GENOME-WIDE HUMAN-SPECIFIC RNA REGULATORY ELEMENTS IN THE BRAIN

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Elisabeth A. Murphy
June 2021

GENOME-WIDE HUMAN-SPECIFIC RNA REGULATORY ELEMENTS IN THE BRAIN

Elisabeth A. Murphy, Ph.D.

The Rockefeller University 2021

Advancements in DNA sequencing technology and the implementation of clinical genome and exome sequencing have allowed for the identification of candidate variants and genes essential in brain development and pathogenesis of neurological diseases, yet there is still much to be learned. It is estimated that the molecular diagnostic rate of whole exome sequencing (WES) of constitutional diseases is between 9-41%; however, it is thought that diagnostic yield could be much improved by gaining a better understanding of individual variation and regulation at the transcript level. Identifying the root cause of diversity between humans and specifically in the human brain has been a long-standing question for the scientific field. At the turn of the 20thcentury, Santiago Ramón y Cajal illustrated thousands of neurons spanning across brain regions to irrefutably demonstrate the morphological complexity and diversity of individual neurons (Levine and Marcillo, 2008). Discoveries spanning centuries continue to drive investigators to question not only why neuronal diversity exists, but how a single genome gives rise to the remarkable heterogeneity observed between single cells, tissue structures, and circuit connections in the brain.

Post-transcriptional gene regulation contributes to organism complexity in eukaryotes. RNA regulation is strictly controlled by RNA-binding proteins that are important for the coordination and regulation of gene expression and eukaryotic cells have evolved intricate post-transcriptional mechanisms that permit for precise spatial and temporal control of RNA steady state levels in any given cell. Alternative RNA processing via RNA binding proteins underlies

distinct phenotypic and functional diversity specifically observed among mammalian cells. Alternative splicing (AS) and alternative polyadenylation (APA) are widely utilized to expand and diversify both transcriptome and proteome. Alternative RNA processing allows one gene to produce multiple transcripts with distinct coding and regulatory sequences giving rise to multifaceted protein function. RNA binding proteins dictate alternative RNA processing which is a key player in most, if not all, biological processes.

Understanding how RNA binding proteins regulate RNA steady state levels and alternative RNA processing requires knowledge of the genomic positions to which these proteins function *in vivo*. Mouse models have facilitated genome-wide, unbiased discovery of RNA regulatory sites, and the link between causation and functional effect with extraordinary resolution and specificity. The inability to translate many RNA regulatory events from the mouse to the human genome emphasizes the need to perform biochemistry on the human brain directly to compare with the mouse. Identification of RNA regulatory regions via CLIP-sequencing provides valuable information as to how human RNA profiles are dictated and controlled in normal and disease states. Studying neuronal splicing factor NOVA elucidates previously unidentified non-coding RNA regulatory sites as functional and unique to human neurons. We show that human expansion in NOVA regulatory elements diversifies the RNA landscape specifically in human neurons and differentially across brain regions. Taken together, these findings demonstrate mouse to human conserved relationships between RNA binding proteins and targets, however the activity, specificity, and usage of RNA regulatory elements is largely species-specific.

To my mom,

Your unwavering love and support has made this possible.

Acknowledgements

First and foremost I would like to thank my advisor Dr. Robert B. Darnell for his guidance and mentorship throughout the course of my thesis work. I want to thank you for welcoming me into the lab wholeheartedly as a second year rotation student and for providing such an innovative (and dog friendly) environment to study RNA biology. I am so grateful for all that I have learned from you. Thank you to my committee, Dr. Winrich Freiwald, Dr. Mary Beth Hatten, and Dr. Richard Lifton for all of your thoughtful insight and advice as my project has evolved over the last five years. I could not have asked for a better group of scientists to help me along the way. Winrich- thank you for being the supportive voice of reason during this journey. Rick- thank you for always challenging me with the tough questions and driving my research forward. Mary Beth- ten years ago, you invited me to your lab so that I could learn how to prepare cerebellar slices and I am so appreciative. Thank you for introducing me to Rockefeller and for continuing to teach me and encourage my scientific career. I especially thank Dr. Adrian Krainer for inspiring much of this work and for his willingness to serve as my external examiner.

One of Bob's many strengths is his ability to attract and hire amazing people and the people in Bob's lab are truly exceptional. They have all made coming to the lab every day enjoyable and I am lucky to have had the opportunity to work with such wonderful scientists and participate in so many engaging projects during my time here at Rockefeller.

Thank you Caryn for all of your contributions to my thesis work and for being my bench mate, therapist, scientific advisor, coding partner, and friend. Your mentorship has been invaluable. Thank you Ezgi and Ruth for all of your friendship, advice, enthusiasm, and help troubleshooting experiments. I like to think of Yuhki and Yoko as team NOVA in the lab. Your outstanding research in the lab and expert knowledge has been instrumental in the navigation of my project and I thank you both for all of your contributions, your insightful conversation, and willingness to swap NOVA findings. Thank you Jak for

being an RNA wizard and for all of your efforts and significant contributions towards my project, the breadth of my research would not be where it is today without you and your talent. Nathalie I am so grateful for all of your guidance and expertise that you have shared with me over the years. Thank you for your significant contributions to my thesis work and your enthusiasm to explore and troubleshoot new systems and protocols with me. These times and experiments have been some of my greatest memories and have greatly shaped my training and time in the lab. To the Darnell clinical team- Nathalie, Salina, Mayu, and Dana- thank you for all of your collaborations, discussions, and for welcoming all my foster dogs into your office :).

I would especially like to acknowledge and thank a former postdoc in the lab who mentored me during my rotation, Claudia Scheckel. Claudia taught me all things CLIP from experimental design, sample preparation, buffers, sequencing approaches, bioinformatics, and the list goes on. The knowledge that Claudia passed on to me greatly influenced my future endeavors in the lab and I am beyond grateful for her time, mentorship, and friendship. Thank you to all past and present Darnell lab members for always listening, advising, and commiserating when needed. Every lab member has contributed to the work presented in my thesis and I would like to express my gratitude for you all.

Part of my thesis work has focused on human model systems and that work would not have been possible without the expertise of Dylan Kwart, from Marc Tessier-Lavigne's lab. I would like to thank Dylan for guiding, teaching, and collaborating. Dylan was invaluable in teaching me CRISPR-Cas9 genome editing, iPSCs, and human neuron culture. I am so appreciative of your contributions and friendship.

I have had the pleasure of collaborating and learning from so many great scientists during my time at Rockefeller. One collaboration that has significantly contributed to the work presented here in the appendix of my thesis is with Sean O'Connor, a graduate student in Paul Cohen's laboratory. Thank you

to Sean and Paul for your continued mentorship and enthusiasm. I have greatly enjoyed my time spent working with you all.

Thank you Jenny Yang, there are not enough words to express the gratitude I have for you, your family, and your friendship. I would not have made it this far without you and I am excited for this next adventure.

Rockefeller Resource Centers are filled with the most brilliant and amazing scientists. All of whom have significantly contributed to my research endeavors during my time here at Rockefeller. Their technical expertise and enthusiasm to collaborate have been instrumental to my training and I am so thankful to them all. I would like to thank Connie and Sophie at the Genomics Core; Alison, Tao, Carlos, and Kate at the BIRC; Tom, Ji-Dung, Matt, and Doug of the Bioinformatics Core. The bioinformatics core at Rockefeller has significantly contributed to the work carried out in my thesis. Their guidance, training, and patience came at an instrumental part of my research, I am indebted to you all.

I am grateful to The Dean's office and for the opportunity to study here at Rockefeller. Cris, Stephanie, Kristen, Marta, Emily, and Sid are such an amazing team and I am thankful for all they do for me and the David Rockefeller Graduate Program. You are all so supportive and I want to thank each of you for making the PhD journey here such a wonderful experience.

Lastly, I would like to thank my family. My mother Marybeth, my sister Melissa, my father Drew, and step-mother Stacey for their endless encouragement. It has been a long path to get where I am and you have all taken it with me. Kevin, thank you for being you, helping me keep things in perspective, and loving me through the ups and downs of graduate school. I cannot wait to see what the next chapter in our lives has in store. Most importantly, I would like to thank my dog, Truro- for providing the emotional support that I never know I need.

Table of Contents

List of Figures

List of Abbreviations

ApA	Alternative polyadenylation
APLP1	Amyloid Precursor-like Protein 1
APLP2	Amyloid Precursor-like Protein 2
APP	Amyloid Precursor Protein
BPTF	Bromodomain PHD finger transcription factor
CB	Cerebellum (cortex)
CTK	CLIP Tool Kit
Cugbp2	Elav-like family member 2
DGN	Disease Gene Net
DN	Dentate Nucleus
DNM1L	Dynamin 1-like
DO	Disease Ontology
EDMD	Emery-Dreifuss muscular dystrophy
ESE or ISE	Splicing enhancer (exon or intron)
ESS or ISS	Splicing silencer (exon or intron)
FGF	Fibroblast Growth Factor
GO	Gene Ontology
GPHN	Gephyrin
GRIK2	Glutamate ionotropic receptor kainate type subunit 2
GSEA	Gene set enrichment analysis
GTF2I	General Transcription Factor II
Hg38	Human genome assembly 38
hiPSC	Human induced pluripotent stem cells
HITS-CLIP	Cross linked-immunoprecipitation followed by high throughput sequencing
iPSC	Induced pluripotent stem cell
Iqr	Interquartile ranges
KH	K homology
KIDINS220	Kinase D-interacting substrate, 220-KD

KMT2E	Lysine methyltransferase 2E
KO	Knockout
LCA	Last common ancestor
LUHMES	The Lund human mesencephalic cell line
NES	Nuclear export signal
NHP	Non-human primates
NLS	Nuclear localization signal
NMD	Nonsense-mediated decay
Nova	Neuro-oncological ventral antigen
NPCs	Neural precursor cells
PAS	Polyadenylation sites
PC	Purkinje-cell
PNDs	Paraneoplastic neurologic degenerations
polyA	polyadenylation
POMA	Paraneoplastic Opsoclonus Myoclonus Ataxia
PSI	Percent splice inclusion
RBPs	RNA binding proteins
RPA	RNase protection analysis
SH-SY5Y	Human neuroblastoma cell line
Slc8a1	Solute carrier family 8 member A1
SMARCE1	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Chromatin, Member 1
SNPs	Single nucleotide polymorphisms
Srgap2	Slit-Robo Rho GTPase-activating protein
SSOs	Splice switch oligonucleotides
SYNE2	Spectrin Repeat Containing Nuclear Envelope Protein 2
UTRs	Untranslated regions
WBS	Williams-Beuren syndrome

Chapter 1: General Introduction

RNA binding proteins and human disease

The central dogma of molecular biology is the flow of information from DNA to RNA to protein. Each cell contains a DNA "blueprint", which is transcribed into RNA by the protein RNA polymerase II. This working copy of RNA is exported from the nucleus to the cytoplasm where the encoded information is translated into protein by ribosomes. So how does the same set of DNA give rise to so many different cell types with such diverse functional properties? Through multiple layers of gene regulation: (1) Transcriptional regulation ensures that genes are turned on/off appropriately. The activity of certain genes determine cellular identity and dictate functional properties across cell-types. (2) Post-transcriptional regulation occurs through RNA processing and modulates gene expression levels. Post-transcriptional RNA-processing is an important process in eukaryotes and allows for precise control over RNA expression levels in time and space. The proteins which regulate these intricate processes are a group of evolutionarily conserved proteins called RNA binding proteins (RBPs). There are ~1,500 RBPs annotated in the human genome and together these RBPs dictate RNA life cycles within cells with contributions across all stages of RNA processing, from production and maturation, to stabilization, degradation, localization, and translation. Although many RBPs are universally expressed, some RBPs are tissue-specific. Tissue-specific RBPs are often identified within more specialized tissues and contribute to cellular diversity and functional complexity of the tissue. Balance in RNA expression networks is important for maintaining stoichiometry in cellular RNA

steady state levels and RNA binding proteins have the capacity to tip the balance from a normal to disease state (Gebauer et al., 2021; Lukong et al., 2008; Musunuru, 2003).

Neurons are long-standing, terminally differentiated, and highly polarized cells, therefore requiring unique molecular demands for maintenance, function, and survival. RNA-binding proteins play a key role in maintaining the neuronal cell state via regulation of RNA metabolism, from transcription to translation. Aberration of RBPs leads to disruptions in RNA processing and subsequent disease. There are multiple ways in which this can occur and examples include altered RBP expression, mislocalization of RBPs, RBP sequestration, and RNA expansions. Taken together, slight changes in an RBP's function, whether it be due to genetic mutation or cellular stress, can have drastic and long-term consequences for the health of a neuron (Gao and Paul Taylor, 2012; Nussbacher et al., 2019; Zhou et al., 2018). Current approaches to identify disease-relevant gene expression changes in human tissue relies heavily on bulk RNA-seq in a healthy and diseased state. Limiting the focus to significant changes in single genes or predicted gene pathways. The increasing evidence for the dysregulation of RBPs and network level disruption in the context of human neurological diseases emphasize the importance of mapping *in vivo* RBP regulatory networks.

Alternative RNA processing mechanisms

Alternative Polyadenylation

In 1971, polyadenylation (polyA) was first discovered in polyribosomal RNA and described as an RNAse resistant adenosine-rich portion of the RNA (Darnell et al., 1971; Edmonds et al., 1971; Lee et al., 1971). Subsequent studies revealed that the majority of eukaryotic mRNAs undergo the two step process of endonucleolytic cleavage and polyadenylation as an essential step in maturation (Proudfoot, 2016; Shi and Manley, 2015). The cleavage site, or poly(A) site (PAS), is delineated by neighboring sequence motifs. The strength of the PAS is determined by the surrounding landscape of PAS motifs. Mammalian sequence motifs include: UGUA, A[A/U]UAAA hexamers, and U-rich motifs (Wang et al., 2018). More recent research has shown that alternative cleavage and polyadenylation sites are prevalent in the mammalian genome and give rise to mRNA isoforms with differential 3'UTR lengths and/or coding sequences (Figure 1-1) (Shepard et al., 2011; Wang et al., 2018). The 3'UTR of an mRNA is hugely concentrated with RNA *cis*-regulatory elements that are necessary for post-transcriptional regulation of expression levels by RNA binding proteins. Alternative polyadenylation can act as a mechanism to promote/inhibit the ability to downregulate expression given that longer 3'UTRs are more subject to microRNA and ARE mediated suppression. In this way alternative cleavage and polyadenylation site usage directly impacts mRNA metabolism including stability localization, and translation. Alternative regulation of 3'UTRs is specifically unique in neurons, as they have been shown to express preferentially long 3'UTRs as a direct consequence of neuronal RBP regulation (Licatalosi et al., 2008; Wang et al., 2008).

Figure 1-1 Alternative polyadenylation sites (PAS)

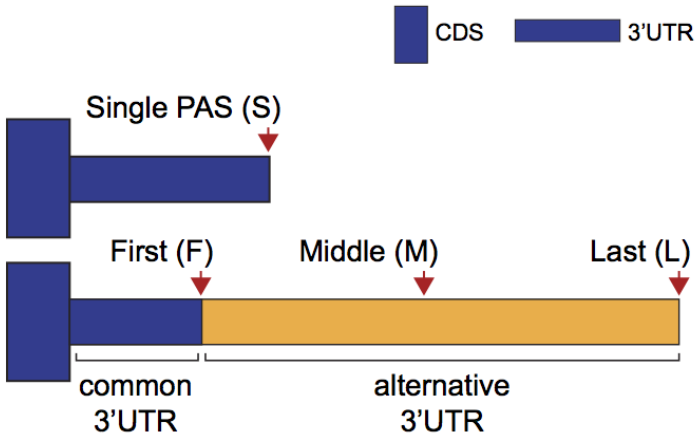


Figure 1-1: Different types of polyadenylation signals in the 3'UTR: single (S); first (F); middle (M); last (L). Typically the area before the first PAS in a 3'UTR is termed common UTR and the region after the first PAS is referred to as the alternative UTR.

Alternative Splicing

In 1977, Richard Roberts, Phil Sharp and colleagues were first to discover introns. Prior to 1977 the assumption was that the total number of genes determines the complexity of an organism, but there were obvious issues with this hypothesis for example: mice, humans, and mustard plants all have about 25,000 total genes (Figure 1-2). It is clear that humans are more complex than mice and mustard plants, however the total number of genes fails to capture such complexity. However when you look at the total number of introns for each organism, the hierarchy is restored. So how might introns give rise to organism complexity? One advantage of introns is that through non-homologous recombination, large sections of the genome can be taken from one area to another without the need to be "perfectly" incorporated since splicing machinery has the ability to cut and paste these regions accordingly. In this way, introns are able to facilitate the evolution of new genes.

Figure 1-2 Introns drive complexity

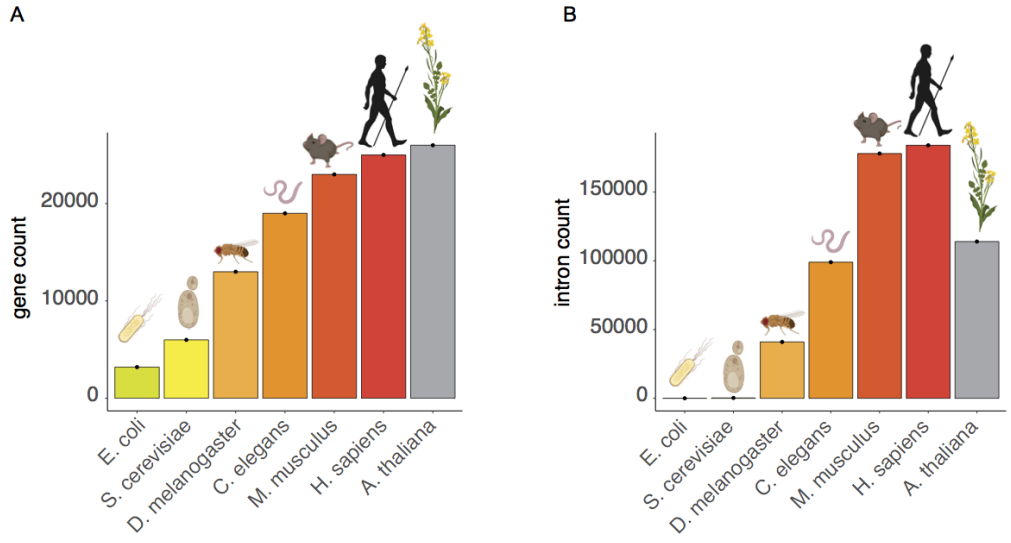


Figure 1-2: (A) Total number of genes per organism. X-axis is ordered from least to most gene counts. (B) Total number of introns per organism.

Alternative splicing was first described in 1980, just a few short years after the discovery of introns. Alternative splicing is another way for eukaryotes, including humans, to compensate for a lack in genic complexity (Figure 1-3). The splicing of pre-mRNA is an intricate process requiring the coordinated actions of a large number of protein factors. In the eukaryotic nucleus pre-mRNA is first recognized by the splice apparatus which includes small nuclear RNAs, snRNAs, and small nuclear ribonucleoproteins, snRNPs (U1, U2, U4, U5, and U6) (Krainer, 1988; Krainer and Maniatis, 1985). The snRNPs interact directly with the pre-mRNA to create the basal machinery mediating splicing. In the first step of splicing the branch site's (conserved adenosine) 2'hydroxyl group attacks the phosphate at 5' splice site which results in 5' splice site cleavage and formation of lariat intermediate, which has 3 phosphates. At this time the splicing machinery holds onto the free 5' exon. The second sequence of events occurs when the 3' hydroxyl, generated during the first step on 5' exon, attacks the phosphate at 3'ss to kick out the intron. Simultaneous to intron excision, the 3' exon is ligated to the 5' exon, and both events are made possible via chemical steps called concerted trans- esterifications. These two steps of splicing are catalyzed by the spliceosome, which consists of 4 major subunits- U1, U2, U4/5/6, NTC (nineteen complex). In short, U1 interacts with 5' splice site, U2 base pairs and recognizes the branch site sequence (A complex), and U4/5/6 comes in to form B complex. One of the Snrps is kicked out and the NTC comes in to finish the splice by removing the intron (Ruskin et al., 1984; Wahl et al., 2009).

Figure 1-3 Alternative splicing

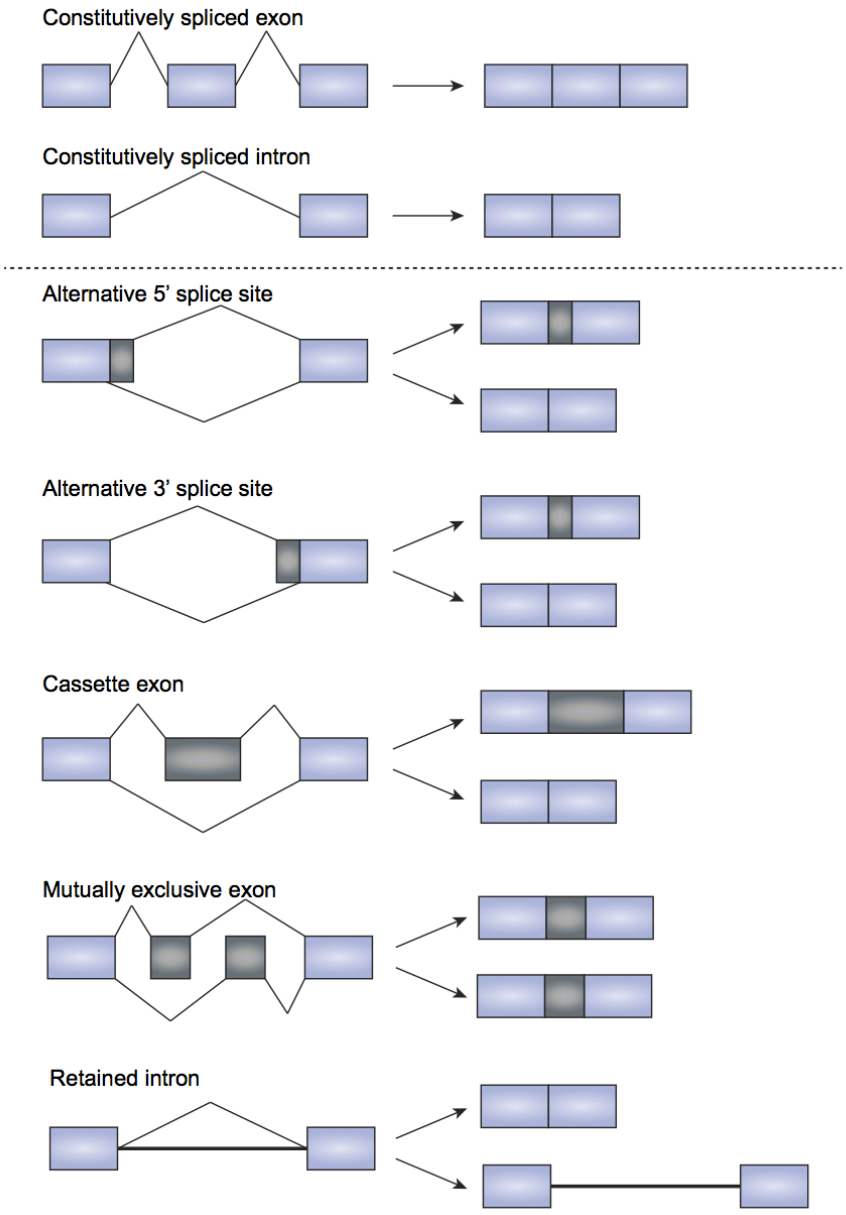


Figure 1-3: Types of commonly observed splicing events. (A) Types of constitutive splice events. (B) Types of alternative splice events.

How does a cell know where the introns are and how is the splice site chosen? The way that introns and exons are recognized in human genes is through exon recognition since the exons tend to be very uniform in their length, the human splice machinery looks for a good match to branch and consensus splice site followed by 25-300nts distance to the 5' splice site. However, this is still not enough information to uniquely identify the exon and also not enough information for alternative splicing to take place. Lee Lim and Chris Burge studied the shortest introns and found that the human introns have far less information in their consensus sequence than yeast introns (Lim and Burge, 2001). The study investigated relative contributions of intron features across different organisms and in budding yeast it all comes from 5'ss, 3'ss, and branch site. However, in plants and humans the 5'ss, 3'ss, and branch site was insufficient to distinguish short introns. Concluding that there must be additional factors at play in humans such as length and composition. Splicing enhancer (ESE or ISE) elements and splicing silencer (ESS or ISS) elements are sequences that often cluster around exons. SR proteins tend to enhance splicing and hnRNP proteins tend to act as silencers, and it's all about upholding the balance between these splicing events to maintain a balanced RNA metabolism (Mayeda et al., 1993; Sun et al., 1993).

Alternative RNA processing both expands and diversifies the transcriptome, thus giving rise to a more heterogeneous proteome. Alternative processes such as splicing and polyadenylation allow for the production of multiple transcripts with distinct coding and regulatory sequences from a single gene, thus impacting cellular function and acting as a key player in most, if not all, biological processes. Unsurprisingly, the human brain endures a higher frequency of alternative events as compared to other mammalian tissues, making neural specific RNA binding proteins uniquely interesting, especially in the context of human neurological disease (Licatalosi and Darnell, 2006, 2010). RNA is not only a template for protein synthesis

but it is also considered a regulatory molecule as it is the foundation for biological complexity. Furthermore, there have been many human hereditary disorders that arise as a direct result of mutations affecting exon inclusion and splice site selection and thus disrupt RNA homeostasis.

Neuro-Oncological Ventral Antigen Proteins: A historical perspective

Nova in the '90s: disease to discovery

Paraneoplastic neurologic degenerations (PNDs) are cancer-associated neurodegenerative disorders in which patients present with a broad spectrum of highly specific neurologic symptoms prior to clinical presentation of a neoplasm, localized outside of the nervous system (Darnell and Posner, 2003a; Posner, 2003). PNDs arise when the tumor begins to express proteins that are normally restricted to the brain, resulting in an immunological response where autoreactive antibodies are made against the antigen, which in this case are neuronal proteins. When the immune response finds its way to the cerebrospinal fluid, the autoantibodies begin to attack the antigen in its normal expressing state- i.e. the brain. Autoantibodies attack and cause neuronal death in the brain, which give rise to presenting neurologic symptoms. PNDs manifest at the juncture of cancer and neurology and have been described as the "best model for naturally occurring tumor immunity in humans" (Darnell and Posner, 2003b). Therefore, in order to better characterize PNDs, antibodies were isolated from patients and used to identify the target antigens.

PND antisera screening of cDNA expression libraries lead to the discovery of Neuro-oncological ventral antigen (Nova) proteins (Buckanovich et al., 1993; Yang et al., 1998). The Nova protein family is composed of two highly conserved neuronal RNA binding proteins

Nova1 and Nova2. Both proteins were first clinically described in Paraneoplastic Opsoclonus Myoclonus Ataxia, or POMA syndrome, which is most commonly associated with lung, breast, and gynecologic tumors. Briefly, Nova expressing cells undergo auto-immune attack and patients present with a loss of inhibitory motor control manifesting as rapid irregular eye movements (opsoclonus), muscle spasms (myoclonus), and inability to coordinate muscles (ataxia) (Buckanovich et al., 1993, 1996; Darnell, 1996).

Buckanovich et al. describe Nova1 as a sequence specific neuronal RNA binding protein containing three K homology (KH) RNA binding domains. Immunohistochemical studies probed Nova1 expression patterns in the mouse brain to show that Nova1 is both nuclear (strong) and cytoplasmic (weak) with expression restricted to subcortical regions of the CNS. Most abundantly expressed regions included the midbrain, brainstem, spinal cord, and cerebellum. All of which correlate with the clinically affected regions in POMA patients. Developmental characterization of Nova1 expression showed a progressive increase in isoform complexity paralleled with an increased restricted patterning of Nova1 expression with age. Findings that indicate high levels of alternative regulation of Nova1 at the RNA level. Altogether, the seminal studies of Buckanovich et al. set the stage for the succeeding decades of Nova research (Buckanovich and Darnell, 1997; Buckanovich et al., 1993, 1996).

Interestingly, the discovery of Nova2 occurred through examining a clinically unique subset of POMA patients with additional neuropathologies including encephalopathy, dementia, and cerebral atrophy (Yang et al., 1998). While the biochemical properties of Nova1 and Nova2 were found to be mostly insignificant, their expression patterns are described as largely

reciprocal throughout the central nervous system. Nova2 is widely expressed, but the strongest expression was found in the cortex, olfactory bulb, and thalamus. Therefore, differential expression patterns between Nova1 and Nova2 may be a contributing factor to the additional phenotypes observed for Nova2 targeting POMAs. By the turn of the century, Nova RNA binding proteins were characterized as neuron-specific regulators of alternative splicing specifically interacting with RNA via 'YCAY' repetitive elements. A role that has defined nearly three decades of Nova protein research.

Linking biochemistry and physiology through mouse models

One primary goal of biological research is to elucidate physiological function from identified proteins. This is a formidable task, especially when it comes to neuronal proteins, because each protein interacts with and functions with thousands of other proteins in every neuron, and there are an estimated 100 billion neurons in the human brain. Furthermore, in order to communicate and maintain plasticity in the brain each neuron makes tens of thousands of contacts with other cells forming highly complex integrated communication networks. These connections are dynamic and are constantly changing the molecular metabolism of each neuron. One approach to surpass the mind-boggling complexity is to remove the cell from the normal environment into a simpler system for experimental purposes. However, while *in vitro* cellular systems are extremely useful and well controlled, the conclusions one can make are hugely restricted.

Technological advancements in genetically modified mouse models have been an instrumental tool for understanding physiological function of proteins. Genetic deletion of a

single gene can provide a great deal of insight towards gene function through careful characterization of the spatial and temporal consequences. Nova knockout mouse (KO) models allowed for the validation of previous *in vitro* biochemical studies in addition to system-wide phenotypic characterization of the functional consequences in the absence of Nova (Dredge and Darnell, 2003; Dredge et al., 2005; Jensen et al., 2000a, 2000b; Lewis et al., 2000). Nova-KO mice were found to recapitulate some of the motor deficits observed in POMA patients. Nova1-KO mice display severe growth retardation and gross motor failure as a result of apoptosis of the brainstem and spinal cord (Jensen et al., 2000b). Nova2 deficiency also causes severe growth retardation in addition to agenesis of the corpus callosum, outgrowth defects in ventral motor neuron axons, and reduction in cochlea innervation (Leggere et al., 2016). Both Nova1-KO and Nova2-KO mice die within two weeks of birth and double Nova-KO mice are born paralyzed and die immediately at birth (Ruggiu et al., 2009). Both models have provided evidence for compensatory mechanisms between Nova1/2 proteins in populations expressing both Novas.

An additional seizure phenotype was observed in Nova2-KO mice that was unique to Nova2. Nova2-KO mice had deficits in activity dependent long-term potentiation of synaptic excitation and slow synaptic inhibition (Huang et al., 2005). The inhibitory/excitatory network dysfunction was dose-sensitive as Nova2 heterozygous mice have frequent bilateral synchronous cortical epileptiform discharges as well as spontaneous occurrence of generalized seizures. Additionally, a nuclear localization signal (NLS) and nuclear export signal (NES) facilitates shuttling of Nova between the nucleus and cytoplasm in an activity dependent manner, which was the first physiological evidence of a non-nuclear function for Nova proteins (Racca et al., 2010).

Overall findings in the mouse conclude that global dysregulation of Nova1/2 leads to widespread neurologic dysfunction; and aberrations in Nova1/2 expression levels lead to widespread changes in RNA expression networks required for proper neuronal development and physiology.

Mapping transcriptome-wide Nova-RNA regulatory networks

In parallel to phenotypic and structural characterization of Nova-KO mice, development of *in vivo* biochemical approaches were greatly improving the feasibility to link molecular mechanisms with phenotypic consequences. Cross-linked immunoprecipitation (CLIP) involved UV-irradiating live cells to "freeze" RNA binding proteins in place. The frozen protein bound RNAs could then be isolated via immunoprecipitation using targeted antibodies. Radioactive labelling of the protein bound RNA allowed for visualization and gel purification of the protein bound RNA "footprints". These "footprints" could then be mapped back to the genome to identify novel functionally bound RNA regulatory elements with nucleotide specificity (Musunuru and Darnell, 2004; Ule et al., 2003, 2005). Coupling the pivotal *in vivo* CLIP approach with microarray proved Nova-dependent splicing "networks" functionally reshape neuronal synapses.

In addition to providing a quantitative link between molecular mechanism and physiological consequences, the resolution that CLIP methodology provided revealed reproducible binding "patterns" relative to alternative exons that directly correlated with observed splice outcome. Thus making Nova proteins the first described position-dependent

regulators of alternative splicing (Ule et al., 2003, 2006). Computational modeling tools demonstrated the utility of CLIP binding maps with target prediction analysis that could predict Nova dependent splice regulation based on Nova binding rules, upstream Nova binding leads to more exon skipping, and more downstream binding leads to more exon inclusion. These binding sites were described as Nova splice "silencers" and "enhancers" (Ule et al., 2003, 2006).

The availability of high throughput sequencing technologies swiftly expanded the breadth of RBP "maps" and the efficiency in which they were able to be generated. Genome-wide Nova-RNA interaction maps rapidly expanded our knowledge of Nova regulated splice networks (Figure 1-4). Additionally, the power of HITS CLIP enabled the discovery of new Nova biology by linking position specific Nova regulation with alternative polyadenylation site usage (Licatalosi et al., 2008). Furthermore context dependent and disease relevant discoveries detail the importance of activity-dependent Nova regulation of cryptic nonsense-mediated decay (NMD) exons to control synaptic protein levels post-seizure (Eom et al., 2013). Improvements in the generation of cell-type specific mouse models and tagging capabilities truly highlighted the specificity and complexity of Nova regulatory networks in the brain by descriptions of Nova networks unique to certain brain regions, developmental timepoints, and neuronal subtypes (Leggere et al., 2016; Saito et al., 2016); (Yuan et al., 2018); (Saito et al., 2019).

Generation of genome-wide binding maps have revealed that Nova regulates functional RNA networks specifically in neurons. The culmination of these studies also suggests that Nova expressing neurons are highly specialized in modulating both cognitive and motor systems. Therefore, given what is known one would hypothesize that Nova's regulatory role is largely

Figure 1-4 Schematic of NOVA-CLIP protocol

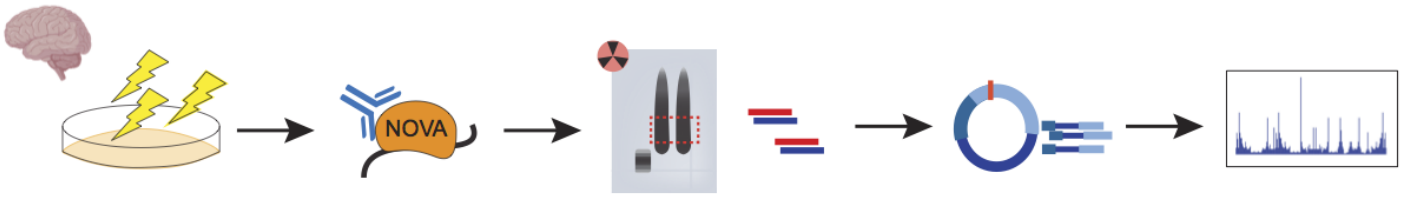


Figure 1-4: Flash frozen post-mortem human brain samples were ground to powder while remaining frozen. The frozen powdered tissue was UV-irradiated on dry ice three times at 400mJ/ cm^2. Sample was lysed and RNA was partially digested. NOVA-RNA complexes are immuno-precipitated, radiolabeled, and purified via SDS-PAGE and transferred to a nitrocellulose membrane for visualization. NOVA-RNA complexes are size-selected for via membrane dissection and NOVA bound RNA fragments are reverse-transcribed, barcoded, circularized and amplified for high-throughput sequencing.

similar from mouse to human and likely functions in a similar capacity since the basic functions and even phenotypes associated with Nova proteins are more or less conserved. However, since the RNA binding capacity of Nova is consistent throughout evolution, in that it interacts with RNAs through YCAY elements, one may conclude that with fixed KH domains, YCAYs would be flexible and change for Nova. So even though these models have facilitated the unbiased discovery of Nova-RNA interactions in the mouse and have been essential towards elucidating Nova biology, it is probable that the exact interaction sites are highly diverged between species. Therefore, translating Nova binding events from mouse to human genomes is challenging and often unreliable, which highlights the inherent need to define exact conserved vs. divergent regulatory elements. These data would facilitate the discovery of new human-specific Nova interactions which could be used to better understand human brain biology and aid in the treatment of neuro-cognitive disorders unique to humans. Addressing the knowledge gap requires biochemistry directly on human brain tissue in order to first independently identify functional Nova-regulatory elements in humans.

Chapter 2: Novel *in vivo* genome-wide NOVA-RNA interaction maps in human brain with regional specificity

Introduction

Identifying the molecular basis for features that make us human is one of the biggest challenges in biological research. Roughly 5-10 million years ago, human ancestors began to diverge from the last common ancestor (LCA) of chimpanzees and in such time nearly 20 million distinct changes in genomic loci have emerged between chimpanzees and modern humans, these genetic changes are referred to as human-specific regions of the genome (Chimpanzee Sequencing and Analysis Consortium, 2005; Pollard et al., 2006). Fossil record provides is evidence of multiple hominin lineage branches emerging after the LCA with non-human primates (NHP), but only one has survived to present day (Kuhlwilm et al., 2016; Wood and Collard, 1999). An estimated 600,000 years ago was the first indication of culturally and anatomically distinct humans. At which point it appears there was a divergence in the "archaic" from "modern" human lineages (Kuhlwilm and Boeckx, 2019).

Advancements in high-throughput sequencing methods and computational approaches has allowed molecular geneticists to apply comparative genomics to identify thousands of genomic loci (~30,000 single nucleotide polymorphisms (SNPs)) that are novel to modern humans, having occurred after the split from archaic human lineage. Analysis reveals about 10% of these changes fall within regulatory regions, consequently altering only 87 proteins, one of which is located within the RNA binding domain of RNA binding protein, NOVA1 (Capra et al.,

2013; Castellano et al., 2014; Kuhlwilm and Boeckx, 2019; Pääbo, 2014). The human-specific genetic change of NOVA1 results in an isoleucine to valine amino acid change in the second K homology (KH) domain. The cocrystal structure of NOVA1-RNA interaction suggests that while the human-specific position may not interact with RNA directly, it is likely important for protein-protein recognition and dimerization related to RNA-looping mechanisms (Teplova et al., 2011). Altogether hinting towards human-specific features of NOVA-RNA interactions in the brain that remain unknown.

Human genetics approaches have been instrumental in identifying genomic features of evolutionary importance. For example, the importance of gene duplication events for human brain evolution is exemplified by the Slit-Robo Rho GTPase-activating protein (Srgap2) (Charrier et al., 2012). There is only one form of the gene in mice, but humans have four forms due to incomplete segmental duplication (Dennis et al., 2012). Overexpression studies of the human paralogs demonstrate phenotypes in radial migration, spine density, and spine maturation, all of which are human-specific characteristics of neuronal development (Bae et al., 2015). Srgap2 is a HITS-CLIP defined Nova target in the mouse brain and exemplifies the need to generate Nova HITS-CLIP binding maps in the human brain (Saito et al., 2019). Human Nova binding maps will provide mechanistic insight towards how RNA regulation has evolved within specific gene families. Complexity at the RNA level in the regulation of duplicated gene expression profiles may even contribute to observed human phenotypes specifically in neurons.

Recent large-scale transcriptomic studies have only just begun to focus on comparing gene expression patterns across species, tissues, and time (Gordon et al., 2021; Hodge et al.,

2019; Pembroke et al., 2021). Quantitative assessment of co-expression networks have revealed a great deal of insight in regards to human-specific divergence patterns. Altogether, findings suggest human-specific features converge within the human cerebral cortex. The cortex governs attentional control and is responsible for working memory, planning, temporal integration, and volition. To modulate complex human behaviors, the cortex is intricately connected to the rest of the brain. It is essential for all interconnected regions to function in sync with one another in order to achieve coordinated functions and specializations. The frontal- subcortical circuitry of the brain controls one's ability to recognize when to initiate or cease any given behavior. If the ability to shift from one state to another is ablated, in the absence of dynamic circuit control, one will present with non-adaptive, rigid behaviors. Unfortunately, studies involving the cortex are challenging as much of it's anatomical structure is unique to humans.

RNA binding proteins dictate expression profiles and drive mechanisms of alternative RNA processing which permits for additional layers of control. Alternative RNA processing allows for transcriptomic expansion, contributes to cellular diversity, and facilitates plasticity in RNA steady state levels. Therefore, this section aims to profile Nova regulation using HITS-CLIP methods in cortical and cerebellar human brain regions. These investigations detail *in vivo* NOVA-RNA interactions with regional specificity, and provide mechanistic insight towards neuronal expression networks unique to the human brain and cognition.

Results

Data generation, complexity, and depth

To characterize NOVA-mediated RNA regulation and define NOVA-RNA interaction sites in the human brain, we performed cross linked-immunoprecipitation followed by high throughput sequencing (HITS-CLIP) for both cortical (BA9 and BA46) and cerebellar (cerebellar cortex and cerebellar dentate nucleus) brain regions using post-mortem brain tissue collected across 13 cognitively normal post-adolescent individuals (ranged 24-36 years of age). For additional sample information see Chapter 9 Methods section. In short, flash frozen post-mortem brain tissue was crosslinked, lysed, and unbound RNA was digested. NOVA-RNA complexes were immunoprecipitated using a polyclonal antibody raised against a c-terminus NOVA fusion protein.(Licatalosi et al., 2008) The immunoprecipitated RNA was radioactively labelled, and RNA-NOVA complexes were SDS- gel purified and transferred to a nitrocellulose membrane for visualization and size selection. RNA was purified from the membrane, reverse transcribed, circularized, PCR amplified, and CLIP libraries were sequenced. The resulting sequencing reads were de-multiplexed, aligned to the human genome (hg38), and bioinformatically processed to identify unique CLIP tags representative of NOVA-RNA interaction events.

Each HITS-CLIP experiment relies on radiolabelling of the RNA in order to visualize Nova-RNA interactions following immunoprecipitation. In order to size-select Nova-RNA interactions accurately, one lysate per immunoprecipitation experiment is treated with high

RNase so that the bulk of RNA is degraded and autoradiogram signal collapses to roughly the size of Nova protein, ~55 kDa (Figure 2-1A, Lane 4). Digesting human brain lysate with a lower concentration of RNase results in non-uniform small RNA fragments, visualized as a smear just above the 52 kDa mark. This approach permits for the downstream identification of Nova binding sites with nucleotide sensitivity. Pearson Correlation coefficients were calculated using Nova-CLIP tags across experimental replicates (Figure 2-1B). Significant positive correlations (R^2 0.73-0.78 between experiments) demonstrate the reproducibility of the experimental approach.

Pooling all unique CLIP tags together (Total = 33,717,995 unique tags) identified 138,323 significant *in vivo* NOVA binding sites in the human brain. Genomic annotation of all significant peaks revealed 35% (53374/138323) of significant binding is localized to introns, 55% (76428/138323) localized to 3'UTRs, and 5% localized to exons (6547/138323), less than 2% of peaks localized to 5'UTRs, downstream 10k, and distal intergenic regions respectively (Figure 2-1C). The distribution of peaks per genomic location is concordant with Nova's established role as a regulator of alternative splicing and polyadenylation. As an initial unbiased approach to identify *de novo* human-specific Nova consensus motifs, we searched for enriched sequences within Nova binding sites. All top enriched motifs are similar and contain the high affinity Nova tetramer 'YCAY' (Figure 2-1D). Again, highlighting the validity and robustness of identifying Nova binding sites using the HITS-CLIP approach.

Figure 2-1 NOVA HITS-CLIP in human brain

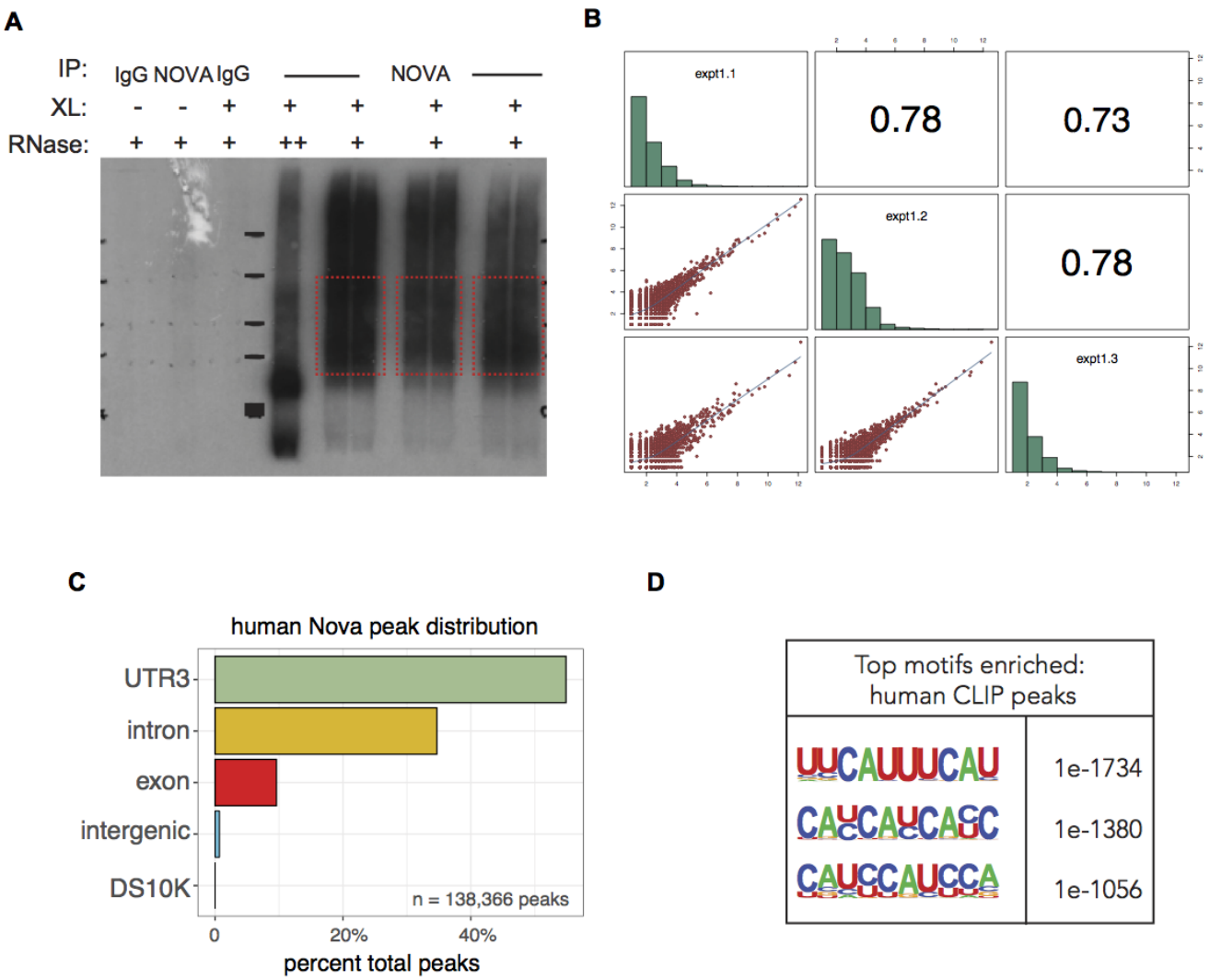


Figure 2-1: **(A)** Autoradiogram of radiolabeled UV-crosslinked NOVA-RNA complexes. Lanes from left to right: (1) non-crosslinked IgG control, (2) non-crosslinked NOVA control, (3) crosslinked IgG control, (4) over-digested control, (5-10) samples used for sequencing. Red box represents the size-selected region that was cut and further subjected to CLIP processing for downstream sequencing. **(B)** Correlation between total number of CLIP reads per peak in three experimental replicates from human brain sample. R= Pearson correlation coefficient. **(C)** Genomic distribution of CLIP peaks, represented as percent of total peaks per genomic feature: 3'UTR (green- 55%); intron (yellow-39%); exon (red-5%); intergenic (blue- 1%); downstream 10k (grey- <1%). **(D)** Significantly enriched *de novo* motifs identified within Nova bound RNA, statistical significance determined via hypergeometric testing.

Characterizing Nova's relationship with YCAY-tetramer in human brain

In order to thoroughly assess Nova binding in the human brain we sought to characterize the relationship between Nova and high affinity Nova tetramer 'YCAY' across targets and peaks. YCAY elements are significantly more abundant in human genic regions than one would expect by chance (1/64 nucleotide) when accounting for total number of nucleotides, or length (Figure 2-2A). In order to assess whether YCAY abundance differed in non-targets as compared to Nova targets, the total number of YCAY occurrences were counted for all genic regions. Each gene was annotated accordingly as "non-target" or "target" based on human brain CLIP data demonstrating that YCAY-tetramers are significantly more abundant in Nova targets relative to non-targets (Figure 2-2B). To further tease apart Nova's relationship with YCAY-tetramer relative to peaks, the boundaries of each significant Nova peak center, or peak "summit", was determined using the peak calling software CLIP Tool Kit (CTK), defined as the region (width = 1-2nt) with the greatest density of overlapping tags.(2016) Specificity of the human Nova binding map is demonstrated by the presence of at least 1 YCAY-tetramer within 50 nucleotides of >88% of all significant Nova peaks. Expansion in the distance from peak summit (100nt, and 200nt) identifies at least 1 YCAY-tetramer counted in >97% and 100% of peaks respectively (Figure 2-2C). Next, each YCAY-tetramer was bioinformatically located within 300nt of each summit and the distance from peak summit to tetramer was determined. In order to globally visualize YCAY enrichment relative to peak summit, the normalized density of all YCAY-tetramers within 300nt were plotted relative to peak summit (Figure 2-2D). These data show enrichment of YCAY density within 100 nt of human Nova-CLIP binding, further demonstrating the specificity and efficacy of HITS-CLIP data.

Nova proteins specifically interact with RNA via YCAY-elements and the number and density of YCAY motifs has the ability to fine tune Nova regulation. In order to determine whether genomic location of Nova binding is associated with changes in local YCAY abundance, CLIP peaks localized to exons, introns, or 3'UTRs were subset accordingly. The total number of YCAY-elements within 100nt of each binding site was determined and comparison across binned groups revealed that YCAY occurrence was significantly greater within intronic peaks as compared to exonic and 3'UTR peaks. Furthermore, 3'UTR peaks demonstrate a significant decrease in YCAY abundance relative to both exonic and intronic peaks (Figure 2-2E).

Figure 2-2 Characterizing high affinity Nova YCAY tetramers in humans

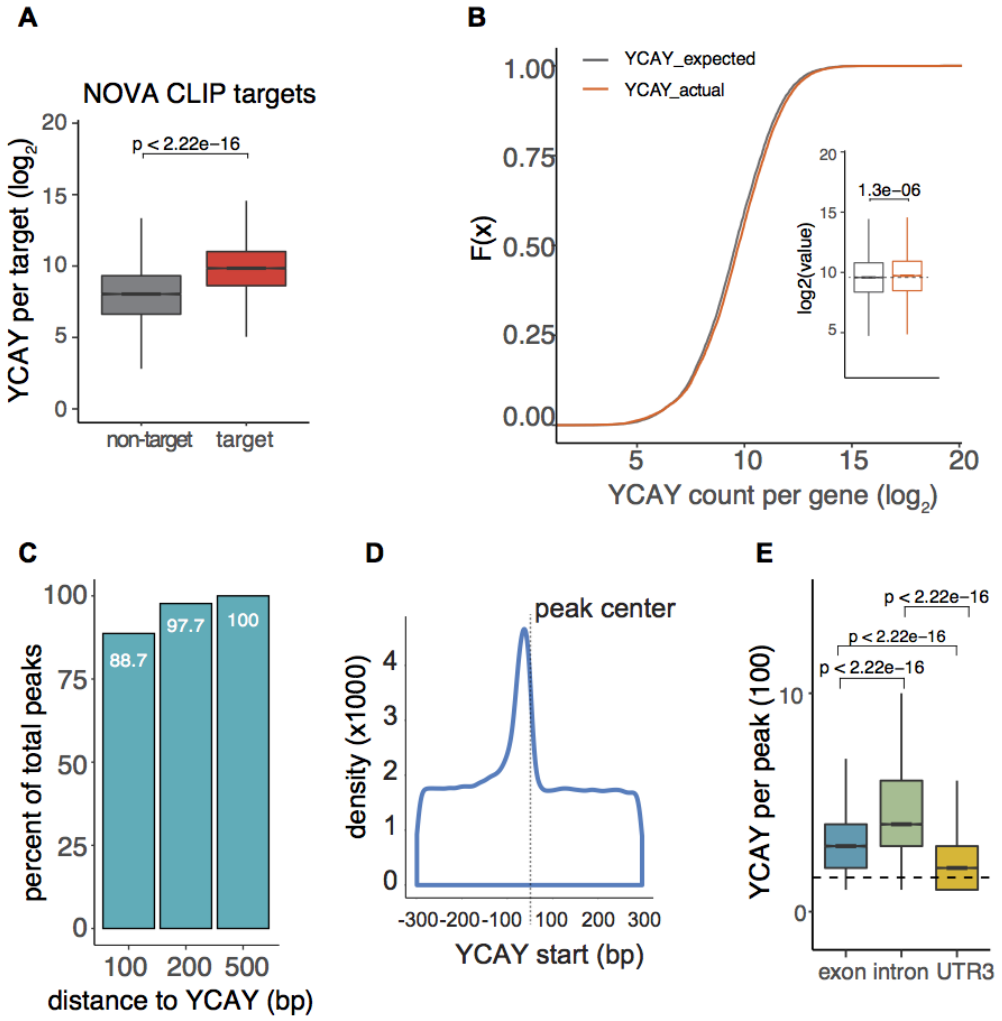


Figure 2-2: (A) Nova targets display higher YCAY abundance than non-targets in the human genome (hg38). Boxplot displaying significant shift in YCAY abundance per human target (red). Significance determined by wilcoxon ranked-sum test. **(B)** Cumulative distribution plots demonstrating on a genome-wide scale, human genes have significantly more YCAY elements per gene than one would expect by chance. Gene length was calculated for all human genes. Expected number of YCAYs (1/64 nucleotides) was determined per gene. Bioinformatic analysis of genomic sequence was carried out to determine actual YCAYs per genic region.Significance determined by wilcoxon ranked sum test. **(C)** Percent of total peaks with at least 1 YCAY tetramer within 50nt, 100nt, and 250nt from CTK defined peak center, at least 1 YCAY elements was identified within 100nt of CLIP peak center in > 97% of all significant peaks. **(D)** Density plot of genome-wide YCAY enrichment relative to distance from Nova peak center. **(E)** Intronic Nova peaks (green) display significantly higher YCAY density within 50 nucleotides of peak center compared to peaks localized to exons (blue) or 3'UTRs (yellow). Significant decrease in YCAY density is observed within 3'UTR peaks compared to peaks in exons or introns. Boxplot displaying significant shifts in YCAY density within 50nt of Nova peak center. Significance determined by wilcoxon ranked-sum test.

In order to better quantify and describe the association of Nova binding depth and YCAY density relative to genomic location, CLIP tags were normalized for read-depth and peak width. Peak width was determined via CTK software's half-peak height algorithm and interquartile ranges (iqr) were calculated in order to bin Nova peaks into categories representative of binding intensity (iqr_1 = low (white); iqr_2 = low-moderate (light pink); iqr_3 = low-high (pink); iqr_4 = high (dark pink)). Normalized Nova CLIP binding was then plotted over YCAY density, correlation coupled with linear regression demonstrates significant positive correlation between YCAY density and Nova binding depth for all genomic regions (exon; intron; 3'UTR; 5'UTR), with the strongest and most significant correlation among intronic Nova binding sites (Figure 2-3A). Comparing YCAY density across interquartile groups representing binding intensity, reveals a significant gradual increase in YCAY abundance with increased Nova binding intensity for exonic and an even stronger relationship among intronic peaks (Figure 2-3B). Global pathway analysis of the top most significantly bound Nova targets reveals Nova specifically regulates neuronal RNAs via YCAY-interaction in the human brain (Figure 2-4). There is large overlap at the network level between processes enriched in the human brain Nova binding map and Nova binding maps derived from mouse brain, suggesting that the data are robust and that Nova target networks are evolutionarily conserved from mouse to human.

