TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl für Proteomik und Bioanalytik

Regulation of Protein Translation and Degradation in Neurons

Ruchi Chauhan

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. M. Klingenspor
Prüfer der Dissertation:	1. Univ.-Prof. Dr. B. Küster
	2. Assist. Prof. Dr. J. Steen

Die Dissertation wurde am 02.02.2015 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 21.05.2015 angenommen.

To those who believed in me and demanded perfection

(Gayatri, Ved, Judith, Late V.K. Jain and Soham)

Table of Contents

Abstract

Regulation of cellular proteome by protein translation and degradation, protein homeostasis, is crucial for the normal functioning of a cell - especially, in a highly metabolic and active cell such as the neuron. Perturbations in protein homeostasis have been implicated in various neuronal diseases. Abnormal regulation of even a single protein e.g. Survival of Motor Neurons (SMN) may cause debilitating neurodegenerative disease such as Spinal Muscular Atrophy (SMA). In this thesis, I undertook global high-throughput quantitative mass-spectrometry based proteomics studies, along with biochemical validation of subset of proteins, to understand protein homeostasis in neurons. The specific aim of this study was to investigate (a) activity dependent protein synthesis in stimulated neurons, (b) global protein degradation by Anaphase Promoting Complex/Cyclosome (APC/C) mediated ubiquitination mechanisms and (c) detailed regulation of SMN protein along with its interacting partner HuD by APC/C in the context of SMA disease. The results from these studies provide novel insights in understanding the regulation of protein translation and degradation in neurons, and may also have potential for developing therapeutics. 'Novel' peptides/translational products were identified from non-coding regions of the genome in depolarized cortical neurons from mouse. The abundance of these peptides was regulated over time similar to that of known proteins, indicating that 'novel' proteome in these neurons may have biological functions. In the protein degradation study, new members of the kinesin family were observed to be undergoing APC/C mediated degradation. The degradation of KIFC1, a kinesin protein, by APC/C was regulated by phosphorylation event. In the final systematic study to identify molecular mechanisms that may regulate SMN protein stability, its novel interaction with the E3 ligase APC/C in neurons was observed using proteomics. Also, we confirmed the earlier known interaction of HuD protein with SMN and provided new evidence that down-regulation of HuD protein concurs with low levels of SMN protein in neurons indicating that HuD may be stabilizing SMN protein. Interestingly, overexpression of HuD in the zebrafish model of SMA partially rescues the disease phenotype, thus suggesting the direct involvement of HuD protein in the SMA. Further, novel interaction of HuD with APC/C was observed using proteomics and SMN and HuD proteins were observed to be undergoing ubiquitination and degradation by APC/C. However, the mechanisms by which APC/C ubiquitinates these proteins were identified to be different. SMN degradation, *in vitro,* was regulated by APC/C-mediated ubiquitination, which in turn depends on phosphorylation of SMN protein. However, HuD

ubiquitination and degradation, *in vitro*, was regulated by the presence of a D-box motif in its sequence. We then employed multi-pronged approaches for inhibition of APC, including a small molecule inhibitor, proTAME, which increase the SMN protein abundance as well as that of HuD protein in mouse neurons. More importantly, the APC/C inhibition in the zebrafish model of SMA was shown to rescue the disease phenotype remarkably, along with a detectable increase in HuD and SMN protein levels. Thus, results from these studies have far-reaching implications for future studies on protein homeostasis in neurons and especially protein synthesis by non-canonical translational events. I also discussed the importance of identifying novel substrates of the APC/C in cancer such as KIFC1 as well as neuronal disorders such as HuD and SMN along with molecular mechanisms crucial in understanding the SMA disease and finding the potential cure.

Zusammenfassung

Proteinhomöostase ist die Regulierung des zellulären Proteoms durch Translation und durch Abbau von Proteinen und ist entscheidend für die normale Funktion der Zelle vorallem in hoch metabolischen und aktiven Zellen wie den Nervenzellen. Störungen der Proteinhomöostase wurden bererits mit verschiedensten neuronalen Erkrankungen in Verbindung gebracht. Abnormale Regulierungen von nur einem Protein, z.B. Survival of Motor Neurons (SMN), können neurodegenerative Krankheiten der spinalen Muskelatrophie (SMA) verursachen. In dieser Arbeit wurde eine auf quantitativer Massenspektrometrie basierenden high-troughput Methoden zur Untersuchung des Proteoms zusammen mit einer detailierten biochemischen Validierung ausgewählter Proteine durchgeführt um die Proteinhomöostase in Nervenzellen zu untersuchen. Das Ziel dieser Arbeit a) die aktivitätsabhängige Proteinsynthese in stimulierten Nervenzellen, b) den globalen Proteinabbau durch den Anaphase Promoting Complex/Cyclosome (APC/C) mit Hilfe von Ubiquitinierung und c) die detaillierte Regulierung des SMN Proteins gemeinsam mit dessen interagierenden Partner HuD durch APC/C im Zusammenhang mit SMA Erkrankungen zu untersuchen. Die erhaltenen Resultate zeigen einen neuen Einblick und führen zu einem besseren Verständnis der Regulierung von Translation und Abbau von Proteinen in Nervenzellen. Des Weiteren können diese Ergebnisse die Entwicklung neuer Therapeutika ermöglichen. Peptide bzw. translatorische Produkte von nicht kodierenden Regionen des Genoms in depolarisierten, kortikalen Nervenzellen der Maus wurden identifiziert. Die Menge an Peptiden wurde über die Zeit ähnlich reguliert wie bei bekannten Proteinen und dies deutet auf die Anwesenheit eines „neuen“ Proteoms hin welches biologische Funktionen besitzt kann. Mit Hilfe der Untersuchung des Proteinabbaus wurden neue Mitglieder der Kinesin Proteine, welche unerstuetzt durch APC/C abgebaut werden, identifiziert und validiert. Weiters wurde dadurch auch gezeigt, dass der Abbau des KIFC1 Proteins mittels APC/C abhängig ist von Phosphorilierungsereignissen. Während der gezielten Untersuchung des Proteoms zur Identifizierung des molekularen Mechanismus, welcher möglicherweise die Stabilität des SMN Protein reguliert, wurde eine neue Interaktion mit der E3 Ligase APC/C in Nervenzellen festgestellt. Die bereits bekannten Interaktionen des HuD Proteins mit SMN wurden ebenfalls bestätigt. Die erhaltenen Ergebnisse zeigen, dass die verminderte Expression des HuD Proteins mit einem geringen Level des SMN Proteins in Nervenzellen einhergeht, welches darauf hinweist dass HuD das SMN Protein stabilisiert. Die Überexpression von HuD im

Zebrafischmodell mit SMA verbessert teilweise den Verlauf des Krankheitsphenotyp und dieses unterstützt die Rolle des HuD Protein in SMA. Weiters wurden mit Hilfe von Proteomics neue Interaktionen von HuD mit APC gezeigt und es wurde entdeckt, dass SMN und HuD Proteine von APC/C ubiquitiniert und abgebaut werden. Jedoch zeigte sich, dass der Mechanismus mit welchem APC die Proteine ubiquitiniert unterschiedlich ist und das der Abbau von SMN in vitro durch Ubiquitinierung reguliert ist welche wiederum abhängig ist von der Phosphorylierung des SMN Proteins. Im Gegensatz dazu is die Ubiquitinierung von HuD und deren Abbau reguliert durch das Vorhandensein eines D-Box Motives in deren Sequenz. Anschliessend wurden mehrstufige Methoden zur Inhibierung von APC angewendet welche unter anderem den Inhibitor proTAME beinhalten. Dieser Inhibitor erhöht die Menge an SMN und HuD Protein in den Nervenzellen der Maus. Von grosser Bedeutung ist auch, dass gezeigt wurde, dass durch die Inhibierung von APC im Zebrafisch Modell mit SMA der Krankheitshenotyp gerettet werden kann einhergehend mit dem detektierbaren Erhöhung des HuD und SMN Proteinlevels. Diese Arbeit ermöglicht somit weitreichende Anwendungen für zukünftige Untersuchungen von Proteinhomöostase in Nervenzellen und im speziellen fuer die Proteinsynthese durch atypische Translationsereignisse. In dieser Arbeit wurde auch die Wichtigkeit der Identifikation von neuen APC/C Substraten in Krebs und neuronalen Störungen diskutiert sowie molekulare Mechanismen welche notwendig sind um SMA Erkrankungen zu untersuchen und potentielle Heilmittel zu finden.

Chapter 1

Introduction to Protein Homeostasis in Neurons

Contents **Page No.**

1. PROTEIN HOMEOSTASIS IN NEURONS

1.1 Introduction to protein homeostasis

The regulation of protein synthesis and degradation is critical to establish a sophisticated balance in the cellular proteome, also known as protein homeostasis, for the normal functioning of a cell. This balance is maintained by combinatorial regulation at multiple layers of 1) transcription - gene transcription, RNA stability and regulatory RNA interactions; 2) translation - RNA binding proteins interactions for transcript processing for protein synthesis and 3) degradation of unused or nonfunctional proteins by the ubiquitin-proteasome machinery. Therefore, protein homeostasis results from the dynamic and coordinated interaction of RNAs and proteins. Protein homeostasis accounts for the efficient rate of RNA translation depending on the accessibility of template RNAs as well as the activity of translation components including ribosomes, tRNA and other associated proteins [1]. Defective mRNAs are quickly cleared by degradation machinery such as exosomes, ribonucleases, nonsense-mediated mRNA decay, nuclear retention and transcriptional silencing [2]. Similarly, in the case of mis-folding events or the presence of degradation markers such as the D-box (RXXLXXN/D/E) recognition motif, proteins are targeted for the ubiquitination and degradation by proteasomes [3]. Failure to regulate these processes results in disruption of molecular and cellular functions, especially in neurons [4].

1.2 Protein homeostasis in neurons and neuronal diseases

Protein homeostasis is fundamental to maintain proper function in neurons as well as other cell types. The integrity and survival of neurons, in particular, is dependent upon the regulation of cellular processes such as dendritic branching, axonal guidance, synaptic plasticity and maintenance of neuromuscular junctions. Dynamic regulation of protein homeostasis must support these structural changes in response to changes in the environment due to extrinsic and intrinsic signals for the proper functioning of these non-dividing cells. As in other cell types, the amount of protein in neurons is regulated by gene transcription, translation and protein degradation. Due to the impact that each step can have on protein abundance and activity, a tight regulation of these steps is important to determine the fate of a given neuronal cell i.e. whether it will be directed to differentiation, death or survival.

In addition, neurons have uniquely polarized morphology presenting a distinct advantage to support the presence of sub-cellular RNAs and protein synthesis 'on demand' at the localized sites of synaptic connections, dendritic arbors and distal growth cones [4-9]. It was generally believed that signal peptides provide cues that allow proteins to travel from their site of synthesis in soma to the site of action in a larger cell like a neuron [10]. However, recently more sophisticated mechanisms such as 'localized translation' involving RNAs in complexes with RNA binding proteins (RBPs) are suggested to play a role in RNA transport across axons [11]. These RNA complexes transport to the 'zip code' destination [12], provide local signaling milieu and bring together several components of cellular function such as post-transcriptional processing, splicing and translation of RNAs [13]. This localized translation may provide an efficient mode of protein availability in specialized cells such as neurons with spatio-temporal precision in an energetically favorable manner to meet the biological need of the cell to maintain its proper functions and health [14]. Besides, it may provide local protein homeostasis in neurons where synaptic plasticity and de-novo synthesis of proteins in long and short information storage processes like learning and memory are ongoing processes that must also sustain at a distance from cell body. However, this field is still emerging and the mechanistic details of local translations are under investigation.

Almost all neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, Huntington's disease and Spinal Muscular Atrophy (SMA) disease are associated with disrupted protein homeostasis due to the mis-regulation at one or many regulatory layers of transcription, translation or degradation. Many of these neurodegenerative disorders are complicated syndromes manifesting as more than one symptomatic disease conditions, making them challenging in identifying a unique cause or cure. Yet, arguably there is hope for diseases like SMA, since it has a well understood genetic component, associated with a single protein 'survival of motor neurons' (SMN) that is implicated in the disease [15]. SMA disease is phenotypically characterized by the selective degeneration of lower motor neurons of the spinal cord and wasting of limb muscles; making this disease a leading cause of mortality in infants. Since the lack of SMN protein is known to be involved in the SMA disease physiology, an

increase in SMN protein abundance has been the primary focus to search for therapeutics for this disease [16].

1.3 Regulation of protein homeostasis

The complexity of an organism, to a large extent, is believed to be determined by its functional proteome. There are various factors that contribute to the abundance and functionality of the proteome such as the diversity of RNA species undergoing translation; regulation of transcription by non-translating RNA species; and alternative splicing of various RNAs.

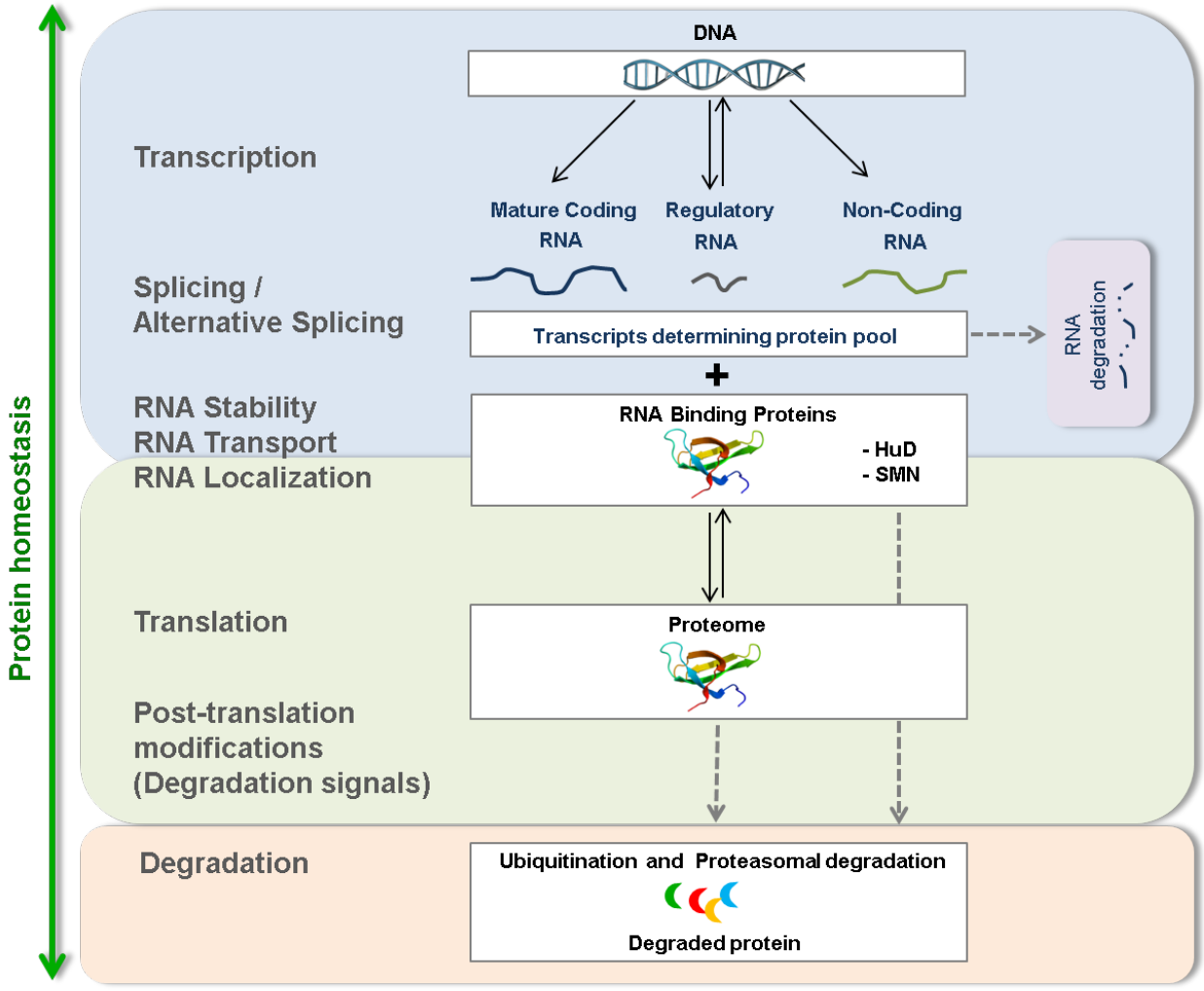


Figure 1: Protein homeostasis reflects a multi-layered regulation paradigm involving gene transcription, post-transcriptional regulation of the transcripts, translation and degradation of various RNA and protein populations.

Further, a number of factors add to the complexity of the cellular proteome such as transcript processing including stability, transport and localization due to interacting RBPs; post-translational modifications; and subsequent degradation of RNAs and proteins, including RBPs. These different aspects of protein homeostasis are illustrated in **Figure 1**. The multi-layered paradigm reflects the complexity associated with maintaining an intricate balance of the proteome in the cell.

In the following sections, I discuss the main components of protein homeostasis regulation in neurons at the RNA as well the protein level.

2. PROTEIN SYNTHESIS IN NEURONS

2.1 Canonical transcription and the RNA-centric view of proteome complexity

Human genome sequencing and subsequent analyses have revealed far fewer genes than expected. These studies have shown that there are only ~20,687 protein coding genes [17] roughly equal to the number of ORFs in mouse [18]; and slightly more than the 19,000 genes in *C. elegans* [19] and 13,600 genes in *Drosophila* [20] suggesting that complexity of humans across evolution requires alternative explanations. Less than 2 % the human genome has been annotated as protein-coding; it is possible that marked regions of the genome as 'gene' may not present a satisfactory model to explain the entire cellular proteome and its functional complexity.

Moreover, all somatic cells in humans carry the same set of genes while possessing different functional capabilities. For example, neurons have distinct capabilities to form networks with multiple cells, connecting diverse anatomical features of the body from sensory to skeletal using synapses. The human nervous system possesses >85 billion neurons in the brain, >13 million neurons in the spinal cord forming 1000-10,000 times more synapses for instant 'firing' (at speed ranging from 0.5 meter/sec to 120 meter/sec) or communication. Also, neurons are highly intricate cells in the human body with unique attributes of functional specialization including formation of synapses and neuromuscular junctions, and distinct morphology [21] including long axons (in some cases these can be one meter long e.g. in the case of spinal cord motor neuron innervating a foot muscle). The functional and physiological uniqueness of neurons may be

attributed to their dynamic proteome resulting from the delicate interplay between transcription, translation and degradation. Yet, it is not clear how neurons achieve their specialized complexity through functional proteome given the limited number of known protein coding genes.

The established central dogma of biology describes the flow of information from DNA to mRNA to protein by transcription and translation (**Figure 2**). This idea has been constantly evolving since its inception to accommodate the RNA-based regulation of cell processes. The presence of regulatory RNAs required to process transcripts into proteins such as ribosomal RNAs, transfer RNAs, small nuclear RNAs, small nucleolar RNAs and RNAs associated with spliceosome are few examples. These classical categories of RNAs are known to be transcribed from ~ 1 % of the total genome. Recent studies provide evidence for the presence of a large diversity of existing RNA species and a multitude of regulatory mechanisms that operate at a post-transcription level to regulate the proteome complexity and cellular functions. It is now clear that transcription in rather widespread involving ~90 % of genome. Taken together, these observations suggest that RNAs may be actively engaged in controlling cellular fate and are not resigned to the status of 'inert' messengers between DNA and protein.

Further, the text book definition of a gene is largely prescribed as a 'coding gene' focusing primarily on a defined structure of exons. The discovery of RNAs transcribed from the 'noncoding' part of genome requires a new definition of a gene; where a gene needs to be considered as a 'chromosomal segment' or 'transcription unit' responsible for making a functional product. This definition helps in bringing forth an unappreciated RNA-centric view of genome output in order to account for the comprehensive coding potential of genome as well as to explain the complexity of functional proteome.

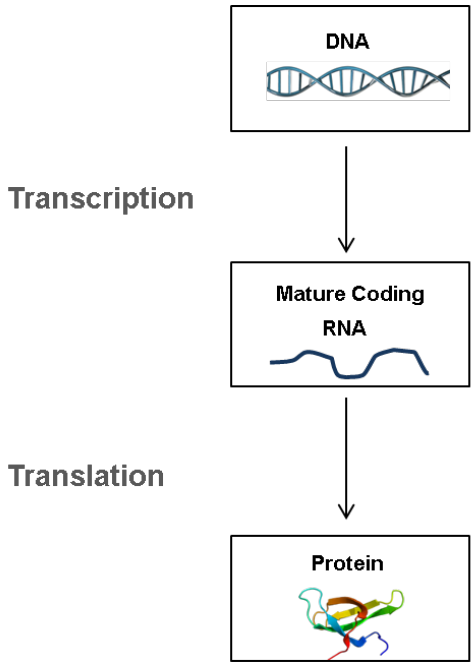


Figure 2: Canonical central dogma of biology. Flow of information from DNA to RNA to protein product by transcription and translation mechanisms.

2.1.1 Alternative splicing in neurons

The RNA-centric view of gene regulation is supported by post-transcriptional processing of pre-mRNAs by alternative splicing. Alternative splicing adds to the myriad RNA populations in order to expand the functional proteome from a limited genome size in response to different signaling events triggered within and from outside of cell [22, 23]. The phenomenon of alternative splicing affects the 74 % of all human genes [24] and is most prevalent in brain cells affecting many important classes of RNAs [23], and is also implicated in 15 % of genetic diseases including neurological disorders [25, 26]. For example, RNA binding proteins such as SMN protein and Hu family members are regulated by alternative splicing, and they in turn determine the downstream regulation of many target RNAs and their subsequent protein abundance in neurons, thus contributing to the complexity of the proteome in neurons. The four members of Hu family, namely HuA/R, HuB, HuC and HuD are isoforms generated by alternative splicing; and inclusion of exon 6 in HuD protein encodes for nuclear export signal, suggesting that alternative splicing may also regulate differential location ensuing interactions with disparate molecular networks of this isoform. Also, Hu proteins alter neuro-oncological

ventral antigen-1 (NOVA-1), a neuron-specific master splicing regulator that regulates the splicing of transcripts involved in the formation and activity of the synapses [27].

2.1.2 Local translation in neurons

The RNA-centric view of gene regulation is also supported by the presence of localized sub-cellular RNAs and protein synthesis. The presence of mRNAs in cytoplasm was believed to be a phenomenon associated with only a few specialized cells such as egg cells [28]. However, recent transcriptome studies have shown a widespread presence of various cytosolic mRNA species in neurons [6, 7, 29-32]. Interestingly, these mRNAs code for important classes of proteins such as receptors, channels, RNA binding proteins, molecular motors, cytoskeletal proteins, growth factors, translation factors, ribosomal proteins, ubiquitination enzymes and degradation-associated proteins etc. [33, 34]. Also, these mRNAs have been quantitatively detected in local compartments of growing axons and in dendrites [7, 35]. A number of these RNAs are shown to be regulated in neurons in an activity-dependent manner [36] and cases that undergo localized translation along with the details of specific cell types, cis-acting elements in 3' or 5' UTR or ORF with length of nucleotides acting as zipcode, trans-acting RNA-binding proteins helping in axonal and dendritic transport have been studied [37]. Some examples include β-actin [38, 39], myelin basic protein MBP [40, 41], kinesin [42-44], CaMKIIα [45, 46], MAP2 [47], BC1 [48], PKMζ [49], GluR2 [50], tau [51], vasopressin [9, 52] and shank [53]. These observations suggest the localized presence of RNAs associated with transport and local translation as well as regulation of genome output to further explain the functional complexity in neurons. However, given the pervasive presence of RNAs and RBPs, these studies are just the beginning and needed to be explored further.

2.2. Non-canonical transcription and their dysfunction in neuronal diseases

Recent investigations have highlighted RNA-based mechanisms that include various small RNA species in order to explain cellular complexity in neurons. These small regulatory RNAs potentially control the genome output by directly interacting with the structure of chromatin. Major players among this class of RNAs are called non-coding RNAs (ncRNAs). This has led to a new understanding of fundamental mechanisms in neurons by which protein synthesis and function is guided by transcriptional cues.

In prokaryotes such as *E. Coli* almost all the genome is composed of closely packed protein coding genes [54]. In contrast, human protein coding genes are interspersed with large tracts of DNA base pairs (97-98 %) assumed to be 'junk DNA' [17, 55, 56]. Even within a gene, there are large tracts of intron regions with demarcated boundaries containing di-nucleotides 5'-GT and AG-3' [57]. However, recent evidence shows that more than 80 % of the genome undergoes non-canonical transcription resulting in various categories of non-coding RNAs (ncRNAs) [58, 59] comprised of transcript products from introns of coding genes as well as introns and exons of non-coding genes.

The ENCODE (Encyclopedia of DNA elements) consortium project, with the goal to characterize functional genome, revealed the presence of thousands of ncRNAs. These ncRNAs were categorized as long non coding RNAs (lncRNAs) above the length of 200 nucleotides and small noncoding RNAs (sncRNAs) molecules below 200 nucleotides in length [60]. Many of these ncRNAs have been quantitatively analyzed [61] and studied for various mode of regulation to determine their role in various cellular processes **(Table 1)**.

ncRNAs	Estimated number	Mode of regulation	References
Micro (miRNAs)	>500	Activity dependent down regulation	[62, 63]
Piwi-interacting (piRNAs)	---	Neural activation, epigenetic regulation	[64]
Small nucleolar (snoRNAs)	1529	Up and down regulation of targets	[65]
Antisense (asRNAs)	4,545	Up and down regulation of targets	[66]
Long non-coding (lncRNAs)	25,000	Alternative splicing, epigenetic regulation	[67] [66]
Pseudogenes	13,419	RNA sponges and cell cycle arrest	[68]

Table 1: Categories of ncRNAs species with their estimated numbers and certain mode of regulation in neurons.

Structurally, many of the long ncRNAs are indistinguishable from mRNAs having 5'$m^7G_{PPP}N$ cap, polyA tail and exon-exon splice junctions; features that are associated with mRNA translation [69, 70]. Further, in-depth analysis of the human genome has been shown to encode

for more lncRNAs than annotated protein-coding genes. Many of these lncRNAs are known to associate with ribosomes [71] indicating their potential to be translated. In addition, the mechanisms involved in mRNA processing such as by polymerase II activity, histone modifications, transcription initiation and elongation (5′ capping and polyadenylation) are similar in lncRNA formation [72]. In contrast with mRNAs, however, lncRNAs are not restricted by the size of ORFs, directionality and strand specificity of transcription. In the brain, they are known to be actively spliced and associated with other ncRNAs and regulatory proteins.

Moreover, the ncRNAs have been shown to have regulatory functions such as inactivation of chromosomes, maintenance of telomere sequence, enhancer functions, transcription, alternative splicing and translation in turn regulating the mRNA and protein turnover at spatial as well as temporal levels. These roles are important to control cell function and survival, cell type identity and development, evolutionary expansion, disease inception and progress. The dysfunction of ncRNAs are associated with a number of neuronal disorders [73]. To list some:

- **Alzheimer's disease:** BACE1-AS is an antisense lncRNA that regulates mRNA and protein expression of β-secretase by preventing a miRNA-induced down-regulation of this enzyme, thus contributing to the severity of Alzheimer's disease [74].
- **Schizophrenia:** Gomafu is a lncRNA linked to aberrant splicing profiles in Schizophrenia by alternative splicing of genes DISC1 and ERBB4 [66] .
- **Autism spectrum disorder:** MSNP1–AS is an antisense transcript that can bind moesin pseudogene regulating the neuronal architecture and its overexpression is linked with Autism spectrum disorder [75].
- **Parkinson's disease:** miRNAs miR-7 and miR-153 inhibit α-synuclein levels, an aggregation prone neural protein, to reduce the pathogenesis in Parkinson's disease by binding to long 3′ UTR [76].
- **Angelman syndrome:** UBE3A-AS is an antisense transcript that regulates epigenetic silencing of paternal allele in neurons of Angelman syndrome; depletion of which activates expression of UBE3A ligase important for reducing pathological symptoms [77].

- **Huntington's disease:** HTT-AS is an antisense transcript that regulates expression of HTT gene, implicated in Huntington's disease due to CAG trinucleotides repeat expansion associated with impaired degradation of misfolded Huntington protein [77, 78].
- **Spinal Muscular Atrophy:** The neuron specific and important miRNA miR-9 controls axon growth and is mis-regulated in stem cell derived motor neurons with SMA. Moreover, the deletion of Dicer, RNase that regulates miRNA biogenesis and function, results in neuromuscular defects similar to that observed in SMA mouse model [79].

2.3 Non-coding RNAs and their protein coding potential

The main function of ncRNAs is believed to be associated with their conservation during evolution and direct influence on the expression state of interacting genes and their conservation during evolution. The ability of these ncRNAs to code for proteins has not been systematically evaluated. However, a number of studies have shown the presence of biologically important short protein coding regions e.g. Drosophila tarsal-less/polished rice gene was originally thought to be a lncRNA [80] but actually encodes a series of short peptides that convert the shaven baby transcription factor from repressor to activator during embryogenesis [81].

The rules describing the protein-coding mRNAs are comprised of minimum length Open Reading Frames (ORFs), biased codon usage and AUG start sites in transcripts [57], and are known to have many exceptions suggesting that many ncRNAs can also serve as a template to code for proteins. Similarly, the translation process itself is known to have alternative mechanisms including internal ribosome entry sites, initiation at non-AUG codons, leaky scanning, translational re-initiation and translational frame-shifts [82, 83] expanding the potential for alternative protein synthesis. Also, the products of translation are influenced by translation initiation at multiple sites in the transcript as well as translational pauses. Additionally, 'sloppiness' of the spliceosome can further support non-canonical translation by presenting substitute mechanisms such as alternative first or last coding exons, intron retention, exon skipping and / or splice site shifts. Taken together, these observations suggest that ncRNAs are capable of translation and may be contributing to 'hidden proteome' with potential functions in neurons.

Yet, the question of which of the potential ncRNAs are actually translated remains largely unknown. Recently, a number of ncRNAs have been shown to bind ribosomes in mouse embryonic cells using ribosomal profiling [84] indicating the presence of translation initiation sites. Mass spectrometry (MS) in recent years has made greater advances, and in particular, state of the art tandem MS can identify peptide sequences with high sensitivity and specificity in a high throughput manner. MS approaches provide a quantitative platform but are restricted to matching peptide spectra to the known and annotated proteome for protein coding genes, thus missing the unannotated noncoding regions of genome. Hence, an approach combined with genomics, termed as proteogenomics, is an important methodology to detect existing peptide population in cells and to trace them to relevant parts of genome. Recently, two large scale studies on human proteome were published using multiple datasets assembled from proteomics studies on human tissues, cell lines and body fluids [85] and the noncoding RNA sequence database NONCODE [86, 87] or the Broad Institute of MIT and Harvard's collection of large intergenic noncoding RNAs and transcripts of unknown coding potential [88]. These studies were able to provide a comprehensive view of various aspects of functional proteome such as estimation of the size of the protein coding genome, identification of tissue specific proteins and the identification of a number of novel peptides from lincRNAs (long intergenic non-coding RNAs) from collective datasets.

In order to ensure the discovery of novel peptides from ncRNAs in a complex mix of peptides from MS/MS proteomics, it is important to compare the complete set of transcriptome and proteome from same biological sample using the advanced technologies from both fields, and also to validate these novel peptides using stringent selection criteria and follow-up biochemistry methods to avoid false positives. We employed such a strategy in an activity dependent paradigm of 'stimulated' neurons that resulted in evidence suggesting the translation of various ncRNAs categories such as pseudogenes, overlapping alternative reading frames, extra-genic transcripts and anti-sense transcripts. I discuss these findings in the section in chapter 2 of the thesis.

3. RNA BINDING PROTEINS IN NEURONS

3.1 Role of RNA binding proteins in RNA stability

Stability of RNAs is a key factor in determining the variety, abundance and localized availability of proteins in neurons. The presence of RNA by itself as a 'naked' molecule is energetically unfavorable. The stability of transcripts is determined by trans-acting RBPs that perform various functions to bind, protect, translocate and scaffold RNAs for downstream post-transcriptional events. Therefore, the low abundance or absence of RBP can be deleterious for cells. In association with RBPs, the half-life of mRNAs can vary from highly unstable (a few minutes) to very stable (many hours) and determines the efficiency of translation. Thus understanding the functions and regulation of RBPs is necessary to understand their modulatory effect on cells in context of their survival and disease.

In neurons, around 323 RBPs have been identified [89]. Recently, another study in embryonic stem cells has significantly expanded the number of RBPs, adding a further 280 proteins to the list [90]. This list includes proteins from various functional categories of splicing regulators, transcription and translation regulators. Interestingly, a novel category of E3 ubiquitin ligases has been included recently as RBPs which are otherwise associated with protein degradation [90]. Immunoprecipitation-Sequencing (IP-Seq) approaches have revealed that some RBPs can bind hundreds of different mRNAs while many RNAs themselves have the potential to engage multiple RBPs using distinct binding motifs [91, 92]. The RBPs perform a variety of functions in the cell. Hence the interplay of their competitive interactions, often with contrasting regulations, can provide a sophisticated mechanism of fine-tuning the abundance and localization of RNAs in response to cell signaling.

HuD is a neuron-specific member of the Hu family and is a well-characterized RBP. HuD may serve as an excellent example to highlight various mechanisms used by RBPs to control protein homeostasis in neurons in general and in the context of SMA disease in particular. HuD RBP is known to bind ~600 mRNA involved in multiple neuronal processes e.g. axon guidance, neurite morphogenesis, neural development, differentiation and synaptic plasticity, which have been identified using IPs and microarrays. Hu family or ELAV (Embryonic Lethal Abnormal Vision)

like proteins are known to influence localization, stability and transport of a range of transcripts [93]. Hu members promote mRNA stability and enhance mRNA translation directly as well as by extending the mRNA half-life, thus regulating the protein levels.

The RBPs identify and bind various target RNAs by recognizing certain sequence elements as well as secondary structures motifs. For example, HuD binds 3'UTRs of target mRNA GAP43 [94] at AU-rich elements (ARE) using its RNA recognizing motifs (RRMs) 1 and 2 and a third RRM that interacts with polyA tail.

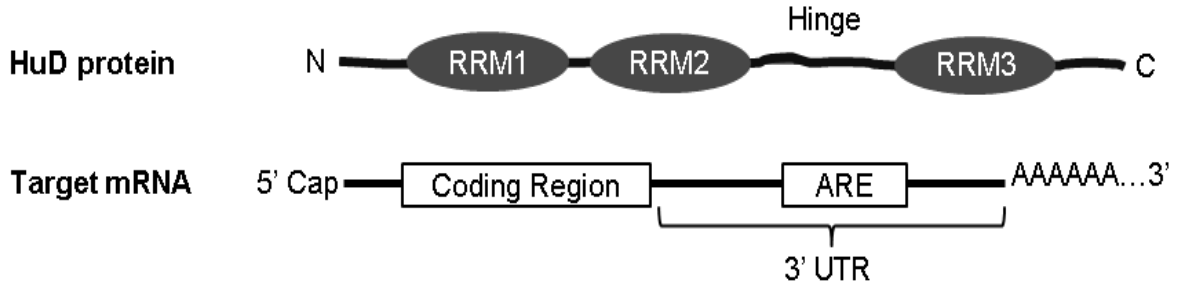


Figure 3: HuD RNA-binding protein has three RNA Recognition Motifs (RRM1-3) to bind target mRNAs. The hinge region is involved in the nucleo-cytoplasmic shuttling of the protein. Figure is adapted from [93].

3.2 Role of RBPs in RNA transport and local translation

RBPs regulate the translation of their interacting RNAs by influencing their stability and localization. Various interactions of HuD with target mRNAs and other RBPs such as SMN are known; facilitating further to investigate its role in regulating the downstream molecular events in neurons. In response to stimuli, a neuron undergoes various intracellular signaling events including activation of specific kinases. The binding of Hu proteins to specific RNA targets followed by their export to the cytoplasm is known to be regulated by various kinases such as MAPK and PKC [95-97] (**Figure 4**). In addition, CARM1 methyltransferase is known to modify HuD and inhibit its interaction with target RNAs like p21, thus affecting the cell decision between differentiation versus cell cycle exit [98]. In the cytoplasm, Hu proteins form comparatively large complexes, which are thought to be assembled in the cell body in RNA-protein granules. Hu proteins bind various mRNAs such as CPG15, GP43, TAU etc. and form ribo-nucleoprotein (RNP) complexes that relay a compendium of signals during differentiation in neurons [99]. Many studies have suggested that the RNA repertoire is dynamically regulated and

consistently changing over time [7, 31]. These RNA-protein complexes interact with components of cytoskeleton such as microtubules and microfilaments to be transported [100] and are also referred to as ‘cargo’. The trafficking of RNA cargo on microtubules utilizes various cytoskeletal motor proteins e.g. tubulin and kinesin along with translation repressors through axons [101, 102]. These cargos or RNA granules may be transported to various destination compartments in neurons such as dendrites and growth cones [103]. HuD complexes with target RNAs have been observed in various subcellular locations in neurons such as dendrites, soma, nucleus and synapses (**Figure 4**).

Interestingly, the events of neuronal activity such as upon KCl stimulation [36] or learning [104] promote translation initiation activity by increasing the association of RNA targets with polysomes and also by facilitating growth of neurites in response to stimuli [103]. Hu proteins closely interact with polysomes in neurons and directly regulate translation of target transcripts [105]. These observations suggest that HuD, and other RBPs, take part in RNA transport and their protein synthesis (**Figure 4**). Thus, a regulatory layer affecting stability and abundance of HuD will greatly determine the outcome in a neuronal cell.

Taken together, the above examples summarize various roles of RBPs: to bridge the processes of transcription and translation, interact closely with splicing machinery and also exchange molecules with protein degradation machinery of proteasome. Therefore, their presence is significant in proper functioning and survival of nervous system. It is plausible that perturbation in this layer of regulation in neurons may result in the altered proteome in neuronal diseases.

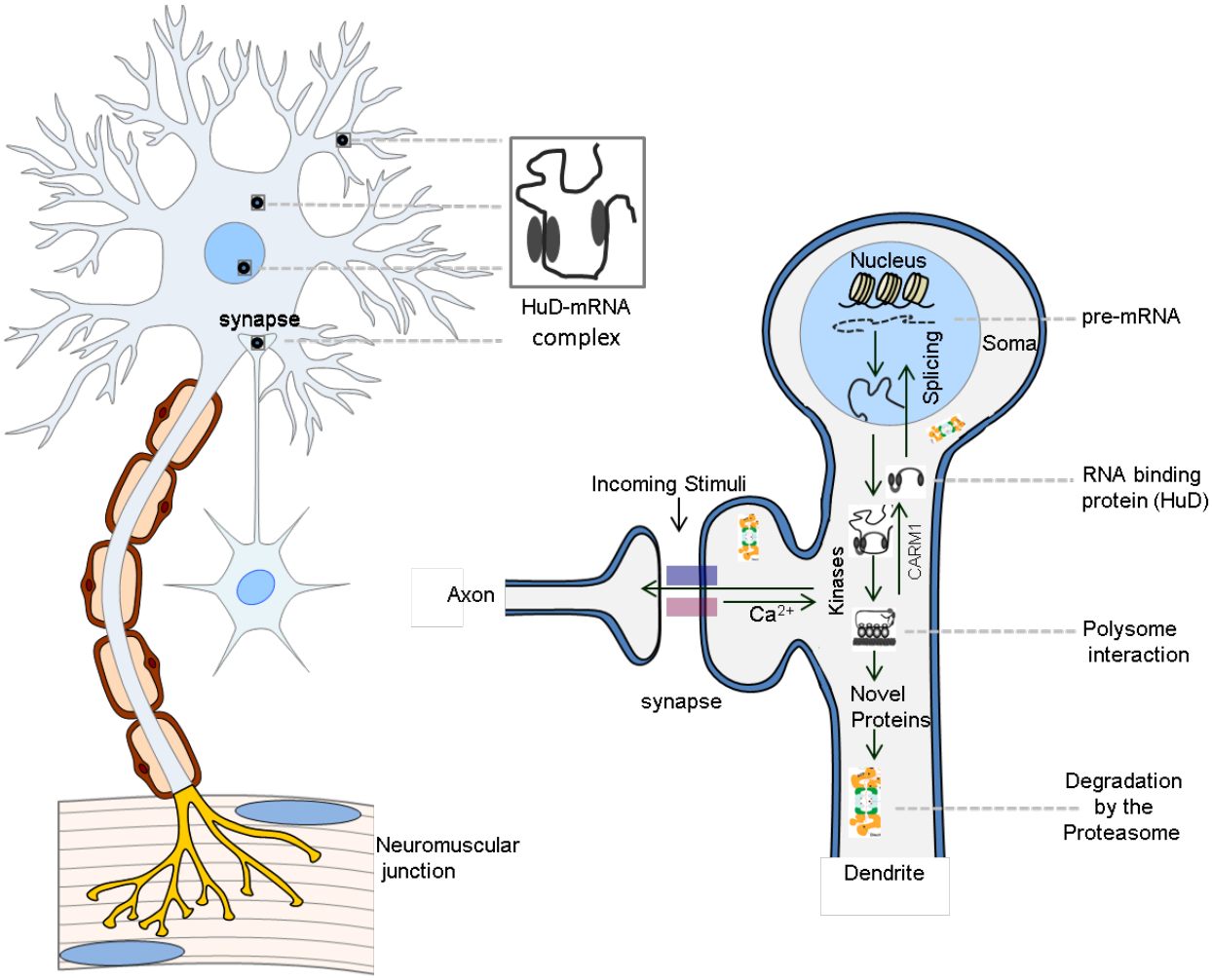


Figure 4: HuD interacts with target mRNAs and regulates protein synthesis and stability. HuD mRNAs complexes occupy various subcellular locations of neurons. HuD-RNA complex interacts with cytoskeleton and translational machinery to synthesize nascent target protein. Phosphorylation of HuD contributes to stabilization of mRNAs while CARM1 methylation suppresses HuD interactions with mRNA. HuD closely interacts with other RBPs including SMN protein in RNA complexes (not shown in Figure) and polysomes that help in protein synthesis of target RNAs. The nascent proteins as well as RBPs are targeted for ubiquitination and proteasomal degradation to maintain protein homeostasis in neurons.

3.3 Role of RBPs in neuronal diseases

Some RBPs may simultaneously interact to process the common target RNA, with similar effects e.g. SMN and HuD or opposite regulation-effects e.g. TIA1 (T-cell-restricted intracellular antigen 1) and HuD, on RNA regulation [106, 107]. The role of RBPs in neurodegenerative diseases is increasingly being investigated. Some RBPs that are found to be linked with SMA disease are discussed below to highlight their inter-connected regulatory actions in neurons, particularly in motor neurons:

- TIA-1 acts as a translation suppressor in RNA-protein aggregates called ‘stress granules’ formed in neurons experiencing ‘stress’ such as exitotoxicity. Interestingly, TIA-1 is also a positive regulator of SMN2 exon 7 splicing, that results in the loss of SMN function in SMA disease. HuD disrupts the translation repression activity of TIA-1 by splicing its mRNA to exclude the 3′-terminal exon. [108, 109].
- FMRP is ubiquitously expressed, yet it has an essential role in neuronal and intellectual development. This protein is associated with stability and translation of several genes in synaptic plasticity and is implicated in Fragile X syndrome, mental retardation, Autism and Parkinson's diseases. FMRP is also known to interact with SMN protein and is linked with SMA disease [110, 111].
- TDP-43 and FUS/TLS proteins play various roles in pre-mRNA splicing, mRNA trafficking, translation as well as in the biogenesis of microRNAs. These proteins are closely linked with motor neuron specific spliceosome defects and protein aggregates in ALS, FTD and SMA diseases [112, 113].
- KSRP promotes the degradation of several target transcripts encoding cytokines, chemokines, transcription factors, proto-oncogenes, cell-cycle regulators and also other RBPs in neurons such as SMN and hnRNP R [114].
- The hnRNP family promotes the degradation of several target transcripts implicated in several processes such as the cell cycle, stress response, apoptosis, transcription, splicing and translation and is thought to recruit mRNAs to the exosomes and proteasomes for degradation; making it another set of example of RBPs bridging translation and degradation machinery along with certain E3 ligases [115]. It is also implicated in ALS and FTD closely associated with protein aggregates, transcription, splicing and translation processes [89].
- SMN is a ubiquitous protein and plays an important role in biogenesis of the spliceosome by assembling small nuclear ribonucleoprotein complexes (snRNPs). The SMN protein facilitates pre-mRNA splicing, RNA stability and interaction with other RBPs such as HuD, hnRNPs to target RNAs for axonal transport and localization [116, 117]. These functions of SMN may be closely associated with survival of motor neurons, neurite-morphogenesis and proper functioning of synapses and neuromuscular junctions. SMN protein is translated from the SMN gene located in duplicated region of chromosome 5 in

human. The telomeric SMN1 gene is degraded by nonsense mediated decay (NMD) due to mutations. The centromeric copy of SMN2 gene shares >99 % sequence homology with SMN1, undergoes alternative splicing and exclusion of exon 7 in the presence of TIA-1 RBP, which is also known to interact with HuD as shown in independent studies [118]. Thus the alternatively-spliced and truncated SMN protein is unable to fold properly and is degraded in the cells. Lack of SMN protein, due to loss of the SMN1 gene and SMN2 protein dysfunction, results in SMA disease. As some of these mechanisms are shared between different motor neuron-specific neurodegenerative disorders it is important to investigate these in order to develop therapies.

- The HuD protein is known to interact with SMN protein using a dimethyl-arginine motif and the Tudor domain. Also, the HuD target of CPG15 mRNA is known to interact with SMN protein and overexpression of CPG15 partially rescues the SMA phenotype. Besides, SMN protein is believed to recruit HuD along with other RBPs and their target mRNAs into RNA granules. Together these proteins may result in chronic destabilization of RNAs, splicing defects and abnormality associated with neurons that is characteristic of SMA disease. It is important to investigate if HuD protein can modulate the levels of SMN protein by affecting its stability and if HuD may be involved in the SMA disease.

In this thesis, I have studied role of HuD and SMN proteins in the context of SMA disease. These proteins are important for motor neuron survival and lack of these proteins may result in the perturbed RNAs and proteins levels participating in the downstream molecular networks. I investigated the degradation mechanisms of these proteins by ubiquitination-proteasome system (UPS) and identified their novel interaction with E3 ubiquitin ligase, Anaphase promoting complex / cyclosome (APC/C). These results show that APC/C-UPS-mediated degradation mechanisms regulate the abundance of these proteins in normal cell and in the SMA disease state. I discuss these studies in chapters 4 and 5.

4. PROTEIN SYNTHESIS AND DEGRADATION AS INTER-RELATED MECHANISMS

The proteasomal link to transcription is better established with several proteins known to as common component in the regulation of transcription as well as processes related to UPS, albeit the exact mechanisms and physiological significance of these interactions are not completely understood. For example, CBP/P300 [119] acts as co-activator of APC/C ubiquitin ligase as well as in transcription, and UCH37 is a component of the Ino80 chromatin remodeler as well as a de-ubiquitinating enzyme [120]. Further, proteolysis limits transcription by targeting transcription factors and chromatin regulators, while ubiquitin ligases have been known to regulate RNA polymerase, histones and transcription co-activators at the site of transcription enabling a switch-like control of protein synthesis and degradation.

The dynamics of protein synthesis and degradation has been studied best in non-neuronal systems that exhibit time-dependent protein turnover e.g. cell cycle and circadian rhythm [121, 122]. Cyclin proteins are key regulators of the cell cycle, known to follow a rhythmic pattern of translation followed by degradation at the specific time in cell division. Although studies in neurons suggest a strong correlation between protein synthesis and degradation processes, a direct structural link between ribosomal and proteasomal complexes is not known. Due to the fast nature of action potential and synaptic transmission, changes at the protein level need to be uniquely time-bound and efficient in neurons [123]. Protein synthesis and degradation may need to be synchronized to maintain cellular homeostasis. Interestingly, eukaryotic translation initiation factor 3 (EIF3) has been shown to form a large complex that includes translation elongation factors, tRNA synthetases, ribosomal subunits, chaperones and the proteasome [124]; this association suggests that proteins involved in translation and degradation are physically linked to carry out their function in cells. In addition, the COP9/signalosome, an evolutionarily conserved complex, is present in dendrites [125] and modulates the exchange of various subunits between various E3 ubiquitin ligases and the EIF3 translation super-complex further supporting the evidence of direct interactions among these functionally disparate protein complexes [126]. Thus, it is plausible that similar interactions may exist to optimize translation in neurons.

5. PROTEIN DEGREDATION IN NEURONS

5.1. Introduction to protein degradation

The tight regulation of protein homeostasis is due to the interplay between protein synthesis and protein degradation. The half–lives of cellular proteins may vary considerably from minutes to several days and are regulated by degradation mechanisms. In eukaryotic cells, two pathways are known for protein degradation: autophagy-lysosomal proteolysis and the ubiquitin-proteasome system (UPS), which are also implicated in neurodegenerative disorders. These pathways carry out rapid turnover of proteins and clear out mis-folded and non-functional proteins to prevent disruption of essential pathways in the cell. Certain neurodegenerative disorders are associated with the accumulation of ubiquitinated protein aggregates that are otherwise rapidly degraded in healthy cells [127]. The autophagy-lysosome pathway is believed to be a nonspecific process involving vesicular trafficking of proteins and cellular organelles by first internalizing the proteins via endocytosis and then digesting them in the presence of protease enzymes. In contrast, UPS pathway involve enzymatic reactions and tightly regulates proteins, by targeting specific recognition sites and PTMs, including proteins involved in transcription, translation, RNA splicing and as RBPs in response to signaling events.

5.1.1 Ubiquitin-proteasome pathway in protein regulation

The UPS pathway is an important mechanism in eukaryotic cells responsible for 80 % of protein degradation by modifying proteins with ubiquitin to signal their degradation. Ubiquitin is a polypeptide of 76 amino acids that canonically modifies target proteins at lysine residues, N-terminus, K6, K11, K27, K29, K33, K46 and K48, but has also been demonstrated at other protein sites such as N-terminal amine group of a protein, the hydroxyl group of serine and threonine amino acids and the thiol group of cysteine amino acid [128]. The ubiquitination process occurs in a cascade of events involving three sequential steps: the E1 enzyme utilizes ATP to activate ubiquitin molecule and forms a high energy thiol-ester intermediate, an E2 enzyme interacts with ubiquitin at cysteine residue forming a thio-ester linkage and carries the activated ubiquitin to a specific E3 ligase enzyme.

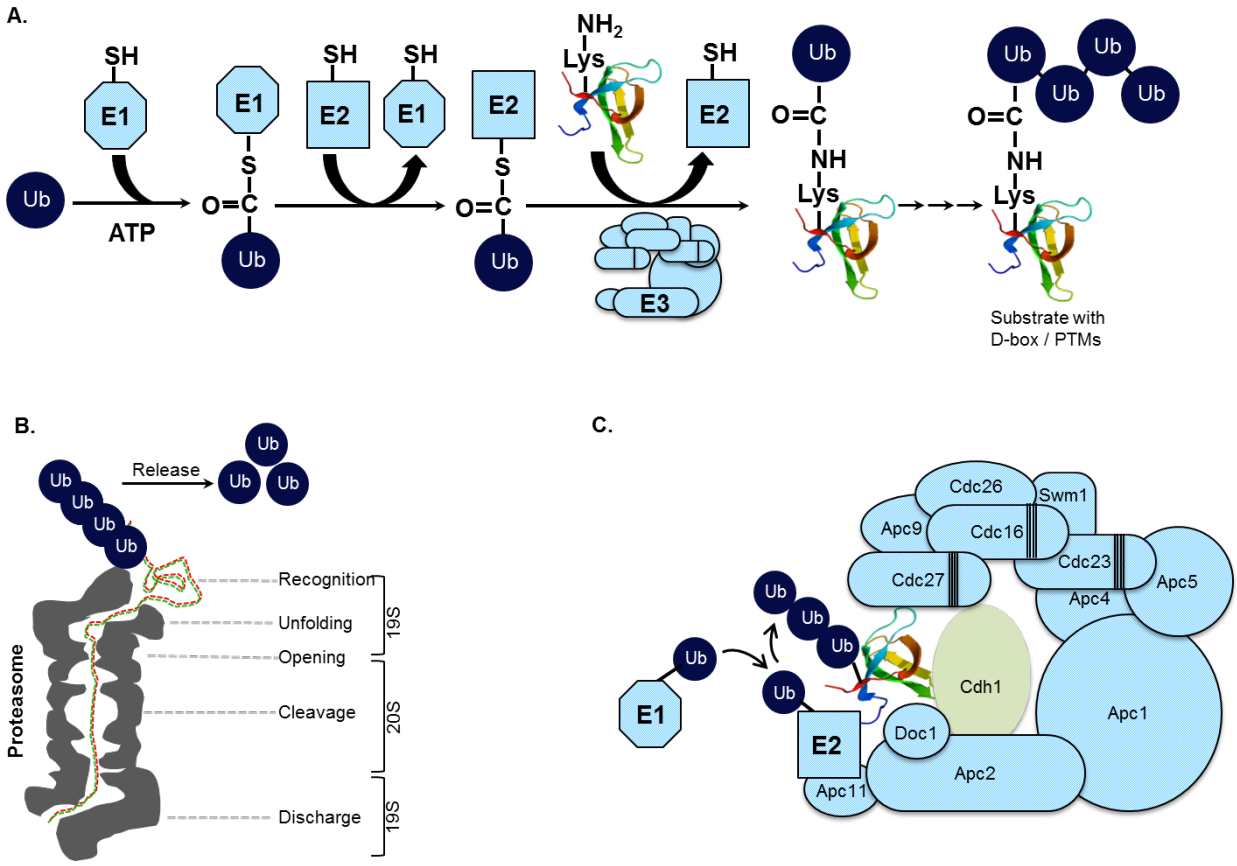


Figure 5: (A) Three enzymes (E1-E3) cascade of ubiquitination for the target substrate. **(B)** Proteasome showing its structural components with their protein recognizing and degradation functions. **(C)** E3 ligase, APC/C complex interacts with E2 and recognizes target proteins with D-box sequence element and / or PTMs in the presence of CDH1 activator to tag them with a chain of ubiquitin molecules. APC/C complex structure is adapted from [129].

E3 ligase enzymes are vital for providing the specificity in this process by recognizing sequence elements called 'recognition motifs' e.g. D-box on substrate proteins. The E3 ligase assists in the transfer of ubiquitin molecules to the substrate for poly-ubiquitination (**Figure 5A**).

The ubiquitinated substrates are recognized by the 26S proteasome, typically with a minimum of four ubiquitin molecules connected in series by isopeptide linkages between the C-terminus of one molecule and lysine of another molecule. Mammalian cells contain a few species-specific E1 enzymes, 40-50 E2s and many hundreds of E3 ligases. The 26S proteasome is a barrel-shaped organelle found ubiquitously in the cells and is constituted of a 20S central complex and two 19S lid complexes. The 19S complexes recognize the ubiquitinated substrates, de-ubiquitinate and unfold proteins using ATP and regulate access to the inner core of 20S subunit. The 20S core

carries out proteolytic activities using different enzymes such as trypsin, chymotrypsin and hydrolases [130-133].

5.1.2 Mechanisms and substrates of APC/C dependent protein regulation

There are four categories of E3 ligases classified as RING-domain, HECT-domain, U-box and PHD-domain families. RING-domain containing E3 ligases directly transfer ubiquitin to substrates, while HECT domain-containing E3 ligases form thiol-ester intermediates. The RING-domain family is the largest one comprised of the Skp1-Cullin-F-box (SCF) complex of four proteins namely Skp1, Rbx1, Cul1 and a variable protein containing F-box; and the Anaphase promoting complex/cyclosome (APC/C) complex [134, 135]. Among the families of E3 ligases available to a eukaryotic cell to carry out substrate specific degradation mechanisms, APC/C or APC is one of the best studied. The APC/C dependent regulation of proteins is also discussed in Chapter 3.

APC/C is a key regulator of cell cycle and is crucial in regulating the cell identity and survival. It is a large protein complex of 11-13 subunits including cullin (APC2) and RING (APC11) proteins along with various CDC proteins. CDC20 and CDH1 are the two known activators of APC/C; their interactions provide substrate specificity and timely regulations of cell cycle proteins in APC/C dependent manner (**Figure 5C**). Recently, APC/C is also shown to be active in human nervous system to regulate various processes (Table 3) in differentiated neuronal cells. APC/C is a member of the RING/CULLIN family of E3 complexes and directs the transfer of ubiquitin molecules from E2 to specific substrates. Substrates are typically identified by APC/C due to the presence of various recognition motifs in the protein sequence such as a KEN box, D-box etc. However, these motifs are neither necessary nor sufficient for degradation via APC/C as some APC/C substrates are targeted for degradation by PTMs like phosphorylation, such as in case of SECURIN and AURKA. During cell division, APC/C regulates various proteins such as CYCLIN and CYCLIN dependent kinases (CDKs) with defined periodicity of degradation events. The APC/C substrates in the cell cycle along with distinct cell processes they are involved in are given in **Table 2**.

Substrate	Function	Reference
ANLN	Cytokinesis	[136]
AURKA	Chromosome segregation and cytokinesis	[137]
AURKB	Chromosome segregation and cytokinesis	[138]
BIRC5	Chromosome segregation and cytokinesis	[139]
BUB1B	Mitosis progression	[140]
CALM2	Cytokinesis progression	[141]
CYCLIN A1	Control of G1/S and G2/M	[142]
CYCLIN A2	Control of S phase and G2/M transition	[142]
CYCLIN B1	Control of G2/M transition	[143]
CCNB2	Control of G2/M transition	[143]
CCNB3	Control of G2/M transition	[139]
CDCA3 (TOME-1)	F-box component of the SCF complex	[144]
CDC6	DNA replication	[145]
CDC20	Mitotic APC/C co-activator	[146]
CDC25	Cell cycle phosphatase	[147]
CDCA5 (Soronin, P35)	Stabilizes cohesin chromatin complex	[138]
P21	CDK inhibitor	[148]
CENPF	Chromosome segregation	[149]
CKAP2	Spindle regulator	[150]
CKS1B	Regulator of CDKs	[151]
CLSPN	Replication associated checkpoint protein	[152]
DNMT1	G2-M phase DNA methylation	[153]
ECT2	Spindle assembly	[139]

FAM64A	Metaphase to anaphase transition	[136]
FOXM1	S- and M-phase progression	[154]
FZR1 (CDH1)	APC/C co-activator	[139]
GEMININ	DNA replication inhibitor	[139]
GTSE1	G2-M phase arrest	[146]
HOXC10	Transcription factor	[139]
NEK2 (NLK1)	Centrosome regulator	[139]
NEK9	Chromosome segregation	[141]
NINL	Microtubule organization	[155]
NUSAP1	Microtubule associated protein	[156]
PFKFB3	Glycolytic enzyme	[157]
PLK1	Mitotic regulator	[158]
PRC1	Cytokinesis	[139]
SECURIN, PPTG1	Co-chaperone and separase inhibitor	[159]
KIF22	Spindle regulator	[160]
RCS1	Chromosome segregation	[161]
RPL30	RNA binding	[141]
RPS6KA4	Phosphorylation	[141]
RRM2	Catalyzes DNA synthesis	[162]
SGOL1	Chromosome cohesion	[163]
SKP2	F-box protein component of the SCF complex	[151]
TK1	dTTP production	[164]
TPX2	Spindle assembly regulator	[165]
TRB3	Akt Kinase inhibitor	[166]

UBE2C (UBCH10)	Ubiquitin-conjugating enzyme (E2)	[159]

Table 2: Substrates of APC/C in cell cycle and their functions. In cases of more than one names are commonly known for a protein, alternative names are given with in brackets.

