Transcriptome Profiling of UPF3B-KO Cells and Its Implication in Alternative Splicing-Coupled Nonsense-Mediated Decay

A thesis submitted in partial fulfilment of the requirement for the award of the degree of

Doctor of Philosophy

by

Pratap Chandra

Roll No. 166106109

Under the supervision of

Dr. Kusum K. Singh

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Assam – 781039, India

June 2024

Transcriptome Profiling of UPF3B-KO Cells and Its Implication in Alternative Splicing-Coupled Nonsense-Mediated Decay

A thesis submitted in partial fulfilment of the requirement for the award of the degree of

Doctor of Philosophy

by

Pratap Chandra

Roll No. 166106109

Under the supervision of

Dr. Kusum K. Singh

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Assam – 781039, India

June 2024

Dedicated

to

my parents

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Statement

I do hereby declare that the research findings in this thesis entitled "Transcriptome Profiling of UPF3B-KO Cells and Its Implication in Alternative Splicing-Coupled Nonsense-Mediated Decay" is the result of work carried out in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, under the supervision of Dr. Kusum Kumari Singh.

In keeping with the general practice of reporting scientific observations, due acknowledgement has been made wherever the work described is based on the findings of other investigators.

Date: May 27, 2024

Pratap Chandra

Roll No. 166106109

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

Certificate

It is certified that the work described in this thesis entitled "Transcriptome Profiling of UPF3B-KO Cells and Its Implication in Alternative Splicing-Coupled Nonsense-Mediated Decay" by Pratap Chandra (Roll No. 166106109) for the award of the degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, IIT Guwahati. The work embodied in this thesis has not been submitted elsewhere for a degree.

Kusum 27/05/2024

Dr. Kusum Kumari Singh

Thesis Supervisor

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Assam-781039, India

Acknowledgements

Firstly, I would like to acknowledge my thesis supervisor, Dr. Kusum Kumari Singh, for giving me the opportunity to work in her laboratory and pursue my dream. I am fortunate to receive her continued support, guidance and scientific insights during every step of my Ph.D. journey. I am grateful to her for giving me the freedom to work independently and helping me grow as a scientist and teaching me the skills to be an independent researcher.

I would also extend my gratitude to the members of my doctoral committee, Dr. Biplab Bose (Chairperson), Prof. B. Anand and Prof. Ashish Anand, for their unbiased opinions and insightful recommendations that have been the driving force required for the completion of my thesis.

I would like to thank the Heads of the Department, Prof. Kannan Pakshirajan, Prof. Latha Rangan and Prof. Rakhi Chaturvedi as well as other faculty members for their support and for lending the facilities of their laboratories.

I would also like to extend my acknowledgement towards lab members of Molecular Endocrinology Lab and Cancer Biology Lab for helping me establish basic protocols in the new laboratory during the initial years of my Ph.D. as second student.

I want to thank all members of the RBP laboratory who have always given me a friendly, supportive and edifying environment in the lab. I would like to acknowledge the help I got from Bhagyashree, Sweta, Ayushi, Priyanka and Sayan for my thesis work. I would like to thank Bhagyashree for optimizing protocols, Sweta for performing Western blotting, Ayushi for checking the manuscripts, Priyanka and Sayan for their help in bacterial cultures and plasmid preparations. I would always cherish the memories we made together as RBP family.

I extend my thanks to the Department of Biosciences and Bioengineering for providing the research facilities to accomplish my Ph.D. thesis. I want to thanks all the staff members of the Department of Biosciences and Bioengineering for providing me with the logistical support essential to perform my research. Further, I would like to acknowledge IIT Guwahati, Ministry of Human Resources Development, India and Indian Council of Medical Research, for financial assistance and funding for my Ph.D. project.

I feel fortunate to have Suvankar, Adhiraj, Ratan, Dipak and Vineet as friends who were constantly with me during the ups and downs of my Ph.D. journey. I will never forget the happy time and friendship we shared over the years. I would also like to thank Adhiraj for his help in writing the scripts and helping in RNA-Seq analysis. I also consider myself lucky to receive much care, love, support and knowledge from Suvankar and Adhiraj.

Finally, I would like to thank my Father, Mother, Brother, Sisters and friends for being the pillars of strength during my Ph.D. It is their innumerable sacrifices that have enabled me to reach this stage in life. This thesis is dedicated to my Father for encouraging me for graduate school, for boosting me up during the failures of my Ph.D. journey and for believing in me.

I also express my gratitude to all others whom I may have missed.

Thank you.

Pratap Chandra

Table of Contents

List of Figures

List of Tables

Abbreviations

Standard Terms

%	Percentage
°C	Degree Celcius
aa	amino acids
ATP	Adenosine Tri-Phosphate
bp	base pairs
Cat	Catalogue
cDNA	complementary DNA
DNA	Deoxyribonucleic Acid
ds	double-stranded
FP	Forward Primer
g	Gravitational force
gDNA	genomic DNA
gm	grams
hrs	hours
kb	kilo-base
L	Litre
M	Molar
min	Minutes
µg	micro-gram
mg	milli-gram
mL	milli-Litre
mM	milli-Molar

µM	micro-Molar
mRNA	messenger RNA
ng	nano-grams
nt	nucleotide
RNA	RiboNucleic Acid
RP	Reverse Primer

Materials and Methods

APS	Ammonium PerSulfate
BSA	Bovine Serum Albumin
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
dNTP	deoxyribonucleotide triphosphate
DPBS	Dulbecco's Phosphate Buffered Saline
DTT	Dithiothreitol
DW	Distilled Water
EDTA	Ethylene Diamine Tetra-acetic Acid
FACS	Fluorescence-Activated Cell Sorting
FBS	Foetal Bovine Serum
FP	Forward Primer
GFP	Green Fluorescent Protein
HEK	Human Embryonic Kidney
HeLa	Henrietta Lacks
LB	Luria-Bertani
MCS	Multiple Cloning Site

MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
NFW	Nuclease-Free Water
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis
PB	PiggyBac
PCR	Polymerase Chain Reaction
pH	potential of Hydrogen
PEI	Poly(EthyleneImine)
PMSF	Phenyl Methyl Sulfonyl Fluoride
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffered Saline Tween-20
PCR	Polymerase Chain Reaction
PenStrep	Penicillin-Streptomycin
qRT-PCR	quantitative Real-Time PCR
RIPA	Radioimmunoprecipitation Assay
RP	Reverse Primer
RT	Room Temperature
SDS	Sodium Dodecyl Sulphate
TBST	Tris-Buffered Saline-Tween-20
TEMED	TetraMethylEthyleneDiamine

Non-standard Terms

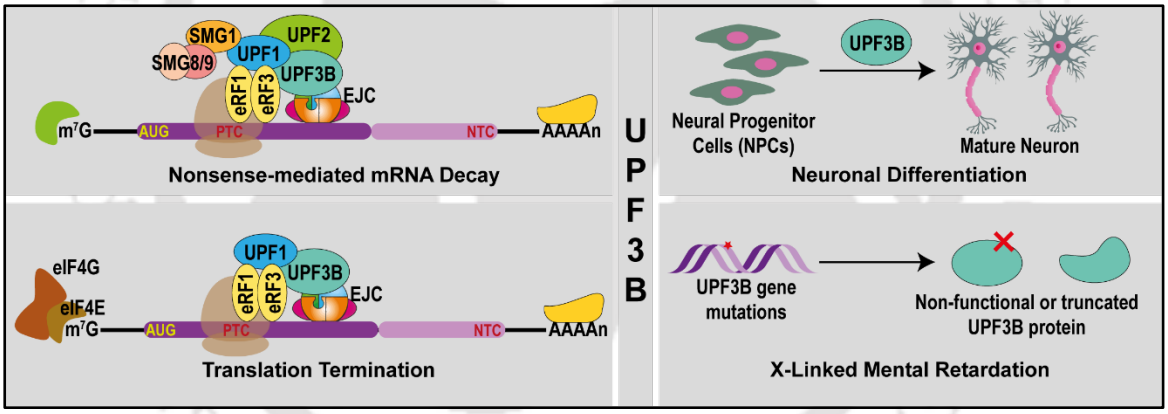
ADHD	Attention Deficient Hyperactivity Disorder
ASD	Autism Spectrum Disorder
CAGE	Cap Analysis Gene Expression

Cas	CRISPR-associated proteins
CASC3	Cancer Susceptibility Candidate 3 Protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
DEG	Differently Expressed Gene
E. coli	*Escherichia coli*
eIF4A3	eukaryotic translation Initiation Factor 4A3
EJC	Exon Junction Complex
eRF	Eukaryotic Release Factor
EST	Expressed Sequence Tag
ID	Intellectual Disability
IRES	Internal Ribosome Entry Site
KO	Knockout
MAGOH	Mago-nashi homolog
MIF4G	Middle domain of eukaryotic initiation factor 4G
NDDs	Neurodevelopmental disorders
NES	Nuclear Export Signal
NHEJ	Non-Homologous End Joining
NLS	Nuclear Localization Sequence
NMD	Nonsense-Mediated mRNA Decay
NTC	Normal Termination Codon
ORF	Open Reading Frame
PAM	Protospacer Adjacent Motif
PTC	Premature Termination Codon

RNA-Seq	RNA sequencing
RNP	Ribonucleoprotein
SAGE	Serial Analysis of Gene Expression
shRNA	short hairpin RNA
SMG	Suppressor of Morphological defects on Genitalia
tracrRNA	trans-crRNA
uORF	upstream Open Reading Frame
UPF	Up-Frameshift Suppressor
UTR	Untranslated Region
XLID	X-linked Intellectual Disability

Chapter 1:

Introduction and Literature Review



The review of literature embodied in this chapter is partially published.

Chandra P, Deka B, Singh KK. Functional roles of human Up-frameshift suppressor 3 (UPF3) proteins: From nonsense-mediated mRNA decay to neurodevelopmental disorders. *Biochimie*. 2021 Jan;180:10-22.

Abstract

Nonsense-mediated mRNA decay (NMD) is a post-transcriptional quality control mechanism that eradicates aberrant transcripts from cells. Aberrant transcripts are recognized by translating ribosomes, eukaryotic release factors (eRFs), and trans-acting NMD factors, leading to their degradation. These trans-factors, UPF1, UPF2, and UPF3 proteins, are conserved among eukaryotes. Intriguingly, in humans, UPF3 exists as paralog proteins, UPF3A and UPF3B. While UPF3 paralogs are known to interact with exon junction complex (EJC) in bridging the NMD machinery to the EJC, there is a growing consensus that the UPF3 proteins are also associated with other critical cellular functions beyond quality control.

This literature review chapter presents the current knowledge on the biochemical functions of UPF3 paralogs in diverse cellular processes, including NMD and translation. I also discuss the role of the UPF3 paralogs in the development and function of the central nervous system and germ cells. Furthermore, significant advances in the past decade have provided new perspectives on the implications of UPF3 paralogs in neurodevelopmental diseases. Genome- and transcriptome-wide sequencing analysis of patient samples has revealed that the loss of UPF3B is associated with brain disorders such as intellectual disability, autism, attention deficit hyperactivity disorder, and schizophrenia. Therefore, the aim of this chapter is to provide insight into the brain diseases associated with loss-of-function mutations of UPF3B.

Chapter 1

1.1 Introduction

Eukaryotic gene expression is a series of events that begin with the transcription of genomic DNA into premature ribonucleotides (pre-mRNAs), splicing of pre-mRNAs to produce mature ribonucleotides (mRNAs), and ultimately the translation of mRNAs into proteins [1]. The splicing (constitutive and alternative splicing) of a single pre-mRNA generates multiple transcripts, thereby increasing the transcriptional diversity of the cell. In such a crowded transcriptional landscape, sophisticated control mechanisms are required to prevent disruptive sense or antisense transcription events for the accumulation of non-functional RNAs that may interfere with the metabolism of functional molecules. These control mechanisms may be post-transcriptional [2] or co-translational [3]. Post-transcriptional regulatory mechanisms function during the maturation of transcripts by identifying the defects, such as capping failure, hairpin formation, splicing defects, export failure [4], and poly(A) shortening [5]. On the other hand, co-translational gene regulation requires a pioneer round of translation to identify erroneous transcripts [5]. These translation-driven regulatory mechanisms include no-go decay (NGD) [6, 7] (Figure 1.1A), non-stop decay (NSD) [8, 9] (Figure 1.1B), and nonsense-mediated mRNA decay (NMD) [10] (Figure 1.1C). Among the translation-dependent pathways, NMD has gained much attention due to its physiological significance and association with human diseases [11, 12].

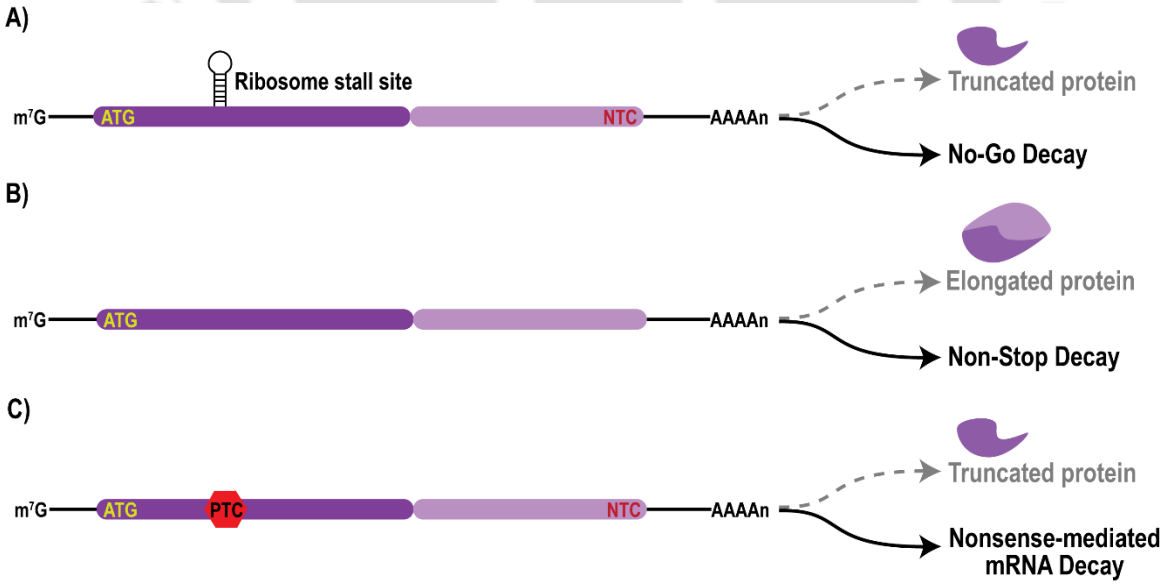


Figure 1.1. Translation-driven mRNA surveillance pathways. A) Transcript with hairpin structure halts the translation progression and leads to No-Go decay, otherwise produces a truncated protein product. B) Loss of stop codon in ORF causes Non-Stop decay of transcript, otherwise forms an elongated protein. C) Nonsense-mediated mRNA decay exploits premature termination codon in transcripts for degradation, if not, translates into truncated protein.

Introduction

NMD is a highly specialized surveillance pathway that barricades the building of aberrant transcripts in the cytoplasm. NMD safeguards cells from the deleterious effects of premature termination codon (PTC)-containing transcripts, which otherwise get translated into potentially toxic truncated proteins. Besides PTC, other cis-acting features (NMD-inducing features) that might elicit NMD include an upstream open reading frame (uORF) [13, 14], mRNAs with long 3' untranslated region (3' UTR) [15, 16], introns in 3' UTR [17], transcripts generated by splicing defects or regulatory alternative splicing coupled with NMD (AS-NMD) [18, 19] (Figure 1.2). Nevertheless, the presence of these canonical determinants on an mRNA does not always trigger NMD. In this regard, many mRNAs harbor NMD-inhibitory sequences, such as AU-rich sequences and RNA stability elements (RSE), which evade NMD [20, 21]. These suggest the intricate nature of NMD in choosing its substrate.

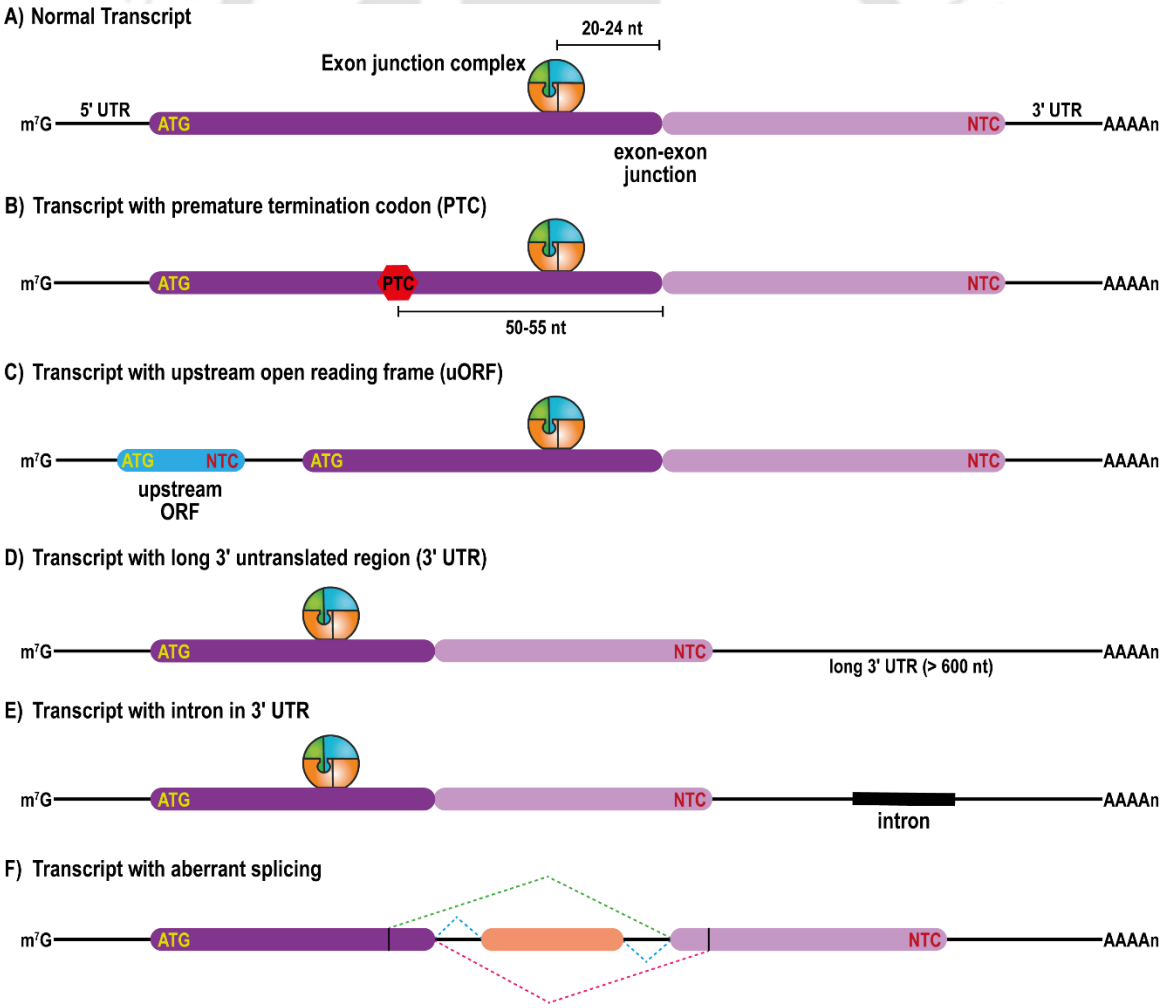


Figure 1.2. Transcript features identified by NMD. A) a normal transcript, B) a transcript containing PTC in coding region upstream to the legitimate stop codon (NTC), C) a transcript containing an upstream open reading frame (uORF), D) a transcript with a long 3' UTR (> 600 nts), E) transcript with an intron in 3' UTR region, F) transcripts generated by splicing defects or regulatory splicing events.

Besides the primary role of NMD in mRNA quality control, NMD also regulates the expression of normal mRNAs and maintains cellular homeostasis. Since organism development is highly dependent on the spatial and temporal synchronization of gene expression, NMD plays a vital role in controlling this dynamic process by regulating the mRNA decay rate [22]. Accordingly, NMD is associated with many physiological or pathophysiological conditions (as described in detail in recent reviews [23-26]). Over the last decade, studies on NMD in humans have increased due to its impact on genetic disorders, including cancer and neuropathological conditions [27-29].

For recognition and efficient degradation of aberrant transcripts, the NMD factors associate with the EJC. During splicing, the spliceosome deposits a multiprotein complex termed EJC onto mRNAs at a distance of 24 nucleotides upstream of an exon-exon junction [30-32]. The subunits that form the core EJC are DEAD-box RNA helicase eIF4A3 and MAGOH-Y14 heterodimer [33, 34]. CASC3 was previously regarded as the fourth EJC core protein, but recent studies showed that CASC3 functions during the late phase of the EJC life cycle [35]. The core EJC aids in nucleating the peripheral EJC factors through dynamic interactions and thereby regulates various post-transcriptional events. Some well-known peripheral EJC proteins and complexes are the splicing regulatory complexes ASAP (ACIN1-RNPS1-SAP18), and PSAP (PNN-RNPS1-SAP18), the transcription export complex (TREX), and UPF3B (Up-frameshift suppressor 3 homolog B) [35-37]. EJC acts as a molecular link between splicing and NMD by serving as a platform for the assembly of the NMD factors [38, 39].

In the literature review section, I will provide a brief introduction to NMD factors, especially UPF proteins. I will discuss the biochemical functions and biology of the UPF3 proteins across various cellular processes in detail. I will also discuss the association between UPF3B mutations and brain disorders, offering insights into the neurological conditions linked to the loss of UPF3B function.

1.2 Literature Review

1.2.1 The UPF Proteins

The initial discovery of mutants in yeast with altered NMD phenotypes resulted from a genetic screen aimed at identifying allosuppressors of the *his4-38* frameshift mutation [40, 41]. The *his4* region of chromosome III in yeast encodes for a protein associated with three nonsequential steps in the histidine biosynthetic pathway [42, 43]. Other mutants affecting the NMD pathway were also discovered in screens that were specifically designed to identify mutations that influence programmed −1 frameshift efficiency or suppress upstream initiation codons [44-46]. Investigation of these alleles showed that mutations in *Upf1*, *Upf2*, and *Upf3* lead to the accumulation of mRNAs with nonsense mutations without affecting the abundance of normal transcripts. The proteins encoded by these genes, Upf1, Upf2, and Upf3, modulate NMD in yeast and are non-essential for cell viability [41, 47-49]. In *Caenorhabditis elegans*, seven suppressors with morphogenic effects on genitalia (*Smg*) genes, *Smg-1-7*, are responsible for NMD activity [50]. Among these genes, *Smg2*, *Smg3,* and *Smg4* serve as the orthologs of the yeast *Upf1*, *Upf2,* and *Upf3* genes. The Upf counterparts in *C. elegans* exhibit minimal effects on genitalia and brood size and they do not significantly impact normal growth and development [50]. In humans, in addition to UPF1 [51] and UPF2 [52], UPF3 exists as paralog proteins, UPF3A and UPF3B [53, 54]. The failure of the NMD pathway or malfunctioning UPF proteins results in embryonic lethality in mice [55] or tissue-specific developmental defects in humans [56].

1.2.1.1 UPF1

UPF1 is an ATP-dependent RNA helicase characterized by conserved core features, including a cysteine/histidine-rich zinc finger-like domain, a nucleoside triphosphate (NTP)-binding domain, and a helicase motif [51, 57] (Figure 1.3A). The functions of UPF1 encompass nucleic acid-dependent ATP hydrolysis and ATP-dependent 5′→3' RNA/DNA helicase activity [58, 59]. Mutations in the *UPF1* gene that affect its ATP hydrolysis and helicase activity disrupt NMD [60]. Studies have demonstrated that UPF1 is required for the decay of aberrant transcripts. In UPF1-deficient cells, unspliced pre-mRNAs and mature mRNAs containing nonsense mutations are stabilized, underscoring UPF1's role in eliminating these transcripts [61]. Despite being evenly distributed in the cytoplasm [54, 62], UPF1 surprisingly bears a sequence resembling the bipartite nuclear localization

sequence (NLS) [51, 62]. However, the role of UPF1-NLS in nuclear import has not yet been elucidated.

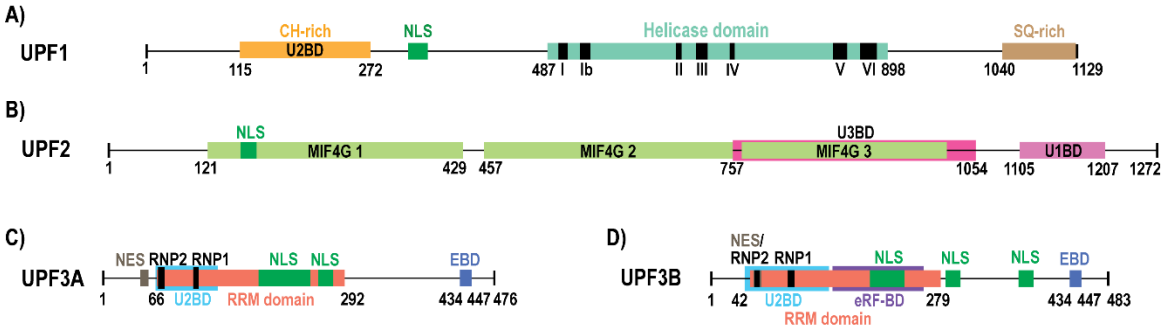


Figure 1.3. Schematic domain architecture of the UPF proteins [52, 53, 63-72]. A) UPF1 is an 1129 amino acid protein with N-terminal UPF2-binding domain (U2BD), nuclear localization signal (NLS), Helicase domain, and C-terminal serine-glutamine (SQ)-rich domain. B) UPF2 is the largest UPF protein, comprising 1272 amino acids. UPF2 contains three MIF4G domains. The third MIF4G domain interacts with UPF3 proteins (U3BD). UPF2 has C-terminal UPF1-binding domain (U1BD). C) UPF3A and D) UPF3B are composed of an RRM domain and EBD. Domains and motifs of UPF1, UPF2, UPF3A, and UPF3B are illustrated with positions of amino acids indicated below. CH – cystine-histidine, EBD – EJC binding domain, eRF-BD – eRF3a binding domain, MIF4G – the middle domain of eukaryotic initiation factor 4G, NES – nuclear export signal, RRM – RNA recognition motif.

1.2.1.2 UPF2

UPF2 is a substantial 148 kDa protein containing an acidic region at its carboxy (C)-terminus and three MIF4G (middle domain of eukaryotic initiation factor 4G)-like domains [67, 70] (Figure 1.3B). The C-terminal acidic region of UPF2 interacts with UPF1, and the third MIF4G domain of UPF2 binds to UPF3. These interactions of UPF2 play a pivotal role in bridging UPF1 to UPF3, initiating the formation of the NMD complex [47, 48]. Notably, removing the UPF1 binding domain from UPF2 has been shown to impair NMD activity [73, 74]. UPF2, identified as a perinuclear protein [54], is recruited to the UPF3-bound ribonucleoprotein (RNP) complex during its export to the cytoplasm [52, 67]. Similar to UPF1, UPF2 also contains an NLS [52], although its significance in nuclear import remains unknown. The primary function of UPF2 is to orchestrate the assembly of the NMD complex through interactions with the crucial NMD factors UPF1 and UPF3 [75].

1.2.1.3 UPF3

UPF3B is a crucial component of the NMD machinery, which is primarily located in the nucleus and shuttles between the nucleus and cytoplasm [54, 76]. The nucleocytoplasmic shuttling behavior of the UPF3B is supported by the presence of three bipartite NLSs and two leucine-rich nuclear export signals (NESs) in the *UPF3B* gene sequence [77, 78] (Figure 1.3D). Besides the conventional function of NES in nuclear export, UPF3B NES plays a role in interacting with UPF2. Mutation studies have

demonstrated that amino acid substitutions in the NES motif impede the UPF3B-UPF2 interaction, ultimately decreasing the NMD activity [79]. Detailed discussions on human UPF3 paralogs and their specific functions will be presented in subsequent sections.

1.2.2 The UPF3 Paralogs

In humans, the UPF3 paralogs are the outcome of gene duplication events during the evolution of vertebrates, resulting in a gene pair with distinct functions [80]. These paralogs are expressed by two different genes located on separate chromosomes. The *UPF3B* gene is situated on the X chromosome (previously known as hUpf3p-X), while the *UPF3A* gene is on chromosome 13 (hUpf3p). Despite their differential chromosomal locations, the *UPF3* paralogs share considerable sequence similarity [64]. Comparative analysis of the expression profiles of UPF3 paralogs reveals that UPF3B is more highly expressed in various cell types than UPF3A. UPF3B exhibits ubiquitous expression, with maximum expression in the fetal brain and testes. Conversely, UPF3A has low expression in many adult organs but is highly expressed in testes [53], where it plays a crucial role in regulating transcripts essential for gametogenesis.

UPF3 proteins feature a conserved amino-terminal with RNA recognition motif (RRM) domain and a carboxy-terminal containing EJC binding domain (EBD) [65, 66, 81] (Figure 1.3C-D). While RRM domains generally facilitate RNA binding, neither the RRM of UPF3A nor that of UPF3B is involved in any RNA binding activity. Instead, the RRM domains of the UPF3 proteins mediate specific protein-protein interactions, particularly with the NMD factor UPF2 [67]. However, evidence from size exclusion chromatography analysis suggests that UPF3B protein can indeed bind RNA [68]. A previous study using iCLIP (individual nucleotide resolution and immunoprecipitation) supported the direct RNA binding ability of UPF3B. iCLIP reads of UPF3B were notably enriched at spliced junctions, exhibiting a weak sequence preference for the 5'-GAAGA-3' sequence motif [32]. An intriguing question that remains is the identification of the specific UPF3B domain responsible for the interaction with RNA.

Recently, Bufton *et al*. conducted a characterization of the middle domain of UPF3 proteins that comprises an RRM-like domain (RRM-L), a NONA/paraspeckle-like domain (NOPS-L), and an extended α-helical domain [82]. The RRM-L, NOPS-L, and the first coiled-coil-like region (CCL1) are essential for UPF3B's RNA binding activity. Using fluorescence anisotropy (FA)-based binding assays, the study demonstrated that UPF3B

exhibits a higher affinity for single-stranded RNA (ssRNA) compared to single-stranded DNA (ssDNA) [82]. Surprisingly, UPF3B binds to double-stranded DNA and RNA with greater affinity than their single-stranded counterparts. Importantly, the RRM-L and NOPS-L domains of UPF3B also interact with the third MIF4G domain of UPF2. The UPF3B-UPF2 interaction is favoured over UPF3B-RNA and UPF2-RNA oligomerization. The UPF3 paralogs share 67.3% sequence identity for the RRM-L-NOPS-L-CCL1 region, indicating a conserved feature of RNA binding and oligomerization in both proteins.

1.2.3 Functions of UPF3B and Its Pathological Implications

In this section, I will discuss the biochemical roles of UPF3B and their physiological significance in neuronal development and neural functions. Furthermore, the section will delve into the impact of perturbations in UPF3B expression on the aetiology of various neurodevelopmental disorders.

1.2.3.1 UPF3B is Required for a Specific Branch of NMD

The available evidence indicates that NMD comprises several branches based on the assembly of different NMD factors (Figure 1.4); each governing distinct subsets of target RNAs. The classical NMD model proposes two modes of NMD activation: EJC-enhanced and EJC-independent NMD [83]. In the EJC-enhanced NMD pathway, EJC serves as a stimulator for the assembly of NMD factors and is critical for the decay of faulty transcripts (Figure 1.4A). Whereas in the EJC-independent NMD pathway, NMD factors do not associate with EJC to carry out NMD activity (Figure 1.4B). Remarkably, NMD can further be categorized into two branches - UPF2-independent and UPF3B-independent NMD pathways [75, 84]. In UPF2-independent NMD, UPF1 interacts with UPF3B-CASC3 to activate UPF1 ATPase and helicase activities, leading to the decay of faulty transcripts [85, 86] (Figure 1.4D). In the absence of UPF3B, RNPS1 interacts with UPF2 and UPF1 to assemble the NMD complex [75] (Figure 1.4C). Alternatively, Serine/Arginine (SR)-rich proteins phosphorylate UPF1 with or without interaction with UPF3B, portraying an SR protein-stimulated branch of the NMD pathway [87, 88] (Figure 1.4E). Recent evidence suggests that AKT1-mediated phosphorylation of UPF1, in addition to phosphorylation by SMG1, activates NMD [89, 90] (Figure 1.4F).

Taken together, the findings suggest that NMD is a branched pathway, and a crucial future objective will be to delineate the target transcripts of different NMD branches. In this context, the focus of the discussion will be on the UPF3B-dependent NMD pathway.

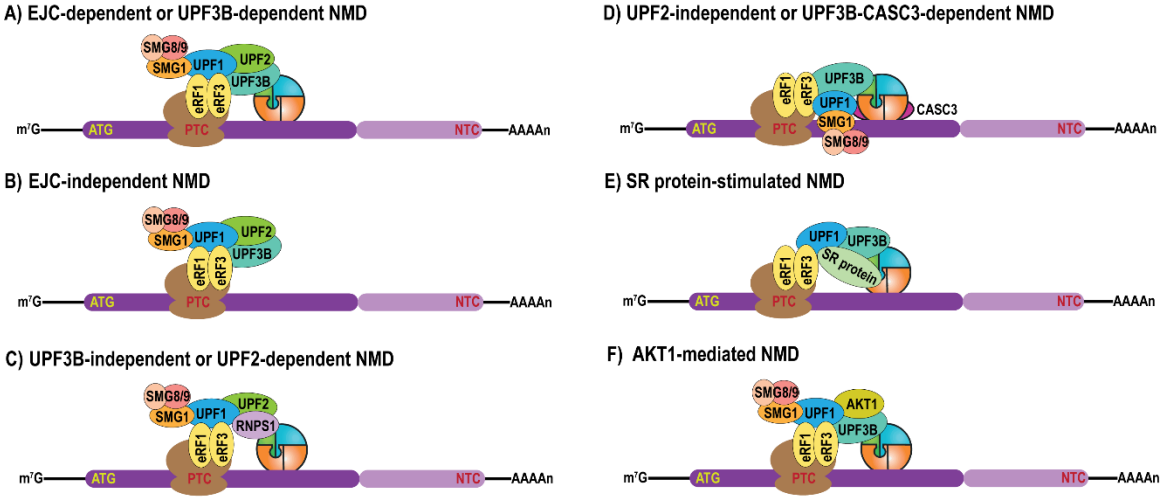


Figure 1.4. The branched nature of the NMD pathway. (A) EJC-dependent NMD pathway requires EJC for the recruitment and assembly of UPF and SMG proteins, B) EJC-independent NMD pathway does not include the involvement of EJC for the assembly of SURF complex, C) UPF3B-independent or UPF2-dependent NMD pathway requires the interaction of EJC with RNPS1 instead of UPF3B, D) UPF2-independent or UPF3B-CASC3-dependent NMD pathway does not require UPF2 for bridging the UPF3B-UPF1 interaction, E) SR protein-stimulated NMD pathway requires SR proteins to phosphorylate UPF1, F) AKT1-mediated NMD pathway is triggered after the phosphorylation of UPF1 by AKT1 with involvement of UPF3B.

Around the time of mRNA export to the cytoplasm, the UPF3B protein associates with core EJC factors, specifically the MAGOH-Y14 heterodimer. The C-terminal region of UPF3B (EBD) plays a critical role in the interaction between UPF3B with MAGOH-Y14. This EBD region in the UPF3B protein is highly conserved across vertebrates [65, 75, 76, 81] (Figure 1.3D). Significantly, even a single mutation within the EBD region of UPF3B compromises NMD activity, emphasizing the importance of the interaction between UPF3B and the MAGOH-Y14 heterodimer for UPF3B-mediated NMD activity [65]. Mutational studies have identified surface residues of both Y14 (Glu82 and Glu83) and MAGOH (Glu68 and Asp66) involved in binding the UPF3B C-terminal region [81]. Additionally, UPF3B interacts with the export factors, ALY/REF and TAP, indicating that UPF3B-bound transcripts are directed for export through the nuclear pore complex [76] (Figure 1.5A). After mRNA export to the cytoplasm, UPF1 is recruited to the mRNA, and CLIP-seq analysis reveals that UPF1 binds along the length of the mRNA [91, 92]. The UPF1-bound regions are rich in guanosine residues [91]. In the cytoplasm, the EJC undergoes a compositional switch, interacting with ASAP or PSAP complexes or CASC3-bound complexes (Figure 1.5B). It is hypothesized that the CASC3-bound form of EJC plays an important role in NMD activation [35]. The initiation of NMD occurs when the translating ribosome encounters a PTC, situated approximately 50-55 nucleotides upstream of an exon-exon junction [25] (Figure 1.5C).

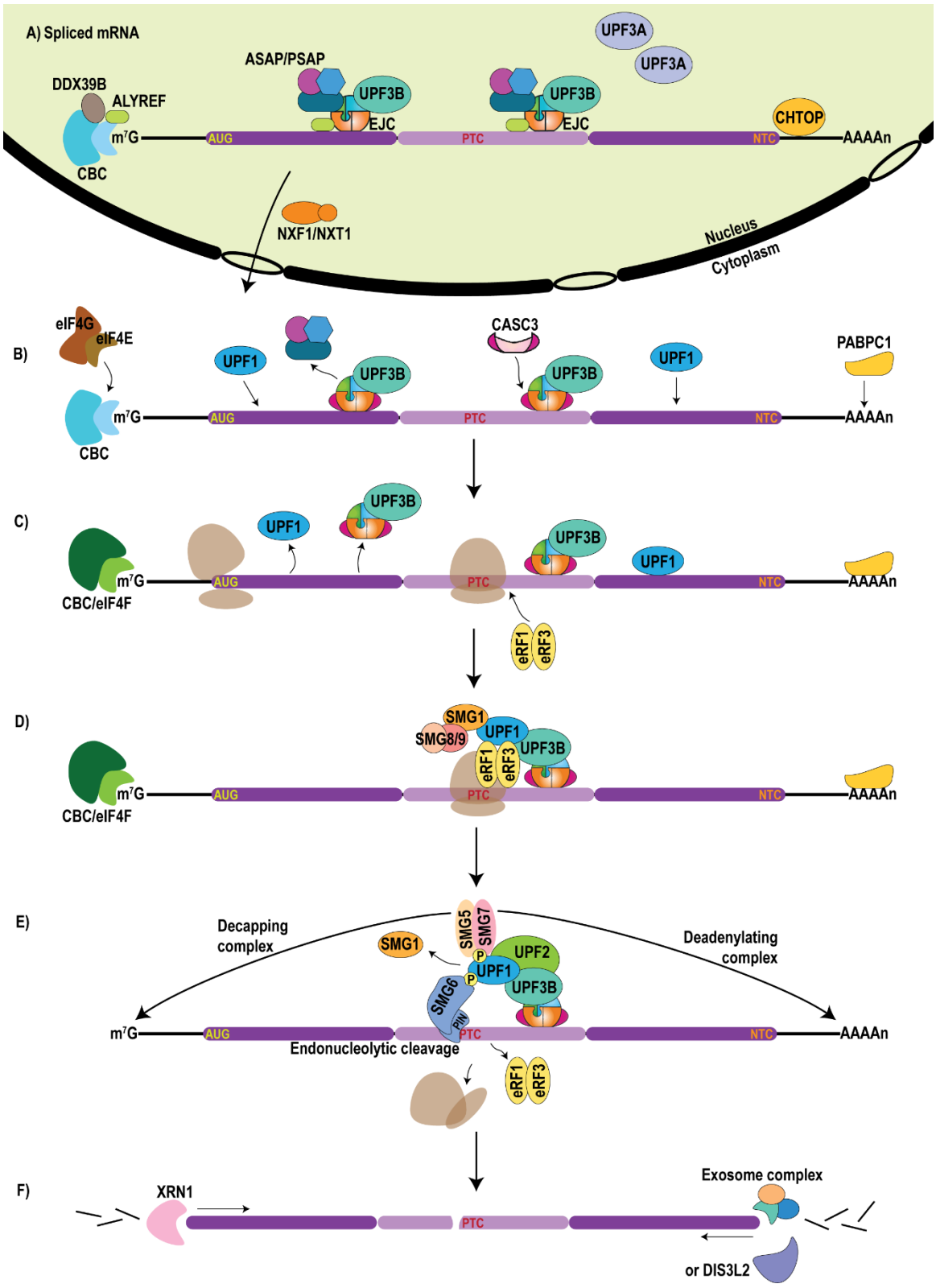


Figure 1.5. UPF3B-dependent NMD pathway. (A) Spliced mRNA contains EJC mark which provides a binding platform for nuclear UPF3B and export factors, (B) cytoplasmic UPF1 gets deposited over mRNA, (C) translating ribosome stalls at PTC, triggers recruitment of UPF1-SMG1 and eRF1-eRF3 dimer, (D) UPF1 interacts with the ternary complex UPF3B-eRF3a-eRF1, (E) the termination complex is dissociated from mRNA, UPF2 joins to form NMD competent complex, SMG1 phosphorylates UPF1, SMG6 binds to phosphorylated UPF1 and perform endonucleolytic cleavage of mRNA near PTC. Phosphorylated UPF1 binds to SMG5-SMG7 heterodimer which recruits deadenylating and decapping complexes leading to (F) exonucleolytic degradation of mRNA via exosome complex, DIS3L2 and XRN1.

The translation termination at a PTC is kinetically slow and may promote a myriad of cellular interactions that favor the assembly of NMD machinery. In a recently modified NMD model, UPF3B, rather than UPF1, is considered the critical factor that reduces termination efficiency at PTC and potentially physically bridges the termination complex with the NMD apparatus [68]. In this context, UPF3B establishes a direct interaction with the eukaryotic translation termination factor, eRF3a (eukaryotic release factors), forming a ternary complex UPF3B-eRF3a-eRF1 [68] (Figure 1.5D). Conversely, UPF1 does not exhibit direct interaction with eRF1 and eRF3a. Therefore, the previously observed association between UPF1 and eRFs, as determined by co-immunoprecipitation experiments, is likely an indirect interaction [68].

Nevertheless, UPF1 does form part of release factor-containing complexes, and this association is likely mediated via direct interaction between UPF1 and UPF3B [68]. Similarly, co-IP experiments have revealed an association between SMG1 and eRFs, but the nature of this interaction, whether direct or indirect remains an open question [69]. It is hypothesized that UPF2 joins the complex after translation termination at the PTC and interacts with both UPF1 and UPF3B, leading to the formation of the UPF1-UPF2-UPF3B-containing NMD competent complex [68, 93]. Subsequently, UPF2 stimulates the kinase activity of SMG1, resulting in the phosphorylation of UPF1 [69, 94] (Figure 1.5E). The kinase activity of SMG1 is negatively regulated by SMG8 and SMG9, which maintain SMG1 in a kinase-inactive conformation [95, 96]. The phosphorylation of UPF1 (p-UPF1) is a crucial step in NMD that instigates a series of events leading to mRNA decay. The phosphorylated UPF1 (p-UPF1) is capable of interacting with the NMD factors SMG6 and the SMG5-SMG7 heterodimer [97-99]. SMG6 mediates the PTC-proximal endonucleolytic cleavage of mRNA through its PIN (PilT N terminus) domain [100]. Additionally, SMG7 recruits the deadenylating CCR4-NOT complex and the decapping protein DCP2 [101, 102]. Furthermore, UPF1 can interact with the decapping factors DCP1, DCP2, and PNRC2 independently of SMG5/SMG7 [101, 103-107]. Following the deadenylation and decapping of mRNA, exonucleolytic degradation of the mRNA is carried out finally by XRN1 (exoribonuclease 1), DIS3L2 (DIS3 Like 3'-5' Exoribonuclease 2), and exonuclease exosome [104, 108-111] (Figure 1.5F).

1.2.3.2. Bi-functional Role of UPF3B in Translation Termination

Evidence suggests that translation termination at a PTC is kinetically slower and less efficient compared to normal termination codons (NTCs) [15, 93, 112-114]. This difference arises from the reduced recruitment of eRFs to the ribosome and the absence of the termination-stimulating protein PABPC1 (cytoplasmic poly(A)-binding protein 1) association with eRFs [113-115]. Neu-Yilik *et al*. employed a fully reconstituted *in vitro* human translation termination system to investigate the mechanistic details of translation termination at a PTC. In this system, they mimicked inefficient translation termination at a PTC by limiting concentration of eRFs and excluding PABPC1 [68]. Surprisingly, UPF3B, among the three human UPF proteins, was found to play a key role in delaying translation termination at a PTC. During the translational delay in the PTC-stimulated condition, UPF3B impedes stop codon recognition and peptidyl-tRNA hydrolysis by eRFs, thereby slowing down translation termination. The underlying mechanism involves UPF3B's interaction with the eRF3a-GTP/eRF1 complex, which sterically hinders stop codon recognition via eRFs. Interestingly, UPF2 inhibits this functional activity of UPF3B, suggesting that UPF2 may act as a modulator of UPF3B activity in translation termination [68]. The physiological significance of UPF3B's ability to delay translation termination at PTC-containing transcripts raises an intriguing question. A likely scenario is that the delay provides a temporal advantage to the NMD machinery, allowing sufficient time for the assembly of NMD-promoting complexes on aberrant mRNA. Therefore, the slowing down of translation termination at PTC may contribute to the assembly of NMD factors, including UPF3B, UPF1, and SMG1.

Interestingly, UPF3B has an additional role in the late phase of translation termination, promoting the dissociation of post-termination ribosomal complexes. In the *in vitro* termination reaction system, Neu-Yilik *et al*. observed that a saturating amount of release factors, in the presence of UPF3B, causes the disassembly of post-termination complexes. UPF3B-induced ribosome dissociation occurs after GTP hydrolysis and the release of nascent peptides through peptidyl-tRNA hydrolysis. This UPF3B function is independent of eRFs binding and is possibly triggered by UPF3B's interaction with the ribosomal subunit at low Mg^{2+} concentrations [33]. Bufton *et al*. demonstrated through surface plasmon resonance (SPR) that the middle domain of UPF3B interacts with both ribosomal subunits (40S and 80S) [82]. However, it remains unclear whether UPF3B's role

in dissociating post-termination ribosomal complex is specific to PTC or is part of a more widespread phenomenon involving both PTC and NTC.

1.2.3.3 Role of UPF3B in Brain Development and Associated Disorders

1.2.3.3.1 *UPF3B is Important for Neural Development and Neuronal Processes*

In addition to the role of NMD in mRNA surveillance, NMD also regulates the expression of ~10% of normal mRNAs [25]. Thus, it comes as no surprise that NMD fine-tunes the expression of genes critical for many developmental events. Accordingly, NMD plays a vital role in neural development, including neuronal differentiation, stemness maintenance, axon guidance, and synaptic plasticity [116-118].

Evidence suggests that the modulation of NMD activity is essential for neuronal differentiation. Lou *et al*. found that a high level of NMD is responsible for maintaining the undifferentiated, stem-like cell state of mouse neural stem cells (mNSCs). NMD retains the stemness of mNSCs by degrading transcripts encoding proliferation inhibitory factors, such as p21, p27, p57, and Mapk6, and neural differentiation factors [118]. One of the well-studied NMD targets is *Smad7* mRNA, which encodes a neural differentiation factor. Smad7 protein is a negative regulator of TGF-β signaling, and several lines of evidence indicate that the repression of TGF-β signaling promotes neural differentiation [119]. In this regard, it was reported that the knockdown of NMD factors like UPF1 and UPF3B in mNSCs upregulated the expression of TGF-β signaling inhibitors, such as Smad7. This reduction in the NMD pathway blocks the TGF-β signaling cascade, thereby triggering the differentiation of neuronal cells. Thus, the downregulation of NMD is essential for stabilizing transcripts necessary for neural differentiation [118]. It was also found that during the maturation of neuronal stem cells, the NMD factors, UPF1, UPF2, and UPF3B, get downregulated [118].

To elucidate the mechanism underlying the regulation of UPF3B and NMD in neural development, Lou *et al*. obtained evidence that the modulation of NMD factors is triggered, at least in part, by neurally expressed miRNAs such as miR-128, miR-124, and miR-9 that target UPF3B [118]. These miRNAs get upregulated in the presence of neural differentiation signals and control the expression of NMD factors. This finding implies that miRNAs are significantly involved in the spatiotemporal regulation of NMD factors. Importantly, UPF3B and UPF1 also negatively regulate miR-128 and miR-9 as deletion of UPF3B or UPF1 in mNSCs induces expression of miRNAs. These results suggest the

presence of an NMD-miRNA negative feedback control loop that directs whether a neural stem cell will retain its stemness or will enter a terminally differentiated state based on the input signal [118].

Interestingly, several reports suggest that NMD or NMD factors have varying effects at different stages of brain development. In support, Jolly *et al*. revealed that UPF3B is required for the differentiation of mouse neural progenitor cells (NPCs) at later phases of neural development. They used mouse NPCs from the E18.5 cortex and found that compromised UPF3B-dependent NMD in NPCs results in increased self-renewal and reduced differentiation of NPCs [120]. On the other hand, Lou *et al*. used mNSCs from the E14 cortex and showed that UPF3B suppresses the differentiation of mNSCs during the initial phases of neural development [118]. Thus, NMD might contribute towards the stemness of multipotent stem cells and is downregulated during the differentiation of stem cells. In contrast, NMD is upregulated to stimulate the differentiation of already committed neural progenitor stem cells at later stages of brain development. These findings suggest that NMD factors are differentially regulated at varying stages of brain development. It will be of interest to identify the regulators of NMD factors at different developmental stages.

Besides regulating mitotic neuronal cells, Jolly *et al*. found that the high activity of UPF3B-dependent NMD is indispensable for post-mitotic neuronal populations, including the mouse adult hippocampus. They showed that loss of UPF3B results in defective neurite growth with reduced axonal length and increased arborization of both axons and dendrites in mouse hippocampal neurons [120]. In line with this, it has been reported that UPF3B is expressed extensively in the mouse hippocampal neurons, especially in dendrites, the vital structure for neuronal signal transmission [121].

Intriguingly, the expression of UPF3B in mature neurons is further regulated by synaptic activity. In this regard, the chronic depolarization of mouse hippocampal neurons reduced the level of UPF3B mRNA [120]. This data suggests that reduced UPF3B might play a vital role in the selective stabilization of specific mRNAs at active synapses. This modulation of NMD factors like UPF3B might contribute towards localized protein translation of selective transcripts at active synapses. It will be exciting to examine transcripts that are targeted by UPF3B at the active synapses. In this regard, Notaras *et al*. revealed that the loss of UPF3B in mouse postsynaptic dendrites results in reduced surface expression of Glutamate Receptor 1 (GLUR1) protein, implying UPF3B regulates GLUR1

signaling. GLUR1 plays an essential role in various forms of synaptic plasticity [122]. Altogether, these results suggest the existence of complex differential expression of UPF3B, which is crucial for both neural development and mature neuronal processes. Nonetheless, despite these observations, the precise molecular mechanism underlying the differential regulation of UPF3B and other NMD factors is still poorly understood.