Figure 2-3 Nova binding depth is positively correlated with YCAY density in introns

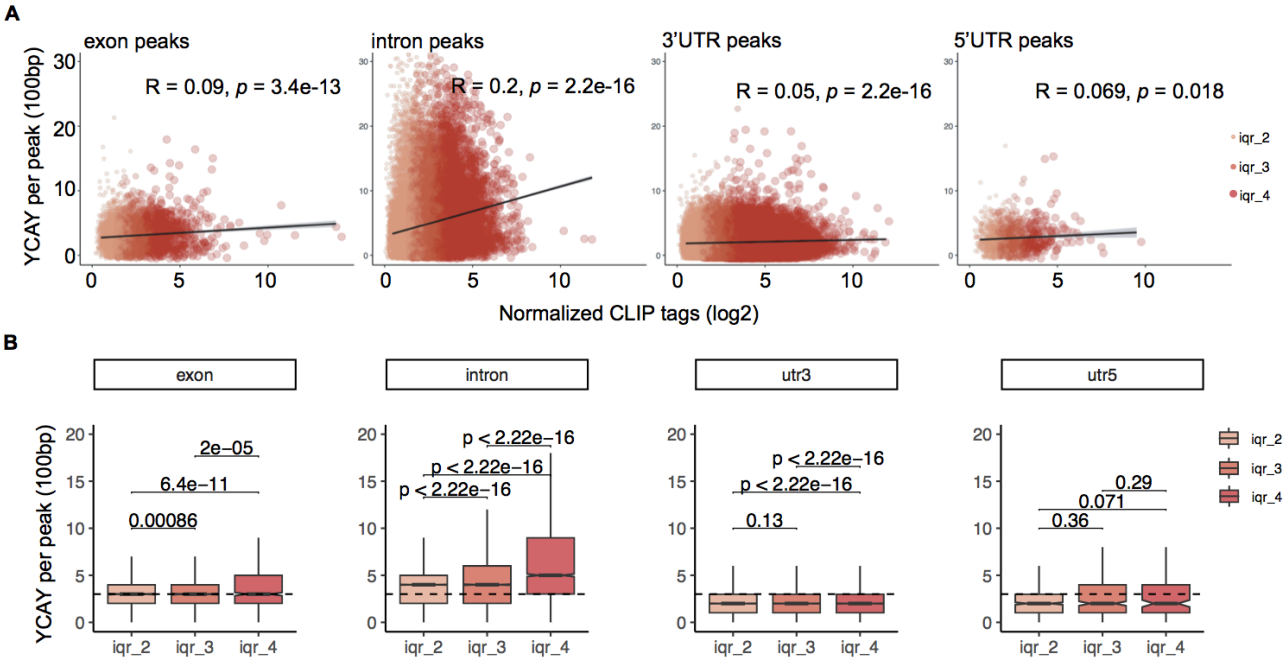


Figure 2-3: (A) YCAY abundance over normalized Nova CLIP tag density per peak across exons, introns, 3'UTRs, and 5'UTRs. Relationship of YCAY abundance relative to Nova binding depth demonstrates a significant, but weak, positive correlation in introns. Relationship of YCAY abundance and Nova binding depth in exons, 3'UTRs, and 5'UTRs is unremarkable. **(B)** Boxplots of YCAY abundance per peak within subdivided by genomic localization. Nova peaks are binned according to increased normalized Nova tag density determined via interquartile ranges of all Nova peaks (iqr_1 = low (white); iqr_2 = low-moderate (light pink); iqr_3 = low-high (pink); iqr_4 = high (dark pink)).

Figure 2-4 Nova specifically regulates RNAs enriched in neuronal functions via YCAY interaction

GO Term	Description	FDR q-value	Enrichment
GO:0045202	synapse	1.15E-30	4.09
GO:0043005	neuron projection	7.15E-21	3.35
GO:0030054	cell junction	2.03E-19	2.71
GO:0051049	regulation of transport	4.93E-18	2.5
GO:0098978	glutamatergic synapse	5.36E-18	4.36
GO:0033267	axon part	3.48E-15	4.01
GO:0032879	regulation of localization	4.75E-15	2.04
GO:0043269	regulation of ion transport	5.50E-15	3.58
GO:0050804	modulation of chemical synaptic transmission	1.24E-14	4.01
GO:0099177	regulation of trans-synaptic signaling	1.55E-14	4.01
GO:0044459	plasma membrane part	2.70E-14	2.1
GO:0034762	regulation of transmembrane transport	2.75E-14	3.68

Figure 2-4: Gene Ontology (GO) analysis carried out on top most significantly bound human NOVA CLIP targets, defined by normalized tags in peaks.

Reproducibility and variance across experiments, individuals, and brain regions

Nova HITS-CLIP methods were utilized to investigate Nova-RNA interaction sites in post-mortem human brain tissue for the generation of human-specific Nova binding maps with regional-specificity. Cross-linked and non-crosslinked brain tissues were processed and prepared as described. As observed previously from the human brain immunoprecipitation, a treatment of high concentration RNase yielded a collapsed band at the expected size of Nova protein, ~55 kDa (Figure 2-5A, Lane 4). All RNA signals were Nova and crosslink dependent, as no signal was observed in any of the non-crosslinked or IgG control lysates (Figure 2-5A, Lanes 1-3). Reaffirming the validity of the HITS-CLIP approach across all human brain regions profiled.

RNA fragments were purified from the membrane and sequencing libraries were prepared and sequenced according to Nova HITS-CLIP protocol described. All raw sequencing reads were filtered for quality, demultiplexed, collapsed for duplicates, and aligned to the human genome (hg38). Cumulatively across all experiments and replicates yielded between 3.55 and 8.67 million unique mappable reads per brain region (Figure 2-5B). All unique tags from biological and experimental replicates were combined per region in order to bioinformatically call significant peaks with region-specificity. Significant peaks were summarised for brain region profiled accordingly (Figure 2-5C). The number of significant peaks identified per brain region were comparable to the total number of unique CLIP tags yielded per region with the greatest number of peaks identified in cortical region, BA46.

Figure 2-5 Region specific NOVA HITS-CLIP in human brain

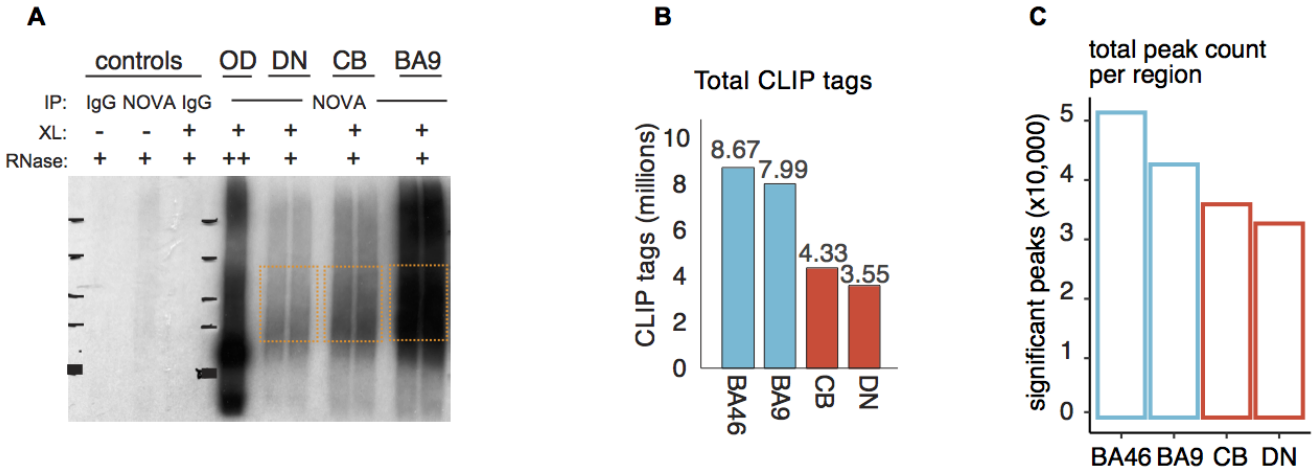


Figure 2-5: (A) Representative autoradiogram from a single human brain NOVA-CLIP experiment including 3 out of 4 brain regions profiled. **(B)** CLIP read depth per brain region, biologic complexity (BC) of cortical regions (blue): BA46 (BC=5) and BA9 (BC=5) and for cerebellar regions (orange): dentate nucleus (DN) (BC=6) and cerebellar cortex (CB) (BC=3). **(C)** Barplots represent the total number per region: of unique CLIP tags per human brain region CLIP'd (left) and significant peaks bioinformatically determined (right).

For the purpose of demonstrating the reproducibility of human Nova-CLIP across experiments unique tags per sample were aligned to all significant peaks. Total tags in peaks were summarised and only peaks with > 5 unique tags were considered for calculating correlation coefficients. Sample to sample comparisons reveal a positive correlation between datasets with R values ranging between 0.52 and 0.94 (Figure 2-6). Unsurprisingly, the strongest correlations were observed between samples prepared from the same brain tissues (cortical (BA46 and BA9) and cerebellar (CB and DN)). The lower correlations were observed across these tissues and may be attributed to differences in cellular environments or expression profiles between brain regions.

Figure 2-6 Reproducibility of Nova-CLIP across human brain regions

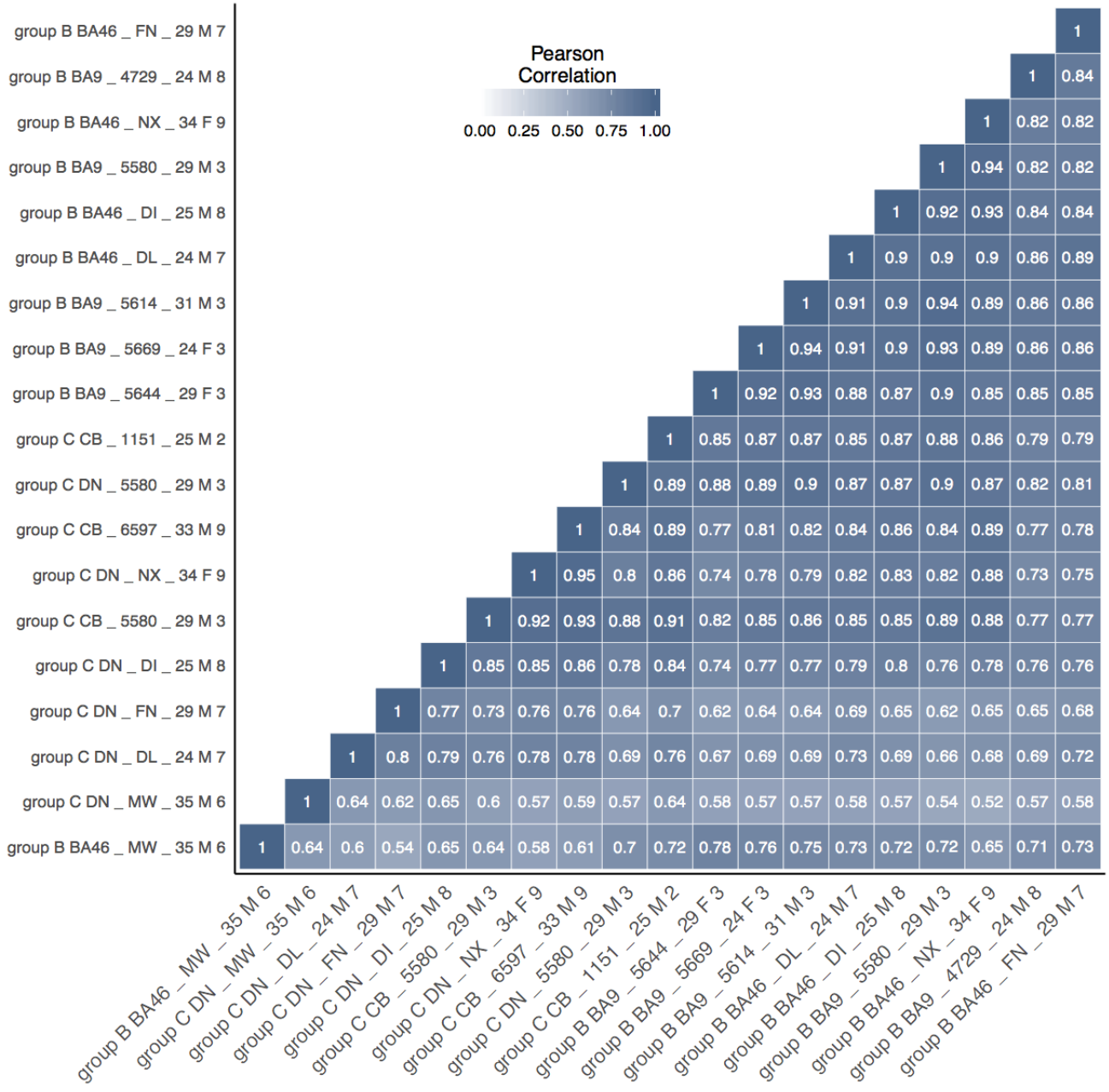


Figure 2-6: Hierarchical clustering of Pearson Correlation Coefficients (R) for total number of CLIP tags in peaks per library. Cortex includes BA46 (BC=5) and BA9 (BC=5), cerebellum refers to dentate nucleus (BC=6) and cerebellar cortex (BC=3). Square color is indicative of strength of sample to sample correlation, and coefficients are shown in white.

In order to assess variance across datasets, Principal Component Analysis (PCA) was carried out on normalized tags in peaks for all samples profiled. PCA identified principal component 1 (PC1) contributing to 37% of observed variance, PC2 contributed 17% of total variance and PC3 contributed 10% of total variance. PC3 is plotted over PC1 in order to visualize PC1's contribution as PC2 confounded the effect (Figure 2-7A). Significance of correlations between PCA scores was determined and categorical variables were tested using Spearman's correlation for continuous variables(Figure 2-7B). The results highlight the top two variables most significantly contributing to PC1 variance is tissue (BA9; BA46; CB; DN) and region (cerebellum vs. cortex). For further evaluation of Nova binding within cerebellar and cortical tissues, the total sum of all tags per peak contributed by each brain region (BA46, BA9, CB, and DN) were determined and pairwise correlation coefficients calculated (Figure 2-7C). The global distribution of Nova tags in peaks from each of the four brain regions were reproducible, with correlation coefficients between 0.54 and 0.93, indicating that correlation of Nova tags in peaks are variable according to brain region. A finding concordant with sample to sample correlation analysis and suggests regional-specificity in the data.

Figure 2-7 Global assessment of variance across samples

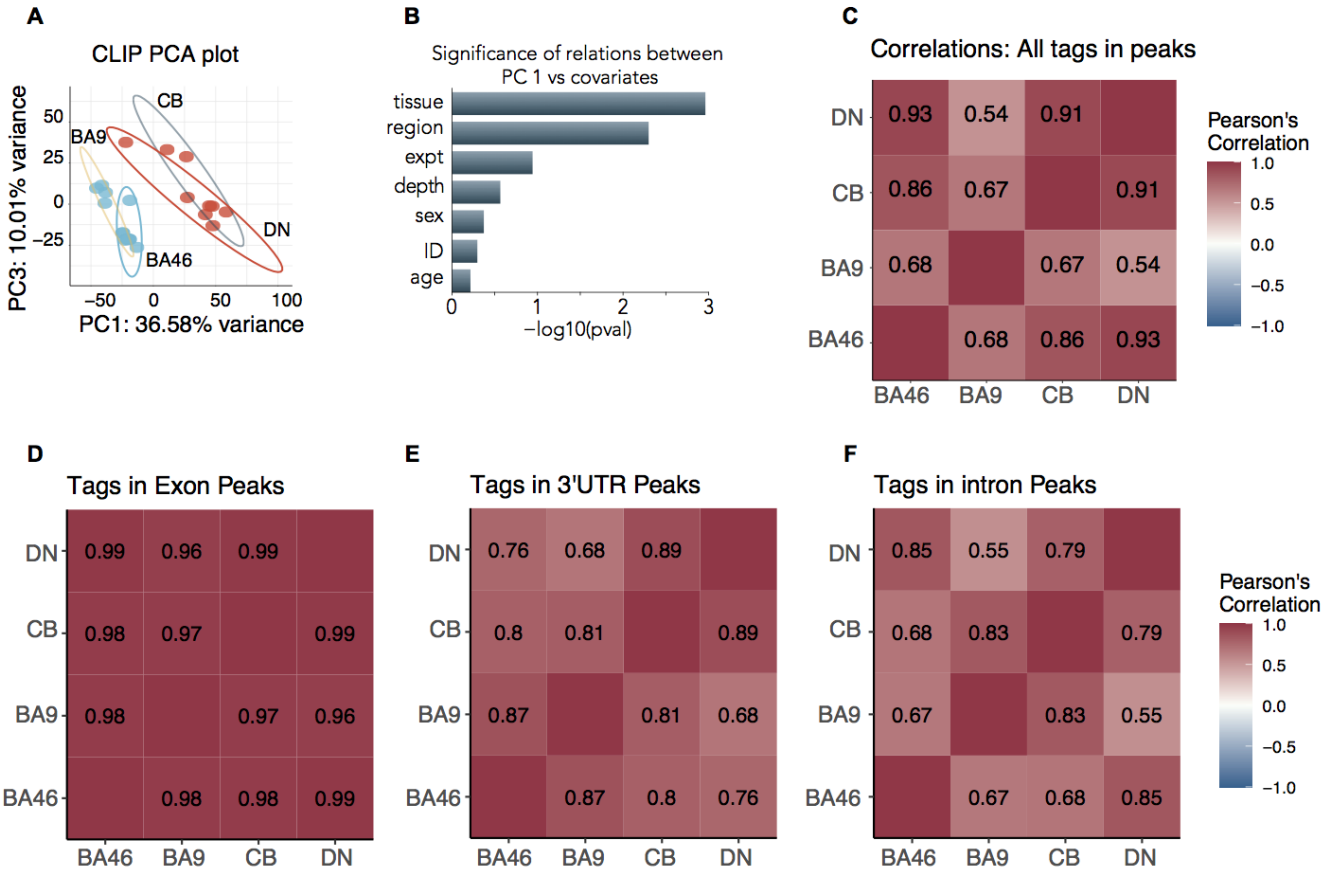


Figure 2-7: (A) Principal component analysis (PCA) plot displaying the projections of sample expression profiles demonstrating overall effects of experimental and batch. Each sample is plotted and spanned by the percent contributions of the first and third principal component. **(B)** computed significance of relations between PC1 and experimental covariates contributing to library variation, using Kruskal-Wallis test for categorical variables. Calculating significance reveals tissue type (cortex vs. cerebellum) and region (BA9, BA46, CB, or DN) are most significantly contributing to PC1 variance. **(C)** Pearson Correlation Coefficients (R) for total number of CLIP tags in peaks per brain region. Cortex includes BA46 (BC=5) and BA9 (BC=5), cerebellum refers to dentate nucleus (BC=6) and cerebellar cortex (BC=3). Only significant correlations shown. **(D-F)** Pearson Correlation Coefficients (R) across brain regions for total number of CLIP tags in **(D)** exonic peaks. **(E)** 3'UTR peaks, and **(F)** intronic peaks. Reproducibility of CLIP tags in peaks across brain regions is highly variable according to genomic localization of the peak. Highest reproducibility observed across regions in exons, R > 0.96 across all brain regions. R coefficients for 3'UTR peaks are slightly decreased with lowest R == 0.68 for BA9 and DN CLIP profiles. Interestingly the greatest shift in global correlations across brain regions was observed for intronic peaks.

To test whether the reproducibility of Nova binding is influenced by the genic location of the binding site, peaks were subset into three categories: (1) exonic, (2) intronic, and (3) 3'UTR binding sites. Pairwise correlation coefficients were calculated per group (Figure 2-7D-F). Tags in exonic peaks per brain region were highly reproducible across all regions, with coefficients between 0.96 and 0.99 (Figure 2-7D). Tags in 3'UTR peaks were moderately reproducible across regions, with coefficients between 0.68 and 0.89 (Figure 2-7E). Finally, tags in intronic peaks were the least reproducible comparatively, with coefficients between 0.55 and 0.85 (Figure 2-7F). These data suggest that Nova binding and thus regulation within exonic space is highly conserved and reproducible across brain regions. However, interactions with 3'UTRs appear to be region-dependent. Furthermore, correlation across regions for Nova binding within introns suggests that utility of precise binding sites may have uniquely diverged across human brain regions.

To view the distribution of Nova regulation across brain regions, each peak was annotated with genomic location and summarised as percent of total peaks per region (Figure 2-8A). As with previous analysis, global distribution of binding within cortical (BA46 and BA9) and cerebellar (CB and DN) regions exhibit patterns of regional specificity in 3'UTRs and introns. Cortical regions display higher percent total peaks localized to 3'UTRs as compared to cerebellar regions (BA9 = 67%; BA46 = 62% vs. CB = 55% ; DN = 55%). For added measure, peaks were combined and those utilized by Nova (i.e. bound by Nova) across all profiled brain regions were subset (n = 11,380) (Figure 2-8B). Genomic distribution of shared Nova regulation was summarised as percent of total peaks (Figure 2-8C). The peak distribution of all shared peaks were as follows: 3'UTR = 70%; intron = 24%; exon = 4.2%; 5'UTR = 0.83%; DS10k =

0.03%. These data suggest that in addition to alternative splicing, Nova might have expanded functional capacity in 3'UTRs. Taken together, Nova-CLIP binding profiles suggest regionally distinct patterns of Nova regulation in the human brain.

Figure 2-8 Genomic distribution of Nova peaks within each brain region profiled

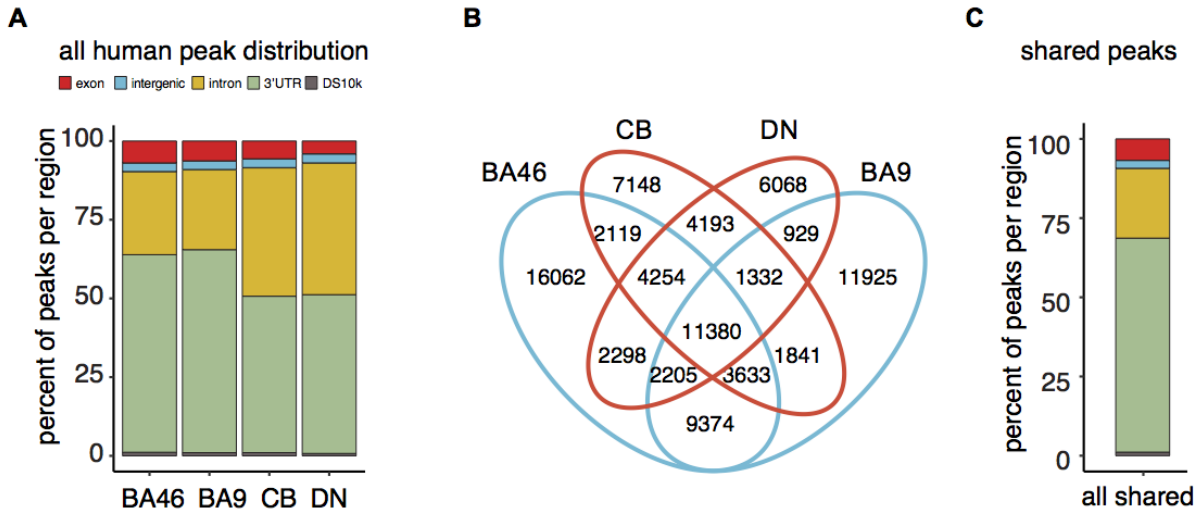


Figure 2-8: (A) Genomic distribution presented as percent of total significant peaks called per human brain region. **(B)** Venn Diagram illustrating the shared overlap in peaks across all brain regions. **(C)** Cumulative genomic distribution of Nova peaks with shared regulation for all brain regions (n = 11380).

Identifying Nova Target networks in human brain CLIP

The identification of Nova targets in humans is of great importance as it will allow for better understanding of the complex regulatory mechanisms driving human brain expression networks. In order to be considered a Nova target, there must be at least one significant peak with 5 or more overlapping tags from a single brain region. For simplicity of interpretation, Nova targets were also filtered for protein coding RNAs. The region with the greatest number of Nova targets is BA46 and the region with the lowest number of Nova targets is DN (Figure 2-9A, left). In total 9,481 Nova targets were identified from human brain CLIP data, indicating more than 70% of all targets were bound per region profiled (Figure 2-9A, right). To determine the percent overlap of Nova targets across all four brain regions, the targets across all four brain regions were intersected to show that 62% of Nova targets are bound by Nova across all regions profiled (Figure 2-9B). Interestingly, in mouse studies Nova targets have been described as highly divergent across brain regions. Conversely these data indicate global Nova target networks are highly similar for all human brain regions profiled in this study.

Figure 2-9 Nova targets per brain region CLIP'd

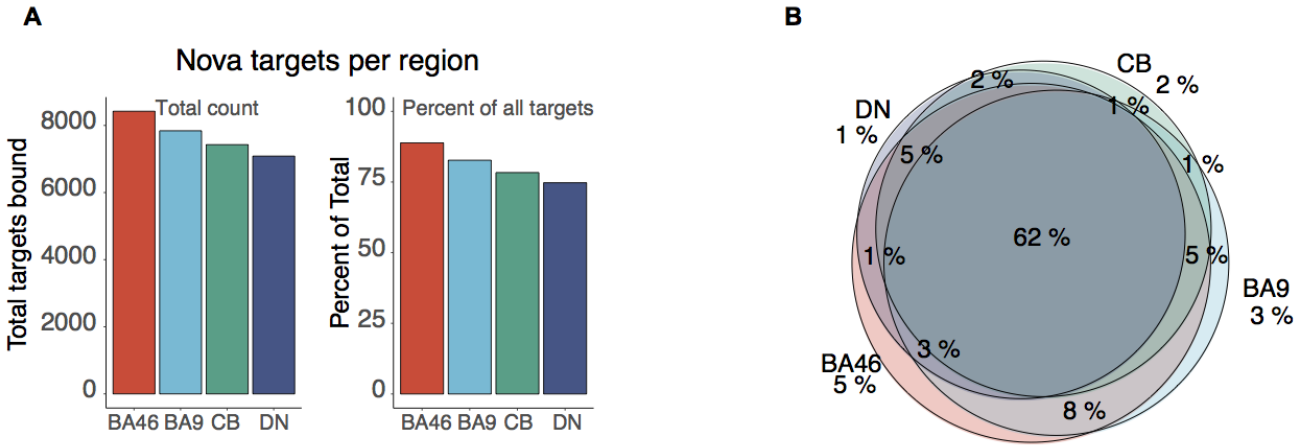


Figure 2-9: (A) Barplot displaying total target count per brain region: BA46 = 8423 (red); BA9 = 7841 (blue); CB= 7427 (green); DN = 7086 (purple) (left). Barplot displaying percent of total targets (n = 9481) per brain region: BA46 = 88.8% (red); BA9 = 82.7% (blue); CB= 78.3% (green); DN = 74.7% (purple) (right). **(B)** Venneuller plot illustrating Nova's shared target network across brain regions. Venn center indicates Nova target is significantly bound by Nova in all four brain regions profiled. Targets were filtered for protein coding RNA, total targets = 9,481. Light grey regions indicate 1% overlap.

More than 75% of targets interact with Nova in at least three out of the four regions profiled. In order to simplify interpretation for the following analysis, BA9 and BA46 clip tags in peaks were grouped together as "cortex" and CB and DN were grouped together as "cerebellum". A normalized binding score for cortex and cerebellum was calculated per target by summarizing total tags in peaks across each target, normalizing for gene length, and scaling for library depth per region. Normalized binding scores for cortex over cerebellum were compared and the top 10 changes in Nova binding were labelled for cortex and cerebellum respectively (Figure 2-10). Color indicates the number of regions with significant binding to target (grey = bound in 3+ regions; yellow = bound in 1-2 regions). Pairwise Fisher's exact test was performed

in order to test for significance and identify most differentially bound targets in the cortex compared to cerebellum (Figure 2-11).

Figure 2-10 Nova target affinity is highly correlated between human cortex and human cerebellum

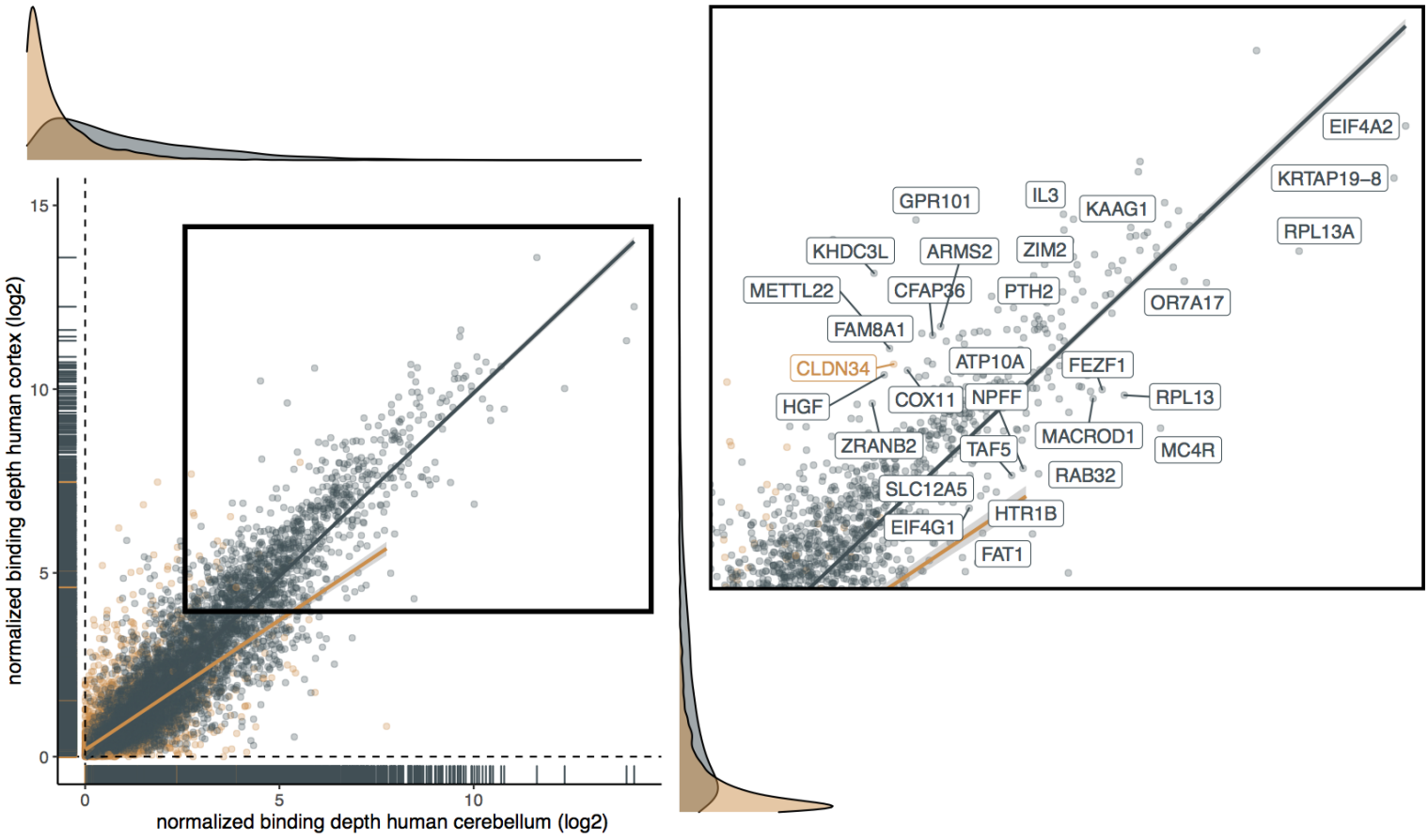


Figure 2-10: Cortex normalized CLIP binding score is plotted over Cerebellum normalized CLIP binding score to further characterize Nova affinity across brain tissues. Cortex = BA46 and BA9 CLIP data and Cerebellum = CB and DN CLIP data. Targets were grouped according into two categories: (1) Nova binding observed in at least three out of the four regions, "shared" (grey), or (2) Nova binding observed in two or less regions, "other" (yellow). Correlation analysis (cortex vs. cerebellum) was performed for each of the two categories (pearson coefficient: R =0.92 (shared) and R = 0.65 (other)) . Linear regression model and density of points per axis is indicated for each category (shared targets (3+ regions) = grey and other targets (1-2 regions) = yellow). Labelled targets represent the largest change in binding score for cortex vs. cerebellum from the top most highly bound in both cortex (Score $\log_2 > 5$) and cerebellum (Score $\log_2 > 5$).

Figure 2-11 Volcano plot to highlight most significantly differentially bound Nova targets Cortex vs. Cerebellum

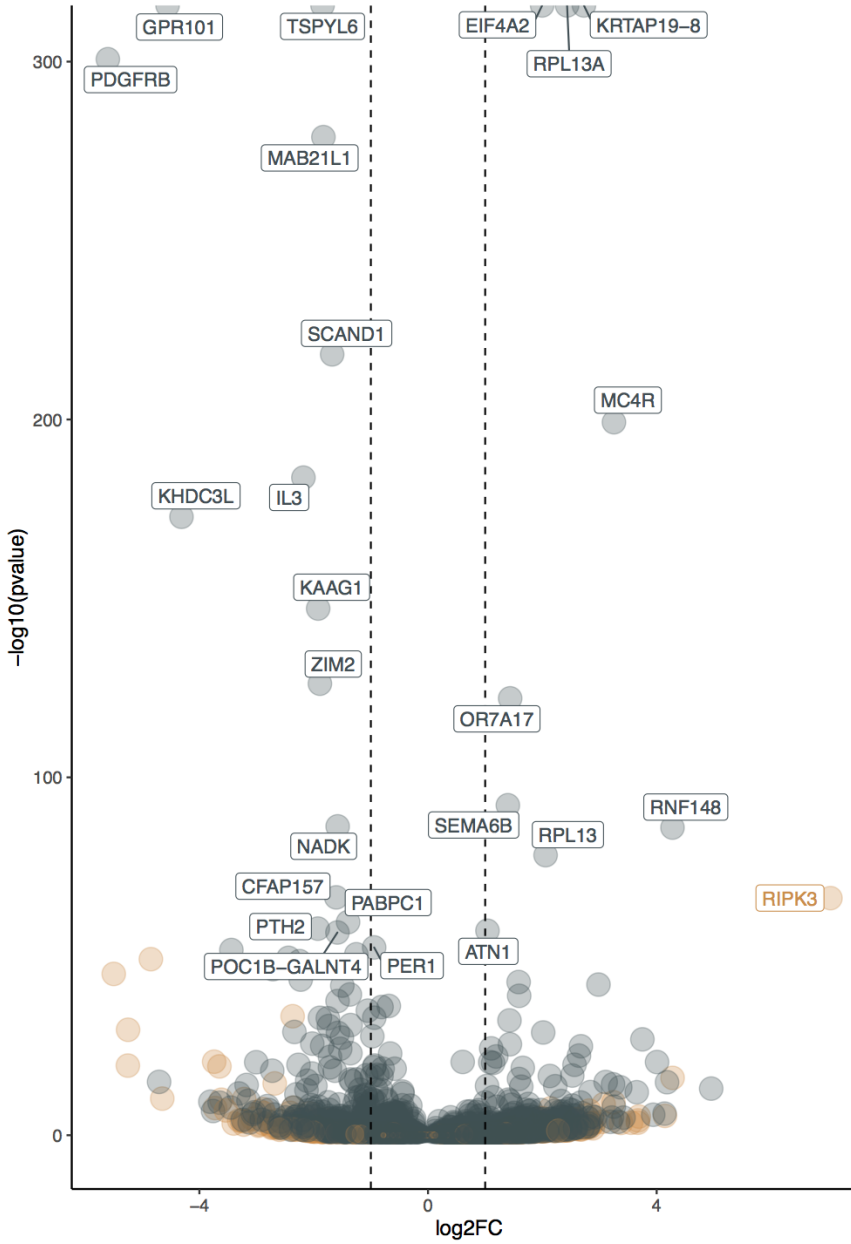


Figure 2-11: Volcano plot of target level Nova CLIP enrichment in cortex (left) vs. cerebellum (right). Vertical dashed line represents threshold of abs(log2FC) > 1. In total, 188 targets significantly enriched in the cerebellum and 188 targets significantly enriched in the cortex. Top 25 enriched targets are labelled with gene symbols (shared targets (3+ regions) = grey and other targets (1-2 regions) = yellow)

Gene set enrichment analysis (GSEA) was performed on all significantly differentially enriched Nova targets to quantitate the enrichment of GO terms within significantly differentially bound targets. The top 3 most enriched terms for Nova regulation in the cerebellum were regulation of histone methylation, postsynaptic density organization and membrane assembly. Whereas the top 3 most enriched terms for Nova regulation in the cortex were forebrain neuron development, hematopoietic stem cell proliferation, and cochlea morphogenesis (Figure 2-12). These data suggest specialized functions for Nova-RNA regulation may have evolved in cortical stem cells specifically in the forebrain. These findings also elude to a number of conserved functions as Nova has been shown to play a role in genesis of the corpus callosum in the developing cortex in addition to Nova being required for efferent innervation and targeting to the cochlea during embryonic development in mice (Saito et al., 2016, 2019).

Figure 2-12 Gene Set Enrichment Analysis: Significantly differentially enriched Nova targets

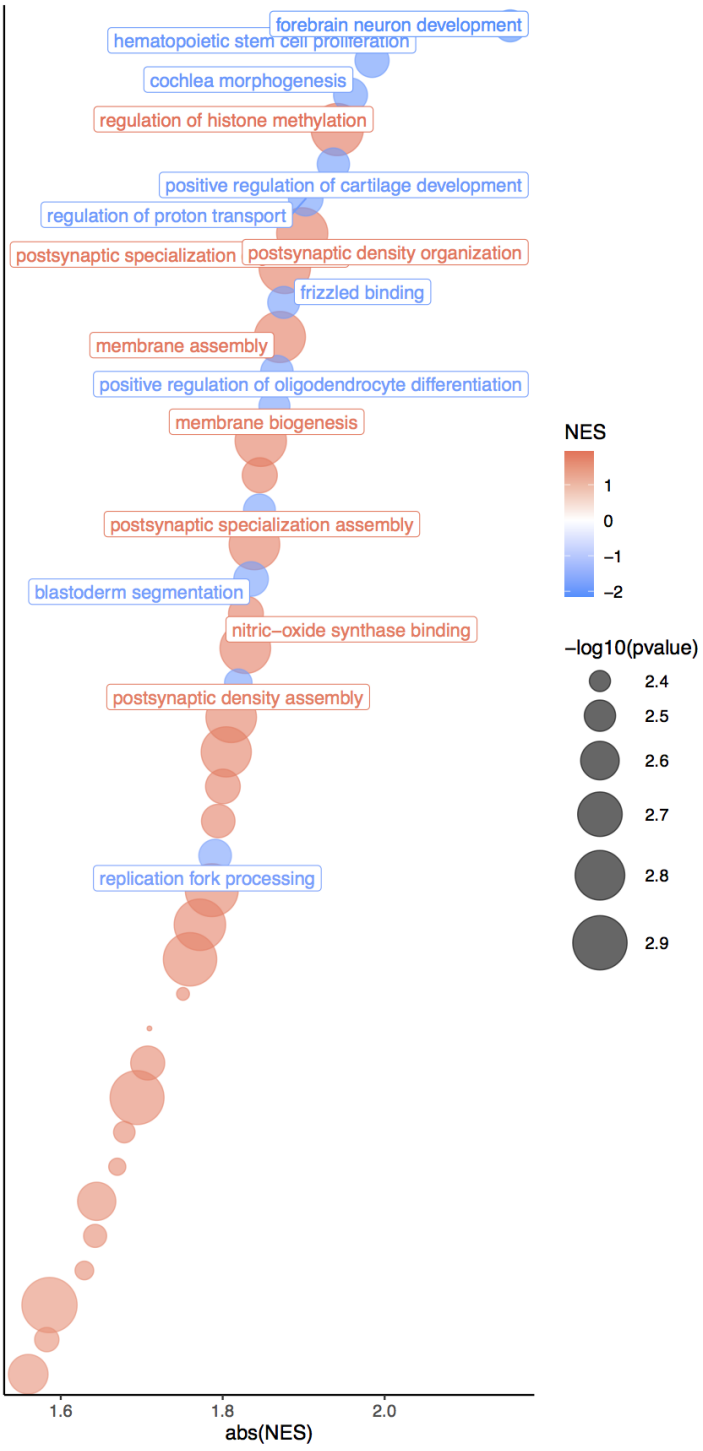


Figure 2-12: Gene Set Enrichment Analysis for all significantly differentially enriched Nova targets in cortex vs. cerebellum. Size of point increases relative to significance and color represents NES (blue = cortical enrichment; orange = cerebellar enrichment). NES is plotted as the absolute value for visual purposes. Top 8 most enriched gene sets for cortex (blue) and for cerebellum (orange) are labeled accordingly.

In order to investigate region-specific regulation and focus specifically on differences in binding site usage, targets that were equally bound by Nova in both brain regions were considered. For the most stringent comparison and in order to control for sample quality and genetic variability across individuals, we focused on regions with samples from matched individuals, the dentate nucleus (DN) and BA46 (Figure 2-13A-B). It is important to note that the significant change in peak localization across regions holds true for inter-individual analysis as well (Figure 2-13C). As previously described, normalized CLIP scores per target and per peak were calculated for DN and BA46 regions. Targets were subset to include only those that were equally bound in both brain regions. The genomic distribution of normalized binding depth in peaks was compared across brain regions. Comparing binding patterns within DN indicates that intronic peaks are more highly utilized compared to 3'UTR peaks, while comparison within BA46 indicates that 3'UTR peaks are more highly utilized compared to intronic peaks (Figure 2-14A-B). It is important to identify targets that have "switched" localization of Nova-RNA interaction sites from intronic to 3'UTR in the cortex, as these observations will inform on human-specific expansions involving Nova protein function within 3'UTRs.

Figure 2-13 Region-specific Nova regulation across individuals

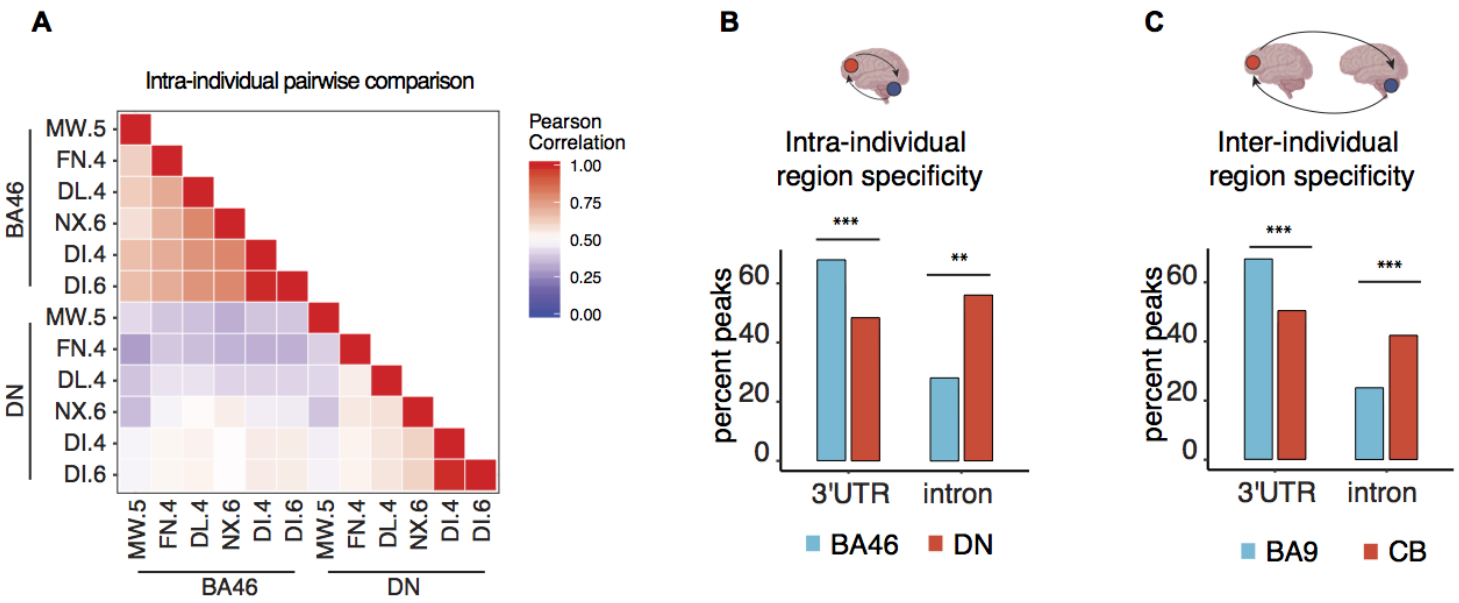


Figure 2-13: (A) Intra-individual pairwise comparison across brain regions to visualize direction of correlation, only significant correlations shown. Calculated coefficient = Pearson. **(B)** Intra-individual regional specificity. Barplot demonstrates a significant difference in the percent of total peaks in matched individuals (BA46 vs. DN) localized to 3'UTRs and introns. ($p < 2.22e^{-16}$, wilcoxon ranked test). **(C)** Inter-individual regional specificity. Barplot demonstrates a significant difference in the percent of total peaks across individuals (BA9 vs. CB) localized to 3'UTRs and introns. ($p < 2.22e^{-16}$, wilcoxon ranked test)

Figure 2-14 Differential preference in 3'UTR and intronic binding within regions

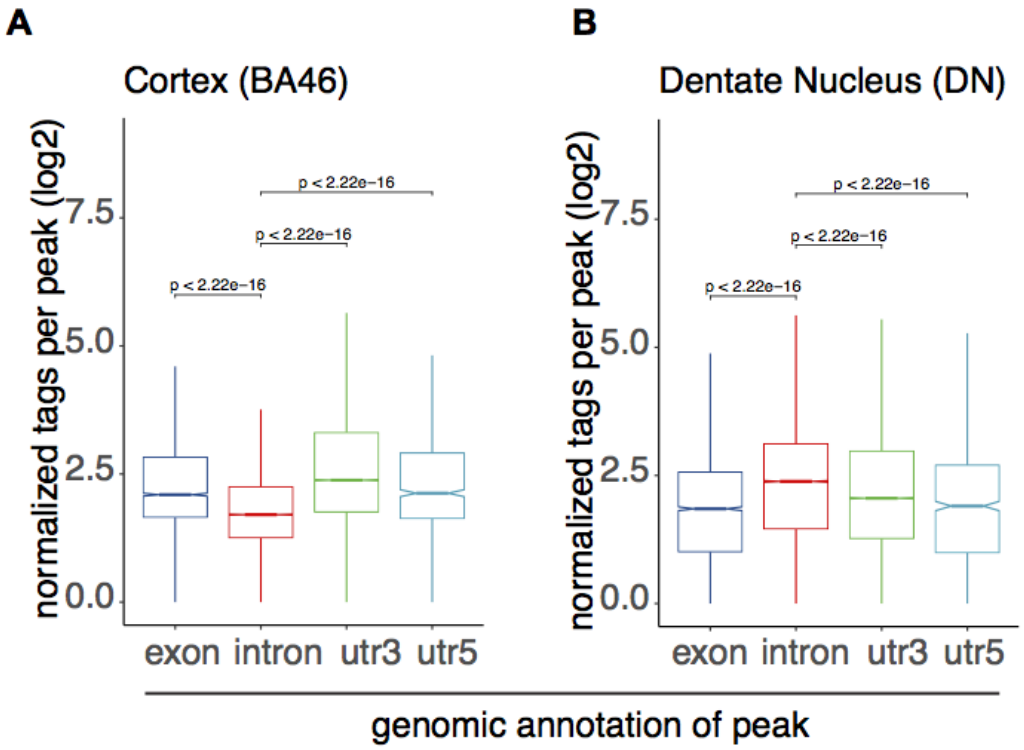


Figure 2-14: Boxplots of normalized binding depth per peak subset according to genomic localization of peak (exon = dark blue; intron = red; 3'UTR = green ; 5'UTR = light blue). **(A)** Dentate nucleus shows Nova is preferentially bound to intronic peaks as compared to all other genomic locations. **(B)** Brodmann Area 46 shows Nova is preferentially bound to 3'UTR peaks as compared to all other genomic locations. Only targets with equal binding for both brain regions are shown. Statistical significance determined using wilcoxon ranked sum test.

Discussion

This study has generated a comprehensive *in vivo* binding map of endogenous human NOVA bound to endogenously expressed RNA. This is only the second instance of such detailed *in vivo* functional regulatory maps for the human brain. It is the first to span four distinct brain regions, and the only *in vivo* human binding map for NOVA (Scheckel et al., 2016). The transcriptome-wide Nova-RNA interaction sites identified provide a valuable resource for researchers. These data illustrate diverse and complex Nova regulatory profiles in the human central nervous system that uniquely highlight RNAs and pathways under the control of multiple layers of alternative regulation with regional specificity in the human brain.

Human brain HITS-CLIP shows per target Nova binding densities are highly correlated across brain regions, conversely enrichment analysis indicates divergent functional networks may have evolved with regional specificity in humans. Furthermore, characterization at the peak level indicates robust differences in genomic localization for Nova-RNA interactions in the cortex compared to cerebellum. Differential utilization of 3'UTR peaks compared to intronic peaks within DN and BA46 implies that in addition to regulating alternative splicing, Nova may have expanded alternative functions in the human cortex pertaining to 3'UTR regulation. The genomic localization of Nova-RNA interactions can provide a great deal of insight towards function. To date, Nova has been widely characterized as a regulator of alternative splicing in the mouse brain. The most notable description of Nova HITS-CLIP is that binding sites are primarily localized to intronic regions, with some data sets describing >90% of all peaks localizing to introns. However, while most targets are bound intronically, not all are subject to Nova regulated alternative splicing (Huang et al., 2005; Licatalosi et al., 2008; Ruggiu et al., 2009).

Licatalosi et al. first revealed Nova-RNA interactions localized to 3'UTRs, specifically in clusters overlapping with polyadenylation sites. A total of 43 transcripts were identified and two candidates were selected for further characterization, Elav-like family member 2 (Cugbp2) and Solute carrier family 8 member A1 (Slc8a1). Microarray coupled with RNase protection analysis (RPA) in Nova2-KO vs. WT revealed an increase in the usage of these polyA sites in the absence of Nova2. Furthermore the increase in polyA usage was correlated with reciprocal decreases in distal polyA site processing. Taken together, these findings point towards changes in the relative levels of alternative polyadenylated RNAs are not only due to changes in stability, but as a result of aberrant ApA in the Nova2-KO. After this realization, an additional 29 Nova target RNAs were validated via qRT-PCR and 12/29 showed significant changes in polyA site usage. Bioinformatic analysis revealed YCAY-rich elements flanking these polyA sites, which were later shown via mutation assays as essential for proper Nova regulation at these sites in the mouse brain. Additionally, the data suggested a position-dependent effect of Nova binding relative to polyA sites as a contributing factor to whether Nova binding suppressed or activated the polyadenylation site. Nova peaks within 50nt of the cleavage site resulted in suppression, whereas polyA sites where Nova was bound to more distal regions activated usage of the polyadenylation site (Licatalosi et al., 2008).

The function of Nova binding at 3'UTRs has been extensively postulated following the initial finding of Licatalosi et al., and biochemical studies have begun to tease apart mechanisms of such roles for a handful of Nova targets. However, one common finding for all studies was that Nova bound 3'UTRs were far less abundant in the mouse brain compared to Nova bound

introns. To briefly summarise these biochemical findings: Nova targets 3'UTRs to (1) control polyadenylation site usage via position dependent binding (proximal = suppression; distal = activation); (2) transport mRNAs from the nucleus to cytoplasm (3) localize mRNAs to distal dendrites/synapses for local translation; (4) regulate translation as Nova1 has been found to associate with the polyribosome in the mouse spinal cord (C. Fraser, unpublished)(Huang et al., 2005; Licatalosi et al., 2008; Racca et al., 2010; Ruggiu et al., 2009).

Our data suggests that Nova function at 3'UTRs may be more abundant in the human brain compared to previous observations in the mouse brain. Although, additional explanations for the observed differences could be attributed to changes in Nova1 vs. Nova2 levels, which will be described in more detail in subsequent chapters. However, it is clear from the current data in humans that Nova-mediated 3'UTR regulation is abundant throughout the brain. Furthermore, it is possible that Nova mediated regulation of alternative polyadenylation usage acts as an additive layer of control and complexity in the human cortex relative to the human cerebellum. These findings may also contribute to the region-specific regulation observed at the network level in the cortex. Altogether these data indicate expanded control over Nova regulated target RNAs, specifically within the human cortex.

Chapter 3: Evolutionary conservation of Nova-CLIP binding profiles

Introduction

In the mouse, global dysregulation of Nova1 and Nova2 result in neurologic dysfunction; and aberrations in Nova1/2 expression levels lead to widespread changes in gene expression networks required for proper neuronal development and function. However HITS-CLIP based approaches have shown that Nova regulation is both specific and dynamic. As an example, Nova2 heterozygous mice exhibit frequent bilateral synchronous cortical epileptiform discharges as well as spontaneous occurrence of generalized seizures (Eom et al., 2013). Mouse seizure models reveal a shift in Nova subcellular localization from nucleus to cytoplasm in cortical and CA1 neurons, but not in cerebellar Purkinje neurons (Eom et al., 2013; Racca et al., 2010). These findings demonstrate the intricacy of Nova protein dynamics as they function in cell-type specific and context dependent manners. Studies later investigated cell-type specific profiling of Nova in the mouse to reveal distinct Purkinje-cell (PC) specific mechanisms driving phenotypic associations in Nova-cKO such as motor coordination defects, PC degeneration, loss of synaptic layer thickness, neuritic swelling, and reduced spine density (Racca et al., 2010; Saito et al., 2019).

Characterizing patterns of conservation and elucidating species- specific mechanisms of RNA regulation significantly aids in understanding how Nova regulation has evolved across species. An important question moving forward is what would be identified by comparing the

human data to the large amount of mouse data available and to ask what is orthogonal or what is uniquely evolved in the human brain? Orthologous genes are determined based on sequence homology of coding exons, which trend towards being conserved in both exon structure and length (Kuhn et al., 2013; United States. Congress. House. Committee on Appropriations. Subcommittee on the Departments of Labor et al., 2004). Exonic regions display high levels of sequence homology given that they encode functional proteins; untranslated regions (UTRs) are also highly homologous between mouse and human as they encode distinct RNA regulatory sequences that are essential for proper post-transcriptional regulation of mature RNA (Yue et al., 2014). Intronic regions of orthologous genes are generally longer in humans and are likely significantly contributing to observed differences in whole genome length between the two species (Breschi et al., 2017). Furthermore, intron-specific expansion suggests that human orthologues have likely gained RNA regulatory capacity, and thus may be subject to more complex post-transcriptional regulation and RNA processing in humans compared to the mouse.

Results

Nova targets are highly orthologous between mouse and human

Conservation of Nova's regulatory network between mouse and human is largely dictated by whether target RNAs are considered orthologous between species. In order to determine which Nova targets are orthologous in human and mouse, all human CLIP targets were matched to the orthologous mouse target; targets without annotation were labeled as "non-orthologous". These data revealed ~95% of human-defined Nova targets are orthologous to the mouse (Figure

3-1A-B). Conservation of functional Nova binding was further characterized by intersecting mouse-defined Nova CLIP targets with human-defined Nova CLIP targets, which elucidated a 56% human to mouse overlap in shared targets (Figure 3-1C). Hypergeometric testing was used to test the significance in overlap between mouse and human CLIP targets (Figure 3-1D). These data highlight the orthologous relationship of Nova target networks across species and demonstrate functional conservation of Nova regulation between mouse and human brain.

Figure 3-1 Nova targets are highly orthologous and bound in mouse and humans

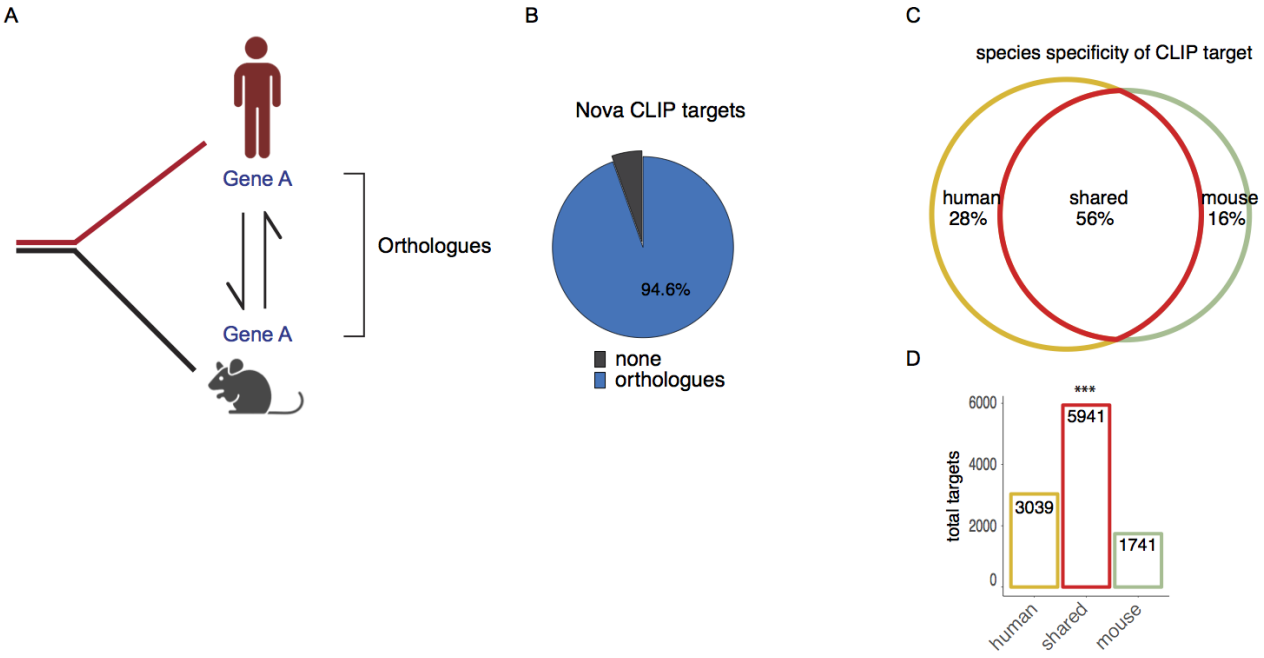


Figure 3-1: (A) Schematic illustrating cross-species orthologues. **(B)** Pie chart representing percent of all human-derived NOVA targets with mapped orthologue in mouse (blue), non-orthologous human targets (grey). **(C)** Intersection of mouse defined and human defined CLIP targets. Venn diagram demonstrating overlap of Nova targets identified in the human brain compared to the mouse brain. **(D)** Significant overlap between mouse and human targets demonstrated using hypergeometric testing.

