THE UNIVERSITY
of EDINBURGH

Genome-Wide and Subcellular Nonsense Mediated Decay

Kathryn A Jackson-Jones

Thesis submitted in fulfilment

of the requirements for the degree of

Doctor of Philosophy

University of Edinburgh

2021

Declaration of authorship

This thesis is the result of my own original research unless otherwise stated and has been written by me. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma or other qualification. Work done in collaboration with, or with assistance of, others, is indicated as such. Inclusion of published figures is under Creative Commons License or specific license agreement with the publisher.

Signed:

Kathryn A. Jackson-Jones

Date: 3rd May 2021

Acknowledgements

Someone once told me that no one wants to read a long acknowledgements section and I should keep it as short as possible.
Someone much wiser once told me that the acknowledgements is the only part of your thesis anyone will read, so make it good!
The list of people to whom I owe enormous gratitude is lengthy. This PhD has undoubtably been the hardest thing I have ever done, and I certainly hope that I will ever do. Whilst I worked incredibly hard and have given it everything, it simply would not have been possible without so much love, support and help.

Firstly, to my parents Bryan and Lynda Jones, who from the beginning told me I could be whatever I wanted to be and have always been there to support me no matter what. Thank you for believing in me and encouraging me to believe in myself.
To Ian Bellamy who first planted the idea of a career in science and Steve Martin who first taught me about genetics with pictures of his (very blonde) children on holiday! Thank you for inspiring my love of science and encouraging me to believe that I could succeed.

To Javier Cáceres and Dáša Longman, thank you for letting me take your project and put my own stamp on it. The Cáceres lab has prepared me well for anything science can throw at me. I am grateful to have been trained by such independent, talented and inspiring women. Thank you Dáša for sharing your never ending enthusiasm for research with me and all your top tips like counting cells by eye(!) Magda thank you for being a great lab mate, and for the help and advice with

setting up the 4sU experiments. Grazie mille Nele for all the NMD knowledge and plethora of technical advice, and for the endless stream of new papers I need to read! Fiona, for my owl of wisdom which was certainly needed and for being a shoulder to cry on. Abby for all the technical support and emotional support too! Ross, thank you for inducting me into the Cáceres lab and for all the lab gadgets that made my life easier. The Germans Michaela, Stephan and Ramon, thanks for being great temporary additions to the lab and allowing me to not be the most junior lab member for a time, and especially Michaela for the CRISPR advice.

I am hugely grateful to the Taylor lab, for being the most nurturing and encouraging environment to do science. Martin Taylor for being an incredibly supportive supervisor but also a passionate and engaging scientist who always made me question bigger. Thank you for taking a chance on a random new student who suggested you co-supervise her without any prior experience or even meeting her before! Enormous thanks to Robert Young for a phenomenal amount of help and support. For teaching everything google couldn't about bioinformatics, for always asking the awkward questions and for being a braw pal. You are the Bella to my Belinda. Thanks for talking me out of quitting. Over and over. Juliet for the endless R help and general office sass. Lana for all the LaTeX help and laughs. Craig for all the pep talks and career advice, for the book recommendations and the cat pictures, and for being the most upbeat person I have ever met. Thomas for the best CRISPR advice, for teaching me how to nanopore and for letting me crash your family Christmas!

Thank you to all the amazing technical services team for always putting up with my last minute requests, incomplete forms and bad time-keeping. Especially to Lizzy and Stacey for never chiding me however late I was with my FACS samples. I'm grateful to everyone in the imaging team for fixing all my microscopy issues and especially Laura for explaining image analysis, multiple times. Thanks to everyone in Evogen for being a fantastic environment to dip my toe into bioin-

formatics and have a load of fun (and Colin cakes!) along the way. Especially to Graeme for getting me out of more than one R pickle. A special mention for JJarvis whose excellent tutoring first got me hooked on coding. Thank you to my thesis committee Ian and Nicky for your scientific input. Thank you to Martin Reijns for you excellent scientific and personal advice. Thanks to the Bickmore lab for adopting me into the gang and letting me crash many social events, so much so that it was often suggested I start attending your lab meetings! Special thanks to Shelagh and Rob for lots of support both personal and scientific. Thank you to everyone in Genome Regulation section for being a great science family and always being ready to say yes to a tea break!

My co-years for being such a fantastic, supportive bunch. I don't think I could have survived these 4 years without our many lunchtime therapy sessions and my vocabulary is definitely much improved thanks to our daily crosswords; so proud I can nearly do a whole cryptic on my own now! Especially to Lauren for always having time to decompress and talk things through. Also, for showing me how to do IF "the best way". Ilya, my original desk mate and long time encourager of coding. I get a lot more work done on the "dark side" although I miss our random chats and Russian lessons. Toby my fellow Welshie, diolch yn fawr for all the bioinformatics top tips and general encouragement! David, thank you for your healthy doses of realism and of course excellent advice. Ioannis, thanks for all the pep talks and general cheer. Your smile got me through many a hard day. To everyone on POGS committee, social committee and health and wellbeing committee; thank you for allowing me to achieve success when science wasn't playing ball and for being a fantastic extended support network! Thanks especially to Dee for being such an amazing cheerleader and sounding board. I'm also extremely grateful to all my friends outside science for being there for me throughout this rollercoaster and listening to my endless science chat!

The final and biggest thanks of all must go to my husband Matthew who has supported me endlessly and to whom I am grateful more than words can ever express.

"Dyfal donc a dyr y garreg"

Welsh proverb

"Nolite te Bastardes Carborundorum"

Margaret Atwood (1985)

Abstract

Nonsense mediated decay (NMD) is a translation-dependent RNA quality control pathway that selectively degrades aberrant transcripts before they lead to truncated proteins. NMD causes rapid degradation of target RNAs by assembling multiple factors to drive poly-deadenylation and decapping of the mRNA.

NBAS (neuroblastoma amplified sequence) protein was identified to be required for NMD in an RNAi screen in *Caenorhabditis elegans*. Subsequent evidence has suggested it is a conserved NMD factor in zebrafish and human cells. NBAS is localised to the endoplasmic reticulum (ER) suggesting it could be involved in a novel branch of the NMD pathway specific to ER. The key NMD factor UPF1 has been shown to partially localise to the ER and interact with NBAS.

NMD targets harbour premature termination codons (PTC)s, upstream open reading frames or long 3' UTRs. However not all transcripts with these features activate NMD, and many NMD targets appear not to contain any of these features. Therefore, the previously defined rules regarding initiation of the NMD pathway do not account for all variation seen in NMD efficiency.

This thesis presents the creation and investigation of a CRISPR cell line containing an endogenously tagged NBAS protein. Immunoprecipitation mass spectrometry on these cells revealed that NBAS interacts with the ER translocon, where the ribosome docks for translation at the ER membrane, suggesting the presence of an NBAS-UPF1-translocon complex. This interaction was confirmed by proximity ligation assay, a microscopy-based approach.

In order to identify the effect of NBAS on transcript expression and stability, RNA-sequencing was performed on total, nascent, cytoplasmic, and membrane-associated fractions, upon siRNA depletion of either NBAS or UPF1 mRNA. Differential expression analysis revealed a correlation in differentially expressed transcripts when NBAS or UPF1 were depleted with an enrichment in ER-associated RNAs specifically in NBAS targets. UPF1 was found to affect expression of a similar number of transcripts in the cytoplasmic and membrane-associated fractions confirming its role in NMD at the ER. Whereas depletion of NBAS led to a 5-fold higher proportion of membrane-associated genes being increased in expression than of non-membrane genes, showing its subcellular specificity.

Using both nascent and total RNA-sequencing datasets, estimates of differential transcript stability upon depletion of either factor were made. Approximately 20% of transcripts increased in expression upon depletion of UPF1 were also found to increase in stability, whereas the remaining 80% of expression changes could be explained by changes in transcription, suggesting these were due to secondary effects. Of those transcripts that changed in expression when NBAS was depleted, none were identified as changed in stability in this study, suggesting NBAS has subtle or negligible effect on transcript stability and is therefore likely to be a peripheral factor in the NMD pathway.

Finally, whole genome sequencing was used in conjunction with total and nascent expression datasets to identify correlation between genome variants and NMD efficiency. I show that it is possible to identify heterozygous variants with allele-specific processing and measure the effect of NMD to identify direct NMD activity. This approach can be used in the future to systematically explore the rules of NMD targeting in human cells.

In sum, this thesis reveals that NBAS interacts with the ER translocon and affects the expression of hundreds of ER-associated transcripts. However, NBAS does not appear to regulate transcript stability suggesting its role at the translocon is distinct from NMD. My findings add to the understanding of NMD quality control at the ER and the genome-wide rules governing NMD efficiency in human cells.

Lay abstract

Deoxyribonucleic acid (DNA) is a large molecule that carries the genetic instructions for all known living organisms. It functions like a cookbook, full of many different recipes. Each recipe is called a gene and gives instructions to produce one of the many different building blocks the cell needs, called proteins. DNA resides within the nucleus of the cell, but the proteins are produced outside of the nucleus in the cytoplasm. Therefore, individual genes are copied, and these copies are taken out of the nucleus to be used in the cytoplasm. The copies are made of ribonucleic acid (RNA) which contains triplets of nucleotide bases, called codons. Each codon is a code for a certain amino acid. Ribosomes read the codons in the RNA and collect the corresponding amino acids. The amino acids are joined together in a chain to create a protein. At the end of the RNA is a stop codon which tells the ribosome that the instructions are finished.

Sometimes when the DNA is copied into RNA, mistakes are made, and a stop codon is present before all the instructions are complete. If the ribosome used this RNA as an instruction, then the resulting protein would be too short and faulty. Therefore, the cell has a quality control mechanism in place to spot RNAs with premature stop codons and destroy them, called the nonsense-mediated decay (NMD) pathway. Although much research has been done on this pathway, it is not yet fully understood.

Recently, a new factor, called NBAS, was identified as essential for the pathway to function, however its exact role is unknown. NBAS is found in a different region of the cell to where NMD normally takes place, called the endoplasmic reticulum (ER). Some RNAs are made into protein at the ER by the ribosome

and it is unknown how these RNAs are quality checked. Furthermore, the key NMD factor UPF1 has also been shown to partially localise to the ER and interact with NBAS. These data suggest NBAS and UPF1 act as part of a separate branch of the NMD pathway, specific to the ER.

This thesis includes the genome editing of human cells to make NBAS fluorescent so it can be visualised. Investigation using these cells showed that NBAS interacts with the ER translocon, a special part of the ER where the translating ribosome attaches, suggesting the presence of an NBAS-UPF1-translocon complex. This interaction was visualised using fluorescent microscopy.

To measure how NBAS affects RNA processing, the level of all RNAs was quantified upon depletion of either NBAS or UPF1. Analysis revealed a correlation in the changed level of RNAs when either NBAS or UPF1 were depleted. Also, there was an enrichment in ER-associated RNA when NBAS was depleted. Next the level of RNA found in the cytoplasm or ER were measured. UPF1 was found to affect a similar number of RNAs in the cytoplasmic and ER fractions confirming its role in NMD at the ER as well as in the cytoplasm. Depletion of NBAS led to a 5-fold higher proportion of ER RNAs being increased in level than of cytoplasmic RNA, showing its specificity for the ER.

Using the level of newly made RNA in conjunction with total RNA levels, estimates of differential RNA stability upon depletion of either factor were made. This is because the NMD pathway makes RNA less stable. Approximately 5% of RNAs increased in level upon depletion of UPF1 were also found to increase in stability, suggesting the remaining changes could be explained by changes in transcription rate. This would mean these RNAs are not changed due to NMD. Of those RNAs that changed when NBAS was depleted, none were identified as changed in stability in this study, which seems to be contradictory. More research is needed to understand this phenomenon.

In conclusion, this thesis reveals that NBAS interacts with the ER translocon and affects the level of hundreds of ER-associated RNAs. However, this work did not show evidence that NBAS regulates transcript stability, suggesting further work is needed to understand the role of NBAS in the NMD pathway.

Contents

Abstract **ix**

Lay abstract **xi**

Abbreviations **xix**

Protein aliases **xxiii**

List of figures **xxiv**

List of tables **xxix**

1 Introduction **1**

1.1 Regulation of gene expression 3

1.1.1 Expression of genetic information 3

1.1.2 Processing of messenger RNA (mRNA) 9

1.1.3 Translation of mRNAs 12

1.1.4 Translation at the endoplasmic reticulum 17

1.2 RNA quality control 21

1.2.1 Surveillance mechanisms regulating RNA quality 21

1.2.2 Mechanism of nonsense mediated decay (NMD) 24

1.2.3 NMD in human genetic disease 29

1.2.4 Global regulation of the transcriptome by NMD 31

1.2.5 Target characteristics 33

1.2.6 Location specific NMD 38

1.3 Neuroblastoma amplified sequence (NBAS) protein 41

1.3.1 Discovery of NBAS as a putative NMD factor 41
1.3.2 The role of NBAS in vesicle transport 44
1.3.3 Impact of NBAS on human health 46
1.3.4 Evidence for NBAS as an NMD factor 47
1.4 Endoplasmic reticulum stress 51
1.4.1 ER homeostasis 51
1.4.2 Mechanisms of ER stress 52
1.4.3 Effect of NMD on ER stress 59
1.4.4 ER-NMD 63
1.5 Aims and research outline 66
1.5.1 Specific motivations and importance 66
1.5.2 Main research questions 66

2 Methodology **69**
2.1 Laboratory techniques 71
2.1.1 Strains and plasmids 71
2.1.2 Cell culture 71
2.1.3 Bacterial DNA transformation 71
2.1.4 Cell transfections 72
2.1.5 RNA extraction 72
2.1.6 Gel electrophoresis 73
2.1.7 Quantitative real time – polymerase chain reaction (qRT-PCR) 73
2.1.8 Western blotting 73
2.1.9 CRISPR strategy and cloning 74
2.1.10 CRISPR targeting and screening 75
2.1.11 Flow cytometry and fluorescence-activated cell sorting (FACS) 76
2.1.12 Immunoprecipitation and co-immunoprecipitation 77
2.1.13 Mass spectrometry 77
2.1.14 Immunofluorescence 78
2.1.15 Proximity ligation assay 79
2.1.16 Metabolic labelling and RNA fractionation 80

2.1.17 Library preparation and RNA-sequencing 81
2.1.18 Materials and reagents 82
2.2 Computational techniques 84
2.2.1 Quality control of RNA-sequencing data 84
2.2.2 Mapping of RNA-sequencing data 84
2.2.3 Differential expression analysis of total RNA-sequencing . 85
2.2.4 Analysis of fractionated RNA-sequencing 85
2.2.5 Characterisation of gene subsets 86
2.2.6 Defining exonic and intronic reads 86
2.2.7 Differential stability analysis using total and nascent RNA-sequencing data 87
2.2.8 Differential stability analysis using only total RNA-sequencing data 89
2.2.9 Measuring expression of HeLa genome variants 90
2.2.10 Identification of NMD-targeted variants genome-wide . . . 91
2.2.11 Characterisation of differentially processed variants 93
2.2.12 Image analysis 93
2.2.13 Statistical analysis 94
2.2.14 Deposited data used 94
2.2.15 Software and packages used 95

3 NBAS interacts with the ER translocon 97
3.1 Introduction 99
3.1.1 Endogenous tagging of protein allows optimised isolation and imaging 99
3.1.2 CRISPR/Cas9 gene editing 101
3.1.3 Aims 104
3.2 Results 105
3.2.1 CRISPR design and strategy 105
3.2.2 Using CRISPR to create a stable tagged-NBAS cell line . . 107
3.2.3 Validation of GFP-NBAS cell line 114

3.2.4 GFP-NBAS interacts with Syntaxin-18 complex and also the ER translocon . . . 119
3.2.5 NBAS interacts with UPF1 in a domain-non-specific manner . . . 124
3.3 Discussion . . . 128
3.3.1 Creation of GFP-FLAG-NBAS cell line . . . 128
3.3.2 Interaction of UPF1, NBAS and the ER translocon . . . 129
3.3.3 Future work . . . 133
3.3.4 Summary . . . 135

4 NBAS regulates a subset of UPF1 targets 137
4.1 Introduction . . . 139
4.1.1 Differential expression as a tool to identify targets of NMD 139
4.1.2 Aims . . . 141
4.2 Results . . . 142
4.2.1 Experimental design . . . 142
4.2.2 Processing of RNA-sequencing data . . . 147
4.2.3 Hundreds of genes are differentially expressed upon depletion of UPF1 or NBAS . . . 151
4.2.4 NBAS and UPF1 co-regulate mRNA expression specifically at the ER . . . 154
4.2.5 UPF1-independent NBAS increases in expression and NBAS decreases in expression . . . 160
4.2.6 Subcellular changes to mRNA expression . . . 161
4.3 Discussion . . . 171
4.3.1 NBAS regulates a subset of UPF1 targets . . . 171
4.3.2 ER-NMD is important for ER homeostasis . . . 174
4.3.3 Future work . . . 175
4.3.4 Summary . . . 176

5 Quantifying changes in RNA stability reveals many expression changes are independent of NMD 177

5.1 Introduction . . . 179
5.1.1 Benefits of differential stability analysis to identify targets of NMD . . . 179
5.1.2 Aims . . . 184
5.2 Results . . . 185
5.2.1 Defining exonic and intronic reads . . . 185
5.2.2 Using exonic and intronic reads from paired nascent and total RNA normalises for mRNA processing and transcription rate . . . 187
5.2.3 DESeq2 provides an appropriate framework to analyse differential transcript stability . . . 192
5.2.4 Comparing exonic and intronic reads normalises for mRNA processing using only total RNA . . . 200
5.2.5 Comparing RNA stability measurements by nascent and total RNA or total exonic and intronic RNA methods indicates a good correlation between the methodologies . . . 208
5.2.6 Comparing changes in RNA expression and RNA stability reveals many abundance changes are dependent on transcription . . . 208
5.3 Discussion . . . 215
5.3.1 It is possible to measure transcript stability at a single time point without transcription inhibition . . . 215
5.3.2 Only a subset of transcripts changed in expression when UPF1 is depleted also change in stability . . . 217
5.3.3 Effect of NBAS on RNA stability . . . 219
5.3.4 Future work . . . 220
5.3.5 Summary . . . 221

6 Matched genome, total RNA and nascent RNA sequencing allows identification of NMD triggering variants genome-wide 225

6.1 Introduction . . . 227

6.1.1 NMD target characteristics . . . 227
6.1.2 Aims . . . 230
6.2 Results . . . 231
6.2.1 Stop gain variants that are destroyed in control cells but survive when NMD is abrogated . . . 231
6.2.2 Measuring expression of HeLa genome variants identifies thousands of biallelic transcripts . . . 234
6.2.3 Assessing NMD efficiency on variants genome-wide . . . 242
6.3 Discussion . . . 256
6.3.1 Measuring NMD efficiency genome-wide using matched genome and RNA-sequencing requires further optimisation . . . 256
6.3.2 Future Work . . . 258
6.3.3 Summary . . . 260

7 Discussion **263**
7.1 Appropriate methodology is vital to study NMD effects . . . 265
7.2 An ER-NMD pathway quality controls transcripts translated at the ER . . . 267
7.3 The role of NBAS in RNA quality control requires further investigation . . . 269
7.4 Closing Remarks . . . 273

Bibliography **275**

Appendices

A Sequence of eGFP-3xFLAG-NBAS repair template **325**

B Total and fractionated RNA-sequencing mapping statistics **327**

C Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum **329**

Abbreviations

Acronym	Meaning
3' UTR	3' untranslated region
4sU	4-thiouridine
5' UTR	5' untranslated region
A	Adenine
ALT	Alternative allele
C	Cytosine
CB	Chromatoid body
CBC	Cap-binding complex
cDNA	Complementary DNA
CDS	Coding sequence
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CI	Confidence interval
CLIP	Cross-linking immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
ddPCR	Digital droplet polymerase chain reaction
DECID	RNA decay inducing complex
DMD	Duchenne muscular dystrophy
DNA	Deoxyribonucleic acid
DSB	Double strand break
EJC	Exon junction complex

ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
FACS	Fluorescence activated cell sorting
FCCS	Fluorescence cross-correlation spectroscopy
FCS	Fluorescence correlation spectroscopy
FLIM	Fluorescence Lifetime Imaging
FRET	Förster resonance energy transfer
G	Guanine
gDNA	Genomic DNA
GFP	Green fluorescent protein
GMD	Glucocorticoid receptor-mediated mRNA decay
GO	Gene ontology
GQC	Golgi quality control
GR	Glucocorticoid receptor
gRNA	Guide RNA
HDR	Homology directed repair
Het	Heterozygous
HL	RNA Half-life
Hom	Homozygous
HR	Homologous recombination
IF	Immunofluorescence
ILFS2	Infantile liver failure syndrome 2
IP	Immunoprecipitation
lncRNA	Long non-coding RNA
KO	Knock out
LRT	Likelihood ratio test
(IP-)MS	(Immunoprecipitation-) Mass spectrometry
MHC	Major histocompatibility complex
miRNA	Micro RNA
mRNA	Messenger RNA

N	Any nucleotide; adenosine, thymine, guanosine or cytosine
NGD	No-go decay
NHEJ	Non-homologous end joining
NITC	Nonsense-induced transcriptional compensation
NMD	Nonsense-mediated decay
NRZ complex	NBAS-RINT1-ZW10 complex
NSD	Non-stop decay
OI	Osteogenesis imperfecta
OR	Odds ratio
PAM	Protospacer adjacent motif
PB	Processing body
PCR	Polymerase chain reaction
PLA	Proximity ligation assay
Pol I	RNA polymerase one
Pol II	RNA polymerase two
Pol III	RNA polymerase three
Poly-A	Poly-adenine tail
Pre-mRNA	Premature messenger RNA
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
PTC	Premature termination (stop) codon
RAPP	Regulation of aberrant protein production
RBP	RNA-binding protein
REF	Reference allele
RIN	RNA integrity number
RMD	Regenase1-mediated mRNA decay
RNA	Ribonucleic acid
RNAi	RNA interference system
(m)RNP	(messenger) Ribonucleoprotein complex
rRNA	Ribosomal ribonucleic acid
SMD	Staufen-mediated mRNA decay

SNARE	Soluble NSF attachment protein receptor
snRNA	small nuclear RNA
SNV	Single nucleotide variant
SOPH	Short stature, optic nerve atrophy, and Pelger-Huet anomaly
SP	Signal peptide
SRP	Signal recognition particle
SRPR	Signal recognition particle receptor
SURF Complex	SMG1-UPF1-Release Factor Complex
T	Thymine
TCRβ	T-cell receptor β
TF	Transcription factor
TMD	Transmembrane domain
TPM	Transcripts per million
TR	Transcription rate
tracrRNA	Trans-activating CRISPR RNA
tRNA	Transfer RNA
U	Uridine
uORF	Upstream open reading frame
UPR	Unfolded protein response
VCF	Variant Call Format
VEP	Variant effect predictor
WGS	Whole-genome sequencing
WT	Wild type

Protein aliases

Protein name used in this thesis	Alternative protein names
NBAS	NAG
UPF1	SMG2, RENT1
UPF3A	UPF3
UPF3B	UPF3X
BNIP1	SEC20, TRG-8
USE1	P31
RINT1	ILFS3, TIP20L
Syntaxin-18	STX18, UFE1
UPF2	RENT2
Y14	RBM8A

List of Figures

1.1 The genetic code 4
1.2 Regulation occurs at every step of gene expression 6
1.3 Detail of a mature ribonucleoprotein (mRNP) particle 9
1.4 Translation of mRNA into nascent peptide 14
1.5 Translation of secretome proteins at the endoplasmic reticulum . . 18
1.6 Canonical nonsense mediated decay pathway 26
1.7 NMD target characteristics 35
1.8 NBAS is located at the endoplasmic reticulum 43
1.9 Evidence of NMD function of NBAS 49
1.10 Mechanisms of endoplasmic reticulum stress 53
1.11 The interplay between NMD and the UPR ER stress pathway . . 60
1.12 Theoretical model of ER-NMD 64

2.1 Schematic of custom analysis pipeline to calculate change in stability of genes in depleted cells compared to control 87

3.1 CRISPR/Cas9 type-II system from Streptococcus pyogenes 102
3.2 CRISPR/Cas9 strategy to target the NBAS genomic locus 106
3.3 Assessment of efficiency of CRISPR gRNAs 109
3.4 Overview of workflow used for CRISPR targeting strategy 111
3.5 Clonal selection and screening of CRISPR targeted cells 112
3.6 Comparison of homozygous and heterozygous GFP-FLAG-NBAS cell line 115
3.7 Validation of homozygous GFP-FLAG-NBAS cell line 117
3.8 GFP-FLAG-NBAS is localised to the ER 118

3.9 Confirmation of interaction between GFP-FLAG-NBAS and UPF1 120
3.10 GFP-NBAS interacts with Syntaxin-18 complex and also the ER translocon 122
3.11 GFP-NBAS interactome is RNA-independent 125
3.12 GFP-NBAS interaction with ER translocon is validated by PLA . 126
3.13 NBAS interacts with UPF1 in a domain-unspecific manner 127
3.14 Summary diagram of ER-NMD mechanism 130

4.1 RNA-sequencing experimental design 143
4.2 siRNA targeting of NBAS results in undetectable level of protein and upregulation of NMD targets 145
4.3 RNA-sequencing experimental design 148
4.4 Poison Exon Alternative Splicing Confirms NMD abrogation . . . 150
4.5 Hundreds of genes are differentially expressed upon depletion of UPF1 or NBAS 152
4.6 NBAS and UPF1 co-regulate mRNA expression and affect ER stress 155
4.7 Targets of NBAS and UPF1 are translated at the ER 159
4.8 UPF1-independent NBAS effects on mRNA expression 162
4.9 Experimental design of fractionation RNA-sequencing 163
4.10 Validation of total mRNA expression changes from fractionation experiment 166
4.11 Correlations and differential expression of RNA fractions 168
4.12 Effect of NMD factor abrogation on expression of membrane-associated genes 169

5.1 Measuring RNA stability requires more than total RNA 180
5.2 Mapping reads with STAR allows identification of intronic read coverage 186
5.3 Schematic of custom analysis pipeline to calculate change in stability of genes in depleted cells compared to control 188

5.4 Measuring stability using paired nascent and total RNA 190
5.5 Differential Stability of Individual Genes 193
5.6 DESeq2 analysis of total and nascent RNA produces a robust measure of stability 198
5.7 Comparing exonic and intronic reads allows RNA stability measurement using only total RNA 202
5.8 DESeq2 analysis of total exonic and intronic RNA identifies hundreds in changes in stability upon UPF1 depletion but not when NBAS is depleted 206
5.9 Comparison of RNA stability measurements by nascent/total RNA and exonic/intronic RNA methods 209
5.10 DESeq2 analysis of differential abundance and differential transcription from STAR alignment 211
5.11 Comparing changes in RNA expression and RNA stability reveals many abundance changes are dependent on transcription 214

6.1 Matched nascent and total RNA-sequencing allows tracking of heterozygous variants 229
6.2 It is possible to detect stop gain variants that are destroyed in control cells but survive when UPF1 is depleted 233
6.3 The haplotype phased genome of HeLa cells 236
6.4 Thousands of variants are biallelic whilst hundreds are multiallelic 237
6.5 Processing and filtering of HeLa genome variants 239
6.6 Further heterozygous variants are present in the HeLa cells used here than in Adey et al. (2013) 241
6.7 Significantly changed variants identified by DESeq2 model 242
6.8 Significantly changed variants are spread across the genome . . . 243
6.9 The stop gain variant within the SMG5 gene in detail 245
6.10 Calculating reference and alternate allele ratios allows comparison of allele-specific RNA processing 247

6.11 Calculating independent odds ratios for si-Scr and si-UPF1 treated samples allows to quantification of changes in allele-specific transcript stability between condition 250
6.12 Using DESeq2 to independently calculate si-Scr and si-UPF1 odds ratios identifies differential allele processing 253

List of Tables

4.1 Total and nascent RNA-sequencing mapping statistics 151

5.1 Mapping efficiency of total and nascent RNA with STAR 187
5.2 SampleTable for DESeq2 total and nascent analysis 195
5.3 SampleTable for DESeq2 total exonic and intronic analysis 205

6.1 Heterozygous stop gain variants identified in mock nascent RNA-sequencing 232

Chapter 1

Introduction

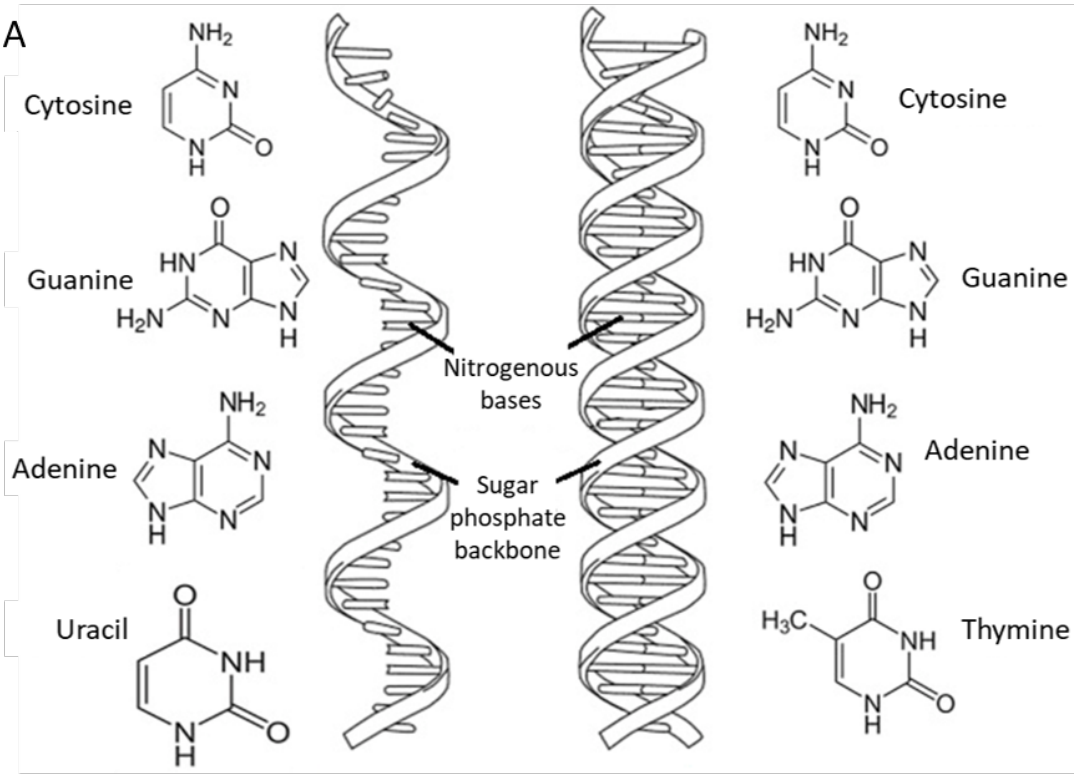
1.1 Regulation of gene expression

1.1.1 Expression of genetic information

Whilst DNA contains the genetic information that makes up all living things, RNA, transcribed from the DNA template, carries this coding information to the site of protein synthesis: the ribosome. Here, the nucleic acid polymer is translated into a chain of amino acids to form a protein. With RNA, as in DNA, genetic information is encoded in the sequence of nucleotides arranged into codons consisting of three bases each (Figure 1.1). Three nucleotides are present in both RNA and DNA; adenine (A), cytosine (C) and guanine (G), DNA also contains thymine (T), whereas in RNA, the base uracil (U) takes the place of thymine. Each codon defines a specific amino acid, except the stop codons that terminate protein synthesis (Matthaei et al., 1962; Nirenberg et al., 1965). Unlike the other 61 triplets of the genetic code, the termination codons UAA, UAG, and UGA instead serve as translation termination signals.

When the human genome project was launched, many scientists were convinced they would soon be able to identify the gene encoding every protein and many mysteries would be solved. However, as well as the astounding discovery that 98% of the human genome does not code for proteins, it was surprising that only ∼19,500 genes were present (Lander et al., 2001; Venter et al., 2001; International Human Genome Sequencing Consortium, 2004). For fewer than 20,000 genes to produce the number different proteins expressed by human cells, currently estimated between 40,000 (Adkins et al., 2002) and several billion (Smith et al., 2013), each expressed spatially and temporally, requires a complex network of processes regulating every step of gene expression (Figure 1.2).

Chromatin structure and accessibility affect the level to which a gene is transcribed. For example, by having enough space between the histone protein complexes that chromatin is wound around for the transcription machinery to assemble. This can also be influenced by signals that promote or repress transcrip-



B

Second Nucleotide

First Nucleotide	U	C	A	G	Third Nucleotide
U	UUU Phe UUC UUA Leu UUG	UCU UCC UCA UCG Ser	UAU Tyr UAC UAA Stop UAG Stop	UGU Cys UGC UGA Stop UGG Trp	U C A G
C	CUU CUC CUA CUG Leu	CCU CCC CCA CCG Pro	CAU His CAC CAA Gln CAG	CGU CGC CGA CGG Arg	U C A G
A	AUU AUC AUA Ile AUG Met	ACU ACC ACA ACG Thr	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU GUC GUA GUG Val	GCU GCC GCA GCG Ala	GAU Asp GAC GAA Glu GAG	GGU GGC GGA GGG Gly	U C A G

Figure 1.1: ***The genetic code***

Figure 1.1: ***The genetic code***

(A) *DNA (right) and RNA (left) have a similar structure and are made up of three of the same nucleotides; cytosine (C), guanine (G) and adenine (A). DNA's fourth nucleotide monomer is thymine (T) whereas RNA utilises uracil (U). DNA is generally double-stranded, whereas RNA is generally single stranded.* ***(B)*** *The code of amino acids from RNA three nucleotide codons. There is significant redundancy in the system including three codons that signal translation termination.*

tion including epigenetic marks on cytosine nucleotides and histone proteins, and whether the gene's promoter(s) is strong or weak according to its affinity for RNA polymerase. Tissue specific enhancers and transcription factors allow expression only in correct cell types. The regulation of gene expression by modulation of the transcription process includes the activity of transcription factors (TFs), which influence expression of their target genes and tissue specific enhancers which drive temporal control of when the gene is transcribed. Although the rules governing enhancer-promoter interactions are not fully understood, the interplay of multiple promoters and enhancers for a single gene allows increased regulatory diversity.

Multiple proteins can be produced from a single DNA gene by way of alternate mRNA isoforms, on average ten per gene in human cells. Alternative splicing describes the differential inclusion of the exons and leads to the inclusion or exclusion of protein domains and active sites, thereby altering the functions of the resulting protein. Alternative start sites within the DNA sequence can also alter the inclusion of multiple exons and it has been shown that the DNA sequence can affect speed of transcription of RNA polymerase II (Pol II) which can also lead to differential exon inclusion (Naftelberg et al., 2015). This ability to produce multiple proteins from the same gene allows for tissue-specific functions in eukaryote cells.

All RNA species that localise outside the nucleus, must pass through the nuclear pore complex from the nucleus to the cytoplasm, often with the aid of export factors. This is a highly regulated process which relies on the assembly of large ribonucleoprotein (RNP) particles.

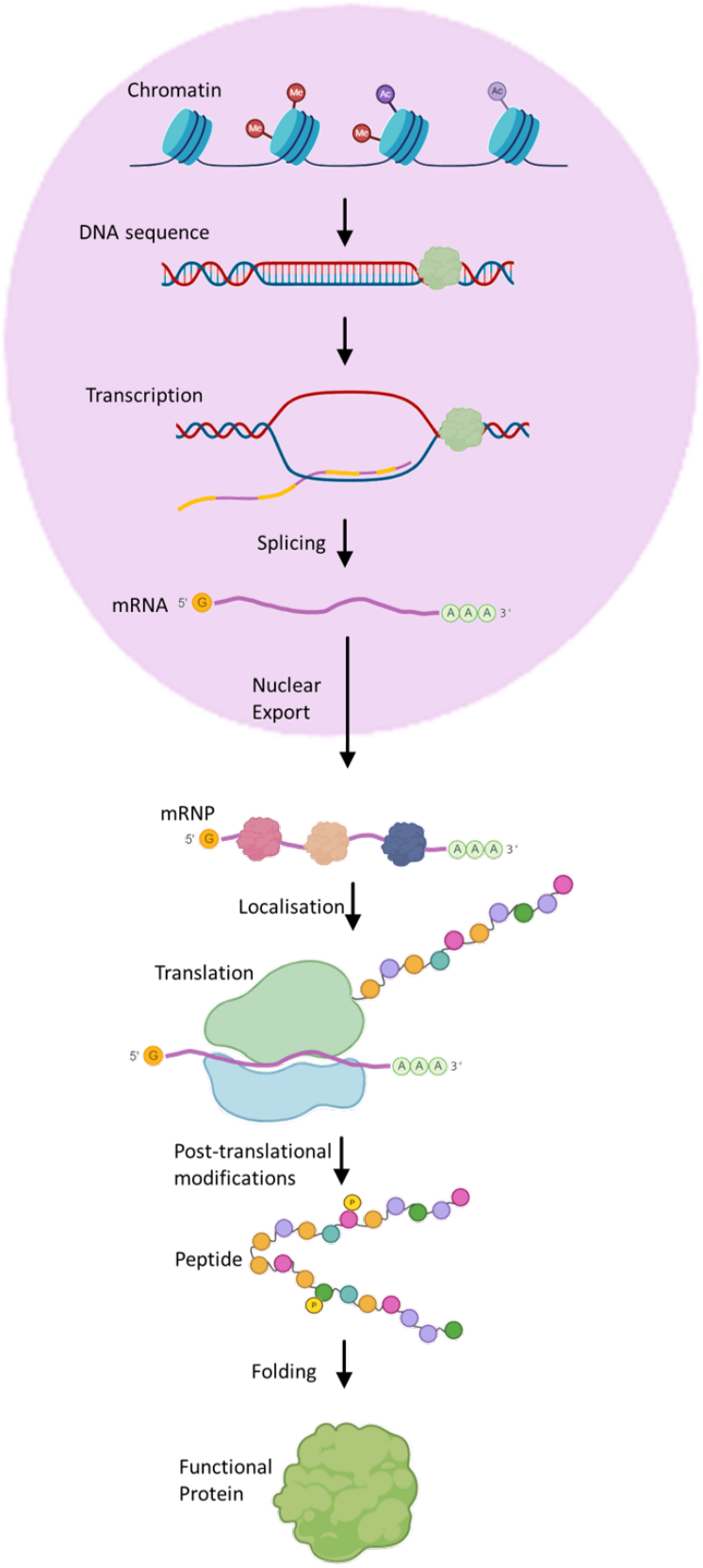


Figure 1.2: ***Regulation occurs at every step of gene expression***

Figure 1.2: ***Regulation occurs at every step of gene expression***
DNA is stored wrapped around histone proteins as chromatin. These histone proteins can be modified by addition of various small chemical groups termed epigenetics which can either promote or suppress transcription. The DNA sequence itself effects gene expression due to the effect on biding efficiency of proteins such as transcription factors. DNA sequence can also effect speed of polymerase II transcription could affect accuracy of the polymerase and influences alternative splicing, levels of transcript produced and/or correct termination of transcription. Enhancers and promoters also control the level of transcription of individual genes both temporally and in different cell types. Splicing removes non-coding intronic regions and joins together coding exonic regions, ready for translation. Alternative splicing leads to the production of different transcripts from the same gene which results in alternative proteins that can have differing functions. Nuclear export and localisation of the mRNA is facilitated by binding of multiple proteins to form the messenger ribonucleoprotein complex (mRNP). Local translation allows spatial control of gene expression and has been demonstrated in a variety of subcellular compartments. Translation of the RNA sequence into the amino acid chain of the equivalent protein is tightly regulated, both on a global and transcript-specific level. Post translational modification to individual amino acids of the peptide chain can activate the protein to carry out a role or alternatively mark the protein for degradation. Folding of the peptide chain into a functional protein enable the specific function of the protein whether enzymatic, structural or otherwise.

Many different mechanisms regulate translation both on a global and transcript-specific level (Gebauer and Hentze, 2004; Sonenberg and Hinnebusch, 2009, Section 1.2.1). Local translation allows spatial control of gene expression and has been demonstrated in a variety of subcellular compartments (Chouaib et al., 2020, Section 1.2.6).

Post translational modifications affect the function and localisation of proteins, leading to massive variation in protein function. This includes addition of various functional groups to amino acid residues including methyl groups, ubiquitin, SUMO, citrulline, phosphate groups and acetyl groups.

Furthermore, RNA itself can regulate the process of gene expression. RNAs carrying information to form a protein are called messenger RNA (mRNA) but the most abundant RNAs in the cell are transfer RNA (tRNAs) and ribosomal RNA (rRNAs), both of which are involved in translation. Small nuclear RNAs (snRNAs) are found within the nucleus and are instrumental in the spliceosome, which splices the introns from pre-mRNA.

Many long noncoding RNAs (lncRNAs) are localised to the nucleus and can recruit chromatin regulatory proteins to genomic DNA to regulate gene expression (Young and Ponting, 2013; Quinodoz and Guttman, 2014). Whereas post transcriptional gene expression is dominated by the action of micro RNAs (miRNAs). miRNAs are short, roughly 22 nucleotides (nt) long, untranslated RNAs whose biogenesis involves folding of RNA by the Drosha protein to form double-stranded RNA and unwinding of this two stranded duplex regulated by the protein Dicer to leave a single-stranded RNA molecule. miRNAs bind to their target mRNA and trigger its degradation via cleavage by the Argonaute family of proteins or by blocking translation. Therefore, miRNA binding ultimately results in little or no protein being produced from the target mRNA. There is extensive literature regarding this aspect of translational regulation, which will not be discussed in detail here.

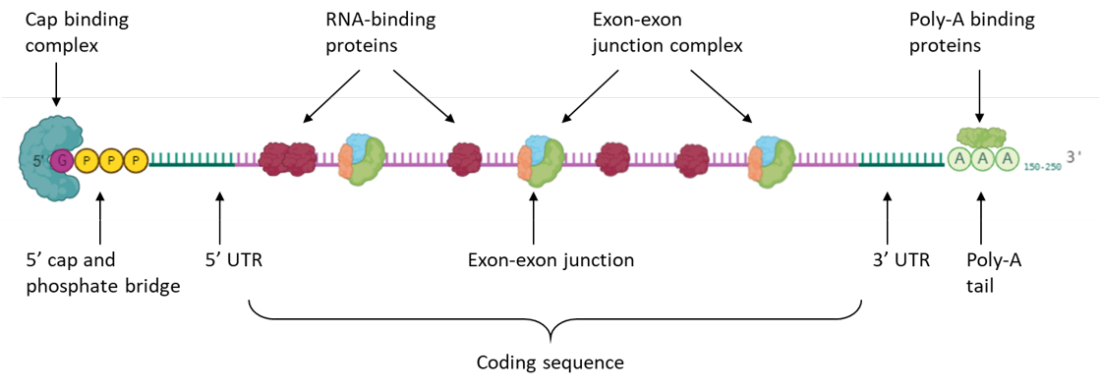


Figure 1.3: ***Detail of a mature ribonucleoprotein (mRNP) particle***
The messenger RNA molecule exists as part of a ribonucleoprotein (RNP) with multiple subtypes of RNA-binding proteins (RBPs). After splicing, only the exonic regions of RNA remain which includes both the protein coding sequence (CDS) and 3' and 5' untranslated regions (UTR). The junctions where two exons have been spliced together are marked by the exon-exon junction complex which is made up of four core proteins and various peripheral factors. The 5' end of the mRNA is protected by the 5' cap consisting of a three phosphate group bridge and a terminal guanine. This is then bound by the cap binding protein complex. The 3' end of the mRNA is protected by up to ∼250 adenine nucleotides called the poly-A tail. This tail is bound by specific poly-A binding proteins.

1.1.2 Processing of messenger RNA (mRNA)

Whilst mRNA is often referred to as a "copy" of the DNA sequence of a gene, the processing of a nascent transcript to a mature mRNA ready for translation is actually multifaceted (Moore and Proudfoot, 2009). The complementary pre-mRNA must be spliced, coated with RNA binding proteins (RBPs), exon-exon junctions marked, a 5' cap added, the resulting RNP complex exported from the nucleus and localised to the ribosome ready for translation (Figure 1.3).

RNA is synthesised by the enzyme RNA polymerase, using the genomic DNA as a template in a process termed transcription. Whereas prokaryotes have a single RNA polymerase to transcribe all their genes, the nuclei of eukaryotes contain three. Each is responsible for transcribing a unique set of RNAs. RNA

polymerase I (Pol I) synthesises most of the rRNA (discussed further in Section 1.1.3), RNA polymerase II (Pol II) synthesises the protein-encoding mRNA and most snRNA, while RNA polymerase III (Pol III) synthesises a variety of small untranslated RNAs, including tRNA and the smaller subunit of rRNA. Although it only transcribes a single gene, RNA polymerase I is highly active, producing ∼50% of the nascent RNA in the cell (Russell and Zomerdijk, 2006) and in growing cells, the combined activities of Pol I and Pol III can exceed 80% of nuclear transcription (Paule and White, 2000).

The newly synthesised RNA molecules are referred to as nascent RNA, and generally require further processing before they become functional. Similar to DNA, RNA base modifications can also occur, primarily in tRNA and rRNA soon after the RNA is transcribed. These modifications are essential for their proper function in protein translation. However, the most significant modification to the nascent RNA is pre-mRNA splicing.

Mammalian genes consist of exonic regions which contain the coding sequence and the regulatory untranslated regions, and intronic regions which are generally non-coding but can contain lncRNAs, miRNAs and cis-regulatory modules. In order for mRNA to mature into a functional message, the correct removal of introns must occur to produce a coding transcript (Will and Lührmann, 2011; Shi, 2017; Wilkinson et al., 2020). Splicing consists of creating a break in the RNA molecule at the 5' splice site and the resulting free OH group of the upstream exon reacting with and ultimately joining to the phosphate group of the downstream exon at the 3' splice site, thereby expelling the intron. These reactions can be self-splicing either through the aid of either a free guanine ribose or an adenine within the intron initially attacking the 5' splice site. However, splicing can also take place with the aid of up to 150 proteins and five small nuclear RNAs (snRNAs), together termed the spliceosome (Shi, 2017; Wilkinson et al., 2020). The five snRNAs, U1, U2, U4, U5 and U6, assembly in an ordered, stepwise manner on the pre-mRNA substrate and facilitate the reaction between the intron embedded adenine and the 5' splice site. The resulting join between two exons is termed an exon-exon junction. The three main splicing sequence triggers are found within the intron;

a donor site at the 5' end of the intron, the branch site towards the 3' end of the intron and an acceptor site at the 3' end of the intron. The splice donor site consists of a highly conserved sequence GU, within a larger, less highly conserved region. The splice acceptor is also highly conserved, with an AG sequence at the 3' end of almost all introns. The branchpoint contains the A nucleotide involved in lariat formation (Wang and Aifantis, 2020).

Despite the strong sequence requirements for canonical splicing, it can sometimes go wrong. For example, intron retention, a type of alternative splicing, can occur in humans when consensus sequences are not recognised, meaning the intron is not spliced out. If the intron that is retained contains stop codons or a shift in the reading frame this could lead to mRNA degradation (Sibley et al., 2016; Wang and Aifantis, 2020, (Section 1.2.2)). In addition, non-canonical splicing events can occur at a low frequency at cryptic splice sites. Such splice sites comprise a consensus akin to those at canonical splice sites, but are located in non-canonical locations within the introns. Considering the volume of intronic sequence in vertebrate genomes, it is unsurprising that this can occur by chance (Sibley et al., 2016).

Eukaryotic messenger RNA does not exist by itself but rather in large complexes consisting of multiple protein factors and non-coding RNAs, together forming large messenger ribonucleotide particles (mRNPs). The combination of molecules ultimately decides the fate of each transcript by influencing virtually every step in gene expression.

One such protein complex loaded onto mRNAs is the exon-exon junction complex (EJC). The EJC is a large protein complex loaded onto transcripts ∼20–24nt upstream of exon-exon junctions as a consequence of the splicing process (Le Hir et al., 2000; Singh et al., 2012; Saulière et al., 2012). The EJC serves as a positional landmark for the intron-exon structure of genes, and also directs post-transcriptional processes in the cytoplasm. Most, although not all (Hauer et al., 2016; Le Hir et al., 2016), exon junctions are loaded with EJC by the splicing machinery, consisting of the hetero-tetramer core of eIF4A3, MAGOH, RBM8A (Y14) and Barentsz (BTZ), as well as auxiliary factors including UPF2, UPF3B,

MLN51 and RNPS1. The protein complex binds to mRNA in a splicing-dependent, but sequence-independent, manner, triggered by the recruitment of eIF4A3 by the spliceosome (Alexandrov et al., 2012; Barbosa et al., 2012; Steckelberg et al., 2012). eIF4A3 is the main RNA-binding protein of the EJC, with the other three core components mainly acting to keep eIF4A3 bound to the RNA and in the correct orientation. The RNA binding of eIF4A3 does not involve the RNA bases but occurs mainly via the ribose-phosphate backbone (Andersen et al., 2006; Bono et al., 2006).

Once a mature RNA molecule is correctly processed, it must be exported from the nucleus into the cytoplasm. Small RNAs (such as tRNAs and miRNAs) follow relatively simple export routes by binding directly to export receptors, whereas larger RNAs (including rRNAs and mRNAs) travel through the nuclear pore complex via export factors (Köhler and Hurt, 2007). Several export proteins are highly conserved and essential for correct mRNA localisation to the cytoplasm (Huang et al., 2003; Carmody and Wente, 2009; Katahira, 2015). Direct interaction with the nuclear pore complex is then possible which promotes export of the mature mRNP to the cytoplasm (Lai and Tarn, 2004) and the carrier proteins are subsequently recycled.

The research presented in this thesis focuses on understanding gene regulation at the RNA level, specifically after nuclear export and at the translation step of expression.

1.1.3 Translation of mRNAs

Once mRNP particles reach the ribosome, their cargo mRNA can be translated into a polypeptide chain. mRNA translation is a highly complex process, which is tightly regulated, both on a global and transcript-specific level (Gebauer and Hentze, 2004; Sonenberg and Hinnebusch, 2009). Fundamentally, transfer RNA (tRNA) molecules which recognise the three base codons bind sequentially to the mRNA and thereby add their bound amino acid (aminoacyl-tRNA) to the peptide chain. The translation process is conducted by the ribosome.

Translation can be divided into three separate consecutive steps (Figure 1.4): initiation, the process of assembly of the ribosomal subunits together with over 70 auxiliary factors; elongation, the actual synthesis of a peptide chain; and termination, which entails the dissociation of the full-length peptide chain and removal of the ribosome from the mRNA transcript. At each stage of the process there is a high degree of both internal and external regulation by various RNA binding proteins.

The majority of regulation occurs during the initiation phase, which requires numerous RNA binding proteins (Sonenberg and Hinnebusch, 2009). Throughout evolution of higher organisms, ribosome size has increased to provide an interactive landing site for auxiliary factors and facilitate customised mRNA translational landscapes (Simsek et al., 2017). Eukaryotes have 80S ribosomes, each consisting of a small (40S) and large (60S) subunit. The 40S subunit is made up of the 18S rRNA and 33 proteins, whilst the large subunit is composed of the 5S rRNA, 28S rRNA, 5.8S rRNA and 46 proteins. During initiation, the small ribosomal subunit, typically bound to an aminoacyl-tRNA containing the first amino acid methionine, recognizes the AUG start codon and binds, which recruits the large ribosomal subunit to form the complete initiation complex.

After initiation, the ribosome proceeds to elongate the initial amino acid into a polypeptide chain. As seen in Figure 1.1, there is a degree of redundancy in the codon code; 18 of the 20 amino acids are encoded by up to 6 different codons. The genetic code is almost universal which means that a given gene is most likely translated into the same protein in any species. However, the frequency of use of particular codons varies strongly between species. With increasing data from comparative analyses of codon usage in different organisms, it has become clear that synonymous codons are not randomly distributed along genomes and genes, and that certain organisms seem to prefer some over others (Grantham et al., 1980; Sharp and Li, 1986; Plotkin and Kudla, 2011). These codon biases have been proposed to contribute to the variation of synonymous codon usage seen

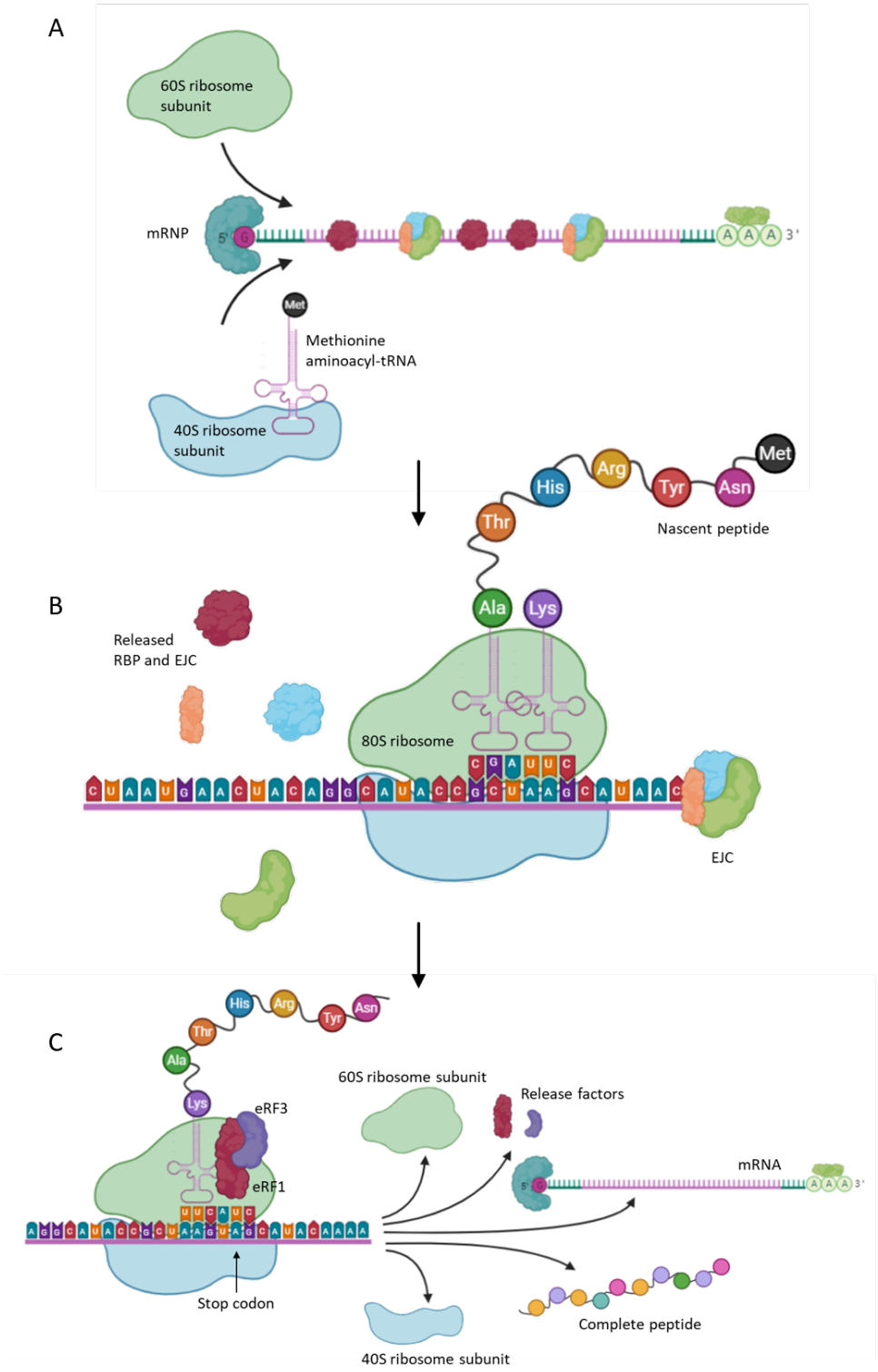


Figure 1.4: ***Translation of mRNA into nascent peptide***

Figure 1.4: ***Translation of mRNA into nascent peptide***
(A) *Translation begins at a methionine "start" codon at the end of the 5' UTR. The small (40S) ribosomal small unit, already bound to a methionine aminocyl-transferRNA (tRNA), recognises the methionine and attaches to the mRNA. The large (60S) ribosomal subunit then binds to assemble the complete (80S) ribosome.* ***(B)*** *As the ribosome translates in a 5'-3' direction, it displaces RBPs including the EJCs. Each three nucleotide codon is bound by a specific tRNA with an attached amino acid residue. These amino acids are then joined to form a nascent peptide chain.* ***(C)*** *When the ribosome reaches a stop codon, there is no corresponding tRNA so instead the release factors eRF1 and eRF3 bind. This triggers the termination of translation and disassociation of the ribosome and complete peptide from the mRNA. The ribosome and release factors are recycled whilst the peptide is further processed to produce a functional protein.*

amongst genes within same genomes (Cannarozzi et al., 2010; Coleman et al., 2008; Tuller et al., 2010). In this way, differential codon usage may act as a regulator of translation dynamics on a genome-wide level, adding an additional layer for fine-tuning gene expression (Quax et al., 2015).

Although translation initiation is commonly acknowledged as the rate limiting step in translation, several studies have also shown that the elongation rate can also vary. In mammalian cells, newly synthesised mRNAs undergo a pioneer round of translation. The cap first binds to the cap-binding protein heterodimer CBP80-CBP20 (Isken and Maquat, 2008) and one or more ribosomes are loaded, depending on the efficiency of translation initiation and the length of the open translational reading frame (Isken and Maquat, 2008; Isken et al., 2008). The CBP80-CBP20 heterodimer is then replaced by eukaryotic translation initiation factor 4E (eIF4E) and steady-state cycles of mRNA translation continue, generating most of the cell's proteins. Translation factors, RNA-binding proteins, and targets of signalling pathways that are particular to the newly synthesised mRNA regulate critical functions of the pioneer round. It is also during the pioneer round of translation that the EJCs are removed from the mRNA (Dostie and Dreyfuss, 2002; Lejeune et al., 2002; Sato and Maquat, 2009).

Because the abundance of the EJC proteins eIF4A3, MAGOH, Y14, and BTZ is limited, and the number of exon-exon junctions is high in the steady-state transcriptome, EJCs must be efficiently disassembled and recycled to ensure that maturing mRNAs can be appropriately tagged (Gehring et al., 2009). Even so, EJCs are not found at every exon-exon junction and have non-uniform occupancy at transcripts (Hauer et al., 2016; Le Hir et al., 2016). Removal of the complexes has been attributed to the activity of the translating ribosome and appears to be sequential as the ribosome progresses along the transcript. The ribosome associated protein PYM binds to the EJC heterodimer MAGOH/Y14 and facilitates dissociation of the EJC (Gehring et al., 2009).

When the ribosome reaches a stop codon, there is no complimentary tRNA so instead the release factors eRF1 and eRF3 bind and initiate translation termination (Zaher and Green, 2009a,b; Dever and Green, 2012). The three different stop codons; amber (UAG), ochre (UAA) and opal (UGA) have been shown to vary in their strength, or ability to successfully terminate translation. UAA is the strongest stop codon, whilst UGA is the weakest signal. Approximately 57% of mammalian genes are estimated to be affected by purifying selection acting on stop codon preference. This proportion varies substantially by codon, with UGA stop codons far more likely to be conserved. Notably, genes with conserved stop codons were found to have longer 3' UTRs and were also associated with shorter mRNA half-life than other genes (Seoighe et al., 2020). Under certain conditions, a near-cognate tRNA can be recruited when the ribosome reaches a stop codon, meaning an amino acid is incorporated and translation continues until a later stop codon is reached. This event is called stop codon readthrough (Palma and Lejeune, 2020). Most eukaryotic genes terminate with multiple stop codons, but if there is a single stop codon this is more likely to happen. Since the least efficient stop codon is the most conserved, it seems probable that readthrough may have useful functional consequences.

Cytoplasmic mRNA degradation generally begins with shortening of the poly-A tail by de-adenylases (Liu et al., 1998; Tucker et al., 2001), followed by de-capping (Lai et al., 2012). The unprotected mRNA is then degraded by exonucleases from both ends. Despite the spatial separation of mRNA synthesis, translation and degradation, several studies have shown evidence that these processes are coupled (Lotan et al., 2005, 2007; Harel-Sharvit et al., 2010). Indeed, compensatory changes in degradation and synthesis have been revealed that indicate buffering of mRNA levels to render gene expression robust in yeast (Sun et al., 2012) and human cells (Slobodin et al., 2020).

1.1.4 Translation at the endoplasmic reticulum

The endoplasmic reticulum (ER) is an expansive, membrane-enclosed organelle composed of smooth peripheral tubules and rough, ribosome-studded central sheets, whose function can be summarised as synthesis and export of proteins and membrane lipids. The binding site of the ribosome on the ER membrane is the translocon (Figure 1.5) (Rapoport et al., 2017). However, the ribosomes are not a stable part of this organelle's structure as they are constantly being bound and released from the ER membrane. A free ribosome only docks at the ER once it has bound the mRNA of a protein destined for the secretory pathway.

Secretory proteins include transmembrane proteins such as receptors and ion channels, as well as secreted proteins including neuropeptides and extracellular matrix components. Many of these proteins have distinct interaction domains on ribosomes that are involved in their targeting and transport. Transcripts are targeted for transport to the ER via various N-terminal hydrophobic signals.

Many secretome proteins encode an initial signal peptide (SP), a short sequence (usually 16-30 amino acids long) that is recognised by signal recognition particle (SRP) (Walter and Blobel, 1981; Wild et al., 2004). SP sequences are extremely heterogeneous, and the efficiency of protein secretion is strongly determined by the specific SP (Kober et al., 2013). Proteins with amino-terminal transmembrane domain are also recognised by SRP resulting in $\sim$15% of cellular transcripts being bound by SRP. Transmembrane domains biochemically resemble a SP but are

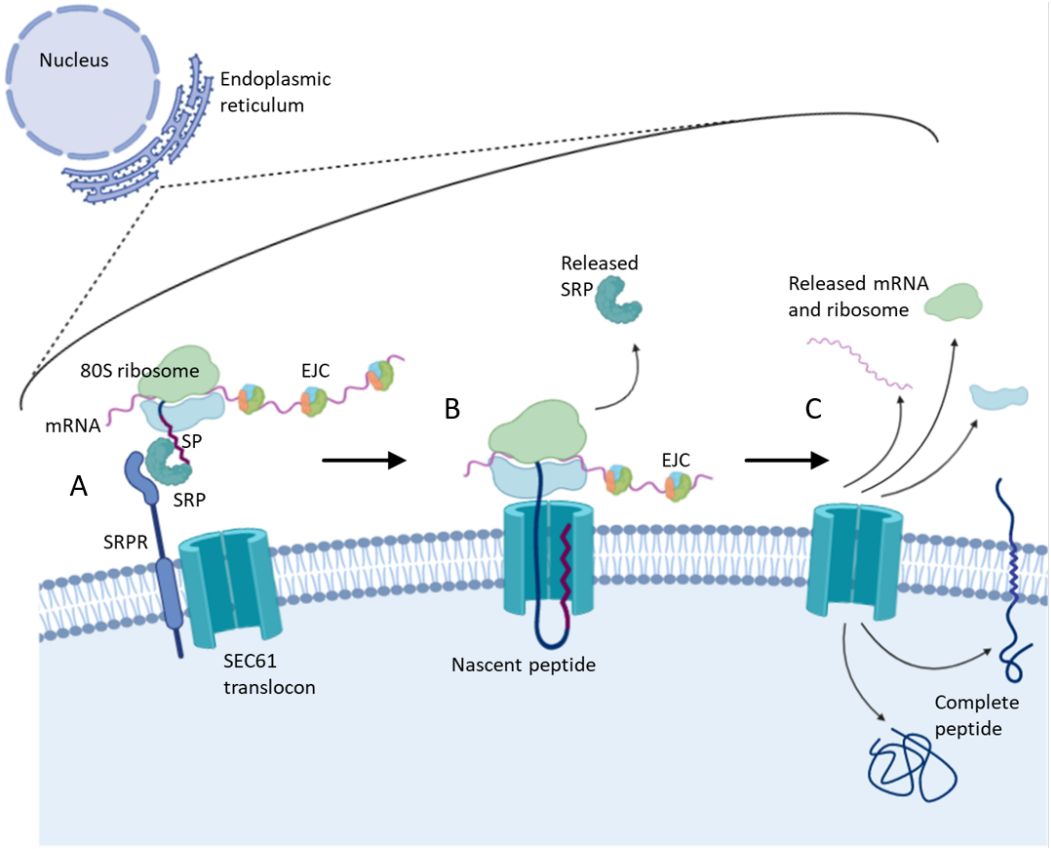


Figure 1.5: ***Translation of secretome proteins at the endoplasmic reticulum***

(A) *When the mRNA of a secretome protein begins to be transcribed, the N-terminal of the peptide chain encodes an ER localisation signal such as a transmembrane domain (TMD) or signal peptide (SP). This triggers the localisation of the ribosome-mRNA complex to the ER membrane. In the case of SP, it is bound by the signal recognition particle (SRP) which is subsequently bound by the signal recognition particle receptor (SRPR), which is anchored in the ER membrane near the SEC61 translocon.* ***(B)*** *The translating ribosome docks at the translocon, the SRP is released and the nascent peptide inserts into the channel. As the ribosome continues to translate the peptide is deposited directly into the ER lumen.* ***(C)*** *On completion of translation, the ribosome and mRNA are released from the ER membrane. The complete peptide is folded and processed in the Er lumen. If one or more TMD are present, the peptide is transferred from the translocon straight to the ER membrane.*

not cleaved after localisation. While secreted proteins are threaded through the channel, transmembrane domains may diffuse across a lateral gate in the translocon to partition into the surrounding membrane. SRP-independent mechanisms of targeting include the transmembrane domain recognition complex of 40 kDa (TRC40) pathway in mammals (Stefanovic and Hegde, 2007; Schuldiner et al., 2008) and the recently discovered SRP-independent targeting pathway, which preferentially targets proteins with transmembrane domains located downstream of the N-terminal (Aviram et al., 2016).

Systematic analysis of the secretome in yeast showed that about 57% of predicted secretory proteins are targeted by SRP whilst about 43% of secretome proteins are SRP-independent (Ast et al., 2013). Regardless of translocation mechanism, the vast majority of secretory proteins have been found to complete translation at the ER *in vivo*. The translation of these proteins begins on a free ribosome but once the localisation signal has been recognised, the ribosome and transcript are transported to the ER membrane (Jan et al., 2014).

There has been some debate about whether the ribosome is stalled during the relocation to the ER. By labelling mRNAs with various fluorescent tags, it is possible to study translation in real time at the single-mRNA level in living cells (Morisaki et al., 2016; Wang et al., 2016; Wu et al., 2016; Yan et al., 2016). By fusing a fluorescent epitope and tag-bearing mRNA with an N-terminal ER targeting signal, it is possible to visualise translation at the ER in live cells. Wu et al. (2016) were able to observe co-translational insertion of the nascent peptide into the ER and also found that the mRNAs were tethered to the ER only during active translation. This technique also allowed observation of a fraction of translating polysomes exhibiting rapid and directed motion in the dendrites of neurons (Wang et al., 2016), suggesting that mRNAs can undergo translation during transport and this is somehow repressed for ER-directed ribosomes.

Once at the ER membrane, the SRP interacts with the signal particle recognition receptor (SPRP) which is closely associated with the translocon channel. The ribosome-transcript complex docks at the translocon and the nascent peptide is secreted directly into the ER lumen. The SEC61 translocon channel is made up

of the SEC61A, SEC61B and SEC61G subunits. Once within the ER lumen, the peptide will be folded and modified, and then released to be transported to the correct cellular location. In addition to secretome proteins, up to 40% of cytosol-encoding mRNAs may be translated at the ER in a typical mammalian cell (Reid and Nicchitta, 2012). This shows the fluid nature of translation localisation, in that for many transcripts, they may be translated at the ER or on a free ribosome in the cytoplasm.

1.2 RNA quality control

1.2.1 Surveillance mechanisms regulating RNA quality

Despite the many layers of regulation discussed thus far, errors still occur and therefore eukaryotic cells utilise quality-control mechanisms that recognise and degrade mRNAs that have not completed nuclear pre-mRNA processing or that fail to encode a polypeptide faithful to the gene sequence. Such aberrant mRNAs are degraded rapidly, before the accumulation of abnormal protein products that could have adverse effects on the cell. The mechanisms that recognise and target aberrant mRNA for degradation take place at every step of gene expression (Isken and Maquat, 2007). Evolution pressure has led to evolving protective mechanisms that prevent appearance of defective proteins by destroying the aberrant proteins themselves or sensing the aberrations in the mRNA templates and degrading them. In this thesis, I will focus on mechanisms of RNA quality control that assess RNA and ensure aberrant transcripts do not lead to faulty proteins.

The No-Go Decay (NGD) mRNA surveillance pathway degrades mRNAs containing stacks of stalled ribosomes. NGD occurs when translation elongation is blocked by the presence of stable intramolecular or intermolecular RNA structures, enzymatic cleavage, chemically damaged sequences or rare codons. Ribosome stalling can result in a wide range of malignancies from protein misfolding to oncogenesis. The best characterised example is mutations in the FMR1 gene, that causes ribosomes to stall and translation to stop (Stefani et al., 2004; Shu et al., 2020), and leads to fragile X syndrome and intellectual disability (Richter and Coller, 2015). Correspondingly, NGD is evolutionarily conserved from yeast and unlike most RNA quality control pathways, NGD is even conserved in prokaryotes including archaea.

Transcripts synthesised without a stop codon due to premature polyadenylation also accumulate stalled ribosomes, that are initially detected by the non-stop decay (NSD) pathway. However, it has been shown that this can also ultimately lead to NGD. Translational readthrough, where the stop codon is skipped over, into the 3' UTR sequence can produce less protein but does not trigger non-stop decay as there is no connection with poly-A sequence. The 3' UTR sequence may be used to control protein levels of such transcripts where readthrough occurs (Arribere et al., 2016).

Staufen1 (STAU1)-mediated mRNA decay (SMD) is an mRNA degradation process in mammalian cells that is mediated by the binding of STAU1 to a STAU1-binding site within the 3' untranslated region (3' UTR) of target mRNAs. These binding sites consist of double-stranded RNA formed by a stem-loop structure in the 3' UTR or an RNA duplex structure formed by intermolecular mRNA–mRNA or mRNA–long noncoding RNA interactions (Ricci et al., 2014; de Lucas et al., 2014; Sugimoto et al., 2015). The ATP-dependent helicase UPF1 is recruited to mRNA via its interaction with STAU1 in a translation-dependent manner (Kim et al., 2005; Park and Maquat, 2013). The recruited UPF1 interacts with proline-rich nuclear receptor coregulatory protein 2 (PNRC2) and triggers rapid mRNA degradation.

Besides SMD, STAUs perform multiple functions in post-transcriptional regulatory processes including alternative splicing, translation and spatial targeting of mRNA (Thomas et al., 2009). STAU1 also promotes efficient replacement of the cap-binding heterodimer CBP20:CBP80 with eIF4E, after the pioneer round of translation (Jeong et al., 2019). This event is inhibited by interaction between STAU1 and UPF1. Increased association between STAU1 and hyperphosphorylated UPF1 has also been seen after cell stress stimuli, which inhibits this replacement. This could be explained by the interplay between UPF1, response to cell stress and inhibition of translation which will be covered in Section 1.4.3.

Regnase-1, a member of the CCCH zinc finger protein family possessing RNase activity, mediates post-transcriptional regulatory activity through degradation of target mRNAs, termed regnase-1–mediated mRNA decay (RMD) (Uehata and Akira, 2013). RMD is particularly important in regulating the magnitude of innate and adaptive immune responses (Uehata and Takeuchi, 2017), and dysfunction of Regnase-1 in mice leads to the development of spontaneous systemic inflammation (Konieczny et al., 2019).

Glucocorticoid receptor (GR) plays a role in rapid mRNA degradation by acting as an RNA-binding protein. GR is loaded onto the 5' UTR of a target mRNA and recruits UPF1 through PNRC2 in a ligand-dependent manner, so as to elicit rapid mRNA degradation, termed GR-mediated mRNA decay (GMD) (Cho et al., 2015). GMD is thought to be a mechanistically unique pathway of UPF1-dependent mRNA decay because it occurs in a translation-independent manner.

It has also been shown that inefficient SRP interaction with the nascent peptide can trigger a unique mRNA quality control mechanism that senses aberrant secretory proteins and degrades their mRNAs preventing synthesis of harmful products (Karamyshev et al., 2014). Termed regulation of aberrant protein production (RAPP), Karamyshev et al., found that mutations in the signal sequence changed the affinity of SRP and instead Ago2 argonaute protein was recruited. Severity of the signal sequence mutations correlated with increased proximity of Ago2 to the nascent peptide and subsequently to mRNA degradation, suggesting a general mechanism of secretome translational quality control.

However, the most studied and well characterised co-translational quality control mechanism is nonsense mediated decay (NMD). This pathway recognises transcripts with a premature termination codon (PTC) and leads to degradation of the aberrant mRNA and the nascent peptide, thus avoiding the production of truncated proteins that could have problematic consequences such as dominant-negative or deleterious gain-of-function effects (Leeds et al., 1991; Jacobson and Peltz, 1996; Neu-Yilik and Kulozik, 2004; Maquat, 2004; Wilkinson, 2005; Chang

et al., 2007; Lykke-Andersen and Jensen, 2015; He and Jacobson, 2015; Kurosaki and Maquat, 2016; Karousis et al., 2016). NMD is evolutionarily conserved through all eukaryotes, although no similar mechanism has been identified in prokaryotes.

Because both SMD and NMD employ UPF1, SMD and NMD are competitive pathways and are mechanistically related. GMD, NMD, and SMD all share the pathway components UPF1 and PNRC2, and both NMD and SMD occur in a translation-dependent manner, raising the question of how distinct these quality control and degradation pathways are from each other in vivo and the level of interplay between them.

1.2.2 Mechanism of nonsense mediated decay (NMD)

Nonsense mediated decay (NMD) was first described in yeast cells, when it was noticed that the presence of nonsense mutations reduced the level of mRNA but did not affect the rate of transcription (Losson and Lacroute, 1979). It was also noted that the strength of this reduction depended on the position of the nonsense codon within the locus. Soon after, whilst investigating a molecular defect identified in four Kurdish Jews with homozygous, mRNA-deficient β-thalassemia, it was found that whilst non-thalassaemic β-globin mRNA was stable during a 30min actinomycin D chase, 30%-75% of the thalassaemic mRNA molecules were degraded during that period (Maquat et al., 1981). It was concluded that the absence of β-globin mRNA in the patients was a result of rapid turnover of the thalassaemic β-globin mRNA molecules. This mechanism was generalised by Peltz et al. (1993) who stated that nonsense mutations in any gene can accelerate the decay rate of the mRNA transcribed from that gene, and termed this phenomenon nonsense-mediated mRNA decay.

The canonical NMD pathway is translation-dependent (Carter et al., 1996; Wu et al., 2020; Karousis et al., 2020) and occurs in the cytoplasm. The prevailing model is that NMD is triggered by a translating ribosome stalling at a PTC (Figure 1.6), however, recent work has shown that NMD activation is not accompanied by stable stalling of ribosomes at stop codons, and in fact, revealed similar

ribosomal occupancy at the stop codons of NMD-sensitive and NMD-insensitive reporter mRNAs (Karousis et al., 2020). The RNA helicase and ATPase UPF1 is the central NMD factor, which binds directly to the RNA being translated (Zünd and Mühlemann, 2013; Singh et al., 2019). Several mechanisms describing the circumstances under which UPF1 binds to RNA have been proposed. Initially UPF1 was thought to be recruited to the mRNA during translation, however UPF1 has since been found to bind lncRNAs (Li et al., 2017a; Zhou et al., 2019). Furthermore, it was originally suggested that UPF1 preferentially binds to the 3' untranslated region (3' UTR) but cross-linking immunoprecipitation (CLIP) studies have shown UPF1 bound along entire mRNAs when cells were treated with translation inhibitors, suggesting it binds non-specifically but transiently to the entire length of RNAs whilst unphosphorylated but is displaced by active translation (Zünd et al., 2013). In contrast, the NMD active phosphorylated UPF1 has been shown to bind predominantly de-adenylated mRNA decay intermediates (Kurosaki et al., 2018).

Translation stalls when the ribosome reaches the stop codon and is recognised by the translation termination factor eukaryotic release factor (eRF) 1 which has a similar structure to an aminoacyl-tRNA. eRF1 forms a complex with, and is activated by the GTPase activity of eRF3, allowing it to interact with UPF1 and later UPF2 (Kashima et al., 2006). Binding of UPF1, eRF1 and eRF3 to SMG1, a phosphatidylinositol-3-kinase-related kinase, forms the surveillance (SURF) complex (Isken and Maquat, 2008). SMG1-mediated phosphorylation of UPF1 is the key step in NMD-activation and is thought to stabilise the binding of UPF1 to the mRNA thereby determining the transcript for decay. Subsequently, UPF1 interacting with EJC components UPF2 and UPF3B induces the formation of the decay-inducing (DECID) complex. UPF1 is phosphorylated at multiple sites along its amino and carboxyterminal regions which allows it to recruit the homodimer endonuclease SMG6 and/or SMG5/SMG7 heterodimer, resulting in endocleavage and exonucleolytic decay of the fragments (Gatfield and Izaurralde, 2004; Huntzinger et al., 2008; Eberle et al., 2009) or promotion of de-capping and de-adenylation of the mRNA (Unterholzner and Izaurralde, 2004; Loh et al., 2013)

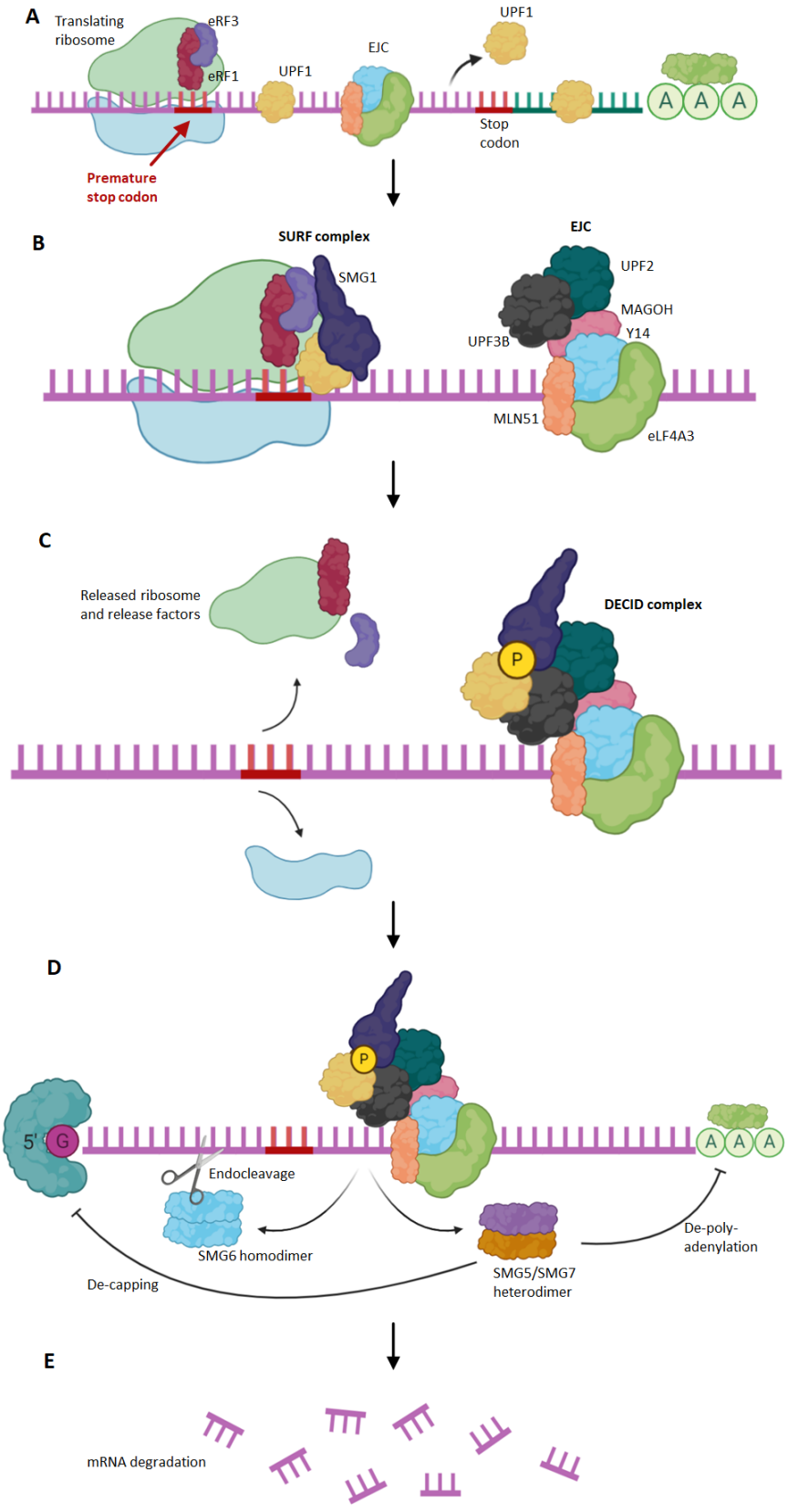


Figure 1.6: ***Canonical nonsense mediated decay pathway***

Figure 1.6: ***Canonical nonsense mediated decay pathway***

(A) *When the ribosome reaches a stop codon it stalls and release factors eRF1 and eRF3 bind. However when the stop codon is premature the ribosome does not dissociate in the usual manner. The RNA binding protein UPF1 usually transiently binds RNA in a dynamic manner.* ***(B)*** *The kinase SMG1 is recruited and facilitates the binding of UPF1 to eRF1 and eRF3 forming the SMG1-UPF1-Release-Factor (SURF) complex. SMG1 phosphorylates UPF1 at multiple sites along its amino and carboxyterminal regions, activating the NMD pathway.* ***(C)*** *Hyperphosphorylated UPF1 and SMG1 transfer to interact with two peripheral proteins of the EJC; UPF2 and UPF3B forming the decay inducing (DECID) complex.* ***(D)*** *The DECID complex formation recruits the SMG6 homodimer and/or SMG5/SMG7 heterodimer. The SMG6 homodimer performs endocleavage to break down the RNA molecule, whilst the SMG5/SMG7 heterodimer promotes de-capping and de-poly-adenylation, making the RNA vulnerable to standard degradation mechanisms.* ***(E)*** *After de-capping, de-poly-adenylation and/or endocleavage, the mRNA is degraded by the exosome and the 5'-3' exoribonuclease XRN1, and any nascent peptide degraded by the proteasome.*

respectively. UPF1 also recruits the de-capping complex itself either directly or via proline-rich nuclear receptor co-activator 2 (PNRC2) (Cho et al., 2009, 2015). Each of these processes reduces the stability of the transcript and targets the RNA for decay, via the 5'-3' exoribonuclease XRN1 and the exosome.

Whilst the NMD pathway core factors are conserved from yeast, more complexity and accordingly more factors are involved in higher order organisms. Seven genes (smg-1–7, for suppressor with morphological effect on genitalia) that are essential for NMD were originally identified in the nematode *Caenorhabditis elegans*, and orthologs of these genes have been found in several species. In *Saccharomyces cerevisiae*, there are three main NMD factors; UPF1, UPF2 and UPF3, that make up the conserved core of the NMD pathway. All three of these factors are trans-acting elements called up-frameshift (UPF) proteins. In mammals, UPF3 has two paralogues; UPF3B which forms part of the EJC and the DECID complex in NMD, and the NMD repressor, UPF3A which results in an increase in NMD activity when depleted (Shum et al., 2016).

There is also strong evidence supporting an NMD autoregulatory circuit, first noticed when mRNAs coding for NMD factors were found to be among NMD-sensitive transcripts (Huang et al., 2011). Many NMD factor mRNAs possess long 3' UTRs, and some of them harbour uORFs. Using reporter gene assays, it has been demonstrated that the long 3' UTRs of UPF1, SMG5, and SMG7 mRNAs are the main NMD-inducing features of these mRNAs (Yepiskoposyan et al., 2011). This negative feedback regulatory network has been identified as acting on all the core NMD factors, although to varying degrees and is conferred by different branches of the NMD pathway in a cell type-specific and developmentally regulated manner (Huang et al., 2011). Reversal of this feedback regulation in response to NMD perturbation is crucial for maintaining NMD (Goetz and Wilkinson, 2017), suggesting the feedback network maintains both RNA surveillance and the homeostasis of normal gene expression in mammalian cells.

1.2.3 NMD in human genetic disease

A pathway to specifically target PTCs is sensible given how abundant and potentially drastic the clinical consequences can be. Approximately 20% of human disease-associated single-nucleotide substitutions form a PTC (Mort et al., 2008) and one third of inherited disorders are caused by a PTC (Bhuvanagiri et al., 2010). PTC mutations can result in severe human genetic diseases such as variants of Duchenne and Becker muscular dystrophies (dystrophin), retinoblastoma (RB1), neurofibromatosis (NF1, NF2), ataxia-telangiectasia (ATM), Tay-Sachs disease (HEXA), cystic fibrosis (CFTR), Wilm's tumour (WT1), β thalassemia (β-globin), haemophilia A (factor VIII) and B (factor IX), von Willebrand disease (Willebrand factor), p53-associated cancers (p53), and numerous others (Holbrook et al., 2004). In some cases, the aberrant mRNA is efficiently targeted by the NMD pathway meaning any truncated nascent peptide is degraded by the proteasome and no protein is present, whilst in others the aberrant transcript escapes NMD meaning dysfunctional protein is produced.

Cystic Fibrosis (CF) can be caused by a plethora of mutations within the CFTR gene leading to a non-functional Na+ ion channel. Class I mutations are those that impair protein production, often nonsense mutations that generate premature stop codons triggering NMD. Common mutations in class I include G542X (the second most common CF-causing mutation), R1162X and W1282X, which accounts for 48% of the mutant alleles among Ashkenazi Jew CF patients. NMD degrades most if not all PTC-containing mRNAs, meaning little to no CFTR protein is produced. If the ribosome does terminate correctly then the resulting protein is truncated and has either reduced or no function. However, some truncated proteins retain at least partial function (Crawford et al., 2000; Kerr et al., 2001) and the amino-terminal portion of CFTR has been shown to form a partially functional Cl- channel (Sheppard et al., 1994). Aminoglycoside antibiotics can suppress premature termination codons, thus permitting translation to continue to the normal end of the transcript. Therefore, the common aminoglycoside antibiotic gentamycin has been administered to CF patients to promote in vivo

read-through of nonsense mutations (Wilschanski et al., 2003). The drug forces the ribosome past the PTC, shielding the nascent peptide from NMD and leads to expression of almost full-length proteins. Whilst these resulting ion channels are altered and have reduced function, they are beneficial to the patient compared to complete absence of protein. Patients carrying stop mutations showed a significant reduction in nasal potential difference upon treatment with gentamycin (Wilschanski et al., 2003), indicating an increase in ion transport, in part due to CFTR function (Rowe et al., 2011).

Duchenne muscular dystrophy (DMD) is also characterised by production of a partial protein which would likely reduce the symptoms of the disease, but the mutant RNAs harboring PTCs are degraded by NMD thereby aggravating the disease. DMD is the most common type of muscular dystrophy with around 10%-15% of cases caused by a point mutation introducing a premature stop codon (Roy and Griggs, 2021). The drug ataluren binds to the ribosome and impedes the recognition of the stop codon, thus promoting read through of the nonsense mutation to produce a full-length functional protein (Siddiqui and Sonenberg, 2016). Ataluren received conditional approval from the European Medicines Agency to treat DMD patients aged five years or older with a nonsense mutation in August 2014, but the most recent phase III trial did not reach statistical significance at the primary endpoint (McDonald et al., 2017).

However, read-through drugs are not the answer to all diseases involving NMD. For this to be an effective strategy, the truncated protein must be at least partially functional, and not harmful. β-thalassemia refers to a group of inherited blood disorders in which there is reduced or absent synthesis of the β chains of haemoglobin. Mutations in both copies of the HBB gene cause the more severe thalassemia major or thalassemia intermedia, whereas patients with heterozygous HBB mutation usually do not have symptoms. The common form of β-thalassemia occurs in homozygotes who possess NMD-competent PTCs in both copies of HBB and the resulting defective β-globin mRNA is degraded by NMD (Kugler et al., 1995). Heterozygous carriers of a single NMD-competent PTC generally produce enough β-globin from the normal allele, however rare forms of

dominant β-thalassemia, are caused by NMD-incompetent PTCs, usually within the last exon of the β-globin gene (Hall and Thein, 1994). These PTCs give rise to a large amount of truncated β-globin that cannot be sufficiently degraded (Thein et al., 1990). This example indicates the protective effect NMD fulfils in preventing many healthy heterozygotes from manifesting clinical disease.
Other NMD-associated phenotypes are due to altered or dysfunctional NMD factors. Mutations in the NMD gene UPF3B cause intellectual disability in humans and are associated with neurodevelopmental disorders, including schizophrenia and autism (Tarpey et al., 2007; Nguyen et al., 2014). Furthermore, copy number variations in several NMD genes have been associated with human neurodevelopmental disorders and other neural diseases (Nguyen et al., 2013, 2014; Johnson et al., 2019).

1.2.4 Global regulation of the transcriptome by NMD

Technologies such as northern blot analysis or qRT-PCR only allow low=throughput measurements of mRNA levels requiring pre-selection of likely candidates or reporter constructs. Therefore, increasingly, genome-wide analyses have been utilised to identify all targets of NMD. This has allowed the discovery that NMD is associated with the regulation of a larger proportion of the transcriptome than first thought. As well as RNA surveillance, NMD has been described as a regulator of normal gene expression through its ability to degrade a subset of functional mRNAs. This conclusion has been drawn from NMD factor knockout and knockdown experiments that result in between 3-20% of transcripts in eukaryotes increasing in expression, suggesting they are regulated by NMD (Peccarelli and Kebaara, 2014; Lykke-Andersen and Jensen, 2015). This additional function of NMD has been identified in all species analysed, including *S. cerevisiae* (Lelivelt and Culbertson, 1999; He et al., 2003; Kawashima et al., 2014), *C. elegans* (Ramani et al., 2009; Longman et al., 2013), *D. melanogaster* (Rehwinkel et al., 2005; Chapin et al., 2014), *D. rerio* (Longman et al., 2013) and *H. sapiens* (Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011; Longman et al., 2013). In humans up to 10% of normal transcripts are thought to be affected in

this way, suggesting NMD can fine tune many physiological processes such as development, cell cycle, cell stress and tumourigenesis by controlling the expression level of specific groups of RNAs (Mendell et al., 2004; Wang et al., 2011a; Lou et al., 2016). This would mean NMD is important for both prevention of disease and also homeostatic gene regulation.

NMD has been found to be a multifaceted pathway with non-canonical mechanisms including UPF2-independent and EJC-independent quality control pathways (Bühler et al., 2006; He and Jacobson, 2015). mRNAs harbouring long 3' UTRs have been shown to be sensitive to NMD, irrespective of the presence of an EJC (Hogg and Goff, 2010). There has even been evidence that NMD can affect lncRNAs (Li et al., 2017a; Zhou et al., 2019) which raises the question if these supposedly non-coding RNAs are being translated or if NMD is working on them in a non-co-translational manner.

In contrast to most known transcripts, T-cell receptor β (TCRβ) transcripts are downregulated by PTCs closer than 50nt from the final EJC (Carter et al., 1996; Wang et al., 2002), raising the possibility that an entity other than the classical EJC can serve as the second signal for TCRβ NMD. Depletion of the NMD factor and EJC component UPF3B, had no effect on stability of T-cell receptor β transcripts containing PTCs. Furthermore, WT genes have been identified that transcribe NMD-targeted mRNAs whose downregulation is hardly or not affected by depletion of UPF3A or UPF3B (Chan et al., 2007b).

It has been suggested that NMD preferentially takes place during the pioneer round of translation (Ishigaki et al., 2001; Maquat et al., 2010a,b), and that a transcript that survives this is then immune to NMD. This is partly due to the EJC being removed in the pioneer round of translation as the translating ribosome moves along the transcript (Lejeune et al., 2002). It is also argued that the cap binding heterodimer CBP80-CBP20, which is replaced with eIF4E after the pioneer round, plays a critical role in NMD because CBP80 interacts directly with UPF1, enhancing the efficiency of the NMD pathway (Isken et al., 2008). However, this model is debated (Durand and Lykke-Andersen, 2013; Rufener and Mühlemann, 2013) and recent findings show that NMD occurs with equal

probability during each round of translation of an mRNA molecule (Hoek et al., 2019). Using a novel real-time visualization of both mRNA translation and NMD of single mRNA molecules in living cells, Hoek et al. produced real-time kinetic measurements of degradation of the 3' mRNA cleavage fragment generated by NMD. They also showed that the probability of NMD occurring is variable and suggest it depends on the exon sequence downstream of the PTC, the PTC-to-intron distance, and the number of introns both upstream and downstream of the PTC. These results call into question the long-accepted link between NMD activation and the EJC since all EJCs are thought to be removed in the pioneer round of translation so would therefore not be present in later rounds.

1.2.5 Target characteristics

Whilst a PTC can result in a truncated protein and potentially cause severe phenotypes (Section 1.2.3), not all PTCs trigger degradation by NMD. The efficiency of NMD varies according to several characteristics, impacting the clinical outcome of genetic mutations. Furthermore, residual levels of many PTC-containing transcripts can be detected in steady-state measurements (Cheng and Maquat, 1993; Cheng et al., 1994; Belgrader et al., 1994; Thermann et al., 1998; Trcek et al., 2013; Boehm et al., 2014; Lindeboom et al., 2016).

The amount of residual mRNA for an NMD substrate appears to vary depending on the gene and the position of the PTC (Lindeboom et al., 2016; Thermann et al., 1998; Cheng et al., 1994). Activation of the NMD pathway is much less likely if the PTC is in the last exon, or less than 50nt from the next exon-exon junction (Nagy and Maquat, 1998). This is commonlyrefered to as the "50nt rule" It is thought this is necessary to allow the NMD machinery to interact with the proteins of the EJC to modulate the SURF complex into the DECID complex. Comparing the NMD efficiency of PTCs located upstream of the last EJC and PTCs inside the last exon, NMD is significantly more inefficient in the latter case

(84.7% versus 1.7% NMD efficiency (Lindeboom et al., 2016). This seems at odds with the finding that show that NMD occurs with equal probability during latter rounds of translation when the EJC would have presumably been removed (Hoek et al., 2019).

Using paired tumour:normal exome data and tumour transcriptomics from the cancer genome atlas (TCGA), Lindeboom et al. (2016), identified 2,840 high-confidence PTCs and investigated their expression using the RNA-sequencing. By normalising to the median expression of the WT transcript across all cancers without the PTC variant, they calculated an "NMD efficiency" score for each PTC and then measured how important each of the previously published NMD target rules was for determining if a PTC would trigger NMD. They found that the most variability could be explained by whether the PTC was in the last exon (45.8%) and the second highest amount was how close the PTC was to the start codon (17.2%). The 50 nucleotide rule often thought to be the most important accounted for only 3.5% of the additional variance after the last exon rule was taken into account. Overall, 74.2% of the variability in NMD response to PTC could be explained by a combination of these rules plus exon length, mRNA half-life, allele frequency, significant motifs and the distance from the PTC to the normal stop codon. They validated these findings using heterozygous germline nonsense variants in lymphoblastoid cell line RNA-sequenceing from healthy individuals (Geuvadis study; Lappalainen et al., 2013).

Furthermore, although the first mRNA substrates identified for NMD encoded PTCs (Losson and Lacroute, 1979; Maquat et al., 1981), it is now appreciated that NMD degrades a variety of mRNAs that arise as a result of either mutations or defects in pre-mRNA processing (Section 1.1.2) (Figure 1.7A), many with EJC-independent mechanisms. NMD that occurs independent of a downstream EJC is often as a consequence of translation termination at the normal termination codon rather than at a PTC (Wang et al. 2002; Bühler et al. 2006; Matsuda et al. 2007; Eberle et al. 2008). Cells define whether or not a termination codon triggers either type of NMD depending on the ability of UPF1 to be recruited and phosphorylated (He and Jacobson 2015; Karousis et al. 2016; Kurosaki et

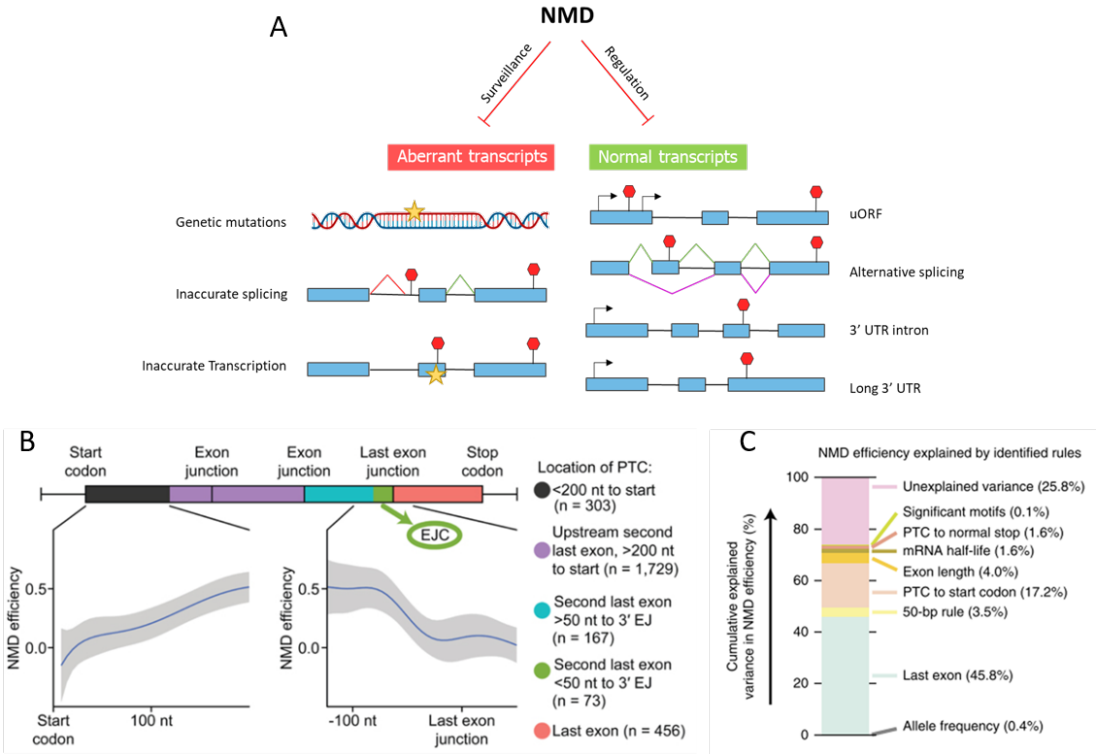


Figure 1.7: ***NMD target characteristics***

al. 2019). The common theme is that substrates of NMD present non-standard spatial relationships between the termination codon and other mRNA features (Baker and Parker, 2004). Features that trigger NMD include alternative splicing leading to inclusion of "poison" exons which contain a PTC (Lareau et al., 2007; Ni et al., 2007; Kurosaki et al., 2019a), mRNAs containing introns due to defective splicing (Sibley et al., 2016; Wang and Aifantis, 2020), pseudogenes (Weischenfeldt et al., 2008; Gatica and Rosa, 2016; Muir et al., 2018), alternative up-stream open reading frames (uORFs) in the 5' UTR (He et al., 2003; Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011), long 3' UTRs (Muhlrad and Parker, 1999; Yepiskoposyan et al., 2011) and the presence of introns in the 3' UTR (Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011). It is likely, these features cause a transcript to engage the NMD machinery due to the continued presence of EJCs or other RBPs downstream of a terminating ribosome.