More recently, Tan *et al*. have shown that UPF3B plays a functional role in the olfactory system, and consequently, Upf3b-null mice display partial olfactory defects [123]. The olfactory system is one of the areas of the nervous system that exhibits lifelong neurogenesis and is hence used as a model system for studying neural development [124]. Tan *et al*. found that the loss of UPF3B causes altered gene expression of mouse mOSNs. One major category of genes regulated by UPF3B in mOSNs is olfactory receptor (Olfr) genes [123]. More than 1000 Olfr genes exist in mice and are involved in detecting a wide range of odor ligands. Interestingly, each OSN expresses only one Olfr gene, a phenomenon known as the "one receptor, one neuron" rule [125-127]. UPF3B might influence the selection of at least 78 Olfr genes so that only a single OLFR be expressed in individual mOSNs. To understand the mechanistic rationale underlying the selection of Olfr genes, Tan *et al*. screened genes to identify regulatory genes that are significantly upregulated in Upf3b-null mice and found *Mafg* and *Irf8* as candidate genes. Accordingly, a model has been proposed in which the transcriptional repressors, IRF8 and MAFG, suppress the transcription of a subset of Olfr genes in OSNs and, thus, regulate the selection of single Olfr genes. *Mafg* and *Irf8* are NMD substrates, so impaired NMD leads to the overexpression of these transcriptional repressors and consequent perturbation in the selection of Olfr genes in OSN [123].

Furthermore, Upf3b-null mice have reduced numbers of horizontal basal cells (HBCs) in the olfactory epithelium, implying that UPF3B functions in the maintenance of reserve stem cells. HBCs are mitotically quiescent stem cells that become proliferative and differentiate into OSN precursor cells only in response to olfactory epithelium injury. Loss of UPF3B also causes alterations in mOSN sub-populations. Taken together, the findings indicate that UPF3B-dependent NMD is critical for cognitive processes such as olfaction [123].

1.2.3.3.2. *Deficient UPF3B and Neurodevelopmental Disorders*

The depletion of UPF3B or generation of non-functional UPF3B is associated with diverse types of abnormalities in the central nervous system. For instance, UPF3B is highly expressed in the hippocampus, and the hippocampus is mainly associated with learning and memory. Hence, it is not surprising that one of the phenotypes observed in patients with deficient UPF3B is intellectual disability (ID).

Approximately 1-2% of the global population suffers from the ID phenotype. X-linked Intellectual Disability (XLID) is a condition of ID due to defects in genes located on the X chromosome, such as UPF3B. XLID occurs predominantly in males. Based on the phenotypes associated with XLID, XLID is subdivided into syndromic and non-syndromic forms. In patients suffering from syndromic XLID, intellectual disability is associated with physical abnormalities or neurological abnormalities. In individuals with non-syndromic XLID, ID is present without any additional abnormalities [128-130].

Tarpey *et al*. made the first significant observation that UPF3B mutations might lead to XLID [130]. A systematic study has been carried out on 250 families with XLID, in which 737 genes present on the X-chromosome were sequenced. Amongst the analyzed samples, patients from four families with mutations in the *UPF3B* gene were identified [130]. These patients are clinically diagnosed with syndromic XLID phenotype [131, 132]. The patients have frameshift mutations leading to PTCs in an alternative reading frame of the *UPF3B* transcript and potentially to a truncated protein with an altered C-terminus (R225fs*20, G290fs*2, and R430*) (Table 1.1) (Figure 1.6). Since these mutations introduce a PTC in the *UPF3B* mRNA, Tarpey *et al*. analyzed the effect of deficient UPF3B on its mutant mRNA. They isolated RNA from lymphoblastoid cell lines (LCLs) of affected individuals to check the *UPF3B* mRNA level. qRT-PCR analysis exhibited decreased PTC-containing *UPF3B* mRNA, indicating the probable degradation of *UPF3B* mRNA via an alternative NMD pathway [130]. Additionally, the truncated UPF3B protein would lack the Y-14 interacting domain [130]. However, Western Blot (WB) analysis did not detect wild-type or truncated UPF3B proteins in LCLs derived from affected individuals. If generated, the truncated UPF3B would lack the ability to interact with EJC and hence, will be unable to execute EJC-enhanced NMD. Thus, these findings indicate the significance of UPF3B-dependent NMD for the proper development and functioning of the brain.

Table 1.1. UPF3B variants and associated neurodevelopmental disorders.

UPF3B Variants (Protein length)	**Mutations/Deletions (Residues changed)**	**Disease-associated**	**References**
Full-length proteins			
Variant 1	478T > G (Y160D)	Non-syndromic ID	[130]
Variant 2	1136G > A (R379H)	Syndromic ID	[121]
Variant 3	1103G > A (R368Q)	Syndromic ID, ASD	[121]
Variant 4	764G > A (R255L)	Schizophrenia	[133]
Variant 5	1101G > C (K367 N)	ID	[134]
Variant 6	883T > A (L295 M)	ID, Infantile spasms	[135]
Other products			
Variant 7 (484 aa)	1310DelC (P437fs*47)	ID	[136]
Variant 8 (429 aa)	1288C > T (R430*)	Syndromic ID	[130, 137]
Variant 9 (364 aa)	1091_1094DelAGAG (E364fs*2)	ID	[138]
Variant 10 (360 aa)	1081C > T (R361*)	Non-syndromic ID	[121]
Variant 11 (292 aa)	IVS8 +1 G-A (L283fs*11)		[139]
Variant 12 (290 aa)	867_868DelAG (G290fs*2)	Syndromic ID	[130]
Variant 13 (263 aa)	697_698DelAG (R233fs*30)	Syndromic ID, ASD	[140]
Variant 14 (263 aa)	684_685DelAA (E230fs*35)	ID	[141]
Variant 15 (245 aa)	674_677DelGAAA (R225fs*20)	Syndromic ID	[130]
Variant 16 (245 aa)	683_686DelAAGA (Q228fs*18)	COS, ADHD, ASD	[142]
Variant 17 (39 aa)	118C > T (Q40*)	Syndromic ID, ASD	[143]
Variant 18	Xq24 loss	Syndromic ID, ASD	[144]

ID – intellectual disability, ASD – autism spectrum disorder, COS – childhood-onset schizophrenia, ADHD – attention deficit hyperactive disorder.

An evaluation of an additional 118 probands by Tarpey *et al*. identified a family with a non-syndromic XLID phenotype. The *UPF3B* gene has a missense mutation (Y160D) in the family [130]. This tyrosine residue is conserved in UPF3B protein across

mammals and plants, and interestingly falls into the eRF3A interaction domain of UPF3B (Table 1.1) (Figure 1.6). Intriguingly, Alrahbeni *et al*. have shown that this missense mutation (Y160D) in UPF3B impairs NMD in rat hippocampal neural stem cells [56]. However, whether this mutation in the eRF3A interaction domain of UPF3B affects proper translation termination at PTC remains to be determined. Another associated question is whether impairment in the translation termination of NMD transcripts results in compromised NMD.

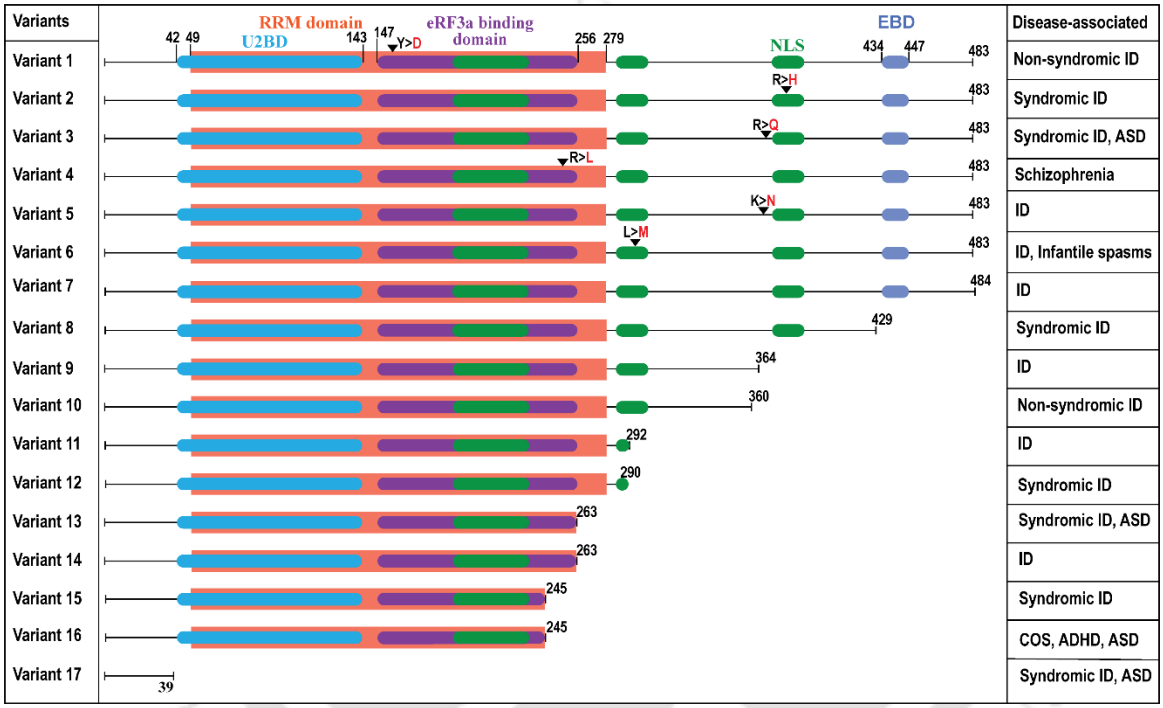


Figure 1.6. Schematic representation of pathogenic variants of UPF3B. Disease-linked missense mutations and truncations due to nonsense/frameshift mutations in UPF3B are depicted. ADHD – attention deficit hyperactive disorder, ASD – autism spectrum disorder, COS – childhood-onset schizophrenia, EBD – EJC binding domain, ID – intellectual disability, NLS – nuclear localization signal, RRM – RNA recognition motif, U2BD – UPF2-binding domain.

Furthermore, a comparative genome hybridization assay identified an Xq21.1e25 (47.7 MB) duplication in a girl with intellectual disability and facial deformities. UPF3B is one of the ten genes linked to the duplication interval [145]. The significance of UPF3B duplication in intellectual disability remains elusive. Perturbation in UPF3B-dependent NMD due to the deregulation of UPF3B (either upregulation or loss of UPF3B) might act as a causative factor for intellectual disability [146].

Several lines of evidence suggest that mutations in the *UPF3B* gene are associated with other neurodevelopmental disorders, including schizophrenia, attention deficit hyperactivity disorder (ADHD), and autism spectrum disorder (ASD). In support of this, two families were identified with two amino acid substitutions (R379H and R368Q) in

UPF3B protein and consequent ID with autism. The amino acid substitutions, R379H and R368Q, are located in NLS and proximal to NLS, respectively [121] (Table 1.1) (Figure 1.6). NLS is mostly required for the shuttling behavior of UPF3B. Although Laumonnier *et al*. did not observe any perceptible change in the localization of this mutant UPF3B protein, it cannot be excluded that a subtle change in the localization of UPF3B might lead to poor neural development. In line with this view, Jolly *et al*. have shown a specific nucleus-cytoplasmic distribution pattern of UPF3B during cortical development in mice [120]. What is the significance underlying differential localization of UPF3B during neurogenesis? One possibility is that the UPF3B-dependent NMD is developmentally regulated by controlling the nuclear-to-cytoplasmic translocation of UPF3B. Hence, any impairment in the regulation of UPF3B-dependent NMD might lead to poor neurogenesis.

Adding up to the findings, the deletion of four bases (683_686DelAAGA) in the *UPF3B* gene caused a frameshift mutation and generation of a PTC-containing transcript (Table 1.1) (Figure 1.6). This mutation is believed to be one of the causative agents of childhood-onset schizophrenia (COS) and ADHD in one child and autism with ADHD in another child from an asymptomatic mother [142]. This *UPF3B* mutant transcript is supposed to be degraded by the NMD pathway. However, it is unclear how the presence of identical UPF3B mutant transcripts leads to the clinical manifestation of different neurodevelopmental disorders.

The diverse nature of *UPF3B* mutations is also observed in another family where two affected siblings show distinct phenotypes. One has microcephaly and remarkable speech delay, while the other suffers from autism [140]. In this regard, deletion in exon 7 of UPF3B (697_698delAG) introduces a PTC, leading to a degraded transcript or truncated product. This protein truncation lacks the EJC binding domain and NLSs [140] (Table 1.1) (Figure 1.6).

The recent development in the Next Generation Sequencing (NGS) has helped researchers identify more and more pathogenic variants of UPF3B [134-139, 141, 143]. However, many UPF3B pathogenic variants identified via high throughput sequencing analysis require additional cohort reports to verify the clinical relevance of the UPF3B variants. Furthermore, functional studies of these variants are necessary to reveal the underlying mechanisms of UPF3B variant pathogenicity. Collectively, Table 1.1 summarizes the pathogenic or likely pathogenic variants of UPF3B and their associated

neurodevelopmental disorders. The fact that many loss-of-function mutations of UPF3B lead to syndromic XLID phenotype suggests that UPF3B might be engaged in regulative functions in neurons that are critical for additional processes, such as neuromuscular processes.

The growing number of evidence indicates loss-of-function mutations of UPF3B as a cause of a broad range of neurodevelopmental disorders, including intellectual disability, schizophrenia, ADHD, and autism. To understand the molecular link between compromised UPF3B-mediated NMD and intellectual disability, Nguyen *et al*. performed transcriptome profiling of LCLs derived from patients with ID [146]. They found that the loss of UPF3B has a direct impact on about 5% of all transcripts. One of the NMD targets, *ARHGAP24* mRNA, is upregulated in LCLs derived from patients with UPF3B loss. ARHGAP24 is a member of the Rho family of the GTPase-activating protein, which upon deregulation in neurons, hinders axon and dendrite outgrowth and branching [146]. Similarly, other differentially expressed genes identified in LCLs include *Robo1* and *Nrcam*, which also have functions in axon guidance and axon growth [146]. As further support, Jolly *et al*. demonstrated the upregulation of *Robo1* and *Nrcam* genes in mouse hippocampal neurons lacking UPF3B [120]. Axon guidance and dendrite branching are imperative for the establishment of functional circuits of the nervous system. Altogether, these results indicate the essential role of UPF3B in axon guidance, axon growth, and dendrite branching. Consequently, loss of UPF3B may lead to aberrant wiring of neurons resulting in various brain disorders.

A significant breakthrough in understanding the molecular rationale underlying UPF3B dysfunction-mediated intellectual disability comes from a study carried out on an Upf3b-null mice model [147]. The study used mice as a model system to understand the molecular basis of UPF3B function in neurodevelopment and neuropathology. These Upf3b-null mice exhibit behavioral defects such as contextual and cued fear-based learning defects but display normal simple working memory and spatial learning. Further, the Upf3b-null mice exhibit reduced acoustic startle response and prepulse inhibition (PPI) defects, which are mostly deficient in patients with schizophrenia and other brain diseases. The startle response is a spontaneous defensive response to an unanticipated stimulus. Interestingly, the morphological analysis of cortical pyramidal neurons of Upf3b-null mice displays reduced dendritic spine density and maturation. This finding raises the possibility

that the abnormal dendritic spine density disrupts the connection between cortical areas and contributes to the observed behavioural defects in Upf3b-null mice. Surprisingly, Upf3b-null mNSCs from the E14.5 cortex differentiate poorly under differentiation conditions [147], which contradicts the previous finding that UPF3B inhibits neural differentiation [118]. Future work will be required to solve this apparent discrepancy.

Upf3b-null mNSCs also display impaired electrical activity, suggesting the integral role of UPF3B in promoting efficient neurogenesis. RNA-Seq analysis of the frontal cortex of Upf3b-null mice identified hundreds of transcripts that could be direct NMD targets. Out of which, 20 transcripts encode proteins that are directly involved in neural processes with known functions like dendritic arborization, promotion of axon growth, synaptic connections, and cellular excitability. Upf3b-KO upregulated genes also include schizophrenia disease risk genes such as *DSCAM*, *HIVEP3*, *MOV10*, *NIN*, *SPEN*, *AKR1C2*, *ARL5C*, *CD59*, and *VWA5B1* [147]. These results may explain the association between deficient UPF3B and brain disease, schizophrenia. In summary, these findings suggest that UPF3B-mediated NMD tightly controls the steady-state levels of specific transcripts involved in neuronal functions and neural development.

All the above findings indicate that UPF3B is involved in neuronal development, and any perturbation of its cellular expression level causes various neuropathologies. Further research towards understanding the role of NMD in healthy brain development will help design NMD modulators for treating neurodevelopmental disorders. For instance, if the truncated or mutated UPF3B protein in a diseased individual is partially functional, then the severity of the neural disease could be lowered by modulators that stabilize the mutated UPF3B transcript in the affected individual. Further research in this direction will provide deeper insights into the effectiveness of using NMD modulators for treating brain disorders.

1.2.4 Functions of UPF3A

This section reviews the functions of the sister paralog of UPF3B, UPF3A.

1.2.4.1 Dual Role of UPF3A in NMD

One of the interesting questions that have concerned researchers was the significance of UPF3 paralog's existence in the genome. Hitherto it has been assumed that the UPF3 paralog pair has a redundant function. Consistent with this, it has been shown that tethering of UPF3A to the 3'UTR of a reporter mRNA elicits NMD albeit less

efficiently than the strong NMD activity of its sister paralog, UPF3B [54, 64]. The findings indicate UPF3A as a weak NMD activator. Subsequently, Chan *et al*. found that UPF3A acts as a redundant NMD factor, partially taking over the NMD activity of UPF3B in response to UPF3B depletion. They also showed that the loss of UPF3B markedly upregulates UPF3A expression by stabilizing UPF3A protein. Mechanistically, free UPF3A is believed to be inherently unstable, and UPF3A remains stable only if bound to UPF2. Notably, the affinity of UPF3B to UPF2 is more efficient than UPF3A, implying the existence of competition between UPF3A and UPF3B for UPF2. As a result, less concentration of UPF3B leads to an increased half-life of UPF3A protein, indicating a unidirectional cross-regulatory relationship between UPF3A and UPF3B [148].

The molecular basis underlying the weak NMD activity of UPF3A is due to a single amino acid change (arginine to alanine) in the EJC binding domain of UPF3A compared to its sister paralog, UPF3B. In fact, substituting arginine in the UPF3A-EJC binding domain increased its NMD activity equivalent to UPF3B [64]. However, this weak NMD activity of UPF3A perplexed many researchers, given that UPF3A has persisted since the origin of vertebrates. Furthermore, the C-terminal sequence of UPF3A harboring the EJC binding domain is highly variable across vertebrates and is even absent in rats and mice [64]. Altogether these raise the question as to why evolution has selected the weak NMD activity of UPF3A if the sole purpose of UPF3A is to balance the gene dosage of UPF3 protein.

Interestingly, Shum *et al*. reassessed the function of UPF3A by performing loss of function studies [80]. If UPF3A functions primarily as an NMD-promoting factor, then the loss of UPF3A should result in the upregulation of NMD substrates due to dampened NMD. On the contrary, it was found that the depletion of UPF3A led to the downregulation of a large number of NMD substrates, indicating that UPF3A stabilizes NMD target RNAs and functions as an anti-NMD factor. To explore the transcriptome-wide effect of UPF3A on NMD, an RNA-seq half-life assay was performed in UPF3A-depleted cells [80]. A significant fraction of NMD targets were destabilized upon depletion of UPF3A, clearly indicating the role of UPF3A as a global repressor of NMD. In an attempt to understand the molecular mechanism underlying the role of UPF3A in NMD repression, Shum *et al*. found that an isoform of UPF3A, lacking a functional UPF2-interacting domain, fails to act as an NMD repressor. This data implies the significance of UPF3A binding to UPF2 in repressing NMD. *In vivo* tethering assay showed that deficient UPF3A did not alter the

NMD activity of tethered UPF2. On the other hand, the co-immunoprecipitation assay indicated that UPF3A restrains the formation of functional NMD-promoting complex as depletion of UPF3A enhances the interaction of EJC with the central NMD factor UPF1. Altogether, these findings supported the notion that UPF3A does not hinder the ability of already assembled UPF2 to recruit other NMD factors, rather UPF3A sequesters freely available UPF2, thereby impairing the NMD machinery [80].

Further, it was found that the EJC binding domain of UPF3A is dispensable for the NMD repression activity of UPF3A [80] (Figure 1.3C). This observation also explains the poorly conserved nature of the C-terminal part of UPF3A, which harbors the EJC binding domain. Nevertheless, the weak EJC binding domain confers an additional benefit to UPF3A in limiting its activity as an NMD-promoting factor. Indeed, a domain-swapping experiment showed that replacing UPF3A's weak EJC binding domain with UPF3B's strong EJC binding domain can transform UPF3A into an NMD stimulating factor [64]. These studies collectively indicate that UPF3A has a dual function in NMD by acting as both a weak NMD activator and a potent NMD repressor. This ambivalent functional aspect of UPF3A is probably controlled by different physiological conditions and stages of development. One of the factors that might influence the dual role of UPF3A in NMD is the 3'UTR length of the target mRNA. Shum *et al*. found that the 3'UTR length of UPF3A-stabilized mRNAs is significantly longer than UPF3A-destabilized mRNAs [80].

More recently, three studies revisited the function of UPF3A in NMD [149-151]. Wallmeroth *et al*. observed that the overexpression or knockout of UPF3A in human HEK-293 and HeLa cell lines does not alter the NMD activity substantially [150]. Simultaneously, Yi *et al*. showed that UPF3A has weak NMD activity [149]. UPF3A partially compensates for the loss of UPF3B in humans, and double knockout of UPF3B paralogs in cultured cells has profound NMD defects [149, 150]. Furthermore, Chen *et al*. generated a conditional *Upf3a*-knockout mouse to study the role of UPF3A in mammalian NMD and established several mouse embryonic stem cells and somatic cells devoid of UPF3A [151]. UPF3A is dispensable in the presence of UPF3B, and UPF3B-knockdown in UPF3A-knockout mESCs causes the accumulation of PTC-containing transcripts [151].

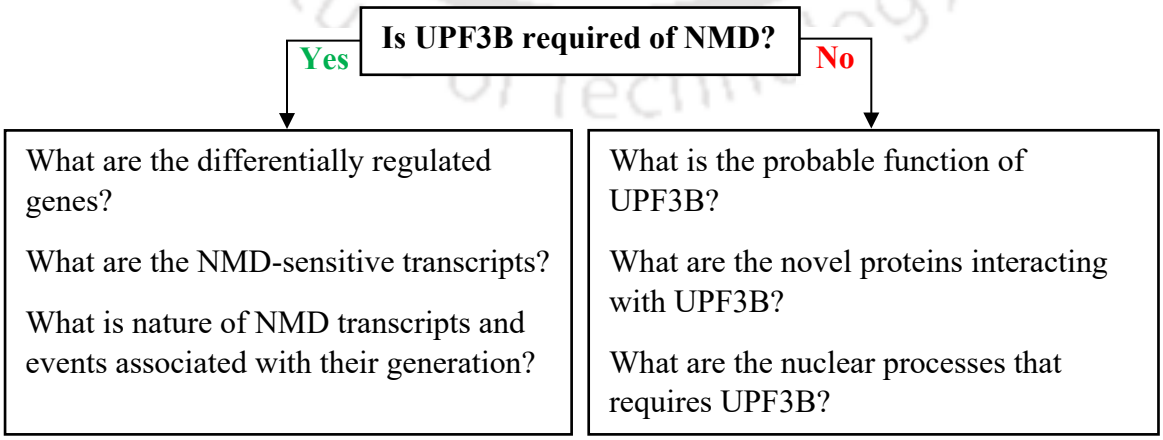
Since the involvement of UPF3 paralogs in NMD is a subject of ongoing debate as some studies consider UPF3A functions as a weak NMD activator in the absence of UPF3B, and its deletion leads to substantial NMD defects. Whereas other studies have

projected that UPF3B does not significantly enhance NMD. These conflicting results leave open research questions about the molecular understanding of UPF3B's involvement in NMD and its influence on the cellular transcriptome.

1.3 Research Gap and Objectives

In order to elucidate the functional role of UPF3B in neuronal growth, I aimed to create knockout (KO) cell lines devoid of UPF3B. While previous studies in mice cells demonstrated neuronal defects upon UPF3B depletion [147], no such investigation had been conducted in human neuronal cell lines. To address this research gap, I sought to establish a UPF3B-KO system to examine the neurodevelopmental implications of UPF3B loss of function mutations. My initial attempts to generate UPF3B-KO cell lines involved the use of the SHSY-5Y neuroblastoma cell line. Unfortunately, these efforts proved unsuccessful as cellular senescence was observed following transfection via lipofection and electroporation methods (APPENDIX I).

Consequently, the HEK-293 cell line, commonly employed in neurobiological studies [152-154], was selected for the generation of UPF3B-KO cells. HEK-293 cells share a neural crest origin with neurons [155], suggesting similarity in biological processes. At the initiation of my research thesis in 2016, the functional role of UPF3B in NMD was a subject of debate, with conflicting findings in the literature. For example, Wallmeroth *et al.* and Yi *et al.* showed that UPF3B is not required for NMD [149, 150], whereas Chan *et al.* asserted that UPF3B is an essential factor [84]. These contrasting data prompted a systematic investigation into UPF3B, leading to the formulation of the following research gaps:



To address these research gaps, I defined the following research objectives:

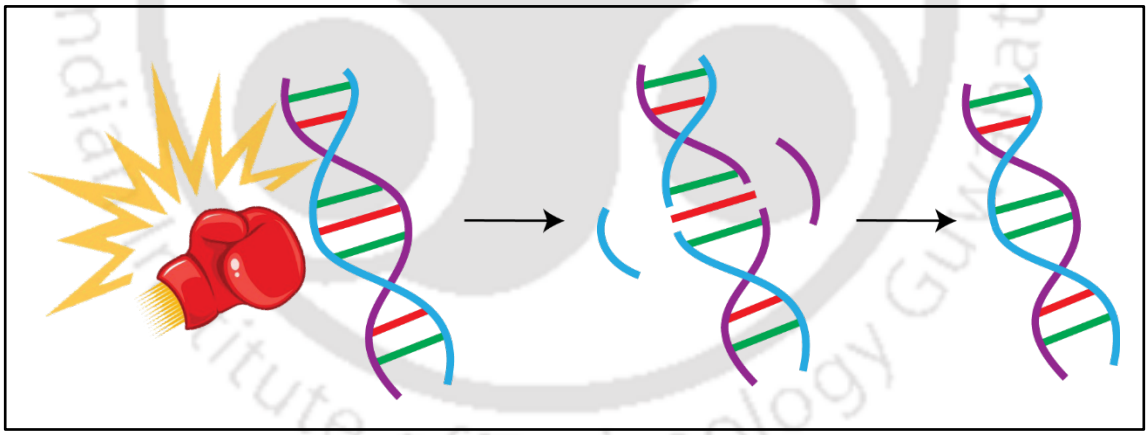
1. To generate UPF3B-knockout HEK-293 cell lines.
2. To perform the global gene expression analysis of UPF3B-KO cells.
3. To investigate the impact of UPF3B-KO on NMD impairment and cell proliferation.
4. To identify NMD-sensitive transcripts targeted by UPF3B-mediated NMD.

For the first objective, the CRISPR/Cas9 gene editing method was employed to generate UPF3B-KO HEK-293 cell lines. To study the effect of UPF3B deletion on gene expression, RNA sequencing of UPF3B-KO cells was performed, and computational analysis was carried out to identify the differentially expressed genes in the second objective.

Since UPF3B is a component of the NMD complex and is involved in several branches of the NMD pathway, in the third objective, NMD impairment was accessed in the absence of UPF3B, and its effect on cell growth was determined. As for impaired NMD activity in UPF3B-KO cells, the fourth objective aimed to identify the NMD-sensitive transcripts that are upregulated and stabilized upon UPF3B-KO.

Chapter 2:

Generation of UPF3B-knockout HEK-293 Cell Line



Chapter 2

Abstract

The UPF3B protein is an integral component of the NMD complex, which assembles over an aberrant transcript for its degradation. Loss-of-function mutations in the *UPF3B* gene are implicated in patients with neurodevelopmental disorders. To investigate the repercussions of UPF3B loss, a UPF3B-knockout HEK-293 cell line was generated using the CRISPR-Cas9 genome editing method. This chapter outlines the *UPF3B* gene knockout strategy, involving the establishment of a stable Cas9-inducible expression system, followed by UPF3B knockout using a single guide RNA. With the implementation of this strategy, two UPF3B-KO clones were generated with 55- and 5-base pair deletions in the *UPF3B* coding sequences. These deletions in the *UPF3B* gene sequence introduce premature termination codons, resulting in a deficiency of the UPF3B protein. This mammalian system devoid of UPF3B protein provides a platform to understand the biological role of UPF3B in human cell lines.

Chapter 2

2.1 Introduction

The UPF3 paralogs, UPF3A and UPF3B, redundantly assemble as part of the SURF complex, facilitating mRNA surveillance via the NMD pathway and contributing to translation fidelity [64, 68]. Pathogenic variants of UPF3B, resulting from mutations and deletions in the *UPF3B* gene, have been identified in patients with neurodevelopmental disorders (NDDs) [156]. These pathogenic variants cause a loss of UPF3B function, leading to the disease condition in patients. However, the specific functional role of UPF3B in NDDs remains undetermined. To gain mechanistic insight into UPF3B's loss-of-function in humans, a mammalian system lacking UPF3B can be generated. Unlike other UPF proteins, knockout of UPF3B is viable in mice and manifests with mild neurological symptoms [147]. Nevertheless, the distinct gene structures and protein functions between mice and humans suggest that UPF3B may have differential functions in the human context.

The first attempt to investigate the role of UPF3B in human cell lines was made by Chan *et al.*, who performed shRNA-mediated knockdown of UPF3B in HeLa cells [84]. However, the knockdown studies may not reflect the loss of protein function, as residual protein levels might still allow normal protein function. Therefore, achieving the absolute loss of UPF3B is essential to precisely delineate its role. At the time of formulating my research objectives, there was a lack of a mammalian system deficient in UPF3B expression. However, in the latter stages of my thesis, two research groups independently generated knockouts in cultured human cells to investigate the NMD functions of UPF3 paralogs [149, 150]. Wallmeroth *et al*. utilized siRNA-mediated knockdown of UPF3B on the CRISPR/Cas9-mediated knockout of UPF3A in HEK-293 and HeLa cells [150], revealing that depletion of both UPF3 paralogs impairs NMD. Yi *et al.* generated two UPF3B-KO clones in HCT-116 cells, one expressing a shorter UPF3B protein lacking the RRM domain and the other with no UPF3B expression [149]. Notably, NMD remained functional upon UPF3B-KO in HEK-293 cells [150], while it showed mild defects in HCT-116 cells [149], suggesting potential cell-type specificity in NMD. These discrepancies in research findings regarding the role of UPF3B in NMD activity necessitate a thorough investigation. Therefore, this study aims to generate UPF3B-knockouts using the CRISPR/Cas9-mediated gene editing method.

The advent of CRISPR/Cas9 genome editing technology has enabled researchers to investigate the loss of protein function in animal and cell line model systems [157, 158]. This technology involves the use of a guide RNA (gRNA) that associates with the Cas9 protein, forming an RNA-protein (RNP) complex. This RNP complex scans the host genome for nucleotide sequences complementary to the gRNA, inducing a double-strand break that is repaired through non-homologous end joining (NHEJ). In this chapter, I have generated a UPF3B-knockout HEK-293 cell line, outlining a detailed procedure for utilizing the CRISPR/Cas9 technique. The knockout strategy involved establishing a stable Cas9-inducible expression system in HEK-293 cells. Subsequently, these cells were transfected with a single guide RNA to knock out UPF3B.

2.2 Materials and Methods

2.2.1 Mammalian Cell Culture

Flp-in T-REx HEK-293 cells (Invitrogen, Cat# R78007) were cultured in DMEM, High glucose (Himedia, Cat# AT007F) supplemented with 10% FBS (Himedia, Cat# RM10432) and 1% Penicillin-Streptomycin solution (Himedia, Cat# A014). Cells were incubated in humidified conditions at 37 °C with a 5% CO_2 supply.

2.2.2 Plasmids

The three plasmids were procured from Addgene and System Biosciences. a) px330 plasmid (Addgene, Cat# 42230) containing the human codon-optimized *Streptococcus pyrogens* Cas9 (hSpCas9) gene sequence. b) PB-CuO-MCS-IRES-GFP-EF1α-CymR-Puro plasmid (System Biosciences, Cat# PBQM812A) containing an inducible promoter, multiple cloning sites (MCS), GFP-tag, and Puromycin resistance gene. c) pCI-FLAG-PB-Transposase plasmid encoding the transposase enzyme.

2.2.3 Recombinant Plasmid Construction

The px330 plasmid was used as a template to amplify the *hSpCas9* gene, utilizing the forward primer containing the NheI restriction site, 5'-CCTAAGCTAGCATGGACTATAAGGACCACGAC-3' and the reverse primer containing the NotI restriction site, 5'-TTCCTGCGGCCGCTCCCCAG-3', under the following conditions: Initial denaturation at 98°C for 2 min, 30 cycles of denaturation at 98°C for 1 min, annealing at 70°C for 45 s, extension at 72°C for 4 min 30 s, and final extension at 72°C for 10 min. The purified PCR product (~4.4 kb size) was digested using a pair of restriction enzymes, NheI and NotI. The digested *hSpCas9* amplicons were ligated

into the PB-CuO-MCS-IRES-GFP-EF1α-CymR-Puro expression vector that had been similarly digested.

2.2.4 Competent Cell Preparation

The glycerol stock of *Escherichia coli* Top10 or DH5α was revived on the LB agar plate. A single colony was picked up and inoculated in 5 mL of LB broth for primary culture. The primary culture was incubated overnight at 37 °C in a shaking incubator at 220 rpm. 500 µL of overnight-grown culture was inoculated into a 250 mL conical flask containing 50 mL of LB media and incubated at 37 °C till the OD reached ~0.35-0.4 at 600 nm. The culture was aliquoted in pre-chilled tubes and harvested by centrifugation at 4000 rpm for 10 min at 4 °C. The cell pellet was resuspended in 25 ml of ice-cold $CaCl_2$ buffer (100 mM $CaCl_2$ and 10 mM Tris pH 7.4) and incubated on ice for 45 minutes. The cells were harvested by centrifugation at 4000 rpm for 10 min at 4 °C. The cell pellet was resuspended in 2.5 mL of ice-cold $CaCl_2$ buffer with glycerol, aliquoted into pre-chilled microcentrifuge tubes, and stored at −80 °C.

2.2.5 Transformation of Recombinant Plasmids into Competent Cells

The recombinant plasmids (20 µL of the ligated product from Section 2.2.3) and control plasmids were added to 100 µL of the competent cells. The tubes were incubated on ice for 30 minutes. Heat shock was given at 42 °C for 45 s, and tubes were snap-cooled on ice for 10 min. The cells were spread on LBA plates containing 100 µg/mL Ampicillin (Himedia, Cat# TC021) and incubated at 37 °C for 16 hrs. Positive colonies were verified by colony PCR and double digestion using NheI and NotI restriction enzymes.

2.2.6 Plasmid Isolation

Plasmid DNA was extracted using the Qiaprep Plasmid Miniprep kit (QIAGEN, Cat# 27106) or the NucleoBond Xtra Midiprep kit (Macherey Nagel, Cat# 740410) according to the manufacturer's instructions. In general, 10 mL of bacterial overnight culture in LB broth supplemented with 100 µg/mL Ampicillin was used for mini-scale plasmid isolation, and 100 mL of bacterial overnight culture was used for midi-scale plasmid isolation.

2.2.7 Transfection of Recombinants to HEK-293 Cells

The day before transfection, 1×10^6 HEK-293 cells were seeded in a 35 mm dish to achieve ~75% confluency on the day of transfection. The medium was changed 2 hours before the transfection. The recombinant plasmid PB-CuO-hSpCas9-IRES-GFP-EF1α-

CymR-Puro (generated in Section 2.2.3) was transfected together with the pCI-FLAG-PB-Transposase plasmid. Transfection was carried out using PEI transfection reagent (Polysciences, Inc., Cat# 24765-1), as per the manufacturer's instructions. In brief, 1.25 µg of PB-CuO-hSpCas9-IRES-GFP-EF1α-CymR-Puro plus 1.25 µg of pCI-FLAG-PB-Transposase were diluted in 5 µL PEI and 115 µL Opti-MEM (Cat# 31985088, Gibco, Thermo Fisher Scientific). The transfection mixture was vortexed, incubated at RT for 15 min, and added dropwise to the cells. The medium of the transfected cells was changed after 6-8 hours.

2.2.8 Generation of Stable Cells Expressing Inducible hSpCas9

72 hours post-transfection, cells were washed with Dulbecco's phosphate-buffered saline (DPBS), trypsinized, and transferred to a new 35 mm dish containing DMEM medium with 2 µg/mL Puromycin (Cat# TC198; Himedia). The cells were grown in the selection medium for 14 days. Single colonies were picked and expanded in the medium supplemented with an increased concentration of Puromycin (3 µg/mL) for 5 days to obtain a stable cell line.

2.2.9 Fluorescent Microscopy

1×10^6 stable cells (explained in Section 2.2.8) were seeded in two 35 mm dishes containing DMEM medium. 300 µg of Cumate (Cat# QM100A-1-SBI; SBI System Biosciences) was used to induce the expression of hSpCas9. 24- and 48-hours post-induction, cells were visualized for GFP signal under the fluorescent microscope (Axiovert A1 FL-LED; ZEISS) at 20x magnification using a blue filter (475 nm).

2.2.10 Preparation of Cell Lysates

Stable cells were induced with or without cumate (300 µg). After 48 hours, the cells were washed twice with DPBS and scraped from the cell culture dish in RIPA buffer [1 M Tris (pH 7.4), 5 M NaCl, Nonidet P-40 (Himedia, Cat# MB143), 10% Sodium Deoxycholate, 10% SDS, 0.5 M EDTA (pH 8.0), 100 mM PMSF, Protease Inhibitor Cocktail, and dH_2O]. The cells were lysed by sonication using an ultrasonic homogenizer (Fisher Scientific). The homogenate was centrifuged at 10,000 rpm in refrigerated conditions, and the supernatant was used for analysis.

2.2.11 Measurement of Protein Concentration

The protein concentration of whole-cell lysates was measured using a Pierce™ Detergent compatible Bradford Assay reagent (Invitrogen, Cat# 23246). 5 µL of lysate was

added to 245 µL of Bradford reagent in a 96-well plate, and absorbance was measured at 595 nm in a microplate reader (Bio-Rad). The RIPA buffer was used as a blank for the measurement.

2.2.12 SDS Polyacrylamide Gel Electrophoresis

A polyacrylamide gel was prepared with 4% stacking gel (500 µL 5x Taurine, 315 µL Tris buffer pH 6.8, 330 µL 30% Acrylamide, 25 µL 10% SDS, 25 µL 10% APS, 6 µL TEMED, 1.3 mL dH_2O) and 12% resolving gel (1 mL 5x Taurine, 1.25 mL Tris buffer pH 8.8, 2 mL 30% Acrylamide, 50 µL 10% SDS, 50 µL 10% APS, 8 µL TEMED, 650 µL dH_2O). The cell lysates (20 µg each) were resolved on SDS-PAGE gel using 1x SDS Running buffer (for 1 L 10x SDS Running buffer: 30.2 g Tris-base, 188 g Glycine, 10 g SDS) in a vertical electrophoresis tank (Bio-Rad) at 220 V for 45 min.

2.2.13 Western Blotting for hSpCas9 Expression

The resolved protein samples were transferred to a nitrocellulose membrane (Axiva, Cat# S020A330RI) from the gel using 1x transfer buffer (for 150 mL: 15 mL 10x Transfer buffer, 30 mL Methanol, 105 mL dH_2O) in a PierceTM Power blotter (Thermo Fisher Scientific) semi-dry system at 19 V and 0.5 A for 25 min. The transferred bands were visualized using Ponceau staining (for a 50 mL staining solution: 0.25g Ponceau S and 0.5 mL acetic acid were added to 49.5 mL of dH_2O). The membrane was washed with TBST (20 mM Tris base, 150 mM, pH 7.5; 0.1% w/v Tween-20) and blocked with 5% (w/v) skim milk in TBST (1 gm milk in 20 mL TBST) for 1 hour. After blocking, the membrane was washed and cut depending on the protein size. The membranes were probed with anti-FLAG (1:2000 dilution, Santa Cruz Biotechnology) and anti-Tubulin (1:10,000 dilution, Sigma) antibodies overnight at 4 °C. Both primary antibodies were diluted in TBST with 1% BSA (0.6 gm BSA in 6 mL TBST). The membranes were washed with TBST, and an anti-mouse secondary antibody (1:10,000 dilution; Cell Signaling Technology) prepared in 1% skim milk in TBST (0.6 gm milk in 6 mL TBST) was used to probe the membranes for 1 hour at room temperature. The membranes were visualized using the Novex™ ECL HRP Chemiluminescent Substrate Reagent Kit (Invitrogen, Cat# WP20005) in the ChemiDoc™ XRS+ system (Bio-Rad) with Image Lab™ software.

2.2.14 Single Guide RNA Transfection

A predesigned CRISPR RNA with sequence 5'-rArCrCrUrCrGrGrGrGrGrArCrArGrCrUrCrCrArArGrUrUrUrUrArGrArGrCrUrArUrGrCrU-3' targeting the *UPF3B* gene

(Cat# Hs.Cas9.UPF3B.1.AA) was ordered from Integrated DNA Technologies. Lyophilized *UPF3B* crRNA (Alt-R CRISPR-Cas9 UPF3B crRNA; 2 nM), negative control crRNA (Alt-R® CRISPR-Cas9 Negative Control crRNA #1, 2 nM; IDT, Cat#1072544) and tracrRNA (Alt-R CRISPR-Cas9 tracrRNA - ATTO 550; 5 nM) were resuspended in Nuclease-Free Duplex Buffer (Cat# 11-01-03-01, IDT) to prepare 100 µM stock solutions. Single guide RNAs (sgRNA) for the negative control and *UPF3B* were assembled by combining the crRNA with trans-crRNA in equimolar concentrations. Here, 1 µL of 100 µM crRNA and tracrRNA were mixed with 98 µL Nuclease-free duplex buffer. The mixture was heated at 95 °C for 5 minutes and lowered to room temperature to form 1 µM sgRNA. Approximately 40,000 stable cells with hSpCas9 integration (generated in Section 2.2.8) were seeded in a 96-well plate and supplemented with DMEM medium containing Cumate. After 24 hours of induction, cells were transfected with negative control sgRNAs and *UPF3B* sgRNAs. Transfection mixtures were prepared by mixing 4.5 µL sgRNA (1 µM) and 1.2 µL Lipofectamine™ RNAiMAX reagent (Cat# 13778150, Invitrogen, Thermo Fisher Scientific) diluted in 44.3 µL Opti-MEM. Scramble sgRNA was used as a negative control, and *UPF3B* sgRNA was used to make double-strand breaks in the *UPF3B* gene. The reaction was mixed by vortexing, incubated for 15 minutes at RT, and added dropwise to the cells.

2.2.15 Selection of Knockout Clones

After 72 hours of sgRNA transfection, cells were serially diluted in a 96-well plate. The wells with single cells were identified and marked. Single cells were grown, expanded, and divided into two dishes. Cells from one dish were stored using a freezing medium (90% FBS and 10% DMSO) in a deep freezer (−80°C), whereas cells from the other dish were used for further analysis.

2.2.16 Western Blotting for Knockout Detection

Colonies grown from a single cell were lysed with RIPA buffer (refer to Section 2.2.10), and the cell lysates were resolved on a 12% polyacrylamide gel. The resolved proteins were transferred to the nitrocellulose membrane. Subsequently, the membrane was blocked with skim milk in TBST. Primary anti-UPF3B (1:2000 dilution) and anti- Histone (1:10,000 dilution; BioBharati) antibodies were used to probe the membrane. The membranes were washed, and a secondary anti-rabbit (1:10,000 dilution, Cell Signaling Technology) antibody was used to probe the membranes for 1 hour at RT. The membranes

were visualized using the Novex™ ECL HRP Chemiluminescent Substrate Reagent Kit in the ChemiDoc™ XRS+ system with Image Lab™ software (as stated in Section 2.2.13).

2.2.17 Genomic DNA Isolation

Colonies obtained from the scrambled negative control and UPF3B-KO were grown in T-25 flasks. Cell homogenates were prepared by adding 1 mL of TRIzol™ reagent (Invitrogen, Cat# 15596018) to the cells and incubated at RT for 5 min. 0.2 mL of chloroform (Himedia, Cat# MB109) was added to each sample, mixed vigorously, and incubated at RT for 5 min. Homogenates were centrifuged at 12,000 x g for 15 minutes at 4 °C. The upper aqueous phase was removed completely. 0.3 mL of ice-cold ethanol (Himedia, Cat# MB228) was added to the remaining organic phase, mixed by inverting and incubated at RT for 5 minutes. Samples were centrifuged at 2000 x g for 5 minutes at 4 °C, and the supernatant was removed. The DNA pellet was washed with 1 mL of sodium citrate/ethanol solution (0.1 M sodium citrate in 10% ethanol; pH 8.5) for 30 minutes at RT with gentle inversion occasionally. The sample was centrifuged at 2000 x g for 5 minutes at 4 °C, and the supernatant was removed. The washing step was repeated twice. After washing, 1.5 mL of 75% ethanol was added to the pellet, incubated for 10-20 minutes at RT, and centrifuged at 2000 x g for 5 minutes at 4 °C. The supernatant was discarded. The DNA pellet was air-dried and resuspended in 100 µL of TE buffer (pH 8.0). The concentration of purified DNA was measured with a nanodrop spectrophotometer (Implen, Inc).

2.2.16 Polymerase Chain Reaction and Sanger Sequencing

Genomic DNA isolated from the colonies (negative and KO) in the previous section was used as a template. PCR amplification of the *UPF3B* gene fragment was performed with the following set of primers: FP, 5'- GCATTTGTTGGGGGCGGAGCTTTGA-3' and reverse primer (RP), 5'-CCCATTGCTCTCAAAATAGGATGCTCTCGA-3'. The amplification was performed under the following reaction conditions: Initial denaturation at 98 °C for 2 min; 35 cycles with denaturation at 98 °C for 45 s, primer annealing at 64 °C for 45 s, extension at 72 °C for 45 s; and final extension at 72 °C for 10 min. PCR amplicons were resolved on a 1% agarose gel. The bands were excised from the gel and purified with the QIAprep gel purification kit, following the manufacturer's protocol. Purified amplicons were sent for Sanger sequencing.

2.3 Results

In an effort to understand the functional role of UPF3B in NMD, a UPF3B-deficient HEK-293 cell line was generated. A CRISPR/Cas9-mediated gene knockout approach was utilized to develop the cell line devoid of UPF3B. The gene-knockout strategy that worked in the laboratory setting is summarized in Figure 2.1.

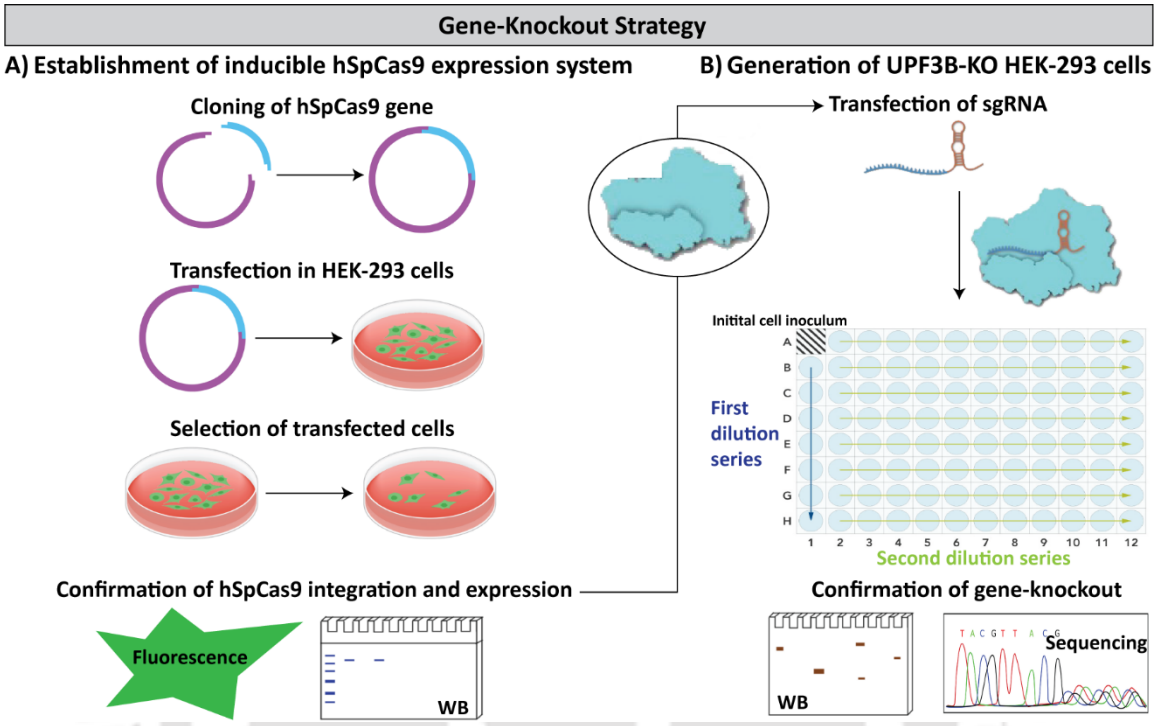


Figure 2.1. Gene-knockout strategy to generate a UPF3B-deficient system. Steps involved in the A) establishment of inducible *hSpCas9* expression system, and B) generation of UPF3B-KO HEK-293 cells are illustrated.

2.3.1 Establishment of Inducible hSpCas9 Expression System

In order to establish an hSpCas9-expressing cell line, the PiggyBac transposon vector system was employed. This system utilizes a PiggyBac transposon vector carrying the gene of interest and a PB-Transposase vector encoding the transposase enzyme (Figure 2.2). The PiggyBac transposon vector functions as a mobile genetic element, readily transposing between vector and chromosome through a "cut and paste" mechanism.

2.3.1.1 Cloning of the *hSpCas9* Gene in PiggyBac Inducible Vector

To clone the gene of interest, the *hSpCas9* gene, into the PiggyBac transposon vector, a px330 plasmid (8.4 kb in size) was used as the template (Figure 2.3A). The px330 plasmid contains 3xFLAG and SV40 nuclear localization signal (NLS) sequences upstream of the *hSpCas9* gene sequence, whereas nucleoplasmin NLS and bGH Poly(A) signal are positioned downstream. For amplifying the *hSpCas9* gene sequence, a specific set of

primers was used (see Materials and Methods; APPENDIX IV). The forward primer, with the NheI restriction site, was complementary to 3xFLAG sequences, and the reverse primer, with the NotI site, was complementary to bGH Poly(A) sequences. PCR amplification produced a 4.4 kb fragment corresponding to FLAG-*hSpCas9* sequences (Figure 2.3B).