Nova has been shown to interact with a functionally diverse set of targets important for normal brain function (Dredge and Darnell, 2003; Dredge et al., 2001, 2005; Jelen et al., 2007; Jensen et al., 2000b; Leggere et al., 2016; Licatalosi et al., 2008; Musunuru and Darnell, 2004; Racca et al., 2010; Saito et al., 2016, 2019; Ule et al., 2005, 2006; Yuan et al., 2018). Therefore, in order to determine whether specific subsets of targets demonstrate patterns of conserved or divergent regulation, all three target groups (human-specific, shared, and mouse-specific) were segregated according to function by module detection using the HumanBase software (Krishnan et al., 2016). For the shared target group, six functional clusters were detected, two of which include over 300 genes (Figure 3-2). The S1 cluster, which contains 311 genes, is highly enriched for genes involved in ion transport and receptor signaling, with the top most significant GO terms being glutamate receptor signaling and potassium ion transport. S2 cluster (457 genes) contains genes involved in cytoskeleton organization and histone modification. The S4 cluster (164 genes) is enriched for genes involved in neuron development and maturation (Figure 3-3). The less robust list of mouse-specific targets clustered into four functional modules (Figure 3-4). The M1 cluster, which contains only 32 genes, is enriched for genes involved in cell morphogenesis involved in neuron differentiation and gamma-aminobutyric acid signaling pathway. The M2 cluster, which represents 26 genes, contains genes involved in regulating the protein catabolic process and progesterone receptor signaling pathway (Figure 3-5). Targets that are uniquely regulated by Nova in humans segregated into five distinct functional clusters (Figure 3-6). The H1 cluster, containing 31 genes, is enriched in GO terms describing regulation of protein localization as well as response to oxidative stress and positive regulation of DNA repair. The H3 cluster, consisting of 42 genes, is enriched for genes involved in mRNA catabolic processes and translation. The H4 cluster, which is the second most robust of the five clusters,

contains 35 genes. The H4 genes are enriched in calcium ion transmembrane transport and negative regulation of cell motility (Figure 3-7).

Figure 3-2 Shared targets segregated into functional clusters

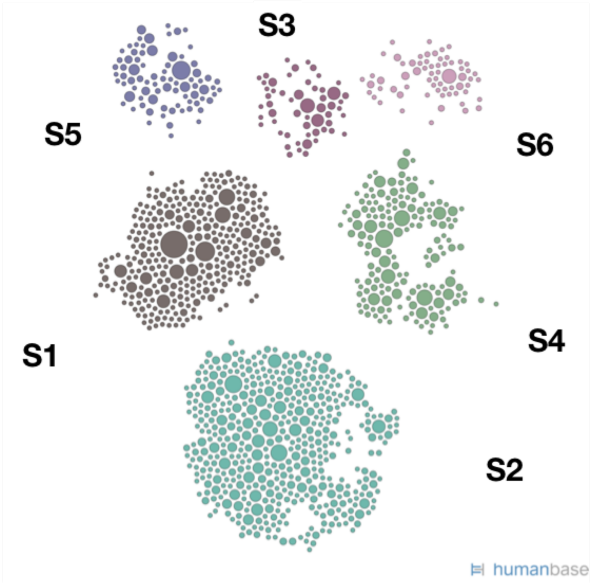


Figure 3-2: Functional clusters of Nova targets that are shared in mouse and human CLIP. Shared Nova targets were clustered with HumanBase software.

Figure 3-3 Top GO terms representing functional clusters of shared Nova targets

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
S1	glutamate receptor signaling pathway	0.00000008	311	124
	potassium ion transport	0.00000021		
	monovalent inorganic cation transport	0.00000021		
	potassium ion transmembrane transport	0.00000021		
	cellular potassium ion transport	0.00000021		
	ionotropic glutamate receptor signaling pathway	0.00000722		
	action potential	0.00037192		
	generation of neurons	0.00094655		
	regulation of membrane potential	0.00094655		
	multicellular organismal signaling	0.00103585		
S2	regulation of microtubule cytoskeleton organization	0.00037192	457	228
	regulation of microtubule-based process	0.00040660		
	regulation of microtubule polymerization or depolymerization	0.00061405		
	regulation of cytoskeleton organization	0.00066931		
	protein polymerization	0.00072515		
	regulation of protein complex assembly	0.00088261		
	histone modification	0.00088261		
	covalent chromatin modification	0.00094655		
	microtubule cytoskeleton organization	0.00148341		
	regulation of supramolecular fiber organization	0.00179140		
S3	establishment of spindle localization	0.00323948	55	36
	spindle localization	0.00352392		
	establishment of spindle orientation	0.02365397		
	positive regulation of calcium ion transmembrane transporter activity	0.02365397		
	negative regulation of plasma membrane bounded cell projection assembly	0.02697247		
	regulation of macroautophagy	0.03065269		
	cellular component disassembly	0.04199910		
	regulation of plasma membrane bounded cell projection assembly	0.04199910		
	regulation of cell projection assembly	0.04282985		
	microtubule cytoskeleton organization involved in mitosis	0.04613353		

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
S4	regulation of neuron differentiation	0.00428902	164	88
	regulation of neuron projection development	0.00639068		
	regulation of neurogenesis	0.00801416		
	regulation of nervous system development	0.01085567		
	neuron projection development	0.01605287		
	negative regulation of neuron differentiation	0.01639143		
	neuron differentiation	0.01698380		
	neuron development	0.02051938		
	adenylate cyclase-activating adrenergic receptor signaling pathway	0.02124616		
	positive regulation of lipopolysaccharide-mediated signaling pathway	0.02124616		
S5	response to starvation	0.01917654	80	23
	Golgi localization	0.02359398		
	response to nutrient levels	0.04372500		
	response to extracellular stimulus	0.04613353		
	regulation of canonical Wnt signaling pathway	0.04706146		
	histone lysine methylation	0.05011666		
	cellular response to external stimulus	0.05492144		
	cellular response to starvation	0.05773501		
	peptidyl-lysine methylation	0.05861943		
	histone methylation	0.06371718		
S6	eye development	0.07891980	65	2
	sensory organ development	0.09540766		

Figure 3-3: Top GO terms representing each functional cluster derived from human and mouse shared Nova targets. GO terms were identified by HumanBase software.

Figure 3-4: Human-specific targets segregated into functional cluster

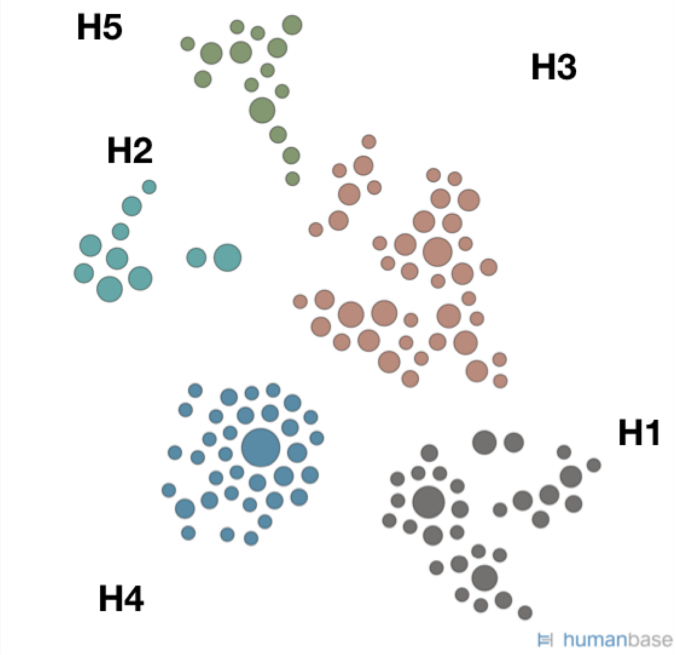


Figure 3-4: Functional clusters of Nova targets that are human-specific. Human-specific Nova targets were clustered with HumanBase software.

Figure 3-5: Top GO terms representing functional clusters human-specific Nova targets

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
H1	regulation of cellular protein localization	0.00722571	31	75
	regulation of protein targeting to mitochondrion	0.00760103		
	regulation of establishment of protein localization to mitochondrion	0.00783658		
	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	0.00824237		
	regulation of intracellular protein transport	0.01416640		
	regulation of protein targeting	0.01416640		
	protein targeting to mitochondrion	0.01805358		
	regulation of intracellular transport	0.01895484		
	response to oxidative stress	0.01895484		
	positive regulation of DNA repair	0.01895484		
H2	RNA phosphodiester bond hydrolysis, endonucleolytic	0.00760103	10	3
	RNA phosphodiester bond hydrolysis	0.01416640		
	nucleic acid phosphodiester bond hydrolysis	0.01895484		
H3	nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	0.01416640	42	31
	positive regulation of mRNA catabolic process	0.01805358		
	amide biosynthetic process	0.01927922		
	positive regulation of mRNA metabolic process	0.02353808		
	nuclear-transcribed mRNA catabolic process	0.02460962		
	regulation of mRNA catabolic process	0.02513254		
	translation	0.03401778		
	peptide biosynthetic process	0.03547903		
	response to antibiotic	0.03680320		
	regulation of cell-substrate adhesion	0.04142250		
H4	plasma membrane bounded cell projection morphogenesis	0.03187666	35	22
	cell projection morphogenesis	0.03187666		
	cell part morphogenesis	0.03493229		
	negative regulation of response to external stimulus	0.04142250		
	calcium ion transmembrane transport	0.04428716		
	negative regulation of cell migration	0.04898135		
	negative regulation of cell motility	0.05077032		
	negative regulation of cellular component movement	0.05162712		
	cellular component disassembly	0.05162712		
	negative regulation of locomotion	0.05396728		
H5	cellular response to hormone stimulus	0.03870697	15	2
	response to hormone	0.04306805		

Figure 3-5: Top GO terms representing each functional cluster derived from human-specific Nova targets. GO terms were identified by HumanBase software.

Figure 3-6: Mouse-specific targets segregated into functional cluster

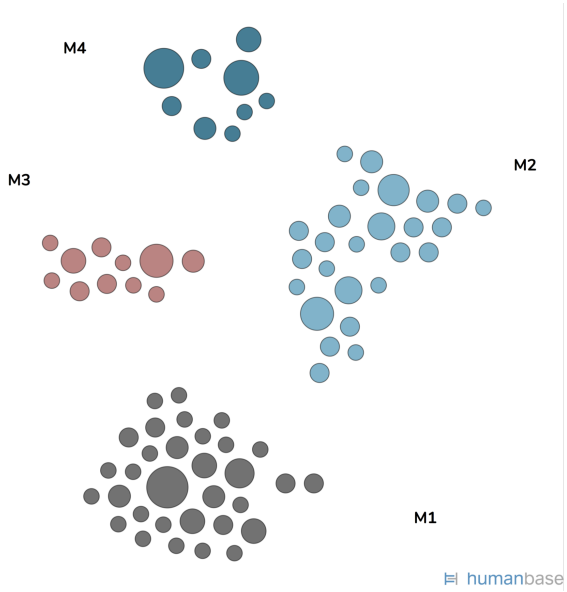


Figure 3-6: Functional clusters of Nova targets that are mouse-specific. Mouse-specific Nova targets were clustered with HumanBase software.

Figure 3-7: Top GO terms representing functional clusters mouse-specific Nova targets

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
M1	cell morphogenesis involved in neuron differentiation	0.00131336	32	114
	gamma-aminobutyric acid signaling pathway	0.00221156		
	positive regulation of axon extension	0.00264876		
	cell morphogenesis involved in differentiation	0.00389177		
	behavior	0.00412032		
	regulation of membrane potential	0.00412032		
	regulation of axon extension	0.00412032		
	regulation of cell morphogenesis involved in differentiation	0.00468562		
	regulation of extent of cell growth	0.00468562		
	positive regulation of axonogenesis	0.00468562		
M2	regulation of protein catabolic process	0.00146037	26	94
	progesterone receptor signaling pathway	0.00221156		
	learning	0.00412032		
	ubiquitin-dependent protein catabolic process	0.00563044		
	modification-dependent protein catabolic process	0.00563044		
	cellular response to radiation	0.00563044		
	modification-dependent macromolecule catabolic process	0.00563898		
	learning or memory	0.00653121		
	protein polyubiquitination	0.00653121		
	regulation of proteolysis involved in cellular protein catabolic process	0.00653121		
M3	protein autophosphorylation	0.01166401	11	4
	positive regulation of cell migration	0.02477051		
	positive regulation of cell motility	0.02584491		
	positive regulation of cellular component movement	0.02631292		
M4	positive regulation of cell migration	0.02172726	9	3
	positive regulation of cell motility	0.02264827		
	positive regulation of cellular component movement	0.02286802		

Figure 3-7: Top GO terms representing each functional cluster derived from mouse-specific Nova targets. GO terms were identified by HumanBase software.

Evolution of YCAY-tetramers in Nova targets

Nova proteins consist of three KH-type RNA binding domains, which bind YCAY-rich sequences with high affinity and have been shown to be necessary and sufficient for Nova-mediated regulation of alternative splicing (Buckanovich et al., 1993; Dredge and Darnell, 2003; Dredge et al., 2005; Jensen et al., 2000a; Lewis et al., 2000)[(Zhang and Darnell, 2011; Zhang et al., 2010, 2013)]. Therefore it is of great interest to further characterize the association between YCAY-tetramers and species-specific changes within genic regions. Globally, human genes are longer compared to mouse genes (Figure 3-8A). In order to compare changes in global YCAY frequency for mouse (mm10) and human (hg38) genomes respectively, the total number of YCAY motifs within genic regions were summarised and normalized for total length. YCAY tetramers are significantly more abundant in human genic regions as compared to mouse genic regions with respect to length (Figure 3-8B). Furthermore, YCAY profiles were specifically enriched in Nova targets compared to non-targets for both respective species (Figure 3-8C). Taken together, these data indicate that humans may have selectively evolved expanded regulatory capacity for Nova regulation.

Figure 3-8: Human genes (hg38) are longer than mouse genes (mm10)

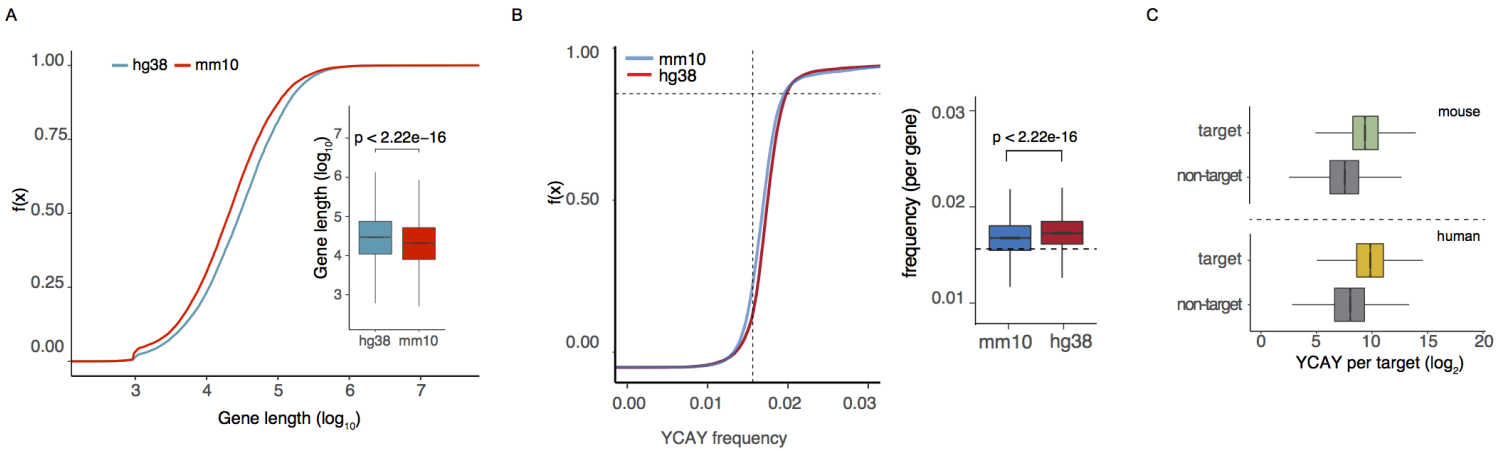


Figure 3-8: (A) Cumulative distribution plots and box plot demonstrating human genes (hg38) are significantly longer than mouse genes (mm10); ($p < 2.22e-16$, Wilcoxon ranked sum test). **(B)** Cumulative distribution plots and box plot to show YCAY frequency calculated in human genic space (hg38, blue) is significantly greater as compared to YCAY frequency calculated within mouse genic space (mm10, red); ($p < 2.22e-16$, Wilcoxon ranked sum test).**(C)** YCAY tetramer is enriched in Nova targets compared to non-Nova targets (grey) for both mouse (green) and human (yellow). Boxplot illustrating per target enrichment in YCAY tetramer for mouse (top) and human (bottom).

For further investigation of the relationship between YCAY tetramers and network-level species-specific changes, Nova targets were subdivided according to previously defined conserved and species-specific targets. Comparative analysis of YCAY tetramers across groups revealed a significant enrichment in conserved targets versus species-specific targets (Figure 3-9A). Additional subset of species-specific targets into mouse-specific and human-specific targets showed no significant difference in YCAY enrichment (Figure 3-9B). It is important to note that since YCAY abundance was calculated for mouse-defined targets using the mouse genome as background and human-defined targets utilized the human genome as background, it is likely that genome-specific differences in YCAY profiles may be influencing these analyses.

Figure 3-9: Conserved Nova targets are more highly enriched in YCAY elements as compared to species-specific and non-targets

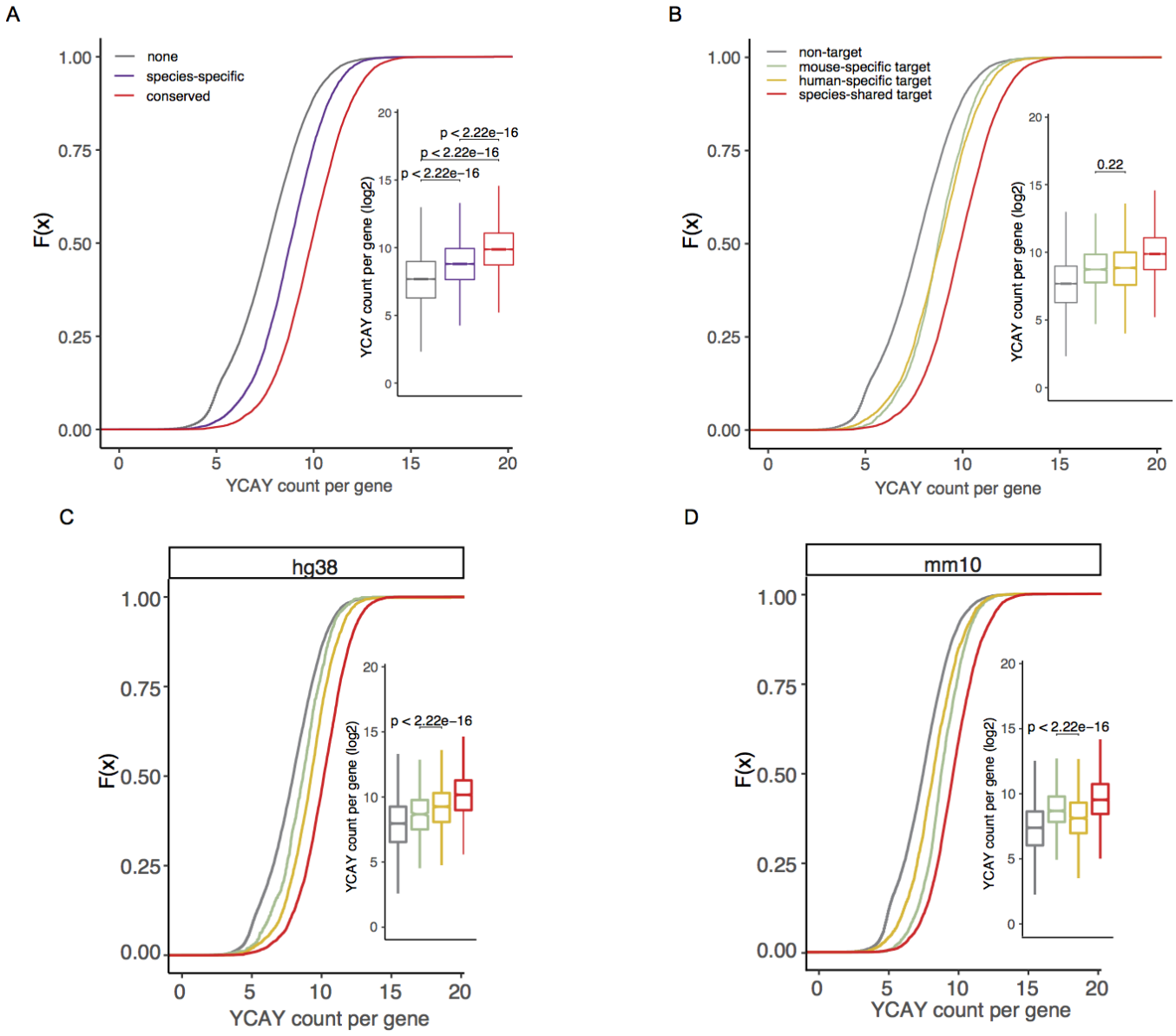


Figure 3-9: (A) Cumulative distribution plots and box plot demonstrating YCAY enrichment in all Nova targets, with the greatest enrichment observed for mouse to human conserved targets ($p < 2.22e-16$, Wilcoxon ranked sum test). **(B)** Cumulative distribution plots and box plot to show no significant difference between YCAY abundance in mouse-specific (mm10) and human-specific (hg38) targets($p = 0.22$, Wilcoxon ranked sum test). **(C)** Cumulative distribution plots and corresponding box plots. Human specific targets are more highly enriched in YCAY elements than mouse specific targets when abundance is quantified using the human genome (hg38) as background. **(D)** Mouse specific targets are more highly enriched in YCAY elements than human specific targets when abundance is quantified using mouse genome (mm10) as background, calculated abundance ($p < 2.22e-16$, Wilcoxon ranked sum test).

In order to compare the landscape of YCAY tetramers across Nova target groups, independent of genomic influence, YCAY abundance was calculated for all targets within each respective genome. Interestingly, when abundance was determined using the human genome (hg38) as background, human-specific targets were significantly more abundant with respect to mouse-specific Nova targets (Figure 3-9C). Indicative of genome-dependent loss in YCAY tetramers and regulatory divergence of these targets within mouse to human lineage. Conversely, comparison utilizing the mouse genome (mm10) as background revealed a greater abundance in YCAY-tetramers for the mouse-specific target group (Figure 3-9D). Pointing towards a genome-dependent gain in YCAY tetramers and thus a human-specific expansion for Nova binding capacity and target network. Taken together, YCAY tetramers have selectively evolved from mouse to humans and are likely a contributing factor to consider for interpreting species specific findings among Nova target networks. In addition, it is not just the abundance of YCAY tetramers that constitute Nova affinity, but rather YCAY clusters and the density of these clusters have been shown to increase Nova binding affinity. Therefore, an important future direction to explore further will be to quantitatively characterize YCAY cluster density within target groups for cross-species comparative analysis.

Bioinformatic approach

CLIP peaks are difficult to reliably map across species. Yet, it is clear that evolutionary genomic changes directly influence cross-species differences at the level of RNA regulation and should not be ignored. Therefore, in order to better tease apart species-specific Nova function and account for genome-dependent changes, pairwise whole-genome alignments, or "chain files", were utilized to map Nova-CLIP binding profiles between genome assemblies (hg38 to mm10 and mm10 to hg38 respectively)(Kuhn et al., 2013). This approach provides a direct comparison of sequence specificity, positional overlap, and infer functional consequences of Nova binding profiles independently derived from both human and mouse brain tissues on a single background genome.

More than 95% of Nova targets are considered orthologous between mouse and humans and Nova binding is heavily localized to non-coding regions such as introns and 3'UTRs. Therefore, characterizing the cross-genome mapping rates of peaks by genomic localization was of great interest. Comparison of differentially localized peaks from human (hg38) to mouse (mm10) genomes revealed 90.2% of all 3'UTR bound sequences successfully aligned, whereas only 45% of intronic binding sites aligned (Figure 3-10A). The robust decline in alignment rates for intronic peaks compared to 3'UTRs in concordant with intronic sequences being subjected to greater speciation events as compared to exons and 3'UTRs. However, the results may not wholly account for speciation as these data are likely influenced by the quality of annotation data available at the time of analysis. Similar results were observed when mouse-derived binding sites were mapped from mouse (mm10) to human (hg38) genomes, with a 94.8% success rate for

3'UTR localized binding and 51.6% success rate for intronically localized binding (Figure 3-10B). Despite the caveats and limitations for cross-genome alignments, these data suggest that the precise location of Nova regulatory sites, specifically within intronic regions, have likely been altered due to intronic speciation events.

Figure 3-10: Cross-species alignment rates of Nova peaks by genomic localization

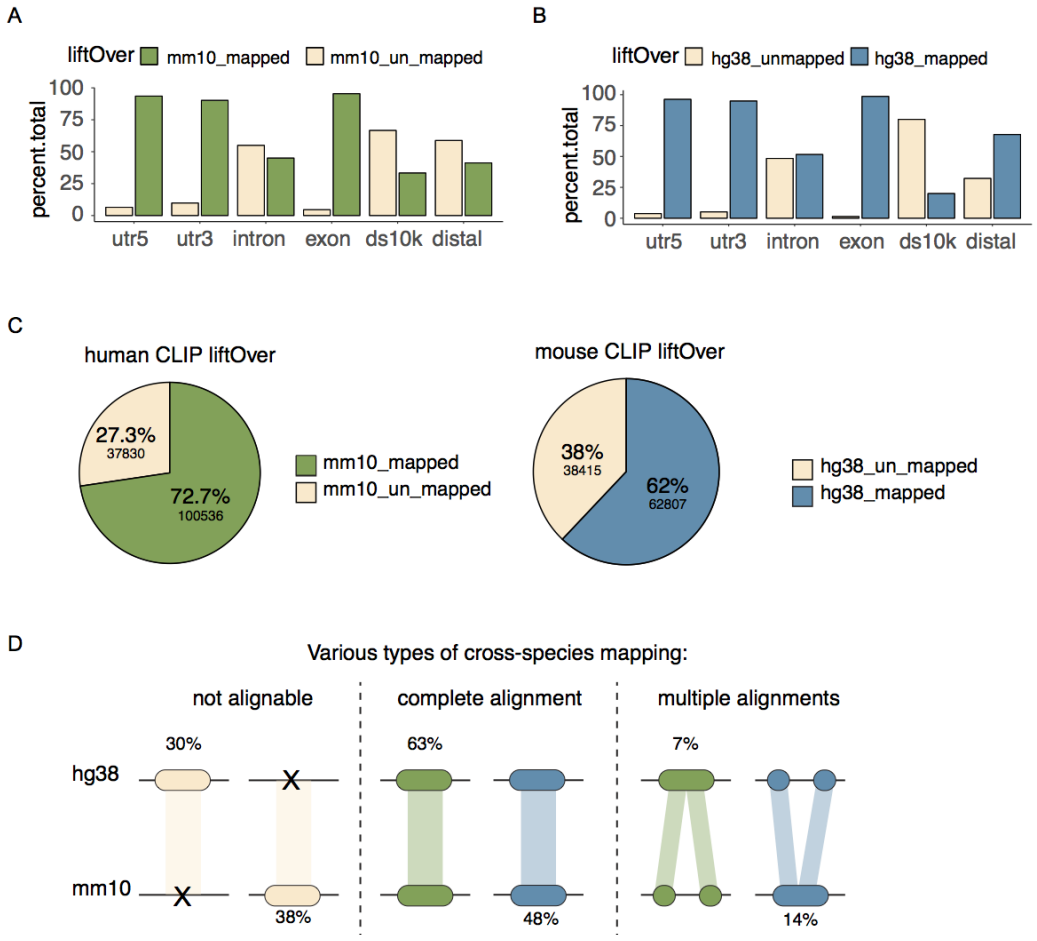


Figure 3-10: (A-B) Genomic annotation of peaks subset by alignment group. Peaks that successfully aligned to both genomes and peaks that were unalignable across genomes (beige). Percent of total represents the total peaks localized within each respective genomic annotation (3'UTR; exon; 5'UTR; intron etc.). **(A)** Human peaks (green) and **(B)** mouse peaks (blue). **(C) (D)** Schematic illustrating homology of human and mouse NOVA binding sites. Percents (top) represent percent of total human peaks for each category. 30% of human peaks were not aligned to mouse genome (no homology), 7% of human peaks aligned to multiple locations within the mouse genome (multi-aligned), 63% of human peaks aligned to the mouse genome. Percents (bottom) represent the percent of total mouse peaks per group. 38% of mouse binding sites did not align to the human genome. 14% of peaks aligned to multiple regions in the human genome, and 48% of mouse peaks aligned one-to-one from mouse to human genome.

Globally, over 90% of both mouse and human genomes can be partitioned into syntenic regions and translated across species. HITS-CLIP in the human brain revealed 138,366 significant NOVA binding sites, of which 72.7% (100536/138366) successfully translated to the mouse genome (Figure 3-10C). An alignment rate of 72.7% is significantly lower than one might expect given that over 90% of each genome can be partitioned into regions of synteny and network-level Nova regulation is significantly conserved. Identical approach was applied to mouse-derived HITS-CLIP data for mapping binding sites to the human genome, revealing an even lower alignment percentage of 62% (62807/101222) (Figure 3-10C). It is reasonable to presume that cross-genome alignment rates of Nova binding sites may be lower than expected due to the nature of Nova binding to non-coding regions, which emphasizes the ongoing need for more diverse and specialized genome conversion approaches.

An additional outcome to consider when mapping Nova HITS-CLIP data between species is the event of a single binding site mapping to multiple locations in the opposing species. In respect to Nova-RNA interactions, duplicated sequences, or multi-aligned sequences, occur within 14% of all mouse peaks, whereas only a 6.9% occurrence is observed for human peaks (Figure 3-10D). It is interesting that multi-mapping rates of mouse Nova binding sites have nearly doubled compared to the reciprocal mapping strategy. One possible explanation for such an asymmetric shift may be the higher incidence of genomic duplication events in the human genome compared to the mouse genome. An additional explanation may be that human-specific insertion events cause an interruption in Nova binding sequence. Both explanations are plausible and may contribute to species-specific patterns in Nova binding profiles and would be worthwhile for future studies to investigate in greater depth.

Species-specific Nova binding maps reveal additional levels of complexity in human RNA regulation

In order to simplify the interpretation of results, the following section only accounts for mouse CLIP data that successfully aligns in a one-to-one relationship from mouse to human genome. Total tags in peaks were summarised per target for each species accordingly. PCA was utilized in order to globally assess variance across datasets and revealed species-derived as the greatest contributing variable across datasets (Figure 3-11A). Subsequent PCA was carried out for all tags in peaks, again highlighting species-derived as the factor with greatest contribution to dataset variance (Figure 3-11B). PCA provides a global unbiased approach to visualize variance across large datasets and while it is not the most stringent approach, these initial findings encouraged efforts to further characterize species-specific differences in Nova binding maps.

Figure 3-11: Global assessment of variance between species

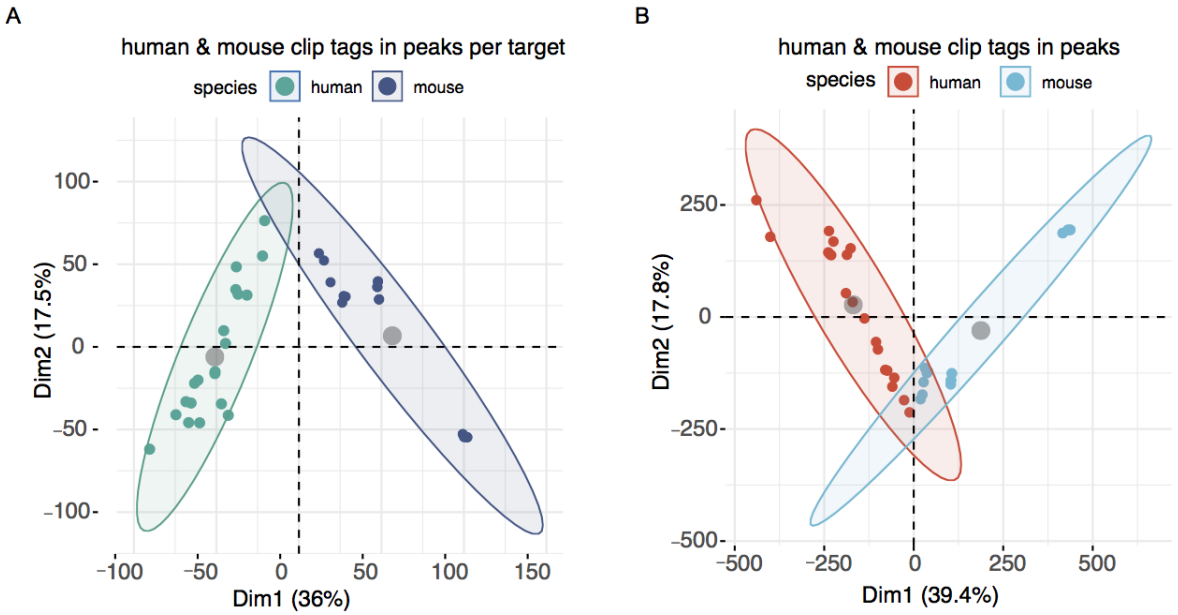


Figure 3-11: (A) Total CLIP tags in peaks per target was summarised across CLIP datasets. Dim1(36%) and Dim2 (17.5%) explain the most variation in the dataset, therefore data points clustered together are more similar to each other, than those distanced apart. Data point color represents the species the HITS-CLIP was derived (human = green and mouse = purple). Shaded ellipse represents 95% confidence interval per species. **(B)** Global assessment of variance in tags in peaks across species. Total CLIP tags in peaks were summarised for across CLIP datasets. Dim1(39.4%) and Dim2 (17.8%) explain the most variation in the dataset, therefore data points clustered together are more similar to each other, than those distanced apart. Data point color represents the species the HITS-CLIP was derived (human = red and mouse = blue). Shaded ellipse represents 95% confidence interval per species.

Conserved occupancy of precise Nova-RNA interaction sites requires that the occupied sequence aligns between genomes and overlaps directly with a human derived peak. Analysis of peak overlap revealed 16% (42724/273993) of all CLIP-peaks (both mouse and human derived) are fully conserved between in sequence, position, and experimentally defined CLIP binding (Figure 3-12A). We were next interested in quantifying the conservation of Nova occupancy with respect to genic location and found that 26.6% (31861/119940) of 3'UTR peaks, 24%

(2265/9445) of exon peaks, and 17.2% (327/1896) of 5'UTR peaks are fully conserved from mouse to human. Over half of all Nova binding sites are localized to introns (50.7%; 138935/273993), however, only 5.9% (8161/138935) are fully conserved in Nova occupancy (Figure 3-12B). These data illustrate location-dependent variance in the conservation of Nova-RNA interaction sites, further exemplifying the divergence of Nova occupancy within introns.

Figure 3-12: Intersection of human and mouse peaks

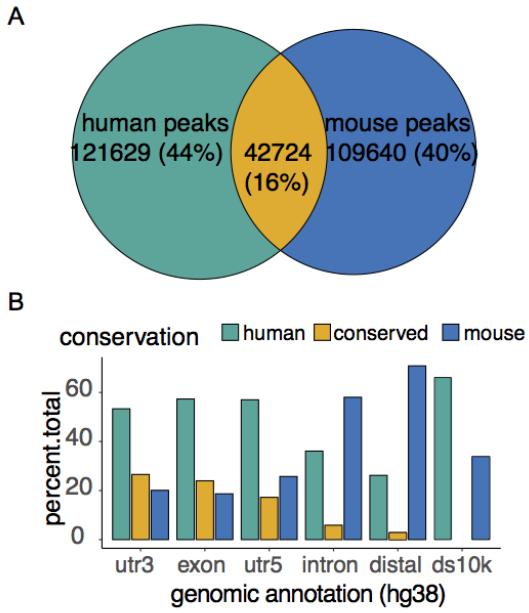


Figure 3-12: (A) Venneuler plot of human derived peak (left) overlapped with mouse derived peak (right). Total peak overlap (yellow) is expressed as percent of all Nova peaks aligned to the human genome (hg38). **(B)** Genomic distribution of mouse to human peak conservation groups. Human-specific peaks (turquoise), conserved peaks (yellow), and mouse-specific peaks (blue). Percent of total represents the total peaks localized within each respective genomic annotation (3'UTR; exon; 5'UTR; intron etc.).

Next, upstream and downstream binding profiles surrounding sites of conserved Nova occupancy were examined. The distance between all human-derived CLIP peaks was calculated relative to the nearest mouse-derived CLIP peak. Enrichment for all human peaks over neighboring mouse peaks reveals a strong signal persists +/-200nt of a mouse peak (Figure 3-13A). Peaks were then subset according to genomic location highlighting the distinct binding landscapes for 3'UTR peaks, exon peaks, and intronic peaks (Figure 3-13B). Binding profiles surrounding conserved 3'UTR peaks suggests that many human binding sites have shifted in location up to 300nt downstream from mouse binding sites. Peak distribution surrounding exons highlights the robustness of precise overlap between species. Lastly, intronic peaks appear to display a similar patterning to 3'UTRs, with a slightly wider range. Subsetting binding sites according to the genomic annotation of peaks allows for assessment of the distributions specifically in exons, 3'UTRs, and introns.

Figure 3-13: Human peak enrichment relative to distance from mouse peak

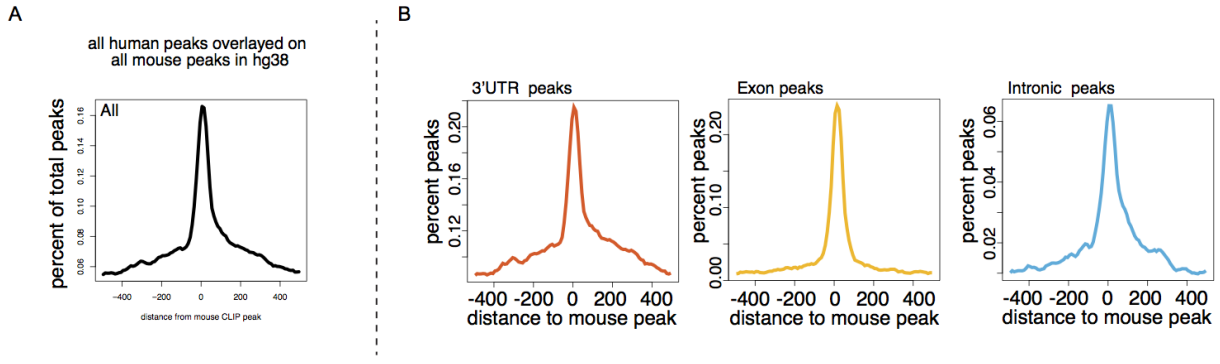


Figure 3-13: (A) Cumulative distribution of all human peaks over mouse peaks. Distance = 0 is an indication of directly overlapping (conserved) binding sites. **(B)** Distribution of human 3'UTR peaks relative to mouse 3'UTR peaks (orange). Distribution of human exon peaks relative to mouse exon peaks (yellow). Distribution of human intron peaks relative to mouse intron peaks (blue).

Given the stringency required for a peak to be considered as fully conserved, and the distribution analysis revealing high density of positional shifts +/-400nt, we next characterized Nova binding sites that may have undergone positional genomic shifts (Figure 3-14A). In order to assess instances of partially conserved peaks, the distance to the nearest mouse peak was calculated for all human peaks. If binding sites were within 50nt of each other, they were considered conserved (~12%) and peaks +/- 1kb from a mouse peak were grouped as partial upstream (14.4%) or downstream (15.3%) conservation (Figure 3-14B). Our analysis identified an additional 29.7% of all human peaks having undergone positional drift. Global distribution of human conserved peaks, both fully conserved and drifted, within the 1kb defined genomic space revealed a distinct enrichment of human peaks within 250nt of a mouse-derived peak, both upstream and downstream (Figure 3-14C). The gradual decline in neighboring peaks +/-500nt reveals that 58.4% of all human Nova binding sites are more than 1kb from a human mapped mouse binding site (Figure 3-14B). Further emphasizing the utility of species specific binding maps. Genomic distribution analysis reveals that more than 50% of those peaks are localized to introns and 40% are localized to 3'UTRs (Figure 3-15A-B). Despite loosening the definition of conserved Nova-RNA interactions, it is evident that human intronic landscapes have undergone significant changes from mouse to human.

Figure 3-14: Schematic illustrating how peak groups were determined

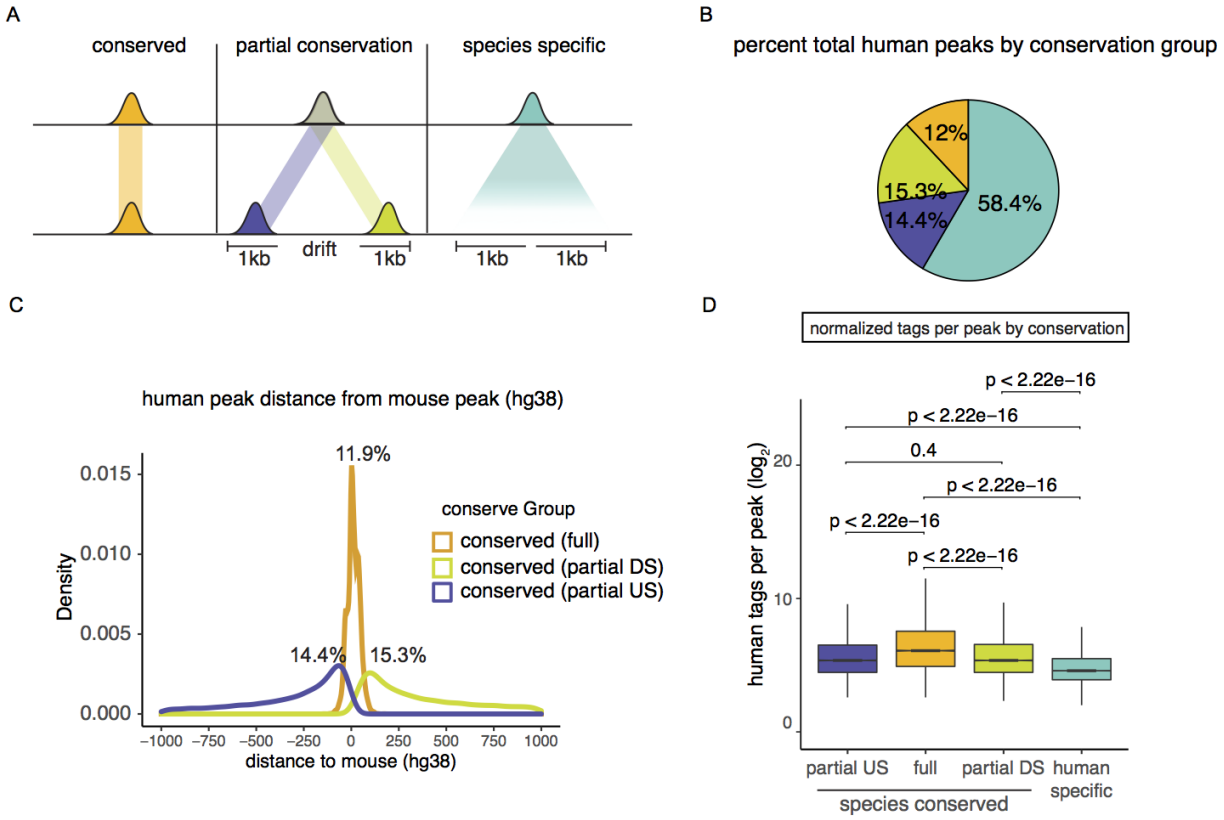


Figure 3-14: (A) Schematic illustration of three different types of peaks on RNA target, cross-species conserved binding (yellow), partially conserved binding- defined as peaks that have drifted in location within 1kb of opposing species binding site (green/purple), and human-specific binding (turquoise). **(B)** Percent total human peaks within each respective category. **(C)** Distribution of conservation group peaks over distance to closest mouse peaks (x-axis). **(D)** Normalized tags per peak subset according to conservation group.

Figure 3-15: Genomic distribution of peaks grouped by conservation

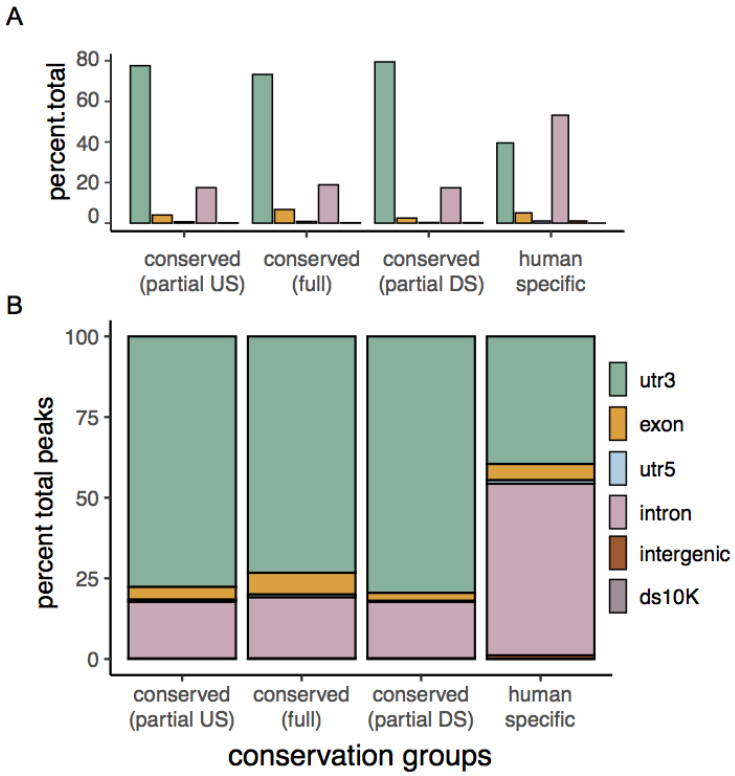


Figure 3-15: (A) percent of total peaks grouped according to conservation groups (partial US, conserved full, partial DS, and human-specific). Percent total represents total peaks for each group. **(B)** Stacked barplot illustrating percent of total peaks grouped according to conservation groups (partial US, conserved full, partial DS, and human-specific). Each conservation group sums to 100%.

Our findings provide a detailed description of the impact of species-specific changes on Nova regulatory landscapes with nucleotide resolution. Concern was raised that our findings may be biased in favor of categorizing human-specific peaks, given that it was carried out using the human genome (hg38) as the background. In order to address this concern and validate the findings, all analysis pipelines were applied using the mouse genome as the background (mm10), which independently confirmed our findings (data not shown). This study has elucidated key

functional RNA regulatory elements that have diverged and/or expanded across species. RNA expression profiles point towards a high level of conservation between species. However, here we demonstrate that the precise mechanisms of RBPs modulating RNA steady state levels may be more complex to discern across species.

Discussion

Nova regulates target RNAs that are highly orthologous in mouse and human. We've shown that more than half of Nova targets identified in both mouse and human are bound in both species. Previous studies have demonstrated that Nova proteins themselves are deeply conserved throughout vertebrate evolution. However, Nova targets have significantly evolved (Jelen et al., 2007). Therefore, it is unsurprising that the core mouse-defined functions of Nova regulation are overall conserved from mouse to humans pertaining to (1) membrane/action potential (neuron firing); (2) glutamate signaling (excitatory synapses); (3) neuron development/differentiation (neurogenesis). One might postulate that while Nova targets RNAs involved in the aforementioned processes in both mouse and human, the exact binding sites and tag profile has likely evolved given that Nova proteins evolve more slowly compared to the precise Nova-RNA interaction sites.

Nova regulation that has diverged from mouse to human (ie. mouse-specific) included functional clusters that overlap with the conserved functional clusters such as: (1) GABA signaling (inhibitory synapses), axonogenesis, membrane potential and (2) modification dependent protein catabolic process, cell response to radiation, polyubiquitination. Taken

together, these findings suggest that RNA targets important for inhibitory neuron function in humans may have diverged significantly in humans resulting in a loss of Nova regulatory capacity. Conversely, our findings could also be explained by divergent expression patterns of Nova in the human brain compared to the mouse as Nova protein expression across cell-types are not well defined in the human brain.

Functional clusters that have recently evolved as Nova regulated processes (i.e. human specific) are (1) positive regulation of mRNA catabolic process- deadenylation dependent decay, amide biosynthetic process, translation. (2) plasma membrane bound cell projection, negative regulation of response to stimulus, calcium ion transmembrane transport, negative regulation of cell motility. Deadenylation is the shortening of the polyA tail, which functions to slow down translation or even prevent translation. The emphasis on human-specific regulation of calcium ion transmembrane transport is most likely coupled with targets important for cell motility since altering calcium gradients aids in the control of cell motility. Higher levels of cognitive ability in the human brain relative to rodents may relate to (1) the possibility of higher degrees of steady state RNA control and translation through Nova regulated processing and (2) regulation of critical cellular processes relating in adult brain (as studied here), to Ca2+ ion regulation and membrane-related phenomenon, plausibly related to neuronal plasticity. Overall, our findings emphasize functional biological processes that have gained Nova regulation and thus evolved higher levels of control specifically in human neurons.

One possible explanation for increased 3'UTR binding is that cytoplasmic Nova proteins remain associated with previously alternatively-spliced mRNAs as a component of exon-exon

junction complex during nuclear export. At some point during mRNA processing/transport, Nova proteins disengage from mRNA complexes and return to the nucleus. Interestingly, it has been shown that 11-22% of cytoplasmic Nova proteins are associated with brain polyribosomes in the mouse brain (C. Fraser, unpublished). Nova dependent changes in polyribosomal associated mRNAs are not robust in the mouse brain but perhaps Nova has evolved to play a more active role in alternative translation processes in the human brain. Our findings highlight potential future directions and research areas worth exploring human-specific features of Nova mediated RNA processing in more detail, which offers the potential to yield insight into what is mechanistically unique in the human brain and contributing to our unique cognitive abilities.

Chapter 4: Developing models for understanding Nova1-dependent splicing in the human

Introduction

There is no ideal human model when it comes to modeling brain function as the ideal system to study is not an experimental. Mouse models are some of the most widely utilized models for neuroscience research. The development of genetic tools for efficiently disrupting gene function in mouse models has fueled decades of research and discoveries pertaining to neurological disease. However, when it comes to human cognition, structure, and physiology of the human brain there are many obvious discrepancies between mice and humans and it is increasingly important to define both the commonalities as well as stark differences in order to authentically interpret human disease modeling using mice. Recent advances have developed highly sophisticated *in vitro* human brain models such as iPSC-derived neuronal cultures (2D) and more recently brain organoids (3D). One caveat to in vitro models is the time spent to develop, maintain, and grow up these complex models. For such reasons, they are not a quick or efficient system and less than ideal for high throughput approaches. The aim of the next series of experiments was to create a fast, convenient, and manageable model to identify and manipulate Nova1 dependent splice changes. While HeLa cells may be far from the biologic context of the brain, they are molecularly "clean" of Nova expression in that they do not endogenously express Nova1 or Nova2. Furthermore, they are human-derived, which was the primary objective for the following set of experiments.

Results

Ectopic expression of Nova1 in HeLa cells

HeLa cells were selected to model Nova-dependent splice events because while HeLa cells do not endogenously express NOVA1, they do express a surprising number of neuronal transcripts. In order to have precise temporal control over Nova expression levels, we created a tet-inducible Nova expressing HeLa cell line. Nova protein expression post-doxycycline induction was confirmed via immunofluorescence staining and quantification of Nova1 expressing cells showed 60% of cells express Nova1 after Day 1, with over 80% of cells expressing Nova1 by 72 hours (Figure 4-1A-B). There was concern that ectopic expression of Nova1 was leading to cell death in cultures, fluorescence activated cell sorting with TOPRO staining showed no significant difference in cell death between cell lines (Figure 4-1C). RNA-sequencing time course further confirmed Nova1 RNA expression in HeLa^{Nova1+}cell-lines (Figure 4-1D). In order to validate the effect of Nova1 expression on alternative splice profiles in a human system on a genome-wide scale, RNA-sequencing was carried out for both cell lines for Day 0, Day 1, Day 2, and Day 3 respectively (n = 3 per timepoints).

Figure 4-1: Ectopic expression of Nova1 in HeLa cells over time

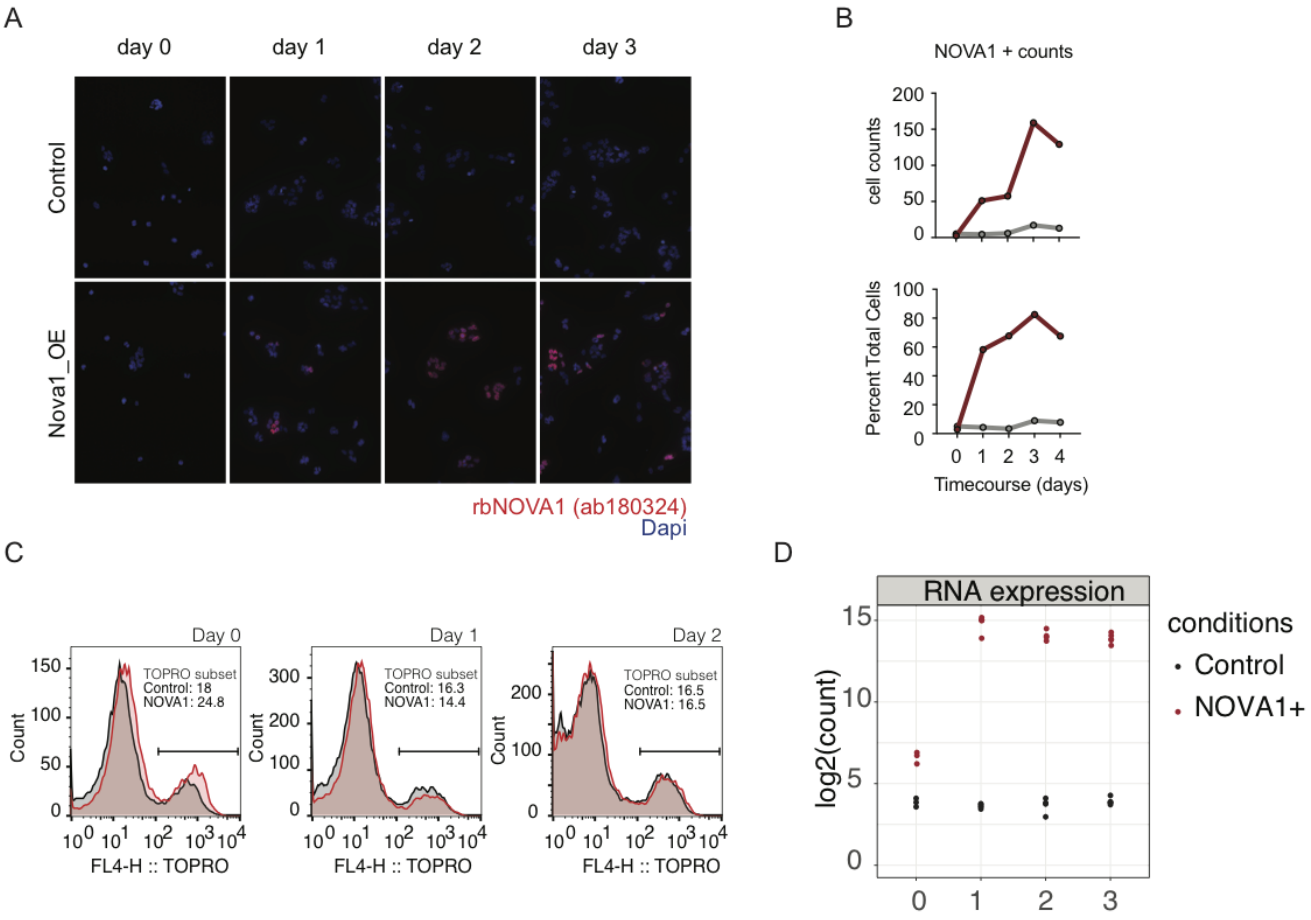


Figure 4-1: (A) Immunofluorescence staining of HeLa^{Nova1-} (top rows) and HeLa^{Nova1+} (bottom rows) confirming ectopic expression of Nova1 (red) post-doxycycline induction, Dapi is shown in blue. **(B)** Quantification of Nova1+ cells over time, average cell counts (left) and percent total cells (right) (n = 10 dishes). Figure is representative validation for 1 of 3 clones selected and validated (each individually matched with control cell line). All cell lines were validated. **(C)** FACS analysis for cell death (TOPRO+) in HeLa^{Nova1+} cells (red) vs. HeLa^{Nova1-} (grey) cell lines post-doxycycline induction over time. No significant differences were detected. **(D)** RNA-seq normalized expression values for NOVA1 steady state RNA levels over time for HeLa^{Nova1-} cell line (black) and HeLa^{Nova1+} cell line (red). (n = 4).

