A Study of Nonsense Mediated mRNA Decay Using Naturally Occurring Genetic Variants and Through the Development of a Synthetic Reporter Transgene

Deepti Dianna Domingo
BSc (Molecular Biology)

Neurogenetics Research Group
The University of Adelaide

Thesis submitted for the degree of
Doctor of Philosophy

In the

Department of Molecular and Biomedical Sciences
School of Biological Sciences
Faculty of Science
The University of Adelaide

November 2019

Table of contents

Abstract

The nonsense mediated mRNA decay pathway (NMD) plays an important role in normal brain development. Genetic variation which disrupts genes encoding key NMD pathway members are implicated in neurodevelopmental disorders such as intellectual disability and autism. The mechanism by which deficient NMD results in neurodevelopmental dysfunction, however, remains unknown. Recently, NMD activity has been recognised to vary across cell types, tissue types and even display inter-individual variability. Yet current methods to quantify NMD activity rely on average cell population measurements and thus lack the resolution needed to capture dynamic changes resulting from the cell and tissue heterogeneity of NMD, as well as its developmental complexity. This thesis aims to further understand NMD by, (1) investigating naturally occurring genetic variants which cause neurodevelopmental disorders and (2) by the development of a synthetic NMD reporter transgene with single cell resolution.

As part of this thesis three novel variants in genes encoding NMD factors which were identified in patients with neurodevelopmental disorders were characterised. The first of these variants was a synonymous single nucleotide variant (SNV) found in a canonical splice region of *UPF3B*. This variant was originally classified as a variant of unknown significance (VUS) and as such overlooked regarding pathogenicity. Molecular investigations in this thesis were able to conclusively resolve this variant as being pathogenic and facilitate patient diagnosis. The remaining two variants were identified within *UPF2*, the first was a novel frameshift variant, which is one of only two SNVs identified to exclusively disrupt *UPF2*. The second was a large copy number variant (CNV) which resulted in the heterozygous deletion of *UPF2* alongside 21 other genes. Investigations into the pathogenicity of these variants supported the involvement of UPF2 in a spectrum of neurodevelopmental disorders which has been concluded from previous studies where *UPF2* has been disrupted by large CNV deletions.

Within this thesis two versions of a fluorescent NMD reporter transgene which can measure NMD activity at a single cell level were also designed. Both transgenes are composed of a number of expression cassettes in 'cis'. The most important of these are the Selection, Control and NMD cassettes. The Control and NMD cassettes co-express distinguishable fluorescent proteins allowing for visual and quantitative real-time output of NMD activity. The Selection cassette enables recombination mediated cassette exchange to take place, allowing the entire transgene to be introduced into the *Col1a1* locus of germ-line competent transgenic mouse embryonic stem cells (mESCs) or transgenic mouse zygotes.

In this thesis I have developed and used experimental pipelines to test the responsiveness of the designed NMD reporter transgenes to NMD inhibition *in vitro*. Unfortunately, following integration into mESCs neither version of the NMD reporter transgene was completely responsive to changes in cellular NMD activity. One version, however, was used to establish a stable and functional NMD reporter HEK293T cell line. These cells can facilitate high-throughput screening tests for drugs or small compounds which alter NMD activity to drive the development of therapeutics or benefit research.

Once the design of an NMD reporter transgene is perfected for use in mESCs or a transgenic mouse line, this technology will provide visual and quantitative tracking of endogenous NMD activity at a single cell level. A possible immediate application would be to track NMD activity across embryonic brain development and into postnatal life. By providing a means to define regions or cell types in the brain most affected by malfunctioning NMD, e.g. due to heritable DNA mutations, the underlying mechanism by which deficient NMD leads to neurodevelopmental dysfunction can be further elucidated. This will support the development and assessment of more targeted therapies for individuals affected with neurodevelopmental disorders due to NMD disrupting genetic variants.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Deepti Dianna Domingo

Bachelor of Science

Student Number: a1628402

DATE 15/11/19

Acknowledgments

I feel very fortunate to have been able to work, learn, and participate in the exciting research at the Neurogenetics Laboratory. I would like to thank my supervisors Professor Jozef Gecz and Dr Lachlan Jolly for providing me with this opportunity and for their ongoing advice and guidance.

I would like to extend my thanks to; Professor Murray Whitelaw and Dr David Bersten for providing me with the transgenic feeder-dependent FLP-in mESC line. Randal Grose and Sophie Watts at the SAHMRI flow and laser scanning cytometry facility for donating their time and expertise to conduct flow cytometry experiments. The clinicians and genetic counsellors for coordinating the genetic material needed for this work, and the patients and their families for giving consent and for donating their unique genetic material.

I would also like to thank past and present members of the Neurogenetics Laboratory, my friends, and my work family at Terry White Chemmart–Rundle mall. A special thanks to; Claire, Tahlia, Matilda, Debrah, Ollie, Thessa, Thuong, Marlie, Sayaka, Atma, Tessa, Renee.S, Renee.C, Clare, Raman and Jean for all the lab/life advice and coffee/wine/food/dancing dates. Without you all to celebrate with (or cry to) it would have been a very empty journey.

A huge thankyou to my parents, Vinutha and Raymond Domingo for always supporting me throughout my studies, whether it be through encouraging words, listening to me complain or providing an abundance of tasty food. Your ongoing love and support have made this thesis possible. And last, but by no means least, thank you to my wonderful partner, Cameron Parker for being on the front line all these years, you went above and beyond to support, motivate and keep me sane. I feel incredibly lucky to have had you in my life.

Abbreviations

Standard terms

%	Percentage
°C	Degrees Celsius
µF	Microfarad
µl	Microlitre
µM	Micromolar
bp	Base pair
cDNA	Complimentary DNA
DNA	Deoxyribonucleic acid
gDNA	Genomic DNA
hnRNP	Heterogenous nuclear ribonucleoprotein
kb	Kilobase
kDa	Kilodalton
M	Molar
Mg	Milligram
miRNA	MicroRNA
mL	Millilitre
mM	Millimolar
mRNA	Messenger RNA
ng	Nanogram
nm	Nanometer
nts	Nucleotides
P	Statistical significance value
RNA	Ribonucleic acid

RPM	Revolutions per minute
siRNA	Small interfering RNA
tRNA	Transfer RNA
x g	Times gravity

Materials and Methods

BSA	Bovine serum albumin
CHX	Cycloheximide
DAPI	4, 6-diaminodino-2-phenylindole
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleotide
DTT	Dithiothreitol
EB	Embryoid body
EDTA	Ethylene diamine tetraacetic acid
FBS	Fetal bovine serum
HEK293T	Human embryonic kidney 293T cell line
HeLa	Henrietta Lack's cervical cancer cell line
HRP	Horse radish peroxidase
LB	Luria broth
MEF	Mouse embryonic fibroblast cell line
MG132	Carbobenzoxy-Leu-Leu-leucinal
NIH3T3	Primary mouse embryonic fibroblast cell line
PAGE	Poly-acrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline with Tween 20
PCR	Polymerase chain reaction

PLL	Poly-L-lysine
PLL/L	Poly-L-lysine/laminin
PMSF	Phenylmethylsulfonyl fluoride
qPCR	Quantitative PCR
RPMI-1640	Roswell Park Memorial Institute 1640 Medium
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulphate
TBE	Tris borate EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween 20
UV	Ultraviolet

Non-standard terms

ADHD	Attention deficit hyperactivity disorder
AS-NMD	Alternative splicing coupled to nonsense mediated mRNA decay
ATP	Adenosine triphosphate
BMD	Becker muscular dystrophy
CBC	Cap binding complex (CBP20-CBP80)
CF	Cystic fibrosis
CFP	Cyan fluorescent protein
CNV	Copy number variant
CRISPR	Clusters of regularly interspaced short palindromic repeats
DMD	Duchenne muscular dystrophy
EGFP	Enhanced green fluorescent protein
EJC	Exon junction complex
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation

ESC	Embryonic stem cell
GTP	Guanosine-5'-triphosphate
HGVS	Human Genome Variation Society
ID	Intellectual disability
Indels	Insertions and deletions
ISR	Integrated stress response
LCL	Lymphoblast cell line
MCDS	Metaphyseal chondrodysplasia, Schmid type
mESC	Mouse embryonic stem cell
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
NLS	Nuclear localisation signal
NMD	Nonsense mediated mRNA decay
ORF	Open reading frame
polyA	Polyadenylation sequence
PTC	Premature termination codon
RMCE	Recombination mediated cassette exchange
SMD	Staufen-mediated decay
SNP	Single nucleotide polymorphism
SR	Serine/arginine rich
SURF	SMG1, UPF1 and ERF1-ERF3 complex
SV40	Simian virus
TAR	Thrombocytopenia with absent radius
TetO	Tetracycline operon
TetR	Tetracycline repressor
uORF	Upstream open reading frame
UPR	Unfolded protein response

UTR	Untranslated region
VUS	Variant of unknown significance
WES	Whole exome sequencing
YFP	Yellow fluorescent protein

Proteins and genes

To facilitate identification of species, human proteins are capitalised, while mouse proteins will only have the first letter capitalised. Genes follow the same nomenclature but are italicised.

ACTB	Actin beta
ARPP21	cAMP regulated phosphoprotein 21
ATF4	Activating transcription factor 4
BRCA1	Breast cancer type 1 susceptibility protein
BRCA2	Breast cancer type 2 susceptibility protein
CASC3	Cancer susceptibility candidate 3
CBP20	CAP-binding protein 20
CBP80	CAP-binding protein 80
CFTR	CF transmembrane conductance regulator
CLN1	Cyclin CLN1
COL10A1	Collagen type X alpha 1 chain
eIF2α	Eukaryotic translation initiation factor 2 alpha
EIF3A	Eukaryotic translation initiation factor 3 subunit A
EIF4A2	Eukaryotic translation initiation factor 4A2
EIF4A3	Eukaryotic translation initiation factor 4A3
EIF4E	Eukaryotic translation initiation factor 4E
ERF1	Eukaryotic translation release factor 1, also known as ETF1
ERF3	Eukaryotic translation release factor 3, also known as GSPT1

GADD45B	Growth arrest and DNA damage inducible beta
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAS5	Growth arrest specific 5
GPR101	G protein-coupled receptor 101
HBB	Haemoglobin subunit beta
HPRT	Hypoxanthine guanine phosphoribosyl transferase
KDM5C	Lysine demethylase 5C
LMNA	Lamin A/C
MST1R	Macrophage stimulating 1 factor
MAGOH	Mago-Nashi homolog, proliferation association
MEN1	Menin 1
MUP	Major urinary protein
NUMB	NUMB endocytic adaptor protein
P53	Tumour protein p53
PABPC1	Poly(A) binding protein cytoplasmic 1
PABPN1	Poly(A) binding protein nuclear 1
PAX6	Paired box protein 6
PP2A	Protein phosphatase 2A
RBM8A	RNA binding motif protein 8A
RNPS1	RNA binding protein with serine rich domain 1
ROBO3	Roundabout guidance receptor 3
SMG1-7	Suppressor of morphological defects of genitalia 1-7
SNORD22	Small nucleolar RNA, C/D box 22
TCR-β	T cell receptor beta chain
TGF-β	Transforming growth factor beta
TPI	Triosephosphate isomerase
UPF1	Up-frameshift suppressor 1 homolog

UPF2	Up-frameshift suppressor 2 homolog
UPF3A	Up-frameshift suppressor 3 homolog A
UPF3B	Up-frameshift suppressor 3 homolog B
XRN1	5'-3' exoribonuclease 1

Species

C. elegans	*Caenorhabditis elegans*
D. melanogaster	*Drosophila melanogaster*
D. rerio	*Danio rerio*
E. coli	*Escherichia coli*
H. sapiens	*Homo sapiens*
S. cerevisiae	*Saccharomyces cerevisiae*

Chapter One: Introduction

1.1 Overview

Nonsense mediated mRNA decay (NMD) is a highly conserved eukaryotic post-transcriptional regulatory pathway. This pathway canonically protects the cell from transcriptome infidelity through identification and degradation of aberrant transcripts with a premature termination codon (PTC). These PTCs can be formed by transcription error or through genetic mutation making NMD a clinically relevant pathway for disorders caused by PTC-type genetic mutations. In such disorders, NMD activity on one hand can have a protective effect by restricting production of harmful, dominant negative truncated proteins, but on other hand, can have an aggravating effect by degrading mRNA encoding truncated proteins which may in fact maintain residual function.

Approximately 5–15% of the normal eukaryotic transcriptome is also sensitive to NMD inhibition (Adachi et al. 2004; McIlwain et al. 2010; Mendell et al. 2004; Nguyen et al. 2012; Weischenfeldt et al. 2012; Yepiskoposyan et al. 2011) and we now know that this impact NMD has on transcriptomes is important for a myriad of cellular processes including the stress response and neuronal differentiation (Bruno et al. 2011; Karam et al. 2015; Karam & Wilkinson 2012; Martin & Gardner 2015). Further highlighting the importance of NMD in the regulation of normal physiology, are the findings that loss of function mutations or copy number variants (CNVs) in core components of this pathway are linked to human genetic disease such as, various cancers (Liu et al. 2014; Lu et al. 2016) and neurodevelopmental disorders (Laumonnier et al. 2010; Linder et al. 2015; Nguyen et al. 2013; Shaheen et al. 2016).

With roles in both degradation of mRNA transcripts featuring aberrant PTCs, and regulation of normal endogenous mRNA transcripts, it is apparent that variations in NMD efficiency can influence and even drive, both disease outcome and normal cell, tissue and organism functions (Miller & Pearce 2014). Relating to disease outcome, manipulation of NMD efficiency has already received much attention as a strategy to develop personalised therapies in cases

involving PTC-type mutations. However, little attention has been dedicated to how NMD controls normal physiology, for example how NMD activity might vary between different cells and tissues, and how such variations may be employed by organisms to coordinate physiological processes.

In this chapter, I will first detail the mechanism of mammalian NMD. I will then elaborate on the complexity and variability of the NMD pathway and its involvement in genetic disease. This will provide a foundation for discussion into the currently used methods to measure NMD activity in cells and the need to derive renewed tools and methods which can be robustly used to interrogate endogenous NMD activity in cells, tissues and organisms.

1.2 The mechanism of NMD

1.2.1 NMD factors

NMD is orchestrated by many proteins acting in defined complexes (Table 1.1). Genetic screens in *Saccharomyces cerevisiae (S.cerevisiae)* first discovered three genes encoding NMD core factors, which upon mutation reduced the rate of decay of mRNA transcripts containing a PTC. As mutations in these genes promoted read-through at the PTC, they were designated up-frameshift (*UPF*) 1, 2 and 3 (Leeds et al. 1992) and to date constitute the evolutionary conserved set of core NMD factors. Similar studies in *Caenorhabditis elegans (C.elegans)* identified orthologs of the *upf1–3* genes (*smg2–smg4*) alongside five other factors which stabilised PTC containing transcripts. Moreover, mutations in these genes also caused morphological defects in the male bursa or hermaphrodite vulva and as such, were designated suppressor of morphological defects on genitalia (*smg1–smg7*) (Cali et al. 1999; Pulak & Anderson 1993). Homologs of *smg1*, *smg5*, *smg6* and *smg7* are only found in metazoans.

In humans, orthologs of all these factors have been identified through sequence conservation studies, and additional NMD factors have been discovered via the findings of genetic and

biochemical assays (Applequist et al. 1997; Denning et al. 2001; Lykke-Andersen et al. 2000; Mendell et al. 2000; Ohnishi et al. 2003; Perlick et al. 1996; Serin et al. 2001; Yamashita et al. 2009; Yamashita et al. 2001). Many of these factors are not restricted to NMD and participate in a range of important biological processes, such as telomere maintenance, cell cycle progression and staufen-mediated mRNA decay (SMD).

Table 1.1: Characteristics of NMD factors and EJC components

Group	Protein	Molecular function	Cellular location	Role in NMD	References
NMD factors	UPF1	RNA helicase, ATPase.	Steady state cytoplasmic with nucleocytoplasmic shuttling.	Essential NMD factor that associates the PTC recognition machinery with the mRNA degradation machinery.	(Applequist et al. 1997; Lykke-Andersen, Shu & Steitz 2000; Ohnishi et al. 2003)
	UPF2	RNA binding, telomeric DNA binding.	Cytoplasmic (perinuclear).	Links UPF1 and UPF3B. In cooperation with UPF3B stimulates UPF1 helicase and ATPase activity to activate NMD.	(Gehring et al. 2005; Kashima et al. 2006; Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000; Serin et al. 2001)
	UPF3A	RNA and protein binding.	Nucleocytoplasmic shuttling.	Links UPF1 and UPF2 to the EJC. In cooperation with UPF2 stimulates UPF1 helicase and ATPase activity to activate weak	(Kim, Kataoka & Dreyfuss 2001; Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000; Serin et al. 2001)

				NMD (compared to that activated by UPF3B).	
	UPF3B	RNA and protein binding.	Steady state nuclear with nucleocytoplasmic shuttling.	Links UPF1 and UPF2 to the EJC. In cooperation with UPF2 stimulates UPF1 helicase and ATPase activity to activate NMD.	(Gehring et al. 2005; Gehring et al. 2003; Kunz et al. 2006; Lykke-Andersen, Shu & Steitz 2000; Ohnishi et al. 2003; Serin et al. 2001)
	SMG1	Serine/threonine protein kinase.	Cytoplasmic and nuclear.	Phosphorylates the N- and C-terminus of UPF1 at various sites.	(Denning et al. 2001; Usuki et al. 2006; Yamashita et al. 2001)
	SMG5/7 complex	Protein binding.	Steady state cytoplasmic with nucleocytoplasmic shuttling.	Interacts with protein phosphatase 2A (PP2A) to promote UPF1 dephosphorylation. Recruits factors needed for exonucleolytic degradation of target mRNA.	(Anders et al. 2003; Gatfield et al. 2003; Ohnishi et al. 2003; Yamashita et al. 2001)

	SMG6	DNA binding, endonuclease, hydrolase, nuclease.	Steady state cytoplasmic with nucleocytoplasmic shuttling.	Interacts with PP2A to promote UPF1 dephosphorylation. Executes endonucleolytic cleavage of target mRNA.	(Fukuhara et al. 2005; Gatfield et al. 2003; Unterholzner & Izaurralde 2004)
	SMG8	Protein binding.	Cytoplasmic and nuclear.	Regulates SMG1 kinase activity by inducing inactivating conformational changes in SMG1.	(Fukuhara et al. 2005; Gatfield et al. 2003; Unterholzner & Izaurralde 2004)
	SMG9	Protein binding.	Cytoplasmic and nuclear.	Regulates SMG1 kinase activity by facilitating efficient association between SMG1 and SMG8.	(Arias-Palomo et al. 2011)
EJC subunits	RBM8A – MAGOH (heterodimer)	RNA and protein binding.	Nucleocytoplasmic shuttling.	Core EJC components which bind mRNA in a splicing-dependent, but sequence-independent, manner. Inhibits ATPase activity of EIFA3	(Gehring et al. 2005; Gehring et al. 2003; Le Hir et al. 2000)

				securing a stable ATP bound EJC to spliced mRNA.	
	EIF4A3	ATP-dependent RNA helicase.	Nucleocytoplasmic shuttling.	Core EJC component which anchors the other EJC proteins to the mRNA.	(Ferraiuolo et al. 2004; Gehring et al. 2005; Lykke-Andersen et al. 2001; Shibuya et al. 2004)
	CASC3	RNA and protein binding.	Nucleocytoplasmic shuttling.	Induces RNA-dependent ATPase and RNA-helicase activity of EIF4A3. Involved in NMD of a small selection of mRNAs. When bound to the EJC, CASC3 is thought to slow down NMD.	(Gehring et al. 2005; Noble & Song 2007; Singh et al. 2008)
	RNPS1	RNA binding.	Nucleocytoplasmic shuttling.	Involved in UPF2-dependent NMD. Cellular concentrations of RNPS1 can modify NMD efficiency.	(Gehring et al. 2005; Le Hir et al. 2000; Lykke-Andersen, Shu & Steitz 2001)

1.2.2 The exon junction complex (EJC) model of NMD

1.2.2.1 The role of the EJC

NMD is a translation dependent process (Belgrader et al. 1993; Carter et al. 1996; Thermann et al. 1998). Despite extensive research over the past three decades, the exact mechanism of NMD remains uncertain, however, the generally accepted model is the exon junction complex (EJC) model summarised in Figure 1.1.

The EJC is a highly dynamic multi-protein complex which at any one time consists of at least 10 proteins, including the core proteins; RBM8A, MAGOH and EIF4A3 (Andersen et al. 2006; Bono et al. 2006). EJCs are deposited by the spliceosome on newly synthesised mRNAs 20-24 nts upstream of the exon-exon junctions. Here the complex can act as a platform for many other factors, including the NMD components UPF2 and UPF3 (Kervestin & Jacobson 2012; Le Hir et al. 2016). During a pioneer round of translation, the ribosome will dislodge these EJCs. For most endogenous transcripts, the termination codon is encoded by the last exon, and as such, by the time the pioneer ribosome encounters the endogenous termination codon all EJCs would have been removed from the transcript. In the EJC model of NMD, it is the presence or absence of an EJC downstream of a termination codon that allows the NMD machinery to differentiate between a physiological termination codon (typically without any downstream EJCs), and a PTC (defined by the presence of a downstream EJC(s)) (Brogna & Wen 2009; Nagy & Maquat 1998). Because of the ribosomal footprint, only PTCs more than 55 nts upstream of an exon-exon junction are recognised (if they are less than 55 nts, the downstream EJC will likely be dislodged by the ribosome, preventing recognition by NMD machinery). This is known as the '55 nt rule of NMD'.

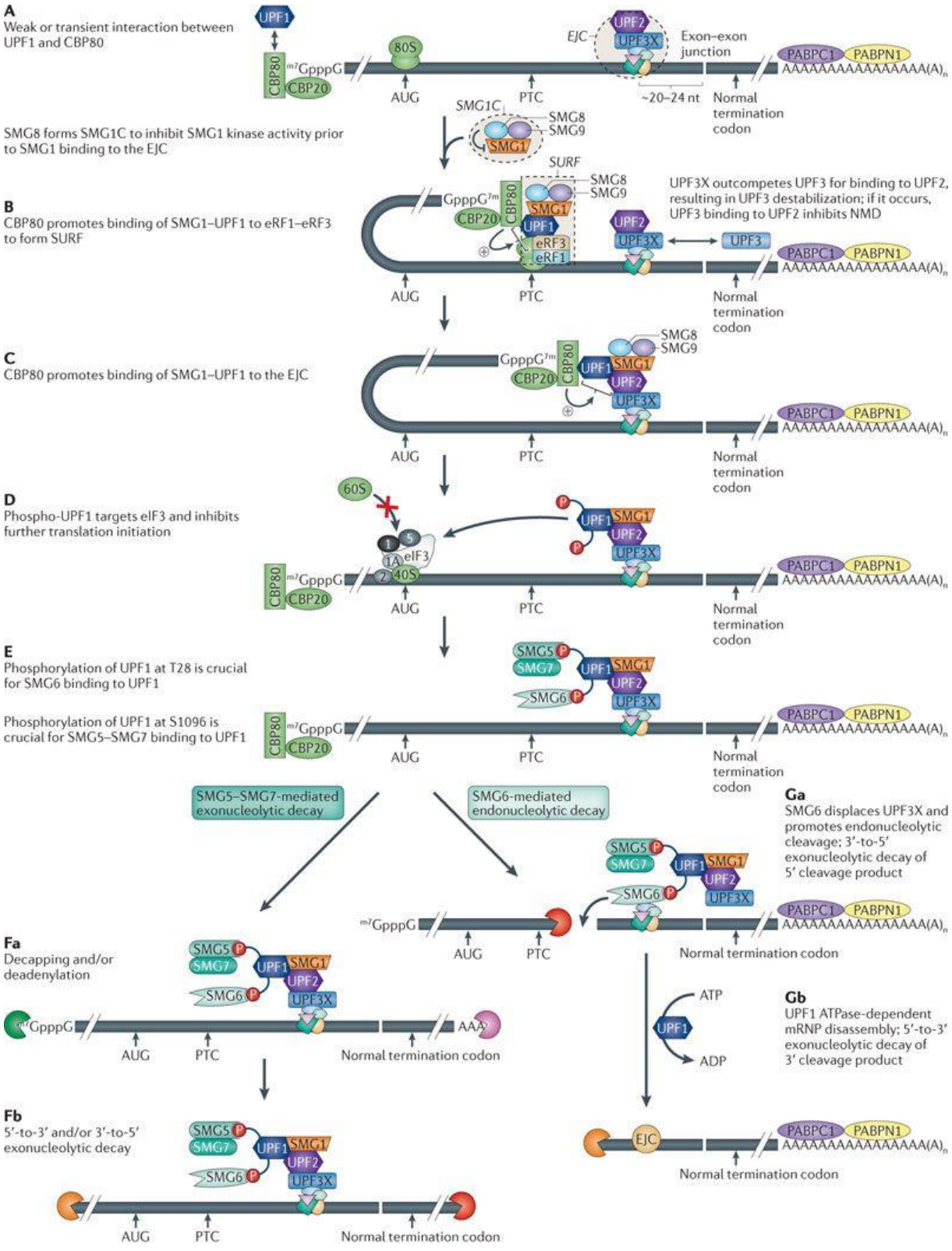

A
Weak or transient interaction between UPF1 and CBP80
UPF1
CBP80
m7GpppG
CBP20
80S
AUG
PTC
EJC
UPF2
UPF3X
Exon–exon junction
~20–24 nt
PABPC1
PABPN1
AAAAAAAAAAAAAAAA(A)n
Normal termination codon
SMG1C
SMG8
SMG9
SMG1
SMG8 forms SMG1C to inhibit SMG1 kinase activity prior to SMG1 binding to the EJC
SURF
B
CBP80 promotes binding of SMG1–UPF1 to eRF1–eRF3 to form SURF
GpppG7m
UPF1
eRF3
eRF1
UPF3
UPF3X outcompetes UPF3 for binding to UPF2, resulting in UPF3 destabilization; if it occurs, UPF3 binding to UPF2 inhibits NMD
C
CBP80 promotes binding of SMG1–UPF1 to the EJC
D
Phospho-UPF1 targets eIF3 and inhibits further translation initiation
60S
eIF3
40S
E
Phosphorylation of UPF1 at T28 is crucial for SMG6 binding to UPF1
Phosphorylation of UPF1 at S1096 is crucial for SMG5–SMG7 binding to UPF1
SMG5
SMG7
SMG6
SMG5–SMG7-mediated exonucleolytic decay
SMG6-mediated endonucleolytic decay
Ga
SMG6 displaces UPF3X and promotes endonucleolytic cleavage; 3′-to-5′ exonucleolytic decay of 5′ cleavage product
Fa
Decapping and/or deadenylation
AAA
Fb
5′-to-3′ and/or 3′-to-5′ exonucleolytic decay
ATP
ADP
Gb
UPF1 ATPase-dependent mRNP disassembly; 5′-to-3′ exonucleolytic decay of 3′ cleavage product
EJC

Figure 1.1: Step-wise degradation of a PTC containing transcript by NMD as per the EJC model (taken from (Schoenberg & Maquat 2012)).

NMD in mammalian cells occurs during the pioneer round of translation and targets newly synthesised mRNAs that are capped by the CBP20-CBP80 complexes (Ishigaki et al. 2001; Lejeune et al. 2002) and contain at least one EJC (Zhang et al. 1998).

(A) Various complexes are involved in facilitating NMD. Firstly, UPF2 and UPF3B are associated with the EJC (Le Hir et al. 2000) which serves as a signal to distinguish a physiological termination codon from a PTC (Kim et al. 2001; Le Hir et al. 2001; Le Hir et al. 2000). Secondly, UPF1 is weakly associated with the CBP20-CBP80 complex and acts to promote the binding of UPF1 and SMG1 to the SURF complex in subsequent steps (Hwang & Maquat 2011). Finally, SMG8 and SMG9 suppress the function of SMG1 in the SMG1-complex, preventing SMG1 from phosphorylating UPF1 before being properly associated with the EJC.

(B) When a ribosome terminates sufficiently upstream of an EJC due to the presence of a PTC, the CBP20-CBP80 complex facilitates SMG1, UPF1 and ERF1-ERF3 to form what is known as the SURF complex (Kashima et al. 2006).

(C) The CBP20-CBP80 complex assists the SURF complex to join directly to the EJC through the bridge formed by UPF1, UPF2 and UPF3 proteins (Chamieh et al. 2008). This facilitates SMG1 mediated phosphorylation of UPF1 (Kashima et al. 2006).

(D) Phosphorylated UPF1 binds directly to EIF3 which prevents the formation of the translational initiation complex, thus, inhibits further translation of the PTC containing transcript (Isken & Maquat 2008).

Figure 1.1 continued…

(E) The phosphorylated sites of UPF1 serve as platforms to recruit SMG6 (to site p.T28) and the SMG5-SMG7 complex (to site p.S1096), causing the ribosome and ERF1-ERF3 to be displaced from the mRNA (Ohnishi et al. 2003; Okada-Katsuhata et al. 2012). The mRNA is then subjected to either **(F)** exonucleoytic decay mediated by the SMG5-SMG7 complex (Unterholzner & Izaurralde 2004) or **(G)** endonucleolytic decay mediated by SMG6 (Eberle et al. 2009; Huntzinger et al. 2008).

(F) Exonucleolytic decay of the PTC containing transcript is initiated by **(Fa)** decapping and deadenylation, followed by **(Fb)** degradation from both 5' and 3' ends (Lejeune et al. 2003).

(G) Endnucleolytic decay of the PTC containing transcript results in a 5' and 3' mRNA product which are respectively degraded by **(Ga)** exonucleases (Eberle et al. 2009; Huntzinger et al. 2008) or **(Gb)** XRN1 mediated degradation via dephosphoryalation of UPF1 and dissociation of the protein complex from the 3' cleaved product (Franks et al. 2010).

1.2.2.2 Translation termination

Following splicing, newly synthesised messenger ribonucleoproteins (mRNPs) bound by the cap-binding protein heterodimer CBP20-CBP80 (CBC) are exported to the cytoplasm where they undergo the first 'quality control' pioneer rounds of translation, which may even begin as they exit the nuclear pore (Ishigaki et al. 2001; Lejeune et al. 2002; Maquat 2004). During this process EJCs are displaced as the ribosome moves 5'–3' along the transcript (Isken & Maquat 2008; Maquat et al. 2010). Normally a ribosome will encounter a termination codon in the proximity of the PABPN1 protein and with no downstream EJCs present. This allows PABPN1 to facilitate recruitment of the eukaryotic release factor 1 (ERF1) in a complex with GTPase and GTP bound ERF3 to the ribosome. ERF1 recognises the termination codon, triggering GTP hydrolysis followed by a conformational change of ERF1 which promotes dissociation of the nascent polypeptide and release of ribosomal subunits i.e. translation termination (Dever & Green 2012; Frolova et al. 1996; Jackson et al. 2010).

A successful pioneer round of translation will end with efficient translation termination, promoting exchange of nuclear PABPN1 with cytoplasmic PABPC1 at the poly(A) tail. Simultaneously, but in a translation-independent manner the CBC is also replaced with the eukaryotic translation initiation factor 4E (EIF4E) (Sato & Maquat 2009). Remodelling of CBC-bound mRNP to EIF4E-bound mRNP allows EIF4E to interact with PABPC1 to form a stable closed loop mRNA which facilitates steady state translation. Translation termination is now led by PABPC1 recruited ERF1, and the close proximity of the mRNA ends in the closed loop structure allows efficient recycling and re-initiation of the ribosome at the start codon (Dias et al. 2009; Gorlich et al. 1996; Sato & Maquat 2009; Wells et al. 1998).

1.2.2.3 Premature translation termination

It was initially proposed that NMD occurs during the pioneer round of translation by interacting exclusively with CBC-bound mRNPs (Isken & Maquat 2008; Maquat, Tarn & Isken 2010). This model is supported by the fact that EJC factors were readily detected in CBC-mRNP immunoprecipitation assays but not in EIF4E-mRNP immunoprecipitation assays (Hosoda et al. 2005; Ishigaki et al. 2001; Lejeune et al. 2002). It has now been shown that NMD machinery can also target EIF4E-bound mRNPs (Durand & Lykke-Andersen 2013; Rufener & Muhlemann 2013) meaning NMD via the EJC model is not limited to the pioneer rounds of translation and can occur during any round of translation. By extension, this means that the EJCs downstream of a PTC are not dislodged during a pioneer round of translation and remain assembled on the mRNA during bulk translation with the potential to initiate NMD (Figure 1.2).

Regardless of the type of mRNP being translated, if the first termination codon encountered by a ribosome is premature i.e. a PTC, translation termination is often less efficient due to ribosomal stalling at said PTC (Peixeiro et al. 2012). This interruption provides the core NMD factor, UPF1 instead of PABPC1 with the opportunity to bind ERF3 thus preventing normal translation termination and re-initiation (Ivanov et al. 2008; Singh, Rebbapragada & Lykke-Andersen 2008; Stalder & Muhlemann 2008). Moreover, binding of UPF1 to ERF3 triggers the formation of the SMG-1-UPF1-ERF1-ERF3 (SURF) complex (Kashima et al. 2006). If there is a recognisable downstream EJC, UPF1 will interact with UPF3B bound UPF2, bridging the SURF complex to the EJC and triggering SMG1 kinase-mediated UPF1 phosphorylation which subsequently releases ERF1 and ERF3 (Chamieh et al. 2008). Also, through these phosphorylation sites the resulting complex recruits SMG6 for endonucleolytic decay of the PTC containing transcript and the SMG5-SMG7 complex for exonucleolytic decay (Eberle et al. 2009; Huntzinger et al. 2008).

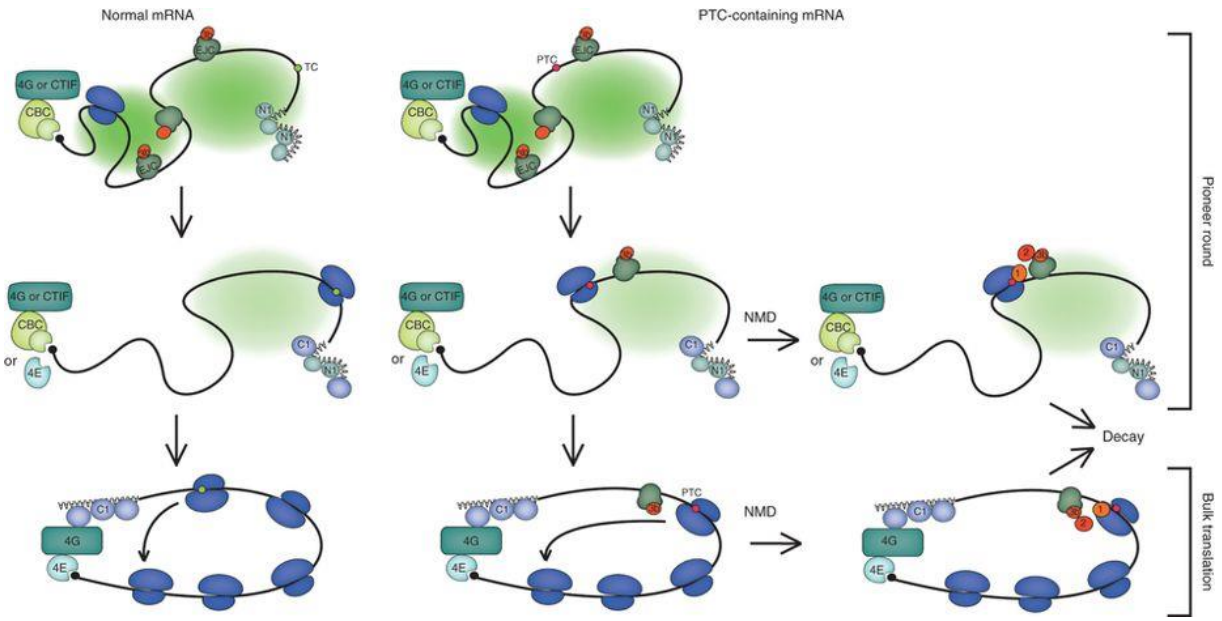


Figure 1.2: NMD targets both CBC and EIF4E-bound mRNAS.

Newly synthesised mRNPs exit the nucleus whilst still bound to the CBC. Once entering the cytoplasm the pioneer round of translation begins where a ribosome scans the transcript dislodging EJCs until it reaches a termination codon. At this point, if the mRNP is processed without interuption the mRNP CBC is replaced with EIF4E and the mRNP forms a stable loop conformation to facilitate bulk translation. If the mRNP contains a PTC that is recognised either during the pioneer round of translation or during bulk translation due to the presence of a downstram EJC, the ribosome stalls allowing UPF1 instead of PABC1 to interact with ERF3 and consequently induce NMD. Figure taken from (Rufener & Muhlemann 2013).

1.2.2.4 An emerging model

The EJC model is the principally accepted mechanism of NMD, and is supported by recent genomic and transcriptomic bioinformatics studies which identify downstream EJCs as the most important predictor of human NMD targeted mRNAs (Lappalainen et al. 2013; Lindeboom et al. 2016). The EJC model however, is not considered a complete model of NMD. As previously mentioned, recent findings challenge that NMD is limited to CBP-bound mRNPs during the pioneer round of translation (Durand & Lykke-Andersen 2013; Rufener & Muhlemann 2013). The EJC model also fails to explain NMD of mRNAs derived from intronless precursors (Rajavel & Neufeld 2001), viral unspliced mRNAs (LeBlanc & Beemon 2004; Quek & Beemon 2014) and spliced mRNAs without EJC components or without necessary EJC spacing (Buhler et al. 2006; Wang et al. 2002; Wen & Brogna 2010; Zhang et al. 1998).

A major hurdle in unravelling the exact mechanism of NMD is the lack of a reliable *in vivo* termination assay. Most recently Neu-Yilik et al. circumvented this requirement using a completely reconstituted *in vitro* translation system to investigate the involvement of the core NMD factors, UPF1, UPF2 and UPF3 in translation termination (Alkalaeva et al. 2006; Neu-Yilik et al. 2017). Contradictory to current understanding, it was found that UPF1, irrespective of phosphorylation status or ATPase activity had no impact on the efficiency of translation termination and showed no measurable evidence of direct binding to ERF1 or ERF3. This finding corroborated with recent findings in yeast that suggest UPF1 has no role in translation elongation, termination or ribosome recycling *in vitro* (Schuller et al. 2018).

Along with UPF1, Neu-Yilik et al. also discovered UPF2 to be functionally dispensable during *in vitro* translation termination. Interestingly, termination delay was shown to be specifically caused by the NMD factor UPF3B. In particular, the UPF3B RNA-recognition domain and

middle domain were found to interfere with termination codon recognition and ribosomal peptide release. UPF3B was also shown to interact directly with ERF3A to form a trimeric complex with both ERF3A and ERF1. Moreover, UPF3B was seen to bind RNA, the ribosome and UPF1 (Neu-Yilik et al. 2017). Together these findings still support a coupled translation termination and NMD mechanism, however, rather than UPF1 being the lead player, UPF3B independently binds the ribosome, RNA, and release factors to influence the efficiency of translation termination whilst also promoting NMD through a direct interaction with UPF1 (Figure 1.3). This study dramatically alters the current view on the NMD factor UPF3B and pushes the possibility of a revised model of NMD.

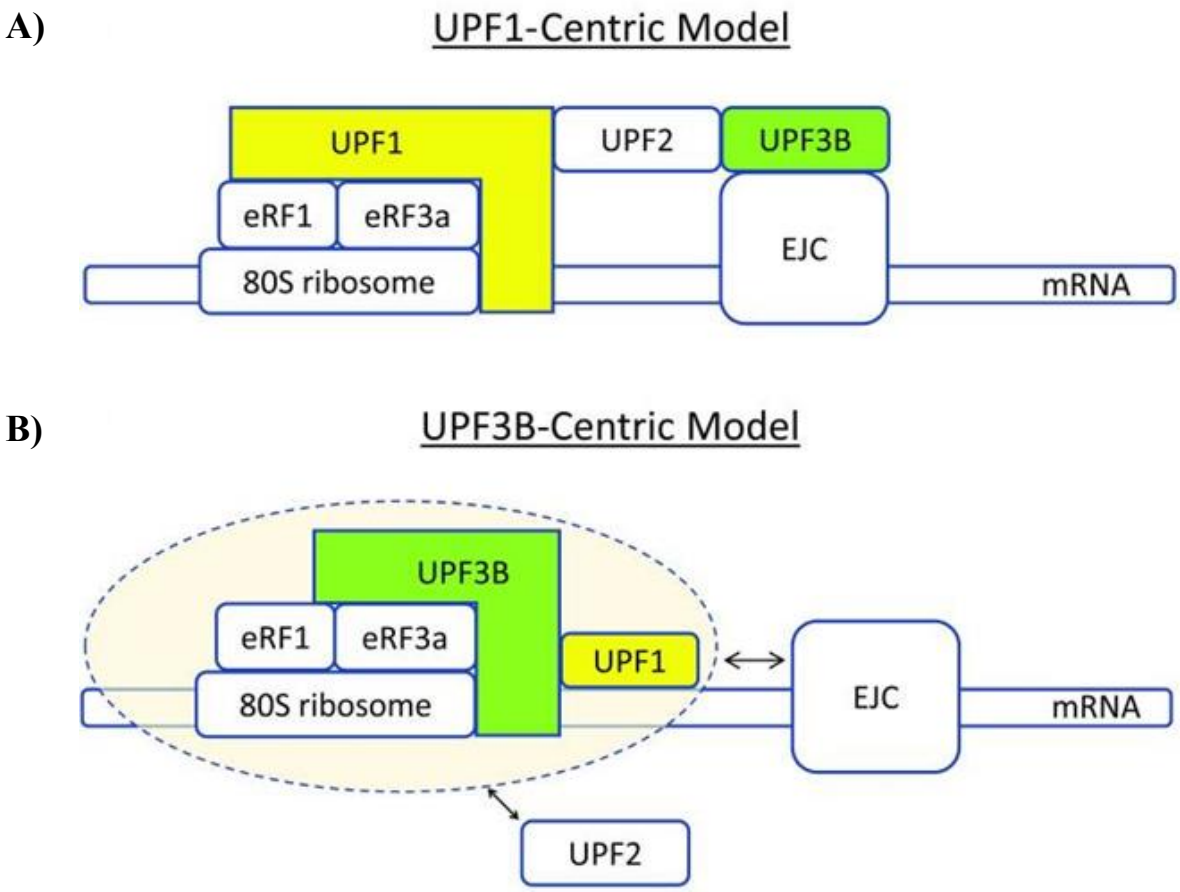


Figure 1.3: Two models that explain coupling of translation termination and NMD.

(A) The classical, UPF1-centric model of NMD describes UPF1 as the central player which couples translation termination with NMD. Here UPF1 interacts with release factors bound to the terminating ribosome stalled at a PTC, UPF1 then recruits UPF2 which is bound by the EJC factor UPF3B effectively bridging the stalled ribosome to the EJC, this results in UPF1 phosphorylation and subsequent NMD of the target mRNA. **(B)** The UPF3B-centric model of NMD provides an alternative to the classical model. In this model, when a ribosome is stalled at a PTC, UPF3B alone links the translation termination machinery to the NMD machinery. A role for the involvement of the EJC and UPF2 in this model have yet to be determined. Figure taken from (Gao et al, 2017).

1.2.3 Alternative NMD branches

Studies investigating the mechanism of NMD revealed that in addition to the classical NMD pathway, distinct 'non-classical' or alternative branches of NMD exist and are characterised by their dependence on specific NMD factors (Chan et al. 2007; Gehring et al. 2005). Such studies have also uncovered evidence that these branches can target both overlapping and distinct subsets of NMD targeted mRNA transcripts, eluding to the possibility that the cellular composition, and even mRNP composition of NMD factors can influence NMD target specificity across different cell and tissue types and across different developmental stages (Miller & Pearce 2014).

1.2.3.1 The UPF2-independent NMD pathway

Like UPF1, UPF2 is a phosphoprotein. In the classical NMD model UPF2 bridges UPF1 and UPF3 via distinct domains. A UPF2-independent pathway was first proposed when tethered function analysis uncovered that the UPF2-UPF3B interaction was not essential for NMD activity (Gehring et al. 2003). This was further supported by evidence from the same group which showed that a subset of NMD substrates remained downregulated following strong UPF2 RNAi depletion. These findings ultimately led to the identification of two functionally distinguishable EJC subgroups, namely, RNPS1-type EJCs which require normal levels of UPF2 to trigger efficient NMD, and RBM8A-MAGOH-EIF4A3-BTZ-type EJCs which are able to activate NMD in UPF2 depleted cells (Figure 1.4) (Gehring et al. 2005). Existence of the latter EJC type and resulting NMD is now supported by the previously described findings by Neu-Yilik et al. who concluded that UPF2 is dispensable for NMD activation as UPF3B can directly associate with UPF1 (Neu-Yilik et al. 2017).

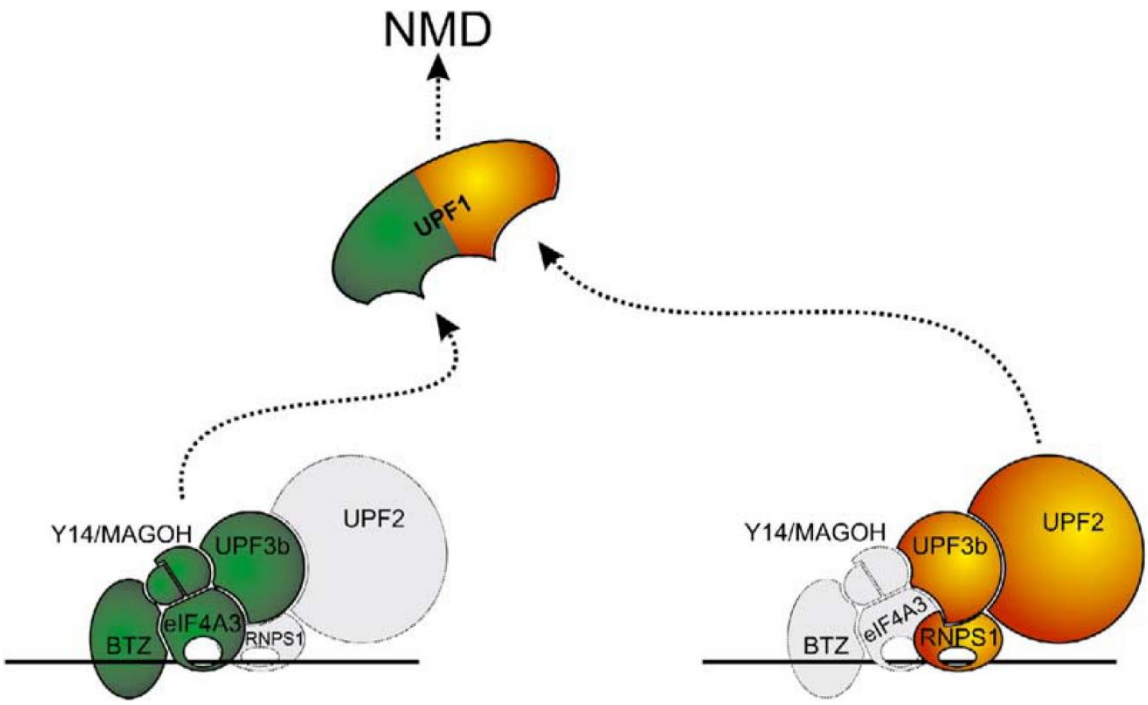


Figure 1.4: UPF2-independent and UPF2-dependent NMD is dictated by EJC composition.

NMD of a targeted mRNP can occur with or without UPF2. The UPF2-dependent route is illustrated on the right and is dictated by an RNPS1-type EJC (yellow) The UPF2-independent route is illustrated on the left and requires a RBM8A-MAGOH-EIF4A3-BTZ-type EJC (green). Proteins that are involved are coloured and those that are not essential are shown in grey. Figure taken from (Gehring et al. 2005).

1.2.3.2 The UPF3-independent pathway

UPF3 is an NMD effector of the EJC. In vertebrates UPF3 is encoded by two genetic paralogs, the X-linked *UPF3B* and the autosomal *UPF3A* genes. Both paralogs share a high sequence similarity, and their encoded proteins compete for UPF2 binding to activate classical NMD. Under normal circumstances, UPF3B is the preferred binding partner of UPF2 and is more effective at triggering NMD than UPF3A (Chan et al. 2009; Chan et al. 2007; Kunz et al. 2006). The presence of a UPF3B-independent NMD pathway was initially observed in patient derived lymphoblast cells which harboured PTC-type *UPF3B* variants. These cells were found to have reduced levels of *UPF3B* mRNA (indicative of active NMD) despite no active UPF3B protein (Tarpey et al. 2007). The presence of both UPF3B-independent and UPF3B-dependent pathways was later confirmed through *in vivo* studies (Huang et al. 2011; Karam et al. 2015). Moreover, using HeLa cells Chan et al. demonstrated that a strong depletion of UPF3B and/or UPF3A had little effect on the downregulation of several selected NMD-targeted mRNAs (Chan et al. 2007). This study demonstrated that NMD can be activated without either UPF3 proteins, suggesting that they are not always required for NMD function.

1.2.4 mRNA transcripts targeted by NMD

Historically, NMD was discovered and characterised as a mechanism that degrades aberrant, PTC containing transcripts which arise via replication, transcription, or genetic error. In mammalian cells, NMD can recognise a PTC when it is located more than 55 nts upstream of an exon-exon junction (Holbrook et al. 2004; Nagy & Maquat 1998). In this capacity NMD serves to protect the cell from the production of deleterious truncated proteins.

Dependent on species, cell or tissue type, inhibition of NMD has been found to impact expression of 5–15% of endogenous transcripts (McIlwain et al. 2010; Mendell et al. 2004; Weischenfeldt et al. 2012; Yepiskoposyan et al. 2011). These endogenous targets of NMD are

important for many cellular processes including cellular homeostasis, the stress response, cell cycle progression and differentiation (Kurosaki & Maquat 2016; Linder, Fischer & Gehring 2015; Lykke-Andersen & Jensen 2015; Ottens & Gehring 2016). Moreover, NMD can target both protein encoding mRNAs and non-functional mRNAs e.g. long non-coding RNAs and those derived from intragenic regions or pseudogenes (He et al. 2003; Lykke-Andersen et al. 2014). Through targeting endogenous transcripts NMD can act as a global regulator of gene expression.

NMD eliciting PTCs can be naturally introduced by; an exon-exon junction downstream of the endogenous termination codon, i.e. the result of an intron in the 3'UTR (Figure 1.5A), alternative splicing events or non-productive genetic loci rearrangements e.g. *TCR* loci, which result in introduction of a PTC (Figure 1.5B) and mRNA features such as; short reading frames upstream of the main open reading frame (uORFs) (Figure 1.5C). or atypically long (>1 kb) 3′ untranslated regions (3'UTRs) (Figure 1.5D) (He et al. 2003; Kebaara & Atkin 2009; Kervestin & Jacobson 2012; McGlincy & Smith 2008; Muhlrad & Parker 1999; Weischenfeldt et al. 2012).

Lastly, mRNAs encoding selenoproteins are also known to be endogenous NMD targets. Selenoproteins are a specific group of proteins which incorporate the unique amino acid selenocysteine. Selenocysteine insertion into the peptide chain occurs during translation and is encoded redundantly by the termination codon 'UGA'. In conditions where selenocysteine is unavailable, the 'UGA' mRNA sequence will encode a termination codon which has the potential to act as a PTC thus stimulating premature translation termination followed by NMD. Comparatively, when selenocysteine is available, the 'UGA' mRNA sequence can recruit selenocysteine creating competition between selenoprotein biosynthesis and NMD (Moriarty et al. 1998; Seyedali & Berry 2014).

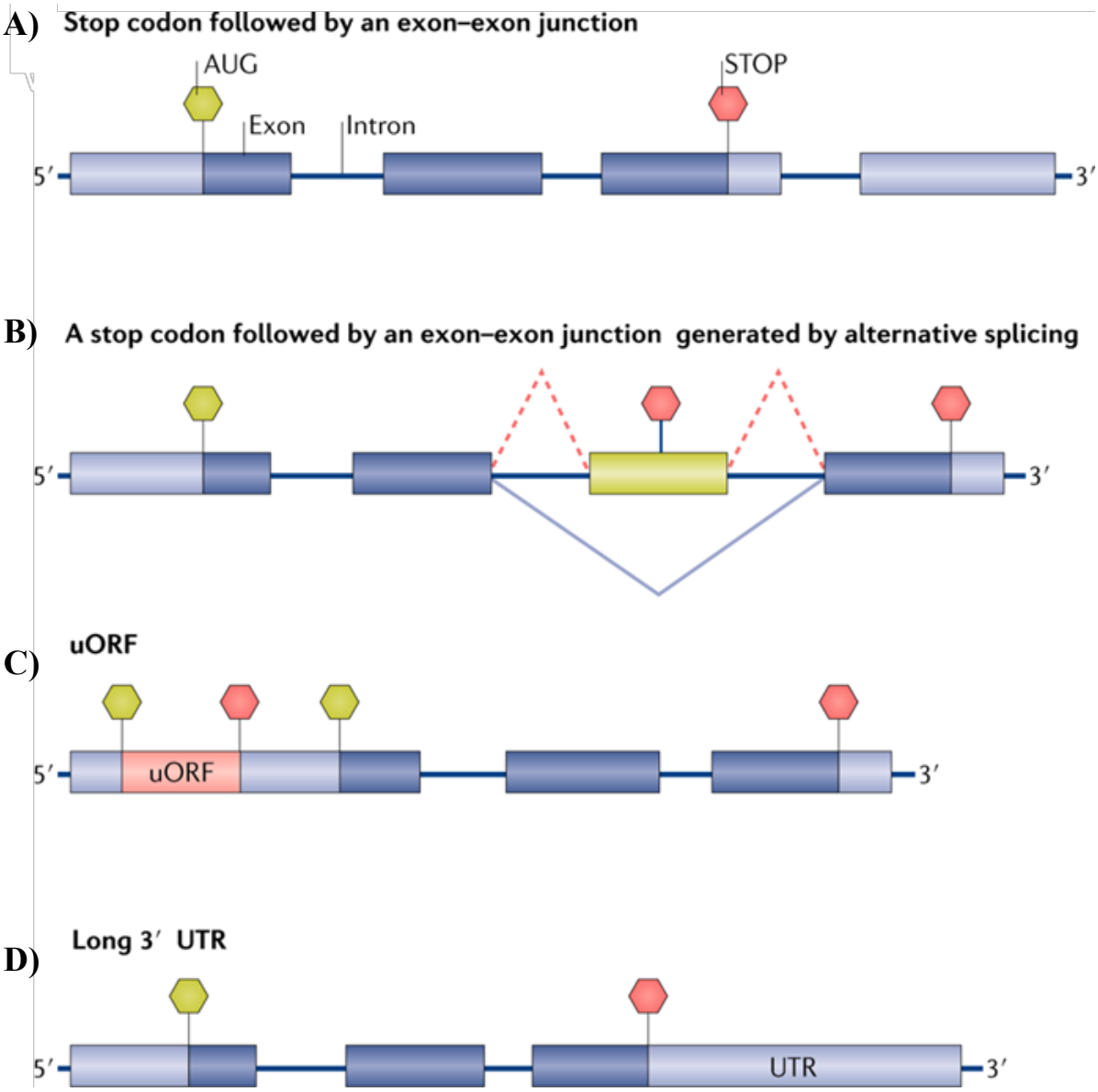


Figure 1.5: NMD inducing features found on endogenous mRNA transcripts.

Endogenous mRNA transcripts can contain an NMD-targeted stop codon (red hexagon) which results in the recruitment of NMD factors and degradation of the transcript. These stop codons are naturally introduced by; **(A)** an exon-exon junction downstream of the naturally occurring termination codon, due to an intron in the 3'UTR, **(B)** alternative splicing events which can introduce a PTC (inclusion of a PTC containing intron shown here) or **(C,D)** short reading frames upstream of the main open reading frame (uORFs). Long 3′ untranslated regions (>1 kb) can also elicit NMD. However, the presence of a uORF or a long 3′ UTR does not necessarily trigger NMD; thus, these NMD-inducing features act only in specific contexts. Figure taken from (Jaffrey & Wilkinson 2018).

1.3 The surveillance role of NMD

An estimated 5–30% of human transcripts are aberrant due to a PTC-type mutation which may encode truncated proteins (Bhuvanagiri et al. 2010; Huang & Wilkinson 2012; Nguyen, Wilkinson & Gecz 2014). The surveillance role of NMD acts to recognise and degrade PTC-containing mRNAs to reduce the production of aberrant and potentially deleterious truncated proteins. Diseases caused by these mutations are three times more likely to come to clinical attention than missense mutations (Krawczak et al. 1998; Miller & Pearce 2014; Mort et al. 2008). In fact, one third of all human genetic or acquired diseases are caused by one or more PTC-type mutations (i.e. nonsense, splice site or frameshift mutations) (Frischmeyer & Dietz 1999; Keeling & Bedwell 2011; Mort et al. 2008). For these affected individuals, the surveillance role of NMD can act in a protective manner by ridding the cell of truncated proteins that are detrimental to cell function even when the other allele is normal and expressed (dominant negative). However, NMD can also act in a disease aggravating manner by ridding the cell of truncated proteins that retain some beneficial wild-type activity. This concept highlights NMD as an important disease driver and modifier and has led to the development of several NMD targeting therapeutics.

1.3.1 The protective role of NMD

The signature phenomenon of NMD is a decreased abundance of mRNAs containing PTC-type mutations. Often NMD prevents the translation of possibly deleterious truncated proteins orchestrating a protective role in the cell. Theoretically this protective role of NMD can act in any disorder caused by a PTC-type mutation which leads to an abnormally truncated protein with a dominant negative effect, but was first discovered through investigations into the genetic disorder β-thalassemia (Chang & Kan 1979). In this study, NMD was found to limit the synthesis of C-terminally truncated β-globin (HBB) polypeptides that might otherwise act in a dominant negative fashion (Chang & Kan 1979).

The activation of NMD involved in β-thalassemia is dependent on PTC location within the *HBB* transcript. PTC-type mutations that reside at least 55 nts upstream of an exon-exon junction are able to trigger efficient NMD resulting in the common recessive mode of inheritance, in which heterozygous carriers of a PTC-type *HBB* mutation are phenotypically normal since protein production from the wild-type allele can compensate for loss of protein production from the variant allele. In contrast, rare PTC-type mutations in the last exon (not followed by a downstream EJC) were unable to trigger NMD, resulting in translation of long, truncated, and non-functional HBB protein. This overwhelms the cellular proteolytic system resulting in toxic precipitation of insoluble globin chains. Heterozygotes with these mutations are affected with an atypical form of dominantly inherited β-thalassemia.(Hall & Thein 1994; Thein et al. 1990).

β-thalassemia was the prototype disorder which first documented the medical importance of NMD. Since then, the dependence of NMD activity on PTC position has been found to influence the clinical severity of several genetic diseases, including but not limited to; Robinow Syndrome, Brachydactyly Type B and Von Willebrand disease (Patton & Afzal 2002; Schneppenheim et al. 2001; Schwabe et al. 2000).

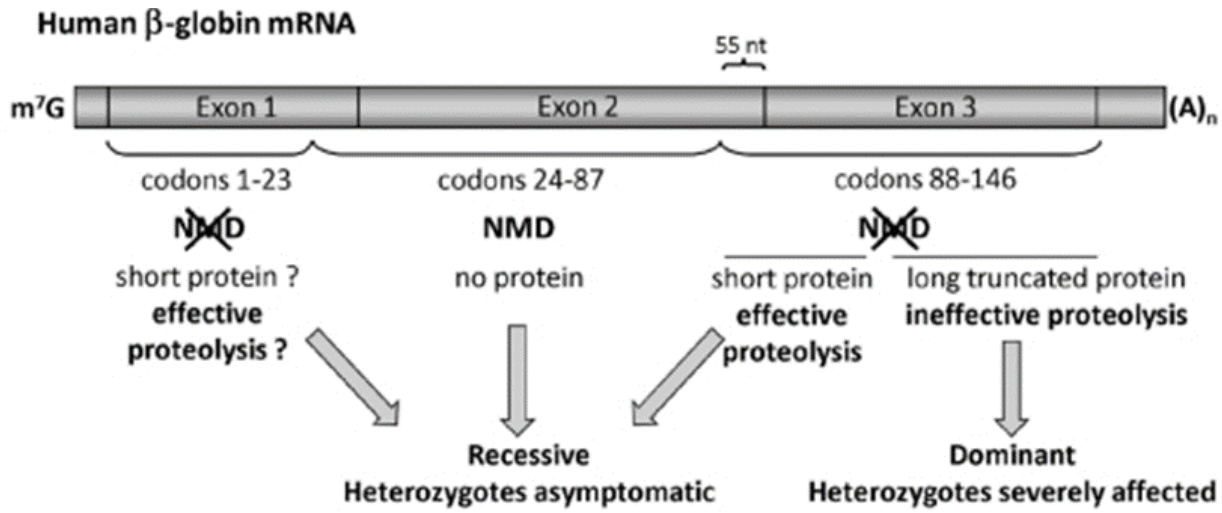


Figure 1.6: NMD of human β-globin mRNA and its impact on β-thalassemia phenotype.

A schematic representation of NMD-resistant and NMD-sensitive regions of human β-globin mRNA. 5' proximal PTCs escape NMD, however heterozygotes are asymptomatic as the translated short β-globin peptides along with the excess α-globin chains are effectively degraded. If a PTC is located downstream of codon 23 and more than 55 nts upstream of the last exon-exon junction, the corresponding transcript is targeted for NMD and heterozygotes are asymptomatic. If the PTC location allows the transcript to escape NMD (i.e. less than 55 nts upstream of the exon-exon junction or not followed by an exon-exon junction), the corresponding transcript can be translated into a truncated protein that is small enough to be efficiently degraded along with the excess α-globin chains resulting in an asymptomatic presentation. Alternatively, the NMD-escaping transcript can be translated into a long, truncated, and non-functional β-globin protein which overwhelms the cellular proteolytic system and results in toxic precipitation of insoluble globin chains leading to the dominantly inherited form of β-thalassemia. Image taken from (Peixeiro et al. 2012).

1.3.2 The disease aggravating role of NMD

Since NMD functions to rid the cell of truncated proteins, it has the potential to aggravate the pathology of certain diseases by preventing the formation of truncated proteins that would otherwise retain residual beneficial activity. This can result in either loss of function or haploinsufficient phenotypes. The role of NMD in dystrophinopathies provides a clear example of this.

PTC-type mutations in transcripts encoded from the X-linked dystrophin gene (*DMD*) can be targeted by NMD preventing the synthesis of a truncated protein which could otherwise confer partial or complete function. When this occurs, NMD activity leads to a loss of functional dystrophin dosage causing the severe form of the disorder, Duchenne muscular dystrophy (DMD). Comparatively, rare PTC-type mutations that evade NMD and give rise to partially functional truncated protein result in the milder, more heterogenous form of the disorder known as Becker muscular dystrophy (BMD) (Kerr et al. 2001). Further examples of this have been documented for Ulrich congenital muscular dystrophy, Ataxia-telangiectasia and Townes-Brocks syndrome (Furniss et al. 2007; Li & Swift 2000; Usuki et al. 2004).

1.3.3 NMD and cancer genetics

Cancerous tumours arise from genetic and epigenetic alterations which disrupt the normal life cycle of a cell and lead to unconstrained growth and evasion of cell death. To facilitate these changes it has been found that by manipulating the NMD pathway, tumour cells are able to alter their own transcriptomes (Popp & Maquat 2018; Wang et al. 2011). These changes allow tumour cells to adapt and survive in the tumour microenvironment. In particular it was found that cells from cases of stomach adenocarcinoma, kidney cancer and colon cancer all presented with a higher proportion of NMD targeted PTC-type mutations when compared to cells from cases of other cancer types with similar mutation frequencies (Hu et al. 2017).

Mutations in oncogenes are predominantly missense, whereas tumour suppressor genes exhibit an increased proportion of PTC-type mutations that are recognised by NMD (Hu, Yau & Ahmed 2017; Mort et al. 2008). NMD of transcripts containing these PTC-type mutations generally occurs to eliminate translation of a possibly dominant negative protein, in cancer genetics however, this often has a detrimental effect. For example, scenarios where an allele of a tumour suppressor gene is rendered non-functional (e.g. due to a heterozygous deletion, chromosomal loss, or another mutation) and the second contains a PTC-type mutation, NMD of transcripts expressed from the PTC containing allele will result in a complete loss of function from that tumour suppressor gene, thus promoting cancer (Figure 1.7). Some examples of this have been documented for PTC-type mutations occurring in genes encoding E-cadherin in stomach cancers (Karam et al. 2008), BRCA1 in breast and ovarian cancer (Perrin-Vidoz et al. 2002), BRCA2 in ovarian cancers (Ware et al. 2006), p53 in breast cancers (Anczukow et al. 2008), pRb in lymphoma (Pinyol et al. 2007) and Wilm tumour protein in kidney cancers (Reddy et al. 1995).

Additionally, a PTC-type mutation in a tumour suppressor gene which evades NMD can also promote cancer through the production of a truncated protein with a dominant negative effect on the function of protein produced from the wild-type allele (Figure 1.7) (Lindeboom, Supek & Lehner 2016).

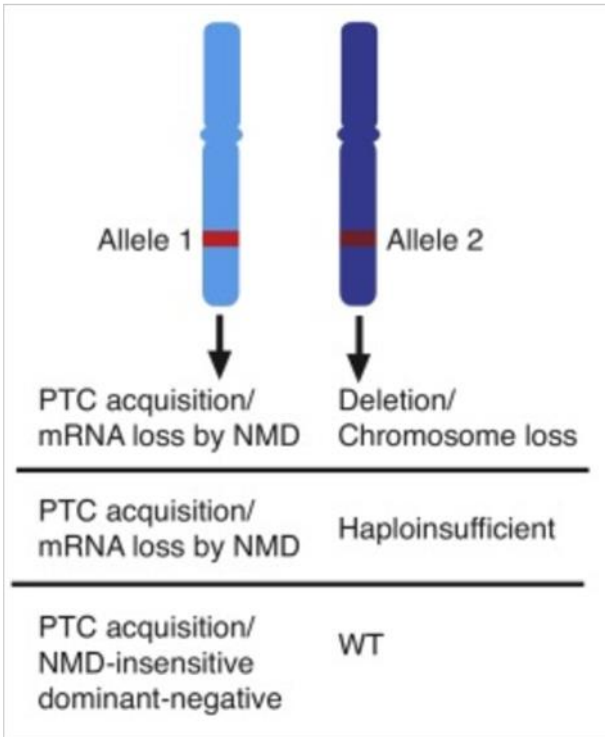


Figure 1.7: Tumour suppressor gene function can be inactivated by PTC-type mutations.

There are several scenarios in which tumour suppressor gene function can be inactivated by NMD, some examples include; introduction of a PTC-type mutation and loss of transcripts from one allele of a tumour suppressor gene via NMD combined with either a deletion/chromosomal loss of the wild-type (WT) allele or haploinusfficient function of protein produced from the remaining wild-type version. Alternatively, a PTC-type mutation located in an NMD-insensitive region may result in production of a dominant-negative allele which interferes with wild-type function Figure adapted from (Popp & Maquat 2018).

1.3.4 Nonsense suppression therapies

As discussed, NMD has been implicated in a long, and still growing list of genetic disorders and is now emerging as a pathway of interest for developing new therapeutics. In theory, NMD suppression and NMD stimulation could both be useful depending on disease context, however, only NMD inhibitors designed for use in diseases where the degraded mutant protein could confer beneficial residual activity have reached the clinic.

The main challenge when developing NMD targeting therapeutics is correcting the disease phenotype without disrupting the many physiological roles of NMD. A popular approach to achieve this has been to identify or develop compounds which rather than inhibiting the NMD pathway, act to promote read-through at the site of a PTC (nonsense suppression). In this way, NMD of the PTC containing transcript can be bypassed and rather than complete loss of protein expression, a full-length protein containing a missense mutation is produced. It is important to note however, that as discussed, there are several instances in which a PTC can be naturally introduced to regulate endogenous cellular gene expression and read-through at these sites may have detrimental consequences for the cell.

1.3.4.1 Suppressor tRNA

One means of suppressing the effect of a PTC-type mutation is the introduction of DNA encoding a chimeric tRNA which can specifically recognise the PTC and compete more efficiently for interaction with the termination machinery than a near-cognate tRNA. This will result in the introduction of an amino acid at the PTC site rather than translation termination thus negating NMD. This approach has been successful *in vitro* and *in vivo* using models of β-thalassemia and DMD (Buvoli et al. 2000; Kiselev et al. 2002; Temple et al. 1982), however, the lack of an efficient delivery method and stable retention and expression of tRNA genes

remains a significant challenge for clinical use of this approach (Atkinson & Martin 1994; Dabrowski et al. 2018).

1.3.4.2 Aminoglycoside treatment

A more popular method of nonsense suppression has been the use of aminoglycosides, such as the commonly used antibiotics, gentamicin and amikacin. In eukaryotes these drugs can also bind to the decoding centre of the ribosome to decrease the accuracy of codon-anticodon pairing. This increases the likelihood of introducing an amino acid at a PTC site rather than initiating translation termination, thus evading NMD and resulting in production of full length protein containing a missense mutation (Recht et al. 1999). Aminoglycoside treatment has resulted in functional improvement *in vitro* and in animal models of cystic fibrosis (CF) and DMD caused by PTC-type mutations (Bedwell et al. 1997; Zsembery et al. 2002).

CFTR is the gene affected in CF. This gene encodes a chloride channel which regulates the salt content of the fluid that covers cell surfaces within the nose and lungs. Transport of ions such as sodium and chloride create an electrical potential which can be measured in terms of nasal potential difference. This measurement is used as a diagnostic tool for CF as affected individuals show a reduction in nasal potential difference compared to unaffected individuals and this difference correlates with disease severity (Rowe et al. 2011). Clinical trials have shown that topical administration of gentamicin resulted in a measurable improvement of nasal potential difference in individuals diagnosed with CF with a subset of individuals showing detectable CFTR protein in nasal epithelial cells (Clancy et al. 2001; Wilschanski et al. 2000; Wilschanski et al. 2003). In contrast, individuals with DMD or BMD were administered prolonged intravenous gentamicin yet showed no significant improvement in functional tests (Politano et al. 2003; Wagner et al. 2001). The variable efficacy of gentamicin (further discussed in Section 1.5.3) and known side effects of prolonged aminoglycoside use, such as

kidney damage and hearing loss ultimately prevented further interest into these drugs as an NMD suppression therapy.

1.3.4.3 Ataluren

Ataluren, formerly known as PTC124 is an 1-2-4 oxadiazole compound which has been developed into the first FDA approved treatment for DMD caused by a PTC-type mutation. Ataluren is believed to interact with the ribosome and stimulate incorporation of near-cognate tRNAs to the PTC site of a translating polypeptide, this results in production of a full-length protein containing a missense mutation (Figure 1.8) (Roy et al. 2016; Siddiqui & Sonenberg 2016). Ataluren was first discovered in a high-throughput screen for its ability to promote read-through of nonsense codons (Welch et al. 2007). Subsequent *in vitro* and *in vivo* studies revealed that ataluren showed selectivity for read-through of PTC-type mutations causing DMD and CF without affecting the processing of normal termination codons (Du et al. 2008; Welch et al. 2007). This paired with its high oral bioavailability and favourable safety profile led it to be developed into a DMD therapy (Hirawat et al. 2007).

Following this success, ataluren was also investigated as a treatment for CF, however, in recent phase 3 clinical trials (NCT01140451) only a subset of individuals who were not using chronic inhaled aminoglycoside tobramycin showed a lower decline in lung function while overall the trial failed to achieve its primary and secondary endpoints (Aslam et al. 2017; Kerem et al. 2014).

The use of NMD suppression therapies have also been of interest to treat inherited eye diseases such as choroideremia, ocular coloboma retinitis pigmentosa, Usher syndrome and Aniridia which are all largely the result of a PTC-type mutation (reviewed in Richardson et al. 2017). In particular, over 600 genetic defects leading to haploinsufficiency of the paired box 6 (PAX6)

transcription factor are known to result in congenital anirdia and of these 72% are PTC-type mutations. Following successful studies in mice where Gregory-Evans et al, used topical administration of ataluren to not only inhibit progression of the disease but reverse the effects of the disorder if treated within a specified time frame (Gregory-Evans et al. 2014), the use of oral ataluren to treat patients with PTC-type mutation aniridia has recently entered phase 2 clinical trials (NCT02647359).

To aid the ongoing development of a nonsense suppression therapy to treat CF alongside countless other disorders caused by PTC-type mutations it is important to fully understand the role NMD plays in the pathology of these disorders, for example is there some cell or tissue specific effect of NMD contributing to the observed phenotypes, or could the observed variable efficiency of treatments be due to underpinning differences in the affected individuals capacity to elicit NMD. A tool allowing the endogenous activity of NMD to be explored at a single cell level can help answer these questions and triage therapies.

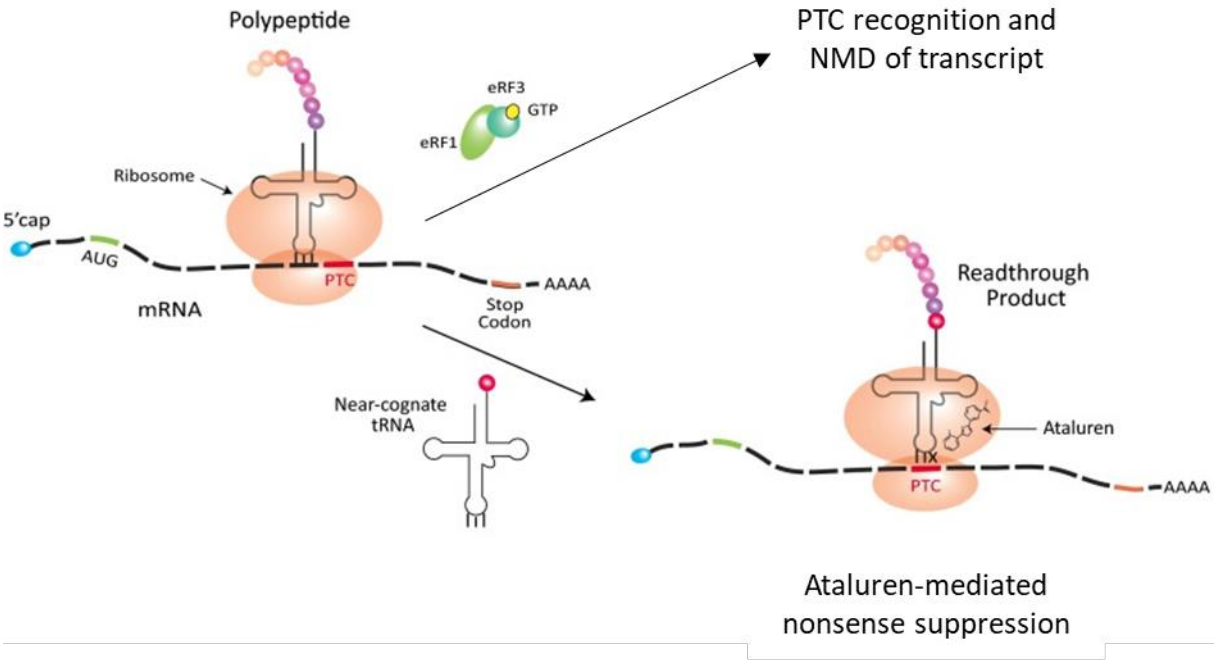


Figure 1.8: A comparison of NMD of a transcript due to PTC recognition and ataluren facilitated translation read-through at a PTC-type mutation.

When a ribsome encounters a PTC it will normally result in premature translation termination and subsequent mRNA degradation via the NMD pathway. This can prevent the formation of truncated proteins that may retain function and is a common mechanism of loss of function disease pathology. The drug atalauren is proposed to bypass premature translation termination by interacting with the ribosome and facilitating recruitment of near-cognate tRNAs to the PTC site. This allows for read-through at the PTC site and production of a full length protein. Figure adapted from (Siddiqui & Sonenberg 2016).

1.4 The Regulatory role of NMD

In addition to its surveillance role, NMD also regulates the stability of approximately 5–15% of normal, physiological transcripts (Adachi et al. 2004; McIlwain et al. 2010; Mendell et al. 2004; Nguyen et al. 2012; Weischenfeldt et al. 2012; Yepiskoposyan et al. 2011). This is exemplified by the many developmental and environmental cues which alter NMD efficiency to influence the expression of endogenous NMD targets and regulate numerous physiological processes in the cell. These include; response to cellular stresses, embryonic cell differentiation, and neurodevelopment (Bruno et al. 2011; Karam et al. 2015; Karam & Wilkinson 2012; Martin & Gardner 2015).

1.4.1 NMD and the stress response pathway operate in synchrony through a negative feedback loop

Cellular stress can be caused through unfavourable environmental or intrinsic conditions. Intrinsic cellular stress is often caused by an accumulation of misfolded proteins in the lumen of the endoplasmic reticulum (ER) and referred to as ER stress. Such misfolded proteins can occur due to several reasons including; protein coding mutations, aberrantly high translation rates or defective protein folding capabilities (Ma & Hendershot 2004; Moore & Hollien 2012; Ron & Walter 2007; Walter & Ron 2011). The unfolded protein response (UPR) is a complex signal transduction pathway which orchestrates cellular adaptation under conditions of ER stress. To initiate the UPR, ER stress is sensed by three main stress sensors, inositol-requiring protein 1 (IRE1), activating transcription factor 6 (ATF6) and protein kinase RNA-like ER kinase (PERK). Each sensor activates a distinct branch of the UPR which transduces information about protein-folding status in the ER and ultimately collaborates to alleviate ER stress (Hetz 2012).

When IRE1 binds an unfolded protein, it undergoes trans-autophosphorylation and then drives non-canonical splicing of the X-box binding protein 1 (XBP1) mRNA This event shifts the open reading frame of the mRNA to generate a stable and active transcription factor known as XBP1s which promotes expression of genes encoding chaperones which modulate protein folding, lipid synthesis proteins which increase the size of the ER membrane to accommodate increased protein load, and ER-associated degradation (ERAD) proteins which are responsible for clearance of misfolded proteins in the ER for cytosolic proteasomal degradation (Goetz & Wilkinson 2017; Hetz & Papa 2018; Ron & Walter 2007).

In response to stress, the chaperon immunoglobin binding protein (BiP) is released from its ATF6 binding site, activating ATF6 and allowing it to be packaged and sent to the Golgi where it is cleaved and transported to the nucleus to activate the transcription of several genes encoding proteins important for protein folding (Hetz & Papa 2018; Walter & Ron 2011).

The third stress sensor PERK undergoes trans-autophosphorylation upon encountering an unfolded protein, this results in phosphorylation of the α-subunit of eukaryotic translation initiation factor 2 (eIF2α). eIF2α phosphorylation leads to a general suppression of protein synthesis to allow the cell to cope with the already high levels of unfolded proteins. eIF2α phosphorylation also leads the paradoxical translational induction of selected stress-related mRNAs including the transcription factor ATF4 (Harding et al. 2000; Ron & Walter 2007; Walter & Ron 2011). Together, these responses aim to alter gene expression patterns to restore cellular homeostasis, however if the stress cannot be mitigated the UPR will trigger apoptosis to eliminate the damaged cell, this can occur through any of the three UPR branches (Goetz & Wilkinson 2017; Ma & Hendershot 2004; Oslowski & Urano 2011).

Environmental stresses include, hypoxia, amino acid deprivation, viral infection and generation of reactive oxygen species, these various stresses are sensed by specialised kinases, known as;

double-stranded RNA-dependent protein kinase (PKR), heme-regulated EIF2α kinase (HRI), and general control non-derepressible 2 (GCN2) which like the UPR stress sensor PERK, act to phosphorylate eIF2α and result in the same downstream effects. In this way PERK is not only a stress sensor of the UPR but also of a second cellular stress response pathway known as the integrated stress response (ISR) which is triggered by activation of any of these four eIF2α kinases (PERK, PKR, HRI, GCN2) (Costa-Mattioli & Walter 2020; Pakos-Zebrucka et al. 2016). Like the UPR, the ISR aims to alleviate stress but if unsuccessful will trigger cell death (Bravo et al. 2013; Pakos-Zebrucka et al. 2016).

Cellular stress responses must be tightly regulated to maximise physiological value while also preventing deleterious effects, such as apoptosis from long term activation. NMD has been recognised as an important pathway in this regulation. The transcripts of several stress-related genes involved in the UPR and ISR including *IRE1*, *PERK*, *ATF4* and *ATF6* are NMD targets which in an unstressed cell would be actively degraded preventing consequences of unnecessary and prolonged UPR or ISR activation (Gardner 2008; Karam et al. 2015; Mendell et al. 2004). This means that in stressed cells NMD activity must be extinguished to allow a rapid and effective UPR or ISR activation. This is achieved by eIF2α phosphorylation which has been shown to be a strong inhibitor of NMD activity (Goetz & Wilkinson 2017; Wang et al. 2011). In other words, NMD and the cellular stress responses act in a negative feedback loop where under normal circumstances stress induced genes are kept at bay, whilst in stressful conditions, the stress response pathways inhibit NMD activity, enhancing their own response until the cellular stress is alleviated (Figure 1.9) (Karam et al. 2015; Martin & Gardner 2015; Oren et al. 2014; Usuki et al. 2013).

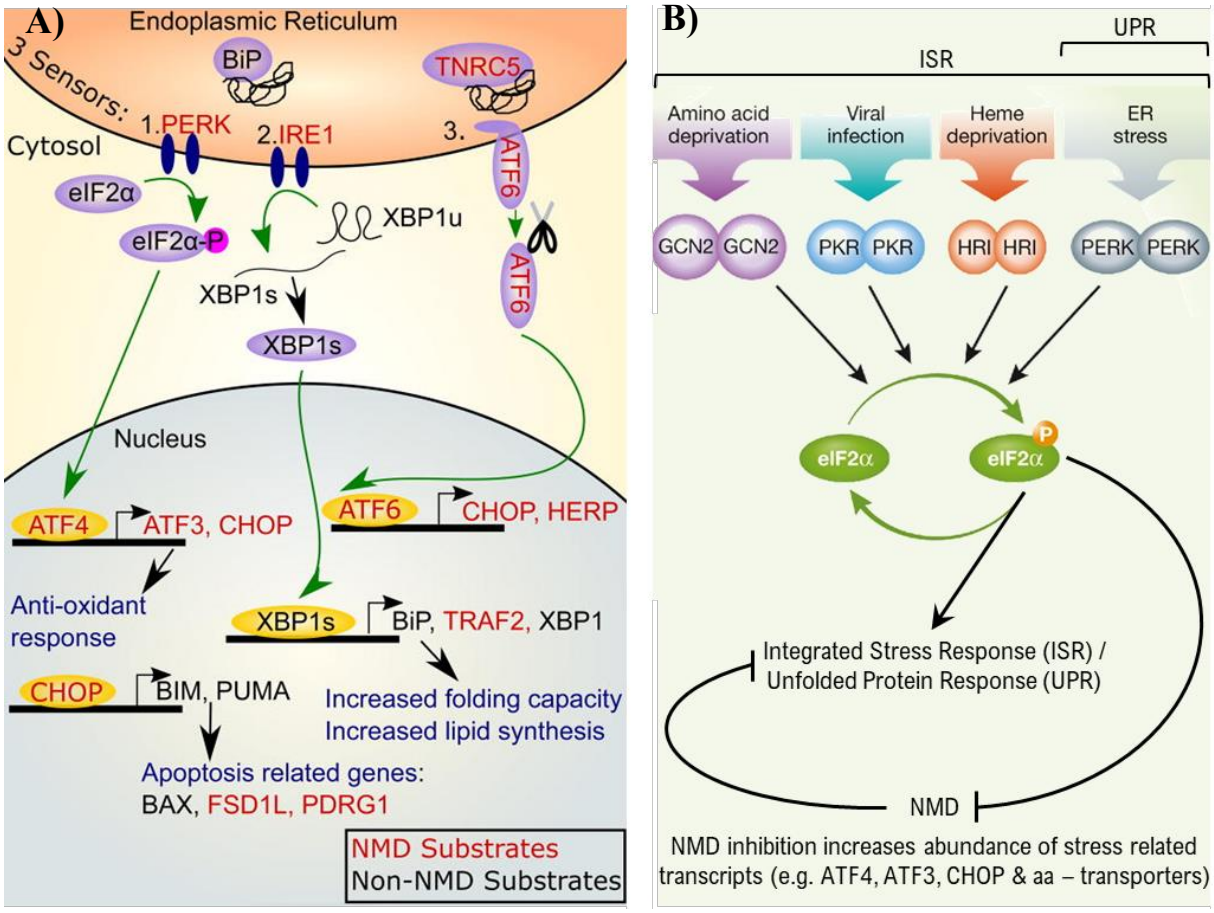


Figure 1.9: NMD and the cellular stress responses operate together in a negative feedback loop.

(A) A simplified diagram depicting the three unfolded protein response (UPR) branches, activated by either the PERK, IRE1 or ATF6 stress sensor. When chaperones including BiP and TNRC5 leave these sensors to bind unfolded proteins, the sensors activate downstream signalling which ultimately alters gene expression patterns to restore cellular homeostasis or if unsuccessful, triggers apoptosis. Stress-related components encoded by high-confidence NMD target mRNAs are shown in red font. Stress-related components not considered to be encoded by NMD target mRNAs are shown in black font. *XBP1u* XBP1 unspliced, *XBP1s* XBP1 spliced isoform. Figure taken from (Goetz & Wilkinson 2017). **(B)** Examples of cellular stresses include, amino acid deprivation, iral infection, heme deprivation and endoplasmic reticulum (ER) stress. To counteract such stresses and re-establish cellular homeostasis, cells activate stress-response pathways. These stress response pathways converge on eIF2α phosphorylation

Figure 1.9 continued…

which has been shown to be an inhibitor of NMD. Many components of the cellular stress response and amino acid (aa) transporters are endogenous substrates of the NMD pathway. Under normal circumstances, NMD keeps expression of stress-induced genes at bay, whilst in stressful conditions the stress response inhibits NMD activity thus enhancing its own response until the cellular stress is alleviated. Figure adapted from (Packos-Zebrucka et al, 2016 and Popp & Maquat 2018).

1.4.2 NMD activity influences embryonic stem cell differentiation and proliferation

Embryonic stem cells (ESCs) have two distinctive traits; they can proliferate infinitely (self-renewal) and they have the potential to differentiate into restricted daughter progenies which can form all three germ layers: ectoderm, endoderm and mesoderm (pluripotency). Transcription factors, epigenetic changes and non-coding RNAs are all known to be involved in both maintaining pluripotency and differentiation of ESCs. Additionally, the factors of NMD, through both NMD-dependent and NMD-independent functions are emerging as important influencers of ESC fate (Han et al. 2018; Li et al. 2015; Lou et al. 2016; Lou et al. 2014) (Figure 1.10).

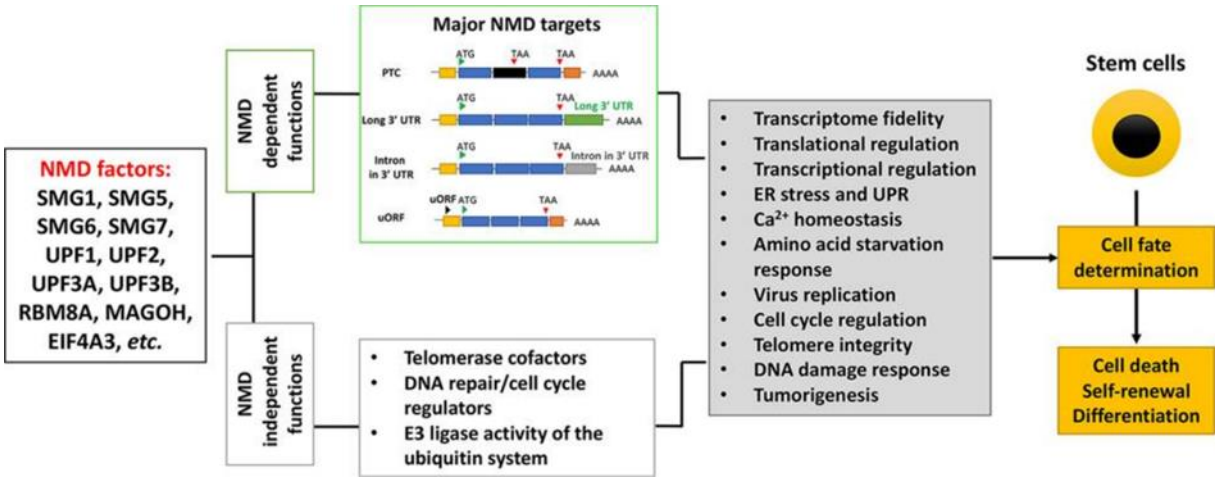


Figure 1.10: The NMD machinery is a regulator of cell fate.

Through both NMD-dependent and NMD-independent functions the NMD machinery can regulate a series of cellular processes (summarised in grey box) which in turn influence stem cell fate i.e. self-renewal, differentiation and cell death (Han et al. 2018).

Most recently, through quantification of NMD factor and NMD target gene expression, Lou et al. discovered that NMD activity of human ESCs is increased during differentiation into definitive endoderm and decreased when differentiated toward the ectoderm or mesoderm (Lou et al. 2016). Furthermore, they discovered that NMD activity correlates with proliferative activity, supporting previous studies which implicate NMD in regulation of the cell cycle (Azzalin & Lingner 2006; Lou et al. 2014; Orford & Scadden 2008). This study also highlights the variation in the role of NMD across different species when compared to a previous study in which loss or depletion of NMD factors in mouse ESCs was seen to inhibit differentiation into all three primary germ layers (Li et al. 2015).

It is hypothesised that the cellular magnitude of NMD drives hESC fate through regulation of transforming growth factor-beta (TGF-β) and bone morphogenetic protein (BMP) signalling. In the cell, TGF-β and BMP ligands induce formation of Type I and Type II receptor complexes. This results in phosphorylation of receptor-regulated (R)-Smads. These complex with Smad4 (common (Co)-Smad), then translocate into the nucleus where interaction with transcription factors can influence gene transcriptional responses, chromatic remodelling and microRNA (miRNA) processing (Figure 1.11A).

TGF-β/BMP signalling is involved in many cellular processes in the adult organism and in the developing embryo. However, regarding hESC differentiation, TGF-β/BMP signalling drives formation of an intermediate lineage between hESCs and definitive endoderm or mesoderm, known as the mesendoderm (Lou et al. 2014). After mesendoderm formation, TGF-β signalling triggers endoderm differentiation whilst BMP signalling elicits mesoderm differentiation (Figure 1.11B) (Guo & Wang 2009; Wang & Chen 2016). Despite evidence that NMD serves as a switch between these lineages by inhibiting TGF-β signalling and activating BMP signalling, further research is needed to determine additional pathways NMD may be regulating to achieve this switch and at which steps(s) of differentiation NMD activity is most important.

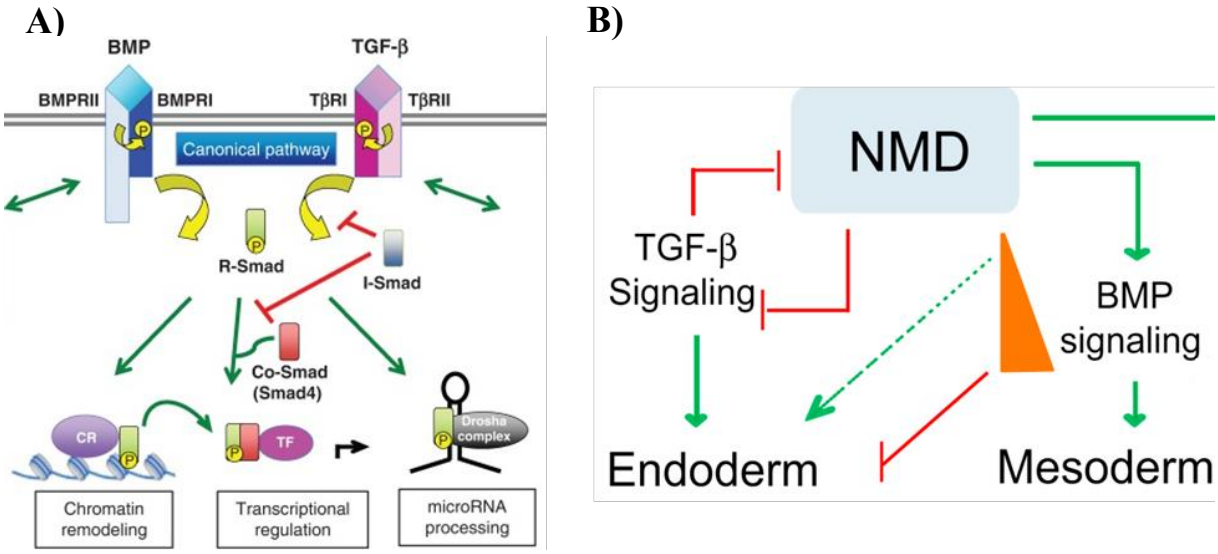


Figure 1.11: NMD drives TGF-β/BMP signalling to influence stem cell fate.

(A) In the canonical TGF-β/BMP signalling pathway, TGF-β and BMP ligands induce formation of heteromeric complex between Type II and Type I receptors. The Type II receptors transphosphorylate the Type I receptors and activate Type I receptor kinases. These kinases transmit the signal to the cell by phosphorylating receptor-regulated (R)-Smads, which form heteromeric complexes with Smad4 (common (Co)-Smad) and translocate to the nucleus where through interaction with other transcription factors can regulate chromatin remodelling, transcription and/or control microRNA processing. Image taken from (Kashima & Hata 2018). **(B)** Endoderm versus mesoderm fate is dictated by NMD through NMD magnitude based regulation of TGF-β/BMP signalling. Image taken from (Lou et al. 2016).

1.4.3 Alternative splicing coupled to NMD regulates splicing

Alternative splicing occurs in nearly 95% of mammalian genes and drives expansion and diversity of the proteasome. One third of these events produce transcripts containing PTC-type mutations which when coupled with NMD can be harnessed to regulate global gene expression, this process is known as alternative splicing coupled to NMD (AS-NMD) (Lewis et al. 2003; Pan et al. 2008; Wang et al. 2008).

Interestingly, AS-NMD influences the expression of transcripts encoding splicing factors that drive alternative splicing events. For example, AS-NMD events are commonly seen amongst members of the serine/arginine rich (SR) and heterogenous nuclear ribonucleoprotein (hnRNP) families of transacting splicing factors (Lareau et al. 2007; Ni et al. 2007; Saltzman et al. 2008; Wollerton et al. 2004). SR proteins and hnRNPs alter exon recognition by binding in a sequence specific manner to enhancer or silencer elements, respectively. These and other core spliceosomal components are able to act on their own transcripts to catalyse the production of a non-productive NMD targeted isoform, as such, self-limiting protein expression is achieved through an autoregulatory negative feedback loop (Lareau et al. 2007; Lykke-Andersen et al. 2014; Ni et al. 2007). More recently, proteins involved in chromatin modification have also been found to be regulated by AS-NMD through a similar autoregulatory negative feedback loop (Izumikawa et al. 2016; Yan et al. 2015).

Since many AS-NMD regulated splicing factors regulate their own cascade of downstream alternative splicing (Hamid & Makeyev 2014; Wollerton et al. 2004), while chromatin modifiers regulated by AS-NMD are involved in transcription and mRNA export (Izumikawa et al. 2016) AS-NMD is able to contribute to the variability and regulation of the mRNA expression levels of proteins involved in a broad range of physiological processes including several cellular differentiation programs (Pimentel et al. 2014; Wong et al. 2013; Zheng 2016).

1.4.4 NMD activity is important during neurodevelopment

1.4.4.1 Mutations in NMD factors and their impact during brain development

NMD has been identified in all eukaryotic organisms. The intricacy of this pathway and its necessity during development appear to increase with organism complexity, specifically the complexity of the organism's nervous system. This is highlighted through numerous animal studies. In lower eukaryotes such as *S. cerevisae* and *C. elegans*, deletion of genes encoding core NMD factors, Upf1, Upf2 or Upf3 led to widespread changes in transcriptome or morphogenic defects limited to the reproductive organs, however, they had no effect on growth and viability (He et al. 2003; Pulak & Anderson 1993). Comparatively, Upf1 and Upf2 were both necessary for viability in *Drosophila melanogaster* (*D. melanogaster*) and *Dani rerio* (*D. rerio*). Moreover, non-lethal deletions in NMD factor genes, *Smg1, Upf2* or *Smg6* disrupted the formation of the neuromuscular junction synapse structure, reduced neurotransmission responses and reduced synaptic vesicle cycling in *D. melanogastor* (Frizzell et al. 2012; Metzstein & Krasnow 2006) and caused aberrant eye and brain patterning in *D. rerio* (Wittkopp et al. 2009). Lastly, in *Mus musculus* (*M. musculus*) inhibition of core NMD factors; Upf1, Upf2 and Smg1 resulted in early embryonic lethality (McIlwain et al. 2010; Medghalchi et al. 2001; Weischenfeldt et al. 2012) while haploinsufficiency of EJC factor Magoh caused microcephaly (Silver et al. 2010).

Perhaps more interesting, is the *Upf3b* knockout mouse model, which to date is the only reported viable constitutive knockout of an NMD factor in mice (Huang et al. 2018). Although viable, UPF3B null mice showed defects in neurogenesis and dendritic spine maturation. These mice also display learning, memory and behavioural abnormalities which are comparable to the spectrum of human neuropsychiatric and neurodevelopmental conditions observed in patients with complete loss of function mutations in *UPF3B* (Addington et al. 2011; Huang et al. 2018). In corroboration with this phenotype data, transcriptome profiling data of LCLs with pathogenic

loss of function *UPF3B* mutations revealed stabilisation of several NMD-feature containing transcripts which are known to be highly expressed in the brain and important for neuronal development (Nguyen et al. 2012).

1.4.4.2 NMD regulates axon guidance

The involvement of NMD in axon guidance is a key example of its importance in the later stages of neural development. Axon guidance is the process by which axons are guided to ultimately form connections with their synaptic targets. A well-studied mechanism of axon guidance in mouse involves the cell-surface roundabout proteins (Robo) located on the tips of elongating axons (Chen et al. 2008; Colak et al. 2013; Jaworski et al. 2010; Long et al. 2004; Sabatier et al. 2004).

Robo3 exists as two isoforms; *Robo3.1* and the alternatively spliced *Robo3.2* which retains a PTC containing intron, thus rendering *Robo3.2* an NMD target (Colak et al. 2013). When axons cross the midline, they are first attracted toward the midline through expression of Robo3.1, while *Robo3.2* remains translationally silent (evades NMD) (Figure 1.12). After midline crossing, the axons are repelled from the midline due to loss of Robo3.1 protein expression and translational activation of the *Robo3.2*. Since *Robo3.2* is an NMD target it is quickly degraded by the NMD machinery, limiting its expression and allowing only short bursts of Robo3.2 protein production which is necessary to support midline repulsion through activating other proteins such as; Robo1 and Robo2 (Figure 1.12) (Colak et al. 2013). This coupling of translational repression and activation with NMD highlights the precise and intricate role NMD activity can play during brain development.

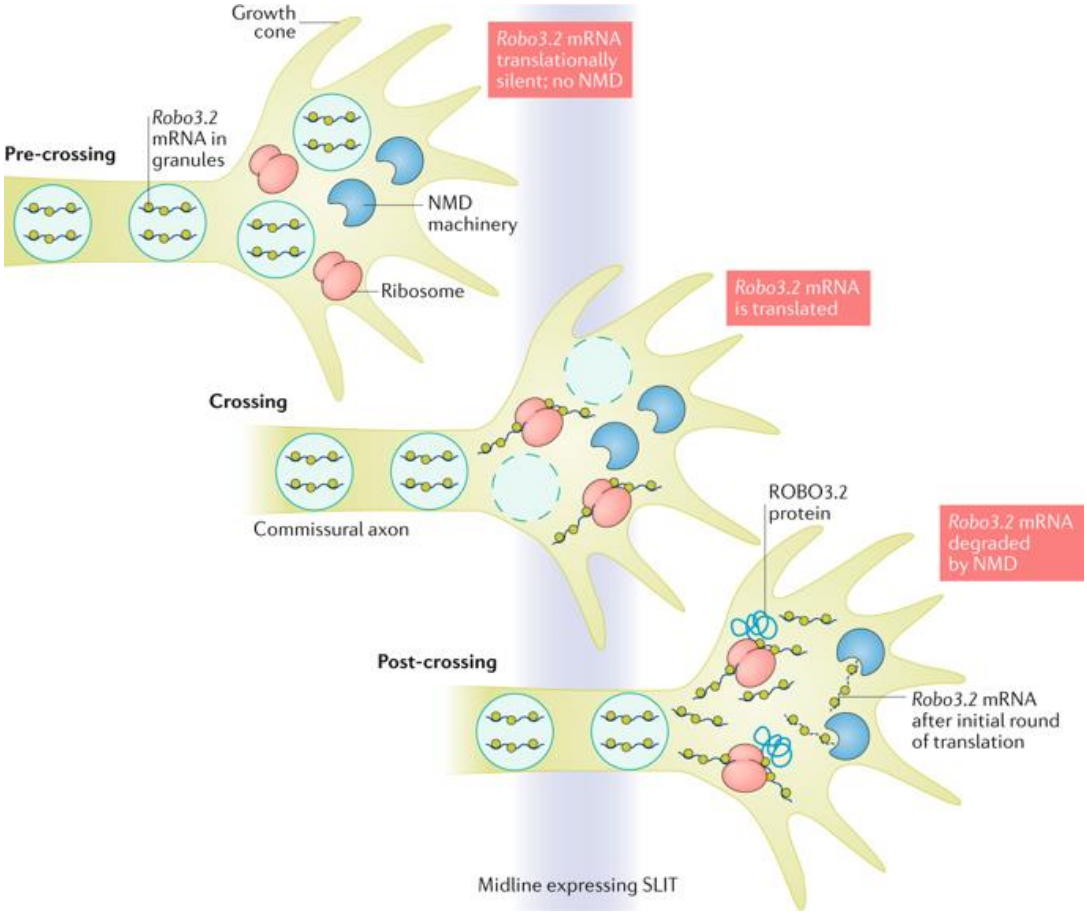


Figure 1.12: NMD controls mRNA levels and protein expression during axon guidance.

In spinal commissural neurons, the axons express Roundabout 3.2 (*Robo3.2*) which is an alternatively spliced transcript encoded by the *Robo3* gene containing a PTC harbouring intron. The ROBO3.2 protein enhances the activity of other ROBO proteins, such as ROBO1 and ROBO2 which mediate repulsion from the midline. When axons are undergoing migration to the midline (pre-crossing, top image), *Robo3.2* mRNA accumulates in RNA granules in a non-translated state. When the axon encounters the midline (middle image), midline-derived factors trigger the translation of *Robo3.2* mRNA. Due to its PTC, translation of *Robo3.2* mRNA also triggers its decay via NMD (post-crossing, bottom image), thereby allowing only a short burst of ROBO3.2 protein production to enhance the function of ROBO1 and ROBO2 and enable the axon to be repelled from the midline. Figure taken from (Jaffrey & Wilkinson 2018).

1.4.4.3 The role of NMD in intellectual disability

The X-linked gene *UPF3B* was the first member of the NMD pathway to be implicated in a neurodevelopmental disorder. Initially individuals with both syndromic and non-syndromic intellectual disability (ID) were identified to carry pathogenic loss of function variants in *UPF3B* (Tarpey et al. 2007). Individuals affected with such genetic variants present with a highly heterogenous range of phenotypes which can vary even within families, these include; autism spectrum disorder, ID, childhood onset schizophrenia and attention deficit hyperactivity disorder (ADHD) (Addington et al. 2010; Laumonnier et al. 2010; Lynch et al. 2012; Tarpey et al. 2007).

The broad range of clinical features and severity observed as a result of *UPF3B* deletions has been hypothesised to be due to a compensatory mechanism involving the *UPF3B* paralog, *UPF3A*. UPF3B protein is more highly expressed than UPF3A and displays a greater binding affinity for UPF2. Therefore, under normal circumstances, UPF3B preferentially binds UPF2 to strongly activate NMD, whilst UPF3A is excluded from UPF2 interactions and subject to rapid turnover (Figure 1.13). Upon depletion of UPF3B, e.g. that observed in individuals carrying loss of function variants in *UPF3B*, UPF3A is stabilised and can act in place of UPF3B to facilitate NMD, albeit less efficiently (Figure 1.13) (Chan et al. 2009). Furthermore, when UPF3B is depleted the extent of UPF3A stabilisation inversely correlates with the extent of transcriptome deregulation and severity of the patient's neurological phenotypes (Chan et al. 2009; Nguyen et al. 2012).

Interestingly, a single amino acid substitution in the EJC-binding domain of UPF3A is sufficient to convert this protein into a potent NMD enhancer and vice versa in UPF3B (Shum et al. 2016). Furthermore, inherently high levels of UPF3A have been observed to have an

antagonistic effect on NMD output through an increased level of UPF3A outcompeting UPF3B for UPF2 binding (Chan et al. 2009; Neu-Yilik et al. 2017).

So far, the function of protein translated from full length *UPF3A* transcript (9 exons) has been discussed, however, a shorter *UPF3A* isoform lacking exon 4 (UPF3A-S) has also been reported. Without inclusion of exon 4, the UPF3A-S protein is unable to bind UPF2 and therefore cannot partake in classical NMD (Andersen et al. 2006; Serin et al. 2001; Shum et al. 2016). Intriguingly, the ratio of *UPF3A* to *UPF3A-S* mRNA was found to vary in different tissues. Together, these examples show that UPF3B and UPF3A act in a tightly controlled regulatory switch which has the potential to regulate NMD activity in a cell and tissue specific manner.

In addition to UPF3B, CNVs in 6 known NMD and EJC genes, namely; *UPF2*, *UPF3A*, *RBM8A*, *SMG6*, *EIF4A3* and *RNPS1* were linked to various neurodevelopmental disorders (Table 1.2) (Nguyen et al. 2013). This suggests that in the broader sense, any alteration to NMD can culminate in the presentation of a neurological disorder. This data also highlights that brain development and function is highly dependent on NMD activity. This brain sensitivity could be due to (1) the sheer complexity of the neuronal transcriptome, meaning a greater volume of NMD targets to be regulated and/or (2) NMD being more active in the brain compared to other tissues. The creation of appropriate tools and methods to investigate endogenous NMD activity across cells and tissues will help resolve this uncertainty and enlighten our understanding on the role of NMD in brain development and function.

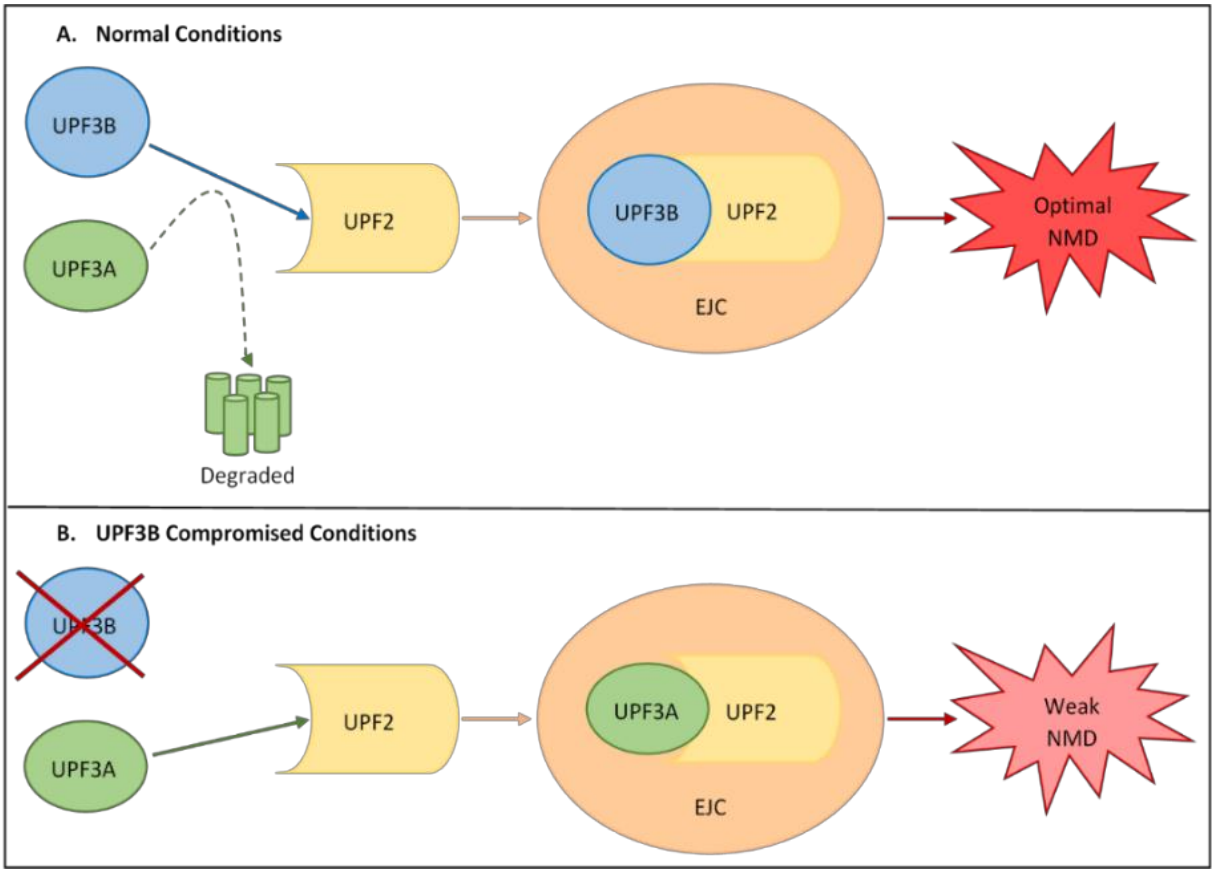


Figure 1.13: UPF3A compensates for loss of UPF3B to induce NMD.

