inclusion of exon 5b (Fig. 11D, E), while Nova1 weakly regulates inclusion of both exons (Fig. 11C). Our observation for Ptbp2 and Mbln2 is slightly different: Ptbp2 promotes only the inclusion of exon 5a (Fig. 11F), while Mbln2 regulates inclusion of both exons with a preference for the inclusion of exon 5a over 5b (Fig. 11G).

Moreover, to assess the co-regulatory control of splicing at the 5a/b site, we performed competition experiments where we co-transfected mSnap25 minigene with 2 different splicing factors in HEK293T cells. For example, a varying amount of mRbfox1 (0x, 1x, 3x) was used with a constant amount of mPtbp2 (1x), and *vice versa*, while the amount of minigene was always constant (1x). We digested the RT-PCR products with NdeI and AvrII and overserved splicing regulation on agarose gel. We observed that there is an antagonistic effect between mRbfox1 and mPtbp2. However, mRbfox1 has strong preference in recognizing and promoting the inclusion of 5b in the presence of mPtbp2. We detected dose-dependent inclusion of 5b and skipping of 5a when co-transfection was carried out with varying amounts of mRbfox1 and constant amount of mPtbp2 (Fig. 12A). On the other hand, co-transfection with varying amounts of mPtbp2 and constant amount of mRbfox1 resulted in an opposite scenario. In this case, we observed a dose-dependent inclusion of 5a and skipping of 5b; interestingly, expression of 5b is much stronger than 5a (Fig. 12B). We have not observed any effect when mNova1 was co-transfected with varying amounts of mRbfox1 (Fig. 12D). However, we found a dose-dependent skipping of 5b when co-transfection with constant mRbfox1 and varying amount of mNova1 was carried out (Fig. 12C). We have not observed any antagonistic

relationship between Nova2 and mRbfox1 but a cooperation. In both co-transfection experiments with constant amounts of mRbfox1 and varying amounts of mNova2, or constant amounts of mNova2 and varying amount of mRbfox1, we observed strong inclusion of 5b (Fig. 12E, F).

2.5 DISCUSSION

In vertebrates Z^+ Agrin-mediated clustering of AChRs at the NMJ is critical for synaptogenesis and muscle stability (Bezakova *et al.,* 2003). However, in invertebrates (the nematode *C. elegans*) the expression of Agrin is very weak and transient in motoneurons in the embryo but not detected in larvae nor in adult stages and its protein sequence best aligns with Z^- Agrin (Hrus *et al.,* 2007), suggesting an Agrin-independent NMJ formation in nematodes. Moreover, as they could not detect any potential alternative exons coding for amino acids resembling the conserved inserts in the Z sites of vertebrates, the authors who cloned and characterized Agrin from *C. elegans* concluded that the "Z alternative spice sites are specific to vertebrates".

A recent study has shown that heparan sulfate -a proteoglycan- is critical for synaptogenesis in *C. elegans* (Lázaro-Peña *et al.,* 2018). Contrary to the findings in *C. elegans*, we have discovered that the Z alternative splice sites are not specific to vertebrates but are also present in invertebrates such as *Ciona*. In mammals, the Z exons of Agrin -termed Z8 and Z11- encode for 8 and 11 AA peptides (Gesemann *et al.,* 1995), respectively, but in *Ciona* the Z exons are even smaller: called Z6 and Z5, they encode for only 6 and 5 AAs, respectively (Fig. 2A). The presence of Z exons in *Ciona* suggests an Agrin/Nova-dependent pathway for NMJ formation, development, and maintenance that is conserved from tunicates to mammals. Using CRISPR/Cas-9 technology, our collaborators Dr. Lionel Christiaen at NYU and Dr. Alberto Stolfi at Georgia Tech generated Agrin KO and Nova KO *Ciona* animals and found that AChRs clustering at the NMJ is significantly reduced in KO animals (data not shown). This is consistent with

findings in Agrin null mice (Gautam *et al.,* 1996; Burgess *et al.,* 1999) as well as in dKO Nova mice (Ruggiu *et al.,* 2009).

The expression of Z^+ Agrin and Nova in *Ciona* is developmentally regulated. We observed a strong expression of Z5 and Z11 in the adult brain and 22.5 hpf at 20° C (swimming larvae) of *Ciona*, respectively. In mammals, the Z8 isoform is the most critical one for formation and development of the NMJ, as it is the most potent in promoting clustering of AChRs (Gesemann *et al.,* 1995; Ruggiu *et al.,* 2009). Interestingly, we detected robust expression of Z^- Agrin in unfertilized eggs (Fig. 3), suggesting that in tunicates Agrin may be maternally expressed and its mRNA deposited in the oocyte, and as such it may be critical for early embryonic development before activation of the zygotic genome. Nova's expression was consistent throughout the development but Z^+ Agrin first appears at 10 hpf during embryonic development, suggesting that Nova alone is not sufficient to induce inclusion of Agrin's Z exons, and that additional layer(s) of complexity in Z^+ Agrin expression that are independent of Nova are likely at play. At NYU in Dr. Christiaen's Lab, fluorescent in situ hybridization (FISH) was performed to validate the expression of Agrin and Nova in *Ciona*. We observed a strong expression of Agrin in the brain and the motor ganglion (MG) of *Ciona* larvae (data not shown). We also discovered that Nova's expression is dynamic throughout development: the expression of Nova is very weak at gastrula stage but is expressed throughout the body (strongest in the brain and MG) at mid-tailbud stage (data not shown), including heart precursor cells, suggesting a possible, yet unknown role for Nova and AS regulation in heart development. Interestingly, a recent paper showed that Agrin is necessary for heart regeneration following myocardial infraction, although

whether AS in general and Z exons in particular are involved in this process is not known. Moreover, we found a dynamic expression of Nova in motor neurons (data not shown): at 15.5 hpf, Nova's expression is upregulated in motor neuron (MN) 2 (data not shown), interestingly, one hour later at 16.5 hpf, Nova's expression disappears in MN2 but is upregulated in MN1 and interneuron1 (data not shown).

We discovered that in *Ciona* Nova uses its first two KH RNA-binding domains to splice Agrin, while KH3 is dispensable. In mammals, Nova uses its KH3 domain for binding to its target pre-mRNAs (Jensen *et al.,* 2000). Moreover, by generating N- and C-terminal we have discovered a unique function of KH3 domain of CiNova deletion constructs. Deletion of either N- or C-terminal or both of them together completely abolishes splicing of CiAgrin's Z exons. Since the effect on splicing is the same, this result suggests that the N-terminus and the C-terminus of Nova are part of the same regulatory domain. Intriguingly, deletion of the KH3 domain rescued the splicing defect of either the N- or the C-terminus deletions, or both. Our findings uncover previously unknown complex intramolecular regulatory elements that modulate Nova's splicing activity. Our interpretation of our result is as follows: The KH3 domain acts as a negative regulator of splicing, while the N- and C-terminus are part of a regulatory domain that negatively regulates the activity of KH3. Deletion of the N- and/or the C-termini eliminates the domain that suppresses the activity of KH3. As KH3 in turn acts to suppress the splicing activity of KH1 and KH2, this explains why a version of Nova that lacks its N- and/or C-terminus can no longer splice Agrin even though the KH1 and KH2 domains of Nova -that is, the two domains that are necessary for Agrin's splicing- are still intact. It's only by deleting the N- and/or C-terminal domains that act as repressors of

KH3 function that the inhibitory activity of KH3 on KH1 and KH2 is unmasked. We speculate that this N- and C-terminus domain may act as a regulatory switch that determines the usage of specific KH domains thus allowing Nova to switch between two distinct splice modalities: one that is KH1- and KH2-dependent and promotes the splicing of targets such as Agrin, and a second modality that is KH3-dependent and may regulate splicing of a different set of targets. The overall picture is likely to be more complex, and it may require further crystallographic and bioinformatical approaches to understand how different domains of Nova modulate its splicing activity. Nevertheless, our findings uncover previously unsuspected layers of regulation of Nova's splicing activity.

We have determined that CiNova regulates AS by specifically recognizing its *bona fide* binding sequence YCAY on CiAgrin's pre-mRNA. It is known that Nova promotes exon inclusion by binding to intronic YCAY clusters downstream of alternative exons; binding to exonic clusters, on the other hand, promote skipping (Ule *et al.,* 2006). We have discovered two Nova-dependent intronic splicing enhancers (NISE) elements (termed NISE1 and NISE2) in the intron downstream of Z5. In doing so we also discovered that at least two, and any two, consecutive YCAY sequences from either NISE1 or NISE2 are needed to promote Z exons inclusion (Fig. 8), thus contributing to deciphering the splicing code that mediates Nova-dependent AS regulation. The YCAY repeats on Nova targets are well conserved in mammals and closely spaced being on average 28 nt apart (Ule *et al.,* 2006). In *Ciona* Agrin the average nt distance between the intronic YCAY repeats is 32 when considering the whole intron, but the average nt distance is reduced to 10 and 22 nt in NISE1 and NISE2, respectively. However, the spacing between single YCAY sequences does not appear to be critical for splicing (Fig.

8F) For example, the closest YCAYs in NICE1, YCAY 6 and 7, are just 2 nt apart and double mutant YAAY 6-7 is still able to splice correctly. Conversely, YCAYs 12 and 13 in NICE2 are also just 2 nt apart, but double mutant YAAY12-13 is unable to splice. At the other end of the spectrum, YCAYs 13 and 14 are 56 nt apart, and double mutant YAAY 13-14 is unable to splice. YCAYs 17 and 18, on the other hand, are 111 nt apart but double mutant YAAY 17-18 is still able to splice correctly. Based on our findings, we propose a model for Nova-dependent splicing of neural Agrin where Nova uses its KH1 and KH2 domains to recognize and bind to NISE1 and NISE2 elements on Agrin's pre-mRNA, respectively or *vice versa*, and the N- and C-terminus act as a regulatory switch to negatively regulate the inhibitory function of KH3 domain (Fig. 10).

2.6 CONCLUSIONS AND FUTURE PERSPECTIVES

In summary, the findings in this chapter provide enough evidence of coevolution between Nova proteins and the cis-regulatory sequences embedded in the downstream intron of Z exons of Agrin. We specifically discovered the regulatory mechanism of Nova-dependent AS of Z exons of Agrin in *Ciona robusta*. We show that mutation in the cis- and trans-regulatory elements of Agrin and Nova could potentially dysregulate AS, which in turn could lead to disorders, particularly neurodegenerative and neuromuscular diseases. Since Nova/Agrin function is critical for NMJ formation, development, and maintenance; our studies here could be relevant to shed light to the understanding AS related diseases including cancer and neurological disorders.

Besides the major accomplishments described in this project, our study has also raised a number of questions that can potentially lead to new findings. For example, mSnap25 is differentially spliced at 5a/b site by 5 different splicing factors including mNova1 and Nova2, mRbfox1, mMbnl2, and mPtbp2. It would be interesting to explore the molecular mechanism of competition between the splicing factors that we see in our study. Finding the answer to the question “How and why Rbfox1 regulate splicing of Z exons of Agrin?” would be another exciting project to investigate.

2.7 TABLES AND FIGURES

2.7.1 TABLES

Table 1. List of all the forward and reverse primers used to clone *Ciona* Agrin constructs used in this study