The APC/C-UPS system of protein degradation is one of the major components of protein homeostasis. Finding novel proteins that may be recognized by the APC/C may provide insights into the regulation mechanisms of cell survival, cell death, malignancies and differentiation disorders. Typical detection of APC/C substrates involves cumbersome screening and laborious *in vitro* biochemistry [141]. However access to MS-based technologies such as multiplexed quantitative proteomics, as I discuss in Chapter 3, have made this task more efficient on a global scale of whole cell analysis. This strategy involves collecting protein 'hits' from different cell cycle phases and grouping them into clusters to identify proteins with similar abundance profiles as known APC/C substrates in the dataset. We identified novel substrates of the APC/C from the kinesin family and performed validation studies to confirm their degradation via the APC/C. In addition, we investigated novel mechanisms, based on the phosphorylation dependent regulation, that in turn control APC/C-mediated ubiquitination in these proteins. These mechanisms may provide important insight into APC/C regulation and may be involved in post-mitotic neurons playing key roles in survival and physiology in relation to neuronal disorders. I further explored and discussed these APC/C regulatory mechanisms in understanding the regulation of protein degradation in the context of neurons and the neurodegenerative disorder of SMA.

5.2. Role of APC/C as a neuronal E3 ligase for protein degradation

APC/C has emerged as one of the most important regulators of biological processes in neurons. As mentioned above, the APC/C E3 ligase a key component that controls transitions between cell cycle phases in dividing cells, yet remarkably APC/C is also expressed and remains active in post-mitotic differentiated neurons. In the nervous system, APC/C has been known to perform diverse functions e.g. regulating axon growth, dendrite branching, maintenance of synapses and neuromuscular junctions and neuronal survival [167].

The presence of cell cycle regulators such as APC/C in post-differentiated neurons was believed to be associated with the abrupt re-induction of anomalous cell proliferation pathways resulting in cell death, specifically in the context of neurodegenerative disorders. Interestingly,

knockdown of CDH1, an endogenous activator of APC/C, promotes robust growth in axons, thus suggesting the role of the APC/C in axonal growth. Also, SNON and ID2 are two direct substrates of APC/C ubiquitination that have been shown to increase axon growth. Another study in drosophila has shown that mutants of APC2 (a subunit of the APC/C complex) results in abnormal neuromuscular junctions. Moreover, in Alzheimer's disease, APC/C has been shown to regulate CYCLIN-B1 levels, which in turn modulate β-amyloid aggregates, resulting in cell death. Some of these examples of APC/C substrates investigated in neurons along with their functions are given (**Table 3**). Many of the these APC/C substrates are involved in neuronal networks with downstream targets, thus indicating the prominence and potential impact of this APC/C E3 ligase in neurons.

Substrate	Function	Reference
CYCLINB1	Cell death	[168]
FAS2	Glial migration	[169]
GLUR1	Homeostatic plasticity	[170]
ID1	Dendrite growth	[171]
ID2	Axon growth	[171]
LIPRIN-α	Regulation of synapse	[172]
NEUROD2	Pre-synaptic differentiation	[173]
PFKFB3	Glycolytic metabolism	[174]
SKP2	Neuronal differentiation	[175]
SNON	Axon growth	[176]

Table 3: Neurons specific substrates of APC/C with their functions.

The above observations indicate that the APC/C plays a crucial role in the regulation of fundamental aspects of neuronal development from morphogenesis to the regulation of mature neuronal connections via synapses (with neurons) and neuromuscular junctions (with muscles). Interestingly, proteasome inhibitors have been shown to have therapeutic potential in neuronal

disease. Bortezomib (Velcade, PS-341), a proteasome inhibitor, is approved by the US food and drug administration and is under investigation for clinical trials in the treatment of multiple myeloma [177]. MG132, a proteasomal inhibitor, has been proposed as a potent therapeutic to stall the degradation of target proteins in neurodegenerative disorders such as SMA. These observations suggest that inhibitors of specific protein degradation are crucial in re-establishing perturbed protein homeostasis in these diseases and need to be investigated as possible therapeutic targets. The specific APC/C inhibition in the context of neurodegenerative disorders may provide a better understanding of its functions in regulating its substrates.

5.3. Dysfunction of protein degradation by the UPS pathway and associated neuronal diseases

With multiple processes and proteins targeted by the ubiquitin pathway, it is not surprising to find that aberrations in the UPS underlie the pathogenesis of many diseases, particularly in relation with neurodegeneration. The pathological states associated with the ubiquitin system can be classified into two groups: 1) those that result from loss of function or mutation in a ubiquitin system enzyme such as an E3 ligase resulting in stabilization of target proteins and 2) those that result from gain of function or abnormal activity in E3 ligase enzymes resulting in accelerated degradation of protein targets. A few examples are discussed as below:

Angelman Syndrome is a brain development disorder characterized by mental retardation, seizures, out of context frequent smiling and laughter and abnormal gait. This disorder is linked with a loss of function mutation in UBE3a gene (a paternally imprinted gene) that codes for the E3 ligase UBE3A enzyme [178].

Muscle Atrophy: Various types of skeletal muscle atrophy and wasting are associated with a dysfunctional UPS resulting in long-term immobilization and denervation. The direct connections between muscle cells and neurons are reflected in their shared pathophysiology of anomalous protein turnover or aggregation, also associated with neurodegenerative disorders.

Neurodegenerative Diseases: In neurodegenerative disorders, neurofibrillary tangles and aggresome formation in Alzheimer's disease [179, 180], Lewy bodies in Parkinson's disease [181], aggregates in Amyotrophic Lateral Sclerosis [182], nucleotide-repeat nuclear inclusions in Huntington's disease [183] and Spino-Cerebellar Ataxias [184] are the featured pathological

signatures associated with a dysfunctional UPS. Inclusion bodies were thought to be product of inherent tendency of abnormal or truncated or mis-folded proteins to concatenate and aggregate. However, accumulation of ubiquitin conjugates along with proteasome and disease-specific proteins in these aggregates also indicate anomalous mechanisms related to UPS.

As described earlier, in SMA, the SMN protein is a known target of the UPS in neurons resulting in low amounts of SMN protein and the disease physiology. Hence, increasing the abundance of SMN protein in cells by stabilizing may be a potential therapy. In Chapter 4 and 5, I demonstrate that SMN protein stability is regulated by its interaction with HuD as well as APC/C, an E3 ubiquitin ligase known to be active in the cell cycle and neurons. APC/C directly targets SMN and HuD proteins for degradation by the UPS. Both ubiquitination assays and degradation assays show that the APC/C directly ubiquitinates SMN and that this ubiquitination is GSK3β dependent. Inhibition of the APC/C in mouse neurons and zebrafish model of SMA increases SMN protein levels and rescues the SMA disease phenotype.

6. AIM OF THESIS

The aim of this thesis is to explore novel mechanisms regulating protein homeostasis, including translation and degradation in neurons. Using stimulated neurons, we investigated the activity dependent response that yields translation products from ncRNAs such as antisense, pseudogenes, 5' UTR and intron retention etc. I used high-throughput global analyses followed by gene-specific validation approaches to verify these mechanisms. I used Spinal Muscular Atrophy as a disease model to establish that RNA processing (stabilizing, splicing and translation) proteins such as HuD and SMN undergo APC/C mediated ubiquitination and degradation.

In Chapter 2, I discuss that non-canonical translation occurs in stimulated neurons resulting in the novel population of peptides or proteins. These peptides or proteins come from various classes of ncRNAs including divergent bidirectional, antisense, intron retention and exon skipping etc. Pseudogenes are also found to be translated in these neurons. We further confirmed the presence of these novel peptides using multiple methods including synthetic peptides profile matching, RT-PCR expression, bioinformatics analyses and biochemical validations. A multiplexed quantitative proteomics approach shows that novel translation products are temporally regulated in a manner similar to known proteins in cells. These findings indicate the presence of unknown proteome that may be tapped in to find their novel putative functions in these neurons.

In Chapter 3, I discuss the APC/C mediated mechanisms involved in ubiquitination of proteins targeted for degradation by proteasomes. Cell cycle in synchronized HeLa cells was used to study the APC/C mediated regulation of proteins as well as mechanisms associated with ubiquitination such as phosphorylation cross-talk and/or presence of specific recognition motifs in protein sequence. The activation and inactivation of APC/C is a hallmark of tightly regulated protein homeostasis by stability and degradation of various protein in cell cycle. I discuss the quantitative high throughput proteomics approach to establish novel class of targets that undergo regulation by APC/C. Also, I validated these targets using *in vitro* biochemistry assays.

Having established that there can be various mechanisms by which APC/C recognizes and targets its substrates for ubiquitination, I investigated neuron specific RBP, HuD, for APC/C mediated degradation in the context of SMA disease. In chapter 4, I describe these results suggesting that HuD is directly regulated by APC/C in neurons. Further, HuD protein undergoes ubiquitination due to the presence of a recognition motif, D-box. The HuD D-box mutant remains stable and partially rescues the SMA disease. So far, only Plastin3 and SMN itself are known to rescue SMA disease *in vivo* zebrafish model; I discuss in Chapter 4 that HuD may act as modifier for SMN protein and SMA disease.

In Chapter 5, I explored SMN protein as direct target for APC/C ubiquitination and degradation. The mechanism by which SMN protein renders itself a substrate is distinct from that of HuD protein. In this study, I show that SMN protein has a regulation similar to kinesin proteins in cell cycle as discuss in Chapter 3. The phosphorylated SMN at Ser4, a site for GSK3β, degrades quickly in the presence of APC/C. This is an important finding as the knowledge of an E3 ligase that can specifically target SMN protein in SMA patients is important. I further tested a small molecule inhibitor of APC/C in various cell lines including stem cell differentiated motor neurons and primary motor neuron cultures from mouse and observed an increase in SMN protein. The APC/C inhibition by employing endogenous inhibitor rescues SMA disease remarkably *in vivo* experiments.

In Chapter 6, I discuss the overall impact of these findings in addressing some of the challenges mentioned in chapter 1. I discuss novel findings of non-canonical translation contributing to protein pool, RNA binding proteins such as SMN and HuD as modifiers of SMA disease and regulatory mechanisms of SMN and HuD stability by APC/C mediated ubiquitination and proteasomal degradation. Results from these studies suggest that the perturbation of protein homeostasis in neurons may involve multi-layer regulation of RNAs and proteins in cells and may result in severe neurodegenerative disorders such as SMA. These results have far reaching implications in revealing the novel mechanisms that will benefit basic neurobiology as well as may be explored further for therapeutic potentials in neuronal diseases.

ABBREVIATIONS

APC/C (APC)	Anaphase Promoting Complex
ARE	AU- rich element
AS	antisense
ATP	Adenosine 5’-triphosphate
BSA	Bovine Serum Albumin
D-box	Degradation box motif
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraaceticacid
HUD (ELAV)	Hu family protein member (Embryonic Lethal Abnormal Vision)
LC-MS/MS (MS)	Liquid Chromatography based Mass-Spectrometry
lncRNA	long non coding RNA
miRNAs	micro RNA
mRNA	messenger RNA
ncRNA	Non coding RNA
NMD	Nonsense mediated decay
piRNAs	Piwi-interacting RNA
PTM	Post Translational Modifications
RBP	RNA binding protein
RRM	RNA recognizing motif
rRNA	ribosomal RNA
SMA	Spinal Muscular Atrophy
SMN	Survival of Motor Neurons
sncRNA	small noncoding RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
tRNA	transfer RNA
UPS	Ubiquitin Proteasome System

REFERENCES

1. Brockmann, R., et al., Posttranscriptional expression regulation: what determines translation rates? PLoS Comput Biol, 2007. **3**(3): p. e57.
2. Eberle, A.B. and N. Visa, Quality control of mRNP biogenesis: networking at the transcription site. Semin Cell Dev Biol, 2014. **32**: p. 37-46.
3. Barford, D., Structural insights into anaphase-promoting complex function and mechanism. Philos Trans R Soc Lond B Biol Sci, 2011. **366**(1584): p. 3605-24.
4. Malgaroli, A., L. Vallar, and V. Zimarino, Protein homeostasis in neurons and its pathological alterations. Curr Opin Neurobiol, 2006. **16**(3): p. 270-4.
5. Campenot, R.B. and H. Eng, Protein synthesis in axons and its possible functions. J Neurocytol, 2000. **29**(11-12): p. 793-8.
6. Taylor, A.M., et al., Axonal mRNA in uninjured and regenerating cortical mammalian axons. J Neurosci, 2009. **29**(15): p. 4697-707.
7. Zivraj, K.H., et al., Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs. J Neurosci, 2010. **30**(46): p. 15464-78.
8. Perry, R.B. and M. Fainzilber, Local translation in neuronal processes--in vivo tests of a "heretical hypothesis". Dev Neurobiol, 2014. **74**(3): p. 210-7.
9. Mohr, E. and D. Richter, Messenger RNA on the move: implications for cell polarity. Int J Biochem Cell Biol, 2001. **33**(7): p. 669-79.
10. Imai, K. and K. Nakai, Prediction of subcellular locations of proteins: where to proceed? Proteomics, 2010. **10**(22): p. 3970-83.
11. Shav-Tal, Y. and R.H. Singer, RNA localization. J Cell Sci, 2005. **118**(Pt 18): p. 4077-81.
12. Singer, R.H., RNA zipcodes for cytoplasmic addresses. Curr Biol, 1993. **3**(10): p. 719-21.
13. Jansen, R.P., mRNA localization: message on the move. Nat Rev Mol Cell Biol, 2001. **2**(4): p. 247-56.
14. Jung, H., et al., Remote control of gene function by local translation. Cell, 2014. **157**(1): p. 26-40.
15. Tiziano, F.D., J. Melki, and L.R. Simard, Solving the puzzle of spinal muscular atrophy: what are the missing pieces? Am J Med Genet A, 2013. **161A**(11): p. 2836-45.
16. Burnett, B.G., et al., Regulation of SMN protein stability. Mol Cell Biol, 2009. **29**(5): p. 1107-15.
17. Lander, E.S., et al., Initial sequencing and analysis of the human genome. Nature, 2001. **409**(6822): p. 860-921.
18. Pennacchio, L.A., Insights from human/mouse genome comparisons. Mamm Genome, 2003. **14**(7): p. 429-36.
19. Consortium, C.e.S., Genome sequence of the nematode C. elegans: a platform for investigating biology. Science, 1998. **282**(5396): p. 2012-8.
20. Adams, M.D., et al., The genome sequence of Drosophila melanogaster. Science, 2000. **287**(5461): p. 2185-95.
21. Lopez-Munoz, F., J. Boya, and C. Alamo, Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramon y Cajal. Brain Res Bull, 2006. **70**(4-6): p. 391-405.

22. Lee, C., et al., ASAP: the Alternative Splicing Annotation Project. Nucleic Acids Res, 2003. **31**(1): p. 101-5.
23. Xu, Q., B. Modrek, and C. Lee, Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res, 2002. **30**(17): p. 3754-66.
24. Johnson, J.M., et al., Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. Science, 2003. **302**(5653): p. 2141-4.
25. Modrek, B. and C. Lee, A genomic view of alternative splicing. Nat Genet, 2002. **30**(1): p. 13-9.
26. Pan, Q., et al., Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet, 2008. **40**(12): p. 1413-5.
27. Ratti, A., et al., Post-transcriptional regulation of neuro-oncological ventral antigen 1 by the neuronal RNA-binding proteins ELAV. J Biol Chem, 2008. **283**(12): p. 7531-41.
28. Lasko, P., mRNA localization and translational control in Drosophila oogenesis. Cold Spring Harb Perspect Biol, 2012. **4**(10).
29. Andreassi, C., et al., An NGF-responsive element targets myo-inositol monophosphatase-1 mRNA to sympathetic neuron axons. Nat Neurosci, 2010. **13**(3): p. 291-301.
30. Cajigas, I.J., et al., The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron, 2012. **74**(3): p. 453-66.
31. Gumy, L.F., et al., Transcriptome analysis of embryonic and adult sensory axons reveals changes in mRNA repertoire localization. RNA, 2011. **17**(1): p. 85-98.
32. Minis, A., et al., Subcellular transcriptomics-dissection of the mRNA composition in the axonal compartment of sensory neurons. Dev Neurobiol, 2014. **74**(3): p. 365-81.
33. Poon, M.M., et al., Identification of process-localized mRNAs from cultured rodent hippocampal neurons. J Neurosci, 2006. **26**(51): p. 13390-9.
34. Zhong, J., T. Zhang, and L.M. Bloch, Dendritic mRNAs encode diversified functionalities in hippocampal pyramidal neurons. BMC Neurosci, 2006. **7**: p. 17.
35. Wu, K.Y., et al., Local translation of RhoA regulates growth cone collapse. Nature, 2005. **436**(7053): p. 1020-4.
36. Tiruchinapalli, D.M., M.D. Ehlers, and J.D. Keene, Activity-dependent expression of RNA binding protein HuD and its association with mRNAs in neurons. RNA Biol, 2008. **5**(3): p. 157-68.
37. Kindler, S., et al., RNA transport and local control of translation. Annu Rev Cell Dev Biol, 2005. **21**: p. 223-45.
38. Eom, T., et al., Localization of a beta-actin messenger ribonucleoprotein complex with zipcode-binding protein modulates the density of dendritic filopodia and filopodial synapses. J Neurosci, 2003. **23**(32): p. 10433-44.
39. Tiruchinapalli, D.M., et al., Activity-dependent trafficking and dynamic localization of zipcode binding protein 1 and beta-actin mRNA in dendrites and spines of hippocampal neurons. J Neurosci, 2003. **23**(8): p. 3251-61.
40. Hoek, K.S., et al., hnRNP A2 selectively binds the cytoplasmic transport sequence of myelin basic protein mRNA. Biochemistry, 1998. **37**(19): p. 7021-9.
41. Shan, J., et al., A molecular mechanism for mRNA trafficking in neuronal dendrites. J Neurosci, 2003. **23**(26): p. 8859-66.
42. Gioio, A.E., et al., Kinesin mRNA is present in the squid giant axon. J Neurochem, 1994. **63**(1): p. 13-8.

43. Kaplan, B.B., et al., beta-Actin and beta-Tubulin are components of a heterogeneous mRNA population present in the squid giant axon. Mol Cell Neurosci, 1992. **3**(2): p. 133-44.
44. Bassell, G.J., et al., Sorting of beta-actin mRNA and protein to neurites and growth cones in culture. J Neurosci, 1998. **18**(1): p. 251-65.
45. Neant-Fery, M., et al., A role for dendritic translation of CaMKIIalpha mRNA in olfactory plasticity. PLoS One, 2012. **7**(6): p. e40133.
46. Huang, Y.S., et al., Facilitation of dendritic mRNA transport by CPEB. Genes Dev, 2003. **17**(5): p. 638-53.
47. Crispino, M., et al., Messenger RNAs in synaptosomal fractions from rat brain. Brain Res Mol Brain Res, 2001. **97**(2): p. 171-6.
48. Muslimov, I.A., et al., Transport of Neuronal BC1 RNA in Mauthner Axons. J Neurosci, 2002. **22**(11): p. 4293-301.
49. Muslimov, I.A., et al., Dendritic transport and localization of protein kinase Mzeta mRNA: implications for molecular memory consolidation. J Biol Chem, 2004. **279**(50): p. 52613-22.
50. Ju, W., et al., Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. Nat Neurosci, 2004. **7**(3): p. 244-53.
51. Aranda-Abreu, G.E., et al., Embryonic lethal abnormal vision-like RNA-binding proteins regulate neurite outgrowth and tau expression in PC12 cells. J Neurosci, 1999. **19**(16): p. 6907-17.
52. Mohr, E., et al., Rat vasopressin mRNA: a model system to characterize cis-acting elements and trans-acting factors involved in dendritic mRNA sorting. Prog Brain Res, 2002. **139**: p. 211-24.
53. Falley, K., et al., Shank1 mRNA: dendritic transport by kinesin and translational control by the 5'untranslated region. Traffic, 2009. **10**(7): p. 844-57.
54. ayashi, T., et al., Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res, 2001. **8**(1): p. 11-22.
55. Venter, J.C., et al., The sequence of the human genome. Science, 2001. **291**(5507): p. 1304-51.
56. Wright, M.W. and E.A. Bruford, Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics, 2011. **5**(2): p. 90-8.
57. Brent, M.R., Genome annotation past, present, and future: how to define an ORF at each locus. Genome Res, 2005. **15**(12): p. 1777-86.
58. Djebali, S., et al., Landscape of transcription in human cells. Nature, 2012. **489**(7414): p. 101-8.
59. Hangauer, M.J., I.W. Vaughn, and M.T. McManus, Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. PLoS Genet, 2013. **9**(6): p. e1003569.
60. Consortium, E.P., et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature, 2007. **447**(7146): p. 799-816.
61. Mudge, J.M., A. Frankish, and J. Harrow, Functional transcriptomics in the post-ENCODE era. Genome Res, 2013. **23**(12): p. 1961-73.
62. Nudelman, A.S., et al., Neuronal activity rapidly induces transcription of the CREB-regulated microRNA-132, in vivo. Hippocampus, 2010. **20**(4): p. 492-8.

63. Vaishnavi, V., et al., Insights on the functional impact of microRNAs present in autism-associated copy number variants. PLoS One, 2013. **8**(2): p. e56781.
64. Rajasethupathy, P., et al., A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell, 2012. **149**(3): p. 693-707.
65. Rogelj, B., et al., Contextual fear conditioning regulates the expression of brain-specific small nucleolar RNAs in hippocampus. Eur J Neurosci, 2003. **18**(11): p. 3089-96.
66. Barry, G., Integrating the roles of long and small non-coding RNA in brain function and disease. Mol Psychiatry, 2014. **19**(4): p. 410-6.
67. Mercer, T.R., et al., Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci, 2010. **11**: p. 14.
68. Johnsson, P., et al., A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. Nat Struct Mol Biol, 2013. **20**(4): p. 440-6.
69. Dinger, M.E., et al., NRED: a database of long noncoding RNA expression. Nucleic Acids Res, 2009. **37**(Database issue): p. D122-6.
70. Derrien, T., et al., The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res, 2012. **22**(9): p. 1775-89.
71. Carninci, P. and Y. Hayashizaki, Noncoding RNA transcription beyond annotated genes. Curr Opin Genet Dev, 2007. **17**(2): p. 139-44.
72. Mercer, T.R., M.E. Dinger, and J.S. Mattick, Long non-coding RNAs: insights into functions. Nat Rev Genet, 2009. **10**(3): p. 155-9.
73. Ng, S.Y., et al., Long noncoding RNAs in development and disease of the central nervous system. Trends Genet, 2013. **29**(8): p. 461-8.
74. Davis, W., Jr., The ATP-binding cassette transporter-2 (ABCA2) increases endogenous amyloid precursor protein expression and Abeta fragment generation. Curr Alzheimer Res, 2010. **7**(7): p. 566-77.
75. Kerin, T., et al., A noncoding RNA antisense to moesin at 5p14.1 in autism. Sci Transl Med, 2012. **4**(128): p. 128ra40.
76. Doxakis, E., Post-transcriptional regulation of alpha-synuclein expression by mir-7 and mir-153. J Biol Chem, 2010. **285**(17): p. 12726-34.
77. Meng, L., et al., Truncation of Ube3a-ATS unsilences paternal Ube3a and ameliorates behavioral defects in the Angelman syndrome mouse model. PLoS Genet, 2013. **9**(12): p. e1004039.
78. Aronin, N. and M. DiFiglia, Huntingtin-lowering strategies in Huntington's disease: antisense oligonucleotides, small RNAs, and gene editing. Mov Disord, 2014. **29**(11): p. 1455-61.
79. Haramati, S., et al., miRNA malfunction causes spinal motor neuron disease. Proc Natl Acad Sci U S A, 2010. **107**(29): p. 13111-6.
80. Tupy, J.L., et al., Identification of putative noncoding polyadenylated transcripts in Drosophila melanogaster. Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5495-500.
81. Kondo, T., et al., Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis. Science, 2010. **329**(5989): p. 336-9.
82. Atkins, J.F. and P.V. Baranov, The distinction between recoding and codon reassignment. Genetics, 2010. **185**(4): p. 1535-6.
83. Namy, O., et al., A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature, 2006. **441**(7090): p. 244-7.

84. Ingolia, N.T., L.F. Lareau, and J.S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell, 2011. **147**(4): p. 789-802.
85. Kim, M.S., et al., A draft map of the human proteome. Nature, 2014. **509**(7502): p. 575-81.
86. Xie, C., et al., NONCODEv4: exploring the world of long non-coding RNA genes. Nucleic Acids Res, 2014. **42**(Database issue): p. D98-103.
87. Wilhelm, M., et al., Mass-spectrometry-based draft of the human proteome. Nature, 2014. **509**(7502): p. 582-7.
88. Cabili, M.N., et al., Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 2011. **25**(18): p. 1915-27.
89. Lenzken, S.C., et al., Neuronal RNA-binding proteins in health and disease. Wiley Interdiscip Rev RNA, 2014. **5**(4): p. 565-76.
90. Kwon, S.C., et al., The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol, 2013. **20**(9): p. 1122-30.
91. Licatalosi, D.D., et al., HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature, 2008. **456**(7221): p. 464-9.
92. Darnell, R.B., RNA protein interaction in neurons. Annu Rev Neurosci, 2013. **36**: p. 243-70.
93. Perrone-Bizzozero, N. and C.W. Bird, Role of HuD in nervous system function and pathology. Front Biosci (Schol Ed), 2013. **5**: p. 554-63.
94. Chung, S., et al., The Elav-like proteins bind to a conserved regulatory element in the 3'-untranslated region of GAP-43 mRNA. J Biol Chem, 1997. **272**(10): p. 6593-8.
95. Winzen, R., et al., The p38 MAP kinase pathway signals for cytokine-induced mRNA stabilization via MAP kinase-activated protein kinase 2 and an AU-rich region-targeted mechanism. EMBO J, 1999. **18**(18): p. 4969-80.
96. Ming, X.F., et al., Parallel and independent regulation of interleukin-3 mRNA turnover by phosphatidylinositol 3-kinase and p38 mitogen-activated protein kinase. Mol Cell Biol, 2001. **21**(17): p. 5778-89.
97. Wang, W., et al., AMP-activated kinase regulates cytoplasmic HuR. Mol Cell Biol, 2002. **22**(10): p. 3425-36.
98. Fujiwara, T., et al., CARM1 regulates proliferation of PC12 cells by methylating HuD. Mol Cell Biol, 2006. **26**(6): p. 2273-85.
99. Keene, J.D., Minireview: global regulation and dynamics of ribonucleic Acid. Endocrinology, 2010. **151**(4): p. 1391-7.
100. Hirokawa, N., mRNA transport in dendrites: RNA granules, motors, and tracks. J Neurosci, 2006. **26**(27): p. 7139-42.
101. Czaplinski, K. and R.H. Singer, Pathways for mRNA localization in the cytoplasm. Trends Biochem Sci, 2006. **31**(12): p. 687-93.
102. Krichevsky, A.M. and K.S. Kosik, Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation. Neuron, 2001. **32**(4): p. 683-96.
103. Smith, C.L., et al., GAP-43 mRNA in growth cones is associated with HuD and ribosomes. J Neurobiol, 2004. **61**(2): p. 222-35.
104. Bolognani, F., et al., Dendritic localization of the RNA-binding protein HuD in hippocampal neurons: association with polysomes and upregulation during contextual learning. Neurosci Lett, 2004. **371**(2-3): p. 152-7.

105. Antic, D. and J.D. Keene, Messenger ribonucleoprotein complexes containing human ELAV proteins: interactions with cytoskeleton and translational apparatus. J Cell Sci, 1998. **111 (Pt 2)**: p. 183-97.
106. Zhu, H., et al., A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators. Mol Biol Cell, 2006. **17**(12): p. 5105-14.
107. Akten, B., et al., Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10337-42.
108. Seo, J., et al., Spinal muscular atrophy: an update on therapeutic progress. Biochim Biophys Acta, 2013. **1832**(12): p. 2180-90.
109. Gilks, N., et al., Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol Biol Cell, 2004. **15**(12): p. 5383-98.
110. Piazzon, N., et al., In vitro and in cellulo evidences for association of the survival of motor neuron complex with the fragile X mental retardation protein. J Biol Chem, 2008. **283**(9): p. 5598-610.
111. Bassell, G.J. and S.T. Warren, Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. Neuron, 2008. **60**(2): p. 201-14.
112. Yamazaki, T., et al., FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. Cell Rep, 2012. **2**(4): p. 799-806.
113. Liu-Yesucevitz, L., et al., ALS-linked mutations enlarge TDP-43-enriched neuronal RNA granules in the dendritic arbor. J Neurosci, 2014. **34**(12): p. 4167-74.
114. Briata, P., et al., KSRP, many functions for a single protein. Front Biosci (Landmark Ed), 2011. **16**: p. 1787-96.
115. Bieler, S., et al., Low dose proteasome inhibition affects alternative splicing. J Proteome Res, 2012. **11**(8): p. 3947-54.
116. Fallini, C., G.J. Bassell, and W. Rossoll, Spinal muscular atrophy: the role of SMN in axonal mRNA regulation. Brain Res, 2012. **1462**: p. 81-92.
117. Sleeman, J., Small nuclear RNAs and mRNAs: linking RNA processing and transport to spinal muscular atrophy. Biochem Soc Trans, 2013. **41**(4): p. 871-5.
118. Kim, H.S., et al., Different modes of interaction by TIAR and HuR with target RNA and DNA. Nucleic Acids Res, 2011. **39**(3): p. 1117-30.
119. Turnell, A.S., et al., The APC/C and CBP/p300 cooperate to regulate transcription and cell-cycle progression. Nature, 2005. **438**(7068): p. 690-5.
120. Yao, T., et al., Distinct modes of regulation of the Uch37 deubiquitinating enzyme in the proteasome and in the Ino80 chromatin-remodeling complex. Mol Cell, 2008. **31**(6): p. 909-17.
121. Johnson, C.H., J.F. Roeber, and J.W. Hastings, Circadian changes in enzyme concentration account for rhythm of enzyme activity in gonyaulax. Science, 1984. **223**(4643): p. 1428-30.
122. Minshull, J., et al., The role of cyclin synthesis, modification and destruction in the control of cell division. J Cell Sci Suppl, 1989. **12**: p. 77-97.
123. Fonseca, R., et al., A balance of protein synthesis and proteasome-dependent degradation determines the maintenance of LTP. Neuron, 2006. **52**(2): p. 239-45.
124. Sha, Z., et al., The eIF3 interactome reveals the translasome, a supercomplex linking protein synthesis and degradation machineries. Mol Cell, 2009. **36**(1): p. 141-52.

125. Djagaeva, I. and S. Doronkin, Dual regulation of dendritic morphogenesis in Drosophila by the COP9 signalosome. PLoS One, 2009. **4**(10): p. e7577.
126. Luke-Glaser, S., et al., CIF-1, a shared subunit of the COP9/signalosome and eukaryotic initiation factor 3 complexes, regulates MEL-26 levels in the Caenorhabditis elegans embryo. Mol Cell Biol, 2007. **27**(12): p. 4526-40.
127. Ghavami, S., et al., Autophagy and apoptosis dysfunction in neurodegenerative disorders. Prog Neurobiol, 2014. **112**: p. 24-49.
128. McDowell, G.S. and A. Philpott, Non-canonical ubiquitylation: mechanisms and consequences. Int J Biochem Cell Biol, 2013. **45**(8): p. 1833-42.
129. Peters, J.M., The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol, 2006. **7**(9): p. 644-56.
130. Korolchuk, V.I., F.M. Menzies, and D.C. Rubinsztein, Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. FEBS Lett, 2010. **584**(7): p. 1393-8.
131. Lowe, J., et al., Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution. Science, 1995. **268**(5210): p. 533-9.
132. Heinemeyer, W., et al., The active sites of the eukaryotic 20 S proteasome and their involvement in subunit precursor processing. J Biol Chem, 1997. **272**(40): p. 25200-9.
133. Sorokin, A.V., E.R. Kim, and L.P. Ovchinnikov, Proteasome system of protein degradation and processing. Biochemistry (Mosc), 2009. **74**(13): p. 1411-42.
134. Ardley, H.C. and P.A. Robinson, E3 ubiquitin ligases. Essays Biochem, 2005. **41**: p. 15-30.
135. Berndsen, C.E. and C. Wolberger, New insights into ubiquitin E3 ligase mechanism. Nat Struct Mol Biol, 2014. **21**(4): p. 301-7.
136. Zhao, W.M. and G. Fang, Anillin is a substrate of anaphase-promoting complex/cyclosome (APC/C) that controls spatial contractility of myosin during late cytokinesis. J Biol Chem, 2005. **280**(39): p. 33516-24.
137. Honda, K., et al., Degradation of human Aurora2 protein kinase by the anaphase-promoting complex-ubiquitin-proteasome pathway. Oncogene, 2000. **19**(24): p. 2812-9.
138. Rankin, S., N.G. Ayad, and M.W. Kirschner, Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. Mol Cell, 2005. **18**(2): p. 185-200.
139. Pines, J., Mitosis: a matter of getting rid of the right protein at the right time. Trends Cell Biol, 2006. **16**(1): p. 55-63.
140. Qi, W. and H. Yu, KEN-box-dependent degradation of the Bub1 spindle checkpoint kinase by the anaphase-promoting complex/cyclosome. J Biol Chem, 2007. **282**(6): p. 3672-9.
141. Merbl, Y. and M.W. Kirschner, Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2543-8.
142. Huang, X., et al., Deubiquitinase USP37 is activated by CDK2 to antagonize APC(CDH1) and promote S phase entry. Mol Cell, 2011. **42**(4): p. 511-23.
143. Casaletto, J.B., et al., Inhibition of the anaphase-promoting complex by the Xnf7 ubiquitin ligase. J Cell Biol, 2005. **169**(1): p. 61-71.
144. Ayad, N.G., et al., Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. Cell, 2003. **113**(1): p. 101-13.

145. Mailand, N. and J.F. Diffley, CDKs promote DNA replication origin licensing in human cells by protecting Cdc6 from APC/C-dependent proteolysis. Cell, 2005. **122**(6): p. 915-26.
146. Pfleger, C.M. and M.W. Kirschner, The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev, 2000. **14**(6): p. 655-65.
147. Donzelli, M., et al., Dual mode of degradation of Cdc25 A phosphatase. EMBO J, 2002. **21**(18): p. 4875-84.
148. Yamada, K., et al., Degradation of p21Cip1 through anaphase-promoting complex/cyclosome and its activator Cdc20 (APC/CCdc20) ubiquitin ligase complex-mediated ubiquitylation is inhibited by cyclin-dependent kinase 2 in cardiomyocytes. J Biol Chem, 2011. **286**(51): p. 44057-66.
149. Gurden, M.D., et al., Cdc20 is required for the post-anaphase, KEN-dependent degradation of centromere protein F. J Cell Sci, 2010. **123**(Pt 3): p. 321-30.
150. Seki, A. and G. Fang, CKAP2 is a spindle-associated protein degraded by APC/C-Cdh1 during mitotic exit. J Biol Chem, 2007. **282**(20): p. 15103-13.
151. Bashir, T. and M. Pagano, Don't skip the G1 phase: how APC/CCdh1 keeps SCFSKP2 in check. Cell Cycle, 2004. **3**(7): p. 850-2.
152. Bassermann, F., et al., The Cdc14B-Cdh1-Plk1 axis controls the G2 DNA-damage-response checkpoint. Cell, 2008. **134**(2): p. 256-67.
153. Ghoshal, K., et al., 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. Mol Cell Biol, 2005. **25**(11): p. 4727-41.
154. Laoukili, J., et al., FoxM1 is degraded at mitotic exit in a Cdh1-dependent manner. Cell Cycle, 2008. **7**(17): p. 2720-6.
155. Wang, Y. and Q. Zhan, Cell cycle-dependent expression of centrosomal ninein-like protein in human cells is regulated by the anaphase-promoting complex. J Biol Chem, 2007. **282**(24): p. 17712-9.
156. Li, L., et al., NuSAP is degraded by APC/C-Cdh1 and its overexpression results in mitotic arrest dependent of its microtubules' affinity. Cell Signal, 2007. **19**(10): p. 2046-55.
157. Tudzarova, S., et al., Two ubiquitin ligases, APC/C-Cdh1 and SKP1-CUL1-F (SCF)-beta-TrCP, sequentially regulate glycolysis during the cell cycle. Proc Natl Acad Sci U S A, 2011. **108**(13): p. 5278-83.
158. Lindon, C. and J. Pines, Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells. J Cell Biol, 2004. **164**(2): p. 233-41.
159. Rape, M. and M.W. Kirschner, Autonomous regulation of the anaphase-promoting complex couples mitosis to S-phase entry. Nature, 2004. **432**(7017): p. 588-95.
160. Feine, O., et al., Human Kid is degraded by the APC/C(Cdh1) but not by the APC/C(Cdc20). Cell Cycle, 2007. **6**(20): p. 2516-23.
161. Zhao, W.M., et al., RCS1, a substrate of APC/C, controls the metaphase to anaphase transition. Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13415-20.
162. Cotto-Rios, X.M., et al., APC/CCdh1-dependent proteolysis of USP1 regulates the response to UV-mediated DNA damage. J Cell Biol, 2011. **194**(2): p. 177-86.
163. Karamysheva, Z., et al., Multiple anaphase-promoting complex/cyclosome degrons mediate the degradation of human Sgo1. J Biol Chem, 2009. **284**(3): p. 1772-80.

164. Ke, P.Y., et al., Control of dTTP pool size by anaphase promoting complex/cyclosome is essential for the maintenance of genetic stability. Genes Dev, 2005. **19**(16): p. 1920-33.
165. Stewart, S. and G. Fang, Destruction box-dependent degradation of aurora B is mediated by the anaphase-promoting complex/cyclosome and Cdh1. Cancer Res, 2005. **65**(19): p. 8730-5.
166. Ohoka, N., et al., Anaphase-promoting complex/cyclosome-cdh1 mediates the ubiquitination and degradation of TRB3. Biochem Biophys Res Commun, 2010. **392**(3): p. 289-94.
167. Puram, S.V. and A. Bonni, Novel functions for the anaphase-promoting complex in neurobiology. Semin Cell Dev Biol, 2011. **22**(6): p. 586-94.
168. Aulia, S. and B.L. Tang, Cdh1-APC/C, cyclin B-Cdc2, and Alzheimer's disease pathology. Biochem Biophys Res Commun, 2006. **339**(1): p. 1-6.
169. Silies, M. and C. Klambt, APC/C(Fzr/Cdh1)-dependent regulation of cell adhesion controls glial migration in the Drosophila PNS. Nat Neurosci, 2010. **13**(11): p. 1357-64.
170. Fu, A.K., et al., APC(Cdh1) mediates EphA4-dependent downregulation of AMPA receptors in homeostatic plasticity. Nat Neurosci, 2011. **14**(2): p. 181-9.
171. Lasorella, A., et al., Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth. Nature, 2006. **442**(7101): p. 471-4.
172. Teng, F.Y. and B.L. Tang, APC/C regulation of axonal growth and synaptic functions in postmitotic neurons: the Liprin-alpha connection. Cell Mol Life Sci, 2005. **62**(14): p. 1571-8.
173. Yang, Y., et al., A Cdc20-APC ubiquitin signaling pathway regulates presynaptic differentiation. Science, 2009. **326**(5952): p. 575-8.
174. Rodriguez-Rodriguez, P., A. Almeida, and J.P. Bolanos, Brain energy metabolism in glutamate-receptor activation and excitotoxicity: role for APC/C-Cdh1 in the balance glycolysis/pentose phosphate pathway. Neurochem Int, 2013. **62**(5): p. 750-6.
175. Zhang, Y., et al., The involvement of down-regulation of Cdh1-APC in hippocampal neuronal apoptosis after global cerebral ischemia in rat. Neurosci Lett, 2011. **505**(2): p. 71-5.
176. Stroschein, S.L., et al., Smad3 recruits the anaphase-promoting complex for ubiquitination and degradation of SnoN. Genes Dev, 2001. **15**(21): p. 2822-36.
177. Chen, D., et al., Bortezomib as the first proteasome inhibitor anticancer drug: current status and future perspectives. Curr Cancer Drug Targets, 2011. **11**(3): p. 239-53.
178. Matsuura, T., et al., De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat Genet, 1997. **15**(1): p. 74-7.
179. Riederer, B.M., et al., The role of the ubiquitin proteasome system in Alzheimer's disease. Exp Biol Med (Maywood), 2011. **236**(3): p. 268-76.
180. Upadhya, S.C. and A.N. Hegde, Role of the ubiquitin proteasome system in Alzheimer's disease. BMC Biochem, 2007. **8 Suppl 1**: p. S12.
181. Ebrahimi-Fakhari, D., P.J. McLean, and V.K. Unni, Alpha-synuclein's degradation in vivo: opening a new (cranial) window on the roles of degradation pathways in Parkinson disease. Autophagy, 2012. **8**(2): p. 281-3.
182. Strong, M.J., S. Kesavapany, and H.C. Pant, The pathobiology of amyotrophic lateral sclerosis: a proteinopathy? J Neuropathol Exp Neurol, 2005. **64**(8): p. 649-64.
183. Imarisio, S., et al., Huntington's disease: from pathology and genetics to potential therapies. Biochem J, 2008. **412**(2): p. 191-209.

184. Wilson, S.M., et al., Synaptic defects in ataxia mice result from a mutation in Usp14, encoding a ubiquitin-specific protease. Nat Genet, 2002. **32**(3): p. 420-5.

Chapter 2

Quantitative Profiling of Peptides from RNAs classified as noncoding

Contents **Page No.**

SUMMARY

Only a small fraction of the mammalian genome codes for messenger RNAs destined to be translated into proteins and it is generally assumed that a large portion of transcribed sequences, including introns and several classes of non-coding RNAs (ncRNAs), do not give rise to peptide products. A systematic examination of translation and physiological regulation of ncRNAs has not been conducted. In this study, we identified the products of non-canonical translation in mouse neurons by analyzing unannotated transcripts in combination with proteomics data. This study supports the existence of non-canonical translation products from both intragenic and extragenic genomic regions, including peptides derived from anti-sense transcripts and introns. Moreover, the studied novel translation products exhibit temporal regulation similar to that of proteins known to be involved in neuronal activity processes. These observations highlight a potentially large and complex set of biologically regulated translational events from transcripts formerly thought to lack coding potential.

INTRODUCTION

Recent genome-wide transcriptome studies have shown that tens of thousands of loci outside of the defined protein coding regions are transcribed [1-3]. The resulting transcripts include a plethora of species such as 5' leader sequences and 3' end regions, introns, micro RNAs (miRNAs) [4], enhancer RNAs (eRNAs) [5], small nuclear RNAs (snRNA) [6], anti-sense transcripts [7-9] and various other short and long RNAs [10]. The identification of these transcripts implies a complexity previously unappreciated and has led to the emergence of significant efforts investigating the roles of RNA species traditionally referred to as non-coding sequences [11, 12]. In parallel to the identification of transcribed regions throughout the genome, proteomic studies have shown that a fraction of the 'high-quality spectra' from mass-spectrometry (MS) based proteomics experiments do not match annotated proteins [13]. This led us to hypothesize that some of these unmatched spectra could represent uncharacterized translation events derived from transcribed regions outside coding genes. Consistent with this hypothesis, recent studies have shown that non-coding transcripts are associated with ribosome [14] and that some non-coding transcripts lead to translation of short open reading frames [15].

Here we report the results of a systematic study that we undertook to investigate and evaluate the existence of non-canonical translation products and their biological regulation under physiological conditions combining RNA-seq and quantitative MS approaches. The results of this study not only indicate the presence of a large number of previously undetected protein translational products but they also show that they are temporally regulated suggesting more complex regulatory biochemical mechanisms that have not been previously observed.

EXPERIMENTAL PROCEDURES

Brief overview of the experimental design and MS analysis

Neuronal cultures were grown, depolarized with potassium chloride (KCl) for 0, 1, 2, 3 and 6 h (three biological repeat experiments were performed independently). Proteins were extracted and digested into peptides from neuronal cell lysates of each time point. Peptides from the 5 time-points in each experiment were labeled with 5 different isobaric MS labels as follows: TMT-127 for 0 h, TMT-128 for 1 h, TMT-129 for 2 h, TMT-130 for 3 h and TMT-131 for 6 h (TMT-126 was used to label pre-stimulated neurons). After labeling, the peptides were pooled together and separated based on their isoelectric point differences into 24 fractions. Hence a total of 24 x 3 = 72 fractions were collected and analyzed by MS. Each sample was analyzed 3 times by LC-MS/MS (at least 3 technical repeats were performed resulting in 238 MS raw data files (with a few re-analyzed). These 238 files were grouped and then searched against mouse database. Un-matched peptides to the mouse database were then exported and searched against the custom RNA seq database in 6 frames. After stringent filtering of the results, peptides designated as novel peptides were further validated by computational and biochemical methods including synthetic peptide matching, RT-PCR, western blotting (WB), Immune-precipitation (IP) confirmed by MS and bioinformatics analysis. Each experiment is described in detail below.

Animals

All experimental procedures were performed in compliance with animal protocols approved by the IACUC at Children's Hospital Boston, Boston, MA. We used embryos at age embryonic day (E) 16.5 from the C57BL6 mouse strain for neuronal cultures.

Mouse cortical cultures

Cortices of the mice embryos (C57BL/6, Charles River) at stage E16.5 were dissected and dissociated in 1x Hank's Balanced Salt Solution (HBSS) (Life Technologies), 100 mM MgCl2, 10 mM Kynurenic acid, 100 mM HEPES, 20 mg/ml trypsin (Worthington Biochemicals) and 0.32 mg/ml L-cysteine (Sigma) for 10 min. Trypsin treatment to dissociate cells was terminated with three 2 min washes in 1x HBSS with 10 mg/ml trypsin inhibitor (Sigma). Cells were triturated with a flame-narrowed Pasteur pipette for complete dissociation. Neurons were seeded at an approximate density of $4x10^7$ on 15 cm dishes. The dishes were pre-coated overnight with 30 µg/ml poly-ornithine (Sigma) in water, washed three times with autoclaved water and washed once with Neurobasal Medium (Life Technologies) before use. Neurons were maintained in 30 ml filtered Neurobasal Medium containing 1 M Glucose, B27 supplement (Invitrogen), penicillin-streptomycin (50 µg/ml penicillin, 50 U/ml Streptomycin, Sigma) and 1 mM Glutamine (Sigma). Neurons were grown *in vitro* for 7 days. Eight ml of the medium was replaced with 10 ml fresh warm medium on the 4th and 6th days.

Potassium chloride depolarization of neurons

Neuronal cultures at day 6 were treated overnight with 1 µM Tetrodotoxin (TTX, Tocris) and 100 µM D(-)-2-amino-5phosphonopentanoic acid (D-AP5, Tocris) and cells were collected at 0, 1, 2, 3 or 6 h after incubation with 55 mM KCl in culture media.

TMT labeling and peptide fractionation

Cells were lysed on ice for 10 min using lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 % NP-40, 1 mM PMSF) containing protease inhibitors (Roche-complete protease inhibitor cocktail tablets) and phosphatase inhibitors (Sigma phosphatase inhibitor cocktails I and II). Cells were passed through a 27 G needle multiple times, sonicated briefly and centrifuged for 30 min, 4 °C at 20,000 g to clear the solution. Clear lysates were collected, transferred to new tubes and protein concentrations estimated using a BCA assay kit (Thermo Scientific). Protein precipitation and digestion was carried out as described by Winter et al., 2011 [16]. Briefly, 100 µg of protein was precipitated from each time point using 1 ml of ice cold chloroform/methanol. The pellets were re-dissolved in 0.1 % Rapigest (Waters) in 100 mM TEAB (triethyl ammonium bicarbonate (Sigma) and incubated at 37 °C for 15 min. Trypsin (Promega) was added to the

samples and incubated at 37 °C for 45 min to dissolve the pellet. Samples were reduced with 20 mM DTT for 60 min at 56 °C and alkylated with 1 % acrylamide for 45 min at room temperature (RT). Further trypsin was added to a final enzyme to protein ratio of 1:100 and the mixture was incubated at 37 °C overnight. Peptide samples from each time point were acidified using 5 µl trifluoroacetic acid (TFA) and incubated for 45 min at 37 °C in order to precipitate the RapiGest followed by centrifugation for 30 min at 20,000 g. Clear supernatants were desalted using Oasis HLB cartridges (Waters). Briefly, the columns were washed twice with 70 % ACN 0.1 % Formic acid and twice with 0.1 % FA. The sample pool was passed twice through each individual column, washed with four times with 0.1 % FA and eluted twice with 30 % ACN 0.1 % FA, twice with 50 % ACN 0.1 % FA and twice with 70 % ACN containing 0.1 % FA. Individual eluate fractions were pooled and samples dried in a vacuum centrifuge. Dried samples were re-suspended in 0.1 M TEAB and peptide concentrations quantified using the BCA assay. For each sample, peptides from different time-points after KCl stimulation were labeled with one of the 5 TMT labels (Thermo Fischer Scientific) for 3 h at room temperature (RT) following the manufacturer's protocol. The samples were combined, partially dried using a vacuum centrifuge and desalted using Oasis HLB cartridges as described above. The dried peptides were resuspended in ampholyte solution (pI 3-10) and fractionated overnight into 24 fractions based on their isoelectric point using an OFFGEL fractionator (Agilent) according to the manufacturer’s instructions. The fractions were desalted and analyzed using LC-MS/MS.

MS analysis

Peptide samples were loaded directly onto an in house packed reverse phase column using 5 µm, 200Å particles (magic C18, Michrom) and PicoTip Emmiters (New Objective) with an autosampler / nanoLC setup (2D nanoLC, Eksigent) at a flow rate of 1 µl/min. After loading the column was washed for 5 min at 1 µl/min at 99 % A (water with 0.2 % FA) 1 % B (acetonitrile with 0.2 % FA) followed by elution with a linear gradient from 1 % B to 35 % B at 400 nl/min in 60 min. Peptides eluting from the column were ionized in the positive ion mode and the 6 most abundant ions were fragmented in the PQD-mode [17] to allow for the detection of low mass range reporter ions. Briefly, the LTQ-Orbitrap was run in positive ion mode. Full scans were carried out with a scan range of 395 to 1200 m/z. Normalized collision energy of 35 was used to activate both the reporter ions and parent ions for fragmentation. Scans were carried out with an

activation time of 30 ms. The isolation window was set to 1.0 m/z.

Testing labeling efficiency

For each of the TMT channels equal amounts of protein, 100 µg, were used. Labeling bias was tested by assessing the $\log_2$ intensities from all the channels. The median across the channels had a standard deviation of 0.16 and a coefficient of variance of 0.02, suggesting that there is no inherent mixing irregularities in total pool of labeled sample and the differential peptide abundance observed among time-points are true observations (**Figure 1**). In addition TMT labeling efficiency was evaluated to be 99.5 % of all unique and high-confidence peptides of which 98.3 % are fully labeled (labeled on N-terminal and internal Lysine residues) with TMT reporter ions. The labeling efficiency of fully labeled peptides and the median intensities calculated for each channel indicate that there is no quantitative bias that would affect the results reported in the profile data.

MS data analysis

The proprietary Thermo Scientific .raw files (238 in total as explained above) were converted into 6 .mgf files and MS/MS data was queried against the Uniprot Feb 2012, (canonical and isoform sequences) protein sequence database, containing common contaminations and concatenated to its decoy version, using MASCOT v2.3 (Matrix Science). TMT peptides were searched with enzyme specificity trypsin, two missed cleavage site, carbamidomethyl (Cys), oxidation, deamidation (N) and Gln to pyro Glu (N-terminal Q), phosphorylations on S/T/Y and TMT-6plex (N-termini and Lys) as variable modifications. Only 226,471 (20 %) of the spectra were matched to the mouse database resulting in 15,516 peptides that were unique. They were grouped into 3,284 proteins with a 1 % protein false-discovery rate (FDR) cut-off for proteins and at least 2 peptides per protein were identified. This corresponded to a Mascot cut-off score of 26.93 and above for each protein.

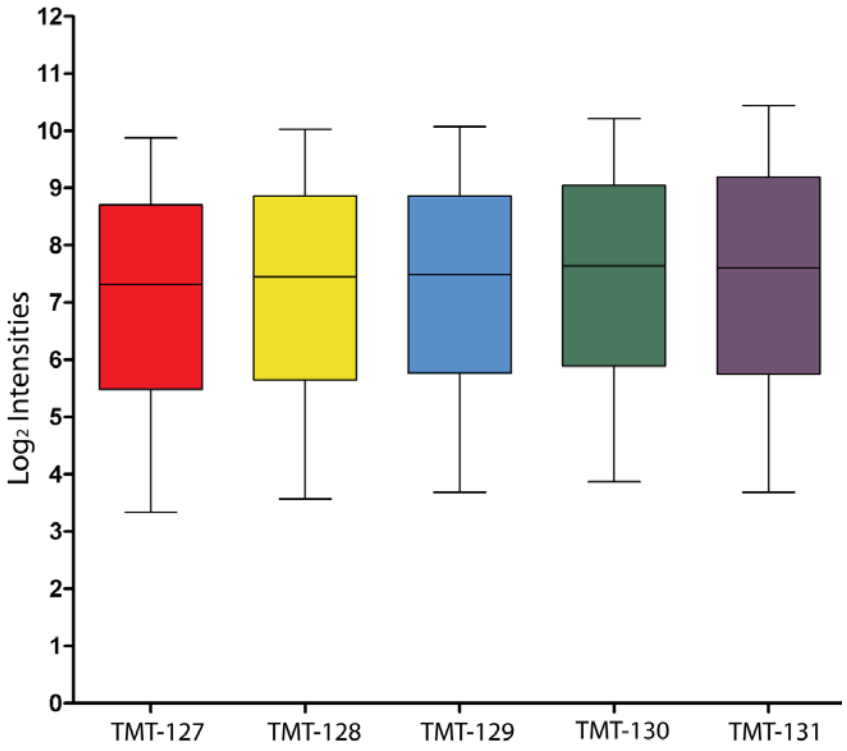


Figure 1. Comparison of median intensities across TMT channels for peptides in the dataset. The box plots from median intensities values of log_2 intensities of all peptides among different TMT channels are represented. The data show that there is no quantitative bias in the labeling for each channel.

The remaining 904,922 (80 %) unmatched spectra were exported and searched against a custom database generated from RNA seq data comprising all the transcribed regions. The nucleotide database was concatenated to common contaminants and its decoy version and searched using MASCOT v2.3 in all 6 frames with the following parameters: enzyme specificity trypsin, two missed cleavage site, carbamidomethyl (Cys), oxidation, deamidation (N) and Gln to pyro Glu (N-terminal Q) and TMT-6plex (N-termini and Lys) as static modifications. A stringent 1 % local FDR was used to select for high confidence matched spectra resulting in 7,722 matched spectra (peptide spectral matches). To further ensure that only high quality matches were included, only the 1,584 matched spectra with an expectation value <0.05 were chosen for further analysis. The expectation value is calculated by the search engine (MASCOT v2.3) and it is equivalent to the E-value in a BLAST search. In total, there were 250 distinct peptides that mapped uniquely to the mouse genome and did not overlap with the 3,284 proteins. This set of

high confidence unmatched peptides that are matched to experimentally defined nominally noncoding transcripts is referred to as novel peptides.

Synthetic Peptide labeling and validation

Synthetic peptides (JPT Peptide Technologies) were dissolved as per instruction. These peptides were pooled in equimolar concentration and categorized in six groups. Each peptide group was mixed with one of six TMT labels (Thermo Scientific) for 3 h at RT. The reactions were quenched with 5 % hydroxylamine in 100 mM TEAB and incubated for 15 min at RT. The samples were desalted with micro spin Silica C18 columns (Nest Group, Inc.) following manufacture's guidelines. Briefly, the columns were washed twice with 70 % ACN (with 0.1 % FA), washed twice with 0.1 % FA. The sample pool was passed twice through each individual column, washed four times with 0.1 % FA and eluted twice with 30 % ACN (with 0.1 % FA), twice with 50 % ACN (with 0.1 % FA), twice with 70 % ACN (with 0.1 % FA). The eluted fractions of each individual pool were combined and dried in a speed-vac (Thermo Scientific) at room temperature. The pellet was re-dissolved in 5 % ACN (with 5 % FA) loading buffer and analyzed on 4 different instruments: a Q-Exactive (Thermo scientific) and on three instruments capable of PQD - an Orbitrap Elite (Thermo scientific) and 2 Orbitrap Velos at Beth Israel and Deaconess Medical Center and at the University of Bonn (Thermo scientific). The .raw files were searched against the custom RNA-seq database.

The spectra and the fragmentation pattern for each of the synthetic peptides were compared and validated to the ones from the experiment, denoted as 'identified'. Briefly, the top 90 % matched fragment ions were compared between the synthetic peptide spectra and identified peptide spectra using Spearman rank-order correlation analysis. All the correlations were checked for a significant p-value of 0.05 or less and positive correlation coefficient for all the instrument types. Finally, the matched y and b ions indicated by their respective fragmentation tables were also mapped.