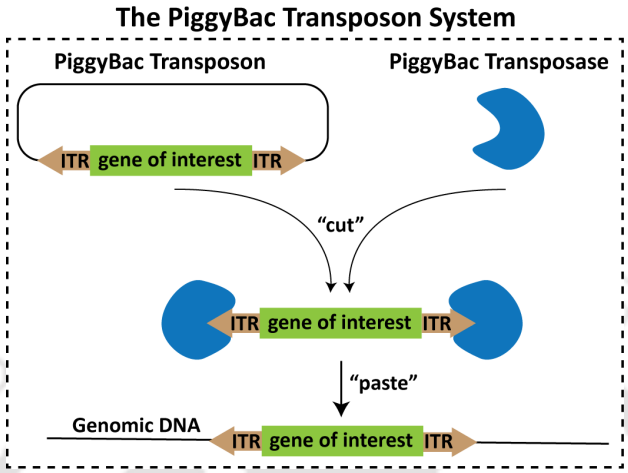


Figure 2.2. The PiggyBac transposon system.

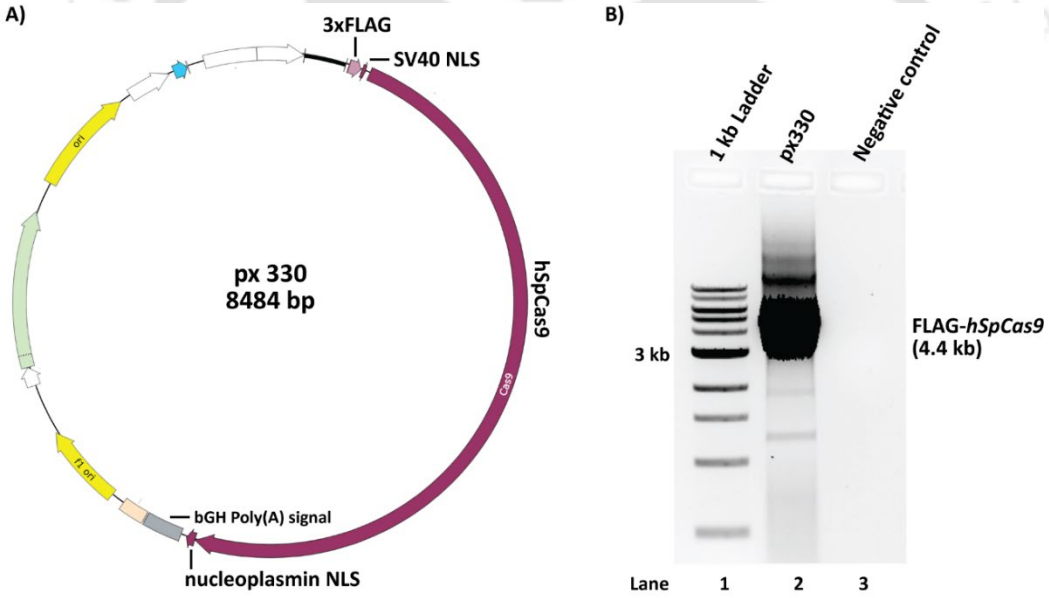


Figure 2.3. Amplification of hSpCas9 gene. A) Schematic diagram of px330 plasmid showing only the sequences used for cloning. B) An agarose gel showing the amplified fragment of FLAG-Cas9 of 4.4 kb in size. Lane 1 – 1 kb DNA Ladder (NEB), Lane 2 – PCR product, Lane 3 – Non-template negative control.

The extracted and purified FLAG-*hSpCas9* fragments obtained from amplification were digested using restriction enzymes NheI and NotI. The PiggyBac transposon vector was similarly digested using the same set of enzymes, as NheI and NotI sites were present in multiple cloning sites (MCS) of the vector (Figure 2.4).

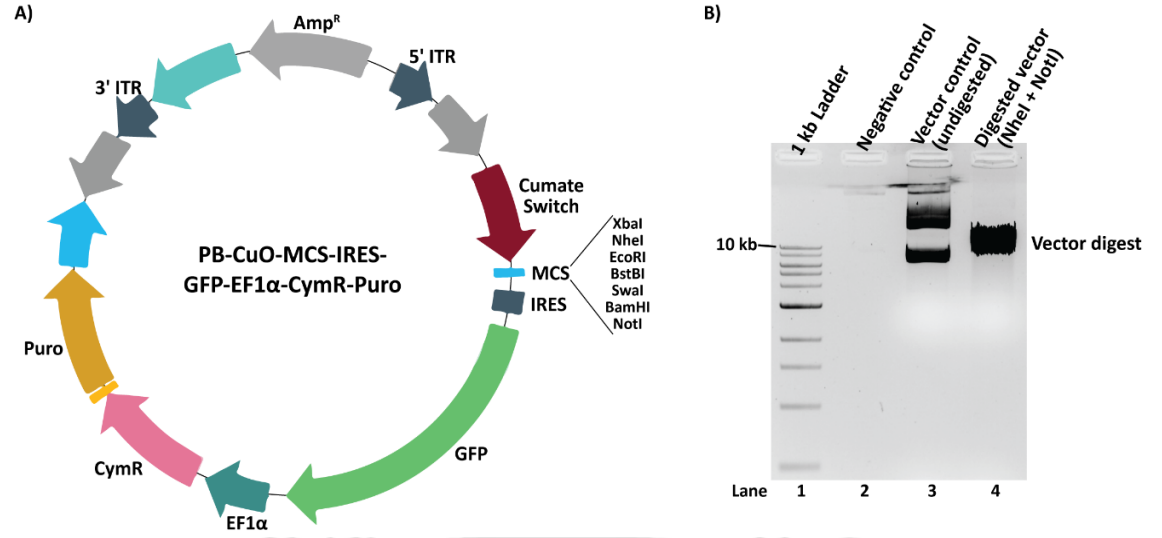


Figure 2.4. Double digestion of PiggyBac vector. A) Schematic diagram of the PiggyBac transposon vector. B) Agarose gel showing PiggyBac transposon vector restriction digestion using NheI and NotI restriction enzymes. Lane 1 – 1 kb DNA Ladder, Lane 2 – non-template negative control, Lane 3 – undigested vector, Lane 4 – double-digested vector.

The double digestion with the pair of restriction enzymes resulted in sticky ends for both the FLAG-*hSpCas9* fragments (4.4 kb) and the linearized PiggyBac transposon vector (8.4 kb). Subsequently, the FLAG-*hSpCas9* insert and PiggyBac transposon vector were ligated using the T4 DNA ligase enzyme (Figure 2.5A). Successful ligation was confirmed through double digestion, yielding a 4.4 kb insert and an 8.4 kb vector digest (Figure 2.5B).

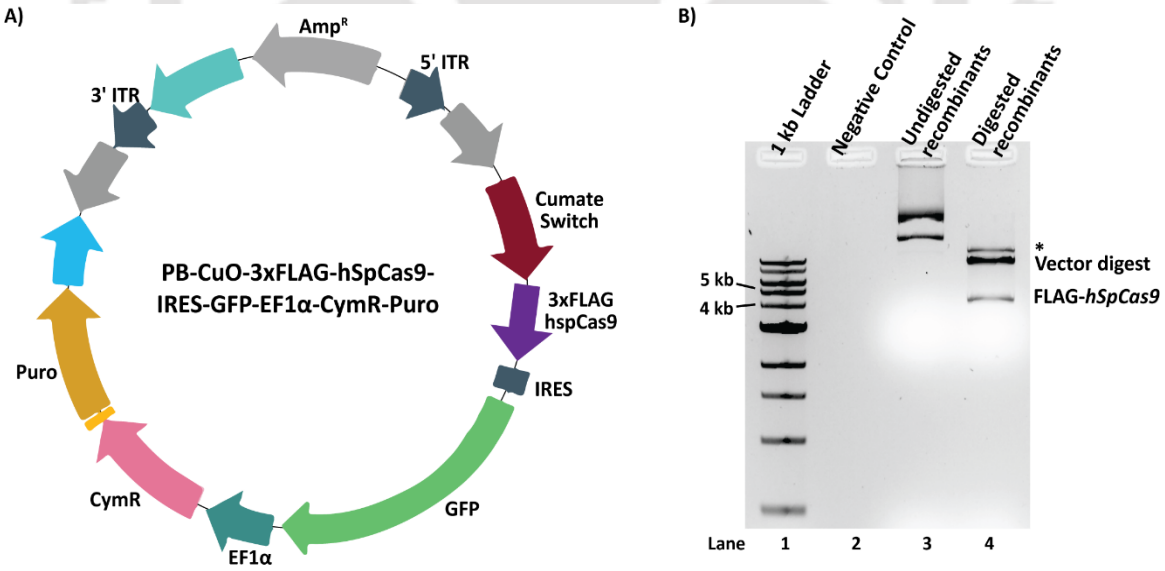


Figure 2.5. Double digestion of recombinant plasmid. A) A schematic diagram of the recombinant plasmid containing FLAG-*hSpCas9* insert. B) Agarose gel showing the double digestion of recombinant plasmid using NheI and NotI restriction enzymes. Lane 1 – 1 kb DNA Ladder, Lane 2 – non-template negative control, Lane 3 – undigested recombinant plasmid, Lane 4 – double-digested recombinant plasmid (*represents undigested or partially digested recombinant plasmid).

2.3.1.2 Generation of Stable Cells Expressing hSpCas9

To integrate the FLAG-*hSpCas9* into the host genome, the recombinant plasmid generated in the previous section (Figure 2.5A) and the PB-Transposase expression plasmid were co-transfected into the HEK-293 cells. The PB-Transposase enzyme recognizes the inverted terminal repeat (ITR) sequences in the recombinant plasmid and integrates its contents, including the FLAG-*hSpCas9* gene, GFP gene, and Puromycin resistance gene into TTAA sites in the chromosomes of HEK-293 cells.

To select the successfully transformed cells, a cell suspension was prepared with the addition of Puromycin antibiotic and allowed to grow (Figure 2.6). The survivors in this culture represent the transformants that have integrated the puromycin resistance gene, indicating the presence of the FLAG-*hSpCas9* and GFP genes.

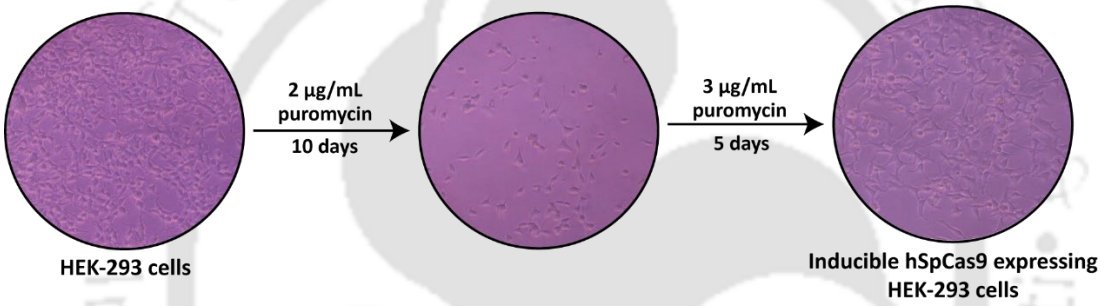


Figure 2.6. Selection of transformants. HEK-293 cells transfected with recombinant plasmid were selected with 2 µg/mL and 3 µg/mL Puromycin for 10 and 5 days, respectively.

2.3.1.3 Confirmation of hSpCas9 Expression by Microscopy and Western Blotting

The stable HEK-293 cells were analyzed to confirm the expression of the hSpCas9 protein. Transcriptional control of the *hSpCas9* and GFP genes is regulated by a Cumate switch. Strong binding of the CymR repressor to the Cumate operator site (CuO) located downstream of the CMV5 promoter, suppresses gene transcription (Figure 2.7A). The introduction of Cumate relieves this suppression, leading to protein expression. Stable HEK-293 cells were treated with or without Cumate to induce protein expression. Fluorescence imaging of cells demonstrated the expression of GFP (Figure 2.7B). Western blot analysis using an anti-FLAG antibody detected the expression of the FLAG-hSpCas9 fusion protein in Cumate-induced cells (Figure 2.7C). However, the faint band in uninduced cells indicated leaky protein expression due to a strong CMV5 promoter.

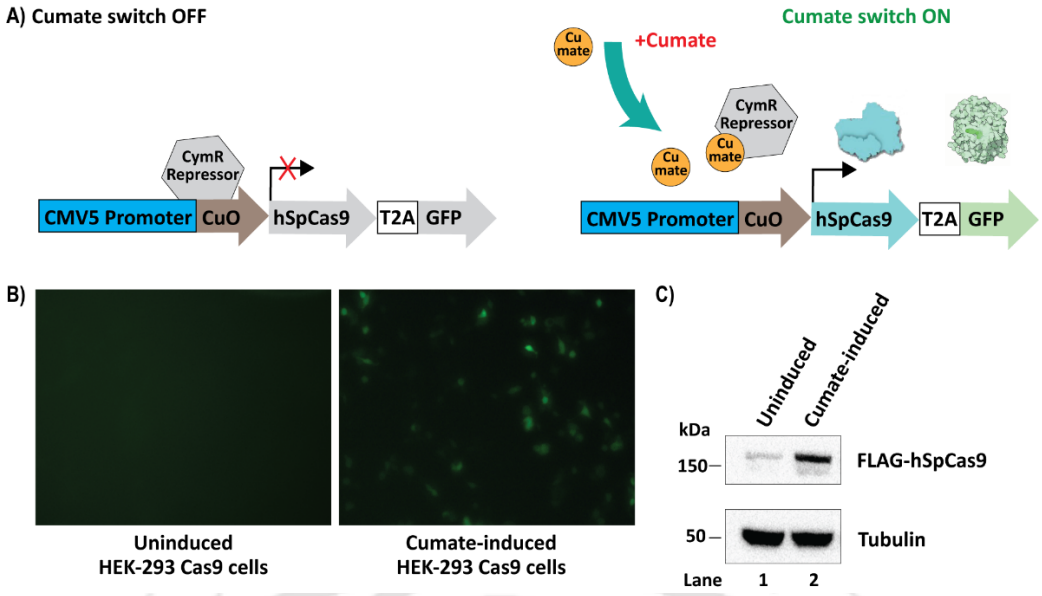


Figure 2.7. Expression of green fluorescent and Cas9 proteins upon Cumate induction. A) A schematic diagram showing the ON and OFF conditions of the Cumate switch. B) Fluorescence imaging of HEK-293 cells using a blue filter (475 nm) for GFP expression with or without Cumate induction. B) Western blot analysis of HEK-293 cells for the detection of Cas9 protein expression. Here anti-FLAG primary antibody was used for the detection of the FLAG-hSpCas9 fusion protein. Tubulin served as an internal control.

In summary, an inducible Cas9-expressing HEK-293 cell line was successfully established to enable controlled endogenous expression of the hSpCas9 protein. Subsequently, this system was employed for the purpose of knocking out the *UPF3B* gene.

2.3.2 Generation of UPF3B-KO HEK-293 Cell Line

2.3.2.1 crRNA Design

A predesigned crRNA targeting the 5' end (exon 1) of the *UPF3B* gene was obtained from IDT (https://www.idtdna.com/site/order/designtool/index/CRISPR_PREDESIGN) (Figure 2.8A). The crRNA was designed to target the 5' end of the *UPF3B* gene, making frameshift mutations likely to introduce PTC, ultimately resulting in the complete loss of the UPF3B protein. The crRNA, with the optimum length of 36 nucleotides, comprised a target-specific 20 nucleotide protospacer domain complementary to exon 1 of the *UPF3B* gene and a 16 nucleotide tracrRNA interaction domain (Figure 2.8B). A negative control crRNA, a computationally designed non-targeting 20 nucleotide protospacer sequence was also used to discard any phenotypic effects caused by crRNA transfection. Both crRNA and tracrRNA collectively formed a single guide RNA (sgRNA), with tracrRNA possessing a hairpin loop structure facilitating the binding of the hSpCas9 protein. The resulting RNP complex scans the genome for target sequences to bind and edit.

2.3.2.2 Delivery of sgRNA and Selection of Prospective KO Clones

The equimolar concentrations of negative control crRNA and UPF3B crRNA were mixed with tracrRNA to assemble their respective sgRNAs. Following the induced expression of the hSpCas9 protein, the sgRNAs were delivered to stable HEK-293 cells by lipofection. The sgRNAs, along with the endogenously expressed hSpCas9 protein, formed the RNP complex, requiring approximately 72 hours to complete genome editing. The cells were then serially diluted to get the single cells, and the resulting clones originating from these single cells were utilized to confirm the knockout.

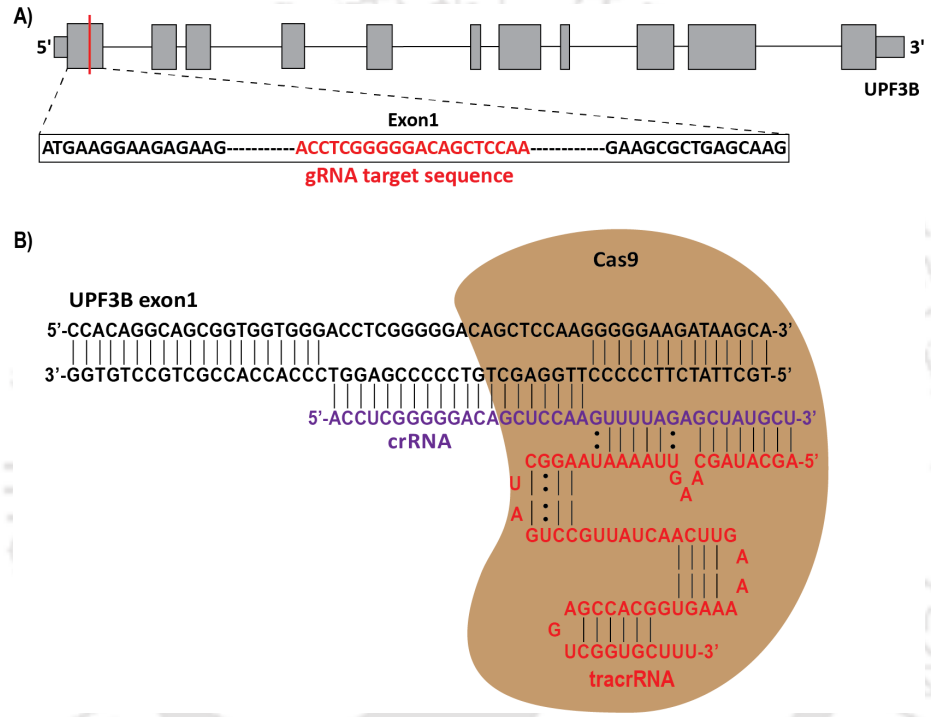


Figure 2.8. crRNA design and formation of RNP complex. A) Schematic representation of UPF3B exon map. Nucleotide sequences of exon 1 show the guide RNA target (in red). B) Aligned crRNA and tracrRNA sequences in the RNP complex. The crRNA is shown (violet) aligned with the tracrRNA (red). The target-specific protospacer domain of the crRNA is aligned with UPF3B exon 1. Cas9 interacts with the hairpin loop structure of tracrRNA.

2.3.2.3 Confirmation of UPF3B-KO Clones

To confirm UPF3B-KO, all prospective clones were examined for the expression of the UPF3B protein. Western blot analysis of five UPF3B-KO clones using an anti-UPF3B primary antibody raised against the c-terminal region (300-483 aa) of UPF3B, revealed a complete loss of UPF3B protein in two clones (Clone1 and Clone3) (Figure 2.9A). Clone2 exhibited traces of UPF3B expression, possibly attributed to a knockout in a single allele.

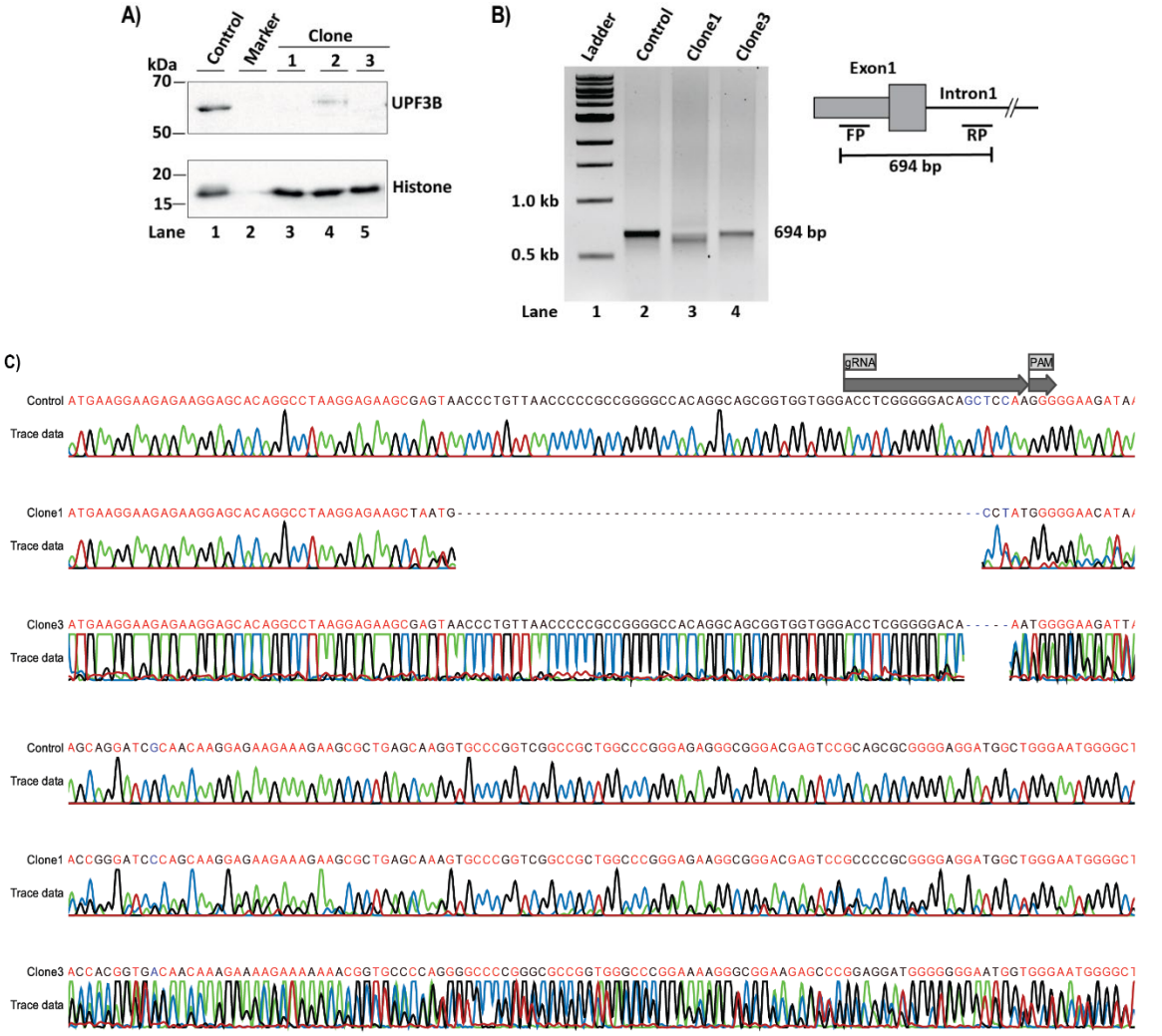


Figure 2.9. Confirmation of UPF3B-KO in HEK-293 cells. A) Western blot analysis of Control and UPF3B-KO clones Histone was used as a loading control. Lane 1 – Control (cells transfected with control sgRNA), Lane 2 – Protein size marker, Lane 3-7 – Prospective clones for UPF3B-KO analysis. B) Agarose gel showing the PCR amplification products from genomic DNA of Control, Clone1, and Clone3 cells. The genomic region amplified is shown in the left panel. Lane 1 – 1 kb DNA Ladder, Lane 2 – Control, Lane 3 – Clone1, Lane 4 – Clone3. C) Sanger sequencing of *UPF3B* exon1 sequences in Clone1 and Clone3.

Next, to identify the deletions causing frameshift mutations in the *UPF3B* gene, a gene fragment spanning the gRNA target sequence was amplified (Figure 2.9B). The primers were designed to amplify the first exonic sequences being complementary to the 5' UTR and intron 1 of the *UPF3B* gene. Sanger sequencing of amplified regions of the *UPF3B* gene showed a 55-bp deletion in Clone1 and a 5-bp deletion in Clone3.

In conclusion, two UPF3B-KO clones were successfully generated, each harboring 55- and 5-bp deletions, respectively, utilizing a single guide RNA.

2.4 Discussion

The study aimed to disrupt the *UPF3B* gene sequence using the CRIPSR/Cas9-based gene-editing tool, in contrast to previous shRNA- or siRNA-mediated knockdown approaches [5, 6], which do not completely abolish protein expression. This CRISPR/Cas9 gene editing tool requires the Cas9 protein and a target-specific guide RNA (gRNA) against UPF3B. In the initial stages, purified Cas9 protein and gRNA were assembled to form an RNP complex, which was then delivered to the cells using lipofection. Despite diligent selection and analysis of clones for UPF3B expression, the generation of UPF3B-KO clones proved elusive. The impediment in achieving gene editing efficacy was hypothesized to arise from challenges associated with Cas9 protein delivery, as microscopic examination revealed the exclusive visualization of guide RNA within the cells. This observation was attributed to the tagging of tracrRNA with a red fluorescent dye. In response to this limitation, an alternative strategy was employed wherein a plasmid construct encoding UPF3B guide RNA and Cas9 protein was transfected to the HEK-293 cells. Despite the meticulous screening of several clones for UPF3B deletion, none exhibited the desired loss of UPF3B protein.

To address the challenges associated with inefficient Cas9 protein delivery and expression from plasmid constructs, a specialized method was adopted. In this approach, an inducible Cas9-expressing HEK-293 cell line was established, where Cas9 expression is controlled by the Cumate switch. This inducible expression system enabled the controlled expression of Cas9 protein, overcoming the issues related to Cas9 delivery. Following the induced expression of Cas9 protein, UPF3B gRNA was delivered to generate knockouts. A negative control gRNA was used to account for the potential impact of gRNA transfection on observed phenotypic changes. However, it is acknowledged that the use of negative control gRNA does not completely rule out the possibility of off-target effects associated with UPF3B gRNA. Following clonal selection from single cells, two distinct clones, namely Clone1 and Clone3, were identified. Both clones demonstrated an absence of detectable UPF3B expression. Sanger sequencing of Clone1 showed a 55-bp deletion, whereas Clone3 demonstrated a 5-bp deletion in exon 1 of the *UPF3B* gene. The primary concern regarding the CRISPR/Cas9 editing method has centred on the potential for off-target effects of gRNA. The optimal strategy to address this concern involves whole genome sequencing, although this approach is more expensive. A cost-effective alternative includes the amplicon sequencing of potential off-target loci. Building upon this idea,

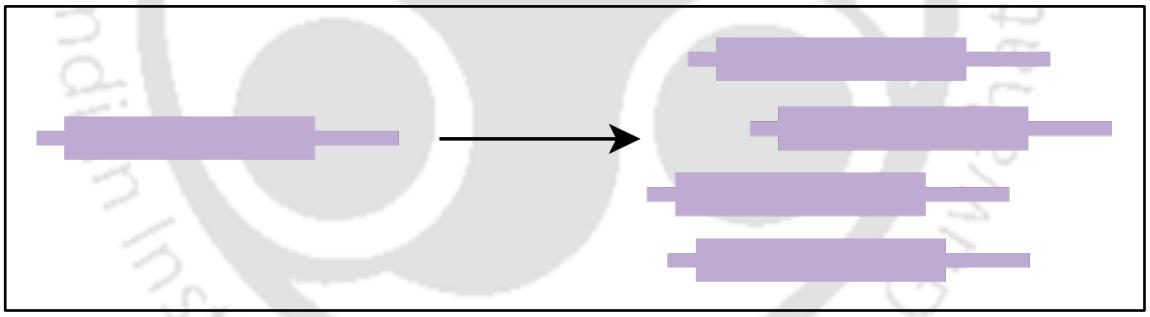
RNA-Seq data from UPF3B-KO cells were analyzed to identify any conceivable alterations in potential off-target sites (refer to APPENDIX VI), revealing an absence of discernible mutations in these loci. This approach contributes insights into the precision and specificity of the CRISPR/Cas9-mediated *UPF3B* gene editing procedure.

2.5 Conclusion

A stable HEK-293 cell line featuring inducible expression of hSpCas9 was successfully established using PiggyBac transposon-mediated genome integration. This controlled expression system served as the platform for the generation of UPF3B-KO cells facilitated by the introduction of a single guide RNA. Consequently, two distinct UPF3B-KO clones were derived, each characterized by deletions of 55-bp and 5-bp in the coding region of the *UPF3B* gene. This achievement underscores the precision and efficacy of the established inducible Cas9 expression system in orchestrating targeted genetic modifications in the context of *UPF3B* gene disruption.

Chapter 3:

Global Gene Expression Analysis of UPF3B-KO Cells



Chapter 3

Abstract

RNA sequencing has become a transformative technology in contemporary biological research, enabling the identification and exploration of defective proteins with consequential impacts on gene expression and disease pathophysiology. In previous chapter, UPF3B-deficient HEK-293 cells were systematically generated utilizing a CRISPR/Cas9-based gene knockout approach. In this chapter, RNA sequencing of UPF3B-KO cells was performed to understand the effect of UPF3B loss on global gene expression. Computational analysis of the RNA-Seq data revealed 5123 differentially expressed genes following the loss of UPF3B. Notably, among the upregulated genes with a log2 fold change greater than 1.2, *GADD45G*, *GPX1*, and *SELENOP*, were identified as potential physiological NMD targets. The observed restoration of transcript levels upon ectopic expression of UPF3B indicates that these genes may function as direct targets of UPF3B. Altogether, these findings collectively suggest a potential compromise in NMD activity subsequent to UPF3B knockout, shedding light on the intricate regulatory landscape governing mRNA decay processes.

Chapter 3

3.1 Introduction

Eukaryotic organisms exhibit intricate transcriptomic complexity, particularly at the transcript level – an intermediary step in the gene expression cascade. The complete set of transcripts within a cell, constituting both coding and non-coding RNAs and representing a particular physiological state, is defined as the transcriptome. A comprehensive comprehension of the cellular transcriptome is imperative for the functional interpretation of genomic elements. It is also essential to determine the molecular constituents of cells across various developmental stages and in pathological conditions. The initial forays into transcriptomics employed low-throughput methodologies, such as Northern blotting [159] and quantitative reverse transcription-polymerase chain reaction (qRT-PCR), which are constrained in their ability to measure individual transcripts. Subsequent advancements have seen the emergence of numerous technologies, including hybridization- and sequence-based approaches, aimed at quantifying the cellular transcriptome. The hybridization-based microarray methods offer high-throughput data acquisition at a comparatively low cost [160]. Nevertheless, these approaches require prior knowledge of the analyzed sequences, encounter challenges related to cross-hybridization in the assessment of highly similar sequences, and display limitations in accurately quantifying genes with both low and high expression levels [161, 162]. Sequence-based methods, such as Expressed Sequence Tags (EST) [163], Serial Analysis of Gene Experssion (SAGE) [164], and Cap Analysis of Gene Experssion (CAGE) [165], determine the transcript sequence to elucidate the transcriptome. A major limitation of sequence-based methods is their inability to discover novel genes. Furthermore, the labor-intensive process of cloning sequence tags, the expensive nature of Sanger sequencing, and the requirement for a substantial quantity of input RNA have restricted their widespread utilization.

Recent developments in high-throughput DNA sequencing technologies have facilitated the cost-effective analysis of genomics in living organisms [166]. These technologies, including next-generation and third-generation sequencing, have been adapted for RNA sequencing (RNA-Seq) to uncover the intricate nature of the transcriptome [167]. Progress in the RNA-Seq workflow has enabled comprehensive profiling of transcriptomes. The analysis of RNA-Seq data requires bioinformatic tools capable of addressing the limitations inherent in sequencing technologies, including the management of sequencing errors [168], considerations for read length biases [169], and the impact of RNA molecule fragmentation [170]. RNA-Seq analysis has significantly

contributed to scientific advances in areas such as the development of novel therapeutics, understanding gene regulation, identification of biomarkers, and the exploration of mutations leading to the generation of pathogenic variants [171].

In the preceding chapter, two UPF3B-KO clones were generated to study the functional consequences of UPF3B loss. This chapter comprises RNA-Seq analysis of UPF3B-KO clones, followed by differential gene expression analysis aimed to identify the broader impacts of UPF3B-KO on global gene expression.

3.2 Materials and Methods

3.2.1 RNA Isolation

Control, Clone1, and Clone3 cells were grown in 10 cm dishes in replicates supplemented with DMEM medium (as described in Section 2.2.1). After removing the growth medium, 1 mL of TRIzol reagent (Invitrogen, Cat# 15596018) was added directly to the cells to prepare a homogenate solution. 0.2 mL of chloroform (Himedia, Cat# MB109) was added to cell homogenate, and samples were centrifuged to separate into the upper aqueous and lower organic phases. The aqueous phase was collected in a new microcentrifuge tube, supplemented with an equal volume of isopropanol (Himedia, Cat# MB063), and centrifuged to obtain RNA as a pellet. The supernatant was removed, and the same volume of 75% ethanol was used to immerse the RNA pellet, followed by centrifugation. The ethanol was removed entirely from the tube, and the pellet was air-dried at RT. The pellet was resuspended in nuclease-free water.

3.2.2 Complementary DNA Library Preparation and Sequencing

The integrity of purified RNA was analyzed using an Agilent TapeStation 2200 (Agilent Technologies) with an RNA Integrity Number (RIN). Poly(A)-containing RNA species were isolated from 1 µg of purified total RNA (RNA integrity number >9.0) using oligo-d(T) attached magnetic beads (NEB, Cat# S1419S). The RNA molecules were fragmented, and the cDNA libraries were prepared with the TruSeq mRNA Sample Preparation Kit (Illumina, Cat# 15031047) according to the manufacturer's instructions. The 3'-ends of cDNA molecules were adenylated, and indexing adapters were ligated to the ds-cDNA fragments. The adapter-ligated cDNA fragments were enriched by PCR amplification and gel purified. For cluster generation, the cDNA library was loaded into a flow cell to be captured by surface-bound oligos complementary to the library adapters. Each library fragment was amplified into distinct, clonal clusters through bridge

amplification. The sequencing was performed with an Illumina NovaSeq 6000 High-Throughput Sequencing System (Illumina) at Macrogen, Seoul, South Korea. The Illumina sequencer uses integrated Real-Time Analysis (RTA) software to generate raw images. The BCL (base calls) binary was converted into FASTQ using the Illumina package bcl2fastq. Adapters were not trimmed away from the reads.

3.2.3 Pre-processing of Sequencing Data

The quality of raw files was checked using the FastQC (v0.12.1) tool (available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and the output files of FastQC were compiled using the multiQC tool [172]. The adaptor contamination was removed with Cutadapt (v3.3) [173], and the sequence quality of trimmed reads was again checked using FastQC and MultiQC. The human genome reference sequence (GRCh38.p13) and comprehensive gene annotation files were obtained from the GENCODE website (https://www.gencodegenes.org/human/). Adaptor-trimmed reads were aligned to the reference human genome (GRCh38.p13) using the RNA-seq aligner STAR (v2.7.8a) [174] in two steps. First, genome indices were prepared with *--runMode genomeGenerate* using human genome reference sequence (GRCh38.primary_assembly.genome.fa) and gene annotation (gencode.v41.primary_assembly.annotation.gtf) files. Second, the alignment was performed using the genome index with *--runMode alignReads* to generate BAM files. The alignment (.bam) files were used to count the reads for each gene with HTSeq-count (v0.13.5) [175].

3.2.4 Differential Gene Expression Analysis

The RStudio (v6.1.0) environment (https://posit.co/download/rstudio-desktop/) with the integration of R-base (version 4.2.1) (cran.r-project.org/bin/windows/base/) was used to analyze the differential gene expression. The DESeq2 package was loaded into the R environment, which stored the read quantification files imported by the function *DESeqDataFromHTSeqCount* [176]. Rows with a zero-count sum were filtered out. Counts were normalized by a regularised log (rlog) transformation and a variance stabilizing transformation. The rlog transformed count data was used to generate sample cluster plots and principle component analysis plots using the ggplot2 package. The variance-transformed count data was used to create boxplots and scatter plots with correlation using the bigPint package [177]. DESeq2 uses negative binomial generalized linear models and shrinkage estimation for dispersions and fold changes to improve the stability and

interpretability of the estimates. It reports a p-value and an adjusted p-value using the Benjamini-Hochberg procedure. Genes with an adjusted p-value of 0.001 were considered differentially expressed. The significant genes were annotated with the GENCODE gtf file. The annotated result data was used to generate the volcano plots using ggplot2.

3.2.5 Rescue Experiment

1×10^6 cells were seeded in 35 mm dishes for UPF3B-KO clones in triplicate. 1 µg of pCI-FLAG-UPF3B plasmid was transfected into each dish using PEI reagent. 24-hour post-transfection cells were harvested using TRIzol reagent.

3.2.6 Total RNA Isolation

1×10^6 cells were seeded in 35 mm dishes for Control, UPF3B-KO, and UPF3B-rescued cells in three biological replicates. Total RNA was isolated from all the samples using the TRIzol reagent (as described in Section 3.2.1).

3.2.7 Removal of Genomic DNA

The contamination of gDNA in RNA samples was removed by DNase treatment. 2 µg of total RNA was incubated with 2 units of DNase I (Promega, Cat# M6101) at 37 °C for 30 min. The activity of DNase I was stopped by adding 1 µL of DNase stop solution and incubating at 65 °C for 10 min.

3.2.8 cDNA Synthesis

The DNA-free RNA samples were used for cDNA synthesis using a Super Reverse Transcriptase (MuLV) kit (Biobharati, Cat# BB-E0043). 2 µg of purified RNA was mixed with 1 µL of each dNTP mix (10 µM) and oligo-d(T) (10 µM) and incubated at 65 °C for 5 min. Next, 4 µL of 5x First strand buffer, 2 µL of 0.1M DTT, and 1 µL of RNase inhibitor were added to the sample. The sample was incubated at RT for 5 min. Finally, 1 µL of reverse transcriptase enzyme was added to the reaction mixture and incubated at 42 °C for 50 min. The reaction was stopped by heating at 70 °C for 15 min.

3.2.9 Quantitative Real-Time PCR

For qRT-PCR, the reaction mixture contained 5 µL of PowerUp SYBR Green Master Mix (Applied Biosystems, Cat# A25742); 1 µL of cDNA template (1:1 or 1:3 or 1:5 dilutions); 0.5 µL of gene-specific primers (0.25 µL forward primer and 0.25 µL reverse primer), and 3.5 µL of nuclease-free water. The relative abundance of genes was quantified using the comparative Ct method (ΔΔCt, where ΔCt = [Ct of target gene] - [Ct of

endogenous control gene]). The actin gene was used as an endogenous control. All the primer sets are listed in APPENDIX IV.

3.2.10 mRNA Stability Assay

1×10^6 Control and UPF3B-KO clones were seeded in 35 mm dishes in replicates. After 24 hours of seeding, cells were treated with 10 µM Actinomycin D. Samples were collected at 0 hr, 3 hr, 6 hr, and 9 hr. Total RNA was isolated from each sample using the TRIzol method, and cDNA was synthesized, followed by qRT-PCR analysis for mRNA quantification. The average of Ct values at each time point was normalized to the average Ct value at t = 0, i.e., ΔCt = average of Ct value at a particular time – average Ct value at 0 hr. The relative abundance of each time point was calculated as mRNA abundance = 2^(-ΔCt). The decay constant (K_{decay}) was derived from the non-linear regression equation between the natural logarithm and time points as the slope of the equation. The mRNA half-lives were calculated using the formula $t_{1/2} = \ln 2/K$.

3.2.11 Statistical Analysis

The data analysis was performed in Microsoft Excel. The tests on the qPCR data were performed in GraphPad Prism 6. The bar plots for qRT-PCR data and non-linear curves for stability assays were generated using GraphPad Prism 6. The Venn diagram was prepared using a web tool developed by Bioinformatics & Evolutionary Genomics (https://bioinformatics.psb.ugent.be/webtools/Venn/).

3.2.12 Data Availability

The RNA-Seq dataset is available at https://iitgoffice-my.sharepoint.com/:f:/g/personal/singhpc90_iitg_ac_in/Eugd7gem4KZFmaxJ7JL45GgBDxcu0vArl9JKWDXhu_2XUw?e=f8xLBr

3.3 Results

3.3.1 RNA Sequencing

RNA-Seq was conducted to elucidate the perturbations in gene expression resulting from the knockout of UPF3B. High-quality RNA samples, characterized by RNA integrity number (RIN) exceeding 9.0, were utilized for cDNA library preparation and subsequent sequencing. A total of 12 raw data files were generated, encompassing Control and UPF3B-KO cells, with four biological replicates from each cell type (Control, Clone1, and Clone3). The salient features of the raw files are listed in Table 3.1. Each raw file contained the read

sequences with an average length of 101 bp, a read depth of 21-24 million, and GC content ranging between 51 to 53%.

Table 3.1. General statistics of RNA-seq output raw files.

S. No.	Sample Name	File Size (GB)	Read Length (bp)	Read Depth (Millions)	GC Content (%)
1.	Control_1.fastq	1.1	101	22.7	51
2.	Control_2.fastq	1.2	101	22.7	52
3.	Control_3.fastq	1.2	101	24.0	51
4.	Control_4.fastq	1.3	101	24.0	52
5.	Clone1_1.fastq	1.1	101	22.6	53
6.	Clone1_2.fastq	1.2	101	22.6	53
7.	Clone1_3.fastq	1.1	101	22.2	53
8.	Clone1_4.fastq	1.2	101	22.2	53
9.	Clone3_1.fastq	1.2	101	23.5	52
10.	Clone3_2.fastq	1.2	101	23.5	52
11.	Clone3_3.fastq	1.0	101	21.0	52
12.	Clone3_4.fastq	1.1	101	21.0	53

3.3.2 Quality Assessment of Raw Files

The quality assessment of the raw files involved a meticulous examination of the read sequences using the Phred quality score (Q). The Phred score, directly proportional to the negative logarithm of the error probability (E), was computed using the formula: $Q = -10\log(E)$. Table 3.2 provides the Phred scores along with the corresponding error probabilities and base call accuracy. Figure 3.1A illustrates that the majority of reads in the RNA-Seq data files demonstrated a Phred score surpassing 30. This signifies that the reads possessed an error probability of less than 0.1% and a base call accuracy exceeding 99.9%.

Table 3.2 The Phred quality scores.

Phred quality score	Probability of incorrect base call	Base call accuracy
10	1 in 10 = 10%	90%
20	1 in 100 = 1%	99%
30	1 in 1000 = 0.1%	99.9%

40	1 in 10000 = 0.01%	99.99%
50	1 in 100000 = 0.001%	99.999%
60	1 in 1000000 = 0.0001%	99.9999%

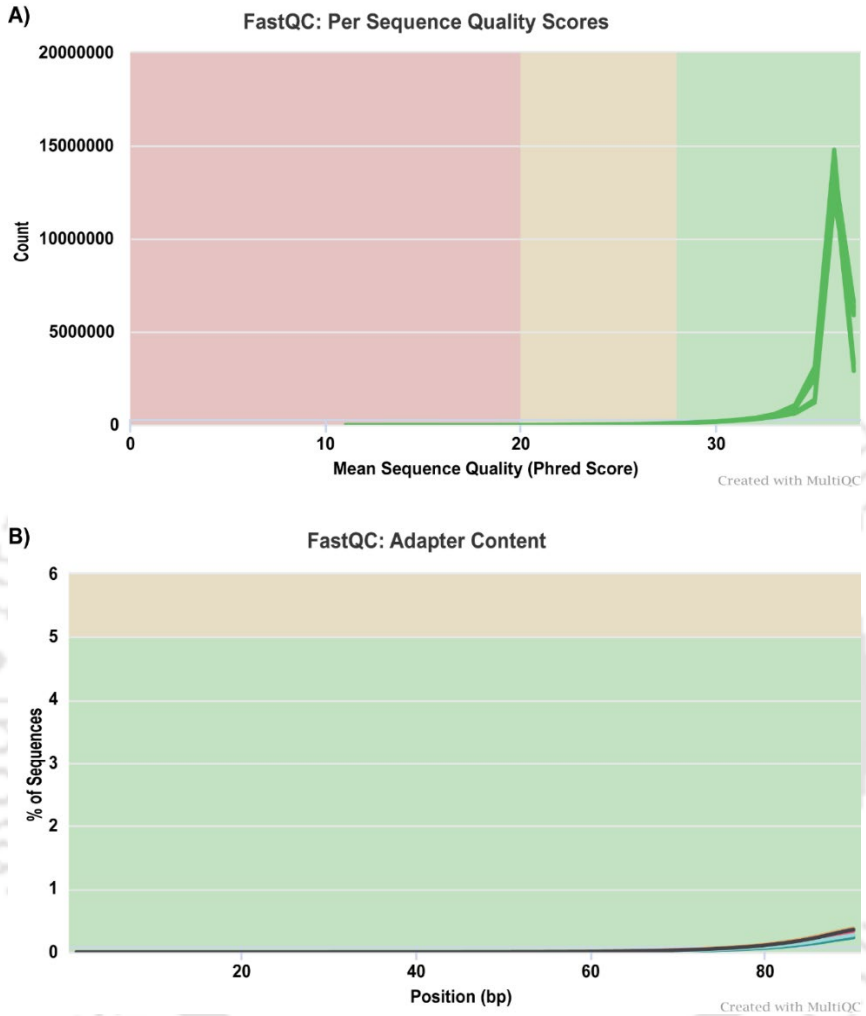


Figure 3.1. Quality assessment of sequences in raw files. A) The Phred quality score per sequence is shown for all 12 raw files. The x-axis shows the mean sequence quality score in terms of the Phred score and the y-axis shows the count of read sequences per sample (raw files). B) The percentage of adapter content is shown for all 12 raw files. The x-axis shows the position of nucleotide bases in read sequences and the y-axis shows the percentage of sequences per sample (raw files).

Within the RNA-Seq workflow, cDNA fragments undergo ligation to an Illumina universal adaptor. Post-sequencing, the analysis of read sequences sought traces of adaptor sequences, as these were not trimmed from the reads. Across all raw files, less than 1% of sequences displayed adaptor content (Figure 3.1B). Subsequently, the adaptor sequences were trimmed from the reads, and ensuing quality checks revealed the absence of samples with adaptor contamination.

3.3.3 Alignment of Processed Reads

To map the read sequences to the genome elements, Spliced Transcripts Alignment to a Reference (STAR) was employed to align the reads to the human reference genome. On average, 92.57% of reads were uniquely aligned to the reference genome, 6.22% were mapped to multiple features or genes, and 1.2% remained unmapped (Figure 3.2). The notably high percentage of reads (>90%) uniquely aligned to the reference genome affirmed the excellent quality of the cDNA library. Specific percentages of aligned reads from the individual sample files are listed in Table 3.3.

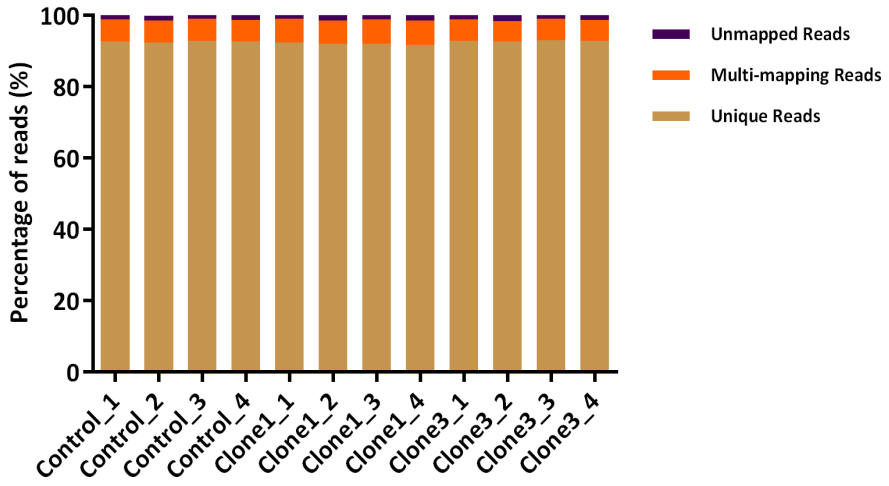


Figure 3.2. Alignment of reads. Pre-processed reads from all 12 raw files (sample replicates) to the reference genome are shown as bar plots. The percentage of unique, multi-mapped, and unmapped reads are shown in brown, orange, and dark violet colors, respectively.

Table 3.3. The percentage of reads aligned to the reference genome.

Sample Name	Aligned Reads (in %)		
	Unique reads	Multi-mapping reads	Unmapped reads
Control_1	92.73	6.17	1.10
Control_2	92.42	6.09	1.49
Control_3	92.93	6.17	0.90
Control_4	92.71	6.00	1.29
Clone1_1	92.47	6.61	0.92
Clone1_2	92.16	6.43	1.41
Clone1_3	92.08	6.90	1.02
Clone1_4	91.80	6.72	1.48
Clone3_1	92.97	5.94	1.09
Clone3_2	92.68	5.78	1.54

Clone3_3	93.08	5.99	0.93
Clone3_4	92.86	5.85	1.29

3.3.4 Counting the Aligned Reads and Normalization of Counts

Quantification of aligned reads was performed by determining the number of reads per feature or gene, resulting in the creation of a count matrix for subsequent data analysis. To assess the distribution of read counts, side-by-side boxplots were generated, showing the central tendency and variability of the dataset. Initially, the read counts from each replicate displayed a skewed distribution across both the Control and knockout samples (Figure 3.3A-B). To mitigate the experiment-wide dependence of variance on the mean, two data transformation approaches were applied: variance stabilizing transformation (vst) and regularized logarithm (rlog). Both approaches generated transformed data on a log2 scale. The application of variance-stabilized read counts resulted in consistent medians with less skewed distributions observed within the sample replicates and across the sample conditions (Figure 3.3C-D).

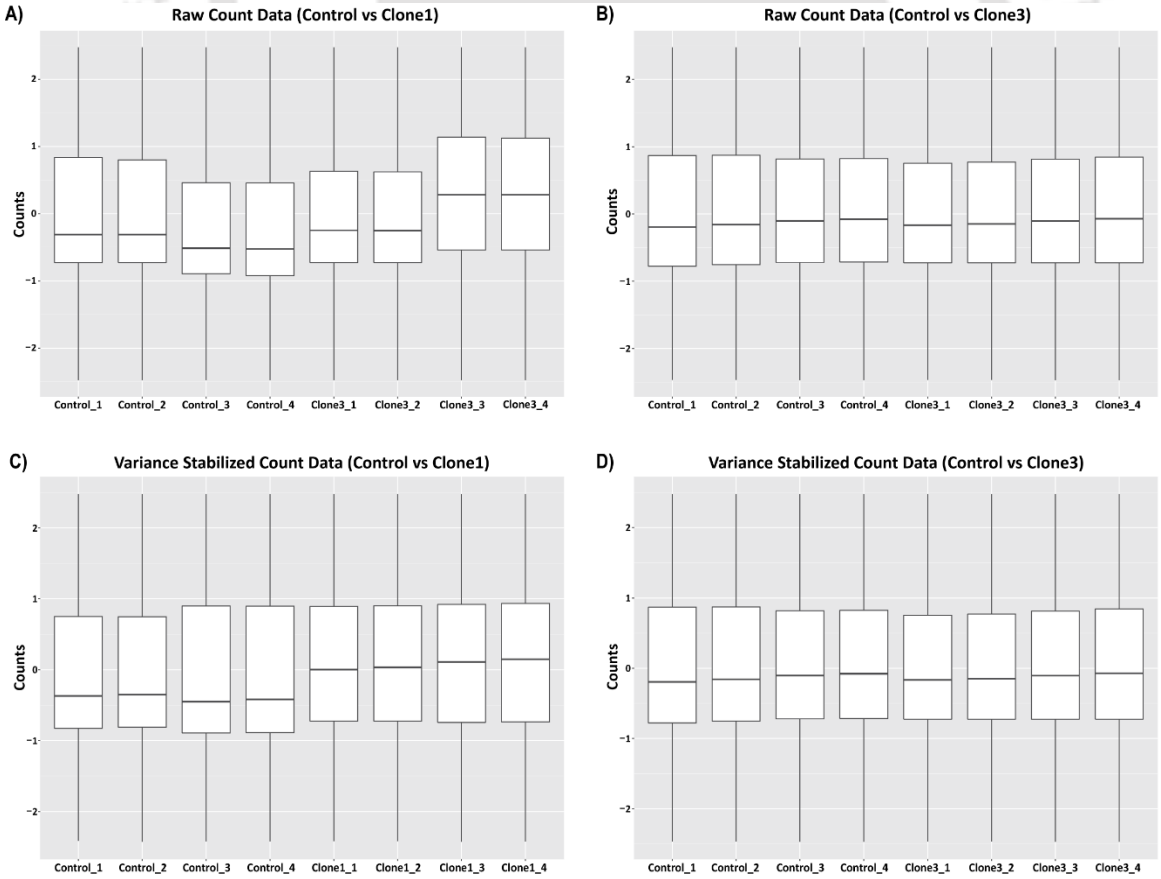


Figure 3.3. Distribution of read counts. The Side-by-side boxplots show the raw count data (A-B) and variance-stabilized count data (C-D) from Control and UPF3B-KO clones. The lower and upper ends of the boxes represent the first quartile (Q1) and third quartile (Q3). The middle line in the box represents the median of the dataset.