(A) When UPF3B protein is accessible it preferentially binds UPF2 (over UPF3A) to incorporate into the EJC, this results in the rapid turnover of UPF3A and optimal NMD activity. **(B)** When UPF3B levels are compromised, such as loss of function *UPF3B* mutations, UPF3A binds UPF2 and is incorporated into the EJC where it can induce weak NMD activity.

Table 1.2: Human neurodevelopmental diseases caused by mutations in NMD factors. Table adapted from (Jaffrey & Wilkinson 2018).

NMD factor	**Disease(s)**	**Evidence**	**References**
UPF2	Neurodevelopmental disorders	CNV gain/loss	(Gulsuner et al. 2013; Nguyen et al. 2013)
UPF3A	Neurodevelopmental disorders	CNV loss	(Coe et al. 2014; Nguyen et al. 2013)
UPF3B	Intellectual disability, autism spectrum disorder, schizophrenia, and attention deficit hyperactivity disorder.	Pedigree analysis Various mutations	(Addington et al. 2011; Laumonnier et al. 2010; Lynch et al. 2012; Tarpey et al. 2007; Xu et al. 2013)
SMG6	Neurodevelopmental disorders	CNV gain	(Coe et al. 2014; Nguyen et al. 2013)
RBM8A	Neurodevelopmental disorders Richieri-Costa-Pereira syndrome TAR syndrome	CNV gain/loss 1q21.1 del 1q21.1 del + RBM8A mutation	(Brunetti-Pierri et al. 2008; Coe et al. 2014; Mefford et al. 2008; Nguyen et al. 2013; Rosenfeld et al. 2012)
EIF4A3	Neurodevelopmental disorders Richieri-Costa-Pereira syndrome	CNV gain 5'UTR repeat	(Coe et al. 2014; Favaro et al. 2014; Nguyen et al. 2013)
RNPS1	Neurodevelopmental disorder	CNV gain	(Coe et al. 2014; Nguyen et al. 2013)

1.5 The dynamic nature of NMD

Through discovery as a cellular surveillance mechanism, NMD was traditionally viewed as a static pathway operating basally in the background of cells. It is now understood that the NMD pathway is in fact highly regulated, and can differ in response to cell state, type or developmental stage can and even vary across individuals. Because of this knowledge, NMD is now identified as a 'gene expression tool' employed to shape transcriptome diversity across different cell and developmental contexts, which better aligns with its complex involvement during cellular response mechanisms and development.

1.5.1 Cell-specific NMD activity

Cell-specific NMD was highlighted when PTC-type mutations targeted by NMD in *CFTR* and *HBB*, as well as five physiologic NMD factors were shown to have differential expression across HeLa, CFP15a, CFP15b, CFP22a and MCF7 cell types and even among cells derived from the same cell type (Linde, Liat et al. 2007; Linde, L. et al. 2007). In corroboration with this, transcriptome studies showed little overlap between physiological transcripts regulated by NMD in LCLs compared to HeLa cells (Nguyen et al. 2012).

One contributor to the difference in NMD regulated transcripts across cell types is the relative abundance of various NMD factors and antagonists in each cell type (Jolly et al. 2013; Viegas et al. 2007). Another contributor, is the activity of cellular regulatory circuits which can operate based on cell type, state or developmental stage to influence NMD activity, an example of such a circuit is the one NMD shares with miRNA-128 (miR-128) (Bruno et al. 2011; Karam & Wilkinson 2012). miR-128 and the core NMD factor UPF1 partake in a conserved negative-feedback loop, where miR-128 represses NMD activity to activate gene expression. This circuit is enriched in neural cells during brain development to promote neural differentiation, and was also identified in cells of the thymus where a similar mechanism involving UPF1 and miR-128

potentially regulates T-cell development (Frischmeyer-Guerrerio et al. 2011; Kisielow et al. 2001).

1.5.2 Tissue-specific NMD activity

Tissue-specific NMD activity has been observed using mouse models carrying NMD targeted PTC-type mutations in genes such as *Men1* and *Cln1* (Thada et al. 2016; Zetoune et al. 2008). In both cases it was found that the relative expression of mRNA from these genes varied widely amongst different tissue samples, however the groups of tissues that showed high and low NMD activity in these studies did not overlap. This indicates that in addition to the tissue specificity of NMD, the level of NMD activity can also be influenced by the transcript being targeted.

The effects of tissue-specific NMD can also be seen through human disease pathology where NMD targeted PTC-type mutations give rise to a phenotype in one tissue and not another. For example, PTC-type mutations in the Collagen X (*COL10A1)* gene can cause haploinsufficiency resulting in metaphyseal chondrodysplasia Schmid type (MCDS). The phenotype of MCDS is restricted to cartilage, where it is shown that NMD of the variant transcript is highly efficient when compared to other non-cartilage cells (Bateman et al. 2003; Chan et al. 1998; Tan et al. 2008). Another example of this has been observed for a family affected by sudden cardiac death, dilated cardiomyopathy, and rhythm disturbances, where affected individuals have been shown to carry a PTC-type mutation in Lamin A (*LMNA*) (Geiger et al. 2008). In this study *LMNA* variant transcript was found to be significantly downregulated compared to its wild-type counterpart in explanted myocardial tissue but not in cultured fibroblasts, which aligned with the disease phenotype.

1.5.3 Inter-individual NMD variation

The efficiency of NMD is also variable across individuals. The main line of evidence for this is the heterogeneity in clinical severity of patients who harbour the same PTC-type mutation. For example, a number of individuals carrying at least one NMD targeted W1282X PTC-type mutation in *CFTR* showed varying responses to treatment with gentamicin, where those with higher basal levels of variant *CFTR* transcript, i.e. lower NMD efficiency, responded more favourably to the treatment (Linde, L. et al. 2007). Another example is two patients who carry the same PTC-type variant in the X-linked dystrophin gene (*DMD*), yet phenotypically one presents with DMD whilst the other presents with the less severe BMD. This difference in severity has been attributed to the less affected individual showing a lower inherent level of NMD resulting in accumulation of the variant transcript and subsequently expression of an increased amount of truncated, partially functional protein (Kerr et al. 2001).

Recently large-scale genome and transcriptome sequencing studies have estimated that any healthy genome contains between 100 and 200 PTC-type loss of function variants. Of these, only 25–32% of the variants which were predicted to trigger NMD were observed to be consistently downregulated in individuals across the sample population, providing further evidence of inter-individual NMD variation (Lappalainen et al. 2013; MacArthur et al. 2012; The Genomes Project et al. 2012). Most of these are common variants suggesting that individual NMD efficiency can contribute to the diversity of normal phenotypic traits. A minor subset of these variants exists at a low frequency within the human population. These are considered to be rare loss of function variants and are predicted to be efficiently targeted by NMD (Lim et al. 2013; Montgomery et al. 2011; Sulem et al. 2015). Interestingly we may be able to implement these sequencing methods to identify and quantify the expression of an individual's rare PTC-type mutations to gauge that individual's unique NMD efficiency. If the individual also carries a heterogenous disease-causing PTC-type mutation we can extrapolate clinical severity to develop personalised therapy (Figure 1.14).

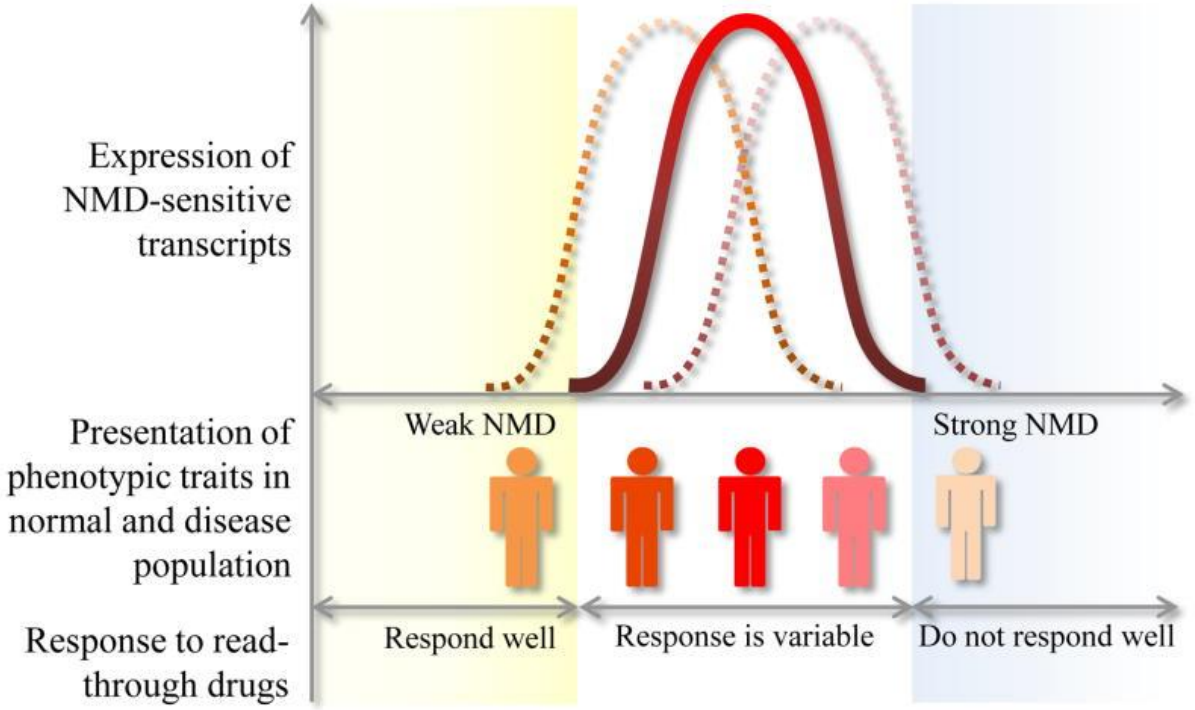


Figure 1.14: A schematic illustration of inter-individual NMD strength and is effect on phenotypic trait.

Basal NMD efficiency is variable across different individuals. This variability can impact the severity of clinical phenotypes and an individual's response to PTC read-through therapies. Figure taken from (Nguyen et al. 2014).

1.6 Current NMD reporter systems

To gain a greater understanding of the involvement of NMD in development and disease it is necessary to have tools which can capture the complex and dynamic nature of NMD. The discussed data highlights indirect evidence of the variability of NMD across cells, tissues, and individuals. Unsurprisingly, many have desired to obtain methods to measure NMD more directly across these scales, predominantly in cells and tissues. The simplest and most common way used to measure NMD activity is to quantify (1) the expression of NMD factors themselves and (2) the expression of known NMD target transcripts. Whilst such approaches have been utilised frequently as a proxy for NMD activity, they are reliant on the assumption that NMD activity can be predicted based on the expression of certain NMD factors, and that selected NMD targeted transcripts are regulated in the same way in different cells. Through the evolution of NMD research discussed in this chapter, both these assumptions are now known to be untrue.

Firstly, the existence of alternative NMD pathways functioning independently of individual NMD factors means that using the expression of any subset of NMD factors as a proxy is fundamentally problematic. This is further influenced by the fact that NMD factors are themselves regulated post-transcriptionally and by differential cellular localisations. Secondly, it is evident that any NMD target, endogenous or caused by genetic error, can be differentially regulated across different cells and tissues, e.g. based even on upstream transcriptional differences. And lastly these techniques represent bulk-cell preparations of cellular biomolecules and therefore lack single cell resolution and the ability to capture live cell dynamic changes. With this in mind, a number of NMD quantification methods and tools have been developed to overcome at least some of these limitations.

1.6.1 RNA quantification based NMD reporters

To study mammalian NMD, several labs have developed synthetic NMD reporter transgenes which enable experimental control on expression levels. The established method has been

transient introduction of a pair of reporter transgenes, one expressing a wild-type transcript and the other expressing the same transcript containing a PTC-type mutation. Typically, the PTC-type mutation has been derived from naturally occurring genetic variants known to cause disease (e.g. within *HBB*, *TPI*, *MUP*) or arise through chromosomal rearrangement during normal development (e.g. *TCR-β*) and have been long identified as NMD targets (Baserga & Benz 1988; Belgrader & Maquat 1994; Buhler et al. 2004; Buhler et al. 2006; Carter et al. 1995; Cheng et al. 1990). Although an enormous amount of knowledge has been gained from this approach, measuring output from a homogenous population of cells based on RNA quantification methods such as northern blot analysis and qPCR, fails to capture the dynamic nature of NMD and therefore limits the information output. Furthermore, it has been observed that not all cell lines efficiently degrade NMD reporter constructs when expressed transiently compared to when expressed following stable genomic integration (Gerbracht et al. 2017).

1.6.2 High-throughput NMD reporters

Since identifying NMD as an important disease modulator, discovering small molecules that alter the efficiency of this pathway has been of considerable pharmaceutical interest. This led to the development of stable fluorescent and bioluminescent high-throughput NMD reporter systems which can measure NMD activity through either flow cytometry, spectrofluorometry, fluorescence microscopy or luciferase assays (Alexandrov et al. 2017; Nickless et al. 2014; Paillusson et al. 2005; Welch et al. 2007). The major advantage of these systems is the ability to facilitate large-scale drug screens for NMD modifiers or forward genetic screens for human NMD factors. To date many NMD inhibitors have been discovered through such NMD reporter systems, including the previously discussed drug, ataluren which resulted in the development of an FDA approved treatment for DMD (Cheruiyot et al. 2018; Durand et al. 2007; Feng et al. 2015; Keeling et al. 2013; Martin et al. 2014; Popp & Maquat 2015; Welch et al. 2007).

1.6.3 NMD reporter systems with single cell resolution

Cell and tissue specific NMD activity is a major contributor to the physiological role of NMD and its involvement in disease severity. To further understand this variability, it is necessary to reliably report on NMD activity at a single cell level. Fluorescence provides an ideal output for single cell quantification as analysis methods such as fluorescent activated cell sorting and fluorescence microscopy can be used. The first dual-fluorescent NMD reporter system utilised the fluorescent proteins TagGFP2 (green) and Katushka (far-red) (Pereverzev et al. 2015). In this system TagGFP2 was encoded by an NMD-targeted mRNA, and Katushka was used as an internal expression control, allowing a ratiometric NMD read-out (Pereverzev et al. 2015). This reporter was used to reveal the heterogeneity in NMD efficiency across several different cell lines (Gerbracht, Boehm & Gehring 2017). However, the potential and reliability of this system is limited by transient transfection efficiency.

A noted issue with stable fluorescent reporters has been a low signal intensity (Paillusson et al. 2005). This was addressed in a recently developed, second dual-fluorescent NMD reporter system, known as the fireworks dual-fluorescent NMD reporter system (Alexandrov, Shu & Steitz 2017). The design featured random chromosomal integration of the reporter transgenes and facilitated fluorescence signal amplification (without toxicity) by incorporating translation of multiple tandemly repeated fluorescent proteins, which in the cell were proteolytically released from the long polypeptide via the encoded tobacco etch virus (TEV) protease. Together fluorescent amplification from this reporter and sequential rounds of enrichment for functional small guide RNAs operated as a functional forward genetic screen to identify factors that modulate NMD in human cells. Although an effective genetic screening technique, a major limitation of this reporter in terms of investigating endogenous NMD levels is that by design it is so far limited to validated use in cancerous HeLa cell line.

Ultimately there is scope for an improved NMD reporter system that ideally allows for, (1) visualisation of NMD at a single cell level in complex tissues (e.g. the brain), (2) live cell measurements, (3) single copy, directed chromosomal integration (i.e. at a so called 'safe-haven loci'), (4) use in non-cancerous cells and (5) use in germ-line competent pluripotent ESCs which opens the potential to study NMD across a variety of developmental processes and cell types; whether via *in vitro* differentiation studies or through the generations of ESC derived animal models.

1.7 Hypothesis and aims

The clinical relevance of NMD in neurodevelopmental disorders is highlighted by the identification of single nucleotide pathogenic variants in the NMD factor *UPF3B* (Laumonnier et al. 2010; Nguyen, Wilkinson & Gecz 2014; Tarpey et al. 2007), and identification of CNVs encompassing *UPF2* or other NMD factors (Johnson et al. 2019; Nguyen et al. 2013; Nguyen, Wilkinson & Gecz 2014). These individuals often present with varying levels of ID, autism spectrum disorder and speech impairments. Although NMD is a crucial post-transcriptional regulator, the mechanism by which dysfunctional NMD leads to impaired brain development and function remains unknown.

Hypothesis 1: The identification and analysis of novel single gene variants which disrupt genes encoding NMD factors can refine the role of NMD in neurodevelopmental disorders.

Aim 1: To use molecular methods to characterise novel genetic variants in genes encoding NMD factors which have been identified in patients with neurodevelopmental disorders and to assess their impact on NMD factor expression and the function of the NMD pathway.

It has been proven that the NMD pathway is involved in neurodevelopmental disorders such as ID and autism spectrum disorder, however, the endogenous role of NMD during brain

development remains largely unexplored. It is likely that the heterogeneity of these disorders and the complex role NMD appears to play in their pathogenicity stems from the cell type, tissue type and inter-individual variation of the NMD pathway. Unfortunately, without an appropriate single cell NMD reporter system it is difficult to investigate this possibility in detail.

Hypothesis 2: The dynamic activity of NMD can be visualised with single cell resolution in live cells, *in vitro*, and *in vivo* using a single NMD reporter transgene.

Aim 2: To design and engineer an NMD reporter transgene which allows visualisation and quantification of endogenous NMD activity at a single cell level *in vitro* within germ-line competent mouse embryonic stem cells and which has the capacity to be developed into a transgenic NMD reporter mouse model.

Chapter Two: Material and Methods

2.1 General solutions

Phosphate-buffered saline (PBS): 0.137 M NaCl, 0.0027 M KCL, 0.01 M Na_2HPO_4, 0.002 M KH_2PO_4, pH 7.4.

Phosphate-buffered saline with Tween 20 (PBST): 0.05% (volume/volume; v/v) Tween 20 in PBS. pH 7.4.

Tris-buffered saline (TBS): 0.05 M Tris-Cl, 0.15 M NaCl, pH 7.4.

Tris-buffered saline with Tween 20 (TBST): 0.05% (v/v) Tween 20 in TBS. pH 7.4.

Tris borate EDTA (TBE): 0.089 M Tris base, 0.089 M boric acid, 0.002 M EDTA, pH 7.6.

Luria broth (LB): 1% (weight/volume; w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl. pH7.4. Autoclaved.

LB agar: 1.5% (w/v) agar in LB. Autoclaved.

NuPage MOPS running buffer (20X) (Cat no. NP001, Invitrogen): 0.05 M MOPS, 0.05 M Tris Base, 0.1% SDS, 0.001 M EDTA, pH 7.7.

Towbin transfer buffer: 0.025 M tris base, 0.192 M glycine, 20% (v/v) methanol.

Mild stripping buffer: 0.2 M glycine, 0.00 4M SDS, l% (v/v) Tween20, pH2.2.

Borate buffer: 0.05 M boric acid, 0.0225 M sodium tetraborate. pH 8.5.

RIPA buffer: 0.0653 M tris, 0.15 M NaCl, 1% (v/v) Nonidet P-40.

2.2 General molecular methods

2.2.1 Genomic DNA isolation

Genomic DNA (gDNA) was extracted from cell pellets of cultured cells using the QuickExtract™ DNA Extraction solution (Cat no. QE09050, Epicentre) as per manufacturer's instructions. For mouse gDNA, quality was assessed using PCR amplification of a genomic region within *KDM5C* using '*KDM5C* F*'* and *'KDM5C* R*'* PCR primers (Table 2.1) under standard conditions as described in Section 2.2.2 and PCR product was visualised by agarose gel electrophoresis (Section 2.2.5).

2.2.2 Polymerase chain reaction (PCR)

Unless otherwise stated PCR was conducted using Taq DNA polymerase (Cat no. 11146173001, Roche, NSW, Australia) as per manufacturer's instructions. Specific single stranded DNA primers used in this thesis are summarised in Table 2.1. PCRs were run on a Mastercycler® nexus GX2 thermocycler (Cat no. 6336000015, Eppendorf, Germany) under the following conditions: initial denaturation at 94°C for 2 minutes followed by 35 cycles of denaturation at 94°C for 20 seconds, annealing at 60°C for 60 seconds, extension at 72°C for 45 seconds and finally extension at 72°C for 7 minutes.

Table 2.1: Description of specific single stranded DNA primers used in this thesis.

Use	**Primer Name**	**Primer sequence (5'–3')**
PCR	*UPF2* F	CAGGAAGAAGTTGGTACGGGC
	UPF2 R	AACCGTCCACATGTCTCCAGC
PCR	*UPF3B* Ex5 F	TGACATCTACTCCAGAGACAC

	UPF3B Ex9 R	GGCTCTTTCATCACTGAGATTC
PCR	*ESTERASE-D* F	GGAGCTTCCCCAACTCATAAATGCC
	ESTERASE-D R	GCATGATGTCTGATGTGGTCAGTAA
PCR	V1-N nosplice1 F	CTTCTCCATCTCCAGCCT
	V1-N splice1 F	CGCTCCGAAAGTTTCCT
	V1-N splice1 R	TTGCCAAAATGATGAGACAG
PCR	V1-N nosplice2 F	ATGGTTGGGATAAGGCTG
	V1-N splice2 F	ATCAGTGTGGAAGTCTCAGG
	V1-Nsplice2 R	AGCACACAGACCAGCACG
PCR	*Col1a1* geno F	AATCATCCCAGGTGCACAGCATTGCGG
	Col1a1 geno R1	TGGACTACTGCGCCCTACAGATCTGC
	Col1a1 geno R2	CTTTGAGGGCTCATGAACCTCCCAGG
PCR	*KDM5C* F	TTCCTTGCTACGCTCTCACTATGA
	KDM5C R	TCAAATGGGCGTGTGTTACAC
qPCR	*GAPDH* F	TGCACCACCAACTGCTTAGC
	GAPDH R	GGCATGGACTGTGGTCATGAG
qPCR	*ACTB* F	ATGGGTCAGAAGGATTCCTATGTG
	ACTB R	TGTTGAAGGTCTCAAACATGATCTGG
qPCR	*HPRT* F	TGACACTGGCAAAACAATGCA
	HPRT R	GGTCCTTTTCACCAGCAAGCT
qPCR	*UPF1* F	CATCATCCTGTCCTGTGTGC
	UPF1 R	GACGCCATACCTTGCTCTG
qPCR	*UPF2* F	GTTGGTACGGGCACTCTTCAT
	UPF2 R	CCCCCTCAGCATGGAACAAA
qPCR	*UPF3B* F	CTTCAGGGCAAAGAATAGAGAGA
	UPF3B R	TTGACACAAGACTTACTCCTCTG

qPCR	*ATF4* F	CACAACATGACCGAGATGAG
	ATF4 R	CGAAGTCAAACTCTTTCAGATCC
qPCR	*GADD45B* F	ATTGACATCGTCCGGGTATC
	GADD45B R	TCCAGGAATCTGTATGACAG
qPCR	*SNORD* F	ACTCTCTGTCCTAGTCCCAG
	SNORD R	CCTCAGACAGTTCCTTCTGGA
qPCR	*GAS5* F	CTTGCCTGGACCAGCTTAAT
	GAS5 R	CAAGCCGACTCTCCATACCT
qPCR	V1_A q F	TGGGAGTTGAGCAGCCTACC
	V1_A q R	AATGACTTGGCGTTGTTCCG
qPCR	V1+2_C F	ACGGCAACGTCTATATCACC
	V1+2_C R	TTGGACTGGGTGCTCAGG
qPCR	V2_R F	GATGACGACGATAAGGGTGG
	V2_R R	TGGTGCAGATGAGCTTCAGG
qPCR	*HBB* F	AAGGTGAAGTTCGTCGTCCAA
	HBB R	GTACAACGTCAGGTTTACCACCTTT
qPCR	*mUpf1* F	CGCAGGCAGGATCATGGATT
	mUpf1 R	AGCTCGACGCACAAGTTGG
qPCR	*mActb* F	ATGAAGATCCTGACCGAGCG
	mActb R	TACTTGCGCTCAGGAGGAGC
qPCR	*mGas5* F	(Jolly et al. 2013)
	mGas5 R	
qPCR	*mATF4* F	(Jolly et al. 2013)
	mATF4 R	
Sequencing	*UPF2* F	CAGGAAGAAGTTGGTACGGGC
Sequencing	*UPF2* R	AACCGTCCACATGTCTCCAGC

Sequencing	*UPF3B* Ex5 F	TGACATCTACTCCAGAGACAC
Sequencing	V1-N F1	TCTTCGCTATTACGCCAGC
Sequencing	V1-N F2	ATGGGACTTTCCTACTTGGC
Sequencing	V1-N F3	CGCTTGGTTTAATGACGGC
Sequencing	V1-N F4	ATGGTAATCGTGCGAGAGG
Sequencing	V1-N F5	TGGGAGTTGAGCAGCCTACC
Sequencing	V1-N F6	TAAAGTGCGAAAGCGGCG
Sequencing	V1-N F7	CAACTACAACAGCCACAACG
Sequencing	V1-N F8	GACGCTTGATGTTTTCTTTCC
Sequencing	V1-N F9	CACATATTGACCAAATCAGGG
Sequencing	V1-N F10	GCTCGCTTTCTTGCTGTCC
Sequencing	V1-S F2	AATGGAAGCGGGTAGGC
Sequencing	V1-S F3	GGTTGAGGACAAACTCTTCG
Sequencing	V1-S F4	TATCATGTCTGGATCTGCG
Sequencing	V1-C F2	GTCAATGACGGTAAATGGC
Sequencing	V1-C3b	GAGTTGCTGAGCACGGC
Sequencing	V1-C F5	CCACAAGTTCAGCGTGTCC
Sequencing	V1-C F6	GCCGACAAGCAGAAGAACG
Sequencing	V1-T F1	GGGTTATTGTCTCATGAGCG
Sequencing	V1-T F5	GAAGAGGAAAGTGGTGAGC
Sequencing	V1-T F6	ATGGCACCGGCAGCACC
Sequencing	V1-T F7	TGTTCCTGTACGGCATGG
Sequencing	*HBB* F2	AACTTCAGGGTGAGTCTATGG
Sequencing	*HBB* F3	CACATATTGACCAAATCAGGG
Sequencing	*HBB* F4	ATTCTGAGTCCAAGCTAGGC
Sequencing	*HBB* R1	CGTCCCATAGACTCACCC

2.2.3 PCR purification

When necessary the MinElute PCR purification kit (Cat no. 28004, QIAGEN, VIC, Australia) was used as per manufacturer's instructions to remove components of the PCR reaction that may interfere with subsequent sequencing reactions. Purified PCR product was quantified using a UV spectrophotometer (Nanodrop 1000, Thermo Fisher Scientific, MA, USA) and stored at 4°C.

2.2.4 Sanger sequencing

BigDye™ Terminator sequencing reactions were used to sequence DNA (plasmid or purified PCR products) and performed using BigDye™ V3.1 (Cat no. 4337455, Applied Biosystems, CA, USA) as per manufacturer's instructions. Reaction products were sent to the Australian Genome Research Facility Ltd. (AGRF, Adelaide, SA, Australia) for precipitation and capillary based sequencing. Sequencing results were interpreted using the free software, ApE (A plasmid editor) (Davis 2019) and/or SeqMan Pro program from the DNASTAR Laser Gene software package version 10.1.2 (DNASTAR Inc. USA). In cases where sequencing traces of purified PCR product was used to quantify the ratio of wild-type cDNA to variant cDNA, analysis was performed using the free online tool, TIDE (Brinkman et al. 2014).

2.2.5 Agarose gel electrophoresis

DNA samples (plasmid DNA, PCR amplified products, restriction endonuclease digestion products or 1 kb plus DNA ladder (Cat no. 10787018, Invitrogen, CA, USA)) were diluted in loading buffer then loaded into the wells of an agarose gel (1–2% (w/v) agarose in TBE-buffer with the addition of 0.2 µg/mL ethidium bromide (Cat no. 1610433, Bio-Rad, CA, USA)). DNA was separated at 100 V using a Mini Sub-cell ® GT electrophoresis tank (Bio-Rad) containing 1X TBE and visualised under UV (Syngene INGENIUS LHR: Gel Documentation System,

LabGear Australia) for imaging, or using a UV transilluminator (Cat no. TFX-20 MX, Sigma-Aldrich, NSW, Australia) for DNA extraction using a scalpel.

2.3 Plasmid generation

The Plasmids described in Table 2.2 were either purchased from Genscript (USA), or engineered in-house using the general methods described below and further detailed in the appropriate results chapters. Once in hand sequence validation was carried out using diagnostic restriction endonuclease digests and Sanger sequencing as described in the following sections (Sections 2.3.1–2.3.6).

2.3.1 DNA modification

All DNA modifying enzymes, including restriction endonucleases, T4 DNA ligases and Shrimp Alkaline Phosphatase (rSAP) were purchased from New England Biolabs (NEB) and used as per manufacturer's instructions.

2.3.2 Gel purification

Purification of DNA excised from agarose gels was achieved using the QIAquick gel extraction kit (Cat no. 28506, QIAGEN) as per manufacturer's instructions. Purified DNA product was quantified using a UV spectrophotometer (Nanodrop 1000, Thermo Fisher Scientific) and stored at 4°C.

2.3.3 Growth and transformation of Escherichia Coli (E. coli)

NEB 5-alpha Competent *E.coli* cells (Cat no. C2987H, NEB, MA, USA), a derivative of DH5α were commercially purchased and grown at 37°C in LB or on LB agar bacterial plates, supplemented with or without antibiotics (i.e. 100 μg/mL ampicillin (Cat no. A0166, Sigma-

Aldrich) or 50 µg/mL kanamycin (Cat no. K1377, Sigma-Aldrich)). Transformation of *E. coli* was achieved using the heat shock method as per manufacturer's instructions (NEB).

2.3.4 Colony PCR

To screen bacterial transformants for the presence of correct plasmid DNA using PCR, transformed bacterial colonies were transferred from an agar plate using a pipette tip and inoculated onto another agar antibiotic plate (patch plate). The remaining bacteria on the pipette tip was used as template in a PCR reaction (Section 2.2.2). PCR products were visualised by agarose gel electrophoresis (Section 2.2.5) and analysed for presence of desired PCR amplified DNA products.

2.3.5 Plasmid DNA isolation

Small scale plasmid DNA isolation was achieved using the Wizard® *Plus* SV Minipreps DNA Purification Systems (Cat no. A1460, Promega, WI, USA) as per manufacturer's instructions. Large scale plasmid DNA isolation was achieved using the EndoFree Plasmid Maxi Kit (Cat no. 12362, QIAGEN) as per manufacturer's instructions. The sequence of isolated plasmid DNA was validated using restriction endonuclease digests and Sanger sequencing methods (Sections 2.2.4 & 2.2.5).

2.3.6 Generating bacterial glycerol stocks

Glycerol stocks were generated for all bacterial transformants of interest by adding 700 µl of fresh bacterial stock from LB cultures to 500 µl of 80% glycerol (Cat no. G5516, Sigma-Aldrich) and immediately storing at -80°C.

Table 2.2: A description of plasmids used in this thesis and the protein(s) they are designed to encode

Plasmid	Supplier	Expression cassette(s) included in plasmid				Encoded Protein(s)	Expected protein size(s)
		Selection (S)	Control (C)	NMD (N)	TetR Responder (T)		
pUC57-$S^{V1.0}$	Genscript	$S^{V1.0}$	-	-	-	-	-
pUC57-$C^{V1.0}$	Genscript	-	$C^{V1.0}$	-	-	CFP^{NLS}	~28 kDa
pUC57-$N^{V1.0}$	Genscript	-	-	$N^{V1.0}$	-	TetR:$EGFP^{NLS}$	~52 kDa
pUC57-TetO-$T^{V1.0}$	Genscript	-	-	-	$T^{V1.0}$	$tdTomato^{NLS}$	~56 kDa
pUC57-$CN^{V1.0}$	In-house	-	$C^{V1.0}$	$N^{V1.0}$	-	CFP^{NLS}, TetR:$EGFP^{NLS}$	~28 kDa, ~52 kDa
pUC57-$SNC^{V1.0}$	In-house	$S^{V1.0}$	$C^{V1.0}$	$N^{V1.0}$	-	CFP^{NLS}, TetR:$EGFP^{NLS}$	~28 kDa, ~52 kDa
pUC57-kan-$SC^{V2.0}$	Genscript	$S^{V2.0}$	$C^{V2.0}$	-	-	$^{HA}CFP^{NLS}$:HBB^{WT}	~46 kDa
pUC57-kan-$N^{V2.0}$	Genscript	-	-	$N^{V2.0}$	-	$^{FLAG}YFP^{NLS}$:HBB^{NS39}	~35 kDa
pUC57-kan-$SCN^{V2.0}$	In-house	$S^{V2.0}$	$C^{V2.0}$	$N^{V2.0}$	-	$^{HA}CFP^{NLS}$:HBB^{WT}, $^{FLAG}YFP^{NLS}$:HBB^{NS39}	~46 kDa, ~35 kDa

pUC57-kan-SC$^{V2.1}$	Genscript	S$^{V2.1}$	C$^{V2.1}$		-	^{HA}CFP^{NLS}, HBB^{WT}	~32 kDa, ~ 16kDa
pUC57-kan-N$^{V2.1}$	Genscript			N$^{V2.1}$	-	^{FLAG}YFP^{NLS}, HBB^{NS39}	~32 kDa, ~ 4 kDa
pUC57-kan-SC$^{V2.2}$	Genscript	S$^{V2.2}$	C$^{V2.2}$		-	^{HA}CFP^{NLS}, HBB$^{WT, \Delta E1}$	~32 kDa, ~13 kDa
pUC57-kan-N$^{V2.2}$	Genscript			N$^{V2.2}$	-	^{FLAG}YFP^{NLS}, HBB$^{NS39, \Delta E1}$	~32 kDa, ~1 kDa
pcDNA3	Invitrogen	-	-	-	-	-	-
pTetR	In-house	-	-	-	-	TetR	~23 kDa
PQCXIP	Clontech	-	-	-	-	Puromycin-*N*-acetyltransferase	~22 kDa
pPGK-FLPo-bpA	Addgene	-	-	-	-	FLP recombinase	~49 kDa

The abbreviations used in the above table are as follows; Version (V), cyan fluorescent protein (CFP), enhanced green fluorescent protein (EGFP), yellow fluorescent protein (YFP), Tetracycline repressor protein (TetR), nuclear localisation signal (NLS), wild-type (WT), β-globin (HBB), HA-tag (HA), FLAG-tag (FLAG), Selection Cassette (S), Control Cassette (C), NMD Cassette (N) or TetR Responsder Cassette (T).

2.4 Ethics statement for human participants

This thesis studied one case with a pathogenic *UPF3B* variant (Table 3.1) and five cases with pathogenic *UPF2* variants (Table 3.3). This research has been approved by the Women's and Children's Hospital Research Network Human Research Ethics Committee under approval number 786/7/2020 and The Human Research Ethics Committee of the Royal Children's Hospital HREC37353. Informed consent was also obtained.

2.5 Cell culture

2.5.1 Cell culture surfaces

The cell culture surfaces used in this thesis are summarised in Table 2.3.

Table 2.3: A summary of cell culture surfaces used in this thesis

Culture dish / flask name	**Growth area**	**Growth medium**	**Cat no.**
25 cm^2 flask	25 cm^2	5 mL	430639, Corning
75 cm^2 flask	75 cm^2	15 mL	430641U, Corning
60 mm dish	21 cm^2	5 mL	430196, Corning
100 mm dish	55 cm^2	12 mL	430167, Corning
100 mm (non-TC coated)	55 cm^2	12 mL	CLS430167, Sigma-Aldrich
6 well plate	9.5 cm^2	2 mL/well	3516, Corning
12 well plate	3.8 cm^2	1 mL/well	3513, Corning
24 well plate	1.9 cm^2	0.5 mL/well	3525, Corning

2.5.2 Coverslip preparations

2.5.2.1 Acid washed coverslips

Coverslips (Cat no. G415 or G404, ProSciTech, QLD, Australia) were placed in a solution of 1 M HCL (Cat no. 20252.244, VWR, PA, USA) at 60°C overnight. The following day coverslips were transferred to a 10 cm dish and washed continuously for 5 minutes in Milli-Q water, they were then washed once in 100% ethanol (Cat no. 20821.321, VWR) and 100% ethanol was again added to the dish. The dish was then moved to a tissue culture hood and using tweezers the coverslips were removed from the ethanol and left to dry. Coverslips were stored at room temperature.

2.5.2.2 Poly-L-lysine (PLL) coated coverslips

Acid washed coverslips were transferred to a 10 cm dish with a solution of 1 mg/mL PLL in 0.1 M borate buffer to immerse the coverslips. These were rocked gently for between 2–16 hours at room temperature. The PLL solution was removed and the coverslips were washed 3 times with PBS and then left to dry prior to use.

2.5.2.3 Poly-L-lysine/laminin (PLL/L) coated coverslips

Acid washed coverslips were transferred to 6 or12 well plates and a solution of 33 µg/mL PLL (1 mg/mL PLL in borate buffer stock diluted in PBS) was added to each well to immerse the coverslip. The plates were incubated at 37°C for between 2–16 hours. The PLL solution was aspirated and the coverslips were washed 3 times with PBS. A 3 µg/mL solution of laminin (Cat no. L2020, Sigma-Aldrich) in PBS was then added to each well and incubated at 37°C for 2 hours. The laminin solution was aspirated, and the dish was used immediately.

2.5.3 Cell culture media preparations

Several cell culture media preparations were used in this thesis, their compositions are summarised once in Table 2.4 and henceforth referred to by their name.

Table 2.4: Composition of cell culture mediums used in this thesis

Medium name	**Composition**
LCL medium	RPMI-1640 (Cat no. R5556, Sigma-Aldrich) supplemented with 10% Fetal Bovine Serum (FBS) (Cat no. 10099141, Gibco, Life Technologies, CA, USA), 100 U/mL Penicillin-Streptomycin (Cat no. 15140122, Gibco, Life Technologies) and 2 mM of Glutamax (Cat no. 35050061, Gibco, Life Technologies).
LCL freezing medium	RPMI-1640 (Cat no. R5556, Sigma-Aldrich) supplemented with 10% Fetal Bovine Serum (FBS) (Cat no. 10099141, Gibco, Life Technologies, CA, USA), 100 U/mL Penicillin-Streptomycin (Cat no. 15140122, Gibco, Life Technologies), 2 mM of Glutamax (Cat no. 35050061, Gibco, Life Technologies) and 10% DMSO (D4540, Sigma-Aldrich).
Standard medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 10% FBS (Cat no. 10099141, Gibco, Life Technologies) +/-100 U/mL Penicillin-Streptomycin (Cat no. 15140-122, Gibco, Life Technologies).
Standard freezing medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 20% FBS (Cat no. 10099141, Gibco, Life Technologies), +/-100 U/mL Penicillin-Streptomycin (Cat no.

	15140-122, Gibco, Life Technologies and 10% DMSO (D4540, Sigma-Aldrich).
FACS medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 5% FBS (Cat no. 10099141, Gibco, Life Technologies),
MEF medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 10% FBS (Cat no. 10099141, Gibco, Life Technologies) and 100 U/mL Penicillin-Streptomycin (Cat no. 15140-122, Gibco, Life Technologies).
MEF freezing medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 25% FBS (Cat no. 10099141, Gibco, Life Technologies) and 10% DMSO (D4540, Sigma-Aldrich).
Incomplete mESC medium	DMEM (Cat no. 1110569-010, Gibco, Life Technologies) supplemented with 15% FBS (Cat no. 10099141, Gibco, Life Technologies), 100 U/mL Penicillin-Streptomycin (Cat no. 15140-122, Gibco, Life Technologies), 1% Non-Essential Amino Acids (Cat no. 11140050, Life-Technologies).
Complete feeder-dependent mESC medium (made fresh to use)	Incomplete mESC medium supplemented with 55 μm (1:1000) β-mercaptoethanol (Cat no. 21985023, Gibco, Life Technologies) and 10^3 U/mL Leukemia inhibitory factor (LIF) (Cat no. ESG1107, Merck Millipore).
Complete feeder-independent mESC medium (made fresh to use)	Incomplete mESC medium supplemented with 55 μm (1:1000) β-mercaptoethanol (Cat no. 21985023, Gibco, Life Technologies) and 10^3 U/mL Leukemia inhibitory factor (LIF) (Cat no. ESG1107, Merck Millipore), 3 μm CHIR99021 (Cat no.

	SML1046, Sigma-Aldrich), 1 μm PD 0325901 (Cat no. PZ0162, Sigma-Aldrich).
mESC EB medium	KnockOut™ DMEM (Cat no. 10829018, Gibco, Life Technologies) supplemented with 15% KnockOut™ Serum Replacement (Cat no. 10828028, Gibco, Life Technologies), 1% Non-Essential Amino Acids (Cat no. 11140050, Life-Technologies) and 1% Glutamax (Cat no. 35050061, Gibco, Life Technologies).
mESC EB medium with 10% FBS	mESC EB medium supplemented with 10% FBS (Cat no. 10099141, Gibco, Life Technologies).
mESC freezing medium	90% FBS (Cat no. 10099141, Gibco, Life Technologies) and 10% DMSO (D4540, Sigma-Aldrich).

2.5.4 Culture of human lymphoblast cell lines (LCLs)

Human lymphoblast cell lines (LCLs) (Neitzel 1986) were maintained in LCL medium. The cells were grown in suspension using an upright 25 or 75 cm^2 flask in a humidified incubator at 37°C with 5% CO_2. Typically, Cultures were subcultured every 72 hours at a density of 2.0 $x10^5$ cells/mL through addition of new media or gentle dissociation of loose aggregates via trituration with a serological pipette. 50% media changes occurred once between subcultures. Frozen cell stocks were maintained in LCL freezing medium.

2.5.5 Culture of primary mouse embryonic fibroblasts (NIH3T3) and human embryonic kidney 293T (HEK293T) cells

NIH3T3 and HEK293T cells were maintained in Standard medium. The cells were grown in a humidified incubator at 37°C with 5% CO_2. Typically, media was changed every other day and cultures were subcultured twice a week using 0.01% Trypsin (Cat no. 15400-054, Gibco, Life Technologies) at an appropriate density (Table 2.5). Frozen cell stocks were maintained in liquid nitrogen in Standard freezing medium.

Table 2.5: A Description of plating densities for cell subculturing

Cell type	**Plating density**
HEK293T	1.33 $x10^4$ cells/cm^2
NIH3T3	1.33 $x10^4$ cells/cm^2

2.5.6 Establishment of stable HEK293T cell lines

Stable HEK293T cell lines (Table 2.6) were generated through random genomic integration of linearised plasmid DNA into the HEK293T cell genome and antibiotic selection. Details of this process are described in the following sections (Sections 2.5.6.1–2.5.6.4).

Table 2.6: A summary of stable transgenic HEK293T cell lines established and used in this thesis

Cell line	**Integrated plasmids**
NMD Reporter$^{V1.0}$ HEK293T cells	pUC57-CN$^{V1.0}$
	pQCXIP
NMD Reporter$^{V2.0}$ HEK293T cells	pUC57-kan-SCN$^{V2.0}$
	pQCXIP

2.5.6.1 Step one: plasmid linearisation

Plasmid DNA (Table 2.6) was linearised by restriction endonuclease digestion as described in Section 2.3.1 and further detailed in Chapter 6.

2.5.6.2 Step two: random DNA integration via transfection

One day prior to transfection of linearised plasmid DNA, HEK293T cells were plated at a density of 6.84 x10^4 cells/cm^2 into 6 well plates in Standard medium without antibiotics. The following day cells were transfected with 1000 ng of linearised plasmid encoding expression cassettes (pUC57-CN$^{V1.0}$ or pUC57-kan-SCN$^{V2.0}$) and 1000 ng of linearised plasmid encoding a puromycin resistance gene (pQCXIP) using lipofectamine 2000 (Cat no. 15338030, Invitrogen) as per manufacturer's instructions. The lipofectamine:DNA ratio used was 3:1. The transfected cells were grown in a humidified incubator at 37°C with 5% CO_2 daily for 48 hours. Medium was changed daily.

2.5.6.3 Step three: puromycin selection

48 hours post transfection 1.5 µg/mL of puromycin (Cat no. P9620-10mL, Sigma-Aldrich) was added to the media. 48 hours later puromycin was withdrawn and the cells were expanded through standard culture methods (Section 2.5.5) for 14 days prior to isolation of fluorescently labelled cells.

2.5.6.4 Step four: flow cytometry – fluorescence activated cell sorting (FACS)

Puromycin resistant cells were passaged as described in Section 2.5.5. To ensure a single cell suspension the cells were washed through a 0.40 µM filter using Standard medium. Cells were centrifuged for 5 minutes at 200 x g and resuspended in 5–15 ml FACs medium at a concentration of $1x10^7$ cells/mL. Cells were sorted to isolate those expressing high levels of cyan fluorescent protein (CFP) using the BDFACSMelody cytometer (ACRF flow and laser-scanning cytometry facility, SAHMRI, Adelaide, Australia). These cells were gated relative to an unlabelled control cell line (parental). Isolated cells were sorted directly into Standard medium and placed on ice. Sorted cells were seeded at high density (approximately $2.0 x10^4$ cells/mL) for continuous culture as described in Section 2.5.5.

2.5.7 Culture of mouse embryonic stem cells (mESCs)

2.5.7.1 Preparation of gelatin coated surfaces

Prior to mESC cell culture, cell culture plates and coverslips were treated with a solution of 0.1% gelatin (Cat no. G9391, Sigma-Aldrich) solution in Milli-Q water (autoclaved and stored at 4°C) was added to culture plates at 1 mL/10 cm^2. Plates were then incubated at 37°C for a minimum of 2 hours. Following incubation, excess gelatin was aspirated, and plates were dried at room temperature. Gelatin coated plates were then used immediately.

2.5.7.2 Culture of feeder-dependent mESCs

2.5.7.2.1 Mouse embryonic fibroblast feeder (MEF) layer

Cryopreserved irradiated MEFs (StemCore, QLD, Australia) were thawed rapidly by addition of pre-warmed (37°C) MEF medium to the cryovial of cells. Cells were then transferred to a 10 mL tube containing pre-warmed (37°C) MEF medium using a transfer pipette and centrifuged at 94 x g for 5 minutes at room temperature. The supernatant was aspirated, and cells were resuspended in MEF medium and replated at 9 $x10^3$ cells/cm^2. The cells were grown in a humidified incubator at 37°C with 5% CO_2. The next day media was replaced with complete feeder-dependent mESC medium and returned to the incubator to equilibrate prior to mESC passage onto these plates.

2.5.7.2.2 Culture conditions for feeder-dependent mESCs

The feeder-dependent FLP-in mESCs (mESCs pre-engineered with *Col1a1* FRT sequences and a promoter-less hygromycin resistance coding sequence, which also lacks a start codon) were kindly gifted from Professor Murray Whitelaw and Doctor David Bersten (The University of Adelaide, SA, Australia) (Bersten et al. 2015). These cells were maintained on a MEF feeder layer in Complete feeder-dependent mESC medium and grown in a humidified incubator at 37°C with 5% CO_2. Media was changed daily, and cells were subcultured every second to third day using Accutase (Cat no. 07920, StemCell Technologies) at a density of 1.3 $x10^4$ cells/cm^2 onto pre-prepared MEF plates.

2.5.7.3 Establishment and culture of feeder-independent mESCs

The experiments detailed in this thesis employ feeder-independent FLP-in mESCs. These cells were established through single cell dissociation of feeder-dependent FLP-in mESCs (Section 2.5.7.2.2) using Accutase (Cat no. 07920, StemCell Technologies, VIC, Australia) and plated onto gelatin coated dishes at a density of 1.3 x 10^4 cells/cm^2 in Complete feeder-independent

mESC medium. The cells were grown in a humidified incubator at 37°C with 5% CO_2. Media was changed daily, and cells were subcultured every second day using Accutase onto gelatin coated plates at a density of 7.6 x10^3 cells/cm^2.

2.5.8 Establishment of stable feeder-independent mESC lines

Stable feeder-independent mESC lines (Table 2.7) were established via Recombination Mediated Cassette Exchange (RMCE) system and antibiotic selection as previously described (Bersten et al. 2015). The methods involved in this process are detailed in the following sections (Sections 2.5.8.1–2.5.8.2).

Table 2.7: A summary of stable transgenic feeder-independent mESC lines established and used in this thesis

Cell line	**Integrated plasmid**
NMD Reporter$^{V1.0}$ mESCs	pUC57-SNC$^{V1.0}$
NMD Reporter$^{V2.0}$ mESCs	pUC57-kan-SCN$^{V2.0}$

2.5.8.1 Step one: electroporation of plasmid DNA

FLP-in feeder-independent mESC were dissociated into a single cell suspension using Accutase (Cat no. 07920, StemCell Technologies). 1.0 x10^7 cells/cuvette (Cat no. 165-2088, Bio-Rad) were pelleted by centrifugation at 300 x g for 5 minutes. Cells were resuspended in PBS, centrifuged again, and resuspended in fresh PBS (800 µl/cuvette) to remove any excess salt. 800 µl of cell/PBS solution, 25 µg of FLP recombinase expression plasmid (pPGK-FLPo-bpA) and 50 µg of reporter plasmid (pUC57-SNC$^{V1.0}$ or pUC57-SCN$^{V2.0}$) was added to the cuvette. Electroporation was conducted at 400 V and 125 µF on the X-Cell Gene Pulser (Bio-Rad). Following electroporation, cells were transferred to a 100 mm gelatin coated dish containing

pre-equilibrated Complete feeder-independent mESC medium. The electroporated cells were grown in a humidified incubator at 37°C with 5% CO_2 for 48 hours. Medium was changed daily.

2.5.8.2 Step two: hygromycin selection

48 hours post electroporation 50 mg/mL hygromycin (Cat no. 10687010, Invitrogen) was added to the media. 6–8 days later, when isolate colonies could be identified, colonies were mechanically dissected and removed from the plate with a pipette and then re-seeded in individual culture wells for further expansion as clonal cell lines. Transgenic feeder-independent mESC lines were expanded through standard culture methods (Section 2.5.7.3) and hygromycin was withdrawn from the media after three passages.

2.5.8.3 Step three: screening transgenic mESCs via PCR amplification of gDNA

gDNA extracted from stable transgenic mESC was used as a template in PCR to confirm that integration of plasmid DNA into the *Col1a1* locus of Flp-in mESCs was in the correct orientation. This PCR was carried out as previously described in Section 2.2.2 with '*Col1a1* geno F' and '*Col1a1* geno R1' PCR primers (Table 2.1). PCR products were visualised using agarose gel electrophoresis (Section 2.2.5). The presence of a 300 bp product would confirm correct genomic orientation of the transgene.

gDNA extracted from transgenic mESCs was also used as a template in PCR to identify the size of integrated plasmid DNA. This PCR was carried out using Expand™ Long template PCR System (Cat no. ELONG-RO, Roche) and '*Col1a1* geno F' and '*Col1a1* geno R2' primers which span the *Col1a1* locus (Table 2.1). The PCR cycle conditions were as follows: initial denaturation at 94°C for 2 minutes then 10 cycles of denaturation at 94°C for 10 seconds, annealing at 60°C for 30 seconds and extension at 68°C for 30 minutes, followed by 25 cycles of denaturation at 94°C for 15 seconds, annealing at 60°C for 30 seconds, extension at 68°C for

30 minutes and lastly a final elongation at 68°C for 7 minutes. PCR products were visualised by agarose gel electrophoresis to assess transgene insertion (Section 2.2.5).

2.5.9 Spontaneous differentiation of mESCs

2.5.9.1 Formation of embryoid bodies

Embryoid bodies (EBs) were established through single cell dissociation of feeder-independent mESCs using Accutase (Cat no. 07920, StemCell Technologies, VIC, Australia), cell suspension was plated into 100 mm non-TC coated (non-adherent) dishes at a density of 2.3 $x10^5$ cells/mL in mESC EB medium. The cells were incubated for 3 days in a humidified incubator at 37°C with 5% CO_2. Medium was changed daily.

2.5.9.2 Spontaneous differentiation to primary germ layers

Following EB formation, EBs were plated in mESC EB medium with 10% FBS onto gelatin coated dishes. EBs were incubated overnight in a humidified incubator at 37°C with 5% CO_2 to allow them to attach to the dish. The following day media was replaced with mESC EB medium. The cells were incubated in a humidified incubator at 37°C with 5% CO_2. Media was replaced every other day. At day 14 after EB differentiation, differentiated cells were collected.

2.5.10 Cell transfection approaches

Several methods were used to introduce DNA and/or siRNA into cells in this thesis, the conditions used for these are summarised in Table 2.8 and detailed in the following sections (Sections 2.4.10.1–2.5.10.3).

Table 2.8: A Summary of cell transfection conditions used in this thesis

Cell Type	**DNA/siRNA delivery method**	**Plating density**	**Concentration of DNA**	**Concentration of siRNA**
HEK293T	Lipofectamine 2000	6.84 x10^4 cells/cm^2	63.16 ng/cm^2	2.11 pmol/cm^2
HEK293T	RNAimax	6.84 x10^4 cells/cm^2	N/A	2.11 pmol/cm
mESCs	Nucleofection	2.0 x10^6 cells/cuvette	2000 ng/cuvette	50 pmol/cuvette
mESC	RNAimax	8.42 x10^2 cells/cm^2	N/A	2.11 pmol/cm^2
NIH3T3	Nucleofection	1.0 x10^6 cells/cuvette	4000 ng/cuvette	30 pmol/cuvette

2.5.10.1 Introduction of siRNA using RNAimax

One day prior to transfection cells were plated into 6 or 12 well plates with or without coverslips in culture medium without antibiotics. The following day cells were transfected using lipofectamine RNAiMAX (Cat no. 13778075, Invitrogen) as per manufacturer's instructions. Media was changed the next day and cells were collected at subsequent time points. siRNA sequences used in this thesis is summarised in Table 2.9.

Table 2.9: A description of all siRNAs used in this thesis

Species	**Target Gene**	**siRNA sequence (5'–3')**	**Life Technologies Cat no.**
Human	Control	N/A	12935-112
Human	UPF1	CCCAACCCGAUAAACCGAUGUUCUU, AAGAACAUCGGUUUAUCGGGUUGGG	UPF1 HSS109172
Mouse	Control	UCAACUUCUUGUCAUAGUCAGCCUC, GAGGCUGACUAUGACAAGAAGUUGA	N/A
Mouse	UPF1	UCACGACUUCUGUUAUACACUGCUC, GAGCAGUGUAUAACAGAAGUCGUGA	N/A

2.5.10.2 DNA transfection/DNA and siRNA co-transfection

One day prior to transfection cells were plated into 6 or 12 well plates with or without coverslips in culture medium without antibiotics. The following day cells were transfected using lipofectamine 2000 (Cat no. 15338030, Invitrogen) as per manufacturer's instructions. The lipofectamine:DNA/RNA ratio used was 3:1. Media was changed the next day and cells were collected at subsequent time points.

2.5.10.3 Nucleofection of NIH3T3 and mESCs (DNA and/or siRNA)

On the day of nucleofection, cells were harvested with 0.01% trypsin for NIH3T3 cells or Accutase for mESCs. Cells were then nucleofected using the 4D-Nucleofector™ X Unit (Cat no. AAF-1002X, Lonza, Basel, Switzerland) as per manufacturer's instructions for NIH3T3 cells (Cat no. AAF-1002X, Lonza) or mESCs (Cat no. V4XP-3024, Lonza). Post nucleofection,

cells from each cuvette were distributed across two wells of a 6-well plate. Media was changed the next day and cells were collected at subsequent time points.

2.5.11 Cycloheximide and MG132 chase assays

One day prior to chase assays, cells were plated as detailed in Table 2.10. The following day media was replaced with media supplemented with 100 mg/mL of the translation inhibitor cycloheximide (CHX) (Cat no. C4859, Sigma-Aldrich), 20 μm of the proteasome inhibitor MG132 (Cat no. C2211, Sigma-Aldrich), or dimethyl sulfoxide (DMSO) as a control. Cells were incubated at chase intervals of 0, 0.5, 1, 3 and 6 hours. At the end of each time point cells were rinsed with PBS and collected for downstream analysis.

Table 2.10: Cell plating densities for cycloheximide and MG132 chase assays

Cell type	**Plating density for chase assays**
HEK293T	5.26 x10^4 cells/cm^2
LCL	8.0 x10^5 cells/mL
mESC	4.5 x10^4 cells/cm^2

2.6 RNA analysis

2.6.1 RNA isolation

2.6.1.1 Trizol/RNeasy RNA extraction

Manual RNA extractions were conducted using a combined method with Trizol Reagent (Cat no. 15596026, Invitrogen), RNeasy Mini Kit (Cat no. 74104, QIAGEN) and RNase-Free DNase set (Cat no. 79254, QIAGEN). Cell pellets collected from cultured cells (thawed on ice if necessary) were homogenized using 1 mL Trizol reagent. Chloroform (200 μl) was then added and mixed shaking vigorously for 30 seconds. The mixture was incubated at room temperature

for 3 minutes followed by centrifugation at 17,900 x g for 1 minute to bind RNA to the membrane. The flow-through was discarded, and the process was repeated until all aqueous phase and ethanol mixture had been passed through. Subsequent handling of the column was conducted as per manufacturer's instructions for the RNeasy Mini Kit. RNA quality was assessed by separating the RNA using agarose gel electrophoresis and visualising the integrity of the 28S and 18S ribosomal RNA bands.

2.6.1.2 Maxwell RNA extraction

Automated RNA extractions were conducted from cell pellets collected from cultured cells using the Maxwell® RSC Instrument (Cat no. AS4500, Promega) as per Maxwell® RSC simply RNA Cell kit manual (Cat no. AS1390, Promega). RNA quality was assessed by separating the RNA using agarose gel electrophoresis and visualising the integrity of the 28S and 18S ribosomal RNA bands.

2.6.1.3 RNA quantitation

Following RNA extraction, RNA was quantified using a UV spectrophotometer at 260 nm (Nanodrop 1000, Thermo Fisher Scientific). RNA was stored at -80°C.

2.6.2 Complementary DNA (cDNA) synthesis

cDNA was synthesised using Superscript III reverse transcriptase (Cat no. 18080-051, Invitrogen) as per manufacturer's instructions using 1–2 µg of RNA and random hexamer primers. cDNA was stored at -20°C. The efficiency of reverse transcription was determined using PCR amplification of the ubiquitously expressed *Esterase D* gene using '*ESTERASE-D* F' and '*ESTERASE-D* R' PCR primers (Table 2.1) under standard conditions as described in Section 2.2.2 and PCR product was visualised by agarose gel electrophoresis (Section 2.2.5). If

cDNA was to be used in subsequent quantitative real time PCR (qPCR) the sample was further diluted as required.

2.6.3 Quantitative real time PCR (qPCR)

Prior to experimental qPCR all primers pairs used were first validated to ensure (1) generation of single size products through melting temperature analysis and (2) efficient product amplification. Efficiency of amplification was determined by comparing Cross Threshold (Ct) values acquired through serial dilutions of a control template. The StepOne V2.3 Software (Applied Biosystems) or the CFX Manager software (Bio-Rad) was then used to calculate % Efficiency. Using this method all primer sets used were found to have an efficiency of 100% +/- 10%.

qPCR reactions conducted in a 96 well plate format were set up as follows; each 20 µl reaction contained 2 µl of cDNA, 0.5 µM of forward and reverse primers, 1x Fast SYBR Green Real-Time PCR master mix (Cat no. 4385612, Applied Biosystems) and water to a final volume of 20 µl. Reactions were run on the Step One Plus Real Time PCR system (Applied Biosystems) using the following conditions: activation at 95°C for 20 seconds followed by 40 cycles of denaturation at 95°C for 3 seconds and extension at 60°C for 30 seconds. Signal emitted from the dye reporter was recorded at the end of each cycle. All samples were analysed in triplicate using the StepOne V2.3 Software (Applied Biosystems). The comparative or ΔΔCt method was then applied for experimental data analysis.

qPCR reactions conducted in a 384 well plate format were set up as follows; Each 10 µl reaction contained 1 µl of cDNA, 0.5 µM of forward and reverse primers, 1x Fast SYBR Green Real-Time PCR master mix (Cat no. 4385612, Applied Biosystems) and water to a final volume of 10 µl. Reactions were run on the CFX384 real time machine (Bio-Rad) using the following conditions: activation at 95°C for 20 seconds followed by 40 cycles of denaturation at 95°C for

3 seconds and extension at 60°C for 30 seconds. Signal emitted from the dye reporter was recorded at the end of each cycle. All samples were analysed in triplicate using the CFX Manager software (Bio-Rad). The comparative or ΔΔCt method was then applied for experimental data analysis.

2.7 Western blot protein analysis

2.7.1 Sample lysis and protein isolation

Cell pellets collected from cultured cells (thawed on ice if necessary) were resuspended in 100–200 µl of ice cold RIPA buffer which per 2 mL was supplemented with 80 µl protease inhibitor cocktail (Cat no. P8340, Sigma-Aldrich), 10 µl of 200 mM Na_2VO_4, 10 µl of 200 nm NaF, and 10 µl of 200 nM phenylmethylsulphonyl fluoride (PMSF). The lysate was either passed through an 18 G needle and 1 mL syringe 10 times or sonicated using the Bioruptor® *Plus* (Cat no. B01020001, Diagenode, NJ, USA) 3 times for 30 seconds with 30 second intervals. Cell lysates were then centrifuged at 17,900 x g at 4°C for 20 minutes before supernatant containing protein was transferred to a new tube and stored at -80°C.

2.7.2 Protein quantitation with Bradford assay

To quantify protein isolated from cultured cells, a diluted (1:5–1:20) sample of each protein isolate was first prepared. The diluted samples were then aliquoted in 10 µl triplicates into the wells of a 98 well plate. Pre-made Bovine Serum Albumin (BSA) standard curve samples at 0, 0.2, 0.4, 0.6, 0.8 and 1.0 mg/mL (Cat no. A7030, Sigma-Aldrich) were also aliquoted in 10 µl triplicates into wells of the same plate. 200 µl of 1x Bradford reagent (Cat no. 500-0006, Bio-Rad) was then added to each occupied well. The plate was read immediately at 560 nm using an automated plate reader (FLUOstar Omega, BMG Lab Tech, VIC, Australia). The standard curve method was used to calculate protein concentration.

2.7.3 Protein separation by SDS-PAGE

10–20 µg of protein samples was prepared in a solution of 10% Dithiothreitol (DTT) reducing reagent, 1x loading buffer (Cat no. NP0007, Invitrogen) and water. The protein solution was then incubated at 95°C for 5 minutes to denature protein. Protein solutions along with Protein Plus Ladder as a reference (Cat no. 1610374, Bio-Rad) were loaded into the wells of a NUPAGE® NOVEX 4–12% bis-tris 12 well (Cat no. NP0322BOX, Invitrogen), 15 well (Cat no. NP0336BOX, Invitrogen), or 17 well gel (Cat no. NP0329BOX, Invitrogen) as part of the Xcell *Surelock*® Mini cell system (Cat no. EI0002, Invitrogen). The inner chamber was filled with cold 1x NuPage MOPS SDS running buffer and 500 µl of NUPAGE antioxidant (Cat no. NP0005, Invitrogen). The outer chamber was filled with only cold 1x MOPS running buffer. Protein was separated using gel electrophoresis at 140 V for 1–3 hours (depending on size of proteins to be resolved).

2.7.4 Transfer of protein to membrane

Protein in the gel was transferred onto Pure Nitrocellulose Blotting Membrane (Cat no. 66485, Biotrace® NTS) using the XCell II™ Blot Module CE Mark Kit (Cat no. EI9051, Invitrogen). The inner chamber was filled with cold Towbin transfer buffer; the outer chamber was filled with cold water. Protein transfer was achieved using a 30 V field for 2 hours. Bulk protein transfer was assessed via Ponceau Staining (Sambrook & Russel 2001). The cellulose membrane was washed with TBST to remove Ponceau Stain.

2.7.5 Immunoblotting and protein detection

2.7.5.1 Blocking

Membranes were blocked for at least 1 hour at room temperature in a blocking solution of 5% skim milk, 5% BSA, and 5% goat serum (Cat no. 16210064, Gibco, Life Technologies) or 5% horse serum (Cat no. 16050130, Gibco, Life Technologies) in TBST.

2.7.5.2 Antibody incubations

All antibodies were diluted in a solution of 1.6% skim milk, 1.6% BSA, and 1.6% goat serum (Cat no. 16210064, Gibco, Life Technologies) or 1.6% horse serum (Cat no. 16050130, Gibco, Life Technologies) in TBST. Membranes were incubated with primary antibody dilutions (Table 2.11) overnight at 4°C with slight agitation. Membranes were rinsed 3 times with TBST, washed at least 3 times for 5 minutes with TBST and then incubated for 1 hour at room temperate with horse radish peroxidase (HRP)-conjugated secondary antibody dilutions (Table 2.12). Membranes were rinsed 3 times with TBST and then washed at least 3 times for 5 minutes with TBST to remove unbound antibody.

2.7.5.3 Chemiluminescent detection, imaging and data analysis

HRP-conjugated antibodies were detected using a chemiluminescence reaction initiated by adding a 1:1 mixture of Clarity Western Peroxide Reagent to Clarity Western Luminol/enhancer Reagent (Cat no. 1705061, Bio-Rad) to the membrane for 1 minute. Excess solution was removed and chemiluminescence was exposed onto X-Ray film (Cat no. 28906836, GE Healthcare, IL, USA) or detected digitally using the ChemiDoc™ (ChemiDoc XRS+, Bio-Rad). Band intensity of target proteins was analysed using Image Lab software (Bio-Rad).

2.7.5.4 Stripping and re-probing

When necessary primary and HRP-conjugated secondary antibodies were stripped from a western blot membrane to allow subsequent re-probing for different proteins. The membrane was washed twice for 10 minutes with mild stripping buffer, twice for 10 minutes with PBS and finally twice for 5 minutes with TBST. All washes were carried out at room temperature with gentle rocking. Once all washes were completed the membrane was ready for the blocking stage (Section 2.7.5.1).

Table 2.11: Primary antibody dilutions used for western blot analysis

Species	Protein	Dilution	Cat no.	Expected Size
Goat	UPF2	1:500	sc-20227, Santa Cruz Biotechnology, TX, USA	148 kDa
Mouse	ACTB	1: 5000	A2228, Sigma-Aldrich	42 kDa
Mouse	FLAG	1: 1000	F1804, Sigma-Aldrich	-
Mouse	HA	1: 5000	H9658, Sigma-Aldrich	-
Mouse	HBB	1: 200	sc-21757, Santa Cruz Biotechnology	-
Mouse	TetR	1:1000	632231, Clontech	-
Rabbit	GFP	1:1000	ab6556, Abcam	
Rabbit	UPF1	1:2000	A301-902, Bethyl Laboratories	150 kDa
Rabbit	UPF3A	1:1000	HPA018325, Sigma-Aldrich	55 kDa
Rabbit	UPF3B	1:250	HPA001800, Sigma-Aldrich	58 kDa
Sheep	UPF3B	1:1000	In house	58 kDa

Table 2.12: Secondary antibody dilutions used for western blot analysis

Origin	Specificity	Conjugate	Dilution	Cat no.
Donkey	Goat/Sheep	HRP	1:2000	AB324P, Millipore
Goat	Rabbit	HRP	1:2000	P0448, DAKO
Goat	Mouse	HRP	1:2000	P0447, DAKO

2.8 Immunofluorescent protein analysis

2.8.1 Sample preparation

Cells were fixed on coverslips using 4% paraformaldehyde (PFA) (Cat no. P6148, Sigma-Aldrich) for 15 minutes at room temperature. Cells were washed 4 times with PBS to remove all PFA and stored in PBS at 4°C.

2.8.2 Immunoblotting and fluorescent detection

2.8.2.1 Cell permeabilisation and blocking

Cells were blocked and permeabilised using a solution of PBS containing 0.2% Tween 20 and 5% horse serum (Cat no. 16050130, Gibco, Life Technologies) or 5% goat serum (Cat no. 16210064, Gibco, Life Technologies).

2.8.2.2 Antibody incubations and nuclear staining

All antibodies were diluted in a solution containing either 0.5% horse serum or 0.5% goat serum. Primary antibody dilutions (Table 2.13) were added to cells and incubated over night at 4°C. Cells were washed 3 times for 5 minutes with PBST. Secondary antibody dilutions (Table 2.14) were added to cells and incubated for 1 hour at room temperature. Cells were then washed a further 3 times for 5 minutes with PBST followed by 3 washes for 5 minutes with PBS. Coverslips were mounted, and nuclei were counterstained using ProLong™ Diamond Antifade Mountant with DAPI (4′,6-diamidino-2-phenylindole) (Cat no. P36961, Invitrogen).

2.8.2.3 Imaging and data analysis

All immunofluorescent images were captured as described in Section.2.9.2. Images were visualised using the AxioVision software Vs4.9.1.0 (Carl Zeiss, Germany) and when necessary fluorescence intensity was quantified using the open source program, Image J.

Table 2.13: Primary antibody dilutions used for immunofluorescence analysis

Species	Protein	Dilution	Cat no.
Mouse	HBB	1:200	sc-21757, Santa Cruz Biotechnology
Mouse	HA	1:1250	H9658, Sigma-Aldrich
Mouse	FLAG	1:500	F1804, Sigma-Aldrich
Rabbit	UPF1	1:500	A301-902, Bethyl Laboratories

Table 2.14: Secondary antibody dilutions used for immunofluorescence analysis

Origin	Specificity	Conjugate	Dilution	Cat no.
Donkey	Rabbit	Alexa 488	1:700	A21206, Invitrogen
Donkey	Mouse	Alexa 555	1:700	A31570, Invitrogen
Donkey	Rabbit	Alexa 647	1:500	A31573, Invitrogen
Goat	Rabbit	Alexa 555	1:800	A21428, Invitrogen

2.9 Microscopy

2.9.1 Inverted microscopy

Phase contrast images and fluorescence of live cells were captured using the Zeiss Vert.A1 inverted microscope fitted with the AxioCam MRm high resolution camera and the AxioVision software Vs4.9.1.0 (Carl Zeiss, Germany).

2.9.2 Fluorescence microscopy

Fluorescence microscopy images were captured using the Zeiss AxioImager M2 fluorescent microscope (Car Zeiss, Germany) fitted with the AxioCam MRm high resolution camera and the AxioVision software Vs4.9.1.0 (Carl Zeiss, Germany).

Chapter Three: Study of Novel Genetic Variants in NMD Factors Associated with Neurodevelopmental Disorders

3.1 Introduction

NMD is a highly conserved mRNA degradation pathway present in all eukaryotes. The dual role of NMD is to target transcripts containing either, an aberrant PTC to protect transcriptome fidelity or a naturally occurring NMD activating feature to regulate global gene expression. During the pioneer rounds of translation, the NMD machinery works to identify and degrade such transcripts prior to bulk translation (Isken & Maquat 2008; Maquat, Tarn & Isken 2010). For this process to be successful many NMD components are required to function in synchrony.

Mutations in most NMD factors are lethal (Hwang & Maquat 2011), however, the identification and study of non-lethal mutations in these factors through a variety of *in vitro* and *in vivo* models has uncovered that NMD has a prominent impact on several aspects of brain development and function. These include, neural differentiation, axonal targeting, cognition, sensorimotor gating and synaptic plasticity (Colak et al. 2013; Giorgi et al. 2007; Guzowski et al. 2005; Huang et al. 2011; Jolly et al. 2013; Lou et al. 2014). Of interest, non-lethal variants in human genes encoding the X-linked NMD factor UPF3B and its autosomal binding partner UPF2 have been shown to result in heterogenous presentations of neurodevelopmental disorders including ID and autism spectrum disorder (Laumonnier et al. 2010; Lynch et al. 2012; Nguyen et al. 2012; Nguyen et al. 2013; Tarpey et al. 2007).

Neurodevelopmental disorders encompass a clinically and genetically heterogenous group of disorders in which the development of the central nervous system is disturbed. This can include developmental brain dysfunction which can manifest as neuropsychiatric problems or impaired motor function, learning, language, or non-verbal communication. Examples of neurodevelopmental disorders include; ID, autism spectrum disorder, schizophrenia and epilepsy, whilst some disorders show overlapping comorbidities (Bitta et al. 2017; Thapar et al. 2017). A diagnosis of any neurodevelopmental disorder has a lifelong mental and physical

impact on the patient while also impacting the family and society in terms of management and resources.

Since *UPF3B* was first implicated in X-linked ID (Tarpey et al. 2007) several studies have focused on the role of this NMD factor in brain development. *In vitro* studies found that UPF3B-dependent NMD promotes differentiation of committed neural progenitor cells and was corroborated by subsequent *in vivo* studies showing poor differentiation of neural stem cells in a *UPF3B*-null mouse (Huang et al. 2018; Jolly et al. 2013; Lou et al. 2014).

In humans, reported pathogenic *UPF3B* variants either completely abolish UPF3B protein expression or interrupt conserved UPF2 or EJC binding domains thus impairing their role in NMD. Affected individuals often present with variable clinical features, including mild to severe ID, autistic features, a slender build, poor musculature, a long face, childhood-onset schizophrenia and ADHD (Laumonnier et al. 2010; Tarpey et al. 2007). Furthermore, the severity of these clinical features can also vary across individuals of the same family who harbour the same genetic mutation (Nguyen et al. 2012). The underlying cause of this clinical variability is likely the remanent level of NMD efficiency within the affected individual, which has been suggested to be proportional to the level at which UPF3A protein is stabilised (Nguyen et al. 2012). Individuals who exhibit a greater magnitude of UPF3A stabilisation are often less affected and show a level of transcriptomic deregulation more comparable to control individuals than those with a low magnitude of UPF3A stabilisation (Nguyen et al. 2012; Nguyen, Wilkinson & Gecz 2014).

In addition to UPF3B, CNVs in known NMD and EJC genes, namely, UPF2, UPF3A, RBM8A, SMG6, EIF4A3 and RNPS1 were found to be highly associated neurodevelopmental disorders (Nguyen et al. 2013). Unfortunately, these variants span large regions of the chromosome making it difficult to determine if the resulting clinical phenotype is due to a disruption in the

gene encoding the NMD factor, another gene in the disrupted region, or a compound effect from several disrupted genes. Identifying and studying additional genetic variants and specifically SNVs exclusively disrupting NMD factor encoding genes will be of great value in resolving their individual contribution to neurodevelopmental disorders. This, together with further studies of *UPF3B* variants is necessary to expand our knowledge on the clinical spectrum associated with neurodevelopmental disorders and to understand how disrupted NMD results in the impaired brain development and function seen in neurodevelopmental disorders.

In this chapter I have used patient derived lymphoblastoid cell lines (LCLs) to investigate the pathogenicity of novel genetic variants in *UPF3B* and *UPF2* which were identified in individuals with neurodevelopmental disorders.

3.2 Results

3.2.1 Characterisation of a variant of unknown significance (VUS) in UPF3B

Individual 54488 is a male child who presented at the age of four years with a neurodevelopmental disorder involving; global developmental delay, autism, motor delay, low muscle tone, and various dysmorphic features (Trivellin et al. 2018). Clinical grade molecular cytogenetics identified a 650 kb chromosome Xq26 microduplication (Table 3.1) in this patient. This finding led to his original referral for possible X-linked acrogigantism (MIM:300942), however, clinical and endocrine investigations failed to support this diagnosis (Trivellin et al. 2018). Further high definition molecular cytogenetics confirmed that the duplication mapped outside the X-linked acrogigantism critical region and subsequent whole exome sequencing (WES) identified a synonymous VUS in *UPF3B* (ENSG00000125351) (Table 3.1) (Trivellin et al. 2018). Herein, I will provide an *in silico* and *in vitro* assessment of the functional effect of this novel *UPF3B* variant.

Table 3.1: A summary of candidate genetic variants identified in individual 54488

Variant location	**Gene(s)**	**HGVS**	**Variant classification**	**Inheritance genotype**	**Consequence**
chrX:134,248,528–134,903,125dup;hg19	LINC00633, BC061642, CXorf48, DKFZp451F083, ZNF75D, ZNF449, AB062081, LINC00086, 5S_rRNA, DDX26B, CT45A1, CT45A3t	NC_000023.10:g 134248528_134903125dup654598	650 kb genomic duplication	Maternal: Hemizygous	Originally suspected to cause X-linked acrogigantism. High density aCGHmapped 750 kb upstream from X-linked acrogigantism critical region.
chrX:118,975,698–118,975,698;hg19	UPF3B	ENST00000276201.2: c.624G>A(P.=)	Synonymous SNV of unknown significance	Maternal: Hemizygous	In annotated regulatory region.

3.2.1.1 The UPF3B c.624G>A variant is predicted to disrupt splicing

The *UPF3B* c.624G>A variant is synonymous, meaning it does not alter the amino acid sequence of the translated UPF3B protein. This variant does however, change the last nucleotide in *UPF3B* exon 6 from 'G' to 'A' (Figure 3.1B). Variants which are located within the first or last codon of an exon have the potential to disrupt a 3' acceptor or 5' donor splice site respectively. The canonical acceptor (NYAG/G) and donor (CAG/GUAAGU) sites are strongly conserved and define exon-intron boundaries that are recognised by elements of the spliceosome (Figure 3.1A) (Anna & Monika 2018). Variation from the canonical sequence may alter interaction between pre-mRNA and proteins involved in intron removal.

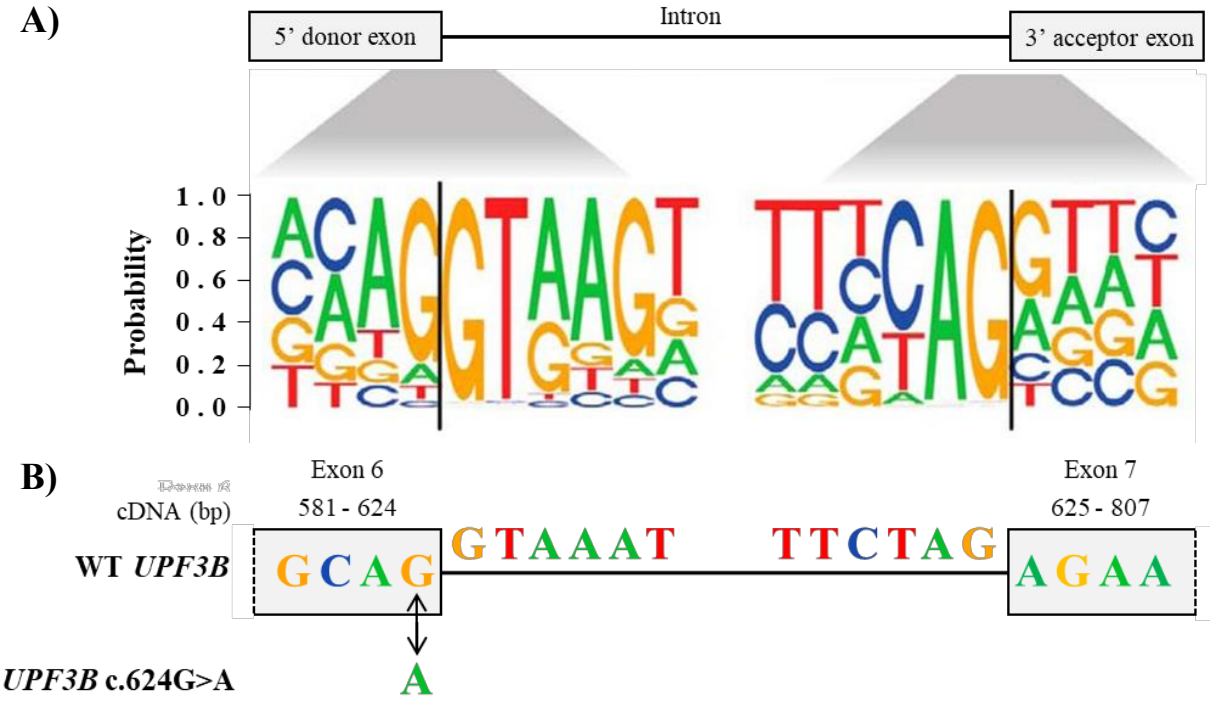


Figure 3.1: UPF3B c.624G>A variant disrupts a canonical 5' donor site within exon 6.

(A) A schematic representation of canonical 5' and 3' splice site sequences on either side of the exon-intron boundaries for all coding genes in the human genome. Figure adapted from (Wang et al. 2018). **(B)** A schematic representation of the canonical splice site sequences on either side of the exon-intron boundaries of exon 6 and exon 7 of wild-type (WT) *UPF3B* and the sequence change to the 5'donor site introduced by the synonymous *UPF3B* c.624G>A variant.

To investigate the possibility that the *UPF3B* c.624G>A variant could disrupt splicing of the *UPF3B* transcript, several *in silico* tools were applied (Table 3.2). These tools all predicted a loss of the wild-type 5'donor splice site. The 'splicing specific' tools; Human Splicing Finder, MaxEnt and Neural Network Splice Predictor all predicted that the mutant splice site would be significantly less efficient than the wild-type site. Furthermore, the *UPF3B* c.624G>A variant obtained a CADD score of 20.4, classifying it amongst the top 1% of deleterious variants.

Table 3.2: In-silico pathogenicity and splicing predictions for UPF3B c.624G>A variant

Prediction Algorithm	**Score**	**Prediction(s)**	**References**
CADD	20.4 (raw score: 2.1)	Top 1% of deleterious variants. Loss of wild-type 5' donor splice site.	(Rentzsch et al. 2019)
Mutation Taster	1 (model: without_aae)	Disease causing. Loss of wild-type 5' donor splice site.	(Schwarz et al. 2014)
MutPred Splice	0.88	Loss of wild-type 5' donor splice site.	(Mort et al. 2014)
Human Splicing Finder	WT: 87.83 / Mutant: 77.26, Variation: -38.02%	Loss of wild-type 5' donor splice site.	(Desmet et al. 2009)
MaxEnt	WT: 8.76 / Mutant: 4.44, Variation: -49.32%	Loss of wild-type 5' donor splice site.	(Desmet et al. 2009)
Neural Network Splice Predictor	WT: 0.97 / Mutant: 0.57	Loss of wild-type 5' donor splice site.	(Reese et al. 1997)

3.2.1.2 UPF3B transcripts containing the UPF3B c.624G>A variant lack exon 6

Given that *in silico* analysis predicted the *UPF3B* c.624G>A variant would result in the loss of the wild-type 5' donor site of exon 6 (Section 3.2.1.1), further *in vitro* investigations were undertaken to assess this variants impact of splicing.

Human *UPF3B* gives rise to two protein encoding isoforms; a long isoform which includes all eleven exons and a dominant short isoform which arises due to an alternative splicing event where the entire sequence of *UPF3B* exon 8 is removed. Both isoforms are predicted to be affected by the *UPF3B* c.624G>A variant.

Variants at canonical 5'donor splice sites often lead to single exon skipping. However, variants at splice sites can also result in the use of alternative 'cryptic' splice sites. Use of an alternative cryptic site in a nearby exon or intron would lead to removal of an exon fragment or inclusion of the intron fragment respectively. To determine if any of these scenarios are a consequence of the *UPF3B* c.624G>A variant, cDNA was reverse transcribed from RNA of LCLs derived from control individuals and individual 54488 (*UPF3B* c.624G>A). This cDNA was then used as a template in a *UPF3B* specific (both isoforms) PCR using primers which span the variant site (Figure 3.2A).

This experiment revealed an obvious reduction in the size of PCR amplified transcripts containing the *UPF3B* c.624G>A variant (Figure 3.2B). Further Sanger sequencing of this PCR product revealed that the observed reduction in *UPF3B* c.624G>A variant transcript size was due to the complete exclusion of *UPF3B* exon 6 (44 bp) from both the long and short isoforms of this transcript (Figure 3.2C). These findings suggest that the mutant donor splice site introduced by the *UPF3B* c.624G>A variant is not recognised by the splicing machinery, at least in LCLs.

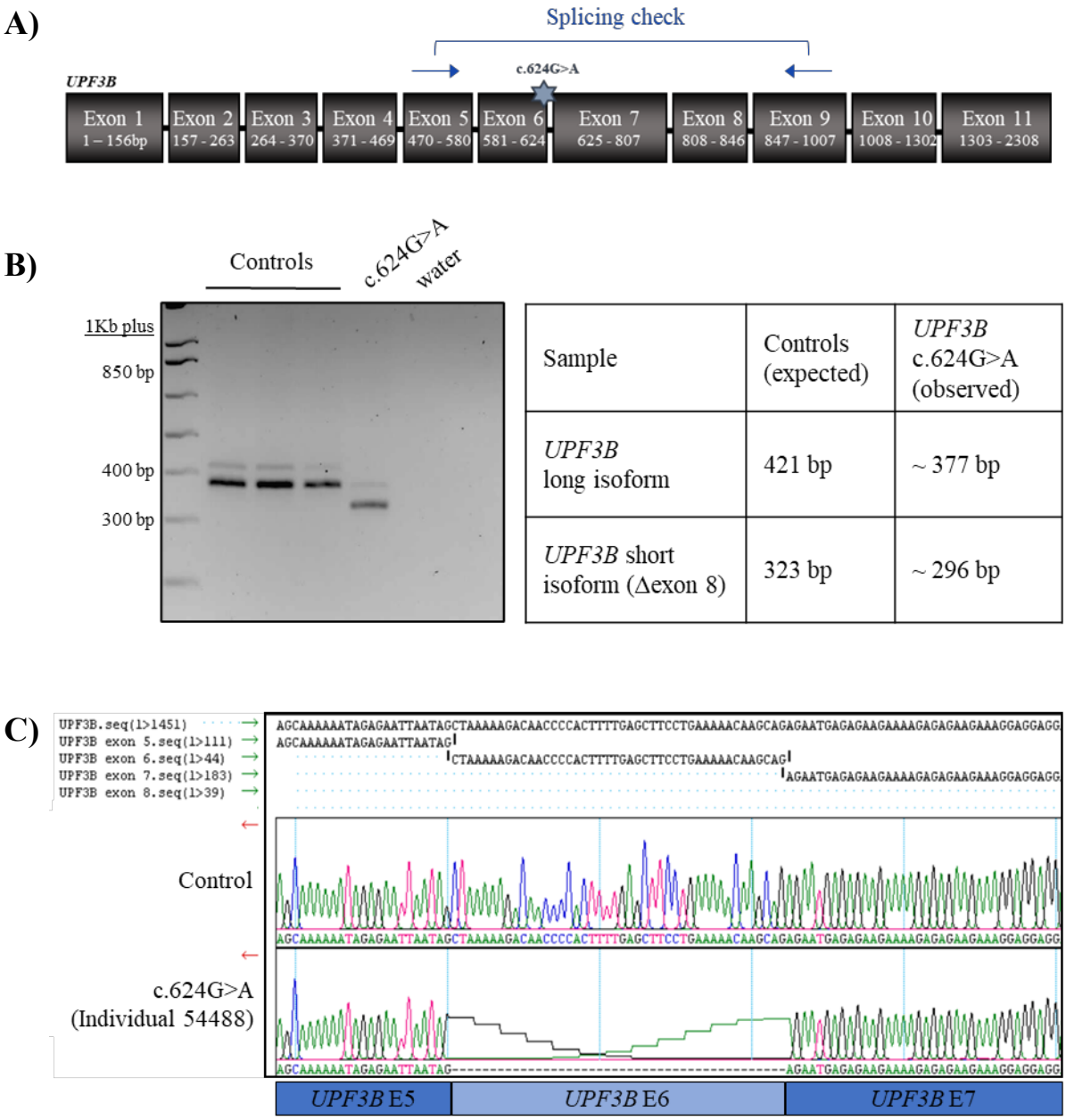


Figure 3.2: The UPF3B c.624G>A variant causes exclusion of exon 6 from UPF3B transcripts in LCLs.

(A) A schematic representation of primers designed to span the variant the *UPF3B* c.624 G>A site (splicing check). **(B)** Reverse transcribed cDNA from the RNA of LCLs derived from control individuals (n=3) and individual 54488 (*UPF3B* c624G>A) was subject to PCR with these primers. The table shows the expected and observed PCR product sizes. **(C)** Sanger sequencing of PCR products confirmed that transcripts produced from the *UPF3B* c624G>A variant identified in individual 54488 harboured a complete exclusion of exon 6 (E6) (44 bp). Green arrows indicate forward sequence orientation, red arrows indicate reverse sequence.

3.2.1.3 The UPF3B c.624G>A variant transcript leads to a frameshift and introduction of a premature termination codon into UPF3B mRNA

Loss of exon 6 (44 bp) from the *UPF3B* transcript is predicted to result in a frameshift and consequent introduction of a PTC within exon 7 of both *UPF3B* isoforms. This PTC is located 174 nts upstream of the next exon-exon junction and is predicted to be recognised by the NMD machinery based on the 55 nt rule of NMD, which states that if a transcript contains a PTC greater than 55 nts upstream of an exon-exon junction it will be targeted for NMD. Since *UPF3B* is X-linked, if this is the case, transcript from the sole copy of *UPF3B* in individual 54488 (*UPF3B* c.624G>A) would be degraded, likely resulting in a complete loss of protein i.e. a loss of function mutation. However, since all PTC-type variants may be uniquely targeted or evade NMD, it was important to undertake studies to resolve this biochemically. If the *UPF3B* c.624G>A variant transcript evades NMD, a truncated protein would be produced. If stable, this truncated protein would maintain a UPF2 binding site, but lack a Y14 binding site (Figure 3.4A), meaning it could act in a dominant negative manner by sequestering UPF2 away from the EJC.

To first investigate UPF3B protein levels produced from the *UPF3B* c.624G>A variant transcript, protein was isolated from LCLs derived from control individuals, individual 54488 (*UPF3B* c.624G>A) and another individual with a known loss of function *UPF3B* variant (*UPF3B* c.867_868delAG) (Tarpey et al. 2007). Proteins were separated by SDS-PAGE and analysed by western blot. Antibodies used to detect the N-terminus of UPF3B showed that only control individuals expressed full length UPF3B, whilst no truncated UPF3B protein was detected from individual 54488 (Figure 3.3B). Further qPCR experiments revealed that individual 54488 also showed a 47.1 % reduction in *UPF3B* transcript compared to male control individuals. These results suggested that NMD of the *UPF3B* c.624G>A variant transcript was occurring, although not at complete efficiency (Figure 3.3D) and the complete loss of protein

observed (Figure 3.3B) is most likely due to any truncated UPF3B protein being unable to fold properly and thus being subject to rapid degradation.

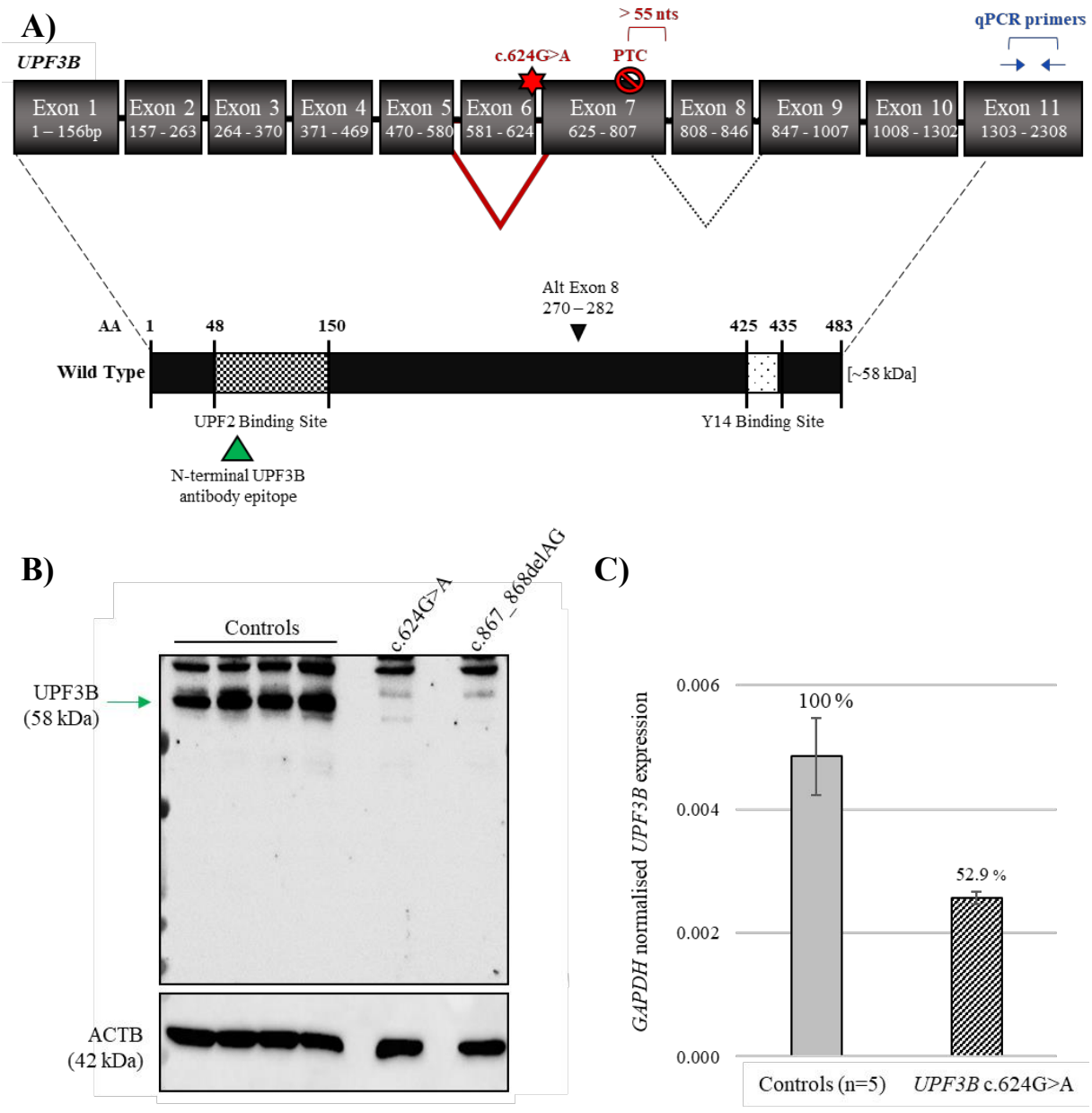


Figure 3.3: The UPF3B c.624G>A variant manifests as loss of function variant in LCLs due to aberrant splicing resulting in introduction of an NMD-targeted PTC.

(A) The *UPF3B* transcript is composed of 11 exons. Two isoforms exist including a long isoform containing all exons and the predominantly expressed short isoform which lacks exon 8 due to alternative splicing (black dotted line). These transcripts (black) are translated into functional wild-type protein. The *UPF3B* c.624G>A variant induces exon 6 skipping (red line) which will introduce a PTC within exon 7. This PTC resides greater than 55 nts upstream of an

Figure 3.3 continued…

exon-exon junction, positioning it to be targeted by NMD. Green arrowhead indicates the binding position of the UPF3B antibody used for western blot analysis in (B) which targets a peptide encoded upstream of the PTC. Blue arrows indicate the primer positions used for qPCR in (C). **(B)** Total was protein isolated from LCLs derived from control individuals (n=4), individual 54488 (*UPF3B* c.624G>A) and an individual with a known loss of function *UPF3B* variant (c.867_868delAG). Proteins were separated by SDS-PAGE and analysed by western blot using antibodies to detect the N-terminus of UPF3B and the loading control ACTB (stripped and re-probed). This analysis showed that wild-type UPF3B can only be detected in controls (indicated by a yellow arrow) and that individuals harbouring *UPF3B* variants showed no evidence of production of a truncated UPF3B protein. Additional larger bands represent non-specific antibody binding. **(C)** cDNA reverse transcribed from RNA of LCLs derived from control individuals (n=5) and individual 54488 (*UPF3B* c.624G>A) (n=1) was used as a template in qPCR to determine mean expression (±standard deviation) of *UPF3B* mRNA. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample.

To provide additional evidence that the *UPF3B* c.624G>A variant transcript from individual 54488 is subjected to NMD, the gold-standard 'cycloheximide chase' experiment was conducted. Cycloheximide is an antibiotic known to inhibit translation in eukaryotes. Since NMD is a translation dependent mRNA degradation pathway cycloheximide also inhibits NMD activity. If the *UPF3B* c.624G>A variant transcript is subject to NMD, levels of this transcript should increase following cycloheximide treatment.

LCLs derived from control individuals and individual 54488 (*UPF3B* c.624G>A) were treated for 6 hours with 100 mg/ml of cycloheximide or DMSO as a control and RNA samples were collected. cDNA was reverse transcribed from the collected RNA samples and used as a template for *UPF3B* specific qPCR. Cycloheximide treatment restored expression of the *UPF3B* c.624G>A variant transcript (Figure 3.4). It was seen that expression of wild-type *UPF3B* transcript also increased in controls following treatment of cycloheximide treatment, however, this increase was not as marked as in individual 54488 (Figure 3.4). The effect of cycloheximide on wild-type *UPF3B* mRNA has been previously documented (Tarpey et al. 2007). This is in line with evidence that wild-type *UPF3B* transcript, alongside transcripts for several other NMD factors is an endogenous target of NMD, forming a negative feedback regulatory network (Huang et al. 2011).

This data shows that the *UPF3B* c.624G>A variant transcript is targeted for NMD, although not completely so, with ~50% of transcript remaining. Importantly, no detectable full-length protein, nor any truncated protein is translated from this transcript (Figure 3.3). Together, these results support the prediction that loss of exon 6 from the *UPF3B* c.624G>A variant transcript results in loss of UPF3B function.

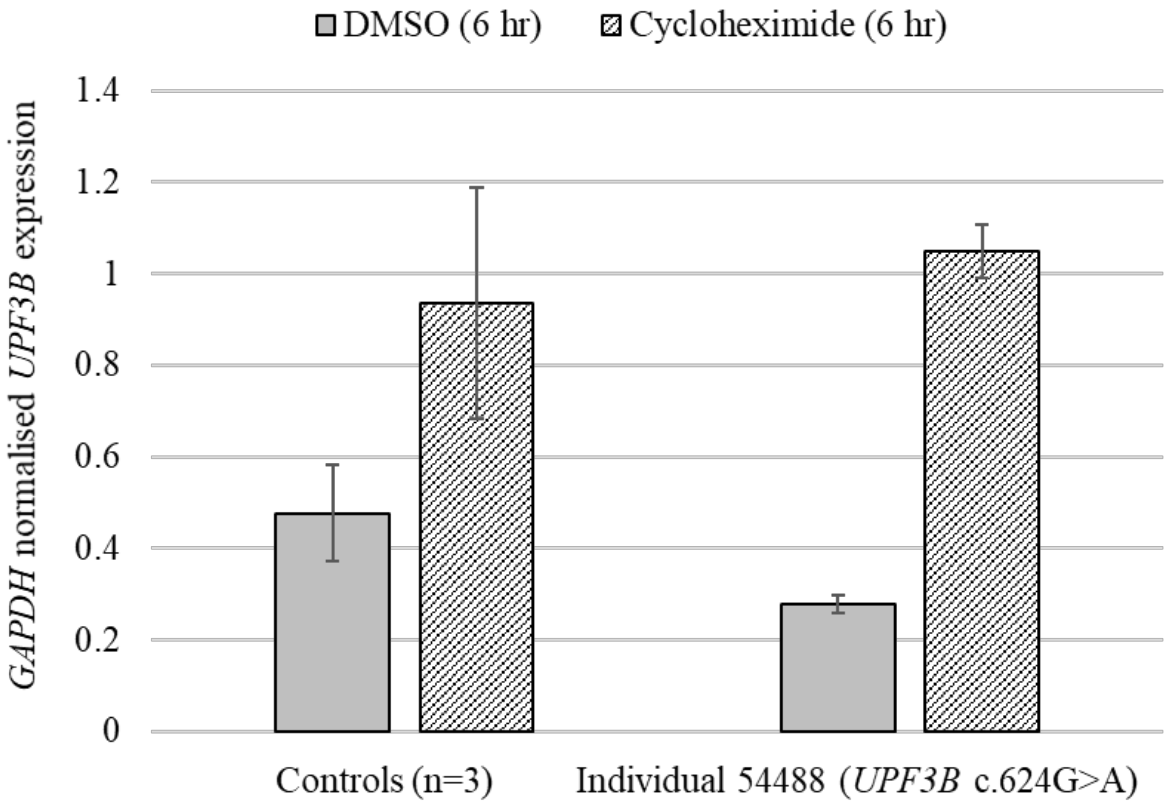


Figure 3.4: Cycloheximide treatment of LCLs reveals that UPF3B c.624G>A variant transcript is degraded by NMD.

LCLs derived from control individuals (n=3) or individual 54488 (*UPF3B* c.624G>A) were treated with cycloheximide to inhibit translation and therefore inhibit NMD. Following treatment, cDNA reverse transcribed from RNA isolated from these cells was used as a template in qPCR to determine mean expression (±standard deviation) of *UPF3B*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample. Note that cycloheximide treatment elevated the expression of the *UPF3B* c.624G>*A* variant mRNA consistent with it being targeted by NMD. Also note elevation of *UPF3B* mRNA in control samples following cycloheximide treatment consistent with reports that wild-type *UPF3B* is an endogenous NMD target (Huang et al, 2011).

3.2.1.4 Classical NMD is less efficient due to the UPF3B c.624G>A variant

Loss of UPF3B protein is typically accompanied by a compensatory increase in UPF3A (Chan et al. 2009). *UPF3A and UPF3B* share a high sequence similarity, however, several studies have shown that UPF3A interacts less efficiently with the EJC than UPF3B and as such, is subjected to rapid turnover in cells. In the absence of UPF3B, however, the UPF3A-UPF2 interaction is permitted, resulting in UPF3A stabilisation. Stabilised UPF3A is able to rescue some, but not all NMD activity lost due to a loss of UPF3B (Buchwald et al. 2010; Kim, Kataoka & Dreyfuss 2001; Kunz et al. 2006). Following from this, it has been observed that the extent of UPF3A protein stabilisation inversely correlates with the extent of transcriptome deregulation and severity of the patients neurological phenotype (Chan et al. 2009; Nguyen et al. 2012).

To investigate the UPF3A levels of individual 54488 (*UPF3B* c.624G>A) protein isolated from LCLs derived from control individuals, individual 54488 and another individual previously shown to have a complete loss of function *UPF3B* mutation (*UPF3B* c.867_868delAG) (Tarpey et al. 2007) was separated by SDS-PAGE and analysed by western blot. The results once again show that wild-type UPF3B was not detectable in individual 54488. Moreover, it was observed that UPF3A is increased in individual 54488 compared to controls and this increase is comparable to that observed for the individual harbouring an established loss of function *UPF3B* variant (Figure 3.5A).

Finally, to assess the impact of the *UPF3B* c.624G>A variant on NMD activity, the expression of three well established NMD target mRNAs (Jolly et al. 2013; Mendell et al. 2004) was assessed via qPCR using cDNA reverse transcribed from RNA of LCLs derived from control individuals and individual 54488 (*UPF3B* c.624G>A). This analysis revealed that NMD target

transcripts were increased by 29–179% in individual 54488, supporting a decrease in the efficiency of classical NMD (Figure 3.5B).

To summarise, the studies conducted here have shown that the *UPF3B* c.624G>A variant identified in individual 54488 is pathogenic and results in loss of UPF3B function accompanied by stabilisation of its paralog UPF3A. Specifically, this variant disrupts *UPF3B* splicing, resulting in the production of an aberrant PTC containing transcript that is subjected to NMD. Further investigations revealed that NMD of the *UPF3B* c.624G>A variant transcript, alongside known NMD target genes was inefficient, supporting the knowledge that stabilised UPF3A cannot entirely compensate for loss of UPF3B (Buchwald et al. 2010; Kim, Kataoka & Dreyfuss 2001; Kunz et al. 2006).

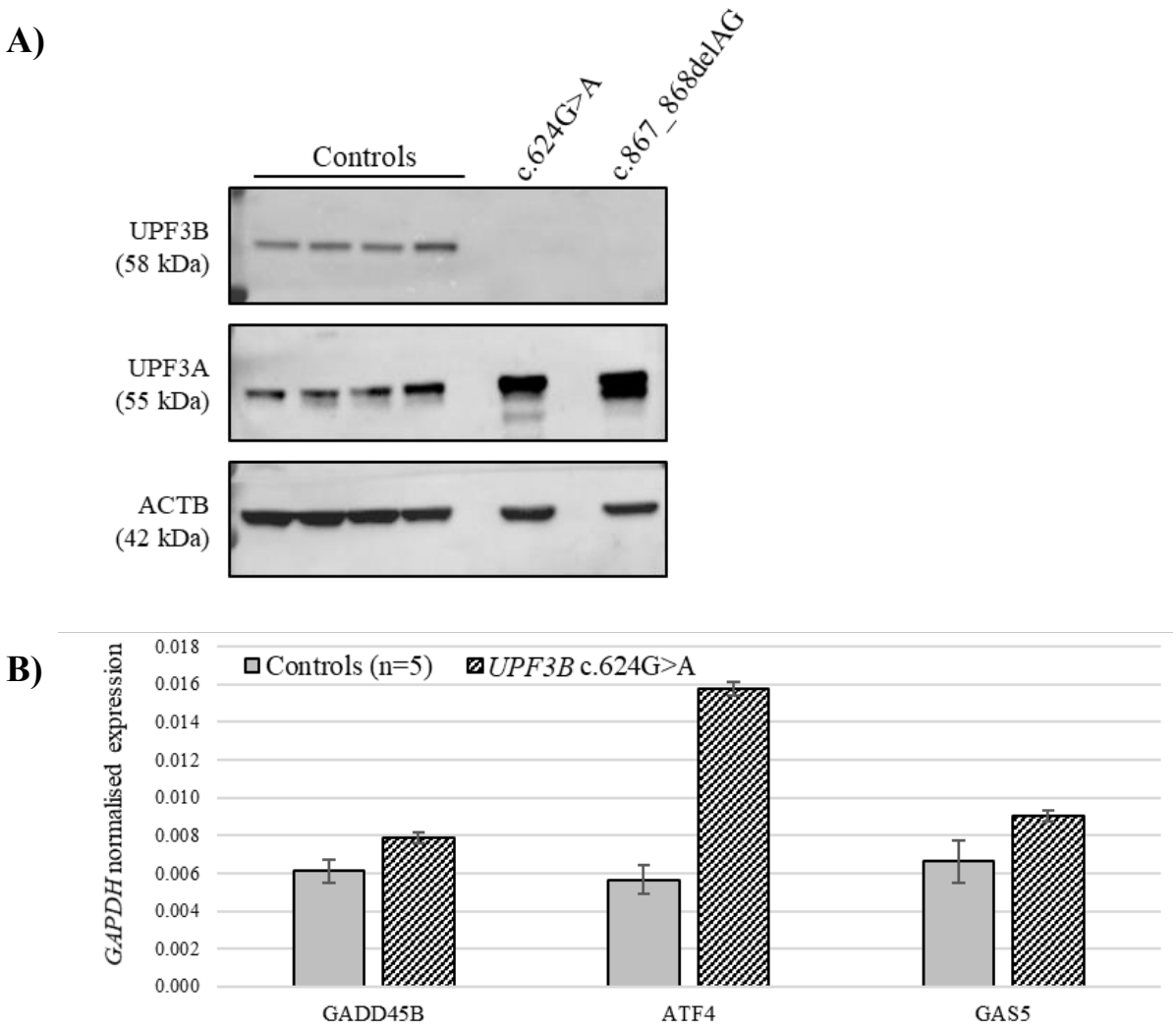


Figure 3.5: UPF3A compensates for a loss of UPF3B in LCLs derived from individual 54488 (UPF3B c.624G>A).

(A) Total protein isolated from LCLs derived from control individuals (n=4), Individual 54488 (c.624G>A) and an individual with an established loss of function *UPF3B* variant causing ID (*UPF3B* c.867_868delAG) were separated by SDS-PAGE and analysed by western blot using antibodies to detect the C-terminus of UPF3B, UPF3A and the loading control ACTB. **(B)** cDNA reverse transcribed from the RNA of LCLs derived from control individuals (n=5) and individual 54488 (*UPF3B* c.624G>A) was used as a template in qPCR to determine mean expression (±standard deviation) of *ATF4, GADD45B and GAS5*. The assay was performed in triplicate per sample. Expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample.

3.2.2 Characterisation of novel variants in UPF2 associated with neurodevelopmental disorders

In the classical model of NMD, UPF2 forms a bridge between the NMD components UPF3B and UPF1. Investigations into the pathology of *UPF3B* X-linked ID causative variants have provided strong evidence that compromised NMD is linked to ID pathology. This finding is supported by a shared clinical presentation of individuals with CNVs affecting additional genes encoding NMD factors or EJC proteins (Nguyen et al. 2012; Nguyen et al. 2013; Tarpey et al. 2007). In this respect, individuals with heterozygous copy number deletions encompassing *UPF2* have been the best studied and show an overlapping deregulated transcriptome when compared to *UPF3B* patients (Nguyen et al. 2013). These CNVs reduce the dosage of UPF2, however, they also contain deleted chromosomal regions including many additional genes which potentially contribute to the clinical phenotypes observed. This makes it difficult to conclude the involvement of *UPF2* in neurodevelopmental disorders and to elucidate phenotypes associated with disrupted NMD. Variants affecting only *UPF2* are vital to help clarify this issue. Through collaboration with clinical genetic diagnostic services, two *de novo* heterozygous single base pair deletions in *UPF2* were discovered (Johnson et al. 2019), one of which was studied in detail in this section. In addition, a novel heterozygous CNV deletion spanning *UPF2* was also included in the analysis.