RT-PCR validation

Total RNA isolation for RT-PCR validation was done using mirVana isolation kit (Ambion). We collected and combined 0, 1, 2, 3, 6 h post-KCl stimulation time points for cortical neuron

cultures that were plated in equal density. Cells were washed with PBS, lysed on ice with vortexing in lysis/binding solution as per the manufacturer's instruction. The lysate was treated with 1/10 volume of miRNA homogenate additive for 10 min on ice and was organically extracted with acid-phenol:chloroform mixture. The aqueous phase was recovered and mixed with 1.25 volumes of 100 % ethanol. The mixture was further purified through filter cartridge, washed and eluted with nuclease free water. The RNA eluate was analyzed (Agilent 2100 bioanalyzer nano series) to determine high quality, intact RNA without DNA contamination and concentration was determined using a nanodrop spectrophotometer (**Figure 2**).

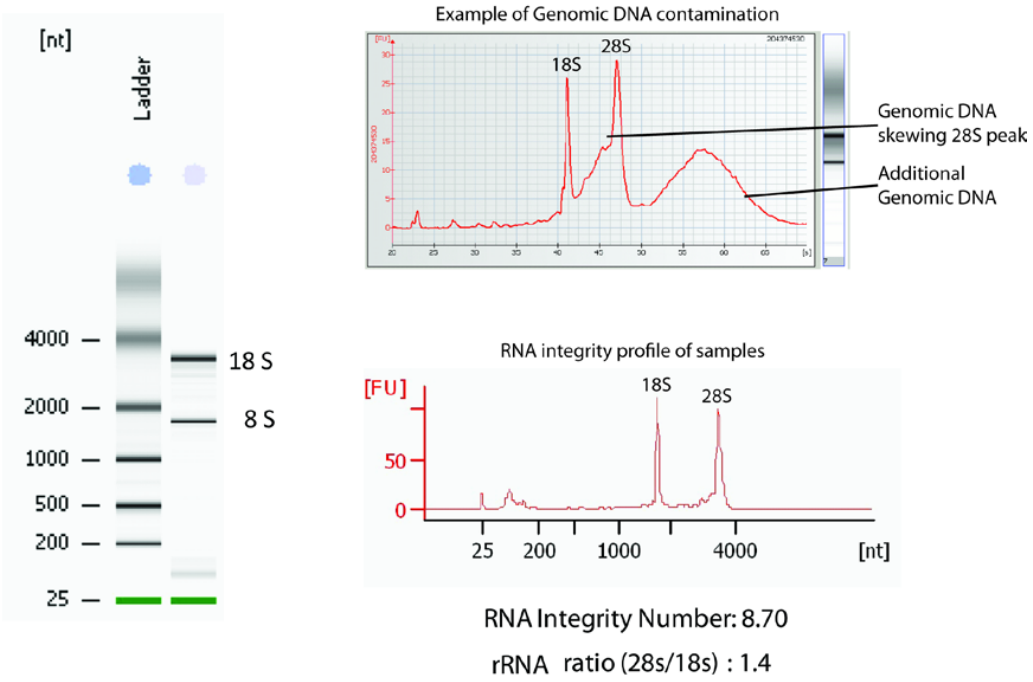


Figure 2. Validation of novel peptides in extragenic regions. RNA integrity profile from bioanalyzer showing DNA-free, high quality and intact RNA. The shaded spectrum is hypothetical example for presence of genomic DNA shown as reference from bio analyzer.

First strand cDNA was synthesized using SuperScript First-Strand synthesis system (Invitrogen). Briefly, total RNA sample was incubated with random hexamer primers to include all RNAs for cDNA synthesis and dNTP mix at 65 ºC for 5 min, then placed on ice for 1 min. The sample was mixed with 2x reaction mix of Reverse transcription buffer, $MgCl_2$, DTT and RNAaseOUT and incubated at room temperature for 10 min, then incubated at 42 ºC for 50 min. The reaction was terminated by heating at 70 ºC for 15 min, then on ice. To ensure the sensitivity of PCR, the

cDNA-RNA hybrid molecule was treated with RNase H at 37 ºC for 20 min. The cDNA was then used to amplify specific novel regions of selected length containing peptide match regions identified using the mass spectra and deep sequencing data. Primers were synthesized with IDT custom oligo, designed using OligoPerfect designer (Invitrogen) and Tm was optimized using Finnzyme calculator (Thermo scientific). **Table 1** lists the regions of interest and primers sequences. PCR was performed using Phusion HF buffer, dNTPs mix, Phusion hot start DNA polymerase with initial denaturation at 98 ºC for 30 sec, denaturation at 98 ºC for 10 sec, annealing at Tm of the lower tm primer (for 20 nucleotides or less) and 3 ºC higher than the Tm of lower tm primer (for more than 20 nucleotides) for 30 sec, extension at 72 ºC for 30 sec / kb template length and 30 cycles. Final extension was done at 72 sec for 10 min, followed by 4 ºC. The PCR products were run on 1 % agarose with ethidium bromide in 1x TAE buffer.

Products were gel purified using QIAQuick PCR purification columns (Qiagen) and sequenced conventionally. The results were mapped to the template region using the ClustalW2 (EMBL-EBI) tool and translated into 6 frames using ExPASy tool to validate for peptides match.

PCR Region	**Forward primer**	**Reverse primer**
Armc9, intron retention	CCCATGGTTCATCTGCTCTT	GCTGGTTACTAGCGGAGACG
Cacna2d3, AS	CTGGGTTGGCAA GAACAAAT	CGGTTCCCAGGTAATAGCAA
Cox17, AS	AGGACACACCACCCTAGTCG	TCTCCAGACCAGAGGCACTT
E130006D01 Rik, overlaps intron of annotated ncRNA	GTAAGCTGCAGGCACATCAA	CTTCATGGCAGAACCTGGAT
Elmo2, divergent	CCCAGCAAGCTAAGAAGTGC	TCTGGTGTGCCTCACTCTTG
Farp 1, novel intron region 3	TAGGAGCCTTCTGTGGCTGT	GAGGTCTCAGGAAGCCCTCT
Farp1, novel intron region 1	GCTGGAGGCAGTAGCAATTC	CGTCTCATCCTTCCTCTTGC
Farp1, novel intron region 2	GGGTTGCCACAGAGAAGGTC	CACCATGGCTGATGGAGAAA
Fars2, AS	ACCCTCACTGGGACTCCTCT	TTCAGCAAGGTGTCAGATGC
Gprk5, AS	GGAAGAATCATGGGCACTGT	TATCAGCCTGACCTGGGAAG
Grpbp10, divergent	CATGCATGTTATCCCAGCAC	TATGGAGCCAGTGGTTGACA
Mlstd2, AS	TAAGCAGCAGCAAAGGGAAT	CTGGGCCCTCTCTCTCTTTT
Nars2, AS	CATGCACACAATGGACACAA	TGGCAGTGGATGTAGCTCAG
PGOMOU00000135506, pseudogene	AATGGTGAAGGTCAGTGTGAAC	ACTCCTTGGAGGCCATGTAG
PGOMOU00000135766, pseudogene	AATGGT GAAGGTCGG TGTG	TACTCCTTGGAGGCCATGTAG

St6galnac3, AS	TCCAGTGTGCCAACATTCAT	CACCATCCCTGTCACTCTCA
Suvn, alternative 3' end	GCCAAACTGTCACCCAAACCC	TTCACATCTCCCCAAACCTC

Table 1: The table lists the regions of interest and primers sequences that were designed to verify the expression of 14 anti-sense RNAs (Figure 7A) along with 3 novel intron regions of FARP1 (Figure 8C) by RT-PCR.

FARP1 Immunoprecipitation (IP) and western blot analysis

Whole brain tissue was lysed in buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1 % NP-40, 0.5 % sodium deoxycholate, 0.1 % SDS, 0.004 % sodium azide, protease inhibitors tablet) using bead beater homogenizer (Precellys) for 6800 rpm, 3x 15 sec with 60 sec rest cycle. The lysate was centrifuged at 14,000 rpm, 10 min. at 4 ºC. FARP1 antibodies K-20 (2 μg per 500 μg of lysate for IP, 1:500 for WB, Santa Cruz Biotechnology) and FARP1 H-300 (2 μg per 500 μg of lysate for IP, 1:500 for WB, Santa Cruz Biotechnology) were tested separately for IP and incubated with lysate for 2 h. The immuno-complexes were pulled-down using protein A/G agarose beads (Santa Cruz) in an overnight incubation with rotation. The beads were pre-blocked with BSA to reduce non-specific immunoglobulin binding. The immune-precipitates were sedimented at 1000 g, 5 min, 4 ºC. The beads were washed 4 times with RIPA buffer containing 150 mM salt. The beads were resuspended in Laemmli buffer containing β-mercaptoethanol and boiled for 3-4 min. The lysate and IP eluate samples were run on 4-12 % SDS-PAGE gel (Invitrogen) and were subjected to western blot analysis using the FARP1 primary antibody K-20 as a probe in dilutions described above. A secondary goat anti-rabbit IgG-HRP antibody (1:5000 dilution was then used for the WB, Santa Cruz Biotechnology). The WB was developed by Super Signal West Pico chemiluminescence kit (Pierce). The commassie stained gel bands were analyzed by MS. The peptides were extracted, reduced with DTT, alkylated with iodoacetamide and digested with trypsin before running on a QE instrument with a 60 min gradient to acquire a base peak chromatogram intensity of e^9 - e^{10}. The raw files were converted into .mgf files and searched using mouse proteome database (Uniprot Feb 2012, (canonical and isoform sequences) protein sequence database, containing common contaminations and concatenated to its decoy version, using MASCOT v2.3 (Matrix Science) appended with mouse FARP1 Uniprot identifiers and intron region (983 bases) in coding frame using ProteinPilot search engine software v4.5.1. The data was analyzed to identify peptides from FARP1.

Bioinformatics analysis

Each of the 250 peptides reported in **Table 2** were inspected manually using the UCSC Genome browser. In addition to using the ChIP-Seq and RNA-Seq data from mouse neurons, the RNA-Seq data from the mouse ENCODE was also used. The initial assignment into different categories was carried out by an algorithm and the assignment was based on the location of the transcribed region and peptide relative to known genomic features. In some cases the category was changed following the manual inspection.

Clustering of novel peptides and known proteins

The intensities for each peptide were normalized by dividing the values at 1 h, 2 h, 3 h and 6 h by the value at 0 h. Next, we carried out unsupervised hierarchical clustering of the $\log_2$-values for the temporal profiles using Matlab's 'clustergram' function. The clustering was based on the Euclidean distance between the intensity profiles and the same settings were used for the known and the novel peptides.

Co-clustering of novel peptides and known proteins

To identify patterns of co-regulation between the known proteins and novel peptides, the known protein abundance changes along with the novel peptide abundance changes were clustered together in an unbiased way using GProX (version 1.1.12) software. Results of the clustering analysis indicate that eight clusters identify discerning patterns of co-regulation of known proteins and novel peptides. The number of novel peptides in each of the clusters is listed in **Figure 11**. All ratios in each experiment were standardized before clustering (the mean was deducted from the values and divided by their standard deviation). The membership scale reflects how well the regulation of a protein matches the consensus profile.

Ribosomal foot-printing analysis

Ribosomal foot-printing data for elongating ribosomes for mouse embryonic stem cells was downloaded from gwips.ucc.ie. To identify peptides with significant levels of ribosomal foot-printing reads, the peptide positions were first converted to mm10 coordinates and then all peptides with 5 or more reads within 200 bps were extracted. Based on manual inspection in the genome browser at gwips.ucc.ie, 34 loci were determined to have evidence of ribosomal

interactions.

Phylogenetic comparison

The MultiZ alignment for 28 species relative to mm9 was downloaded from the UCSC website. Around 142 peptides were found in alignment blocks by MultiZ and the orthologous sequences for all other species that aligned to the same block were extracted. For each aligned peptide, the codons with one substitution were classified as either synonymous (S) or non-synonymous (N) and the total numbers of synonymous and non-synonymous events were recorded. For each peptide, the ratio dN/dS were calculated and a value <1 is characteristic of purifying selection.

RESULTS

Work flow and analysis of RNA-seq data

To systematically investigate the existence of non-canonical translation products and their biological regulation under physiological conditions, we compared the proteome and transcriptome from the same experimental context of mouse cortical neurons that were depolarized by KCl to induce activity-dependent expression changes. The quantitative MS proteomic data and total RNA-seq transcriptome data including polyA and non-polyA species after rRNA depletion [5] were collected at multiple time-points (**Figure 3A**). To identify transcribed regions from the total RNA data, a *de novo* transcript-calling algorithm was used [18]. This algorithm is well suited to discover unannotated translation products, since it identifies unspliced transcripts, has no sequence biases and is specifically designed to detect lowly expressed regions of the genome. The algorithm identified 26,169 transcribed regions, 12,108 of which overlapped with annotated protein coding genes (RefSeq, mm9). Here, regions corresponding to unspliced transcripts from total RNA data are searched; thus in combination with proteomics analysis, our study has the potential to detect a wider range of non-canonical protein products and differs from previous work that was confined to either spliced transcripts [15] or transcripts attached to ribosomes [19].

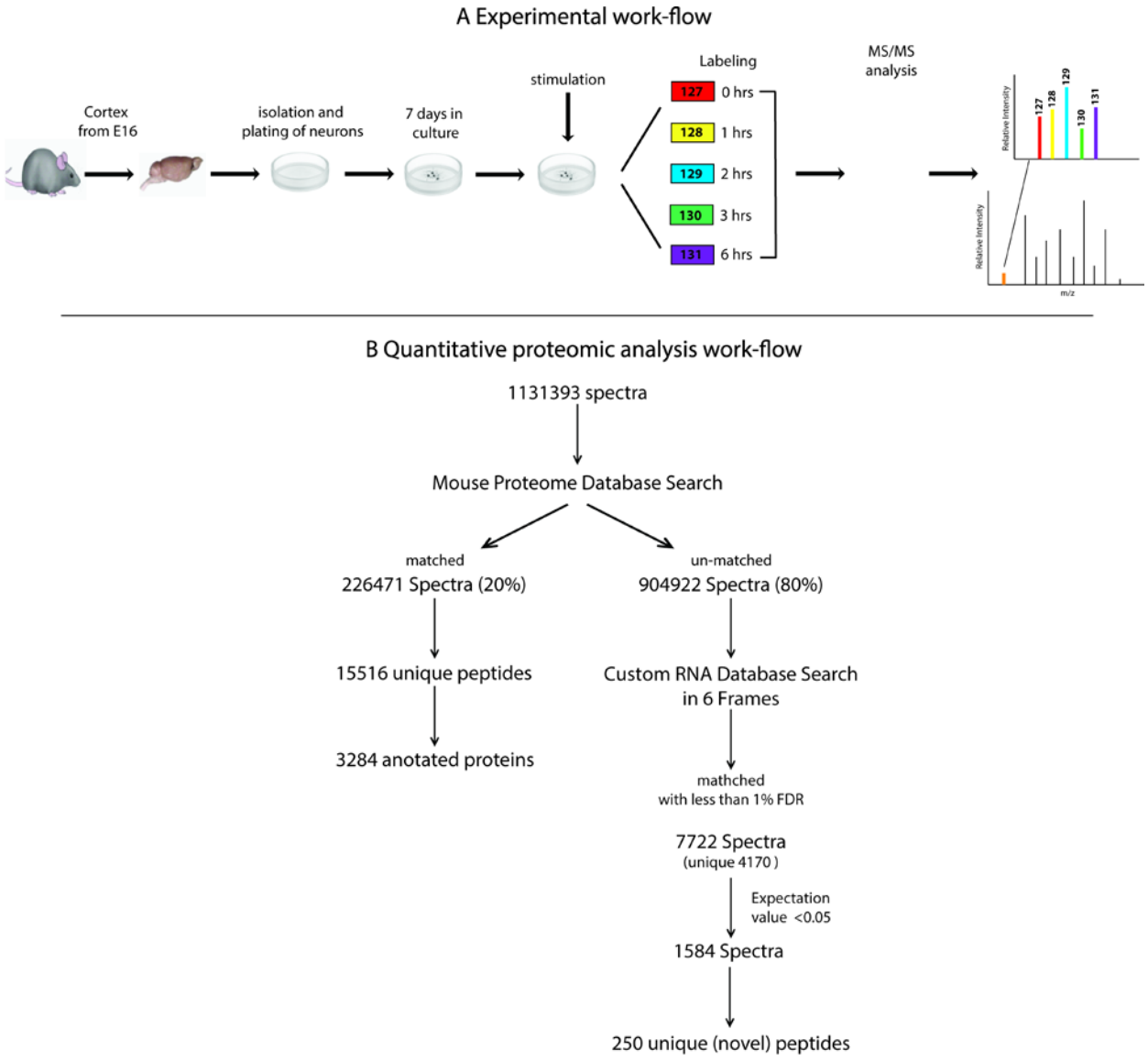


Figure 3. Overview of the experimental procedure and peptide identification. (A) Work flow of the KCl depolarization experiment illustrating the collection of total RNA and protein from mouse cortical neurons. Protein samples from multiple time points after depolarization were digested, labeled with isobaric tags and analyzed by quantitative proteomic analyses (LC-MS). **(B)** The resulting spectra were subjected to the database search procedure for filtering and matching peptide spectra. Twenty percent of the spectra matched to the known mouse proteome (Uniprot 12 December, 2011, incl. canonical and isoform sequences). The remaining 904,922 spectra were searched against the custom RNA-Seq transcriptome database in 6 frames and the resulting matched spectra were filtered using stringent criteria as mentioned in the text and of these 250 novel peptides were selected for further analyses.

Identification of novel non-canonical translational products

A multiplexed quantitative MS proteomic data set comprised of 3 biological replicates with 3 technical replicates each, with a temporal profile similar to that of the transcriptomic data yielded 1,131,393 MS spectra. We used an iterative procedure to investigate these peptide fragmentation spectra. The rationale for this iterative strategy was to reduce the random matching spectra derived from annotated peptides to non-canonical sequences, thereby increasing the chances of identifying true spectral matches for non-canonical transcripts. First, the spectra were searched with MASCOT (v2.3) against an annotated mouse protein database (Uniprot Feb 2012, canonical and isoform sequences) to identify matches to annotated peptides; 226,471 (20 %) of the spectra matched annotated sequences. These matched spectra (peptide spectral matches, PSMs) corresponded to 15,516 unique peptides and were grouped into 3,284 proteins using a 1 % false-discovery rate (FDR) cut-off and a minimum of two peptides per protein (**Figure 3B**). Since the focus in this study was to investigate non-canonical translational events, the spectra matching the annotated proteome were then excluded before searching against the custom transcriptome database.

While unmatched good quality spectra may be derived from peptides which are post-translationally modified and are not accounted for in the initial searches against the canonical mouse protein sequence database, we asked whether some of the unmatched good quality spectra could constitute translation products from transcripts classified as non-coding RNAs (ncRNAs). To investigate this possibility, the remaining un-matched spectra were searched against a custom database, which was generated by translating the transcribed regions from total RNA-sequence data. This custom database contained transcripts derived from total RNA, including unspliced RNAs and also annotated genes. Each transcript was searched in all six reading frames using MASCOT. A 1 % FDR was used to select for high confidence PSMs resulting in 7,722 PSMs (**Figure 3B**) (of which 54 % were non-redundant). To further restrict subsequent analyses to high quality matches, only the 1,584 matches with a MASCOT expectation value <0.05 were chosen for further analysis. The expectation value is equivalent to the E-value in a BLAST search and thus gives each PSM a probability estimate for being a random match. These parameters provide confidence that these hits represent true positives, even though in some cases only a single instance of a peptide is detected [20]. The set of 1,584 PSMs comprised 250 distinct peptides that

mapped uniquely to regions of the mouse genome considered non-coding (**Table 2**). Seven pairs of peptides (**Table 3**) matched two different ncRNAs that differed only by an isobaric I/L amino acid, while the unambiguous identification of the source of these peptides is not possible, it is clear that these peptides sequences are not found in the canonical mouse protein sequences. Given that RNA sequence information was obtained for both I and L containing peptides, it is possible that both the amino acid I and the L containing peptides are present; therefore, all 14 peptides were included for further analysis.

Novel Peptide	No. of peptides identified	Chromosome number	Chromosome Position	Ribosome_hits	Overlapping gene	RNA Category
AAAAGPGAASEHR	1	1	91478037	1	Centg2	Novel internal exon
AADIQGASLGPRK	1	7	148086257	0	Cox8b	Short alternative ORF
AAEDEEVPAFFK	10	14	67529554	0	-	Alternative 5' start
AAPPPICPSTALCPCLAR	1	2	33420519	0	Lmx1b	Overlaps intron
AAVSIAEER	1	1	91567439	0	Centg2	AS
ACQSMLNTGIIIFK	1	2	165150427	0	Elmo2	Divergent
AGVTPASQNR	1	X	20939669	0	EG331392	Pseudogene (BLAST)
AILQALVK	1	X	84544638	0	Il1rapl1	AS
AILQGLLK	1	5	136734659	0	Cutl1	Overlaps intron
AIVLIPK	1	9	60250275	0	Thsd4	Overlaps intron
ALIQAVIK	5	7	120704164	0	Mlstd2	AS
ALLQALVK	1	7	104182852	0	Nars2	AS
ALVLIPK	1	2	32434758	0	Pip5kl1	AS
ANLGGLWLL	1	18	64518528	0	Onecut2	AS
APEGSREGVR	1	9	121115688	0	Ulk4	Overlaps intron
APFRLCPLPVEK	1	5	67827299	0	EG666938	Putative ncRNA
ASGAPSGSATAPAER	2	2	25066765	1	730008L03Rik	Alternative 5' start
ASGEPSVESSR	6	14	67529332	0	-	Alternative 5' start
ASSSILINEAEPTTDIQIR	1	5	49944247	0	EG433885	Pseudogene
ATWAVARR	1	10	70891149	0	-	AS
AVCLLSNTTAIAEAWAR	1	12	52234717	0	-	Pseudogene (BLAST)
AYHEQLSVAEITNVCFEPTNQMVK	1	17	5111198	1	Arid1b	Pseudogene
CDFPYLVLWVQFR	2	16	42107151	0	Lsamp	AS
CLLAAGRGLR	1	16	23133852	0	-	Putative ncRNA
CPFLLMRTLR	1	8	80130867	0	Tmem34	Overlaps intron
CPGNLFEGGR	2	X	103416080	0	-	Short alternative ORF

CPTDCTRTLR	1	14	66988529	0	-	Overlaps intron
CSSPPHCTR	1	3	32436586	1	Mfn1	AS
DAVAGIR	1	2	168038058	1	Dpm1	AS
DDLGLGAR	1	7	91104662	0	930013L23Rik	Overlaps intron
DDQECLLLCQR	1	1	37813851	0	Tsga10	Overlaps intron
DEDGLGLR	1	3	153224947	0	St6galnac3	Overlaps intron
DEDVAGLR	2	16	3969222	0	Nlrc3	Alternative 3' end
DELGALDQGR	1	4	101508277	0	B020004J07Rik	AS
DGIRPSNPQPPQPSTGPASR	1	14	121642789	1	Farp1	Overlaps intron
DKPANVQPYYLYGSK	2	11	46053502	0	Cyfip2	UCSC (BLAST)
DLYLPLLLQRNILR	2	4	130370613	0	Sdc3	eRNA
DMCISSDTLGVSDR	1	3	157733235	0	-	Alternative 3' end
DQLVDAR	1	18	80860859	0	Nfatc1	AS
DVAHLGPDPHR	10	14	67529377	0	-	Alternative 5' start
DVQQYILHR	5	3	38356427	0	-	Novel internal exon
EDSLLMQILR	1	14	87169352	0	Diap3	Overlaps intron
EDYYNPSR	1	4	28748296	0	Epha7	AS
EEEMRDTNESTNTR	1	13	36082831	0	LOC100042773	Pseudogene (BLAST)
EIGGGGEIR	1	1	84400192	0	Dner	AS
EMESRDEEVEEAR	1	11	77659932	0	Myo18a	UCSC (BLAST)
EQDGLFSFCLR	1	6	64474513	0	Grid2	Overlaps intron
EQSTGGTLLR	1	1	161841514	0	Tnr	AS
ESTEGGPGITR	1	1	194948951	0	AA408296	Overlaps intron
ESVLVVPPTPPHPQL	1	5	118629779	0	Hrk	Overlaps intron
ETVEECCILAR	1	5	149071107	0	C130038G02Rik	Overlaps intron
EVAGRSGGR	1	X	132395342	0	-	Alternative 3' end
EYEEDGSR	1	1	36756941	0	Actr1b	UCSC
FEDEDFILK	3	5	92235820	1	-	Pseudogene
FLVIMALGMSRAR	4	8	109372919	0	Tmco7	AS
FQGPLDWLR	3	18	77462336	0	St8sia5	Overlaps intron
FSDEDATR	1	3	55337524	0	Dclk1	UCSC
FSSHNSLLLSPR	1	X	70764755	0	Atp2b3	AS
FTQAQLDSGLVLFSHR	3	9	56740276	0	Cspg4	UCSC (BLAST)
FWEVISDEHGIDPTGSYHG DSDLQLER	1	12	11889329	0	-	Pseudogene (BLAST)
GAELVDSVLDVVR	5	13	63069659	1	-	Pseudogene (BLAST)
GAHLVSSPAR	1	13	23875947	0	Trim38	Putative ncRNA
GASQQAVPLR	1	4	125290089	0	Grik3	AS
GDASPAVLR	4	8	96458995	0	Gnao1	eRNA
GDSVVNLR	3	17	79010529	0	Vit	AS
GERGRAGGGR	5	5	112174502	0	-	Putative ncRNA
GFILLRWGWGGGGVGLIS R	1	4	149833438	0	Rere	Overlaps intron
GGAMQGGLR	2	19	61104526	0	Gprk5	AS
GGDGLGSLTVGLFK	2	7	53669782	0	Kcnc1	AS
GGGCVSSALSILSK	1	5	99285415	0	Bmp3	Overlaps intron
GGGGGGGGGGGGGGGG GGRER	1	9	28395874	0	Opcml	AS

GGGGGGGGGGRGRGGSR	1	18	12042894	0	Cables1	AS
GGHGGIR	4	11	43365085	0	Ccnjl	Overlaps intron
GIAILRR	2	2	170281125	0	-	Putative ncRNA
GILTGSRR	1	4	149034084	0	Pik3cd	Overlaps intron
GISYVRR	1	3	104417607	0	-	Putative ncRNA
GLAIIRR	5	19	50306195	0	Sorcs1	Overlaps intron
GLSYVRR	1	14	21587505	1	Camk2g	Overlaps intron
GLTLILRR	1	7	20428525	0	-	Putative ncRNA
GLTLLLRR	1	15	23362683	0	-	Overlaps intron
GMIEALLVR	1	2	22195185	0	-	Overlaps intron
GPVTCLPR	1	14	78011663	1	D230005D02Rik	Overlaps intron
GSGQVVAQQR	6	14	67529410	0	-	Alternative 5' start
GSILASLSTFQQMWISK	1	8	47315377	0	-	Pseudogene (BLAST)
GSLLSNWACVQLSPGR	1	3	132560178	0	Npnt	Overlaps intron
GSSDEEMPL	2	8	121653077	0	Cdh13	AS
GTLTDGDSNNQRR	2	9	14412177	0	Amotl1	Overlaps intron
GVIVRCR	1	3	152989615	0	St6galnac3	AS
GVLVRCR	1	17	51925415	0	Satb1	Overlaps intron
GVMVGMGQK	1	19	34326318	0	Acta2	UCSC
HELQANCYEEAVKDR	1	10	12566236	1	Utrn	Pseudogene
HFECLLSSSPICK	1	9	50798314	0	Snf1lk2	Overlaps intron
HGDPGDAAQQEAK	1	1	191101600	0	LOC100042424	Pseudogene (BLAST)
HPENYQWENWSLENIATILAR	3	1	57439615	0	1700066M21Rik	UCSC (BLAST)
HSLEGKGCRLLEQNFK	1	15	47903515	0	Csmd3	AS
HSSLIDDMFR	1	14	121643143	1	Farp1	Overlaps intron
HTGPEILSMANAGPNTNGSQFFICTSK	2	X	21883037	0	-	Pseudogene
HTGPGILSMANAGLNTNGSQFFICTAK	1	2	83230557	0	-	Pseudogene
IAKPLSSLTPLIAAAK	5	7	105059625	0	Pak1	UCSC (BLAST)
IEGDMIICAAYAHELPK	8	6	83551442	0	-	Pseudogene
IGALSCPRRR	1	12	87860099	1	Esrrb	Overlaps intron
IHFPIAT	3	7	107107329	0	-	AS
IIISLAPR	2	6	37869262	0	Trim24	Overlaps intron
IILMPSLPAR	18	16	38351276	0	Cox17	AS
ILAILDAV	1	1	154406413	0	Rgl1	Overlaps intron
ILDAGNGTIR	1	1	173302431	0	Pvrl4	Overlaps intron
ILDDDTIITTLENLK	13	12	111896626	0	Dync1h1	UCSC (BLAST)
ILDDDTIITTLENLKR	7	12	111896626	0	Dync1h1	UCSC (BLAST)
ILELDIEDLLLR	3	3	96586664	0	Zfp364	Overlaps intron
ILISPLAR	1	1	164637478	0	Bat2d	AS
ILLLSPAR	1	X	108666439	0	-	Pseudogene
ILNNLNNILR	1	3	106941027	0	-	Putative ncRNA
ILSLLPAR	2	10	79093067	0	Shc2	Overlaps intron
ILSSPLTFQLLASLAPR	2	1	88138943	0	Armc9	Overlaps intron
ILVTLIH	4	13	49058457	0	BC010304	Overlaps intron
IPGLLGEDLLCGSR	1	16	37956669	0	Gpr156	Overlaps intron

IQSLAEER	1	18	76314644	0	-	Alternative 3' end
IQTMFEQLALVDHPNIVK	1	15	75921106	1	-	Alternative 5' start
IRQGMPLILR	1	3	146552234	0	Ttll7	Putative ncRNA
IVLLPQDFVIPN	6	1	148724097	0	B830045N13Rik	AS
IVVLGIR	2	17	73210269	0	-	Alternative 3' end
KLLMMAGIEDCYTSAR	1	2	170779768	0	-	Pseudogene
KPPPPASPGSSDSSAR	2	8	73438387	0	Mtap1s	UCSC (BLAST)
KQEEQMETEQ	2	4	83109494	0	Snapc3	UCSC
KTLQLYIQAIDSLA	1	18	69838152	0	Tcf4	AS
LAPLMLLPAR	5	13	36301302	0	Fars2	AS
LAQVALELK	2	9	106733048	0	Vprbp	Overlaps intron
LAVDTEGQRR	1	5	74029635	0	Sgcb	Overlaps intron
LEDALMENR	2	17	25321158	1	-	Novel internal exon
LELMVILLPQTLK	1	2	12896039	1	Pter	Putative ncRNA
LFLSPAIQGLLLPAR	2	17	8994702	1	Pde10a	Novel internal exon
LGIALNF	15	X	98344394	0	-	Overlaps intron
LGTESRDTVTENER	1	7	141163514	0	Adam12	Overlaps intron
LIEPVLIIQK	15	14	122877279	0	Zic2	Putative ncRNA
LIGLPLQILM	1	16	76368640	0	Nrip1	Overlaps intron
LILLGLGDR	1	4	101034924	0	-	Pseudogene
LINDLPAR	1	16	43497573	0	-	Overlaps intron
LINISEILK	1	16	40822205	0	-	Overlaps intron
LIYIEMIIIR	2	2	134941721	0	Plcb1	AS
LLAALLHNPQLVER	1	16	31948474	1	0610012G03Rik	UCSC
LLETDLILRN	2	8	84434162	0	Inpp4b	AS
LLLSPLAR	1	3	51960827	0	-	Putative ncRNA
LLLVSSLAPR	1	14	18167137	0	-	Putative ncRNA
LLSIPLAR	1	10	8083470	0	Ust	AS
LLSLIPAR	2	9	74888348	0	Arpp19	Overlaps intron
LLSNLSAPR	1	1	68680567	0	Erbb4	AS
LLVFLPFILGGP	1	6	50122947	0	Mpp6	Overlaps intron
LLVMVLINR	1	10	69298622	0	Ank3	AS
LLWPGGLLLLP	1	4	82096819	0	Nfib	AS
LNKFYLYICVFTR	1	19	16784133	1	Vps13a	AS
LPPSFLLPAR	5	14	8673935	0	Flnb	AS
LTTPTYGDLNHLVSATMS EVTTCLR	127	13	63069299	1	-	Pseudogene
LVELLLRN	1	13	21289375	1	-	Alternative 3' end
LVIDGKPITIFQER	8	11	3851927	1	-	Pseudogene
LVITDVVLNR	1	2	157840512	0	-	Overlaps intron
LYIVFSLLILCR	2	3	97557685	0	Pde4dip	Overlaps intron
MDSFDEDLAR	2	4	125855101	0	Thrap3	UCSC
MFQYDSTHGK	4	11	99539579	0	-	UCSC (BLAST)
MSIFGH	4	1	179180269	0	-	Overlaps intron
MVILELRR	1	15	76514928	0	Ppp1r16a	AS
MVVSEEYHLSSISK	1	5	106313626	0	Zfp326	AS
NAFLSSSALDK	3	19	12534993	0	-	Alternative 3' end
NALQELQQIIITPIK	6	7	29191631	0	Samd4b	UCSC (BLAST)

NCLPGLQLQEESTGLER	3	19	3819420	0	-	Alternative 3' end
NDSDSVLISILEK	1	3	35805209	0	Dcun1d1	Overlaps intron
NLISAGGQGPR	1	8	124806573	0	Zfpm1	Overlaps intron
NMETPLSVHHCFVK	1	3	37122645	0	Il21	Putative ncRNA
NSNILMENR	2	4	49601752	0	2810432L12Rik	AS
NSSYFVEWIPNDVK	17	4	24166130	0	-	Pseudogene
NVELTYYFLFHLVILAR	1	14	117673918	0	Gpc6	Overlaps intron
NVVHQLSVTLEDLYDGATR	12	5	21458463	0	Reln	Pseudogene
PFSPLMENR	1	9	117145898	0	Rbms3	AS
PLCILTGCQPLR	1	14	29813217	0	Cacna2d3	Overlaps intron
QARSLDPLALIR	1	4	139799648	0	Igsf21	Divergent
QCWKMLQMASGYGTTMR	1	1	68551590	0	Erbb4	Overlaps intron
QDLDVAR	1	X	35701835	0	-	Putative ncRNA
QEAVAGSLSK	4	12	59839129	0	-	Putative ncRNA
QEFDTMEDHAGDYTLLQDQEGDMDHGLK	1	11	104143714	0	Mapt	UCSC (BLAST)
QGAEELAR	1	15	59195487	0	E430025E21Rik	Overlaps intron
QGLILYLR	1	9	56157990	0	C230081A13Rik	AS
QIIEINPR	1	11	47353320	0	-	Pseudogene
QLALEEQHER	2	5	121704012	0	EG545802	UCSC
QNNILIVANNSCVLK	1	1	74490131	0	Usp37	AS
QPALGTALK	5	17	27624222	0	Grm4	AS
QPTHLPR	1	8	122445320	0	Usp10	AS
QQVLLTMSEAPACK	2	9	67118687	0	Tln2	AS
QTERGGGGGGGGGGGGGGR	1	14	30226844	0	Cacna2d3	AS
RAPADTHSSSAESIEGSPR	1	1	91478172	1	Centg2	Novel internal exon
RASGEPSVESSR	19	14	67529332	0	-	Alternative 5' start
RIEAGSCPR	1	5	121599352	0	Ptpn11	Overlaps intron
RLCQGSIVEIIK	2	16	40798627	0	-	AS
RLESCAPR	2	1	189610538	0	-	Putative ncRNA
RPETVIGVDINPYLHMR	1	9	122098075	0	Tmem16k	Alternative 3' end
RQLGNNLLLR	1	X	74870093	0	Tbl1x	Overlaps intron
RQQLQELR	1	5	143652771	0	Fbxl18	AS
RRVLELGDLER	1	3	122821453	0	Synpo2	Overlaps intron
SAPADGADLSAHLWAR	1	14	31944491	1	-	Alternative 5' start
SCSLLVLNLK	1	14	12560810	0	Ptprg	AS
SDLHALVVASSVPLLFR	1	8	126552336	0	AK122209	UCSC (BLAST)
SHFEQWGTLTNCVVMR	5	X	138732208	0	-	UCSC (BLAST)
SIVETLVLSPSTVGSGR	1	19	45796161	0	-	Putative ncRNA
SLFDFLLSRN	5	1	32403440	0	Khdrbs2	AS
SLFSSPASLAFPYSPVAR	12	14	104083469	1	Slain1	UCSC
SMTEAEQQQLIDDHFLFDKPVSPLLLASSMAR	2	16	38483886	1	Cd80	Pseudogene
SRVGLAEAER	1	8	75251940	0	Sin3b	Overlaps intron
SSSWLVVQSLEAR	9	11	81715748	0	Accn1	Overlaps intron
STYEDQSAR	1	6	71833517	0	Ptcd3	UCSC

TEGRETGGRER	3	16	63563279	0	Epha3	AS
TFCVFILTLR	5	X	72401812	0	4930428E23Rik	Pseudogene
TFSVMPSPK	3	13	63069500	1	-	Pseudogene
TGSGFTSLLMER	3	2	151285127	0	-	Pseudogene (BLAST)
TIDFDSLSVGR	8	14	67529440	0	-	Alternative 5' start
TLILGH	1	1	95590235	0	Bok	Overlaps intron
TLLIGH	1	6	111488471	0	Grm7	Overlaps intron
TLQAEQFIVYRLDQKAK	2	12	76925629	0	Syne2	Overlaps intron
TLSQITITNR	1	13	18082330	0	1600012F09Rik	Putative ncRNA
TLWASVQLLLTLISK	1	11	69727257	1	0610025P10Rik	Alternative 3' end
TRLSSDSQSSACFCLLR	1	12	33692187	0	-	Putative ncRNA
TSFLEMLMGALLLVISDCC LLVCSFLPPHSGAGEFQL	1	2	74632604	0	-	Putative ncRNA
TSTVDLPIESQLLWQLDR	14	11	53080494	0	Hspa4	UCSC (BLAST)
TTSMNLARNVHIIMPISK	1	4	19987497	0	LOC667301	Overlaps intron
VAEVLFDAADANAIEEVN LAYENVK	4	12	111855891	0	Dync1h1	UCSC (BLAST)
VAILFLALVSVLGAQS	1	11	56929950	0	Gria1	AS
VAVLQALASTVNR	7	6	71855088	0	Ptcd3	UCSC (BLAST)
VDNDEDEHQLSLR	1	10	27000624	0	Lama2	AS
VEEQVFPLLDLR	1	13	21360267	1	-	Putative ncRNA
VGLLLVFPLIQSQSR	3	5	121704129	0	EG545802	UCSC (BLAST)
VIHDNFGIAEGLMTMVHAI TATQK	1	8	89183665	1	Lonp2	Putative ncRNA
VIHDNFGIVEGLMTTVCAI TATQK	6	9	109733333	1	-	Pseudogene
VIHDNFGIVEGLMTTVHGI TATQK	12	14	103770128	1	-	Pseudogene
VIIILWLK	1	14	62005615	0	Kpna3	Overlaps intron
VIISTPSANAPMFVMGVNH EK	3	11	109021088	0	-	Pseudogene (BLAST)
VILLVV	1	1	133393996	0	Srgap2	Overlaps intron
VILLVVVMGMGNLR	1	9	47671285	0	-	AS
VIPELDGK	2	11	3852386	1	-	Pseudogene
VLIHLAK	2	5	40139120	0	Hs3st1	Overlaps intron
VLIIWILK	5	3	113786318	0	Col11a1	AS
VLLHLAK	2	5	102402044	0	Wdfy3	Overlaps intron
VLLIIWIK	1	19	58945723	1	-	Alternative 5' start
VLLLWLLK	2	18	46761781	0	Eif1a	Overlaps intron
VLTISLLGH	1	6	98988990	0	Foxp1	Overlaps intron
VSQHFPGEDVLQRTR	1	5	113191543	0	Myo18b	AS
VTHTVPIYEGYALPHAILR	8	13	81204153	0	-	Pseudogene (BLAST)
VTSIILLMFS	1	5	5563174	0	-	Divergent
VVDLLACR	2	X	131081028	0	Btk	AS
VWVSQGSDAPR	1	15	80226262	0	Cacna1i	Alternative 3' end
WEGGMQRGGGGMGGGSE R	1	16	81459937	0	Ncam2	AS
WMADDLR	2	16	59592942	0	LOC672884	AS

WWGRGGEGVGVGVGVGE GWGRLGEGR	1	5	89346189	0	Slc4a4	Overlaps intron
YNAAGTGLQVEDCR	1	19	59990938	0	Rab11fip2	Overlaps intron

Table 2: The table lists all of the high-quality novel peptides which were mapped to the RNA-seq database. **Bold peptides** were validated by synthetic peptides. The data represents peptide sequence, number of times peptide was identified, mapped chromosome number and position in genome, validation of peptides using ribosome profiling database where 1 indicates match and 0 indicates no-match, overlapping genes and category of ncRNAs.

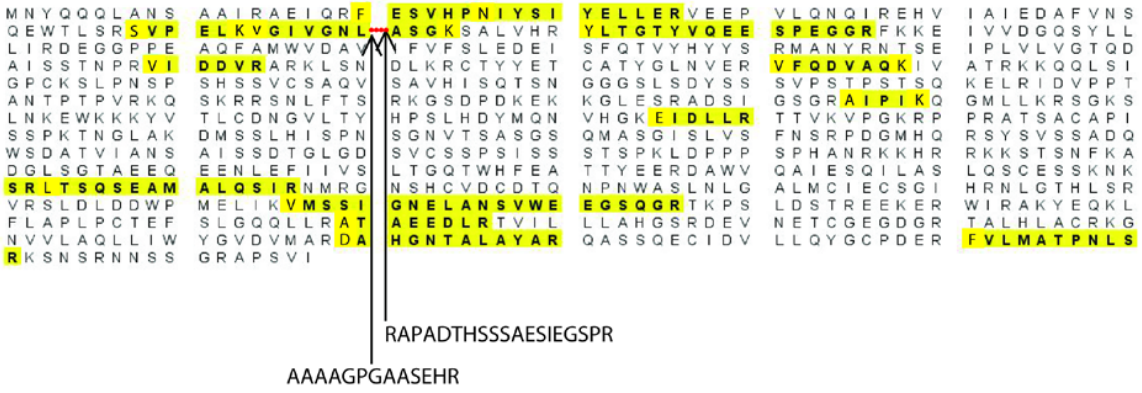
GIAILRR	GLAIIRR
GISYVRR	GLSYVRR
GLTLILRR	GLTLLLRR
GVIVRCR	GVLVRCR
ILISPLAR	ILLLSPAR
TLILGH	TLLIGH
VLIIWILK	VLLLWLLK

Table 3: The Table lists the seven pairs of peptides which matched two different non-coding RNAs that differed only by an isobaric I/L amino acid, while the unambiguous identification of the source of these peptides is not possible, it is clear that these peptides sequences are not found in the canonical mouse protein sequences.

The 250 high confidence peptides that were aligned to experimentally defined transcripts are referred to as novel peptides throughout the chapter. Interestingly, both annotated and novel peptides were found for some proteins. For example, the initial database search against the annotated Uniprot database resulted in the identification of 9 annotated, unique peptides for Centg2 (Agap1) with 17 % protein sequence coverage (**Figure 4A**). One of the canonical peptide's (VGIVGNLASGK) spectrum from Centg2 along with its fragmentation pattern is shown in **Figure 4A**. However, the database search against the custom transcriptome database resulted in the identification of two novel peptides from the first intron of Centg2 (Agap1) (**Figure 4B**). Validation of one of these novel peptides (AAAAGPGAASEHR) by comparing with its synthetic peptide is shown in Figure 5.

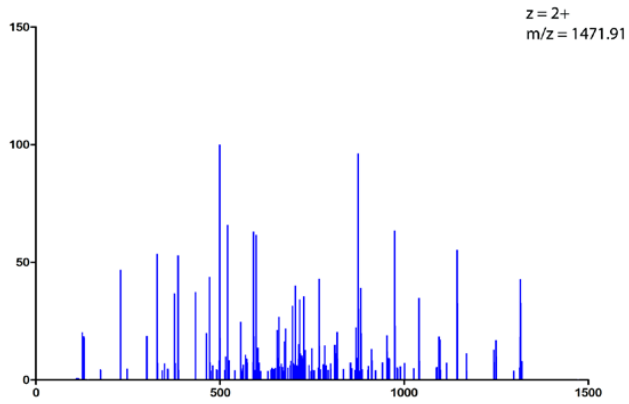
A

sp|Q8BXK8|AGAP1_MOUSE (100%), 94,412.1 Da
Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 1 OS=Mus musculus GN=Agap1 PE=2 SV=1
9 unique peptides, 10 unique spectra, 23 total spectra, 143/857 amino acids (17% coverage)



Identified Canonical Peptide

VGIVGNLASGK



Fragmentation Table

B	B Ions	B+2H	B-NH3	B-H2O	AA	Y Ions	Y+2H	Y-NH3	Y-H2O	Y
1	329.24				V+229	1,472.92	736.96	1,455.89	1,454.91	11
2	386.26				G	1,144.69	572.85	1,127.66	1,126.68	10
3	499.34				I	1,087.67	544.34	1,070.64	1,069.66	9
4	598.41				V	974.58	487.8	957.56	956.57	8
5	655.43				G	875.51	438.26	858.49	857.5	7
6	769.48	385.24	752.45		N	818.49	409.75	801.47	800.48	6
7	882.56	441.78	865.53		L	704.45		687.42	686.44	5
8	953.6	477.3	936.57		A	591.37		574.34	573.36	4
9	1,040.63	520.82	1,023.60	1,022.62	S	520.33		503.3	502.32	3
10	1,097.65	549.33	1,080.63	1,079.64	G	433.3		416.27		2
11	1,472.92	736.96	1,455.89	1,454.91	K+229	376.28		359.25		1

B

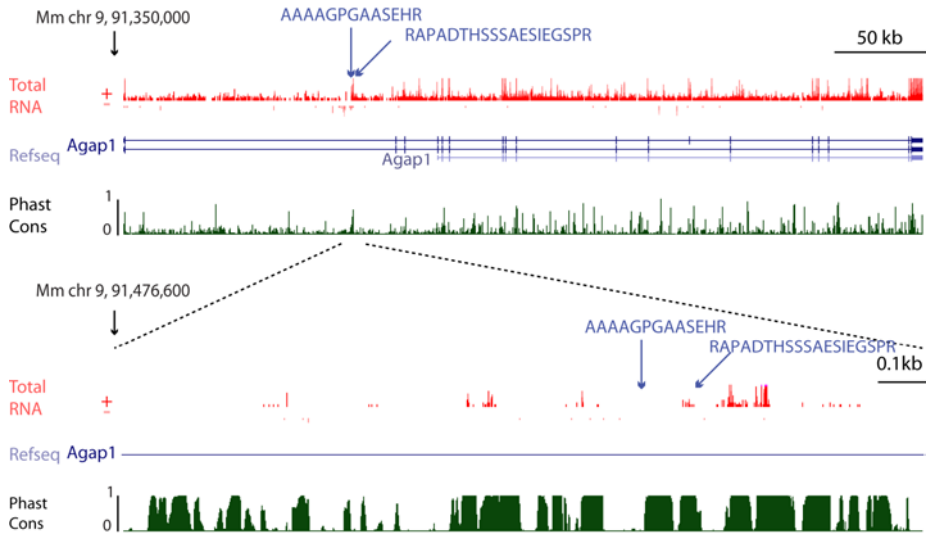


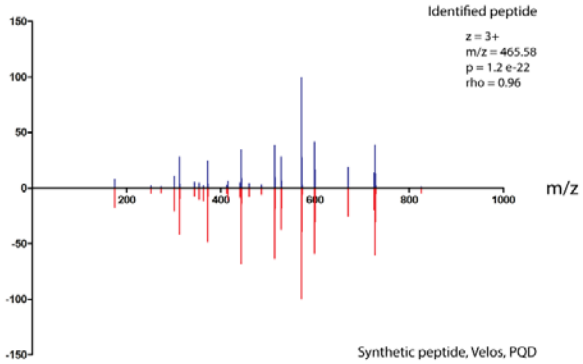
Figure 4. Known and novel peptides mapping to Centg2 (Agap1) (A) Amino acid sequence for Centg2 (Agap1), canonical peptides that were detected covering the amino acid sequence (17 %) are highlighted in yellow. The protein sequence is formatted into 10 residue blocks for readability and the location of the gap is a coincidence. The red dash indicated with double arrows show the novel exon region where the two novel peptides are identified. Spectrum of a canonical peptide VGIVGNLASGK along with its fragmentation pattern is shown. **(B)** Screenshot from the UCSC genome browser of the Centg2 (Agap1) locus for mouse where two novel peptides were identified overlapping the first intron.

Evaluation and validation of novel peptides

In addition to the high stringency cut-offs used to select peptides, several experiments were performed to evaluate the novelty of the peptides and the validity of the PSMs. First, each of the 250 novel peptide sequences was searched in non-redundant protein sequences from mouse using BLASTP and the UCSC genome browser. The results showed that 31 of the peptides corresponded to either pseudogenes or unannotated genes, which had been predicted to be translated, based on computational and transcriptional evidence, but lacked any supporting protein evidence. None of the remaining 214 peptides showed a similarity to the annotated proteome in Table 2. Second, each of the 250 novel peptide sequences was searched against both the peptide atlas (http://www.peptideatlas.org) and NIST-Libraries of Peptide Tandem Mass Spectra (http://peptide.nist.gov) databases. None of these novel peptides were identified in those databases, confirming that the novel peptides had not been reported in publicly available databases. Third, synthetic peptides were used to chemically validate the identity of the novel peptides. Fragmentation patterns of the synthetic peptides were compared with the mouse

peptides identified from the primary dataset. This is the most stringent method available for validating an identified peptide as it constitutes a chemically synthesized positive control. The fragmentation patterns of 45 peptides are presented (**Figure 5**). The similarity between the identified and synthetic spectra was evaluated by computing the rank-ordered Spearman Correlation coefficient. All the peptides showed strong correlation with a significant p-value ($p<0.05$). In addition, the average number of common fragment ions identified was greater than 90 %. Fourth, the 250 novel peptides were compared against published results from a ribosomal foot-printing experiment [14]. Of note, the ribosomal profiling data were performed on mouse embryonic stem cells, which have a different expression profile from mouse neurons. Nonetheless, 34 of the novel peptides were observed to be associated with ribosomes, providing independent evidence consistent with non-canonical translation events. These analyses and validation procedures lend support to the current proof-of-existence of novel peptides from non-canonical translational events.