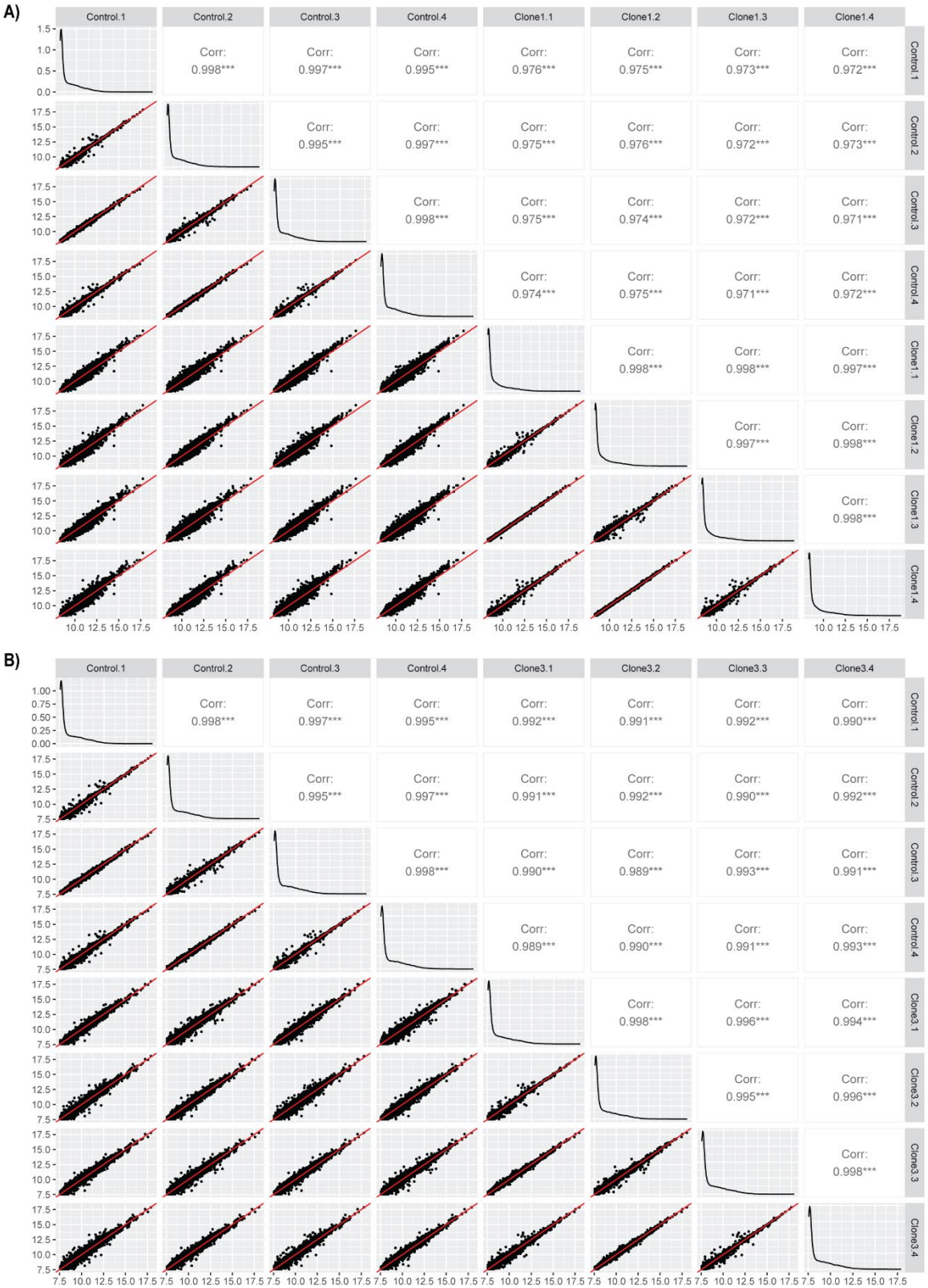


Figure 3.4. Sample correlation scatter plots. The scatterplot matrices show the transcriptome structure of RNA sequencing data. The scatterplot matrices show the correlation between A) Control and Clone1 sample replicates, and B) Control and Clone3 sample replicates. Most genes (dots) in scatterplots are aligned along the x-y line. Genes deviating more from the x-y line can be observed in UPF3B-KO clones, i.e. Clone1 and Clone3.

Scatterplot matrices, illustrating the distribution of variance-stabilized read counts, revealed increased variability across the experimental conditions (Control and UPF3B-KO) in contrast to the variation observed within the sample replicates (Figure 3.4A-B). Similarly, the principal component analysis demonstrated that the transcriptome signatures of Clone1 and Clone3 formed distinct clusters from the Control, while the sample replicates clustered together (Figure 3.5). This indicates that the differences across the experimental conditions are more prominent than the variations among the sample replicates. Moreover, the transcriptome signature of two UPF3B-KO clones were found to be clustered separately, indicating that these clones have different levels of gene expression.

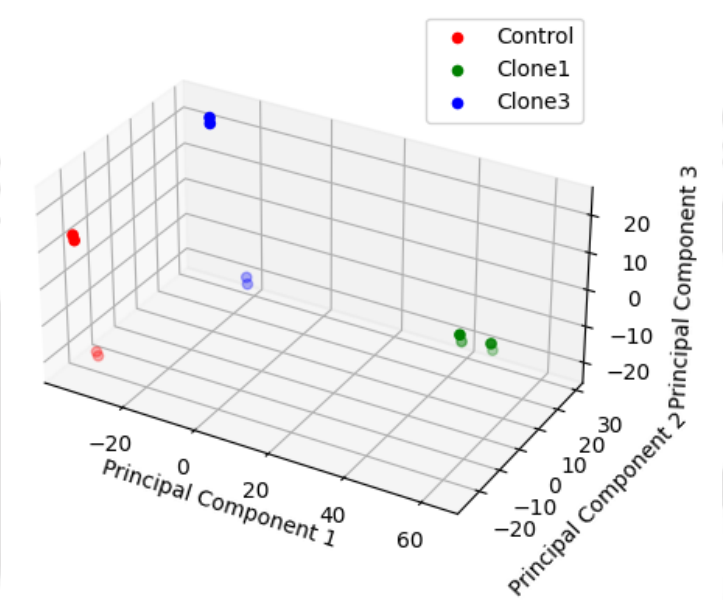


Figure 3.5. Principal component analysis. 3D PCA plot of RNA-Seq dataset for Clone1 and Clone3 compared to the Control cells, with each point representing the biological replicates.

3.3.5 Differential Gene Expression Analysis

In order to quantify gene expression, differential gene expression analysis was performed using the DESeq2 tool. The differentially expressed genes (DEGs) in UPF3B-KO clones (Adj. p-value < 0.05) were compared with the recent study by Wallmeroth *et al.* to assess whether the deregulated genes correlate with recent studies [149, 150]. Interestingly, 797 DEGs aligned with the RNA-Seq analysis of UPF3B-KO HEK-293 cells in the aforementioned study (Figure 3.6), suggesting that UPF3B regulates a set of physiological transcripts in UPF3B-KO cells.

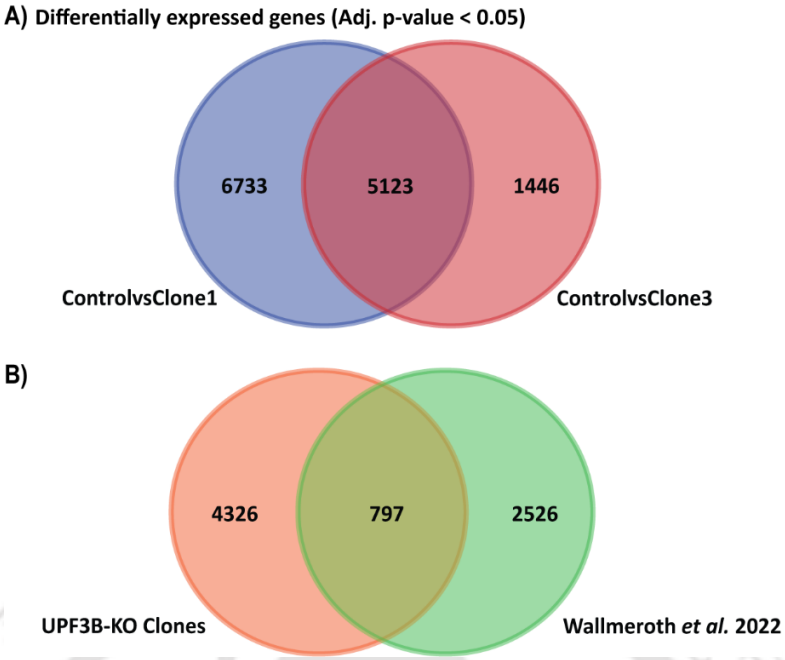


Figure 3.6. Common DEGs in UPF3B-KO cells. A) Common DEGs in UPF3B-KO clones, i.e. Clone1 and Clone3. B) The Venn diagram shows overlapping DEGs from UPF3B-KO clones compared to UPF3B-KO HEK-293 cells generated by Wallmeroth *et al*. 2022.

Furthermore, to consider statistically more significant and biologically relavent genes, DEGs were filtered with adjusted p-value < 0.001 and log2FC > |1.2|. The volcano plots displayed the filtered DEGs using aforesaid parameters (Red dots, Figure 3.7A-B). To analyze the impact of UPF3B-deficiency on biological processes, Gene Ontology (GO) classification of DEGs (red dots in volcano plots) was performed using DAVID online tool. The largest set of genes are associated with regulation of transcription in both the knockout clones (Figure 3.8A-B). Additionally, DEGs are associated with cell division and cell cycle in Clone1, whereas cell differentiation and cell proliferation in Clone3. This suggest that UPF3B might have significant role in transcriptional regulation and cell growth. Apart from these biological processes, UPF3B is an integral component of several branches of NMD pathway and there could be stabilization of NMD genes in knockout cells. Therefore, the upregulated genes (280 in Clone1 and 125 in Clone3) among DEGs were further analysed.

3.3.6 Upregulation of Physiological NMD Substrates

Further investigation of differential gene expression data revealed that the mRNA levels of *GADD45G* and *GPX1* were enriched in UPF3B-KO cells (Figure 3.9A-B). *GADD45G* transcripts are routinely degraded via NMD to sustain cell viability [178]. *GADD45G* mRNA level was upregulated by ~1.35 (p-value 0.016) and ~5 (p-value 0.0001) folds in Clone1 and Clone3, respectively. *GPX1* mRNA contains UGA codon 105 nts

upstream of the sole exon-exon junction, which incorporates selenocysteine into the polypeptide chain and is otherwise degraded by NMD if UGA is recognized as nonsense [179, 180]. *GPX1* mRNA level was elevated by ~1.3 fold (p-values 0.0002 and 0.002 in Clone1 and Clone3, respectively) in UPF3B-KO clones. Additionally, the mRNA half-life assay using Actinomycin-D showed the stabilization of *GPX1* mRNA with an increase in half-lives by 0.55 hours and 0.68 hours in Clone1 and Clone3, respectively (Figure 3.9C).

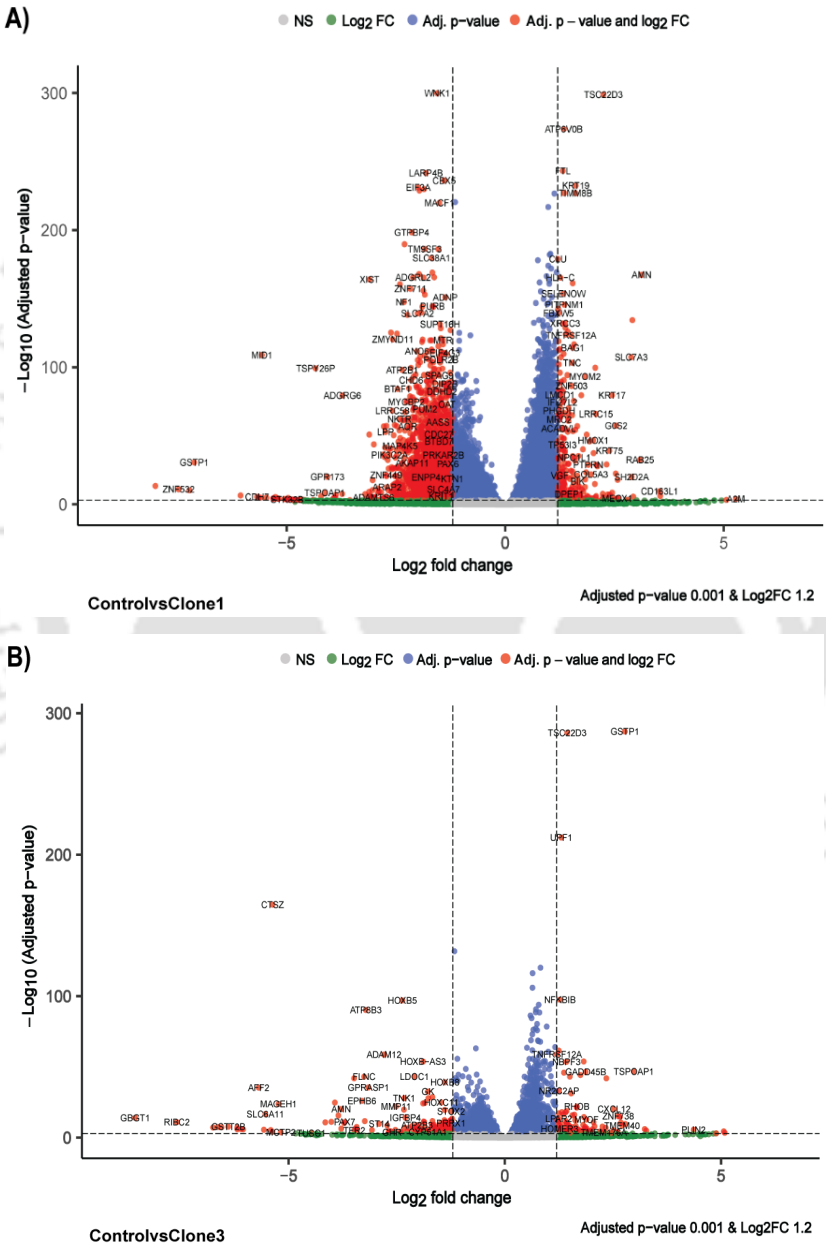


Figure 3.7. Differential gene expression analysis. The volcano plots show the differential gene expression of RNA-Seq data of A) Clone1 vs Control, and B) Clone3 vs Control. The log2FoldChange is plotted against the –log10 adjusted p-value.

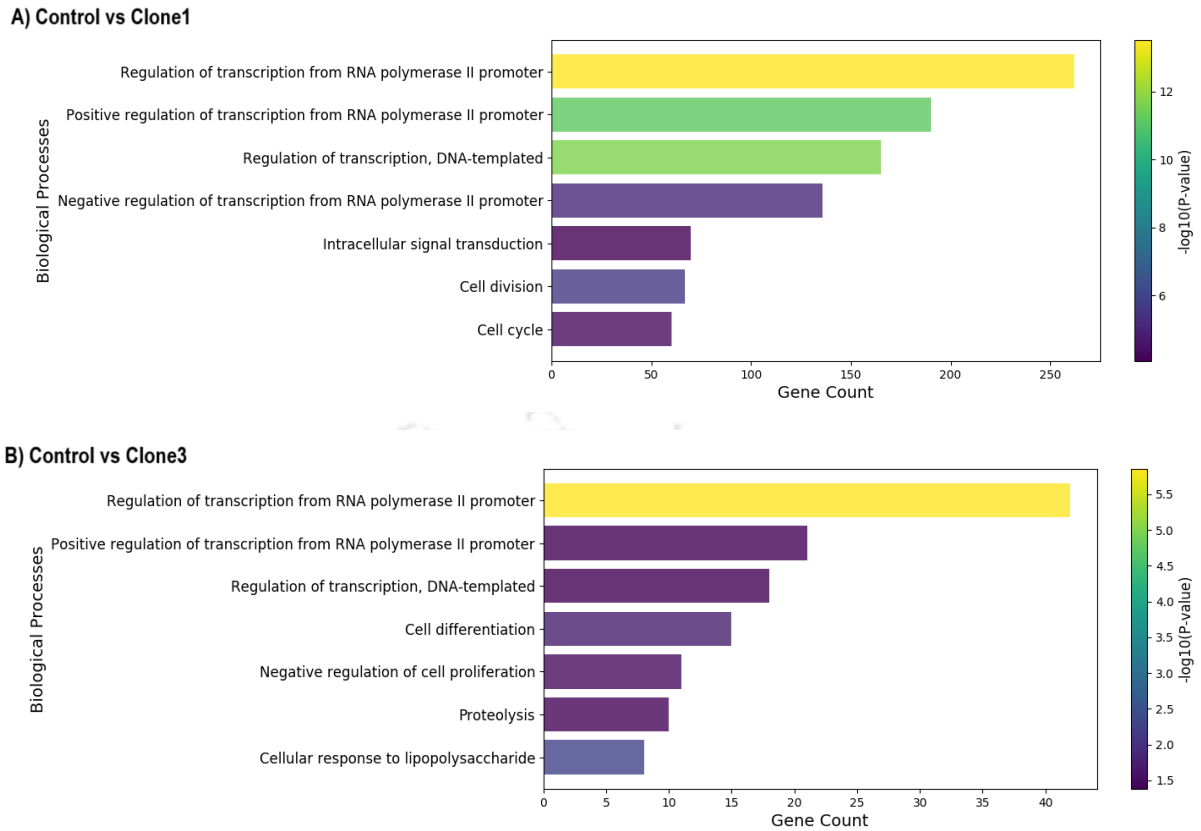


Figure 3.8. Gene Ontology classification. GO analysis of DEGs obtained from A) Control vs Clone1, and B) Control vs Clone3 analyses. The horizontal bar plots are arranged according to decreased order of -log10(p-value). The length of bars indicates the number of genes associated with biological processes.

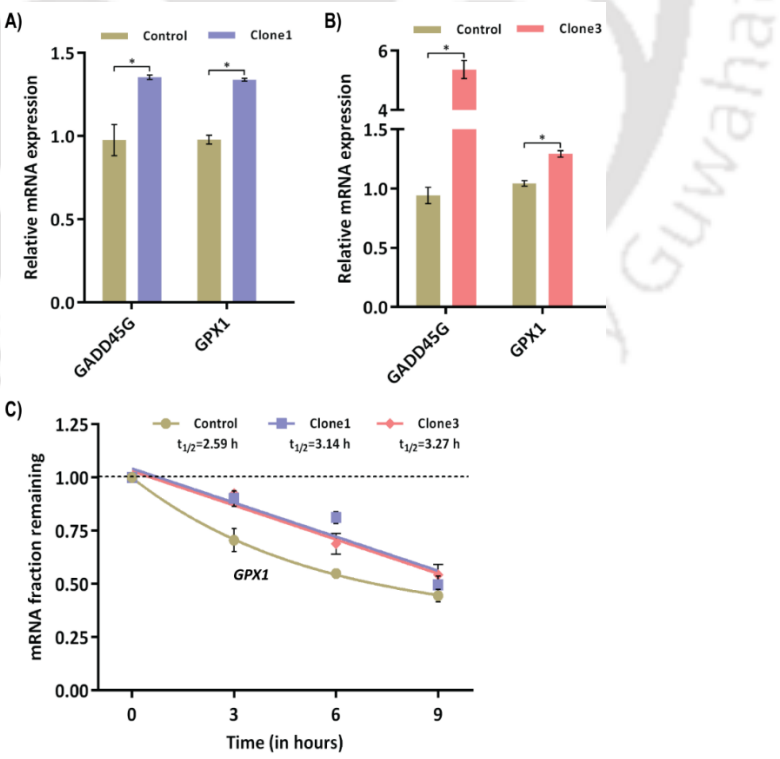


Figure 3.9. Transcript expression and stability of physiological NMD substrates. The qPCR data shows the mRNA levels of *GADD45G* and *GPX1* in A) Clone1 and B) Clone3 compared to Control. C) Non-linear curve showing the decay of *GPX1* mRNA in *UPF3B*-KO clones.

To investigate whether the upregulated substrates are directly under the regulatory influence of UPF3B, UPF3B was ectopically expressed in KO cells (Figure 3.10A). The quantification of transcript levels revealed a complete rescue of *GADD45G* and *GPX1* upon ectopic expression of UPF3B in Clone1 (Figure 3.10B-C). Another interesting candidate, *SELENOP* mRNA, encodes a selenoprotein and contains multiple UGA codons, rendering it susceptible to the NMD pathway. The qRT-PCR data demonstrated a remarkable upregulation of ~9-fold (p-value 0.0001) in *SELENOP* mRNA, specifically in Clone1 (Figure 3.10D). In *SELENOP* transcripts, the first selenocysteine codon is located upstream of introns and has the potential to be identified as a PTC by the NMD machinery [151]. Notably, to ascertain whether *SELENOP* is a direct target of UPF3B, transcript abundance was determined upon UPF3B rescue. Partial rescue of *SELENOP* transcripts was observed in Clone1, suggesting the involvement of NMD along with other pathways in regulating *SELENOP*.

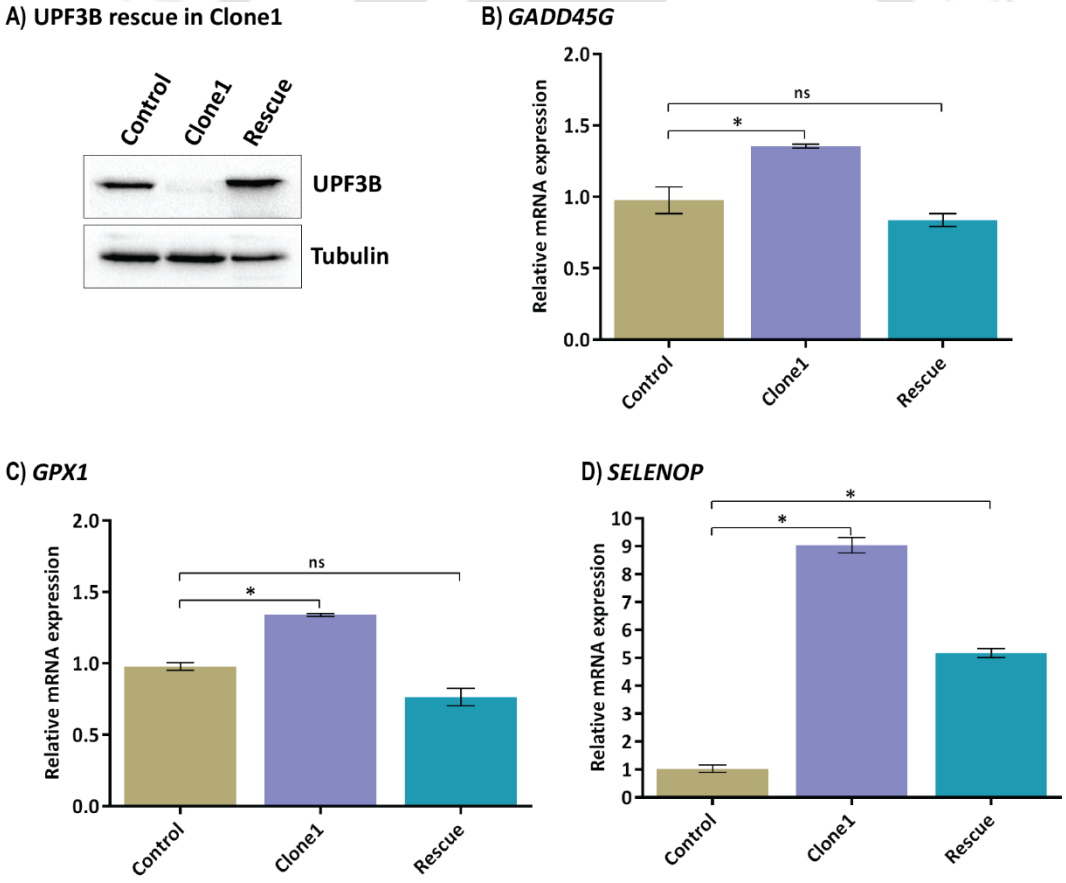


Figure 3.10. Rescue of physiological NMD substrates. A) Western blot analysis of UPF3B protein in Control, Clone1 and rescued cells (ectopic expression of UPF3B in Clone1). Tubulin was used as the internal control. The qPCR data shows the mRNA levels of B) *GADD45G*, C) *GPX1,* and D) *SELENOP* in Control, Clone1, and UPF3B-rescued Clone1 cells. (Unpaired Student's t-test, n=3). Error bars indicate the standard error of the mean. ns – not significant, *p-value < 0.05.

In conclusion, the knockout of UPF3B results in the upregulation of physiological transcripts like *GADD45G*, *GPX1*, and *SELENOP*, which are typically targeted by NMD. These pioneer findings underscore the intricate nature of UPF3B in regulation of physiological NMD substrates.

3.4 Discussion

RNA sequencing and subsequent analysis of sequenced data for differential gene expression identified several genes deregulated in UPF3B-KO cells compared to the Control cells. The difference in gene expression might stem from the varying deletions in the UPF3B gene sequence, which could introduce genetic variability in the two knockout clones. The changes in gene structure at the micro level might have affected the epigenetic landscape of the cells, leading to differential gene expression patterns, which is beyond the scope of this thesis to investigate. Since the two UPF3B-KO clones, Clone1 and Clone3, are devoid of UPF3B, the common shared genes between both the clones were analyzed to avoid the ambiguity in expression profile. A total of 5123 genes were differentially expressed in both KO clones. A recent study by Wallmeroth *et al.* performed a differential gene expression analysis of UPF3B-KO HEK-293 cells, and identified 3323 differentially expressed genes [150]. Remarkably, 797 genes overlapped with the aforementioned study, indicating the regulatory influence of UPF3B on a subset of genes with physiological importance. After filtering the genes with log2FoldChange > 1.0, Wallmeroth *et al.* reported 74 genes upregulated in UPF3B-KO cells. In contrast, the current study identified 280 and 125 upregulated genes in Clone1 and Clone3, respectively, with log2FoldChanges > 1.2. This difference in gene expression profiles might stem from different deletions in the *UPF3B* gene, exerting varying effects on the cellular transcriptome.

Importantly, the loss of UPF3B is functionally associated with neurodevelopmental disorders (Reviewed in [156]). RNA-Seq analysis of the KO clones revealed the upregulation of *GADD45G* and *SELENOP*, both of which are known NMD targets and play crucial roles in neuronal function. In Xenopus, Gadd45g expression in primary neuron precursors induces cell cycle exit and promotes terminal differentiation, emphasizing its role in neuronal development [190]. SELENOP is critical for neuronal survival and is linked to neurological dysfunction [191, 192]. The upregulation of these transcripts underscores the importance of UPF3B in regulating neural-significant transcripts.

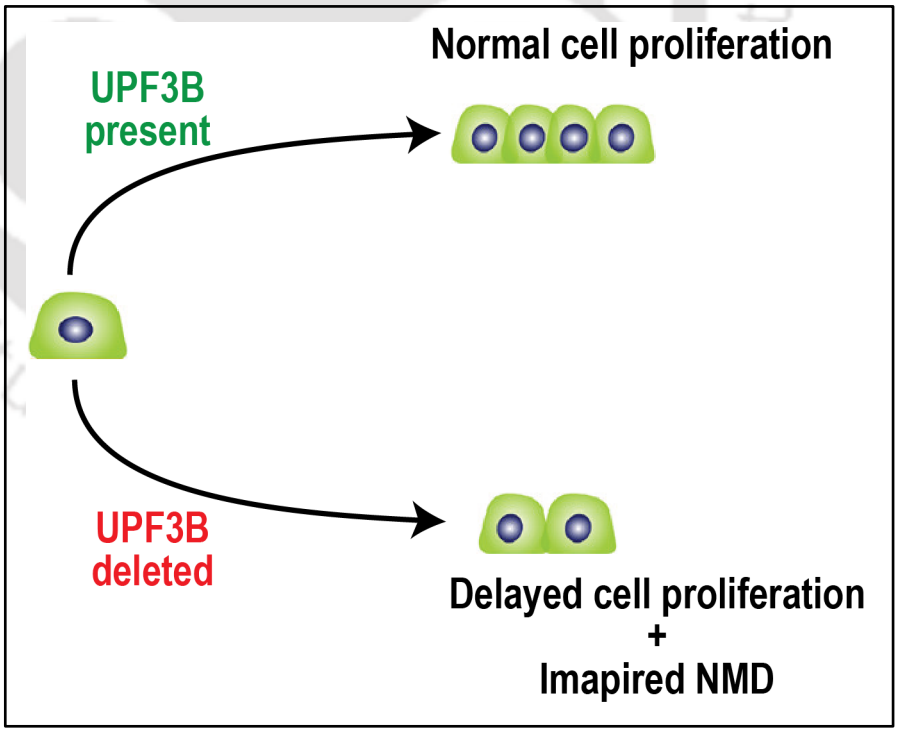
Altogether, the increased expression of physiological NMD substrates indicates plausible impairment in NMD activity upon UPF3B-KO. Additionally, it is worth to note that a greater number of genes were found to be downregulated in UPF3B-KO clones compared to upregulated genes, prompting the prospect of investigating the role of UPF3B in gene downregulation as a potential avenue for future research.

3.5 Conclusion

The analysis of differential gene expression in UPF3B-KO clones identified 5123 genes, with 797 genes coinciding with a separate study involved in UPF3B-KO. Notably, three noteworthy NMD-associated genes, *GADD45G*, *GPX1*, and *SELENOP*, were found to be upregulated in KO clones. It was demonstrated for the first time that these physiological genes are regulated by UPF3B-mediated NMD as the ectopic expression of UPF3B restored the mRNA levels of these genes in KO cells. The elevation of these NMD-associated physiological genes suggests a potential impairment of the NMD pathway in the absence of UPF3B.

Chapter 4:

Impact of UPF3B-KO on NMD Impairment and Cell Proliferation



Chapter 4

Abstract

UPF3B, a constituent of the Nonsense-Mediated Decay (NMD) complex, plays a pivotal role in recruiting other NMD factors onto spliced mRNA across various alternative NMD pathways. Despite conflicting evidence regarding the necessity of UPF3B for NMD activity, an investigation into the influence of UPF3B on the NMD pathway was conducted through the analysis of UPF3B-knockout (KO) cells. This scrutiny involved assessing the expression levels of canonical NMD substrates and an NMD reporter transcript. The discernible elevation in the expression of these transcripts upon UPF3B-KO indicates an impaired NMD activity. Consequently, this alteration in NMD activity triggers the autoregulatory feedback mechanism of NMD factors. However, the concomitant upregulation of NMD factors fails to compensate for the loss of NMD activity in the absence of UPF3B. Additionally, the compromised NMD pathway is implicated in an extension of cell's doubling time and an impact on cell cycle progression, potentially attributable to reduced expression of cell cycle genes. In summary, the deletion of UPF3B leads in a consequential loss of NMD activity, subsequently contributing to a slowdown in cell growth and proliferation.

Chapter 4

4.1 Introduction

NMD represents a complex branched pathway, orchestrated by the dynamic combinatorial recruitment of various proteins, including UPF, SMG, EJC, and additional accessory proteins, culminating in the formation of the NMD complex [75, 83-90]. Notably, among the components of the NMD complex, UPF3B consistently associates with the majority of alternative NMD pathways (Chapter1, Figure 1.4). Despite being an essential component of NMD, the functional significance of UPF3B in NMD activity remains a subject of contention. The depletion of UPF3B in HeLa cells using short hairpin RNA (shRNA) interference resulted in impaired NMD activity, indicating that a sufficiently low level of UPF3B is required for NMD inhibition [84]. It is important to note, however, that shRNA-mediated knockdown does not achieve complete removal of UPF3B protein expression. The advent of CRISPR/Cas genome editing technology, has enabled the eradication of protein expression by gene knockout. Two research groups, utilizing CRISPR/Cas9-mediated gene editing, independently generated UPF3B-KO cell lines to elucidate the effect of UPF3B on NMD activity [149, 150]. Intriguingly, they demonstrated that the simultaneous KO of UPF3B and UPF3A results in potent NMD inhibition, whereas the sole KO of UPF3B yields either functional NMD or mild deficiency in NMD. Together, these data indicate that UPF3B may be dispensable for overall NMD activity. However, further investigation is warranted to meticulously delineate the exact role of UPF3B in the intricate processes underlying NMD.

The intricate regulatory role of NMD in cellular processes is underscored by its influence on approximately 10% of normal transcripts. A disruption in NMD activity is poised to reverberate throughout the cellular transcriptome, thereby affecting essential biological functions [146, 193-195]. The robustness of the NMD pathway is sustained by a buffering mechanism intrinsic to the NMD process. Consequently, repression of NMD prompts the upregulation of NMD factors, a phenomenon orchestrated by a feedback regulatory loop [14, 84, 196]. Transcripts augmented during this feedback mechanism directly represent NMD targets, harbouring features that induce NMD [197]. The escalation in the expression of NMD factors serves as a reliable indicator of perturbations within the NMD pathway. Beyond its autoregulatory feedback loop, NMD plays a significant role in regulating the cell cycle progression. Compromised NMD pathway integrity, evidenced by the loss of UPF1 and UPF3B, results in the accumulation of cells, including human embryonic stem cells (hESCs), in the G1 phase of the cell cycle [22, 118]. The depletion

of NMD factors has been observed to induce alterations in cell proliferation and causes cell cycle arrest at the G2/M phase in *Drosophila* [194]. Similarly, the downregulation of UPF1 in HeLa cells elicits an early S phase arrest [198]. These observations collectively underscore the indispensability of NMD for proper cell cycle progression. Despite these insights, the molecular intricacies governing the regulation of the cell cycle by NMD remain elusive.

In the preceding chapter, the identification of the upregulated physiological NMD targets through a differential gene expression analysis served a poignant indicator of compromised NMD activity. In this chapter, a focused examination is undertaken to elucidate the repercussions of UPF3B-KO on both NMD activity and cell cycle progression.

4.2 Materials and Methods

4.2.1 Quantitative Real-Time PCR

Please refer to Chapter 3, Section 3.2.9.

4.2.2 Reporter Assay

1×10^6 Control and UPF3B-KO clones were seeded in 35 mm dishes in three biological replicates. Cells were transfected with 2 µg PTC39 β-globin reporter plasmid using the PEI method. 24-hours post-transfection, total RNA was isolated using TRIzol reagent, and cDNA synthesis was performed, followed by qPCR analysis.

4.2.3 Western Blotting

1×10^6 cells from Control, Clone1, and Clone3 were seeded in 35 mm dishes in three biological replicates followed by lysate preparation, SDS-PAGE, and Western blot analysis. Please refer to Chapter 2, Section 2.2.10-2.2.13.

4.3.4 Rescue Experiment

Please refer to Chapter 3, Section 3.2.5.

4.3.5 Microscopy

All the cell types (Control, Clone1, Clone3, UPF3B-rescued) were seeded in 35 mm dishes in equal number (1×10^6 cells per dish). After 36 hours of seeding, cells were visualized under the microscope (Axiovert A1, Zeiss) with 10X magnification.

4.3.6 Cell Proliferation Assay

1×10^5 of Control, UPF3B-KO clones, and UPF3B-rescued cells were seeded in 24-well plates in triplicates. Relative cell proliferation was determined by cell counting at intervals of 12, 24, 36, and 48 hours using the Trypan blue (Sigma, Cat# 302643) staining method in the hemocytometer.

4.3.7 MTT Assay

MTT assay was used to determine the viability of cells after UPF3B-KO and UPF3B-rescue. Control, Clone1, Clone3, and UPF3B-rescued cells were harvested and seeded in 96-well plates at the density of 5×10^4 per well. The cells were incubated in humidified condition (37°C, 5% CO_2) to allow them to attach. After 36 hours, 50 µL of 5g/mL MTT (Himedia, Cat# TC191) was added to each well. After 4 hours of incubation at 37°C, the culture medium was removed, followed by the addition of 150 µL DMSO to each well to resolve the formazan product. After incubating the cells at room temperature for 15 min, the absorbance of solubilized formazan was measured using a Biorad microplate reader at a wavelength of 570 nm.

4.3.8 Flow Cytometric Analysis of Cell Cycle

1×10^6 cells from the Control and UPF3B-KO cell lines were seeded in 60 mm dishes. After 24 hours of incubation, cells were harvested and transferred to a 1.5 mL microcentrifuge tube. After centrifugation, cell pellets were washed with DPBS and resuspended in 150 µL DPBS. Cell fixation and permeabilization were done by using ice-cold 70% ethanol at -20°C overnight. The cells were recovered as pellets after centrifuging the fixed cells at 700 x g for 15 min. The pellets were treated with RNase solution (150 µL of 200 µg/mL of stock solution) at 37°C for 30 min to remove the chances of RNA staining by propidium iodide (PI; Life Technologies, Cat# V13241). The cells were suspended in 150 µL of PI (60 µg/mL stock solution) to stain the DNA. The samples were incubated for 15 min and analyzed by flow cytometry (BD Biosciences). FCS Express 5 software was used to analyze the output data.

4.3.9 Cell Apoptosis Assay

Control and UPF3B-KO cells SH-SY5Y cells were plated in a six-well plate. At 70% confluency, cells were harvested by centrifugation. The pellets were washed with DPBS with EDTA, and staining was performed with fluorescein isothiocyanate (FITC) – Annexin V/PI (Life Technologies) following the manufacturer's protocol. Briefly, cells

were resuspended in Annexin V binding buffer. Next, the cells were stained with FITC-Annexin V for 30 min at 4°C in the dark. PI was added to the stained cells and analyzed by flow cytometry (BD Biosciences).

4.3.10 KEGG Pathway Analysis

The gage package was used for pathway analysis [199]. The *gage()* function requires a named vector of fold changes, where the names of the values are the Entrez gene IDs. The gageData package has pre-compiled databases mapping genes to KEGG pathways and GO terms for common organisms. kegg.sets.hs is a named list of 229 elements. Each element is a character vector of member gene Entrez IDs for a single KEGG pathway. The input file for analysis was prepared as the list of DEGs with padj > 0.001 and log2FC > 1.2 and < -1.2. Finally, the *pathview()* function in the pathview package makes the plots.

4.2.10 Statistical Analysis

Please refer to Chapter 3, Section 3.2.11.

4.3 Results

4.3.1 Impact of UPF3B-KO on NMD Activity

To assess the impact of UPF3B-KO on NMD activity, an analysis of the transcript levels of endogenous NMD substrates, namely *GAS5*, *ITGAE*, *SRSF2*, and *ZFAS1*, was conducted [200] (Figure 4.1A-B). *GAS5* and *ITGAE* transcripts, characterized by retained introns, are recognized as NMD-sensitive targets. Subsequent qRT-PCR analysis demonstrated a significant upregulation in the levels of *GAS5* (~1.75-fold with a p-value of 0.0025 in Clone1 and ~3.5-fold with a p-value of 0.0001 in Clone3) and *ITGAE* (~1.5-fold with a p-value of 0.0011 in Clone1 and ~1.6-fold with a p-value of 0.0012 in Clone3), both of which are NMD-sensitive transcripts. SRSF2, a splicing factor featuring an SR-rich domain, exhibits autoregulation of its mRNA levels through the inclusion of exon 3, resulting in the generation of an NMD-sensitive isoform [201]. Notably, a ~2-fold increase in the abundance of the *SRSF2* NMD-sensitive transcript was observed (p-values 0.0006 and 0.0005 in Clone1 and Clone3, respectively) compared to its protein-coding counterpart. *ZFAS1*, an NMD substrate belonging to the small nucleolar RNA family and harboring a short PTC-containing ORF [202], displayed a significant ~2-fold (p-value 0.0024) elevation in the RNA levels specially in Clone3, whereas no discernible differential regulation was evident in Clone1. These findings collectively underscore the impact of

UPF3B-KO on the NMD activity, as reflected by the differential expression of established NMD-sensitive targets.

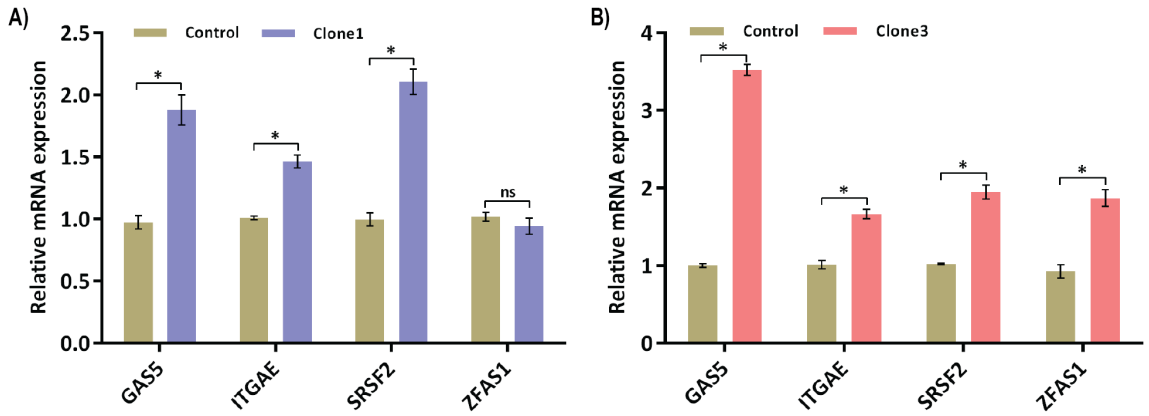


Figure 4.1. Transcript abundance of canonical NMD substrates. The qPCR analysis showing the abundance of NMD-sensitive transcripts, *GAS5*, *ITGAE*, *SRSF2*, and *ZFAS1* in A) Clone1 and B) Clone3. The expression levels of transcripts were normalized to that of the Actin transcripts. For *SRSF2* only, the ratio of NMD isoform to canonical isoform was calculated. The values shown are the average fold change relative to the control cells. Error bars indicate standard error of mean. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

In addition to the analysis of endogenous targets, a bona fide NMD reporter, namely PTC39 β-globin [150], was ectopically introduced into both UPF3B-KO clones (Figure 4.2A). Quantitative assessment of PTC39 β-globin mRNA levels unveiled a substantial ~2.5-fold (p-value 0.0079) upregulation in Clone1 and even more pronounced ~5-fold (p-value 0.0136) upregulation in Clone3 (Figure 4.2B). The observed augmentation in both canonical NMD substrates and PTC39 β-globin mRNAs within the UPF3B-KO clones serves as indicative evidence of compromised NMD activity. This dual manifestation further reinforces the impact of UPF3B-KO on the regulation of NMD, emphasizing the perturbation of the decay pathway for transcripts harbouring premature termination codons.

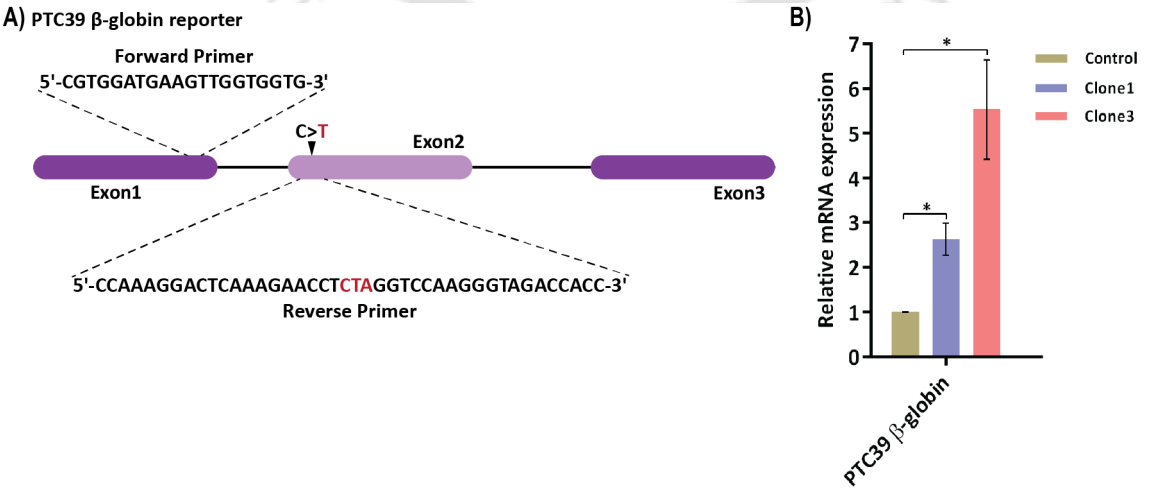


Figure 4.2. Quantification of NMD reporter. The qPCR analysis showing the abundance of PTC39 β-globin reporter in *UPF3B*-KO clones. [Unpaired Student's t-test, n=3, *p-value < 0.05]

4.3.2 Effect of UPF3B deletion on expression of other NMD factors

Given the pivotal role of UPF3B within the SURF complex, acting as a crucial mediator in bridging the PTC-stalled ribosome to the EJC, an investigation was conducted to determine the potential impact of UPF3B absence on other essential NMD factors. Analysis of RNA-Seq data revealed a significant upregulation of *UPF1*, with log2FoldChanges of 0.77 in Clone1 and 1.3 in Clone3, respectively. To corroborate these findings, both qRT-PCR and Western blot analyses were conducted, consistently affirming the increased expression of *UPF1* at both the transcript level (~1.21-fold with a p-value of 0.0199 in Clone1 and ~1.73-fold with a p-value of 0.0088 in Clone3) and protein level (~1.53-fold with a p-value of 0.0033 in Clone1 and ~1.72-fold with a p-value of 0.0004 in Clone3) (Figure 4.3A-B). These results collectively signify a compensatory upregulation of UPF1 in response to the absence of UPF3B, suggesting a potential interplay between these critical components in the intricate regulatory network of the NMD pathway.

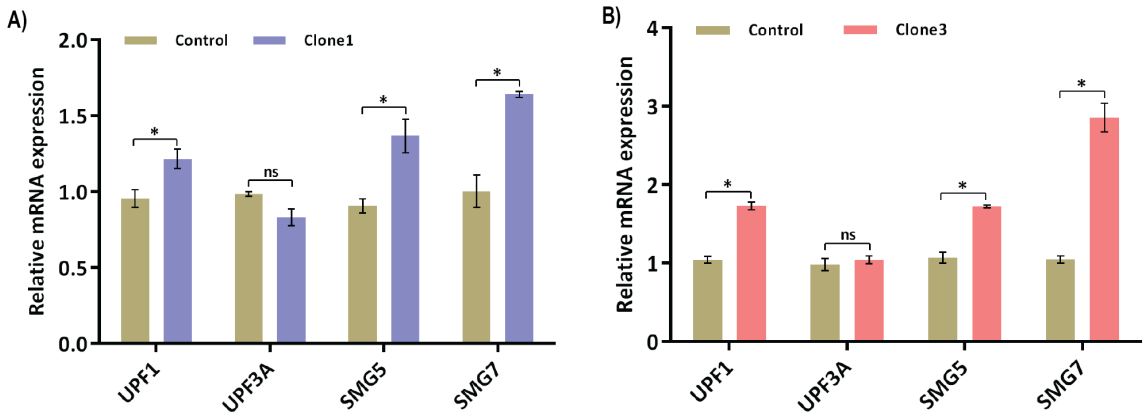


Figure 4.3 Transcript expression of NMD factors. The qPCR analysis of *UPF1*, *UPF3A*, *SMG5*, and *SMG7* in A) Clone1 and B) Clone3. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

The *UPF3* gene family comprises paralogous members – UPF3A and UPF3B. To comprehensively examine the impact of UPF3B-KO, an investigation into the influence on UPF3A expression was undertaken. Consistent with prior literature [142], mRNA quantification revealed an unaltered *UPF3A* transcript level following UPF3B-KO (Figure 4.3A-B). In contrast, UPF3A protein levels exhibited a discernible increase, indicating a stabilization effect of approximately ~1.41-fold (p-value 0.0032) in Clone1 and ~1.26-fold (p-value 0.0003) in Clone3 (Figure 4.4A-B). These findings align with earlier research indicating that UPF3A tends to experience stabilization under UPF3B-deficient conditions

[120, 142]. This implies a partial compensatory role of UPF3A in the absence of UPF3B, although it is noteworthy that this compensatory mechanism appears insufficient to instigate the degradation of canonical NMD targets. The nuanced interplay between UPF3A and UPF3B underscores the intricacies of their regulatory dynamics within the context of the NMD pathway.