Figure 1.7: ***NMD target characteristics***

(A) *The NMD pathway targets both aberrant transcripts that could lead to disease phenotypes and also normal transcripts. Aberrant transcripts can occur due to mutations or damage on the DNA, aberrant splicing leading to retained PTC-containing-introns or inaccurate transcription leading to accidental incorporation of a PTC. Targeting of normal transcripts has been identified in the case of a 5' upstream open reading frame (uORF), alternative splicing that leads to inclusion of a "poison" exon that contains a PTC, presence of a 3' UTR intron and when transcripts have a long 3'UTR (Figure adapted from Goetz and Wilkinson, 2017).* ***(B)*** *Schematic with color-coded gene regions showing different NMD efficiencies; boxes represent exons. Plots below show NMD efficiency trends (individual data points not shown, 95% confidence intervals shown in grey). Bottom left; increase in NMD efficiency across the first 200 coding nucleotides. Bottom right; variation in NMD efficiency around the last exon junction, the presumed location of the EJC is demarcated at the top of the plot (Figure from Lindeboom et al., 2016).* ***(C)*** *The amount of NMD efficiency variance explained by the identified rules for NMD target discrimination according to an integrated model incorporating multiple rules. The total variance explained is 74.2%. Each identified rule explains different amounts of variance in NMD efficiency between PTCs. The predictive power of each feature was determined by sequentially introducing the features to random forest models, on the basis of the order of the features suggested by an initial analysis of the entire feature set (Figure from Lindeboom et al., 2016).*

The long 3' UTR model postulates that the failure of polyA-binding protein (PABP) to interact with the terminating ribosome results in inefficient termination which triggers NMD. Based on current knowledge, a termination codon is recognised as "normal" because of the interactions between the terminating ribosome complex and the proteins that are bound to the 3' UTR and the poly-adenine (poly-A) tail. Specifically, Singh et al. (2008) showed that if the interaction between the cytoplasmic poly-A binding protein PABPC1 and the release factors eRF3 is impaired, eRF3 is able to interact with the core NMD factor UPF1, thus triggering the NMD response. Therefore, if the ribosome stalls too far away from the 3' UTR, it cannot terminate properly. Accordingly, mRNAs with substantially extended 3' UTRs are targeted by NMD in yeast, human, drosophila and plant cells (Kebaara and Atkin, 2009). More recently, Wu et al. (2020) showed that readthrough efficiencies for a PTC increased in response to 3' UTR lengthening and suggest that inefficient termination is a trigger for NMD. This model explains previous observation that EJC-independent degradation occurs in nonsense mRNA having a long 3' UTR. However, others have found that mRNAs having long 3' UTRs were not enriched in direct UPF1 targets (Tani et al., 2012). Intron retention is a type of alternative splicing where an intron is not correctly spliced out of the pre-mRNA. If the intron that is retained contains stop codons or a shift in the reading frame, this could lead to mRNA degradation (Sibley et al., 2016; Wang and Aifantis, 2020). Using RNA-sequencing and exome data from cancer patients to identify single nucleotide variants (SNVs) that disrupt splicing, Jung et al. (2015) found that 97% and 86% of SNVs causing intron retention and intronic cryptic site activation, respectively, generated PTCs, suggesting why this alternative splicing is such a strong trigger for NMD. For this reason, intron-retained transcripts had been thought to be non-functional, however, advances in next-generation sequencing technologies have enabled the detection of numerous transcripts that retain introns, suggesting intron-retaining mRNAs play an essential conserved role in normal physiology and an emergent role in diverse diseases (Lareau et al., 2007; Wong et al., 2013, 2016).

Pseudogenes are non-functional genes or gene fragments that have accumulated nonsense mutations by genetic drift, making them likely targets of NMD. They have been identified as one of several structural classes of substrates targeted by the NMD pathway including transcripts encoded by transposable elements and bicistronic mRNAs, where the DNA codes for two gene products (He et al., 2003). Pseudogene RNA was found to be specifically unregulated in NMD-deficient *C. elegans* (Mitrovich and Anderson, 2005), indicating that NMD is required for the downregulation of pseudogene expression.

However, these features do not characterise all known NMD targets and transcripts meeting one or more of these rules have been found to escape NMD. For example, in transcripts with intron-bearing 3' UTRs, PTCs in the penultimate exon may also strongly evade NMD despite the presence of a downstream EJC (Lindeboom et al., 2016), suggesting that the EJC in 3' UTRs may be less potent in initiating NMD. Furthermore, not all mRNAs with uORFs are substrates for NMD (Rehwinkel et al., 2006). It has been proposed that some yeast uORF-containing transcripts harbour elements that stabilise their expression and allow them to escape NMD regulation (Ruiz-Echevarría and Peltz, 2000). In plants, the uORF length determines whether the transcript is going to be degraded by the NMD machinery (Nyikó et al., 2009). In mammals, the mechanism that allows some uORFs containing transcripts to evade NMD regulation has not yet been characterised (Rehwinkel et al., 2006). It has been suggested that some of the described NMD pathways and target characteristics are cell type specific. For example, purified spermatocytes from UPF2 knock out have long 3' UTR-dependent NMD compared to spermatocytes from control mice, and no detectable EJC-dependent NMD (Bao et al., 2016).

1.2.6 Location specific NMD

NMD is a co-translational process and therefore is likely to occur at all sites of translation. However, not all translation occurs in the cytoplasm. Local translation has been observed from bacteria and yeast to humans (Blower, 2013; Jung et al., 2014; Eliscovich and Singer, 2017; Bovaird et al., 2018). It is commonly in-

volved in the delivery of mature proteins to specific cellular compartments, while allowing local regulation. In mammals, mRNA localization is involved in cell polarization and motility, mainly through the localization of actin and related mRNAs at the leading edge (Lawrence and Singer, 1986), and it is also involved in axonal growth and synaptic plasticity of neurons (Van Driesche and Martin, 2018). Translation of mRNAs at specific sites may also be important for the assembly of protein complexes (Pichon et al., 2016) or to avoid the deleterious effects of releasing free proteins at inappropriate places (Müller et al., 2013).
Localised mRNAs are often subjected to spatial control of translation (Besse and Ephrussi, 2008). RNA localisation and local translation provide ways to direct the synthesis of proteins with spatial and temporal precision: at subcellular locations where they are needed or in response to the timing of extracellular cues (Jung et al., 2012). In the case of *ash1* mRNA in yeast and β-actin mRNA in human neurons, translation is repressed during transport and is activated at their final location by phosphorylation-dependent mechanisms (Hüttelmaier et al., 2005; Paquin et al., 2007). This spatial regulation of translation provides an additional layer of control ensuring that mRNAs are translated only at the desired location. The level of NMD in a particular cell can be modulated by the level of expression of the NMD repressor UPF3A (Shum et al., 2016). High levels of UPF3A in a cell, would strongly repress NMD and thus robustly stabilise NMD target transcripts, a subset of which could encode proteins critical for that tissue. Conversely, a tissue with low levels of UFP3A would be expected to have strong NMD, which would robustly destabilise NMD substrates that might otherwise cause aberrant events in that tissue. Both these examples could also be temporally regulated whether through development, or in response to stimuli. In mice, UPF3A mRNA levels have been found to differ in adult tissues, with particularly high expression in the testis (Zetoune et al., 2008). UPF3A protein is also abundant in the adult testis but is largely undetectable in other adult tissues. UPF3A mRNA and protein are both expressed in a stage-specific manner during spermatogenesis, with expression highest in spermatocytes, the stage of germ cell development in which meiosis occurs (Shum et al., 2016).

Most studies have investigated mechanisms that promote local translation, but NMD has been shown to limit the translation of a localised mRNA in a mechanism that operates with temporal and spatial specificity within the cell (Colak et al., 2013). Neurons, being highly polarised cells, are well suited for studying localised subcellular events and localised translation has been identified as a key process in neuronal growth (Brittis et al., 2002). NMD has been shown to regulate levels of the synaptic protein ARC (Giorgi et al., 2007), suggesting that NMD might be localised to individual synapses.

Processing bodies (P-bodies) are cytoplasmic granules primarily composed of translationally repressed mRNAs and proteins related to mRNA decay, suggesting roles in post-transcriptional regulation. Nonsense-mutation-containing mRNPs transit through P-bodies before undergoing NMD (Durand et al., 2007), however, the function of P-bodies in translational repression and/or mRNA decay remains contentious. It has been shown that P-body formation can be a consequence, rather than a cause, of RNA-mediated decay and repression events (Eulalio et al., 2007). Furthermore, activation of PTC readthrough has been shown to occur concomitantly with the appearance of cytoplasmic foci containing UPF proteins and mRNAs with nonsense mutations but lacking P-body markers (Jia et al., 2017).

Germ cells don't generally contain P-bodies but studies have found that several NMD factors are highly concentrated in chromatoid bodies (CBs), and there is little to no UPF1 in the cytosol, raising the possibility that CBs are sites of highly active NMD (Meikar et al., 2014; Fanourgakis et al., 2016). CBs were also found to have several characteristics in common with P-bodies (Kotaja et al., 2006) and have been shown to support NMD (Jia et al., 2017). Together, these observations lead to the intriguing model that CBs are, in some respects, the germ cell version of P-bodies and that both are sites of active NMD.

NMD activity is regulated in a stage-, tissue-, and cell type-specific manner, allowing it to accordingly alter the level natural target mRNAs (Karam et al., 2013; Lou et al., 2016). The evidence described here suggest NMD is also regulated at a subcellular level.

1.3 Neuroblastoma amplified sequence (NBAS) protein

1.3.1 Discovery of NBAS as a putative NMD factor

A genome-wide RNA interference (RNAi) screen for novel NMD factors in *C. elegans*, identified two novel proteins as involved in NMD named *smgl-1* and *smgl-2* (Longman et al., 2007). Transgenic worms carrying a GFP reporter with a PTC were produced and the expression of GFP-PTC reporter mRNA monitored by GFP fluorescence and semi-quantitative RT–PCR. The worms lack GFP expression since the PTC means the transcript is broken down. However, GFP expression is restored upon the inactivation of NMD by depletion of the positive control, core NMD factor *smg-2*. In the GFP-PTC strain, the level of reporter mRNA was significantly increased upon depletion of the NMD core factors *smgl-2* or *smgl-1*. WT GFP reporter was not unregulated upon depletion of either factor. Unlike previously identified smg NMD genes, both *smgl-1* and *smgl-2* were found to be essential for viability in *C. elegans*. Both nematode proteins are evolutionarily conserved and have functional homologues in humans; neuroblastoma amplified sequence (NBAS) and DExH-Box Helicase 34 (DHX34) respectively.

The human homologue of *smgl-1* is NBAS (neuroblastoma amplified sequence), previously known as NAG (Neuroblastoma Amplified Gene). The protein is so named because it was first identified as upregulated in some neuroblastoma cell lines (Wimmer et al., 1999; Frühwald et al., 2000). However, the increased NBAS expression resulted from a large-scale genomic amplification of chromosome 2 containing both NBAS and the oncogene N-Myc. It has since been suggested that only the 5' region of NBAS is co-amplified and no link between NBAS and oncogenesis has been found (Scott et al., 2003).

NBAS is a large protein of 273kDa (2371 amino acids; Figure 1.8A), encoded by a long gene of 52 exons. The protein is broadly expressed, with the highest levels in heart and skeletal muscle, retinal ganglion cells, epidermal skin cells, and leukocytes, and the lowest levels in liver, small intestine and thymus (Wimmer et al., 1999; Frühwald et al., 2000). The structure of human NBAS has not been resolved, but from the sequence, functional domains have been identified. The first 420 amino acids of the N-terminal form a β-propeller structure; a type of all-β protein architecture characterised by four to ten highly symmetrical blade-shaped β sheets arranged toroidally around a central axis (Chaudhuri et al., 2008). Protein-protein interactions can form with the top, bottom, central channel, and side faces of the β-propeller but they are also involved in a diverse set of functions (Chen et al., 2011). The β-propeller of NBAS is made up of WD40 repeats; a short structural motif which has been shown to be involved in coordinating multi-protein complex assemblies in other proteins. Therefore, it is highly likely that the N-terminal domain of NBAS is involved in formation of protein complexes. β-propellers have also been shown to function catalytically and even to bind DNA (Vandermarliere et al., 2009; Chuankhayan et al., 2010; Chen et al., 2011). NBAS also contains a sec39-like domain between 734–1355 amino acids, so called because of its homology with the sec39 protein in yeast which has been shown to be required for Golgi-ER retrograde transport (Sasvari et al., 2013). The C-terminal half of the protein is arranged in an α-solenoid; a protein fold composed of repeating α helix subunits, arranged in antiparallel fashion to form a superhelix (Fournier et al., 2013). α-solenoids are generally involved in protein-protein interactions. The rare combination of a β-propeller and an α-solenoid in the same protein has been previously identified in vesicle-coating and nuclear pore complex components, including the clathrin heavy chain, the α subunit of COPI complexes, and a subset of nucleoporins (Leksa et al., 2009; Brohawn and Schwartz, 2009; Devos et al., 2004). NBAS has also been predicted

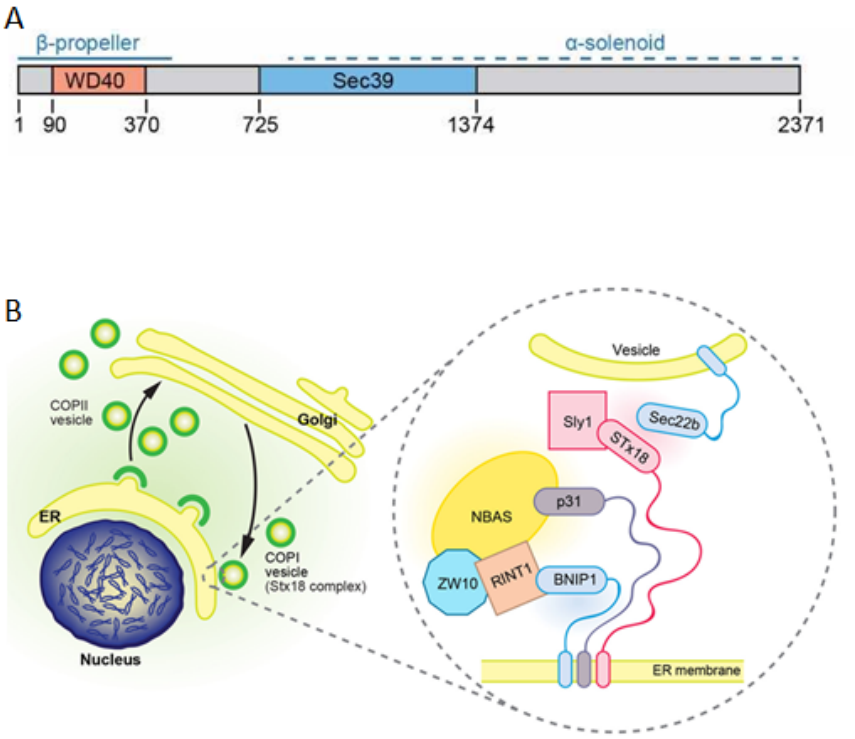


Figure 1.8: ***NBAS is located at the endoplasmic reticulum***

(A) *Depiction of the NBAS protein including it's annotated functional domains and structural regions.* ***(B)*** *NBAS is an endoplasmic reticulum (ER) localised protein with a role in ER secretion. Inset shows NBAS as part of the Syntaxin-18 complex, which facilitates Golgi-to-ER retrograde transport of COPI vesicles (Figure from Longman et al., 2020).*

to contains two leucine zippers and a ribosomal S14 motif, coiled coils and two or three transmembrane domains (Scott et al., 2003). Leucine zippers are protein-protein oligomerisation motifs (Landschulz et al., 1988), whereas the ribosomal S14 motif functions in the assembly of the ribosome (Chan et al., 1993).

NBAS protein is localised to the ER outer membrane. This is in contrast to most other identified NMD factors which are found throughout the cytoplasm. Notable exceptions are immunofluorescence studies that have shown yeast UPF3 is both nuclear and cytoplasmic (Moreno et al., 2013), and cellular fractionation experiments that have identified that the majority of UPF2 is located in the cytoplasmic perinuclear region and the remainder is detected in the intranuclear

region (Tatsuno et al., 2016). This is likely due to both UPF2 and UPF3B having a role as part of the EJC which is loaded onto mRNA within the nucleus during splicing and is only removed during translation outside of the nucleus. They are both also thought to have functions within the nucleus and when cytoplasm-localised (Shirley et al., 2002).

1.3.2 The role of NBAS in vesicle transport

NBAS has been described to function in Golgi-ER retrograde transport. Since there is no continuous membrane between the endoplasmic reticulum (ER) and the Golgi apparatus, membrane-bound transport vesicles are necessary to shuttle proteins between these two compartments. Vesicles are surrounded by coating proteins called COPI and COPII (Barlowe and Helenius, 2016; Gomez-Navarro and Miller, 2016). COPII targets vesicles to the Golgi apparatus and COPI marks them to be brought back to the rough ER. The rough ER works in concert with the Golgi complex to target new proteins to their proper destinations. The COPI coated vesicles leave the cis end of the Golgi complex and are actively transported on microtubules, most likely by dynein (Chen et al., 2005). The vesicles contain proteins found in the lumen of the Golgi complex that need to be transported to the lumen of the ER and transmembrane proteins which reside in the ER (Cosson and Letourneur, 1997; Stornaiuolo et al., 2003; Ma and Goldberg, 2013). This pathway also allows misfolded or aberrant proteins to be sent back to the ER for refolding and quality checking (Spang, 2013). Endocytosed proteins may travel from the plasma membrane via the Golgi to the ER via this pathway (Spang and Schekman, 1998). Retrograde transport is part of the organellar homeostasis pathway essential to maintaining the identity of the ER and of the Golgi apparatus.

When the vesicles reach the ER they encounter the tethering complex DSL1, also referred to as the Syntaxin-18 complex. Syntaxins are a family of soluble NSF attachment protein receptor (SNARE) proteins implicated in membrane fusion (Hong, 2005; Jahn and Scheller, 2006). Mammalian syntaxin members involved in transport between the ER and Golgi are Syntaxin-5 and Syntaxin-18,

which are localised in the Golgi and the ER, respectively (Dascher et al., 1994; Hatsuzawa et al., 2000). The presence of subdomains in the ER enables this organelle to perform a variety of functions. By mediating the fusion of retrograde membrane carriers with the ER membrane, Syntaxin-18 has an important role in ER subdomain organization (Iinuma et al., 2009). Syntaxin-18 interacts with three SNAREs (BNIP1, USE1 and SEC22B) and three peripheral membrane proteins (SLY1, ZW10 and RINT1) (Hirose et al., 2004; Nakajima et al., 2004). NBAS forms part of the NBAS-RINT1-ZW10 (NRZ) tethering complex which associates with the SNAREs Syntaxin-18, USE1, BNIP1 and SEC22B (Figure 1.8B) and is believed to play a role in assembly of the multi-SNARE complex (Hatsuzawa et al., 2000; Burri et al., 2003; Hirose et al., 2004; Tagaya et al., 2014). The homologous complex to NRZ in yeast, Dsl1, also comprises three subunits and has been shown using x-ray crystallography to form a two-legged structure with a central hinge (Travis et al., 2020).

NBAS has been shown to interact directly with RINT1 and ZW10 somewhere in its C-terminal region between 1036–2371 amino acids (Civril et al., 2010), although the direct interaction between NBAS and RINT1 has been disputed (Tagaya et al., 2014). Stability of NBAS appears to be linked to the stability of ZW10, so that the levels of both proteins, as detected by western blotting, were greatly decreased at 96h after siRNA transfection against ZW10 (Civril et al., 2010). Although there was no reciprocal reduction in the stability of ZW10 when NBAS was depleted. Both ZW10 and RINT1 have diverse cellular functions in addition to retrograde transport including endosome-to-Golgi transport, cytokinesis, cell cycle checkpoint and autophagy.

NBAS is structurally and evolutionarily related to a protein involved in mitotic chromosome segregation ROD (Rough Deal). Both have N-terminal WD40 β-propellers and sec39-like domains. ZW10 binds both ROD and NBAS but this is mutually exclusive (Civril et al., 2010).

1.3.3 Impact of NBAS on human health

Multiple studies have identified patients with compound heterozygous mutations in NBAS, mainly located throughout the two annotated domains of the protein, that cause several complex syndromes, including short stature, optic nerve atrophy, and Pelger-Huet anomaly (SOPH). SOPH is an autosomal recessive syndrome characterised by severe postnatal growth retardation, facial dysmorphism with senile face, small hands and feet, normal intelligence, abnormal nuclear shape in neutrophil granulocytes (Pelger-Huet anomaly), and optic atrophy with loss of visual acuity and colour vision (Maksimova et al., 2010; Lacassie et al., 2020).
Osteogenesis imperfecta (OI), the most common inherited bone fragility disorder, affects 1 in 15,000 live births resulting in frequent fractures and reduced mobility, with significant impact on quality of life. Whole exome sequencing in patients with OI identified, in two patients with a multi-system phenotype, compound heterozygous variants in NBAS (Balasubramanian et al., 2017). One of the homozygous missense variants found to cause SOPH syndrome was also found in one OI patient on one allele and a nonsense variant in the other allele.
Infantile liver failure syndrome 2 (ILFS2) is also caused by mutations affecting NBAS. ILFS2 is a life-threatening disorder of hepatic function that manifests with acute liver failure in the first few months of life. Clinical features include anaemia, renal tubulopathy, developmental delay, seizures, failure to thrive, and liver steatosis and fibrosis (Haack et al., 2015). These NBAS mutations could be effecting the liver specifically due its generally low wild type expression in hepatocytes (Wimmer et al., 1999; Frühwald et al., 2000), as reduction of functional protein would therefore have a bigger effect.
NBAS mutations have also been found in unnamed multisystem diseases affecting the liver, eye, immune system, connective tissue, and bone. Clinical manifestations include a progeroid appearance, short stature, slender bones, epiphyseal dysplasia with multiple phalangeal pseudo-epiphyses, cervical instability,

myelopathy, elevated transaminases, hypogammaglobulinemia, reduced natural killer cells, Pelger-Huet anomaly of granulocytes which describes an abnormal nuclear shape, and in some cases retinal dystrophy and optic atrophy (Segarra et al., 2015).

Whilst the evidence suggests that most of the NBAS mutations cause reduced level of NBAS protein, it is not currently known by what mechanism the mutations in NBAS lead to these phenotypes. They could be due to a lack of retrograde transport in the secretome pathways, or a break down in NMD regulation leading to production of truncated proteins, or a combination of these pathways.

1.3.4 Evidence for NBAS as an NMD factor

As well as *C. elegans* where it was first identified, NBAS has been described as an NMD factor in *Danio rerio* (zebrafish) and human cells (Anastasaki et al., 2011; Longman et al., 2013, 2020).

Zebrafish goldenb1 mutants carry a nonsense mutation in exon 5 of the slc24a5 mRNA. Expression of slc24a5 transcript was increased in melanocyte precursor cells when either UPF1 or NBAS depleted, whilst the WT allele was unaffected, suggesting NMD-mediated degradation (Anastasaki et al., 2011). In HeLa cells, depletion of NBAS by siRNA causes an increase in the level of mRNA of a β-globin NMD reporter containing a PTC. This increase can be rescued by the addition of exogenous overexpressed FLAG-NBAS (Longman D., unpublished; Figure 1.9A). This is surprising because β-globin is not a secreted protein and so may not be expected to be translated at the ER, and therefore not be affected by NBAS activity. However, as described in Section 1.1.4, translation localisation is fluid and many non-secretome proteins can be translated at the ER. Both of these experiments use change in level of mRNA as an indicator or changed NMD activity. Whilst this has been common practice for identifying NMD effects, it is not a true measure of NMD activity as NMD degrades the mRNA and therefore specifically affects mRNA stability, whereas global abundance changes could be due to other factors or processes. For the experiment in HeLa cells, an improve-

ment could be to co-transfect the PTC-containing reporter and a WT β-globin reporter, however introducing exogenous reporters and overexpressing protein from transfected plasmid both have caveats as they create a non-physiological state.

Significant co-regulation of endogenous transcripts was observed between core NMD factor UPF1 and NBAS in *C. elegans*, *D. rerio* and human cells (Longman et al., 2013). However the correlation was much stronger in zebrafish (r=0.74) than HeLa cells (r=0.55) or *C. elegans* (r=0.42; all Spearman's correlation tests). This could be because the zebrafish were targeted for knock down using morpholino (MO) which consists of antisense oligomers against the mRNA of interest and was commonly used in the absence of tools for efficient targeted mutagenesis in a range of models in the early 2000s. However, concerns have since arisen about the off-target effects of MOs (Stainier et al., 2017). MO-induced phenotypes are often more severe than those of corresponding genomic mutants, which could be the result of various phenomena including off-target effects of the MO (Kok et al., 2015) or genetic compensation in genetically mutated but not MO-injected animals (Rossi et al., 2015). It has been suggested that the major off-targeting effect is mediated through p53 activation (Robu et al., 2007). This could explain the much higher number of targets and stronger correlation between UPF1 effects and NBAS effects seen in zebrafish compared to *C. elegans* or human cells.

It was also shown that upregulation of transcripts encoding NMD factors in response to NMD abrogation included NBAS, suggesting that the conserved autoregulatory NMD circuit is mediated by NBAS in *C. elegans* and zebrafish in addition to HeLa cells (Longman et al., 2013). More recently, separation of function experiments have shown a block of constitutive secretion does not affect NMD activity (Longman et al., 2020), indicating that depletion of NBAS does not inactivate NMD simply by impairing ER secretion, leading to cell stress.

Transfected, overexpressed NBAS and UPF1 co-immunoprecipitate and the interaction between NBAS and UPF1 has also been shown by multiple techniques including Förster resonance energy transfer (FRET), fluorescence lifetime imaging microscopy (FLIM), fluorescence cross-correlation spectroscopy (FCCS) and

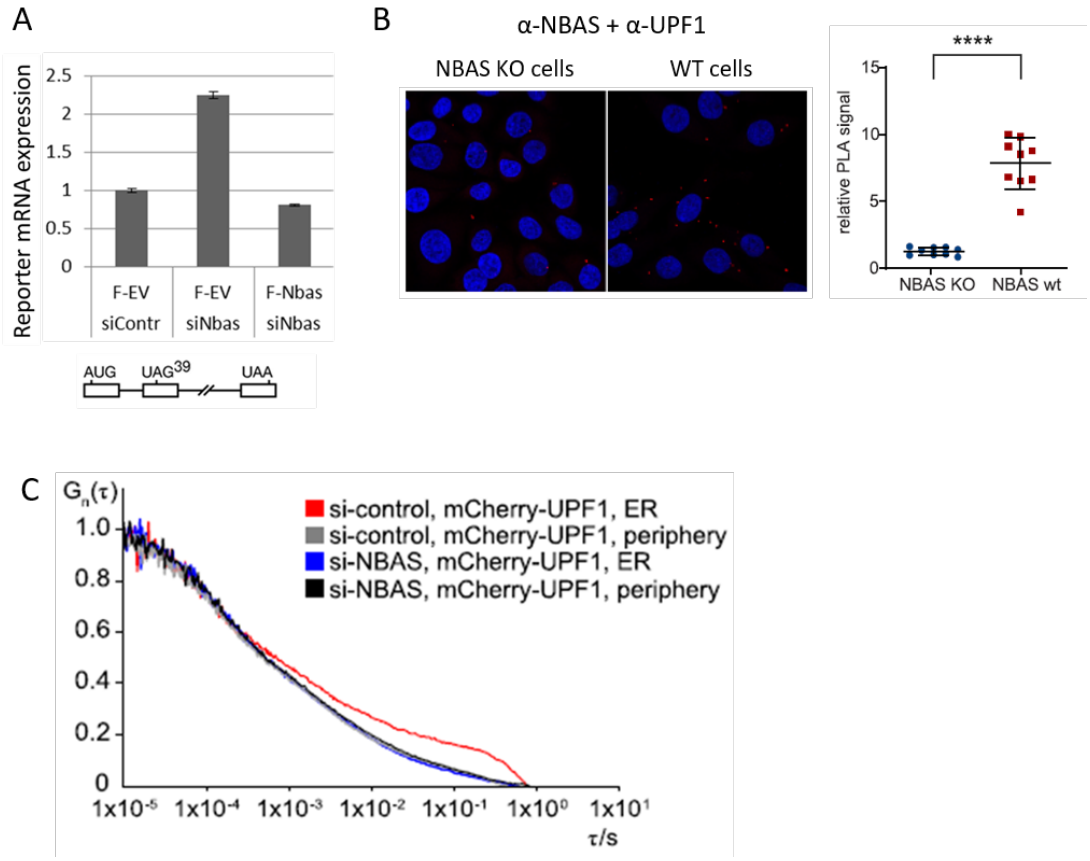


Figure 1.9: ***Evidence of NMD function of NBAS***

(A) *Level of a β-globin NMD reporter containing a PTC was increased in HeLa cells upon depletion of NBAS by siRNA. This was rescued to control levels by addition of exogenous FLAG-NBAS (unpublished data, Longman D.).* ***(B)*** *A direct interaction between endogenous NBAS and UPF1 was determined by PLA using antibodies against endogenous UPF1 and NBAS proteins in HeLa cells. The PLA signal in wild-type (WT) HeLa cells was compared with a HeLa NBAS knockout (KO) cells, as a negative control. The PLA signal was quantified in the graph where each point represents mean PLA count in one captured frame, relative to NBAS KO negative control. Significance was determined by two-tailed Mann-Whitney test p value* < 0.0001 *(Figure from Longman et al., 2020).* ***(C)*** *NBAS recruits UPF1 to the ER. Fluorescence correlation spectroscopy (FCS) produces average autocorrelation curves of a fluorescent protein in a given volume. The curve of mCherry-UPF1 is shown at the ER or cell periphery in the presence (si-control) or absence of NBAS (si-NBAS) (Figure from Longman et al., 2020).*

biochemistry techniques using overexpressed GFP-NBAS or FLAG-NBAS and proximity ligation assay (PLA) of endogenous proteins (Figure 1.9B). FRET quantifies the distance between two proteins by measuring the fluorescent signal given off by an acceptor molecule which is tagged to one protein of interest. The acceptor emits signal when it comes into close contact (typically 10-100Å) with a donor molecule that is tagged to the other protein of interest and is excited by laser during imaging (Liput et al., 2020). FLIM is one way of measuring FRET and is often carried out in conjunction. FLIM measures the decrease in lifetime of the donor molecule as FRET occurs. A shorter average lifetime indicates the donor is nearer the acceptor. A shorter average lifetime indicates the donor is nearer the acceptor. PLA is an antibody-based technique that indicates when two proteins are in close contact. It uses specific antibodies against the two proteins of interest and secondary antibodies with either a plus (+) or minus (-) DNA oligo attached. When the secondary antibodies are less than approximately 40nm apart, the oligos can anneal and rolling DNA synthesis can take place. These nascent DNA clusters can be visualised with fluorescent marker which is imaged as a red spot. However, this does not exclude the possibility that the two proteins may be within the same protein complex and not interacting directly. FCCS is an extension of fluorescence correlation spectroscopy (FCS) which measures fluorescence intensity fluctuations in a given volume. FCCS measures two labelled particles' movement through a predefined confocal volume to identify if they are moving together and therefore likely to be interacting. FCS of mCherry-UPF1 at the ER and at the cellular periphery show that UPF1 diffuses more slowly at the ER than in the cytoplasm (Figure 1.9C), suggesting it is tethered to the ER in some manner. Furthermore, upon depletion of NBAS with siRNA, the diffusion rate of UPF1 at the ER was increased to the same level as the cellular periphery, indicating it is likely NBAS is tethering UPF1 to the ER membrane.

NBAS has also been shown to interact with the NMD factors SMG5, SMG6, SMG7 and DHX34 by immunoprecipitation followed by western blot and probing with specific antibodies (Longman et al., 2020).

1.4 Endoplasmic reticulum stress

1.4.1 ER homeostasis

As touched on in Section 1.1.4, the ER has many roles within the cell: lipid biosynthesis, glucose homeostasis, a Ca^{2+} store, protein export, and contacting and regulating other organelles (Loi et al., 2018). ER homeostasis is therefore an important part of maintaining overall cell stability. Cells respond to internal and external cellular stressors by activating stress-response pathways to mitigate the effects and re-establish balance. Physiological, chemical, and pathological factors that compromise ER homeostasis lead to ER stress. If homeostasis is not achieved in a timely manner, the stress pathways trigger apoptosis to preserve organism integrity (Oyadomari et al., 2002; Chen et al., 2011; Wolff et al., 2014; Schneider and Bertolotti, 2015). Therefore, it is important for the cell to regulate ER homeostasis in order to maintain survival.

Many enzymes, foldases, and chaperones assist with protein folding and other post-translational modifications, which ensure the maintenance of ER homeostasis (Sicari et al., 2020). ER stress occurs when the balance between the protein folding demand exceeds the capabilities of the ER, leading to an abnormal accumulation of improperly folded proteins within the ER lumen (Schröder and Kaufman, 2005). Cells have evolved elaborate mechanisms to ensure the accuracy with which secreted, and membrane proteins are folded and assembled. The correct folding of the nascent peptide is necessary before the protein can be transported to the Golgi and localised to the correct cellular location. However, if a peptide is misfolded or cannot be folded at all, which could be due to the presence of a PTC causing the peptide to be truncated, this can lead to formation of aggregates within the ER lumen. Despite the multitude of different proteins recruited to stabilise folding intermediates and overcome energy barriers (Balchin et al., 2016), protein folding is inherently error prone, especially in the ER. 12–15% of

newly synthesised polypeptides in human cells are co-translationally eliminated by the ubiquitin-proteasome system (Duttler et al., 2013; Wang et al., 2013), however since at the ER, the nascent peptide is deposited directly into the lumen, misfolded proteins must first be exported to the cytoplasm.

Many pathologies are characterised by formation of aggregates, especially neurodegenerative diseases such as Alzheimer's Disease. These protein aggregates trigger ER stress (Zhu et al., 2020; Nilsberth et al., 2001). It has also been shown that ER overload of pancreatic β cells in conditions such as hyperglycaemia, obesity, and long-term treatment with sulfonylureas cause ER stress, which leads to apoptosis and can result in type 2 diabetes (Oyadomari et al., 2002).

To counteract protein misfolding and maintain protein homeostasis in the secretory pathway, multiple quality control mechanisms have evolved (Figure 1.10), including ER-associated degradation (ERAD), ER-phagy and the Golgi quality control pathway (GQC). Each process is regulated to some degree by the unfolded protein response (UPR) and/or the heat shock response, which adjust protein synthesis, chaperone levels, and the activity of protein degradation pathways (Travers et al., 2000; Bravo et al., 2013; Li et al., 2017b,c; Preissler and Ron, 2019; Karagöz et al., 2019).

1.4.2 Mechanisms of ER stress

Considered the first line of defence, misfolded and damaged proteins, as well as surplus single subunits of multi-subunit proteins are recognised by the ER-associated protein degradation (ERAD) pathway (Roth and Zuber, 2017; Vembar and Brodsky, 2008). These proteins are dislocated to the cytoplasm and eventually degraded by the canonical ubiquitin–proteasome system (Bonifacino and Weissman, 1998; Dikic, 2017; Grumati and Dikic, 2018; Varshavsky, 2017). The ERAD has a conserved mechanism: misfolded proteins are recognised and ubiquitinated through either the ERAD-L (ERAD of substrates with misfolded lesions within the ER lumen), ERAD-C (cytoplasm), or ERAD-M (membrane) pathways, depending on the location of the folding lesion (Huyer et al., 2004; Vashist and Ng, 2004; Carvalho et al., 2006; Denic et al., 2006).

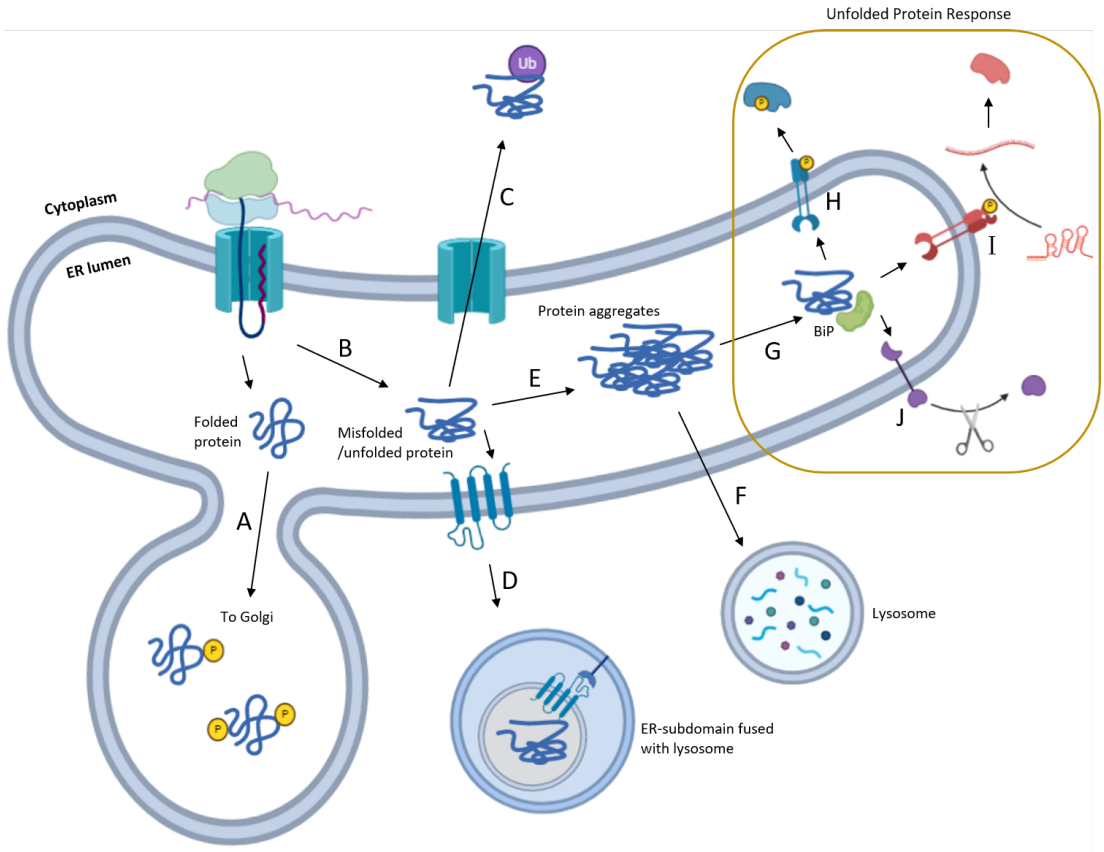


Figure 1.10: ***Mechanisms of endoplasmic reticulum stress***

Figure 1.10: ***Mechanisms of endoplasmic reticulum stress***

Translation at the SEC61 translocon leads to the nascent peptide being deposited directly into the ER lumen. ***(A)*** *If the protein is folded correctly then post-translational modifications can be added and the protein continues to the Golgi.* ***(B)*** *However if the peptide can not be folded or is misfolded then then quality control mechanisms are activated to remove the peptide and prevent ER stress.* ***(C)*** *The first line of defence is ER-associated degradation (ERAD), which dislocates proteins to the cytoplasm, possibly by retrotranslocation through the SEC61 channel, where they are ubiquitinated and degraded.* ***(D)*** *Proteins that are too large for the translocon channel or are otherwise ERAD-resistant, are bound by ER-phagy receptors which then produce ER subdomains that pinch off from the ER network and fuse with lysosomes to facilitate degradation.* ***(E)*** *If neither ERAD or ER-phagy controls the level of misfolded protein they may begin to form aggregates and activate ER stress.* ***(F)*** *Protein aggregation can itself be sufficient to trigger lysosomal degradation.* ***(G)*** *Upon accumulation of unfolded proteins, the chaperone BiP dissociates from the three unfolded protein response (UPR) sensors and binds the aberrant proteins instead. This activates the three branches of the UPR;* ***(H)*** *eukaryotic translation initiation factor 2 alpha kinase (PERK),* ***(I)*** *inositol requiring enzyme 1 (IRE1) and* ***(J)*** *activating transcription factor 6 (ATF6). Via three different mechanisms all three lead to block of translation, autophagy and activation of transcription factors that upregulate genes involved in antioxidant response, increased folding capacity and apoptosis.*

Since ubiquitin ligase activity occurs in the cytoplasm, ERAD substrates with lumenal lesions (those handled by the ERAD-L pathway), must be retrotranslocated (Schoebel et al., 2017). Based on its function as the ER translocation channel, SEC61 has been suggested to also act in reverse and function as a retrotranslocon. Evidence supporting this hypothesis includes experiments demonstrating interaction between SEC61 and ERAD substrates, as well as the proteasome (Wiertz et al., 1996; Kalies et al., 2005; Schäfer and Wolf, 2009). In addition, Apolipoprotein B is retrotranslocated and targeted for degradation when forward translocation through SEC61 is interrupted (Mitchell et al., 1998), indicating that the two functions for this channel, translocation and retrotranslocation, can be uncoupled. Other studies have also suggested the HRD1 ubiquitin ligase (Baldridge and Rapoport, 2016) and multi-spanning membrane protein DER1 (Mehnert et al., 2014) promote movement of misfolded proteins through the ER membrane. Retrotranslocated membrane proteins contain hydrophobic transmembrane domains and are consequently aggregation prone (Wang et al., 2011b).

In addition to misfolded proteins, some native proteins deemed unnecessary are degraded by the ERAD pathway. For these substrates, adaptor protein binding, ligand binding, and/or the display of a degron is essential. In this way, the ERAD pathway can be hijacked by pathogens. The human cytomegalovirus encodes two membrane adaptors, US2 and US11, which bind newly synthesised major histocompatibility complex (MHC) class I molecules and recruit the ERAD machinery (Wiertz et al., 1996; Stagg et al., 2009). Cell-surface display of MHC class I is reduced, thereby disabling the host immune response. A similar strategy is used by the HIV-encoded adaptor Vpu, which directs CD4 to the proteasome (Fujita et al., 1997; Schubert et al., 1998).

While most misfolded proteins in the ER are efficiently handled by the ERAD pathway, the diameter of the pore in the retrotranslocon may preclude misfolded oligomeric or aggregated proteins from being exported from the ER. Because these substrates are toxic, the ER is equipped with another mechanism to protect homeostasis; ER-phagy (Bernales et al., 2006, 2007). ER-phagy also disposes of

ERAD-resistant misfolded proteins (Fregno and Molinari, 2019). In yeast, atg40 localises primarily to the cytoplasmic ER, resulting in autophagic sequestration of ER subdomains. Expression of an aggregation-prone ERAD substrate induces the expression of an alternate COPII subunit, lst1, which functions with Atg40 to select ER-phagy sites. An lst1 homologue, SEC24C, similarly marks ER-phagy sites in mammals (Cui et al., 2019). The discovery of various ER-phagy receptors in mammals highlights the potential for substrate selection (Khaminets et al., 2015; Fumagalli et al., 2016; Grumati et al., 2017; Smith and Wilkinson, 2018; An et al., 2019; Chen et al., 2019; Chino et al., 2019). Each receptor contains a domain that facilitates interaction with an autophagy-requiring factor. Even though the identification of these ER-phagy receptors has helped define the nature of cargo selection, how the ER-phagy machinery degrades most receptor-bound cargo proteins remains an important open question (Wilkinson, 2019).

SEC62 is an ER transmembrane component of the SEC61/SEC62/SEC63 translocation machinery involved in the import of newly synthesised proteins into the ER lumen (Section 1.1.4). SEC62 was identified as the ER-phagy receptor that regulates ER turnover after conclusion of a transient ER stress. This function is independent of its role during protein translocation (Fumagalli et al., 2016). SEC62-regulated "recovER-phagy" resets ER volume and content after ER stress and employs the autophagy core machinery.

The Golgi apparatus is the organelle encountered by secretory proteins after leaving the ER and is where post-translational modifications occur including glycosylation, acetylation, phosphorylation as well as proteolytical activation. The Golgi also serves a quality control function (GQC) (Arvan et al., 2002; Potelle et al., 2015). Misfolded proteins that escape ERAD, are recognised in the Golgi and either delivered to the lysosome for degradation or are proteasome-targeted, requiring retrograde transport to the ER (Vashist et al., 2001, Section 1.3.2). For example, misfolded proteins with mutated transmembrane domains are specif-

ically returned to the ER from the Golgi via a retrieval receptor, RER1 (Letourneur and Cosson, 1998; Sato et al., 2003). COPI vesicle transport has been shown to play an important role in retrieving unassembled T cell receptor α subunit (TCRα) to the ER (Yamamoto et al., 2001).

Targeting for lysosomal degradation requires the sortilin receptor (for misfolded lumenal/soluble proteins) or a ubiquitin ligase complex (for misfolded membrane proteins). Sortilin resides primarily in the Golgi and plasma membrane and mediates lysosomal cargo sorting (Nielsen et al., 1999, 2001; Lefrancois et al., 2003; Amengual et al., 2018). While protein aggregation is sufficient to trigger lysosomal degradation, trafficking is sortilin independent, suggesting that protein aggregation is necessary, but not sufficient, for sortilin-mediated sorting.

When ERAD is unable to maintain ER homeostasis, the unfolded protein response (UPR) is induced. The UPR is a highly conserved ER stress response, that consists of a set of intracellular signalling pathways. One of the most abundant proteins within the ER is the Hsp70-type chaperone, BiP (binding-immunoglobulin protein; GRP-78). BiP binds to nascent polypeptide chains initially to prevent their aggregation and has a role in correct peptide folding. Under homeostatic conditions, BiP also binds the luminal domain of the UPR transmembrane sensors, keeping them inactive. However, upon accumulation of unfolded proteins, BiP preferentially binds to these in an attempt to refold them, releasing the sensors and allowing activation of the three branches of the UPR. The three ER transmembrane sensors are inositol requiring enzyme 1 (IRE1), eukaryotic translation initiation factor 2 alpha kinase (PERK) and activating transcription factor 6 (ATF6). The UPR functions to restore normal function of the cell by global translational repression, expansion of the ER membrane network, activating the signalling pathways that lead to increasing the production of molecular chaperones and folding catalysts involved in protein folding, and enhancement of ERAD, ER-phagy, and GQC (Travers et al., 2000; Yoshida et al., 2003, 2006; Smith and Wilkinson, 2017; Smith et al., 2018; Karagöz et al., 2019).

Together, these events increase the folding capacity of the ER, limit the entry of newly synthesised proteins into the ER, and facilitate the clearance of those that have accumulated; however, if the stress is not alleviated, the UPR triggers cell death.

IRE1 is the most conserved transducer of the UPR and can either lead to adaptive or death signals. This occurs through both unconventional splicing of XBP1 mRNA and regulated IRE1-dependent decay of mRNA (RIDD) (Hollien and Weissman, 2006). Whereas XBP1 mRNA splicing is cytoprotective in response to ER stress, RIDD can either preserve ER homeostasis or induce cell death. RIDD involves IRE1 RNase activity, which cleaves ER-localised RNA in a sequence-specific manner at XBP1-like consensus sites. The free 5' and 3' ends of the generated mRNA fragments are then substrates for cellular exoribonucleases leading to their rapid degradation. Evidence supports altered IRE1 function in various diseases such as diabetes, inflammatory and neurodegenerative disorders and, more generally, in diseases associated with proteostasis imbalance (Hetz and Glimcher, 2011; Hetz et al., 2011). PERK phosphorylates eIF2a which leads to activation of the transcription factors ATF4 and CHOP which then upregulate several apoptotic genes, whilst part of the ATF6 receptor self-cleaves and acts as a TF on similar genes. All three mechanisms lead to block of translation, autophagy and activation of transcription factors that upregulate genes involved in anti-oxidant response, increased folding capacity and apoptosis.

While physiologically beneficial, the UPR and other ER stress pathways require tight regulation to provide a beneficial outcome and avoid lethal consequences. Chronic activation of the UPR is linked to neurodegeneration (Yan et al., 2019; Torkzaban et al., 2020; Li et al., 2020), malignancy (Giampietri et al., 2015; Rodvold et al., 2016), prion diseases (Shah et al., 2017) and viral infection (Blázquez et al., 2014; Fung and Liu, 2014; Jheng et al., 2014; Torkzaban et al., 2020). In these circumstances, cellular stress pathways can contribute to disease. Therefore, the inhibition or modulation of the UPR represents a potential treatment for several diseases and conditions.

1.4.3 Effect of NMD on ER stress

Several studies have demonstrated that NMD serves as a regulator of the UPR pathway (Figure 1.11) by preventing inappropriate UPR activation and promoting the timely termination of the UPR to avoid apoptosis. NMD also regulates responses to non-ER stressors, including hypoxia (Usuki et al., 2019), amino-acid deprivation (Wengrod et al., 2013), and pathogen infection (Ajamian et al., 2008). NMD regulates stress responses in species across the phylogenetic scale (Goetz and Wilkinson, 2017), suggesting that it has conserved roles in shaping stress responses.

A link between UPR regulation and RNA stability was first suggested when ATF3 and ATF4 mRNA were found to be stabilised when NMD was suppressed (Mendell et al., 2004), suggesting both transcripts are targets of the NMD pathway. ATF3, which is downstream of the PERK branch of the UPR, regulates several genes critical for responding to cellular stress. ATF4 also acts downstream of PERK and activates the transcription of genes encoding other UPR pathway factors, including the ATF3 gene. Because NMD represses both factors, it could dramatically downregulate the magnitude of the UPR. This was further supported by the finding that the magnitude of the UPR was increased when the NMD factor UPF1 was depleted (Gardner, 2008). The level of several UPR factors, including ATF3, ATF4, CHOP, and GADD34, in the PERK and ATF6 branches of UPR, were increased in the absence of UPF1 whereas overexpression of UPF1 decreased expression of ATF3, ATF4 and CHOP, indicating that hyperactivation of the NMD pathway suppresses UPR magnitude.

Microarray analysis identified several candidates of other UPR mRNAs might be targeted by NMD (Karam et al., 2015). Depletion of the core NMD factor UPF1 and the branch-specific factor, UPF3B led to upregulation and stabilisation of several UPR components, indicating that they are likely targets of the UPF3B-dependent branch of NMD. This included the previously identified ATF3 and ATF4, and also components of the other two UPR branches; IRE1α, TRAF2, FSD1L, and TNRC5. In the same study, they found that HeLa cells depleted of

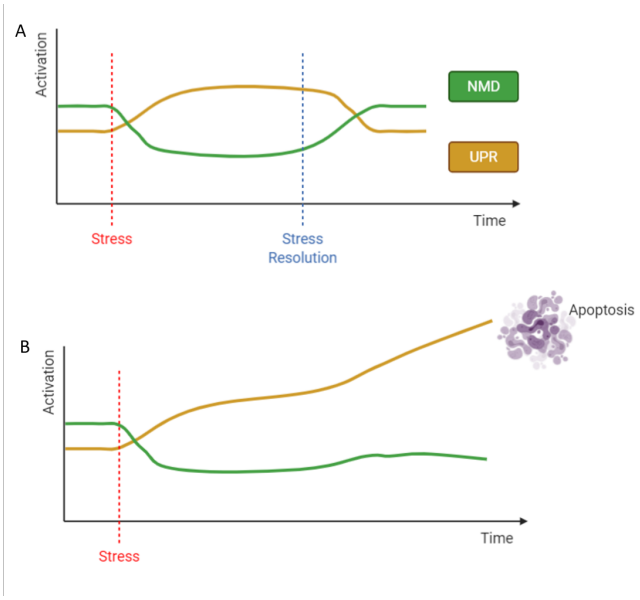


Figure 1.11: ***The interplay between NMD and the UPR ER stress pathway***

(A) *In basal conditions, NMD constantly represses the UPR via the degradation of the mRNA of several components involved in the pathway. Upon ER stress, the NMD pathway is downregulated to allow a robust UPR activation. However, because sustained UPR activity can lead to apoptosis, when the ER stress is resolved, NMD acts to downregulate UPR genes and thus impacts cell fate after ER stress.* ***(B)*** *If ER stress is not remedied, or NMD is not correctly restarted, uncontrolled UPR leads to autophagy and apoptosis. (Figure adapted from Goetz and Wilkinson, 2017)*

UPF3B exhibited UPR activation in response to low level ER stress that triggered little or no response in control cells, suggesting the ability of NMD to suppress the UPR confers a physiological benefit. UPF3B depletion increased the activation of XBP1, phosphorylation of IRE1α, and cleavage of ATF6, providing evidence that NMD can suppress all three branches of the UPR pathway. Furthermore, UPF3B depleted and therefore NMD-deficient mice had a lower threshold for UPR activation in hepatocytes than control mice. Altogether, these results show that NMD protects cells from overreacting to harmless levels of ER stress.

Several studies have indicated that NMD protects cells from stress-induced apoptosis. SMG6-depleted HeLa cells treated with the ER-stress inducing drug tunicamycin exhibited $\sim$50% reduced cell survival compared to control cells, while SMG6 overexpression increased HeLa cell viability (Sakaki et al., 2012). Termination of the UPR after stress has been resolved is vital because of the deleterious effects of extended stress response. NMD-deficient HeLa cells exhibited a prolonged stress response when exposed to moderate doses of the ER stressor tunicamycin, as measured by XBP1s, BIP and CHOP mRNA levels 15–18hrs post exposure (Karam et al., 2015). In vivo mouse studies showed that UPF3B-null mouse hepatocytes also exhibited an abnormally long UPR in response to moderate doses of tunicamycin (Karam et al., 2015). These results indicate that NMD also protects cells from prolonged UPR in response to bona fide stress.

Further evidence of the link between mRNA decay and stress response focuses on their cytoplasmic structures. Although stress granules contain translation initiation factors and P-bodies are associated with miRNA/siRNA silencing and mRNA degradation, they have many parallels. They share some protein components, both can be induced by cellular stress (Teixeira et al., 2005; Raaben et al., 2007) and both are in exchange with polysomes and contain translationally stalled mRNAs that can re-engage in translation (Cougot et al., 2004; Bhattacharyya et al., 2006). The two types of cytoplasmic granule have been shown to

come into direct contact with each other (Kedersha et al., 2005; Souquere et al., 2009). Under certain types of stress, P-bodies frequently dock with stress granules, and overexpress certain proteins that localise to both structures can cause P-body/stress granule fusion (Franks and Lykke-Andersen, 2007).

Studies in *D. melanogaster*, zebrafish, and mammalian cell lines have shown the importance of NMD in neural development and neural stem cell self-renewal vs. differentiation decision (Lou et al., 2016; Jolly et al., 2013; Metzstein and Krasnow, 2006; Wittkopp et al., 2009). The UPF3B-dependent branch of NMD in particular, has been shown to function in neural differentiation and brain development (Tarpey et al., 2007; Jolly et al., 2013; Nguyen et al., 2013). Taken together these results suggest that the UPF3B-dependent branch of NMD may regulate the transient UPR activation that occurs during normal neural differentiation (Zhao et al., 2005; Luo et al., 2006; Weng et al., 2011), as well as the chronic UPR activation that accompanies some forms of neural disease (Yan et al., 2019; Torkzaban et al., 2020; Li et al., 2020). As highly secretory cells, neurons have a highly active ER and therefore are particularly sensitive to ER stress. Supporting this line of thought, neural-specific miRNAs have been identified that target NMD factors and thereby repress NMD (Bruno et al., 2011; Lou et al., 2014).

It has been suggested that UPF1 hyperphosphorylation is involved in the replacement of cap-binding complex from mRNA by eIF4E under stress conditions. UPF1 becomes hyperphosphorylated by PI 3-kinase-related kinase (PIKK) family kinases in circumstances such as when DNA replication is compromised (Kaygun and Marzluff, 2005; Choe et al., 2014). In addition, members of PIKK family ATM and SMG1, both of which induce UPF1 hyperphosphorylation, also function in mitogenic and stress-induced signalling pathways in eukaryotic cells (Brumbaugh et al., 2004). SMG1 which phosphorylates UPF1 to activate the NMD pathway (Section 1.2.2) is also characterised as a genotoxic stress-activated protein kinase and is required for optimal p53 activation after cellular exposure to genotoxic stress.

In response to strong ER stress, NMD is inhibited by the UPR to allow for a full-magnitude UPR response (Figure 1.11. Illustrating the complex interplay between ER stress and NMD, mild ER stress amplified NMD suppression. However, NMD suppression, evidenced by upregulation of NMD-sensitive mRNAs and a decrease in UPF1 phosphorylation, was observed upon either oxidative or ER stress in myogenic cells, cerebral cortical neuronal cells, and astroglial cells (Usuki et al., 2019).

1.4.4 ER-NMD

Depletion of NBAS appears to cause a reduction in NMD activity in human cells, however NBAS is located at the ER outer membrane, whilst canonical NMD occurs in the cytoplasm. This indicates that if NBAS is directly partaking in NMD, it is via a distinct pathway at the ER. Due to the movement of ribosome-transcript complexes to the ER, the window of opportunity for cytoplasmic NMD factors to assemble on these transcripts is narrow, suggesting that if a transcript translated at the ER contains a PTC, it may be resistant to cytoplasmic NMD. Together, these imply suggests an alternate NMD pathway specific to the ER would be necessary to target transcripts with PTCs that are translated at the translocon (Figure 1.12).

As described in Section 1.1.4, between 30% and 50% of transcripts are translated at the ER. Therefore, it is reasonable to assume that NMD must take place at the ER to regulate these transcripts. Although Staufen1 (STAU1), which regulates SMD quality control, is localised to the ER, this does not protect against translation of transcripts with PTCs.

Further supporting the hypothesis of an ER-specific NMD pathway, HaloTag fusion UPF1 has been shown to interact with ER membrane proteins (Brannan et al., 2016) and a fraction of UPF1 has also been shown to localise to the ER outer membrane and interact directly with SEC61B, a component of the ER translocon (Longman et al., 2020). As well as NBAS being located at the ER membrane (Section 1.3.2) another putative NMD factor identified in *C. elegans*, SEC13, is also localised to the ER (Casadio et al., 2015). Similar to NBAS, SEC13 comprises

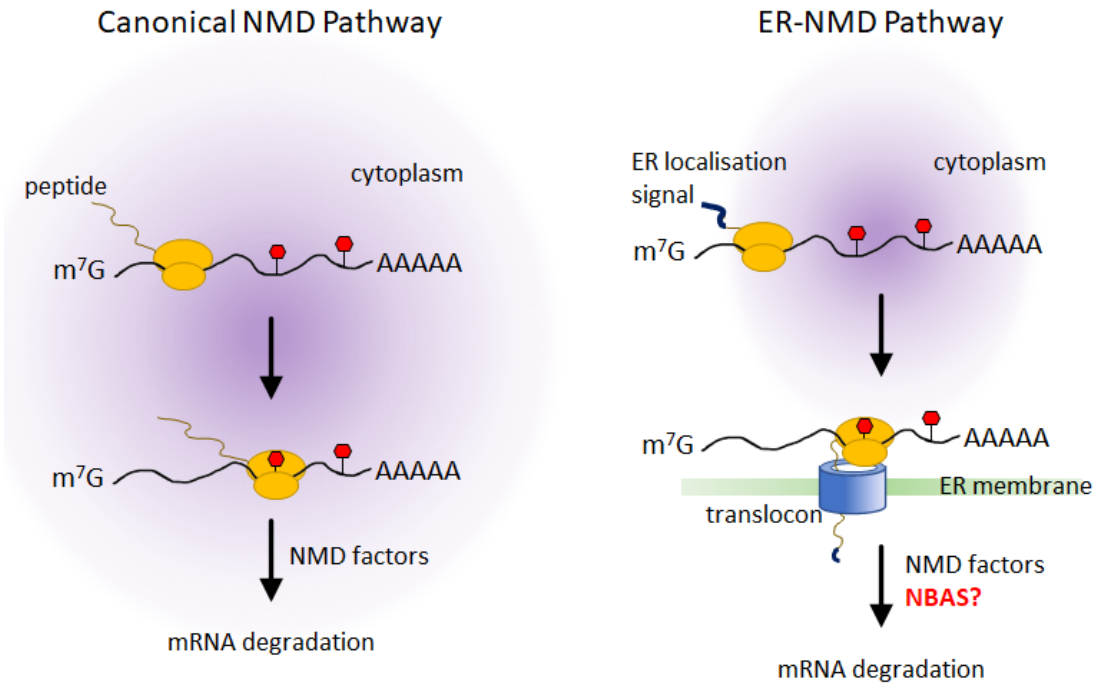


Figure 1.12: ***Theoretical model of ER-NMD***

NBAS has been shown to function in the NMD pathway and localises to the membrane of the endoplasmic reticulum (ER). NBAS recruits the core NMD factor UPF1 and tethers it to the membrane of the ER. This suggest a local NMD response that targets for degradation mRNAs that are translated at the ER.

six WD40 domains and is located on transport vesicles budding from the ER as well as being a constituent of the nuclear pore complex (Shaywitz et al., 1995). Whereas NBAS functions in Golgi-to-ER vesicular transport, SEC13 forms part of the COPII vesicle coat which promotes the formation of transport vesicles that carry secretory proteins from the ER to the Golgi (Salama et al., 1997).

As already discussed more generally, NMD plays an important role in preventing translation of truncated proteins. Protein aggregates made up of truncated or misfolded proteins, which have likely escaped NMD regulation, trigger ER stress which underlies many diseases (Blázquez et al., 2014; Fung and Liu, 2014; Jheng et al., 2014; Giampietri et al., 2015; Rodvold et al., 2016; Shah et al., 2017; Yan et al., 2019; Torkzaban et al., 2020; Li et al., 2020). The mutations underlying several diseases associated with NMD occur in secretome genes. For example, since CFTR is an ion channel, it has several transmembrane domains and is likely translated at the ER. Therefore, quality control of RNA is arguably even more important at the ER than in the cytoplasm.

1.5 Aims and research outline

1.5.1 Specific motivations and importance

NBAS appears to be involved in NMD in human cells but unlike most other identified NMD factors, is localised to the ER. This suggests that as well as the canonical NMD pathway located in the cytoplasm and regulated by UPF1 and other NMD factors, there is also a specific NMD pathway localised to the ER that is regulated by NBAS as well as UPF1 and other NMD factors. This branch of the NMD pathway would specifically target those transcripts that are translated at the ER, although this is not exclusive to secretome proteins (Section 1.1.4). Understanding more about NMD at the ER is important for knowledge of RNA quality control and associated diseases.

More generally, there is still a lot of conflicting literature and differing opinions about the NMD pathway and how its targets are identified. Therefore, further clarifying NMD target characteristics will add to the understanding of both the remit and the limits of this quality control pathway. This could be beneficial in the area of designing therapeutics to force readthrough of stop codons to produce truncated protein (Section 1.2.2).

Since several patient phenotypes have been caused by mutations in NBAS, it is beneficial to understand more about the functions NBAS has within the cell, both relating to NMD and how this is linked or distinct from its role in retrograde transport. No mechanism is known by which these patient mutations cause their phenotypes, so understanding more about NBAS may shed light on these cases and potentially reveal therapeutic targets.

1.5.2 Main research questions

Throughout this thesis I aim to address the following questions:

1 What is the role of NBAS in the NMD pathway?

(a) What proteins does NBAS interact with?

(b) What transcripts are targeted by NBAS?

(c) How does this subset differ from general NMD targets?

(d) Is NMD quality control occurring at the ER?

2 What are the characteristics of NMD targets and how do we define them?

(a) How does defining targets by change in stability compare with change in expression?

(b) Is it possible to directly measure change in stability of individual transcripts?

I will address the first main question through identification of the protein binding partners of NBAS in human cells (Chapter 3) and identification and characterisation of transcripts increased in expression and/or stability when NBAS is depleted (Chapter 4 and 5).

The second main question will be addressed through analysis of the differentially expressed transcripts compared with differentially stabilised transcripts upon depletion of both NBAS and the core NMD factor UPF1 (Chapter 4 and 5). Furthermore, I will investigate the measurement of specific transcripts' stability as a means to identify direct NMD targets (Chapter 6).

Chapter 2

Methodology

2.1 Laboratory techniques

2.1.1 Strains and plasmids

All experiments were carried out using HeLa cells (ATCC), obtained from lab stocks in liquid nitrogen storage. 3xFLAG-tagged expression vectors for UPF1 and NBAS were generated by subcloning the 3xFLAG tag into the multiple cloning site of pcDNA3 (Life Technologies). UPF1 deletion constructs (ΔCT, ΔNT, ΔCH) were a generous gift from Andreas Kulozik (University of Heidelberg). NBAS deletion constructs (ΔC1, ΔN2, Δsec39 and ΔWD40+sec39) were generated in house using the 3xFLAG pcDNA3 backbone. Small interfering RNAs (siRNAs) used were ON-TARGETplus control pool (si-scramble; si-Scr), hUPF1 pool (si-UPF1) and hNBAS pool (si-NBAS) (Dharmacon).

2.1.2 Cell culture

HeLa cells were grown in 1X Dulbecco's Modified Eagle Medium (DMEM) with high glucose, GlutaMAX supplement and pyruvate (Gibco Life technologies; 10569010) supplemented with 10% fetal calf serum (in-house technical services). Cells were maintained in 5% CO_2 at 37°C. HeLa 3xFLAG-eGFP-NBAS cells were selected in the same conditions and media with the addition of 1g/ml of Puromycin (Gibco; A11138043), and maintained in the same conditions and media with the addition of 1% Penicillin/Streptomycin (in-house technical services).

2.1.3 Bacterial DNA transformation

For all cloning applications, DNA was transformed into sub-cloning efficiency DH5α *Escherichia coli* cells (Invitrogen). 50μl thawed cells were used per transformation. After addition of DNA, cells were incubated on ice for 30min then subjected to heat shock at 42°C for 45sec, followed by 2min on ice. 500μl of SOC

medium [2% (w/v) Tryptone, 0.5% (w/v) Yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgCl2, 20mM Glucose], was added and tubes were incubated at 37°C for 1h with shaking at 225 rpm. Transformants were plated on either Ampicillin or Kanamycin agar plates, as appropriate, and incubated overnight at 37°C.

2.1.4 Cell transfections

All transfections were carried out in Opti-MEM reduced serum medium (Gibco, 31985047). Reverse transfections of siRNA oligos were carried out using DharmaFECT 1 transfection reagent (Dharmacon, T-2001-03) following manufacturer's protocol, with cells at 50-60% confluency. For total RNA-sequencing and subcellular fractionation, cells were reverse transfected in 6-well plates with 30pmol of indicated siRNAs. 48h later, cells were expanded into 10cm plates and reverse transfected with 1μmol of the same siRNAs. Cells were harvested for analysis 24h after the second depletion. Plasmids were transfected using Lipofectamine 2000 (ThermoFisher Scientific, 11668019) following manufacturer's protocol on cells at 80% confluency. The amount of DNA used was dependent on the cell vessel. For co-immunoprecipitation, cells were co-transfected with 2.4g total mix of T7-tagged and FLAG-tagged expression vectors for 24-48h before harvesting.

2.1.5 RNA extraction

Cells were harvested by addition of TRIzol reagent (ThermoFisher Scientific; 15596018) and scraping. Samples were frozen at -80°C until further processing. Total RNA was isolated by phenol-chloroform extraction and treated with TURBO DNA-freeTM DNase I kit (Invitrogen Ambion; AM1907) according to the manufacturer's instructions. RNA concentration was measured using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific) using the manufacturer's broad-spectrum RNA assay.

2.1.6 Gel electrophoresis

All agarose gels for electrophoresis were prepared using 1x TBE buffer [1M Tris base, 1M Boric acid, 0.02M EDTA] supplemented with ethidium bromide (2 μl/100 ml). 1xTBE was used as the running buffer and the gel percentage was adjusted to suit application.

2.1.7 Quantitative real time – polymerase chain reaction (qRT-PCR)

Complementary DNA (cDNA) synthesis and qRT–PCR were performed using Super-ScriptIII One-Step RT–PCR Kit (Invitrogen) following the manufacturer's instructions. To measure depletion of NMD factors and the corresponding block of NMD, qRT–PCR analysis was performed to compare RNA levels corresponding to a control mRNA, the mRNA targeted by siRNA and potential NMD target mRNAs. Each gene was assayed in separate wells and four technical replicates of each reaction were prepared for all samples. Primers used were designed using Roche Real-Time Ready Configurator and combined with Roche Universal Probe Library. All primer sequences are listed in Section 2.1.18. qRT-PCR was run on an LC480 lightcycler (Roche) following the program: reverse transcription at 55°C for 30min, melting at 95°C for 2min, then 40 cycles of 95°C for 30sec, 55°C for 20sec, 70°C for 20sec followed by the plate read step, and finishing with 5min at 70°C. Gene expression data was analysed by the delta delta Ct method, with each gene normalised to housekeeping gene POL2RJ. Student's unpaired t-test was used for statistical validations.

2.1.8 Western blotting

Cells were washed and harvested in ice-cold PBS, centrifuged to separate PBS and then cell pellets were lysed with immunoprecipitation (IP) buffer [20mM Tris-HCl pH 8, 150mM NaCl, 1mM EDTA, 1% NP-40, 0.2% NaDeoxycholate, Complete Protease Inhibitor (Roche), PhosphoSTOP (Roche), 1mM DTT] for 20min on

ice. Proteins were resolved by SDS-PAGE using NuPAGE 3-8% Tris-Acetate precast gels (Novex, Life Technologies, Invitrogen) run for 1h at 170V in 1x NuPAG Tris-Acetate SDS Running Buffer (Invitrogen, LA0041). Protein transfer was achieved using the iBlot2 Gel Horizontal Transfer Device (Invitrogen). Nitrocellulose membranes were blocked in 5% BSA in PBST (PBS; Tween 20 0.1%) for a minimum of 1h at room temperature and probed with the appropriate primary antibody diluted in blocking solution, overnight at 4°C. Secondary horseradish peroxidase (HRP) antibodies were used for developing the signal diluted 1:10,000. Antibodies used for western blot analyses are listed in Section 2.1.18. Blots were developed using Pierce ECL Western Blotting Substrate (Thermofisher Scientific, 32209) and imaged with an ImageQuantLAS 4000 chemiluminescent camera (GE Healthcare).

2.1.9 CRISPR strategy and cloning

CRISPR targetting was carried out using pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid, which encodes the Cas9 protein and the transactivating CRISPR RNA (tracrRNA), and custom, sequence-specific CRISPR RNAs (crRNAs). The design of the crRNAs was undertaken as in Ran et al. (2013b). Using 100 bp genomic DNA sequence around the appropriate gene locus (Ensembl genome browser), crRNAs were designed with the help of the two online tools sgRNA Designer CRISPRko (Broad Institute, portals.broadinstitute.org/gpp/public/analysis-tools /sgrna-design) and the Cas-Designer (RGEN Tools, rgenome.net/cas-designer/). The suggested 20 nucleotide crRNA sequences required for Cas9 specific gene targeting were preferentially selected by closest proximity to the start codon and the highest score determined by the design programs. For each crRNA, the complementary sequence was determined, and a BbsI restriction site added to both oligos. Designed crRNAs were then ordered as custom single stranded DNA oligos with an appended 5' phosphate to facilitate ligation into the targeting vector (Ran et al., 2013b). A table with crRNA sequences is given in Section 2.1.18. Top and bottom strands of ordered crRNAs were resuspended in dH_20 and annealed at a concentration of 100μM at 95°C for 5min and subsequently allowed to slowly

cool to room temperature. pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid #62988). Annealed crRNAs were cloned into the PX459 vector using BbsI restriction cloning, directly before the tracrRNA to form functional guide RNA (gRNA) sequences. 1μl of a 1:250 dilution of annealed crRNAs was ligated using T4 DNA Ligase (NEB) into 36ng of the PX459 vector. Ligation was performed at room temperature for 2hr. To assess cutting efficiency of each gRNA/Cas9, each plasmid was transfected into HeLa cells (protocol as in Section 2.1.4) and RNA extracted as in Section 2.1.5. PCR over the target region was performed with custom primers (Section 2.1.18) and the resulting products assessed by gel electrophoresis (Section 2.1.6). The PCR products were also sent for Sanger sequencing (MRC PPU, Dundee) and the traces analysed using Tracking of Indels by DEcomposition (TIDE Brinkman et al., 2014) and Inference of CRISPR Edits (ICE; Synthego) online tools to give exact cutting efficiencies for each guide.

The repair template was designed using SnapGene (GSL Biotech) by combining the genomic sequence for eGFP and 3xFLAG (downloaded from SnapGene repository) and integrating short linker sequences; a five amino acid linker between eGFP and 3xFLAG (sequence: tccggactcagatct) and a six amino acid linker between 3xFLAG and the start of the NBAS coding sequence (sequence: ggtaccagcggcggaagc). Homology arms of length 400 nucleotide were deemed necessary as a consequence of the large size of the tag (Chan et al., 2007a). The full repair template sequence was ordered from Integrated DNA Technologies Inc. on a pUCIDT-AMP plasmid backbone. After bacterial amplification, the plasmid was extracted using Plasmid Maxi Prep kit (Qiagen) following manufacturer's protocol.

2.1.10 CRISPR targeting and screening

The gRNA/Cas9 plasmid and repair template plasmid were transfected into HeLa cells as described in Section 2.1.4, at a ratio of 2:1 ($\sim$900ng : 500ng per well of 6-well plate). After 24h media was supplemented with 1g/ml puromycin to select for cells expressing the gRNA/Cas9 plasmid. Once control cells were >90% dead,

surviving cells' fluorescence was measured by fluorescence-activated cell sorting (FACS) and single cells deposited into wells of a 96-well plate. After approx. two weeks of clonal expansion, wells containing growing colonies were treated with 30μl pre-warmed trypsin and subsequently quenched with 70μl of appropriate cell culture media. Dispersed clones were then transferred into a new 96-well plate containing 100μl fresh media. To remove residual trypsin, the media was replaced after 24hr. After replica plating, cells were grown in 96-well plates for 1-2 days until growth media turned yellow within 24h. To extract genomic DNA, cells were washed twice with PBS and subsequently lysed using the DirectPCR Lysis Reagent (Cell) supplemented with 0.5 g/l Proteinase K (Roche). Following the manufacturer's instructions, lysates were incubated in a wet chamber placed in a rocket shaker at 55°C overnight. After Proteinase K inactivation at 85°C for 45min, crude lysates were frozen until used for genotyping. Colony PCRs were performed by adding primers and SYBR green Master Mix to each well. The following programme was carried out: initial denaturation at 98°C for 5min, followed by 35 cycles of: denaturation at 98°C for 30sec, annealing at 55°C for 90sec and extension at 72°C for 30sec. 1% TBE agarose gels supplemented with ethidium bromide were used for visualisation of the PCR products and genotyping.

2.1.11 Flow cytometry and fluorescence-activated cell sorting (FACS)

Cells were trypsinised, washed with cold PBS and placed on ice. FACS was performed by the IGMM FACS core facility on an Aria II cell sorter (BD biosciences). Cells were sorted based on GFP fluorescence, analysed by FACS BD AccuriT (488nm excitation laser, FL-1: 533/30nm emission filter) running the software BD AccuriTM C6. Gates were set using a non-fluorescent control. GFP expression was analysed by gating the single cell population in a SSC-H/SSC-A dot plot, followed by debris-exclusion gate in a SSC-A/FSC-A dot plot. 10,000 single cells were analysed for each sample. The mean GFP fluorescence was calculated with FlowJoTM Software (version 10.6.0).

2.1.12 Immunoprecipitation and co-immunoprecipitation

Cells were washed and harvested in ice-cold PBS before pellets were lysed with immunoprecipitation (IP) buffer [20mM Tris-HCl pH 8, 150mM NaCl, 1mM EDTA, 1% NP-40, 0.2% NaDeoxycholate, Complete Protease Inhibitor (Roche), PhosphoSTOP (Roche), 1mM DTT] for 20min on ice. Lysates were precleared with Dynabeads Protein-G (Novex, Life Technologies) for 1h, rotating at 4°C. Appropriate antibody coupled magnetic beads were washed three times with IP buffer before incubation with the precleared lysate overnight rotating at 4°C. Beads were washed five times, each for 5min with IP buffer and then bound protein was eluted by boiling in 4X SDS sample buffer supplemented with reducing agent for 5min. Antibodies used for immunoprecipitations are listed in Section 2.1.18.

2.1.13 Mass spectrometry

Cells were harvested and lysed as in Section 2.1.12. α-GFP antibody-coupled magnetic beads (Sigma) were equilibrated with IP buffer. Lysates were resuspended in 500μl IP buffer for capture of GFP-NBAS bound proteins and subsequent mass spectrometry analysis. Immunoprecipitation and washing was performed on a KingfisherDuo robot (Thermo) for 4h at 4°C. Beads were then transferred for two washes in IP buffer and three washes in TBS (300μl each). After transfer into 100μl 2M urea, 100mM Tris, 1mM DTT containing 0.3g trypsin per sample, beads were incubated at 27°C for 30min with mixing to achieve limited proteolysis. The beads were then removed, and tryptic digest of the released peptides was allowed to continue for 9h at 37°C. Following this, peptides were alkylated by adding iodoacetamide to 50mM and incubating at room temperature for 30min. Finally, peptides were acidified by addition of 8μl 10% TFA. An estimated 10mg of the resulting peptide solution was loaded onto an activated (20μl methanol), equilibrated (50μl 0.1% TFA) C18 StAGE tip, and washed with 50μl

0.1% trifluoroacetic acid (TFA).The bound peptides were eluted into a 96-well plate with 20μl 80% acetonitrile (ACN), 0.1% TFA and concentrated to less than 4μl in a vacuum concentrator. The final volume was adjusted to 15μl with 0.1% TFA.

Mass spectrometry (MS) processing and analysis of raw data was carried out by the IGMM Mass Spectrometry core facility. Liquid chromatography was performed using a Dionex RSLC Nano (Thermo Fisher Scientific). 5g peptides were injected onto a C18 packed emitter and eluted over a gradient of 2%-80% acetonitrile in 48min, with 0.5% acetic acid throughout. Eluting peptides were ionised at +2.2kV before data-dependent analysis on a Thermo Q-ExactivePlus. First spectrometer measurement (MS1) was acquired with mz range 300–1650 and resolution 70,000, and top 12 ions were selected for fragmentation with normalised collision energy of 26, and an exclusion window of 30s. Second spectrometer measurements (MS2) were collected with resolution 17,500. Raw MS data were analysed using MaxQuant (v1.5.6.5) (Max Planck Institute of Biochemistry) in conjunction with UniProt human reference proteome release 201611 (uniprot), with match between runs (MS/MS not required), label-free quantification (LFQ) with 1 peptide required, and statistical analyses performed using Wasim Aftab's LIMMA Pipeline Proteomics (github.com/wasimaftab/ LIMMA-pipeline-proteomics) implementing a Bayes-moderated method. Interactome analysis including gene ontology was carried out by inputting protein hits into STRING (string-db.org Szklarczyk et al., 2019).

2.1.14 Immunofluorescence

Cells were grown on coverslips, fixed with 4% paraformaldehyde, washed with PBS and permeabilised with 0.5% Triton X-100. Coverslips were then incubated for 1h with block buffer (1% BSA, 0.01% Triton X-100 in PBS). Primary antibodies (Section 2.1.18) were diluted 1:500 in block buffer. Primary antibodies were incubated with coverslips in a humidified chamber overnight at 4°C. Coverslips were washed 3 times with wash buffer (0.01% Triton X-100 in PBS). Secondary antibodies (Section 2.1.18) were diluted 1:1000 in block buffer and incubated with

coverslips in a dark, humidified chamber for 1h at room temperature. Coverslips were washed twice with wash buffer, incubated with DAPI diluted in wash buffer (1:1000) and mounted in Vectrashield (Vector) sealed with clear nail varnish. Where cells were treated with digitonin, media was aspirated, and cells incubated for 10min in ice-cold PBS, on ice. Then cells were incubated in digitonin buffer (110mM KoAc, 25mMK-HEPES, 2.5mM MgCl2, 1mM EGTA) with 0.01% digitonin for 5min on ice, washed in room temperature PBS and fixed with 4% paraformaldehyde as above. Images were acquired on a Nikon Confocal A1R confocal microscope using a Plan Apochromat 100x1.4NA objective. A piezo-electrically driven objective mount (PIFOC model P-721, Physik Instrumente) was used to control movement in the z dimension. Step size for z-stacks was set at 0.2μm. Hardware control and image capture were performed using NIS Elements AR software (Nikon Instruments Europe). Image analysis was performed using FIJI/ImageJ software (USA National Institute of Health)(Section 2.2.12).

2.1.15 Proximity ligation assay

Cells were grown, fixed and permeabilised as in Section 2.1.14. Coverslips were then incubated for 1h with PLA block buffer from Duolink In-Situ PLA Kit (Olink Bioscience) at room temperature and incubated with primary antibodies as for immunofluorescence. Coverslips were washed three times with wash buffer A (Duolink In-Situ PLA Kit) and intermolecular interaction explored according to the manufacturer's instructions. The PLA reaction occurs between target proteins that are located in close proximity (<40nm). PLA probes used were goat PLUS (DUO92003, Sigma-Aldrich), rabbit MINUS (DUO92005, Sigma-Aldrich) and mouse MINUS (DUO92004, Sigma-Aldrich). Cover slips were incubated with DAPI diluted 1:1000 in 0.01X wash buffer B (Duolink In-Situ PLA Kit) and mounted in Vectrashield (Vector). Images for PLA experiments were acquired as for immunofluorescence experiments.

2.1.16 Metabolic labelling and RNA fractionation

Sub-confluent HeLa cells in 10cm dishes were labelled with the uridine analogue 4-thiouridine (4sU) (SigmaAldrich; T4509), for 40min by mixing the 4sU into normal media at a final concentration of 100μM. After 40min incubation at 37°C, media was aspirated, and cells harvested by addition of 5ml TRIzol reagent (ThermoFisher Scientific; 15596018) and scraping. Samples were frozen at -20°C until further processing. Total RNA was isolated by phenol-chloroform extraction (Section 2.1.5) Nascent RNA pull-down was completed by adapting the method from Russo et al. (2017). Biotinylation was achieved by assembling a 150ml reaction containing 100μg of 4sU labelled total RNA, 15μl 10x biotinylation buffer (100mM HEPES [pH 7.5], 10mM EDTA) and 10μl MTSEA-biotin-XX (Biotium; 90064) 1mg/mL dissolved in dimethylformamide, and incubating at room temperature for 2hr. Excess biotin was removed by addition of an equal volume of chloroform and centrifugation in MaXtract High Density phase-lock gel tubes (Quiagen; 129065) at 13,000rpm for 5min at 4°C. RNA was precipitated using 0.1x volume 5M NaCl and 1x volume isopropanol and 1μg RNA kept as "total RNA". MACS streptavidin magnetic beads (Miltenyi; 130-074-101) were added 1:1 to RNA and incubated at room temperature for 15min in the dark. Beads were isolated from each sample by applying the sample to pre-equilibrated MACS magnetic columns and washing with wash buffer [100mM Tris Ph7.5, 10mM EDTA, 1MNaCl, 0.1% Tween 20]. Flow through of "pre-existing RNA" was collected and "nascent RNA" was eluted from streptavidin beads by addition of 100mM dithiothreitol (DTT, Thermo Scientific; 20291) and precipitation with 100% ethanol and 5M NaCl at -80°C for 20min. Total RNA quantity and quality were assessed using an RNA 6000 pico chip on an Agilent 2100 Bioanalyzer at the Wellcome Trust Clinical Research Facility, Edinburgh. The Agilent Bioanalyzer reports RNA integrity numbers (RINs), which indicate how intact an RNA sample is (1 meaning completely degraded and 10 meaning completely intact). Sample RINs ranged from 8.6 to 10, indicating that the RNA quality was sufficient for sequencing.

2.1.17 Library preparation and RNA-sequencing

RNA-sequencing was performed on both total and nascent RNA samples for each condition. Libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs; E7420) and Agencourt AMPure XP Beads (Beckman Coulter; A63881), following manufacturer's procedure. Total samples were fragmented for 15min and nascent samples for 5min, both at 94°C. Samples were ligated to primers 1-40 from NEBNext Multiplex Oligos for Illumina (Index Primers Set 1, 2 and 3; New England Biolabs; E7335, E7500, E7710) and analysed using DNA High Sensitivity chip on an Agilent 2100 Bioanalyzer before being pooled. Library molarity for sequencing was calculated using Qubit dsDNA high specificity quantification and fragment size information from Bioanalyzer. 150 base pair, paired-end sequencing was performed using an S2 chip on the NovaSeq platform (Illumina Inc.). Sequencing was performed by Edinburgh Genomics, Edinburgh.

2.1.18 Materials and reagents

Primary antibodies

Target Protein	Species	Company	Product Number
GFP	Mouse	Roche	11814460001
FLAG	Mouse	Sigma-Aldrich	#F3165
NBAS	Rabbit	Abcam	#Ab122370
Calnexin	Rabbit	Enzo Life Sciences	ADI-SPA-860-D
SEC61B	Rabbit	Proteintech	15087-1-AP
UPF1	Goat	Bethyl Laboratories	A300-038A
Tubulin	Mouse	Sigma-Aldrich	#4026
T7	Mouse	Novagen	#69522

Fluorescent secondary antibodies

Target Species	Fluorophore	Company	Product Number
Rabbit	Alexa FlourR 488	Molecular Probes	A11070
Mouse	Alexa FlourR 594	Molecular Probes	A21203
Mouse	Alexa FlourR 488	Molecular Probes	A11054
Rabbit	Alexa FlourR 594	Molecular Probes	A31632

HRP-linked secondary antibodies

Target Species	Company	Product Number
Mouse	GeneTex	GTX221667
Goat	GeneTex	GTX628547
Rabbit	GeneTex	GTX221666

PCR & sequencing primers

Primer Name	Sequence	Melting Temperature
KJJ_5	TGACTTGGCCGATCTCATGT	59.1°C
KJJ_6	AGCGACACTAACAATCCCCT	58.7°C

Guide RNAs

Name	Sequence	Strand	Oligo	Sequence
gRNA #1	GCCTGAGT CCTCGGCG AACA	sense	top	/5Phos/CACCGggcctgagtcctcggcgaaca
			bottom	/5Phos/AAACtgttcgccgaggactcaggccC
gRNA #2	TGAGTCCT CGGCGAAC ATGG	sense	top	/5Phos/CACCGgtgagtcctcggcgaacatgg
			bottom	/5Phos/AAACccatgttcgccgaggactcacC
gRNA #4	GGGGGCCG CCATGTTC GCCG	antisense	top	/5Phos/CACCGgggggccgccatgttcgccg
			bottom	/5Phos/AAACcggcgaacatggcggccccccC
gRNA #6	GCCATGTT CGCCGAGG ACTC	antisense	top	/5Phos/CACCGggccatgttcgccgaggactc
			bottom	/5Phos/AAACgagtcctcggcgaacatggccC

qRT-PCR primers

Gene	Sequence	UPL Probe
POLR2J	CTGTGAGCCCCGTTCCTAC	#1
	GTCGGTGTCAGGGTGAGG	
UPF1	CCTGCACACCAAGCTCTACC	#39
	TTCTCGCGTCCCTGAAAG	
CCL20	TTGAGCTAAAAACCATGTGCTG	#71
	CGCAGAGGTGGAGTAGCAG	
INHBA	CAATACGCACTGCATGAGAAG	#85
	CCCACATGCATGATATCCAA	

*NBAS qRT-PCR was carried out using Roche RealTime ready Custom Single Assay.

2.2 Computational techniques

2.2.1 Quality control of RNA-sequencing data

The `.fastq.gz` files containing raw read sequences from the total/nascent experiment were downloaded from the Edinburgh Genomics delivery server and md5sums were checked to identify any file corruptions or truncations. The `.fastq.gz` files containing raw read sequences from the membrane/cytoplasm experiment were retrieved from an external hard drive and md5sums were checked. `.fastq` files were quality control checked for base and sequence quality scores, and adapter contamination using fastQC (v0.11.7; Babraham Bioninformatics).

2.2.2 Mapping of RNA-sequencing data

Since no experimental purification of mRNA was used, many ribosomal RNA (rRNA) reads were present in total RNA. Therefore, the rDNA reference U13369 sequence was downloaded from NCBI (www.ncbi.nlm.nih.gov/nuccore/555853/?report=fasta) and Basic Local Alignment Search Tool (BLAST) used to identify coordinates of high identity and long overlaps in the reference genome `GRCh38/hg38`. These regions were then masked by changing the sequence to runs of Ns, and a single copy of the U13369 sequence was added as an extra chromosome to the genome. Reads were aligned and psedudoaligned separately. Alignment was done using Spliced Transcripts Alignment to a Reference (STAR, v2.5.1b) (Dobin et al., 2013) to the masked reference genome in paired-end mode and default settings. Samtools flagstat (Li et al., 2009) was used to get information about numbers of reads and bam files were sorted by name using samtools sort (Li et al., 2009). Pseudoalignment was done using kallisto (v0.43.1) (Bray et al., 2016). Kallisto index was created by combining "all basic gene annotation" and "long non-coding RNA gene annotation" fasta files, downloaded from Gen-

code (www.gencodegenes.org/ releases/current.html), release 27 (`GRCh38.p10`). Kallisto was run with 100 bootstraps. Reads from the same sample but sequenced on separate lanes were initially processed individually to ensure that they correlate well with each other, and then were merged for further analysis.

2.2.3 Differential expression analysis of total RNA-sequencing

Reads per feature in STAR output files were counted using HTSeq (v0.9.1) (Anders et al., 2015). These output files and kallisto output abundance files were then analysed for fold changes using DESeq2 (v1.14.1) differential expression analysis, using wald test and standard parameters. Target lists were defined as all genes significantly increased in expression in treated cells compared to control cells (Benjamini-Hochberg false discovery rate multiple testing corrected p value < 0.05).

2.2.4 Analysis of fractionated RNA-sequencing

Reads per feature in STAR output files was counted using HTSeq (v0.9.1) as above. The resulting raw counts were transformed using the DESeq2 vst function to perform variance stabilizing transformations. These transformed counts were then averaged across replicates and gene expression in the two fractions from control cells was compared. Membrane-associated genes were defined as those >2-fold higher expression in membrane fraction than cytoplasmic fraction in these control cells, all other genes were termed non-membrane. Enrichment for membrane genes was validated by overlap of these newly defined "membrane-associated genes" with three experimental ER datasets (Section 2.2.14).

Standard DESeq2 differential expression analysis using wald test and standard parameters was then carried out on "membrane-associated genes" in the membrane fraction and on "non-membrane genes" in the cytoplasmic fraction. Genes were determined as regulated by UPF1, NBAS or UPF2 if they were significantly (Benjamini-Hochberg false discovery rate multiple testing corrected p value < 0.05) increased in expression when the relevant factor was depleted.

2.2.5 Characterisation of gene subsets

Gene Ontology (GO) analysis was carried out using Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) (cbl-gorilla.cs.technion.ac.il/) (Eden et al., 2009) and Database for Annotation, Visualization and Integrated Discovery (DAVID) (david.ncifcrf.gov/home.jsp) (Dennis et al., 2003). To investigate genes of specific GO terms, genes were annotated using the biomaRt Bioconductor R package (Durinck et al., 2005) and filtered for the relevant categories. Overlap with experimental ER-localisation was carried out by extracting relevant gene lists from supplementary data of the following articles; APEX-seq (Fazal et al., 2019), ER Fractionation (Reid and Nicchitta, 2012) and ER Ribosome Profiling (Jan et al., 2014) (Section 2.2.14). For APEX-seq and ER Fractionation-Seq the published ER gene lists were used. Gene names from ER Ribosome Profiling were converted using DAVID and a threshold of 1.5-fold enrichment at ER applied.

2.2.6 Defining exonic and intronic reads

Genomic and exonic intervals from gencode.v27.annotation.gtf (EBI Gencode; gencode-genes.org/human/) were used to define intronic and exonic regions of each gene and Bedtools intersect function to determine categorisation of each read (Figure 2.1). For this analysis, transcripts were merged, and all possible exons were classified as exonic regions, producing obligate exons and facultative introns. Reads that overlapped any intronic region by one base or more were

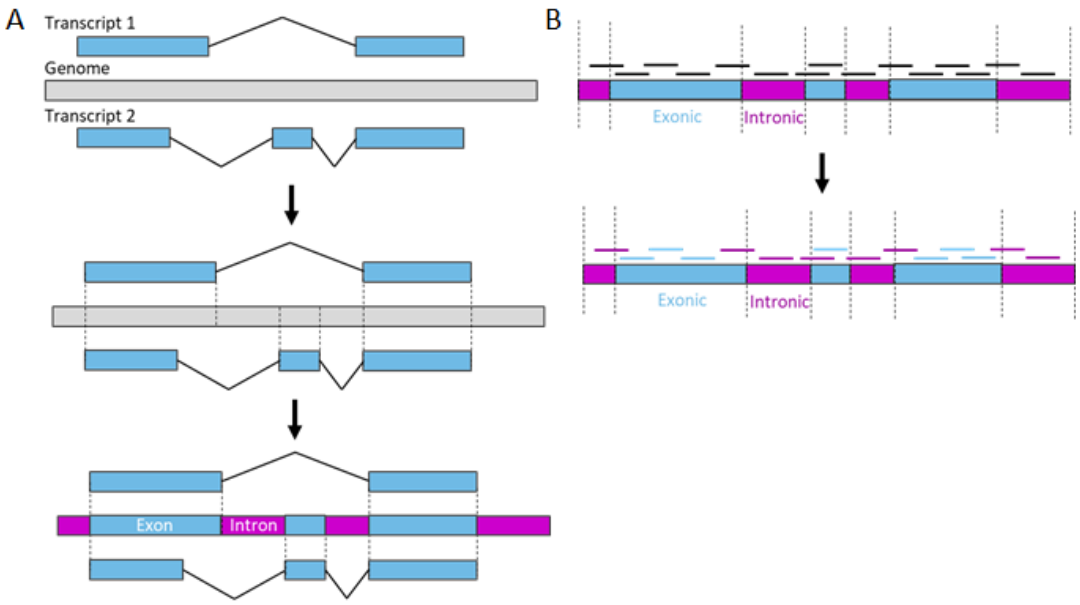


Figure 2.1: ***Schematic of custom analysis pipeline to calculate change in stability of genes in depleted cells compared to control***

(A) *"Exonic" regions (blue) were defined as any part of the gene body that is defined as an exon in one or more transcripts and "intronic" regions (purple) as those parts of the gene which are never defined as an exon, resulting in facultative exons and obligate introns.* ***(B)*** *"Exonic" reads were defined as those that exclusively overlapped with "exonic" regions and "intronic" reads as those which even partially overlapped with defined "intronic" regions.*

classed as intronic reads since it was presumed they must originate from unspliced pre-mRNA. Only reads that were fully within exonic regions were termed exonic reads. The number of exonic and intronic reads were summed for each gene for downstream analysis.

2.2.7 Differential stability analysis using total and nascent RNA-sequencing data

To normalise for transcription rate, nascent RNA sequencing data as well as total was analysed to extract exonic and intronic read counts per gene. An odds ratio (OR) of each gene was calculated according to the follow equation:

$$OR = (total\ exonic\ (mature\ mRNA)/total\ intronic\ (pre-mRNA))$$
$$/(nascent\ exonic\ (mature\ mRNA)/nascent\ intronic\ (pre-mRNA))$$

This produced a single value per gene in each condition which is a proxy for the stability of the mRNA. Therefore, the stability (OR) of each gene could be compared between conditions. To test significance of these changes, a paired t-test was carried out for each gene between the two groups of ORs (five replicates for each condition). This gave the magnitude and significance of the change in OR, which is a proxy for change in stability. Replicates were then combined to calculate a mean change in OR for each gene.
To analyse this data using DESeq2, a two-factor design was used to account for total or nascent fraction and si-Scr or si-UPF1 treatment. A reduced likelihood ratio test (LRT) was used to determine if the increased likelihood of the data using the extra terms in the full model was more than expected if those extra terms were truly zero. The following commands were run, where "assay" represents the siRNA treatment and "condition" represents the total or nascent fraction.

```
$ dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, design=
~ assay + condition + assay:condition)
$ dds <- DESeq(dds, test="LRT", reduced= ~ assay + condition)
$ results <- results(dds)
```

Genes were determined to be differentially stabilised upon depletion of UPF1 or NBAS if they were significantly changed after Benjamini-Hochberg false discovery rate multiple testing correction ($p < 0.05$).

2.2.8 Differential stability analysis using only total RNA-sequencing data

To calculate change in stability using only total RNA, the ratio between exonic reads (mature RNA) and intronic (pre-mRNA) reads for each gene was analysed. To measure the effect of UPF1 or NBAS depletion on stability, an odds ratio (OR) of each gene was calculated using the following equation:

$$OR = (treated\,exonic\,(mature\,mRNA)/treated\,intronic\,(pre-mRNA))$$
$$/(control\,exonic\,(mature\,mRNA)/control\,intronic\,(pre-mRNA))$$

Unlike the previous OR analysis, this OR produced a proxy for change in stability so that si-UPF1, si-NBAS and si-Scr were all considered "treated" and mock cells used as the "control" for all three. The stability (OR) values of si-UPF1 and si-NBAS were then compared to those of si-Scr to identify changes beyond technical noise.

These values were also analysed using DESeq2 using a two-factor design and a reduced likelihood ratio test (LRT) to determine if the increased likelihood of the data using the extra terms in the full model was more than expected if those extra terms were truly zero. The following commands were run, where "assay" represents the siRNA treatment and "condition" represents exonic or intronic read count.

```
$ dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, design=
~ assay + condition + assay:condition)
$ dds <- DESeq(dds, test="LRT", reduced= ~ assay + condition)
$ results <- results(dds)
```

Genes were determined to be differentially stabilised upon depletion of UPF1 or NBAS if they were significantly changed after Benjamini-Hochberg false discovery rate multiple testing correction ($p < 0.05$).