Construct Name	Forward(F)/Reverse(R) Promers	Primer Sequence
pCi-CiAgrin-E40-41	Ci Agr_E40F1	CAAGATGAACAAGCGACTGC
	Ci Agr_E41R1	GGACCAGAACCAAGGTCAAA
pCi-CiAgrin-E40-YAAY	CiNova E40_YAAY_F	CAAGCGACTGCAATTTATTTGGATG
	CiNova E40_YAAY_R	TTCATCTTGTAGCAAGTCAG
pCi-CiAgrin-E41-YAAY1-2	CiNova E41_YAAY_1-2_F	ACTGCTAATGGTGGGCAAAGCGAGGGA
	CiNova E41_YAAY_1-2_R	AGACCGTTACGAGCAGTAGTGCGGAAC
pCi-CiAgrin-I40-YAAY3-4	CiNova E41_YAAY_3-4_F	CGTCTTAATCTTCGTTTTGACCTTGGTTC
	CiNova E41_YAAY_3-4_R	ACCGTCATTAATTGCGAGTGCAATATAATC
pCi-CiAgrin-I40-YAAY1-3	CiNova I40_YAAY_1-3_F	TTAACATCGTGCAAATAATTAGTTAGTTTTAGTGTCATTTG
	CiNova I40_YAAY_1-3_R	ATTATATGTTGATGCAATTGCAGTTTAACCTAAACATTTCTTAC
pCi-CiAgrin-I40-YAAY4-7	CiNova I40_YAAY_4-7_F	ATTAATGTCAATAGGAATGTGCGCACGGGA
	CiNova I40_YAAY_4-7_R	TCATTACAAATTACACTAAAACTAACTAATGATTTGCACGATGTTAAATG
pCi-CiAgrin-I40-YAAY8-10	CiNova I40_YAAY_8-10_F	TTGTAATTTTGAATTAATGTCAATAGTGCTGTATGAAAAG
	CiNova I40_YAAY_8-10_R	ACTAACTGTTAAATATTATCACAAAAAATAGGTCTTTAAAAAATG
pCi-CiAgrin-I40-YAAY11-13	CiNova I40_YAAY_11-13_F	ATAATATTAATGCTAACATTGAGAATGCTG
	CiNova I40_YAAY_11-13_R	GTTTTAGATTAATTGACCATAAAAAATAAACATAAAG
pCi-CiAgrin-I40-YAAY14	CiNova I40_YAAY_14_F	CTCAAAGCAGTAATAAAAACATTG
	CiNova I40_YAAY_14_R	CAAGATGTAATATATTTCGAAATC
pCi-CiAgrin-I40-YAAY15-16	CiNova I40_YAAY_15-16_F	AAAACCCAACTAACTTATTGTGAGTCCAAC
	CiNova I40_YAAY_15-16_R	ACCTTAATTAGTTTTTAGAACCCATCTATATATAAAG
pCi-CiAgrin-I40-YAAY17	CiNova I40_YAAY_17_F	GTAGGCCAAATAATATAACTCTATAAC
	CiNova I40_YAAY_17_R	TAAGGTATTCTCTGGGGTTG
pCi-CiAgrin-I40-YAAY18	CiNova I40_YAAY_18_F	AATTTCAATATAATTTTTTTTTTGTTTAGGC
	CiNova I40_YAAY_18_R	TTTAGATTTAGGGTGTGTG
pCi-CiAgrin-I40-YAAY4-5	CiAgr I40_YAAY_4-5_F	TGTAATGAATTCATGTCCATAGG
	CiAgr I40_YAAY_4-5_R	AATTACACTAAAACTAACTAATGATTTG
pCi-CiAgrin-I40-YAAY5-6	CiAgr I40_YAAY_5-6_F	ATTAATGTCCATAGGAATGTGCGC
	CiAgr I40_YAAY_5-6_R	TCATTACAAATGACACTAAAACTAACTAATGATTTG
pCi-CiAgrin-I40-YAAY6-7	CiAgr I40_YAAY_6-7_F	TGTCAATAGGAATGTGCGCACGGGA
	CiAgr I40_YAAY_6-7_R	TTAATTCATGACAAATGACACTAAAACTAACTAATGATTTGC
pCi-CiAgrin-E41-YAAY1-2	CiAgr I40_YAAY_1-2_F	AACATATAATTTAACATCGTGCAAATCATTAG
	CiAgr I40_YAAY_1-2_R	GATGCAATTGCAGTTTAACCTAAACATTTCTTAC
pCi-CiAgrin-I40-YAAY2-3	CiAgr I40_YAAY_2-3_F	GTGCAAATAATTAGTTAGTTTTAGTGTCATTTG
	CiAgr I40_YAAY_2-3_R2	GATGTTAAATTATATGTTGATGCAATGGCAG
pCi-CiAgrin-I40-YAAY3	CiAgr I40_YAAY_3_F	ATCGTGCAAATAATTAGTTAGTTTTAG
	CiAgr I40_YAAY_3_R	GTTAAATGATATGTTGATGCAATG
pCi-CiAgrin-I40-YAAY12	CiAgr I40_YAAY_12_F	ATCTAAAACATAATATTCATGCTAAC
	CiAgr I40_YAAY_12_R	GAATTGACCATAAAAAATAAACATAAAG
pCi-CiAgrin-I40-YAAY12-13	CiAgr I40_YAAY_12-13_F	TTAATGCTAACATTGAGAATGCTG
	CiAgr I40_YAAY_12-13_R	TATTATGTTTTAGATGAATTGACCATAAAAAATAAAC
pCi-CiAgrin-I40-YAAY4	CiAgr I40_YAAY_4_F	AGTTTTAGTGTAATTTGTCATGAATTC
	CiAgr I40_YAAY_4_R	AACTAATGATTTGCACGATG
pCi-CiAgrin-I40-YAAY11	CiAgr I40_YAAY_11_F	ATGGTCAATTAATCTAAAACATCATATTC
	CiAgr I40_YAAY_11_R	AAAAAATAAACATAAAGAGTTTTTTTAAATTC
pCi-CiAgrin-I40-YAAY13	CiAgr I40_YAAY_13_F	ACATCATATTAATGCTAACATTGAG
	CiAgr I40_YAAY_13_R	TTTAGATGAATTGACCATAAAAAATAAAC
pCi-CiAgrin-I40-YAAY3-4	CiAgr I40_YAAY_3-4_F	TTTAGTGTAATTTGTCATGAATTCATGTCC
	CiAgr I40_YAAY_3-4_R	ACTAACTAATTATTTGCACGATGTTAAATG
pCi-CiAgrin-I40-YAAY11-12	CiAgr I40_YAAY_11-12_F	AACATAATATTCATGCTAACATTGAGAATG
	CiAgr I40_YAAY_11-12_R	TTAGATTAATTGACCATAAAAAATAAACATAAAG

Table 2. List of all the forward and reverse primers used to clone *Ciona* Nova constructs used in this study

Construct Name	Forward(F)/Reverse(R) Promers	Primer Sequence
pEGFP-CiNova_MMM	Ci Nova_MMM_F1_EcoRI	ACAGTGGAATTCTATGATGATGACGGCCGTAGTACC
	Ci Nova_R1_BamHI	ACAGTGGGATCCCTACAGTAACTTAGCCTGCTGTGC
pEGFP-CiNova_MEY	Ci Nova_MEY_F1_EcoRI	ACAGTGGAATTCTATGGAGTATGAATGCCAGTACAATGC
	Ci Nova_R1_BamHI	ACAGTGGGATCCCTACAGTAACTTAGCCTGCTGTGC
pEGFP-CiNova_MLN	Ci Nova_MLN_F1_EcoRI	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	Ci Nova_R1_BamHI	ACAGTGGGATCCCTACAGTAACTTAGCCTGCTGTGC
pEGFP-CiNova_MLN-KH1-GDDG	CiNova KH1_GDDG_F	GGTGATCGGGGACGACGGTCAGATTATTGTACAACTTCAGAAAGATTCAGGG
	CiNova KH1_GDDG_R	GCCCCCGCAGCGTACCCC
pEGFP-CiNova_MLN-KH2-GDDG	CiNova KH2_GDDG_F	GGTAATAGGAGACGACGGCGCAACGATAAAG
	CiNova KH2_GDDG_R	AGTCCCGCAGTTGTGTTG
pEGFP-CiNova_MLN-KH3-GDDG	CiNova KH3_GDDG_F	AGTCCTCGGAGACGACGGAAGGACACTG
	CiNova KH3_GDDG_R	GCTCCGATCAGGTTTTCG
pEGFP-CiNova_MLN-KH1/2-GDDG	Ci Nova_MLN_F1_EcoRI	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	CiNova KH2_GDDG_R	AGTCCCGCAGTTGTGTTG
pEGFP-CiNova_MLN-KH1/3-GDDG	Ci Nova_MLN_F1_EcoRI	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	CiNova KH3_GDDG_R	GCTCCGATCAGGTTTTCG
pEGFP-CiNova_MLN-KH2/3-GDDG	Ci Nova_MLN_F1_EcoRI	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	CiNova KH3_GDDG_R	GCTCCGATCAGGTTTTCG
pEGFP-CiNova_MLN-KH1/2/3-GDDG	Ci Nova_MLN_F1_EcoRI	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	CiNova KH3_GDDG_R	GCTCCGATCAGGTTTTCG
pEGFP-CiNova ΔN	CiNova ΔN_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova ΔN_R	ACAGTGGGATCCCTACAGTAACTTAGCCTGCTGTGC
pEGFP-CiNova ΔNΔC	CiNova ΔNΔC_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova ΔNΔC_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA
pEGFP-CiNova ΔNΔKH3	CiNova ΔNΔKH3_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova ΔNΔKH3_R	ACAGTGGGATCCCTACAGTAACTTAGCCTGCTGTGC
pEGFP-CiNova ΔNΔKH3ΔC	CiNova ΔNΔKH3ΔC_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova ΔNΔKH3ΔC_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA
pEGFP-CiNova MLN ΔC-ter	CiNova MLN ΔC-ter_F	ACAGTGGAATTCTATGCTAAATGCAATGGAGTATGAATGCCAGTACAATGC
	CiNova MLN ΔC-ter_R	ACAGTGGGATCCCTACGTAATAAGAAACTGCGCAGTCTGT
pEGFP-CiNova MLN ΔKH3	CiNova MLN ΔKH3_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova MLN ΔKH3_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA
pEGFP-CiNova MLN ΔKH3ΔC	CiNova MLN ΔKH3ΔC_F	ACAGTGGAATTCTATGATTCTTAAAGTTCTAATACCGGGGTACGC
	CiNova MLN ΔKH3ΔC_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA
pEGFP-CiNova MMM ΔC-ter	CiNova MMM ΔC-ter_F	ACAGTGGAATTCTATGATGATGACGGCCGTAGTACC
	CiNova MMM ΔC-ter_R	ACAGTGGGATCCCTACGTAATAAGAAACTGCGCAGTCTGT
pEGFP-CiNova MLN ΔKH3ΔC	CiNova MLN ΔKH3ΔC_F	ACAGTGGAATTCTATGATGATGACGGCCGTAGTACC
	CiNova MLN ΔKH3ΔC_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA
pEGFP-CiNova MMM ΔKH3ΔC	CiNova MMM ΔKH3ΔC_F	ACAGTGGAATTCTATGATGATGACGGCCGTAGTACC
	CiNova MMM ΔKH3ΔC_R	ACAGTGGATCCCTAGCTTGACTTTTCGATGCTTAGGATACTCA

Table 3. List of all the forward and reverse primers used to clone mouse Nova and Agrin constructs used in this study

Construct Name	Forward(F)/Reverse(R) Promers	Primer Sequence
pEGFP-mNova1	mNova1_F_HindIII	ATGCTCAAGCTTCGATGATGGCGGCAGCTCCCATTC
	mNova1_R_KpnI	ATGCTCGGTACCTCAACCCACTTTCTGAGGATTGGCA
pEGFP-mNova2	mNova2_F_EcoRI	AGCTTCGAATTCTATGGAGCCCGAGGCCCCGG
	mNova2_R_BamHI	ATGGTCGGATCCTCATCCCACTTTCTGTGGGTTTGAAGCCCTCC
pEGFP-mNova1-KH1-GDDG	mN1 KH1_GDDG_F	TATAATTGGGGACGACGGACAGACAATTGTTCAG
	mN1 KH1_GDDG_R	GATCCAGCAGCATAACTAG
pEGFP-mNova1-KH2-GDDG	mN1 KH2_GDDG_F	GATAATAGGGGACGACGGTGCTACTGTGAAGGC
	mN1 KH2_GDDG_R	AGACCTGCTGTGCTGTTG
pEGFP-mNova1-KH3-GDDG	mN1 KH3_GDDG_F	AATACTTGGCGACGACGGGAAAACCTTAGTG
	mN1 KH3_GDDG_R	GCACCAACTAAGTTTTCTG
pEGFP-mNova1-KH1/2-GDDG	mNova1_F_HindIII	ATGCTCAAGCTTCGATGATGGCGGCAGCTCCCATTC
	mN1 KH2_GDDG_R	AGACCTGCTGTGCTGTTG
pEGFP-mNova1-KH1/3-GDDG	mNova1_F_HindIII	ATGCTCAAGCTTCGATGATGGCGGCAGCTCCCATTC
	mN1 KH3_GDDG_R	GCACCAACTAAGTTTTCTG
pEGFP-mNova1-KH2/3-GDDG	mNova1_F_HindIII	ATGCTCAAGCTTCGATGATGGCGGCAGCTCCCATTC
	mN1 KH3_GDDG_R	GCACCAACTAAGTTTTCTG
pEGFP-mNova1-KH1/2/3-GDDG	mNova1_F_HindIII	ATGCTCAAGCTTCGATGATGGCGGCAGCTCCCATTC
	mN1 KH3_GDDG_R	GCACCAACTAAGTTTTCTG
pEGFP-mNova2-KH1-GDDG	mN2 KH1_GDDG_F	AATCATCGGCGACGACGGCCAGACCATCCAAGCTTCG
	mN2 KH1_GDDG_R	GAGCCGGCGGCGTAGCTG
pEGFP-mNova2-KH2-GDDG	mN2 KH2_GDDG_F	GATCATCGGTGACGACGGAGCGACCGTGAAGG
	mN2 KH2_GDDG_R	AGTCCTGCCGTGCTGTTG
pEGFP-mNova2-KH3-GDDG	mN2 KH3_GDDG_F	CATCCTGGGCGACGACGGCAAGACGCTGG
	mN2 KH3_GDDG_R	GCCCCCACCAGGTTCTCG
pEGFP-mNova2-KH1/2-GDDG	mNova2_F_EcoRI	AGCTTCGAATTCTATGGAGCCCGAGGCCCCGG
	mN2 KH2_GDDG_R	AGTCCTGCCGTGCTGTTG
pEGFP-mNova2-KH1/3-GDDG	mNova2_F_EcoRI	AGCTTCGAATTCTATGGAGCCCGAGGCCCCGG
	mN2 KH3_GDDG_R	GCCCCCACCAGGTTCTCG
pEGFP-mNova2-KH2/3-GDDG	mNova2_F_EcoRI	AGCTTCGAATTCTATGGAGCCCGAGGCCCCGG
	mN2 KH3_GDDG_R	GCCCCCACCAGGTTCTCG
pEGFP-mNova2-KH1/2/3-GDDG	mNova2_F_EcoRI	AGCTTCGAATTCTATGGAGCCCGAGGCCCCGG
	mN2 KH3_GDDG_R	GCCCCCACCAGGTTCTCG

Table 4. List of all the forward and reverse primers used to clone other minigene and cDNA constructs used in this study