3.2.2.1 Clinical phenotype of individuals harbouring UPF2 variants

The two *de novo* heterozygous single base pair deletions in *UPF2* (ENSG00000151461) were identified in a male (case 1) and female (case 2) paediatric patient (Table 3.3). Both cases presented with low-average IQ, low-average ability in receptive and expressive language and a phonological speech sounds disorder. A third male individual (case 3) who harbours a larger deletion including *UPF2* was also identified (Table 3.3). This individual presented with severe ID, autism spectrum disorder and was non-verbal (Johnson et al. 2019). Blood samples needed

to establish patient LCLs were only available from cases 1 and 3. Using these resources and molecular methods, the variant exclusively disrupting *UPF2* in case 1 was assessed and compared to case 3 and two previously characterised CNV deletion cases (Table 3.3, cases 4 and 5) (Nguyen et al. 2013) with the aim to resolve the specific contribution of heterozygous loss of function *UPF2* variants to neurodevelopmental disorders and to study the impact of such variants on the NMD pathway.

Table 3.3: Description of UPF2 variants investigated in this chapter and their clinical phenotypes

Case number	Variant	Sex	Age (y,m)	Inheritance	Genes affected	Clinical phenotype	Reference
Case 1	NM_080599.1: c.1940delA (p.Asp647Valfs*23)	Male	11, 5	*De novo*	*UPF2*	Clinodactyly of 5th digit, low average IQ, severe phonological disorder that resolved to mild at the time of testing (expressive language is more impaired than receptive language) and mild impairment in literacy.	(Johnson et al. 2019)
Case 2	NM_080599.1: / c.986delC, p.(Ser329Metfs*8)	Female	8, 11	*De novo*	*UPF2*	Born at 36 weeks, cleft lip and cleft palate, VACTERL syndrome, right hemifacial microsomia, low average IQ, moderate dysphasia (expressive language is more impaired than receptive language) and mild impairment in literacy.	(Johnson et al. 2019)
Case 3	chr10:10,901,194-13,286,601del;hg19	Male	8, 8		*LINC00710, CELF2, CELF2-AS2, AF007147, USP6NL, ECHDC3, PROSER2, PROSER2-AS1,* ***UPF2****, DHTKD1, SEC61A2, NUDT5, CDC123, CAMK1D, MIR4480, LOC283070, CCDC3, 5S_rRNA, AK311458, OPTN, MCM10, UCMA*	Macrocephaly, severe intellectual disability, non-verbal (expressive and receptive language affected), and severely autistic.	(Johnson et al. 2019)

Case 4	chr10:1-12,878,994 del;hg19	Female	2, 0	*De novo*	*TUBB8, ZMYND11, DIP2C, 5S_rRNA, DIP2C, MIR5699, PRR26, LARP4B, BC127786, GTPBP4, IDI2, IDI2-AS1, IDI1, BC046483, WDR37, AX748285, LINC00200, ADARB2, ADARB2-AS1, LINC00700, LINC00701, PFKP, PITRM1, PITRM1-AS1, BC039685, BC037918, KLF6, AK055803, AK095699, U6, LINC00704, LINC00705, AKR1E2, AKR1C6P, AKR1C2, AKR1C3, AKR1CL1, AKR1C4, UCN3, TUBAL3, NET1, CALML5, AK128534, CALML3, ASB13, FAM208B, C10orf18, GDI2, TRNA_Val, ANKRD16, FBXO18, IL15RA, IL2RA, RBM17, MIR3155A, MIR3155B, PFKFB3, LOC399715, PRKCQ, AX748236, PRKCQ-AS1, LINC00707, SFMBT2, ITIH5, ITIH2, KIN, ATP5C1, TAF3, GATA3-AS1, GATA3, BC031880, 5S_rRNA, HV745896, HV745902, HV745905, HV745899, HV745905, BC032914, SFTA1P, LINC00710, CELF2, CELF2-AS2, AF007147, USP6NL, ECHDC3, PROSER2, PROSER2-AS1,* ***UPF2****, DHTKD1, SEC61A2, NUDT5, CDC123, CAMK1D, MIR4480, LOC283070*	Intellectual disability and dysmorphic features.	(Nguyen et al. 2013)
Case 5	chr10:11,418,767-14,745,833del;hg19	Female	12, 10		*USP6NL, ECHDC3, PROSER2, PROSER2-AS1,* ***UPF2****, DHTKD1, SEC61A2, NUDT5, CDC123, CAMK1D, MIR4480, LOC283070, CCDC3, 5S_rRNA, AK311458, OPTN, MCM10, UCMA, PHYH, SEPHS1, BEND7, PRPF18, AK055017, FRMD4A, MIR1265, FAM107B*	Developmental delay, ADHD, learning disabilities, major depression and sensory issues.	(Nguyen et al. 2013)

3.2.2.2 The UPF2 c.1940delA frameshift variant introduces an NMD-targeted PTC

UPF2 resides on chromosome 10 and is composed of 21 exons. The novel *UPF2* c.1940delA variant occurs in exon 8 causing a frameshift and consequently introducing a PTC within exon 9. This PTC resides 56 nts upstream of the next exon-exon junction and was therefore predicted to be degraded via NMD, albeit only just satisfying the 55 nt rule. To test this prediction, cDNA reverse transcribed from RNA of LCLs derived from controls and case 1 (*UPF2* c.1940delA) was first subject to a PCR designed to amplify the variant region, the purified PCR product was then Sanger sequenced and a ratio of the *UPF2* wild-type and c.1940delA variant cDNA species was established using a software program called Tracking of Indels by Decomposition (TIDE) (Brinkman et al. 2014) (Figure 3.6).

TIDE is a web tool/R code designed to interrogate heterogenous genome editing events (e.g. CRISPR). It helps identify and quantify the types of insertions and deletions (indels) across a pool of heterogenous loci using analysis of Sanger sequencing traces (Brinkman et al. 2014). This tool can also be applied to identify and quantify different species of cDNA (like it does for alleles) using sequencing traces from a sample of interest and comparing it to that of a reference sequence.

Since *UPF2* is an autosomal gene, TIDE should detect two separate cDNA species from a sequencing trace of case 1 i.e. the wild-type transcript and the *UPF2* c.1940delA variant transcript. If the variant transcript is subject to NMD as predicted, the quantity of variant transcript will be detected by TIDE analysis at a much lower level than its wild-type counterpart. If the prediction is incorrect, then both transcripts would be expressed at equal levels. TIDE analysis of transcripts derived from case 1 showed that the *UPF2* c.1940delA variant transcript was lowly expressed (13.1% of detectable transcripts) compared to wild type transcripts (83.1%). This suggests that variant transcript is degraded by NMD (Figure 3.6B).

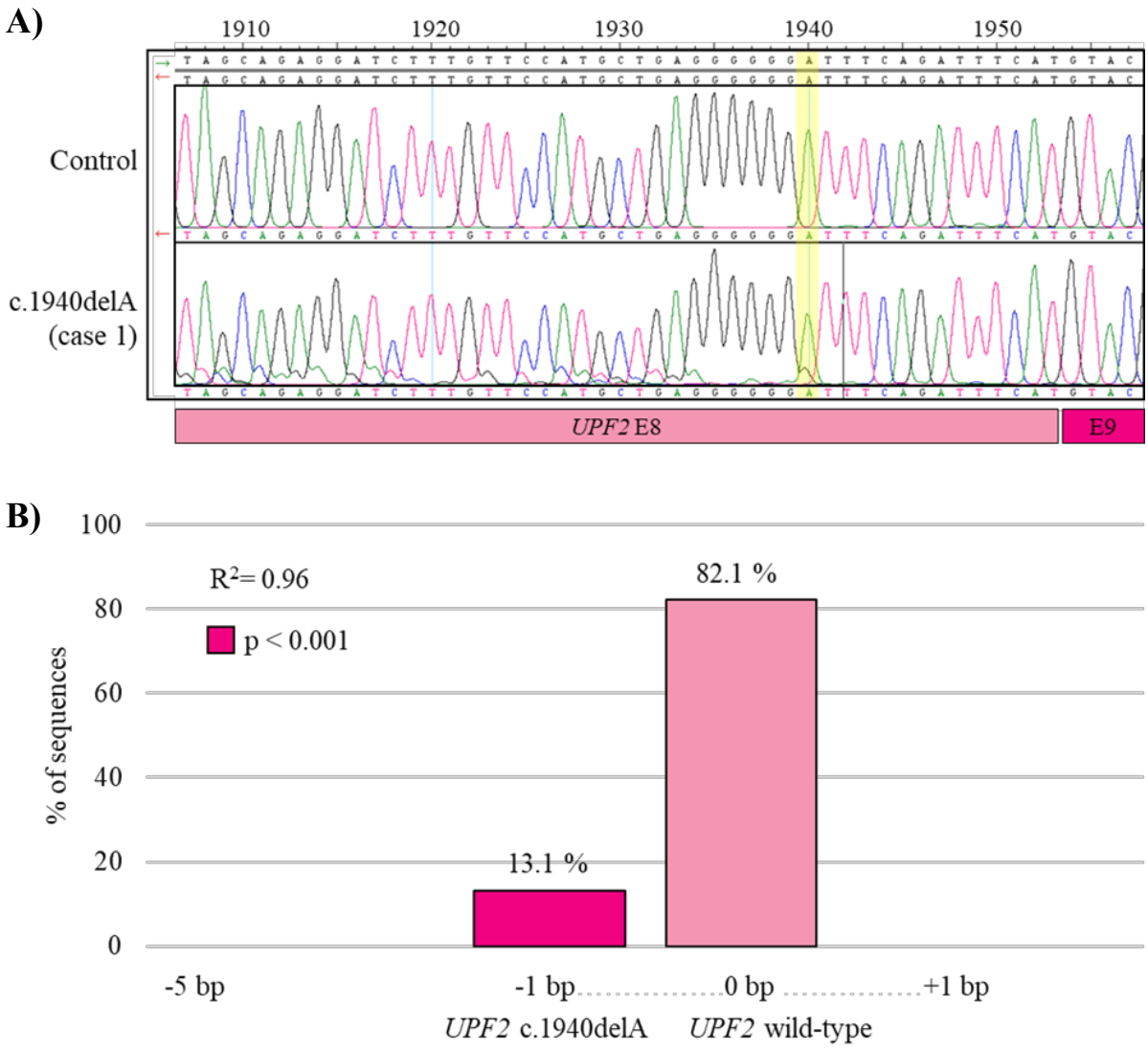


Figure 3.6: Expression of the UPF2 c.1940delA variant transcript is reduced in LCLs.

(A) Sanger sequencing of cDNA reverse transcribed from RNA of LCLs derived from a control individual and case 1 (*UPF2* c.1940delA) reveals the presence of a second lowly expressed *UPF2* variant transcript which diverges from the wild-type transcript within exon 8 (E8) at the variant position c.1940 (yellow shading). Green arrows indicate forward sequence orientation, red arrows indicate reverse sequence orientation. **(B)** These sequencing traces were also quantified using the TIDE web tool. Analysis showed that the *UPF2* variant transcript differs from the wild-type transcript by the loss of a single base pair and accounts for 13.1% of total detectable transcript. Based on the provided wild-type and variant sequence traces the software provides the R^2 value as a measure of goodness of fit and calculates the statistical significance (*p*-value) for each indel detected.

To further investigate the consequence of the *UPF2* c.1940delA variant on *UPF2* expression in the cell, cDNA reverse transcribed from RNA of LCLs derived from control individuals, case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletions) was used as a template in *UPF2* specific qPCR. It was observed that the level of *UPF2* expression of case 1 was reduced to approximately half of control levels and comparable to the reduction of *UPF2* expression observed for cases 3–5 where the entire *UPF2* gene is deleted as part of large CNV (Figure 3.7A).

To support that the observed loss of *UPF2* transcript levels in case 1 is due to NMD of the *UPF2* c.1940delA variant transcripts, LCLs derived from control individuals or case 1 (*UPF2* c.1940delA) were treated for 6 hours with 100 mg/ml of cycloheximide to inhibit translation (and thus NMD) as previously described (Section 3.2.1.3). If the *UPF2* c.1940delA variant transcript is subject to NMD, levels of this transcript should increase following cycloheximide treatment resulting in an increase of detected total *UPF2* transcripts.

cDNA reverse transcribed from the RNA of cycloheximide or DMSO treated cells was used as a template for *UPF2* specific qPCR. Treatment with cycloheximide elevated expression of *UPF2* transcripts in LCLs from case 1 (Figure 3.7B). It was seen that expression of wild-type *UPF2* transcripts also increased in control individuals following cycloheximide treatment, however, this increase was not as marked as the increase in *UPF2* transcripts observed in case 1 (Figure 3.7B). These results were similar to previous results obtained using cycloheximide treatment of cells harbouring the *UPF3B* loss of function variant (Figure 3.4). The increase in *UPF2* levels in control individuals is consistent with data showing that *UPF2* is an endogenous NMD target, this mechanism is described as a negative feedback loop that buffers NMD activity under normal conditions (Huang et al. 2011).

Collectively these data are consistent with the hypothesis that the *UPF2* c.1940delA frameshift variant results in a PTC containing transcript which is recognised and degraded by the NMD machinery. Since case 1 (*UPF2* c.1940delA) showed an overall loss of *UPF2* mRNA dosage which was comparable to that observed for cases 3–5 which harboured large CNV deletions encompassing *UPF2*, NMD of the *UPF2* c.1940delA variant transcripts appears to be efficient, despite UPF2 being an NMD factor.

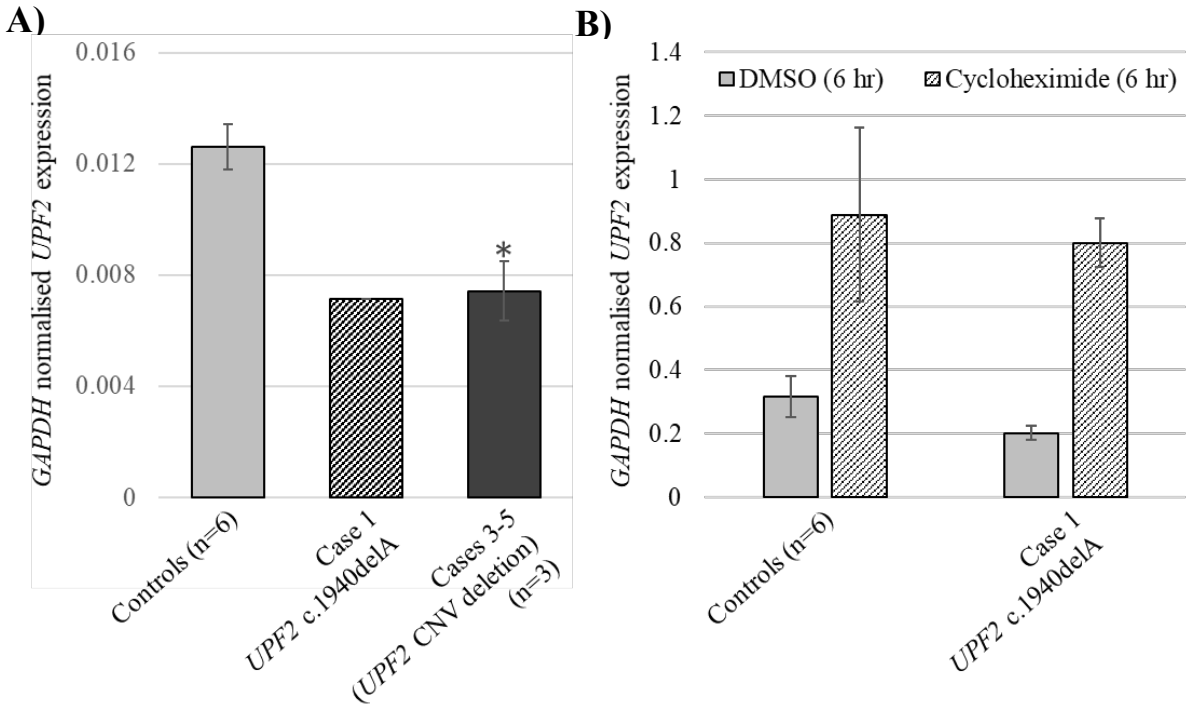


Figure 3.7: The UPF2 c.1940delA variant transcript is subject to NMD in LCLs.

(A) cDNA reverse transcribed from the RNA of LCLs derived from control individuals (n=6), case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletion, n=3) was used as a template in qPCR to determine mean expression (±standard deviation) of *UPF2*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample. **P*<0.05 by Student's two-tailed t-test. **(B)** LCLs derived from control individuals (n=3) and case 1 (*UPF2* c.1940delA) were treated for 6 hours with cycloheximide to inhibit translation and thus inhibit translation dependent NMD. Following treatment, cDNA reverse transcribed from the RNA of these cells was used as a template in qPCR to determine mean expression (±standard deviation) of *UPF2*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample. Note that cycloheximide treatment elevated the expression of the *UPF2* mRNA consistent with the *UPF2* c.1940delA variant mRNA being targeted by NMD. Also note the elevation of *UPF2* mRNA in control samples following cycloheximide treatment, however this increase was not as marked as in case 1.

3.2.2.3 The UPF2 c.1940delA variant reduces cellular UPF2 protein levels but does not impact the efficiency of classical NMD.

To investigate the effect of the *UPF2* c.1940delA variant on UPF2 cellular protein levels, protein isolated from LCLs derived from control individuals, case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletions) was separated by SDS-PAGE and analysed by western blot (Figure 3.8A). Case 1 showed a 78% reduction in UPF2 expression compared to control individuals and this reduction was comparable to that of heterozygous *UPF2* encompassing CNV deletion cases (cases 3–5) when compared to control individuals (Figure 3.8). Additional NMD factors were also analysed, interestingly, UPF3A was reduced by ~66% in all affected cases when compared to control individuals (Figure 3.8).

To next determine if reduced levels of UPF2 caused by the *UPF2* c.1940delA variant of case 1 has an impact on the efficiency of classical NMD, cDNA reverse transcribed from RNA of LCLs derived from control individuals, case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletion) was used as a template in qPCR to determine the expression of three well established NMD target mRNAs, namely, ATF4, GADD45B and GAS5 (Jolly et al. 2013; Mendell et al. 2004). These were selected as they express in LCLs and are known to respond to a loss of classical NMD dictated by a loss of UPF1 or UPF3B (Imamachi et al. 2012; Nguyen et al. 2012; Nguyen et al. 2013; Tarpey et al. 2007).

This analysis revealed that *ATF4* and *GAS5* showed a slight (~20%) but significant increase in expression in CNV cases 3–5 compared to control individuals, however, neither of these mRNAs were upregulated as a consequence of the *UPF2* c.1940delA variant of case 1 (Figure 3.9). To note, expression from the target gene *GADD45B* appeared slightly upregulated in all cases but did not reach significance in the *UPF2* CNV deletion cases (Figure 3.9).

These results show that although the *UPF2* c.1940delA variant appears to drastically reduce UPF2 protein expression in the cell, this reduction does not impact the efficiency of the classical NMD pathway, at least as measured by the expression of selected genes.

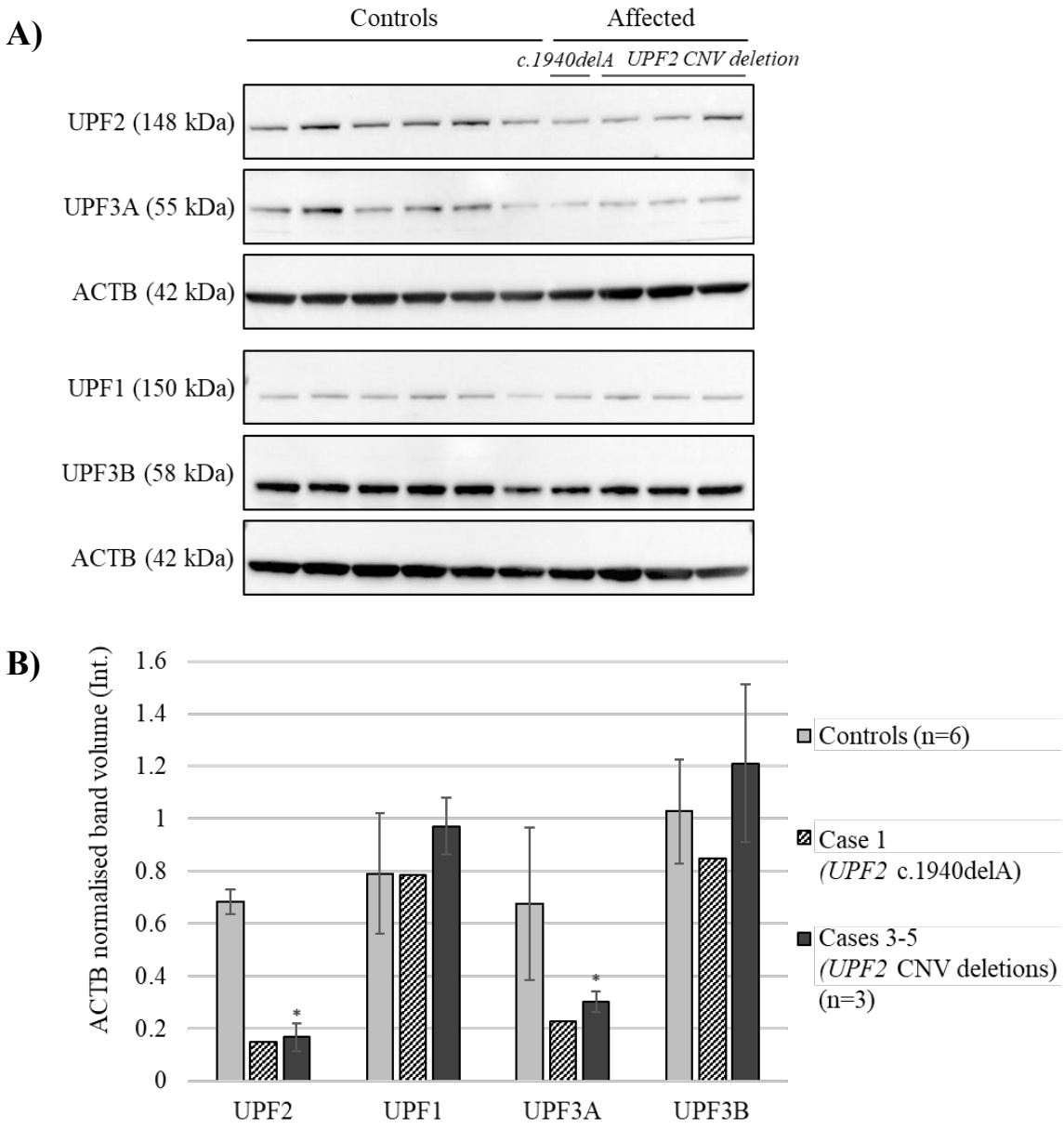


Figure 3.8: The UPF2 c.1940delA variant results in reduced UPF2 protein expression in LCLs.

(A) Total protein isolated from LCLs derived from control individuals (n=6), case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletions) was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF2, UPF3A, UPF1, UPF3B and the loading control ACTB. **(B)** Densitometric analysis of these western blots shows that UPF2 is reduced as a result of the *UPF2* c.1940delA variant in case 1 and this reduction is comparable to the significant decrease in protein expression due to large CNVs resulting in the heterozygous deletion of UPF2 (cases 3–5). Furthermore, in samples where UPF2 expression is reduced UPF3A expression is also reduced. *P<0.05 by Student's two-tailed t-test.

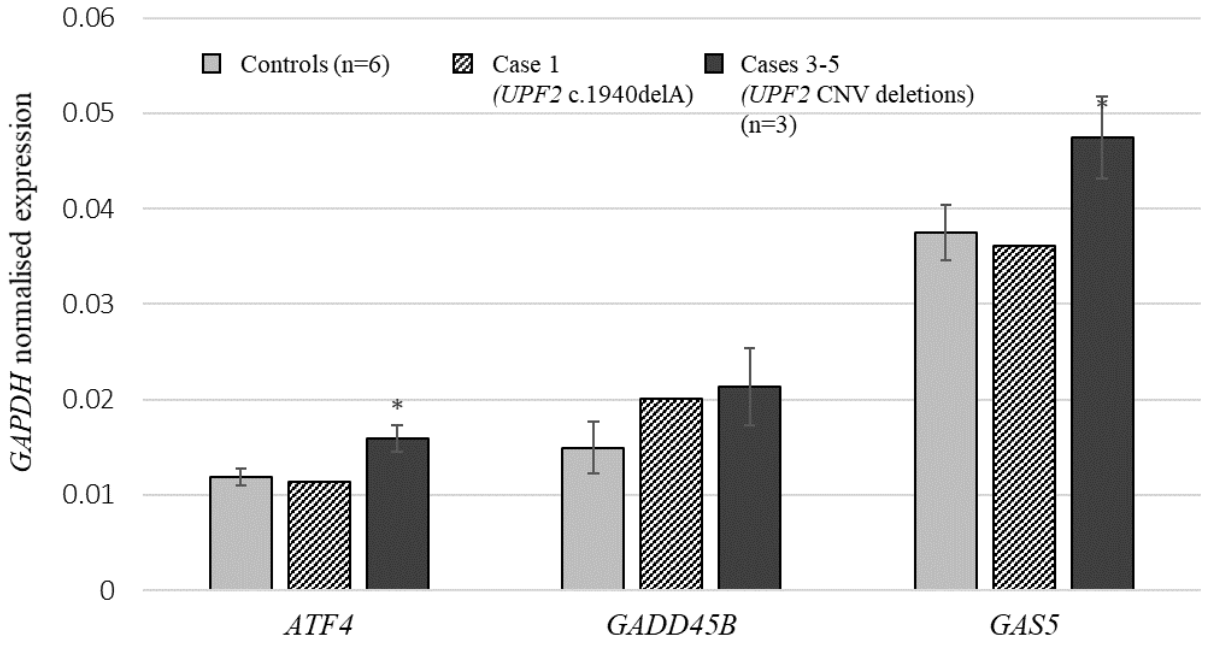


Figure 3.9: Classical NMD target gene expression remains unchanged in response to the UPF2 c.1940delA variant in LCLs.

cDNA reverse transcribed from the RNA of LCLs derived from control individuals (n=6), case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletion, n=3) was used as a template in qPCR to determine mean expression (±standard deviation) of *ATF4, GADD45B* and *GAS5*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample. **P*<0.05 by Student's two-tailed t-test.

3.3 Discussion

Using molecular methods, I have characterised three novel variants within genes encoding the NMD factors UPF3B and UPF2 , these were identified in individuals who present with varying presentations of neurodevelopmental disorders (Johnson et al. 2019). The variant disrupting *UPF3B* is a hemizygous SNV found in a canonical splice region (*UPF3B* c 624G>A, individual 54488). This variant was originally classified as a VUS and as such overlooked regarding pathogenicity. The two other variants studied impacted *UPF2* and were heterozygous. One was a novel frameshift inducing SNV in *UPF2* (*UPF2* c.1940delA, case 1) which is one of only two mutations identified to exclusively disrupt *UPF2*. The other was a CNV resulting in the heterozygous deletion of *UPF2* along with 21 other genes (case 3).

Investigating the pathology of genetic variants is challenging and several approaches are commonly employed, each with its own limitations. Introduction of mutations through manipulating cell lines (i.e. genome editing, gene silencing or gene overexpression) is often used when material from the affected individual(s) is unavailable, however this has inherent problems, such as off target effects or differences in underlying genetic backgrounds. Another common approach is modelling genetic changes in animals however, this is a significant investment and can be limiting due to differences between human and animal physiology.

A valuable approach has been the use of patient derived cell lines such as LCLs to provide a resource of patient biomolecules (DNA, RNA and protein) and a cellular model in which to assess alteration of cell function due to a genetic variant. Intriguingly, although derived from blood, LCLs share a similar gene expression profile to neurons and as such they have been employed as a surrogate cell model in many molecular and pharmacogenetic studies of neurological disorders (Nguyen et al. 2012; Sie et al. 2009; Wheeler & Dolan 2012). Having access to LCLs, or to blood samples from affected individuals (to derive an LCL cell line) has

provided a unique resource to investigate variant pathogenicity while maintaining the affected individual's native biology.

3.3.1 The synonymous UPF3B c.624G>A variant disrupts a 5'donor splice site resulting in exon skipping and a pathogenic loss of function

Synonymous SNVs were historically considered 'silent' and often neglected in pathogenic variant discovery. It is now widely acknowledged that synonymous SNVs can contribute to a phenotype through altering splicing patterns, miRNA binding, pre-mRNA structure, and translation dynamics, as such, these types of variants have been linked to a plethora of human diseases (Bartoszewski et al. 2010; Brest et al. 2011; Cartegni et al. 2002; Chamary et al. 2006; Duan et al. 2003; Ito et al. 2017; Kimchi-Sarfaty et al. 2007; Tsai et al. 2008; Wang et al. 2015).

Splicing changes are perhaps the best-studied effect of functional synonymous SNVs, however, in a clinical diagnostic setting synonymous variants that create, modify, or eliminate splice sites are often classified as a VUS. This is likely due to an imperfect understanding of RNA splice signals and a lack of commercially available bioinformatic algorithms which can successfully predict pathology of synonymous SNVs (Ito et al. 2017; Macaya et al. 2009; Richards et al. 2015).

I have reported here on a male paediatric patient (individual 54488) who presents with global developmental delay, autism, motor delay, low muscle tone and various dysmorphic features (Trivellin et al. 2018). This individual was found to carry two maternally inherited variants; a 650 kb Xq26.3 duplication suggestive of X-linked acrogigantism and a novel synonymous SNV in the X linked gene *UPF3B* which was designated as a VUS (*UPF3B* c.624G>A).

Individuals affected with X-linked acrogigantism have germline or somatic Xq26.3 duplications and all present with growth hormone and prolactin-secreting pituitary tumours or hyperplasia that are the cause of their gigantism. The size of these microduplications varies amongst affected individuals, however, the smallest region of overlap was recently shown to encompass only one protein-encoding gene, *GPR101*. *GPR101* encodes for an orphan G protein coupled receptor that is over-expressed in the pituitary lesions of affected individuals (Trivellin et al. 2018). Following clinical and endocrine evaluation individual 54488 did not display features of gigantism or acromegaly typical of patients with X-linked acrogigantism. Furthermore, the microduplication in this individual did not encompass *GPR101* and as such a diagnosis of X-linked acrogigantism was rejected and the individual was referred for further molecular investigations into the possible pathogenicity of the VUS in *UPF3B*.

In silico investigations predicted that the *UPF3B* c.624G>A variant would disrupt wild-type splicing of *UPF3B* transcripts. To functionally assess the involvement of this variant in the clinical presentation of individual 54488, LCL cultures were first established for this individual. Molecular investigations were then able to confirm the *in silico* prediction by revealing that the *UPF3B* variant transcripts in individual 54488 lacked exon 6 of *UPF3B*. Furthermore, these investigations showed the frameshift caused by exon 6 skipping, introduces an NMD-targeted PTC resulting in a complete loss of UPF3B protein and stabilised levels of UPF3A protein which is characteristic of UPF3B X-linked ID (Nguyen et al. 2012).

Stabilised UPF3A cannot entirely compensate for loss of UPF3B in the NMD pathway (Nguyen et al. 2012; Tarpey et al. 2007). This is reflected in individual 54488 where the *UPF3B* c.624G>A variant ultimately resulted in inefficient NMD as measured by expression of known classical NMD target transcripts. Interestingly, inefficient NMD of the *UPF3B* variant transcript itself was also observed. NMD is known to reduce the level of a PTC containing transcripts to approximately 5–25% of the nonsense free level (Isken & Maquat 2007), since

UPF3B is X-linked and the affected individual is male at least a 75% reduction in transcript is expected, however only a 50% reduction was observed.

These molecular findings along with a clinical phenotype of X-linked ID akin to previously described individuals harbouring loss of function *UPF3B* variants (Tarpey et al. 2007) has led to a confident genetic diagnosis of X-linked ID for individual 54488 and will facilitate appropriate downstream counselling and treatment. Additionally, this study highlights the importance of both clinical and molecular evaluation in the interpretation of VUS.

3.3.2 The novel UPF2 c.1940delA frameshift variant reduces UPF2 protein expression in LCLs but does not disrupt classical NMD

Individuals harbouring large CNV deletions which include the autosomal NMD factor gene *UPF2*, show an overlapping clinical phenotype and transcriptomic profile to those with *UPF3B* X-linked ID pathogenic variants (Johnson et al. 2019; Nguyen et al. 2013). Compromised NMD is known to be pathogenic for UPF3B X-linked ID (Jolly et al. 2013; Nguyen et al. 2012; Tarpey et al. 2007). As such, a strong theory is that compromised NMD due to a loss of UPF2 is also causative for the neurodevelopmental disorders observed in individuals with large *UPF2* CNV deletions. It is difficult however, to determine whether clinical presentations are caused exclusively by a loss of UPF2, are a consequence of another gene in the impacted region, or are a compound effect from several disrupted genes. In this chapter I have used patient derived LCLs and molecular methods to characterise and compare the impact of the first heterozygous *UPF2* frameshift variant (case 1, *UPF2* c.1940delA) against a group of heterozygous large CNV deletions encompassing *UPF2* (cases 3–5).

The *UPF2* c.1940delA frameshift variant was predicted to introduce an NMD-targeted PTC within the expressed transcript. Molecular investigations were able to support this by revealing

a reduction in overall *UPF2* transcription levels and reduced protein levels akin to that observed in *UPF2* CNV deletion cases. Furthermore, *UPF2* transcript levels could be elevated through inhibition of translation, suggesting that the *UPF2* c.1940delA variant transcript was efficiently targeted by NMD.

An interesting observation was that all individuals with reduced UPF2 protein levels also showed a reduction in UPF3A protein levels. In the classical model of NMD, UPF2 forms a bridge between the NMD components UPF3B and UPF1. Under normal circumstances UPF3A is outcompeted by UPF3B for UPF2 binding and destabilised (Chan et al. 2009). As such, it is possible that when UPF2 protein is limited in LCLs, UPF3A's binding may be further hindered leading to a more rapid destabilisation.

To investigate the effect of loss of UPF2 on the activity of classical NMD, the expression of selected NMD target transcripts which are known to respond to a loss of NMD factors *UPF1* and *UPF3B* (Jolly et al. 2013; Mendell et al. 2004) was assessed in case 1 (*UPF2* c.1940delA) and cases 3–5 (*UPF2* CNV deletions) by qPCR. *ATF4* and *GAS5* transcripts were seen to significantly increase as a consequence of *UPF2* CNV deletions (cases 3–5). Interestingly, these transcripts were not upregulated as a consequence of the *UPF2* c.1940delA frameshift variant (case 1). Moreover, as mentioned, the *UPF2* c.1940delA variant transcript itself was shown to be efficiently degraded by NMD. Together these results suggest that, either classical NMD is not perturbed in response to a heterozygous loss of UPF2 or that the selected NMD mRNA targets investigated are targeted by a UPF2-independent pathway.

The presence of a UPF2-independent pathway was first suggested when tethered *UPF3B* mutants lacking a UPF2-interacting domain were shown to be able to elicit NMD (Kunz et al. 2006). Since then a number of studies have documented further evidence of not only a UPF2-independent NMD branch, but also NMD branches independent of UPF3A, UPF3B or other

EJC components, however it remains unclear if these pathways constitute as regular pathways or reflect transcript-specific or cell type-specific processes (Chan et al. 2007; Gehring et al. 2005; Gehring et al. 2009; Ivanov et al. 2008; Metze et al. 2013).

This study has shown that the *UPF2* c.1940delA frameshift variant, introduces an NMD-targeted PTC resulting in what appears to be a complete loss of protein expression from the variant allele. Unlike large CNVs which result in a heterozygous deletion of *UPF2*, the *UPF2* c.1940delA variant does not appear to reduce the efficiency of classical NMD, at least as measured by the expression of selected mRNAs. Thus, this study alone cannot conclude that a perturbed NMD pathway due to a reduction in UPF2 is the underlying cause of the neurodevelopmental disorders documented for cases 1–5 (Table 3.3). Nevertheless, an interesting observation to motivate further investigations is that the individual harbouring the *UPF2* c.1940delA frameshift variant (case 1) who displayed no evidence of a disrupted NMD pathway compared to the large CNV deletion cases (cases 3–5) was also clinically less severely affected than the large CNV deletion cases, perhaps suggesting that the degree to which NMD is affected correlates with clinical phenotype.

Another path to explore would be investigating the expression of a larger selection of known NMD target transcripts within case 1 (*UPF2* c.1940delA). Although frequently used, *ATF4*, *GAS5* and *GADD45B* may not always serve as a proxy for NMD activity. NMD targeted transcripts are known to be influenced by the cell or tissue type being tested (Jolly et al. 2013; Shum et al. 2016) or due to external influences on transcript expression from alternative cellular pathways, such as the integrated stress response pathway and SMD (Goetz & Wilkinson 2017; Park & Maquat 2013). One method to overcome any bias introduced by selecting a limited number of NMD target transcripts is to conduct whole transcriptome studies (Nguyen et al. 2012). As such, RNA samples isolated from individuals in this study have been prepared for RNA sequencing to further investigate the effect of these *UPF2* variants on whole

transcriptomes. However, the identification and establishment of patient derived LCLs from additional individuals who harbour *UPF2* variants (specifically, those which do not disrupt any other genes) will be necessary to elucidate the global impact of these variants on the NMD pathway and/or branches thereof.

3.3.3 Chapter conclusions and future directions

In this chapter I have used patient derived LCLs to characterise novel variants within the genes encoding NMD factors *UPF3B* or *UPF2*. These variants were identified in individuals who present with varying neurodevelopmental disorders. Through molecular methods a synonymous SNV found in a canonical splice region of *UPF3B* and originally classified as a VUS, was able to be identified as pathogenic and result in patient diagnosis. This study highlighted the need to value any variants identified in *UPF3B* and other genes which when disrupted are known to result in neurodevelopmental disorders, particularly those in splice site regions. Moreover, this study showed that patient derived LCLs serve as an effective starting point to explore variant pathogenicity.

Comparatively, molecular investigations using patient derived LCLs from one of only two variants identified to exclusively disrupt *UPF2* was unable to conclude that perturbed NMD due to a reduction in UPF2 was the underlying cause of neurodevelopmental disorders identified in this patient and by extension, in patients harbouring heterozygous CNV deletions encompassing *UPF2*. This study highlighted the value of large sample sizes which can be subjected to whole transcriptome studies to powerfully investigate the effect of genetic variants on the global transcriptome scale. Without a large sample size, it is important to have ample evidence to support any conclusion. As such, this study would have benefited from the availability of alternative methods to measure NMD activity, such as an NMD reporter system which could compliment or refute findings from conventional RNA expression-based approaches.

Chapter Four: Design and Testing of a Fluorescent NMD Reporter System with Single Cell Resolution (Version 1.0)

4.1 Introduction

NMD is a highly conserved eukaryotic post-transcriptional regulatory pathway. Canonically NMD protects the cell from transcriptome infidelity through identification and degradation of aberrant PTC containing transcripts which arise by transcription error or genetic mutation. It is now known that NMD also plays a second role in maintaining normal cellular homeostasis. Approximately 5–15% of the normal eukaryotic transcriptome is sensitive to inhibition of NMD (Adachi et al. 2004; McIlwain et al. 2010; Mendell et al. 2004; Nguyen et al. 2012; Weischenfeldt et al. 2012; Yepiskoposyan et al. 2011). The impact NMD has on transcriptomes is important for a myriad of cellular processes such as the stress response and neuronal differentiation (Bruno et al. 2011; Karam et al. 2015; Karam & Wilkinson 2012; Martin & Gardner 2015). Both roles of NMD have been implicated in human disease and have thus sparked interest in the pathway as a possible therapeutic target. (Bechara et al. 2013; Ghigna et al. 2005; Hall & Thein 1994; Kerr et al. 2001; Tarpey et al. 2007)

NMD was first discovered in 1979 (Chang & Kan 1979; Losson & Lacroute 1979) and since then the field has gained a considerable level of knowledge into its underlying mechanisms, albeit far from complete. One way to delve deeper into the molecular mechanisms and biological roles of NMD is to implement a technology which can accurately quantify cellular NMD activity levels. As such, both transient and stable NMD reporter systems have been developed and used to facilitate the identification of, (1) cis-acting NMD-target mRNA features (2) Trans-acting NMD factors and (3) small molecules or drugs which can alter the efficiency of NMD (Alexandrov, Shu & Steitz 2017; Boelz et al. 2006; Nickless et al. 2014; Paillusson et al. 2005; Welch et al. 2007). Across these NMD reporter systems, RNA quantification, bioluminescence and fluorescence have all been utilised to provide a quantifiable 'NMD activity' output.

The most used NMD reporter systems are RNA quantification-based reporters. These compare the mRNA levels produced from transgenes encoding a wild-type (NMD refractory) expression cassette to that from transgenes encoding an NMD-targeted expression cassette, both contain sequences from known NMD target genes such as; *HBB*, *TPI*, *TCR-β* and *MUP* (Baserga & Benz 1988; Belgrader & Maquat 1994; Buhler, Paillusson & Muhlemann 2004; Buhler et al. 2006; Carter et al. 1995; Cheng, Fogel-Petrovic & Maquat 1990). RNA quantification-based reporters provide a simple system to measure NMD activity, however they are reliant on transient transfection approaches which are limited by variable transfection efficiencies and can accommodate only fixed endpoint analysis which may involve lengthy RNA quantification methods.

More recently, stable fluorescent and bioluminescent NMD reporter systems have been developed (Nickless et al. 2014; Paillusson et al. 2005; Welch et al. 2007). Since fluorescence and bioluminescence can be quantified using technologies for high-throughput screening, these systems have facilitated the identification of several small molecules that can modify NMD activity (Cheruiyot et al. 2018; Durand et al. 2007; Feng et al. 2015; Keeling et al. 2013; Martin et al. 2014; Popp & Maquat 2015; Welch et al. 2007). Since NMD has been identified as a disease modulator, these findings have been of considerable pharmaceutical interest and were responsible for the identification of the nonsense suppression drug, ataluren (Welch et al. 2007).

A major limitation of both RNA quantification based NMD reporters and stable high-throughput fluorescent and bioluminescent NMD reporters is the inability to quantify NMD activity at a single cell level. Recent developments in NMD research suggests that NMD activity varies across different cell types, states and developmental stages, which can be a major driver of disease presentation and severity (Bateman et al. 2003; Gerbracht, Boehm & Gehring 2017; Thada et al. 2016). To date, only two NMD reporter systems with single cell resolution have been developed. The first is a transient transfection based, dual-fluorescent NMD reporter

system (Pereverzev et al. 2015) and more recently (published half way through candidature) a stable dual-fluorescent NMD reporter system. (Alexandrov, Shu & Steitz 2017). In this type of system, an NMD-refractive mRNA encodes the first fluorescent protein and serves as an internal control. A second fluorescent protein is also encoded by another mRNA which contains an NMD-targeting feature. NMD activity can thus be calculated as a ratio of fluorescence expressed from the NMD-targeted mRNA normalised to fluorescence expressed from the NMD-refractive mRNA. Although able to resolve NMD at a single cell level, the transient NMD reporter system described by Pereverez et al. 2015 is still limited by transfection efficiency and involves co-transfection of two separate constructs. Comparatively the NMD reporter system described by Alexandrov et al. 2017 is designed for chromosomal integration, however, is thus far limited to use in a cancer cell line (HeLa), and involves random chromosomal integration with variable copy numbers.

In summary, major limitations of the current NMD reporter systems include, transfection efficiency of one or more constructs, restricted to use in a homogenous population of cells, and lack of single cell resolution. Due to these design limitations, very few studies have applied an NMD reporter system to explore the dynamic nature of NMD during development or to investigate the cell, tissue, or species-specific activity of NMD. Investigating these complex functions of NMD are key to understanding both its endogenous role in the cell and its role as a disease modulator. With this in mind, there is a need for an improved NMD reporter system which can provide accurate information on the endogenous activity of the NMD pathway *in vitro* and *in vivo* with single cell resolution and in different cells and tissue types undergoing dynamic changes such as during development.

To address the shortcomings of existing NMD reporter systems. I envisaged a stable system which provides a quantifiable, fluorescent output of NMD activity with single cell resolution. Moreover, targeted integration of this system into the genome of pluripotent stem cells would

provide the opportunity to study NMD in any cell type. In this way, NMD activity could also be tracked across differentiation pathways involved in the development of many cell types. Furthermore, being an integrated genomic system, it would not rely on transient transfection efficiency and could be readily used in high-throughput assays.

Such a system can be engineered for use in mouse embryonic stem cells (mESCs). This approach begins with the design of a fluorescent NMD reporter transgene followed by functional *in vitro* validation of the transgene and targeted integration into genome of mESCs. The resulting NMD reporter mESC line can be used to investigate NMD as described in Figure 4.1. Moreover, an exciting possibility which stems from this line of experiments is the potential to generate a transgenic NMD reporter mouse model. This would allow NMD activity to be investigated and visualised *in vivo* during development and into adult life (Figure 4.1)

Chapters 4–6 will describe the scientific techniques used to design, engineer, and experimentally test two fluorescent NMD reporter transgenes which aim to report on endogenous NMD activity at the single cell level. This chapter will focus on the first version of this transgene.

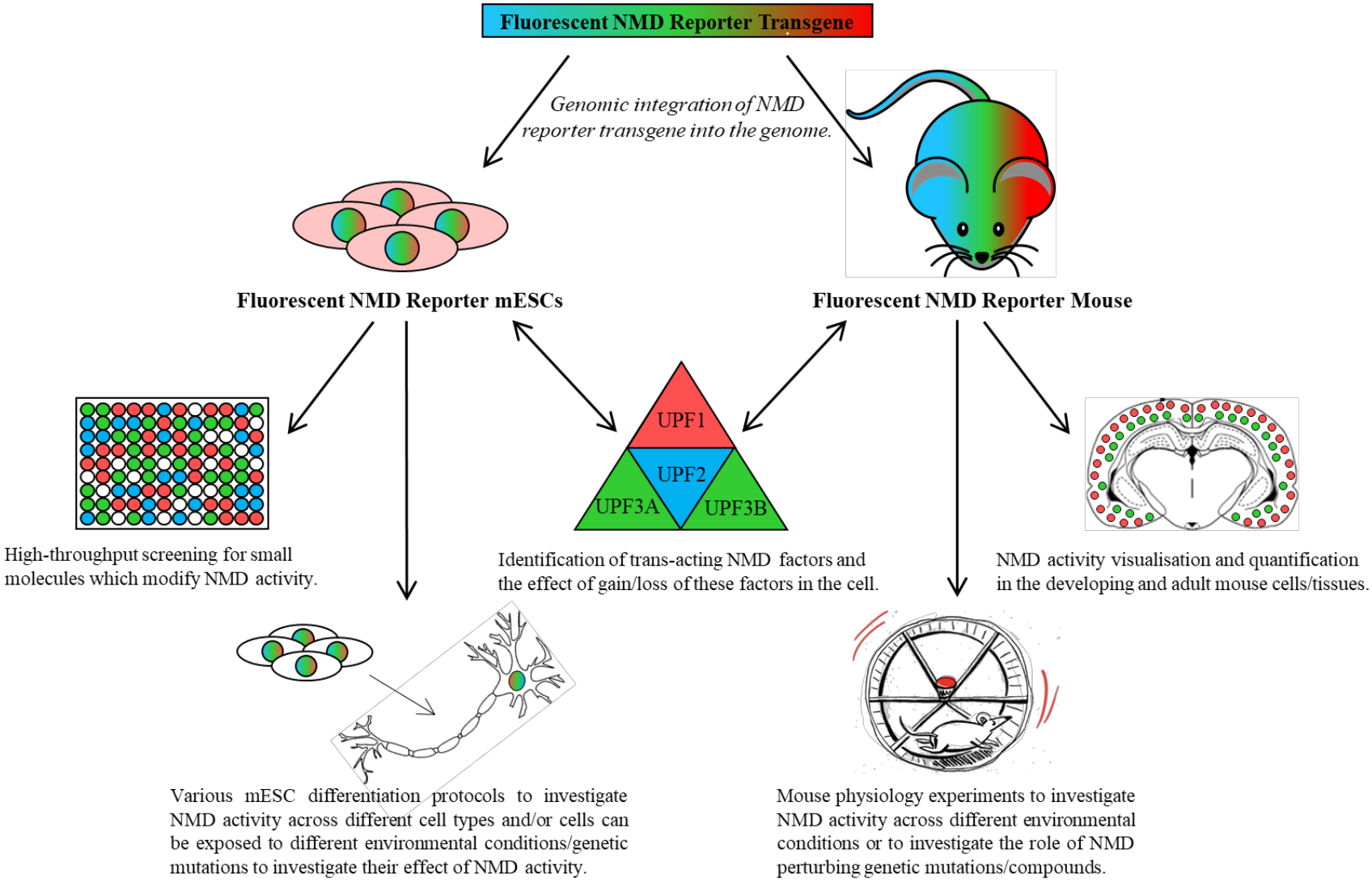


Figure 4.1: A schematic summary of the potential of an NMD reporter transgene which can be stably integrated into the genome of mESCs and/or mice.

4.2 Design

4.2.1 A summary of the design and expected output of Transgene$^{V1.0}$

A functional NMD reporter transgene was designed with the end goal being to investigate endogenous NMD activity within cells and in a developing NMD reporter mouse. This transgene was required to encode a dual-fluorescent NMD reporter system with a visual and quantifiable output of NMD at a single cell level. This transgene must also incorporate features which enable targeted integration into the genome of mESCs, or into mouse zygotes using genetic modification approaches.

Fluorescent output was chosen as it allows single cell quantification using fixed or live cells through fluorescence microscopy and/or FACS. This will enable NMD activity to be tracked across development, giving insight into when, where and in which cell types NMD is most important. Tracking NMD across neural development is of particular interest given that a compromised NMD pathway due to genetic variants disrupting NMD factor genes *UPF2* and *UPF3B* results in neurodevelopmental disorders (Chapter 3).

The first design and expected output of an NMD reporter transgene (Transgene Version 1.0, Transgene$^{V1.0}$) is illustrated in Figure 4.2. It employs four expression cassettes, namely the; Selection Cassette$^{V1.0}$ (Selection$^{V1.0}$), Control Cassette$^{V1.0}$ (Control$^{V1.0}$), NMD Cassette$^{V1.0}$ (NMD$^{V1.0}$) and the TetR Responder Cassette$^{V1.0}$ (TetR Responder$^{V1.0}$). NMD$^{V1.0}$ and TetR Responder$^{V1.0}$ are linked via a Tetracycline repressor (TetR)-based gene circuit. NMD$^{V.1.0}$ is designed to constitutively express an NMD-targeted mini gene encoding TetR fused to nuclear localised enhanced green fluorescent protein (TetR:EGFP^{NLS}), while TetR Responder$^{V1.0}$ will encode nuclear localised red fluorescent protein (tdTomato^{NLS}) under the transcriptional control of Tet operon (TetO) sequences. In cells where NMD activity is low, TetR:EGFP^{NLS} expression will be favoured, whereas in cells where NMD activity is high a switch to tdTomato^{NLS} expression will occur.

Control$^{V1.0}$ was designed to constitutively express a nuclear localised cyan fluorescent protein (CFP^{NLS}). This cassette was included to serve as an internal control for the NMD reporter system to correct for inherent noise in gene expression, i.e. non-NMD based influences on cassette expression.

Finally, to facilitate integration of the entire transgene into cells and/or mouse zygotes Selection$^{V1.0}$ was included. Selection$^{V1.0}$ enables Recombination Mediated Cassette Exchange (RMCE) as previously described (Bersten et al. 2015). Its function is to target Transgene$^{V1.0}$ to the Collagen 1a1 (*Col1a1)* locus of genetically modified FLP-in mESCs or FLP-in mouse zygotes. The design of each cassette will be further detailed in Section 4.2.2.

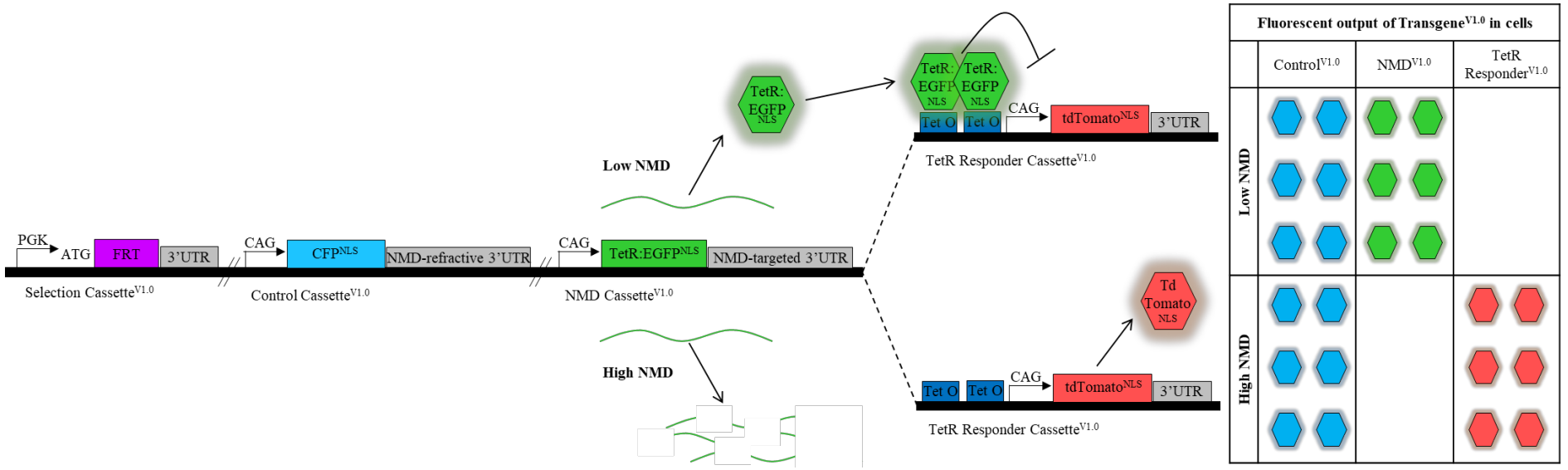


Figure 4.2: A schematic representation of Transgene $^{V1.0}$ and its fluorescence output in cells with low or high NMD activity.

Control$^{V1.0}$ encodes CFP^{NLS}. Following splicing, transcripts expressed from this cassette are not targeted by NMD, and thus serve as an internal control which expresses a constant level of CFP^{NLS} regardless of NMD activity. NMD$^{V1.0}$ encodes the fusion protein TetR:EGFP^{NLS}. Following splicing, transcripts expressed from this cassette are targeted for NMD. In cells with low NMD activity, translation of TetR:EGFP^{NLS} can occur. TetR:EGFP^{NLS} will bind to the TetO sites of TetR Responder$^{V1.0}$ and repress expression of tdTomato^{NLS}. These cells will therefore fluoresce green. In cells with high NMD activity transcripts expressed from NMD$^{V1.0}$ are degraded by NMD and the absence of TetR:EGFP^{NLS} allows uninhibited tdTomato^{NLS} expression from TetR Responder$^{V1.0}$. These cells will therefore fluoresce red. Ratios of EGFP:tdTomato:CFP provide an NMD activity reading. This can be measured at the single cell level due to nuclear localisation signals (NLS). Selection$^{V1.0}$ facilitates genomic integration into FLP-in mESCs via RMCE.

4.2.2 A detailed description of the design and expected output of Transgene$^{V1.0}$

4.2.2.1 Expression plasmids encoding the cassettes of Transgene$^{V1.0}$

Transgene$^{V1.0}$ contains four cassettes; Selection$^{V1.0}$, Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$. The sequences for these cassettes were designed using the plasmid editor 'ApE' (Davis 2019), and synthesised individually under a fee-for-service contract with Genscript (Genscript, USA). Following synthesis, Genscript provided these sequences ligated into the multiple cloning site of the plasmid backbone pUC57, these plasmids are described in Table 4.1 and illustrated in Figure 4.3. By beginning with the sequence for each cassette within its own plasmid, the unique features of these cassettes can be functionally validated prior to assembly into Transgene$^{V1.0}$ (Figure 4.4).

To facilitate assembly of all four cassettes into Transgene$^{V1.0}$, the design of each cassette included several restriction endonuclease recognition sites which can be used in ligase based cloning methods. Each cassette was also designed with two homology arms on either end to permit transgene assembly via Gibson isothermal assembly. Additional restriction sites were also included within all cassettes to facilitate swapping in and out of entire cassettes or unique features or to facilitate introduction of sequence modifications if necessary (Figure 4.4).

Table 4.1: A description of Version 1.0 expression plasmids and the proteins they encode

Plasmid	Supplier	Expression cassette(s) included in plasmid				Encoded Protein(s)	Expected protein size(s)
		Selection (S)	Control (C)	NMD (N)	TetR Responder (T)		
pUC57-$S^{V1.0}$	Genscript	$S^{V1.0}$	-	-	-	-	-
pUC57-$C^{V1.0}$	Genscript	-	$C^{V1.0}$	-	-	CFP^{NLS}	~28 kDa
pUC57-$N^{V1.0}$	Genscript	-	-	$N^{V1.0}$	-	TetR:$EGFP^{NLS}$	~52 kDa
pUC57-TetO-$T^{V1.0}$	Genscript	-	-	-	$T^{V1.0}$	$tdTomato^{NLS}$	~56 kDa
pUC57-$CN^{V1.0}$	In-house	-	$C^{V1.0}$	$N^{V1.0}$	-	CFP^{NLS}, TetR:$EGFP^{NLS}$	~28 kDa, ~52 kDa
pUC57-$SNC^{V1.0}$	In-house	$S^{V1.0}$	$C^{V1.0}$	$N^{V1.0}$	-	CFP^{NLS}, TetR:$EGFP^{NLS}$	~28 kDa, ~52 kDa

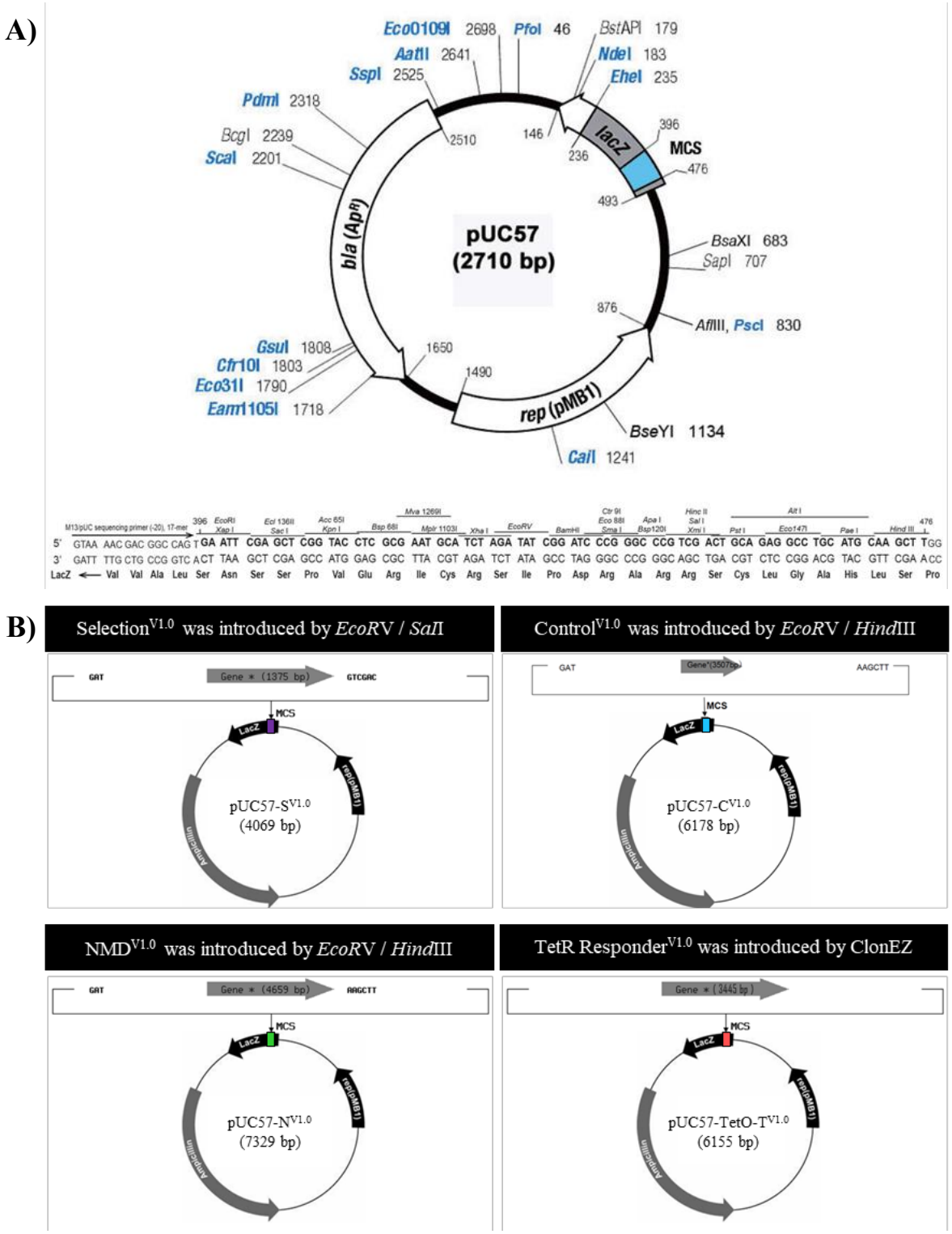


Figure 4.3: Preparation of expression plasmids encoding the cassettes of Transgene$^{V1.0}$ via the contracted services of Genscript.

(A) A plasmid map of the pUC57 plasmid backbone detailing the sequence of the multiple cloning site (MCS) and restriction endonuclease recognition sites. **(B)** A schematic representation of the restriction endonucleases/cloning process used by Genscript to introduce synthesised cassette DNA sequences into the MCS of pUC57.

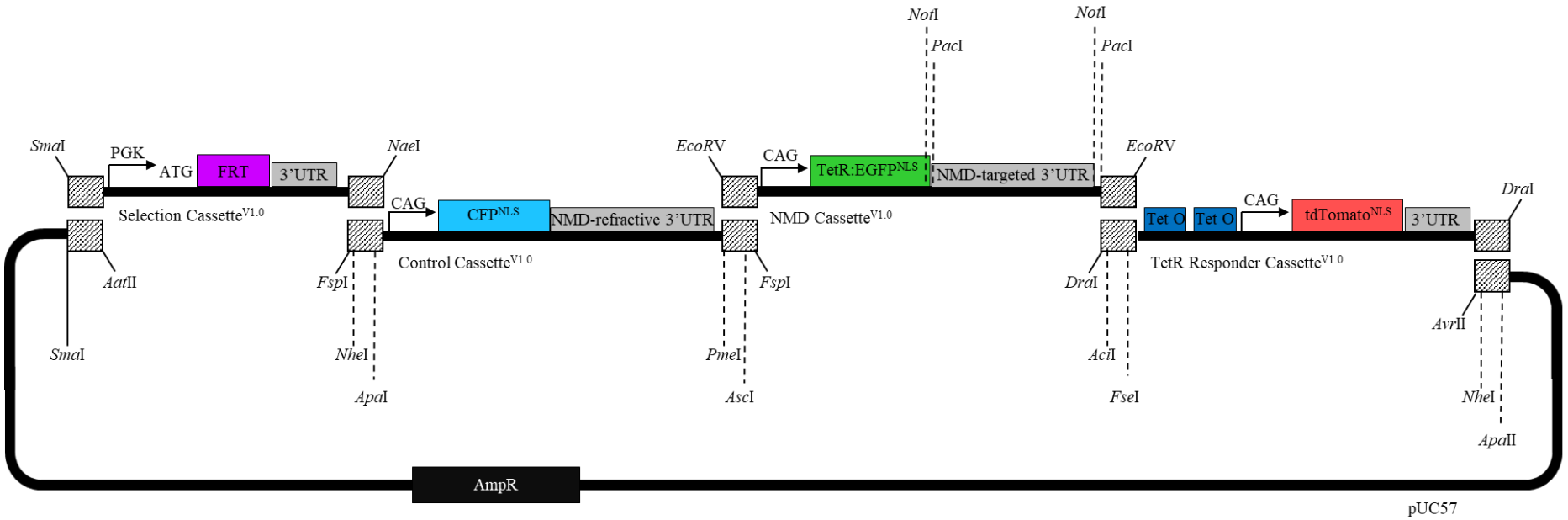


Figure 4.4: A detailed schematic of Transgene$^{V1.0}$.

Transgene$^{V1.0}$ is composed of four cassettes in cis; namely the, Selection$^{V1.0}$, Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$. The plasmid backbone used to house this transgene is pUC57. All cassettes apart from Selection$^{V1.0}$ encode protein. Each cassette is driven by its own promoter and flanked by two homology arms (shaded boxes). The homology arms or the restriction sites they contain can facilitate directed assembly of the cassettes and the pUC57 backbone through Gibson isothermal assembly or through conventional restriction endonuclease mediated cloning methods, respectively. Restriction sites within these homology arms have also been engineered to drop in and out entire cassettes (bold lines) or specific features within a cassette (dashed lines) to allow modification of the transgene if necessary.

.

4.2.2.2 Design and function of Selection$^{V1.0}$

Selection$^{V1.0}$ (Figure 4.5A) is one of four cassettes which make up Transgene$^{V1.0}$. This cassette acts as a tool to facilitate integration of the entire Transgene$^{V1.0}$ as single stable copy into the *Col1a1* locus of FLP-in mESCs (or FLP-in mouse zygotes) using RMCE (Figure 4.5B). This system was previously described by our collaborators Professor Murray Whitelaw and Dr David Bersten (Bersten et al. 2015).

To establish a stable NMD reporter mESC line, Transgene$^{V1.0}$ is required to be introduced via co-transfection alongside an expression vector for flippase (FLP) recombinase. FLP recombinase recognises the flippase recognition target (FRT) sites present within Selection$^{V1.0}$ and the *Col1a1* locus of FLP-in mESCs. This results in excision of the amino 3'-glycosyl phosphotransferase (*NeoR)* gene from the *Col1a1* locus and linearisation of, and subsequent exchange with Transgene$^{V1.0}$. This modification to the *Col1a1* locus positions a promoter and start codon (encoded by Selection$^{V1.0}$) upstream of a *HygroR* gene (encoded in the *Col1a1* locus of FLP-in mESCs) to drive its expression. *HygroR* encodes Hygromycin B Phosphotransferase, which confers selectable resistance to the antibiotic hygromycin, and thus provides an efficient means to select correct transgene integration events (Figure 4.5B).

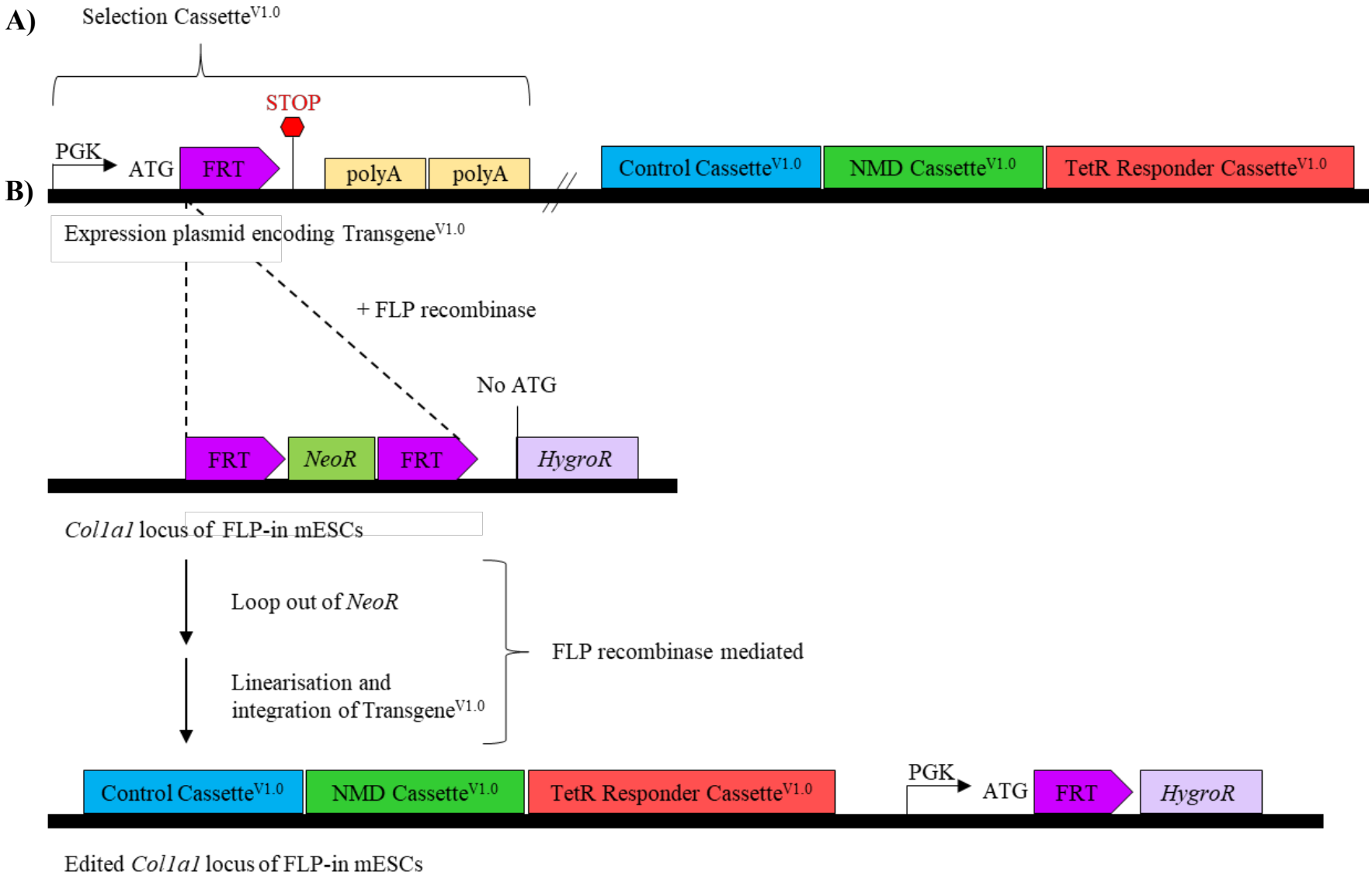

A)
Selection Cassette V1.0
STOP
PGK
ATG
FRT
polyA
polyA
Control Cassette V1.0
NMD Cassette V1.0
TetR Responder Cassette V1.0
B)
Expression plasmid encoding Transgene V1.0
+ FLP recombinase
No ATG
FRT
NeoR
FRT
HygroR
Col1a1 locus of FLP-in mESCs
Loop out of NeoR
Linearisation and integration of Transgene V1.0
FLP recombinase mediated
Control Cassette V1.0
NMD Cassette V1.0
TetR Responder Cassette V1.0
PGK
ATG
FRT
HygroR
Edited Col1a1 locus of FLP-in mESCs

Figure 4.5: Selection$^{V1.0}$ facilitates Recombination Mediated Cassette Exchange (RMCE) of Transgene$^{V1.0}$ into the Col1a1 locus of FLP-in mESCs.

(A) A schematic representation of Selection$^{V1.0}$. Selection$^{V1.0}$ does not encode any protein. It contains a PGK promoter sequence followed by a flippase recognition target (FRT) sequence and two SV40 poly adenylation (polyA) sequences. This cassette is designed to facilitate RMCE into the *Col1a1* locus of FLP-in mESCs (mESCs pre-engineered with *Col1a1* FRT sequences and a promoter-less hygromycin resistance coding sequence (*HygroR*), which also lacks a start codon). **(B)** A schematic representation of Selection$^{V1.0}$ mediated RMCE of Transgene$^{V1.0}$ into the *Col1a1* locus of FLP-in mESCs. Following co-transfection of an expression plasmid encoding Transgene$^{V1.0}$ and an expression plasmid encoding FLP recombinase into FLP-in mESCs, FLP recombinase will recognise the FRT sites present within Selection$^{V1.0}$ and the *Col1a1* locus of FLP-in mESCs. This will result in excision of *NeoR* from the *Col1a1* locus and linearisation of, and exchange with Transgene$^{V1.0}$. This modification to the *Col1a1* locus positions the PGK promoter and start codon encoded by Selection$^{V1.0}$ upstream and in-frame with the hygromycin B Phosphotransferase (*HygroR*) coding region, therefore, reconstituting a functional *HygroR* gene. Expression of *HygroR* enables selection of correct recombination events by providing cellular resistance to hygromycin.

4.2.2.3 Design and function of Control$^{V1.0}$ and NMD$^{V1.0}$

Control$^{V1.0}$ is expressed independent of NMD activity levels in the cell therefore serving as an internal control for Transgene$^{V1.0}$. In comparison, NMD$^{V1.0}$ contains a gene feature which targets its transcripts for NMD. This feature is referred to as the 'NMD-targeted 3' untranslated region (3'UTR)'. The design of this feature is a minor variation of the 'NMD-refractive 3'UTR' encoded by Control$^{V1.0}$ (Figure 4.6). Both of these 3'UTRs are based on a previously published NMD reporter system design (Pereverzev et al. 2015).

To escape or undergo NMD, Control$^{V1.0}$ and NMD$^{V1.0}$ are designed to exploit the 55 nt rule of NMD. This rule states that only PTCs residing greater than 55 nts upstream of the final exon-exon junction are recognised by NMD. If this distance is less than 55 nts, the downstream EJC will be dislodged by the ribosome (due to the ribosomal footprint) and the transcript will escape NMD recognition. The termination codons of CFP^{NLS} and TetR:EGFP^{NLS} encoded by Control$^{V1.0}$ and NMD$^{V1.0}$ respectively are placed upstream of an exon-exon junction, which is deposited following splicing due to the presence of an intron within the 3'UTR of both cassettes. These 3'UTRs are derived from a sequence of *HBB* which incorporates exon 2–intron 2–exon 3. However, differing lengths of *HBB* exon 2 were included in Control$^{V1.0}$ and NMD$^{V1.0}$ to adjust the distance between the termination codon and the exon-exon junction.

The NMD-targeted 3'UTR contains 537 bp of *HBB* exon 2. As such, the TetR:EGFP^{NLS} stop codon is placed greater than 55 nts upstream of the exon-exon junction, rendering the transcript sensitive to NMD. When NMD activity is high TetR:EGFP^{NLS} expression will be low and vice versa (Figure 4.6). Conversely, the NMD-refractive 3'UTR is shortened to contain only 35 bp of *HBB* exon 2. In this case, the CFP^{NLS} stop codon is placed less than 55 nts upstream of the exon-exon junction thus blinding transcript expressed from Control$^{V1.0}$ to the NMD recognition machinery and allowing constitutive, NMD-independent expression of CFP^{NLS} (Figure 4.6).

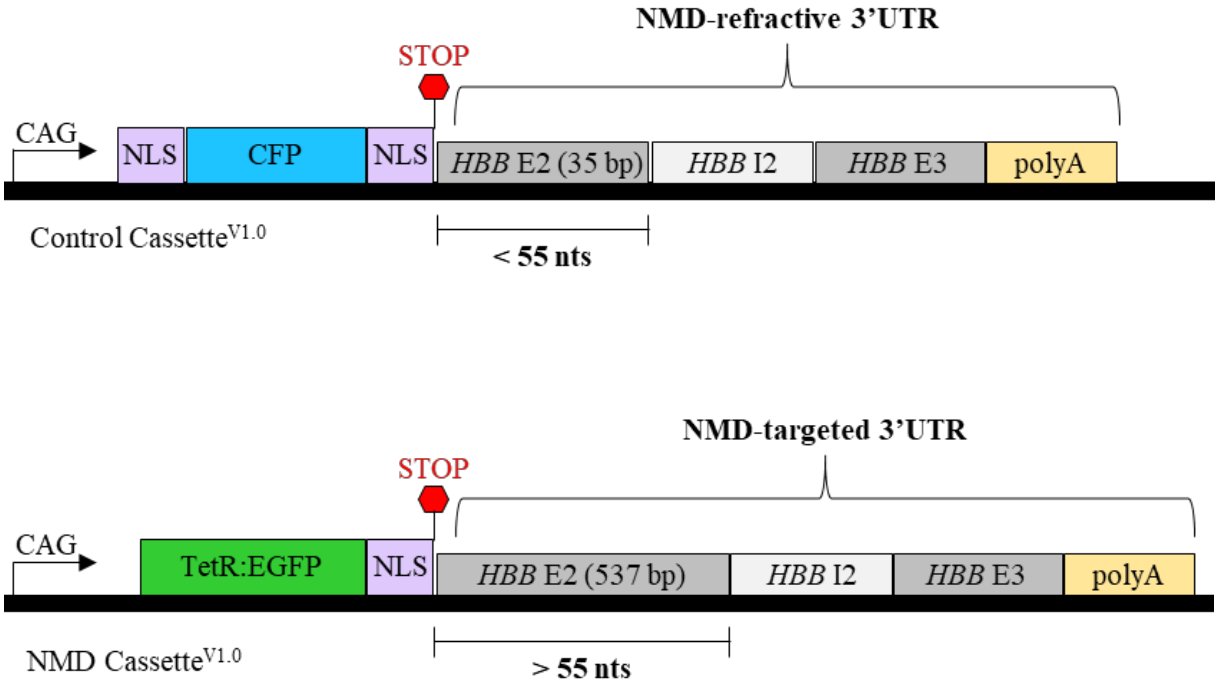


Figure 4.6: A schematic comparison between the design of Control$^{V1.0}$ and NMD$^{V1.0}$.

Expression from Control$^{V1.0}$ and NMD$^{V1.0}$ is driven by a CAG promoter. Control$^{V1.0}$ is designed to function as an internal control for Transgene$^{V1.0}$. Specifically, the NMD-refractive 3'UTR of Control$^{V1.0}$ acts as a direct control to the NMD-targeted 3'UTR of NMD$^{V1.0}$. Both 3'UTRs include an exon 2–intron 2–exon 3 sequence from *HBB*, subjecting the transcript to splicing and introduction of an exon junction complex (EJC). *HBB* exon 2 sequence of the NMD-refractive 3'UTR is only 35 bp long. Splicing will therefore place the CFP^{NLS} stop codon < 55 nts upstream of the exon-exon junction, as such the resulting EJC is dislodged by the ribosome and transcripts expressed from Control$^{V1.0}$ will escape NMD recognition. Comparatively, *HBB* exon 2 of the NMD-targeted 3'UTR is 537 bp long. Following splicing, this will position the TetR:EGFP^{NLS} stop codon in the context of a PTC i.e. > 55 nts upstream of an exon-exon junction. Consequently, transcripts expressed from NMD$^{V1.0}$ will be subject to NMD. Proteins encoded by Control$^{V1.0}$ and NMD$^{V1.0}$ both contain at least one nuclear localisation signal (NLS) to direct proteins into the nucleus and enable single cell quantification of fluorescence, even in complex tissues.

4.2.2.4 Design and function of a Tetracycline repressor (TetR)-based gene circuit between NMD$^{V1.0}$ and TetR Responder$^{V1.0}$

Expression from Control$^{V1.0}$ and NMD$^{V1.0}$ can function together to report on NMD efficiency, however, if NMD is found be highly efficient in certain contexts, interpreting high levels of NMD based on a lack of EGFP signal may become inaccurate. To address this, a reciprocal fluorescent signal which is expressed in a directly proportional manner to NMD efficiency was also designed. The rationale of reporting not only on the presence of NMD activity, but also on its absence, was that such a design might facilitate more accurate and robust transgene function across a spectrum of different NMD activity levels, dynamics, and contexts.

To achieve this TetR Responder$^{V1.0}$ was designed to exploit the TetR-based gene circuit and work in synchrony with NMD$^{V1.0}$ (Figure 4.7A). This system relies on two elements, the first is expression of TetR protein which is provided through expression of the TetR:EGFP^{NLS} fusion protein encoded by NMD$^{V1.0}$. The second element needed is a TetR responsive cassette, for example, one containing consensus TetR binding sites (TetO) within the promoter, which in the presence of TetR will be occupied to repress downstream expression. This was provided by TetR Responder$^{V1.0}$ which was designed to incorporate a series of TetOs upstream of a CAG promoter and drives expression of the nuclear localised fluorescent protein tdTomato^{NLS}. In this way, transcription of tdTomato^{NLS} is repressed by TetR binding the TetO sites within the CAG promoter. In cells harbouring both NMD$^{V1.0}$ and TetR Responder$^{V1.0}$, low levels of NMD activity will result in increased expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ which can occupy the TetO binding sites of TetR Responder$^{V1.0}$ to inhibit transcription of tdTomato^{NLS}. The opposite is true for cells with high NMD activity levels (Figure 4.7B).

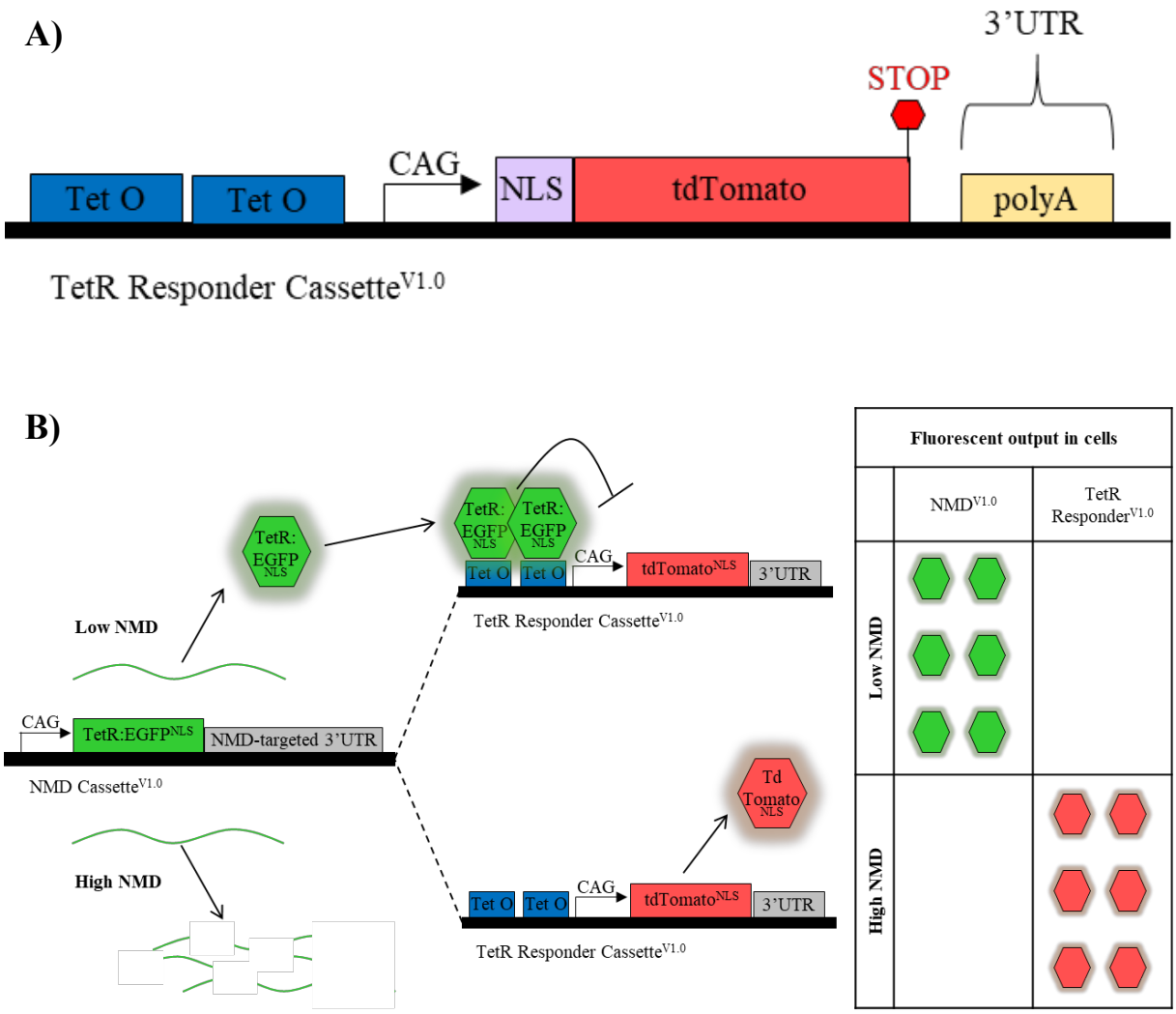


Figure 4.7: A schematic representation of TetR Responder$^{V1.0}$ and its function within a tetracycline repressor-based gene circuit with NMD$^{V1.0}$ in cells with low and high NMD activity.

(A) A schematic representation of TetR Responder$^{V1.0}$. This cassette encodes nuclear localised tdTomato (tdTomato^{NLS}) from a CAG promoter controlled by Tet operon sites (TetO). When these sites are bound by TetR, transcription of tdTomato^{NLS} is inhibited and vice versa. **(B)** A schematic representation of fluorescent output from cells simultaneously expressing NMD$^{V1.0}$ and TetR Responder$^{V1.0}$. If these cells have low NMD activity, TetR:EGFP^{NLS} expressed from NMD$^{V1.0}$ will bind the two TetO sites of TetR Responder$^{V1.0}$ to inhibit tdTomato^{NLS} expression. If NMD activity is high, transcripts from NMD$^{V1.0}$ are degraded by NMD and TetR:EGFP^{NLS} expression is lost. This results in un-repressed expression of tdTomato^{NLS} from TetR Responder$^{V1.0}$. In this way tdTomato^{NLS} expression is proportional to NMD activity in the cell.

4.3 Results

4.3.1 Sequence validation of Genscript synthesised expression plasmids

To ensure that the contracted synthesis of all DNA sequences for the cassettes of Transgene$^{V1.0}$ was completed without error, sequence validation was initially carried out via diagnostic restriction endonuclease digestion of expression plasmids encoding Selection$^{V1.0}$, Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$ (pUC57-S$^{V1.0}$, pUC57-C$^{V1.0}$ pUC57-N$^{V1.0}$ and pUC57-TetO-T$^{V1.0}$ respectively). The digested DNA products, alongside undigested plasmid DNA for comparison were visualised by agarose gel electrophoresis (Figure 4.8). The observed product sizes matched the expected digest product sizes suggesting that there were no large sequence errors. Selected constructs were then subject to Sanger sequencing (data not shown) to ensure the fidelity of each cassette prior to use in subsequent experiments. In each case the sequence of the entire cassette i.e. Seclection$^{V1.0}$, Control$^{V1.0}$, NMD$^{V1.0}$ or TetR Responder$^{V1.0}$ (excluding the pUC57-kan backbone) was sequenced and verified.

Plasmid	**pUC75-N$^{V1.0}$**	**pUC75-S$^{V1.0}$**	**pUC75-C$^{V1.0}$**	**pUC75-TetO-T$^{V1.0}$**
Cassette	NMD$^{V1.0}$	Selection$^{V1.0}$	Control$^{V1.0}$	TetR Responder$^{V1.0}$
Restriction endonucleases	*Pvu*I + *BamH*I	*Pvu*I + *Xho*I	*Pvu*I + *Sal*I	*Pvu*I + *Sal*I
Expected product sizes (bp)	3534, 2899, 896	8968, 3435, 738	5083, 896, 200	4203, 600

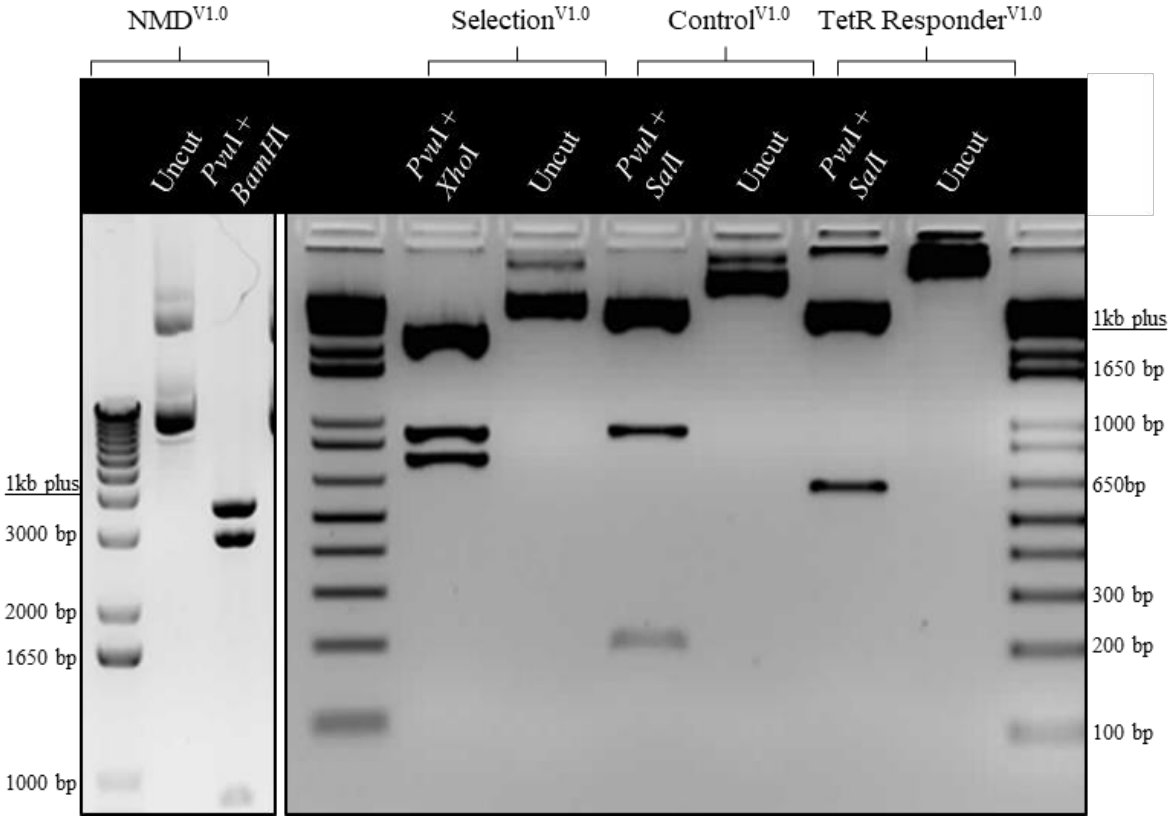


Figure 4.8: Diagnostic restriction endonuclease digests supports correct sequence synthesis of expression plasmids encoding the cassettes of Transgene$^{V1.0}$.

Genscript synthesised expression plasmids (pUC57-N$^{V1.0}$, pUC57-S$^{V1.0}$, pUC57-C$^{V1.0}$ and pUC57-TetO-T$^{V1.0}$) were subject to diagnostic restriction endonuclease digests. The table summarises the plasmids, the cassette they express, the restriction endonucleases used for digestion and the expected digest product sizes following digestion of a correctly synthesised plasmid. The digest product alongside undigested (uncut) plasmid DNA was visualised by agarose gel electrophoresis. The expected band sizes were observed for all expression plasmids.

4.3.2 Functional testing of Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$ using expression plasmids in a transient setting

To establish a stable NMD reporter system, Transgene$^{V1.0}$ was designed to be integrated into the genome of FLP-in mESCs (or FLP-in mouse zygotes) using the previously established RMCE system which involves Selection$^{V1.0}$ (Bersten et al. 2015). Prior to assembly of Transgene$^{V1.0}$ however, it is important to assess the function of Control$^{V1.0}$, NMD$^{V1.0}$, and TetR Responder$^{V1.0}$. To enable functional testing of each cassette, expression plasmids encoding either Control$^{V1.0}$, NMD$^{V1.0}$ or TetR Responder$^{V1.0}$ (pUC57-C$^{V1.0}$, pUC57-N$^{V1.0}$ or pUC57-TetO-T$^{V1.0}$ respectively) were transfected into cells in culture. Initial experiments were carried out using a mouse embryonic fibroblast cell line (NIH3T3) as a precursor to functional testing in mESCs or using human HEK293T cells. Transfected cells were treated with Control or *UPF1* siRNA which was used to reduce cellular NMD activity levels. To enable downstream analysis, the transfected cells were either fixed onto coverslips or used to isolate protein and RNA (Figure 4.9).

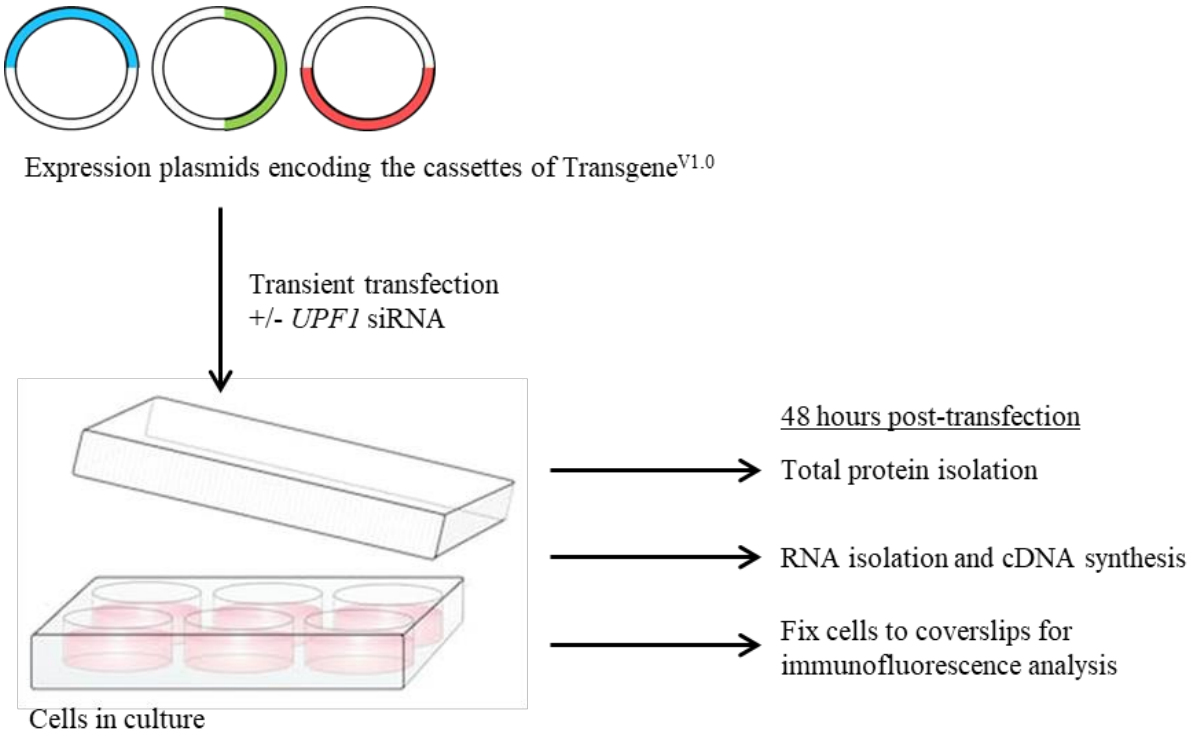


Figure 4.9: The experimental pipeline used for functional testing of expression plasmids encoding the cassettes of Transgene$^{V1.0}$.

A schematic representation of the basic molecular pipeline used to test the features of Control$^{V1.0}$, NMD$^{V1.0}$, and TetR Responder$^{V1.0}$.

4.3.2.1 Using fluorescence expression to assess function of basic cassette features

Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$ share several identical or similar sequences encoding basic features of an expression cassette, these are, the promoter, fluorescent proteins, and nuclear localisation signals (NLSs) (Table 4.2). Fluorescence microscopy analysis of NIH3T3 cells transiently transfected with expression plasmids encoding either Control$^{V1.0}$, NMD$^{V1.0}$ or TetR Responder$^{V1.0}$ (pUC57-C$^{V1.0}$, pUC57-N$^{V1.0}$ or pUC57-TetO-T$^{V1.0}$) revealed that fluorescent proteins were expressed from these plasmids and were restricted to the nuclear compartments (Figure 4.10). This concluded that the promoters are functional, pre-mRNA splicing events can generate translationally competent protein encoding mRNA, and that the fluorescent proteins are functional as are their NLSs.

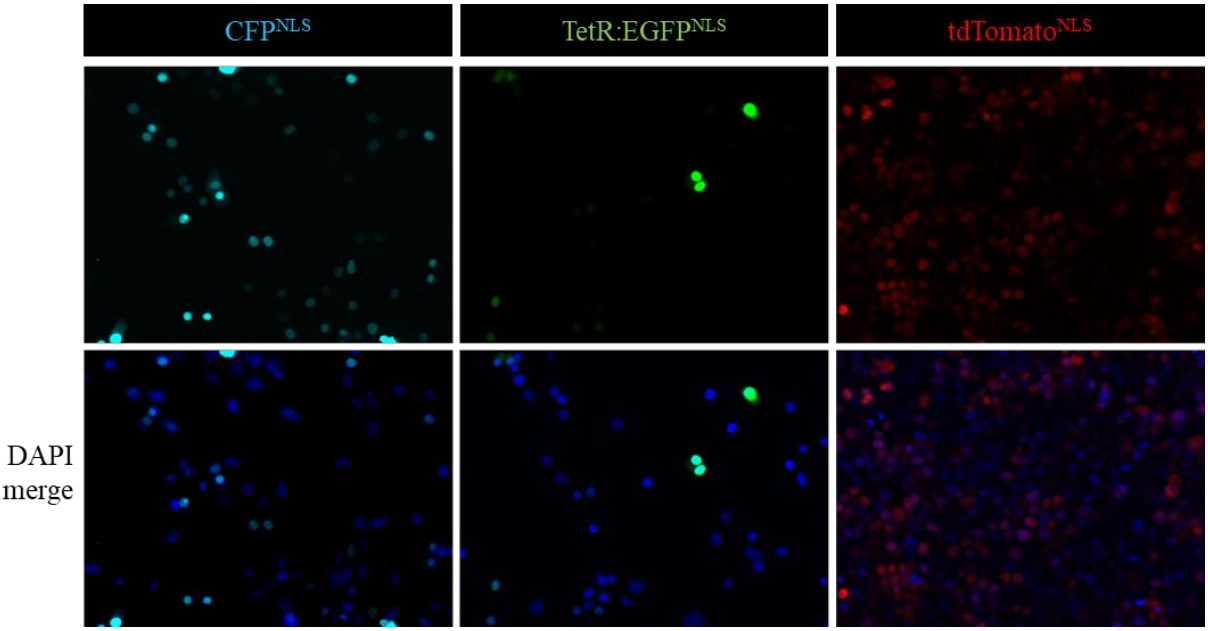


Figure 4.10: Fluorescent imaging of cells expressing Control$^{V1.0}$, NMD $^{V1.0}$ and TetR Reponder$^{V1.0}$ supports correct function of basic features.

Representative fluorescence microscopy images of NIH3T3 cells transfected with expression plasmids encoding Control$^{V1.0}$ (pUC57-C$^{V1.0}$), NMD$^{V1.0}$ (pUC57-N$^{V1.0}$) or TetR Repsonder$^{V1.0}$ (pUC57-TetO-T$^{V1.0}$). These cells express CFP^{NLS} from Control$^{V1.0}$ (cyan), TetR:EGFP^{NLS} from NMD$^{V1.0}$ (green) or tdTomato^{NLS} from TetR Responder $^{V1.0}$ (red). Cell nuclei were counterstained with DAPI (blue).

Table 4.2: Common features across cassettes of Transgene$^{V1.0}$ and their descriptions

Feature	**Function**	**Logic**
CAG promoter	Drives expression of genes encoded by an expression cassette of Transgene$^{V1.0}$.	The CAG promoter is a strong synthetic promoter known to drive moderate constitutive expression in all cell types. It is composed of (C) the cytomegalovirus (CMV) early enhancer element, (A) the promoter, the first exon and the first intron of the chicken beta-actin gene and (G) the splice acceptor of the rabbit beta-globin gene.
Fluorescent protein expression sequences	Encodes a fluorescent protein which is self-sufficient to form a visible wavelength chromophore.	Fluorescence is both visual and quantifiable at a single cell level through various methods. This makes it ideal to visualise and quantify NMD activity in cells. Three fluorescent proteins were designed to express from Transgene$^{V1.0}$; CFP, EGFP and tdTomato.
Nuclear localisation signal (NLS)	Fused to fluorescence proteins to limit their expression to the nucleus of any cell.	NLS were added to the fluorescence proteins so that in the case where many cells and/or cell types are visualised simultaneously the fluorescence expression can be easily attributed to nucleus of one specific cell i.e. to provide single cell resolution.

Homology arms	Overlapping sequences which also contain a unique set of restriction endonucleases.	All cassettes have been designed with two flanking homology arms which also contain a specific set of restriction endonuclease sites to allow easy transgene assembly through Gibson isothermal assembly or standard recombinant DNA cloning methods respectively. Restriction sites also facilitate swapping in and out new cassettes and/or or modifications if required.
SV40 poly adenylation signal (polyA)	Terminator sequence that signals the end of a transcriptional unit.	SV40 polyA sequences are placed at the end of each cassette to prevent read through transcription from the promoter of one cassette into the next cassette.

4.3.2.2 Testing fluorescent spectral resolution of proteins expressed from Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$

When designing a fluorescent reporter, the choice of fluorescent protein(s) used is influenced by several factors. These include; sufficient brightness to be easily distinguished from autofluorescence, high photostability for lengthy imaging experiments, and if the fluorescent protein is to be part of a fusion protein it should not be able to oligomerise easily (Shaner et al. 2005). CFP, EGFP and tdTomato were the three fluorescent proteins chosen for use within Transgene$^{V1.0}$, their properties are summarised in Table 4.3, and their emission and excitation plots shown in Figure 4.11A.

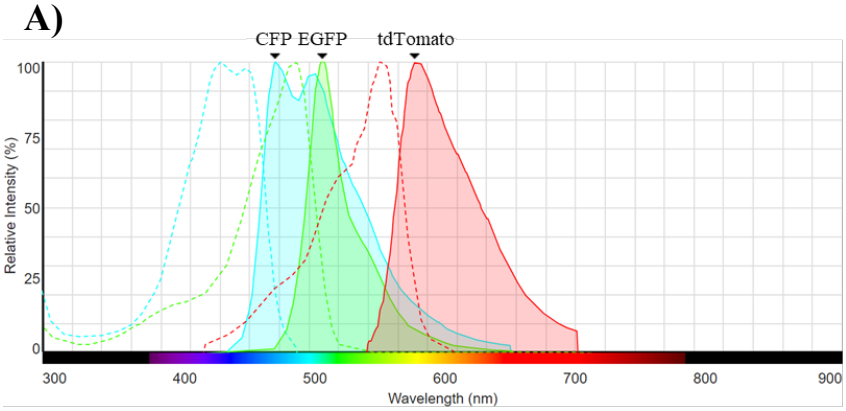
The Zeiss AxioImager M2 fluorescent microscope (Carl Zeiss, Germany) was used in fluorescent imaging analysis, the filter sets available for this microscope were able to spectrally resolve tdTomato and EGFP from each other and from CFP (Figure 4.11B). CFP and EGFP share similar emission and excitation wavelengths (Table 4.3) as such, the filter set available to detect CFP (emission filter 436/20, excitation filter 480/40) was unable to completely spectrally resolve CFP (expressed from pUC57-C$^{V1.0}$) from EGFP (expressed from pUC57-N$^{V1.0}$) (Figures 4.11B & 4.11C). This observation means that in scenarios where Control$^{V1.0}$ and NMD$^{V1.0}$ are expressed simultaneously, signals from EGFP fluorescent molecules can also be detected to an appreciable degree using the filter set specific for CFP, i.e. EGFP signal expressed from NMD$^{V1.0}$ can compound the levels of CFP visualised and detected from Control$^{V1.0}$ using this microscope and filter set. In such situations it will therefore be ideal to use additional methods of protein quantification such as western blot analysis.

Table 4.3: Properties of the fluorescent proteins designed to express from Transgene$^{V1.0}$

Class	**Protein**	**Excitation[a]**	**Emission[b]**	**Brightness[c]**	**Photostability[d]**	**Oligomerisation**	**Molecular weight**	**References**
Cyan	Cerulean (CFP)	433 nm	475 nm	27	36 $t_{1/2}$ (s)	Weak dimer	26.8 kDa	(Lelimousin et al. 2009; Shaner, Steinbach & Tsien 2005)
Green	Enhanced GFP (EGFP)	488 nm	507 nm	34	50.1 $t_{1/2}$ (s)	Weak dimer	26.9 kDa	(Shaner, Steinbach & Tsien 2005; Zhong et al. 2019)
Red	tdTomato	554 nm	581 nm	95	70 $t_{1/2}$ (s)	Tandem dimer	54.2 kDa	Shaner, Steinbach & Tsien 2005)

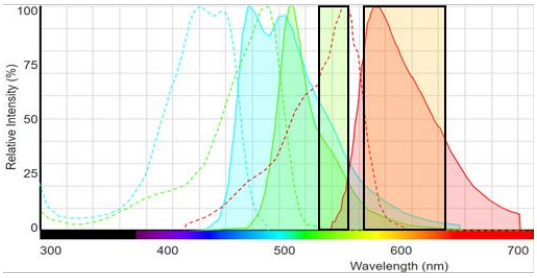
(a) Major excitation peak. (b) Major emission peak. (c) Product of extinction coefficient and quantum yield at pH 7.4 (d) Time for bleaching from an initial emission rate of 1000 photons/s down to 500 photons/s ($t_{1/2}$) in seconds (s).

A)

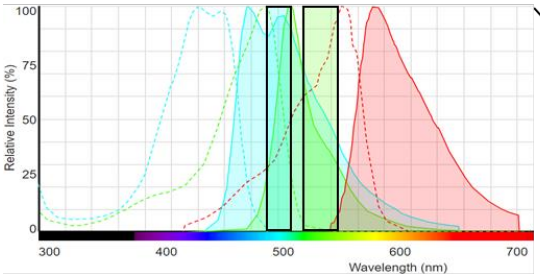


B)

Excitation Filter: 545/25, Emission Filter: 605/70 – detects tdTomato only

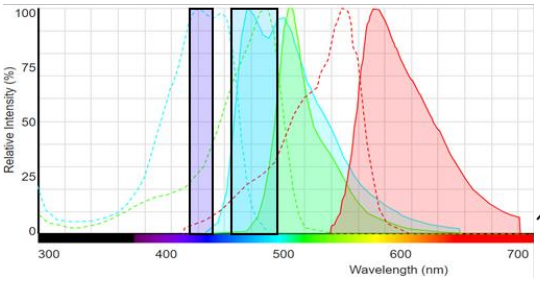


Excitation Filter: 500/20, Emission Filter: 535/30 – detects EGFP only



Excitation Filter: 436/20, Emission Filter: 480/40 – detects CFP

+ some EGFP



C)

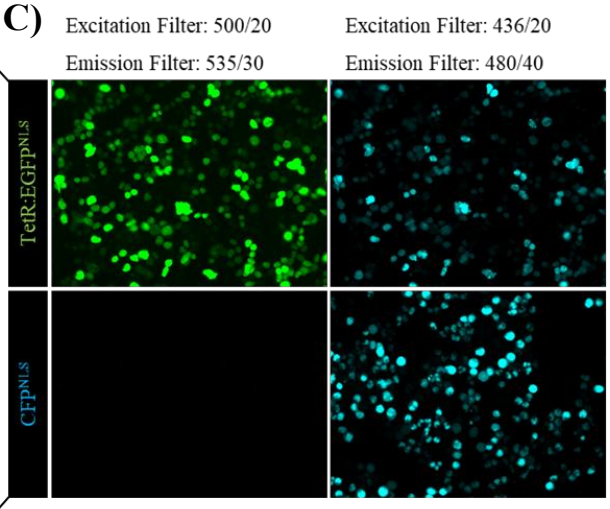


Figure 4.11: CFP cannot be spectrally resolved from EGFP.

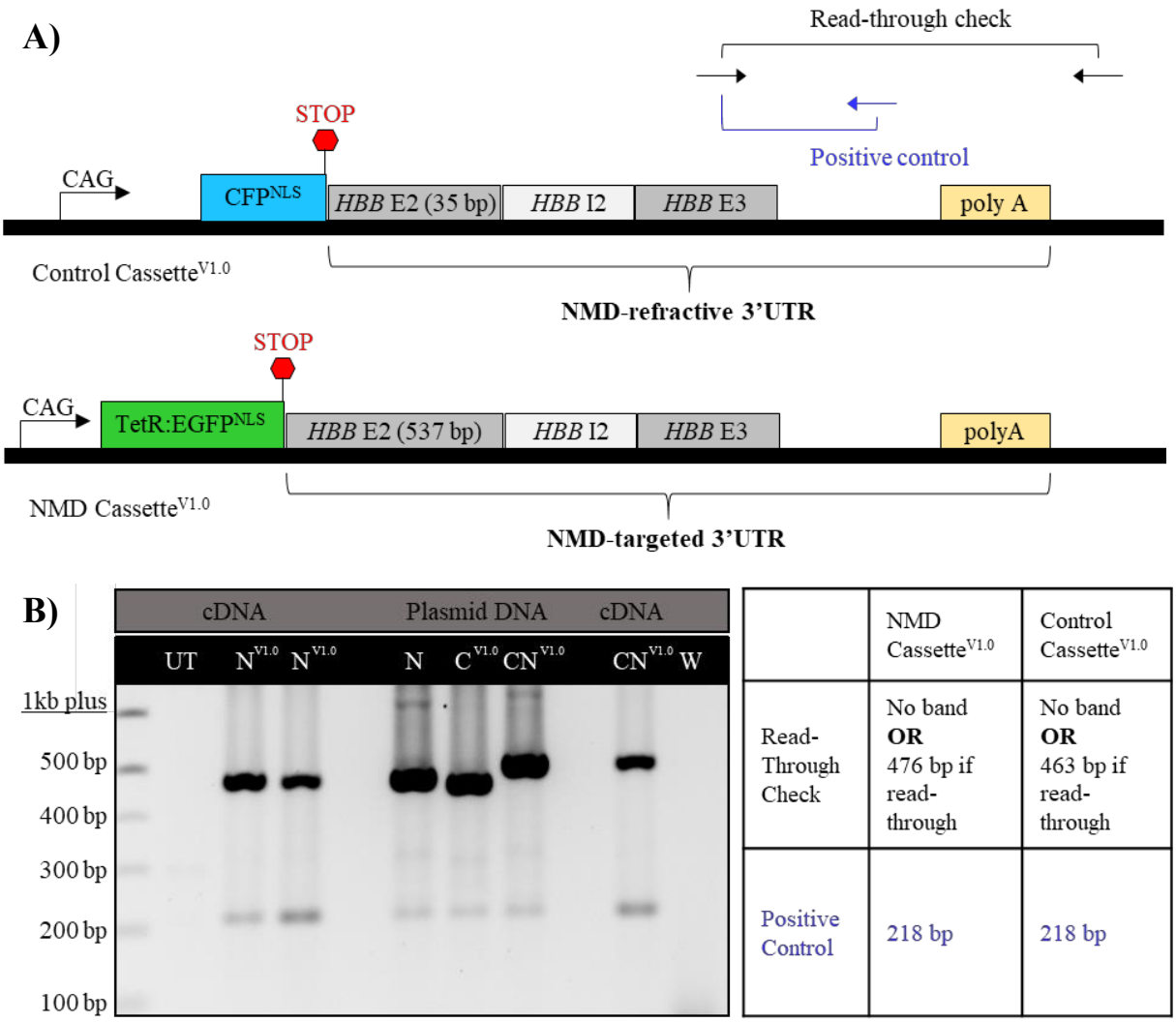
(A) Emission and excitation plots for CFP, EGFP and tdTomato. **(B)** Overlap of excitation and emission filter sets used to resolve tdTomato (excitation filter: 545/25, emission filter: 605/70) EGFP (excitation filter: 500/20, emission filter: 535/30) and CFP (excitation filter: 436/20, emission filter: 480/40). **(C)** Representative fluorescence microscopy images of HEK293T cells transfected with expression plasmids encoding $Control^{V1.0}$ (pUC57-$C^{V1.0}$) or $NMD^{V1.0}$ (pUC57-$N^{V1.0}$). These cells express CFP^{NLS} from $Control^{V1.0}$ (cyan) and TetR:$EGFP^{NLS}$ from $NMD^{V1.0}$ (green). Using excitation filter 436/20 and emission filter 480/40 CFP cannot be completely spectrally resolved from EGFP.