2.2.9 Measuring expression of HeLa genome variants

HeLa genome datasets were downloaded from the US National Institutes of Health (NIH) server. Data from Adey et al. (2013) included a haplotype resolved variant call format (VCF) file, containing reference and alternate alleles. The variants were originally called by comparing the HeLa genome to `GRCh37/hg19`, therefore variant positions were first lifted over to `GRCh38/hg38` since this was the genome used in the rest of the analysis. This was completed using LiftOver program (UCSC) and a pre-generated `GRCh37/hg19` to `GRCh38/hg38` `over.chain` file downloaded from hgdownload.soe.ucsc.edu/goldenPath/hg19/liftOver/.

Since this analysis was specifically focused on single nucleotide variants (SNVs), `.fastq.gz` files from total and nascent RNA-sequencing were re-mapped using STAR and a re-implementation of the allele-specific analysis method WASP (Van De Geijn et al., 2015) as an additional filter. WASP substitutes the SNV base with the alternative genotype in allelic reads and re-aligns those reads to correct for the reference bias. By excluding the allelic reads that are affected by different genotypes, WASP obtains extremely low false positive rate when identifying allele-specific expression SNVs. Only reads that passed WASP filtering were taken forward.

To find variants expressed in the total/nascent RNA-sequencing dataset, samtools mpileup was run on the STAR output `.BAM` files for each sample, using the STAR indexed genome file used for alignment and a position file containing the chromosome and genomic position of each variant from the `GRCh38/hg38` VCF file. The mpileup output files were then passed through a custom script written by Craig Anderson which counted the number of reads showing each of the four bases across each variant and, using these values, annotated each variant as being expressed as monoallelic (if only one base represented), biallelic (if two bases represented) or multiallelic (if three or more bases represented). Variants were also filtered for Phred quality score of 20 or above (equating to a 99% base

call accuracy) and indels removed. Instances of a variant falling in an intronic region, with a split read covering the region but not the intronic sequence, and also where the variant falls within the insert region between the two paired end reads were removed and subtracted from the coverage estimates.

Every variant was assigned a unique `variant_id` comprised of the chromosome number and the position on that chromosome, which was used henceforth. All nascent samples were compared and only variants expressed >10 reads in every replicate of both si-Scr nascent and si-UPF1 nascent were kept. Variants were further filtered to only keep those that were annotated as biallelic in at least one replicate of both si-Scr nascent and si-UPF1 nascent.

Heterozygous variants were identified from the `GRCh38/hg38` VCF by filtering for those variants where both the reference allele (REF) and the alternate allele (ALT) were present in the whole genome dataset, as opposed to homozygous for the alternative allele. Only variants that were annotated as biallelic in my RNA-sequencing data and heterozygous in the `GRCh38/hg38` VCF were taken forward. Finally, the variants were filtered to keep only those that had one read mapped in at least one replicate of both si-Scr total and si-UPF1 total.

2.2.10 Identification of NMD-targeted variants genome-wide

Differentially processed variants were initially identified using DESeq2 analysis of reference and alternate allele reads in both total and nascent from each replicate of si-UPF1 and si-Scr treated cells. DESeq2 was run with a three-factor design to account for all variables, and a reduced likelihood ratio test (LRT) to determine if the increased likelihood of the data using the extra terms in the full model was more than expected if those extra terms were truly zero. The following commands were run, where "condition" represents the siRNA treatment, "fraction" indicates total or nascent RNA and "allele" codes for REF or ALT nucleotide.

```
$ dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, design=
condition + fraction + allele + condition:fraction + condition:allele)
$ dds <- DESeq(dds, test="LRT", reduced=   condition + fraction + allele)
$ results <- results(dds)
```

Variants were determined to be differentially processed upon depletion of UPF1 if they were significantly changed after Benjamini-Hochberg false discovery rate multiple testing correction ($p < 0.05$).

Manual calculation of differential allele-specific processing was carried out by first dividing raw counts over each variant by library size to produce a transcripts per million (TPM) value. Using these values, an odds ratio (OR) was calculated for both si-Scr samples and si-UPF1 samples independently using the following equation:

$$OR = ((median\ ALT\ total\ TPM / median\ ALT\ nascent\ TPM) \\ /(median\ REF\ total\ TPM / median\ REF\ nascent\ TPM))$$

Variants behaving in line with an NMD-targeted variant were determined by a si-UPF1 OR at least 3-fold higher than si-Scr OR and a si-UPF1 OR between –1 and 1.

To analyse this data using DESeq2, a two-factor design was used to account for REF or ALT allele and total or nascent fraction. A reduced likelihood ratio test (LRT) was used to determine if the increased likelihood of the data using the extra terms in the full model was more than expected if those extra terms were truly zero. The following commands were run, where "allele" represents the REF or ALT allele and "fraction" represents the total or nascent fraction.

```
$ dds <- DESeqDataSetFromHTSeqCount(sampleTable = sampleTable, design=
~ allele + fraction + allele:fraction)
$ dds <- DESeq(dds, test="LRT", reduced= ~ allele + fraction)
$ results <- results(dds)
```

This was carried out for both si-Scr samples and si-UPF1 samples independently and Benjamini-Hochberg false discovery rate multiple testing correction applied ($p < 0.05$). Variants were determined to be differentially processed upon depletion of UPF1 if they were significantly changed upon si-Scr depletion but not significantly changed upon si-UPF1 depletion.

2.2.11 Characterisation of differentially processed variants

The ensembl ID for the gene(s) and specific transcript(s) each variant is located within, as well as the strand (+ or -) and co-ordinates of the transcript(s) were added to each variant using BEDTools intersect (Quinlan & Hall, 2010) and genomic intervals of transcripts from the genome gene transfer format (GTF) file used to create the STAR index. Official gene symbols were added to each variant using bioMart rest API (Ensembl). The predicted effect of all analysed variant was found by creating a new VCF file from the DESeq2 results and analysing with the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016) using the inbuilt `GRCh38/hg38` database.

2.2.12 Image analysis

Image analysis was carried out using the FIJI/ImageJ software (National Institutes of Health, Bethesda, MD, USA). All Z-stacks were flattened using maximum intensity projection. PLA signal was measured by thresholding using MaxEntropy settings and analysing particles to get a count per image. This PLA signal count was normalised by dividing by number of nuclei in each image. Significance of change in PLA signal was determined by two-tailed Mann-Whitney test.

2.2.13 Statistical analysis

For information about the number of replicates, the meaning of error bars (e.g., standard error of the mean) and other relevant statistical analysis see the corresponding figure legend. For information about how data was analysed and/or quantified, see the relevant section in methods and/or the figure legend. RStudio was used for all statistical analysis. Specific packages and algorithms used are listed in Section 2.2.15. Unless otherwise stated, a multiple testing corrected value of $p < 0.05$ was considered significant. Unless otherwise stated, multiple testing correction was performed using Bonferroni correction.

2.2.14 Deposited data used

Total, membrane-associated and cytoplasmic RNA-sequencing data described in this thesis have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession number GSE152437. Comparison of si-UPF1 differential expression targets with literature was carried out by downloading relevant supplementary data from (Tani et al., 2012; Lykke-Andersen et al., 2014; Schmidt et al., 2015). Overlap with experimental ER-localisation was carried out by downloading relevant supplementary data; APEX-seq (Fazal et al., 2019, GEO: GSE116008), ER fractionation sequencing (Reid and Nicchitta, 2012, PMID: 22199352) and ER proximity-specific ribosome profiling (Jan et al., 2014, PMID: 25378630). HeLa genome data from Adey et al. (2013) is maintained on the American National Institute of Health (NIH, USA) server. Access was granted by the director of the NIH in December 2018 and relevant files were downloaded directly from the server.

2.2.15 Software and packages used

Resource	Source	Identifier
FASTQC v0.11.7	Babraham Bioinformatics Team (2018)	RRID:SCR_014583
STAR v2.5.1b	Dobin et al. (2013)	RRID:SCR_015899
HTSeq v0.9.1	Anders et al. (2015)	RRID:SCR_005514
SAMtools v1.6	Li et al. (2009)	PMID: 19505943
BEDTools v2.27.1	Quinlan and Hall (2010)	RRID: SCR_006646
kallisto v0.43.1	Bray et al. (2016)	RRID:SCR_016582
R v3.4.1	R Core Team (2018)	https://www.r-project.org/
Tidyverse v1.3.0	Wickham et al. (2019)	https://www.tidyverse.org/
Bioconductor v3.6	Gentleman et al. (2004)	RRID: SCR_006442
DESeq2 v1.14.1	Love et al. (2014)	RRID:SCR_015687
Integrated Genomics Viewer	Robinson et al. (2011)	RRID:SCR_011793
biomaRt	Durinck et al. (2005)	RRID:SCR_002987
Database for Annotation, Visualization and Integrated Discovery (DAVID)	Dennis et al. (2003)	RRID:SCR_001881
Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GOrilla)	Eden et al. (2009)	RRID:SCR_006848
STRING	Szklarczyk et al. (2019)	RRID:SCR_005223
ImageJ	Schneider et al. (2012)	RRID:SCR_003070

Chapter 3

NBAS interacts with the ER translocon

3.1 Introduction

NBAS has been shown to be localised to the endoplasmic reticulum (ER) as part of the multi-protein Syntaxin-18 complex (Aoki et al., 2009), however, its mechanistic role in mRNA surveillance remains to be discovered. NBAS forms part of the NBAS-RINT1-ZW10 (NRZ) tethering complex which associates with the soluble NSF attachment protein receptors (SNAREs) Syntaxin-18, USE1, BNIP1 and SEC22B, and is believed to play a role in assembly of the multi-SNARE complex (Hatsuzawa et al., 2000; Burri et al., 2003; Hirose et al., 2004; Tagaya et al., 2014). In addition, both the α-solenoid at NBAS' C-terminal and its N-terminal β-propeller structure are generally involved in protein-protein interactions (Chen et al., 2011, Section 1.3.1). This suggests that NBAS is proficient at protein complex assembly, which is a fundamental process in the NMD pathway. Therefore, I wanted to know what other proteins NBAS interacts with, especially those involved in the NMD pathway.

Previous work has used transfected, tagged NBAS to study localisation and function (Longman et al., 2020, Section 1.3.4). Interaction studies show that overexpressed NBAS interacts with UPF1, preferentially in its unphosphorylated state (Longman et al., 2020) suggesting NBAS is involved early in the NMD pathway. In order to further investigate the function of NBAS in a more physiological context, I wanted to identify protein binding partners of endogenous NBAS using immunoprecipitation (IP) followed by mass spectrometry (IP-MS).

3.1.1 Endogenous tagging of protein allows optimised isolation and imaging

Whilst transfecting tagged protein allows simple and fast expression in a variety of cell types, interactions determined using endogenously tagged proteins are more likely to recapitulate physiological interactions than overexpressed protein, especially because excess protein could form complexes in a non-physiological

manner, since overexpression or additional expression of endogenous and transfected protein leads to altered stoichiometry, whilst tagging the endogenous protein leaves this unaltered. Antibodies against human NBAS are commercially available, however these are only recommended for western blot probing, whilst reagents for numerous biochemical experiments have been optimised for various protein tags, such as green fluorescent protein (GFP). Highly specific purification tools have been developed for GFP which guarantee an efficient enrichment of the bait. Endogenous tagging of the protein is also superior to transfection as it removes the complication of both tagged and WT protein being present and therefore allows the isolation of all NBAS proteins. Furthermore, since the tagged cells will be clonal, they provide a homogenous population with the same number of copies of NBAS gene as the WT cells.

Tagging of proteins with GFP has been used for decades to examining the subcellular localization of these proteins in living cells (Wang and Hazelrigg, 1994). Several hundred human cDNAs have since been cloned as N- or C-terminal gene fusions and this methodology is still used widely to enable visualisation and use of optimised reagents (Deal et al., 2020). Furthermore, the only available α-NBAS antibodies are all raised in rabbit whereas GFP antibodies are available from multiple species allowing greater compatibility with other antibodies for immunofluorescence. eGFP has a greater folding efficiency at 37°C (Cormack et al., 1996) making it more practical for use in mammalian cells. Other tags commonly used, such as FLAG and HA, are much smaller than eGFP and therefore are small additions to a eGFP-tag and unlikely to interfere with folding or function. Furthermore, a second tag allows an alternate method for isolating and probing the tagged protein, and permits tandem affinity purification. The 3xFLAG system is an improvement upon the original system by fusing three adapted tandem FLAG epitopes with improved affinity and expression. Detection of fusion proteins containing 3xFLAG is enhanced up to 200 times more than any other system (Hernan et al., 2000). Therefore, I decided to generate a cell line with endogenously eGFP-3xFLAG tagged NBAS.

HeLa cells are known to be genetically unstable and non-physiological in many ways (Mittelman and Wilson, 2013; Frattini et al., 2015). However, there are numerous advantages to using them. They are relatively amenable to transfection (Kumar et al., 2019) and to genome editing using CRISPR techniques (Koch et al., 2018). They also have a strong NMD response, unlike some other cell lines (Baird and Hogg, 2018). Finally, all previous work with NBAS and NMD in human cells, has used HeLa cells (Longman et al., 2013, 2020), so protocols have already been optimised for the cells and it is preferable for this work to be as comparable as possible.

Tags can be added to either the N- or C-terminal of the protein of interest. Often the C-terminal is targeted because this is the last part of the protein to be synthesised, so it limits the chance of the tag interfering with translation and protein localisation. However, the specific domains and functions of the target protein must also be considered. NBAS has been shown to bind p31 and Syntaxin-18 somewhere in the amino half of the protein, whereas the C-terminal half of NBAS was associated with ZW10 and RINT1 (Aoki et al., 2009, Section 1.3.1). It is not known what part of NBAS interacts with UPF1. However, the plasmid-based GFP-NBAS and FLAG-NBAS constructs previously used, both have the tag at the N-terminal. These constructs have been validated and shown to interact with UPF1. Therefore, I decided to add the tag to the N-terminal of endogenous NBAS.

3.1.2 CRISPR/Cas9 gene editing

CRISPR-associated RNA-guided endonuclease allows straightforward genome editing of virtually any genomic locus. The endonuclease proteins can be loaded with a short sequence of RNA, called a guide RNA (gRNA), which directs it to the complementary genomic destination to cleave a specific DNA target (Figure 3.1). There are multiple examples of similar creation of human cells stably expressing a C- or N-terminal GFP tag on the gene of interest using CRISPR technology, followed by α-GFP IP and mass spectrometry analysis of endogenous protein-protein interactions (Shen et al., 2009; Tachie-Menson et al., 2020).

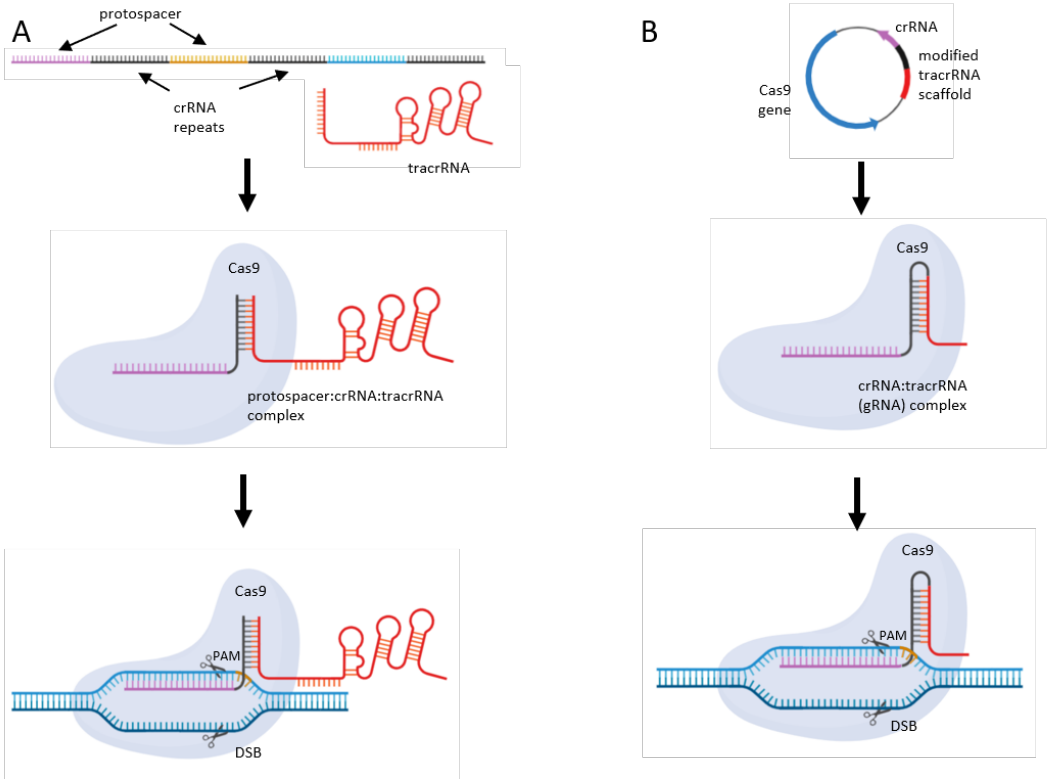


Figure 3.1: ***CRISPR/Cas9 type-II system from Streptococcus pyogenes*** ***(A)*** *The CRISPR/Cas system is a form of innate immune system in S. pyogenes. In type-II CRISPR systems, the RNA guide exists within a duplex formed of tracrRNA:crRNA that forms base pairs with a target DNA sequence. Once in contact with the target, the nuclease induces a double-stranded break (DSB) in the bound DNA. Importantly, for cleavage to occur the complex must recognise a short consensus region termed a Protospacer adjacent motif (PAM) site downstream of the base-paired DNA sequence to facilitate recognition of a correct target.* ***(B)*** *In the repurposed CRISPR/Cas9 toolkit, the modified crRNA and tracRNA are fused into a single chimeric RNA that has a designed sequence specific gRNA to direct nuclease activity. The RNA hybrid and Cas9 gene are transfected into target cells.*

Although there are various natural and engineered CRISPR/Cas systems, the first to be harnessed for gene editing applications and the most popular CRISPR/Cas methodology utilises a type-II system from *Streptococcus pyogenes* that utilises a Cas9 nuclease (Lewis and Ke, 2017). CRISPR is an acronym for "Clustered Regularly Interspaced Short Palindromic Repeats." This refers to the DNA repeats in the bacterial genome between which short fragments of viral genome are encoded as DNA, termed a protospacer, as a type of immune memory. Together the viral sequence and the protospacer make up the CRISPR RNA (crRNA) which forms a duplex with a trans-activating crRNA (tracrRNA). The entire RNA complex is called the guide RNA (gRNA). Once the Cas9 protein has been recruited, the gRNA forms base pairs with the target DNA sequence and the Cas9 nuclease induces a double-stranded break (DSB) in the bound DNA (Doudna and Charpentier, 2014). Importantly, for cleavage to occur the complex must recognise a short consensus region, termed the protospacer adjacent motif (PAM) site, downstream of the base-paired DNA sequence to facilitate recognition of the correct target.

In the repurposed CRISPR/Cas9 system, the bacterial tracrRNA and a custom, sequence-specific crRNA are fused into a single gRNA that can direct nuclease activity which is highly advantageous from both a technical and design standpoint (Cong et al., 2013; Ran et al., 2013a,b; Hsu et al., 2013, 2014). The sequence-specific crRNA must be 18-22nt in length and have perfect sequence complementarity with the genomic target site. The suitability of a genomic locus is dependent on the proximity of a PAM site to facilitate cleavage by Cas9. The canonical Cas9 PAM is the sequence 5'-NGG-3' (Anders et al., 2014). CRISPR/Cas9 cuts 4nt upstream from the PAM site.

After a DSB is made, the cell proceeds with DNA repair by one of two mechanisms: non-homologous end joining (NHEJ) or homologous directed repair (HDR). NHEJ ligates the free ends of DNA directly, often after paring back free ends, leading to repair in an error prone manner and frequent occurrence of insertions or deletions (indels) (Bothmer et al., 2013, 2017). In contrast, HDR requires a homologous template to guide repair of the break, often the sister chromosome

during S/G2 phase of the cell cycle. This pathway proceeds with high fidelity although it can still lead to errors such as the use of sequence-similar regions distinct from the cut site being used as a template. Nevertheless, HDR can be harnessed for insertion of genetic sequence into a specific target site by providing the cell with a repair template that has sequence complementarity to the targeted genomic locus and includes the desired extra genetic information.

3.1.3 Aims

In this chapter I aim to facilitate an understanding of NBAS function in NMD by creating a cell line with an NBAS protein that can be directly visualised and effectively isolated.

To do this, I will design CRISPR/Cas9 gene editing tools to target the N-terminal of endogenous NBAS and integrate eGFP and 3xFLAG tags. I will use these tools to target HeLa cells to generate a cell line endogenously expressing eGFP-3xFLAG-NBAS. To identify the putative protein binding partners of NBAS, I will characterise the NBAS interactome by performing immunoprecipitation coupled with mass spectrometry analysis (IP-MS) and generating interaction profiles. I will also treat cells with RNase prior to IP-MS to determine if protein interactions are RNA dependent and identify direct protein binding partners. I will validate novel interacting partners identified by IP-MS using the imaging technique proximity specific ligation (PLA).

In summary, the main aims of this chapter are:

a) To design and optimise a CRISPR/Cas9-based strategy to create an endogenously tagged NBAS cell line

b) To investigate the localisation and interacting proteins of NBAS in a cell system

3.2 Results

3.2.1 CRISPR design and strategy

Transiently transfected GFP-NBAS and 3xFLAG-NBAS have both been used previously (Longman et al., 2013, 2020), with both constructs tagging NBAS at the N-terminus. These transfected constructs are not cytotoxic or lethal and do not disrupt NBAS binding with UPF1 or it's localisation to the ER. Although C-terminal tags have traditionally been more popular since they are less likely to interfere with transcription or translation, N-terminal GFP tags have been used successfully to tag many endogenous proteins (Hoffman et al., 2019; Ohashi et al., 2020). Consequently, I designed the CRISPR strategy to target the endogenous tag to the N-terminus of the NBAS protein in this cell line (Figure 3.2A).

To add the eGFP-3xFLAG tag to the NBAS locus, the CRISPR/Cas system with the Cas9 nuclease enzyme was used. The eGFP tag allows for direct visualisation of the protein and in conjunction with well validated α-GFP antibodies (Peters et al., 2020) can be used for protein purification based assays. A 3xFLAG tag was also incorporated since it is a small tag compared to eGFP and would allow an alternate method for isolating tagged NBAS.

Four sequence-specific crRNAs were designed to target the NBAS N-terminus (Figure 3.2B). The crRNAs were designed with the help of two online tools; sgRNA Designer CRISPRko (Doench et al., 2016) and the Cas-Designer (Bae et al., 2014), using 100 base-pair genomic DNA sequence around the NBAS start codon retrieved from the Ensembl genome browser (Yates et al., 2020). Both tools give suggested crRNA sequences based on PAM sequences in the provided sequence and rank them based on a variety of parameters including off-target matches, on-target efficacy score and number of mismatches and GC content. Results were compared between tools and crRNAs ranked by scores from each output. crRNAs were selected taking into account these scores as well as the proximity of the cut site to the start codon, since that was where the insert was

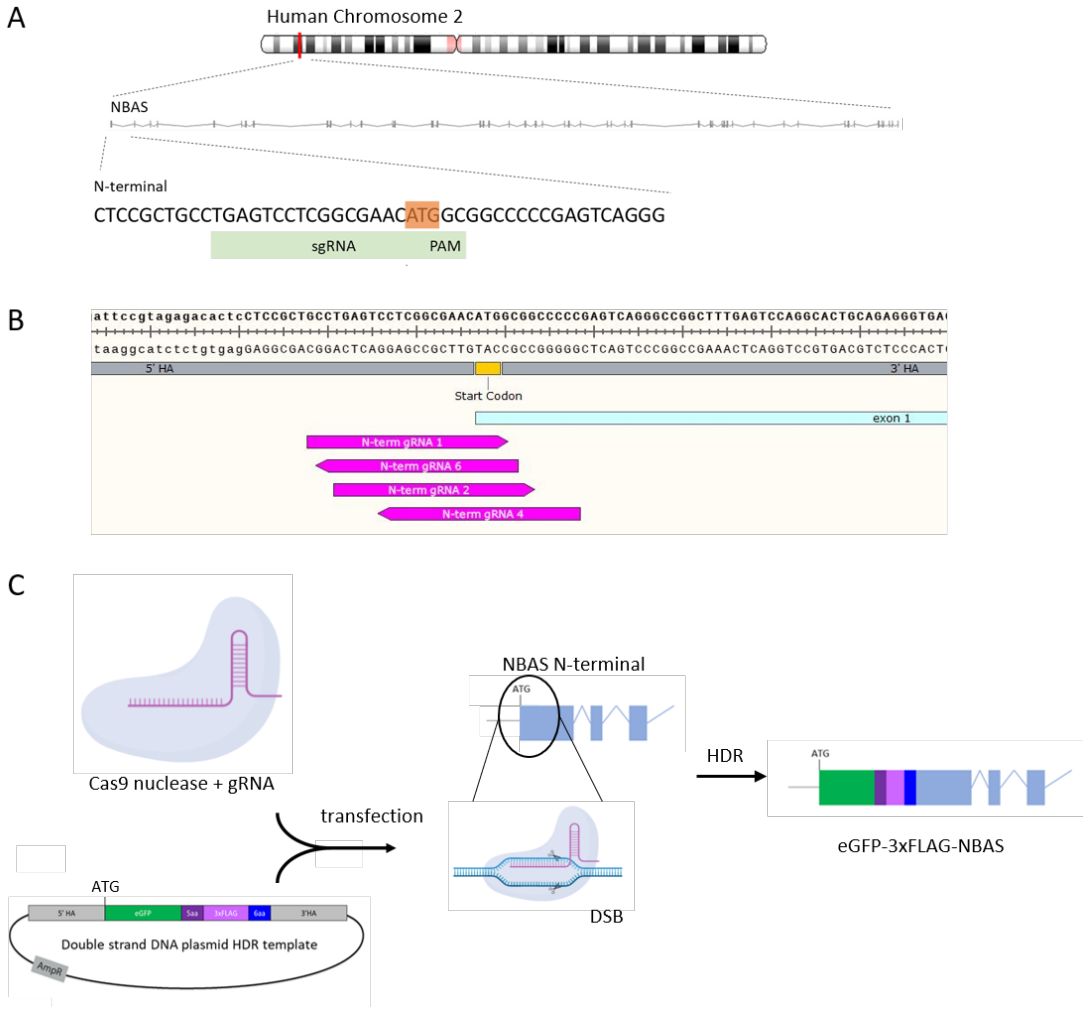


Figure 3.2: ***CRISPR/Cas9 strategy to target the NBAS genomic locus*** ***(A)*** *gRNAs were designed to target the N-terminal of the NBAS locus with a cut site close to the start codon.* ***(B)*** *The four gRNAs chosen from the CRISPR tools all cut near the NBAS start codon.* ***(C)*** *Summary schematic of targeting design used to generate clones.*

targeted and the cut site should not be more than 10nt from the site of repair for best HR efficiency. It was also considered whether the PAM could be mutated so the CRISPR/Cas9 construct would not re-cut, without altering the amino acid sequence. Four crRNAs were selected, and BbsI-compatible overhangs added to both ends to aid integration into the Cas9 vector (crRNA sequences are listed in Section 2.1.18).

The design of the repair template is essential to successful HDR. Short amino acid linkers were included between the eGFP and 3xFLAG tags and between 3xFLAG and NBAS exon one (Figure 3.2C; Appendix A), to prevent misfolding by ensuring physical space between the peptides of each protein. These two linkers had been previously used and validated (Nele Hug & Michaela Raab, personal communication). Homology arms flanking the double-tag-linker sequence aid homology directed repair (HDR) of the new DNA into the genomic locus. Homology arms of length 400 nucleotide were deemed necessary as a consequence of the large size of the tag (Chan et al., 2007a). Due to the size of the repair template, it was synthesised as a complete sequence within a standard plasmid (IDT).

3.2.2 Using CRISPR to create a stable tagged-NBAS cell line

The four crRNAs were cloned into a pSpCas9(BB)-2A-Puro (PX459) V2.0 plasmid (Ran et al., 2013b) containing the scaffold section of the gRNA and Cas9 nuclease open reading frame, as well as a puromycin resistance gene (Figure 3.3A). Plasmids containing each of the four gRNAs were separately transfected into HeLa cells to allow comparison of the cutting efficiency of each guide. Genomic DNA (gDNA) was extracted from the transfected cells and PCR amplified using custom primers KJJ_5 and KJJ_6 (primer sequences detailed in Section 2.1.18) targeting the NBAS N-terminal region. The PCR products were run on an agarose gel to visualise the size of the products produced (Figure 3.3B, C). The amplified region was 655nt (lane 2 - uncut control). Any cutting by the Cas9 guided by the specific gRNA would lead to smaller bands or smears of various

sizes of gDNA. These are caused by cutting of the gDNA followed by deletions of differing lengths followed by re-annealing of the gDNA ends, in a heterogeneous manner. Cells treated with gRNA #1, #4 and #6 showed smears of smaller PCR products as well as a band at the uncut size. Whereas, gRNA #2 clearly had the strongest effect as no gDNA appeared to remain full length.

To confirm these results, and quantify the cutting efficiencies, the PCR products were subjected to Sanger sequencing. The resulting sequence reads contain mixed signals due to the mixed population of cells used in the PCR reaction. Due to the different sized deletions produced at the cut site, there is a loss of phase synchrony of the fluorescent signals during the Sanger sequencing resulting in a mixed signal. Using the online tools Inference of CRISPR Edits (ICE; Synthego) and Tracking of Indels by DEcomposition (TIDE Brinkman et al., 2014), it is possible to compare the heterogeneous cut sequence with the uncut control, to analyse the cutting efficiency and effects of each gRNA. The two tools use different algorithms to achieve the same outcome. Each algorithm reconstructs the spectrum of indels from the mixed sequence traces. The identity of the detected indels and their frequencies are reported as well as a cutting efficiency (Figure 3.3D). The identity and frequency of indels would be more important if a loss of function mutation was the desired outcome. In this case cutting efficiency was the most important parameter. In general, both tools were proficient at identifying the cut site of each gRNA. ICE gave similar cutting efficiencies for gRNAs #1, #4 and #6 (34%-38%) but a higher efficiency of 47% for gRNA #2. TIDE gave the highest efficiency score (65.8%) to gRNA #4 and failed to report a score for gRNA #2. However, TIDE is not as efficient as ICE when dealing with complex indels (Martin Reijns, personal communication), which the highly smeared DNA gel (Figure 3.3C) suggests may be the case for gRNA #2. I therefore decided to carry out all further targeting with gRNA #2.

With the exact cut site and PAM sequence determined, the repair template sequence could be finalised and ordered. The PAM site of gRNA #2 overlaps with the original start codon so the inserted tag would mean the target sequence was disrupted and would no longer be recognised by the gRNA. A new ATG start

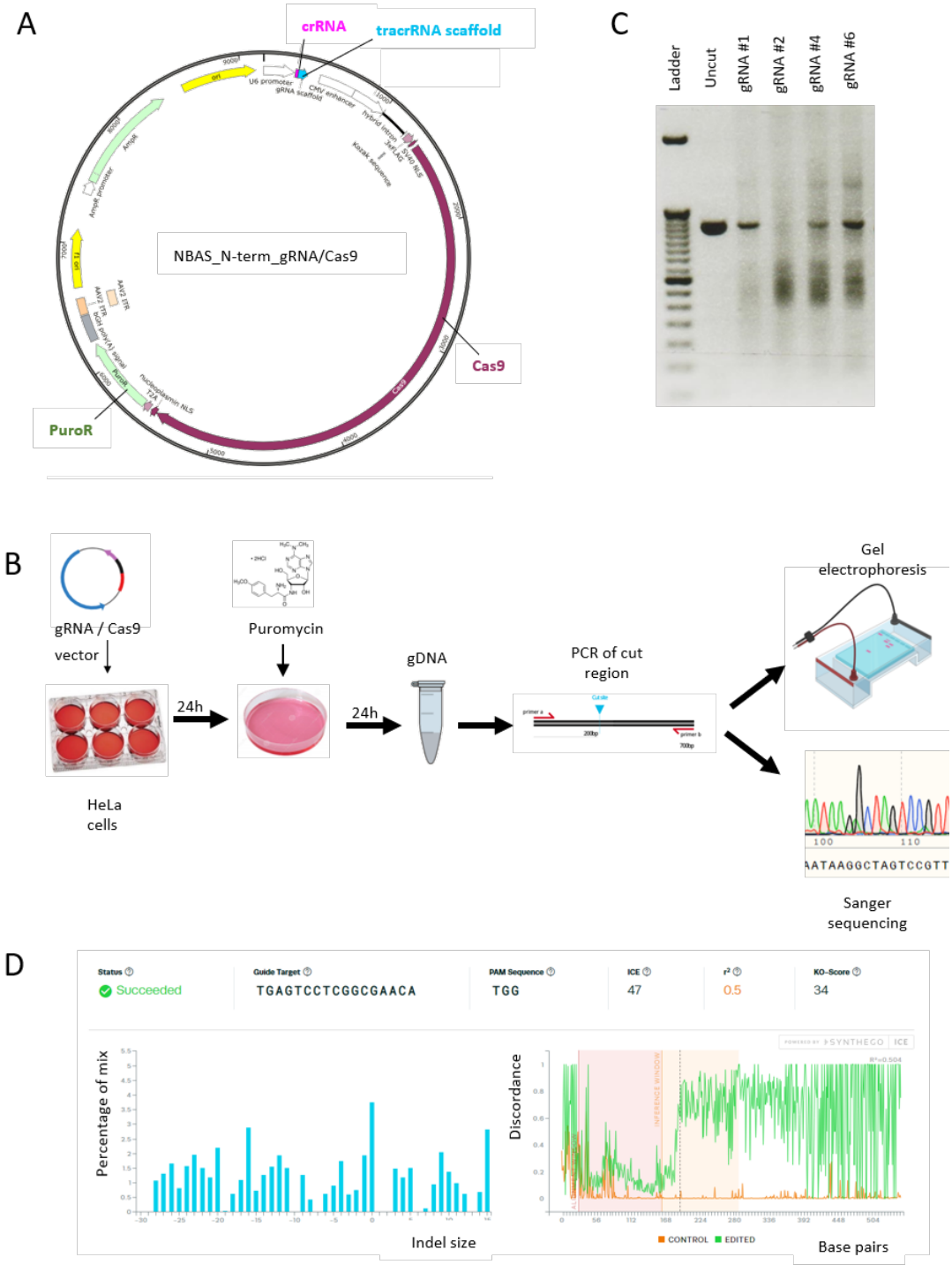


Figure 3.3: ***Assessment of Efficiency of CRISPR gRNA***

Figure 3.3: ***Assessment of efficiency of CRISPR gRNAs***

(A) *gRNAs were cloned into a PX459 plasmid containing the scaffold section of the gRNA and Cas9 nuclease open reading frame. The plasmid also contains a puromycin resistance cassette.* ***(B)*** *After transfection of the gRNA/Cas9 plasmid, cells were selected by treatment with puromycin. Once >90% control cells were dead, gDNA was extracted and the NBAS N-terminal locus amplified by PCR. PCR products were analysed by gel electrophoresis or Sanger sequencing.* ***(C)*** *Agarose gel showing PCR products from uncut control and cells trated with each of the four gRNAs.* ***(D)*** *Example output from Synthego ICE online tool. The right plot shows the discordance value of Sanger trace from uncut control and cells treated with gRNA #2. The dotted line indicates the estimated cut point. The left plot shows estimated frequency of indels of various sizes. The ICE score is an indicator of how efficient the gRNA is.*

codon was added before the eGFP and 3xFLAG sequences (Figure 3.4A). Specific homology arms were added that matched the sequence 400nt either side of the cut site. The full template was custom synthesised (Integrated DNA Technologies Inc.) and provided within a pUCIDT-AMP plasmid backbone (Appendix A).

The gRNA/Cas9 plasmid and the repair template plasmid were co-transfected at a ratio of 2:1 into HeLa cells followed by selection for successful transfection using puromycin treatment (Figure 3.4B). Once >90% of control mock treated cells were deemed dead by visual inspection, transfected cells were allowed to recover in standard media for 24h. Cells were then sorted using fluorescence-activated cell sorting (FACS) gated on GFP expression level (Figure 3.5A) to deposit single GFP-positive cells into 96 well plates. Cells transfected with only the repair template plasmid were used to set the GFP expression threshold. Only 0.1% of cells met the GFP threshold, nevertheless 1152 single cells were seeded across two rounds of screening.

Cells were then grown into colonies with media changes every 48h for ∼2 weeks, and 43.7% of single cells resulted in colonies. All wells containing colonies were dispersed using trypsin, split into replicate plates and allowed to recover for 24h. One set of cells were frozen in freezing media, whilst the matching set were lysed,

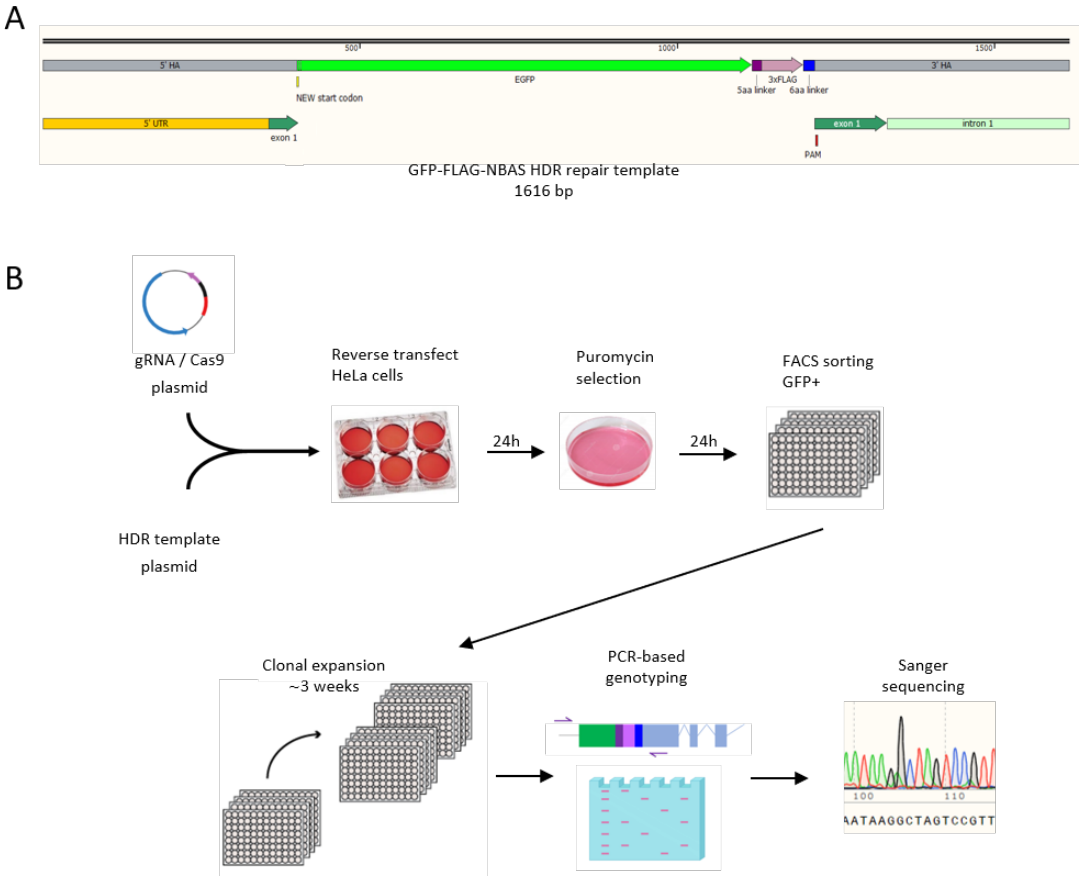


Figure 3.4: ***Overview of workflow used for CRISPR targeting strategy*** ***(A)*** *Final HDR template including insertion of new start codon.* ***(B)*** *Successfully transfected cells were selected for using fluorescence-activated cell sorting (FACS) for GFP+ cells, normalised to HDR-plasmid only cells. Clones were expanded from single cells over a maximum of three weeks, depending on growth rate, before genomic DNA was screened for the insert using PCR and downstream sequencing analysis to verify clones.*

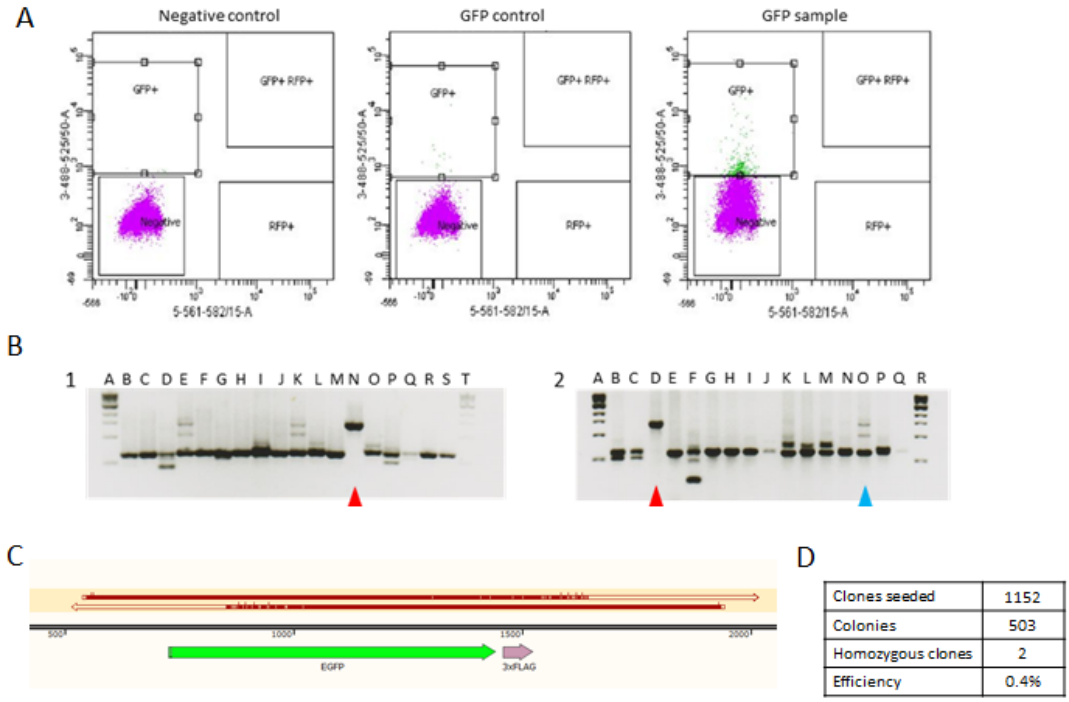


Figure 3.5: ***Clonal selection and screening of CRISPR targeted cells*** ***(A)*** *Cells targeted using CRISPR to insert an eGFP tag were sorted using fluorescence-activated cell sorting (FACS) gated on GFP expression level. Control cells were untreated, whilst HDR-only cells were transfected with the GFP-containing repair template but not the gRNA/Cas9 plasmid.* ***(B)*** *Example screening gels showing products of PCR over the N-terminal of the NBAS locus run on agarose gel; panel 1 and panel 2. Ladders are shown in panel 1 lanes A and T, and panel 2 lanes A and R. Bands can be seen at expected sizes for WT (e.g. panel 1 lane B, panel 2 lane H), full tag incorporation (panel 1 lane N and panel 2 lane D, red arrows) as well as various intermediates. Clone taken forward as "heterozygous" can be seen in panel 2 lane O, blue triangle.* ***(C)*** *Sanger sequencing reads showing accurate insertion of complete tag with no mutations in clone from lane 1N in panel C.* ***(D)*** *Number of GFP positive clones seeded, that successfully formed colonies and that were confirmed to be homozygous for the full tag. Efficiency is calculated as the number of homozygous clones as a percentage of the total number of GFP+ cell seeded.*

and genomic DNA (gDNA) extracted. Using the same custom primers KJJ_5 and KJJ_6, the NBAS N-terminal locus from each clone was amplified by PCR and the resulting products run on agarose gel to visualise the DNA fragment sizes (Figure 3.5B). WT product was 655nt and full incorporation of the tag by HDR would yield a product of 1468nt. Most clones gave bands discernible as WT. Two clones with one larger band had no WT band present (Figure 3.5B; panel 1 lane N and panel 2 lane D, red arrows). No samples cleanly produced bands at both WT and full tag sizes, however several clones produced bands at WT and full tag sizes but also intermediate sizes (eg. panel 1 lanes E and K, panel 2 lane O). To confirm sequences, both putative homozygous and two of the "heterozygous" clones were Sanger sequenced (Figure 3.5C). Both homozygous clones were shown to have complete tags incorporated with no mutations. The "heterozygous" samples resulted in mixed signals suggesting the presence of different PCR products from different alleles. Cytogenetic analysis of HeLa cells by spectral karyotyping revealed the presence of three copies of chromosome 2 (Macville et al., 1999). This was also supported by sequencing over the NBAS locus which also suggests three copies (Landry et al., 2013). However, since HeLa cells are genomically unstable, it is not known how many copies of chromosome 2 these HeLa cells have or indeed how many copies of the NBAS locus. Therefore, it was concluded that in the "heterozygous" cells one or more alleles were untargeted, one or more targeted efficiently and one or more inefficiently targeted. Of the 1152 puromycin resistant, GFP positive cells seeded, only two homozygous clones were detected. This represents a very low efficiency of 0.4% (Figure 3.5D), since the cutting efficiency of the gRNA/Cas9 was high (Figure 3.3C), this is assumed to be due to low HDR efficiency.

The two homozygous clones and one "heterozygous" clone were thawed, and cells allowed to recover for 24h. The "heterozygous" clone, and only one of the two homozygous clones recovered from the freeze-thaw process. Whole cell lysates were made and visualised by western blot (Figure 3.6A). Probing with α-NBAS antibody revealed that both the homozygous and the "heterozygous" clones expressed only tagged NBAS protein at detectable levels. Tagged NBAS could

also be detected by α-GFP and α-FLAG antibodies. Tagged NBAS produced a band at the expected size, $\sim$30 kDa larger than WT control NBAS. Probing with α-GFP shows successful integration of the GFP tag in both homozygous and "heterozygous" cells. The different sensitivities of α-NBAS and α-GFP antibodies can clearly be seen, confirming the greater sensitivity and stronger signal of α-GFP compared with α-NBAS. NBAS protein expression appears to be higher in the homozygous cells than the "heterozygous" cells, consistent with only some of the NBAS alleles being successfully targeted in these cells. However, expression is not conclusive as western blots are not quantitative.

GFP expression, quantified by flow cytometry (Figure 3.6B), showed narrow peaks for both "heterozygous" and homozygous cell populations, suggesting homogenous, clonal populations of cells. Homozygous GFP expression was distinctly higher than heterozygous which in turn was higher than WT, in agreement with higher expression of GFP-NBAS protein in homozygous cells, suggested by western blot. At this point, I decided to only take the homozygous cells forward for exploratory experiments. Henceforth, these cells will be referred to as "GFP-NBAS cells".

3.2.3 Validation of GFP-NBAS cell line

Immunoprecipitation (IP) using α-GFP antibody followed by western blot (IP-WB) resulted in a strong enrichment of GFP-NBAS in the IP sample compared to input samples (Figure 3.7A) showing efficient pull down of GFP-NBAS using α-GFP beads. This blot also showed the stronger signal of the commercially optimised α-GFP antibody compared to the α-NBAS antibody. A single clean band was seen with both antibodies indicating GFP-NBAS cells express no WT NBAS or independent GFP protein. To further check for off-target integration of the GFP tag, an IP-WB was performed on the GFP-NBAS cells and the whole membrane probed with α-GFP antibody. As well as taking images at optimal exposure, I also overexposed the membrane to detect faint bands (Figure 3.7B). Although there were several faint bands, all of these were also present in the parental HeLa cell line lanes, suggesting weak off-target binding of the α-GFP

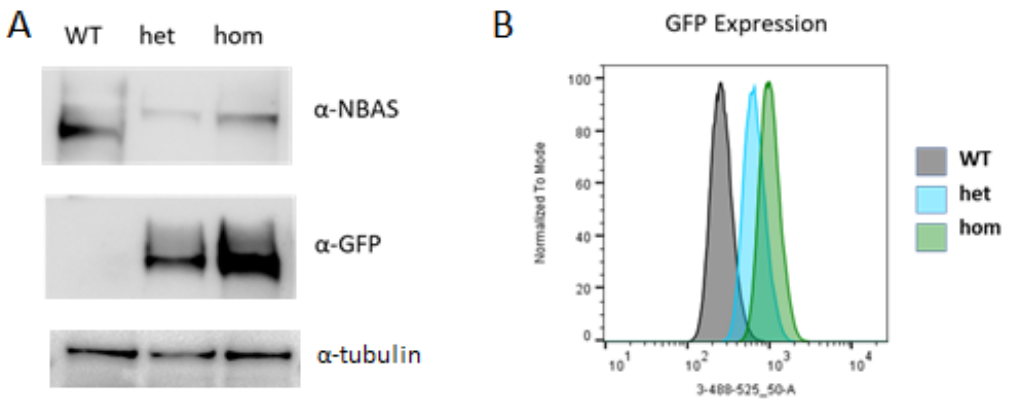


Figure 3.6: ***Comparison of homozygous and heterozygous GFP-FLAG-NBAS cell line***

(A) *Western blot of whole cell extract from CRISPR targeted homozygous (hom) and genetically "heterozygous" (het) clones, and the WT parental HeLa cell line. Probing with α-NBAS shows both het and hom cells express only one form of NBAS which is ∼30 kDa larger than WT NBAS. Probing with α-GFP shows successful integration and expression of the eGFP tag in both het and hom cells. The different sensitivities of α-NBAS and α-GFP antibodies can clearly be seen. NBAS protein expression appears to be higher in hom cells than het cells.* ***(B)*** *Flow cytometry measurement of GFP fluorescence of WT, het and hom cells. Both het and hom result in narrow peaks indicating clonal populations. GFP expression in hom cells is higher than het cells, in agreement with higher expression of GFP-NBAS protein in hom cells.*

antibody, rather than off-target integration of the GFP-tag. Whilst this result does not confirm the absence of off-target integration into the genome which could be non-expressed or non-translated, it does show that there is not substantial off-target protein expression of the GFP antigen recognised by α-GFP, especially since the other bands were all much fainter than the GFP-NBAS band. Therefore, I judged the binding of the α-GFP antibody was suitably specific to NBAS.

To visualise GFP-NBAS within the cells, cells were subject to imaging, initially without antibodies, showing that it was possible to detect native GFP-NBAS (Figure 3.7C). The signal was similar in location and dispersion but clearer and stronger when α-GFP and an appropriate secondary antibody were used, and so all further imaging was done following immunofluorescence (IF) protocol for visualising GFP-NBAS.

GFP-NBAS was located mainly in the perinuclear region, consistent with ER localisation (Aoki et al., 2009) (Varma et al., 2006), but to explore this further I also probed for calnexin, an ER marker. GFP-NBAS showed a clear overlap of signal with calnexin (Figure 3.8A). To directly test ER membrane localisation, cells were treated with the drug digitonin before fixing, to permeabilise the plasma membrane but leave the ER membrane intact (Niklas et al., 2011). Thus, only proteins tethered to the ER remain, and free protein disperses and is washed off during the IF protocol. Upon digitonin permeabilisation, GFP-NBAS signal was slightly reduced but still clearly present and overlapping with the signal of the ER marker SEC61B (Figure 3.8B), confirming GFP-FLAG-NBAS is appropriately localised to the ER.

I next wanted to confirm whether GFP-FLAG-NBAS reproduced the previously identified interaction between NBAS and UPF1 (Section 1.3.4). Proximity ligation assay (PLA) is an antibody-based technique that uses specific antibodies against the two proteins of interest and secondary antibodies with either a plus (+) or minus (-) DNA oligo attached (Figure 3.9A) (Söderberg et al., 2006). When the secondary antibodies are less than approximately 40nm apart, the oligos can anneal and rolling DNA synthesis can take place. These nascent DNA clusters can be visualised with fluorescent marker which is imaged as a red spot.

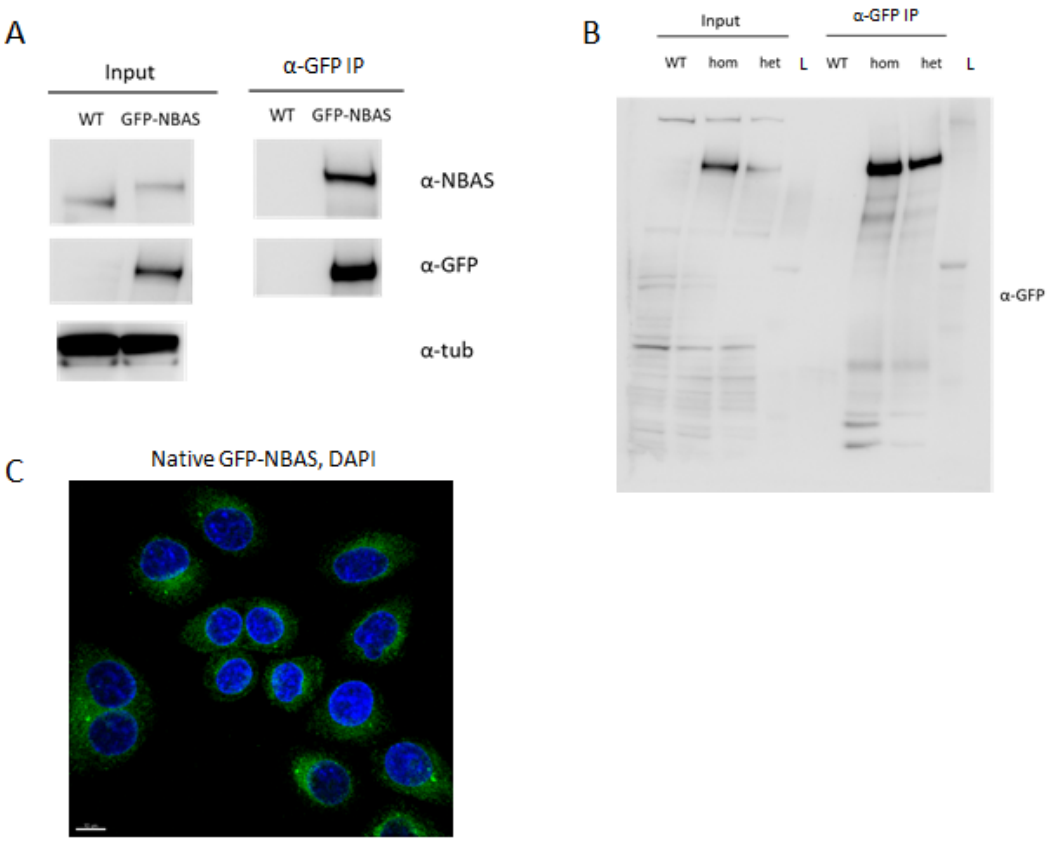


Figure 3.7: ***Validation of homozygous GFP-FLAG-NBAS cell line*** ***(A)*** *Immunoprecipitation with α-GFP antibody of GFP-NBAS CRISPR cells and WT parental HeLa cells. Western blot was probed with α-NBAS, α-GPF and α-tubulin antibodies and appropriate secondary antibodies.* ***(B)*** *Immunoprecipitation with α-GFP antibody of homozygous and genetically "heterozygous" GFP-NBAS CRISPR cells and WT parental HeLa cells. Whole membrane was probed with α-GPF antibody and appropriate secondary antibody.* ***(C)*** *Confocal image of native GFP-NBAS signal. Nuclei stained with DAPI.*

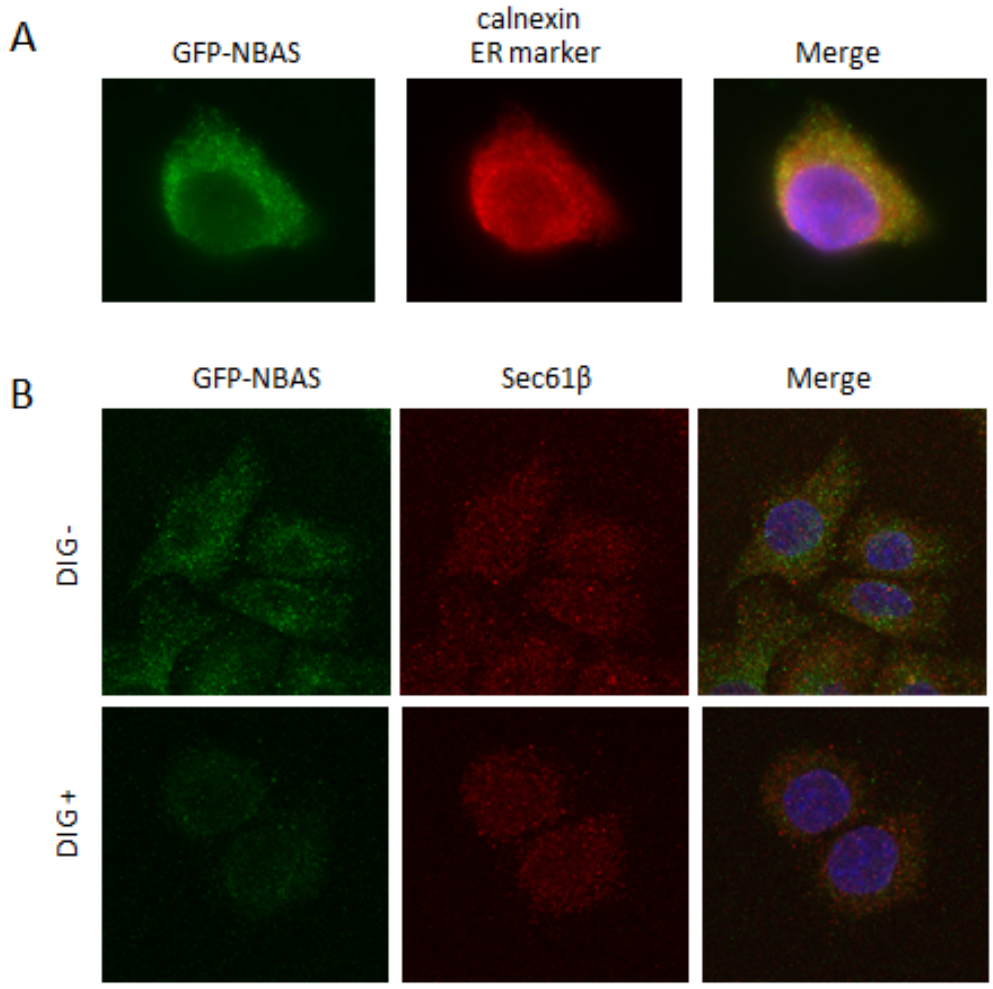


Figure 3.8: ***GFP-FLAG-NBAS is localised to the ER***

(A) *Wide field immunofluorescence of GFP-NBAS and the ER marker calnexin. Nuclei stained with DAPI.* ***(B)*** *Confocal immunofluorescence of GFP-NBAS and the ER translocon component SEC61B. Digitonin (DIG)+ cells were treated with digitonin permeabilisation agent prior to fixation. DIG- control cells were mock treated with PBS. Nuclei stained with DAPI.*

To test the interaction of the endogenous proteins, I performed PLA with antibodies recognising UPF1 and GFP (Figure 3.9B). PLA signal was quantified as number of red spots in a single frame divided by the number of nuclei. There was a significant increase in PLA signal in the GFP-NBAS cells compared to the WT control parental HeLa cell line, indicating GFP-NBAS was less than 40nm from UPF1, suggesting direct interaction (Figure 3.9B, C). Altogether these results suggest that the addition of the GFP-3xFLAG tags hasn't disrupted the known localisation or interactions of endogenous NBAS.

3.2.4 GFP-NBAS interacts with Syntaxin-18 complex and also the ER translocon

To investigate global protein binding partners of NBAS, the GFP-NBAS cells were subjected to immunoprecipitation-mass spectrometry (IP-MS). Efficient pull down of GFP-NBAS using α-GFP beads was established (Figure 3.7A). A stronger signal in the pull-down compared to input showed there was an enrichment of NBAS in the immunoprecipitate. Therefore, I was confident to proceed with IP-MS. IP of GFP-NBAS cell lysate and the parental HeLa cell lysate was carried out using α-GFP beads as before, and resulting proteins digested into peptides. Samples were then submitted to the IGMM Mass Spectrometry core facility who carried out MS processing and analysis of raw data. To identify binding of proteins to GFP-NBAS, the abundance of each protein in the GFP-NBAS cells sample was normalised to the control cells sample to give a fold change and significance value. This was carried out by Jimi Wills of the IGMM Mass Spectrometry core facility, using Wasim Aftab's LIMMA Pipeline Proteomics (Section 2.1.13).
The most significantly enriched protein identified was NBAS ($p = 5.8\text{x}10^{-8}$), confirming success of the experiment (Figure 3.10A). The next twelve most enriched proteins detected have all been previously identified as part of the Syntaxin-18 complex (Figure 3.10B; Section 1.3.2), including both RINT1 and ZW10, the two proteins that with NBAS make up the NBAS-RINT1-ZW10 (NRZ) complex, confirming not only GFP-NBAS is in the correct subcellular location but that it is interacting with the expected complexes. Binding studies show that the N- and

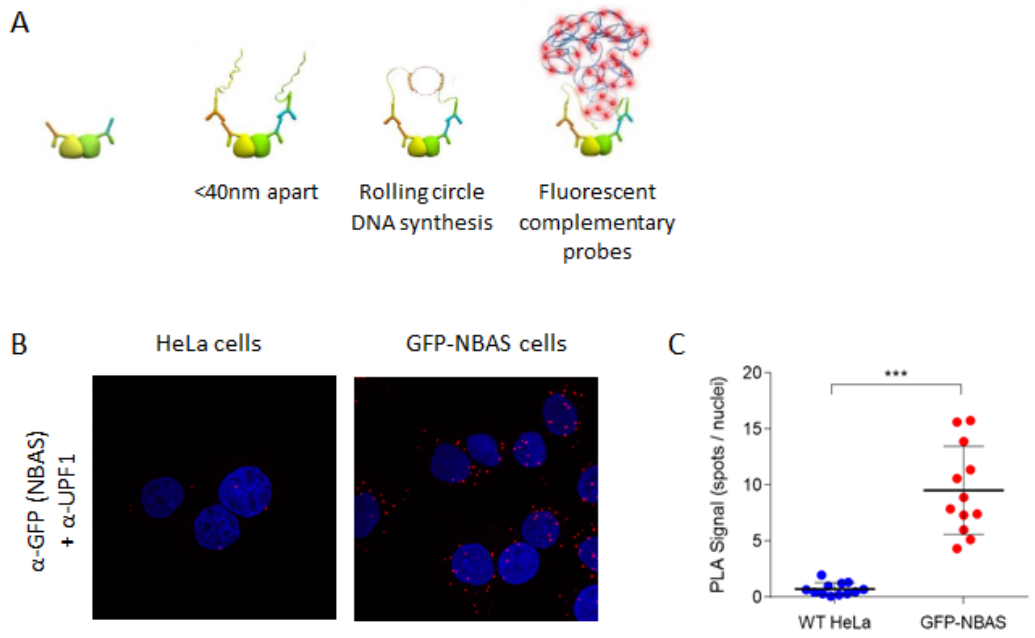


Figure 3.9: ***Confirmation of interaction between GFP-FLAG-NBAS and UPF1***

(A) *Proximity ligation assay (PLA) is an antibody-based technique that uses specific antibodies against the two proteins of interest and secondary antibodies with either a plus (+) or minus (-) DNA oligo attached. When the secondary antibodies are less than approximately 40nm apart, the oligos can anneal and rolling DNA synthesis can take place. These nascent DNA clumps can be visualised with fluorescent marker which is imaged as a red spot.* ***(B)*** *Representative PLA signal (red) when antibodies against GFP and UPF1 were used in GFP-NBAS cells and the parental HeLa cell line. DAPI staining shows cell nuclei (blue).* ***(C)*** *Quantification of all replicates of PLA in parental WT HeLa cells and tagged GFP-NBAS cells. Each spot is the PLA signal in a single image, calculated as the number of red spots divided by the number of nuclei present. The mean and standard deviation are indicated by the bold line and whiskers. A significant increase in PLA signal was seen in the GFP-NBAS cells compared to WT HeLa cells (two-tailed Mann-Whitney test; *** indicates* $p < 0.001$*).*

the C-terminal regions of NBAS interact with p31 and Syntaxin-18, and ZW10 and RINT-1, respectively (Civril et al., 2010; Tagaya et al., 2014). Therefore, that interactions with all four of these proteins are detected, suggests that the N-terminal tagged NBAS is able to interact correctly with its established partners. No known NMD factors were identified as significantly enriched.

In addition to identifying known binding partners of NBAS in the Syntaxin-18 complex, there was also a significant enrichment of SEC61A (log(fold change)=2.66, p=0.035; Figure 3.10C), a component of the trimeric SEC61 complex which makes up the core channel of the ER translocon where the ribosome docks when translating directly into the ER lumen (Hartmann et al., 1994) (Figure 3.10D; Section 1.1.4). In addition to these robust interactions, there was a more suggestive (non-significant) enrichment of SEC61B another subunit of the SEC61 translocon channel (Figure 3.10D), SEC63; which is thought to work with the SEC61 complex in post-translational protein translocation into the ER, signal recognition particle 72 (SRP72), and signal recognition particle receptor (SRPR); which are involved in transcript recruitment to the ER translocon for translation there (Section 1.1.4). Analysis with the protein network tool STRING (Szklarczyk et al., 2019) shows that these translocon complex components have previously been shown to interact with members of the Syntaxin-18 complex (Figure 3.10E). Although, STRING uses multiple sources and most of the interactions between the Syntaxin-18 complex and the translocon components are based on co-expression data (black lines). However, the interaction between the Vesicle Fusing ATPase NSF and SEC61A has been experimentally determined (pink line). NSF is required for vesicle-mediated transport and functions as part of the COPI-dependent Golgi-to-ER retrograde traffic system.

IP-MS was also performed after treating GFP-NBAS cell lysates with RNase, to identify whether identified GFP-NBAS interactions are RNA-dependent. Calculating fold change of proteins in RNase- GFP-NBAS compared to RNase+ GFP-NBAS showed that the most significantly depleted proteins were all RNA binding proteins (RBPs), indicating success of the experiment (Figure 3.11A).

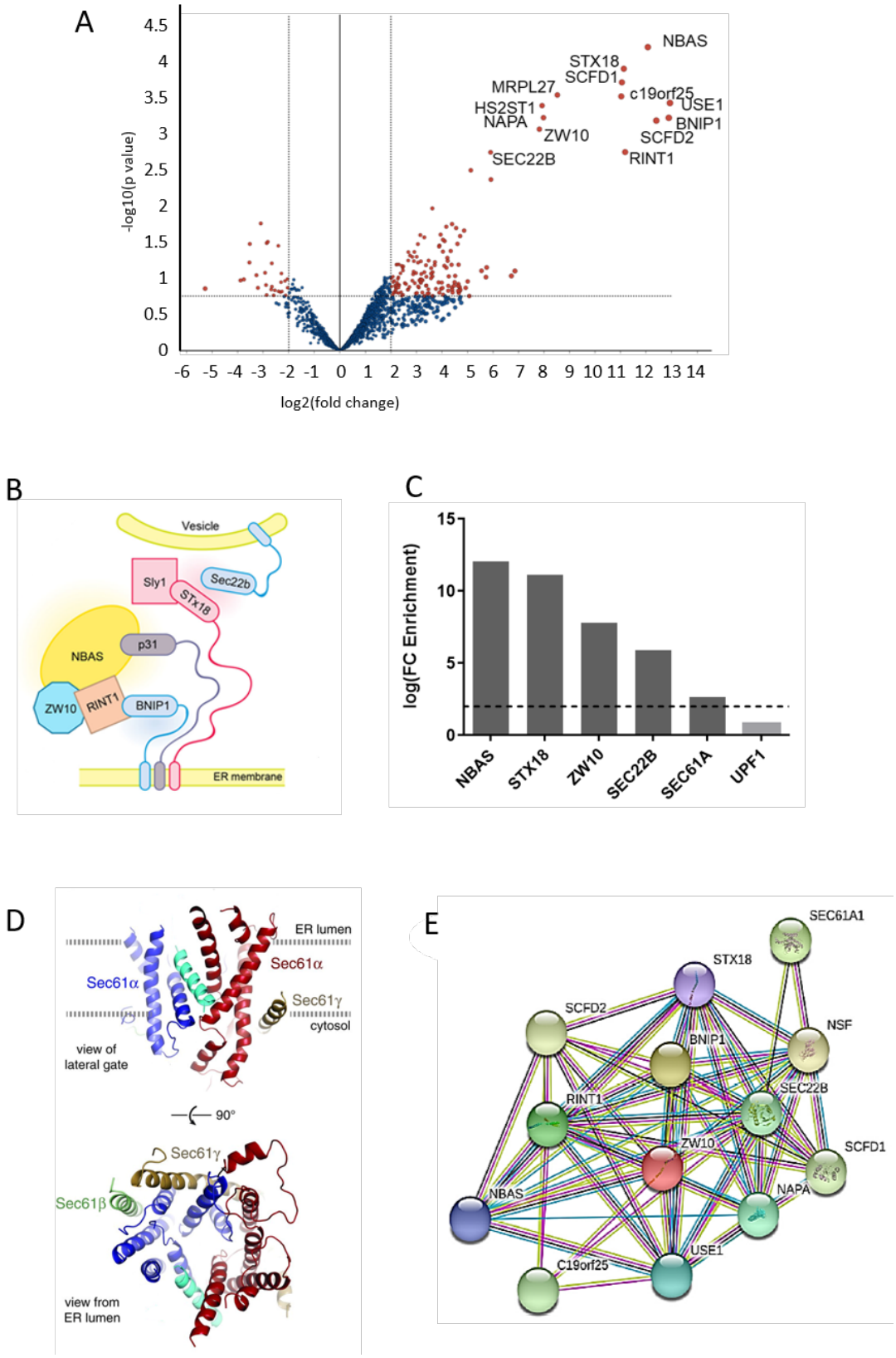


Figure 3.10: ***GFP-NBAS interacts with Syntaxin-18 complex and also the ER translocon***

Figure 3.10: ***GFP-NBAS interacts with Syntaxin-18 complex and also the ER translocon***

(A) *Enrichment of proteins from mass spectrometry of α-GFP immunoprecipitation of GFP-NBAS cells compared to WT parental HeLa cell line. Top 13 most significantly enriched proteins are labelled, all are associated with the Syntaxin-18 complex. Dotted lines indicate thresholds for significance ($p < 0.05$) and enrichment (>2-fold increase or decrease in detection).* ***(B)*** *Schematic of the NBAS-RINT1-ZW10 (NRZ) complex within the larger Syntaxin-18 complex located at the endoplasmic reticulum (ER) membrane. Figure adapted from Longman et al. (2020).* ***(C)*** *Comparison of enrichment of translocon channel component SEC61A and core NMD factor UPF1 with enrichment of Syntaxin-18 complex components by comparitive mass spectrometry. Dotted line indicates threshold for enrichment (>2-fold increase in detection in GFP-NBAS sample compared to WT parental HeLa cells).* ***(D)*** *Structure of the SEC61 complex which makes up the endoplasmic reticulum (ER) translocon channel where the translating ribosome docks. View of the lateral gate (top) or from the ER lumen (bottom) of the SEC61 complex bound to the preprolactin signal peptide (cyan). The mobile regions of SEC61A are blue; the comparatively immobile regions are red. The SEC61B and SEC61G subunits are pale green and tan, respectively. (Figure adapted from Voorhees and Hegde, 2016).* ***(E)*** *STRING interaction analysis showing evidence of interaction between proteins of the Syntaxin-18 complex and SEC61A.*

The majority of the identified interactome was unchanged upon treatment with RNase including Syntaxin-18 components and SEC61A (Figure 3.11B) suggesting that NBAS interacts with both the Syntaxin-18 complex and the ER translocon independently of the presence of RNA.

The interaction between GFP-NBAS and the SEC61 translocon was validated using PLA. Figure 3.12 shows a significant increase in the PLA signal in GPF-NBAS cells compared to the parental HeLa cells when α-GFP and α-SEC61B antibodies are used, indicating that GFP-NBAS and SEC61B are <40nm apart, either directly binding or within the same complex.

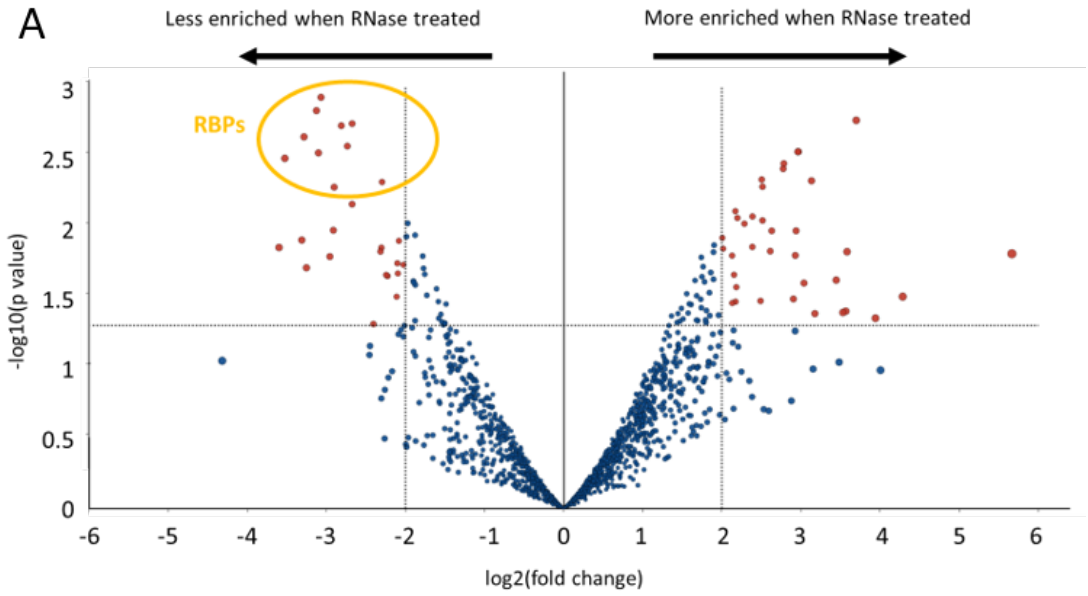
IP-MS did not result in a robust enrichment of any NMD factors. Even UPF1 was not enriched in any replicates despite being robustly shown to be directly interacting with endogenous GFP-NBAS by PLA, and also by published work using FRET/FLIM, PLA, FCCS and biochemistry techniques (Longman et al., 2020).

3.2.5 NBAS interacts with UPF1 in a domain-non-specific manner

NBAS has been shown to interact directly with RINT1 and ZW10 (Civril et al., 2010) somewhere in its C-terminal region between 1036–2371 amino acids (Tagaya et al., 2014), but interaction domains of other proteins have not been determined. Both the α-solenoid at NBAS' C-terminal and its N-terminal β-propeller structure are generally involved in protein-protein interactions (Chen et al., 2011). To understand the interplay between NBAS, UPF1 and the NMD machinery and the Syntaxin-18 complex, I wanted to map the interaction domains of NBAS and UPF1.

To investigate what NBAS domains are necessary for it to interact with UPF1, I simultaneously co-expressed T7-epitope tagged full length UPF1 and FLAG-tagged deletion mutants of NBAS and performed co-IP with α-FLAG magnetic beads to pull down tagged protein mutants and any interacting proteins. Mutant NBAS protein and co-IP full length UPF1 proteins were visualised by western blot (Figure 3.13A). I also carried out co-IP of T7-tagged full length NBAS and

A



B

Gene Symbol	GFP-NBAS v WT Rnase -		GFP-NBAS v WT Rnase +	
	logFC	q value	logFC	q value
USE1	12.93	1.2E-06	12.89	7.3E-08
BNIP1	12.89	2.8E-06	12.45	5.3E-07
SCFD2	12.40	3.3E-06	11.69	4.5E-07
NBAS	12.07	5.8E-08	11.99	1.9E-09
RINT1	11.18	1.8E-05	10.95	2.9E-06
STX18	11.13	1.9E-07	11.17	6.0E-09
SCFD1	11.05	4.0E-07	11.06	2.4E-08
C19orf25	11.03	8.7E-07	11.47	8.7E-09
MRPL27	8.52	8.0E-07	N/A	N/A
NAPA	7.98	2.8E-06	7.28	1.5E-06
HS2ST1	7.92	1.4E-06	6.85	5.9E-07
ZW10	7.82	5.3E-06	7.80	4.8E-07
CALD1	6.86	1.3E-02	6.94	1.1E-03
ACTN1	6.72	1.7E-02	8.17	5.6E-04
CEP72	5.92	8.3E-05	4.73	1.2E-05
SEC22B	5.91	1.9E-05	5.85	3.2E-06

Figure 3.11: ***GFP-NBAS interactome is RNA-independent***

(A) *Enrichment of proteins from mass spectrometry of α-GFP immunoprecipitation of GFP-NBAS cells treated with RNase compared to untreated GFP-NBAS cells. RNA binding proteins (RBPs) whose enrichment was significantly less in RNase treated cells are highlighted in yellow ring.* ***(B)*** *Comparison of GFP-NBAS RNase+ interactome with GFP-NBAS RNase- interactome, both normalised to WT parental HeLa cells. Proteins are ordered by fold enrichment in RNase- cells. Components of the Syntaxin-18 complex are in bold.*

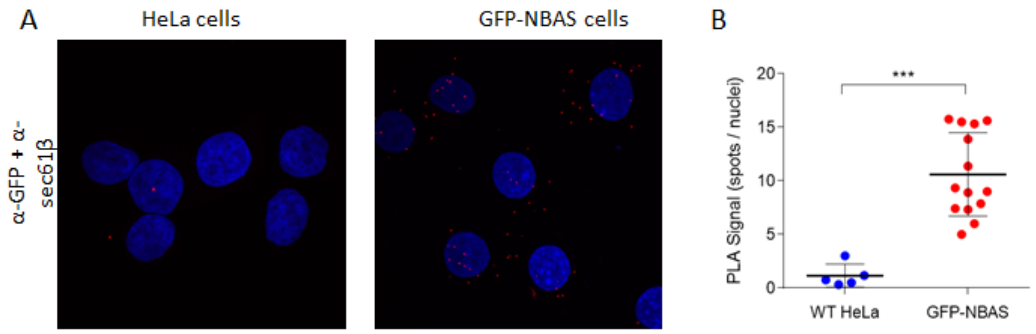


Figure 3.12: ***GFP-NBAS interaction with ER translocon is validated by PLA***

(A) *Representative PLA signal (red) when probes against GFP and SEC61B were used in GFP-NBAS cells and the parental HeLa cell line. DAPI staining shows cell nuclei (blue).* ***(B)*** *Quantification of all replicates of PLA shown in (A). The mean and standard deviation are indicated by the bold line and whiskers. Significance was determined by two-tailed Mann-Whitney test. *** indicates $p<0.001$.*

FLAG-tagged deletion mutants of UPF1 to probe which domains of UPF1 are important for the interaction (Figure 3.13B). Full length UPF1 co-IP with all NBAS deletion mutants used and full length NBAS co-IP with all UPF1 deletion mutants. Although the UPF1 deletion constructs only test a small portion of the protein, the NBAS deletion mutants span almost the entirety of its length and both identified domains. This could suggest that no single portion of NBAS is solely required for interaction with UPF1, or it forms a dimer or higher order complex. Alternatively, the small sections of NBAS that are present in all mutants constructs could be the binding domain. The C-terminal of NBAS consists of many α-solenoid repeats, which are known to be strong protein-protein interaction domains so even the small portion present in the NBAS C-terminal mutants could be sufficient.

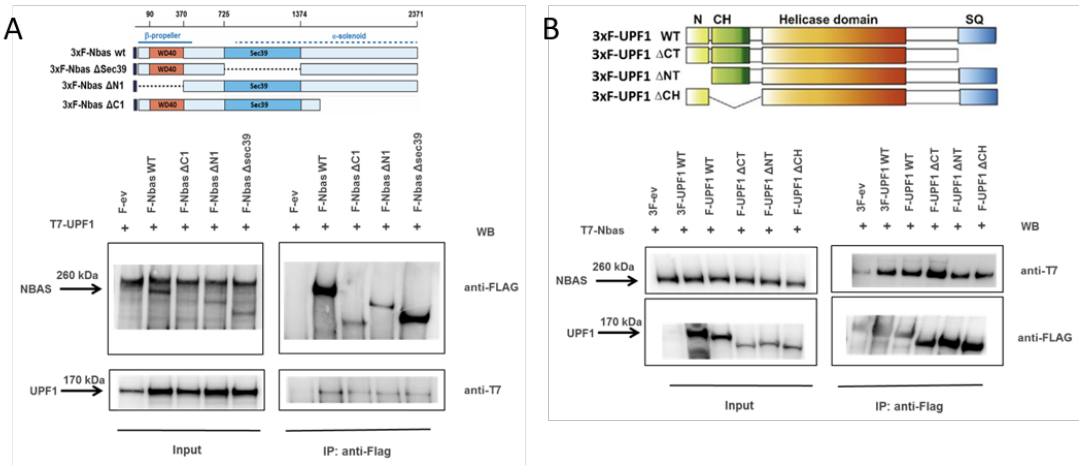


Figure 3.13: ***NBAS interacts with UPF1 in a domain-unspecific manner***

(A) *Schematic of deletion mutant constructs of NBAS and immunoprecipitation using α-FLAG. Cells were co-transfected with T7-UPF1 and FLAG-tagged NBAS either full-length WT or the mutant indicated. Western blot is probed with α-FLAG and α-T7 antibodies as indicated.* ***(B)*** *Schematic of deletion mutant constructs of UPF1 and immunoprecipitation using α-FLAG. Cells were co-transfected with T7-NBAS and FLAG-tagged UPF1 either full-length WT or the mutant indicated. Western blot is probed with α-FLAG and α-T7 antibodies as indicated.*

3.3 Discussion

3.3.1 Creation of GFP-FLAG-NBAS cell line

NBAS has been shown to be involved in Golgi-to-ER retrograde transport; the function is proposed to depend on its association in the NRZ complex which is believed to play a role in SNARE assembly at the ER (Aoki et al., 2009). The role of NBAS in vesicular transport is thought to be independent of its function in RNA regulation (Longman et al., 2020), although little is understood regarding the mechanism of its NMD activity.

Here, work aimed to create an endogenous model system in human cells to investigate the details of the NMD function of NBAS. A CRISPR/Cas9 strategy was successfully designed and executed to endogenously tag the N-terminal of NBAS with a 3xFLAG epitope and eGFP tag. Both homozygous and "heterozygous" GFP-FLAG-NBAS cell lines were generated, and homozygous cells characterised. These cells exhibit an ER staining pattern by immunofluorescence analogous to the ER marker calnexin (Figure 3.8A) and IP-MS confirms interaction of GFP-FLAG-NBAS with members of the Syntaxin-18 complex (Figure 3.10A).

The efficiency of the CRISPR targetting screen was very low efficiency (Figure 3.5D). Since the cutting efficiency of the gRNA/Cas9 was high (Figure 3.3C), this is assumed to be due to low HDR efficiency. There are several techniques that could be employed in the future to potentially increase success rate leading to multiple clones that can be used as biological replicates. For example, using cell synchronisation to increase HDR frequency because HDR is more active in S/G2 than other cell cycle phases (Lin et al., 2014). Furthermore, many "heterozygous" clones appeared to have multiple products from the PCR of the N-terminal region of the NBAS locus, strongly suggesting more than two NBAS loci in the HeLa cells used. Targeting in more physiological cells such as mouse embryonic stem cells may improve the efficiency of this technique as only two successful incorporations need to occur to produce a homozygous clone.

Whether the tagged NBAS in these cells is capable of full NMD function is to be determined. To interrogate the function of GFP-NBAS in NMD, qRT-PCR could be used to compare abundance of mRNA of previously identified NBAS NMD targets (Longman et al., 2007, 2013, Chapter 4), after siRNA knockdown of GFP-NBAS or WT NBAS in the GFP-NBAS cells compared to the parental untagged HeLa cells.

In this chapter, I produced a useful tool cell line that can be used for future study of NBAS and to further investigate ER-NMD. Tagged GFP-3xFLAG-NBAS behaved as expected in all validation experiments carried out. Further validation that could be carried out on the cell line would be IP-WB to probe interaction of NBAS with known interactors including Syntaxin-18 components and UPF1, and the novel interactors identified in the IP-MS study including SEC61B and SRPR.

3.3.2 Interaction of UPF1, NBAS and the ER translocon

The results of the IP-MS and PLA in this chapter suggest NBAS interacts directly with the SEC61 translocon where translating ribosomes dock at the ER membrane. This would bring NBAS into close proximity to the transcripts being translated. Taken together with evidence that UPF1 interacts with NBAS directly and also with SEC61B (Longman et al., 2020), this strongly suggests that ER-NMD functions via a complex involving NBAS, UPF1 and the translocon (Figure 3.14).

No robust enrichment of any NMD factors was observed, despite these interactions previously being demonstrated by IP-WB (Longman et al., 2020, Appendix C, Figure S6, S7). This may be due to the more transient nature of interactions between NBAS and NMD factors compared to components of the Syntaxin-18 complex, meaning their interactions are not conserved through the IP-MS protocol. Repeating the IP-MS after crosslinking the cells could stabilise these interac-

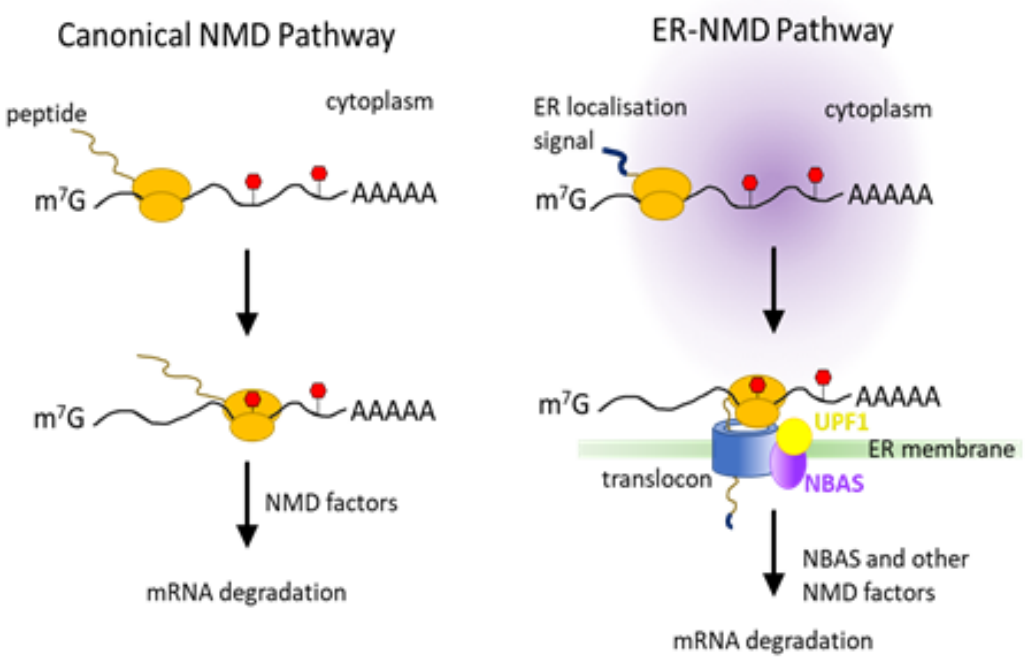


Figure 3.14: ***Summary diagram of ER-NMD mechanism***

Canonical NMD takes place in the cytoplasm where UPF1 and other factors target a transcript with a premature termination codon for degradation. When a transcript has an endoplasmic reticulum (ER) localisation signal the ribosome and transcript move to the ER and translation takes place at the translocon. Both NBAS and UPF1 interact with the translocon channel and each other. This suggests a model where NBAS, UPF1 and the ER translocon form a complex that is involved in NMD quality control of transcripts translated at the ER.

tions and aid in identifying the more transient interactions of NBAS. It could also be that although transiently expressed tagged NBAS has been shown to interact with UPF1, SMG5, SMG6 and SMG7 (Longman et al., 2020), expression of the endogenous tag disrupts the interaction of NBAS with these proteins.

IP-MS studies of UPF1 have identified NMD interactions, although not with all known NMD proteins. UPF2 and UPF3B have been found by published MS studies on FLAG-UPF1 (Flury et al., 2014). Furthermore, the EJC components CASC3 (MLN51) and eIF4A3 were identified. However, the UPF1 phosphorylating kinase SMG1, its regulators SMG8 and SMG9, as well as the downstream NMD proteins SMG5, SMG6, and SMG7 were absent from the enriched proteins. The previously reported interaction with Staufen homologue 2 (STAU2) was also identified as an mRNA independent UPF1 interactor, while the more studied homologue STAU1 was not identified. IP-MS of GFP-tagged NMD factor could clarify whether it is the GFP tag that is preventing detection of NMD complexes in this system. Alternatively, a more recent technique BioID uses proximity-dependent trans-biotinylation by the Escherichia coli biotin ligase BirA mutant R118G (BirA*) (Roux et al., 2012, 2013). This allows stringent streptavidin affinity purification of proximal proteins and provides an alternative to co-IP or IP-MS to identify protein-protein interactions. BioID using known NMD factors as HA-BirA*-tagged bait proteins revealed new putative NMD-associated proteins that presumably interact only transiently with the NMD complex (Schweingruber et al., 2016). However, even this technique did not detect all known NMD protein interactions. For example, HA-BirA*-UPF1 selectively biotinylated UPF2, but not UPF3B, and biotinylated SMG1, SMG6 and SMG7 were observed just above the detection limit. HA-BirA*-UPF2 was found to biotinylate UPF3B and HA-BirA*-SMG5 biotinylated SMG7, but SMG5 was not detected in the HA-BirA*-UPF1 or HA-BirA*-UPF2 samples. MS is known to be influenced by expression level of proteins. Even with ample input proteome, MS measurements are heavily biased toward high abundance proteins, often omitting a majority of the proteome in routine analyses (Picotti et al., 2007; Li et al., 2017a).

Brannan et al. (2016) performed quantitative IP-MS on a HaloTag fusion UPF1 and showed that UPF1 interacts with ZW10, a member of the NRZ complex which NBAS is also a subunit of. This suggests that perhaps the entire NRZ complex is involved in ER-NMD. However, knockdown of ZW10 appeared to have no impact on NMD (Longman et al., 2020). This could suggest UPF1 interacts with ZW10 as part of a separate function to NMD since UPF1 has diverse non-NMD roles. These include regulation of RNA on multiple levels such as RNA splicing, transport, translation, and mRNA turnover (Atkin et al., 1995; Lykke-Andersen et al., 2000; Mendell et al., 2002; Wilkinson, 2005; Nazarenus et al., 2005). For example, UPF1 has been identified as interacting with SRSF1 which plays a role in preventing exon skipping, ensuring the accuracy of splicing and regulating alternative splicing (Aznarez et al., 2018), and UPF1 stabilizes primary retroviral RNA so that UPF1 downregulation by siRNA decreases HIV RNA and diminishes the synthesis of virion structural proteins (Ajamian et al., 2008). The interaction of UPF1 and ZW10 could be indirect, via a larger complex that could include NBAS, or even components of the translocon and Syntaxin-18 complex.

Full length UPF1 co-IP with all NBAS deletion mutants used and full length NBAS co-IP with all UPF1 deletion mutants. Although the UPF1 deletion constructs only test a small portion of the protein, the NBAS deletion mutants span almost the entirety of its length and both identified domains. This could suggest that no single portion of NBAS is solely required for interaction with UPF1, perhaps because NBAS forms a binding platform that brings UPF1 to the proximity of translocon at ER. Alternatively, NBAS could form a dimer or higher order complex, and overexpressed truncated proteins are functionally complemented by the endogenous WT NBAS. To test this further, WT endogenous NBAS should be degraded or knocked out and the experiment repeated with expression of each of the mutants with full length UPF1. Whilst the interactions sites between UPF1 and many NMD factors has been identified, the direct interaction between UPF1 and UPF3B also still remains to be mapped (Lavysh & Neu-Yilik, 2020).

3.3.3 Future work

The putative protein interactors of NBAS identified in the mass spectrometry analysis will need to be confirmed by IP-WB analysis, and could be followed up by PLA to validate via a different approach. Furthermore, to determine which of the interactions are mediated by RNA, RNase treatment of control parental cell line followed by IP-MS is necessary to allow normalisation of the RNase treated GFP-NBAS sample to account for changes in solubility due to the lack of RNA (Trevino et al., 2007). To verify whether the interactions are direct or mediated by other proteins, in vitro binding assays could be performed and potentially deletion mutants of NBAS and identified interacting proteins could then be developed to map the protein regions/domains important for binding. Interacting proteins could also be knocked down and NMD activity measured to assess whether they are involved in the ER-NMD response.

The next question regarding involvement of NBAS in ER-NMD is its precise function in the pathway. NBAS tethers UPF1 to the ER and potentially recruits UPF1 initially (Longman et al., 2020). This suggests that NBAS has a structural scaffold role in bringing UPF1 to the translocon and into close proximity with the translating ribosome. In this case, the question arises whether NBAS only recruits and tethers UPF1 when the ribosome stalls or whether NBAS recruits UPF1 constitutively, but UPF1 will only act on NMD when specific complexes are formed and UPF1 is phosphorylated by SMG1. Perhaps the interaction between NBAS and UPF1 is stabilised by the stalling of the ribosome in a similar manner to how this appears to stabilise UPF1 binding to mRNA. Alternatively, UPF1 could be trafficked to the ER translocon with the mRNA and ribosome but is then kept at the ER by interaction with NBAS. One way to investigate this would be comparing the transcripts bound by UPF1 in WT cells compared to cells were NBAS is depleted. If transcripts translated at the ER are less bound in the absence of NBAS, it would suggest NBAS stabilises UPF1 binding to these transcripts. To do this, a crosslinking immunoprecipitation (CLIP) protocol could be used, which consists of crosslinking cells with UV to stabilise

interactions between protein and RNA, followed by immunoprecipitation of the protein of interest and RNA-sequencing of the co-immunoprecipitated RNA. This comparative CLIP experiment would also have the likely advantage of producing a result with less background signal as it would allow normalisation to wild type. Furthermore, CLIP could be combined with subcellular fractionation (Brugiolo 2017) to specifically identify transcripts bound by UPF1 in the vicinity of the ER.

It would also be interesting to test if NBAS and UPF1 bind the translocon at the same time, and whether UPF1's slower diffusion at the ER (Longman et al., 2020) is due to it binding specifically to the translocon or to NBAS or both. To test this, the interaction between UPF1 and SEC61B could be probed by IP-WB in cells depleted of NBAS. If NBAS is required for the recruiting of UPF1 to the translocon or stabilising the interaction, I would expect the interaction to be less or abrogated when NBAS was depleted.

Another linked question is whether NBAS binds the transcripts that it helps regulate by NMD. Although NBAS does not have an annotated RNA binding domain, such as an RNA-recognition motif, double-stranded RNA-binding motif or zinc-finger motif, more than a third of RNA binding proteins (RBPs) have no prior RNA binding related homology or annotation (Castello et al., 2013) and little is known about the specific functions of the identified NBAS protein domains (Section 1.3.1). Neither NBAS nor any of its paralogues feature in the RNA-Binding Protein DataBase (Berglund et al., 2008; Cook et al., 2011), and NBAS is not among the nearly 900 RBPs identified in human HeLa cells using interactome capture; UV cross-linking of proteins to RNA followed by oligo(dT) capture and mass spectrometry (Beckmann et al., 2015; Castello et al., 2012; Conrad et al., 2016). However preliminary evidence suggests NBAS may bind RNA. Messenger RNA capture assay uses UV crosslinking to stabilise protein-RNA interactions and oligo(dT) beads to isolate all RNA species with poly-A tails, thereby purifying RBPs. The presence of NBAS in mRNA-protein (mRNP) complexes was revealed by western blot with α-NBAS antibody, suggesting that NBAS does bind mRNA (Longman et al., 2020, Appendix C, Figure S6D). However, this

could also indicate that NBAS is part of a protein complex and one or members of the complex directly bind RNA, not NBAS, leading to indirect pull-down of NBAS. This could be validated by IP of NBAS followed by degradation of protein and DNA, and visualisation of RNA.

If NBAS is an RBP, then it would next be interesting to investigate whether it is the transcripts regulated by ER-NMD that NBAS binds to or if this RNA binding function is an independent role. A CLIP experiment coupled with RNA-sequencing would enable identification of the RNAs that are directly bound by NBAS. It is widely accepted that performing CLIP on some proteins is much more challenging than others and a strong enrichment of the protein of interest in the immunoprecipitation is essential. Therefore, a pertinent approach would be to perform CLIP on the GFP tagged NBAS cell line described in this chapter so an α-GFP antibody, which has already been shown to be successful (Meyer et al., 2017; Tang et al., 2020), could be used to isolate NBAS.

3.3.4 Summary

In conclusion, in this chapter I describe the endogenously tagged NBAS cell line created, which will be a useful tool for future studies. Using this cell line, I have shown that NBAS interacts with components of the ER translocon where the ribosome docks when translating at the ER. Treatment with RNase determined that protein interactions between NBAS, the Syntaxin-18 complex and the translocon were RNA-independent. The novel interaction between NBAS and the SEC61 translocon was validated by proximity specific ligation (PLA). Taken together with the evidence that NBAS interacts with UPF1 and UPF1 also interacts with the ER translocon suggests that UPF1 and NBAS form a complex with the translocon to regulate the ER-NMD.

The key findings presented in this chapter have been published in the following article, which is attached as Appendix C:

Longman D, ***Jackson-Jones KA****, Maslon MM, Murphy, LC, Young, RS, Stoddart, JJ, Hug, N, Taylor, MS, Papadopoulos, DK, & Cáceres, JF (2020) Identification of a localised nonsense-mediated decay pathway at the endoplasmic reticulum Genes & Dev. 2020. 34: 1075-1088. doi:10.1101/gad.338061.120*

Chapter 4

NBAS regulates a subset of UPF1 targets

4.1 Introduction

Since NBAS is localised to the endoplasmic reticulum (ER), it has been hypothesised that it is involved in a branch of nonsense mediated decay (NMD) specific to transcripts translated at the ER (Longman et al., 2020). This ER-NMD pathway would likely require the presence and activity of the core NMD factor UPF1. If this is the case, a subset of UPF1 targets is expected to be translated at ER and also regulated by NBAS. Therefore, I wanted to carry out an unbiased transcriptomics approach to identify mRNA targets of NBAS and compare them to UPF1 targets.