High-throughput identification of Nova1 dependent exons

Experimental approach for high-throughput identification of Nova1-dependent splicing events is illustrated in Figure 4-2A. Briefly, differential splice changes between HeLa^{Nova1+}compared to matched control HeLa^{Nova1-} were measured using rMATs software, identifying 2,223 significant Nova1-dependent splice changes (Figure 4-2B). In order to determine whether the high-throughput *in vitro* model could recapitulate NOVA-dependent splicing changes identified *in vitro* events were filtered for those with significant *in vivo* intronic binding as defined via human brain CLIP. Culminating in 1,376 events within 991 RNAs. The greatest number of changed splicing events were skipped exons (SE) (Figure 4-2C).

Figure 4-2: High-throughput identification of Nova1 dependent splice changes

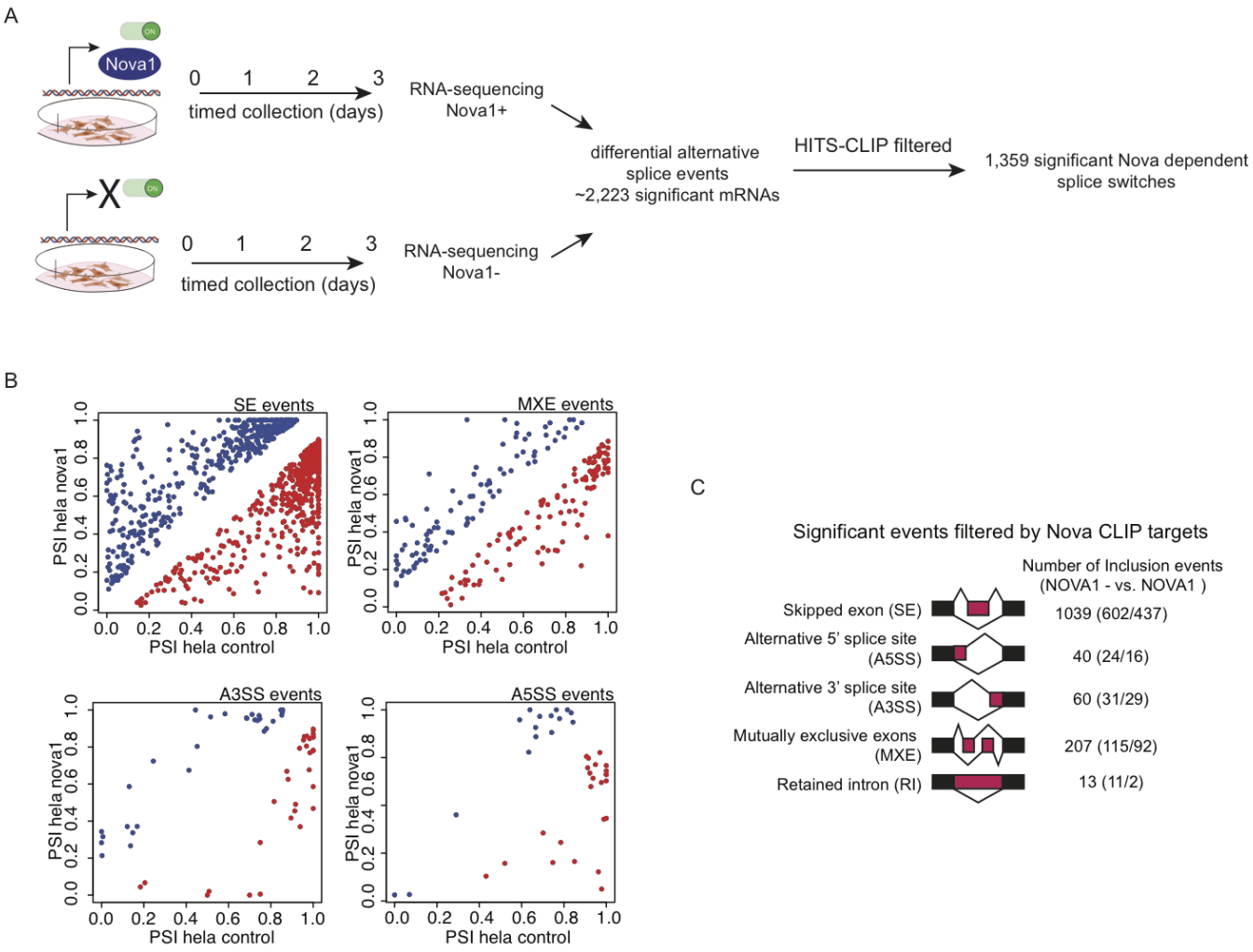


Figure 4-2: (A) Schematic illustrating experimental approach. HeLa^{Nova1-} and HeLa^{Nova1+} cells were grown in parallel and treated with doxycycline on Day 0. Cells were harvested in pairs at Day 0, Day 1, Day 2, and Day3 and flash frozen for batch preparation of RNA-sequencing libraries (left). Libraries were sequenced and differential splicing analysis identified over 2,000 significant splice changes. Events were filtered by *in vivo* HITS-CLIP targets to identify "high-confidence" targets. **(B)** Percent spliced inclusion plots (single exon events; mutually exclusive events; Alternative 3' splice site; Alternative 5' splice site): HeLa^{Nova1+} PSI values over HeLa^{Nova1-} PSI values. **(C)** In order to identify brain-relevant splicing events, data was filtered for significant Nova targets identified via human brain HITS-CLIP.

Gene ontology analysis showed these RNAs to be enriched in molecular functions pertaining to neurons such as regulation of post- synaptic membrane, axon part, and receptor localization to synapse (Figure 4-3A-B). Due to the robust changes observed in neuronally-relevant target pathways, we were interested in whether Nova1 regulated exons were encoding for relevant functional domains. Enrichment analysis across splice targets revealed Nova regulated splice events were enriched for PH-like domains, SH3 domains, and CH domains (Figure 4-3C). The unexpected enrichment of neuronal splice targets prompted further investigation to examine conservation of these events in the mouse. To discern Nova-splice targets that are conserved and species-specific between mouse and human, the target lists were intersected revealing 384 common splice-targets, a 15% overlap between species (Figure 4-3D). It is important to note that given the disparity in RNA profiles between experimental models utilized, the calculated percent conserved is likely an under-representation of the true cross-species overlap.

Figure 4-3: Nova1 splice target enrichment

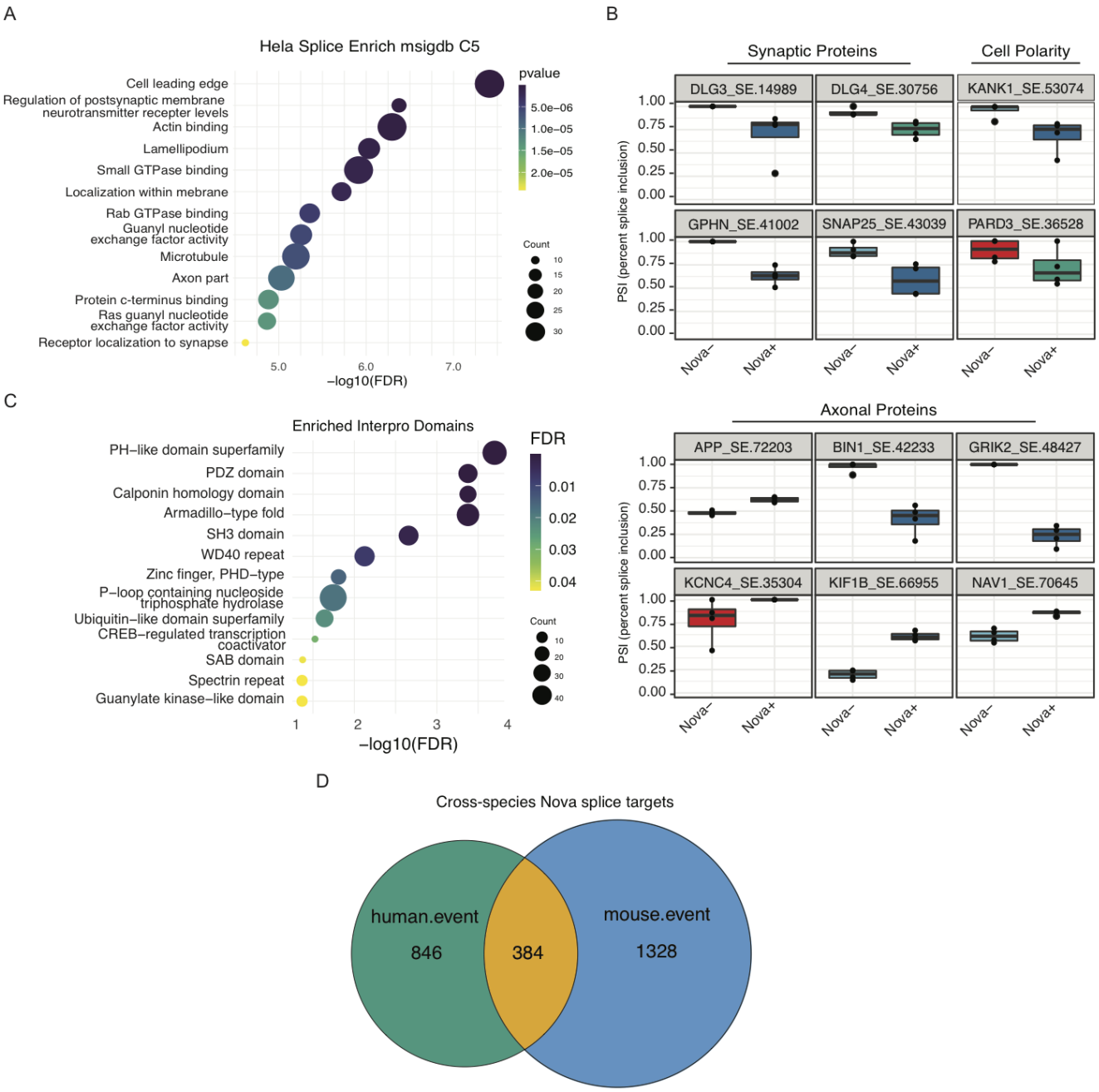


Figure 4-3: (A) Nova1 splice targets in HeLa cells are enriched in targets important for neuronal function. **(B)** Boxplots of neuro-related significant HeLa^{Nova1-} and HeLa^{Nova1+} splice changes PSI values determined utilizing rMATs software. **(C)** Nova1 splice targets are enriched in domains important for proper cellular function. **(D)** Venn diagram illustrating the intersection of Nova splice events identified with ectopic expression of Nova1 in HeLa cells (green) with Nova mediated splicing elucidated using knockout mouse models (blue).

Top most significant Nova1-dependent events

As an additional approach to validate alternative splice targets identified via high throughput sequencing methods, the top most significant Nova1 dependent splice switches were selected for further validation (Figure 4-4A). Cumulative analysis of all events that were screened with semi-quantitative PCR methods significantly correlated (Pearson; $R = 0.95$, $p < 2.2e-16$) with high-throughput PSI values (Figure 4-4B). One interesting observation is that 12/12 of the top most significant events are conserved Nova dependent splice events in mouse and human (Figure 4-4C). One hypothesis for this may be that the conserved Nova regulated events are likely the most important or physiologically relevant given that Nova itself is highly conserved. Although all events were previously identified as Nova dependent in the mouse brain, the binding profile within introns flanking the events were divergent between mouse and human CLIP binding maps. Taken together, these data aid in the validation of Nova1 dependent splice events in a relevant human experimental model.

Figure 4-4: Top most significant splice changes with ectopic Nova1 expression in HeLa cells

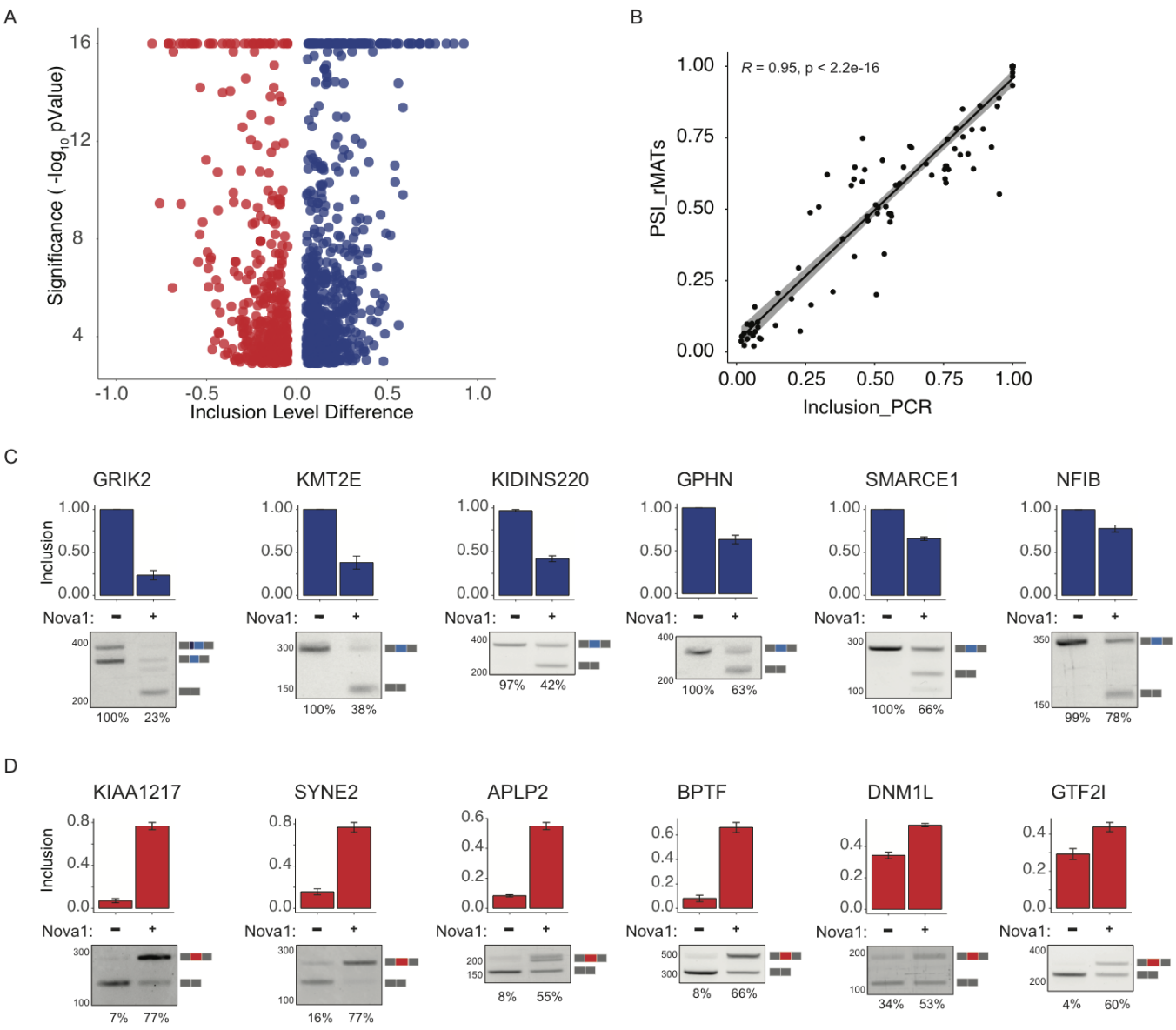


Figure 4-4: (A) Volcano plot of significant splice inclusion level differences (x-axis) identified from ectopic expression of Nova1 in HeLa cells compared to matched control cells lines at 72 hours (n = 4). **(B)** Correlation analysis demonstrated utility of semi-quantitative approach. Significant positive correlation between rMATs calculated PSI values (y-axis) and semi-quantitative PCR methods utilizing imageJ for measurement of band intensity, correlation coefficient, Pearson R = 0.95. **(C)** Semi-quantitative PCR validation of top most significant events (n = 6) where Nova1 promotes silencing of target exon relative to control. **(D)** Semi-quantitative PCR validation of top most significant events (n = 6) where Nova1 promotes inclusion of target exon relative to control.

Discussion

Here we've developed an *in vitro* experimental system to validate Nova binding sites identified by post-mortem human brain HITS-CLIP. We used a human cell line, HeLa, that expresses a number of human brain targets. We find great concordance in the data, such that 100% of the top significant splice changes validated via semi-quantitative PCR methods. There are clear limitations to the model at hand as it is obviously not neuronal in nature. In the next chapters we explore and characterize several in vitro models that are more "neuronal" in nature. However there are distinct advantages to using HeLa cells as a model system as it permits us to focus on several key targets identified from our human brain HITS-CLIP experiments in an experimentally addressable system.

Potential disease relevance for top most significant Nova1-dependent splice events

General Transcription Factor II (GTF2I)

One of the top events identified is Nova1-mediated inclusion of exon 10 in General Transcription Factor II (GTF2I). GTF2I is a striking Nova target due to its association with the neurodevelopmental disorder, Williams-Beuren syndrome (WBS). WBS is a multisystem disorder that is characterized by distinct facial features, mild to moderate intellectual disability, and unique personality traits. GTF2I has specifically been associated with the cognitive phenotypes of WBS rather than the physical characteristics. The cognitive phenotypes include difficulty with visual-spatial tasks such as drawing and puzzles, however they tend to do well on vocal tasks and memorization. There is an interesting social phenotype linked with WBS as

affected individuals are highly social with engaging and outgoing personalities. Due to these unique characteristics, GTF2I is often referred to as the "pro-social" gene. Our study highlights Nova-mediated inclusion of GTF2I exon 10 as a mouse to human conserved Nova regulated exon. However, we've taken an exceptional interest in Nova1's regulation of this exon because the affinity of Nova1 to this site is human-specific, with regionally distinct binding patterns. Unpublished data using a humanized Nova1 mouse model identified markedly increased Nova binding in the humanized Nova1 mouse compared to WT mouse at a relevant exon 10 binding site (unpublished). Additionally, a recent study characterizing archaic Nova1 compared to modern Nova1 using human brain organoids has independently arrived at the same finding in regards to human-specific Nova1 regulation (Trujillo et al., 2021). Taken together, the human-specific increase in Nova1 regulation of GTF2I exon 10 would be an interesting target to work up for future studies, especially in the context of human-specific social behaviors and syndromes.

Lysine methyltransferase 2E (KMT2E)

KMT2E is a histone methyltransferase that mono- and dimethylates 'Lys-4' of histone H3 (H3K4me1 and H3K4me2) and truncating mutations in exons 26-27 leads to global neurodevelopmental delay and epilepsy phenotypes (O'Donnell-Luria et al., 2019). Our study highlights exon 27 as a conserved Nova1 regulated exon. Thus, KMT2E may be of interest for future investigations using disease-relevant mouse-models of epilepsy for better phenotypic characterization.

Kinase D-interacting substrate, 220-KD (KIDINS220)

Nova1-mediated alternative regulation of KIDINS220 exon 25 is conserved in the mouse. KIDINS220 is a scaffolding protein that directs maturation of axons and dendrites through coordinating neurotrophin signaling pathways. Mutations in KIDINS220 that give rise to truncated protein isoforms have been found to be causative for a spasticity phenotype observed in spastic paraplegia patients carrying the mutation. Additional phenotypes observed in these patients include intellectual disability, nystagmus, and obesity (Josifova et al., 2016).

SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily E, Member 1 (SMARCE1)

SMARCE1 is a chromatin remodelling SWI/SNF gene and causative mutations in this gene have been identified in patients with Coffin-Siris syndrome. Coffin-Siris syndrome is a rare genetic disease characterized by mental retardation, delayed development, and coarse facial features (Boycott et al., 2013). Nova1 binds upstream of SMARCE1 exon 4 to promote silencing of the exon, human brain HITS-CLIP data identifies several highly bound and targettable intronic Nova regulatory sites and thus may be beneficial in modulating expression levels with transcript specificity in the context of disease.

Spectrin Repeat Containing Nuclear Envelope Protein 2 (SYNE2)

Disease-causing germline mutations in SYNE2, Spectrin Repeat Containing Nuclear Envelope Protein 2, have been identified in patients with childhood onset Autosomal dominant Emery-Dreifuss muscular dystrophy (EDMD). EDMD is a debilitating and slowly progressive muscle wasting disease with limited options available for management of the disease (Heller et al., 2020). Our data identifies SYNE2 as a conserved Nova1 target and future study into

Nova1-specific regulation of SYNE2 expression levels would be imperative towards development of therapeutic interventions and treatment options.

Bromodomain PHD finger transcription factor (BPTF)

Nova1-mediated inclusion of BPTF exon 5 is a mouse to human conserved Nova regulated event with fully conserved Nova-RNA interaction sites downstream of the exon. Additionally, specifically in our human datasets, Nova1 promotes inclusion of downstream exon 18. Taken together, Nova1 alternative regulation of BPTF appears to have expanded in human systems. Interestingly, *de novo* variants spanning BPTF have been identified in patients with variable degrees of developmental delay and intellectual disability (Stankiewicz et al., 2017). Three variants lie within a region that is highly regulated by Nova, therefore developing ways to modulate Nova function at these splice sites would be hugely beneficial towards developing therapeutics for the treatment of such rare disorders.

Dynamin 1-like (DNM1L)

Nova1 modulates several alternative exons within DNM1L. Briefly Nova1 promotes silencing of exon 3 and exon 7, while promoting inclusion of exon 9, exon 16, and exon 17. DNM1lL mediates membrane remodeling in addition to regulation fission and fusion equilibrium in mitochondria. Dysfunction in the balance of mitochondria fusion and fission has been linked to diseases such as Charcot-Marie-Tooth Type 2A, optic myopathy, Alzheimer's, and Parkinson's. It has also been reported that de novo mutation in exon 3 results in a condition of rapid cognitive decline (Hogarth et al., 2018).

Gephyrin (GPHN)

Our data and previous mouse studies have identified GPHN as a Nova1 splice target, specifically alternative regulation of exon 8 (Dredge and Darnell, 2003). The synaptic protein Gephyrin is a dendritically localized protein that dynamically controls synaptic strength and plasticity at inhibitory synapses. Aberration in Gephyrin protein function has been implicated in both Alzheimer's disease and epilepsy, further solidifying the important role of Nova1 and Gephyrin at the inhibitory synapse (Groeneweg et al., 2018). The conserved and reproducible mechanisms of Nova mediated regulation of GPHN indicate that it may be an attractive target for *in vivo* targeting as a way to modulate inhibitory/excitatory circuits in the mouse brain.

Chapter 5: Nova regulation in iPS cells

Introduction

Nova regulation in the human brain highlights the robust expansion of Nova1 expression and Nova-RNA interactions identified in the human cortex. Additionally, the increase inNova-RNA interaction sites localized to 3'UTRs hint towards human-specific Nova biology. Advances in experimental modeling for studying the human brain have come a long way in recent years and initial intentions for functional human-relevant Nova studies were to leverage these advances. We sought to utilize human derived induced pluripotent stem cells (hiPSCs) as an experimental model for human-relevant functional studies and extensively characterized Nova binding map, target network, and expression patterns over time.

Shinya Yamanaka and colleagues were the first to show that somatic cells could be reprogrammed into an embryonic-like stem cell state (Takahashi and Yamanaka, 2006). Their ground-breaking research showed that the introduction of four transcription factors, Oct4, Klf4, Sox2, and C-Myc, were sufficient for the transformation. Expression of these factors would promote dedifferentiation of the somatic cells into iPSCs. The capacity for self-renewal of iPSCs is limitless and they are able to differentiate into all three germ layers: ectoderm, mesoderm, and endoderm. The technology developed by Dr. Yamanaka earned the Nobel Prize in Medicine in 2012. iPSC technology allowed for better modeling approaches on a per-patient basis because researchers could now create a cell line for a single patient that could be differentiated into any relevant cell type in order to study disease-causing mutations, all from a single biopsy. Shortly

after the development of iPSC-technology, cortical neuron differentiation methods were established for the differentiation of iPSCs into cortical neurons (Chambers et al., 2009). Seminal work opened up new possibilities in the neuroscience field as one could now study neurologic disorders in physiologically functional human neurons by collecting skin biopsies of living patients and reprogramming them into iPSCs.

Results

Human induced pluripotent stem cells: hIPS-cells

HITS-CLIP characterization of Nova regulation in hIPS-derived cortical neurons

To characterize NOVA-mediated RNA regulation and define NOVA-RNA interaction sites in human induced pluripotent stem cell derived neurons (hiPSC), we performed cross linked-immunoprecipitation followed by high throughput sequencing (HITS-CLIP) in differentiated cortical neuron cultures. Over digested (OD) and partially digested RNA fragments can be visualized just above the 50kDa mark on the representative HITS-CLIP autoradiogram (Figure 5-1A). There was no detectable signal from non-crosslinked Nova IP controls. Average Nova binding frequency at intron-exon junctions shows enrichment of Nova binding just proximal to the junction sites (Figure 5-1B). Interestingly, the binding distribution is indicative of stronger downstream intronic binding relative to upstream introns (Figure 5-1B). Additionally, hiPSC-cortical neuron CLIP had an additional enriched site on the exon region that was not observed among other CLIP data sets. Genomic distribution of Nova CLIP peaks showed ~45% of peaks were localized to introns and ~49% of peaks were localized to 3'UTRs (Figure 5-1C). It

is encouraging that the genomic distributions are more similar to human brain distributions compared to mouse brain distributions.

Figure 5-1 Nova HITS-CLIP in human IPS-cell derived cortical neurons

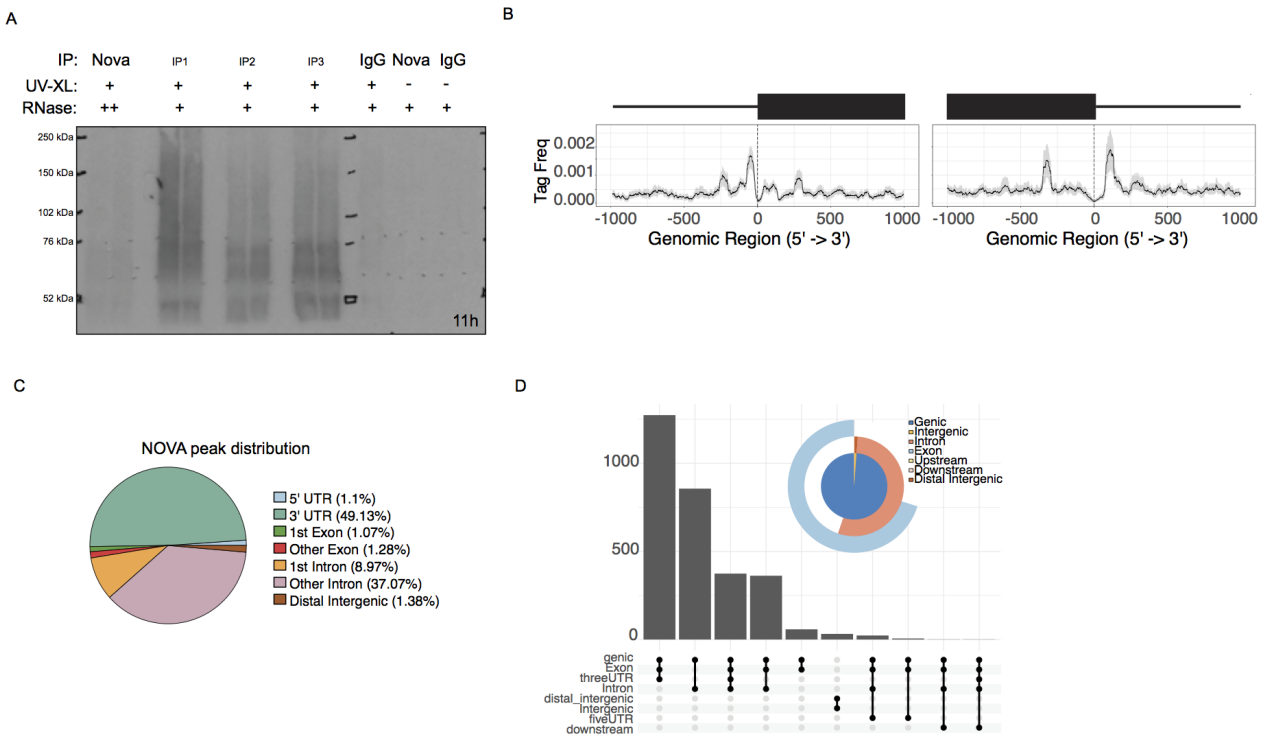


Figure 5-1: (A) Representative autorad from NOVA-HITS CLIP in human iPSc cortical neurons. **(B)** Average normalized tag distribution at intron-exon junctions. Average normalized tag distribution across introns and exons for Nova CLIP in hIPS-derived neurons. **(C)** Genomic distribution of significant Nova peaks in hIPSc cortical neurons. **(D)** Visualization of overlap in genomic annotation of Nova IPSc cortical neuron peaks. Upset plot demonstrating the annotation overlap of Nova binding sites. Conceptual overlap visualized in venn pie diagram.

One characteristic of Nova regulation of alternative RNA processing is that the regions bound by Nova frequently can be annotated by multiple categories. One aspect of alternative splicing regulation is that Nova binds alternative exons to promote exon silencing, or skipping.

Therefore, at events such as these the peak may be annotated as an exon or an intron. Additionally, alternative polyadenylation sites can be located in regions other than 3'UTRs such as 5'UTRs, exons, introns. Due to the possibility of overlap depending upon the transcript considered, we utilized an upset plot as a way to resolve and visualize associations between these events (Figure 5-1D). This approach revealed >1000 binding sites in regions that are annotated as both 3'UTRs and exons, depending upon the context. Additionally, >400 binding sites can be considered exon, intron, and 3'UTR regions depending on the transcript in question. Pathway analysis was performed on all Nova bound targets in order to globally characterize Nova regulation in hiPSC-derived neurons and highlight Reactome Pathways that could be studied using this experimental model for studying Nova regulation (Figure 5-2A). Unsurprisingly, the most significant and most enriched pathway highlighted was Neuronal System, secondly was the regulation of expression of SLITs and ROBOs which is especially encouraging for the study of human-specific SRGAP protein family members. In order to assess the experimental model for disease relevance we looked for significant enrichment of Disease Gene Net (DGN) terms within Nova targets in hiPSC-derived neurons (Figure 5-2B). These data were exceptionally promising as the disorders included Major Depressive Disorder, Autism Spectrum Disorder, and Attention deficit hyperactivity disorder.

Figure 5-2 Global Pathway Analysis Nova targets in hIPS-derived neurons

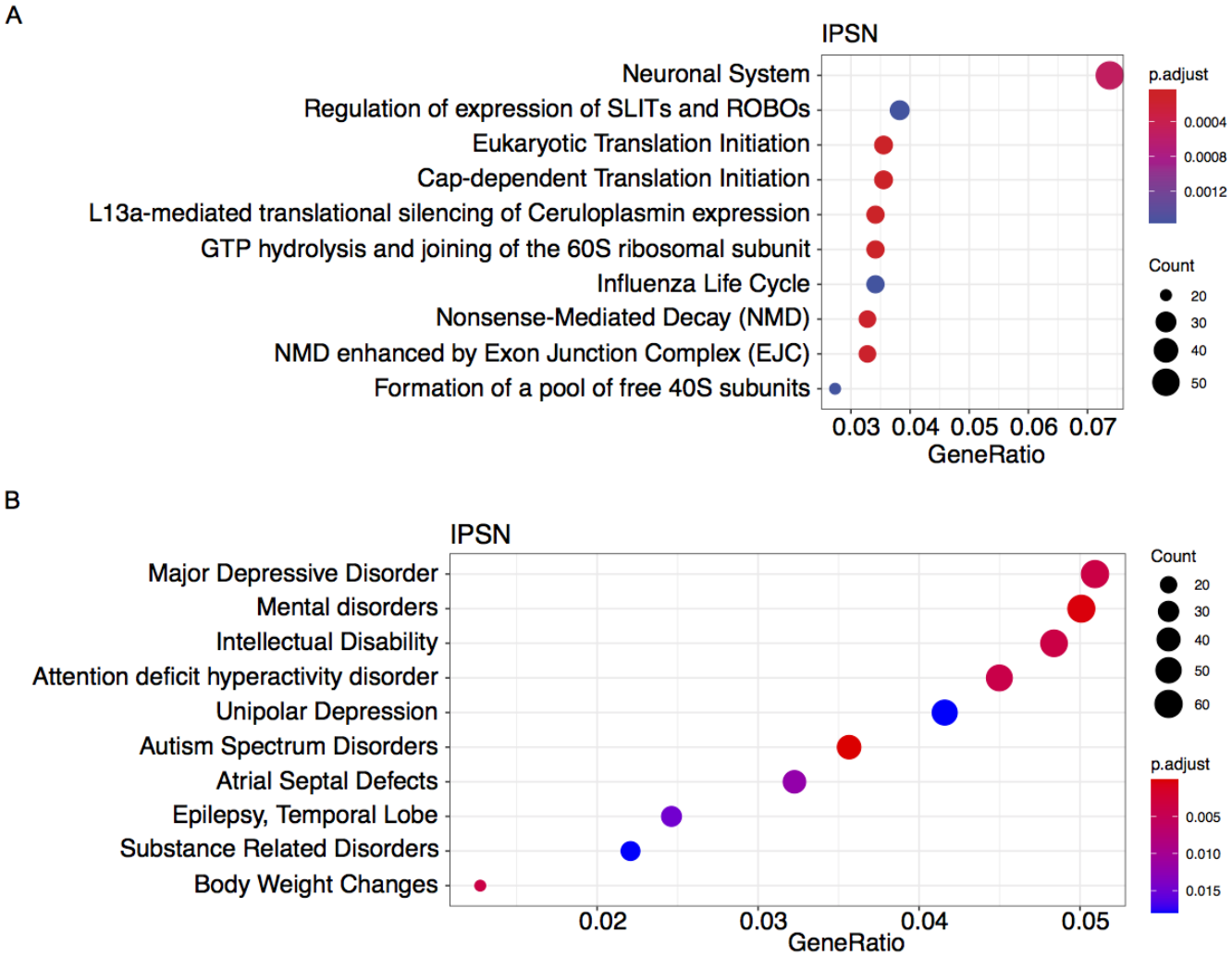


Figure 5-2: (A) Reactome pathway analysis of Nova targets in hIPS-derived cortical neuron HITS-CLIP. **(B)** Disease GeneNet (DGN) pathway enrichment analysis of Nova targets in hIPS-derived cortical neuron HITS-CLIP.

Characterizing Nova expression in human neurons over time

Developmental expression patterns of Nova1 and Nova2 are different in the human cortex compared to the mouse cortex (Figure 5-3A-B). We utilize human induced pluripotent stem cells (hiPSC) differentiated into cortical neurons in order to thoroughly characterize Nova expression patterns and regulation networks over the course of human cortical neuron development (Figure 5-4). High-throughput characterization of Nova1 and Nova2 levels via RNA-sequencing showed Nova1 to be more highly expressed than Nova2 during cortical neuron development, which is concordant with expression patterns observed from RNA-sequencing of the human cortex (Figure 5-4). Expression markers for differentiation were plotted at time points profiled to confirm neuron differentiation status (Figure 5-4). Additionally, Nova protein expression was confirmed at selected time points via western blot analysis (Figure 5-4). Differential analysis of gene expression profiles between pluripotent iPSC compared to differentiated neurons reveal > 5,000 significantly differentially expressed RNAs (Figure 5-4).

Figure 5-3 Developmental expression patterns of Nova are different in mouse and human brain

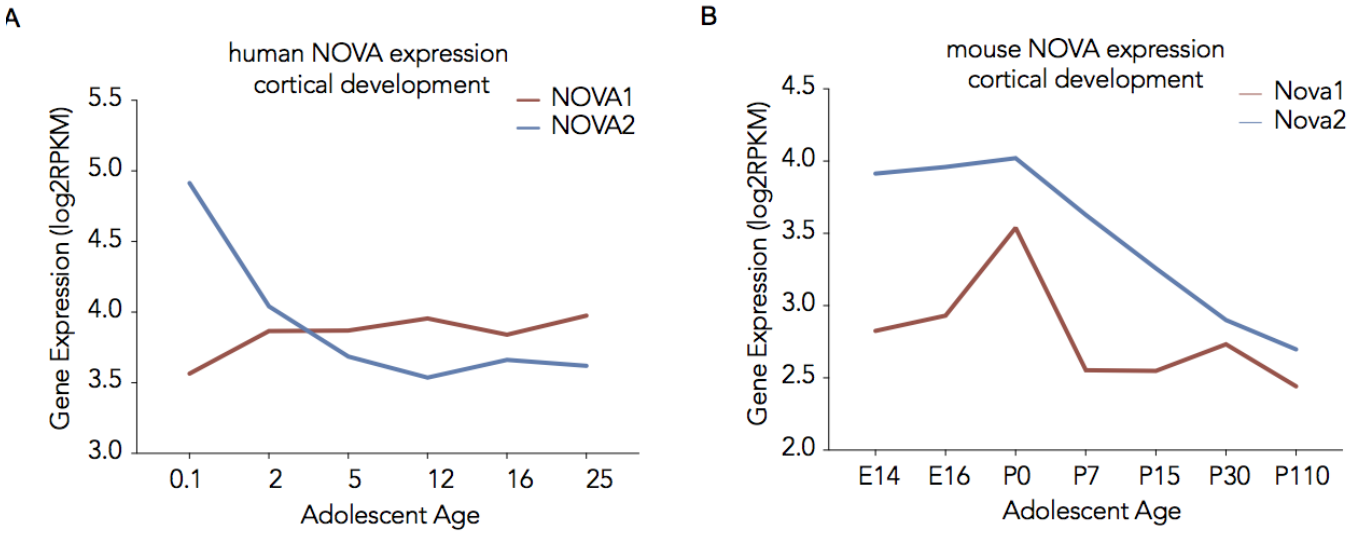


Figure 5-3: (A) Nova1 (red) and Nova2 (blue) RNA steady state levels during human cortical development. X-axis indicates developmental time points and Y-axis is log2-transformed RPKM values from RNA-sequencing. **(B)** Nova1 (red) and Nova2 (blue) RNA steady state levels during mouse cortical development. X-axis indicates developmental time points and Y-axis is log2-transformed RPKM values from RNA-sequencing.

Figure 5-4 Assessing changes in Nova expression and RNA steady state levels in stem cells and differentiated neurons

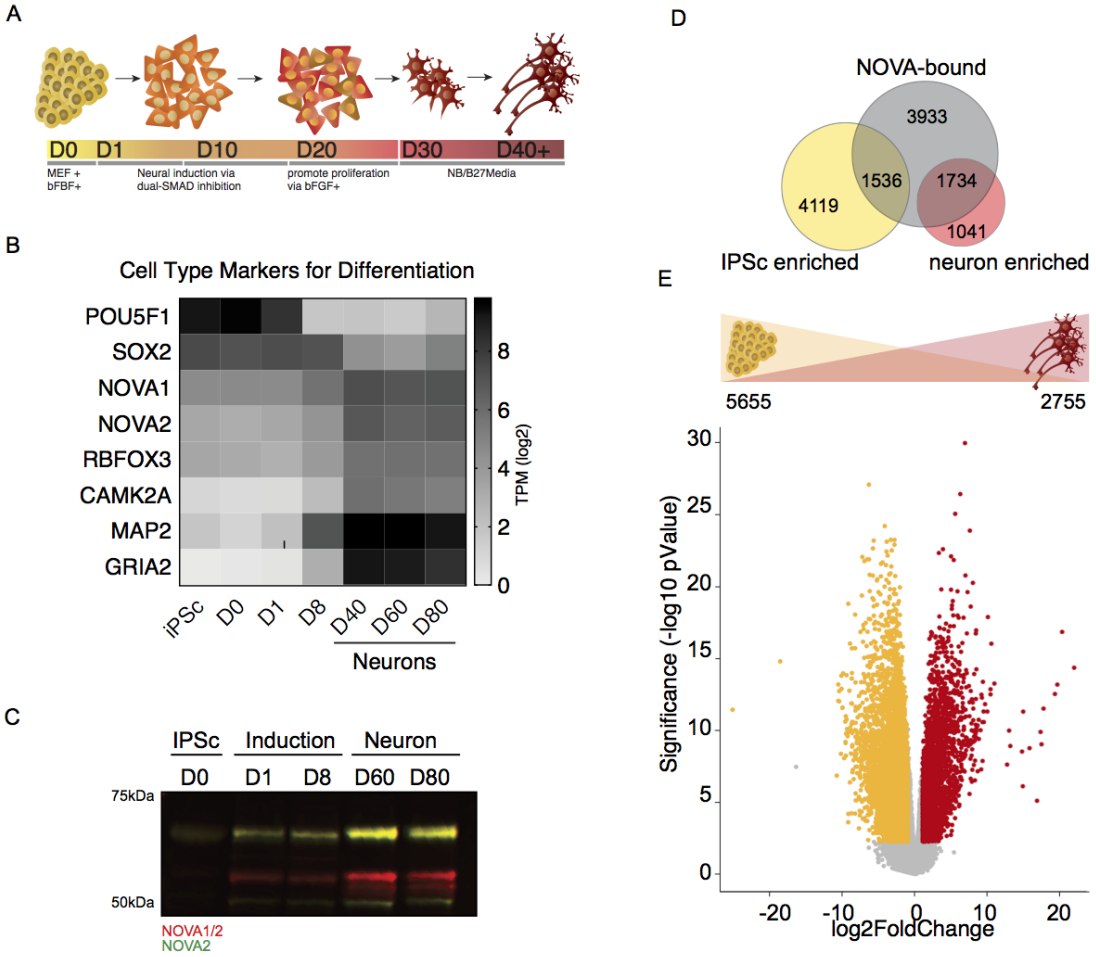


Figure 5-4: (A) Schematic illustration of methodology and time to differentiation of human IPScs into cortical neurons *in vitro*. **(B)** Heatmap of normalized expression values for markers of differentiation across time-course. **(C)** Western blot characterizing temporal expression of Nova protein during human neuron differentiation (20ug protein/well). **(D)** Venneuler plot. Nova binds stem cell enriched targets as well as neuron enriched targets in human neurons. **(E)** Volcano plot of differential analysis to assess changes in RNA steady state levels in IPS-cells compared to differentiated neurons. Quantification calculated using DESeq2 software (padj < 0.05 & abs(log2FC) > 1).

Intersection of differentially expressed RNAs and human brain derived Nova targets elucidated that in addition to neuronal targets, Nova also interacts with stem-cell enriched RNAs. Furthermore, analysis of differential transcript usage across neuronal differentiation revealed a significant switch in Nova1 transcript usage in stem cells compared to differentiated neurons (Figure 5-5). The Nova1 transcripts that are prominent and more highly utilized during a stem-cell state likely alter the functionality of Nova1 as an RNA binding protein. The last exon of Nova1 encodes for two out of three KH domains and the first exon encodes the nuclear localization signal (NLS). Taken together, these data emphasize the robust changes in RNA metabolism endured during differentiation and reveal that Nova1 is also temporally regulated at the RNA-level during these complex transitions.

Figure 5-5 Differential Nova1 transcript usage during neuron differentiation

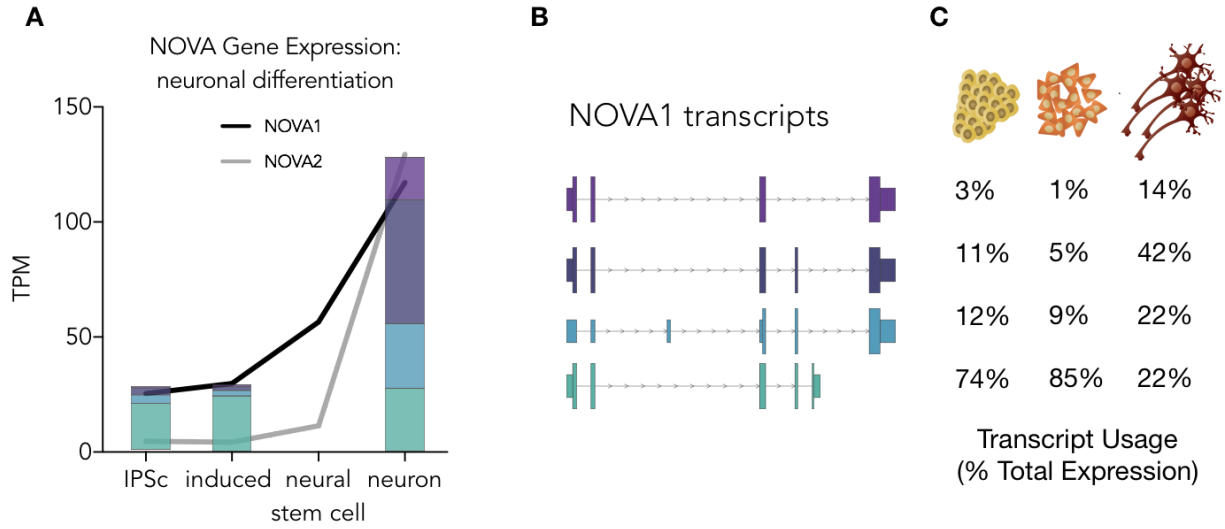


Figure 5-5: (A) Nova1 expression levels during differentiation time course, block colors (green, blue, navy, purple) correspond to percent of total expression per Nova1 transcript. Black line refers to Nova1 expression, grey indicates Nova2 expression. **(B)** Schematic drawing of differential transcripts expressed during neuron differentiation, transcript colors correspond to plot in Fig 5-5A. **(C)** Per transcript usage calculated as percent of total Nova1 expression for time points: (1) iPSC, (2) induced stem cell, and (3) differentiated neuron.

Associating changes in Nova1 levels with global changes in Nova target network

In order to identify groups of mRNAs that are coordinated in expression levels during human neuron development, all differentially expressed mRNAs were grouped by normalized expression levels at each time point accordingly. First, K means clustering method was used to determine the optimal number of clusters (Figure 5-6). Hierarchical clustering analysis grouped RNAs according to like expression patterns during development. Cluster 1 RNAs are most upregulated during the "induction" phase. Cluster 2 expression is steady and decreases post-induction, i.e. these are RNAs that are lowly expressed in neurons. Cluster 3 RNA levels are downregulated at the beginning of induction. Most notably, cluster 4 includes RNAs that are upregulated post-induction, i.e. RNA's that are upregulated in the neuronal lineage (Figure 5-6). Correlation analysis of Nova1 association per cluster showed a significant positive correlation between Nova1 and cluster 4, whereas a negative correlation was observed between Nova1 and cluster 2 (Figure 5-7A). Furthermore, Nova targets were found to be significantly enriched in cluster 4 accordingly (Figure 5-7B). Our findings are concordant with Nova regulating neuronal RNAs.

Figure 5-6 Clustering analysis to identify co-expression networks during human neuron differentiation

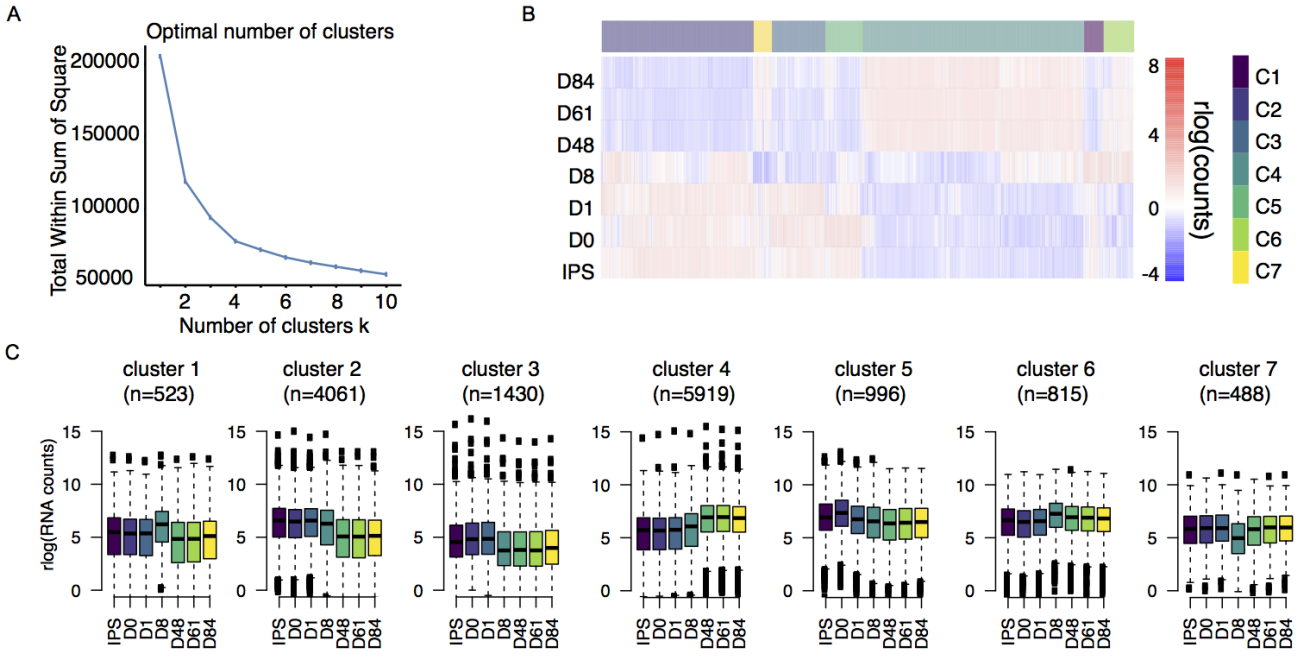
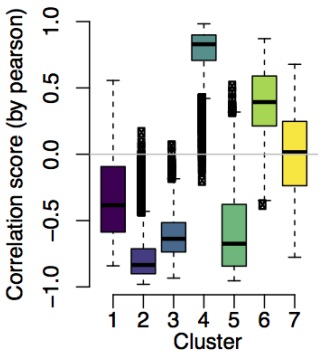


Figure 5-6: (A) Optimal number of clusters was determined by k-means method and revealed 4-8 clusters were optimal for gene expression changes during neuronal differentiation time course dataset. **(B)** Heatmap of normalized expression values per cluster over time. All genes grouped by cluster are represented here as rlog(Counts) are centred across each gene. **(C)** Normalized counts for clusters 1-7 at profiled time points. Box plots of clusters 1-7 displaying normalized RNA counts across all samples at select time points across differentiation of human cortical neurons (n = total number of genes per cluster).

Figure 5-7 High confidence Nova targets are significantly enriched in cluster 4

A



B

cluster	enrichment	p.value (fisher)
1	0.42428	0.000202572
2	0.24771	1.78E-59
3	0.12414	8.20E-28
4	1.93976	1.03E-127
5	0.13367	7.91E-19
6	0.99831	1
7	0.81847	0.356610248

C

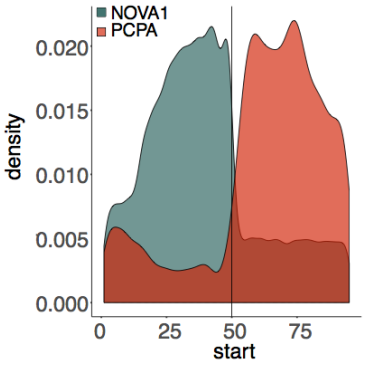


Figure 5-7: (A) Nova1 levels most highly correlate with clusters 1 and 4. Correlation analysis of all expressed genes relative to Nova1 expression levels over time. Nova1 expression robustly correlated with expression pattern of RNAs belonging to cluster 1 and cluster 4. Cluster 2 and cluster 3 are highly negatively correlated with Nova1 expression levels over time. Pearson correlation coefficient was calculated to test association of differentially expressed genes with Nova1 expression and plotted based on previously assigned clusters. **(B)** Enrichment analysis of high confidence Nova targets are significantly enriched in cluster 4. **(C)** High affinity Nova 6mer is highly enriched within 50nt of annotated premature polyadenylation signals.

Using IPSc model to investigate Nova biology

Human Nova-RNA interactions are enriched within 200nt of annotated human polyadenylation sites (Figure 5-7C). Given what has been shown in the mouse regarding Nova's relationship with polyadenylation and in order to leverage the neuronal time course data to demonstrate it's utility, we asked whether mRNA steady state levels of Nova targets were bound within 200nt of a polyadenylation site were temporally regulated during neuron development. As controls for the analysis we also tested the following groups for temporal changes in RNA steady state levels: (1) All expressed RNAs as background; (2) RNAs without Nova binding; (3) RNAs that have Nova binding but not at polyadenylation site. Given the differential distribution of Nova binding around polyadenylation sites we reasoned sites in different genic locations may

exhibit various relationships with RNA steady state levels. Therefore, polyadenylation sites were subset accordingly for the analysis. RNAs with only a single annotated polyadenylation signal were referred to as “single (S)”, RNAs with multiple polyadenylation signals were grouped as follows: “first (F)”, “middle (M)”, or “last (S)”.

Figure 5-8 in vivo Nova-RNA interactions in the human brain are enriched at annotated polyadenylation sites (PAS)

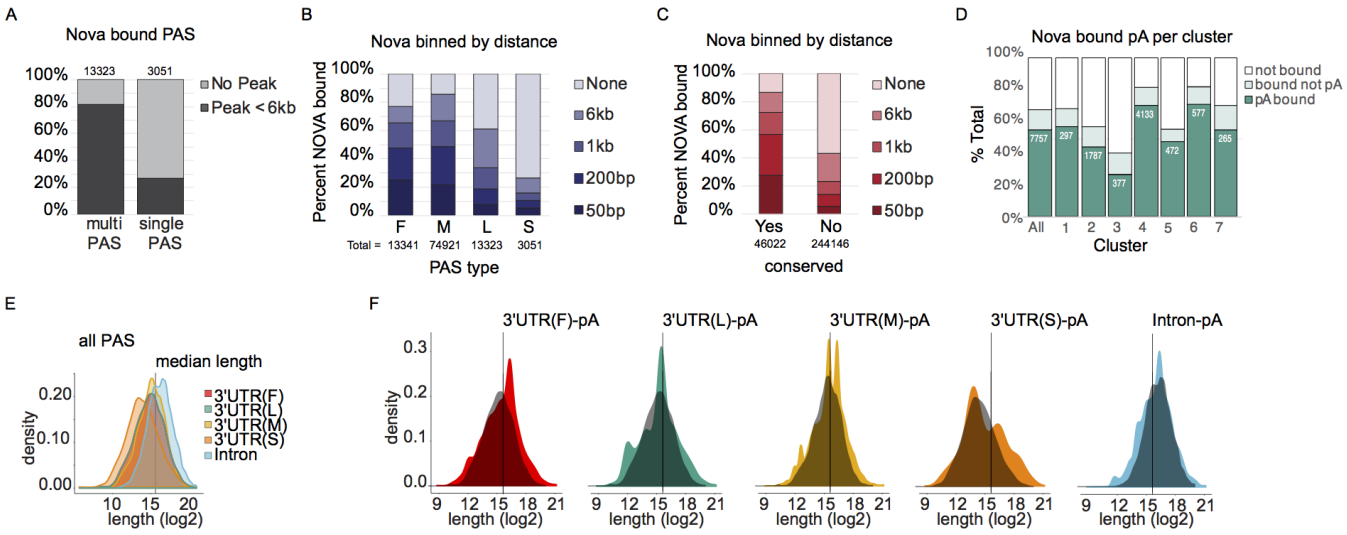


Figure 5-8: (A) Roughly 80% of all RNAs with multiple PAS are also bound by Nova in the human brain. ~20% of all single PAS RNAs are bound by Nova in the human brain. **(B)** Nova binding is dispersed across all types of alternative PAS, with middle and first PAS types the most (see Figure 1-2 for details on PAS type). **(C)** Majority of conserved polyA sites are Nova-bound (>80% of all annotated) and ~40% human-specific (non-conserved) polyA sites are bound by Nova within 6kb. **(D)** Nova binds polyadenylation sites *in vivo* of targets changing in expression during differentiation. Nova targets bound at polyadenylation site (green), bound but not overlapping with polyadenylation site (light green), and unbound non-targets (white) per co-expression clusters. Data is expressed as percent of total genes per cluster. **(E)** Distribution of total genic length for all annotated PAS sites in 3’UTRs (F, M, L, and S) and introns. **(F)** Gene lengths of all NOVA bound polyA sites, line indicates median gene length (log2) for all NOVA bound genes. PAS types subset by type, grey density = all annotated PAS for that type, colored density = all Nova bound PAS for the specified type in order to assess changes in length distributions.