4.3.2.3 Testing if a single SV40 polyA sequence can efficiently terminate transcription downstream of a CAG promoter

Control$^{V1.0}$, NMD $^{V1.0}$ and TetR Responder$^{V1.0}$ all contain a single SV40 polyA sequence at the end of their 3'UTR. The SV40 polyA is a sequence-based transcriptional terminator which functions to define the end of a transcriptional unit and initiate the release of newly synthesised RNA from the transcription machinery. If the SV40 polyA is unable to efficiently terminate transcription, RNA polymerase read-through at the SV40 polyA site of one cassette could influence expression of the downstream cassette.

To assess the efficiency of transcription termination from Control$^{V1.0}$ and NMD$^{V1.0}$, three PCR primers (one forward and two reverse primers) were designed to perform in the same reaction. Two primers either amplify a sequence within the transcriptional unit to act as a positive control for PCR conditions (positive control) or span the SV40 polyA site to assess if transcription termination is efficient (read-through check) (Figure 4.12A). Read-through check primers will only amplify a product if RNA polymerase read-through beyond the SV40 polyA site is occurring i.e. transcription termination is not efficient. cDNA reverse transcribed from the RNA of HEK293T cells transfected with an expression plasmid encoding NMD$^{V1.0}$ (pUC57-N$^{V1.0}$) or both Control$^{V1.0}$and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) (see Section 4.3.2.4.3) was subject to PCR using these primers.

PCR product was then visualised by agarose gel electrophoresis. The positive control primers returned a clear product at the expected size, confirming that the PCR conditions used were appropriate (Figure 4.12B). It was also seen that read-through check primers returned an abundance of amplified product at a size indicating that RNA polymerase read-through beyond the SV40 polyA site was occurring. It was concluded that a single SV40 polyA sequence was insufficient to terminate transcription from the CAG promoter in this context (Figure 4.12B).



	NMD Cassette$^{V1.0}$	Control Cassette$^{V1.0}$
Read-Through Check	No band **OR** 476 bp if read-through	No band **OR** 463 bp if read-through
Positive Control	218 bp	218 bp

Figure 4.12: A single SV40 polyA sequence downstream of a CAG promoter cannot efficiently terminate transcription.

(A) A schematic representation of primers designed to assess the efficiency of transcription termination from Control$^{V1.0}$ and NMD$^{V1.0}$ which both contain a single SV40 polyA sequence within their 3'UTR. **(B)** cDNA reverse transcribed from the RNA of HEK293T cells transfected with expression plasmids encoding NMD$^{V1.0}$(pUC57-N$^{V1.0}$, N$^{V1.0}$) or both Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$, CN$^{V1.0}$) alongside cDNA reverse transcribed from untransfected HEK293T cells (UT) and water (W) as negative controls were subject to PCR using the three described primers. Plasmid DNA was also subject to the same PCR as controls to identify products indicative of RNA polymerase read-through at the SV40 polyA sequence(s). The table outlines the possible PCR products. Banding patterns confirm that transcription termination was inefficient from all cassettes tested.

4.3.2.4 Testing expression from Control$^{V1.0}$ and NMD$^{V1.0}$ in response to changes in NMD activity levels of NIH3T3 cells.

4.3.2.4.1 Using Upf1 siRNA to establish low NMD activity in NIH3T3 cells

Following experimental testing of the basic cassette features, the unique features of each cassette were tested. NMD$^{V1.0}$ contains the NMD responsive element of the transgene, i.e. the NMD-targeted 3'UTR. To test the function of this feature in NIH3T3 cells it was necessary to establish conditions of high and low NMD activity *in vitro*. The most common method of cellular NMD inhibition is to reduce the levels of the core NMD factor UPF1 via small interfering RNA (siRNA). siRNAs are designed to interfere with the expression of specific target genes (i.e. those which harbour a complementary nucleotide sequence to that of the siRNA by degrading their mRNA after transcription and thus preventing downstream translation into protein).

Either Control or *Upf1* siRNA was transiently introduced into NIH3T3 cells and 48 hours later levels of Upf1 protein were assessed by immunofluorescence and western blot. By both analyses, introduction of *Upf1* siRNA resulted in an almost complete loss of Upf1 protein expression (Figure 4.13A & 4.13B). Furthermore, to investigate the impact of this loss of Upf1 on NMD activity, the expression of three known NMD target mRNAs (Jolly et al. 2013; Mendell et al. 2004) was assessed via qPCR. cDNA reverse transcribed from RNA of Control or *Upf1* siRNA treated NIH3T3 cells was used as a template in these qPCR investigations. This revealed that accompanying a decrease in *Upf1* expression was an increase in the expression of NMD target genes *Atf4, Gas5* and *Snord22*. This data shows that an siRNA mediated loss of *Upf1* in NIH3T3 cells diminishes NMD activity (Figure 4.13C).

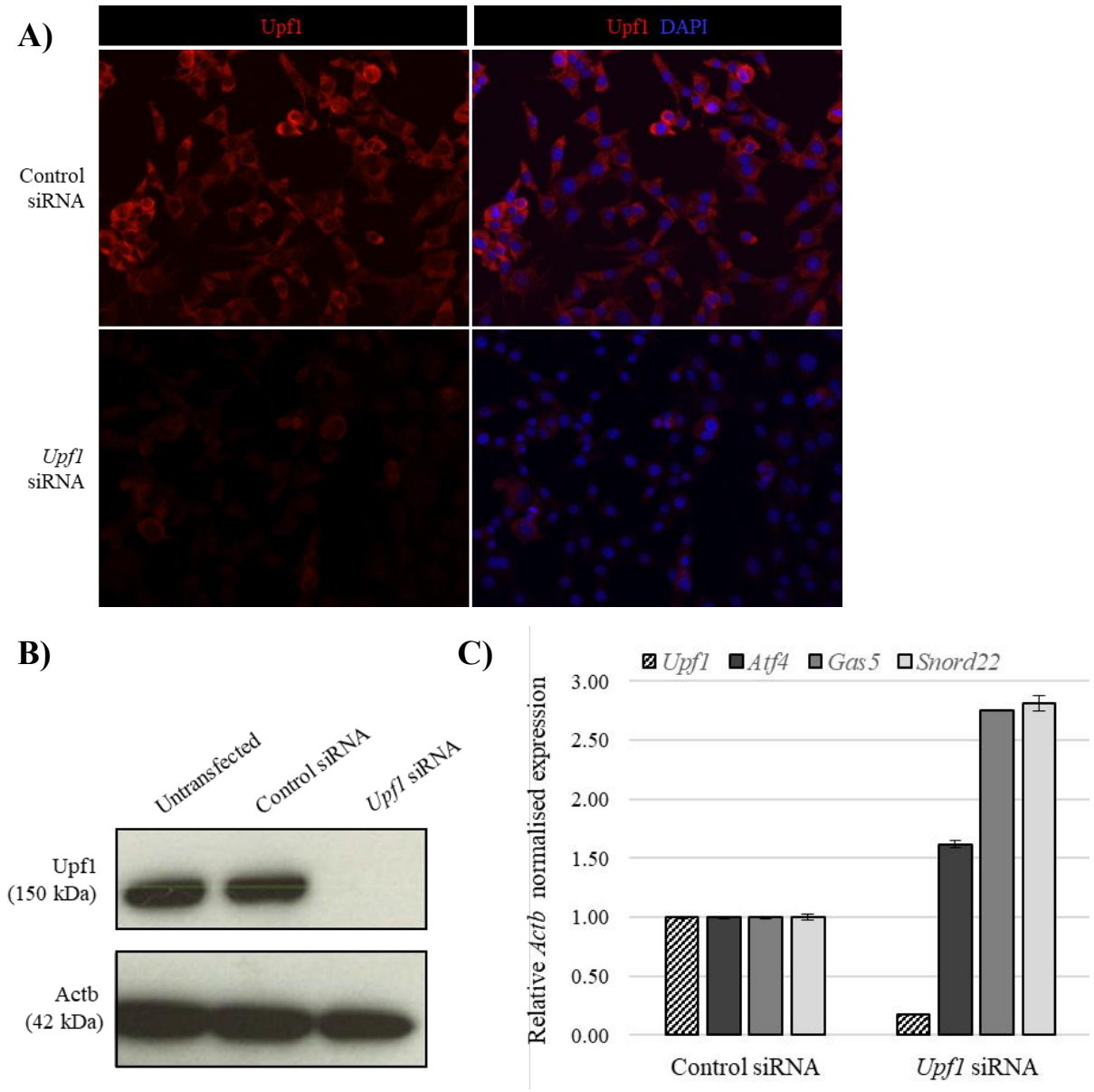


Figure 4.13: Upf1 siRNA reduces NMD activity in NIH3T3 cells.

NIH3T3 cells were transfected with Control or *Upf1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. Upf1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells and from untransfected NIH3T3 cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect Upf1 and the loading control Actb. **(C)** cDNA reverse transcribed from the RNA of these cells was used as a template in qPCR to determine mean expression (±standard deviation) of *Upf1, Atf4, Gadd45B* and *Snord22*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *Actb* mRNA expression in the same sample.

4.3.2.4.2 Testing protein expression from NMD$^{V1.0}$ of pUC57-N$^{V1.0}$ in response to an siRNA mediated reduction of Upf1 levels in NIH3T3 cells

Initial experiments to test if the NMD-targeted 3'UTR of NMD$^{V1.0}$ was responsive to cellular NMD activity were conducted by co-transfecting the expression plasmid encoding NMD$^{V1.0}$ (pUC57-N$^{V1.0}$) alongside either Control or *Upf1* siRNA into NIH3T3 cells. Under Control siRNA conditions, transcripts expressed from NMD$^{V1.0}$ should be degraded by NMD, and thus following a Upf1 reduction (i.e. a reduction in NMD activity) an increase in expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ is expected.

Protein isolated from NIH3T3 cells co-transfected with an expression plasmid encoding NMD$^{V1.0}$ and either Control or *Upf1* siRNA was visualised by western blot or immunofluorescence. In the presence of *Upf1* siRNA, the expected increase in TetR:EGFP^{NLS} expression was not observed (Figure 4.14). Considering that the NMD-targeted 3'UTR feature being tested was based on a previous publication (Pereverzev et al. 2015) it was reasoned that factors other than the feature itself, such as, a difference in transfection efficiencies in cells treated with either Control or *Upf1* siRNA could be confounding the results. To combat this, I aimed to ligate Control$^{V1.0}$ and NMD$^{V1.0}$ together and use this construct for further testing. In this way an internal control will be introduced to the system.

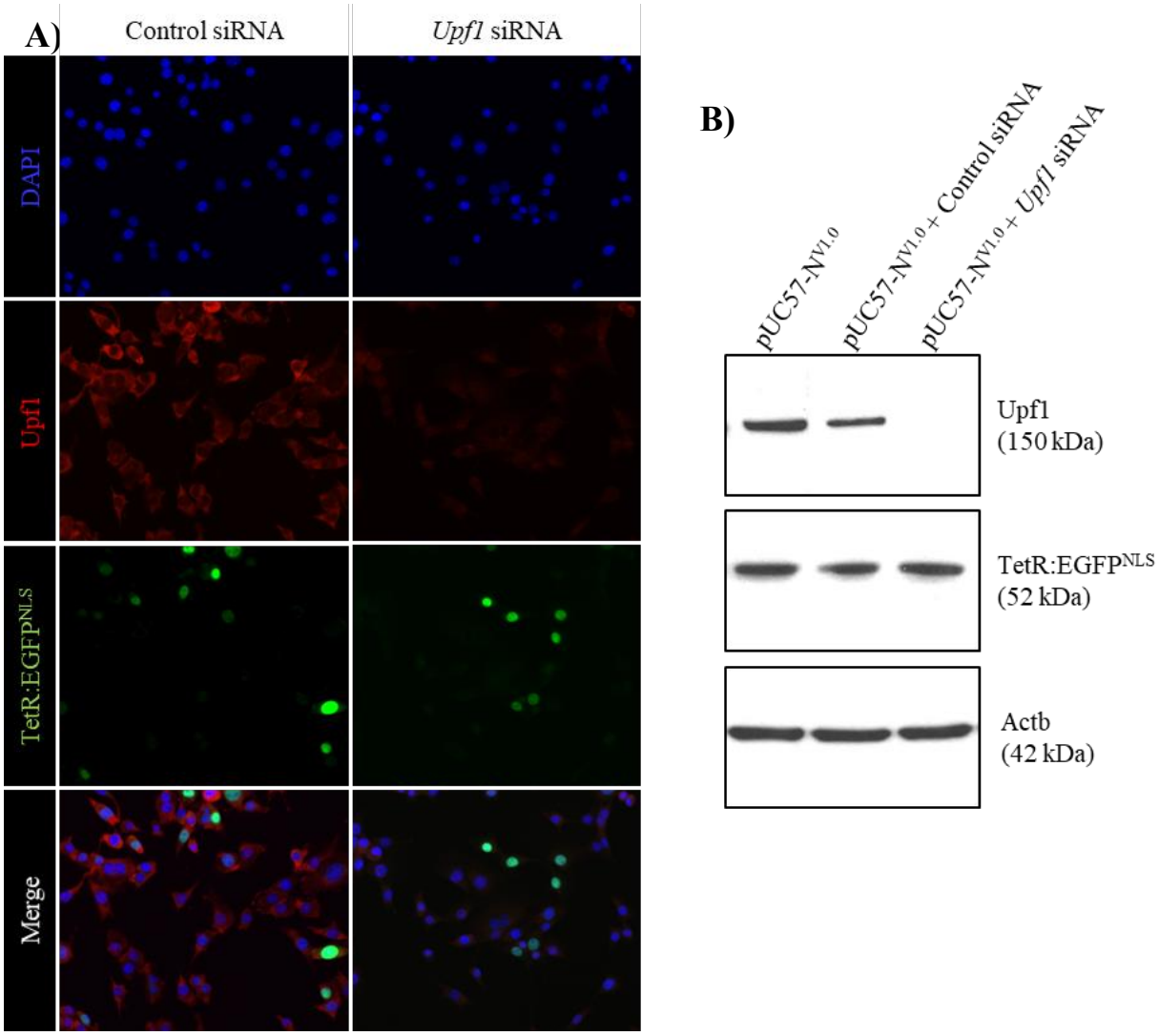


Figure 4.14: Protein expression from NMD$^{V1.0}$ is not responsive to a reduction in UPF1 protein levels following transient transfection in NIH3T3 cells.

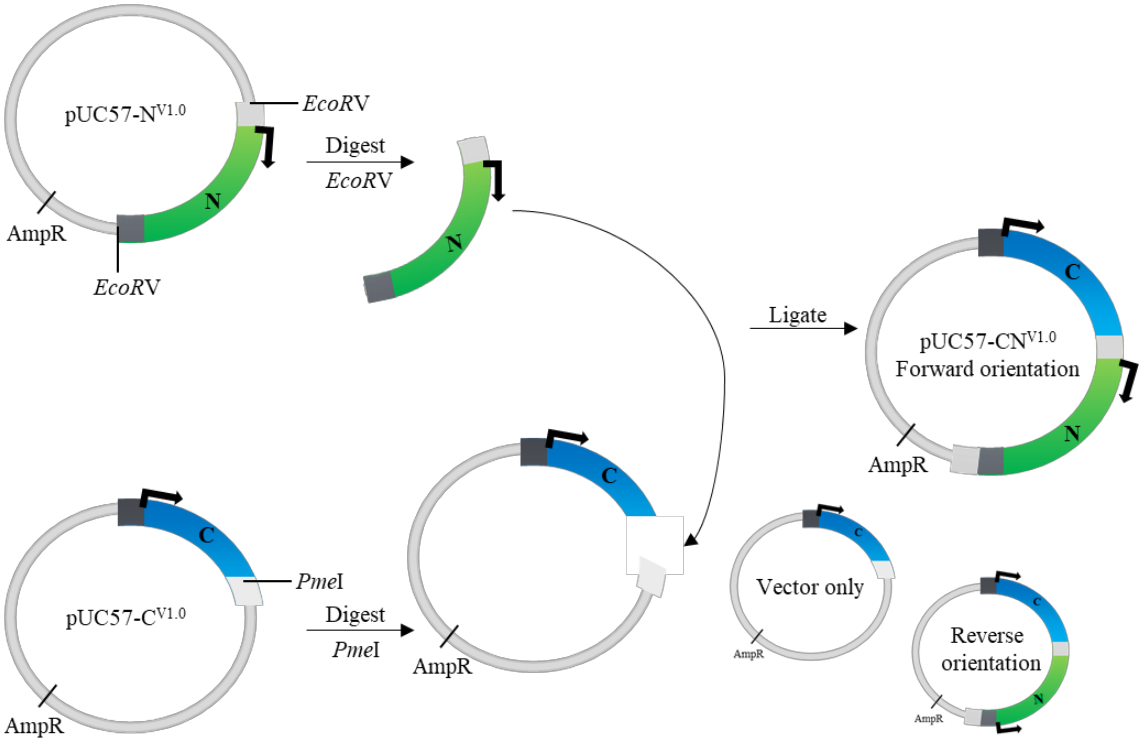
NIH3T3 cells were co-transfected with an expression plasmid encoding NMD$^{V1.0}$ (pUC57-N$^{V1.0}$) and either Control or *Upf1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express TetR:EGFP^{NLS} from NMD$^{V1.0}$ (green), Upf1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells and from NIH3T3 cells transfected only with pUC57-N$^{V1.0}$ was separated by SDS-PAGE and analysed by western blot using antibodies to detect Upf1, TetR and the loading control Actb.

4.3.2.4.3 Engineering pUC57-CN$^{V1.0}$—an expression plasmid encoding Control$^{V1.0}$ and NMD$^{V1.0}$

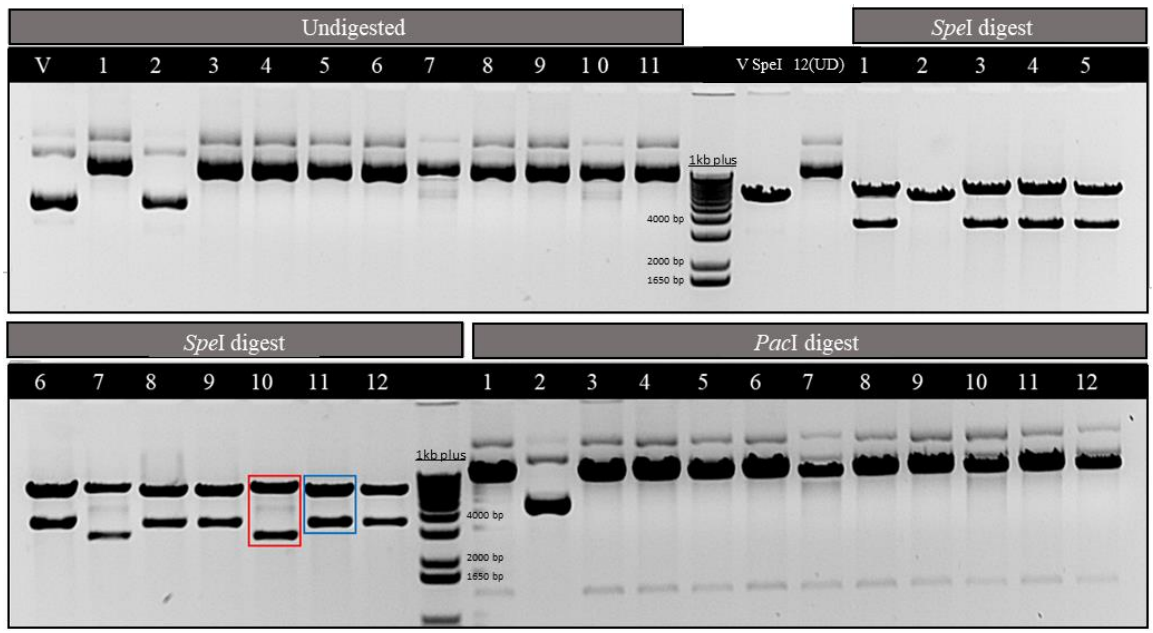
To engineer Control$^{V1.0}$ and NMD$^{V1.0}$ together standard recombinant DNA cloning methods were applied. NMD$^{V1.0}$ was excised from its pUC57 backbone using flanking *EcoR*V restriction sites, and then ligated into the *Pme*I site of pUC57-C$^{V1.0}$ (Figure 4.15A). Ligation reactions were then transformed into competent bacterial cells and subject to antibiotic selection with ampicillin. Since the restriction endonucleases used produce blunt-end DNA fragments (i.e. no 5' or 3' overhangs) it was possible for insertion of NMD$^{V1.0}$ to occur in the reverse orientation. Furthermore, re-ligation of pUC57-C$^{V1.0}$ would also confer ampicillin resistance. Therefore, following antibiotic selection, successful bacterial transformants could harbour any of three possible ligation products; the desired pUC57-CN$^{V1.0}$ (forward orientation), pUC57-NC$^{V1.0}$ (reverse orientation) or pUC57-C$^{V1.0}$ (vector only) (Figure 4.15A).

To confirm inclusion and orientation of NMD$^{V1.0}$ within pUC57-C$^{V1.0}$, plasmid DNA isolated from several resistant bacterial transformants was subject to diagnostic restriction endonuclease digests (Figure 4.15B). Colony 11 displayed insertion of NMD$^{V1.0}$ in the desired forward orientation (Figure 4.15B) and was subject to Sanger sequencing (data not shown) to ensure sequence fidelity prior to use in downstream experiments.

A)



B)



	Expected band sizes (bp)	
	*Spe*I Digest	*Pac*I Digest
pUC57-C$^{V1.0}$/ Vector (V)	6178	N/A
pUC57-CN$^{V1.0}$ →→ Forward orientation	7359 + 4372	9509 +1302
pUC57-CN$^{V1.0}$ →← Reverse orientation	8048 + 2783	9509 +1302

Figure 4.15: Restriction endonuclease digest screening for bacterial colonies containing pUC57-CN$^{V1.0}$ plasmid DNA.

(A) A schematic representation of the ligase based cloning method used to engineer pUC57-CN$^{V1.0}$. **(B)** Following ligation, bacterial transformation and antibiotic selection, plasmid DNA was isolated from 12 colonies and digested with the restriction endonucleases *Spe*I or *Pac*I to check orientation and inclusion of NMD$^{V1.0}$ within pUC57-C$^{V1.0}$ respectively. Expected digest product sizes are summarised in the table. Colony 2 is an example of vector only (pUC57-C$^{V1.0}$), colony 10 indicates an example of NMD$^{V1.0}$ inserted into pUC57-C$^{V1.0}$ in the reverse orientation (red box) and colony 11 is an example of a successful ligation containing NMD$^{V1.0}$ inserted into pUC57-C$^{V1.0}$ in the forward orientation (blue box).

4.3.2.4.4 Testing protein expression from Control$^{V1.0}$ and NMD$^{V1.0}$ of pUC57-CN$^{V1.0}$ in response to an siRNA mediated reduction of Upf1 levels in NIH3T3 cells

To better test if the NMD-targeted 3'UTR of NMD$^{V1.0}$ was responsive to NMD activity in NIH3T3 cells, tests were conducted in the presence of an internal control (Control$^{V1.0}$). The expression plasmid pUC57-CN$^{V1.0}$ encoding both Control$^{V1.0}$ and NMD$^{V1.0}$ was co-transfected alongside either Control or *Upf1* siRNA into NIH3T3 cells. Under Control siRNA conditions, transcripts expressed from NMD$^{V1.0}$ should be degraded by NMD. Therefore, following an siRNA mediated reduction of Upf1 (i.e. a reduction in NMD activity) an increase in expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ is expected. Comparatively, transcripts expressed from Control$^{V1.0}$ should evade NMD, resulting in no change in expression of CFP^{NLS} following Upf1 reduction.

Protein from NIH3T3 cells co-transfected with pUC57-CN$^{V1.0}$ and Control or *Upf1* siRNA was visualised by fluorescence microscopy and quantified using Image J (Figure 4.16). This showed that following Upf1 reduction there was an ~11% increase in TetR:EGFP^{NLS} expressed from NMD$^{V1.0}$ when normalised to CFP^{NLS} expressed from Control$^{V1.0}$ in the same cell. This did not reach significance. Published work in HEK293T cells suggest that a more potent and significant increase in normalised TetR:EGFP^{NLS} i.e. ~170% was expected (Pereverzev et al. 2015).

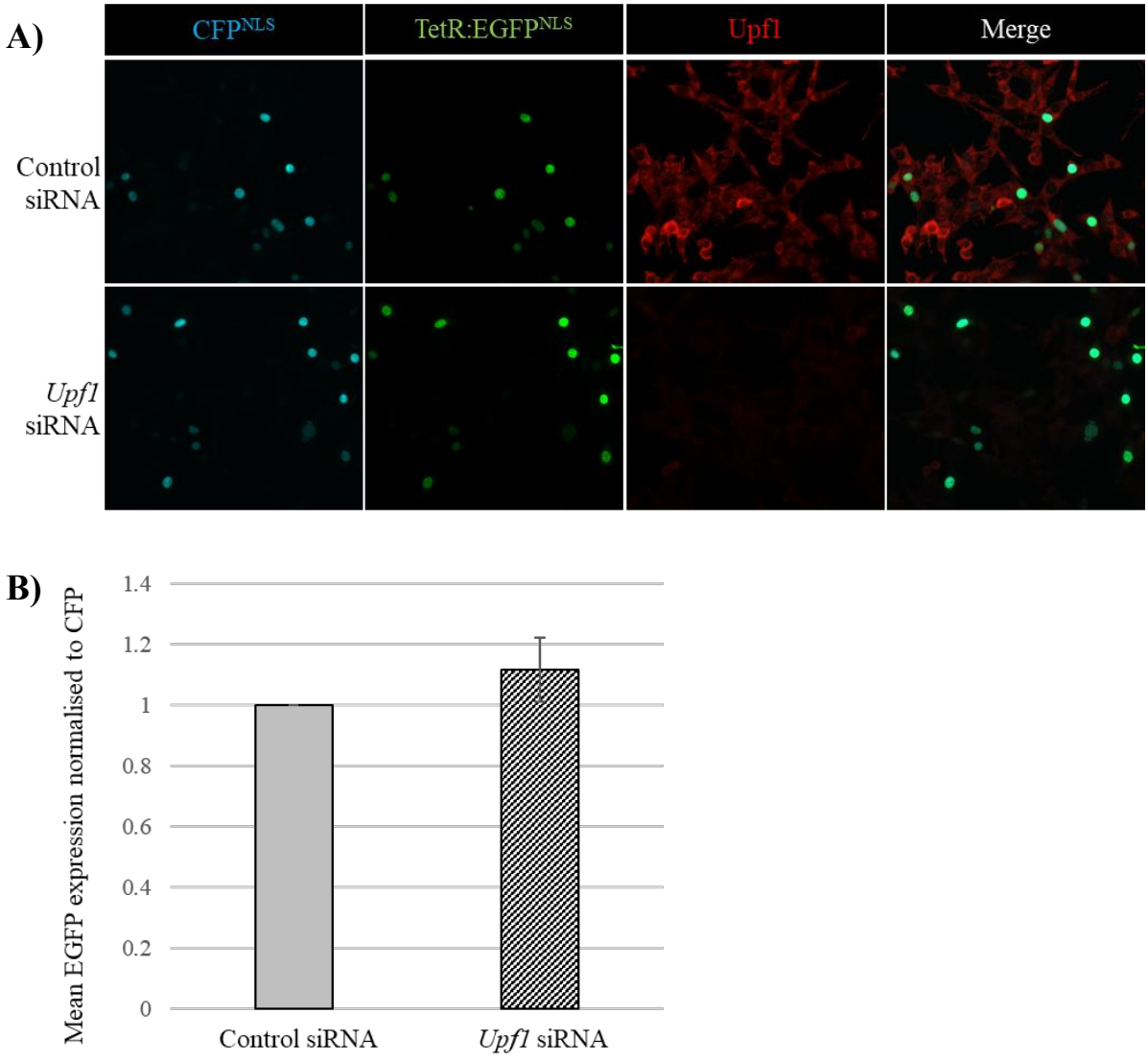


Figure 4.16: Fluoresence output from pUC57-CN$^{V1.0}$ is not responsive to a reduction in UPF1 protein levels following transient transfection in NIH3T3 cells.

NIH3T3 cells were co-transfected with an expression plasmid encoding both Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) and either Control or *Upf1* siRNA across three separate experiments. **(A)** Representative fluorescence microscopy images of these cells. The cells express CFP^{NLS} from Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from NMD$^{V1.0}$ (green). Upf1 was detected by immunofluorescence (red). **(B)** Fluorescence microscopy images were analysed using Image J. Nuclear localised EGFP fluorescence was normalised to CFP expression in the same nucleus. *P<0.05 by Student's two-tailed t-test.

4.3.2.5 Testing expression from Control$^{V1.0}$ and NMD$^{V1.0}$ of pUC57-CN$^{V1.0}$ in response to an siRNA mediated reduction of UPF1 levels in HEK293T cells

4.3.2.5.1 Using UPF1 siRNA to establish low NMD activity in HEK293T cells

The NMD-refractory and NMD-targeted 3'UTRs of Control$^{V1.0}$ and NMD$^{V1.0}$ are based on a previously published NMD reporter system design, where a striking and significant difference in expression was observed in compromised HEK293T cells (Pereverzev et al. 2015). Data from the same paper also suggested that compared to mouse cells (MEFs), human HEK293T and HeLa cells showed a higher inherent level of NMD activity. To investigate whether Control$^{V1.0}$ and NMD$^{V1.0}$ show a stronger response to NMD manipulation in cells with a higher inherent level of NMD activity than NIH3T3 cells, the expression plasmid pUC57-CN$^{V1.0}$ was subject to testing in HEK293T cells.

To replicate high and low NMD conditions in HEK293T cells siRNA targeting mRNA of the core NMD component *UPF1* was employed. Either Control or *UPF1* siRNA was transiently introduced into HEK293T cells, and levels of UPF1 protein were visualised by immunofluorescence and western blot. In both cases, introduction of *UPF1* siRNA resulted in a dramatic reduction in UPF1 protein expression (Figures 4.17A & 4.17B). To further investigate the impact of loss of UPF1 on NMD activity within these cells, the expression of three known NMD target mRNAs was assessed via qPCR using cDNA reverse transcribed from RNA of Control or *UPF1* siRNA treated HEK293T cells. This revealed that accompanying a decrease in *UPF1* expression was a significant increase in the expression of known NMD target genes *ATF4* and *GAS5* (Jolly et al. 2013; Mendell et al. 2004). This data confirms that a loss of *UPF1* in these cells diminishes NMD activity (Figure 4.17C).

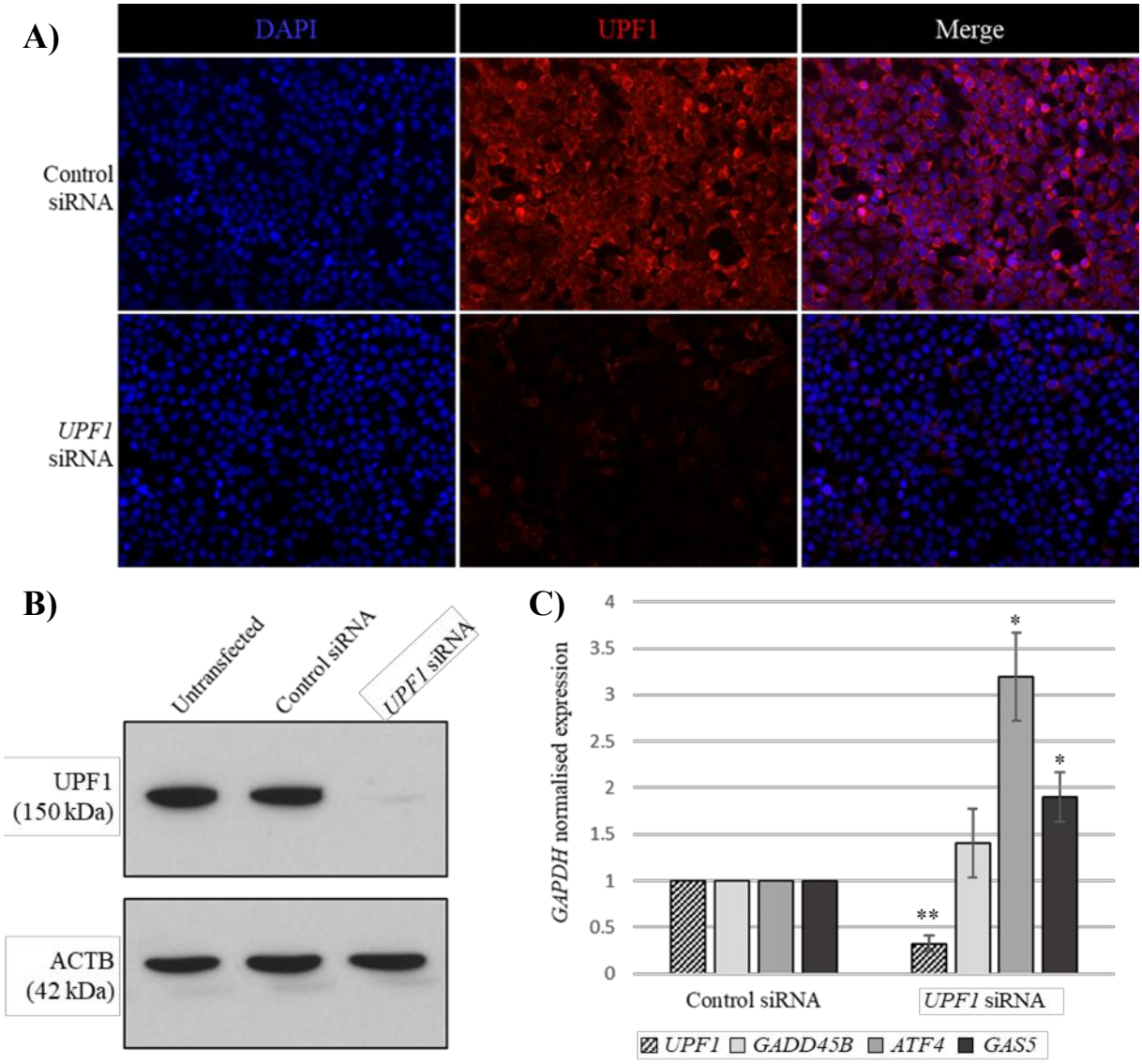


Figure 4.17: UPF1 siRNA reduces NMD activity in HEK293T cells.

HEK293T cells were transfected with Control or *UPF1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells and from untransfected HEK293T cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1 and the loading control ACTB. **(C)** cDNA reverse transcribed from the RNA of these cells was used as a template in qPCR to determine mean expression (±standard deviation) of *UPF1, GADD45B, ATF4,* and *GAS5*. The assay was performed in triplicate per sample and expression was measured using the relative standard curve method and normalised against *GAPDH* mRNA expression in the same sample. *P < 0.05 and ** P< 0.01 by Student's two tailed t-test.

4.3.2.5.2 Testing protein expression from Control$^{V1.0}$ and NMD$^{V1.0}$ of pUC57-CN$^{V1.0}$ in response to an siRNA mediated reduction of UPF1 levels in HEK293T cells

To test if the NMD-targeted 3'UTR of NMD$^{V1.0}$ was responsive to NMD activity in HEK293T cells, the expression plasmid encoding both Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) was co-transfected into HEK293T cells alongside either Control or *UPF1* siRNA. Under Control siRNA conditions, transcripts expressed from NMD$^{V1.0}$ should be degraded by NMD, thus following UPF1 reduction (i.e. a reduction in NMD activity) an increase in expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ is expected. Comparatively, transcripts expressed from Control$^{V1.0}$ should evade NMD, resulting in no change in expression of CFP^{NLS} following a reduction in UPF1.

Protein from HEK293T cells co-transfected with pUC57-CN$^{V1.0}$ and Control or *UPF1* siRNA was visualised through fluorescence microscopy and western blot (Figures 4.18B & 4.18C). Following a reduction in UPF1, these cells showed an expected ~91% upregulation of TetR:EGFP^{NLS} expressed from NMD$^{V1.0}$, however CFP^{NLS} expressed from Control$^{V1.0}$ which should remain unresponsive to changes in NMD also showed an unexpected upregulation of ~47% (Figure 4.18C). It was noted that the magnitude of these upregulations was variable across the three replicates and thus did not reach significance (Figure 4.18B).

Published work in HEK293T cells suggested that following a decrease in NMD activity, the level of reporter fluorescence normalised to control fluorescence should increase by ~170% when compared to control conditions (Pereverzev et al. 2015). Since similar results were not observed for protein expression from NMD$^{V1.0}$ and Control$^{V1.0}$ in the experiments so far, testing of pUC57-CN$^{V1.0}$ was considering unsuccessful.

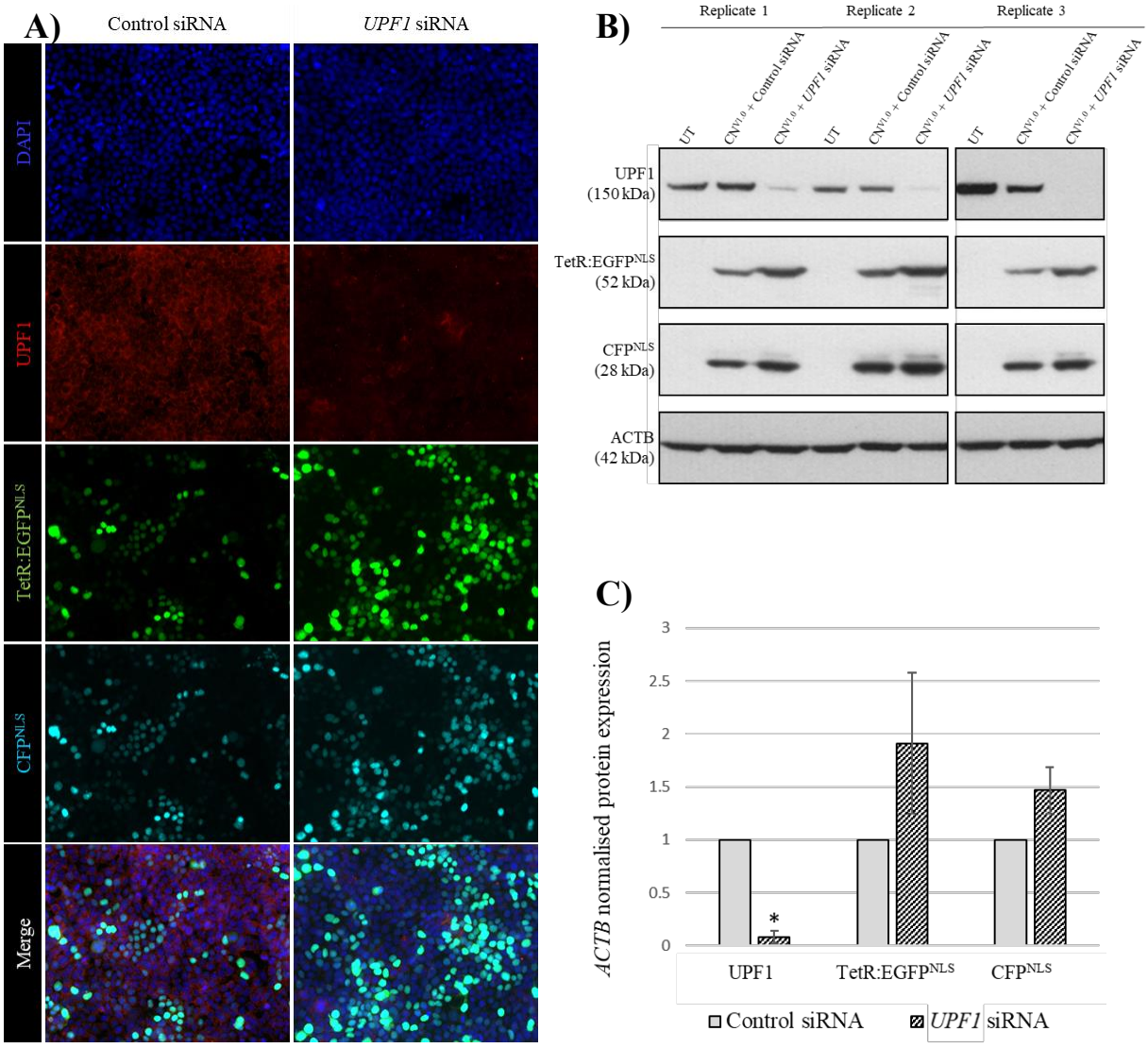


Figure 4.18: Expression of Control$^{V1.0}$ and NMD$^{V1.0}$ following transient transfection of pUC57-CN$^{V1.0}$ into HEK293T cells is elevated in response to UPF1 depletion.

HEK293T cells were co-transfected with the expression plasmid encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) alongside either Control or *UPF1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express CFP^{NLS} from Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from NMD$^{V1.0}$ (green). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells and from untransfected HEK293T cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, GFP and the and the loading control ACTB (stripped and re-probed). **(C)** Densitometric analysis of this western blot showed that following UPF1 reduction, expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ increased by ~91%, while expression of CFP^{NLS} from Control$^{V1.0}$ increased by ~47%.*P < 0.05 by Student's two tailed t-test.

4.3.2.5.3 Sequence investigations reveal that the design of the NMD-refractive and NMD-targeted 3'UTR of Control$^{V1.0}$ and NMD$^{V1.0}$ is based on an alternative HBB transcript which disrupts expected splicing patterns

The NMD-refractive and NMD-targeted 3'UTR sequences should harbour sequences from the predominant *HBB* isoform, *HBB-201* (ENST00000335295.4) (Pereverzev et al. 2015). After functional validation of pUC57-CN$^{V1.0}$ was unsuccessful in NIH3T3 and HEK293T cells, the original sequence design was re-analysed. It was found that the 3'UTR of Control$^{V1.0}$ and NMD$^{V1.0}$ was designed using sequences from the alternative *HBB-204* (ENS0000485743.1) isoform. The main difference between these isoforms which impacts 3'UTR design is that the second exon of *HBB-204* is composed of both exonic and intronic sequence from *HBB-201* (i.e. 223 bp of *HBB-201* exon 2 followed by the first 314 bps of *HBB-201* intron 2) (Figure 4.19).

Compared to the original 3'UTR design of NMD$^{V1.0}$, the use of sequences from *HBB-204* will shorten the length of exon 2 from 537 bp to 223 bp and increase the length of intron 2 by 314 bps (Figure 4.20). Following splicing and EJC deposition, the exon-exon junction within NMD$^{V1.0}$ will remain greater than 55 nts downstream from the TetR:EGFP^{NLS} stop codon and is still predicted be recognised by the NMD machinery. Therefore, regardless of the use of *HBB-204* sequence, the 3'UTR of NMD is predicted to remain 'NMD-targeted' (Figure 4.20).

In the case of Control$^{V1.0}$, the use of sequences from *HBB-204* in the 3'UTR will have a substantial impact on the design of this cassette, however the 3'UTR is still predicted to remain 'NMD-refractive'. The NMD-refractive 3'UTR originally included only the last 35 bp of *HBB-201* exon 2, the use of the last 35 bp of *HBB-204* exon 2 means that 35 bp of *HBB-201* intronic, rather than exonic sequence was included (Figure 4.20). This will result in loss of the desired exon-intron-exon structure and introduction of an intron-exon structure within the 3'UTR. This

change is predicted to inhibit splicing and exon junction deposition thus preventing NMD recognition of transcripts expressed from Control$^{V1.0}$.

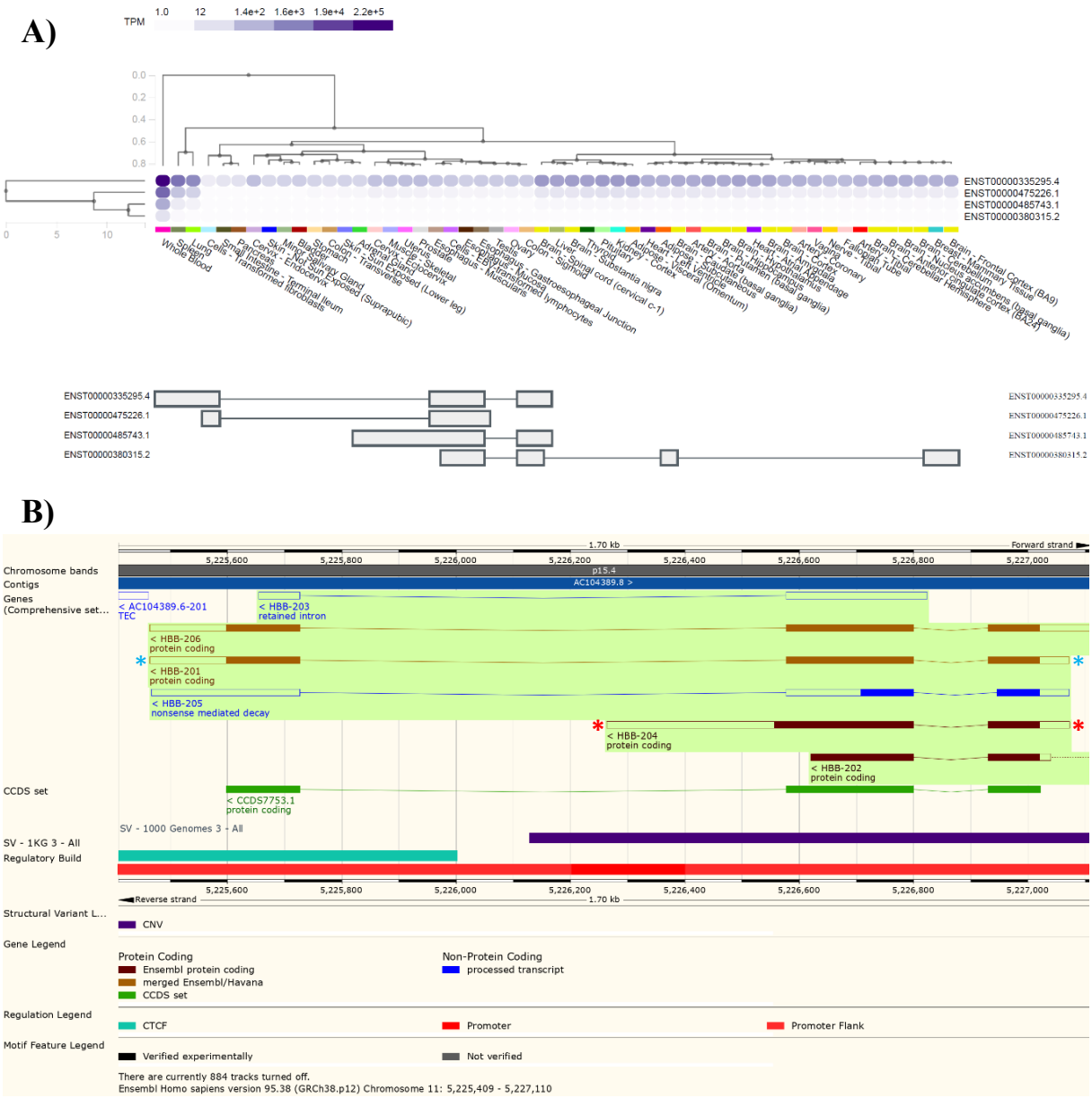


Figure 4.19: HBB-201 is the dominant transcript expressed from the human HBB gene.

(A) Gtex screenshot identifying ENST00000335295.4 (*HBB-201*) as the dominant *HBB* transcript in all analysed tissues. **(B)** Ensembl screenshot depicting the various transcripts predicted to be expressed from *HBB* (reverse orientation) and their exon/intron composition. The dominant *HBB-201* isoform is indicated by blue asterisks and the alternative *HBB-204* isoform (ENS0000485743.1) is indicated by red asterisks. The second exon of *HBB-204* includes both exon and intron sequence of the dominant *HBB-201* transcript.

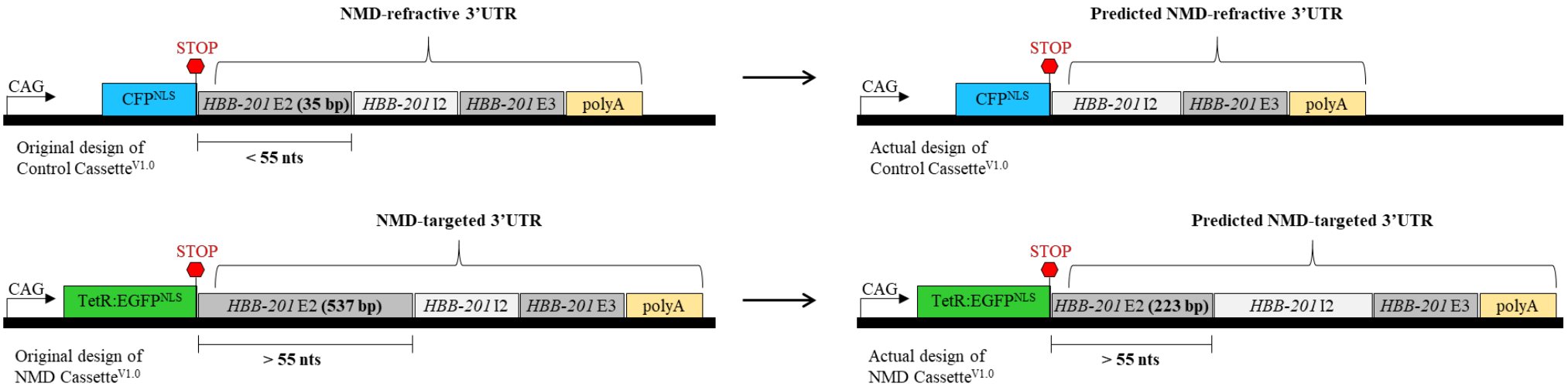


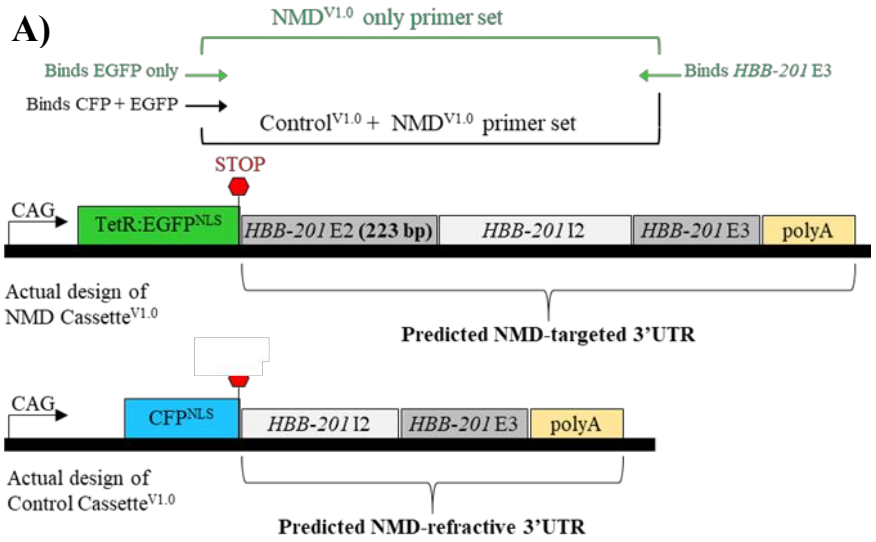
Figure 4.20: Schematic representation of the updated design of Control$^{V1.0}$ and NMD$^{V1.0}$ due to the use of sequences from HBB-204 instead of HBB-201.

The NMD-refractive and NMD-targeted 3'UTR of Control$^{V1.0}$ and NMD$^{V1.0}$ respectively, were both designed to contain an exon 2-intron 2-exon 3 sequence from the dominant *HBB-201* isoform (left of the arrows). Upon re-analysis however, it was found that sequences from the alternative *HBB-204* isoform were used instead (right of the arrows). The general 55 nt rule of NMD states that a PTC must be greater than 55 nts upstream of an exon-exon junction to be recognised and targeted by NMD. Originally, splicing would place the stop codon within NMD$^{V1.0}$ in the context of a PTC, while the stop codon within Control$^{V1.0}$ would be unrecognised by the NMD machinery. Despite the use of sequences from *HBB-204*, the 3'UTR of NMD$^{V1.0}$ is predicted to remain NMD-targeted, as following splicing the stop codon will remain greater than 55 nts upstream of an exon-exon junction and thus trigger NMD. Following the use of sequences from *HBB-204*, the 3'UTR of Control$^{V1.0}$ is also predicted to remain NMD-refractive, however, this is because the change in design will disrupt the exon-intron-exon structure of the 3'UTR, which is predicted to prevent splicing and thus the deposition of any EJCs. Without the presence of an EJC NMD cannot proceed.

To investigate the effect of these undesired changes to the original design on 3'UTR splicing of Control$^{V1.0}$ and NMD$^{V1.0}$, cDNA reverse transcribed from the RNA of HEK293T transfected with an expression plasmid for either Control$^{V1.0}$ (pUC57-C$^{V1.0}$) or NMD$^{V1.0}$ (pUC57-N$^{V1.0}$) was first generated. This cDNA was then subject to PCR using primers which flank the 3'UTR intronic sequence of both cassettes (splicing check) (Figure 4.21A). If the NMD-refractive or NMD-targeted 3'UTR sequence is efficiently spliced this PCR will amplify only a single product.

PCR amplified products were visualised via agarose gel electrophoresis and it was observed that despite the use of *HBB-204* sequences within the 3'UTR of NMD$^{V1.0}$, transcripts expressed from this cassette were observed to be predominantly spliced correctly (Figure 4.21B). Comparatively, the use of sequences from *HBB-204* within the 3'UTR of Control$^{V1.0}$ resulted in 3'UTR of this cassette remaining largely unspliced, with a small amount of spliced transcript detected (Figure 4.21B).

Although not as per the original design, both the spliced and unspliced 3'UTR of transcripts expressed from Control$^{V1.0}$ are predicted to remain 'NMD-refractive'. When transcripts expressed from Control$^{V1.0}$ are spliced, the exon-exon junction introduced will be less than 55 nts downstream from the CFP^{NLS} stop codon, as such, this stop codon is predicted to evade NMD based on the general 55 nt rule. When transcripts expressed from Control$^{V1.0}$ remain unspliced, they will not have had the opportunity for EJC deposition. Without the presence of an EJC, NMD cannot proceed under the 'downstream EJC' canonical mechanism. Furthermore, due to the intron-exon structure introduced, the 3'UTR of unspliced transcripts expressed from Control$^{V1.0}$ will be lengthened to 814 nts. This remains shorter than the minimum 1 kb required to act as a 'long 3'UTR-type' NMD triggering feature (Ge et al. 2016). However, this transcript will, differ to the original design by inclusion of 314 nts of intronic sequence of unknown relevance (Figure 4.20).



	N^V1.0		C^V1.0	
	Spliced*	Unspliced*	Spliced*	Unspliced*
NMD^V1.0 only primer set	557 bp	1407 bp	N/A	N/A
Control^V1.0 and NMD^V1.0 primer set	418 bp	1297 bp	229 bp	800 bp

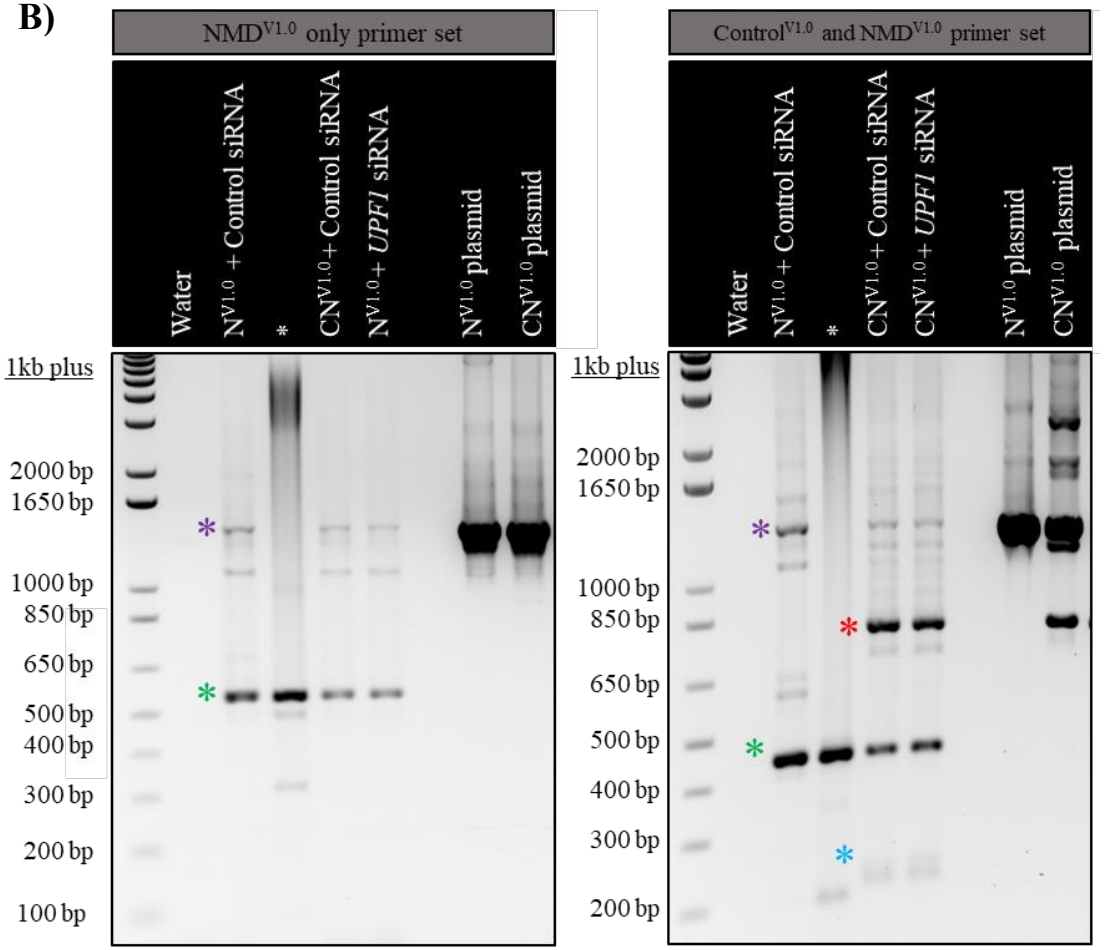


Figure 4.21: The use of sequences from HBB-204 in the design of Control$^{V1.0}$ and NMD$^{V1.0}$ prevents splicing of the NMD-refractive 3'UTR of Control$^{V1.0}$ in HEK293T cells.

(A) A schematic representation of two primer sets designed to assess splicing of the 3'UTR of transcripts expressed from Control$^{V1.0}$ and NMD$^{V1.0}$. The table outlines the expected PCR product sizes (these are colour coded in reference to the asterisks on the agarose gel image in (B)). (**B)** cDNA reverse transcribed from the RNA of HEK293T cells transfected with expression plasmids encoding NMD$^{V1.0}$ (pUC57-N$^{V1.0}$, N$^{V1.0}$) or both Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$, CN$^{V1.0}$) was subject to PCR using the described primers. PCR amplified products were visualised by agarose gel electrophoresis, the sample in lane 3 (*) remained 'stuck' in the well and was excluded from analysis. Banding patterns show that most transcripts expressed from NMD$^{V1.0}$ are correctly spliced, while a large proportion of transcripts expressed from Control$^{V1.0}$ remain unspliced.

4.3.2.6 Testing the performance of the TetO-CAG promoter of TetR Responder$^{V1.0}$ expressed from pUC57-TetO-T$^{V1.0}$ in HEK293T cells

The CAG promoter of TetR Responder$^{V1.0}$ was re-engineered to introduce two TetO sequences between the TATA box and initiator sequences, these are cognate binding sites for TetR. Binding of TetR to these TetO sequences was envisaged to result in transcriptional repression of the engineered TetO-CAG promoter. The rationale was that if cellular NMD activity was high, decreased expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ would promote tdTomato^{NLS} expression downstream of the TetO-CAG promoter in TetR Responder$^{V1.0}$. In this way, a reciprocal fluorescent signal which is expressed in a proportional manner to NMD efficiency would be introduced. This would facilitate not only low NMD activity, but also high NMD activity to be visualised and quantified via fluorescence.

To first test the sensitivity of this novel TetO-CAG promoter to TetR, HEK293T cells were co-transfected with varying ratios (1:1–1:12.5) of an expression plasmid encoding TetR Responder$^{V1.0}$ (pUC57-TetO-T$^{V1.0}$) and either a TetR expression plasmid (pTetR) or a control plasmid (pcDNA3). If the TetO-CAG promoter is functioning as expected, tdTomato^{NLS} expression from TetR Responder$^{V1.0}$ should diminish upon addition of TetR. Protein from co-transfected cells was visualised by immunofluorescent analysis and quantified using Image J.

It was observed that when TetR was present in a 12.5 times higher proportion than TetR Responder$^{V1.0}$, expression of tdTomato^{NLS} from TetR Responder$^{V1.0}$ was reduced by ~55%. In comparison when TetR was present in a 1:1 ratio with TetR Responder$^{V1.0}$, expression of tdTomato^{NLS} from TetR Responder$^{V1.0}$ was only reduced by ~30% (Figure 4.22). These results suggest that the TetO-CAG promoter is responsive to TetR and that this can be visualised and quantified by fluorescence. However, needing an assumed 12.5-fold (based on amount of plasmid added) difference in TetR protein to repress only ~55% of expression from TetR

Reponder$^{V1.0}$ suggests that the system TetO-CAG system may be 'leaky'. This level of sensitivity may not be optimal for use within Transgene$^{V1.0}$. To determine if this is true, similar experiments would need to be carried out using a single expression plasmid for both TetR Responder$^{V1.0}$ and NMD$^{V1.0}$. This line of experiments was put on hold until a functional NMD responsive cassette could first be established.

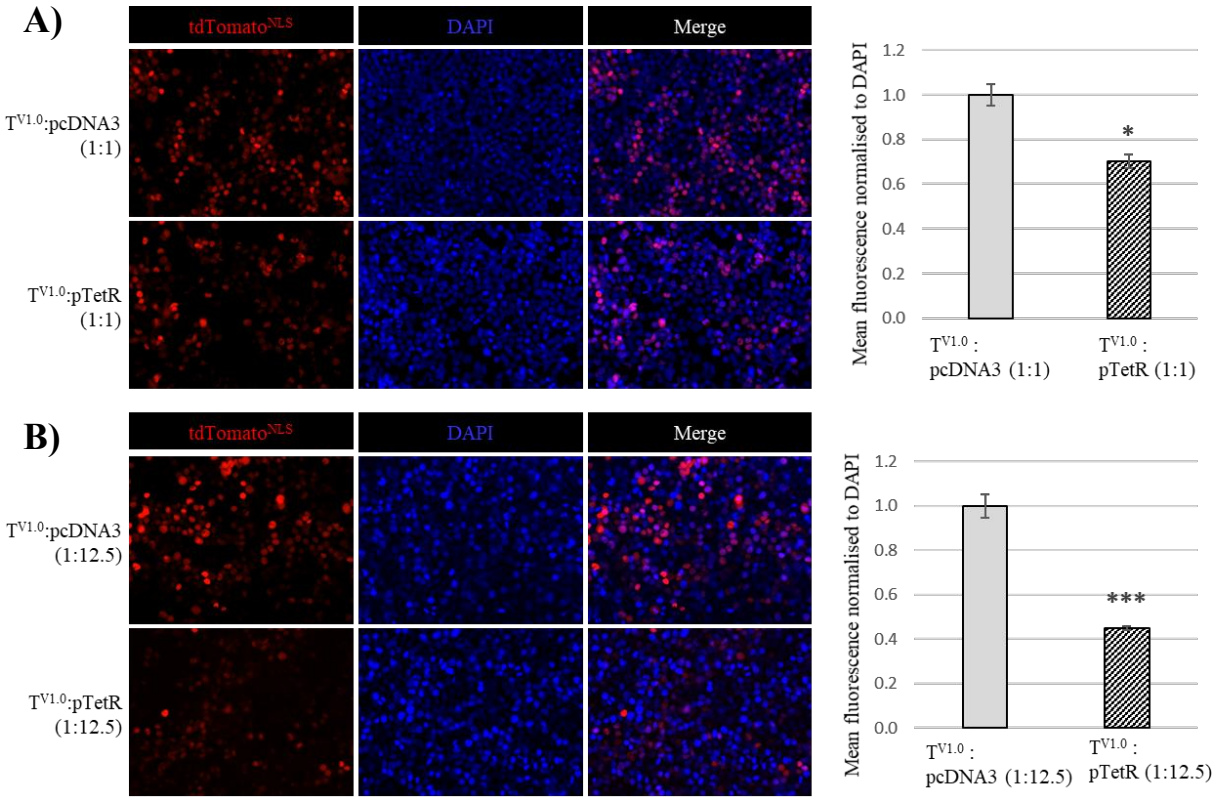


Figure 4.22: High levels of TetR are needed to repress transcription of tdTomato^{NLS} from TetR Respnder$^{V1.0}$ in HEK293T cells.

HEK293T cells were co-transfected with a, **(A)** 1:1 or **(B)** 1:12.5 ratio of an expression plasmid encoding TetR Responder$^{V1.0}$ (pUC57-TetO-T$^{V1.0}$, T$^{V1.0}$) and either an expression plasmid encoding Tetracycline repressor (pTetR) or the control plasmid (pcDNA3). The left panels show fluorescence microscopy images of these cells. The cells express tdTomato^{NLS} from TetR Responder$^{V1.0}$ (red) and cell nuclei were counterstained with DAPI (blue). The right panels show mean fluorescence of tdTomato^{NLS} expression normalised against DAPI staining calculated using Image J. *P<0.05, **P<0.001 by Student's two-tailed t test.

4.3.3 Functional testing of Selection$^{V1.0}$, Control $^{V1.0}$ and NMD$^{V1.0}$ following stable genomic integration of Transgene$^{V1.0}$ into FLP-in mESCs

4.3.3.1 Functional validation of Selection$^{V1.0}$ through establishing NMD Reporter$^{V1.0}$ mESCs

Selection$^{V1.0}$ remained largely unchanged from published work (Bersten et al. 2015). Minor non-functional sequence modifications included the addition of homology arms and insertion into the pUC57 backbone. One way to confirm that this cassette can still facilitate RMCE into FLP-in mESCs, was to engineer Selection$^{V1.0}$, Control$^{V1.0}$ and NMD$^{V1.0}$ into a single transgene and then attempt to integrate this transgene into FLP-in mESCs via Selection$^{V1.0}$ mediated RMCE.

If the transgene is successfully incorporated the function of Selection$^{V1.0}$ and the RMCE method can be validated. Furthermore, the resulting mESC line can be used to investigate the function of Control$^{V1.0}$ and NMD$^{V1.0}$ in a stable environment. Previous experiments have been transient transfection based, these types of experiments can drive spurious results due to excessively high expression, and saturation or sequestration of gene expression machinery. Genomic integration of Control$^{V1.0}$ and NMD$^{V1.0}$ as a single copy per cell may alleviate such effects.

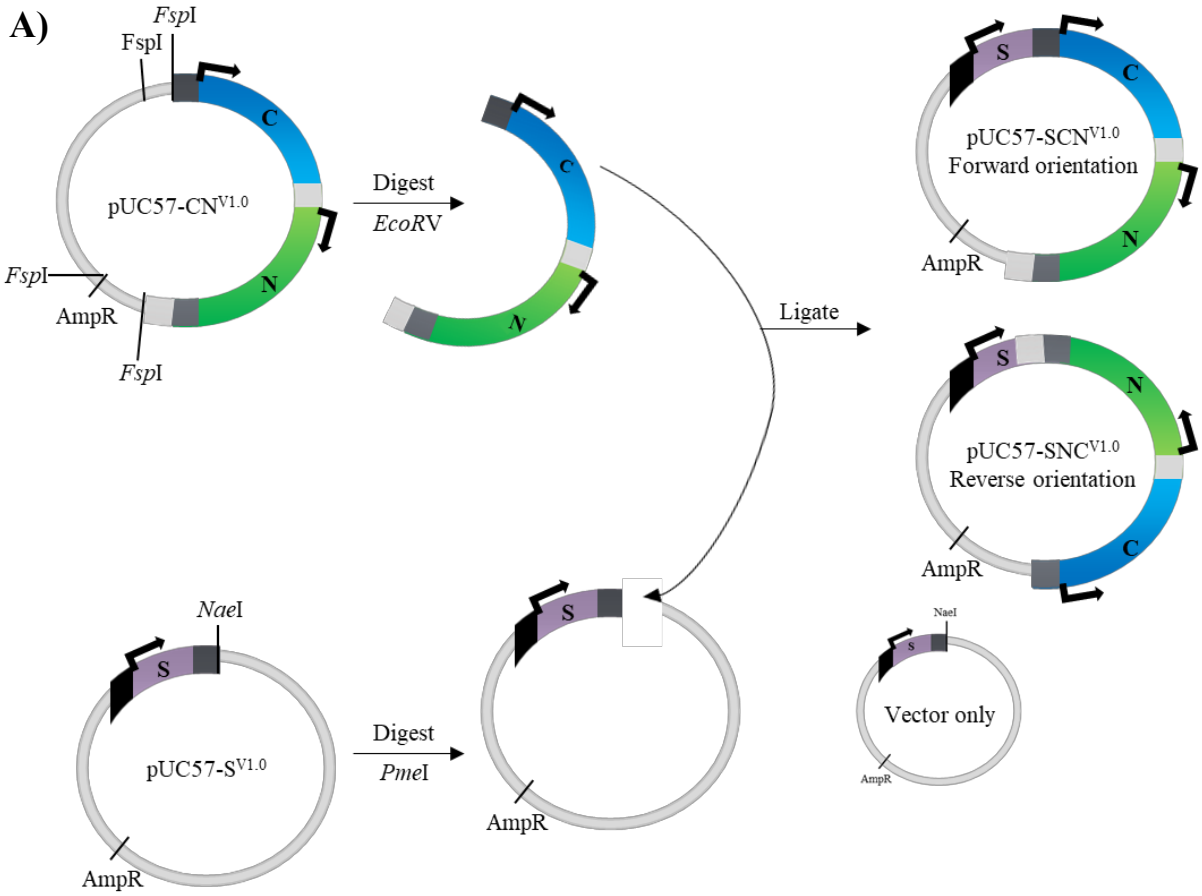
4.3.3.1.1 Engineering pUC57-SNC$^{V1.0}$—an expression plasmid encoding Selection$^{V1.0}$, Control$^{V1.0}$ and NMD$^{V1.0}$

To engineer Control$^{V1.0}$ and NMD$^{V1.0}$ into the expression plasmid encoding Selection$^{V1.0}$ (pUC57-S$^{V1.0}$) standard recombinant DNA cloning methods were applied. Using flanking *Fsp*I restriction sites, Control$^{V1.0}$ and NMD$^{V1.0}$ were excised as a single DNA fragment from pUC57-CN$^{V1.0}$ and then ligated into the *Nae*I site of pUC57-S$^{V1.0}$ (Figure 4.23A). Ligation reactions were then transformed into competent bacterial cells and subject to antibiotic selection with ampicillin. Since the restriction endonucleases used produce blunt-end DNA fragments (i.e. no 5' or 3' overhangs) it was possible for insertion of Control$^{V1.0}$ and NMD$^{V1.0}$ to occur in the

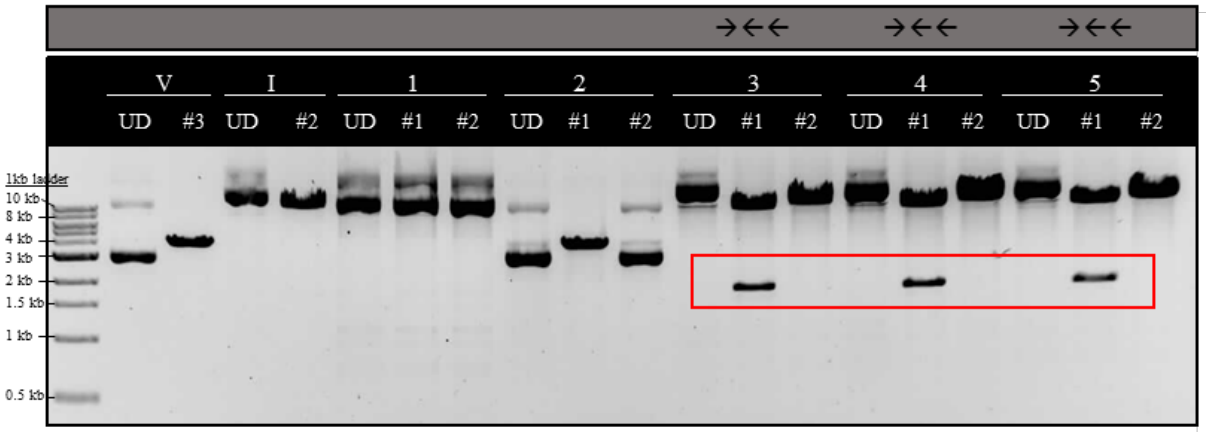
reverse orientation. Furthermore, re-ligation of pUC57-$S^{V1.0}$ would also confer ampicillin resistance. As such, following antibiotic selection, successful bacterial transformants could harbour any of three possible ligation products; the desired pUC57-$SCN^{V1.0}$ (forward orientation), pUC57-$SNC^{V1.0}$ (reverse orientation), or pUC57-$S^{V1.0}$ (vector only) (Figure 4.23A).

To confirm inclusion and orientation of $Control^{V1.0}$ and $NMD^{V1.0}$ within pUC57-$S^{V1.0}$, plasmid DNA isolated from several successful bacterial transformants was subject to diagnostic restriction endonuclease digests. It was observed that all successful colonies displayed only reverse orientation of $Control^{V1.0}$ and $NMD^{V1.0}$ (i.e. pUC57-$SNC^{V1.0}$) (Figure 4.23B). In this case, the reverse orientation of the insert is not predicated to affect the function of any cassettes. Plasmid DNA from one such bacterial colony was subject to Sanger sequencing (data not shown) to ensure sequence fidelity. This plasmid was used in subsequent RMCE experiments.

A)



B)



		Expected digest product sizes (bp)			
	Restriction Endonucleases	Vector (V)	Insert (I)	V1. SCA →→→	V1. SAC →←←
Digest #1	*BamH*I	-	10831	12223	12223
Digest #2	*Cla*I + *BamH*I	4069	-	7418 + 4805	10311 + 1912
Digest #3	*Eco*RI	4069			

Figure 4.23: Restriction endonuclease digest screening for bacterial colonies containing pUC57-SNC$^{V1.0}$ plasmid DNA.

(**A)** A schematic representation of the cloning method used to engineer pUC57-SNC$^{V1.0}$. **(B)** Following ligation, bacterial transformation and antibiotic selection, plasmid DNA was isolated from 14 colonies and digested with the restriction endonucleases, *Cla*I and *BamH*I to check both inclusion and orientation of Control $^{V1.0}$ and NMD$^{V1.0}$ within pUC57-S$^{V1.0}$. The digested product was visualised by agarose gel electrophoresis. Expected digest product sizes are summarised in the table. All successful colonies analysed were observed to have Control$^{V1.0}$ and NMD$^{V1.0}$ inserted into pUC57-S$^{V1.0}$ in the reverse orientation, this is represented here by colonies 3-5 (red box, represents unique 1913 bp diagnostic band).

4.3.3.1.2 Using RMCE to establish NMD Reporter$^{V1.0}$ mESCs which express a single copy of the Control$^{V1.0}$ and NMD$^{V1.0}$ from the Col1a1 locus

To establish a stable NMD reporter system, the expression plasmid encoding Selection$^{V1.0}$, Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-SNC$^{V1.0}$) has the potential to be integrated into the genome of FLP-in mESCs using the previously established RMCE system which involves Selction$^{V1.0}$ (Section 4.2.2.2) (Bersten et al. 2015). This method first involved co-electroporation of pUC57-SNC$^{V1.0}$, alongside an expression plasmid encoding FLP recombinase (pPGK-FLPo-bpA) into FLP-in mESCs. 48 hours post electroporation, electroporated cells were subject to hygromycin selection (Figure 4.24A). Following selection, hygromycin resistant clones which expressed CFP^{NLS} from genomic Control$^{V1.0}$ were isolated based on visible CFP expression and expanded to establish NMD Reporter$^{V1.0}$ mESC lines (Figure 4.24B). In these cells, TetR:EGFP^{NLS} expressed from genomic NMD$^{V1.0}$ was also visible by fluorescence microscopy (Figure 4.24B). Correct integration into the *Col1a1* locus was confirmed via PCR of gDNA (Chapter 5).

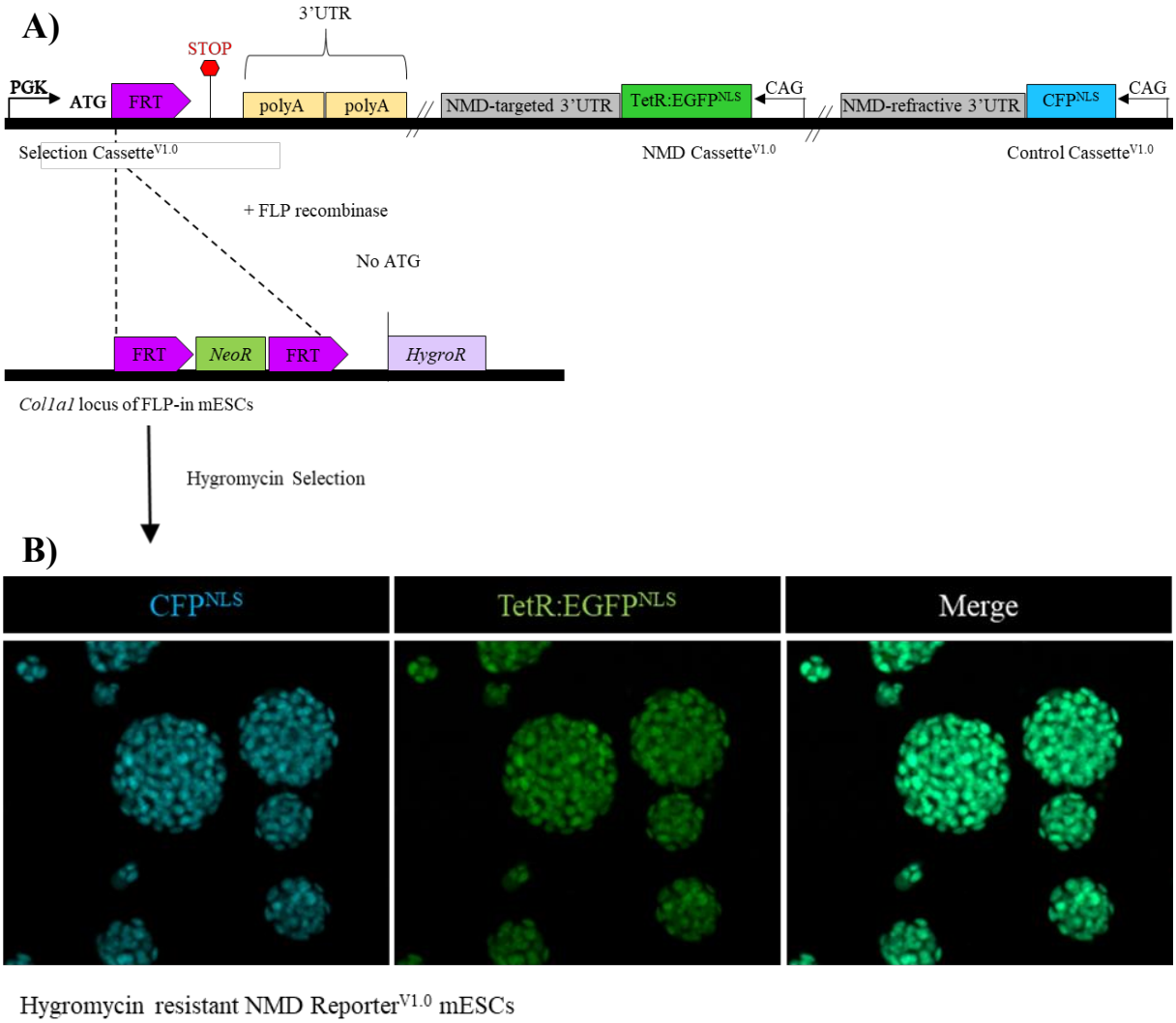


Figure 4.24: Generating an NMD Reporter$^{V1.0}$ mESC line via Selection$^{V1.0}$ mediated RMCE.

(A) A schematic representation of *Col1a1* RMCE targeting strategy. FLP-in mESCs (mESCs pre-engineered with *Col1a1* FRT sequences and a promoter-less hygromycin resistance coding sequence (*HygoR*), which also lacks a start codon) are co-electroporated with the expression plasmid encoding Selection$^{V1.0}$, Control $^{V1.0}$ and NMD$^{V1.0}$ (pUC57-SNC$^{V1.0}$) alongside an expression plasmid encoding FLP recombinase (pPGK-FLPo-bpA). Successful RMCE results in gain of hygromycin resistance and FRT-directed, site specific genomic integration of Control$^{V1.0}$ and NMD$^{V1.0}$ as a single copy into cells. Following hygromycin selection and expansion, these cells are referred to as NMD Reporter$^{V1.0}$ mESCs. **(B)** Representative fluorescence microscopy images of NMD Reporter$^{V1.0}$ mESCs following hygromycin selection. These cells express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ (green).

4.3.3.1.3 Testing protein expression from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs in response to an siRNA mediated reduction of Upf1 levels

To test expression of genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCS to changes in NMD activity, siRNA knockdown of *Upf1* was employed. First, the efficiency of *Upf1* siRNA in mESCs was tested. Either Control or *Upf1* siRNA was transiently introduced into FLP-in mESCs and levels of Upf1 protein expression were visualised via immunofluorescence and western blot. Both methods detected a reduction in Upf1 protein expression following introduction of *Upf1* siRNA.

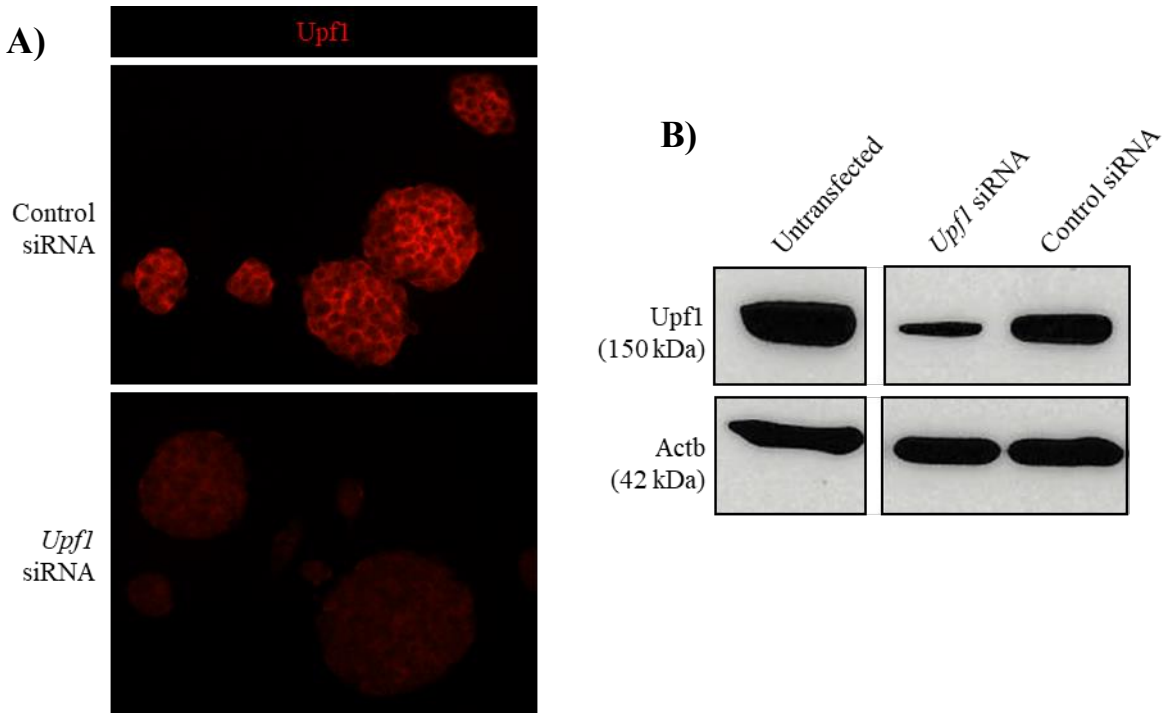


Figure 4.25: Upf1 siRNA reduces Upf1 protein levels in FLP-in mESCs.

FLP-in mESCs were transfected with Control or *Upf1* siRNA. **(A)** Representative fluorescence microscopy images of these cells showing immunofluorescent detection of Upf1 (red). **(B)** Total protein isolated from these cells, and from untransfected FLP-in mESCs was separated by SDS-PAGE and analysed by western blot using antibodies to detect Upf1 and the loading control Actb.

Next, Control or *Upf1* siRNA were transiently introduced into NMD Reporter$^{V1.0}$ mESCs. If these cells are functional, expression of TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ should increase following a reduction in Upf1, whilst CFP^{NLS} expression from genomic Control$^{V1.0}$ should remain unchanged.

Protein from siRNA treated NMD Reporter$^{V1.0}$ mESCs was visualised by fluorescence microscopy and western blot (Figure 4.26). It was observed that following a reduction in Upf1, expression of TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ was increased by ~67%, whilst expression of CFP^{NLS} from genomic Control$^{V1.0}$ was seen to decrease by ~10% (Figure 4.26C). Interestingly, these results were more comparable to those observed when NMD$^{V1.0}$ and Control$^{V1.0}$ were transiently tested in NIH3T3 cells rather than HEK293T cells.

Ultimately, even though these results are in line with the expected results, the magnitude of response was underwhelming. Any changes in expression could not be easily observed by fluorescence microscopy and did not compare to the magnitude of published results using similar NMD targeting features (Pereverzev et al. 2015). These results paired with errors in the underlying design of Control$^{V1.0}$ and NMD$^{V1.0}$ made the design of Transgene$^{V1.0}$ unsuitable for use as a dual fluorescent NMD reporter system with single cell resolution.

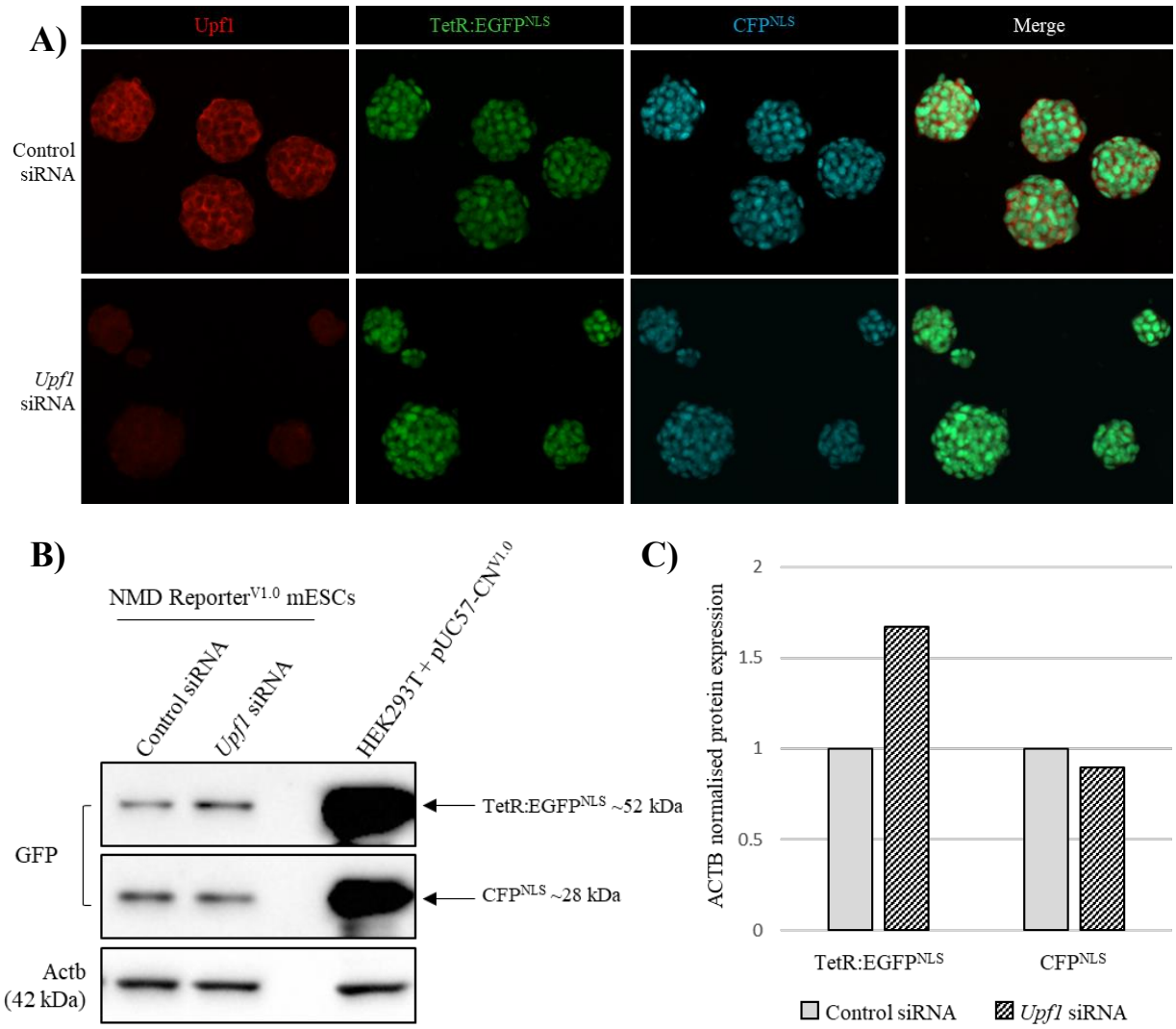


Figure 4.26: Protein expression from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs is slightly elevated following a reduction in Upf1 protein.

NMD Reporter$^{V1.0}$ mESCs were transfected with Control or *Upf1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ (green). Upf1 was detected by immunofluorescence (red). **(B)** Total protein isolated from these cells and HEK293T cells transiently transfected with an expression plasmid for Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) for comparison was separated by SDS-PAGE and analysed by western blot using antibodies to detect GFP and the loading control Actb. **(C)** Densitometric analysis of this western blot revealed that following Upf1, expression of TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ was increased by ~67% and expression of CFP^{NLS} from genomic Control$^{V1.0}$ was reduced by ~10%.

4.4 Discussion

The aim of this chapter was to design, engineer, and functionally validate an NMD reporter transgene (Transgene$^{V1.0}$). Transgene$^{V1.0}$ was being developed to address the deficiencies of current NMD reporter systems to report on endogenous NMD activity, its cell and tissue specificity, and its role in development and disease.

The design of Transgene$^{V1.0}$ required the incorporation of several basic features and their functional validation. Fluorescence was chosen to visualise and quantify NMD activity levels within cells. This decision was based on the success of previous fluorescent NMD reporter systems. (Alexandrov, Shu & Steitz 2017; Pereverzev et al. 2015) and the fact that fluorescence can be quantified in fixed or live cells, allowing it to be tracked across development via a variety of analysis methods including FACS and fluorescence microscopy. Through these techniques, when, where and in which cells NMD is most active can be identified. Moreover, by choosing to restrict fluorescence to the nucleus via the addition of NLSs to sequences encoding fluorescent proteins, single cell resolution can be achieved. In this way, when there are many cells and cell types present, for example in complex tissue architectures such as neural cell cultures or brain tissue, the fluorescent transgene read-out can be easily attributed to a specific cell or cell type.

As discussed, the overall goal of the envisaged NMD reporter system was to be integrated into the genome of pluripotent stem cells and eventually a mouse zygote which can give rise to an NMD reporter mouse model. To facilitate this, Selection$^{V1.0}$ was designed as tool which can allow an entire transgene to be introduced as a single, stable copy into the genetically modified *Col1a1* locus of FLP-in mESCs or FLP-in mouse zygotes using the previously described RMCE system (Bersten et al. 2015). Following integration, NMD activity could be potentially studied *in vitro* by subjecting pluripotent NMD reporter mESCs to a variety of differentiation protocols, or *in vivo*, within the cells of the developing and adult NMD reporter mouse. For such studies

to be informative the NMD reporter system needs to express and function efficiently in all cell types. To achieve this, the synthetic CAG promoter was chosen to drive expression of the cassettes within Transgene$^{V1.0}$. The CAG promoter is known to drive strong constitutive expression in all mammalian cell types and is composed of the CMV early enhancer element, the promoter, the first exon and the first intron of the chicken beta-actin gene and the splice acceptor of the rabbit beta-globin gene (Niwa et al. 1991; Sakaguchi et al. 2014).

The next step in the design process was to incorporate features which can report on NMD activity within the cell. At the time of design, the most recent and impressive NMD reporter to be published was a transient dual fluorescent system (Pereverzev et al. 2015). This system consists of two separate fluorescent protein expressing constructs. The first construct expresses transcripts targeted for splicing-dependent NMD whilst the second, expresses NMD-refractive control transcripts. The difference between these constructs was in the design of their 3'UTRs. The 3'UTR of the NMD-targeted construct, contains exon 2, intron 2 and the first 233 bp of exon 3 from *HBB*. The removal of intron 2 and deposition of an EJC via splicing results in an exon-exon junction residing greater than 55 nts downstream of the stop codon. This will position the stop codon in the context of a PTC and thus render transcripts expressed from the NMD targeted construct subject to NMD. The 3'UTR of the control construct of this system was identical to the NMD-targeted construct apart from containing a shortened 35 bp fragment of *HBB* exon 2, in this case removal of intron 2 due to splicing will introduce an exon-exon junction less than 55 nts from the stop codon rendering the expressed transcripts unrecognisable to the NMD machinery (Pereverzev et al. 2015).

The NMD-targeted 3'UTR and the NMD-refractive 3'UTR developed by Pereverzev et al, 2015 was adopted into the design of NMD$^{V1.0}$ and Control$^{V1.0}$ respectively. NMD$^{V1.0}$ encodes the TetR:EGFP^{NLS} fusion protein while Control$^{V1.0}$ encodes CFP^{NLS}. In the design of Transgene$^{V1.0}$, these cassettes were placed in cis to facilitate an internally controlled system, where expression

of CFP^{NLS} from Control$^{V1.0}$ remains unchanged in situations of high or low NMD activity. Comparatively, expression of NMD$^{V1.0}$ is predicted to increase as NMD activity decreases and vice versa.

An additional cassette, named the TetR Responder$^{V1.0}$ was also designed. This cassette expresses tdTomato^{NLS} and was engineered to report on the presence of NMD. In this way, cells with high NMD activity can be identified not only by a lack of TetR:EGFP^{NLS} expressed from NMD$^{V1.0}$, but also by an increase in tdTomato^{NLS} expressed from TetR Responder$^{V1.0}$. To achieve this, the CAG promoter of TetR Responder$^{V1.0}$ was re-engineered to introduce two TetO sequences between the TATA box and initiator sequences of the CAG promoter. These TetO sequences are cognate binding sites for TetR. The rationale was that in cells with low NMD activity increased expression of TetR:EGFP^{NLS} from NMD$^{V1.0}$ would bind the TetO sequences and repress transcription from TetR Responder$^{V1.0}$, and vice versa in cells with high NMD activity. In this way, tdTomato^{NLS} would express in a proportional manner to NMD activity.

The final design of the Transgene$^{V1.0}$ was composed of four cassettes; Selection$^{V1.0}$, Control$^{V1.0}$, NMD$^{V1.0}$ and TetR Responder$^{V1.0}$. In this chapter I used molecular methods to conduct a series of functional tests to investigate if each of these cassettes performed as designed. The experimental validation runs through sequence confirmation, basic cassette feature testing and unique cassette feature testing. Following this process, I was able to confirm that all sequences had been synthesised as per the design and many of the basic cassette features, including the CAG promoter, NLSs and fluorescent proteins were functional within this design. Furthermore, by establishing NMD Reporter$^{V1.0}$ mESCs via RMCE, I was able to validate that the unique features of Selection$^{V1.0}$ were functional following minor non-functional sequence modifications from its published design (Bersten et al. 2015) and as part of a larger transgene containing Control$^{V1.0}$ and NMD$^{V1.0}$.

4.4.1 Following a reduction in NMD activity, protein expression from Control$^{V1.0}$ was seen to unexpectedly increase in human, but not mouse cells

To test the unique 3'UTR sequences of Control$^{V1.0}$ (NMD-refractive 3'UTR) and NMD$^{V1.0}$ (NMD-targeted 3'UTR), high and low NMD activity needed to be simulated within cells. To do this, siRNA targeting mRNA of the core NMD factor UPF1 was employed. This method successfully reduced levels of UPF1 protein in HEK293T, NIH3T3 and mESCs. Moreover, this was accompanied by a concomitant increase in NMD targeted endogenous mRNAs.

Following a reduction in UPF1 protein, expression from Control$^{V1.0}$ was expected to remain unchanged while expression from NMD$^{V1.0}$ was expected to dramatically increase. When these cassettes were expressed transiently from mouse NIH3T3 or from stable NMD Reporter$^{V1.0}$ mESCs (under stable integration conditions), protein expression from Control$^{V1.0}$ did not respond to a reduction in Upf1, while protein expression from NMD$^{V1.0}$ did increase, albeit only slightly. When cassettes were transiently expressed in HEK293T cells, an expected 91% increase in expression was observed from NMD$^{V1.0}$ following UPF1 reduction. However, expression of CFP^{NLS} from Control$^{V1.0}$ was also seen to increase by ~47%.

Further investigations uncovered some short comings in the design that to some extent may have contributed to the unforeseen results observed from functional testing of Control$^{V1.0}$ and NMD$^{V1.0}$. The first was the inability to spectrally resolve CFP from EGFP using the fluorescent microscopy filter sets available. This has the potential to confound fluorescence-based analysis of experiments in which CFP and EGFP are expressed simultaneously. As such, experiments where CFP^{NLS} encoded by Control$^{V1.0}$ and TetR:EGFP^{NLS} encoded by NMD$^{V1.0}$ were expressed in the same cell relied heavily on additional methods of analysis, such as western blots to be informative. To develop a successful fluorescence based NMD reporter system, this issue

would need to be resolved by either obtaining more stringent filter sets or replacing CFP and/or EGFP with fluorescent proteins that are further away from one another on an emission graph.

Another unpredicted issue was the inability of a single SV40 polyA sequence to efficiently terminate transcription downstream of a CAG promoter. The design of Transgene$^{V1.0}$ involves several expression cassettes in cis, each driven by their own promoter. As such, RNA polymerase reading through the SV40 polyA site of one cassette has the capacity to influence transcription of the following cassette.

Both fluorescent microscopy and western blot analysis revealed that following a reduction of UPF1 protein expression, HEK293T cells transiently transfected with the expression plasmid for Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) showed an unexpected increase in CFP^{NLS} expressed from Control$^{V1.0}$. Since Control$^{V1.0}$ is the first (i.e. upstream) cassette in this plasmid, and results were also observed by western blot analysis, neither compromised CFP spectral resolution nor inefficient transcription termination could explain this observation.

Further investigations which revisited the original sequence design, revealed that the 3'UTR of both Control$^{V1.0}$ and NMD$^{V1.0}$ was based on an alternative *HBB* transcript (*HBB-204*) rather than the dominant transcript (*HBB-201*). This error does not compromise the design of NMD$^{V1.0}$, however, the exon-intron-exon structure within the NMD-refractive 3'UTR of Control$^{V1.0}$ is replaced with an intron-exon structure. It was shown that due to this, transcripts expressed from Control$^{V1.0}$ remain predominantly unspliced. Without undergoing splicing there will be no opportunity for EJC deposition on these transcripts. As such, transcripts expressed from Control$^{V1.0}$ cannot be recognised by the NMD machinery and should remain 'NMD-refractive'.

To note, transcripts expressed from Control$^{V1.0}$ now contain an elongated 3'UTR (804 nts) which contains a stretch of intronic sequence of unknown relevance. Long 3'UTRs themselves can act as an NMD-targeting feature, which could explain the increase in expression from Control$^{V1.0}$ following a reduction in UPF1 of HEK293T cells, however, experimental evidence suggests that for this to occur the 3'UTR must be at least 1 kb in length (Ge et al. 2016; Jaffrey & Wilkinson 2018).

Another explanation considers that UPF1 is not restricted to functioning within the NMD pathway. SMD is another mRNA degradation pathway which acts in competition with NMD through UPF1 binding. SMD is mediated by the binding of STAU1 to a STAU1-binding site within the 3'UTR of target transcripts (Park & Maquat 2013). It is possible that such a binding site could be introduced through the retained intronic sequence within the elongated 3'UTR of Control$^{V1.0}$. Interestingly expression from Control$^{V1.0}$ did not increase following a Upf1 reduction when expressed transiently in mouse NIH3T3 cells or from stable NMD Reporter$^{V1.0}$ mESCs. This discrepancy in Control$^{V1.0}$ processing could be explained by a species dependent definition of degradation targets, a phenomena observed for both NMD and SMD pathways (Li et al. 2015; Lucas et al. 2018; Mendell et al. 2004).

4.4.2 Following a reduction in NMD activity, protein expression from NMD$^{V1.0}$ is more notably increased in human compared to mouse cells.

Following a loss of UPF1, protein expression from NMD$^{V1.0}$ was seen to increase by ~91% when this cassette was transiently expressed from human HEK293T cells. Comparatively, this response was not as pronounced when NMD$^{V1.0}$ was expressed transiently from mouse NIH3T3 or from stable NMD Reporter$^{V1.0}$ mESCs. The NMD targeting feature of NMD$^{V1.0}$ is it's NMD-targeted 3'UTR. This feature has been shown to successfully target mRNAs for NMD in both human and mouse cells as part of a published dual fluorescent NMD reporter system

(Pereverzev et al. 2015). It is thus unlikely, but not impossible, that some species dependent NMD recognition of the NMD-targeted 3'UTR could also explain the differential response of $NMD^{V1.0}$ to changes in NMD activity in mouse cells compared to human cells.

Another more likely explanation considers the context in which the NMD-targeted 3'UTR is being expressed. $NMD^{V1.0}$ contains many features that differ or are not included in the NMD reporter system developed by Pereverzev et al, 2015, these include the specific fluorescent proteins, the CAG promoter, and the addition of sequences encoding TetR. It is possible that by either disrupting splicing, silencing transcription, or promoting post transcriptional decay due to cell toxicity the sequences encoding any of these features could interfere with processing and NMD of transcripts expressed from $NMD^{V1.0}$. Moreover, such disruptions may be more likely to occur in one species compared to another i.e. in mouse versus human cells.

4.4.3 Chapter conclusions and future directions

In this chapter I have aided in the design of $Transgene^{V1.0}$ which is a single, internally controlled, fluorescent based NMD reporter transgene. This transgene would for the first time, allow a visual and quantifiable read-out of endogenous NMD activity at a single cell level and would be conducive to studying dynamic changes in NMD within complex tissue environments both *in vitro* and *in vivo*.

Following the design and synthesis of this transgene, functional validation was required. I have developed an experimental pipeline which allows validation of basic transgene features and NMD responsive features in both transient transfection-based experiments and stable genomic integration systems. Through this pipeline I was able to validate several design features and identify complications with others. Ultimately, $Transgene^{V1.0}$ was unsuccessful, with evidence suggesting that the NMD-refractive and NMD-targeted features employed may not be well suited to function within a system designed for use in stable mESC lines.

Collectively, the studies conducted in this chapter support the utility of a single cell dual-fluorescent NMD reporter transgene that can be used in mESCs. These pioneer studies have identified advantages and disadvantages of several key features within Transgene$^{V1.0}$. This knowledge can now be used to drive the development of a second version (Transgene$^{V2.0}$) with the purpose of fulfilling functionality.

Chapter Five: Design and Testing of a Fluorescent NMD Reporter System with Single Cell Resolution (Version 2.0)

5.1 Introduction

NMD is an intrinsic regulatory pathway essential for proper cell function. Chapter 4 focused on addressing the deficiencies of current NMD reporter systems to report on endogenous NMD activity by engineering an NMD reporter transgene (Transgene$^{V1.0}$) which can be integrated into the genome of mESCs. After extensive testing however, Transgene$^{V1.0}$ was found to be largely unsuccessful with a major concern being that the NMD targeting feature of the design appeared unresponsive to changes in NMD activity, particularly in mouse cells. It was hypothesised that although the feature had been shown to successfully target mRNAs for NMD in both human and mouse cells as part of a published dual fluorescent NMD reporter system (Pereverzev et al. 2015), the sheer complexity in the design of Transgene$^{V1.0}$ may have hindered the function of the NMD targeted feature. Moving forward it was decided to build on the knowledge acquired through the development and testing of Transgene$^{V1.0}$ to design and engineer a second NMD reporter transgene (Transgene$^{V2.0}$) with a simplified and improved design.

The signature phenomenon of NMD is a decreased abundance of mRNAs containing PTC-type mutations. This was first discovered through investigations into the genetic disorder β-thalassemia where NMD was found to limit the synthesis of C-terminally truncated proteins that might otherwise act in a dominant negative fashion (Chang & Kan 1979). The activation of such NMD was dependent on PTC location within the *HBB* transcript. PTC-type mutations that resided at least 55 nts upstream of an exon-exon junction were able to trigger efficient NMD of the variant transcripts. In these cases, protein production from the normal allele could compensate for loss of protein production from the variant allele such that heterozygous carriers of PTC-type mutations were phenotypically normal and escaped β-thalassemia which typifies the recessive form. In contrast, rare PTC-type mutations in the last exon (not followed by a downstream EJC following splicing) were unable to trigger NMD and resulted in translation of long, truncated and non-functional HBB proteins. As such, heterozygotes for PTC-type

mutations in *HBB* which cannot trigger NMD are affected with an atypical form of dominantly inherited β-thalassemia (Hall & Thein 1994; Thein et al. 1990).

β-thalassemia was the prototype disorder which documented the medical importance of NMD. In particular, the PTC-type mutation which affects codon 39 of the HBB protein (Nonsense 39, aka NS39 PTC) and results in production of HBB^{NS39} encoding transcripts, was found to trigger a robust and highly reproducible decrease in variant transcripts through NMD activation (Maquat 1995; Maquat et al. 1981; Thermann et al. 1998). In this way, the NS39 PTC can be considered the pioneering NMD activating mutation and has featured in the design of several published NMD reporter systems (Alexandrov, Shu & Steitz 2017; Baird et al. 2018; Pereverzev et al. 2015). The most commonly used of these reporters is a chemiluminescence-based NMD reporter system (Boelz et al. 2006). This reporter system consists of two in-frame *Renilla* luciferase/HBB fusion constructs; one encoding *Renilla* luciferase fused to wild-type HBB (HBB^{WT}) and one encoding *Renilla* luciferase fused to HBB^{NS39}. Through co-transfection of one of these constructs with a second control construct expressing firefly luciferase, NMD activity can be monitored through a ratio between *Renilla* luciferase expressed from the HBB^{NS39} fusion protein and *Renilla* luciferase expressed from the HBB^{WT} fusion protein, each normalised to firefly luciferase (Figure 5.1).

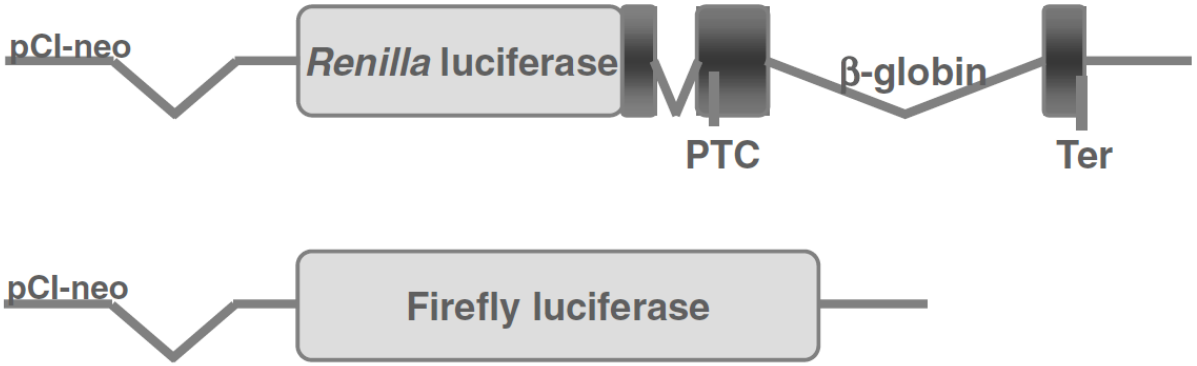


Figure 5.1: A chemiluminescence-based reporter system to monitor NMD.