4.1.1 Differential expression as a tool to identify targets of NMD

The NMD pathway degrades the target RNA transcript, and the nascent peptide is presumed to also be degraded, presumably by the proteasome, although the mechanism is not known. Therefore, identifying differentially expressed transcripts when specific NMD factors are depleted has been used as a standard for identifying targets of those factors.

Many microarray analyses and, more recently, RNA-sequencing analyses of the transcriptome of human cells after expression of UPF1 has been knocked down have been carried out (Wang et al., 2011a; Tani et al., 2012; Hurt et al., 2013; Lykke-Andersen et al., 2014; Schmidt et al., 2015; Lou et al., 2016). In line with the variable nature of NMD, overlap of identified UPF1 targets between datasets is generally low. Of the 533 targets identified by Lou et al. (2016), only 27% had been previously identified by other studies. There have also been similar analyses carried out when other NMD factors have been knocked down (Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011). Identification of how the targets of various factors differ from those of canonical NMD have been used to identify subclasses of NMD (Section 1.2.3).

NMD activity of NBAS has only previously been investigated by microarray analysis of mRNA upon depletion of NBAS (Longman et al., 2013). This study showed a significant correlation between transcripts increased in abundance when NBAS was depleted and those increased when UPF1 was depleted, indicating co-regulation by the two factors in *Caenorhabditis elegans*, *Danio rerio* (zebrafish) and HeLa cells (Longman et al., 2013).

Whilst microarray analysis is considered to be genome-wide, it is limited to those transcripts for which probes are included in the microarray and ability to identify specific transcripts is limited (Wang et al., 2009; Wilhelm and Landry, 2009). In contrast, RNA-seq is unbiased and can detect novel transcripts, gene fusions, single nucleotide variants, indels, and other previously unknown changes that arrays cannot detect. Furthermore, RNA-seq can quantify expression across a wider dynamic range since it produces discrete, digital sequencing read counts (Wang et al., 2009; Wilhelm and Landry, 2009; Zhao et al., 2014; Rao et al., 2018) which allows detection of rare and low-abundance transcripts. Compared to microarrays, RNA-seq technology can detect a higher percentage of differentially expressed genes, especially genes with low expression (Wang et al., 2014; Liu et al., 2015), resulting in higher specificity and sensitivity. Several studies have investigated the comparability between the microarray platform and RNA-seq (Fu et al., 2009; Zhao et al., 2014; van der Kloet et al., 2020). Correlation of the two techniques have been estimated between $r = 0.62$ and $r = 0.75$ (Fu et al., 2009), with RNA-seq found to provide better estimates of absolute expression levels, as background hybridisation and probe saturation in microarrays resulted in limited sensitivity in both low and high expression levels (Zhao et al., 2014).

Performing RNA-seq on cells depleted of NBAS will allow more complete and unbiased identification of transcripts regulated by NBAS. This will enable greater understanding of the role NBAS has in NMD and the types of transcripts it regulates. Comparing these results to those of the core NMD factor UPF1 will indicate whether a distinct NMD pathway is occurring at the ER.

4.1.2 Aims

In this chapter I aim to understand the similarities and differences of NBAS-dependent gene expression changes and UPF1-dependent changes by implementing RNA-seq upon depletion of either factor and then characterising and comparing the results.

To do this, I will deplete HeLa cells of either UPF1 or NBAS and subject the total RNA to RNA-seq. Differential expression analysis will reveal transcripts increased in expression when NBAS is depleted. To determine whether UPF1 and NBAS targets are translated at the ER, I will compare the target lists to three external, experimental datasets which identified ER-localised transcripts. Analysis of RNA-seq data from fractionated cells will allow comparison of membrane and cytoplasmic effects of NBAS compared to UPF1 and UPF2.

In summary, the main aims of this chapter are:

a) To identify putative NBAS NMD targets by determining transcripts increased in expression upon depletion of NBAS

b) To compare NBAS targets with those of core NMD factor UPF1

c) To assess the effects of UPF1 on transcripts translated at the ER

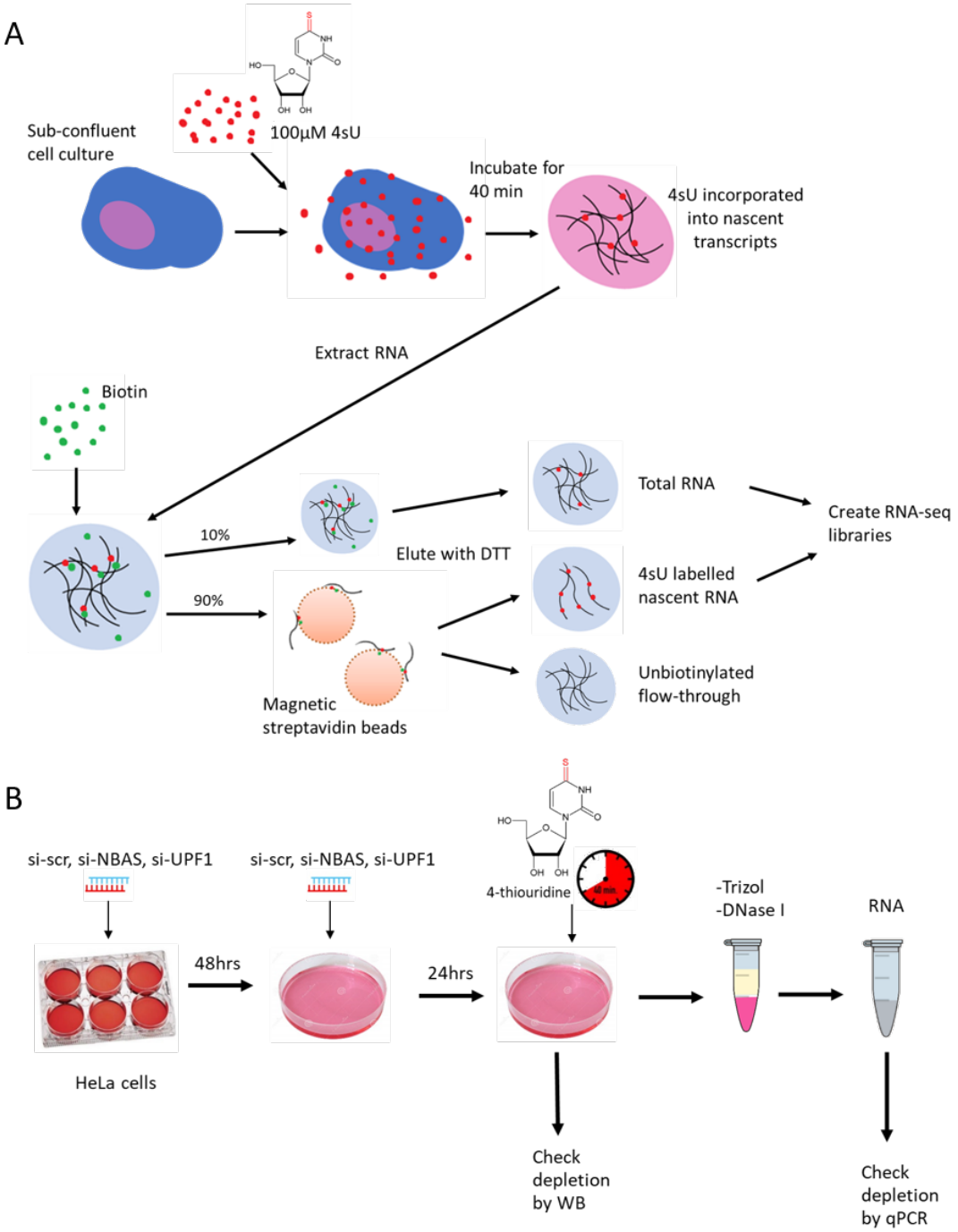
4.2 Results

4.2.1 Experimental design

As well as differential expression, I wanted a single design that would allow the discrimination of transcriptional changes from transcript stability changes, as direct NMD effects are specifically expected to be exerted through altered transcript stability. Measurement of stability required both total RNA and nascent RNA from the same cells (described fully in Chapter 5) therefore as well as extracting total RNA and sequencing it to measure differential expression, cells were treated with the uridine analogue 4-thiouridine (4sU) and nascent RNA fractionated (Figure 4.1A). RNA-seq libraries were prepared from both total RNA and 4sU labelled nascent RNA from the same cells. For comparison to previous gene expression-based studies of NMD, only the total RNA was used in all analyses in this chapter.

HeLa cells were depleted of either UPF1 or NBAS using siRNA pools. Control cells were treated with a scrambled siRNA pool (si-Scr) or mock depleted as negative controls (Figure 4.1B). Five biological replicates were treated, and libraries prepared resulting in a total of 40 sequencing libraries and corresponding datasets (Figure 4.1C).

Figure 4.2A shows the level of depletion of NBAS and UPF1 mRNA in each sample that was prepped for RNA-sequencing, normalised to mock treated cells. UPF1 mRNA was depleted to 29.2% on average ($p < 0.0001$) when cells were treated with si-UPF1, whilst NBAS was depleted to 23.4% on average ($p = 0.0016$) of untreated levels when si-NBAS was used. Although not meeting statistical significance, the level of NBAS mRNA was slightly increased upon UPF1 is depletion, suggesting NBAS is part of the NMD factor positive feedback loop described by others (Longman et al., 2013; Huang et al., 2011; Yepiskoposyan et al., 2011). Western blot of depleted samples shows protein levels of both factors were undetectable (Figure 4.2B). Since ∼25% of WT levels of mRNA were



C

	mock	si-Scr	si-UPF1	si-NBAS
Total	n=5	n=5	n=5	n=5
Nascent	n=5	n=5	n=5	n=5

Figure 4.1: ***RNA-sequencing experimental design***

Figure 4.1: ***RNA-sequencing experimental design***

(A) *Labelling with the uridine analogue 4-thiouridine (4sU) allows isolation of nascent RNA from total. 4sU is incorporating into nascent transcripts during the labelling period and then cells are lysed with Trizol and RNA purified. Biotin is added which strongly binds to the sulphur of the 4sU. 10% of the sample is separated as "total RNA". Incubation with streptavidin beads allow purification of the biotin tagged "nascent RNA" from flow-through. Libraries can then be prepared from matched total and nascent RNA originating from the same cells.* ***(B)*** *Time course of the experiment proceeded with reverse transfection of HeLa cells with appropriate siRNAs. 48hrs later, cells were expanded and depleted for a second time with the same siRNA pools. 24hrs later, cells were labelled with 4sU for 40min before being harvested and RNA extracted. Depletion of all samples was carried out by qRT-PCR before library preparation.* ***(C)*** *Each condition was replicated 5 times and then total and nascent RNA extracted from each sample, resulting in 40 RNA-sequencing libraries.*

detected by qPCR, this suggests that the that the protein level of knockdown may have been stronger than measured by qPCR or could be an indicator that qPCR detection limits are lower than western blot probing. However, it is difficult to directly compare qPCR results with western blot, as western blots are not linearly quantitative. Analysis using qRT-PCR was used to assess the effect of the depletions on previously identified UPF1 and NBAS NMD targets (Longman et al., 2013, Figure 4.2C). Although the level of depletion of both UPF1 and NBAS was similar, the increase in abundance of previously identified target mRNAs was generally lower than seen previously (Longman et al., 2013).

RNA quantity and quality were assessed by Bioanalyzer. Sample RNA integrity numbers (RINs) ranged from 8.6 to 10, indicating that the RNA quality was sufficient for sequencing. RNA-sequencing libraries were produced from total RNA and nascent RNA from the same samples. 1μg of each sample was reserved as "total RNA", then 4sU-labelled nascent RNA was purified using biotinylation

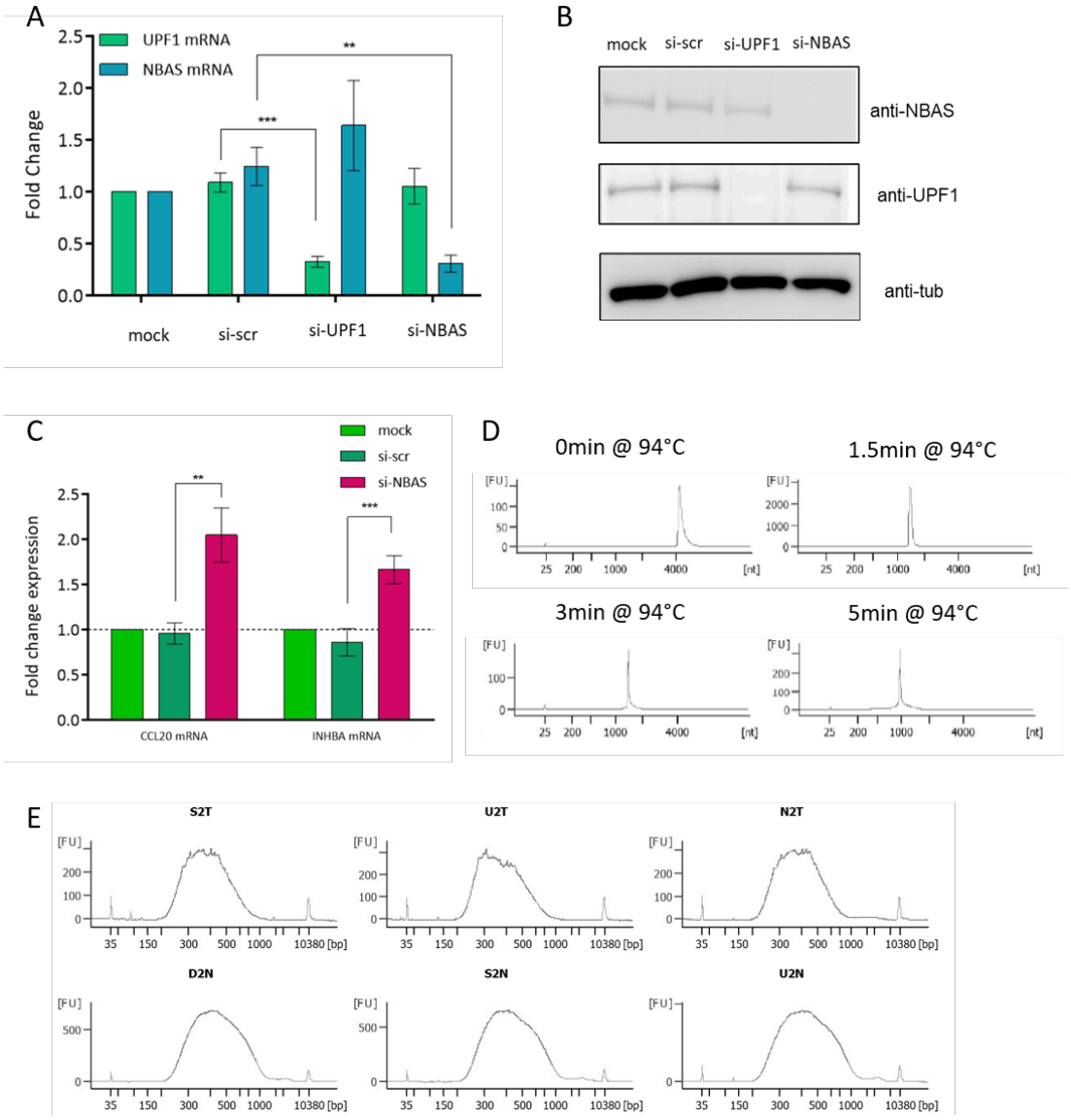


Figure 4.2: ***siRNA targeting of NBAS results in undetectable level of protein and upregulation of NMD targets***

Figure 4.2: ***siRNA targeting of NBAS results in undetectable level of protein and upregulation of NMD targets***

(A) *qRT-PCR analysis shows relative expression of NBAS and UPF1 mRNA in each sample that was prepped for RNA-sequencing, normalised to mock treated cells (each condition n=5). UPF1 mRNA was depleted to 29.2% on average ($p < 0.0001$) when cells were treated with si-UPF1, whilst NBAS was depleted to 23.4% on average ($p = 0.0016$) when si-NBAS was used.* ***(B)*** *Representative western blot of depleted samples shows protein levels of both factors were undetectable.* ***(C)*** *qRT-PCR analysis shows relative expression of mRNA of previously identified NBAS NMD targets; CCL20 was increased 2.05-fold ($p = 0.0068$) and INHBA was increased 1.66-fold ($p = 0.0008$).* ***(D)*** *Bioanalyzer plots show decreasing size of fragments upon sonication for increasing periods of time.* ***(E)*** *Representative bioanalyzer plots of three total and three nascent RNA libraries show the single peak spanning 300-1000nt.*

and streptavidin beads. Ribosomal RNA (rRNA) depletion was not performed on any samples due to the very low concentration of nascent RNA samples (1.96-7.48ng/μl; total samples concentration 69.4-180ng/μl) and the desire to process nascent and total RNA samples in the same way.

Integrity analysis of nascent RNA using Bioanalyzer implied very little RNA present in nascent samples, and correspondingly samples were assigned very low RIN values. Since bioanalyzer analysis relies on levels of rRNA present, this was not surprising since very little rRNA is expected to be transcribed in a 40min window. However, this raised the issue of choosing the correct fragmentation time of RNA prior to library preparation, since this is usually calculated from the RIN obtained with the Bioanalyzer. Therefore, to estimate optimal RNA fractionation time, excess of two more concentrated nascent samples (mock and si-scramble) were pooled to increase the concentration of RNA and split into aliquots. These replicates were sonicated for 1.5, 3 or 5mins and the resulting fragment lengths were analysed by Bioanalyzer (Figure 4.2D). Fragmentation for 1.5, 3 and 5min all decreased the average fragment length from >4000nt to ∼1000nt. Even though

after 5min the fragments were larger than the optimal fragment size of ∼200nt, 5min was chosen as optimal because it gave a clean peak at the lowest size of the times tested and a longer fragmentation time could be too long for the other samples with lower concentrations.

Libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina, following manufacturers protocol with NEBNext Multiplex Oligos with unique barcodes for each library. After completing library preparation, libraries were again assessed by Bioanalyzer (Figure 4.2E). All libraries produced a single broad peak at ∼300nt-700nt in length. All 40 libraries were pooled and submitted to Edinburgh Genomics for paired-end sequencing.

4.2.2 Processing of RNA-sequencing data

Reads were de-barcoded by Edinburgh Genomics and data received as fastq format files for each sample, split by lane of flow cell and by each end of the reads. Average number of reads per library was ∼100 million, with a similar level of variance across all samples (Figure 4.3A). Quality control analysis with fastqc showed no tile biases, high per sequence quality scores (Figure 4.3B) and high per base sequence quality (Figure 4.3C).

Each lane of each replicate was mapped using the quantification tool kallisto (Bray et al., 2016) to a genome index consisting of protein coding genes and long non-coding RNAs from `GRCh38/hg38` human genome (Figure 4.3D). Kallisto pseudo-aligns reads on a transcript level without alignment of individual bases, and produces counts data, minimising downstream analysis steps. Mapping efficiency was consistent across all samples with an average of 31.7% of reads being successfully pseudo-aligned and counted (Table 4.1). This was expected as the kallisto index excluded all non-coding regions except long non-coding RNAs and many reads would likely have originated from those regions, especially rRNA and tRNA since these are highly abundant in the cell.

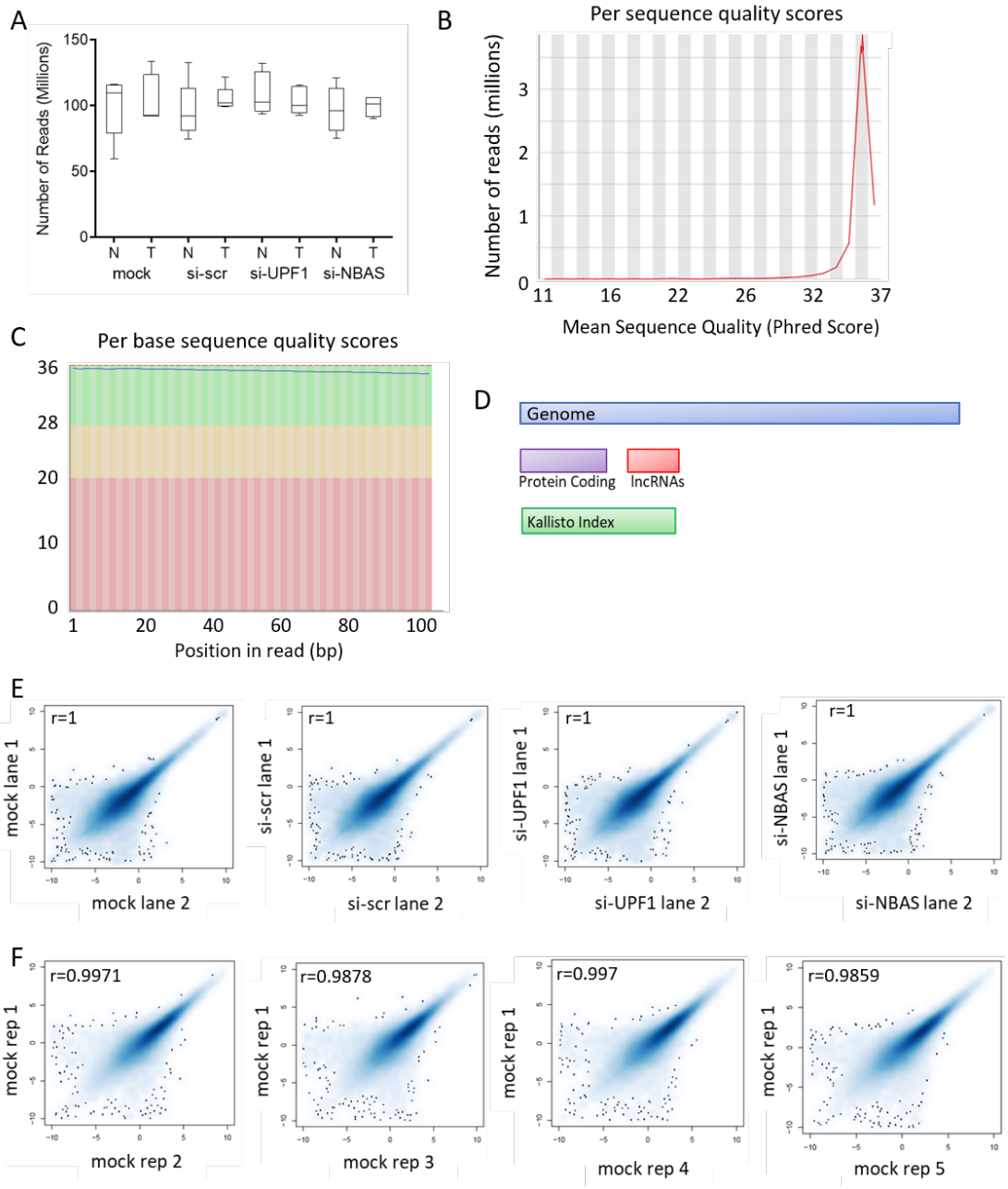


Figure 4.3: ***RNA-sequencing experimental design***

Figure 4.3: ***RNA-sequencing experimental design***

(A) *Number of reads per library in millions ($n = 5$) for each condition and nascent (N) or total (T) RNA.* ***(B)*** *Representative per sequence quality score plot output from fastqc.* ***(C)*** *Representative per base sequence quality score plot output from fastqc.* ***(D)*** *Diagram illustrating building of a kallisto index. Rather than use the full genome, specific regions of interest, in this case protein coding regions and long non-coding RNA (lncRNA) regions, are assigned k-mers to produce the index for pseudoalignment.* ***(E)*** *Representative plots of log_{10} transformed counts of technical replicates.* ***(F)*** *Representative plots of log_{10} transformed counts of biological replicates.*

Technical replication across lanes was uniformly high ($r \geq 0.9$) verifying all outputs were of consistent quality (Figure 4.3E), therefore resulting abundance files from each lane were merged to produce one abundance file per sample ($n = 40$). Plotting log-transformed counts of each replicate showed high correlation between replicates across all conditions (Figure 4.3F).

To visualise coverage of reads across the genome, each replicate was mapped to the `GRCh38/hg38` human genome assembly using STAR and visualised using the integrated genome viewer. Evidence of successful NMD abrogation was identified by way of increased inclusion of "poison" exons in a number of transcripts including SRSF3 and TRA2A. These alterative unproductive isoforms are common in splicing factors and have been found in all conserved members of the SR family of splice regulators (Lareau et al., 2007; Ni et al., 2007). The exon, which is normally skipped, contains a PTC, so when it is included via alternative splicing it targets the transcript for degradation by NMD. Poison exons are therefore a mechanism of post transcriptional gene regulation by way of the NMD pathway (Kurosaki et al., 2019b). In samples where UPF1 was depleted, an increased number of reads aligned to exon 3 and exon 4 of TRA2A and SRSF3 respectively, indicating a reduction in the decay of transcripts containing these exons (Figure 4.4).

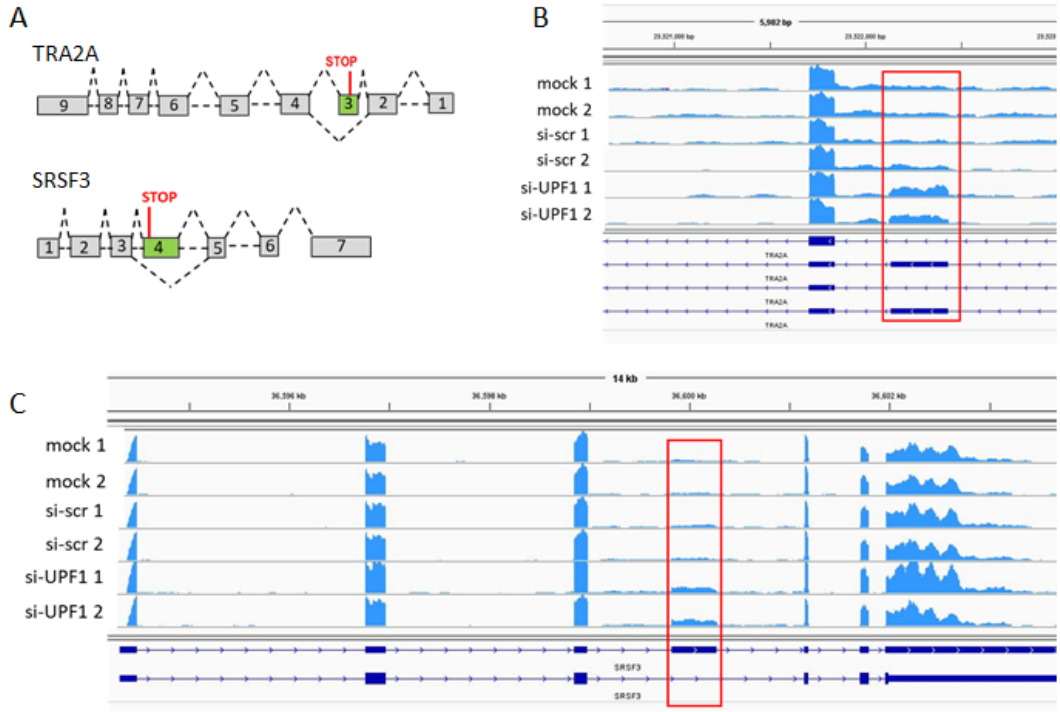


Figure 4.4: ***Poison Exon Alternative Splicing Confirms NMD abrogation***

(A) *Exon diagrams of splicing factors TRA2A and SRSF3 showing alternative splicing that leads to inclusion of poison exon containing premature termination codon. Coverage over the relevant exons of* ***(B)*** *TRA2A and* ***(C)*** *SRSF3 are shown for two replicates of mock treated, si-Scr treated or si-UPF1 treated displaying the increased coverage over the poison exon in UPF1 depleted samples.*

	Total analysed read (millions)	Aligned reads (millions)	Pseudoaligning efficiency (%)
mock total	99.6	29.6	29.7%
mock nascent	102.7	36.1	35.2%
si-scr total	96.0	27.8	28.9%
si-scr nascent	104.9	36.9	35.2%
si-UPF1 total	109.1	30.1	27.6%
si-UPF1 nascent	103.6	36.3	35.1%
si-NBAS total	96.8	26.0	26.8%
si-NBAS nascent	99.1	34.7	35.0%

Table 4.1: ***Total and nascent RNA-sequencing mapping statistics*** *RNA-sequencing libraries were pseudoaligned to a custom kallisto index of* `GRCh38/hg38` *human genome coding regions and long non-coding RNA regions. Table shows average ($n = 5$) number of reads analysed and aligned in millions and as a percentage mapping efficiency for each library condition.*

4.2.3 Hundreds of genes are differentially expressed upon depletion of UPF1 or NBAS

DESeq2 was used to identify differentially expressed genes in si-UPF1 cells or si-NBAS cells compared to si-scramble treated cells. Both NBAS and UPF1 were the most decreased mRNA in the relevant samples (fold change -1.85, $p < 2.3\times10^{-16}$ and fold change -3.28, $p < 1.14\times10^{-148}$ respectively; Figure 4.5A, B). Whilst confirming the successful depletion of each of these factors, this is surprising considering NBAS depletion was actually slightly stronger than UPF1 depletion as measured by qPCR. Furthermore, neither of the targets identified as increased in expression by qPCR, CCL20 and INHBA, were identified as significantly changed in expression upon depletion of UPF1 or NBAS.

Depletion of UPF1 affected the expression of 4756 genes, 2411 of which were increased, deemed UPF1 targets. A comparison of the UPF1 targets identified with previously published UPF1 targets (Tani et al., 2012; Lykke-Andersen et al., 2014; Schmidt et al., 2015), revealed an extensive overlap as well as led to the

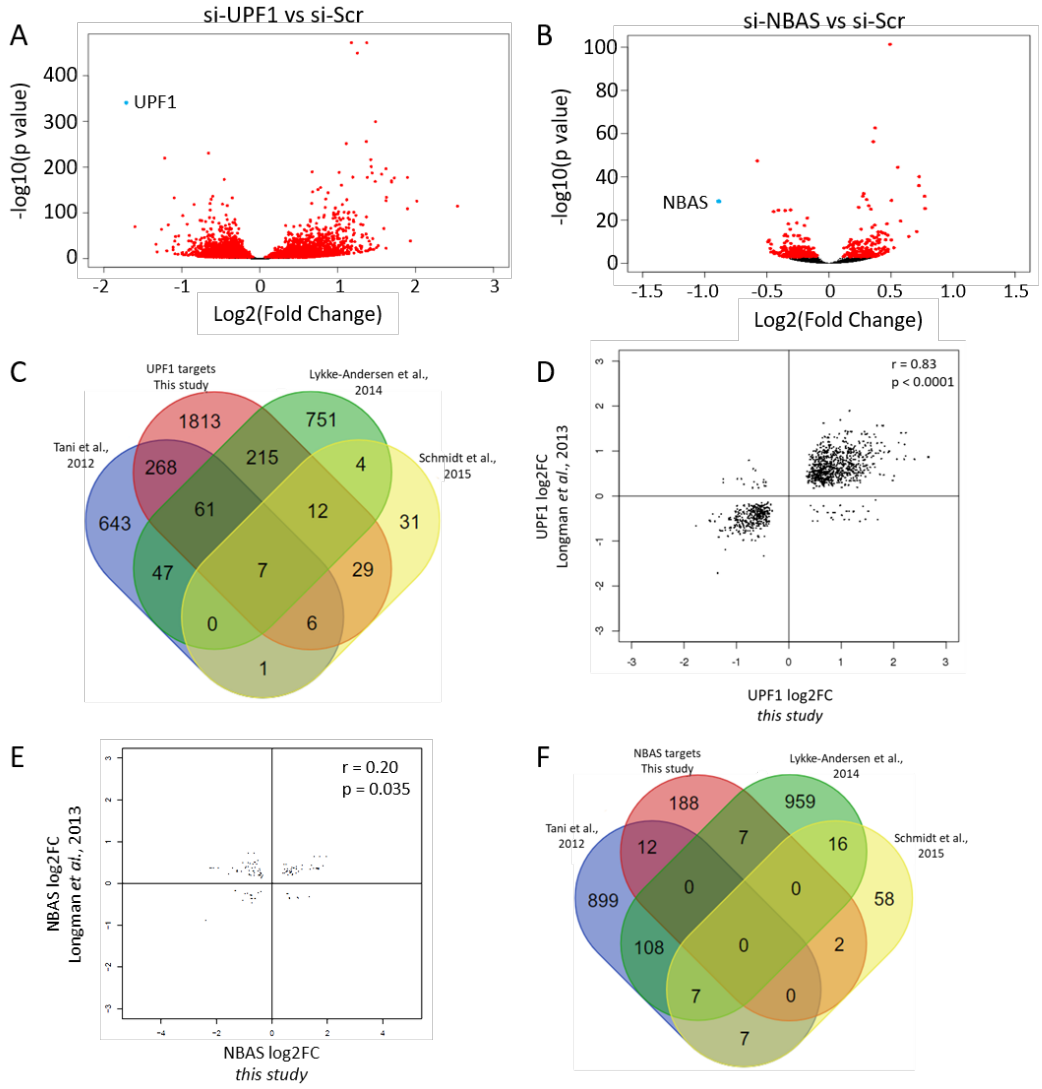


Figure 4.5: ***Hundreds of genes are differentially expressed upon depletion of UPF1 or NBAS***

Figure 4.5: ***Hundreds of genes are differentially expressed upon depletion of UPF1 or NBAS***

Differential expression analysis using DESeq2 on kallisto abundance estimates shows genes changed in expression when ***(A)*** *UPF1 or* ***(B)*** *NBAS was depleted by siRNA in HeLa cells. Genes significantly changed in expression marked in red ($p < 0.05$). The siRNA targeted gene indicated in blue and labelled.* ***(C)*** *A comparison of the UPF1 targets identified in this analysis with previously published UPF1 targets (Tani et al., 2012; Lykke-Andersen et al., 2014; Schmidt et al., 2015), revealed an extensive overlap as well as identifying many novel NMD targets. Only significant genes are shown ($p < 0.05$), full details in Sections 2.2.5 and 2.2.15.* ***(D)*** *Comparison of differentially expressed UPF1 targets identified in this dataset with those identified in the previous similar study (Longman et al., 2013). Analysis with Pearson's correlation test produces $r = 0.83, p < 0.0001$.* ***(E)*** *Comparison of differentially expressed NBAS targets identified in this dataset with those identified in the previous similar study (Longman et al., 2013). Analysis with Pearson's correlation test produces $r = 0.20, p = 0.035$.* ***(F)*** *A comparison of the NBAS targets identified in this analysis with previously published UPF1 targets.*

identification of many novel NMD targets (Figure 4.5C). Indeed, the overlap between this study and others is stronger than between some previous studies. Comparison of differentially expressed genes identified in this dataset with those identified in the previous similar study (Longman et al., 2013) shows a high correlation ($r = 0.83, p < 0.0001$, Pearson's correlation coefficient; Figure 4.5D). This indicates that despite the difference in sequencing technology (RNA-sequencing compared to microarray), the effect of UPF1 depletion can be reasonably reproducible by using the same siRNA pool and cell line.

Depletion of NBAS altered the mRNA levels of 421 genes. Of these, 209 genes were increased in abundance, suggesting they are targets of NBAS that are normally broken down. This is the first RNA-sequencing experiment assessing the effects of NBAS depletion. Comparison with the only other genome wide NBAS profiling (Longman et al., 2013) shows a weak correlation that is not statistically significant ($r = 0.20, p = 0.035$, Pearson's correlation coefficient; Figure 4.5E). This is in agreement with the variable results seen by qPCR upon depletion of NBAS, demonstrating that the effects of NBAS depletion on gene expression are more inconsistent than that of UPF1 depletion. Correspondingly, a comparison of the NBAS targets identified with previously published UPF1 targets (Tani et al., 2012; Lykke-Andersen et al., 2014; Schmidt et al., 2015), revealed very limited overlap (Figure 4.5F).

4.2.4 NBAS and UPF1 co-regulate mRNA expression specifically at the ER

A robust co-regulation of mRNAs was observed when UPF1 or NBAS was depleted (Pearson's correlation $r = 0.67, p < 0.001$) (Figure 4.6A), indicating that UPF1 and NBAS function in a common pathway. Of all NBAS targets, 95 are common targets of NBAS and UPF1 (45%), implying these are specifically targeted by the ER-NMD pathway.

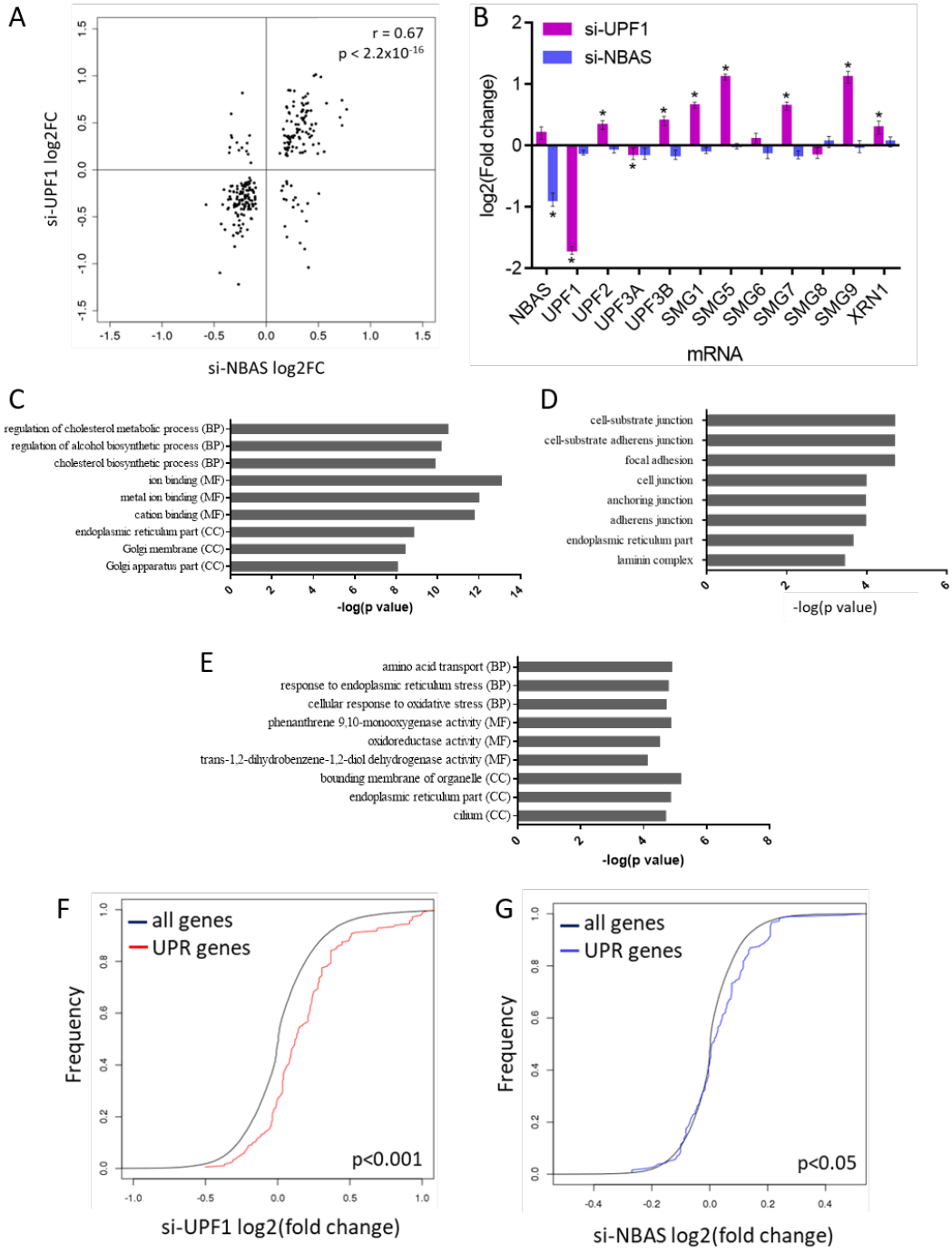


Figure 4.6: ***NBAS and UPF1 co-regulate mRNA expression and affect ER stress***

Figure 4.6: ***NBAS and UPF1 co-regulate mRNA expression and affect ER stress***

(A) *Comparing fold change of genes when UPF1 or NBAS were depleted reveals a correlation (Pearson's correlation* $r = 0.67, p < 0.001$*).* ***(B)*** *Differential expression of NMD factors upon depletion of NBAS or UPF1. Change in abundance shown as log2(fold change) and standard deviation from DESeq2. Stars indicate significant changes (*$p < 0.05$*).* ***(C)*** *Gene Ontology (GO) analysis of all NBAS targets. Top three terms are shown of biological process (BP), molecular function (MF) and cellular component (CC) GO terms.* ***(D)*** *Cellular component GO analysis of mRNAs regulated by both NBAS and UPF1.* ***(E)*** *GO analysis of all UPF1 targets as for NBAS targets.* ***(F)*** *Cumulative fold change of genes associated with the GO term ER unfolded protein response (UPR) (GO:0006986) when UPF1 depleted or* ***(G)*** *when NBAS depleted compared to the global fold change when each factor was depleted. Significant increase in fold change of these genes was seen upon depletion of UPF1 or NBAS (*$p < 0.001$ *and* $p < 0.05$ *respectively; Wilcoxon rank sum test).*

It has been shown that transcripts encoding NMD factors are themselves regulated by the NMD pathway, thus forming a feed-back auto-regulatory loop (Huang et al., 2011; Yepiskoposyan et al., 2011; Longman et al., 2013, Section 1.2.2). Therefore, I was interested to see how the RNA of other NMD factors was changed upon depletion of UPF1 and NBAS. Differential expression in either condition compared to si-Scr treated samples showed that several NMD factors were increased in expression upon depletion of UPF1 (Figure 4.6B, stars indicate significant ($p < 0.05$). The core NMD factors UPF2, UPF3B, SMG1, SMG5, SMG7 and SMG9 were all significantly increased in expression as well as the exonuclease XRN1. The NMD repressor UPF3A was significantly decreased in expression. NBAS mRNA was slightly increased in expression upon depletion of UPF1 although it did not reach significance. A significant change in expression was not detected for any of the NMD factors upon depletion of NBAS, suggesting it does not trigger the NMD autoregulatory feedback loop.

Gene Ontology (GO) analysis of NBAS targets identified ion binding and regulation of metabolic processes as common functions of proteins encoded by transcripts (Figure 4.6C). Significant cellular component terms include endoplasmic reticulum part (GO:0005783), Golgi membrane (GO:0000139) and Golgi apparatus part (GO:0005794), demonstrating NBAS target transcripts encode for secretome proteins. To identify the localisation of common targets, GO analysis of the mRNAs regulated by both NBAS and UPF1 was carried out (Figure 4.6D), resulting in all significant GO terms in the cellular compartment category relating to either secreted, extracellular proteins or membrane-associated proteins, revealing a strong enrichment for mRNAs encoding secretome proteins in the common targets, which are preferentially translated at the ER translocon (Ast et al., 2013, Section 1.1.4). GO analysis of UPF1 targets revealed enrichment of genes associated with the GO terms response to endoplasmic reticulum stress (GO:0034976), localised to endoplasmic reticulum part (GO:0005783) and bounding membrane of organelle (GO:0098588), suggesting they are secretome proteins, likely to be translated at the ER (Figure 4.6E). Whilst some secretome related terms are

significant when GO analysis is carried out on all UPF1 targets, they account for only a subset of the cellular component terms, indicating a subset of UPF1 targets are translated at the ER. Taken together this suggests that transcripts at the ER that are regulated by UPF1 are also affected by NBAS.

There is a growing body of evidence that ER stress is regulated by NMD (Section 1.4.3; Mendell et al., 2004; Gardner, 2008; Sakaki et al., 2012; Karam et al., 2015; Usuki et al., 2019), so I was particularly intrigued to know if NBAS also had links to ER stress. To investigate effect on these pathways, genes associated with ER-stress GO terms were analysed. There was a significant increase in fold change in genes associated with the ER unfolded protein response (UPR) (GO:0006986) when either NBAS or UPF1 was depleted, compared to the global fold change when each factor was depleted ($p < 0.05$ and $p < 0.001$ respectively; Wilcoxon rank sum test, Figure 4.6F, G), showing that ER stress response is increased when either UPF1 or NBAS is depleted. This suggests that ER homeostasis is perturbed when ER-NMD is not functional, supporting the hypothesis that an ER-specific NMD pathway is essential for cellular homeostasis.

To further characterize the target mRNAs, the presence of gene annotations was assessed across the three groups of UPF1 targets, NBAS targets and common targets. Surprisingly, only 20% of UPF1 targets had an annotated NMD transcript, whilst this was even lower amongst NBAS targets (10%; Figure 4.7A). Annotation of signal sequence and transmembrane helix, two common markers of secretome proteins translated at the ER, were both higher in NBAS targets than UPF1 targets, although still very low (Figure 4.7B, C).

To gain a more physiological characterisation of the targets, they were intersected with three experimental datasets identifying transcripts localised and/or translated at the ER: APEX-seq, ER fractionation sequencing and ER proximity specific ribosome profiling (Section 2.2.15). APEX-seq, a method for RNA sequencing based on direct proximity labelling of RNA led to the identification of 1077 mRNAs localised at the ER (Fazal et al., 2019). Cell fractionation followed by ribosome foot-printing coupled with deep sequencing identified 486 mRNAs (Reid and Nicchitta, 2012). Finally, proximity-specific ribosome profiling, based on ER

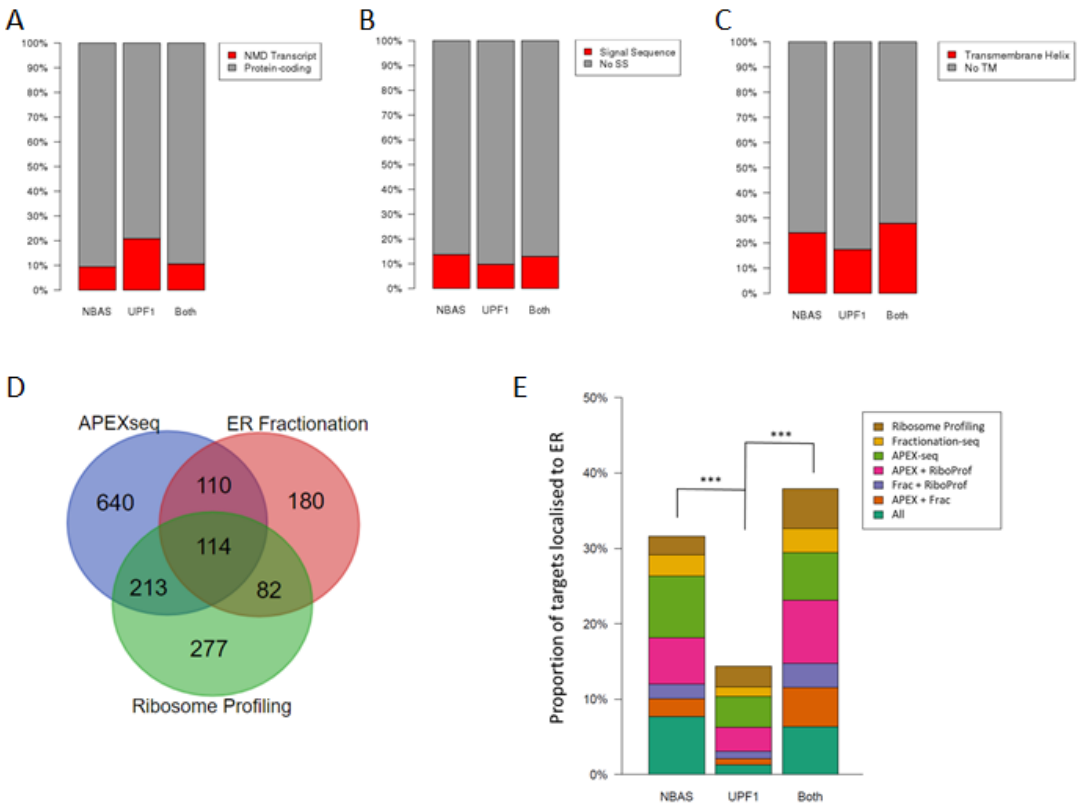


Figure 4.7: ***Targets of NBAS and UPF1 are translated at the ER***

Enrichment of gene annotations was assessed across the three groups of UPF1 targets, NBAS targets and common targets. Red area represents percentage of these targets that are annotated as ***(A)*** *having an NMD transcript,* ***(B)*** *containing a signal sequence and* ***(C)*** *containing a transmembrane helix.* ***(D)*** *Overlap of the three literature datasets identifying transcripts localised and/or translated at the ER: APEX-seq, ER fractionation sequencing and ER proximity specific ribosome profiling (Section 2.2.15).* ***(E)*** *Enrichment of each of the three ER datasets in the three groups of UPF1 targets, NBAS targets and common targets. Both NBAS targets and common targets were significantly more enriched than UPF1 targets* ($2.4 - fold, p < 0.001$ *and* $2.6 - fold, p < 0.001$ *respectively, Exact Binomial test).*

membrane proximity labelling of ribosomes identified 686 mRNAs translated at the ER (Jan et al., 2014). Surprisingly, overlap between the transcripts identified as ER localised from the three datasets was only 7% even though all were carried out in the same cell line; HEK293T (114 transcripts identified in all three studies of 1616 total RNAs identifed; Figure 4.7D), however this low rate of intersection could be partly due to inclusion thresholds used by each technique. Therefore, mRNAs deemed "significant" by one group could be almost but not quite significant for another. By overlapping with all three datasets, NBAS targets were found to be strongly enriched for experimentally validated ER genes (31.6%), as compared to 14.4% of UPF1 targets. Furthermore, 18.2% of NBAS targets were found in two or more of the datasets, as compared to only 6.3% of UPF1 targets. Considering NBAS and UPF1 common targets, 37.9% of targets were found in at least one dataset, whilst 23.2% were found in two or more datasets (Figure 4.7E). NBAS-regulated targets were strongly enriched for experimentally identified ER-localised genes ($oddsratio = 6.07, p < 0.001$, Fisher's Exact test), whereas UPF1 targets were only modestly enriched ($oddsratio = 2.44, p < 0.001$, Fisher's Exact test). Compared to UPF1 targets, those targets which were also regulated by NBAS were 2.6-fold more likely than expected by chance to be found at the ER ($p < 0.001$, Exact Binomial test), confirming that NBAS and UPF1 together regulate NMD targets specifically at the ER.

4.2.5 UPF1-independent NBAS increases in expression and NBAS decreases in expression

It is accepted that UPF1 is the core NMD factor and therefore all NMD targets are regulated by UPF1 (Zünd and Mühlemann, 2013; Singh et al., 2019). Therefore only those NBAS targets that are also UPF1 targets are presumed to be *bona fide* ER-NMD targets. However, since NBAS is not fully characterised I was interested to profile the NBAS-dependent, UPF1-independent changes in transcript expression.

Gene ontology analysis of the 114 genes increased in expression only when NBAS was depleted revealed membrane-associated and extracellular matrix-associated cellular component, molecular function and biological process terms (Figure 4.8A). Overlapping these genes with the three previously described ER datasets revealed 33.04% are also identified in these datasets (Figure 4.8B), a similar overlap to common targets (37.9%) that were increased in expression when NBAS and UPF1 were depleted. NBAS does have a known role in processing of secretome proteins, but it was surprising to find that NBAS appears to also affect expression of mRNAs that specifically encode secretome proteins in an NMD-independent manner. 212 genes were significantly downregulated upon depletion of NBAS. Overlapping these genes with the three ER datasets revealed just 9.91% were identified in these datasets (Figure 4.8B). This intersection is lower than that of UPF1 targets (14.4%), suggesting that the majority of these transcripts decreased in abundance are not ER associated and therefore are likely to be indirect effects of NBAS depletion. GO analysis of these downregulated genes resulted in disperse cellular component and molecular function terms (Figure 4.8C). However, in the biological processes category metabolic processes and RNA localisation to cajal bodies were identified. Cajal bodies are regions within the nucleus that are enriched in proteins and RNAs involved in mRNA processing (Hebert and Poole, 2017). This suggests a downregulation of RNA processing upon depletion of NBAS which is a standard cellular stress response.

4.2.6 Subcellular changes to mRNA expression

Evidence from GO analysis and comparison with ER datasets suggested that the majority of NBAS regulated mRNAs and a proportion of UPF1 regulated mRNAs are translated at the ER. However, GO analysis relies on generalisations and often predicted rather than experimentally determined function and the three ER datasets used were all performed in a different cell line to my RNA-sequencing experiment. Carrying out RNA-sequencing on cells that have been experimentally

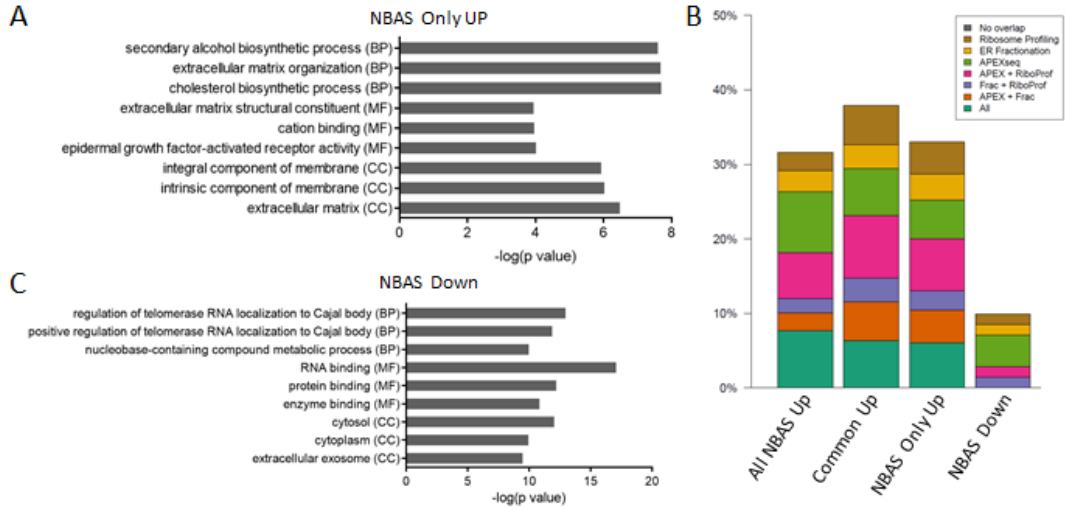
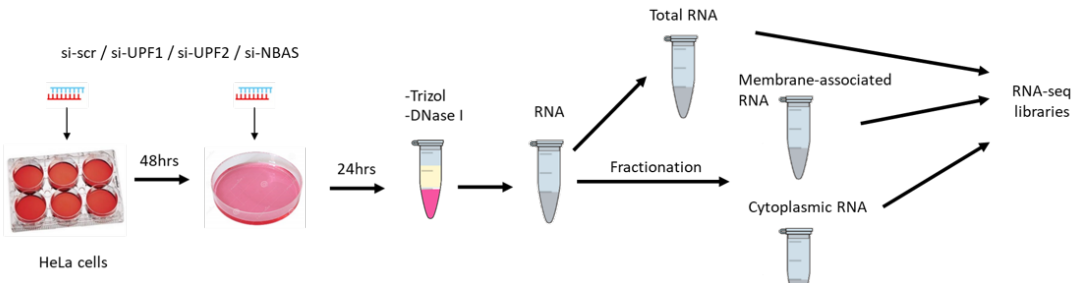


Figure 4.8: ***UPF1-independent NBAS effects on mRNA expression*** ***(A)*** *Gene ontology (GO) analysis of the 114 genes increased in expression only when NBAS was depleted. Top three terms are shown of biological process (BP), molecular function (MF) and cellular component (CC) GO terms.* ***(B)*** *Overlaps with the three literature datasets identifying transcripts localised and/or translated at the ER: APEX-seq, ER fractionation sequencing and ER proximity specific ribosome profiling (Section 2.2.15). Bars show percentage overlap of all genes increased in expression when NBAS depleted (All NBAS Up), genes increased in expression when NBAS and UPF1 depleted (Common Up), genes increased in expression when NBAS depleted but not when UPF1 depleted (NBAS Only Up) and all genes decreased in expression when NBAS depleted (NBAS Down).* ***(C)*** *GO analysis of the 212 genes decreased in expression when NBAS depleted.*



	si-scr	si-UPF1	si-NBAS	si-UPF2
Total	n=3	n=3	n=3	n=3
Membrane	n=3	n=3	n=3	n=3
Cytoplasm	n=3	n=3	n=3	n=3

Figure 4.9: ***Experimental design of fractionation RNA-sequencing*** *Schematic of experimental design of fractionation RNA-sequencing experiment. Cells were double depleted with siRNA pools, fractionated and libraries prepared from total RNA as well as membrane-associated and cytoplasmic fractions. All conditions were sequenced in triplicate.*

fractionated would give a more realistic picture of the localisation of transcripts regulated by NBAS. Therefore, I next examined how NBAS depletion affected transcripts specifically at the ER compared to the cytoplasm and how this differs from depletion of core NMD factors UPF1 and UPF2.

Magda Maslon carried out fractionation of HeLa cells depleted of NBAS, UPF1, UPF2 or mock-depleted control cells and then prepared RNA-sequencing libraries from the cytoplasmic and membrane fractions, and total RNA. Cellular fractionation was performed as described before (Jagannathan et al., 2011) with some modifications (Longman et al., 2020). Briefly, cells were permeabilised with digitonin to extract cytosolic fraction, then lysed and the membrane fraction recovered by centrifugation (Figure 4.9). Paired-end illumina sequencing was performed by BGI Genomics, reads de-barcoded and fastq files returned for analysis.

After quality control and mapping to `GRCh38/hg38` using STAR (Section 4.2.2; Appendix B), HTSeq was used to count reads per gene feature and differential gene expression was analysed using DESeq2 (Figure 4.10A, B, C). In this dataset, depletion of UPF1 significantly affected the expression of 8475 genes, 4397 of which were increased, and 4078 were decreased. Although the numbers of genes

changing up and down in expression are similar, a skew can be seen with generally larger fold changes in genes increased in expression (Figure 4.10A). Comparison between dataset 1, described earlier in this chapter, and dataset 2, the fractionated experiment descried here, reveals a strong correlation between targets significantly changed upon UPF1 depletion in both datasets ($r = 0.88, p < 2.2\text{x}10^{-16}$; Figure 4.10D), demonstrating the comparable results of the two experiments. In response to NBAS depletion 4813 genes were significantly changed in expression in this dataset, far more than in dataset 1. However more genes were decreased in expression (2614) than increased (2199). Despite the smaller number of NBAS targets significantly changed in both datasets, compared to UPF1 targets, there was also a strong correlation of differential expression with dataset 1 upon depletion of NBAS ($r = 0.71, p < 2.2\text{x}10^{-16}$; Figure 4.10E). The correlations of UPF1-mediated changes, and especially NBAS-mediated changes, between the two RNA-sequencing datsets is much stronger than either dataset compared to the UPF1 and NBAS microarray (Longman et al., 2013, Figure 4.5D, E) Depletion of UPF2 significantly affected the expression of 1210 genes, far less than when NBAS was depleted, despite UPF2 often being classed as a core NMD factor. However, of these, there was a clear skew to increased expression with 818 significantly increased in expression and 392 decreased (Figure 4.10C).

As with the first dataset in this chapter, I was interested to see how the mRNA of other NMD factors was changed upon depletion of UPF1, UPF2 and NBAS. Differential expression in depleted cells compared to si-Scr treated samples showed that several NMD factors were increased in expression upon depletion of UPF1 and UPF2 (Figure 4.10F), stars indicate significant ($p < 0.05$). The core NMD factors UPF3B, SMG5 and SMG9 were significantly increased in expression upon depletion of UPF1 or UPF2. In contrast to datset 1 discussed earlier inn this chapter, the NMD repressor UPF3A was also significantly increased in expression upon depletion of UPF1 or UPF2. The core factors SMG1 and SMG6 were also increased in expression upon depletion of UPF1 only. Both UPF2 and NBAS

mRNA were slightly increased in expression upon depletion of UPF1 although neither reached significance. UPF1, UPF3B, SMG1 and SMG9 were all significantly decreased in expression upon depletion of NBAS, suggesting it does not trigger the NMD autoregulatory feedback loop.

Despite the strong correlation of expression changes in response to depletion of NBAS and UPF1 between this dataset and the one described earlier in this chapter, the correlation between NBAS and UPF1 changes in mRNA expression was weak, although still significant, unlike the same comparison in the first dataset ($r = 0.11, p = 4.6\text{x}10^{-8}$; Figure 4.11A). Whereas the correlation between changes in response to depletion of UPF1 and UPF2 was strong ($r = 0.89, p < 2.2\text{x}10^{-16}$; Figure 4.11B).

Next, differential expression analysis was carried out on the cytoplasmic and membrane fractions of each condition (Figure 4.11C-H). These analyses largely reflect the total RNA, UPF1 has by far the largest effect and causes the largest fold changes, especially for genes in the cytoplasmic fraction increased in expression (Figure 4.11E). Depletion of NBAS or UPF2 led to fewer changes in gene expression, compared with UPF1. NBAS has more of an effect in the membrane fraction (17.81% membrane genes vs. 3.51% non-membrane genes; Figure 4.11C, F), whilst UPF2 has more of an effect in the cytoplasmic fraction (6.60% membrane genes vs. 16.74% non-membrane genes; Figure 4.11E, H), However, many genes were found to be not only present in both membrane and cytoplasmic fractions but differentially expressed in both. Therefore, a classification of genes as membrane-associated or non-membrane was necessary.

Gene expression in both fractions in control cells was compared by using variant stabilised transformation of transcripts per million (TPM), averaged across replicates. A group of "membrane-associated" genes were defined that showed $\geq$2-fold higher expression in the membrane fraction, whereas all other expressed genes were deemed "non-membrane" genes (Figure 4.12A). The validity of cell fractionation was assessed by identifying genes in the previously described experimentally validated ER datasets (Section 4.2.3) in this dataset. As expected, the majority of genes present in any two ER datasets were found in the membrane-

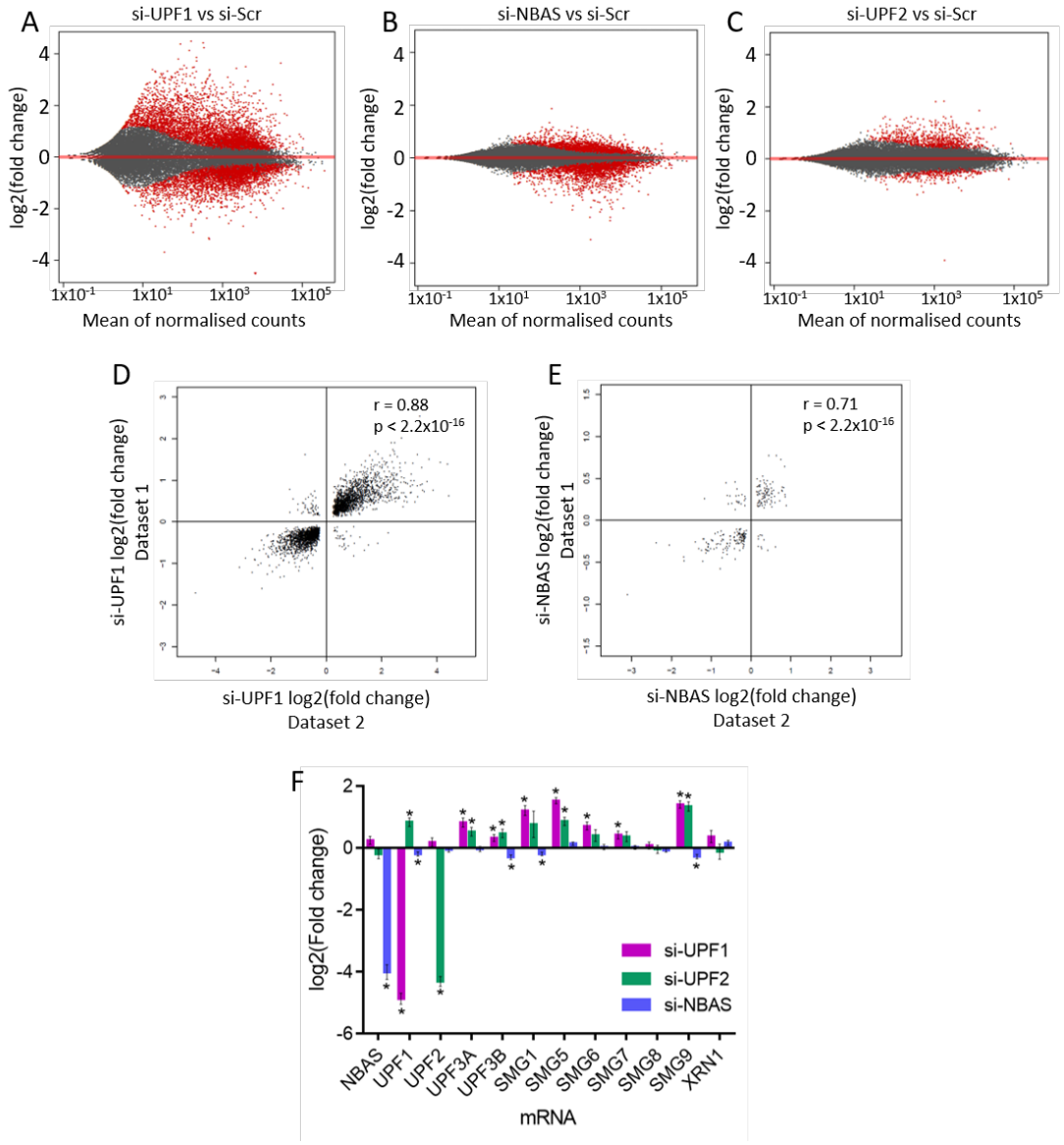


Figure 4.10: ***Validation of total mRNA expression changes from fractionation experiment***

Figure 4.10: ***Validation of total mRNA expression changes from fractionation experiment***

Differential gene expression compared to si-Scr control cells when cells were depleted of ***(A)*** *UPF1,* ***(B)*** *NBAS or* ***(C)*** *UPF2. MA plots show log2 fold-change of gene expression plotted over mean of normalised counts (gene expression). Significantly changed genes ($p < 0.05$) are indicated in red. Comparison between dataset 1, described earlier in this chapter, and dataset 2, the fractionated experiment reveals strong correlations between differentially expressed genes upon depletion of* ***(D)*** *UPF1 and* ***(E)*** *NBAS. Pearson's correlation coefficient and p value indicated in each plot.* ***(F)*** *Differential expression of NMD factors upon depletion of NBAS, UPF1 or UPF2. Change in abundance shown as log2(fold change) and standard deviation from DESeq2. Stars indicate significant changes ($p < 0.05$).*

associated set ($OR = 66.94, p < 0.001$, Fishers Exact test) (Figure 4.12B). Regulation by each NMD factor was measured by percentage of 'membrane associated' genes in membrane fraction, and of 'non-membrane' genes in cytoplasmic fraction that were increased in expression when the factor was depleted (Figure 4.12C, D). Knock-down of UPF1 led to a robust increase in the abundance of mRNAs in both membrane and cytoplasmic fractions, whereas depletion of NBAS led to an upregulation of a higher percentage of membrane-associated genes than of non-membrane genes ($5-$fold increase$, p < 0.001$, Exact Binomial test). In contrast, depletion of UPF2 led to preferential upregulation of 'non-membrane' genes in the cytoplasm ($>$ 2.5-fold increase$, p < 0.001$, Exact Binomial test). Of the 356 'membrane associated' genes in membrane fraction that were increased upon NBAS depletion, 48.6% were also increased upon UPF1 depletion. The gene most increased in abundance upon NBAS depletion was Matrix Gla protein (MGP) which is interesting since MGP plays a role in bone organization and NBAS mutations have been shown to lead to osteogenesis imperfecta, the commonest inherited bone fragility disorder (Balasubramanian et al., 2017).

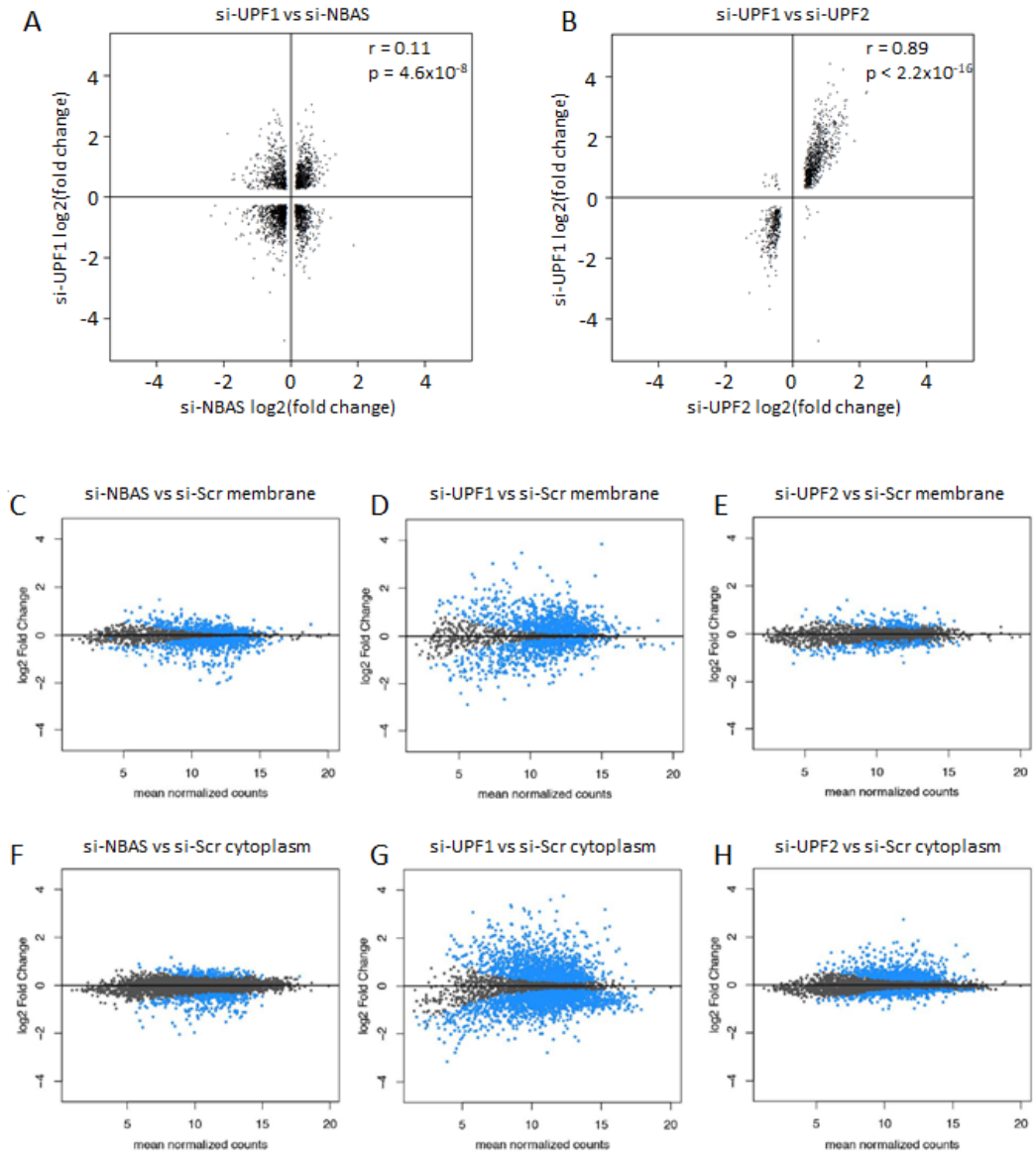
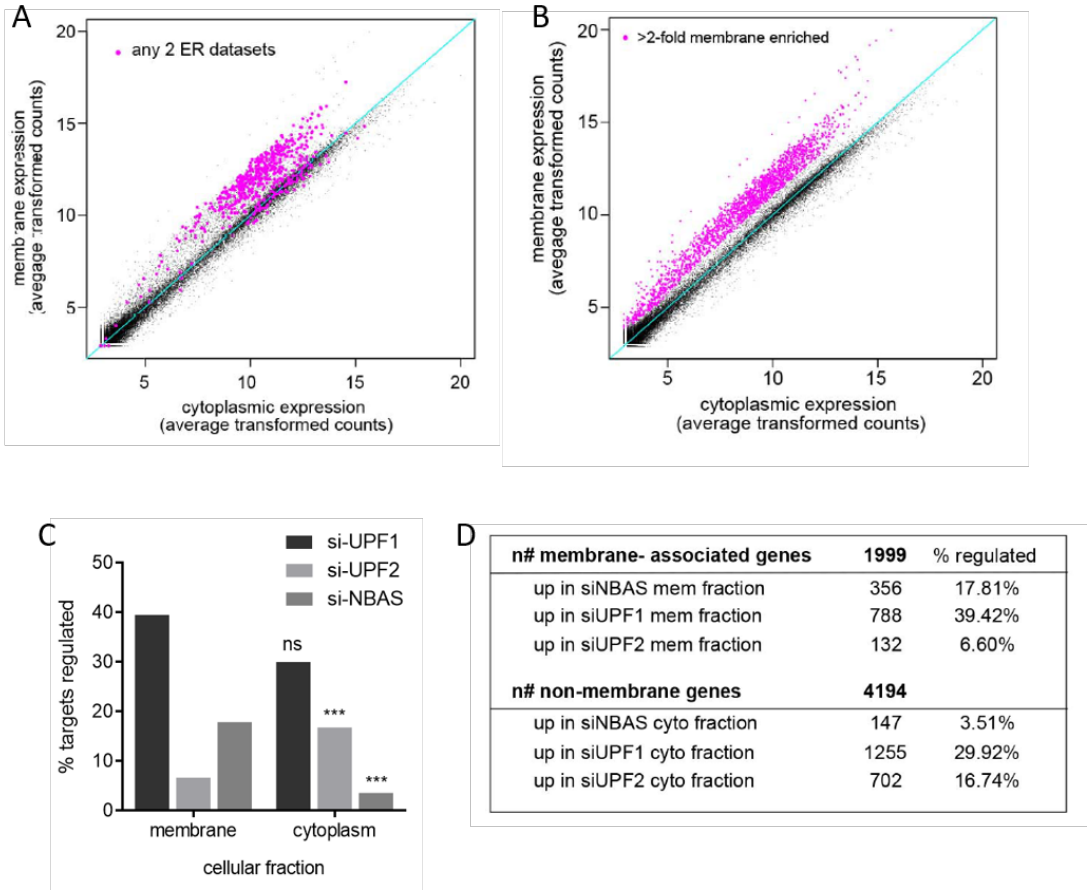


Figure 4.11: ***Correlations and differential expression of RNA fractions*** ***(A)*** *Comparing fold change of genes upon depletion of UPF1 and NBAS or* ***(B)*** *UPF1 and UPF2. Pearson's correlation coefficient and p value indicated in each plot. Differential gene expression analysis compared to si-Scr control cells in membrane fraction* ***(C,D,E)*** *and cytoplasmic fraction* ***(F,G,H)*** *when cells were depleted of* ***(C,F)*** *NBAS,* ***(D,G)*** *UPF1 or* ***(E,H)*** *UPF2. MA plots show log2 fold-change of gene expression plotted over mean of normalised counts (level of gene expression). Significantly changed genes ($p < 0.05$) are indicated in blue*



n# membrane- associated genes	1999	% regulated
up in siNBAS mem fraction	356	17.81%
up in siUPF1 mem fraction	788	39.42%
up in siUPF2 mem fraction	132	6.60%
n# non-membrane genes	**4194**	
up in siNBAS cyto fraction	147	3.51%
up in siUPF1 cyto fraction	1255	29.92%
up in siUPF2 cyto fraction	702	16.74%

Figure 4.12: ***Effect of NMD factor abrogation on expression of membrane-associated genes***

Figure 4.12: ***Effect of NMD factor abrogation on expression of membrane-associated genes***

(A) *Experimentally validated ER genes are present in the membrane fraction. Scatter plot of gene expression in membrane vs cytoplasmic fractions in mock-depleted cells, with experimentally validated ER genes present in at least 2 of the 3 ER datasets described in Fig. 2B are indicated in magenta. (Enrichment* $OR = 66.94, p < 0.0001$*, Fishers Exact test).* ***(B)*** *Scatter plot shows average transformed counts of each gene in membrane and cytoplasmic fractions in mock-depleted cells. Transcripts >2-fold increased in expression in membrane fraction over cytoplasmic section are shown in magenta and identified as "membrane-associated genes".* ***(C)*** *NBAS preferentially regulates membrane-associated targets. Plot shows percentage of "membrane-associated" genes in the membrane fraction and of "non-membrane" genes in the cytoplasm that are regulated (*$p < 0.05$*) by each NMD factor. NBAS regulates over fivefold higher percentage of membrane-associated genes than non-membrane genes (*$p < 0.001$*, exact binomial test). In contrast, UPF1 regulates a similar percentage of both membrane-associated and non-membrane genes (1.3-fold higher percentage of membrane-associated genes,* $p < 0.001$*, exact binomial test), and UPF2 regulates 2.5-fold higher percentage of non-membrane genes than membrane genes (*$p < 0.001$*, exact binomial test).* ***(D)*** *Summary table of number and percentage of genes regulated by NMD factors in each cellular fraction.*

4.3 Discussion

4.3.1 NBAS regulates a subset of UPF1 targets

Although comparing the UPF1 targets identified here and those identified in previous studies produces a small overlap, this was not unexpected as overlap between all NMD datasets is weak, evidence of the heterogeneous nature of the NMD response and the variability in measuring it. In fact, better overlap is seen between this study and others than between previous studies (Figure 4.5C). Added to that, as well as NBAS having at least one non-NMD function in Golgi-ER retrograde transport, UPF1 has several known functions beyond its role in NMD. These include regulation of RNA on multiple levels such as RNA splicing, transport, translation, and mRNA turnover (Atkin et al., 1995; Lykke-Andersen et al., 2000; Mendell et al., 2002; Wilkinson, 2005; Nazarenus et al., 2005). UPF1 shuttles between the nucleus and cytoplasm, indicating that the protein has potential roles in both the nucleus and the cytoplasm (Singh et al., 2019). For example, UPF1 has been identified as interacting with SRSF1 which plays a role in preventing exon skipping, ensuring the accuracy of splicing and regulating alternative splicing (Aznarez et al., 2018), and UPF1 stabilizes primary retroviral RNA so that UPF1 downregulation by siRNA decreases HIV RNA and diminishes the synthesis of virion structural proteins (Ajamian et al., 2008). In addition, the interaction of UPF1 with Staufen1-mediated mRNA decay directly affects RNA stability (Kim et al., 2005; Park and Maquat, 2013). All these roles affect RNA processing and so could be the cause of changes to transcription rate of various transcripts, which would confound NMD studies of transcript expression after UPF1 depletion.

Identification of NBAS targets has only previously been carried by microarray analysis of NBAS depleted cells (Longman et al., 2013). Correlation between the effects measured in that study and the effects described here was not strong but was significant (Figure 4.5E). However, the difference in technical approach is

likely to account for a large amount of the variability seen. Although the level of knock down of NBAS achieved was similar to that of Longman et al. (2013) (23.4% compared to ∼20%), the number of targets identified in this study was ∼36% of the number identified there and the magnitude of change generally lower. However, the number of NBAS targets that were also found to be UPF1 targets was very similar (106 and 95). Intersecting these two lists of joint targets reveals only three mRNAs upregulated in all four experiments: NCOA7, SNAP25, MVD. Whilst NCOA7 is a nuclear receptor coactivator, both MVD and SNAP25 are involved in intracellular transport, MVD in transport to the Golgi and SNAP25 in vesicle fusion with the cell membrane to bring about exocytotic events.

NBAS has previously been shown to contribute to the autoregulatory loop affecting levels of transcripts that encode NMD factors. RT-qPCR analysis of total cellular RNA from NBAS-depleted HeLa cells showed a significant increase in mRNA of UPF1, DHX34, UPF2, SMG5, SMG7 and SMG8 (Longman et al., 2013). However, neither of the datasets analysed in this chapter identified significant increases in the mRNA level of any NMD factors, indeed in the second dataset, NBAS depletion lead to significant decrease in mRNA level of several NMD factors including UPF1. It was also shown by Longman et al. (2013) that NBAS mRNA was significantly increased upon UPF1 depletion, but this was also not replicated by either of the differential expression analyses. Longman et al. (2013) only showed changes measured by qPCR, so this difference in result could be due to technical differences between the two techniques, however both the RNA-sequencing datasets analysed here reproduced the qPCR results showing increased mRNA of NMD factors upon depletion of UPF1.

Taken together the caveats discussed here show that identifying *bona fide* NMD targets using only steady-state gene expression measures is challenging. Since a similar knock down of NBAS by siRNA resulted in variable effect, an elegant improvement to this experiment would be the creation of a cell line where NBAS has an auxin-inducible degron tag. This could be created using CRISPR/Cas9 technology, utilising the methodology I have optimised (Chapter 3). The addition

of this tag would allow acute and efficient knockdown of NBAS at the protein level (Li et al., 2019). These cells could then be assessed for differential expression by RNA-sequencing but a more accurate measure of NMD activity would be measuring change of RNA stability since this is the process NMD directly affects. I have shown that there is a significant overlap between genes increased in expression when NBAS is depleted and those increased in expression when the core NMD factor UPF1 is depleted, in agreement with NBAS' previously documented role as an NMD factor (Longman et al., 2007; Anastasaki et al., 2011; Longman et al., 2013, 2020). NBAS affects a smaller number of genes than UPF1 but of those 45.5% are commonly affected. Characterising the overlapping genes showed an enrichment for secretome genes, translated at the ER, compared with UPF1 targets. This suggests NBAS and UPF1 are working together to regulate transcripts specifically translated at the ER.

Although the three ER-localisation studies were performed in HEK293T cells, the overlap between NBAS targets and each datasets is equally strong as the overlap between each of the datasets, showing that NBAS targets are transcripts not just part of the secretome, but actively translated at the ER translocon. The overlap of the common UPF1 and NBAS targets with these data is even higher, confirming that the NBAS targets also regulated by UPF1 are precisely those that are translated at the ER. This indicates that the NBAS targets not regulated by UPF1 are likely not NMD targets but linked to NBAS' role in Golgi-ER retrograde transport.

GO analysis of UPF1-independent NBAS targets, resulted in membrane and secretome terms being significant in all three categories. This is not unexpected since NBAS is involved in post-translational processing of secretome proteins and is localised to the ER so is likely to affect transcripts also involved in this pathway. However, since these changes are UPF1-independent, it is likely they are NMD-independent and therefore it is not clear how NBAS is affecting mRNA abundance rather than protein abundance and perhaps indicates a feedback mechanism that links reduction in secretome protein processing to increased expression of the

mRNA of those genes. Compared to all mRNAs that increased in expression upon depletion of NBAS, those that decreased in expression were significantly less enriched in the three experimental ER datasets, suggesting that they are indirect effects since NBAS is located at the ER.

It would also be interesting to repeat this profiling in different cell lines, specifically more physiological cell lines than HeLa cells and those where secretion is more active such as U2OS cells which are bone osteosarcoma epithelial cells. This would be especially relevant as NBAS compound heterozygous variants have been identified in patients as a cause of atypical osteogenesis imperfecta (Balasubramanian et al., 2017).

4.3.2 ER-NMD is important for ER homeostasis

Evidence of a link between UPF1 or NBAS depletion and increased unfolded protein response (UPR) is important because previous work has suggested a connection between NMD and ER stress and the UPR (Section 1.4.3 Mendell et al., 2004; Gardner, 2008; Karam et al., 2015). It has previously been shown that the NMD factors UPF1 and UPF2 decreased the risk of death in an *in vitro* model of Amyotrophic Lateral Sclerosis (ALS) (also known as motor neuron disease (MND)) (Barmada et al., 2015). Patients commonly harbour a mutant form of TDP43 that leads to aberrant cleavage and aggregation of the C-terminal region. Primary rat neurons expressing human TDP43 with the patient mutation, were found to have better survival when either UPF1 or UPF2 was overexpressed, but not by expression of an RNA helicase defective UPF1 mutant protein, possibly indicating a neuroprotective function of NMD (Barmada et al., 2015). It seems likely that protection against aggregates would have increased importance at the ER, where build-up of misfolded protein is of particular risk, and therefore it is likely that the ER-NMD pathway is important for ER homeostasis and protection against ER stress.

Recent evidence (Karam et al., 2015; Carreras-Sureda and Hetz, 2015; Li et al., 2017c) supports a role for NMD in modulating the ER stress response by ensuring appropriate activation of the UPR; a pathway that senses and responds to excessive amounts of misfolded proteins in the ER. Consequently, an appropriate level of NMD activity is required to protect cells from detrimental stress response activation. In particular, pathological UPR activation is recognised as a major causal mechanism in neurodegenerative diseases including Alzheimer's disease, Parkinson's disease and prion diseases (Shah et al., 2017; Yan et al., 2019; Torkzaban et al., 2020; Li et al., 2020). Therefore, modulation of the activity of the ER-NMD pathway could be crucial to regulate the response to ER stress. It is tempting to speculate that increasing the activity of the ER-NMD pathway will have a neuroprotective effect by limiting excessive UPR signalling in neurodegeneration.

4.3.3 Future work

As with all genome-wide studies it would be prudent to validate my findings using orthogonal approaches. Firstly, the confirmation of expression changes of identified UPF1 and NBAS targets upon depletion of the relevant factors by qPCR. Although the extent of change may be different due to the difference in measurement techniques, the direction and relative magnitude should be reproducible. In fact, it would be interesting to see how consistent the change in magnitude was between genes, considering the seemingly low negative fold change in both UPF1 and NBAS measured by DE and the ∼80% decrease in mRNA level measure by qPCR of the same samples. However, qRT-PCR only reports the average mRNA levels of an NMD target in a population of cells and thus preclude insight into differences within mRNA populations. Furthermore, like total RNA differential expression analysis, it is reliant on steady-state mRNA levels which are likely affected by differences in nuclear export rates and differences in the onset and efficiency of translation of different mRNAs, in addition to NMD degradation. These confounding factors mean differentiating legitimate RNA targets of NMD

from downstream effects is near impossible, even with the high number of replicates and sequencing depth in this study. Therefore, to understand fully the effects of any given NMD factor, RNA stability should be measured and change in stability of transcript used to define NMD targets.

4.3.4 Summary

In conclusion, in this chapter I have shown that NBAS regulates a subset of NMD targets that are specifically membrane-associated and translated at the ER. I have also shown that a subset of UPF1 targets are ER-translated, and that these transcripts are also regulated by NBAS. Altogether, these results add to the previous work indicating the existence of a specific ER-NMD pathway that targets for degradation mRNAs that are translated at the ER.

The key findings presented in this chapter have been published in the following article, which is attached as Appendix C:

Longman D, ***Jackson-Jones KA****, Maslon MM, Murphy, LC, Young, RS, Stoddart, JJ, Hug, N, Taylor, MS, Papadopoulos, DK, & Cáceres, JF (2020) Identification of a localised nonsense-mediated decay pathway at the endoplasmic reticulum Genes & Dev. 2020. 34: 1075-1088. doi:10.1101/gad.338061.120*

Chapter 5

Quantifying changes in RNA stability reveals many expression changes are independent of NMD

5.1 Introduction

Measurements of mature RNA are in themselves poorly informative of the real transcriptional state of genes (Furlan et al., 2020). For instance, the detection of a mature RNA is typically taken as indication that the corresponding gene is transcriptionally active. This is not necessarily the case for highly stable RNAs, which might persist long after the gene has become silent. Furthermore, identifying the cause of a change in mature RNA abundance is difficult without determining the corresponding RNA dynamics. Therefore, several studies have attempted to extend traditional total RNA differential expression by using integrative analyses of premature and mature RNA abundances (Zeisel et al., 2011; Gray et al., 2014; La Manno et al., 2018; Bergen et al., 2020). However, the key limitation of all these studies is having considered intronic expression as a proxy of synthesis rates. RNA abundance is determined by the combined action of three main processes: the synthesis of premature RNA, its processing into mature RNA and finally degradation of the mature mRNA (Figure 5.1A). The dynamics of transcripts metabolism are set by the rates governing the kinetics of those steps, ultimately setting the abundance of premature and mature RNA species (Orphanides and Reinberg, 2002; Shalem et al., 2008; Friedel and Dölken, 2009; Rabani et al., 2011; Gray et al., 2014; Eser et al., 2016; de Pretis et al., 2017). Therefore, it is important to incorporate both RNA synthesis and RNA maturation when trying to elucidate RNA degradation from total RNA abundance.

5.1.1 Benefits of differential stability analysis to identify targets of NMD

NMD is defined as rapid degradation of target RNAs (Gatfield and Izaurralde, 2004; Kebaara and Atkin, 2009), meaning those transcripts will be unstable. Therefore, transcripts with increased stability upon NMD factor depletion are likely to be direct NMD targets regulated by that factor. Whilst traditionally,

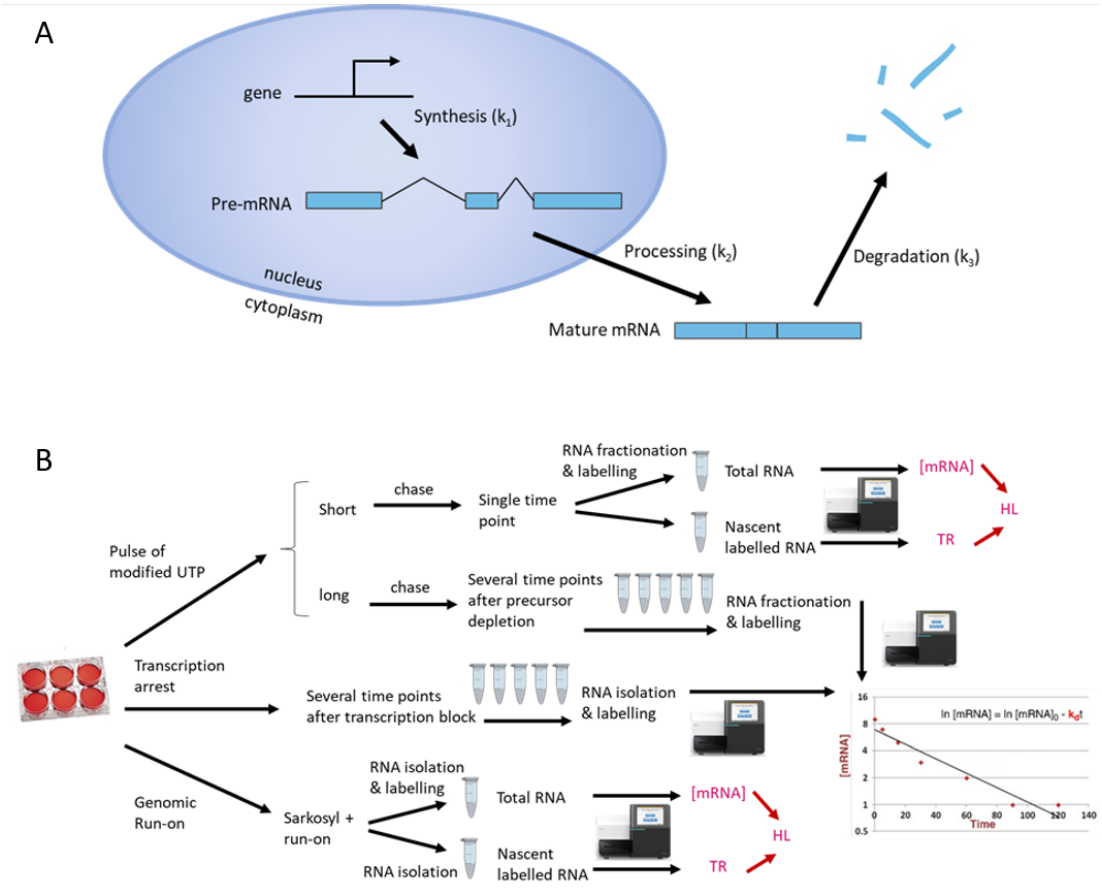


Figure 5.1: ***Measuring RNA stability requires more than total RNA***

Figure 5.1: ***Measuring RNA stability requires more than total RNA*** ***(A)*** *RNA abundance is determined by three main processes: the synthesis of premature RNA, its processing into the mature form, and the degradation of the mature RNA. The dynamics of transcripts metabolism and ultimately the abundance of both premature and mature RNA species, are set by the rates governing the kinetics of those steps (k_1, k_2, k_3). Figure adapted from Furlan et al. (2020).* ***(B)*** *Three main types of experiment have been used to measure RNA stability on a genomic scale. Top - in vivo metabolic labelling with uridine analogues using either time course measurements or RNA fractionation to produce values for steady state RNA abundance [mRNA] and transcription rate (TR). This allows calculation of RNA half life (HL). Middle - inhibition of transcription followed by RNA abundance measurements over time. Decreasing abundance can then be plotted to provide a decay curve from which kd and HL can be determined. Bottom - genomic run-on techniques which combine metabolic labelling with inhibition of new transcription initiation using sarkosyl. RNA fractionation then allows for steady state RNA abundance [mRNA] and transcription rate (TR) to used to calculate of RNA half life (HL). Figure adapted from Pérez-Ortín et al. (2013).*

differential gene expression upon depletion of a putative NMD factor has been used to identify probable targets of that factor, as described in the previous chapter, there are several drawbacks to this method. Most significantly, the fact that NMD does not affect transcription directly, and therefore increases or decreases in gene expression could be due to altered transcription rate caused by secondary or downstream effects. Accordingly, measures of transcript stability have been more recently used to identify transcripts targeted by NMD. Many different techniques have been employed to measure differential stability genome wide, upon depletion of an NMD factor (Figure 5.1B).

Traditional measurements of transcript stability have relied on transcription block followed by multiple sampling times to track decay of signal over time. This consists of measuring abundance of a transcript at multiple time points to give a decay constant from which a half-life can be calculated. This half-life can then be compared between different conditions such as control cells and cells where an NMD factor such as UPF1 has been depleted.

However, these methods are highly invasive, affect cell viability, and could alter various pathways including translation and RNA decay, and confound interpretation (Friedel and Dölken, 2009; Wada and Becskei, 2017; Russo and Angelini, 2014). Furthermore, it has been suggested that transcription and decay are coupled processes, such that inhibiting one is likely to directly influence the rate of the other (Sun et al., 2012; Haimovich et al., 2013). Therefore, there has been a search for less disruptive techniques to measure RNA stability.

The majority of recent methods use metabolic labelling, which consists of the addition of a uridine analogue to cells which is specifically incorporated into newly transcribed, or nascent, RNA without needing to block transcription. Run-on techniques involve pulse labelling for a set time period with the uridine analogue which is then incorporated into the nascent RNA and then a wash out step so endogenous uridine is once again the only option. In this way only the RNA transcribed in the set time period will be measured at the multiple time points, thereby allowing the RNA abundance to be measured without being affected by transcription rate.

Global run-on sequencing (GRO-seq) is a modified Run-On protocol to isolate nascent RNA allowing unbiased mapping of nascent transcripts genome-wide (Core et al., 2008). GRO-seq combines metabolic labelling with transcription block. This method uses 5'-bromo-uridine (BrU) to label newly synthesised RNA, then nascent RNA is isolated and ultimately converted into an RNA-sequencing library. After isolation of nuclei, sarkosyl is added, which prevents the attachment of RNA polymerase to the DNA and stops *de novo* assembly of the pre-initiation complex. Therefore, only transcripts that were initiated at the time of nuclei isolation will be further elongated by engaged RNA polymerase, to allow incorporation of radioactive nucleotides. Due to this, GRO-seq is a sensitive method to estimate transcriptional activity throughout the entire genome and has generated detailed results, for example, on RNA polymerase II density at different classes of protein coding genes and evidence that that RNA is synthesised bi-directionally at most mammalian promoters (Huang et al., 1996; Core et al., 2008; Andersson et al., 2014; Duttke et al., 2015).

Metabolic labelling of RNA has increasingly been used to measure stability by fractionation of labelled RNA and isolation of nascent transcripts **without** the use of transcription inhibition (Johnson et al., 1991; Dölken et al., 2008; Schwanhäusser et al., 2011). This is because under steady-state conditions, RNA transcription is balanced with RNA decay to maintain stable total transcript levels. Therefore, *de novo* synthesis rates are higher for unstable transcripts. Thus, assuming steady-state conditions, mRNA decay rates can be determined by analysing the ratio of nascent to total RNA derived from the same sample (Dölken et al., 2008). BrU immunoprecipitation chase-deep sequencing (BRIC-seq) also uses BrU to label nascent RNA but in physiologically undisturbed conditions since no inhibitors of transcription or translation are used. The "chase" here refers to a subsequent removal of BrU after the initial treatment, leading to chronological decreases of BrU-labeled RNAs. The RNA half-life of each transcript is calculated from the decreasing number of BrU-labeled RNA sequence tags measured by deep sequencing of BrU-labeled RNAs at multiple time points (Imamachi et al., 2014).

Another common uridine analogue is 4-thiouridine (4sU). This can be used to label nascent RNA in the same way as BrU (Dölken et al., 2008; Miller et al., 2011; Rabani et al., 2011; Sun et al., 2012; Fuchs et al., 2014; Russo and Angelini, 2014; de Pretis et al., 2015; Marzi et al., 2016; Schwalb et al., 2016; de Pretis et al., 2017; Michel et al., 2017; Wissink et al., 2019). Russo and Angelini (2014) used 4sU labelling and then purified total and nascent RNA from the same cells. Using digital droplet PCR (ddPCR) they showed that by using both RNA fractions it was possible to measure RNA stability at a single time point and without the use of transcription inhibitors. I extended this approach by using RNA-sequencing to identify NMD targets genome-wide, by identifying transcripts increased in stability when NMD factors are depleted.

5.1.2 Aims

In this chapter I aim to understand how NBAS affects transcript stability by identifying targets which increase in stability when NBAS is depleted and comparing these with UPF1-dependent increases in stability.

To do this, I will further analyse the total RNA described in Chapter 4 by considering both intronic and exonic reads and also used matched nascent RNA to assess stability independent of gene expression. To identify genuine NMD targets, I will compare changes in stability with changes in expression to determine transcripts whose stability and expression were increased upon depletion of UPF1 and/or NBAS. This will allow a more accurate comparison of NBAS NMD targets with those of the core NMD factor UPF1.

In summary, the main aims of this chapter are:

a) To use exonic and intronic reads from matched total and nascent RNA-sequencing to determine stability changes at a single time point without transcription inhibition.

b) To compare the resulting differentially stabilised transcripts with the differentially expressed targets identified in Chapter 4 to determine *bona fide* NMD targets.