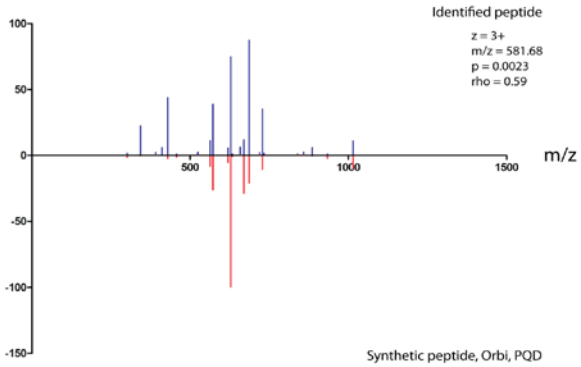
AAAAGPGAASEHR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	301.21	151.11	A+229	1,394.73	697.87	13
2	372.24	186.63	A	1,094.53	547.77	12
3	443.28	222.14	A	1,023.50	512.25	11
4	514.32	257.66	A	952.46	476.73	10
5	571.34	286.17	G	881.42	441.21	9
6	668.39	334.7	P	824.4	412.7	8
7	725.41	363.21	G	727.35	364.18	7
8	796.45	398.73	A	670.33	335.67	6
9	867.49	434.25	A	599.29	300.15	5
10	954.52	477.76	S	528.25	264.63	4
11	1,083.56	542.29	E	441.22	221.11	3
12	1,220.62	610.81	H	312.18	156.59	2
13	1,394.73	697.87	R	175.12	88.06	1

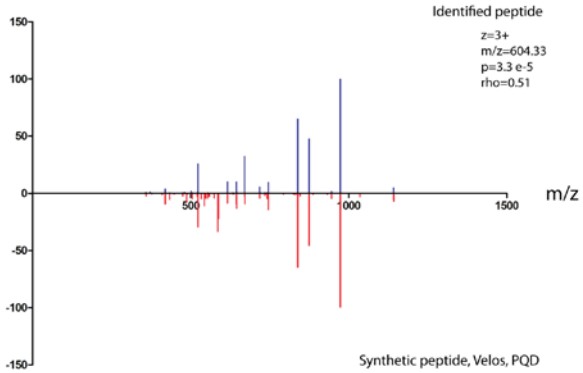
AADIQGASLGPRK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	301.21	151.11	A+229	1,742.03	871.52	13
2	372.24	186.63	A	1,441.83	721.42	12
3	487.27	244.14	D	1,370.80	685.9	11
4	600.36	300.68	I	1,255.77	628.39	10
5	728.41	364.71	Q	1,142.68	571.85	9
6	785.44	393.22	G	1,014.63	507.82	8
7	856.47	428.74	A	957.6	479.31	7
8	943.5	472.26	S	886.57	443.79	6
9	1,056.59	528.8	L	799.54	400.27	5
10	1,113.61	557.31	G	686.45	343.73	4
11	1,210.66	605.84	P	629.43	315.22	3
12	1,366.76	683.89	R	532.38	266.69	2
13	1,742.03	871.52	K+229	376.28	188.64	1

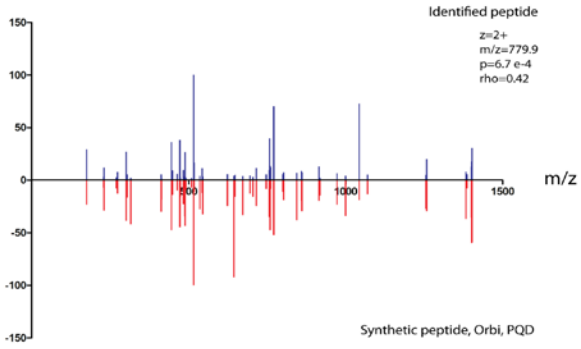
AAEDEEVPAFFK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	301.21	151.11	A+229	1,810.96	905.99	12
2	372.24	186.63	A	1,510.76	755.89	11
3	501.29	251.15	E	1,439.73	720.37	10
4	616.31	308.66	D	1,310.68	655.85	9
5	745.36	373.18	E	1,195.66	598.33	8
6	874.4	437.7	E	1,066.61	533.81	7
7	973.47	487.24	V	937.57	469.29	6
8	1,070.52	535.76	P	838.5	419.75	5
9	1,141.56	571.28	A	741.45	371.23	4
10	1,288.63	644.82	F	670.41	335.71	3
11	1,435.69	718.35	F	523.34	262.18	2
12	1,810.96	905.99	K+229	376.28	188.64	1

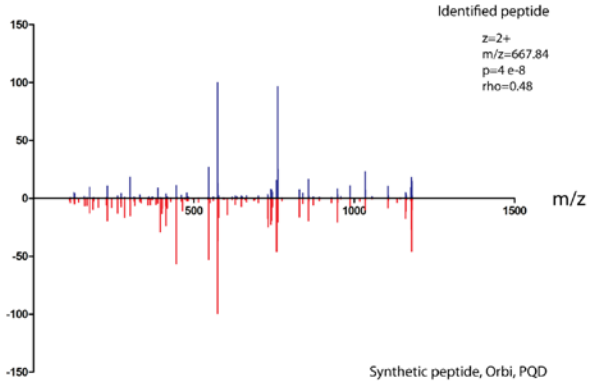
ASGAPSGSATAPAER



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	301.21		A+229	1,558.80	779.9	15
2	388.24		S	1,258.60	629.8	14
3	445.26		G	1,171.57	586.29	13
4	516.3		A	1,114.55	557.78	12
5	613.35		P	1,043.51	522.26	11
6	700.38	350.69	S	946.46	473.73	10
7	757.4	379.21	G	859.43	430.22	9
8	844.44	422.72	S	802.41	401.71	8
9	915.47	458.24	A	715.37	358.19	7
10	1,016.52	508.76	T	644.34	322.67	6
11	1,087.56	544.28	A	543.29		5
12	1,184.61	592.81	P	472.25		4
13	1,255.65	628.33	A	375.2		3
14	1,384.69	692.85	E	304.16		2
15	1,558.80	779.9	R	175.12		1

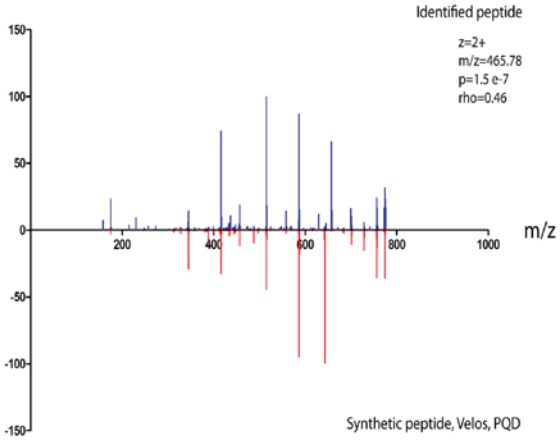
ASGEPSVESSR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	301.21	151.1	A+229	1,334.67	667.84	11
2	388.24	194.62	S	1,034.47	517.74	10
3	445.26	223.13	G	947.44	474.23	9
4	574.3	287.65	E	890.42	445.71	8
5	671.36	336.18	P	761.38	381.19	7
6	758.39	379.7	S	664.33	332.67	6
7	857.46	429.23	V	577.29	289.15	5
8	986.5	493.75	E	478.23	239.61	4
9	1,073.53	537.27	S	349.18	175.09	3
10	1,160.56	580.79	S	262.15	131.57	2
11	1,334.67	667.84	R	175.12	88.06	1

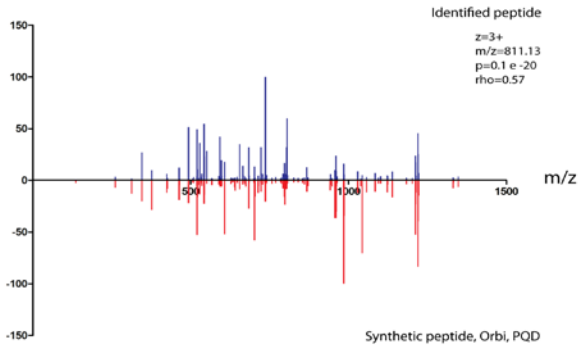
DAVAGIR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	345.2		D+229	930.56	465.78	7
2	416.23		A	586.37	293.69	6
3	515.3		V	515.33		5
4	586.34		A	416.26		4
5	643.36		G	345.22		3
6	756.45	378.73	I	288.2		2
7	930.56	465.78	R	175.12		1

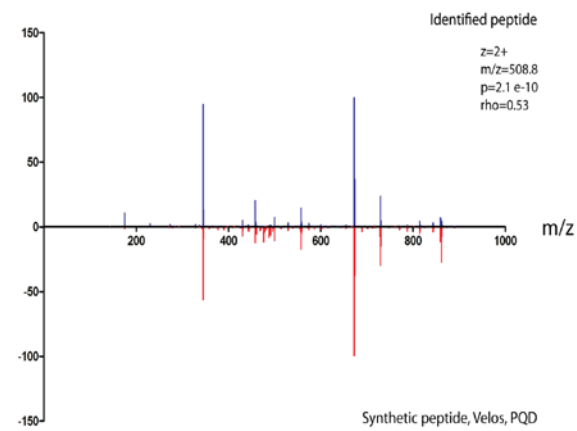
DKPANVQPYYLYGSK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	345.2	173.1	D+229	2,430.36	1,215.69	15
2	702.46	351.73	K+229	2,086.17	1,043.59	14
3	799.51	400.26	P	1,728.92	864.96	13
4	870.55	435.78	A	1,631.86	816.44	12
5	984.59	492.8	N	1,560.83	780.92	11
6	1,083.66	542.33	V	1,446.78	723.9	10
7	1,211.71	606.36	Q	1,347.71	674.36	9
8	1,308.77	654.89	P	1,219.66	610.33	8
9	1,471.83	736.42	Y	1,122.60	561.81	7
10	1,634.89	817.95	Y	959.54	480.27	6
11	1,747.98	874.49	L	796.48	398.74	5
12	1,911.04	956.02	Y	683.39	342.2	4
13	1,968.06	984.54	G	520.33	260.67	3
14	2,055.10	1,028.05	S	463.31	232.16	2
15	2,430.36	1,215.69	K+229	376.28	188.64	1

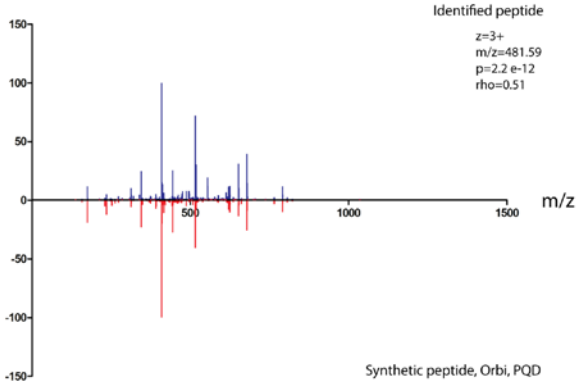
DDVLGLR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	345.2		D+229	1,016.59	508.8	7
2	460.22		D	672.4	336.71	6
3	559.29		V	557.38		5
4	672.38		L	458.31		4
5	729.4		G	345.22		3
6	842.48	421.74	L	288.2		2
7	1,016.59	508.8	R	175.12		1

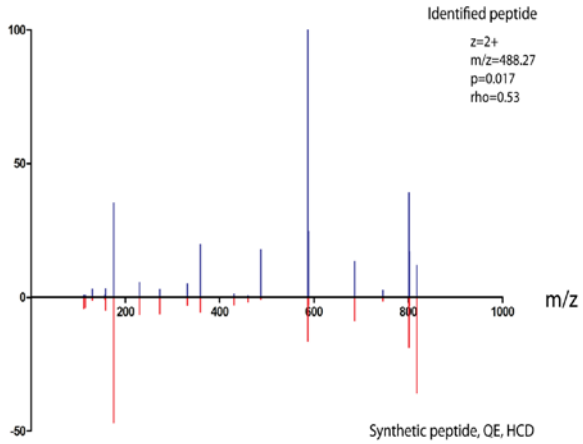
DVAHLGPDPHR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	345.2	173.1	D+229	1,442.77	721.89	11
2	444.27	222.64	V	1,098.58	549.79	10
3	515.3	258.15	A	999.51	500.26	9
4	652.36	326.68	H	928.47	464.74	8
5	765.45	383.23	L	791.42	396.21	7
6	822.47	411.74	G	678.33	339.67	6
7	919.52	460.26	P	621.31	311.16	5
8	1,034.55	517.78	D	524.26	262.63	4
9	1,131.60	566.3	P	409.23	205.12	3
10	1,268.66	634.83	H	312.18	156.59	2
11	1,442.77	721.89	R	175.12	88.06	1

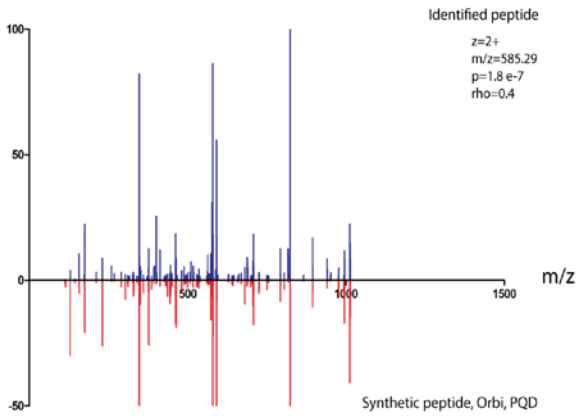
EEVVDR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	359.21		E+229	975.53	488.27	6
2	488.26		E	617.33		5
3	587.32		V	488.28		4
4	686.39		V	389.21		3
5	801.42		D	290.15		2
6	975.53	488.27	R	175.12		1

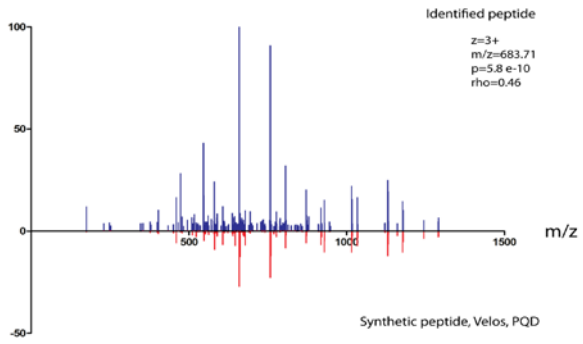
FSDEDATR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	377.24		F+229	1,169.56	585.29	8
2	464.27		S	793.33	397.17	7
3	579.3		D	706.3	353.65	6
4	708.34		E	591.27		5
5	823.37		D	462.23		4
6	894.4	447.71	A	347.2		3
7	995.45	498.23	T	276.17		2
8	1,169.56	585.29	R	175.12		1

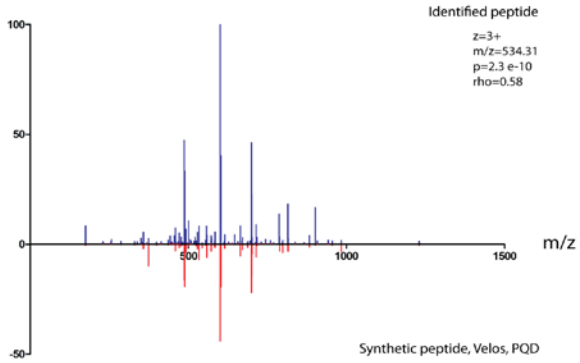
FTQAQLDSGLVLFSHR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	377.24	189.12	F+229	2,048.11	1,024.56	16
2	478.29	239.65	T	1,671.88	836.44	15
3	606.34	303.68	Q	1,570.83	785.92	14
4	677.38	339.19	A	1,442.78	721.89	13
5	805.44	403.22	Q	1,371.74	686.37	12
6	918.52	459.77	L	1,243.68	622.34	11
7	1,033.55	517.28	D	1,130.60	565.8	10
8	1,120.58	560.8	S	1,015.57	508.29	9
9	1,177.61	589.31	G	928.54	464.77	8
10	1,290.69	645.85	L	871.51	436.26	7
11	1,389.76	695.38	V	758.43	379.72	6
12	1,502.84	751.92	L	659.36	330.18	5
13	1,649.91	825.46	F	546.28	273.64	4
14	1,736.94	868.97	S	399.21	200.11	3
15	1,874.00	937.5	H	312.18	156.59	2
16	2,048.11	1,024.56	R	175.12	88.06	1

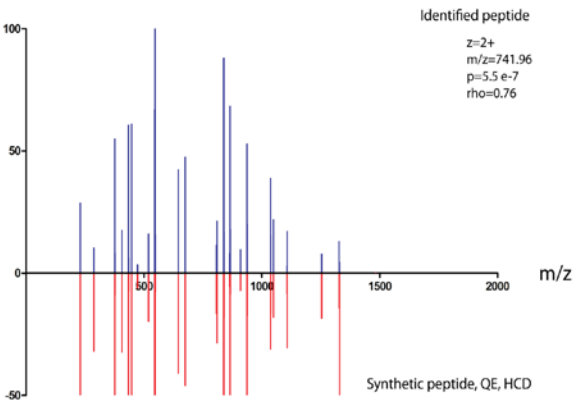
GAELVDSVLDVVR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	287.19	144.1	G+229	1,600.91	800.96	13
2	358.23	179.62	A	1,314.73	657.87	12
3	487.27	244.14	E	1,243.69	622.35	11
4	600.36	300.68	L	1,114.65	557.83	10
5	699.42	350.22	V	1,001.56	501.28	9
6	814.45	407.73	D	902.49	451.75	8
7	901.48	451.25	S	787.47	394.24	7
8	1,000.55	500.78	V	700.44	350.72	6
9	1,113.64	557.32	L	601.37	301.19	5
10	1,228.66	614.83	D	488.28	244.65	4
11	1,327.73	664.37	V	373.26	187.13	3
12	1,426.80	713.9	V	274.19	137.6	2
13	1,600.91	800.96	R	175.12	88.06	1

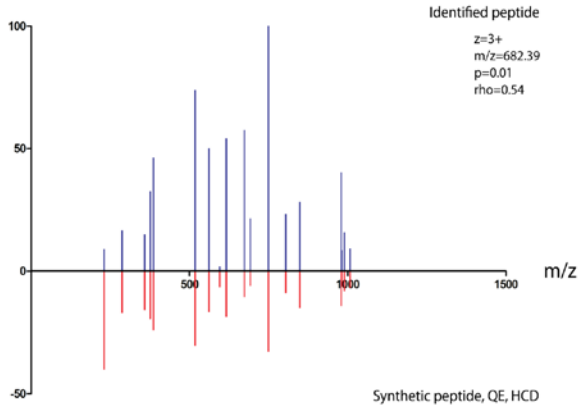
GFLAGYVVAK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	287.19		G+229	1,482.91	741.96	10
2	434.26		F	1,196.72	598.87	9
3	547.34		L	1,049.66	525.33	8
4	618.38		A	936.57	468.79	7
5	675.4		G	865.53	433.27	6
6	838.47	419.74	Y	808.51		5
7	937.53	469.27	V	645.45		4
8	1,036.60	518.81	V	546.38		3
9	1,107.64	554.32	A	447.31		2
10	1,482.91	741.96	K+229	376.28		1

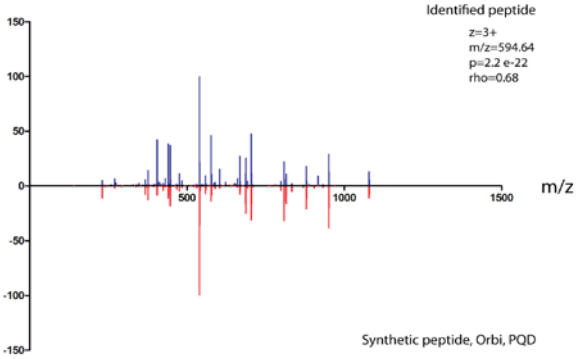
GVMVGMGQK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	287.19		G+229	1,364.78	682.89	9
2	386.26		V	1,078.60	539.8	8
3	517.3		M	979.53	490.27	7
4	616.37		V	848.49	424.75	6
5	673.39		G	749.42		5
6	804.43	402.72	M	692.4		4
7	861.45	431.23	G	561.36		3
8	989.51	495.26	Q	504.33		2
9	1,364.78	682.89	K+229	376.28		1

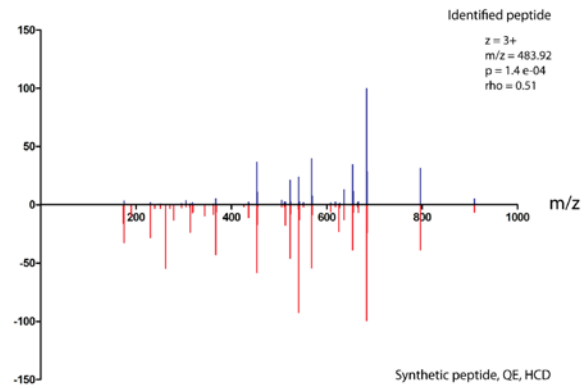
HGDPGDAAQQEAK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	367.23	184.12	H+229	1,781.92	891.46	13
2	424.25	212.63	G	1,415.70	708.35	12
3	539.28	270.14	D	1,358.68	679.84	11
4	636.33	318.67	P	1,243.65	622.33	10
5	693.35	347.18	G	1,146.60	573.8	9
6	808.38	404.69	D	1,089.57	545.29	8
7	879.42	440.21	A	974.55	487.78	7
8	950.45	475.73	A	903.51	452.26	6
9	1,078.51	539.76	Q	832.47	416.74	5
10	1,206.57	603.79	Q	704.41	352.71	4
11	1,335.61	668.31	E	576.36	288.68	3
12	1,406.65	703.83	A	447.31	224.16	2
13	1,781.92	891.46	K+229	376.28	188.64	1

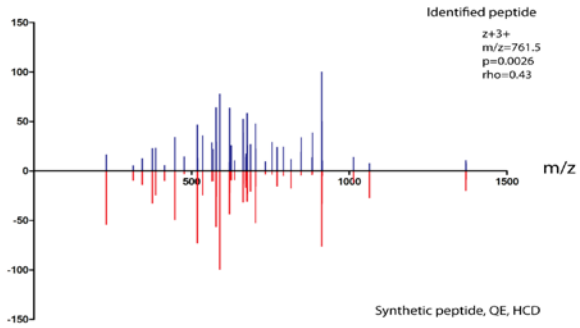
HSSLIDDMFR



Overlapped Fragmentation Table

B	B Ions	AA	Y Ions	Y
1	367.23	H+229	1,449.74	10
2	454.26	S	1,083.51	9
3	541.29	S	996.48	8
4	654.38	L	909.45	7
5	767.46	I	796.37	6
6	882.49	D	683.28	5
7	997.52	D	568.25	4
8	1,128.56	M	453.23	3
9	1,275.62	F	322.19	2
10	1,449.74	R	175.12	1

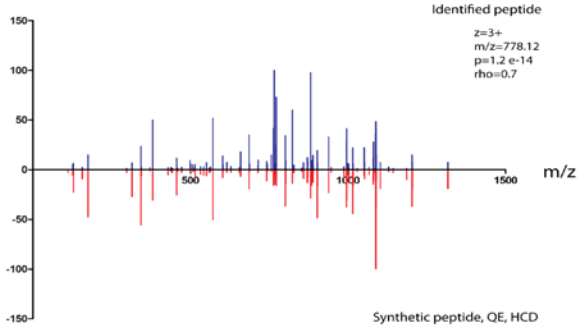
IAKPLSSLTPLIAAAK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25	172.13	I+229	2,281.48	1,141.25	16
2	414.29	207.65	A	1,939.24	970.12	15
3	771.55	386.28	K+229	1,868.20	934.6	14
4	868.6	434.8	P	1,510.94	755.97	13
5	981.69	491.35	L	1,413.89	707.45	12
6	1,068.72	534.86	S	1,300.80	650.91	11
7	1,155.75	578.38	S	1,213.77	607.39	10
8	1,268.83	634.92	L	1,126.74	563.87	9
9	1,369.88	685.44	T	1,013.66	507.33	8
10	1,466.93	733.97	P	912.61	456.81	7
11	1,580.02	790.51	L	815.56	408.28	6
12	1,693.10	847.06	I	702.47	351.74	5
13	1,764.14	882.57	A	589.39	295.2	4
14	1,835.18	918.09	A	518.35	259.68	3
15	1,906.21	953.61	A	447.31	224.16	2
16	2,281.48	1,141.25	K+229	376.28	188.64	1

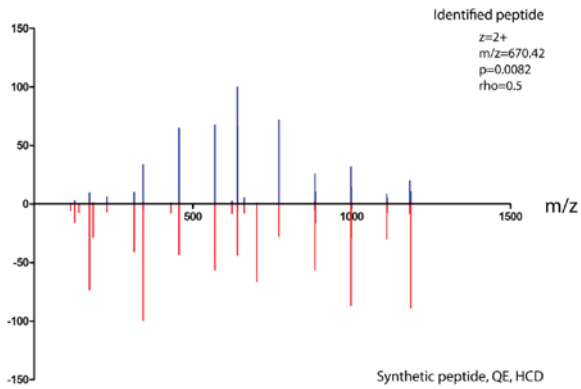
ILDDDTIITTLENLKR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25	172.13	I+229	2,331.35	1,166.18	16
2	456.34	228.67	L	1,989.11	995.06	15
3	571.37	286.19	D	1,876.02	938.52	14
4	686.39	343.7	D	1,761.00	881	13
5	801.42	401.21	D	1,645.97	823.49	12
6	902.47	451.74	T	1,530.94	765.97	11
7	1,015.55	508.28	I	1,429.89	715.45	10
8	1,128.64	564.82	I	1,316.81	658.91	9
9	1,229.68	615.35	T	1,203.73	602.37	8
10	1,330.73	665.87	T	1,102.68	551.84	7
11	1,443.81	722.41	L	1,001.63	501.32	6
12	1,572.86	786.93	E	888.55	444.78	5
13	1,686.90	843.95	N	759.5	380.26	4
14	1,799.98	900.5	L	645.46	323.23	3
15	2,157.24	1,079.12	K+229	532.38	266.69	2
16	2,331.35	1,166.18	R	175.12	88.06	1

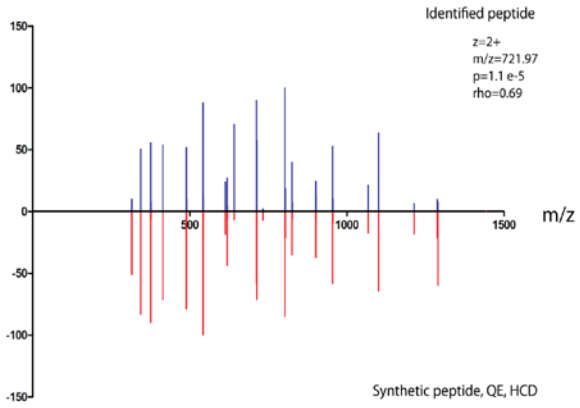
IILMPSLPAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25		I+229	1,339.83	670.42	10
2	456.34		I	997.59	499.3	9
3	569.42		L	884.5	442.75	8
4	700.46		M	771.42	386.21	7
5	797.52		P	640.38	320.69	6
6	884.55	442.78	S	543.32		5
7	997.63	499.32	L	456.29		4
8	1,094.68	547.85	P	343.21		3
9	1,165.72	583.36	A	246.16		2
10	1,339.83	670.42	R	175.12		1

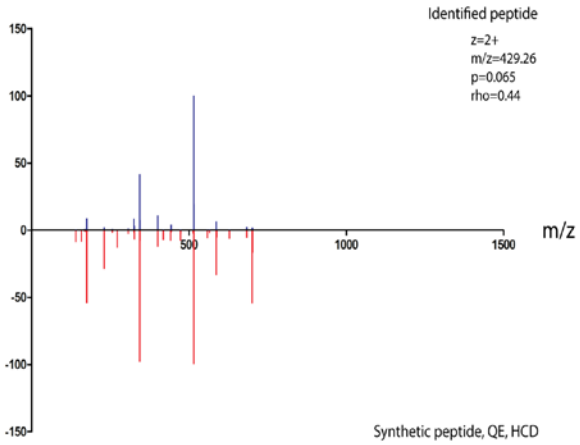
LAQVALELK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25		L+229	1,442.93	721.97	9
2	414.29		A	1,100.69	550.85	8
3	542.35		Q	1,029.65	515.33	7
4	641.42		V	901.59	451.3	6
5	712.46		A	802.52		5
6	825.54	413.27	L	731.49		4
7	954.58	477.79	E	618.4		3
8	1,067.67	534.34	L	489.36		2
9	1,442.93	721.97	K+229	376.28		1

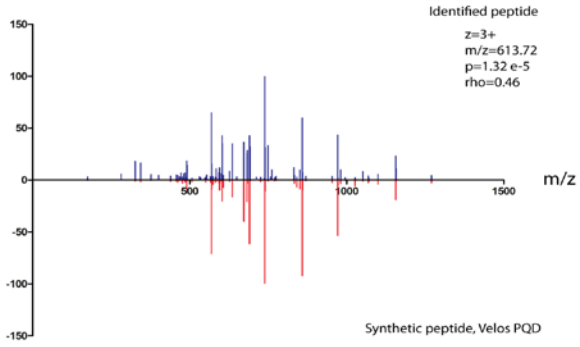
LGDAPR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25		L+229	857.5	429.26	6
2	400.28		G	515.26		5
3	515.3		D	458.24		4
4	586.34		A	343.21		3
5	683.39		P	272.17		2
6	857.5	429.26	R	175.12		1

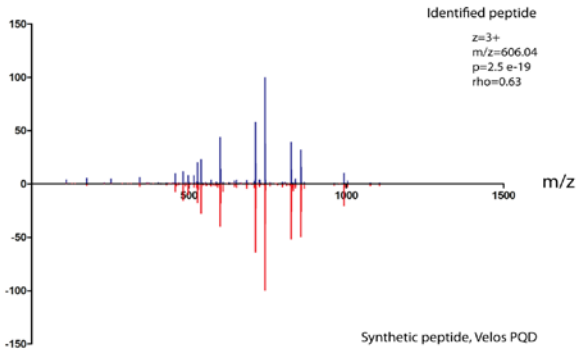
LFLSPAIQGLLLPAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25	172.13	L+229	1,838.15	919.58	15
2	490.32	245.66	F	1,495.90	748.45	14
3	603.41	302.21	L	1,348.83	674.92	13
4	690.44	345.72	S	1,235.75	618.38	12
5	787.49	394.25	P	1,148.72	574.86	11
6	858.53	429.77	A	1,051.66	526.33	10
7	971.61	486.31	I	980.63	490.82	9
8	1,099.67	550.34	Q	867.54	434.27	8
9	1,156.69	578.85	G	739.48	370.24	7
10	1,269.78	635.39	L	682.46	341.73	6
11	1,382.86	691.93	L	569.38	285.19	5
12	1,495.95	748.48	L	456.29	228.65	4
13	1,593.00	797	P	343.21	172.11	3
14	1,664.04	832.52	A	246.16	123.58	2
15	1,838.15	919.58	R	175.12	88.06	1

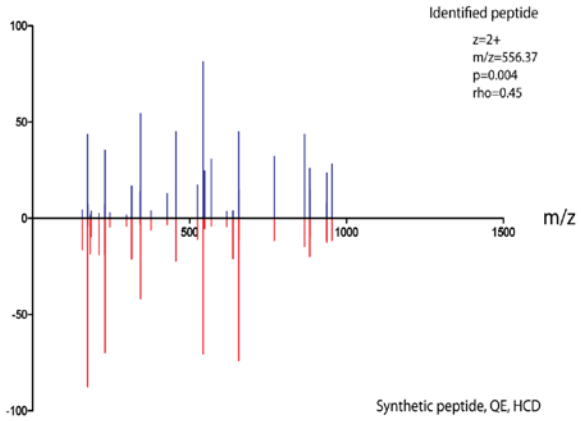
LLAALLHNPQLVER



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25	172.13	L+229	1,816.10	908.55	14
2	456.34	228.67	L	1,473.85	737.43	13
3	527.38	264.19	A	1,360.77	680.89	12
4	598.41	299.71	A	1,289.73	645.37	11
5	711.5	356.25	L	1,218.70	609.85	10
6	824.58	412.79	L	1,105.61	553.31	9
7	961.64	481.32	H	992.53	496.77	8
8	1,075.68	538.34	N	855.47	428.24	7
9	1,172.74	586.87	P	741.43	371.22	6
10	1,300.79	650.9	Q	644.37	322.69	5
11	1,413.88	707.44	L	516.31	258.66	4
12	1,512.95	756.98	V	403.23	202.12	3
13	1,641.99	821.5	E	304.16	152.58	2
14	1,816.10	908.55	R	175.12	88.06	1

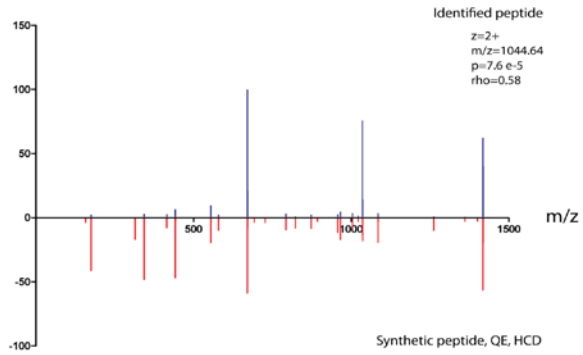
LLSIPLAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25		L+229	1,111.74	556.37	8
2	456.34		L	769.49	385.25	7
3	543.37		S	656.41	328.71	6
4	656.45		L	569.38		5
5	769.54		I	456.29		4
6	866.59	433.8	P	343.21		3
7	937.63	469.32	A	246.16		2
8	1,111.74	556.37	R	175.12		1

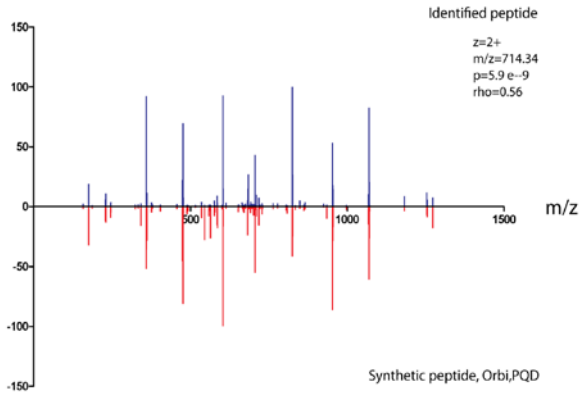
LVIDGKPITIFQER



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	343.25		L+229	2,087.26	1,044.14	14
2	442.32		V	1,745.02	873.01	13
3	555.41		I	1,645.95	823.48	12
4	670.43		D	1,532.86	766.94	11
5	727.46		G	1,417.84	709.42	10
6	1,084.71	542.86	K+229	1,360.82	680.91	9
7	1,181.77	591.39	P	1,003.56	502.28	8
8	1,294.85	647.93	I	906.5	453.76	7
9	1,395.90	698.45	T	793.42	397.21	6
10	1,508.98	754.99	I	692.37		5
11	1,656.05	828.53	F	579.29		4
12	1,784.11	892.56	Q	432.22		3
13	1,913.15	957.08	E	304.16		2
14	2,087.26	1,044.14	R	175.12		1

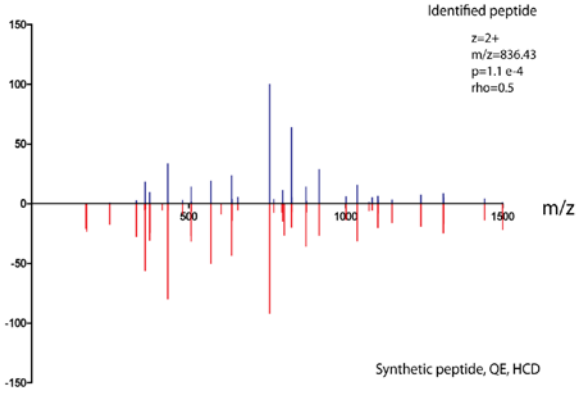
MDSFDEDLAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	361.21		M+229	1,427.67	714.34	10
2	476.24		D	1,067.46	534.24	9
3	563.27		S	952.44	476.72	8
4	710.34		F	865.41	433.21	7
5	825.37		D	718.34	359.67	6
6	954.41	477.71	E	603.31	302.15	5
7	1,069.43	535.22	D	474.27	237.63	4
8	1,182.52	591.76	L	359.24	180.12	3
9	1,253.56	627.28	A	246.16		2
10	1,427.67	714.34	R	175.12		1

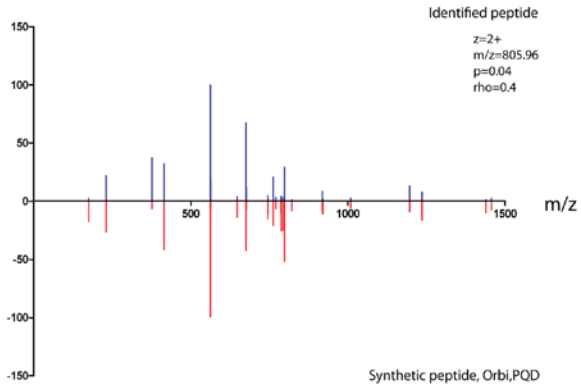
MFQYDSTHGK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	361.21		M+229	1,671.86	836.43	10
2	508.28		F	1,311.65	656.33	9
3	636.34		Q	1,164.58	582.8	8
4	799.4		Y	1,036.53	518.77	7
5	914.43		D	873.46	437.24	6
6	1,001.46	501.23	S	758.44	379.72	5
7	1,102.51	551.76	T	671.4	336.21	4
8	1,239.57	620.29	H	570.36	285.68	3
9	1,296.59	648.8	G	433.3		2
10	1,671.86	836.43	K+229	376.28		1

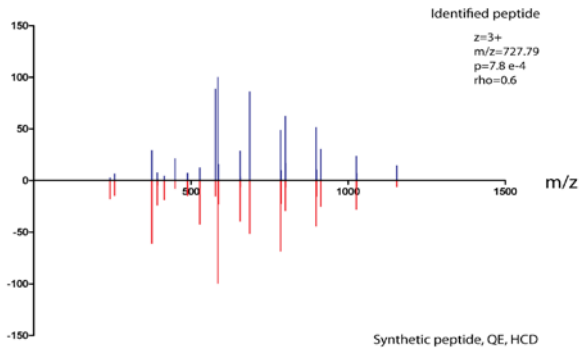
NAFLSSSALDK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	344.21		N+229	1,610.92	805.96	11
2	415.25		A	1,267.71	634.36	10
3	562.32		F	1,196.67	598.84	9
4	675.4		L	1,049.60	525.31	8
5	762.43		S	936.52	468.76	7
6	849.47	425.24	S	849.49	425.25	6
7	936.5	468.75	S	762.46		5
8	1,007.54	504.27	A	675.42		4
9	1,120.62	560.81	L	604.39		3
10	1,235.65	618.33	D	491.3		2
11	1,610.92	805.96	K+229	376.28		1

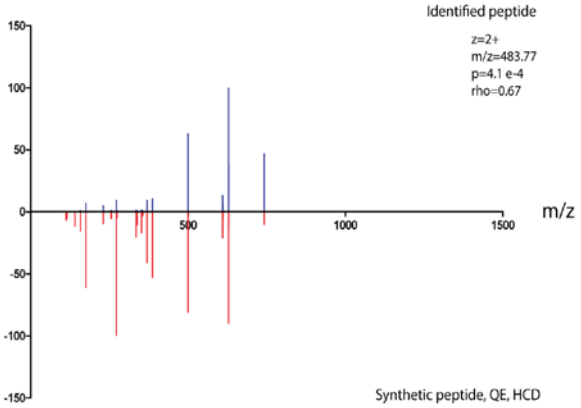
NALQELQQIIITPIK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	344.21	172.61	N+229	2,180.34	1,090.67	15
2	415.25	208.13	A	1,837.14	919.07	14
3	528.33	264.67	L	1,766.10	883.55	13
4	656.39	328.7	Q	1,653.02	827.01	12
5	785.44	393.22	E	1,524.96	762.98	11
6	898.52	449.76	L	1,395.91	698.46	10
7	1,026.58	513.79	Q	1,282.83	641.92	9
8	1,154.64	577.82	Q	1,154.77	577.89	8
9	1,267.72	634.36	I	1,026.71	513.86	7
10	1,380.80	690.91	I	913.63	457.32	6
11	1,493.89	747.45	I	800.54	400.78	5
12	1,594.94	797.97	T	687.46	344.23	4
13	1,691.99	846.5	P	586.41	293.71	3
14	1,805.07	903.04	I	489.36	245.18	2
15	2,180.34	1,090.67	K+229	376.28	188.64	1

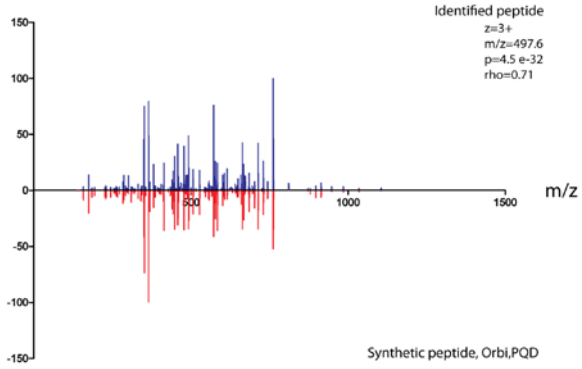
QIIEINPR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	341.2		Q+212	1,194.70	597.86	8
2	454.29		I	854.51	427.76	7
3	567.37		I	741.43	371.22	6
4	696.41		E	628.34		5
5	809.5		I	499.3		4
6	923.54	462.27	N	386.21		3
7	1,020.59	510.8	P	272.17		2
8	1,194.70	597.86	R	175.12		1

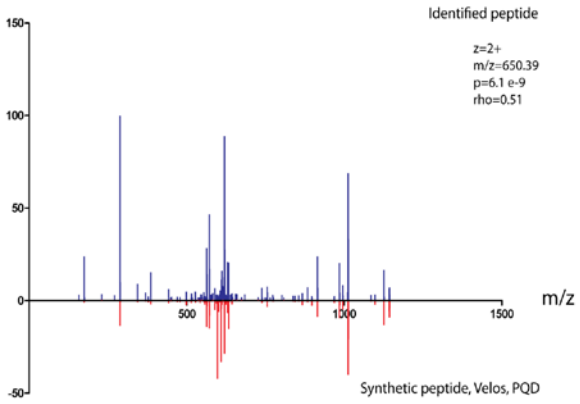
RASGEPSVESSR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	386.27	193.64	R+229	1,490.78	745.89	12
2	457.31	229.16	A	1,105.51	553.26	11
3	544.34	272.67	S	1,034.47	517.74	10
4	601.36	301.18	G	947.44	474.23	9
5	730.4	365.71	E	890.42	445.71	8
6	827.46	414.23	P	761.38	381.19	7
7	914.49	457.75	S	664.33	332.67	6
8	1,013.56	507.28	V	577.29	289.15	5
9	1,142.60	571.8	E	478.23	239.62	4
10	1,229.63	615.32	S	349.18	175.1	3
11	1,316.66	658.84	S	262.15	131.58	2
12	1,490.78	745.89	R	175.12	88.06	1

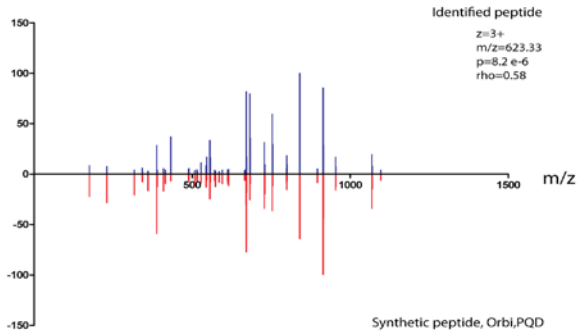
RQQLQELR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	386.27	193.64	R+229	1,299.77	650.39	8
2	514.33	257.67	Q	914.51	457.76	7
3	642.39	321.7	Q	786.45	393.73	6
4	755.47	378.24	L	658.39		5
5	883.53	442.27	Q	545.3		4
6	1,012.57	506.79	E	417.25		3
7	1,125.66	563.33	L	288.2		2
8	1,299.77	650.39	R	175.12		1

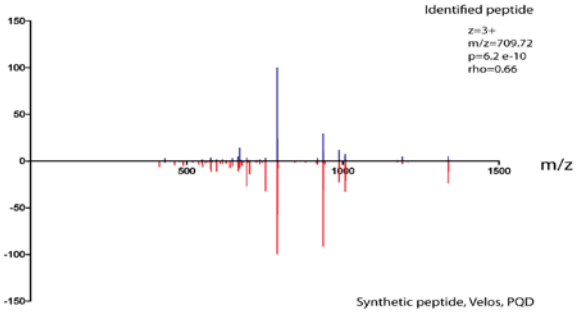
SAPADGADLSAHLWAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	317.2	159.1	S+229	1,866.97	933.99	16
2	388.24	194.62	A	1,550.77	775.89	15
3	485.29	243.15	P	1,479.73	740.37	14
4	556.33	278.67	A	1,382.68	691.84	13
5	671.36	336.18	D	1,311.64	656.33	12
6	728.38	364.69	G	1,196.62	598.81	11
7	799.41	400.21	A	1,139.60	570.3	10
8	914.44	457.72	D	1,068.56	534.78	9
9	1,027.53	514.27	L	953.53	477.27	8
10	1,114.56	557.78	S	840.45	420.73	7
11	1,185.59	593.3	A	753.42	377.21	6
12	1,322.65	661.83	H	682.38	341.69	5
13	1,435.74	718.37	L	545.32	273.16	4
14	1,621.82	811.41	W	432.24	216.62	3
15	1,692.85	846.93	A	246.16	123.58	2
16	1,866.97	933.99	R	175.12	88.06	1

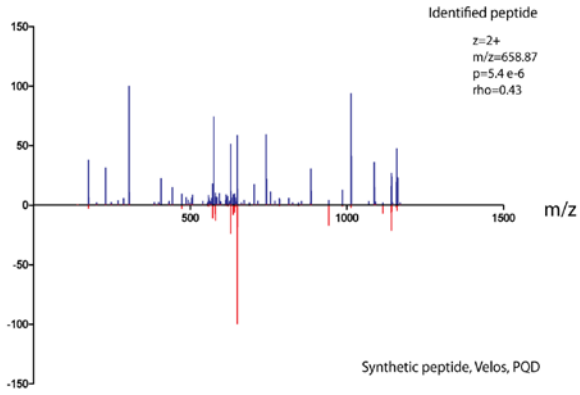
SLFSSPASLAFPYSPVAR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	317.2	159.1	S+229	2,126.15	1,063.58	18
2	430.29	215.65	L	1,809.95	905.48	17
3	577.35	289.18	F	1,696.87	848.94	16
4	664.39	332.7	S	1,549.80	775.4	15
5	751.42	376.21	S	1,462.77	731.89	14
6	848.47	424.74	P	1,375.74	688.37	13
7	919.51	460.26	A	1,278.68	639.85	12
8	1,006.54	503.77	S	1,207.65	604.33	11
9	1,119.62	560.32	L	1,120.61	560.81	10
10	1,190.66	595.83	A	1,007.53	504.27	9
11	1,337.73	669.37	F	936.49	468.75	8
12	1,434.78	717.9	P	789.43	395.22	7
13	1,597.85	799.43	Y	692.37	346.69	6
14	1,684.88	842.94	S	529.31	265.16	5
15	1,781.93	891.47	P	442.28	221.64	4
16	1,881.00	941	V	345.22	173.12	3
17	1,952.04	976.52	A	246.16	123.58	2
18	2,126.15	1,063.58	R	175.12	88.06	1

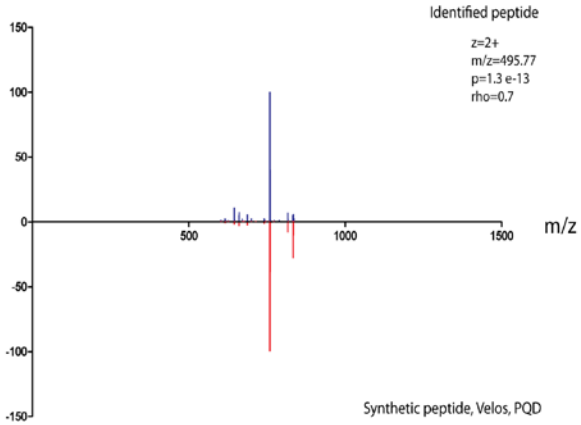
SRVGLAEAER



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	317.2		S+229	1,316.75	658.88	10
2	473.3	237.16	R	1,000.55	500.78	9
3	572.37	286.69	V	844.45	422.73	8
4	629.39	315.2	G	745.38	373.2	7
5	742.48	371.74	L	688.36	344.68	6
6	813.51	407.26	A	575.28		5
7	942.56	471.78	E	504.24		4
8	1,013.59	507.3	A	375.2		3
9	1,142.64	571.82	E	304.16		2
10	1,316.75	658.88	R	175.12		1

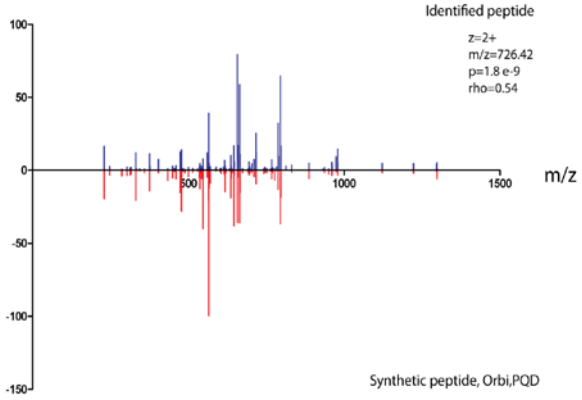
TELADGR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	331.22		T+229	990.54	495.77	7
2	460.26		E	660.33	330.67	6
3	573.34		L	531.29		5
4	644.38		A	418.2		4
5	759.41		D	347.17		3
6	816.43	408.72	G	232.14		2
7	990.54	495.77	R	175.12		1

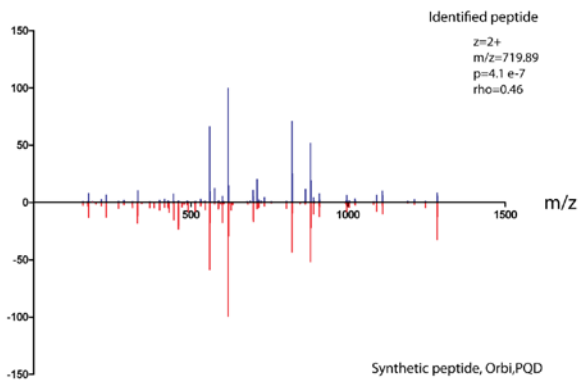
TFSVMPSPK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	331.22		T+229	1,451.83	726.42	9
2	478.29		F	1,121.62	561.32	8
3	565.32		S	974.55	487.78	7
4	664.39		V	887.52	444.26	6
5	795.43		M	788.45		5
6	892.48	446.74	P	657.41		4
7	979.51	490.26	S	560.36		3
8	1,076.56	538.79	P	473.33		2
9	1,451.83	726.42	K+229	376.28		1

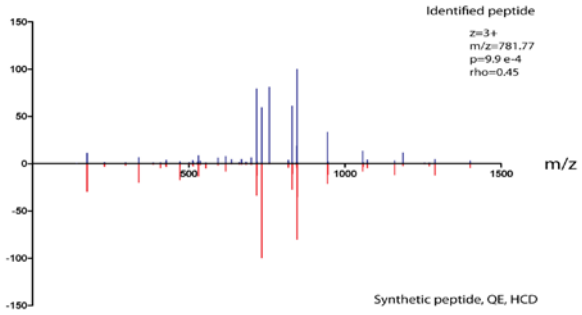
TIDFDSLSVGR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	331.22		T+229	1,438.77	719.89	11
2	444.3		I	1,108.56	554.79	10
3	559.33		D	995.48	498.24	9
4	706.4		F	880.45	440.73	8
5	821.42		D	733.38	367.2	7
6	908.46	454.73	S	618.36	309.68	6
7	1,021.54	511.27	L	531.32		5
8	1,108.57	554.79	S	418.24		4
9	1,207.64	604.32	V	331.21		3
10	1,264.66	632.83	G	232.14		2
11	1,438.77	719.89	R	175.12		1

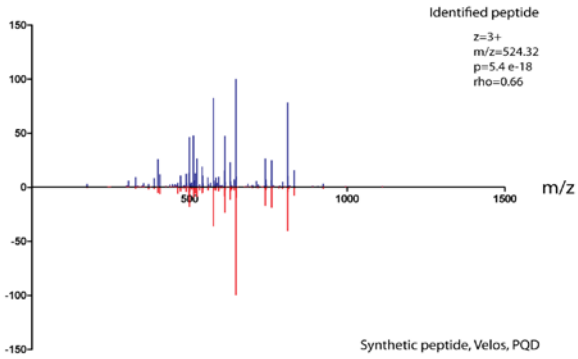
TSTVDLPIESQLLWQLDR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	331.22	166.11	T+229	2,343.28	1,172.14	18
2	418.25	209.63	S	2,013.07	1,007.04	17
3	519.3	260.15	T	1,926.03	963.52	16
4	618.37	309.69	V	1,824.99	913	15
5	733.39	367.2	D	1,725.92	863.46	14
6	846.48	423.74	L	1,610.89	805.95	13
7	943.53	472.27	P	1,497.81	749.41	12
8	1,056.61	528.81	I	1,400.75	700.88	11
9	1,185.66	593.33	E	1,287.67	644.34	10
10	1,272.69	636.85	S	1,158.63	579.82	9
11	1,400.75	700.88	Q	1,071.59	536.3	8
12	1,513.83	757.42	L	943.54	472.27	7
13	1,626.92	813.96	L	830.45	415.73	6
14	1,812.99	907	W	717.37	359.19	5
15	1,941.05	971.03	Q	531.29	266.15	4
16	2,054.14	1,027.57	L	403.23	202.12	3
17	2,169.16	1,085.09	D	290.15	145.58	2
18	2,343.28	1,172.14	R	175.12	88.06	1

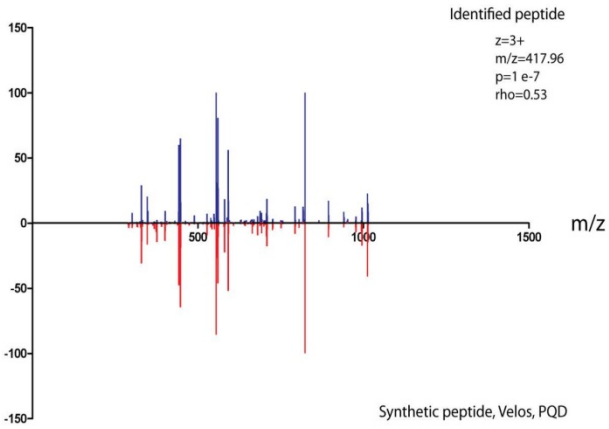
VAVLQALASTVNR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	329.24	165.12	V+229	1,570.95	785.98	13
2	400.28	200.64	A	1,242.72	621.86	12
3	499.34	250.18	V	1,171.68	586.34	11
4	612.43	306.72	L	1,072.61	536.81	10
5	740.49	370.75	Q	959.53	480.27	9
6	811.52	406.27	A	831.47	416.24	8
7	924.61	462.81	L	760.43	380.72	7
8	995.65	498.33	A	647.35	324.18	6
9	1,082.68	541.84	S	576.31	288.66	5
10	1,183.72	592.37	T	489.28	245.14	4
11	1,282.79	641.9	V	388.23	194.62	3
12	1,396.84	698.92	N	289.16	145.08	2
13	1,570.95	785.98	R	175.12	88.06	1

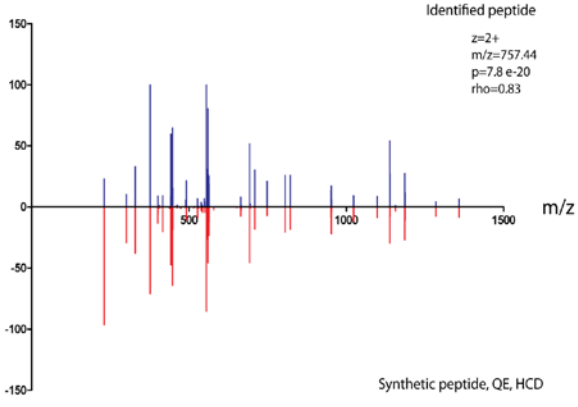
VLLHLAK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	329.24	165.12	V+229	1,251.86	626.43	7
2	442.32	221.66	L	923.62	462.32	6
3	555.41	278.21	L	810.54	405.77	5
4	692.47	346.74	H	697.46	349.23	4
5	805.55	403.28	L	560.4	280.7	3
6	876.59	438.8	A	447.31	224.16	2
7	1,251.86	626.43	K+229	376.28	188.64	1

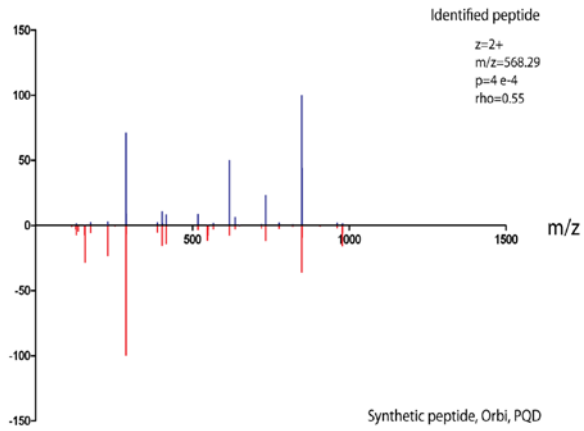
VSFELFADK



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	329.24		V+229	1,513.87	757.44	9
2	416.27		S	1,185.64	593.32	8
3	563.34		F	1,098.60	549.81	7
4	692.38		E	951.54	476.27	6
5	805.47		L	822.49		5
6	952.53	476.77	F	709.41		4
7	1,023.57	512.29	A	562.34		3
8	1,138.60	569.8	D	491.3		2
9	1,513.87	757.44	K+229	376.28		1

WMADDLR



Overlapped Fragmentation Table

B	B Ions	B+2H	AA	Y Ions	Y+2H	Y
1	416.25		W+229	1,135.58	568.30	7
2	547.29		M	720.33	360.7	6
3	618.33		A	589.29		5
4	733.35		D	518.26		4
5	848.38		D	403.23		3
6	961.47	481.24	L	288.2		2
7	1,135.58	568.29	R	175.12		1

Figure 5. Validation of identified spectral matches using synthetic peptides. Spectra of 44 novel peptides (top) were compared and statistically validated to corresponding synthetic peptides (bottom). A single peptide of ILDDDTIITTLENLK is not shown here. All these peptides show as strong positive rank-ordered Spearman correlation coefficient with a significant p-value of less than 0.05. The fragmentation table indicates the y and b ion series for the peptides, ion depicted in green color indicate that they were identified in both the spectra, ion depicted in blue color were identified only in 'identified novel peptide spectra' (blue) and ions in pink color were identified in only in synthetic peptide spectra (pink). Novel peptides include the ones described earlier as - for the peptide LVIDGKPITIFQER from the pseudogene PGOMOU00000135506 on chromosome 11 (described in Figure 7B and number 13 in RTPCR gel Figure 7A); for the peptide IILMPSLPAR from extragenic regions overlapping the anti-sense region of Cox17; for the peptide AAAAGPGAASEHR from Centg2 (Agap1) (Figure 4)**;** and for the peptide HSSLIDDMFR from Farp1 (Figure 8).

Mapping of novel peptides to genomic regions. The novel peptides were mapped to their genomic origins and compared with existing annotations and the RNA-seq data. These comparisons led to a categorization of the novel peptides into 5 sub-groups (**Figure 6**).

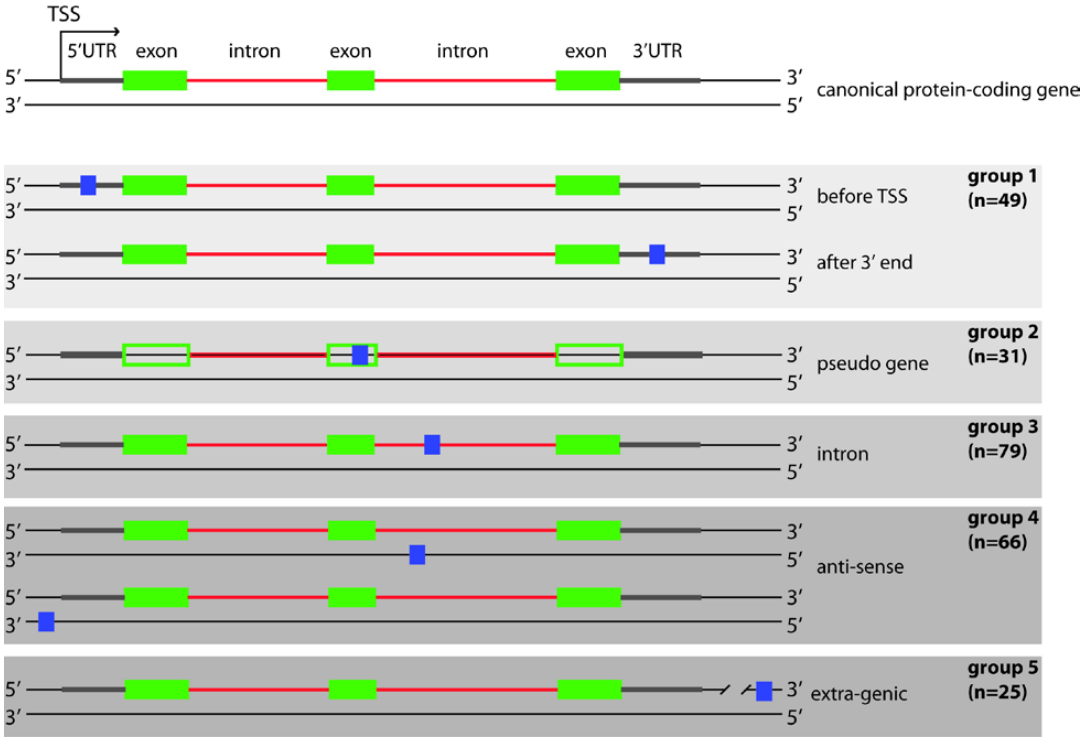
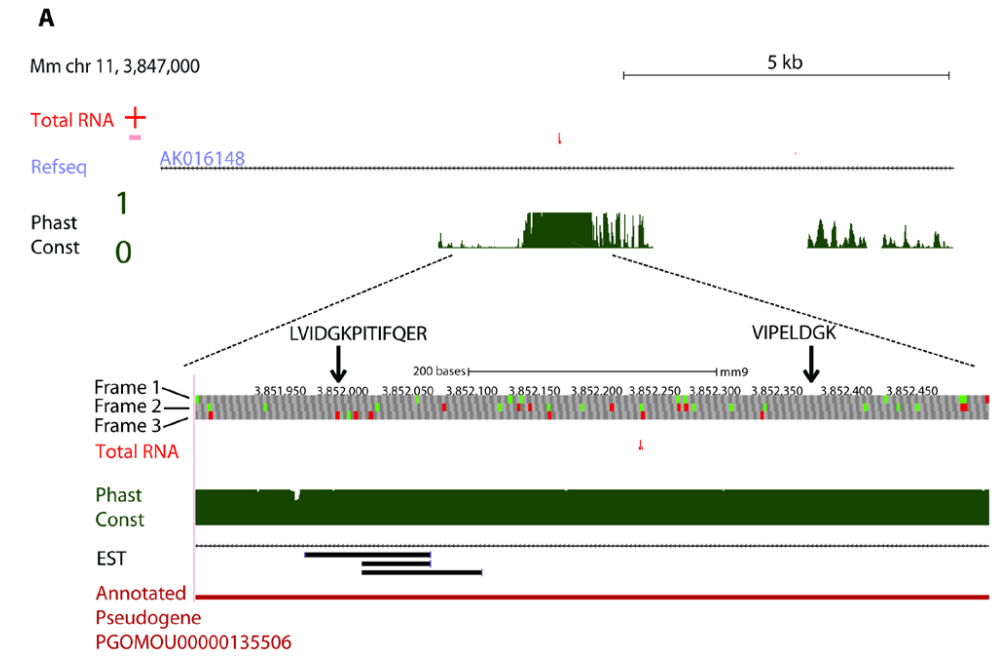
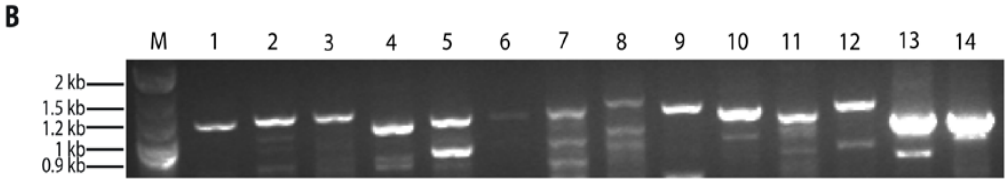


Figure 6. Classification of novel peptides based on their genomic location. The top panel depicts the two DNA strands of a canonical protein-coding gene with exons (green boxes), introns (red lines) and the transcribed sequences before the first exon and after the last exon (gray lines, here referred to as 5'UTR and 3'UTR). The 250 novel peptides were divided into 5 groups based on their location relative to known genes. The peptide location is denoted here by blue boxes. The number of peptides in each group is indicated. In the last group (extragenic) the line is cut to indicate that there is a large distance (>10 kb) from the protein-coding gene. TSS is transcription start site.

Forty-nine of the novel peptides belonged to the set of peptides identified in the BLASTP search or were derived from 5' leader sequences or 3' end regions (Group 1). A second group comprised 31 peptides that mapped to annotated pseudogenes (Group 2). Ten of the peptides in this group were also identified in the BLASTP search as having computational evidence suggesting protein-coding potential. This result is consistent with previous studies showing that pseudogenes can be translated [21] (**Figure 7A**). A third group, (Group 3) included 79 peptides that mapped onto introns. These examples may reflect the existence of unannotated exons or intron inclusion events. One such example is the FARP1 locus discussed above. A particularly intriguing and unexpected group of 66 peptides (Group 4) mapped onto the reverse strand from (n = 63) or near (n = 3) known exons and introns in protein coding genes. Three of the peptides were found in the region upstream of the promoter, suggesting that they are derived from upstream anti-sense RNAs [8] in mouse cortical neurons. To validate the expression levels of a subset of identified 'hit' transcripts, total RNA was isolated, tested for its integrity and reverse transcribed (**Table 1)** and Figure 2. Sequence verification confirmed the presence of the targeted anti-sense transcripts (**Figure 7B**). One example is Flnb (**Figure 7C**), which shows evidence of a ~5 kb long anti-sense transcript based on our RNA-seq data as well as RNA-seq data in both the mouse liver and kidney [22]. One of the novel peptides anti-sense to Cox17 (IILMPSLPAR), located at position 38,351,276 on chromosome 16 overlapping but anti-sense to its second intron was validated by synthetic peptide validation as discussed above in Figure 5. All of the novel peptides groups described above were located within or in the vicinity of known protein-coding genes. Finally, a second unexpected set consisted of 25 peptides that mapped onto extragenic regions removed from any known protein-coding gene were found (Group 5). Three of these peptides were also found in the ribosomal foot-printing data [14].