A schematic representation of the published chemiluminescence-based, mammalian NMD reporter system. The reporter gene consists of an in-frame *Renilla* luciferase/HBB fusion construct with or without a nonsense mutation at codon 39 of the HBB open reading frame (PTC for premature termination codon or Ter for wild type stop signal). Figure adapted from (Boelz et al. 2006).

Being reliant on transient transfection and a chemiluminescent read out, the chemiluminescence-based NMD reporter is unable to report on endogenous NMD activity at a single cell level. However, since its development, this system has been used considerably more than that developed by Pereverzev et al. 2015 to report across different branches of NMD in both human and mouse cells (Bhuvanagiri et al. 2014; Boelz et al. 2006; Huang et al. 2011; Karam et al. 2015; Keeling et al. 2013; Lou et al. 2014; Shum et al. 2016). As such the possibility of adapting this system to develop a dual fluorescent NMD reporter construct with single cell resolution emerged.

If successfully developed, an NMD reporter system which can monitor NMD at a single cell level and which can be integrated into the genome of pluripotent stem cells or an animal zygote will allow NMD activity to be tracked across differentiation and development. This can provide insight into the emerging area of cell and tissue specific NMD. In this chapter I will describe the scientific techniques used to design, engineer, and experimentally test Transgene$^{V2.0}$ which is inspired by both Transgene$^{V1.0}$ and the chemiluminescence-based NMD reporter system developed by Boelz et al. 2006.

5.2 Design

5.2.1 A summary of the design and expected output of Transgene$^{V2.0}$

Prior to investigating the endogenous activity of NMD in cells or development of an NMD reporter mouse, a functional NMD reporter transgene must be established. Transgene$^{V2.0}$ was envisaged to encode a dual fluorescent NMD reporter system which can be used to visualise and quantify NMD activity at a single cell level. Once validated, this transgene, through genetic modification systems, can be introduced into the genome of transgenic FLP-in mESCs or into transgenic FLP-in mouse embryos at the zygote stage.

Fluorescent output was chosen as it allows single cell quantification of fixed or live cells using fluorescence microscopy and/or FACS. This will enable NMD activity to be tracked across development, giving insight into when, where and in which cell types NMD is most important. Tracking NMD across neural development is of particular interest given that a compromised NMD pathway due to genetic variants in NMD factor genes *UPF2* and *UPF3B* results in neurodevelopmental disorders (Chapter 3).

The design and output of Transgene$^{V2.0}$ is illustrated in Figure 5.2. It employs three expression cassettes, namely, Selection Cassette$^{V2.0}$ (Selection$^{V2.0}$), Control Cassette$^{V2.0}$ (Control$^{V2.0}$), and NMD Cassette$^{V2.0}$ (NMD$^{V2.0}$). Control$^{V2.0}$ and NMD$^{V2.0}$ are largely based on the design published by Boelz et al, 2006, however, these have been adapted to provide fluorescent rather than a chemiluminescent output. These cassettes have also been adapted to function within a single transgene, where Control$^{V2.0}$ will constitutively express the nuclear localised and HA-epitope tagged cyan fluorescent protein/HBB fusion construct (^{HA}CFP^{NLS}:HBB^{WT}). Similarly, NMD$^{V2.0}$ encodes a nuclear localised and FLAG-epitope tagged yellow fluorescent protein/HBB fusion construct, in this instance however, the sequence encoding HBB contains the NMD targeted NS39 PTC (^{FLAG}YFP^{NLS}:HBB^{NS39}). This means that expression of NMD$^{V2.0}$ should be influenced by cellular NMD activity such that, cells with low NMD activity should

efficiently express the $^{FLAG}YFP^{NLS}$:HBB^{NS39} fusion protein from $NMD^{V2.0}$, whereas in cells with high NMD activity, transcripts expressed from $NMD^{V2.0}$ are predicted to be degraded by NMD resulting in reduced $^{FLAG}YFP^{NLS}$:HBB^{NS39} expression.

Finally, to facilitate integration of the entire transgene into cells or mouse zygotes, $Selection^{V2.0}$, which through the previously described RMCE system (Bersten et al. 2015) will target the transgene to the genetically modified *Col1a1* locus of FLP-in mESCs or FLP-in mouse embryos. The design of each cassette will be further detailed in Section 5.2.2.

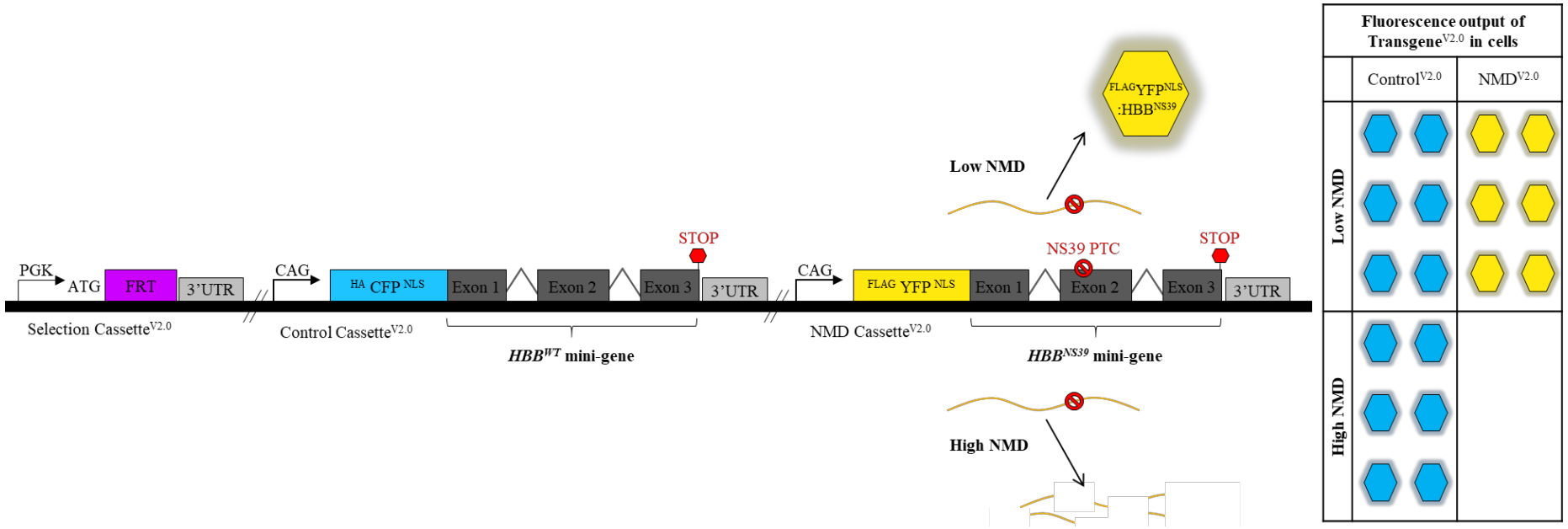


Figure 5.2: A schematic representation of Transgene $^{V2.0}$ and its fluorescence output in cells with low or high NMD activity.

Control$^{V2.0}$ encodes the fluorescent protein/HBB fusion construct ^{HA}CFP^{NLS}:HBB^{WT}. Transcripts expressed from this cassette are not targeted by NMD and thus serve as an internal control expressing a constant level of ^{HA}CFP^{NLS}:HBB^{WT} regardless of cellular NMD activity. NMD$^{V2.0}$ encodes the fluorescent protein/HBB fusion construct ^{FLAG}YFP^{NLS}:HBB^{NS39}. Transcripts expressed from this cassette are targeted for NMD due to the presence of the NS39 PTC within the sequence encoding *HBB* (*HBB^{NS39}* mini-gene). Therefore, cells with low NMD activity will express greater levels of ^{FLAG}YFP^{NLS}:HBB^{NS39} compared to cells with high NMD activity. Levels of YFP can be normalised to CFP to provide an NMD activity reading. This can be quantified at a single cell level due to nuclear localisation signals (NLS). Selection$^{V2.0}$ facilitates genomic integration into FLP-in mESCs via RMCE.

5.2.2 A detailed description of the design and expected output of Transgene$^{V2.0}$

5.2.2.1 Expression plasmids encoding the cassettes of Transgene$^{V2.0}$

Transgene$^{V2.0}$ contains three cassettes, Selection$^{V2.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$. Several features of Control$^{V2.0}$ and NMD$^{V2.0}$ have been re-designed based on observations from functional testing of Transgene$^{V1.0}$ (Chapter 4). These updates have been summarised in Table 5.1.

The sequences for all cassettes were designed using the plasmid editor 'ApE' (Davis 2019), and synthesised under a fee-for-service contract with Genscript (Genscript, USA). Since the features of Selection$^{V2.0}$ had been previously validated (Chapter 4), the sequences for all three cassettes of Transgene$^{V2.0}$ were obtained within just two plasmids using the pUC57-kan backbone. The first plasmid contained sequences for both Selction$^{V2.0}$ and Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) while the second plasmid contained the sequence for NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$) (Table 5.2, Figure 5.3). By beginning with the sequences for Control$^{V2.0}$ and NMD$^{V2.0}$ within separate plasmids, the updated features of each cassette can be functionally validated prior to assembly into Transgene$^{V2.0}$ (Figure 5.4).

To facilitate assembly of the two cassettes into Trangene$^{V2.0}$, the design of each of these cassettes included *Not*I and *BstB*I restriction endonuclease sites which allow either plasmid to be used as the 'vector backbone' or 'insert fragment' in ligase based recombinant DNA cloning methods (Figure 5.4). Additional restriction sites were also included to facilitate swapping in and out of entire cassettes, editing of unique features or introduction of transgene sequence modifications if necessary (Figure 5.4).

Table 5.1: The updated features of Control $^{V2.0}$ and NMD$^{V2.0}$ from Transgene$^{V1.0}$

Updated Feature	**Function**	**Rationale**
Three consecutive SV40 polyA sequences within the 3'UTR.	Terminator sequences that signal the end of a transcriptional unit.	A single polyA sequence could not efficiently terminate transcription downstream of the CAG promoters in Transgene$^{V1.0}$.
Sequences encoding YFP (mVenus) in the design on NMD$^{V2.0}$.	Fluorescent protein used to visualise expression from NMD$^{V2.0}$.	EGFP expressed from NMD$^{V1.0}$ interfered with spectral resolution of CFP expressed from Control$^{V1.0}$. The spectral properties of YFP do not overlap with CFP which will be used in the design of Control$^{V2.0}$.
N-terminal HA and FLAG epitope tags in the design of Control$^{V2.0}$ and NMD$^{V2.0}$ respectively.	To allow robust protein detection via western blot and immunofluorescent techniques.	Control$^{V2.0}$ and NMD$^{V2.0}$ encode proteins that are similar in size. Unique epitope tags on these proteins should enable their individual detection.
Translated fluorescent protein/HBB fusion constructs based on the design of a previously published NMD reporter system (Boelz et al, 2006).	To facilitate differential fluorescent expression of Control$^{V1.0}$ and NMD$^{V1.0}$ in response to cellular NMD.	Since the NMD feature used in Transgene$^{V1.0}$ appeared unresponsive to a reduction in NMD activity, the NMD feature from the most commonly used NMD reporter was adopted into the design of Transgene$^{V2.0}$.

Table 5.2:A description of Version 2.0 expression plasmids and the proteins they encode

Plasmid	Supplier	Expression cassette(s) included in plasmid			Encoded Protein(s)	Expected protein size(s)
		Selection (S)	Control (C)	NMD (N)		
pUC57-kan-SC$^{V2.0}$	Genscript	S$^{V2.0}$	C$^{V2.0}$	-	^{HA}CFP^{NLS}:HBB^{WT}	~46 kDa
pUC57-kan-N$^{V2.0}$	Genscript	-	-	N$^{V2.0}$	^{FLAG}YFP^{NLS}:HBB^{NS39}	~35 kDa
pUC57-kan-SCN$^{V2.0}$	In-house	S $^{V2.0}$	C $^{V2.0}$	N $^{V2.0}$	^{HA}CFP^{NLS}:HBB^{WT}, ^{FLAG}YFP^{NLS}:HBB^{NS39}	~46 kDa, ~35 kDa

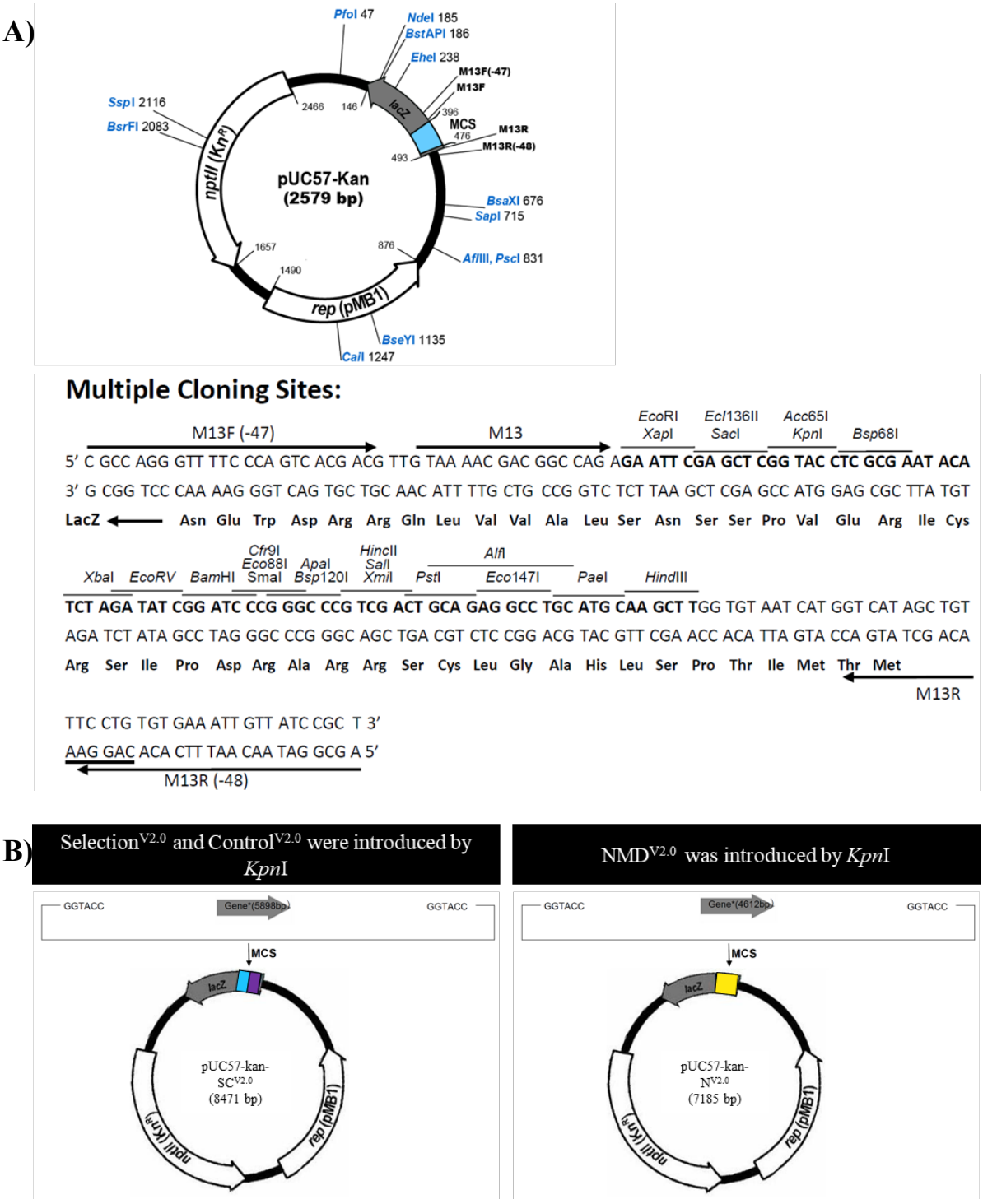


Figure 5.3: Preparation of expression plasmids encoding the cassettes of Transgene$^{V2.0}$ via the contracted services of Genscript.

(A) A plasmid map of the pUC57-kan plasmid backbone detailing the sequence of the multiple cloning site (MCS) and restriction endonuclease recognition sites. **(B)** A schematic representation of the restriction endonucleases used by Genscript to introduce synthesised cassette DNA sequences into the MCS of pUC57-kan.

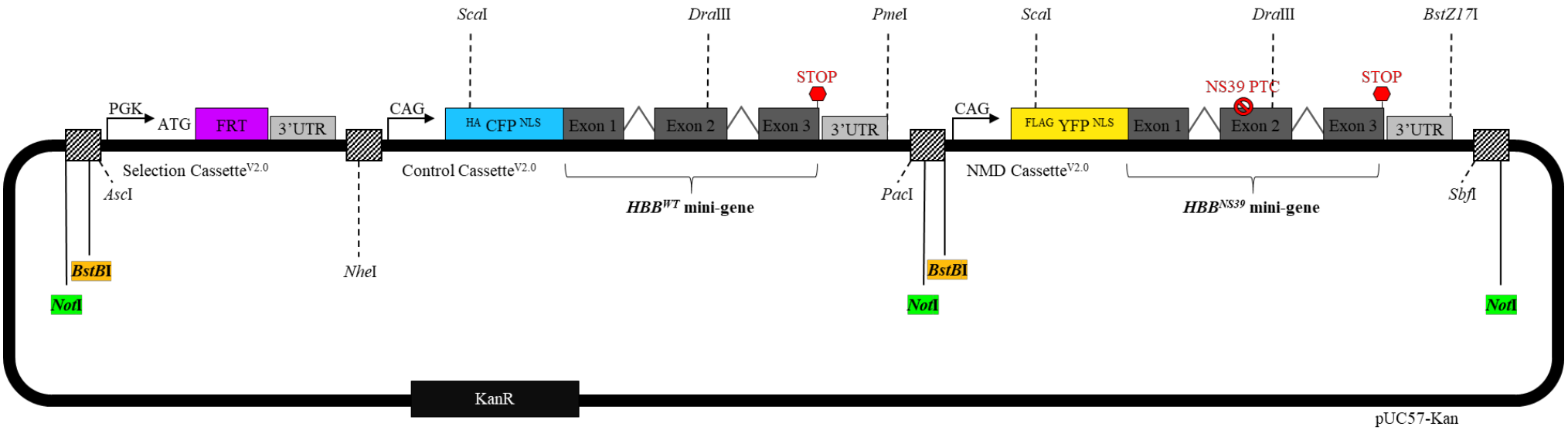


Figure 5.4: A detailed schematic of Transgene$^{V2.0}$.

Transgene$^{V2.0}$ is composed of three cassettes in cis, namely, Selection$^{V2.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$. The plasmid backbone used to harbour Transgene$^{V2.0}$ is pUC57-kan. All cassettes apart from Selection$^{V1.0}$ encode protein. Control$^{V2.0}$ and NMD$^{V2.0}$ encode an N-terminally epitope tagged and nuclear localised fluorescent protein/HBB fusion construct. Each cassette is driven by its own promoter and flanked by at least one restriction endonuclease site which allows assembly of the cassettes into a single vector with a pUC57-kan backbone via standard recombinant DNA cloning methods (bold lines). Additional restriction sites have also been included to drop in and out entire cassettes or to edit specific features within a cassette (dashed lines).

5.2.2.2 Design and function of Selection$^{V2.0}$

Selection$^{V2.0}$ is one of two cassettes encoded by pUC57-kan-SC$^{V2.0}$. The sequence of this cassette remains largely unchanged from Selection$^{V1.0}$ (Chapter 4) which has undergone successful functional validation. The role of Selection$^{V2.0}$ is to facilitate integration of the entire Transgene$^{V2.0}$ as a single, stable copy into the *Col1a1* locus of FLP-in mESCs (or FLP-in mouse zygotes) using the previously described RMCE system (Chapter 4) (Bersten et al. 2015).

5.2.2.3 Design and function of Control$^{V2.0}$ and NMD$^{V2.0}$

Control$^{V2.0}$ and NMD$^{V2.0}$ both encode a nuclear localised and N-terminally epitope tagged fluorescent protein/HBB fusion construct and are based on the pioneering chemiluminescent-based NMD reporter system (Boelz et al. 2006). The *HBB* mini-gene of both cassettes contains exonic and intronic sequences to ensure that the expressed transcripts undergo pre-mRNA splicing which results in the deposition of EJCs necessary for NMD recognition. Specifically, Control$^{V2.0}$ encodes ^{HA}CFP^{NLS}:HBB^{WT} protein (Figure 5.5). By design, transcripts expressed from this cassette contains no NMD-targeting features and are predicted to express at a constant level regardless of cellular NMD activity. In this way expression from Control$^{V2.0}$ acts as an internal control for Transgene$^{V2.0}$.

NMD$^{V2.0}$ encodes ^{FLAG}YFP^{NLS}:HBB^{NS39}. The main point of difference between NMD$^{V2.0}$ and Control$^{V2.0}$ is that the sequence of the *HBB* mini-gene within NMD$^{V2.0}$ contains the NMD-targeted NS39 PTC, as such transcripts expressed from this cassette are subject to NMD. Therefore, in cells with high NMD activity low levels of ^{FLAG}YFP^{NLS}:HBB^{NS39} will be expressed from NMD$^{V2.0}$, and vice versa. In this design NMD activity is quantified by the level of fluorescence emitted from NMD$^{V2.0}$ normalised to that from Control$^{V2.0}$.

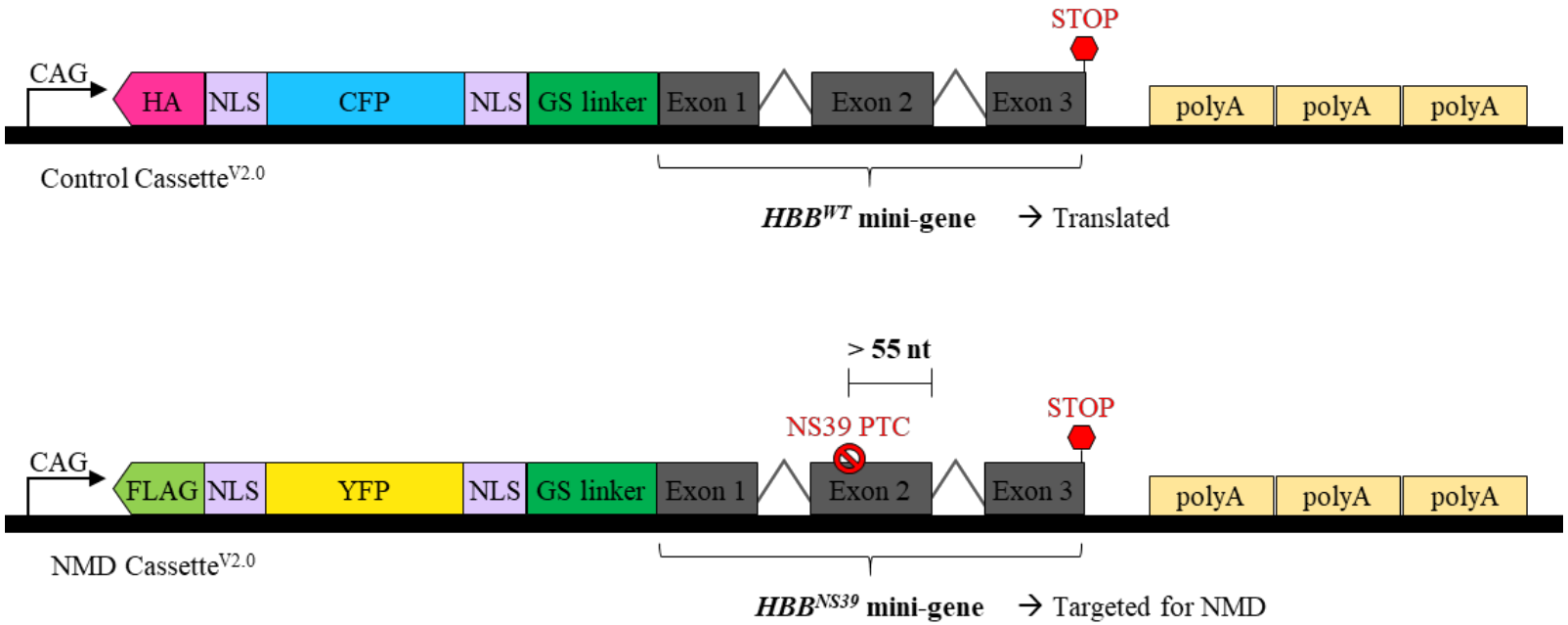


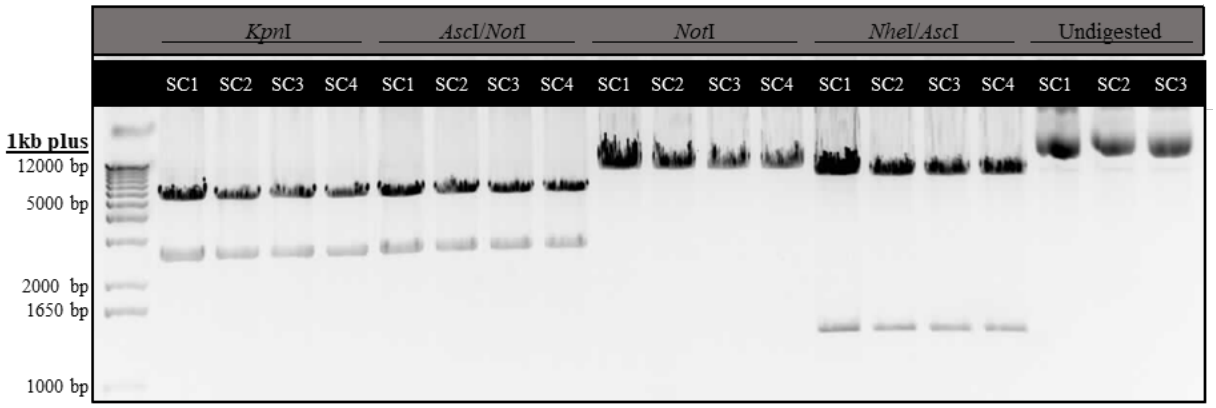
Figure 5.5: A schematic comparison between the design of Control$^{V2.0}$ and NMD$^{V2.0}$.

Control$^{V2.0}$ and NMD$^{V2.0}$ are both driven by a CAG promoter and encode an N-terminally tagged, fluorescent protein/HBB fusion construct linked by a glycine serine (GS) linker sequence and followed by a 3'UTR containing three SV40 polyA sequences (polyA). Control$^{V2.0}$ is designed to function as an internal control for Transgene$^{V2.0}$, specifically this cassette contains a *HBB^{WT}* mini-gene which does not contain any NMD targeting features. Comparatively, NMD$^{V2.0}$ contains a *HBB* mini-gene encoding the NMD targeted NS39 PTC (*HBB^{NS39}* mini-gene). Transcripts expressed from NMD$^{V2.0}$ are therefore subject to NMD. Both Control$^{V2.0}$ and NMD$^{V2.0}$ contain two nuclear localisation signals (NLS) to enable single cell quantification of fluorescence.

5.3 Results

5.3.1 Sequence validation of Genscript synthesised expression plasmids

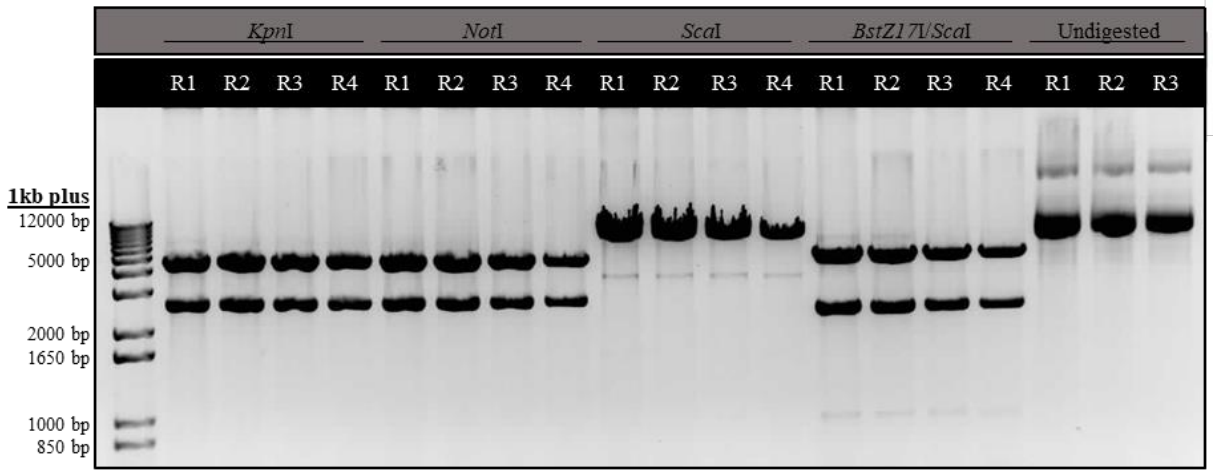
To ensure that the contracted synthesis of all DNA sequences for the cassettes of Transgene$^{V2.0}$ was completed without error, sequence validation was initially carried out via diagnostic restriction endonuclease digestion of expression plasmids encoding Selection $^{V2.0}$ and Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) or NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$). The digested DNA products alongside undigested plasmid DNA for comparison were visualised by agarose gel electrophoresis (Figures 5.6 & 5.7). The observed product sizes matched the expected digest product sizes, suggesting that there were no large sequence errors. Selected constructs were then subject to Sanger sequencing (data not shown) to ensure the fidelity of each cassette prior to use in subsequent experiments. In each case the sequence of the entire cassette i.e. Selection$^{V2.0}$, Control$^{V2.0}$ or NMD$^{V2.0}$ (excluding the pUC57-kan backbone) was sequenced and verified.



	***Kpn*I**	***Asc*I / *Not*I**	***Not*I**	***Asc*I / *Nhe*I**	**Undigested**
Expected outcome	Excise Selection$^{V2.0}$ and Control$^{V2.0}$ from pUC57-kan-SC$^{V2.0}$	Excise Selection$^{V2.0}$ and Control$^{V2.0}$ from pUC57-kan-SC$^{V2.0}$	Linearise pUC57-kan-SC$^{V2.0}$	Digest within pUC57-kan-SC$^{V2.0}$	Control
Expected digestion product size	5892 + 2579 bp	5866 + 2605 bp	8471 bp	7195 + 1276 bp	8471 bp

Figure 5.6: Diagnostic restriction endonuclease digests support correct sequence synthesis of pUC57-kan-SC$^{V2.0}$.

The Genscript synthesised expression plasmid encoding Selection$^{V2.0}$ and Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) was subjected to diagnostic restriction endonuclease digests. The table summarises the restriction endonucleases used and the expected digest product sizes following digestion of a correctly synthesised plasmid. The digested product alongside undigested plasmid DNA for comparison was visualised by agarose gel electrophoresis. The expected band sizes were observed for pUC57-kan-SC$^{V2.0}$.



	***Kpn*I**	***Not*I**	***Sca*I**	***BstZ17*I / *Sca*I**	**Undigested**
Expected outcome	Excise $NMD^{V2.0}$ from pUC57-kan-$N^{V2.0}$	Excise $NMD^{V2.0}$ from pUC57-kan-$N^{V2.0}$	Linearise pUC57-kan-$N^{V2.0}$	Digest within pUC57-kan-$N^{V2.0}$	Control
Expected digestion product size	4606 + 2579 bp	4592 + 2593 bp	7185 bp	4841 + 2334 bp	7185 bp

Figure 5.7: Diagnostic restriction endonuclease digests support correct sequence synthesis of pUC57-kan-$N^{V2.0}$.

The Genscript synthesised expression plasmid encoding $NMD^{V2.0}$ (pUC57-kan-$N^{V2.0}$) was subjected to diagnostic restriction endonuclease digests. The table summarises the restriction endonucleases used and the expected digest product sizes following digestion of a correctly synthesised plasmid. The digested product alongside undigested plasmid DNA for comparison was visualised by agarose gel electrophoresis. The expected band sizes were observed for pUC57-kan-$N^{V2.0}$.

5.3.2 Functional testing of Control$^{V2.0}$ and NMD$^{V2.0}$ using expression plasmids in a transient setting

To establish a stable NMD reporter system, Transgene$^{V2.0}$ is designed to be integrated into the genome of FLP-in mESCs (or FLP-in mouse zygotes) using the previously established RMCE system which involves Selction$^{V2.0}$ (Bersten et al. 2015). To enable functional testing of Control$^{V2.0}$ and NMD$^{V2.0}$ prior to assembly into Transgene$^{V2.0}$, the experimental pipeline which was successfully used to test the cassettes of Transgene$^{V1.0}$ (Chapter 4) was used. Firstly, expression plasmids encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) or NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$) were co-transfected into HEK293T cells, alongside either Control or *UPF1* siRNA to reduce cellular NMD activity. To enable downstream analysis, the transfected cells were then either fixed onto coverslips or used to isolate protein and RNA (Figure 5.8).

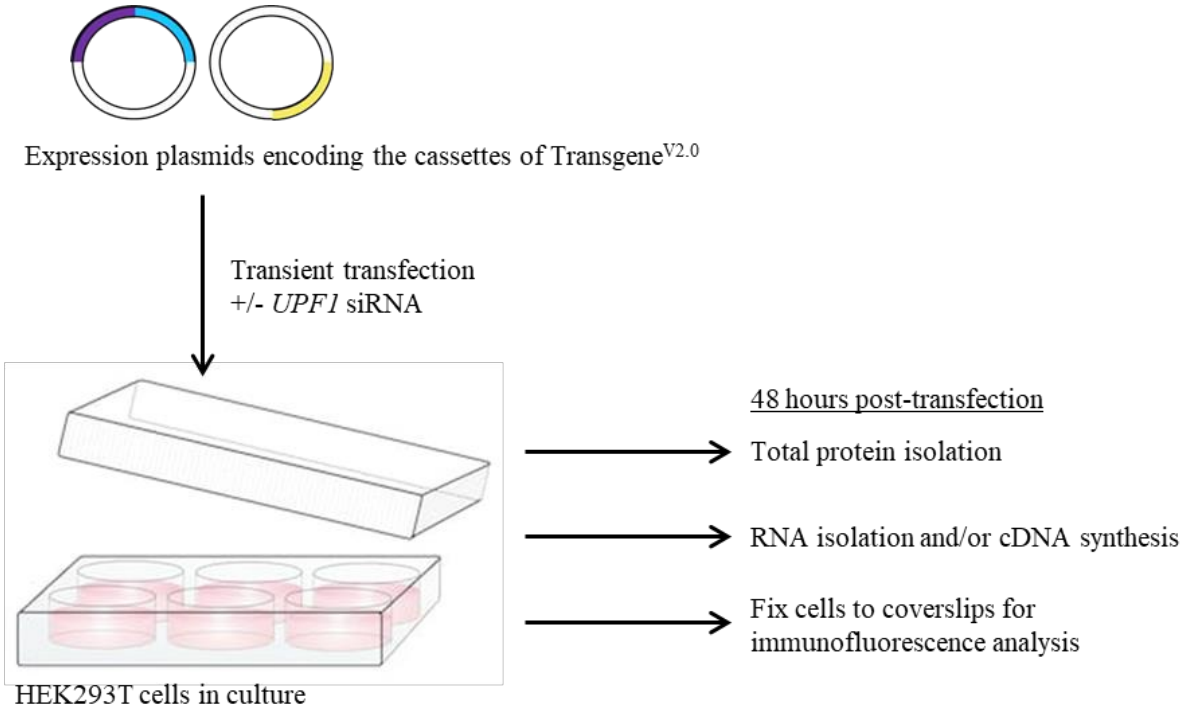


Figure 5.8: The experimental pipeline used for functional testing of expression plasmids encoding the cassettes of Transgene$^{V2.0}$.

A schematic representation of the basic molecular pipeline used to functionally test the features of Control$^{V2.0}$ and NMD$^{V2.0}$.

5.3.2.1 Using fluorescence expression to assess function of basic cassette features

Control $^{V2.0}$ and NMD$^{V2.0}$ share several identical or similar sequences encoding basic features of an expression cassette. These are, the promoter, fluorescence proteins, and NLSs. Fluorescence microscopy analysis of HEK293T cells transiently transfected with expression plasmids encoding either Control$^{V2.0}$ or NMD$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$) revealed that fluorescence expressed from both plasmids was restricted to the nuclear compartments (Figure 5.9). This concluded that the promoters are functional, pre-mRNA splicing events can generate translationally competent protein encoding mRNA, and that the fluorescent proteins are functional as are their NLSs.

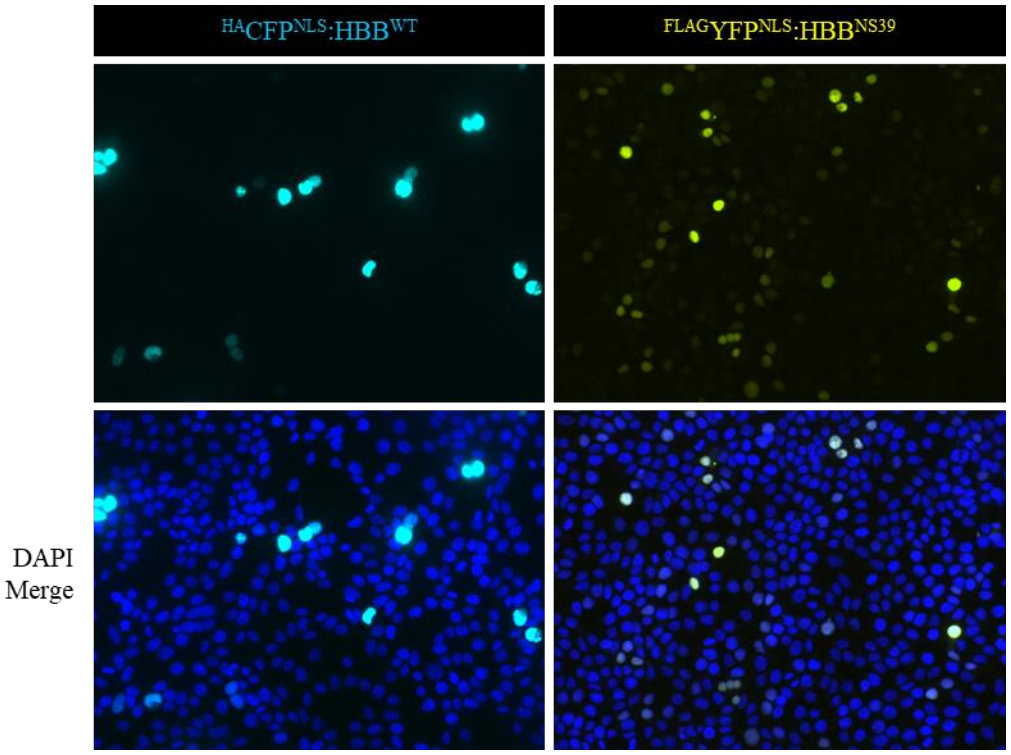


Figure 5.9: Fluorescent imaging of cells expressing Control$^{V2.0}$ and NMD$^{V2.0}$ supports correct function of basic cassette features.

Representative fluorescence microscopy images of HEK293T cells transfected with expression plasmids encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) or NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$). These cells express ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ (yellow). Cell nuclei were counterstained with DAPI (blue).

5.3.2.2 Testing fluorescent spectral resolution of proteins expressed from Control$^{V2.0}$ and NMD$^{V2.0}$

When designing a fluorescent reporter, the choice of fluorescent protein(s) used is influenced by several factors, including; sufficient brightness to be easily distinguished from autofluorescence, high photostability for lengthy imaging experiments, and if the fluorescent protein is to be part of a fusion protein it should not be able to oligomerise easily (Shaner, Steinbach & Tsien 2005).

Through functional testing of the expression cassettes of Transgene$^{V1.0}$ in Chapter 4, it was found that CFP and EGFP signals could not be completely spectrally resolved from one another using the fluorescence microscope and filter sets available. CFP (Cerulean) was again used in the design of Control$^{V2.0}$, however, to address the previously observed complications with spectral resolution, YFP (mVenus) rather than EGFP was used in the design of NMD$^{V2.0}$. YFP has a greater excitation and emission wavelength than EGFP which places it further away from CFP on an emissions graph (Figure 5.10A). As such, YFP is less likely to interfere with the spectral resolution of CFP than EGFP. Moreover, YFP is brighter, more photostable, and less likely to oligomerise than EGFP, making it a better choice in all regards. The properties of CFP and YFP are summarised in Table 5.3 and their emission and excitation plots shown in (Figure 5.10A). The properties of EGFP were previously summarised in Chapter 4 (Table 4.3).

Through fluorescent microscopy, it was observed that the filter sets available to us for the Zeiss AxioImager M2 fluorescent microscope (Car Zeiss, Germany) were able to spectrally resolve CFP (expressed from pUC57-kan-SC$^{V2.0}$) from YFP (expressed from pUC57-kan-N$^{V2.0}$), and vice versa (Figure 5.10B).

Table 5.3: Properties of the fluorescent proteins designed to express from Transgene$^{V2.0}$

Class	**Protein**	**Excitation**	**Emission**	**Brightness**	**Photostability**	**Oligomerization**	**Molecular Weight**	**References**
Cyan	Cerulean (CFP)	433 nm	475 nm	27	36 $t_{1/2}$ (s)	Weak dimer	26.8 kDa	(Lelimousin et al. 2009; Shaner, Steinbach & Tsien 2005)
Yellow	mVenus (YFP)	515 nm	527 nm	66.56	60 $t_{1/2}$ (s)	Monomer	26.9 kDa	(Shaner, Steinbach & Tsien 2005; Zhong et al. 2019)

(a) Major excitation peak. (b) Major emission peak. (c) Product of extinction coefficient and quantum yield at pH 7.4 (d) Time for bleaching from an initial emission rate of 1000 photons/s down to 500 photons/s ($t_{1/2}$) in seconds (s).

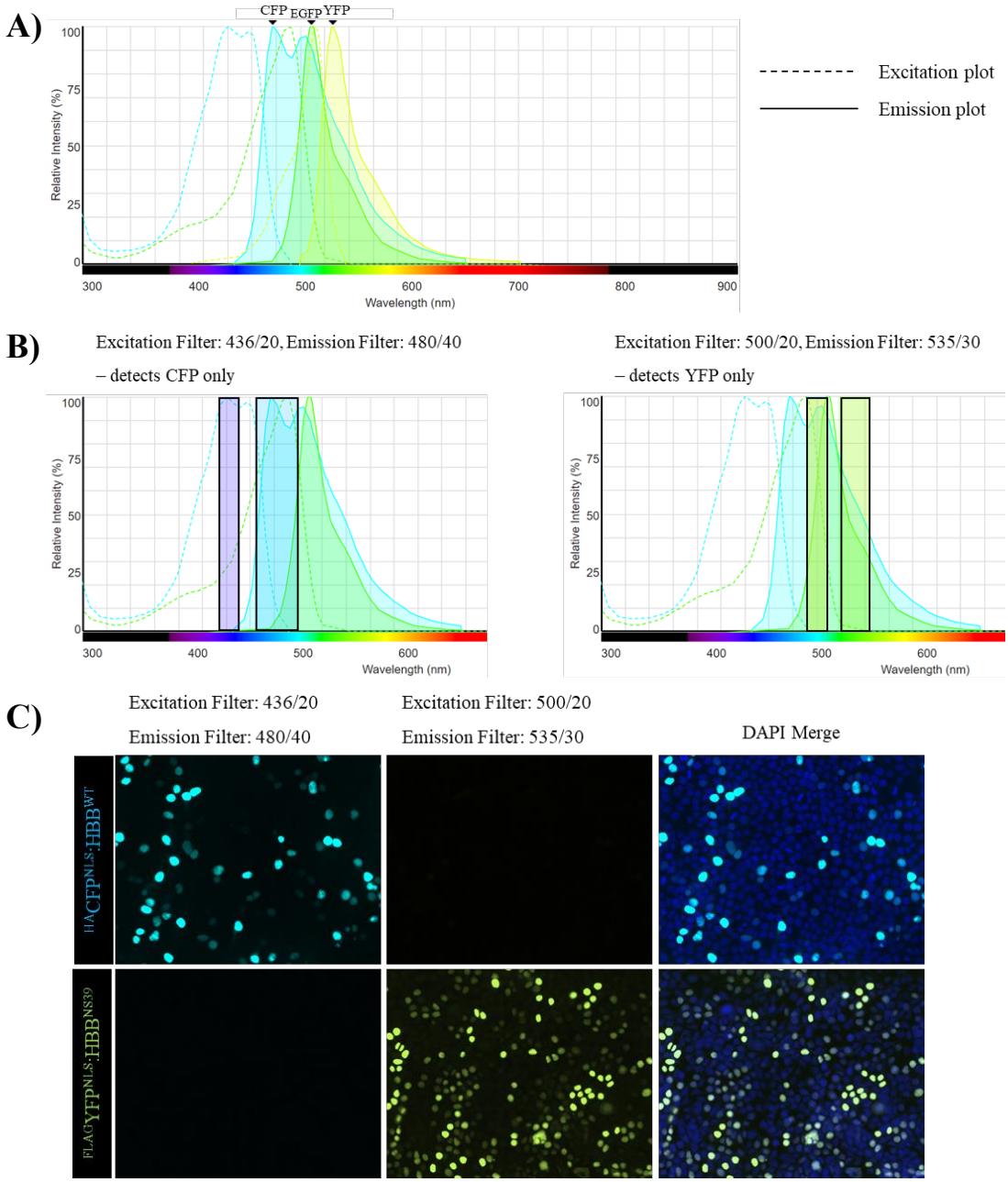


Figure 5.10: CFP and YFP can be spectrally resolved from one another.

(A) Emission and excitation plots for CFP, EGFP and YFP. **(B)** Overlap of excitation and emission filter sets used to resolve CFP (excitation filter: 436/20, emission filter: 480/40) and YFP (excitation filter 500/20: emission filter: 535/30). **(C)** Representative fluorescence microscopy images of HEK293T cells transfected with expression plasmids encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) or NMD$^{V2.0}$ (pUC57-kan-N$^{V1.0}$). These cells express ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ (yellow). Cell nuclei were counterstained with DAPI (blue). The filter sets available allow spectral resolution of CFP and YFP from one another.

5.3.2.3 Testing if three SV40 polyA sequences can efficiently terminate transcription downstream of a CAG promoter

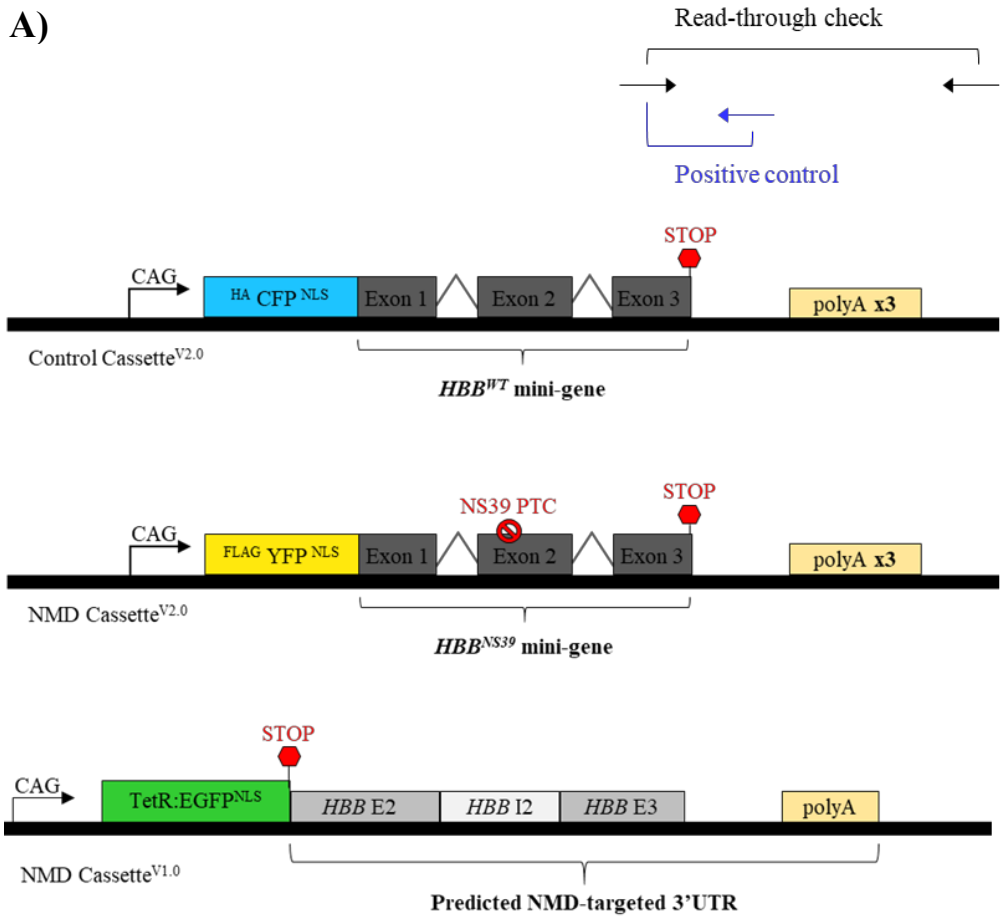
Control$^{V1.0}$ and NMD$^{V1.0}$ of Transgene$^{V1.0}$ contained only a single SV40 polyA sequence within their 3'UTRs which was unable to efficiently terminate transcription. To address this issue, the 3'UTR of both Control$^{V2.0}$ and NMD$^{V2.0}$ was designed to harbour three SV40 polyA sequences in tandem.

To assess the efficiency of transcription termination from Control$^{V2.0}$ and NMD$^{V2.0}$, three PCR primers (one forward and two reverse primers) were designed to perform in the same reaction. Two primers either amplify a sequence within the transcriptional unit to act as a positive control for the PCR conditions (positive control) or span the SV40 poly sequences to assess if transcription termination is efficient (read-through check) (Figure 5.11A). Read-through check primers will only amplify a product if RNA polymerase read-through beyond the SV40 polyA sites is occurring i.e. transcription termination is not efficient. cDNA reverse transcribed from the RNA of HEK293T cells transfected with an expression plasmid encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$), NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$) or both Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) (see Section 5.3.2.4.2) was used as a template in this PCR. cDNA reverse transcribed from the RNA of NMD Reporter$^{V1.0}$ mESCs or NMD Reporter $^{V2.0}$ mESCs (see Section 5.3.3) was also subject to the same PCR for comparison.

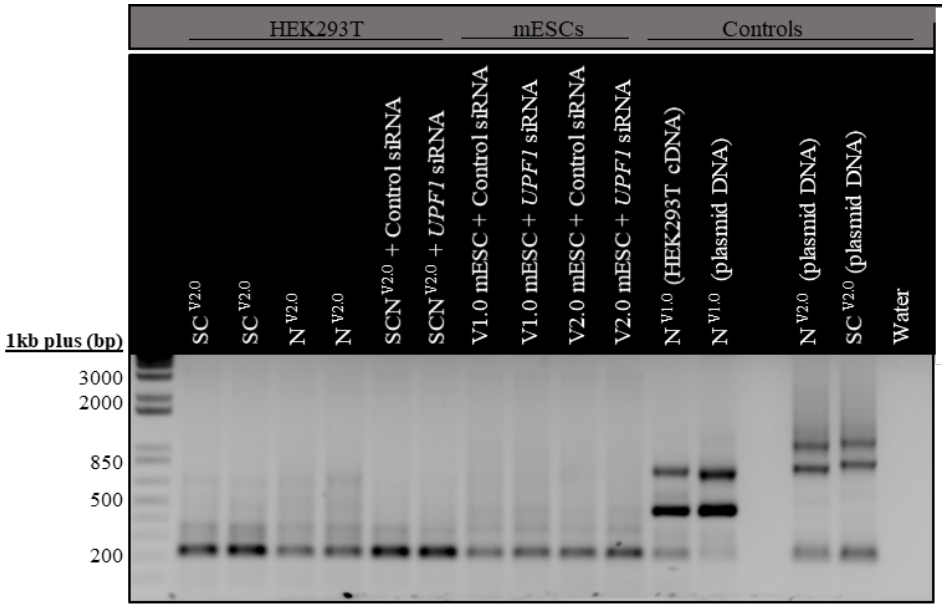
PCR product was visualised by agarose gel electrophoresis. In all cases transcription termination was seen to be efficient from Control$^{V2.0}$ and NMD$^{V2.0}$ (Figure 5.11B). This suggests that three SV40 polyA sequences can efficiently terminate transcription downstream of a CAG promoter.

Efficient transcription termination was also seen from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of stable NMD Reporter$^{V1.0}$ mESCs which both contained only one SV40 polyA sequence within their 3'UTR (Figure 5.11B). This was interesting as previous transient transfection-based investigations in HEK293T cells showed that a single SV40 polyA sequence was unable to efficiently terminate transcription (Chapter 4). This data suggests that, at least for the expression cassettes designed thus far, the efficiency of transcriptional termination may be influenced by cell type (HEK293T cells versus mESCs) or differ based on transient versus stable expression of cassettes.

A)



B)



	SC$^{V2.0}$	**N$^{V2.0}$**	**SCN$^{V2.0}$**	**mESC$^{V1.0}$**	**mESC$^{V2.0}$**	**N$^{V1.0}$**
Read-through check	No band **OR** 797 bp if read-through	No band **OR** 866 bp if read-through	No band **OR** 797 +866 bp if read-through	No band **OR** 476 bp if read-through	No band **OR** 797 +866 bp if read-through	No band **OR** 476 bp if read-through
Positive control	218 bp	218 bp	218 bp	218 bp	218 bp	218 bp

Figure 5.11: Three SV40 polyA sequences downstream of a CAG promoter can efficiently terminate transcription.

(A) A Schematic representation of primers designed to assess the efficiency of transcription termination from Control$^{V2.0}$, NMD$^{V2.0}$ and NMD$^{V1.0}$. NMD$^{V1.0}$ contains a single SV40 polyA sequence within its 3'UTR, while Control$^{V2.0}$ and NMD$^{V2.0}$ contain three SV40 polyA sequences in tandem within their 3'UTRs. **(B)** cDNA reverse transcribed from the RNA of NMD Reporter$^{V1.0}$ mESCs (V1.0 mESC), NMD Reporter$^{V2.0}$ mESCs (V2.0 mESC) and cDNA reverse transcribed from the RNA of HEK293T cells transfected with expression plasmids encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$, SC$^{V2.0}$), NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$, N$^{V2.0}$) or both Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$, SCN$^{V2.0}$) was subject to PCR using the three described read-through primers. Plasmid DNA of these constructs and cDNA from HEK293T cells transfected with an expression plasmid encoding NMD$^{V1.0}$ (pUC57-N$^{V1.0}$, N$^{V1.0}$) was also subject to the same PCR as controls to identify products indicative of RNA polymerase read-through at the SV40 polyA sequence(s). PCR amplified products were visualised by agarose gel electrophoresis. The table outlines the possible product sizes from this PCR. Banding patterns reveal that three polyA sequences can efficiently terminate transcription downstream of a CAG promoter, whilst a single polyA sequence can efficiently terminate transcription when Version 1.0 cassettes are expressed from stable NMD Reporter$^{V1.0}$ mESCS but not when transiently expressed in HEK293T cells.

5.3.2.4 Testing the expression from Control$^{V2.0}$ and NMD$^{V2.0}$ in response to changes in NMD activity levels of HEK293T cells

5.3.2.4.1 Testing protein expression from Control$^{V2.0}$ and NMD$^{V2.0}$ in response to an siRNA mediated reduction of UPF1 levels in HEK293T cells

To test if Control$^{V2.0}$ or NMD$^{V2.0}$ are sensitive to NMD, their respective expression plasmids (pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$) were co-transfected separately into HEK293T cells alongside either Control or *UPF1* siRNA. This *UPFI* siRNA had been previously shown to reduce NMD activity in these cells (Chapter 4). Under Control siRNA conditions, only transcripts expressed from NMD$^{V2.0}$ should be degraded by NMD. Therefore, following a UPF1 reduction (i.e. a reduction in NMD activity) an increase in expression of ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ is expected. Comparatively, transcripts expressed from Control$^{V2.0}$ should evade NMD, resulting in no change in expression of ^{HA}CFP^{NLS}:HBB^{WT} following a reduction in UPF1

Protein isolated from these co-transfected HEK293T cells was visualised by western blot and fluorescence microscopy. These analysis showed that following UPF1 reduction, ^{HA}CFP^{NLS}:HBB^{WT} protein expression from Control$^{V2.0}$ remained unchanged in two out of three replicates, and was reduced in one (Figures 5.12 & 5.13A). This suggests that Control$^{V2.0}$ was functioning largely as expected with an outlier result obtained for 'replicate 2', likely involving some experimental error or variance. The same analysis also showed an obvious, but inconsistent increase in magnitude of ^{FLAG}YFP^{NLS}:HBB^{NS39} protein expressed from NMD$^{V2.0}$ across all three replicates (Figures 5.12 & 5.13B).

It was noted that ^{FLAG}YFP^{NLS}:HBB^{NS39} was unable to be detected by western blot when using a HBB antibody (Figure 5.12). The epitope recognised by this antibody could not be disclosed by the manufacture and therefore the observed result was inconclusive as it could mean that the

recognised HBB epitope resides downstream of the NS39 PTC or although less llikely, folding of $^{FLAG}YFP^{NLS}$:HBB^{NS39} may have rendered the recognised epitope inaccessible.

It was also noted that the FLAG antibody used to detect $^{FLAG}YFP^{NLS}$:HBB^{NS39} also detected an unexpected 46 kDa protein only in protein samples isolated from pUC57-kan-$SC^{V2.0}$ transfections (Figure 5.12). This protein was not present in protein samples isolated from pUC57-kan-$N^{V2.0}$ transfections. Together this information suggests that despite the lack of a FLAG tag, the FLAG antibody may be detecting the $^{HA}CFP^{NLS}$:HBB^{WT} fusion protein expressed from $Control^{V2.0}$ which is expected to be ~46 kDa.

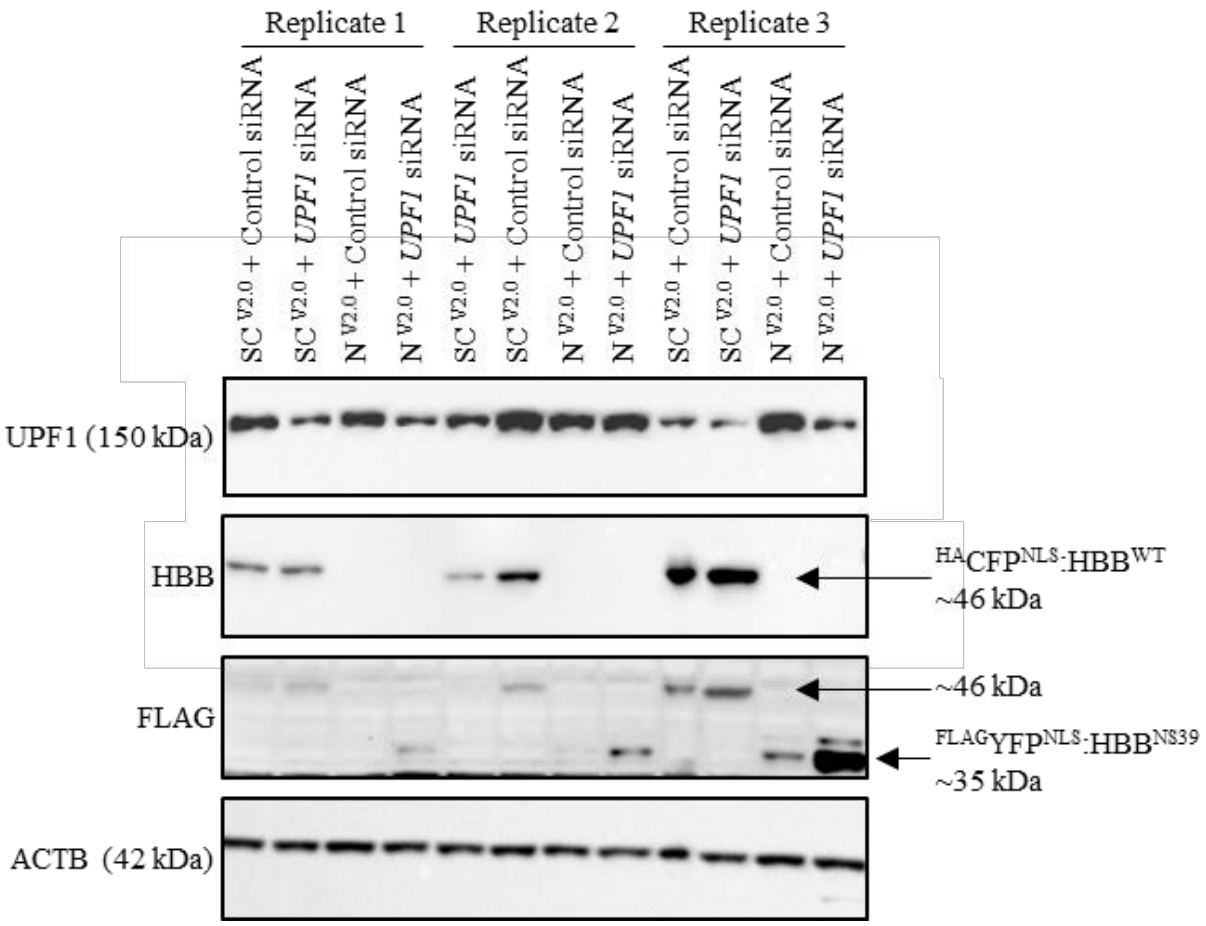


Figure 5.12: Protein expression from NMD$^{V2.0}$ is responsive to a reduction in UPF1 protein levels, whilst protein expression from Control$^{V2.0}$ remains predominantly unchanged following transient transfection in HEK293T cells.

HEK293T cells were co-transfected with expression plasmids encoding either Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$, SC$^{V2.0}$) or NMD$^{V2.0}$ (pUC57-N$^{V2.0}$, N$^{V2.0}$) alongside Control or *UPF1* siRNA. These cells express ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$. Total protein isolated from these cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, HBB (separate blot), FLAG and the loading control ACTB (stripped and re-probed).

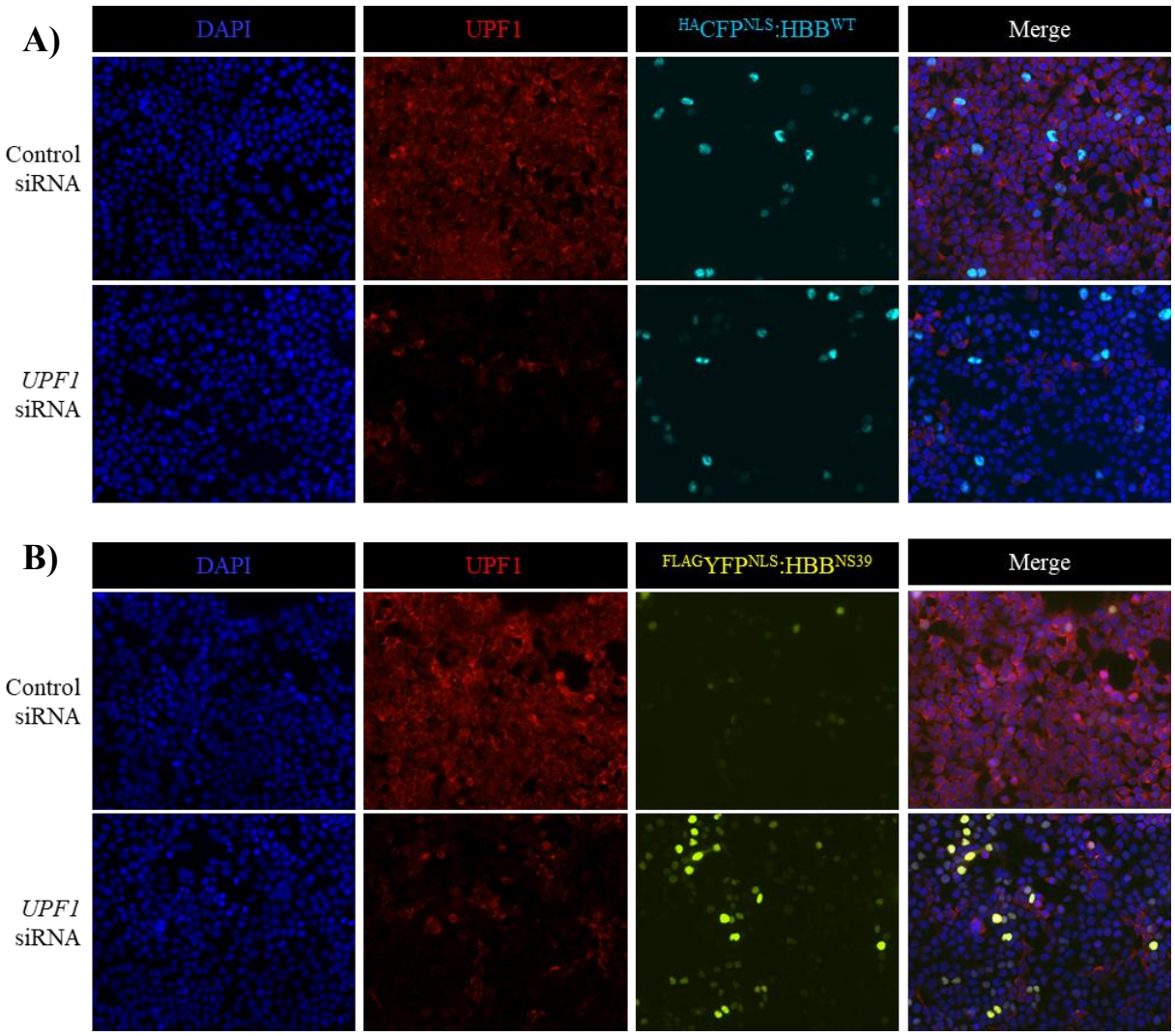


Figure 5.13: Fluorescence expression from Control$^{V2.0}$ is not responsive to a reduction in UPF1 protein levels, whilst fluorescence expressed from NMD$^{V2.0}$ increases following transient transfection in HEK293T cells.

(A) Representative fluorescence microscopy images of HEK293T cells co-transfected with an expression plasmid encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$) and Control or *UPF1* siRNA. These cells express ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ (cyan). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Representative fluorescence microscopy images of HEK293T cells co-transfected with an expression plasmid encoding NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$) and Control or *UPF1* siRNA. These cells express ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ (yellow). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue).

5.3.2.4.2 Engineering pUC57-kan-SCN$^{V2.0}$—an expression plasmid encoding Selection$^{V2.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$

To engineer Transgene$^{V2.0}$ which encodes; Selection$^{V2.0}$, Control$^{V2.0}$, and NMD$^{V2.0}$ in cis, standard recombinant DNA cloning methods were applied. NMD$^{V2.0}$ was excised from its pUC57-kan backbone using flanking *Not*I sites, and then ligated into the *Not*I site of pUC57-kan-SC$^{V2.0}$ (Figure 5.14A). Ligation reactions were then transformed into competent bacterial cells and subject to antibiotic selection with kanamycin. Since only *Not*I was used, it was possible for insertion of NMD$^{V2.0}$ to occur in the reverse orientation. Furthermore, re-ligation of pUC57-kan-SC$^{V2.0}$ would also confer kanamycin resistance. Therefore, following antibiotic selection, successful bacterial transformants could harbour any of three possible ligation products; the desired pUC57-kan-SCN$^{V2.0}$ (forward orientation), pUC57-kan-SCN$^{V2.0}$ (reverse orientation) or pUC57-SC$^{V2.0}$ (vector only) (Figure 5.14A).

To confirm inclusion and orientation of NMD$^{V2.0}$ within pUC57-Kan-SC$^{V2.0}$, plasmid DNA isolated from several successful bacterial transformants was subject to diagnostic restriction endonuclease digests (Figure 5.14B). Colony 1 displayed forward orientation (Figure 5.14B) and was subject to Sanger sequencing (data not shown) to ensure fidelity prior to use in downstream experiments.

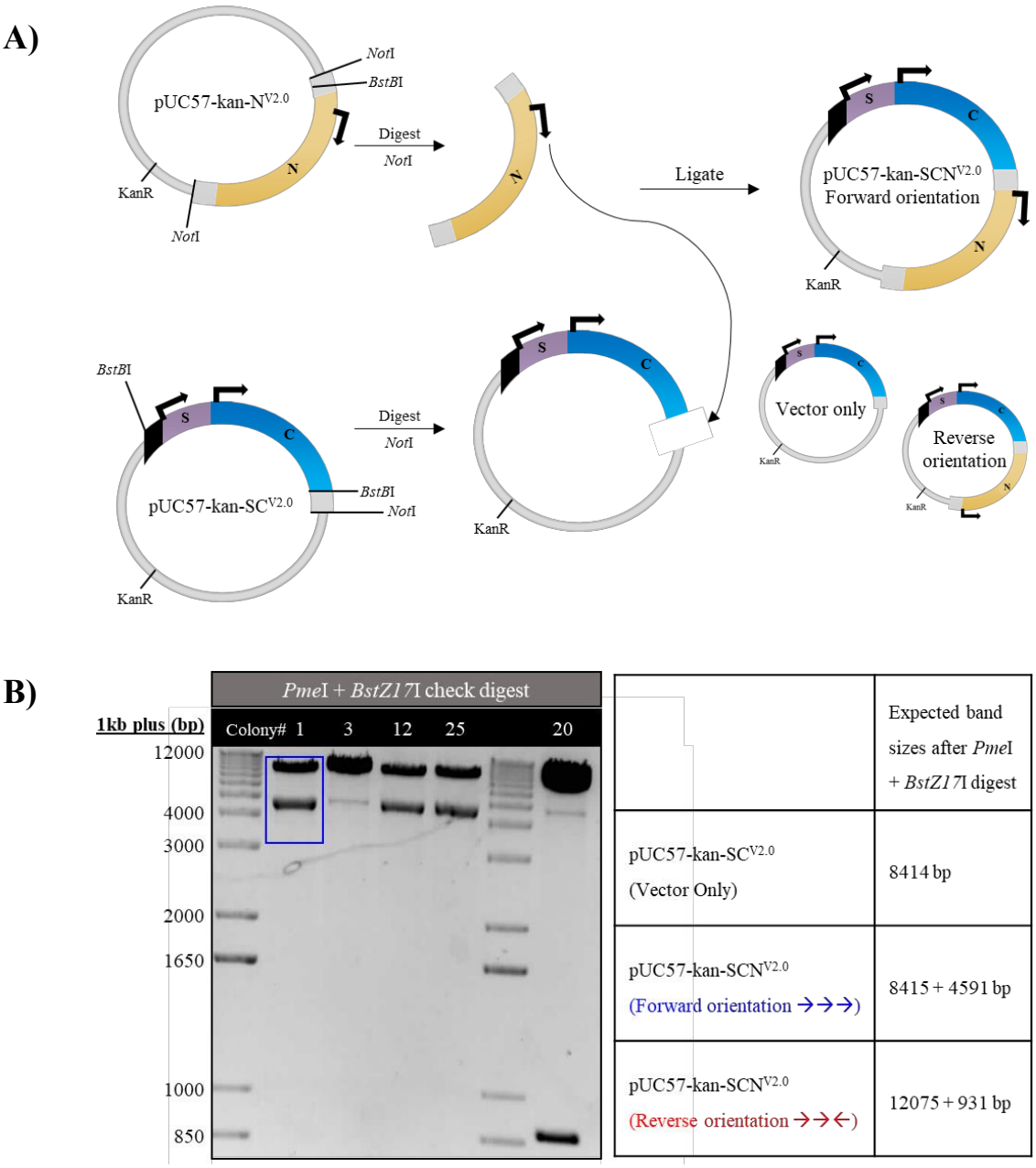


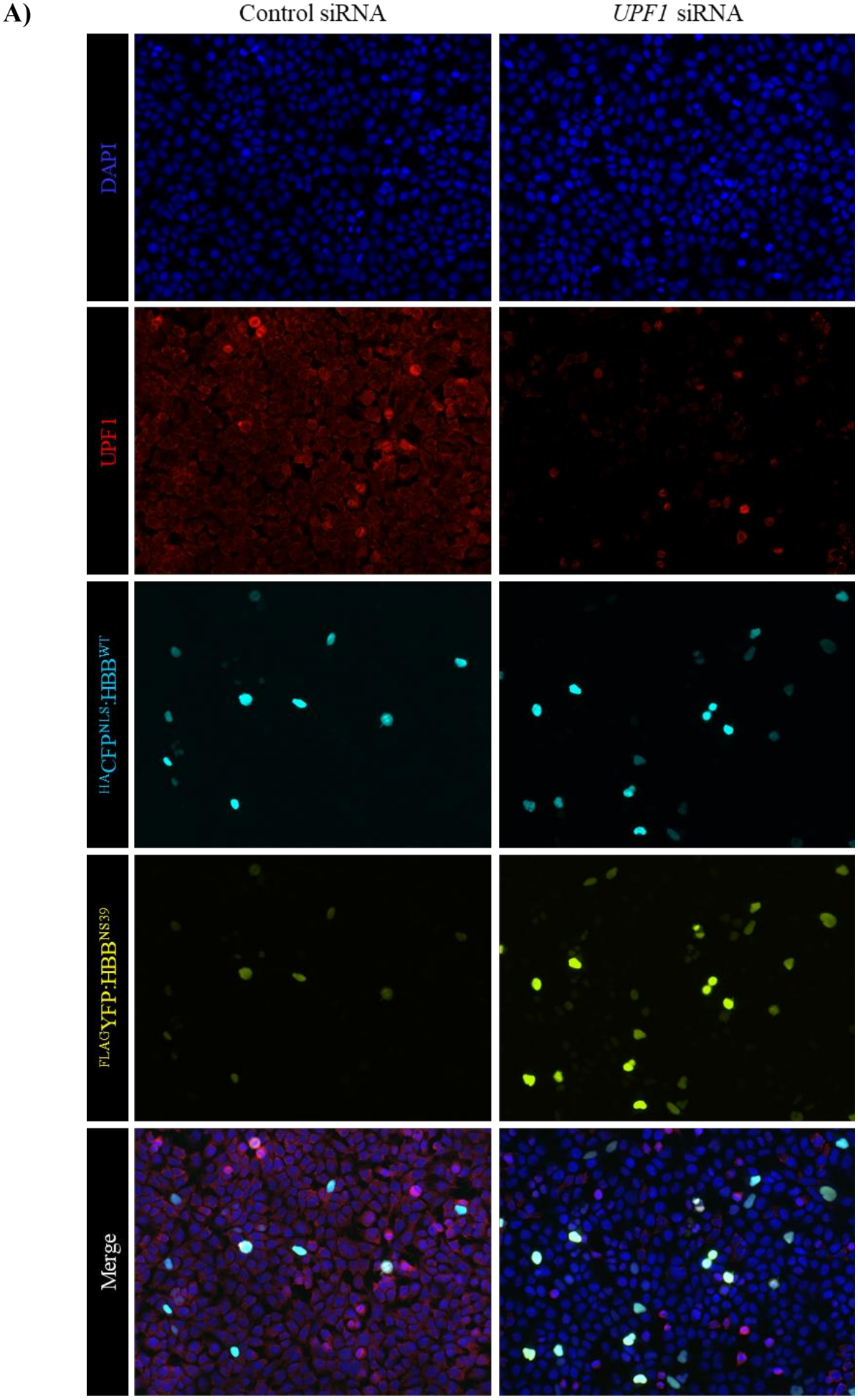
Figure 5.14: Restriction endonuclease digest screening for bacterial colonies containing pUC57-kan-SCN$^{V2.0}$ plasmid DNA.

(A) A Schematic representation of the ligase based cloning method used to engineer pUC57-kan-SCN$^{V2.0}$ which encodes Transgene$^{V2.0}$. **(B)** Following ligation, bacterial transformation, and antibiotic selection, plasmid DNA was isolated from several colonies and digested with the restriction endonucleases *Pme*I and *BstZ17*I to check inclusion and orientation of NMD$^{V2.0}$ within pUC57-kan-SC$^{V2.0}$. Five digested products were visualised by agarose gel electrophoresis. Expected digest product sizes are summarised in the table. Colony 1 (blue box) is an example of a successful ligation containing NMD$^{V2.0}$ inserted in the forward orientation.

5.3.2.4.3 Testing protein expression from Control$^{V2.0}$ and NMD$^{V2.0}$ of pUC57-SCN$^{V2.0}$ in response to an siRNA mediated reduction of UPF1 levels in HEK293T cells

To investigate if Control$^{V2.0}$ and NMD$^{V2.0}$ remain functional when part of a single transgene, the expression plasmid encoding Transgene$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) alongside Control or *UPF1* siRNA was co-transfected into HEK293T cells. Once again, we would expect that following a reduction in UPF1 (i.e. a reduction in NMD activity), protein expression from Control$^{V2.0}$ should remain unchanged whilst protein expression from NMD$^{V2.0}$ should notably increase.

Protein from HEK293T cells co-transfected with pUC57-kan-SCN$^{V2.0}$ and Control or *UPF1* siRNA was visualised through fluorescence microscopy and western blot. This showed that following UPF1 reduction, there was a clear upregulation of ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from NMD$^{V2.0}$, whilst expression of ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ appeared unresponsive (Figure 5.15). These results mirrored expectations. meaning that when pUC57-SCN$^{V2.0}$ is transiently introduced into HEK293T cells, the cassettes of Transgene$^{V2.0}$ appear to respond as expected to a loss of cellular NMD activity.


A)
Control siRNA
UPF1 siRNA
DAPI
UPF1
HACFPNLS:HBBWT
FLAGYFP:HBBNS39
Merge

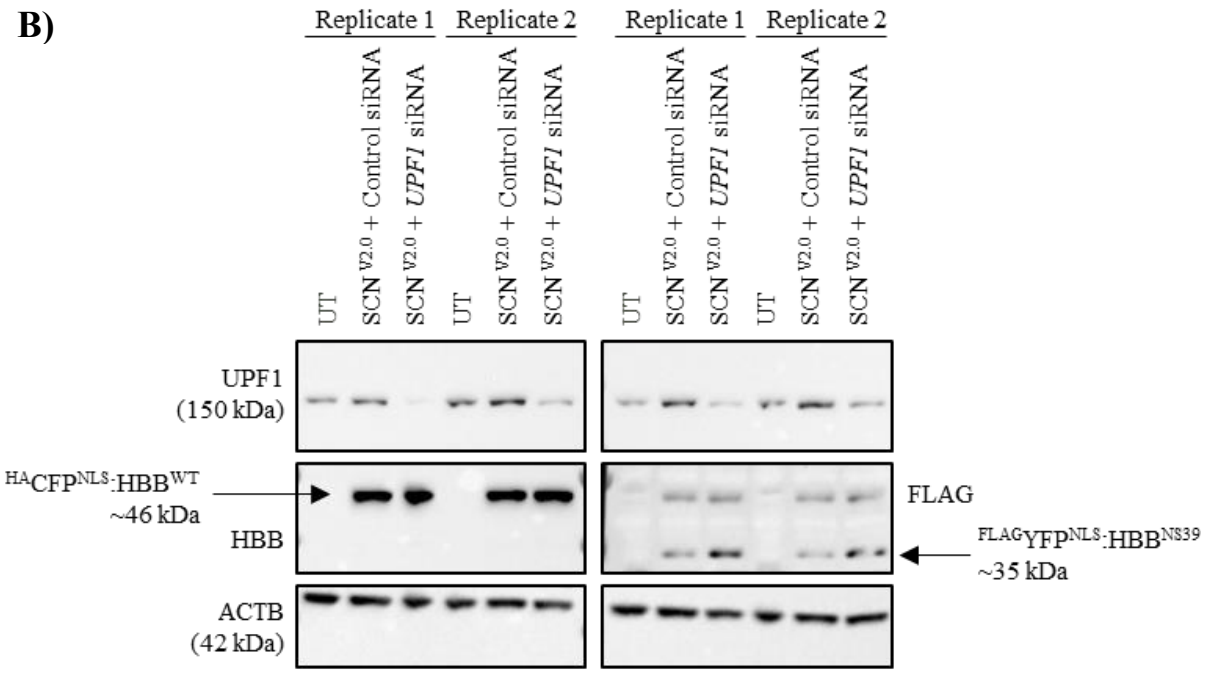


Figure 5.15: Protein expressed from NMD$^{V2.0}$ but not Control$^{V.20}$ (encoded by pUC57-SCN$^{V2.0}$) is increased in response to a reduced level of UPF1 protein following transient transfections in HEK293T cells.

HEK293T cells were co-transfected with an expression plasmid encoding Transgene$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$, SCN$^{V2.0}$) and either Control or *UPF1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express ^{HA}CFP^{NLS}:HBB^{WT} from Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ (yellow). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells, and from untransfected HEK293T cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, HBB, FLAG and the loading control ACTB (stripped and re-probed).

5.3.3 Functional testing of Selection$^{V2.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$ following stable genomic integration of Transgene$^{V2.0}$ into FLP-in mESCs

5.3.3.1 Using RMCE to establish NMD Reporter$^{V2.0}$ mESCs which stably express a single copy of Transgene$^{V2.0}$ from the Col1a1 locus

Following transient transfection-based validation of the functional elements within Transgene$^{V2.0}$ in HEK293T cells, the next step toward a single cell fluorescent NMD reporter transgene is to test the performance of Transgene$^{V2.0}$ in an integrated genomic environment. To integrate Transgene$^{V2.0}$ into the *Col1a1* locus of FLP-in mESCs, the previously described RMCE method was employed (Chapter 4) (Bersten et al. 2015). This method first involves co-electroporation of the expression plasmid encoding Transgene$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) alongside an expression plasmid encoding FLP recombinase (pPGK-FLPo-bpA) into FLP-in mESCs. 48 hours post electroporation, electroporated cells were subject to hygromycin selection (Figure 5.16). Following selection, hygromycin resistant clones which expressed ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ were isolated based on visible expression of CFP and expanded to establish NMD Reporter$^{V2.0}$ mESC lines (Figure 5.16).

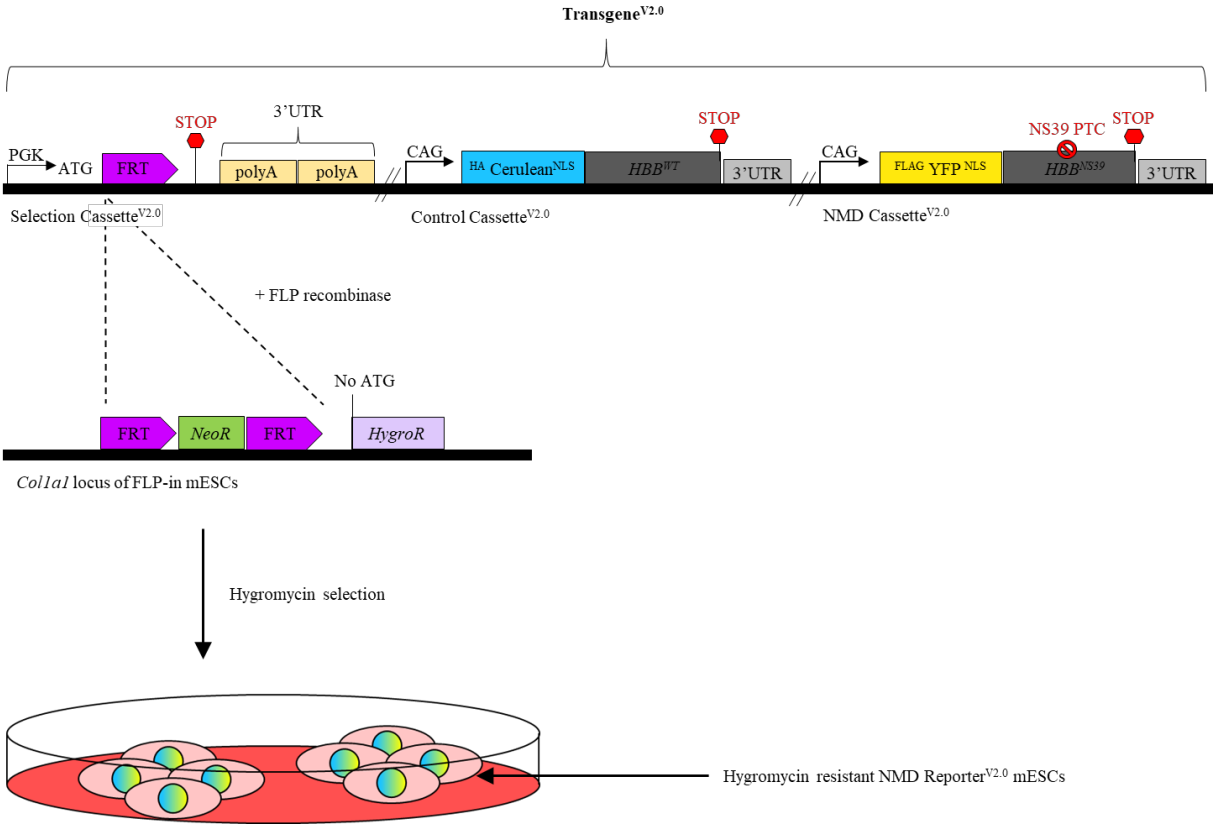


Figure 5.16: Generating an NMD Reporter$^{V2.0}$ mESC line via Selection$^{V2.0}$ mediated RMCE.

A schematic representation of *Col1a1* FLP-RMCE targeting strategy. FLP-in mESCs (mESCs pre-engineered with *Col1a1* FRT sequences and a promoter-less hygromycin resistance coding sequence (*HygroR*), which also lacks a start codon) are co-electroporated with an expression plasmid encoding Transgene$^{V2.0}$ (pUC57-SCN$^{V2.0}$) alongside an expression plasmid encoding FLP recombinase (pPGK-FLPo-bpA). Successful RMCE results in gain of hygromycin resistance and FRT-directed, site specific genomic integration of Control$^{V2.0}$ and NMD$^{V2.0}$ as a single stable copy in all cells. Following hygromycin selection these cells are referred to as NMD Reporter$^{V2.0}$ mESCs.

It was noted that fluorescence expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs was markedly dimmer than that expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs (Figure 5.17). To provide further evidence of correct transgene integration, gDNA isolated from NMD Reporter$^{V2.0}$ mESCs and both FLP-in mESCs and NMD Reporter$^{V1.0}$ mESCs (used as controls) were subject to PCR-based investigations using two previously described PCR primer sets (Bersten et al. 2015). The first primer set was designed to flank the transgene insert site within the targeted *Col1a1* locus of FLP-in mESCs to identify if any insertion had occurred (insertion check). The second primer set was designed to determine the orientation of any insertion into the *Col1a1* locus of FLP-in mESCs (orientation check).

PCR amplified products were visualised by agarose gel electrophoresis. Due to the large size of expected products from the targeted *Col1a1* allele of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs (>15 kb) PCR conditions using 'insertion check' primers could not be optimised to amplify these products (Figure 5.18). Using these primers however, a loss of the original 5 kb targeted allele band which was amplified in FLP-in mESCs was observed for both transgenic mESC lines (Figure 5.18). This suggests that some large insertion within the targeted *Col1a1* locus had occurred. Moreover, 'orientation check' primers showed that in NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs, this insertion had occurred in the correct orientation (Figure 5.18).

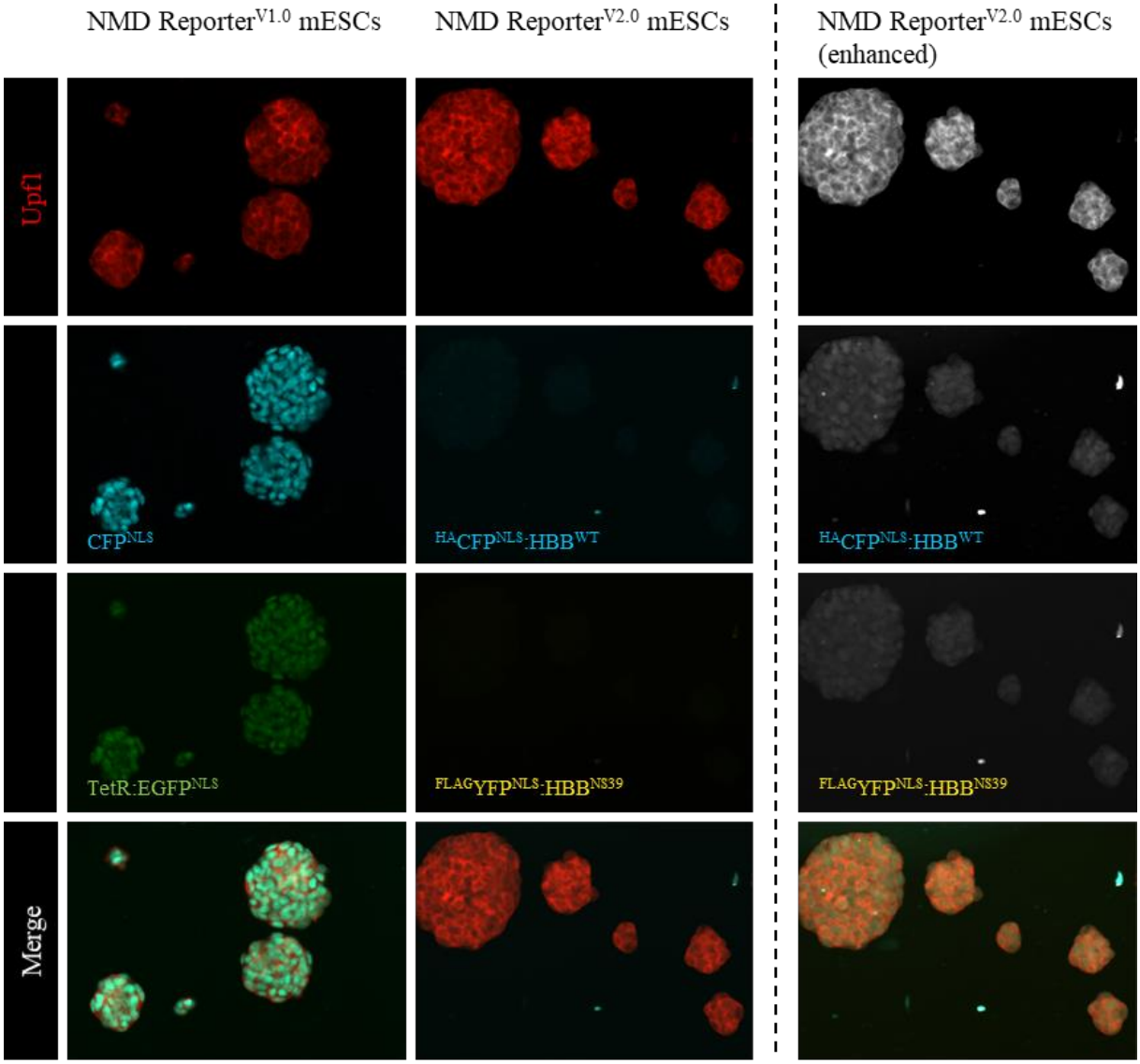
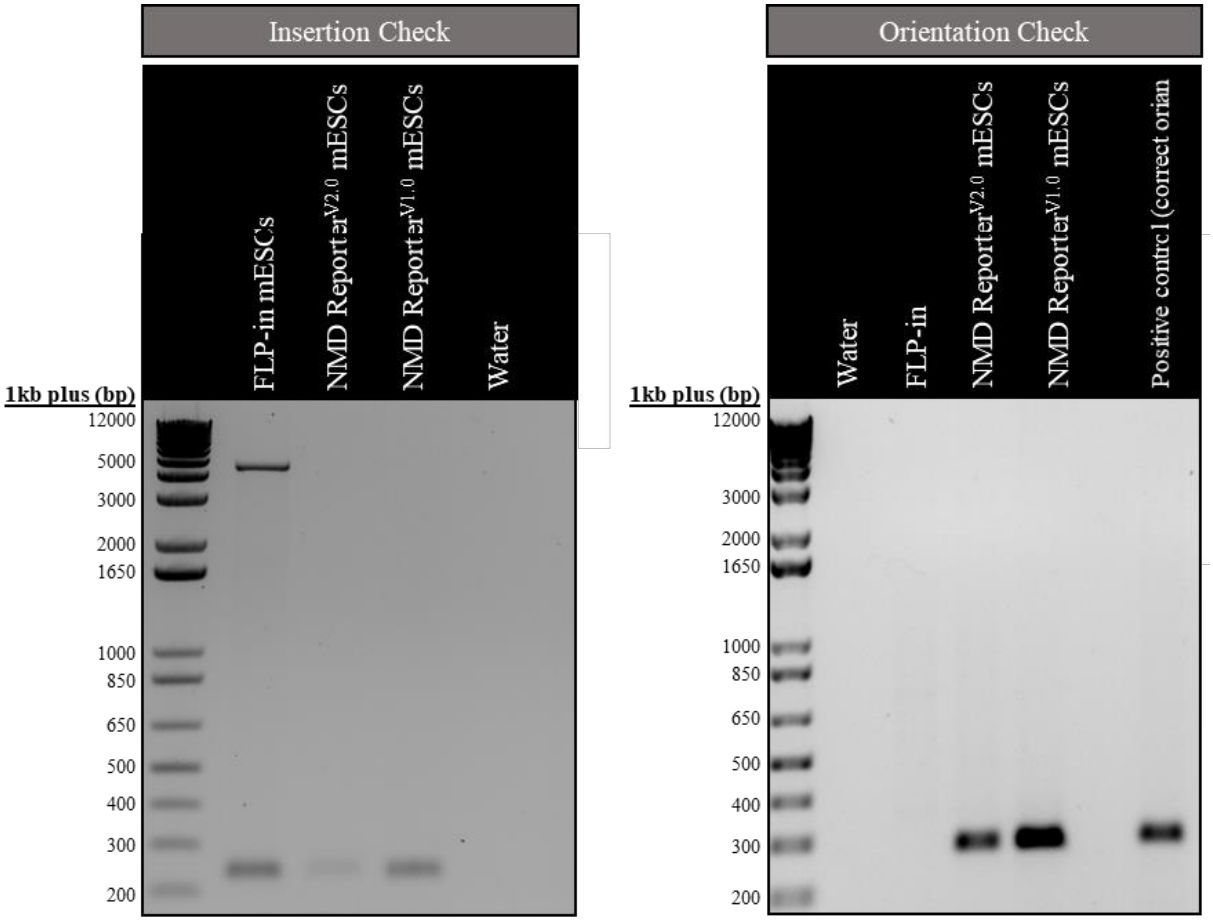


Figure 5.17: Fluorescence expression from NMD Reporter$^{V2.0}$ mESCs is considerably lower than fluorescence expression from NMD Reporter$^{V1.0}$ mESCs.

The left panels show representative fluorescence microscopy images of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs. NMD Reporter$^{V1.0}$ mESCs express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ (green). NMD Reporter$^{V2.0}$ mESCs express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (yellow). Upf1 was detected by immunofluorescence (red). The right panel shows the same fluorescence microscopy images of NMD Reporter$^{V2.0}$ mESCs in grayscale and enhanced only for visual purposes to demonstrate that a low level of fluorescence is present and nuclear localised.



	Insertion check	**Orientation Check**
FLP-in mESCs	WT allele = 240 bp Targeted allele = ~5 kb	No insert, thus no band
NMD Reporter$^{V2.0}$ mESCs	WT allele = 240 bp Targeted allele = ~18 kb	Correct orientation = 300 bp
NMD Reporter$^{V1.0}$ mESCs	WT allele = 240 bp Targeted allele = ~17.2 kb	Correct orientation = 300 bp

Figure 5.18: Genomic PCR of the targeted Col1a1 locus confirms insertion and correct orientation of Transgene$^{V2.0}$ within the Col1a1 locus of NMD Reporter$^{V2.0}$ mESCs.

To provide evidence of correct transgene integration, gDNA isolated from FLP-in mESCs (FLP-in), NMD Reporter$^{V2.0}$ mESCs and NMD Reporter$^{V1.0}$ mESCs was subject to PCR with two previously designed primers (Bersten et al. 2015). The first primer set was designed to flank the transgene insert site within the targeted *Col1a1* locus of FLP-in mESCs to identify if

Fig 5.18 continued…

any insertion had occurred (insertion check). The second primer set was designed to determine the orientation of any insertion into the *Col1a1* locus of FLP-in mESCs (orientation check). PCR amplified products were visualised by agarose gel electrophoresis. The table summarises the expected PCR product sizes. Due to the large size of expected products from the targeted allele of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs (>15 kb) PCR conditions using insertion check primers could not be optimised to amplify these products. Using these primers however, a loss of the original 5 kb targeted allele band which was amplified in the original FLP-in mESCs was observed for both transgenic mESC lines. This suggests that some large insertion within the targeted *Col1a1* locus had occurred. Moreover, Orientation check primers showed that in both cell lines, this insertion had occurred in the correct orientation.

5.3.3.2 *Assessing HBB mini-gene splicing of transcripts expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs*

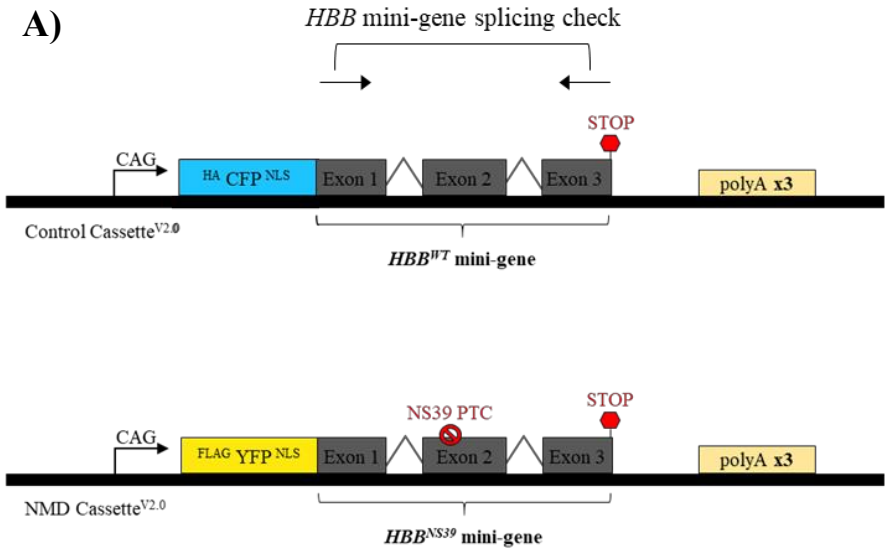
Control$^{V2.0}$ and NMD$^{V2.0}$ both encode fluorescent protein/HBB fusion constructs. In these cassettes, sequence encoding *HBB* contains all three exons and two introns of endogenous *HBB*. This design could introduce unwanted mis-splicing. If most transcripts are in fact mis-spliced, correct expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs could be impaired which may explain the low fluorescence observed from these cells. To investigate this hypothesis, cDNA was first reverse transcribed from RNA of NMD Reporter$^{V2.0}$ mESCs which were transfected with either Control or *Upf1* siRNA. This cDNA was then used as a template in PCR with primers designed to bind within the first and last exon of the *HBB* mini-genes to assess splicing patterns (Figure 5.19A). Since fluorescence was easily detected in HEK293T cells transiently transfected with expression plasmids encoding Control$^{V2.0}$ or NMD$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$), cDNA reverse transcribed from the RNA of these cells was also subject to the same PCR as a control. Correct splicing of the *HBB* mini-gene should produce a single PCR product at the expected size of 244 bp (Figure 5.19A).

PCR product was visualised by agarose gel electrophoresis, which showed that in both HEK293T cells and mESCs the dominant PCR products resembled transcripts produced from correct splicing of *HBB* mini-genes, however, additional PCR products were detected from both cell types (Figure 5.19B). In NMD Reporter$^{V2.0}$ mESCs the production of both unspliced and 'intron 1 retained' transcripts was observed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$. To note, these products were also detected when Control$^{V2.0}$ and NMD$^{V2.0}$ were expressed transiently from HEK293T cells and at a much higher proportion.

Intron 1 retention following mis-splicing of the *HBB* mini-gene is predicted to introduce an NMD targeted PTC into the design of Control$^{V2.0}$ and NMD$^{V2.0}$ (Figure 5.19C). Expression

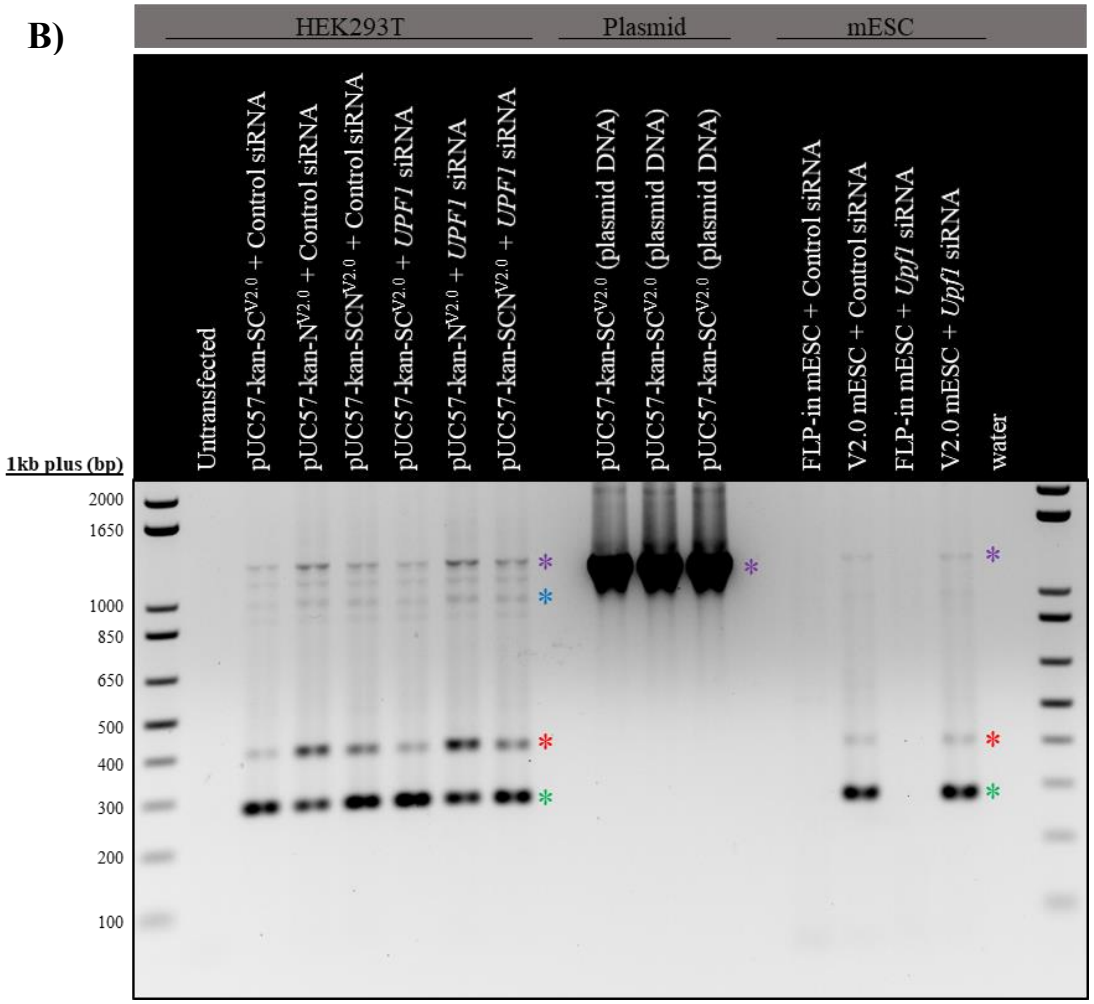
from Control$^{V2.0}$ was intended to be unresponsive to changes in NMD efficiency to act as a direct internal control to the NMD-sensitive NMD$^{V2.0}$. If a significant number of transcripts expressed from Control$^{V2.0}$ retain intron 1 and thus contain a PTC, protein expression from this cassette has the potential to be influenced by cellular NMD activity. Thus far, evidence of this occurring has not been observed in transient transfection-based experiments (Section 5.3.2.4). Moreover, although mis-splicing of the *HBB* mini-gene may bring its own concerns, it is unlikely that this mis-splicing is the sole cause of low fluorescence observed from NMD Reporter$^{V2.0}$ mESCs.

A)



	Unspliced/ plasmid*	Intron 2 retained*	Intron 1 retained*	Exon 2 skipped*	Correct splicing*
HBB mini-gene	1224 bp	1094 bp	373 bp	57 bp	244 bp

B)



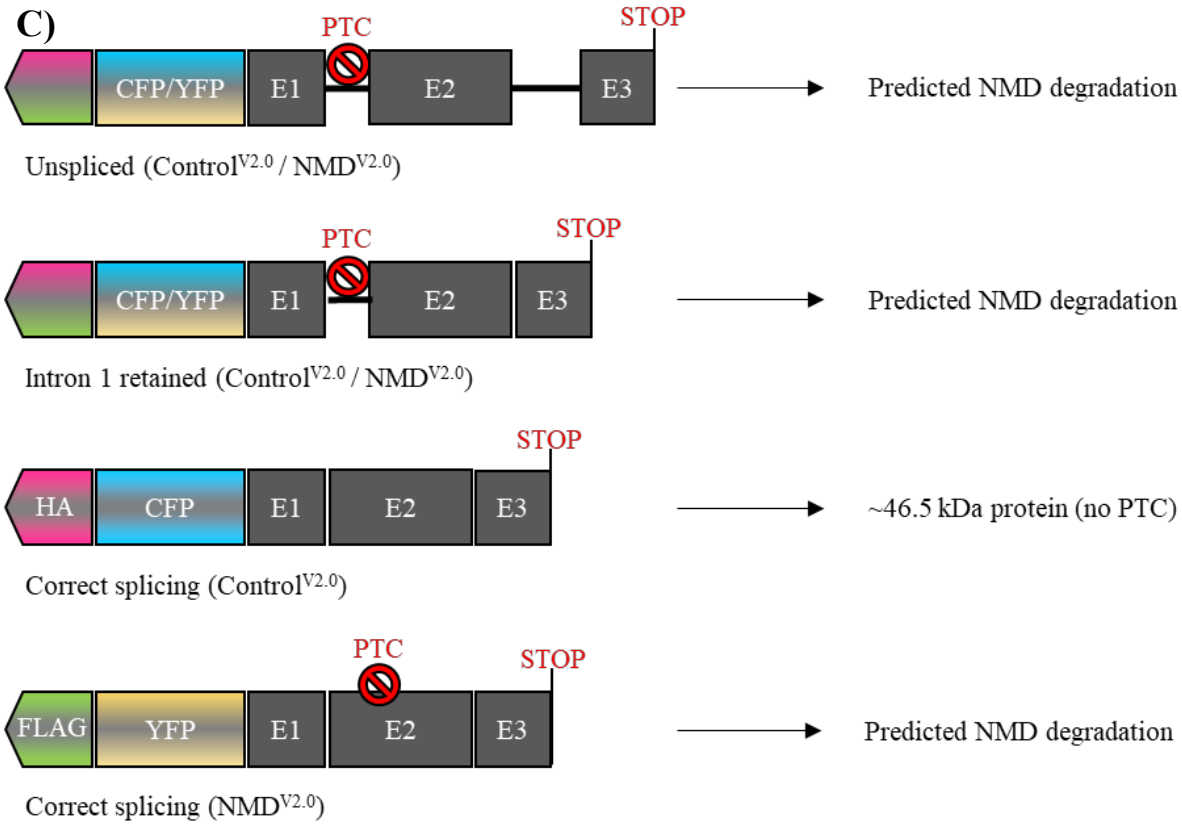


Figure 5.19: Transcripts expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ show low levels of intron retention when expressed transiently from HEK293T cells or from stable NMD Reporter$^{V2.0}$ mESCs.

(A) A schematic representation of primers designed to assess splicing of the *HBB* mini-gene (grey boxes) of Control$^{V2.0}$ and NMD$^{V2.0}$ (*HBB* mini-gene splicing check). The table outlines the PCR product sizes of all potential *HBB* mini-gene splice products when using these primers (these are colour coded in reference to asterisks on the agarose gel image in (B)). **(B)** FLP-in mESCs and NMD Reporter$^{V2.0}$ mESCs (V2.0 mESC) were transfected with Control or *Upf1* siRNA. HEK293T cells were co-transfected with expression plasmids encoding either Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$), NMD$^{V2.0}$ (pUC57-kan-N$^{V2.0}$) or Transgene$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) alongside either Control or *UPF1* siRNA. cDNA reverse transcribed from the RNA of these cells was subject to PCR with '*HBB* mini-gene splicing check' primers and the PCR amplified product was visualised by agarose gel electrophoresis. The coloured asterisks on the gel identify a band as a particular splice product of *HBB*. **(C)** A schematic representation of predicted processing of splice variants detected in NMD Reporter$^{V2.0}$ mESCs.

5.3.3.3 Testing protein expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs in response to an siRNA mediated reduction of Upf1 levels

Mis-spliced transcripts were produced from Control$^{V2.0}$ and NMD$^{V2.0}$ when they were expressed from stable NMD Reporter$^{V2.0}$ mESCs or transiently in HEK293T cells (Section 5.3.3.2). Under transient conditions, the presence of mis-spliced transcripts did not appear to affect the response of Control$^{V2.0}$ or NMD$^{V2.0}$ to changes in cellular NMD activity (Section 5.3.2.4). Therefore, despite low fluorescence expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs, these cassettes could respond as expected to changes in cellular NMD activity. Specifically, fluorescence expression from genomic NMD$^{V2.0}$ may increase to detectable levels following a reduction in NMD efficiency. To investigate if this is true, NMD Reporter$^{V2.0}$ mESCs were transfected with either Control or *Upf1* siRNA, which has been previously shown to reduce Upf1 protein in these cells (Chapter.4).