5.2 Results

5.2.1 Defining exonic and intronic reads

I first wanted to differentiate changes of RNA stability from changes in transcription. To do this, reads needed to be defined as originating from pre-mRNA or mature mRNA. Normally, the majority of reads produced by RNA-sequencing are exonic and this is further emphasised by mapping using annotation files that only map to coding regions. However, in this analysis I decided to normalise each gene to its own internal control of pre-mRNA level. For this reason, it was necessary to produce and keep intronic reads. Therefore, no poly-A selection protocol was used when preparing the RNA libraries. Furthermore, the nascent RNA samples had a very low amount of RNA and there were concerns that any manipulation would reduce this further. Therefore, no ribosomal RNA (rRNA) depletion was carried out on the samples. Total RNA samples were treated the same as nascent RNA, even though there was an abundance of RNA in these samples, to ensure the samples were as matched as possible for downstream analysis.

Unlike for differential expression (Chapter 4) kallisto pseudoalignment is not appropriate for this analysis as reads need to be mapped precisely so they can be defined as exonic or intronic rather than counted per transcript. Therefore, each lane of each replicate was mapped to `GRCh38/hg38` human genome assembly separately using STAR. STAR was used because it is a splice-aware aligner, meaning the tool can deal with reads that span across introns (Dobin et al., 2013). Mapping efficiency was high across all samples with an average of 89.2% of reads being successfully aligned (Figure 5.2A; Table 5.1). Since this required defining both intronic and exonic reads and standard transcriptomic mapping doesn't readily accept intronic reads, alignment to `GRCh38/hg38` with STAR was performed without an annotation file, with default parameters. This is because when provided with an annotation of gene regions, an aligner will only keep reads that map to

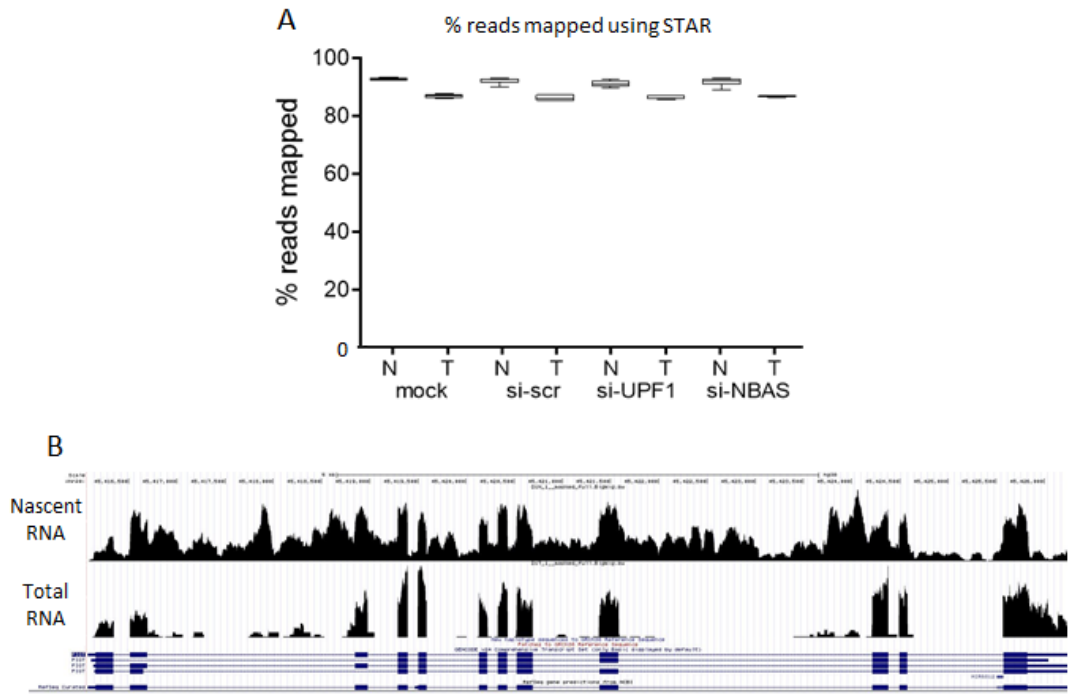


Figure 5.2: ***Mapping reads with STAR allows identification of intronic read coverage***

(A) *Percentage of reads per library successfully mapped to* `GRCh38/hg38` *(n=5) for each condition and nascent (N) or total (T) RNA.* ***(B)*** *Representative read pile up tracks show the high level of intronic reads in nascent RNA, and lower level but still present intronic reads in total RNA.*

these defined regions and may even preferentially map reads to these regions than non-coding regions even if the probability is higher. Visualising the read pileups revealed a high amount of intronic signal in nascent RNA samples, and a lower amount but still visible intronic signal in total RNA samples (Figure 5.2B).

Gene structure annotation was extracted from Gencode v27 to define intronic and exonic regions of each gene. In the case of multiple overlapping transcripts for the same gene, overlapping exon coordinates were merged into union exons. Subtraction of these from transcripts extents produced obligate intron annotations (Figure 5.3A). Reads that overlapped obligate intron regions were classed

	Total analysed read (millions)	Aligned reads (millions)	Mapping efficiency (%)	Multimapping (%)
mock total	99.6	92.3	92.6%	3.54%
mock nascent	102.7	89.1	86.8%	8.93%
si-scr total	96.0	87.5	91.1%	4.31%
si-scr nascent	104.9	90.7	86.5%	8.97%
si-UPF1 total	109.1	99.9	91.6%	4.05%
si-UPF1 nascent	103.6	89.8	86.7%	8.82%
si-NBAS total	96.8	89.1	92.0%	4.18%
si-NBAS nascent	99.1	85.4	86.2%	9.06%

Table 5.1: ***Mapping efficiency of total and nascent RNA with STAR*** *Average number of reads per library in each condition, number and percentage of reads mapped to* `GRCh38/hg38` *using STAR and no annotation file, and percentage of reads failed due to multi-mapping (n=5).*

as intronic reads since it was presumed they must originate from unspliced pre-mRNA. Only reads that were fully within union exonic regions were counted as exonic reads (Figure 5.3B). Reads in union exon and obligate intron regions were counted per gene using bedtools.

5.2.2 Using exonic and intronic reads from paired nascent and total RNA normalises for mRNA processing and transcription rate

Measuring stability of RNA using analysis of premature and mature RNA abundances is an improvement on the assumptions and caveats of using total RNA-sequencing alone (Section 5.1.1) and is being increasingly used in genome-wide studies (Zeisel et al., 2011; Gray et al., 2014; La Manno et al., 2018; Bergen et al., 2020). However, the key limitation in this type of analysis is assuming intronic expression as a proxy of synthesis rates, when intronic reads actually result from the joint action of two processes: the synthesis of premature RNA and its pro-

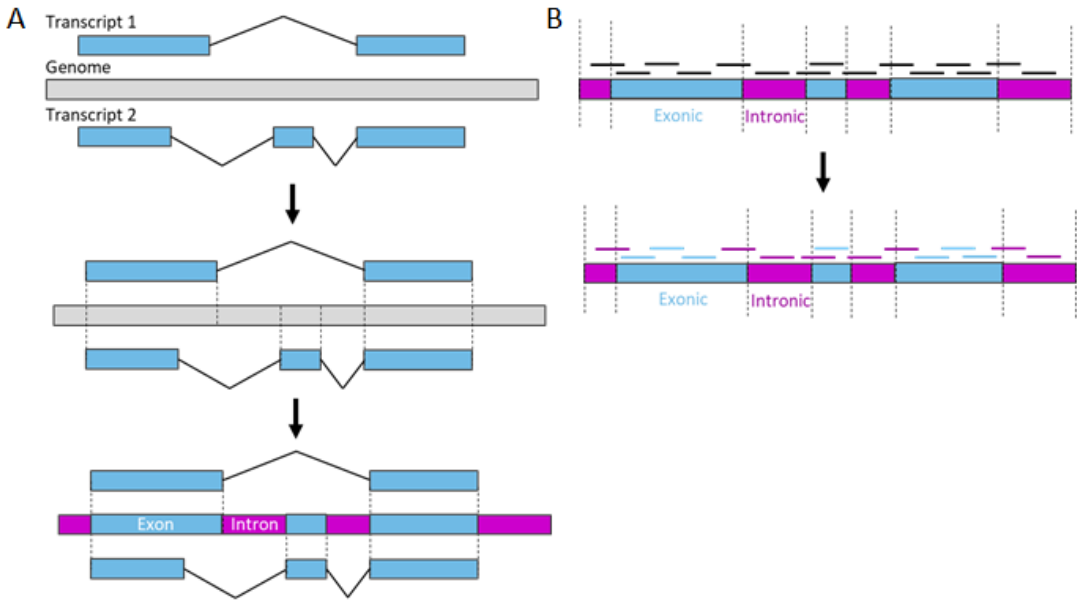


Figure 5.3: ***Schematic of custom analysis pipeline to calculate change in stability of genes in depleted cells compared to control***

(A) *"Exonic" regions (blue) were defined as any part of the gene body that is ever defined as an exon and "intronic" regions (purple) as those parts of the gene which are never defined as an exon, resulting in facultative exons and obligate introns.* ***(B)*** *"Exonic" reads were defined as those that exclusively overlapped with "exonic" regions and "intronic" reads as those which even partially overlapped with defined "intronic" regions.*

cessing into mature RNA (Figure 5.1A). Therefore, to measure more accurately the kinetics affecting each RNA species, direct measurement of transcription rate should be included in calculations in addition to the use of intronic reads to account for RNA processing.

Once counts for exonic and intronic reads in each sample had been determined, a dataset with total exonic, total intronic, nascent exonic and nascent intronic read counts for each gene was produced for each sample, to give four values for each gene in each sample (Figure 5.4A). Odds ratio (OR) calculations were performed using these four values, according to the follow equation:

$$OR = (\textit{total exonic (mature mRNA)} / \textit{total intronic (pre-mRNA)}) / (\textit{nascent exonic (pre-mRNA)} / \textit{nascent intronic (pre-mRNA)})$$

This produced a single number for each gene. This allowed measurement of change in OR when either NBAS or UPF1 was depleted, or when cells were treated with si-Scr as a control, compared to mock treated cells.

Globally, an increase in stability was observed when either UPF1 or NBAS were depleted which was significant compared to treatment with si-Scr (Figure 5.4B). The shift when UPF1 was depleted was a 1.14-fold increase in stability compared to si-Scr ($p < 0.0001$; paired t-test) and when NBAS was depleted the shift compared to si-Scr was a 1.04-fold increase ($p < 0.0001$; paired t-test), consistent with my previous results that NBAS affects fewer transcripts than UPF1 and the resulting changes are more subtle.

To calculate significance, paired t-tests were carried out on a per gene basis. This resulted in 859 transcripts being identified as increased in stability and 199 decreased in stability upon depletion of UPF1 (Figure 5.4C). This represents a 4.31-fold enrichment of increased stability ($p < 0.001$; Fisher's exact test) when UPF1 is depleted, in agreement with UPF1's role as a core NMD factor. Upon depletion of NBAS, 202 transcripts were increased in stability whereas 278 transcripts were decreased in stability (Figure 5.4D), showing a slight enrichment in

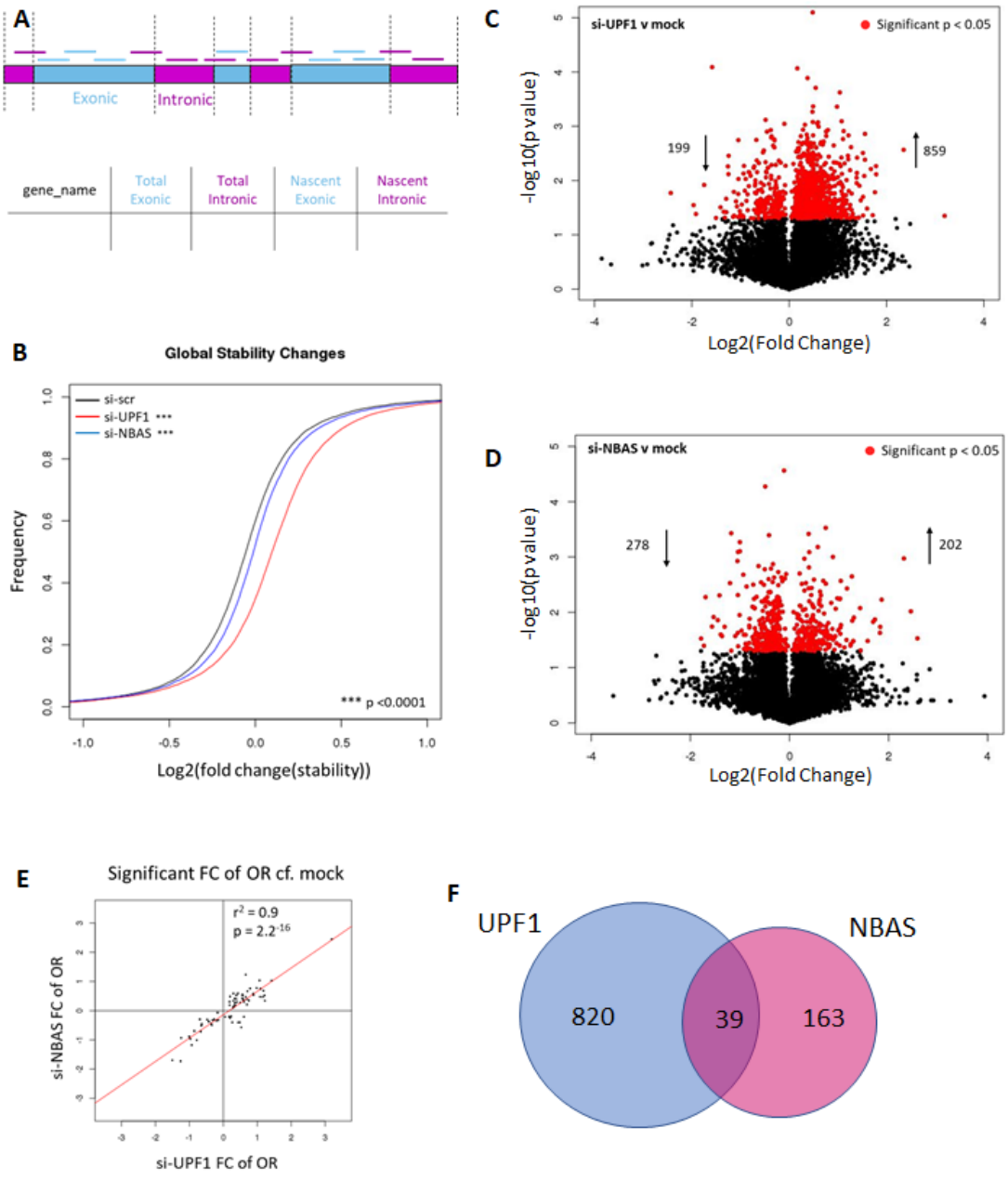


Figure 5.4: ***Measuring stability using paired nascent and total RNA***

Figure 5.4: ***Measuring stability using paired nascent and total RNA*** ***(A)*** *Total exonic, total intronic, nascent exonic and nascent intronic read counts for each gene were produced for each sample, to give four values for each gene in each sample. OR calculations were performed using these four values.* ***(B)*** *RNA stability (OR) for each gene was normalised to mock treated cells to give a measure of fold change upon si-Scr, si-UPF1 or si-NBAS treatment. Cummulative distributions show the global change in stability in si-Scr control (black line), si-NBAS (blue line) and si-UPF1 (red line) treated cells. Both UPF1 and NBAS depletion resulted in a significant global increase in RNA stability ($p < 0.0001$; paired t-test). Volcano plots show log2 fold change in stability and –log10 p-value of transcripts upon depletion of* ***(C)*** *UPF1 and* ***(D)*** *NBAS. Significant changes are indicated in red ($p < 0.05$, not multiple testing corrected, paired t-test).* ***(E)*** *Correlation of significant (not multiple testing corrected) stability changes upon depletion of NBAS and UPF1. Pearson's correlation coefficient and p value are indicated.* ***(F)*** *Overlap of significantly increased (not multiple testing corrected) stability changes upon depletion of NBAS and UPF1.*

decreased stability (non-significant). Plotting transcripts significantly changed in stability upon UPF1 and NBAS depletion revealed a strong correlation in both stabilised and destabilised transcripts (Figure 5.4E), although only 39 transcripts were increased in stability when either factor was depleted (Figure 5.4F).
However, after correcting for multiple testing using a Bonferroni correction, no RNAs showed a significant change in stability upon UPF1 depletion or NBAS depletion. This is likely due to the relatively low number of replicates in genomic data terms and that this methodology does not account for the variance between replicates and samples. This reasoning is further supported by looking at individual genes (Figure 5.5). Whilst the decrease in stability of both NBAS (Figure 5.5A) and UPF1 (figure 5.5B) mRNA can be seen in the appropriate depleted sample, only NBAS is significant even before multiple testing correction, which is likely due to the variability within conditions, in this case UPF1 mRNA in control mock cells. Figures 5.5C and D show examples of a gene with increased

stability when UPF1 was depleted (MDK) and a gene with increased stability when either UPF1 or NBAS was depleted (EDARADD). Only the difference between EDARADD mRNA stability in si-UPF1 treated and mock control cells is significant ($p < 0.05$).

5.2.3 DESeq2 provides an appropriate framework to analyse differential transcript stability

The package DESeq2 (Love et al., 2014) was developed to identify deferentially expressed genes in RNA-sequencing datasets, given two conditions eg. control and experimental treatment, in this case represented by si-Scr and si-NBAS. Before calculating differential expression, DESeq2 estimates dispersion of each gene which considers the variance and the expression level. The dispersion estimates are inversely related to the mean, and therefore reflect the variance in gene expression for a given mean value. When replicate number per condition is low (3-6 replicates), the estimates of variation for each gene are often unreliable due to the large differences in dispersion for genes with similar means. To address this problem, DESeq2 shares information across genes to generate more accurate estimates of variation based on the mean expression level of the gene. The dispersion for each gene is estimated using a maximum likelihood estimation which assumes that genes with similar expression levels have similar dispersion and calculates the most likely estimate of dispersion. Therefore, using the DESeq2 algorithm should account for the within-group variability seen in my dataset (Figure 5.5) and provide a more robust measure of stability.

Furthermore, DESeq2 can be utilised to find differential changes when there are more than one factor influencing the counts, for example total or nascent as well as treated or untreated. This ability to analyse multi-factor experiments has been suggested as useful for analysing ribosome profiling where two conditions are compared whilst also accounting for the comparison of the bound and input samples for each condition (Love et al., 2014).

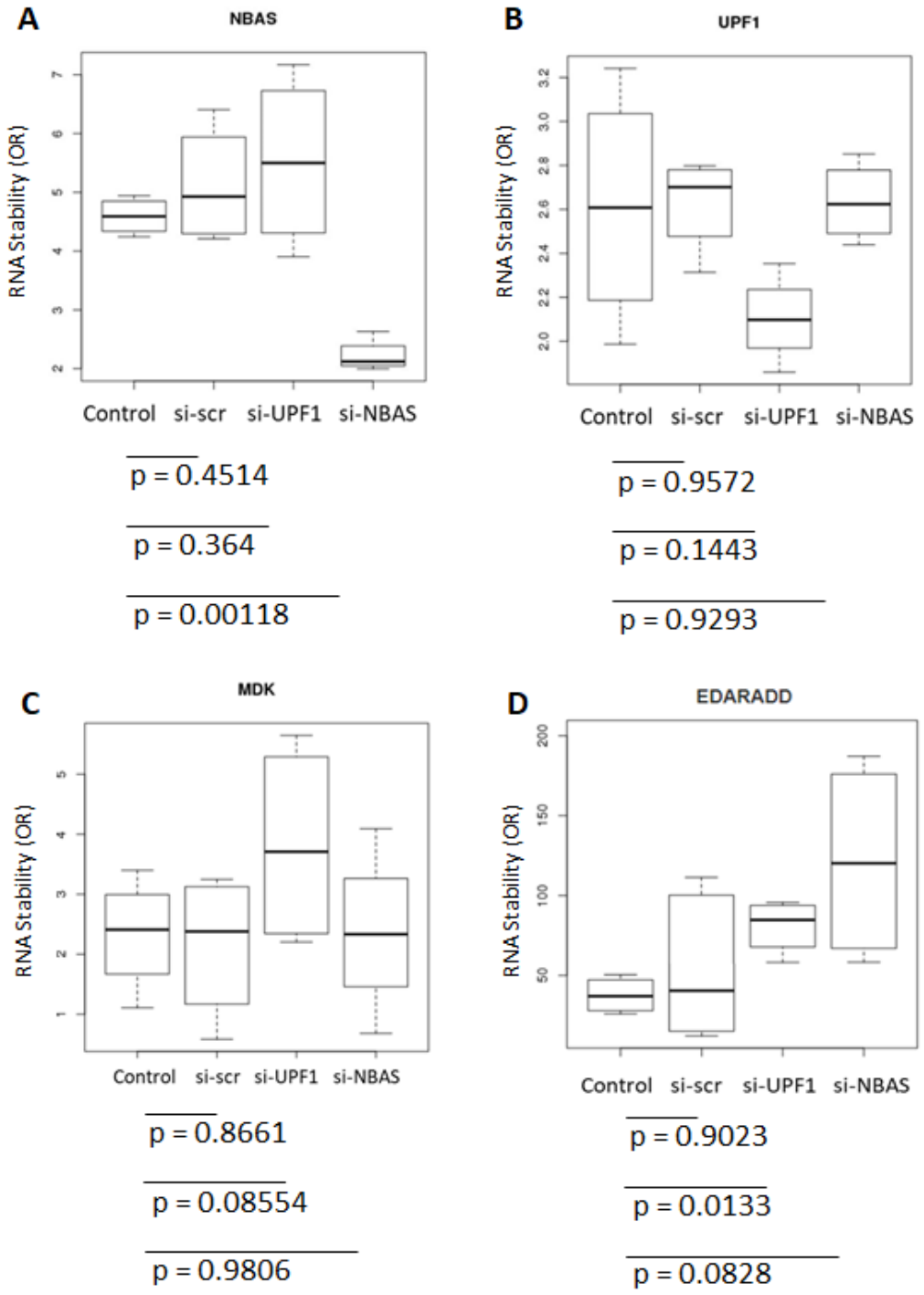


Figure 5.5: ***Differential Stability of Individual Genes***

Figure 5.5: ***Differential Stability of Individual Genes***
OR of individual genes in each condition (n=5) shows variability within conditions and changes in RNA stability between conditions. The decrease in stability of both ***(A)*** *NBAS and* ***(B)*** *UPF1 mRNA can be seen in the appropriate depleted sample.* ***(C)*** *Example of a gene with increased stability when UPF1 depleted and* ***(D)*** *example of gene with increased stability when either UPF1 or NBAS depleted. Results of paired t-tests between each condition and control are indicated below each plot. None of the statistical tests are multiple testing corrected.*

In regards to building multi-factor variables into the DESeq2 model, the DESeq2 vignette states:

"Can I use DESeq2 to analyse paired samples? Yes, you should use a multi-factor design which includes the sample information as a term in the design formula. This will account for differences between the samples while estimating the effect due to the condition. The condition of interest should go at the end of the design formula, e.g. $\sim$ *subject + condition."*

This adaptation is similar in setup to my manual analysis of measuring a ratio of ratios. However, because DESeq2 measures the change between conditions, I decided to run DESeq2 on the total and nascent reads for each gene in si-Scr and si-UPF1 treated conditions (Table 5.2), as measured by STAR mapping and HTSeq counts per gene. In this way, the effect of knockdown on expression can be measured normalised to transcription resulting in differential stability changes being identified.
I first looked at stability changes when UPF1 was knocked down as a positive control. The results of this analysis showed one gene destabilised when UPF1 was depleted, which was UPF1 ($FC = 1.26, p < 0.05$). This reflects the decreased stability of the target mRNA when treating with siRNA. 86 genes were significantly increased in stability (Figure 5.6A). The most significantly stabilised

sample	condition	assay
N1T	total	si-NBAS
N2T	total	si-NBAS
N3T	total	si-NBAS
N4T	total	si-NBAS
N5T	total	si-NBAS
S1T	total	si-Scr
S2T	total	si-Scr
S3T	total	si-Scr
S4T	total	si-Scr
S5T	total	si-Scr
N1N	nascent	si-NBAS
N2N	nascent	si-NBAS
N3N	nascent	si-NBAS
N4N	nascent	si-NBAS
N5N	nascent	si-NBAS
S1N	nascent	si-Scr
S2N	nascent	si-Scr
S3N	nascent	si-Scr
S4N	nascent	si-Scr
S5N	nascent	si-Scr

Table 5.2: ***SampleTable for DESeq2 total and nascent analysis*** *The DESeq2 model requires a SampleTable detailing the variables describing each sample, in this case whether each sample is total or nascent and si-Scr control or si-NBAS treated.*

was CROCCP2 (ENSG00000215908), a pseudogene with six annotated processed transcripts and one transcribed unprocessed transcript. 30 of the 86 genes have an annotated splice variant and 39 are labelled with the keywords "alternative splicing".

Overlap of expression targets and stability targets identified 84 of the 86 transcripts increased in stability were also increased in expression (Figure 5.6B), strongly suggesting they are *bona fide* NMD targets. Both of the stability targets not also found in expression targets, VIPR1 and TXNRD2, were not identified as significantly changed in expression. Looking just at the fold changes, VIPR1 was decreased in total RNA and in nascent RNA but was more decreased in nascent suggesting that the transcripts could be stabilised. Furthermore, 11 of the 28 annotated transcripts of VIPR1 are not thought to be translated as are 11 of the 32 annotated TXNRD2 transcripts, strongly suggesting that these are also NMD targets. Overlap of UPF1 expression and stability targets from a similar study that used BRIC-seq to calculate stability, found 76 RNAs that increased in expression and stability (Figure 5.6B Tani et al., 2012). Comparison of change in RNA stability upon UPF1 depletion from this study with relative half-life upon depletion of UPF1 from Tani et al. (2012), results in a significant but weak correlation ($r = 31, p < 2.2\text{x}10^{-16}$, Pearson's correlation test; Figure 5.6C). However, this accounts for all measured changes as Tani et al. did not include p values. Overlap of the 76 targets from Tani et al. that increased in expression and stability, with the 84 transcripts identified in this study resulted in eight common targets between the two analyses.

Gene ontology (GO) analysis of the 84 targets increased in expression and stability revealed many terms related to the immune system and chemotaxis (Figure 5.6D), suggesting it is mainly cellular response genes that are being destabilised in an UPF1-dependent manner. Comparing these 84 targets with the previously described ER datasets (Section 2.2.15, Section 4.2.4; Fazal et al., 2019; Reid and Nicchitta, 2012; Jan et al., 2014) identified two targets, CYB561 and POFUT2,

that were defined by APEX-seq to be ER translated and two targets, TMEM222 and LEMD2, that were defined by ER proximity-specific ribosome profiling to be ER translated, equating to 6.35% of the UPF1 targets being experimentally defined ER translated.

The clean results seen here with no apparent indirect effects, indicate that using DESeq2 and matched total and nascent RNA sequencing allows stability calculations at a single time point without the use of transcription inhibition.

The same analysis when NBAS was depleted showed one gene to be significantly changed in stability (Figure 5.6E). This gene was NBAS which was destabilised ($FC = -1.85, p < 0.01$). The further increase in negative fold change compared to previous analyses shows that this method further accounts for noise in the data. This result coupled with the results of the positive control UPF1, strongly indicate that NBAS has no effect on RNA stability. Whilst there could be changes that are too subtle to be detected using this analysis, the combination of number of replicates, depth of sequencing and strong statistical framework suggests that another technique is necessary to identify any changes in stability that may be occurring when NBAS is depleted.

It has been shown that transcripts encoding NMD factors are themselves regulated by the NMD pathway, thus forming a feed-back auto-regulatory loop (Huang et al., 2011; Yepiskoposyan et al., 2011; Longman et al., 2013, Section 1.2.2). Therefore, I was interested to see how the RNA of other NMD factors was changed in stability upon depletion of UPF1 and NBAS. Differential stability in either condition compared to si-Scr treated samples showed that several NMD factors were stabilised upon depletion of UPF1 (Figure 5.6F, stars indicate significant ($p < 0.05$), although only the changes of SMG5 and SMG9 reached significance. A significant change in stability was not detected for any of the NMD factors upon depletion of NBAS.

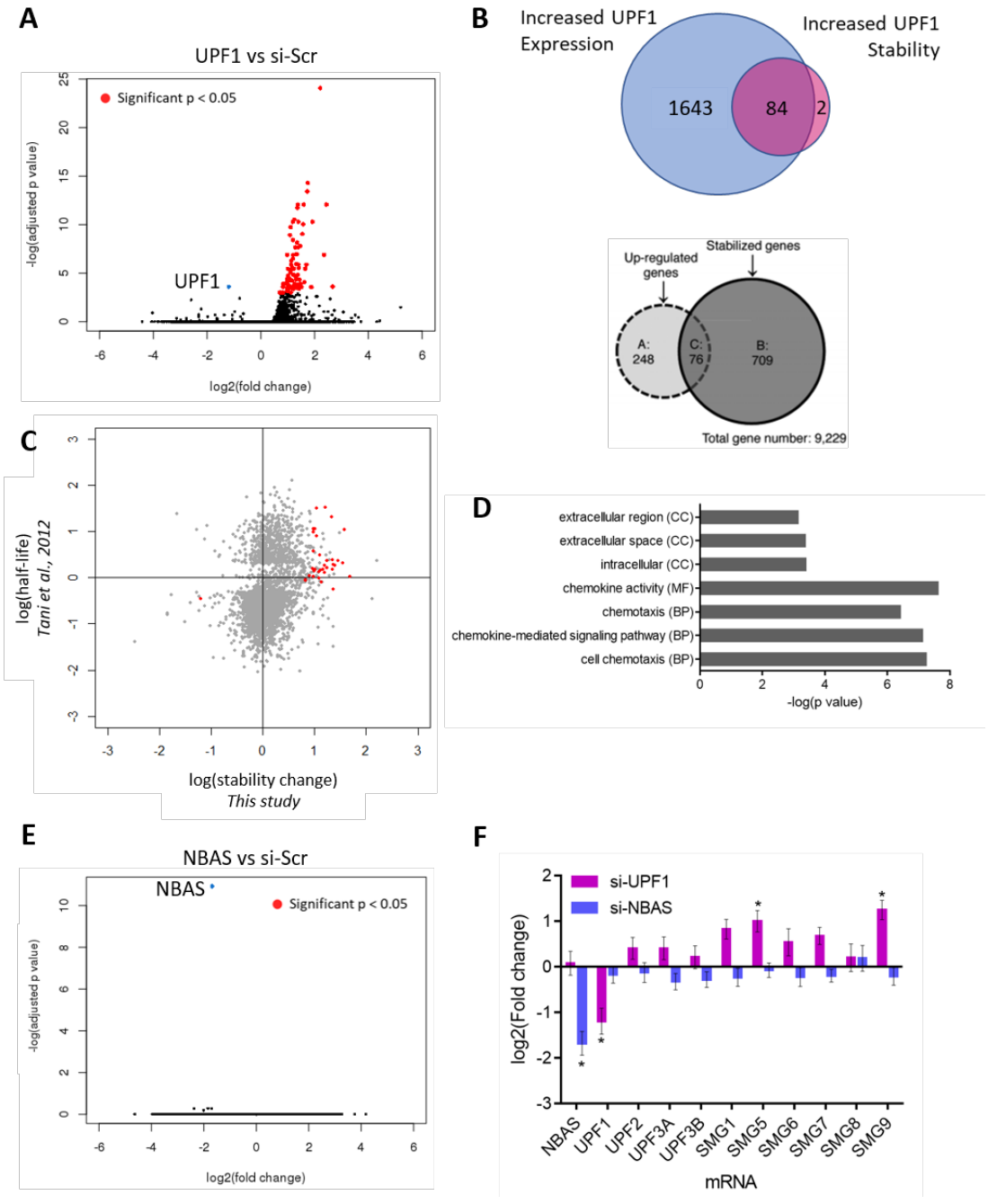


Figure 5.6: ***DESeq2 analysis of total and nascent RNA produces a robust measure of stability***

Figure 5.6: ***DESeq2 analysis of total and nascent RNA produces a robust measure of stability***

(A) *Differential stability analysis using DESeq2 on total and nascent RNA shows genes changed in stability when UPF1 was depleted by siRNA in HeLa cells. Genes significantly changed in stability are marked in red ($p < 0.05$). UPF1 is indicated in blue* ***(B)*** *Overlap of transcripts increased in expression and transcripts increased in stability upon depletion of UPF1. For comparison, a similar overlap of UPF1 expression and stability targets is shown from Tani et al. (2012).* ***(C)*** *Comparison of change in RNA stability upon UPF1 depletion from this study with relative half-life upon depletion of UPF1 from Tani et al. (2012). Genes significantly changed in stability in this study are marked in red ($p < 0.05$).* ***(D)*** *Gene Ontology (GO) analysis of all transcripts increased in stability when UPF1 depleted. Top three terms are shown of biological process (BP), molecular function (MF) and cellular component (CC) GO terms.* ***(E)*** *Differential stability analysis using DESeq2 on total and nascent RNA shows genes changed in stability when NBAS was depleted by siRNA in HeLa cells. Genes significantly changed in stability marked in red ($p < 0.05$). NBAS is indicated in blue.* ***(F)*** *Differential stability of NMD factors upon depletion of NBAS or UPF1. Change in stability shown as log2(fold change) and standard deviation from DESeq2. Stars indicate significant changes ($p < 0.05$).*

5.2.4 Comparing exonic and intronic reads normalises for mRNA processing using only total RNA

Although not as elegant as using both total and nascent RNA, measuring stability of RNA using exonic and intronic reads from total RNA-sequencing as a proxy for mature mRNA and pre-mRNA is often used in genome-wide studies to determine degradation rates (Zeisel et al., 2011; Gray et al., 2014; La Manno et al., 2018; Bergen et al., 2020). In this study, it has the advantage of using only the total RNA libraries which have been validated against other datasets (Figure 4.5 and 4.10).

Using just the read counts for total exonic and total intronic in each sample gives a ratio of the mature mRNA and pre-mRNA levels, which can be used as a proxy for transcript stability. To measure the effect of UPF1 or NBAS depletion on RNA stability, the change of this ratio was calculated between treated and control cells. An OR for each gene was calculated using the following equation:

$$OR = (\textit{treated exonic (mature mRNA)} \,/\, \textit{treated intronic (pre-mRNA)}) \,/\, (\textit{control exonic (mature mRNA)} \,/\, \textit{control intronic (pre-mRNA)})$$

Unlike the previous OR analysis, rather than producing one value per condition, this OR gives a measure of the **change** in stability. Negative control si-Scr, as well as si-UPF1 and si-NBAS, were each defined as "treated" and all three were measured against mock treated "control" samples, making them directly comparable. The OR for si-UPF1 and si-NBAS was then compared to that of si-Scr to identify changes beyond technical noise. An OR of 1 indicates that the likelihood of an event to occur is not more than would be expected by chance. Since we assume transcription is not affected by NMD abrogation (Carter et al., 1996; Wu et al., 2020; Karousis et al., 2020), and each condition is normalised to the same control, changes to the counts of mature mRNA (exonic) reads will drive changes in OR. In this way, an increase in the OR compared to si-Scr indicates an increase in stability of the mRNA. I compared the cumulative stabilities of four subsets of genes, as defined in Chapter 4, in each condition; "UPF1 targets" which

were increased in expression upon UPF1 depletion, "NBAS targets" which were increased in expression upon NBAS depletion, "common targets" which were increased in expression upon UPF1 **and** NBAS depletion, and "UPF1-independent NBAS targets" which were increased in expression upon NBAS depletion **but not** upon UPF1 depletion. Using random sampling with replacement, 95% confidence intervals (CI) were calculated and added to each curve to allow statistical determination of changes to global stability (Figure 5.7).

UPF1 targets were significantly increased in stability when UPF1 was depleted, as seen by the shift to the right of the si-UPF1 line compared to the si-Scr line. This is agreement with these transcripts being genuine NMD targets that are increased in abundance and stability. There was no discernible difference in the stability of UPF1 targets when NBAS was depleted, indicated by the highly overlapping blue and black lines (Figure 5.7A). This could be explained by the fact that only some UPF1 targets are also regulated by NBAS. However, a similar result was seen for NBAS targets, where as a group, the transcripts were significantly increased in stability when UPF1 was depleted but not changed in stability when NBAS was depleted (Figure 5.7B). This indicates that these genes are increased in abundance but do not have altered stability when NBAS is depleted.

As expected, given these results, common targets which increased in abundance when either factor was depleted were increased in stability when UPF1 was depleted but were largely unchanged when NBAS was depleted with the small shift compared to si-Scr within the 95% CI (Figure 5.7C). Considering the UPF1-independent NBAS targets, the 95% CI of all three lines overlap indicating no significant changes in stability take place for these transcripts upon depletion of either factor (Figure 5.7D), suggesting that UPF1 depletion does not increase stability of all transcripts globally.

The cumulative distribution plots compare the global changes in stability upon depletion of each factor, so I next wanted to compare the change in stability for individual mRNAs. Plotting the log2 change in OR of UPF1 targets and NBAS targets when NBAS was depleted against the change in OR when UPF1 was depleted, reveals that some mRNAs do appear to change in stability when

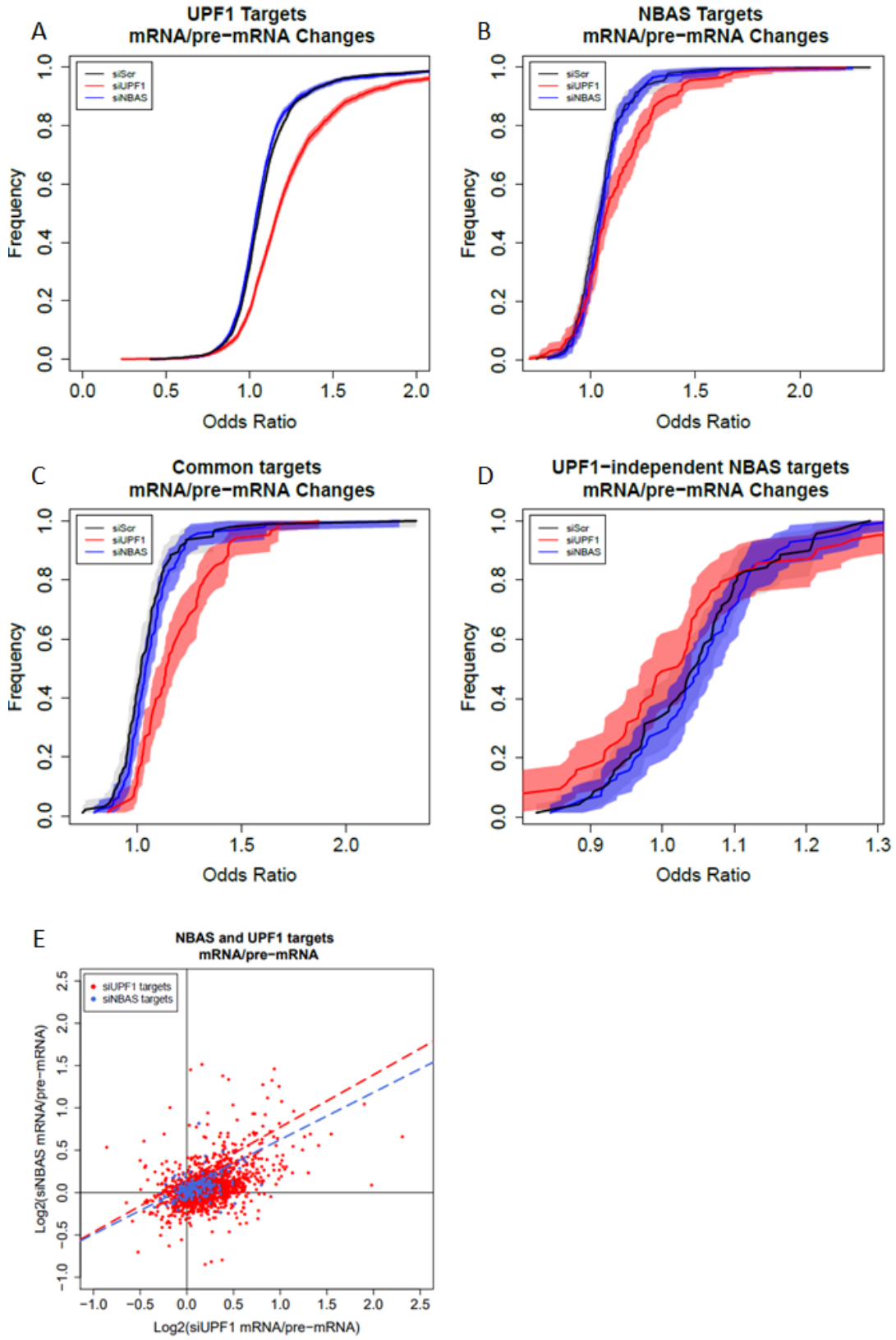


Figure 5.7: ***Comparing exonic and intronic reads allows RNA stability measurement using only total RNA***

Figure 5.7: ***Comparing exonic and intronic reads allows RNA stability measurement using only total RNA***

Using read counts for total exonic and total intronic in each sample gives a ratio of the mature mRNA and pre-mRNA levels, which can be used as a proxy for transcript stability. By calculating the change in this ratio when cells are treated with siRNA compared to mock treated cells, the effect of each treatment on RNA stability can be measured. Cumulative stabilities of subsets of genes, as defined in Chapter 4, in each condition; ***(A)*** *"UPF1 targets" which were increased in expression upon UPF1 depletion,* ***(B)*** *"NBAS targets" which were increased in expression upon NBAS depletion,* ***(C)*** *"common targets" which were increased in expression upon UPF1 and NBAS depletion, and* ***(D)*** *"UPF1-independent NBAS targets" which were increased in expression upon NBAS depletion but not upon UPF1 depletion. Using random sampling with replacement, 95% confidence intervals (CI) were calculated and added to each curve to allow statistical determination of changes to global stability.* ***(E)*** *The log2 change in stability of UPF1 targets (red) and NBAS targets (blue) when NBAS was depleted against the change in OR when UPF1 was depleted. The dashed line shows correlation of each set of targets.*

NBAS is depleted (Figure 5.7E), although this plot does not account for statistical significance. As shown by the dashed lines, the correlation of change in OR upon NBAS depletion or UPF1 depletion for both NBAS (blue line) and UPF1 (red line) targets is skewed towards UPF1-mediated changes in stability.

As with the previous analysis using total and nascent RNA, this methodology does not account for the variance between replicates and samples. Therefore, I next analysed total exonic and total intronic samples using the DESeq2 framework (Table 5.3). This resulted in 284 genes being identified as significantly increased in stability when UPF1 was depleted, and 109 genes significantly decreased in stability (Figure 5.8A), showing an enrichment for increased stability ($OR = 2.6$ compared to null hypothesis which would be equal numbers increased and decreased, $p < 0.0001$; fishers exact test). Similar to the total and nascent DESeq2 RNA analysis (Section 5.2.3), the enrichment of increased stability is in line with UPF1's established role as an NMD factor regulating mRNA stability. UPF1 was decreased in stability ($FC = -0.498, p = 0.0024$) as expected due to the siRNA treatment, although the most significantly decreased mRNA was large ribosomal subunit protein L37A (RPL37A; ENSG00000197756) ($FC = -1.39, p < 0.0001$), suggesting a global decrease in translation. The gene most significantly increased in stability was DNAJC25 (ENSG00000059769) ($FC = 1.87, p < 0.0001$). DNAJC25 is a member of the heat shock protein family (Hsp40) which interestingly, is located at the ER and failure of hsp40-mediated protein refolding machinery is known to cause aggregation of poly-glutamine proteins commonly found in neurodegenerative disorders (Muchowski et al., 2000).

Conversely, the same analysis looking at changes when NBAS was depleted showed only two genes changed in stability (Figure 5.8B). One of these was NBAS which was decreased in stability ($FC = -0.9, p < 0.001$) as expected since it was specifically targeted with siRNA. The other gene was WASH6P, a pseudogene which was increased in stability ($FC = 4.1, p < 0.05$). This gene has 15 splice variants, of which 14 are annotated as processed transcripts and the other one is annotated as transcribed unprocessed pseudogene.

Sample	condition	assay
N1T	exonic	siNBAS
N2T	exonic	siNBAS
N3T	exonic	siNBAS
N4T	exonic	siNBAS
N5T	exonic	siNBAS
D1T	exonic	mock
D2T	exonic	mock
D3T	exonic	mock
D4T	exonic	mock
D5T	exonic	mock
N1T	intronic	siNBAS
N2T	intronic	siNBAS
N3T	intronic	siNBAS
N4T	intronic	siNBAS
N5T	intronic	siNBAS
D1T	intronic	mock
D2T	intronic	mock
D3T	intronic	mock
D4T	intronic	mock
D5T	intronic	mock

Table 5.3: ***SampleTable for DESeq2 total exonic and intronic analysis*** *The DESeq2 model requires a SampleTable detailing the variables describing each sample, in this case whether each sample is total exonic or total intronic and si-Scr control or si-NBAS treated.*

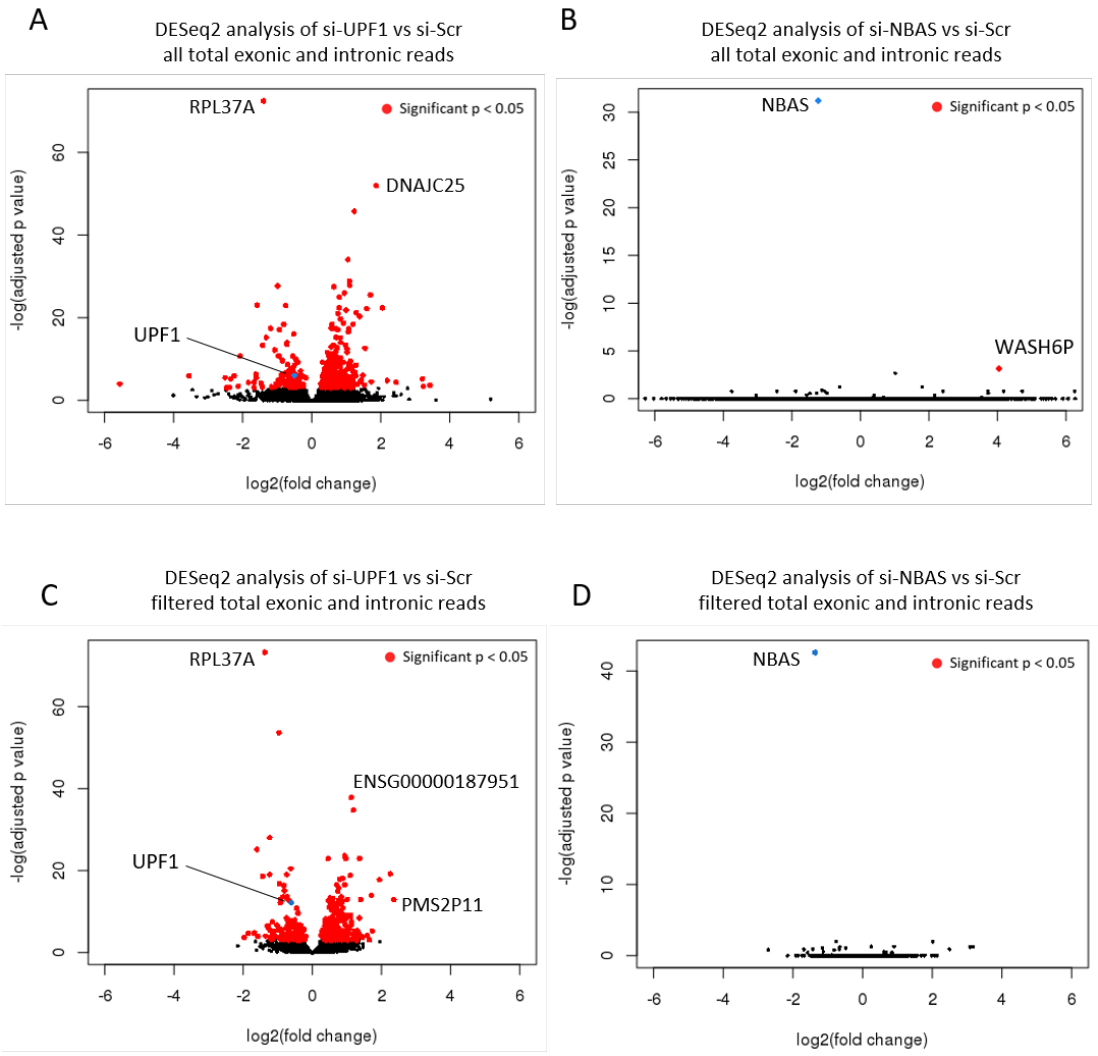


Figure 5.8: ***DESeq2 analysis of total exonic and intronic RNA identifies hundreds in changes in stability upon UPF1 depletion but not when NBAS is depleted***

DESeq2 analysis of RNA stability using all total exonic and intronic reads of ***(A)*** *si-UPF1 vs si-Scr and* ***(B)*** *si-NBAS vs si-Scr. DESeq2 analysis of RNA stability using total exonic and intronic reads of only universally expressed mRNAs comparing* ***(C)*** *si-UPF1 vs si-Scr and* ***(D)*** *si-NBAS vs si-Scr. In all plots, mRNA significantly changed in stability are marked in red ($p < 0.05$) and the siRNA targeted mRNA is indicated in blue.*

This analysis included all genes that had at least one read in at least one sample. Itwas therefore possible that the large number of genes with low expression might be masking subtle changes leading to this lack of significant results in si-NBAS. Therefore, I next filtered the input for only genes that were detected (at least one read) in all replicates of all conditions and repeated the DESeq2 analysis.

This analysis found 425 genes changed in stability when UPF1 was depleted compared to si-Scr treated cells (Figure 5.8C). 282 transcripts showed an increase in stability and 143 showed a reduction in stability, an enrichment for increased stability ($OR = 1.98$ compared to null hypothesis which would be equal numbers increased and decreased, $p < 0.0001$; fishers exact test). UPF1 mRNA was identified as decreased in stability ($FC = -0.61, p < 0.0001$) although the most significantly decreased mRNA was again RPL37A ($FC = -1.37, p < 0.0001$). The mRNA most significantly increased in stability was the pseudogene ENSG00000187951 ($FC = 1.12, p < 0.0001$) which is annotated as a lncRNA by ensembl but as microRNA mir-1255 by Rfam, either way suggesting it is not translated. The RNA with the largest increase in stability was the pseudogene PMS2P11 (ENSG00000241350) ($FC = 2.36, p < 0.0001$).

When si-NBAS was analysed compared to si-Scr, only one gene was significantly changed in stability when NBAS was depleted (Figure 5.8D). This was NBAS which was decreased in stability ($FC = -1.4, p < 0.001$). The difference in fold change using this filtered dataset compared to the previous analysis demonstrates the effect of only testing, and correcting significance thresholds for, relevant genes rather than all genes. However, instead of resulting in subtler effects on stability by NBAS being revealed, this analysis appears to confirm that NBAS does not significantly affect stability of any transcript measured here.

5.2.5 Comparing RNA stability measurements by nascent and total RNA or total exonic and intronic RNA methods indicates a good correlation between the methodologies

The two methodologies used in this chapter to analyse RNA stability gave broadly similar results. Both identified NBAS and UPF1 as destabilised in the relevant samples and both show UPF1 depletion affected the stability of many transcripts whereas both showed NBAS depletion appears to have limited effect on RNA stability. To compare the two methods more specifically, I plotted the estimated change in stability for individual genes upon UPF1 depletion produced by each method (Figure 5.9). Both methods identified transcripts as changed in stability that were not identified by the other method. However, the correlation of estimated change in stability of those genes identified as significantly changed by both methodologies (black dots) was high ($r = 0.81, p = 1.32\text{x}10^{-8}$; Pearson's correlation). Both UPF1 and NBAS were identified as significantly destabilised in the appropriate samples, and the results of UPF1 knockdown were in line with previous studies and understanding of UPF1's function on transcript stability. The clean results seen from the analysis of total and nascent RNA using the DESeq2 framework, with no apparent indirect effects, indicate that using DESeq2 and matched total and nascent RNA sequencing allows stability calculations at a single time point without the use of transcription inhibition.

5.2.6 Comparing changes in RNA expression and RNA stability reveals many abundance changes are dependent on transcription

To further investigate the effect of transcription and stability on RNA expression, DESeq2 was used to analyse differential transcription using only nascent RNA samples, mapped with STAR and counted per gene by HTSeq. Upon depletion of UPF1, 18 mRNAs were differentially transcribed (Figure 5.10A), two were de-

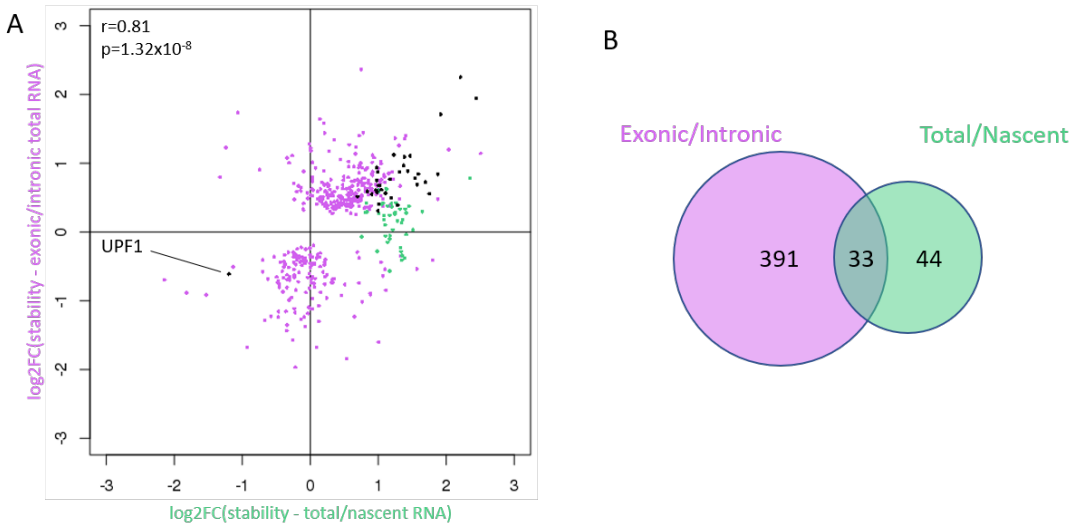


Figure 5.9: ***Comparison of RNA stability measurements by nascent/total RNA and exonic/intronic RNA methods***

(A)) *Estimated change in stability for individual genes upon UPF1 depletion produced by DESeq2 analysis of total and nascent RNA (x-axis, green) and by DESeq2 analysis of total exonic and total intronic RNA (y-axis, purple). Black dots indicate genes identified as significantly changed by both methodologies, correlation of these points was r=0.81 ($p = 1.32x10^{-8}$; Pearson's correlation).* ***(B)*** *Overlap of mRNAs identified as significantly increased in stability upon depletion of UPF1 by both methodologies.*

creased in transcription and 16 increased. The two genes decreased in transcription were the complement factor C3 ($FC = -2.58, p < 0.0001$) and a lncRNA located in an intron of C3 (ENSG00000276980) ($FC = -4.80, p = 0.0298$). The gene most significantly increased in transcription upon UPF1 depletion was the lncRNA ENSG00000272068 ($FC = 2.30, p < 0.0001$). The UPF1 gene was not changed in transcription, further validating effective targetting of UPF1 mRNA by siRNA treatment. Upon depletion of NBAS, no genes were identified as changed in transcription, including NBAS (Figure 5.10B).

Since the differential expression analysis carried out in Chapter 4 used DESeq2 on estimated counts per gene produced by kallisto pseudoaligning, I wanted to repeat the differential expression analysis using DESeq2 on STAR aligned reads and HTSeq counts per gene, so that the results from total RNA and nascent RNA would be as comparable as possible.

This analysis identified 1968 transcripts as differentially expressed when UPF1 was depleted (Figure 5.10C). 1727 genes were identified as increased in abundance, whereas 241 genes were identified as decreased in abundance, representing an enrichment for increased abundance ($OR = 7.16$ compared to null hypothesis which would be equal numbers increased and decreased, $p < 0.0001$; fishers exact test). UPF1 mRNA was the most significantly decreased in abundance ($FC = -1.54, p < 0.0001$). The gene most significantly increased in abundance upon UPF1 depletion was the pseudogene CROCCP2 (ENSG00000215908) ($FC = 2.43, p < 0.0001$), whereas the gene with the biggest increase in abundance (and second most significant increase), was the lncRNA ENSG00000272068 ($FC = 3.70, p < 0.0001$) which was the RNA most significantly changed in transcription, suggesting this change in abundance is driven by transcription.

Comparing NBAS depleted cells to si-Scr using this analysis identified 17 genes as changed in abundance (Figure 5.10D). 7 genes were increased in expression whereas 10 were decreased in expression, a non-significant enrichment for a decrease in abundance ($OR = 1.41$ compared to null hypothesis which would be equal numbers increased and decreased, $p = 0.732$; fishers exact test). NBAS mRNA was the most significantly decreased in abundance ($FC = -1.70, p <$

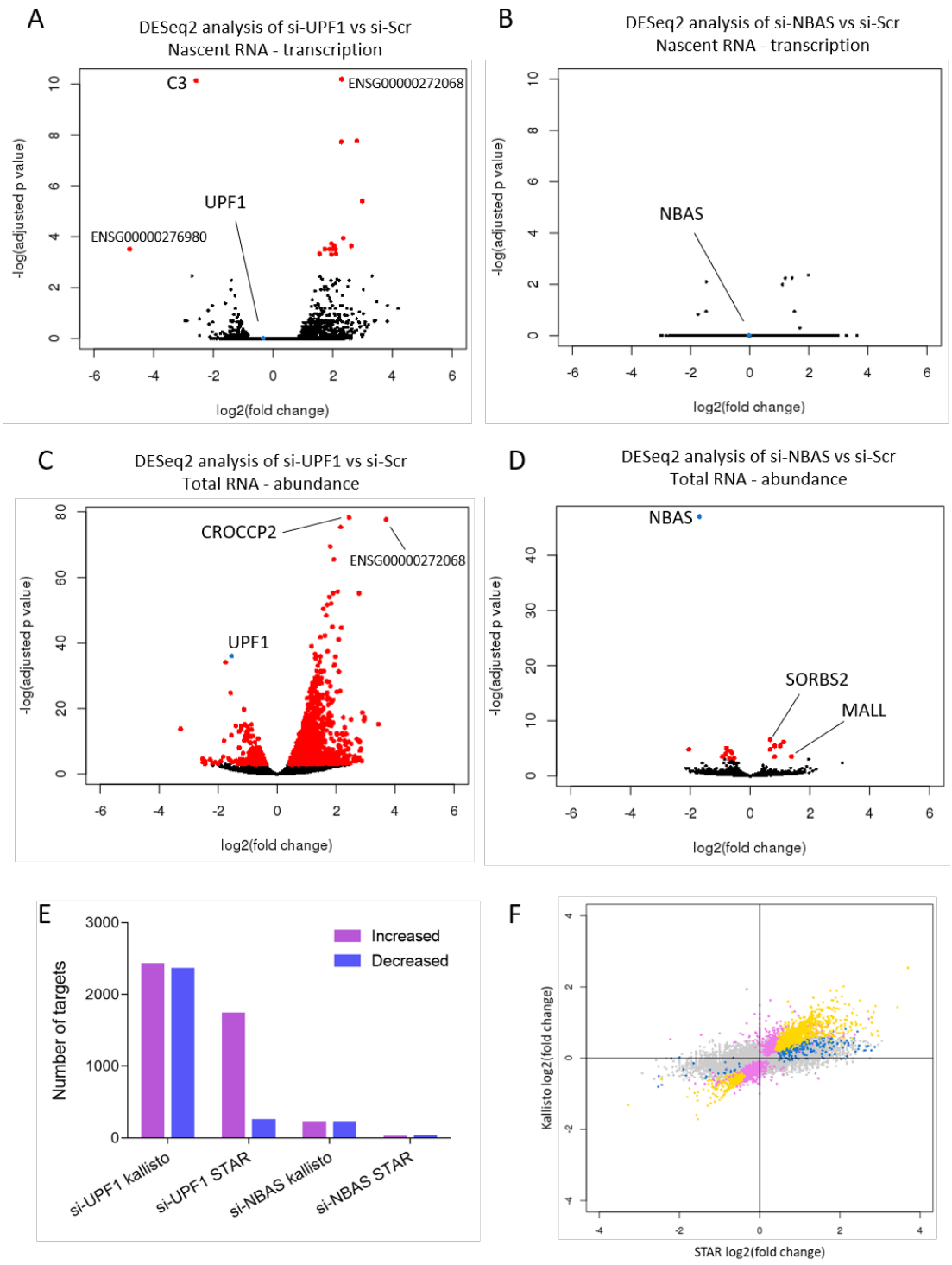


Figure 5.10: ***DESeq2 analysis of differential abundance and differential transcription from STAR alignment***

Figure 5.10: ***DESeq2 analysis of differential abundance and differential transcription from STAR alignment***

DESeq2 analysis of change in transcription using nascent reads of ***(A)*** *si-UPF1 vs si-Scr and* ***(B)*** *si-NBAS vs si-Scr. DESeq2 analysis of differential RNA abundance using total reads, aligned using STAR and counted using HTSeq, of* ***(C)*** *si-UPF1 vs si-Scr and* ***(D)*** *si-NBAS vs si-Scr. In all plots, mRNA significantly changed in stability are marked in red ($p < 0.05$) and the siRNA targeted mRNA is indicated in blue. (E) Number of targets identified as significantly ($p < 0.05$) increased or decreased in expression by kallisto-DESeq2 pipeline and STAR-HTSeq-DESeq2 pipeline upon depletion of UPF1 or NBAS. (F) Comparison of fold change calculated by kallisto-DESeq2 pipeline and STAR-HTSeq-DESeq2 pipeline upon depletion of UPF1. Colours indicate all genes analysed (grey), those called as significant by STAR (blue), those called as significant by kallisto (pink) and those called as significant by both (yellow).*

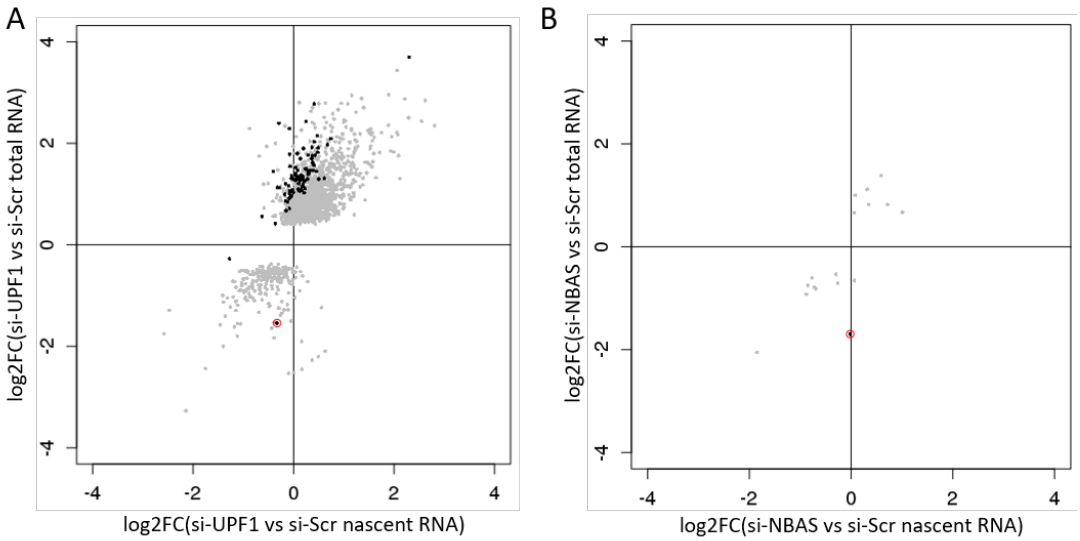
0.0001). The gene most significantly increased in abundance upon NBAS depletion was the cytoskeleton structural protein SORBS2 (ENSG00000154556) ($FC = 0.67, p = 0.0014$), whereas the gene with the largest increase in abundance was the membrane trafficking protein MALL (ENSG00000144063) ($FC = 1.39, p = 0.0297$).

Surprisingly, the results produced by this STAR-HTSeq-DESeq2 pipeline were considerably different to those resulting from the kallisto-DESeq2 pipeline (Figure 5.10E). As well as identifying fewer transcripts as differentially changed, the STAR-HTSeq-DESeq2 pipeline also showed a marked difference in the number of targets increased and decreased in expression, whereas the kallisto-DESeq2 pipeline resulted in a similar number of genes identified as increased and decreased in expression. To investigate this discrepancy, I compared the log2(fold change) calculated by the kallisto-DESeq2 pipeline and the STAR-HTSeq-DESeq2 pipeline in si-UPF1 treated cells compared to si-Scr treated cells (Figure 5.10F). Whilst the correlation between the fold change calculated by the two analyses of all genes was significant ($r = 0.59, p < 2.2\text{x}10^{-16}$, Pearson's correlation test), the STAR-HTSeq-DESeq2 calculated fold changes as larger in general. Identifying

those called as significant by STAR (blue), by kallisto (pink) and by both (yellow) revealed further differences between the methods. The large number of pink dots at near zero log2(fold change)m indicate that the kallisto-DESeq2 pipeline leads to much smaller fold changes being identified as significant. This effect was more pronounced for transcripts decreased in expression which accounts for the change in number of targets seen in (Figure 5.10E).

The transcription and abundance changes from STAR-HTSeq-DESeq2 analysis of total or nascent RNA were then compared to visualise how the kinetic processes of RNAs changed upon depletion of UPF1 (Figure 5.11A) and NBAS (Figure 5.11B). For UPF1 depletion, as expected, all genes significantly changed in stability (black dots) clustered around the y-axis indicating their change in expression was not correlated with their change in transcription, whereas genes only changed in expression (grey dots) splayed to each side of the y-axis showing their change in abundance was driven by change in transcription, even if this change was not identified as significant by the DESeq2 analysis run on nascent RNA only (Figure 5.10A). UPF1 is indicated by a red circle. For NBAS, since so few genes were identified as changed in abundance (Figure 5.10D), the plot is quite empty even with a generous significance cut off of $p < 0.05$. The only black dot is NBAS, which was the only gene that was changed in stability (Figure 5.6E), and is indicated by a red circle. It is on the x=0 line indicating no change in transcription which is as expected since NBAS mRNA was targetted by siRNA which does not affect transcription. The majority of the grey dots are splayed to each side of the y-axis showing their change in abundance was driven by change in transcription, even if this change was not identified as significant by the DESeq2 analysis run on nascent RNA only (Figure 5.10B).

Whilst the UPF1 stability changes measures are in line with previous studies (Tani et al., 2012), it was surprising to observe that all NBAS-dependent changes in stability appeared to be due to transcriptional changes (Figure 5.11C) since this does not agree with NBAS' previously studied role as an NMD factor (Longman et al., 2007; Anastasaki et al., 2011; Longman et al., 2013, 2020).



C

	Expression up	Expression down	Stability up	Stability down	Expression & stability up	Expression & stability down
UPF1	1727	241	86	1	84	1
NBAS	7	10	0	1	0	1

Figure 5.11: ***Comparing changes in RNA expression and RNA stability reveals many abundance changes are dependent on transcription***

(A) *Comparison of abundance change in total RNA and change in nascent RNA of si-UPF1 compared to si-Scr control cells as measured by DESeq2 analysis. Genes significantly changed in total RNA abundance are marked in grey and those significantly changed in RNA stability (using total and nascent RNA) are marked in black. UPF1 is marked with a red circle.* ***(B)*** *The equivalent plot for si-NBAS depletion.* ***(C)*** *Summary table of number of genes changed in expression and stability when either factor was depleted and the overlap of these.*

5.3 Discussion

5.3.1 It is possible to measure transcript stability at a single time point without transcription inhibition

In this chapter I have shown that it is possible to achieve significant and specific measurements of changing transcript stability by RNA-sequencing a single time point without inhibiting transcription (Figure 5.8A).

This has important consequences for the future of measuring genome-wide transcript stability. The current standard stability experiment relies on taking measurements of transcript abundance at multiple time points. This requires complex experimental set up and also many individual samples, pushing up the cost of sequencing which can still be prohibitive to some researchers. Removing the need for multiple time points means that money can be spent on deeper sequencing and/or more replicates, both of which increase the meaningful results that can be gleaned from these experiments. From a biological standpoint, it is also important to be able to measure transcript stability without the need for transcription inhibition because transcription and decay are related processes and transcription inhibition can have negative effects on the cell (Section 5.1.1).

The results in this chapter contrast with those of Chapter 4 as RNA abundance changes resulting from the STAR-HTSeq-DESeq2 pipeline were markedly different from those resulting from the kallisto-DESeq2 pipeline described in Chapter 4. As shown in (Figure 5.10F), this seems to be due to the kallisto-DESeq2 pipeline leading to smaller fold changes being identified as significant. This has recently been noted by others who found only approximately half of the differentially expressed transcripts detected using kallisto were also identified by STAR or LAST (Shrestha and Frith, 2013) and in general, disagreement in low-expression

genes was responsible for most of the discordance between methods (Kuo et al., 2020). This was suggested to be due to kallisto tending to overestimate some low-expression genes compared to other methods (Kuo et al., 2020; Everaert et al., 2017).

The methodology used here is not however perfectly optimised. The uridine analogue 4sU has been shown to cause harmful effects, whereas the alternative analogue BrU does not appear to have these drawbacks (Friedel and Dölken, 2009). It has been shown that the labelled RNA fraction can become contaminated with unlabelled (pre-existing) RNA molecules (Furlan et al., 2020). A model based on a constant rate of 4sU incorporation into nascent transcripts did not fit data measuring the amount of labelled RNA that could be recovered with pulses of 4sU lasting from 10min to 2h. Whereas a model assuming a constant contamination rate, not dependent on the 4sU pulse length fit the data better (Furlan et al., 2020). The rate of contamination has been reported as 30% (Baptista and Dölken, 2018; Furlan et al., 2020), although the rate is likely to depend on the cell type and on the specific protocol used. There was a much lower level of reads mapping to rRNA in the nascent libraries compared to the high level of rRNA reads in the matched total RNA libraries. This could be evidence that 4sU is not incorporated as efficiently by RNA polymerase I specifically. However, previous studies have seen up to 30% of nascent RNA mapping to rRNA (Rabani et al., 2011), although in that study the 4sU labelling time was longer.

Due to the need for intronic reads for this method, poly-A enrichment should not be carried out, since this would deplete the samples of pre-mRNA. This means deeper sequencing will be necessary to achieve the same depth of reads over exonic regions. In this experiment, rRNA depletion was not carried out mainly because the amount of RNA starting material in nascent RNA samples was low and total RNA samples were treated the same as nascent RNA samples. However, in hindsight, since rRNA reads accounted for a much lower percentage of reads in nascent samples, it could be argued that rRNA depletion of only total

samples would be appropriate since there was an excess of RNA starting material for these samples and it would mean many more usable reads from total samples. Indeed, if I were to repeat this experiment, I would deplete rRNA from the total samples, but not the nascent samples, before preparation of the cDNA libraries. Finally, this experiment was unusual in that both mock treated samples and si-Scr treated samples were sequenced as negative controls. There were a small number of significant differences between these two conditions. This is to be expected as siRNA treatment induces a level of stress to the cell even when not targeting a specific gene. Although normalising si-UPF1 and si-NBAS changes to si-Scr would show changes beyond those caused by non-specific siRNA, the ability to normalise all siRNA treatments to mock treated cells allows a measurement of the magnitude of change caused by the specific knockdown of the gene of interest.

5.3.2 Only a subset of transcripts changed in expression when UPF1 is depleted also change in stability

It has previously been suggested that UPF1 regulates up to 10% of the human transcriptome (Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011; Longman et al., 2013). However, these studies used increase in RNA abundance as an indicator of an NMD target. The results presented in this chapter suggest a large proportion of expression changes seen when UPF1 is depleted are NMD-independent, caused by transcriptional changes, rather than bona fide NMD targets characterised by an increase in transcript stability. Tani et al. (2012), who also carried out differential expression and stability analyses upon depletion of UPF1, suggest that many NMD targets are stabilised yet not unregulated, however the results presented here suggest the vast majority of transcripts increased in stability also increase in expression. There are many possible explanations for this result. Cellular pathways are highly interconnected and UPF1 destabilising a gene via the NMD pathways may in turn downregulate transcription of downstream targets of these targets. Furthermore, UPF1 has many differ-

ent roles within the cell which could account for its non-stability effects on gene expression (Section 4.3.1). To further characterise the targets identified here, the transcripts themselves could be analysed, for example is there an enrichment in GC content, RNA length or secondary structures.

It should also be noted that the stability analysis carried out here merged transcripts to give one measurement per gene. Since alternative splicing can lead to transcripts being NMD targets, this methodology could mask more subtle transcript-specific change in stability. Furthermore, transcripts with a retained intron are known to trigger NMD, and these are not accounted for in this analysis. A retained intron would lead to reads over that intronic region being wrongly assigned to the pre-mRNA when they actually originate from the mature RNA. However, intron retention is a very rare event and has only ever been seen for a single intron within a transcript. Therefore, since I merged read counts across all introns for each gene, one intron being retained is unlikely to have a strong affect. Future analysis of this dataset could identify any retained introns by locating introns that have an increased read count compared to the other intronic regions from the same gene.

The results of this study, build on previous literature suggesting that differential stability is a far more accurate method to identify NMD targets than differential RNA expression (Tani et al., 2012; Furlan et al., 2020). However there are differences in approach. Tani et al., utilised BRIC-seq to measure RNA stability (discussed in Section 5.1.1) which is similar to the method used in this study in that it does not block transcription, however multiple time points are used. Making use of these multiple time points could explain why that study identified many more stabilised genes than this study (Figure 5.6B). However, Tani et al. also used only two replicates for their BRIC-seq experiment, whereas this study utilised five replicates which is likely to result in a more reproducibly affects as greater variability is accounted for. Despite the exact genes identified not matching between the two studies (Figure 5.6C), both clearly show that an increase in RNA abundance measured by total RNA-sequencing does not represent

an increase in RNA stability and therefore should not be used as a measure of NMD activity. It would be very interesting to apply this methodology or indeed BRIC-seq upon knockdown of other NMD factors including UPF2, UPF3B and DHX34.

5.3.3 Effect of NBAS on RNA stability

Since previous work has suggested NBAS is an NMD factor in human cells (Longman et al., 2007; Anastasaki et al., 2011; Longman et al., 2013, 2020), my expectation was that depletion of NBAS would lead to increased stability of a subset of the genes increased in stability when UPF1 was depleted. Here I have used four methodologies to determine change in RNA stability; manually using total exonic, total intronic, nascent exonic and nascent intronic reads, manually using just total exonic and intronic reads, using the DESeq2 statistical framework on total exonic and intronic reads, and the cleanest and most robust method, using DESeq2 to analyse total and nascent RNA reads. Whilst all four approaches identified many mRNAs increased in stability upon depletion of UPF1, there was no evidence that NBAS affects RNA stability. This suggests that NBAS is not directly involved in the NMD pathway. Perhaps NBAS does aid UPF1 to perform NMD at the ER but it is neither necessary nor sufficient for ER-NMD.

Depletion of NBAS has previously been shown to delay the decay of UPF1 mRNA (Longman et al., 2013), suggesting it would normally be involved in degradation of this transcript. However, this experiment used actinomycin D to block transcription which, as previously discussed (Section 5.1.1), means that degradation kinetics are also altered. Actinomycin D is also toxic and it has been shown that HeLa cells are extremely sensitive to rapid actinomycin D-induced cell death, even at concentrations as low as 1μg/ml (Sawicki and Godman, 1971). Furthermore, as seen in Figure 4.2C, a non significant but noticeable increase in NBAS mRNA was identified by qPCR when UPF1 was depleted but there was no change to UPF1 mRNA levels when NBAS was depleted.

There is a possibility that the methodologies used in this work are not sensitive enough to identify subtle changes in stability due to NBAS depletion. However, this seems unlikely due to both the subtle changes identified as significant when UPF1 was depleted (Figure 5.8A), and also that no changes to stability were identified when NBAS was depleted even non-significantly (Figure 5.8E).
Since none of the changes in expression when NBAS is depleted, identified in Chapter 4, were also identified as changes in stability (Figure 5.11C), these changes are likely due to altered transcription, or perhaps RNA processing (Figure 5.1A) as downstream effects of NBAS depletion.
While these findings suggest a non-NMD function for NBAS, since its depletion does not affect RNA stability, they do not completely rule out the possibility of NBAS being involved in the ER-NMD mechanism. Perhaps the knockdown of NBAS was sufficient to cause effects at the ER but not strong enough to sufficiently impair its NMD functions. It could also be that NBAS has a redundant role in the ER-NMD mechanism so that upon depletion, its role is fulfilled by another protein.

5.3.4 Future work

Since this is the first study to make use of single time-point metabolic labelling without transcription inhibition to measure RNA stability genome-wide, it would be worth validating the results using a second method that has been previously utilised. One such experimental technique is BRIC-seq (Imamachi et al., 2014, discussed in Section 5.1.1) which is similar to the one used in this study in that it does not block transcription. Mathematical models, with various assumptions, have also been developed specifically to estimate RNA stability such as cDTA (Sun et al., 2012), DRiLL (Rabani et al., 2014), INSPEcT (de Pretis et al., 2015) and pulseR (Uvarovskii and Dieterich, 2017), which utilise the quantification of both nascent and total RNA species. Furthermore, there are also other approaches that do not require the quantification of nascent RNA to estimate

transcript stability (Zeisel et al., 2011; Gray et al., 2014; La Manno et al., 2018; Furlan et al., 2020). These techniques focus on calculating the kinetic rates associated with RNA lifetime from only the total RNA levels to measure RNA stability (Figure 5.1A).

Since only one other study (that I am aware of) has measured genome-wide change in RNA stability upon depletion of UPF1 (without using transcription inhibitors), it is currently difficult to extend the limited comparisons made in this thesis. Nevertheless, it would be interesting to take into account the duration of the depletion of UPF1 before cells were harvested and transcripts analysed. Furthermore, a meta analysis bringing together any genome-wide measurements of RNA stability could provide insight as to the reproducibility of the various techniques being utilised and any biases between these. Another extension to this analysis could be to compare the targets with UPF1-mediated changes in stability, with transcripts found to be bound by UPF1 using CLIP experiments (Kurosaki et al., 2018; Zünd et al., 2013). Since UPF1 binds many transcripts transiently, and it's phosphorylation is thought to stabilise these interactions, CLIP specifically of P-UPF1 would be likely give a more accurate description of NMD targets (Kurosaki et al., 2018).

As detailed in the introduction to this chapter (Section 5.1.1), there is an intricate link between transcription and RNA decay and for this reason it would not be prudent to carry out stability experiments that make use of transcription inhibition as it is known these do not provide an accurate picture of in vivo RNA processing.

5.3.5 Summary

In this chapter I present the first method allowing genome wide measurement of stability of RNA at a single time point without transcription inhibition. This has important consequences for the future of measuring genome-wide transcript stability. I also build on previous literature suggesting that differential stability is a far more accurate method to identify NMD targets than differential RNA expression and suggest that a large proportion of expression changes seen when UPF1

is depleted are NMD-independent changes caused by transcriptional changes. Finally, I find no evidence that NBAS regulates RNA stability, which whilst not disproving its function as an NMD factor does call into question it's direct role in RNA quality control.

Chapter 6

Matched genome, total RNA and nascent RNA sequencing allows identification of NMD triggering variants genome-wide

6.1 Introduction

6.1.1 NMD target characteristics

Even though differential expression studies have been extensively used to identify targets of NMD factors (Mendell et al., 2004; Wittmann et al., 2006; Yepiskoposyan et al., 2011; Longman et al., 2013), there are always indirect effects caused by downstream factors, such as an NMD-targeted TF that then affects expression of many other transcripts, and these are challenging to differentiate from *bona fide* direct NMD targets. Furthermore, although upon NMD-factor depletion an increase in expression would be expected, there is often many decreases in steady state transcript levels upon NMD-depletion. The number of transcripts that decrease in expression are usually roughly equal in number to those that increase but are often disregarded or not even mentioned (Lykke-Andersen et al., 2014).
Studies assessing stability are an improvement on differential expression analyses because NMD degrades mRNA but doesn't directly affect transcription, so measuring decay of transcript is a more direct measure of NMD activity. However, this is still inferring NMD activity by change of stability and does not explain **why** a transcript is being targeted by NMD.
Whole-genome sequencing of individual genomes allow detection of variants that result in a premature termination codon (PTC) which are likely to make the transcript an NMD target. In healthy individuals, these transcripts would be rare since they would be degraded by the NMD machinery, however, under some disease circumstances, NMD activity is altered. Using paired tumour:normal exome data and tumour transcriptomics from the cancer genome atlas (TCGA), Lindeboom et al. (2016), identified 2,840 high-confidence PTCs in the exome data and investigated their expression using the matched RNA-sequencing data. By normalising to the median expression of the WT transcript across all cancers without the PTC variant, they calculated an "NMD efficiency" score for each PTC and then measured how important each of the previously published NMD target rules

was for determining to what extent a PTC would trigger NMD. They validated these findings using heterozygous germline nonsense variants in lymphoblastoid cell line RNA-sequencing from healthy individuals (Geuvadis study; Lappalainen et al., 2013).

As discussed in Section 1.2.4, there is marked variability in the rules determining whether a transcript will trigger NMD. Lindeboom et al. (2016) found that the biggest contribution to the NMD efficiency variability could be explained by whether the PTC was in the last exon (45.8%) and the second highest amount was how close the PTC was to the start codon (17.2%). The "50-nucleotide rule" (Section 1.2.4), that states a PTC that is less than 50nt upstream of an EJC will not trigger NMD, often thought to be the most important determinant, accounted for only 3.5% of the additional variance after the last exon rule was taken into account. Overall, 74.2% of the variability in NMD response to PTC could be explained by a combination of these rules plus exon length, mRNA half-life, allele frequency, significant motifs and the distance from the PTC to the normal stop codon. This still leaves 25.8% variability unexplained. Uncovering the underlying cause of this variability, as well as the factors that influence it, would provide important insights into the mechanisms that control NMD efficiency and would provide a better understanding of the clinical outcome of disease-associated genetic mutations.

Most NMD datasets comprise either total or nascent RNA, whether at one time point or several. Conversely, the dataset I have produced contains matched nascent and total RNA at high sequence coverage. Therefore, it is possible to track a variant within an individual transcript from transcription to steady state levels in the same cells. This allows identification of transcripts that are actively transcribed, and therefore represented in the nascent RNA samples, but are broken down by NMD, and are therefore absent in total RNA samples. For example, if there were a heterozygous variant that triggered NMD, it could be the case that one allele of a transcript is an NMD target whilst the other is not a target and so survives. In this instance, both alleles would be detectable in the nascent RNA sample but only the allele that is not an NMD target will be present at

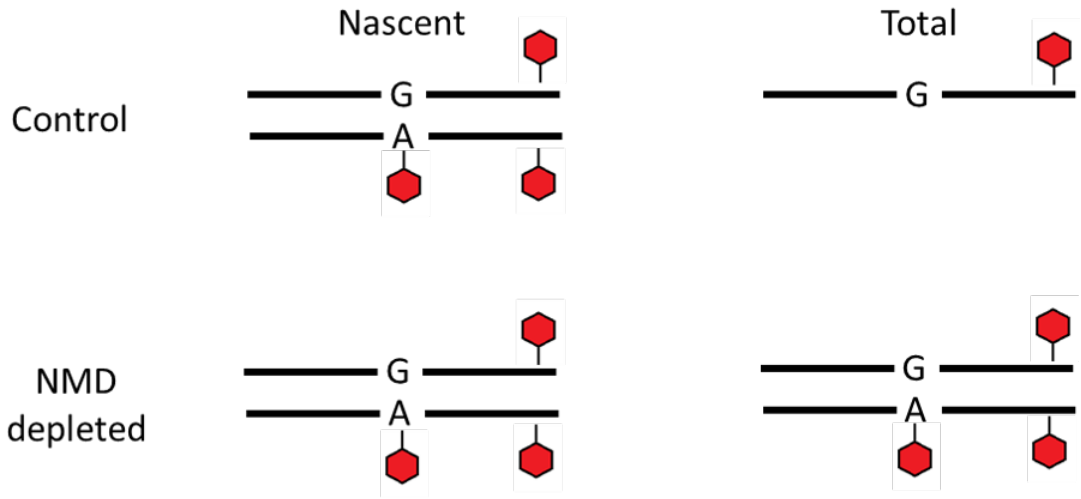


Figure 6.1: ***Matched nascent and total RNA-sequencing allows tracking of heterozygous variants***

Diagram showing a hypothetical heterozygous variant that causes a stop gain. Both alleles would be transcribed and therefore reads mapping to both the reference and alternative allele would be detectable in both control and depleted nascent RNA-sequencing. During translation, the gain of stop codon would lead to one allele being targeted by NMD and therefore in control total RNA-sequencing, only reads mapping to the reference allele would be detectable. However, upon UPF1 depletion, NMD would be abrogated and the stop gain transcript would survive, therefore reads mapping to both the reference and alternative allele would be detectable.

a comparable level in the total RNA sample. When NMD is abrogated due to depletion of the core NMD factor UPF1, the NMD target allele would survive and therefore now also be detectable in the total RNA sample (Figure 6.1). By comparing the read counts over the variant in the four samples; si-Scr and si-UPF1 treated, total and nascent RNA, the extent of NMD effect on a single variant can be measured.

6.1.2 Aims

I am interested in directly measuring stability of individual transcripts genome-wide upon abrogation of NMD, using heterozygous variants that result in an NMD-triggering allele and a wild type allele that will be processed differently. To do this I will find HeLa heterozygous genomic variants that are expressed biallelically in my dataset by measuring the expression of each allele at each variant location in the nascent RNA samples. I will then measure the abundance of each allele in the total samples and use this to calculate differential stability of the two alleles. Using the si-UPF1 depleted samples, I will measure the effect of NMD abrogation on the differential stability of the heterozygous alleles.
In summary the main aims of this chapter are:

a) Determine whether it is possible to directly measure differentially stability of heterozygous alleles when NMD is depleted

b) Use this method to identify endogenous aberrant transcripts that are NMD substrates

6.2 Results

6.2.1 Stop gain variants that are destroyed in control cells but survive when NMD is abrogated

To test whether it was possible to identify distinct alleles that are differentially regulated in my dataset, I conducted a proof of principle study. I used the mock treated nascent RNA samples since these are untreated and the least processed sample and therefore provide the most accurate representation of transcription. I called variants by comparing these reads to the human reference genome `GRCh38/hg38` using the variant caller strelka as it has an optional argument `--rna` whcih allows it to run on transcriptomic data. I then filtered the identified variants for those that were expressed in a biallelic manner ie. those where reads containing both the reference and alternate allele were present. I used the variant effect predictor (VEP) from Ensembl to annotate each variant with a predicted effect and filtered for those causing a stop gain ie. a PTC. This resulted in the identification of nine variants (Table 6.1).

I then manually checked each of these variants in the control (mock) and depleted (si-UPF1), total and nascent RNA samples by visualising the region of interest using the integrated genome viewer (IGV). Four of the positions only had reads mapping to the reference strand in the mock treated total RNA samples, suggesting they had been broken down by NMD. Of these, two variants were identified as biallelic in mock and si-UPF1 nascent RNA, monoallelic in mock treated total RNA, but biallelic in si-UPF1 treated total RNA (Figure 6.2), suggesting that the transcripts containing the alternate allele are normally degraded by NMD and therefore when UPF1 is depleted the alternate allele containing transcripts can survive.

Chromosome	Ref	Alt	VEP Annotation	Gene ID
chr12	A	T	Stop gain	ENSG00000120805
chr14	G	A	Stop gain	ENSG00000092295
chr17	C	T	Stop gain	ENSG00000278259
chr17	C	T	Stop gain	ENSG00000182687
chr1	T	A	Stop gain	ENSG00000243725
chr2	T	A	Stop gain	ENSG00000132321
chr7	G	T	Stop gain	ENSG00000213413
chr9	G	A	Stop gain	ENSG00000119403
chrX	C	T	Stop gain	ENSG00000215301

Table 6.1: ***Heterozygous stop gain variants identified in mock nascent RNA-sequencing***

Variants were called in mock nascent RNA-sequencing dataset compared to human reference genome `GRCh38/hg38` *using the variant caller strelka and the argument* `--rna` *to allow it to run on transcriptomic data. Variants were filtered for those that appeared to be biallelic ie. those where reads containing both the reference and alternate allele were present. The variant effect predictor (VEP) from Ensembl was used to annotate each variant with a predicted effect and those predicted to cause a stop gain are listed.*

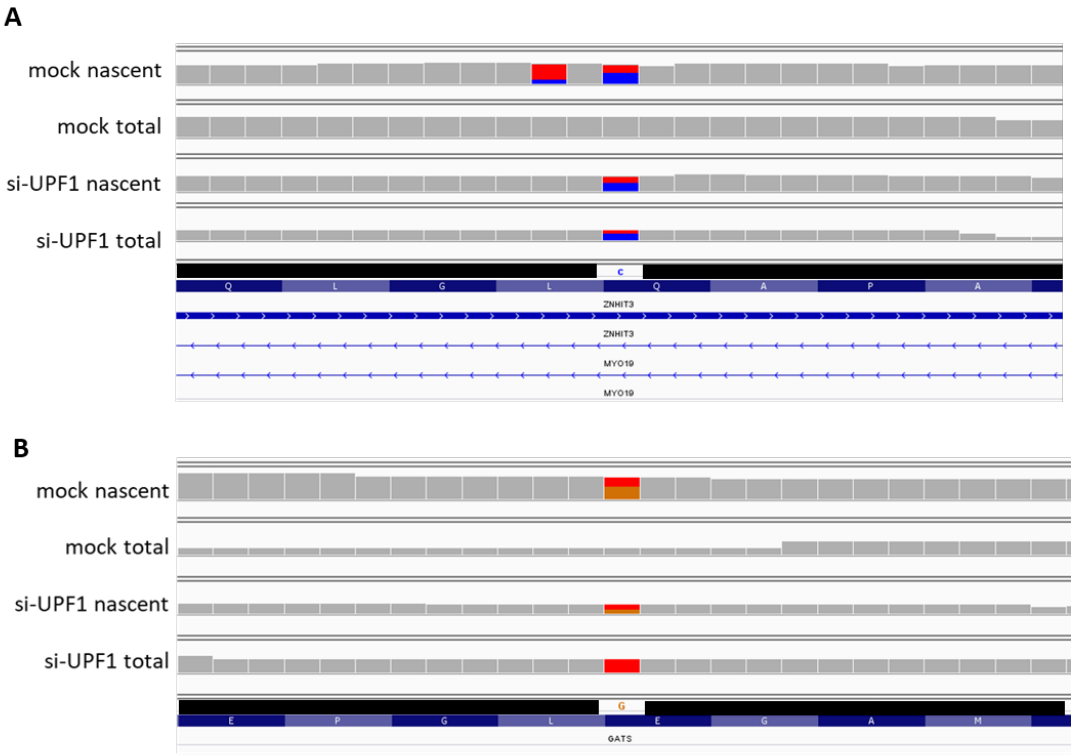


Figure 6.2: ***It is possible to detect stop gain variants that are destroyed in control cells but survive when UPF1 is depleted***

Read pile-ups from RNA-sequencing of nascent and total RNA in mock treated and si-UPF1 treated cells, visualised using the integrated genome viewer (IGV), across the locus of variants in ***(A)*** *ZNHIT3 (ENSG00000278259) on chromosome 17 and* ***(B)*** *GATS (ENSG00000213413) on chromosome 7. Both are annotated as stop gain variants by the variant effect predictor (VEP).*

6.2.2 Measuring expression of HeLa genome variants identifies thousands of biallelic transcripts

This proof of principle study shows that it is possible to identify biallelic variants that have differing fates dependent on NMD activity. However, variants are normally called using genome sequences rather than transcriptomics, since variants at the RNA level could be due to errors in transcription or splicing, or RNA-level editing rather than a reflection of a DNA change. Whole genome sequencing of HeLa is tightly regulated due to the wishes and legal case of the Lacks family. However, two research group have sequenced HeLa genomes and published their findings (Adey et al., 2013; Landry et al., 2013). Access to the data requires approval from the director of the American National Institute of Health (NIH, USA). Permission was granted to access and analyse the data in January 2019. However, the genomic data is restricted access and so identifiable genomic information such as chromosomal co-ordinates and rs number (RSID) accession identifiers must be redacted.

Adey et al. (2013) performed whole-genome sequencing (WGS) of the HeLa CCL-2 strain, identified point- and indel-mutation variations, mapped copy-number variations and loss of heterozygosity regions, and phased variants across chromosome arms into "haplotype A" and "haplotype B" (Figure 6.3). They then called variants compared to human reference genome `GRCh37/hg19`. By comparing variation and copy-number profiles for the HeLa S3 cell line and eight additional HeLa strains, they also showed that HeLa is relatively stable in terms of point variation, with few new mutations accumulating after early passaging. This is important because without WGC my HeLa cells, it is not possible to quantify the similarity of the "strain" of HeLa cells I have used to the cells used in this study. However, it is likely that the genome of he cells I have used is more similar to the HeLa genome descried in this paper than to the human reference genome, and I will validate the heterozygous genotypes by only considering regions that are confirmed by RNA-sequencing. Therefore, I can consider my data to be made up

of matched WGS, total RNA and nascent RNA datasets. For clarity, I will use "homozygous" and "heterozygous" to refer to the genomic variant calls made by Adey et al and "monoallelic" and "biallelic" to refer to the transcriptomic variant calls in my transcriptomics datasets.

Since I was primarily interested in single nucleotide variants as these will generate putative NMD-inducing PTCs, mapping accuracy was extremely important as technical errors could be wrongly interpreted as biallelic expression. Therefore, I re-mapped all `.fastq` files from my total/nascent RNA-sequencing experiment using STAR and a re-implementation of the allele-specific analysis method WASP (Van De Geijn et al., 2015) as an additional filter. WASP substitutes the SNV base with the alternative genotype in putative allelic-specific reads and re-aligns those reads to correct for the reference bias. By excluding the allelic reads that map discrepantly when alternate alleles are edited-in, WASP obtains an extremely low false positive rate when identifying allele-specific expression SNVs. Only reads that passed WASP filtering were taken forward. Percentage of reads mapping was consistently high with an average of 89.27% of reads being successfully mapped across all samples (Figure 6.4A).

The variants identified in the HeLa genome by Adey et al. (2013) were called against `GRCh37/hg19`, whereas all the analysis done in this thesis has used `GRCh38/hg38`. Therefore, I performed a lift-over to produce `GRCh38/hg38` coordinates for each variant. I could then identify which variants were expressed in my cells and detectable in nascent RNA data by using samtools mpileup to find reads covering each variant from the `.BAM` file output from STAR alignment (Section 2.2.9). Using samtools mpileup and custom python scripts written by Craig Anderson (Aitken et al., 2020), I next identified how many reads showed each of the four nucleotides at the position of each variant, producing tables with the number of total reads, A reads, C reads, G reads and T reads for each variant in each sample. Samtools mpileup also identified reads that were split across two exons, and therefore covered the genomic interval of intronic variants but did

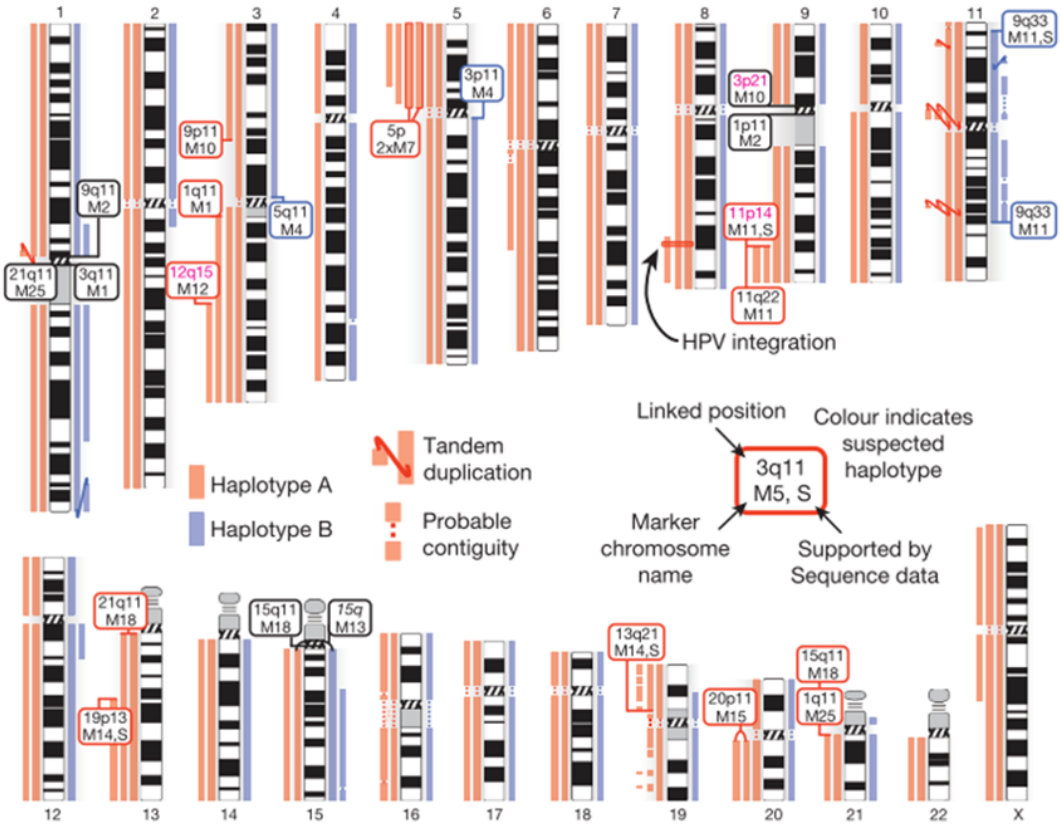


Figure 6.3: ***The haplotype phased genome of HeLa cells***

Haplotype-resolved copy-number profile of HeLa cancer cell line genome from Adey et al. (2013). Haplotypes are shown as orange and blue bars beside each chromosome, indicating copy number and specific regions present. Red links denote likely contiguity and tandem duplications. Boxes indicate marker chromosomes identified by copy number breakpoints. Boxes are coloured by haplotype (orange or blue) or black indicating unknown haplotype. Pink text represents uncertain locations and an "S" indicates links confirmed by mate-pair sequencing.