Construct Name	Forward(F)/Reverse(R) Promers	Primer Sequence
mSnap25 minigene	mSnap25_I4F1_XhoI	ACAGTGCTCGAGCTTGCAGTTTCCCCAACTTGGT
	mSnap25_I5R1_BamHI	ACAGTGGGATCCATCTGAGCGACTGCTTCCTGTTAG
mPtbp2	mPtbp2_F_AgeI	ACAGTGACCGGTATGGACGGAATTGTCACTGAGGT
	mPtbp2_R_SalI	ACAGTGGTCGACTTAGATTGTTGACTTGGAGAAAGACACTCTCAG
mRbfox1	RbFox1_F1_XhoI	ACAGTGCTCGAGCTAATTGTGAAAGAGAGCAGCTGAGG
	RbFox1_R1_SalI	ACAGTGGTCGACTTAAGTGGCACCAACGCCG
mMbnl2	Mbnl2_F1_XhoI	ACAGTGCTCGAGCTGCCTTGAACGTTGCCCC
	Mbnl2_R1_SalI	ACAGTGGTCGACTTACTTAAGTTTCAGAATTATCTGATTGGCTGTGG

Table 5. List of all the forward and reverse primers used to detect spliced isoforms of Agrin, Snap25, Dcc minigenes by semi-quantitative RT-PCR analysis

Minigenes	RT-PCR primers(F-Forward, R=Reverse)	Sequence	PCR cycles	Anealing temperature
Ciona Agrin	pCI_RT_F	GTGTCCACTCCCAGTTCAATTACAG	27	60°C
	pCI_RT_R	TGTCTGCTCGAAGCATTAACCC		
Mouse Snap25	V1_F	TCTGAGTCACCTGGACAACC	35	55°C
	V2_R	ATCTCAGTGGTATTTGTGAGC		
Mouse Agrin	mAgr_31F2	TTTGATGGGCGGACCTACATCG	30	60°C
	3xFLAG_R	GCCGTCGTGGTCTTTGTAGTCTCTA		
Mouse Dcc	mDcc_E16F1	TCTCATTATGTAATCTCCTTAAAAGC	30	52°C
	mDcc_E17R1	CTGCCCAGCTGACCCTCACAG		

Table 6. Co-transfection protocol of minigene and splicing factor cloned in pCi-neo and pEGFP-C1 vector respectively

pCi-neo_minigene (µL)	1x = 0.5 µg	1x = 0.5 µg	1x = 0.5 µg
pEGFP_Splicing factor (µL)	0x = 0.0 µg	1x = 0.5 µg	4x = 2.0 µg
pEGFP-C1_Empty vector (µL)	4x = 2.0 µg	3x = 1.5 µg	0x = 0.0 µg
Total DNA (µg)	2.5 µg	2.5 µg	2.5 µg
PEI (µL)	7.5 µL	7.5 µL	7.5 µL
Opti-MEM (µL)	Upto 200 µL	Upto 200 µL	Upto 200 µL

Table 7. Co-transfection protocol of minigene and splicing factor cloned in pSPL3 and pCAGGS-3x-Flag vector respectively

pSP_minigene (µL)	1x = 0.5 µg	1x = 0.5 µg	1x = 0.5 µg
pCAGGS_Splicing factor (µL)	0x = 0.0 µg	1x = 0.5 µg	4x = 2.0 µg
pcDNA3_Empty vector (µL)	4x = 2.0 µg	3x = 1.5 µg	0x = 0.0 µg
Total DNA (µg)	2.5 µg	2.5 µg	2.5 µg
PEI (µL)	7.5 µL	7.5 µL	7.5 µL
Opti-MEM (µL)	Upto 200 µL	Upto 200 µL	Upto 200 µL

2.7.2 FIGURES AND FIGURE LEGENDS

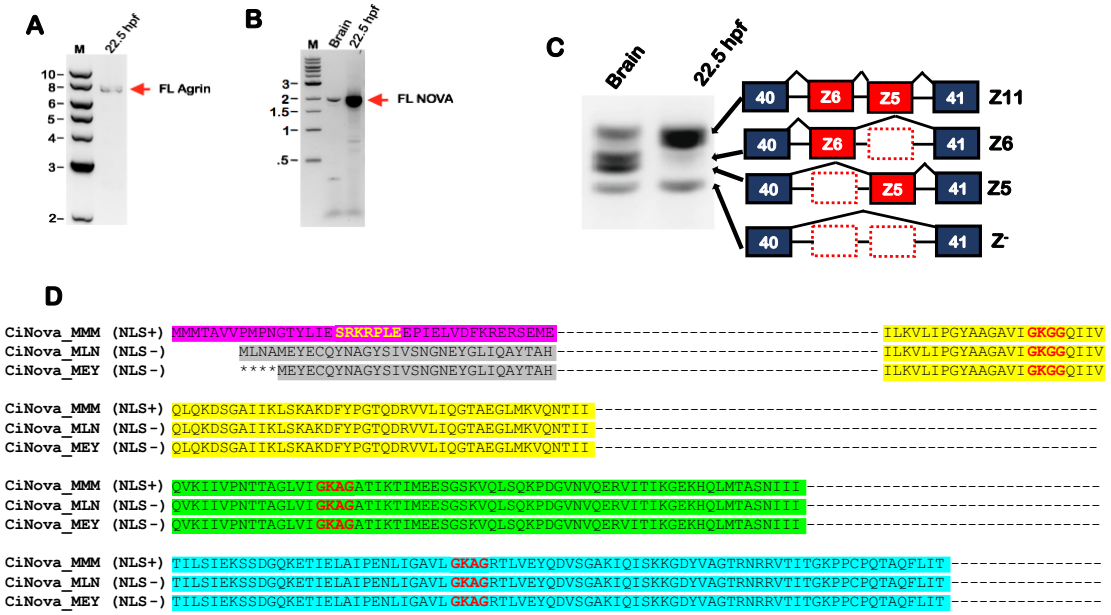


Figure 1. Cloning and characterization of Nova and Agrin from *Ciona robusta*.

(A, B) Full-length (FL) Nova and Agrin from larvae at 22.5 hpf at 20° C and adult brain were detected by semi-quantitative RT-PCR. FL Agrin and Nova are approximately 8 kilobases and 2 kilobases, respectively. (C) Detection of all the possible Z isoforms (Z^+ and Z^-) of Agrin from *Ciona* adult brain and larvae at 22.5 hpf at 20° C by semi-quantitative RT-PCR. (D) Amino acid (AA) sequence alignment of different CiNova isoforms; the isoforms are named after the first 3 AA. Alternative first exon 1a usage by CiNova_MMM is highlighted as purple with NLS bolded and shown in yellow. Usage of alternative exon 1b by CiNova_MLN and CiNova_MEY is highlighted as light gray; both isoforms lack a canonical NLS. CiNova_MLN and CiNova_MEY use different starting AUG codons from the same pre-mRNA. Asterisks represent missing AA from CiNova MEY and hyphens represent sequence identity beside KH domains. The KH1, KH2, and KH3 domains are highlighted as yellow, green, and turquoise, respectively. The GXXG motifs present in each KH domain are bolded and shown in red.

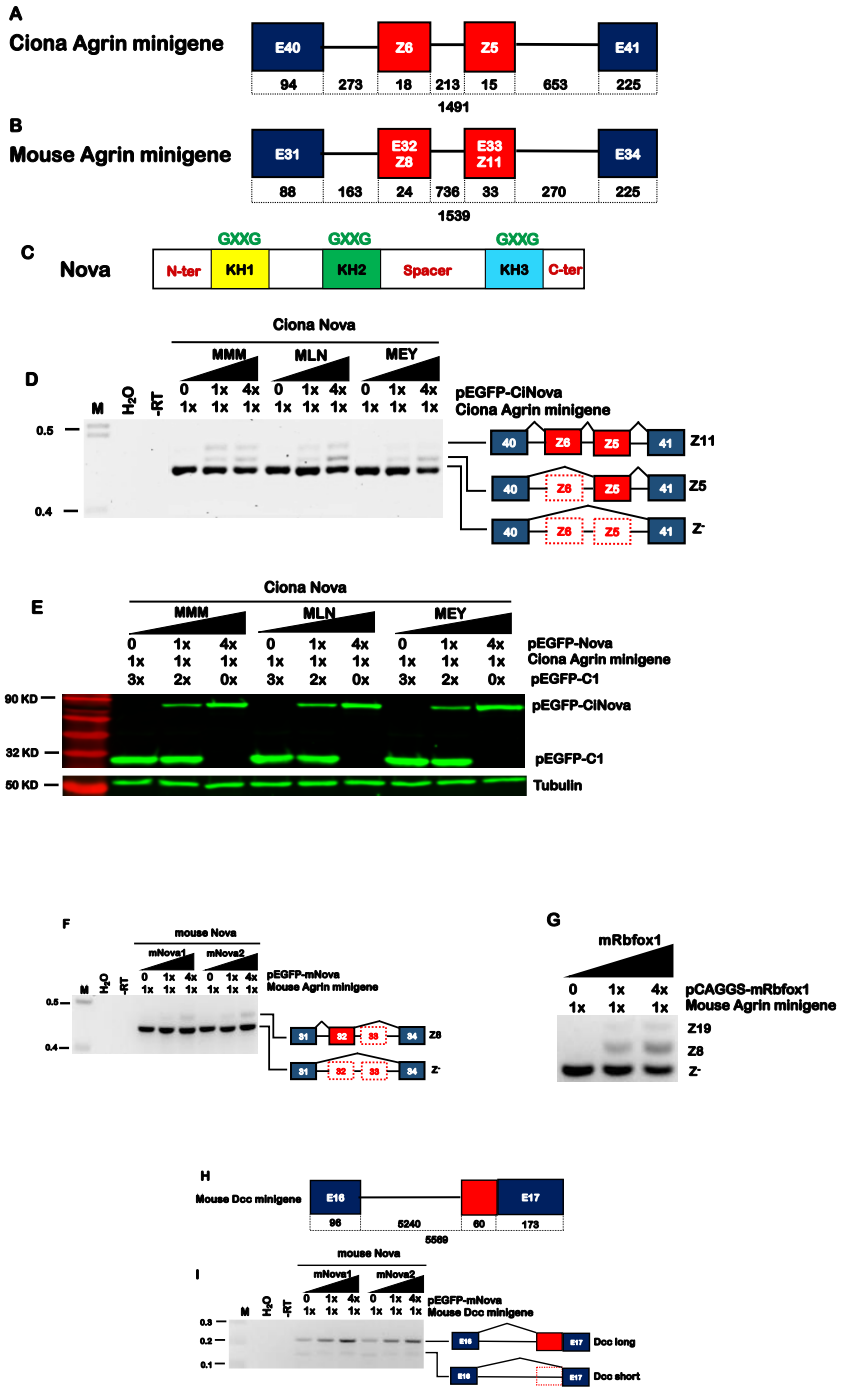


Figure 2. Nova promotes inclusion of Agrin's Z exons in a dose-dependent manner.

(A, B) Schematic illustration of *Ciona* Agrin minigenes containing the genomic region between exons 40 and 41 in *Ciona*, and exons 31 and 33 in mouse. The constitutive exons

(exon 40 and 41 in *Ciona*; exon 31 and 34 in mouse) are shown in dark blue, the alternative Z exons (Z6 and Z5 in *Ciona*; exon 32/Z8 and 33/Z11 in mouse) are shown in red. Thick lines between exons represent introns. Intron and exon sizes are shown in nucleotides. (C) Schematic representation of Nova protein structure. Nova harbors 3 KH RNA-binding domains and a spacer sequence in between KH2 and KH3. N/C-terminus and GXXG motifs in each KH domain are also shown. (D) CiAgrin minigene splicing assay. Constant amounts of *Ciona* Agrin (0.5 μg/well; 1x) minigene were co-transfected in HEK293T cells in 6-well plates with increasing amounts of EGFP-CiNova (0 μg = 0x, 0.5 μg = 1x, 2 μg = 4x) constructs (MMM, MLN, and MEY). Total RNA was extracted 48h after transfection and subjected to semi-quantitative RT-PCR analysis. All 3 CiNova isoforms are able to promote inclusion of CiAgrin's Z exons in a dose-dependent manner, giving rise to Z11 (Z6+Z5) and Z5 splice forms. The identity of the RT-PCR products was confirmed by cloning and direct sequencing. (E) The expression of all 3 EGFP-CiNova protein was detected by western blotting. All 3 EGFP CiNova constructs were co-transfected with CiAgrin minigene in HEK293T cells as in D. The anti-EGFP antibody recognizes both EGFP-CiNova and EGFP from EGFP-C1 empty vector. All 3 EGFP-CiNova proteins were robustly expressed in a dose-dependent manner. (F) AS of mouse Agrin minigene and 2 cDNA constructs of mNova. Only Z8 isoform of mouse Agrin was detected by semi-quantitative RT-PCR when mouse Agrin minigene was individually co-transfected with mNova1 and mNova2. mNova promotes inclusion of Z8 exon of mouse Agrin in a dose-dependent manner. (G) mRbfox1 strongly promotes inclusion of Z exons. Z8 and Z19 isoforms of Z^+ Agrin were detectable by RT-PCR. (H) Schematic representation of the genomic region of Dcc between exons 16 and 17 used to generate a minigene construct. Intron and exon sizes are shown in nucleotides. Constitutive exons are shown as dark blue boxes, while an alternative version of exon 17 is shown as a red box. Use of an alternative 3' splice site at exon 17 gives rise to a longer version of Dcc. Thick lines between exons represent introns. (I) Total RNA from HEK293T cells co-transfected with mDcc minigene and mouse Nova1 or Nova2 was extracted and subjected to semi-quantitative RT-PCR analysis. Both Nova1 and Nova2 promote the usage of the distal 3' splice site and the generation of the Dcc long isoform in a dose-dependent manner.