1 Suvn (g1) - 1325 bp
2 Armc9 (g3) - 1396 bp
3 Grpbp10 (g4) - 1473 bp
4 Elmo2 (g4) - 1177 bp
5 E13006D01Rik (g5) - 1310 bp
6 Cacna23d (g4) - 1278 bp
7 Gprk5 (g4) - 1394 bp
8 Nars2 (g4) - 1549 bp
9 Cox17 (g4) - 1367 bp
10 Mlstd2 (g4) - 1243 bp
11 Fars2 (g4) - 1155 bp
12 St6galnac3 (g4) - 1295 bp
13 PGOMOU00000135506 (g2) - 1002 bp
14 PGOMOU00000135766 (g2) - 1008 bp

C

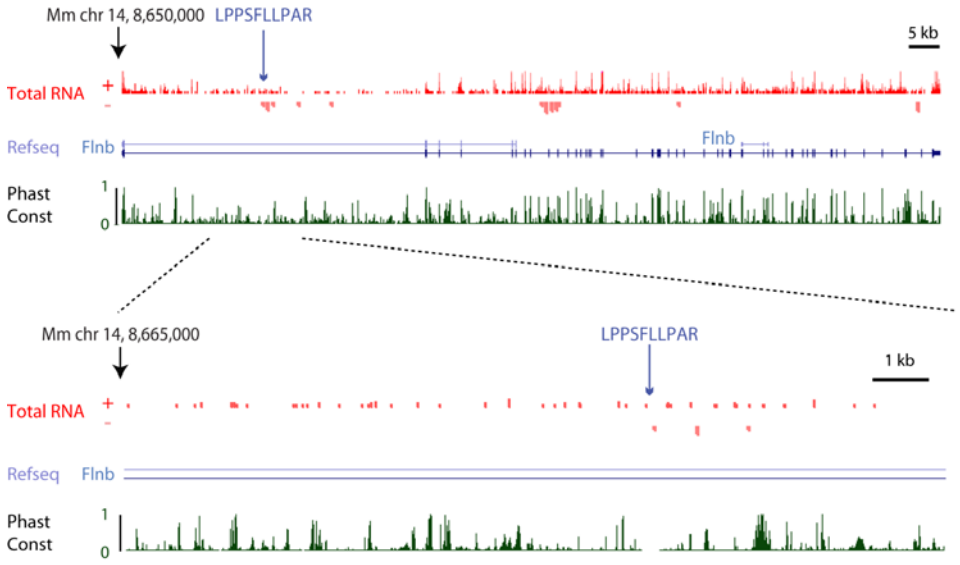


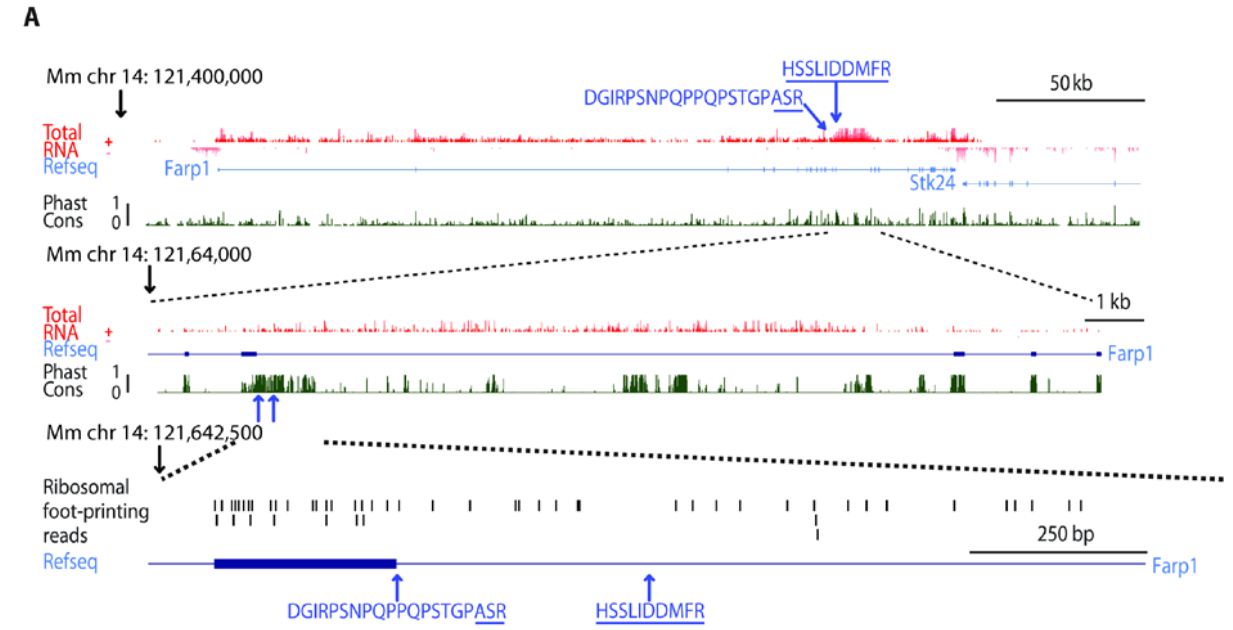
Figure 7: (A) Validation of novel peptides in extragenic regions. RT-PCR for 14 different anti-sense loci. Sequence verification confirmed the presence of the targeted anti-sense transcripts. **(B) Spectral validation data for a pseudogene.** Screenshot showing two peptides from the pseudogene PGOMOU00000135506 on chromosome 11 (number 13 in RTPCR gel Figure 7A). Experimental and synthetic spectra validation for the first peptide LVIDGKPITIFQER is shown in Figure 5. **(C) Screenshot from the UCSC genome browser of the Flnb locus from mouse.**

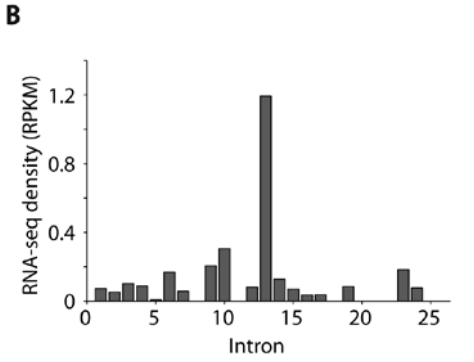
Molecular characterization of FARP1. Further in-depth molecular characterization of non-canonical translation was carried out for the FARP1 protein. Two unique novel peptides DGIRPSNPQPPQPSTGPASR and HSSLIDDMR, which were in frame with the rest of the protein, were mapped to the FARP1 intron 13 (**Figure 8A**). The RNA-seq read density across this intron is ~10-fold higher than that of the other introns and the RNA levels for this intron are comparable to the levels observed for the exons (**Figure 8B**). In addition, RT-PCR from three regions of this intron showed good expression in cortical cells from same experimental set up (**Figure 8C**). Closer inspection of the FARP1 intron 13 sequence reveals that the first 983 bps do not contain a stop codon within the reading frame used by the preceding exon. Thus, we hypothesized that there could be a novel isoform present with ~330 additional amino acids.

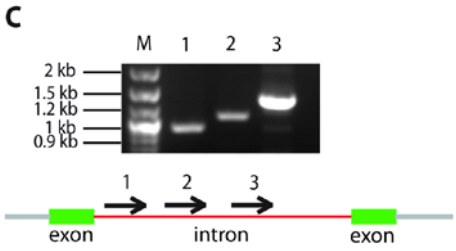
Western blot (WB) validation was performed using FARP1 (F8VPU2) primary antibody in brain lysates. FARP1 is FERM, Rho-GEF, pleckstrin domain containing protein with several different splice variants and varying molecular weights ranging from 51 kDa to 420 kDa. To validate that

FARP1 isoforms are present in the protein bands corresponding to the WB, an immuno-precipitation (IP) from mouse brain tissue using the FARP1 (F8VPU2) primary antibody followed by MS analysis was performed. Briefly, when the IP of FARP1 is separated by size on an SDS PAGE gel and stained using commassie blue, protein bands are observed at molecular weight ranging from 51 kDa to > 420 kDa as observed from WB results of the whole cell lysate (**Figure 8D**). These bands were excised and analyzed by LC-MS/MS. The resulting data was searched against the Uniprot mouse proteome database that was appended with the FARP1 intron sequence that was identified as a translation product in the global proteomics experiment. The novel peptides from the presumed intron were identified in three of the bands as indicated. Thus, this extensive validation using IP and MS confirms that the bands observed in the WB do indeed contain various splice variants of FARP1, some of which include the novel intron.

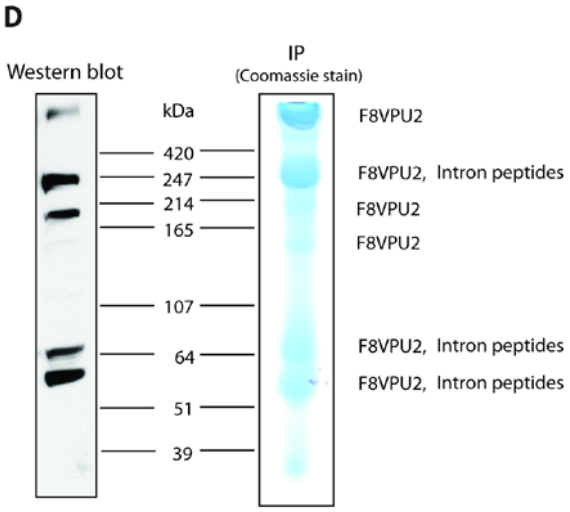
Furthermore, the fragmentation pattern of one of the novel peptides (HSSLIDDMFR) identified for FARP1, in Figure 5, matches that of its synthetic peptide. Taken together, the transcriptional and translational evidence suggest that this novel intron region is part of a larger unannotated variant of FARP1. The WB validation lends support to the genome-wide strategy for discovering novel non-canonical translation events. To predict the functional consequences of this large novel intron in the FARP1 protein, a disorder prediction analysis (http://iupred.enzim.hu) and domain mapping (SMART) was carried out. These analyses predicted a highly disordered region for this novel intron and domain mapping using ELM (http://elm.eu.org) indicated numerous regulatory regions in this novel intron sequence when compared to the whole sequence. **Figure 8E** illustrates the quantitative estimation of a subset of regulatory regions which is enriched in the novel intron as compared to the entire protein sequence suggesting that this region is possibly a hotspot for regulation by post-translational modifications. Additionally, ribosomal foot-printing data and RNA-seq data from HeLa cells provide further support for the expression of this FARP1 intron (**Figure 8F**).







1. 918 bp from genome position121642849
2. 1072 bp from genome position121643866
3. 1318 bp from genome position121645156



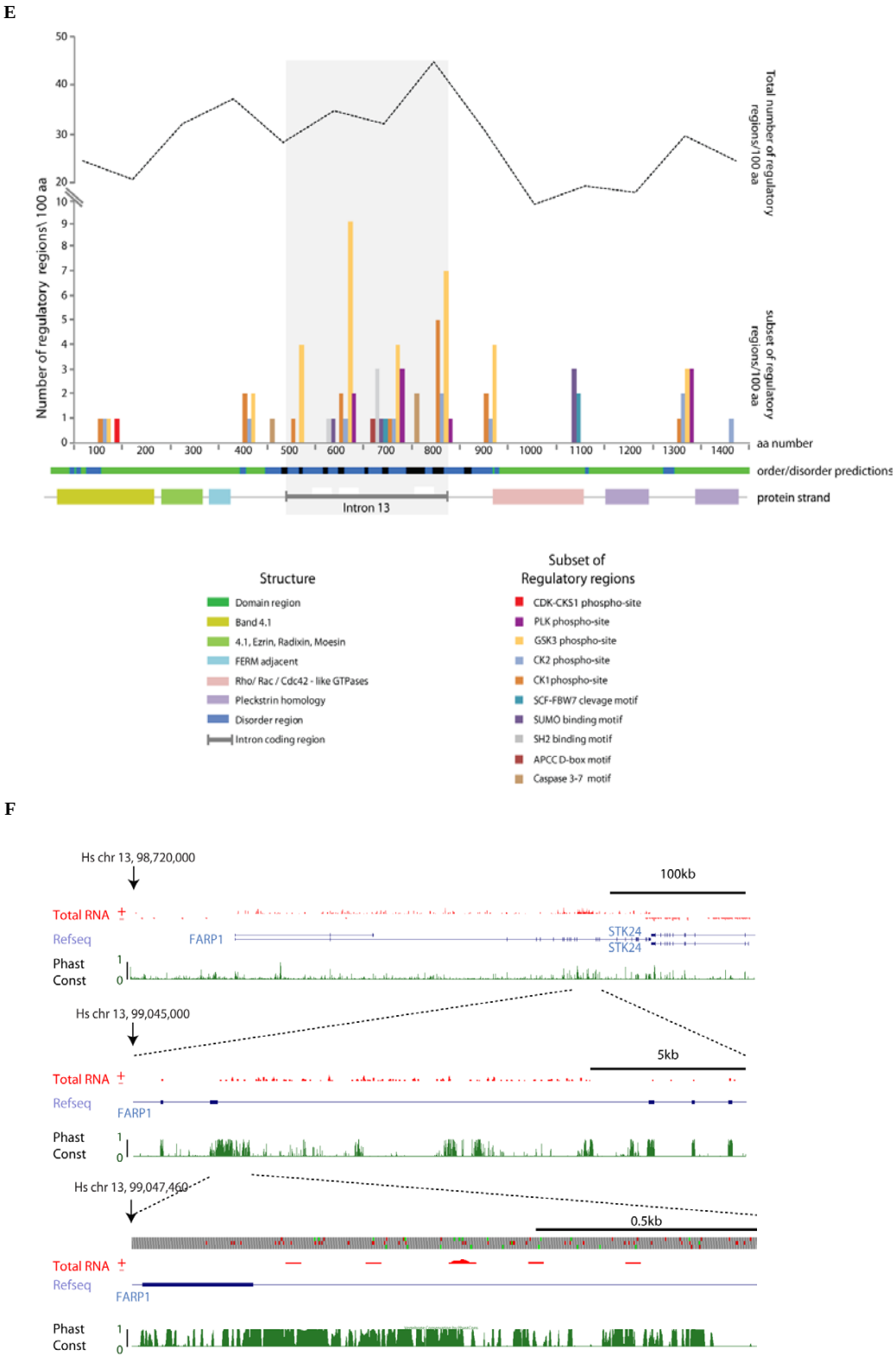


Figure 8. Identification and validation of two novel peptides at the FARP1 locus. (A) Screenshot from the UCSC genome browser showing total RNA, ribosomal footprints, Refseq gene structure and degree of conservation (Phast Cons). The bottom part is a zoomed in version spanning intron 13. The arrows indicate the position of the two

novel peptides. The first peptide (DGIRPSNPQPPQPSTGPASR) spans an exon-intron boundary; the underlined amino acids are located within the intron. The figure also indicates peptides found from ribosomal foot-printing studies **(B)** Average read density (RPKM) for all FARP1 introns. **(C)** RT-PCR gel showing RNA expression of three regions within FARP1 intron from depolarized cortical cells. **(D)** Western blot analysis of mouse brain lysate showing FARP1 protein bands of different molecular weights. Coommassie stained SDS-PAGE gel showing IP of FARP1 from mouse brain lysate. The bands were analyzed using LC-MS/MS and confirmed as FARP1 (F8VPU2) protein along with peptides from intron. **(E)** A model showing the 'domains', 'disordered regions' and 'regulatory regions' prediction in FARP1 amino acid sequence including the novel coding region of intron 13 by ELM (http://elm.eu.org/). The total number of regulatory regions per 100 amino acids is shown by line graph, out of which the distribution of a subset of functionally important regions per 100 amino acids are indicated by color bars. The number of regulatory regions is enriched in Intron 13 coding region **(F)** A screenshot from the UCSC genome browser of the FARP1 locus from human with RNA-seq data from HeLa cells. Intron 13 (magnified) is similar to mouse with an elevated level of expression. The top track in the bottom panel indicates stop codons (red) and start codons (green). The third frame has an amino acid sequence which is consistent with the longer exon.

Quantitative analysis of the novel peptides. Having determined the genomic context of the novel peptides, we set out to determine whether they were physiologically regulated in response to enhanced neuronal activity. Neuronal activation induces a gene expression program involving several hundred genes as well as non-coding RNAs [5].

In the experimental setup in Figure 1, mouse cortical neurons were stimulated using KCl and TMT labeling were used to quantify the relative abundance of each peptide at different time-points following KCl stimulation. For each TMT channel the same amount of protein, 100 µg, was used and labeling bias was tested by assessing the median of $\log_2$ intensities from all the channels (Figure 1). The medians across the channels had a standard deviation of 0.16 and a coefficient of variance of 0.02. This suggests that there is no inherent mixing irregularities in the total pool of labeled sample and the differential peptide abundance observed among time-points are true observations and hence could be biologically relevant.

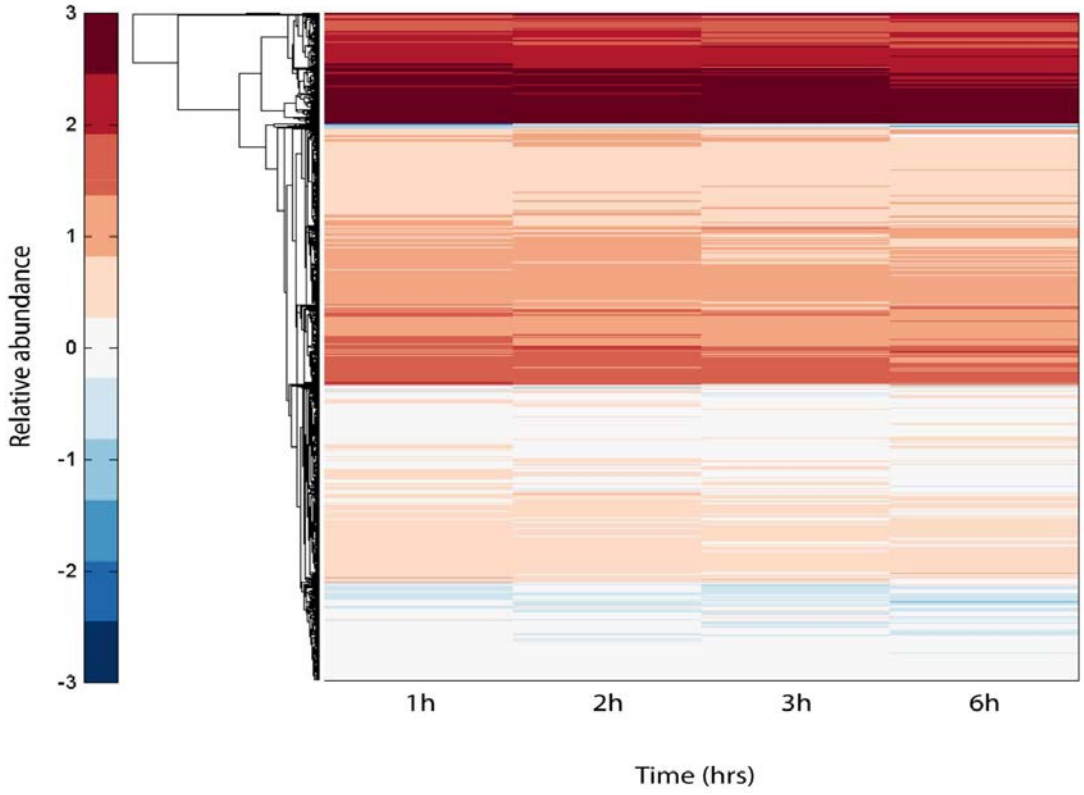


Figure 9. Quantitative regulation of known proteins. Hierarchical clustering of the intensity temporal profiles for known proteins (3284). Each row represents one protein. The four columns indicate abundance at 1 h, 2 h, 3 h and 6 h post KCl stimulation. The colors represent up- or down-regulation with respect to the 0 h time-point (see color map on left, arbitrary units). The clustering reveals several distinct regulatory patterns present both for known.

In addition, TMT labeling efficiency was evaluated to be 99.5 % of all unique and high-confidence peptides of which 98.3 % are fully labeled (labeled on N-terminal and internal Lysine residues) with TMT reporter ions indicating that there is no significant loss of peptides that would affect quantitation. All the above validations highlight our careful and comprehensive assessment of the quantitative changes of the novel peptides and hence indicate that the abundances of the known proteins (**Figure 9**) and novel peptides (**Figure 10)** are modulated by KCl depolarization.

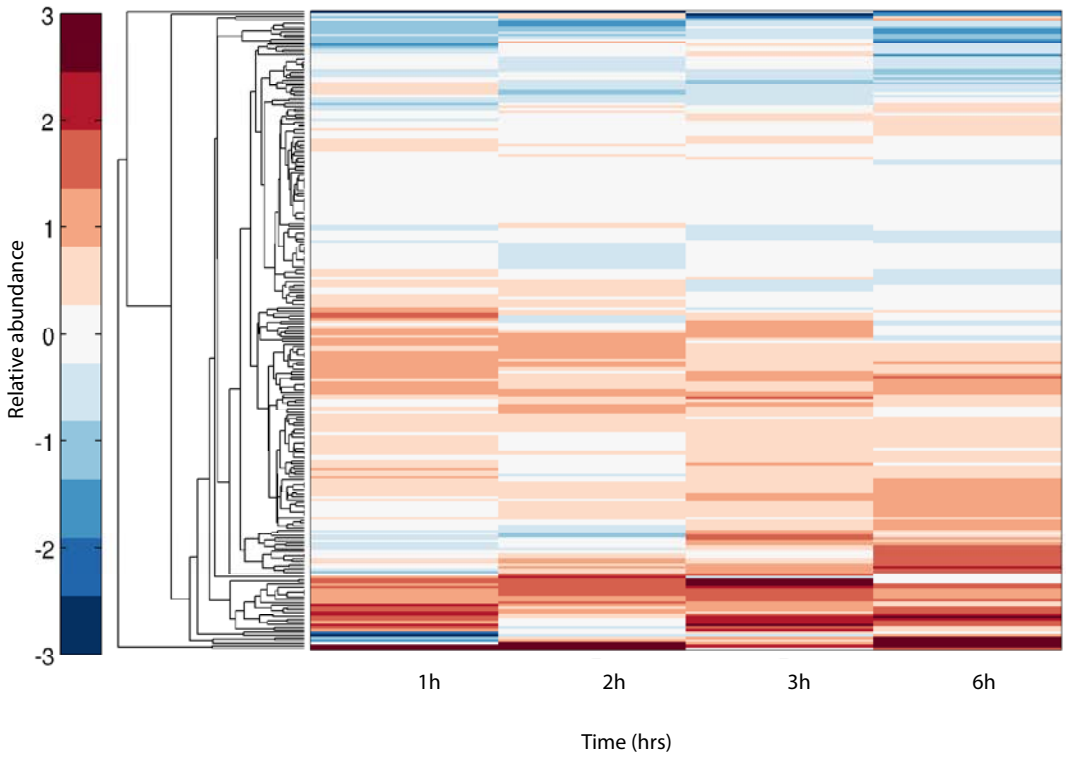


Figure 10. Quantitative regulation of novel peptides. Hierarchical clustering of the relative abundance temporal profiles for novel peptides (250 peptides). Each row represents one peptide. The four columns indicate abundance at 1 h, 2 h, 3 h and 6 h post KCl stimulation. The colors represent up- or down-regulation with respect to the 0 h time-point (see color map on left, a.u.). The clustering reveals several distinct regulatory patterns present both for known and novel peptides. These regulatory patterns are similar to those shown for known proteins.

To identify patterns of co-regulation between the known proteins and novel peptides, the known protein abundance changes along with the novel peptide abundance changes were clustered together in an unbiased way using GProX (version 1.1.12) software. Results of the clustering analysis indicate that eight clusters identify discerning patterns of co-regulation of known proteins and novel peptides (**Figure 11**). The number of novel peptides in each of the clusters is listed in Figure 11. Extrapolating from this data and from our own previous analyses [23, 24], it can be predicted that these novel peptides are co-regulated in similar manner as the known proteins in each of the clusters. Some of these known proteins include CamKIId, Gria1 and Shank2 that are known to be temporally regulated with KCl stimulation in neurons.

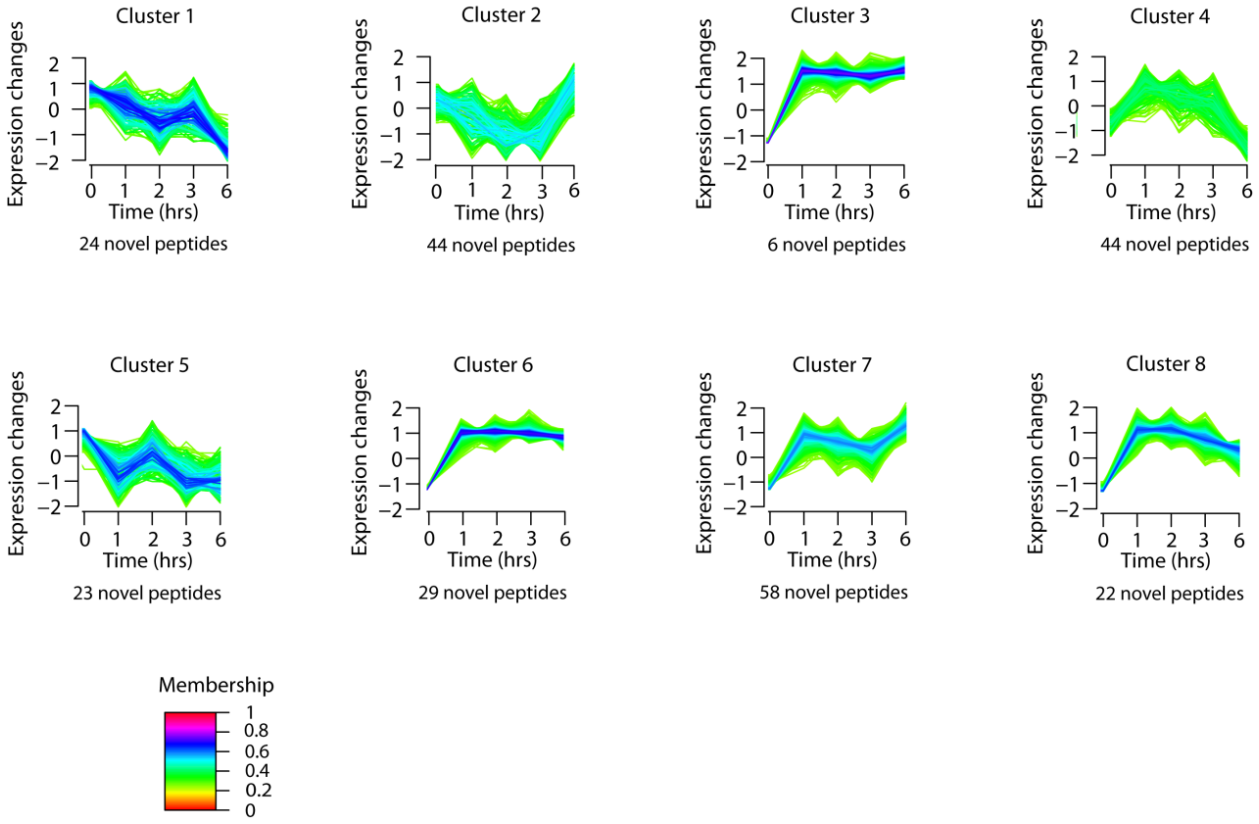


Figure 11. Clusters of both known proteins and novel peptides showing similar temporal regulation post KCl stimulation. The numbers of novel peptides in each cluster is indicated. The membership scale reflects how well the regulation of a protein matches the consensus profile.

DISCUSSION

Using a custom processed transcriptome database from the same experimental context as reported here for the unbiased examination of the proteome, is a robust means to identify the existence of peptides derived from regions classified as non-coding throughout the genome. Although it is possible that some of these non-canonical peptides are the result of biological noise, this interpretation would require “errors” at multiple levels including: transcriptional sloppiness by RNA polymerase, followed by the stable survival of non-polyA transcripts, export of these species into the cytoplasm, recognition by the ribosomal machinery and translation. The consistency of the findings, which are derived from pooling large numbers of cells, together with the validation, cast doubt on an interpretation purely based on biological noise. Instead, the

discovery of non-canonical translation products and the validation of 250 novel peptides is consistent with several different lines of evidence, including ribosomal profiling studies [14], characterization of short open reading frames [21, 25-27] and studies examining translation of pseudogenes [28], which have pointed to a more extensive proteome than what is described by current databases. Most MS analyses, including the one presented here, are not saturating and the number of detected protein products is much lower than the number of mRNAs detected by RNA-seq. Our RNA-seq data identified ~12,000 transcripts corresponding to canonical protein sequences, the proteomics data gave evidence for ~25 % of the canonical transcripts. Extrapolating from this observation, it can be estimated that at least 4 times as many novel peptides are yet to be discovered under the current experimental conditions. Furthermore, it can be speculated that examining other tissues and biological conditions will uncover an even richer set of non-canonical translational events.

The biological function, if any, of the novel peptides described here is not known. Several studies have suggested that evolutionarily conserved regions are more likely to be functionally important. Sequence conservation for the 250 novel peptides was investigated in 28 vertebrate species using the MultiZ alignment [29]. The results show that there are 30 peptides with significant evidence of purifying selection (**Table 2**) and that the translated sequences found in many species are more likely to show evidence of purifying selection consistent with studies of proto-genes in yeast [30]. The observation that peptides derived from non-canonical translational events show dynamic regulatory patterns under physiological conditions similar to those characterized for well-established proteins further suggests the possibility that these peptides may also play functional roles. Future examination of other tissues and conditions combined with functional studies will shed light on the biological roles of the uncharted proteome.

ACKNOWLEDGEMENTS

The author acknowledges the contribution of Dr. Sudhakaran Prabakaran for the mass spectrometry data analysis, Dr. Martin Hemberg for RNA-seq data analysis.

ABBREVIATIONS

AS	antisense
ACN	Acetonitrile
FA	Formic acid
dNTPs	Deoxynucleotide triphosphates
IP	Immunoprecipitation
MS	Mass spectrometry
ncRNAs	Non coding RNAs
PQD	Pulsed Q collision-induced dissociation
RNA-seq	RNA sequencing
RT-PCR	Reverse Transcription- Polymerase chain reaction
TFA	Trifluoroacetic acid
TMT	Tandem mass tagging
WB	Western blotting

REFERENCES

1. Djebali, S., et al., Landscape of transcription in human cells. Nature, 2012. **489**(7414): p. 101-8.
2. Jacquier, A., The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet, 2009. **10**(12): p. 833-44.
3. Kapranov, P., et al., Large-scale transcriptional activity in chromosomes 21 and 22. Science, 2002. **296**(5569): p. 916-9.
4. Liu, Q. and Z. Paroo, Biochemical principles of small RNA pathways. Annu Rev Biochem, 2010. **79**: p. 295-319.
5. Kim, T.K., et al., Widespread transcription at neuronal activity-regulated enhancers. Nature, 2010. **465**(7295): p. 182-7.
6. Wright, M.W. and E.A. Bruford, Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. Hum Genomics, 2011. **5**(2): p. 90-8.
7. Seila, A.C., et al., Divergent transcription from active promoters. Science, 2008. **322**(5909): p. 1849-51.
8. Werner, A., M. Carlile, and D. Swan, What do natural antisense transcripts regulate? RNA Biol, 2009. **6**(1): p. 43-8.
9. Wu, X. and P.A. Sharp, Divergent transcription: a driving force for new gene origination? Cell, 2013. **155**(5): p. 990-6.
10. Ponting, C.P., P.L. Oliver, and W. Reik, Evolution and functions of long noncoding RNAs. Cell, 2009. **136**(4): p. 629-41.
11. Brent, M.R., Genome annotation past, present, and future: how to define an ORF at each locus. Genome Res, 2005. **15**(12): p. 1777-86.
12. Dinger, M.E., et al., Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput Biol, 2008. **4**(11): p. e1000176.
13. Koenig, T., et al., Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. J Proteome Res, 2008. **7**(9): p. 3708-17.
14. Ingolia, N.T., L.F. Lareau, and J.S. Weissman, Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell, 2011. **147**(4): p. 789-802.
15. Slavoff, S.A., et al., Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol, 2013. **9**(1): p. 59-64.
16. Winter, D. and H. Steen, Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. Proteomics, 2011. **11**(24): p. 4726-30.
17. Bantscheff, M., et al., Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.
18. Hemberg, M., et al., Integrated genome analysis suggests that most conserved non-coding sequences are regulatory factor binding sites. Nucleic Acids Res, 2012. **40**(16): p. 7858-69.
19. Menschaert, G., et al., Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol Cell Proteomics, 2013. **12**(7): p. 1780-90.
20. Serang, O., et al., A non-parametric cutout index for robust evaluation of identified proteins. Mol Cell Proteomics, 2013. **12**(3): p. 807-12.

21. Bazzini, A.A., et al., Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J, 2014. **33**(9): p. 981-93.
22. Shen, Y., et al., A map of the cis-regulatory sequences in the mouse genome. Nature, 2012. **488**(7409): p. 116-20.
23. Kirchner, M., et al., Computational protein profile similarity screening for quantitative mass spectrometry experiments. Bioinformatics, 2010. **26**(1): p. 77-83.
24. Singh, S.A., et al., Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J, 2014. **33**(4): p. 385-99.
25. Magny, E.G., et al., Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. Science, 2013. **341**(6150): p. 1116-20.
26. Oyama, M., et al., Diversity of translation start sites may define increased complexity of the human short ORFeome. Mol Cell Proteomics, 2007. **6**(6): p. 1000-6.
27. Pauli, A., et al., Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science, 2014. **343**(6172): p. 1248636.
28. Brosch, M., et al., Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. Genome Res, 2011. **21**(5): p. 756-67.
29. Blanchette, M., et al., Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res, 2004. **14**(4): p. 708-15.
30. Carvunis, A.R., et al., Proto-genes and de novo gene birth. Nature, 2012. **487**(7407): p. 370-4.

Chapter 3

Co-regulation Proteomics Reveals Substrates and Mechanisms of APC/C complex dependent degradation

Contents **Page No.**

SUMMARY

We analyzed cell cycle dependent changes of the human proteome to identify putative Anaphase Promoting Complex / Cyclosome (APC/C) substrates during different phases of synchronized cell division. Using multiplexed quantitative proteomics, we identified >4400 proteins, each with a 6 time point abundance profile across the cell cycle. Hypothesizing, that proteins with similar abundance profiles are co-regulated, we clustered the proteins with abundance profiles most similar to known APC/C substrates that resulted in a shortlist enriched in kinases and kinesins. Biochemical studies on the kinesins confirmed KIFC1, KIF18A, KIF2C and KIF4A as APC/C^{CDH1} substrates. The targeted phospho-mutant assays showed that the KIFC1 degradation is modulated by a stabilizing CDK1 dependent phosphorylation site within the D-box motif of KIFC1, a mechanism also known in canonical APC/C substrates such as SECURIN. The regulation of KIFC1 degradation provides insights into the mechanism of substrate degradation by the APC/C during mitosis.

INTRODUCTION

The mammalian cell cycle is controlled by the distinct quantitative oscillations of a subset of proteins that are ubiquitinated by the E3 ligase Anaphase-Promoting Complex/Cyclosome (APC/C) and subsequently degraded by the proteasome [1-6]. Over the last two decades, the study of APC/C and its targets resulted in the identification of a set of proteins whose APC/C-mediated turnover is critical for ordered progression through mitosis and the cell cycle in general. Strategies to identify and validate novel APC/C targets have always been a matter of great interest particularly due to the strong link between cell cycle regulation and human diseases, most notably cancer and recently neurological disorders. Many of the proteins that regulate chromosome segregation and cytokinesis have been recognized as cancer-promoting factors, as well as therapeutic targets [7], include protein kinases as well as motor proteins from the kinesin family [8]. To date, systematic screens for APC/C targets typically rely on laborious *in vitro* based methods [9-11]. Recent high-throughput methods have greatly simplified the search for APC/C targets [12-14]; however, the *in vitro* based screen designs for ubiquitination without the contexts of degradation are not readily confirmed.

Here we exploit the versatility of multiplexed quantitative proteomic strategies to develop an efficient *in vivo* based strategy for the identification of novel APC/C substrates and also to dissect the mechanisms by which the degradation of these APC/C substrates are regulated. We conducted a time course experiment to identify substrates degraded by the APC/C during cell cycle. The basis for the identification of novel APC/C substrates by co-regulation proteomics resembles the approach taken in many transcriptomics studies in which transcripts with similar quantitative profiles are identified to find co-regulated genes [15-17]. The premise is that genes involved in related pathways are co-regulated at the transcriptional level, such that novel transcripts whose profiles are similar to those of known function are likely to function similarly. These genes are then followed up in-depth using *in vitro* assays to validate their degradation and ubiquitination mechanisms.

We employed a 6-plex tandem mass tagging (TMT) based quantitative proteomic strategy [18] that provides relative endogenous protein abundances. Synchronized cells were sampled before, during and post-mitosis, where the APC/C displays activity specifically from prometaphase to late G1 phase [19-23] (**Figure 1A**). Many of the APC/C substrates such as SECURIN and CYCLIN B display a distinct protein abundance profile during the cell cycle which has become a hallmark for APC-dependent regulation (**Figure 1A**). We hypothesized that all proteins ubiquitinated by the APC/C and then degraded by the proteasome in a cell cycle dependent manner, would show similar abundance profiles over the course of the cell cycle. Thus, in order to identify novel substrates of the APC, we determined reference protein abundance profiles based on known APC/C substrates and compared them to the abundance profiles of all the other proteins we detected. This approach generated a shortlist of putative APC/C substrates, some of which were confirmed as APC/C substrates using the conventional *in vitro* degradation assays. Selecting the newly identified APC/C substrate KIFC1 for detailed follow-up studies, we used biochemistry of D-box mutant and the FLEXIQuant methodology [24], to profile the dynamics of the post-translational modifications (PTMs) of KIFC1. These experiments revealed insights into the interplay between phosphorylation and APC-dependent degradation, pointing towards a more prevalent, though underappreciated mechanism for the degradation of APC/C substrates.

EXPERIMENTAL PROCEDURES

Cell culture

Reagents were purchased from Sigma-Aldrich unless otherwise noted. HeLa S3 cells (ATCC) were maintained in DMEM supplemented with 10 % fetal bovine serum, penicillin and streptomycin (Cellgro). For the mitosis TMT experiment, HeLa S3 cells were grown in 15 cm culture dishes until confluent, split in a 1 to 6 ratio and grown for 12 h. Cells were then incubated with 2 mM thymidine for 20-22 h, washed and released into fresh media for 8 h, incubated again with 2 mM thymidine for 20-22 h and washed and released for up to 15 h. The time points sampled for the TMT labeling included, 0 h/S-phase (immediately before the second thymidine release) and 7, 8, 9, 10 and 15 h post-thymidine release.

For *in vitro* degradation and ubiquitination assays, HeLa S3 cells were grown in suspension until the population reached a density of 2.5 - 3 x 10^5 cells/ml. Cells were then incubated with 2 mM thymidine for 20-22 h, washed and released into fresh media for 3 h and blocked again with 0.1 µg/ml nocodazole for 11 h. Cells were washed twice with fresh warm media, released for four h and then harvested to obtain a G1 population. Synchronization for all experiments was validated using Flow Cytometry (5-laser LSRII, BD Biosciences) after propidium iodide staining of ethanol-fixed cells.

TMT labeling and peptide fractionation

Three biological replicates were performed for each cell cycle experiment. Cells were lysed and the proteins extracted as outlined previously [25], digested using trypsin (Promega) and labeled with the TMT 6-plex reagent (Thermo Scientific), combined and desalted using Oasis Hlb 1cc columns [26]. The desalted peptide samples were fractionated based on their isoelectric point in a range of pI 3 to 10 in 24 fractions using an OFFGEL fractionator (Agilent). The fractions were extracted, dried with a table top speed vacuum (Thermo Scientific), resuspended in 20µl 5 % acetonitrile (ACN) 5 % formic acid (FA) and analyzed by LC-MS/MS in triplicate (for details see below), resulting in nine replicates for each time point when considering all three biological repeats.

Mass spectrometry

Peptide samples were analyzed with the high resolution/accuracy LTQ-Orbitrap mass spectrometer (Thermo Scientific Classic model). The reverse phase columns were packed in-house using Magic C18 particles (5 µm, 200 Å; Michrom) and PicoTip Emitters (New Objective). Peptides were eluted with a 60 minute linear gradient from 95 % A (water with 0.2 % FA) 5 % B (ACN with 0.2 % FA) to 65 % A 35 % B for TMT samples. The data were acquired in the data dependent mode, fragmenting the 6 most abundant peptide species. For the TMT analyses, peptides were fragmented in the PQD-mode [27].

The proprietary Thermo Scientific .raw files were converted into mgf files [28] and MS/MS data queried either against the IPI Human v.3.69 protein sequence database, containing common contaminations and concatenated to its decoy version using MASCOT version 2.1 (Matrix Science). TMT peptides were searched with enzyme specificity trypsin, one missed cleavage site, propionamide (Cys) as a fixed modification; and TMT6plex (N-termini and Lys), oxidation [29], deamidation (N) and Gln to pyroGlu (N-terminal Q) as variable modifications. Using the target-decoy approach, peptides were filtered at a false discovery rate of 1 % and proteins assigned based on at least two unique peptide sequences using in house software.

Quantitative MS analysis and Protein Profile Similarity Screening

Pre-clustering steps – Peptide-spectrum matched TMT ion intensities were extracted from the fragment ion spectra without applying MS1 or MS/MS-based cut-offs. Prior to further data analysis we corrected for loading variation. Next, each TMT ion channel intensity was divided by the sum of all sister channel intensities (sum normalization), which resulted in the peptide abundance profiles for the clustering, protein profiling and co-regulation analysis.

Clustering steps - The profiles of the APC/C substrates identified in our screen were extracted and co-clustered. This procedure yielded our six reference groups (**Figure 1B, 2A**). We subsequently determined a representative average protein trace for each reference group and determined the similarity of each observed protein trace to these reference protein traces. Proteins were then ranked according to decreasing similarity for every reference group. The

profiles of 45 proteins most similar to each of the 6 different reference profiles (top percentile, i.e. 270 protein descriptions in total) were chosen to select for more frequent protein hits.

Plasmids

Human clones of KIFC1, KIF2C, KIF18A, KIF14, KIF20A, KIF22, KIF23 and KLC1 were purchased from Open Biosystems (OB), the KIF4A clone was a gift from the T. Mitchison Lab. The genes were subcloned into the CellFree Sciences (CFS) vector, pEU-E01-MCS, for *in vitro* degradation assays. The various D-box and serine mutations were generated using QuikChange site-directed mutagenesis strategy (Stratagene).

***In vitro* degradation and ubiquitination assays**

G1 population of HeLa S3 cells was synchronized and validated using Flow Cytometry (5-laser LSRII, BD Biosciences) after propidium iodide staining of ethanol-fixed cells. Lysates were tested for activity using SECURIN substrates. Substrates were subcloned and *in vitro* transcribed and expressed in the presence of ^{35}S-labeled methionine. The APC/C/ inhibitors EMI1 protein and SECURIN protein were recombinantly expressed, His-tag purified and tested for activity. The APC/C activator CDH1 was *in vitro* transcribed and expressed using wheat germ extracts.

Degradation assays - Extracts of HeLa S3 cells were prepared as described previously (Ayad et al, 2003). Degradation assays were done in 20 µl cell extracts supplemented with 1 µl of energy regenerating mix (20 mM ATP, 150 mM Creatine Phosphate, 2 mM EGTA pH 8.0, 20 mM $MgCl^2$, 0.1 µg/ml Ubiquitin (Boston Biochem), 0.1 µg/ml Cyclohexamide), 1 µl of 1 mg/ml recombinant 6xHis-tag UBE2S (Boston Biochem), 1 µl of 1 mg/ml recombinant UBCH10 (Boston Biochem), 5 µl of 6xHis-tag purified recombinant of either 0.5 mg/ml EMI1 protein or 0.5 mg/ml SECURIN protein or PBS for no inhibitor control and 1-2 µl radiolabelled *in vitro* synthesized substrate expressed in reticulocytes (Promega) with ^{35}S-labeled methionine.

Ubiquitination assays were done by pulling down the APC/C, maintaining the complex on the beads and supplementing the APC/C with recombinant CDH1 [9] and reagents described above. Samples were incubated at 30 ºC and aliquots were sampled every 15 to 30 minutes over a two h period. Aliquots were added to Laemmli sample buffer (Bio-Rad), incubated at 95 ºC for five

min and flash-frozen on dry ice. Samples were resolved by SDS–PAGE and visualized by autoradiography.

RESULTS

We generated synchronous cell populations at 6 different points along the cell cycle using double thymidine block, as this blocking agent shows less pleiotropic effects than other commonly applied nocodazole/taxol-based synchronization protocols, which activate the mitotic checkpoints [30]. After lysis of cell pellets from each time point, proteins were trypsinized in solution. The resulting peptides from each time point were subsequently labeled with one of six TMT labels and combined into one sample (**Figure 1A**).

The peptide mixtures were fractionated into 24 fractions by isoelectric focusing prior to LC-MS/MS analysis. To minimize under-sampling of this complex sample thereby ensuring a robust dataset, three biological replicates were analyzed in triplicate by LC/MS resulting in 216 LC/MS/MS runs matching 264,386 MS/MS spectra identifying 44,085 unique peptides at a 1 % false discovery rate (FDR). For subsequent protein grouping, we required a 1 % FDR cut-off and a minimum of at least two unique peptides with complete TMT-reporter ion traces. Applying these criteria, this experiment identified and quantified 4,470 non-redundant proteins.

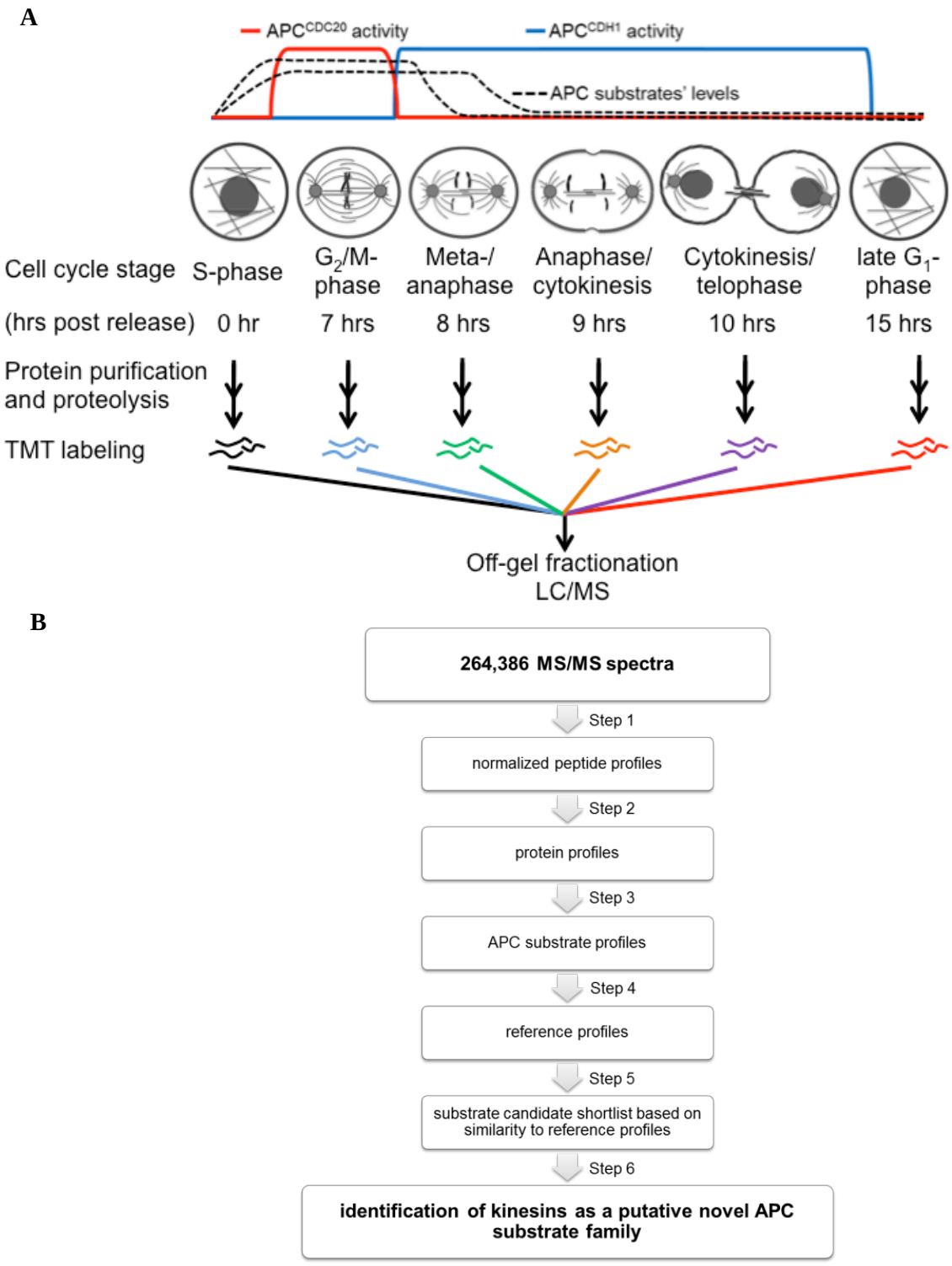


Figure 1. (A) Overview of cell cycle proteomics experiments of synchronized HeLa cells. Upper Panel: A schematic of APC/C activity with respect to prototypical substrate level profiles. APC/C can complex with either CDC20 or CDH1 as co-activator. Lower panel: Workflow of the cell synchronization experiment and the subsequent

TMT-based quantitative proteomics experiment. **(B) Protein clustering approach and results.** The protein profile similarity screening workflow leading to the identification of novel APC/C substrates.

In our data set, we identified 24 known APC/C substrates known to be degraded by the APC/C, while the abundance profile of GAPDH remained same across time-points, hence serving as negative control. In our dataset, 17 of 24 identified known APC/C substrates, showed the canonical abundance profile for APC/C substrates, i.e. an initial increase during G2-phase followed by marked decrease in abundance, reaching the lowest level in G1 phase. To identify co-regulated proteins in the quantitative proteomics dataset, a clustering analysis was performed using the previously reported strategy for screening protein profile similarities [31] to identify the proteins whose abundance profiles resemble those of known APC/C substrates (**Figure 1B**). We generated six APC/C reference clusters to reflect different classes of APC/C substrate with slightly different abundance profiles based on, for example, their respective apices and/or the steepness of their ascending and/or descending parts of their curves. To prioritize the best matching candidates, proteins found in the 1st percentile of each reference cluster were selected that included numerous kinases (**Figure 1B**). Interestingly, many of these protein kinases have been described as APC/C substrates.

In contrast to the expected high number of kinase hits, the 18-fold enrichment of kinesins in the 1st percentile (between 2 and 5 kinesins per reference cluster, **Figure 2A**) in comparison to the entire dataset (20/4470) was surprising. Only a single kinesin family member in higher eukaryotes, KIF22, had been described as an APC/C substrate previous to this study [32]; in addition, two kinesins (KIP1 and CIN8) had already been described as substrates in yeast [33, 34]. KIF22 was indeed within the 1st percentile along with other kinesins including KIF18A, KIF2C, KIF23, KIF14 and KIFC1. Recently, KIF18A was described as an APC/C substrate by Sedgwick et al., independently confirming our results from co-regulation proteomics based strategy [35].

Following the lead from the co-regulation analysis, we investigated the possibility that kinesins are indeed common APC/C substrates. To this end, we ranked the protein profiles of all identified kinesins in each reference cluster and subsequently calculated the median percentile for each kinesin (**Figure 2A**). To expand our short list of kinesins to be tested for APC/C-

dependent degradation, we accepted all kinesins within the 5th percentile as potential APC/C substrates, thereby adding KIF13A and KIF4A to the list of candidate substrates. The mean TMT-based protein profiles for the candidate kinesins are displayed in **Figure 2B**.

Of the eight kinesins selected by the screen KIF22, KIF14 and KIF13A were not tested for degradation as KIF22 was a known substrate [32] and KIF14 and KIF13A did not easily translate *in vitro* due to their molecular weights of ≥ 185 kDa. Although KIFC1 had been suggested to be regulated by ubiquitination during the cell cycle [36], this protein has not been studied further in the context of APC/C degradation and thus was included in the degradation assays.

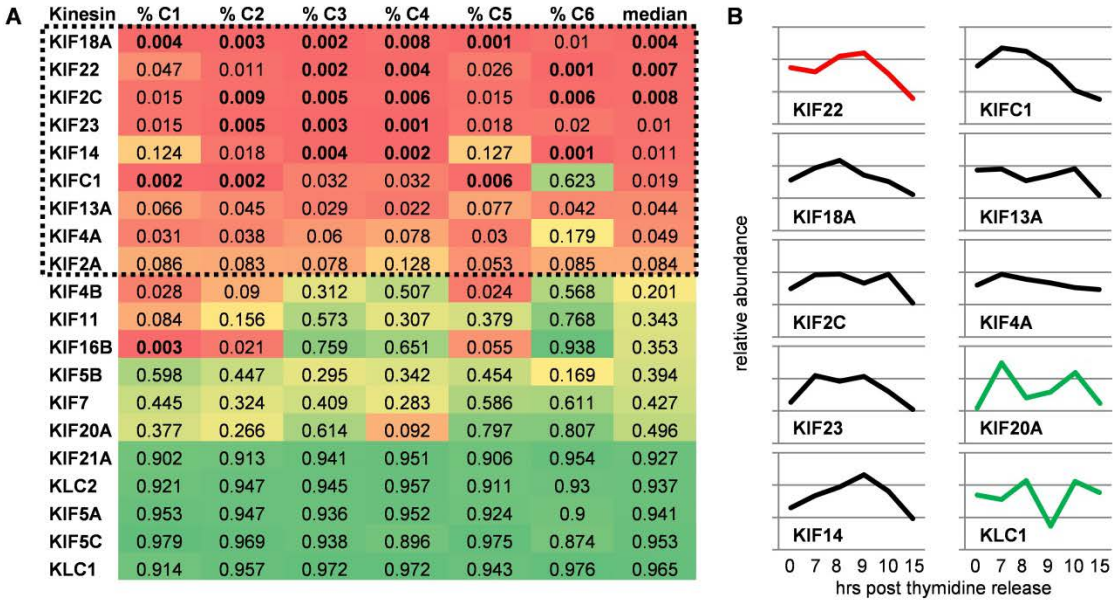


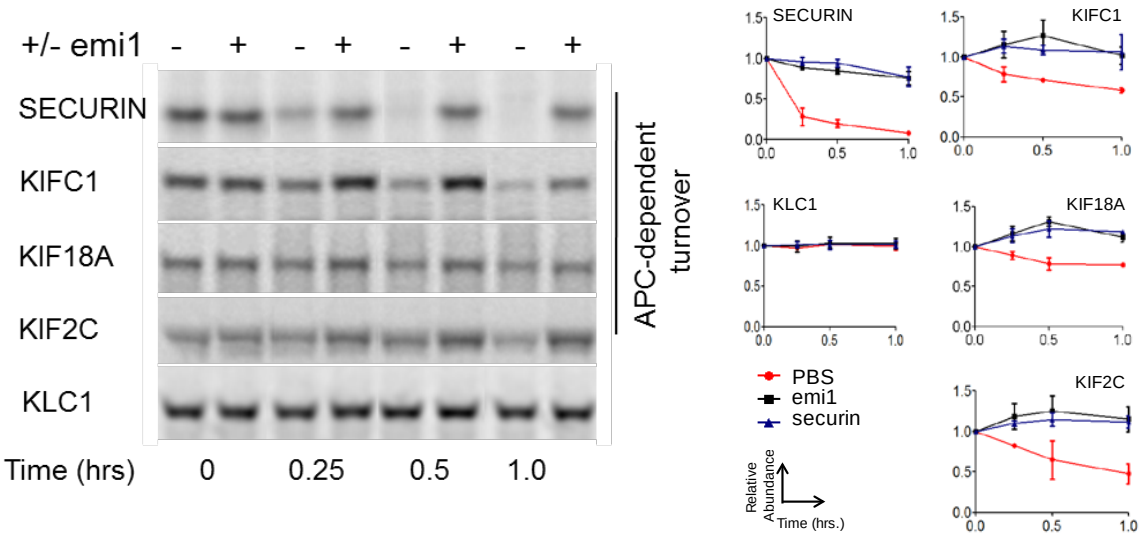
Figure 2. (**A**) **Co-regulation analysis reveals kinesins as putative novel APC/C targets.** Heat map to show the percentile (%) in which the kinesins are ranked in the six reference clusters (C1 to C6) based on abundance profiles of known APC/C substrates. Boxed kinesins were flagged as putative APC/C targets based on their median percentile. Bold values indicate ranks within the 1st percentile of each cluster. (**B**) Individual protein abundance profiles for kinesins flagged as candidate (black traces) and non-candidate (green) APC/C substrates. KIF22, a known APC/C substrate, is indicated in red. The y-axes indicate relative abundance ranging from 0 to 0.3.

We performed *in vitro* degradation assays using G1 HeLa cell extracts [5], which maintain high APC/C^{CDH1} activity, to test whether this E3 complex targets KIF18A, KIF2C, KIF23, KIFC1 and KIF4A proteins for degradation by the proteasome. We added two additional kinesins as negative controls: KIF20A at the 50th percentile and KLC1 at the 96th percentile. All seven kinesins were expressed in an *in vitro* transcription/translation system in the presence of ^{35}S-

methionine and incubated with HeLa S3 G1-phase extracts and the reaction conditions under which, SECURIN, a prototypic APC/C substrate, is readily degraded. In addition, two APCC/C inhibition reactions were also performed simultaneously to confirm the APC/C-specific degradation: 1) addition of the recombinantly expressed and His-tag purified endogenous EMI1 as a non-competitive inhibitor and 2) addition of an excess of recombinant expressed and His-tag purified non-radioactive labeled SECURIN as a competitive inhibitor **(Figure 3**) [10, 37]. If degradation of the substrate is observed in the presence of either inhibitor, it would indicate APC/C-independent degradation modulated by e.g. an alternative E3 ligase.