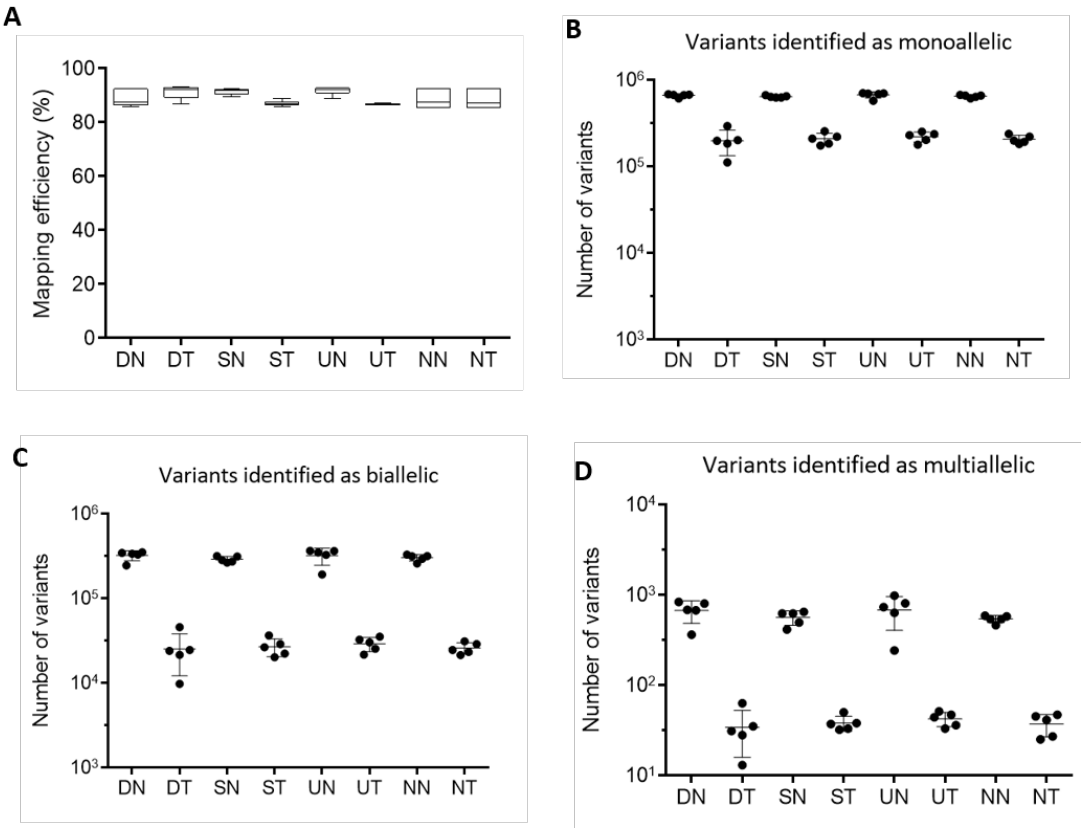


Figure 6.4: ***Thousands of variants are biallelic whilst hundreds are multiallelic***

(A) *Mapping of mock (D), si-Scr treated (S), si-UPF1 treated (U) and si-NBAS treated (N) nascent (N) and total (T) RNA-sequencing using STAR with WASP filter applied that ensures high accuracy of single nucleotide variants (SNVs) in reads. After mapping, samtools mpileup was used to identify the nucleotides present at each variant position identified by Adey et al. (2013). For each sample, number of variants identified as* ***(B)*** *monoallelic,* ***(C)*** *biallelic and* ***(D)*** *multiallelic are shown.*

not provide nucleotide information, and also instances where the variant was in the insert section, between the two paired end reads. As it was impossible to determine which allele these reads originated form, they were removed from the analysis. This resulted in 1,173,446 variants being taken forward.

From the read counts of each nucleotide, I was able to call each variants' expression as monoallelic, meaning only one nucleotide was represented, biallelic, where two nucleotides were expressed at the variant position, and in some rarer cases, multiallelic, where three or even four nucleotides were identified at the position of the variant (Figure 6.4B, C, D). The number of variants found to be expressed was higher for all cases in nascent RNA which was unsurprising given the higher number of rRNA-originating reads in total RNA, and that nascent RNA represents all transcribed regions more evenly compared to total RNA which mostly consists of mature transcripts which are more uneven in their distribution. . An average of 307,236 variants were identified as biallelically expressed in nascent samples, whereas an average of 26,610 were identified as biallelic in total samples. Surprisingly, several hundred variants were identified as multiallelic in nascent samples, and an average of 38 variants were identified as multiallelic in total samples. The identification of multiallelic reads suggests the presence of (at least) three copies of the genomic locus with three different alleles represented. Whilst interesting, this phenomenon is outwith the scope of this analysis and so these variants were discarded.

Variants were next filtered (Figure 6.5), first to only retain those with 10 or more reads in all si-Scr and si-UPF1 nascent samples, to remove lowly expressed and more variable variants. Next, variants were filtered for those identified as biallelic at the transcription level in at least one replicate of both si-Scr nascent and si-UPF1 nascent RNA samples. This resulted in a total of 124,120 expressed biallelic variants.

I also wanted to ensure that all biallelic variants were labelled as heterozygous at the genome level in the original VCF file from Adey et al. (2013), as opposed to being labelled homozygous ALT. Surprisingly, almost 50,000 variants that were identified as expressed biallelically in the transcriptomics dataset were

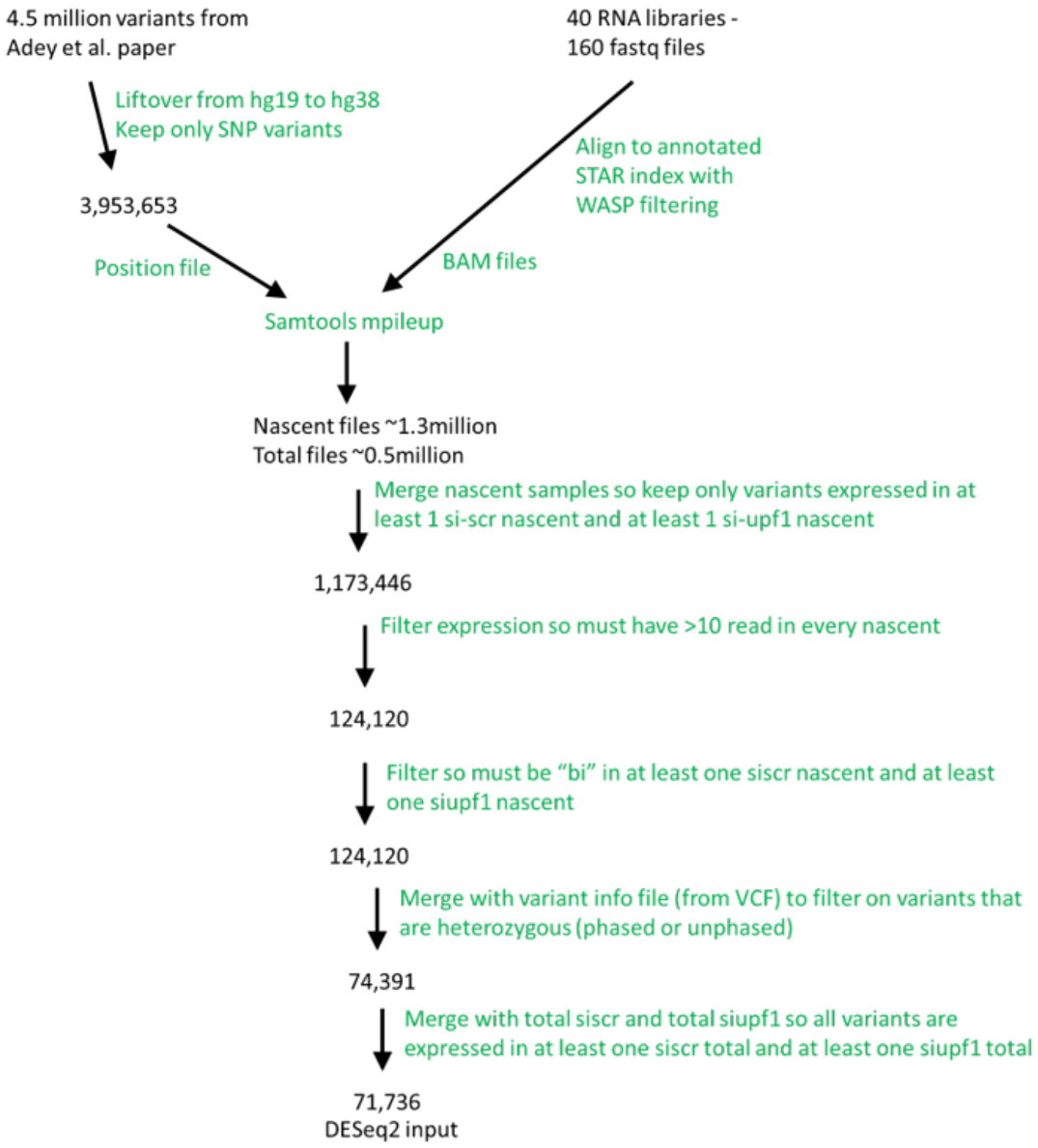


Figure 6.5: ***Processing and filtering of HeLa genome variants***

To identify heterozygous variants in the HeLa genome from Adey et al. (2013) that were expressed in the RNA-sequencing transcriptomics from this thesis, the output `.BAM` *files from STAR-WASP were analysed with samtools mpileup using the position of each variant from the variant calling format (VCF) file. Variants were then filtered for those that were covered by 10 or more reads in every si-Scr and si-UPF1 nascent RNA sample, that were identified as biallelic in at least one si-Scr and one si-UPF1 nascent RNA sample. Variants also had to be identified as heterozygous in the original VCF and that were expressed in at least si-Scr and one si-UPF1 total RNA sample.*

not labelled as heterozygous in the original VCF. Comparing the number of heterozygous variants and biallelic variants on each chromosome revealed that all heterozygous variants identified by Adey et al. (2013) that met the expression threshold of at least ten reads in all si-Scr and si-UPF1 nascent RNA samples, were expressed biallelically. Conversely, further heterozygous variants were identified on every chromosome in the HeLa cells used in this thesis compared to those used by Adey et al. (Figure 6.6). Since all the variant positions were taken from the original VCF file from Adey et al. (2013), these variants must have been identified as homozygous ALT by WGS, suggesting that in the cells I have used, either one allele has mutated to a second alternative nucleotide, or regressed to the REF. This would be an interesting question to follow up, but for this analysis I could not rule out the possibility of technical error being the source and so to be conservative these variants were discarded.

Finally, the variants were filtered to keep only those that were expressed at all in at least one replicate of both si-Scr total and si-UPF1 total, meaning the effect of NMD on them could be measured. This resulted in 71,736 heterozygous, biallelically expressed variants that were then analysed for NMD effect (Figure 6.5).

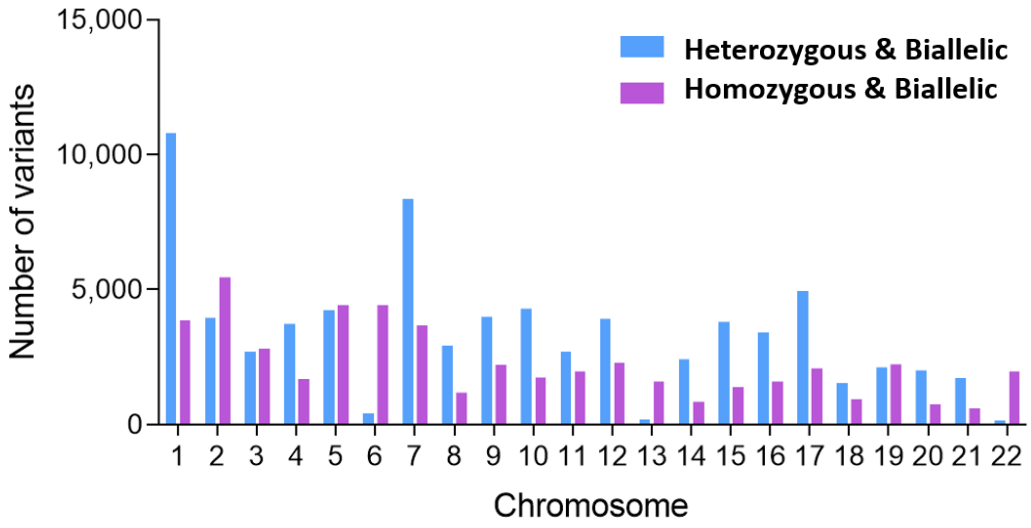


Figure 6.6: ***Further heterozygous variants are present in the HeLa cells used here than in Adey et al. (2013)***

Bars show number of variants passing expression filters on each chromosome, that are identified as heterozygous in Adey et al. (2013) and identified as biallelically expressed in the nascent RNA-sequencing described in this thesis (blue), and those that are biallelic but were not identified as heterozygous by Adey et al. (purple). Variants were filtered for those that were covered by 10 or more reads in every si-Scr and si-UPF1 nascent RNA sample. All heterozygous variants from Adey et al. that met this threshold were expressed biallelically in nascent RNA from this study.

6.2.3 Assessing NMD efficiency on variants genome-wide

To identify variants where the two alleles were differentially processed by NMD, I extended the best method I used for measuring RNA stability (Section 5.2.3) by creating a DESeq2 model that accounts for three variables. The model design was:

```
design= ~ condition + fraction + allele
+ condition:fraction + condition:allele
```

Where "condition" represents the siRNA treatment, "fraction" indicates total or nascent RNA and "allele" codes for REF or ALT nucleotide. The interaction terms define that I am interested in differences between the si-Scr and si-UPF1 conditions and that the fraction and allele are co-variables. This analysis resulted in 42 variants being identified as significantly changed in expression (Figure 6.7). These were spread across 13 chromosomes (Figure 6.8).

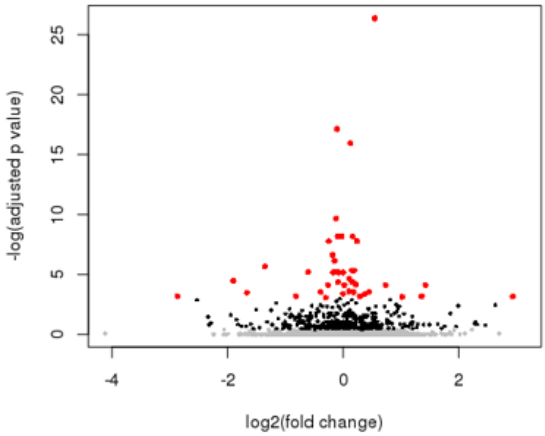


Figure 6.7: ***Significantly changed variants identified by DESeq2 model*** *Differential expression analysis using DESeq2 on read numbers of reference and alternate alleles in nascent and total RNA when UPF1 was depleted compared to si-Scr control cells. Genes nominally significantly marked in black ($p < 0.05$) and those reaching multiple testing correction threshold marked in red ($adjustedp < 0.05$).*

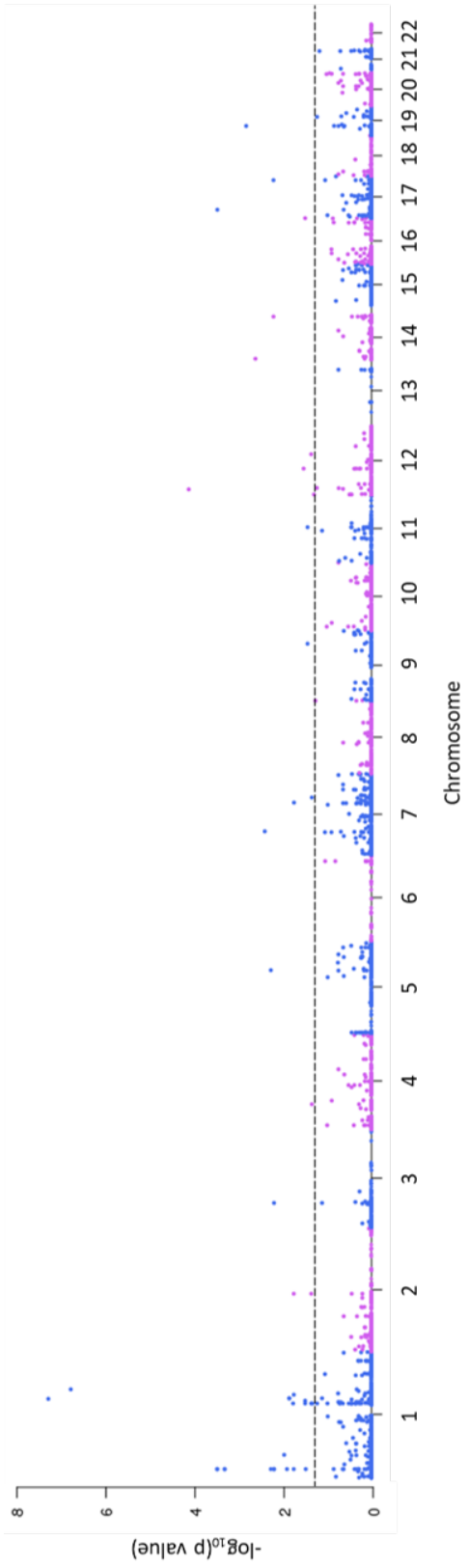


Figure 6.8: ***Significantly changed variants are spread across the genome***

Figure 6.8: ***Significantly changed variants are spread across the genome*** *Manhattan plot of the DESeq2 results identifying significantly changed variants. The y-axis represents –log10(adjusted p value) from DESeq2. The horizontal dashed line represents the threshold for significance (adjusted* $p < 0.05$*).*

I first wanted to investigate stop gain variants that would cause a PTC. Therefore, the predicted effect of each variant was found by creating a new VCF file from the DESeq2 results and analysing with the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). The output from VEP contained details about the transcript(s) that each variant is predicted to effect as well as the severity of the variant effect. The majority of variants were predicted to effect multiple transcripts and have differing effects in each. I used this annotation to subset those variants predicted to cause a stop gain in at least one of the transcripts they were contained within. Of the 42 variants identified as significantly changed by DESeq2, one was annotated as a stop gain variant by VEP. The variant was on chromosome 1 and falls within the SMG5 gene (Figure 6.9A).

I wanted to examine the underlying values of this variant to understand if the DESeq2 model was correctly identifying variants where the ratio of REF/ALT was changed in the si-UPF1 total, normalised to the si-Scr total and both nascent RNA samples. To do this, I plotted the sum of all reads over the variant, as well as the number mapping to the reference allele (REF) and alternate allele (ALT) in each condition (Figure 6.9B). This shows a greater number of reads in the si-UPF1 nascent compared to the si-Scr nascent and also a greater number of reads in the si-UPF1 total compared to the si-Scr total. The proportion of the REF and ALT reads in each condition looked roughly equivalent, but the ratio was hard to estimate. Therefore, I next plotted the number of REF reads against the number of ALT reads for each replicate of each sample (Figure 6.9C). By eye, all the points looked to be along a similar ratio, however plotting a linear regression line for si-UPF1 total RNA showed a strong skew towards REF reads.

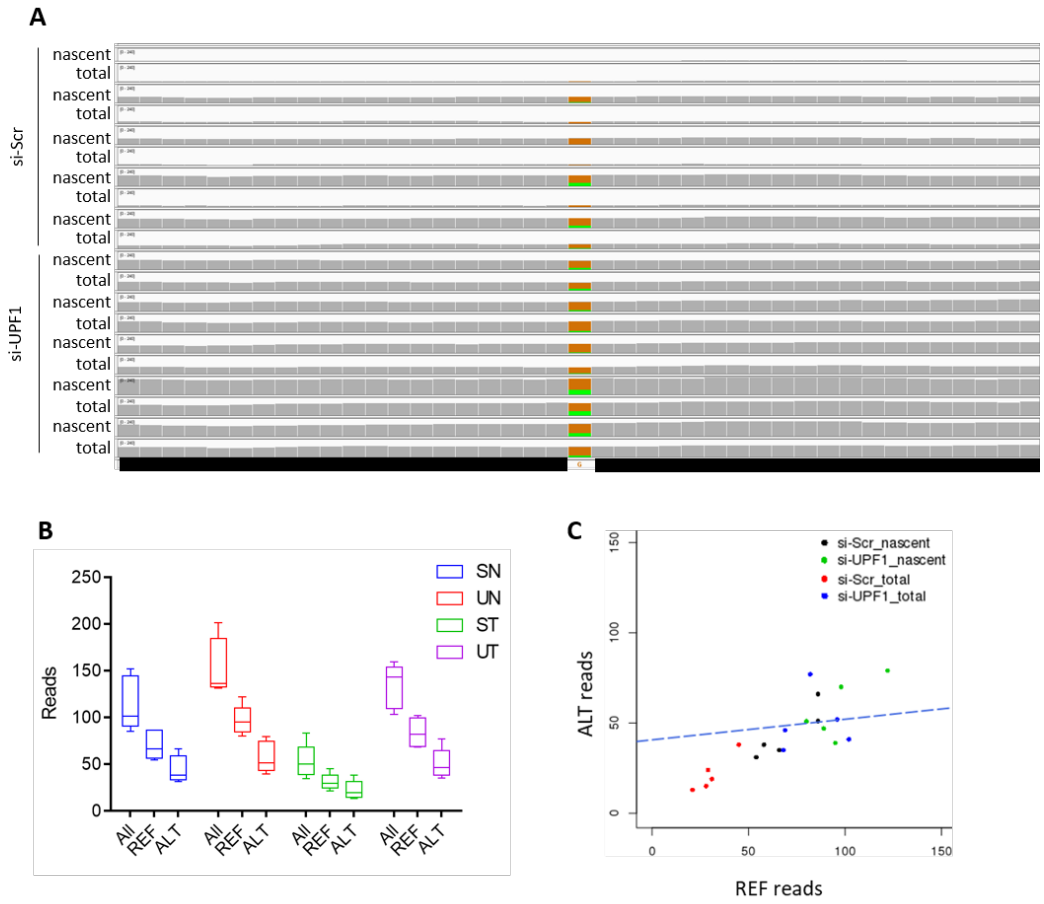


Figure 6.9: ***The stop gain variant within the SMG5 gene in detail***

Figure 6.9: ***The stop gain variant within the SMG5 gene in detail***

(A) *Read pile-ups from RNA-sequencing of nascent and total RNA in mock treated and si-UPF1 treated cells, visualised using the integrated genome viewer (IGV), across the locus of the identified stop gain variant within the SMG5 gene. Grey bars indicate nucleotides where all reads match the reference. Coloured bars indicate where two or more nucleotides have been identified.* ***(B)*** *Sum of all reads (All), number mapping to the reference allele (REF) and number mapping to the alternate allele (ALT) in si-scramble nascent RNA (SN), si-UPF1 nascent RNA (UN), si-scramble total RNA (ST) and si-UPF1 total RNA (UT). Each condition shows mean and standard deviation of n=5 replicates.* ***(C)*** *Ratio of REF reads to ALT reads in each replicate (n=5) of each condition; si-scramble nascent RNA, si-UPF1 nascent RNA, si-scramble total RNA and si-UPF1 total RNA . Dashed blue line indicates linear regression line of si-UPF1 total RNA samples.*

However, this was not an accurate way of comparing the RNA processing in each treatment, since number of counts could vary due to technical difference. Therefore, I decided to manually calculate the change in ratio of ALT to REF reads for every variant in each condition.

To do this, I first converted the read count over each variant into a transcripts per million (TPM) value to make the libraries comparable. Since this analysis is looking at single nucleotide variants, I did not normalise to transcript length, only library size. I then took the median ALT TPM and REF TPM from each condition (n=5) and calculated a ratio of total RNA ALT TPM to nascent RNA ALT TPM and ratio of total RNA REF TPM to nascent RNA REF TPM for both si-Scr treated cells and si-UPF1 treated cells (Figure 6.10A). The si-UPF1 ratios (purple) are overlaid on the si-Scr ratios (green) for comparison. In general, the two distributions look very similar, however the density of points makes interpretation difficult. Therefore, linear regression lines for si-Scr and si-UPF1 were added. The coefficient of the lines were significantly different, however the regression for si-UPF1 was actually lower than that of si-Scr ($0.48 +/- 0.004$ compared to $0.56 +/- 0.004$; coefficient and standard error, linear regression), suggesting a shift towards more REF allele than ALT allele when UPF1 is depleted.

Since only a small proportion of the variants will trigger NMD, I next used VEP to annotate the effect caused by each of the 71,736 variants and then compared the REF and ALT ratios for only those variants predicted to cause stop gains (Figure 6.10B). The linear regression line for si-UPF1 was steeper than the si-Scr line, suggesting more of the PTC-containing variants are present in total RNA when UPF1 is depleted. However, the coefficients of the lines were not significantly different ($0.87 +/- 0.092$ si-UPF1 compared to $0.84 +/- 0.086$ si-Scr; coefficient and standard error, linear regression), likely due to the small number of variants in this analysis.

To investigate the stop gain variant within the SMG5 gene, median TPM mapping to REF and ALT, in si-Scr total and nascent RNA samples and si-UPF1 total and nascent RNA samples was plotted (Figure 6.10C). The lines represent the gradient between the total and nascent REF/ALT ratios. The nascent ALT/REF

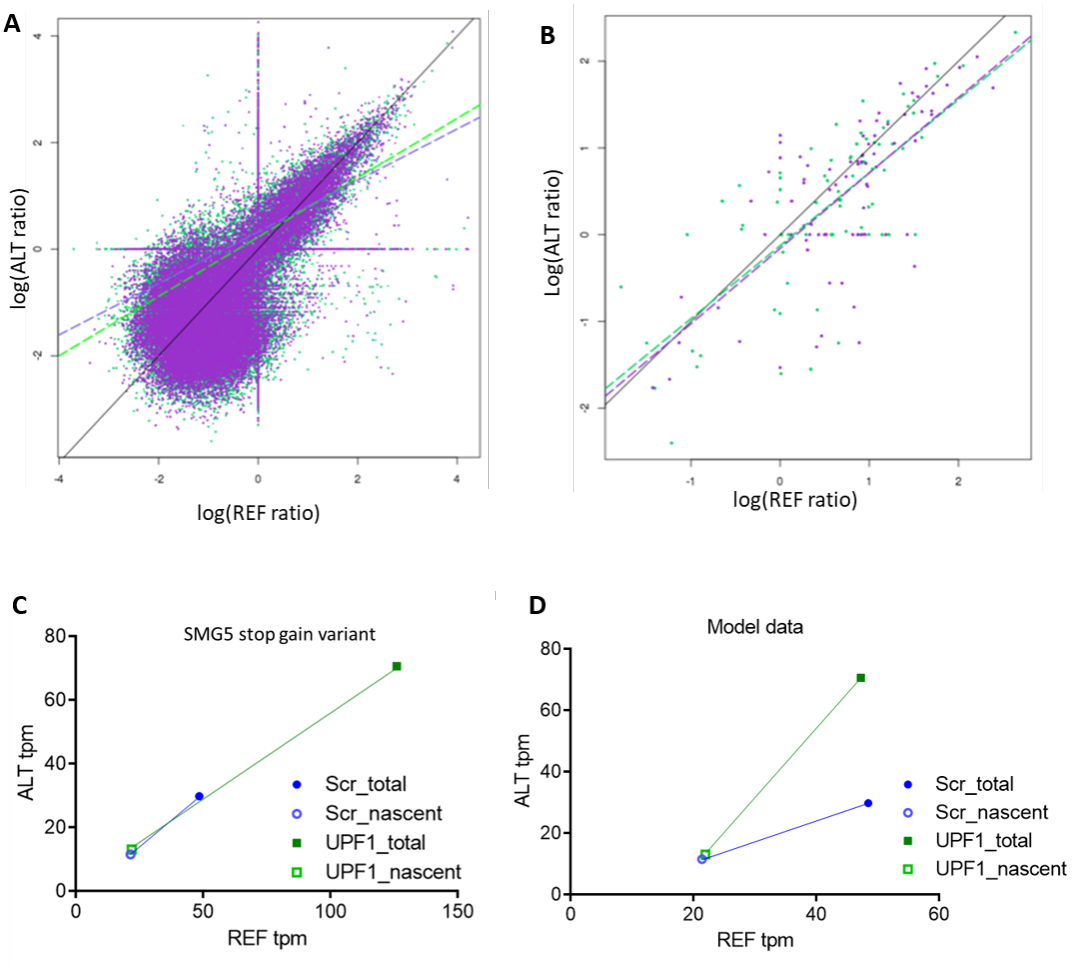


Figure 6.10: ***Calculating reference and alternate allele ratios allows comparison of allele-specific RNA processing***

Figure 6.10: ***Calculating reference and alternate allele ratios allows comparison of alelle-specific RNA processing***

(A) *Comparison of ratio of total RNA alternate allele (ALT) transcripts per million (TPM) to nascent RNA ALT TPM on the y-axis to ratio of total RNA reference allele (REF) TPM to nascent RNA REF TPM on the x-axis. Plotted for both si-Scr treated cells (green) and si-UPF1 treated cells (purple). Black line represents y=x. Dashed lines show linear regression model for each treatment.* ***(B)*** *Comparison of REF and ALT ratios for only those variants predicted to cause stop gains by the variant effect predictor (VEP). Plotted for both si-Scr treated cells (green) and si-UPF1 treated cells (purple). Black line represents y=x. Dashed lines show linear regression model for each treatment.* ***(C)*** *Median TPM (n=5) mapping to REF and ALT of the stop gain variant within the SMG5 gene, in si-Scr total and nascent RNA samples and si-UPF1 total and nascent RNA samples. Line represent gradient between total and nascent ratios.* ***(D)*** *Model data showing expected median TPM mapping to REF and ALT if a variant causes the ALT transcript to be specifically targeted by NMD.*

ratios are very similar and so are the gradients of the lines, although the gradient is slighter steeper for si-Scr towards ALT (si-Scr gradient = 0.6861709, si-UPF1 = gradient 0.5573066). However, the TPM of ALT and REF in si-UPF1 total RNA is much higher than in si-Scr total RNA, suggesting that the stop gain variant in SMG5 was identified by DESeq2 as significant due to the stabilisation of the SMG5 transcript. Conversely, Figure 6.10D shows model data indicating the expected values if a variant triggers NMD meaning the ALT transcript specifically is degraded. The ALT/REF ratios in nascent si-Scr and si-UPF1 samples are very similar showing there is no change in transcription, whereas the ratio is skewed toward REF in si-Scr total showing that the ALT transcript is being degraded in an allele specific manner. This effect is rescued when UPF1 is depleted showing that the ALT transcript is likely an NMD target. The discrepancy between the SMG5 plot and the model data confirms that the DESeq2 model used is not identifying the changes that I am looking for in this analysis. Therefore, I decided to use the manually calculated REF/ALT ratios to identify candidate variants.

DESeq2 is well suited to the analysis of count-based sequence data, identifying differences in expression between conditions (a ratio; Section 4.2.3) and can also be adapted to capture a change in ratio between conditions (a ratio of ratios; Section 5.2.3). However, to quantify changes in allele-specific transcript stability between conditions, requires the identification of a change in ratio between fractions, influenced by a treatment (a ratio of ratios of ratios). Figure 6.11A illustrates the different variables involved. Total RNA (representing steady state abundance) is normalised to nascent RNA (representing transcription). Calculating this ratio for both ALT and REF allows comparison of the processing of the allele-specific transcripts. If a variant caused a transcript to become an NMD-target, the ALT transcript would be degraded and reduced in total RNA, therefore the ratio of total to nascent for REF would be greater than the ratio of total to nascent for ALT. However, upon depletion of UPF1, NMD would be abrogated and so the ALT transcript would not be degraded and the ratio of total to nascent for REF and ALT would be similar.

To visualise this model, an odds ratio (OR) can be calculated for both si-Scr samples and si-UPF1 samples independently and then these compared by plotting an x-y distribution (Figure 6.11B). The OR for each condition was calculated using the following equation:

$$OR = ((\text{median ALT total TPM} / \text{median ALT nascent TPM}) / (\text{median REF total TPM} / \text{median REF nascent TPM}))$$

The majority of variants would be located along the y=x axis since they are not processed in a differentially allele-specific manner upon depletion of UPF1 (black oval). However, for variants that are affected by UPF1, their si-UPF1 OR would be near zero, whilst their si-Scr OR would indicate the normal allele-specific bias in RNA processing (orange oval). Those variants with a negative si-Scr OR and a near zero si-UPF1 OR are likely to be targeted by NMD.

Calculating these OR using the median TPM for each sample revealed the majority of variants were located along the y=x axis as predicted (Figure 6.11C). The stop gain variant identified in the SMG5 gene is indicated in red and shows a near zero OR for both si-Scr and si-UPF1, in line with the absence of allele-

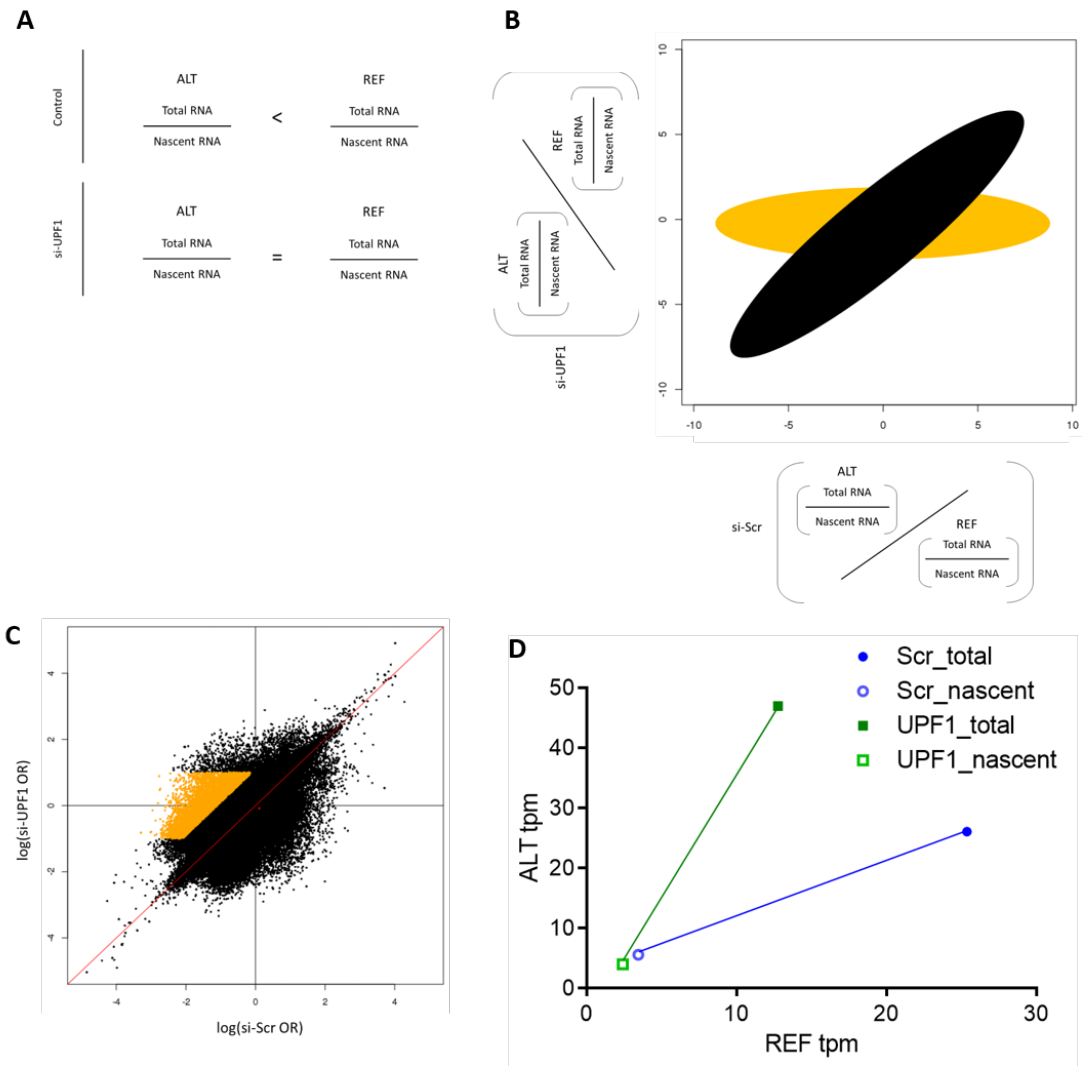


Figure 6.11: ***Calculating independent odds ratios for si-Scr and si-UPF1 treated samples allows to quantification of changes in allele-specific transcript stability between conditions***

Figure 6.11: ***Calculating independent odds ratios for si-Scr and si-UPF1 treated samples allows to quantification of changes in allele-specific transcript stability between conditions***

(A) *Pseudo equation illustrating the different variables involved in quantifying changes in allele-specific transcript stability between conditions. Total RNA (representing steady state abundance) is normalised to nascent RNA (representing transcription). Calculating this ratio for both ALT and REF allows comparison of the processing of the allele-specific transcripts.* ***(B)*** *Diagram showing plot of model data comparing the odds ratio (OR) calculated for both si-Scr samples and si-UPF1 samples independently.* ***(C)*** *Plot of 71,736 heterozygous HeLa genome variants following calculation of the OR values outlined in (B) using the median TPM for each sample (n=5). Red line shows y=x axis. The stop gain variant identified within the SMG5 gene is indicated in red. Orange points are variants with 3-fold higher si-UPF1 OR than si-Scr OR and si-UPF1 OR between -1 and 1.* ***(D)*** *Example of orange highlighted variant from (C) that recapitulates the model data showing differential allele-specific RNA processing upon depletion of UPF1. Median TPM (n=5) mapping to REF and ALT, in si-Scr total and nascent RNA samples and si-UPF1 total and nascent RNA samples. Line represent gradient between total and nascent ratios.*

specific processing seen for this variant (Figure 6.10C). To identify those variants behaving in line with an NMD-targeted variant, I highlighted those with a si-UPF1 OR at least 3-fold higher than si-Scr OR and with a si-UPF1 OR between –1 and 1. Although a complete loss of NMD activity would predict a y-axis value very close to zero, since knock down of UPF1 was ∼80% (Section 3.2.1), loss of NMD may be incomplete, so I extended this to include variants with a si-UPF1 OR between –1 and 1. This resulted in 5,314 variants being highlighted. Through manual inspection of the median TPMs of these variants, I identified a variant on chromosome 17 that recapitulated the model data (Figure 6.11D), indicating that this subset of variants contains the allele-specific UPF1-mediated differentially processed targets I am interested in. This variant is not predicted to cause a stop gain by VEP so is likely not the causative variant, but the change in ALT/REF ratio in si-UPF1 total RNA compared to si-Scr total RNA is highly suggestive of the transcript being targeted by NMD.

This analysis showed that it is possible to identify UPF1-targeted variants using independent si-Scr and si-UPF1 ORs, and this is similar to the previous analysis I have carried out using DESeq2 (Section 5.2.3). Therefore, I next carried out DE-Seq2 differential analysis of ALT/REF ratio of total RNA, normalised to nascent RNA of si-Scr and si-UPF1 independently. Plotting the log2(fold change) of the si-Scr analysis against the log2(fold change) of the si-UPF1 analysis revealed three distinct distributions along the y=x axis (Figure 6.12A). The majority of variants are near zero on both axis, since they are not processed in a differentially allele-specific manner upon depletion of UPF1. The higher and lower concentrations of data points likely reflect ploidy. Two copies of the maternal haplotype and one copy of the paternal would be expected to give a corresponding log2 expression ratio of 1 or -1 depending on whether the maternal was the REF or ALT. However, there is a clear bias to negative OR in both si-Scr and si-UPF1 samples, indicating a systematic bias towards lower ALT expression. This is understandable given that many of the variants are predicted by VEP to have high impact on protein function.

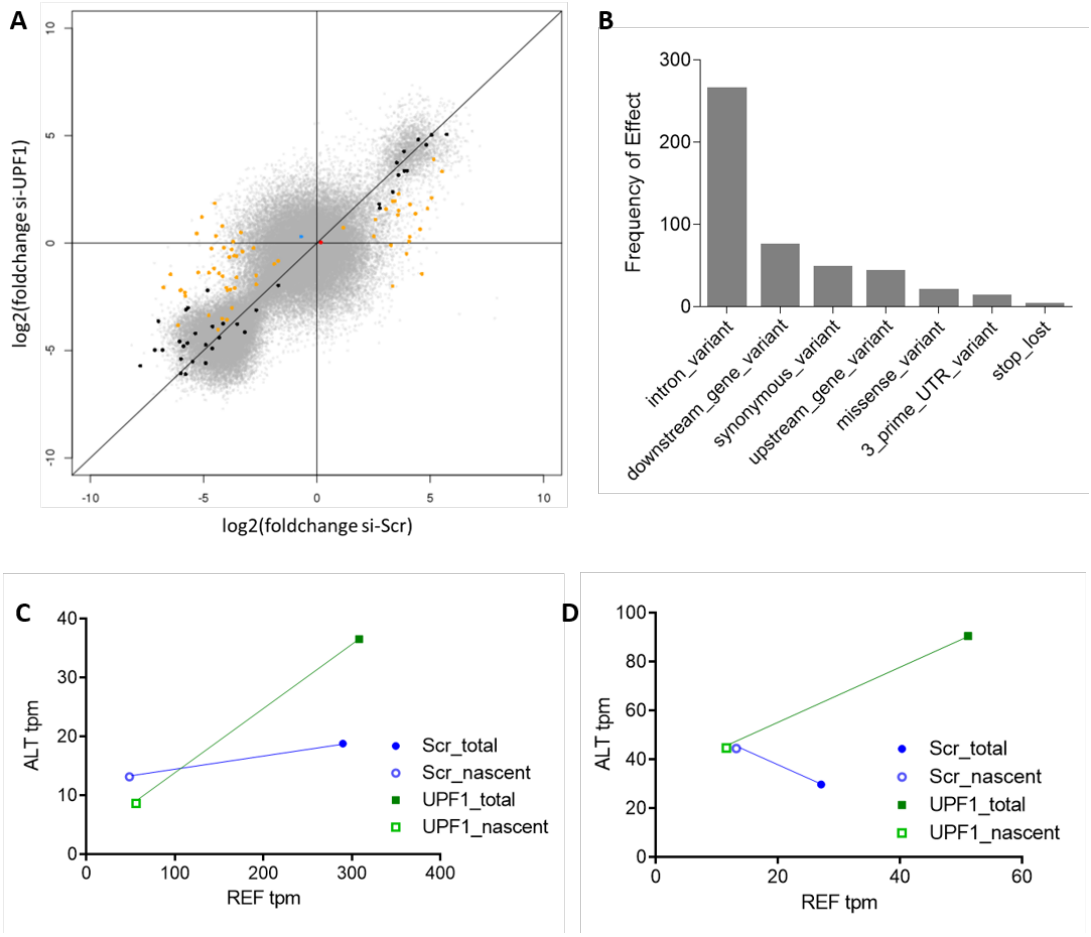


Figure 6.12: ***Using DESeq2 to independently calculate si-Scr and si-UPF1 odds ratios identifies differential allele processing***

Figure 6.12: ***Using DESeq2 to independently calculate si-Scr and si-UPF1 odds ratios identifies differential allele processing***

(A) *Comparison of log2(fold change) of si-Scr DESeq2 analysis against log2(fold change) of si-UPF1 DESeq2 analysis. All variants are shown in grey, and coloured dots indicate specific genes; the stop gain variant identified in the SMG5 gene (red), the example variant that recapitulates model data showing differential allele-specific RNA processing upon depletion of UPF1 (blue). Black points indicate variants significant in si-Scr DESeq2 analysis, whereas orange points indicate variants significant in si-Scr DESeq2 analysis but not significant in si-UPF1 DESeq2 analysis* ***(B)*** *Variant effect predictions of the 42 variants shown in orange in (A) and with a negative si-Scr log2(fold change). Since the majority of variants were predicted to effect multiple transcripts and have differing effects in each, the number of predicted effects is greater than the number of variants.* ***(C & D)*** *Example variants that appear to be NMD targets since they show evidence of differential allele-specific RNA processing upon depletion of UPF1. Median TPM (n=5) mapping to REF and ALT, in si-Scr total and nascent RNA samples and si-UPF1 total and nascent RNA samples. Line represent gradient between total and nascent ratios.*

Onto the background of all variants (grey), the stop gain variant identified within the SMG5 gene is indicated in red. As with the manual OR analysis (Figure 6.11C) this variant shows a near zero fold change for both si-Scr and si-UPF1, in line with the absence of allele-specific processing seen for this variant. The variant identified in the manual OR analysis that behaves like an NMD target is indicated in blue. It is above the x=y axis showing an increased ALT in total RNA in si-UPF1 treated cells compared to si-Scr treated cells. However, this variant was not identified as significantly changed in the si-Scr DESeq2 analysis (p=0.443). Variants that were significant in the si-Scr DESeq2 analysis are indicated in black whilst those significant in the si-Scr DESeq2 analysis but specifically **not significant** in the si-UPF1 deseq2 analysis are shown in orange. The distribution of these points recapitulates the model data in figure 6.11B. 42 of these variants also had a negative log2(fold change) in si-Scr indicating that in control cells the ALT allele is degraded. VEP annotation of these 42 variants did not identify any predicted stop gain effects, however it is possible there are PTCs elsewhere in the transcripts, possibly caused by the variants identified. 10 of the variants were predicted to be located within NMD-target transcripts. The majority of predicted effects were within non-coding regions (Figure 6.12B), including introns and the 3' UTR. The variants in regulatory regions of the transcript could affect splicing leading to alternate isoforms or retained introns for example, making them NMD targets. Since the majority of variants were predicted to effect multiple transcripts and have differing effects in each, the number of predicted effects is greater than the number of variants. Despite the lack of stop gain annotation, plotting the median REF and ALT TPM, indicated many of the variants are NMD targets since they show evidence of differential allele-specific RNA processing upon depletion of UPF1 (Figure 6.12C, D).

6.3 Discussion

6.3.1 Measuring NMD efficiency genome-wide using matched genome and RNA-sequencing requires further optimisation

In this chapter I have shown that it is possible to identify expression of heterozygous genome variants in nascent RNA-sequencing data and, by combining this with total RNA-sequencing data, identify allele-specific RNA processing. The analyses in this chapter provide a first look at the possibility of measuring direct NMD effect on individual transcripts utilising expression of heterozygous variants. This analysis provides interesting insights into differential processing of heterozygous variants, however, it requires further optimisation to utilise the dataset to its full potential.

Firstly, as shown by Figure 6.10C, the original integrated DESeq2 model I designed did not identify the changes I was interested in. This does not mean DESeq2 is not an appropriate way to perform this analysis. It is a statistical framework that is well suited to the analysis of count-based sequence data with replicates and has been successfully adapted to capture differences in expression between conditions as well as change in allele ratio between conditions. However, it is not readily adapted to capture the multiple variables required to quantify changes in transcript stability between conditions. Therefore, instead of applying a single unified framework, I developed a hierarchical strategy of quantifying allele specific differences in stability for each condition (si-Scr and si-UPF1 knockdown) and then classifying the discordance between conditions. This was a pragmatic and robust solution, but the lack of variance sharing between the analysis of each condition limits the power of this statistical test, and suggests that further work to develop a unified framework for the joint analysis of a ratio of ratios of ratios (Figure 6.11A) would be a useful endeavour.

As the nature of NMD targets is that they are degraded, this could mean that there are no reads over an ALT variant in control total RNA. Appropriately dealing with these informative zero values is essential for identifying the most efficiently targeted NMD targets. In the analysis presented in this chapter, I added a pseudocount to all median TPM values to account for this issue, but a more elegant solution would be identifying those variants with a zero count of ALT in total RNA specifically, but both ALT and REF reads represented in nascent RNA as these are likely NMD-targets. Finally, since this analysis requires a complicated statistical model, it could be that this dataset is underpowered. For example, the genotype-tissue expression GTEx project only shows summary analyses of genotypes with a minimum of 70 replicates. Making use of the haplotype information determined by Adey et al. (2013) would increase the power of these analyses as it would allow the phasing of the variants to identify which variants are within the same transcript and allow addition of the read counts across the whole transcript. It may also be possible to then identify the causative variant within a transcript that triggers NMD. Furthermore, this would allow the various triggers for NMD to be measured in terms of their likelihood to result in NMD degradation. It may even be possible to identify new characteristics or rules regarding when and how a transcript triggers the NMD pathway.

There are also various elements to this dataset that were discounted in this first pass analysis that would be interesting to follow up. Such as the "novel" biallelic variants identified in the transcriptomic data that were not heterozygous in the HeLa WGS (Adey et al., 2013). These genomic positions were taken originally from the HeLa genomic variants list, so must have been identified as homozygous ALT by WGS. This could suggest errors in the genotype calls, or that in the cells I have used, either one allele has mutated to a second alternative nucleotide, or regressed to the REF. More likely, this could represent loss of heterozygosity in the HeLa cells sequenced by Adey et al., that has been retained in the HeLa cells used in this thesis. Visualising exactly where on each chromosome these variants are located could identify specific chromosomes or large regions of them where this has occurred.

There were also a number of indels in the HeLa genome identified by Adey et al., that could lead to frameshifts and thereby creation of PTCs or aberrant splicing that could trigger NMD. Therefore, heterozygous indels could be analysed in a similar way to the SNVs to identify any that are targeted by NMD and are therefore processed in an allele-specific manner.

A fundamental disadvantage of this analysis is that I do not know how similar the HeLa cells that were WGS and the HeLa cells I performed RNA-sequencing on are. Since WGS of HeLa cells is so tightly regulated, a more controlled experiment would be to repeat this analysis in another cell line and carry out exome sequencing as well as total and nascent RNA-sequencing in the same cells. It would also be interesting to extend this to an *in vivo* model of NMD abrogation. UPF1 has been shown to be commonly mutated in certain cancers (Liu et al., 2014) so RNA-sequencing of these tumours would likely contain many NMD-targeted transcripts that have survived due to lack of UPF1 activity. However, since many pathways and processes are dysregulated in cancer, a more controlled experiment would be a conditional knock out of UPF1 in an animal model which would allow study not only of the effect of acute loss of NMD activity but also the differing effect of NMD in different tissues.

6.3.2 Future Work

It has recently been proposed that there exists a mechanism of nonsense-induced transcriptional compensation (NITC) (El-Brolosy et al., 2019; Ma et al., 2019). Both papers show that NMD degradation of a transcript leads to increased transcription of sequence similar genes and in the case of heterozygous mutants, increased transcription of the WT allele. Dosage compensation was first reported in Drosophila in 1932 (Muller and J, 1932), and genetic compensation in response to a gene knockout was first reported in yeast in 1965 (Mukherjee and Beermann, 1965). Since then, genetic compensation has been documented many times in a number of model organisms (Heard and Disteche, 2006; Raj et al., 2010; El-Brolosy and Stainier, 2017; Ghanbarian et al., 2017), but the mechanism remained elusive.

Using zebrafish and mouse mutants that harbour a PTC or have their last exon deleted, El-Brolosy et al. (2019) showed six zebrafish genes and four mouse genes that, when mutated, trigger the NITC mechanism. (Ma et al., 2019) use the term genetic compensation response (GCR) and studied knockdown and knockout models of two zebrafish genes. They demonstrated that NITC is dependent on both the presence of a PTC and the nucleotide sequence of the NMD-targeted mRNA.

El-Brolosy et al. (2019) propose that mRNA fragments generated by NMD bind selectively to complementary nucleotide sequences in related genes and carry unknown regulatory factors to induce transcription. One of the factors suggested was UPF3A, which Ma et al. show is crucial for the transcriptional compensatory mechanism, suggesting that UPFA is binding to the NMD degradation fragments, being transported with them into the nucleus and using them to identify sequence similar genes which are then upregulated. The increased transcription of gene paralogues was accompanied by increased occupancy of the genes' promoter regions by the transcription inducting epigenetic mark H3K4me3 (3 methyl groups added onto the lysine in the fourth position from N-terminus, on the tail of the H3 histone). Ma et al. (2019) found that UPF3A interacts with COMPASS — a protein complex that has been shown to activate gene transcription by modifying H3 histone proteins (Wysocka et al., 2006). Both groups find that COMPASS is integral to the genetic compensatory response. Interestingly, whilst El-Brolosy et al. found that knockout of UPF1 prevented NITC, Ma et al. observed that NITC was not impaired by knockdown of UPF1, UPF2 or UPF3B, suggesting the mechanism could be distinct from the canonical NMD pathway. This apparent discrepancy could result from the use of different approaches for perturbing UPF1, because of the different genes studied, or from other factors.

Although the details of this mechanism are currently rudimentary, this process could explain why knock-out models that cause the target mRNA to become an NMD target sometimes show no phenotype (El-Brolosy and Stainier, 2017; Rossi et al., 2015). Ma et al. showed that capn3a-knockdown zebrafish embryos had small livers, and nid1a-knockdown embryos had reduced body lengths, capn3a-

null and nid1a-null mutants appeared normal. The null zebrafish were created by introducing a PTC into the gene meaning all RNA was degraded by NMD, whereas the knockdown zebrafish were targeted at the RNA level in an NMD-independent manner. Therefore, the phenotypic differences were attributed to the upregulation of other genes in the same families, specifically upon NMD degradation.

Whilst intriguing, these studies are small scale and only show data for a handful of genes. It remains unclear whether NITC is applicable to most genes. It would be interesting to investigate how widespread and at what magnitude this phenomenon is, and whether it is influenced by factors such as the characteristics that mark a target as an NMD target. Although not designed to test this model, the dataset I have created does provide a genome-wide description of change in transcription upon depletion of the core NMD factor UPF1. Therefore, by using the si-UPF1 treated cells as a control condition, it could be investigated whether there is increased transcription of the WT allele or sequence similar genes of an NMD target in the si-Scr treated cells. To do this, I would use the heterozygous variants identified as targeted by UPF1 and compare the nascent RNA ALT and REF transcription to determine if depletion of the ALT transcript leads to an increase in REF transcription. I could also extend this to look at whether transcription of sequence-similar genes is also affected due to NMD degradation.

6.3.3 Summary

In this chapter, I develop a methodology to identify NMD-triggering variants using matched genome and transcriptome datasets. I show that it is possible to directly measure differentially stability of heterozygous alleles when NMD is depleted and use this methodology to identify endogenous aberrant transcripts that are processed in an allele-specific manner. However, further optimisation of the analysis is necessary to distinguish NMD substrates genome-wide and their relative NMD-triggering efficiency.

Chapter 7

Discussion

7.1 Appropriate methodology is vital to study NMD effects

My analyses clearly show the advantages of using differential RNA stability to measure direct effects of NMD activity compared with simply using differential abundance of total RNA. Whilst analysis of differential abundance of RNA upon depletion of UPF1 resulted in many transcripts being identified that decreased as well as increased, analysis of differential stability identified only one mRNA that was destabilised which was UPF1 itself. This suggests that many changes identified in total RNA differential expression analysis are indirect effects.

My results also indicate that a large proportion of the RNA abundance changes seen by differential expression analysis were not due to changes in RNA stability and therefore are NMD-independent effects. It seems likely that the large amount of variability in the identification of NMD targets is due, in part, to these effects being mistakenly categorised as NMD-dependent changes. Therefore, any future studies on NMD activity should not use differential abundance as a read out of NMD activity but instead measure change in RNA stability. An increasing number of techniques are being developed and optimised in this area including SLAM-seq and TT-seq making RNA stability measurements easier and more accessible. Furthermore, my results add to the body of evidence that transcription inhibition leads to altered results compared with undisturbed cellular conditions, and that to understand clearly the processes governing NMD it is important to use as biologically undisturbed a system as possible.

Despite this, RNA abundance and transcription inhibition continue to be used as techniques to measure and characterise the NMD pathwya and identify targets (Raxwal, 2020). Genome-wide measures of RNA stability are mostly found in studies where RNA degradation kinetics or mechanism are of interest and the affect of NMD is not addressed (Sorenson 2018).

From my results and published work looking at RNA stability changes upon depletion of UPF1 (Tani et al., 2012), it appears that UPF1 does not regulate as many genes as previously estimated from differential expression analysis. The identified stabilised genes, and increased in abundance account for$\sim$1% of the transcriptome. However, this does not indicate that NMD is unimportant, since targetting of one factor in a complex pathway could alter the levels of many different proteins and have strong affects on cellular signalling. Furthermore, there is much evidence for the importance of NMD in development, differentiation and stress response. Therefore, perhaps future studies on NMD should utilise appropriate cell types or external stimuli to recreate these physiological contexts that NMD may be most active.

It is also important to consider the kinetics of NMD, since harvesting RNA only a short period after UPF1 depletion will give a more clear picture of the primary effects, whereas at 72hr post depletion (as used here) the bigger picture is being captured including downstream effects. This could also mean more transient effects that are quickly mitigated by the cell are also undetected. Many gene regulation networks are very rapid for example some RNA-binding proteins bind and regulate their own mRNA. To truly appreciate the waves of affects caused by depletion of a multi-functional protein such as UPF1 many time points should be utilised.

7.2 An ER-NMD pathway quality controls transcripts translated at the ER

The results in this thesis support a role for UPF1 in regulation of transcripts at the ER. Since NMD is co-translational, and these transcripts are translated when the ribosome is docked at the translocon and UPF1 has been shown to interact with the SEC61 channel, it seems sensible to conclude that UPF1-mediated quality control is occurring at the ER membrane. Since not only secretome transcripts are translated at the ER, this mechanism would regulate expression of a notable proportion of cellular mRNAs. However, a lot about ER-NMD is still unknown. For example, does UPF1 get carried to the ER whilst bound to mRNA or is it specifically recruited when necessary? Longman et al. (2020) showed that UPF1 is tethered to the ER by NBAS which suggests it could always be present or that perhaps the NBAS-UPF1 interaction stabilises the RNA at the ER membrane. Also, in cytoplasmic NMD the incomplete nascent peptide is readily accessible to the degradation machinery in the cytoplasm, whereas nascent peptide translated at the translocon is deposited straight into the ER lumen so it must then be recognised as incomplete and then actively exported out of ER. Is the ERAD pathway then activated to retrotranslocate the peptide to the cytoplasm so it can be degraded by ubiquitination or is another mechanism utilised?

Considering all the evidence thus far, it seems likely that a quality control mechanism **does** occur at the ER and since UPF1 is localised to the translocon and affects both abundance and, crucially, stability of ER translated transcripts, it seems probable that this is an NMD pathway at the ER membrane. What is not clear is the exact mechanism by which this occurs. In this thesis, I showed that UPF2 seems to have a very limited effect on expression of transcripts in the membrane-associated fraction (Chapter 4), suggesting that ER-NMD could be occurring in an UPF2-independent manner. Since UPF2 is the bridge between UPF1 and UPF3B, this would likely mean that the EJC would not be involved in

ER-NMD and the DECID complex would not form. Since the EJC can be fairly heterogeneous in its composition, perhaps this suggests that EJC loaded onto secretome transcripts are less likely to include UPF2. Or perhaps the non-uniform distribution of EJC deposition on transcripts (Le Hir et al., 2016) is skewed away from secretome transcripts. Furthermore, recent findings show that NMD occurs with equal probability during each round of translation of an mRNA molecule (Hoek et al., 2019) and since all EJCs are thought to be removed in the pioneer round of translation so would therefore not be present in later rounds, this calls into question whether NMD activation requires the EJC to proceed.

To understand more fully the ER-NMD mechanism, it would be interesting to understand more about the other putative NMD factor that localises there; SEC13. Like NBAS, the *C. elegans* homologue of SEC13 was also found in an NMD factor screen (Casadio et al., 2015). Like NBAS, SEC13 has been shown to increase the abundance of the NMD reporter HBB construct with a PTC, and seems to have a stronger effect than NBAS. However confusingly, the HBB gene is not part of the secretome. One study that aimed to add localisation signals to transfected genes found that a WT β-globin reporter was partially found in the membrane fraction (Partridge et al., 1999), but it was not identified in any of the three experiential datasets used in this thesis to denote ER-translocation. Whilst it has been shown that up to 40% of cytosol-encoding mRNAs can be translated at the ER (Reid and Nicchitta, 2012), this assumed to occur at low frequency. Therefore, the increase in abundance of PTC-containing HBB mRNA upon depletion of NBAS or SEC13 seems contradictory. A more clear measurement of ER-specific NMD activity would be the change in **stability** of a PTC-containing gene with and without an ER localisation signal, upon depletion of these factors as well as core NMD factors such as UPF1 and UPF2. This would allow measurement of how specific to ER-translated mRNAs the effects are.

7.3 The role of NBAS in RNA quality control requires further investigation

The question of whether NBAS is an NMD factor is still not conclusively answered but the results presented here seem to point to a different mechanism that would explain the seemingly contradictory results.

An overwhelming amount of evidence agrees that NBAS and UPF1 interact, which makes the model of NBAS having a function in human NMD appealing. However, as already discussed, UPF1 has many known cellular functions and interacts with a variety of protein subtypes, the vast majority of which are not thought to be involved in NMD. Therefore, this evidence is not sufficient on it's own to classify NBAS as an NMD factor. In this thesis I show that, as well as UPF1, NBAS interacts with the SEC61 translocon (Chapter 3) where translation takes place at the ER. However it is not clear whether these interactions are simultaneous and this would be an interesting point for further study. Three-colour immunofluorescence and super-resolution microscopy may be able to determine whether the three proteins are found in close proximity or whether their inter-protein interactions are independent.

It has been argued that NBAS recruits UPF1 to the ER, however this has not been shown conclusively and since UPF1 binds RNA it is likely that it could be transported to the ER as part of the mRNP when an the RNA contains an ER-localisation signal. One way to investigate the role of NBAS on UPF1's localisation to the ER would be to carry out a PLA on UPF1 and SEC61B in the presence and absence of NBAS.

NBAS does appears to tether UPF1 at the ER, evidenced by UPF1's faster diffusion at the ER region in the absence of NBAS. An interesting experiment to investigate the effect of NBAS on UPF1's NMD activity at the ER, would be to carry out UPF1 CLIP in the presence and absence of NBAS. UPF1 is a known RBP and successful CLIP experiments have been performed that indicate the

transcripts it binds. Therefore, comparing UPF1 binding in WT cells to those depleted of NBAS would indicate the effect NBAS had on recruiting and/or stabilising UPF1 on target RNAs translated at the ER. Since UPF1 binds many transcripts transiently, and it's phosphorylation is thought to stabilise these interactions, CLIP specifically of P-UPF1 would be likely give a more accurate description of NMD targets. The results of this experiment may shed light on the mechanism by which NBAS affects abundance but not stability of ER translated transcripts.

Transcripts with increased abundance when NBAS was depleted were identified in the first RNA-sequencing experiment discussed in this thesis (Chapter 4). These transcripts overlapped with UPF1 targets (Figure 4.6A), however these were concluded to be lowly expressed RNAs with small fold changes (Figure 5.10F), and the vast majority did not reach significance in the STAR-HTSeq-DESeq2 re-analysis of this dataset (Chapter 5, Figure 5.10D). A large number of transcripts were identified as increased in abundance in the second dataset analysed using the STAR-HTSeq-DESeq2 pipeline (Chapter 4), however these correlated very weakly with UPF1 targets from that experiment (Figure 4.11A). Whilst depletion of UPF2 or UPF1 in either experiment led to increased abundance and stability of several NMD factors, depletion of NBAS did not lead to increase in abundance or stability of any NMD factors in either dataset (Chapter 4, 5). Finally, no transcripts were identified as increased in stability by any of the four methods used (Chapter 5), strongly suggesting that the few targets that did change in abundance are affected in an NMD-independent manner.

NBAS has a known role in secretome protein localisation and post-translational processing, so it is feasible that disrupting these processes would affect other cellular systems such as RNA processing or transcription. Therefore, the loss of NBAS from the vesicular transport system could be the cause of the changes in RNA abundance seen. Since UPF1 has been shown to be localised to the ER, the correlation of the lowly expressed transcripts with increased abundance when either UPF1 or NBAS was depleted (Figure 4.6A) could represent on disruption of a non-NMD mechanism at the ER. This would concur with the finding that

many of the targets were ER-associated (Figure 4.7E). It would be interesting to see if NBAS depletion affects mRNA localisation for example reducing the likelihood that transcripts localise to the ER. To test this, the abundance of specific transcripts could be compared in the membrane fraction and cytoplasmic fraction of the fractionated RNA-sequencing dataset from this thesis (Chapter 4). This could reveal whether re-localisation is occurring between the ER and the cytoplasm upon depletion of NBAS and/or UPF1.

UPF1 has been shown to be localised to the ER and a regulator of ER stress. Therefore, the interaction between NBAS and UPF1 could be as part of an ER stress pathway. UPF1 has recently been shown to promotes aggresome formation (Park et al., 2020). Aggresomes are perinuclear non-membranous compartments that contain potentially toxic misfolded polypeptides that are prone to aggregation. The polypeptide aggregates are selectively recognized by aggresome-targeting cellular factors and transported to the microtubule-organizing centre (MTOC) via microtubule-mediated retrograde movement, and eventually cleared from the cell via the autophagy pathway. Since the outer nuclear membrane is continuous with the ER membrane, NBAS could be in close proximity to these aggresomes, event potentially recruiting or tethering UPF1.

As well as being the site of ER translation, the SEC61 translocon has been shown to be able to retrotranslocate faulty nascent proteins out of the ER lumen into the cytoplasm for ubiquitination and degradation as part of the ERAD ER homeostasis pathway. This is another avenue that would be interesting to follow up in terms of NBAS function. Since NBAS appears to have no effect on transcript stability, it would be interesting to investigate if it is involved in aberrant protein degradation by facilitating the efflux of these proteins via the SEC61 translocon. Since faulty proteins must be moved from the ER lumen to the cytoplasm to be degraded, the ERAD and ER-NMD pathways are likely to be inherently linked. Therefore, it seems plausible that abrogation of the ERAD machinery could have a negative effect on the function and efficiency of ER-NMD.

Since NBAS has a known role in protein complex assembly as part of Golgi-to-ER retrograde transport, I think the most compelling explanation is that NBAS functions to facilitate assembly of other protein complexes too. It could have a role in the secure docking of the ribosome to allow translation to occur, in the assembly of the SURF complex upon stalling of the ribosome at a PTC or in modulating the SEC61 channel to facilitate retrotranslocation of faulty, nascent peptides. To identify other avenues of investigation in understanding the other functions of NBAS, further pathway analysis could be utilised to identify the mechanisms that the targets identified as affected by NBAS are involved in. For example, identifying any transcription factors that are affected by NBAS could elucidate downstream effects due to altered levels of that transcription factor.

As discussed in Section 1.3.3, several patients have been identified to have harbour in their NBAS gene. Phenotypes are complex and include short stature, optic nerve atrophy, bone fragility, abnormal nuclear shape in neutrophil granulocytes, small hands and feet, developmental delay, seizures and reduced natural killer cells (Maksimova et al., 2010; Lacassie et al., 2020; Balasubramanian et al., 2017; Haack et al., 2015; Wimmer et al., 1999; Frühwald et al., 2000; Segarra et al., 2015). Whilst the evidence suggests that most of the NBAS mutations cause reduced level of NBAS protein, it is not currently known by what mechanism the mutations in NBAS lead to these phenotypes. The results in this thesis suggest that it is likely not through a direct role in RNA stability.

7.4 Closing Remarks

This project set out to investigate the mechanism by which the putative NMD factor NBAS was involved in the pathway and whether a dedicated NMD pathway at the ER was present in human cells. The conclusions in this thesis may seem to contradict other work that has been done on NBAS and the ER-NMD pathway. However, my findings taken together with the previous work do support the existence of an ER-specific NMD pathway. This pathway may well be mechanistically different to canonical NMD, for example in an UPF2-independent manner. However, given the conflicting findings regarding the role of NBAS in ER-NMD it seems unlikely that NBAS is integral to the ER-NMD pathway and it's role appears to be largely redundant.

Utilising paired nascent and total RNA-sequencing to determine genome-wide steady-state RNA expression levels and RNA decay rates is a powerful approach for identifying *bona fide* targets regulated by RNA decay factors, including, but not limited to, NMD factors. This approach could be applied to and shed new light on various processes including quality control, checkpoint, and proofreading mechanisms that ensure the fidelity of the genetic information flow.

Bibliography

Adey, A., Burton, J. N., Kitzman, J. O., et al. (2013). The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature*, 500(7461):207–211.

Adkins, J. N., Varnum, S. M., Auberry, K. J., et al. (2002). Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. *Mol. Cell. Proteomics*, 1(12):947–955.

Aitken, S. J., Anderson, C. J., Connor, F., et al. (2020). Pervasive lesion segregation shapes cancer genome evolution. *Nature*, 583(7815):265–270.

Ajamian, L., Abrahamyan, L., Milev, M., et al. (2008). Unexpected roles for UPF1 in HIV-1 RNA metabolism and translation. *RNA*, 14(5):914–927.

Alexandrov, A., Colognori, D., Shu, M.-D., and Steitz, J. A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc. Natl. Acad. Sci. U. S. A.*, 109(52):21313–21318.

Amengual, J., Guo, L., Strong, A., et al. (2018). Autophagy is required for Sortilin-Mediated degradation of apolipoprotein B100. *Circ. Res.*, 122(4):568–582.

An, H., Ordureau, A., Paulo, J. A., et al. (2019). TEX264 is an endoplasmic Reticulum-Resident ATG8-Interacting protein critical for ER remodeling during nutrient stress. *Mol. Cell*, 74(5):891–908.e10.

Anastasaki, C., Longman, D., Capper, A., et al. (2011). Dhx34 and nbas function in the NMD pathway and are required for embryonic development in zebrafish. *Nucleic Acids Res.*, 39(9):3686–3694.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the cas9 endonuclease. *Nature*, 513(7519):569–573.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169.

Andersen, C. B. F., Ballut, L., Johansen, J. S., et al. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science*, 313(5795):1968–1972.

Andersson, R., Gebhard, C., Miguel-Escalada, I., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461.

Aoki, T., Ichimura, S., Itoh, A., et al. (2009). Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in golgi-to-endoplasmic reticulum retrograde transport. *Mol. Biol. Cell*, 20(11):2639–2649.

Arribere, J. A., Cenik, E. S., Jain, N., et al. (2016). Translation readthrough mitigation. *Nature*, 534(7609):719–723.

Arvan, P., Zhao, X., Ramos-Castaneda, J., and Chang, A. (2002). Secretory pathway quality control operating in golgi, plasmalemmal, and endosomal systems. *Traffic*, 3(11):771–780.

Ast, T., Cohen, G., and Schuldiner, M. (2013). A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell*, 152(5):1134–1145.

Atkin, A. L., Altamura, N., Leeds, P., and Culbertson, M. R. (1995). The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm. *Mol. Biol. Cell*, 6(5):611–625.

Aviram, N., Ast, T., Costa, E. A., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature*, 540(7631):134–138.

Aznarez, I., Nomakuchi, T. T., Tetenbaum-Novatt, J., et al. (2018). Mechanism of Nonsense-Mediated mRNA decay stimulation by splicing factor SRSF1. *Cell Rep.*, 23(7):2186–2198.

Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of cas9 RNA-guided endonucleases. *Bioinformatics*, 30(10):1473–1475.

Baird, T. D. and Hogg, J. R. (2018). Using Tet-Off cells and RNAi knockdown to assay mRNA decay. In Lamandé, S. R., editor, *mRNA Decay: Methods and Protocols*, pages 161–173. Springer New York, New York, NY.

Baker, K. E. and Parker, R. (2004). Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr. Opin. Cell Biol.*, 16(3):293–299.

Balasubramanian, M., Hurst, J., Brown, S., et al. (2017). Compound heterozygous variants in NBAS as a cause of atypical osteogenesis imperfecta. *Bone*, 94:65–74.

Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016). In vivo aspects of protein folding and quality control. *Science*, 353(6294):aac4354.

Baldridge, R. D. and Rapoport, T. A. (2016). Autoubiquitination of the hrd1 ligase triggers protein retrotranslocation in ERAD. *Cell*, 166(2):394–407.

Bao, J., Vitting-Seerup, K., Waage, J., et al. (2016). UPF2-Dependent Nonsense-Mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3'UTR transcripts. *PLoS Genet.*, 12(5):e1005863.

Baptista, M. A. P. and Dölken, L. (2018). RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. *Nat. Methods*, 15(3):171–172.

Barbosa, I., Haque, N., Fiorini, F., et al. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat. Struct. Mol. Biol.*, 19(10):983–990.

Barlowe, C. and Helenius, A. (2016). Cargo capture and bulk flow in the early secretory pathway. *Annu. Rev. Cell Dev. Biol.*, 32:197–222.

Barmada, S. J., Ju, S., Arjun, A., et al. (2015). Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc. Natl. Acad. Sci. U. S. A.*, 112(25):7821–7826.

Beckmann, B. M., Horos, R., Fischer, B., et al. (2015). The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nat. Commun.*, 6(1):10127.

Belgrader, P., Cheng, J., Zhou, X., et al. (1994). Mammalian nonsense codons can be cis effectors of nuclear mRNA half-life. *Mol. Cell. Biol.*, 14(12):8219–8228.

Bergen, V., Lange, M., Peidli, S., et al. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.*

Berglund, A.-C., Sjölund, E., Ostlund, G., and Sonnhammer, E. L. L. (2008). InParanoid 6: eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Res.*, 36(Database issue):D263–6.

Bernales, S., McDonald, K. L., and Walter, P. (2006). Autophagy counterbalances endoplasmic reticulum expansion during the unfolded protein response. *PLoS Biol.*, 4(12):e423.

Bernales, S., Schuck, S., and Walter, P. (2007). ER-phagy: selective autophagy of the endoplasmic reticulum. *Autophagy*, 3(3):285–287.

Besse, F. and Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat. Rev. Mol. Cell Biol.*, 9(12):971–980.

Bhattacharyya, S. N., Habermacher, R., Martine, U., et al. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell*, 125(6):1111–1124.

Bhuvanagiri, M., Schlitter, A. M., Hentze, M. W., and Kulozik, A. E. (2010). NMD: RNA biology meets human genetic medicine. *Biochem. J*, 430(3):365–377.

Blázquez, A.-B., Escribano-Romero, E., Merino-Ramos, T., et al. (2014). Stress responses in flavivirus-infected cells: activation of unfolded protein response and autophagy. *Front. Microbiol.*, 5:266.

Blower, M. D. (2013). Molecular insights into intracellular RNA localization. *Int. Rev. Cell Mol. Biol.*, 302:1–39.

Boehm, V., Haberman, N., Ottens, F., et al. (2014). 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep.*, 9(2):555–568.

Bonifacino, J. S. and Weissman, A. M. (1998). Ubiquitin and the control of protein fate in the secretory and endocytic pathways. *Annu. Rev. Cell Dev. Biol.*, 14:19–57.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*, 126(4):713–725.

Bothmer, A., Phadke, T., Barrera, L. A., et al. (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat. Commun.*, 8:13905.

Bothmer, A., Rommel, P. C., Gazumyan, A., et al. (2013). Mechanism of DNA resection during intrachromosomal recombination and immunoglobulin class switching. *J. Exp. Med.*, 210(1):115–123.

Bovaird, S., Patel, D., Padilla, J.-C. A., and Lécuyer, E. (2018). Biological functions, regulatory mechanisms, and disease relevance of RNA localization pathways. *FEBS Lett.*, 592(17):2948–2972.

Brannan, K. W., Jin, W., Huelga, S. C., et al. (2016). SONAR discovers RNA-Binding proteins from analysis of Large-Scale Protein-Protein interactomes. *Mol. Cell*, 64(2):282–293.

Bravo, R., Parra, V., Gatica, D., et al. (2013). Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration. *Int. Rev. Cell Mol. Biol.*, 301:215–290.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, 34(5):525–527.

Brinkman, E. K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, 42(22):e168.

Brittis, P. A., Lu, Q., and Flanagan, J. G. (2002). Axonal protein synthesis provides a mechanism for localized regulation at an intermediate target. *Cell*, 110(2):223–235.

Brohawn, S. G. and Schwartz, T. U. (2009). Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. *Nat. Struct. Mol. Biol.*, 16(11):1173–1177.

Brumbaugh, K. M., Otterness, D. M., Geisen, C., et al. (2004). The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol. Cell*, 14(5):585–598.

Bruno, I. G., Karam, R., Huang, L., et al. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. *Mol. Cell*, 42(4):500–510.

Bühler, M., Steiner, S., Mohn, F., et al. (2006). EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat. Struct. Mol. Biol.*, 13(5):462–464.

Burri, L., Varlamov, O., Doege, C. A., et al. (2003). A SNARE required for retrograde transport to the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U. S. A.*, 100(17):9873–9877.

Cannarozzi, G., Schraudolph, N. N., Faty, M., et al. (2010). A role for codon order in translation dynamics. *Cell*, 141(2):355–367.

Carmody, S. R. and Wente, S. R. (2009). mRNA nuclear export at a glance. *J. Cell Sci.*, 122(Pt 12):1933–1937.

Carreras-Sureda, A. and Hetz, C. (2015). RNA metabolism: putting the brake on the UPR. *EMBO Rep.*, 16(5):545–546.

Carter, M. S., Li, S., and Wilkinson, M. F. (1996). A splicing-dependent regulatory mechanism that detects translation signals. *EMBO J.*, 15(21):5965–5975.

Carvalho, P., Goder, V., and Rapoport, T. A. (2006). Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell*, 126(2):361–373.

Casadio, A., Longman, D., Hug, N., et al. (2015). Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. *EMBO Rep.*, 16(1):71–78.

Castello, A., Fischer, B., Eichelbaum, K., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell*, 149(6):1393–1406.

Castello, A., Horos, R., Strein, C., et al. (2013). System-wide identification of RNA-binding proteins by interactome capture. *Nat. Protoc.*, 8(3):491–500.

Chan, W., Costantino, N., Li, R., et al. (2007a). A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic Acids Res.*, 35(8):e64.

Chan, W.-K., Huang, L., Gudikote, J. P., et al. (2007b). An alternative branch of the nonsense-mediated decay pathway. *EMBO J.*, 26(7):1820–1830.

Chan, Y. L., Suzuki, K., Olvera, J., and Wool, I. G. (1993). Zinc finger-like motifs in rat ribosomal proteins S27 and S29. *Nucleic Acids Res.*, 21(3):649–655.

Chang, Y.-F., Chan, W.-K., Imam, J. S., and Wilkinson, M. F. (2007). Alternatively spliced t-cell receptor transcripts are up-regulated in response to disruption of either splicing elements or reading frame. *J. Biol. Chem.*, 282(41):29738–29747.

Chapin, A., Hu, H., Rynearson, S. G., et al. (2014). In vivo determination of direct targets of the nonsense-mediated decay pathway in drosophila. *G3*, 4(3):485–496.

Chaudhuri, I., Söding, J., and Lupas, A. N. (2008). Evolution of the beta-propeller fold. *Proteins*, 71(2):795–803.

Chen, C. K. M., Chan, N. L., and Wang, A. H. J. (2011). The many blades of the β-propeller proteins: Conserved but versatile.

Chen, J.-L., Xu, W., and Stamnes, M. (2005). In vitro reconstitution of ARF-regulated cytoskeletal dynamics on golgi membranes. *Methods Enzymol.*, 404:345–358.

Chen, Q., Xiao, Y., Chai, P., et al. (2019). ATL3 is a tubular ER-Phagy receptor for GABARAP-Mediated selective autophagy. *Curr. Biol.*, 29(5):846–855.e6.

Cheng, J., Belgrader, P., Zhou, X., and Maquat, L. E. (1994). Introns are cis effectors of the nonsense-codon-mediated reduction in nuclear mRNA abundance. *Mol. Cell. Biol.*, 14(9):6317–6325.

Cheng, J. and Maquat, L. E. (1993). Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. *Mol. Cell. Biol.*, 13(3):1892–1902.

Chino, H., Hatta, T., Natsume, T., and Mizushima, N. (2019). Intrinsically disordered protein TEX264 mediates ER-phagy. *Mol. Cell*, 74(5):909–921.e6.

Cho, H., Kim, K. M., and Kim, Y. K. (2009). Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. *Mol. Cell*, 33(1):75–86.

Cho, H., Park, O. H., Park, J., et al. (2015). Glucocorticoid receptor interacts with PNRC2 in a ligand-dependent manner to recruit UPF1 for rapid mRNA degradation. *Proc. Natl. Acad. Sci. U. S. A.*, 112(13):E1540–9.

Choe, J., Ahn, S. H., and Kim, Y. K. (2014). The mRNP remodeling mediated by UPF1 promotes rapid degradation of replication-dependent histone mRNA. *Nucleic Acids Res.*, 42(14):9334–9349.

Chouaib, R., Safieddine, A., Pichon, X., et al. (2020). A dual Protein-mRNA localization screen reveals compartmentalized translation and widespread co-translational RNA targeting. *Dev. Cell*, 54(6):773–791.e5.

Chuankhayan, P., Hsieh, C.-Y., Huang, Y.-C., et al. (2010). Crystal structures of aspergillus japonicus fructosyltransferase complex with donor/acceptor substrates reveal complete subsites in the active site for catalysis. *J. Biol. Chem.*, 285(30):23251–23264.

Civril, F., Wehenkel, A., Giorgi, F. M., et al. (2010). Structural analysis of the RZZ complex reveals common ancestry with multisubunit vesicle tethering machinery. *Structure*, 18(5):616–626.

Colak, D., Ji, S.-J., Porse, B. T., and Jaffrey, S. R. (2013). Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. *Cell*, 153(6):1252–1265.

Coleman, J. R., Papamichail, D., Skiena, S., et al. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science*, 320(5884):1784–1787.

Cong, L., Ran, F. A., Cox, D., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823.

Conrad, T., Albrecht, A.-S., de Melo Costa, V. R., et al. (2016). Serial interactome capture of the human cell nucleus. *Nat. Commun.*, 7(1):11212.

Cook, K. B., Kazan, H., Zuberi, K., et al. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, 39(Database issue):D301–8.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848.

Cormack, B. P., Valdivia, R. H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). In *Gene*, volume 173, pages 33–38. Elsevier B.V.

Cosson, P. and Letourneur, F. (1997). Coatomer (COPI)-coated vesicles: role in intracellular transport and protein sorting. *Curr. Opin. Cell Biol.*, 9(4):484–487.

Cougot, N., Babajko, S., and Séraphin, B. (2004). Cytoplasmic foci are sites of mRNA decay in human cells. *J. Cell Biol.*, 165(1):31–40.

Crawford, G. E., Faulkner, J. A., Crosbie, R. H., et al. (2000). Assembly of the dystrophin-associated protein complex does not require the dystrophin COOH-terminal domain. *J. Cell Biol.*, 150(6):1399–1410.

Cui, Y., Parashar, S., Zahoor, M., et al. (2019). A COPII subunit acts with an autophagy receptor to target endoplasmic reticulum for degradation. *Science*, 365(6448):53–60.

Dascher, C., Matteson, J., and Balch, W. E. (1994). Syntaxin 5 regulates endoplasmic reticulum to golgi transport. *J. Biol. Chem.*, 269(47):29363–29366.

de Lucas, S., Oliveros, J. C., Chagoyen, M., and Ortín, J. (2014). Functional signature for the recognition of specific target mRNAs by human staufen1 protein. *Nucleic Acids Res.*, 42(7):4516–4526.

de Pretis, S., Kress, T., Morelli, M. J., et al. (2015). INSPEcT: a computational tool to infer mRNA synthesis, processing and degradation dynamics from RNA- and 4su-seq time course experiments. *Bioinformatics*, 31(17):2829–2835.

de Pretis, S., Kress, T. R., Morelli, M. J., et al. (2017). Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. *Genome Res.*, 27(10):1658–1664.

Deal, J., Pleshinger, D. J., Johnson, S. C., et al. (2020). Milestones in the development and implementation of FRET-based sensors of intracellular signals: A biological perspective of the history of FRET. *Cell. Signal.*, 75:109769.

Denic, V., Quan, E. M., and Weissman, J. S. (2006). A luminal surveillance complex that selects misfolded glycoproteins for ER-associated degradation. *Cell*, 126(2):349–359.

Dennis, Jr, G., Sherman, B. T., Hosack, D. A., et al. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.*, 4(5):P3.

Dever, T. E. and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.*, 4(7):a013706.

Devos, D., Dokudovskaya, S., Alber, F., et al. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.*, 2(12):e380.

Dikic, I. (2017). Proteasomal and autophagic degradation systems. *Annu. Rev. Biochem.*, 86:193–224.

Dobin, A., Davis, C. A., Schlesinger, F., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Doench, J. G., Fusi, N., Sullender, M., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, 34(2):184–191.

Dölken, L., Ruzsics, Z., Rädle, B., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*, 14(9):1959–1972.

Dostie, J. and Dreyfuss, G. (2002). Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr. Biol.*, 12(13):1060–1067.

Doudna, J. A. and Charpentier, E. (2014). Genome editing. the new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213):1258096.

Durand, S., Cougot, N., Mahuteau-Betzer, F., et al. (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.*, 178(7):1145–1160.

Durand, S. and Lykke-Andersen, J. (2013). Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat. Struct. Mol. Biol.*, 20(6):702–709.

Durinck, S., Moreau, Y., Kasprzyk, A., et al. (2005). BioMart and bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, 21(16):3439–3440.

Duttke, S. H. C., Lacadie, S. A., Ibrahim, M. M., et al. (2015). Human promoters are intrinsically directional. *Mol. Cell*, 57(4):674–684.

Duttler, S., Pechmann, S., and Frydman, J. (2013). Principles of cotranslational ubiquitination and quality control at the ribosome. *Mol. Cell*, 50(3):379–393.

Eberle, A. B., Lykke-Andersen, S., Mühlemann, O., and Jensen, T. H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.*, 16(1):49–55.

Eden, E., Navon, R., Steinfeld, I., et al. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10:48.

El-Brolosy, M. A., Kontarakis, Z., Rossi, A., et al. (2019). Genetic compensation triggered by mutant mRNA degradation. *Nature*, 568(7751):193–197.

El-Brolosy, M. A. and Stainier, D. Y. R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genet.*, 13(7):e1006780.

Eliscovich, C. and Singer, R. H. (2017). RNP transport in cell biology: the long and winding road. *Curr. Opin. Cell Biol.*, 45:38–46.

Eser, P., Wachutka, L., Maier, K. C., et al. (2016). Determinants of RNA metabolism in the schizosaccharomyces pombe genome. *Mol. Syst. Biol.*, 12(2):857.

Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007). P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol. Cell. Biol.*, 27(11):3970–3981.

Everaert, C., Luypaert, M., Maag, J. L. V., et al. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.*, 7(1):1559.

Fanourgakis, G., Lesche, M., Akpinar, M., et al. (2016). Chromatoid body protein TDRD6 supports long 3' UTR triggered nonsense mediated mRNA decay. *PLoS Genet.*, 12(5):e1005857.

Fazal, F. M., Han, S., Parker, K. R., et al. (2019). Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell*, 178(2):473–490.e26.

Flury, V., Restuccia, U., Bachi, A., and Mühlemann, O. (2014). Characterization of phosphorylation- and RNA-dependent UPF1 interactors by quantitative proteomics. *J. Proteome Res.*, 13(6):3038–3053.

Fournier, D., Palidwor, G. A., Shcherbinin, S., et al. (2013). Functional and genomic analyses of alpha-solenoid proteins. *PLoS One*, 8(11):e79894.

Franks, T. M. and Lykke-Andersen, J. (2007). TTP and BRF proteins nucleate processing body formation to silence mRNAs with AU-rich elements. *Genes Dev.*, 21(6):719–735.

Frattini, A., Fabbri, M., Valli, R., et al. (2015). High variability of genomic instability and gene expression profiling in different HeLa clones. *Sci. Rep.*, 5(1):15377.

Fregno, I. and Molinari, M. (2019). Proteasomal and lysosomal clearance of faulty secretory proteins: ER-associated degradation (ERAD) and ER-to-lysosome-associated degradation (ERLAD) pathways. *Crit. Rev. Biochem. Mol. Biol.*, 54(2):153–163.

Friedel, C. C. and Dölken, L. (2009). Metabolic tagging and purification of nascent RNA: Implications for transcriptomics.

Frühwald, M. C., O'Dorisio, M. S., Rush, L. J., et al. (2000). Gene amplification in PNETs/medulloblastomas: mapping of a novel amplified gene within the MYCN amplicon. *J. Med. Genet.*, 37(7):501–509.

Fu, X., Fu, N., Guo, S., et al. (2009). Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics*, 10(1):161.

Fuchs, G., Voichek, Y., Benjamin, S., et al. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.*, 15(5):R69.

Fujita, K., Omura, S., and Silver, J. (1997). Rapid degradation of CD4 in cells expressing human immunodeficiency virus type 1 env and vpu is blocked by proteasome inhibitors. *J. Gen. Virol.*, 78 (Pt 3):619–625.

Fumagalli, F., Noack, J., Bergmann, T. J., et al. (2016). Translocon component sec62 acts in endoplasmic reticulum turnover during stress recovery. *Nat. Cell Biol.*, 18(11):1173–1184.

Fung, T. S. and Liu, D. X. (2014). Coronavirus infection, ER stress, apoptosis and innate immunity. *Front. Microbiol.*, 5:296.

Furlan, M., Galeota, E., Gaudio, N. D., et al. (2020). Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling. *Genome Res.*, 30(10):1492–1507.

Gardner, L. B. (2008). Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. *Mol. Cell. Biol.*, 28(11):3729–3741.

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

Gatica, L. V. and Rosa, A. L. (2016). A complex interplay of genetic and epigenetic events leads to abnormal expression of the DUX4 gene in facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.*, 26(12):844–852.

Gebauer, F. and Hentze, M. W. (2004). Molecular mechanisms of translational control.

Gehring, N. H., Lamprinaki, S., Kulozik, A. E., and Hentze, M. W. (2009). Disassembly of exon junction complexes by PYM. *Cell*, 137(3):536–548.

Gentleman, R. C., Carey, V. J., Bates, D. M., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, 5(10):R80.

Ghanbarian, H., Wagner, N., Michiels, J.-F., et al. (2017). Small RNA-directed epigenetic programming of embryonic stem cell cardiac differentiation. *Sci. Rep.*, 7:41799.

Giampietri, C., Petrungaro, S., Conti, S., et al. (2015). Cancer microenvironment and endoplasmic reticulum stress response.

Giorgi, C., Yeo, G. W., Stone, M. E., et al. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell*, 130(1):179–191.

Goetz, A. E. and Wilkinson, M. (2017). Stress and the nonsense-mediated RNA decay pathway. *Cell. Mol. Life Sci.*, 74(19):3509–3531.

Gomez-Navarro, N. and Miller, E. A. (2016). COP-coated vesicles. *Curr. Biol.*, 26(2):R54–R57.

Grantham, R., Gautier, C., and Gouy, M. (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.*, 8(9):1893–1912.

Gray, J. M., Harmin, D. A., Boswell, S. A., et al. (2014). SnapShot-Seq: a method for extracting genome-wide, in vivo mRNA dynamics from a single total RNA sample. *PLoS One*, 9(2):e89673.

Grumati, P. and Dikic, I. (2018). Ubiquitin signaling and autophagy. *J. Biol. Chem.*, 293(15):5404–5413.

Grumati, P., Morozzi, G., Hölper, S., et al. (2017). Full length RTN3 regulates turnover of tubular endoplasmic reticulum via selective autophagy. *Elife*, 6.

Haack, T. B., Staufner, C., Köpke, M. G., et al. (2015). Biallelic mutations in NBAS cause recurrent acute liver failure with onset in infancy. *Am. J. Hum. Genet.*, 97(1):163–169.

Haimovich, G., Medina, D. A., Causse, S. Z., et al. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell*, 153(5):1000–1011.

Hall, G. W. and 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*, 83(8):2031–2037.

Harel-Sharvit, L., Eldad, N., Haimovich, G., et al. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. *Cell*, 143(4):552–563.

Hartmann, E., Sommer, T., Prehn, S., et al. (1994). Evolutionary conservation of components of the protein translocation complex. *Nature*, 367(6464):654–657.

Hatsuzawa, K., Hirose, H., Tani, K., et al. (2000). Syntaxin 18, a SNAP receptor that functions in the endoplasmic reticulum, intermediate compartment, and cis-golgi vesicle trafficking. *J. Biol. Chem.*, 275(18):13713–13720.

Hauer, C., Sieber, J., Schwarzl, T., et al. (2016). Exon junction complexes show a distributional bias toward alternatively spliced mRNAs and against mRNAs coding for ribosomal proteins. *Cell Rep.*, 16(6):1588–1603.

He, F. and Jacobson, A. (2015). Nonsense-Mediated mRNA decay: Degradation of defective transcripts is only part of the story. *Annu. Rev. Genet.*, 49:339–366.

He, F., Li, X., Spatrick, P., et al. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell*, 12(6):1439–1452.

Heard, E. and Disteche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.*, 20(14):1848–1867.

Hebert, M. D. and Poole, A. R. (2017). Towards an understanding of regulating cajal body activity by protein modification.

Hernan, R., Heuermann, K., and Brizzard, B. (2000). Multiple epitope tagging of expressed proteins for enhanced detection. *Biotechniques*, 28(4):789–793.

Hetz, C. and Glimcher, L. H. (2011). Protein homeostasis networks in physiology and disease.

Hetz, C., Martinon, F., Rodriguez, D., and Glimcher, L. H. (2011). The unfolded protein response: integrating stress signals through the stress sensor IRE1α. *Physiol. Rev.*, 91(4):1219–1243.

Hirose, H., Arasaki, K., Dohmae, N., et al. (2004). Implication of ZW10 in membrane trafficking between the endoplasmic reticulum and golgi. *EMBO J.*, 23(6):1267–1278.

Hoek, T. A., Khuperkar, D., Lindeboom, R. G. H., et al. (2019). Single-Molecule imaging uncovers rules governing Nonsense-Mediated mRNA decay. *Mol. Cell*, 75(2):324–339.e11.

Hoffman, H. K., Fernandez, M. V., Groves, N. S., et al. (2019). Genomic tagging of endogenous human ESCRT-I complex preserves ESCRT-mediated membrane-remodeling functions. *J. Biol. Chem.*, 294(44):16266–16281.

Hogg, J. R. and Goff, S. P. (2010). Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell*, 143(3):379–389.

Holbrook, J. A., Neu-Yilik, G., Hentze, M. W., and Kulozik, A. E. (2004). Nonsense-mediated decay approaches the clinic. *Nat. Genet.*, 36(8):801–808.

Hollien, J. and Weissman, J. S. (2006). Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science*, 313(5783):104–107.

Hong, W. (2005). SNAREs and traffic. *Biochim. Biophys. Acta*, 1744(3):493–517.

Hsu, P. D., Lander, E. S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering.

Hsu, P. D., Scott, D. A., Weinstein, J. A., et al. (2013). DNA targeting specificity of RNA-guided cas9 nucleases. *Nat. Biotechnol.*, 31(9):827–832.

Huang, H., Paliouras, M., Rambaldi, I., et al. (2003). Nonmuscle myosin promotes cytoplasmic localization of PBX. *Mol. Cell. Biol.*, 23(10):3636–3645.

Huang, L., Lou, C.-H., Chan, W., et al. (2011). RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol. Cell*, 43(6):950–961.

Huang, W., Wong, J. M., and Bateman, E. (1996). TATA elements direct bi-directional transcription by RNA polymerases II and III. *Nucleic Acids Res.*, 24(6):1158–1163.

Huntzinger, E., Kashima, I., Fauser, M., et al. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA*, 14(12):2609–2617.

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

Hüttelmaier, S., Zenklusen, D., Lederer, M., et al. (2005). Spatial regulation of beta-actin translation by src-dependent phosphorylation of ZBP1. *Nature*, 438(7067):512–515.

Huyer, G., Piluek, W. F., Fansler, Z., et al. (2004). Distinct machinery is required in saccharomyces cerevisiae for the endoplasmic reticulum-associated degradation of a multispanning membrane protein and a soluble luminal protein. *J. Biol. Chem.*, 279(37):38369–38378.

Iinuma, T., Aoki, T., Arasaki, K., et al. (2009). Role of syntaxin 18 in the organization of endoplasmic reticulum subdomains. *J. Cell Sci.*, 122(Pt 10):1680–1690.

Imamachi, N., Tani, H., Mizutani, R., et al. (2014). BRIC-seq: a genome-wide approach for determining RNA stability in mammalian cells. *Methods*, 67(1):55–63.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945.

Ishigaki, Y., Li, X., Serin, G., and Maquat, L. E. (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*, 106(5):607–617.

Isken, O., Kim, Y. K., Hosoda, N., et al. (2008). Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. *Cell*, 133(2):314–327.

Isken, O. and Maquat, L. E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, 21(15):1833–1856.

Isken, O. and Maquat, L. E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat. Rev. Genet.*, 9(9):699–712.

Jacobson, A. and Peltz, S. W. (1996). Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells. *Annu. Rev. Biochem.*, 65:693–739.

Jagannathan, S., Nwosu, C., and Nicchitta, C. V. (2011). Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation. *Methods Mol. Biol.*, 714:301–321.

Jahn, R. and Scheller, R. H. (2006). SNAREs–engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.*, 7(9):631–643.

Jan, C. H., Williams, C. C., and Weissman, J. S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science*, 346(6210):1257521.

Jeong, K., Ryu, I., Park, J., et al. (2019). Staufen1 and UPF1 exert opposite actions on the replacement of the nuclear cap-binding complex by eIF4E at the 5' end of mRNAs. *Nucleic Acids Res.*, 47(17):9313–9328.

Jheng, J.-R., Ho, J.-Y., and Horng, J.-T. (2014). ER stress, autophagy, and RNA viruses. *Front. Microbiol.*, 5:388.

Jia, J., Werkmeister, E., Gonzalez-Hilarion, S., et al. (2017). Premature termination codon readthrough in human cells occurs in novel cytoplasmic foci and requires UPF proteins. *J. Cell Sci.*, 130(18):3009–3022.

Johnson, J. L., Stoica, L., Liu, Y., et al. (2019). Inhibition of Upf2-Dependent Nonsense-Mediated decay leads to behavioral and neurophysiological abnormalities by activating the immune response. *Neuron*, 104(4):665–679.e8.

Johnson, T. R., Rudin, S. D., Blossey, B. K., et al. (1991). Newly synthesized RNA: Simultaneous measurement in intact cells of transcription rates and RNA stability of insulin-like growth factor i, actin, and albumin in growth hormone-stimulated hepatocytes (4-thiourlidine/cultured hepatocytes/mercurated agarose affinity chromatography). *Biochemistry*, 88:5287–5291.

Jolly, L. A., Homan, C. C., Jacob, R., et al. (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.*, 22(23):4673–4687.

Jung, H., Gkogkas, C. G., Sonenberg, N., and Holt, C. E. (2014). Remote control of gene function by local translation. *Cell*, 157(1):26–40.

Jung, H., Lee, D., Lee, J., et al. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.*, 47(11):1242–1248.

Jung, H., Yoon, B. C., and Holt, C. E. (2012). Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat. Rev. Neurosci.*, 13(5):308–324.

Kalies, K.-U., Allan, S., Sergeyenko, T., et al. (2005). The protein translocation channel binds proteasomes to the endoplasmic reticulum membrane. *EMBO J.*, 24(13):2284–2293.

Karagöz, G. E., Acosta-Alvear, D., and Walter, P. (2019). The unfolded protein response: Detecting and responding to fluctuations in the Protein-Folding capacity of the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.*, 11(9).