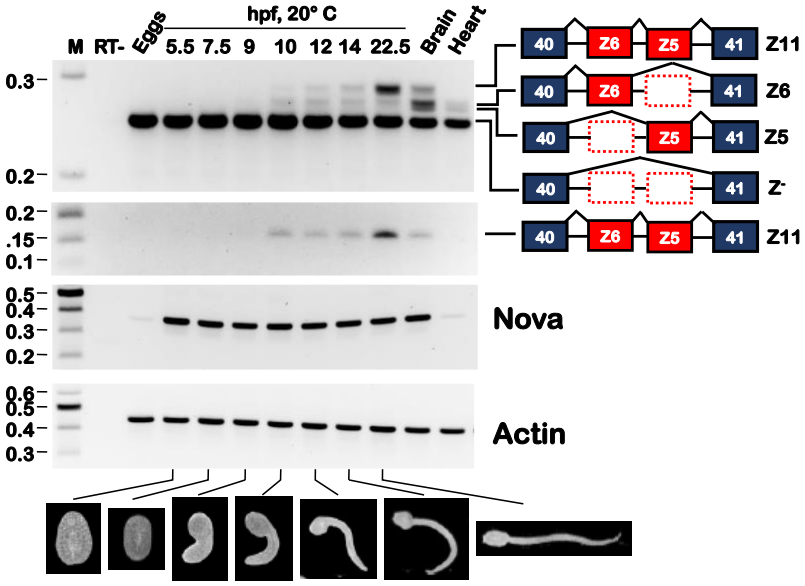


Figure 3. Expression of CiAgrin's Z isoforms is developmentally regulated.

Total RNA from unfertilized eggs and embryos at different developmental stages (hpf at 20° C), plus brain and heart from adult animals, was extracted and subjected to semi-quantitative RT-PCR analysis using gene-specific primers. Primers located in constitutive exon 40 and 41 surrounding the Z site detect all isoforms of CiAgrin. While Agrin is robustly expressed in all samples analyzed including unfertilized eggs, Z^+ isoforms start to be expressed at 10 hpf and later stages, including adult brain, with strongest expression in swimming larvae (top panel).This result was confirmed using primers that specifically detect the Z11 (Z6+Z5) isoform of CiNova, with strongest expression in swimming larvae (second panel). CiNova is expressed at low levels in fertilized eggs and adult heart, and robustly throughout *Ciona* larval development (third panel). Actin was used as internal control as it is robustly in all samples analyzed (fourth panel). 5.5 hpf at 20° C: Late gastrula/early neurula; 7.5 hpf at 20° C: Tailbud; 22.5 hpf at 20° C: Swimming larvae.

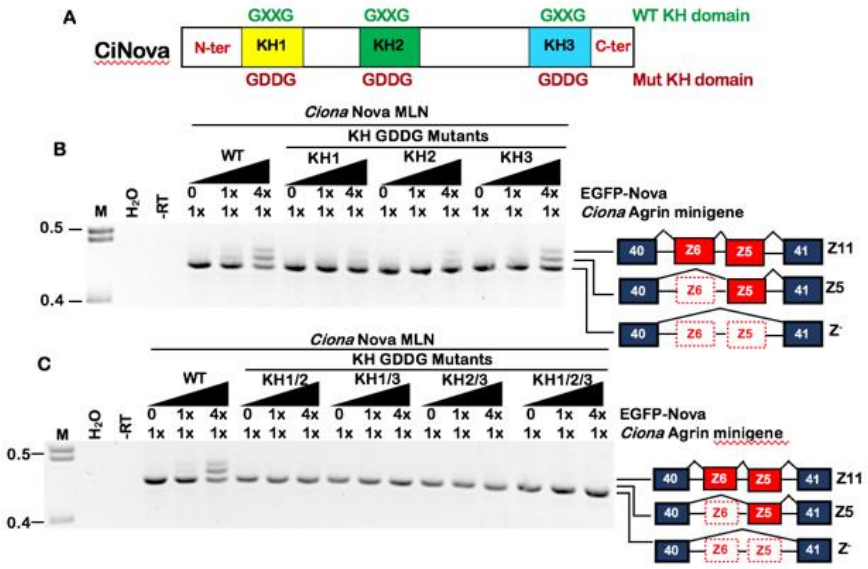


Figure 4. CiNova requires both KH1 and KH2 domains for splicing.

(A) Schematic representation of CiNova protein. Its 3 KH domains are labeled in different colors. GXXG motifs (green) present in WT protein were mutated to GDDG (dark red). (B) CiAgrin minigene splicing assays were performed in HEK293T cells as described in Figure 2, using WT and single KH GDDG mutants of CiNova. WT CiNova promotes inclusion of Agrin's Z exons, while KH1 and KH2 GDDG mutants failed to promote Z exons inclusion. KH3 GDDG mutant is indistinguishable from WT. (C) CiAgrin minigene splicing assay using WT and double/triple KH GDDG mutants. Double and triple KH GDDG mutants fail to promote inclusion of Agrin's Z exons.

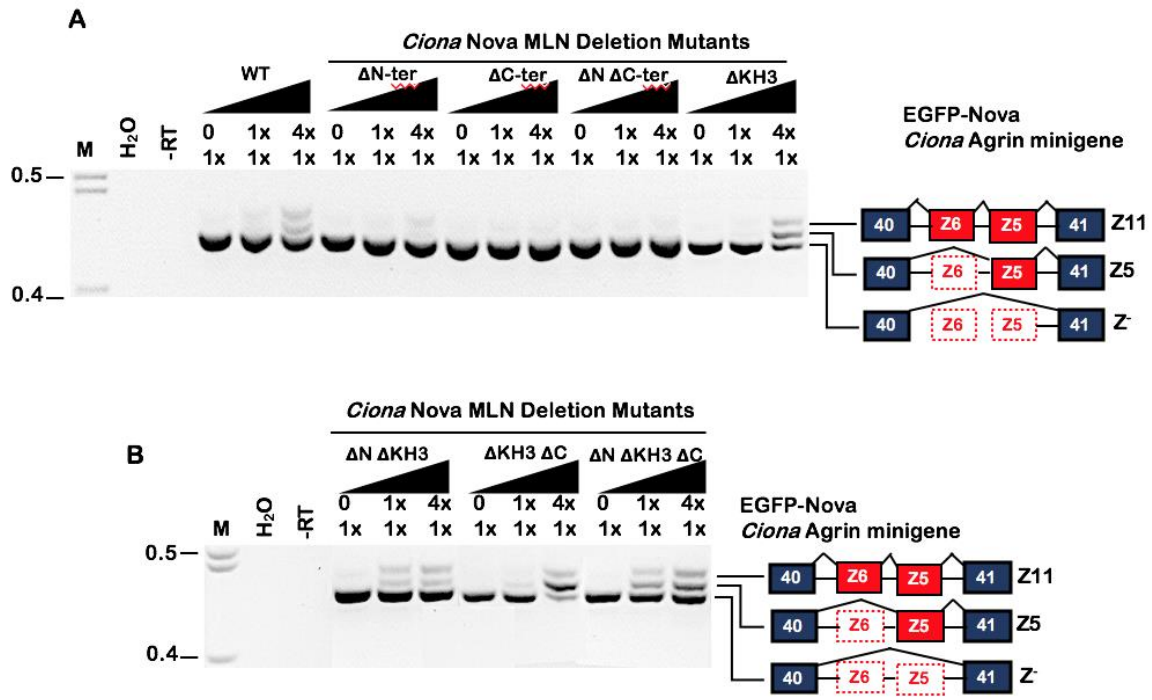


Figure 5. Novel function of N/C-terminals and KH3 domain of CiNova.

HEK293T cells were co-transfected with CiAgrin minigene and CiNova WT and 7different deletion mutants. (A) AS of WT CiNova and its 4 deletion mutants: ΔN-ter, ΔC-ter, ΔNΔC-ter, and ΔKH3. All these deletion mutants are unable to promote inclusion of Agrin's Z exons with the exception of ΔKH3 mutant, which is indistinguishable from WT. (B) Minigene splicing assay of WT CiNova and N/C-ter deletion mutants coupled with KH3 deletion. Surprisingly, deletion of KH3 rescues the splicing defects of N/C-ter deletion mutants.

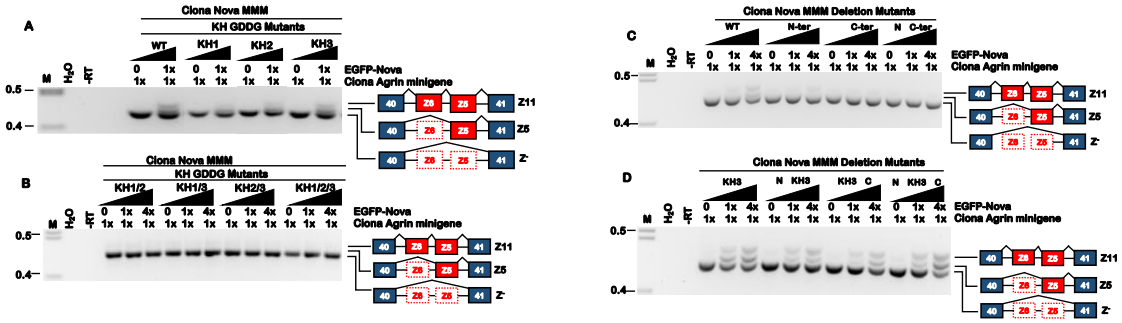


Figure 6. Alternative splicing of CiAgrin with WT and mutant CiNova_MMM constructs.

(A) CiAgrin minigene splicing assays were performed in HEK293T cells as described in Figure 2 (except 4x Nova transfection was omitted), using WT and single KH GDDG mutants of CiNova. Cells were co-transfected with constant amounts of *Ciona* Agrin minigenes and increasing amounts of CiNova (WT vs. KH GDDG mutants). Total RNA was extracted and subjected to semi-quantitative RT-PCR analysis. WT CiNova promotes inclusion of Agrin's Z exons, while KH1 and KH2 GDDG mutants failed to promote Z exons inclusion. KH3 GDDG mutant is indistinguishable from WT. (B) CiAgrin minigene splicing assay using WT and double/triple KH GDDG mutants. Double and triple KH GDDG mutants fail to promote inclusion of Agrin's Z exons. (C) Alternative splicing of WT CiNova_MMM and its 3 deletion mutants: ΔN-ter, ΔC-ter, and ΔNΔC-ter. All these deletion mutants are unable to promote inclusion of Agrin's Z exons. (D) Minigene splicing assay of KH3 deletion and N/C-ter deletion mutants coupled with KH3 deletion. KH3 deletion mutant is indistinguishable from WT in its ability to promote splicing of the Z exons. Deletion of KH3 rescues the splicing defects of N/C-ter deletion mutants.

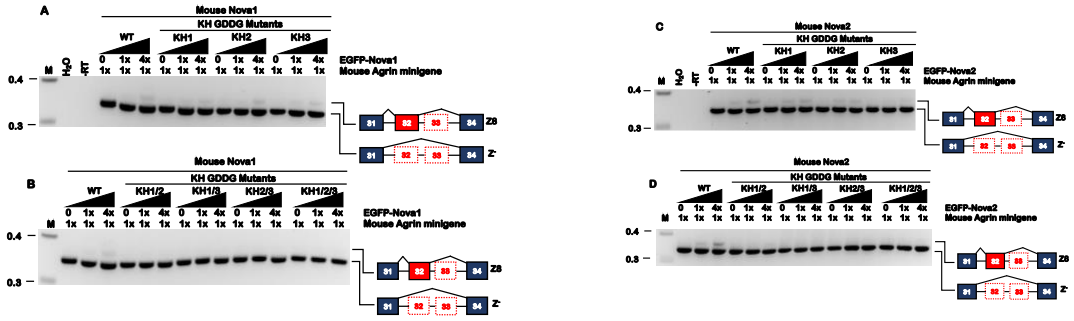
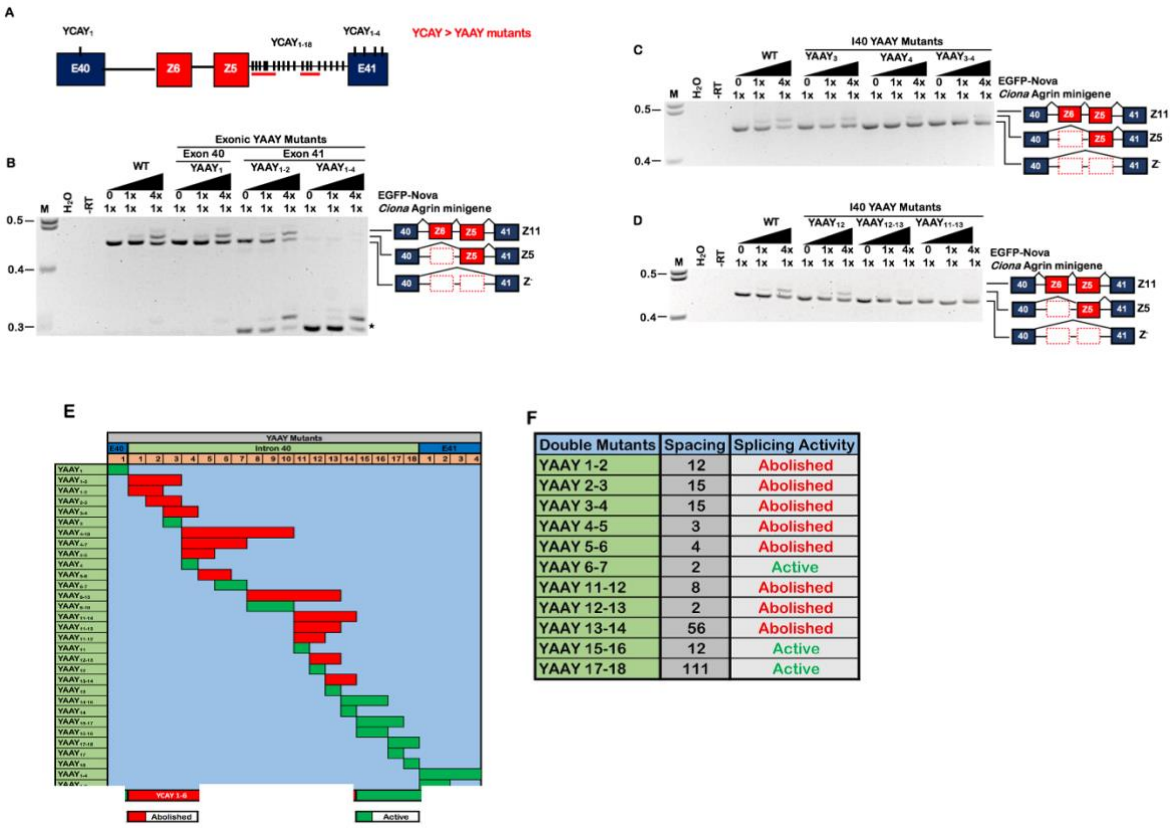


Figure 7. Mouse Nova2 requires its third KH3 domain to splice mouse Agrin.