Our *in vitro* degradation assays showed a time dependent degradation for four of the five potential substrates tested: KIF18A, KIF2C, KIFC1 and KIF4A were readily degraded, however remained stable in the presence of either APC/C inhibitor (**Figure 3A**). In contrast, inconclusive results were obtained for KIF23 and degradation was not observed for the two negative controls KIF20A and KLC1 (**Figure 3B**).

A



B

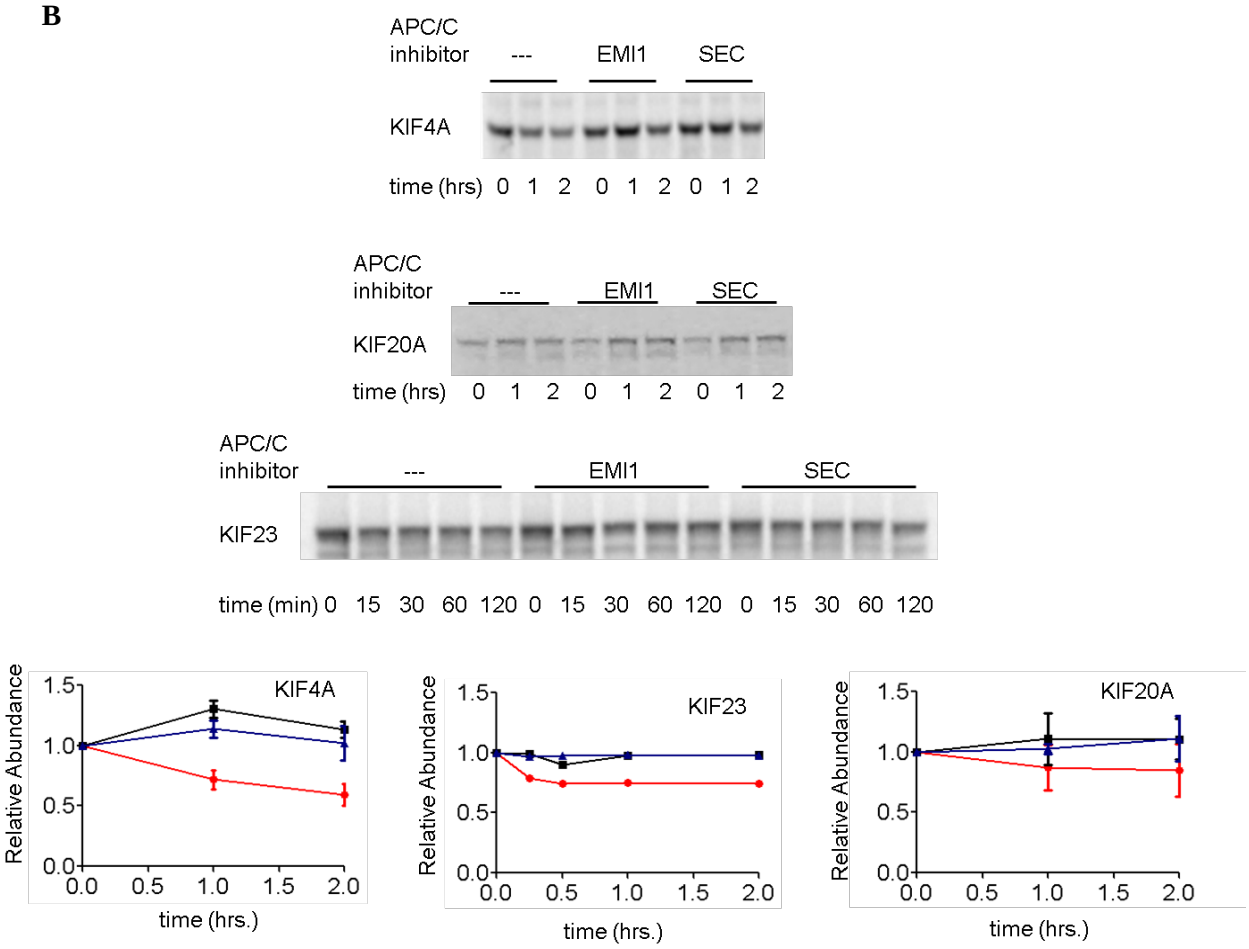


Figure 3. (A) *In vitro* APC/C-dependent degradation assays on selected kinesins. Results of the degradation assays of ^{35}S-labeled kinesins KIFC1, KIF18A and KIF2C as a function of incubation with G1, i.e. APC/C^{CDH1}-active, HeLa S3 cell extracts in the presence or absence of the APC/C inhibitors EMI1 and SECURIN. The known APC/C substrate SECURIN and kinesin light chain 1 (KLC1) serve as a positive and negative control, respectively. The densitometry-based quantification results of three independent degradation assays are shown. **(B) Additional panels for *in vitro* APC/C-dependent degradation assays on candidate APC/C substrates and controls.** 35S-labeled KIF4A, KIF20A and KIF23 were incubated in G1 HeLa S3 cell extracts in the absence/presence of the APC/C inhibitors EMI1 and SECURIN and their time dependent degradation was assayed by SDS-PAGE and autoradiography. KIF20A was identified as 2nd non-APC/C substrates in our study and used as a negative control. The densitometry based quantifications are based on three independent degradation assays for KIF4A and KIF20A and a single result for KIF23A.

APC/C substrates have two possible recognition motifs: the destruction box (D-box) and the KEN-box. However, neither motif is sufficient or necessary for degradation and thereby making it challenging to develop sensitive and specific bioinformatics approaches for the prediction of APC/C substrates. The D-box consensus sequence was initially characterized as RXXL [38, 39]. For follow-up experiments, we selected KIF18A and KIFC1 with 5 and 6 candidate D-box

motifs, respectively, to verify which of their multiple RXXL motifs serve as the bona fide D-boxes. These kinesins represent two distinct classes of kinesins: KIF18A is a conventional kinesin with an N-terminal motor domain, followed by the coiled-coil stalk and the C-terminal head domain and KIFC1 is one of three 'inverted' kinesins with an N-terminal head-domain followed by the stalk and a C-terminal motor domain [40]. Site-directed mutagenesis of each individual candidate D-box (RXXL -> GXXV) revealed that mutation of a single motif close to the N-terminus (G5V8) rendered KIFC1 stable in G1 extracts with APC/C active, whereas either one of two RXXL mutations towards the C-terminus, G283V286 or G668V671, rendered KIF18A stable in the assay (**Figures 4**).

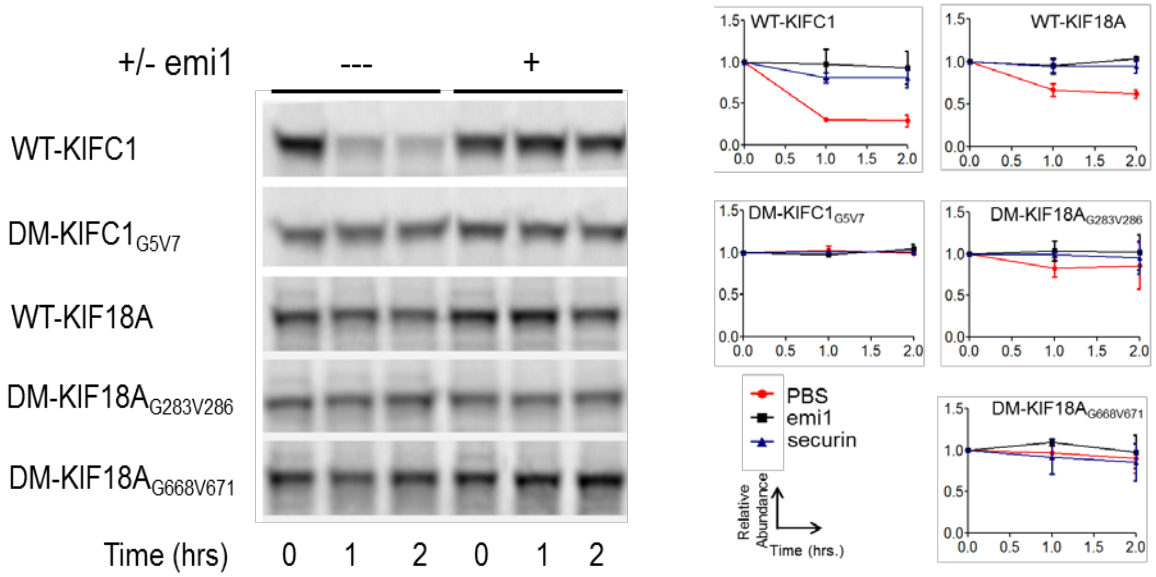


Figure 4. ***In vitro*** **APC/C-dependent degradation assays on selected kinesins.** Results of the degradation assays of ^{35}S-labeled WT or D-box mutants (DM) of KIFC1, KIF18A and SECURIN (positive control) in G1 cell extracts in the presence or absence of the APC/C inhibitors EMI1 and SECURIN. The densitometry-based quantification results of three independent degradation assays are shown.

Given that KIFC1 has an important role in spindle assembly and the bi-polar arrangement of mitotic microtubules [41, 42], we further investigated the APC/C-dependent ubiquitination of KIFC1. We further confirmed, *in vitro*, that ubiquitination of WT-KIFC1 by the APC/C was reduced when the D-box mutation was introduced (**Figure 5**).

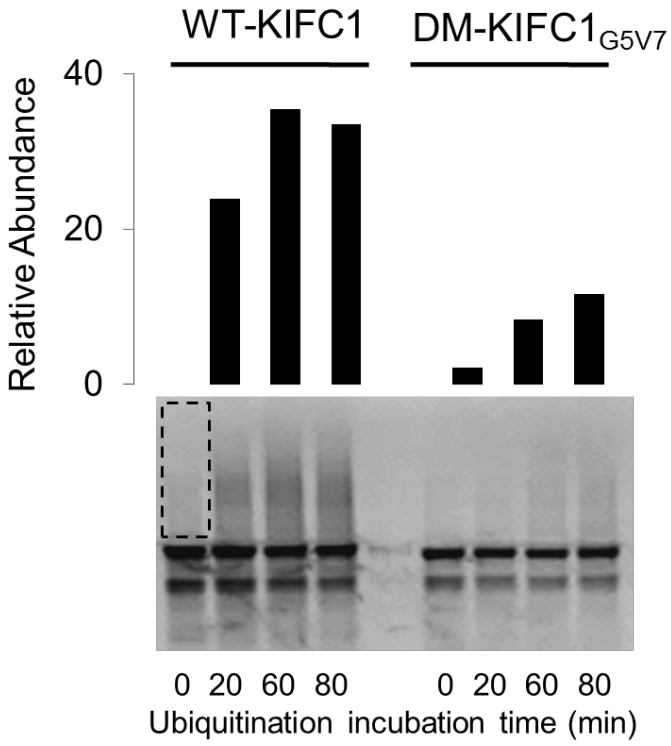
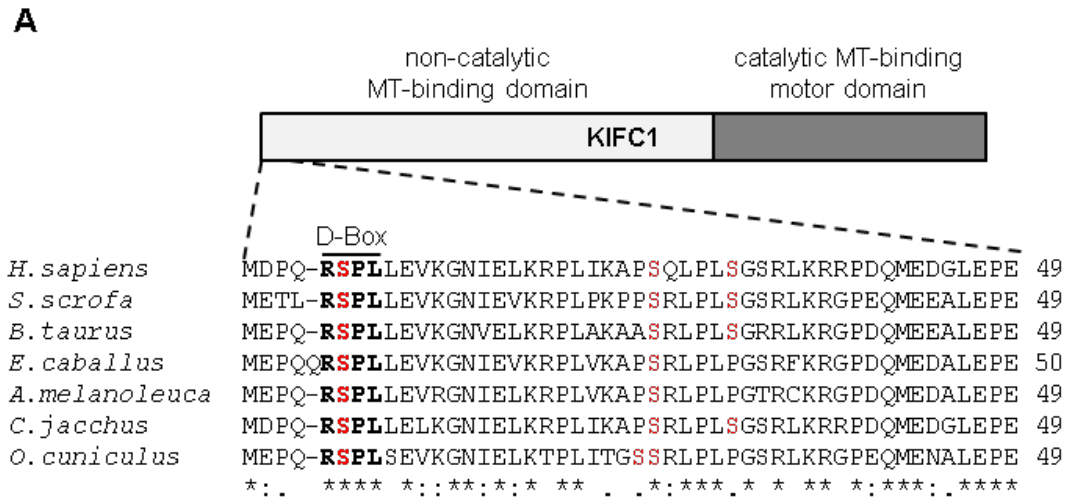


Figure 5. *In vitro* APC/C-dependent ubiquitination assays on KIFC1. APC/C-dependent *in vitro* ubiquitination of either ^{35}S-labeled WT or D-box mutated KIFC1. The high molecular weight smearing in the gel indicates ubiquitinated ^{35}S-KIFC1 (the dotted black box indicates the quantified area). The poly-ubiquitinated forms of wild-type and mutant KIFC1 were quantified and normalized to the 0 h time point.

It is not clear which cellular cues promote ubiquitination of a particular substrate by the APC/C. However, previous studies have demonstrated phosphorylation-dependent inhibition of the APC/C mediated degradation of AURORA kinase A (AURKA) [43] and SECURIN [44]. Presence of phosphorylation sites close to the degradation motifs is believe to sharpen the switch between different phases of cell cycle by negatively regulating APC/C mediated degradation until necessary [45, 46]. The phosphorylation of Ser6 site in KIFC1 has been described before in several phosphoproteomics studies [47-49]. Interestingly, this particular peptide also contains part of KIFC1's D-box motif (**Figure 6A**). Further analysis using FLEXIQuant in KIFC1 identified Ser26 and Ser31 as being phosphorylated on a tryptic peptide APSQLPLSGSR, albeit without a cell cycle dependent dynamics [50]. All these phosphorylation sites have been described previously [48, 51]. Given the cell cycle dependence of the Ser6 phosphorylation site, we hypothesized that this phosphorylation within the KIFC1 D-box motif could protect KIFC1 from premature degradation by the APC/C similar to AURORA kinase A (AURKA) and

SECURIN substrates. To test the latter hypotheses, we performed *in vitro* degradation assays on KIFC1 mutants of ser6 carrying S6D and S6A mutations. As functional controls, we also included the KIFC1 mutants carrying other two phosphorylation sites Ser26 and Ser31 to alanine or aspartate. Additional positive and negative controls included wildtype KIFC1 and its D-box mutant.

The only phosphorylation site mutation that rendered KIFC1 resistant to degradation similar to the D-box mutant was the Ser6 to Asp substitution (S6D), mimicking constitutive phosphorylation. All other mutations showed the same APC/C-dependent degradation as the wildtype KIFC1 (**Figure 6B**). Although Ser6 is located within the active D-box, its mutation does not affect the functionality of the D-box as demonstrated by the unaltered APC/C-dependent degradation of the S6A mutant. The degradation of this S6A mutant clearly shows that Ser6 is not part of the D-box function itself, but has a unique and distinct role in the phosphorylation-induced degradation inhibition mechanism. The Ser6 in KIFC1 is followed by a proline residue. The cell cycle dependent phosphorylation and stabilization of KIFC1 may be regulated by the SP-motif that is well-known site for CDK1 kinase/CYCLIN B action, similar to what has been described for SECURIN [44].



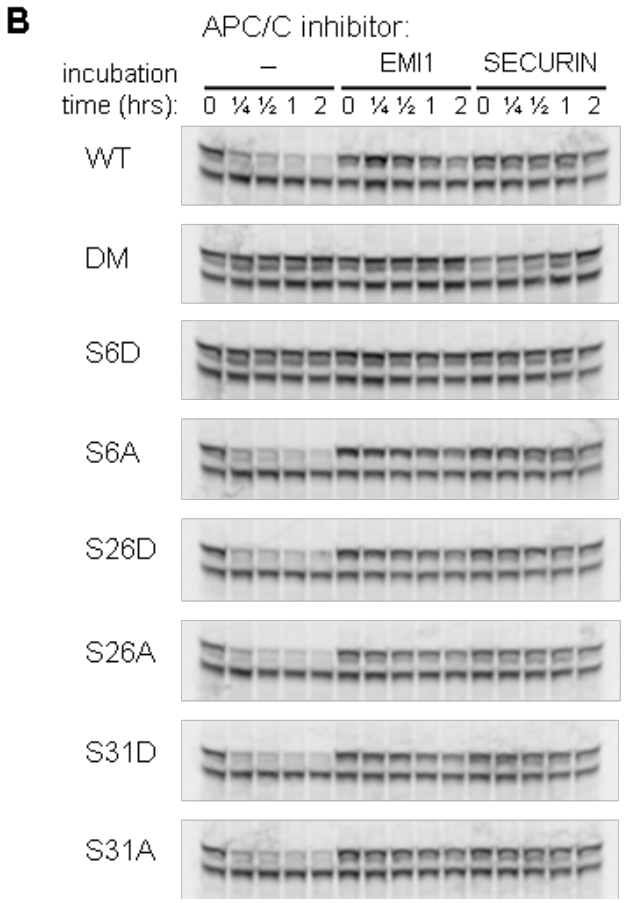


Figure 6. ***In vitro*** **degradation analysis of KIFC1 phosphorylations.** (**A)** KIFC1 D-box is conserved and N-terminally located in the protein. (**B)** Degradation assays of ^{35}S-labeled WT-KIFC1, DM-KIFC1 and various serine-to-alanine and serine-to-aspartic acid mutants in G1 cell extracts in the absence/presence of recombinant EMI1 and SECURIN. KIFC1 D-box mutant (DM) and S6D are resistant to APC/C-dependent degradation. These assays were performed n=3.

DISCUSSION

In this study we employed high throughput screening of cell cycle dependent substrates of APC/C using an approach that identifies novel substrates based on the abundance profile of already known substrates using protein profile similarity screening [52]. Our hypothesis was that the mitotic abundance profiles of yet unknown APC/C substrates will resemble those of the quantified known APC/C substrates identified in our study. This approach identified putative APC/C substrates from a quantitative proteomics dataset that profiles 4470 proteins as a function of the cell cycle. This study evaluated the biologically relevant *in vivo* protein degradation as a function of the cell cycle in an unbiased manner followed by *in vitro* validations of a subset of

identified substrates. As such, we monitored protein profiles across mitosis using the TMT strategy which permitted the simultaneous analysis of six cell cycle stages before, during and after mitosis. Using the abundance profiles of the identified known APC/C substrates as references we performed co-regulation analysis using our protein profile similarity screening strategy to identify other proteins with similar abundance profiles. The top 1 % of proteins from the protein profile similarity screen showed an 18-fold enrichment of kinesins which led us to investigate whether kinesins are more commonly degraded in an APC/C dependent manner than previously thought.

Kinesins are microtubule-associated motor proteins, some of which are known to regulate microtubule dynamics during cell division. As such, some kinesins have been described as potential targets for cancer drug development. For example, small molecule inhibitors of the kinesins KIF11 (EG5) and KIF10 (CENPE) are both in clinical trials as cancer therapies [8]. Considering the massive rearrangement of the microtubule cytoskeleton and chromosomes during mitosis, it is surprising that only two yeast kinesins, KIP1 and CIN8 and two human kinesins, KIF18A and KIF22, have been identified as APC/C substrates thus far [33-35, 53].

Given that several kinesins showed such tight co-regulation in a large-scale experiment where more than 4,400 proteins were profiled across mitosis and that these proteins could be potential targets for anti-cancer drugs, we tested whether these candidate proteins were indeed bona fide APC/C substrates. We used the well-established biochemical *in vitro* degradation assay to validate the short list of potential APC/C targets. Four of five tested candidate kinesins, KIFC1, KIF18A, KIF2C and KIF4A exhibited APC/C-dependent degradation profiles in our assay system. KIF23 resulted in inconclusive degradation profiles under the applied assay conditions.

We then focused on understanding the mechanisms behind APC/C-mediated degradation of KIFC1 and KIF18A. These two kinesins both play clear but distinct roles in cell division. For instance, KIFC1 and its opposing motor KIF11 are involved in organizing microtubule spindle structure observed in mitosis. First, we identified their D-box motifs: one was identified at the N-terminus of KIFC1 and two were identified within the C-terminal half of KIF18A. Liu et al. [54] published an *in silico* study aimed at computationally screening for APC/C substrates based on

D-box and KEN-box motifs that predicted that a number of kinesins are candidate APC/C substrates with KIFC1 and KIF4A as predicted D-box dependent substrates and KIF18A, KIF23 and KIF14A as KEN-box dependent substrates. We confirmed that the R5SPL8 sequence of KIFC1 is indeed a D-box motif. After successfully identifying the relevant D-boxes, we did not investigate any putative KEN-box motifs. However, it is now apparent that a given substrate can be APC/C-regulated by multiple recognition sites as demonstrated previously [45] and recently by an excellent structural-guided biochemical study [55]. KIF18A for example appears to possess at least two D-box motifs necessary for degradation based on our experiments, in addition to the predicted tentative KEN box [54]. The biological relevance of possessing multiple recognition sites has not been fully investigated in the field in general and is beyond the scope of this study. It is nonetheless plausible that multiple sites may generate a kinetic proofreading mechanism that regulates the susceptibility and timing of degradation.

A number of studies have investigated the roles and mechanisms by which KIFC1 regulates bi-polarization of mitotic spindles [41, 56, 57]. APC/C substrates are degraded in a very precisely ordered fashion to ensure the fidelity of cell division [5]. We hypothesized that PTMs are involved in fine-tuning this timing precision. We tested this hypothesis of cell cycle dependent PTMs using FLEXIQuant approach which allows for an unbiased identification of peptides that harbor PTM(s) and the quantification of the extent of the PTM [24, 30, 58]. One specific advantage of this method is that it prioritizes potential PTM sites for subsequent follow up experiments. While the standard qualitative MS data revealed three KIFC1 phosphorylation sites on Ser6, Ser26 and Ser31, FLEXIQuant demonstrated that only the abundance profile of the unmodified peptide S6PLLEVK12 is clearly modulated during the transition through mitosis, whereas the peptide harboring Ser26 and Ser31 demonstrated negligible changes during mitosis clearly indicating low mitosis dependency of these two phosphorylation sites [50]. Given the timing of phosphorylation and placement of Ser6 within the D-box motif, a clearer picture emerged when considering two previous studies that demonstrated the role for specific phosphorylation sites in protecting the respective protein from APC/C-mediated degradation [41, 44]. In each study, a phosphorylation site proximal to confirmed APC/C recognition motifs rendered their given substrates, SECURIN and AURORA A, resistant to degradation. In the case of KIFC1, however, the stabilizing phosphorylation site is located within the D-box itself.

Site directed mutagenesis of three identified phosphorylation sites was performed. Only the S6D mutation could stabilize KIFC1 levels whereas all other serine mutants displayed normal degradation profiles similar to the wildtype. Overall these findings are consistent with the growing notion that phosphorylation at specific sites within a substrate play a crucial role in preventing premature recognition and ubiquitination by the APC/C and hence the general timing of events in mitosis [44, 46]. We and others [59, 60] have shown that CDK1 plays a major role in mitosis and thus may be the kinase involved in stabilization of KIFC1. Our study suggests that PTMs such as phosphorylation fine tune the degradation of KIFC1.

ACKNOWLEDGEMENTS

The author acknowledges the contribution of Dr. Dominic Winter for the MS based proteomics analysis, Dr. Marc Kirchner for the candidate proteins co-regulation analysis using bioinformatics and Dr. Sasha Singh for the KIFC1 FLEXI-quant based phosphopeptides analysis.

ABBREVIATIONS

ACN	Acetonitrile
APC/C	Anaphase-Promoting Complex/Cyclosome
FA	Formic acid
MS	Mass spectrometry
PQD	Pulsed Q collision-induced dissociation
PBS	Phosphate Buffered Saline
PTMs	Post-translational modifications
TMT	Tandem mass tagging

REFERENCES

1. Hadwiger, J.A., et al., A family of cyclin homologs that control the G1 phase in yeast. Proc Natl Acad Sci U S A, 1989. **86**(16): p. 6255-9.
2. King, R.W., et al., How proteolysis drives the cell cycle. Science, 1996. **274**(5293): p. 1652-9.
3. Murray, A.W. and M.W. Kirschner, Cyclin synthesis drives the early embryonic cell cycle. Nature, 1989. **339**(6222): p. 275-80.
4. Murray, A.W., M.J. Solomon, and M.W. Kirschner, The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. Nature, 1989. **339**(6222): p. 280-6.
5. Peters, J.M., The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol, 2006. **7**(9): p. 644-56.
6. Richardson, H.E., et al., An essential G1 function for cyclin-like proteins in yeast. Cell, 1989. **59**(6): p. 1127-33.
7. Mazumdar, M. and T. Misteli, Chromokinesins: multitalented players in mitosis. Trends Cell Biol, 2005. **15**(7): p. 349-55.
8. Rath, O. and F. Kozielski, Kinesins and cancer. Nature Reviews Cancer, 2012. **12**(8): p. 527-539.
9. Ayad, N.G., et al., Tome-1, a trigger of mitotic entry, is degraded during G1 via the APC. Cell, 2003. **113**(1): p. 101-13.
10. Ayad, N.G., et al., Identification of ubiquitin ligase substrates by in vitro expression cloning. Methods Enzymol, 2005. **399**: p. 404-14.
11. McGarry, T.J. and M.W. Kirschner, Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell, 1998. **93**(6): p. 1043-53.
12. Kim, W., et al., Systematic and quantitative assessment of the ubiquitin-modified proteome. Mol Cell, 2011. **44**(2): p. 325-40.
13. Merbl, Y. and M.W. Kirschner, Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2543-8.
14. Wagner, S.A., et al., A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics, 2011. **10**(10): p. M111 013284.
15. Brady, S.M., et al., A high-resolution root spatiotemporal map reveals dominant expression patterns. Science, 2007. **318**(5851): p. 801-6.
16. Brady, S.M., et al., Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Physiol, 2007. **143**(1): p. 172-87.
17. Reinhold, W.C., et al., Identification of a predominant co-regulation among kinetochore genes, prospective regulatory elements, and association with genomic instability. PLoS One, 2011. **6**(10): p. e25991.
18. Dayon, L., et al., Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem, 2008. **80**(8): p. 2921-31.
19. Draetta, G., et al., Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF. Cell, 1989. **56**(5): p. 829-38.
20. King, R.W., et al., A 20S complex containing CDC27 and CDC16 catalyzes the mitosis-specific conjugation of ubiquitin to cyclin B. Cell, 1995. **81**(2): p. 279-88.
21. Zou, H., et al., Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science, 1999. **285**(5426): p. 418-22.

22. Clute, P. and J. Pines, Temporal and spatial control of cyclin B1 destruction in metaphase. Nat Cell Biol, 1999. **1**(2): p. 82-7.
23. Hagting, A., et al., Human securin proteolysis is controlled by the spindle checkpoint and reveals when the APC/C switches from activation by Cdc20 to Cdh1. J Cell Biol, 2002. **157**(7): p. 1125-37.
24. Singh, S., et al., FLEXIQuant: a novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. J Proteome Res, 2009. **8**(5): p. 2201-10.
25. Winter, D. and H. Steen, Optimization of cell lysis and protein digestion protocols for the analysis of HeLa S3 cells by LC-MS/MS. PROTEOMICS, 2011. **11**(24): p. 4726-30.
26. Campbell, C.C., et al., Neurovascular injury and displacement in type III supracondylar humerus fractures. J Pediatr Orthop, 1995. **15**(1): p. 47-52.
27. Bantscheff, M., et al., Robust and sensitive iTRAQ quantification on an LTQ orbitrap mass spectrometer. Molecular & Cellular Proteomics, 2008. **7**(9): p. 1702-1713.
28. Renard, B.Y., et al., When less can yield more - Computational preprocessing of MS/MS spectra for peptide identification. PROTEOMICS, 2009. **9**(21): p. 4978-4984.
29. Djebali, S., et al., Landscape of transcription in human cells. Nature, 2012. **489**(7414): p. 101-8.
30. Steen, J.A.J., et al., Different phosphorylation states of the anaphase promoting complex in response to antimitotic drugs: A quantitative proteomic analysis. Proc Natl Acad Sci U S A, 2008. **105**(16): p. 6069-6074.
31. Kirchner, M., et al., Computational protein profile similarity screening for quantitative mass spectrometry experiments. Bioinformatics, 2010. **26**(1): p. 77-83.
32. Feine, O., et al., Human kid is degraded by the APC/C-Cdh1 but not by the APC/C-Cdc20. Cell Cycle, 2007. **6**(20): p. 2516-2523.
33. Gordon, D.M. and D.M. Roof, Degradation of the kinesin Kip1p at anaphase onset is mediated by the anaphase-promoting complex and Cdc20p. Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12515-12520.
34. Hildebrandt, E.R. and M.A. Hoyt, Cell cycle-dependent degradation of the Saccharomyces cerevisiae spindle motor Cin8p requires APC(Cdh1) and a bipartite destruction sequence. Mol Biol Cell, 2001. **12**(11): p. 3402-16.
35. Sedgwick, G.G., et al., Mechanisms controlling the temporal degradation of Nek2A and Kif18A by the APC/C-Cdc20 complex. EMBO J, 2013. **32**(2): p. 303-14.
36. Zhao, W.M., et al., RCS1, a substrate of APC/C, controls the metaphase to anaphase transition. Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13415-20.
37. Schmidt, A., et al., Cytostatic factor: an activity that puts the cell cycle on hold. J Cell Sci, 2006. **119**(Pt 7): p. 1213-8.
38. Glotzer, M., A.W. Murray, and M.W. Kirschner, Cyclin is degraded by the ubiquitin pathway. Nature, 1991. **349**(6305): p. 132-8.
39. Pfleger, C.M. and M.W. Kirschner, The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev, 2000. **14**(6): p. 655-65.
40. Miki, H., Y. Okada, and N. Hirokawa, Analysis of the kinesin superfamily: insights into structure and function. Trends Cell Biol, 2005. **15**(9): p. 467-76.
41. Cai, S., et al., Kinesin-14 family proteins HSET/XCTK2 control spindle length by cross-linking and sliding microtubules. Mol Biol Cell, 2009. **20**(5): p. 1348-59.

42. Walczak, C.E., S. Verma, and T.J. Mitchison, XCTK2: a kinesin-related protein that promotes mitotic spindle assembly in Xenopus laevis egg extracts. J Cell Biol, 1997. **136**(4): p. 859-70.
43. Littlepage, L.E. and J.V. Ruderman, Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev, 2002. **16**(17): p. 2274-85.
44. Holt, L.J., A.N. Krutchinsky, and D.O. Morgan, Positive feedback sharpens the anaphase switch. Nature, 2008. **454**(7202): p. 353-7.
45. Crane, R., A. Kloepfer, and J.V. Ruderman, Requirements for the destruction of human Aurora-A. J Cell Sci, 2004. **117**(Pt 25): p. 5975-83.
46. Littlepage, L.E., et al., Identification of phosphorylated residues that affect the activity of the mitotic kinase Aurora-A. Proc Natl Acad Sci U S A, 2002. **99**(24): p. 15440-5.
47. Malik, R., et al., Quantitative analysis of the human spindle phosphoproteome at distinct mitotic stages. J Proteome Res, 2009. **8**(10): p. 4553-63.
48. Olsen, J.V., et al., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal, 2010. **3**(104): p. ra3.
49. Wang, D.H. and W.X. Yang, Molecular cloning and characterization of KIFC1-like kinesin gene (es-KIFC1) in the testis of the Chinese mitten crab Eriocheir sinensis. Comp Biochem Physiol A Mol Integr Physiol, 2010. **157**(2): p. 123-31.
50. Singh, S.A., et al., Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J, 2014. **33**(4): p. 385-99.
51. Dephoure, N., et al., A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A, 2008. **105**(31): p. 10762-7.
52. Kirchner, M., et al., libfbi: a C++ implementation for fast box intersection and application to sparse mass spectrometry data. Bioinformatics, 2011. **27**(8): p. 1166-7.
53. Feine, O., et al., Human Kid is degraded by the APC/C(Cdh1) but not by the APC/C(Cdc20). Cell Cycle, 2007. **6**(20): p. 2516-23.
54. Liu, Z., et al., GPS-ARM: computational analysis of the APC/C recognition motif by predicting D-boxes and KEN-boxes. PLoS One, 2012. **7**(3): p. e34370.
55. Tian, W., et al., Structural analysis of human Cdc20 supports multisite degron recognition by APC/C. Proc Natl Acad Sci U S A, 2012. **109**(45): p. 18419-24.
56. Cai, S., et al., Chromosome congression in the absence of kinetochore fibres. Nat Cell Biol, 2009. **11**(7): p. 832-8.
57. Mountain, V., et al., The kinesin-related protein, HSET, opposes the activity of Eg5 and cross-links microtubules in the mammalian mitotic spindle. J Cell Biol, 1999. **147**(2): p. 351-66.
58. Singh, S.A., et al., FLEXIQinase, a mass spectrometry-based assay, to unveil multikinase mechanisms. Nat Methods, 2012. **9**(5): p. 504-8.
59. Kettenbach, A.N., et al., Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. Sci Signal, 2011. **4**(179): p. rs5.
60. Ozlu, N., et al., Binding partner switching on microtubules and aurora-B in the mitosis to cytokinesis transition. Mol Cell Proteomics, 2010. **9**(2): p. 336-50.

Chapter 4

The APC/C complex Regulates HuD protein interaction with SMN protein in SMA disease

Contents **Page No.**

SUMMARY

Low abundance of SMN protein is implicated in the Spinal Muscular Atrophy, a neurodegenerative disorder. SMA disease is characterized by the specific degeneration of motor axons and neuromuscular junctions. SMN is an essential protein with ubiquitous presence in all cells; therefore its distinct molecular interactions with other RNA binding proteins such as HuD and their RNA targets may play a critical role in regulating the survival and / or proper functions of these neurons. HuD, neuron specific RBP, is known to stabilize target mRNAs, and interacts with SMN protein in a complex with its RNA targets. It is important to investigate if HuD levels can directly alter the abundance of SMN protein in neurons and if HuD is involved in the SMA disease. In this study, we investigated these aspects and observed that HuD protein regulates the stability of SMN protein and partially rescues the SMA disease in zebrafish upon overexpression. Also, we identified the novel interaction of HuD protein with an E3 ubiquitin ligase, Anaphase Promoting Complex / Cyclosome (APC/C), which regulates the stability and abundance of HuD protein by targeting the D-box motif in its sequence. Inhibition of the APC/C in mouse neuron cultures stabilizes HuD protein levels and the overexpression of D-box mutant of HuD protein partially rescues the SMA disease phenotype in zebrafish model.

INTRODUCTION

SMA is a neurodegenerative disease with affected spinal cord motor neurons and associated muscles caused by low abundance of SMN protein. The disease is typically characterized by loss of motor axonal extensions and path-finding to muscles. SMN is an essential protein present in all cells, thus, the distinct vulnerability of motor neurons may depend on their unique morphology, with very long axons, important for establishing and maintaining neuromuscular junctions. SMN protein, is a key molecule in assembling small nuclear ribonucleoproteins (snRNPs) and RNA components of spliceosome, along with its interacting partners such as Gemin 2-8 proteins [1]. Additionally, SMN has been recently recognized to be moonlighting in cytoplasm [2] implying to its multi-functionality in neurons such as - post-transcriptional regulation of RNAs by binding to RNA binding proteins [3], axonal transport of RNAs to growth cones [4] and translation regulation by interacting with polysomes [5, 6]. Various biological interactions of SMN protein in pre-RNA splicing, binding to other proteins using Tudor domain

and interactions with important RNA binding proteins (RBPs) are important for its role in the SMA disease, which is associated with aberrant splicing and SMN protein mutations in Tudor domain [7]. In particular, interaction of SMN protein with RBPs and their target mRNAs, may recruit these RNA-Protein complexes for various downstream signaling and molecular events in neurons. These complexes are believed to interact with cytoskeletal protein components and transport through axons to reach distal target sites including growth cones [4]. One such example has been reported in the case of β-actin transport by KSRP and hnRNP R, RBPs that have been shown to interact with SMN protein [3].

SMN protein has high affinity to interact with other proteins containing di-methyl arginine motifs using its Tudor domain such as snRNPs and Sm components of the spliceosome and HuD protein [8]. HuD is a neuron specific RNA binding protein (RBP) and is known to be associated with a variety of neuronal functions. It has been shown earlier that SMN interacts with HuD protein in a complex that is sensitive to RNAase treatment, suggesting that these proteins have common downstream RNA targets [9]. One of the mRNA targets of HuD, CPG15 (neuritin), was found to be interacting with SMN protein [10]. This study observed that a knock-down of SMN results in the reduction of CPG15 mRNAs, suggesting that SMN may be directly interacting with CPG15 mRNAs, and that overexpression of CPG15 results in a partial rescue of the disease phenotype in a zebrafish model of SMA [11]. However, it is not known if a direct knock-down of HuD gene expression will affect the SMN protein abundance or if HuD expression is directly associated with motor neurons specific morphology in the SMA disease. In addition, mechanisms that may regulate the stability or abundance of HuD protein are not known.

HuD provides stability and helps in the translation of mRNA targets; about ~50-60 of its targets are known to be associated with cellular process such as cell cycle progression, early neuronal differentiation, maturation of neurons, long term potentiation in synaptic plasticity, growth of neurites and cell death in neurons. Also, a number of these mRNAs have been found to be modulated in neuronal disorders including Alzheimer's disease, Huntington's disease and Schizophrenia [12]. Moreover, HuD regulates mRNAs such as GAP43 that is important for axon growth and morphogenesis; tau, which is a cytoskeletal protein crucial for axon mobility and cytoplasmic transport; Nova-1, which is a master regulator of splicing [9, 10, 13-16]. These

interactions are facilitated by the presence of kinases [17, 18]; while HuD methylation by CARM1 hinders them [19] [20]. Moreover, low levels of CARM1 expression are important for cell-cycle exit and proper differentiation of neuronal cells; while its high levels are associated with SMA disease [5] that may result in aberrant reactivation of cell-division in differentiated neurons resulting in cell death. Taken together, above findings suggest that HuD may be involved in SMA pathophysiology in more than one way; therefore, HuD dependent regulation of SMN protein stability and its direct involvement in the SMA disease should be investigated further. Plastin3 is the only protective modifier of SMA validated until now [21]. Hence, studying the direct involvement of HuD in stabilizing SMN protein and identifying the mechanisms that may regulate HuD protein in neurons are important to understand the role of HuD as a potential modifier of the SMA disease.

In this systematic study, we confirmed that HuD protein interacts with SMN protein, and also, we found the novel interaction of HuD with an E3 ligase, APC/C. This interaction was validated reciprocally in APC/C interactome using MS based affinity purified complexes of the APC/C as well as WB. We showed that APC/C ubiquitinates and modifies HuD protein for proteasomal degradation *in vitro* and APC/C inhibition in primary neuron cultures renders HuD protein stable. We observed that HuD levels are regulated via APC/C by recognizing the presence of a D-box motif in its sequence. We found that HuD knock-down in primary neuron cultures concurs with decreased levels of SMN protein and overexpression of HuD-D box mutant in zebrafish model of SMA results in a partial rescue of the disease phenotype.

EXPERIMENTAL PROCEDURES

Animals

All experimental procedures were performed in compliance with animal protocols approved by the IACUC at Children's Hospital Boston, Boston, MA.

MS and WB analyses of HuD, SMN and APC/C Co-IP experiments

The mouse spinal cord and brain tissues at age E18 were dissected and lysed using bead beater instrument (Precellys) for co-IP. Briefly, beads were conjugated with antibodies against

APC/C^{cdc27} (AF38.1, mouse, Santa Cruz), SMN (mouse, BD Biosciences), HuD (E1, mouse, Santa Cruz) and IgG (mouse, Santa Cruz) individually for 2 h at 4 ^{o}C with gentle rotation. The beads were then washed three times in PBST buffer and three times in PBST buffer containing either 150 mM KCl. The supernatant was then pre-cleared with protein A beads (Bio-Rad) for 2 h at 4 ^{o}C with constant rotation. The beads coated with primary antibody or non-specific IgG were incubated with the pre-cleared lysates overnight at 4 ^{o}C with constant rotation and were collected by centrifugation at 2,000 rpm for 15 min. at 4 ^{o}C. For WB, the protein was eluted by boiling the beads in 80 µl of Laemmli Sample Buffer (Invitrogen). The lysates were run on a 4-12 % MES Tris Gel (NuPage) and WB was performed using antibodies against APC/C^{cdc27}, HuD and SMN.

For Mass Spectrometry, to obtain high sequence coverage, the entire IP complex was run on an SDS-PAGE gel and stained with coomassie blue. The gel lanes were sliced into multiple bands and peptides from each slice were extracted, reduced with DTT, alkylated with iodoacetamide and digested with trypsin before analyzing in QE instrument with a 60 min. gradient per fraction to acquire the base peak chromatogram intensity of e^{9} - e^{10}. The raw files were converted into .mgf files and searched using mouse proteome database (Uniprot Feb 2012, (canonical and isoform sequences) protein sequence database), containing common contaminations and concatenated to its decoy version, using MASCOT v2.3 (Matrix Science) and ProteinPilot (ABSciex) software.

HuD *in vitro* degradation assays

The 1-2 µl of ^{35}S methionine labeled protein were incubated with 20 µl HelaS3 extracts enriched for G1 phase, 2 µl expressed CDH1, 1 µl of 1µg/µl E1 (Boston Biochem), 1 µl of 1 µg/µl UbcH10 (Boston Biochem), 1 µl of 1µg/µl UbcH5 (Boston Biochem), 1 µl of 1 µg/µl Ube2S (Boston Biochem), 0.1 µg/ml cycloheximide, 0.1 µg/ml ubiquitin, energy regenerating mix containing 20 mM ATP, 150 mM creatine phosphate, 2 mM EGTA pH 8.0 for a control reaction. For the reactions in presence of APC/C inhibitors either added 5 µl of 0.5 µg/µl Ni-NTA purified His tag-EMI1 or 5 µl of 0.5 µg/µl of Ni-NTA purified His tag-SECURIN or 12 µM TAME instead of same volume of PBS in control reaction. The reactions were incubated at 30 ºC for time points and stopped using Laemmli buffer (Invitrogen) containing β-mercaptoethanol

followed by heating at 70 ºC for 10 min total. The samples were resolved on 4-12 % SDS-PAGE. Gels were heat dried under vacuum, exposed to Phosphor image screen (Molecular dynamics) and scanned by Typhoon scanner (Amersham Biosciences).

HuD *in vitro* ubiquitination assays

All the constructs were cloned in pEU-E01-MCS vector containing tag free sp6 promoter (cell free sciences). The ^{35}S methionine-labeled proteins were expressed using above constructs in TNT coupled reticulocyte lysate *in vitro* transcription and translation system (Promega). The APC/C complex was immunoprecipitated using specific antibody on sepharose beads and incubated with expressed CDH1 and reagents similar to described *in vitro* degradation assays. In addition, ub-aldehyde and MG132 were used in these reactions as an inhibitor of de-ubiquitinating enzymes and proteasome inhibitor respectively. The reactions were incubated at 30 ^{o}C and samples were resolved on 4-12 % SDS-PAGE. Gels were heat dried under vacuum, exposed to Phosphor image screen and scanned by Typhoon scanner.

sh-RNA silencing of HuD in mouse cortical cultures

A set of five HuD sh-RNAs designed against different regions of HuD gene (NM_010488) was purchased from Sigma (SHCLNG glycerol stocks of constructs in PLKO.1-puro vector). The purified DNA constructs was used to transiently transfect HEK293T cells using Lipofectamine-2000 (Invitrogen) for packaging viral particles. The transfected HEK293T cells were tested for the downregulation of HuD protein by western blot (WB) using mouse HuD antibody (Santa Cruz Biotechnology), goat anti-mouse IgG-HRP secondary antibody (Santa Cruz Biotechnology) and developed by Super Signal West Pico Chemiluminescence kit (Pierce). The selected constructs HuD sh2RNA (target sequence: CCGGCCAACCTCATCGTCAACTATTCTCGAGAATAGTTGACGATGAGGTTGGTTTTTG) and HuD sh4RNA (target sequence: CGGGTGTTGCAAGTTTCCTTTAAACTCGAGTTTAAAGGAAACTTGCAACACTTTTTG) lentiviruses were used to transfect mouse cortical cultures at DIV1 of cultures.

E18 mouse cortical cells were cultured in density of 1 million cells per well of 6 well plates (Nunc) in Neurobasal media containing penicillin-streptomycin (Invitrogen), B27 supplement (Invitrogen), l-glutamine (Invitrogen). The cells were maintained for 9 days post infection with

half media replacement at day 4. The cells were collected in the lysis buffer from mirVana miRNA isolation kit (Ambion) for RNA isolation and in Laemmli buffer containing β-mercaptoethanol for WB in three different experiments. RNA was organically extracted, purified using protocol given with mirVana miRNA isolation kit (Ambion) for total RNA isolation, treated with DNAase and measured by nanodrop quantitation at optical density ratio of A260/280. RNA was reverse transcribed using oligo $(dT)_{12-18}$ primers of Superscript First Strand synthesis system (Invitrogen). The cDNA was treated with RNaseH and amplified with sybr green PCR master mix (Applied Biosystems) using primers for GAPDH, HuD, cpg15, GAP43 and Tau (in **Table 1**). The signals were measured at reaction conditions of 50 ^{o}C for 2 min, 95 ^{o}C for 10 min, followed by 39 cycles of 95 ^{o}C for 15 sec, 60 ^{o}C for 30 sec, 72 ^{o}C for 30 sec using CFX connect real-time system (Biorad). Signals were collected after the melting curve reaction of 95 ^{o}C for 1 min, 55 ^{o}C for 1 min with melting curve from 55 ^{o}C to 95 ^{o}C incrementing at 5 ^{o}C for 5 sec to ensure that measurements of Cq intensity values came from specific products without any contaminants. The values were normalized with GAPDH as internal control.

RTPCR Region	Forward primer 5' to 3'	Reverse primer 5' to 3'
GAPDH	TGTGTCCGTCGTGGATCTGA	CCTGCTTCACCACCTTCTTGA
Cpg15	ATGGGACTTAAGTTGAACGG	TCAGAAGGAAAGCCAGGTCG
GAP43	CCAGCCAAGGAGGAGCCTAAA	AGCCTCGGGGTCTTCTTTACC
Tau	CACCCAGCCCAAAGACTCCTC	TGGTGCTTCAGGTTCTCAGTG
HuD	ATAAGTAAGGGTGAGAAATTCAGG	TGCTTAATATGGCCTTATGGCG

Table 1: The table showing RT-PCR primers to amplify signals of specific genes.

Inhibition of APC/C activity using EMI1 competitive inhibitor

The EMI1 viruses were used to infect HEK293T cells and to analyze levels of CYCLIN-B using WB. The embryonic stage of E18 from wild type mouse was used to dissect cortical cultures and infected with EMI1 viruses at day 1 with lentiviruses that contained either EMI1 or the empty control CMV promoter. The cells were observed under microscope for GFP^{+} cells for effective transfections and collected at day 9 for WB of tubulin and HuD.

Inhibition of APC/C activity by small molecule inhibitor proTAME

The embryonic stage of E18 from wild type mouse was used to dissect cortical cultures. The cells were treated with 6 μM and 12 μM proTAME for 24 h and collected for WB of tubulin and HuD.

For WB, cells were collected in sample Buffer containing β-mercaptoethanol. The lysates were run on a 4-12 % SDS-PAGE gel.

Overexpression of HuD in the zebrafish SMA model

Human HuD cDNA was PCR amplified and cloned in $pCS2^{+}$ vector and modified to create D-box HuD mutant using *in vitro* mutagenesis. Plasmids were sequence verified and tested for expression using wheat germ expression system and WB to validate the expression. Each plasmid was linearized with Not1 and capped RNA was generated using the sp6 mMESSAGE mMACHINE kit (Ambion).

For HuD overexpression experiments, 1-2 cell stage Tg embryos (hb9:GFP) were injected with 9 ng of control morpholino (MO), 9 ng of SMN MO, 9 ng of SMN MO with 100 pg HuD or 100 pg HuD-D box mutant RNA individually using an MPPI-2 pressure injector (Applied scientific instrumentation). Fifty embryos per experiment were analyzed from three separate experiments at 28 h post fertilization. The live embryos were anesthetized with tricaine, mounted on glass coverslip and observed with Zeiss axio-plan2 microscope. Motor neurons were scored as described previously [6]. In brief, motor neurons innervating the mid-trunk myotomes 7-16 were examined under the microscope. Normal motor axon morphology is stereotyped along the trunk targeting myotomes while decreasing SMN levels result in the motor axon truncations and abnormal branching. The abnormal branching is characterized as the extra branches along the entire length of the motor axon or excessive branching at the distal region of the motor axon. The fish were classified as having severe, moderate, mild, or no defects based on the character of the motor axon defects. Fish with at least 20 % severe axon defects (i.e. truncations or truncation with abnormal branching) or 40 % moderate defects (i.e. abnormal branching without truncations, “wishboned” axons) were classified as severe. Fish classified as moderate typically

shows 10 % severe defects, 20-40 % moderate defects, or >40 % mild defects (i.e. axon lacking stereotyped morphology, but not abnormally branched). Fish classified as mild shows 10 % moderate defects and 20-40 % mild defects. Each side of the fish was scored for total 20 motor axons and this classification was used from the combined motor neuron defects. Data was analyzed and statistical significance was determined using the mann-whitney nonparametric rank test.

RESULTS

While our previous investigation of HuD interactome identified SMN protein along with various members of the APC/C complex (unpublished observation), we were interested in determining if the reciprocal analysis would identify HuD in APC/C IPs from neurons. Also, one way to determine the identity of substrates of an E3 ubiquitin ligase is to study its interactome [22]. Therefore, a co-IP of the APC/C complex was performed, using antibody against one of its subunit cdc27, from the developing neurons of mouse embryonic E18 brain and spinal cord tissue. The samples were analyzed using MS based proteomics. Co-IP experiments were performed using high stringency conditions to isolate tightly bound complex and analyzed using a rigorous 1 % false discovery rate (FDR) cutoff for the database searches to ensure a minimal false positive rate in the protein identification process. IgG controls were used to eliminate non-specific interactions and to exclude common proteins from the list of protein hits established from the target IPs. The LC-MS/MS analysis identified several core subunits of APC/C complex including the cullin-based scaffold protein APC2, APC5 and CDC27. Other interacting partners included various ribosomal subunits, splicing proteins, proteins of SMN complex and RNA binding proteins such as HuD.

We then used WB analyses as an orthogonal approach to MS to specifically detect interacting proteins in reciprocal co-IP experiments. The WB analyses of the endogenous IP complexes from mouse embryonic spinal cord and brain tissue detected the presence of SMN, HuD and APC/C^{cdc27} proteins in reciprocal IPs **(Figure 1)**, suggesting that these proteins are interacting *in vivo* in mouse neurons. Validation of the interaction of APC/C with SMN and HuD proteins in nervous tissue prompted us to study the role of the APC/C in these complexes with SMN and HuD proteins.

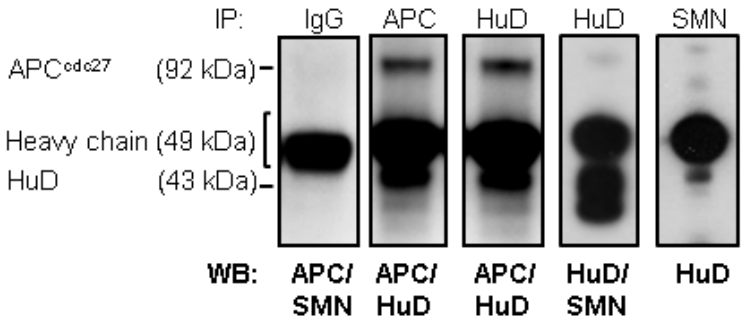


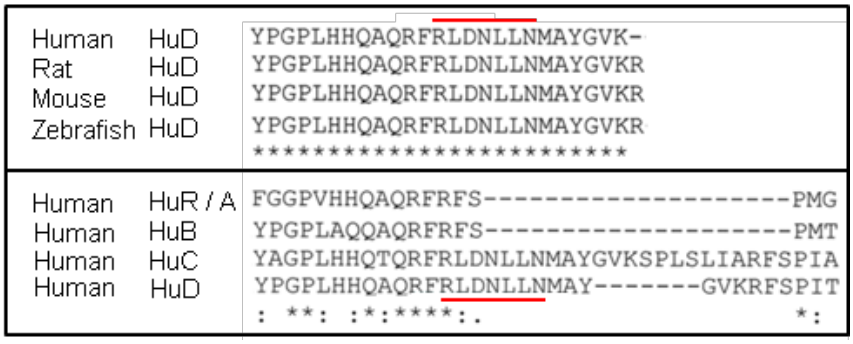
Figure 1: APC/C interacts with HuD and SMN. WB analyses of co-IP experiments from the E18 mouse spinal cord and brain using antibodies directed against APC/C^{cdc27}, SMN and HuD. These reciprocal co-IP experiments are annotated above the lanes and antibodies used for the detection of specific proteins are indicated below the image.

The APC/C is an E3 ubiquitin ligase that promotes protein degradation by directly interacting and catalyzing the ubiquitination of target molecules in the presence of co-activators such as CDH1 and CDC20 [23]. We hypothesized that APC/C may regulate HuD protein stability in neurons. There are usually tight regulatory mechanisms that govern the degradation of substrates by E3 ubiquitin ligases and the regulation often involves specific motifs and/or post-translational control of the degradation event. APC/C targets often carry a destruction motif called D-box (RXXLXXN/D/E), KEN-box (KENXXXN) [24], however other APC/C recognition motifs have been described, including the A-box [25] and O-box [26] or CRY box [27]. These motifs are neither necessary nor sufficient for the APC/C dependent protein turnover, yet indicative of a putative substrate. We noted that out of the Hu-family members identified in the co-IP experiments, HuC and HuD proteins contain D-box motif that remains particularly well-conserved in the HuD sequence among vertebrates (**Figure 2A**).

To investigate if HuD is regulated by the direct action of the APC/C E3 ubiquitin ligase, we performed *in vitro* degradation assays. Briefly, lysates are made from HeLa cells synchronized in the G1 phase for enzymatically active APC/C fractions. These lysates were first tested for endogenous APC/C activity using SECURIN or KIFC1 proteins, well characterized substrates of the APC/C [28]. KIFC1 was used as control to compare the degradation mechanism and kinetics of potential neuronal substrates of APC/C. The specificity of the degradation observed in the G1 extract is tested by the addition of recombinant expressed and purified protein inhibitors of the

APC/C such as SECURIN, a substrate and a competitive inhibitor of APC/C [28] or EMI1, an endogenous inhibitor of APC/C [29]**.** In addition, TAME a small molecule inhibitor of the APC/C can be used to show the specific inhibition of the APC/C [30]. In our assay HuD showed a robust APC/C-dependent degradation, where HuD protein is degraded within 1-2 h and this degradation is inhibited by the specific inhibitors of APC/C, EMI1 and SECURIN (**Figure 2B**). The degradation of HuD protein was quantified over time in presence of APC/C and its inhibitors (n=3) and p values were calculated (**Figure 2B)**. To assess if HuD is a D-box dependent substrate of APC/C, we generated two point mutation in D-box of HuD (HuD-DM) by modifying Arg243 to Gly and Leu246 to Val that must render mutant protein stable in presence of APC/C in comparison to the wildtype HuD protein, as is shown in the case of KIFC1 mutant. Interestingly, the D-box mutant of HuD remains stable in the presence of APC/C as well as its inhibitors, suggesting that HuD protein is a D-box dependent target of APC/C complex (**Figure 2B**).

A. D-box motif in HuD sequence alignment



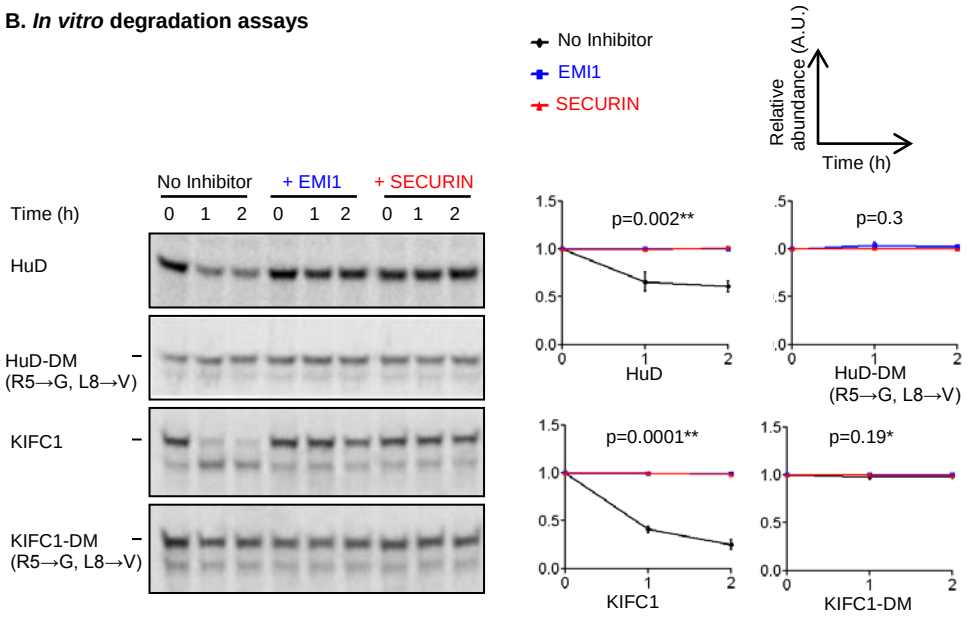


Figure 2: (A). Sequence alignment of canonical D-box motif of HuD gene. HuD has a typical APC/C detectable D-box (outlined with red) is conserved among vertebrates. **(B) APC/C dependent *in vitro* degradation of HuD protein.** ^{35}S methionine labeled HuD protein degrades in presence of active APC/C, whereas mutations in the APC/C recognized D-box of HuD, HuD-DM, stabilize the protein. The degradation assays of the substrates is performed in the presence of APC/C specific inhibitors such as EMI1 and SECURIN to ensure that the degradation is not due to other E3 ubiquitin ligase activities in the lysates. KIFC1 was used as a control to test the APC/C activity in the lysate and degrades within 1 h, while KIFC1-D box mutant remains stable. Quantitation of bands is measured over time by densitometry shown in graphs for control reaction without inhibitor-black and with inhibitor EMI1-blue, SECURIN-red. The data represents the mean ± SEM of three experiments.