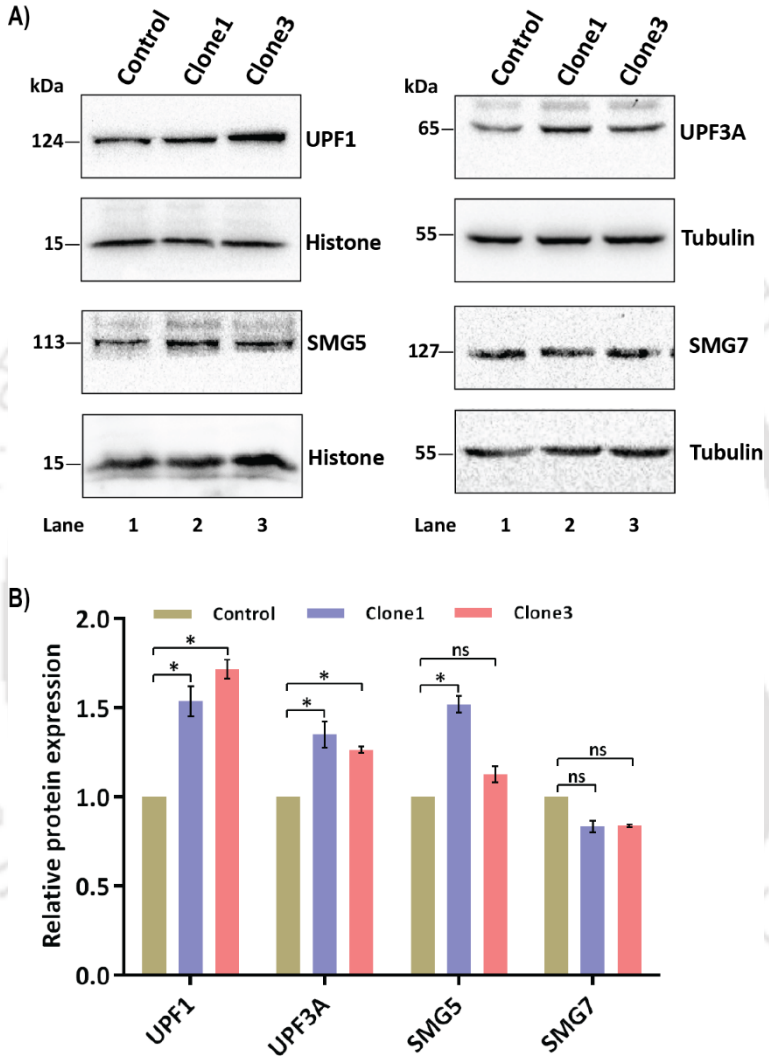


Figure 4.4. Protein expression of NMD factors. A) Western blot analysis of UPF1, UPF3A, SMG5 and SMG7 in UPF3B-KO clones. Histone and tubulin are used as internal control. B) Quantification of UPF1, UPF3A, SMG5 and SMG7 proteins. Protein expression was normalized with histone and tubulin. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

Subsequently, I conducted an evaluation of downstream effectors within the NMD pathway, specifically examining SMG5 and SMG7, which form the SMG5-SMG7 heterodimer and interact with UPF1. Analysis of the RNA-Seq data revealed a significant upregulation of SMG5 in both Clone1 and Clone3. Correspondingly, qRT-PCR analysis showed a substantial elevation in *SMG5* mRNA levels by approximately ~1.37-fold (p-

value 0.0204) in Clone1 and ~1.72-fold (p-value 0.012) in Clone3 (Figure 4.3A-B). Further validation through Western blot analysis confirmed these findings, revealing an approximately ~1.52-fold increase in SMG5 protein levels in Clone1 and a ~1.12-fold increase in Clone3 (Figure 4.4A-B). Regarding *SMG7*, both transcriptome data and qRT-PCR analysis demonstrated significant increases in transcript expression, exhibiting fold changes of 1.64 (p-value 0.0192) in Clone1 and 2.86 (p-value 0.0012) in Clone3 (Figure 4.3A-B). However, alterations in SMG7 protein expression were deemed insignificant, displaying only marginal fold changes (~0.83-fold) in UPF3B-KO clones (Figure 4.4A-B). These results illuminate the regulatory dynamics of SMG5 and SMG7 in response to UPF3B deficiency, highlighting potential compensatory mechanisms at the transcript level, albeit with disparities at the protein level.

I further examined SMG6, an additional NMD factor within the NMD pathway, and observed no significant difference at the transcript level (Figure 4.5A). A modest upregulation was, however, noted at the protein level (Figure 4.5B-C). Collectively, these results suggest that, beyond its impact on physiological and NMD-sensitive transcripts, UPF3B may exert regulatory influence on the mRNA levels of other NMD factors, evident at both the transcript and protein levels.

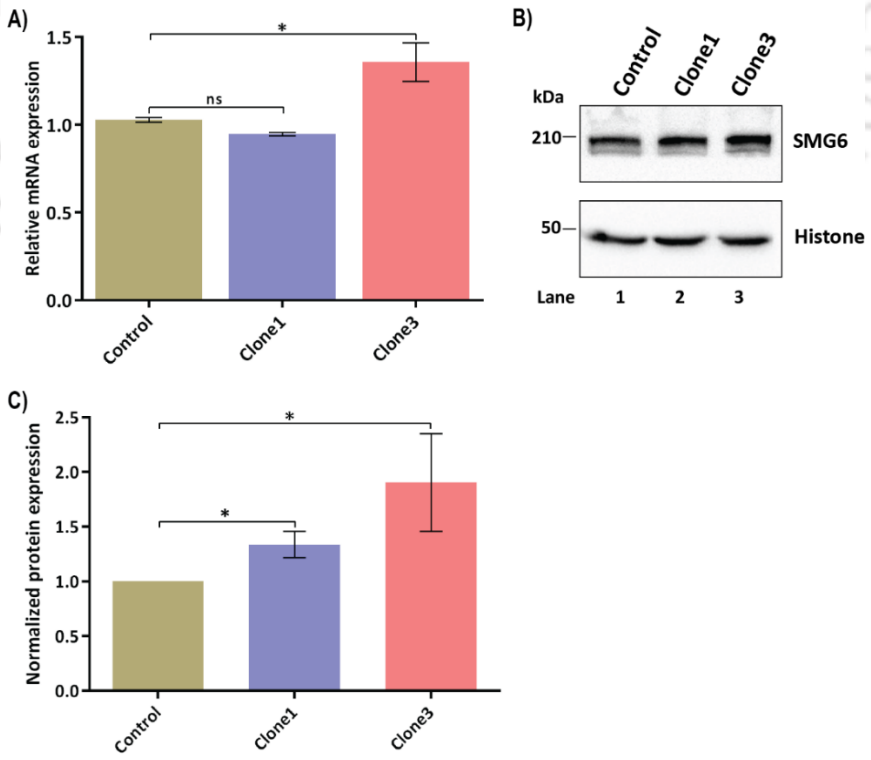


Figure 4.5. Expression of SMG6. A) The qPCR data showing the *SMG6* mRNA levels in Control, Clone1 and Clone3 cells. B) Western blot image and C) quantification of SMG6 protein in Control, Clone1 and Clone3. Histone was used as internal control. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

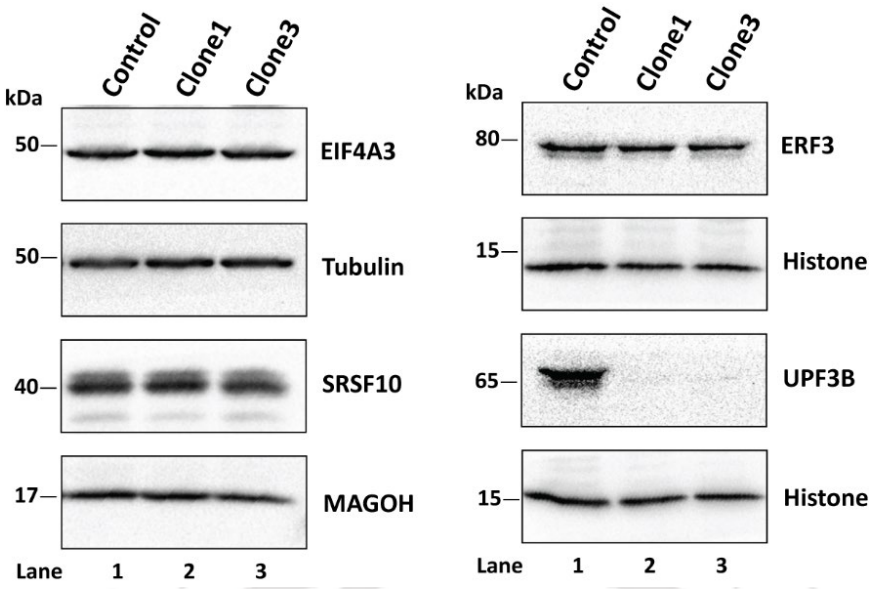


Figure 4.6. Expression of EJC and other factors. Western blot analysis of eIF4A3, SRSF10, MAGOH, eRF3, and UPF3B protein in Control and UPF3B-KO clones. Tubulin was used as the internal control for eIF4A3, SRSF10, and MAGOH. Histone served as the control for eRF3 and UPF3B.

Considering the involvement of EJC factors in NMD, subsequent Western blot analyses were conducted to assess the impact of UPF3B-KO on core EJC factors, specifically eIF4A3 and MAGOH (Figure 4.6). No significant alterations were observed in the expression of these core EJC factors. Additionally, an examination of the eukaryotic release factor eRF3 and a splicing factor SRSF10, revealed comparable protein levels in both Control and KO clones. These findings collectively indicate that UPF3B-knockout may induce deregulation of mRNA and protein levels of NMD factors leaving the expression of EJC proteins unaffected.

4.3.3 UPF3B-KO delays cell proliferation rate

During the course of cell culture experiments, a distinctive disparity in the growth pattern of UPF3B-KO clones became apparent. Notably, the cell growth of UPF3B-KO cells exhibited a delayed trajectory compared to the Control cells (Figure 4.7A). To explicate this discrepancy in cell growth dynamics, a cell proliferation assay using Trypan blue staining was performed, revealing a decelerated proliferation rate characterized by an increased population doubling time. Specifically, among the UPF3B-KO clones, Clone1 manifested an approximately 1.5 times reduction in proliferation, attaining a doubling time of 42 hours, while Clone3 exhibited a doubling time of 30 hours in contrast to the 24 hours observed in Control cells (Figure 4.7B). Furthermore, a cell viability assay employing MTT reagent demonstrated a significant reduction in cell number exclusively in Clone1 (Figure

4.7C). Consequently, an in-depth examination of Clone1 was undertaken to a more comprehensive understanding to the observed delay in cell proliferation.

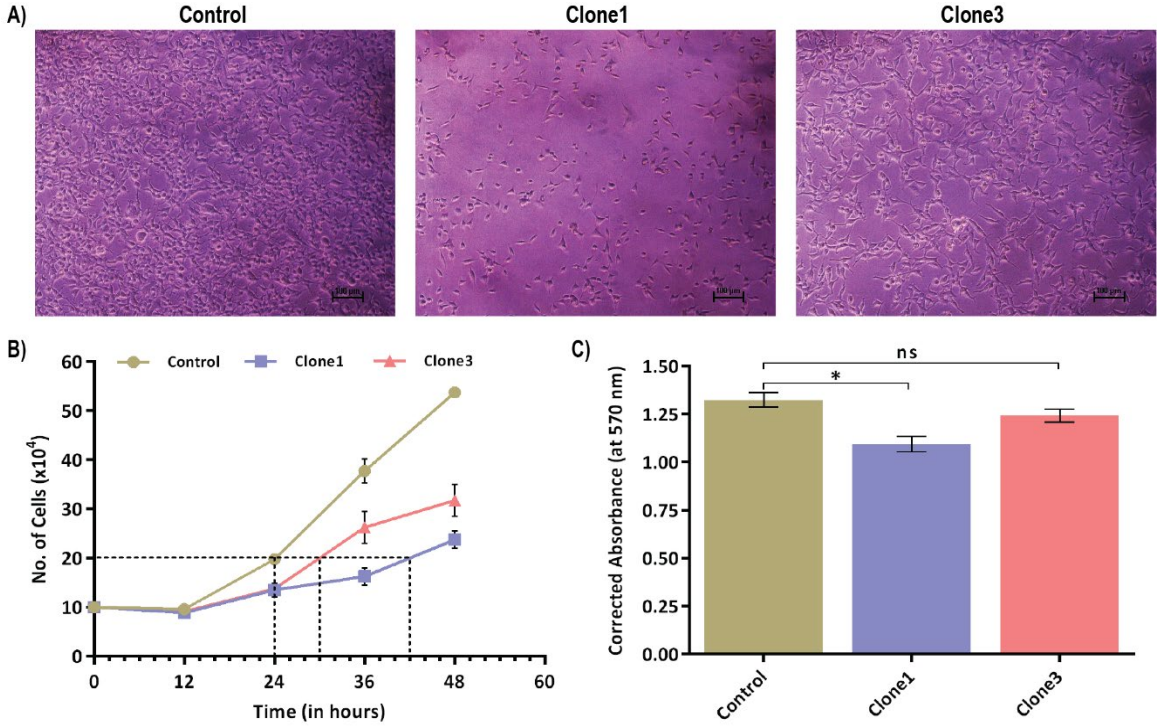


Figure 4.7. UPF3B-KO delays cell growth and proliferation. A) Microscopic images showing the difference in cell population in Control and UPF3B-KO clones. Started with equal number of cells and visualized after 36 hours of incubation (10X objective). B) Cell proliferation assay using Trypan blue staining showed increase in doubling time of UPF3B-KO clones compared to Control. The number of cells (y-axis) are plotted with the time (in hours; x-axis). Dashed line on y-axis shows the number of cells twice to the seeding density and that on x-axis indicates population doubling times of respective samples. C) The corrected absorbance of Control and UPF3B-KO clones measured at 570 nm wavelength. The absorbance is directly proportional to the number of cells. Error bars indicate standard deviation. [Unpaired Student's t-test, n=5, ns – not significant, *p-value < 0.05]

4.3.4 Effect of UPF3B-KO on Cell Cycle Progression

Subsequently, to elucidate the mechanism underlying the decelerated proliferation of UPF3B-KO cells, an examination of cell cycle progression was conducted using flow cytometry (Figure 4.8). Mitotic phase analysis showed a notable shift, with a higher proportion of UPF3B-KO cells (53.2%) occupying the S phase compared to Control cells (41.5%). Conversely, a diminished presence of UPF3B-KO (37.1%) was observed in the G1 phase, in contrast to the higher representation of Control cells (49.8%). These findings suggest a potential link of UPF3B in modulating cell cycle progression. Given the reduced cell number in Clone1 (Figure 4.7C), it became imperative to discern whether this decline stemmed from a decelerated proliferation rate or enhanced cell death. Therefore, Annexin-V staining was employed to investigate apoptosis. The flow cytometry data showed no discernible difference in apoptotic rates between UPF3B-KO and Control cells, affirming

no direct involvement of UPF3B in the apoptotic pathway (Figure 4.9). This implies that the reduced cell number in Clone1 is primarily attributed to a slower proliferation rate rather than an increase in apoptotic events.

4.3.5 Expression of cell cycle genes upon UPF3B deletion

To elucidate the molecular basis underlying slower proliferation rate in UPF3B-KO cells, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DEGs (shown in volcano plots of Figure 3.7 as red dots) was conducted. This analysis identified 17 genes associated with the cell cycle pathway (hsa04110), wherein all 16 genes, with the exception of

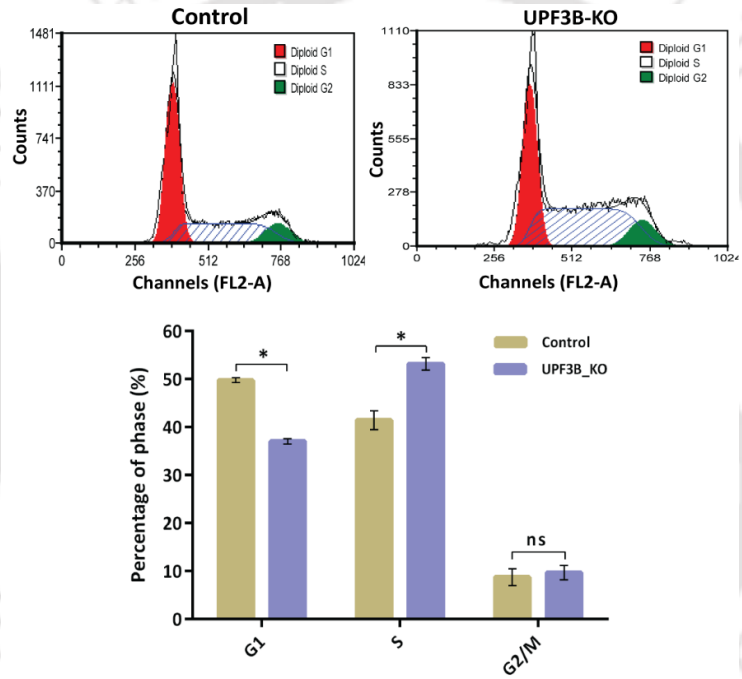


Figure 4.8. Effect of UPF3B-KO on mitotic phases. Flow cytometry analysis for cell cycle distribution of Control and UPF3B-KO cells. The result of one representative assay from three similar independent experiments is shown. x- and y-axes denote DNA content and cell number, respectively. Each phase was calculated by using the FCS express5 software. The percentages of cells in G1, S, and G2/M were also shown as indicated. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

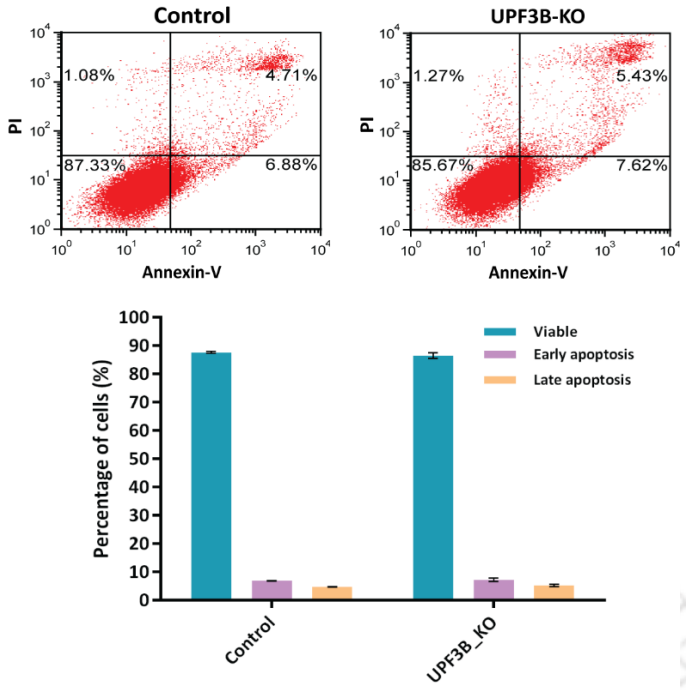


Figure 4.9. Apoptosis remains unaffected upon loss of UPF3B. Apoptosis rate of Control and UPF3B-KO cells was tested using flow cytometry (upper panel). The corresponding linear diagram of flow cytometry was also shown (lower panel).

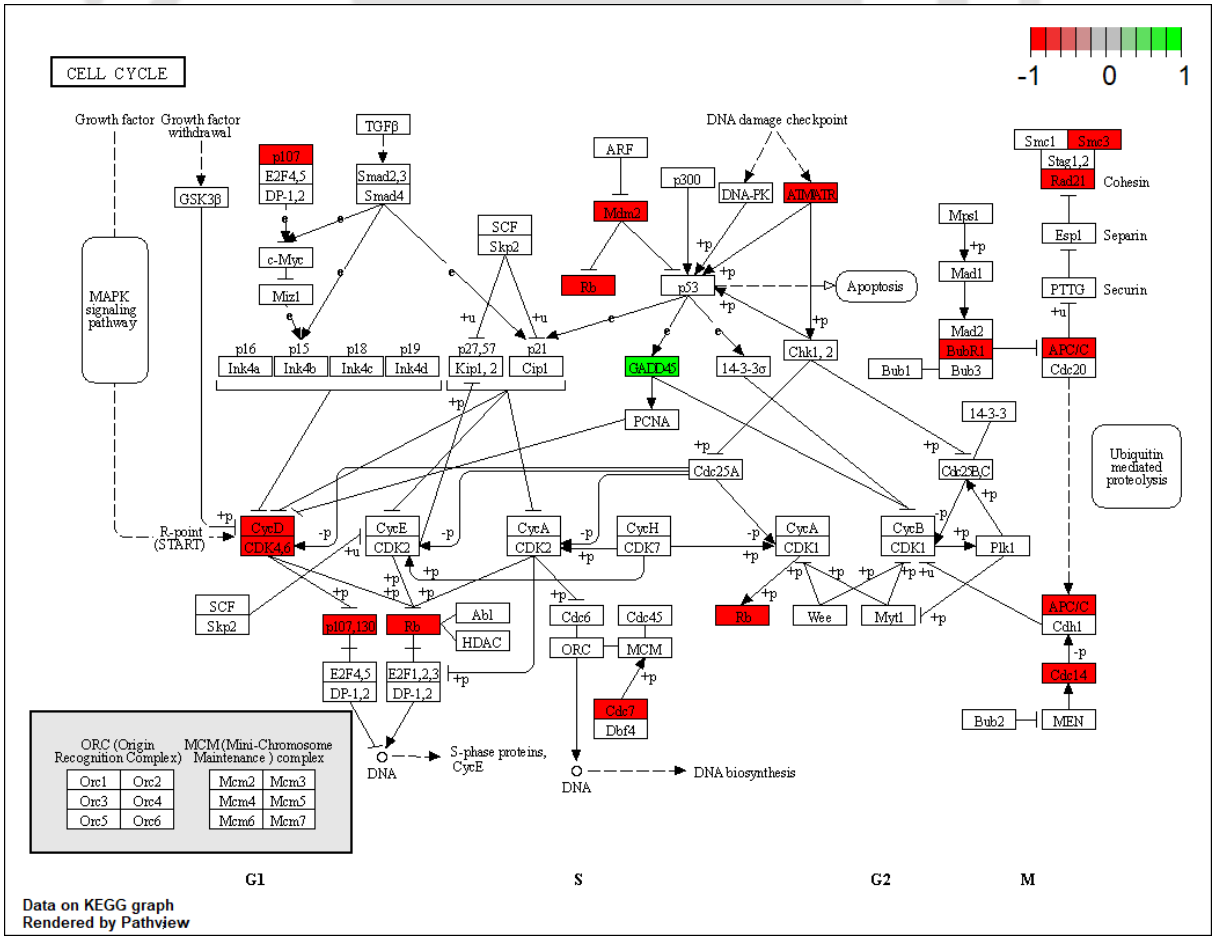


Figure 4.10. The cell cycle pathway. KEGG pathway analysis of DEGs (Adjusted p-value < 0.001; log2FC > |-1.2|). Graph of genes and processes involved in the pathway hsa04110: Cell cycle. The scheme shows genes and relationships between them. Additionally, the diagram indicates the genes up (green) or downregulated (red) in RNA-Seq data. Solid arrows indicate a direct effect on specific gene expression while dotted arrows indicate gene involvement in a process or pathway.

GADD45G, exhibited downregulation (Figure 4.10). As discussed in the preceding chapter, *GADD45* transcripts are known to undergo routine degradation to facilitate cell viability [178], and the observed upregulation in UPF3B-KO cells may contribute to the delayed growth. Also, the transcript abundance of a number of cell cycle genes, including *BUB1B*, *RBL1*, *RBL2*, *CDC7*, *CDK6*, *SMC3*, *CDC27*, and *PRKDC*, showed significant decreases in expression in UPF3B-KO cells (Figure 4.11A-B), thereby establishing a link to the slower proliferation rate in the KO cells. Importantly, the mRNA expression of DNA damage checkpoint genes, *ATM* and *ATR*, remained unaltered, suggesting that the observed effects are not attributed to DNA damage. To determine the direct regulatory influence of UPF3B on cell cycle genes, the KO cells were transfected with plasmid expressing the UPF3B. The UPF3B-rescued cells were analyzed for the transcript expression. The results showed complete rescue exclusively for *GADD45G*, as illustrated in Chapter 3, Figure 3.9A.

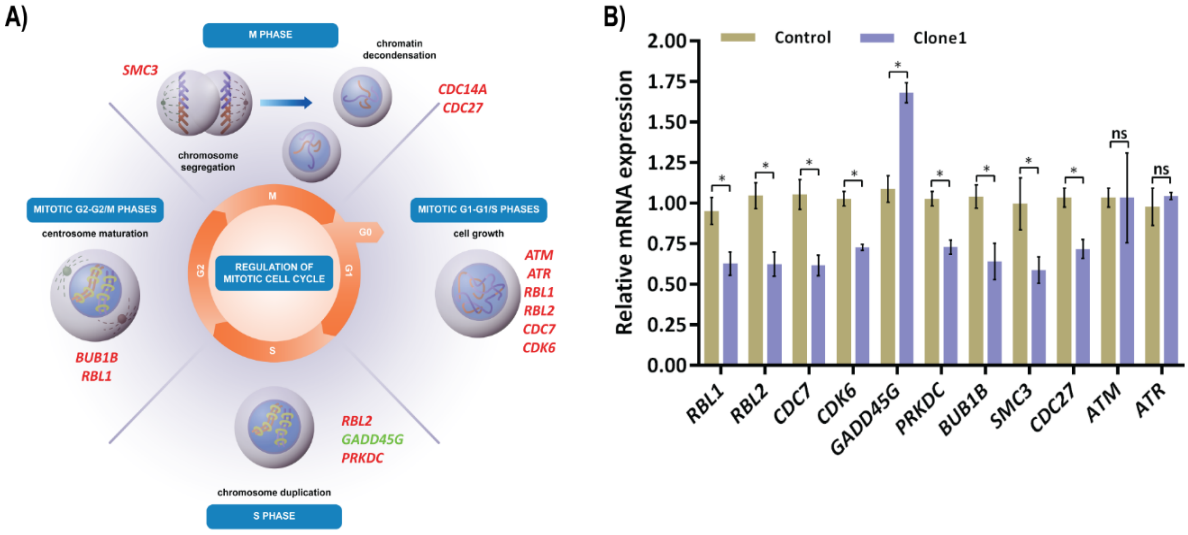


Figure 4.11. Expression of cell cycle genes. A) Schematic representation of cell cycle genes associated with various mitotic phases. B) Bar graphs showing the transcript expression of cell cycle genes. [Unpaired Student's t-test, n=3, ns – not significant, *p-value < 0.05]

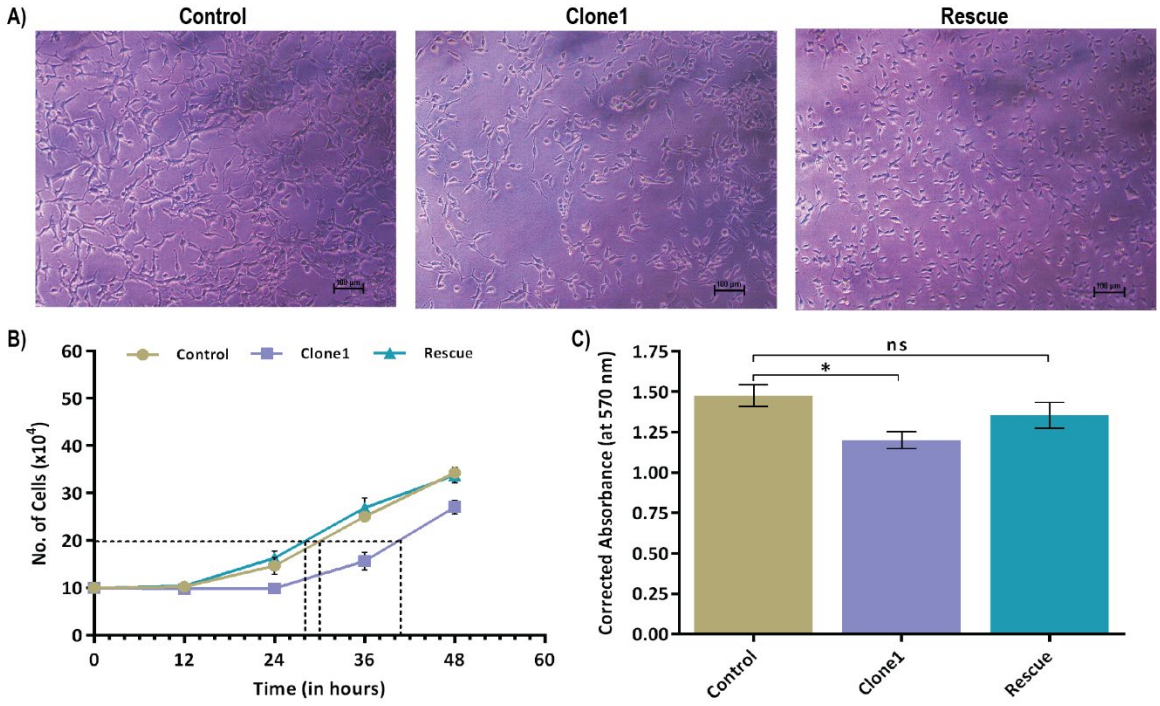


Figure 4.12. UPF3B expression rescues the proliferation of knockout cells. A) Microscopic images showing the difference in cell population in Control and UPF3B-KO and UPF3B-rescued cells. Started with equal number of cells and visualized after 36 hours of incubation (10X objective). B) Cell proliferation assay using Trypan blue staining showed increase in doubling time of UPF3B-KO clones and restoration upon ectopic expression of UPF3B compared to Control. The number of cells (y-axis) are plotted with the time (in hours; x-axis). Dashed line on y-axis shows the number of cells twice to the seeding density and that on x-axis indicates population doubling times of respective samples. C) The corrected absorbance of Control, UPF3B-KO, and UPF3B-rescued cells measured at 570 nm wavelength. Error bars indicate standard deviation. [Unpaired Student's t-test, n=5, ns – not significant, *p-value < 0.05]

Furthermore, the restoration of UPF3B expression in knockout cells resulted in a complete recovery of cell proliferation (Figure 4.12A-B). The doubling time of UPF3B-KO cells with ectopic UPF3B expression closely resembled that of Control cells, underscoring the crucial role of UPF3B in cell growth and proliferation. However, other cell cycle genes like *RBL1*, *CDC7*, *CDK6* could not be rescued upon ectopic expression of UPF3B, suggesting that these genes are part of complex regulatory network and UPF3B alone may not be sufficient to override the regulatory mechanisms and are not directly regulated by UPF3B.

4.4 Discussion

This study reveals that the knockout of UPF3B in HEK-293 cells abolishes NMD activity, as evidenced by the upregulation of canonical NMD substrates and the NMD reporter PTC-39 β-globin transcripts in both KO clones. These results are consistent with prior investigation where siRNA knockdown of UPF3B similarly led to the upregulation of NMD reporter PTC-39 β-globin transcripts (as depicted in Figure 1c of Chan *et al.* 2007 [84]). Likewise, UPF3B knockout in HCT-116 cells showed increased levels of NMD substrates [149], indicating the disruption of NMD pathways following UPF3B deletion. Notably, the observed elevation of canonical NMD substrates upon UPF3B knockout in this study contrasts with the findings of Wallermoth *et al*., where stabilization of canonical NMD transcripts was reported only in the absence of both UPF3A and UPF3B factors (Figure 4c in [150]). The incomplete KO of UPF3B in the Wallmeroth *et al.*, may possibly result in the residual or truncated version of UPF3B protein, which may cause these discrepancies in the NMD activity. An upregulation of NMD factor transcripts, including *UPF1*, *SMG5*, *SMG6*, and *SMG7*, was observed in the UPF3B-KO mouse models studied by Huang *et al.* [196]. Consistently, my work and analysis also demonstrated elevated transcript expression of NMD factors *UPF1*, *SMG5*, and *SMG7* in UPF3B-KO clones. This suggests the existence of an autoregulatory circuit within the NMD pathway.

The investigation displayed a notable reduction in cell growth in UPF3B-KO cells compared to the Control cells. Cell cycle analysis further showed a distinctive distribution of knockout cells, with an increased presence in the S phase and a decreased presence in the G1 phase compared to the Control cells. The diminished distribution of UPF3B-KO cells in the G1 phase indicates a bypassing of the preparatory phase for DNA replication. In contrast, the elevated distribution of UPF3B-KO cells in the S phase, indicates heightened efforts in DNA replication and transcription processes, potentially due to the deregulation of many transcription factors. KEGG pathway analysis identified several downregulated genes associated with the cell cycle pathway in UPF3B-KO cells. Notably, key regulators of the cell cycle pathway, including *BUB1B*, *RBL1*, *RBL2*, *CDC7*, *CDK6*, *PRKDC*, *CDC27*, and *SMC3*, showed significant downregulation in UPF3B-KO cells. BUB1B, also known as BUBR1, is an essential component of the mitotic checkpoint, has been implicated in accelerating cell entry into the M phase when depleted in the G2 phase [203]. BUB1B is highly upregulated in various cancer types, including lung cancer cells [204], gastric cancer [205], breast cancer cell lines and primary tumor samples [206]; thus,

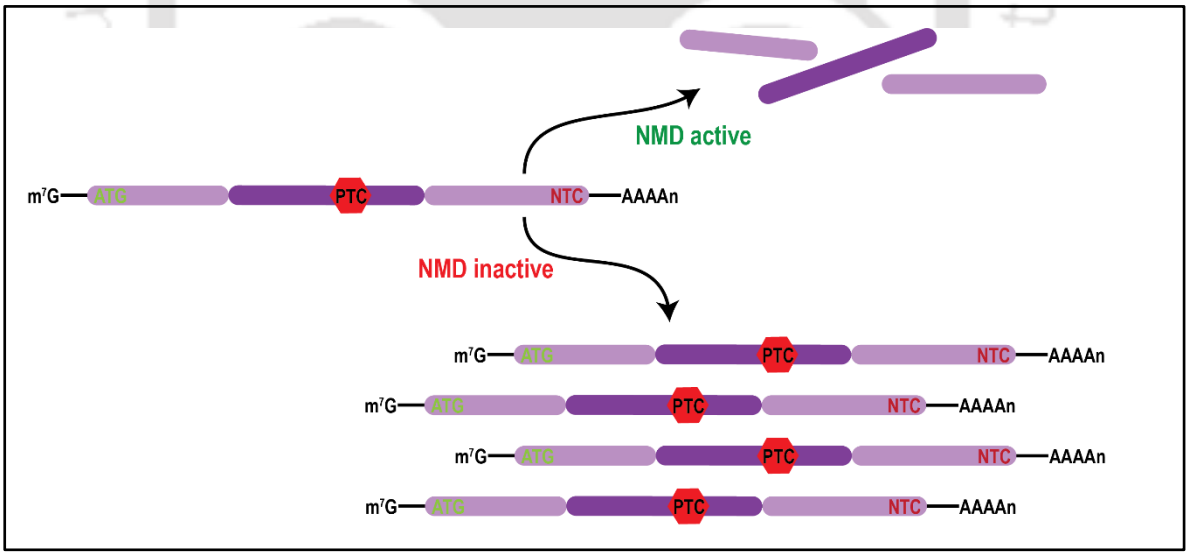
reflecting active cell division. Therefore, the decreased expression of BUB1B might be associated with the compromised growth of UPF3B-KO cells. Other downregulated cell cycle genes, such as *RBL1*, *RBL2*, *CDC7*, *CDK6*, and *SMC3*, are involved in the regulation of cell cycle progression. Mutations in *CDK6* and *SMC3* are associated with primary autosomal recessive microcephaly [207] and cornelia de Lange syndrome with predominant mental retardation [208], respectively, suggesting potential implications for neuronal processes and defects in brain function and development. Conversely, the upregulation of *GADD45G*, a growth regulator, in knockout cells indicates its possible role in delayed growth of cells. Importantly, *Gadd45g* overexpression has been demonstrated to inhibit cell proliferation in *Xenopus* and medaka fish [190, 209], emphasizing its conserved function in regulating cell proliferation in mammals. While previous studies identified the upregulation of *GADD45* in NMD-deficient cells leading to cell death [178]. The present study did not observe any significant effect on cell viability upon UPF3B loss. This suggests that the increased expression of *GADD45G* may not be sufficient to exert its negative effect on the viability of UPF3B-KO cells. The unaltered expression of DNA damage sensors, *ATM* and *ATR*, aligns with the observation that apoptosis was not elevated in UPF3B-KO cells. The reported delayed growth of UPF3B-depleted cells [84], was effectively restored by rescuing UPF3B expression in UPF3B-KO cells, indicating an active engagement of UPF3B in cell growth, although the detailed molecular mechanism underlying this phenomenon needs further exploration. In summary, these results suggest that UPF3B may function as an important regulator of cell growth and proliferation.

4.5 Conclusion

The knockout of UPF3B in HEK-293 cells resulted in impaired Nonsense-Mediated mRNA Decay (NMD) activity. This impairment, in turn, triggered the activation of the autoregulatory circuitry within the NMD pathway. As a pioneer piece of evidence, the deletion of UPF3B influenced cell proliferation without undergoing apoptosis. Notably, the preliminary insight into the expression of the cell cycle genes displayed significant downregulation in their mRNA levels in UPF3B-KO cells. This suggests that UPF3B plays a regulatory role in a set of NMD transcripts and normal physiological transcripts, contributing to the proper functioning of cells. The observed alterations in NMD activity and cell cycle progression underscore the multifaceted regulatory functions of UPF3B in cellular processes that requires further investigations.

Chapter 5:

Identification of Novel NMD-Sensitive Transcripts



Chapter 5

Abstract

UPF3B-knockout demonstrated a subtle impairment in NMD activity in HEK-293 cells, as evidenced by the upregulation of canonical NMD substrates. This upregulation of NMD substrates enables the identification of novel NMD-sensitive transcripts regulated by UPF3B. To identify the NMD-sensitive transcripts, computational analysis was performed on RNA-Seq data obtained from UPF3B-KO clones. Differential transcript expression analysis using DESeq2 and differential transcript usage analysis using IsoformSwitchAnalyzeR identified 165 NMD transcripts and 153 novel PTC+ transcripts, respectively, in UPF3B-KO clones, highlighting their regulation by UPF3B-mediated NMD. Subsequently, isoform switching analysis demonstrated an increased usage of NMD-sensitive splice isoforms, such as *SAT1* and *HNRNPA2B1*, underscoring the splicing-associated regulatory role of UPF3B. These findings suggest that UPF3B is associated with alternative splicing coupled with NMD (AS-NMD) and functions to restrain unproductive splicing events. Altogether, my results reveal novel NMD substrates of the UPF3B-dependent surveillance pathway, indicating that UPF3B is required for NMD function in a context-dependent manner.

Chapter 5

5.1 Introduction

The spliced mRNA is exported to the cytoplasm in the form of an RNP complex [76]. UPF3B interacts with the core EJC factors, MAGOH and Y14 [65, 76], thereby being transported as a component of the RNP complex through protein-protein interaction. The presence of RNA recognition motif (RRM) domain in UPF3B (Chapter 1, Figure 1.3), is required for RNA-binding activity, indicates its possible interaction with RNA. However, direct binding of UPF3B to RNA was not initially reported. Neu-Yilik *et al*. established the RNA-binding activity of UPF3B *in vitro* using size-exclusion chromatography [68]. Later, an investigation employing the iCLIP experiment corroborated the direct binding of UPF3B to RNA molecules [32]. A recent study used a fluorescence anisotropy-based binding assay to demonstrate the binding of UPF3B with single stranded RNA (ssRNA) and single stranded DNA (ssDNA) [82]. Interestingly, UPF3B has a higher affinity for double stranded RNA and double stranded DNA compared to ssRNA and ssDNA. The RNA-binding activity of UPF3B is associated with the RRM-L, NOPS-L, and CCL1 regions, which are also required for UPF3B's interaction with UPF2. The UPF3B-UPF2 interaction is stronger than the UPF3B-RNA interaction, indicating that the RNA binding activity of UPF3B prevails in the absence or depletion of UPF2. Altogether, these findings drove the researchers to identify the transcripts regulated by UPF3B.

A microarray analysis was performed on UPF3B-depleted and UPF1-depleted HeLa cells, in comparison to control HeLa cells, to identify the wild-type transcripts regulated by UPF3B [84]. The analysis identified that approximately 4% and 2% of transcripts were upregulated by at least two-fold (p-value < 0.01) in UPF1-depleted and UPF3B-depleted HeLa cells, respectively [84]. Many of the UPF1-regulated transcripts did not show significant regulation by UPF3B, indicating their specificity as substrates for the alternative NMD pathway. Therefore, it becomes intriguing to determine the novel transcripts regulated by UPF3B-mediated NMD. Subsequently, UPF3B-KO cell lines were generated in HCT116 and HEK-293 cells by two independent research groups to study the effect of UPF3B on NMD and gene expression [149, 150]. Transcriptome-wide analysis of UPF3B-depleted HCT116 cells revealed 819 upregulated transcripts, indicative of their association with UPF3B-dependent NMD [149]. RNA-Seq analysis of UPF3B-KO HEK-293 cells identified 74 upregulated genes, with only 5 upregulated transcripts containing PTC [150]. This observation highlights the cell type-dependent role of UPF3B in influencing gene expression and transcript regulation. Further investigation is required to

disclose the features of NMD transcripts regulated by UPF3B and to delineate the associated biological processes.

In this chapter, differential transcript expression analyses on UPF3B-KO clones were performed to identify novel deregulated transcripts in the absence of UPF3B. Subsequently, splicing enrichment analysis and differential transcript usage analysis were performed that revealed NMD-inducing splicing events and the usage patterns of NMD transcripts in UPF3B-KO cells.

5.2 Materials and Methods

5.2.1 Differential Transcript Expression Analysis

The wicked-fast decoy-aware transcript assembly was performed using Salmon (v1.10.2) [210]. A database was prepared from annotation available in gencode.v41.primary_assembly.annotation.gtf file obtained from the GENCODE database using the GenomeFeatures package [211] to link transcript names with gene names. The transcript quantification files were imported to the RStudio environment using the Tximport package [212] with the following parameters: *txIn = TRUE, txOut = TRUE, tx2gene = tx2gene, ignoreAfterBar = TRUE*. Differential transcript expression analysis was performed using the DESeq2 R package [176]. The transcripts with zero counts were filtered out from the DESeq2 analysis. The Benjamini-Hochberg method was used to correct the Wald test p-values for multiple testing. Transcripts with thresholds log2FoldChange > |1.2| and adjusted p-value < 0.05 were considered as significant. The transcript biotypes of differentially expressed transcripts were obtained from the GENCODE annotation. Transcripts annotated as nonsense-mediated decay were selected, and gene ontology analysis was performed using common NMD transcripts in both the UPF3B-KO clones using the ShinyGO tool (v0.77) [213].

5.2.2 mRNA Stability Assay

Please refer to Chapter 3, section 3.2.10.

5.2.3 Alternative Splicing and Isoform Switching Analysis

The estimates of full-length transcript abundance were generated by the Salmon tool. These quantification files (quants.sf) were imported to the IsoformSwitchAnalyzeR package [214] via the *importIsoformExpression()* function. The resulting count and abundance matrix was used to generate a switchAnalyzeRlist with the *importRdata()* function by supplying an isoform annotation file

(gencode.v41.primary_assembly.annotation.gtf.gz) and a nucleotide fasta file (gentrome.fa.gz). Genes with single isoform and non-expressed isoforms were filtered out by using the *preFilter()* function with the following parameters: *geneExpressionCutoff = 1, isoformExpressionCutoff = 0, removeSingleIsoformGenes = TRUE*. The filtered transcripts with dIF > |0.1| and adjusted P-value < 0.05 were analyzed for isoform switching using DEXSeq method [175, 215]. The isoforms were annotated with the ORFs using *analyzeORF()* function. Knowing the CDS/ORF of transcripts, the nucleotide and amino acid sequences were extracted as fasta files. Nucleotide fasta files were used in CPC2 to predict the coding potential of transcripts [216]. The amino acid sequences were used to predict the domains using Pfam [217] and signal peptides using SignalP [218] tools. The analyses of external tools were imported into IsoformSwitchAnalyzeR to analyze alternative splicing and predict isoform switching.

5.2.4 Determining the transcripts with PTC

IsoformSwitchAnalyzeR also determined the PTC status of isoforms annotated with ORF using the 50-nt rule of NMD [214, 219]. The transcript isoforms having no annotation in GENCODE were designated “NA” in the PTC analysis.

5.2.5 Quantitative Real-Time PCR

Please refer to Chapter 3, section 3.2.9.

5.2.6 Statistical Analysis

Please refer to Chapter 3, Section 3.2.11.

5.3 Results

5.3.1 Differential Expression of NMD Transcripts

The results from the previous chapter suggest an impairment in NMD activity upon UPF3B-KO, manifested by the upregulation of canonical NMD targets. In order to pinpoint the transcripts regulated by UPF3B-dependent NMD, transcriptome-wide analysis was performed on KO clones. Differential transcript expression (DTE) analysis identified 9447 and 2725 differentially regulated transcripts in Clone1 and Clone3, respectively, with a log2FoldChange > |1.2| and an adjusted p-value < 0.05. These transcripts were further annotated to categorise them into major isoform biotypes such as protein-coding, retained introns, processed transcripts, long non-coding RNAs, and nonsense-mediated decay RNAs (Figure 5.1A). In line with the objective of discerning UPF3B-specific NMD targets, isoform characterization segregated 1030 (~11% of the 9447 transcripts differentially

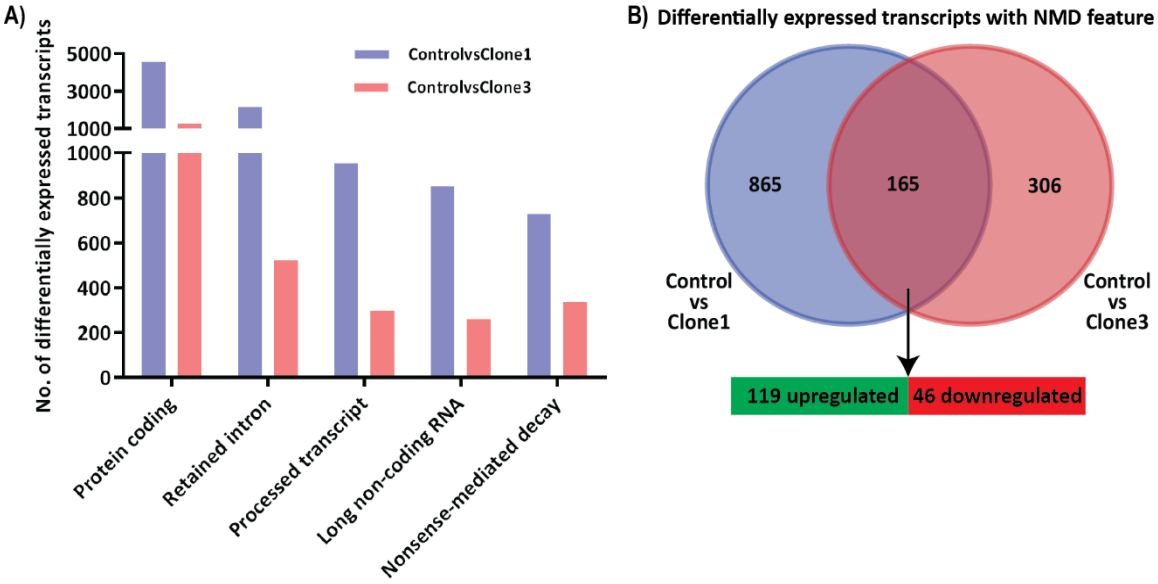


Figure 5.1. NMD transcripts in UPF3B-KO cells. A) A bar graph showing the isoform biotypes of differentially expressed transcripts. B) A Venn diagram showing the Ensembl-annotated NMD transcripts in UPF3B-KO clones.

expressed) and 471 (~17% of the 2725 transcripts differentially expressed) NMD transcripts in Clone1 and Clone3, respectively (Figure 5.1B). These results indicate that NMD transcripts are differentially regulated in both the KO clones (Figure 5.1).

5.3.2 Features of NMD Transcripts

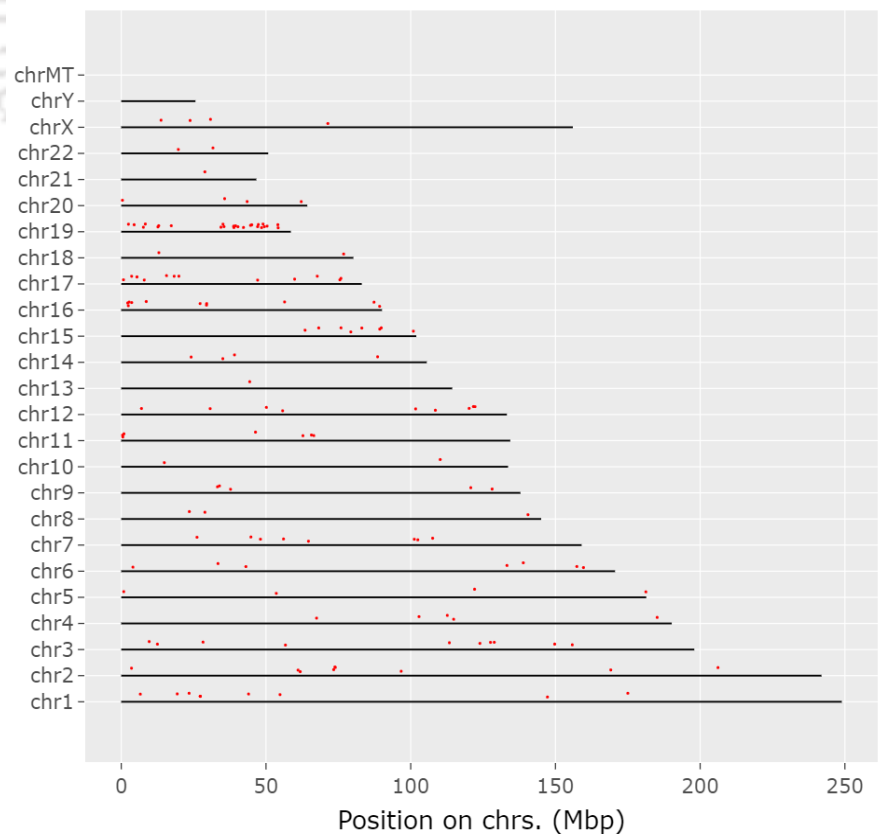


Figure 5.2. Positions of NMD transcripts in the different chromosomes.

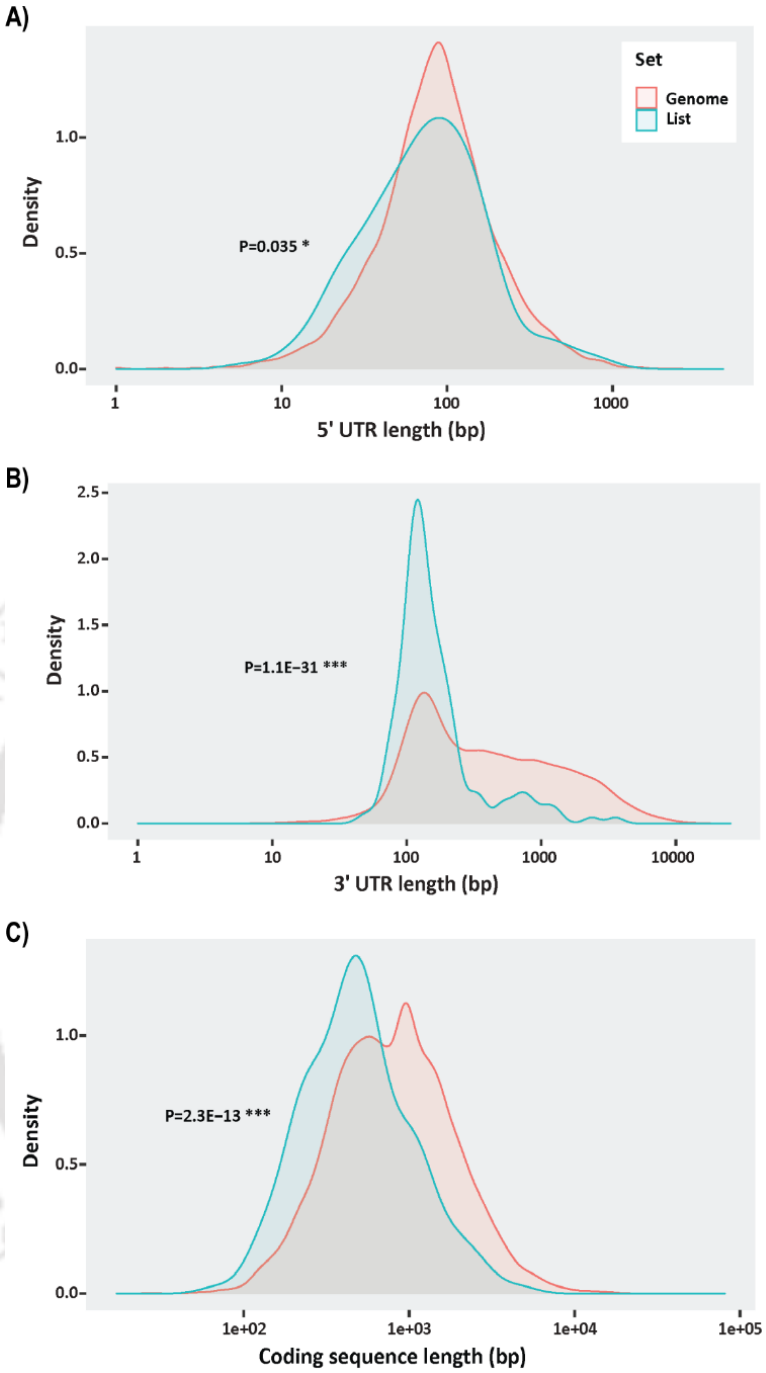


Figure 5.3. Features of NMD transcripts. (A-C) The distribution of 165 NMD transcripts common in KO clones compared to the human genome. The x-axis of graph represents 3'-UTR length (A), 5'-UTR length (B), and coding sequence length (C), whereas y-axis represents the density of transcripts.