(A, C) Mouse Agrin minigene splicing assays were performed in HEK293T cells as described in Figure 2, using WT and KH GDDG mutants of mNova1 (A, B) and Nova2 (C, D). Cells were co-transfected with constant amounts of mouse Agrin minigenes and increasing amounts of Nova (WT vs. KH GDDG mutants). Total RNA was extracted and subjected to semi-quantitative RT-PCR analysis. (A) The expected Z8 isoforms were detected when mouse Agrin minigene was co-transfected with WT and single KH GDDG mutants of Nova1. (B) Double and triple KH GDDG mutants of Nova1 completely abolish splicing of Z exons of mouse Agrin, suggesting that all 3 KH domains of Nova1 cooperate to splice Agrin. (C) The expected Z8 isoforms were detected when mouse Agrin minigene was co-transfected with WT and single KH1 and KH2 GDDG mutants of Nova2, while Nova2 KH3 GDDG mutant is unable to splice mouse Agrin. (D) The expected Z8 isoforms were detected when mouse Agrin minigene was co-transfected with WT Nova2, but double and triple KH GDDG mutants of Nova2 completely abolish splicing of Z exons of mouse Agrin, suggesting that Nova2 primarily uses KH3 to splice Agrin, but KH1 and KH2 also contribute to its splicing activity.



Double Mutants	Spacing	Splicing Activity
YAAY 1-2	12	Abolished
YAAY 2-3	15	Abolished
YAAY 3-4	15	Abolished
YAAY 4-5	3	Abolished
YAAY 5-6	4	Abolished
YAAY 6-7	2	Active
YAAY 11-12	8	Abolished
YAAY 12-13	2	Abolished
YAAY 13-14	56	Abolished
YAAY 15-16	12	Active
YAAY 17-18	111	Active

Figure 8. Exonic YCAY sequences are not required for Nova-dependent inclusion of Agrin's Z exons, while a bipartite Nova-dependent intronic splicing enhancer mediates inclusion of Agrin's Z exons.

(A) Schematic illustration of *Ciona* Agrin minigene containing the genomic region between exons 40 and 41. The constitutive exons (exon 40 and 41) are shown in dark blue color and the alternative Z exons (Z5 and Z6) are shown in red color. YCAY (Y=C/U) sequences (black bars indicate YCAY sequence) are *bona fide* Nova binding sequences and are exclusively located downstream of exon Z5 and within exons 40 and 41. Two clusters of YCAY sequences (red lines downstream of exon Z5) are the Nova binding sites and are called Nova intronic splicing enhancer 1 and 2 (NISE1, YCAY1-6 and NISE2, YCAY11-14. (B) CiAgrin minigene splicing assays were performed in HEK293T cells as described in Figure 2 using WT and exonic YAAY mutant CiAgrin minigenes and WT CiNova. Exonic YAAY mutations don't abolish splicing, suggesting that exonic YCAY sequences are not required for splicing. *: RT-PCR products form exon 41 mutants run aberrantly on agarose gel. (C, D) CiAgrin minigene splicing assays were performed using WT and 3 NISE1 (C) and NISE 2 (D) YAAY mutants with WT CiNova. For CiNova to effectively splice Z exons of CiAgrin, at least two consecutive YCAY sequences from each NISE clusters are needed. Single YAAY CiAgrin mutants are indistinguishable from WT, while double YAAY mutants are unable to include CiAgrin's Z exons. (E) Two NISE YCAY clusters are required for CiNova to promote splicing of CiAgrin's Z exons. Heat map represents the Nova-dependent splicing activity, where a green rectangle means splicing is active and a red rectangle means splicing is

abolished. X axis represents the position and number of YCAY sequences on exon 40, intron 40, and exon 41. Y axis represents all 32 YAAY mutant minigenes of CiAgrin (3 exonic and 29 intronic) generated in this study. Two NISE clusters (NISE1, YCAY1-6 and NISE2, YCAY11-14) of YCAY sequences are critical for CiNova-dependent splicing of CiAgrin's Z exons. Single YAAY mutations from either NISE clusters cannot disrupt splicing but two or more than two mutations in YCAY completely abolish splicing, suggesting that 2 is the minimum number of YCAY sequences required in each cluster to be splicing-competent. (F) Spacing between the YCAY sequences does not appear to be critical for splicing.

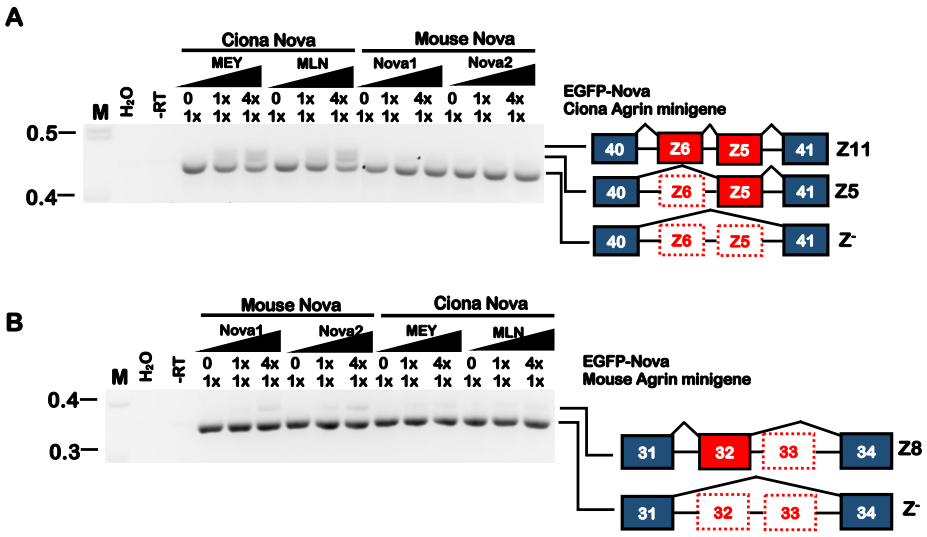


Figure 9. Splicing of Z exons of Agrin is species-specific.

(A) Total RNA from HEK293T cells co-transfected with CiAgrin minigene and Nova from either *Ciona* or mouse was extracted and subjected to semi-quantitative RT-PCR analysis as described in Figure 2. CiNova promotes inclusion of the Z exons of CiAgrin but mouse Nova cannot splice the Z exons of *Ciona* Agrin. (B) Total RNA from HEK293T cells co-transfected with mouse Agrin minigene and Nova from either mouse or *Ciona* extracted and subjected to semi-quantitative RT-PCR analysis. Mouse Nova promotes inclusion of the Z exons of mouse Agrin, while CiNova cannot splice the Z exons of mouse Agrin.

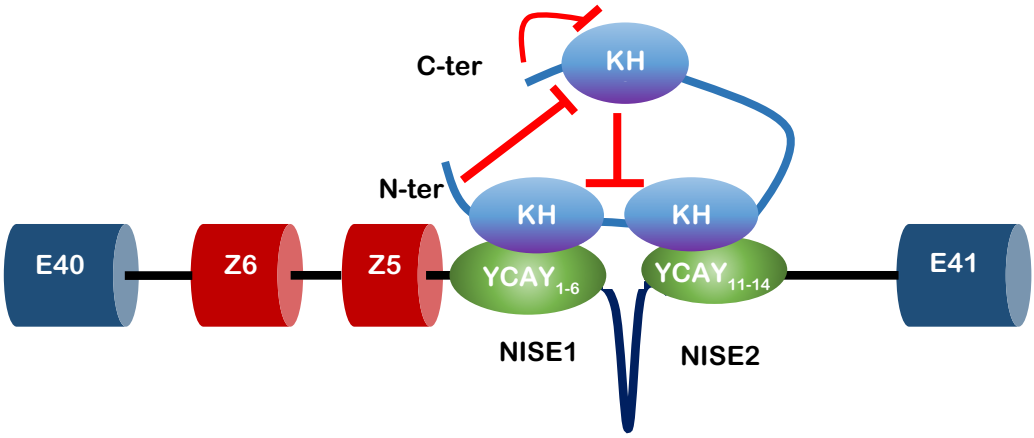


Figure 10. A model of CiNova protein action on CiAgrin pre-mRNA.

KH1 and KH2 domains of Nova interact with NISE1 and NISE2 on CiAgrin's pre-mRNA, (respectively or *vice versa*). The KH3 domain is a negative regulator of Agrin's splicing, while both N- and C-terminus domains block the inhibitory activity of KH3. We suggest that Nova in *Ciona* may regulate splicing using two different modalities: one mode requires KH1 and KH2 domains to regulate AS of targets such as Agrin; the other mode requires the KH3 domain to regulate AS of other, yet to be identified non-Agrin targets. The N/C-terminus domain acts as a regulatory switch between the two splice modalities.

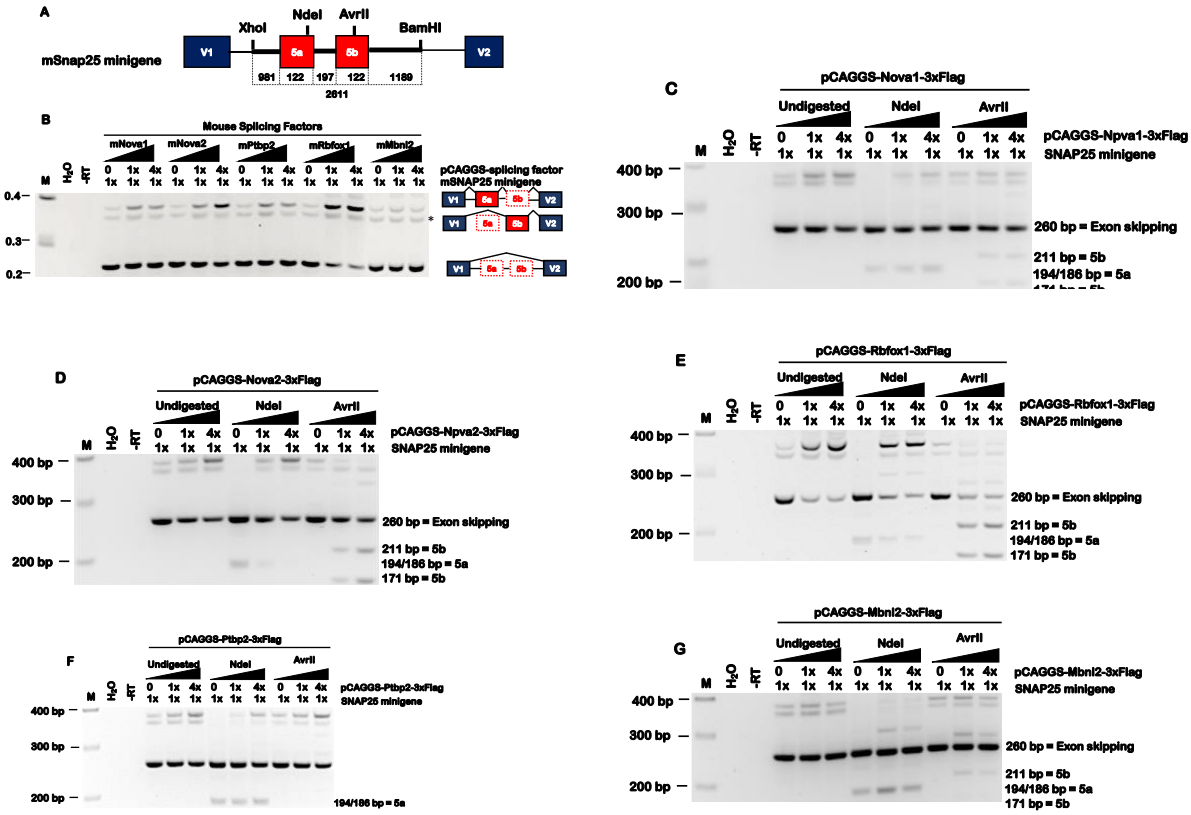


Figure 11. Nova1, Nova2, Ptbp2, and Rbfox1 regulate alternative splicing of Snap25.