Protein expression from these cells was visualised by fluorescence microscopy and western blot analysis. Fluorescence microscopy showed that following a reduction in Upf1 there was no striking increase in ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs (Figure 5.20). Moreover detection of ^{FLAG}YFP^{NLS}:HBB^{NS39} required long exposure times to be detected by western blot analysis and when detected, showed no difference in expression levels following Upf1 reduction (Figure 5.21). These results suggest that when expressed from NMD Reporter$^{V2.0}$ mESCs, any response in expression from genomic NMD$^{V2.0}$ to reduced NMD activity is not detectable at the protein level.

To note, western blot analysis also showed that following Upf1 reduction, ^{HA}CFP^{NLS}:HBB^{WT} expression from genomic Control$^{V2.0}$ of NMD reporter$^{V2.0}$ mESCs remained unchanged (Figure 5.21).

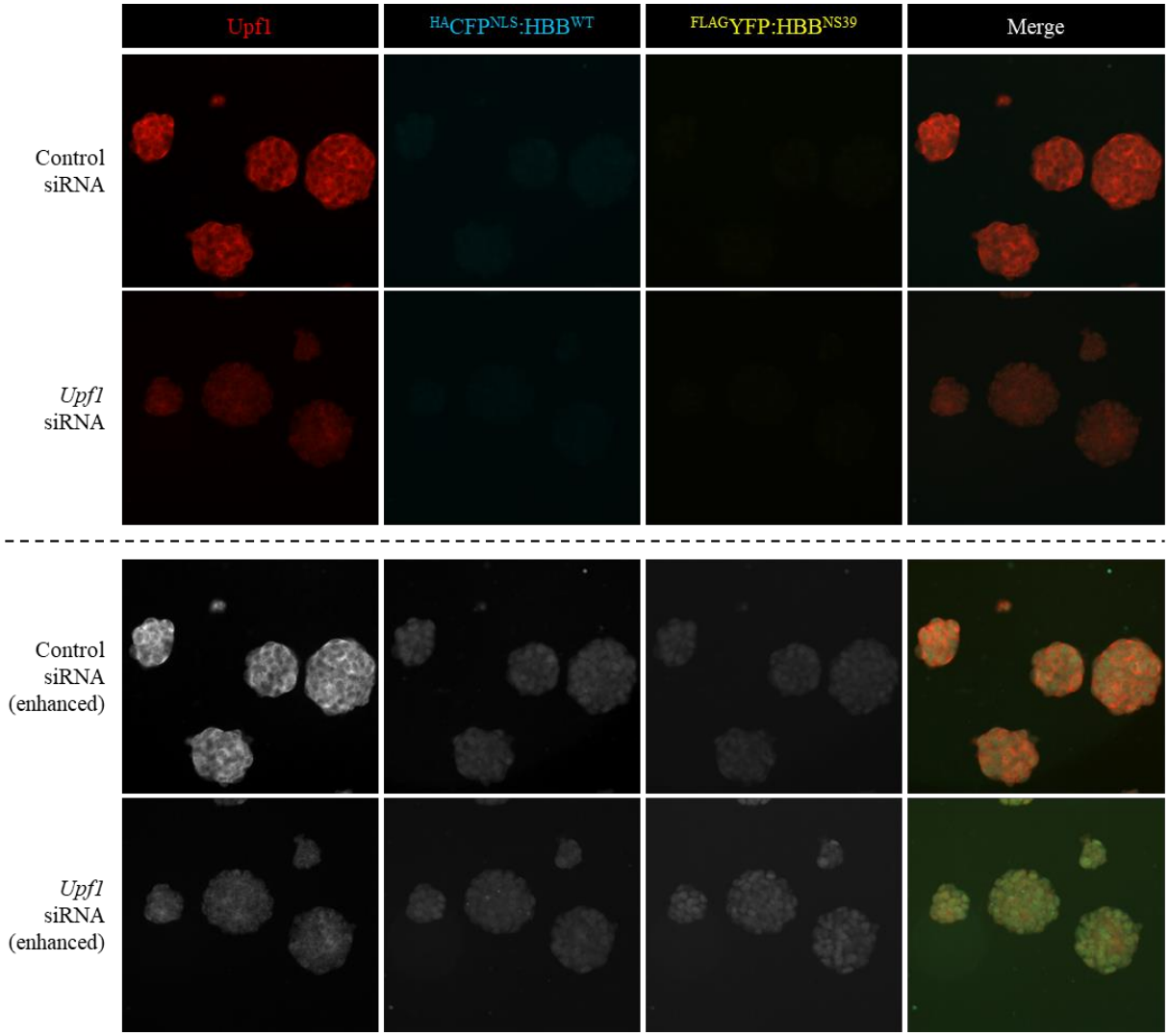


Figure 5.20: Fluorescence expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs remains unchanged following a reduction in Upf1 protein .

The upper panels show representative fluorescence microscopy images of NMD Reporter$^{V2.0}$ mESCs transfected with Control or *Upf1* siRNA. These cells express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (yellow). Upf1 was detected by immunofluorescence (red). The lower panels show the same fluorescence microscopy images in grayscale and enhanced only for visual purposes to demonstrate that low levels of cyan and yellow fluorescence signal is present and nuclear localised.

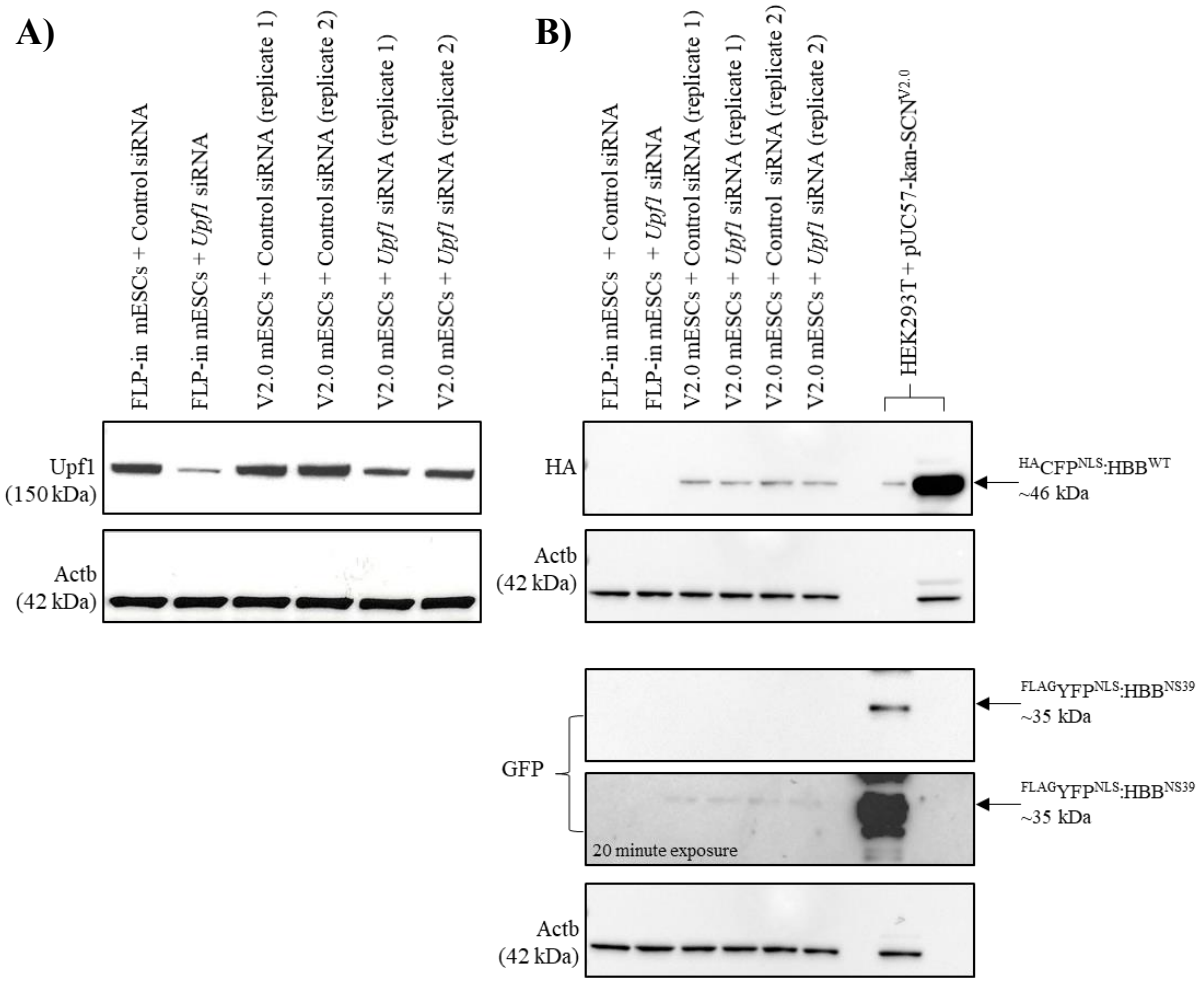


Figure 5.21: Protein expression from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs remains unchanged following a reduction in Upf1 protein.

FLP-in mESCs and NMD Reporter$^{V2.0}$ mESCs (V2.0 mESCs) were transfected with Control or *Upf1* siRNA. NMD Reporter$^{V2.0}$ mESCs express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$. Total protein isolated from these cells was separated by SDS-PAGE and analysed with antibodies to detect; **(A)** Upf1 and the loading control Actb and **(B)** HA, GFP and the loading control Actb (stripped and re-probed).

5.3.3.4 Assessing stability and proteasomal degradation of protein expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs

Another hypothesis that could explain the low levels of fluorescence expression from NMD Reporter$^{V2.0}$ mESCs is the possibility that the fluorescent protein/HBB constructs produced from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of these cells are inherently unstable or targeted by some post-translational degradation pathway.

To investigate this hypothesis, NMD Reporter$^{V2.0}$ mESCs were treated with either cycloheximide or MG132. Since protein expression from NMD Reporter$^{V1.0}$ mESCs was robustly detected by fluorescence microscopy and western blot analysis (Chapter 4) these cells were subjected to the same treatments and used as a positive control. Cycloheximide inhibits translation allowing protein stability to be assayed, while MG132 inhibits the proteasome (protein complexes which can post-transcriptionally target and degrade proteins) therefore stabilising proteins that would otherwise be degraded by the proteasome. As such, an unstable protein should rapidly decrease in abundance following treatment with cycloheximide, whilst a protein targeted for proteasomal degradation should rapidly increase in abundance following treatment with MG132. If these treatments are successful, expression of the endogenous protein β-catenin (a known target of rapid degradation by the proteasome) will be seen to decrease following cycloheximide treatment and increase following MG132 treatment.

Proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs are predicted to be both stable and not targeted for proteasomal degradation. If this is true protein expression from Control$^{V1.0}$ and NMD$^{V1.0}$ will remain unchanged following cycloheximide or MG132 treatment of NMD Reporter$^{V1.0}$ mESCs. Following these treatments, protein isolated

from NMD Reporter$^{V1.0}$ mESCs was visualised by western blot, the results of which confirmed expression of stable proteins from genomic Control$^{V1.0}$ and NMD$^{V1.0}$(Figure 5.22A).

Comparatively western blot analysis of protein isolated from treated NMD Reporter$^{V2.0}$ mESCs, showed that expression of ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ was rapidly decreased following cycloheximide treatment and rapidly increased following MG132 (Figure 5.22B). These results are typical of proteins that are targeted for efficient proteasomal degradation rendering them unstable. This finding explains the low levels of fluorescence observed from NMD Reporter$^{V2.0}$ mESCs. Moreover, proteasomal degradation of ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs has the potential to mask (at the protein level) any response of transcripts expressed from NMD$^{V2.0}$ to changes in cellular NMD activity and thus may explain why ^{FLAG}YFP^{NLS}:HBB^{NS39} was not seen to increase following a loss of Upf1 in NMD Reporter$^{V2.0}$ mESCs (Section 5.3.3.3).

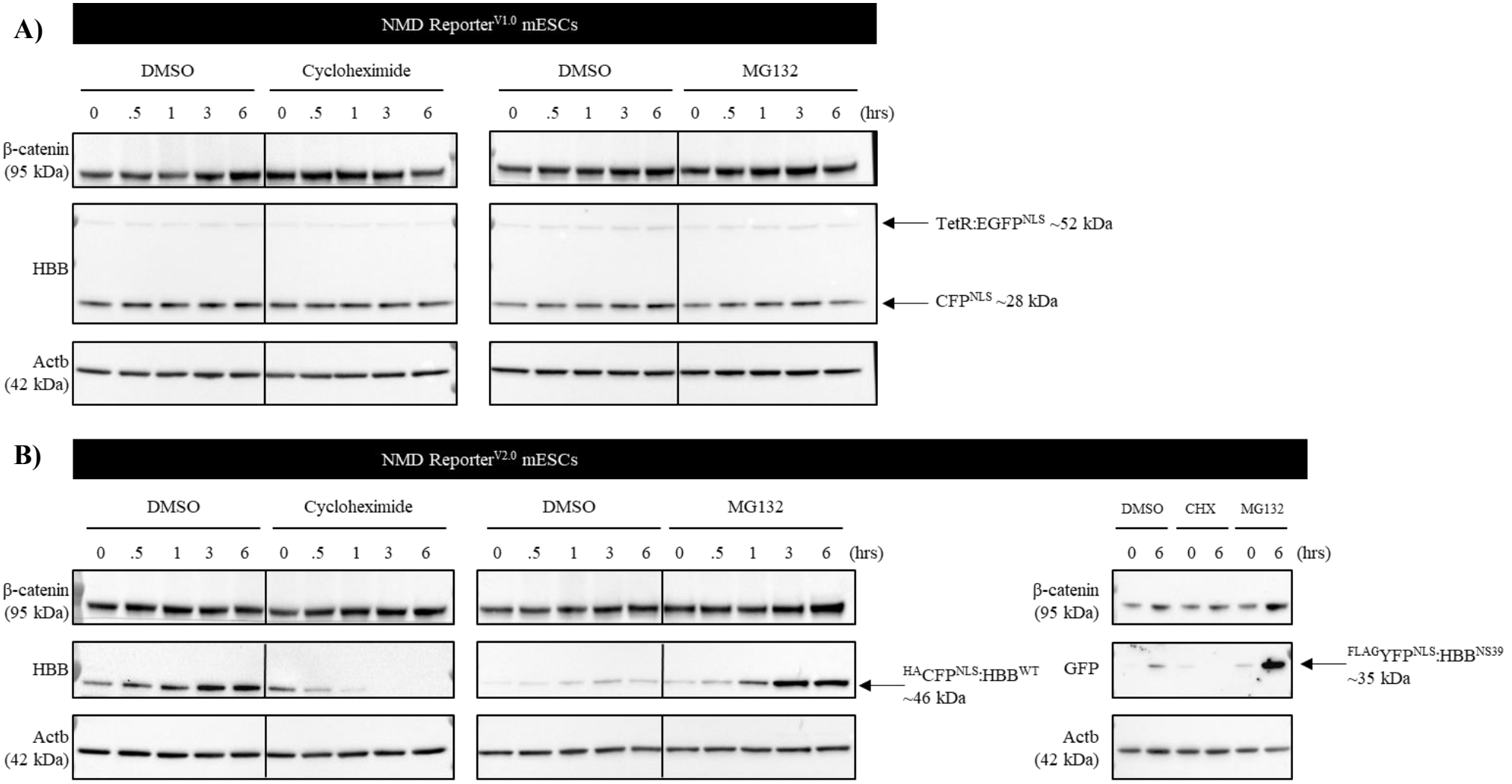

A)
NMD Reporter V1.0 mESCs
DMSO
Cycloheximide
DMSO
MG132
0 .5 1 3 6 0 .5 1 3 6 0 .5 1 3 6 0 .5 1 3 6 (hrs)
β-catenin (95 kDa)
HBB
Actb (42 kDa)
TetR:EGFP NLS ~52 kDa
CFP NLS ~28 kDa
B)
NMD Reporter V2.0 mESCs
DMSO
Cycloheximide
DMSO
MG132
0 .5 1 3 6 0 .5 1 3 6 0 .5 1 3 6 0 .5 1 3 6 (hrs)
β-catenin (95 kDa)
HBB
Actb (42 kDa)
HA CFP NLS :HBB WT ~46 kDa
DMSO
CHX
MG132
0 6 0 6 0 6 (hrs)
β-catenin (95 kDa)
GFP
Actb (42 kDa)
FLAG YFP NLS :HBB NS39 ~35 kDa

Figure 5.22: Proteins expressed from genomic Control$^{V2.0}$ and NMD $^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs are unstable due to degradation by the proteosome.

NMD Reporter$^{V2.0}$ mESCs were subject to treatment with either cycloheximide (CHX) or MG132 to inhibit translation or the proteasome, respectively. The same treatments of NMD Reporter$^{V1.0}$ mESCs served as a positive control. NMD Reporter$^{V1.0}$ mESCs express CFP^{NLS} from genomic Control$^{V1.0}$ and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$. NMD Reporter$^{V2.0}$ mESCs express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$. Following treatment, total protein was isolated from these cells at different time points, separated by SDS-PAGE and analysed by western blot using antibodies to detect to βcatenin, HBB, GFP and the loading control Actb (stripped and re-probed). **(A)** Proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs were seen to be stable and unaffected by inhibition of the proteasome. **(B)** Proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs were observed to be unstable due to rapid degradation by the proteasome.

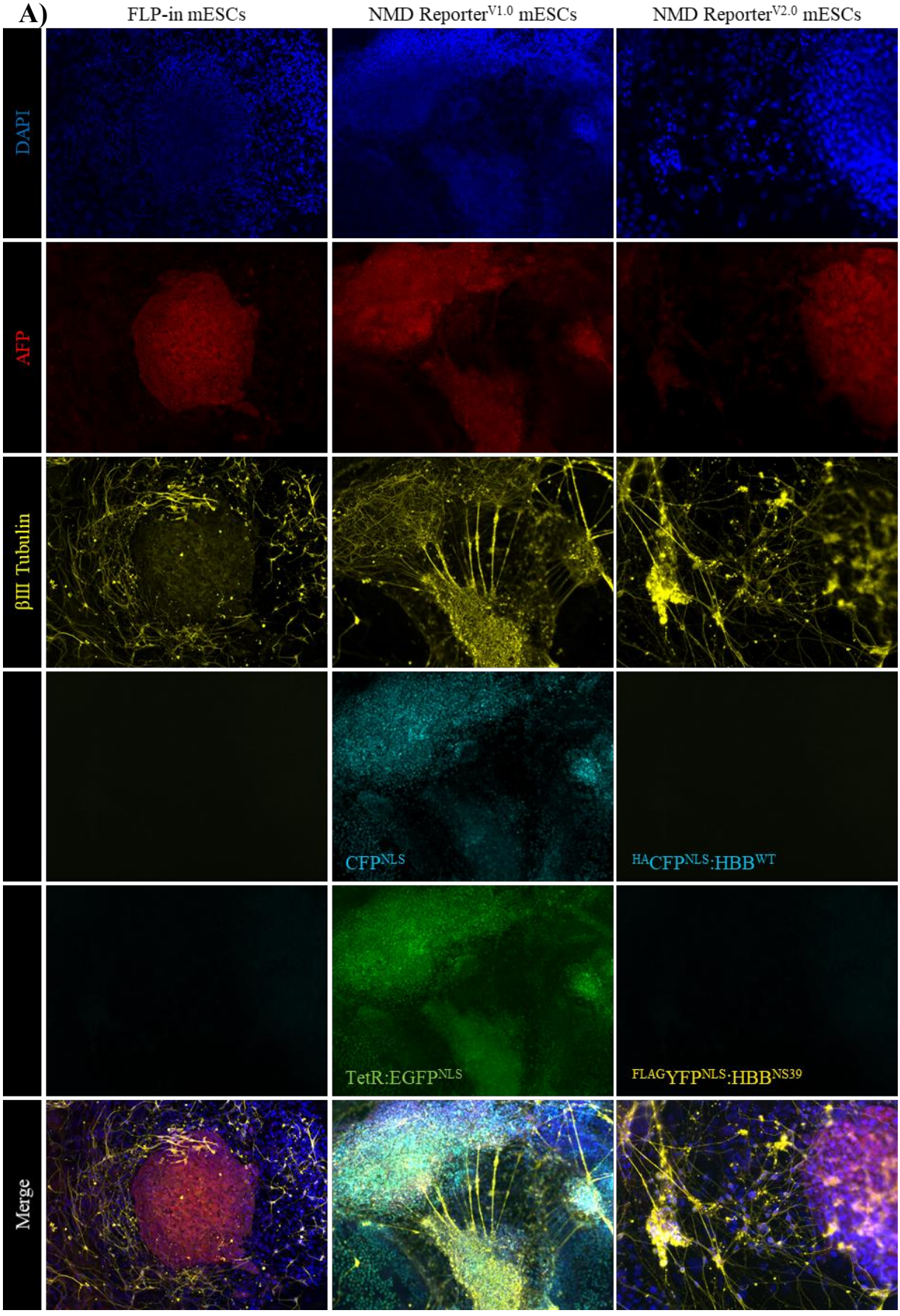
5.3.3.5 *Investigating if proteasomal degradation of proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ is restricted to mESCs via differentiation of NMD Reporter$^{V2.0}$ mESCs.*

The proteasome is composed of selective degradation machinery which targets proteins that have been covalently labelled with the small protein ubiquitin. The upstream activity of ubiquitin ligases as well as the proteasomes are variable and can be differentially localised in the nucleus or cytoplasm of eukaryotic cells, as such, a wide variety of proteins are targeted by the proteasome. The altered activity and intracellular distribution of ubiquitin ligase and proteasomes depends on the cell or tissue type (Grigoreva et al. 2015). It therefore remained a possibility that although proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ were degraded in NMD Reporter$^{V2.0}$ mESCs, they may not encounter the same fate in other cell types.

To determine if proteasomal degradation of ^{HA}CFP^{NLS}:HBB^{WT} expressed from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs is restricted to mESCs, a spontaneous differentiation of these cells into the three germ layers (mesoderm, endoderm and ectoderm) was performed. The same protocol was performed on NMD Reporter$^{V1.0}$ mESCs and FLP-in mESCs as positive controls. If the observed proteasomal degradation is restricted to mESCs, it is expected that following differentiation, fluorescence expressed from the genomic cassettes of NMD Reporter$^{V2.0}$ mESCs will be restored to levels similar to that from NMD Reporter$^{V1.0}$ mESCs as proteins expressed from these cells was observed to be stable and not targeted by the proteasome (Section 5.3.3.4).

Following differentiation, cells were visualised by immunofluorescence analysis. All mESC lines successfully differentiated into the three germ layers as seen by expression of germline specific markers (Figure 5.23). Following differentiation, fluorescence expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs was not elevated (Figure 5.23).

This suggests that $^{HA}CFP^{NLS}$:HBB^{WT} and $^{FLAG}YFP^{NLS}$:HBB^{NS39} expressed from NMD Reporter$^{V2.0}$ mESCs are likely targeted by the proteasome across many different cell and tissue types, making this an impractical NMD reporter system for use in mESCs.



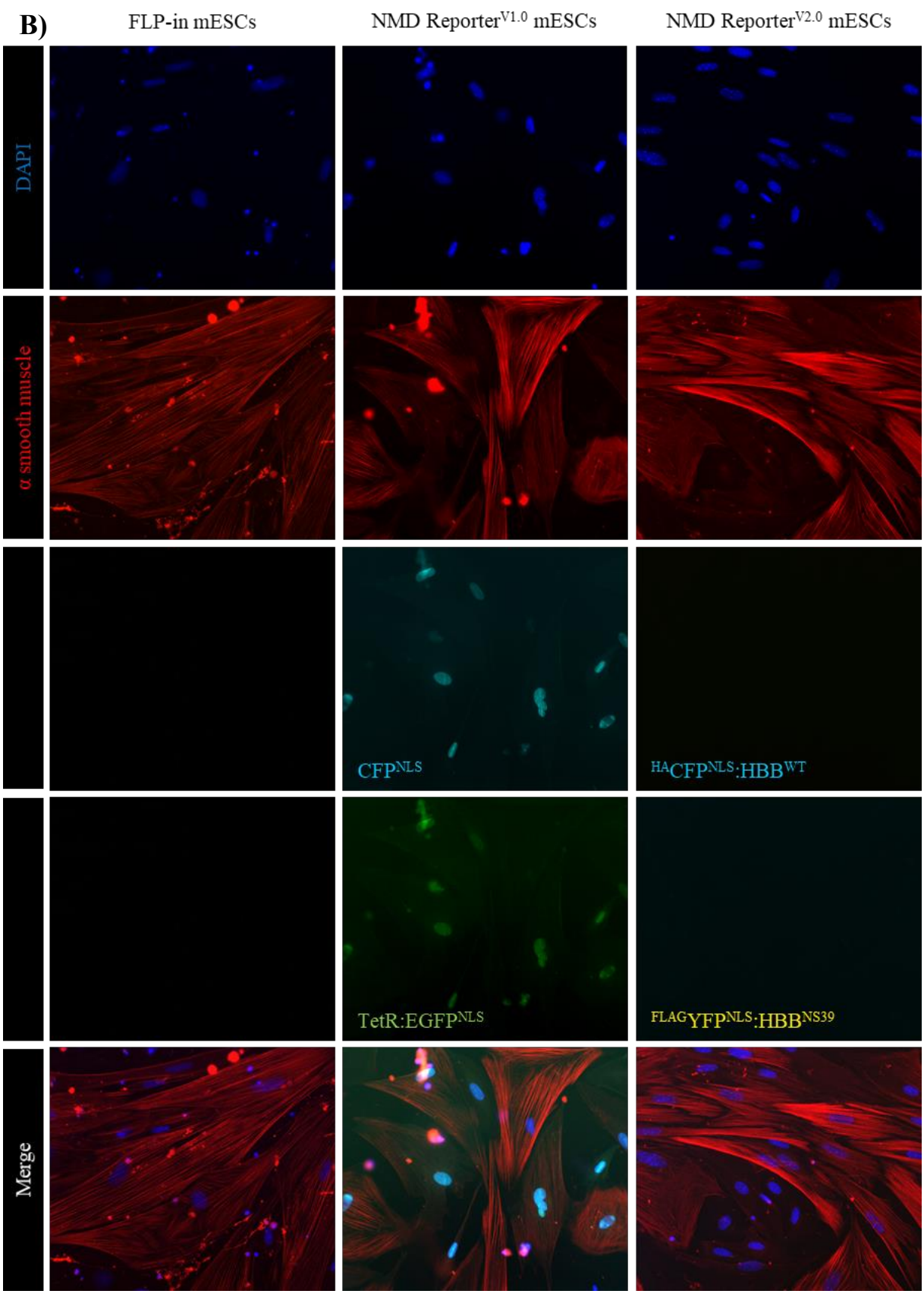

B)
FLP-in mESCs
NMD Reporter V1.0 mESCs
NMD Reporter V2.0 mESCs
DAPI
α smooth muscle
CFP NLS
HA CFP NLS :HBB WT
TetR:EGFP NLS
FLAG YFP NLS :HBB NS39
Merge

Figure 5.23: Spontaneous differentiation of NMD Reporter$^{V2.0}$ mESCs reveals that proteosomal degradation of proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ are not restricted to mESCs.

Representative fluorescence microscopy images of FLP-in mESCs, NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs which were allowed to spontaneously differentiate as embryoid bodies in suspension followed by adherent culture for 14 days. NMD Reporter$^{V1.0}$ mESCs express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ (green). NMD Reporter$^{V2.0}$ mESCs express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (yellow). **(A)** Alpha -Fetoprotein (AFP), a marker of the endoderm (red) and βIII tubulin, a marker of the ectoderm (yellow) were detected by immunofluorescence. Cell nuclei were counterstained for DAPI (blue). **(B)** α smooth muscle, a marker of the mesoderm (red) was detected by immunofluorescence. Cell nuclei were counterstained for DAPI (blue).

5.3.4 Design and functional testing of Control$^{V2.1}$, Control $^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ using expression plasmids in a transient setting

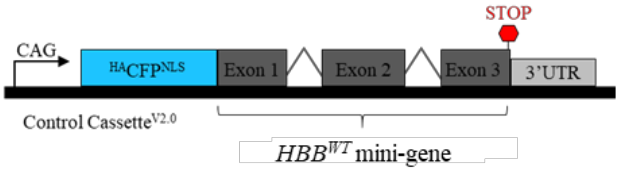
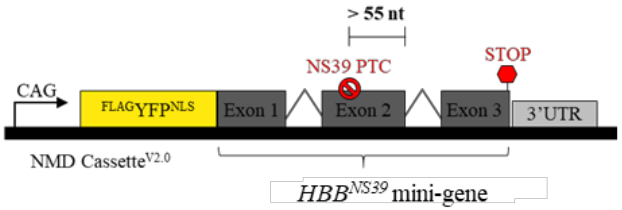
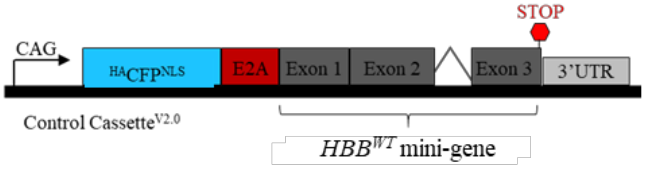
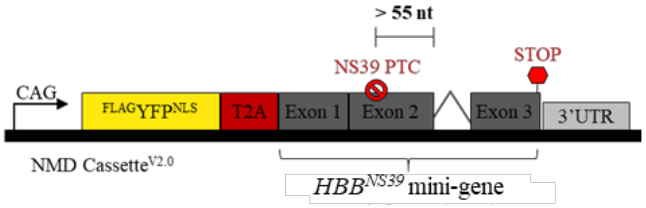
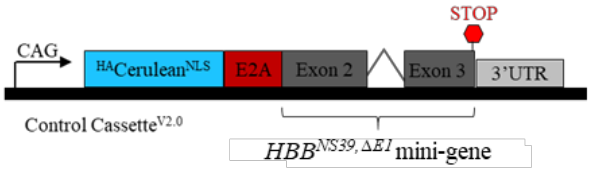
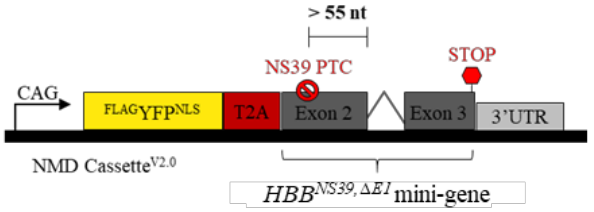
5.3.4.1 A detailed description of the design and processing of Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$

To summarise the findings so far, Control$^{V2.0}$ and NMD$^{V2.0}$ were observed to respond as expected to a reduction in NMD when transiently expressed in HEK293T cells. However, when these cassettes were expressed from stable NMD Reporter$^{V2.0}$ mESCs, fluorescence expression was considerably low. Further investigations revealed that protein expressed from genomic Control$^{V2.0}$ or NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs was targeted for proteasomal degradation. This degradation was not restricted to mESCs and observed in the cells of all three germ layers following differentiation of NMD Reporter$^{V2.0}$ mESCs. The end goal was to engineer a fluorescent NMD reporter that allows single cell resolution so that NMD can be tracked across cellular development pathways or to establish an NMD reporter mouse. This cannot be achieved if proteins encoded by the transgene itself are subject to proteasomal degradation in cells.

Additionally, mis-splicing of the *HBB* mini-gene of Control$^{V2.0}$ and NMD$^{V2.0}$ was observed in transient transfection-based HEK293T cell experiments and in experiments using stable NMD Reporter$^{V2.0}$ mESCs. This mis-splicing resulted in production of 'intron 1 retained' and unspliced transcripts, which are predicted to trigger NMD. Although this did not appear to affect the function of Control$^{V2.0}$ and NMD$^{V2.0}$ when expressed transiently in HEK293T cells, it could cause complications since transcripts predicted to trigger NMD are being produced from Control$^{V2.0}$, the expression of which should remain unchanged in response to NMD activity.

To combat mis-splicing and proteasomal degradation, two updated versions (Version 2.1 and Version 2.2.) of Control$^{V2.0}$ and NMD$^{V2.0}$ were designed. These modified cassettes are referred to as Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ and are summarised in Table 5.4.

Table 5.4: A summary of the modifications designed for Control$^{V2.0}$ (i.e. Control$^{V2.1}$ and Control$^{V2.2}$) and NMD$^{V2.0}$ (i.e. NMD$^{V2.1}$ and NMD$^{V2.2}$)

Version	Control Cassette	NMD Cassette	Modifications
V2.0			
V2.1			- Removal of *HBB* intron 1 - Addition of E2A or T2A self cleavage sequence
V2.2			- Removal of *HBB* exon 1 (Δ E1) and *HBB* intron 1 - Addition of E2A or T2A self cleavage sequence

Proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs were not degraded by the proteasome. The main difference between these proteins, and those expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ is the presence of an in-frame fusion of the translated *HBB* mini-gene with a fluorescent protein. This *HBB* mini-gene encodes all three exons and two introns of endogenous *HBB*. It was reasoned that expression of this protein outside of its natural cellular environment (i.e. red blood cells) could be a trigger for proteasomal degradation. To address this, Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ were updated from either Control$^{V2.0}$ and NMD$^{V2.0}$ to include sequences for a self-cleaving 2A peptide sequence which would intervene the fluorescent protein and *HBB* translation product.

In the case of Control$^{V2.1}$ and Control$^{V2.2}$ the E2A (equine rhinitis A virus) self-cleavage peptide sequence was used. Whilst for NMD$^{V2.1}$ and NMD$^{V2.2}$ the T2A (Thosea asigna virus 2A) self-cleavage peptide sequence was used (Table 5.4). These self-cleavage 2A sequences function as a 'ribosomal-skip mechanism' which impairs normal peptide bond formation between the last glycine and proline that they encode without impacting normal translation of the downstream sequences (Szymczak et al. 2004). This will allow the fluorescent protein expressed from any of the cassettes to be separated from (instead of fused to) HBB protein expressed from the same cassette. By physically dissociating the production of HBB and fluorescent proteins, the possibility of any post translational degradation of HBB impacting the fluorescent NMD read out of a cassette should be removed without affecting NMD recognition and degradation of PTC containing transcripts during the pioneer round of translation (Figure 5.24).

To minimise the formation of mis-spliced variants, intron 1 of the *HBB* mini-gene was also excluded in all updated cassettes (Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$) (Table 5.4). Intron 2 was maintained in these cassettes to ensure that transcripts produced would be subject to splicing and subsequent EJC deposition necessary for PTC recognition by NMD.

Additionally, the design of Control$^{V2.2}$ and NMD$^{V2.2}$ also excluded *HBB* exon 1 (and 1 bp of *HBB* exon 2 to maintain reading frame of the *HBB* mini-gene) (Table 5.4). This was done as the exclusion of *HBB* exon 1 may further minimise issues with mis-splicing and by not translating full-length, functional HBB, the translated protein may be less likely to be targeted for proteasomal degradation.

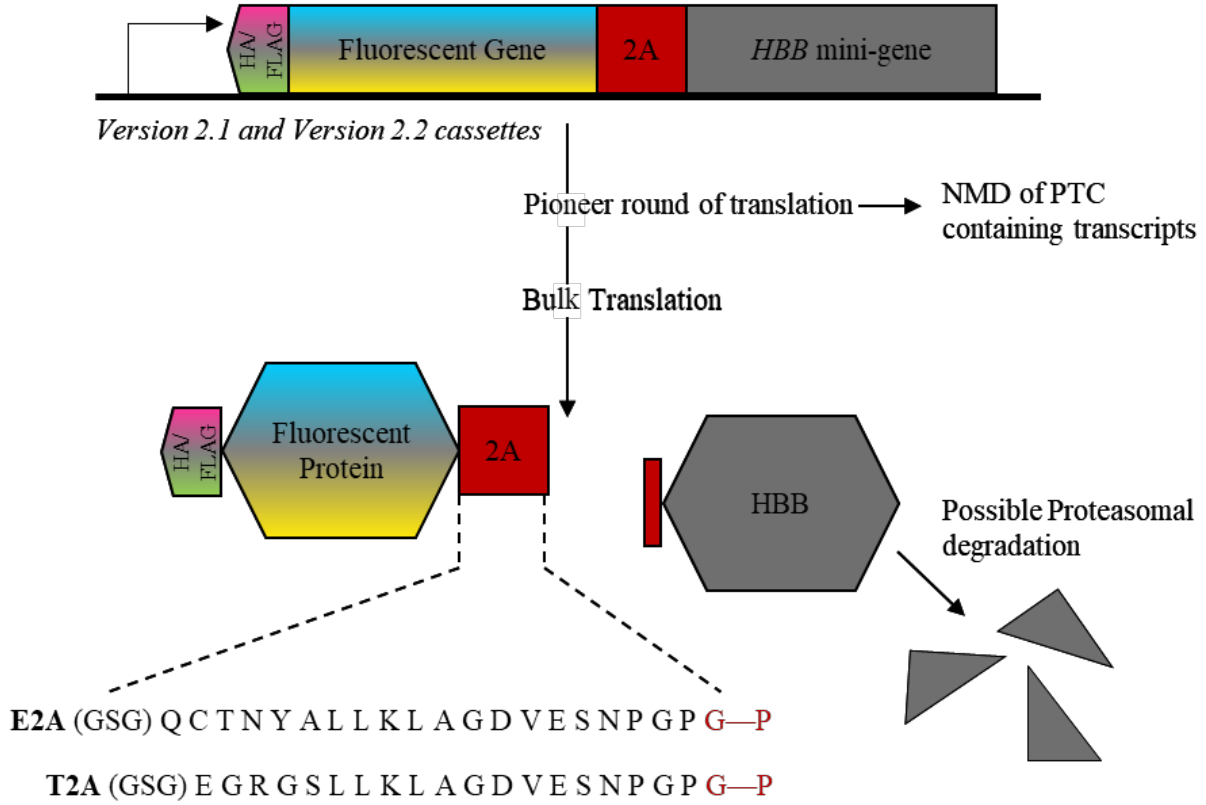


Figure 5.24: A schematic representation of the basic design of Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ and how a 2A self-cleavage sequence influences protein expression.

The design of all Version 2.1 or Version 2.2 cassettes include sequences encoding either an E2A (Control$^{V2.1}$ and Conrol$^{V2.2}$) or T2A (NMD$^{V2.1}$ and NMD$^{V2.2}$) self-cleavage peptide sequence. Following translation, cleavage will occur at the last glycyl-prolyl (G—P) peptide bond encoded by the 2A self-cleavage sequence. This will result in separation of the fluorescent protein from HBB protein encoded by the same cassette. This will not affect NMD of PTC containing transcripts during the pioneer round of translation. Since no longer fused, any downstream processing of the HBB protein, such as proteasomal degradation should not impact the stability and processing of the fluorescent protein.

5.3.4.2 *Engineering pUC57-kan-SC$^{V2.1}$, pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ and pUC57-kan-N$^{V2.2}$—expression plasmids encoding Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ respectively*

To introduce the discussed modifications into Control$^{V2.0}$ and NMD$^{V2.0}$, DNA sequence spanning from the end of the fluorescent protein to the end of the *HBB* mini-gene was re-designed and synthesised into the pUC57-kan plasmid backbone through the services of Genscript (pUC57-kan-SCmod$^{2.1}$, pUC57-kan-SCmod$^{2.2}$, pUC57-kan-Nmod$^{2.1}$ and pUC57-kan-Nmod$^{2.1}$) (Table 5.5). The re-designed regions contain overlapping restriction endonuclease sites to the original design allowing changes to be introduced as modifications into the expression plasmids encoding Control$^{V2.0}$ or NMD$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$) via standard recombinant DNA cloning methods.

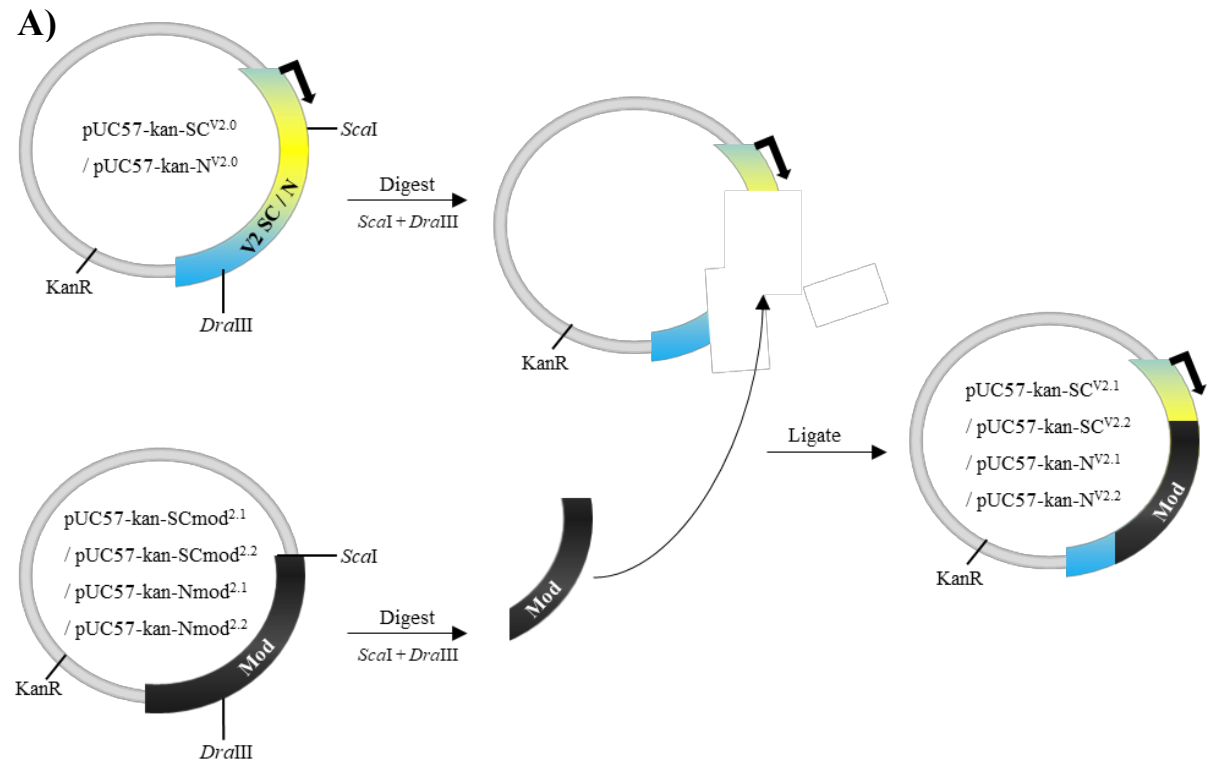
All modifications were introduced into pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$ using the following ligase-based method. *Sca*I and *Dra*III sites flanking the re-designed sequence of all plasmids was used to exchange the original sequences from pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$ with sequences from the desired modification encoding plasmids (pUC57-kan-SCmod$^{2.1}$, pUC57-kan-SCmod$^{2.2}$, pUC57-kan-Nmod$^{2.1}$ or pUC57-kan-Nmod$^{2.2}$) (Figure 5.25A). Ligation reactions were then transformed into competent bacterial cells and subject to antibiotic selection with kanamycin. Since *Dra*III leaves a 3' overhang, while *Sca*I produces blunt end DNA fragments (i.e. no 5' or 3' overhangs) the vector should not be able to re-ligate, and insertion of a modification should be directional. As such, following antibiotic selection, successful bacterial transformants should express one of; pUC57-kan-SC$^{V2.1}$, pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ or pUC57-kan-N$^{V2.2}$ (Figure 5.25A and Table 5.5).

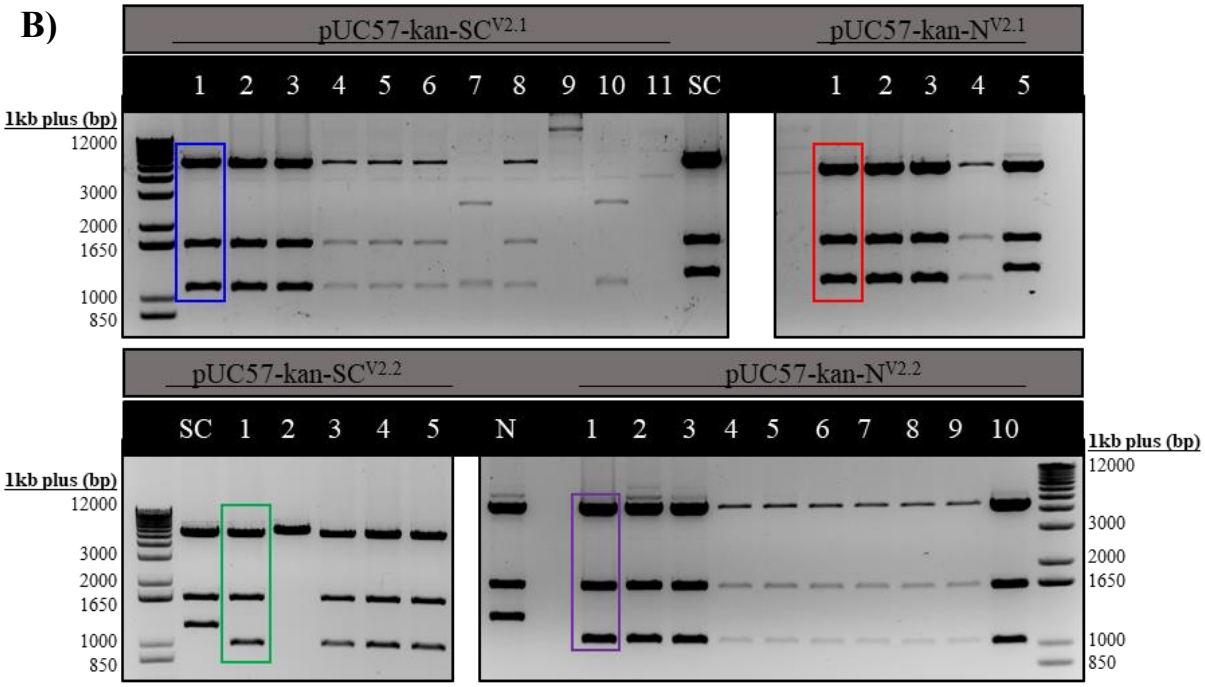
To confirm that ligation of the desired modification was successful, plasmid DNA isolated from several resistant bacterial transformants was subject to diagnostic restriction endonuclease

digests (Figure 5.25B). Clone 1 of all ligations was seen to match the expected band sizes, this suggested that there were no large sequence errors (Figure 5.25B).

Table 5.5: A description of Version 2.1 and Version 2.2 expression plasmids and the proteins they encode

Version	Plasmid	Origin	Expression cassette(s) included in plasmid			Encoded Protein(s)	Expected protein size (s)
			Selection	Control	NMD		
V2.1	pUC57-kan-SCmod$^{2.1}$	Genscript	-	-	-	-	
	pUC57-kan-Nmod$^{2.1}$	Genscript	-	-	-	-	
	pUC57-kan-SC$^{V2.1}$	In-house	S$^{V2.0}$	C$^{V2.1}$		^{HA}CFP^{NLS}, HBB^{WT}	~32 kDa, ~ 16kDa
	pUC57-kan-N$^{V2.1}$	In-house	-	-	N $^{2.1}$	^{FLAG}YFP^{NLS}, HBB^{NS39}	~32 kDa, ~ 4 kDa
V2.2	pUC57-kan-SCmod$^{2.2}$	Genscript	-	-		-	
	pUC57-kan-Nmod$^{2.2.}$	Genscript	-	-	-	-	
	pUC57-kan-SC$^{V2.1}$	In-house	S$^{V2.0}$	C$^{V2.2}$		^{HA}CFP^{NLS}, HBB$^{WT, \Delta E1}$	~32 kDa, ~13 kDa
	pUC57-kan-N$^{V2.1}$	In-house	-	-	N $^{2.2}$	^{FLAG}YFP^{NLS}, HBB$^{NS39, \Delta E1}$	~32 kDa, ~1 kDa





Expected band sizes following *ScaI* / *BamHI* diagnostic restriction enzyme digestion					
pUC57-kan-SC$^{V2.0}$ (vector, SC)	pUC57-kan-N$^{V2.0}$ (vector, N)	pUC57-kan-SC$^{V2.1}$	pUC57-kan-N$^{V2.1}$	pUC57-kan-SC$^{V2.2}$	pUC57-kan-N$^{V2.2}$
5620 bp 1639 bp 1212 bp	4339 bp 1634 bp 1212 bp	5620 bp 1639 bp 1099 bp	4339 bp 1634 bp 1109 bp	5620 bp 1639 bp 1009 bp	4339 bp 1634 bp 1019 bp

Figure 5.25: Restriction endonuclease digest screening for bacterial colonies containing pUC57-kan-SC$^{V2.1}$ pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ or pUC57-kan-N$^{V2.2}$ plasmid DNA.

(A) A Schematic representation of the ligase based cloning method used to assemble pUC57-kan-SC$^{V2.1}$, pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ and pUC57-kan-N$^{V2.2}$. **(B)** Following ligation, bacterial transformation, and antibiotic selection. DNA was isolated from several colonies and digested with the restriction endonucleases, *Sca*I and *BamH*I to check if inclusion of the plasmid modification was successful. Expected digest product sizes are summarised in the table (these are colour coded to refer to the boxes on the agarose gel image). The digested products were visualised by agarose gel electrophoresis, coloured boxes indicate colonies harbouring a successful ligation.

5.3.4.3 *Testing sequence fidelity of expression plasmids encoding Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$*

5.3.4.3.1 *Sanger sequencing reveals frameshift inducing deletions in expression plasmids encoding Control$^{V2.1}$ and Control$^{V2.2}$*

To ensure sequence fidelity prior to downstream experiments, plasmid DNA isolated from successful bacterial transformants was subject to Sanger sequencing. This confirmed that expression plasmids encoding NMD$^{V2.1}$ and NMD$^{V2.2}$ (pUC57-kan-N$^{V2.1}$ and pUC57-kan-N$^{V2.2}$) both contained no errors and encoded the genetic mutation for the *HBB* NS39 PTC (Figure 5.26). Sanger sequencing also revealed that expression plasmids encoding Control$^{V2.1}$ and Control$^{V2.2}$ (pUC57-kan-SC$^{V2.1}$ and pUC57-kan-C$^{V2.2}$) both contained a 16 bp deletion at the start of *HBB* exon 2 (Figure 5.26).

These 16 bp deletions are predicted to introduce a frameshift within the open reading frames (ORFs) of Control$^{V2.1}$ and Control$^{V2.2}$. This would result in introduction of a PTC within *HBB* exon 2 of the *HBB* mini-genes. These PTCs will reside greater than 55 nts upstream of an exon-exon junction and are thus predicted to be targeted for NMD based on the general 55 nt rule of NMD. Further investigations found that these deletions were present in the original modification carrying plasmids (pUC57-kan-SCmod$^{2.1}$ and pUC57-kan-SCmod$^{2.2}$). Therefore, to rectify this issue these plasmids would need to be re-synthesised.

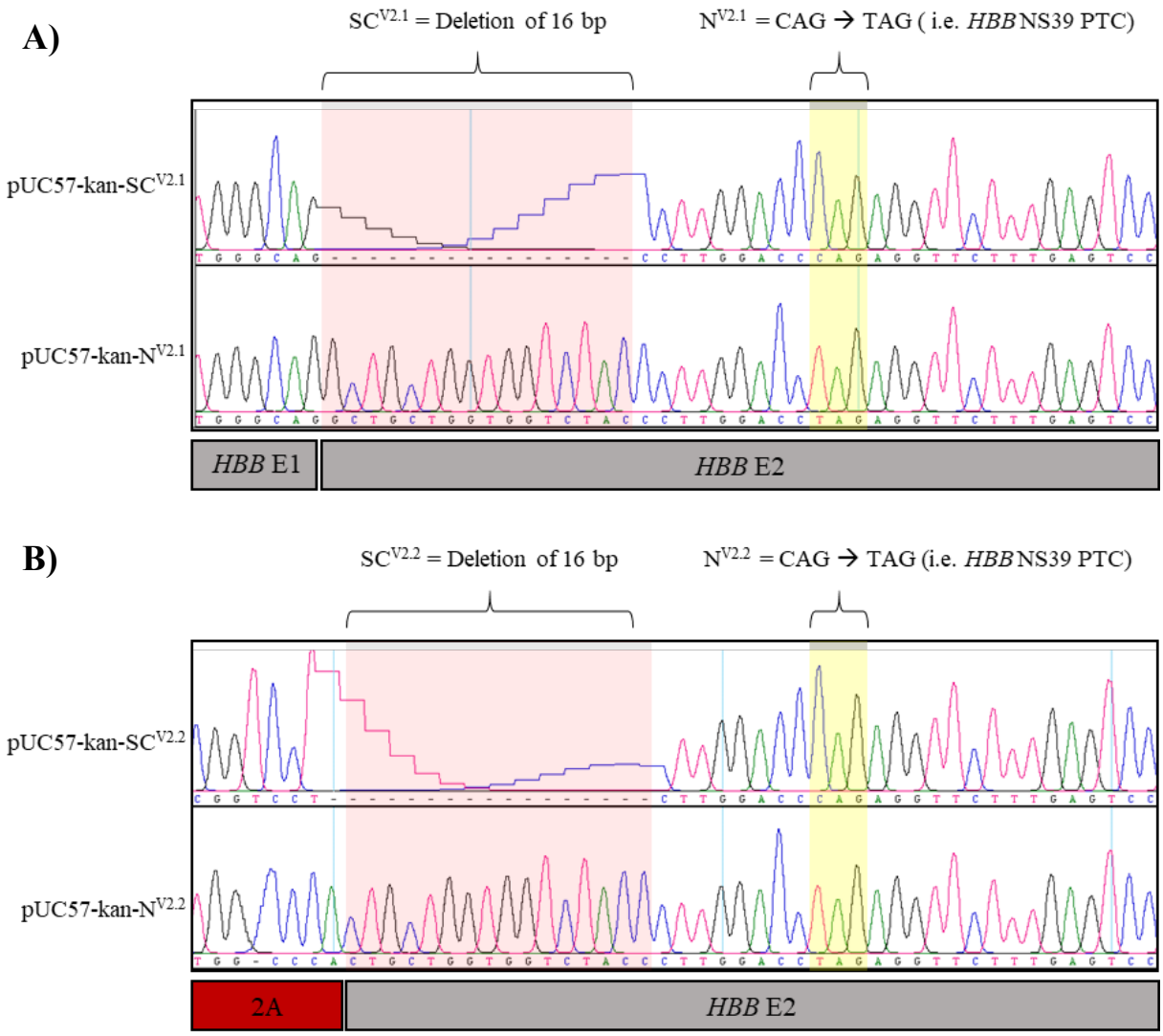


Figure 5.26: Sanger sequencing identifies frameshift inducing deletions in pUC57-kan-SC$^{V2.1}$ and pUC57-kan-SC$^{V2.2}$.

(A) Plasmid DNA encoding Selection$^{V2.1}$ and Control$^{V2.1}$ (pUC57-kan-SC$^{V2.1}$) or NMD$^{V2.1}$ (pUC57-kan-N$^{V2.1}$) was subject to Sanger sequencing. Sequence alignment revealed that pUC57-kan-SC$^{V2.1}$ contained a 16 bp deletion at the beginning of *HBB* exon 2 (E2) (red shading). This also showed that pUC57-kan-N$^{V2.1}$ contained no errors and included the *HBB* NS39 PTC encoding sequence (yellow shading). **(B)** Plasmid DNA encoding Selection$^{V2.2}$ and Control$^{V2.2}$ (pUC57-kan-SC$^{V2.2}$) or NMD$^{V2.2}$ (pUC57-kan-N$^{V2.2}$) was subject to Sanger sequencing. Sequence alignment revealed that pUC57-kan-SC$^{V2.2}$ contained a 16 bp deletion at the beginning of *HBB* exon 2 (E2) (red shading). This also showed that pUC57-kan-N$^{V2.2}$ contained no errors and included the *HBB* NS39 PTC encoding sequence (yellow shading).

5.3.4.3.2 *Transient expression of plasmids encoding Control$^{V2.1}$ and Control$^{V2.2}$ reveals that 2A self-cleavage sequences render nearby glycyl-prolyl peptide bonds susceptible to cleavage in HEK293T cells*

An interesting observation was that when expression plasmids encoding Control$^{V2.1}$ and Control$^{V2.2}$ which both contained 16 bp deletions within *HBB* exon 2 of their *HBB* mini-gene (pUC57-kan-SC$^{V2.1}$ and pUC57-kan-SC$^{V2.2}$) were transiently introduced into HEK293T cells, several different sized protein products were detected by western blot analysis (Figures 5.27B & 5.27C). Firstly, this suggests that these transcripts are not efficiently targeted by NMD as was predicted by the introduction of a PTC due to a frameshifted ORF. Secondly, this shows that bulk translation of the PTC containing transcripts was occurring.

Since the 16 bp deletions detected in Control$^{V2.1}$ and Control$^{V2.2}$ do not disrupt the E2A self-cleavage sequence, ribosomal skipping at the final glycyl-prolyl peptide bond encoded by this sequence should still occur. If this is the case, following evasion of NMD two separate proteins should be produced from Control$^{V2.1}$ and Control$^{V2.2}$; ^{HA}CFP^{NLS} and a protein encoded by the frameshifted *HBB* mini-gene, referred to here as HBB^{PTC}.

Following transfection of pUC57-kan-SC$^{V2.1}$ or pUC57-kan-SC$^{V2.2}$ into HEK293T cells, western blot analysis detected a ~32 kDa protein product indicative of correctly cleaved ^{HA}CFP^{NLS}. Another ~38 kDa protein product was also detected in both cases, this is likely the uncleaved ^{HA}CFP^{NLS}:HBB^{PTC} fusion protein (Figure 5.27B). These results suggest that at least under the conditions tested, (albeit not ideal due to errors in the DNA sequence encoding Control$^{V2.1}$ and Control$^{V2.2}$) ribosomal skipping from the E2A sequence is inefficient.

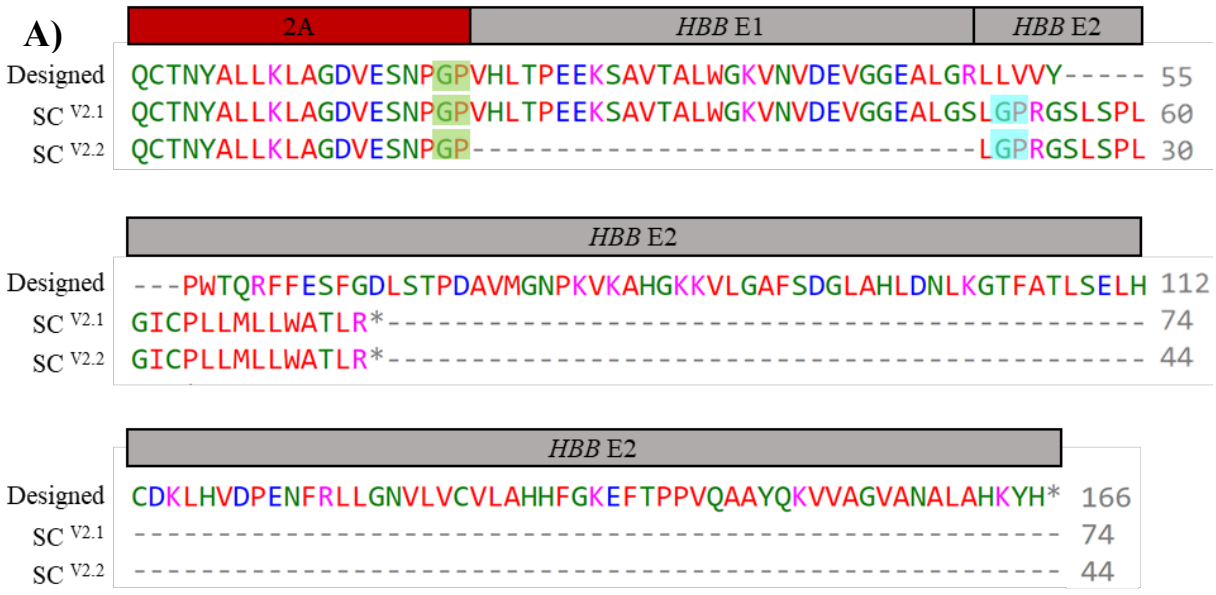
Interestingly, an additional ~35 kDa protein product was also detected by western blot when pUC57-kan-SC$^{V2.1}$ was transiently expressed in HEK293T cells (Figure 5.27B). This product

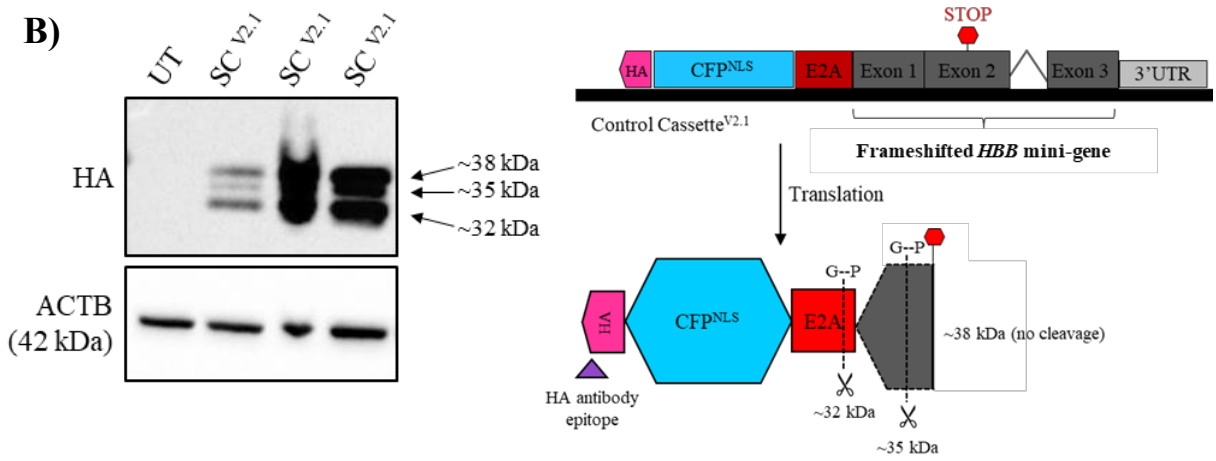
could be explained by considering the amino acid sequence resulting from the frameshifted ORF of Control$^{V2.1}$ (Figure 5.27A). The disrupted *HBB* mini-gene now encodes a glycyl-prolyl peptide bond 33 amino acids downstream of that encoded by the E2A sequence (Figure 5.27A). The observed ~35 kDa protein suggests that ribosomal skipping may also be occurring between this additional glycyl-prolyl peptide bond (Figure 5.27B). Suggesting that 2A self-cleavage sequences may have the potential to render nearby glycyl-prolyl peptide bonds susceptible to ribosomal skipping in HEK293T cells.

When pUC57-kan-SC$^{V2.2}$ was transiently expressed in HEK293T cells, evidence of ribosomal skipping between the additional glycyl-prolyl peptide bond within protein produced from Control$^{V2.2}$ was not detected by western blot (Figure 5.27C). If such ribosomal skipping was occurring, the resulting protein would weigh ~32.1 kDa. By western blot this would be difficult to distinguish from the ~32 kDa correctly cleaved protein. As such, ribosomal skipping could still be occurring at the aberrant glycyl-prolyl peptide but remain un-detectable by western blot.

If this phenomena is proven to be correct, these results would indicate that when using a 2A self-cleavage sequence it is important to closely examine the amino acid sequence of the protein downstream from the final 2A glycine-proline residues for additional glycine-proline residues which may lead to unwanted ribosomal skipping events.

Moving forward, it was decided that prior to re-engineering expression plasmids for Control$^{V2.1}$ and Control$^{V2.2}$ the function of NMD$^{V2.1}$ and NMD$^{V2.2}$ would first be tested. In this way, features shared between Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ and the response to changes in cellular NMD activity of NMD$^{V2.1}$ and NMD$^{V2.2}$ can be assessed. The results of these experiments were viewed as prerequisites to determine if Control$^{V2.1}$ and Control$^{V2.2}$ should be re-synthesised as per the original designs or if these cassettes would require further modifications.


A)
2A
HBB E1
HBB E2
Designed
SC V2.1
SC V2.2
QCTNYALLKLAGDVESNPGPVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVY----- 55
QCTNYALLKLAGDVESNPGPVHLTPEEKSAVTALWGKVNVDEVGGEALGSLGPRGSLSPL 60
QCTNYALLKLAGDVESNPGP----------LGPRGSLSPL 30
HBB E2
---PWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELH 112
GICPLLMLLWATLR*---------- 74
GICPLLMLLWATLR*---------- 44
HBB E2
CDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH* 166
---------- 74
---------- 44


B)
UT
SC V2.1
SC V2.1
SC V2.1
HA
ACTB
(42 kDa)
~38 kDa
~35 kDa
~32 kDa
STOP
HA
CFPNLS
E2A
Exon 1
Exon 2
Exon 3
3'UTR
Control CassetteV2.1
Frameshifted HBB mini-gene
Translation
VH
HA antibody
epitope
G--P
~32 kDa
~35 kDa
~38 kDa (no cleavage)

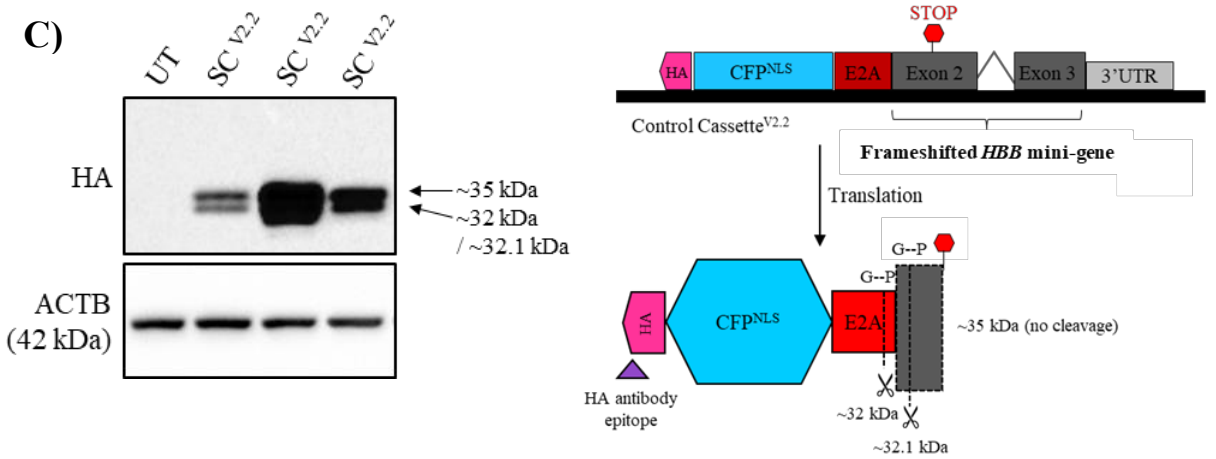

C)
UT
SC V2.2
SC V2.2
SC V2.2
HA
ACTB
(42 kDa)
~35 kDa
~32 kDa
/ ~32.1 kDa
STOP
HA
CFPNLS
E2A
Exon 2
Exon 3
3'UTR
Control CassetteV2.2
Frameshifted HBB mini-gene
Translation
VH
HA antibody
epitope
G--P
~32 kDa
~32.1 kDa
~35 kDa (no cleavage)

Figure 5.27: 2A self-cleavage sequences render nearby glycyl-prolyl peptide bonds susceptible to cleavage in HEK293T cells.

(A) Amino acid sequence encoded by the original design for Control$^{V2.1}$ (designed) and the actual amino acid sequence encoded by expression plasmids encoding Control$^{V2.1}$ (pUC57-kan-SC$^{V2.1}$, SC$^{V2.1}$) or Control$^{V2.2}$ (puc57-kan-SC$^{V2.2}$, SC$^{V2.2}$) due to a 16 bp frameshift deletion at the beginning of *HBB* exon 2 (E2). The original designs encode a single glycyl-prolyl peptide bond (GP) at the end of the 2A sequence (shaded in green) however, the actual sequences for Control$^{V2.1}$ and Control$^{V2.2}$ now encode a second glycyl-prolyl peptide bond at the beginning of *HBB* exon 2 (shaded in blue). To note, other than exclusion of amino acids encoded by *HBB* exon 1, the amino acid sequence encoded by Control$^{V2.2}$ is identical to that encoded by Control$^{V2.1}$. **(B)** Total protein isolated from HEK293T cells transfected with an expression plasmid encoding Control$^{V2.1}$ (pUC57-kan-SC$^{V2.1}$, SC$^{V2.1}$) was separated by SDS-PAGE and analysed by western blot using antibodies to detect HA and the loading control ACTB (stripped and re-probed). Based on detection of three bands, the schematic to the right depicts that proteins encoded by Control$^{V2.1}$ are likely cleaved inefficiently at both glycyl-peptide bonds (G--P). **(C)** Total protein isolated from HEK293T cells transfected with an expression plasmid encoding Control$^{V2.2}$ (pUC57-kan-SC$^{V2.2}$, SC$^{V2.2}$) was separated by SDS-PAGE and analysed by western blot using antibodies to detect HA and the loading control ACTB (stripped and re-probed). Based on observations from (B), the schematic to the right depicts that proteins encoded by Control$^{V2.2}$ is likely cleaved inefficiently at both glycyl-peptide bonds (G-P), however due to an ~0.1 kDa difference between the two smaller products, these cannot be resolved by SDS-PAGE resulting in detection of what appears to be only two products.

5.3.4.4 Assessing HBB mini-gene splicing of transcripts expressed from Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$

Control$^{V2.0}$ and NMD$^{V2.0}$ encoded a fluorescent protein/HBB fusion construct. Originally, the *HBB* mini-gene contained all three exons and two introns of endogenous *HBB*. This design resulted in a small proportion of incorrect transcript splicing which could possibly interfere with accurate NMD quantification. To attempt to correct mis-splicing, Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ were all designed without intron 1 of *HBB*.

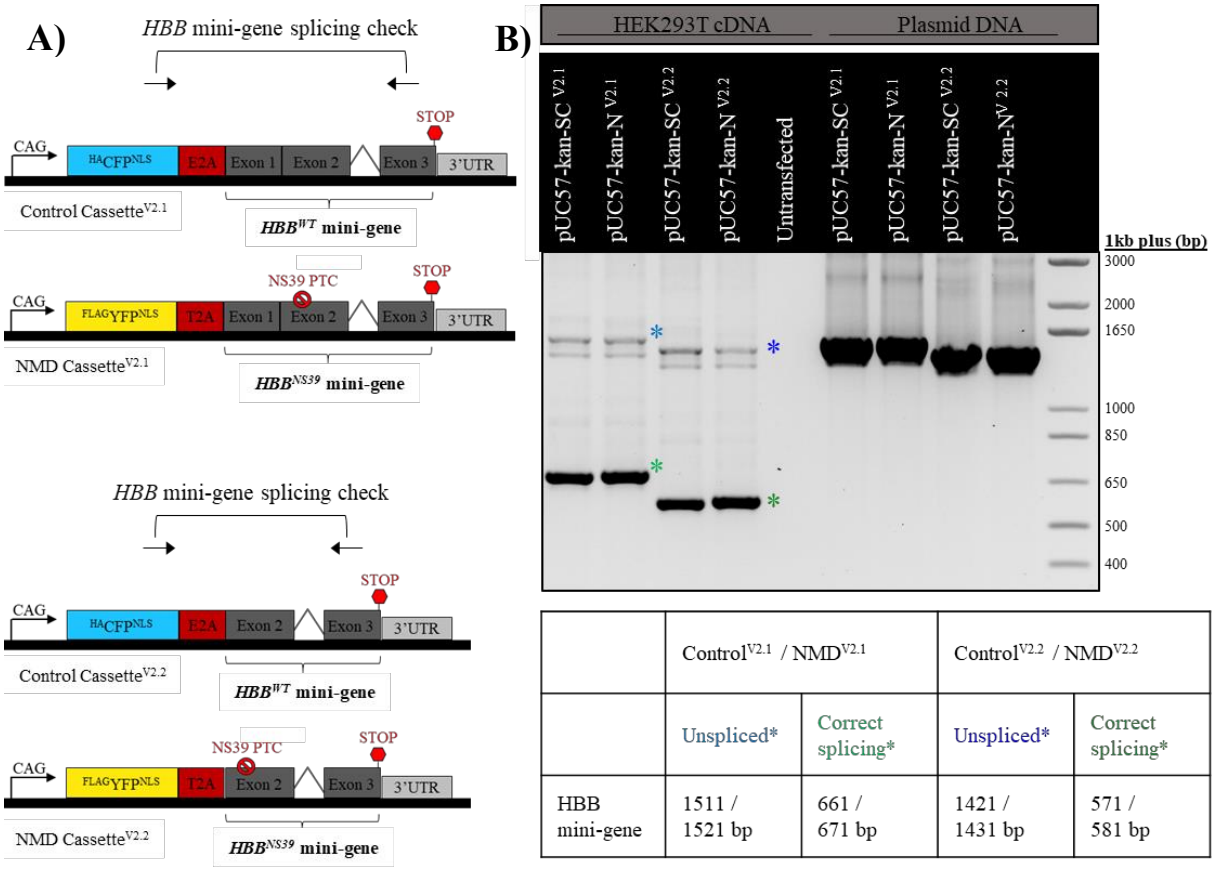
To investigate if this was successful, cDNA reverse transcribed from the RNA of HEK293T cells transiently transfected with an expression plasmid encoding Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ or NMD$^{V2.2}$ (pUC57-kan-SC$^{V2.1}$, pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ or pUC57-kan-N$^{V2.2}$ respectively) was subject to PCR using primers which bind within the sequences encoding the fluorescent protein (CFP or YFP) and the last exon of the *HBB* mini-gene to assess *HBB* mini-gene splicing (Figure 5.28A). pUC57-kan-SC$^{V2.1}$, pUC57-kan-SC$^{V2.2}$, pUC57-kan-N$^{V2.1}$ and pUC57-kan-N$^{V2.2}$ plasmid DNA was also subject to the same PCR as a control to identify any unspliced products. Correct splicing of the *HBB* mini-gene in these cells should produce a single PCR product at the expected size (Figure 5.28B).

PCR products were visualised via agarose gel electrophoresis, which showed that in these conditions the dominant transcripts produced from Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$ appeared to be correctly spliced. However, lowly expressed transcripts indicative of unspliced (intron 2 retained) were also detected from all cassettes (Figure 5.28B).

Retention of *HBB* intron 2 will result in the introduction of a PTC in the last exon of the *HBB* mini-gene in all cassettes tested (i.e. Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$). PTCs in the last exon of a gene often evade NMD as following splicing, they lack a downstream exon-

exon junction and thus cannot be recognised by the NMD machinery. If unspliced transcripts produced from $NMD^{V2.1}$ and $NMD^{V2.2}$ were to escape NMD and be translated, the resulting protein would be larger than expected. Such a protein was not observed (Figure 5.29B), suggesting that the observed unspliced transcripts may be an artefact of overexpression which, at least in this context, does not appear to be translated into protein.

This result was in line with previous investigations which have shown that unspliced products appear to be more prominent when cassettes are expressed transiently in cells, compared to when they are expressed following stable genomic integration (Section 5.3.3.2).



	ControlV2.1 / NMDV2.1		ControlV2.2 / NMDV2.2	
	Unspliced*	Correct splicing*	Unspliced*	Correct splicing*
HBB mini-gene	1511 / 1521 bp	661 / 671 bp	1421 / 1431 bp	571 / 581 bp

Figure 5.28: Transcripts expressed from ControlV2.1, ControlV2.2, NMDV2.1 and NMDV2.2 are predominantly spliced correctly when expressed transiently from HEK293T cells.

(A) A schematic representation of primers designed to assess splicing of the *HBB* mini-gene of transcripts expressed from Version 2.1 and Version 2.2 cassettes (*HBB* mini-gene splicing check). **(B)** cDNA reverse transcribed from the RNA of HEK293T cells transfected with expression plasmids encoding ControlV2.1, ControlV2.2, NMDV2.1 or NMDV2.2 (pUC57-kan-SCV2.1, pUC57-kan-SCV2.2, pUC57-kan-NV2.1 or pUC57-kan-NV2.2 respectively) were subject to PCR using the described primers. Plasmid DNA of these constructs was also subject to the same PCR as a control to identify unspliced transcripts. The table outlines the possible PCR product sizes (these are colour coded in reference to asterisks on the agarose gel image). PCR amplified products were visualised by agarose gel electrophoresis. The coloured asterisks on the gel identify a band as a particular splice product of the *HBB* mini-gene.

5.3.4.5 Investigating the cleavage efficiency of the T2A self-cleavage sequence within NMD$^{V2.1}$ and NMD$^{V2.2}$ in HEK293T cells

To investigate whether the T2A self-cleavage peptide within NMD$^{V2.1}$ and NMD$^{V2.2}$ was functional, expression plasmids encoding NMD$^{V2.1}$ and NMD$^{V2.2}$ (pUC57-kan-N$^{V2.1}$ and pUC57-kan-N$^{V2.2}$) were transiently transfected into HEK293T cells. Protein isolated from these cells was then analysed by western blot using a FLAG, GFP or HBB antibody to detect different regions of the encoded proteins (Figure 5.29B). It was predicted that transcripts expressed from the NMD$^{V2.1}$ and NMD$^{V2.2}$ would be efficiently degraded by NMD as they encoded the NS39 PTC. If this prediction were true, bulk translation of these transcripts would not occur and there would be no detected protein. This was not observed, instead it appeared that NMD$^{V2.1}$ and NMD$^{V2.2}$ both encoded a single protein indicative of an uncleaved product at ~36 kDa (Figure 5.29B). This suggests that (1) NMD of transcripts expressed from NMD$^{V2.1}$ and NMD$^{V2.2}$ is not efficient and (2) The T2A self-cleavage peptide is not functional.

If the T2A self-cleavage peptide was functioning as expected, ribosomal skipping should occur at the peptide bond between the last glycine and proline encoded by the T2A sequence. This would result in detection of two separate proteins; ^{FLAG}YFP^{NLS} (~32 kDa) and HBB^{NS39} from NMD$^{V2.1}$ (~4 kDa) or HBB$^{NS39,\ \Delta E1}$ from NMD$^{V2.2}$ (~1 kDa). If as previously suspected, HBB protein product is subject to degradation (Section 5.3.3.4), then only ^{FLAG}YFP^{NLS} would be detected. Since the size of the protein detected was larger than the ~32 kDa expected for ^{FLAG}YFP^{NLS}, western blot analysis suggested that protein produced from NMD$^{V2.1}$ and NMD$^{V2.2}$ remained uncleaved (Figures 5.29A & 5.29B).

To note, ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from NMD$^{V2.1}$ was predicted to have a molecular weight of ~36 kDa and ^{FLAG}YFP^{NLS}:HBB$^{NS39,\ \Delta E1}$ expressed from NMD$^{V2.1}$ was predicted to have a molecular weight of ~33 kDa. However when expression plasmids encoding these

cassettes were transiently transfected into HEK293T cells both $^{FLAG}YFP^{NLS}$:HBB^{NS39} and $^{FLAG}YFP^{NLS}$:$HBB^{NS39,\ \Delta E1}$ appeared to detected at the same size (Figure 5.29B).

The same HEK293T cells transiently transfected with expression plasmids encoding $NMD^{V2.1}$ or $NMD^{V2.2}$ (pUC57-kan-$N^{V2.1}$ or pUC57-kan-$N^{V2.2}$) were also analysed through immunoblotting and fluorescence microscopy. It was observed that proteins expressed from the *HBB* mini-gene of $NMD^{V2.1}$ and $NMD^{V2.2}$ were visible (i.e. likely not subject to proteasomal degradation). Moreover, the cellular localisation of HBB proteins overlapped with YFP signal expressed from the same plasmid (Figure 5.29C). In line with western blot analysis, the overlapping cellular localisation of HBB proteins and YFP signal suggests that protein produced from the *HBB* mini-gene of $NMD^{V2.1}$ or $NMD^{V2.2}$ remains uncleaved from YFP expressed from the same cassette.

Together, these data show that when transiently expressed in HEK293T cells, at least some transcripts expressed from $NMD^{V2.1}$ and $NMD^{V2.2}$ escape NMD and are translated into a protein which remains uncleaved due to inefficient or non-existent activity from its T2A self-cleavage sequence.

A)

	Efficient NMD (expected)	Inefficient NMD leading to translation with efficient cleavage	Inefficient NMD leading to translation with inefficient cleavage
pUC57-kan-$N^{V2.1}$ ($N^{V2.1}$)	No protein produced	$^{FLAG}YFP^{NLS}$ (~32 kDa) + HBB^{NS39} (~ 4 kDa)	$^{FLAG}YFP^{NLS}$:HBB^{NS39} (~36)
pUC57-kan-$N^{V2.2}$ ($N^{V2.2}$)	No protein produced	$^{FLAG}YFP^{NLS}$ (~ 32 kDa) + $HBB^{NS39,\ \Delta E1}$ (~1 kDa)	$^{FLAG}YFP^{NLS}$:$HBB^{NS39,\ \Delta E1}$ (~33)

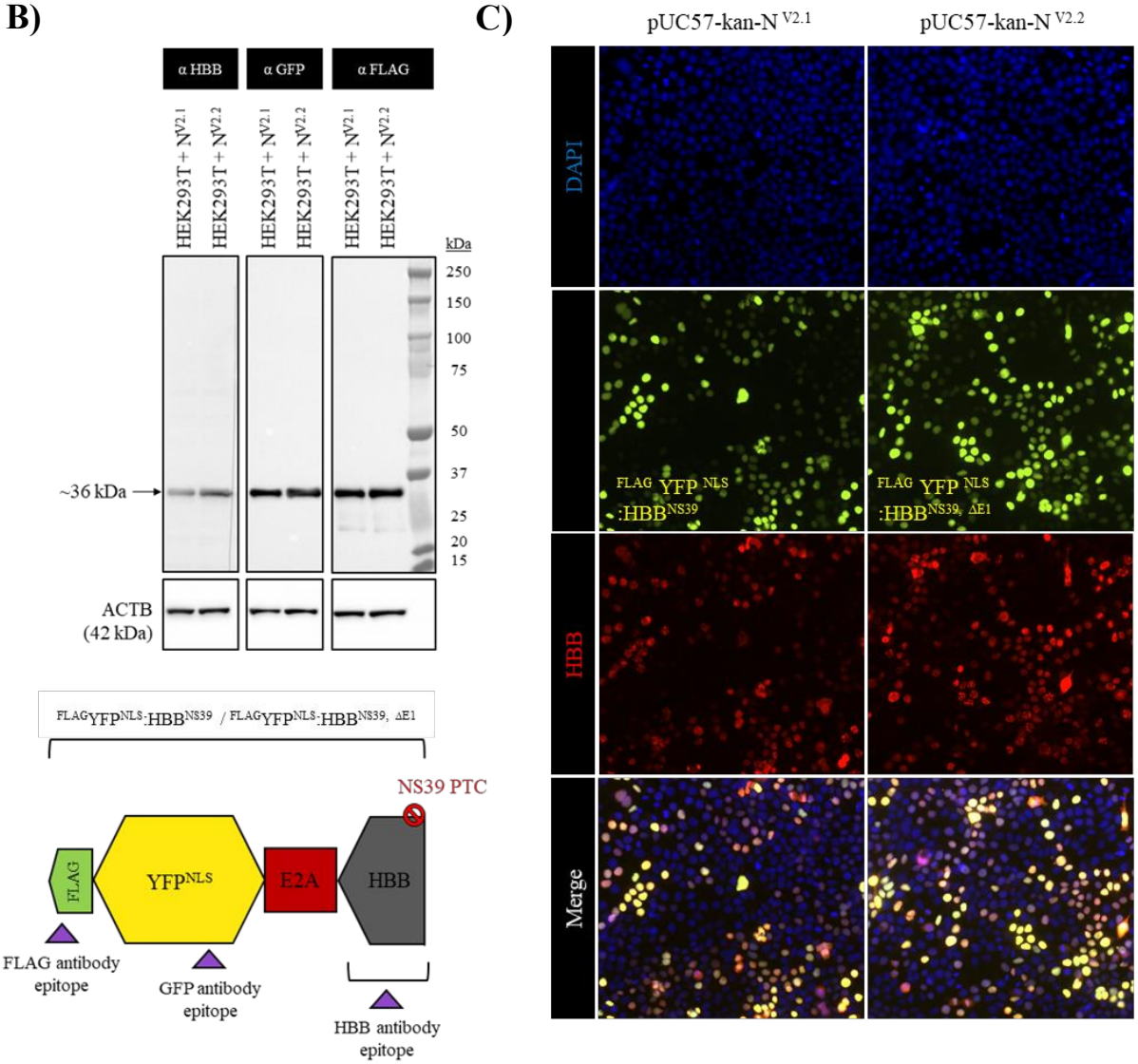


Figure 5.29: T2A self-cleavage sequence shows inefficient cleavage when used within $NMD^{V2.1}$ and $NMD^{V2.2}$ following transient expression from HEK293T cells.

HEK293T cells were transiently transfected with expression plasmids encoding $NMD^{V2.1}$ (pUC57-kan-$N^{V2.1}$, $N^{V2.1}$) or $NMD^{V2.2}$ (pUC57-kan-$N^{V2.2}$, $N^{V2.2}$). **(A)** The table summarises the possible effects of NMD on transcripts expressed from $NMD^{V2.1}$ and $NMD^{V2.2}$ and the result this will have on protein production from $NMD^{V2.1}$ and $NMD^{V2.2}$. **(B)** Total protein isolated

Figure 5.29 continued…

from these cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect HBB, GFP, FLAG and the loading control ACTB (stripped and re-probed). Based on detection of only one band at ~36 kDa, the schematic below depicts the likely, uncleaved protein produced from $NMD^{V1.0}$ ($^{FLAG}YFP^{NLS}$:HBB^{NS39}) and $NMD^{V2.0}$ ($^{FLAG}YFP^{NLS}$:$HBB^{NS39,\ \Delta E1}$). This schematic also outlines the positions of the epitopes recognised by antibodies used in western blot analysis. **(C)** Representative fluorescence microscopy images of these cells. The cells express $^{FLAG}YFP^{NLS}$:HBB^{NS39} from $NMD^{V2.1}$ or $^{FLAG}YFP^{NLS}$:$HBB^{NS39,\ \Delta E1}$ from $NMD^{V2.2}$ (yellow). HBB was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue).

5.3.4.6 Testing protein expression from NMD$^{V2.1}$ and NMD$^{V2.2}$ in response to an siRNA mediated reduction of UPF1 levels in HEK293T cells

Although cleavage from the T2A sequence of NMD$^{V2.1}$ and NMD$^{V2.2}$ was found to be inefficient (Section 5.3.4.5), it was still possible for transcripts expressed from these cassettes to be sensitive to changes in cellular NMD activity due to their encoded NMD targeted NS39 PTC. To test if this was true, expression plasmids encoding NMD$^{V2.1}$ and NMD$^{V2.2}$ (pUC57-kan-N$^{V2.1}$and pUC57-kan-N$^{V2.2}$) were co-transfected separately into HEK293T cells, alongside Control or *UPF1* siRNA. If transcripts expressed from NMD$^{V2.1}$ or NMD$^{V2.2}$ are sensitive to NMD, a loss of UPF1 should result in increased expression from the cassette.

Following co-transfection, fluorescence microscopy and western blot analysis was used to assess protein expression in these cells. Through these analyses, it was seen that following a reduction in UPF1, the level of ^{FLAG}YFP^{NLS}:HBB^{NS39} protein expressed from NMD$^{V2.1}$ and ^{FLAG}YFP^{NLS}:HBB$^{NS39,\Delta E1}$ expressed from NMD$^{V2.2}$ remained unchanged (Figures 5.30 & 5.31). This suggests that transcripts produced from NMD$^{V2.1}$ and NMD$^{V2.2}$ are escaping NMD when these cassettes are transiently expressed in HEK293T cells. This outcome was unexpected as prior to the updated modifications (i.e. introduction of T2A and deletion of HBB exon 1 and/or intron 1) NMD$^{V2.0}$ was sensitive to NMD when tested transiently in HEK293T cells (Section 5.3.2.4).

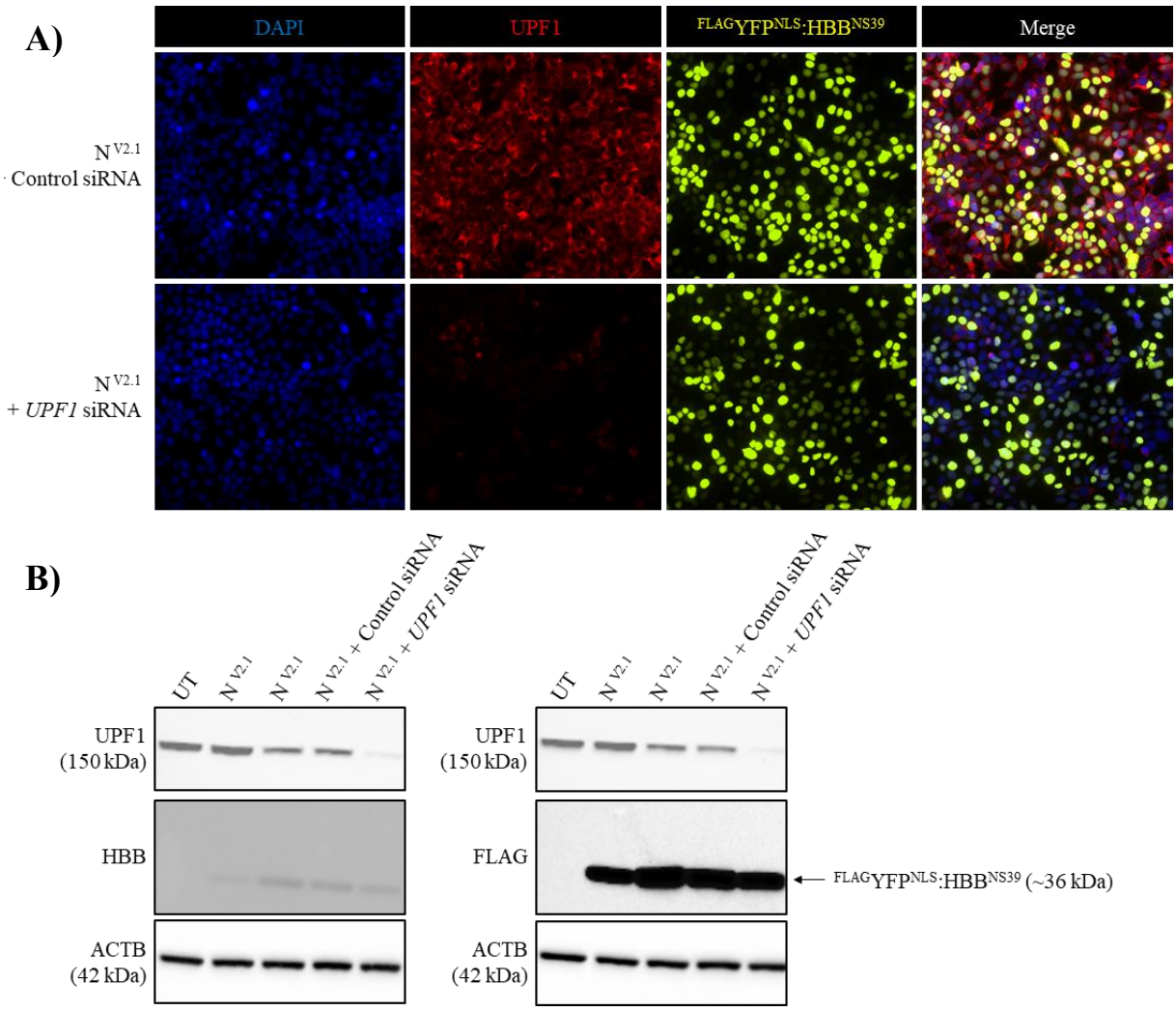


Figure 5.30: Protein expressed from NMD$^{V2.1}$ does not respond to reduced levels of UPF1 protein following transient transfections in HEK293T cells.

HEK293T cells were co-transfected with an expression plasmid encoding NMD$^{V2.1}$ (pUC57-kan-N$^{V2.1}$, N$^{V2.1}$) and Control or *UPF1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.1}$ (yellow). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells, untransfected HEK293T cells (UT) and HEK293T cells transfected with only pUC57-kan-N$^{V2.1}$ (N$^{V2.1}$), was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, HBB, FLAG and the loading control ACTB (stripped and re-probed).

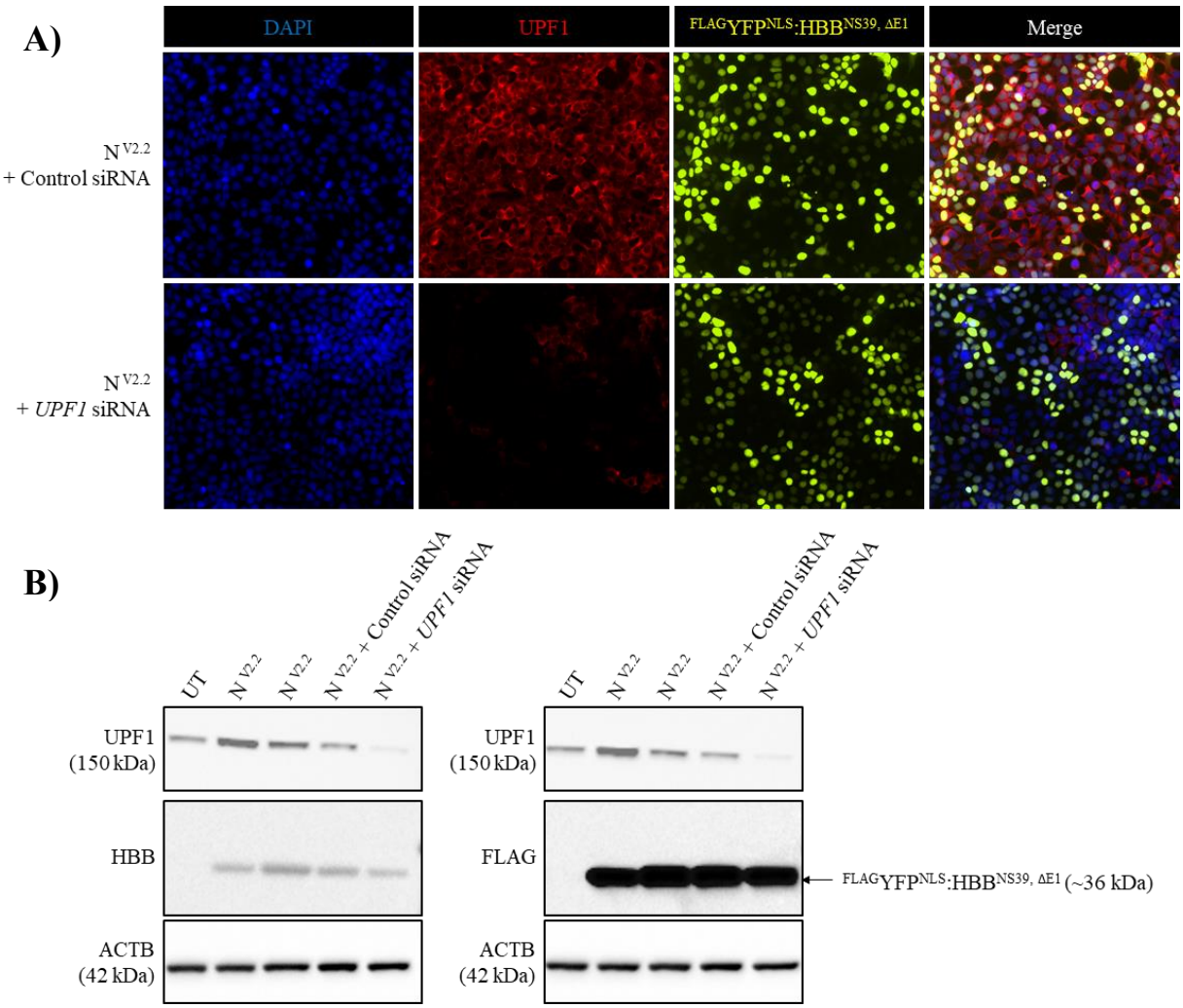


Figure 5.31: Protein expressed from NMD$^{V2.2}$ does not respond to reduced levels of UPF1 protein following transient transfections in HEK293T cells.

HEK293T cells were co-transfected with an expression plasmid encoding NMD$^{V2.2}$ (pUC57-kan-N$^{V2.2}$, N$^{V2.2}$) and Control or *UPF1* siRNA. **(A)** Representative fluorescence microscopy images of these cells. The cells express ^{FLAG}YFP^{NLS}:HBB$^{NS39,\ \Delta E1}$ from NMD$^{V2.2}$ (yellow). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue). **(B)** Total protein isolated from these cells, untransfected HEK293T cells (UT) and HEK293T cells transfected with only pUC57-kan-N$^{V2.2}$ (N$^{V2.2}$), was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, HBB, FLAG and the loading control ACTB (stripped and re-probed).

The next step in functional validation would be to use standard recombinant DNA cloning methods to extract the sequence encoding $NMD^{V2.1}$ or $NMD^{V2.2}$ from their respective expression plasmids (pUC57-kan-$N^{V2.1}$ or pUC57-kan-$N^{V2.2}$) and re-introduce them into an expression cassette encoding a validated Selection Cassette (i.e. pUC57-$S^{V1.0}$ of pUC57-kan-$SC^{V2.0}$). In this way, recombinant mESC lines which express $NMD^{V2.1}$ or $NMD^{V2.2}$ from the *Col1a1* locus could be established via the previously described RMCE method (Bersten et al, 2015). These cells could then be used to test if the features of $NMD^{V2.1}$ or $NMD^{V2.2}$, i.e. response to NMD, T2A self-cleavage, *HBB* mini-gene splicing, are functional following genomic integration.

To attempt to pursue this line of experiments, previously successful cloning methods were applied (Figure 5.14A), however, despite numerous in-house attempts, followed by paid for services by the Gene Silencing and Expression core facility (Robinson Research Institute, Adelaide) an expression plasmid encoding a Selection Cassette ($Selection^{V1.0}$ or $Selection^{V2.0}$) and either $NMD^{V2.1}$ or $NMD^{V2.2}$ could not be engineered. Furthermore, since the modifications introduced into the cassettes of Version 2.0 to establish Version 2.1 and Version 2.2 cassettes appeared to introduce more complications than solutions, functional validation of Version 2.1 and Version 2.2. cassettes were concluded at this point.

5.4 Discussion

The aim of this chapter was to design and validate Transgene$^{V2.0}$ in order to overcome the shortcomings of Transgene$^{V1.0}$ and report on endogenous NMD activity, including its cell and tissue specificity, and its role in development and disease.

The design of Transgene$^{V2.0}$ required several basic features to be chosen and functionally validated. Like Transgene$^{V1.0}$, nuclear localised fluorescence was chosen to quantify NMD activity levels in Transgene$^{V2.0}$. As discussed (Chapter 4), this method ensures that NMD activity can be quantified with single cell resolution in both fixed or live cells using fluorescence microscopy or FACs. Moreover, restricting fluorescence to the nucleus of a cell ensures that in circumstances where there are many cells or cell types present, for example; in neural cell cultures or brain tissue, fluorescent output can be easily attributed to a specific cell or cell type.

Following successful use of CFP^{NLS} in the design of Control$^{V1.0}$, this fluorescent protein was also designed to express from Control$^{V2.0}$ as part of a fluorescent protein/HBB fusion construct. Originally, EGFP^{NLS} was designed to express from NMD$^{V1.0}$, however, since fluorescent signal from this protein was seen to interfere with spectral resolution of CFP^{NLS}, NMD$^{V2.0}$ was designed to express YFP^{NLS} as part of a fluorescent protein/HBB fusion construct. The greater excitation and emission wavelength of YFP compared to EGFP allowed efficient spectral resolution of CFP^{NLS} and YFP^{NLS} from one another. Furthermore, in the design of Control$^{V2.0}$ and NMD $^{V2.0}$, nuclear localised fluorescent proteins were also N-terminally epitope tagged to enable robust non-fluorescent protein detection via methods such as, western blot analysis and immunofluorescence analysis.

Since the end goal was to incorporate Transgene$^{V2.0}$ into mESCs, an integration method was needed. To facilitate this, the previously validated Selection$^{V1.0}$ underwent minor non-functional sequence changes to develop Selection$^{V2.0}$ and was synthesised as part of an expression plasmid also encoding Control$^{V2.0}$ (pUC57-kan-SC$^{V2.0}$). Selection$^{V2.0}$ provides a tool which facilitates the integration of an entire transgene into the genetically modified *Col1a1* locus of FLP-in mESCs (or FLP-in mouse zygotes) as a single stable copy via the previously described RMCE system (Bersten et al. 2015).

Following integration of Transgene$^{V2.0}$ into pluripotent mESCs, NMD activity across differentiation protocols or throughout mouse development can be studied, however, for this to be informative Transgene$^{V2.0}$ needs to express and function efficiently in all cell types. To achieve this, another feature carried forward from Transgene$^{V1.0}$ was the synthetic CAG promoter. This promoter is known to drive strong constitutive expression in all mammalian cell types (Niwa, Yamamura & Miyazaki 1991; Sakaguchi et al. 2014), and was designed to drive expression from Control$^{V2.0}$ and NMD$^{V2.0}$ of Transgene$^{V2.0}$.

Previous investigations uncovered that when expression was driven from the CAG promoter of the Control$^{V1.0}$ or NMD$^{V1.0}$ a single SV40 polyA sequence was unable to efficiently terminate transcription. To overcome this, the design of Control$^{V2.0}$ and NMD$^{V2.0}$ both included three SV40 polyA sequences in tandem within the 3'UTR. This change was successful in achieving efficient transcriptional termination of transcripts produced from Control$^{V2.0}$ and NMD$^{V2.0}$ when expressed transiently from HEK293T cells or from Stable NMD Reporter$^{V2.0}$ mESCs.

The next step in the design process was to incorporate features which can report on cellular NMD activity levels. The design of Control$^{V2.0}$ and NMD$^{V2.0}$ is based the pioneering chemiluminescent NMD reporter system (Boelz et al. 2006). Both cassettes encode a fluorescent protein/HBB fusion construct. The main difference between NMD$^{V2.0}$ and

Control$^{V2.0}$ is that the *HBB* mini-gene of NMD$^{V2.0}$ encodes the HBB containing the NMD-targeted NS39 PTC (HBB^{NS39}), whilst the *HBB* mini-gene of Control$^{V2.0}$ encodes wild-type HBB (HBB^{WT}). In the design of NMD$^{V2.0}$, the sequence encoding the NS39 PTC resides within the second exon of the *HBB* mini-gene and following splicing, this PTC will reside more than 55 nts upstream of an exon-exon junction. Therefore, based on the general 55 nt rule of NMD, when NMD activity is high, transcripts expressed from NMD$^{V2.0}$ should be efficiently degraded by NMD resulting in low levels of ^{FLAG}YFP^{NLS}:HBB^{NS39} protein production, and vice versa. In comparison, ^{HA}CFP^{NLS}:HBB^{WT} designed to express from Control$^{V2.0}$ should express constitutively, regardless of cellular NMD activity levels and thus serve as an internal control.

5.4.1 Transcripts expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ are predominantly spliced correctly and the encoded proteins respond as expected to a reduction in UPF1 levels following transient expression in HEK293T cells

Functional testing of Control $^{V2.0}$ and NMD$^{V2.0}$ was first carried out transiently in HEK293T cells. High and low cellular NMD activity was generated in these cells using Control or *UPF1* siRNA respectively. Following a reduction in UPF1, protein expression from Control$^{V2.0}$ remained unchanged, whilst protein expression from NMD$^{V2.0}$ was observed to increase. These results matched the expected output from these cassettes and was observed when Control$^{V2.0}$ and NMD$^{V2.0}$ were expressed from separate plasmids (pUC57-kan-SC$^{V2.0}$ or pUC57-kan-N$^{V2.0}$) or when they were expressed simultaneously from a single expression plasmid (pUC57-kan-SCN$^{V2.0}$).

Control$^{V2.0}$ and NMD$^{V2.0}$ both encode a fluorescent protein/HBB fusion construct which contains all three exons and two introns of endogenous *HBB*. Although protein analysis suggested that these cassettes were functioning as expected when transiently expressed in

HEK293T cells, further investigations into the splicing of transcripts expressed from either Control$^{V2.0}$ or NMD$^{V2.0}$ in these cells revealed the formation of a number of mis-spliced transcripts alongside the dominant, correctly spliced product. Most markedly was the presence of *HBB* intron 1 retained transcripts.

Intron retention is generally caused by weak splice sites flanking short introns (< 274bp) (Sakabe & de Souza 2007; Zhan 2013). Both these features are true for the retained intron 1 of *HBB* within transcripts expressed from Control$^{V1.0}$ and NMD$^{V1.0}$. This intron is only 130 bp long and the flanking 5' donor site classifies as an alternative isoform/cryptic splice site compared to the remaining stronger constitutive splice sites within the *HBB* mini-gene (Wang & Marin 2006). Under normal conditions *HBB* transcripts can splice efficiently, however in the design of Control$^{V2.0}$ and NMD$^{V2.0}$, the addition of upstream features such as the sequence encoding fluorescent proteins has the potential to introduce splicing silencer motifs which may inhibit correct splicing through the recruitment of splicing silencer proteins. An alternative explanation is that these mis-spliced products are an artefact of overexpression and is supported by the fact that they were less prevalent when Control$^{V2.0}$ and NMD$^{V2.0}$ were expressed from stable NMD Reporter$^{V2.0}$ mESCs.

In the case of Control$^{V2.0}$ or NMD$^{V2.0}$, retention of *HBB* intron 1 is predicted to introduce an NMD targeted PTC. Expression from Control$^{V2.0}$ was intended to be unresponsive to changes in NMD efficiency, as such it could act as a direct internal control for the NMD-sensitive NMD$^{V2.0}$. If a significant number of transcripts expressed from Control$^{V2.0}$ retain intron 1 and as a result contain a PTC, transcripts and proteins expressed from this cassette have the potential to be influenced by cellular NMD activity and impair accurate quantification of NMD activity from this reporter system.

Transient experiments in HEK293T cells showed no evidence that protein expression from Control$^{V2.0}$ was influenced by NMD. However, to minimise the formation of mis-spliced variants, intron 1 of the *HBB* mini-gene was excluded in the subsequent designs of Control$^{V2.1}$, Control$^{V2.2}$, NMD$^{V2.1}$ and NMD$^{V2.2}$. Intron 2 of the *HBB* mini-gene was maintained in the design of all updated cassettes to ensure that transcripts underwent splicing and EJC deposition necessary for PTC recognition by NMD. It was observed that in transient transfection-based HEK293T experiments, removal of *HBB* intron 1 reduced the number and abundance of mis-spliced products from the *HBB* mini-gene, however, could not eliminate a small amount of unspliced product. These unspliced products were not seen to be translated into detectable protein and were likely an artefact of overexpression in a transient setting.

5.4.2 Proteins expressed from genomic Control $^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs are targeted for proteasomal degradation

The role of Selection$^{V2.0}$ is to act as a tool to integrate Transgene$^{V2.0}$ as a single, stable copy into the genetically modified *Col1a1* locus of FLP-in mESCs. In this chapter Selection$^{V2.0}$ facilitated the generation of NMD Reporter$^{V2.0}$ mESCs. This result confirmed that the unique features of Selection$^{V2.0}$ were still functional as part of a large transgene containing sequences encoding Control$^{V2.0}$ and NMD$^{V2.0}$.

Despite successful establishment of NMD Reporter$^{V2.0}$ mESCs, fluorescence expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of these cells was extremely low when compared to that from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs. Furthermore, protein expression from genomic NMD$^{V2.0}$ showed no response to a reduction in cellular NMD activity. This does not necessarily mean that the transcripts expressed from genomic NMD$^{V2.0}$ are not subject to NMD, however, without detectable fluorescence this NMD reporter system cannot provide a visual output of NMD activity at a single cell level.

Further investigations revealed that proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs were being targeted for proteasomal degradation. Moreover, fluorescence from these cassettes was not restored following a spontaneous differentiation of NMD Reporter$^{V2.0}$ mESCs into the three primary germ layers suggesting that this proteasomal degradation was not limited to mESCs:

An obvious difference between the Control and NMD cassettes of Version 1.0 and Version 2.0 is the inclusion of HBB encoding sequences in the translated region of Control$^{V2.0}$ and NMD$^{V2.0}$ but not Control$^{V1.0}$ and NMD$^{V1.0}$. Considering that protein expression from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ mESCs was not targeted by the proteasome and easily detected in these mESCs and the cells of all three germ layers derived from them, it was likely that translation of HBB protein from Control$^{V2.0}$ and NMD$^{V2.0}$ was having a negative effect on fluorescent protein expression from the same cassette.

human *HBB* is expressed from the β-globin locus on chromosome 11 which contains the five β-like globin genes; ε, ^{G}γ, ^{A}γ, δ, and β. Two β-like subunits combine with two of the α-like globins (ζ or α) to form haemoglobin (Hb), a tetramer which constitutes ~98% of protein in the erythrocyte cytoplasm and is responsible for oxygen transport in the blood (Iarovaia et al. 2018).

In humans, embryonic (Hb Gower 1; $\zeta_2\varepsilon_2$), fetal (HbF; $\alpha_2\gamma_2$) and adult haemoglobins (HbA; $\alpha_2\gamma_2$) are sequentially expressed in developing erythroblasts during ontogeny. Comparatively, murine erythropoiesis begins when primitive red blood cells, expressing the early εy and βh1 genes appear in the yolk sac blood islands at ~E7.5, and only a single switch occurs by mid-gestation (~E12.5) where expression of the early genes declines and a complementary increase in adult β^{maj} and β^{min} genes occurs (McColl & Vadolas 2016).