Karam, R., Lou, C.-H., Kroeger, H., et al. (2015). The unfolded protein response is shaped by the NMD pathway. *EMBO Rep.*, 16(5):599–609.

Karam, R., Wengrod, J., Gardner, L. B., and Wilkinson, M. F. (2013). Regulation of nonsense-mediated mRNA decay: Implications for physiology and disease.

Karamyshev, A. L., Patrick, A. E., Karamysheva, Z. N., et al. (2014). Inefficient SRP interaction with a nascent chain triggers a mRNA quality control pathway. *Cell*, 156(1-2):146–157.

Karousis, E. D., Gurzeler, L.-A., Annibaldis, G., et al. (2020). Human NMD ensues independently of stable ribosome stalling. *Nat. Commun.*, 11(1):4134.

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

Kashima, I., Yamashita, A., Izumi, N., et al. (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.*, 20(3):355–367.

Katahira, J. (2015). Nuclear export of messenger RNA. *Genes*, 6(2):163–184.

Kawashima, T., Douglass, S., Gabunilas, J., et al. (2014). Widespread use of non-productive alternative splice sites in saccharomyces cerevisiae. *PLoS Genet.*, 10(4):e1004249.

Kaygun, H. and Marzluff, W. F. (2005). Regulated degradation of replication-dependent histone mRNAs requires both ATR and upf1. *Nat. Struct. Mol. Biol.*, 12(9):794–800.

Kebaara, B. W. and Atkin, A. L. (2009). Long 3'-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in saccharomyces cerevisiae. *Nucleic Acids Res.*, 37(9):2771–2778.

Kedersha, N., Stoecklin, G., Ayodele, M., et al. (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.*, 169(6):871–884.

Kerr, T. P., Sewry, C. A., Robb, S. A., and Roberts, R. G. (2001). Long mutant dystrophins and variable phenotypes: evasion of nonsense-mediated decay? *Hum. Genet.*, 109(4):402–407.

Khaminets, A., Heinrich, T., Mari, M., et al. (2015). Regulation of endoplasmic reticulum turnover by selective autophagy. *Nature*, 522(7556):354–358.

Kim, Y. K., Furic, L., Desgroseillers, L., and Maquat, L. E. (2005). Mammalian staufen1 recruits upf1 to specific mRNA 3'UTRs so as to elicit mRNA decay. *Cell*, 120(2):195–208.

Kober, L., Zehe, C., and Bode, J. (2013). Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol. Bioeng.*, 110(4):1164–1173.

Koch, B., Nijmeijer, B., Kueblbeck, M., et al. (2018). Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. *Nat. Protoc.*, 13(6):1465–1487.

Köhler, A. and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm.

Kok, F. O., Shin, M., Ni, C.-W., et al. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev. Cell*, 32(1):97–108.

Konieczny, P., Lichawska-Cieslar, A., Kwiecinska, P., et al. (2019). Keratinocyte-specific ablation of mcpip1 impairs skin integrity and promotes local and systemic inflammation. *J. Mol. Med.*, 97(12):1669–1684.

Kotaja, N., Bhattacharyya, S. N., Jaskiewicz, L., et al. (2006). The chromatoid body of male germ cells: similarity with processing bodies and presence of dicer and microRNA pathway components. *Proc. Natl. Acad. Sci. U. S. A.*, 103(8):2647–2652.

Kugler, W., Enssle, J., Hentze, M. W., and Kulozik, A. E. (1995). Nuclear degradation of nonsense mutated beta-globin mRNA: a post-transcriptional mechanism to protect heterozygotes from severe clinical manifestations of beta-thalassemia? *Nucleic Acids Res.*, 23(3):413–418.

Kumar, P., Nagarajan, A., and Uchil, P. D. (2019). Transfection of mammalian cells with calcium Phosphate-DNA coprecipitates. *Cold Spring Harb. Protoc.*, 2019(10).

Kuo, T. C. Y., Hatakeyama, M., Tameshige, T., et al. (2020). Homeolog expression quantification methods for allopolyploids. *Brief. Bioinform.*, 21(2):395–407.

Kurosaki, T. and Maquat, L. E. (2016). Nonsense-mediated mRNA decay in humans at a glance. *J. Cell Sci.*, 129(3):461–467.

Kurosaki, T., Miyoshi, K., Myers, J. R., and Maquat, L. E. (2018). NMD-degradome sequencing reveals ribosome-bound intermediates with 3'-end non-templated nucleotides. *Nat. Struct. Mol. Biol.*, 25(10):940–950.

Kurosaki, T., Myers, J. R., and Maquat, L. E. (2019a). Defining nonsense-mediated mRNA decay intermediates in human cells. *Methods*, 155:68–76.

Kurosaki, T., Popp, M. W., and Maquat, L. E. (2019b). Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nature Reviews Molecular Cell Biology*, 20(7):406–420.

La Manno, G., Soldatov, R., Zeisel, A., et al. (2018). RNA velocity of single cells. *Nature*, 560(7719):494–498.

Lacassie, Y., Johnson, B., Lay-Son, G., et al. (2020). Severe SOPH syndrome due to a novel NBAS mutation in a 27-year-old woman-review of this pleiotropic, autosomal recessive disorder: Mystery solved after two decades. *Am. J. Med. Genet. A*, 182(7):1767–1775.

Lai, M.-C. and Tarn, W.-Y. (2004). Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J. Biol. Chem.*, 279(30):31745–31749.

Lai, T., Cho, H., Liu, Z., et al. (2012). Structural basis of the PNRC2-mediated link between mrna surveillance and decapping. *Structure*, 20(12):2025–2037.

Lander, E. S., Linton, L. M., Birren, B., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Landry, J. J. M., Pyl, P. T., Rausch, T., et al. (2013). The genomic and transcriptomic landscape of a HeLa cell line. *G3*, 3(8):1213–1224.

Landschulz, W. H., Johnson, P. F., and McKnight, S. L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, 240(4860):1759–1764.

Lappalainen, T., Sammeth, M., Friedländer, M. R., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468):506–511.

Lareau, L. F., Inada, M., Green, R. E., et al. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, 446(7138):926–929.

Lawrence, J. B. and Singer, R. H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell*, 45(3):407–415.

Le Hir, H., Izaurralde, E., Maquat, L. E., and Moore, M. J. (2000). The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.*, 19(24):6860–6869.

Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.*, 17(1):41–54.

Leeds, P., Peltz, S. W., Jacobson, A., and Culbertson, M. R. (1991). The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.*, 5(12A):2303–2314.

Lefrancois, S., Zeng, J., Hassan, A. J., et al. (2003). The lysosomal trafficking of sphingolipid activator proteins (SAPs) is mediated by sortilin. *EMBO J.*, 22(24):6430–6437.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L. E. (2002). The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.*, 21(13):3536–3545.

Leksa, N. C., Brohawn, S. G., and Schwartz, T. U. (2009). The structure of the scaffold nucleoporin nup120 reveals a new and unexpected domain architecture. *Structure*, 17(8):1082–1091.

Lelivelt, M. J. and Culbertson, M. R. (1999). Yeast upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol. Cell. Biol.*, 19(10):6710–6719.

Letourneur, F. and Cosson, P. (1998). Targeting to the endoplasmic reticulum in yeast cells by determinants present in transmembrane domains. *J. Biol. Chem.*, 273(50):33273–33278.

Lewis, K. M. and Ke, A. (2017). Building the class 2 CRISPR-Cas arsenal. *Mol. Cell*, 65(3):377–379.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, J., Snyder, E. Y., Tang, F. H. F., et al. (2020). Nna1 gene deficiency triggers purkinje neuron death by tubulin hyperglutamylation and ER dysfunction. *JCI Insight*, 5(19).

Li, L., Geng, Y., Feng, R., et al. (2017a). The human RNA surveillance factor UPF1 modulates gastric cancer progression by targeting long Non-Coding RNA MALAT1. *Cell. Physiol. Biochem.*, 42(6):2194–2206.

Li, S., Prasanna, X., Salo, V. T., et al. (2019). An efficient auxin-inducible degron system with low basal degradation in human cells. *Nat. Methods*, 16(9):866–869.

Li, W., Zhu, J., Dou, J., et al. (2017b). Phosphorylation of LAMP2A by p38 MAPK couples ER stress to chaperone-mediated autophagy. *Nat. Commun.*, 8(1):1763.

Li, Z., Vuong, J. K., Zhang, M., et al. (2017c). Inhibition of nonsense-mediated RNA decay by ER stress. *RNA*, 23(3):378–394.

Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife*, 3:e04766.

Lindeboom, R. G. H., Supek, F., and Lehner, B. (2016). The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.*, 48(10):1112–1118.

Liput, D. J., Nguyen, T. A., Augustin, S. M., et al. (2020). A guide to fluorescence lifetime microscopy and förster's resonance energy transfer in neuroscience. *Curr. Protoc. Neurosci.*, 94(1):e108.

Liu, H. Y., Badarinarayana, V., Audino, D. C., et al. (1998). The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively. *EMBO J.*, 17(4):1096–1106.

Liu, X., Shi, X., Chen, C., and Zhang, L. (2015). Improving RNA-Seq expression estimation by modeling isoform- and exon-specific read sequencing rate. *BMC Bioinformatics*, 16:332.

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

Loi, M., Fregno, I., Guerra, C., and Molinari, M. (2018). Eat it right: Er-phagy and recover-phagy.

Longman, D., Hug, N., Keith, M., et al. (2013). DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and caenorhabditis elegans. *Nucleic Acids Res.*, 41(17):8319–8331.

Longman, D., Jackson-Jones, K. A., Maslon, M. M., et al. (2020). Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum. *Genes Dev.*, 34(15-16):1075–1088.

Longman, D., Plasterk, R. H. A., Johnstone, I. L., and Cáceres, J. F. (2007). Mechanistic insights and identification of two novel factors in the c. elegans NMD pathway. *Genes Dev.*, 21(9):1075–1085.

Losson, R. and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. *Proc. Natl. Acad. Sci. U. S. A.*, 76(10):5134–5137.

Lotan, R., Bar-On, V. G., Harel-Sharvit, L., et al. (2005). The RNA polymerase II subunit rpb4p mediates decay of a specific class of mRNAs. *Genes Dev.*, 19(24):3004–3016.

Lotan, R., Goler-Baron, V., Duek, L., et al. (2007). The rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J. Cell Biol.*, 178(7):1133–1143.

Lou, C.-H., Dumdie, J., Goetz, A., et al. (2016). Nonsense-Mediated RNA decay influences human embryonic stem cell fate. *Stem Cell Reports*, 6(6):844–857.

Lou, C. H., Shao, A., Shum, E. Y., et al. (2014). Posttranscriptional control of the stem cell and neurogenic programs by the nonsense-mediated RNA decay pathway. *Cell Rep.*, 6(4):748–764.

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

Luo, S., Mao, C., Lee, B., and Lee, A. S. (2006). GRP78/BiP is required for cell proliferation and protecting the inner cell mass from apoptosis during early mouse embryonic development. *Mol. Cell. Biol.*, 26(15):5688–5697.

Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2000). Human upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell*, 103(7):1121–1131.

Lykke-Andersen, S., Chen, Y., Ardal, B. R., et al. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.*, 28(22):2498–2517.

Lykke-Andersen, S. and Jensen, T. H. (2015). Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes.

Ma, W. and Goldberg, J. (2013). Rules for the recognition of dilysine retrieval motifs by coatomer. *EMBO J.*, 32(7):926–937.

Ma, Z., Zhu, P., Shi, H., et al. (2019). PTC-bearing mRNA elicits a genetic compensation response via upf3a and COMPASS components. *Nature*, 568(7751):259–263.

Macville, M., Schröck, E., Padilla-Nash, H., et al. (1999). Comprehensive and definitive molecular cytogenetic characterization of HeLa cells by spectral karyotyping. *Cancer Res.*, 59(1):141–150.

Maksimova, N., Hara, K., Nikolaeva, I., et al. (2010). Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huët anomaly. *J. Med. Genet.*, 47(8):538–548.

Maquat, L. E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.*, 5(2):89–99.

Maquat, L. E., Hwang, J., Sato, H., and Tang, Y. (2010a). CBP80-promoted mRNP rearrangements during the pioneer round of translation, nonsense-mediated mRNA decay, and thereafter. *Cold Spring Harb. Symp. Quant. Biol.*, 75:127–134.

Maquat, L. E., Kinniburgh, A. J., Rachmilewitz, E. A., and Ross, J. (1981). Unstable beta-globin mRNA in mRNA-deficient beta o thalassemia. *Cell*, 27(3 Pt 2):543–553.

Maquat, L. E., Tarn, W. Y., and Isken, O. (2010b). The pioneer round of translation: Features and functions.

Marzi, M. J., Ghini, F., Cerruti, B., et al. (2016). Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res.*, 26(4):554–565.

Matthaei, J. H., Jones, O. W., Martin, R. G., and Nirenberg, M. W. (1962). Characteristics and composition of RNA coding units. *Proc. Natl. Acad. Sci. U. S. A.*, 48(4):666–677.

McDonald, C. M., Campbell, C., Torricelli, R. E., et al. (2017). Ataluren in patients with nonsense mutation duchenne muscular dystrophy (ACT DMD): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet*, 390(10101):1489–1498.

McLaren, W., Gil, L., Hunt, S. E., et al. (2016). The ensembl variant effect predictor. *Genome Biol.*, 17(1):122.

Mehnert, M., Sommer, T., and Jarosch, E. (2014). Der1 promotes movement of misfolded proteins through the endoplasmic reticulum membrane. *Nat. Cell Biol.*, 16(1):77–86.

Meikar, O., Vagin, V. V., Chalmel, F., et al. (2014). An atlas of chromatoid body components. *RNA*, 20(4):483–495.

Mendell, J. T., ap Rhys, C. M. J., and Dietz, H. C. (2002). Separable roles for rent1/hupf1 in altered splicing and decay of nonsense transcripts. *Science*, 298(5592):419–422.

Mendell, J. T., Sharifi, N. A., Meyers, J. L., et al. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.*, 36(10):1073–1078.

Metzstein, M. M. and Krasnow, M. A. (2006). Functions of the nonsense-mediated mRNA decay pathway in drosophila development. *PLoS Genet.*, 2(12):e180.

Meyer, K., Köster, T., Nolte, C., et al. (2017). Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. *Genome Biol.*, 18(1):204.

Michel, M., Demel, C., Zacher, B., et al. (2017). TT-seq captures enhancer landscapes immediately after t-cell stimulation. *Mol. Syst. Biol.*, 13(3):920.

Miller, C., Schwalb, B., Maier, K., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.*, 7(1):458.

Mitchell, D. M., Zhou, M., Pariyarath, R., et al. (1998). Apoprotein B100 has a prolonged interaction with the translocon during which its lipidation and translocation change from dependence on the microsomal triglyceride transfer protein to independence. *Proc. Natl. Acad. Sci. U. S. A.*, 95(25):14733–14738.

Mitrovich, Q. M. and Anderson, P. (2005). mRNA surveillance of expressed pseudogenes in c. elegans. *Curr. Biol.*, 15(10):963–967.

Mittelman, D. and Wilson, J. H. (2013). The fractured genome of HeLa cells. *Genome Biol.*, 14(4):111.

Moore, M. J. and Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell*, 136(4):688–700.

Moreno, A. B., Martínez de Alba, A. E., Bardou, F., et al. (2013). Cytoplasmic and nuclear quality control and turnover of single-stranded RNA modulate post-transcriptional gene silencing in plants. *Nucleic Acids Res.*, 41(8):4699–4708.

Morisaki, T., Lyon, K., DeLuca, K. F., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. *Science*, 352(6292):1425–1429.

Mort, M., Ivanov, D., Cooper, D. N., and Chuzhanova, N. A. (2008). A meta-analysis of nonsense mutations causing human genetic disease. *Hum. Mutat.*, 29(8):1037–1047.

Muchowski, P. J., Schaffar, G., Sittler, A., et al. (2000). Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl. Acad. Sci. U. S. A.*, 97(14):7841–7846.

Muhlrad, D. and Parker, R. (1999). Aberrant mRNAs with extended 3' UTRs are substrates for rapid degradation by mRNA surveillance. *RNA*, 5(10):1299–1307.

Muir, V. S., Gasch, A. P., and Anderson, P. (2018). The substrates of Nonsense-Mediated mRNA decay in caenorhabditis elegans. *G3*, 8(1):195–205.

Mukherjee, A. S. and Beermann, W. (1965). Synthesis of ribonucleic acid by the x-chromosomes of drosophila melanogaster and the problem of dosage compensation. *Nature*, 207(998):785–786.

Muller and J, H. (1932). Further studies on the nature and causes of gene mutations. *Proc. Sixth Int. Cong. Genet., Ithaca, New York, USA*, 1:213–255.

Müller, D., Lasfargues, C., El Khawand, S., et al. (2013). 4E-BP restrains eIF4E phosphorylation. *Translation (Austin)*, 1(2):e25819.

Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.*, 84:165–198.

Nagy, E. and Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.*, 23(6):198–199.

Nakajima, K.-I., Hirose, H., Taniguchi, M., et al. (2004). Involvement of BNIP1 in apoptosis and endoplasmic reticulum membrane fusion. *EMBO J.*, 23(16):3216–3226.

Nazarenus, T., Cedarberg, R., Bell, R., et al. (2005). Upf1p, a highly conserved protein required for nonsense-mediated mRNA decay, interacts with the nuclear pore proteins nup100p and nup116p. *Gene*, 345(2):199–212.

Neu-Yilik, G. and Kulozik, A. E. (2004). mRNA metabolism and hereditary disorders: a tale of surveillance and escape. *Klin. Padiatr.*, 216(6):304–314.

Nguyen, L. S., Kim, H.-G., Rosenfeld, J. A., et al. (2013). Contribution of copy number variants involving nonsense-mediated mRNA decay pathway genes to neurodevelopmental disorders. *Hum. Mol. Genet.*, 22(9):1816–1825.

Nguyen, L. S., Wilkinson, M. F., and Gecz, J. (2014). Nonsense-mediated mRNA decay: Inter-individual variability and human disease.

Ni, J. Z., Grate, L., Donohue, J. P., et al. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.*, 21(6):708–718.

Nielsen, M. S., Jacobsen, C., Olivecrona, G., et al. (1999). Sortilin/neurotensin receptor-3 binds and mediates degradation of lipoprotein lipase. *J. Biol. Chem.*, 274(13):8832–8836.

Nielsen, M. S., Madsen, P., Christensen, E. I., et al. (2001). The sortilin cytoplasmic tail conveys golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.*, 20(9):2180–2190.

Niklas, J., Melnyk, A., Yuan, Y., and Heinzle, E. (2011). Selective permeabilization for the high-throughput measurement of compartmented enzyme activities in mammalian cells. *Anal. Biochem.*, 416(2):218–227.

Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., et al. (2001). The 'arctic' APP mutation (E693G) causes alzheimer's disease by enhanced abeta protofibril formation. *Nat. Neurosci.*, 4(9):887–893.

Nirenberg, M., Leder, P., Bernfield, M., et al. (1965). RNA codewords and protein synthesis, VII. on the general nature of the RNA code. *Proc. Natl. Acad. Sci. U. S. A.*, 53(5):1161–1168.

Nyikó, T., Sonkoly, B., Mérai, Z., et al. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol. Biol.*, 71(4-5):367–378.

Ohashi, T., Hegi, S., Fukunaga, T., et al. (2020). Golgi localization of glycosyltransferases requires gpp74p in schizosaccharomyces pombe. *Appl. Microbiol. Biotechnol.*, 104(20):8897–8909.

Orphanides, G. and Reinberg, D. (2002). A unified theory of gene expression.

Oyadomari, S., Koizumi, A., Takeda, K., et al. (2002). Targeted disruption of the chop gene delays endoplasmic reticulum stress-mediated diabetes. *J. Clin. Invest.*, 109(4):525–532.

Palma, M. and Lejeune, F. (2020). Deciphering the molecular mechanism of stop codon readthrough. *Biol. Rev. Camb. Philos. Soc.*

Paquin, N., Ménade, M., Poirier, G., et al. (2007). Local activation of yeast ASH1 mRNA translation through phosphorylation of khd1p by the casein kinase yck1p. *Mol. Cell*, 26(6):795–809.

Park, E. and Maquat, L. E. (2013). Staufen-mediated mRNA decay. *Wiley Interdiscip. Rev. RNA*, 4(4):423–435.

Partridge, K. A., Johannessen, A., Tauler, A., et al. (1999). Competition between the signal sequence and a 3'UTR localisation signal during redirection of beta-globin mRNA to the endoplasmic reticulum: implications for biotechnology. *Cytotechnology*, 30(1-3):37–47.

Paule, M. R. and White, R. J. (2000). Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res.*, 28(6):1283–1298.

Peccarelli, M. and Kebaara, B. W. (2014). Regulation of natural mRNAs by the nonsense-mediated mRNA decay pathway. *Eukaryot. Cell*, 13(9):1126–1135.

Peltz, S. W., Brown, A. H., and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes Dev.*, 7(9):1737–1754.

Pérez-Ortín, J. E., Alepuz, P., Chávez, S., and Choder, M. (2013). Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. *J. Mol. Biol.*, 425(20):3750–3775.

Peters, D. K., Erickson, K. D., and Garcea, R. L. (2020). Live cell microscopy of murine polyomavirus subnuclear replication centers. *Viruses*, 12(10):1123.

Pichon, X., Bastide, A., Safieddine, A., et al. (2016). Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J. Cell Biol.*, 214(6):769–781.

Picotti, P., Aebersold, R., and Domon, B. (2007). The implications of proteolytic background for shotgun proteomics. *Mol. Cell. Proteomics*, 6(9):1589–1598.

Plotkin, J. B. and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1):32–42.

Potelle, S., Klein, A., and Foulquier, F. (2015). Golgi post-translational modifications and associated diseases. *J. Inherit. Metab. Dis.*, 38(4):741–751.

Preissler, S. and Ron, D. (2019). Early events in the endoplasmic reticulum unfolded protein response. *Cold Spring Harb. Perspect. Biol.*, 11(4).

Quax, T. E. F., Claassens, N. J., Söll, D., and van der Oost, J. (2015). Codon bias as a means to Fine-Tune gene expression. *Mol. Cell*, 59(2):149–161.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842.

Quinodoz, S. and Guttman, M. (2014). Long noncoding RNAs: An emerging link between gene regulation and nuclear organization.

Raaben, M., Groot Koerkamp, M. J. A., Rottier, P. J. M., and de Haan, C. A. M. (2007). Mouse hepatitis coronavirus replication induces host translational shutoff and mRNA decay, with concomitant formation of stress granules and processing bodies. *Cell. Microbiol.*, 9(9):2218–2229.

Rabani, M., Levin, J. Z., Fan, L., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.*, 29(5):436–442.

Rabani, M., Raychowdhury, R., Jovanovic, M., et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell*, 159(7):1698–1710.

Raj, A., Rifkin, S. A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature*, 463(7283):913–918.

Ramani, A. K., Nelson, A. C., Kapranov, P., et al. (2009). High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective caenorhabditis elegans. *Genome Biol.*, 10(9):R101.

Ran, F. A., Hsu, P. D., Lin, C.-Y., et al. (2013a). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, 154(6):1380–1389.

Ran, F. A., Hsu, P. D., Wright, J., et al. (2013b). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, 8(11):2281–2308.

Rao, M. S., Van Vleet, T. R., Ciurlionis, R., et al. (2018). Comparison of RNA-Seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from Short-Term rat toxicity studies. *Front. Genet.*, 9(JAN):636.

Rapoport, T. A., Li, L., and Park, E. (2017). Structural and mechanistic insights into protein translocation. *Annu. Rev. Cell Dev. Biol.*, 33(1):369–390.

Rehwinkel, J., Letunic, I., Raes, J., et al. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. *RNA*, 11(10):1530–1544.

Rehwinkel, J., Raes, J., and Izaurralde, E. (2006). Nonsense-mediated mRNA decay: target genes and functional diversification of effectors.

Reid, D. W. and Nicchitta, C. V. (2012). Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *J. Biol. Chem.*, 287(8):5518–5527.

Ricci, E. P., Kucukural, A., Cenik, C., et al. (2014). Staufen1 senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol.*, 21(1):26–35.

Richter, J. D. and Coller, J. (2015). Pausing on polyribosomes: Make way for elongation in translational control. *Cell*, 163(2):292–300.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.*, 29(1):24–26.

Robu, M. E., Larson, J. D., Nasevicius, A., et al. (2007). p53 activation by knockdown technologies. *PLoS Genet.*, 3(5):e78.

Rodvold, J. J., Mahadevan, N. R., and Zanetti, M. (2016). Immune modulation by ER stress and inflammation in the tumor microenvironment.

Rossi, A., Kontarakis, Z., Gerri, C., et al. (2015). Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature*, 524(7564):230–233.

Roth, J. and Zuber, C. (2017). Quality control of glycoprotein folding and ERAD: the role of n-glycan handling, EDEM1 and OS-9. *Histochem. Cell Biol.*, 147(2):269–284.

Roux, K. J., Kim, D. I., and Burke, B. (2013). BioID: a screen for protein-protein interactions. *Curr. Protoc. Protein Sci.*, 74(1):19.23.1–19.23.14.

Roux, K. J., Kim, D. I., Raida, M., and Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, 196(6):801–810.

Rowe, S. M., Clancy, J. P., and Wilschanski, M. (2011). Nasal potential difference measurements to assess CFTR ion channel activity. *Methods Mol. Biol.*, 741:69–86.

Roy, B. and Griggs, R. (2021). Advances in treatments in muscular dystrophies and motor neuron disorders. *Neurol. Clin.*, 39(1):87–112.

Rufener, S. C. and Mühlemann, O. (2013). eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat. Struct. Mol. Biol.*, 20(6):710–717.

Ruiz-Echevarría, M. J. and Peltz, S. W. (2000). The RNA binding protein pub1 modulates the stability of transcripts containing upstream open reading frames. *Cell*, 101(7):741–751.

Russell, J. and Zomerdijk, J. C. B. M. (2006). The RNA polymerase I transcription machinery. *Biochem. Soc. Symp.*, 73(73):203–216.

Russo, F. and Angelini, C. (2014). RNASeqGUI: a GUI for analysing RNA-Seq data. *Bioinformatics*, 30(17):2514–2516.

Russo, J., Heck, A. M., Wilusz, J., and Wilusz, C. J. (2017). Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. *Methods*, 120:39–48.

Sakaki, K., Yoshina, S., Shen, X., et al. (2012). RNA surveillance is required for endoplasmic reticulum homeostasis. *Proc. Natl. Acad. Sci. U. S. A.*, 109(21):8079–8084.

Salama, N. R., Chuang, J. S., and Schekman, R. W. (1997). Sec31 encodes an essential component of the COPII coat required for transport vesicle budding from the endoplasmic reticulum. *Mol. Biol. Cell*, 8(2):205–217.

Sasvari, Z., Gonzalez, P. A., Rachubinski, R. A., and Nagy, P. D. (2013). Tombusvirus replication depends on sec39p endoplasmic reticulum-associated transport protein. *Virology*, 447(1-2):21–31.

Sato, H. and Maquat, L. E. (2009). Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes Dev.*, 23(21):2537–2550.

Sato, K., Sato, M., and Nakano, A. (2003). Rer1p, a retrieval receptor for ER membrane proteins, recognizes transmembrane domains in multiple modes. *Mol. Biol. Cell*, 14(9):3605–3616.

Saulière, J., Murigneux, V., Wang, Z., et al. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat. Struct. Mol. Biol.*, 19(11):1124–1131.

Sawicki, S. G. and Godman, G. C. (1971). On the differential cytotoxicity of actinomycin D. *J. Cell Biol.*, 50(3):746–761.

Schäfer, A. and Wolf, D. H. (2009). Sec61p is part of the endoplasmic reticulum-associated degradation machinery. *EMBO J.*, 28(19):2874–2884.

Schmidt, S. A., Foley, P. L., Jeong, D.-H., et al. (2015). Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Res.*, 43(1):309–323.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nat. Methods*, 9(7):671–675.

Schneider, K. and Bertolotti, A. (2015). Surviving protein quality control catastrophes–from cells to organisms. *J. Cell Sci.*, 128(21):3861–3869.

Schoebel, S., Mi, W., Stein, A., et al. (2017). Cryo-EM structure of the protein-conducting ERAD channel hrd1 in complex with hrd3. *Nature*, 548(7667):352–355.

Schröder, M. and Kaufman, R. J. (2005). The mammalian unfolded protein response. *Annu. Rev. Biochem.*, 74(1):739–789.

Schubert, U., Antón, L. C., Bacík, I., et al. (1998). CD4 glycoprotein degradation induced by human immunodeficiency virus type 1 vpu protein requires the function of proteasomes and the ubiquitin-conjugating pathway. *J. Virol.*, 72(3):2280–2288.

Schuldiner, M., Metz, J., Schmid, V., et al. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell*, 134(4):634–645.

Schwalb, B., Michel, M., Zacher, B., et al. (2016). TT-seq maps the human transient transcriptome. *Science*, 352(6290):1225–1228.

Schwanhäusser, B., Busse, D., Li, N., et al. (2011). Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342.

Schweingruber, C., Soffientini, P., Ruepp, M.-D., et al. (2016). Identification of interactions in the NMD complex using Proximity-Dependent biotinylation (BioID). *PLoS One*, 11(3):e0150239.

Scott, D. K., Board, J. R., Lu, X., et al. (2003). The neuroblastoma amplified gene, NAG: genomic structure and characterisation of the 7.3 kb transcript predominantly expressed in neuroblastoma. *Gene*, 307(1-2):1–11.

Segarra, N. G., Ballhausen, D., Crawford, H., et al. (2015). NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina. *Am. J. Med. Genet. A*, 167A(12):2902–2912.

Seoighe, C., Kiniry, S. J., Peters, A., et al. (2020). Selection shapes synonymous stop codon use in mammals. *J. Mol. Evol.*, 88(7):549–561.

Shah, S. Z. A., Zhao, D., Hussain, T., and Yang, L. (2017). The role of unfolded protein response and mitogen-activated protein kinase signaling in neurodegenerative diseases with special focus on prion diseases.

Shalem, O., Dahan, O., Levo, M., et al. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.*, 4:223.

Sharp, P. M. and Li, W. H. (1986). Codon usage in regulatory genes in escherichia coli does not reflect selection for 'rare' codons. *Nucleic Acids Res.*, 14(19):7737–7749.

Shaywitz, D. A., Orci, L., Ravazzola, M., et al. (1995). Human SEC13Rp functions in yeast and is located on transport vesicles budding from the endoplasmic reticulum. *J. Cell Biol.*, 128(5):769–777.

Shen, J., Ambrosone, C. B., and Zhao, H. (2009). Novel genetic variants in microRNA genes and familial breast cancer. *Int. J. Cancer*, 124(5):1178–1182.

Sheppard, D. N., Ostedgaard, L. S., Rich, D. P., and Welsh, M. J. (1994). The amino-terminal portion of CFTR forms a regulated cl- channel. *Cell*, 76(6):1091–1098.

Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat. Rev. Mol. Cell Biol.*, 18(11):655–670.

Shirley, R. L., Ford, A. S., Richards, M. R., et al. (2002). Nuclear import of upf3p is mediated by importin-alpha/-beta and export to the cytoplasm is required for a functional nonsense-mediated mRNA decay pathway in yeast. *Genetics*, 161(4):1465–1482.

Shrestha, A. M. S. and Frith, M. C. (2013). An approximate bayesian approach for mapping paired-end DNA reads to a reference genome. *Bioinformatics*, 29(8):965–972.

Shu, H., Donnard, E., Liu, B., et al. (2020). FMRP links optimal codons to mRNA stability in neurons. *Proc. Natl. Acad. Sci. U. S. A.*

Shum, E. Y., Jones, S. H., Shao, A., et al. (2016). The antagonistic gene paralogs upf3a and upf3b govern Nonsense-Mediated RNA decay. *Cell*, 165(2):382–395.

Sibley, C. R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing.

Sicari, D., Delaunay-Moisan, A., Combettes, L., et al. (2020). A guide to assessing endoplasmic reticulum homeostasis and stress in mammalian systems. *FEBS J.*, 287(1):27–42.

Siddiqui, N. and Sonenberg, N. (2016). Proposing a mechanism of action for ataluren. *Proc. Natl. Acad. Sci. U. S. A.*, 113(44):12353–12355.

Simsek, D., Tiu, G. C., Flynn, R. A., et al. (2017). The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell*, 169(6):1051–1065.e18.

Singh, A. K., Choudhury, S. R., De, S., et al. (2019). The RNA helicase UPF1 associates with mRNAs co-transcriptionally and is required for the release of mRNAs from gene loci. *Elife*, 8.

Singh, G., Kucukural, A., Cenik, C., et al. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell*, 151(4):750–764.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol.*, 6(4):e111.

Slobodin, B., Bahat, A., Sehrawat, U., et al. (2020). Transcription dynamics regulate Poly(A) tails and expression of the RNA degradation machinery to balance mRNA levels. *Mol. Cell*, 78(3):434–444.e5.

Smith, L. M., Kelleher, N. L., and Consortium for Top Down Proteomics (2013). Proteoform: a single term describing protein complexity. *Nat. Methods*, 10(3):186–187.

Smith, M. and Wilkinson, S. (2017). ER homeostasis and autophagy. *Essays Biochem.*, 61(6):625–635.

Smith, M. D., Harley, M. E., Kemp, A. J., et al. (2018). CCPG1 is a non-canonical autophagy cargo receptor essential for ER-Phagy and pancreatic ER proteostasis. *Dev. Cell*, 44(2):217–232.e11.

Smith, M. D. and Wilkinson, S. (2018). CCPG1, an unconventional cargo receptor for ER-phagy, maintains pancreatic acinar cell health. *Mol Cell Oncol*, 5(5):e1441631.

Söderberg, O., Gullberg, M., Jarvius, M., et al. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods*, 3(12):995–1000.

Sonenberg, N. and Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: Mechanisms and biological targets.

Souquere, S., Mollet, S., Kress, M., et al. (2009). Unravelling the ultrastructure of stress granules and associated p-bodies in human cells. *J. Cell Sci.*, 122(Pt 20):3619–3626.

Spang, A. (2013). Retrograde traffic from the golgi to the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.*, 5(6).

Spang, A. and Schekman, R. (1998). Reconstitution of retrograde transport from the golgi to the ER in vitro. *J. Cell Biol.*, 143(3):589–599.

Stagg, H. R., Thomas, M., van den Boomen, D., et al. (2009). The TRC8 E3 ligase ubiquitinates MHC class I molecules before dislocation from the ER. *J. Cell Biol.*, 186(5):685–692.

Stainier, D. Y. R., Raz, E., Lawson, N. D., et al. (2017). Guidelines for morpholino use in zebrafish. *PLoS Genet.*, 13(10):e1007000.

Steckelberg, A.-L., Boehm, V., Gromadzka, A. M., and Gehring, N. H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.*, 2(3):454–461.

Stefani, G., Fraser, C. E., Darnell, J. C., and Darnell, R. B. (2004). Fragile X mental retardation protein is associated with translating polyribosomes in neuronal cells. *Journal of Neuroscience*, 24(33):7272–7276.

Stefanovic, S. and Hegde, R. S. (2007). Identification of a targeting factor for post-translational membrane protein insertion into the ER. *Cell*, 128(6):1147–1159.

Stornaiuolo, M., Lotti, L. V., Borgese, N., et al. (2003). KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and golgi complex. *Mol. Biol. Cell*, 14(3):889–902.

Sugimoto, Y., Vigilante, A., Darbo, E., et al. (2015). hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by staufen 1. *Nature*, 519(7544):491–494.

Sun, M., Schwalb, B., Schulz, D., et al. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res.*, 22(7):1350–1359.

Szklarczyk, D., Gable, A. L., Lyon, D., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.*, 47(D1):D607–D613.

Tachie-Menson, T., Gázquez-Gutiérrez, A., Fulcher, L. J., et al. (2020). Characterisation of the biochemical and cellular roles of native and pathogenic amelogenesis imperfecta mutants of FAM83H. *Cell. Signal.*, 72:109632.

Tagaya, M., Arasaki, K., Inoue, H., and Kimura, H. (2014). Moonlighting functions of the NRZ (mammalian dsl1) complex. *Front Cell Dev Biol*, 2:25.

Tang, N. H., Kim, K. W., Xu, S., et al. (2020). The mRNA decay factor CAR-1/LSM14 regulates axon regeneration via mitochondrial calcium dynamics. *Curr. Biol.*, 30(5):865–876.e7.

Tani, H., Imamachi, N., Salam, K. A., et al. (2012). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol.*, 9(11):1370–1379.

Tarpey, P. S., Raymond, F. L., Nguyen, L. S., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. *Nat. Genet.*, 39(9):1127–1133.

Tatsuno, T., Nakamura, Y., Ma, S., et al. (2016). Nonsense-mediated mRNA decay factor upf2 exists in both the nucleoplasm and the cytoplasm. *Mol. Med. Rep.*, 14(1):655–660.

Teixeira, D., Sheth, U., Valencia-Sanchez, M. A., et al. (2005). Processing bodies require RNA for assembly and contain nontranslating mRNAs. *RNA*, 11(4):371–382.

Thein, S. L., Hesketh, C., Taylor, P., et al. (1990). Molecular basis for dominantly inherited inclusion body beta-thalassemia. *Proc. Natl. Acad. Sci. U. S. A.*, 87(10):3924–3928.

Thermann, R., Neu-Yilik, G., Deters, A., et al. (1998). Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J.*, 17(12):3484–3494.

Thomas, M. G., Martinez Tosar, L. J., Desbats, M. A., et al. (2009). Mammalian staufen 1 is recruited to stress granules and impairs their assembly. *J. Cell Sci.*, 122(Pt 4):563–573.

Torkzaban, B., Mohseni Ahooyi, T., Duggan, M., et al. (2020). Cross-talk between lipid homeostasis and endoplasmic reticulum stress in neurodegeneration: Insights for HIV-1 associated neurocognitive disorders (HAND). *Neurochem. Int.*, 141:104880.

Travers, K. J., Patil, C. K., Wodicka, L., et al. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell*, 101(3):249–258.

Travis, S. M., Damico, K., Yu, I.-M., et al. (2020). Structural basis for the binding of SNAREs to the multisubunit tethering complex dsl1. *J. Biol. Chem.*, 295(30):2020.04.07.029496.

Trcek, T., Sato, H., Singer, R. H., and Maquat, L. E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev.*, 27(5):541–551.

Trevino, S. R., Scholtz, J. M., and Pace, C. N. (2007). Amino acid contribution to protein solubility: Asp, glu, and ser contribute more favorably than the other hydrophilic amino acids in RNase sa. *J. Mol. Biol.*, 366(2):449–460.

Tucker, M., Valencia-Sanchez, M. A., Staples, R. R., et al. (2001). The transcription factor associated ccr4 and caf1 proteins are components of the major cytoplasmic mRNA deadenylase in saccharomyces cerevisiae. *Cell*, 104(3):377–386.

Tuller, T., Waldman, Y. Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci. U. S. A.*, 107(8):3645–3650.

Uehata, T. and Akira, S. (2013). mRNA degradation by the endoribonuclease Regnase-1/ZC3H12a/MCPIP-1. *Biochim. Biophys. Acta*, 1829(6-7):708–713.

Uehata, T. and Takeuchi, O. (2017). Regnase-1 is an endoribonuclease essential for the maintenance of immune homeostasis. *J. Interferon Cytokine Res.*, 37(5):220–229.

Unterholzner, L. and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell*, 16(4):587–596.

Usuki, F., Yamashita, A., and Fujimura, M. (2019). Environmental stresses suppress nonsense-mediated mRNA decay (NMD) and affect cells by stabilizing NMD-targeted gene expression. *Sci. Rep.*, 9(1):1279.

Uvarovskii, A. and Dieterich, C. (2017). pulser: Versatile computational analysis of RNA turnover from metabolic labeling experiments. *Bioinformatics*, 33(20):3305–3307.

Van De Geijn, B., Mcvicker, G., Gilad, Y., and Pritchard, J. K. (2015). WASP: Allele-specific software for robust molecular quantitative trait locus discovery.

van der Kloet, F. M., Buurmans, J., Jonker, M. J., et al. (2020). Increased comparability between RNA-Seq and microarray data by utilization of gene sets. *PLoS Comput. Biol.*, 16(9):e1008295.

Van Driesche, S. J. and Martin, K. C. (2018). New frontiers in RNA transport and local translation in neurons. *Dev. Neurobiol.*, 78(3):331–339.

Vandermarliere, E., Bourgois, T. M., Winn, M. D., et al. (2009). Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. *Biochem. J*, 418(1):39–47.

Varma, D., Dujardin, D. L., Stehman, S. A., and Vallee, R. B. (2006). Role of the kinetochore/cell cycle checkpoint protein ZW10 in interphase cytoplasmic dynein function. *J. Cell Biol.*, 172(5):655–662.

Varshavsky, A. (2017). The ubiquitin system, autophagy, and regulated protein degradation. *Annu. Rev. Biochem.*, 86:123–128.

Vashist, S., Kim, W., Belden, W. J., et al. (2001). Distinct retrieval and retention mechanisms are required for the quality control of endoplasmic reticulum protein folding. *J. Cell Biol.*, 155(3):355–368.

Vashist, S. and Ng, D. T. W. (2004). Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. *J. Cell Biol.*, 165(1):41–52.

Vembar, S. S. and Brodsky, J. L. (2008). One step at a time: endoplasmic reticulum-associated degradation. *Nat. Rev. Mol. Cell Biol.*, 9(12):944–957.

Venter, J. C., Adams, M. D., Myers, E. W., et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–1351.

Voorhees, R. M. and Hegde, R. S. (2016). Structure of the sec61 channel opened by a signal sequence. *Science*, 351(6268):88–91.

Wada, T. and Becskei, A. (2017). Impact of methods on the measurement of mRNA turnover. *Int. J. Mol. Sci.*, 18(12).

Walter, P. and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J. Cell Biol.*, 91(2 Pt 1):557–561.

Wang, C., Han, B., Zhou, R., and Zhuang, X. (2016). Real-Time imaging of translation on single mRNA transcripts in live cells. *Cell*, 165(4):990–1001.

Wang, D., Zavadil, J., Martin, L., et al. (2011a). Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. *Mol. Cell. Biol.*, 31(17):3670–3680.

Wang, E. and Aifantis, I. (2020). RNA splicing and cancer.

Wang, F., Durfee, L. A., and Huibregtse, J. M. (2013). A cotranslational ubiquitination pathway for quality control of misfolded proteins. *Mol. Cell*, 50(3):368–378.

Wang, F., Song, W., Brancati, G., and Segatori, L. (2011b). Inhibition of endoplasmic reticulum-associated degradation rescues native folding in loss of function protein misfolding diseases. *J. Biol. Chem.*, 286(50):43454–43464.

Wang, J., Vock, V. M., Li, S., et al. (2002). A quality control pathway that down-regulates aberrant t-cell receptor (TCR) transcripts by a mechanism requiring UPF2 and translation. *J. Biol. Chem.*, 277(21):18489–18493.

Wang, S. and Hazelrigg, T. (1994). Implications for bcd mRNA localization from spatial distribution of exu protein in drosophila oogenesis. *Nature*, 369(6479):400–403.

Wang, Y., Ma, Y., and Carroll, R. J. (2009). Variance estimation in the analysis of microarray data. *J. R. Stat. Soc. Series B Stat. Methodol.*, 71(2):425–445.

Wang, Z., Murigneux, V., and Le Hir, H. (2014). Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol.*, 15(12):551.

Weischenfeldt, J., Damgaard, I., Bryder, D., et al. (2008). NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev.*, 22(10):1381–1396.

Weng, W.-C., Lee, W.-T., Hsu, W.-M., et al. (2011). Role of glucose-regulated protein 78 in embryonic development and neurological disorders. *J. Formos. Med. Assoc.*, 110(7):428–437.

Wengrod, J., Martin, L., Wang, D., et al. (2013). Inhibition of nonsense-mediated RNA decay activates autophagy. *Mol. Cell. Biol.*, 33(11):2128–2135.

Wickham, H., Averick, M., Bryan, J., et al. (2019). Welcome to the tidyverse. *JOSS*, 4(43):1686.

Wiertz, E. J., Tortorella, D., Bogyo, M., et al. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature*, 384(6608):432–438.

Wild, K., Halic, M., Sinning, I., and Beckmann, R. (2004). SRP meets the ribosome. *Nat. Struct. Mol. Biol.*, 11(11):1049–1053.

Wilhelm, B. T. and Landry, J.-R. (2009). RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, 48(3):249–257.

Wilkinson, M. E., Charenton, C., and Nagai, K. (2020). RNA splicing by the spliceosome. *Annu. Rev. Biochem.*, 89(1):359–388.

Wilkinson, M. F. (2005). A new function for nonsense-mediated mRNA-decay factors. *Trends Genet.*, 21(3):143–148.

Wilkinson, M. F. (2019). Genetic paradox explained by nonsense. *Nature*, 568(7751):179–180.

Will, C. L. and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, 3(7):1–2.

Wilschanski, M., Yahav, Y., Yaacov, Y., et al. (2003). Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations. *N. Engl. J. Med.*, 349(15):1433–1441.

Wimmer, K., Zhu, X. X., Lamb, B. J., et al. (1999). Co-amplification of a novel gene, NAG, with the n-myc gene in neuroblastoma. *Oncogene*, 18(1):233–238.

Wissink, E. M., Vihervaara, A., Tippens, N. D., and Lis, J. T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nat. Rev. Genet.*, 20(12):705–723.

Wittkopp, N., Huntzinger, E., Weiler, C., et al. (2009). Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. *Mol. Cell. Biol.*, 29(13):3517–3528.

Wittmann, J., Hol, E. M., and Jäck, H.-M. (2006). hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell. Biol.*, 26(4):1272–1287.

Wolff, S., Weissman, J. S., and Dillin, A. (2014). Differential scales of protein quality control. *Cell*, 157(1):52–64.

Wong, J. J.-L., Au, A. Y. M., Ritchie, W., and Rasko, J. E. J. (2016). Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. *Bioessays*, 38(1):41–49.

Wong, J. J.-L., Ritchie, W., Ebner, O. A., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*, 154(3):583–595.

Wu, B., Eliscovich, C., Yoon, Y. J., and Singer, R. H. (2016). Translation dynamics of single mRNAs in live cells and neurons. *Science*, 352(6292):1430–1435.

Wu, Q., Wright, M., Gogol, M. M., et al. (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. *EMBO J.*, 39(17):e104763.

Wysocka, J., Swigut, T., Xiao, H., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, 442(7098):86–90.

Yamamoto, K., Fujii, R., Toyofuku, Y., et al. (2001). The KDEL receptor mediates a retrieval mechanism that contributes to quality control at the endoplasmic reticulum. *EMBO J.*, 20(12):3082–3091.

Yan, C., Liu, J., Gao, J., et al. (2019). IRE1 promotes neurodegeneration through autophagy-dependent neuron death in the drosophila model of parkinson's disease. *Cell Death Dis.*, 10(11):800.

Yan, X., Hoek, T. A., Vale, R. D., and Tanenbaum, M. E. (2016). Dynamics of translation of single mRNA molecules in vivo. *Cell*, 165(4):976–989.

Yates, A. D., Achuthan, P., Akanni, W., et al. (2020). Ensembl 2020. *Nucleic Acids Res.*, 48(D1):D682–D688.

Yepiskoposyan, H., Aeschimann, F., Nilsson, D., et al. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA*, 17(12):2108–2118.

Yoshida, H., Matsui, T., Hosokawa, N., et al. (2003). A time-dependent phase shift in the mammalian unfolded protein response. *Dev. Cell*, 4(2):265–271.

Yoshida, H., Oku, M., Suzuki, M., and Mori, K. (2006). pXBP1(U) encoded in XBP1 pre-mRNA negatively regulates unfolded protein response activator pXBP1(S) in mammalian ER stress response. *J. Cell Biol.*, 172(4):565–575.

Young, R. S. and Ponting, C. P. (2013). Identification and function of long non-coding RNAs. *Essays Biochem.*, 54(1):113–126.

Zaher, H. S. and Green, R. (2009a). Fidelity at the molecular level: lessons from protein synthesis. *Cell*, 136(4):746–762.

Zaher, H. S. and Green, R. (2009b). Quality control by the ribosome following peptide bond formation. *Nature*, 457(7226):161–166.

Zeisel, A., Köstler, W. J., Molotski, N., et al. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.*, 7(1):529.

Zetoune, A. B., Fontanière, S., Magnin, D., et al. (2008). Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. *BMC Genet.*, 9:83.

Zhao, L., Longo-Guess, C., Harris, B. S., et al. (2005). Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. *Nat. Genet.*, 37(9):974–979.

Zhao, S., Fung-Leung, W.-P., Bittner, A., et al. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*, 9(1):e78644.

Zhou, Y., Li, Y., Wang, N., et al. (2019). UPF1 inhibits the hepatocellular carcinoma progression by targeting long non-coding RNA UCA1. *Sci. Rep.*, 9(1):6652.

Zhu, X., Zhang, H., and Mendell, J. T. (2020). Ribosome recycling by ABCE1 links lysosomal function and iron homeostasis to 3' UTR-Directed regulation and Nonsense-Mediated decay. *Cell Rep.*, 32(2):107895.

Zünd, D., Gruber, A. R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat. Struct. Mol. Biol.*, 20(8):936–943.

Zünd, D. and Mühlemann, O. (2013). Recent transcriptome-wide mapping of UPF1 binding sites reveals evidence for its recruitment to mRNA before translation. *Translation (Austin)*, 1(2):e26977.

Appendices

Appendix A

Sequence of eGFP-3xFLAG-NBAS repair template

Sequence of repair template ordered on plasmid backbone from IDT. eGFP (green) and 3xFLAG (red) sequences are indicated

```
   1 AAGCTTGAGC TCCTGTAGCA TCCTTCCTCT GCTACCTTTC ATTGGATAAC
  51 TTTAGTTGTT TATTTTAAAA CCACTGTTTG AACTCTCTCT ACGTGTTAGG
 101 CAAAGAATAT ACAAATGAAC ATTAGCATTT CAGTGGGGTA CCTTTACCAT
 151 TTCTGTTTTA TGGAAACTGA GTTATAAGAG AGGCTCAATG ACTTGGCCGA
 201 TCTCATGTAG ATTAATGAAC AAGGAGTGTA TCAGGACCGG GACACAAGTC
 251 GTTCCCAAGA CAAGCAATAG TCCCTGTCCT TCCTACTCAT AAAAGGGTTA
 301 ACCTAGCATT TCCTTGCCGT GAAACAAAAG CGGCCGCGGA TAGAGAAGAT
 351 TGTCCTGTGA GGGATTCCGT AGAGACACTC CTCCGCTGCC TGAGTAATCG
 401 GCGAACATGG TGAGCAAGGG CGAGGAGCTG TTCACCGGGG TGGTGCCCAT
 451 CCTGGTCGAG CTGGACGGCG ACGTAAACGG CCACAAGTTC AGCGTGTCCG
 501 GCGAGGGCGA GGGCGATGCC ACCTACGGCA AGCTGACCCT GAAGTTCATC
 551 TGCACCACCG GCAAGCTGCC CGTGCCCTGG CCCACCCTCG TGACCACCCT
 601 GACCTACGGC GTGCAGTGCT TCAGCCGCTA CCCCGACCAC ATGAAGCAGC
 651 ACGACTTCTT CAAGTCCGCC ATGCCCGAAG GCTACGTCCA GGAGCGCACC
 701 ATCTTCTTCA AGGACGACGG CAACTACAAG ACCCGCGCCG AGGTGAAGTT
 751 CGAGGGCGAC ACCCTGGTGA ACCGCATCGA GCTGAAGGGC ATCGACTTCA
 801 AGGAGGACGG CAACATCCTG GGGCACAAGC TGGAGTACAA CTACAACAGC
 851 CACAACGTCT ATATCATGGC CGACAAGCAG AAGAACGGCA TCAAGGTGAA
 901 CTTCAAGATC CGCCACAACA TCGAGGACGG CAGCGTGCAG CTCGCCGACC
```

```
 951 ACTACCAGCA GAACACCCCC ATCGGCGACG GCCCCGTGCT GCTGCCCGAC
1001 AACCACTACC TGAGCACCCA GTCCGCCCTG AGCAAAGACC CCAACGAGAA
1051 GCGCGATCAC ATGGTCCTGC TGGAGTTCGT GACCGCCGCC GGGATCACTC
1101 TCGGCATGGA CGAGCTGTAC AAGTCCGGAC TCAGATCTGA CTACAAAGAC
1151 CATGACGGTG ATTATAAAGA TCATGACATC GACTACAAGG ATGACGATGA
1201 CAAGGGTACC AGCGGCGGAA GCGCAGCCCC CGAGTCAGGG CCGGCTTTGA
1251 GTCCAGGCAC TGCAGAGGGT GAGGAGGAGA CGATTCTCTA TGACTTGTTG
1301 GTCAACACCG AGTGGCCACC GGAGACTGAA GTACAGGTGA ACCAGGCCCA
1351 TCTACAGCAG CAGCCAGCTT GGCGCGGGCG CTTCGTGGGT AGAGCCACGT
1401 AGCAGGAGCC ACGTAGGGGA GTGGAGACGG GTGGGTGGGG GACTAGGGCC
1451 GGTTGGCTTG TGCGGTCGCC TTGGCAACGG CACCTTGGGC TAGCGGGGGT
1501 CTAGGGCGGG GTGGGGTGGT GGTCCCTGAG ATTCTTCCTG GGCTCAGAAT
1551 TCGGTTGTTG GGGGCTGGAA CTTCGTGCTC TGGGGAGCGT ACGATCCCAC
1601 TTTGGTTCTC AGTTTCGCTC TCGGATCC
```

Appendix B

Total and fractionated RNA-sequencing mapping statistics

Total number of reads, number and percentage mapped by STAR and number and percentage of those aaaigned to gene features by HTSeq. Sample name is comprised of replicate number (E1, E2 or E3), siRNA treatment (si_scr, si_U1, si_NB or si_U2) and fraction (total, cytoplasm or membrane).

Sample	Input reads	% mapped	n# mapped	% assigned	Assigned
E1_si_scr_T	63,926,590	94.14%	60,180,492	93.94%	56,534,784
E1_si_scr_C	70,609,179	93.36%	65,920,730	93.19%	61,434,110
E1_si_scr_M	80,863,384	94.48%	76,399,725	93.55%	71,469,771
E1_si_U1_T	50,587,345	94.17%	47,638,103	93.67%	44,624,912
E1_si_U1_C	77,610,793	94.52%	73,357,722	92.98%	68,208,680
E1_si_U1_M	71,010,168	94.94%	67,417,053	93.03%	62,715,263
E1_si_NB_T	62,048,402	94.25%	58,480,619	93.60%	54,740,331
E1_si_NB_C	57,361,263	93.42%	53,586,892	93.25%	49,969,195
E1_si_NB_M	68,588,570	94.58%	64,871,070	93.62%	60,734,004
E1_si_U2_T	50,643,340	94.87%	48,045,337	93.51%	44,926,150
E1_si_U2_C	61,007,760	94.05%	57,377,798	93.09%	53,411,322
E1_si_U2_M	73,040,738	94.42%	68,965,065	93.36%	64,388,779
E2_si_scr_T	71,707,270	94.47%	67,741,858	93.60%	63,408,237
E2_si_scr_C	60,230,641	93.87%	56,538,503	93.09%	52,631,314
E2_si_scr_M	73,024,049	94.64%	69,109,960	93.40%	64,550,441
E2_si_U1_T	73,836,453	94.14%	69,509,637	93.26%	64,822,586
E2_si_U1_C	70,893,266	94.39%	66,916,154	92.72%	62,046,475
E2_si_U1_M	80,134,386	94.53%	75,751,035	92.87%	70,347,788
E2_si_NB_T	64,443,164	94.41%	60,840,791	93.61%	56,952,199
E2_si_NB_C	54,577,081	93.46%	51,007,740	93.22%	47,547,538
E2_si_NB_M	58,816,011	94.39%	55,516,433	93.44%	51,872,442
E2_si_U2_T	60,307,676	94.07%	56,731,431	93.62%	53,112,425
E2_si_U2_C	64,734,634	94.02%	60,863,503	93.22%	56,736,655
E2_si_U2_M	69,923,392	94.64%	66,175,498	93.23%	61,697,706
E3_si_scr_T	65,495,052	94.58%	61,945,220	93.88%	58,152,423
E3_si_scr_C	65,311,574	93.52%	61,079,384	92.98%	56,792,304
E3_si_scr_M	64,335,042	94.59%	60,854,516	93.56%	56,937,464
E3_si_U1_T	53,794,770	94.51%	50,841,437	93.05%	47,310,482
E3_si_U1_C	51,736,706	93.84%	48,549,725	92.80%	45,052,890
E3_si_U1_M	64,970,487	94.44%	61,358,128	93.04%	57,084,893
E3_si_NB_T	58,838,494	94.56%	55,637,680	93.63%	52,092,846
E3_si_NB_C	72,622,531	93.43%	67,851,231	93.26%	63,277,354
E3_si_NB_M	68,284,583	94.42%	64,474,303	93.64%	60,370,683
E3_si_U2_T	57,346,902	94.58%	54,238,700	93.40%	50,657,168
E3_si_U2_C	71,052,473	94.02%	66,803,535	93.00%	62,126,871
E3_si_U2_M	58,135,978	94.46%	54,915,245	93.54%	51,367,924

Appendix C

Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum

Longman D, **Jackson-Jones KA**, Maslon MM, Murphy LC, Young RS, Stoddart JJ, Hug N, Taylor MS, Papadopoulos DK, & Cáceres JF

Genes & Development, Volume 34
August 1st 2020

Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum

Dasa Longman, Kathryn A. Jackson-Jones, Magdalena M. Maslon, Laura C. Murphy, Robert S. Young,[1] Jack J. Stoddart, Nele Hug, Martin S. Taylor, Dimitrios K. Papadopoulos, and Javier F. Cáceres

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom

Nonsense-mediated decay (NMD) is a translation-dependent RNA quality control mechanism that occurs in the cytoplasm. However, it is unknown how NMD regulates the stability of RNAs translated at the endoplasmic reticulum (ER). Here, we identify a localized NMD pathway dedicated to ER-translated mRNAs. We previously identified NBAS, a component of the Syntaxin 18 complex involved in Golgi-to-ER trafficking, as a novel NMD factor. Furthermore, we show that NBAS fulfills an independent function in NMD. This ER–NMD pathway requires the interaction of NBAS with the core NMD factor UPF1, which is partially localized at the ER in the proximity of the translocon. NBAS and UPF1 coregulate the stability of ER-associated transcripts, in particular those associated with the cellular stress response. We propose a model where NBAS recruits UPF1 to the membrane of the ER and activates an ER-dedicated NMD pathway, thus providing an ER-protective function by ensuring quality control of ER-translated mRNAs.

[*Keywords*: nonsense-mediated decay (NMD); RNA quality control; UPF1; NBAS; ER stress; UPR]

Supplemental material is available for this article.

Received March 3, 2020; revised version accepted June 5, 2020.

The nonsense-mediated decay (NMD) pathway is a highly conserved surveillance mechanism that targets mRNAs harboring premature termination codons (PTCs) for degradation. In doing so, it prevents the accumulation of truncated proteins and modulates the phenotypic outcome of genetic disorders that arise due to the presence of PTCs (Bhuvanagiri et al. 2010). Importantly, NMD also controls the stability of a large number of endogenous RNAs and fine-tunes many physiological processes (for review, see Karousis and Mühlemann 2019; Kurosaki et al. 2019).

Core NMD factors were initially identified in genetic screens in *C. elegans* and in *S. cerevisiae* and were later shown to be involved in NMD in other species, including *Arabidopsis*, *Drosophila*, and mammals. Human orthologs include SMG1, UPF1, UPF2, UPF3, SMG5, SMG6, and SMG7. Additional NMD factors have been identified using a variety of different experimental approaches, including interactome studies, RNAi screens in nematodes, and CRISPR screens in mammalian cells (Hug et al. 2016; Alexandrov et al. 2017; Baird et al. 2018). Mechanistically, NMD is tightly linked to mRNA translation and is initiated by the recognition of a PTC by the surveillance (SURF) complex within which the RNA helicase UPF1 and its associated kinase, SMG1, bind to the ribosomal release factors eRF1 and eRF3 (Karousis and Mühlemann 2019). Subsequently, components of the SURF complex interact with the core NMD factors UPF2 and UPF3B, and with an exon junction complex (EJC) located downstream from the PTC, to form the decay-inducing complex (DECID) that triggers UPF1 phosphorylation by SMG1, and dissociation of eRF1 and eRF3 (Kashima et al. 2006; Singh et al. 2007). This leads to the recruitment of mRNA degradation factors that trigger RNA decay. Substrate selection for NMD occurs not only in an EJC-dependent manner, but alternatively via an EJC-independent mechanism, which targets transcripts harboring very long 3′ UTRs (Metze et al. 2013; Kurosaki et al. 2019).

Using RNAi screens in *C. elegans*, we previously identified several novel NMD factors that were also essential for viability, suggesting that they fulfill additional functions in nematodes, where this pathway is not essential (Longman et al. 2007; Casadio et al. 2015). One of these novel NMD factors is encoded by *smgl-1* that corresponds

[1]**Present address: Centre for Global Health Research, Usher Institute, University of Edinburgh, Old Medical School, Edinburgh EH8 9AG, UK**
Corresponding author: javier.caceres@igmm.ed.ac.uk
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.338061.120. Freely available online through the *Genes & Development* Open Access option.

to the human gene *NBAS* (neuroblastoma amplified sequence, also known as *NAG*, for neuroblastoma amplified gene). *NBAS* was first identified as a gene that is coamplified with the *N-myc* gene in human neuroblastomas; however, no clear role in the disease has been reported (Wimmer et al. 1999; Scott et al. 2003). We previously showed that NBAS acts in concert with UPF1 to coregulate a large number of transcripts not only in nematodes but also in zebrafish and human cells (Anastasaki et al. 2011; Longman et al. 2013). *NBAS* encodes a peripheral ER membrane protein that is a component of the Syntaxin 18 complex, which functions in Golgi-to-ER retrograde transport (Aoki et al. 2009). A series of loss-of-function mutations in *NBAS* have been found in several human conditions, including biallelic mutations in patients with a multisystem disease involving liver, eye, immune system, connective tissue, and bone (Haack et al. 2015; Segarra et al. 2015). Compound heterozygous variants in *NBAS* were also identified as a cause of atypical osteogenesis imperfecta (Balasubramanian et al. 2017) and in a short stature with optic atrophy and Pelger-Huët anomaly (SOPH) syndrome (Maksimova et al. 2010). Currently, it remains unclear whether the phenotypes observed in patients with mutations in *NBAS* are due to a compromised NMD response, defects in Golgi-to-ER retrograde transport, or a combination of both.

Despite initial controversy concerning the intracellular location of NMD in mammalian cells, it has been conclusively demonstrated that decay of a PTC-containing β-globin NMD reporter occurs in the cytoplasm (Trcek et al. 2013). The ER is a major site of localized protein synthesis, with approximately a third of all mRNAs being translated there, in particular, those encoding proteins entering the secretory pathway. It has become increasingly evident that ER and cytosol constitute different environment for protein translation and posttranscriptional gene regulation (Reid and Nicchitta 2015). Current efforts have focused on the mechanism and regulation of cytoplasmic NMD; however, it is largely unknown how this mechanism operates on mRNAs that are translated at the ER, which, due to their intrinsic localized translation, will not have sufficient exposure to cytoplasmic NMD. There is a precedent for a localized NMD response in neurons, where NMD regulates the expression of both dendritic and axonal mRNAs upon their activation of localized mRNA translation (Giorgi et al. 2007; Colak et al. 2013). Both NBAS and a second novel NMD factor identified in our RNAi screens, SEC13 (nuclear pore and COPII coat complex component), localize to the membrane of the ER, raising the possibility that they could be involved in an ER-localized NMD pathway (Longman et al. 2007; Casadio et al. 2015).

Here, we present evidence that reveals a central role for NBAS, acting together with UPF1, in an NMD response that is associated with the ER. We show that NBAS has dual roles in NMD and Golgi-to-ER retrograde transport; but importantly, these functions act independent of each other. We demonstrate that NBAS recruits the core NMD factor UPF1 to the membrane of the ER and promotes the degradation of NMD substrates that are translated at the ER.

Results

A dual role of NBAS in Golgi-to-ER transport and NMD

NBAS encodes a 2371 amino acid protein containing WD40 repeats and a SEC39 domain, present in proteins involved in the secretory pathway (Fig. 1A). Together with RINT1 and ZW10, NBAS forms part of the evolutionarily conserved NRZ complex, which functions as a tethering complex for retrograde trafficking of COPI vesicles from the Golgi to the ER and is part of the larger Syntaxin 18 complex (Aoki et al. 2009; Çivril et al. 2010).

To dissect whether NBAS had separate roles in NMD and Golgi-to-ER transport, or whether these processes influence each other, we first tested whether abrogating the NMD pathway influences ER trafficking. We used a modified flow cytometry-based assay that relies on the expression of an eGFP fluorescent reporter for measuring constitutive secretion. This assay is based on the property of mutant FKBP (FK506-binding protein) (F36M) to form large aggregates that, when expressed in the ER, cannot be secreted. However, upon incubation with an FKBP (F36M) ligand, D/D solubilizer, these aggregates are solubilized, leading to efficient secretion (Gordon et al. 2010). We used this reporter in combination with siRNA-mediated knockdown of individual NMD factors or of components of the Syntaxin 18 complex (Fig. 1B; Supplemental Fig. S1A). Fluorescence of clonal HeLa C1 cells stably expressing the eGFP reporter decreased upon addition of the ligand, D/D solubilizer, to the level of control cells, as was observed previously (Supplemental Fig. S1A, left panel; Gordon et al. 2010). As expected, siRNA-mediated knockdown of components of the Syntaxin 18 complex, including STX18, p31, and NBAS, resulted in an increase of the remaining fluorescence due to an interference with the secretion process (Fig. 1B). We observed some effect upon depletion of ZW10, but almost no effect with the knockdown of RINT1, perhaps reflecting a differential contribution of these components to the secretion process. Importantly, we observed that a strong reduction in the levels of individual NMD factors did not interfere with secretion. This was the case for UPF1 and UPF2, but also for the RNA helicase DHX34, and for SEC13, another NMD factor that localizes to the ER (Fig. 1B; Supplemental Fig. S1A). From this experiment, we conclude that NMD activity is not required for ER secretion.

Next, we investigated whether interfering with constitutive secretion had any effect on the NMD pathway. For this, HeLa cells stably expressing a well-characterized β-globin NMD reporter harboring a nonsense mutation at position 39 (HBB NS39) (Trecartin et al. 1981), or its wild-type (WT) HBB counterpart, were depleted of NMD factors (UPF2 or NBAS) or of secretion factors (STX18 or STX5) using specific siRNAs (Supplemental Fig. S1D). As expected, the level of HBB WT mRNA remained unchanged upon UPF1, NBAS, STX18, or STX5 depletion (Fig. 1C, right panel). In contrast, depletion of UPF2 or of NBAS resulted in an increased level of the HBB NS39 NMD reporter mRNA, compared with mock-depleted cells (Fig. 1C, left panel). Importantly,

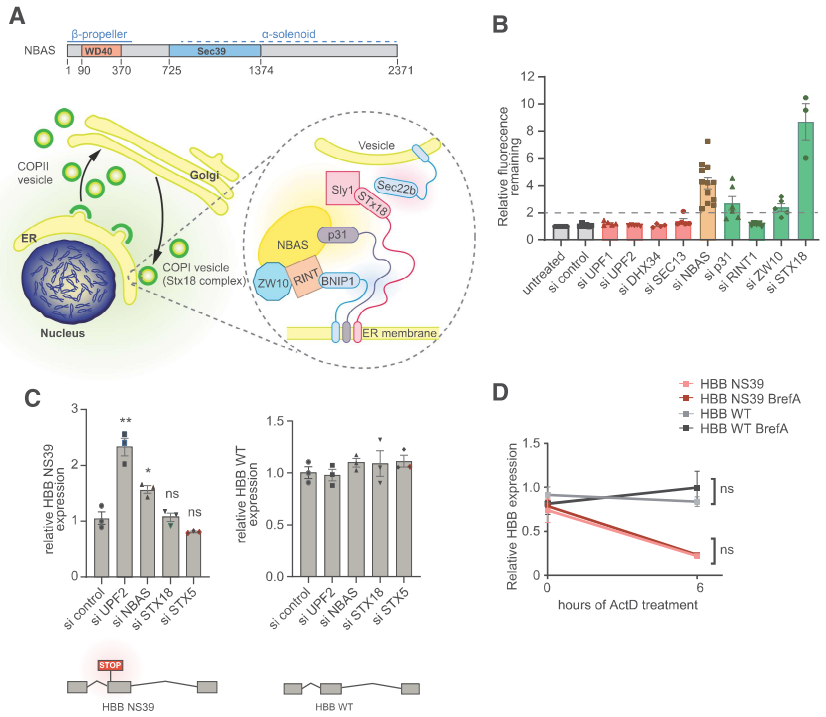


Figure 1. NBAS is an ER-localized protein with dual, independent, functions in NMD and ER secretion. (*A*) Cartoon depicting the functional domains of NBAS and its intracellular localization at the endoplasmic reticulum (ER), as part of the Syntaxin 18 complex. (*B*) NMD activity is not required for constitutive secretion. Depletion of NMD factors, with the exception of NBAS, does not affect the ability of HeLa C1 cells to secrete a GFP-based reporter construct. GFP accumulation was measured by flow cytometry and a secretion defect was defined as a twofold accumulation of GFP reporter (after the addition of 1 µM D/D solubilizer [ligand] compared with untreated HeLa C1 cells [dashed line]). Each point represents one biological replica, bars indicate mean with SEM. (*C*) NMD activity is not affected by depletion of secretion factors, STX18 and STX5. In contrast, depletion of UPF2 and NBAS led to an up-regulation of the HBB NS39 NMD-sensitive reporter. HeLa cells stably expressing the HBB NS39 or wild-type (WT) reporters were depleted of the indicated factors and the steady state level of the reporter mRNA was measured by qRT-PCR and normalized to POLR2J expression. Each point represents one biological replica, bars indicate mean with SEM. Significance was determined by two-tailed unpaired *t*-test. (**) $P < 0.005$; (*) $P < 0.05$; (ns) not significant. (*D*) NMD activity is not affected by blocking constitutive secretion. HeLa cells expressing HBB WT or HBB NS39 reporters were treated with actinomycin D to block transcription and Brefeldin A to block constitutive secretion. HBB expression was measured as described in *C*. Each point represents the mean and SEM of three biological replicas. The significance of Brefeldin A treatment was determined by two-tailed unpaired *t*-test. (ns) Not significant. See Supplemental Figure S1C for the effect of Brefeldin A on secretion.

depletion of either STX18 or STX5 did not affect the level of β-globin NMD reporter mRNA (Fig. 1C), indicating that normal ER secretion is not required for NMD. These results were confirmed using a fluorescent NMD reporter (NMD$^+$) that quantifies NMD activity at the single cell level. Cells carrying the NMD$^+$ reporter were identified by the constitutive expression of red fluorescence, whereas NMD activity was determined by the mean green fluorescence, which is subject to NMD, in all red cells (Pereverzev et al. 2015). Here again, whereas knockdown of UPF2 and NBAS resulted in increased levels of the NMD reporter (measured by increased green fluorescence), depletion of STX18 or STX5 had no effect (Supplemental Fig. S1B). We extended these observations to show that NMD activity is not affected by blocking constitutive secretion with the use of Brefeldin A (Supplemental Fig. S1C; Aridor et al. 1995). HeLa cells stably expressing either HBB WT or HBB NS39 were treated with actinomycin D to block transcription in the presence or absence of Brefeldin A. We observed that the stability of the HBB NS39 NMD reporter mRNA was not increased by blocking ER secretion (Fig. 1D). Altogether, these experiments show that NMD activity and ER secretion are not functionally linked, strongly suggesting that NBAS has two independent roles in Golgi-ER retrograde transport and in NMD.

NBAS regulates a subset of NMD targets specifically translated at the ER

Previously, we conducted RNA profiling experiments that revealed a large proportion of NBAS mRNA targets are coregulated by the core NMD factor UPF1, with a significant enrichment for genes involved in the cellular stress response (Longman et al. 2013). Here, we extended this analysis by RNA sequencing to profile changes in mRNA abundance upon depletion of NBAS or UPF1. Both NBAS and UPF1 were the most significantly down-regulated genes in the relevant samples (fold change −1.85, $P < 3.51 \times 10^{-13}$, and fold change −3.28, $P < 1.14 \times 10^{-148}$, respectively) (Supplemental Table S1). Depletion of UPF1 significantly affected the mRNAs of 4756 genes, with 2411 genes displaying an increased mRNA expression upon UPF1 knockdown. Depletion of NBAS increased the mRNA levels of 209 genes (Supplemental Table S1), consistent with the view that NBAS regulates a subset of the UPF1 targets. We observed a robust coregulation of mRNAs when UPF1 or NBAS was depleted (Pearson's correlation $r = 0.67$, $P < 0.0001$) (Fig. 2A), indicating that UPF1 and NBAS function in a common pathway. Gene ontology (GO) analysis of the mRNAs regulated by both NBAS and UPF1 revealed a strong enrichment for the secretome, which is preferentially translated at the ER translocon (Supplemental Table S2). Interestingly,

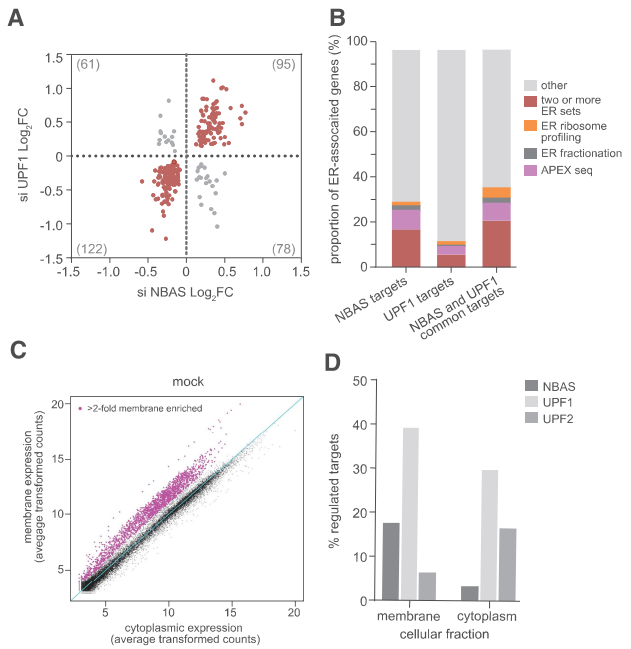


Figure 2. NBAS-regulated targets are enriched for ER-localized transcripts. (*A*) Scatter plot of differentially expressed transcripts shows a significant positive correlation between NBAS and UPF1-regulated targets (Pearson's correlation $r = 0.67$; $P < 0.0001$). Numbers in brackets indicate the number of regulated transcripts in each quadrant. (*B*) NBAS-regulated targets are enriched for experimentally identified ER-localized transcripts. Bar charts show the proportion of targets that are ER-localized by APEX-seq, ER fractionation, and ER proximity-specific ribosome profiling. The ER enrichment of NBAS targets is significantly higher than that of UPF1 targets ($P < 0.0001$, exact binomial test), whereas the difference between NBAS targets and NBAS-UPF1 common targets is not statistically significant ($P = 0.1867$, exact binomial test). (*C*) RNA-seq of samples from subcellular fractions (membrane and cytoplasm) allows identification of membrane-associated genes. Scatter plot shows average transformed counts of each gene in membrane and cytoplasmic fractions in mock-depleted cells. (*D*) NBAS preferentially regulates membrane-associated targets. Plot shows percentage of "membrane-associated" genes in the membrane fraction and of "nonmembrane" genes in the cytoplasm that are regulated ($P < 0.05$) by each NMD factor. NBAS regulates over fivefold higher percentage of membrane-associated genes than nonmembrane genes ($P < 0.001$, exact binomial test). In contrast, UPF1 regulates a similar percentage of both membrane-associated and nonmembrane genes (1.3-fold higher percentage of membrane-associated genes, $P < 0.001$, exact binomial test), and UPF2 regulates 2.5-fold higher percentage of nonmembrane genes than membrane genes ($P < 0.001$, exact binomial test).

we found a significant increase in fold change in genes associated with the ER unfolded protein response (UPR) (GO: 0006986) when either NBAS or UPF1 was depleted ($P < 0.05$ and $P < 0.001$, respectively; Wilcoxon rank sum test), showing that the ER stress response is perturbed when the NMD response at the ER is not functional (Supplemental Fig. S2). This supports our hypothesis that an ER-specific NMD pathway is essential for cellular homeostasis.

To further characterize genes regulated by NBAS and UPF1, we intersected targets regulated by either NBAS or UPF1 and also common targets, with three experimental data sets identifying genes localized and/or translated at the ER. APEX-seq, a method for RNA sequencing based on direct proximity labeling of RNA led to the identification of 1077 mRNAs localized at the ER (Fazal et al. 2019), cell fractionation followed by ribosome footprinting identified 486 mRNAs (Reid and Nicchitta 2012), and proximity-specific ribosome profiling, based on ER membrane proximity labeling of ribosomes, identified 686 mRNAs translated at the ER (Jan et al. 2014). We found that NBAS targets are strongly enriched for experimentally validated ER genes (31.6%), as compared with 14.4% of UPF1 targets. Furthermore, 18.2% of NBAS targets were found in two or more of the data sets, as compared with only 6.3% of UPF1 targets. Considering NBAS and UPF1 common targets, 37.9% of targets were found in at least one data set, while 23.2% were found in two or more data sets (Fig. 2B). NBAS-regulated targets were strongly enriched for experimentally identified ER-localized genes (OR = 6.07, $P < 0.001$, Fisher's exact test), whereas UPF1 targets were only modestly enriched (OR = 2.44, $P < 0.001$, Fisher's exact test). Compared with UPF1 targets, we found that those targets that were also regulated by NBAS were 2.6-fold more likely than expected by chance to be found at the ER ($P < 0.001$, exact binomial test), confirming that NBAS and UPF1 together regulate NMD targets specifically at the ER.

We next examined how NBAS depletion affected transcripts specifically at the ER compared with the cytoplasm, and how this differs from depletion of core NMD factors UPF1 and UPF2. We performed RNA sequencing on cytoplasmic and membrane fractions of HeLa cells depleted for each factor or mock-depleted control cells, following subcellular fractionation. By comparing expression of genes in each fraction in control samples we defined a group of "membrane-associated" genes, as those that in the control sample showed a twofold higher expression in the membrane fraction (magenta points), whereas all other expressed genes were deemed "nonmembrane" genes (Fig. 2C). We assessed the validity of cell fractionation by identifying genes in the previously described experimentally validated ER data sets in our data. As expected, we observed that the majority of genes present in two or more ER data sets were found in the membrane-associated set (OR = 66.94, $P < 0.0001$, Fisher's exact test) (Supplemental Fig. S3A). Next, we analyzed changes in gene expression of "membrane-associated" genes in the membrane fraction, and of "nonmembrane" genes in the cytoplasmic fraction, upon depletion of individual NMD factors (Fig. 2D; Supplemental Fig. S3B,C; Supplemental Table S3). Knockdown of UPF1 led to a robust increase in the abundance of mRNAs in both membrane and cytoplasmic fractions, whereas depletion of NBAS led to an up-regulation of a higher percentage of membrane-associated genes than of nonmembrane genes (fivefold increase, $P < 0.001$, exact binomial test). In contrast, depletion of UPF2 led to preferential up-regulation of "nonmembrane" genes in the cytoplasm (>2.5-fold,

$P < 0.001$, exact binomial test). Altogether, these results suggest that NBAS is a crucial component of an ER–NMD pathway that targets for degradation mRNAs that are translated at the ER, constituting a novel localized NMD response.

Site of NMD for transcripts translated at the ER

Single-molecule RNA fluorescent in situ hybridization (smRNA FISH) was used previously to localize the NMD response of a β-globin NMD reporter to the cytoplasm (Trcek et al. 2013). Here, we used smRNA FISH to spatially map the location of mRNA degradation of mRNAs that are translated at the ER and that we also showed to be up-regulated upon UPF1 and/or NBAS depletion (Longman et al. 2013). We used a set of fluorescent probes that label the full-length of two endogenous mRNAs targeted by NMD, which are translated either in the cytoplasm or at the ER. We selected *SETD4* (SET Domain Containing 4) mRNA, which encodes a lysine methyltransferase and is translated in the cytoplasm (Faria et al. 2013). As an example of an mRNA that is translated at the ER, we selected *FAP* mRNA (also known as seprase), which encodes fibroblast activation protein α, a 170-kDa membrane-bound gelatinase (Goldstein et al. 1997). First, we showed that levels of *SETD4* mRNA were robustly increased upon knockdown of UPF1, whereas NBAS depletion had only a marginal stabilizing effect (Fig. 3A). In contrast, the ER-localized *FAP* mRNA was comparably up-regulated upon knockdown of UPF1 and of NBAS (Fig. 3A).

RNA FISH of *SETD4* mRNA in UPF1-depleted cells revealed a strong increase of uniformly distributed fluorescent signal throughout the cell, consistent with NMD taking place in the cytoplasm. As expected, depletion of NBAS led only to a slight up-regulation of *SETD4* mRNA (Fig. 3B). Importantly, RNA FISH of the ER-translated *FAP* mRNA in both UPF1 and NBAS-depleted cells revealed a strong increase of the fluorescent signal that clustered to the perinuclear region of the cell (Fig. 3C, green signal). The RNA FISH fluorescence overlapped with ER staining (Fig. 3C, right panels), thus spatially mapping the mRNA degradation of an NBAS NMD target to the ER. This is consistent with an NBAS-dependent NMD-mediated degradation occurring at the ER. In order to represent the differences in the distribution of the smRNA FISH signal within cells, we plotted the distance of the fluorescent FISH signal from the edge of the nucleus (defined by DAPI staining), for each experimental condition. This demonstrated that UPF1 depletion leads to the accumulation of *SETD4* mRNA that is distributed widely within the cell, which is significantly different

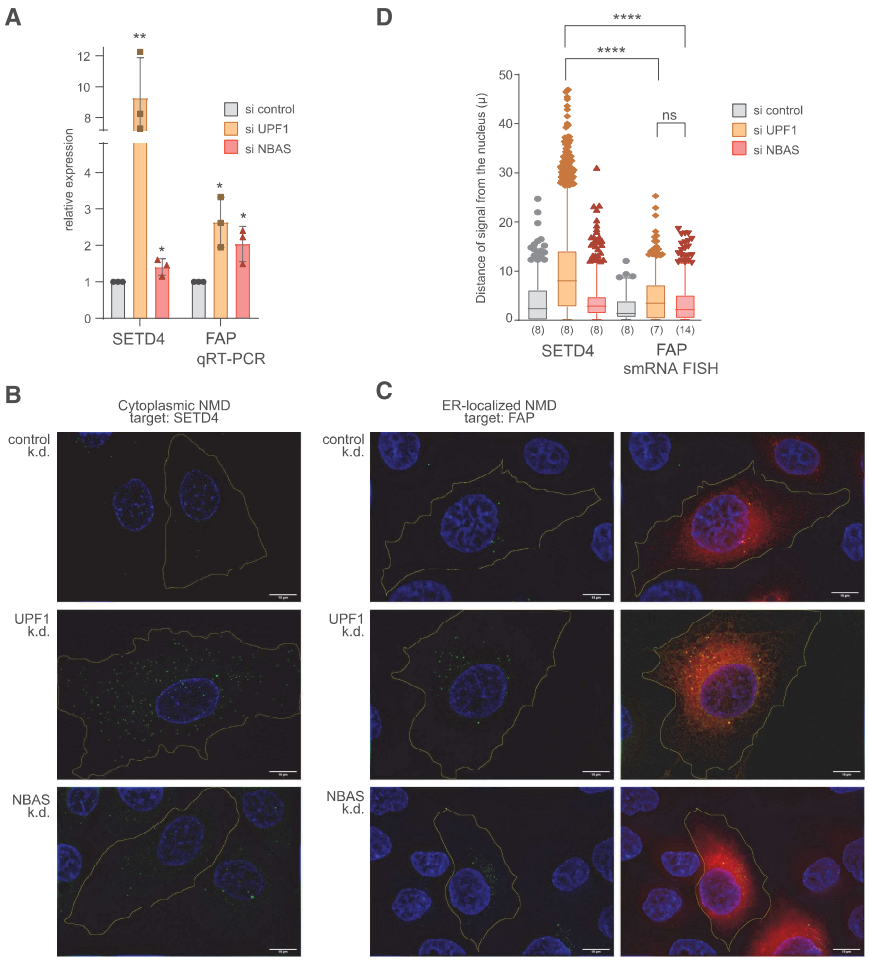


Figure 3. ER-localized NMD targets are degraded at the ER. (*A*) Relative changes in SETD4 and FAP expression upon UPF1 or NBAS depletion were measured by qRT-PCR and normalized to POLR2J expression. Each point represents one biological replica, bars indicate mean with SEM. Significance was determined by two-tailed unpaired *t*-test. (**) $P < 0.005$; (*) $P < 0.05$. (*B*,*C*) Single-molecule (sm) RNA FISH was used to visualize the site of accumulation of a cytoplasmic NMD target SETD4 (*B*), or an ER-localized NMD target FAP (*C*), within the cell when NMD was abrogated by either UPF1 or NBAS knockdown. Individual RNA molecules were visualized by a set of transcript-specific Quasar 570-labeled Stellaris probes and pseudo-colored green. The cell outline was visualized by staining with MemBright 640-nm probe and is indicated by the yellow line. (*Right* column) The ER compartment was visualized by mEmeral-ER-3 cotransfected plasmid and is pseudocolored red. Scale bars, 10 µm. (*D*) Distribution of FISH signal within imaged cells is represented by the distance of signal from the nucleus in each condition. Graph shows distances as 5–95 percentile box plots, numbers in brackets indicate the number of analyzed cells. Data were acquired in three independent experiments. Statistical significance was determined by one-way ANOVA test. (****) $P\_adj < 0.0001$; (ns) not significant.

from the distribution pattern of the ER-associated *FAP* mRNA signal that clusters close to the nuclear periphery upon depletion of both NBAS and UPF1 (Fig. 3D). Finally, using smRNA FISH signal counts we confirmed the increase of *SETD4* and *FAP* transcripts in response to UPF1 and NBAS depletion (Supplemental Fig. S4). Altogether, these results provide strong evidence that NBAS has a role in NMD regulation of mRNAs translated at the ER.

UPF1 is present at the ER

The involvement of the ER-associated factor NBAS in the NMD response, together with its coregulation of RNA targets with UPF1, led us to probe whether UPF1 localizes to the ER and interacts with NBAS. It has been previously reported that UPF1 localizes to the cytoplasm both in yeast (Atkin et al. 1995) and human cells (Applequist et al. 1997; Serin et al. 2001). However, there is evidence that UPF1 is also localized at the ER. First, UPF1 was found associated with cytoplasmic, but also with ER-bound polysomes (Jagannathan et al. 2014a). Moreover, a large-scale protein–protein interactome revealed that UPF1 interacts with components of the Syntaxin 18 complex, ZW10, p31 and STX18, where NBAS also resides (Aoki et al. 2009; Brannan et al. 2016). First, we tagged endogenous NBAS with an eGFP/ 3xFlag tag at its N terminus and following anti-GFP immunoprecipitation, we revealed the interacting proteins using mass spectrometry. This resulted in the identification of several components of the Syntaxin 18 complex as well as of SEC61A1, a component of the trimeric SEC61 complex at the ER translocon that interact with NBAS independently of the presence of RNA (Supplemental Fig. S5A; Supplemental Table S4; Hartmann et al. 1994). We did not observe a robust enrichment of NMD factors despite having demonstrated these interactions by other experimental approaches (see Fig. 5, below; Supplemental Figs. S6, S7). We hypothesize that this is due to the more transient nature of interactions between NBAS and NMD factors compared with components of the Syntaxin 18 complex.

Immunofluorescence analysis of HeLa cells transiently transfected with Flag-UPF1 showed that UPF1 localizes to the cytoplasm, as previously suggested. However, upon incubation of HeLa cells with digitonin, which permeabilizes the plasma membrane and consequently releases cytosolic components that are not anchored to cellular membranes, we observed a population of UPF1 that was resistant to digitonin treatment and colocalized with the ER marker calnexin, indicating ER membrane association (Fig. 4A). We next used the proximity ligation assay (PLA) to probe for interactions of UPF1 and NBAS, with SEC61B, a component of the SEC61 channel-forming translocon complex that is a central component of the protein translocation apparatus at the ER membrane. PLA has been extensively used to detect interactions of many cellular proteins, including the core NMD factors UPF1 and UPF2 (Tatsuno et al. 2016). We detected a robust PLA interaction between UPF1 and SEC61B in HeLa cells, indicating that UPF1 localizes at the site of mRNA translation at the ER (Fig. 4B). PLA analysis of the endogenously NBAS-tagged cell line revealed that NBAS is also colocalized with SEC61B (Supplemental Fig. S5B).

Altogether, these results show that both UPF1 and NBAS are present at the translocon, bringing them into close proximity with mRNAs being translated at the ER. Even though UPF1 is the core factor essential for NMD function, it is likely that other NMD factors will, at times, also partially localize to the ER. Previously, immunofluorescence was used to colocalize SMG6 and UPF3B with GRP78, a member of the heat shock protein 70 (HSP70) family that is present in the lumen of the ER (Sakaki et al. 2012). More recently, a large human interactome study identified UPF3B, in complex with UPF2 and UPF1, to be associated with SEC61A1 (Hein et al. 2015).

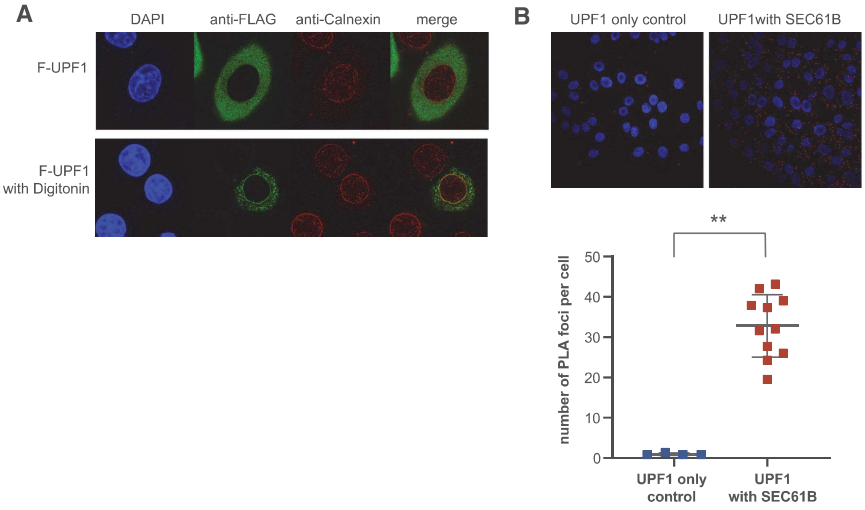


Figure 4. UPF1 localizes at the ER. (*A*) Immunofluorescence of HeLa cells transiently expressing Flag-tagged UPF1 (F-UPF1, in green) together with the ER marker Calnexin (in red). (*Top*) Cell nuclei were visualized by DAPI staining. UPF1 predominantly shows a diffused cytoplasmic localization. (*Bottom*) Partial cell permeabilization with digitonin revealed that a fraction of UPF1 is anchored at the ER membrane and colocalizes with calnexin. (*B*) UPF1 is localized in the close proximity of the SEC61B translocon component at the ER. Proximity ligation assay (PLA) using antibodies against endogenous UPF1 and SEC61B proteins generated a discrete signal (red spots), indicating that the proteins are <40 nm apart. The graph shows the quantification of PLA signal. Each point represents mean PLA count per cell in one captured frame. Significance was determined by two-tailed Mann-Whitney test. (**) $P<0.005$. Data showing the colocalization of NBAS and UPF2 with SEC61B are shown in Supplemental Figure S5.

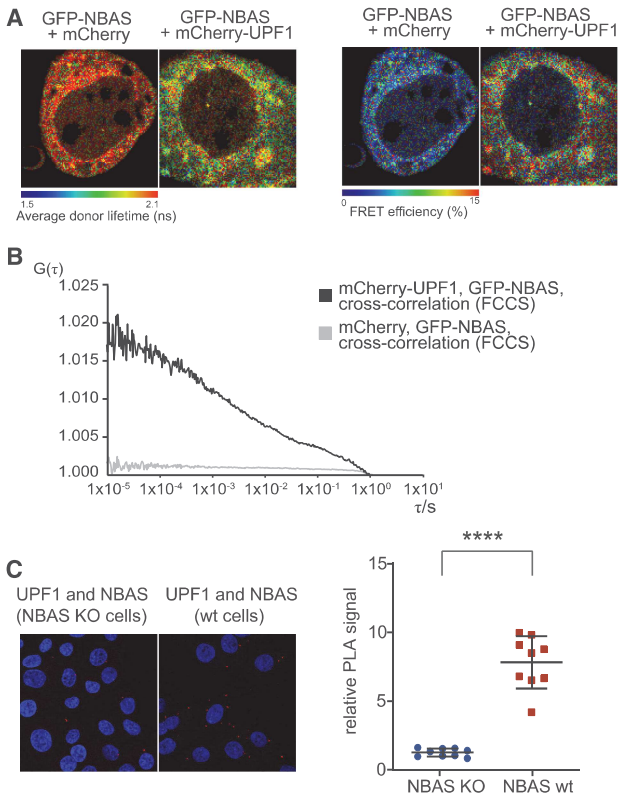


Figure 5. NBAS and UPF1 interact directly. (*A*) The interaction between NBAS and UPF1 was measured by FRET-FLIM. The average fluorescence lifetime of GFP-NBAS donor molecules was measured in HeLa cells expressing GFP-tagged NBAS together with mCherry-UPF1 or mCherry, as a control. Cells were pseudo-colored by average lifetime (ns) where red represents longer lifetime and green corresponds to a shorter lifetime. (*Right* panel) Average fluorescence lifetime was used to calculate the FRET efficiency. Cells were pseudo-colored by FRET efficiency (percentage). (*B*) The interaction between NBAS and UPF1 was measured by fluorescence cross-correlation spectroscopy (FCCS) in HeLa cells transiently expressing GFP-NBAS and mCherry-UPF1 or GFP-NBAS and mCherry, as a control. Increased G(τ) amplitude of the average cross-correlation curves corresponds to higher cross-correlation (higher fraction of codiffusing molecules) of GFP-NBAS and mCherry-UPF1 (in black), as compared with GFP-NBAS and mCherry average cross-correlation (in gray). (*C*) A direct interaction between endogenous NBAS and UPF1 was determined by PLA using antibodies against endogenous UPF1 and NBAS proteins in HeLa cells. The PLA signal in wild-type (WT) HeLa cells was compared with a HeLa NBAS knockout (KO) cells, as a negative control. The PLA signal was quantified in the graph where each point represents mean PLA count in one captured frame, relative to NBAS KO negative control. Significance was determined by two-tailed Mann-Whitney test. (****) $P < 0.0001$.

To investigate whether other components of the NMD pathway with previously established cytoplasmic localization are also present at the ER, we probed for the colocalization of UPF2 with SEC61B. We observed a modest PLA signal indicating that a fraction of UPF2 is also associated with the translocon at the ER in HeLa cells (Supplemental Fig. S5C). This modest colocalization of UPF2 with the ER is consistent with our previous observation that UPF2 plays only a minor role in the regulation of ER-translated mRNAs (Fig. 2D).

NBAS interacts with UPF1 and preferentially associates with the SURF complex

The presence of NBAS and UPF1 in proximity with the ER translocon, together with the coregulation of RNA targets by NBAS and UPF1 as part of the ER–NMD response strongly suggested an interaction of these two NMD factors. To investigate this we first tested for the copurification of NBAS with UPF1 by coimmunoprecipitation in HeLa cells, in the presence or absence of RNases. We found that Flag-tagged UPF1 coimmunoprecipitated with cotransfected T7-NBAS (Supplemental Fig. S6A) and we also detected the interaction of endogenous UPF1 with Flag-tagged NBAS, even in the absence of RNA (Supplemental Fig. S6B). Since UPF1 phosphorylation is a later step in NMD activation it can be used as a diagnostic tool to infer the timing of recruitment of a particular protein to the NMD complex. We used two UPF1 mutants that resemble the hypophosphorylated state of UPF1 when present in the early surveillance (SURF) complex (C126S), or the hyperphosphorylated UPF1 present in the late decay-inducing (DECID) complex (K498A) (Weng et al. 1996; Kashima et al. 2006). As observed previously, the C126S mutation blocks UPF1 interaction with UPF2, yet it binds to T7-tagged NBAS, whereas the K498A UPF1 mutant displayed no NBAS binding (Supplemental Fig. S6C). This strongly suggests that NBAS is preferentially associated with the initial surveillance SURF complex, where UPF1 is hypophosphorylated. Interestingly, we observed previously similar results for the RNA helicase DHX34 (Hug and Cáceres 2014), suggesting that most of the regulatory steps of the NMD pathway occur in the early stages of the NMD response. The observed biochemical interaction of NBAS with UPF1 was not RNA-dependent (Supplemental Fig. S6A,B). Nevertheless, we wanted to investigate the possibility that NBAS directly recruits NMD targets for degradation. We tested whether NBAS directly binds to mRNA using an mRNA capture assay, which relies on in-situ UV cross-linking, followed by affinity selection of mRNPs by oligo-dT cellulose (Piñol-Roma and Dreyfuss 1992; Sanford et al. 2005). Affinity selection of mRNPs by oligo-dT cellulose showed that NBAS binds to mRNA (Supplemental Fig. S6D), opening the possibility that RNA-binding by NBAS contributes to selection of NMD targets at the ER.

Next, we probed for the interaction of endogenous NBAS with other core NMD factors. We detected an interaction of transiently expressed Flag-tagged SMG5, SMG6, SMG7, and UPF1 with endogenous NBAS, even in the absence of RNA (Supplemental Fig. S7A). Additionally, we observed that endogenous NBAS can also be copurified with full-length tagged RNA helicase DHX34, but not with a truncated DHX34 version lacking an essential OB-like domain (Supplemental Fig. S7B), which is required for helicase function (Melero et al. 2016).