To assess overlap between previously defined co-expression clusters and polyadenylation sites, a simple overlap between polyadenylation gene sets and clusters was calculated. One caveat of this approach was that some groups may appear enriched purely due to the size of the overlapping group. To remedy this, a Chi-squared test was used to test for deviation from the expected distribution given the number of genes in each of the genesets. Residuals help to assess how much overlap contributes to any differences observed from the expected. Chi-square residuals revealed a differential pattern across clusters (Figure 5-9A). Cluster 4, which is the cluster that NOVA1 and an overwhelming majority of its targets belong to indicates that Nova binding polyadenylation sites is associated with increase in RNA steady state levels and the absence of Nova is associated with a decrease in RNA levels. Whereas all other clusters demonstrated the inverse relationship. The transformed residuals were plotted in order to visualize how much of the significance observed is provided by any given overlap (Figure 5-9B). Interestingly, the relationships between the polyadenylation sites and the absence of Nova binding were the associations identified as most significant associations. In other words, the absence of Nova binding to alternative polyadenylation sites in cluster 4 is significantly associated with decreases in RNA steady state levels. Whereas, the absence of Nova binding at alternative polyadenylation sites for cluster 2 RNAs leads to a significant increase in RNA steady state levels. In order to visualize Nova binding relative to PAS binding densities were plotted for each PAS type accordingly for the top 3 most significant clusters (Figure 5-9C-E). Furthermore, the absence of Nova binding to single PAS targets is associated with a significant increase in RNA steady state levels. All other cluster group associations were unremarkable in significance. Here we have demonstrated the utility of intersecting *in vivo* human binding maps with *in vitro*

experimental systems to model relationships between global expression profiles and Nova binding events to further characterize human Nova biology.

Figure 5-9 Testing association between Nova binding alternative PAS and global expression changes in clusters

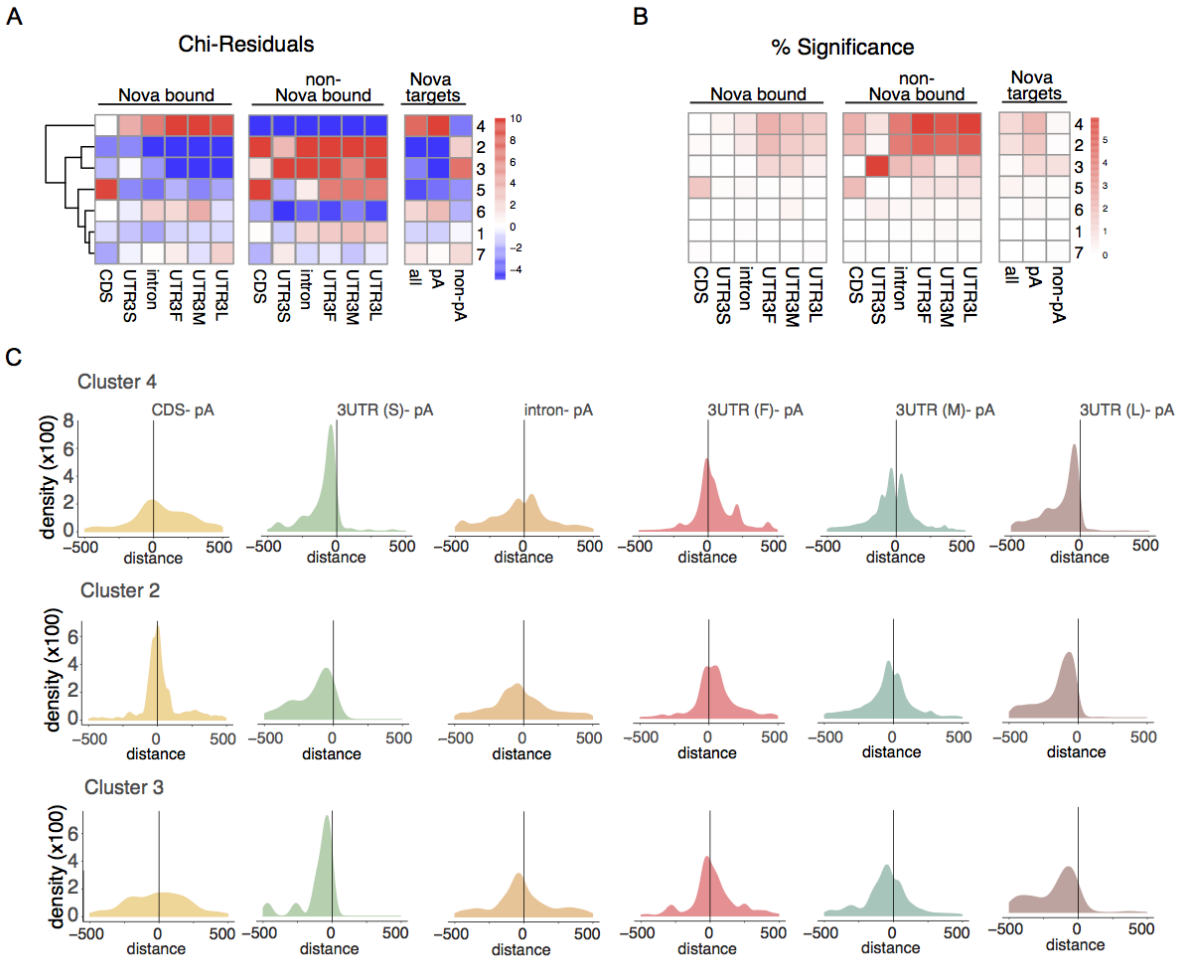


Figure 5-9: (A) Heatmap depicting results from Chi-square test of groups deviating from the expected distribution, given the number of genes in each gene set. Chi residuals are plotted which indicates how much each overlap contributes to any differences observed from the expected. **(B)** Transformation of Chi residual to show how much of the significance we observe is provided by a given overlap between polyA groups and clusters. **(C)** Cluster 4 Nova target density maps of *in vivo* Nova-RNA interaction sites in the human brain within 500nt of an annotated PAS. Plots are subset according to the genic location of PAS. **(D)** Cluster 2 Nova target density maps of *in vivo* Nova-RNA interaction sites in the human brain within 500nt of an annotated PAS. Plots are subset according to the genic location of PAS. **(E)** Cluster 3 Nova target density maps of *in vivo* Nova-RNA interaction sites in the human brain within 500nt of an annotated PAS. Plots are subset according to the genic location of PAS.

Discussion

Human iPS-cells are a powerful model to study the functional effect of human-specific Nova regulation. Pathway analysis demonstrates that network level Nova regulation is recapitulated in hIPS-derived neuron cultures. Additionally, the model system at hand permits for neuronal specificity which allows one to bypass many technical issues that arise when studying molecular changes in bulk tissues such as changes being driven or masked by contamination from neighboring cell-types. However, the ability to study interactions and relationships between neurons and neighboring cells types is not attainable in the model system. hIPS-derived neuronal models enable for precise control over time, which allows for experimental modeling to answer questions relating to human-specific neuron development. Unfortunately, investigations of migration are not well addressed by this model. Technical challenges include highly complex and difficult differentiation protocols for specialized neuron subtypes, such as dopaminergic neurons and motor neurons to name a few.

Despite challenges in the model system, our efforts showed that Nova1 undergoes alternative RNA processing resulting in a switch in transcript usage following commitment to neuronal lineage. We hypothesize that the observed switch may be a mechanism to modulate Nova1 function through sequestration during critical transition periods. We also highlight a robust number of Nova targets within stem-cell enriched populations suggesting that Nova is important for RNA regulation in human stem cells.

It is clear that human brain data is a heterogeneous mix of diverse neuronal cell types at various developmental stages. Here we demonstrate the utility of using IPS-derived neurons to

identify co-expression networks over time as a way to deconvolute bulk tissue HITS-CLIP data. We are able to delineate associations between Nova-bound poly(A) sites and global RNA steady state levels over time to connect changes in RNA steady state levels with lack of Nova bound polyadenylation sites. Here we have taken an integrative approach to combine *in vivo* human brain RNA binding maps and models of temporal expression using relevant experimental systems to characterize human-specific Nova biology.

Chapter 6: Additional systems for validating human brain HITS-CLIP peaks

Introduction

In addition, an intended approach was to utilize cell-lines that endogenously expressed Nova for our functional studies. Unfortunately, each model proved uniquely complicated with unexpected caveats for our intended purposes. Few human cell-lines endogenously express Nova, however we identified two and characterized them for both Nova expression and binding profiles: (1) SH-SY5Y: Human neuroblastoma cell line and (2) LUHMES: The Lund human mesencephalic cell line (Lotharius et al., 2002, 2005; Ross et al., 1983). Nova HITS-CLIP coupled with RNA-sequencing was carried out for both cell systems and the following sections aim to describe the data, which may provide a useful resource for future researchers and experimental design.

The neuroblastoma SH-SY5Y cell line is a subclone of the SK-N-SH cell line, which was derived from a bone marrow biopsy of a metastatic neuroblastoma from a 4 year-old female patient. Therefore, the cell line exhibits multiple genetic aberrations due to the cancerous origin. Nonetheless its use has been very popular in the neuroscience field due to neuronal-like characteristics and the ability to differentiation capabilities. Specifically, they have been well utilized in the context of Parkinson's Disease research due to dopaminergic properties. SH-SY5Y cells are one of only a few human cell lines that express both Nova1 and Nova2 proteins, which made them a particularly attractive model for our studies.

LUHMES cells were derived from a human female 8 week old mesencephalon. They are highly proliferative and therefore are advantageous for studies requiring large scale rapid expansion of cells. Furthermore, they have the ability to be differentiated into dopaminergic-like neurons and are a particularly popular model for studying neurodegenerative disease. The property that was most attractive for our purpose of studying Nova in LUHMES was the robust and exclusive expression of Nova1 protein, and absence of Nova2 protein. Careful characterization and establishment of human relevant experimental models is important to consider as unknown limitations of a model system can significantly influence experimental design and the interpretation of any results.

Results

The human neuroblastoma SH-SY5Y cell-line

Nova protein expression was validated in SH-SY5Y cells by western blot analysis in triplicates (Figure 6-1A). Following validation, Nova HITS-CLIP was carried out using identical conditions to human brain HITS-CLIP. One property of SH-SY5Y HITS CLIP that was troublesome was the efficiency of the polyclonal antibody utilized to generate the human brain Nova-CLIP map was not as efficient in purifying Nova-RNA interactions from SH-SY5Y cells. Experimental comparison of the polyclonal Nova antibody compared to a monoclonal Nova1 (Abcam) antibody showed a much more efficient pull-down with monoclonal Nova1 antibody (data not shown). This was surprising as the most highly expressed Nova protein in SH-SY5Y cells is Nova2. However, because the goal was to overlay the SH-SY5Y CLIP with the human brain Nova CLIP, we chose to move forward using the polyclonal Nova antibody. Initial experiments revealed significant autorad signal in the non-targeted control IgG IP in cross-linked cells. A parallel non-crosslink control sample demonstrated that the observed signal was cross-linking dependent. We observed no detectable signal from non-crosslinked Nova IP control (Figure 6-1B). In order to increase autorad signal, the number of cells utilized for IP was increased and all negative controls were carried out for each replicate CLIP experiment, library preparation and sequencing was performed in parallel. Non-specific unique Nova tags were bioinformatically filtered from the dataset. Final CLIP libraries yielded roughly 300,000 unique tags (n = 4) culminating in 5,082 unique binding sites.

Figure 6-1 Characterizing Nova and binding profile in SH-SY5Y cell line

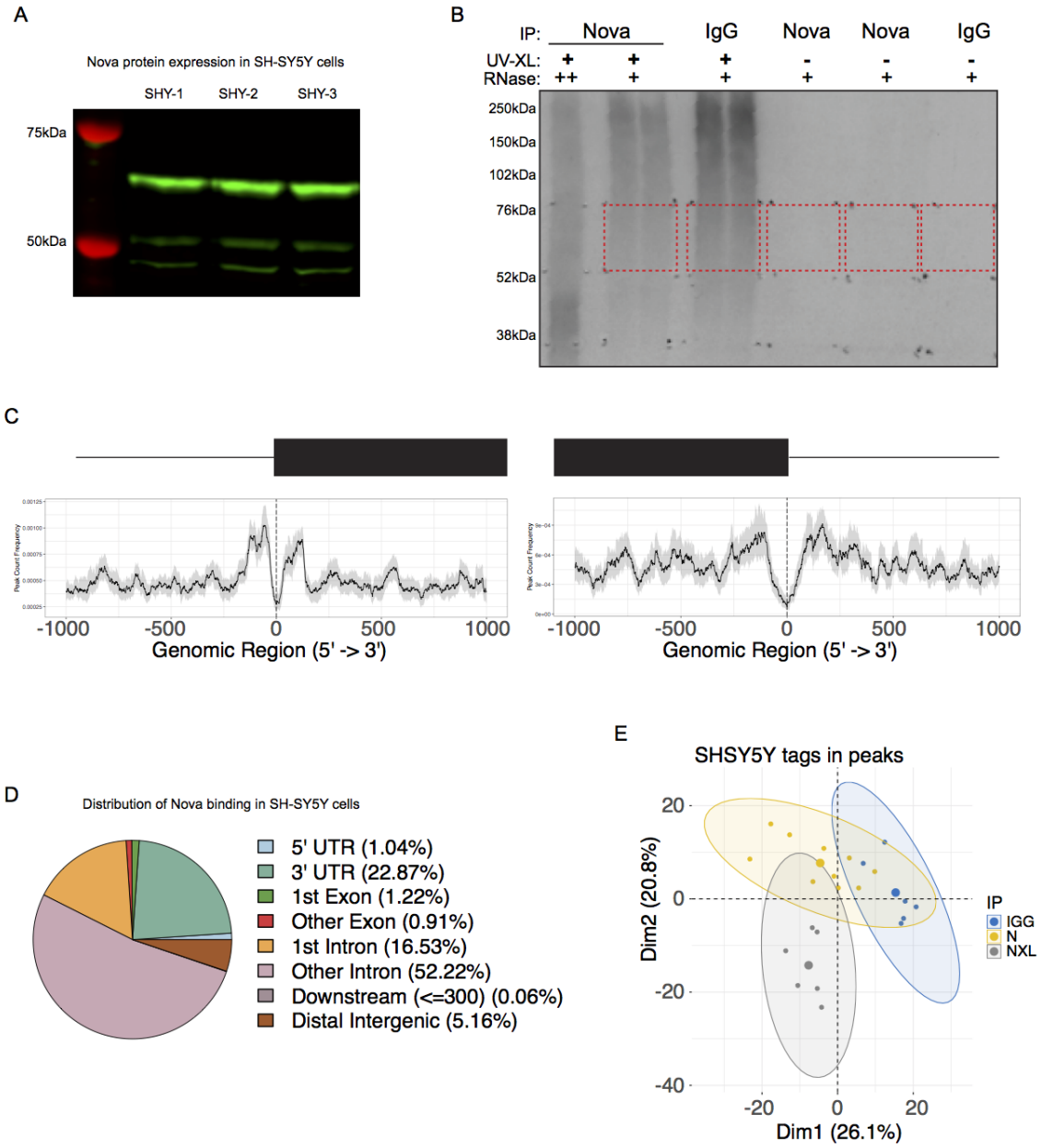


Figure 6-1: (A) Western blot analysis Nova protein in SH-SY5Y cell lysates, rbNova 1:1000 (n = 3). **(B)** Nova HITS-CLIP autorad. Region excised from autorad for purification and library preparation is indicated in red. Exposure time = 16 hours. Autorad is representative of 4 separate HITS-CLIP experiments. **(C)** Principal Component Analysis to assess variance between NOVA HITS-CLIP experiments vs. controls. **(D)** Genomic Distribution. Pie chart displaying genomic distribution of significant Nova-RNA interactions in SH-SY5Y cells. **(E)** Average normalized tag distribution across introns and exons. X-axis represents distance from exon start (left), or distance from intron start (right). Start site indicated by dashed line.

Average Nova binding frequency at intron-exon junctions reveals a moderate enrichment in Nova binding within 250nt of junction site (Figure 6-1C, left). Downstream intronic enrichment in average binding frequency showed a much less robust Nova binding signal (Figure 6-1C, right). Genomic distribution of Nova CLIP peaks showed ~68% of peaks were localized to introns and ~23% of peaks were localized to 3'UTRs (Figure 6-1D). Genomic distributions reflective of binding patterns typically observed in mouse Nova CLIP compared to human Nova CLIP. Variance across experimental and control datasets was assessed via PCA (prior to read filtering) and while libraries formed distinct clusters according to experimental (Nova) vs. control (NXL and IgG) IP conditions. The confidence intervals were not convincing (Figure 6-1E). Further demonstrating the utility of having the negative controls per CLIP experiment for downstream bioinformatic assessment and filtering purposes.

One characteristic of Nova alternative splicing regulation is that Nova binds the exon resulting in more exon silencing, or skipping. Many genomic regions can overlap depending upon the transcript considered, therefore we utilized an upset plot as a way to resolve and visualize these binding sites (Figure 6-2A). This approach revealed ~400 binding sites in regions that are annotated as both introns and exons, depending upon context. In the context of Nova, one might hypothesize these are "silenced" Nova exons. Additionally, over 700 binding sites are both exons and 3'UTRs, which may relate to Nova-dependent alternative RNA regulation. Pathway analysis was performed on all Nova bound targets in order to globally characterize Nova regulation in SH-SY5Y cells and highlight reactome pathways that could be studied using this model in the context of Nova (Figure 6-2B). The most significant and most enriched pathway highlighted was Signaling Receptor Tyrosine Kinases.

Figure 6-2 Global patterns observed in SH-SY5Y Nova HITS-CLIP data

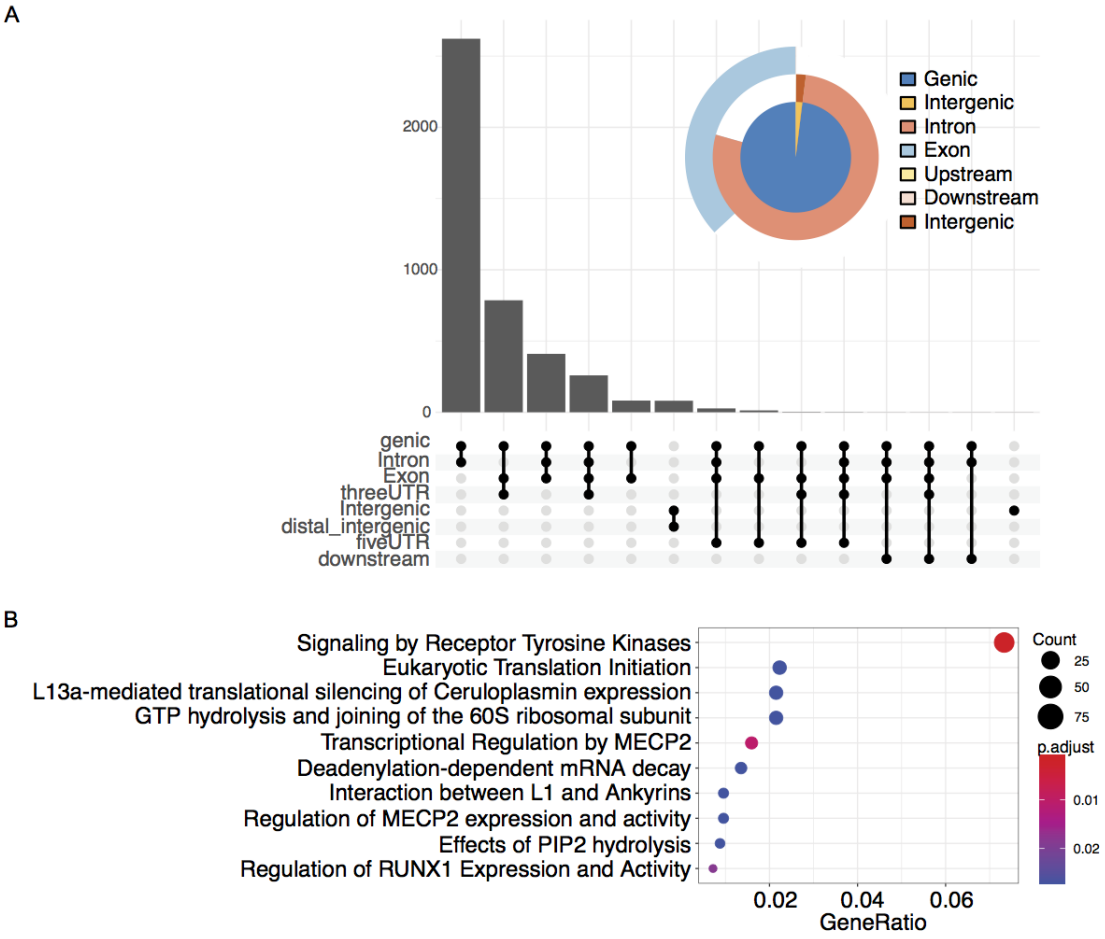


Figure 6-2: (A) Upset plot demonstrating the annotation overlap of Nova binding sites. Conceptual overlap visualized in venn pie diagram (top right). **(B)** Reactome Pathway Analysis of SH-SY5Y Nova targets.

The Lund human mesencephalic LUHMES cell-line

Due to exclusive expression of Nova1 in LUHMES cells, HITS-CLIP a monoclonal Nova1 specific antibody was used to increase efficiency of HITS-CLIP experiments. Global distribution plots reveal that average frequency of Nova1 binding across introns and exons is not as robust in LUHMES HITS-CLIP as previous HITS-CLIP data sets (Figure 6-3A). Nova1 HITS-CLIP in LUHMES yielded a total of 14,513 unique peaks (n=2). Roughly 35% of significant peaks were localized to introns, and 48% of peaks localized to 3'UTRs (Figure 6-3B). Over 4,500 peaks are in regions that can be annotated as either "exons" or "3'UTR", and about 1,500 peaks that are annotated as "exons", "3'UTR", or "intron" (Figure 6-3C). The discrepancy in annotation and overlapping features could be interpreted as Nova binding in regions of RNA that are highly alternatively regulated. Global reactome pathway analysis was performed for all LUHMES HITS-CLIP targets (Figure 6-3D).

Figure 6-3 Nova HITS-CLIP genomic distribution in LUHMES cells

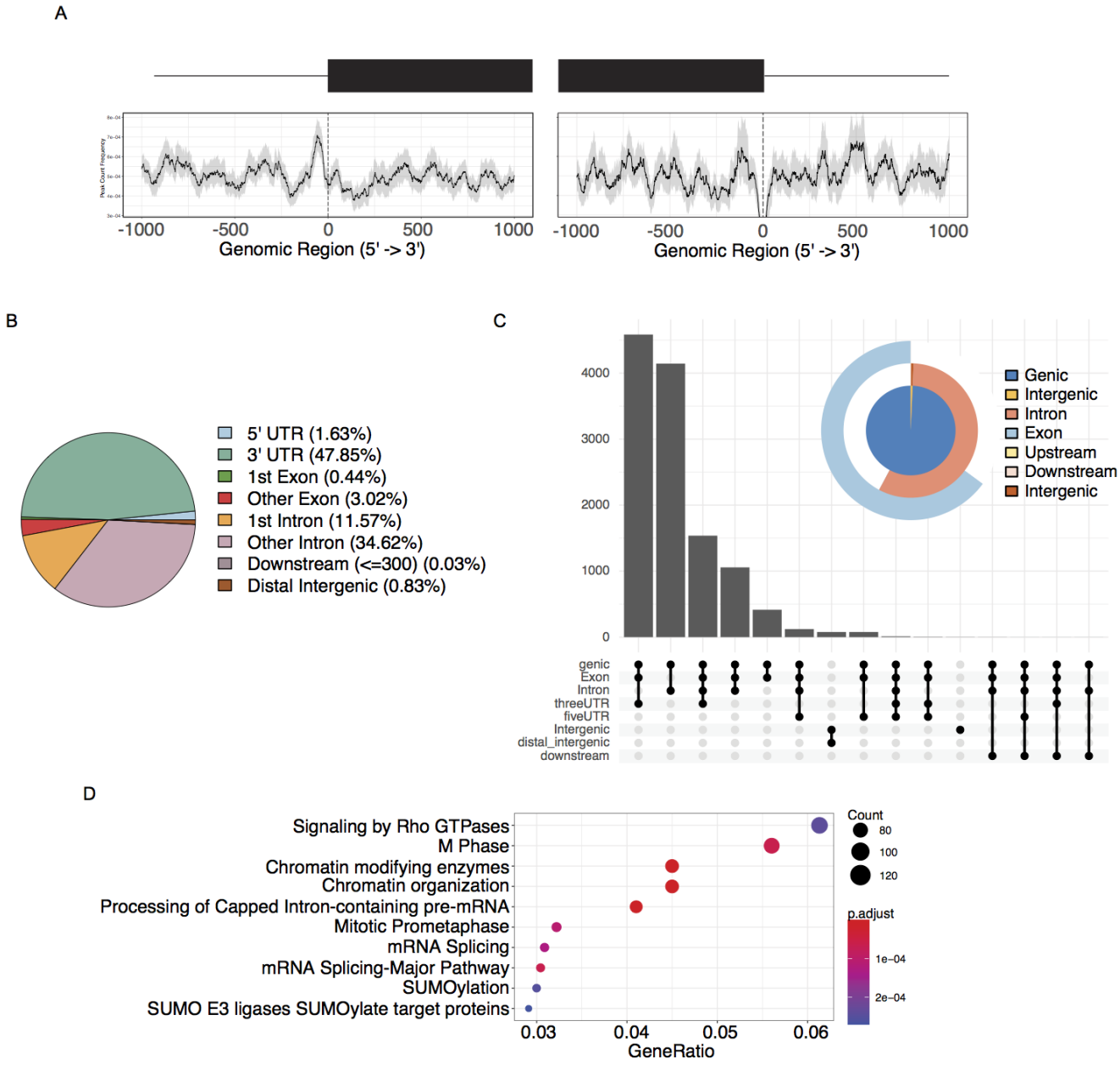


Figure 6-3: (A) Genomic distribution of significant Nova-RNA interactions in LUHMES cells. **(B)** Average normalized tag distribution across introns and exon junctions for Nova CLIP in LUHMES cells. **(C)** Upset plot demonstrating annotation overlap association within Nova binding sites in LUHMES cells. Conceptual overlap visualized in venn pie diagram (top right). **(D)** Reactome Pathway Analysis of LUHMES Nova targets.

Comparative pathway analysis across Nova HITS-CLIP datasets was performed in order to assess global similarities and caveats within the various models that were screened via Nova HITS-CLIP. Comparing Nova regulated pathways and processes across experimental models allows for the identification of biological functions/processes where Nova regulation is likely recapitulated and thus may be a suitable model for functional experimentation. Kegg Pathway analysis was performed in order to compare Nova binding profiles (1) *in vitro* vs. *in vivo* (2) cancer vs. neuronal (3) human vs. mouse (Figure 6-4). SH-SY5Y cells were only able to significantly model *in vivo* Nova regulation in the context of focal adhesion processes. Whereas Nova targets in LUHMES cells were abundant for targets important in functions that are also regulated by Nova in the brain (both human and mouse comparatively). These functionally recapitulated processes include, Insulin signaling pathway, Thyroid hormone signaling pathway, Neurotrophin signaling pathway, cGMP-PKG signaling pathway, and longevity regulating pathway to name a few. Altogether, the inability to model Nova's regulatory network *in vitro* is striking and displays the unique context-dependent nature of Nova function in specialized neuronal networks and functions.

Figure 6-4 Comparative analysis of enriched pathways across HITS-CLIP defined Nova networks

Figure 6-4: Comparative enrichment analysis Nova targets associated with Kegg Pathways. Each target network is individually experimentally defined using Nova HITS-CLIP methods.

In order to determine potential models for the study of disease-relevant pathologies, we also performed comparative enrichment analysis against disease ontology (DO) terms (Figure 6-5). Disease Ontology enrichment indicates that SH-SY5Y cells and LUHMES cells are poor disease-relevant models to study changes in Nova regulation as there were no terms enriched for either dataset. However, autism spectrum disorder and epilepsy syndrome were both identified as significantly enriched within the IPS-derived cortical neuron Nova target network and therefore may provide a relevant system to extrapolate Nova dysregulation with disease using a physiologically relevant model system. Interestingly, none of the human-derived HITS-CLIP data showed enrichment in motor neuron disease, which was enriched in mouse derived HITS-CLIP data. One explanation for this may be due to the human brain regions profiled. Motor neurons are a highly specialized neuronal subtype and therefore we hypothesize that Nova regulation may need to be independently defined within the population. Taken together, our extensive Nova HITS-CLIP binding maps have further emphasized the importance and utility of defining RNA binding protein networks across diverse and physiologically relevant systems.

Figure 6-5 Comparative analysis of enriched pathways across HITS-CLIP defined Nova networks

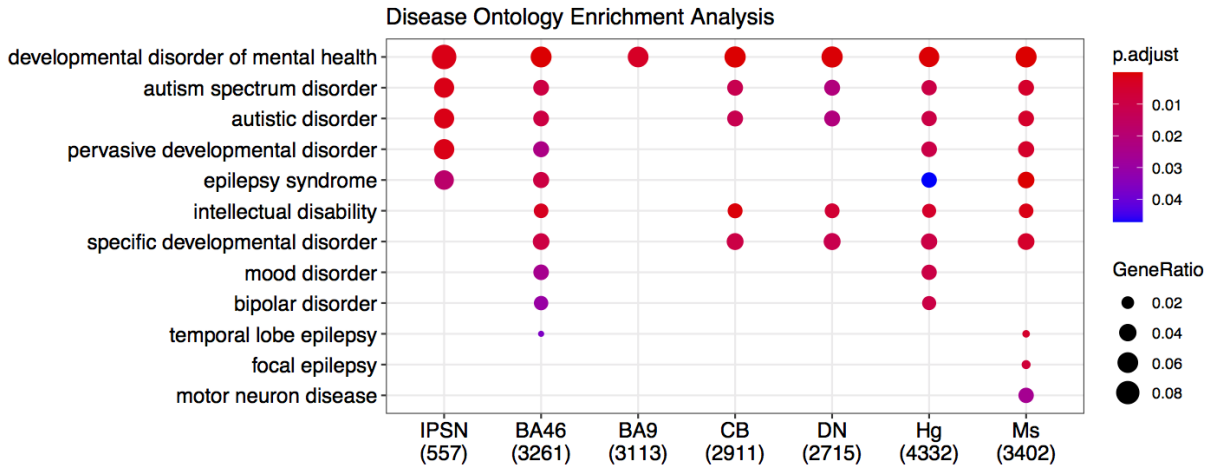


Figure 6-5: Comparative enrichment analysis Nova targets associated with specific Disease Ontologies. Each target network is individually experimentally defined using Nova HITS-CLIP methods. There were no significantly enriched disease ontology terms for SH-SY5Y or LUHMES derived Nova targets.

Discussion

SH-SY5Y cells are a widely utilized cell line for neuroscience research and the generation of a comprehensive Nova HITS-CLIP binding map will be a valuable resource for researchers when choosing an experimental model or design. There were several Nova targets that overlapped with splice changes in HeLa cells in addition to functional processes regulated by Nova in the human brain, such as Focal adhesion. SH-SY5Y cells performed poorly in their ability to recapitulate Nova regulatory networks. One rationale for SH-SY5Y Nova targets to be enriched in focal adhesion processes may be attributed to the unique morphology of SH-SY5Y cells as they are sometimes described as "neuronal" in structure. Nova regulation of the cytoskeleton is important for neuronal viability, therefore Nova regulation over cytoskeletal RNAs may likely be an important responsibility for Nova proteins in SH-SY5Y cells. An intriguing direction for future researchers may include assaying for detectable changes in Nova's regulatory network upon SH-SY5Y "differentiation" as they have the potential to expand the applicability of SH-SY5Y models in neuroscience research and in the context of Nova research.

LUHMES cells were enriched in far more biological pathways/functions that were uniquely enriched in the human brain CLIP and not enriched in the mouse brain CLIP including: FoxO signaling pathway, C-type lectin receptor signaling pathway, and Chronic myeloid leukemia. These findings hint towards the possibility that human-specific Nova regulation may be active in stem-like cells, compared to mouse studies where Nova regulation has been specialized within mature neurons. Nova regulation of FoxO signaling pathway is of exceptional interest as recent studies have shown that FOXO1 expression profiles in the brain have significantly and robustly diverged from mouse to humans (Pembroke et al., 2021).

Physiological processes that have been associated with FoxO signaling including apoptosis, cell-cycle control, glucose metabolism, and longevity. Therefore LUHMES cells may be an interesting human-experimental model for investigating the physiological relevance of Nova1-specific regulation of RNA processing within the FoxO pathway.

Ultimately, our comparative analysis across human model systems emphasizes that the ideal approach for investigating Nova-regulated biological processes is an *in vivo* approach and that the mouse model is superior to any *in vitro* model tested in this study, including IPS-derived neurons. Although it is important to note that possible explanations to account for poor correlation between IPS-derived neurons and the human brain are the (1) inability to model a 3D network in a 2D culture system and (2) neuronal age. The human brains that were used to generate our human binding map were adult brains, and at the IPSc-neuron time point (Day 80) that was utilized for HITS-CLIP, the neurons are still rather immature comparatively. Neuronal maturity is an important factor to consider when studying Nova regulation. Nova expression and regulation of RNA processing is quite diverse during development and aging processes and because IPSc-derived neurons are terminally differentiated and slow, they may be better suited for studying altered Nova regulation in the context of disease pathologies.

Chapter 7: Human Brain CLIP & connections to Human Brain Disease

Introduction

NOVA was first discovered in the autoimmune sera from patients with paraneoplastic opsoclonus-myoclonus ataxia, termed POMA (Buckanovich et al., 1993; Darnell et al., 2011; Yang et al., 1998). POMA patients present with lung or gynecologic tumors accompanied with the development of excessive motor movements due to motor inhibition failure, with the observation that some POMA cases lead to encephalopathy (Darnell et al., 2011). More recently a 9 year-old patient with global developmental delay and relatively low IQ score was found to have a 2.3MB *de novo* deletion at 14Q12 resulting in haplo-insufficiency of *NOVA1*(unpublished data). Primary indication for concern began at the age of 2 years, and stemmed from an observed low muscle tone and delay in both gross and fine motor skills, which have persisted throughout development. Neuropsychological evaluation concluded the patient to be significantly disinhibited, displaying deficits associated with attentional shifting. The ability to shift attention lies at the cornerstone for conceptual and abstract thought. In order to develop a concept one must be able to consider the same information from a variety of different viewpoints, in the absence of this ability, thoughts and behaviors remain concrete. It is evident that Nova expressing cells in the human brain are important for proper excitation/inhibition control of circuitry imbalance of higher-order motor and cognitive function.

Gaining a deeper understanding of the role of NOVA regulated exons in disease provides targetable treatment options, such as the use of oligonucleotides. Taken together with the robust specificity of our recently generated *in vivo* Nova-RNA interaction map of the human brain, we have created a detailed resource providing mechanism insight for thousands of actionable, disease relevant exons. In recent years, research has emphasized the benefits of therapeutic approaches aimed at targeting expressed genes, or mRNA, as opposed to other modalities (Havens and Hastings, 2016). RNA based medicines have the potential to be applied as a broad-based therapeutic approach to multi-genic neurologic disorders. Treatment of both developmental and degenerative disease have benefited from RNA based modalities as there is a great deal of overlap between genes involved in intellectual disability and genes regulating cognitive decline. In fact, there are several therapeutics for the treatment of neurodegenerative disease that are performing extremely well in modulating exon inclusion/exclusion (Havens and Hastings, 2016).

The most renowned and encouraging example for successful application of antisense oligonucleotides for the treatment of the neurodegenerative disease Spinal muscular atrophy (SMA). SMA is a devastating disease that is characterized by muscle atrophy and weakness as a direct result of motor neuron degeneration in the spinal cord and brainstem due to inadequate levels of survival motor neuron protein (SMN). Due to mutation, SMA patients do not have a functional SMN1 gene and therefore have to rely on the highly homologous SMN2 gene for survival. However, the SMN2 gene has a single nucleotide change that converts a splice enhancer to a splice silencer (Wu et al., 2017). Thus, the single nucleotide difference results in skipping of exon 7 in the majority of transcripts, giving rise to a nonfunctional SMN protein

(Castro and Iannaccone, 2014). Groundbreaking research by Adrian Krainer and colleagues discovered the splicing mechanisms regulating exon 7 skipping and developed antisense oligonucleotide technologies to inhibit the exon 7 skipping and produce a functional SMN protein (Staropoli et al., 2015). This pivotal discovery resulted in the first FDA-approved antisense oligonucleotide treatment for SMA patients (Rigo et al., 2012).

More recently it was shown that with understanding of the pathogenicity of a single mutation, a customized SSO therapeutic can be effectively designed, tested, and delivered on an individualized basis for rapid treatment of genetic disease (Kim et al., 2019). Paving the way towards an era of personalized medicine for patients with rare causative genetic mutations.

It is vital to determine the molecular underpinnings of neurological disease in order to develop new technologies and approaches for treatment. It is clear that the ideal model system to understand neurologic disease resides in human tissue. Single cell based genomics has eluded that the spectrum of diversity lying between individual cell-types is immense and largely unknown (Johnson and Walsh, 2017). Single neuron resolution has further encouraged the idea that distinct cellular populations could dictate the etiologic outcomes of patients with seemingly identical diagnosis, while also highlighting phenotypically relevant molecular mechanisms across diseases (Zhang et al., 2016). The research here addresses an integral step towards utilizing patient-relevant molecular based discovery methods in order to develop novel treatment options with incredible transcript specificity. For the purpose of demonstrating utility of human brain specific Nova binding maps and connecting Nova-RNA interaction sites with

function, this section aims to validate the effect of precise *in vivo* binding sites on *in vitro* splice switches.

One of the most dramatic molecular effects of Nova1-dependent splicing events observed in HeLa cells was exon 16 of Glutamate ionotropic receptor kainate type subunit 2 (GRIK2). GRIK2 functions to mediate excitatory synaptic transmission in the brain with specific roles in modulating slow and long-lasting presynaptic neurotransmitter release. Numerous studies have shown genetic association between GRIK2 and autism (Dutta et al., 2007; Jamain et al., 2002; Shuang et al., 2004). In addition, aberration of GRIK2 has also been associated with Schizophrenia and mild mental retardation (Beneyto et al., 2007; De Haan et al., 2015; Lanore et al., 2012; Motazacker et al., 2007; Shibata et al., 2002). Last, GRIK2 has been studied in the context of the neurodegenerative disorder, Huntington's disease as a potential contributing factor to disease onset and progression (Adelmann et al., 1999; Lee et al., 2012). Given the role of GRIK2 in mediating synaptic transmission and excitatory network balance in the brain, coupled with its significant role in cognition and human neurologic disease, the ability to modulate Nova's effect may have therapeutic applications in the treatment of disease.

The second set of targets were selected based on relevance to human-specific disease. Amyloid Precursor Protein family consists of three distinct homologous genes: Amyloid Precursor Like Protein 1 (APLP1), Amyloid Precursor Like Protein 2 (APLP2), and Amyloid Precursor Protein (APP). APLP2 and APP are both broadly expressed across tissues, are highly homologous in sequence, and are conserved in domain structure. However, one notable difference between the two is that APLP2 does not produce an Aβ peptide like APP.

Furthermore, the APP family is highly evolutionarily conserved and it has been postulated that the emergence of the protein family coincides with the evolution of the nervous system and lipoproteins(van der Kant and Goldstein, 2015). Knockout mouse models have shown that ablation of a single member results in unique and measurable phenotypes, but not embryonic lethality. However, double knockout mouse studies have demonstrated that APLP2 offers an essential function for life that only the combination of APP and APLP1 can compensate for. Interestingly, APLP2 has the ability to compensate in the absence of both APP and APLP1 (van der Kant and Goldstein, 2015; Lazarov and Tesco, 2016). Aberration in the APP gene family gives rise to deficits in synapse formation, function, and plasticity(Ring et al., 2007). Extensive characterization of phenotypes related to brain size malformation, seizures, and learning and memory have illustrated the wide spectrum of functions that the APP family plays a role (Dawson et al., 1999; Seabrook et al., 1999; Smith et al., 2007, 2010; Zheng et al., 1995). It is increasingly clear that both APP and APLP2 expression patterns are highly context specific and are essential for healthy brain function and cognition. Therefore, teasing apart how APP and APLP2 expression is controlled is highly relevant to their role in neurons and disease. Furthermore, developing a method to precisely titer the levels of APP and/or APLP2 could have diverse and widespread clinical application.

Results

To further characterize the function of Nova1-RNA interactions identified via human brain CLIP and to examine the utility of generating human specific Nova binding maps, bioinformatic analysis was performed to identify human brain Nova peaks proximal to *in vitro* Nova1-mediated splicing events identified using our robust HeLa cell model system. Targets of exceptional interest were selected for further characterization of the specific function of Nova-RNA interaction sites relative to splicing events. CLIP peak data was used to design splice switch oligonucleotides (SSOs) specifically targeted to Nova binding sites to inhibit function at each respective interaction site. The functional outcome of Nova SSO- inhibition was assayed with semi-quantitative PCR methods (Figure 7-1).

Figure 7-1 Schematic of model for experimental approach

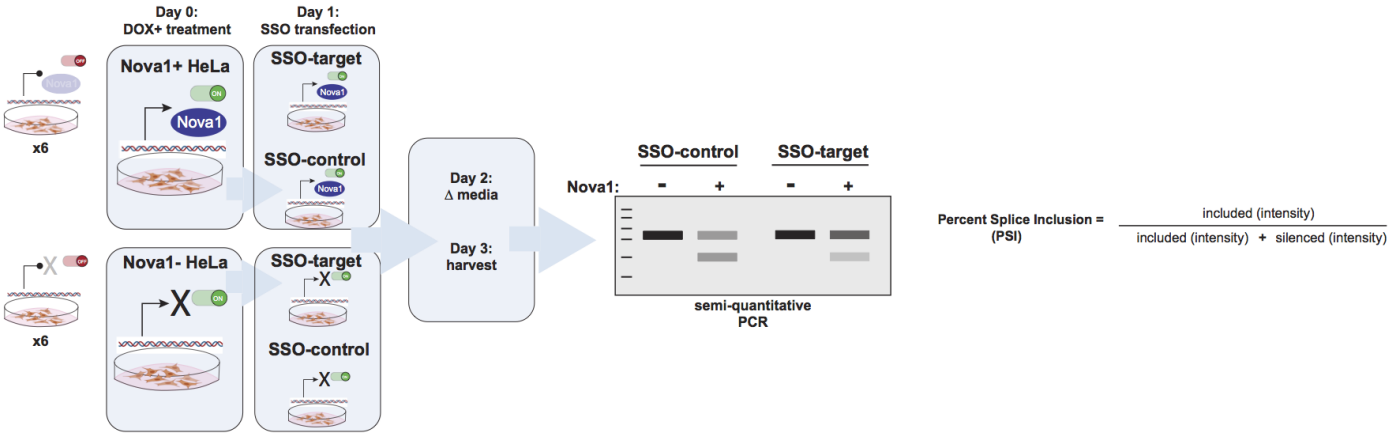


Figure 7-1: Nova1+ expressing cell-lines and Nova1- control cell-lines were plated in parallel 16 hours prior to doxycycline treatment (Day 0), after 24-hours cells were transfected with either target SSO or control SSO (Day 1). Media was changed 24-hours after transfection (Day 2) and Cells were harvested 48 hours post-transfection (Day 3). RNA was purified from each condition, PCR amplification, and run on agarose gel for size shift analysis and quantification, 3 replicates per transfection.

Glutamate ionotropic receptor kainate type subunit 2 (GRIK2)

Ectopic expression Nova1 expression in HeLa cells results in a significant switch in alternative splicing patterns of GRIK2 (Figure 7-2A). Human brain HITS- CLIP data identifies a mouse-human conserved Nova-RNA interaction site spanning the exon-intron junction of the alternatively spliced exon (Figure 7-2B). Our data demonstrates Nova1-dependent silencing of the target exon 16, which is concordant with previously described position specific Nova binding rules. Inclusion of the Nova-silenced exon 16 encodes a premature stop codon which would result in truncation of the GRIK2 cytoplasmic region, impacting GRIK2 protein function as well as protein-protein interactions in the absence of Nova mediated inhibition(Barbon et al., 2001). To demonstrate the utility of human brain specific Nova binding map and to inhibit Nova's function at GRIK2 exon 16, the HITS-CLIP identified peak was used as a template for SSO design (Figure 7-2C). Transfection of the SSO into HeLa cells significantly decreases exon 16 silencing, with dose dependent effectiveness (Figure 7-2D-E). These data support the notion that SSO-mediated inhibition of a human brain identified Nova binding site in HeLa cells, is a robust and practical first step towards SSO screening for therapeutic applications.

Figure 7-2 Nova1 promotes silencing of GRIK2 exon 17

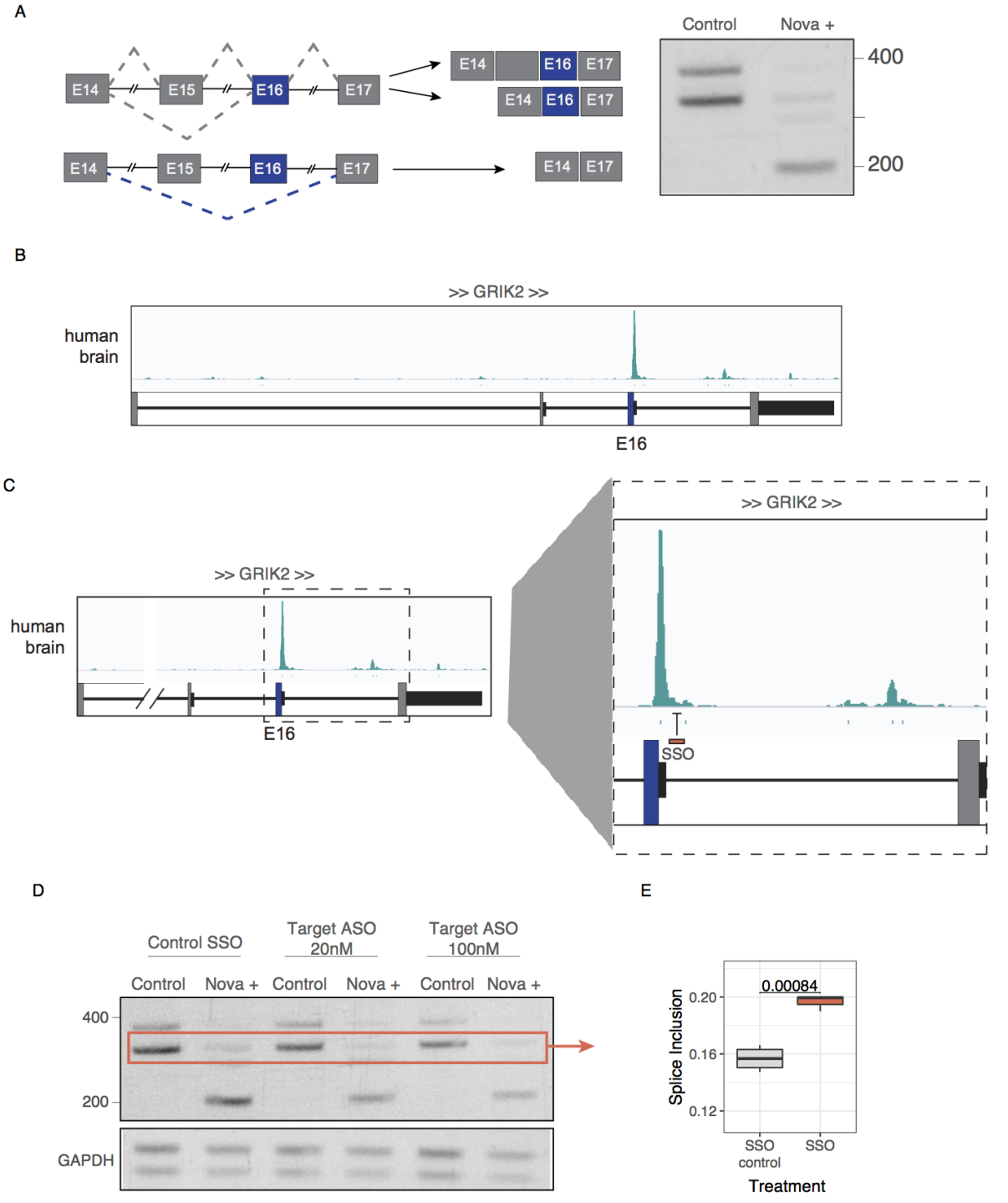


Figure 7-2: (A) Nova1 promotes silencing of GRIK2 exon 17 **(B)** *in vivo* Nova binding at 3' intron-exon junction is conserved in the mouse and human brain. **(C)** SSO is designed against *in vivo* Nova-RNA interaction site to inhibit Nova function at exon 17 splice. **(D)** *in vitro* transfection of SSO decreases the effect of Nova-dependent silencing at GRIK2 alternative exon 17, (Gel image is representative of n = 4).

Amyloid Precursor Like Protein 2 (APLP2)

An additional Nova target of exceptional interest is APLP2, which was identified as having two Nova-dependent splicing events. In the context of Nova regulation, APLP2 demonstrates splicing regulation that is conserved across species, but with divergent binding profiles as defined from mouse and human brain HITS-CLIP. Nova1-dependent inclusion of exon 14 was selected to validate as it was the more significant event (Figure 7-3A). Nova HITS-CLIP binding track shows a robust mouse Nova binding site, whereas the human brain HITS-CLIP track depicts a more dispersed binding pattern downstream of the target exon 14 (Figure 7-3B). Due to the human-specific focus of this study, the peak with the highest density of human brain CLIP tags was utilized for SSO design (Figure 7-3C). Transfection of the SSO into Nova1 expressing HeLa cells did not result in a significant change in calculated splice inclusion (Figure 7-3D-E). Here, our data does not suggest that inhibition of the human-specific Nova binding site via SSO results in a "switching" effect. One hypothesis could be that it is the cumulative effect of all downstream Nova binding that results in more inclusion of exon 14 and therefore one may need to inhibit all downstream Nova binding sites with multiple SSOs in order to detect a significant effect. Altogether it is clear that in order to delineate the relationship between Nova binding downstream of exon 14 more experimentation is required. Additional possible explanations could be that inhibiting Nova1 function at the human-specific location may have a secondary effect on splice outcome and thus may (1) lead to more Nova binding at the proximal and conserved location; (2) alter secondary structure of the splice event leading to variation in splice outcome; (3) cause activation of nonsense-mediated decay of the pre-mRNA. These findings highlight the need for future experimentation and approaches in order to further elucidate the mechanisms of APLP2 exon 14 splicing.

Figure 7-3 Nova1 promotes inclusion of APLP2 exon 14

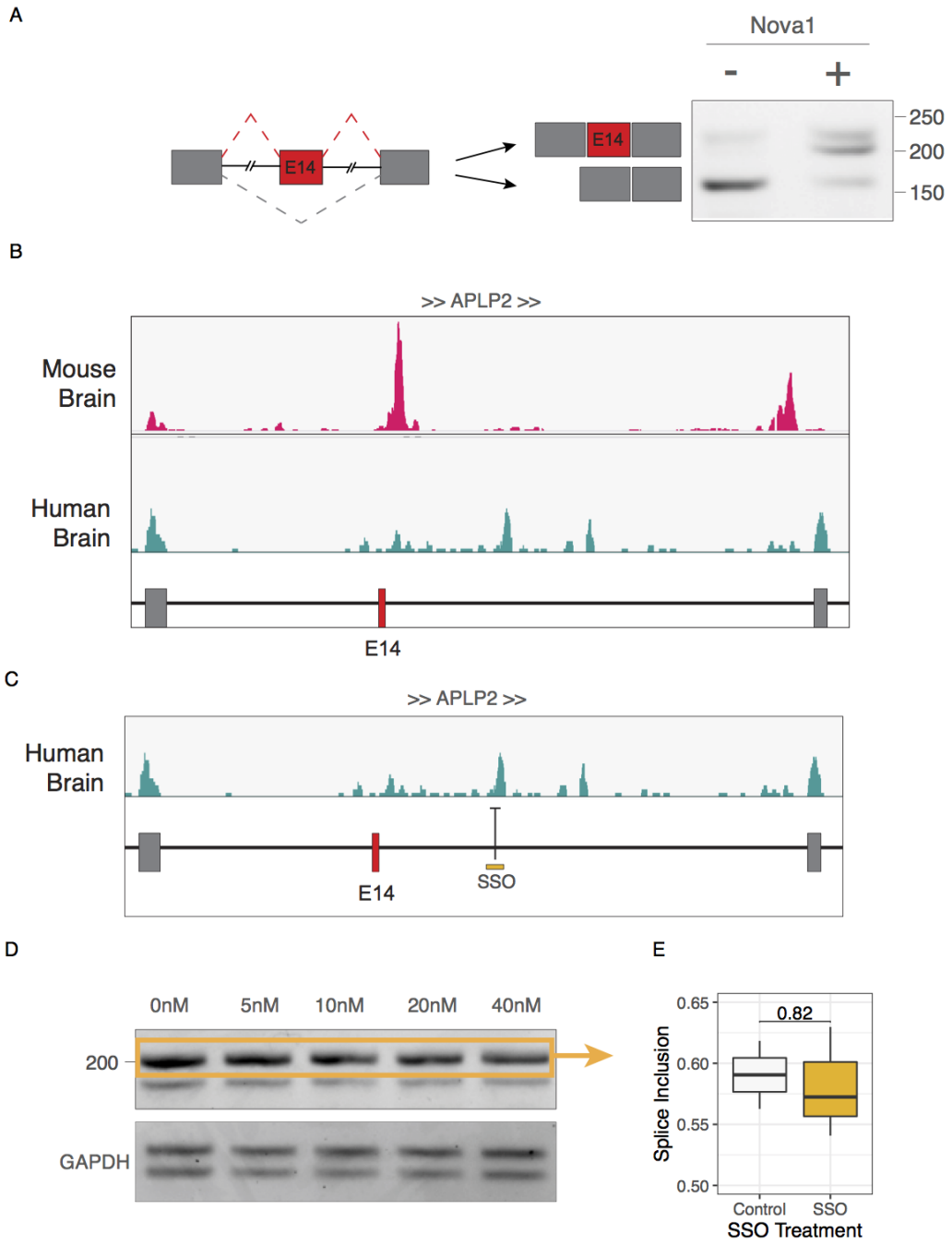


Figure 7-3: (A) Nova1 promotes inclusion of APLP2 exon 14 **(B)** *in vivo* Nova binding downstream of exon 14 (red) is conserved in the mouse and human brain. HITS-CLIP binding track; mouse CLIP mapped to human genome (top, pink), human CLIP (bottom, turquoise), and Nova1-dependent exon 14 (red) **(C)** SSO is designed against *in vivo* downstream Nova-RNA interaction site to inhibit Nova function at exon 14 splice. **(D)** *in vitro* transfection of SSO does not significantly alter the inclusion of APLP2 alternative exon 14 (Gel image is representative of n = 3).

Amyloid Precursor Protein (APP)

One of the most disease-relevant targets to come out of our combinatorial *in vitro*/ *in vivo* approaches was Nova1-dependent splicing of APP. Proper APP regulation and function is extremely important for proper synaptic connectivity in the human brain. However, APP is most notably recognized for its contribution to the pathogenesis of Alzheimer's disease. Effective and efficient methodology to modulate APP levels with isoform specificity would be of great importance. Ectopic expression of Nova1 showed Nova1-dependent silencing of APP exon 7 with both high-throughput sequencing methods and semi-quantitative PCR (Figure 7-4A). These data also show a Nova1 dependent increase in inclusion of APP exon 8. However, because of the human-specific upstream binding pattern observed in CLIP data, we focused SSO efforts on the peaks localized to the human-specific region of intron 6 (Figure 7-4B). Nova-CLIP binding tracks show two significant binding sites upstream of Nova regulated exon 7 (Figure 7-4C).