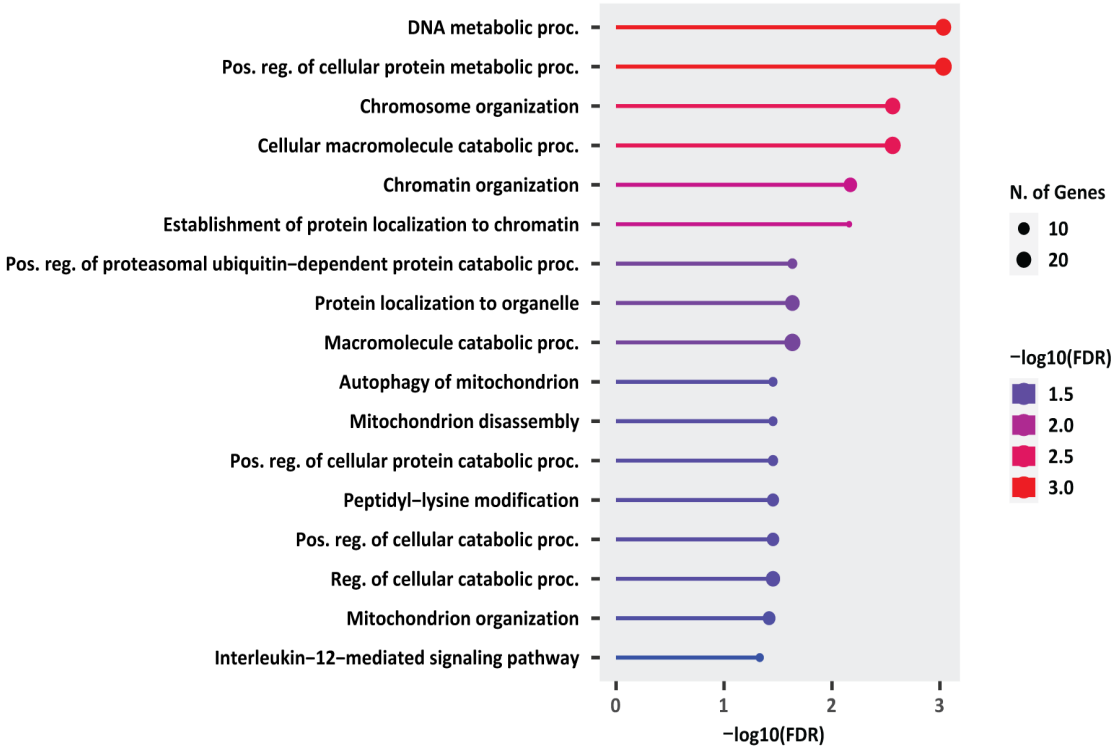


Figure 5.4. Gene ontology analysis for biological processes. Bar dot plot of GO analysis of biological processes of NMD transcripts common in UPF3B-KO clones.

Since both the KO clones are devoid of UPF3B expression, further analysis was conducted focussing on the common transcripts. Remarkably, 165 NMD transcripts were identified as shared between both KO clones (Figure 5.1B). These common NMD transcripts exhibited an even distribution across different chromosomes (Figure 5.2), suggesting a global regulatory influence of UPF3B on NMD transcripts. To characterize the NMD features of these 165 transcripts, the ShinyGO tool was used. Intriguingly, these NMD transcripts demonstrated shorter 5'-UTR and 3'-UTR lengths, along with a reduced coding sequence length (Figure 5.3A-C). This observation suggests that the loss of UPF3B might contribute to the deregulation of transcription, resulting in the production of shorter transcripts. To understand the biological processes associated with these transcripts, a Gene Ontology (GO) analysis was carried out. The majority of the NMD transcripts were linked to DNA metabolic processes, positive regulation of protein metabolic processes, and chromosome organization (Figure 5.4). This suggests that UPF3B might regulate NMD transcripts that impact both nuclear and cytoplasmic processes.

5.3.3 NMD Transcripts are Stabilized in UPF3B-KO Clones

Out of the 165 NMD transcripts, 119 were upregulated, while 46 were downregulated (Figure 5.1B). The observed upregulation of these NMD-sensitive targets in the context of

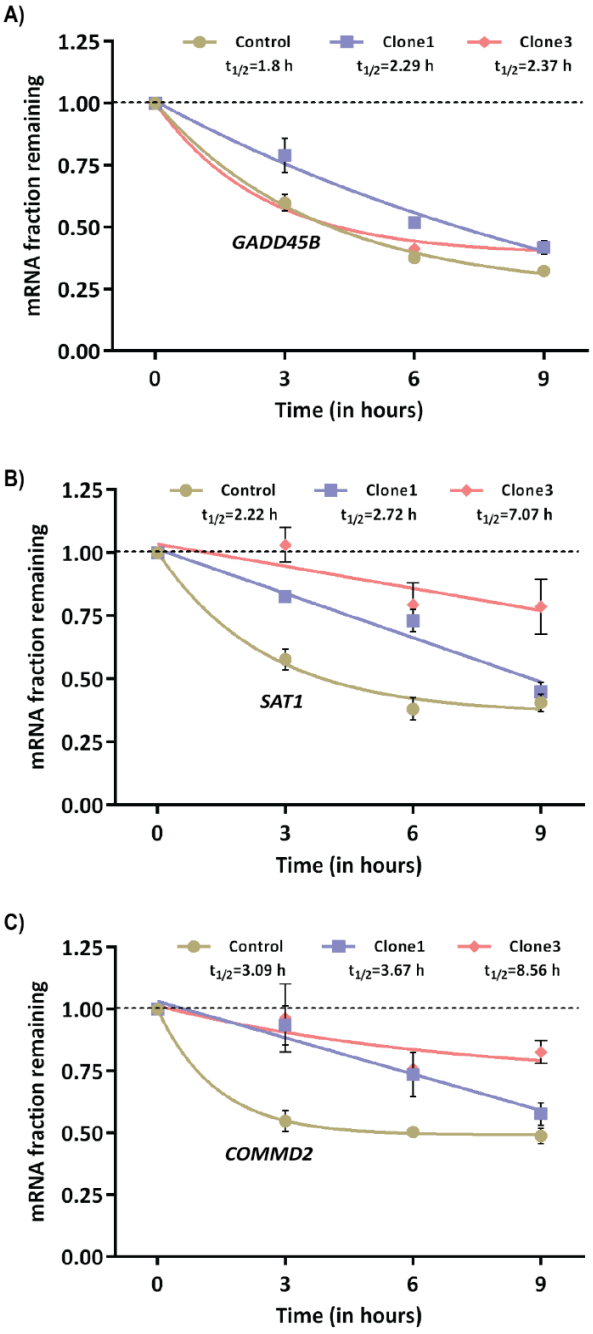


Figure 5.5. Stability of NMD transcripts. Non-linear regression curve showing decay of A) GADD45B B) SAT1, and C) COMMD2 mRNAs in UPF3B-KO clones. For the mRNA stability assays, the average Ct value of transcripts at each time-points was normalized to the average Ct value at t = 0. The relative abundance of each time-point was calculated as mRNA abundance = 2^(-ΔCt). Decay constant (K) was derived from the non-linear regression equation between the natural logarithm and time-points as slope of the equation. Half-lives were calculated using the formula, t1/2 = ln2/K.

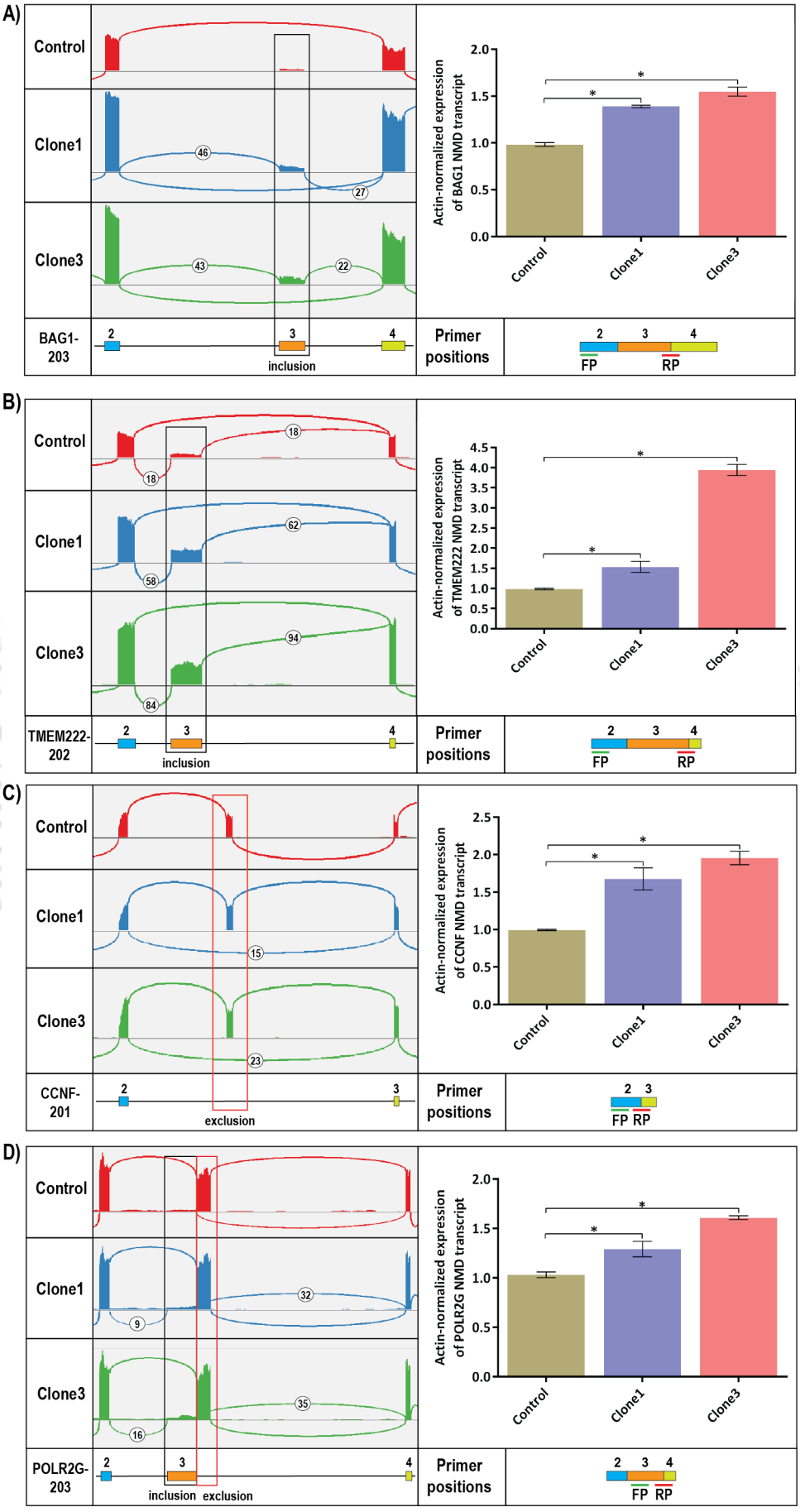


Figure 5.6. Validation of NMD transcripts. Integrated genome viewer visualization of corresponding exons of NMD transcripts, qPCR validation and primer locations of *BAG1*, *TMEM222*, *CCNF*, and *POLR2G*. a) In *BAG1*, inclusion of exon 3 introduces PTC in exon 5, resulting in a NMD-sensitive transcript. b) In case of TMEM222, inclusion of exon 3 introduces a PTC in exon 3 itself. c) In *CCNF*, exclusion of a particular exon results the introduction of PTC in subsequence exon (exon 3) and forms an NMD transcript. d) In *POLR2G*, inclusion of intronic sequences (intron 2) and exclusion of subsequent exon results in a NMD-sensitive transcript. Here, the intronic sequences contains a PTC.
The qPCR analysis of all these NMD-sensitive transcripts revealed significant upregulation in both KO clones.

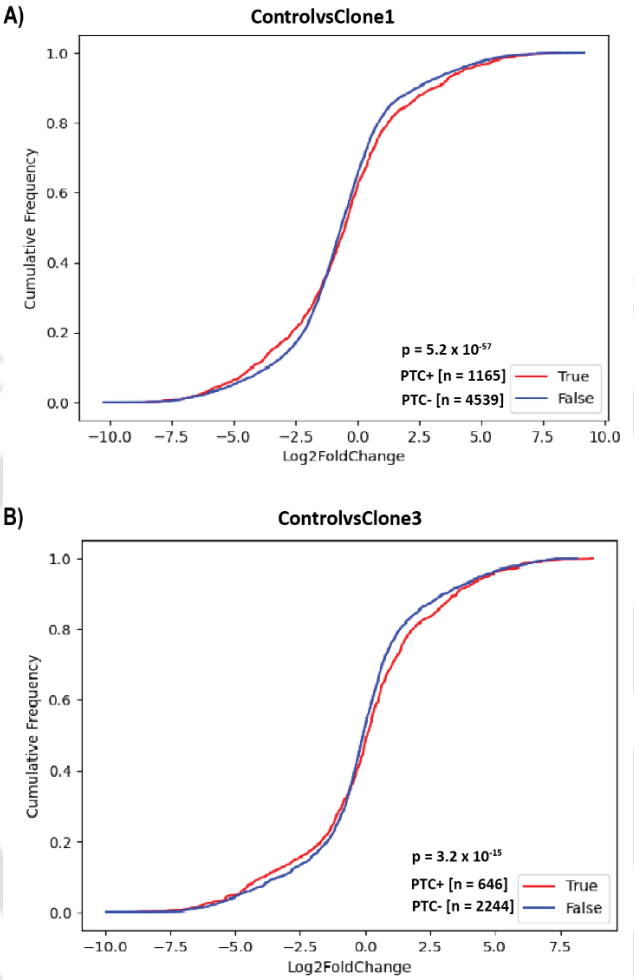


Figure 5.7. Distribution of novel PTC+ transcripts in UPF3B-KO clones. Cumulative Distribution Function (CDF) plots of PTC+ isoforms and PTC− transcripts in Control vs Clone1 (A) and Control vs Clone3 (B). X-axis represents log2FoldChange in UPF3B-KO cells versus Control cells. Number of transcripts in each set (n) and P-value from Kolmogorov–Smirnov (KS) test comparing the two distributions are shown.

impaired NMD may be attributed to the stabilization of these transcripts. Notably, half-life of mRNA stability analysis revealed the upregulation of three noteworthy transcripts: *GADD45B*, *SAT1*, and *COMMD2*. Degradation of *GADD45* mRNA via NMD has been demonstrated to be essential for cell viability [178]. SAT1, encoding Spermidine/spermine N1-acetyltransferase 1, is a rate-limiting enzyme in the polyamine metabolic pathway. The inclusion of an exon in the unproductive splice isoform of *SAT1* introduces a PTC, rendering it susceptible to degradation via NMD [220, 221]. COMMD2 inhibits the NF-KappaB transcription factor, associated with various cellular processes, including cell cycle

regulation [222]. Further investigation into the stability of *GADD45B*, *SAT1*, and *COMMD2* transcripts upon UPF3B KO revealed their significant stabilization (Figure 5.5A-C). As envisaged, the half-life of *GADD45B* was increased by 0.49 h in Clone1 and 0.57 h in Clone3 (Figure 5.5A). Similarly, the half-life of *SAT1* increased by 0.5 h in Clone1 and 4.85 h in Clone3 (Figure 5.5B), and *COMMD2* half-life increased by 0.58 h in Clone1 and 5.47 h in Clone3 (Figure 5.5C). These results strongly indicate the accumulation of NMD transcripts upon UPF3B-KO, potentially exerting deleterious effects on the cellular processes.

The stabilization of NMD transcripts promoted a closer examination of other upregulated NMD-sensitive targets, including *BAG1*, *TMEM222*, *CCNF* and *POLR2G* (Figure 5.6A-D). The inclusion of exon 3 sequences in BAG1 and TMEM222 transcripts introduced a PTC, as illustrated in Figure 5.6A and 5.6B. Similarly, the exclusion of a particular exon in CCNF transcripts resulted in the formation of an NMD-sensitive isoform (Figure 5.6C). For the POLR2G transcript, inclusion of intronic sequences and exclusion of an adjacent exon rendered it NMD-sensitive (Figure 5.6D). The qPCR analysis of these NMD-sensitive transcripts showed a significant upregulation in both KO clones (Figure 5.6A-D). This upregulation of NMD transcripts may be attributed to the alterations in aforementioned biological processes in the context of NMD impairment.

5.3.4 Identification of Novel PTC+ Transcripts

The aforementioned 165 NMD transcripts identified through differential transcript expression (DTE) are already annotated. Therefore, in order to investigate novel NMD-sensitive transcripts, the RNA-Seq data were subjected to analysis using IsoformSwitchAnalyzeR, focusing on identifying isoforms with PTC in both the KO clones. Intriguingly, it was observed that PTC+ transcripts were expressed with significant log2FoldChanges, resulting in a shifted curve to the left or right, compared to PTC− transcripts (Figure 5.7A-B).

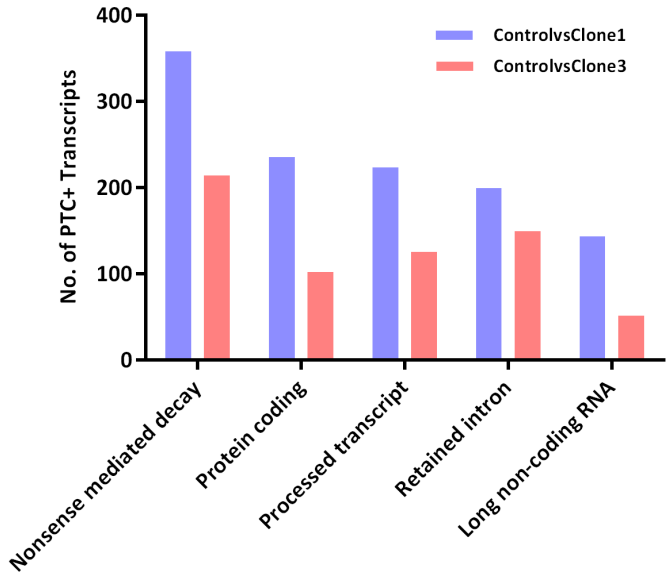


Figure 5.8. Bar graph showing the isoform biotypes of novel PTC+ transcripts.

In Clone1 and Clone3, 1165 and 646 PTC-containing transcripts were identified, respectively (Figure 5.7A-B). To characterize the nature of PTC+ transcripts, further analysis was conducted on their transcript biotypes. A considerable fraction of these transcripts possessed NMD features, followed by protein coding, processed transcripts, retained intron, and lncRNA features (Figure 5.8). Notably, 153 of these PTC-containing transcripts, associated with 69 genes, were common in both KO clones (Figure 5.9). Remarkably, PTC+ transcripts tended to have shorter 3'-UTR lengths (Figure 5.10A), while the coding sequence length remained unaffected (Figure 5.10B). The analysis of PTC+ transcript sequences revealed early termination of transcription leading to shortened transcript length compared to the coding transcripts of same gene. IGV snapshot of two genes, *TMEM248* and *CENPU*, displays shorter 3'-UTR lengths in PTC+ transcripts compared to the protein coding transcripts (Figure 5.11A-B).

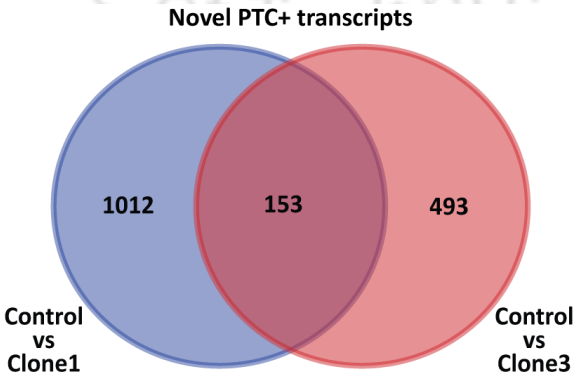


Figure 5.9. A Venn diagram showing the PTC+ transcripts in UPF3B-KO clones.

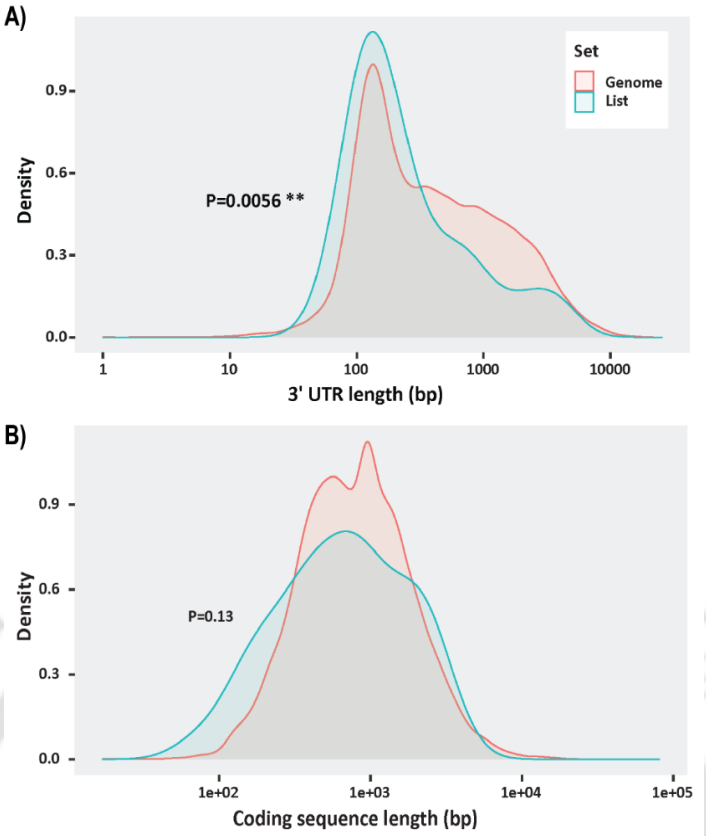


Figure 5.10. Features of novel PTC+ transcripts. (A-B) The distribution of 153 NMD transcripts common in KO clones compared to the human genome. The x-axis of graph represents 3'-UTR length (A) and coding sequence length (B), whereas y-axis represents the density of transcripts.

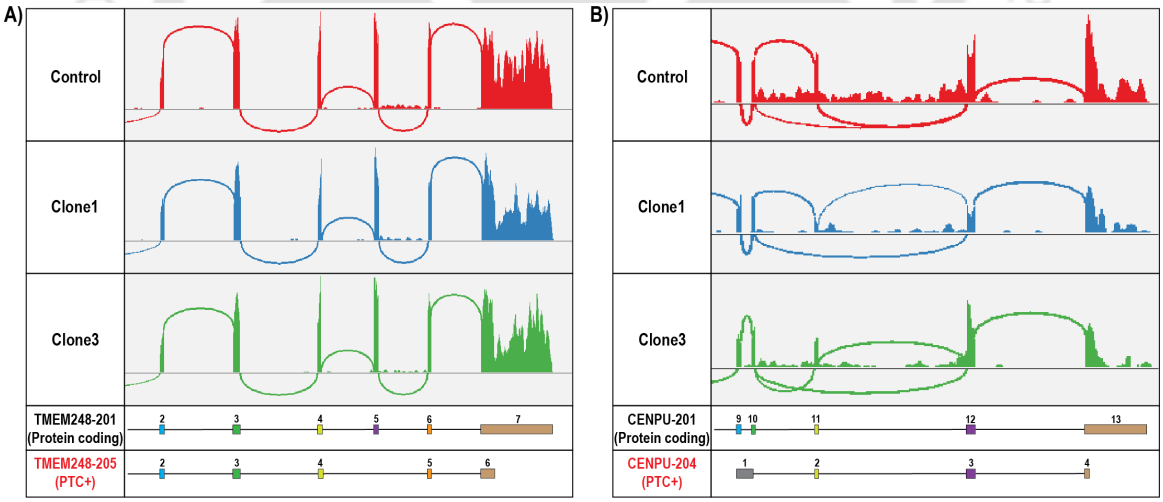


Figure 5.11. IGV visualization of PTC+ transcripts. IGV of corresponding exons of NMD transcripts, TMEM248 and CENPU, with shorter 3′UTR lengths in UPF3B-KO clones.

Altogether, a total of 153 novel PTC+ transcripts were identified in the KO clones, potentially representing substrates specific to UPF3B-dependent NMD.

5.3.5 Increased Usage of NMD-Sensitive Transcripts in KO cells

In the preceding section, both NMD transcripts and novel PTC+ transcripts were characterized by shorter 3'-UTR lengths. Additionally, PTC+ transcripts exhibited diverse isoform biotypes, potentially stemming from alternative splicing events. Consequently, I aimed to investigate the splicing events associated with PTC-containing transcripts.

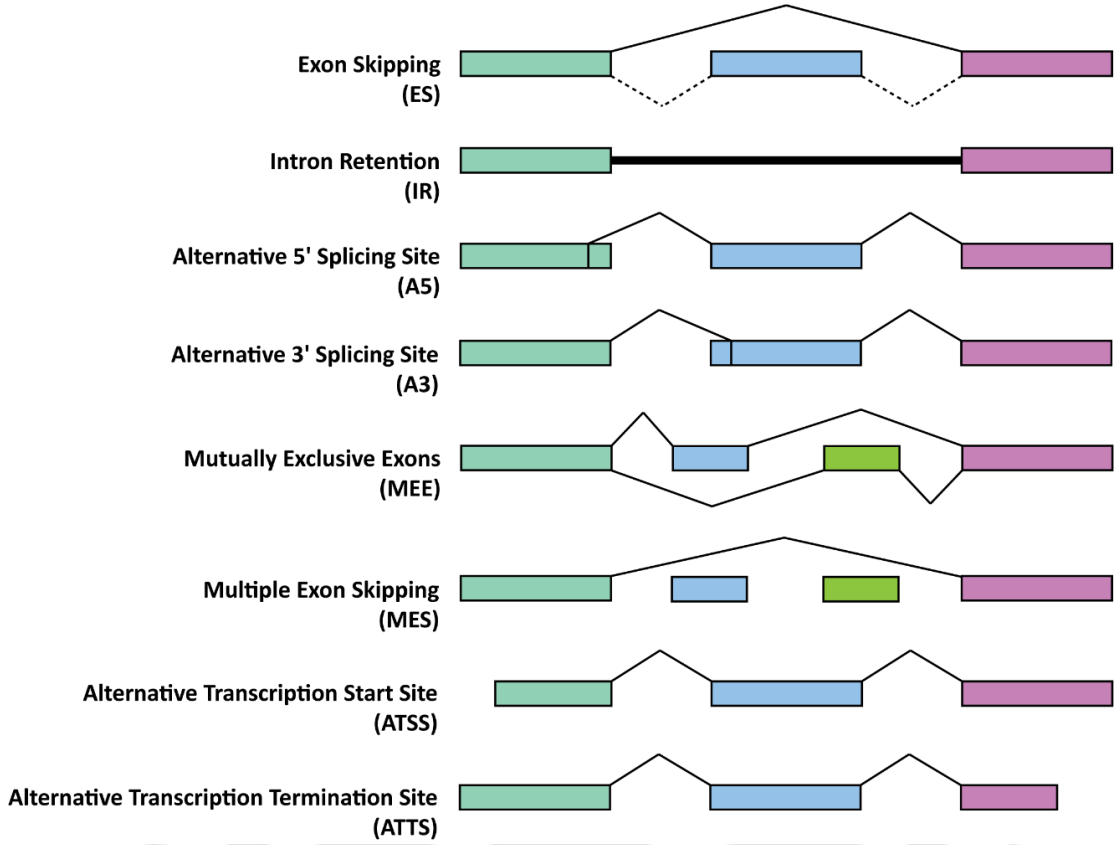


Figure 5.12. Schematic representation of alternative splicing events.

Therefore, I conducted a comprehensive examination of transcriptome-wide alternative splicing events in the KO clones (Figure 5.12). To identify the abundance of isoform switches associated with splicing events, we performed splicing enrichment analysis. The transcript switches from intron retention (IR) gain to IR loss and that of from alternative transcription termination site (ATTS) loss to ATTS gain were significantly enriched in the KO clones (Figure 5.13). Consequently, I aimed to investigate the splicing events associated with PTC-containing transcripts. Interestingly, I found that PTC+ transcripts were frequently linked to multiple splicing events, with a majority of these transcripts (138 in Clone1 and 77 in Clone3) demonstrating alternative transcription start site (ATSS) and alternative transcription termination site (ATTS) events (Figure 5.14).

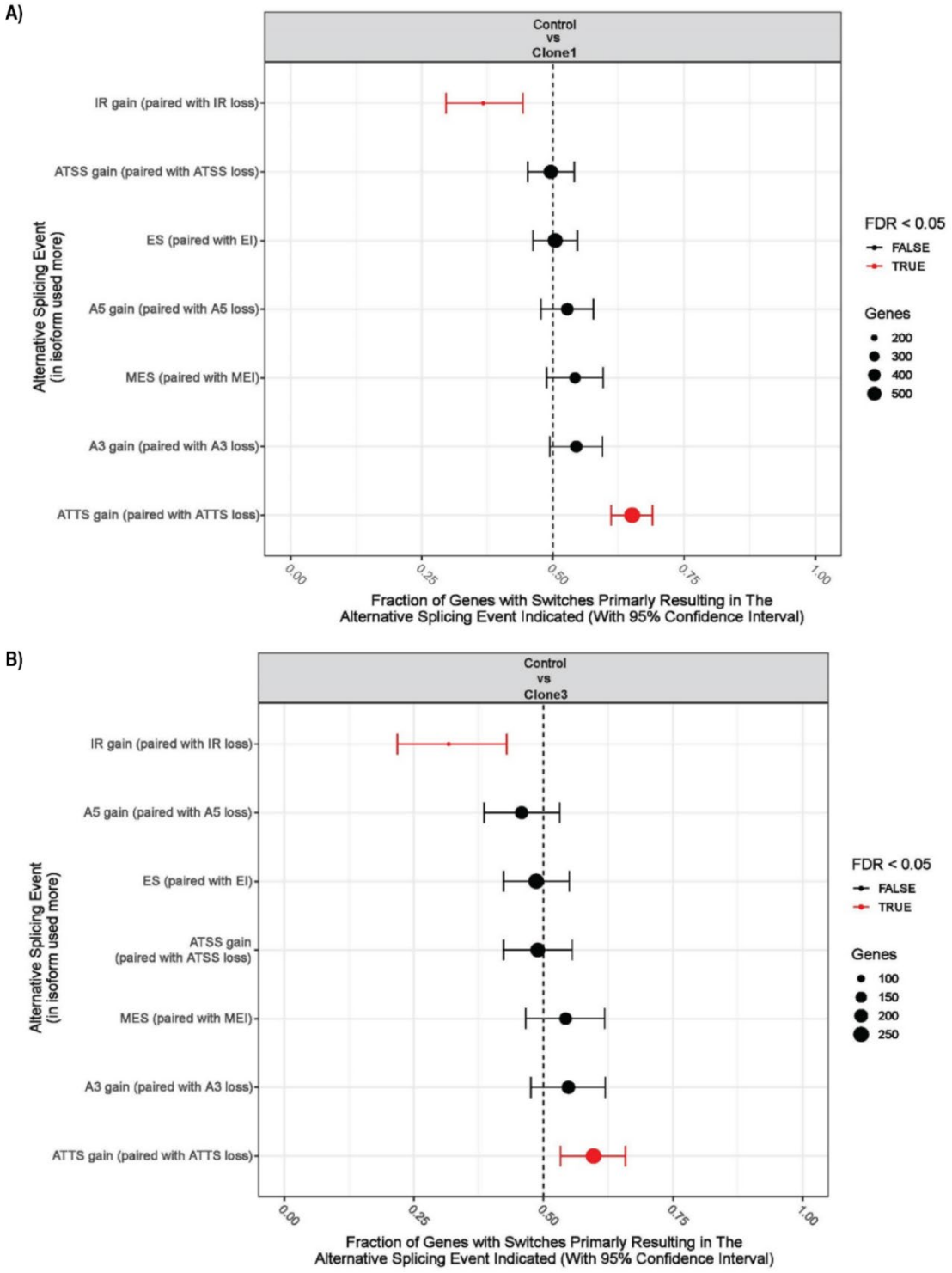


Figure 5.13. Splicing enrichment analysis. Enrichment of alternative splicing events in transcriptome of A) Clone1 and B) Clone3 cells compared to Control cells.

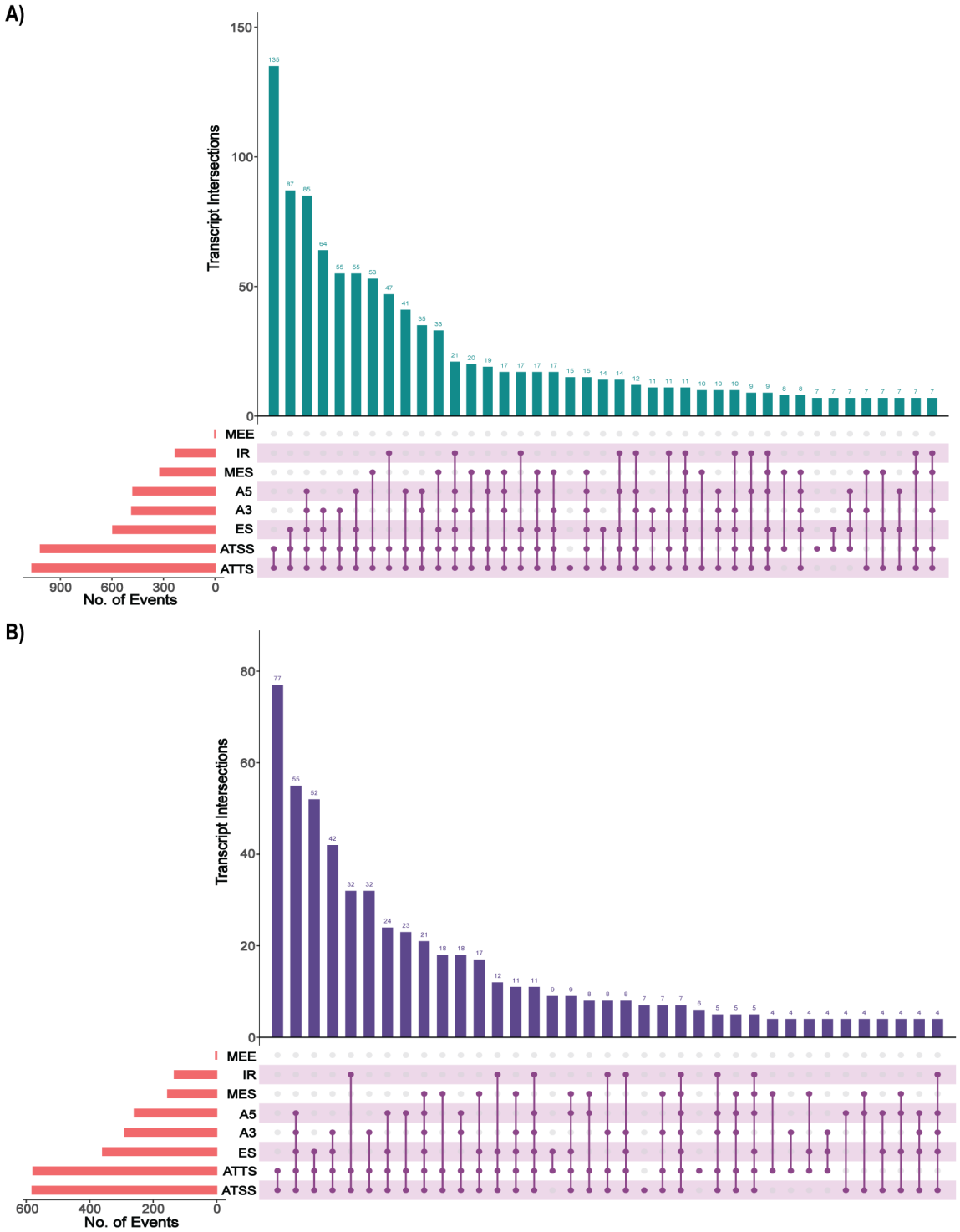


Figure 5.14. PTC+ transcripts associated with alternative splicing events. The UpSet plots of intersections between sets of PTC+ transcripts with alternative splicing events in A) ControlvsClone1 and B) ControlvsClone3. The horizontal bar graph on the left indicates the number of alternative splicing events. The violet dots respresent the intersection points of alternative splicing events and PTC+ transcripts possessed the alternative splicing events. The vertical bar graph indicates the intersection size between PTC+ transcripts associated with one or more alternative splicing event(s). MEE – Mututally Exclusive Exons, IR – Intron Retention, MES – Multiple Exon Skipping, A5 – Alternative 5'-splice Site, A3 – Alternative 3'-splice Site, ES – Exon Skipping, ATTS – Alternative Transcription Termination Site, ATSS – Alternative Transcription Start Site.

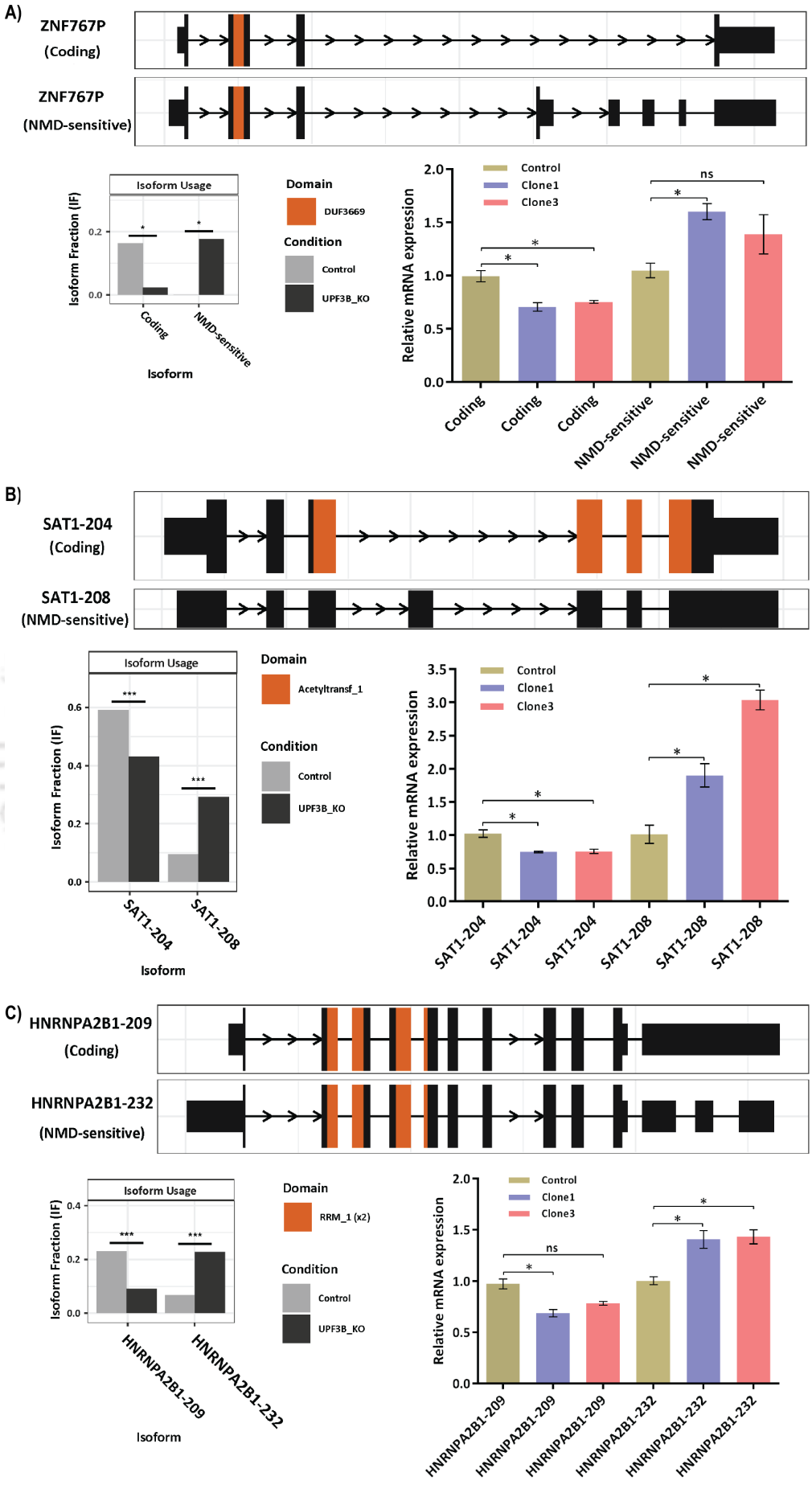


Figure 5.15. Differential transcript usage of NMD-sensitive transcripts. Isoform switching analysis for the c) *ZNF767P* d) *SAT1* and e) *HNRNPA2B1* genes. The exon map for each gene, featuring coding and NMD-sensitive isoforms are depicted. The protein domain information is indicated with orange color. The isoform fractions of coding and NMD-sensitive transcripts in both Control and UPF3B-KO clones are presented below the exon map. To validate the isoform switching analysis, qPCR was performed, and mRNA levels are normalized to the housekeeping Actin transcripts. The values depicted represent the average fold change relative to the Control cells, as determined by Unpaired Student's t-test (n=3). Error bars indicate the standard error of mean, with 'ns' indicating not significant, and '*' denoting a P-value < 0.05.

To identify which transcript features were more commonly utilized in the KO clones, I performed isoform switching analysis, revealing a total of 185 genes with predicted switch consequences in Clone1 and 35 genes in Clone3. Among these genes with switch consequences, several switched to NMD-sensitive transcripts in the KO clones. For instance, *ZNF767P*, a pseudogene with an uncharacterized function, exhibited a switch to an NMD-sensitive isoform in the KO clones. This NMD-sensitive splice isoform of *ZNF767P* was upregulated by approximately 1.6-fold (p-value 0.006) in Clone1 and about 1.39-fold in Clone3 (Figure 5.15A). In the case of the *SAT1* transcript, the inclusion of an exon introduced a PTC, resulting in an NMD-sensitive transcript. This PTC+ transcript was upregulated by approximately 1.9-fold (p-value 0.0088) in Clone1 and roughly 3.0-fold (p-value 0.0002) in Clone3 (Figure 5.15B). Furthermore, *HNRNPA2B1* transcripts are known to undergo autoregulation through alternative splicing coupled with NMD [223]. Alternative splicing events within the 3'-UTR of the *HNRNPA2B1* gene generate an NMD-sensitive splice isoform, which was upregulated by approximately 1.4-fold (p-value 0.0168) in Clone1 and about 1.43-fold (p-value 0.0055) in Clone3 (Figure 5.15C).

In summary, the isoform switching analysis provides further evidence of the crucial role played by UPF3B in alternative splicing-mediated NMD. These results underscore the importance of UPF3B in the degradation of NMD-sensitive transcripts, thereby mitigating their potential detrimental effects within cells.

5.4 Discussion

UPF3B-KO HEK-293 cells showed a marked reduction in NMD activity, which points to the deregulation of NMD-sensitive transcripts upon loss of UPF3B. The differential transcript expression analysis of RNA-Seq data identified 1030 and 471 significantly expressed NMD transcripts in Clone1 and Clone3, respectively, with notable fold changes (log2FoldChange > 1.2). Recently, Yi *et al.* carried out RNA-Seq on UPF3B-KO HCT-116 cells, revealing the overexpression of 819 UPF3B-dependent NMD targets at the transcript level [149]. To eliminate clonal variation, 165 NMD transcripts shared by the two UPF3B-KO clones were examined. Dysregulation was observed in NMD transcripts encoding transcription factors such as POLR2G, ZNF236, HMBOX1, TAF1, POLD1, and POLG. The variable expression of NMD transcripts may alter the ability of the encoded proteins to bind DNA, affecting DNA metabolic processes. However, in-depth mechanistic studies are required to gain a better understanding of these processes.

In another approach to identify the NMD-sensitive transcripts, Yi *et al.* focused on transcripts with PTC located ≥50 nucleotides upstream of an exon-exon junction [149]. They identified 3492 PTC+ transcript in UPF3B-KO HCT-116 cells. Simultaneously, Wallmeroth *et al.* utilized the 50-nucleotide rule of NMD to identify PTC+ transcripts in UPF3B-KO HEK-293 cells and identified a total of 402 PTC+ transcripts [150]. Following a similar approach, this study identified 1165 and 646 PTC+ transcripts in Clone1 and Clone3, respectively. Among these PTC+ transcripts, 153 transcripts were common in both UPF3B-KO clones, characterized by the presence of shorter 3'UTRs. Given that long 3'-UTRs are typically associated with NMD induction, the presence of these shorter 3'-UTRs suggests a distinct set of transcripts regulated by UPF3B-mediated NMD. This observation not only implies that UPF3B governs a unique group of transcripts but also hints at a potential role in transcription termination, warranting further investigation.

Given the upregulation of NMD and PTC+ transcripts in KO clones, an analysis of isoform switching was conducted on RNA-Seq data. Isoform switching analysis revealed that several genes, including *ZNF767P*, *SAT1*, and *HNRNPA2B1*, switched to their NMD-sensitive isoforms in absence of UPF3B. The enhanced usage of NMD-sensitive splice isoforms in UPF3B knockouts suggests that UPF3B plays a role in alternative splicing, suppressing the formation of mis-spliced transcripts via alternative splicing coupled with NMD. In summary, UPF3B serves as a key player in maintaining cellular homeostasis by

facilitating the degradation of both physiological and NMD-sensitive transcripts. Further research is needed to fully elucidate the intricate mechanisms underlying these processes.

5.5 Conclusion

Differential transcript expression analysis identified 165 NMD transcripts and 153 novel PTC+ transcripts in UPF3B-KO clones. These NMD and PTC+ transcripts possessed shorter 3′-UTR length compared to genome, indicating an unprecedented characteristic feature of transcripts regulated by UPF3B-mediated NMD. The majority of these transcripts are generated through the alternative splicing events by utilizing the alternative transcription start and termination sites. The increased usage of NMD-sensitive transcripts associated with the alternative splicing events suggests the regulatory function of UPF3B via alternative splicing-coupled NMD.

Chapter 6:
Summary and Future Perspectives

Chapter 6

6.1 Summary

The outcomes of this thesis work clearly demonstrate that UPF3B is integral to nonsense-mediated mRNA decay (NMD), as knockout of UPF3B leads to the upregulation of canonical NMD substrates (like *GAS5*, *ITGAE*, *SRSF2*, *ZFAS1*), and NMD reporter, such as PTC39 β-globin transcripts. These findings corroborate with the previous study where siRNA-mediated knockdown of UPF3B upregulated PTC39 β-globin NMD reporter and a set of endogenous NMD transcripts [84]. In my study, the absolute loss of UPF3B has a subtle effect on the expression of NMD-associated factors. For example, UPF3A, the paralog of UPF3B, is stabilized in the absence of UPF3B. However, stabilized UPF3A was unable to restore NMD activity. Also, the upregulation of UPF1, SMG5, SMG6 and SMG7 at both transcript and protein levels in UPF3B-knockouts (KOs) suggests that UPF3B deletion triggers the activation of an autoregulatory circuit within the NMD pathway, as previously reported [196].

Differential gene expression (DGE) analysis identified a subset of NMD-sensitive genes was significantly upregulated despite the stabilization of UPF3A and upregulation of NMD-associated factors, UPF1, SMG5, SMG6 and SMG7, indicating that the substrates are regulated by UPF3B and not by NMD-associated factors. Among the upregulated differentially expressed genes (DEGs), the increased mRNA levels of two NMD targets, *GADD45G* and *SELENOP*, were particularly notable. The upregulation of *GADD45G* transcripts in NMD-deficient cells inhibits cell proliferation [190, 224], as these transcripts are typically targeted for degradation by the NMD machinery to maintain proper cell viability [178]. Expression profiling of *GADD45G* in *Xenopus* and medaka fish demonstrated that *GADD45G* promotes neuronal differentiation by inhibiting cell proliferation [190, 224]. *SELENOP* transcripts are degraded via NMD when the in-frame UGA codon, which also encodes for selenocysteine, is recognized as a stop codon. SELENOP is essential for neuronal survival and is linked to neurological dysfunction [191, 192]. Exogenous expression of UPF3B in knockout cells completely rescued the expression of *GADD45G* transcripts and partially rescued *SELENOP* transcripts, indicating that UPF3B-mediated NMD regulates the physiological levels of these genes and underscores critical role of UPF3B in neuronal functions.

Additionally, KEGG pathway analysis of DEGs revealed significant downregulation of genes associated with axon guidance pathway. The DEGs, *SEMA3D*,

SEMA6D, *PLXNC1*, *ARHGEF12*, and *ITGB1*, are involved in semaphorin interactions. SEMA3D is a member of class III semaphorins which are secreted signaling proteins, whereas SEMA6D, the class VI semaphorin, is a transmembrane signaling protein [225]. Semaphorins interact with plexin receptors to induce semaphorin signaling. The plexin receptor, Plexin C1 (PLXNC1), interacts with semaphorins via its sema domain to execute the signaling cascade. Semaphorin-plexin signaling regulates neural circuit development, including axon growth, axon bundling, and pruning [226]. Other DEGs, *NTN4*, *PTPN11*, *TRPC4*, and *ROBO1*, are components of the Netrin signaling pathway. Netrins are secreted signaling molecules that attract axons to the midline during brain commissure formation [227]. Semaphorins and netrins are two conserved families of axon guidance molecules, and changes in their expression affect neural circuits, leading to neurological disorders. Another important neuron-specific gene, *ARC*, is a master regulator of synaptic plasticity in neurons. Depletion of UPF3B leads to the stabilization of *ARC* mRNA by UPF3A [80]. The upregulation of ARC mRNA in UPF3B-KO cells underscores the role of UPF3B in regulating a subset of mRNAs involved in neuronal functions.