Transgenic mice expressing human HBB can recapitulate the developmental timing of the HbF-to-HbA switching that occurs in humans (Behringer et al. 1990; Enver et al. 1990; McColl et al. 2014; McConnell et al. 2011). Studies using such models has determined that the switch from early to adult globin expression is controlled by a complex regulatory network including; autonomous regulation of the early genes mediated by nearby sequences, cis competition between early and adult genes for interaction with the upstream locus control region and the presence and interaction of trans-acting factors (Hu et al. 2007; Liu et al. 2018; Sankaran & Orkin 2013; Wilber et al. 2011; Zhou et al. 2010) .

The proteasome is a selective degradation pathway present in both the nucleus and cytoplasm which targets ubiquitinated proteins for degradation (Grigoreva et al. 2015; Jang 2018; Morozov & Karpov 2018). In terms of both localisation and developmental timing, genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs introduces unnatural expression of the human adult HBB protein into mESCs. The presence and accumulation of which, are a likely cause for the observed proteolytic degradation of fluorescent protein/HBB fusion constructs expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs.

Preferential degradation of HBB is a phenomenon also observed in reticulocytes from patients with α thalassemia, in which the cell attempts to normalise alpha to beta globin chain levels (Sancar et al. 1981). Further evidence of HBB being a target for proteasomal degradation comes from *in vitro* studies which show that free α, β, and γ globin chains can be ubiquitinated and degraded either co-translationally of shortly after release of the nascent polypeptide from ribosomes (Adachi et al. 2004).

Another possibility is that when Control$^{V2.0}$ and NMD$^{V2.0}$ are expressed from NMD Reporter$^{V2.0}$ mESCs the fluorescent protein/HBB fusion constructs encoded fail to fold correctly and thus result in an accumulation of aberrant misfolded proteins causing ER stress within the cell. The

cells response to such elevated levels of ER stress would be to activate the unfolded protein response (UPR). The UPR consists of three distinct branches which act in synchrony to attempt to alleviate ER stress. If this is the case, a complex process activated by the downstream effects of the UPR and known as ER-associated degradation (ERAD) would work to first recognise the misfolded fluorescent protein/HBB fusion constructs, and subject them to ubiquitination and retrotranslocation into the cytosol where they would then be targeted for proteasomal degradation (Goetz & Wilkinson 2017; Hetz & Papa 2018; Ron & Walter 2007).

In any case, targeting of the fluorescent protein/HBB fusion constructs to the proteasome disrupts the fluorescence signal emitted from either $^{HA}CFP^{NLS}$:HBB^{WT} or $^{FLAG}YFP^{NLS}$:HBB^{NS39} expressed from $Control^{V2.0}$ or $NMD^{V2.0}$ respectively. This confounds their function within an NMD reporter system. To overcome this issue, the subsequent designs of $Control^{V2.1}$, $Control^{V2.2}$, $NMD^{V2.1}$ and $NMD^{V2.2}$ included a 2A self-cleavage peptide between the fluorescent protein and HBB protein. During translation (following any NMD of PTC containing transcripts) this self-cleavage peptide should stimulate the ribosome to skip the formation of a glycyl-prolyl peptide bond at the C-terminus of the 2A peptide thus separating the fluorescent protein from the HBB protein (Kim et al. 2011; Liu et al. 2017; Szymczak et al. 2004).

Contradictory to the literature, cleavage observed from the chosen 2A self-cleavage sequences was inefficient under the transient conditions tested in HEK293T cells (Kim et al. 2011; Liu et al. 2017; Szymczak et al. 2004). Furthermore, the introduction of the 2A cleavage sequence alongside modifications to the *HBB* mini-gene (deletion of exon 1 and/or intron 1) resulted in protein expression from $NMD^{V2.1}$ and $NMD^{V2.2}$ no longer being responsive to a reduction in NMD activity when transiently tested in HEK293T cells. These changes were also not compatible with previously successful cloning methods which would have allowed assembly of $NMD^{V2.1}$ or $NMD^{V2.2}$ into a single transgene with $Selection^{V2.0}$ (or $Selection^{V1.0}$). This prevented the establishment of mESC lines which expressed a single genomic copy of $NMD^{V2.1}$

or NMD$^{V2.2}$ and therefore the response to NMD of these cassettes could not be tested in a genomic context. Ultimately, the inclusion of 2A sequences as a modification to the original Control$^{V2.0}$ and NMD$^{V2.0}$ proved an unsuccessful approach to combat proteasomal degradation of proteins expressed from these cassettes.

5.4.3 Chapter conclusions and future directions

In this chapter I have engineered and tested the cassettes of Version 2.0, Version 2.1 and Version 2.2 of a fluorescent NMD reporter transgene. These cassettes were intended to provide a visual and quantifiable read out of endogenous NMD activity. Through functional testing via an experimental pipeline, several design features of these cassettes were validated whilst complications with others were identified.

Transgene$^{V2.0}$ showed promising results when expressed transiently in HEK293T cells and thus has the potential to be tailored into a stable HEK293T NMD reporter system to provide a fluorescent read out of NMD at a single cell level. These systems are useful for high-throughput experiments as they do not rely on transfection efficiency and can be quickly and easily analysed using FACS.

Unfortunately, when Transgene$^{V2.0}$ was used to establish NMD Reporter$^{V2.0}$ mESCs the encoded proteins were actively targeted for proteasomal degradation. This was attributed to the fact that by design human HBB proteins were expressed in unnatural cell types and developmental time points. I designed modifications to combat the observed proteasomal degradation, however these were ultimately unsuccessful.

Moving forward, there are several experiments that could be conducted to further investigate and rectify the issues faced with proteasomal degradation however, with the aim to make a functional NMD reporter which can map NMD activity across mouse brain development, these

efforts would be better directed toward the replacement of *HBB* in the NMD reporter transgene with another gene that can reliably express in the nucleus of mouse cells across all stages of development or towards engineering a fluorescent protein encoding gene itself to be efficiently targeted for degradation by NMD.

Chapter Six: Establishing Dual Fluorescent NMD Reporter HEK293T Cell Lines

6.1 Introduction

NMD is a selective and highly conserved mRNA turnover mechanism. In the past, the molecular mechanism of NMD has been characterised using pairs of reporter constructs that express the same mRNA with or without an NMD-targeted PTC in transient transfection-based studies. Although an enormous amount of knowledge has been gained through this approach, measuring output from a homogenous population of cells based on RNA quantification methods such as northern blot analysis and qPCR, fails to capture the dynamic nature of NMD and therefore limits the information output of such NMD reporter systems. Furthermore, it has been observed that not all cell lines efficiently degrade NMD reporter constructs when expressed transiently compared to when they are expressed following stable genomic integration (Gerbracht, Boehm & Gehring 2017).

So far two NMD reporter transgenes have been designed in this thesis, Version 1.0, and Version 2.0. Both versions were designed with an NMD-insensitive and an NMD-sensitive cassette in cis within the same transgene. Unlike many of the current NMD reporters, this method allows for an internally controlled NMD reporting system which bypasses the need to introduce two separate constructs into cells. In both designs, in order to escape or undergo NMD the Control and NMD cassettes are designed to exploit the general 55 nt rule of NMD, which states only PTCs greater than 55 nts upstream of an exon-exon junction are recognised and targeted for degradation by the NMD machinery.

In Version 1.0 (Chapter 4), Control$^{V1.0}$ is designed to express CFP^{NLS} regardless of cellular NMD levels and thus function as an internal control. Specifically, its NMD-refractive 3'UTR acts as a direct control for the NMD-targeted 3'UTR of NMD$^{V1.0}$, both of which are based on a previously published design (Pereverzev et al. 2015). The NMD-targeted 3'UTR contains intronic sequence that is designed to be spliced out. Following splicing and EJC introduction,

the stop codon will be positioned in the context of an NMD-targeted PTC i.e. greater than 55 nts upstream of an exon-exon junction. As such, transcripts expressed from NMD$^{V1.0}$ will be subject to NMD. Therefore, in cells with low NMD activity an increased level of TetR:EGFP^{NLS} will be expressed from NMD$^{V1.0}$ compared to cells with high NMD activity (Figure 6.1).

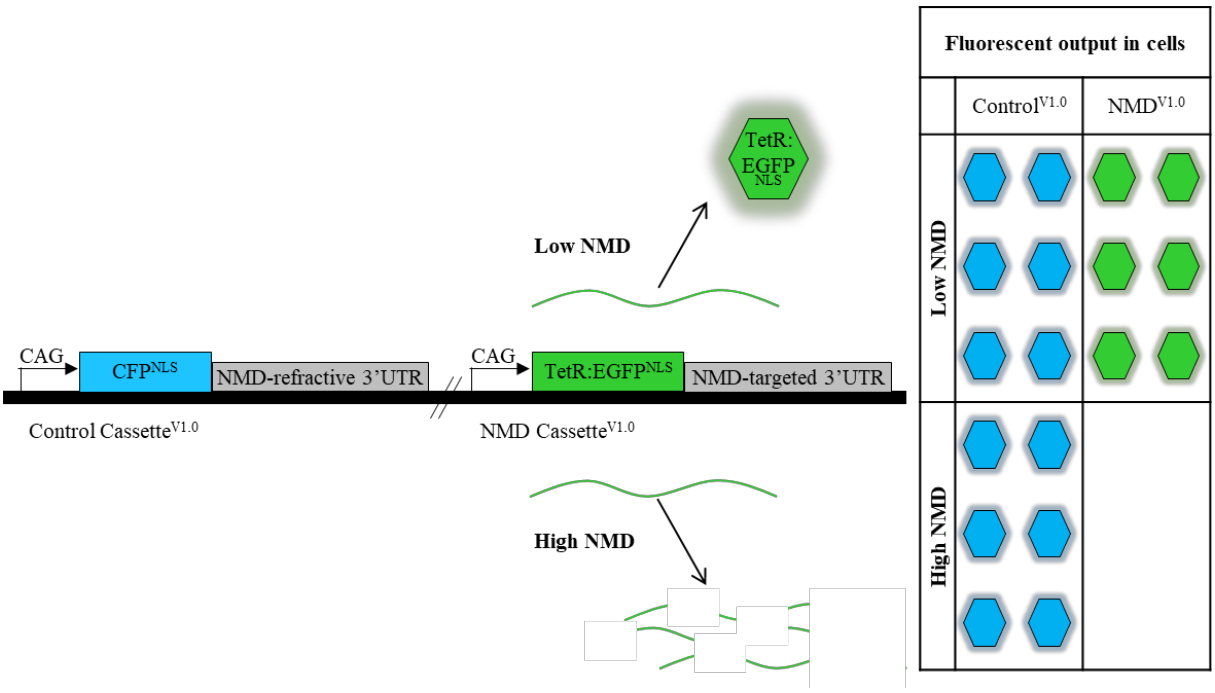


Figure 6.1: A simplified schematic representation of the design and fluorescent output of Control$^{V1.0}$ and NMD$^{V1.0}$ in cells with low and high NMD activity.

Control$^{V1.0}$ encodes CFP^{NLS}. Transcripts expressed from this cassette are not targeted by NMD, as such, this cassette serves as an internal control which expresses a constant level of CFP^{NLS} regardless of cellular NMD activity levels. NMD$^{V1.0}$ encodes the fusion protein TetR:EGFP^{NLS}. Transcripts expressed from this cassette are targeted for NMD due to the NMD-targeted 3'UTR. During splicing, intronic sequence within this 3'UTR is removed which will position the stop codon in the context of a PTC, thus subjecting the transcript to NMD. Therefore, in cells with low NMD activity, ongoing translation of TetR:EGFP^{NLS} encoding transcripts can occur and cells will express more green fluorescence compared to scenarios where cellular NMD levels are high. Ratios of EGFP:CFP provide an NMD activity reading, which can be taken at the single cell level due to nuclear localisation signals (NLS).

In Version 2.0 (Chapter 5), Control$^{V2.0}$ and NMD$^{V2.0}$ are largely based on a previously published design, however were adapted to provide a fluorescent rather than chemiluminescent output (Boelz et al. 2006). These cassettes have also been adapted to function within a single transgene, where Control$^{V2.0}$ will constitutively express the fluorescent protein/HBB fusion construct, ^{HA}CFP^{NLS}:HBB^{WT}. NMD$^{V2.0}$ also expresses a fluorescent protein/HBB fusion construct, in this case however, the *HBB* mini-gene encodes the NMD-targeted NS39 PTC, i.e. ^{FLAG}YFP^{NLS}:HBB^{NS39}. This means that expression from NMD$^{V2.0}$ should be influenced by cellular NMD activity levels, i.e. in cells where NMD activity is low, ^{FLAG}YFP^{NLS}:HBB^{NS39} should be efficiently expressed from NMD$^{V2.0}$, whereas in cells with high NMD activity, transcripts expressed from NMD$^{V2.0}$ should be degraded by NMD resulting in reduced expression of ^{FLAG}YFP^{NLS}:HBB^{NS39} (Figure 6.2).

Both Version 1.0 and Version 2.0 were originally designed to be integrated into the genome of mESCs. Stable systems combat problems associated with transfection efficiency and reproducibility inherent to transient transfection-based systems. Moreover, transient transfection-based reporter systems are only able to provide a snapshot of information from a large population of cells. This approach lacks temporal resolution and often, measurement of population averages also does not provide the necessary resolution to decipher molecular or cellular behaviour. Furthermore, it has been suggested that NMD substrates may be generally less efficiently degraded when expressed from transiently transfected reporter constructs compared to when expressed following stable genomic integration into a cell line (Gerbracht, Boehm & Gehring 2017). A stable NMD reporter system with single cell resolution will allow for endogenous NMD activity to be observed, and measured uninterrupted across lengthy experiments (e.g. differentiation protocols) and in high-throughput assays (e.g. drug and small molecule screens). This will allow differential responses to cellular NMD activity of single cells in distinct states to be investigated.

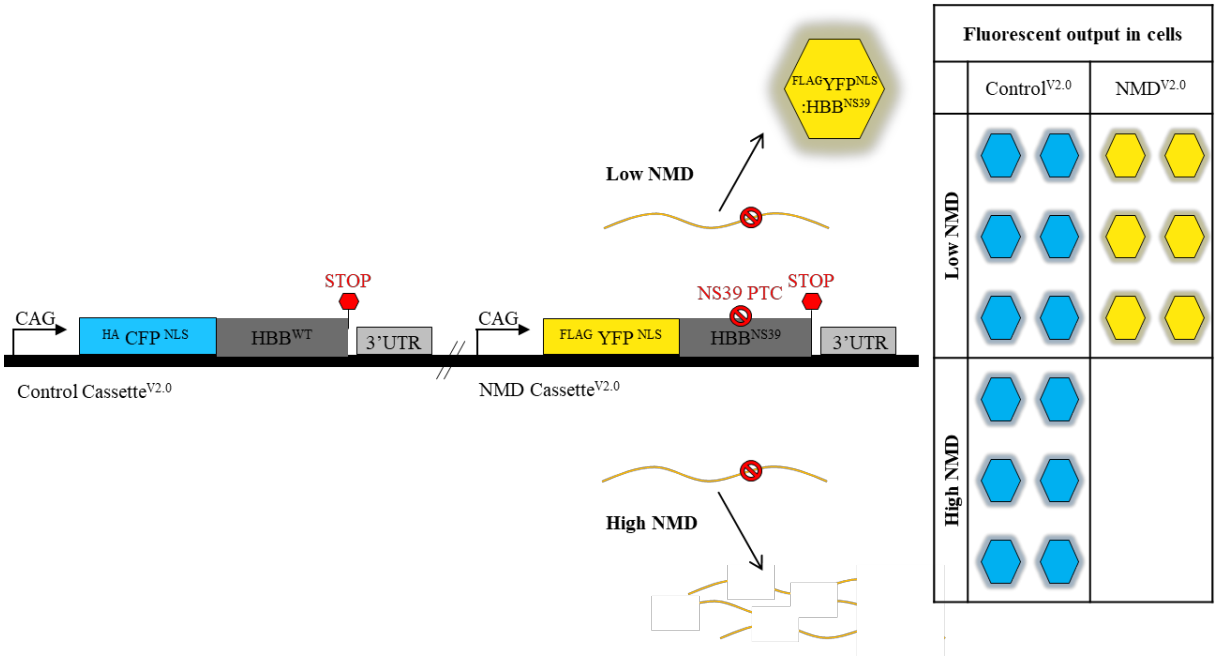


Figure 6.2: A simplified schematic representation of the design and fluorescent output of Control$^{V2.0}$ and NMD$^{V2.0}$ in cells with low and high NMD activity.

Control$^{V2.0}$ encodes the fluorescent protein/HBB fusion construct ^{HA}CFP^{NLS}:HBB^{WT}. Transcripts expressed from this cassette are not targeted by NMD and as such, Control$^{V2.0}$ expresses a constant level of ^{HA}CFP^{NLS}:HBB^{WT} regardless of cellular NMD activity. NMD$^{V2.0}$ encodes the fluorescent protein/HBB fusion construct, ^{FLAG}YFP^{NLS}:HBB^{NS39}. Transcripts expressed from this cassette are targeted for NMD due to an encoded NS39 PTC within exon 2 of the *HBB* mini-gene, therefore cells with low NMD activity will express greater levels of ^{FLAG}YFP^{NLS}:HBB^{NS39} from NMD$^{V2.0}$ compared to cells with high NMD activity. Levels of YFP can be normalised to CFP to provide an NMD activity reading. This can be taken at the single cell level due to nuclear localisation signals (NLS).

Unfortunately, investigations in Chapters 4 and 5 concluded that after establishing NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ mESCs, expression from genomic Control and NMD cassettes of these cells was unable to faithfully quantify NMD. Interestingly, through cassette testing in a transient setting (Chapters 4 and 5) it was found that NMD$^{V1.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$ did respond as expected to a reduction in NMD activity when transiently expressed from HEK293T cells. This perhaps suggested some preference of these cassettes to function in human rather than mouse cell types. To expand on these findings, this chapter aims to determine if the Control and NMD cassettes of Version 1.0 and Version 2.0 can be used to establish stable NMD Reporter HEK293T cell lines. And if so, determine whether these cells can provide a means to faithfully visualise and quantify endogenous NMD activity at a single cell level.

6.2 Results

6.2.1 Using random genomic transgene integration to establish NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cell lines

The cassettes of Transgene$^{V1.0}$ or Transgene$^{V2.0}$ were not designed for loci-targeted stable integration into HEK293T cells. However, transgenes can be randomly integrated into cellular genomes at a low frequency upon transfection. In this process, a transgene of interest, preferably also encoding a selectable marker, is introduced into cells as it would be for transient transfection-based experiments. Under these conditions, a small percentage of introduced DNA (which can be increased by using linear instead of circular DNA) will be integrated into locations within the genome of a cell at random. A population of cells with successful genomic integration(s) can then be isolated via expression of the integrated selectable marker and expanded to establish a stable cell line. Therefore, in order to establish stable NMD Reporter HEK293T cell lines, expression plasmids encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) were transiently transfected into HEK293T cells and these cells were then screened for random integration events.

Firstly, expression plasmids encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) alongside an expression plasmid encoding *PurR* (pQCXIP) were linearised using restriction endonuclease digestion to promote integration of an entire uninterrupted transgene (Figure 6.3A). The *PurR* gene expresses puromycin-*N*-acetyltransferase to confer a selectable resistance to puromycin. These plasmids and their encoded products are summarised in Table 6.1.

Table 6.1: A Description of expression plasmids used to establish stable NMD reporter HEK293T cell lines and the proteins they encode

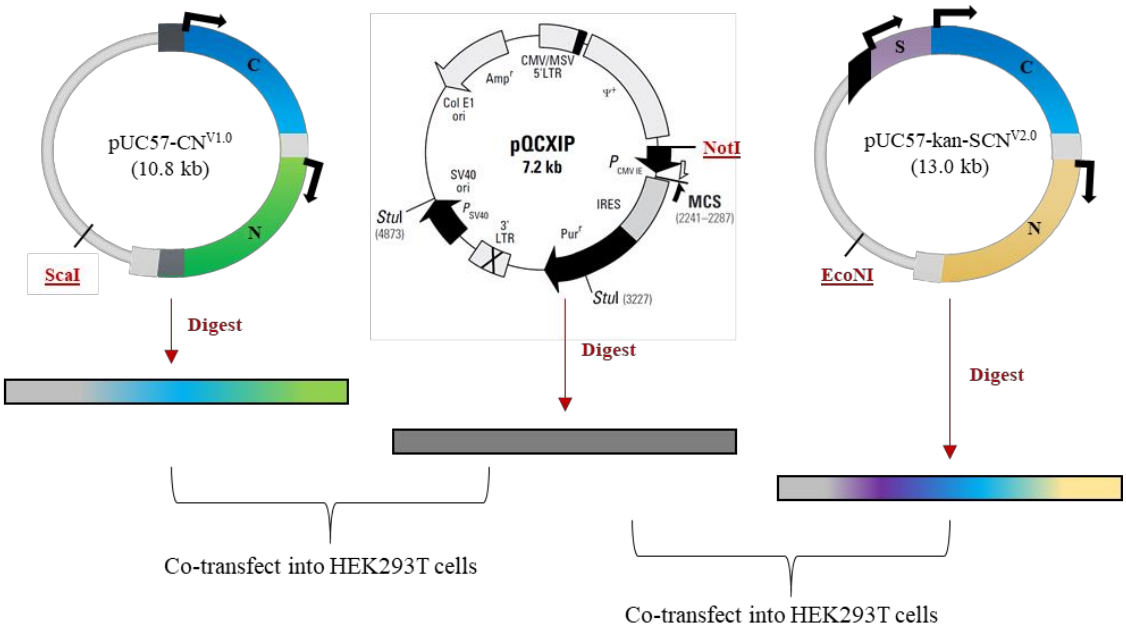
Cell line	**Integrated plasmids**	**Function**	**Encoded protein(s)**	**Expected protein size(s)**
NMD Reporter$^{V1.0}$ HEK293T cells	pUC57-CN$^{V1.0}$	Expresses Control$^{V1.0}$ and NMD$^{V1.0}$ of Transgene$^{V1.0}$.	CFP^{NLS}, TetR:EGFP^{NLS}	~28 kDa, ~52 kDa
	pQCXIP	Encodes *PurR* to confer puromycin resistance.	Puromycin-*N*-acetyltransferase	~22 kDa
NMD Reporter$^{V2.0}$ HEK293T cells	pUC57-kan-SCN$^{V2.0}$	Expresses Selection$^{V2.0}$, Control$^{V2.0}$ and NMD$^{V2.0}$ of Transgene$^{V2.0}$.	^{HA}CFP^{NLS}:HBB^{WT}, ^{FLAG}YFP^{NLS}:HBB^{NS39}	~46 kDa, ~35 kDa
	pQCXIP	Encodes *PurR* to confer puromycin resistance.	Puromycin-*N*-acetyltransferase	~22 kDa

Linearised cassette expression plasmids, alongside the linearised pQXCIP plasmid was co-transfected into HEK293T cells. 48 hours post transfection, cells underwent puromycin selection for 48 hours followed by expansion for 14 days (Figure 6.3B).

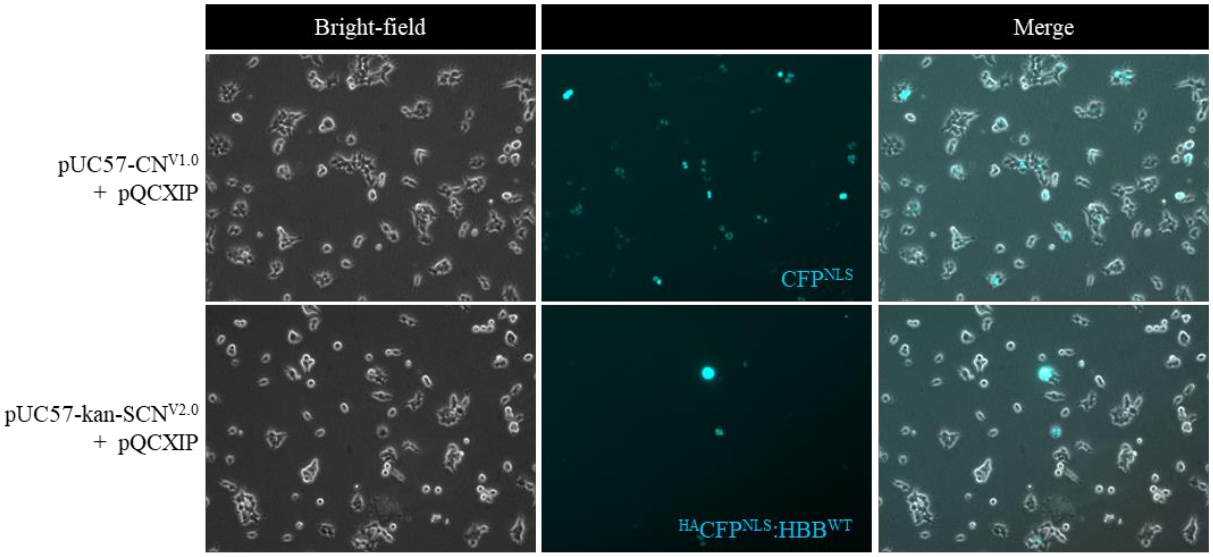
At this point there were two possible populations of cells; (1) cells resistant to puromycin due to genomic integration(s) of the linearised pQCXIP DNA sequence and (2) cells resistant to puromycin due to genomic integration(s) of the linearised pQCXIP DNA sequence which also express fluorescent signals from genomic integration(s) of the linearised DNA sequence of either pUC57-CN$^{V1.0}$ (CFP and EGFP) or pUC57-kan-SCN$^{V2.0}$ (CFP and YFP) (Figure 6.3B). The latter population of cells which have successfully integrated Control and NMD Cassettes of either Version 1.0 or Version 2.0 into their genomes was desired. To isolate these cells FACS of live cells was conducted based on CFP expression (Figure 6.3C).

It was found that ~10% of cells co-transfected with linearised pUC57-CN$^{V1.0}$ and linearised PQCXIP expressed CFP, while ~0.5% of cells co-transfected with linearised pUC57-kan-SCN$^{V2.0}$ and linearised PQCXIP expressed CFP. Following sorting, these cells were expanded to establish either NMD Reporter$^{V1.0}$ HEK293T cells or NMD Reporter$^{V2.0}$ HEK293T cells respectively (Figure 6.3D). To note, it was observed that despite two flanking nuclear localisation signals, CFP expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells was not completely restricted to the nucleus (Figure 6.3D).

A) **STEP 1: Linearisation and introduction of expression plasmids into HEK293T cells**



B) **STEP 2: Puromycin selection and expansion of a mixed population of co-transfected HEK293T cells**



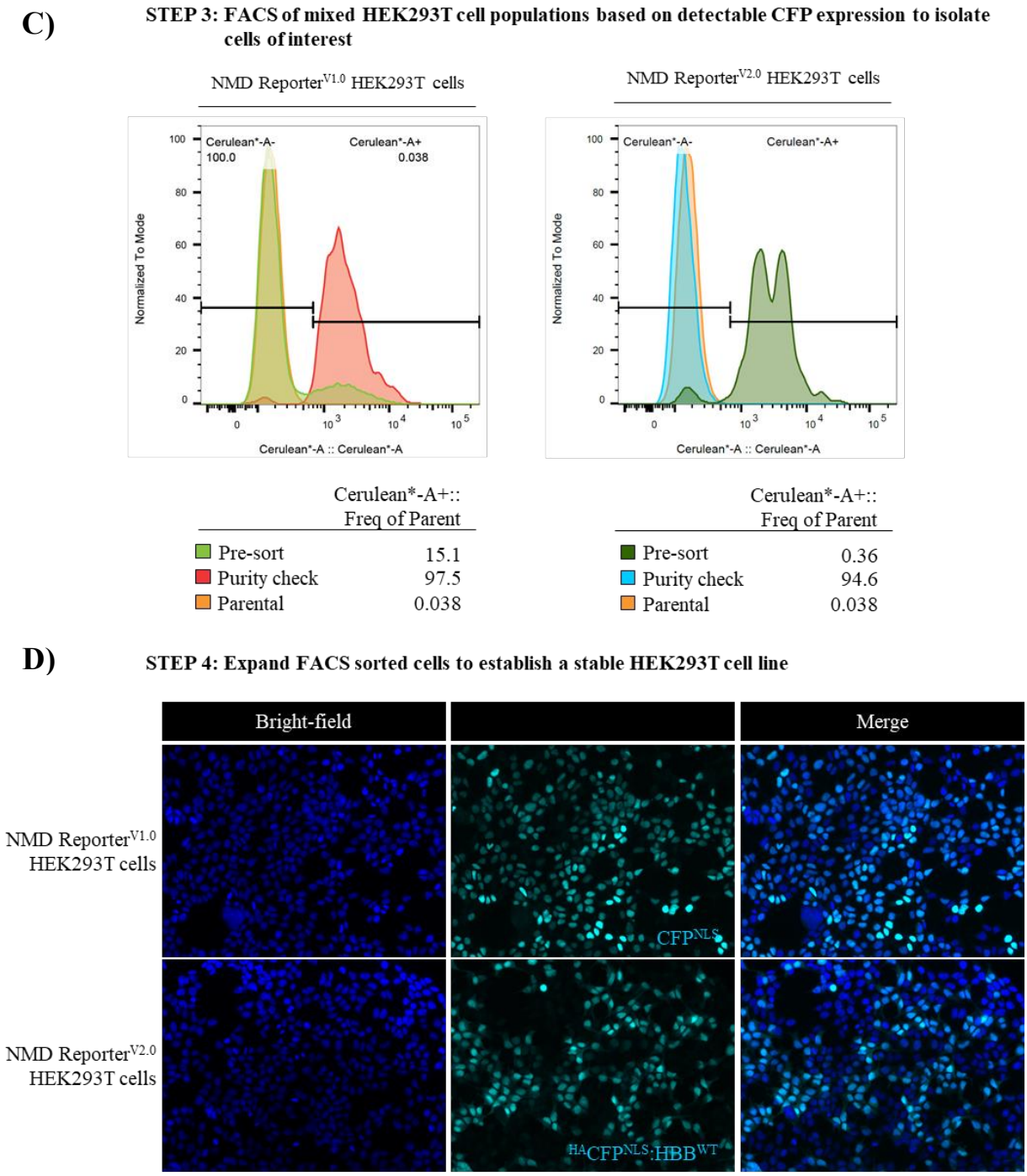


Figure 6.3: A step-by-step description of the workflow used to randomly integrate expression plasmids encoding Control$^{V1.0}$ and NMD$^{V1.0}$ or Control$^{V2.0}$ and NMD$^{V2.0}$ into the genome of HEK293T cells.

(A) A schematic representation of the workflow used to facilitate random genomic integration of Control$^{V1.0}$ and NMD$^{V1.0}$ or Control$^{V2.0}$ and NMD$^{V2.0}$ into the genome of HEK293T cells. **(B)** HEK293T cells were co-transfected with a linearised expression plasmid for a gene that confers puromycin resistance (*Pur^{r}*) (pQCXIP) and either a linearised expression plasmid for

Figure 6.3 continued…

Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$). These cells were then subject to 48 hours of puromycin selection and 14 days of expansion. After expansion, representative brightfield and fluorescence microscopy images were taken, cells expressing only *Pur^{r}* due to random integration(s) of linearised pQCXIP into their genome show no fluorescence. Cells expressing *Pur^{r}* and CFP^{NLS} due to random integration(s) of linearised pQCXIP and pUC57-CN$^{V1.0}$ into their genome express cyan fluorescence (cyan) and cells expressing *Pur^{r}* and ^{HA}CFP^{NLS}:HBB^{WT} due to random genomic integration(s) of linearised pQCXIP and pUC57-kan-SCN$^{V2.0}$ also express cyan fluorescence (cyan). **(C)** To isolate the cells expressing CFP and *Pur^{r}* (i.e. NMD Reporter$^{V1.0}$ or NMD Reporter$^{V2.0}$ HEK293T cells) from those expressing only *Pur^{r}* (no fluorescence) FACs analysis based on CFP (cerulean) expression was conducted. FACs histograms summarise the results from this analysis. **(D)** Representative fluorescence microscopy images from NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells following FACS and expansion. NMD Reporter$^{V1.0}$ HEK293T cells express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$ (not shown). NMD Reporter$^{V2.0}$ HEK293T cells express ^{HA}CFP^{NLS}:HBB^{WT} from the genomic Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (not shown). Cell nuclei were counterstained with DAPI (blue).

6.2.2 Functional testing of genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells

6.2.2.1 Testing if a single and/or three SV40 polyA sequences can efficiently terminate transcription downstream of a CAG promoter

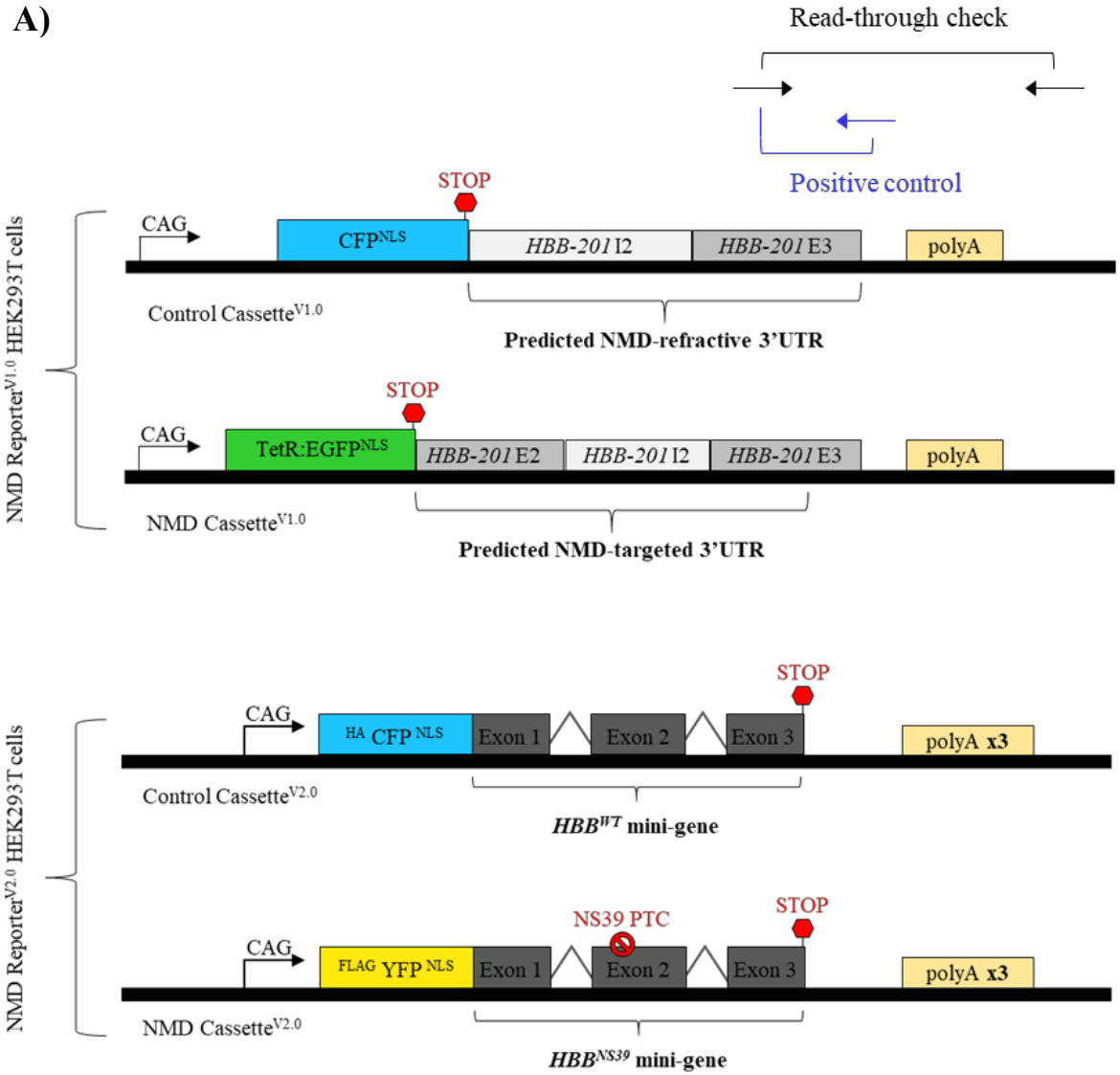
The SV40 polyA is a sequence-based terminator, its role is to define the end of a transcriptional unit and initiate the release of newly synthesised RNA from the transcription machinery. If the SV40 polyA(s) cannot efficiently terminate transcription from the designed cassettes, transcriptional read-through at the SV40 polyA site(s) of one cassette could influence expression of the downstream cassette. Genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells contain only one SV40 polyA within their 3'UTR. Previous studies (Chapter 4) have shown that when these same cassettes were transiently expressed in HEK293T cells, transcription termination was inefficient. In comparison, when these cassettes were expressed following genomic integration in NMD Reporter$^{V1.0}$ mESCs, transcription termination was efficient.

To assess the efficiency of transcription termination from Control$^{V1.0}$ and NMD$^{V1.0}$ when these cassettes are integrated into the genome of NMD Reporter$^{V1.0}$ HEK293T cells, three PCR primers (one forward and two reverse primers) were designed to perform in the same reaction. Two primers either amplify a sequence within the transcriptional unit to act as a positive control for the PCR conditions (positive control) or span the SV40 polyA site to assess if transcription termination is efficient (read-through check) (Figure 6.4A). Read-through check primers will only amplify a product if RNA polymerase read-through beyond the SV40 polyA sequence is occurring i.e. transcription termination is not efficient. cDNA reverse transcribed from the RNA of NMD Reporter$^{V1.0}$ HEK293T cells, was subject to PCR using these primers.

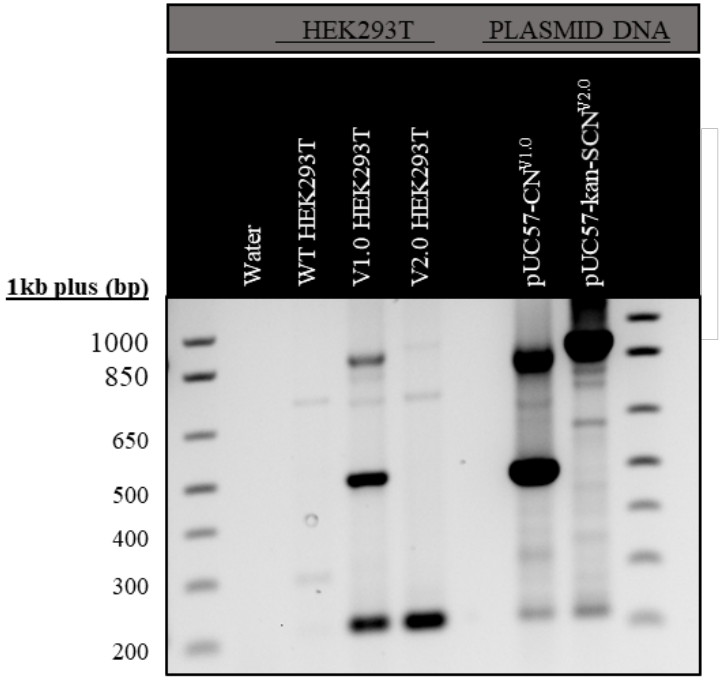
PCR product was then visualised by agarose gel electrophoresis. Both sets of primers clearly amplified a product concluding that one SV40 polyA sequence was unable to efficiently terminate transcription downstream of the CAG promoter in these cells (Figure 6.4B). These results were comparable to those observed from transient experiments using HEK293T cells, but unlike the results observed from NMD Reporter$^{V1.0}$ mESCs cells. This may suggest, that at least in this context, efficient transcription termination could be a cell or species type-specific event.

The same PCR was performed using cDNA reverse transcribed from the RNA of NMD Reporter$^{V2.0}$ HEK293T cells as a template. It was observed that in these cells, read-through check primers did not amplify a strong product (Figure 6.4B). This means that in these cells the presence of three SV40 polyA sequences can facilitate efficient transcription termination downstream of a CAG promoter. These results align with previous findings (Chapter 5) which showed that three SV40 polyA sequences could efficiently terminate transcription of Control$^{V2.0}$ and NMD$^{V2.0}$ when these cassettes were expressed transiently in HEK293T cells, and when they were expressed from stable NMD Reporter$^{V2.0}$ mESCs.

A)



B)



	NMD Reporter V1.0 HEK293T cells (V1.0 HEK293T)		NMD Reporter V2.0 HEK293T cells (V2.0 HEK293T)	
	C V1.0	**N V1.0**	**C V2.0**	**N V2.0**
Read-through check	No band **OR** 476 bp if read-through		No band **OR** 797 bp if read-through	No band **OR** 866 bp if read-through
Positive control	218 bp		218 bp	218 bp

Figure 6.4: Three SV40 polyA sequences downstream of a CAG promoter can efficiently terminate transcription in stable NMD Reporter$^{V2.0}$ HEK293T cells.

(A) A schematic representation of primers designed to assess transcriptional read through at the SV40 polyA site(s) of the genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ HEK293T cells (C$^{V1.0}$ and N$^{V1.0}$) and NMD Reporter$^{V2.0}$ HEK293T cells (C$^{V2.0}$ and N$^{V2.0}$). **(B)** Reverse transcribed cDNA from the RNA of wild-type HEK293T cells (WT HEK293T), NMD Reporter$^{V1.0}$ HEK293T cells (V1.0 HEK293T) or NMD Reporter$^{V2.0}$ HEK293T cells (V2.0 HEK293T) was subject to PCR using the three described primers. Plasmid DNA encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) was subject to the same PCR to identify products indicative of RNA polymerase read-through at the SV40 polyA site(s). PCR amplified products were visualised by agarose gel electrophoresis. The table outlines the possible PCR product sizes. Banding patterns confirm that transcription termination is efficient for genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells which contain three SV40 polyA sequences within their 3'UTRs, but not for genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells which contain only one SV40 polyA sequence within their 3'UTR.

6.2.2.2 Assessing splicing of transcripts expressed from genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells

Genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells all contain intronic sequences that should be removed during splicing of the pre-mRNA. A failure of these transcripts to undergo splicing will prevent the introduction of EJCs onto the transcripts needed for NMD recognition of a PTC. On the other hand, inaccurate splicing can lead to the formation of mRNA containing an aberrant NMD-targeted PTC. Either of these outcomes can interfere with accurate fluorescent NMD quantitation.

Genomic Control $^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells, contain a partial sequence of *HBB* within their 3'UTR. Previous transient investigations in HEK293T cells showed that the predicted NMD-refractive 3'UTR of Control$^{V1.0}$ remained largely unspliced, while the predicted NMD-targeted 3'UTR of NMD$^{V1.0}$ was spliced efficiently.

To now assess 3'UTR splicing of genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells, these cells were first transfected with Control or *UPF1* siRNA and cDNA was then reverse transcribed from the RNA of these cells. This cDNA was used as a template in PCR-based investigations with primers designed to flank the 3'UTR intronic sequence (splicing check) (Figure 6.5A). Plasmid DNA encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) was also subject to the same PCR as a control to identify unspliced product. If the predicted NMD-refractive or NMD-targeted 3'UTR sequence is efficiently spliced this PCR will amplify a single product at the expected size for each cassette (Figure 6.5B).

PCR amplified product was visualised via agarose gel electrophoresis. Genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter HEK293T cells were each observed to produce two splice products, one indicative of correctly spliced transcripts and the other of unspliced transcripts (Figure

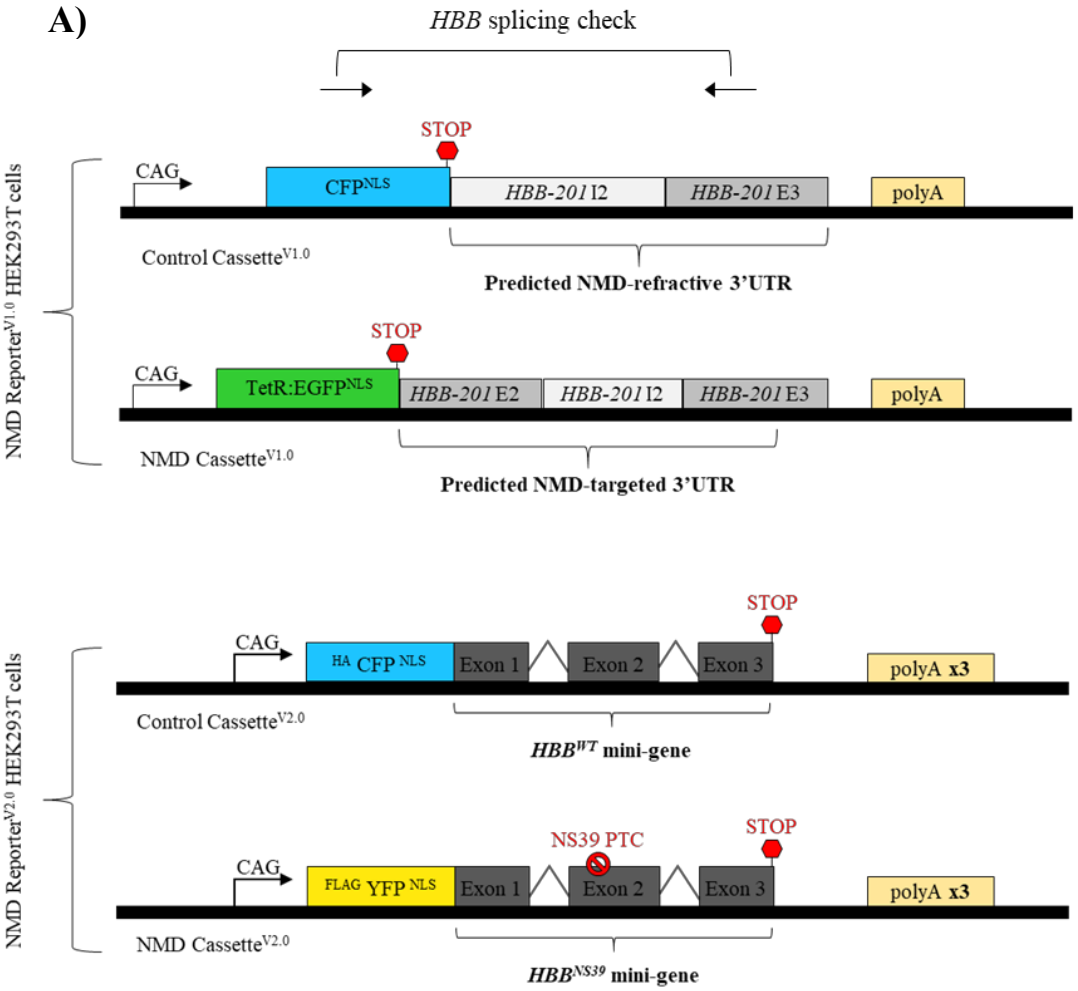
6.5B). This was in line with splicing in HEK293T cells under transient conditions, where the majority of transcripts expressed from genomic Control$^{V1.0}$ appeared unspliced, whilst the majority of transcripts expressed from genomic NMD$^{V1.0}$ appeared correctly spliced (Figure 6.5B). The observed splicing did not appear to be influenced by cellular UPF1 levels (Figure 6.5B).

Genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells, both encode a fluorescent protein/ HBB fusion construct. In both cases the *HBB* mini-gene contains all three exons and two introns of endogenous *HBB*. Previous transient transfection-based investigations in HEK293T cells showed that the transcripts from both cassettes were predominantly spliced correctly, however 'intron 1 retained' transcripts, and less prominently, 'intron 2 retained' and unspliced transcripts were also detected.

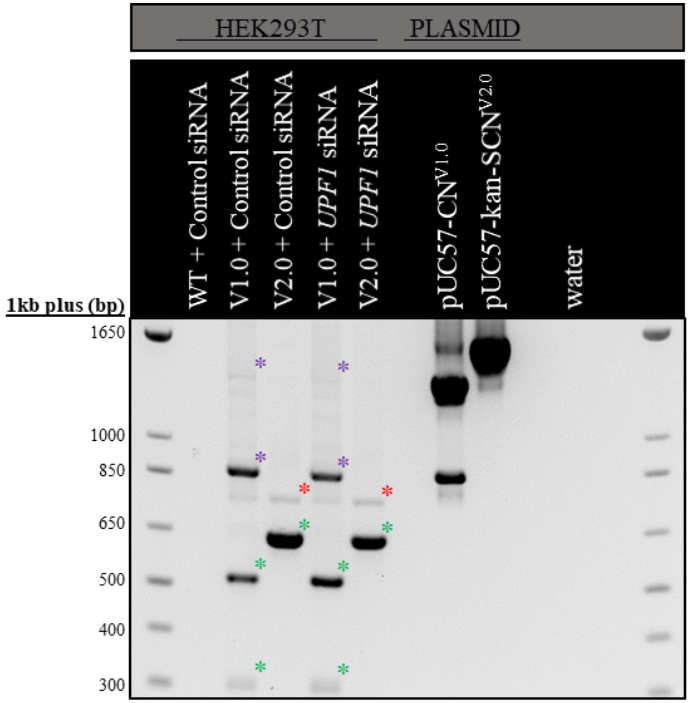
To now assess splicing of transcripts expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells, these cells were first transfected with Control or *UPF1* siRNA and cDNA was reverse transcribed from the RNA of these cells. This cDNA was used as a template in PCR-based investigations with primers designed to flank the *HBB* mini-gene (splicing check) (Figure 6.5A). Plasmid DNA encoding Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-SCN$^{V2.0}$) was also subject to the same PCR as a control to identify unspliced product. If the *HBB* mini-gene is efficiently spliced this PCR will amplify a single product at the expected size of 644 bp (Figure 6.5B).

PCR amplified product was visualised by agarose gel electrophoresis. Genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T were each observed to produce two splice products. The dominant product was indicative of correctly spliced transcripts while a second product was suggestive of 'intron 1 retained' transcripts was also detected (Figure 6.5B). Splicing did not appear to be influenced by cellular UPF1 levels (Figure 6.5B).

A)



B)



	NMD Reporter V1.0 HEK293T cells (V1.0)		NMD Reporter V2.0 HEK293T cells (V2.0)	
	C V1.0	**N V1.0**	**C V2.0**	**N V2.0**
Unspliced / plasmid *	879 bp	1376 bp	1624 bp	
Intron 2 retained*	-	-	1494 bp	
Intron 1 retained*	-	-	774 bp	
Exon 2 skipped*	-	-	421 bp	
Correct splicing*	308 bp	526 bp	644 bp	

Figure 6.5: Investigation of pre-mRNA splicing reveals that all cassettes apart from Control$^{V1.0}$ are predominantly spliced correctly upon stable genomic integration into HEK293T cells.

(A) A schematic representation of primers designed to assess splicing of transcripts expressed from genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells (HBB splicing check). **(B)** Wildtype (WT), NMD Reporter$^{V1.0}$ (V1.0) and NMD Reporter$^{V2.0}$ (V2.0) HEK293T cells were transfected with Control or *UPF1* siRNA. cDNA reverse transcribed from the RNA of these cells was subject to PCR with HBB splicing check primers. Plasmid DNA encoding Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) was subject to the same PCR to identify unspliced transcripts. PCR amplified products were visualised by agarose gel electrophoresis. The table outlines the PCR product sizes of all potential splice products and is colour coded in reference to the asterisks on the agarose gel image. Banding patterns confirm that transcripts expressed from NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells all show some level of mis-splicing, however, only transcripts produced from genomic Control$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells appear to be predominantly unspliced.

6.2.2.3 *Assessing stability and proteasomal degradation of proteins expressed from genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells*

Proteasomes refer to protein complexes which target and degrade ubiquitin-tagged proteins. Proteins expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of stable NMD Reporter$^{V2.0}$ mESCs, were found to be unstable due to degradation by the proteasome. This was not observed for proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of stable NMD Reporter$^{V1.0}$ mESCs.

To investigate if any proteins expressed from genomic Control or NMD cassettes of NMD Reporter$^{V1.0}$ or NMD Reporter$^{V2.0}$ HEK293T cells were targeted for proteasomal degradation, these cells were treated with either cycloheximide or MG132. Cycloheximide inhibits translation allowing protein stability to be assayed, while MG132 inhibits the proteasome therefore stabilising proteins that would normally be degraded by this pathway. As such, an unstable protein should rapidly decrease in abundance following treatment with cycloheximide, whilst a protein targeted for proteasomal degradation should rapidly increase in abundance following treatment with MG132. If these treatments are successful, expression of the endogenous protein β-catenin (a known target of rapid degradation by the proteasome) will decrease following cycloheximide treatment and increase following MG132 treatment.

Following treatments, protein isolated from NMD Reporter$^{V1.0}$ or NMD Reporter$^{V2.0}$ HEK293T cells were analysed by western blot. It was observed that proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells were stable, and not targeted for proteasomal degradation (Figure 6.6A). Comparatively, proteins expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells were seen to be unstable due to efficient proteasomal degradation (Figure 6.6B). Unfortunately, the antibodies used were unable to detect proteins produced from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells.

The results observed for proteins expressed from genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells, largely mirrored what was seen for proteins expressed from genomic Control and NMD cassettes of NMD Reporter$^{V1.0}$ and NMD Reporter $^{V2.0}$ mESCs respectively (Chapter 4 & 5). This suggests that proteasomal targeting and degradation of at least $^{HA}CFP^{NLS}$:HBB^{WT} can occur in both human and mouse cells.

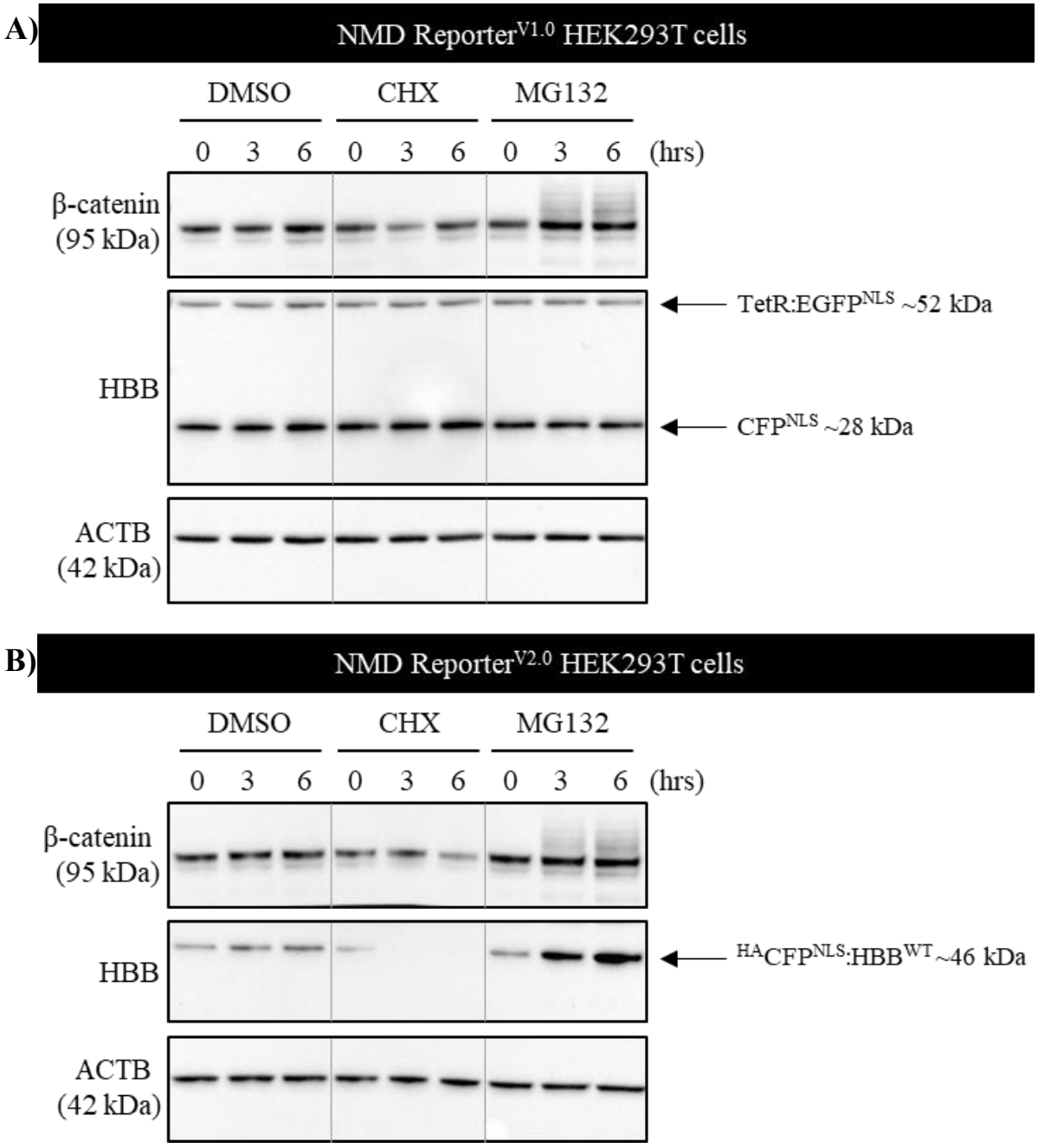


Figure 6.6: Protein expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells is unstable due to proteasomal degradation.

NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells were subjected to treatment with either cycloheximide (CHX) or MG132 to inhibit translation or the proteasome respectively. NMD Reporter$^{V1.0}$ HEK293T cells express CFP^{NLS} from genomic Control$^{V1.0}$ and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$. NMD Reporter$^{V2.0}$ HEK293T cells express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$. Following treatments, total protein was isolated from these cells at different time points, separated by SDS-PAGE and analysed by western blot with antibodies to detect to βcatenin, HBB, GFP and the loading control ACTB (stripped and re-probed). **(A)** Protein expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells was seen to be stable and unaffected by inhibition of the proteasome. **(B)** Protein expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells was seen to be unstable due to rapid degradation by the proteasome. Protein expressed from genomic NMD$^{V2.0}$ of these cells was unable to be detected.

6.2.2.4 Testing protein expression from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells in response to an siRNA mediated reduction in UPF1

So far, transcripts expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells have shown similar levels of RNA polymerase read-through at the SV40 polyA site and similar splicing patterns when compared to their transient expression in HEK293T cells. Furthermore, proteins translated from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells are stable and not targeted for proteasomal degradation.

Transcripts expressed from NMD$^{V1.0}$ contain a stop codon which following splicing of the 3'UTR and EJC deposition, is positioned in the context of an NMD targeted PTC i.e. 55 nts upstream of an exon-exon junction. As such, these transcripts are predicted to undergo NMD which should ultimately result in a loss of TetR:EGFP^{NLS} protein expression from NMD$^{V1.0}$. Comparatively, transcripts expressed from Control$^{V1.0}$ contain a stop codon that should not be recognised by the NMD machinery. Additionally, pre-mRNA splicing investigations showed that transcripts expressed from Control$^{V1.0}$ remain predominantly unspliced and thus free of any EJCs further preventing NMD recognition (Section 6.2.2.2). These transcripts should therefore result in constitutive expression of CFP^{NLS} regardless of cellular NMD activity levels.

Previous investigations revealed that protein expression from NMD$^{V1.0}$ showed some sensitivity to reduced levels of UPF1 (i.e. reduced NMD activity) when expressed transiently in HEK293T cells and when expressed from stable NMD Reporter$^{V2.0}$ mESCs (Chapter 4). Comparatively, protein expressed from Control$^{V1.0}$ was only unresponsive to reduced levels of UPF1 when expressed from stable NMD Reporter$^{V1.0}$ mESCs.

To now investigate the response of these proteins, when expressed from stable NMD Reporter$^{V1.0}$ HEK293T cells, these cells were transfected with either Control or *UPF1* siRNA.

Protein isolated from these cells was analysed by western blot. This showed that following an almost complete loss of UPF1 protein expression, TetR:EGFP^{NLS} expressed from genomic NMD$^{V1.0}$ was significantly increased by 3.78-fold (Figure 6.7). Comparatively, CFP^{NLS} expressed from genomic Control$^{V1.0}$ remained unchanged (Figure 6.7).

Fluorescence emitted from these cells was also analysed via FACS and visualised by fluorescence microscopy. The results obtained supported western blot analysis, and showed that following a loss of UPF1, reporter output, as measured by levels of EGFP expressed from genomic NMD$^{V1.0}$ (TetR:EGFP^{NLS}) normalised to CFP expressed from genomic Control$^{V1.0}$ (CFP^{NLS}) was significantly increased by ~2-fold (Figure 6.8).

Collectively, these results indicate that when expressed from stable NMD Reporter$^{V1.0}$ HEK293T cells, genomic Control$^{V1.0}$ and NMD$^{V1.0}$ are functional in terms of response to reduced NMD activity.

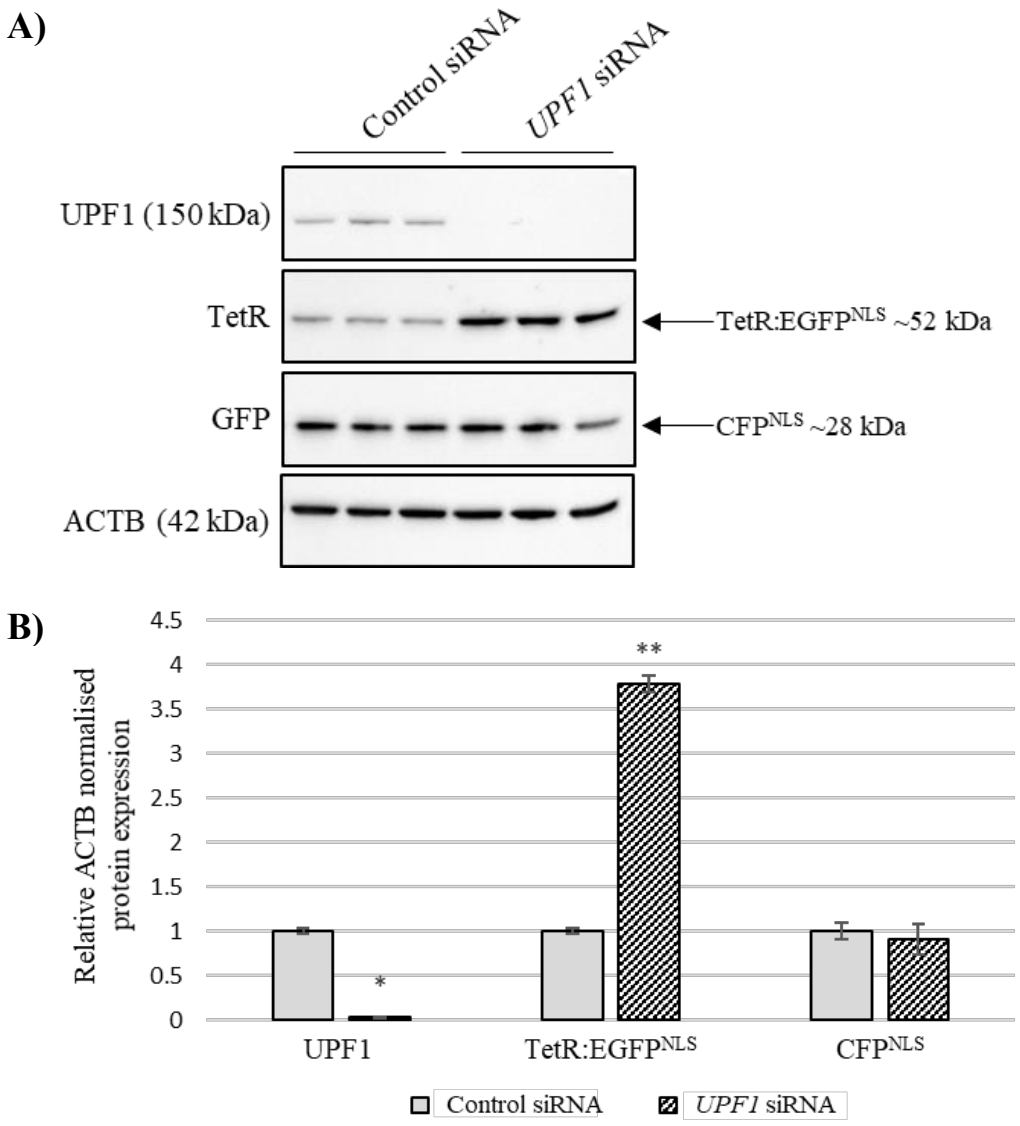


Figure 6.7: Proteins expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells respond as expected to an siRNA mediated reduction in UPF1.

(A) Three separate replicates of NMD Reporter$^{V1.0}$ HEK293T cells were transfected with Control or *UPF1* siRNA. These cells express CFP^{NLS} from genomic Control$^{V1.0}$ and TetR:EGFP^{NLS} from genomic NMD$^{V1.0}$. Total protein isolated from these cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1, TetR, GFP and the loading control ACTB (stripped and re-probed). **(B)** Densitometric analysis of this western blot revealed a significant reduction in UPF1 levels to approximately 3% of normal. This was accompanied by a significant 278% increase in TetR:EGFP^{NLS} expressed from genomic NMD$^{V1.0}$ and a 10% decrease in CFP^{NLS} expressed from genomic Control$^{V1.0}$ which did not reach significance. *P < 0.05 and ** P< 0.01 by Student's two tailed t-test.

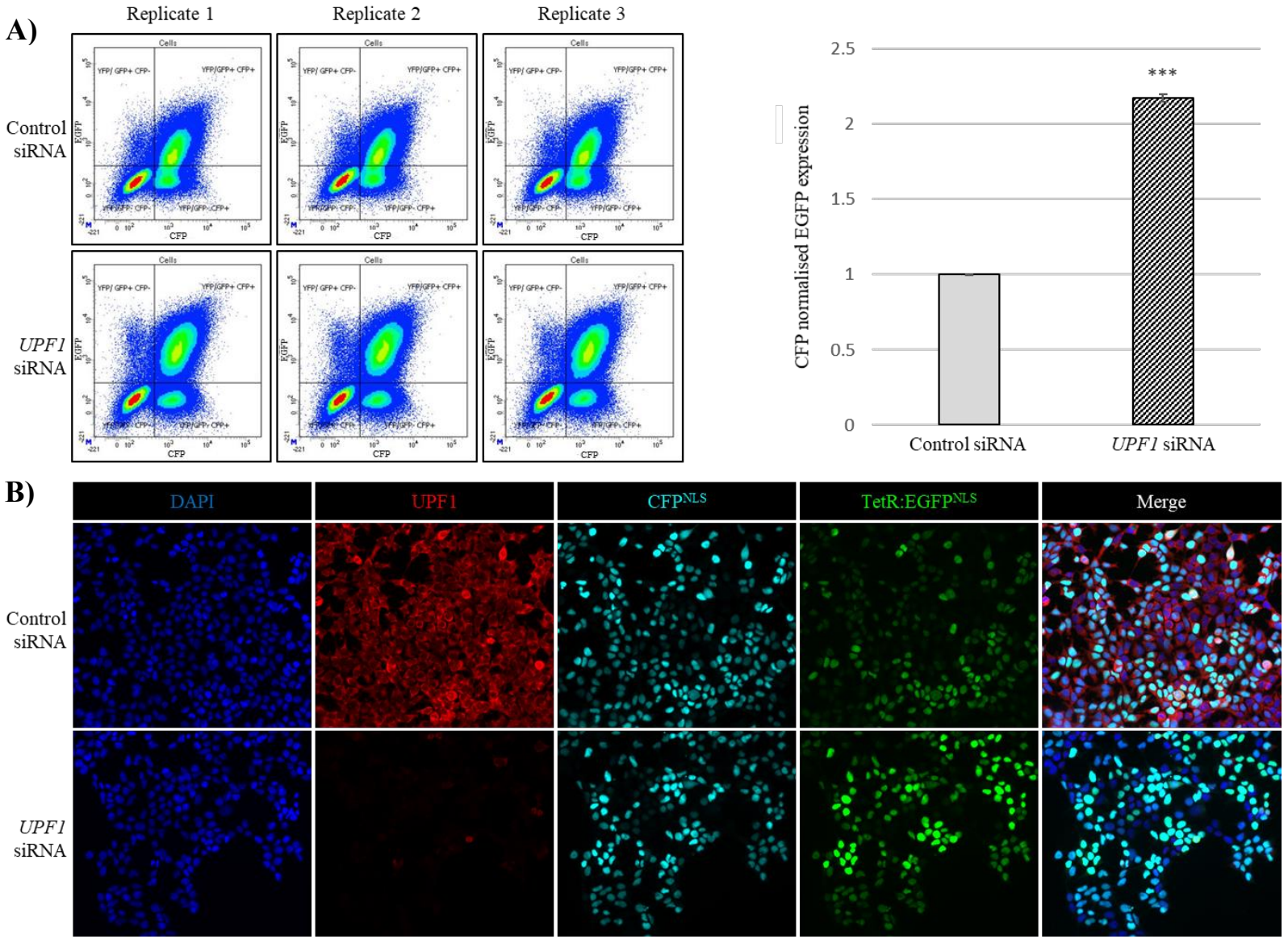

A)
Replicate 1
Replicate 2
Replicate 3
Control siRNA
UPF1 siRNA
Cells
YFP/ GFP+ CFP-
YFP/GFP+ CFP+
YFP/GFP- CFP-
YFP/GFP- CFP+
EGFP
CFP
CFP normalised EGFP expression
2.5
2
1.5
1
0.5
0
***
Control siRNA
UPF1 siRNA
B)
DAPI
UPF1
CFPNLS
TetR:EGFPNLS
Merge
Control siRNA
UPF1 siRNA

Figure 6.8: Fluorescence expressed from genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells responds as expected to an siRNA mediated reduction in UPF1.

Three separate replicates of NMD Reporter$^{V1.0}$ HEK293T cells were transfected with Control or *UPF1* siRNA. These cells express CFP^{NLS} from genomic Control$^{V1.0}$ and TetR:EGFP^{NLS} from genomic NMD$^{V2.0}$. **(A)** Fluorescence expression from these cells was quantified by FACS analysis. The left panel shows EGFP vs CFP density plots across all three replicates. The right panel compiles this information into a graph representing CFP fluorescence expression normalised EGFP. These results depict a significant 217% increase in CFP normalised EGFP fluorescence expression from NMD Reporter$^{V1.0}$ HEK293T cells following a loss of UPF1 protein. *** P< 0.01 by Student's t-test. **(B)** Representative fluorescence microscopy images of these cells. The cells express CFP^{NLS} from genomic Control$^{V1.0}$ (cyan) and TetR:EGFP^{NLS} from genomic NMD$^{V2.0}$ (green). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue).

6.2.2.5 Testing protein expression from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells in response to an siRNA mediated reduction in UPF1

Control$^{V2.0}$ and NMD$^{V2.0}$ both encode fluorescent protein/HBB fusion constructs. Specifically, Control$^{V2.0}$ encodes wild-type HBB, while NMD$^{V2.0}$ encodes HBB containing the NMD-targeted NS39 PTC. Under control circumstances, this is designed to facilitate constitutive protein expression from Control$^{V2.0}$, while resulting in reduced protein expression from NMD$^{V2.0}$ due to NMD targeting and degradation of the NS39 PTC containing transcripts.

Previous investigations (Chapter 5) revealed that protein expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ responded as expected to a reduction of UPF1 upon transient transfection into HEK293T cells. Unfortunately, when these cassettes were expressed from stable NMD Reporter$^{V2.0}$ mESCs, any response to changes in cellular NMD activity was masked by proteasomal degradation of the proteins produced from genomic Control$^{V1.0}$ and NMD$^{V1.0}$. Proteins expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells were also found to be targeted for proteasomal degradation (Section 6.2.2.3), as such it is likely that proteins expressed from genomic NMD$^{V1.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells are targeted for proteasomal degradation. Therefore, it was not predicted that NMD Reporter$^{V2.0}$ HEK293T cells would be able to faithfully report on changes in cellular NMD activity.

To test if this prediction was true, NMD Reporter$^{V2.0}$ HEK293T cells were transfected with either Control or *UPF1* siRNA to reduce NMD activity. Western blot analysis showed that across three replicates, this method significantly reduced UPF1 protein levels to approximately 2.5% of that in control conditions (Figure 6.9). The cells were then analysed by FACs and fluorescence microscopy. Both methods showed that following a reduction in UPF1, protein expression of ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ were both slightly increased, i.e. no change in reporter output, as measured

by levels of YFP expressed from genomic NMD$^{V2.0}$ normalised to CFP expressed from genomic Control$^{V2.0}$ (Figure 6.10). Overall, these results identified the system as being unsurprisingly non-responsive to a reduction in NMD activity.

It was noted, however, that despite encoding nuclear localised CFP, CFP expression from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells was not restricted to the nucleus, and appeared to provide a much stronger fluorescence signal than YFP expressed from genomic NMD$^{V2.0}$ (Figure 6.10B).

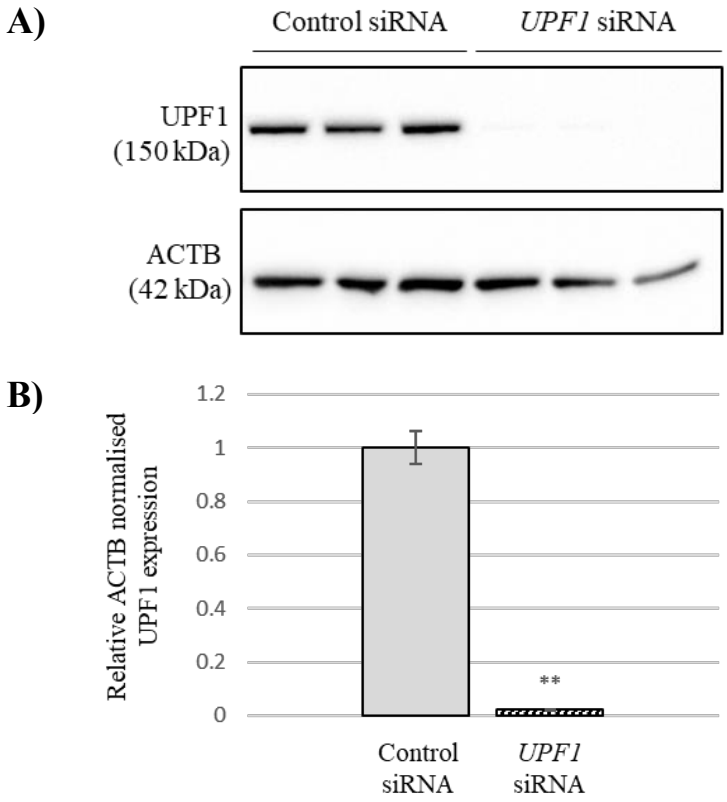


Figure 6.9: UPF1 siRNA effectively reduces UPF1 protein levels in NMD Reporter$^{V2.0}$ HEK293T cells.

Three separate replicates of NMD Reporter$^{V2.0}$ HEK293T cells were transfected with Control or *UPF1* siRNA. **(A)** Total protein was isolated from these cells, separated by SDS-PAGE and analysed by western blot using antibodies to detect UPF1 and the loading control ACTB. **(B)** Densitometric analysis of this western blot revealed a significant 98% decrease in UPF1 protein levels following transfection. **P < 0.01 by Student's two tailed t-test.

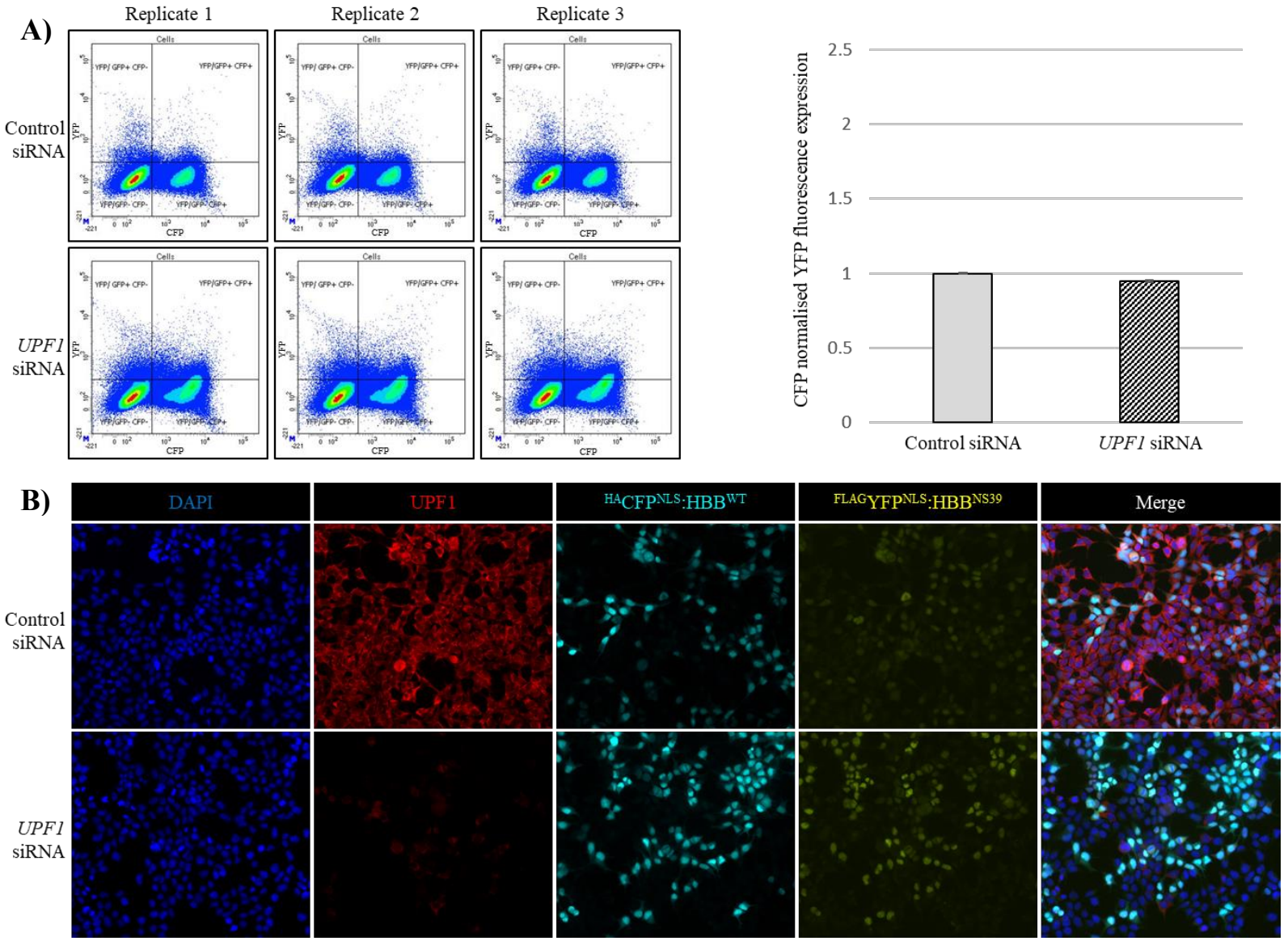

A)
Replicate 1
Replicate 2
Replicate 3
Control siRNA
UPF1 siRNA
Cells
YFP/ GFP+ CFP-
YFP/GFP+ CFP+
YFP/GFP- CFP-
YFP/GFP- CFP+
YFP
CFP
CFP normalised YFP fluorescence expression
2.5
2
1.5
1
0.5
0
Control siRNA
UPF1 siRNA
B)
DAPI
UPF1
HACFPNLS:HBBWT
FLAGYFPNLS:HBBNS39
Merge
Control siRNA
UPF1 siRNA

Figure 6.10: Protein expressed from genomic Control$^{V2.0}$ and NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HE293T cells is unresponsive to reduced levels of UPF1 protein.

Three separate replicates of NMD Reporter$^{V2.0}$ HEK293T cells were transfected with Control or *UPF1* siRNA. These cells express ^{HA}CFP^{NLS}:HBB^{NLS} from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$. **(A)** Fluorescence expression from these cells was quantified by FACS. The left panel shows YFP vs CFP density plots across all three replicates. The right panel compiles this information into a graph representing CFP normalised YFP fluorescence expression. These results depict no change in fluorescence expression from NMD Reporter$^{V2.0}$ HEK293T following a loss of UPF1 protein. **(B)** Representative fluorescence microscopy images of these cells. The cells express ^{HA}CFP^{NLS}:HBB^{NLS} from genomic Control$^{V2.0}$ (cyan) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (yellow). UPF1 was detected by immunofluorescence (red) and cell nuclei were counterstained with DAPI (blue).

Protein isolated from NMD Reporter$^{V2.0}$ HEK293T cells in the first replicate of these experiments was subject to further investigations via western blot analysis (Figure 6.11). Similar to results from fluorescence analysis, it was seen that levels of $^{HA}CFP^{NLS}$:HBB^{WT} expressed from genomic Control$^{V2.0}$ and $^{FLAG}YFP^{NLS}$:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ both increased following treatment with *UPF1* siRNA (Figure 6.11).

Moreover, a smaller than expected ~28 kDa protein was also detected by HA and GFP antibodies (Figure 6.11A). This information suggests that this protein product may be a truncated protein expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells which encodes at least, an N-terminal HA epitope tag (detected by the HA antibody) and CFP (detected by the GFP antibody). As such, this protein product is henceforth referred to as ^{HA}CFP.

Through protein detection with the GFP antibody, it was observed that truncated ^{HA}CFP appeared to show a much higher level of expression than $^{HA}CFP^{NLS}$:HBB^{WT} which is also expressed from genomic Control$^{V2.0}$ and $^{FLAG}YFP^{NLS}$:HBB^{NS39} which is expressed from genomic NMD$^{V2.0}$ (Figure 6.11A).

One possibility that could explain these discussed results and the observed cytoplasmic CFP signal from NMD Reporter$^{V2.0}$ HEK293T cells (Figures 6.3 & 6.10B), is that perhaps ^{HA}CFP is (1) not targeted for proteasomal degradation, thus explaining the high levels of protein expression and (2) not nuclear localised, explaining the cytoplasmic CFP signal.

Interestingly, ^{HA}CFP also appeared to increase in expression following a loss of UPF1 (Figure 6.11). However, with only one replicate included in this western blot analysis this cannot be commented on further.

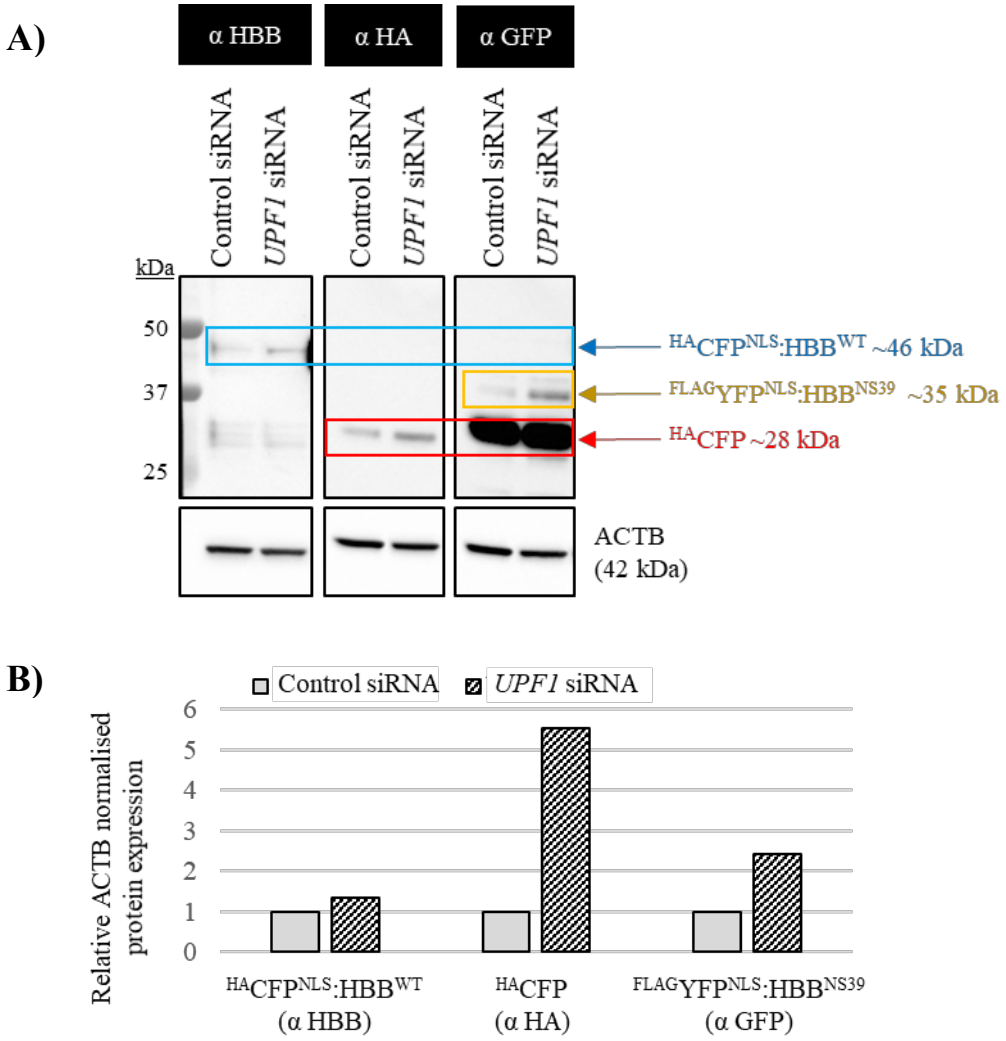


Figure 6.11: Western blot analysis of NMD Reporter$^{V2.0}$ HEK293T cells reveals a small and unexpected protein expressed from genomic Control$^{V2.0}$.

(A) NMD Reporter$^{V2.0}$ HEK293T cells were transfected with Control or *UPF1* siRNA. Total protein isolated from these cells was separated by SDS-PAGE and analysed by western blot using antibodies to detect HBB, HA, GFP and the loading control ACTB (stripped and re-probed). NMD Reporter$^{V2.0}$ HEK293T cells are expected to express ^{HA}CFP^{NLS}:HBB^{WT} from genomic Control$^{V2.0}$ (blue box) and ^{FLAG}YFP^{NLS}:HBB^{NS39} from genomic NMD$^{V2.0}$ (yellow box). Antibodies used to detect HA and GFP also detected a third protein (red box), since a HA tag is only encoded from Control$^{V2.0}$ the GFP antibody must be detecting CFP expressed from Control$^{V2.0}$, this protein species is referred to as '^{HA}CFP'. **(B)** Following a reduction in UPF1 expression, densitometric analysis of this western blot showed a 35% increase from ^{HA}CFP^{NLS}:HBB^{WT}, a 454% increase from ^{HA}CFP and a 142% increase in ^{FLAG}YFP^{NLS}:HBB^{NS39}.

6.3 Discussion

Previous investigations revealed that when Control$^{V1.0}$ and NMD$^{V1.0}$ or Control$^{V2.0}$ and NMD$^{V2.0}$ were expressed from stable NMD Reporter$^{V1.0}$ or NMD Reporter$^{V2.0}$ mESCs, respectively, NMD could not be faithfully quantified by either cell line. However, when expression plasmids for these cassettes were transiently expressed from HEK293T cells NMD$^{V1.0}$, Control$^{V2.0}$ and NMD$^{V2.0,}$ were functional. This led to the aims of this chapter, which was to (1) Establish stable NMD Reporter$^{V1.0}$ HEK293T and NMD Reporter$^{V2.0}$ HEK293T cells and (2) determine if these cells can faithfully quantify endogenous NMD levels of HEK293T cells.

To address the first aim, the engineered expression plasmids for Control$^{V1.0}$ and NMD$^{V1.0}$ (pUC57-CN$^{V1.0}$) or Control$^{V2.0}$ and NMD$^{V2.0}$ (pUC57-kan-SCN$^{V2.0}$) expression plasmids were chosen to be integrated into the genome of HEK293T cells to establish stable NMD Reporter$^{V1.0}$ HEK293T and NMD Reporter$^{V2.0}$ HEK293T cells respectively.

The plasmid pUC57-CN$^{V1.0}$ expresses Control$^{V1.0}$ and NMD$^{V1.0}$. Control$^{V1.0}$ serves as an internal control to NMD$^{V1.0}$ by expressing CFP^{NLS} regardless of NMD activity level in the cell. In comparison, NMD$^{V1.0}$ contains an NMD-targeting feature which should subject transcripts expressed from this cassette to NMD and thus influence expression of the encoded TetR:EGFP^{NLS} protein. This NMD-targeting feature is referred to as the NMD-targeted 3'UTR and is directly comparable to the NMD-refractory 3'UTR of Control$^{V1.0}$ (Chapter 4).

The plasmid pUC57-SCN$^{V2.0}$ contains three cassettes, namely, Selection$^{V2.0}$, Control $^{V2.0}$ and NMD$^{V2.0}$ in the case of NMD Reporter$^{V2.0}$ HEK293T cells only genomic Control$^{V2.0}$ and NMD$^{V2.0}$ are necessary to report on NMD activity. Control$^{V2.0}$ encodes ^{HA}CFP^{NLS}:HBB^{WT}. This cassette contains no NMD-targeting features and is predicted to express at a constant level

regardless of cellular NMD activity, in this way it serves as an internal control to the NMD-sensitive $NMD^{V2.0}$. Transcripts expressed from $NMD^{V2.0}$ encode $^{FLAG}YFP^{NLS}$:HBB^{NS39}, due to the presence of the NS39 PTC these transcript are subject to NMD which should influence protein expression from $NMD^{V2.0}$ (Chapter 5).

Neither pUC57-$CN^{V1.0}$ or pUC57-$SCN^{V2.0}$ were designed for loci-targeted stable integration into HEK293T cells. As such these plasmids were integrated into the genome of HEK293T via random genomic transgene integration. This method involves linearisation of the plasmid DNA and, as the name suggests, random integration into the genome.

Although this method was successful in establishing NMD $Reporter^{V1.0}$ HEK293T and NMD $Reporter^{V2.0}$ HEK293T cell lines. These cells will contain an unknown number of copies of linearised pUC57-$CN^{V1.0}$ or pUC57-$SCN^{V2.0}$ at various genomic locations. This is not an ideal situation as the number of copies of linearised expression plasmids integrated into any cell may influence the intensity of its fluorescent signal from genomic Control and NMD cassettes. Moreover, integration into unknown regions of the genome has the potential to disrupt genes that may be important for correct cellular function. Ideally, new 'FLP-in HEK293T' cell lines which like FLP-in mESCs can work with the designed selection cassettes to facilitate single copy, directed chromosomal integration into a 'safe-haven loci' would be constructed, however, this would have required significant time and effort that was beyond the scope of this PhD.

6.3.1 NMD $Reporter^{V1.0}$ HEK293T cells respond as expected to reduced levels of NMD activity, however, transcription termination and 3'UTR splicing of genomic $Control^{V1.0}$ and $NMD^{V1.0}$ is inefficient

Once NMD $Reporter^{V1.0}$ HEK293T cells were established, I used molecular methods to conduct a series of functional tests on the cell line. Through fluorescent microscopy of NMD

Reporter$^{V1.0}$ HEK293T cells, CFP^{NLS} expressed from genomic Control$^{V1.0}$ and TetR:EGFP^{NLS} expressed from genomic NMD$^{V1.0}$ was detectable and nuclear localised, this confirmed that many of the basic features of these cassettes including; the promoter, nuclear localisation signals and fluorescent proteins were functional when expressed from an integrated genomic loci.

Genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells both contain an intron within their 3'UTR which should be removed during splicing of the pre-mRNA. Investigations revealed that the NMD-refractive 3'UTR of genomic Control$^{V1.0}$ remained largely unspliced whilst the NMD-targeted 3'UTR of genomic NMD$^{V1.0}$ was predominantly spliced correctly. These results mirrored the 3'UTR splicing observed when Control$^{V1.0}$ and NMD$^{V1.0}$ were transiently expressed from HEK293T cells. As previously discussed (Chapter 4) an unspliced NMD-refractive 3'UTR would not contain any EJCs and therefore, like its spliced counterpart, is not predicted to trigger NMD. In this regard, unspliced transcripts expressed from Control$^{V1.0}$ should not interfere with NMD quantitation from NMD Reporter$^{V1.0}$ HEK293T cells, however, the unspliced transcript will contain a 314 nt stretch of intronic sequence of unknown relevance.

Also consistent with results from transient transfection experiments in HEK293T cells, was the presence of inefficient transcriptional termination at the single SV40 polyA site within the 3'UTR of genomic Control$^{V1.0}$ and NMD$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells. Interestingly, when Control$^{V1.0}$ and NMD$^{V1.0}$ were expressed from stable NMD Reporter$^{V1.0}$ mESCs, a single SV40 polyA sequence was able to promote efficient transcription termination. This information suggests a fundamental difference in the mechanism of transcription termination, at least for Control$^{V1.0}$ and NMD$^{V1.0}$ in human versus mouse cells. Furthermore, this highlights that the type and number of terminator sequences used is an important consideration when designing a transgene.

Although evidence of mis-splicing and inefficient transcription termination was identified, when *UPF1* siRNA was used to reduce NMD activity levels in NMD Reporter$^{V1.0}$ HEK293T cells, a significant increase in protein expression from genomic NMD$^{V1.0}$ compared to that from genomic Control$^{V1.0}$ was observed. Compared to transient expression from HEK293T cells and expression from stable NMD Reporter$^{V1.0}$ mESCs, this is the first time in which Control$^{V1.0}$ and NMD$^{V1.0}$ have both functioned as designed.

Through transient transfection-based experiments in HEK293T cells no difference could be detected between protein expression from Control$^{V1.0}$ and NMD$^{V1.0}$ following a loss of UPF1, however when expressed stably from NMD Reporter$^{V1.0}$ HEK293T cells the cassettes responded as expected (albeit, at a considerably low magnitude) (Chapter 4). Similar findings were published by Gerbracht et al. 2017, who showed that several different cell lines including HEK-293 Flp-In T-Rex (293 FT) cells were unable to efficiently degrade NMD reporter mRNAs during transient transfection-based experiments. However, when they integrated the same reporter constructs into the genome of 293T FT cells, levels of PTC containing reporter mRNA was efficiently reduced. This study suggested that NMD of target mRNA is generally less efficient when the target mRNA is expressed episomally during transient transfection-based experiments compared to when expressed from integrated genomic loci (Gerbracht, Boehm & Gehring 2017).

Like NMD Reporter$^{V1.0}$ HEK293T cells, NMD Reporter$^{V1.0}$ mESCs also harboured genomic Control$^{V1.0}$ and NMD$^{V1.0}$. In these cells, however the cassettes were unable to show a strong response to reduced cellular Upf1. These results suggest that alongside transient versus genomic expression, cell type is an important driver of NMD reporter function. This supports the idea that NMD is a complex process and the activity and target recognition of which can differ based on cell type (Linde, Liat et al. 2007; Viegas et al. 2007).

6.3.2 NMD Reporter$^{V2.0}$ HEK293T cells do not respond as expected to reduced levels of NMD activity, likely due to proteasomal degradation of encoded proteins

In this chapter, NMD Reporter$^{V2.0}$ HEK293T cells were also established and subject to functional testing. Fluorescent microscopy of these cells initially revealed that CFP^{NLS} (^{HA}CFP^{NLS}:HBB^{WT}) expressed from genomic Control$^{V1.0}$ and YFP^{NLS} (^{FLAG}YFP^{NLS}:HBB^{NS39}) expressed from genomic NMD$^{V1.0}$ was detectable, however it was observed that CFP expression was not restricted to the nucleus. This phenomenon had not been previously documented when Control$^{V1.0}$ was transiently expressed from HEK293T cells or from stable NMD Reporter$^{V1.0}$ mESCs. Further investigations through western blot analysis revealed that it was likely that ^{HA}CFP^{NLS}:HBB^{WT} expressed from genomic Control$^{V1.0}$ of NMD Reporter$^{V1.0}$ HEK293T cells was truncated or cleaved within these cells, resulting in a smaller than expected protein product (^{HA}CFP) which may have compromised nuclear localisation.

Alongside truncation issues, proteasomal degradation of ^{HA}CFP^{NLS}:HBB^{WT} expressed from genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells was detected. This same degradation was observed for ^{HA}CFP^{NLS}:HBB^{WT} expressed from genomic Control$^{V2.0}$ and ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ mESCs, and thus does not appear to be species dependent.

The proteasome is selective degradation pathway present in both the nucleus and cytoplasm which targets ubiquitinated proteins for degradation (Grigoreva et al. 2015; Jang 2018; Morozov & Karpov 2018). It is likely, that as previously discussed (Chapter.5), inclusion of a *HBB* mini-gene in the translated sequence of Control$^{V2.0}$ may be having a negative effect on protein stability due to forced expression of HBB outside its native cell type and sub-cellular location. Although not detected, based on previous results from NMD Reporter$^{V2.0}$ mESCs, it

was predicted that ^{FLAG}YFP^{NLS}:HBB^{NS39} expressed from genomic NMD$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells was also targeted for proteasomal degradation.

When considering the discussed issues with truncation and proteasomal degradation of proteins expressed from the genomic Control$^{V2.0}$ of NMD Reporter$^{V2.0}$ HEK293T cells, it was unsurprising that following UPF1 reduction in these cells, this NMD reporter system did not respond as expected. Control$^{V2.0}$ and NMD$^{V2.0}$ are largely based on the design of a previously published chemiluminescent NMD reporter system which was designed to function when expressed transiently in cells (Boelz et al, 2006). Interestingly, the only time protein expressed from Control$^{V2.0}$ and NMD$^{V2.0}$ was observed to respond as expected to a reduction in NMD activity levels was when these cassettes were transiently expressed in HEK293T cells (Chapter 5). This suggests that although the *HBB*-based NMD reporting system may be able to quantify NMD activity when expressed transiently, it may not be an ideal choice for a genome integrated NMD reporter system.

6.3.3 Chapter conclusions and future directions

Traditionally, the molecular mechanism of NMD has been characterised using pairs of reporter constructs that express the same mRNA with or without a PTC through transient transfection. However, when introduced into the cell via transient transfection, these transcripts are expressed from a high number of extrachromosomal plasmids which has been reported to yield less pronounced and less robust results than stable reporter cell lines (Gerbracht, Boehm & Gehring 2017).

In this chapter I have established and experimentally tested NMD Reporter$^{V1.0}$ and NMD Reporter$^{V2.0}$ HEK293T cells. The findings here identified NMD Reporter$^{V1.0}$ HEK293T cell as responsive to a reduction in NMD activity achieved by *UPF1* siRNA, whilst NMD Reporter$^{V2.0}$ HEK293T cells were found to be non-functional due to post-translational processing issues.

Moving forward, NMD Reporter$^{V1.0}$ HEK293T cells could be developed for use in high-throughput experiments to identify drugs or small compounds that can alter NMD efficiency to drive therapeutics development or to aid research and investigate the effects of a range of external stimuli e.g. cellular stresses such as hypoxia, on NMD activity. Unfortunately, since HEK293T cells are not pluripotent, this system cannot be used to report on endogenous NMD activity across different cells, tissues or developmental paths.

Ultimately there have been advantages and disadvantages of Version 1.0 and Version 2.0 of the NMD reporter system designed in this thesis and it is likely that an improved design could follow from the findings in this chapter and the previous chapters of this thesis. For example, some of the successful design features from the Version 2.0 NMD reporter system, such as; three polyA SV40 sequences to combat inefficient transcription termination, N-terminal epitope tags to allow more robust western blot analysis and the use of YFP instead of EGFP to allow more informative fluorescent microscopy could be combined with the responsive NMD-targeting feature of Version 1.0. Moreover, the development of FLP-in HEK293T cells (or other cell types of interest) could be beneficial. In this way an improved NMD reporter system could be used to develop both easy to use targeted stable cell lines for high throughput experiments and stable germline competent mESCs cell lines for investigations into NMD activity across differentiation and development.

Chapter Seven: Final Discussion

7.1 Significance of the results

My studies in patient derived LCLs (Chapter 3) characterised three novel variants within genes encoding NMD factors *UPF3B* or *UPF2*. These variants were identified in individuals who show varying presentations of neurodevelopmental disorders. The first of these variants was a synonymous SNV found in a canonical splice region of *UPF3B*. This variant was originally classified as a VUS and as such overlooked regarding pathogenicity. An in-depth assessment of pathogenicity was performed for this variant and the findings were able to conclusively resolve this variant as pathogenic. This outcome resulted in a confident patient diagnosis whilst also highlighting the need to value any variants identified in *UPF3B* and other genes which when disrupted are known to result in neurodevelopmental disorders, particularly those in canonical splice site regions. The remaining two variants were identified in *UPF2*, the first was a novel frameshift mutation in *UPF2,* which is one of only two single nucleotide variants identified to exclusively disrupt *UPF2*. The second was a CNV which resulted in the heterozygous deletion of *UPF2* along with 21 other genes. These studies supported the involvement of UPF2 in a spectrum of neurodevelopmental disorders which have been observed for these cases, and previous cases where UPF2 has been disrupted as a consequence of large CNV deletions, however conclusively attributing the clinical presentation of these individuals to impaired NMD due to a loss of UPF2 was not achieved due to a limited sample size. These studies show that patient derived LCLs serve as an effective starting point to explore the involvement of NMD in the pathology of the still growing list of genetic disorders caused by variants in NMD factors or by NMD targeted PTC mutations.

The studies in this thesis also, for the first time, have attempted to engineer a single NMD reporter transgene that can be integrated into the genome of pluripotent mESCs and provide a visual and quantifiable measurement of NMD with single cell resolution (Chapters 4 & 5). The development of such a tool would allow single cell visualisation and quantification of NMD

activity across many differentiation pathways and in many different cell types. Being an integrated transgenic system, this NMD reporter system would not be confounded by differences in transfection efficiencies between experiments and therefore can also be easily used in high-throughput assays. Although the development of an NMD reporter mESC line was ultimately unsuccessful, an abundance of knowledge was gained through the process of design and functional validation of two separate versions of an NMD reporter transgene. This will provide an important basis for the design and testing approach of an improved NMD reporter transgene and has also highlighted the complexity of the NMD pathway across different cell types.

The investigations in this thesis also led to the development of a functional dual fluorescent NMD reporter HEK293T cell line with single cell resolution (NMD Reporter$^{V1.0}$ HEK293T cells) (Chapter 6). Currently, there a number of stable fluorescent and luminescent NMD reporter systems (Alexandrov, Shu & Steitz 2017; Nickless et al. 2014; Paillusson et al. 2005; Welch et al. 2007) however, only one of these is able to provide single cell resolution and is limited to use in the HeLa cancer cell line (Alexandrov, Shu & Steitz 2017). NMD Reporter$^{V1.0}$ HEK293T cells can now be used in high-throughput experiments to identify drugs and small molecules that alter NMD activity, investigate the effects of a range of external stimuli on NMD activity and thus broaden our understanding on the mechanics of NMD.

.

7.2 Limitations of the study

There are some important limitations in these studies that should be further noted and discussed. The first was having LCLs from only one individual harbouring a variant exclusively disrupting *UPF2*. The findings in this thesis showed that this novel *UPF2* frameshift variant presented as a loss of function mutation with reduced transcript and protein levels comparable to large CNVs which delete *UPF2* amongst many other genes. It has been shown that in such CNV cases, the

heterozygous loss of UPF2 disrupts classical NMD which is thought to result in a spectrum of neurodevelopmental phenotypes. Based on this, it was hypothesised that the *UPF2* frameshift variant would also disrupt classical NMD.

Contradictory to the hypothesis, molecular investigations revealed that in comparison to CNVs, the *UPF2* frameshift variant did not drastically reduce the efficiency of classical NMD, at least as measured by the expression of selected mRNAs tested in LCLs. Without more patient derived LCLs harbouring variants exclusively disrupting *UPF2* it was not possible to confidently conclude or dispute that a disruption in classical NMD was caused by a heterozygous loss of UPF2 or that such a loss of UPF2 is the underlying cause of the range of neurodevelopmental phenotypes observed in individuals harbouring genetic variants resulting in heterozygous loss of UPF2. This highlighted the need to continue to identify and collect genetic material from individuals harbouring variants of interest as large sample sizes are essential to give confidence in the results of molecular investigations. Moreover, large sample sizes can drive powerful whole transcriptomic studies which, in this case can assess the effects of genetic variants on NMD activity at a global level.

Chapters 4 and 5 detailed the process of designing, engineering and experimentally testing two versions of a transgene encoding an NMD reporter system which aimed to report on endogenous NMD activity at the single cell level in mESCs. Through these studies, several complications caused by transgene design were uncovered and discussed at length (Chapters 4–6). In Chapters 4 and 6, a splicing dependent NMD reporter system (Version 1.0) was developed based on a published design (Pereverzev et al, 2015). An important discovery was that the ability of the designed system to report on NMD appeared to be dependent on whether it was expressed transiently in cells or from a stable cell line, and moreover, if the cell type it was expressed in was derived from human or mouse. As such, the expression cassettes of Version 1.0 were able to faithfully report on NMD from stable NMD Reporter$^{V1.0}$ HEK293T cells, however, were

unable to do this when expressed from stable NMD Reporter$^{V1.0}$ mESCs or when used in transient transfection-based experiments. Observations such as this exemplified the complexity of NMD, which prior to this research was not well appreciated.

In Chapter 5 and 6, a second NMD reporter transgene was designed and tested (Version 2.0, Transgene$^{V2.0}$), this time based on a published transient and chemiluminescent-based NMD reporter system (Boelz et al, 2006). When Transgene$^{V2.0}$ was transiently expressed from HEK293T cells, the NMD reporter system responded as expected to a reduction in NMD activity. However, following integration of this transgene into the genome of human HEK293T cells or mESCs, the encoded proteins were targeted for proteasomal degradation. Since Transgene$^{V2.0}$ was designed to measure NMD activity via fluorescent output, i.e. protein expression rather than target mRNA expression, this unpredicted degradation was ultimately the downfall of Transgene$^{V2.0}$ preventing its function in a stable genomic context. This observation highlighted the impact other cellular processing and degradation pathways can have on the activity of a fluorescent NMD reporter system which may have been overlooked prior to this research. Moreover, the findings from Chapters 4–6 showed that perhaps this research relied too heavily on the published results of others as a starting point, and the design of an improved NMD reporter system would now be better built anew from the knowledge acquired from in-house experiments conducted in this thesis.

7.3 Future directions

This thesis is the first to use molecular approaches to investigate a genetic variant which exclusively disrupts *UPF2* and results in a neurodevelopmental disorder. It was demonstrated that this variant resulted in loss of UPF2 protein comparable to previous large heterozygous CNV deletion cases which disrupted *UPF2*. However, further work is required to fully understand if this variant disrupts classical NMD efficiency and if so, the mechanism by which

disrupted NMD can result in the neurodevelopmental phenotypes associated with neurodevelopmental disorders. Furthermore, identifying additional individuals who harbour variants that exclusively disrupt *UPF2* (or other members of the NMD pathway) and collecting genetic material will drive powerful whole transcriptomic studies which can potentially elucidate the effect of such variants on NMD activity at a global level, and also give insight into specific cellular pathways that may be disrupted as a result. Lastly, development of additional models such as induced pluripotent stem cells (iPSCs) which harbour genetic variants in NMD factor encoding genes would be invaluable to investigate the neurodevelopmental phenotypes observed in neurodevelopmental disorders caused by genetic variants in NMD factors as these cells can be differentiated into various neuronal cell types.

After completing this study, the idea of an NMD reporter transgene that can be integrated into the genome of mESCs or mouse embryos to report on endogenous NMD activity at a single cell level remains intriguing and is still believed to be achievable, however, the complexity of such a task has become apparent. Through this study a considerable amount of knowledge has been gained in terms of understanding how to effectively design and test a reporter transgene. As such, the studies in this thesis have established many basic features needed to engineer a successful NMD reporter transgene and have optimised many protocols needed to test said transgene, these include; feeder-independent FLP-in mESCs which can be successfully edited via an optimised RMCE system, A CAG/3x SV40 polyA expression system which facilitates efficient transcription termination, a streamlined experimental pipeline to assemble and test reporter transgenes and methods to visualise and quantify nuclear localised expression of simultaneously expressed fluorescent proteins. The design of an improved fluorescent NMD reporter transgene will build on this knowledge.

Finally, as NMD Reporter$^{V1.0}$ HEK293T cells were found to respond as expected to a reduction in NMD activity, this cell line can be further developed for use in high-throughput screens for

drugs or small compounds which can modify NMD activity. The identification of such compounds can not only drive research into therapeutics but also to identify new ways to manipulate NMD to benefit research methods.

7.4 Final concluding remarks

My results support the knowledge that variants in NMD factors can result in heterogenous presentations of neurodevelopmental disorders. In addition, the findings in this thesis suggest that the role NMD plays in these disorders may not be as straightforward as predicted and could involve different branches of NMD and be influenced by the inter-individual variability of NMD activity. One way to further understand NMD, especially during neurodevelopment is to develop NMD reporter systems which can quantify NMD activity in pluripotent stem cells and monitor its activity across a neural differentiation.

In this study two versions of such an NMD reporter system were designed, engineered, and tested. Although ultimately unsuccessful, the knowledge gained through the investigations within this thesis have laid a foundation that will lead to developing a new and improved fluorescent NMD reporter system with single cell resolution. Such a system will not only aid in understanding the role of NMD during neurodevelopment but can also be used to explore the countless aspects of biology and disease that NMD is involved in, and support the development of new therapeutic approaches.

Appendix: Publications

Publications from this work

Domingo D, Nawaz U, Corbett M, Espinoza JL, Tatton-Brown K, Coman D, Wilkinson M, Gecz J, Jolly L (**in press**), 'A synonymous mutation in UPF3B causes loss-of-function due to aberrant splicing in an individual with motor delay and absence of speech.', *Human Molecular Genetics*.