Figure 7-4 Nova1 regulates alternative processing of APP via upstream binding

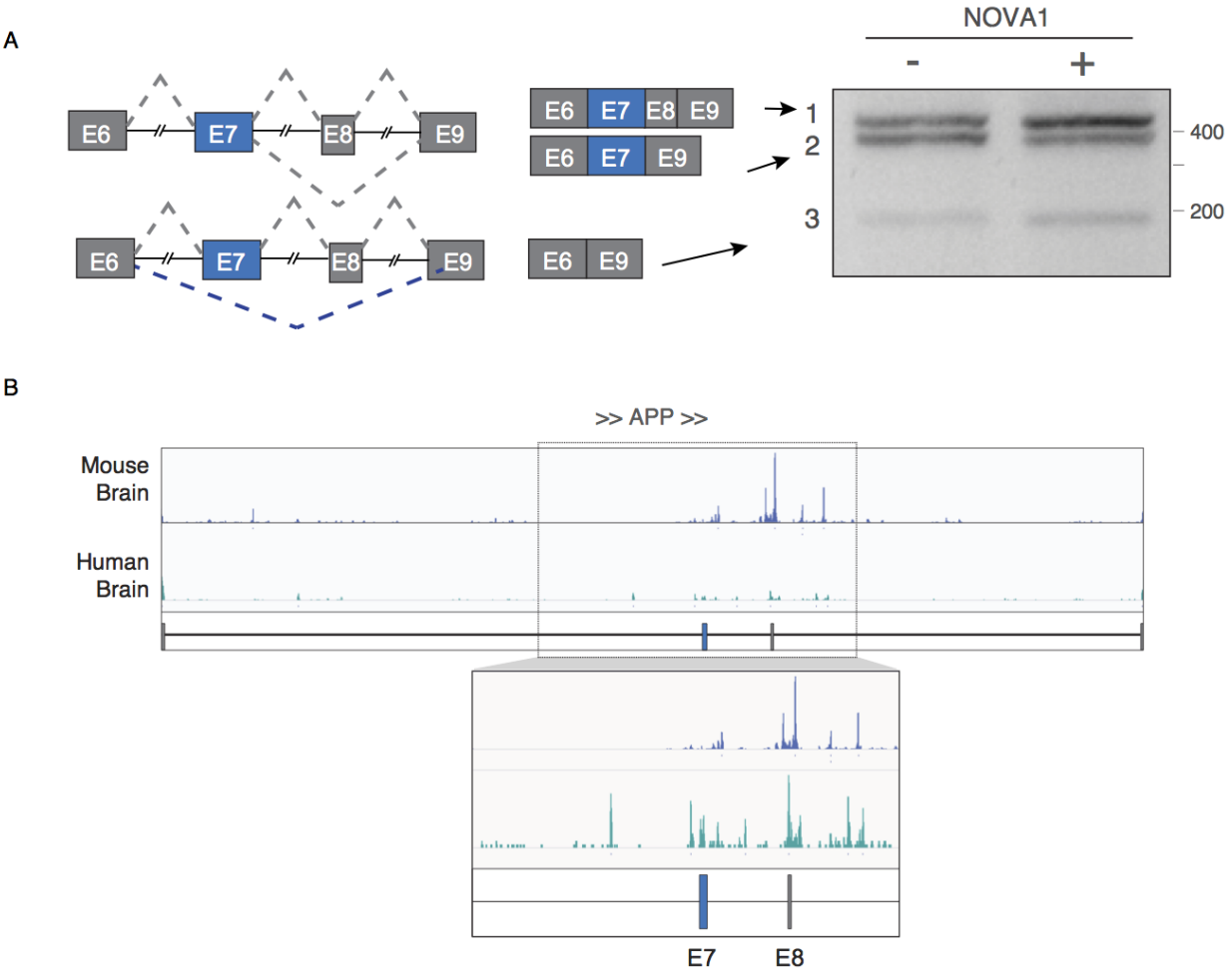


Figure 7-4: (A) Nova1 promotes silencing of APP exon 7. Diagram illustrating the three most prominent splice patterns of APP exons 7-8 and the splice outcome of exon structure. Splice outcomes (1-3) are aligned with corresponding gel bands (1-3, top to bottom). Semi-quantitative PCR gel shows Nova1-dependent increase in APP_exon7/8+ (band 1) and APP_exon7/8- (band 3). **(B)** *in vivo* Nova binding upstream of exon 7 is specific to the human brain. HITS-CLIP binding track; mouse CLIP mapped to human genome (top, blue), human CLIP (bottom, turquoise), and Nova1-dependent exon 7 (blue). **(C)** Zoomed image of human and mouse brain Nova HITS-CLIP binding profiles on hg38 genome.

We first focused on SSO targeting of the most proximal binding site with SSO-77035 (Figure 7-5A). Transfection of SSO-77035 in HeLa cells resulted in a dose dependent and reproducible increase in exon 7 inclusion (Figure 7-5B). While a significant reduction in the silencing of exon 7 (band 3) was observed with treatment of SSO-77035, the decrease did not reproducibly exhibit a dosage effect (data not shown).

Figure 7-5 Nova1 regulates alternative processing of APP via upstream binding

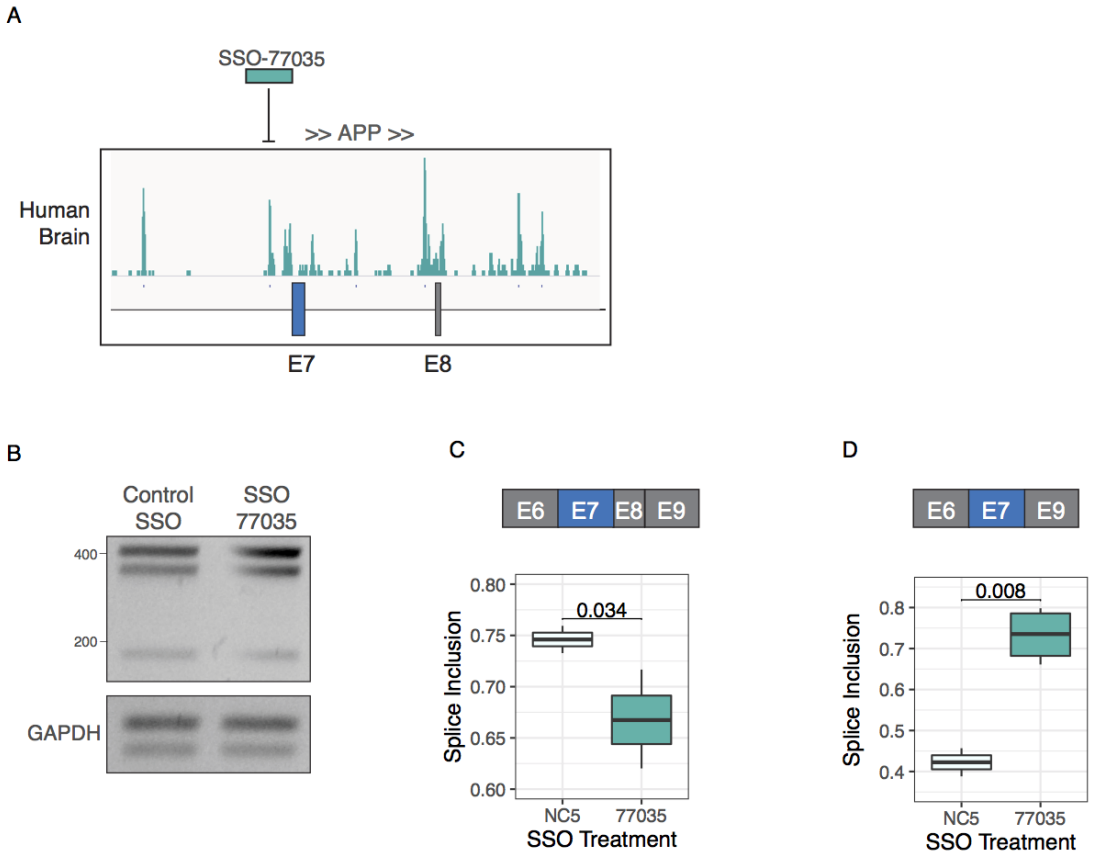


Figure 7-5: (A) HITS-CLIP binding track of human brain CLIP (turquoise), and Nova1-dependent exon 7 (blue). SSO (green) is designed against *in vivo* upstream Nova-RNA interaction site to inhibit Nova binding at peak site. **(B)** *in vitro* transfection of SSO-77035 increases the inclusion of alternative exon 7. (Gel image is representative of n = 14). **(C)** Quantitative analysis of the percent splice inclusion of APP isoform 1 Treatment with SSO-77035 (green) compared to SSO-control (white), SSO-77035 significantly decreases isoform 1 spliced inclusion respectively. **(D)** Nova1+ HeLa cells treated with SSO-77035 express significantly more of APP isoform 2 compared to SSO-control treated Nova1+ HeLa cells.

Next, to examine the functional impact of Nova binding to the human specific region of intron 6, an SSO was designed against the distal Nova binding site, SSO-77036 (Figure 7-6A). Transfection of SSO-77036 resulted in a reproducible and significant decrease in Nova-mediated silencing of exon 7, the effect was dose dependent (Figure 7- 6B).

Figure 7-6 Nova1 regulates alternative RNA processing of APP via secondary upstream binding site

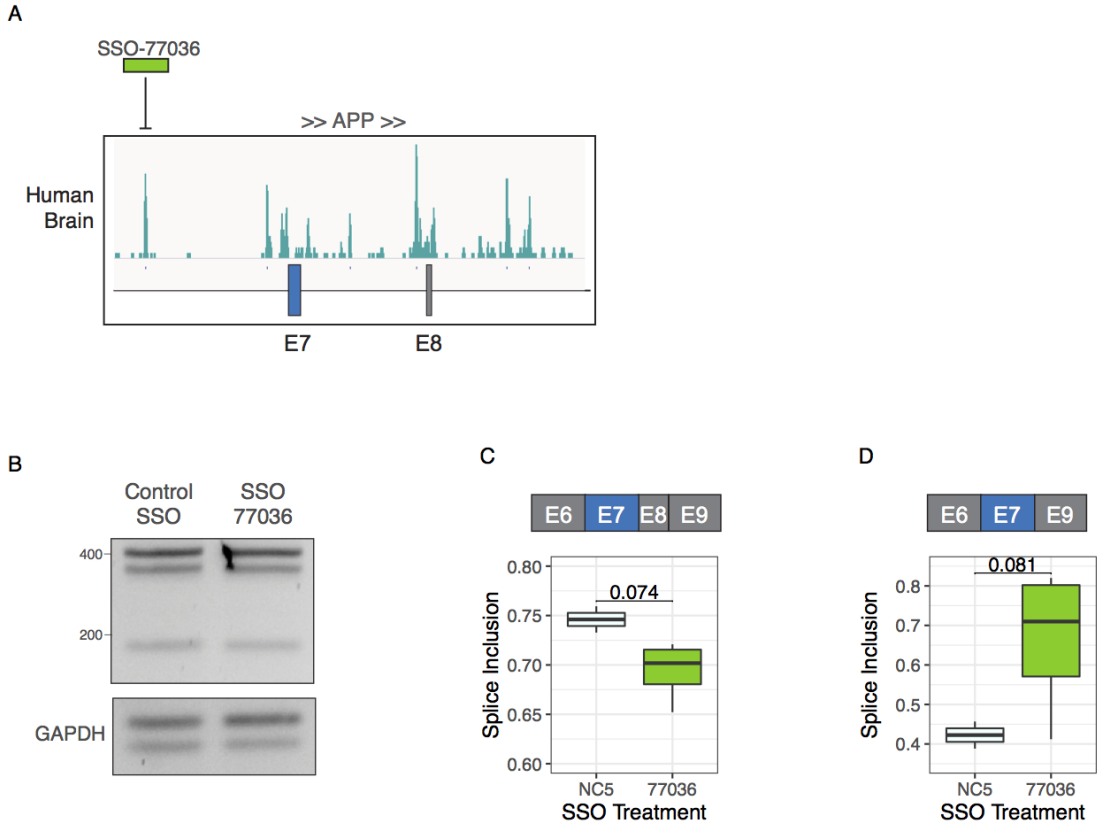


Figure 7-6: (A) *in vivo* Nova binding upstream of exon 7 is specific to the human brain. HITS-CLIP binding track of human brain CLIP (turquoise), and Nova1-dependent exon 7 (blue). SSO (bright green) is designed against *in vivo* upstream Nova-RNA interaction site to inhibit Nova binding at that position. **(B)** *in vitro* transfection of SSO-control and SSO-77036 Nova1+ cells. (Gel image is representative of n = 18). **(C)** Quantitative analysis of the percent splice inclusion for APP isoform 1. Treatment with SSO-77036 (bright green) compared to SSO-control (white) does not significantly alter APP isoform 1 expression. **(D)** Quantitative analysis of the percent splice inclusion for APP isoform 2. Treatment with SSO-77036 (bright green) compared to SSO-control (white) does not significantly alter APP isoform 1 expression.

Since individual transfection of SSO-77035 or SSO-77036 effectively alter exon 7 splicing, albeit with subtle differences in strength, we asked whether transfection of both SSO-77035 (more exon 7 inclusion) and SSO-77036 (less exon 7 silencing) would lead to a stronger effect on exon 7 splice outcome (Figure 7-7A). To answer this, HeLa cells were transfected with both SSOs revealing a dose-dependent effect on exon 7 silencing (band 3) (Figure 7-7C). The effect of Nova1 as well as the effect of the SSO on exon 7 inclusion is subtle and suggests that Nova is likely not the only factor regulating the splicing of exon 7 in HeLa cells. In neurons, band 3 (or silencing of exon 7-8) is the preferred APP isoform and therefore screening our SSO targets in a neuronal system would be the ideal next step to determining the efficiency of our SSO designs. Once again emphasizing the strength of a biologically relevant system in order to properly discern strength of splicing mechanisms.

Figure 7-7 SSO-77035 & SSO-77036 suppress Nova's inhibitory effect

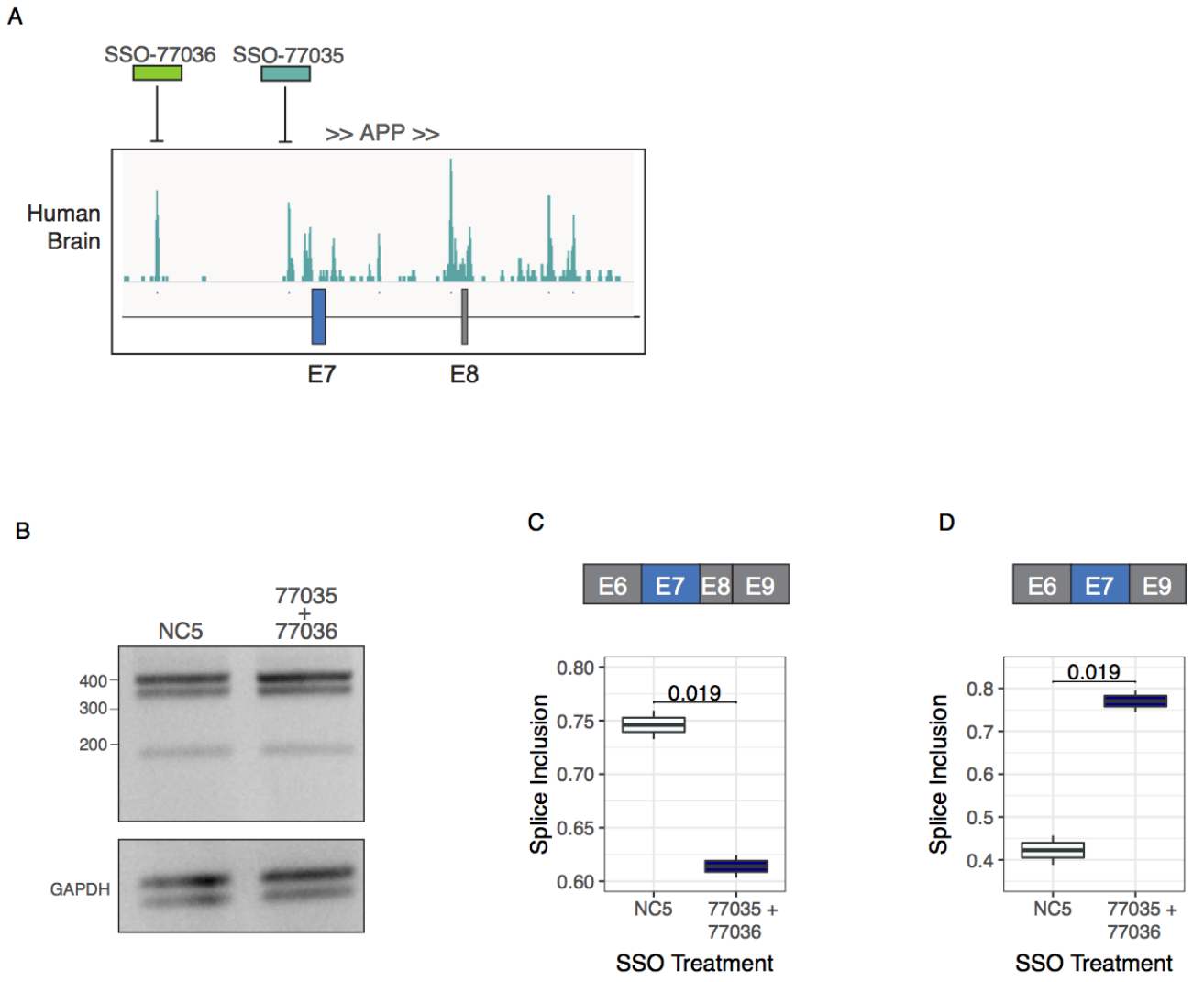


Figure 7-7: (A) HITS-CLIP binding track of human brain CLIP (turquoise), and Nova1-dependent exon 7 (green). SSO (blue) targets *in vivo* upstream Nova-RNA interaction site to inhibit Nova binding at site. **(B)** *in vitro* transfection of SSO-control compared with dual transfection of SSO-77035 & SSO-77036. Gel is representative of n =4. **(C)** Quantitative analysis of the percent splice inclusion for APP isoform 1. Treatment with both SSOs (SSO-77035 and SSO-77036) (blue) compared to SSO-control (white) significantly and reproducibly decreases APP isoform 1 expression. **(D)** Quantitative analysis of the percent splice inclusion for APP isoform 2. Treatment with both SSOs (SSO-77035 and SSO-77036) (blue) compared to SSO-control (white) significantly and reproducibly increases APP isoform 2 expression.

Discussion

Due to the recognized potential for RNA therapeutics, the pharmaceutical industry has pioneered numerous options for delivery and have set up pipelines for large based screening platforms (Nafee and Gouda, 2017). While the design of RNA-based therapeutics is relatively uncomplicated, finding a specific and targetable sequence, with low probability of off-target effects proves to be a rate-limiting step for discovery efficiency. In this context NOVA HITS-CLIP reliably narrows down the need for tiling strategies to identify splice regulatory site sequences. Furthermore, the high-throughput approach of HITS-CLIP enables the identification of hundreds of thousands of regulatory sites at once. Furthermore, the utility of a HITS-CLIP binding map provides a resource that can now be applied to screening of disease-relevant datasets and patient populations for expanding targeting strategies and possibilities. The specificity and resolution of Nova's HITS-CLIP defined regulatory network has demonstrated immense potential in the identification of novel non-coding regulatory sequences as actionable therapeutic targets in the treatment of disease.

Chapter 8: General Discussion

Nova's target network has robustly expanded in the human brain, and we know that Nova modulates expression levels of targets that are important for specialized neuronal functions and properties so it is important to continue to elucidate such species-specific differences in order to better understand the molecular underpinnings of disease.

Our findings demonstrate that Nova targets evolve "quickly" and through intronic changes. Nova proteins are evolutionarily constrained so it would make sense that the target network would evolve instead of evolution of Nova itself. In evolution if you examine the crystal structure of Nova it is invariant and almost identical between Nova1 and Nova2. The only variation occurs within the linker regions. One might hypothesize that it worked so well, it was locked and resistant to any change. Therefore, since the RNA binding capacity of Nova is consistent throughout evolution, in that it interacts with RNAs through YCAY elements, one may conclude that with fixed KH domains, YCAYs would be flexible and change for Nova. Global human-specific YCAY evolution of Nova target network is shown to be unique to intronic regions. In other words, Nova targets have selectively evolved YCAY tetramers within introns to modulate NOVA binding/regulatory capacity. This logic makes sense as introns actually accelerate the evolution of new genes and facilitate organismal complexity. But if YCAY tetramers have not evolved in 3'UTRs, then why do we observe so much more 3'UTR occupancy in the human cortical Nova binding maps?

Nova regulation of RNA expression profiles has widespread applications within the human cortex, which is a brain region that has become highly specialized in primates and

humans. And Nova target networks are largely shared between the human cortex and human cerebellum. Yet, global distribution of precise Nova-RNA interaction sites has diverged between the human cortex and human cerebellum. Furthermore, the human cortex has become highly specialized in primates and humans. Since most targets are bound by Nova in borth brain regions, there are several reasonable hypotheses that could account for the described changes and are all important directions to consider for future research. (1) First, Nova may be sequestered in the cytoplasm vs. nucleus as a consequence of alternative processing of Nova (we show that Nova undergoes a developmental switch in transcript usage that would alter expression of NLS/NES in human neurons). (2) Additionally, expression levels of Nova1 and Nova2 are different in humans compared to mice and changes in levels across brain regions cortex vs. cerebellum may lead to altered binding distributions. In the mouse, Nova1 has more 3'UTR localized regulatory sites and Nova2 is highly bound at introns. (3) Alternative RNA processing of Nova1 may alter RNA binding capabilities (alternative splicing controls Nova1 transcript usage during development giving rise to Nova1 isoforms with KH1, KH2/3, or KH1/2/3. (4) Less likely but still plausible possibilities include discrepancies in neuronal subtypes within the brain regions such as pyramidal vs. heavily myelinated vs. Purkinje cells as they all have very different physiological properties and therefore very different molecular properties and RNA metabolism. (5) Last, differences in proliferative capacity between brain regions such as developing vs. migrating vs. immature vs. mature neuronal states. We demonstrate that Nova temporally regulates targets during neuron development and therefore it is possible that Nova is actively regulating RNA in cells that require different forms of alternative control such as high plasticity areas vs. specialized neurons contributing to complexity vs. size and neuron structure.

Our data elucidates human-specific functional processes where RNA expression is subject to Nova control and more tightly regulated in human neurons. As expected we see neuron specific functions like neuron differentiation, synapse formation/assembly, axonal processes such as myelination and outgrowth and motility, circuit level control like membrane potential (capacity, activated, and resting). But we also observe robust Nova regulation among molecular processes such as DNA repair pathways, histone modifications, alternative RNA processing, stability/decay pathways, and translation. Altogether indicating higher levels of surveillance and control over these pathways and processes in human neurons, specifically in human cortical neurons. Our human brain CLIP data indicates the presence of more complex layers of RNA regulation and RNA processing mechanisms in the human brain as compared to the mouse brain.

Mouse models prove to be an indispensable *in vivo* system for studying NOVA proteins. They have demonstrated exceptional capacity for extrapolating disease relevance based on phenotype, function, binding profile, and target network. Mouse models offer unique cell-type specificity and *in vivo* disease models for comparative analysis. However, the robust differences observed between mouse-derived NOVA CLIP data does not accurately depict the specificity and resolution observed within our human brain derived NOVA CLIP binding maps. A level of detail that is essential in order to leverage these maps for therapeutic discovery and development.

Nova HITS-CLIP profiles in hIPS-cell derived neurons indicate they are attractive and relevant models to study relationships between NOVA target networks and disease. IPS-cells demonstrate promising capacity to answer questions involving neuron development and delineating Nova1/2 function since expression patterns are concordant with human brain

expression patterns. Although it may not be attainable to investigate the effect of Nova regulation on cellular motility and neuronal migration. Unfortunately IPS-derived neurons are not a quick model to set up as they require a great deal of hands on time and maintenance. Furthermore, once neurons are differentiated they are very slow to mature and therefore may not be efficient experimental models for NOVA regulation in some contexts such as aging and neurodegenerative disease.

Recent evidence suggests human brain organoids could be promising for studying neuron migration in addition to cell-cell interactions between neurons and neighboring cell-types such as glial, astrocytes, etc. However, the same studies have also shown that RNA steady state levels in the mouse brain are far more similar to *in vivo* human steady state levels compared to organoid systems (Pembroke et al., 2021). But that is not to say that the mechanisms giving rise to RNA steady state levels are also equivalent. Therefore, additional studies may be required to determine utility of the model as it pertains to RNA binding proteins and RNA-processing. Unpublished findings from our lab using humanized mouse models coupled with recently published data from human organoid models, independently demonstrate that the human-specific Nova1 protein isoform may indirectly alter the affinity of Nova-RNA interactions at highly specific regulatory sites (Trujillo et al., 2021). The cumulative sum of human-specific Nova1 binding depth at certain peaks, could be a contributing factor towards characteristics and behaviors that are unique to humans.

Chapter 9: Methods

Procurement of post-mortem human brain sample

All post-mortem human brain samples were procured from the National Institutes of Health (NIH) NeuroBioBank. All samples received from NIH NeuroBioBank and associated sample information are described in Table 9-1. Experiments which utilized each tissue are noted respectively. Samples that underwent CLIP-sequencing are highlighted in green.

Acknowledgement of Procurement:

Human tissue was obtained from the NIH Neurobiobank at the University of Maryland, Baltimore, MD; Harvard University, Boston, MA; Miami University, Miami, FL.

Figure 9-1 List of all human brain samples utilized in study

NIH NeuroBioBank Source	Distributive diagnosis	Sample ID	Age (years)	Sex	Tissue Subtype	NOVA-CLIP	RNA-seq	RIN
MARYLAND	Cognitively Normal	1599	2	F	BA9		TRUE	6.5
	Cognitively Normal				Cerebellum		TRUE	6.5
	Cognitively Normal				Dentate Nuclei		TRUE	6.5
	Cognitively Normal	4321	2	F	BA9		TRUE	6.9
	Cognitively Normal				Cerebellum		TRUE	6.9
	Cognitively Normal				Dentate Nuclei		TRUE	6.9
	Cognitively Normal	4729	24	M	BA9	TRUE	TRUE	
	Cognitively Normal	5161	10	F	BA9		TRUE	8.2
	Cognitively Normal				Cerebellum		TRUE	8.2
	Cognitively Normal				Dentate Nuclei		TRUE	8.2
	Cognitively Normal	5387	12	M	BA9		TRUE	7.5
	Cognitively Normal				Cerebellum		TRUE	7.5
	Cognitively Normal				Dentate Nuclei		TRUE	7.5
	Cognitively Normal	5408	6	M	BA9		TRUE	7.6
	Cognitively Normal				Cerebellum		TRUE	7.6
	Cognitively Normal				Dentate Nuclei		TRUE	7.6
	Cognitively Normal	5558	5	M	BA9		TRUE	5.9
	Cognitively Normal				Cerebellum		TRUE	5.9
	Cognitively Normal				Dentate Nuclei		TRUE	5.9
	Cognitively Normal	5576	2	M	BA9		TRUE	5.7
	Cognitively Normal				Cerebellum		TRUE	5.7
	Cognitively Normal				Dentate Nuclei		TRUE	5.7
	Cognitively Normal	5580	29	M	BA9	TRUE	TRUE	7
	Cognitively Normal				Cerebellum	TRUE	TRUE	7
	Cognitively Normal				Dentate Nuclei	TRUE	TRUE	7
	Cognitively Normal	5614	31	M	BA9	TRUE	TRUE	7
	Cognitively Normal	5644	29	F	BA9	TRUE	TRUE	6.7
	Cognitively Normal	5647	2	M	BA9		TRUE	5.5
	Cognitively Normal				Cerebellum		TRUE	5.5
	Cognitively Normal				Dentate Nuclei		TRUE	5.5
	Cognitively Normal	5669	24	F	BA9	TRUE	TRUE	
	Cognitively Normal	5871	2	M	BA9		TRUE	6.4
	Cognitively Normal				Cerebellum		TRUE	6.4
	Cognitively Normal				Dentate Nuclei		TRUE	6.4
	Cognitively Normal	5925	7	F	BA9		TRUE	6.4
	Cognitively Normal				Cerebellum		TRUE	6.4
	Cognitively Normal				Dentate Nuclei		TRUE	6.4
HARVARD	Cognitively Normal	6597	33	M	Cerebellum	TRUE	TRUE	
	Cognitively Normal	11551	24	M	Cerebellum	TRUE	TRUE	
	Cognitively Normal	15240	36	F	Cerebellum		TRUE	
	Cognitively Normal	16665	36	M	Cerebellum	TRUE	TRUE	
MIAMI	Cognitively Normal	HCTDI	25	M	Dentate Nuclei	TRUE	TRUE	
	Cognitively Normal				BA46	TRUE	TRUE	
	Cognitively Normal	HCTDL	24	M	Dentate Nuclei	TRUE	TRUE	
	Cognitively Normal				BA46	TRUE	TRUE	
	Cognitively Normal	HCTFN	29	M	Dentate Nuclei	TRUE	TRUE	
	Cognitively Normal				BA46	TRUE	TRUE	
	Cognitively Normal	HCTMW	35	M	Dentate Nuclei	TRUE	TRUE	
	Cognitively Normal				BA46	TRUE	TRUE	
	Cognitively Normal	HCTNX	34	F	Dentate Nuclei	TRUE	TRUE	
	Cognitively Normal				BA46	TRUE	TRUE	

RNA-isolation

Total RNA was prepared via Trizol extraction (Life Technologies, Carlsbad, CA), precipitation with ethanol, and yields determined by absorption spectroscopy using a NanoDrop (NanoDrop Products, Wilmington, DE).

RNA-sequencing

RNA-seq samples were treated with RQ1 RNase-free DNase (Promega) prior to RNA isolation and library preparation. RNA was further purified for polyadenylated RNA by using Dynabeads mRNA Purification Kit (Ambion). The libraries were prepared by TruSeq RNA Sample Preparation Kit v2 (Illumina) following manufacturer's instructions. High-throughput sequencing was performed on HiSeq (Illumina) to obtain 100 nucleotide paired-end reads.

High-throughput sequencing cross-linked immunoprecipitation (HITS-CLIP)

Sample preparation & UV-crosslinking: **Cells:** Media was carefully removed from wells and washed 3x with 1mL 1X PBS. Cells remained in third 1X PBS wash for crosslinking. Plates were places on ice, the cover was removed and plate was irradiated in Strategene for 400mJ/cm^2(4,000 μJ/cm^2). Plate was rotated and irradiated one more time for 200mJ/cm^2(2,000 μJ/cm^2). Cells were collected via cell scraper into 1ml 1XPBS and placed into Eppendorf tube which was pelleted at 2500rpm for 5 minutes at 4C. Liquid was removed via pipette and pellet was flash frozen in liquid nitrogen and stored at -80C for future CLIP processing. **Post-mortem human brain tissue:** All sample preparation was carried out on dry ice to prevent any freeze/thaw of the samples. Frozen tissue samples were placed in a cell crusher (http://cellcrusher.com/tissuepulverizer/) that was pre-cooled in liquid nitrogen. Tissue was

pulverized with a hammer and tissue power was collected with liquid-nitrogen precooled spoons into a precooled (dry ice) tissue culture dish. Plate was UV irradiated on dry ice 3x at 400mJ/cm^2, plate was rotated in between irradiations. Lysis buffer (PXL+) was added to (frozen) powdered tissue. Lysate was pipetted several times before being flash frozen in liquid nitrogen. UV-crosslinked tissue powders were stored at -80C for future CLIP processing.

Bead preparation: Protein A Dynabeads (Invitrogen) were resuspended and washed three times in antibody binding buffer (1x PBS, 0.02% Tween-20, pH 7.4). 200µl of beads were used per sample per IP. Beads were pre-coated with rabbit polyclonal antibodies against c-terminus Nova Fusion Protein (citation). 15µl of antibody was added to 185µl of antibody binding buffer per IP which was then added to the beads and rotated at room temperature for at least 45 minutes.

Brain lysate preparation: Previously prepared and frozen tissue powder samples (~100mg each) were thawed and lysed in 800µl 1X PXL lysis buffer (1X PBS [Ambion], 0.1% SDS, 0.5% Na-DOC, 0.5% NP-40 with 1 tablet/10 mL cOmplete protease inhibitor cocktail (Roche)). Lysates were then incubated on ice for 5-10 minutes. 60µl of RQ1 DNAse (Promega, M6101) was added to each tube and incubated at 37°C thermomixing at 1100 rpm for 5 minutes.

Partial RNA digestion: Samples were then treated with RnaseA, which was first diluted to the indicated concentration by volume (1:100 or 1:10,000) and then added at 10µl per lysate accordingly. It is important to determine the amount of RNAse to be added prior to CLIP experiment as the dilution will vary depending on tissue type. The RNAse digestion is essential for interpreting CLIP autorads. Lysates were then subjected to an additional round of thermomixing for 5 minutes at 37°C at 1100 rpm.

Ultracentrifugation & Immunoprecipitation: Lysates in Eppendorf tubes were spun low-high speed in order to remove particulates from the lysates.

10µl of supernatant was removed and stored at -80°C for future immunoblot analysis as input. The remaining supernatants were pre-cleared via 1 hour IP with unbound sepharose-A beads at 4C according to manufacturer instructions. The supernatants were then added to the prepared tube of beads. Bead/lysate mixture was rotated for 3 hours at 4°C. Following IP, a small amount of post-IP lysate was kept. Beads were washed sequentially twice each with 1X PXL, 5x PXL (same as 1X but using 5X PBS), and 1X PNK buffer (50mM Tris-HCl, pH 7.5, 10mM MgCl2, 0.5% NP-40).

Dephosphorylation and labeling of RNA fragments: To prevent RNA circularization of RNA tags, IPs were treated with alkaline phosphatase. Beads were resuspended in 80µl containing 1x dephosphorylation buffer, 3U of CIAP (Roche), RNasin inhibitor (Promega), and thermomixed for 20 minutes at 37°C, shaking at 1100rpm for 15s ever 2 minutes. Samples were then washed as above sequentially in 1X PNK, 1X PNK plus 20mM EGTA, and twice with 1X PNK. Radiolabeled linkers for 3' linker ligation per sample were set up by preparing a T4 phosphonucleotide kinase (PNK, NED) reaction according to the manufacturers instructions using 2.5µl 32P-γ-ATP and an RNA linker, either 20pmol L32 or L34 RNA linker (Dharmacon or IDT) and incubating for 30 minutes at 37°C. To drive the reaction to completion, 0.5µl of 1mM ATP was added and incubated for an additional 5 minutes. Linkers were purified from free nucleotides using G-25 columns following the manufacturers instructions (GE Healthcare). SDS-PAGE and nitrocellulose transfer were carried out for NOVA without the addition of 10X Reducing Agent (NuPage) during elution of complexes from the bead.

Purification of RNA fragments from nitrocellulose membrane: Tags were collected from nitrocellulose as described (irCLIP paper, Zarnegar et al Nature Methods 13, 489-492, 2016) with the following exceptions: Phenol:Chloroform:IAA, 25:24:1 pH 6.6 was used for extraction, and tags were precipitated with a standard NaOAC precipitation. Cloning was performed using the BrdU-CLIP protocol as described (Moore et al., 2018) with a few exceptions. Briefly, the RT primer contains a 14-nt degenerate linker (a 3-nt degenerate sequence, a 4-nt multiplexing index, and a 7-nt unique molecular identifier), a 5′linker for PCR amplification, a spacer to prevent rolling circle amplification after circularization, and the reverse-complementary sequence of the 3′linker for reverse transcription. BrdUTP-labeled cDNA was specifically isolated via two sequential BrdUTP immunoprecipitations (with Abcam AB8955), and circularized with CircLigase II (Epicenter CL9025K).

Reverse transcription and PCR amplification: RNA was purified as in NOVA HITS-CLIP from excised membrane fragments corresponding to crosslinked lanes at ~50-60kD. Pelleted RNA recovered from the first phenol:chloroform extraction were resuspended in 6.25!l H2O and denatured at 65C for 5min with no shaking, then placed on ice. The polyA tailing reaction mix was prepared on ice by adding 1µl 10X E-PAP Buffer (Promega), 1.5µl 1mM ATP, 1.5µl RNAsin Plus (diluted to 5U/µl) and 1µl E-PAP (3U/µl, NEB M0276) to the purified RNA. The reaction was incubated for 15min at 37C, then 20min at 65C in a pre-warmed thermocycler. 2.5µl of each polyA-tailed reaction was taken for a –RT control with an additional 2.5µl H2O. Reverse transcription was carried out using 5µl polyA tailed RNA, 1µl 0.752M Tris, 1µl 8.2mM dATP, 1µl 8.2mM dCTP, 1µl 8.2mM dGTP (all Invitrogen), 1µl 8.2mM Br-dUTP (Sigma), 1µl 25µM RT Primer and 1µl H2O. This mix was incubated for 3min at 75C then ramped down to and held at 48C. 1µl 82mM DTT, 1µl 10U/µl RNasin and 1µl SuperscriptIII (or H2O for –RT)

pre-warmed to 48C were added to the mix, then incubated for 45min 48C, 15min 55C, 5mi 85C then held at 4C. RT-primers with different index sequences were used, one per replicate.

PCR Reaction Purification: PCR product was purified using Agencourt AMPure XP beads (Beckman Coulter) according to manufacturer's standard protocol. Beads were equilibrated to room temperature. PCR reaction volume was measured and beads were added at 1.8x PCR reaction volume and gently mixed via pipette. Reaction was left for 5' on the benchtop for beads to bind followed by a 15' on magnet incubation. 200µl 70% EtOH was added directly to the bottom of the tube for at least 30 seconds and repeated for a second wash. Beads were air-dried for 5' and 40µl EB buffers were directly added to beads, which were pipette up/down at least 10 times to mix. The sample was eluted for 5' at room temperature. Supernatant was collected via magnet and all purified samples were submitted for tape station analysis.

All multiplexed NOVA-CLIP libraries were pooled according to tape station quantification and diluted to 10nM and sequenced on HiSeq (Illumina) using standard Read 1 sequencing primer to obtain 100-nt single-end reads.

HITS-CLIP Adaptors and Primers:

Pre-Adenylated L32 adaptor:

L32: /5rApp/GTGTCAGTCACTTCCAGCGG/3ddc/

DNA Primers (IDT):

RT primers (2nd generation, with 6 nt barcodes that correspond to the TruSeq barcode sequences)

RT-1T: ATCACG (CGTGAT)

/5Phos/DDDCGTGATNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-2T: CGATGT (ACATCG)

/5Phos/DDDACATCGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-3T: TTAGGC (GCCTAA)

/5Phos/DDDGCCTAANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-4T: TGACCA (TGGTCA)

/5Phos/DDDTGGTCANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-5T: ACAGTG (CACTGT)

/5Phos/DDDCACTGTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-6T: GCCAAT (ATTGGC)

/5Phos/DDDATTGGCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-7T: CAGATC (GATCTG)

/5Phos/DDDGATCTGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-8T: ACTTGA (TCAAGT)

/5Phos/DDDTCAAGTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-9T: GATCAG (CTGATC)

/5Phos/DDDCTGATCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-10T: TAGCTT (AAGCTA)

/5Phos/DDDAAGCTANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-11T: GGCTAC (GTAGCC)

/5Phos/DDDGTAGCCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-12T: CTTGTA (TACAAG)

/5Phos/DDDTACAAGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-13T: ATCACG (CGTGAT)

/5Phos/DDDCGTGATNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-14T: AGTTCC (GGAACT)

/5Phos/DDDGGAACTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-15T: ATGTCA (TGACAT)

/5Phos/DDDTGACATNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-16T: CCGTCC (GGACGG)

/5Phos/DDDGGACGGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-18T: GTCCGC (GCGGAC)

/5Phos/DDDGCGGACNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-19T: GTGAAA (TTTCAC)

/5Phos/DDDTTTCACNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-20T: GTGGCC (GGCCAC)

/5Phos/DDDGGCCACNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-21T: GTTTCG (CGAAAC)

/5Phos/DDDCGAAACNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-22T: CGTACG (CGTACG)

/5Phos/DDDCGTACGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-23T: GAGTGG (CCACTC)

/5Phos/DDDCCACTCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-25T: ACTGAT (ATCAGT)

/5Phos/DDDATCAGTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT-27T: ATTCCT (AGGAAT)

/5Phos/DDDAGGAATNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/**CCG**

CTGGAAGTGACTGAC-3

RT Primers (1st generation):

RT-1 (ATCG):

/5Phos/DDDCGATNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-2 (GCTA):

/5Phos/DDDTAGCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-3 (CGAT):

/5Phos/DDDATCGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-4 (TAGC):

/5Phos/DDDGCTANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-5 (CTAG):

/5Phos/DDDCTAGNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-6 (GATC):

/5Phos/DDDGATCNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-7 (AGCT):

/5Phos/DDDAGCTNNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

RT-8 (TCGA):

/5Phos/DDDTCGANNNNNNNAGATCGGAAGAGCGTCGT/iSp18/CACTCA/iSp18/CCGCT

GGAAGTGACTGAC-3

PCR primers:

DP5-PE: (Allowing standard Read1 sequencing primer)

5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

SP3-PE

5’-CAAGCAGAAGACGGCATACGAGATCTCGGCATTCCTGCCGCTGGAAGTGACTGACAC

Sequencing primers:

The 1st Sequencing primer (Standard Illumina, Read 1):

5’-ACACTCTTTCCCTACACGACGCTCTTCCGATCT

The 2nd Sequencing primer (Custom, SSP2):

5’-TCGGCATTCCTGCCGCTGGAAGTGACTGACAC

Processing and alignment:

CLIP reads were processed using CLIP Tool Kit (CTK) (Moore et al., 2014; Shah et al., 2017). Briefly, raw reads were filtered for quality and de-multiplexed using indexes introduced during the reverse transcription reaction. PCR duplicates were collapsed and adapter sequences removed. Reads were mapped to the hg38 RefSeq genome using novo-align. Mapped reads were further collapsed for potential PCR duplicates by coordinates and taking into consideration the degenerate barcodes introduced during the reverse transcription. Only reads that were uniquely mapped in the appropriate direction were retained as “unique”.

Fluorescence Activated Cell Sorting (FACS)

Cell surface expression was measured using fluorescently-conjugated antibodies according to the manufacturer’s instructions. Cells were resuspended in 100µl FACS buffer (PBS supplemented with 1% PHS, 1% FBS), 1µg of each antibody was added and incubated for 10 minutes at room temperature in the dark. Cells were washed twice with FACS buffer and collected immediately on a BD FACS caliber machine and analyzed using Flowjo software (Treestar).

Immunohistochemistry (IHC)

Cells were washed 3x with 1X PBS at room temperature. They were fixed with 4% paraformaldehyde for 2 hours shaking at room temperature. Fixed cells were first washed three times with PBS at room temperature (RT), incubated with 0.2% Triton X- 100/PBS at RT, blocked with 1.5% normal donkey serum (NDS)/PBS at RT, and then incubated overnight at 4°C with primary antibodies in 1.5% NDS/PBS followed by incubation with Alexa 488 and 555 conjugated donkey secondary antibodies (1:1000) in 1.5% NDS/PBS. Images of immunostained specimens were taken using BZ-X700 (KEYENCE) microscope. All antibodies utilized are listed below.

Splice-switch oligonucleotide (SSO) transfections

HeLa cells were cultured in 1X DMEM supplemented with 10% FBS in antibiotic free media. 24 hours after plating cells were treated with doxycycline to induce Nova1 expression. On day 2 cells were transfected in triplicate using Lipofectamine 2000 (Life Technologies) following the manufacturer's instructions using either 20 nM splice-switching oligonucleotides (SSOs) or, for the dose-response curve, 1–80 nM. The 2′-MOE phosphorothioate backbone SSOs were synthesized by IDT. Media was changed 24 hours post transfection and cells were harvested 48 h after transfection. RNA was isolated using Trizol extraction (Life Technologies, Carlsbad, CA) and reverse-transcribed using iScript Reverse Transcriptase (Bio-rad). PCR was performed using *Taq* DNA Polymerase with Standard *Taq* Buffer (NEB) and analyzed by agarose gel electrophoresis. Bands were quantified using ImageJ.

All SSOs were designed using the hg38 reference genome and human NOVA CLIP peaks.

APP_77036_5

/52MOErA/*/i2MOErT/*/i2MOErG/* /i2MOErG/*/i2MOErG/*/i2MOErA/*

/i2MOErT/*/i2MOErA/*/i2MOErA/* /i2MOErA/*/i2MOErA/*/i2MOErG/*

/i2MOErA/*/i2MOErA/*/i2MOErT/* /i2MOErG/*/i2MOErA/*/i2MOErT/*

/i2MOErG/*/32MOErA/

APP_77035_4

/52MOErG/*/i2MOErA/*/i2MOErT/* /i2MOErG/*/i2MOErA/*/i2MOErG/*

/i2MOErT/*/i2MOErA/*/i2MOErT/* /i2MOErG/*/i2MOErA/*/i2MOErA/*

/i2MOErA/*/i2MOErT/*/i2MOErG/* /i2MOErA/*/i2MOErC/*/i2MOErG/*

/i2MOErG/*/32MOErA/

APP_77031_5

/52MOErA/*/i2MOErG/*/i2MOErT/* /i2MOErG/*/i2MOErG/*/i2MOErA/*

/i2MOErT/*/i2MOErG/*/i2MOErA/* /i2MOErT/*/i2MOErG/*/i2MOErA/*

/i2MOErA/*/i2MOErA/*/i2MOErG/* /i2MOErT/*/i2MOErG/*/i2MOErG/*

/i2MOErT/*/32MOErA/

APP_77032_3

/52MOErG/*/i2MOErA/*/i2MOErA/* /i2MOErA/*/i2MOErA/*/i2MOErA/*

/i2MOErT/*/i2MOErT/*/i2MOErA/* /i2MOErA/*/i2MOErA/*/i2MOErT/*

/i2MOErG/*/i2MOErA/*/i2MOErT/* /i2MOErC/*/i2MOErT/*/i2MOErG/*

/i2MOErT/*/i2MOErG/*/32MOErA/

APLP2_30164_5

/52MOErG/*/i2MOErT/*/i2MOErA/* /i2MOErA/*/i2MOErC/*/i2MOErA/*

/i2MOErC/*/i2MOErT/*/i2MOErT/* /i2MOErA/*/i2MOErG/*/i2MOErT/*

/i2MOErA/*/i2MOErA/*/i2MOErA/* /i2MOErG/*/i2MOErT/*/i2MOErT/*

/i2MOErG/*/32MOErT/

GPHN_37985_4

/52MOErA/*/i2MOErA/*/i2MOErC/* /i2MOErA/*/i2MOErA/*/i2MOErC/*

/i2MOErT/*/i2MOErA/*/i2MOErT/* /i2MOErG/*/i2MOErA/*/i2MOErA/*

/i2MOErT/*/i2MOErA/*/i2MOErT/* /i2MOErA/*/i2MOErG/*/i2MOErT/*

/i2MOErG/*/32MOErA/

GPHN_37987_5

/52MOErA/*/i2MOErA/*/i2MOErT/* /i2MOErG/*/i2MOErA/*/i2MOErG/*

/i2MOErA/*/i2MOErG/*/i2MOErC/* /i2MOErT/*/i2MOErC/*/i2MOErA/*

/i2MOErA/*/i2MOErT/*/i2MOErG/* /i2MOErA/*/i2MOErA/*/i2MOErT/*

/i2MOErG/*/32MOErA/

GRIK2_108055_10

/52MOErA/*/i2MOErG/*/i2MOErT/* /i2MOErG/*/i2MOErT/*/i2MOErA/*

/i2MOErA/*/i2MOErA/*/i2MOErT/* /i2MOErT/*/i2MOErA/*/i2MOErG/*

/i2MOErT/*/i2MOErA/*/i2MOErG/* /i2MOErT/*/i2MOErG/*/i2MOErG/*
/i2MOErG/*/32MOErT/

NC5_2'MOE
/52MOErG/*/i2MOErC/*/i2MOErG/* /i2MOErA/*/i2MOErC/*/i2MOErT/*
/i2MOErA/*/i2MOErT/*/i2MOErA/* /i2MOErC/*/i2MOErG/*/i2MOErC/*
/i2MOErG/*/i2MOErC/*/i2MOErA/* /i2MOErA/*/i2MOErT/*/i2MOErA/*
/i2MOErT/*/32MOErG/

Semi-quantitative RT-PCR

Triplicate dishes of HeLa-WT and NOVA1 Tet-On cells were treated for three days with 10ng/ml doxycycline. Cell pellets were thawed and lysed in PXL+ (previously described), underwent DNase treatment with RQ1 DNase for 10' shakings @ 37C. Total RNA was purified from DNase treated lysates by Trizol extraction (Life Technologies, Carlsbad, CA) and reverse-transcribed using iScript Reverse Transcriptase (Bio-rad). PCR was performed using *Taq* DNA Polymerase with Standard *Taq* Buffer (NEB) and analyzed by agarose gel electrophoresis. Bands were quantified using ImageJ.

The ratio of included isoforms was determined using ImageJ analysis of scanned x-ray film. All primers were designed using the hg38 reference genome and Primer3Plus.

<u>Primers for validation:</u>

APLP2 F-GAG CTC TGA GGA GAG TGA GGA G

R-TGT TTT CAT CCA CTT TAT TCT TAC TG

APP	F-TTG GGA TGC TTC TTC GGT GT
	R-TTG GCT TTC TGG AAA TGG GC
BPTF	F-AGC AGA AGC CGA CAG TGA TT
	R-CCT GGG CTG GAA TGA AGT AA
DNM1L	F-GAG AAA CAG GCT AGC CAG AGA A
	R-GTT GCC TGT GGT TGG TTC TT
GAPDH	F-CAA TGC CTC CTG CAC CAC
	R-GCC ATC ACG CCA CAG TTT
GRIK2	F-TTC AGT TTT TGT GGC AGT GG
	R-CTG GCA ACC TTC TGT CGT TA
GPHN	F-CAG TGG TGT TGC TTC AAC AGA
	R-AGA ATG TTC TCC TTG CTG CTG
GTF2I	F-GGC AGC TGT GAC AGT AAA GGA
	R-GGG AAT CAC CTT TTA TTT CCA A
KIAA1217	F-GGA GCG CTG CAG GAT TAT
	R-CTA TCT GTG TGG TTC CTG TTG G
KIDINS220	F-GTG TGC TCT TCC ACG TCC TT
	R-TGG TCC AGC CCT TCT ATT TG
KMT2E	F-CTG CCT CAA CTT CAG TGG AAC
	R-TGA GAA GTA TCT GGG GAT GGA
MARK3	F-CAT GAA GCC ACA CCA TTG TC
	R-GTG GTT TTC ATG CTC CAG GT
NFIB_E7	F-AAG CCA TTG TTC AGC TCT GC

	R-CCG GTA AGA TGG GTG TCC TA
NOVA1	F-ACC TGA TGG GAT CAA CTT GC
	R-ATT GGA ATT TGC CAC TGG AC
PPFIBP1	F-GAA CAA CTA GAA GAA AAG GAA TCT GAA
	R-TTT CTT CAT TTG CTG CCA TC
PPP1R12B	F-TAT ACC CGG CAG CTA TGG AG
	R-TTG GCA GTG CTA GGA AGA GG
PTPRF	F-GAC CCA CTC TCC GTC CTC TA
	R-GAG AAC TTG AGG CCA TCG TT
SMARCC2	F-AAG TTT GGG CCC TTC TGA AC
	R-AGA TGT CTG GCT GGC TCC T
SMARCE1	F-GGA ACT GAG AAC TTA ATC TTC CAA A
	R-ACC TCA TGT AGG GCA TCA GC
SUPT5H	F-TGA GTA TGA GGA CGA GGA CCA
	R-CCC AGA ACG ATC TTC ATC CA
SYNE2	F-CCT CTC ACG AAG AGG ACG AG
	R-CAG GTG GAA CAT TCC TGT CA

In vitro models

Human neuroblastoma SH-SHY5Y cells were obtained from the American Type Culture Collection (ATCC) and cultured in a 50:50 mix of minimal essential media (MEM) and F12 nutrient mixture supplemented with 10% fetal bovine serum (FBS). SH-SY5Y cells were grown

(~1.2-2 × 10$_7$) in 150 mm dishes per replicate to ~80% confluency at which time they were either passaged for maintenance or harvested for high-throughput sequencing experiments accordingly.

Lund human mesencephalic (LUHMES) cells were obtained from the American Type Culture Collection (ATCC) (catalog no. CRL-2927) and cultured according to supplier's recommendations. All experiments were carried out with cells at a passage of 2 to 5 from the ATCC stock. Briefly, cell culture flasks and dishes were sequentially coated with poly-l-ornithine (Sigma-Aldrich) overnight at room temperature followed by fibronectin (Sigma-Aldrich) overnight at 37°C, plates were rinsed with phosphate-buffered saline (PBS), and allowed to fully air dry at room temperature before cell plating. LUHMES cells were maintained in a proliferation medium (Dulbecco modified Eagle medium [DMEM]-F12; ATCC) containing 1% N2 supplement (ThermoFisher Scientific), 1× penicillin-streptomycin-glutamine solution (ThermoFisher Scientific), and recombinant human fibroblast growth factor (FGF)-basic added to proliferation medium immediately before use at a 50-ng/ml final concentration. LUHMES cells were passaged at 80% confluency.

iPSC lines utilized for the studies described here were a generous gift from the Marc Tessier-Lavigne laboratory.

iPS Cell culture iPSCs were maintained on irradiated MEFs (Globalstem) plated on cell culture plates coated with 0.1% gelatin and grown in HUESM (Knock-out Dulbecco's modified Eagle's Medium (KO-DMEM), 20% knock-out serum, 0.1 mM non-essential amino acids, 2 mM Glutamax, 100 U/mL-0.1 mg/mL penicillin- streptomycin (all Life Technologies), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich), 10 ng/mL FGF2 (Stemgent), at 37 oC with 5% CO2. Prior

to transfection, iPSCs were transferred to Geltrex-coated (Life Technologies) cell culture plates and grown in MEF-conditioned HUESM containing 10 μM ROCK inhibitors (Stemgent).

Cortical neuron differentiation iPSC-derived cortical neurons were generated as described in (Shi et al., 2012) with modifications. Specifically, to generate neural precursor cells (NPCs), iPSCs were plated on 12-well tissue culture plates coated with Geltrex (Life Technologies) in MEF-conditioned HUESM with ROCK inhibitor. When cells were 100% confluent, medium was replaced with neural induction (NI) medium (day in vitro 0 (DIV0)) and maintained for 8 days. On DIV8 day cells were dissociated using Accutase (Life Technologies) and resuspended in NI medium at 30 million cells/mL. Cells were plated on dried poly-L-ornithine (Sigma-Aldrich) and laminin-coated (Life Technologies) 6-well plates in 200 μL spots. Cells were left to adhere for ~45 min and NI medium was added. On DIV10 NI was replaced with neural maintenance (NM) medium. Upon the appearance of neural rosettes, 20 ng/mL FGF2 was added for 2 days. When neurons started to form (~DIV21), rosettes were isolated manually after treatment using STEMdiff Neural Rosette Selection Reagent (STEMCELL Technologies) for 1 h. Rosettes were washed and plated on poly-L-ornithine/laminin-coated 6-well plates. 7 days after rosette selection and plating, the NPCs were expanded, again with STEM diff Neural Rosette Selection Reagent (STEMCELL Technologies) for 1.5hr. NPCs were washed and plated on poly-L-ornithine/laminin coated 6-well plates. Between DIV35 and DIV 38 NPCs were either frozen in NM supplemented with 10% DMSO and 20 ng/mL FGF2 or culture was continued for cortical neuron maturation.

For cortical neuron maturation, ~600,000-1,000,000 NPCs were plated on 6- well poly-L-ornithine/laminin-coated plates and maintained in Neurobasal medium supplemented with B-27 serum-free supplement, 2 mM Glutamax and 100 U/mL-0.1 mg/mL

penicillin-streptomycin (all Life Technologies). During the first 7 days after plating, cells were treated with 10 μM DAPT (Sigma-Aldrich) to augment neuronal maturation.

Generation of Tet-inducible HeLa cell lines ePB vectors (piggyback Tet-On + puroR) were utilized for HeLa cell line generation. (Lacoste et al., 2009) Vectors were subcloned with NOVA1 (NOVA1+) or without NOVA1 (control) and Puromycin for selection were expressed separately in HeLA cells (ATCC) using transient transfection with Lipofectamine 3000 (Thermofisher) using manufacture recommendations when cells were at 60-70% confluency. (Figure 9-1) Selection of clones expressing the vector of interest was performed with standard HeLa cell media containing increasing concentrations of 1-2ug/ml of Puromycin (Thermofisher) over a 2 week period. After 2 weeks of selection, individual colonies were picked and transferred to 24-well dishes and expanded in media containing 1ug/ml of Puromycin. Cells were frozen using standard media with 10% DMSO without puromycin.