To validate the degradation results and to determine if HuD is a substrate of the APC/C E3 ligase, we performed *in vitro* ubiquitination assays using IP purified complex of enzymatically active APC/C isolated from synchronized HeLa cell lysates using specific antibodies [31]. In addition, the reactions were performed in the presence of proteasome inhibitor MG132 and the de-ubiquitinating enzyme inhibitor ub-aldehyde to allow accumulation of modified protein. The *in vitro* expressed and ^{35}S labeled proteins were added to the reaction mixture and samples were analyzed at 0, 20, 40 and 80 minutes on an SDS-PAGE gel. The assays show that the HuD is ubiquitinated as observed by the existence of higher molecular weight bands and smear appearing within 1 h while HuD D-box mutant remains unmodified confirming the D-box dependent regulation by APC/C. The modified protein was quantified by densitometry of a rectangle selection drawn above the normal size band of unmodified protein (n=3). The values were normalized to the zero time point for graph plots and p values were calculated **(Figure 3).**

These results indicate that HuD protein is a novel neuron-specific substrate of APC/C E3 ligase and depends on ubiquitination mediated regulation for its stability.

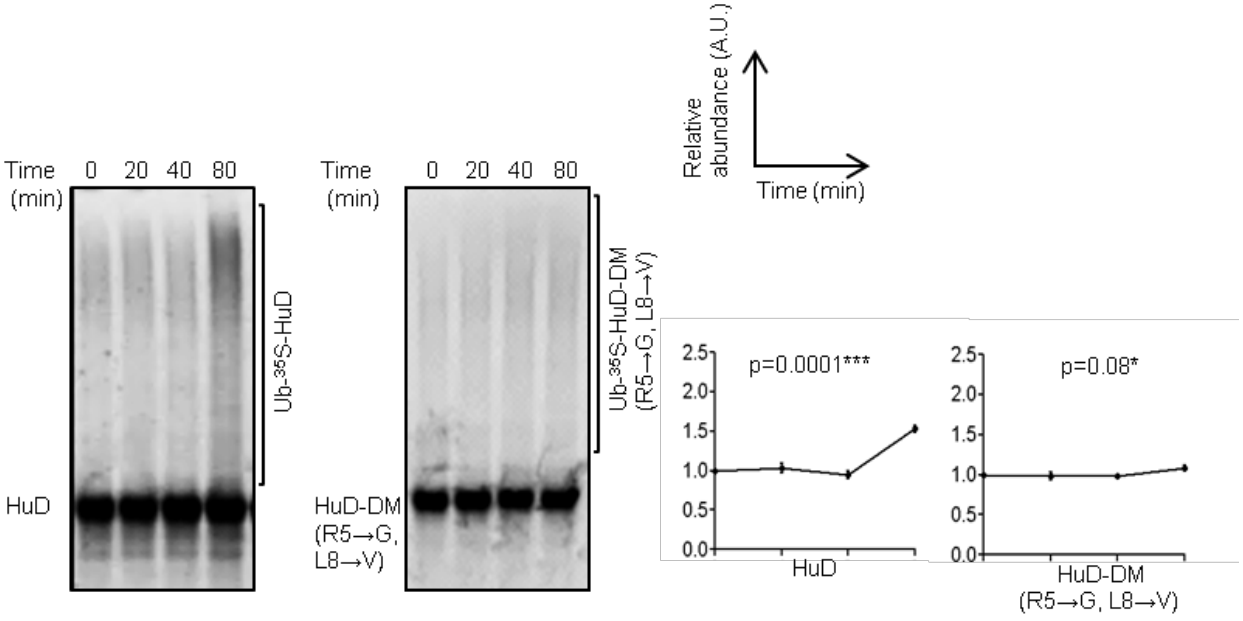


Figure 3: APC/C dependent *in vitro* ubiquitination assays. ^{35}S methionine labeled HuD protein is tagged with ubiquitin in presence of immunopurified and active APC/C as shown by high molecular bands and smear of modified protein. HuD-DM is not modified as expected. Images were quantified by drawing a rectangle box (area indicated by a bracket on the side of images) above the unmodified protein band. The data represent the mean ± SEM of three experiments.

We followed the regulation of the HuD protein by APC/C in the physiological context of primary cortical neuron cultures by overexpression of EMI1 inhibitor and treatment with small molecule inhibitor, proATME. First, we cultured the primary cortical neurons from E18 stage of mouse and after 1 DIV (days *in vitro*) lentivirally infected with EMI1 gene construct containing CMV promoter and GFP reporter. GFP expression was used to observe the efficiency of transfection and CMV plasmid was used as control. Also, the constructs were transfected in HEK293T cells prior to cortical cells to optimize the method by testing the effect of APC/C inhibition on a cell-cycle substrate, CYCLIN-B. HEK293T cells that were infected with the empty control vector had no detectable CYCLIN-B levels as expected in the presence of APC/C activity while cells infected with EMI1 exhibited increase levels of CYCLIN-B in the WB **(Figure 4A)**. Similarly, the cortical neurons showed increased presence of HuD protein **(Figure 5A)** upon APC/C inhibition by 9 DIV (days *in vitro*). Since levels of HuD protein are relatively

high at this time point, we collected lysates from neurons after the treatment with a translational inhibitor cycloheximide (CHX) for 4 h, which allowed for a better visualization of subtle differences as detected by WB. The results were quantified (n=3) using densitometry. These results suggest that endogenous inhibition of APC/C by EMI1 results in the stabilization of HuD protein.

Further, we tested pharmacological inhibition of APC/C by using a small molecule inhibitor proTAME, a pro-drug version of TAME (tosyl-L-arginine methyl ester) molecule that is more readily internalized though cell membrane and is known to induce mitotic arrest in developing cells by direct inhibition of APC/C activity. We cultured the primary cortical neurons from E18 stage of mouse and treated them with DMSO vehicle and proTAME. We tested the optimal concentration of drug to inhibit the APC/C activity in post-mitotic cortical neuron culture and found that a dose of 12-15 μM, which is effective in cell cycle, is sufficient to increase the HuD protein expression as analyzed by WB after 24 h. The results were quantified (n=3) using densitometry and p values were calculated (**Figures 4B)**. The optimal dose was selected to avoid any cytotoxic effects as studied by monitoring cell survival and morphology, and a decrease in the dose to 6 μM did not result in detectable increases in the levels of HuD protein. These results confirm that APC/C activity is required for the degradation of these proteins and that inhibition of this activity in neurons leads to an increase in levels of HuD protein.

Taken together these observations suggest that APC/C complex is associated with the clearance of HuD protein in these cells.

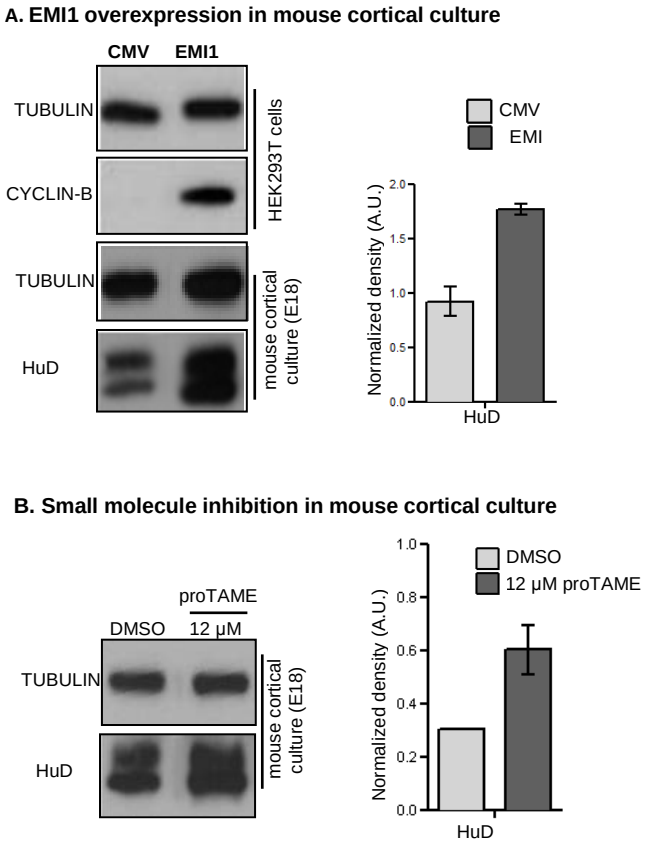


Figure 4: (A) Over-expression of EMI1 in cortical cultures increases HuD. Cortical cells from wild type mouse were transfected using CMV promoter directed lentivirus construct that leads to the accumulation of CYCLIN-B in HEK293T cells in a parallel experiment (upper panel) and of HuD in cortical cells (lower panel). Histogram shows a densitometry analysis of bands (n=3). **(B) proTAME inhibition of APC/C in cortical cultures increases HuD**. Cortical cells were treated with proTAME that leads to increase in levels of HuD at a dose of 12 µM in cultured cortical neurons as compared to DMSO. Histograms show a densitometry analysis of bands (n=3).

As discussed earlier, HuD and SMN proteins closely interact, thus in this study we explored their direct association in the context of SMA disease. We investigated if HuD knockdown, in primary cortical neuron cultures using sh-RNA affects its downstream RNA targets and SMN protein levels. A subset of known interacting mRNAs of HuD such as CPG15, GAP43 and TAU along with GAPDH as control were investigated in cells transfected with shRNA-HuD. These RNA levels were detected by quantitative RT-PCR. GAPDH levels were used as internal control for normalization. The measurements were analyzed in biological replicates (n=3) and plotted as fold difference of Cq intensity values **(Figure 5A).** These results provided further evidence that HuD regulation results in the drastically reduced levels of its mRNA targets as compared to the experimental cell controls. In addition, these cells were analyzed for the protein abundance of

HuD and SMN to investigate if reduced levels of HuD proteins would destabilize SMN levels. The WB analysis of HuD and SMN protein levels in these samples show a decrease in HuD as well as SMN protein **(Figure 5B).** These findings further support that SMN and HuD not only closely associate but also that HuD protein stabilizes SMN levels in a complex; that may include mRNA targets of these proteins.

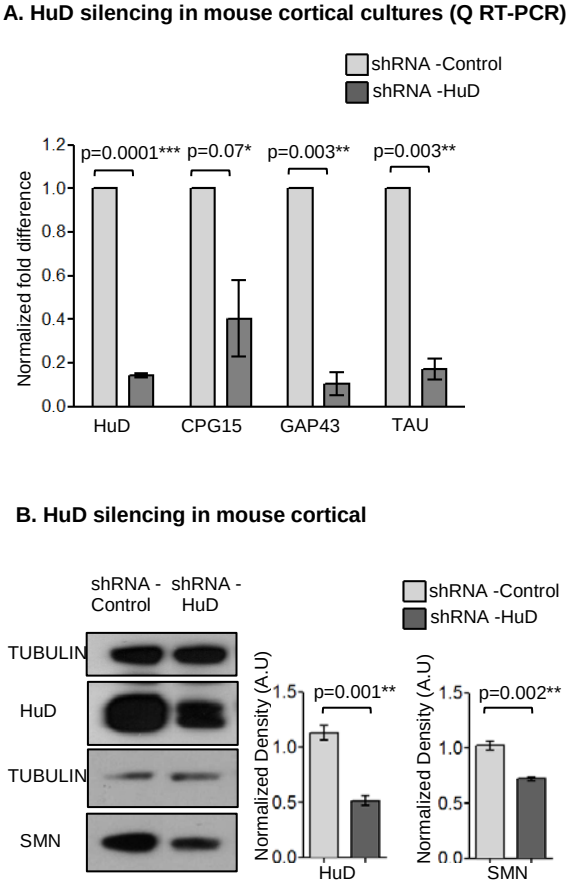


Figure 5: (A) HuD downregulation results in SMN transcript level in mouse cortical cultures. Quantitative RT-PCR shows silencing of HuD using sh-RNA resulting in reduced levels of target transcripts including HuD mRNAs in mouse cortical cultures. The GAPDH as control is used as control for RT-PCR and average fold difference were normalized to cells treated with control shRNA. **(B) HuD downregulation results in SMN protein level in mouse cortical cultures.** WB of samples collected from HuD-shRNA shows reduced levels of HuD and SMN protein. TUBULIN is used as internal standard. The data represent the mean ± SEM of three experiments.

Having established the direct interaction of HuD with SMN and APC/C, we were interested in understanding the importance of these interactions in the context of SMA disease. Since HuD provides stability to SMN protein, we investigated if its overexpression in SMN MO background of zebrafish model of SMA [32] would alter the disease phenotype. Also, we explored the HuD-

D box mutant along with wild type HuD in these assays to validate our earlier observations. The 1-2 cell stage embryos in transgenic zebrafish with GFP-positive motor neurons were injected with SMN morpholino (MO) with or without HuD as well as HuD-D box mutant RNAs and analyzed at 28 h post-fertilization in comparison to control MO. The end points were analyzed for the levels of SMN protein along with axon phenotypes of SMA motor neurons characterized as truncated axon ends with aberrations that are untargeted towards the myotomes. The progeny is then scored over a 'severity of disease' based classification system of severe defects, mild defects, moderate defects, no defects or normal phenotype. The HuD expression in morphants was validated using WB analysis. The co-injections of HuD with SMN MO resulted in the partial rescue of SMA phenotype as observed by imaging and quantitation (n=3). The SMA morphants were quantified in the percentage values of severe 27.49 ± 0.3, moderate 57.5 ± 1.8, mild 13.76 ± 0.9 and no defect 1.24 ± 0.3 phenotypes. After injection of HuD RNA, the morphants were quantified as percentage values of severe 6.68 ± 1.2, moderate 42.66 ± 1.4, mild 38.66 ± 0.3, no defect 12 ± 1.7 percent phenotypes while injection of HuD-DM RNA resulted in a rescue in each class depicted as percentage values of severe 5.12 ± 1.3, moderate 39.73 ± 0.8, mild 41.04 ± 1.2, no defect 14.12 ± 0.3 percent phenotypes **(Figure 6)**. The HuD-D box mutant resulted in a better rescue than HuD not only supports that stabilized HuD protein in the post-mitotic motor neurons may help in rescue of the SMA phenotype but also that APC/C mediated regulation of HuD protein is important in understanding the molecular network of this disease.

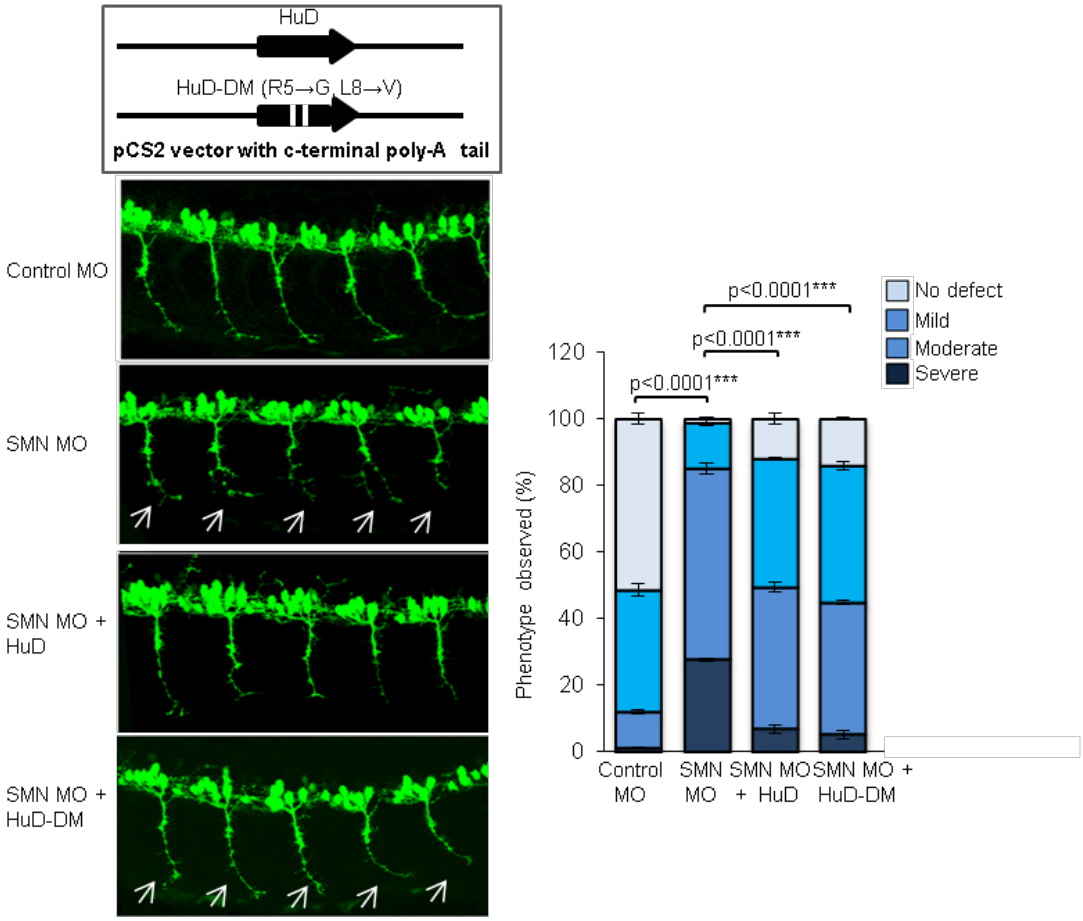


Figure 6: HuD overexpression partially rescues SMA phenotype in zebrafish SMA model. Overexpression of HuD in zebrafish SMA morphants results in rescued phenotype of characteristic truncations and aberrated branches in motor neurons and better path-finding shown as representative image. The effect is more pronounced using HuD-D box mutant, an APC/C degradation resistant form of HuD. The quantitative phenotype of morphants is shown in graphs from various classes of severe, mild, moderate and no defect categories. The data represent the mean ± SEM of three experiments. Upper panel shows construct design for this experiment. Arrows indicate the differences in phenotype upon rescue.

DISCUSSION

The levels of SMN protein in spinal motor neurons connected with muscles as 'motor units' are clinically important and the determining factor of diagnosis and therapeutics in debilitating SMA disorder. We confirmed that HuD interacts with SMN protein using proteomics that also revealed the novel interaction of HuD with APC/C complex in mouse embryonic spinal cord and brain. Although, it has been shown that HuD interacts with various mRNAs and SMN protein in subcellular compartments of neurons such as axons. It is not known if HuD may directly be

involved in regulation of SMA disease and if silencing of HuD expression may result in destabilization of SMN protein. In this study, we investigated these novel aspects of HuD's involvement in stability of SMN protein and SMA disease regulation. Besides, we explored the mechanisms controlling the stability and abundance of HuD protein itself that are not known yet.

APC/C has been gaining attention for its potential involvement in the regulation of various neuronal substrates separate from its earlier well characterized role in regulating cell cycle proteins. We found APC/C to be interacting in a complex with HuD along with SMN in neurons and hypothesized that being a potent E3 ubiquitin ligase; APC/C might be involved in regulating the abundance of HuD protein. We noted that HuD protein has a D-box motif that can be altered to investigate APC/C mediated ubiquitination mechanism. The *in vitro* biochemistry assays revealed that HuD undergoes direct ubiquitination and degradation in the presence of enriched APC/C but remains stable in the presence of APC/C inhibitors such as EMI1 and SECURIN. While, the D-box mutant of HuD remains stable in presence or absence of APC/C or its inhibitors. These results confirmed that HuD is a novel neuron specific target of APC/C. The APC/C inhibition in primary neuron cultures from mouse showed that inactivation of APC/C, leads to stabilization and increased levels of endogenous HuD protein, by the overexpression of EMI1 and treatment with small molecule inhibitor, proTAME.

Further, we observed that knock-down of HuD gene expression concurs with the low levels of SMN protein in mouse primary cortical cultures. This suggests that HuD and SMN proteins are co-regulated in neurons and HuD may be involved in stability of SMN protein. Therefore, we analyzed the overexpression of HuD RNA in a zebrafish model of SMA generated by SMN morpholino. Overexpression of HuD, in particular HuD-D box, results in the partial rescue of SMA disease phenotype. These studies collectively confirmed that HuD is an effective regulator of SMN protein stability in the context of SMA disease. Also, we showed that HuD is a novel target of APC/C. This regulation may potentially fine-tune the abundance and functions of HuD affecting various downstream targets in neurons. This study provides a novel understanding of molecular modifiers associated with SMN protein in the SMA disease.

ACKNOWLEDGEMENT

The author acknowledges the contribution of Dr. Bikem Akten for the initial experiments on APC/C inhibition in cortical cultures, Dr. Hao Le for introducing the constructs in zebrafish SMA model and imaging the embryos.

ABBREVIATIONS

APC/C (APC)	Anaphase Promoting Complex
D-box	Degradation box motif
HUD (ELAV)	Hu family protein member (Embryonic Lethal Abnormal Vision)
IP	Immunoprecipitation
LC-MS/MS [20]	Liquid Chromatography based Mass-Spectrometry
PBS	Phosphate Buffered Saline
PTM	Post Translational Modifications
Q (RT-PCR)	Quantitative Reverse Transcription- Polymerase chain reaction
RBP	RNA binding protein
SMA	Spinal Muscular Atrophy
SMN	Survival of Motor Neurons
UPS	Ubiquitin Proteasome System

REFERENCES

1. Pellizzoni, L., et al., A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell, 1998. **95**(5): p. 615-24.
2. Stump, A.D., M. Dillon-White, and S. Gu, Molecular evolution of the moonlighting protein SMN in metazoans. Comp Biochem Physiol Part D Genomics Proteomics, 2013. **8**(3): p. 220-30.
3. Rossoll, W., et al., Specific interaction of Smn, the spinal muscular atrophy determining gene product, with hnRNP-R and gry-rbp/hnRNP-Q: a role for Smn in RNA processing in motor axons? Hum Mol Genet, 2002. **11**(1): p. 93-105.
4. Piper, M. and C. Holt, RNA translation in axons. Annu Rev Cell Dev Biol, 2004. **20**: p. 505-23.
5. Sanchez, G., et al., A novel function for the survival motoneuron protein as a translational regulator. Hum Mol Genet, 2013. **22**(4): p. 668-84.
6. Carrel, T.L., et al., Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis. J Neurosci, 2006. **26**(43): p. 11014-22.
7. Cusco, I., et al., Detection of novel mutations in the SMN Tudor domain in type I SMA patients. Neurology, 2004. **63**(1): p. 146-9.
8. Hubers, L., et al., HuD interacts with survival motor neuron protein and can rescue spinal muscular atrophy-like neuronal defects. Hum Mol Genet, 2011. **20**(3): p. 553-79.
9. Fallini, C., et al., The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly(A) mRNA in primary motor neuron axons. J Neurosci, 2011. **31**(10): p. 3914-25.
10. Wang, Z.H., et al., HuD regulates the cpg15 expression via the 3'-UTR and AU-rich element. Neurochem Res, 2011. **36**(6): p. 1027-36.
11. Akten, B., et al., Interaction of survival of motor neuron (SMN) and HuD proteins with mRNA cpg15 rescues motor neuron axonal deficits. Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10337-42.
12. Perrone-Bizzozero, N. and C.W. Bird, Role of HuD in nervous system function and pathology. Front Biosci (Schol Ed), 2013. **5**: p. 554-63.
13. Hinman, M.N. and H. Lou, Diverse molecular functions of Hu proteins. Cell Mol Life Sci, 2008. **65**(20): p. 3168-81.
14. Bolognani, F., T. Contente-Cuomo, and N.I. Perrone-Bizzozero, Novel recognition motifs and biological functions of the RNA-binding protein HuD revealed by genome-wide identification of its targets. Nucleic Acids Res, 2010. **38**(1): p. 117-30.
15. Tiruchinapalli, D.M., M.D. Ehlers, and J.D. Keene, Activity-dependent expression of RNA binding protein HuD and its association with mRNAs in neurons. RNA Biol, 2008. **5**(3): p. 157-68.
16. Pascale, A., et al., Increase of the RNA-binding protein HuD and posttranscriptional up-regulation of the GAP-43 gene during spatial memory. Proc Natl Acad Sci U S A, 2004. **101**(5): p. 1217-22.
17. Wang, W., et al., AMP-activated protein kinase-regulated phosphorylation and acetylation of importin alpha1: involvement in the nuclear import of RNA-binding protein HuR. J Biol Chem, 2004. **279**(46): p. 48376-88.

18. Winzen, R., et al., Distinct domains of AU-rich elements exert different functions in mRNA destabilization and stabilization by p38 mitogen-activated protein kinase or HuR. Mol Cell Biol, 2004. **24**(11): p. 4835-47.
19. Fujiwara, T., et al., CARM1 regulates proliferation of PC12 cells by methylating HuD. Mol Cell Biol, 2006. **26**(6): p. 2273-85.
20. Adams, M.D., et al., The genome sequence of Drosophila melanogaster. Science, 2000. **287**(5461): p. 2185-95.
21. Oprea, G.E., et al., Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. Science, 2008. **320**(5875): p. 524-7.
22. Sowa, M.E., et al., Defining the human deubiquitinating enzyme interaction landscape. Cell, 2009. **138**(2): p. 389-403.
23. Visintin, R., S. Prinz, and A. Amon, CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. Science, 1997. **278**(5337): p. 460-3.
24. Pfleger, C.M. and M.W. Kirschner, The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev, 2000. **14**(6): p. 655-65.
25. Littlepage, L.E. and J.V. Ruderman, Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev, 2002. **16**(17): p. 2274-85.
26. Okudaira, K., et al., Transcriptional regulation of the Drosophila orc2 gene by the DREF pathway. Biochim Biophys Acta, 2005. **1732**(1-3): p. 23-30.
27. Reis, A., et al., The CRY box: a second APCcdh1-dependent degron in mammalian cdc20. EMBO Rep, 2006. **7**(10): p. 1040-5.
28. Singh, S.A., et al., Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J, 2014. **33**(4): p. 385-99.
29. Reimann, J.D., et al., Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. Cell, 2001. **105**(5): p. 645-55.
30. Zeng, X., et al., Pharmacologic inhibition of the anaphase-promoting complex induces a spindle checkpoint-dependent mitotic arrest in the absence of spindle damage. Cancer Cell, 2010. **18**(4): p. 382-95.
31. Passmore, L.A., D. Barford, and J.W. Harper, Purification and assay of the budding yeast anaphase-promoting complex. Methods Enzymol, 2005. **398**: p. 195-219.
32. McWhorter, M.L., et al., Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding. J Cell Biol, 2003. **162**(5): p. 919-31.

Chapter 5

The APC/C complex Regulates SMN protein and its inhibition Rescues SMA disease

Contents **Page No.**

SUMMARY

Spinal Muscular Atrophy is a neurodegenerative disorder caused by the low abundance of Survival of Motor Neuron (SMN) protein. Normal levels of SMN protein are crucial for motor neuron survival; therefore, increasing the stability and abundance of the SMN protein is the driving force of the discovery of potential therapeutics. Earlier, we showed that SMN protein stability, and also the SMA phenotype, are in part regulated by the HuD protein in neurons. Also, it is known that SMN protein undergoes ubiquitination and degradation by proteasome. However, the mechanism of SMN ubiquitination is not known. In this study, we showed that SMN protein is a direct substrate of the Anaphase Promoting Complex / Cyclosome (APC/C), an E3 ubiquitin ligase, for ubiquitination and degradation. Therefore, we report the novel interaction of SMN protein with APC/C that regulates the SMN levels in neurons and that SMN ubiquitination is facilitated by a GSK3β phosphorylation site. Importantly, the inhibition of the APC/C in mouse cell cultures and *in vivo* zebrafish model of SMA increases SMN protein levels. Finally, the APC/C inhibition in this SMA disease model rescues the SMA phenotype remarkably.

INTRODUCTION

SMA is a neuromuscular disease with selective loss of spinal motor neurons with degenerating axons and synapses resulting in the dysfunction of skeletal muscles. SMA is one of the leading genetic causes of infant mortality with incidence of 1 in 6000 live births [1]. The SMA disease pathogenesis, diagnosis and severity based classification are dependent on the functional SMN gene and the abundance of its protein product [2, 3]. In humans, duplication of SMN1 gene in telomeric region resulted in SMN2 gene at centromere region located on chromosome 5. SMA patients are typically diagnosed with deleted SMN1 gene copy thus depending on the SMN2 gene translation products for survival that acts as a disease modifier. However, SMN2 gene is known to alternatively spliced due to a single nucleotide C to T mutation resulting in the exclusion of exon 7, (SMN Δ7), such that ~90 % of the SMN protein product is truncated and unstable, thus non-functional. SMN2 gene does, however, produce some full length protein product albeit at low levels. The copy number of SMN2 gene can vary among SMA patients to partially compensate for the loss of SMN. Three or more SMN2 gene copies result in a mild

disease phenotype while four-five copies are sufficient to rescue the phenotype, indicating that small increase in SMN protein dosage is critical for survival [4-6].

Research into therapeutic interventions of SMA has largely focused on increasing the SMN protein by targeted gene therapy using antisense oligos [7] and stem cell therapy [8]. However, unavailability of efficient delivery mechanisms without side-effects presents great challenges to these approaches. In this study, we have investigated the possibility of increasing the stability of full-length SMN protein. SMN protein is known to be degraded by Ubiquitin Proteasome System (UPS) and ubiquitination, a cascade of enzymatic reactions that ubiquitinate substrates post-translationally in the presence of E1 (ubiquitin-activating), E2 (ubiquitin-conjugating) and E3 (ubiquitin-ligase). However, the details of ubiquitination mechanism are not known. Also, the specific E3 ubiquitin ligase for SMN protein ubiquitination is not known. The use of generic proteasome inhibitor, MG132 and bortezomib have been shown to increase the SMN levels in patient fibroblast cell lines and embryonic stem cell derived motor neurons in a drug screening experiment [9-12]. However, the nonspecific inhibition of proteasome is not ideal and may perturb cellular homeostasis due to various off-target effects. Hence, it is critical to identify specific E3 ligase resulting in recognition and modification of SMN protein by ubiquitination, which can potentially be inhibited in cells and with fewer ramifications on non SMN functions. Recently, it was reported that the overexpression of an E3 ubiquitin ligase Mib1 in cell culture increases the ubiquitination of the SMN Δ7 protein to a larger extent than full length functional SMN protein and knockdown of Mib1 improves the pharyngeal pumping defect in *C. elegans* [12]. As the sequence in the exon 7 is critical for SMN protein transport down the axons [13] in motor neurons and also motor axon outgrowth [14], along with the fact that full length SMN protein determines the normal functionality of neurons; I sought to investigate other E3 ligases that may target full length functional form of SMN protein for degradation.

In this study, we observed that SMN protein interacts with APC/C using reciprocal co-IP experiments followed by MS and WB analysis. The APC/C is a well characterized multi-protein complex that acts as E3 ubiquitin ligase and is active in the cell-cycle and in post-mitotic neurons. In the cell cycle, this complex ubiquitinates over 50 proteins including SECURIN and CYCLIN-B, priming those for UPS mediated degradation [15, 16]. In neurons, APC/C substrates

have been identified to play roles in neuronal survival, dendritic morphogenesis and synaptic plasticity [17-19]. Of particular note was the observation that loss of function mutants of APC/C in *Drosophila* and *C. elegans* resulted in neuromuscular junction defects [20, 21]. Taken together, our co-IP data and the literature evidence warranted further investigation into the role of APC/C activity to regulate SMN proteins stability and if this interaction has a relevance to SMA disease.

EXPERIMENTAL PROCEDURES

Animals

All experimental procedures were performed in compliance with animal protocols approved by the IACUC at Children's Hospital Boston, Boston, MA.

MS and WB analyses of APC/C and SMN Co-IP experiments

E18 mouse spinal cord and brain tissues were dissected and lysed using bead beater instrument (Precelleys) for co-IP. Briefly, beads were conjugated with antibodies against APC/C^{cdc27} (AF38.1, mouse, Santa Cruz), SMN (mouse, BD Biosciences) and IgG (mouse, Santa Cruz) individually for 2 h at 4 ^{o}C with gentle rotation. The beads were then washed three times in PBST buffer and three times in PBST buffer containing either 150 mM KCl. The supernatant was then pre-cleared with protein A beads (Bio-Rad) for 2 h at 4 ^{o}C with constant rotation. The beads coated with primary antibody or non-specific IgG were incubated with the pre-cleared lysates overnight at 4 ^{o}C with constant rotation and were collected by centrifugation at 2,000 rpm for 15 min. at 4 ^{o}C. For WB, the protein was eluted by boiling the beads in 80 µl of Laemmli Sample Buffer (BioRad). The lysates were run on a 4-12 % MES Tris Gel (NuPage) and WB was performed using antibodies against either APC^{cdc27} or SMN.

For Mass Spectrometry (MS), to obtain high sequence coverage, the entire immunoprecipitated complex was run on an SDS-PAGE gel and stained with coomassie blue. The gel was sliced at different molecular weight intervals and peptides from each slice were extracted, reduced with DTT, alkylated with iodoacetamide and digested with trypsin before running in QE instrument for 60 min. gradient per fraction to acquire the base peak chromatogram intensity of e^{9} - e^{10}. The

raw files were converted into .mgf files and searched using mouse proteome database (Uniprot Feb 2012, (canonical and isoform sequences) protein sequence database, containing common contaminations and concatenated to its decoy version, using MASCOT v2.3 (Matrix Science) and ProteinPilot (ABSciex) software.

***in vitro* degradation assays**

The 1-2 µl of ^{35}S methionine labeled protein were incubated with 20 µl HelaS3 extracts enriched for G1 phase, 2 µl expressed CDH1, 1 µl of 1µg/µl E1 (Boston Biochem), 1 µl of 1 µg/µl UbcH10 (Boston Biochem), 1 µl of 1µg/µl UbcH5 (Boston Biochem), 1 µl of 1 µg/µl Ube2S (Boston Biochem), 0.1 µg/ml cycloheximide, 0.1 µg/ml ubiquitin, energy regenerating mix containing 20 mM ATP, 150 mM creatine phosphate, 2 mM EGTA pH 8.0 for a control reaction. For the reactions in presence of APC inhibitors either added 5 µl of 0.5 µg/µl Ni-NTA purified His tag-EMI1 or 5 µl of 0.5 µg/µl of Ni-NTA purified His tag-SECURIN or 12 µM TAME instead of same volume of PBS in control reaction. The reactions were incubated at 30 ºC for time points and stopped using Laemmli buffer (Invitrogen) containing β-mercaptoethanol followed by heating at 70 ºC for 10 min total. The samples were resolved on 4-12 % SDS-PAGE. Gels were heat dried under vacuum, exposed to Phosphor image screen (Molecular dynamics) and scanned by Typhoon scanner (Amersham Biosciences).

***in vitro* ubiquitination assays**

All the constructs were cloned in pEU-E01-MCS vector containing tag free sp6 promoter (cell free sciences) except SMN del7 and SMN S4D that were cloned in T7 promoter vector. The ^{35}S methionine-labeled proteins were expressed using above constructs in TNT coupled reticulocyte lysate *in vitro* transcription and translation system (Promega). The APC complex was immunoprecipitated using specific antibody on sepharose beads and incubated with expressed CDH1 and reagents similar to described *in vitro* degradation assays. In addition, ub-aldehyde and MG132 were used in these reactions as an inhibitor of de-ubiquitinating enzymes and proteasome inhibitor respectively. The reactions were incubated at 30 ^{o}C and samples were resolved on 4-12 % SDS-PAGE. Gels were heat dried under vacuum, exposed to Phosphor image screen and scanned by Typhoon scanner.

Inhibition of APC/C activity using EMI1 competitive inhibitor

The EMI1 viruses were used to infect HEK293T cells and to analyze levels of CYCLIN-B using WB. The embryonic stage of E18 from wild type mouse was used to dissect cortical cultures and infected with EMI1 viruses at day 1 with lentiviruses that contained either EMI1 or the empty control CMV promoter. The cells were observed under microscope for GFP+ cells for effective transfections and collected at day 9 for WB of tubulin and SMN.

Inhibition of APC/C activity by small molecule inhibitor proTAME

The embryonic stage of E18 from wild type mouse was used to dissect cortical cultures. The cells were treated with 6 μM and 12 μM proTAME for 24 h and collected for WB of tubulin and SMN.

For WB, cells were collected in sample Buffer containing β-mercaptoethanol. The lysates were run on a 4-12 % SDS-PAGE gel.

Human fibroblasts culture and proTAME treatment

Human fibroblast carrier cell line (3814, mother, Coriell cell repositories) and SMA cell line (3813, Coriell cell repositories) were obtained and cultured in DMEM medium containing 1 % penicillin-streptomycin, 1 % glutamine, 10 % fetal bovine serum. The cells were treated with either DMSO or proTAME and used for WB analysis.

Embryonic stem cell derived motor neurons and proTAME treatment

P3 mouse embryo irradiated fibroblasts (MEF, Global stem cell) were used as a feeder layer and grown on 0.1 % gelatin coating in DMEM media supplemented with 10 % FBS, 1 % glutamine and 1 % penicillin-streptomycin. The WT, GFP containing, embryonic stem cells [22] were cultured on MEFs layer in embryomax DMEM media (Millipore) supplemented with 15 % FBS (Hyclone ES grade, Thermo fischer scientific), 1 % glutamine (Invitrogen) and 1 % penicillin-streptomycin (Invitrogen), 1 % embryomax non-essential amino acids (Millipore), 1 % nucleosides, 0.1 % 2-mercaptoethanol (1000x stock, Sigma) and 0.01 % LIF (ESGRO, Miilipore). The confluent plates of ESC were treated with trypsin, centrifuged at 800 rpm for 5 min for splitting. To make embryoid bodies, the ESCs were plated in density of 2×10^6 on 15 cm

plate in embryomax DMEM/ / F12 media (Millipore) supplemented with 50 % neurobasal medium (Invitrogen), 10 % knock out serum (Invitrogen), 1 % glutamine (Invitrogen) and 1 % penicillin-streptomycin (Invitrogen) and 100 μM β-mercaptoethanol (Sigma). At day 2, embryoid bodies were detached and collected in above medium containing 100 nM retinoic acid (Sigma) and 100 nM 1 μM hedgehog agonist 1.3 (Millipore) on treated plates to remove any MEFs. At day 3, embryoid bodies were collected and treated with above medium containing 100 nM retinoic acid and 1 μM hedgehog agonist 1.3. At day 6, embryoid bodies were dissociated with papain (Worthington biochemical), treated with DNAase (Worthington biochemical) and albumin-ovomucoid inhibitor (Worthington biochemical) mix, filtered using 40 μ filter to remove clumps. The dissociated cells were counted for GFP^+ cells using haemocytometer and plated in density of 20,000 cells per well in 96 well plate coated with 25 μg/ml polyornithine (Sigma-Aldrich). The cells were cultured in DMEM/F12 medium containing 0.3 5 BSA, 2 % fetal bovine serum (Hyclone ES grade, Thermo fischer scientific), 2 mM glutamine, 1% penicillin-streptomycin, 0.1 mg/ml apo-transferrin (Sigma), 16 μg/ml putrescine, 160 ng/ml sodium selenite, 10 μg/ml insulin, 60 ng/ml progesterone, 2 % B27 supplement, 10 μg/ml insulin, 10 ng/ml recombinant human brain-derived neurotrophic factor (Peprotech), 10 ng/ml recombinant rat cilliary neurotrophic factor (Peprotech), 10 ng/ml recombinant rat glial-derived neurotrophic factor (Peprotech).

The cells were treated with DMSO and proTAME at different concentration of 12 μM and 15 μM for 48 h. The WB was performed using above described methods. For in cell western, the cells were fixed with 4 % PFA for 20 min, washed three times with PBS, blocked in normal goat serum with 0.1 % triton X-100, stained with primary SMN antibody and secondary antibody Alexa red and Hoechst for nuclei staining. The images were acquired at scanning microscope and images were analyzed using volocity software v6.3 (Perkin Elmer) to quantitate SMN protein. The images were analyzed in 10 frames / well using automation for high content imaging. The SMN intensity was measured in comparison to GFP intensity in GFP^+ cells and normalized with DMSO treated cells. The quantitative values were plotted in graphs using prism software.

Neurites outgrowth in Primary mouse motor neuron cultures

Wild type HB9-GFP pregnant mice were dissected at the E12.5 embryonic stage. The spinal cords form embryos were cut into sections and collected in HBSS (with phenol red, Invitrogen) on ice. The sections were washed with HBSS, treated with 0.1 % trypsin in HBSS for 10 min at 37 ºC. The tissues were gently triturated until fragmentation in complete L15 medium (containing 3.6 mg/ml glucose, 1 % penicillin-streptomycin (Invitrogen), 6.25 ng/ml progesterone (Sigma-Aldrich), 5 μg/ml insulin (Sigma-Aldrich), 16 μg/ml putrescine (Sigma-Aldrich), 5.1 ng/ml sodium selenite (Sigma-Aldrich), 2 % horse serum (Invitrogen), 10 % heat inactivated filter-sterile FBS and 100 μg/ml DNAase (Worthington Biochemical). The dissociated cells were collected after trituration and until cell suspension was homogenous. The cell suspension was centrifuge over a 4 % BSA (dialyzed with L15 media) cushion at 600 g for 3 min. Pellet was re-suspended in complete L15 medium and centrifuge over 6 % optiprep density gradient solution (diluted in L15 medium) at 800 g for 10 min. The white interface enriched of motor neuron population (about 70 %) was harvested and centrifuge over a 4 % BSA cushion at 600 g for 3 min. The pellet was re-suspended in complete neurobasal medium (Invitrogen) containing 2 % B27 supplement (Invitrogen), 0.5 mM glutamine (Invitrogen), 0.04 % 2-mercaptoethanol (Sigma), 2 % heat inactivated horse serum (Invitrogen), 10 ng/ml recombinant human brain-derived neurotrophic factor (Peprotech), 10 ng/ml recombinant rat cilliary neurotrophic factor (Peprotech), 10 ng/ml recombinant rat glial-derived neurotrophic factor (Peprotech). Cells were counted using haemocytometer and plated in 24 well plates (Nunc). The coverslips were washed in ethanol, flame sterilized, coated with 3 μg/ml poly-ornithine in water and 2 μg/ml laminin in PBS overnight at room temperature, washed three times with HBSS and once with Neurobasal medium.

Cells were plated on to coverslips at a density of 2000-4000 cells per well and cultured for 1-2 days until a neurites start to grow; and treated with DMSO or proTAME at 12 μM every day and incubated for 48 h in triplicates and three different experiments. The cells were fixed in 4 % PFA and 1.5 % sucrose for 10 min, washed three times with PBS for 10 min, blocked in 3 % BSA with 0.5 % triton-X100 for 2 h. Mouse anti-neurofilament (Hybridoma bank) or mouse anti-SMN (BD Biosciences) and goat anti-CHAT (Millipore) overnight at dilution 1:100 were used to treat cells. The samples were washed three times with PBS and treated with secondary antibody

Alexa mouse red, Alexa goat far red for 2 h at dilution 1:400; and mounted on slides with Dapi fluoromount G (Southern biotech). The cells were observed 80i microscope (Nikon) at 10 x magnification and analyzed with NeuronJ plugin appended to ImageJ software to trace outgrowth of neurites from cell body. The lengths were scored and plotted using prism software.

Overexpression of EMI1 in the zebrafish SMA model

Human EMI1 cDNA was subcloned with GFP, HB9 promoter (3.5 kb) and HSP70 promoter (1.5 kb) in pbluescript vector to investigate the effects of APC inhibition in a targeted manner in the motor neurons. However, that resulted in a low expression of GFP protein that was suboptimal for further analysis in this assay. Separately, human EMI1-GFP gene in $pCS2^{+}$ vector (a gift from Dr Petros Marangos) was used to generate mRNAs with polyA tail. The plasmids were sequence verified and tested for expression using wheat germ expression system and WB to validate expression. The overexpression of EMI1 from mRNAs injections did not contribute to cytotoxicity. On the contrary, in co-injections with SMN MO, overexpression of EMI1 resulted in surprisingly healthy phenotype of fish.

Briefly, the plasmid was linearized with Not1 and capped RNA was generated using the sp6 mMESSAGE mMACHINE kit (Ambion). For EMI1 overexpression, One or two cell-stage Tg (hb9:GFP) embryos were injected with ~9 ng of control MO (morpholino), 9 ng of SMN MO or 9 ng of SMN MO with 600 ng GFP-EMI1 RNA individually using an MPPI-2 pressure injector (Applied scientific instrumentation). Fifty embryos per experiment were analyzed from three separate experiments at 28 h post fertilization. The live embryos were anesthetized with tricaine, mounted on glass coverslip and observed with Zeiss axio-plan2 microscope. Motor neurons were scored as described previously [14]. Briefly, motor neurons innervating the mid-trunk myotomes 7-16 were examined under the microscope. Normal motor axon morphology is stereotyped along the trunk targeting myotomes while decreasing SMN levels result in the motor axon truncations and abnormal branching. The abnormal branching is characterized as the extra branches along the entire length of the motor axon or excessive branching at the distal region of the motor axon. The fish were classified as having severe, moderate, mild, or no defects based on the character of the motor axon defects. Fish with at least 20 % severe axon defects (i.e. truncations or truncation with abnormal branching) or 40 % moderate defects (i.e. abnormal branching without

truncations, “wishboned” axons) were classified as severe. Fish classified as moderate typically shows 10 % severe defects, 20-40 % moderate defects, or >40 % mild defects (i.e. axon lacking stereotyped morphology, but not abnormally branched). Fish classified as mild shows 10 % moderate defects and 20-40 % mild defects. Each side of the fish was scored for total 20 motor axons and this classification was used from the combined motor neuron defects. Data was analyzed and statistical significance was determined using the mann-whitney nonparametric rank test.

For WB, 20 embryos per treatment group were collected from three independent experiments at 2 days post fertilization. Samples were generated by boiling in 60 µl buffer (63 mM Tris pH 6.8, 5mM EDTA, 10 % SDS). The total volume of 10 µl, equivalent to 3 embryos, was mixed with equal volume of loading buffer (100 mM Tris pH 6.8, 0.2 % phenol-blue, 20 % glycerol, 200 mM DTT). The samples were run on a 10 % SDS PAGE gel, probed with mouse anti-SMN antibody, re-probed after stripping with mouse anti-HuD (Santa Cruz Biotechnology) and re-probed after stripping with mouse anti-β-actin (Sigma) and detected by chemiluminescence of bound HRP-conjugated mouse antibody.

RESULTS

In an independent investigation of SMN interactome, our lab identified various members of the APC/C complex (data unpublished). In this study, I was interested in determining if the reciprocal analysis of APC/C interactome would identify the SMN protein or its associated proteins complex in neurons. Also, one way to determine the identity of substrates of an E3 ubiquitin ligase is to study its interactome [23]. Therefore, a co-IP of the APC/C complex was performed using antibody against one of its subunit cdc27 in the developing neurons from mouse embryonic E18 brain and spinal cord tissue and was analyzed using mass spectrometry based proteomics. Co-IP experiments were performed using high stringency conditions to isolate tightly bound complex and analyzed using a rigorous 1 % false discovery rate (FDR) cutoff for the database searches to ensure a minimal false positive rate in the protein identification process. IgG controls were used to eliminate non-specific interactions and to exclude common proteins from the list of protein hits established from the target IPs. The LC-MS/MS analysis identified several subunits of APC/C complex including the cullin-based scaffold protein APC2, APC5 and

CDC27. Other interacting partners included various ribosomal subunits, splicing proteins, proteins of SMN complex and various RNA binding proteins including HuD. A table **(Table 1)** is given showing some of the subunits of APC/C complex, SMN complex including Gemins, SMN-d, Sm proteins, various snRNPs, proteins associated with splicing etc. (shown in blue highlights) along with number of identified peptides. APC/C IP also identified interacting Hu proteins including HuD (shown in pink highlights). Moreover, in a similar co-IP experiments from HeLa cell cultures from S and M phases of cell cycle independently showed SMN and HuR (a non-neuronal pan isoform of HuD) in SMN interactome. This table also highlights (shown in green) some of the known substrates of APC/C from neurons and cell-cycle that were found to be interacting in these co-IP experiments.

Mouse embryonic brain and spinal cord	**No. of peptides**	**HeLa cells (mitosis)**	**No. of peptides**	**HeLa cells (S phase)**	**No. of peptides**
CDC27	53	CDC27	117	CDC27	139
APC1	81	APC1	133	APC1	156
APC2	28	APC16	2	APC16	1
APC4	39	APC4	54	APC4	65
APC5	71	APC5	91	APC5	85
APC7	32	APC7	106	APC7	116
APC10	5	APC10	16	APC10	26
CDC16	34	CDC16	71	CDC16	75
CDC23	34	CDC23	71	CDC23	74
CDH1	11	CDH1	10	CDH1	13
SMN splicing factor 30	1	SMN	76	SMN	77
Splicing factor 3B	25	Splicing factor 3B	4	Splicing factor 3B	3
snRNPN	6	GEMIN2	58	GEMIN2	48
GEMIN2	-	GEMIN4	180	GEMIN4	182
GEMIN4	2	GEMIN5	13	GEMIN5	26
GEMIN5	4	GEMIN6	10	GEMIN6	9
Sm protein D1	1	GEMIN7	9	GEMIN7	9
Sm protein D2	4	GEMIN8	14	GEMIN8	14
Sm protein D3	3	Sm protein D1	17	Sm protein D1	17
PLASTIN 3	2	Sm protein D2	22	Sm protein D2	21
GSK3β	7	Sm protein D3	6	Sm protein D3	7
HuR	14	HuR	2	HuR	1
HuD	12	KIF22	3	KIF22	1
HuB	13	MAP7	1	MAP7	0

HuC	12	AURK8	5	AURK8	2
LIPRIN-α	16	DNMT1	1	DNMT1	0
		CDC20	25	CDC20	15
		PLK1	3	PLK1	0
		PRC1	1	PRC1	0
		TPX2	2	TPX2	0
		BUB1B	25		18

Table 1: LC-MS/MS based interactome of APC/C. Table of protein hits with identified number of peptides showing APC/C^{cdc27} interacting proteins in embryonic nervous system and HeLa cells at mitosis and S phase. This list includes identified subunits of APC/C complex, known SMN complex and interacting proteins (blue highlight) including Hu proteins (pink highlight) along with known APC/C substrates (green highlight).

We then used WB analyses as an orthogonal approach to mass spectrometry to validate the presence of specific proteins in reciprocal co-IP experiments. Co-IP experiments were performed using antibodies against SMN and APC/C to isolate endogenous complexes from mouse embryonic spinal cord and brain tissue. WB analyses of the endogenous IP complexes from above experiments detected the presence of SMN and APC/C^{cdc27} proteins in reciprocal IPs **(Figure 1)** suggesting that these proteins are interacting endogenously in these tissues.

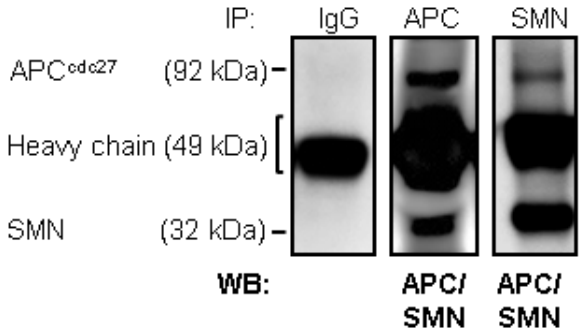
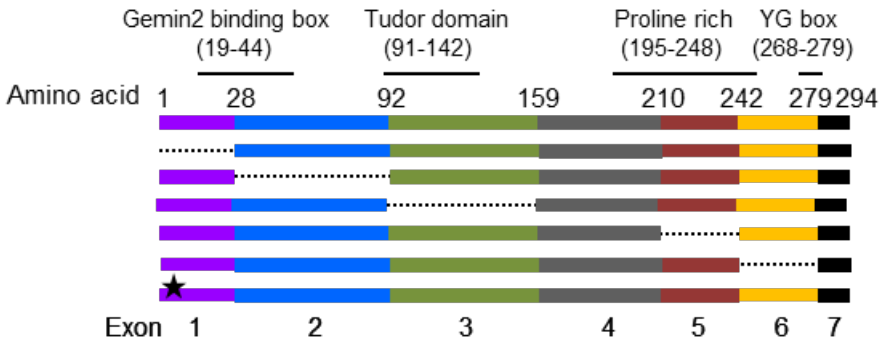
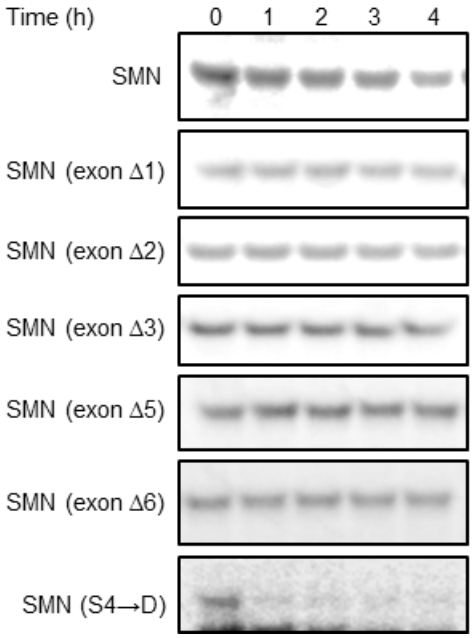


Figure 1: APC/C interacts with HuD and SMN. WB analyses of co-IP experiments from the E18 mouse spinal cord and brain using antibodies directed against APC/C^{cdc27}and SMN. These reciprocal co-IP experiments are annotated above the lanes and antibodies used for the detection of specific proteins are indicated below the image.

The above observation of the APC/C and SMN interaction in nervous tissue prompted us to investigate the role of the APC/C in regulating the SMN protein. The APC/C is an E3 ubiquitin ligase that promotes protein degradation by directly interacting and catalyzing the ubiquitination of target molecules in the presence of co-activators such as CDH1 and CDC20 [24]. To investigate if SMN is regulated by the action of APC/C E3 ubiquitin ligase, we performed *in vitro* degradation assays. Briefly, lysates are made from HeLa cells synchronized in the G1 phase

for enzymatically active fractions of APC/C. These lysates were first tested for APC/C activity using SECURIN or KIFC1 proteins as well characterized substrates of the APC/C [25]. KIFC1 was used as control in assays probing the potential neuronal substrates of APC/C, which may have unique profiles of degradation over time. The specificity of the APC/C dependent degradation observed in these assays was tested by the addition of recombinantly expressed and purified inhibitors of the APC/C such as SECURIN (a substrate and a competitive inhibitor of APC/C) [25] or EMI1 (an endogenous inhibitor of APC/C) [26]**.** In addition, TAME a small molecule inhibitor of the APC/C can be used to show the specific inhibition of the APC/C [27]. The results from these experiments suggest that SMN protein is reproducibly degraded at later time points than 2 h in comparison to KIFC1 control (**Figure 2A, 2B**). However, this degradation is APC/C dependent and inhibited in the presence of APC/C inhibitors. We also performed degradation assays of SMN Δ7 to see if this protein might also be degraded by the APC/C. SMN Δ7 was extremely unstable in the degradation assay, however the degradation appeared to be independent of the APC/C (**Figure 2B**). There are usually tight regulatory mechanisms that govern the degradation of substrates by E3 ubiquitin ligases and the regulation often involves specific motifs and/or post-translational control of the degradation event. APC/C targets often carry a destruction motif called D-box (RXXLXXN/D/E), KEN-box (KENXXXN) [28], however other APC/C recognition motifs have been described, including the A-box [29] and O-box [30] or CRY box [31]. These motifs are neither necessary nor sufficient for APC/C dependent protein turnover as the process can be also regulated by the presence of post-translational modifications (PTMs) e.g. phosphorylation. While SMN does not contain any of these degradation motifs, the moderate amount of degradation of SMN protein in these assays prompted us to investigate if the modular structure of SMN protein or PTMs may be involved in the degradation process. These results indicate that SMN protein with independent deletion of exon 1, 2, 3, 5 and 6 show no degradation of truncated proteins, suggesting that modular structure and primary sequence of these SMN mutants may not contribute to protein turnover (**Figure 2**).