Johnson, JL, Stoica, L, Liu, Y, Zhu, PJ, Bhattacharya, A, Buffington, SA, Huq, R, Eissa, NT, Larsson, O, Porse, BT, **Domingo, D**, Nawaz, U, Carroll, R, Jolly, L, Scerri, TS, Kim, H-G, Brignell, A, Coleman, MJ, Braden, R, Kini, U, Jackson, V, Baxter, A, Bahlo, M, Scheffer, IE, Amor, DJ, Hildebrand, MS, Bonnen, PE, Beeton, C, Gecz, J, Morgan, AT & Costa-Mattioli, M **2019**, 'Inhibition of Upf2-Dependent Nonsense-Mediated Decay Leads to Behavioral and Neurophysiological Abnormalities by Activating the Immune Response', *Neuron*.

Publications arising from collaborative work

Johnson, BV, Kumar, R, Oishi, S, Alexander, S, Kasherman, M, Vega, MS, Ivancevic, A, Gardner, A, **Domingo, D**, Corbett, M, Parnell, E, Yoon, S, Oh, T, Lines, M, Lefroy, H, Kini, U, Van Allen, M, Grønborg, S, Mercier, S, Küry, S, Bézieau, S, Pasquier, L, Raynaud, M, Afenjar, A, Billette de Villemeur, T, Keren, B, Désir, J, Van Maldergem, L, Marangoni, M, Dikow, N, Koolen, DA, Vanhasselt, PM, Weiss, M, Zwijnenburg, P, Sa, J, Reis, CF, López-Otín, C, Santiago-Fernández, O, Fernández-Jaén, A, Rauch, A, Steindl, K, Joset, P, Goldstein, A, Madan-Khetarpal, S, Infante, E, Zackai, E, Mcdougall, C, Narayanan, V, Ramsey, K, Mercimek-Andrews, S, Pena, L, Shashi, V, Schoch, K, Sullivan, JA, Pinto E Vairo, F, Pichurin, PN, Ewing, SA, Barnett, SS, Klee, EW, Perry, MS, Koenig, MK, Keegan, CE, Schuette, JL, Asher, S, Perilla-Young, Y, Smith, LD, Rosenfeld, JA, Bhoj, E, Kaplan, P, Li, D, Oegema, R,

van Binsbergen, E, van Der Zwaag, B, Smeland, MF, Cutcutache, I, Page, M, Armstrong, AE, Lin, MA, Steeves, ND, Hollander, MJV, Hoffer, MRF, Reijnders, S, Demirdas, DC, Koboldt, D, Bartholomew, TM, Mosher, SE, Hickey, C, Shieh, PA, Sanchez-Lara, JM, Graham, K, Tezcan, GB, Schaefer, NR, Danylchuk, A, Asamoah, KE, Jackson, N, Yachelevich, M, Au, LA, Pérez-Jurado, T, et al. **2019**, 'Partial Loss of USP9X Function Leads to a Male Neurodevelopmental and Behavioral Disorder Converging on Transforming Growth Factor β Signaling', *Biological Psychiatry*.

Cheng, H, Dharmadhikari, AV, Varland, S, Ma, N, **Domingo, D**, Kleyner, R, Rope, AF, Yoon, M, Stray-Pedersen, A, Posey, JE, Crews, SR, Eldomery, MK, Akdemir, ZC, Lewis, AM, Sutton, VR, Rosenfeld, JA, Conboy, E, Agre, K, Xia, F, Walkiewicz, M, Longoni, M, High, FA, van Slegtenhorst, MA, Mancini, GM., Finnila, CR, van Haeringen, A, Den Hollander, N, Ruivenkamp, C, Naidu, S, Mahida, S, Palmer, EE, Murray, L, Lim, D, Jayakar, P, Parker, MJ, Giusto, S, Stracuzzi, E, Romano, C, Beighley, JS, Bernier, RA, Küry, S, Nizon, M, Corbett, MA, Shaw, M, Gardner, A, Barnett, C, Armstrong, R, Kassahn, KS, Van Dijck, A, Vandeweyer, G, Kleefstra, T, Schieving, J, Jongmans, MJ, de Vries, BB., Pfundt, R, Kerr, B, Rojas, SK, Boycott, KM, Person, R, Willaert, R, Eichler, EE, Kooy, RF, Yang, Y, Wu, JC, Lupski, JR, Arnesen, T, Cooper, GM, Chung, WK, Gecz, J, Stessman, HA., Meng, L & Lyon, GJ **2018**, 'Truncating Variants in NAA15 Are Associated with Variable Levels of Intellectual Disability, Autism Spectrum Disorder, and Congenital Anomalies', *The American Journal of Human Genetics*, vol. 102, no. 5, pp. 985–994.

Bridges, CR, Tan, M-C, Premarathne, S, Nanayakkara, D, Bellette, B, Zencak, D, **Domingo, D**, Gecz, J, Murtaza, M, Jolly, LA & Wood, SA **2017**, 'USP9X deubiquitylating enzyme maintains RAPTOR protein levels, mTORC1 signalling and proliferation in neural progenitors', *Scientific Reports*, vol. 7, no. 1, pp. 1–15.

Reijnders, M., Zachariadis, V, Latour, B, Jolly, L, Mancini, G, Pfundt, R, Wu, K, Van ravenswaaij-Arts, C., Veenstra-Knol, H, Anderlid, B-M, Wood, S, Cheung, S, Barnicoat, A, Probst, F, Magoulas, P, Brooks, A, Malmgren, H, Harila-Saari, A, Marcelis, C, Vreeburg, M, Hobson, E, Sutton, V rei., Stark, Z, Vogt, J, Cooper, N, Lim, J, Price, S, Lai, A, **Domingo, D**, Reversade, B, Gecz, J, Gilissen, C, Brunner, H, Kini, U, Roepman, R, Nordgren, A & Kleefstra, T **2016**, 'De Novo Loss-of-Function Mutations in USP9X Cause a Female-Specific Recognizable Syndrome with Developmental Delay and Congenital Malformations', *The American Journal of Human Genetics*, vol. 98, no. 2, pp. 373–381.

Jolly LS, Nguyen LS, **Domingo D**, Sun Y, Barry S, Hancarova M, Plevova P, Vlckova M, Havlovicova M, Kalscheuer VM, Graziano C, Pippucci T, Bonora E, Sedlacek Z & Gecz J **2015**, 'HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain', *Human Molecular Genetics*, vol. 24, no. 12, pp. 3335–3347.

Abstracts arising from this work

Domingo.D, Gecz.J, Berstern.D, Wilkinson.M and Jolly, L. 'Visualising the Nonsense (mediated decay)' Proceedings for **Lorne Genome 2018**, Lorne, Australia. *Poster presentation.*

Domingo.D, Gecz.J, Berstern.D, Wilkinson.M and Jolly, L. 'Visualising the Nonsense (mediated decay)' Proceedings for **Combio 2017**, Adelaide, Australia. *Oral presentation.*

Domingo.D, Gecz.J, Berstern.D, Wilkinson.M and Jolly, L. 'Visualising the Nonsense (mediated decay)' Proceedings for **Australasian Neuroscience Society 2016,** Hobart, Australia. *Poster presentation.*

References

Adachi, K, Lakka, V, Zhao, Y & Surrey, S 2004, 'Ubiquitylation of nascent globin chains in a cell-free system', *J Biol Chem*, vol. 279, no. 40, Oct 1, pp. 41767-41774.

Addington, AM, Gauthier, J, Piton, A, Hamdan, FF, Raymond, A, Gogtay, N, Miller, R, Tossell, J, Bakalar, J, Germain, G, Gochman, P, Long, R, Rapoport, JL & Rouleau, GA 2010, 'A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders', *Molecular Psychiatry*, vol. 17, no. 4, pp. 468-468.

Addington, AM, Gauthier, J, Piton, A, Hamdan, FF, Raymond, A, Gogtay, N, Miller, R, Tossell, J, Bakalar, J, Inoff-Germain, G, Gochman, P, Long, R, Rapoport, JL & Rouleau, GA 2011, 'A novel frameshift mutation in UPF3B identified in brothers affected with childhood onset schizophrenia and autism spectrum disorders', *Mol Psychiatry*, vol. 16, no. 3, Mar, pp. 238-239.

Alexandrov, A, Shu, MD & Steitz, JA 2017, 'Fluorescence Amplification Method for Forward Genetic Discovery of Factors in Human mRNA Degradation', *Mol Cell*, vol. 65, no. 1, Jan 5, pp. 191-201.

Alkalaeva, EZ, Pisarev, AV, Frolova, LY, Kisselev, LL & Pestova, TV 2006, 'In vitro reconstitution of eukaryotic translation reveals cooperativity between release factors eRF1 and eRF3', *Cell*, vol. 125, no. 6, Jun 16, pp. 1125-1136.

Anczukow, O, Ware, MD, Buisson, M, Zetoune, AB, Stoppa-Lyonnet, D, Sinilnikova, OM & Mazoyer, S 2008, 'Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and p53 proteins?', *Hum Mutat*, vol. 29, no. 1, Jan, pp. 65-73.

Anders, KR, Grimson, A & Anderson, P 2003, 'SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A', *Embo j*, vol. 22, no. 3, Feb 3, pp. 641-650.

Andersen, CB, Ballut, L, Johansen, JS, Chamieh, H, Nielsen, KH, Oliveira, CL, Pedersen, JS, Seraphin, B, Le Hir, H & Andersen, GR 2006, 'Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA', *Science*, vol. 313, no. 5795, Sep 29, pp. 1968-1972.

Anna, A & Monika, G 2018, 'Splicing mutations in human genetic disorders: examples, detection, and confirmation', *J Appl Genet*, vol. 59, no. 3, pp. 253-268.

Applequist, SE, Selg, M, Raman, C & Jack, HM 1997, 'Cloning and characterization of HUPF1, a human homolog of the Saccharomyces cerevisiae nonsense mRNA-reducing UPF1 protein', *Nucleic Acids Res*, vol. 25, no. 4, Feb 15, pp. 814-821.

Arias-Palomo, E, Yamashita, A, Fernandez, IS, Nunez-Ramirez, R, Bamba, Y, Izumi, N, Ohno, S & Llorca, O 2011, 'The nonsense-mediated mRNA decay SMG-1 kinase is regulated by large-scale conformational changes controlled by SMG-8', *Genes Dev*, vol. 25, no. 2, Jan 15, pp. 153-164.

Aslam, AA, Higgins, C, Sinha, IP & Southern, KW 2017, 'Ataluren and similar compounds (specific therapies for premature termination codon class I mutations) for cystic fibrosis', *Cochrane Database Syst Rev*, vol. 1, Jan 19, p. Cd012040.

Atkinson, J & Martin, R 1994, 'Mutations to nonsense codons in human genetic disease: implications for gene therapy by nonsense suppressor tRNAs', *Nucleic Acids Res*, vol. 22, no. 8, Apr 25, pp. 1327-1334.

Azzalin, CM & Lingner, J 2006, 'The human RNA surveillance factor UPF1 is required for S phase progression and genome stability', *Curr Biol*, vol. 16, no. 4, Feb 21, pp. 433-439.

Baird, TD, Cheng, KC, Chen, YC, Buehler, E, Martin, SE, Inglese, J & Hogg, JR 2018, 'ICE1 promotes the link between splicing and nonsense-mediated mRNA decay', *Elife*, vol. 7, Mar 12.

Bartoszewski, RA, Jablonsky, M, Bartoszewska, S, Stevenson, L, Dai, Q, Kappes, J, Collawn, JF & Bebok, Z 2010, 'A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein', *J Biol Chem*, vol. 285, no. 37, Sep 10, pp. 28741-28748.

Baserga, SJ & Benz, EJ, Jr. 1988, 'Nonsense mutations in the human beta-globin gene affect mRNA metabolism', *Proc Natl Acad Sci U S A*, vol. 85, no. 7, Apr, pp. 2056-2060.

Bateman, JF, Freddi, S, Nattrass, G & Savarirayan, R 2003, 'Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage', *Hum Mol Genet*, vol. 12, no. 3, Feb 1, pp. 217-225.

Bechara, EG, Sebestyen, E, Bernardis, I, Eyras, E & Valcarcel, J 2013, 'RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation', *Mol Cell*, vol. 52, no. 5, Dec 12, pp. 720-733.

Bedwell, DM, Kaenjak, A, Benos, DJ, Bebok, Z, Bubien, JK, Hong, J, Tousson, A, Clancy, JP & Sorscher, EJ 1997, 'Suppression of a CFTR premature stop mutation in a bronchial epithelial cell line', *Nat Med*, vol. 3, no. 11, Nov, pp. 1280-1284.

Behringer, RR, Ryan, TM, Palmiter, RD, Brinster, RL & Townes, TM 1990, 'Human gamma-to beta-globin gene switching in transgenic mice', *Genes Dev*, vol. 4, no. 3, Mar, pp. 380-389.

Belgrader, P, Cheng, J & Maquat, LE 1993, 'Evidence to implicate translation by ribosomes in the mechanism by which nonsense codons reduce the nuclear level of human triosephosphate isomerase mRNA', *Proc Natl Acad Sci U S A*, vol. 90, no. 2, Jan 15, pp. 482-486.

Belgrader, P & Maquat, LE 1994, 'Nonsense but not missense mutations can decrease the abundance of nuclear mRNA for the mouse major urinary protein, while both types of mutations can facilitate exon skipping', *Mol Cell Biol*, vol. 14, no. 9, Sep, pp. 6326-6336.

Bersten, DC, Sullivan, AE, Li, D, Bhakti, V, Bent, SJ & Whitelaw, ML 2015, 'Inducible and reversible lentiviral and Recombination Mediated Cassette Exchange (RMCE) systems for controlling gene expression', *PLoS One*, vol. 10, no. 3, p. e0116373.

Bhuvanagiri, M, Lewis, J, Putzker, K, Becker, JP, Leicht, S, Krijgsveld, J, Batra, R, Turnwald, B, Jovanovic, B & Hauer, C 2014, '5-azacytidine inhibits nonsense-mediated decay in a MYC-dependent fashion', *EMBO Mol Med*, vol. 6, no. 12, pp. 1593-1609.

Bhuvanagiri, M, Schlitter, AM, Hentze, MW & Kulozik, AE 2010, 'NMD: RNA biology meets human genetic medicine', *Biochem J*, vol. 430, no. 3, Sep 15, pp. 365-377.

Bitta, M, Kariuki, SM, Abubakar, A & Newton, C 2017, 'Burden of neurodevelopmental disorders in low and middle-income countries: A systematic review and meta-analysis', *Wellcome Open Res*, vol. 2, p. 121.

Boelz, S, Neu-Yilik, G, Gehring, NH, Hentze, MW & Kulozik, AE 2006, 'A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay', *Biochemical and biophysical research communications*, vol. 349, no. 1, pp. 186-191.

Bono, F, Ebert, J, Lorentzen, E & Conti, E 2006, 'The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA', *Cell*, vol. 126, no. 4, Aug 25, pp. 713-725.

Bravo, R, Parra, V, Gatica, D, Rodriguez, AE, Torrealba, N, Paredes, F, Wang, ZV, Zorzano, A, Hill, JA, Jaimovich, E, Quest, AF & Lavandero, S 2013, 'Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration', *Int Rev Cell Mol Biol*, vol. 301, pp. 215-290.

Brest, P, Lapaquette, P, Souidi, M, Lebrigand, K, Cesaro, A, Vouret-Craviari, V, Mari, B, Barbry, P, Mosnier, JF, Hebuterne, X, Harel-Bellan, A, Mograbi, B, Darfeuille-Michaud, A & Hofman, P 2011, 'A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease', *Nat Genet*, vol. 43, no. 3, Mar, pp. 242-245.

Brinkman, EK, Chen, T, Amendola, M & van Steensel, B 2014, 'Easy quantitative assessment of genome editing by sequence trace decomposition', *Nucleic Acids Res*, vol. 42, no. 22, pp. e168-e168.

Brogna, S & Wen, J 2009, 'Nonsense-mediated mRNA decay (NMD) mechanisms', *Nat Struct Mol Biol*, vol. 16, no. 2, Feb, pp. 107-113.

Brunetti-Pierri, N, Berg, JS, Scaglia, F, Belmont, J, Bacino, CA, Sahoo, T, Lalani, SR, Graham, B, Lee, B, Shinawi, M, Shen, J, Kang, SH, Pursley, A, Lotze, T, Kennedy, G, Lansky-Shafer, S, Weaver, C, Roeder, ER, Grebe, TA, Arnold, GL, Hutchison, T, Reimschisel, T, Amato, S, Geragthy, MT, Innis, JW, Obersztyn, E, Nowakowska, B, Rosengren, SS, Bader, PI, Grange, DK, Naqvi, S, Garnica, AD, Bernes, SM, Fong, CT, Summers, A, Walters, WD, Lupski, JR, Stankiewicz, P, Cheung, SW & Patel, A 2008, 'Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities', *Nat Genet*, vol. 40, no. 12, Dec, pp. 1466-1471.

Bruno, IG, Karam, R, Huang, L, Bhardwaj, A, Lou, CH, Shum, EY, Song, HW, Corbett, MA, Gifford, WD, Gecz, J, Pfaff, SL & Wilkinson, MF 2011, 'Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay', *Molecular Cell*, vol. 42, no. 4, May 20, pp. 500-510.

Buchwald, G, Ebert, J, Basquin, C, Sauliere, J, Jayachandran, U, Bono, F, Le Hir, H & Conti, E 2010, 'Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex', *Proc Natl Acad Sci U S A*, vol. 107, no. 22, Jun 1, pp. 10050-10055.

Buhler, M, Paillusson, A & Muhlemann, O 2004, 'Efficient downregulation of immunoglobulin mu mRNA with premature translation-termination codons requires the 5'-half of the VDJ exon', *Nucleic Acids Res*, vol. 32, no. 11, pp. 3304-3315.

Buhler, M, Steiner, S, Mohn, F, Paillusson, A & Muhlemann, O 2006, 'EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length', *Nat Struct Mol Biol*, vol. 13, no. 5, May, pp. 462-464.

Buvoli, M, Buvoli, A & Leinwand, LA 2000, 'Suppression of nonsense mutations in cell culture and mice by multimerized suppressor tRNA genes', *Mol Cell Biol*, vol. 20, no. 9, May, pp. 3116-3124.

Cali, BM, Kuchma, SL, Latham, J & Anderson, P 1999, 'smg-7 is required for mRNA surveillance in Caenorhabditis elegans', *Genetics*, vol. 151, no. 2, Feb, pp. 605-616.

Cartegni, L, Chew, SL & Krainer, AR 2002, 'Listening to silence and understanding nonsense: exonic mutations that affect splicing', *Nature Reviews Genetics*, vol. 3, 04/01/online, p. 285.

Carter, MS, Doskow, J, Morris, P, Li, S, Nhim, RP, Sandstedt, S & Wilkinson, MF 1995, 'A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro', *J Biol Chem*, vol. 270, no. 48, Dec 1, pp. 28995-29003.

Carter, MS, Li, S & Wilkinson, MF 1996, 'A splicing-dependent regulatory mechanism that detects translation signals', *Embo j*, vol. 15, no. 21, Nov 1, pp. 5965-5975.

Chamary, JV, Parmley, JL & Hurst, LD 2006, 'Hearing silence: non-neutral evolution at synonymous sites in mammals', *Nature Reviews Genetics*, vol. 7, 02/01/online, p. 98.

Chamieh, H, Ballut, L, Bonneau, F & Le Hir, H 2008, 'NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity', *Nat Struct Mol Biol*, vol. 15, no. 1, Jan, pp. 85-93.

Chan, D, Weng, YM, Graham, HK, Sillence, DO & Bateman, JF 1998, 'A nonsense mutation in the carboxyl-terminal domain of type X collagen causes haploinsufficiency in schmid metaphyseal chondrodysplasia', *J Clin Invest*, vol. 101, no. 7, Apr 1, pp. 1490-1499.

Chan, WK, Bhalla, AD, Le Hir, H, Nguyen, LS, Huang, L, Gecz, J & Wilkinson, MF 2009, 'A UPF3-mediated regulatory switch that maintains RNA surveillance', *Nat Struct Mol Biol*, vol. 16, no. 7, Jul, pp. 747-753.

Chan, WK, Huang, L, Gudikote, JP, Chang, YF, Imam, JS, MacLean, JA, 2nd & Wilkinson, MF 2007, 'An alternative branch of the nonsense-mediated decay pathway', *Embo j*, vol. 26, no. 7, Apr 4, pp. 1820-1830.

Chang, JC & Kan, YW 1979, 'beta 0 thalassemia, a nonsense mutation in man', *Proceedings of the National Academy of Sciences*, vol. 76, no. 6, p. 2886.

Chen, Z, Gore, BB, Long, H, Ma, L & Tessier-Lavigne, M 2008, 'Alternative splicing of the Robo3 axon guidance receptor governs the midline switch from attraction to repulsion', *Neuron*, vol. 58, no. 3, May 8, pp. 325-332.

Cheng, J, Fogel-Petrovic, M & Maquat, LE 1990, 'Translation to near the distal end of the penultimate exon is required for normal levels of spliced triosephosphate isomerase mRNA', *Mol Cell Biol*, vol. 10, no. 10, Oct, pp. 5215-5225.

Cheruiyot, A, Li, S, Nickless, A, Roth, R, Fitzpatrick, JAJ & You, Z 2018, 'Compound C inhibits nonsense-mediated RNA decay independently of AMPK', *PLoS One*, vol. 13, no. 10, p. e0204978.

Clancy, JP, Bebok, Z, Ruiz, F, King, C, Jones, J, Walker, L, Greer, H, Hong, J, Wing, L, Macaluso, M, Lyrene, R, Sorscher, EJ & Bedwell, DM 2001, 'Evidence that systemic gentamicin suppresses premature stop mutations in patients with cystic fibrosis', *Am J Respir Crit Care Med*, vol. 163, no. 7, Jun, pp. 1683-1692.

Coe, BP, Witherspoon, K, Rosenfeld, JA, van Bon, BW, Vulto-van Silfhout, AT, Bosco, P, Friend, KL, Baker, C, Buono, S, Vissers, LE, Schuurs-Hoeijmakers, JH, Hoischen, A, Pfundt, R, Krumm, N, Carvill, GL, Li, D, Amaral, D, Brown, N, Lockhart, PJ, Scheffer, IE, Alberti, A, Shaw, M, Pettinato, R, Tervo, R, de Leeuw, N, Reijnders, MR, Torchia, BS, Peeters, H, O'Roak, BJ, Fichera, M, Hehir-Kwa, JY, Shendure, J, Mefford, HC, Haan, E, Gecz, J, de Vries, BB, Romano, C & Eichler, EE 2014, 'Refining analyses of copy number variation identifies specific genes associated with developmental delay', *Nat Genet*, vol. 46, no. 10, Oct, pp. 1063-1071.

Colak, D, Ji, SJ, Porse, BT & Jaffrey, SR 2013, 'Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay', *Cell*, vol. 153, no. 6, Jun 6, pp. 1252-1265.

Costa-Mattioli, M & Walter, P 2020, 'The integrated stress response: From mechanism to disease', *Science*, vol. 368, no. 6489, Apr 24.

Dabrowski, M, Bukowy-Bieryllo, Z & Zietkiewicz, E 2018, 'Advances in therapeutic use of a drug-stimulated translational readthrough of premature termination codons', *Molecular Medicine*, vol. 24, no. 1, 2018/05/29, p. 25.

Davis, WM 2019, *ApE, A plasmid Editor*, <http://jorgensen.biology.utah.edu/wayned/ape/>.

Denning, G, Jamieson, L, Maquat, LE, Thompson, EA & Fields, AP 2001, 'Cloning of a novel phosphatidylinositol kinase-related kinase: characterization of the human SMG-1 RNA surveillance protein', *J Biol Chem*, vol. 276, no. 25, Jun 22, pp. 22709-22714.

Desmet, FO, Hamroun, D, Lalande, M, Collod-Beroud, G, Claustres, M & Beroud, C 2009, 'Human Splicing Finder: an online bioinformatics tool to predict splicing signals', *Nucleic Acids Res*, vol. 37, no. 9, May, p. e67.

Dever, TE & Green, R 2012, 'The elongation, termination, and recycling phases of translation in eukaryotes', *Cold Spring Harb Perspect Biol*, vol. 4, no. 7, Jul 1, p. a013706.

Dias, SM, Wilson, KF, Rojas, KS, Ambrosio, AL & Cerione, RA 2009, 'The molecular basis for the regulation of the cap-binding complex by the importins', *Nat Struct Mol Biol*, vol. 16, no. 9, Sep, pp. 930-937.

Du, M, Liu, X, Welch, EM, Hirawat, S, Peltz, SW & Bedwell, DM 2008, 'PTC124 is an orally bioavailable compound that promotes suppression of the human CFTR-G542X nonsense allele in a CF mouse model', *Proc Natl Acad Sci U S A*, vol. 105, no. 6, Feb 12, pp. 2064-2069.

Duan, J, Wainwright, MS, Comeron, JM, Saitou, N, Sanders, AR, Gelernter, J & Gejman, PV 2003, 'Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor', *Hum Mol Genet*, vol. 12, no. 3, Feb 1, pp. 205-216.

Durand, S, Cougot, N, Mahuteau-Betzer, F, Nguyen, CH, Grierson, DS, Bertrand, E, Tazi, J & Lejeune, F 2007, 'Inhibition of nonsense-mediated mRNA decay (NMD) by a new chemical molecule reveals the dynamic of NMD factors in P-bodies', *J Cell Biol*, vol. 178, no. 7, Sep 24, pp. 1145-1160.

Durand, S & Lykke-Andersen, J 2013, 'Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells', *Nat Struct Mol Biol*, vol. 20, no. 6, Jun, pp. 702-709.

Eberle, AB, Lykke-Andersen, S, Muhlemann, O & Jensen, TH 2009, 'SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells', *Nat Struct Mol Biol*, vol. 16, no. 1, Jan, pp. 49-55.

Enver, T, Raich, N, Ebens, AJ, Papayannopoulou, T, Costantini, F & Stamatoyannopoulos, G 1990, 'Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice', *Nature*, vol. 344, no. 6264, Mar 22, pp. 309-313.

Favaro, FP, Alvizi, L, Zechi-Ceide, RM, Bertola, D, Felix, TM, de Souza, J, Raskin, S, Twigg, SR, Weiner, AM, Armas, P, Margarit, E, Calcaterra, NB, Andersen, GR, McGowan, SJ, Wilkie, AO, Richieri-Costa, A, de Almeida, ML & Passos-Bueno, MR 2014, 'A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects', *Am J Hum Genet*, vol. 94, no. 1, Jan 2, pp. 120-128.

Feng, D, Su, RC, Zou, L, Triggs-Raine, B, Huang, S & Xie, J 2015, 'Increase of a group of PTC(+) transcripts by curcumin through inhibition of the NMD pathway', *Biochim Biophys Acta*, vol. 1849, no. 8, Aug, pp. 1104-1115.

Ferraiuolo, MA, Lee, CS, Ler, LW, Hsu, JL, Costa-Mattioli, M, Luo, MJ, Reed, R & Sonenberg, N 2004, 'A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay', *Proc Natl Acad Sci U S A*, vol. 101, no. 12, Mar 23, pp. 4118-4123.

Franks, TM, Singh, G & Lykke-Andersen, J 2010, 'Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense- mediated mRNA decay', *Cell*, vol. 143, no. 6, Dec 10, pp. 938-950.

Frischmeyer-Guerrerio, PA, Montgomery, RA, Warren, DS, Cooke, SK, Lutz, J, Sonnenday, CJ, Guerrerio, AL & Dietz, HC 2011, 'Perturbation of thymocyte development in nonsense-mediated decay (NMD)-deficient mice', *Proc Natl Acad Sci U S A*, vol. 108, no. 26, Jun 28, pp. 10638-10643.

Frischmeyer, PA & Dietz, HC 1999, 'Nonsense-mediated mRNA decay in health and disease', *Hum Mol Genet*, vol. 8, no. 10, pp. 1893-1900.

Frizzell, KA, Rynearson, SG & Metzstein, MM 2012, 'Drosophila mutants show NMD pathway activity is reduced, but not eliminated, in the absence of Smg6', *Rna*, vol. 18, no. 8, Aug, pp. 1475-1486.

Frolova, L, Le Goff, X, Zhouravleva, G, Davydova, E, Philippe, M & Kisselev, L 1996, 'Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase', *Rna*, vol. 2, no. 4, Apr, pp. 334-341.

Fukuhara, N, Ebert, J, Unterholzner, L, Lindner, D, Izaurralde, E & Conti, E 2005, 'SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway', *Mol Cell*, vol. 17, no. 4, Feb 18, pp. 537-547.

Furniss, D, Critchley, P, Giele, H & Wilkie, AO 2007, 'Nonsense-mediated decay and the molecular pathogenesis of mutations in SALL1 and GLI3', *Am J Med Genet A*, vol. 143a, no. 24, Dec 15, pp. 3150-3160.

Gardner, LB 2008, 'Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response', *Mol Cell Biol*, vol. 28, no. 11, Jun, pp. 3729-3741.

Gatfield, D, Unterholzner, L, Ciccarelli, FD, Bork, P & Izaurralde, E 2003, 'Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways', *Embo j*, vol. 22, no. 15, Aug 1, pp. 3960-3970.

Ge, Z, Quek, BL, Beemon, KL & Hogg, JR 2016, 'Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway', *Elife*, vol. 5, Jan 8.

Gehring, NH, Kunz, JB, Neu-Yilik, G, Breit, S, Viegas, MH, Hentze, MW & Kulozik, AE 2005, 'Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements', *Mol Cell*, vol. 20, no. 1, Oct 7, pp. 65-75.

Gehring, NH, Lamprinaki, S, Hentze, MW & Kulozik, AE 2009, 'The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay', *PLoS Biol*, vol. 7, no. 5, May, p. e1000120.

Gehring, NH, Neu-Yilik, G, Schell, T, Hentze, MW & Kulozik, AE 2003, 'Y14 and hUpf3b form an NMD-activating complex', *Mol Cell*, vol. 11, no. 4, Apr, pp. 939-949.

Geiger, SK, Bar, H, Ehlermann, P, Walde, S, Rutschow, D, Zeller, R, Ivandic, BT, Zentgraf, H, Katus, HA, Herrmann, H & Weichenhan, D 2008, 'Incomplete nonsense-mediated decay of mutant lamin A/C mRNA provokes dilated cardiomyopathy and ventricular tachycardia', *J Mol Med (Berl)*, vol. 86, no. 3, Mar, pp. 281-289.

Gerbracht, JV, Boehm, V & Gehring, NH 2017, 'Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells', *Sci Rep*, vol. 7, no. 1, 2017/09/06, p. 10616.

Ghigna, C, Giordano, S, Shen, H, Benvenuto, F, Castiglioni, F, Comoglio, PM, Green, MR, Riva, S & Biamonti, G 2005, 'Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene', *Mol Cell*, vol. 20, no. 6, Dec 22, pp. 881-890.

Giorgi, C, Yeo, GW, Stone, ME, Katz, DB, Burge, C, Turrigiano, G & Moore, MJ 2007, 'The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression', *Cell*, vol. 130, no. 1, Jul 13, pp. 179-191.

Goetz, AE & Wilkinson, M 2017, 'Stress and the nonsense-mediated RNA decay pathway', *Cell Mol Life Sci*, vol. 74, no. 19, Oct, pp. 3509-3531.

Gorlich, D, Kraft, R, Kostka, S, Vogel, F, Hartmann, E, Laskey, RA, Mattaj, IW & Izaurralde, E 1996, 'Importin provides a link between nuclear protein import and U snRNA export', *Cell*, vol. 87, no. 1, Oct 4, pp. 21-32.

Gregory-Evans, CY, Wang, X, Wasan, KM, Zhao, J, Metcalfe, AL & Gregory-Evans, K 2014, 'Postnatal manipulation of Pax6 dosage reverses congenital tissue malformation defects', *J Clin Invest*, vol. 124, no. 1, Jan, pp. 111-116.

Grigoreva, TA, Tribulovich, VG, Garabadzhiu, AV, Melino, G & Barlev, NA 2015, 'The 26S proteasome is a multifaceted target for anti-cancer therapies', *Oncotarget*, vol. 6, no. 28, Sep 22, pp. 24733-24749.

Gulsuner, S, Walsh, T, Watts, AC, Lee, MK, Thornton, AM, Casadei, S, Rippey, C, Shahin, H, Nimgaonkar, VL, Go, RC, Savage, RM, Swerdlow, NR, Gur, RE, Braff, DL, King, MC & McClellan, JM 2013, 'Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network', *Cell*, vol. 154, no. 3, Aug 1, pp. 518-529.

Guo, X & Wang, XF 2009, 'Signaling cross-talk between TGF-beta/BMP and other pathways', *Cell Res*, vol. 19, no. 1, Jan, pp. 71-88.

Guzowski, JF, Timlin, JA, Roysam, B, McNaughton, BL, Worley, PF & Barnes, CA 2005, 'Mapping behaviorally relevant neural circuits with immediate-early gene expression', *Curr Opin Neurobiol*, vol. 15, no. 5, Oct, pp. 599-606.

Hall, GW & Thein, S 1994, 'Nonsense codon mutations in the terminal exon of the beta-globin gene are not associated with a reduction in beta-mRNA accumulation: a mechanism for the phenotype of dominant beta-thalassemia', *Blood*, vol. 83, no. 8, Apr 15, pp. 2031-2037.

Hamid, FM & Makeyev, EV 2014, 'Regulation of mRNA abundance by polypyrimidine tract-binding protein-controlled alternate 5' splice site choice', *PLoS Genet*, vol. 10, no. 11, Nov, p. e1004771.

Han, X, Wei, Y, Wang, H, Wang, F, Ju, Z & Li, T 2018, 'Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology', *Nucleic Acids Res*, vol. 46, no. 3, Feb 16, pp. 1038-1051.

Harding, HP, Zhang, Y, Bertolotti, A, Zeng, H & Ron, D 2000, 'Perk is essential for translational regulation and cell survival during the unfolded protein response', *Mol Cell*, vol. 5, no. 5, May, pp. 897-904.

He, F, Li, X, Spatrick, P, Casillo, R, Dong, S & Jacobson, A 2003, 'Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast', *Mol Cell*, vol. 12, no. 6, Dec, pp. 1439-1452.

Hetz, C 2012, 'The unfolded protein response: controlling cell fate decisions under ER stress and beyond', *Nat Rev Mol Cell Biol*, vol. 13, no. 2, Jan 18, pp. 89-102.

Hetz, C & Papa, FR 2018, 'The Unfolded Protein Response and Cell Fate Control', *Mol Cell*, vol. 69, no. 2, Jan 18, pp. 169-181.

Hirawat, S, Welch, EM, Elfring, GL, Northcutt, VJ, Paushkin, S, Hwang, S, Leonard, EM, Almstead, NG, Ju, W, Peltz, SW & Miller, LL 2007, 'Safety, tolerability, and pharmacokinetics of PTC124, a nonaminoglycoside nonsense mutation suppressor, following single- and multiple-dose administration to healthy male and female adult volunteers', *J Clin Pharmacol*, vol. 47, no. 4, Apr, pp. 430-444.

Holbrook, JA, Neu-Yilik, G, Hentze, MW & Kulozik, AE 2004, 'Nonsense-mediated decay approaches the clinic', *Nat Genet*, vol. 36, no. 8, Aug, pp. 801-808.

Hosoda, N, Kim, YK, Lejeune, F & Maquat, LE 2005, 'CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells', *Nat Struct Mol Biol*, vol. 12, no. 10, Oct, pp. 893-901.

Hu, X, Eszterhas, S, Pallazzi, N, Bouhassira, EE, Fields, J, Tanabe, O, Gerber, SA, Bulger, M, Engel, JD, Groudine, M & Fiering, S 2007, 'Transcriptional interference among the murine beta-like globin genes', *Blood*, vol. 109, no. 5, Mar 1, pp. 2210-2216.

Hu, Z, Yau, C & Ahmed, AA 2017, 'A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay', *Nat Commun*, vol. 8, p. 15943.

Huang, L, Lou, CH, Chan, W, Shum, EY, Shao, A, Stone, E, Karam, R, Song, HW & Wilkinson, MF 2011, 'RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD', *Mol Cell*, vol. 43, no. 6, Sep 16, pp. 950-961.

Huang, L, Shum, EY, Jones, SH, Lou, CH, Dumdie, J, Kim, H, Roberts, AJ, Jolly, LA, Espinoza, JL, Skarbrevik, DM, Phan, MH, Cook-Andersen, H, Swerdlow, NR, Gecz, J & Wilkinson, MF 2018, 'A Upf3b-mutant mouse model with behavioral and neurogenesis defects', *Mol Psychiatry*, vol. 23, no. 8, Aug, pp. 1773-1786.

Huang, L & Wilkinson, MFJWIRR 2012, 'Regulation of nonsense-mediated mRNA decay', vol. 3, no. 6, pp. 807-828.

Huntzinger, E, Kashima, I, Fauser, M, Sauliere, J & Izaurralde, E 2008, 'SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan', *Rna*, vol. 14, no. 12, Dec, pp. 2609-2617.

Hwang, J & Maquat, LE 2011, 'Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: to die or not to die, that is the question', *Curr Opin Genet Dev*, vol. 21, no. 4, Aug, pp. 422-430.

Iarovaia, OV, Kovina, AP, Petrova, NV, Razin, SV, Ioudinkova, ES, Vassetzky, YS & Ulianov, SV 2018, 'Genetic and Epigenetic Mechanisms of beta-Globin Gene Switching', *Biochemistry (Mosc)*, vol. 83, no. 4, Apr, pp. 381-392.

Imamachi, N, Tani, H & Akimitsu, N 2012, 'Up-frameshift protein 1 (UPF1): multitalented entertainer in RNA decay', *Drug Discov Ther*, vol. 6, no. 2, Apr, pp. 55-61.

Ishigaki, Y, Li, X, Serin, G & Maquat, LE 2001, 'Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20', *Cell*, vol. 106, no. 5, Sep 7, pp. 607-617.

Isken, O & Maquat, LE 2007, 'Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function', *Genes Dev*, vol. 21, no. 15, Aug 1, pp. 1833-1856.

Isken, O & Maquat, LE 2008, 'The multiple lives of NMD factors: balancing roles in gene and genome regulation', *Nat Rev Genet*, vol. 9, no. 9, Sep, pp. 699-712.

Ito, K, Patel, PN, Gorham, JM, McDonough, B, DePalma, SR, Adler, EE, Lam, L, MacRae, CA, Mohiuddin, SM, Fatkin, D, Seidman, CE & Seidman, JG 2017, 'Identification of pathogenic gene mutations in LMNA and MYBPC3 that alter RNA splicing', *Proc Natl Acad Sci U S A*, vol. 114, no. 29, Jul 18, pp. 7689-7694.

Ivanov, PV, Gehring, NH, Kunz, JB, Hentze, MW & Kulozik, AE 2008, 'Interactions between UPF1, eRFs, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways', *Embo j*, vol. 27, no. 5, Mar 5, pp. 736-747.

Izumikawa, K, Yoshikawa, H, Ishikawa, H, Nobe, Y, Yamauchi, Y, Philipsen, S, Simpson, RJ, Isobe, T & Takahashi, N 2016, 'Chtop (Chromatin target of Prmt1) auto-regulates its expression level via intron retention and nonsense-mediated decay of its own mRNA', *Nucleic Acids Res*, vol. 44, no. 20, Nov 16, pp. 9847-9859.

Jackson, RJ, Hellen, CU & Pestova, TV 2010, 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nat Rev Mol Cell Biol*, vol. 11, no. 2, Feb, pp. 113-127.

Jaffrey, SR & Wilkinson, MF 2018, 'Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease', *Nature Reviews Neuroscience*, 2018/11/08.

Jang, HH 2018, 'Regulation of Protein Degradation by Proteasomes in Cancer', *J Cancer Prev*, vol. 23, no. 4, Dec, pp. 153-161.

Jaworski, A, Long, H & Tessier-Lavigne, M 2010, 'Collaborative and specialized functions of Robo1 and Robo2 in spinal commissural axon guidance', *J Neurosci*, vol. 30, no. 28, Jul 14, pp. 9445-9453.

Johnson, JL, Stoica, L, Liu, Y, Zhu, PJ, Bhattacharya, A, Buffington, S, Huq, R, Eissa, NT, Larsson, O, Porse, BT, Domingo, D, Nawaz, U, Carroll, R, Jolly, L, Scerri, TS, Kim, HG, Brignell, A, Coleman, MJ, Braden, R, Kini, U, Jackson, V, Baxter, A, Bahlo, M, Scheffer, IE, Amor, DJ, Hildebrand, MS, Bonnen, PE, Beeton, C, Gecz, J, Morgan, AT & Costa-Mattioli, M 2019, 'Inhibition of Upf2-Dependent Nonsense-Mediated Decay Leads to Behavioral and Neurophysiological Abnormalities by Activating the Immune Response', *Neuron*, Sep 27.

Jolly, LA, Homan, CC, Jacob, R, Barry, S & Gecz, J 2013, 'The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth', *Hum Mol Genet*, vol. 22, no. 23, Dec 1, pp. 4673-4687.

Karam, R, Carvalho, J, Bruno, I, Graziadio, C, Senz, J, Huntsman, D, Carneiro, F, Seruca, R, Wilkinson, MF & Oliveira, C 2008, 'The NMD mRNA surveillance pathway downregulates aberrant E-cadherin transcripts in gastric cancer cells and in CDH1 mutation carriers', *Oncogene*, vol. 27, no. 30, Jul 10, pp. 4255-4260.

Karam, R, Lou, CH, Kroeger, H, Huang, L, Lin, JH & Wilkinson, MF 2015, 'The unfolded protein response is shaped by the NMD pathway', *EMBO Rep*, vol. 16, no. 5, May, pp. 599-609.

Karam, R & Wilkinson, M 2012, 'A conserved microRNA/NMD regulatory circuit controls gene expression', *RNA Biology*, vol. 9, no. 1, Jan, pp. 22-26.

Kashima, I, Yamashita, A, Izumi, N, Kataoka, N, Morishita, R, Hoshino, S, Ohno, M, Dreyfuss, G & Ohno, S 2006, '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 Dev*, vol. 20, no. 3, Feb 1, pp. 355-367.

Kashima, R & Hata, A 2018, 'The role of TGF-β superfamily signaling in neurological disorders', *Acta Biochim Biophys Sin (Shanghai)*, vol. 50, no. 1, Jan, pp. 106-120.

Kebaara, BW & Atkin, AL 2009, 'Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae', *Nucleic Acids Res*, vol. 37, no. 9, May, pp. 2771-2778.

Keeling, KM & Bedwell, DM 2011, 'Suppression of nonsense mutations as a therapeutic approach to treat genetic diseases', *Wiley Interdiscip Rev RNA*, vol. 2, no. 6, Nov, pp. 837-852.

Keeling, KM, Wang, D, Dai, Y, Murugesan, S, Chenna, B, Clark, J, Belakhov, V, Kandasamy, J, Velu, SE, Baasov, T & Bedwell, DM 2013, 'Attenuation of nonsense-mediated mRNA decay enhances in vivo nonsense suppression', *PLoS One*, vol. 8, no. 4, p. e60478.

Kerem, E, Konstan, MW, De Boeck, K, Accurso, FJ, Sermet-Gaudelus, I, Wilschanski, M, Elborn, JS, Melotti, P, Bronsveld, I, Fajac, I, Malfroot, A, Rosenbluth, DB, Walker, PA, McColley, SA, Knoop, C, Quattrucci, S, Rietschel, E, Zeitlin, PL, Barth, J, Elfring, GL, Welch, EM, Branstrom, A, Spiegel, RJ, Peltz, SW, Ajayi, T & Rowe, SM 2014, 'Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial', *Lancet Respir Med*, vol. 2, no. 7, Jul, pp. 539-547.

Kerr, TP, Sewry, CA, Robb, SA & Roberts, RG 2001, 'Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay?', *Hum Genet*, vol. 109, no. 4, Oct, pp. 402-407.

Kervestin, S & Jacobson, A 2012, 'NMD: a multifaceted response to premature translational termination', *Nat Rev Mol Cell Biol*, vol. 13, no. 11, Nov, pp. 700-712.

Kim, JH, Lee, SR, Li, LH, Park, HJ, Park, JH, Lee, KY, Kim, MK, Shin, BA & Choi, SY 2011, 'High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice', *PLoS One*, vol. 6, no. 4, p. e18556.

Kim, VN, Kataoka, N & Dreyfuss, G 2001, 'Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex', *Science*, vol. 293, no. 5536, Sep 7, pp. 1832-1836.

Kimchi-Sarfaty, C, Oh, JM, Kim, IW, Sauna, ZE, Calcagno, AM, Ambudkar, SV & Gottesman, MM 2007, 'A "silent" polymorphism in the MDR1 gene changes substrate specificity', *Science*, vol. 315, no. 5811, Jan 26, pp. 525-528.

Kiselev, AV, Ostapenko, OV, Rogozhkina, EV, Kholod, NS, Seit Nebi, AS, Baranov, AN, Lesina, EA, Ivashchenko, TE, Sabetskii, VA, Shavlovskii, MM, Rechinskii, VO, Kiselev, LL & Baranov, VC 2002, '[Suppression of nonsense mutations in the Dystrophin gene by a suppressor tRNA gene]', *Mol Biol (Mosk)*, vol. 36, no. 1, Jan-Feb, pp. 43-47.

Kisielow, J, Nairn, AC & Karjalainen, K 2001, 'TARPP, a novel protein that accompanies TCR gene rearrangement and thymocyte education', *Eur J Immunol*, vol. 31, no. 4, Apr, pp. 1141-1149.

Krawczak, M, Ball, EV & Cooper, DN 1998, 'Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes', *Am J Hum Genet*, vol. 63, no. 2, Aug, pp. 474-488.

Kunz, JB, Neu-Yilik, G, Hentze, MW, Kulozik, AE & Gehring, NH 2006, 'Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation', *Rna*, vol. 12, no. 6, Jun, pp. 1015-1022.

Kurosaki, T & Maquat, LE 2016, 'Nonsense-mediated mRNA decay in humans at a glance', *J Cell Sci*, vol. 129, no. 3, Feb 1, pp. 461-467.

Lappalainen, T, Sammeth, M, Friedlander, MR, t Hoen, PA, Monlong, J, Rivas, MA, Gonzalez-Porta, M, Kurbatova, N, Griebel, T, Ferreira, PG, Barann, M, Wieland, T, Greger, L, van Iterson, M, Almlof, J, Ribeca, P, Pulyakhina, I, Esser, D, Giger, T, Tikhonov, A, Sultan, M, Bertier, G, MacArthur, DG, Lek, M, Lizano, E, Buermans, HP, Padioleau, I, Schwarzmayr, T, Karlberg, O, Ongen, H, Kilpinen, H, Beltran, S, Gut, M, Kahlem, K, Amstislavskiy, V, Stegle, O, Pirinen, M, Montgomery, SB, Donnelly, P, McCarthy, MI, Flicek, P, Strom, TM, Lehrach, H, Schreiber, S, Sudbrak, R, Carracedo, A, Antonarakis, SE, Hasler, R, Syvanen, AC, van Ommen, GJ, Brazma, A, Meitinger, T, Rosenstiel, P, Guigo, R, Gut, IG, Estivill, X & Dermitzakis, ET 2013, 'Transcriptome and genome sequencing uncovers functional variation in humans', *Nature*, vol. 501, no. 7468, Sep 26, pp. 506-511.

Lareau, LF, Brooks, AN, Soergel, DA, Meng, Q & Brenner, SE 2007, 'The coupling of alternative splicing and nonsense-mediated mRNA decay', *Adv Exp Med Biol*, vol. 623, pp. 190-211.

Laumonnier, F, Shoubridge, C, Antar, C, Nguyen, LS, Van Esch, H, Kleefstra, T, Briault, S, Fryns, JP, Hamel, B, Chelly, J, Ropers, HH, Ronce, N, Blesson, S, Moraine, C, Gecz, J & Raynaud, M 2010, 'Mutations of the UPF3B gene, which encodes a protein widely expressed in neurons, are associated with nonspecific mental retardation with or without autism', *Mol Psychiatry*, vol. 15, no. 7, Jul, pp. 767-776.

Le Hir, H, Gatfield, D, Izaurralde, E & Moore, MJ 2001, 'The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay', *Embo j*, vol. 20, no. 17, Sep 3, pp. 4987-4997.

Le Hir, H, Izaurralde, E, Maquat, LE & Moore, MJ 2000, 'The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions', *Embo j*, vol. 19, no. 24, Dec 15, pp. 6860-6869.

Le Hir, H, Sauliere, J & Wang, Z 2016, 'The exon junction complex as a node of post-transcriptional networks', *Nat Rev Mol Cell Biol*, vol. 17, no. 1, Jan, pp. 41-54.

LeBlanc, JJ & Beemon, KL 2004, 'Unspliced Rous sarcoma virus genomic RNAs are translated and subjected to nonsense-mediated mRNA decay before packaging', *J Virol*, vol. 78, no. 10, May, pp. 5139-5146.

Leeds, P, Wood, JM, Lee, BS & Culbertson, MR 1992, 'Gene products that promote mRNA turnover in Saccharomyces cerevisiae', *Mol Cell Biol*, vol. 12, no. 5, May, pp. 2165-2177.

Lejeune, F, Ishigaki, Y, Li, X & Maquat, LE 2002, 'The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling', *Embo j*, vol. 21, no. 13, Jul 1, pp. 3536-3545.

Lejeune, F, Li, X & Maquat, LE 2003, 'Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities', *Mol Cell*, vol. 12, no. 3, Sep, pp. 675-687.

Lelimousin, M, Noirclerc-Savoye, M, Lazareno-Saez, C, Paetzold, B, Le Vot, S, Chazal, R, Macheboeuf, P, Field, MJ, Bourgeois, D & Royant, A 2009, 'Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield', *Biochemistry*, vol. 48, no. 42, Oct 27, pp. 10038-10046.

Lewis, BP, Green, RE & Brenner, SE 2003, 'Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans', *Proc Natl Acad Sci U S A*, vol. 100, no. 1, Jan 7, pp. 189-192.

Li, A & Swift, M 2000, 'Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients', *Am J Med Genet*, vol. 92, no. 3, May 29, pp. 170-177.

Li, T, Shi, Y, Wang, P, Guachalla, LM, Sun, B, Joerss, T, Chen, YS, Groth, M, Krueger, A, Platzer, M, Yang, YG, Rudolph, KL & Wang, ZQ 2015, 'Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay', *Embo j*, vol. 34, no. 12, Jun 12, pp. 1630-1647.

Lim, ET, Raychaudhuri, S, Sanders, SJ, Stevens, C, Sabo, A, MacArthur, DG, Neale, BM, Kirby, A, Ruderfer, DM, Fromer, M, Lek, M, Liu, L, Flannick, J, Ripke, S, Nagaswamy, U, Muzny, D, Reid, JG, Hawes, A, Newsham, I, Wu, Y, Lewis, L, Dinh, H, Gross, S, Wang, LS, Lin, CF, Valladares, O, Gabriel, SB, dePristo, M, Altshuler, DM, Purcell, SM, State, MW, Boerwinkle, E, Buxbaum, JD, Cook, EH, Gibbs, RA, Schellenberg, GD, Sutcliffe, JS, Devlin, B, Roeder, K & Daly, MJ 2013, 'Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders', *Neuron*, vol. 77, no. 2, Jan 23, pp. 235-242.

Linde, L, Boelz, S, Neu-Yilik, G, Kulozik, AE & Kerem, B 2007, 'The efficiency of nonsense-mediated mRNA decay is an inherent character and varies among different cells', *European Journal Of Human Genetics*, vol. 15, 07/11/online, p. 1156.

Linde, L, Boelz, S, Nissim-Rafinia, M, Oren, YS, Wilschanski, M, Yaacov, Y, Virgilis, D, Neu-Yilik, G, Kulozik, AE, Kerem, E & Kerem, B 2007, 'Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin', *J Clin Invest*, vol. 117, no. 3, Mar 1, pp. 683-692.

Lindeboom, RG, Supek, F & Lehner, B 2016, 'The rules and impact of nonsense-mediated mRNA decay in human cancers', *Nat Genet*, vol. 48, no. 10, Oct, pp. 1112-1118.

Linder, B, Fischer, U & Gehring, NH 2015, 'mRNA metabolism and neuronal disease', *FEBS Lett*, vol. 589, no. 14, Jun 22, pp. 1598-1606.

Liu, C, Karam, R, Zhou, Y, Su, F, Ji, Y, Li, G, Xu, G, Lu, L, Wang, C, Song, M, Zhu, J, Wang, Y, Zhao, Y, Foo, WC, Zuo, M, Valasek, MA, Javle, M, Wilkinson, MF & Lu, Y 2014, 'The UPF1 RNA surveillance gene is commonly mutated in pancreatic adenosquamous carcinoma', *Nat Med*, vol. 20, no. 6, Jun, pp. 596-598.

Liu, N, Hargreaves, VV, Zhu, Q, Kurland, JV, Hong, J, Kim, W, Sher, F, Macias-Trevino, C, Rogers, JM, Kurita, R, Nakamura, Y, Yuan, GC, Bauer, DE, Xu, J, Bulyk, ML & Orkin, SH 2018, 'Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch', *Cell*, vol. 173, no. 2, Apr 5, pp. 430-442.e417.

Liu, Z, Chen, O, Wall, JBJ, Zheng, M, Zhou, Y, Wang, L, Ruth Vaseghi, H, Qian, L & Liu, J 2017, 'Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector', *Sci Rep*, vol. 7, no. 1, 2017/05/19, p. 2193.

Long, H, Sabatier, C, Ma, L, Plump, A, Yuan, W, Ornitz, DM, Tamada, A, Murakami, F, Goodman, CS & Tessier-Lavigne, M 2004, 'Conserved roles for Slit and Robo proteins in midline commissural axon guidance', *Neuron*, vol. 42, no. 2, Apr 22, pp. 213-223.

Losson, R & Lacroute, F 1979, 'Interference of nonsense mutations with eukaryotic messenger RNA stability', *Proceedings of the National Academy of Sciences*, vol. 76, no. 10, p. 5134.

Lou, C-H, Dumdie, J, Goetz, A, Shum, EY, Brafman, D, Liao, X, Mora-Castilla, S, Ramaiah, M, Cook-Andersen, H & Laurent, L 2016, 'Nonsense-mediated RNA decay influences human embryonic stem cell fate', *Stem cell reports*, vol. 6, no. 6, pp. 844-857.

Lou, CH, Shao, A, Shum, EY, Espinoza, JL, Huang, L, Karam, R & Wilkinson, MF 2014, 'Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway', *Cell Rep*, vol. 6, no. 4, Feb 27, pp. 748-764.

Lu, J, Plank, TD, Su, F, Shi, X, Liu, C, Ji, Y, Li, S, Huynh, A, Shi, C, Zhu, B, Yang, G, Wu, Y, Wilkinson, MF & Lu, Y 2016, 'The nonsense-mediated RNA decay pathway is disrupted in inflammatory myofibroblastic tumors', *J Clin Invest*, vol. 126, no. 8, Aug 1, pp. 3058-3062.

Lucas, BA, Lavi, E, Shiue, L, Cho, H, Katzman, S, Miyoshi, K, Siomi, MC, Carmel, L, Ares, M, Jr. & Maquat, LE 2018, 'Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay', *Proc Natl Acad Sci U S A*, vol. 115, no. 5, Jan 30, pp. 968-973.

Lykke-Andersen, J, Shu, MD & Steitz, JA 2000, 'Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon', *Cell*, vol. 103, no. 7, Dec 22, pp. 1121-1131.

Lykke-Andersen, J, Shu, MD & Steitz, JA 2001, 'Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1', *Science*, vol. 293, no. 5536, Sep 7, pp. 1836-1839.

Lykke-Andersen, S, Chen, Y, Ardal, BR, Lilje, B, Waage, J, Sandelin, A & Jensen, TH 2014, 'Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes', *Genes Dev*, vol. 28, no. 22, Nov 15, pp. 2498-2517.

Lykke-Andersen, S & Jensen, TH 2015, 'Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes', *Nat Rev Mol Cell Biol*, vol. 16, no. 11, Nov, pp. 665-677.

Lynch, SA, Nguyen, LS, Ng, LY, Waldron, M, McDonald, D & Gecz, J 2012, 'Broadening the phenotype associated with mutations in UPF3B: two further cases with renal dysplasia and variable developmental delay', *Eur J Med Genet*, vol. 55, no. 8-9, Aug-Sep, pp. 476-479.

Ma, Y & Hendershot, LM 2004, 'The role of the unfolded protein response in tumour development: friend or foe?', *Nat Rev Cancer*, vol. 4, no. 12, Dec, pp. 966-977.

MacArthur, DG, Balasubramanian, S, Frankish, A, Huang, N, Morris, J, Walter, K, Jostins, L, Habegger, L, Pickrell, JK, Montgomery, SB, Albers, CA, Zhang, ZD, Conrad, DF, Lunter, G, Zheng, H, Ayub, Q, DePristo, MA, Banks, E, Hu, M, Handsaker, RE, Rosenfeld, JA, Fromer, M, Jin, M, Mu, XJ, Khurana, E, Ye, K, Kay, M, Saunders, GI, Suner, MM, Hunt, T, Barnes, IH, Amid, C, Carvalho-Silva, DR, Bignell, AH, Snow, C, Yngvadottir, B, Bumpstead, S, Cooper, DN, Xue, Y, Romero, IG, Wang, J, Li, Y, Gibbs, RA, McCarroll, SA, Dermitzakis, ET, Pritchard, JK, Barrett, JC, Harrow, J, Hurles, ME, Gerstein, MB & Tyler-Smith, C 2012, 'A systematic survey of loss-of-function variants in human protein-coding genes', *Science*, vol. 335, no. 6070, Feb 17, pp. 823-828.

Macaya, D, Katsanis, SH, Hefferon, TW, Audlin, S, Mendelsohn, NJ, Roggenbuck, J & Cutting, GR 2009, 'A synonymous mutation in TCOF1 causes Treacher Collins syndrome due to mis-splicing of a constitutive exon', *Am J Med Genet A*, vol. 149a, no. 8, Aug, pp. 1624-1627.

Maquat, LE 1995, 'When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells', *Rna*, vol. 1, no. 5, Jul, pp. 453-465.

Maquat, LE 2004, 'Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics', *Nat Rev Mol Cell Biol*, vol. 5, no. 2, Feb, pp. 89-99.

Maquat, LE, Kinniburgh, AJ, Rachmilewitz, EA & Ross, J 1981, 'Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia', *Cell*, vol. 27, no. 3 Pt 2, Dec, pp. 543-553.

Maquat, LE, Tarn, WY & Isken, O 2010, 'The pioneer round of translation: features and functions', *Cell*, vol. 142, no. 3, Aug 6, pp. 368-374.

Martin, L & Gardner, LB 2015, 'Stress-induced inhibition of nonsense-mediated RNA decay regulates intracellular cystine transport and intracellular glutathione through regulation of the cystine/glutamate exchanger SLC7A11', *Oncogene*, vol. 34, no. 32, Aug 6, pp. 4211-4218.

Martin, L, Grigoryan, A, Wang, D, Wang, J, Breda, L, Rivella, S, Cardozo, T & Gardner, LB 2014, 'Identification and characterization of small molecules that inhibit nonsense-mediated RNA decay and suppress nonsense p53 mutations', *Cancer Res*, vol. 74, no. 11, Jun 1, pp. 3104-3113.

McColl, B, Kao, BR, Lourthai, P, Chan, K, Wardan, H, Roosjen, M, Delagneau, O, Gearing, LJ, Blewitt, ME, Svasti, S, Fucharoen, S & Vadolas, J 2014, 'An in vivo model for analysis of developmental erythropoiesis and globin gene regulation', *Faseb j*, vol. 28, no. 5, May, pp. 2306-2317.

McColl, B & Vadolas, J 2016, 'Animal models of β-hemoglobinopathies: utility and limitations', *J Blood Med*, vol. 7, pp. 263-274.

McConnell, SC, Huo, Y, Liu, S & Ryan, TM 2011, 'Human globin knock-in mice complete fetal-to-adult hemoglobin switching in postnatal development', *Mol Cell Biol*, vol. 31, no. 4, Feb, pp. 876-883.

McGlincy, NJ & Smith, CW 2008, 'Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?', *Trends Biochem Sci*, vol. 33, no. 8, Aug, pp. 385-393.

McIlwain, DR, Pan, Q, Reilly, PT, Elia, AJ, McCracken, S, Wakeham, AC, Itie-Youten, A, Blencowe, BJ & Mak, TW 2010, 'Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay', *Proc Natl Acad Sci U S A*, vol. 107, no. 27, Jul 6, pp. 12186-12191.

Medghalchi, SM, Frischmeyer, PA, Mendell, JT, Kelly, AG, Lawler, AM & Dietz, HC 2001, 'Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability', *Hum Mol Genet*, vol. 10, no. 2, Jan 15, pp. 99-105.

Mefford, HC, Sharp, AJ, Baker, C, Itsara, A, Jiang, Z, Buysse, K, Huang, S, Maloney, VK, Crolla, JA, Baralle, D, Collins, A, Mercer, C, Norga, K, de Ravel, T, Devriendt, K, Bongers, EM, de Leeuw, N, Reardon, W, Gimelli, S, Bena, F, Hennekam, RC, Male, A, Gaunt, L, Clayton-Smith, J, Simonic, I, Park, SM, Mehta, SG, Nik-Zainal, S, Woods, CG, Firth, HV, Parkin, G, Fichera, M, Reitano, S, Lo Giudice, M, Li, KE, Casuga, I, Broomer, A, Conrad, B, Schwerzmann, M, Raber, L, Gallati, S, Striano, P, Coppola, A, Tolmie, JL, Tobias, ES, Lilley, C, Armengol, L, Spysschaert, Y, Verloo, P, De Coene, A, Goossens, L, Mortier, G, Speleman, F, van Binsbergen, E, Nelen, MR, Hochstenbach, R, Poot, M, Gallagher, L, Gill, M, McClellan, J, King, MC, Regan, R, Skinner, C, Stevenson, RE, Antonarakis, SE, Chen, C, Estivill, X, Menten, B, Gimelli, G, Gribble, S, Schwartz, S, Sutcliffe, JS, Walsh, T, Knight, SJ, Sebat, J, Romano, C, Schwartz, CE, Veltman, JA, de Vries, BB, Vermeesch, JR, Barber, JC, Willatt, L, Tassabehji, M & Eichler, EE 2008, 'Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes', *N Engl J Med*, vol. 359, no. 16, Oct 16, pp. 1685-1699.

Mendell, JT, Medghalchi, SM, Lake, RG, Noensie, EN & Dietz, HC 2000, 'Novel Upf2p orthologues suggest a functional link between translation initiation and nonsense surveillance complexes', *Mol Cell Biol*, vol. 20, no. 23, Dec, pp. 8944-8957.

Mendell, JT, Sharifi, NA, Meyers, JL, Martinez-Murillo, F & Dietz, HC 2004, 'Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise', *Nat Genet*, vol. 36, no. 10, Oct, pp. 1073-1078.

Metze, S, Herzog, VA, Ruepp, MD & Mühlemann, O 2013, 'Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways', *Rna*, vol. 19, no. 10, Oct, pp. 1432-1448.

Metzstein, MM & Krasnow, MA 2006, 'Functions of the nonsense-mediated mRNA decay pathway in Drosophila development', *PLoS Genet*, vol. 2, no. 12, Dec, p. e180.

Miller, JN & Pearce, DA 2014, 'Nonsense-mediated decay in genetic disease: friend or foe?', *Mutat Res Rev Mutat Res*, vol. 762, Oct-Dec, pp. 52-64.

Montgomery, SB, Lappalainen, T, Gutierrez-Arcelus, M & Dermitzakis, ET 2011, 'Rare and common regulatory variation in population-scale sequenced human genomes', *PLoS Genet*, vol. 7, no. 7, Jul, p. e1002144.

Moore, KA & Hollien, J 2012, 'The unfolded protein response in secretory cell function', *Annu Rev Genet*, vol. 46, pp. 165-183.

Moriarty, PM, Reddy, CC & Maquat, LE 1998, '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*, vol. 18, no. 5, May, pp. 2932-2939.

Morozov, AV & Karpov, VL 2018, 'Biological consequences of structural and functional proteasome diversity', *Heliyon*, vol. 4, no. 10, Oct, p. e00894.

Mort, M, Ivanov, D, Cooper, DN & Chuzhanova, NA 2008, 'A meta-analysis of nonsense mutations causing human genetic disease', *Hum Mutat*, vol. 29, no. 8, Aug, pp. 1037-1047.

Mort, M, Sterne-Weiler, T, Li, B, Ball, EV, Cooper, DN, Radivojac, P, Sanford, JR & Mooney, SD 2014, 'MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing', *Genome Biol*, vol. 15, no. 1, January 13, p. R19.

Muhlrad, D & Parker, R 1999, 'Aberrat mRNAs with extended 3'UTRs are substrates for rapid degradation by mRNA surveillance', *RNA*, vol. 5, pp. 1299-1307.

Nagy, E & Maquat, LE 1998, 'A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance', *Trends Biochem Sci*, vol. 23, no. 6, Jun, pp. 198-199.

Neitzel, H 1986, 'A routine method for the establishment of permanent growing lymphoblastoid cell lines', *Hum Genet*, vol. 73, no. 4, August 01, pp. 320-326.

Neu-Yilik, G, Raimondeau, E, Eliseev, B, Yeramala, L, Amthor, B, Deniaud, A, Huard, K, Kerschgens, K, Hentze, MW, Schaffitzel, C & Kulozik, AE 2017, 'Dual function of UPF3B in early and late translation termination', *Embo j*, vol. 36, no. 20, Oct 16, pp. 2968-2986.

Nguyen, LS, Jolly, L, Shoubridge, C, Chan, WK, Huang, L, Laumonnier, F, Raynaud, M, Hackett, A, Field, M, Rodriguez, J, Srivastava, AK, Lee, Y, Long, R, Addington, AM, Rapoport, JL, Suren, S, Hahn, CN, Gamble, J, Wilkinson, MF, Corbett, MA & Gecz, J 2012, 'Transcriptome profiling of UPF3B/NMD-deficient lymphoblastoid cells from patients with various forms of intellectual disability', *Mol Psychiatry*, vol. 17, no. 11, Nov, pp. 1103-1115.

Nguyen, LS, Kim, HG, Rosenfeld, JA, Shen, Y, Gusella, JF, Lacassie, Y, Layman, LC, Shaffer, LG & Gecz, J 2013, 'Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neuro-developmental disorders', *Hum Mol Genet*, vol. 22, no. 9, May 1, pp. 1816-1825.

Nguyen, LS, Wilkinson, MF & Gecz, J 2014, 'Nonsense-mediated mRNA decay: inter-individual variability and human disease', *Neurosci Biobehav Rev*, vol. 46 Pt 2, Oct, pp. 175-186.

Ni, JZ, Grate, L, Donohue, JP, Preston, C, Nobida, N, O'Brien, G, Shiue, L, Clark, TA, Blume, JE & Ares, M, Jr. 2007, 'Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay', *Genes Dev*, vol. 21, no. 6, Mar 15, pp. 708-718.

Nickless, A, Jackson, E, Marasa, J, Nugent, P, Mercer, RW, Piwnica-Worms, D & You, Z 2014, 'Intracellular calcium regulates nonsense-mediated mRNA decay', *Nat Med*, vol. 20, no. 8, Aug, pp. 961-966.

Niwa, H, Yamamura, K & Miyazaki, J 1991, 'Efficient selection for high-expression transfectants with a novel eukaryotic vector', *Gene*, vol. 108, no. 2, Dec 15, pp. 193-199.

Noble, CG & Song, H 2007, 'MLN51 stimulates the RNA-helicase activity of eIF4AIII', *PLoS One*, vol. 2, no. 3, p. e303.

Ohnishi, T, Yamashita, A, Kashima, I, Schell, T, Anders, KR, Grimson, A, Hachiya, T, Hentze, MW, Anderson, P & Ohno, S 2003, 'Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7', *Mol Cell*, vol. 12, no. 5, Nov, pp. 1187-1200.

Okada-Katsuhata, Y, Yamashita, A, Kutsuzawa, K, Izumi, N, Hirahara, F & Ohno, S 2012, 'N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD', *Nucleic Acids Res*, vol. 40, no. 3, Feb, pp. 1251-1266.

Oren, YS, McClure, ML, Rowe, SM, Sorscher, EJ, Bester, AC, Manor, M, Kerem, E, Rivlin, J, Zahdeh, F, Mann, M, Geiger, T & Kerem, B 2014, 'The unfolded protein response affects readthrough of premature termination codons', *EMBO Mol Med*, vol. 6, no. 5, May, pp. 685-701.

Orford, KW & Scadden, DT 2008, 'Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation', *Nat Rev Genet*, vol. 9, no. 2, Feb, pp. 115-128.

Oslowski, CM & Urano, F 2011, 'Measuring ER stress and the unfolded protein response using mammalian tissue culture system', *Methods Enzymol*, vol. 490, pp. 71-92.

Ottens, F & Gehring, NH 2016, 'Physiological and pathophysiological role of nonsense-mediated mRNA decay', *Pflugers Arch*, vol. 468, no. 6, Jun, pp. 1013-1028.

Paillusson, A, Hirschi, N, Vallan, C, Azzalin, CM & Mühlemann, O 2005, 'A GFP-based reporter system to monitor nonsense-mediated mRNA decay', *Nucleic Acids Res*, vol. 33, no. 6, p. e54.

Pakos-Zebrucka, K, Koryga, I, Mnich, K, Ljujic, M, Samali, A & Gorman, AM 2016, 'The integrated stress response', *EMBO Rep*, vol. 17, no. 10, Oct, pp. 1374-1395.

Pan, Q, Shai, O, Lee, LJ, Frey, BJ & Blencowe, BJ 2008, 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet*, vol. 40, no. 12, Dec, pp. 1413-1415.

Park, E & Maquat, LE 2013, 'Staufen-mediated mRNA decay', *Wiley Interdiscip Rev RNA*, vol. 4, no. 4, Jul, pp. 423-435.

Patton, MA & Afzal, AR 2002, 'Robinow syndrome', *J Med Genet*, vol. 39, no. 5, May, pp. 305-310.

Peixeiro, I, Inacio, A, Barbosa, C, Silva, AL, Liebhaber, SA & Romao, L 2012, 'Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations', *Nucleic Acids Res*, vol. 40, no. 3, Feb, pp. 1160-1173.

Pereverzev, AP, Gurskaya, NG, Ermakova, GV, Kudryavtseva, EI, Markina, NM, Kotlobay, AA, Lukyanov, SA, Zaraisky, AG & Lukyanov, KA 2015, 'Method for quantitative analysis of nonsense-mediated mRNA decay at the single cell level', *Sci Rep*, vol. 5, Jan 12, p. 7729.

Perlick, HA, Medghalchi, SM, Spencer, FA, Kendzior, RJ, Jr. & Dietz, HC 1996, 'Mammalian orthologues of a yeast regulator of nonsense transcript stability', *Proc Natl Acad Sci U S A*, vol. 93, no. 20, Oct 1, pp. 10928-10932.

Perrin-Vidoz, L, Sinilnikova, OM, Stoppa-Lyonnet, D, Lenoir, GM & Mazoyer, S 2002, 'The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons', *Hum Mol Genet*, vol. 11, no. 23, Nov 1, pp. 2805-2814.

Pimentel, H, Parra, M, Gee, S, Ghanem, D, An, X, Li, J, Mohandas, N, Pachter, L & Conboy, JG 2014, 'A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis', *Nucleic Acids Res*, vol. 42, no. 6, Apr, pp. 4031-4042.

Pinyol, M, Bea, S, Pla, L, Ribrag, V, Bosq, J, Rosenwald, A, Campo, E & Jares, P 2007, 'Inactivation of RB1 in mantle-cell lymphoma detected by nonsense-mediated mRNA decay pathway inhibition and microarray analysis', *Blood*, vol. 109, no. 12, Jun 15, pp. 5422-5429.

Politano, L, Nigro, G, Nigro, V, Piluso, G, Papparella, S, Paciello, O & Comi, LI 2003, 'Gentamicin administration in Duchenne patients with premature stop codon. Preliminary results', *Acta Myol*, vol. 22, no. 1, May, pp. 15-21.

Popp, MW & Maquat, LE 2015, 'Attenuation of nonsense-mediated mRNA decay facilitates the response to chemotherapeutics', *Nat Commun*, vol. 6, p. 6632.

Popp, MW & Maquat, LE 2018, 'Nonsense-mediated mRNA Decay and Cancer', *Curr Opin Genet Dev*, vol. 48, Feb, pp. 44-50.

Pulak, R & Anderson, P 1993, 'mRNA surveillance by the Caenorhabditis elegans smg genes', *Genes Dev*, vol. 7, no. 10, Oct, pp. 1885-1897.

Quek, BL & Beemon, K 2014, 'Retroviral strategy to stabilize viral RNA', *Curr Opin Microbiol*, vol. 18, Apr, pp. 78-82.

Rajavel, KS & Neufeld, EF 2001, 'Nonsense-mediated decay of human HEXA mRNA', *Mol Cell Biol*, vol. 21, no. 16, Aug, pp. 5512-5519.

Recht, MI, Douthwaite, S & Puglisi, JD 1999, 'Basis for prokaryotic specificity of action of aminoglycoside antibiotics', *Embo j*, vol. 18, no. 11, Jun 1, pp. 3133-3138.

Reddy, JC, Morris, JC, Wang, J, English, MA, Haber, DA, Shi, Y & Licht, JD 1995, 'WT1-mediated transcriptional activation is inhibited by dominant negative mutant proteins', *J Biol Chem*, vol. 270, no. 18, May 5, pp. 10878-10884.

Reese, MG, Eeckman, FH, Kulp, D & Haussler, D 1997, 'Improved splice site detection in Genie', *J Comput Biol*, vol. 4, no. 3, Fall, pp. 311-323.

Rentzsch, P, Witten, D, Cooper, GM, Shendure, J & Kircher, M 2019, 'CADD: predicting the deleteriousness of variants throughout the human genome', *Nucleic Acids Res*, vol. 47, no. D1, Jan 8, pp. D886-d894.

Richards, S, Aziz, N, Bale, S, Bick, D, Das, S, Gastier-Foster, J, Grody, WW, Hegde, M, Lyon, E, Spector, E, Voelkerding, K & Rehm, HL 2015, 'Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College

of Medical Genetics and Genomics and the Association for Molecular Pathology', *Genet Med*, vol. 17, no. 5, May, pp. 405-424.

Richardson, R, Smart, M, Tracey-White, D, Webster, AR & Moosajee, M 2017, 'Mechanism and evidence of nonsense suppression therapy for genetic eye disorders', *Exp Eye Res*, vol. 155, Feb, pp. 24-37.

Ron, D & Walter, P 2007, 'Signal integration in the endoplasmic reticulum unfolded protein response', *Nat Rev Mol Cell Biol*, vol. 8, no. 7, Jul, pp. 519-529.

Rosenfeld, JA, Traylor, RN, Schaefer, GB, McPherson, EW, Ballif, BC, Klopocki, E, Mundlos, S, Shaffer, LG & Aylsworth, AS 2012, 'Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes', *Eur J Hum Genet*, vol. 20, no. 7, Jul, pp. 754-761.

Rowe, SM, Clancy, JP & Wilschanski, M 2011, 'Nasal potential difference measurements to assess CFTR ion channel activity', *Methods Mol Biol*, vol. 741, pp. 69-86.

Roy, B, Friesen, WJ, Tomizawa, Y, Leszyk, JD, Zhuo, J, Johnson, B, Dakka, J, Trotta, CR, Xue, X, Mutyam, V, Keeling, KM, Mobley, JA, Rowe, SM, Bedwell, DM, Welch, EM & Jacobson, A 2016, 'Ataluren stimulates ribosomal selection of near-cognate tRNAs to promote nonsense suppression', *Proc Natl Acad Sci U S A*, vol. 113, no. 44, Nov 1, pp. 12508-12513.

Rufener, SC & Muhlemann, O 2013, 'eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells', *Nat Struct Mol Biol*, vol. 20, no. 6, Jun, pp. 710-717.

Sabatier, C, Plump, AS, Le, M, Brose, K, Tamada, A, Murakami, F, Lee, EY & Tessier-Lavigne, M 2004, 'The divergent Robo family protein rig-1/Robo3 is a negative regulator of slit responsiveness required for midline crossing by commissural axons', *Cell*, vol. 117, no. 2, Apr 16, pp. 157-169.

Sakabe, NJ & de Souza, SJ 2007, 'Sequence features responsible for intron retention in human', *BMC Genomics*, vol. 8, p. 59.

Sakaguchi, M, Watanabe, M, Kinoshita, R, Kaku, H, Ueki, H, Futami, J, Murata, H, Inoue, Y, Li, SA, Huang, P, Putranto, EW, Ruma, IM, Nasu, Y, Kumon, H & Huh, NH 2014, 'Dramatic increase in expression of a transgene by insertion of promoters downstream of the cargo gene', *Mol Biotechnol*, vol. 56, no. 7, pp. 621-630.

Saltzman, AL, Kim, YK, Pan, Q, Fagnani, MM, Maquat, LE & Blencowe, BJ 2008, 'Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay', *Mol Cell Biol*, vol. 28, no. 13, Jul, pp. 4320-4330.

Sambrook, J & Russel, DW 2001, *Molecular cloning: a laboratory manual.*, 3 edn, vol. 2, Cold Spring Harbour Laboratory Press, New York.

Sancar, GB, Cedeno, MM & Rieder, RF 1981, 'Rapid destruction of newly synthesized excess beta-globin chains in HbH disease', *Blood*, vol. 57, no. 5, May, pp. 967-971.

Sankaran, VG & Orkin, SH 2013, 'The switch from fetal to adult hemoglobin', *Cold Spring Harb Perspect Med*, vol. 3, no. 1, Jan, p. a011643.

Sato, H & Maquat, LE 2009, 'Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta', *Genes Dev*, vol. 23, no. 21, Nov 1, pp. 2537-2550.

Schneppenheim, R, Budde, U, Obser, T, Brassard, J, Mainusch, K, Ruggeri, ZM, Schneppenheim, S, Schwaab, R & Oldenburg, J 2001, 'Expression and characterization of von Willebrand factor dimerization defects in different types of von Willebrand disease', *Blood*, vol. 97, no. 7, Apr 1, pp. 2059-2066.

Schoenberg, DR & Maquat, LE 2012, 'Regulation of cytoplasmic mRNA decay', *Nature Reviews Genetics*, vol. 13, no. 4, p. 246.

Schuller, AP, Zinshteyn, B, Enam, SU & Green, R 2018, 'Directed hydroxyl radical probing reveals Upf1 binding to the 80S ribosomal E site rRNA at the L1 stalk', *Nucleic Acids Res*, vol. 46, no. 4, Feb 28, pp. 2060-2073.

Schwabe, GC, Tinschert, S, Buschow, C, Meinecke, P, Wolff, G, Gillessen-Kaesbach, G, Oldridge, M, Wilkie, AO, Komec, R & Mundlos, S 2000, 'Distinct mutations in the receptor tyrosine kinase gene ROR2 cause brachydactyly type B', *Am J Hum Genet*, vol. 67, no. 4, Oct, pp. 822-831.

Schwarz, JM, Cooper, DN, Schuelke, M & Seelow, D 2014, 'MutationTaster2: mutation prediction for the deep-sequencing age', *Nat Methods*, vol. 11, 03/28/online, p. 361.

Serin, G, Gersappe, A, Black, JD, Aronoff, R & Maquat, LE 2001, 'Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4)', *Mol Cell Biol*, vol. 21, no. 1, Jan, pp. 209-223.

Seyedali, A & Berry, MJ 2014, 'Nonsense-mediated decay factors are involved in the regulation of selenoprotein mRNA levels during selenium deficiency', *Rna*, vol. 20, no. 8, Aug, pp. 1248-1256.

Shaheen, R, Anazi, S, Ben-Omran, T, Seidahmed, MZ, Caddle, LB, Palmer, K, Ali, R, Alshidi, T, Hagos, S, Goodwin, L, Hashem, M, Wakil, SM, Abouelhoda, M, Colak, D, Murray, SA & Alkuraya, FS 2016, 'Mutations in SMG9, Encoding an Essential Component of Nonsense-Mediated Decay Machinery, Cause a Multiple Congenital Anomaly Syndrome in Humans and Mice', *Am J Hum Genet*, vol. 98, no. 4, Apr 7, pp. 643-652.

Shaner, NC, Steinbach, PA & Tsien, RY 2005, 'A guide to choosing fluorescent proteins', *Nat Methods*, vol. 2, no. 12, Dec, pp. 905-909.

Shibuya, T, Tange, TO, Sonenberg, N & Moore, MJ 2004, 'eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay', *Nat Struct Mol Biol*, vol. 11, no. 4, Apr, pp. 346-351.

Shum, EY, Jones, SH, Shao, A, Dumdie, J, Krause, MD, Chan, WK, Lou, CH, Espinoza, JL, Song, HW, Phan, MH, Ramaiah, M, Huang, L, McCarrey, JR, Peterson, KJ, De Rooij, DG, Cook-Andersen, H & Wilkinson, MF 2016, 'The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay', *Cell*, vol. 165, no. 2, Apr 7, pp. 382-395.

Siddiqui, N & Sonenberg, N 2016, 'Proposing a mechanism of action for ataluren', *Proc Natl Acad Sci U S A*, vol. 113, no. 44, Nov 1, pp. 12353-12355.

Sie, L, Loong, S & Tan, E 2009, 'Utility of lymphoblastoid cell lines', *Journal of neuroscience research*, vol. 87, no. 9, pp. 1953-1959.

Silver, DL, Watkins-Chow, DE, Schreck, KC, Pierfelice, TJ, Larson, DM, Burnetti, AJ, Liaw, HJ, Myung, K, Walsh, CA, Gaiano, N & Pavan, WJ 2010, 'The exon junction complex component Magoh controls brain size by regulating neural stem cell division', *Nat Neurosci*, vol. 13, no. 5, May, pp. 551-558.

Singh, G, Rebbapragada, I & Lykke-Andersen, J 2008, 'A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay', *PLoS Biol*, vol. 6, no. 4, Apr 29, p. e111.

Stalder, L & Muhlemann, O 2008, 'The meaning of nonsense', *Trends Cell Biol*, vol. 18, no. 7, Jul, pp. 315-321.

Sulem, P, Helgason, H, Oddson, A, Stefansson, H, Gudjonsson, SA, Zink, F, Hjartarson, E, Sigurdsson, GT, Jonasdottir, A, Jonasdottir, A, Sigurdsson, A, Magnusson, OT, Kong, A, Helgason, A, Holm, H, Thorsteinsdottir, U, Masson, G, Gudbjartsson, DF & Stefansson, K 2015, 'Identification of a large set of rare complete human knockouts', *Nat Genet*, vol. 47, no. 5, May, pp. 448-452.

Szymczak, AL, Workman, CJ, Wang, Y, Vignali, KM, Dilioglou, S, Vanin, EF & Vignali, DAA 2004, 'Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide–based retroviral vector', *Nature Biotechnology*, vol. 22, 04/04/online, p. 589.

Tan, JT, Kremer, F, Freddi, S, Bell, KM, Baker, NL, Lamandé, SR & Bateman, JF 2008, 'Competency for nonsense-mediated reduction in collagen X mRNA is specified by the 3' UTR and corresponds to the position of mutations in Schmid metaphyseal chondrodysplasia', *Am J Hum Genet*, vol. 82, no. 3, Mar 3, pp. 786-793.

Tarpey, PS, Raymond, FL, Nguyen, LS, Rodriguez, J, Hackett, A, Vandeleur, L, Smith, R, Shoubridge, C, Edkins, S, Stevens, C, O'Meara, S, Tofts, C, Barthorpe, S, Buck, G, Cole, J, Halliday, K, Hills, K, Jones, D, Mironenko, T, Perry, J, Varian, J, West, S, Widaa, S, Teague, J, Dicks, E, Butler, A, Menzies, A, Richardson, D, Jenkinson, A, Shepherd, R, Raine, K, Moon, J, Luo, Y, Parnau, J, Bhat, SS, Gardner, A, Corbett, M, Brooks, D, Thomas, P, Parkinson-Lawrence, E, Porteous, ME, Warner, JP, Sanderson, T, Pearson, P, Simensen, RJ, Skinner, C, Hoganson, G, Superneau, D, Wooster, R, Bobrow, M, Turner, G, Stevenson, RE, Schwartz, CE, Futreal, PA, Srivastava, AK, Stratton, MR & Gecz, J 2007, 'Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation', *Nat Genet*, vol. 39, no. 9, Sep, pp. 1127-1133.

Temple, GF, Dozy, AM, Roy, KL & Kan, YW 1982, 'Construction of a functional human suppressor tRNA gene: an approach to gene therapy for beta-thalassaemia', *Nature*, vol. 296, no. 5857, Apr 8, pp. 537-540.

Thada, V, Miller, JN, Kovács, AD & Pearce, DA 2016, 'Tissue-specific variation in nonsense mutant transcript level and drug-induced read-through efficiency in the Cln1(R151X) mouse model of INCL', *J Cell Mol Med*, vol. 20, no. 2, Feb, pp. 381-385.

Thapar, A, Cooper, M & Rutter, M 2017, 'Neurodevelopmental disorders', *Lancet Psychiatry*, vol. 4, no. 4, Apr, pp. 339-346.

The Genomes Project, C, McVean, GA, Altshuler, DM, Durbin, RM, Abecasis, GR, Bentley, DR, Chakravarti, A, Clark, AG, Donnelly, P, Eichler, EE, Flicek, P, Gabriel, SB, Gibbs, RA, Green, ED, Hurles, ME, Knoppers, BM, Korbel, JO, Lander, ES, Lee, C, Lehrach, H, Mardis, ER, Marth, GT, McVean, GA, Nickerson, DA, Schmidt, JP, Sherry, ST, Wang, J, Wilson, RK, Gibbs, RA, Dinh, H, Kovar, C, Lee, S, Lewis, L, Muzny, D, Reid, J, Wang, M, Wang, J, Fang, X, Guo, X, Jian, M, Jiang, H, Jin, X, Li, G, Li, J, Li, Y, Li, Z, Liu, X, Lu, Y, Ma, X, Su, Z, Tai, S, Tang, M, Wang, B, Wang, G, Wu, H, Wu, R, Yin, Y, Zhang, W, Zhao, J, Zhao, M, Zheng, X, Zhou, Y, Lander, ES, Altshuler, DM, Gabriel, SB, Gupta, N, Flicek, P, Clarke, L, Leinonen, R, Smith, RE, Zheng-Bradley, X, Bentley, DR, Grocock, R, Humphray, S, James, T, Kingsbury, Z, Lehrach, H, Sudbrak, R, Albrecht, MW, Amstislavskiy, VS, Borodina, TA, Lienhard, M, Mertes, F, Sultan, M, Timmermann, B, Yaspo, M-L, Sherry, ST, McVean, GA, Mardis, ER, Wilson, RK, Fulton, L, Fulton, R, Weinstock, GM, Durbin, RM, Balasubramaniam, S, Burton, J, Danecek, P, Keane, TM, Kolb-Kokocinski, A, McCarthy, S, Stalker, J, Quail, M, Schmidt, JP, Davies, CJ, Gollub, J, Webster, T, Wong, B, Zhan, Y, Auton, A, Gibbs, RA, Yu, F, Bainbridge, M, Challis, D, Evani, US, Lu, J, Muzny, D, Nagaswamy, U, Reid, J, Sabo, A, Wang, Y, Yu, J, Wang, J, Coin, LJM, Fang, L, Guo, X, Jin, X, Li, G, Li, Q, Li, Y, Li, Z, Lin, H, Liu, B, Luo, R, Qin, N, Shao, H, Wang, B, Xie, Y, Ye, C, Yu, C, Zhang, F, Zheng, H, Zhu, H, Marth, GT, Garrison, EP, Kural, D, Lee, W-P, Fung Leong, W, Ward, AN, Wu, J, Zhang, M, Lee, C, Griffin, L, Hsieh, C-H, Mills, RE, Shi, X, von Grotthuss, M, Zhang, C, Daly, MJ, DePristo, MA, Altshuler, DM, Banks, E, Bhatia, G, Carneiro, MO, del Angel, G, Gabriel, SB, Genovese, G, Gupta, N, Handsaker, RE, Hartl, C, Lander, ES, McCarroll, SA, Nemesh, JC, Poplin, RE, Schaffner, SF, Shakir, K, Yoon, SC, Lihm, J, Makarov, V, Jin, H, Kim, W, Cheol Kim, K, Korbel, JO, Rausch, T, Flicek, P, Beal, K, Clarke, L, Cunningham, F, Herrero, J, McLaren, WM, Ritchie, GRS, Smith, RE, Zheng-Bradley, X, Clark, AG, Gottipati, S, Keinan, A, Rodriguez-Flores, JL, Sabeti, PC, Grossman, SR, Tabrizi, S, Tariyal, R, Cooper, DN, Ball, EV, Stenson, PD, Bentley, DR, Barnes, B, Bauer, M, Keira Cheetham, R, Cox, T, Eberle, M, Humphray, S, Kahn, S, Murray, L, Peden, J, Shaw, R, Ye, K, Batzer, MA, Konkel, MK, Walker, JA, MacArthur, DG, Lek, M, Sudbrak, Amstislavskiy, VS, Herwig, R, Shriver, MD, Bustamante, CD, Byrnes, JK, De La Vega, FM, Gravel, S, Kenny, EE, Kidd, JM, Lacroute, P, Maples, BK, Moreno-Estrada, A, Zakharia, F, Halperin, E, Baran, Y, Craig, DW, Christoforides, A, Homer, N, Izatt, T, Kurdoglu, AA, Sinari, SA, Squire, K, Sherry, ST, Xiao, C, Sebat, J, Bafna, V, Ye, K, Burchard, EG, Hernandez, RD, Gignoux, CR, Haussler, D, Katzman, SJ, James Kent, W, Howie, B, Ruiz-Linares, A, Dermitzakis, ET, Lappalainen, T, Devine, SE, Liu, X, Maroo, A, Tallon, LJ, Rosenfeld, JA, Michelson, LP, Abecasis, GR, Min Kang, H, Anderson, P, Angius, A, Bigham, A, Blackwell, T, Busonero, F, Cucca, F, Fuchsberger, C, Jones, C, Jun, G, Li, Y, Lyons, R, Maschio, A, Porcu, E, Reinier, F, Sanna, S, Schlessinger, D, Sidore, C, Tan, A, Kate Trost, M, Awadalla, P, Hodgkinson, A, Lunter, G, McVean, GA, Marchini, JL, Myers, S, Churchhouse, C, Delaneau, O, Gupta-Hinch, A, Iqbal, Z, Mathieson, I, Rimmer, A, Xifara, DK, Oleksyk, TK, Fu, Y, Liu, X, Xiong, M, Jorde, L, Witherspoon, D, Xing, J, Eichler, EE, Browning, BL, Alkan, C, Hajirasouliha, I, Hormozdiari, F, Ko, A, Sudmant, PH, Mardis, ER, Chen, K, Chinwalla, A, Ding, L, Dooling, D, Koboldt, DC, McLellan, MD, Wallis, JW, Wendl, MC, Zhang, Q, Durbin, RM, Hurles, ME, Tyler-Smith, C, Albers, CA, Ayub, Q, Balasubramaniam, S, Chen, Y, Coffey, AJ, Colonna, V, Danecek, P, Huang, N, Jostins, L, Keane, TM, Li, H, McCarthy, S, Scally, A, Stalker, J, Walter, K, Xue, Y, Zhang, Y, Gerstein, MB, Abyzov, A, Balasubramanian, S, Chen, J, Clarke, D, Fu, Y, Habegger, L, Harmanci, AO, Jin, M, Khurana, E, Jasmine Mu, X, Sisu, C, Li, Y, Luo, R, Zhu, H, Lee, C, Griffin, L, Hsieh, C-H, Mills, RE, Shi, X, von Grotthuss, M, Zhang, C, Marth, GT, Garrison, EP, Kural, D, Lee, W-P, Ward, AN, Wu, J, Zhang, M, McCarroll, SA, Altshuler, DM, Banks, E, del Angel, G, Genovese, G, Handsaker, RE, Hartl, C, Nemesh, JC, Shakir, K, Yoon, SC, Lihm, J, Makarov, V, Degenhardt, J, Flicek, P, Clarke, L, Smith, RE, Zheng-Bradley, X, Korbel, JO, Rausch, T, Stütz, AM, Bentley, DR, Barnes, B, Keira Cheetham, R, Eberle, M, Humphray, S, Kahn, S, Murray, L, Shaw, R, Ye, K, Batzer, MA, Konkel, MK, Walker, JA, Lacroute, P, Craig, DW, Homer, N, Church, D, Xiao, C, Sebat, J, Bafna, V, Michaelson, JJ, Ye,

K, Devine, SE, Liu, X, Maroo, A, Tallon, LJ, Lunter, G, McVean, GA, Iqbal, Z, Witherspoon, D, Xing, J, Eichler, EE, Alkan, C, Hajirasouliha, I, Hormozdiari, F, Ko, A, Sudmant, PH, Chen, K, Chinwalla, A, Ding, L, McLellan, MD, Wallis, JW, Hurles, ME, Blackburne, B, Li, H, Lindsay, SJ, Ning, Z, Scally, A, Walter, K, Zhang, Y, Gerstein, MB, Abyzov, A, Chen, J, Clarke, D, Khurana, E, Jasmine Mu, X, Sisu, C, Gibbs, RA, Yu, F, Bainbridge, M, Challis, D, Evani, US, Kovar, C, Lewis, L, Lu, J, Muzny, D, Nagaswamy, U, Reid, J, Sabo, A, Yu, J, Guo, X, Li, Y, Wu, R, Marth, GT, Garrison, EP, Fung Leong, W, Ward, AN, del Angel, G, DePristo, MA, Gabriel, SB, Gupta, N, Hartl, C, Poplin, RE, Clark, AG, Rodriguez-Flores, JL, Flicek, P, Clarke, L, Smith, RE, Zheng-Bradley, X, MacArthur, DG, Bustamante, CD, Gravel, S, Craig, DW, Christoforides, A, Homer, N, Izatt, T, Sherry, ST, Xiao, C, Dermitzakis, ET, Abecasis, GR, Min Kang, H, McVean, GA, Mardis, ER, Dooling, D, Fulton, L, Fulton, R, Koboldt, DC, Durbin, RM, Balasubramaniam, S, Keane, TM, McCarthy, S, Stalker, J, Gerstein, MB, Balasubramanian, S, Habegger, L, Garrison, EP, Gibbs, RA, Bainbridge, M, Muzny, D, Yu, F, Yu, J, del Angel, G, Handsaker, RE, Makarov, V, Rodriguez-Flores, JL, Jin, H, Kim, W, Cheol Kim, K, Flicek, P, Beal, K, Clarke, L, Cunningham, F, Herrero, J, McLaren, WM, Ritchie, GRS, Zheng-Bradley, X, Tabrizi, S, MacArthur, DG, Lek, M, Bustamante, CD, De La Vega, FM, Craig, DW, Kurdoglu, AA, Lappalainen, T, Rosenfeld, JA, Michelson, LP, Awadalla, P, Hodgkinson, A, McVean, GA, Chen, K, Tyler-Smith, C, Chen, Y, Colonna, V, Frankish, A, Harrow, J, Xue, Y, Gerstein, MB, Abyzov, A, Balasubramanian, S, Chen, J, Clarke, D, Fu, Y, Harmanci, AO, Jin, M, Khurana, E, Jasmine Mu, X, Sisu, C, Gibbs, RA, Fowler, G, Hale, W, Kalra, D, Kovar, C, Muzny, D, Reid, J, Wang, J, Guo, X, Li, G, Li, Y, Zheng, X, Altshuler, DM, Flicek, P, Clarke, L, Barker, J, Kelman, G, Kulesha, E, Leinonen, R, McLaren, WM, Radhakrishnan, R, Roa, A, Smirnov, D, Smith, RE, Streeter, I, Toneva, I, Vaughan, B, Zheng-Bradley, X, Bentley, DR, Cox, T, Humphray, S, Kahn, S, Sudbrak, R, Albrecht, MW, Lienhard, M, Craig, DW, Izatt, T, Kurdoglu, AA, Sherry, ST, Ananiev, V, Belaia, Z, Beloslyudtsev, D, Bouk, N, Chen, C, Church, D, Cohen, R, Cook, C, Garner, J, Hefferon, T, Kimelman, M, Liu, C, Lopez, J, Meric, P, O'Sullivan, C, Ostapchuk, Y, Phan, L, Ponomarov, S, Schneider, V, Shekhtman, E, Sirotkin, K, Slotta, D, Xiao, C, Zhang, H, Haussler, D, Abecasis, GR, McVean, GA, Alkan, C, Ko, A, Dooling, D, Durbin, RM, Balasubramaniam, S, Keane, TM, McCarthy, S, Stalker, J, Chakravarti, A, Knoppers, BM, Abecasis, GR, Barnes, KC, Beiswanger, C, Burchard, EG, Bustamante, CD, Cai, H, Cao, H, Durbin, RM, Gharani, N, Gibbs, RA, Gignoux, CR, Gravel, S, Henn, B, Jones, D, Jorde, L, Kaye, JS, Keinan, A, Kent, A, Kerasidou, A, Li, Y, Mathias, R, McVean, GA, Moreno-Estrada, A, Ossorio, PN, Parker, M, Reich, D, Rotimi, CN, Royal, CD, Sandoval, K, Su, Y, Sudbrak, R, Tian, Z, Timmermann, B, Tishkoff, S, Toji, LH, Tyler-Smith, C, Via, M, Wang, Y, Yang, H, Yang, L, Zhu, J, Bodmer, W, Bedoya, G, Ruiz-Linares, A, Zhi Ming, C, Yang, G, Jia You, C, Peltonen, L, Garcia-Montero, A, Orfao, A, Dutil, J, Martinez-Cruzado, JC, Oleksyk, TK, Brooks, LD, Felsenfeld, AL, McEwen, JE, Clemm, NC, Duncanson, A, Dunn, M, Green, ED, Guyer, MS, Peterson, JL, Abecasis, GR, Auton, A, Brooks, LD, DePristo, MA, Durbin, RM, Handsaker, RE, Min Kang, H, Marth, GT & McVean, GA 2012, 'An integrated map of genetic variation from 1,092 human genomes', *Nature*, vol. 491, 10/31/online, p. 56.

Thein, SL, Hesketh, C, Taylor, P, Temperley, IJ, Hutchinson, RM, Old, JM, Wood, WG, Clegg, JB & Weatherall, DJ 1990, 'Molecular basis for dominantly inherited inclusion body beta-thalassemia', *Proc Natl Acad Sci U S A*, vol. 87, no. 10, May, pp. 3924-3928.

Thermann, R, Neu-Yilik, G, Deters, A, Frede, U, Wehr, K, Hagemeier, C, Hentze, MW & Kulozik, AE 1998, 'Binary specification of nonsense codons by splicing and cytoplasmic translation', *Embo j*, vol. 17, no. 12, Jun 15, pp. 3484-3494.

Trivellin, G, Sharwood, E, Hijazi, H, Carvalho, CMB, Yuan, B, Tatton-Brown, K, Coman, D, Lupski, JR, Cotterill, AM, Lodish, MB & Stratakis, CA 2018, 'Xq26.3 Duplication in a Boy

With Motor Delay and Low Muscle Tone Refines the X-Linked Acrogigantism Genetic Locus', *J Endocr Soc*, vol. 2, no. 10, Oct 1, pp. 1100-1108.

Tsai, CJ, Sauna, ZE, Kimchi-Sarfaty, C, Ambudkar, SV, Gottesman, MM & Nussinov, R 2008, 'Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima', *J Mol Biol*, vol. 383, no. 2, Nov 7, pp. 281-291.

Unterholzner, L & Izaurralde, E 2004, 'SMG7 acts as a molecular link between mRNA surveillance and mRNA decay', *Mol Cell*, vol. 16, no. 4, Nov 19, pp. 587-596.

Usuki, F, Fujimura, M & Yamashita, A 2013, 'Endoplasmic reticulum stress preconditioning attenuates methylmercury-induced cellular damage by inducing favorable stress responses', *Sci Rep*, vol. 3, p. 2346.

Usuki, F, Yamashita, A, Higuchi, I, Ohnishi, T, Shiraishi, T, Osame, M & Ohno, S 2004, 'Inhibition of nonsense-mediated mRNA decay rescues the phenotype in Ullrich's disease', *Ann Neurol*, vol. 55, no. 5, May, pp. 740-744.

Usuki, F, Yamashita, A, Kashima, I, Higuchi, I, Osame, M & Ohno, S 2006, 'Specific inhibition of nonsense-mediated mRNA decay components, SMG-1 or Upf1, rescues the phenotype of Ullrich disease fibroblasts', *Mol Ther*, vol. 14, no. 3, Sep, pp. 351-360.

Viegas, MH, Gehring, NH, Breit, S, Hentze, MW & Kulozik, AE 2007, 'The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway', *Nucleic Acids Res*, vol. 35, no. 13, pp. 4542-4551.

Wagner, KR, Hamed, S, Hadley, DW, Gropman, AL, Burstein, AH, Escolar, DM, Hoffman, EP & Fischbeck, KH 2001, 'Gentamicin treatment of Duchenne and Becker muscular dystrophy due to nonsense mutations', *Ann Neurol*, vol. 49, no. 6, Jun, pp. 706-711.

Walter, P & Ron, D 2011, 'The unfolded protein response: from stress pathway to homeostatic regulation', *Science*, vol. 334, no. 6059, Nov 25, pp. 1081-1086.

Wang, D, Zavadil, J, Martin, L, Parisi, F, Friedman, E, Levy, D, Harding, H, Ron, D & Gardner, LB 2011, 'Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis', *Mol Cell Biol*, vol. 31, no. 17, Sep, pp. 3670-3680.

Wang, ET, Sandberg, R, Luo, S, Khrebtukova, I, Zhang, L, Mayr, C, Kingsmore, SF, Schroth, GP & Burge, CB 2008, 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, vol. 456, no. 7221, Nov 27, pp. 470-476.

Wang, J, Gudikote, JP, Olivas, OR & Wilkinson, MF 2002, 'Boundary-independent polar nonsense-mediated decay', *EMBO Rep*, vol. 3, no. 3, Mar, pp. 274-279.

Wang, L & Chen, YG 2016, 'Signaling Control of Differentiation of Embryonic Stem Cells toward Mesendoderm', *J Mol Biol*, vol. 428, no. 7, Apr 10, pp. 1409-1422.

Wang, M & Marin, A 2006, 'Characterization and prediction of alternative splice sites', *Gene*, vol. 366, no. 2, Feb 1, pp. 219-227.

Wang, X, Codreanu, SG, Wen, B, Li, K, Chambers, MC, Liebler, DC & Zhang, B 2018, 'Detection of Proteome Diversity Resulted from Alternative Splicing is Limited by Trypsin Cleavage Specificity', *Molecular & Cellular Proteomics*, vol. 17, no. 3, p. 422.

Wang, Y, Qiu, C & Cui, Q 2015, 'A Large-Scale Analysis of the Relationship of Synonymous SNPs Changing MicroRNA Regulation with Functionality and Disease', *Int J Mol Sci*, vol. 16, no. 10, Sep 30, pp. 23545-23555.

Ware, MD, DeSilva, D, Sinilnikova, OM, Stoppa-Lyonnet, D, Tavtigian, SV & Mazoyer, S 2006, 'Does nonsense-mediated mRNA decay explain the ovarian cancer cluster region of the BRCA2 gene?', *Oncogene*, vol. 25, no. 2, Jan 12, pp. 323-328.

Weischenfeldt, J, Waage, J, Tian, G, Zhao, J, Damgaard, I, Jakobsen, JS, Kristiansen, K, Krogh, A, Wang, J & Porse, BT 2012, 'Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns', *Genome Biol*, vol. 13, no. 5, May 24, p. R35.

Welch, EM, Barton, ER, Zhuo, J, Tomizawa, Y, Friesen, WJ, Trifillis, P, Paushkin, S, Patel, M, Trotta, CR, Hwang, S, Wilde, RG, Karp, G, Takasugi, J, Chen, G, Jones, S, Ren, H, Moon, YC, Corson, D, Turpoff, AA, Campbell, JA, Conn, MM, Khan, A, Almstead, NG, Hedrick, J, Mollin, A, Risher, N, Weetall, M, Yeh, S, Branstrom, AA, Colacino, JM, Babiak, J, Ju, WD, Hirawat, S, Northcutt, VJ, Miller, LL, Spatrick, P, He, F, Kawana, M, Feng, H, Jacobson, A, Peltz, SW & Sweeney, HL 2007, 'PTC124 targets genetic disorders caused by nonsense mutations', *Nature*, vol. 447, no. 7140, May 3, pp. 87-91.

Wells, SE, Hillner, PE, Vale, RD & Sachs, AB 1998, 'Circularization of mRNA by eukaryotic translation initiation factors', *Mol Cell*, vol. 2, no. 1, Jul, pp. 135-140.

Wen, J & Brogna, S 2010, 'Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe', *Embo j*, vol. 29, no. 9, May 5, pp. 1537-1551.

Wheeler, HE & Dolan, ME 2012, 'Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation', *Pharmacogenomics*, vol. 13, no. 1, Jan, pp. 55-70.

Wilber, A, Nienhuis, AW & Persons, DA 2011, 'Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities', *Blood*, vol. 117, no. 15, Apr 14, pp. 3945-3953.

Wilschanski, M, Famini, C, Blau, H, Rivlin, J, Augarten, A, Avital, A, Kerem, B & Kerem, E 2000, 'A pilot study of the effect of gentamicin on nasal potential difference measurements in cystic fibrosis patients carrying stop mutations', *Am J Respir Crit Care Med*, vol. 161, no. 3 Pt 1, Mar, pp. 860-865.

Wilschanski, M, Yahav, Y, Yaacov, Y, Blau, H, Bentur, L, Rivlin, J, Aviram, M, Bdolah-Abram, T, Bebok, Z, Shushi, L, Kerem, B & Kerem, E 2003, 'Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations', *N Engl J Med*, vol. 349, no. 15, Oct 9, pp. 1433-1441.

Wittkopp, N, Huntzinger, E, Weiler, C, Saulière, J, Schmidt, S, Sonawane, M & Izaurralde, E 2009, 'Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival', *Mol Cell Biol*, vol. 29, no. 13, Jul, pp. 3517-3528.

Wollerton, MC, Gooding, C, Wagner, EJ, Garcia-Blanco, MA & Smith, CW 2004, 'Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay', *Mol Cell*, vol. 13, no. 1, Jan 16, pp. 91-100.

Wong, JJ, Ritchie, W, Ebner, OA, Selbach, M, Wong, JW, Huang, Y, Gao, D, Pinello, N, Gonzalez, M, Baidya, K, Thoeng, A, Khoo, TL, Bailey, CG, Holst, J & Rasko, JE 2013, 'Orchestrated intron retention regulates normal granulocyte differentiation', *Cell*, vol. 154, no. 3, Aug 1, pp. 583-595.

Xu, X, Zhang, L, Tong, P, Xun, G, Su, W, Xiong, Z, Zhu, T, Zheng, Y, Luo, S, Pan, Y, Xia, K & Hu, Z 2013, 'Exome sequencing identifies UPF3B as the causative gene for a Chinese non-syndrome mental retardation pedigree', *Clin Genet*, vol. 83, no. 6, Jun, pp. 560-564.

Yamashita, A, Izumi, N, Kashima, I, Ohnishi, T, Saari, B, Katsuhata, Y, Muramatsu, R, Morita, T, Iwamatsu, A, Hachiya, T, Kurata, R, Hirano, H, Anderson, P & Ohno, S 2009, '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 Dev*, vol. 23, no. 9, May 1, pp. 1091-1105.

Yamashita, A, Ohnishi, T, Kashima, I, Taya, Y & Ohno, S 2001, 'Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay', *Genes Dev*, vol. 15, no. 17, Sep 1, pp. 2215-2228.

Yan, Q, Weyn-Vanhentenryck, SM, Wu, J, Sloan, SA, Zhang, Y, Chen, K, Wu, JQ, Barres, BA & Zhang, C 2015, 'Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators', *Proc Natl Acad Sci U S A*, vol. 112, no. 11, Mar 17, pp. 3445-3450.

Yepiskoposyan, H, Aeschimann, F, Nilsson, D, Okoniewski, M & Muhlemann, O 2011, 'Autoregulation of the nonsense-mediated mRNA decay pathway in human cells', *Rna*, vol. 17, no. 12, Dec, pp. 2108-2118.

Zetoune, AB, Fontanière, S, Magnin, D, Anczuków, O, Buisson, M, Zhang, CX & Mazoyer, S 2008, 'Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues', *BMC Genet*, vol. 9, p. 83.

Zhan, L-L 2013, 'Recent advances of studies on alternative intron retention', *Trends in Evolutionary Biology*, vol. 5, no. 1, p. 1.

Zhang, J, Sun, X, Qian, Y & Maquat, LE 1998, 'Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm', *Rna*, vol. 4, no. 7, Jul, pp. 801-815.

Zheng, S 2016, 'Alternative splicing and nonsense-mediated mRNA decay enforce neural specific gene expression', *Int J Dev Neurosci*, vol. 55, Dec, pp. 102-108.

Zhong, S, Rivera-Molina, F, Rivetta, A, Toomre, D, Santos-Sacchi, J & Navaratnam, D 2019, 'Seeing the long tail: A novel green fluorescent protein, SiriusGFP, for ultra long timelapse imaging', *J Neurosci Methods*, vol. 313, Feb 1, pp. 68-76.

Zhou, D, Liu, K, Sun, C-W, Pawlik, KM & Townes, TM 2010, 'KLF1 regulates BCL11A expression and γ- to β-globin gene switching', *Nat Genet*, vol. 42, 08/01/online, p. 742.

Zsembery, A, Jessner, W, Sitter, G, Spirli, C, Strazzabosco, M & Graf, J 2002, 'Correction of CFTR malfunction and stimulation of Ca-activated Cl channels restore HCO3- secretion in cystic fibrosis bile ductular cells', *Hepatology*, vol. 35, no. 1, Jan, pp. 95-104.