Figure 9-2 Plasmids used for HeLa cell line generation

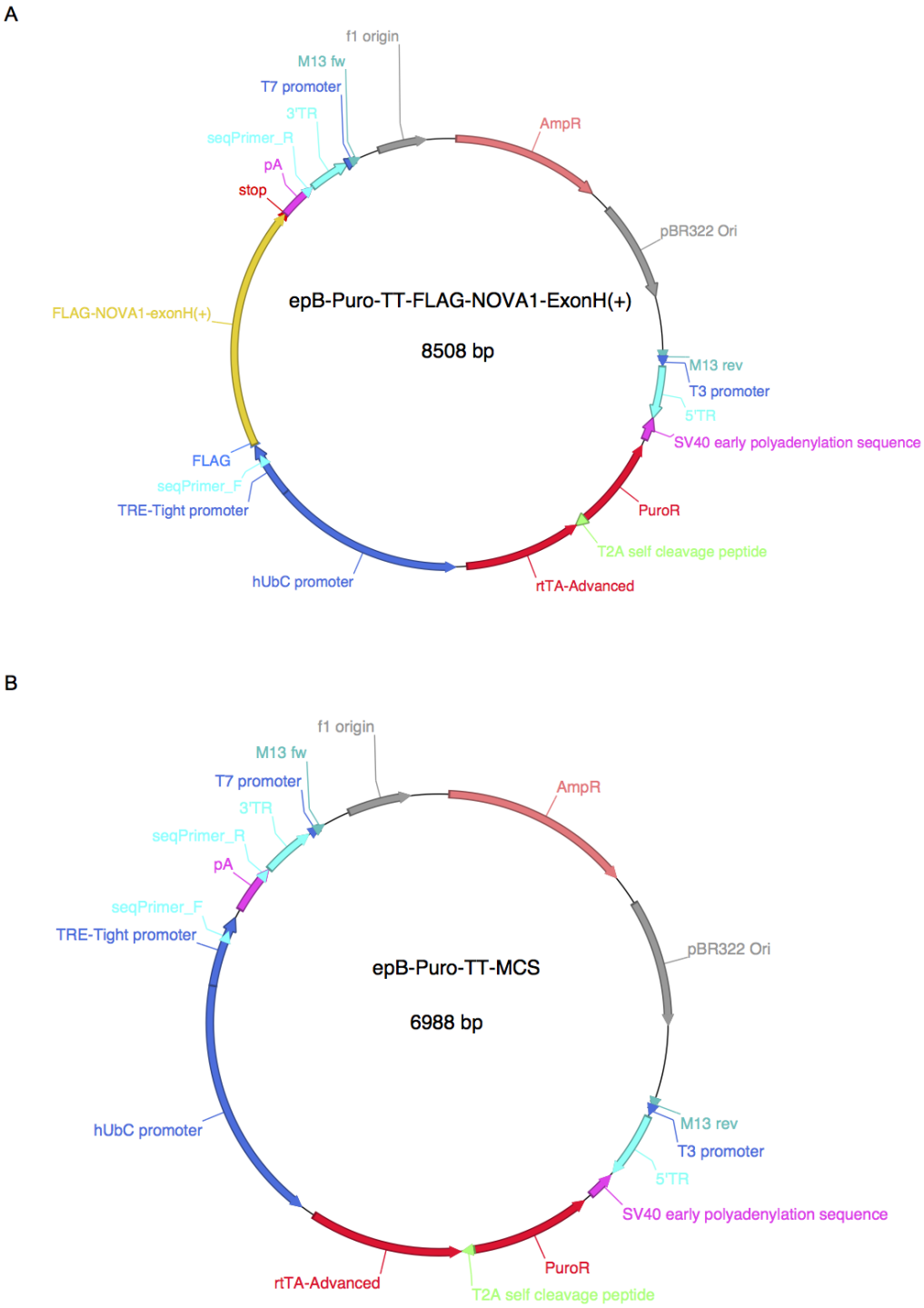


Figure 9-2: (A) ePB-Puro-TT-FLAG-NOVA1-ExonH(+) vector utilized to generate NOVA1 expressing Hela-Nova1(+) cell lines. **(B)** ePB-Puro-TT-MCS vector utilized to generate control Hela-Nova1(-) cell lines.

Western blotting

All gels were blotted using standard techniques and transferred onto nitrocellulose membrane (LI-COR). Membranes were blocked with a LI-COR blocking buffer and were probed with primary antibodies listed accordingly. IRDye conjugated secondary antibodies (LI-COR) were used at 1:10,000. Proteins were detected using the Odyssey CLx Imaging System (LI-COR).

Antibodies

Primary antibodies used for CLIP, immunofluorescent staining (IF), and western blots (WB) were: rabbit anti-Nova (Darnell lab, 15ug/IP); goat anti-NOVA2 (Santa Cruz, sc-10546, IF; 1/500, WB; 1/2,000, CLIP), rabbit anti-NOVA1 [EPR13847] (abcam, ab183024, IF: 1/1,1000), human anti-pan NOVA (anti-NOVA paraneoplastic human serum, WB: 1/5,000); mouse anti-GAPDH (Santa Cruz, sc-32233, WB: 1/1,000)

Bioinformatics & Statistical Analysis

CLIP read processing and alignment: CLIP reads were processed as described previously using CLIP Tool Kit (CTK) (Moore et al., 2014; Shah et al., 2017). Briefly, raw reads were filtered for quality and demultiplexed using indexes introduced during the reverse transcription reaction. PCR duplicates were collapsed and adapter sequences removed. Reads were mapped to the mm10 RefSeq genome using novo-align. Mapped reads were further collapsed for potential PCR duplicates by coordinates and taking into consideration the degenerate barcodes introduced during the reverse transcription. Only reads that were uniquely mapped in the appropriate direction were retained.

Differential Peak Analysis: To identify NOVA-mRNA binding sites disproportionally enriched in one of two compared cell lines, the total number of tags per cluster from each cell line were analyzed in a pair-wise fashion, normalized to the total number of tags in clusters in that cell line. Fold change and associated p- value for each cluster were calculated using scripts developed in-house by Chaolin Zhang for the Galaxy bioinformatics platform (Hillman-Jackson et al., 2012), and an FDR derived using the Benjamini-Hochberg correction for multiple hypothesis testing.

Nova normalization scores: Normalized CLIP scores were calculated for each biological replicate individually. Nova-CLIP tags mapped to peaks were normalized first for transcript length and then by sequencing depth (scaled to 1,000,000 tags) in order to generate a tpm-like value for each transcript. For each transcript the total number of reads which overlapped with a significant peak were summarised for each sample CLIP'd. Addition of a degenerate sequence in the RT primers enabled collapse of PCR duplicates. These steps result in the size of the libraries to have differing read depths relative to the number of CLIP tags isolated in the initial immunoprecipitation.

Functional Clustering of Nova target groups: Functional Module Detection using the HumanBase software was used to determine functional clusters of Nova targets per conservation group. Enriched GO terms were calculated by the software. (Krishnan et al., 2016)

Global analysis of target groups: GSEA analysis was performed with the fgsea R package. (Sergushichev, 2016) GO analysis was performed with the goseq R package. (Young et al., 2010) PCA analysis was performed using pcaExplorer. (Marini et al., 2020)

RNA-sequencing processing and alignment: RNA-sequencing reads were aligned to human genome hg38 using Salmon. DGE with batch correction was performed using DESeq2 for all described comparisons.

Differential splice analysis: Splicing analysis was performed using rMATs (Shen et al., 2014), considering both junction counts and exon coverage and the maser R package was used for visualization.

General statistics: Wilcoxon ranked sum test was applied for significance testing unless otherwise indicated, threshold of $P < 0.05$ or $Padj < 0.05$ to be considered statistically significant across all analysis.

All experimental data were analyzed using R version 3.6.3 and RStudio 1.1.463.

Data availability

All raw data, tables, bioinformatics code and analysis details for the research outlined in these chapters can be accessed: https://github.com/bmurph13/ThesisDarnell.

Appendix A: AGO HITS-CLIP reveals distinct miRNA regulation of white and brown adipose tissue identity

Contributions:

The following sections of research were designed, planned, and carried out in equal collaboration with Sean O'Connor, a graduate student in Paul Cohen's Laboratory of Molecular Metabolism at The Rockefeller University. These studies aim to elucidate miRNA regulatory networks and pathways modulating RNA steady state levels across fat depots with differential energetic capacities. The benchwork for CLIP experiments and sequencing of CLIP libraries were carried out by Elisabeth in the Darnell Laboratory. Data processing and bioinformatic analysis for all sequencing experiments (HITS-CLIP, RNA-sequencing, and microRNA-sequencing) were designed and carried out by Elisabeth in the Darnell Laboratory. All tissue collections and *in vitro* experiments were designed, performed, and analyzed by Sean O'Connor with the significant contributions of Cohen Laboratory members: Sarah Szwed and Francois Marchildon).

Abstract:

MicroRNAs (miRNAs) are short, noncoding RNAs that associate with Argonaute (AGO) to influence mRNA stability and translation, thereby regulating cellular determination and phenotype. While several individual miRNAs have been shown to control adipocyte function, including energy storage in white fat and energy dissipation in brown fat, a comprehensive analysis of miRNA activity in these tissues has not been performed. We used high-throughput sequencing of RNA isolated by cross-linking immunoprecipitation (HITS-CLIP) to comprehensively characterize the network of high-confidence, in vivo mRNA:miRNA

interactions across white and brown fat, revealing >20,000 unique AGO binding sites. When coupled with miRNA and mRNA sequencing, we found an inverse correlation between depot-enriched miRNAs and their targets. To illustrate the functionality of our HITS-CLIP data set in identifying specific miRNA:mRNA interactions, we show that miR-29 is a novel regulator of leptin, an adipocyte-derived hormone that coordinates food intake and energy homeostasis. Two independent miR-29 binding sites in the leptin 3′ UTR were validated using luciferase assays, and miR-29 gain and loss of function modulated leptin mRNA and protein secretion in primary adipocytes. This work represents the only experimentally generated miRNA targetome in adipose tissue and identifies multiple regulatory pathways that may specify the unique identities of white and brown fat.

Introduction:

Worldwide obesity rates have nearly tripled over the last four decades, and in the U.S. nearly 50% of adults are projected to be obese by 2030 (Ward et al. 2019). Obesity is characterized by excess white adipose tissue (WAT), which accumulates in either subcutaneous or visceral depots. While visceral obesity is strongly associated with a constellation of chronic diseases, including type 2 diabetes, cardiovascular disease, nonalcoholic fatty liver disease, and certain cancers, subcutaneous obesity is generally less deleterious (Neeland et al. 2019; Jayedi et al. 2020). Mammals also possess brown adipose tissue (BAT) (Rosen and Spiegelman 2014), which can dissipate energy via adaptive thermogenesis and may offset many of the risks associated with visceral adiposity. In mice, trans-plantation of BAT decreases insulin resistance, improves glucose tolerance, and lowers total fat mass (Stanford et al. 2013). Adrenergic

stimulation of BAT, such as by cold exposure, similarly improves glucose uptake, in-creases energy expenditure, and protects against diet-induced obesity in mice (Kajimura et al. 2015). Cold exposure additionally induces the activation of beige fat cells in subcutaneous WAT, which engage in thermogenesis and confer many of the same metabolic benefits as BAT (Wu et al. 2012). While BAT and beige fat may not be as abundant in humans, recent studies suggest these thermogenically active fat cells can be activated in response to cold or other stimuli (Cypess et al. 2009; van Marken Lichtenbelt et al. 2009; Virtanen et al. 2009), presenting an opportunity for therapeutic intervention.

Much of the work on adipocyte differentiation and phe-notype over the past few decades has focused on transcriptional regulators of cellular identity. These include Pparg2, which is both necessary and sufficient for the development and maintenance of white and brown adipocytes (Tontonoz et al. 1994; Tontonoz and Spiegelman 2008), and Prdm16, which regulates brown and beige adi-pocyte identity (Seale et al. 2007; Cohen et al. 2014). More recently, it has become clear that post-transcription-al regulators also play a key role in influencing adipocyte phenotype. Studies with adipose-specific dicer KO mice demonstrate that microRNAs (miRNAs) are essential in regulating adipogenesis, insulin sensitivity, and nonshivering thermogenesis (Mori et al. 2014; Reis et al. 2016). Several individual miRNAs have been shown to contribute to these phenotypes. For example, miR-133 represses BAT function by directly targeting Prdm16 and inhibiting expression of thermogenic genes (Trajkovski et al. 2012), miR-196a induces browning of WAT by targeting Hoxc8 (Mori et al. 2012), miR-155 regulates the differentiation of brown and beige adipocytes (Chen et al. 2013), and miR-26 protects from diet-induced obesity by blocking adipogenesis (Acharya et al. 2019). These studies, and most others in adipose tissue,

rely on computational predictions and low-throughput validation to identify miRNA targets. Although high-throughput cross-linking techniques to map the miRNA targetome have been applied to liver, brain, heart, and other organs (Chi et al. 2009; Spengler et al. 2016; Luna et al. 2017), thus far, no comprehensive targetome has been reported for adipose tissue.

Here, we performed miRNA sequencing to comprehensively profile miRNA levels across adipose depots, followed by HITS-CLIP to characterize miRNA binding behavior in iBAT and eWAT. These data are complemented by RNA sequencing to assess the effects of miRNA binding on steady-state mRNA levels. Our data provide a rich resource of miRNA expression coupled with in vivo miRNA targeting behavior, which we used to characterize the role of miR-29 in suppressing leptin production. Moreover, these studies provide a wealth of data for future investigations of miRNA regulatory interactions in white and brown fat.

Results:

Global profiling of microRNAs across fat depots

In order to broadly profile miRNAs across mouse adipose tissues, we sequenced small RNAs covering the full spectrum of adipocyte phenotype, ranging from white to brown fat. These included two visceral fat depots, epididymal WAT (eWAT) and mesenteric WAT (mWAT); two subcutaneous fat depots, inguinal WAT (iWAT) and axillary WAT (aWAT); and interscapular BAT (iBAT) (Figure 11-1A). We first compared the four WAT depots with iBAT and identified dozens of WAT-enriched miRNAs in each depot (Figure 11-2A). Global clustering analysis of the top 100 expressed miRNAs identified depot-specific clusters, with a particularly distinct profile observed in iBAT (Figure 11-1B). By comparison, the two subcutaneous fat depots, which demonstrate the greatest plasticity and have a molecular phenotype intermediate from the visceral/BAT extremes (Wu et al. 2012; Rosenwald et al. 2013), cluster together and most closely match the mean miRNA expression profile (Figure 11-1B). Given that many of the regulatory mechanisms governing subcutaneous fat function may only be revealed fol-lowing perturbation, we also sequenced small RNAs from iWAT of mice that had been cold-exposed (1 wk at 4°C) to stimulate beige fat biogenesis (cWAT). There were significant changes in the abundance of several miRNAs in response to cold exposure, including an in-crease in let-7 and miR-30 and a decrease in miR-125 and miR-27 (Figure 11-1C-D), all of which have been implicated in brown and beige adipocyte function (Fu et al. 2014; Sun and Trajkovski 2014; Giroud et al. 2016a,b) . To assess the effects of cold exposure on a global scale, we performed a principal component analysis (PCA) with all samples (Figure 11-1E). Cold exposure did not significantly change the

Figure 11-1 miRNA-seq identifies divergent miRNA profiles in eWAT and iBAT depots

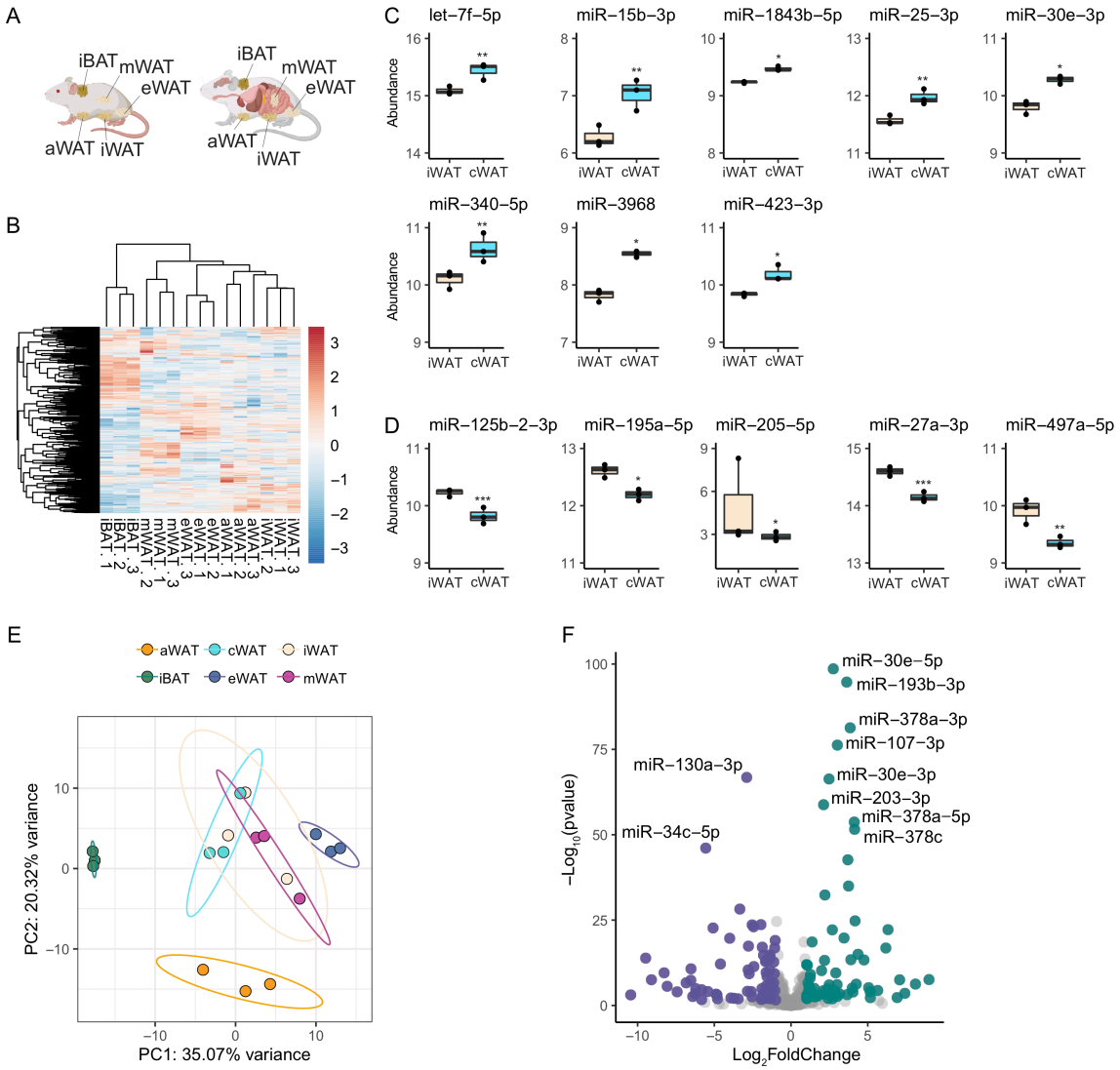


Figure 11-1: miRNA-seq identifies divergent miRNA profiles in eWAT and iBAT depots **(A)** Anatomical schematic of sampled adipose depots (left) and their locations relative to other organs (right) in the male mouse. **(B)** Global cluster analysis of miRNA-seq expression profile of top 100 expressed miRNAs clustered by adipose depot from 13-week old male mice housed at room temperature (3 mice pooled per group, n = 3 biological replicates per depot). **(C,D)** miRNAs with a significant increase **(C)** or decrease **(D)** in inguinal adipose tissue in response to cold exposure at 4ÅãC for 1 week prior to sacrifice (cold exposed WAT = cWAT). Expressed as normalized miRNA abundance (*p < 0.05, **p < 0.005, ***p < 0.0005; black horizontal bar is median, box shows 25th-75th percentile, and whiskers show max and min). **(E)** Principal component analysis of miRNA expression profiles by adipose depots defined in (A), including cWAT. Samples are color-coded based on depot. **(F)** Volcano plot of significantly differentially expressed miRNAs (DESEQ2) between iBAT and eWAT. Green = miRNAs enriched in iBAT relative to eWAT. Blue = miRNAs enriched in eWAT relative to iBAT. The most significantly differentially expressed miRNAs are highlighted.

miRNA profile of iWAT along the first two principal components (Figure 11-1E). In contrast, eWAT and iBAT samples clustered closely among biological replicates and diverged significantly along the first principal component, suggesting an important role for miRNAs in controlling phenotypic differences between these two types of adipose tissue. A volcano plot analysis shows that many of the miRNAs that differ most significantly between eWAT and iBAT have been shown to impact iBAT function or regulate adipocyte metabolism, including miR-30 (Hu et al. 2015), miR-193b (Sun et al. 2011), miR-34 (Fu et al. 2014), and miR-107 (Trajkovski et al. 2011).

Figure 11-2 Enriched miRNAs across white adipose depots

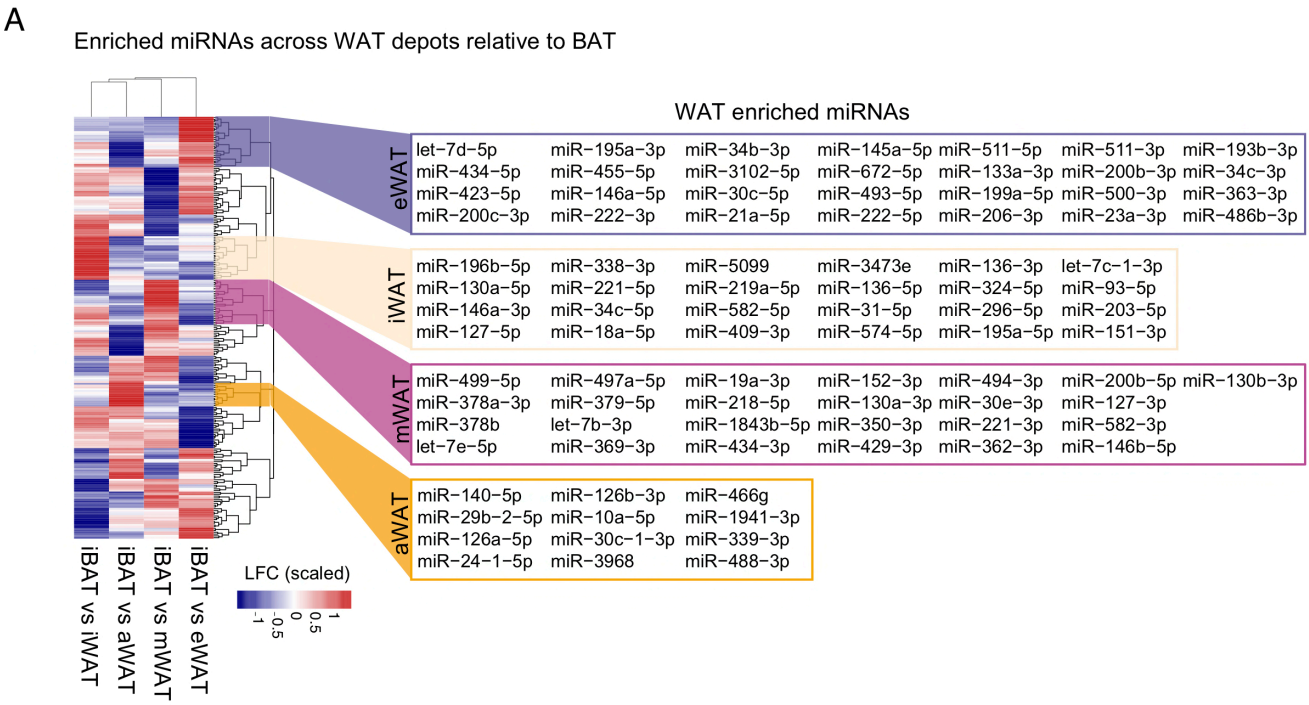


Figure 11-2: Enriched miRNAs across white adipose depots. **(A)** Differential analysis of miRNAs enriched across sampled WAT depots relative to iBAT (3 mice pooled per group, n = 3 biological replicates per depot). Heat map of significantly changed microRNAs for iBAT vs. all other depots. Values represent scaled log2FoldChanges.

Differential AGO regulation in iBAT and eWAT depots

As eWAT and iBAT are the two most divergent types of adipose tissue, both by phenotype and miRNA expression profiles, we examined in vivo miRNA targeting in these two tissues. To capture a broad, unbiased map of miRNA:mRNA interactions, we performed HITS-CLIP to biochemically identify AGO/miRNA regulatory sites in mRNAs in a high-throughput manner. Following UV cross-linking, AGO:RNA complexes were immunoprecipitated (Figure 11-3A), radiolabeled, and separated by gel electrophoresis. iBAT and eWAT samples contained a range of RNA fragment lengths, illustrated by a dense smear on the autoradiograph, compared with IgG and non cross-linked controls (Figure 11-3B). After excising AGO:RNA complexes of ~110–170 kDa and sequencing AGO-associated RNAs, iBAT and eWAT samples separated effectively by PCA (Figure 11-4A) and showed low inter-sample variability by Pearson correlation (Figure 11-3C). mRNA tags (sequencing reads corresponding to AGO-associated mRNA fragments) were clustered into significant peaks, revealing a total of 20,697 unique peaks across 5248 genes, with 14,562 unique iBAT peaks and 928 unique eWAT peaks, consistent with the greater number of iBAT unique RNA reads (Table 11-2).

Figure 11-3 miRNA binding sites are variable across iBAT and eWAT

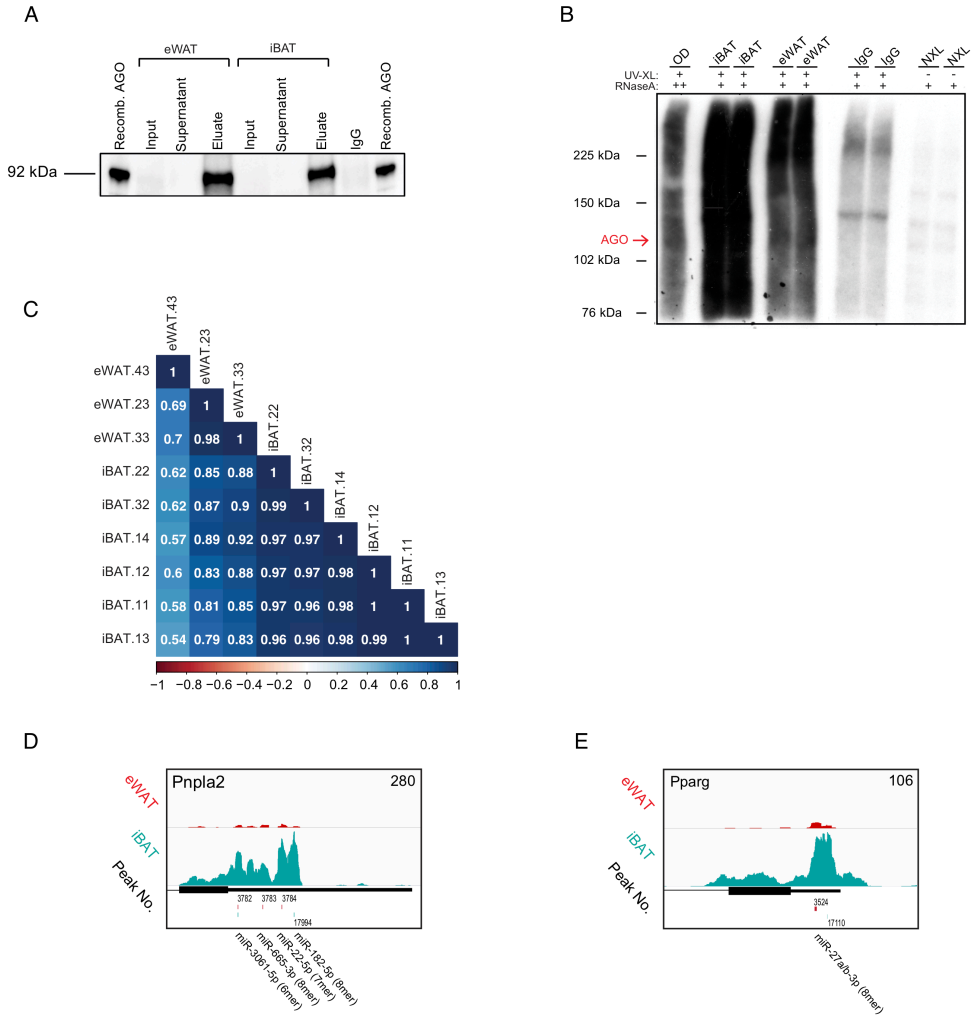


Figure 11-3: miRNA binding sites are variable across iBAT and eWAT. **(A)** Immunoprecipitation of AGO from eWAT and iBAT tissues (image cropped to display AGO band more clearly). **(B)** Autoradiogram showing NuPAGE separation of radio-labeled Ago-mRNA complexed, over digested control (OD), iBAT, eWAT, an IgG control and a no-crosslink control (NXL). Region excised: 110 – 170 kDa. **(C)** Pearson correlations for n = 6 groups for iBAT, n = 3 groups for eWAT, two mice pooled per group, samples clustered based on hierarchical clustering order. (D,E) Gene tracks of miRNAdependent differentially regulated targets Pnpla2 **(D)** and Pparg **(E)** showing HITS-CLIP tags plotted by genomic location. Purple = significant peaks called from aggregated reads, red = significant peaks called from eWAT reads only, teal = significant peaks called from iBAT reads only, scale for each gene track was normalized by the tallest peak.

A slightly higher percentage of iBAT-specific peaks (59%) localized to 3′ untranslated regions (UTRs) than eWAT-specific peaks (55%), and most of the remaining peaks localized to exons (Figure 11-4B). Next, we identified enriched KEGG pathways among genes with increased AGO binding in iBAT and eWAT (Figure 11-4C), revealing that several genes involved in thermogenesis, including Pparg and Pnpla2, show in-creased AGO binding in iBAT. miR-27, a known suppressor of beige fat function is among the miRNAs targeting an iBAT-specific peak in Pparg (Sun and Trajkovski 2014), while miR-182, a highly iBAT-enriched miRNA, targets an iBAT-specific peak in Pnpla2 (Figure 11-3 D-E; the complete set of miRNA seed-to-peak matches used for analysis in this study is in Table 11-4 and is based on a valley depth of 0.9 for peak disaggregation; as an additional resource, Table 11-5 shows a less stringent analysis using a valley depth of 0.5).

Next, to identify peaks with depot-specific regulation while accounting for differences in total AGO binding, peaks enriched in iBAT or eWAT were subsetted from genes with equal levels of total tags, resulting in 226 genes with eWAT-specific peaks, 122 genes with iBAT-specific peaks, and 42 genes with both iBAT- and eWAT-specific peaks (Figure 11-4D; Table 11-3). Among these were several genes known to influence adipocyte metabolism, including Pik3r1 (Figure 11-4E) and Irs1 (Figure 11-4F), both of which are involved in insulin signal transduction. Interestingly, miR-10b-5p is a highly expressed eWAT-enriched miRNA that binds to peak 1293 in Irs1, suggesting a potential role in promoting insulin resistance in visceral fat.

Figure 11-4 HITS-CLIP identifies differential AGO regulation in iBAT and eWAT depots

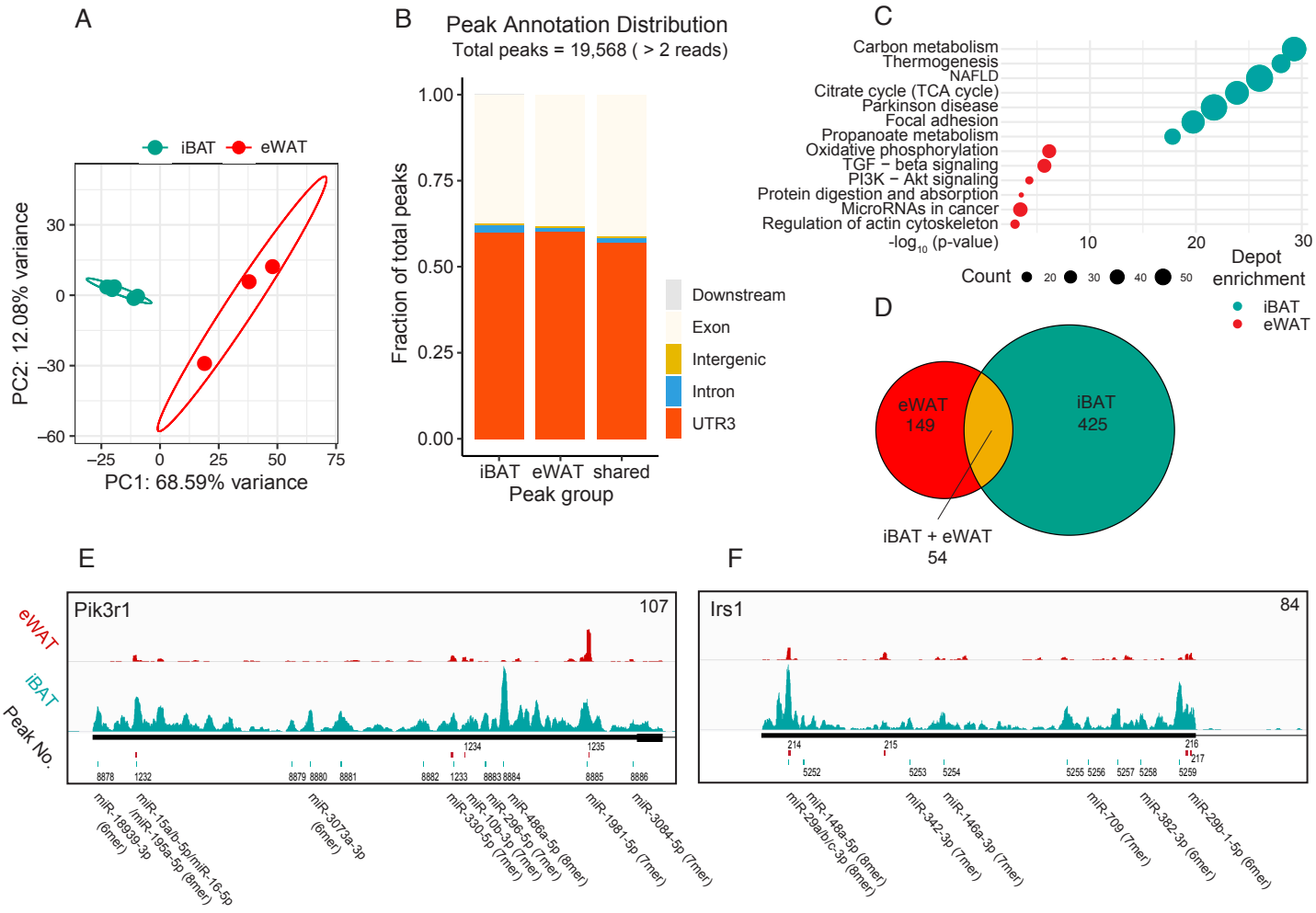


Figure 11-4: HITS-CLIP identifies differential AGO regulation in iBAT and eWAT depots. **(A)** PCA analysis of HITS-CLIP peaks from iBAT and eWAT (15-week old male mice, 2 mice pooled per group, n = 6 groups for BAT, n = 3 groups for eWAT). **(B)** Genomic location of depot specific HITS-CLIP peak distribution by annotated region and depot, expressed as fraction of total peaks; “Shared” indicates that peak is significant in both iBAT and eWAT, "All" indicates that peak is significant when iBAT and eWAT reads are pooled before scanning for peaks. **(C)** KEGG gene-set enrichment analysis of genes with iBAT and eWAT specific peaks (significant differential peak defined as |Log2FC| >1 and padj < 0.05). **(D)** Venn diagram depicting differential peak usage across depots. 122 genes contain iBAT but no eWAT specific peaks, 226 genes contain eWAT but no iBAT specific peaks, and 42 genes contain both iBAT and eWAT specific peaks. **(E,F)** Gene tracks of miRNA-dependent differentially regulated targets Pik3r1 **(E)** and Irs1 **(F)** showing 3’UTR localized HITS-CLIP tags. Purple = significant peaks called from aggregated reads, red = significant peaks called from eWAT reads only, teal = significant peaks called from iBAT reads only; scale for each gene track was normalized by the tallest peak.

CLIP peak-score elucidates AGO-driven changes in RNA steady-state levels across iBAT and eWAT depots

AGO regulation is just one way in which a cell can control gene expression in the setting of a dynamically changing transcriptional profile. To account for changes in the transcriptome, we normalized CLIP tags for transcript abundance, allowing AGO targets to be ranked by relative binding. We first performed RNA sequencing of eWAT and iBAT to measure transcript abundance for each depot. Next, we calculated a "CLIP score" to define AGO bind-ing affinity at each peak site based on both cross-linked events determined via CLIP and transcript abundance determined via RNA-seq, as described previously (Sawicka et al. 2019). We summarized CLIP tag density for each peak (per replicate) and correlated it with transcript abundance (Figure 11-5A). We then used this to generate a "CLIP peak score" for each binding site, which quantifies AGO binding relative to all other transcripts of comparable abundance (Table 11-3). For each CLIP replicate, a linear regression was fitted through a plot of CLIP tags versus RNA-seq tags per million (TPMs), and the position of each peak was determined relative to the fitted line. For a given peak, the greater the distance above the regression line, the higher the CLIP peak score and the greater the inferred AGO binding.

Figure 11-5 miR-29a modulates leptin levels in vitro

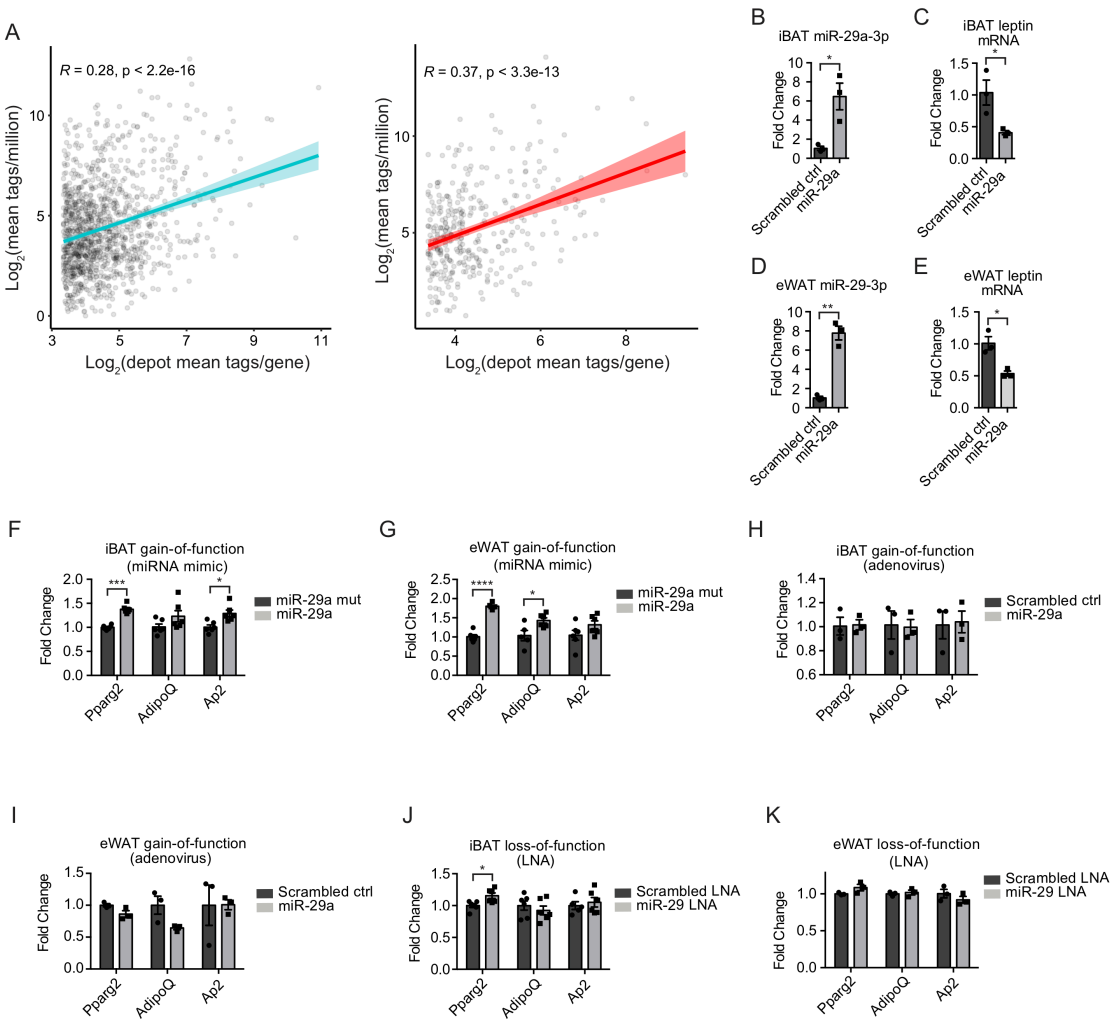


Figure 11-5: miR-29a modulates leptin levels in vitro. **(A)** CLIP tags per gene are plotted over the normalized gene expression value for each depot (blue = iBAT and red = eWAT) **(B, C)** iBAT miR29a and leptin mRNA in primary adipocyte cultures transduced with adenovirus containing primary miR29a or a scrambled sequence (n = 3). **(D, E)** eWAT miR-29a and leptin mRNA in primary adipocyte cultures transduced with adenovirus containing primary miR-29a or a scrambled sequence (n = 3). **(F-K)** qPCR of differentiation markers in primary iBAT and eWAT adipocyte cultures transfected with miR-29a mimic (F, G), transduced with miR-29a adenovirus (F, G) or transfected with miR-29 LNA (J,K; n = 3 for F, G and H, n = 6 for I, J, and K). All results are presented as mean ±SEM and Student's t test used for all two-way comparisons (*p < 0.05, **p < 0.005, ****p < 0.00005).

Our study aims to characterize differential AGO regulation across iBAT and eWAT depots. Therefore, we asked how changes in AGO binding compare with changes in mRNA abundance across these two depots. When we examined the relationship between CLIP peak score and mRNA abundance, we found that many of the high affinity iBAT binding sites are in genes that are more highly expressed in iBAT relative to eWAT (Figure 11-6A, quadrant I). Likewise, a portion of high-affinity eWAT binding sites are localized to mRNAs that are more highly expressed in eWAT relative to iBAT (Figure 11-6A, quadrant III). These rep-resent depot-specific AGO binding sites likely linked with transcriptional regulation. In addition to identifying AGO-regulated peaks with depot-specificity, we identified a subset of regulatory sites that display an increase in AGO affinity in eWAT despite higher mRNA levels in iBAT (Figure 11-6A, quadrant IV), as well as peaks with an increase in AGO affinity in iBAT despite higher mRNA levels in eWAT (Figure 11-6A, quadrant II; Table 11-3).

Figure 11-6 CLIP peak-score for predicting AGO-driven changes in RNA steady state levels

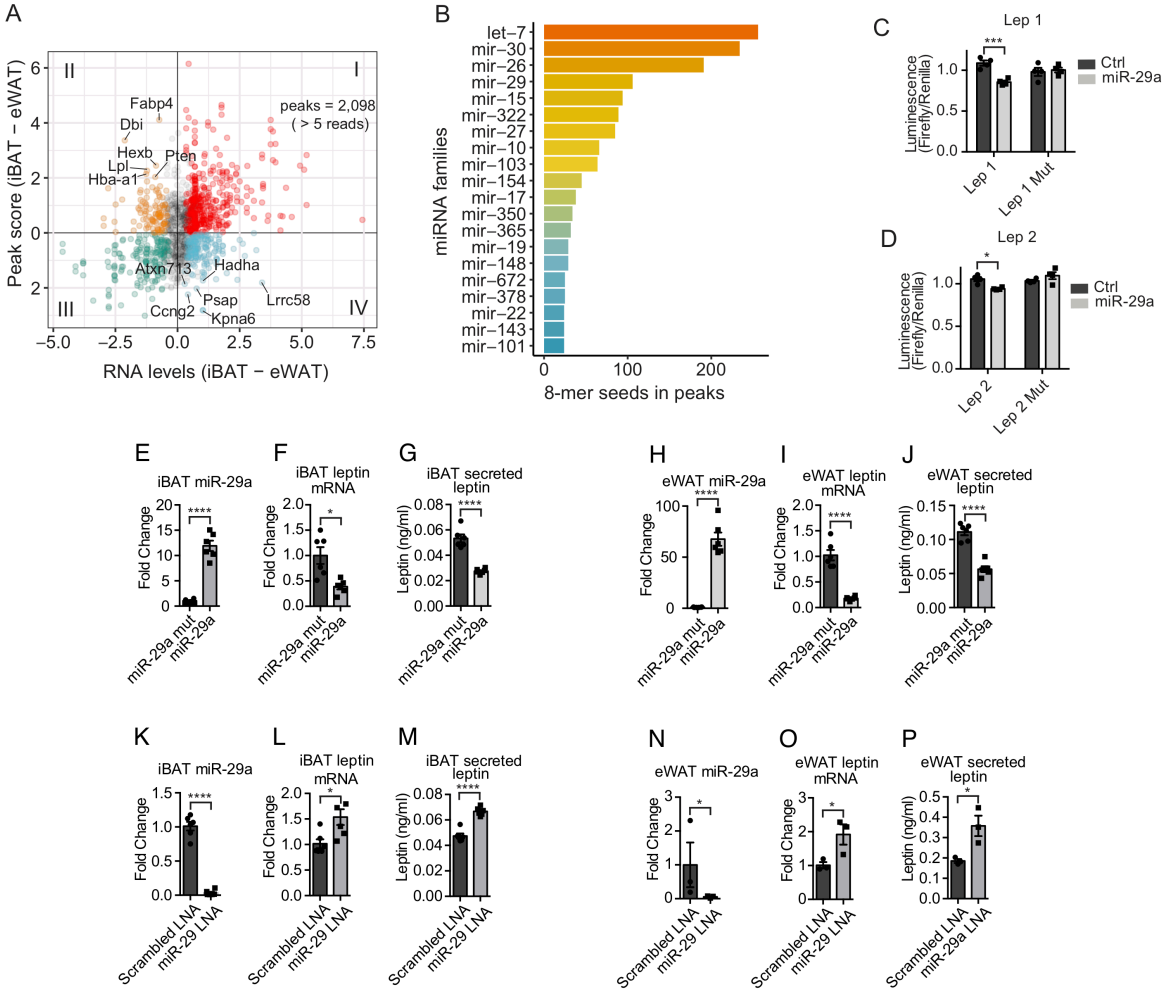


Figure 11-6: CLIP peak-score for predicting AGO-driven changes in RNA steady state levels. (A) Difference in mRNA abundance (iBAT- eWAT mRNA) plotted against difference in CLIP peak score (iBAT - eWAT) for AGO-binding sites (quadrant I: mRNA enriched in iBAT and a higher iBAT CLIP score than eWAT, quadrant II: mRNA enriched in eWAT and a higher iBAT CLIP score than eWAT, quadrant III: mRNA enriched in eWAT and a higher eWAT CLIP score than iBAT, quadrant IV: mRNA enriched in iBAT and a higher eWAT CLIP score than iBAT). (B) Bar plot of top 20 miRNA families with 8-mer miRNA seed matched sites in Ago peaks. (C, D) Dual luciferase assay validation conducted by co-transfecting HEK- 293A cells with plasmid containing luciferase enzymes and either a scrambled control miRNA or miR-29a. Mut = luciferase plasmid containing point mutations in seed sequence binding site. Lep 1 and Lep 2 refer to the miR-29 binding sites in leptin. Luciferase/renilla luminescence is normalized to scrambled control transfected cells and presented as a fold change (n=4). (E-G) iBAT miR-29a, leptin mRNA, and conditioned media (CM) leptin concentration from primary adipocyte cultures transfected with a miR-29a mimic or a mutated miR-29 mimic (n = 6). (H-J) eWAT miR-29a, leptin mRNA, and conditioned media (CM) leptin concentration from primary adipocyte cultures transfected with a miR-29a mimic or a mutated miR-29 mimic (n = 6). (K-M) iBAT miR-29a, leptin mRNA, and CM leptin concentration in primary adipocyte cultures transfected with a

miR-29 LNA or a scrambled LNA (n = 6). (N-P) eWAT miR-29a, leptin mRNA, and CM leptin concentration in primary adipocyte cultures transfected with a miR-29 LNA or a scrambled LNA (n = 3). Results for C-P are presented as mean Å}SEM. 2-way ANOVA used for miR-29a to control comparisons in C and D and student's t tests used for comparisons in E-P (*p < 0.05, ***p < 0.0005, ****p < 0.00005).

To demonstrate the utility of our HITS-CLIP data set in exploring specific miRNA:mRNA interactions, we first identified miRNAs most responsible for regulating gene expression in adipose tissue by ranking candidates by high-confidence seed matches (8-mer binding sites within the top 1000 peaks) (Figure 11-6B). Many of the top ranked candidates have been thoroughly studied in adipocytes, including let-7 and miR-30 (Ventayol et al. 2014; Hu et al. 2015; Giroud et al. 2016a; Koh et al. 2018). For this reason, we focused on miR-29, an abundantly expressed miRNA that has been described previously as a regulator of the glucocorticoid receptor (GR) in adipocytes (Glantschnig et al. 2019) but has otherwise not been well studied in fat cells (our data show a miR-29 binding site in the iBAT GR, but not eWAT). Among miR-29's top binding sites, as ranked by peak score, was a peak localized to the leptin 3′ UTR (Table 11-4). Additionally, a second sequence complementary to the miR-29 seed was identified in the leptin 3′ UTR, distal to the first. To validate these sites, miR-29a mimics were co-transfected into HEK-293A cells with a dual luciferase plasmid containing miRNA binding sites cloned into the 3′ UTR of firefly luciferase. Normalized luminescence was significantly decreased with miR-29a, but not a scrambled control, and mutations introduced specifically within the miR-29-3p seed sequence binding sites were sufficient to block this effect (Figure 11-6C,D).

To assess the function of these binding sites, we transfected primary adipocytes with a miR-29a mimic. Seven-ty-two hours post-transfection, miR-29a-3p levels were substantially increased in both brown and white adipocytes (Figure 11-6E,H). Leptin mRNA levels correspondingly decreased by 40% in brown adipocytes ($P = 0.0054$) and 80%in white adipocytes ($P < 0.00001$), relative to cells transfected with a mutated miR-29a mimic (Figure 11-6F,I). Leptin secretion from primary adipocytes was lower than that from adipocytes in vivo; however, the conditioned media (CM) of primary fat cells contained a sufficient amount of leptin to be detected by ELISA. We observed a 50%decrease in leptin protein in CM from both brown and white adipocytes (Figure 11-6G,J). As an alternative gain-of-function approach, primary cells were transduced with an adenovirus expressing miR-29a or a scrambled miRNA control. Consistent with the miRNA mimic experiments, an increase in miR-29a caused a 50%–60% decrease in leptin mRNA (Figure 11-5B,E). Next, primary cells were transfected with a locked nucleic acid (LNA) to knock down endogenous miR-29a-3p. Both primary iBAT and eWAT adipocytes showed a significant decrease in miR-29a-3p levels (Figure 11-6K,N) and corresponding in-creases in leptin mRNA of 55% ($P = 0.015$) and 90% ($P = 0.044$), respectively (Figure 11-6L,O). In CM from both brown and white adipocytes, secreted leptin protein significantly increased in parallel with leptin mRNA (Figure 11-6M,P). For each experiment, RNA levels of the general adipocyte differentiation markers Pparg2, Adipoq, and Ap2 were measured. The expression of these differentiation-dependent genes was modestly increased in adipocytes transfected with miRNA mimics (Figure 11-5F,G), consistent with previous reports that miR-29 promotes adipogenesis in human cells (Zhang et al. 2016). Importantly, leptin levels decreased in mimic-transfected cells despite the increase in markers of differentiation.

Adenoviral gain-of-function and LNA-based loss of function did not alter markers of adipocyte differentiation (Figure 11-5H–K).

In obese mice and humans, leptin mRNA and circulating protein are generally increased compared with lean controls, in proportion to body fat mass (Frederich et al. 1995; Maffei et al. 1995). We hypothesized that miR-29 levels would be inversely correlated with adiposity in models of obesity. Mice fed a HFD for 15 wk were 60%heavier than chow-fed controls (Figure 11-7A) and had a fourfold increase in leptin mRNA in iBAT (basal levels of leptin in iBAT are very low), a 12-fold increase in eWAT, and an eightfold increase in serum leptin (Figure 11-7B,C). miR-29a-3p levels were correspondingly reduced by ~30% in iBAT ($P = 0.027$) and 60% in eWAT ($P < 0.00001$) (Figure 11-7D). To confirm our findings, we next measured miR-29a abundance in a genetic model of obesity. Ob/ob mice were hyperphagic and obese relative to littermate controls (Figure 11-7E) due to a nonsense mutation in the leptin coding se-quence that led to high levels of leptin mRNA, but no functional protein (Figure 11-7F,G). miR-29a in iBAT and eWAT of these mice was reduced by 24% ($P = 0.017$) and 31% ($P = 0.011$), respectively (Figure 11-7H). We also considered the possibility that circulating leptin levels might regulate miR-29a expression. To that end, 60-d-old ob/ob mice were treated with recombinant leptin (300 ng/h via osmotic mini pump) for 2 d. To control for leptin's effects on food intake, a group of ob/ob mice was treated with PBS and pair-fed to the leptin-treated group. After 2 d, both groups of animals lost a comparable amount of weight (Figure 11-8A,B). We found no significant change in miR-29a abundance between these groups in iBAT or eWAT (Figure 11-7I,J).

Figure 11-7 miR-29a and leptin levels are inversely correlated in obese mice

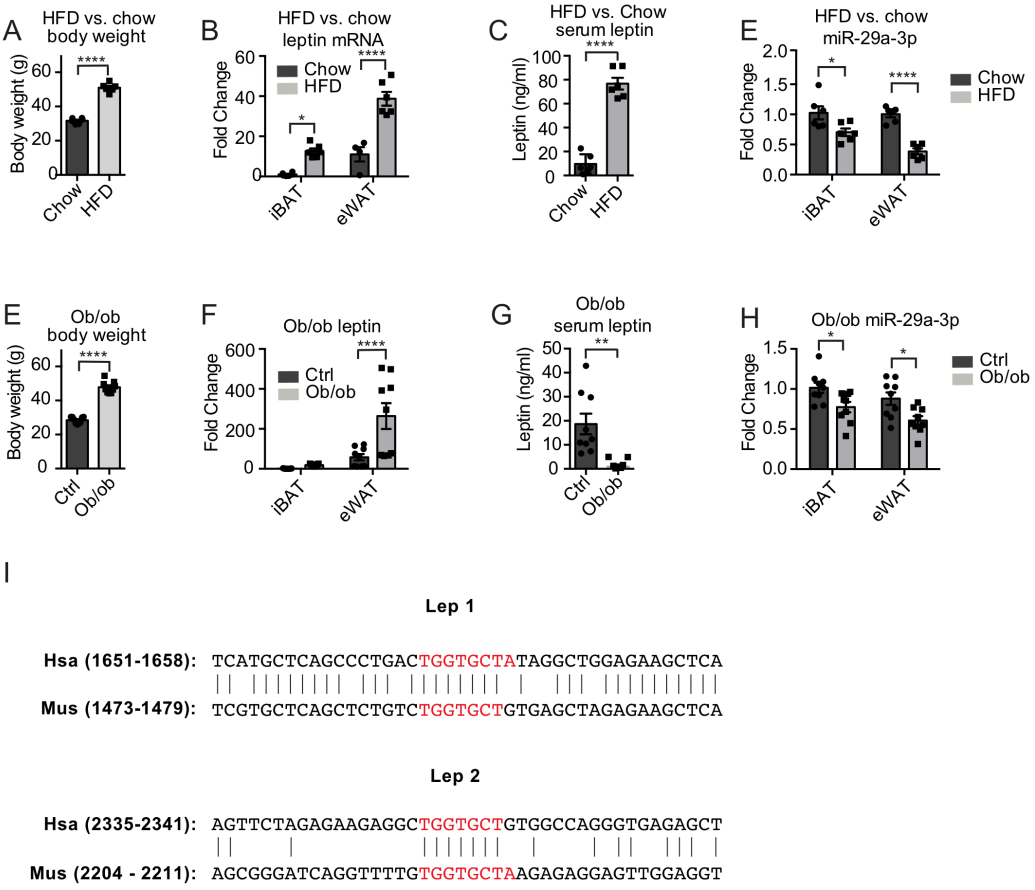


Figure 11-7: miR-29a and leptin levels are inversely correlated in obese mice. (A-D), 21-week old HFD vs. chow mice (n = 6). Body weights **(A)**, leptin mRNA **(B)**, serum leptin **(C)** and miR-29a levels **(D)**. (E-H) 9-week old ob-/- vs. ob+/- (n = 9), body weights **(E)**, leptin mRNA **(F)**, serum leptin **(G)** and miR-29a levels **(H)**. (I,J) miR-29a levels in iBAT **(I)** and eWAT **(J)** from ob/ob mice treated with leptin or saline (PBS) and pair-fed to mice treated with recombinant leptin (n = 3 for PBS treatment, n = 4 for leptin treatment) **(K)** miR-29 binding seed sites (red) within the leptin 3'-UTR in mouse and human; number in parentheses indicates location from start of 3'-UTR. All results are presented as mean ±SEM. 2-way ANOVA used for leptin mRNA and miR-29a comparisons in B, E, F and H and Student's t test used for all other comparisons (*p < 0.05, **p < 0.005, ****p < 0.00005).

Human miR-29a-3p is identical in sequence to murine miR-29a-3p, and human leptin contains two TargetScan predicted binding sites (Agarwal et al. 2015) that are moderately well conserved with those in mice (Figure 11-7K). If miR-29a-3p serves as a physiologically relevant regulator of leptin in humans, our findings suggest that miR-29a-3p should be reduced in obese patients. Indeed, previously published work demonstrated that in a cohort of 66 men and women, ranging from lean to obese, miR-29a-3p in visceral and subcutaneous WAT is negatively correlated with BMI (Glantschnig et al. 2019). A second human study found an increase in miR-29a-3p in subcutaneous WAT from 19 obese individuals following diet- and exercise-induced weight loss (Kristensen et al. 2017).

Figure 11-8 Body weight and food intake of leptin treated ob/ob mice

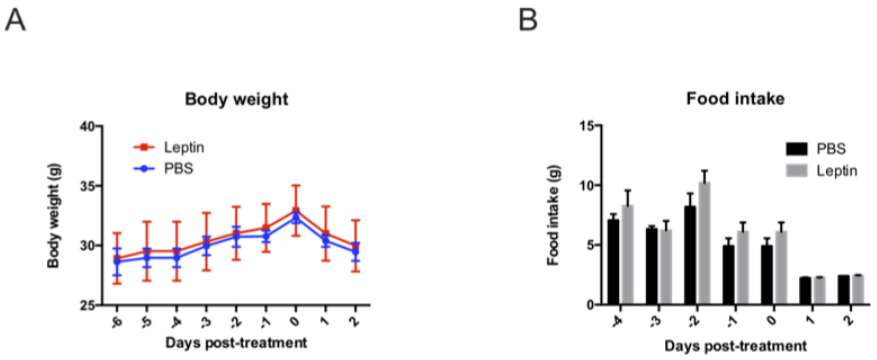


Figure 11-8: Body weight and food intake of leptin treated ob/ob mice. A separate group of animals was treated with saline (PBS) and pair-fed to the leptin treated animals. **(A)** Body weight and **(B)** food intake of mice treated with 300 ng/hr recombinant leptin beginning on day 0 (n = 3 for PBS treatment, n = 4 for leptin treatment). All results are presented as mean ±SEM.