Differential transcript expression (DTE) analysis identified the upregulation and stabilization of 165 NMD-sensitive transcripts in UPF3B-KO cells. These transcripts are rapidly degraded in normal cells, and their increased expression in diseased condition alters cellular functions [228]. Additionally, the PTC (premature termination codon) status of transcripts was determined using the 50-nucleotide rule, resulting in the identification of 153 novel PTC+ transcripts in UPF3B-KO clones. Both the 165 NMD-sensitive and 165 PTC+ transcripts were characterized by short 3′-untranslated regions (3′-UTRs). Further characterization of these transcripts revealed that alternatively spliced transcripts with PTCs are more prevalent in UPF3B-KO cells. Isoform switching analysis showed that many genes, including *ZNF767P* and *HNRNPA2B1*, have increased usage of NMD-sensitive transcripts compared to the coding transcripts in UPF3B-KO cells. *ZNF767P*, despite being a pseudogene, has been identified as a potential biomarker for glioblastoma [229]. HNRNPA2B1 is a splicing factor and is closely associated with nuclear RNA processing [230]. In neurons, animal models and human Alzheimer brain tissues, HNRNPA2B1 proved to associate with endogenous oligomeric Tau (oTau) protein, where it functions as a linker that connects oTau with N6-methyladenosine (m6A) modified RNA transcripts [231]. The increased prevalence of NMD-sensitive splice isoforms of genes with

neuronal importance in UPF3B-KO cells suggests that UPF3B plays a significant role in the degradation of misspliced transcripts via alternative splicing coupled with NMD.

6.2 Future Perspectives

Surprisingly, DGE analysis of UPF3B-KO clones revealed a higher number of downregulated genes compared to upregulated ones. Gene Ontology analysis indicated that these downregulated genes are associated with regulation of transcription from RNA Polymerase II promoter. Future research should focus on elucidating the role of downregulated genes in RNA Polymerase II promoter regulation. This raises the question about UPF3B's involvement in transcription, and its influence on gene expression patterns. Preliminary experiments aimed at understanding UPF3B's involvement in transcription showed an upregulation of the transcription termination factor, TTF1, in UPF3B-KO cells. However, co-immunoprecipitation (co-IP) assays did not detect a direct interaction between TTF1 and UPF3B. Interestingly, co-IP experiments identified a weak interaction between UPF3B and U2AF1, a protein involved in both transcription and splicing. This suggests a potential role for UPF3B in regulating transcription termination and alternative splicing events. Detailed investigations are necessary to further elucidate UPF3B's function in these nuclear processes.

Using UPF3B-KO clones as a model, researchers can explore fundamental cellular processes such as transcription, splicing, translation, and the cell cycle in human neuronal cells. Additionally, KEGG pathway analysis of DEGs revealed the downregulation of genes involved in the axon guidance pathway, which integrates signaling pathways to mediate axon guidance. To study UPF3B's role in the axon guidance pathway, UPF3B-KO HEK-293 cells could be co-cultured with SH-SY5Y cells, similar to previous studies that used mixed cultures of HEK-293 cells and rat hippocampal neurons to investigate neuronal synapse formation [232]. This approach may uncover the molecular mechanisms underlying the progression of neurodevelopmental disorders resulting from the functional loss of UPF3B.

References

1. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **227**(5258): p. 561-563.
2. Zhao, B.S., I.A. Roundtree, and C. He, *Post-transcriptional gene regulation by mRNA modifications.* Nature Reviews Molecular Cell Biology, 2017. **18**(1): p. 31-42.
3. Shoemaker, C.J. and R. Green, *Translation drives mRNA quality control.* Nat Struct Mol Biol, 2012. **19**(6): p. 594-601.
4. Houseley, J. and D. Tollervey, *The many pathways of RNA degradation.* Cell, 2009. **136**(4): p. 763-76.
5. Jacobson, A. and S.W. Peltz, *Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells.* Annu Rev Biochem, 1996. **65**: p. 693-739.
6. Doma, M.K. and R. Parker, *Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.* Nature, 2006. **440**(7083): p. 561-564.
7. Harigaya, Y. and R. Parker, *No-go decay: a quality control mechanism for RNA in translation.* WIREs RNA, 2010. **1**(1): p. 132-141.
8. Frischmeyer, P.A., et al., *An mRNA surveillance mechanism that eliminates transcripts lacking termination codons.* Science, 2002. **295**(5563): p. 2258-61.
9. van Hoof, A., et al., *Exosome-Mediated Recognition and Degradation of mRNAs Lacking a Termination Codon.* Science, 2002. **295**(5563): p. 2262-2264.
10. Maquat, L.E., *Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics.* Nature Reviews Molecular Cell Biology, 2004. **5**(2): p. 89-99.
11. Sun, L., J. Mailliot, and C. Schaffitzel *Nonsense-Mediated mRNA Decay Factor Functions in Human Health and Disease*. Biomedicines, 2023. **11**, DOI: 10.3390/biomedicines11030722.
12. Frischmeyer, P.A. and H.C. Dietz, *Nonsense-Mediated mRNA Decay in Health and Disease.* Human Molecular Genetics, 1999. **8**(10): p. 1893-1900.
13. Oliveira, C.C. and J.E. McCarthy, *The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae.* J Biol Chem, 1995. **270**(15): p. 8936-43.

14. Mendell, J.T., et al., *Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise.* Nature Genetics, 2004. **36**(10): p. 1073-1078.
15. Eberle, A.B., et al., *Posttranscriptional Gene Regulation by Spatial Rearrangement of the 3′ Untranslated Region.* PLOS Biology, 2008. **6**(4): p. e92.
16. Bühler, M., et al., *EJC-independent degradation of nonsense immunoglobulin-μ mRNA depends on 3′ UTR length.* Nature Structural & Molecular Biology, 2006. **13**(5): p. 462-464.
17. Bicknell, A.A., et al., *Introns in UTRs: Why we should stop ignoring them.* BioEssays, 2012. **34**(12): p. 1025-1034.
18. McGlincy, N.J. and C.W.J. Smith, *Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?* Trends in Biochemical Sciences, 2008. **33**(8): p. 385-393.
19. Saltzman, A.L., et al., *Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay.* Molecular and Cellular Biology, 2008. **28**(13): p. 4320-4330.
20. Toma, K.G., et al., *Identification of elements in human long 3′ UTRs that inhibit nonsense-mediated decay.* RNA, 2015. **21**(5): p. 887-897.
21. Withers, J.B. and K.L. Beemon, *The structure and function of the rous sarcoma virus RNA stability element.* Journal of Cellular Biochemistry, 2011. **112**(11): p. 3085-3092.
22. Lou, C.-H., et al., *Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate.* Stem Cell Reports, 2016. **6**(6): p. 844-857.
23. Jones, S.H. and M. Wilkinson, *RNA decay, evolution, and the testis.* RNA Biology, 2017. **14**(2): p. 146-155.
24. Nasif, S., L. Contu, and O. Mühlemann, *Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression.* Seminars in Cell & Developmental Biology, 2018. **75**: p. 78-87.
25. Kurosaki, T., M.W. Popp, and L.E. Maquat, *Quality and quantity control of gene expression by nonsense-mediated mRNA decay.* Nature Reviews Molecular Cell Biology, 2019. **20**(7): p. 406-420.
26. Jaffrey, S.R. and M.F. Wilkinson, *Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease.* Nature Reviews Neuroscience, 2018. **19**(12): p. 715-728.

27. Pawlicka, K., U. Kalathiya, and J. Alfaro *Nonsense-Mediated mRNA Decay: Pathologies and the Potential for Novel Therapeutics*. Cancers, 2020. **12**, DOI: 10.3390/cancers12030765.
28. Popp, M.W. and L.E. Maquat, *Nonsense-mediated mRNA Decay and Cancer.* Current Opinion in Genetics & Development, 2018. **48**: p. 44-50.
29. Supek, F., B. Lehner, and R.G.H. Lindeboom, *To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases.* Trends in Genetics, 2021. **37**(7): p. 657-668.
30. Le Hir, H., et al., *The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions.* EMBO Journal, 2000. **19**(24): p. 6860-6869.
31. Singh, G., et al., *The Cellular EJC Interactome Reveals Higher-Order mRNP Structure and an EJC-SR Protein Nexus.* Cell, 2012. **151**(4): p. 750-764.
32. Hauer, C., et al., *Exon Junction Complexes Show a Distributional Bias toward Alternatively Spliced mRNAs and against mRNAs Coding for Ribosomal Proteins.* Cell Reports, 2016. **16**(6): p. 1588-1603.
33. Tange, T.Ø., et al., *Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core.* RNA, 2005. **11**(12): p. 1869-1883.
34. Ballut, L., et al., *The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity.* Nature Structural & Molecular Biology, 2005. **12**(10): p. 861-869.
35. Schlautmann, L.P. and N.H. Gehring *A Day in the Life of the Exon Junction Complex*. Biomolecules, 2020. **10**, DOI: 10.3390/biom10060866.
36. Hir, H.L., J. Saulière, and Z. Wang, *The exon junction complex as a node of post-transcriptional networks.* Nature Reviews Molecular Cell Biology, 2016. **17**(1): p. 41-54.
37. Deka, B. and K.K. Singh, *Multifaceted Regulation of Gene Expression by the Apoptosis- and Splicing-Associated Protein Complex and Its Components.* International Journal of Biological Sciences, 2017. **13**(5): p. 545-560.
38. Le Hir, H., et al., *The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay.* EMBO Journal, 2001. **20**(17): p. 4987-4997.

39. Buchwald, G., et al., *Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex.* Proceedings of the National Academy of Sciences, 2010. **107**(22): p. 10050-10055.
40. Culbertson, M.R., K.M. Underbrink, and G.R. Fink, *Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors.* Genetics, 1980. **95**(4): p. 833-53.
41. Leeds, P., et al., *Gene products that promote mRNA turnover in Saccharomyces cerevisiae.* Mol Cell Biol, 1992. **12**(5): p. 2165-77.
42. Keesey, J.K., R. Bigelis, and G.R. Fink, *The product of the his4 gene cluster in Saccharomyces cerevisiae. A trifunctional polypeptide.* Journal of Biological Chemistry, 1979. **254**(15): p. 7427-7433.
43. Bigelis, R., J.K. Keesey, and G.R. Fink, *The yeast his4 multifunctional protein. Immunochemistry of the wild type protein and altered forms.* J Biol Chem, 1981. **256**(10): p. 5144-52.
44. Dinman, J.D. and R.B. Wickner, *Translational maintenance of frame: mutants of Saccharomyces cerevisiae with altered -1 ribosomal frameshifting efficiencies.* Genetics, 1994. **136**(1): p. 75-86.
45. Lee, S.I., J.G. Umen, and H.E. Varmus, *A genetic screen identifies cellular factors involved in retroviral -1 frameshifting.* Proc Natl Acad Sci U S A, 1995. **92**(14): p. 6587-91.
46. Pinto, I., et al., *cis- and trans-acting suppressors of a translation initiation defect at the cyc1 locus of Saccharomyces cerevisiae.* Genetics, 1992. **132**(1): p. 97-112.
47. Cui, Y., et al., *Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon.* Genes Dev, 1995. **9**(4): p. 423-36.
48. He, F. and A. Jacobson, *Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen.* Genes Dev, 1995. **9**(4): p. 437-54.
49. Lee, B.S. and M.R. Culbertson, *Identification of an additional gene required for eukaryotic nonsense mRNA turnover.* Proc Natl Acad Sci U S A, 1995. **92**(22): p. 10354-8.
50. Pulak, R. and P. Anderson, *mRNA surveillance by the Caenorhabditis elegans smg genes.* Genes Dev, 1993. **7**(10): p. 1885-97.

51. Perlick, H.A., et al., *Mammalian orthologues of a yeast regulator of nonsense transcript stability.* Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10928-32.
52. Mendell, J.T., et al., *Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes.* Molecular and Cellular Biology, 2000. **20**(23): p. 8944-8957.
53. Serin, G., et al., *Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4).* Molecular and Cellular Biology, 2001. **21**(1): p. 209-223.
54. Lykke-Andersen, J., M.-D. Shu, and J.A. Steitz, *Human Upf Proteins Target an mRNA for Nonsense-Mediated Decay When Bound Downstream of a Termination Codon.* Cell, 2000. **103**(7): p. 1121-1131.
55. Han, X., et al., *Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology.* Nucleic Acids Research, 2018. **46**(3): p. 1038-1051.
56. Alrahbeni, T., et al., *Full UPF3B function is critical for neuronal differentiation of neural stem cells.* Molecular Brain, 2015. **8**.
57. Cheng, Z., et al., *Structural and functional insights into the human Upf1 helicase core.* Embo j, 2007. **26**(1): p. 253-64.
58. Czaplinski, K., et al., *Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation.* Rna, 1995. **1**(6): p. 610-23.
59. Applequist, S.E., et al., *Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein.* Nucleic Acids Res, 1997. **25**(4): p. 814-21.
60. Weng, Y., K. Czaplinski, and S.W. Peltz, *Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein.* Mol Cell Biol, 1996. **16**(10): p. 5477-90.
61. He, F., et al., *Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant.* Proc Natl Acad Sci U S A, 1993. **90**(15): p. 7034-8.
62. Atkin, A.L., et al., *The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm.* Mol Biol Cell, 1995. **6**(5): p. 611-25.
63. Bhattacharya, A., et al., *Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay.* Rna, 2000. **6**(9): p. 1226-35.
64. Kunz, J.B., et al., *Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation.* RNA, 2006. **12**(6): p. 1015-1022.

65. Gehring, N.H., et al., *Y14 and hUpf3b Form an NMD-Activating Complex.* Molecular Cell, 2003. **11**(4): p. 939-949.

66. Karousis, E.D., S. Nasif, and O. Mühlemann, *Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact.* WIREs RNA, 2016. **7**(5): p. 661-682.

67. Kadlec, J., E. Izaurralde, and S. Cusack, *The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3.* Nature Structural and Molecular Biology, 2004. **11**(4): p. 330-337.

68. Neu-Yilik, G., et al., *Dual function of UPF3B in early and late translation termination.* The EMBO Journal, 2017. **36**(20): p. 2968-2986.

69. Kashima, I., et al., *Binding of a novel SMG-1–Upf1–eRF1–eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay.* Genes & Development, 2006. **20**(3): p. 355-367.

70. Kadlec, J., et al., *Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1.* RNA, 2006. **12**(10): p. 1817-1824.

71. Ajamian, L., et al. *HIV-1 Recruits UPF1 but Excludes UPF2 to Promote Nucleocytoplasmic Export of the Genomic RNA*. Biomolecules, 2015. **5**, 2808-2839 DOI: 10.3390/biom5042808.

72. Clerici, M., et al., *Structural and functional analysis of the three MIF4G domains of nonsense-mediated decay factor UPF2.* Nucleic Acids Research, 2014. **42**(4): p. 2673-2686.

73. He, F., A.H. Brown, and A. Jacobson, *Interaction between Nmd2p and Upf1p is required for activity but not for dominant-negative inhibition of the nonsense-mediated mRNA decay pathway in yeast.* Rna, 1996. **2**(2): p. 153-70.

74. He, F., A.H. Brown, and A. Jacobson, *Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway.* Mol Cell Biol, 1997. **17**(3): p. 1580-94.

75. Gehring, N.H., et al., *Exon-Junction Complex Components Specify Distinct Routes of Nonsense-Mediated mRNA Decay with Differential Cofactor Requirements.* Molecular Cell, 2005. **20**(1): p. 65-75.

76. Kim, V.N., N. Kataoka, and G. Dreyfuss, *Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex.* Science, 2001. **293**(5536): p. 1832-1836.

77. Shirley, R.L., et al., *A factor required for nonsense-mediated mRNA decay in yeast is exported from the nucleus to the cytoplasm by a nuclear export signal sequence.* Journal of Cell Science, 1998. **111**(21): p. 3129-3143.

78. Shirley, R.L., et al., *Nuclear import of Upf3p is mediated by importin-alpha/-beta and export to the cytoplasm is required for a functional nonsense-mediated mRNA decay pathway in yeast.* Genetics, 2002. **161**(4): p. 1465-82.

79. Shirley, R.L., M.R. Richards, and M.R. Culbertson, *Using the cre-lox recombination system to assess functional impairment caused by amino acid substitutions in yeast proteins.* Biological Procedures Online, 2004. **6**(1): p. 209-219.

80. Shum, Eleen Y., et al., *Nonsense-Mediated RNA Decay.* Cell, 2016. **165**(2): p. 382-395.

81. Chamieh, H., et al., *NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity.* Nature Structural & Molecular Biology, 2008. **15**(1): p. 85-93.

82. Bufton, J.C., et al., *Structures of nonsense-mediated mRNA decay factors UPF3B and UPF3A in complex with UPF2 reveal molecular basis for competitive binding and for neurodevelopmental disorder-causing mutation.* Nucleic Acids Research, 2022. **50**(10): p. 5934-5947.

83. Metze, S., et al., *Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways.* RNA, 2013. **19**(10): p. 1432-1448.

84. Chan, W.-K., et al., *An alternative branch of the nonsense-mediated decay pathway.* The EMBO Journal, 2007. **26**(7): p. 1820-1830.

85. Ivanov, P.V., et al., *Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways.* The EMBO Journal, 2008. **27**(5): p. 736-747.

86. Gehring, N.H., et al., *The Hierarchy of Exon-Junction Complex Assembly by the Spliceosome Explains Key Features of Mammalian Nonsense-Mediated mRNA Decay.* PLOS Biology, 2009. **7**(5): p. e1000120.

87. Aznarez, I., et al., *Mechanism of Nonsense-Mediated mRNA Decay Stimulation by Splicing Factor SRSF1.* Cell Reports, 2018. **23**(7): p. 2186-2198.

88. Rahman, M.A., et al., *Recurrent SRSF2 mutations in MDS affect both splicing and NMD.* Genes & Development, 2020. **34**(5-6): p. 413-427.

89. Palma, M., et al., *A role for AKT1 in nonsense-mediated mRNA decay.* Nucleic Acids Res, 2021. **49**(19): p. 11022-11037.

90. Cho, H., et al., *AKT constitutes a signal-promoted alternative exon-junction complex that regulates nonsense-mediated mRNA decay.* Mol Cell, 2022. **82**(15): p. 2779-2796.e10.

91. Hurt, J.A., A.D. Robertson, and C.B. Burge, *Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay.* Genome Research, 2013. **23**(10): p. 1636-1650.

92. Zünd, D., et al., *Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs.* Nature Structural & Molecular Biology, 2013. **20**(8): p. 936-943.

93. Raimondeau, E., J.C. Bufton, and C. Schaffitzel, *New insights into the interplay between the translation machinery and nonsense-mediated mRNA decay factors.* Biochemical Society Transactions, 2018. **46**(3): p. 503-512.

94. Melero, R., et al., *Structures of SMG1-UPFs Complexes: SMG1 Contributes to Regulate UPF2-Dependent Activation of UPF1 in NMD.* Structure, 2014. **22**(8): p. 1105-1119.

95. Yamashita, A., et al., *SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay.* Genes & Development, 2009. **23**(9): p. 1091-1105.

96. Arias-Palomo, E., et al., *The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8.* Genes & Development, 2011. **25**(2): p. 153-164.

97. Fukuhara, N., et al., *SMG7 Is a 14-3-3-like Adaptor in the Nonsense-Mediated mRNA Decay Pathway.* Molecular Cell, 2005. **17**(4): p. 537-547.

98. Okada-Katsuhata, Y., et al., *N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD.* Nucleic Acids Research, 2012. **40**(3): p. 1251-1266.

99. Ohnishi, T., et al., *Phosphorylation of hUPF1 Induces Formation of mRNA Surveillance Complexes Containing hSMG-5 and hSMG-7.* Molecular Cell, 2003. **12**(5): p. 1187-1200.

100. Eberle, A.B., et al., *SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells.* Nature Structural & Molecular Biology, 2009. **16**(1): p. 49-55.

101. Loh, B., S. Jonas, and E. Izaurralde, *The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2.* Genes & Development, 2013. **27**(19): p. 2125-2138.

102. Unterholzner, L. and E. Izaurralde, *SMG7 Acts as a Molecular Link between mRNA Surveillance and mRNA Decay.* Molecular Cell, 2004. **16**(4): p. 587-596.

103. Lykke-Andersen, J., *Identification of a human decapping complex associated with hUpf proteins in nonsense-mediated decay.* Molecular and Cellular Biology, 2002. **22**(23): p. 8114-8121.

104. Lejeune, F., X. Li, and L.E. Maquat, *Nonsense-Mediated mRNA Decay in Mammalian Cells Involves Decapping, Deadenylating, and Exonucleolytic Activities.* Molecular Cell, 2003. **12**(3): p. 675-687.

105. Cho, H., K.M. Kim, and Y.K. Kim, *Human Proline-Rich Nuclear Receptor Coregulatory Protein 2 Mediates an Interaction between mRNA Surveillance Machinery and Decapping Complex.* Molecular Cell, 2009. **33**(1): p. 75-86.

106. Fenger-Grøn, M., et al., *Multiple Processing Body Factors and the ARE Binding Protein TTP Activate mRNA Decapping.* Molecular Cell, 2005. **20**(6): p. 905-915.

107. Lai, T., et al., *Structural Basis of the PNRC2-Mediated Link between mRNA Surveillance and Decapping.* Structure, 2012. **20**(12): p. 2025-2037.

108. Huntzinger, E., et al., *SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan.* RNA, 2008. **14**(12): p. 2609-2617.

109. Gatfield, D. and E. Izaurralde, *Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila.* Nature, 2004. **429**(6991): p. 575-578.

110. Kurosaki, T., et al., *NMD-degradome sequencing reveals ribosome-bound intermediates with 3'-end non-templated nucleotides.* Nature Structural & Molecular Biology, 2018. **25**(10): p. 940-950.

111. da Costa, P.J., et al., *A role for DIS3L2 over natural nonsense-mediated mRNA decay targets in human cells.* Biochemical and Biophysical Research Communications, 2019. **518**(4): p. 664-671.

112. Amrani, N., et al. *Aberrant termination triggers nonsense-mediated mRNA decay*. in *Biochemical Society Transactions*. 2006.

113. Peixeiro, I., et al., *Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations.* Nucleic Acids Research, 2012. **40**(3): p. 1160-1173.

114. Ivanov, A., et al., *PABP enhances release factor recruitment and stop codon recognition during translation termination.* Nucleic Acids Research, 2016. **44**(16): p. 7766-7776.

115. Amrani, N., et al., *A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay.* Nature, 2004. **432**(7013): p. 112-118.

116. Giorgi, C., et al., *The EJC Factor eIF4AIII Modulates Synaptic Strength and Neuronal Protein Expression.* Cell, 2007. **130**(1): p. 179-191.

117. Colak, D., et al., *Regulation of Axon Guidance by Compartmentalized Nonsense-Mediated mRNA Decay.* Cell, 2013. **153**(6): p. 1252-1265.

118. Lou, Chih H., et al., *Posttranscriptional Control of the Stem Cell and Neurogenic Programs by the Nonsense-Mediated RNA Decay Pathway.* Cell Reports, 2014. **6**(4): p. 748-764.

119. Seuntjens, E., et al., *Transforming Growth Factor type β and Smad family signaling in stem cell function.* Cytokine & Growth Factor Reviews, 2009. **20**(5): p. 449-458.

120. Jolly, L.A., et al., *The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth.* Human Molecular Genetics, 2013. **22**(23): p. 4673-4687.

121. Laumonnier, F., et al., *Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism.* Molecular Psychiatry, 2010. **15**(7): p. 767-776.

122. Notaras, M., et al., *UPF2 leads to degradation of dendritically targeted mRNAs to regulate synaptic plasticity and cognitive function.* Molecular Psychiatry, 2020. **25**(12): p. 3360-3379.

123. Tan, K., et al., *The role of the NMD factor UPF3B in olfactory sensory neurons.* eLife, 2020. **9**: p. e57525.

124. Brann, J.H. and S.J. Firestein, *A lifetime of neurogenesis in the olfactory system.* Frontiers in Neuroscience, 2014. **8**.

125. Serizawa, S., K. Miyamichi, and H. Sakano, *One neuron–one receptor rule in the mouse olfactory system.* Trends in Genetics, 2004. **20**(12): p. 648-653.

126. Serizawa, S., et al., *Negative Feedback Regulation Ensures the One Receptor-One Olfactory Neuron Rule in Mouse.* Science, 2003. **302**(5653): p. 2088-2094.

127. Chess, A., et al., *Allelic inactivation regulates olfactory receptor gene expression.* Cell, 1994. **78**(5): p. 823-834.

128. Ropers, H.H. and B.C.J. Hamel, *X-linked mental retardation.* Nature Reviews Genetics, 2005. **6**(1): p. 46-57.

129. Raymond, F.L. and P. Tarpey, *The genetics of mental retardation.* Human Molecular Genetics, 2006. **15**(suppl_2): p. R110-R116.

130. Tarpey, P.S., et al., *Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation.* Nature Genetics, 2007. **39**(9): p. 1127-1133.

131. Lujan, J.E., et al., *A form of X-linked mental retardation with marfanoid habitus.* American Journal of Medical Genetics, 1984. **17**(1): p. 311-322.

132. Fryns, J.-P., et al., *X-linked mental retardation with marfanoid habitus.* American Journal of Medical Genetics, 1987. **28**(2): p. 267-274.

133. Szyszka, P., et al., *A nonconservative amino acid change in the UPF3B gene in a patient with schizophrenia.* Psychiatric Genetics, 2012. **22**(3): p. 150-151.

134. Tzschach, A., et al., *Next-generation sequencing in X-linked intellectual disability.* European Journal of Human Genetics, 2015. **23**(11): p. 1513-1518.

135. Zhang, Y., et al., *Gene Mutation Analysis in 253 Chinese Children with Unexplained Epilepsy and Intellectual/Developmental Disabilities.* PLOS ONE, 2015. **10**(11): p. e0141782.

136. Hu, H., et al., *X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes.* Molecular Psychiatry, 2016. **21**(1): p. 133-148.

137. Xu, X., et al., *Exome sequencing identifies UPF3B as the causative gene for a Chinese non-syndrome mental retardation pedigree.* Clinical Genetics, 2013. **83**(6): p. 560-564.

138. Soden, S.E., et al., *Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders.* Science Translational Medicine, 2014. **6**(265): p. 265ra168-265ra168.

139. Chérot, E., et al., *Using medical exome sequencing to identify the causes of neurodevelopmental disorders: Experience of 2 clinical units and 216 patients.* Clinical Genetics, 2018. **93**(3): p. 567-576.

140. Lynch, S.A., et al., *Broadening the phenotype associated with mutations in UPF3B: Two further cases with renal dysplasia and variable developmental delay.* European Journal of Medical Genetics, 2012. **55**(8): p. 476-479.

141. Yavarna, T., et al., *High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders.* Human Genetics, 2015. **134**(9): p. 967-980.

142. Addington, A.M., et al., *A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders.* Molecular Psychiatry, 2011. **16**(3): p. 238-239.

143. Tejada, M.I., et al., *Molecular and clinical characterization of a novel nonsense variant in exon 1 of the upf3b gene found in a large spanish basque family (Mrx82).* Frontiers in Genetics, 2019. **10**(OCT).

144. Lovrečić, L., et al., *Diagnostic efficacy and new variants in isolated and complex autism spectrum disorder using molecular karyotyping.* Journal of Applied Genetics, 2018. **59**(2): p. 179-185.

145. Jin, Z., et al., *A novel 47.2Mb duplication on chromosomal bands Xq21.1–25 associated with mental retardation.* Gene, 2015. **567**(1): p. 98-102.

146. Nguyen, L.S., et al., *Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability.* Molecular Psychiatry, 2012. **17**(11): p. 1103-1115.

147. Huang, L., et al., *A Upf3b-mutant mouse model with behavioral and neurogenesis defects.* Molecular Psychiatry, 2018. **23**(8): p. 1773-1786.

148. Chan, W.-K., et al., *A UPF3-mediated regulatory switch that maintains RNA surveillance.* Nature Structural & Molecular Biology, 2009. **16**(7): p. 747-753.

149. Yi, Z., et al., *Mammalian UPF3A and UPF3B can activate nonsense-mediated mRNA decay independently of their exon junction complex binding.* The EMBO Journal, 2022. **41**(10): p. e109202.

150. Wallmeroth, D., et al., *Human UPF3A and UPF3B enable fault-tolerant activation of nonsense-mediated mRNA decay.* The EMBO Journal, 2022. **41**(10): p. e109191.

151. Chengyan, C., et al., *UPF3A is dispensable for nonsense-mediated mRNA decay in mouse pluripotent and somatic cells.* Life Science Alliance, 2023. **6**(6): p. e202201589.

152. Zamudio, R.M., et al., *Human embryonic kidney (HEK-293) cell line: An alternative for rabies virus diagnosis and research.* J Virol Methods, 2021. **294**: p. 114195.

153. Hening, P., et al., *The neuroprotective effect of Ocimum sanctum Linn. ethanolic extract on human embryonic kidney-293 cells as in vitro model of neurodegenerative disease.* Vet World, 2018. **11**(9): p. 1237-1243.

154. Horibata, S., et al., *Host gene expression modulated by Zika virus infection of human-293 cells.* Virology, 2021. **552**: p. 32-42.

155. Shaw, G., et al., *Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells.* Faseb j, 2002. **16**(8): p. 869-71.

156. Deka, B., P. Chandra, and K.K. Singh, *Functional roles of human Up-frameshift suppressor 3 (UPF3) proteins: From nonsense-mediated mRNA decay to neurodevelopmental disorders.* Biochimie, 2021. **180**: p. 10-22.

157. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11): p. 2281-2308.

158. Jacobi, A.M., et al., *Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes.* Methods, 2017. **121-122**: p. 16-28.

159. He, S.L. and R. Green, *Northern blotting.* Methods Enzymol, 2013. **530**: p. 75-87.

160. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-70.

161. Casneuf, T., Y. Van de Peer, and W. Huber, *In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation.* BMC Bioinformatics, 2007. **8**: p. 461.

162. Shendure, J., *The beginning of the end for microarrays?* Nature Methods, 2008. **5**(7): p. 585-587.

163. Adams, M.D., et al., *Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project.* Science, 1991. **252**(5013): p. 1651-1656.

164. Velculescu, V.E., et al., *Serial Analysis of Gene Expression.* Science, 1995. **270**(5235): p. 484-487.

165. Shiraki, T., et al., *Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage.* Proceedings of the National Academy of Sciences, 2003. **100**(26): p. 15776-15781.

166. Shendure, J. and H. Ji, *Next-generation DNA sequencing.* Nature Biotechnology, 2008. **26**(10): p. 1135-1145.

167. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature Reviews Genetics, 2009. **10**(1): p. 57-63.

168. Le, H.S., et al., *Probabilistic error correction for RNA sequencing.* Nucleic Acids Res, 2013. **41**(10): p. e109.

169. Oshlack, A. and M.J. Wakefield, *Transcript length bias in RNA-seq data confounds systems biology.* Biol Direct, 2009. **4**: p. 14.

170. Tuerk, A., G. Wiktorin, and S. Güler, *Mixture models reveal multiple positional bias types in RNA-Seq data and lead to accurate transcript concentration estimates.* PLoS Comput Biol, 2017. **13**(5): p. e1005515.

171. Han, Y., et al., *Advanced Applications of RNA Sequencing and Challenges.* Bioinform Biol Insights, 2015. **9**(Suppl 1): p. 29-46.

172. Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-3048.

173. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data AnalysisDO - 10.14806/ej.17.1.200, 2011.

174. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

175. Anders, S., A. Reyes, and W. Huber, *Detecting differential usage of exons from RNA-seq data.* Genome Research, 2012. **22**(10): p. 2008-2017.

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

177. Rutter, L. and D. Cook, *bigPint: A Bioconductor visualization package that makes big data pint-sized.* PLOS Computational Biology, 2020. **16**(6): p. e1007912.

178. Nelson, J.O., et al., *Degradation of Gadd45 mRNA by nonsense-mediated decay is essential for viability.* Elife, 2016. **5**.

179. Moriarty, P.M., C.C. Reddy, and L.E. Maquat, *Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA.* Mol Cell Biol, 1998. **18**(5): p. 2932-9.

189. Sun, X., et al., *Nonsense-mediated decay of mRNA for the selenoprotein phospholipid hydroperoxide glutathione peroxidase is detectable in cultured cells but masked or inhibited in rat tissues.* Mol Biol Cell, 2001. **12**(4): p. 1009-17.

190. de la Calle-Mustienes, E., et al., *Xiro homeoproteins coordinate cell cycle exit and primary neuron formation by upregulating neuronal-fate repressors and downregulating the cell-cycle inhibitor XGadd45-gamma.* Mech Dev, 2002. **119**(1): p. 69-80.

191. Forceville, X., et al., *Selenoprotein P, rather than glutathione peroxidase, as a potential marker of septic shock and related syndromes.* Eur Surg Res, 2009. **43**(4): p. 338-47.

192. Takemoto, A.S., M.J. Berry, and F.P. Bellinger, *Role of selenoprotein P in Alzheimer's disease.* Ethn Dis, 2010. **20**(1 Suppl 1): p. S1-92-5.

193. He, F., et al., *Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast.* Mol Cell, 2003. **12**(6): p. 1439-52.

194. Rehwinkel, J., et al., *Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets.* Rna, 2005. **11**(10): p. 1530-44.

195. Wittmann, J., E.M. Hol, and H.M. Jäck, *hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay.* Mol Cell Biol, 2006. **26**(4): p. 1272-87.

196. Huang, L., et al., *RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD.* Mol Cell, 2011. **43**(6): p. 950-61.

197. Yepiskoposyan, H., et al., *Autoregulation of the nonsense-mediated mRNA decay pathway in human cells.* Rna, 2011. **17**(12): p. 2108-18.

198. Azzalin, C.M. and J. Lingner, *The human RNA surveillance factor UPF1 is required for S phase progression and genome stability.* Curr Biol, 2006. **16**(4): p. 433-9.

199. Luo, W., et al., *GAGE: generally applicable gene set enrichment for pathway analysis.* BMC Bioinformatics, 2009. **10**(1): p. 161.

200. Boehm, V., et al., *SMG5-SMG7 authorize nonsense-mediated mRNA decay by enabling SMG6 endonucleolytic activity.* Nature Communications, 2021. **12**(1): p. 3965.

201. Sureau, A., et al., *SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs.* The EMBO Journal, 2001. **20**(7): p. 1785-1796.

202. Lykke-Andersen, S., et al., *Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes.* Genes Dev, 2014. **28**(22): p. 2498-517.

203. Park, S.Y., et al., *Depletion of BubR1 promotes premature centrosomal localization of cyclin B1 and accelerates mitotic entry.* Cell Cycle, 2009. **8**(11): p. 1754-64.

204. Haruki, N., et al., *Molecular analysis of the mitotic checkpoint genes BUB1, BUBR1 and BUB3 in human lung cancers.* Cancer Lett, 2001. **162**(2): p. 201-5.

205. Grabsch, H., et al., *Overexpression of the mitotic checkpoint genes BUB1, BUBR1, and BUB3 in gastric cancer--association with tumour cell proliferation.* J Pathol, 2003. **200**(1): p. 16-22.

206. Yuan, B., et al., *Increased expression of mitotic checkpoint genes in breast cancer cells with chromosomal instability.* Clin Cancer Res, 2006. **12**(2): p. 405-10.

207. Hussain, M.S., et al., *CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly.* Hum Mol Genet, 2013. **22**(25): p. 5199-214.

208. Deardorff, M.A., et al., *Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation.* Am J Hum Genet, 2007. **80**(3): p. 485-94.

209. Candal, E., et al., *Medaka as a model system for the characterisation of cell cycle regulators: a functional analysis of Ol-Gadd45gamma during early embryogenesis.* Mech Dev, 2004. **121**(7-8): p. 945-58.

210. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nature Methods, 2017. **14**(4): p. 417-419.

211. Lawrence, M., et al., *Software for Computing and Annotating Genomic Ranges.* PLOS Computational Biology, 2013. **9**(8): p. e1003118.

212. Soneson, C., M.I. Love, and M.D. Robinson, *Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences.* F1000Res, 2015. **4**: p. 1521.

213. Ge, S.X., D. Jung, and R. Yao, *ShinyGO: a graphical gene-set enrichment tool for animals and plants.* Bioinformatics, 2020. **36**(8): p. 2628-2629.

214. Vitting-Seerup, K. and A. Sandelin, *IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences.* Bioinformatics, 2019. **35**(21): p. 4469-4471.

215. Reyes, A., et al., *Drift and conservation of differential exon usage across tissues in primate species.* Proceedings of the National Academy of Sciences, 2013. **110**(38): p. 15377-15382.

216. Kang, Y.-J., et al., *CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features.* Nucleic Acids Research, 2017. **45**(W1): p. W12-W16.

217. Sonnhammer, E.L., S.R. Eddy, and R. Durbin, *Pfam: a comprehensive database of protein domain families based on seed alignments.* Proteins, 1997. **28**(3): p. 405-20.

218. Teufel, F., et al., *SignalP 6.0 predicts all five types of signal peptides using protein language models.* Nature Biotechnology, 2022. **40**(7): p. 1023-1025.

219. Weischenfeldt, J., et al., *Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns.* Genome Biol, 2012. **13**(5): p. R35.

220. Hyvönen, M.T., et al., *Polyamine-regulated unproductive splicing and translation of spermidine/spermine N1-acetyltransferase.* Rna, 2006. **12**(8): p. 1569-82.

221. Hyvönen, M.T., et al., *Tissue-specific alternative splicing of spermidine/spermine N1-acetyltransferase.* Amino Acids, 2012. **42**(2): p. 485-493.

222. Burstein, E., et al., *COMMD proteins, a novel family of structural and functional homologs of MURR1.* J Biol Chem, 2005. **280**(23): p. 22222-32.

223. McGlincy, N.J., et al., *Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay.* BMC Genomics, 2010. **11**: p. 565.

224. Candal, E., et al., *Medaka as a model system for the characterisation of cell cycle regulators: a functional analysis of Ol-Gadd45γ during early embryogenesis.* Mechanisms of Development, 2004. **121**(7): p. 945-958.

225. Alto, L.T. and J.R. Terman, *Semaphorins and their Signaling Mechanisms*. Methods Mol Biol, 2017. **1493**: p. 1-25.

226. Pasterkamp, R.J., *Getting neural circuits into shape with semaphorins.* Nat Rev Neurosci, 2012. **13**(9): p. 605-18.

227. Lai Wing Sun, K., J.P. Correia, and T.E. Kennedy, *Netrins: versatile extracellular cues with diverse functions.* Develop-ment, 2011. **138**(11): p. 2153-69.

228. Karousis, E.D., Gypas, F., Zavolan, M. et al. *Nanopore sequencing reveals endogenous NMD-targeted isoforms in human cells.* Genome Biol 22, **223** (2021).

229. Li, B., et al., *A Novel Pseudogene Methylation Signature to Predict Temozolomide Outcome in Non-G-CIMP Glioblastomas.* J Oncol, 2022. 2022: p. 6345160.

230. Alarcón, C.R., et al., *HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events.* Cell, 2015. **162**(6): p. 1299-308.

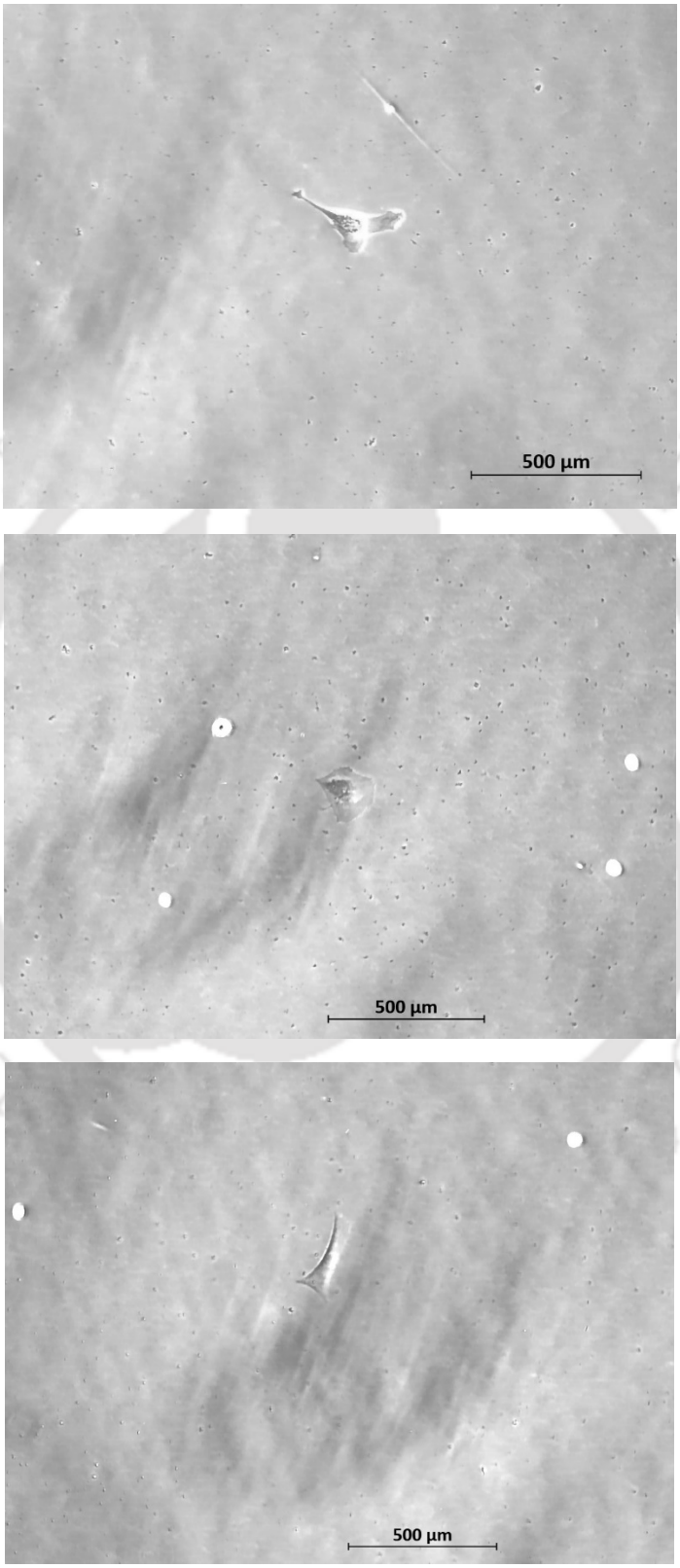
231. Jiang L, et al. *Interaction of tau with HNRNPA2B1 and N6-methyladenosine RNA mediates the progression of tauopathy.* Mol Cell. 2021 Oct 21;**81**(20):4209-4227.e12.

232. Biederer, T. and P. Scheiffele, *Mixed-culture assays for analyzing neuronal synapse formation.* Nature Protocols, 2007. **2**(3): p. 670-676.

APPENDICES

APPENDIX I

SHSY-5Y cells (40 days' post-transfection)



APPENDIX II

List of cell lines

Name	Parental cell line	Genome alteration	Source
Flp-In T-REx HEK-293			Invitrogen; Cat # R78007
Control	Flp-In T-REx HEK-293	Genome integration of PB-CuO-FLAG-Cas9-GFP-Puro plasmid	Generated in the lab
Clone1	Control	Deletion of the *UPF3B* gene	Generated in the lab
Clone3	Control	Deletion of the *UPF3B* gene	Generated in the lab

APPENDIX III

List of plasmids

Name	Vector backbone	Source
px330		Addgene; Cat # 42230
PB-CuO-MCS-IRES-GFP-EF1α-CymR-Puro		System Biosciences; Cat # PBQM812A
pCI-FLAG		Promega; Cat # E1731
pCI-FLAG-UPF3B	pCI-FLAG	From Prof. Niels Gehring
pCI-FLAG-PB-Transposase	pCI-FLAG	From Prof. Niels Gehring
PB-CuO-3xFLAG-Cas9-IRES-GFP-EF1α-CymR-Puro	PB-CuO-MCS-IRES-GFP-EF1α-CymR-Puro	Generated in our lab
pCI-globin_PTC39	pCI-neo	From Prof. Neils Gehring

APPENDIX IV

List of primers

Oligo Name	Sequence (5'→3')	Purpose
Amplification of the *hSpCas9* gene		
FP-3xFLAG	CCTAAGCTAGCATGGACTATAAGGACCACGAC	PCR
RP-BGH-polyA	TTCCTGCGGCCGCTCCCCAG	PCR
Amplification of the *UPF3B* gene fragment for knockout detection		
FP-UPF3B-5'UTR	GCATTTGTTGGGGGCGGAGCTTTGA	PCR
RP-UPF3B-I1	CCCATTGCTCTCAAAATAGGATGCTCTCGA	PCR
Primers for gene expression		
FP-Actin	CTGTACGCCAACACAGTGCT	qPCR
RP-Actin	GCTCAGGAGGAGCAATGATC	qPCR
FP-GADD45G	CGTCTACGAGTCAGCCAAAGTC	qPCR
RP-GADD45G	CGATGTCGTTCTCGCAGCA	qPCR
FP-GPX1	GTGCTCGGCTTCCCGT	qPCR
RP-GPX1	CTCGAAGAGCATGAAGTTGG	qPCR
FP-SELENOP	CATCTTGGTTTGCCTTTTTCC	qPCR
RP-SELENOP	ACAGTAGCCAAAGATACACGTTTAC	qPCR
FP-GAS5	TCGGCTTGACTACACTGTGT	qPCR
RP-GAS5	TGCTTGCTTGTTGTGGTCAT	qPCR
FP-ITGAE	GAGTCAGGAGCGCGCTT	qPCR
RP-ITGAE	CACCAAGGATCTGCAGTTCA	qPCR
FP-SRSF2	CGGTCTCCAGATCTCGTTCG	qPCR
RP-SRSF2	CAGGAGACCGCAGCATTTTC	qPCR
FP-ZFAS1	ATGCGGGTGTTGGAAGTAGA	qPCR
RP-ZFAS1	ACAAGGCAGACTGAATCAAGC	qPCR
FP-β-globin	CGTGGATGAAGTTGGTGGTG	qPCR
RP-PTC39 β-globin	CCAAAGGACTCAAAGAACCTCTAGGTCCAAGGGTAGAC CACC	qPCR
FP-UPF1	AACGAGCACCAAGGCATTGG	qPCR
RP-UPF1	GGCTGCTTTGATAGTGCCTTCG	qPCR
FP-UPF3A	CCATGGAGAGACTGGGAAG	qPCR
RP-UPF3A	TGCCTTGCAGATCCTCCTG	qPCR
FP-SMG5	CCTCTTTTCCCACCTCGTCAA	qPCR
RP-SMG5	ACAGGTTCCTTGGACTCTGGT	qPCR
FP-SMG6	GATGGTCTTGCCATTCGCAG	qPCR
RP-SMG6	GCTGTATCACTGGCTTGCTC	qPCR

FP-SMG7	TTTCAGGAGGCAGTGGTGG	qPCR
RP-SMG7	CAAACTCCTCTGGAAGTGGTGTC	qPCR
FP-RBL1	CGAACTGACAGTGGGAGTCTTC	qPCR
RP-RBL1	TCTCTTAGCACTCCCTGCGGTA	qPCR
FP-RBL2	GAAGTTCGTGCTGATACTGGAGG	qPCR
RP-RBL2	AATAGCCGCCTTCTGGTAGTGC	qPCR
FP-CDC7	GGAAAACTGCCAGTTCTTGCC	qPCR
RP-CDC7	GGCACTTTGTCAAGACCTCTGG	qPCR
FP-CDK6	GGATAAAGTTCCAGAGCCTGGAG	qPCR
RP-CDK6	GCGATGCACTACTCGGTGTG	qPCR
FP-PRKDC	GCGCCATATCTGTCATCTGC	qPCR
RP-PRKDC	TTATAGCGGCGCTTCAGGTC	qPCR
FP-BUB1B	TGGGTCCTTCTGGAAACTTAGC	qPCR
RP-BUB1B	AAGCTCCCCAAGAACAGACA	qPCR
FP-SMC3	ATGCGTGGAAGTCACTGCTG	qPCR
RP-SMC3	GGCAGAAAAGTAACCTCTCCAGG	qPCR
FP-CDC27	AGCAGCAGCAGAAGGTTTGA	qPCR
RP-CDC27	TTGGCACAGTACCCAACCAG	qPCR
FP-ATM	CTCTGAGTGGCAGCTGGAAGA	qPCR
RP-ATM	TTTAGGCTGGGATTGTTCGCT	qPCR
FP-ATR	GGAGATTTCCTGAGCATGTTCG	qPCR
RP-ATR	GGCTTCTTTACTCCAGACCAATC	qPCR
Primers for NMD transcripts		
FP-GADD45B	GTCTCCTGGTCACGAACCC	qPCR
RP-GADD45B	GTGGGTTGGGTTGTTTTGG	qPCR
FP-COMMD2	TTGCTTCAGCTTTATCTGGACAAC	qPCR
RP-COMMD2	GCAAGCTTTAGAAAGGAGTGAAGG	qPCR
FP-SAT1	CCGGAAGGTTACAGTCTCTAGC	qPCR
RP-SAT1	GCCAAAGCCTCTATAATCACTCATC	qPCR
FP-ZNF767P	AGGGCAGGGGCAGTCA	qPCR
RP-ZNF767P	CACGGAGAAATCGGAGCTG	qPCR
FP-HNRNPA2B1	CTTTAGCATTGATATGAACCATGGAC	qPCR
RP-HNRNPA2B1	GACACCTTCCATTTTACTACCAAC	qPCR
Primers for protein coding transcripts		
FP-SAT1	GACTCCGGAAGGACACAGC	qPCR
RP-SAT1	GCCAAAGCCTCTATAATCACTCATC	qPCR
FP-ZNF767P	GCGGCGACGACTTTATCTG	qPCR
RP-ZNF767P	CACGGAGAAATCGGAGCTG	qPCR

FP-HNRNPA2B1	GGTAGGTAAAAGTTAGGCTGCAAG	qPCR
RP-HNRNPA2B1	CTCTTCAGGGTAAGACACTGAAC	qPCR
FP-SRSF2	CGGTCTCCAGATCTCGTTCG	qPCR
RP-SRSF2	CTGCTTGCCGATACATCATTTTC	qPCR

APPENDIX V

List of antibodies

Name of antibody	Dilution	Source
Primary antibodies		
Mouse Anti-FLAG	1:2000	Santa Cruz Biotechnology, Cat#sc-166355
Rabbit polyclonal anti-UPF3B	1:2000	Eurogentech, custom-made, raised against C-terminal UPF3B (300-483) and affinity purified; From Prof. Neils Gehring's Lab
Goat polyclonal anti-UPF1	1:1000	Bethyl Laboratories Inc. Cat# A300-038A-T
Mouse monoclonal anti-UPF3A	1:2000	Sigma-Aldrich, Cat# SAB1402625-100UG
Rabbit polyclonal anti-SMG5	1:800	Invitrogen, Cat# PA5-44295
Rabbit polyclonal anti-SMG6	1:2000	antibodies.com, Cat# A16786
Rabbit polyclonal anti-SMG7	1:2000	antibodies.com, Cat# A88094
Rabbit polyclonal anti-EIF4A3	1:2000	Bethyl Laboratories Inc. Cat# A302-980A
Rabbit polyclonal anti-MAGOH	1:2000	Invitrogen, Cat# PA5-80788
Rabbit monoclonal anti-SRSF10	1:1000	Cell Signalling Technology, Cat#46905
Rabbit polyclonal anti-ERF3	1:1000	Cell Signalling Technology, Cat#14980
Mouse monoclonal anti-β-Tubulin	1:10,000	Sigma, Cat# T8328-100UL
Rabbit polyclonal anti-Histone	1:10,000	BioBharati, Cat# BB-AB0055
Secondary antibodies		
Anti-mouse IgG, HRP-linked Antibody	1:5000	Cell Signalling Technology, Cat#7076
Anti-rabbit IgG, HRP-linked Antibody	1:5000	Cell Signalling Technology, Cat#7074
Anti-goat IgG, HRP-linked Antibody	1:5000	Santa Cruz Biotechnology, Cat#sc-2354