HEK293T cells were co-transfected with constant amount of mSnap25 (1x) and 5 different splicing factors including mNova1/2, mPtbp2, mRbfox1, and mMbnl2. Different amounts of splicing factor (0 µg = 0x, 0.5 µg = 1x, 2 µg = 4x) were used. (A) Schematic illustration of mSnap25 minigene cloned into pSPL3 exon trapping vector. Vector exons V1 and V2 are depicted as dark blue boxes and alternative exons 5a/5b of mSnap25 are shown in red boxes. The thick dark line represents the introns of mSnap25, and the thin line represents the vector intron. The sizes of intronic and exon regions from Snap25 genomic region are indicated. The location of the restriction sites NdeI on exon 5a and AvrII on exon 5b are indicated, alongside the restriction sites used for cloning. (B) Minigene splicing assay of mSnap25 minigene and 5 expression constructs of mNova1, mNova2, mPtbp2, mRbfox1, and mMbnl2. A dose-dependent inclusion of alternative 5a/5b exons was detected by semi-quantitative RT-PCR (*shadow band of 5a/5b) in all samples. (C, D, E, F, G) Total RNA from HEK293T cells were extracted after co-transfection with mSnap25 minigene and subjected to semi-quantitative RT-PCR analysis. RT-PCR products were digested with NdeI and AvrII to determine which exon inclusion is promoted by mNova1, mNova2, mRbfox1, mMbnl2, and mPtbp2. If exon 5a

is included, digestion of the RT-PCR products with the exon 5a-specific restriction enzyme NdeI generates two fragments of 194 and 186 bp; conversely, if exon 5b is included, digestion of the RT-PCR products with the exon 5b-specific restriction site AvrII generates two fragments of 211 and 171 bp. (C) mNova1 very weakly promotes the inclusion of both 5a and 5b in a dose-dependent manner, while mNova2 (D) and mRbfox1 (E) promotes the inclusion of 5b and skipping of 5a of mSnap25 in a dose-dependent manner. (F) mPtbp2 only promotes the inclusion of 5a of mSnap25 minigene (G) mMbln2 regulates inclusion of both exons but the inclusion of 5a is stronger than 5b.

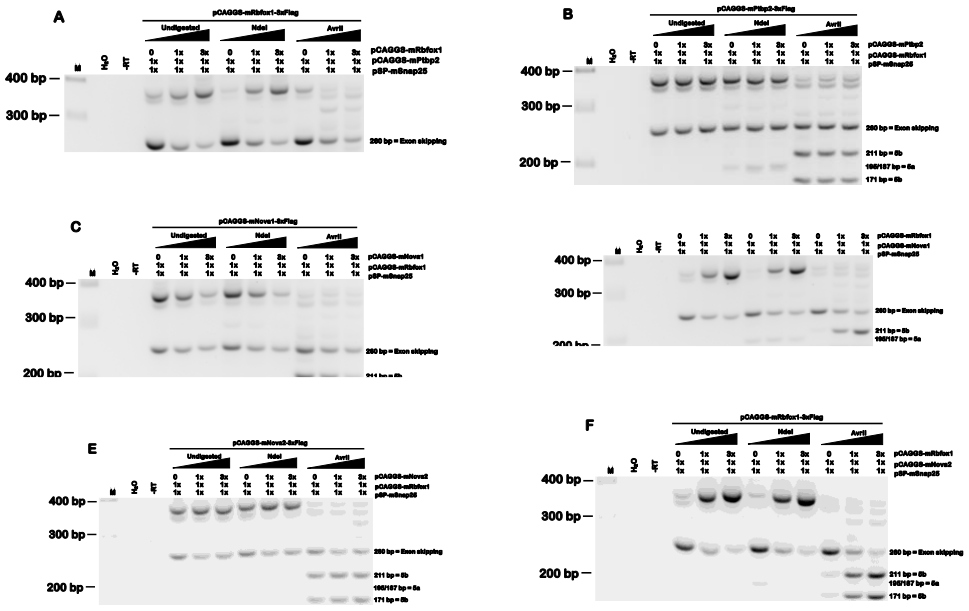


Figure 12. Competition between splicing factors determines splicing outcome of 5a and 5b of mSnap25.

(A, B, C, D, E, F) HEK293T cells were co-transfected with mSnap25 minigene and 2 different splicing factors (one with varying amounts and other one with constant amount). Total RNA was extracted 48 hours after transfection and subjected to semi-quantitative RT-PCR analysis. RT-PCR products were digested with NdeI and AvrII to determine which exon inclusion or skipping is promoted by each splicing factor. (A) RT-PCR analysis of cells co-transfected with constant amounts of mPtbp2 in combination with increasing amounts of mRbfox1. mRbfox1 promotes inclusion of 5b when competing with mPtbp2 for mSnap25 splicing. (B) mPtbp2 promotes the inclusion of 5a and skipping of 5b in a dose-dependent manner when competing with constant amount of mRbfox1. (C) mNova1 promotes skipping of 5b in a dose-dependent manner when competing with constant amount of mRbfox1. (D) mRbfox1 promotes inclusion of 5b when competing with constant amount of mNova1. (E, F) No competition is observed between mRbfox1 and mNova2.

CHAPTER 3

INVESTIGATING HOW MUTATIONS IN *SLC25A10* MAY AFFECT SPLICING IN A PATIENT WITH EPILEPTIC ENCEPHAPLOPATHY

3.1 ABSTRACT

Mitochondrial diseases are a plethora of inherited neuromuscular disorders sharing defects in mitochondrial respiration, but largely different from one another for genetic basis and pathogenic mechanism. The laboratory of Dr. De Grassi performed whole exome sequencing in a familiar trio (trio-WES) with a child affected by severe epileptic encephalopathy associated with respiratory complex I deficiency and mitochondrial DNA depletion in skeletal muscle. By trio-WES they identified biallelic mutations in *SLC25A10*, a nuclear gene encoding a solute carrier protein that transports molecules for Kreb's cycle and is a part of complex I in mitochondria. The patient inherited 3 mutations from his parents: 1 from his mother and 2 from his father. The maternal-derived mutation introduces a stop codon in exon 3. Mutations from the paternal allele are located in exon 9 and intron 10. Although the exonic mutation is a synonymous mutation, the patient had very low levels of *SLC25A10* mRNA and was devoid of protein. Using minigene splicing studies, we discovered that paternal-derived mutations cause aberrant splicing, providing evidence for the mechanism that leads to the failure to make a functional protein product in the patient. Our work underlies the

importance of splicing in neurodegenerative disorders and proposes a molecular mechanism that explains disease pathology.

3.2 BACKGROUND

The mitochondrial respiratory complex I couples the transfer of electrons from NADH to ubiquinone and the translocation of protons from the mitochondrial matrix to the intermembrane space, contributing to oxidative phosphorylation. Clinical presentation of complex I deficiency is extremely heterogeneous, ranging from fatal neonatal disease to adult-onset neurodegenerative disorders (Fassone *et al.,* 2012), and often associates to epilepsy (Khurana *et al.,* 2008). The vast genetic heterogeneity of isolated complex I deficiency is caused by mutations in mitochondrial DNA (mtDNA) or, in most cases, in nuclear-DNA genes encoding structural subunits, assembly factors, or other proteins with apparently complex I unrelated functions (see OMIM 252010). Further, the activity of complex I is susceptible to environmental factors, such as oxidative stress (Musatov *et al.,* 2012). *SLC25A10* (also known as *DIC*) transports dicarboxylates and phosphate across the inner mitochondrial membrane and is conserved from yeast to mammals (Palmieri *et al.,* 1996; Fiermonte *et al.,* 1998). As reviewed in (Lash *et al.,* 2015), *SLC25A10* inhibition has been repeatedly reported to cause reduced levels of mitochondrial glutathione (GSH) and impairment of complex I activity in rat neurons (Kamga *et al.,* 2010). Here we report the first human mutations abolishing *SLC25A10* function in a patient affected by a progressive form of epileptic encephalopathy and severe hypotonia associated with complex I deficiency.

3.2.1 Case description

This study was approved by the Pediatric Ethics Committee of the Tuscany Region, Italy (in the context of the Project DESIRE) and informed consent was signed by the patient's parents. The proband is a 9-year-old boy born at term from nonconsanguineous healthy parents of two additional healthy children. Hypospadias, bilateral hydrocele and unilateral right hearing loss were noticed in the patient. At 3 months, generalized jerking and infantile spasms appeared in multiple per day episodes. Clinical examinations showed severe hypotonia, absent eye tracking and poor spontaneous movements. EEG revealed multifocal epileptiform discharges and a suppression burst pattern. MRI, initially normal, showed high signal intensity of the white matter at 1 year and also thinning of the CC at 4 years of age. Clinical conditions evolved in intractable tonic spasms and focal seizures and quadriparesis, which progressively became spastic and dyskinetic. Growth parameters have always been within the normal range and neuropsychiatric evaluations revealed intellectual ability in the normal-high range. Multiple metabolic investigations, cerebrospinal-fluid amino acids, neurotransmitter levels and visual evoked potentials were normal. Blood analyses revealed microcytic anemia, which was improved after iron supplementation, and increased lactate (3.64 mM) and lactate/pyruvate (25.58) levels. The analysis of the four mitochondrial respiratory complexes from muscle homogenates, performed at 18 months, indicated reduced respiratory complex I activity (27% of the mean control value) and decreased mtDNA content (40% lower than the mean control value). Multiple antiepileptic drug trials and ketogenic diet were ineffective against seizures.

3.2.2 Compound heterozygous mutations in *SLC25A10*

Pathogenic mutations in the mitochondrial genome were excluded by Sanger sequencing. Three independent whole exome sequencing experiments (trio-WES) were performed on the genomic DNA of the patient and his parents. Compound heterozygous mutations were identified in *SLC25A10*. A heterozygous *SLC25A10* nonsense mutation was inherited from the mother (NM_001270888.1: c.304A>T, p.Lys102*, Fig.). This mutation, absent in the ExAC database (September 2017), generates a prematurely truncated protein lacking ~70% of the amino acid sequence. Two heterozygous mutations were inherited from the father: a synonymous mutation (NM_001270888.1: c.684C>T, p.Pro228Pro) annotated in the dbSNP146 database (rs114621664) with frequency of 0.0014 in the ExAC database and an intronic mutation (NM_001270888.1: c.790–37G>A) annotated in the dbSNP146 database (rs200706742) with frequency of 0.0011 in the ExAC database. Mutations have been submitted to the NCBI ClinVar database (www.ncbi.nlm.nih.gov/clinvar/, accessions SCV000611119 and SCV000611120).
The quantitative Real Time-PCR (qRT-PCR) analysis, conducted on cDNA obtained from cultured fibroblasts, showed a 10-fold decrease in *SLC25A10* transcript level in the patient relative to control. PCR primer-pairs were designed to amplify two distinct portions of the *SLC25A10* cDNA, which span exons 3–8 and exons 3–11, respectively. The former was detected in both the patient and control. The sequence of this fragment corresponds to the paternally inherited allele, suggestive of nonsense mediated decay of the maternally inherited one. The fragment spanning exons 3–11 is instead virtually absent in the patient cells, suggesting abnormal RNA splicing in-between exons 9 and 11. The paternally inherited mutations are predicted to break an exon splicing enhancer in exon 9 and to create a new one in intron 10, respectively.

3.3 MATERIALS AND METHODS

3.3.1 Transfection with *SLC25A10* minigene

The day before the transfection 0.6 x 10^6 HEK293T cells were seeded per well in a 6-well plate (USA Scientific). On the day transfection, a total of 1 μg DNA of *SLC25A10* minigene was used to transfect each of 6 well plate(s) and 3 μL of PEI (1mg/mL) was used in a ratio of 1:3 (DNA : PEI). The total volume of the DNA mixture was 200 μL. First, the exact amount of DNA in μL was pipetted in 1.5 mL Eppendorf tube (Eppendorf) and Opti-MEM media was used to bring the volume to 197 μL. Then the mixture was vortexed thoroughly. Finally, 3 μL of PEI were added, vortexed, and centrifuged briefly. The mixture was then incubated for 15 minutes at room temperature. In the meantime, medium in the cells was aspirated and 2 mL new DMEM medium was added. After 15 minutes of incubation 200 μL of reaction mixture was added to the cell and the plate was cross-shanked gently. The plate was then Incubated for 48 hours at 37 ° C.

For other materials and methods please refer to "2.2 MATERIALS AND METHODS" section in Chapter 2.

3.4 RESULTS

3.4.1 *SLC25A10* mutations cause RNA depletion and aberrant RNA splicing

To investigate if paternally derived mutations lead to aberrant splicing, 5 pair of WT and mutant minigenes were generated (Fig. 13J) containing the genomic region of *SLC25A10* encompassing flanking exons/introns. The mutant minigenes include the intronic mutation (G>A) alone (with and without 3′ UTR), the exonic mutation (C>T) alone, and both intronic and exonic mutations together. The genomic region was amplified by PCR and cloned into the mammalian expression vector pCi-neo. We transfected constant amounts of our minigene construct in triplicate in HEK293T cells. Total RNA was extracted, quantified, normalized and subjected to semi-quantitative RT-PCR analysis.

Our splicing data revealed a change in splicing pattern for the intronic mutation alone (with and without 3′ UTR) (Fig.13B, D) but not for the synonymous exonic mutation alone (Fig.13F), suggesting a primary role of the intronic mutation in inducing aberrant RNA splicing. Intriguingly, however, splicing was severely disrupted in a minigene hosting both exonic and intronic mutations (Fig.13H): the exonic mutation alone has no effect but it appears to exacerbate the effect of intronic mutation, thus having an additive effect on aberrant pre-mRNA splicing. RT-PCR quantification of RNA splicing isoforms showed that mutant *SLC25A10* promotes a shift toward shorter splicing isoforms, thus favoring the exclusion of both exon 10 and intron 10, when compared with the WT allele (Fig.13I). Our splicing investigation on the paternally-derived mutant *SLC25A10* allele provides a convincing explanation for the absence of protein in the patient.