Discussion:

In conclusion, we have conducted the first comprehensive biochemical mapping of miRNA:mRNA binding in white and brown adipose tissue, providing an easily accessible resource for future studies. The stark phenotypic differences between iBAT and eWAT are mirrored by highly divergent miRNA expression profiles (Figure 11-1E,F) and by AGO/miRNA regulatory behavior (Figure 11-4C,D). iBAT-specific peaks indicate a role for miRNAs in controlling the citric acid cycle and thermogenesis, two pathways known to display significant heterogeneity across adipose depots (Rosen and Spiegelman 2014), suggesting that miRNAs may be important contributors to these phenotypic differences. Similarly, many eWAT-enriched peaks belong to the PI3K-Akt and insulin signaling pathways, consistent with eWAT being more prone to exhibiting insulin resistance (Hardy et al. 2012).

To further increase the utility of our data, we combined HITS-CLIP with mRNA-seq data to generate peak scores, which reflect AGO:mRNA affinity and al-low for a ranking of miRNA targets (Figure 11-6A). Comparisons between iBAT and eWAT revealed a number of peaks with peak scores that were the inverse of mRNA steady-state levels, indicative of depot-specific regulation (Figure 11-6A). For example, our data show that Lpl, an eWAT-enriched gene that plays a central role in lipolysis, is more strongly bound by AGO in iBAT. In total, the 1079 peaks with inverse peak score/mRNA levels provide a rich source of previously unexplored, depot-specific interactions (Table 11-3) that will serve as a valuable tool for future investigations into adipose tissue biology.

Finally, we focused our attention on the miR-29:leptin interaction to illustrate the functionality of the HITS-CLIP data set. While lncRNAs have been shown to modulate leptin expression (Lo et al. 2018; Dallner et al. 2019), much of leptin's regulatory control remains unexplored. Our data describe miR-29 as a novel post-transcriptional regulator of leptin. These findings complement the previously described role of miR-29 in controlling mammalian metabolism, acting in the liver, heart, pancreas, and other organs (Pullen et al. 2011; Kurtz et al. 2015; Caravia et al. 2018). In addition to the leptin-miR-29 interaction, our data provide thousands of high-affinity miRNA: mRNA associations, demonstrating the power of HITS-CLIP in identifying novel regulatory relationships in adipose tissue and providing a valuable resource for future discovery.

Methods:

Mouse experiments

Animal studies were performed according to procedures approved by the Institutional Animal Care and Use Committee at Rockefeller University. Unless stated otherwise, all mice were males purchased from The Jackson Laboratory and on the C57BL/6 background, including ob/ob (stock no. 000632), WT (stock no. 000664), and high-fat diet (HFD)-fed mice (stock no. 380050). Controls for ob/ob mice were heterozygous littermates. Mice were group housed at five mice per cage (four for HFD) and maintained on 12-h light:dark cycles at 23°C. HFD mice were fed a diet composed of 60% dietary fat (Research Diets D12492) from 6 wk of age. All other animals were maintained on a chow diet, ad libitum. Mice chronically exposed to cold were individually housed at 4°C and closely monitored by laboratory personnel. For leptin experiments, 8.5-wk ob/ob singly housed female mice were treated with recombinant leptin (R&D 498OB05M) at 300 ng/h (ALZET osmotic mini pump #2002) for 2 d (n = 4). The control group (n = 3) of ob/ob mice was treated with PBS and pair-fed to the leptin-treated group.

CLIP sequencing

iBAT and eWAT tissues were dissected from 15-wk-old male C57BL/6J mice and immediately flash-frozen in liquid nitrogen. Frozen iBAT and eWAT pairs were then pooled into groups of two, ground with a tissue pulverizer (Cellcrusher), and cooled in liquid nitrogen. Frozen samples were spread to a thin layer on Petri dishes precooled on dry ice and UV cross-linked three times at 400 mJ/cm2 using a Stratalinker 2400 (Stratagene). All samples were stored at −80°C until tissue lysis.

For each sample, 400 μL of protein A Dynabeads (Invitrogen) was pooled, then prepared by washing three times with 1 mL of antibody binding (AB) buffer (AB; PBS, 0.02% Tween). Following the final wash, beads were resuspended in 400 μL of AB per sample, 50 μL of rabbit anti-mouse bridging antibody (Jackson ImmunoResearch 315005008) was added, and beads were rotated for 1 h at room temperature. Beads were then washed three times with 1 mL of AB, resuspended in AB, and incubated for 2 h at room temperature with either 12 μL of pan-Ago antibody (clone A28)(Moore et al. 2014), or 12 μL of mouse normal IgG (Santa Cruz Biotechnology sc-2343).

Tissue pellets were lysed with 800 μL of lysis buffer (PXL[1× PBS, 0.1% SDS, 0.5% Na-DOC, 0.5% NP-40], RNase inhibitor [Promega N2615, one Complete EDTA-free protease inhibitor cocktail tablet [Roche 11873580001] per 10 mL) and subjected to DNase treatment (24407355), using 30 μL of DNase (Promega) and incubated by shaking at 1100 rpm for 5 min at 37°C. Samples were then centrifuged at 20,000g for 20 min, and a needle was used to withdraw 800 μL of aqueous protein lysate while avoiding the lipid layer and pellet. Lysates were then treated with 20 U/mL RNase A (Affymetrix), first diluted to the indicated concentration by volume (e.g., 1:100 for "high RNase," or 1:10,000 for "low RNase") in lysis buffer and then added at 8 μL/mL of lysate. Then samples were digested by shaking at 1100 rpm for 5 min at 37°C. Following RNase digestion, samples were added directly to prepared antibody-bound beads and rotated for 3 h at 4°C.

After IP, the following washes were performed: twice with lysis buffer, twice with high-salt lysis buffer (5× PBS, 1% Igepal, 0.5%deoxycholate, 0.1% SDS), twice with stringent wash buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% NaDOC, 0.1% SDS, 120 mM NaCl, 25 mM KCl), twice with high-salt wash buffer (15 mM Tris at pH 7.5, 5 mM EDTA, 2.5 mM EGTA, 1% Triton X-100, 1% NaDOC, 0.1% SDS, 1 M NaCl), twice with low-salt wash buffer (15 mM Tris at pH 7.5, 5 mM EDTA), and twice with PNK wash buffer (50 mM Tris at pH 7.4, 10 mM MgCl2, 0.5% NP-40). The second of each wash was rotated for 2–3 min at room temperature. Tags were dephosphorylated as described (Moore et al. 2014) and subjected to over-night 3′ ligation at 16°C with a preadenylated linker (Moore et al. 2018) with the following ligation reaction: 2 μL of 25 uM linker, 2 μL of T4 RNA ligase 2, truncated K227Q (NEB), 1× ligation buffer (supplied with ligase), 2 μL of superasin RNase inhibitor (Promega), and 8 μL of PEG8000 (supplied with ligase). The resulting beads were washed, 32P-labeled, and subjected to SDS-PAGE and transferred as described (Moore et al. 2014).

Tags were collected from nitrocellulose using phenol:chloro-form:IAA (25:24:1 at pH 6.6) and tags were precipitated with a standard NaOAC precipitation. Cloning was performed using the BrdU-CLIP protocol as described (Moore et al. 2018) with a few exceptions. Briefly, the RT primer contained a 14-nt degenerate linker (a 3-nt degenerate sequence, a 4-nt multiplexing index, and a 7-nt unique molecular identifier), a 5′ linker for PCR amplification, a spacer to prevent rolling circle amplification after circularization, and the reverse complementary sequence of the 3′ linker for reverse transcription. BrdUTP-labeled cDNA was specifically isolated via two sequential Brd-UTP immunoprecipitations (with Abcam AB8955) and

circularized with CircLigase II (Epicenter CL9025K). Ago-CLIP libraries were sequenced by MiSeq (Illumina) to obtain 75-nt single-end reads.

CLIP data analysis (CTK)

Sequencing files were demultiplexed (barcode.txt), linkers were trimmed, and reads were collapsed. Reads were aligned to mm10 genome assembly using Novoalign. All raw CLIP read processing and significant peak calling were performed using the CTK software package (Shah et al. 2017). CLIP peaks and CIMS analysis was carried out as previously described (Moore et al. 2014).

CLIP data analysis (in R)

CLIP peak and unique tag output files were analyzed in R (Chan 2018) using Bioconductor (Gentleman et al. 2004).

RNA sequencing

For miRNA-seq, total RNA was extracted from fat pads of 13-wk-old C57BL/6J male mice housed at room temperature and from inguinal fat pads of 13-wk-old mice housed for 1 wk at 4°C prior to sacrifice. Following extraction, RNAwas pooled into three groups of three mice for iWAT, eWAT, iBAT, aWAT, and cWAT fat (mWAT was not pooled). The TruSeq small RNA library prepara-tion kit (Illumina) was used to prepare libraries for sequencing, which was performed by LC Sciences on an Illumina Hiseq platform.

For mRNA-seq, iBAT and eWAT tissue was dissected from 15-wk-old C57BL/6J male mice and immediately flash-frozen in liq-uid nitrogen (dissected from the same cohort of mice used for

HITS-CLIP). RNA was extracted and pooled into groups of two, a library was prepared using the Illumina TruSeq stranded total RNA library preparation kit, and sequencing was performed on an Illumina NextSeq 500.

Primary adipocyte culture

Primary adipocytes were isolated from 6-wk-old male C57BL/6J eWAT and iBAT adipose depots. Dissected adipose depots were incubated in 10 mL of digestion buffer for 16 min (eWAT) or 40 min (iBAT) in a 37°C shaking water bath (140 rpm). iBAT digestion buffer was prepared as follows: 15 mg of collagenase B (Roche 11088831001) was dissolved in 5 mL of 2× master mix consisting of 125 mM NaCl, 5 mM KCl, 1.3 mM CaCl2, 5 mM glucose, 1%penicillin-streptomycin, 4% BSA in H2O, and 5 mL of PBS. eWAT digestion buffer was prepared as follows: 100 mg of collagenase D (Roche 11088882001) was dissolved in a solution of 10 mL of PBS, 200 μL of 120 mg/mL dispase II (Roche 04942078001), and 40 μL of 2.5 M CaCl2. Following digestion, the stromal vascular fraction was isolated by centrifugation and filtration at 100 μm and 70 μm, then plated in 12-well collagen-coated plates.

All primary adipocyte cultures were maintained at 37°C with 10% CO2. Prior to differentiation, iBAT cells were cultured in GlutaMAX media (Gibco 10565) supplemented with 10% FBS and 1% penicillin-streptomycin and eWAT cells were cultured in ITS media (ITS media: 56.65% 1 g/L glucose DMEM [Gibco11885-084], 37.65% 1× MCDB210 at pH 7.25 [Sigma M6770], 2% FBS, 2% ITS premix [Corning 354352], 1% 10 mM L-ascorbic acid 2-phosphate in DMEM, 0.01% 100 μg/mL bFGF BD Bio 13256-029], 0.5% penicillin-streptomycin, 0.2% primo-cin). Following induction of differentiation, iBAT cells and eWAT cells were cultured in GlutaMAX media.

When iBAT preadipocytes reached 100% confluency and eWAT preadipocytes reached 95% confluency, differentiation was induced using a standard cocktail of 0.5 mM IBMX, 1 mM dexamethasone, 850 nM insulin, and 1 mM rosiglitazone dis-solved in GlutaMAX media. After 48 h, media was replaced with fresh media containing only 850 nM insulin and 1 mM rosiglitazone and after 96 h media was replaced with fresh media con-taining only insulin.

Dual-luciferase assays

5′-phosphorylated oligonucleotides containing predicted binding sites of miR-29 target genes, 6 bp of flanking DNA, and SacI/XhoI restriction site overhands were annealed then ligated into the dual luciferase reporter plasmid (Promega E1960). For mutated sequences, three to four points were induced in the miR-29a seed binding site using the Q5 site-directed mutagenesis kit (NEB E0554S). Twenty nanograms of recombinant dual-luciferase plasmid and 6 pmol of either miR-29a mimic or scrambled control were mixed with 100 µL of Opti-MEM containing 1 µL of lipofect-amine RNAiMax (Invitrogen 13778) in 24-well plates. Reverse transfection was initiated by adding 90,000 HEK-293A cells/well in 500 µL. After 48 h, firefly activity was measured by luminescence and normalized to Renilla activity.

LNA and miRNA mimic transfections

All cells were transfected on 12-well plates 4 d after inducing dif-ferentiation. For each well, 3 µLof 10 µM double-stranded mature miRNA mimic (Dharmacon) or 20 µM LNA (Qiagen) was mixed with 3 µL of lipofectamine RNAiMAX in 100 µL of Opti-MEM re-duced serum media and added to cells following a 15-min incuba-tion. RNA was collected in TRIzol (Invitrogen) 72 h after transfection.

RNA extraction, reverse transcriptase PCR, and quantitative PCR

RNA was extracted from tissues and cells using a TRIzol (Invitrogen)/chloroform extraction, followed by purification with RNeasy mini kits (Qiagen). To retain small RNAs during RNA purification, 1.5× volumes of 100% ethanol was used to precipi-tate RNA prior to column loading and RW1 buffer was substituted with RWT (Qiagen). For cDNA synthesis of mRNAs, the Applied Biosystems high-capacity cDNA synthesis kit was used with 1 μg of purified RNA. For cDNA synthesis of miRNAs, re-verse transcription using the universal primer 5′-CAGGTCC AGTTTTTTTTTTTTTTTVN-3′ was used as previously reported (Balcells et al. 2011). qPCR for both mRNAs and miRNAs was performed on a QuantStudio 6 Flex machine (Applied Biosystems) using SYBR Green fluorescent dye (Applied Biosystems). qPCR reactions were carried out in 384-well plates with a volume of 10 μL containing 0.5 μM forward and reverse primers, 1× qPCR master mix, and 5 ng of mRNA-based cDNA or 1 ng of miRNA-based cDNA. MiRprimer2.0 was used to design primers for miRNA qPCR (Busk 2014).

Leptin ELISA

A mouse leptin ELISA kit (Crystal Chem 90030) was used to mea-sure leptin concentrations in serum (5 μL per assay) or CM (100 μL per assay) according to the manufacturer's instructions.

HEK-293A cell line

HEK-293A cells (Invitrogen, Cat#R70507) were maintained at 37°C with 5% CO2 in DMEM (4.5 g/L glucose, 0.58 g/L L-glutamine, 110 mg/L sodium pyruvate) supplemented with 10% FBS (Gemini BioProducts) and 1% penicillin-streptomycin (Gibco).

Adenovirus cloning, amplification and transduction

Adenoviruses for miR-29a and the scrambled miRNA control were cloned using the AdEasy system (Agilent, Cat#240010) according to the protocol provided. Briefly, primary miRNAs for miR-29a and miR-16 were amplified from genomic DNA by PCR and cloned into the pAdTrack shuttle (Addgene, Plasmid#16404). The miR-16 pAdTrack plasmid was then mutated by site-directed mutagenesis to induce 3 point-mutations into the miR-16 seed sequence, creating the scrambled control. Both the miR-29a and the scrambled control pAdTrack plasmids were linearized with PmeI and transformed into BJ5183-AD-1 competent cells for recombination with the adenoviral backbone. Successful clones were amplified in the XL10-Gold bacterial strain. Generation of the initial adenoviral stock was performed by digesting the adenoviral plasmid with PacI, then transfecting HEK-293A cells. Several rounds of amplification were performed by infecting HEK-293A cells with low titer adenovirus. For transduction of primary adipocytes, iBAT and eWAT cells were treated with a final adenoviral concentration of 7 x 107 IFU/mL on the 4th day of differentiation. 72 hours post-transduction, adipocytes were inspected for GFP expression, then total RNA was collected.

Western blot analysis

Samples and beads were prepared as described above. Following immunoprecipitation and washes, samples were immediately eluted, separated by SDS-PAGE, transferred to a nitrocellulose membrane and incubated with anti-Ago 2A8. The Ago positive control was prepared from cell lysate of HEK-293T cells transfected with a plasmid expressing mouse Ago2. Protein was detected by incubation with HRPconjugated sheep anti-mouse secondary antibody (Jackson ImmunoResearch) and blots were imaged on a Biorad Chemidoc Imaging System.

Data availability

All raw data from HITS-CLIP and mRNA/miRNA sequencing are available through the Gene Expression Omnibus (GSE142677).

All raw data, data tables, bioinformatics code and analysis details for this research can be accessed: https://github.com/bmurph13/CohenLab.

Notes on statistical analyses

All measurements and analyses were performed using distinct samples. All t-tests were two-tailed with an assumption of normality. Significance cutoffs for HITS-CLIP tag differential comparisons (tags per gene and tags per peak) were adjusted for multiple comparisons.

Packages used for CLIP data Analysis ®

"AnnotationDbi", "annotatr", "Biobase", "BiocGenerics", "BiocManager", "BiocParallel", "Biostrings", "BSgenome", "BSgenome.Mmusculus.UCSC.mm10", "DelayedArray", "DESeq2", "dplyr", "edgeR", "GenomeInfoDb", "GenomicFeatures", "GenomicRanges", "ggplot2", "ggpubr", "ggsignif", "gridExtra", "IRanges", "limma", "magrittr", "matrixStats", "mirbase.db", "org.Mm.eg.db", "plyr", "Rsamtools", "rtracklayer", "S4Vectors", "stringr", "SummarizedExperiment", "tibble", "tidyr", "TxDb.Mmusculus.UCSC.mm10.knownGene", "tximport", "VariantAnnotation", "wesanderson", "XVector". R version 3.6.1 Patched (2019-08-29 r77095).

Hervé Pagès, Marc Carlson, Seth Falcon and Nianhua Li (2019). AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. R package version 1.46.1.

Cavalcante RG, Sartor MA. annotatr: genomic regions in context. Bioinformatics. (2017) 33(15):2381-2383. doi:10.1093/bioinformatics/btx183

Orchestrating high-throughput genomic analysis with Bioconductor. W. Huber, V.J. Carey, R. Gentleman, ..., M. Morgan Nature Methods, 2015:12, 115.
Martin Morgan (2018). BiocManager: Access the Bioconductor Project Package Repository. R package version 1.30.4. https://CRAN.R-project.org/package=BiocManager

Martin Morgan, Valerie Obenchain, Michel Lang, Ryan Thompson and Nitesh Turaga (2019). BiocParallel: Bioconductor facilities for parallel evaluation. R package version 1.18.1. https://github.com/Bioconductor/BiocParallel

H. Pagès, P. Aboyoun, R. Gentleman and S. DebRoy (2019). Biostrings: Efficient manipulation of biological strings. R package version 2.52.0.

Hervé Pagès (2019). BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs. R package version 1.52.0.

The Bioconductor Dev Team (2014). BSgenome.Mmusculus.UCSC.mm10: Full genome sequences for Mus musculus (UCSC version mm10). R package version 1.4.0.

Hervé Pagès, with contributions from Peter Hickey and Aaron Lun (2019). DelayedArray: A unified framework for working transparently with on-disk and in-memory array-like datasets. R package version 0.10.0.

Love, M.I., Huber, W., Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 Genome Biology 15(12):550 (2014)

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.3. https://CRAN.R-project.org/package=dplyr

Robinson MD, McCarthy DJ and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140

McCarthy DJ, Chen Y and Smyth GK (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Research 40, 4288-4297

Sonali Arora, Martin Morgan, Marc Carlson and H. Pagès (2019). GenomeInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style. R package version 1.20.0.

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, et al. (2013) Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 9(8): e1003118. doi:10.1371/journal.pcbi.1003118

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Alboukadel Kassambara (2019). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.2.3. https://CRAN.R-project.org/package=ggpubr

Constantin Ahlmann-Eltze (2019). ggsignif: Significance Brackets for 'ggplot2'. R package version 0.6.0. https://CRAN.R-project.org/package=ggsignif

Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3. https://CRAN.R-project.org/package=gridExtra

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, et al. (2013) Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 9(8): e1003118. doi:10.1371/journal.pcbi.1003118

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47.

Stefan Milton Bache and Hadley Wickham (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. https://CRAN.R-project.org/package=magrittr

Henrik Bengtsson (2019). matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors). R package version 0.55.0. https://CRAN.R-project.org/package=matrixStats

James F. Reid (2013). mirbase.db: miRBase: the microRNA database. R package version 1.2.0.

Marc Carlson (2019). org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.8.2.

Hadley Wickham (2011). The Split-Apply-Combine Strategy for Data Analysis. Journal of Statistical Software, 40(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Morgan M, Pagès H, Obenchain V, Hayden N (2020). Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.4.0, http://bioconductor.org/packages/Rsamtools

M. Lawrence, R. Gentleman, V. Carey: "rtracklayer: an (Acharya et al.) package for interfacing with genome browsers". Bioinformatics 25:1841-1842.

H. Pagès, M. Lawrence and P. Aboyoun (2019). S4Vectors: Foundation of vector-like and list-like containers in Bioconductor. R package version 0.22.1.

Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr

Martin Morgan, Valerie Obenchain, Jim Hester and Hervé Pagès (2019). SummarizedExperiment: SummarizedExperiment container. R package version 1.14.1.

Kirill Müller and Hadley Wickham (2019). tibble: Simple Data Frames. R package version 2.1.3. https://CRAN.R-project.org/package=tibble

Hadley Wickham and Lionel Henry (2019). tidyr: Tidy Messy Data. R package version 1.0.0. https://CRAN.R-project.org/package=tidyr

Bioconductor Core Team and Bioconductor Package Maintainer (2019). TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s). R package version 3.4.7.

Charlotte Soneson, Michael I. Love, Mark D. Robinson (2015): Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research

Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M (2014). “VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants.” _Bioinformatics_, *30*(14), 2076-2078. doi: 10.1093/bioinformatics/btu168 (URL: https://doi.org/10.1093/bioinformatics/btu168).

Karthik Ram and Hadley Wickham (2018). wesanderson: A Wes Anderson Palette Generator. R package version 0.3.6. https://CRAN.R-project.org/package=wesanderson

Hervé Pagès and Patrick Aboyoun (2019). XVector: Foundation of external vector representation and manipulation in Bioconductor. R package version 0.24.0.

References:

Acharya A, Berry DC, Zhang H, Jiang Y, Jones BT, Hammer RE, Graff JM, Mendell JT. 2019. miR-26 suppresses adipocyte progenitor differentiation and fat production by targeting Fbxl19. Genes Dev.

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. Elife 4.

Balcells I, Cirera S, Busk PK. 2011. Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. BMC Biotechnol 11: 70.

Busk PK. 2014. A tool for design of primers for microRNA-specific quantitative RT-qPCR. BMC Bioinformatics 15: 29.

Caravia XM, Fanjul V, Oliver E, Roiz-Valle D, Moran-Alvarez A, Desdin-Mico G, Mittelbrunn M, Cabo R, Vega JA, Rodriguez F et al. 2018. The microRNA-29/PGC1alpha regulatory axis is critical for metabolic control of cardiac function. PLoS Biol 16: e2006247.

Chan BKC. 2018. Data Analysis Using R Programming. Adv Exp Med Biol 1082: 47-122.

Chen Y, Siegel F, Kipschull S, Haas B, Fröhlich H, Meister G, Pfeifer A. 2013. miR-155 regulates differentiation of brown and beige adipocytes via a bistable circuit. Nat Commun 4: 1769.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460: 479-486.

Cohen P, Levy JD, Zhang Y, Frontini A, Kolodin DP, Svensson KJ, Lo JC, Zeng X, Ye L, Khandekar MJ et al. 2014. Ablation of PRDM16 and beige adipose causes metabolic dysfunction and a subcutaneous to visceral fat switch. Cell 156: 304-316.

Cypess AM, Lehman S, Williams G, Tal I, Rodman D, Goldfine AB, Kuo FC, Palmer EL, Tseng YH, Doria A et al. 2009. Identification and importance of brown adipose tissue in adult humans. N Engl J Med 360: 1509-1517.

Dallner OS, Marinis JM, Lu YH, Birsoy K, Werner E, Fayzikhodjaeva G, Dill BD, Molina H, Moscati A, Kutalik Z et al. 2019. Dysregulation of a long noncoding RNA reduces leptin leading to a leptin-responsive form of obesity. Nat Med 25: 507-516.

Frederich RC, Hamann A, Anderson S, Lollmann B, Lowell BB, Flier JS. 1995. Leptin levels reflect body lipid content in mice: evidence for diet-induced resistance to leptin action. Nat Med 1: 1311-1314.

Fu T, Seok S, Choi S, Huang Z, Suino-Powell K, Xu HE, Kemper B, Kemper JK. 2014. MicroRNA-34a inhibits beige and brown fat formation in obesity in part by suppressing adipocyte fibroblast growth factor 21 signaling and SIRT1 function. Mol Cell Biol 34: 4130-4142.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

Giroud M, Karbiener M, Pisani DF, Ghandour RA, Beranger GE, Niemi T, Taittonen M, Nuutila P, Virtanen KA, Langin D et al. 2016a. Let-7i-5p represses brite adipocyte function in mice and humans. Sci Rep 6: 28613.

Giroud M, Pisani DF, Karbiener M, Barquissau V, Ghandour RA, Tews D, Fischer-Posovszky P, Chambard JC, Knippschild U, Niemi T et al. 2016b. miR-125b affects mitochondrial biogenesis and impairs brite adipocyte formation and function. Mol Metab 5: 615-625.

Glantschnig C, Koenen M, Gil-Lozano M, Karbiener M, Pickrahn I, Williams-Dautovich J, Patel R, Cummins CL, Giroud M, Hartleben G et al. 2019. A miR-29a-driven negative feedback loop regulates peripheral glucocorticoid receptor signaling. FASEB J 33: 5924-5941.

Hardy OT, Czech MP, Corvera S. 2012. What causes the insulin resistance underlying obesity? Curr Opin Endocrinol Diabetes Obes 19: 81-87.

Hu F, Wang M, Xiao T, Yin B, He L, Meng W, Dong M, Liu F. 2015. miR-30 promotes thermogenesis and the development of beige fat by targeting RIP140. Diabetes 64: 2056-2068.

Jayedi A, Soltani S, Zargar MS, Khan TA, Shab-Bidar S. 2020. Central fatness and risk of all cause mortality: systematic review and dose-response meta-analysis of 72 prospective cohort studies. Bmj 370: m3324.

Kajimura S, Spiegelman BM, Seale P. 2015. Brown and Beige Fat: Physiological Roles beyond Heat Generation. Cell Metab 22: 546-559.

Koh EH, Chernis N, Saha PK, Xiao L, Bader DA, Zhu B, Rajapakshe K, Hamilton MP, Liu X, Perera D et al. 2018. miR-30a Remodels Subcutaneous Adipose Tissue Inflammation to Improve Insulin Sensitivity in Obesity. Diabetes 67: 2541-2553.

Kristensen MM, Davidsen PK, Vigelso A, Hansen CN, Jensen LJ, Jessen N, Bruun JM, Dela F, Helge JW. 2017. miRNAs in human subcutaneous adipose tissue: Effects of weight loss induced by hypocaloric diet and exercise. Obesity (Silver Spring) 25: 572-580.

Kurtz CL, Fannin EE, Toth CL, Pearson DS, Vickers KC, Sethupathy P. 2015. Inhibition of miR-29 has a significant lipid-lowering benefit through suppression of lipogenic programs in liver. Sci Rep 5: 12911.

Lo KA, Huang S, Walet ACE, Zhang ZC, Leow MK, Liu M, Sun L. 2018. Adipocyte Long-Noncoding RNA Transcriptome Analysis of Obese Mice Identified Lnc-Leptin, Which Regulates Leptin. Diabetes 67: 1045-1056.

Luna JM, Barajas JM, Teng KY, Sun HL, Moore MJ, Rice CM, Darnell RB, Ghoshal K. 2017. Argonaute CLIP Defines a Deregulated miR-122-Bound Transcriptome that Correlates with Patient Survival in Human Liver Cancer. Mol Cell 67: 400-410 e407.

Maffei M, Halaas J, Ravussin E, Pratley RE, Lee GH, Zhang Y, Fei H, Kim S, Lallone R, Ranganathan S et al. 1995. Leptin levels in human and rodent: measurement of plasma leptin and ob RNA in obese and weight-reduced subjects. Nat Med 1: 1155-1161.

Moore MJ, Blachere NE, Fak JJ, Park CY, Sawicka K, Parveen S, Zucker-Scharff I, Moltedo B, Rudensky AY, Darnell RB. 2018. ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. Elife 7.

Moore MJ, Zhang C, Gantman EC, Mele A, Darnell JC, Darnell RB. 2014. Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat Protoc 9: 263-293.

Mori M, Nakagami H, Rodriguez-Araujo G, Nimura K, Kaneda Y. 2012. Essential role for miR-196a in brown adipogenesis of white fat progenitor cells. PLoS Biol 10: e1001314.

Mori MA, Thomou T, Boucher J, Lee KY, Lallukka S, Kim JK, Torriani M, Yki-Jarvinen H, Grinspoon SK, Cypess AM et al. 2014. Altered miRNA processing disrupts brown/white adipocyte determination and associates with lipodystrophy. J Clin Invest 124: 3339-3351.

Neeland IJ, Ross R, Després JP, Matsuzawa Y, Yamashita S, Shai I, Seidell J, Magni P, Santos RD, Arsenault B et al. 2019. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. Lancet Diabetes Endocrinol 7: 715-725.

Pullen TJ, da Silva Xavier G, Kelsey G, Rutter GA. 2011. miR-29a and miR-29b contribute to pancreatic beta-cell-specific silencing of monocarboxylate transporter 1 (Mct1). Mol Cell Biol 31: 3182-3194.

Reis FC, Branquinho JL, Brandao BB, Guerra BA, Silva ID, Frontini A, Thomou T, Sartini L, Cinti S, Kahn CR et al. 2016. Fat-specific Dicer deficiency accelerates aging and mitigates several effects of dietary restriction in mice. Aging (Albany NY) 8: 1201-1222.

Rosen ED, Spiegelman BM. 2014. What we talk about when we talk about fat. Cell 156: 20-44.
Rosenwald M, Perdikari A, Rulicke T, Wolfrum C. 2013. Bi-directional interconversion of brite and white adipocytes. Nat Cell Biol 15: 659-667.

Sawicka K, Hale CR, Park CY, Fak JJ, Gresack JE, Van Driesche SJ, Kang JJ, Darnell JC, Darnell RB. 2019. FMRP has a cell-type-specific role in CA1 pyramidal neurons to regulate

autism-related transcripts and circadian memory. Elife 8.

Seale P, Kajimura S, Yang W, Chin S, Rohas LM, Uldry M, Tavernier G, Langin D, Spiegelman BM. 2007. Transcriptional control of brown fat determination by PRDM16. Cell Metab 6: 38-54.

Shah A, Qian Y, Weyn-Vanhentenryck SM, Zhang C. 2017. CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. Bioinformatics 33: 566-567.

Spengler RM, Zhang X, Cheng C, McLendon JM, Skeie JM, Johnson FL, Davidson BL, Boudreau RL. 2016. Elucidation of transcriptome-wide microRNA binding sites in human cardiac tissues by Ago2 HITS-CLIP. Nucleic Acids Res 44: 7120-7131.

Stanford KI, Middelbeek RJ, Townsend KL, An D, Nygaard EB, Hitchcox KM, Markan KR, Nakano K, Hirshman MF, Tseng YH et al. 2013. Brown adipose tissue regulates glucose homeostasis and insulin sensitivity. J Clin Invest 123: 215-223.

Sun L, Trajkovski M. 2014. MiR-27 orchestrates the transcriptional regulation of brown adipogenesis. Metabolism 63: 272-282.

Sun L, Xie H, Mori MA, Alexander R, Yuan B, Hattangadi SM, Liu Q, Kahn CR, Lodish HF. 2011. Mir193b-365 is essential for brown fat differentiation. Nat Cell Biol 13: 958-965.

Tontonoz P, Hu E, Spiegelman BM. 1994. Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. Cell 79: 1147-1156.

Tontonoz P, Spiegelman BM. 2008. Fat and beyond: the diverse biology of PPARgamma. Annu Rev Biochem 77: 289-312.

Trajkovski M, Ahmed K, Esau CC, Stoffel M. 2012. MyomiR-133 regulates brown fat differentiation through Prdm16. Nat Cell Biol 14: 1330-1335.

Trajkovski M, Hausser J, Soutschek J, Bhat B, Akin A, Zavolan M, Heim MH, Stoffel M. 2011. MicroRNAs 103 and 107 regulate insulin sensitivity. Nature 474: 649-653.

van Marken Lichtenbelt WD, Vanhommerig JW, Smulders NM, Drossaerts JM, Kemerink GJ, Bouvy ND, Schrauwen P, Teule GJ. 2009. Cold-activated brown adipose tissue in healthy men. N Engl J Med 360: 1500-1508.

Ventayol M, Vinas JL, Sola A, Jung M, Brune B, Pi F, Mastora C, Hotter G. 2014. miRNA let-7e targeting MMP9 is involved in adipose-derived stem cell differentiation toward epithelia. Cell Death Dis 5: e1048.

Virtanen KA, Lidell ME, Orava J, Heglind M, Westergren R, Niemi T, Taittonen M, Laine J, Savisto NJ, Enerback S et al. 2009. Functional brown adipose tissue in healthy adults. N Engl J Med 360: 1518-1525.

Ward ZJ, Bleich SN, Cradock AL, Barrett JL, Giles CM, Flax C, Long MW, Gortmaker SL. 2019. Projected U.S. State-Level Prevalence of Adult Obesity and Severe Obesity. N Engl J Med 381: 2440-2450.

Wu J, Bostrom P, Sparks LM, Ye L, Choi JH, Giang AH, Khandekar M, Virtanen KA, Nuutila P, Schaart G et al. 2012. Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. Cell 150: 366-376.

Zhang XM, Wang LH, Su DJ, Zhu D, Li QM, Chi MH. 2016. MicroRNA-29b promotes the adipogenic differentiation of human adipose tissue-derived stromal cells. Obesity (Silver Spring) 24: 1097-1105.

Chapter 10: References

Adelmann, G., Jonas, P., and Monyer, H. (1999). Ionotropic Glutamate Receptors in the CNS (Springer Science & Business Media).

Bae, B.-I., Jayaraman, D., and Walsh, C.A. (2015). Genetic changes shaping the human brain. Dev. Cell *32*, 423–434.

Barbon, A., Vallini, I., and Barlati, S. (2001). Genomic organization of the human GRIK2 gene and evidence for multiple splicing variants. Gene *274*, 187–197.

Beneyto, M., Kristiansen, L.V., Oni-Orisan, A., McCullumsmith, R.E., and Meador-Woodruff, J.H. (2007). Abnormal glutamate receptor expression in the medial temporal lobe in schizophrenia and mood disorders. Neuropsychopharmacology *32*, 1888–1902.

Boycott, K.M., Vanstone, M.R., Bulman, D.E., and MacKenzie, A.E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat. Rev. Genet. *14*, 681–691.

Breschi, A., Gingeras, T.R., and Guigó, R. (2017). Comparative transcriptomics in human and mouse. Nat. Rev. Genet. *18*, 425–440.

Buckanovich, R.J., and Darnell, R.B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. Mol. Cell. Biol. *17*, 3194–3201.

Buckanovich, R.J., Posner, J.B., and Darnell, R.B. (1993). Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. Neuron *11*, 657–672.

Buckanovich, R.J., Yang, Y.Y., and Darnell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. J. Neurosci. *16*, 1114–1122.

Capra, J.A., Erwin, G.D., McKinsey, G., Rubenstein, J.L.R., and Pollard, K.S. (2013). Many human accelerated regions are developmental enhancers. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20130025.

Castellano, S., Parra, G., Sánchez-Quinto, F.A., Racimo, F., Kuhlwilm, M., Kircher, M., Sawyer, S., Fu, Q., Heinze, A., Nickel, B., et al. (2014). Patterns of coding variation in the complete exomes of three Neandertals. Proc. Natl. Acad. Sci. U. S. A. *111*, 6666–6671.

Castro, D., and Iannaccone, S.T. (2014). Spinal muscular atrophy: therapeutic strategies. Curr. Treat. Options Neurol. *16*, 316.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of

SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Charrier, C., Joshi, K., Coutinho-Budd, J., Kim, J.-E., Lambert, N., de Marchena, J., Jin, W.-L., Vanderhaeghen, P., Ghosh, A., Sassa, T., et al. (2012). Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. Cell *149*, 923–935.

Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature *437*, 69–87.

Darnell, R.B. (1996). Onconeural antigens and the paraneoplastic neurologic disorders: at the intersection of cancer, immunity, and the brain. Proc. Natl. Acad. Sci. U. S. A. *93*, 4529–4536.

Darnell, R.B., and Posner, J.B. (2003a). Paraneoplastic syndromes involving the nervous system. N. Engl. J. Med. *349*, 1543–1554.

Darnell, R.B., and Posner, J.B. (2003b). Observing the invisible: successful tumor immunity in humans. Nat. Immunol. *4*, 201.

Darnell, J.E., Wall, R., and Tushinski, R.J. (1971). An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. Proc. Natl. Acad. Sci. U. S. A. *68*, 1321–1325.

Darnell, R., Darnell, R.B., and Posner, J.B. (2011). Paraneoplastic Syndromes (OUP USA).

Dawson, G.R., Seabrook, G.R., Zheng, H., Smith, D.W., Graham, S., O'Dowd, G., Bowery, B.J., Boyce, S., Trumbauer, M.E., Chen, H.Y., et al. (1999). Age-related cognitive deficits, impaired long-term potentiation and reduction in synaptic marker density in mice lacking the beta-amyloid precursor protein. Neuroscience *90*, 1–13.

De Haan, L., Schirmbeck, F., and Zink, M. (2015). Obsessive-Compulsive Symptoms in Schizophrenia (Springer).

Dennis, M.Y., Nuttle, X., Sudmant, P.H., Antonacci, F., Graves, T.A., Nefedov, M., Rosenfeld, J.A., Sajjadian, S., Malig, M., Kotkiewicz, H., et al. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. Cell *149*, 912–922.

Dredge, B.K., and Darnell, R.B. (2003). Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. Mol. Cell. Biol. *23*, 4687–4700.

Dredge, B.K., Polydorides, A.D., and Darnell, R.B. (2001). The splice of life: alternative splicing and neurological disease. Nat. Rev. Neurosci. *2*, 43–50.

Dredge, B.K., Stefani, G., Engelhard, C.C., and Darnell, R.B. (2005). Nova autoregulation reveals dual functions in neuronal splicing. EMBO J. *24*, 1608–1620.

Dutta, S., Das, S., Guhathakurta, S., Sen, B., Sinha, S., Chatterjee, A., Ghosh, S., Ahmed, S., Ghosh, S., and Usha, R. (2007). Glutamate receptor 6 gene (GluR6 or GRIK2) polymorphisms in

the Indian population: a genetic association study on autism spectrum disorder. Cell. Mol. Neurobiol. *27*, 1035–1047.

Edmonds, M., Vaughan, M.H., Jr, and Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. Proc. Natl. Acad. Sci. U. S. A. *68*, 1336–1340.

Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J.L., and Darnell, R.B. (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. Elife *2*, e00178.

Gao, F.-B., and Paul Taylor, J. (2012). RNA-binding proteins in neurological disease. Brain Research *1462*, 1–2.

Gebauer, F., Schwarzl, T., Valcárcel, J., and Hentze, M.W. (2021). RNA-binding proteins in human genetic disease. Nature Reviews Genetics *22*, 185–198.

Gordon, A., Yoon, S.-J., Tran, S.S., Makinson, C.D., Park, J.Y., Andersen, J., Valencia, A.M., Horvath, S., Xiao, X., Huguenard, J.R., et al. (2021). Long-term maturation of human cortical organoids matches key early postnatal transitions. Nat. Neurosci. *24*, 331–342.

Groeneweg, F.L., Trattnig, C., Kuhse, J., Nawrotzki, R.A., and Kirsch, J. (2018). Gephyrin: a key regulatory protein of inhibitory synapses and beyond. Histochem. Cell Biol. *150*, 489–508.

Havens, M.A., and Hastings, M.L. (2016). Splice-switching antisense oligonucleotides as therapeutic drugs. Nucleic Acids Res. *44*, 6549–6563.

Heller, S.A., Shih, R., Kalra, R., and Kang, P.B. (2020). Emery-Dreifuss muscular dystrophy. Muscle Nerve *61*, 436–448.

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68.

Hogarth, K.A., Costford, S.R., Yoon, G., Sondheimer, N., and Maynes, J.T. (2018). DNM1L Variant Alters Baseline Mitochondrial Function and Response to Stress in a Patient with Severe Neurological Dysfunction. Biochem. Genet. *56*, 56–77.

Huang, C.S., Shi, S.-H., Ule, J., Ruggiu, M., Barker, L.A., Darnell, R.B., Jan, Y.N., and Jan, L.Y. (2005). Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell *123*, 105–118.

Jamain, S., Betancur, C., Quach, H., Philippe, A., Fellous, M., Giros, B., Gillberg, C., Leboyer, M., Bourgeron, T., and Paris Autism Research International Sibpair (PARIS) Study (2002). Linkage and association of the glutamate receptor 6 gene with autism. Mol. Psychiatry *7*, 302–310.

Jelen, N., Ule, J., Zivin, M., and Darnell, R.B. (2007). Evolution of Nova-dependent splicing

regulation in the brain. PLoS Genet. *3*, 1838–1847.

Jensen, K.B., Musunuru, K., Lewis, H.A., Burley, S.K., and Darnell, R.B. (2000a). The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. Proc. Natl. Acad. Sci. U. S. A. *97*, 5740–5745.

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y., and Darnell, R.B. (2000b). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron *25*, 359–371.

Johnson, M.B., and Walsh, C.A. (2017). Cerebral cortical neuron diversity and development at single-cell resolution. Curr. Opin. Neurobiol. *42*, 9–16.

Josifova, D.J., Monroe, G.R., Tessadori, F., de Graaff, E., van der Zwaag, B., Mehta, S.G., DDD Study, Harakalova, M., Duran, K.J., Savelberg, S.M.C., et al. (2016). Heterozygous KIDINS220/ARMS nonsense variants cause spastic paraplegia, intellectual disability, nystagmus, and obesity. Hum. Mol. Genet. *25*, 2158–2167.

van der Kant, R., and Goldstein, L.S.B. (2015). Cellular functions of the amyloid precursor protein from development to dementia. Dev. Cell *32*, 502–515.

Kim, J., Hu, C., Moufawad El Achkar, C., Black, L.E., Douville, J., Larson, A., Pendergast, M.K., Goldkind, S.F., Lee, E.A., Kuniholm, A., et al. (2019). Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. N. Engl. J. Med. *381*, 1644–1652.

Krainer, A.R. (1988). Pre-mRNA splicing by complementation with purified human U1, U2, U4/U6 and U5 snRNPs. Nucleic Acids Res. *16*, 9415–9429.

Krainer, A.R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell *42*, 725–736.

Krishnan, A., Zhang, R., Yao, V., Theesfeld, C.L., Wong, A.K., Tadych, A., Volfovsky, N., Packer, A., Lash, A., and Troyanskaya, O.G. (2016). Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nature Neuroscience *19*, 1454–1462.

Kuhlwilm, M., and Boeckx, C. (2019). A catalog of single nucleotide changes distinguishing modern humans from archaic hominins. Sci. Rep. *9*, 1–14.

Kuhlwilm, M., Gronau, I., Hubisz, M.J., de Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H.A., Lalueza-Fox, C., de la Rasilla, M., et al. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. Nature *530*, 429–433.

Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief. Bioinform. *14*, 144–161.

Lacoste, A., Berenshteyn, F., and Brivanlou, A.H. (2009). An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic

stem cells. Cell Stem Cell *5*, 332–342.

Lanore, F., Labrousse, V.F., Szabo, Z., Normand, E., Blanchet, C., and Mulle, C. (2012). Deficits in morphofunctional maturation of hippocampal mossy fiber synapses in a mouse model of intellectual disability. J. Neurosci. *32*, 17882–17893.

Lazarov, O., and Tesco, G. (2016). Genes, Environment and Alzheimer's Disease (Academic Press).

Lee, J.-H., Lee, J.-M., Ramos, E.M., Gillis, T., Mysore, J.S., Kishikawa, S., Hadzi, T., Hendricks, A.E., Hayden, M.R., Morrison, P.J., et al. (2012). TAA repeat variation in the GRIK2 gene does not influence age at onset in Huntington's disease. Biochem. Biophys. Res. Commun. *424*, 404–408.

Lee, S.Y., Mendecki, J., and Brawerman, G. (1971). A Polynucleotide Segment Rich in Adenylic Acid in the Rapidly-Labeled Polyribosomal RNA Component of Mouse Sarcoma 180 Ascites Cells. Proceedings of the National Academy of Sciences *68*, 1331–1335.

Leggere, J.C., Saito, Y., Darnell, R.B., Tessier-Lavigne, M., Junge, H.J., and Chen, Z. (2016). NOVA regulates Dcc alternative splicing during neuronal migration and axon guidance in the spinal cord. Elife *5*.

Lewis, H.A., Musunuru, K., Jensen, K.B., Edo, C., Chen, H., Darnell, R.B., and Burley, S.K. (2000). Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. Cell *100*, 323–332.

Licatalosi, D.D., and Darnell, R.B. (2006). Splicing regulation in neurologic disease. Neuron *52*, 93–101.

Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. *11*, 75–87.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. Proc. Natl. Acad. Sci. U. S. A. *98*, 11193–11198.

Lotharius, J., Barg, S., Wiekop, P., Lundberg, C., Raymon, H.K., and Brundin, P. (2002). Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line. J. Biol. Chem. *277*, 38884–38894.

Lotharius, J., Falsig, J., van Beek, J., Payne, S., Dringen, R., Brundin, P., and Leist, M. (2005). Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. J. Neurosci. *25*, 6329–6342.

Lukong, K.E., Chang, K.-W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends in Genetics *24*, 416–425.

Marini, F., Linke, J., and Binder, H. (2020). ideal: an R/Bioconductor package for interactive differential expression analysis. BMC Bioinformatics *21*, 565.

Mayeda, A., Helfman, D.M., and Krainer, A.R. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. Mol. Cell. Biol. *13*, 2993–3001.

Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. Nat. Protoc. *9*, 263–293.

Motazacker, M.M., Rost, B.R., Hucho, T., Garshasbi, M., Kahrizi, K., Ullmann, R., Abedini, S.S., Nieh, S.E., Amini, S.H., Goswami, C., et al. (2007). A defect in the ionotropic glutamate receptor 6 gene (GRIK2) is associated with autosomal recessive mental retardation. Am. J. Hum. Genet. *81*, 792–798.

Musunuru, K. (2003). Cell-Specific RNA-Binding Proteins in Human Disease. Trends in Cardiovascular Medicine *13*, 188–195.

Musunuru, K., and Darnell, R.B. (2004). Determination and augmentation of RNA sequence specificity of the Nova K-homology domains. Nucleic Acids Res. *32*, 4852–4861.

Nafee, N., and Gouda, N. (2017). Nucleic Acids-based Nanotherapeutics Crossing the Blood Brain Barrier. Curr. Gene Ther. *17*, 154–169.

Nussbacher, J.K., Tabet, R., Yeo, G.W., and Lagier-Tourenne, C. (2019). Disruption of RNA Metabolism in Neurological Diseases and Emerging Therapeutic Interventions. Neuron *102*, 294–320.

O'Donnell-Luria, A.H., Pais, L.S., Faundes, V., Wood, J.C., Sveden, A., Luria, V., Abou Jamra, R., Accogli, A., Amburgey, K., Anderlid, B.M., et al. (2019). Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy. Am. J. Hum. Genet. *104*, 1210–1222.

Pääbo, S. (2014). The human condition-a molecular approach. Cell *157*, 216–226.

Pembroke, W.G., Hartl, C.L., and Geschwind, D.H. (2021). Evolutionary conservation and divergence of the human brain transcriptome. Genome Biol. *22*, 52.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167–172.

Posner, J.B. (2003). Immunology of paraneoplastic syndromes: overview. Ann. N. Y. Acad. Sci. *998*, 178–186.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926–aad9926.

Racca, C., Gardiol, A., Eom, T., Ule, J., Triller, A., and Darnell, R.B. (2010). The Neuronal Splicing Factor Nova Co-Localizes with Target RNAs in the Dendrite. Front. Neural Circuits *4*, 5.

Rigo, F., Hua, Y., Krainer, A.R., and Bennett, C.F. (2012). Antisense-based therapy for the treatment of spinal muscular atrophy. J. Cell Biol. *199*, 21–25.

Ring, S., Weyer, S.W., Kilian, S.B., Waldron, E., Pietrzik, C.U., Filippov, M.A., Herms, J., Buchholz, C., Eckman, C.B., Korte, M., et al. (2007). The secreted beta-amyloid precursor protein ectodomain APPs alpha is sufficient to rescue the anatomical, behavioral, and electrophysiological abnormalities of APP-deficient mice. J. Neurosci. *27*, 7817–7826.

Ross, R.A., Spengler, B.A., and Biedler, J.L. (1983). Coordinate morphological and biochemical interconversion of human neuroblastoma cells. J. Natl. Cancer Inst. *71*, 741–747.

Ruggiu, M., Herbst, R., Kim, N., Jevsek, M., Fak, J.J., Mann, M.A., Fischbach, G., Burden, S.J., and Darnell, R.B. (2009). Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing. Proc. Natl. Acad. Sci. U. S. A. *106*, 3513–3518.

Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317–331.

Saito, Y., Miranda-Rottmann, S., Ruggiu, M., Park, C.Y., Fak, J.J., Zhong, R., Duncan, J.S., Fabella, B.A., Junge, H.J., Chen, Z., et al. (2016). NOVA2-mediated RNA regulation is required for axonal pathfinding during development. Elife *5*.

Saito, Y., Yuan, Y., Zucker-Scharff, I., Fak, J.J., Jereb, S., Tajima, Y., Licatalosi, D.D., and Darnell, R.B. (2019). Differential NOVA2-Mediated Splicing in Excitatory and Inhibitory Neurons Regulates Cortical Development and Cerebellar Function. Neuron *101*, 707–720.e5.

Scheckel, C., Drapeau, E., Frias, M.A., Park, C.Y., Fak, J., Zucker-Scharff, I., Kou, Y., Haroutunian, V., Ma'ayan, A., Buxbaum, J.D., et al. (2016). Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. Elife *5*.

Seabrook, G.R., Smith, D.W., Bowery, B.J., Easter, A., Reynolds, T., Fitzjohn, S.M., Morton, R.A., Zheng, H., Dawson, G.R., Sirinathsinghji, D.J., et al. (1999). Mechanisms contributing to the deficits in hippocampal synaptic plasticity in mice lacking amyloid precursor protein. Neuropharmacology *38*, 349–359.

Shah, A., Qian, Y., Weyn-Vanhentenryck, S.M., and Zhang, C. (2017). CLIP Tool Kit (CTK): a flexible and robust pipeline to analyze CLIP sequencing data. Bioinformatics *33*, 566–567.

Shen, S., Park, J.W., Lu, Z.-X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y. (2014). rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq

data. Proc. Natl. Acad. Sci. U. S. A. *111*, E5593–E5601.

Shepard, P.J., Choi, E.-A., Lu, J., Flanagan, L.A., Hertel, K.J., and Shi, Y. (2011). Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA *17*, 761–772.

Shi, Y., and Manley, J.L. (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev. *29*, 889–897.

Shi, Y., Kirwan, P., and Livesey, F.J. (2012). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. *7*, 1836–1846.

Shibata, H., Shibata, A., Ninomiya, H., Tashiro, N., and Fukumaki, Y. (2002). Association study of polymorphisms in the GluR6 kainate receptor gene (GRIK2) with schizophrenia. Psychiatry Res. *113*, 59–67.

Shuang, M., Liu, J., Jia, M.X., Yang, J.Z., Wu, S.P., Gong, X.H., Ling, Y.S., Ruan, Y., Yang, X.L., and Zhang, D. (2004). Family-based association study between autism and glutamate receptor 6 gene in Chinese Han trios. Am. J. Med. Genet. B Neuropsychiatr. Genet. *131B*, 48–50.

Smith, K.D.B., Kallhoff, V., Zheng, H., and Pautler, R.G. (2007). In vivo axonal transport rates decrease in a mouse model of Alzheimer's disease. Neuroimage *35*, 1401–1408.

Smith, K.D.B., Peethumnongsin, E., Lin, H., Zheng, H., and Pautler, R.G. (2010). Increased Human Wildtype Tau Attenuates Axonal Transport Deficits Caused by Loss of APP in Mouse Models. Magn. Reson. Insights *4*, 11–18.

Stankiewicz, P., Khan, T.N., Szafranski, P., Slattery, L., Streff, H., Vetrini, F., Bernstein, J.A., Brown, C.W., Rosenfeld, J.A., Rednam, S., et al. (2017). Haploinsufficiency of the Chromatin Remodeler BPTF Causes Syndromic Developmental and Speech Delay, Postnatal Microcephaly, and Dysmorphic Features. Am. J. Hum. Genet. *101*, 503–515.

Staropoli, J.F., Li, H., Chun, S.J., Allaire, N., Cullen, P., Thai, A., Fleet, C.M., Hua, Y., Bennett, C.F., Krainer, A.R., et al. (2015). Rescue of gene-expression changes in an induced mouse model of spinal muscular atrophy by an antisense oligonucleotide that promotes inclusion of SMN2 exon 7. Genomics *105*, 220–228.

Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R., and Rottman, F.M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes & Development *7*, 2598–2608.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Teplova, M., Malinina, L., Darnell, J.C., Song, J., Lu, M., Abagyan, R., Musunuru, K., Teplov, A., Burley, S.K., Darnell, R.B., et al. (2011). Protein-RNA and protein-protein recognition by dual KH1/2 domains of the neuronal splicing factor Nova-1. Structure *19*, 930–944.

Trujillo, C.A., Rice, E.S., Schaefer, N.K., Chaim, I.A., Wheeler, E.C., Madrigal, A.A.,

Buchanan, J., Preissl, S., Wang, A., Negraes, P.D., et al. (2021). Reintroduction of the archaic variant of in cortical organoids alters neurodevelopment. Science *371*.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.-S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. *37*, 844–852.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature *444*, 580–586.

United States. Congress. House. Committee on Appropriations. Subcommittee on the Departments of Labor, Health, Human Services, Education, and Agencies, R. (2004). Departments of Labor, Health and Human Services, Education, and Related Agencies Appropriations for 2005: Hearings Before a Subcommittee of the Committee on Appropriations, House of Representatives, One Hundred Eighth Congress, Second Session.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, R., Zheng, D., Yehia, G., and Tian, B. (2018). A compendium of conserved cleavage and polyadenylation events in mammalian genes. Genome Res. *28*, 1427–1441.

Wood, B., and Collard, M. (1999). The human genus. Science *284*, 65–71.

Wu, X., Wang, S.-H., Sun, J., Krainer, A.R., Hua, Y., and Prior, T.W. (2017). A-44G transition in SMN2 intron 6 protects patients with spinal muscular atrophy. Hum. Mol. Genet. *26*, 2768–2780.

Yang, Y.Y., Yin, G.L., and Darnell, R.B. (1998). The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. Proc. Natl. Acad. Sci. U. S. A. *95*, 13254–13259.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. *11*, R14.

Yuan, Y., Xie, S., Darnell, J.C., Darnell, A.J., Saito, Y., Phatnani, H., Murphy, E.A., Zhang, C., Maniatis, T., and Darnell, R.B. (2018). Cell type-specific CLIP reveals that NOVA regulates cytoskeleton interactions in motoneurons. Genome Biol. *19*, 117.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome.

Nature *515*, 355–364.

Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat. Biotechnol. *29*, 607–614.

Zhang, C., Frias, M.A., Mele, A., Ruggiu, M., Eom, T., Marney, C.B., Wang, H., Licatalosi, D.D., Fak, J.J., and Darnell, R.B. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. Science *329*, 439–443.

Zhang, C., Lee, K.-Y., Swanson, M.S., and Darnell, R.B. (2013). Prediction of clustered RNA-binding protein motif sites in the mammalian genome. Nucleic Acids Res. *41*, 6793–6807.

Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., et al. (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell *166*, 1147–1162.e15.

Zheng, H., Jiang, M., Trumbauer, M.E., Sirinathsinghji, D.J., Hopkins, R., Smith, D.W., Heavens, R.P., Dawson, G.R., Boyce, S., Conner, M.W., et al. (1995). beta-Amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity. Cell *81*, 525–531.

Zhou, Y., Dong, F., and Mao, Y. (2018). Control of CNS functions by RNA-binding proteins in neurological diseases. Curr Pharmacol Rep *4*, 301–313.

(2016). CLIP Tool Kit (CTK): a Flexible and Robust Pipeline to Analyze CLIP Sequencing Data.