Gemin2 binding box
(19-44)
Tudor domain
(91-142)
Proline rich
(195-248)
YG box
(268-279)
Amino acid
1
28
92
159
210
242
279
294
Exon
1
2
3
4
5
6
7


Time (h)
0
1
2
3
4
SMN
SMN (exon Δ1)
SMN (exon Δ2)
SMN (exon Δ3)
SMN (exon Δ5)
SMN (exon Δ6)
SMN (S4→D)

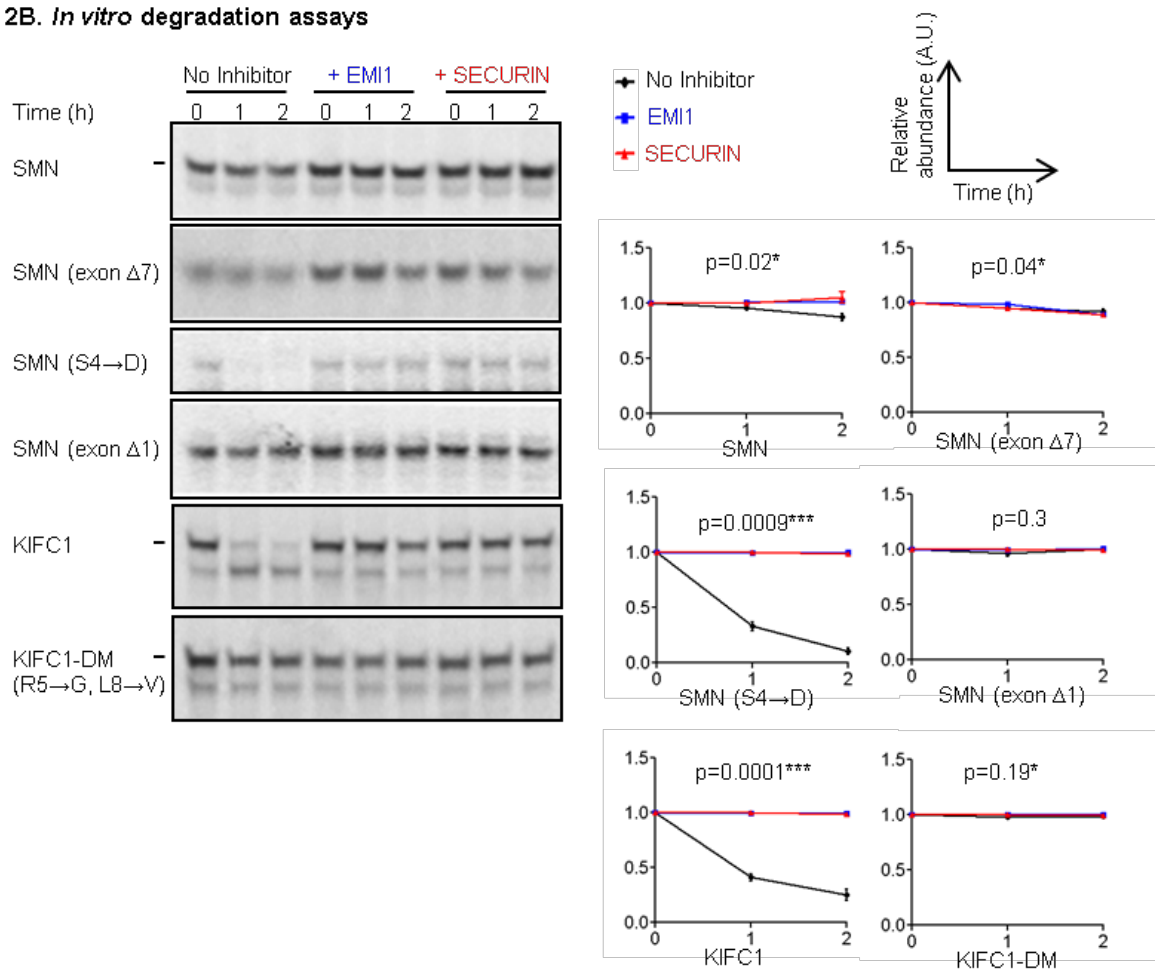


Figure 2: (A). ***In vitro*** **degradation of SMN and mutants.** A representation of SMN protein mutants showing the structural domains with number of amino acids and deleted exons (grey line) in upper panel. Lower panel shows APC/C mediated *in vitro* degradation of sequentially exon-deleted mutants of SMN as SMN Δ1, SMN Δ2, SMN Δ3, SMN Δ5, SMN Δ6 and SMNΔ7 in comparison to full length and SMN-S4D phospho-mimic. SMN-S4D mutation is showed as star on first exon. **(B) APC/C dependent *in vitro* degradation of SMN proteins.** ^{35}S methionine labeled SMN protein shows a slight decrease by 2 h in the APC/C degradation assays, whereas SMN Δ7 is generally unstable while SMN Δ1 is non-degradable. The SMN-S4D phospho-mimic shows remarkable degradation relative to KIFC1 controls. The degradation assays of the substrates is performed in the presence of the APC/C specific inhibitors such as EMI1 and SECURIN to ensure that the degradation is not due to other E3 ubiquitin ligase activities in the lysates. KIFC1 was used as a control to test the APC/C activity in the lysate and degrades within 1 h, while KIFC1-D box mutant (DM) remains stable. Quantitation of bands is measured over time by densitometry shown in graphs for control reaction without inhibitor-black and with inhibitor EMI1-blue, SECURIN-red. The data represents the mean ± SEM of three experiments.

We then investigated the phosphoregulation of ubiquitination in these assays. Previously, we have mapped a GSK3β kinase phosphorylation site, Ser4 at the N-terminus of SMN, which was shown to destabilize SMN protein. It is also known that GSK3β inhibitors may increase the

SMN protein levels from a dug screen study performed in cell cultures [32]. To determine if this phospho-modification is regulated by APC/C, we used a phospho-mimic mutant of SMN (SMN Ser4 to Asp; S4D) in the APC/C degradation assays.

The results indicate that S4D phospho-mimic of SMN was a strikingly effective target of APC/C dependent degradation with faster degradation kinetics in comparison to KIFC1. Moreover, SMN-S4D mutant is stable in the presence of APC/C inhibitors such as EMI1 and the small molecule inhibitor TAME. In contrast, the deletion of this phosphorylation site in SMN mutant (SMN Δ1) protein did not result in protein degradation. Taken together, these experiments indicate that APC/C along with the effect of GSK3β in concert may destabilize SMN protein. The degradation of SMN substrates was quantified over time in presence of APC/C and its inhibitors (n=3) and p values were calculated (**Figure 2)**.

To validate the degradation results and to determine if SMN is indeed ubiquitinated by the APC/C, we performed *in vitro* ubiquitination assays using IP purified complex of enzymatically active APC/C from synchronized HeLa cell lysates [33]. In addition, the reactions were performed in the presence of proteasome inhibitor MG132 and the de-ubiquitinating enzyme inhibitor ub-aldehyde to allow accumulation of modified protein. The *in vitro* expressed and ^{35}S labeled proteins were added to the reaction mixture and samples were analyzed at 0, 20, 40 and 80 minutes separated by molecular weight on an SDS-PAGE gel. These results showed that the full length SMN is directly ubiquitinated by the APC/C at end-point of 80 min as observed by the existence of higher molecular weight bands and smear appearing in reactions over time. SMN Δ7 was slightly ubiquitinated in comparison. In line with degradation analysis of substrates, the SMN ubiquitination was further facilitated in the presence of GSK3β and when assessed for phospho-mimic form of SMN protein (S4D mutant). The phospho-dependent ubiquitination assays also showed the depletion of original protein band at endpoint of 80 min indicating that reaction was in progress for protein modification (**Figure3**). The modified protein was quantified by densitometry of a rectangle selection drawn above the normal size band of unmodified protein (n=3). The values were normalized to the zero time point for graph plots and p-values were calculated **(Figure 3).** These results indicate that SMN protein interacts with

APC/C that ubiquitinates this protein and regulates its abundance in phosphorylation-dependent manner.

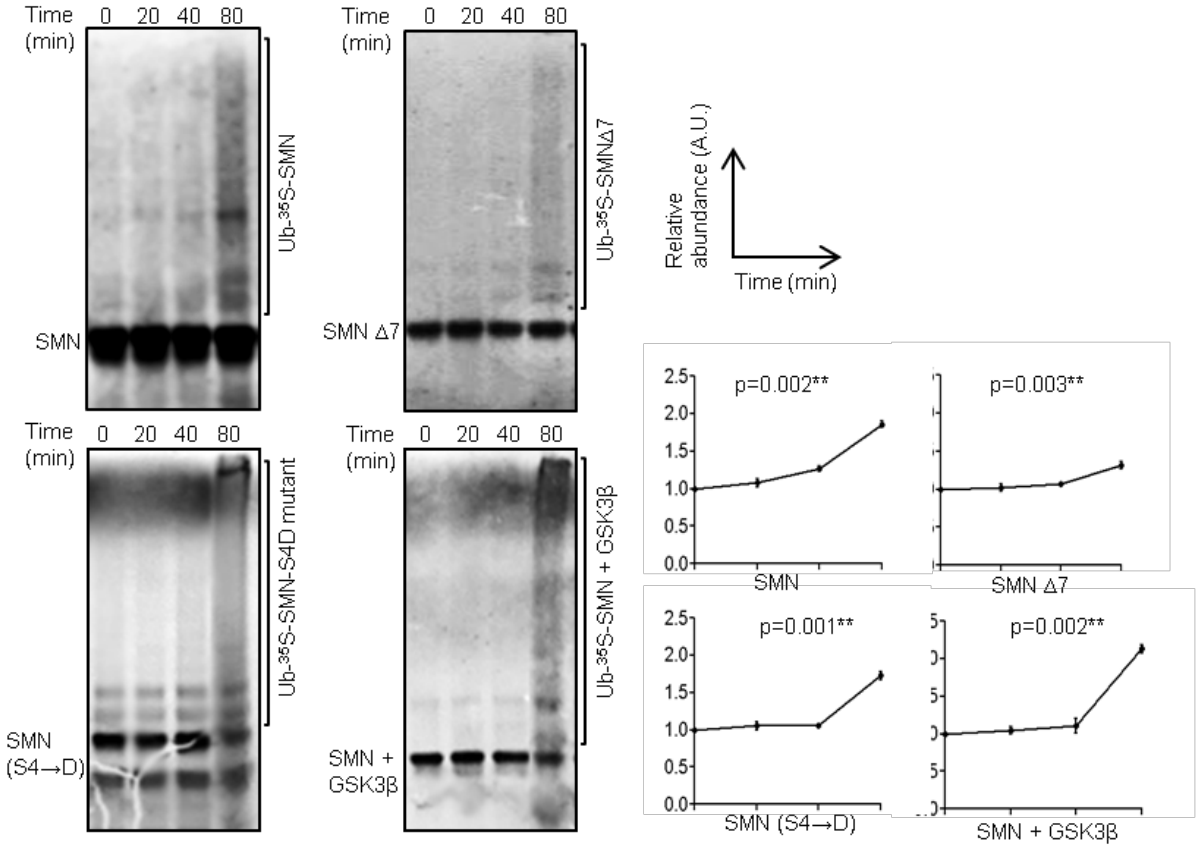


Figure 3: APC/C dependent *in vitro* ubiquitination assays. ^{35}S methionine labeled SMN and SMN Δ7 are differently modified at 80 min. SMN in the presence of GSK3β kinase and SMN-S4D phospho-mimic show modified protein along with depletion of unmodified band. Images were quantified by drawing a rectangle box (area indicated by a bracket on the side of images) above the unmodified protein band. The data represent the mean ± SEM of three experiments.

We followed the APC/C dependent regulation of SMN protein in the physiological context of primary cortical neurons cultures, SMA patient fibroblast cell line and motor neurons differentiated from embryonic stem cells after inhibiting APCC/C using EMI1 overexpression or small molecule drug, proTAME.

First, we cultured the mouse primary cortical neurons from E18 embryonic stage and lentivirally transfected with EMI1 gene construct containing a CMV promoter and GFP reporter. GFP expression was used to observe the efficiency of transfection. The cells were allowed to grow 9

DIV (days *in vitro*) to permit the expression of EMI1. The translation was inhibited using cycloheximide (CHX) for 4 h prior to collecting the samples. The lysates from CMV control transfected and EMI1 expressing cells were collected for WB analysis of SMN protein and TUBULIN as internal control. The results show increased levels of SMN protein upon APC/C inhibition. The results were quantified reproducibly (n=3) using densitometry **(Figure 4B).**

These results were further confirmed using pharmacological mechanism of APC inhibition using a small molecule inhibitor. The proTAME, a pro-drug version of TAME (tosyl-L-arginine methyl ester) molecule, is more readily internalized though cell membrane and is known to induce mitotic arrest in developing cells by direct inhibition of APC/C activity. We tested the optimal concentration of drug molecule to inhibit the APC/C activity in post-mitotic cortical neuron culture and found that a dose of 12-15 μM which is effective in cell cycle is sufficient to increase the SMN protein expression, as analyzed by WB after 24 h. The results were quantified reproducibly (n=3) using densitometry and p values were calculated (**Figures 4C)**. The optimal dose was selected to avoid any cytotoxic effects of overdose as studied by monitoring cell survival and morphology and a decrease in the dose to 6 μM did not result in detectable increase in the levels of SMN proteins. These results confirm that APC/C activity is required for the degradation of these proteins and that inhibition of this activity in neurons leads to an increase in levels of SMN protein.

A well-employed tool in small molecule drug screening for SMA disease is to use WT (mother, carrier) and SMA (son, Type I) fibroblast cell lines to test the efficacy and viability of a potential therapeutics. We cultured these cell lines and treated with 12 μM proTAME (a higher dose of proTAME above 18 μM shows cytoskeleton defects in the fibroblast cultures). The results show a slight increase in SMN levels in both cell types at 24 h as compared to DMSO treated cells as detected by WB **(Figure 4C).**

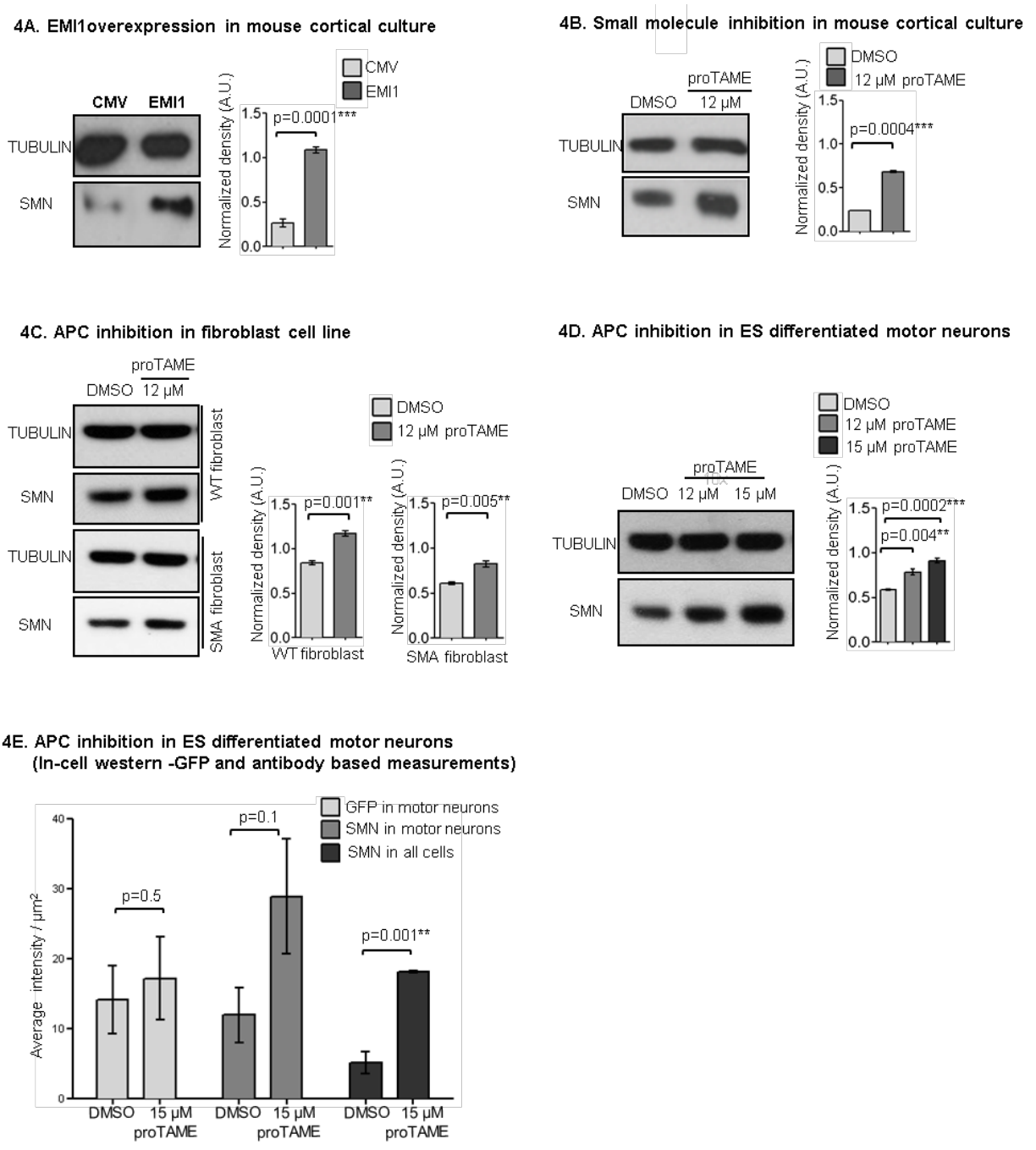


Figure 4: (A). APC/C inhibition by EMI1 overexpression in mouse cortical cultures results in increased SMN. Over-expression of EMI1 using CMV promoter directed lentivirus construct leads to the accumulation SMN protein in cortical neurons. Histogram shows a densitometry analysis of bands (n=3). **(B). APC/C inhibition by proTAME treatment in mouse cortical cultures results in increased SMN.** Pharmacological inhibition of APC/C activity by small molecule proTAME leads to increase in levels of SMN at a dose of 12 µM after 24 h in cultured cortical neurons as compared to DMSO. Histograms show a densitometry analysis of bands (n=3). **(C). APC/C inhibition results in increased SMN in SMA fibroblast culture.** Fibroblast cell line 03813 (SMA patient), 0314 (carrier mother) in culture show an increase in SMN upon proTAME treatment at a dose of 12 µM after 24 h. Histograms show a densitometry analysis of bands normalized values. **(D). APC/C inhibition results in increased SMN**

protein in motor neurons differentiated from embryonic stem cells. The differentiated cells show increase in SMN protein at 12 μM and 15 μM proTAME after 48 h as compared to DMSO by WB. Histograms show a densitometry analysis of bands normalized values. **(E). APC/C inhibition in motor neurons differentiated from embryonic stem cells (GFP positive) by in-cell WB.** The differentiated cells were immunostained for SMN, imaged in 96 well plates, processed by automative analysis for high GFP signal/noise cut-off for motor neurons. Histograms show averaged intensity of GFP and SMN after normalization against area (μm^2). The GFP levels remain same in proTAME treated cells as compared to DMSO while SMN levels increase in GFP-positive motor neurons as well as all cells. The data represents the mean ± SEM of three experiments.

These results were quantitated using densitometry (n=3) and suggest that patient cell line can possibly be altered to increase SMN levels to an extent using a targeted APC/C inhibitor.

Additionally, we used another system where embryonic stem cells from WT mouse were differentiated in to motor neurons (GFP-HB9) with 30 % efficiency as measured by GFP expression. These cells were then treated with 12 μM and 15 μM proTAME for 48 h that reproducibly (n=3) showed an increase in SMN levels as compared to DMSO treated cells in WB. The results were quantitated using densitometry (n=3) and confirm that APC/C inhibition directly effects SMN protein levels in the programmed stem cells **(Figure 4D).** We further investigated this system using in-cell western strategy and cells were cultured in 96 well plates in the biological replicates, treated with either DMSO or proTAME, incubated with SMN antibody and subjected to microscopy. The scanning microscope was used with automated high-throughput imaging that collected images at 10 random positions per well and analyzed by image analysis software Volocity v6.3. For quantification, a score cut off for GFP intensity was determined based on high signal/noise ratio and applied globally to all images. The output above the cut-off was defined as GFP positive cells, therefore motor neurons and measurements below the cut-off were selected as non-GFP positive cells, non-motor neurons. The intensities of all SMN channel measurements were normalized against area (μm^2) and resultant ratios from proTAME treated cells were compared to the DMSO. The average intensity of SMN was plotted along with GFP intensities. In comparison to the GFP intensity, measured over hundreds of observation (n=3), the SMN levels were increased in motor neurons as Ill as in all cells; as shown in figure at a dose of 15 μM proTAME **(Figure 4E).** Taken together, these results suggest that APC inhibition may be a potential modifier of SMN protein in neurons and may be further investigated for therapeutics in SMA disease. Also, these results indicate to the efficacy of

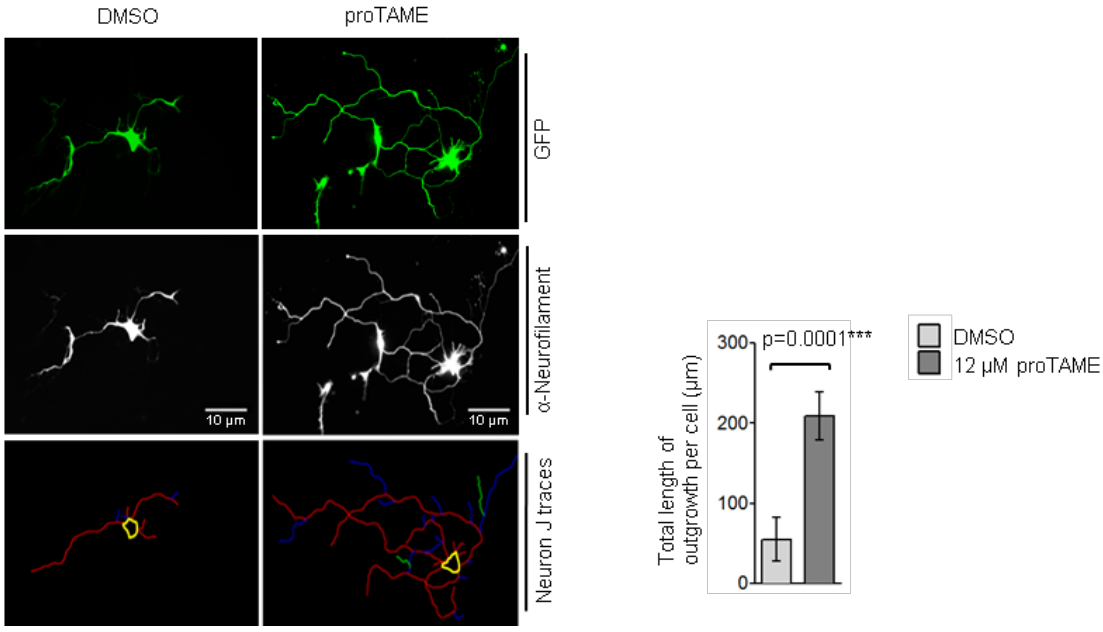
proTAME as a small molecule inhibitor of APC/C, its efficiency and optimal dose may vary depending of cell-types.

In order to test the effect of APC/C inhibition on WT mouse motor neurons from spinal cord, I dissected and cultured the mouse primary motor neurons containing GFP expression from spinal cords of embryonic stage E12.5 [34] and treated with either DMSO or 12 μM proTAME at 1 DIV for 24 h. Due to the limitation of total protein amount obtained from these sensitive cultures, I opted to investigate the morphological effects of APC inhibition on the length of neurites [19] in primary motor neurons. The cells were stained with neurofilament antibody and used to measure the combined length including secondary and tertiary branching. The imageJ v1.47 plugin NeuronJ was used to trace the total length of neurites outgrowth from the cell body of cells. These cells displayed an increase in the total length from neurons treated with proTAME as compared to DMSO **(Figure 5A).** The cells were observed for n=3 biological replicates and measurements were plotted in graph for total 30 cells, 15 random cells each per condition, to show the length differences. These results suggest that inhibition of APC/C E3 ligase has a potential to regulate length of neurites in mouse motor neurons.

The results from these experiments showed a remarkable rescue of disease phenotype with extended and targeted motor axons as explained below. The SMA morphants were quantified in percentage values of severe 28.43 ± 3.4, moderate 38.7 ± 2.4, mild 22.5 ± 4.7, no defect 10.23 ± 5.7 phenotypes. After injection of EMI1 RNA at a low dosage of 250 ng the morphants were quantified as percentage values of severe 0 ± 0, moderate 14.5 ± 1.5, mild 43.4 ± 9.2, no defect 42.1 ± 9.5 percent phenotypes while injection of EMI1 RNA at a dose of 600 ng resulted in better rescue and percentage values of severe 1.3 ± 1.3, moderate 10.3 ± 4.6, mild 24.7 ± 3.4, no defect 63.8 ± 8.8 percent phenotypes. Notably, no side-effects such as apoptosis or cytotoxicity were observed at these doses (**Figure 5B**). To establish that these morphants indeed have stabilized levels of HuD and SMN proteins resulting in the rescue of phenotype, the end-points were collected to detect HuD and SMN protein levels by WB.

This experiment led us to observe the effects of APC inhibition in motor neurons in zebrafish model of SMA disease. As discussed in method section, the embryos were co-injected with SMN MO along with EMI1 construct for overexpression starting with a smaller dose to carefully rule out any apoptotic or cytoskeleton defects.

5A. Neurites length in mouse primary motor neuron culture



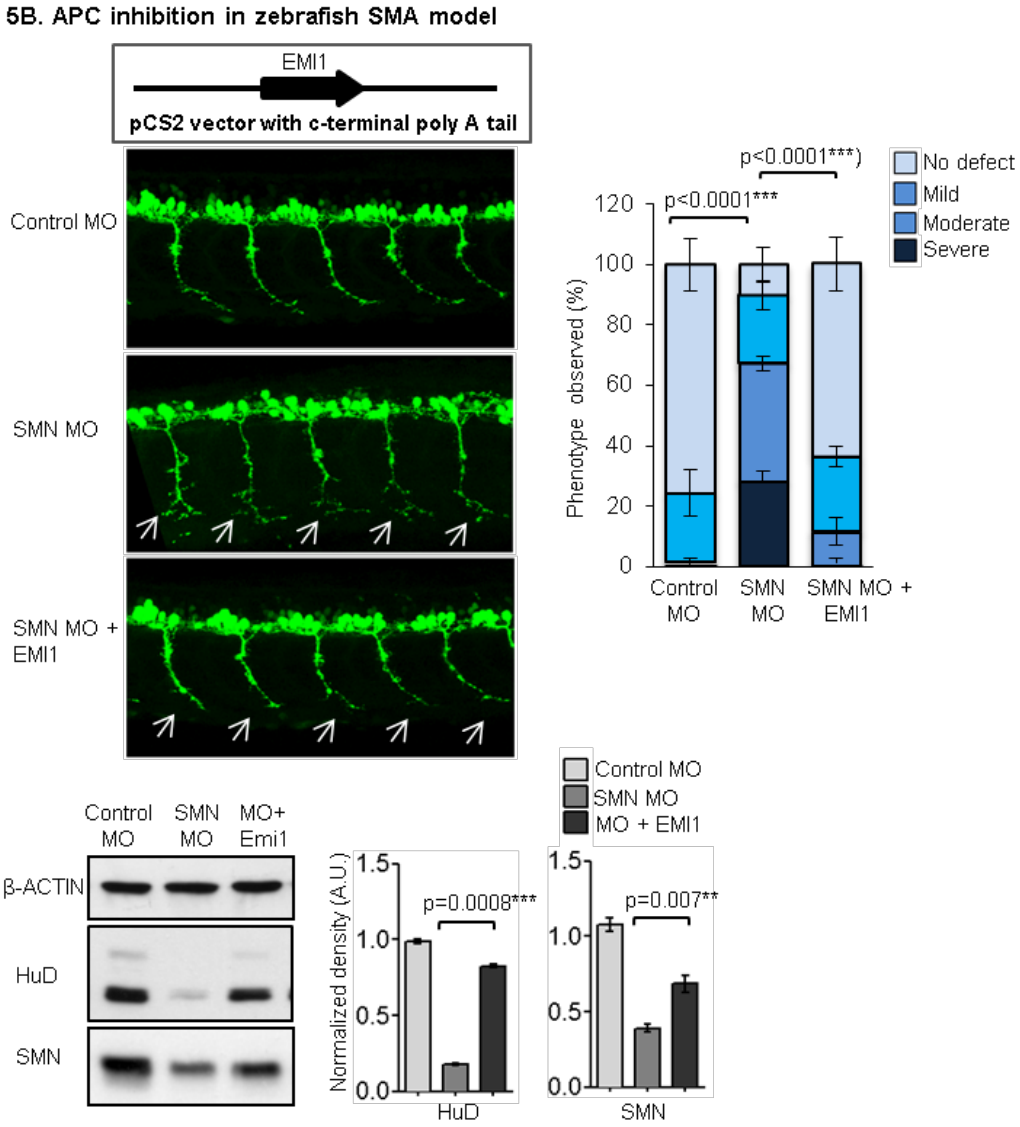


Figure 5: (A) APC/C inhibition leads to extended neurites outgrowth in motor neurons. Mouse wild type primary motor neuron cultures treated with proTAME results in the increased total length of neurites as compare to DMSO as shown in a representative image. The top panel is GFP-positive cell of motor neuron, middle panel is neurofilament staining and bottom panel is traces of neurites by Neuron J. The total length of is measured for outgrowth from cell body including branching. The phenotype was observed in three experiments and graphs were plotted for random n=15 cells to show mean ± SEM length of three experiments. **(B) APC/C inhibition leads to rescue of SMA motor neurons phenotype in zebrafish.** Overexpression of APC/C inhibitor EMI1 in zebrafish SMA morphants results in rescued phenotype of characteristic truncations and aberrated branches in motor neurons and better path-finding shown as representative image. The dose effect of EMI1 tested with two different amounts show remarkable rescue of disease phenotype with no side defects of cytoskeleton and cytotoxicity. The quantitative rescue of morphants is classified in to severe, mild, moderate and no defect categories as shown in the graph. The data represent the mean ± SEM of three experiments. WB of end-point morphants shows an increase in SMN and HuD protein levels in the presence of APC/C inhibitor as compared to SMA background of SMN-morpholino. The data represents the mean ± SEM of three experiments. Upper panel shows construct design for this experiment. Arrows indicate the differences in phenotype upon rescue.

These observations provide further evidence that APC/C inhibition regulates morphology of motor neurons and serves as direct regulator of HuD and SMN proteins. These observations factor in the importance of APC/C, or its inhibition, in regulating the SMA disease phenotype.

Finally, I present a model based on our results explaining the mechanisms of APC/C mediated regulation of SMA phenotype. Depending on our observations, phosphorylated SMN protein due to the activity of GSK3β kinase is an amenable target for ubiquitination and degradation by active APC/C E3 ligase. Similarly, HuD is a novel and neuron-specific substrate of active APC/C that interacts with SMN protein closely in a stabilizing complex. These results have been discussed in previous chapter 4. HuD interactions with SMN proteins to regulate various downstream RNAs, and in concert with APC/C may play a key role in SMA disease physiology. Absence of APC/C activity conclusively checks the degradation of HuD protein as well as SMN protein and has a potential to be modifier for SMA disease.

6. Model depicting the APC/C regulatory mechanism of SMN and HuD stability

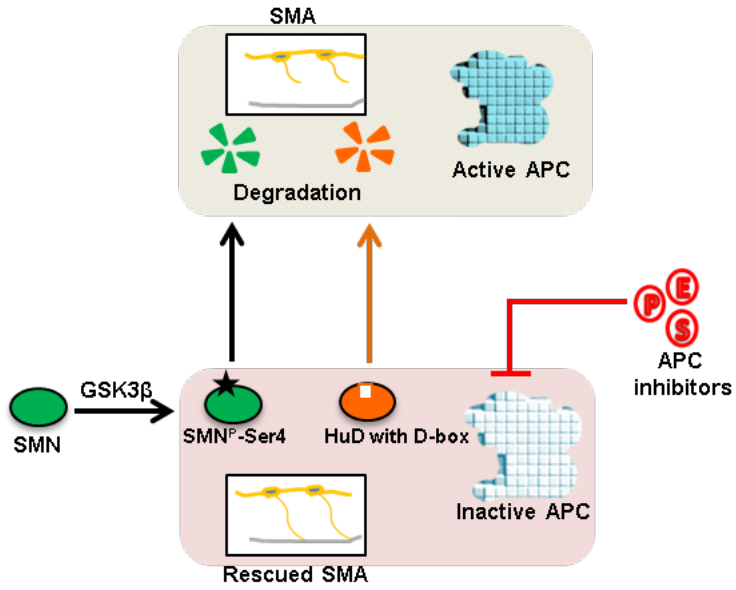


Figure 6: A model to explain the mechanism of APC/C mediated regulation of SMA phenotype. Depending on our observations from this study, phosphorylated SMN protein in the presence of GSK3β kinase is an amenable target for ubiquitination and proteasome degradation by APC/C as E3 ligase. Similarly, HuD is also a novel and neuron-specific substrate of APC/C and interacts with SMN closely in a stabilizing complex and may be playing a role in SMA disease physiology. Absence of APC/C or inhibition of its activity conclusively checks the degradation of HuD, and more importantly, SMN protein. Furthermore, APC/C inhibition has potential to be a modifier for SMA disease.

DISCUSSION

In this study, I investigated a novel mechanism of SMN protein regulation by APC/C, an E3 ubiquitin ligase, using *in vitro* biochemistry assays, primary neuron cultures and *in vivo* rescue of a zebrafish model of SMA. APC/C has been the subject of increasing interest for its involvement in the regulation of post-mitotic neuronal physiology and protein homeostasis separate from its canonical role in regulation of mitosis and cytokinesis. The reciprocal co-IPs of APC/C and SMN show that these proteins interact in neurons. SMN protein is known to be ubiquitinated and degraded by the 26S proteasome, however mechanism of ubiquitination is not known. The latter is important not only to understand molecular mechanisms in neurons but also for the therapeutics that depend on specific targeting of regulating E3 ligases responsible for ubiquitination.

We observed that the APC/C E3 ligase directly regulates SMN protein by ubiquitination and degradation; that in turn is facilitated by phosphorylation in SMN protein at Ser4 by GSK3β. It is relevant to note that inhibitors of GSK3β have been shown to increase levels of SMN proteins in a drug screen designed to identify regulators of SMN protein. APC/C inhibition, in the presence of endogenous inhibitor such as EMI1 and small molecule inhibitor such as proTAME, was found to be associated with increase in levels of SMN protein in a range of cell types such as primary mouse cortical neurons, SMA patient fibroblasts and motor neurons differentiated from embryonic stem cell. SMA disease is associated with motor axon degeneration resulting in dysfunctional motor units unable to innervate muscles. Primary motor neurons from mouse spinal cord demonstrate an increase in length of neurites with proTAME inhibitor treatment indicating that APC/C activity may be able to potentially regulate the SMA disease-specific vulnerability of motor neurons. When tested in the zebrafish SMA model directly, APC/C inhibition results in remarkable rescue of SMA disease including improved length of motor axons with rescued branching as well as improved path-finding targeted to myotomes. This was achieved along with increased levels of SMN as well as HuD proteins in motor neurons *in vivo*. There are no off-target effects of cytoskeletal defects or cytotoxicity, an improvement over previous generic inhibition of proteasome machinery to stabilize SMN levels. In contrast, APC/C inhibition in zebrafish resulted in the observable healthy phenotype in SMN MO injected fish.

These findings indicate that APC/C is important in determining the turnover and availability of SMN protein, a limiting factor in SMA disease and may be investigated further for therapeutics.

ACKNOWLEDGEMENT

The author acknowledges Dr. Bikem Akten for the initial experiments on APC/C regulation of SMN in cortical cultures, Dr. Hao Le for introducing the constructs in zebrafish SMA model and imaging the embryos.

ABBREVIATIONS

APC/C (APC)	Anaphase Promoting Complex/Cyclosome
D-box	Degradation box motif
HUD (ELAV)	Hu family protein member (Embryonic Lethal Abnormal Vision)
IP	Immunoprecipitation
LC-MS/MS (MS)	Liquid Chromatography based Mass-Spectrometry
PBS	Phosphate Buffered Saline
PTM	Post Translational Modifications
Q (RT-PCR)	Quantitative Reverse Transcription- Polymerase chain reaction
RBP	RNA binding protein
SMA	Spinal Muscular Atrophy
SMN	Survival of Motor Neurons
UPS	Ubiquitin Proteasome System
WB	Western blotting

REFERENCES

1. Coovert, D.D., et al., The survival motor neuron protein in spinal muscular atrophy. Hum Mol Genet, 1997. **6**(8): p. 1205-14.
2. Burghes, A.H. and C.E. Beattie, Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? Nat Rev Neurosci, 2009. **10**(8): p. 597-609.
3. McAndrew, P.E., et al., Identification of proximal spinal muscular atrophy carriers and patients by analysis of SMNT and SMNC gene copy number. Am J Hum Genet, 1997. **60**(6): p. 1411-22.
4. Prior, T.W., et al., Homozygous SMN1 deletions in unaffected family members and modification of the phenotype by SMN2. Am J Med Genet A, 2004. **130A**(3): p. 307-10.
5. Seo, J., et al., Spinal muscular atrophy: an update on therapeutic progress. Biochim Biophys Acta, 2013. **1832**(12): p. 2180-90.
6. Lefebvre, S., et al., Correlation between severity and SMN protein level in spinal muscular atrophy. Nat Genet, 1997. **16**(3): p. 265-9.
7. Rigo, F., et al., Pharmacology of a central nervous system delivered 2'-O-methoxyethyl-modified survival of motor neuron splicing oligonucleotide in mice and nonhuman primates. J Pharmacol Exp Ther, 2014. **350**(1): p. 46-55.
8. Ciechanover, A., Proteolysis: from the lysosome to ubiquitin and the proteasome. Nat Rev Mol Cell Biol, 2005. **6**(1): p. 79-87.
9. Chang, H.C., et al., Degradation of survival motor neuron (SMN) protein is mediated via the ubiquitin/proteasome pathway. Neurochem Int, 2004. **45**(7): p. 1107-12.
10. Hsu, S.H., et al., Ubiquitin carboxyl-terminal hydrolase L1 (UCHL1) regulates the level of SMN expression through ubiquitination in primary spinal muscular atrophy fibroblasts. Clin Chim Acta, 2010. **411**(23-24): p. 1920-8.
11. Burnett, B.G., et al., Regulation of SMN protein stability. Mol Cell Biol, 2009. **29**(5): p. 1107-15.
12. Kwon, D.Y., et al., The E3 ubiquitin ligase mind bomb 1 ubiquitinates and promotes the degradation of survival of motor neuron protein. Mol Biol Cell, 2013. **24**(12): p. 1863-71.
13. Zhang, H.L., et al., Active transport of the survival motor neuron protein and the role of exon-7 in cytoplasmic localization. J Neurosci, 2003. **23**(16): p. 6627-37.
14. Carrel, T.L., et al., Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis. J Neurosci, 2006. **26**(43): p. 11014-22.
15. Castro, A., et al., The anaphase-promoting complex: a key factor in the regulation of cell cycle. Oncogene, 2005. **24**(3): p. 314-25.
16. Glotzer, M., A.W. Murray, and M.W. Kirschner, Cyclin is degraded by the ubiquitin pathway. Nature, 1991. **349**(6305): p. 132-8.
17. van Roessel, P., et al., Independent regulation of synaptic size and activity by the anaphase-promoting complex. Cell, 2004. **119**(5): p. 707-18.
18. Teng, F.Y. and B.L. Tang, APC/C regulation of axonal growth and synaptic functions in postmitotic neurons: the Liprin-alpha connection. Cell Mol Life Sci, 2005. **62**(14): p. 1571-8.
19. Yang, Y., A.H. Kim, and A. Bonni, The dynamic ubiquitin ligase duo: Cdh1-APC and Cdc20-APC regulate neuronal morphogenesis and connectivity. Curr Opin Neurobiol, 2010. **20**(1): p. 92-9.

20. Chan, Y.B., et al., Neuromuscular defects in a Drosophila survival motor neuron gene mutant. Hum Mol Genet, 2003. **12**(12): p. 1367-76.
21. Kowalski, J.R., et al., The Anaphase-Promoting Complex (APC) ubiquitin ligase regulates GABA transmission at the C. elegans neuromuscular junction. Mol Cell Neurosci, 2014. **58**: p. 62-75.
22. Bantscheff, M., et al., Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. Mol Cell Proteomics, 2008. **7**(9): p. 1702-13.
23. Sowa, M.E., et al., Defining the human deubiquitinating enzyme interaction landscape. Cell, 2009. **138**(2): p. 389-403.
24. Visintin, R., S. Prinz, and A. Amon, CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. Science, 1997. **278**(5337): p. 460-3.
25. Singh, S.A., et al., Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J, 2014. **33**(4): p. 385-99.
26. Reimann, J.D., et al., Emi1 is a mitotic regulator that interacts with Cdc20 and inhibits the anaphase promoting complex. Cell, 2001. **105**(5): p. 645-55.
27. Zeng, X., et al., Pharmacologic inhibition of the anaphase-promoting complex induces a spindle checkpoint-dependent mitotic arrest in the absence of spindle damage. Cancer Cell, 2010. **18**(4): p. 382-95.
28. Pfleger, C.M. and M.W. Kirschner, The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. Genes Dev, 2000. **14**(6): p. 655-65.
29. Littlepage, L.E. and J.V. Ruderman, Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev, 2002. **16**(17): p. 2274-85.
30. Okudaira, K., et al., Transcriptional regulation of the Drosophila orc2 gene by the DREF pathway. Biochim Biophys Acta, 2005. **1732**(1-3): p. 23-30.
31. Reis, A., et al., The CRY box: a second APCcdh1-dependent degron in mammalian cdc20. EMBO Rep, 2006. **7**(10): p. 1040-5.
32. Makhortova, N.R., et al., A screen for regulators of survival of motor neuron protein levels. Nat Chem Biol, 2011. **7**(8): p. 544-52.
33. Passmore, L.A., D. Barford, and J.W. Harper, Purification and assay of the budding yeast anaphase-promoting complex. Methods Enzymol, 2005. **398**: p. 195-219.
34. Rossoll, W., et al., Smn, the spinal muscular atrophy-determining gene product, modulates axon growth and localization of beta-actin mRNA in growth cones of motoneurons. J Cell Biol, 2003. **163**(4): p. 801-12.

Chapter 6

General Discussion and Future Perspectives

Contents **Page No.**

1. Quantitative Profiling of Peptides from RNAs classified as non-coding

Role of ncRNAs in cellular processes is not well understood. Recent studies have shown their involvement in regulating the genome in the context of neurodegenerative disease [1]. While various categories of ncRNAs may have multiple functions in cells; their potential to directly code for proteins is almost unknown in-spite of having a similar structure to mRNAs.

The advancements in RNA-seq technology has made it possible to obtain a complete set of transcriptome with high sensitivity and deep coverage from any cell types, including neurons. Although, the presence of a transcript does not necessarily imply the presence of its protein product, it is nonetheless important to investigate transcripts to obtain an understanding of the post-transcriptional processing of RNAs, both from coding as well as non-coding genome, including their splicing, stability and translation [2, 3]. On the other hand, the high-throughput global proteomics based approaches aim to produce a comprehensive and quantitative map of the functional proteome from acquired MS/MS spectra [4]. The availability of good chemical labeling tags such as TMT for multiplexing the proteomics experiment provides a great tool to process multiple biological samples, such as time-dependent cellular profiling, for quantitative comparisons [5].

The success of a proteomics experiment is largely determined by the state-of-the-art mass spectrometer with high resolution and mass accuracy, good experimental design and a reliable pipeline for databases search. The identification of the sequenced peptides, typically, relies on the spectra matches using databases comprised of annotated protein coding genes, and may exclude the protein/peptide output from non-coding genome. The use of proteomics data to identify novel protein coding regions and to refine gene annotations has been discussed since early days of proteomics [6]. However, it is now feasible to use integrative approaches combining MS/MS proteomics and RNA-seq data to compare proteome and transcriptome from cells. To that end, the number of ‘peptide-to-spectrum’ matches can be improved using custom designed databases, with the inclusion of next generation RNA-seq based transcriptome [7]. This strategy provides an advantage of including all cellular transcripts - representing the expressible

set of genome - as well as information on single nucleotide polymorphisms, mutations, splicing and RNA editing [7-9] to identify novel translation products and protein forms [10-12].
In our study, we used total RNA-seq and TMT-labeled quantitative MS/MS proteomics from activity dependent paradigm of depolarized primary cortical neurons from mouse. We systematically analyzed the identified peptides against canonical databases and mapped the unmatched peptide 'hits' against custom database containing six frames of transcriptome data. This approach indicated the presence of hundreds of peptides translated from the non-coding regions of genome suggesting the coding potential of ncRNAs. The six frame translations of transcriptome data alone may exclude the protein products from various alternative translation mechanisms such as presence of internal ribosome entry sites, leaky scanning, ribosomal frame-shifting, multiple ORFs per transcript template, non-AUG start codons and stop codon read through [13-15]. However, in combination with proteomics analysis, this method enables the mapping of protein products to the templates of transcripts. The ribosome profiling [16, 17], in combination with RNA-seq, further provides a better understanding of the coding potential of transcriptome.

We validated the 'novel' peptides identified in our study using multiple approaches including bioinformatics approaches and spectral comparison with synthetic peptides. Also, we mined the ribosomal profiling data and found that many of these peptides were shown to be a match, indicating they may undergo ribosomal dependent translation. Since we investigated only identified 'novel' peptides, our analysis excludes the possibility of finding non-coding transcripts that are untranslatable despite being associated with ribosomes known for some ncRNAs. The ribosome profiling based sequencing, however, does not suggest the evidence of mature proteins and/or their stability [18-20]. We performed IP experiment in tandem with LC-MS/MS based proteomics that identified the presence of 'novel' peptide from non-coding region of intron in endogenous FARP1 protein, a protein that was observed with multiple isoforms of varying size. Taken together, these observations indicate that some of these novel peptides may contribute to the complexity of cellular proteome. Also, we observed that these novel peptides are regulated in abundance similar to that of known proteins. This indicates that they are actively regulated and may have some unknown biological functions. Thus our in-depth proteogenomics study has

provided strong evidence of peptides/proteins from ncRNAs that were largely termed as ‘junk DNA’.

The functional role and associated biological networks of these 'novel’ peptides are yet to be explored. Many of the novel peptides may be investigated further to identify their role in activity dependent functions of neurons such as synaptic plasticity. These novel peptides using MS/MS proteomics may not have been observed earlier because of their lower abundance and also because of the conventional design of the proteomic studies. The low abundance proteins may, however, play crucial role in cells such as transcription factors. Our study indicates that cellular proteome may be more complex than appreciated. The unknown proteome from ncRNAs may have tremendous scope for functional implications in healthy or diseased neurons along with potential for pharmacological interventions.

2. Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation

A number of neurodegenerative diseases have been linked with faulty protein degradation mechanisms. In non-neuronal cells, the turnover and abundance of functional proteome is known to be regulated by (a) autophagy–lysosomal associated degradation mechanisms and (b) ubiquitination-proteasome mediated protein degradation. In neurons, the role of autophagy regulated protein degradation is well studied and there are promising therapeutic autophagy targets for neurodegenerative diseases like Alzheimer’s [21]. The regulation of protein degradation by ubiquitination-proteasome mediated degradation is not well understood. However, there are some evidences linking the loss of function in important E3 ubiquitin ligases associated with accumulation of mis-folded proteins that perturb cellular homeostasis e.g. in Angelman syndrome [22], while gain of function in E3 ubiquitin ligases may create a paucity of essential target proteins such as in some case of poly-glutamine diseases [23].

The APC/C E3 ligase, in particular, has been investigated with increasing interest to identify its substrates during cell proliferation and in post mitotic differentiated neurons, since many of its targets are implicated in cancer and neurological disorders. The discovery of these substrates along with their varying abundance is important to understand various biological processes e.g.

regulation of cell cycle and certain disease specific regulators in neurons. The *in vitro* high-throughput ubiquitination screens [24] along with our quantitative global proteome approaches [25] have greatly simplified the search for APC/C substrates. However, confirmation of APC/C targets typically relies on *in vitro* biochemistry methods. In addition, these substrates may employ various complex regulatory mechanisms including PTMs such as phosphorylation as well as the structural elements such as recognition motifs for ubiquitination by APC/C. The APC/C is a well-characterized E3 ligase with known enzyme-substrate dynamics, endogenous activators and specific inhibitors, which may help in probing its functions.

In our study, we acquired a temporal map of quantitative changes of more than 4,400 proteins across different cell-cycle phases using 6-point TMT-labels. Co-regulation screening of the data based on the profiles of canonical APC/C substrates identified novel APC/C substrate candidates including various members of kinesin family. We validated many of these kinesin proteins including KIFC1, KIF18A, KIF2C and KIF4A using biochemical methods. We performed *in vitro* assays that confirmed KIFC1, among others, as a novel substrate of APC/C. We investigated the D-box motifs present in KIFC1 as well as KIF18A to establish the mechanisms by which APC/C recognize and modifies these substrates. Further, we conducted a thorough examination of phosphorylation dependent degradation of KIFC1 and observed that CDK-site for Ser6 phosphorylation in KIFC1 results in the inhibition of the APC/C-mediated degradation of KIFC1. The mechanism of phosphorylation dependent degradation is known in only two other substrates, SECURIN [26] and AURKA [27], of APC/C in cell cycle. This is one of the first comprehensive quantitative proteomics study undertaken to investigate APC/C regulation in cell cycle along with extensive *in vitro* validation that enables us to observe degradation profiles of important kinesin proteins. Moreover, the mechanisms involved in the degradation of KIFC1 were dissected that provide us insight into the APC/C dependent regulation of cellular proteins.

3. The APC/C regulates SMN and HuD levels and rescues the SMA disease phenotype

SMA is a neuronal disease characterized by the motor neurons specific degeneration and denervation of skeletal muscles due to the lack of a RNA processing protein called SMN. SMN protein is known to interact with various other RBPs and their RNA targets to regulate splicing, transport and translation of RNAs in neurons. Also, SMN is important in assembly of

spliceosome and ubiquitously present in all cells; therefore, its low abundance may potentially perturb many molecular pathways in neurons. Stabilizing the endogenous levels of SMN protein by identifying its interacting modifier proteins is important to understand underlying molecular mechanism and finding therapeutics in the SMA disease. Recently, HuD protein was shown to interact with SMN protein along with its target RNAs [28]. However, the functional consequences of this association are not known. In this study we investigated the regulation of SMN and its interacting protein HuD in neuron specific context of the SMA disease.

We observed that low availability of HuD protein may destabilize SMN protein in neurons and overexpression of HuD results in a partial rescue of SMA disease phenotype in zebrafish model. This study suggests that HuD may act as modulator of SMN protein and directly participates in molecular networks associated with SMA pathophysiology. Hence, it is important to explore the mechanisms by which HuD protein is regulated in neurons. We identified novel interaction of HuD with APC/C and observed that HuD is a direct target of ubiquitination/degradation via APC/C-UPS and is recognized by a sequence element of D-box motif. The APC/C inhibition results in increased levels of HuD proteins in primary mouse neurons, thus confirming that HuD abundance in neurons is subjected to the regulation by APC/C, while the D-box mutant protein-form of HuD is not degraded in the presence of APC/C and its overexpression partially rescues the SMA disease phenotype in zebrafish model.

SMN protein is believed to be degraded by proteasome; however, mechanism of ubiquitination is not known. A class of proteasomal inhibitors has been identified to increase the levels of SMN protein [29], albeit their use may have off-target effects perturbing the cellular homeostasis, indicating the importance to identify specific E3 ligases that may ubiquitinate SMN protein. Our co-IP based proteomics analysis reveals that APC/C interacts with SMN protein in neurons as well as other cell types such as HeLa cells. This finding prompted us to investigate if SMN levels are directly controlled by APC/C. Unlike HuD, SMN protein does not have D-box or other structural elements recognized APC/C. Therefore we investigated exons-deleted mutants SMN protein and PTMs including a GSK3β kinase phosphorylation site at Ser4 in SMN. We observed that SMN is readily ubiquitinated/degraded in presence of APC/C due to the presence of phospho-site Ser4; thus, regulating the stability and abundance of SMN protein. It is relevant to

note that inhibitors of GSK3β are known to increase levels of SMN proteins in a drug screen designed to identify modifiers of SMN protein [29], providing evidence in support of our observations.

Further, we tested the effects of APC/C inhibition in dissociated primary cortical cultures from mouse; motor neurons differentiated from mouse ES cells and patient fibroblast cell line and observed that APC/C inhibition results in increasing levels of SMN protein in these cells types. It is known that SMA disease is associated with degeneration of motor neurons and axons resulting in dysfunctional motor units that are unable to innervate muscles. In our study, the primary motor neurons from mouse spinal cord show increase in length of neurites in the presence of APC/C inhibitor, suggesting that APC/C activity may potentially regulate SMA disease specific vulnerability of motor neurons. When tested in zebrafish SMA model, the APC/C inhibition by overexpression of EMI1, results in a remarkable rescue of SMA disease showing improvements in the morphology of motor neurons - including length, branching and ends in axons as well as improved path-finding targeted to myotomes. These zebrafish were observed to have increased levels of endogenous SMN and HuD proteins, and looked healthy upon APC/C inhibition.

Together, these observations indicated that HuD is a positive modulator of SMN protein and a co-regulatory molecule involved in SMA disease. Also, the APC/C is important in determining the turnover and availability of SMN protein, a limiting factor in SMA disease. This study is important to identify the underlying molecular mechanisms in the regulations of the SMN protein abundance in the SMA disease. Identification of a specific E3 ligase for SMN ubiquitination may enable specific therapeutics in regulating SMN protein abundance. A further advancement in SMA and other neurodegenerative disorder will be to identify large scale neuronal proteins that undergo ubiquitination and proteasomal degradation. The candidate proteins from these studies along with their regulatory mechanisms may provide better insight into the neurodegeneration and its cure.

REFERENCES

1. Barry, G., Integrating the roles of long and small non-coding RNA in brain function and disease. Mol Psychiatry, 2014. **19**(4): p. 410-6.
2. Selbach, M., et al., Widespread changes in protein synthesis induced by microRNAs. Nature, 2008. **455**(7209): p. 58-63.
3. Sonenberg, N. and A.G. Hinnebusch, New modes of translational control in development, behavior, and disease. Mol Cell, 2007. **28**(5): p. 721-9.
4. Bantscheff, M., et al., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
5. Dayon, L., et al., Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem, 2008. **80**(8): p. 2921-31.
6. Choudhary, J.S., et al., Interrogating the human genome using uninterpreted mass spectrometry data. Proteomics, 2001. **1**(5): p. 651-67.
7. Woo, S., et al., Proteogenomic database construction driven from large scale RNA-seq data. J Proteome Res, 2014. **13**(1): p. 21-8.
8. Ning, K., D. Fermin, and A.I. Nesvizhskii, Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res, 2012. **11**(4): p. 2261-71.
9. Wang, X., et al., Protein identification using customized protein sequence databases derived from RNA-Seq data. J Proteome Res, 2012. **11**(2): p. 1009-17.
10. Beck, M., et al., The quantitative proteome of a human cell line. Mol Syst Biol, 2011. **7**: p. 549.
11. Djebali, S., et al., Landscape of transcription in human cells. Nature, 2012. **489**(7414): p. 101-8.
12. Low, T.Y., et al., Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. Cell Rep, 2013. **5**(5): p. 1469-78.
13. Touriol, C., et al., Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell, 2003. **95**(3-4): p. 169-78.
14. Michel, A.M., et al., Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res, 2012. **22**(11): p. 2219-29.
15. Namy, O., et al., Reprogrammed genetic decoding in cellular gene expression. Mol Cell, 2004. **13**(2): p. 157-68.
16. Ingolia, N.T., Genome-wide translational profiling by ribosome footprinting. Methods Enzymol, 2010. **470**: p. 119-42.
17. Menschaert, G., et al., Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol Cell Proteomics, 2013. **12**(7): p. 1780-90.
18. Bazzini, A.A., et al., Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J, 2014. **33**(9): p. 981-93.
19. Volders, P.J., et al., LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res, 2013. **41**(Database issue): p. D246-51.
20. Guttman, M. and J.L. Rinn, Modular regulatory principles of large non-coding RNAs. Nature, 2012. **482**(7385): p. 339-46.

21. Orr, M.E. and S. Oddo, Autophagic/lysosomal dysfunction in Alzheimer's disease. Alzheimers Res Ther, 2013. **5**(5): p. 53.
22. Matsuura, T., et al., De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat Genet, 1997. **15**(1): p. 74-7.
23. Ciechanover, A. and P. Brundin, The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. Neuron, 2003. **40**(2): p. 427-46.
24. Merbl, Y. and M.W. Kirschner, Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. Proc Natl Acad Sci U S A, 2009. **106**(8): p. 2543-8.
25. Singh, S.A., et al., Co-regulation proteomics reveals substrates and mechanisms of APC/C-dependent degradation. EMBO J, 2014. **33**(4): p. 385-99.
26. Holt, L.J., A.N. Krutchinsky, and D.O. Morgan, Positive feedback sharpens the anaphase switch. Nature, 2008. **454**(7202): p. 353-7.
27. Littlepage, L.E. and J.V. Ruderman, Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev, 2002. **16**(17): p. 2274-85.
28. Fallini, C., et al., The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly(A) mRNA in primary motor neuron axons. J Neurosci, 2011. **31**(10): p. 3914-25.
29. Makhortova, N.R., et al., A screen for regulators of survival of motor neuron protein levels. Nat Chem Biol, 2011. **7**(8): p. 544-52.

List of Publications
Acknowledgements

List of Publications

1. Sasha Singh*, Dominic Winter*, Marc Kirchner*, **Ruchi Chauhan***, Saima Ahmed, Nurhan Ozlu, Amit Tzur, Judith A. Steen#, Hanno Steen#. Co-regulation Proteomics Reveals Substrates and Mechanisms of APC/C-dependent degradation. The EMBO Journal, Vol 33, No. 4, 2014.
   *These authors should be considered shared first author. #These authors contributed equally to the work.

2. Sudhakaran Prabakaran*, Martin Hemberg*, **Ruchi Chauhan***, Dominic Winter, Ry Tweedie-Cullen, Christian Dittrich, Elizabeth Hong, Jeremy Gunawardena, Hanno Steen, Gabriel Kreiman# and Judith A. Steen#. Quantitative Profiling of Peptides from RNAs classified as non-coding. Nature Communication, Nov 18; 5:5429, 2014.
   ***These authors should be considered shared first author. #These authors contributed equally to the work.

3. **Ruchi Chauhan**, Hao Le, Bikem Akten**,** Zhigang He, Randall King, Christine Beattie, Hanno Steen, Judith A. Steen. The Anaphase Promoting Complex Regulates SMN protein abundance and rescues SMA disease phenotype (Manuscript to be submitted).

Acknowledgements

I am thankful to Dr. Judith Steen for giving me the opportunity to work on this project and for her excellent mentoring. I am grateful to Dr. Bernhard Kuster for his guidance and help in formalizing this work into doctorate thesis. I am obliged to Dr. Hanno Steen for providing me an opportunity to work in Steen lab and his valuable advice on this research work.

I express thanks to my former and present colleagues of Steen lab, especially Waltraud Mair, Sebastian Berger, Dr. Michaela Helmel, Dr. Ertugrul Cansizoglu, Dr. Kevin Broadbelt, Dr. Jan Muntel, Dr. Shaojun Tang, Dr. John Froehlich and Dr. Hui Chen for their dynamic scientific discussions at weekly lab meetings & journal clubs and for readiness to share expertise and resources during this work. I am thankful to vibrant technical staff of Steen lab especially Ceren Uncu, John Sauld and Saima Ahmed for their timely support in animal care, lab management and routine functioning of lab.

I am thankful to my collaborators Dr. Gabriel Kreiman, Dr. Martin Hemberg, Dr. Christine Beattie, Dr. Dominic Winter, Dr. Sasha Singh, Dr. Bikem Akten, Dr. Hao Le, Dr. Rubin Lee and Dr. Randall King for their resourcefulness and a great work experience with them during the experiments.

I am indebted to Dr. Kristin Krukenberg and Dr. Gary Mcdowell for critically proof-reading this manuscript.

I thank Late Dr. VK Jain and all my other teachers for teaching me perseverance in pursuit of knowledge among others. I am grateful to my parents, late parents-in law and sisters Nidhi Chauhan and Divya Chauhan for their unconditional support and encouragement. I thank Dr. Sudhakaran Prabakaran for his critical comments on this manuscript, his inspirational dedication to science and for being a strong support. Finally, I thank Soham Sudhakaran for being co-operative and patient with me all this while and to be a happy person.