APPENDIX VI

List of off-target loci for UPF3B gRNA

S. No.	Sequence	PAM	Score	#MM	Gene	Locus
0.	ACCTCGGGGGACAGCTCCAA	GGG	N/A		UPF3B	chrX:-119852824
1.	CCCT-GGGGAACAGCTCCAA	AAG	22	3		chr9:-136079930
2.	ATCT-TGGGGACAGCTCCAA	GAG	23	3		chr19:+15675664
3.	ACCGCGGAGGGCAGCTCCAA	AGG	23	3	FAM110A	chr20:+833791
4.	TCCTC-AGGGACAGCTCCAA	AGG	26	3		chr2:+238460529
5.	ACCT--AGGGACAGCTCCAA	GAG	29	3	RNF144A	chr2:+7058170
6.	ACCTAGATGAACAGCTCCAA	TAG	32	4		chr13:-65013115
7.	TTCTCGGGCAACAGCTCCAA	GGG	38	4	SCAMP2	chr15:-74850606
8.	ACATCAGAGCACAGCTCCAA	AGG	39	4		chr11:+15697879
9.	ACTTCTAGGGACAGCTCCTA	GGG	40	4		chr12:-88953449
10.	ACCTAGGAGGGACAGCTCCAG	TGG	41	3	NOVA2	chr19:+45939222
11.	ACCT--GGGGACAGCTCCTA	CAG	42	3		chr15:-77824953
12.	ACCAGGGAAGACAGCTCCAA	AGG	43	4		chr2:-19874598
13.	GCCTGGTGGGCCAGCTCCAA	AGG	44	4		chr9:+114608385
14.	ATCTCAGGGAACAGCTCCTA	GAG	45	4		chr5:-93735415
15.	ACATCTGGGGA-AGCTCCAA	TAG	48	3		chr1:+203296049
16.	ACCTTGTCGGGGACAGCTCCAA	GAG	48	3		chr10:-1559311
17.	ACCT-GGTGGCCAGCTCCAA	GAG	49	3		chr14:+92954390
18.	TCCTCCAGGGACAGCTCCCA	GGG	50	4	SPIRE2	chr16:-89828607
19.	TCCTCCAGGGACAGCTGCAA	TGG	50	4		chr9:+129186455
20.	TCCTCCGAGGACAGCTCCCA	AGG	51	4		chr4:+6022791
21.	CCCCCGGGGGACAGCTCCTT	CAG	51	4		chr20:+45407040
22.	CCCTCTGAGGACAGCTTCAA	AGG	52	4		chrX:+106872635
23.	ACCT--GGGGACAACTCCAA	AAG	53	3		chr1:-238585818
24.	TCATCGAGGGACATCTCCAA	CAG	53	4		chr16:-79623428
25.	TCCTCTGGGGACAGCTCCAG	TGG	56	3	ALS2CL	chr3:+46685347
26.	ACCT-GGGTGCCAGCTCCAA	GGG	58	3		chr1:+89604429
27.	GCCTCTTGGGACAGCTTCAA	AGG	58	4		chr12:+47883305
28.	GCCTGGGGGAACAGCTCCCA	GGG	58	4		chr8:+35808363
29.	ACCTGGAGGGGCTGCTCCAA	AAG	59	4		chr15:+46210069
30.	AACTCTGGGAATAGCTCCAA	TAG	60	4		chr1:+169884061
31.	ATCTCTGGTGGCAGCTCCAA	AAG	60	4		chr8:-68510471
32.	ACCTAGGGGGACGGCACCAA	TGG	61	3		chr19:-30106824
33.	ACCTCATAGGACAGCTACAA	GGG	62	4		chr2:-234846997
34.	GCCTCGGGAGGACAGCTCCCA	AGG	64	3		chr1:+227913995
35.	ACCT--GGGGACAGCTCCCA	AGG	66	3		chr1:+18597511
36.	ACCTCGGCGGAGCAGCTCCCA	CGG	67	3		chr6:-157236565
37.	GCCTCTGTGGACAGCTCCAT	AGG	67	4	LOC105379098	chr5:-98093542
38.	ACCCAGGGGGGCAGGTCCAA	GGG	67	4		chrX:-80561673
39.	ACAGCGGGTGTCAGCTCCAA	GAG	68	4		chr19:+42108170
40.	ACCTGGAGGAACAGCTCAAA	GGG	69	4		chrX:-17545088
41.	TCCTCGGTCGACAGCTCCCA	CAG	69	4	TYK2	chr19:+10366524
42.	GCCTAGGGGCACAGCCCCAA	GAG	70	4		chr5:-77310413
43.	AACTCGGAGGCCAGCACCAA	AGG	70	4		chr9:-113747905
44.	ACCACGGGGTGCAGCTCCCA	TGG	71	4		chr1:+7143132
45.	TCCTAGGGGGCCAGCTCCAG	AGG	71	4		chr9:-133102717
46.	ACCTCAGGGG-CAGCTCCAG	AGG	73	3		chr20:+21046891
47.	ATCTCGAGAGCCAGCTCCAA	AGG	74	4	H1FOO	chr3:-129554435
48.	ACCTCGGGGGA-AGCTCCAC	AGG	75	2		chr10:-128009534
49.	ACCCCGGGGGACAG-TCCAG	GAG	76	3		chr11:+2026254
50.	ACCTCCAGGGACAACTCCCA	GAG	76	4		chr21:+44332140
51.	GCCTCAGGGGGCACCTCCAA	GGG	77	4		chr7:-142969252
52.	ACCACGGGGCACAGCTCCTT	AGG	77	4		chr3:+32885367
53.	ACTTCTGGGTACAGATCCAA	AGG	78	4		chr2:-41009511
54.	ACATCAGGAGAAAGCTCCAA	AAG	78	4		chr4:+160665139

55.	ACATCAGGAGAAAGCTCCAA	AAG	78	4		chr20:+23814593
56.	ACATCAGGAGAAAGCTCCAA	AAG	78	4		chr7:-23859308
57.	ACATCAGGAGAAAGCTCCAA	AAG	78	4		chr5:-113623017
58.	ACCGCTGGGGGCAGCTCCCA	GGG	79	4		chr7:+102464711
59.	TCCTGGGGGGACAGCTCCGT	GGG	79	4	LOC105371809	chr17:-48642901
60.	ACCTGTGGTGACAGCTACAA	GAG	79	4		chr12:+31875082
61.	ACCTGAGGGGACAGCACCAG	AAG	79	4		chr16:-67653000
62.	ACCT-GCGGGACAGCTGCAA	GGG	80	3		chr12:-10253284
63.	ACCTCGCAGGACAGCTACAA	AGG	80	3		chr16:-57522669
64.	ACCACCGGCGACAGCGCCAA	GAG	80	4	PHLDB1	chr11:+118628597
65.	ACCATGGGGGAGAGCTCCAG	AAG	83	4		chr11:+30295450
66.	ACCTGGGAGGACAGCACCAC	CGG	84	4	TXNRD2	chr22:-19895453
67.	CCCTCTGGGGAAAGCTCCAG	AGG	84	4		chr15:-91005632
68.	ACCTGGAGGGACAGATCCAG	GGG	84	4		chr17:-44916392
69.	GCCTCCGGGGACAGCTCCCT	GAG	84	4		chr4:-8000694
70.	GCCCCGGGGGCCAGCTCCAC	TGG	84	4		chr14:+104437893
71.	CCCTCAGGGGAAACCTCCAA	CAG	85	4		chr11:-96818252
72.	ACCTGGTGAGACAGCTCCCA	GGG	85	4		chr16:+71629101
73.	ACCTCCTGGCAGAGCTCCAA	GGG	85	4		chr22:+29737409
74.	ACCTTGGGGGACAGCCCCCA	GGG	86	3		chr17:-9230232
75.	ACCT-GGGGGCCAGCTCCAC	AGG	87	3		chr15:-81101761
76.	CCCTCAGGAGACAGCTCCCA	GAG	87	4		chr4:-11416400
77.	ACCAGGGGAGACAGGTCCAA	GAG	87	4		chr3:+14644127
78.	GCCTTGGGGGACAGTTGCAA	GAG	87	4		chr14:-54219531
79.	GCCTCGGTGGACAGTTTCAA	TAG	88	4		chr12:-125326860
80.	ACCACTGGGGACAACTGCAA	AGG	88	4		chr2:-192102501
81.	AACTCTGGTGACAGCTCAAA	TGG	88	4		chr2:+80975035
82.	ACATCTGGTGACACCTCCAA	TAG	88	4		chr8:+140257569
83.	AGCTGGGGCCACAGCTCCAA	AGG	90	4		chr1:+190858493
84.	ACCTCAGGCCACTGCTCCAA	AAG	90	4		chr16:-47792171
85.	ACATCTGGGGACAGCTTCCA	CGG	90	4		chr6:+33591074
86.	ACCTATGGGGACAAATCCAA	GGG	92	4		chrX:+30718291
87.	TCCTCGGGGTCCAGCTCAAA	GGG	92	4	PARVA	chr11:-12473803
88.	ACCTTGGGGGAAGGCTCCAT	CAG	93	4	LOC105372791	chr21:+34354238
89.	ACCTGAGGGGTCAGCTCCAT	TGG	93	4		chr22:-35539321
90.	ACCTCCAGGGCCAGCTCAAA	AAG	94	4		chr2:+219348050
91.	ACCTCTGGGGACACCTTCAA	GAG	95	3		chr4:-106842760
92.	ACCTTCCGGGACAGCTCAAA	AAG	95	4		chr6:-39994510
93.	ACCTCGGGGTCCAGCTCCTG	AGG	95	4		chr16:+9798509
94.	AGCTGGTGGGACAGCTCCCA	GAG	96	4	OPN4	chr10:+86654861
95.	AGCCCGGTGGACAGCTGCAA	GGG	98	4		chr16:+22751452
96.	ACTTCTGGGGAAAGTTCCAA	AAG	99	4		chr2:-233402249
97.	CCCTCGGGACACAGCTCCCA	TGG	99	4		chr10:-75352996
98.	TCCTCGGGGCACAGTTTCAA	TAG	100	4		chrX:-8663430
99.	AACTCGGGGAACAGCACCAC	TGG	100	4		chr9:+103104480

APPENDIX VII

List of upregulated genes in Clone1

(Adjusted p-value < 0.001 & Log2FoldChange > 1.2)

Ensemble ID	Gene Name	Log2FoldChange	Ensemble ID	Gene Name	Log2FoldChanges
ENSG00000175899	A2M	5.277317969	ENSG00000139725	RHOF	1.389477771
ENSG00000162711	NLRP3	5.061472197	ENSG00000198840	MT-ND3	1.387071379
ENSG00000261441	POLG-DT	3.547620694	ENSG00000232956	SNHG15	1.383996445
ENSG00000177675	CD163L1	3.532142934	ENSG00000182687	GALR2	1.383877247
ENSG00000166126	AMN	3.119986722	ENSG00000220201	ZGLP1	1.381750584
ENSG00000132698	RAB25	3.083522007	ENSG00000167703	SLC43A2	1.378222735
ENSG00000215478	CES5AP1	2.929294743	ENSG00000126215	XRCC3	1.372375405
ENSG00000253123		2.910110117	ENSG00000166106	ADAMTS15	1.370363526
ENSG00000027869	SH2D2A	2.892008991	ENSG00000178381	ZFAND2A	1.369479058
ENSG00000165349	SLC7A3	2.881396027	ENSG00000064300	NGFR	1.366074403
ENSG00000265386	RN7SL219P	2.796617461	ENSG00000042493	CAPG	1.364460232
ENSG00000143125	PROK1	2.580921765	ENSG00000248498	ASNSP1	1.361143806
ENSG00000149575	SCN2B	2.571772667	ENSG00000249116	IRX4-AS1	1.356741755
ENSG00000005102	MEOX1	2.560851603	ENSG00000179066		1.35500185
ENSG00000123689	G0S2	2.539450793	ENSG00000181019	NQO1	1.353467148
ENSG00000250722	SELENOP	2.528234923	ENSG00000110697	PITPNM1	1.352957061
ENSG00000232363		2.516782486	ENSG00000121577	POPDC2	1.352915855
ENSG00000205879	FAM90A2P	2.507874963	ENSG00000226824		1.35268406
ENSG00000116774	OLFML3	2.447794996	ENSG00000245522	LINC02709	1.349182083
ENSG00000128422	KRT17	2.444396947	ENSG00000161671	EMC10	1.348480669
ENSG00000002933	TMEM176A	2.400248118	ENSG00000186523	FAM86B1	1.347234647
ENSG00000172789	HOXC5	2.393687348	ENSG00000198712	MT-CO2	1.346436366
ENSG00000170454	KRT75	2.383608145	ENSG00000264107	MIR4733HG	1.346244768
ENSG00000259156	CHEK2P2	2.351122154	ENSG00000139899	CBLN3	1.3461519
ENSG00000128310	GALR3	2.328442717	ENSG00000025434	NR1H3	1.345432233
ENSG00000174460	ZCCHC12	2.3195582	ENSG00000197982	C1orf122	1.344705457
ENSG00000157514	TSC22D3	2.245779792	ENSG00000261342		1.341853024
ENSG00000173714	WFIKKN2	2.23450741	ENSG00000117410	ATP6V0B	1.341361495
ENSG00000179855	GIPC3	2.230828858	ENSG00000133874	RNF122	1.340659128
ENSG00000125910	S1PR4	2.195474348	ENSG00000185338	SOCS1	1.340613213
ENSG00000166143	PPP1R14D	2.109189063	ENSG00000152475	ZNF837	1.338936877
ENSG00000266601		2.094854117	ENSG00000256269	HMBS	1.336696931
ENSG00000287064		2.08127884	ENSG00000221821	C6orf226	1.336313522
ENSG00000149418	ST14	2.079651204	ENSG00000172273	HINFP	1.332218362
ENSG00000172061	LRRC15	2.075559539	ENSG00000092850	TEKT2	1.331591678
ENSG00000289423		2.067567802	ENSG00000168071	CCDC88B	1.329779692
ENSG00000237149	ZNF503-AS2	2.059759042	ENSG00000178980	SELENOW	1.32911568
ENSG00000171401	KRT13	2.049788909	ENSG00000106610	STAG3L4	1.329010166
ENSG00000170608	FOXA3	2.021515889	ENSG00000133878	DUSP26	1.327829194
ENSG00000100292	HMOX1	2.015597281	ENSG00000107175	CREB3	1.326376431
ENSG00000232265	LINC02805	2.00772982	ENSG00000171223	JUNB	1.325069389
ENSG00000261241	LINC02128	1.966322532	ENSG00000164077	MON1A	1.323786233
ENSG00000080573	COL5A3	1.964874948	ENSG00000104368	PLAT	1.320012943
ENSG00000175287	PHYHD1	1.93006951	ENSG00000164087	POC1A	1.318818839
ENSG00000158516	CPA2	1.90191582	ENSG00000087086	FTL	1.318769845
ENSG00000054356	PTPRN	1.899899385	ENSG00000143774	GUK1	1.317489497
ENSG00000181097	LINC02878	1.869203891	ENSG00000152137	HSPB8	1.316764239
ENSG00000111405	ENDOU	1.865892502	ENSG00000243449	C4orf48	1.316565653

ENSG00000102109	PCSK1N	1.863589494	ENSG00000100908	EMC9	1.310568403
ENSG00000153303	FRMD1	1.851706439	ENSG00000149295	DRD2	1.309272522
ENSG00000139637	MYG1	1.843076958	ENSG00000119632	IFI27L2	1.309253687
ENSG00000092051	JPH4	1.841228393	ENSG00000115129	TP53I3	1.307591441
ENSG00000275719		1.840743548	ENSG00000214212	C19orf38	1.304433245
ENSG00000185215	TNFAIP2	1.830281107	ENSG00000163472	TMEM79	1.304164438
ENSG00000196154	S100A4	1.826781276	ENSG00000106245	BUD31	1.303431401
ENSG00000036448	MYOM2	1.822016131	ENSG00000164535	DAGLB	1.302984695
ENSG00000103740	ACSBG1	1.809289006	ENSG00000256008	CABP1-DT	1.301971127
ENSG00000101222	SPEF1	1.787659818	ENSG00000162496	DHRS3	1.298770199
ENSG00000198113	TOR4A	1.77879691	ENSG00000165272	AQP3	1.296893224
ENSG00000172020	GAP43	1.759667221	ENSG00000255020	TDH-AS1	1.29668348
ENSG00000287828		1.75646091	ENSG00000203722	RAET1G	1.296441961
ENSG00000272734	ADIRF-AS1	1.750642652	ENSG00000141933	TPGS1	1.296343827
ENSG00000153531	ADPRHL1	1.740938962	ENSG00000288908		1.295016344
ENSG00000177363	LRRN4CL	1.728483121	ENSG00000125775	SDCBP2	1.294357557
ENSG00000159884	CCDC107	1.723298103	ENSG00000179859	RNF227	1.293584289
ENSG00000164082	GRM2	1.703697301	ENSG00000204934	ATP6V0E2-AS1	1.292892952
ENSG00000141738	GRB7	1.695156696	ENSG00000196754	S100A2	1.290356768
ENSG00000253746		1.684894059	ENSG00000196517	SLC6A9	1.285182841
ENSG00000142513	ACP4	1.684048957	ENSG00000279792		1.282960626
ENSG00000230426	LINC01036	1.679055662	ENSG00000204444	APOM	1.278019531
ENSG00000140675	SLC5A2	1.676840322	ENSG00000204252	HLA-DOA	1.277389674
ENSG00000141968	VAV1	1.674908645	ENSG00000114988	LMAN2L	1.273723031
ENSG00000128591	FLNC	1.672774251	ENSG00000168062	BATF2	1.273663959
ENSG00000188404	SELL	1.665940195	ENSG00000142546	NOSIP	1.273529359
ENSG00000100290	BIK	1.655867623	ENSG00000263818	RDM1P5	1.272393024
ENSG00000130222	GADD45G	1.649572288	ENSG00000256802		1.271236879
ENSG00000198576	ARC	1.647692795	ENSG00000179528	LBX2	1.269286207
ENSG00000137440	FGFBP1	1.637736567	ENSG00000089116	LHX5	1.26569006
ENSG00000105369	CD79A	1.637475796	ENSG00000124074	ENKD1	1.265641312
ENSG00000083814	ZNF671	1.633705534	ENSG00000117148	ACTL8	1.263981577
ENSG00000129214	SHBG	1.629566828	ENSG00000204525	HLA-C	1.26171438
ENSG00000184414	IRS3P	1.616789758	ENSG00000178860	MSC	1.260107239
ENSG00000171345	KRT19	1.615093689	ENSG00000235863	B3GALT4	1.259088847
ENSG00000271155		1.609222479	ENSG00000125734	GPR108	1.256026005
ENSG00000281571		1.608790245	ENSG00000179148	ALOXE3	1.253732251
ENSG00000150779	TIMM8B	1.607688566	ENSG00000150990	DHX37	1.25345747
ENSG00000144655	CSRNP1	1.60726767	ENSG00000149922	TBX6	1.25190093
ENSG00000285043		1.599392854	ENSG00000163156	SCNM1	1.251886613
ENSG00000225151	GOLGA2P7	1.597999273	ENSG00000071282	LMCD1	1.250041754
ENSG00000141744	PNMT	1.582971142	ENSG00000128564	VGF	1.248657903
ENSG00000139410	SDSL	1.574569012	ENSG00000181649	PHLDA2	1.247866223
ENSG00000015520	NPC1L1	1.574529194	ENSG00000167723	TRPV3	1.247474747
ENSG00000183696	UPP1	1.568955195	ENSG00000176933	TOB2P1	1.247406778
ENSG00000240065	PSMB9	1.566490096	ENSG00000130766	SESN2	1.244504925
ENSG00000173991	TCAP	1.562505901	ENSG00000102996	MMP15	1.242901363
ENSG00000260123	CARMAL	1.559640908	ENSG00000115661	STK16	1.24162324
ENSG00000286011		1.557062201	ENSG00000221869	CEBPD	1.240650894
ENSG00000171462	DLK2	1.548911332	ENSG00000090266	NDUFB2	1.240521029
ENSG00000147394	ZNF185	1.548890653	ENSG00000284948		1.239715943
ENSG00000270885	RASL10B	1.544851909	ENSG00000178623	GPR35	1.236796342
ENSG00000107262	BAG1	1.544810869	ENSG00000181404	WASHC1	1.235454013
ENSG00000197989	SNHG12	1.544058109	ENSG00000011028	MRC2	1.234829942
ENSG00000106565	TMEM176B	1.5394951	ENSG00000092621	PHGDH	1.231366189

ENSG00000287055		1.533234766	ENSG00000182768	NGRN	1.22941063
ENSG00000041982	TNC	1.532325852	ENSG00000104522	GFUS	1.228547145
ENSG00000205309	NT5M	1.531581687	ENSG00000171159	BBLN	1.227927321
ENSG00000125730	C3	1.519880121	ENSG00000162772	ATF3	1.227816027
ENSG00000165655	ZNF503	1.518142983	ENSG00000175634	RPS6KB2	1.22647525
ENSG00000185633	NDUFA4L2	1.508323261	ENSG00000203724	C1orf53	1.226249358
ENSG00000006327	TNFRSF12A	1.505866716	ENSG00000184281	TSSC4	1.225891236
ENSG00000205485	LINC03009	1.505811514	ENSG00000275329		1.224960262
ENSG00000275395	FCGBP	1.502611743	ENSG00000168517	HEXIM2	1.224061428
ENSG00000267372		1.489739552	ENSG00000168894	RNF181	1.223272325
ENSG00000178977	LINC00324	1.487436434	ENSG00000127399	LRRC61	1.222994489
ENSG00000108551	RASD1	1.486305782	ENSG00000272183		1.22280357
ENSG00000231107	LINC01508	1.485175346	ENSG00000132205	EMILIN2	1.222707501
ENSG00000272888	CHASERR	1.480412977	ENSG00000149557	FEZ1	1.222621941
ENSG00000188859	FAM78B	1.469816841	ENSG00000072778	ACADVL	1.220914134
ENSG00000015413	DPEP1	1.462830257	ENSG00000162144	CYB561A3	1.220420945
ENSG00000243024		1.455305831	ENSG00000226210	WASH8P	1.220057094
ENSG00000197019	SERTAD1	1.451688269	ENSG00000159069	FBXW5	1.219910756
ENSG00000226415	TPI1P1	1.451605305	ENSG00000172602	RND1	1.21866034
ENSG00000115350	POLE4	1.448984799	ENSG00000147123	NDUFB11	1.218641789
ENSG00000130529	TRPM4	1.448047322	ENSG00000197568	HHLA3	1.217003688
ENSG00000134259	NGF	1.447682615	ENSG00000283633		1.216868384
ENSG00000148290	SURF1	1.447611337	ENSG00000110080	ST3GAL4	1.215831702
ENSG00000177106	EPS8L2	1.44613012	ENSG00000146707	POMZP3	1.213759632
ENSG00000161955	TNFSF13	1.440643249	ENSG00000204394	VARS1	1.21261582
ENSG00000106404	CLDN15	1.435509439	ENSG00000149792	MRPL49	1.212007909
ENSG00000198835	GJC2	1.433417076	ENSG00000181029	TRAPPC5	1.210718686
ENSG00000149150	SLC43A1	1.430883486	ENSG00000267454	ZNF582-DT	1.209216812
ENSG00000135929	CYP27A1	1.42911985	ENSG00000227473	TSSK5P	1.208554322
ENSG00000099860	GADD45B	1.428585768	ENSG00000198727	MT-CYB	1.207758487
ENSG00000165507	DEPP1	1.421903329	ENSG00000186994	KANK3	1.205988405
ENSG00000175213	ZNF408	1.420394894	ENSG00000182902	SLC25A18	1.205880136
ENSG00000179750	APOBEC3B	1.416481017	ENSG00000147813	NAPRT	1.205760825
ENSG00000167136	ENDOG	1.413078192	ENSG00000130734	ATG4D	1.205390307
ENSG00000260456	C16orf95	1.410276479	ENSG00000120885	CLU	1.204471514
ENSG00000168748	CA7	1.4016983	ENSG00000279457	WASH9P	1.201991964
ENSG00000255837	TAS2R20	1.391438575	ENSG00000168393	DTYMK	1.201846789

List of upregulated genes in Clone3

(Adjusted p-value < 0.001 & Log2FoldChange > 1.2)

Ensemble ID	Gene Name	Log2FoldChange	Ensemble ID	Gene Name	Log2FoldChanges
ENSG00000288253		5.085674795	ENSG00000184194	GPR173	1.534414867
ENSG00000268621	IGFL2-AS1	5.061537083	ENSG00000287853		1.50976792
ENSG00000184845	DRD1	4.888988256	ENSG00000248905	FMN1	1.507213136
ENSG00000147872	PLIN2	4.364468616	ENSG00000248498	ASNSP1	1.504093265
ENSG00000268785	RPL7P50	3.272110289	ENSG00000111799	COL12A1	1.501152885
ENSG00000163520	FBLN2	3.235560961	ENSG00000196754	S100A2	1.493365799
ENSG00000005379	TSPOAP1	2.983349276	ENSG00000265485	LINC01915	1.487662985
ENSG00000167767	KRT80	2.786007834	ENSG00000275395	FCGBP	1.471573818
ENSG00000084207	GSTP1	2.77757534	ENSG00000091513	TF	1.468306849
ENSG00000088726	TMEM40	2.709616879	ENSG00000141738	GRB7	1.454090834
ENSG00000196653	ZNF502	2.683131413	ENSG00000117586	TNFSF4	1.45153018

ENSG00000197134	ZNF257	2.666220087	ENSG00000117595	IRF6	1.451059803
ENSG00000172687	ZNF738	2.621572741	ENSG00000157514	TSC22D3	1.443182143
ENSG00000231367	LINC02613	2.608492157	ENSG00000267277		1.435559611
ENSG00000254561		2.580326384	ENSG00000256229	ZNF486	1.435135646
ENSG00000267737		2.568240829	ENSG00000262712		1.433152874
ENSG00000107562	CXCL12	2.524773632	ENSG00000267454	ZNF582-DT	1.430302008
ENSG00000137142	IGFBPL1	2.509670625	ENSG00000142794	NBPF3	1.425052956
ENSG00000287407		2.453060591	ENSG00000027869	SH2D2A	1.420948528
ENSG00000182986	ZNF320	2.348021697	ENSG00000135046	ANXA1	1.419590779
ENSG00000227165	WDR11-DT	2.333801523	ENSG00000271755		1.419028908
ENSG00000180316	PNPLA1	2.29790189	ENSG00000101605	MYOM1	1.413428752
ENSG00000002933	TMEM176A	2.287788018	ENSG00000181097	LINC02878	1.412327278
ENSG00000257150	PGAM1P5	2.283791838	ENSG00000173227	SYT12	1.410743395
ENSG00000105369	CD79A	2.257507194	ENSG00000206535	LNP1	1.405412203
ENSG00000244752	CRYBB2	2.228025796	ENSG00000173535	TNFRSF10C	1.394358128
ENSG00000176046	NUPR1	2.200704188	ENSG00000160712	IL6R	1.367286832
ENSG00000128510	CPA4	2.074791472	ENSG00000165370	GPR101	1.362697672
ENSG00000170454	KRT75	2.052781409	ENSG00000110195	FOLR1	1.362673879
ENSG00000182912	TSPEAR-AS2	2.024060458	ENSG00000285043		1.356970764
ENSG00000204666		2.002232665	ENSG00000197813		1.346062812
ENSG00000268460		1.957389694	ENSG00000205143	ARID3C	1.334692431
ENSG00000092607	TBX15	1.912143116	ENSG00000240583	AQP1	1.332053465
ENSG00000249550	LINC01234	1.894626145	ENSG00000276570		1.320340258
ENSG00000138119	MYOF	1.887141664	ENSG00000134531	EMP1	1.319971641
ENSG00000142619	PADI3	1.885165051	ENSG00000130222	GADD45G	1.314935797
ENSG00000099860	GADD45B	1.868907175	ENSG00000005007	UPF1	1.304869579
ENSG00000256612	CYP2B7P	1.860134778	ENSG00000237118	CYP2F2P	1.304230998
ENSG00000060558	GNA15	1.8584263	ENSG00000259146	SIPA1L1-AS1	1.285867561
ENSG00000167157	PRRX2	1.840820444	ENSG00000165168	CYBB	1.285709515
ENSG00000162873	KLHDC8A	1.838936612	ENSG00000101115	SALL4	1.280978511
ENSG00000175697	GPR156	1.83793745	ENSG00000166394	CYB5R2	1.280320163
ENSG00000138798	EGF	1.829669566	ENSG00000104825	NFKBIB	1.275793186
ENSG00000196878	LAMB3	1.825138984	ENSG00000178623	GPR35	1.271549656
ENSG00000104413	ESRP1	1.810928436	ENSG00000051128	HOMER3	1.266189012
ENSG00000237440	ZNF737	1.757442354	ENSG00000123201	GUCY1B2	1.265181308
ENSG00000261373	VPS9D1-AS1	1.74849569	ENSG00000179761	PIPOX	1.263614956
ENSG00000121577	POPDC2	1.705280252	ENSG00000251129	LINC02506	1.258822642
ENSG00000081138	CDH7	1.668859047	ENSG00000184162	NR2C2AP	1.251149432
ENSG00000143878	RHOB	1.666816669	ENSG00000257135	ODC1-DT	1.249929703
ENSG00000283403		1.663442864	ENSG00000213626	LBH	1.248548565
ENSG00000269404	SPIB	1.663426903	ENSG00000198626	RYR2	1.248150381
ENSG00000021488	SLC7A9	1.661206372	ENSG00000064547	LPAR2	1.247175693
ENSG00000171401	KRT13	1.640503723	ENSG00000127241	MASP1	1.242434314
ENSG00000175832	ETV4	1.628111288	ENSG00000178093	TSSK6	1.237513231
ENSG00000204613	TRIM10	1.608614201	ENSG00000129214	SHBG	1.223475655
ENSG00000156966	B3GNT7	1.607265615	ENSG00000267767	LINC01801	1.221359336
ENSG00000106538	RARRES2	1.585135044	ENSG00000249069	LINC01033	1.212287574
ENSG00000084636	COL16A1	1.56761224	ENSG00000167554	ZNF610	1.21068181
ENSG00000104894	CD37	1.560312878	ENSG00000184261	KCNK12	1.206406687
ENSG00000244558	KCNK15-AS1	1.558126445	ENSG00000143184	XCL1	1.206047272
ENSG00000135346	CGA	1.544921706	ENSG00000006327	TNFRSF12A	1.203749366
ENSG00000182132	KCNIP1	1.542609658			

APPENDIX VIII

List of NMD transcripts common in Clone1 & Clone3

(Adjusted p-value < 0.05 & Log2FoldChange > |1.2|)

Ensemble ID	Gene Name	Transcript Name	Ensemble ID	Gene Name	Transcript Name
ENST00000587556			ENST00000429769	EEF1B2	EEF1B2-205
ENST00000436445	RPS9	RPS9-206	ENST00000686548	MED12	MED12-215
ENST00000510237	CEP192	CEP192-207	ENST00000423763	SUMF2	SUMF2-207
ENST00000601053	PIH1D1	PIH1D1-220	ENST00000453535	GSK3A	GSK3A-203
ENST00000543455	TMED3	TMED3-204	ENST00000403542	RBM39	RBM39-206
ENST00000547787	CERS5	CERS5-211	ENST00000420440	SIRT2	SIRT2-207
ENST00000585359	GADD45B	GADD45B-202	ENST00000676496	PHB2	PHB2-215
ENST00000509622	LARP7	LARP7-210	ENST00000509705	NFKBIB	NFKBIB-203
ENST00000415942	RBCK1	RBCK1-209	ENST00000360117	ZNF107	ZNF107-202
ENST00000542364	MORN3	MORN3-203	ENST00000676790	GSK3A	GSK3A-210
ENST00000455843			ENST00000525455	POLR2G	POLR2G-203
ENST00000573550	CDC27	CDC27-215	ENST00000575112	RPAIN	RPAIN-215
ENST00000493230	HRAS	HRAS-210	ENST00000650721	CHD1L	CHD1L-212
ENST00000547235	NPIPB12	NPIPB12-201	ENST00000423566	AP3S2	AP3S2-202
ENST00000571540	CNTROB	CNTROB-206	ENST00000683325	EYA4	EYA4-216
ENST00000601252	TIMM50	TIMM50-214	ENST00000569022	FBXO22	FBXO22-207
ENST00000464720	TMEM222	TMEM222-202	ENST00000293968	CCNF	CCNF-201
ENST00000409657	CCDC14	CCDC14-202	ENST00000649769	DHCR24	DHCR24-211
ENST00000683055	OFD1	OFD1-213	ENST00000391881	CYTH2	CYTH2-202
ENST00000491617	COMMD2	COMMD2-207	ENST00000415933	PPIA	PPIA-202
ENST00000493476	SERP2	SERP2-203	ENST00000683954	TAF1	TAF1-227
ENST00000565903	PRSS27	PRSS27-204	ENST00000547173	RPLP0	RPLP0-207
ENST00000538686	USP3	USP3-203	ENST00000518214	ZDHHC14	ZDHHC14-207
ENST00000666746	POLG	POLG-226	ENST00000649772		
ENST00000475814	ASAP3	ASAP3-205	ENST00000572919		
ENST00000565593	FBXO31	FBXO31-204	ENST00000590466	UBXN6	UBXN6-206
ENST00000485808	LRWD1	LRWD1-208	ENST00000689477	FOXRED1	FOXRED1-232
ENST00000552884	SARNP	SARNP-205	ENST00000568556	BOLA2-SMG1P6	BOLA2-SMG1P6-203
ENST00000506974	NDUFS4	NDUFS4-204	ENST00000522468	HMBOX1	HMBOX1-209
ENST00000453186	SANBR	SANBR-204	ENST00000683717	UBAP2	UBAP2-223
ENST00000436842	RNASEH1	RNASEH1-202	ENST00000449367	FIS1	FIS1-204
ENST00000481109	PRPF4B	PRPF4B-206	ENST00000676977	PPIA	PPIA-211
ENST00000582611	AKAP10	AKAP10-210	ENST00000526457	LRRK1	LRRK1-205
ENST00000527479	PTDSS2	PTDSS2-207	ENST00000552215	CHPT1	CHPT1-208
ENST00000688618	TMEM231	TMEM231-216	ENST00000334841	CMC1	CMC1-201
ENST00000592290	FXYD5	FXYD5-213	ENST00000466813	ZBTB48	ZBTB48-204
ENST00000418996	DGUOK	DGUOK-203	ENST00000648570	L3MBTL1	L3MBTL1-222
ENST00000395560	UPP1	UPP1-202	ENST00000543926	ZNF236	ZNF236-203
ENST00000594247	BABAM1	BABAM1-204	ENST00000580977	C19orf47	C19orf47-205
ENST00000379707	BAG1	BAG1-203	ENST00000678916	PTGES2	PTGES2-220
ENST00000461670	ATP6V0B	ATP6V0B-202	ENST00000465229	EXOSC3	
ENST00000586313	UBA2	UBA2-203	ENST00000699750	OGT	
ENST00000593114	RAD23A	RAD23A-210	ENST00000489394	SAT1	SAT1-208
ENST00000684388			ENST00000555726	PPP2R3C	
ENST00000600859	POLD1	POLD1-211	ENST00000528815	AP2A2	AP2A2-213
ENST00000651280	SART3	SART3-217	ENST00000440610	LMAN2L	LMAN2L-205
ENST00000546260	SOD2	SOD2-215	ENST00000367677	MRPS14	MRPS14-201

ENST00000689093	OCRL	OCRL-208	ENST00000506882	CENPC	
ENST00000432485	PRR14L	PRR14L-205	ENST00000524458	BBS1	
ENST00000491736	SLC25A4	SLC25A4-202	ENST00000495265	BRD9	
ENST00000378331	SUV39H2	SUV39H2-205	ENST00000581375	NOL11	
ENST00000684312	STAMBP	STAMBP-260	ENST00000414294	SAE1	SAE1-204
ENST00000652323	MXI1	MXI1-234	ENST00000597212	DMKN	DMKN-264
ENST00000592207	MARK4	MARK4-208	ENST00000638076	CLN6	CLN6-224
ENST00000511473	RACK1	RACK1-224	ENST00000685980	KCNA3	KCNA3-202
ENST00000578638	LLGL2	LLGL2-211	ENST00000602265	TIMM50	TIMM50-218
ENST00000460212	N6AMT1	N6AMT1-203	ENST00000643561	CISD2	CISD2-204
ENST00000465659	ARHGEF3	ARHGEF3-204	ENST00000560600	EMC9	EMC9-207
ENST00000290075	SLC25A37	SLC25A37-201	ENST00000531684	GEMIN2	GEMIN2-211
ENST00000187830	PVR	PVR-201	ENST00000570769	DNASE1	DNASE1-207
ENST00000460681	DHX34	DHX34-202	ENST00000431346	REPS1	REPS1-206
ENST00000482371	BCAP29	BCAP29-212	ENST00000453291	FBXW2	FBXW2-201
ENST00000418113	FAM161A	FAM161A-204	ENST00000467906	KLC4	KLC4-209
ENST00000595856	NDUFA7	NDUFA7-203	ENST00000649153	LRP2	LRP2-206
ENST00000584582	TOP3A	TOP3A-217	ENST00000596247	RUVBL2	RUVBL2-208
ENST00000699948	CLCN7	CLCN7-215	ENST00000511227	ACAD9	ACAD9-206
ENST00000639739	LOX	LOX-206	ENST00000674228	CAPZB	CAPZB-210
ENST00000679670	TSEN2	TSEN2-219	ENST00000163678	METTL22	METTL22-201
ENST00000652487	ALMS1	ALMS1-211	ENST00000537553	CAPRIN2	CAPRIN2-208
ENST00000388749	HNRNPL	HNRNPL-202	ENST00000447144	MTMR14	MTMR14-210
ENST00000678962	HNRNPA2B1	HNRNPA2B1-241	ENST00000489525	MDK	MDK-209
ENST00000494759	TRIM16	TRIM16-207	ENST00000447062	WDTC1	WDTC1-203
ENST00000644142	RNASEH2C	RNASEH2C-210	ENST00000393400	TPRA1	TPRA1-203
ENST00000597068	STXBP2	STXBP2-210	ENST00000523609	AGO2	AGO2-209
ENST00000330736	ANKRD11	ANKRD11-202	ENST00000495509	PHF1	PHF1-211
ENST00000563790	HDGFL3	HDGFL3-204	ENST00000649731	ZC3H14	ZC3H14-226
ENST00000451517	NDUFA3	NDUFA3-209	ENST00000569236	NSMCE1	NSMCE1-214
ENST00000425834	WDR83	WDR83-202	ENST00000468581	SLC33A1	SLC33A1-203
ENST00000504854	NDST4	NDST4-202	ENST00000700618	PLD3	PLD3-227
ENST00000472940	RPS6KB1	RPS6KB1-205	ENST00000467101	MTG2	MTG2-206
ENST00000467136	TAB3	TAB3-206	ENST00000342392	DIABLO	DIABLO-202
ENST00000574581	GLOD4	GLOD4-208	ENST00000691183		
ENST00000565687	MLST8	MLST8-222			

APPENDIX IX

List of novel PTC+ transcripts common in Clone1 & Clone3

Ensemble ID	Gene Name	Log2FoldChange	Ensemble ID	Gene Name	Log2FoldChange
ENST00000233505	CENPA	8.185067271	ENST00000554563	PPP2R3C	-0.524390459
ENST00000562170	GANC	5.874690374	ENST00000702537	RPLP0P2	-0.558054195
ENST00000463567	ZNF767P	5.638601506	ENST00000565764	TFIP11-DT	-0.592761777
ENST00000493743	SEMA3F	5.478981161	ENST00000506278	PRIMPOL	-0.639844237
ENST00000537668	ENSG00000256594	5.46512623	ENST00000511028	SPATA18	-0.725337441
ENST00000420514	FAM66C	5.148055895	ENST00000685711	FAM66C	-0.754462687
ENST00000643950	FUT1	4.805315579	ENST00000552453	ZDHHC17	-0.785595938
ENST00000515152	PRIMPOL	4.303537214	ENST00000574448	CRYM	-0.816484602
ENST00000664314	SLC16A1-AS1	4.08357009	ENST00000578055	OSBPL1A	-0.848272118
ENST00000433271	TMEM248	3.967914217	ENST00000684939	ZNF767P	-0.863766845
ENST00000508095	CENPU	3.962828244	ENST00000456135	FAM66C	-0.907900777
ENST00000535567	FAM66C	3.893143722	ENST00000546778	ZDHHC17	-0.917714392
ENST00000593239	ODAD4	3.732138463	ENST00000562859	GANC	-0.94583395
ENST00000402675	TRNT1	3.58772036	ENST00000460612	ECHDC2	-1.039470629
ENST00000690890	SCAND2P	3.256528689	ENST00000477833	ARHGEF3	-1.043503535
ENST00000498225	KIF17	2.907818248	ENST00000434583	TRNT1	-1.058481286
ENST00000588512	ZNF527	2.739543957	ENST00000462858	KIF17	-1.092906599
ENST00000414115	SHC1	2.730699188	ENST00000428411	SLC16A1-AS1	-1.107572099
ENST00000553273	PPP2R3C	2.578099151	ENST00000533947	TCIRG1	-1.212902417
ENST00000597220	FUT1	2.23001195	ENST00000465835	REV1	-1.225815801
ENST00000456950	PLEKHH3	2.1692686	ENST00000591639	ZNF506	-1.319255777
ENST00000513796	UNC5C	1.503628102	ENST00000589189.5	GALNT1	-1.389646208
ENST00000591662	CNTNAP1	1.473880031	ENST00000673460	POLG2	-1.403370398
ENST00000487741	EFCAB14	1.230009882	ENST00000530343	CAT	-1.508865104
ENST00000529657	TCIRG1	1.210590627	ENST00000450415	REV1	-1.603482085
ENST00000510042	CBR4	1.098782025	ENST00000475662	CENPA	-1.705138479
ENST00000507752	CBR4	1.068016855	ENST00000511335	MMS22L	-1.725187559
ENST00000578091	OSBPL1A	1.042712037	ENST00000481762	ZNF767P	-1.736746763
ENST00000506808	CBR4	0.884012131	ENST00000486492	ZNF767P	-1.878331317
ENST00000407797	IRF1-AS1	0.829720997	ENST00000698408	TRNT1	-1.957012221
ENST00000700580	MAP3K7	0.823099926	ENST00000438366	REV1	-1.966239185
ENST00000483428	RBM18	0.794424496	ENST00000479745	EFCAB14	-2.04445464
ENST00000685822	ENSG00000284719	0.791052519	ENST00000493198	ZNF767P	-2.146929654
ENST00000488386	CDHR3	0.710869552	ENST00000419525	CENPA	-2.251677558
ENST00000416193	SLC16A1-AS1	0.70141845	ENST00000641632	ENSG00000284719	-2.28316151
ENST00000591196	PLEKHH3	0.68492205	ENST00000658265	RIPK2-DT	-2.320267835
ENST00000470245	NAIF1	0.646933842	ENST00000560467	LDLR	-2.481529034
ENST00000534674	GLYATL1	0.643331057	ENST00000530471	SLC35C1	-2.484677593
ENST00000586801	CNTNAP1	0.619271906	ENST00000636898	CHRNA7	-2.571203342
ENST00000504145	RIPK2-DT	0.611228484	ENST00000502461	CENPU	-2.584478845
ENST00000481495	OVGP1	0.533621759	ENST00000552693	ZDHHC17	-2.59910797
ENST00000606621	TBX18	0.522107082	ENST00000524598	TCIRG1	-2.663744851
ENST00000505320	SPATA18	0.515503141	ENST00000454799	FAM66C	-2.667085145
ENST00000530449	TCIRG1	0.500343508	ENST00000335756	CENPA	-2.736449858
ENST00000541558	FAM66C	0.472286358	ENST00000654334	SLC16A1-AS1	-2.737700839
ENST00000606325	TBX18	0.399635352	ENST00000598723	ZNF611	-2.766785351

ENST00000524870	TCIRG1	0.378318644	ENST00000687204	RIPK2-DT	-2.962598101
ENST00000534673	TCIRG1	0.356742752	ENST00000443883	CACNA2D1	-3.074764792
ENST00000460030	CENPA	0.270492358	ENST00000544531	ECHDC2	-3.106550683
ENST00000650839	TRNT1	0.205513403	ENST00000533514	MSANTD2	-3.139101904
ENST00000450338	SEMA3F	0.032671292	ENST00000498026	INPP4A	-3.303713509
ENST00000687422	STAG3L1	0.022252936	ENST00000676677	KIF5C	-3.31335454
ENST00000510146	CENPU	-0.036871482	ENST00000685145	RIPK2-DT	-3.455709136
ENST00000698256	TCIRG1	-0.061985207	ENST00000555630	PPP2R3C	-3.505903783
ENST00000443905	ZNF506	-0.08507451	ENST00000580490	POLG2	-3.521416771
ENST00000501463	CCNO	-0.098595351	ENST00000414301	SEMA3F	-3.66037692
ENST00000531791	SLC25A45	-0.138384649	ENST00000529962	SLC25A45	-3.829279156
ENST00000609327	SATB1-AS1	-0.183940926	ENST00000533864	GLYATL1	-4.168758146
ENST00000474330	CDHR3	-0.207595849	ENST00000627923	SATB1-AS1	-4.466119516
ENST00000508001	PRIMPOL	-0.215258394	ENST00000481719	REV1	-4.522699074
ENST00000559340	LDLR	-0.239610358	ENST00000482887	REV1	-4.560613086
ENST00000582645	OSBPL1A	-0.27201497	ENST00000494622	OVGP1	-4.609154655
ENST00000586260	ZNF506	-0.294306436	ENST00000699719	ADCY5	-4.641070312
ENST00000472212	ZNF767P	-0.296557117	ENST00000478959	RPLP0P2	-4.653251271
ENST00000591658	ODAD4	-0.30847312	ENST00000698259	TCIRG1	-4.704010295
ENST00000493896	ECHDC2	-0.323631123	ENST00000525608	GLYATL1	-4.718664971
ENST00000496096	TIA1	-0.336662278	ENST00000485487	REV1	-4.834864559
ENST00000578997	POLG2	-0.357383479	ENST00000601931	FUT1	-4.863727466
ENST00000413697	REV1	-0.357467395	ENST00000434663	GSN	-5.039215447
ENST00000527221	FSCN2	-0.369489888	ENST00000450621	KIF5C	-5.598962445
ENST00000420168	SLC16A1-AS1	-0.370748892	ENST00000465818	VPS8	-5.688146623
ENST00000472116	CDHR3	-0.41788784	ENST00000557278	PPP2R3C	-5.808664945
ENST00000552186	CYP27B1	-0.43581482	ENST00000465086	REV1	-5.880415223
ENST00000625515	SATB1-AS1	-0.443674954	ENST00000576703	CRYM	-6.361665376
ENST00000700581	MAP3K7	-0.464276749	ENST00000531684	GEMIN2	-6.858640825
ENST00000525214	COPB1	-0.485627246	ENST00000515266	MBD4	-7.02580334
ENST00000533629	SLC25A45	-0.485836457			

APPENDIX X

List of publications

1. **Chandra P**, Kumari S, Rehman A, Anand A and Singh KK. *Impaired NMD upon UPF3B-Knockout accumulates alternatively spliced transcripts with premature termination codons.* **[Under revision]**
2. Kumari S, Rehman A, **Chandra P**, Singh KK. *Functional role of SAP18 Protein: From transcriptional repression to splicing regulation.* Cell Biochemistry & Function (https://doi.org/10.1002/cbf.3830) 2023 July 24.
3. Deka B, **Chandra P**, Yadav P, Rehman A, Kumari S, Kunnumakkara AB, Singh KK. *RNPS1 functions as an oncogenic splicing factor in cervical cancer cells.* IUBMB Life (doi.org/10.1002/iub.2686) 2022 October 27.
4. **Chandra P**, Deka B, Kumari S, Rehman A, Yadav P and Singh KK. *CRISPR/Cas9-Mediated Gene-Knockout of UPF3B Alters Expression of Cell Cycle and Neuron-Specific Genes.* Preprints (doi: 10.20944/preprints202111.0013.v1) 2021 November 1.
5. Rehman A, **Chandra P**, Singh KK. *The MAGOH paralogs - MAGOH, MAGOHB and their multiple isoforms.* Gene Reports (doi.org/10.1016/j.genrep.2021.101214) 2021 September.
6. **Chandra P***, Deka B*, Singh KK. *Functional roles of human Up-frameshift suppressor 3 (UPF3) proteins: From nonsense-mediated mRNA decay to neurodevelopmental disorders.* Biochimie (doi.org/10.1016/j.biochi.2020.10.11) 2020 Oct 24. [*Equal contribution]
7. **Chandra P,** Roy S, Rehman A, Singh KK. *A minigene-based approach to identify the cis-element that facilitates the exon inclusion in long-isoform of UPF3B*. **[In preparation]**

List of conferences and workshops

Poster presentations at conferences

1. **Chandra P**, Deka B, Kumari S, Singh KK. *Transcriptome-wide analysis of UPF3B-KO HEK-293 cells generated by CRISPR/Cas9 gene-editing method.* 11th RNA Group Meeting organized by National Centre for Cell Science, Pune during December 1-3, 2022.

2. **Chandra P**, Deka B, Kumari S, Rehman A, Yadav P, Singh KK. *UPF3B-KO in HEK-293 cells delays cell proliferation and deregulates the expression of neuron-specific genes.* North-East Research Conclave: Sustainable Science and Technology organized by Indian Institute of Technology, Guwahati during May 20-22, 2022.
3. **Chandra P**, Deka B, Kumari S, Rehman A, Yadav P, Singh KK. *CRISPR/Cas9-mediated gene-knockout of UPF3B in HEK-293 cells alters the expression of cell cycle and axon guidance genes.* Systems Biology: Global Regulation of Gene Expression organized by Cold Spring Harbor Laboratory (CSHL) during March 9-12, 2022 in virtual mode.
4. **Chandra P**, Deka B, Kumari S, Singh KK. *CRISPR/Cas9-mediated gene-knockout of UPF3B in HEK-293 cells alters the expression of cell-cycle and axon guidance genes.* CRISPR/Cas: From Biology to Technology organized by the Institute of Bioinformatics and Applied Biology (IBAB) and SRM University during November 25-27, 2021 in virtual mode. **[Best Poster Award]**

Conferences attended

1. **Chandra P**. Participated in "2nd Departmental Retreat (Biotech Express)" organized by the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati. 2021 August 21 (virtual mode).
2. **Chandra P**. Participated in "1st Departmental Retreat (Biotech Express)" organized by the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati. 2019 December 17.

Workshop attended

1. **Chandra P.** Participated in three-day workshop on "Genome/Transcriptome Sequence Analysis" organized by Accelerator program for Discovery in Brain disorders using Stem cells (ADBS) and Institute of Bioinformatics and Applied Biotechnology (IBAB) during November 13-15, 2019, at NCBS, Bangalore.