3.5 DISCUSSION

Here, we used our functional minigene splicing assay to explain disease pathology in a patient with severe epileptic and progressive encephalopathy. The proband inherited 3 mutations from his parents including 1 from his mother and 2 from his father. The maternal-derived mutation introduces a stop codon in exon 3, while parental-derived mutations are located in exon 9 and intron 10. Although the exonic mutation is a synonymous mutation, the patient had very low levels of *SLC25A10* mRNA and was devoid of protein. Using minigene splicing assays, we showed that the mutant DNA promotes a shift toward shorter splicing isoforms, thus favoring the exclusion of *SLC25A10* exon 10 and intron 10, when compared with the wild-type (Fig. 13H, I). The RNA quantitative change was also observed in other minigene hosting the intronic mutation alone (Fig. 13B, D), but not in minigene hosting the synonymous mutation alone (Fig. 13F), however, the effect of exonic mutation exacerbate the effect of intronic mutation, suggesting a primary role of the intronic mutation in inducing the aberrant RNA splicing (Fig. 13H). Our work underlies the importance of splicing in neurodegenerative disorders and proposes a molecular mechanism that explains disease pathology.

3.6 CONCLUSIONS AND PERSPECTIVES

In summary, we report first *SLC25A10* recessive disease-causing mutations associated with abnormal splicing. Albeit additional *SLC25A10* mutations in unrelated individuals will be required as standard proof of causality. Further work is warranted to exploit the *SLC25A10*-mediated molecular mechanisms that protect cellular respiration

and redox state. An interesting future project would be do design antisense oligonucleotides to restore the expression of correct isoform of *SLC25A10*.

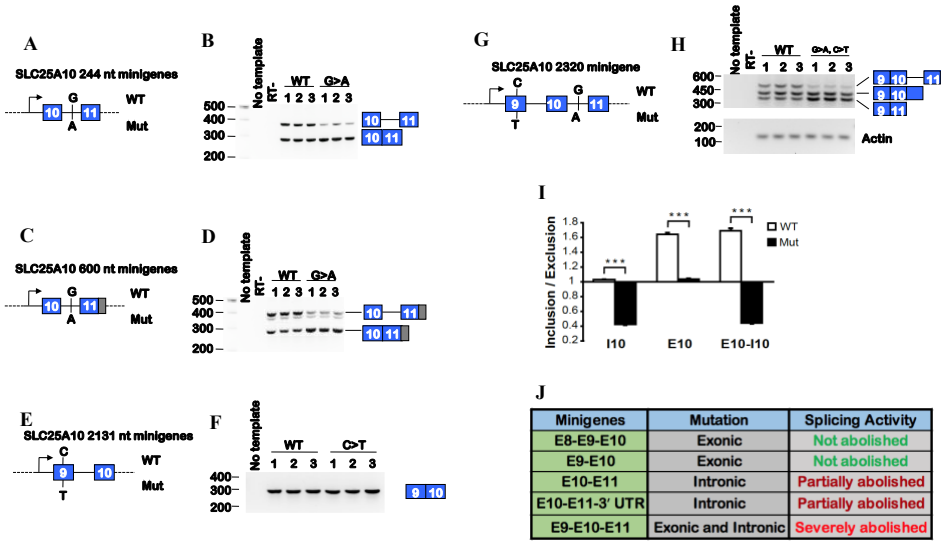
3.7 TABLES AND FIGURES

3.7.1 TABLES

Table 8. List of all the forward and reverse primers used to clone SLC25A10 minigene constructs used in this study

Construct Name	Forward(F)/Reverse(R) Promers	Primer Sequence
SLC25A10 244 nt minigene	SLC25A10_E9F1_XhoI	ACAGTGCTCGAGGCCGCTGGTGACGAGC
	SLC25A10_E11R1_XbaI	ACAGTGTCTAGAGGATGGCACTTTGATGCCAAAG
SLC25A10 600 nt minigene	SLC25A10_E9F1_XhoI	ACAGTGCTCGAGGCCGCTGGTGACGAGC
	SLC25A10_E11R2_XbaI	ACAGTGTCTAGACCTCGATGGAAAGTGCTGGAAGAT
SLC25A10 2131 nt minigene	SLC25A10_I8F1	TAGGAGTCAGGTGGAGGTTCTGG
	SLC25A10_E10R1_XbaI	ACAGTGTCTAGACTTGTAAAAGGCCAGAGGCCC
SLC25A10 2320 nt minigene	SLC25A10_I8F2	TAAGTGGCCGGCATGGCTA
	SLC25A10_E11R1_XbaI	ACAGTGTCTAGAGGATGGCACTTTGATGCCAAAG
SLC25A10 2489 nt minigene	SLC25A10_E8F1_XhoI	ACAGTGCTCGAGCTGTCCTGCTACGACCAGG
	SLC25A10_E10R1_XbaI	ACAGTGTCTAGACTTGTAAAAGGCCAGAGGCCC

3.7.2 FIGURES AND FIGURE LEGENDS



Minigenes	Mutation	Splicing Activity
E8-E9-E10	Exonic	Not abolished
E9-E10	Exonic	Not abolished
E10-E11	Intronic	Partially abolished
E10-E11-3′ UTR	Intronic	Partially abolished
E9-E10-E11	Exonic and Intronic	Severely abolished

Figure 13. A functional splicing analysis for SLC25A10 minigenes.

(A, C, E, G) Schematic representations of the minigene constructs; minigenes are named based on the number of nucleotides (nt) (blue box: coding sequence with exon number; gray: 3'UTR), along with the sequence identity of the cloned DNA (WT: wild-type; Mut: mutant). (B, D, F, H) RT-PCR analysis of total RNA extracted from HEK293T cells transfected with specific minigenes in triplicate. The intronic mutants without (B) and with 3' UTR (D) promote a small change in splicing by excluding intron 10 (I10) when compared to WT minigene. However, the minigene hosting the intronic and exonic mutations together (G) severely disrupt the splicing program (H). (I) Ratio analysis between the splicing fragments from figure H shows a significant reduction of I10, E11 (exon11), and E10 (exon10)-I10. Actin served as experimental control (H). No difference is observed in the two minigenes hosting the synonymous mutation (C>T) alone (F). (J) Table representing splicing activity of all the different combination of exonic and intronic mutant minigenes. Splicing is severely affected when both intronic and exonic mutations are present in the minigene. Data are presented as mean+SEM of at least three experiments; ***: non parametric Wilcoxon test P-value<0.05.(minigene cartoons are not drawn to scale).

REFERENCES

Allen, Summer E., Robert B. Darnell, and Diane Lipscombe. "The neuronal splicing factor Nova controls alternative splicing in N-type and P-type CaV2 calcium channels." *Channels* 4.6 (2010): 483-489.

Ars, E., *et al.,* "Recurrent mutations in the NF1 gene are common among neurofibromatosis type 1 patients." *Journal of Medical Genetics* 40.6 (2003): e82-e82.

Barbosa-Morais, Nuno L., Maria Carmo-Fonseca, and Samuel Aparício. "Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion." *Genome Research* 16.1 (2006): 66-77.

Bentley, Barry, et al. "The multilayer connectome of Caenorhabditis elegans." *PLoS Computational Biology* 12.12 (2016): e1005283.

Berget, Susan M., Claire Moore, and Phillip A. Sharp. "Spliced segments at the 5′ terminus of adenovirus 2 late mRNA." *Proceedings of The National Academy of Sciences* 74.8 (1977): 3171-3175.

Bezakova, Gabriela, and Markus A. Ruegg. "New insights into the roles of agrin." *Nature Reviews Molecular Cell Biology* 4.4 (2003): 295.

Bhalla, Kavita, et al. "The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene." *Journal of Human Genetics* 49.6 (2004): 308-311.

Black, Douglas L. "Mechanisms of alternative pre-messenger RNA splicing." *Annual Review of Biochemistry* 72.1 (2003): 291-336.

Buckanovich, Ronald J., and Robert B. Darnell. "The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo." *Molecular and Cellular Biology* 17.6 (1997): 3194-3201.

Buckanovich, Ronald J., Jerome B. Posner, and Robert B. Darnell. "Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system." *Neuron* 11.4 (1993): 657-672.

Buckanovich, Ronald J., Y. Y. Yang, and Robert B. Darnell. "The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies." *Journal of Neuroscience* 16.3 (1996): 1114-1122.

Burden, Steven, Maartje Huijbers, and Leonor Remedio. "Fundamental molecules and mechanisms for forming and maintaining neuromuscular synapses." *International Journal of Molecular Sciences* 19.2 (2018): 490.

Burge, C. B., T. Tuschl, and P. A. Sharp. "*The RNA World*." (1999): 525-560.

Burnett, Barrington G., et al. "Regulation of SMN protein stability." *Molecular and Cellular Biology* 29.5 (2009): 1107-1115.

Cai, Xin, *et al.,* "A human-specific schizophrenia risk tandem repeat affects alternative splicing of a human-unique isoform AS3MT d2d3 and mushroom dendritic spine density." *Schizophrenia Bulletin* 47.1 (2021): 219-227.

Cartegni, Luca, and Adrian R. Krainer. "Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1." *Nature Genetics* 30.4 (2002): 377-384.

Cartegni, Luca, Shern L. Chew, and Adrian R. Krainer. "Listening to silence and understanding nonsense: exonic mutations that affect splicing." *Nature Reviews Genetics* 3.4 (2002): 285-298.

Charizanis, Konstantinos, et al. "Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy."*Neuron* 75.3 (2012): 437-450.

Chow, Louise T., *et al.,* "An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA." *Cell* 12.1 (1977): 1-8.

Dillman, Allissa A., *et al.,* "mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex." *Nature Neuroscience* 16.4 (2013): 499-506.

Darnell, Robert B. "Onconeural antigens and the paraneoplastic neurologic disorders: at the intersection of cancer, immunity, and the brain." *Proceedings of the National Academy of Sciences* 93.10 (1996): 4529-4536.

Darnell, Robert B., and Jerome B. Posner. "Paraneoplastic syndromes involving the nervous system." *New England Journal of Medicine* 349.16 (2003): 1543-1554.

Del Gatto-Konczak, Fabienne, *et al.,* "hnRNP A1 recruited to an exon in vivo can function as an exon splicing silencer." *Molecular and Cellular Biology* 19.1 (1999): 251-260.

Delsuc, Frédéric, et al. "Tunicates and not cephalochordates are the closest living relatives of vertebrates." *Nature* 439.7079 (2006): 965.

Denzer, Alain J., et al. "An amino-terminal extension is required for the secretion of chick agrin and its binding to extracellular matrix." *The Journal of Cell Biology* 131.6 (1995): 1547-1560.

Dredge, B. Kate, et al. "Nova autoregulation reveals dual functions in neuronal splicing." *The EMBO Journal* 24.8 (2005): 1608-1620.

Duman-Scheel, M. "Netrin and DCC: axon guidance regulators at the intersection of nervous system development and cancer." *Current Drug Targets* 10.7 (2009): 602-610.

Eizirik, Décio L., et al. "The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines." *PLoS Genetics* 8.3 (2012): e1002552.

Elia, J., et al. "Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes." *Molecular Psychiatry* 15.6 (2010): 637-646.

Engel, Andrew G., Xin-Ming Shen, and Duygu Selcen. "The unfolding landscape of the congenital myasthenic syndromes." *Annals of The New York Academy of Sciences* 1413.1 (2018): 25.

Eom, Taesun, et al. "NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure." *Elife* 2 (2013): e00178.

Fassone, Elisa, and Shamima Rahman. "Complex I deficiency: clinical features, biochemistry and molecular genetics." *Journal of Medical Genetics* 49.9 (2012): 578-590.

Fiermonte, Giuseppe, et al. "The sequence, bacterial expression, and functional reconstitution of the rat mitochondrial dicarboxylate transporter cloned via distant homologs in yeast and Caenorhabditis elegans." *Journal of Biological Chemistry* 273.38 (1998): 24754-24759.

Fukuda, Hiroyuki, et al. "A novel missense SNAP25b mutation in two affected siblings from an Israeli family showing seizures and cerebellar ataxia." *Journal of Human Genetics* 63.5 (2018): 673-676.

Gautam, Medha, et al. "Defective neuromuscular synaptogenesis in agrin-deficient mutant mice." *Cell* 85.4 (1996): 525-535.

Gehman, Lauren T., et al. "The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain." *Nature Genetics* 43.7 (2011): 706-711.
Gesemann, Matthias, Alain J. Denzer, and Markus A. Ruegg. "Acetylcholine receptor-aggregating activity of agrin isoforms and mapping of the active site." *The Journal of Cell Biology*128.4 (1995): 625-636.

Giampietro, Costanza, et al. "The alternative splicing factor Nova2 regulates vascular development and lumen formation." *Nature Communications* 6.1 (2015): 1-15.

Gilbert, Walter. "Why genes in pieces?." *Nature* 271.5645 (1978): 501-501.

Grishin, Nick V. "KH domain: one motif, two folds." *Nucleic Acids Research* 29.3 (2001): 638-643.

Groffen, Alexander JA, et al. "Primary structure and high expression of human agrin in basement membranes of adult lung and kidney." *European Journal of Biochemistry* 254.1 (1998): 123-128.

Hamshere, Marian Lindsay, et al. "Genetic utility of broadly defined bipolar schizoaffective disorder as a diagnostic concept." *The British Journal of Psychiatry* 195.1 (2009): 23-29.

Hollingworth, David, et al. "KH domains with impaired nucleic acid binding as a tool for functional analysis." *Nucleic Acids Research* 40.14 (2012): 6873-6886.

Hrus, Ana, et al. "C. elegans agrin is expressed in pharynx, IL1 neurons and distal tip cells and does not genetically interact with genes involved in synaptogenesis or muscle function." *PLoS One* 2.8 (2007): e731.

Huang, Cindy Shen, et al. "Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition." *Cell* 123.1 (2005): 105-118.

Irimia, Manuel, *et al.,* "A highly conserved program of neuronal microexons is misregulated in autistic brains." *Cell* 159.7 (2014): 1511-1523.

Jensen, Kirk B., et al. "Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability." *Neuron* 25.2 (2000): 359-371.