We used Förster resonance energy transfer (FRET) to probe the interaction of NBAS and UPF1 in HeLa cells in culture. This approach is based on the transfer of energy from the excited state of a donor fluorophore to an adjacent acceptor fluorophore when the two molecules are in the correct orientation and <10 nm apart. FRET was detected by a reduction in the amount of energy that the donor releases as fluorescence, measured by fluorescence lifetime imaging microscopy (FLIM) (Day et al. 2001; Ellis et al. 2008). Cotransfection of GFP-NBAS and mCherry-UPF1 in HeLa cells resulted in a reduction of the average donor fluorescence lifetime and an increase in the FRET efficiency, as compared with transfection of GFP-NBAS and mCherry (Fig. 5A). This experiment strongly suggests that NBAS and UPF1 interact directly. Next, we performed fluorescence cross-correlation spectroscopy (FCCS) to probe the interaction of NBAS and UPF1 in HeLa cells transiently expressing GFP-NBAS and mCherry-UPF1 or GFP-NBAS and mCherry, as a control. The cross-correlation signal is a direct indication of both molecules moving together, where a low amplitude of cross-correlation signal indicates that the labeled molecules diffuse separately, whereas high amplitude of cross-correlation signal is only achieved when both molecules are bound and diffuse together. The high cross-correlation signal observed in HeLa cells transiently expressing GFP-NBAS and mCherry-UPF1 indicated that NBAS and UPF1 interact directly (Fig. 5B, black curve), in contrast to the low amplitude of cross-correlation signal observed in cells expressing GFP-NBAS with mCherry, as a negative control (Fig. 5B, gray curve). Finally, we confirmed the interaction of endogenous NBAS and UPF1 using PLA. We observed a prominent PLA signal using antibodies against endogenous UPF1 and NBAS proteins, which was absent in cells that do not express NBAS (NBAS KO) (Fig. 5C). These results robustly show that NBAS and UPF1 interact directly in situ.

NBAS recruits UPF1 to the membrane of the ER

To probe the functional consequences of the NBAS–UPF1 interaction in vivo we used fluorescence correlation spectroscopy (FCS). This is a noninvasive method with single-molecule sensitivity that allows the analysis of the dynamic behavior of fluorescent molecules with high temporal resolution and at low, physiologically relevant concentrations in live cells (Liu et al. 2015; Papadopoulos et al. 2019). When the recorded fluorescence intensity fluctuations are caused by molecular movement, FCS measurements can be used to measure molecular mobility/diffusion rate of fluorescent molecules in a subfemtoliter detection volume in live cells. Thus, free, fast diffusing molecular movement is reflected by autocorrelation curves that display decays in shorter characteristic times (Fig. 6A, green curve), whereas bound or slow diffusing molecules are characterized by autocorrelation curves with slower decay times (Fig. 6A, orange curve). We performed FCS measurements to derive the mobility of transiently transfected mCherry-UPF1, or mCherry control at the ER, in the perinuclear region of the cell, or in the cytoplasm at the cell periphery, in HeLa cells (Fig. 6A, right panel). We observed that the mobility of mCherry-UPF1 in the cell periphery was very similar, irrespective of the levels of NBAS protein (Fig. 6B, cf. si-control, mCherry-UPF1, and periphery [in gray] with si-NBAS, mCherry-UPF1, and periphery [in black]). In contrast, a significantly slower mobility of mCherry-UPF1was observed at the ER in control cells bearing physiological levels of NBAS, as

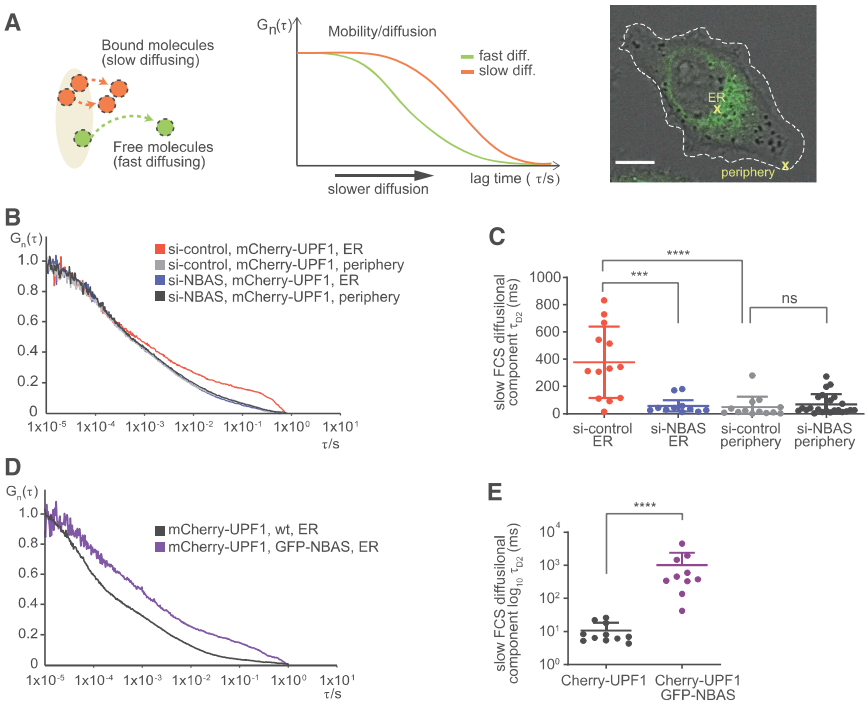


Figure 6. NBAS recruits UPF1 to the ER. (*A*) Schematic representation of the principle of fluorescence correlation spectroscopy (FCS). See the Materials and Methods for details. (*B*) Average autocorrelation curves of mCherry-UPF1 at the ER or cell periphery in the presence (si-control) or absence of NBAS (si-NBAS) are displayed. (*C*) Graph showing the derived characteristic decay times τ_{D2} (slow FCS component) of mCherry-UPF1, after fitting of the autocorrelation curves with a two-component model for diffusion (plus triplet correction). Significance was determined by the two-tailed Mann-Whitney test. (****) $P < 0.0001$; (***) $P < 0.001$; (ns) not significant. (*D*) Average autocorrelation curves of mCherry-UPF1 at the ER with physiological (WT) or increased (GFP-NBAS) level of NBAS expression. (*E*) Characteristic decay times τ_{D2} (slow FCS component), derived as in *C* of mCherry-UPF1 at the ER. Significance was determined by the two-tailed Mann-Whitney test. (****) $P < 0.0001$. Data showing FCS control experiments are in Supplemental Figure S8.

reflected by a marked shift of the autocorrelation curve toward longer characteristic times (Fig. 6B, red line). Depletion of NBAS increased the mobility of mCherry-UPF1 at the ER (Fig. 6B, blue line), so that the diffusion time of mCherry-UPF1 was uniform within the cell. In contrast, the mobility of mCherry alone was uniform throughout the cell and was not affected by the level of NBAS expression (Supplemental Fig. S8A,B). Fitting of the autocorrelation curves with a two-component diffusion model revealed that the characteristic decay time τ_{D2} of mCherry-UPF1 was considerably longer at the ER in the presence of NBAS (Fig. 6C, in red), whereas NBAS depletion caused significant reduction of the τ_{D2} mCherry-UPF1 decay times (Fig. 6C, in blue). The τ_{D2} decay time of mCherry-UPF1 was markedly shorter at the cell periphery than at the ER and was not affected by the depletion of NBAS (Fig. 6C). The observed reduced mobility of mCherry-UPF1 at the ER strongly suggests that UPF1 is part of a slower diffusing protein complex, and that NBAS is required for the retention of UPF1 at the ER. In agreement, overexpression of NBAS led to even slower mCherry-UPF1 mobility at the ER (Fig. 6D, cf. WT NBAS level, mCherry-UPF1, and ER [in black] with overexpressed GFP-NBAS, mCherry-UPF1, and ER [in purple]). In accordance, the characteristic decay time τ_{D2} of mCherry-UPF1 was significantly increased by the addition of GFP-NBAS (Fig. 6E). These results show that ER-localized NBAS acts to recruit UPF1 to the ER to activate a localized NMD response (Fig. 7).

Discussion

An NMD response at the ER

The targeting of mRNAs to specific subcellular sites for local translation has an important role in many cellular processes and during development (Decker and Parker 2006). In particular, translation of secreted and integral membrane proteins occurs on ER-bound ribosomes, whereas cytosolic protein synthesis mostly occurs on ribosomes dispersed throughout the cytoplasm. A subset of transcripts that are targeted for translation at the ER encode a leading signal peptide that is recognized after emerging from the ribosome by the signal recognition particle (SRP). These SRP-bound transcripts are subsequently targeted to the ER, and the resulting translated peptides are translocated across or inserted into the membrane by the SEC61 translocation channel (Walter and Johnson 1994). However, recent evidence showed that mRNAs encoding cytosolic proteins can also be translated by ER-associated ribosomes, suggesting a more diverse role for the ER in mRNA translation (Jagannathan et al. 2014b; Reid and Nicchitta 2015).

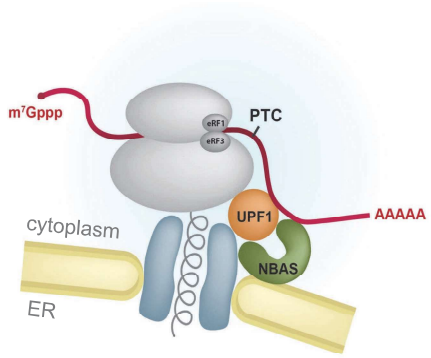


Figure 7. Model for an ER–NMD pathway. NBAS is an NMD factor that localizes to the membrane of the endoplasmic reticulum (ER). NBAS recruits the core NMD factor UPF1 and tethers it to the membrane of the ER to activate a local NMD response that targets for degradation mRNAs that are translated at the ER.

Despite the ER representing a specialized environment for the translation of a large proportion of cellular mRNAs, it remains unclear how RNA quality control pathways in general, and NMD in particular, operate on those mRNAs. Since NMD is dependent on active translation, transcripts coding for secreted and integral membrane proteins will not have sufficient exposure to the NMD quality control in the cytoplasm due to the intrinsic nature of ER-associated translation, which is spatially and temporally distinct from mRNA translation in the cytosol. In fact, ER targeted transcripts are only translationally active once they encounter the ER membrane (Wu et al. 2016). Thus, ER targeted transcripts will fail to undergo NMD in the cytoplasm, until their translation is activated at the ER membrane. We propose that these transcripts are instead targeted for degradation by an ER–NMD dedicated pathway that is responsible for the quality control of cellular mRNAs that are translated at the ER.

We showed previously that NBAS together with core NMD factors regulates the stability of a large number of endogenous RNA targets that are preferentially linked to cellular stress and membrane trafficking in nematodes, zebrafish, and human cells (Anastasaki et al. 2011; Longman et al. 2013). We also showed that NBAS contributes to a negative feedback regulatory network, in which the NMD pathway controls the levels of transcripts encoding NMD factors (Huang et al. 2011; Yepiskoposyan et al. 2011; Longman et al. 2013). Here, we used RNA-sequencing to profile changes in mRNA abundance upon depletion of NBAS or UPF1. Importantly, this analysis revealed that a large proportion of NBAS regulated targets were coregulated by UPF1, establishing them as candidate NMD targets (Fig. 2). These results, together with evidence provided by FCS experiments demonstrates that NBAS, which is anchored at the membrane of the ER, promotes the recruitment of the core NMD factor UPF1 and thereby activates the NMD response at the ER (Fig. 6). The decay step in NMD involves the recruitment of nucleases along two pathways that involve endonucleolytic decay mediated by SMG6 (Huntzinger et al. 2008; Eberle et al. 2009), or alternatively exonucleolytic decay mediated by the SMG5–SMG7 heterodimer (Loh et al. 2013). Interestingly, despite NBAS being preferentially associated with hypophosphorylated UPF1 in the SURF complex, we also detected its interaction with SMG5, SMG6, and SMG7 (Supplemental Fig. S7), which are required for mRNA degradation at the later stages of NMD.

Altogether, this suggests that the presence of several core NMD factors is required for an active NMD pathway in the vicinity of the ER. In summary, we identified a localized NMD response at the ER that acts on transcripts localized and translated at the ER, which we term ER–NMD (Fig. 7).

Biological role of ER–NMD pathway

Recent evidence supports a role for NMD in modulating the ER stress response by ensuring appropriate activation of the unfolded protein response (UPR) (Goetz and Wilkinson 2017). This pathway is able to sense and respond to excessive amounts of misfolded proteins in the ER (Walter and Ron 2011). Interestingly, mRNAs encoding several UPR components are targeted by NMD, including the UPR sensor IRE1α, as well as ATF-4 and CHOP that are activated by PERK branch signaling. Thus, NMD can act to control the threshold of cellular stress that is necessary to activate the UPR (Gardner 2008; Karam et al. 2015; Sieber et al. 2016). As a consequence, appropriate level of NMD activity is required to protect cells from detrimental stress response activation. In particular, pathological UPR activation is recognized as a major causal mechanism in neurodegenerative diseases including Alzheimer's, Parkinson's, and prion diseases, which are associated with the accumulation of misfolded disease-specific proteins (Hetz and Mollereau 2014). A chronic activation of PERK/eIF2α-P signaling results in a sustained reduction of global protein synthesis rates in neurons, caused by the phosphorylation and inactivation of eIF2α, which blocks translation at the level of initiation. Inappropriate eIF2α phosphorylation contributes to disease pathogenesis, and is observed in patients' brains, and in mouse models of protein-misfolding neurodegenerative diseases (Hoozemans et al. 2009; Moreno et al. 2012). Thus, limiting the chronic activation of the UPR could have a neuroprotective effect in neurodegenerative diseases.

We hypothesize that modulation of the activity of the ER–NMD pathway could be crucial to regulate the response to ER stress. It is tempting to speculate that increasing the activity of the ER–NMD pathway will have a neuroprotective effect by limiting excessive UPR signaling in neurodegeneration. Indeed, a protective role for UPF1 has already been shown in primary neuronal models of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) induced by overexpression of the RNA-binding proteins, TDP43 and FUS (Barmada et al. 2015). Moreover, NMD has been shown to protect against the effects of hexanucleotide repeat expansion in C9orf72 (C9-HRE) found in *Drosophila* and cellular models of ALS/FTD (Xu et al. 2019; Ortega et al. 2020).

In summary, we uncovered a localized NMD response at the ER, which provides spatial and temporal separation from cytoplasmic NMD acting on mRNAs that are translated at the ER. This could have important implications to protect cells from ER stress/dysfunction that results from the accumulation of aberrantly processed RNAs, leading to the production of truncated proteins and affecting cellular homeostasis. This could be of particular relevance for many disease processes, such as neurodegenerative diseases, where an exacerbated ER stress is observed. Future studies will aim to further characterize this novel ER–NMD pathway; investigating its physiological role, as well as the biological consequences of manipulating its activity.

Material and methods

NMD reporter assays

HeLa cells stably expressing HBB WT or HBB NS39 reporters were mock-depleted or depleted twice of UPF2, NBAS, STX18, and STX5, and harvested 5 d after the first depletion. Alternatively, HeLa HBB NS39/WT-expressing cells were treated with 2 µg/mL actinomycin D (Sigma A9415) to block transcription and/or 2 µg/mL Brefeldin A (Sigma B5936) to block secretion for 6 h. The stability of HBB reporters was determined by qRT-PCR. The fluorescence-based NMD assay was performed as described previously (Pereverzev et al. 2015). NMD activity was determined by quantitative measurement of red and green fluorescence using FACS BD LSR-Fortessa X-20 SORP with BD FACSDiva software. Gates were set using a nonfluorescent control. Compensation was applied in BD FACSDiva software using single-transfected cells. Green fluorescence expression (488-525/50) was analyzed by gating the single–cell population in SSC-H/SSC-A scatter plot, followed by debris exclusion gate in SSC-A/FSC-A scatter plot, followed by gating only the red fluorescent-positive cells in 405-450/50 (autofluorescence) and 561-610/20 dot plot. Gate settings were kept constant during the experiment. Data were analyzed with FlowJo software (version 10.6.0). Between 4500 and 12,000 red fluorescent cells were analyzed in each biological replica. NMD disruption was determined by the increase of green fluorescence in red cells, normalized to mock-depleted cells.

Single-molecule (sm) RNA FISH

smRNA FISH was performed following the Stellaris RNA FISH protocol for adherent cells (http://www.biosearchtech.com/stellarisprotocols) using RNase-free reagents. Transcript-specific Quasar 570-labeled Stellaris probes were designed with the Stellaris RNA FISH probe designer (Biosearch Technologies, Inc.; http://www.biosearchtech.com/stellarisdesigner), and resuspended in TE buffer to make 12.5 µM probe stock. Probe sequences are listed in Supplemental Table S5. HeLa cells were depleted as described above, with the addition of mEmerald-ER-3 (Addgene plasmid 54082; http://n2t.net/addgene:54082; RRID: Addgene 54082) at 1 µg per well in six-well plates during the second round of depletions. Cells were seeded on high-precision coverslips (Marienfeld 0107052), washed twice with PBS, and fixed with 4% formaldehyde for 10 min at room temperature. Cells were permeabilized in 70% ethanol for at least 1 h and hybridization of probes was performed in a humidified chamber for 15 h at 37°C. Probe hybridization and all subsequent steps were carried out in the dark. Following hybridization, coverslips were washed as described in the protocol. Nuclei were counterstained with 5 ng/mL of DAPI in buffer A for 30 min. Cell outlines were determined by staining with the addition of 20 nM MemBright 640-nm dye (Idylle MCO-MEM-640-1902) (Collot et al. 2019) in buffer B for 15 min at room temperature. Coverslips were mounted in VectaShield and sealed with nail polish.

Proximity ligation assay (PLA)

PLA assay was performed following Duolink PLA fluorescence protocol (Sigma-Aldrich). Cells were grown, fixed, and permeabilized as for immunofluorescence. Coverslips were then incubated for 1 h with PLA block buffer provided with the Duolink in situ PLA probes at room temperature and incubated with primary antibodies. Coverslips were then washed three times with Duolink wash buffer A and Duolink probe incubation, ligation, and PLA signal amplification were performed using Duolink in situ detection reagent Red kit according to the manufacturer's instructions. In the last wash, coverslips were incubated with DAPI at 50 ng/mL in 0.01× wash buffer B for 5 min, mounted in VectraShield, and sealed with nail varnish.

FRET-FLIM

HeLa cells transiently expressing GFP-NBAS with mCherry or mCherry-UPF1 were grown on slides for 48 h after transfection, fixed in 4% PFA for 10 min, washed in PBS, and mounted in VectaShield with DAPI to visualize cell nuclei. Fluorescence lifetime images were acquired on a Leica SP5 SMD confocal laser-scanning microscope fitted with a time-correlated single photon counting module (PicoHarp 300) using a 63/1.4 numeric aperture HCX PL Apo oil immersion objective lens, as previously described (Saleeb et al. 2019). The donor EGFP was excited using a tunable white light supercontinuum laser operating at 488 nm and pulsing at 40 MHz. Emission was detected with an external single-photon avalanche diode (MicroPhoton Devices). Single-pixel fluorescence lifetime analyses were carried out with SymPhoTime version 5.4.4 (PicoQuant).

FCS and FCCS

FCS measurements were performed by recording fluorescence intensity fluctuations in a very small, approximately ellipsoidal observation volume element (OVE) (~0.2 µm wide and 1 µm long) that is generated in HeLa cells by focusing the laser light through the microscope objective and by collecting the fluorescence light through the same objective using a pinhole in front of the detector to block out-of-focus light. The fluorescence intensity fluctuations caused by fluorescently labeled molecules passing through the OVE were analyzed using temporal autocorrelation analysis. HeLa cells were mock-depleted or depleted of NBAS in two rounds of depletions. During the second round of depletions cells were cotransfected with 1 µg of mCherry control or of mCherry-UPF1plasmids. For NBAS overexpression, cells were cotransfected with GFP-NBAS and mCherry or mCherry-UPF1 (1 µg of each). Prior to FCS/FCCS measurement, cells were grown overnight on chambered coverslips (µ-slide, eight-well, Ibidi) and the growth medium was replaced with L-15 medium (Leibovitz) (Sigma-Aldrich L1518) immediately prior to FCS/FCCS measurements. Where appropriate, ER was visualized by the addition of 1 µM ER tracker (green) dye (Thermo Fisher E34251). Measurements were made only in weakly expressing cells. Fluorescence microscopy imaging of HeLa cells and FCS measurements were performed on a PicoQuant modified Leica SP5 microscope, which features Avalanche PhotoDiodes that enable close to single-photon detection (Vukojević et al. 2008). Fluorescence intensity fluctuations were recorded for 50 or 100 sec. Autocorrelation curves were analyzed using the SymphoTime 64 software package (PicoQuant). Control FCS measurements to assess the detection volume were routinely performed prior to data acquisition, using dilute solutions (10 nM and 20 nM, respectively) of Alexa488 and Alexa568 dyes (Petrášek and Schwille 2008). The variability between independent measurements reflects the variability between cells, rather than imprecision of FCS measurements.

Quantification and statistical analysis

For information about the number of replicates, the meaning of error bars (e.g., standard error of the mean) and other relevant statistical analysis, see the corresponding figure legend. For information about how data were analyzed and/or quantified, see the relevant section in the Materials and Methods and/or the figure legend. GraphPad Prism software was used for the statistical analysis in Figures 1, C and D; 3A; 4B; 5C; and 6, C and E, and Supplemental Figures S1, B and C; S4; and S5, A and B. RStudio was used for the statistical analysis in Figure 2, A, B, and D, and Supplemental Figures S2 and S3.

Data availability

All RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE152437.

Acknowledgments

We are grateful to Andrew Peden (University of Sheffield) for the gift of HeLa C1 cells, a secretion reporter, and for advice on secretion experiments, and to Konstantin Lukyanov (Moscow) for the gift of the NMD$^+$ reporter. We thank Matthew Pearson (advanced imaging), Elisabeth Freyer (flow cytometry), and Craig Nicol (graphics) from the MRC Institute of Genetics and Molecular Medicine core support facilities for technical assistance and advice; Shelagh Boyle for her help with the smRNA FISH setup; and Alison Dun and Rory Duncan at the MRC-funded Edinburgh Super-Resolution Imaging Consortium (ESRIC) for expertise and infrastructure. We are grateful to our MRC Human Genetics Unit colleagues, Andrew Jackson, Ian Adams, and Wendy Bickmore for critical reading of the manuscript. This work was supported by a Chancellor's fellowship of the University of Edinburgh (to D.K.P.) and an Institutional Strategic Research Fund (ISSF3; to J.J.S.) and core funding to the MRC Human Genetics Unit from the Medical Research Council (to J.F.C.).

Author contributions: D.L. and J.F.C. conceived and designed the project. D.L. performed secretion assays, NMD assays, RNA FISH, and cell-biological, FRET-FLIM, and FCS experiments. K.A.J.-J. and M.M.M. carried out RNA-seq profiling of NBAS and UPF1 knockdown cells, K.A.J.-J. generated the endogenously tagged NBAS cell line and together with R.S.Y. and M.S.T performed all bioinformatic and statistical analysis. N.H. analyzed the interaction of NBAS with core NMD factors. L.C.M. performed the analysis of FISH signal distribution. D.K.P. and J.J.S. performed, analyzed, and interpreted the FCS and FCCS experiments. The manuscript was cowritten by all of the authors.

References

Alexandrov A, Shu M-D, Steitz JA. 2017. Fluorescence amplification method for forward genetic discovery of factors in human mRNA degradation. *Mol Cell* **65:** 191–201. doi:10.1016/j.molcel.2016.11.032

Anastasaki C, Longman D, Capper A, Patton EE, Cáceres JF. 2011. Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. *Nucleic Acids Res* **39:** 3686–3694. doi:10.1093/nar/gkq1319

Aoki T, Ichimura S, Itoh A, Kuramoto M, Shinkawa T, Isobe T, Tagaya M. 2009. Identification of the neuroblastoma-

amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport. *Mol Biol Cell* **20:** 2639–2649. doi:10.1091/mbc.e08-11-1104

Applequist SE, Selg M, Raman C, Jäck HM. 1997. Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein. *Nucleic Acids Res* **25:** 814–821. doi:10.1093/nar/25.4.814

Aridor M, Bannykh SI, Rowe T, Balch WE. 1995. Sequential coupling between COPII and COPI vesicle coats in endoplasmic reticulum to Golgi transport. *J Cell Biol* **131:** 875–893. doi:10.1083/jcb.131.4.875

Atkin AL, Altamura N, Leeds P, Culbertson MR. 1995. The majority of yeast UPF1 co-localizes with polyribosomes in the cytoplasm. *Mol Biol Cell* **6:** 611–625. doi:10.1091/mbc.6.5.611

Baird TD, Cheng KC-C, Chen Y-C, Buehler E, Martin SE, Inglese J, Hogg JR. 2018. ICE1 promotes the link between splicing and nonsense-mediated mRNA decay. *Elife* **7:** e33178. doi:10.7554/eLife.33178

Balasubramanian M, Hurst J, Brown S, Bishop NJ, Arundel P, DeVile C, Pollitt RC, Crooks L, Longman D, Caceres JF, et al. 2017. Compound heterozygous variants in NBAS as a cause of atypical osteogenesis imperfecta. *Bone* **94:** 65–74. doi:10.1016/j.bone.2016.10.023

Barmada SJ, Ju S, Arjun A, Batarse A, Archbold HC, Peisach D, Li X, Zhang Y, Tank EMH, Qiu H, et al. 2015. Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc Natl Acad Sci* **112:** 7821–7826. doi:10.1073/pnas.1509744112

Bhuvanagiri M, Schlitter AM, Hentze MW, Kulozik AE. 2010. NMD: RNA biology meets human genetic medicine. *Biochem J* **430:** 365–377. doi:10.1042/BJ20100699

Brannan KW, Jin W, Huelga SC, Banks CAS, Gilmore JM, Florens L, Washburn MP, Van Nostrand EL, Pratt GA, Schwinn MK, et al. 2016. SONAR discovers RNA-binding proteins from analysis of large-scale protein-protein interactomes. *Mol Cell* **64:** 282–293. doi:10.1016/j.molcel.2016.09.003

Casadio A, Longman D, Hug N, Delavaine L, Vallejos Baier R, Alonso CR, Cáceres JF. 2015. Identification and characterization of novel factors that act in the nonsense-mediated mRNA decay pathway in nematodes, flies and mammals. *EMBO Rep* **16:** 71–78. doi:10.15252/embr.201439183

Çivril F, Wehenkel A, Giorgi FM, Santaguida S, Di Fonzo A, Grigorean G, Ciccarelli FD, Musacchio A. 2010. Structural analysis of the RZZ complex reveals common ancestry with multisubunit vesicle tethering machinery. *Structure* **18:** 616–626. doi:10.1016/j.str.2010.02.014

Colak D, Ji S-J, Porse BT, Jaffrey SR. 2013. Regulation of axon guidance by compartmentalized nonsense-mediated mRNA decay. *Cell* **153:** 1252–1265. doi:10.1016/j.cell.2013.04.056

Collot M, Ashokkumar P, Anton H, Boutant E, Faklaris O, Galli T, Mély Y, Danglot L, Klymchenko AS. 2019. MemBright: a family of fluorescent membrane probes for advanced cellular imaging and neuroscience. *Cell Chem Biol* **26:** 600–614.e7. doi:10.1016/j.chembiol.2019.01.009

Day RN, Periasamy A, Schaufele F. 2001. Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus. *Methods* **25:** 4–18. doi:10.1006/meth.2001.1211

Decker CJ, Parker R. 2006. CAR-1 and trailer hitch: driving mRNP granule function at the ER? *J Cell Biol* **173:** 159–163. doi:10.1083/jcb.200601153

Eberle AB, Lykke-Andersen S, Mühlemann O, Jensen TH. 2009. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* **16:** 49–55. doi:10.1038/nsmb.1530

Ellis JD, Llères D, Denegri M, Lamond AI, Cáceres JF. 2008. Spatial mapping of splicing factor complexes involved in exon and intron definition. *J Cell Biol* **181:** 921–934. doi:10.1083/jcb.200710051

Faria JA, Corrêa NC, de Andrade C, de Angelis Campos AC, Dos Santos de Almeida R, Rodrigues TS, de Goes AM, Gomes DA, Silva FP. 2013. SET domain-containing protein 4 (SETD4) is a newly identified cytosolic and nuclear lysine methyltransferase involved in breast cancer cell proliferation. *J Cancer Sci Ther* **5:** 058–065.

Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY. 2019. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* **178:** 473–490.e26. doi:10.1016/j.cell.2019.05.027

Gardner LB. 2008. Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. *Mol Cell Biol* **28:** 3729–3741. doi:10.1128/MCB.02284-07

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* **130:** 179–191. doi:10.1016/j.cell.2007.05.028

Goetz AE, Wilkinson M. 2017. Stress and the nonsense-mediated RNA decay pathway. *Cell Mol Life Sci* **74:** 3509–3531. doi:10.1007/s00018-017-2537-6

Goldstein LA, Ghersi G, Piñeiro-Sánchez ML, Salamone M, Yeh Y, Flessate D, Chen WT. 1997. Molecular cloning of seprase: a serine integral membrane protease from human melanoma. *Biochim Biophys Acta–Mol Basis Dis* **1361:** 11–19. doi:10.1016/S0925-4439(97)00032-X

Gordon DE, Bond LM, Sahlender DA, Peden AA. 2010. A targeted siRNA screen to identify SNAREs required for constitutive secretion in mammalian cells. *Traffic* **11:** 1191–1204. doi:10.1111/j.1600-0854.2010.01087.x

Haack TB, Staufner C, Köpke MG, Straub BK, Kölker S, Thiel C, Freisinger P, Baric I, McKiernan PJ, Dikow N, et al. 2015. Biallelic mutations in NBAS cause recurrent acute liver failure with onset in infancy. *Am J Hum Genet* **97:** 163–169. doi:10.1016/j.ajhg.2015.05.009

Hartmann E, Sommer T, Prehn S, Görlich D, Jentsch S, Rapoport TA. 1994. Evolutionary conservation of components of the protein translocation complex. *Nature* **367:** 654–657. doi:10.1038/367654a0

Hein MY, Hubner NC, Poser I, Cox J, Nagaraj N, Toyoda Y, Gak IA, Weisswange I, Mansfeld J, Buchholz F, et al. 2015. A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163:** 712–723. doi:10.1016/j.cell.2015.09.053

Hetz C, Mollereau B. 2014. Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases. *Nat Rev Neurosci* **15:** 233–249. doi:10.1038/nrn3689

Hoozemans JJM, van Haastert ES, Nijholt DAT, Rozemuller AJM, Eikelenboom P, Scheper W. 2009. The unfolded protein response is activated in pretangle neurons in Alzheimer's disease hippocampus. *Am J Pathol* **174:** 1241–1251. doi:10.2353/ajpath.2009.080814

Huang L, Lou C-H, Chan W, Shum EY, Shao A, Stone E, Karam R, Song H-W, Wilkinson MF. 2011. RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol Cell* **43:** 950–961. doi:10.1016/j.molcel.2011.06.031

Hug N, Cáceres JF. 2014. The RNA helicase DHX34 activates NMD by promoting a transition from the surveillance to the

decay-Inducing complex. *Cell Rep* **8:** 1845–1856. doi:10.1016/j.celrep.2014.08.020

Hug N, Longman D, Cáceres JF. 2016. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* **44:** 1483–1495. doi:10.1093/nar/gkw010

Huntzinger E, Kashima I, Fauser M, Saulière J, Izaurralde E. 2008. SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* **14:** 2609–2617. doi:10.1261/rna.1386208

Jagannathan S, Hsu JC-C, Reid DW, Chen Q, Thompson WJ, Moseley AM, Nicchitta CV. 2014a. Multifunctional roles for the protein translocation machinery in RNA anchoring to the endoplasmic reticulum. *J Biol Chem* **289:** 25907–25924. doi:10.1074/jbc.M114.580688

Jagannathan S, Reid DW, Cox AH, Nicchitta CV. 2014b. De novo translation initiation on membrane-bound ribosomes as a mechanism for localization of cytosolic protein mRNAs to the endoplasmic reticulum. *RNA* **20:** 1489–1498. doi:10.1261/rna.045526.114

Jan CH, Williams CC, Weissman JS. 2014. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346:** 1257521. doi:10.1126/science.1257521

Karam R, Lou C-HC-H, Kroeger H, Huang L, Lin JH, Wilkinson MF. 2015. The unfolded protein response is shaped by the NMD pathway. *EMBO Rep* **16:** 599–609. doi:10.15252/embr.201439696

Karousis ED, Mühlemann O. 2019. Nonsense-mediated mRNA decay begins where translation ends. *Cold Spring Harb Perspect Biol* **11:** a032862. doi:10.1101/cshperspect.a032862

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* **20:** 355–367. doi:10.1101/gad.1389006

Kurosaki T, Popp MW, Maquat LE. 2019. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* **20:** 406–420. doi:10.1038/s41580-019-0126-2

Liu Z, Lavis LD, Betzig E. 2015. Imaging live-cell dynamics and structure at the single-molecule level. *Mol Cell* **58:** 644–659. doi:10.1016/j.molcel.2015.02.033

Loh B, Jonas S, Izaurralde E. 2013. The SMG5–SMG7 heterodimer directly recruits the CCR4–NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* **27:** 2125–2138. doi:10.1101/gad.226951.113

Longman D, Plasterk RHA, Johnstone IL, Cáceres JF. 2007. Mechanistic insights and identification of two novel factors in the *C. elegans* NMD pathway. *Genes Dev* **21:** 1075–1085. doi:10.1101/gad.417707

Longman D, Hug N, Keith M, Anastasaki C, Patton EE, Grimes G, Cáceres JF. 2013. DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and *Caenorhabditis elegans*. *Nucleic Acids Res* **41:** 8319–8331. doi:10.1093/nar/gkt585

Maksimova N, Hara K, Nikolaeva I, Chun-Feng T, Usui T, Takagi M, Nishihira Y, Miyashita A, Fujiwara H, Oyama T, et al. 2010. Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huët anomaly. *J Med Genet* **47:** 538–548. doi:10.1136/jmg.2009.074815

Melero R, Hug N, López-Perrote A, Yamashita A, Cáceres JF, Llorca O. 2016. The RNA helicase DHX34 functions as a scaffold for SMG1-mediated UPF1 phosphorylation. *Nat Commun* **7:** 10585. doi:10.1038/ncomms10585

Metze S, Herzog VA, Ruepp M-D, Mühlemann O. 2013. Comparison of EJC-enhanced and EJC-independent NMD in human cells reveals two partially redundant degradation pathways. *RNA* **19:** 1432–1448. doi:10.1261/rna.038893.113

Moreno JA, Radford H, Peretti D, Steinert JR, Verity N, Martin MG, Halliday M, Morgan J, Dinsdale D, Ortori CA, et al. 2012. Sustained translational repression by eIF2α-P mediates prion neurodegeneration. *Nature* **485:** 507–511. doi:10.1038/nature11058

Ortega JA, Daley EL, Kour S, Samani M, Tellez L, Smith HS, Hall EA, Esengul YT, Tsai Y-H, Gendron TF, et al. 2020. Nucleocytoplasmic proteomic analysis uncovers eRF1 and nonsense-mediated decay as modifiers of ALS/FTD C9orf72 toxicity. *Neuron* **106:** 90–107.e13. doi:10.1016/j.neuron.2020.01.020

Papadopoulos DK, Skouloudaki K, Engström Y, Terenius L, Rigler R, Zechner C, Vukojević V, Tomancak P. 2019. Control of Hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch. *Development* **146:** dev168179. doi:10.1242/dev.168179

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* **5:** 7729. doi:10.1038/srep07729

Petrášek Z, Schwille P. 2008. Precise measurement of diffusion coefficients using scanning fluorescence correlation spectroscopy. *Biophys J* **94:** 1437–1448. doi:10.1529/biophysj.107.108811

Piñol-Roma S, Dreyfuss G. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355:** 730–732. doi:10.1038/355730a0

Reid DW, Nicchitta CV. 2012. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *J Biol Chem* **287:** 5518–5527. doi:10.1074/jbc.M111.312280

Reid DW, Nicchitta CV. 2015. Diversity and selectivity in mRNA translation on the endoplasmic reticulum. *Nat Rev Mol Cell Biol* **16:** 221–231. doi:10.1038/nrm3958

Sakaki K, Yoshina S, Shen X, Han J, DeSantis MR, Xiong M, Mitani S, Kaufman RJ. 2012. RNA surveillance is required for endoplasmic reticulum homeostasis. *Proc Natl Acad Sci* **109:** 8079–8084. doi:10.1073/pnas.1110589109

Saleeb RS, Kavanagh DM, Dun AR, Dalgarno PA, Duncan RR. 2019. A VPS33A-binding motif on syntaxin 17 controls autophagy completion in mammalian cells. *J Biol Chem* **294:** 4188–4201. doi:10.1074/jbc.RA118.005947

Sanford JR, Ellis JD, Cazalla D, Cáceres JF. 2005. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor 2/alternative splicing factor. *Proc Natl Acad Sci* **102:** 15042–15047. doi:10.1073/pnas.0507827102

Scott DK, Board J, Lu X, Pearson AD, Kenyon RM, Lunec J. 2003. The neuroblastoma amplified gene, NAG: genomic structure and characterisation of the 7.3 kb transcript predominantly expressed in neuroblastoma. *Gene* **307:** 1–11. doi:10.1016/S0378-1119(03)00459-1

Segarra NG, Ballhausen D, Crawford H, Perreau M, Campos-Xavier B, van Spaendonck-Zwarts K, Vermeer C, Russo M, Zambelli P-Y, Stevenson B, et al. 2015. NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver,

immune system, and retina. *Am J Med Genet Part A* **167:** 2902–2912. doi:10.1002/ajmg.a.37338

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* **21:** 209–223. doi:10.1128/MCB.21.1.209-223.2001

Sieber J, Hauer C, Bhuvanagiri M, Leicht S, Krijgsveld J, Neu-Yilik G, Hentze MW, Kulozik AE. 2016. Proteomic analysis reveals branch-specific regulation of the unfolded protein response by nonsense-mediated mRNA decay. *Mol Cell Proteomics* **15:** 1584–1597. doi:10.1074/mcp.M115.054056

Singh G, Jakob S, Kleedehn MG, Lykke-Andersen J. 2007. Communication with the exon-junction complex and activation of nonsense-mediated decay by human Upf proteins occur in the cytoplasm. *Mol Cell* **27:** 780–792. doi:10.1016/j.molcel.2007.06.030

Tatsuno T, Nakamura Y, Ma S, Tomosugi N, Ishigaki Y. 2016. Nonsense-mediated mRNA decay factor Upf2 exists in both the nucleoplasm and the cytoplasm. *Mol Med Rep* **14:** 655–660. doi:10.3892/mmr.2016.5331

Trcek T, Sato H, Singer RH, Maquat LE. 2013. Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev* **27:** 541–551. doi:10.1101/gad.209635.112

Trecartin RF, Liebhaber SA, Chang JC, Lee KY, Kan YW, Furbetta M, Angius A, Cao A. 1981. β Zero thalassemia in Sardinia is caused by a nonsense mutation. *J Clin Invest* **68:** 1012–1017. doi:10.1172/JCI110323

Vukojević V, Heidkamp M, Ming Y, Johansson B, Terenius L, Rigler R. 2008. Quantitative single-molecule imaging by confocal laser scanning microscopy. *Proc Natl Acad Sci* **105:** 18176–18181. doi:10.1073/pnas.0809250105

Walter P, Johnson AE. 1994. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. *Annu Rev Cell Biol* **10:** 87–119. doi:10.1146/annurev.cb.10.110194.000511

Walter P, Ron D. 2011. The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334:** 1081–1086. doi:10.1126/science.1209038

Weng Y, Czaplinski K, Peltz SW. 1996. Genetic and biochemical characterization of mutations in the ATPase and helicase regions of the Upf1 protein. *Mol Cell Biol* **16:** 5477–5490. doi:10.1128/mcb.16.10.5477

Wimmer K, Zhu XX, Lamb BJ, Kuick R, Ambros PF, Kovar H, Thoraval D, Motyka S, Alberts JR, Hanash SM. 1999. Co-amplification of a novel gene, NAG, with the N-myc gene in neuroblastoma. *Oncogene* **18:** 233–238. doi:10.1038/sj.onc.1202287

Wu B, Eliscovich C, Yoon YJ, Singer RH. 2016. Translation dynamics of single mRNAs in live cells and neurons. *Science* **352:** 1430–1435. doi:10.1126/science.aaf1084

Xu W, Bao P, Jiang X, Wang H, Qin M, Wang R, Wang T, Yang Y, Lorenzini I, Liao L, et al. 2019. Reactivation of nonsense-mediated mRNA decay protects against C9orf72 dipeptide-repeat neurotoxicity. *Brain* **142:** 1349–1364. doi:10.1093/brain/awz070

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Mühlemann O. 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17:** 2108–2118. doi:10.1261/rna.030247.111

Supplemental Material

Identification of a localized Nonsense-Mediated Decay pathway at the Endoplasmic Reticulum

Dasa Longman,[1] Kathryn A. Jackson-Jones,[1] Magdalena M. Maslon,[1] Laura C. Murphy,[1] Robert S. Young,[1,2] Jack J. Stoddart,[1] Nele Hug[1], Martin S. Taylor,[1] Dimitrios K. Papadopoulos,[1] and Javier F. Cáceres[1,3*]

[1]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XU, UK

[2]Present address: Centre for Global Health Research, Usher Institute, University of Edinburgh, Old Medical School, Teviot Place, Edinburgh EH8 9AG, UK

[*]**Corresponding author: Javier.Caceres@igmm.ed.ac.uk**

Running title: NMD regulation at the ER

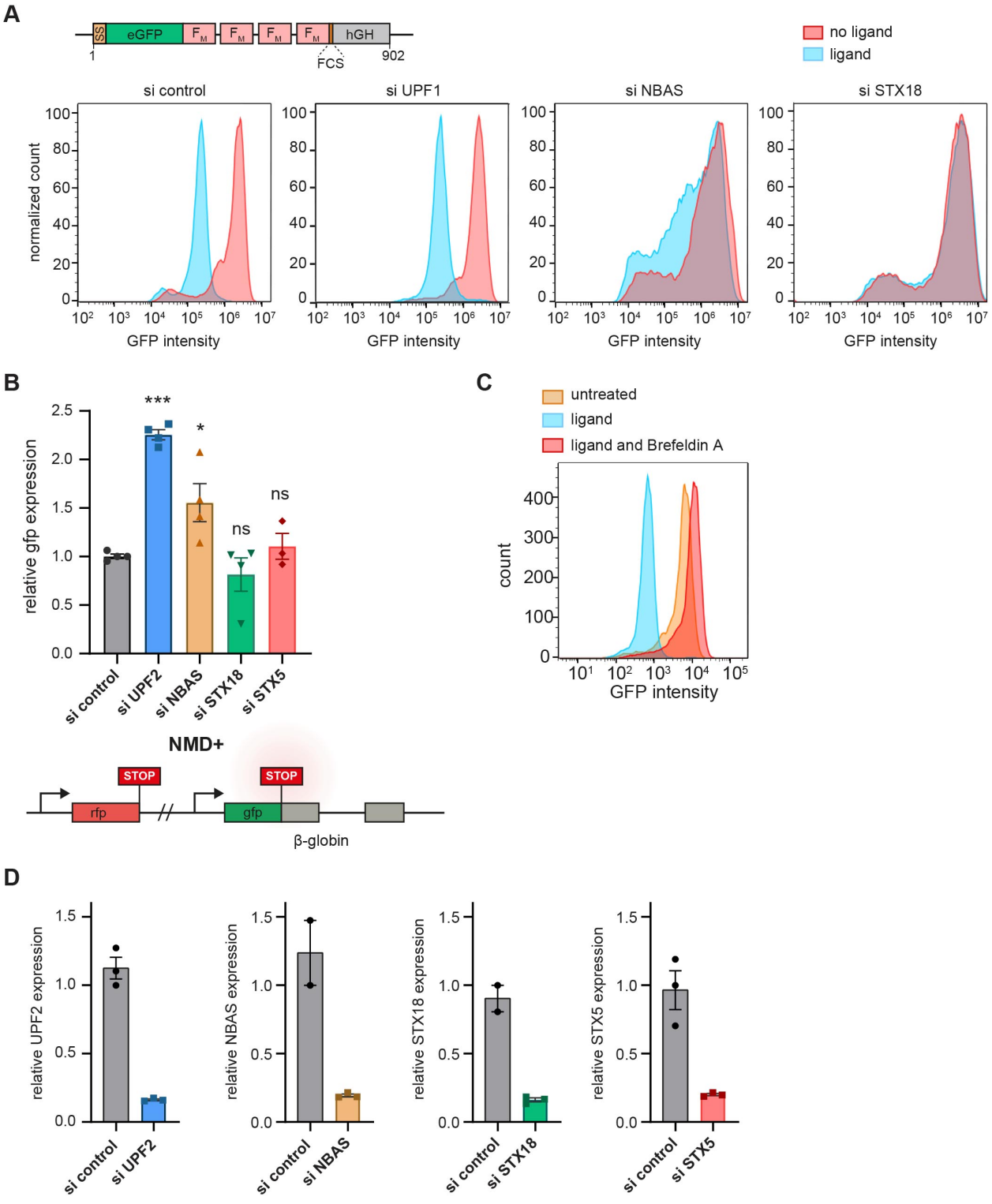
[*Keywords:* Nonsense-mediated decay (NMD); RNA quality control; UPF1; NBAS; ER stress, UPR]

Supplemental Material includes:

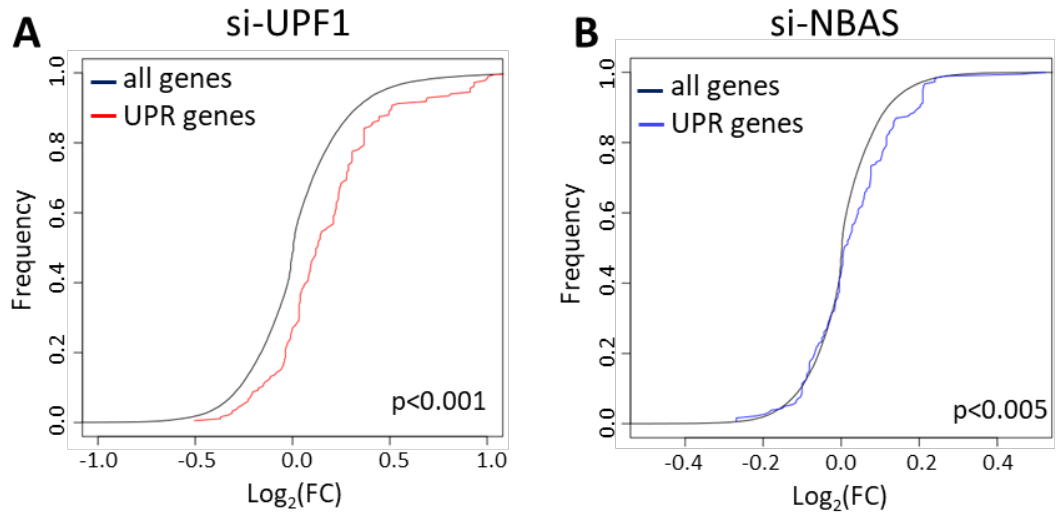
Supplemental Figures: S1-S8

Supplemental Tables: S1-S5 (All uploaded separately as Excel files)

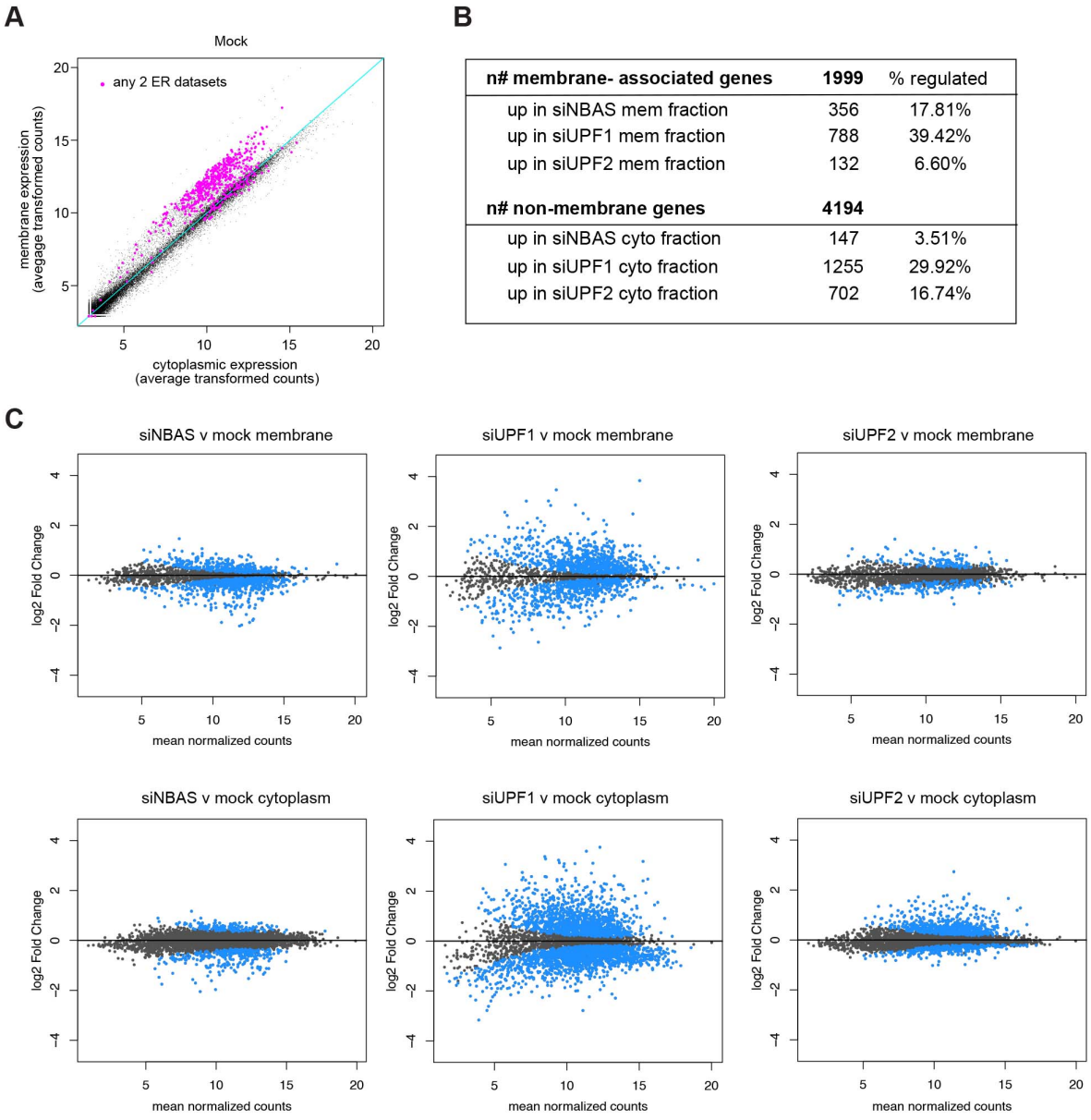
Additional Materials and methods


A
SS
eGFP
F_M
F_M
F_M
F_M
hGH
1
FCS
902
no ligand
ligand
si control
si UPF1
si NBAS
si STX18
normalized count
GFP intensity
B
relative gfp expression
***
*
ns
ns
si control
si UPF2
si NBAS
si STX18
si STX5
NMD+
STOP
STOP
rfp
gfp
β-globin
C
untreated
ligand
ligand and Brefeldin A
count
GFP intensity
D
relative UPF2 expression
relative NBAS expression
relative STX18 expression
relative STX5 expression
si control
si UPF2
si NBAS
si STX18
si STX5

Supplemental Figure S1. Independent functions of NBAS in NMD and Golgi-to-ER transport, Related to Fig. 1. (*A*) Top panel shows a schematic representation of the GFP-based reporter expressed in HeLa C1 cells used to measure constitutive secretion. SS: signal sequence; eGFP: enhanced green fluorescent protein; FM: FKBP mutated; FCS: furin cleavage sequence; hGH: human growth hormone, and numbers represent amino acid residues. Defects in secretion are manifested by the inability of cells to lose GFP fluorescence even after addition of the ligand. Lower panels show examples of flow-cytometry before (in red), and after the addition of ligand (in blue), for control-depleted cells, cells depleted of NMD factors UPF1 and NBAS, and of secretion factor, STX18. (*B*) NMD activity is not affected by the depletion of secretion factors STX18 and STX5. HeLa cells were transfected with a fluorescent NMD+ reporter and depleted of NMD factors UPF2 or NBAS, or secretion factors STX18 or STX5. Depletion of both NMD factors increased the mean green fluorescence in comparison to mock-depleted cells, as expected, whereas STX18 and STX5 depletion did not affect the mean green fluorescence. Each point represents one biological replica, bars indicate mean with SEM. Significance was determined by two-tailed unpaired t-test: ***: $p<0.0001$; *: $p<0.05$; ns: not significant. (*C*) Brefeldin A treatment blocks constitutive secretion. HeLa C1 cells carrying a GFP-based secretion reporter were FACS-sorted to monitor their ability to secrete GFP. HeLa C1 cells (in orange) were able to secrete a GFP-based reporter following addition of ligand (in blue). Brefeldin A treatment led to a block in secretion and a concomitant accumulation of GFP even in the presence of ligand (in red). (*D*) Representative efficiency of depletion of NMD and secretion factors was determined by qRT-PCR relative to mock-depleted cells. In each case, relative expression of the depleted factor was normalized to POLR2J. Each point represents one biological replica, bars indicate mean with SEM.



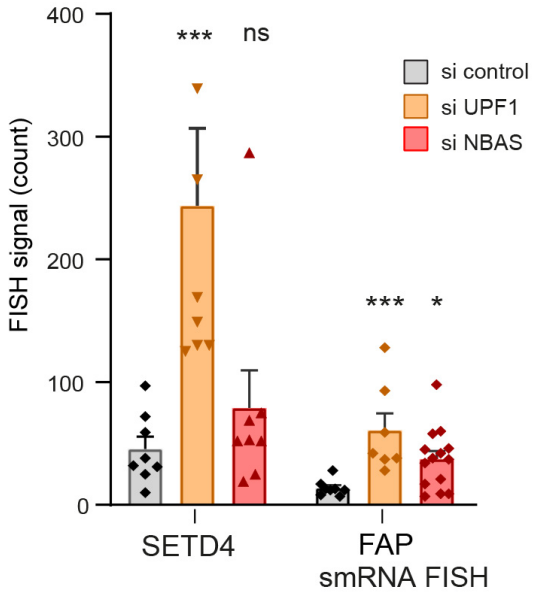
Supplemental Figure S2. Both NBAS and UPF1 regulate the ER stress response, Related to Fig. 2. Genes associated with the ER unfolded protein response (UPR) (GO: 0006986) show a significant increase in fold change compared to all genes, when either UPF1 (A) or NBAS (B) were depleted ($p < 0.001$ and $p < 0.005$ respectively; Wilcoxon rank sum test).



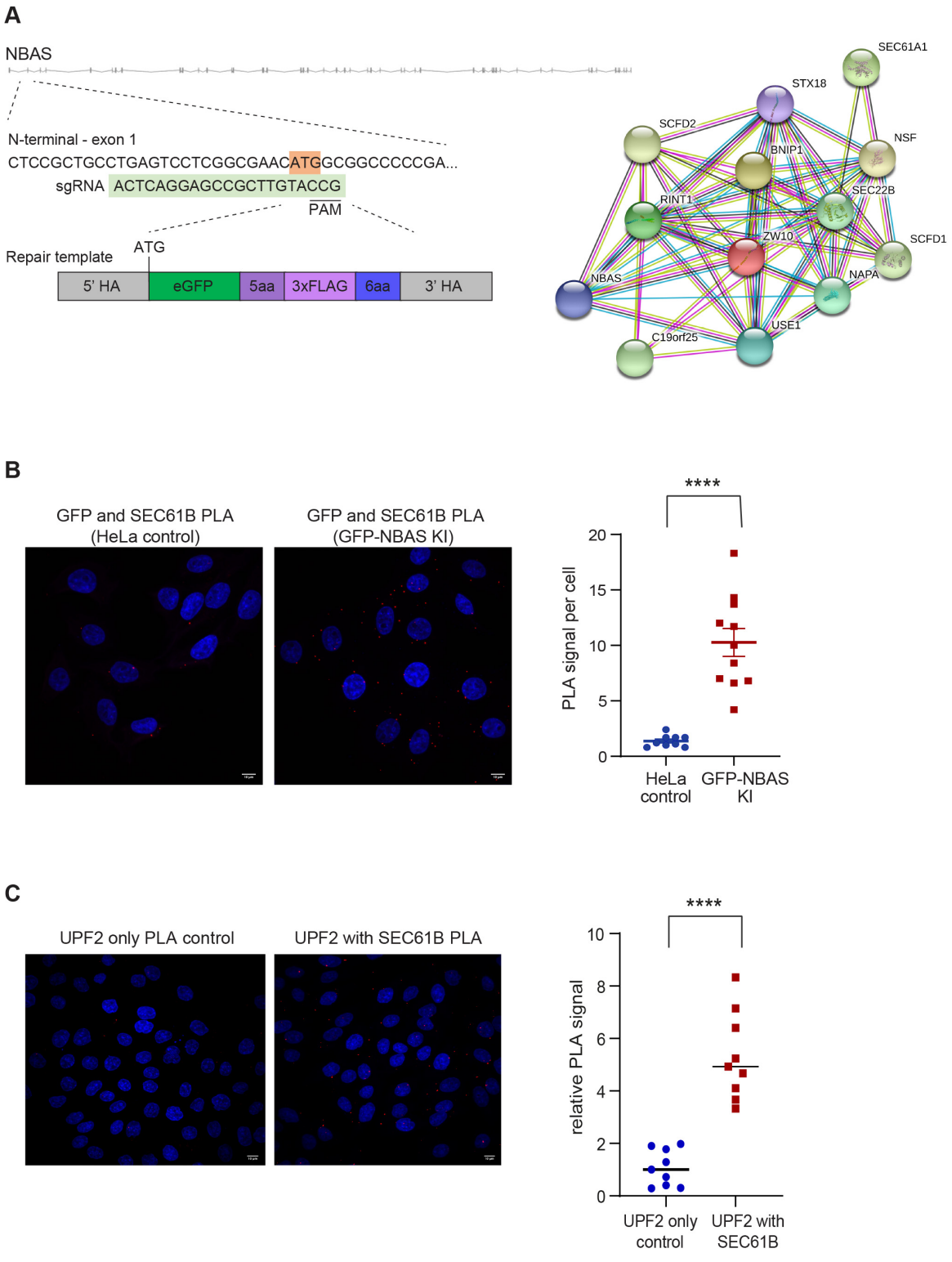
n# membrane- associated genes	1999	% regulated
up in siNBAS mem fraction	356	17.81%
up in siUPF1 mem fraction	788	39.42%
up in siUPF2 mem fraction	132	6.60%
n# non-membrane genes	**4194**	
up in siNBAS cyto fraction	147	3.51%
up in siUPF1 cyto fraction	1255	29.92%
up in siUPF2 cyto fraction	702	16.74%

Supplemental Figure S3. NBAS regulates membrane-associated RNA targets, Related to Fig. 2. (*A*) Experimentally validated ER genes are present in the membrane fraction. Scatter plot of gene expression in membrane vs cytoplasmic fractions in mock-depleted cells, with experimentally validated ER genes present in at least 2 of the 3 ER datasets described in Fig. 2B are indicated in magenta. (Enrichment OR = 66.94, p-value <0.0001, Fishers Exact test). (*B*) Summary table of genes

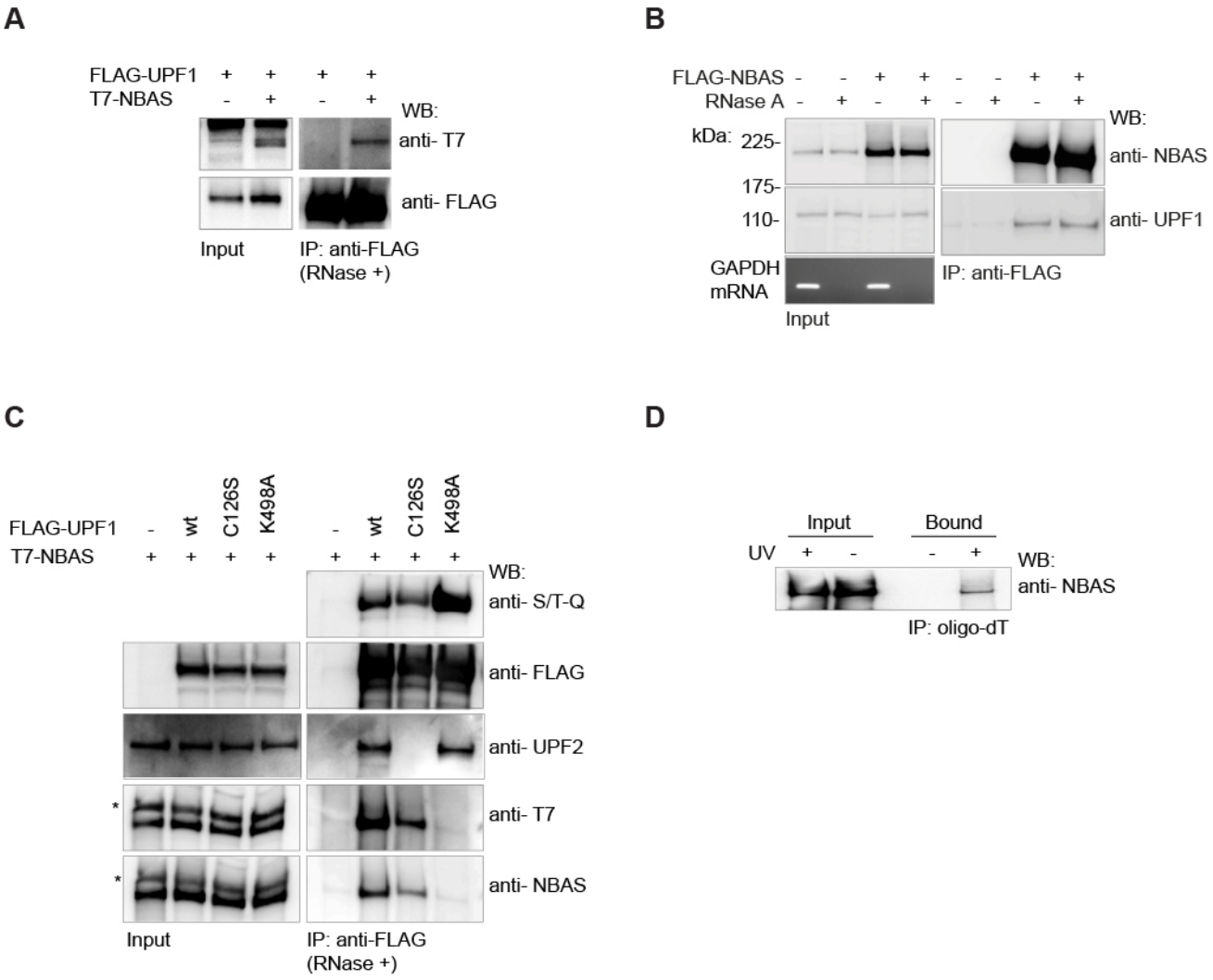
regulated by NMD factors in each cellular fraction. Genes were defined as membrane-enriched or non-membrane-enriched as described in Figure 2C. Genes were classified as regulated by the respective NMD factor if they were significantly ($p<0.05$) upregulated in that fraction, when the NMD factor was depleted. (*C*) Differential expression analysis of genes regulated by NMD factors in membrane fractions and in the cytoplasm. MA plots show log2 fold-change of gene expression plotted over mean of normalized counts (gene expression). Significantly changed genes ($p<0.05$) are indicated in blue.



Supplemental Figure S4. Quantitation of *SETD4* and *FAP* transcripts upon UPF1 and NBAS depletion in smRNA FISH experiments, Related to Fig. 3. Relative changes in the expression of a cytoplasmic NMD target (SETD4) or an ER-localized NMD target (FAP) were determined by counting discreet FISH signal in individual cells. The FISH signal of FAP upon UPF1 or NBAS depletion is likely underestimated, as clusters of FISH dots were counted as one. Each point represents counts in one cell, bars indicate mean with SEM. Significance was determined by Mann-Whitney test: ***: $p<0.001$; *: $p<0.05$; ns: not significant.

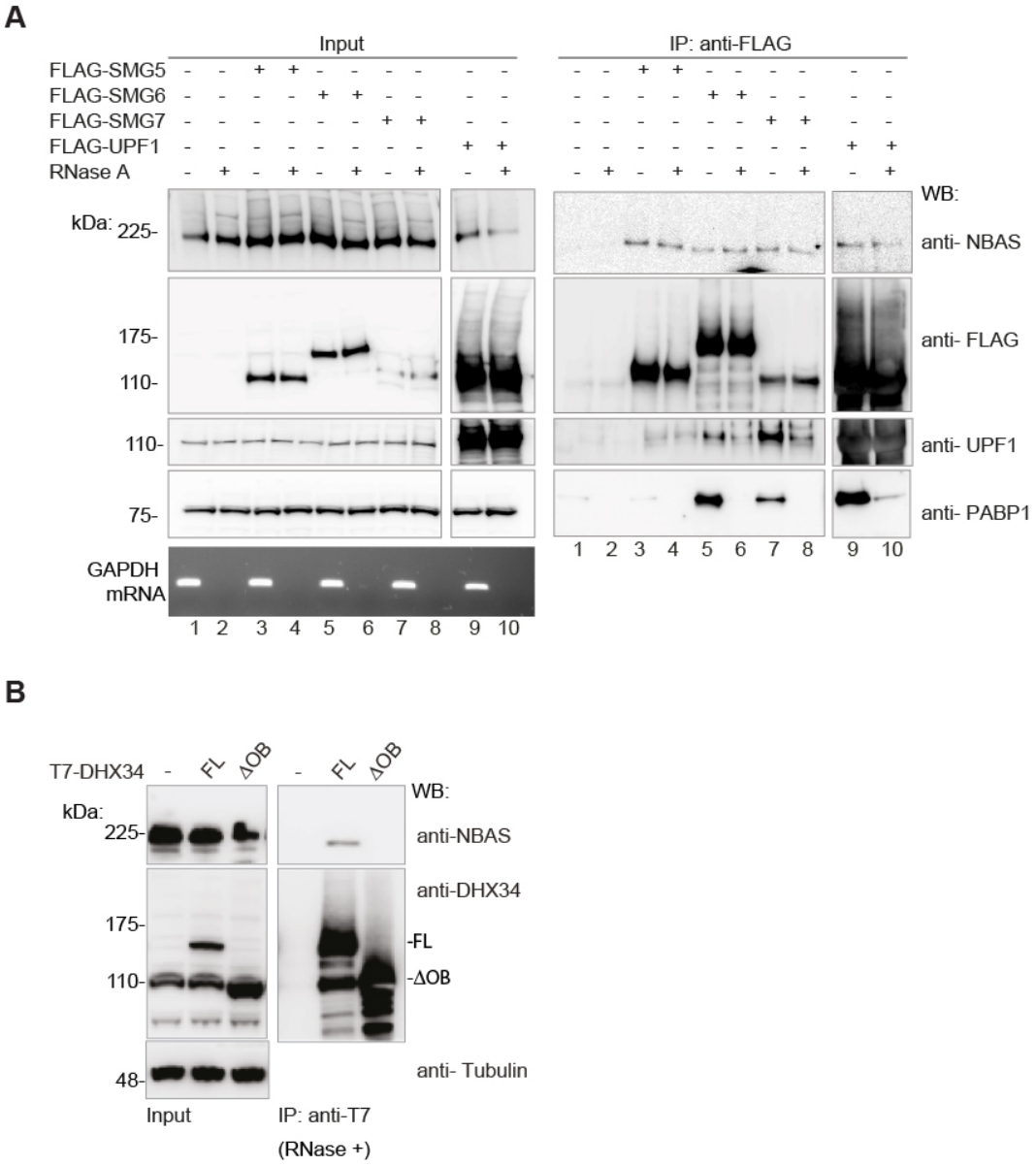

A
NBAS
N-terminal - exon 1
CTCCGCTGCCTGAGTCCTCGGCGAACATGGCGGCCCCCGA...
sgRNA ACTCAGGAGCCGCTTGTACCG
PAM
Repair template
ATG
5' HA
eGFP
5aa
3xFLAG
6aa
3' HA
SEC61A1
STX18
SCFD2
NSF
BNIP1
SEC22B
RINT1
ZW10
SCFD1
NBAS
NAPA
USE1
C19orf25
B
GFP and SEC61B PLA (HeLa control)
GFP and SEC61B PLA (GFP-NBAS KI)
****
PLA signal per cell
20
15
10
5
0
HeLa control
GFP-NBAS KI
C
UPF2 only PLA control
UPF2 with SEC61B PLA
****
relative PLA signal
10
8
6
4
2
0
UPF2 only control
UPF2 with SEC61B

Supplemental Figure S5. Localization of endogenous NBAS and UPF2 proteins to the ER, Related to Fig. 4. (*A*) NBAS is part of the Syntaxin 18 complex. Left panel depicts the strategy used to knock-in an eGFP/3xFLAG tag at the N-terminus of the NBAS locus, using CRISPR/Cas9-mediated genome editing in HeLa cells. Right panel shows the top NBAS interacting partners, and SEC61A1, identified by Mass spectrometry of GFP-pulled down cell lysates. The protein-protein interaction cluster was generated by STRING (https://string-db.org). The full list of proteins identified by Mass Spectrometry is presented in Supplemental Table T4. *(B)* Direct interaction between endogenous epitope-tagged NBAS and SEC61B proteins was determined by PLA using anti-GFP and anti-SEC61B antibodies, in HeLa KI (epitope tag knock-in) cells or wild-type HeLa cells, as a control. The graph shows the quantification of the PLA signal. Each point represents mean PLA count per cell in one captured frame, relative to the HeLa negative control. Significance was determined by Mann-Whitney test: ****: $p<0.0001$. (*C*) UPF2 is localized in the close proximity of the SEC61 translocon component at the ER. Direct interaction between endogenous UPF2 and SEC61B proteins was determined by PLA using the anti-UPF2 and anti-SEC61B antibodies, in HeLa cells. The graph shows the quantification of PLA signal. Each point represents mean PLA signal per cell in one captured frame, relative to the HeLa negative control. Significance was determined by Mann-Whitney test: ****: $p<0.0001$.



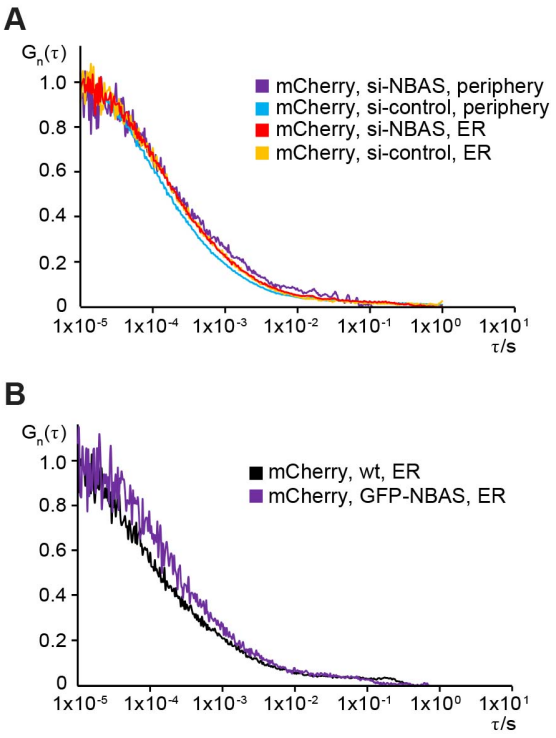
Supplemental Figure S6. NBAS interacts with UPF1 and preferentially associates with the SURF complex, Related to Fig. 5. (*A*) T7-tagged NBAS was co-expressed with FLAG-tagged UPF1 in HeLa cells. Cell lysates were subjected to Immunoprecipitation (IP) with anti-FLAG antibody in the presence of RNase. The NBAS-UPF1 interaction was revealed by Western Blot with anti-T7 antibody. (*B*) Interaction of endogenous UPF1 with FLAG-tagged NBAS in the absence or presence of RNase A. FLAG-tagged-NBAS was transiently expressed in HEK293T cells and immunopurified

with anti-FLAG antibody. Co-Immunopurification of endogenous UPF1 was analyzed by Western Blot with anti-UPF1 antibody. RNA digestion was confirmed by RT-PCR for GAPDH. (*C*) T7-NBAS was co-expressed in HeLa cells with FLAG-tagged wild-type UPF1; C126S hypo-phosphorylated mutant UPF1 predominantly present in the SURF complex, or with K498A ATP-binding mutant UPF1, which is mainly associated with the DECID complex. Anti-FLAG IPs were performed in the presence of RNase and subjected to Western Blot analysis with the indicated antibodies. To detect phosphorylated UPF1, anti-FLAG-IPs were probed with a phospho-(Ser/Thr) ATM/ATR substrate antibody (anti-S/T-Q). (*D*) NBAS is bound to mRNAs. Hela cells were UV crosslinked and mRNP complexes were purified using oligo dT beads. The presence of NBAS in mRNP complexes was revealed by Western Blot with anti-NBAS antibody.



Supplemental Figure S7. Interaction of endogenous NBAS with core NMD factors, SMG5, 6, 7, UPF1 and DHX34, Related to Fig. 5. (*A*) Endogenous NBAS interacts with several NMD factors independently of RNA. HEK293T cells were transiently transfected with FLAG-tagged SMG5,

SMG6, SMG7 and UPF1. Anti-FLAG Immunoprecipitations (IPs) were performed in the absence or presence of RNase A. Inputs and anti-FLAG IPs were subjected to Western Blot (WB) analysis with the indicated antibodies. The interaction with poly(A)-binding protein (PABP1) was used as a control for the RNase treatment. Successful RNA digestion of the cell lysate was further confirmed by RT-PCR for the GAPDH housekeeping gene and analyzed on agarose gels. (*B*) T7-tagged full-length DHX34 interacts with endogenous NBAS. HEK293T cells were transfected with full-length (FL) and a truncated version of DHX34, lacking its OB-like domain (ΔOB). T7-DHX34 (FL and ΔOB) were immunopurified with anti-T7 antibody in the presence of RNase A. Inputs and T7-IPs were analyzed with anti-NBAS and anti-DHX34 antibodies. Anti-Tubulin antibody served as a loading control.



Supplemental Figure S8. FCS control experiments, Related to Fig. 6. (*A*) Average autocorrelation curves obtained for FCS measurements indicate that mCherry alone shows an almost complete overlap of decay times upon depletion of NBAS by RNAi, or between the ER and the cytoplasmic periphery. (*B*) Overexpression of GFP-NBAS does not affect the mobility of mCherry control at the ER.

Supplemental Table S1. Gene expression analysis.

Supplemental Table S2. GO term analysis of data presented in Supplemental Table S1.

Supplemental Table S3. Gene expression analysis of membrane and cytoplasmic fractions.

Supplemental Table S4. IP-Mass spectrometry protein interactors of an endogenously-tagged NBAS cell line.

Supplemental Table S5. List of oligonucleotides, smRNA FISH probes and siRNAs.

Additional materials and methods

Cell Culture and transfections.

HeLa cells were maintained in DMEM media with high glucose, GlutaMAX™ Supplement, pyruvate (Gibco Life technologies; 10569010) supplemented with 10% FCS, at 37°C in the presence of 5% CO_2. HeLa C1 cells and 3xFLAG-eGFP-NBAS cells were maintained in the same media with the addition of 1μg/ml of Puromycin. Cells were grown without antibiotic prior to transfections, which were carried out in Opti-MEM reduced serum medium (Gibco, 31985047). Transfections of siRNA oligos were done using DharmaFECT 1 (Dharmacon, T-2001-03) following manufacturer's protocol. See Supplemental Table S5 for siRNA reagents information. Co-transfection of plasmids and siRNA oligos was achieved using Lipofectamine 2000 (ThermoFisher Scientific, 11668019) following manufacturer's instructions. For depletions, cells were plated in 12-well plates and transfected with 30 pmol of specific siRNAs. Cells were then expanded into 6-well plates and transfected with 50 pmol of the same siRNAs on the third day and were harvested for analysis on day 5 after the first depletion. For fluorescence based NMD assays HeLa cells were treated as above, with the addition of 50 ng NMD+ reporter that was co-transfected with siRNAs during the second round of depletion. For total RNA-sequencing and subcellular fractionation, cells were plated in 6-well plates and transfected with 30 pmol of indicated siRNAs. Cells were expanded into 10 cm plates the following day and were transfected with 30 pmol of the same siRNAs on day 3 and were harvested for analysis on day 4 after the first depletion.

Antibodies

Anti-UPF1, Bethyl # A300-036A, RRID:AB_203272; Anti UPF1, Bethyl # A300-038A, RRID:AB_2288326; Anti-GFP, Roche # 11814460001, RRID:AB_390913; Anti-Sec61β,

Proteintech # 15087-1-AP, RRID:AB_2186411; Anti-NBAS, Abcam # Ab122370, RRID:AB_11129367; Anti-UPF2, Santa Cruz Biotechnology # sc-374230, RRID:AB_10988267; Anti-Calnexin, Enzo Life Sciences # ADI-SPA-860-D, RRID:AB_2038898; Anti-T7, Novagen # 69522; Anti-FLAG, Sigma-Aldrich # F3165, RRID:AB_259529; Anti-Phospho-(Ser/Thr) ATM/ATR Substrate (anti-S/T-Q), Cell Signalling # 2851, RRID:AB_330318; Anti-Tubulin, Sigma-Aldrich # 4026, RRID:AB_477577, Anti-PABP1 PABP1 (4992, Cell Signaling), Anti- DHX34 is a peptide-specific antibody raised against human DHX34 obtained from Eurogentec (Hug and Cáceres 2014).

Flow cytometry-based assay for measuring constitutive secretion

The secretion assay was performed as previously described (Gordon et al. 2010). HeLa C1 cells constitutively express a secreted eGFP reporter fused to mutant FKBP proteins (FK506-binding protein), which form large aggregates that cannot be secreted. The F36M substitution leads to formation of dimers that can be dissociated by ligand binding, leading to efficient secretion (Gordon et al., 2010). HeLa C1 cells mock-depleted or depleted of UPF1, UPF2, DHX34, SEC13, NBAS, p31, RINT, ZW10 and STX18 were incubated with 1µM D/D Solubilizer (ligand) (Clontech 6350540 for 1.5 hours to facilitate secretion or left untreated for comparison. Cells were trypsinized, washed with cold PBS and placed on ice to stop further secretion. To measure the amount of GFP reporter remaining, cells were analyzed by FACS BD AccuriTM (488nm excitation laser, FL-1: 533/30nm emission filter) running the software BD AccuriTM C6. Gates were set using a non-fluorescent control. GFP expression was analyzed by gating the single cell population in SSC-H/SSC-A dot plot, followed by debris-exclusion gate in SSC-A/FSC-A dot plot. Between 5000 and 10000 single cells were analyzed for each sample. The mean GFP fluorescence

was calculated with FlowJo™ Software (Version 10.6.0). The relative GFP fluorescence remaining after secretion was calculated as a ratio between the mean fluorescence of depleted cells incubated with the ligand and depleted cells. The relative fluorescence remaining of untreated C1 cells was set to 1, and the threshold for a secretion defect was set to two.

Cell fractionation

HeLa cells were mock-depleted or depleted twice of UPF1, UPF2 and NBAS as described above. Cellular fractionation was performed as described before (Jagannathan et al. 2011) with some modifications. Briefly, four days after the first depletion, cells were detached using trypsin and washed twice in 1ml of ice-cold PBS (500g, 10 min, 4°C). Cellular pellets were resuspended in 0.4 ml of permeabilization buffer (110 mM KOAc, 25 mM K-HEPES pH 7. 2, 2.5 mM $Mg(OAc)_2$, 1mM EGTA, 0.015% digitonin, 1mM DTT, 1× Complete Protease Inhibitor Cocktail, 40 U/mL RNaseOUT™) and incubated for 5 min on a rotating wheel at 4°C. The resulting cytosolic fractions were recovered by centrifugation at 2000g for 10 minutes, 4°C. Cells were then lysed in 0.4 ml of NP-40 lysis buffer (400 mM KOAc, 25 mM K-HEPES pH 7.2, 15 mM $Mg(OAc)_2$, 1% (v/v) NP-40, 1 mM DTT, 1× Complete Protease Inhibitor Cocktail, 40 U/mL RNaseOut) for 30 min on ice. The membrane fraction was recovered by centrifugation at 7000g for 10 min at 4°C. Both cytosolic and membrane fractions were clarified by centrifugation at 7500g for 10 min at 4°C. Digitonin, DTT, Complete Protease Inhibitor Cocktail and RNaseOUT™ were added fresh to the buffers. RNA from both fractions was isolated using PureLink RNA Mini Kit (Life Technologies) according to manufacturer's instructions. DNA was removed using TURBO DNA-free™ DNase I kit (Invitrogen Ambion; AM1907).

Gene expression profiling: RNA extraction, library preparation and RNA-sequencing

Total RNA was isolated from depleted cells by phenol-chloroform extraction and treated with TURBO DNA-free™ DNase I kit (Invitrogen Ambion; AM1907). Libraries were prepared following NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs; #E7420) and Agencourt AMPure XP Beads (Beckman Coulter; A63881). Samples were ligated to barcode primers 1-20 of NE Next Index Primers for Illumina sets 1 and 2 (New England Biolabs) and libraries analyzed using DNA High Sensitivity chip on an Agilent 2100 Bioanalyzer before being pooled. 150 base pair, paired-end sequencing was performed using the S2 flow cell on the NovaSeq 6000 System (Illumina Inc.). Library molarity for sequencing was calculated using Qubit dsDNA quantification results and fragment size information from Bioanalyzer results. Sequencing was performed by Edinburgh Genomics, Edinburgh, UK.

RNA-sequencing analysis

FASTQ files were quality control checked for base and sequence quality scores, and adapter contamination using fastQC (v0.11.7; Babraham Bioinformatics). Reads were aligned using Spliced Transcripts Alignment to a Reference (STAR, v2.5.1b) (Dobin et al. 2013) or pseudoaligned using kallisto (v0.43.1; (Bray et al. 2016)). Kallisto index was created by combining "all basic gene annotation" and "long non-coding RNA gene annotation" fasta files, downloaded from Gencode (www.gencodegenes.org/releases/current.html), Release 27 (GRCh38.p10). Kallisto was run with 100 bootstraps. Reads per feature in STAR output files was counted using HTSeq (v0.9.1; ((Anders et al. 2015)). Abundance files were then analyzed for fold changes by running DESeq2 (v1.14.1) differential expression analysis (Love et al. 2014). Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) (http://cbl-gorilla.cs.technion.ac.il/) (Eden et al. 2009)) and Database for

Annotation, Visualization and Integrated Discovery (DAVID) (https://david.ncifcrf.gov/home.jsp) (Huang et al. 2009) were used for Gene Ontology (GO) analysis. To investigate genes linked to specific GO terms, genes were annotated using biomaRt Bioconductor R package (Durinck et al. 2005). For NBAS and UPF1 targets overlap with experimentally ER-localized genes, we first downloaded relevant supplementary data. For APEX-seq data (Fazal et al. 2019), ER genes were defined as all those labeled "ERM_Gene" and for ER Fractionation sequencing data (Reid and Nicchitta 2012) those genes that were labelled "ER". For ER Proximity-specific Ribosome Profiling data (Jan et al. 2014), gene names were converted from uniprot.protein.id to official gene symbols using DAVID (https://david.ncifcrf.gov/home.jsp) (Huang et al., 2009), and a 1.5-fold enrichment at ER cut off applied. For fractionated RNA-sequencing, gene expression in both fractions in control cells was compared by using variant stabilized transformation of tpm, averaged across replicates. Membrane-associated genes were defined as those >2-fold higher expression in membrane fraction than cytoplasmic fraction, all other genes were termed non-membrane. Membrane-enrichment was confirmed by overlap with three experimental ER datasets (see above). Changes in gene expression resulting from depletion of individual NMD factors was measured by percentage of 'membrane-associated' genes in membrane fraction, and of 'non-membrane' genes in cytoplasmic fraction regulated. Genes were determined as regulated if they were significantly ($p<0.05$) increased in expression when the relevant factor was depleted.

Quantitative RT-PCR

Total RNA was isolated using RNeasy Mini kit and resuspended in nuclease-free water at 100ng/µl. qRT–PCR was performed using SuperScriptIII One-Step RT–PCR Kit (Invitrogen) following the manufacturer's instructions. All RT-PCRs were run on the CFX96 Real-Time System (Bio-Rad

machine, following this program: RT at 50°C for 30 min, 95°C for 2 min, then 40 cycles of 95°C for 30 sec, 55°C for 20 sec, 70°C for 20 sec followed by the plate read step. Each sample was run in 3 technical replicates. Primers were designed using Roche Real-Time Ready Configurator and combined with Roche Universal Probe Library (See Supplemental Table S4). Gene expression data was analysed by the delta Ct method, with each gene normalised to housekeeping gene *POL2RJ*. Unpaired two-tailed t-test was used for statistical analysis.

Immunofluorescence

Cells were grown on coverslips, fixed with 4% paraformaldehyde at room temperature for 10 min, washed with PBS and permeabilized with 0.5% Triton X-100 at room temperature for 10 min. Coverslips were then incubated for 1 hour with block buffer (1% BSA, 0.01% Triton X-100 in PBS). Primary antibodies (diluted 1:500 in block buffer) were incubated with coverslips in humidified chamber overnight at 4°C. Coverslips were washed 3 times with wash buffer (0.01% Triton X-100 in PBS). Secondary antibodies (Alexa Flour® 488 or Alexa Flour® 594, Molecular Probes, diluted 1:1000 in block buffer) were incubated with coverslips in a dark, humidified chamber for 1 hour at room temperature. Coverslips were then washed 3 times with wash buffer and stained with 4,6-diaminidino-2-phenylidole (DAPI) at 50ng/ml, mounted in Vectashield (Vector) and sealed with nail varnish. For Digitonin treatment, media was aspirated, and cells incubated for 10 min on ice in cold PBS and later incubated in Digitonin buffer (110mM KoAc, 25mM K-HEPES, 2.5mM $MgCl_2$, 1mM EGTA) with 0.01% Digitonin for 5 min on ice, washed in room temperature PBS and fixed with 4% paraformaldehyde, as above.

Design and screening of CRISPR cell lines

The design of the guide RNAs (gRNAs) was undertaken, as previously described (Ran et al. 2013). gRNAs were designed using sgRNA Designer CRISPRko (Broad institute, https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design) and Cas-Designer (RGEN Tools, http://www.rgenome.net/cas-designer/). For the tagged-NBAS cell line, four gRNAs were selected by closest proximity to the start codon and highest predicted efficiency. For each guide RNA, the complementary sequence was determined and a BbsI restriction site added to both oligos. Designed gRNAs were ordered as custom single stranded DNA oligos with an extra G at the 5' end, and 5' phosphate at the reverse complement DNA strand (IDT). Top and bottom strands of gRNAs were annealed at a concentration of 100μM and cloned into the px459 V2.0 vector using BbsI restriction cloning. 1 μl of a 1:250 dilution of annealed gRNAs was ligated with the T4 DNA Ligase (NEB) into 36 ng of the px459 V2.0 vector. To assess cutting efficiency of gRNAs, each plasmid was transfected into HeLa cells and RNA extracted as above. PCR over the target region was performed with custom primers (Supplemental Table S4) and the resulting products assessed by DNA electrophoresis. PCR products were also sent for Sanger sequencing and traces were analyzed by TIDE (REF) and ICE (REF) to give exact cutting efficiencies for each guide. Repair template was ordered as a custom plasmid from IDT on a pUCIDT-AMP backbone. For the generation of NBAS knock out (KO) HeLa cells, two guides (NBAS KO_A and NBAS KO_B) targeting the 5th exon of the NBAS gene were cloned into the pSpCas9n(BB)-2A-GFP vector encoding the (D10A nickase mutant (PX461), as described above. pSpCas9n(BB)-2A-GFP (PX461) (Addgene plasmid # 48140; http://n2t.net/addgene:48140; RRID: Addgene_48140). The gRNA/Cas9 plasmid and repair template plasmid were transfected into HeLa cells as described above. After 24hr media was supplemented with 1.5 ug/ml puromycin to select for cells expressing the gRNA/Cas9 plasmid. Once control cells

were dead, surviving cells' fluorescence was measured by FACS and single cells deposited into wells of a 96-well plate. After approx. 2 weeks clonal expansion, genomic DNA was extracted by lysing with DirectPCR Lysis Reagent Cell (Peqlab, VWR) supplemented with 0.5 µg/µl Proteinase K (Invitrogen), at 55°C O/N. PCR of the N-terminal locus of NBAS was performed using previously mentioned custom primers and SYBR green Master Mix. PCR products were resolved on an agarose gel for genotyping. The correct repair template integration was validated by genomic DNA sequencing and expression of tagged-NBAS was validated by Western blotting. Screening for the NBAS KO clones was performed as above. HeLa cells co-transfected with NBAS KO_A and _B pSpCas9n(BB)-2A-GFP plasmids were FACS-sorted into 96-well plates 48 h after transfection. Growing colonies were screened by PCR of genomic DNA as described, using NBAS KO F and R primers. NBAS expression was assessed by Western Blotting and genomic DNA of the NBAS KO was checked by sequencing.

Image Capture and analysis

For smRNA FISH experiment, epifluorescent images were acquired using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A1 fluorescence microscope with a Plan Apochromat 100x 1.4NA objective, a Nikon Intensilight Mercury based light source (Nikon UK Ltd, Kingston-on-Thames, UK) and either Chroma #89014ET (3 colour) or #89000ET (4 colour) single excitation and emission filters (Chroma Technology Corp., Rockingham, VT) with the excitation and emission filters installed in Prior motorized filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Step size for z stacks was set at 0.2 µm. Hardware control, image capture and analysis were performed using Nikon Nis-Elements software (Nikon UK Ltd, Kingston-on-Thames, UK).

Images were deconvolved using a calculated point spread function with the constrained iterative algorithm of Volocity (PerkinElmer Inc, Waltham MA). Image analysis was carried out using the FIJI/ImageJ software (2.0.0-rc-69/1.52p) (Schindelin et al. 2012). To measure the distribution of FISH signal within cells, a maximum intensity projection of each deconvolved z-stack was created and the DAPI signal was thresholded to create a mask of the nucleus. This mask was inverted and a euclidean distance map was created using the "Distance Map" function. This resulted in an image where all pixels that were inside the nucleus had a value of zero and the other pixels had values based on their distance from the closest nucleus edge (in pixels). To detect the centroids of FISH spots in the FISH signal channel, the ImageJ function "Find Maxima" was used with a prominence setting of 800. The x-y co-ordinates of each maximum point was measured on the distance map, giving a distance of the FISH spots to the edge of the nucleus. These values were converted to microns using the pixel spacing value of the original image. For immunofluorescence and PLA experiments images were acquired on a Nikon Confocal A1R confocal microscope using a Plan Apochromat 100x 1.4NA objective. The microscope comprises a Nikon Eclipse TiE inverted microscope with Perfect Focus System and is equipped with 405nm diode, 457/488/514nm Multiline Argon, 561nm DPSS and 638nm diode lasers. Detection is via four Photomultiplier tubes (2x standard Photomultiplier tubes and 2x GaAsP PMTs). Data were acquired using NIS Elements AR software (Nikon Instruments Europe, Netherlands). Z-stacks of images were acquired with a 0.2 µm step, scan size 1024x1024, 1.2x zoom and 2x frame averaging. Image analysis was carried out using the FIJI/ImageJ software.

Mass Spectrometry

Cells were harvested and lysed as in immunoprecipitation protocol (see below). α-GFP antibody-coupled magnetic beads (Sigma) were equilibrated with IP buffer. Lysates were resuspended in 500µl

IP buffer for capture of GFP-NBAS bound proteins and subsequent mass spectrometry analysis. Immunoprecipitation was performed on Kingfisher Duo robot (Thermo) for 4 h. All steps were carried out at 4°C. Beads were then transferred for two washes in IP buffer and three washes in TBS (300 µl each). After transfer into 100 µl 2M urea, 100 mM Tris, 1 mM DTT containing 0.3 µg trypsin per sample, beads were incubated at 27°C for 30 min with mixing to achieve limited proteolysis. The beads were then removed, and tryptic digest of the released peptides was allowed to continue for 9 h at 37°C. Following this, peptides were alkylated by adding iodoacetamide to 50 mM and incubated at room temperature for 30 min. Finally, peptides were acidified by addition of 8 µl 10% TFA. An estimated 10 µg of the resulting peptide solution was loaded onto an activated (20 µl methanol), equilibrated (50 µl 0.1% TFA) C18 StAGE tip, and washed with 50 µl 0.1% trifluoroacetic acid (TFA). The bound peptides were eluted into a 96-well plate (Axygen, Corning Inc., Corning, NY, USA) with 20 µl 80% acetonitrile (ACN), 0.1% TFA and concentrated to less than 4 µl in a vacuum concentrator. The final volume was adjusted to 15 µl with 0.1% TFA. Mass spectrometry was carried out by IGMM Mass Spectrometry core facility. Online LC was performed using a Dionex RSLC Nano (Thermo Fisher Scientific). Following the C18 clean-up, 5 µg peptides were injected onto a C18 packed emitter and eluted over a gradient of 2%-80% ACN in 48 minutes, with 0.5% acetic acid throughout. Eluting peptides were ionised at +2.2kV before data-dependent analysis on a Thermo Q-Exactive Plus. MS1 was acquired with mz range 300–1650 and resolution 70,000, and top 12 ions were selected for fragmentation with normalised collision energy of 26, and an exclusion window of 30 seconds. MS2 were collected with resolution 17,500. Raw MS data were analyzed using MaxQuant (v 1.5.6.5) (Max Planck Institute of Biochemistry) in conjunction with UniProt human reference proteome release 2016_11 (uniprot.com), with match between runs (MS/MS not required), LFQ with 1 peptide required, and statistical analyses performed in R (RStudio 1.1.453 / R x64 3.4.4)

(rstudio.com) using Wasim Aftab's LIMMA Pipeline Proteomics (github.com/wasimaftab/LIMMA-pipeline-proteomics) implementing a Bayes-moderated method. Interactome analysis including gene ontology was carried out by inputting protein list into STRING (string-db.org/).

Immunoprecipitation and Western Blotting

Cells were washed and harvested in ice-cold PBS before pellets were lysed with immunoprecipitation (IP) buffer [20 mM Tris-HCl pH 8, 150 mM NaCl, 1mM EDTA, 1% NP-40, 0.2% Deoxycholate, Complete Protease Inhibitor (Roche), Phosphor STOP (Roche), 1 mM DTT] for 20 min on ice. Cell lysates were treated with 40–80 mg/ml RNase A per 1 ml of extract. Lysates were precleared with Dynabeads Protein-G (Novex, Life Technologies) for 1 h, rotating at 4°C. Anti-FLAG M2 antibody coupled magnetic beads (Sigma-Aldrich M8823) were washed x3 with IP buffer before incubation with the precleared lysate overnight rotating at 4°C. Beads were washed x5, each for 5 min with IP buffer and then bound protein was eluted by boiling in SDS sample buffer supplemented with reducing agent for 5min. Proteins were resolved by SDS-PAGE using NuPAGE 3-8% Tris-Acetate precast gels (Novex, Life Technologies) run for ~1 h at 170V in 1x Tris-Acetate running buffer. Protein transfer was achieved using the iBlot™ 2 Gel Horizontal Transfer Device (Invitrogen). Nitrocellulose membranes were blocked in 5% BSA in PBS/Tween 20 (0.1%) for a minimum of 1h at room temperature and probed with the appropriate primary antibody diluted in blocking solution 1:3000. FLAG antibody was used at 1:10,000 dilution and T7 antibody was used at 1:5,000 dilution. HRP-conjugated secondary antibodies (BioRAD) were used at 1:10,000 and blots developed with ChemiGlow detection reagent and visualized using ImageQuant LAS 4000 chemiluminescent camera (GE Healthcare).

In situ UV cross-linking mRNP capture assay

In situ UV cross-linking mRNP capture protocol was adapted from (Piñol-Roma and Dreyfuss 1992). Briefly, cells, grown in 150mm plates, were irradiated with 0.15Jcm^{-2} at 254-nm UV light, on ice, in ice-cold PBS. Cells were then scraped from plates in PBS and pelleted. Cells were lysed in 1ml of ice-cold lysis buffer (20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 0.5% LiDS, 1 mM EDTA and 5 mM DTT), and then sonicated 30s on/30s off for 5 cycles at 4°C. After 10 min incubation at 4°C, the lysates were fractionated by centrifuge (15,000 xg, 4°C, 15 min), 10% sample was stored as input. 100µl per sample oligo(dT)25 magnetic beads were equilibrated in lysis buffer and incubated with lysate overnight at 4°C with rotation. Beads were washed once with 1ml of lysis buffer, twice with Buffer 1 (20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 0.1% LiDS, 1 mM EDTA and 5 mM DTT), twice with Buffer 2 (20 mM Tris-HCl (pH 7.5), 500 mM LiCl, 1 mM EDTA and 5 mM DTT) and twice with Buffer 3 (20 mM Tris-HCl (pH 7.5), 200 mM LiCl, 1 mM EDTA and 5 mM DTT). All washes for 5 min at 4°C. Captured mRNPs were eluted from the beads with 100µl elution buffer (20 mM Tris-HCl (pH 7.5) and 1 mM EDTA) for 3 min at 55°C and treated with 200 U of RNase A in 500 µl of 10× RNase buffer, for 1h at 37 °C. Liberated mRNA binding proteins were mixed 1:1 with SDS PAGE sample buffer (1x LDS sample buffer, 1x reducing agent), and resolved by 3-12% Tris-acetate SDS-PAGE and analyzed by Western blotting.

Supplemental References

Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**: 3439–40.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**: 48 (2009) doi:10.1186/1471-2105-10–48.

Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY. 2019. Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**: 473-490.e26.

Gordon DE, Bond LM, Sahlender DA, Peden AA. 2010. A Targeted siRNA Screen to Identify SNAREs Required for Constitutive Secretion in Mammalian Cells. *Traffic* **11**: 1191–1204.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.

Hug N, Cáceres JF. 2014. The RNA Helicase DHX34 Activates NMD by Promoting a Transition from the Surveillance to the Decay-Inducing Complex. *Cell Rep* **8**: 1845–1856.

Jagannathan S, Nwosu C, Nicchitta C V. 2011. Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation. *Methods Mol Biol* **714**: 301–21.

Jan CH, Williams CC, Weissman JS. 2014. Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. *Science* **346**: 1257521.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Piñol-Roma S, Dreyfuss G. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**: 730–2.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281–2308.

Reid DW, Nicchitta C V. 2012. Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. *J Biol Chem* **287**: 5518–5527.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682.