Jensen, Kirk B., et al. "The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain." *Proceedings of the National Academy of Sciences* 97.11 (2000): 5740-5745.

Johansson, Jenny U., et al. "An Ancient Duplication of Exon 5 in the Snap 25 Gene Is Required for Complex Neuronal Development/Function." *PLoS Genetics* 4.11 (2008): e1000278.

Juan-Mateu, Jonàs, et al. "Neuron-enriched RNA-binding proteins regulate pancreatic beta cell function and survival." *Journal of Biological Chemistry* 292.8 (2017): 3466-3480.

Kamga, Christelle K., Shelley X. Zhang, and Yang Wang. "Dicarboxylate carrier-mediated glutathione transport is essential for reactive oxygen species homeostasis and normal respiration in rat brain mitochondria." *American Journal of Physiology-Cell Physiology* 299.2 (2010): C497-C505.

Kashima, Tsuyoshi, and James L. Manley. "A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy." *Nature Genetics* 34.4 (2003): 460-463.

Khoo, Bernard, and Adrian R. Krainer. "Splicing therapeutics in SMN2 and APOB." *Current Opinion in Molecular Therapeutics* 11.2 (2009): 108.

Khurana, D. S., et al. "Epilepsy and respiratory chain defects in children with mitochondrial encephalopathies." *Neuropediatrics* 39.01 (2008): 8-13.

Kim, Namshin, *et al.,* "The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species." *Nucleic Acids Research* 35.suppl_1 (2007): D93-D98.

Lash, Lawrence H. "Mitochondrial glutathione in diabetic nephropathy." *Journal of Clinical Medicine* 4.7 (2015): 1428-1447.

Lázaro-Peña, María I., et al. "synaptogenesis is modulated by heparan sulfate in Caenorhabditis elegans." *Genetics* 209.1 (2018): 195-208.
Leggere, Janelle C., et al. "NOVA regulates Dcc alternative splicing during neuronal migration and axon guidance in the spinal cord." *Elife* 5 (2016): e14264.

Le-Niculescu, H., et al. "Convergent functional genomics of genome-wide association data for bipolar disorder: Comprehensive identification of candidate genes, pathways and mechanisms." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 150.2 (2009): 155-181.

Lenzken, Silvia Carolina, *et al.,* "Neuronal RNA-binding proteins in health and disease." *Wiley Interdisciplinary Reviews*: RNA 5.4 (2014): 565-576.

Letunic, Ivica, Richard R. Copley, and Peer Bork. "Common exon duplication in animals and its role in alternative splicing." *Human Molecular Genetics* 11.13 (2002): 1561-1567.

Lewis, Hal A., et al. "Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome." *Cell* 100.3 (2000): 323-332.

Licatalosi, Donny D., et al. "HITS-CLIP yields genome-wide insights into brain alternative RNA processing." *Nature* 456.7221 (2008): 464-469.

Licatalosi, Donny D., et al. "Ptbp2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain."*Genes & Development* 26.14 (2012): 1626-1642.

Lipscombe, Diane, Arturo Andrade, and Summer E. Allen. "Alternative splicing: functional diversity among voltage-gated calcium channels and behavioral

consequences." *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1828.7 (2013): 1522-1529.

Lorson, Christian L., et al. "SMN oligomerization defect correlates with spinal muscular atrophy severity." *Nature Genetics* 19.1 (1998): 63-66.

Lorson, Christian L., et al. "A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy." *Proceedings of the National Academy of Sciences* 96.11 (1999): 6307-6311.

Lunde, Bradley M., Claire Moore, and Gabriele Varani. "RNA-binding proteins: modular design for efficient function." *Nature Reviews Molecular Cell Biology* 8.6 (2007): 479-490.

Martin, Christa Lese, et al. "Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 144.7 (2007): 869-876.

Mattioli, Francesca, et al. "De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder." *The American Journal of Human Genetics* 106.4 (2020): 438-452.

Mattioli, Francesca. Identification of novel genetic causes of monogenic intellectual disability. *Diss. Université de Strasbourg*, 2018.

Meldolesi, Jacopo. "Alternative Splicing by NOVA Factors: From Gene Expression to Cell Physiology and Pathology." *International Journal of Molecular Sciences* 21.11 (2020): 3941.

Mikulak, Joanna, et al. "Dual REST-dependence of L1CAM: from gene expression to alternative splicing governed by Nova2 in neural cells." *Journal of Neurochemistry* 120.5 (2012): 699-709.

Mills, James Dominic, and Michal Janitz. "Alternative splicing of mRNA in the molecular pathology of neurodegenerative diseases." *Neurobiology of Aging* 33.5 (2012): 1012-e11.

Misra, Sima, *et al*. "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review." *Genome Biology* 3.12 (2002): 1-22.

Musatov, Andrej, and Neal C. Robinson. "Susceptibility of mitochondrial electron-transport complexes to oxidative damage. Focus on cytochrome c oxidase." *Free Radical Research* 46.11 (2012): 1313-1326.

Musco, Giovanna, et al. "Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome." *Cell* 85.2 (1996): 237-245.

Nisson, Paul E., Abdul Ally, and Paul C. Watkins. "Protocols for trapping internal and 3'-terminal exons." *Genome Research* 4.1 (1994): S24-S39.

Nitkin, Ralph M., et al. "Identification of agrin, a synaptic organizing protein from Torpedo electric organ." *The Journal of Cell Biology* 105.6 (1987): 2471-2478.

O'leary, Ben, Richard S. Finn, and Nicholas C. Turner. "Treating cancer with selective CDK4/6 inhibitors." *Nature reviews Clinical oncology* 13.7 (2016): 417-430.

Palmieri, Luigi, et al. "Identification by bacterial expression and functional reconstitution of the yeast genomic sequence encoding the mitochondrial dicarboxylate carrier protein." *FEBS Letters* 399.3 (1996): 299-302.

Pan, Qun, *et al.,* "Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing." *Nature Genetics* 40.12 (2008): 1413-1415.

Parikshak, Neelroop N., *et al.,* "Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism." *Nature* 540.7633 (2016): 423-427.

Pellizzoni, Livio, *et al.,* "A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing." *Cell* 95.5 (1998): 615-624.

Pennisi, Elizabeth. "Comparative biology joins the molecular age." *Science* (2002): 1792-1795.

Polydorides, Alexandros D., et al. "A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing." *Proceedings of the National Academy of Sciences* 97.12 (2000): 6350-6355.

Porter, Robert S., Farris Jaamour, and Shigeki Iwase. "Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders." *Molecular and Cellular Neuroscience* 87 (2018): 35-45.

Quesnel-Vallières, Mathieu, *et al.,* "Misregulation of an activity-dependent splicing network as a common mechanism underlying autism spectrum disorders." *Molecular Cell* 64.6 (2016): 1023-1034.

Rahman, M. A., and K. Ohno. "Splicing aberrations in congenital myasthenic syndromes." *Journal of Investigative Genomics* 2.6 (2015): 111-118.

Ruggiu, Matteo, et al. "Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing." *Proceedings of the National Academy of Sciences* 106.9 (2009): 3513-3518.

Rupp, Fabio, et al. "Structure and expression of a rat agrin." *Neuron* 6.5 (1991): 811-823.

Ryan, Kerrianne, Zhiyuan Lu, and Ian A. Meinertzhagen. "The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling." *Elife* 5 (2016): e16962.

Saito, Yuhki, et al. "Differential NOVA2-mediated splicing in excitatory and inhibitory neurons regulates cortical development and cerebellar function." *Neuron* 101.4 (2019): 707-720.

Saito, Yuhki, et al. "NOVA2-mediated RNA regulation is required for axonal pathfinding during development." *Elife* 5 (2016): e14371.

Satoh, Noriyuki. Developmental biology of ascidians. Cambridge University Press, 1994.
Sebat, Jonathan, et al. "Strong association of de novo copy number mutations with autism." *Science* 316.5823 (2007): 445-449.
Schmucker, Dietmar, *et al.,* "Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity." *Cell* 101.6 (2000): 671-684.

Shen, Haihong, Julie LC Kan, and Michael R. Green. "Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly." *Molecular Cell* 13.3 (2004): 367-376.

Shen, Xin-Ming, et al. "Mutant SNAP25B causes myasthenia, cortical hyperexcitability, ataxia, and intellectual disability." *Neurology* 83.24 (2014): 2247-2255.

Siomi, Haruhiko, et al. "Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome." *Cell* 77.1 (1994): 33-39.

Siomi, Haruhiko, et al. "The pre-mRNA binding K protein contains a novel evolutionary conserved motif." *Nucleic Acids Research* 21.5 (1993): 1193-1198.

Smith, Martin A., et al. "Isolation and characterization of a cDNA that encodes an agrin homolog in the marine ray." *Molecular and Cellular Neuroscience* 3.5 (1992): 406-417.
Sørensen, Jakob B., et al. "Differential control of the releasable vesicle pools by SNAP-25 splice variants and SNAP-23." *Cell* 114.1 (2003): 75-86.

Stamm, Stefan, *et al.,* "A sequence compilation and comparison of exons that are alternatively spliced in neurons." *Nucleic Acids Research* 22.9 (1994): 1515-1526.

Stoeckli, Esther T. "Understanding axon guidance: are we nearly there yet?." *Development* 145.10 (2018).

Sugimoto, Yoichiro, et al. "Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions." *Genome Biology* 13.8 (2012): 1-13.

Sugnet, Charles W., *et al.,* "Transcriptome and genome conservation of alternative splicing events in humans and mice." *Biocomputing* 2004. 2003. 66-77.

Teraoka, Sharon N., *et al.,* "Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences." *The American Journal of Human Genetics*64.6 (1999): 1617-1631.

Thalhammer, Agnes, *et al.,* "Alternative splicing of P/Q-type Ca2+ channels shapes presynaptic plasticity." *Cell Reports*20.2 (2017): 333-343.

Thierry-Mieg, Danielle, and Jean Thierry-Mieg. "AceView: a comprehensive cDNA-supported gene and transcripts annotation." *Genome Biology* 7.1 (2006): 1-14.

Tompson, Stuart W., and Terri L. Young. "Assaying the effects of splice site variants by exon trapping in a mammalian cell line." *Bio-protocol* 7.10 (2017).

Tsim, Karl WK, et al. "cDNA that encodes active agrin." *Neuron* 8.4 (1992): 677-689.

Ule J, et al. (2006) An RNA map predicting Nova-dependent splicing regulation. *Nature* 444:580–586.

Ule, Jernej, et al. "CLIP identifies Nova-regulated RNA networks in the brain." *Science* 302.5648 (2003): 1212-1215.

Ule, Jernej, *et al.,* "Nova regulates brain-specific splicing to shape the synapse." *Nature Genetics* 37.8 (2005): 844-852.

Underwood, Jason G., et al. "Homologues of the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals." *Molecular and Cellular Biology* 25.22 (2005): 10005-10016.

Valverde, Roberto, Laura Edwards, and Lynne Regan. "Structure and function of KH domains." *The FEBS Journal* 275.11 (2008): 2712-2726.

Vernia, Santiago, et al. "An alternative splicing program promotes adipose tissue thermogenesis." *Elife* 5 (2016): e17672.

Villate, Olatz, et al. "Nova1 is a master regulator of alternative splicing in pancreatic beta cells." *Nucleic Acids Research* 42.18 (2014): 11818-11830.

Voineagu, Irina, et al. "Transcriptomic analysis of autistic brain reveals convergent molecular pathology." *Nature* 474.7351 (2011): 380-384.

Wang, Kai, et al. "Common genetic variants on 5p14. 1 associate with autism spectrum disorders." *Nature* 459.7246 (2009): 528-533.

Wintle, Richard F., et al. "A genotype resource for postmortem brain samples from the Autism Tissue Program." *Autism Research* 4.2 (2011): 89-97.

Xiong, Hui Y., *et al.,* "The human splicing code reveals new insights into the genetic determinants of disease." *Science* 347.6218 (2015).

Xu, Bin, et al. "Strong association of de novo copy number mutations with sporadic schizophrenia." *Nature Genetics* 40.7 (2008): 880-885.

Xu, Qiang, Barmak Modrek, and Christopher Lee. "Genome-wide detection of tissue-specific alternative splicing in the human transcriptome." *Nucleic Acids Research* 30.17 (2002): 3754-3766.

Yang, Yolanda YL, Guang Lin Yin, and Robert B. Darnell. "The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia." *Proceedings of the National Academy of Sciences* 95.22 (1998): 13254-13259.

Yano, Masato, et al. "Nova2 regulates neuronal migration through an RNA switch in disabled-1 signaling." *Neuron* 66.6 (2010): 848-858.

Zhang, Bin, et al. "LRP4 serves as a coreceptor of agrin." *Neuron* 60.2 (2008): 285-297.

Zhang, Chaolin, et al. "Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls." *Science* 329.5990 (2010): 439-443.

Zhou, Zhaolan, *et al.,* "Comprehensive proteomic analysis of the human spliceosome." *Nature* 419.6903 (2002): 182-185.

Zong, Yinong, and Rongsheng Jin. "Structural mechanisms of the agrin–LRP4–MuSK signaling pathway in neuromuscular junction differentiation." *Cellular and Molecular Life Sciences* 70.17 (2013): 3077-3088.

Zong, Yinong, et al. "Structural basis of agrin–LRP4–MuSK signaling." *Genes & Development* 26.3 (2012): 247-258.

Vita

Name	*Md Faruk Hossain*
Degrees	*BSc and MSc, University of Dhaka, Dhaka, Bangladesh, Major: Biochemistry and Molecular Biology*
Date Graduated	*December, 2013 and 2014*
Other Degrees and Certificates	*Master of Science, St. John's University, Queens, Major: Biology*
Date Graduated	*December, 2016*